2024 年度
数理統計学序論１
数理統計学１

大坂公立大学 大学院理学研究科 数学専攻

今野　良彦

2024 年 3 月 18 日

目次

はじめに

この講義録は, 数理統計学の基本的な事項についてまとめたものである. 第 I 部では, 統計推測理論に必要な確率・確率変数・期待値等の概念と性質を丁寧い説明した. 性質についてはできるだけ証明を与えておいた. ただし, 証明に測度論の知識が必要な個所については, 補遺を参照のこと. 第 II 部では, 統計的推測理論の基本的な事項について説明をした. 第 III 部は補遺である.

2024 年 1 月 22 日更新

変更すべきこと

1. $\overset{\text{i.i.d.}}{\sim}$ とかく.
2. $\overset{n\to\infty}{\longrightarrow}$ と書く.
3. 微分は $\dfrac{\mathrm{d}f}{\mathrm{d}x}, \dfrac{\mathrm{d}^n y}{\mathrm{d}x^n}$ と書く.
4. 積分は $\int_a^b f(x)\,\mathrm{d}x$ と書くこと.
5. 分布名: Poisson 分布は $\mathsf{Po}(\lambda)$ と書く. 幾何分布は $\mathsf{Geo}(p)$ と書く. 母数が未知の時は Greek letters にする.
6. 組み合わせ数は ${}^nC_r = \begin{pmatrix} n \\ r \end{pmatrix}$ と書く.
7. 関数は $f : A \to B$ と $f : x \mapsto y$ と書く.
8. 同値等は $\Leftrightarrow, \Rightarrow, \Leftarrow$ と書く.
9. $\mathsf{P}^{\otimes n}$ の記号をどうするか? P^n とするか?
10. 標本空間をどうかくか? i.i.d. の場合, 元の空間を $\mathbb{X}$ と書き, i.i.d. の確率ベクトルの標本空間を $\mathbb{X}^n$ とかくのか?
11. p_X を p^X と書き直す. P_X も P^X に変更.
12. 補遺 D の内積の記号を $(\cdot|\cdot)_m$ に統一する.

記号について

ここに掲げる規則は Jordan Stoyanov (2023, IMS Bulletin **52**, Issue 4, 25-26) の借用である．また, 記号の字体等は Notation list-Cmabridge International Mathematics Qualifications[1]を参考にした.

3 つの基本規則

- 規則 1: ひとつの項目や対象に対してひとつの文字か記号を使用すること.
- 規則 2: 対象の同じグループに対しては同じ字体を使用すること.
- 規則 3: 対象の異なるグループに対しては異なる字体を使用すること.

派生規則

- 確率, 確率分布, 期待値, 分散, 共分散には `\mathsf` を使用する．すなわち, Pr, P, E, Var, Cov のように記す.
- 重要な空間には `\mathbb` を使用する.
- 集合族には `\mathcal` を使用する.
- 略称/略語は「p.m.f.」等のように書く.
- 分布名は `\mathsf` を使用する．たとえば, $\mathsf{Poi}(\lambda)$ で母数 $\lambda > 0$ の Poisson 分布を表す.
- 母数と母数空間等の未知のものは Greek letters を使用する.
- ベクトルと行列は小文字と大文字の `\mathbf` を使用する.
- 内部等の数学の概念は `\mathrm` を使用する．たとえば, 母数空間 $\mathsf{\Theta}$ の内部を $\mathrm{int}\mathsf{\Theta}$ (`\mathrm{int}\mathsf{\Theta}`) と書く.

[1] https://www.cambridgeinternational.org/Images/420009-mathematics-notation-list-.pdf

記号一覧

記号	説明
$\mathbb{N}$	自然数全体の集合
$\mathbb{Z}$	整数全体の集合
$\mathbb{Q}$	有理数全体の集合
$\mathbb{R}$	実数全体の集合
$\mathbb{R}^n$	有限 n 個の実数空間の直積集合
$\mathbb{C}$	複素全体の集合
$\mathbb{R}^n$	有限 n 個の実数空間の直積集合
$\mathrm{Mat}(n, m; \mathbb{R})$	n 行 m 列の実行列全体のなす集合. $\mathrm{Mat}(n, n; \mathbb{R})$ を $\mathrm{Mat}(n; \mathbb{R})$ と書く.
$\mathrm{Sym}(n, \mathbb{R})$	n 行 n 列の対称行列全体のなす集合.
$\mathrm{Sym}^+(n, \mathbb{R})$	n 行 n 列の半正値対称行列全体のなす集合.
$\mathrm{Sym}^{++}(n, \mathbb{R})$	n 行 n 列の正値対称行列全体のなす集合.
$\mathrm{Herm}(n, \mathbb{C})$	n 行 n 列のエルミート行列全体のなす集合.
$\mathrm{Herm}^+(n, \mathbb{C})$	n 行 n 列の半正定値エルミート行列全体のなす集合.
$\mathrm{Herm}^{++}(n, \mathbb{C})$	n 行 n 列の正定値エルミート行列全体のなす集合.
$\mathbf{I}_n$	n 次の単位行列
$\mathbf{0}_n$	$\mathbb{R}^n$ の零ベクトル
$\mathbf{A}^\top$	m 行 n 列の行列 $\mathbf{A}$ の転置行列.
$\mathbf{A}^{-1}$	正方行列 $\mathbf{A}$ の逆行列
$\mathcal{A}, \mathcal{B}, \mathcal{C}$	σ 集合族.

第I部

宴の支度編: 確率・確率変数・確率分布・標本分布

第0章　準備

この章では, 集合・写像・位相の基本的な事項をまとめている. 詳しいは [35, 9, 16, 34] を参照のこと. 第 0.1 節では, 集合の言葉使いと数の集合についての基礎事項をまとめている. 第 0.2 節では, 写像の性質についての基礎事項をまとめている. 第 0.3 節では, 可算集合と非可算集合の基礎事項をまとめている. 第 0.4 節では, 距離空間と位相空間の定義と基礎事項をまとめている. 第 0.5 節では, 実数の拡張である $\mathbb{R}\cup\{-\infty, \infty\}$ における演算の規則についてまとめた. これらのことは, よく理解している場合には, この章を読み飛ばしてもよい.

0.1　集合論の言葉使い

0.1.1　数の集合

数の集合に対して, 以下の記号を用いることにする.

$\mathbb{N}$: 正の整数の集合 (0 を除く)
$\mathbb{Z}$: 整数の集合
$\mathbb{Q}$: 有理数の集合
$\mathbb{R}$: 実数の集合
$\mathbb{C}$: 複素数の集合

とする.

0.1.2　集合の記号

空集合を $\varnothing$ と記し, 集合 X のすべての部分集合をの族を 2^X と記す. すわわち

$$2^X := \big\{E;\, E \subset X\big\}$$

である.

$\mathcal{E}$ を集合 X の部分集合族としたとき

$$\bigcup_{E\in\mathcal{E}} E := \bigl\{x\in X;\ \text{ある } E\in\mathcal{E} \text{ が存在して}, x\in E\bigr\},$$
$$\bigcap_{E\in\mathcal{E}} E := \bigl\{x\in X;\ \text{すべての } E\in\mathcal{E} \text{ に対して}, x\in E\bigr\}$$

と定める. 通常, 添え字集合を用いて, $\mathcal{E} = \{E_\alpha;\ \alpha\in A\} = \{E_\alpha\}_{\alpha\in A}$ と書いたとき

$$\bigcup_{\alpha\in A} E_\alpha; \qquad \bigcap_{\alpha\in A} E_\alpha$$

と表す． $\alpha\neq\beta\,(\alpha,\beta\in A)$ に対して, $E_\alpha\cap E_\beta=\varnothing$ のとき, 集合族 $\{E_\alpha\}_{\alpha\in A}$ は**互いに排反**という. 部分集合族が $\mathbb{N}$ で添え字付けられるとする. すなわち, $\{E_n\}_{n\in\mathbb{N}}$ を考える. このとき, この集合族の**上極限**と**下極限**をそれぞれ

$$\limsup E_n := \bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty} E_n; \qquad \liminf E_n := \bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty} E_n$$

と定める. すると

$$\limsup E_n = \bigl\{x\in X;\ \text{高々可算個の } n \text{ に対して}, x\in E_n\bigr\}$$
$$\liminf E_n = \bigl\{x\in X;\ \text{有限個の } n \text{ を除いたすべてに対して}, x\in E_n\bigr\}$$

と書き直せることがわかる.

部分集合 $E,\,F\subset X$ のとき, $E\setminus F$ によって, それらの差を表す. すなわち

$$E\setminus F = \bigl\{x\in X;\ x\in E \text{ かつ } x\notin F\bigr\}$$

である. 集合 $E\subset X$ の**補集合**を E^c で表す. すなわち

$$E^c = X\setminus E$$

である. このとき, De Morgan の法則

$$\Bigl(\bigcup_{\alpha\in A} E_\alpha\Bigr)^c = \bigcap_{\alpha\in A} E_\alpha^c; \qquad \Bigl(\bigcap_{\alpha\in A} E_\alpha\Bigr)^c = \bigcup_{\alpha\in A} E_\alpha^c$$

が成立する.

集合 X と Y の**直積**を $X\times Y$ と書く. すなわち

$$X\times Y = \bigl\{(x,\,y):\ x\in X,\, y\in Y\bigr\}$$

である. 有限個の集合 $X_1,\,X_2,\,\ldots,\,X_n$ の直積

$$X_1\times X_2\times\cdots\times X_n\ \Bigl(\prod_{j=1}^{n} X_j \text{ とも表す}\Bigr)$$

も同様に定義する. 特に, $X_1=X_2=\cdots=X_n=X$ のとき, $X_1,\,X_2,\,\ldots,\,X_n$ を X^n とも書く.

0.2　写像

0.2.1　写像の定義と性質

定義 0.1. (1) X, Y を空でない集合とする．このとき

- X の任意の元に対して
- その元に対応する Y のある元がただ 1 つ与えられる

とする．このことを

$$f: X \to Y$$

と表し, f を X から Y への**写像**という．
(2) X を写像 f の**定義域**, Y を写像 f の**終域**という．

注意 0.2. (1) X, Y を空でない集合とし, $y_0 \in Y$ を 1 つ選んで固定しておく．このとき, 写像 $f \to Y$ を

$$f(x) = y_0 \qquad (x \in X)$$

により定める．この f を**定値写像**という．
(2) X を空でない集合とし, $A \subset X$ とする．このとき, X から $\{0, 1\}$ への写像 $\mathbb{1}_A: X \to \{0, 1\}$ を

$$\mathbb{1}_A(x) = \begin{cases} 1 & (x \in A) \\ 0 & (x \in X \setminus A) \end{cases}$$

により定める．この $\mathbb{1}_A$ を**定義写像** (または**特性写像**) という．
(3) X, Y を空でない集合とし, $X \subset Y$ とする．このとき, 写像 $\imath: X \to Y$ を

$$\imath(x) = x \qquad (x \in X)$$

により定める．$\imath$ を**包含写像**という．特に, $X = Y$ のとき, $\imath$ を id_X と表し, **恒等写像**という．
(4) X, Y を空でない集合, $f: X \to Y$ を写像とし, $A \subset X$, $A \neq \varnothing$ とする．このとき, 写像 $f\big|_A: A \to Y$ を

$$f\big|_A(x) = f(x) \qquad (x \in A)$$

により定める．この $f\big|_A$ を f の A への**制限** (または**制限写像**) という．

定義 0.3. f, g を写像とする．f と g の定義域が等しく，f と g の値域も等しく，さらに f, g の定義域の任意の元 x に対して，$f(x) = g(x)$ がなりたつとき，$f = g$ と表し，写像 f と g は**等しい**という．また，写像 f と g が等しくないとき，$f \neq g$ と表す．

定義 0.4. X, Y を空でない集合，$f : X \to Y$ を写像とする．このとき，$G(f) \subset X \times Y$ を

$$G(f) := \bigl\{\bigl(x, f(x)\bigr);\, x \in X\bigr\}$$

により定め，これを写像 f の**グラフ**という．

0.2.2　像と逆像

定義 0.5. X, Y を空でない集合とし，$f : X \to Y$ を写像とする．
(1) X の部分集合 A に対して，$\bigl\{f(x);\, x \in A\bigr\}$ を写像 f による部分集合 A の**像**といい，$f(A)$ と書く．$f(X)$ を写像 f の**値域**という．
(2) Y の部分集合 B に対して，写像 f による B の**逆像** $f^{-1}(B)$ を

$$f^{-1}(B) = \{x \in X;\, f(x) \in B\}$$

で定める．略記して，$\{f \in B\}$ とも書く．
(3) $f(X) = Y$ のとき，f は**全射**という．
(4) $x, x' \in X$, $x \neq x'$ ならば，$f(x) \neq f(x')$ であるとき，f は**単射**という．
(5) f が全射かつ単射のとき，**全単射**という．

命題 0.6. X, Y を空でない集合とし，$f : X \to Y$ を写像とし，$A, A_1, A_2 \subset X$, $B, B_1, B_2 \subset Y$ とする．このとき，次の (1) ～ (10) がなりたつ．
(1) $A_1 \subset A_2$ ならば，$f(A_1) \subset f(A_2)$.
(2) $f(A_1 \cup A_2) = f(A_1) \cup f(A_2)$.
(3) $f(A_1 \cap A_2) \subset f(A_1) \cap f(A_2)$.
(4) $f(A_1 \setminus A_2) \supset f(A_1) \setminus f(A_2)$.
(5) $B_1 \subset B_2$ ならば，$f^{-1}(B_1) \subset f^{-1}(B_2)$.
(6) $f^{-1}(B_1 \cup B_2) = f^{-1}(B_1) \cup f^{-1}(B_2)$.
(7) $f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2)$.
(8) $f^{-1}(B_1 \setminus B_2) = f(B_1) \setminus f(B_2)$.
(9) $f^{-1}\bigl(f(A)\bigr) \supset A$.
(10) $f\bigl(f^{-1}(B)\bigr) \subset B$.

Proof. 証明は [35, pp.32-33] を参照のこと．□

命題 0.6(3)(4)(9)(10) において，等号が成立しない例をあげる．

注意 0.7. 写像 $f:\mathbb{R}\to\mathbb{R}$ を

$$f(x)=x^2 \qquad (x\in\mathbb{R})$$

により定める.
(1) $A_1=(-1,\,0]$, $A_2=[0,\,1)$ とすると

$$f(A_1\cap A_2)=\{0\}; \qquad f(A_1)\cap f(A_2)=[0,\,1);$$
$$f(A_1\setminus A_2)=(0,\,1); \qquad f(A_1)\setminus f(A_2)=\varnothing$$

となる.
(2) $A=[0,\,1)$, $B=(-1,\,1)$ とすると

$$f^{-1}\bigl(f(A)\bigr)=(-1,\,1); \qquad f\bigl(f^{-1}(B)\bigr)=[0,\,1)$$

となる. □

命題 0.8. X, Y を空でない集合とし, $f:X\to Y$ を写像とし, A, A_1, $A_2\subset X$, $B\subset Y$ とする. このとき, 次の (1) $\sim$ (10) がなりたつ.
(1) f が単射ならば, $f(A_1\cap A_2)=f(A_1)\cap f(A_2)$.
(2) f が単射ならば, $f(A_1\setminus A_2)=f(A_1)\setminus f(A_2)$.
(3) f が単射ならば, $f^{-1}\bigl(f(A)\bigr)=A$.
(4) f が全射ならば, $f\bigl(f^{-1}(B)\bigr)=B$.

Proof. 証明は [35, p.37] を参照のこと. □

0.2.3　合成写像

定義 0.9. X, Y, Z を空でない集合とし, $f:X\to Y$, $g:Y\to Z$ を写像とする. このとき, X から Z への写像 $g\circ f$ を

$$(g\circ f)(x)=g\bigl(f(x)\bigr) \qquad (x\in X)$$

により定める. $g\circ f$ を f と g の**合成** (または**合成写像**) という.

命題 0.10. X, Y, Z, W を空でない集合, $f:X\to Y$, $g:Y\to Z$, $h:Z\to W$ を写像とする. このとき

$$h\circ(g\circ f)=(h\circ g)\circ f$$

である. 特に, $h\circ(g\circ f)$ と $(h\circ g)\circ f$ をともに $h\circ g\circ f$ と表す.

Proof. 証明は [35, p.40] を参照のこと. □

命題 0.11. X, Y, Z を空でない集合, $f : X \to Y$, $g : Y \to Z$ を写像とする．このとき, (1), (2) がなりたつ．
(1) f, g が全射ならば, $g \circ f$ も全射．
(2) f, g が単射ならば, $g \circ f$ も単射．
(3) f, g が全単射ならば, $g \circ f$ も全単射．

Proof. 証明は [35, p.40] を参照のこと． □

0.3　可算集合・非可算集合・濃度

定義 0.12. (1) X, Y を空でない集合とする．X から Y への全単射が存在するとき, X と Y は**濃度が等しい**という．
(2) 有限の元から構成される集合を**有限**集合という．無限の元から構成される集合を**無限**集合という．
(3) 自然数全体の集合 $\mathbb{N}$ と濃度が等しい集合を**可算**集合という．
(4) 集合 X が有限集合または可算集合のとき, X は**高々可算集合**という．
(5) 可算集合でない無限集合を**非可算**集合という．

命題 0.13. (1) 整数全体の集合 $\mathbb{Z}$ と有理数全体の集合 $\mathbb{Q}$ は可算集合である．
(2) 自然数 d に対して, $\mathbb{N}^d$, $\mathbb{Z}^d$, $\mathbb{Q}^d$ は可算集合である．
(3) 実数全体の集合 $\mathbb{R}$ から $\mathbb{N}$ の冪集合 $2^{\mathbb{N}}$ は濃度が等しい．
(4) X を空でない集合とする．このとき, X から X の冪集合 2^X への全射は存在しない．
(5) d を自然数とする．このとき, $\mathbb{R}$ と $\mathbb{R}^d$ は濃度が等しい．

Proof. 証明は [35, pp.68-80] を参照のこと． □

0.4　距離と位相

0.4.1　距離空間

定義 0.14. $\mathbb{X}$ を空でない集合とする．写像 $d : X \times \mathbb{X} \to \mathbb{R}$ が次の 3 条件をみたすとき, d を $\mathbb{X}$ 上の**距離関数**といい, 組 $(\mathbb{X}, d)$ を**距離空間**という．
(1) 任意の $x, y \in \mathbb{X}$ に対して, $d(x, y) \geq 0$ である．特に, $d(x, y) = 0 \Leftrightarrow x = y$ である．
(2) 任意の $x, y \in \mathbb{X}$ に対して, $d(x, y) = d(y, x)$.
(3) 任意の $x, y, z \in \mathbb{X}$ に対して, $d(x, z) \leq d(x, y) + d(y, z)$.

注意 0.15. d を自然数とし, $\mathbb{X}=\mathbb{R}^d$ とする. 距離関数の例として, Euclid ノルム

$$|\boldsymbol{x}|_{2,\,d}=\sqrt{\sum_{j=1}^{d}x_j^2}\qquad(\boldsymbol{x}=(x_1,\,x_2,\,\ldots,\,x_d)^\top\in\mathbb{R}^d)$$

から定まる Euclid の距離 $d(\boldsymbol{x},\,\boldsymbol{y})=|\boldsymbol{x}-\boldsymbol{y}|_{2,\,d}\,(\boldsymbol{x},\,\boldsymbol{y}\in\mathbb{R}^d)$ である. □

定義 0.16. $(\mathbb{X},\,d)$ を距離空間とする.
(1) $x\in\mathbb{X}$, $r>0$ に対して

$$B(x;\,r)=\{y\in\mathbb{X};\,d(x,\,y)<r\}$$

を, 中心 x, 半径 r の**開球**という.
(2) $\mathbb{X}$ の部分集合 O が**開集合**であるとは, 任意の $x\in O$ に対して, $\exists r>0$ を選んで, $B(x;\,r)\subset A$ とできることである.
(3) 開集合の補集合を**閉集合**という.
(4) A が**有界**であるとは, $\exists x\in\mathbb{X}$ と $\exists r>0$ を選んで $A\subset B(x;\,r)$ とできることである.

定義 0.17. $(\mathbb{X},\,d)$ を距離空間とし, $\{x_n\}_{n\in\mathbb{N}}$ を $\mathbb{X}$ の点列とする.
(1) $x\in\mathbb{X}$ とする. 点列 $\{x_n\}_{n\in\mathbb{N}}$ が x に収束するとは

$$\lim_{n\to\infty}d(x_n,\,x)=0$$

がなりたつことである.
(2) 点列 $\{x_n\}_{n\in\mathbb{N}}$ が Cauchy 列であるとは, 任意の $\epsilon>0$ に対して, ある $N\in\mathbb{N}$ が存在して, $m\geq N$, $n\geq N$ ならば, $d(x_m,\,x_n)<\epsilon$ となることである.
(3) Cauchy 列が常に収束するとき, 距離空間 $\mathbb{X}$ は**完備**であるという.

定義 0.18. $(\mathbb{X},\,d_X)$ と $(\mathbb{Y},\,d_Y)$ を距離空間とする. 写像 $f:\mathbb{X}\to\mathbb{Y}$ が点 $x\in\mathbb{X}$ で**連続**であるとは, 任意の $\epsilon>0$ に対して, ある $\delta>0$ が存在して, 任意の $x'\in\mathbb{X}$ に対して

$$d_X(x,\,x')<\delta\;\Rightarrow\;d_Y(f(x),\,f(x'))<\epsilon$$

をみたすときをいう. 任意の $x\in\mathbb{X}$ で f が連続ならば, 単に f は連続という.
(2) 写像 f が一様連続であるとは, 任意の $\epsilon>0$ に対して, ある $\delta>0$ が存在して, 任意の $x,\,x'\in\mathbb{X}$ に対して

$$d_X(x,\,x')<\delta\;\Rightarrow\;d_Y(f(x),\,f(x'))<\epsilon$$

がなりたつときをいう.

定義 0.19. X を集合, $(\mathbb{Y}, d)$ を距離空間とする. $f, f_1, f_2, \ldots, f_n \ldots$ を X から $\mathbb{Y}$ への写像の列とする.
(1) $\{f_n\}_{n\in\mathbb{N}}$ が f に**各点収束**するとは, 任意の $x \in X$ に対して

$$\lim_{n\to\infty} d(f_n(x),\, f(x)) = 0$$

であることをいう.
(2) X の部分集合 $A\,(\neq \varnothing)$ に対して, $\{f_n\}_{n\in\mathbb{N}}$ が f に A 上で**一様収束**するとは

$$\lim_{n\to\infty} \sup_{x\in A} d(f_n(x),\, f(x)) = 0$$

がなりたつときをいう.

0.4.2　位相空間

定義 0.20. $\mathbb{X}$ を空でない集合とする. $\mathbb{X}$ 上の集合族 $\mathcal{O} \subset 2^{\mathbb{X}}$ が次の条件 (1) ～ (3) をみたつとき, $\mathcal{O}$ を $\mathbb{X}$ の**位相**という. 組 $(\mathbb{X}, \mathcal{O})$ を**位相空間**という. $\mathcal{O}$ の元を $\mathbb{X}$ の開集合という.
(1) $\varnothing, \mathbb{X} \in \mathcal{O}$.
(2) $A, B \in \mathcal{O}$ ならば, $A \cap B \in \mathcal{O}$.
(3) $\{A_\lambda\}_{\lambda\in\Lambda}$ を　$\mathcal{O}$ の元から成る集合族としたとき

$$\bigcup_{\lambda\in\Lambda} A_\lambda \in \mathcal{O}$$

である.

定義 0.21. $(\mathbb{X}, \mathcal{O})$ を位相空間とする.
(1) $A \subset \mathbb{X}$ とする. A のすべての部分開集合からなる族を $\{O_\lambda\}_{\lambda\in\Lambda}$ とする. このとき, A の**内部** A° を

$$A^\circ = \bigcup_{\lambda\in\Lambda} O_\lambda$$

により定める. また, A° の元を A の**内点**という.
(2) $A \subset \mathbb{X}$ とする. A を含むすべての $\mathbb{X}$ の部分閉集合から成る族を $\{C_\lambda\}_{\lambda\in\Lambda}$ とする. このとき, A の**閉包** $\mathrm{cl}(A)$ を

$$\mathrm{cl}(A) = \bigcap_{\lambda\in\Lambda} C_\lambda$$

により定める. また, $\mathrm{cl}(A)$ の元を A の**触点**という.
(3) $A \subset \mathbb{X}$ とする. $\partial A = \mathrm{cl}(A) \setminus A^\circ$ とし, A の**境界**という.

(4) $A \subset \mathbb{X}$ とする. $x \in A^\circ$ のとき, A は x の**近傍**という.
(5) $A \subset \mathbb{X}$ とする. $\mathcal{O}_A = \{O \cap A; O \in \mathcal{O}\}$ とすると, $\mathcal{O}_A$ は A の位相となる. これを A 上の**相対位相**という.
(6) $A \subset \mathbb{X}$ とする. A が $\mathbb{X}$ において**稠密**であるとは, $\mathrm{cl}(A) = \mathbb{X}$ となることである.
(7) $\mathbb{X}$ が**可分**であるとは, $\mathbb{X}$ の高々可算な部分集合 A で, $\mathbb{X}$ において稠密なものが存在するときである.
(8) $\mathbb{X}$ が**コンパクト**であるとは, $\mathbb{X}$ に任意の**開被覆** $\{O_\lambda\}_{\lambda\in\Lambda}$, すなわち, $\mathbb{X} = \bigcup_{\lambda\in\Lambda} O_\lambda$ に対して, Λ の有限部分集合 Λ_0 が存在して, $\mathbb{X} = \bigcup_{\lambda\in\Lambda_0} O_\lambda$ となるときをいう.

定義 0.22. $\mathbb{X}$, $\mathbb{Y}$ を位相空間とする.
(1) 写像 $f : \mathbb{X} \to \mathbb{Y}$ が**連続**であるとは, $\mathbb{Y}$ の任意の開集合 B に対して, $f^{-1}(B)$ は $\mathbb{X}$ の開集合となるときをいう. この定義は距離空間の連続写像の定義と整合している.
(2) 全単射な連続写像 $f : \mathbb{X} \to \mathbb{Y}$ で逆写像 f^{-1} も連続のとき, $\mathbb{X}$ と $\mathbb{Y}$ は**同相**であるという. また, f を**同相写像**という.

命題 0.23. (1) $\mathbb{X}$, $\mathbb{Y}$ を位相空間とし, $f : \mathbb{X} \to \mathbb{Y}$ を連続写像とする. $\mathbb{X}$ のコンパクト部分集合 A に対して, $f(A)$ は $\mathbb{Y}$ のコンパクト部分集合となる.
(2) $(\mathbb{X}, d_X)$ をコンパクト距離空間, $(\mathbb{Y}, d_Y)$ を距離空間とする. $\mathbb{X}$ から $\mathbb{Y}$ への連続写像は一様連続である.

Proof. 証明は省略. □

0.5　$\mathbb{R}$ の拡張

集合 $\mathbb{R} \cup \{-\infty, \infty\}$ を $\overline{\mathbb{R}}$ と書くことにする. $\overline{\mathbb{R}}$ での演算を以下のように定める.

- 実数に関する演算は通常通り.
- $a \in \mathbb{R}$ に対して, $-\infty < a < \infty$.
- $a \in \mathbb{R}$ に対して, 以下のように定める.
  - $a + (\pm\infty) = \pm\infty$, $\pm\infty + a = \pm\infty$.
  - $a > 0$ のとき, $a \times (\pm\infty) = \pm\infty$, $\pm\infty \times a = \pm\infty$.
  - $a + (\pm\infty) = \pm\infty$, $\pm\infty + a = \pm\infty$.
  - $a < 0$ のとき, $a \times (\pm\infty) = \mp\infty$, $\pm \times a = \mp\infty$.
  - $0 \times (\pm\infty) = 0$, $\pm\infty \times 0 = 0$.

  - $0/(\pm\infty) = 0$.

- $(\pm\infty) + \pm(\infty) = \pm\infty$, $(\pm\infty) - (\mp\infty) = \pm\infty$.
- $(\pm\infty) \times (\pm\infty) = +\infty$, $(\pm\infty) \times (\mp\infty) = -\infty$.
- $(\pm\infty)/(\pm\infty)$, $(\pm\infty)/(\mp\infty)$ は定義されないものとする.
- $|\pm\infty| = +\infty$.
- $\overline{\mathbb{R}}$ の部分集合 $A\,(\neq \varnothing)$ に対して, 上限 $\sup A$ と下限 $\inf A$ が定まり, $\overline{\mathbb{R}}$ に値をとる.
- $a,\, b \in \overline{\mathbb{R}}$ について, 閉区間 $[a,\, b]$ と開区間 $(a,\, b)$ を

$$[a,\, b] = \{x \in \overline{\mathbb{R}};\, a \leq x \leq b\}, \qquad (a,\, b)\{x \in \overline{\mathbb{R}};\, a < x < b\}$$

  により定める.

0.6　章末注釈と参考文献

第 0.1 節は [34] を借用した. 第 0.2 節は [35, pp.26 – 29] を借用した. 第 0.3 節は [34] を借用した. 第 0.4 節は [34] と [35] を参照した. 第 0.5 節は [34, pp.18 – 19] を借用した.

第1章　確率・確率変数の基本事項

この章では, 統計的推測理論を理解する上で必要な確率論の基礎事項を簡潔に述べる. したがって, 証明はしない場合がある. また, 測度論の内容には立ち入らない. 数学的な厳密性を犠牲にして, 直観的な理解を目指す記述を心掛けた. 節 1.1 では, 確率に関わる基本的な概念を導入する. 節 1.2 では, 抽象的な確率空間と数量の世界を結ぶ確率変数を導入し, 確率変数・確率分布・累積分布関数の基本的な性質を説明する. 節 1.3 では, 累積分布関数の逆関数のようなものである分位点関数を導入する. 節 1.4 では, 1 変数確率変数の確率モデルを説明する. 節 1.5 では, 2 次元確率変数の同時分布, 周辺分布, 条件付き分布を説明する. 節 1.6 では, 多変数確率変数の確率モデルを導入する. 節 1.7 では, 正規分布から誘導される重要な分布であるカンマ分布, χ^2 分布, t 分布, F 分布を導入する.

1.1　基礎的な確率規則

確率論はランダムな現象を扱う数学理論である. 確率論で扱う行為を**試行**という. 試行のあり得る結果すべてを集めた集合を**標本空間**といい, Ω と記すことにする. Ω の部分集合[1]を**事象**という. 事象には標本空間 Ω と空事象 $\varnothing$(何も起こらないという事象) も含める. 事象をすべて集めた集合族を $\mathcal{A}$ と記す. $\mathcal{A}$ は以下で述べる σ **加法性** (完全加法性) をみたすことにする.

定義 1.1. Ω を空でない集合とし, $\mathcal{A}$ を Ω の部分集合族とする. $\mathcal{A}$ が次の 3 条件をみたすとき, σ **加法族**と呼ばれる.

(1) $\Omega \in \mathcal{A}$.

(2) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$.

(3) $A_n \in \mathcal{A}\,(n = 1, 2, \ldots) \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

[1] Ω が可算集合ならば, σ 加法族をすべての部分集合としてよいが, Ω が連続濃度のときには, すべての部分集合とすると不都合が生じることになる. これについては $[2, p.21]$ を参照のこと.

ただし $A^c = \{\omega \in \Omega;\, \omega \notin A\}$ である. Ω と $\mathcal{A}$ の組 $(\Omega, \mathcal{A})$ を**可測空間**と呼ぶ

補題 1.2. $(\Omega, \mathcal{A})$ を可測空間とする. このとき以下が成立する.

(1) $\varnothing \in \mathcal{A}$.

(2) $A_n \in \mathcal{A}\,(n = 1, 2, \ldots) \Rightarrow \bigcap_{n=1}^{\infty} A_n \in \mathcal{A}$.

Proof. (1) $\varnothing = \Omega^c \in \mathcal{A}$.
(2) $A_n^c \in \mathcal{A}\,(n = 1, 2, \ldots)$ と De Morgan の法則から

$$\bigcap_{n=1}^{\infty} A_n = \bigcap_{n=1}^{\infty} (A_n^c)^c = \left(\bigcup_{n=1}^{\infty} A_n^c\right)^c \in \mathcal{A}.$$

最後は定義 1.1(3) を用いた. □

注意 1.3. $\mathcal{C}$ を Ω の部分集合族とする. $\mathcal{C}$ は σ 加法性をみたしてなくともよい. このとき, 集合族 $\sigma[\mathcal{C}]$ を

$$\sigma[\mathcal{C}] := \bigcap \left\{\mathcal{A};\, \mathcal{A} \supset \mathcal{C},\, \mathcal{A} \text{ は } \sigma \text{ 加法族}\right\}$$

で定める. すると $\sigma[\mathcal{C}]$ は σ 加法族となることを確かめることができる. さらに $\mathcal{G}$ を $\mathcal{A}$ を含む σ 加法族としたとき

$$\sigma[\mathcal{C}] \subseteq \mathcal{G}$$

となることが直にわかる. すなわち $\sigma[\mathcal{C}]$ は $\mathcal{C}$ を含む最小 (包含関係の意味) の σ 加法族となる. □

問 1.1. 集合族 $\mathcal{C}$ に対して

$$\sigma[\mathcal{C}] = \bigcap \left\{\mathcal{A};\, \mathcal{A} \supset \mathcal{C},\, \mathcal{A} \text{ は } \sigma \text{ 加法族}\right\}$$

と定める. このとき, 以下を証明せよ.
(1) $\sigma[\mathcal{C}]$ は σ 集合族になることを示せ.
(2) $\mathcal{G}$ を $\mathcal{C}$ を含む任意の σ 集合族としたき

$$\sigma[\mathcal{C}] \subset \mathcal{G}$$

を示せ.

定義 1.4. $\Omega = \mathbb{R}$ とし

$$\mathcal{O} = \{O \subset \mathbb{R};\, O \text{ は } \mathbb{R} \text{ の開集合 }\}$$

とする．$\sigma[\mathcal{O}]$ を $\mathbb{R}$ の **Borel 集合族**と呼び，$\mathcal{B}(\mathbb{R})$ と記す．また

$$\mathcal{C} = \{(-\infty, x) \subset \mathbb{R};\, x \in \mathbb{R}\}$$

とする．このとき $\mathcal{C}$ は $\mathcal{O}$ の真部分集合であるが $\sigma[\mathcal{C}] = \sigma[\mathcal{O}]$ となる[2]．

注意 1.5. Ω が高々可算集合のときは，$\mathcal{F} = 2^{\Omega}$ と取る．ただし 2^{Ω} は Ω のすべての部分集合からなる集合族で冪(べき)集合という． □

定義 1.6. $(\Omega, \mathcal{A})$ を可測空間とする．$\mathcal{A}$ 上の関数

$$\mathsf{Pr} : \mathcal{A} \ni A \mapsto \mathsf{Pr}(A) \in [0, 1]$$

が次の 2 条件をみたすとき，$(\Omega, \mathcal{A})$ 上の**確率測度**[3]と呼ばれる．

(1) $\mathsf{Pr}(\Omega) = 1$ である．

(2) 互いに排反[4]な事象列 $A_n \in \mathcal{A}\,(n = 1, 2, \ldots)$ に対して

$$\mathsf{Pr}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathsf{Pr}(A_n) := \lim_{N\to\infty} \sum_{n=1}^{N} \mathsf{Pr}(A_n)$$

をみたす．

これらの 3 つの組 $(\Omega, \mathcal{A}, \mathsf{Pr})$ を**確率空間**という．

補題 1.7. $(\Omega, \mathcal{A}, \mathsf{Pr})$ を確率空間とする．このとき以下が成立する．

(1) $\mathsf{Pr}(\varnothing) = 0$ である．

(2) $A \in \mathcal{A}$ に対して，$\mathsf{Pr}(A^c) = 1 - \mathsf{Pr}(A)$ となる．

(3) $N \in \mathbb{N}$ とする．$\{A_n\}_{n=1}^{N} \subset \mathcal{A}$ が互いに排反ならば

$$\mathsf{Pr}\left(\bigcup_{n=1}^{N} A_n\right) = \sum_{n=1}^{N} \mathsf{Pr}(A_n)$$

となる．

(4) $A, B \in \mathcal{A}, A \subset B \Rightarrow \mathsf{Pr}(B \setminus A) = \mathsf{Pr}(B) - \mathsf{Pr}(A)$ となる．よって，$A \subset B \;\Rightarrow\; \mathsf{Pr}(A) \leq \mathsf{Pr}(B)$ が成立する．

[2] $\sigma[\mathcal{C}] \subset \sigma[\mathcal{O}]$ は明らかであるが，逆の包含関係も示すことができる．この逆の証明は，定理 1.16 と Euclid 位相の事実を用いて証明ができる．

[3] 簡単に Ω 上の確率測度ともいう．

[4] $m \neq n$ ならば，$A_m \cap A_n = \varnothing$ が成立していること．

(5) (Boole の定理/ユニオン・バウンド) $A_n \in \mathcal{A}\,(n = 1, 2, \ldots)$ に対して

$$\mathsf{Pr}\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mathsf{Pr}(A_n)$$

となる.

(6) $A_n \in \mathcal{A}$ が $A_n \subset A_{n+1}\,(n = 1, 2, \ldots)$ をみたすならば

$$\lim_{n\to\infty} \mathsf{Pr}(A_n) = \mathsf{Pr}\left(\bigcup_{n=1}^{\infty} A_n\right)$$

となる.

(7) $A_n \in \mathcal{A}$ が $A_n \supset A_{n+1}\,(n = 1, 2, \ldots)$ をみたすならば

$$\lim_{n\to\infty} \mathsf{Pr}(A_n) = \mathsf{Pr}\left(\bigcap_{n=1}^{\infty} A_n\right)$$

となる.

Proof. (1) $F_1 := \Omega,\ F_n := \varnothing\,(n \geq 2)$ とおくと $\{F_n\}_{n=1}^{\infty}$ は互いに排反な事象列となる. 定義 1.6(2) を用いると

$$\mathsf{Pr}(\Omega) = \mathsf{Pr}\left(\bigcup_{n=1}^{\infty} F_n\right) = \sum_{n=1}^{\infty} \mathsf{Pr}(F_n) = \mathsf{Pr}(\Omega) + \sum_{n=2}^{\infty} \mathsf{Pr}(\varnothing)$$

を得る. よって

$$\mathsf{Pr}(\varnothing) = 0$$

がわかる.

(2) $F_1 := A,\ F_2 := A^c,\ F_n = \varnothing\,(n \geq 3)$ とおく. すると $\bigcup_{n=1}^{\infty} F_n = \Omega$ かつ $\{F_n\}_{n=1}^{\infty}$ は互いに排反な事象列となるので, 定義 1.6(1), (2) を用いると

$$\begin{aligned}1 = \mathsf{Pr}(\Omega) = \mathsf{Pr}\left(\bigcup_{n=1}^{\infty} F_n\right) &= \sum_{n=1}^{\infty} \mathsf{Pr}(F_n) = \mathsf{Pr}(A) + \mathsf{Pr}(A^c) + \sum_{n=3}^{\infty} \mathsf{Pr}(\varnothing)\\ &= \mathsf{Pr}(A) + \mathsf{Pr}(A^c) \qquad (\because (1))\end{aligned}$$

がわかる. よって, (2) は示せた.

(3) $F_i := A_i\,(i = 1, 2, \ldots, N)$ と $F_i = \varnothing\,(i \geq N+1)$ とおくと $\{F_n\}_{n=1}^{\infty}$ は互いに排反な事象列で $\bigcup_{n=1}^{\infty} F_n = \bigcup_{i=1}^{N} A_i$ となる. 定義 1.6(2) を用いると

$$\begin{aligned}\mathsf{Pr}\left(\bigcup_{i=1}^{N} A_i\right) = \mathsf{Pr}\left(\bigcup_{n=1}^{\infty} F_n\right) = \sum_{i=1}^{\infty} \mathsf{Pr}(F_i) &= \sum_{i=1}^{N} \mathsf{Pr}(A_i) + \sum_{i=N+1}^{\infty} \mathsf{Pr}(\varnothing)\\ &= \sum_{i=1}^{N} \mathsf{Pr}(A_i)\end{aligned}$$

を得る．よって, (3) は示された.
(4) $B \setminus A = B \cap A^c$ かつ $B = (B \cap A^c) \cup A$ である．このことに注意して, $F_1 := B \cap A^c$, $F_2 = A$, $F_n = \varnothing\,(n \geq 3)$ とおくと $\{F_n\}_{n=1}^{\infty}$ は互いに排反な事象列で $\bigcup_{n=1}^{\infty} F_n = B$ となる．定義 1.6(2) を用いると

$$\begin{aligned}\mathsf{Pr}(B) &= \mathsf{Pr}\left(\bigcup_{n=1}^{\infty} F_n\right) = \sum_{n=1}^{\infty} \mathsf{Pr}(F_n) = \mathsf{Pr}(B \cap A^c) + \mathsf{Pr}(A) + \sum_{n=3}^{\infty} \mathsf{Pr}(\varnothing) \\ &= \mathsf{Pr}(B \cap A^c) + \mathsf{Pr}(A)\end{aligned}$$

がわかる．よって, (4) は示せた.
(5) $\mathsf{Pr}(A_1 \cup A_2) \leq \mathsf{Pr}(A_1) + \mathsf{Pr}(A_2)$ を示せばよい．まず, $A_1 \cup A_2 = A_1 \cup \Big(A_2 \setminus (A_1 \cap A_2\Big)$ かつ $A_1 \bigcap \Big(A_2 \setminus (A_1 \cap A_2)\Big) = \varnothing$ であることに注意する．(3) から

$$\begin{aligned}\mathsf{Pr}\big(A_1 \cup A_2) &= \mathsf{Pr}(A_1) + \mathsf{Pr}\Big(A_2 \setminus (A_1 \cap A_2)\Big) \\ &= \mathsf{Pr}(A_1) + \mathsf{Pr}(A_2) - \mathsf{Pr}(A_1 \cap A_2) \\ &\qquad \big(\because A_2 \supset A_1 \cap A_2 \text{ なので (4) を用いた}\big) \\ &\leq \mathsf{Pr}(A_1) + \mathsf{Pr}(A_2) \qquad \big(\because \mathsf{Pr}(A_1 \cap A_2) \geq 0\big)\end{aligned}$$

がわかる.
(6) $F_1 := A_1$, $F_2 = A_2 \setminus A_1$, $\ldots$, $F_{n+1} := A_{n+1} \setminus A_n$ とおく．$\{F_n\}_{n=1}^{\infty}$ は互いに排反であり

$$A_n = \bigcup_{i=1}^{n} F_i \ \Rightarrow\ A := \bigcup_{i=1}^{\infty} F_i = \bigcup_{n=1}^{\infty} A_n$$

となる．よって

$$\begin{aligned}\mathsf{Pr}(A) &= \mathsf{Pr}\left(\bigcup_{i=1}^{\infty} F_i\right) = \sum_{i=1}^{\infty} \mathsf{Pr}(F_i) \quad (\because \text{定義 } 1.6(2) \\ &= \lim_{n\to\infty} \sum_{i=1}^{n} \mathsf{Pr}(F_i) = \lim_{n\to\infty} \mathsf{Pr}\left(\bigcup_{i=1}^{n} F_i\right) \quad (\because (3)) \\ &= \lim_{n\to\infty} \mathsf{Pr}(A_n)\end{aligned}$$

を得る．よって, (6) は示せた.
(7) $A = \bigcap_{n=1}^{\infty} A_n$ とおく．すると $\{A_1 \setminus A_n\}_{n=1}^{\infty}$ は $A_1 \setminus A_2 \subset A_1 \setminus A_3 \subset A_1 \setminus A_n \subset \cdots$ となる．このことに注意して分配法則を用いると

$$\bigcup_{n=1}^{\infty} (A_1 \setminus A_n) = \bigcup_{n=1}^{\infty} (A_1 \cap A_n^c) = A_1 \cap \left(\bigcap_{n=1}^{\infty} A_n\right)^c = A_1 \cap A^c = A_1 \setminus A$$

となる. 次に (4) と (6) を用いると

$$\begin{aligned}\mathsf{Pr}(A_1) - \mathsf{Pr}(A) &= \mathsf{Pr}(A_1 \setminus A) \quad (A_1 \supset A\ なので, (4)\ を用いた) \\ &= \lim_{n\to\infty} \mathsf{Pr}(A_1 \setminus A_n) \quad (\because (6)\ を用いた) \\ &= \lim_{n\to\infty} \Big(\mathsf{Pr}(A_1) - \mathsf{Pr}(A_n)\Big) \\ &= \mathsf{Pr}(A_1) - \lim_{n\to\infty} \mathsf{Pr}(A_n)\end{aligned}$$

がわかる. □

注意 1.8. 補題 1.7(6)(7) において, $\{\mathsf{Pr}(A_n)\}_{n=1}^{\infty}$ は有界な非減少列もしくは非増加列なので, 左辺は必ず収束することに注意せよ.

定義 1.9. $(\Omega, \mathcal{A}, \mathsf{Pr})$ を確率空間とし, $A, B \in \mathcal{A}$ とする.

(1) (独立性):

$$A\ と\ B\ は\textbf{独立} \Leftrightarrow \mathsf{Pr}(A \cap B) = \mathsf{Pr}(A)\mathsf{Pr}(B).$$

(2) (条件付き確率)] $\mathsf{Pr}(B) > 0$ のとき, B を与えたときの A の**条件付き確率** $\mathsf{Pr}(A|\,B)$ を

$$\mathsf{Pr}(A|\,B) := \frac{\mathsf{Pr}(A \cap B)}{\mathsf{Pr}(B)}$$

で定める.

注意 1.10. (1) $A, B \in \mathcal{A}$ とし, $\mathsf{Pr}(B) > 0$ とする. A と B が独立のとき

$$\mathsf{Pr}(A|\,B) = \mathsf{Pr}(A)$$

が成立する.
(2) $\mathsf{Pr}(B) > 0$ のとき

$$\mathsf{Pr}(A \cap B) = \mathsf{Pr}(B)\mathsf{Pr}(A|\,B)$$

となる. これを**乗法の公式**という.
(3) $\mathsf{Pr}(B) > 0$ のとき, 関数 $\mathsf{Pr}(\,\cdot\,|B) : \mathcal{A} \to [0, 1]$ は定義 1.6(1) – (3) をみたす. すなわち $(\Omega, \mathcal{A})$ 上の確率測度である. □

問 1.2. 注意 1.10(1)(3) を証明せよ.

補題 1.11. (全確率の法則) $N \in \mathbb{N}$ とする．$A_1, A_2, \ldots, A_N \in \mathcal{A}$ は Ω の分割とする．すなわち，すべての $m \neq n\,(m, n \in \{1, 2, \ldots, N\})$ に対し，$A_m \cap A_n = \varnothing$ で，$\bigcup_{n=1}^{N} A_n = \Omega$ が成立している．このとき任意の $B \in \mathcal{A}$ に対して

$$\mathsf{Pr}(B) = \mathsf{Pr}\left(\bigcup_{n=1}^{N}(A_n \cap B)\right) = \sum_{n=1}^{N} \mathsf{Pr}(A_n \cap B)$$

となる．

Proof. $B = B \cap \Omega = B \cap \left(\bigcup_{n=1}^{N} A_n\right) = \bigcup_{n=1}^{N}(A_n \cap B)$ と $\{A_i \cap B\}_{i=1}^{n}$ は互いに排反であることに注意して，定義 1.6(2) を用いればよい． □

定理 1.12. (Bayes の定理) (1) $N \in \mathbb{N}$ とする．$A_1, A_2, \ldots, A_N, B \in \mathcal{A}$ とし，$A_1, A_2, \ldots, A_N$ は Ω の**分割**とする．$j = 1, 2, \ldots, N$ に対して，$\mathsf{Pr}(B) > 0$, $\mathsf{Pr}(A_j) > 0$ のとき

$$\mathsf{Pr}(A_j |\, B) = \frac{\mathsf{Pr}(A_j)\mathsf{Pr}(B |\, A_j)}{\sum_{n=1}^{N} \mathsf{Pr}(A_n)\mathsf{Pr}(B |\, A_n)}$$

が成立する．
(2) $A, B \in \mathcal{A}$ とし，$\mathsf{Pr}(A) > 0$, $\mathsf{Pr}(B) > 0$ とする．このとき

$$\mathsf{Pr}(A |\, B)\mathsf{Pr}(B) = \mathsf{Pr}(B |\, A)\mathsf{Pr}(A)$$

が成立する．

Proof. 条件付き確率の定義，補題 1.11, 定義 1.4(2) を用いればよい． □

例 1.13. 病気 D に対する検査の結果を $+$ と $-$ とし，確率が以下であるとする．

$$\mathsf{Pr}(+ |\, D) = 0.9, \quad \mathsf{Pr}(- |\, D^c) = 0.9, \quad \mathsf{Pr}(D) = 0.01.$$

Bayes の定理を用いて，検査で $+$ に人が本当に D である確率を求めると

$$\begin{aligned}\mathsf{Pr}(D |\, +) = \frac{\mathsf{Pr}(+ \cap D)}{\mathsf{Pr}(+)} &= \frac{\mathsf{Pr}(D)\mathsf{Pr}(+ |\, D)}{\mathsf{Pr}(D)\mathsf{Pr}(+|, D) + \mathsf{Pr}(D^c)\mathsf{Pr}(+ |\, D^c)} \\ &= \frac{0.01 \times 0.9}{0.01 \times 0.9 + (1 - 0.01) \times \{1 - 0.9\}} \approx 0.83\end{aligned}$$

となる．

1.2　確率変数

前節では, 確率と事象を記述する数学的なモデルを導入した. しかし, 統計学が扱う現実の現象は, 事象には直接結びつかないかもしれない数量的な情報である. 以下で定義する確率変数は, 事象と数量の間の橋渡しをする.

定義 1.14. $(\Omega, \mathcal{A})$ と $(\mathbb{X}, \mathcal{B})$ を可測空間とする. 関数 $X : \Omega \to \mathbb{X}$ は $(\Omega, \mathcal{A})$ から $(\mathbb{X}, \mathcal{B})$ への**可測写像**であるとは

$$X^{-1}(B) := \bigl\{\omega \in \Omega : X(\omega) \in B\bigr\} \in \mathcal{A} \qquad \bigl(\forall B \in \mathcal{B}\bigr)$$

をみたすときをいう.

注意 1.15. (i). $d \geq 2\,(d \in \mathbb{N})$ とする. $\bigl(\mathbb{X}, \mathcal{B}\bigr) = \bigl(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\bigr)$ のとき, X は**確率ベクトル**と呼ばれる.
(ii). $d = 1$ のとき, X は**確率変数**と呼ばれる. □

定理 1.16. $\bigl(\mathbb{X}, \mathcal{B}\bigr)$ を可測空間とし, $X : \Omega \to \mathbb{X}$ を写像とし, $\mathcal{C}$ を $\mathbb{X}$ の集合族とする. $\forall C \in \mathcal{C}$ に対して, $\bigl\{\omega \in \Omega; X(\omega) \in C\bigr\} \in \mathcal{A}$ であり, $\mathcal{C}$ が $\mathcal{B}$ を生成する[5]とき, X は可測となる.

Proof. $B \in \mathcal{B}$ に対して, $\bigl\{\omega \in B : X(\omega) \in B\bigr\}$ を $\{X \in B\}$ と書くことにする. $\{B\}, \{B_n\}_{n=1}^{\infty} \subset \mathcal{B}$ に対して

$$\left\{X \in \bigcup_{n=1}^{\infty} B_n\right\} = \bigcup_{n=1}^{\infty}\{X \in B_n\}$$
$$\bigl\{X \in B^c\bigr\} = \bigl\{X \in B\bigr\}^c$$

となる. したがって, 集合族 $\mathcal{D} := \{B \subset \mathcal{B}; \{X \in B\} \in \mathcal{A}\}$ は σ 集合族となる. よって, $\mathcal{D} \subset \mathcal{C}$ であり, $\mathcal{C}$ は $\mathcal{B}$ を生成するので, $\mathcal{D} \subset \mathcal{B}$ となる. □

注意 1.17. (i). $\bigl(\mathbb{X}, \mathcal{B}\bigr) = \bigl(\mathbb{R}, \mathcal{B}(\mathbb{R})\bigr)$ に対して, 定理 1.16 における $\mathcal{C}$ の選択として, $\{(-\infty, r] : r \in \mathbb{R}\}$ と $\{(-\infty, q] : q \in \mathbb{Q}\}$ などがある.
(ii). $d \geq 2\,(d \in \mathbb{N})$ とする. $\bigl(\mathbb{X}, \mathcal{B}\bigr) = \bigl(\mathbb{R}^d, \mathcal{B}(\mathbb{R})\bigr)$ に対して, 定理 1.16 における $\mathcal{C}$ の選択として

$$\{(a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_d, b_d) : -\infty < a_i < b_i < \infty\,(i = 1, 2, \ldots, d)\}$$

がある. □

定理 1.18. $\bigl(\Omega, \mathcal{A}\bigr)$, $\bigl(\mathbb{X}, \mathcal{B}\bigr)$, $\bigl(\mathbb{Y}, \mathcal{C}\bigr)$ を可測空間とする. $X : \Omega \to \mathbb{X}$ と $f : \mathbb{X} \to \mathbb{Y}$ は可測写像のとき, $f(X) : \Omega \to \mathbb{Y}$ は可測写像となる.

[5] $\mathcal{B}$ は $\mathcal{C}$ を含む最小の σ 集合族.

Proof. $C \in \mathcal{C}$ に対して

$$\big\{\omega \in \Omega : f(X(\omega)) \in C\big\} = \big\{\omega \in \Omega : X(\omega) \in f^{-1}(C)\big\} \in \mathcal{A}$$

となる．なぜならば, f の可測性から $f^{-1}(C) \in \mathcal{B}$ となることからわかる．
□

定理 1.19. $n \in \mathbb{N}$ とし, $X_1, X_2, \ldots, X_n$ は確率変数とし, $f : \mathbb{R}^n \to \mathbb{R}$ は可測とする．このとき, $f(X_1, X_2, \ldots, X_n)$ は確率変数となる．

Proof. 定理 1.18 から $(X_1, X_2, \ldots, X_n)$ は確率ベクトルであることを示せばよい．$A_1, A_2, \ldots, A_n \in \mathcal{B}(\mathbb{R})$ に対して

$$\big\{(X_1, X_2, \ldots, X_n) \in A_1 \times A_2 \times \cdots \times A_n\big\} = \bigcap_{i=1}^{n} \{X_i \in A_i\} \in \mathcal{A}$$

である．さらに, 集合族 $\{A_1 \times A_2 \times \cdots \times A_n : A_i \in \mathcal{B}(\mathbb{R})\,(i = 1, 2, \ldots, n)\}$ は $\mathcal{B}(\mathbb{R}^n)$ を生成[6]する．したがって, 定理 1.16 から $\big(\mathbb{X}, \mathcal{B}\big) = \big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)\big)$ に対して, $(X_1, X_2, \ldots, X_n)$ は確率ベクトルであることがわかる．□

定理 1.20. $X_1, X_2, \ldots, X_n$ は確率変数のとき, $X_1 + X_2 + \cdots + X_n$ も確率変数となる．

Proof. 定理 1.19 から $\mathbb{R}^n$ 上の実数値関数 $f(x_1, x_2, \ldots, x_n) = x_1 + x_2 + \cdots + x_n$ が可測であることを示せばよい．$\forall r \in \mathbb{R}$ に対して

$$\{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : x_1 + x_2 + \cdots + x_n < r\}$$

は $\mathbb{R}^n$ の開集合となるので

$$\{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : x_1 + x_2 + \cdots + x_n < r\} \in \mathcal{B}(\mathbb{R}^n)$$

となる．集合族 $\{(-\infty, r) : r \in \mathbb{R}\}$ は $\mathcal{B}(\mathbb{R})$ を生成[7]するので, f は可測である．□

注意 1.21. 同様の議論から, X_1, X_2 は確率変数のとき, X_1X_2 も確率変数であることがわかる．さらに, $X_2 \neq 0$ のとき, X_1/X_2 も確率変数であることもわかる．□

[6] この事実の証明には, 測度論の議論が必要なので, 信じることにしよう．

[7] この事実の証明には, 測度論の議論が必要となる．信じることにする．丁寧な証明は

`https://mcm-www.jwu.ac.jp/~konno/pdf/note_book=20230831.pdf`

にある．

定理 1.22. $X_1, X_2, \ldots$ は確率変数列のとき

$$\inf_n X_n \quad \sup_n X_n \quad \limsup_n X_n \quad \liminf_n X_n$$

も確率変数となる.

Proof. $r \in \mathbb{R}$ に対して

$$\left\{\omega \in \Omega : \inf_n X_n(\omega) < r\right\} = \bigcup_{n=1}^{\infty} \left\{\omega \in \Omega : X_n(\omega) < r\right\} \in \mathcal{A}$$

が $\inf_n X_n$ は確率変数であることがわかる. 同様に

$$\left\{\omega \in \Omega : \sup_n X_n(\omega) < r\right\} = \bigcap_{n=1}^{\infty} \left\{\omega \in \Omega : X_n(\omega) < r\right\}$$

から $\sup_n X_n$ も確率変数であることがわかる. 次に

$$\liminf_n X_n = \sup_n \left(\inf_{m \geq n} X_m \right)$$

に注意する. $Y_n := \inf_{m \geq n} X_m$ は各 $n \in \mathbb{N}$ に対して確率変数なので, $\sup_n Y_n$ も確率変数となる. 同様に

$$\limsup_n X_n = \inf_n \left(\sup_{m \geq n} X_m \right)$$

から $\limsup_n X_n$ も確率変数であることはわかる. □

注意 1.23. (i). $X_1, X_2, \ldots\ldots$ を確率変数列とする. 定理 1.22 と 定理 1.19 から

$$\begin{aligned}\Omega_0 :=& \left\{\omega \in \Omega : \lim_{n\to\infty} X_n \text{ は存在}\right\} \\ =& \left\{\omega \in \Omega : \limsup_{n\to\infty} X_n - \liminf_{n\to\infty} X_n = 0\right\}\end{aligned}$$

は可測集合となる. $\mathrm{Pr}(\Omega_0) = 1$ のとき, 確率変数列 $\{X_n\}_{n=1}^{\infty}$ は**ほとんど確実に収束**するといい, a.s. と記す. $\mathbb{R} \cup \{\pm\infty\}$ に値を取る確率変数を X_∞ とおくと

$$X_\infty = \lim_{n\to\infty} X_n \qquad \text{(a.s.)}$$

と書くける.

(ii). 任意の $a, b \in \mathbb{R}$ とし, $\{-\infty\} \cup (-\infty, a)$, (a, b), $(b, \infty) \cup \{+\infty\}$ によって生成される σ 集合族を $\mathcal{B}(\mathbb{R} \cup \{\pm\infty\})$ と書く. 集合 $D \subset \Omega$ が定義域で, $\mathbb{R} \cup \{\pm\infty\}$ を値域ともつ関数 X は可測であるとは, $\forall B \in \mathcal{B}(\mathbb{R} \cup \{\pm\infty\})$ に対して

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A}$$

をみたすときをいう. □

定義 1.24. (1) 確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率変数 X に対して

$$\mathsf{F}^X(x) := \mathsf{Pr}(X \leq x) := \mathsf{Pr}(\{\omega \in \Omega;\, X(\omega) \leq x\}) \qquad (x \in \mathbb{R})$$

を X **の累積分布関数** (cumulative distribution function(c.d.f.)) という．また

$$\mathsf{P}^X(B) := \mathsf{Pr}(X \in B) := \mathsf{Pr}(\{\omega \in \Omega;\, X(\omega) \in B\}) \qquad (B \in \mathcal{B}(\mathbb{R}))$$

を X の**分布**という．したがって, $\mathsf{P}^X((-\infty, x]) = \mathsf{F}^X(x)$ である．
(2) 確率変数 X, Y を $(\Omega, \mathcal{A}, \mathsf{Pr})$ の確率変数とし, それぞれの c.d.f. を $\mathsf{F}^X, \mathsf{F}^Y$ とする．このとき

$$\mathsf{F}^X(x) = \mathsf{F}^Y(x) \qquad (\forall x \in \mathbb{R})$$

が成立するとき, X と Y の分布は同じである[8]という．これを $X \overset{d}{=} Y$ と書く．
(3) 確率変数 X が c.d.f. F を持つとき, $X \sim \mathsf{F}$ と書く．

注意 1.25. 文脈から, どの確率変数の確率分布または累積分布関数であることがわかる場合には, 簡単に P または F と書くこともある．　□

注意 1.26. (1) P^X は可測空間 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上の確率測度である．すなわち P^X は定義 1.6 をみたすことがわかる．
(2) c.d.f. F が与えられると

$$\mathsf{F}(x) = \mathsf{P}((-\infty, x]) \qquad (x \in \mathbb{R})$$

をみたす $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上の確率測度 P が一意的に定まることが知られている．証明は補遺の命題 C.29 を参照せよ．このことにより X の c.d.f. と分布を同一視する．さらに $X \sim \mathsf{F}$ とし, c.d.f. F から定まる確率測度を P としたとき, $X \sim \mathsf{P}$ とも書く．
(3) 確率変数 X と Y の分布が同じとき

$$\mathsf{P}^X(B) = \mathsf{P}^Y(B) \qquad (\forall B \in \mathcal{B}(\mathbb{R}))$$

も成立する．

問 1.3. 注意 1.26(1) を示せ．逆像の性質を利用すること．

定理 1.27. 確率変数 X を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率変数とし, F を X の c.d.f. とする．このとき F は次をみたす．

[8] $\mathsf{F}^X(x) = \mathsf{F}^Y(x)\,(\forall x \in \mathbb{R}) \Leftrightarrow \mathsf{P}^X(B) = \mathsf{P}^Y(B)\,(\forall B \in \mathcal{B}(\mathbb{R})$ となることが知られている．注意 1.26(2) を参照のこと．

(1) F は非減少関数: $x < y \Rightarrow \mathsf{F}(x) \leq \mathsf{F}(y)$.

(2) $\lim_{x\to\infty} \mathsf{F}(x) = 1$: $\lim_{x\to-\infty} \mathsf{F}(x) = 0$.

(3) F は右連続関数: $\lim_{y\to x+0} \mathsf{F}(y) = \mathsf{F}(x)$.

Proof. 確率変数 X の分布を P とする．すなわち $\mathsf{P}(B) := \mathsf{Pr}(X \in B)$ $\big(\forall B \in \mathcal{B}(\mathbb{R})\big)$ である.
(1) $(-\infty, x] \subset (-\infty, y]$ に注意して, 補題 1.7(4) を P に適用すればよい.
(2) $A_n = (-\infty, n]$ と $A_n = (-\infty, -n]$ として, P に補題 1.7(6)(7) を適用すればよい.
(3) $A_n = \left(-\infty, x + \dfrac{1}{n}\right]$ として, P に補題 1.7(7) を適用すればよい． □

問 1.4. 固定した $x \in \mathbb{R}$ に対して

$$(-\infty, x] = \bigcap_{n=1}^{\infty} \left(-\infty, x + \frac{1}{n}\right)$$

を示せ.

問 1.5. 定理 1.27(1) – (3) の証明を詳しく書き下せ.

定理 1.28. 確率変数 X を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率変数とし, X の c.d.f. を F とする．このとき次の (1) – (3) は同値である.

(1) F は $\mathbb{R}$ 上の連続関数.

(2) $\mathsf{F}(x) = \mathsf{F}(x-)\,(\forall x \in \mathbb{R})$．ただし, $\mathsf{F}(x-) := \sup\{\mathsf{F}(y);\, y < x\}$ とした.

(3) $\mathsf{Pr}(X = x) = 0\,(\forall x \in \mathbb{R})$.

Proof. (1) $\Rightarrow$ (2) の証明: $x_n = x - \dfrac{1}{n}\,(n = 1, 2, \ldots)$ とすれば

$$\lim_{n\to\infty} \mathsf{F}(x_n) = F(x-)$$

である．この式から, F が連続ならば, $\mathsf{F}(x) = \mathsf{F}(x-)$ となることがわかる.
(2) $\Rightarrow$ (3) の証明:

$$\{X \leq x_n\} \subset \{X \leq x_{n+1}\}\,(n = 1, 2, \ldots), \quad \bigcup_{n=1}^{\infty}\{X \leq x_n\} = \{X < x\}$$

であるので, 補題 1.7(6) より

$$\begin{aligned}\mathsf{F}(x-) &= \lim_{n\to\infty} \mathsf{F}(x_n) = \lim_{n\to\infty} \mathsf{Pr}(X \leq x_n) = \mathsf{Pr}\left(\bigcup_{n=1}^{\infty}\{X \leq x_n\}\right)\\ &= \mathsf{Pr}(X < x)\end{aligned}$$

を得る．さらに補題 1.7(4) より

$$\begin{aligned}\mathsf{Pr}(X=x) &= \mathsf{Pr}\Big(\{X \le x\} \setminus \{X < x\}\Big) = \mathsf{Pr}(X \le x) - \mathsf{Pr}(X < x) \\ &= F(x) - F(x-) \end{aligned} \tag{1.1}$$

であることからわかる．よって, $\mathsf{Pr}(X=0)=0$ となる．

(3) ⇒ (1) の証明: (1.1) から

$$\mathsf{Pr}(X=x) \Rightarrow \mathsf{F}(x) = \mathsf{F}(x-)$$

である．このことと F は右連続であることから F は連続となる．以上から 3 つの条件は同値であることが示せた． □

問 1.6. 定理 1.28 の記号を踏襲する．固定した $x \in \mathbb{R}$ に対して

$$\begin{aligned}\bigcup_{n=1}^{\infty}\left\{X \le x - \frac{1}{n}\right\} &= \bigcup_{n=1}^{\infty}\left\{\omega \in \Omega;\, X(\omega) \le x - \frac{1}{n}\right\} = \{\omega \in \Omega;\, X(\omega) < x\} \\ &= \{X < x\}\end{aligned}$$

を示せ．

注意 1.29. F が $\mathbb{R}$ 上で連続のとき, $\mathsf{Pr}(X=x) = 0\,(\forall x \in \mathbb{R})$ なので, $a<b$ に対して

$$\mathsf{Pr}(a \le X \le b) = \mathsf{Pr}(a < X < b) = \mathsf{Pr}(a < X \le b) = \mathsf{Pr}(a \le X < b)$$

である． □

定義 1.30. 確率変数 X を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率変数とし, その c.d.f. と分布を F と P と書く．

(1) X が高々可算個の集合 $\{x_1, x_2, \ldots\}$ 上にしか値を取らないとき, X を**離散型**であるという．この場合には

$$\mathsf{p}(x) := \mathsf{Pr}(X=x) = \mathsf{F}(x) - \mathsf{F}(x-) \qquad (x \in \{x_1, x_2, \ldots\})$$

で X の分布が特徴付けられる[9]．その p を X の**確率関数** (probability mass function(p.m.f.)) と呼ぶ．

(2) $\mathsf{Pr}(X=x) = 0\,(\forall x \in \mathbb{R})$ のとき, X を連続型であるという．さらにある非負値関数 p で

$$\mathsf{F}(x) = \int_{-\infty}^{x} \mathsf{p}(t)\,\mathrm{d}t \qquad (\forall x \in \mathbb{R})$$

[9] $\mathsf{p}(x) = 0\,(x \notin \{x_1, x_2, \ldots\})$ となるので, p は $\mathbb{R}$ 上の関数であり, $0 \le \mathsf{p}(x) \le 1$ となる．

をみたすものが存在するとき, p を X の**確率密度関数** (probability density function(p.d.f.)) という. 特に F がほとんどいたるところ[10]で微分可能ならば, ほとんどいたるところで

$$\dot{\mathsf{F}}(x) = \frac{\mathrm{d}\mathsf{F}}{\mathrm{d}x}(x) = \mathsf{p}(x)$$

となる.

注意 1.31. 確率変数 X の分布を P とする.
(1) X を離散型とし, $S = \{x \in \mathbb{R};\, \mathsf{p}(x) > 0\}$ とおく[11]. すると任意の Borel 集合 $B \subset \mathbb{R}$ に対して

$$\mathsf{P}(B) = \mathsf{Pr}(X \in B) = \sum_{x \in B \cap S} \mathsf{p}(x)$$

が成り立つことを示すことができる.
(2) X を連続型とし, その p.d.f. p が定義されるとする. このとき, 任意の Borel 集合 $B \subset \mathbb{R}$ に対して

$$\mathsf{P}(B) = \mathsf{Pr}(X \in B) = \int_B \mathsf{p}(x)\,\mathrm{d}x$$

が成り立つことを示すことができる. 厳密な証明は測度論の知識が必要となるので, 証明は省略する.

例 1.32. (1) $\Omega = \{0, 1\}$, $\mathcal{A} = 2^\Omega$, $\mathsf{Pr}(\{0\}) = \mathsf{Pr}(\{1\}) = 1/2$ とし

$$X : \Omega \ni \omega \mapsto X(\omega) = \omega \in \mathbb{R}$$

と定義すれば

$$\mathsf{Pr}(X = 0) = \mathsf{Pr}(X = 1) = \frac{1}{2}$$

となる. このとき, X の c.d.f. F と p.m.f. p はそれぞれ

$$\mathsf{F}(x) = \begin{cases} 0 & (x < 0) \\ \frac{1}{2} & (0 \le x < 1) \\ 1 & (x \ge 1) \end{cases} \quad ; \qquad \mathsf{p}(x) = \begin{cases} \frac{1}{2} & (x = 0,\, 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

となる.
(2) $\Omega = [0, 1]$, $\mathcal{A} = \mathcal{B}(\mathbb{R}) \cap [0, 1]$ とし

$$\mathsf{Pr}((a, b]) = b - a \qquad (0 \le a < b \le 1)$$

---

[10] この授業では, $\mathbb{R}$ から可算個の点を除いた集合上で微分可能と理解して差し支えない.

[11] 累積分布関数 F は有界で非減少なので, その不連続点は高々可算個であるから, 下記の式の右辺の和記号が正当化される.

とする．さらに

$$X : \Omega \ni \omega \mapsto X(\omega) = \omega \in \mathbb{R}$$

と定義すれば, X の c.d.f. F と p.d.f. p はそれぞれ

$$\mathsf{F}(x) = \begin{cases} 0 & (x < 0) \\ x & (0 \leq x \leq 1) \\ 1 & (x > 1) \end{cases} \quad ; \quad \mathsf{p}(x) = \begin{cases} 1 & (0 \leq x \leq 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

となる．

1.3 分位点関数

定義 1.33. c.d.f. F に対して, **分位点関数** (quantile function) $\mathsf{F}^- : (0, 1) \to \mathbb{R}$ を

$$\mathsf{F}^-(y) := \inf\bigl\{x \in \mathbb{R} : \mathsf{F}(x) \geq y\bigr\} \qquad (0 < y < 1)$$

で定義する．また, $y \in (0, 1)$ に対して, $\mathsf{F}^{-1}(y)$ を F **の** y **分位点**とよぶ. $1/2$ 分位点を**メディアン** (median) と呼ぶ.

注意 1.34. c.d.f. F が連続かつ $\operatorname{supp}\mathsf{F} = \{x \in \mathbb{R}; 0 < \mathsf{F}(x) < 1\}$ 上で狭義単調増加のとき, F^- は F の逆関数になる． □

注意 1.35. F^- の定義から, 任意の $0 < y < 1$ に対して, 単調減少列 $\{x_n\}_{n=1}^\infty$ が存在して

$$\mathsf{F}(x_n) \geq y \quad \text{かつ} \quad \lim_{n\to\infty} x_n = \mathsf{F}^{-1}(y)$$

とできる．このとき, $\{x_n\}_{n=1}^\infty$ の単調減少性と F の右連続性から

$$\mathsf{F}\bigl(\mathsf{F}^-(y)\bigr) = \mathsf{F}\Bigl(\lim_{n\to\infty} x_n\Bigr) = \lim_{n\to\infty} \mathsf{F}(x_n) \geq y \tag{1.2}$$

となることがわかる．よって, F^- の定義における inf は達成される． □

注意 1.36. c.d.f. F は連続とする．c.d.f. の性質

$$\lim_{n\to\infty} \mathsf{F}(x) = 1 \quad \text{かつ} \quad \lim_{x\to-\infty} \mathsf{F}(x) = 0$$

と F の連続性から, 中間値の定理を用いると任意の $0 < y < 1$ に対して, ある $x \in \mathbb{R}$ が存在して

$$\mathsf{F}(x) = y$$

をみたす．ここで, F^{-} の定義から

$$\mathsf{F}^{-}(y) \leq x \tag{1.3}$$

となる．なぜならば, $\mathsf{F}^{-}(y)$ は条件 $\mathsf{F}(z) \geq y$ をみたす $z \in \mathbb{R}$ の inf である．一方, $\mathsf{F}(x) = y$ だから $x \in \{z \in \mathbb{R} : \mathsf{F}(z) \geq y\}$ なので, $x \geq \mathsf{F}^{-}(y)$ がわかる．(1.3) と F の非減少性から

$$\mathsf{F}\big(\mathsf{F}^{-}(y)\big) \leq \mathsf{F}(x) = y \tag{1.4}$$

となる．一方, F^{-} の定義から

$$\mathsf{F}\big(\mathsf{F}^{-}(y)\big) \geq y \tag{1.5}$$

となる．(1.4) と (1.5) を合わせると

$$\mathsf{F}\big(\mathsf{F}^{-}(y)\big) = y$$

を得る．したがって

$$\mathsf{F}\text{ が連続 } \Rightarrow \mathsf{F}\big(\mathsf{F}^{-}(y)\big) = y\,(\forall 0 < y < 1)$$

となる．

注意 1.37. F が不連続ならば

$$\mathsf{F}\big(\mathsf{F}^{-}(y)\big) = y$$

とは限らない．反例は, 次の例 1.39 を参照せよ．□

例 1.38. $a < b\,(a,\, b \in \mathbb{R})$ とし, $X \sim \mathsf{U}(a,\, b)$ とする．すると X の c.d.f. は

$$\mathsf{F}(x) = \begin{cases} 0 & (x \leq a) \\ \dfrac{x-a}{b-a} & (a < x < b) \\ 1 & (x \geq b) \end{cases}$$

である．よって, F の分位点関数は

$$\mathsf{F}^{-}(y) = a + (b-a)y \qquad (y \in (0,\, 1))$$

となる．□

例 1.39. $\mathsf{Pr}(X=0)=\mathsf{Pr}(X=1)=\dfrac{1}{2}$ とする．このとき，X の c.d.f. は

$$\mathsf{F}(x)=\begin{cases} 0 & (x<0) \\ \dfrac{1}{2} & (0\leq x<1) \\ 1 & (x\geq 1) \end{cases}$$

となる．よって，F の分位点関数は

$$\mathsf{F}^{-}(y)=\begin{cases} 0 & \left(0<y\leq \dfrac{1}{2}\right) \\ 1 & \left(\dfrac{1}{2}<y<1\right) \end{cases}$$

となる．この場合，$y\neq\dfrac{1}{2}$ のとき

$$\mathsf{F}\big(\mathsf{F}^{-}(y)\big)\neq y$$

となる． □

定理 1.40. 分位点関数は以下の性質 (1) ～ (3) をもつ．
(1) F^{-} は非減少である．
(2) F^{-} は左連続である．すなわち，$\{y_n\}_{n=1}^{\infty}$ を任意の非減少列で $\lim_{n\to\infty}y_n=y$ としたとき

$$\lim_{n\to\infty}\mathsf{F}^{-}(y_n)=\mathsf{F}^{-}(y)$$

が成立する．
(3) $\mathsf{F}^{-}(y)\leq x \ \Leftrightarrow\ y\leq\mathsf{F}(x)$ が成立する．

Proof. (1) $y_1<y_2\,(y_1,\,y_2\in(0,\,1))$ に対して

$$\big\{x\in\mathbb{R}:\,\mathsf{F}(x)\geq y_1\big\}\supset\big\{x\in\mathbb{R}:\,\mathsf{F}(x)\geq y_2\big\}$$

となる．両辺の inf をとると

$$\mathsf{F}^{-}(y_1)=\inf\big\{x\in\mathbb{R}:\,\mathsf{F}(x)\geq y_1\big\}\leq\inf\big\{x\in\mathbb{R}:\,\mathsf{F}(x)\geq y_2\big\}=\mathsf{F}^{-}(y_2)$$

となる．よって，F^{-} の非減少性が証明できた．
(2) $x_n:=\mathsf{F}^{-}(y_n)\,(n\in\mathbb{N})$ とおく．数列 $\{y_n\}_{n=1}^{\infty}$ は非減少なので，(1) から $\{x_n\}_{n=1}^{\infty}$ も非減少列となる．さらに，$y_n\leq y$ から

$$x_n=\mathsf{F}^{-}(y_n)\leq\mathsf{F}^{-}(y)=:x_0 \tag{1.6}$$

となる．よって, (1.6) の両辺の極限を取ると

$$x := \lim_{n\to\infty} x_n \le x_0 \tag{1.7}$$

となる．ここで, $x < x_0$ を仮定して矛盾を導く．そのために

$$\epsilon := \frac{1}{2}(x_0 - x) > 0 \Leftrightarrow x = x_0 - 2\epsilon$$

とおく．F^- の定義から導出された (1.5) に注意すると

$$y_n \le \mathsf{F}\big(\mathsf{F}^-(y_n)\big) \le \mathsf{F}(x_n) \le \mathsf{F}(x+\epsilon) = \mathsf{F}(x_0 - \epsilon)$$

となる．すると

$$y = \lim_{n\to\infty} y_n \le \mathsf{F}(x_0 - \epsilon)$$

となる．しかし, F^- の定義から

$$\mathsf{F}(x_0 - \epsilon) < y$$

となる．なぜならば, $\mathsf{F}^-(y) = x_0$ なので, x_0 は条件 $\mathsf{F}(z) \ge y$ をみたす z の inf である．$x_0 > x_0 - \epsilon$ だから, $x_0 - \epsilon \not\in \{z \in \mathbb{R} : \mathsf{F}(z) \ge y\}$ となる．よって, $\mathsf{F}(x_0 - \epsilon) < y$ となることがわかる．以上から $y \le \mathsf{F}(x_0 - \epsilon)$ かつ $y > \mathsf{F}(x_0 - \epsilon)$ となるので, 矛盾．よって

$$\mathsf{F}^-(y) = x_0 = x = \lim_{n\to\infty} x_n = \lim_{n\to\infty} \mathsf{F}^-(y_n)$$

となるので, F^- の左連続性がわかる．

(3) F^- の非減少性から

$$y \le \mathsf{F}(x) \Rightarrow \mathsf{F}^-(y) \le \mathsf{F}^-\big(\mathsf{F}(x)\big) \le x$$

となる．最後の不等号は

$$\mathsf{F}^-\big(\mathsf{F}(x)\big) = \inf\big\{z \in \mathbb{R} : \mathsf{F}(z) \ge \mathsf{F}(x)\big\}$$

である．しかし, $x \in \big\{z \in \mathbb{R} : \mathsf{F}(z) \ge \mathsf{F}(x)\big\}$ であるので, $x \ge \mathsf{F}^-\big(\mathsf{F}(x)\big)$ がわかる．

逆に, $\mathsf{F}^-(y) \le x$ ならば, (1.5) と F の非減少性から

$$y \le \mathsf{F}\big(\mathsf{F}^-(y)\big) = \mathsf{F}(x)$$

がわかる．よって, (3) は証明された．□

系 1.41. F を c.d.f. とする．このとき, $U \sim \mathsf{U}(0,1)$ に対して

$$X := \mathsf{F}^{-}(U) \sim \mathsf{F}$$

となる．

Proof. 定理 1.40(3) から, $\forall x \in \mathbb{R}$ に対して

$$\{X \le x\} \Leftrightarrow \{U \le \mathsf{F}(x)\}$$

となる．よって

$$\mathsf{Pr}(X \le x) = \mathsf{Pr}(U \le \mathsf{F}(x)) = \mathsf{F}(x)$$

を得る． □

1.4　主な 1 次元分布

1.4.1　離散型確率変数

Bernoulli 分布

$0 \le \theta \le 1$ とする．確率変数 X は**母数 θ の Bernoulli 分布**に従うとは, X の p.m.f. $\mathsf{p}(\cdot\,|\,\theta)$ が

$$\mathsf{p}(x|\,\theta) = p(x) = \begin{cases} \theta^x(1-\theta)^{1-x} & (x = 0,\, 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

のときをいう．このことを $X \sim \mathsf{Ber}(\theta)$ と記す．

2 項分布

$n \in \mathbb{N}$, $0 \le \theta \le 1$ とする．確率変数 X は**母数 (n,θ) の 2 項分布**に従うとは, X の p.m.f. $\mathsf{p}(\cdot\,|\,n,\theta)$ が

$$\mathsf{p}(x|\,n,\theta) = \mathsf{p}(x) = \begin{cases} \begin{pmatrix} n \\ x \end{pmatrix} \theta^x(1-\theta)^{n-x} & (x = 0,\,1,\,\ldots,\,n) \\ 0 & (\text{その他の場合}) \end{cases}$$

のときをいう．ただし

$$\begin{pmatrix} n \\ x \end{pmatrix} = \frac{n!}{x!\,(n-x)!}, \qquad 0! = 1$$

である．このことを $X \sim \mathsf{Bino}(n,\theta)$ と記す．

問 1.7. $\mathsf{p}(x|\,n,\,\theta)$ を 2 項分布 $\mathsf{Bino}(n,\,\theta)\,(0<\theta<1,\,n\in\mathbb{N})$ p.m.f. としたとき, $\sum_{x=0}^{n}\mathsf{p}(x|\,n,\,\theta)=1$ を確認せよ.

幾何分布

$0<\theta<1$ とする. 確率変数 X は (**母数 θ の幾何分布**に従うとは, X の p.m.f. $\mathsf{p}(\,\cdot\,|\,\theta)$ が

$$\mathsf{p}(x|\,\theta)=\mathsf{p}(x)=\begin{cases}\theta(1-\theta)^{x-1} & (x=1,\,2,\,\ldots)\\ 0 & (\text{その他の場合})\end{cases}$$

のときをいう. このことを $X\sim\mathsf{Geo}(\theta)$ と記す.

問 1.8. $\mathsf{p}(x|\,n,\,\theta)$ を幾何分布 $\mathsf{Geo}(\theta)\,(\theta>0)$ の p.m.f. としたとき, $\sum_{x=0}^{\infty}\mathsf{p}(x|\,\theta)=1$ を確認せよ.

Poisson 分布

$\theta>0$ とする. 確率変数 X は**母数 θ の Poisson 分布**に従うとは, X の p.m.f. $\mathsf{p}(\,\cdot\,|\,\theta)$ が

$$\mathsf{p}(x|\,\theta)=\mathsf{p}(x)=\begin{cases}e^{-\theta}\dfrac{\theta^x}{x!} & (x=0,\,1,\,2,\,\ldots)\\ 0 & (\text{その他の場合})\end{cases}$$

のときをいう. このことを $X\sim\mathsf{Po}(\theta)$ と記す.

問 1.9. $\mathsf{p}(x|\,\theta)$ を Poisson 分布 $\mathsf{Po}(\theta)\,(\theta>0)$ の p.m.f. としたとき, $\sum_{x=0}^{\infty}\mathsf{p}(x|\,\theta)=1$ を確認せよ.

1.4.2　連続型確率変数

正規分布

$\mu\in\mathbb{R},\,0<\sigma<\infty$ とする. 確率変数 X は**平均 μ, 分散 σ^2 の正規分布**[12]に従うとは, X の p.d.f. $\mathsf{p}(\,\cdot\,|\,\mu,\,\sigma^2)$ が

$$\mathsf{p}(x|\,\mu,\,\sigma^2)=\mathsf{p}(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\qquad(-\infty<x<\infty)$$

のときをいう. ただし $\exp(x)=e^x$ である. このことを $X\sim\mathsf{N}(\mu,\,\sigma^2)$ と記す. $\mu=0,\,\sigma=1$ のときの分布を**標準正規分布**という.

[12]母数 $(\mu,\,\sigma^2)$ を平均と分散となぜ呼ぶかは第 2 章で判明する.

注意 1.42. p.d.f. $\mathsf{p}(x|\,\mu,\,\sigma^2)$ のグラフは単峰型であり, μ に関して左右対称となる．さらに, μ が p.d.f. $\mathsf{p}(x|\,\mu,\,\sigma^2)$ のグラフの峰の位置に対応することがわかる．また, σ が大きくなると p.d.f. $\mathsf{p}(x|\,\mu,\,\sigma^2)$ のグラフの裾が長くなることがわかる.

注意 1.43. よく知られている事実として

$$\int_0^\infty e^{-t^2}\,\mathrm{d}t = \sqrt{\pi}$$

がある．$\mathbb{R} \ni t \mapsto e^{-t^2} \in (0,\,\infty)$ は偶関数であることに注意して, $t = z/\sqrt{2}$ と変換すると

$$\int_{-\infty}^\infty e^{-z^2/2}\,\mathrm{d}z = \sqrt{2\pi}$$

となることがわかる．さらに, $x = (z-\mu)/\sigma$ と変数変換をすると

$$\int_{-\infty}^\infty \mathsf{p}(x|\,\mu,\,\sigma^2)\,\mathrm{d}x = 1$$

となることが確認できる. □

ガンマ分布

$\alpha > 0, \beta > 0$ とする．確率変数 X は**母数** $(\alpha,\,\beta)$ **のガンマ分布**に従うとは, X の p.d.f. $\mathsf{p}(\,\cdot\,|\,\alpha,\,\beta)$ が

$$\mathsf{p}(x|\,\alpha,\,\beta) = \mathsf{p}(x) = \begin{cases} \dfrac{1}{\beta^\alpha\Gamma(\alpha)}x^{\alpha-1}e^{-x/\beta} & (x>0) \\ 0 & (\text{その他の場合}) \end{cases}$$

のときをいう．ただし

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\,\mathrm{d}x \qquad (\alpha > 0)$$

である．このことを $X \sim \mathsf{Ga}(\alpha,\,\beta)$ と記す．$\lambda > 0$ とし, $\alpha = 1,\, \beta = \dfrac{1}{\lambda}$ とおく．このとき, $\mathsf{Ga}(1,\,1/\lambda)$ を**母数** λ **の指数分布**といい, $\mathsf{Exp}(\lambda)$ と書く．また, $p \in \mathbb{N}$ とし, $\alpha = \dfrac{p}{2},\, \beta = 2$ とおく．このとき, $\mathsf{Ga}(p/2,\,2)$ を**自由度** p **の** χ^2 **分布**といい, χ_p^2 と記す.

問 1.10. $\mathsf{p}(x|\,\alpha,\,\beta)$ をガンマ分布 $\mathsf{Ga}(\alpha,\,\beta)\,(\alpha>0,\,\beta>0)$ の p.d.f. としたとき, $\int_0^\infty \mathsf{p}(x|\,\alpha,\,\beta)\,\mathrm{d}x = 1$ を示せ.

1.5　2 次元の分布

1.5.1　同時確率関数と確率密度関数

確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の離散型確率変数 (X, Y) の対に対して, 同時 p.m.f. p を

	Y = 0	Y = 1	
X = 0	1/9	2/9	1/3
X = 1	2/9	4/9	2/3
	1/3	2/3	

で定める. たとえば

$$\mathsf{p}(1, 1) = \mathsf{Pr}(X = 1, Y = 1) = \frac{4}{9}$$

である.

定義 1.44. 確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率変数 X, Y は連続型とする. $\mathbb{R}^2$ 上の非負値実数値関数 p が確率ベクトル (X, Y) の **同時確率密度関数** (同時 p.d.f.) であるとは, 次の条件をみたすときをいう.

(1) $\mathsf{p}(x, y) \geq 0 \, (\forall (x, y) \in \mathbb{R}^2)$.

(2) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathsf{p}(x, y) \, \mathrm{d}x \mathrm{d}y = 1$.

(3) $\forall A \in \mathcal{B}(\mathbb{R}^2)$ に対して

$$\mathsf{Pr}((X, Y) \in A) = \int\int_A \mathsf{p}(x, y) \, \mathrm{d}x \mathrm{d}y.$$

注意 1.45. (X, Y) の同時累積分布関数 (同時 c.d.f.) F を

$$\begin{aligned} \mathsf{F}(x, y) &:= \mathsf{Pr}(X \leq x, Y \leq y) \qquad (1.8) \\ &:= \mathsf{Pr}(\{\omega \in \Omega;\, X(\omega) \leq x, Y(\omega) \leq y\}) \qquad (\forall (x, y) \in \mathbb{R}^2) \end{aligned}$$

で定義する. (X, Y) の同時分布 P を

$$\mathsf{P}(A) = \mathsf{Pr}\big((X, Y) \in A\big) \quad (\forall A \in \mathcal{B}(\mathbb{R}^2)) \qquad (1.9)$$

で定義する. どの確率変数の c.d.f. または確率分布であるかを明示したいときには, $\mathsf{F}^{(X,Y)}$ または $\mathsf{P}^{(X,Y)}$ と記す.

細かなことであるが, (1.8) の F によって $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ 上の確率測度 P を一意的に定めていることができる. この点に関しては補遺 C 章の測度論の知識が必要となる. □

例 1.46. 連続型確率変数のベクトル (X, Y) は同時 p.d.f.

$$\mathsf{p}(x, y) = \begin{cases} x + y & (0 < x < 1,\ 0 < y < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

を持つとする．このとき集合 $\{(x, y) \in \mathbb{R}^2;\ \mathsf{p}(x, y) > 0\}$ に注意すれば

$$\begin{aligned} \int_0^1 \int_0^1 \mathsf{p}(x, y)\,\mathrm{d}x\mathrm{d}y &= \int_0^1 \left[\int_0^1 x\,\mathrm{d}x\right] \mathrm{d}y + \int_0^1 \left[\int_0^1 y\,\mathrm{d}y\right] \mathrm{d}x \\ &= \int_0^1 \frac{1}{2}\,\mathrm{d}y + \int_0^1 \frac{1}{2}\,\mathrm{d}x = 1 \end{aligned}$$

となる． □

1.5.2 周辺分布

定義 1.47. (X, Y) を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上で定義された確率ベクトルとする．
(1) (X, Y) が離散型で同時 p.m.f. p を持つとする．X の**周辺確率関数** (周辺 p.m.f.) を

$$\mathsf{p}^X(x) = \mathsf{Pr}(X = x) = \sum_{y \in S_Y} \mathsf{Pr}(X = y,\ Y = y) = \sum_{y \in S_Y} \mathsf{p}(x, y) \qquad (\forall x \in S_X)$$

で定義する．ただし

$$\begin{aligned} S_X &:= \{x \in \mathbb{R};\ \text{ある } y \in \mathbb{R} \text{ に対して},\ \mathsf{p}(x, y) > 0\} \\ S_Y &:= \{y \in \mathbb{R};\ \text{ある } x \in \mathbb{R} \text{ に対して},\ \mathsf{p}(x, y) > 0\} \end{aligned}$$

である．
(2) (X, Y) は連続型とし，同時 p.d.f. p を持つとする．このとき X の**周辺確率密度関数** (周辺 p.d.f.) を

$$\mathsf{p}^X(x) = \int_{-\infty}^{\infty} \mathsf{p}(x, y)\,\mathrm{d}y \qquad (\forall x \in \mathbb{R})$$

で定義し，Y の**周辺確率密度関数** (周辺 p.d.f.) を

$$\mathsf{p}^Y(x) = \int_{-\infty}^{\infty} \mathsf{p}(x, y)\,\mathrm{d}x \qquad (\forall y \in \mathbb{R})$$

で定義する．

注意 1.48. 連続型確率ベクトル (X, Y) に対して

$$\begin{aligned}\mathsf{F}^X(x) &:= \mathsf{Pr}(X \leq x) = \int\int_{(s,\,t)\in\mathbb{R}^2;\, s\leq x} \mathsf{p}(s,\,t)\,\mathrm{d}s\mathrm{d}t \\ &= \int_{-\infty}^{x} \left\{ \int_{-\infty}^{\infty} \mathsf{p}(s,\,t)\,\mathrm{d}s \right\} \mathrm{d}t = \int_{-\infty}^{x} \mathsf{p}^X(s)\,\mathrm{d}s\end{aligned}$$

となるので, X の周辺 p.d.f. と X の p.d.f. は同じである. □

注意 1.49. F を確率ベクトル (X, Y) の同時 c.d.f. とし, F^X を X の周辺 c.d.f. とする. このとき $x \in \mathbb{R}$ と $y \in \mathbb{R}$ に対して

$$\mathsf{F}^X(x) = \lim_{y\to\infty} \mathsf{F}(x,\,y); \qquad \mathsf{F}^Y(y) = \lim_{x\to\infty} \mathsf{F}(x,\,y)$$

が成立する. □

問 1.11. 注意 1.49 の式を証明せよ.

1.5.3　独立性な分布と条件付き分布

定義 1.50. 確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の 2 つの確率変数 X と Y は**独立**であるとは, $\forall A, B \in \mathcal{B}(\mathbb{R})$ に対して

$$\mathsf{Pr}(X \in A,\, Y \in B) = \mathsf{Pr}(X \in A)\mathsf{Pr}(Y \in B)$$

が成り立つときをいう. 独立でないとき X と Y は**従属**であるという.

定理 1.51. p を確率ベクトル (X, Y) の同時 p.d.f. とし, p^X と p^Y を X と Y それぞれの周辺 p.d.f. とする. このとき

$$X \text{ と } Y \text{ は独立} \Leftrightarrow \mathsf{p}(x,\,y) = \mathsf{p}^X(x)\mathsf{p}^Y(y) \quad (\forall x,\, y \in \mathbb{R}).$$

となる.

Proof. 独立ならば, 同時 p.d.f. が周辺 p.d.f. の積で表現されることの証明は易しい. 逆はこの講義の範囲 (測度の拡張の議論が必要となる!) を超えるので, 信じることにする. □

注意 1.52. (1) 連続型の場合には, 測度 0 の集合を除いて[13]p.d.f. は定義されるので, 定理 1.51 の書き方はやや数学的な厳密性に欠ける.
(2) 離散型確率変数の場合には, p.d.f. を p.m.f. に置き換えればよい. □

[13] この用語は定義されていない. 「有限個の点を除いて」と理解してもこの講義の内容の範囲では問題ない.

例 1.53. 連続型確率変数 X と Y は独立で同時 p.d.f.

$$\mathsf{p}(x) = \begin{cases} 2x & (0 < x < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

を持つとする．このとき確率 $\mathsf{Pr}(X + Y \leq 1)$ を求めてみよう．独立性より (X, Y) の同時 p.d.f. は

$$\mathsf{p}^{(X,Y)}(x, y) = \begin{cases} 4xy & (0 < x < 1,\, 0 < y < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

となる．したがって

$$\begin{aligned} \mathsf{Pr}(X + Y \leq 1) &= \int\int_{x+y\leq 1} \mathsf{p}^{(X,Y)}(x, y)\,\mathrm{d}x\mathrm{d}y = 4\int_0^1 x\left\{\int_0^{1-x} y\,\mathrm{d}y\right\}\mathrm{d}x \\ &= 4\int_0^1 x\frac{(1-x)^2}{2}\,\mathrm{d}x = \frac{1}{6} \end{aligned}$$

となる．□

定理 1.54. 連続型確率ベクトル (X, Y) は同時 p.d.f. $\mathsf{p}(x, y)$ を持つとする．ただし $\{(x, y) \in \mathbb{R}^2; \mathsf{p}(x, y) > 0\}$ は矩形[14]とする．このときある非負値関数 q と r が存在して

$$\mathsf{p}(x, y) = q(x)r(y)$$

と書けるとき, X と Y は独立である．

Proof. 積分をして確かめればよい．□

注意 1.55. 定理 1.54 の主張も「ほとんど至る所」でよい．□

例 1.56. 連続型確率ベクトル (X, Y) は同時 p.d.f.

$$\mathsf{p}(x, y) = \begin{cases} 2e^{-(x+y)} & (x > 0,\, y > 0) \\ 0 & (\text{その他の場合}) \end{cases}$$

を持つとする．$\{(x, y) \in \mathbb{R}^2; \mathsf{p}(x, y) > 0\} = [0, \infty)^2$ なので

$$q(x) = \begin{cases} 2e^{-x} & (x > 0) \\ 0 & (\text{その他の場合}) \end{cases};$$

$$r(y) = \begin{cases} e^{-y} & (x > 0) \\ 0 & (\text{その他の場合}) \end{cases}$$

[14] 有界でなくともよい．

と取れば

$$\mathsf{p}(x,\,y) = q(x)r(y)$$

となるので, X と Y は独立である. □

定義 1.57. (1) 離散型確率ベクトル $(X,\,Y)$ は同時 p.m.f. $\mathsf{p}(x,\,y)$ を持つとする. $\mathsf{p}_Y(y) > 0$ なる y に対して, $Y = y$ を与えたときの X の**条件付き確率関数** (条件付き p.m.f.) を

$$\mathsf{p}^{X|Y}(x|\,y) = \mathsf{Pr}(X = x|\,Y = y) = \frac{\mathsf{Pr}(X = x,\,Y = y)}{\mathsf{Pr}(Y = y)} = \frac{\mathsf{p}(x,\,y)}{\mathsf{p}^Y(y)} \quad (x \in \mathbb{R})$$

で定義する.

(2) 連続型確率ベクトル $(X,\,Y)$ は同時 p.d.f. $\mathsf{p}(x,\,y)$ を持つとする. $\mathsf{p}^Y(y) > 0$ なる y に対して, $Y = y$ を与えたときの X の**条件付き確率密度関数** (条件付き p.d.f.) を

$$\mathsf{p}^{X|Y}(x|\,y) = \frac{\mathsf{p}(x,\,y)}{\mathsf{p}^Y(y)} \quad (x \in \mathbb{R})$$

で定義する.

注意 1.58. $(X,\,Y)$ が連続型確率ベクトルのとき, $\mathsf{p}^Y(y) > 0$ なる $y \in \mathbb{R}$ に対して, $Y = y$ を与えたときの事象 $\{X \in A\}$ の条件付き確率を

$$\mathsf{Pr}(X \in A|\,Y = y) := \int_A \mathsf{p}^{X|Y}(x|\,y)\,\mathrm{d}x \qquad \bigl(\forall A \in \mathcal{B}(\mathbb{R})\bigr)$$

と形式的に定義する. □

例 1.59. 連続型確率ベクトル $(X,\,Y)$ は同時 p.d.f.

$$\mathsf{p}(x,\,y) = \begin{cases} x + y & (0 < x < 1,\, 0 < y < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

を持つとする. このとき $\mathsf{Pr}(X \leq 1/4|\,Y = 1/3)$ を求めてみよう. まず $0 < y < 1$ に対して

$$\mathsf{p}^Y(y) = \int_{-\infty}^{\infty} \mathsf{p}(x,\,y)\,\mathrm{d}x = y + \int_0^1 y\,\mathrm{d}x = y + \frac{1}{2}$$

となる. したがって

$$\mathsf{p}^Y(y) = \begin{cases} y + \frac{1}{2} & (0 < y < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

となる．よって $0<y<1$ と $0<x<1$ に対して

$$\mathsf{p}^{X|Y}(x|\,y)=\frac{\mathsf{p}(x,\,y)}{\mathsf{p}^{Y}(y)}=\frac{x+y}{y+\frac{1}{2}}$$

となる．これより $0<y<1$ のとき

$$\mathsf{p}^{X|Y}(x|\,y)=\begin{cases}\dfrac{x+y}{y+\frac{1}{2}} & (0<x<1)\\ 0 & (\text{その他の場合})\end{cases}$$

となる．注意 1.58 より

$$\mathsf{Pr}\left(X\leq\frac{1}{4}\middle|\,Y=\frac{1}{3}\right)=\int_{0}^{1/4}p^{X|Y}\left(x\middle|\,\frac{1}{3}\right)\mathrm{d}x=\int_{0}^{1/4}\frac{x+\frac{1}{3}}{\frac{1}{3}+\frac{1}{2}}\,\mathrm{d}x=\frac{11}{80}$$

となる． □

1.6　多次元分布と i.i.d. 標本

$X_1,\,X_2,\,\ldots,\,X_n$ を確率空間 $(\Omega,\,\mathcal{A},\,\mathsf{Pr})$ 上の確率変数とし

$$\boldsymbol{X}=(X_1,\,X_2,\,\ldots,\,X_n)^{\top}$$

と書く．$\boldsymbol{X}$ を確率ベクトルという．

注意 1.60. 本講義録では, ベクトルは縦ベクトルとする． □

定義 1.61. $X_1,\,X_2,\,\ldots,\,X_n$ が独立であるとは，すべての Borel 集合 $A_1,\,A_2,\,\ldots,\,A_n\in\mathcal{B}(\mathbb{R})$ に対して

$$\mathsf{Pr}(X_1\in A_1,\,X_2\in A_2,\,\ldots,\,X_n\in A_n)=\prod_{j=1}^{n}\mathsf{Pr}(X_j\in A_j) \tag{1.10}$$

が成立するときである．

注意 1.62. $\boldsymbol{X}$ の同時 p.d.f. を $\mathsf{p}(x_1,\,x_2,\,\ldots,\,x_n)$ と書き，各 $X_j\,(j=1,\,2,\,\ldots,\,n)$ の周辺 p.d.f. を p^{X_j} と書くことにする．(1.10) を示すためには

$$\mathsf{p}(x_1,\,x_2,\,\ldots,\,x_n)=\prod_{j=1}^{n}\mathsf{p}^{X_j}(x_j)\quad(\forall x_1,\,x_2,\,\ldots,\,x_n\in\mathbb{R})$$

を示せばよい． □

定義 1.63. $X_1, X_2, \ldots, X_n$ は独立で, 各 $X_j\,(j = 1, 2, \ldots, n)$ は同じ c.d.f. F を持つとき, $X_1, X_2, \ldots, X_n$ は**独立同一分布に従う** (i.i.d. =identically and independently distributed) といい

$$X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{F}$$

と書く[15]. $X_1, X_2, \ldots, X_n$ は累積分布関数 F から標本の大きさが n のランダム標本ともいう.

1.6.1　重要な多次元分布モデル

この本で取り上げる代表的な多次元分布モデルをあげておく. 離散型分布は多項分布, 連続型分布は多変量正規分布である.

多項分布

2 項分布を多変量にしたものが多項分布である. $d, n \in \mathbb{N}$ とし, $\boldsymbol{p} = (p_1, p_2, \ldots, p_d)$, $p_j\,(0 \le p_j \le 1; , j = 1, 2, \ldots, d)$, $\sum_{j=1}^d p_j = 1$ とする. $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)$ は**母数** $(d, \boldsymbol{p})$ **の多項分布**に従うとは, $\boldsymbol{X}$ の p.m.f. が

$$\mathsf{p}(\boldsymbol{x} | n, d, \boldsymbol{p}) = \begin{cases} \begin{pmatrix} d \\ x_1, x_2, \ldots, x_d \end{pmatrix} p_1^{x_1} p_2^{x_2} \times \cdots \times p_d^{x_d} & \begin{pmatrix} x_j = 0, 1, \ldots, n; \\ j = 1, 2, \ldots, d \end{pmatrix} \\ 0 & (\text{その他の場合}) \end{cases}$$

で与えられたときをいう. ただし

$$\begin{pmatrix} d \\ x_1 x_2, \ldots, x_d \end{pmatrix} = \frac{n!}{x_1! x_2! \cdots x_d!}, \quad 0! = 1$$

である. これを $\boldsymbol{X} \sim \mathsf{Multi}_d(n, \boldsymbol{p})$ と記す.

補題 1.64. $\boldsymbol{X} \sim \mathsf{Multi}_d(n, \boldsymbol{p})$ とする. ただし, $\boldsymbol{p} = (p_1, p_2, \ldots, p_d)$ と $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)$ とする. このとき $X_j\,(j = 1, 2, \ldots, d)$ の周辺分布は $\mathsf{Bino}(n, p_j)$ である.

Proof. x_2 から x_d について同時 p.m.f. の和をとればよい. □

[15]p.d.f. p を使い

$$X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{p}$$

と書く. $\sim$ の右側には, 確率測度/確率分布/p.d.f./p.m.f./$\mathsf{N}(0, 1)$ など分布を特定するものを書いてよいことにする.

多変量正規分布

$Z_1,\, Z_2,\, \ldots,\, Z_d \overset{\text{i.i.d.}}{\sim} \mathsf{N}(0,\, 1)$ に対して

$$\boldsymbol{Z} := \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_d \end{pmatrix} \tag{1.11}$$

と定める[16]. すると $\boldsymbol{Z}$ の同時 p.d.f. は

$$\begin{aligned} \mathsf{p}^{\boldsymbol{Z}}(\boldsymbol{z}) = \mathsf{p}^{\boldsymbol{Z}}(z_1,\, z_2,\, \ldots,\, z_d) &= \prod_{j=1}^{d} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} z_j^2\right\} \\ &= \frac{1}{(2\pi)^{d/2}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{d} z_j^2\right\} \\ &= \frac{1}{(2\pi)^{d/2}} \exp\left\{-\frac{1}{2}\boldsymbol{z}^\top \boldsymbol{z}\right\} \quad (\boldsymbol{z} \in \mathbb{R}^d) \end{aligned}$$

と書ける．ただし $\boldsymbol{z} = (z_1,\, z_2,\, \ldots,\, z_d)^\top$ である．これを $\boldsymbol{Z} \sim \mathsf{N}_d(\mathbf{0},\, \boldsymbol{I}_d)$ と記す．ただし $\boldsymbol{I}_d$ は d 次単位行列である．さらに定義より

$$\int \cdots \int_{\mathbb{R}^d} \mathsf{p}^{\boldsymbol{Z}}(z_1,\, z_2,\, \ldots,\, z_d)\, \mathrm{d}z_1 \mathrm{d}z_2 \cdots d\mathrm{z}_d = 1$$

となっていることもわかる．

次に

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_d \end{pmatrix} \in \mathbb{R}^d, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1d} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{d1} & \sigma_{d2} & \cdots & \sigma_{dd} \end{pmatrix}, \quad \begin{array}{l} \sigma_{ij} = \sigma_{ji} \\ (i,\, j = 1,\, 2, \ldots,\, d) \end{array}$$

とする．ただし $\boldsymbol{\Sigma}$ は正定値対称行列[17]とする．　このとき, ある $d \times d$ の正則行列 $\mathbf{A}$ が存在して

$$(1)\ \mathbf{A}\ \text{は対称行列},\ (2)\ \boldsymbol{\Sigma} = \mathbf{A}^2$$

と取れる．これを $\boldsymbol{\Sigma}$ の平方根といい, $\boldsymbol{\Sigma}^{1/2}$ と書くことにする．これを用いて

$$\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{Z} \tag{1.12}$$

[16] この節ではベクトルは縦とする．下の式では縦ベクトルと横ベクトルを混ぜて表現している．これは記号の乱用であるが, $f((z_1,\, z_2,\, \ldots,\, z_d)^\top)$ などと書くのは煩わしい．

[17] $\forall \boldsymbol{x} \in \mathbb{R}^d\, (\boldsymbol{x} \neq \mathbf{0}_d)$ に対して, $\boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x} > 0$ が成立する．したがって, 逆行列 $\boldsymbol{\Sigma}^{-1}$ が存在する．

と定めたとき, $\boldsymbol{X}$ の分布を $\mathsf{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ と記す．このとき $\boldsymbol{X}$ の同時 p.d.f. は

$$\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{d/2}\det(\boldsymbol{\Sigma})^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\} \quad (\boldsymbol{x} \in \mathbb{R}^d) \tag{1.13}$$

で与えられる．ただし, $\det(\,\cdot\,)$ は行列式を表す．この分布を**平均ベクトル $\boldsymbol{\mu}$, 分散共分散行列 Σ の d 変量正規分布**という．

注意 1.65. (1.13) の導出は以下のように行うことができる．一般に $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_d)^\top$ を d 次元確率ベクトルとし, その同時 p.d.f. を $\mathsf{p}^{\boldsymbol{Z}}$ と書くことにする．さらに $\mathbb{X} := \{\boldsymbol{z} \in \mathbb{R}^d;\, \mathsf{p}^{\boldsymbol{Z}}(\boldsymbol{z}) > 0)\}$ とし, 関数

$$\boldsymbol{h}(\,\cdot\,) = (h_1(\,\cdot\,), h_2(\,\cdot\,), \ldots, h_d(\,\cdot\,))^\top : \mathbb{X} \to \boldsymbol{h}(\mathbb{X})$$

は 1 対 1 とし, $j = 1, 2, \ldots, d$ に対して

$$X_j = h_j(Z_1, Z_2, \ldots, Z_d); \quad \boldsymbol{X} = (X_1, X_2, \ldots, X_d)^\top$$

とおく．$\boldsymbol{h}$ は 1 対 1 なので, $\boldsymbol{h}$ の逆写像を

$$\boldsymbol{h}^{-1} = (h_1^{-1}, h_2^{-1}, \ldots, h_d^{-1})^\top : \qquad \boldsymbol{h}(\mathbb{X}) \to \mathbb{X}$$

が存在して, $\boldsymbol{Z} = \boldsymbol{h}^{-1}(\boldsymbol{X})$ となる．$\boldsymbol{X}$ の同時 p.d.f. を求めるために, $\boldsymbol{h}^{-1}(\boldsymbol{x})$ の Jacobian $\mathbf{J}_{\boldsymbol{h}^{-1}}(\boldsymbol{x})$ を

$$\mathbf{J}_{\boldsymbol{h}^{-1}}(\boldsymbol{x}) = \det\begin{pmatrix} \frac{\partial h_1^{-1}(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial h_1^{-1}(\boldsymbol{x})}{\partial x_d} \\ \vdots & \ddots & \vdots \\ \frac{h_d^{-1}(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial h_d^{-1}(\boldsymbol{x})}{\partial x_d} \end{pmatrix}$$

で定め, $\mathbf{J}_{\boldsymbol{h}^{-1}}(\boldsymbol{x}) \neq 0\,(x \in \mathbb{X})$ と仮定[18]する．このとき

$$\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}) = \left|\mathbf{J}_{\boldsymbol{h}^{-1}}(\boldsymbol{x})\right|\mathsf{p}^{\boldsymbol{Z}}(h^{-1}(\boldsymbol{x})) \tag{1.14}$$

となることが補題の定理 C.84 からわかる．

$\boldsymbol{h}(\boldsymbol{z}) = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{z} \Leftrightarrow \boldsymbol{h}^{-1}(\boldsymbol{x}) = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu})$ より

$$\mathbf{J}_{\boldsymbol{h}^{-1}}(\boldsymbol{x}) = \det(\boldsymbol{\Sigma})^{-1/2} = \frac{1}{\sqrt{\det(\boldsymbol{\Sigma})}}$$

を (1.14) に代入すれば, (1.13) はわかる．以上の議論から

$$\int\int\cdots\int_{\mathbb{R}^d}\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_d = 1$$

となっていることもわかる． □

[18] 実は, a.e $\mathbb{X}$ でよい.

問 1.12. $h^{-1}(\boldsymbol{x}) = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x}-\boldsymbol{\mu})$ のとき

$$J_{h^{-1}}(\boldsymbol{x}) = \frac{1}{\sqrt{\det\boldsymbol{\Sigma}}}$$

を示せ.

注意 1.66. $\boldsymbol{\Sigma}$ を半正定値としたときに, 多変量正規分布を (1.11) と変換 (1.12) を用いて定義することがことができる. しかし, $\det(\boldsymbol{\Sigma}) = 0$ のときは, $\forall B \in \mathcal{B}(\mathbb{R}^d)$ に対して, 確率 $\mathsf{Pr}(\boldsymbol{X} \in B)$ は定義できるが, $\boldsymbol{X}$ は同時 p.d.f. を持たないことが知られている. このような分布を退化した多変量正規分布という. □

定理 1.67. $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)^\top \sim \mathsf{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ とする. ただし $\boldsymbol{\mu} \in \mathbb{R}^d$ で, $\boldsymbol{\Sigma}$ は $d \times d$ の正定値行列である. このとき次が成立する.

(1) $X_j\,(j = 1, 2, \ldots, d) \sim \mathsf{N}(\mu_j, \sigma_{jj})$. ただし μ_j は $\boldsymbol{\mu}$ の第 j 成分, σ_{jj} は $\boldsymbol{\Sigma}$ の第 j 対角成分である.

(2) $X_\ell = x_\ell\,(\ell = 1, 2, \ldots, d)$ を与えたときの $X_j\,(j \neq \ell)$ の条件付き分布は

$$X_j|\,X_\ell = x_\ell \sim \mathsf{N}\left(\mu_j + \frac{\sigma_{\ell\ell}}{\sigma_{jj}}(x_\ell - \mu_\ell),\, \sigma_{jj} - \frac{\sigma_{j\ell}\sigma_{\ell j}}{\sigma_{\ell\ell}}\right)$$

となる.

(3) 定数ベクトル $\boldsymbol{c} \in \mathbb{R}^d\,(\boldsymbol{c} \neq \boldsymbol{0}_d)$ に対して

$$\boldsymbol{c}^\top\boldsymbol{X} \sim \mathsf{N}(\boldsymbol{c}^\top\boldsymbol{\mu},\, \boldsymbol{c}^\top\boldsymbol{\Sigma}\boldsymbol{c})$$

となる. ただし, $\boldsymbol{0}_d$ は $\mathbb{R}^d$ の零ベクトルである.

(4) $V := (\boldsymbol{X}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu}) \sim \chi^2_d$ である.

Proof. (1) $j = 1$ として示す. $\boldsymbol{\Sigma}$ の $(i, j)\,(i, j = 1, 2, \ldots, d)$ 成分を σ_{ij} と書き, $(d-1)\times(d-1)$ 行列 $\boldsymbol{\Sigma}_1$ と $(d-1)\times 1$ ベクトル $\boldsymbol{\sigma}_1$ をそれぞれ

$$\boldsymbol{\Sigma}_1 = (\sigma_{ij})_{i,j=2,3,\ldots,d}, \quad \boldsymbol{\sigma}_1 = (\sigma_{21}, \sigma_{31}, \ldots, \sigma_{d1})^\top$$

で定める. すると

$$\boldsymbol{\Sigma} = \left(\begin{array}{c|c} \sigma_{11} & \boldsymbol{\sigma}_1^\top \\ \hline \boldsymbol{\sigma}_1 & \boldsymbol{\Sigma}_2 \end{array}\right)$$

となる．このとき

$$\boldsymbol{\Sigma} = \left(\begin{array}{c|c} 1 & \mathbf{0}_{d-1}^\top \\ \hline \sigma_{11}^{-1}\boldsymbol{\sigma}_1 & \boldsymbol{I}_{d-1} \end{array}\right)\left(\begin{array}{c|c} \sigma_{11} & \mathbf{0}_{d-1}^\top \\ \hline \mathbf{0}_{d-1} & \boldsymbol{\Sigma}_2 - \sigma_{11}^{-1}\boldsymbol{\sigma}_1\boldsymbol{\sigma}_1^\top \end{array}\right)\left(\begin{array}{c|c} 1 & \sigma_{11}^{-1}\boldsymbol{\sigma}_1^\top \\ \hline \mathbf{0}_{d-1} & \boldsymbol{I}_{d-1} \end{array}\right)$$

と書ける．ただし $\boldsymbol{I}_{d-1}$ は $(d-1)$ 次の単位行列で, $\mathbf{0}_{d-1}$ は $(d-1)\times 1$ の零ベクトルである．$\boldsymbol{\Sigma}$ が正定値ならば, $\boldsymbol{\Sigma}_2 - \sigma_{11}^{-1}\boldsymbol{\sigma}_1\boldsymbol{\sigma}_1^\top$ も正定値であるので

$$\begin{aligned}\boldsymbol{\Sigma}^{-1} = &\left(\begin{array}{c|c} 1 & -\sigma_{11}^{-1}\boldsymbol{\sigma}_1^\top \\ \hline \mathbf{0}_{d-1} & \boldsymbol{I}_{d-1} \end{array}\right)\left(\begin{array}{c|c} \sigma_{11}^{-1} & \mathbf{0}_{d-1}^\top \\ \hline \mathbf{0}_{d-1} & \left\{\boldsymbol{\Sigma}_2 - \sigma_{11}^{-1}\boldsymbol{\sigma}_1\boldsymbol{\sigma}_1^\top\right\}^{-1} \end{array}\right) \\ &\times \left(\begin{array}{c|c} 1 & \mathbf{0}_{d-1}^\top \\ \hline -\sigma_{11}^{-1}\boldsymbol{\sigma}_1 & \boldsymbol{I}_{d-1} \end{array}\right)\end{aligned} \tag{1.15}$$

となる．$\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_d)^\top$ と $\boldsymbol{x} = (x_1, x_2, \dots, x_d)^\top$ と書いたときに, $\boldsymbol{\mu}_1 = (\mu_2, \mu_3, \dots, \mu_d)^\top$ と $\boldsymbol{x}_1 = (x_2, x_3, \dots, x_d)^\top$ とおく．(1.15) から

$$\begin{aligned}\mathrm{tr}\,(\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) &= \frac{1}{\sigma_{11}}(x_1-\mu_1)^2 \\ &\quad + \mathrm{tr}\left[\left(\boldsymbol{x}_1 - \boldsymbol{\mu}_1 - \frac{x_1-\mu_1}{\sigma_{11}}\boldsymbol{\sigma}_1\right)^\top\left\{\boldsymbol{\Sigma}_2 - \sigma_{11}^{-1}\boldsymbol{\sigma}_1\boldsymbol{\sigma}_1^\top\right\}^{-1}\right. \qquad (1.16) \\ &\qquad \left.\times\left(\boldsymbol{x}_1 - \boldsymbol{\mu}_1 - \frac{x_1-\mu_1}{\sigma_{11}}\boldsymbol{\sigma}_1\right)\right] \\ &=: \frac{1}{\sigma_{11}}(x_1-\mu_1)^2 + g_{d-1}(\boldsymbol{x}_1) \qquad (1.17)\end{aligned}$$

となる．ただし

$$\begin{aligned}g_{d-1}(\boldsymbol{x}_1) = \mathrm{tr}&\left[\left(\boldsymbol{x}_1 - \boldsymbol{\mu}_1 - \frac{x_1-\mu_1}{\sigma_{11}}\boldsymbol{\sigma}_1\right)^\top\left\{\boldsymbol{\Sigma}_2 - \sigma_{11}^{-1}\boldsymbol{\sigma}_1\boldsymbol{\sigma}_1^\top\right\}^{-1}\right. \\ &\left.\times\left(\boldsymbol{x}_1 - \boldsymbol{\mu}_1 - \frac{x_1-\mu_1}{\sigma_{11}}\boldsymbol{\sigma}_1\right)\right]\end{aligned}$$

である．このことから

$$\begin{aligned}&\int_{\mathbb{R}^{d-1}} \exp\left[-\frac{1}{2}\mathrm{tr}\,(\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]\mathrm{d}x_2\,\mathrm{d}x_3\,\times\,\mathrm{d}x_d \\ &\quad = \exp\left[-\frac{(x_1-\mu_1)^2}{2\sigma_{11}}\right]\int_{\mathbb{R}^{d-1}}\exp\left[-\frac{1}{2}g_{d-1}\right]\mathrm{d}x_2\,\mathrm{d}x_3\,\times\,\mathrm{d}x_d \\ &\quad = \exp\left[-\frac{(x_1-\mu_1)^2}{2\sigma_{11}}\right]\det\left(2\pi\big(\boldsymbol{\Sigma}_2 - \sigma_{11}^{-1}\boldsymbol{\sigma}_1\boldsymbol{\sigma}_1^\top\big)\right)^{1/2}\end{aligned}$$

がわかる．よって

$$\int_{\mathbb{R}^{d-1}} \mathsf{p}^{\boldsymbol{X}}\big(\boldsymbol{x}|\, \boldsymbol{\mu},\, \boldsymbol{\Sigma}\big)\, \mathrm{d}x_2\, \mathrm{d}x_3,\, \cdots \mathrm{d}x_d = \frac{1}{\sqrt{2\pi\sigma_{11}}} \exp\left[\frac{(x_1-\mu_1)^2}{2\sigma_{11}}\right]$$

がわかる．

(2) (1.17) からわかる．

(3) $\boldsymbol{c}^\top \boldsymbol{X}$ の積率を計算すればよい．

(4) (3) から

$$\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{X}-\boldsymbol{\mu}) \sim \mathsf{N}_d(\boldsymbol{0}_d,\, \boldsymbol{I}_d)$$

となる．あとは χ^2 分布の定義からわかる． □

(1.14) を用いた例題を述べておく

例 1.68. 連続型確率ベクトル $(X_1,\, X_2)$ は同時 p.d.f.

$$\mathsf{p}(x_1,\, x_2) = \begin{cases} 1 & (0 < x_1 < 1,\, 0 < x_2 < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

を持つとする．さらに

$$\begin{gathered} y_1 = h_1(x_1,\, x_2) = x_1 + x_2,\, y_2 = h_2(x_1,\, x_2) = x_2 \\ \Leftrightarrow x_1 = h_1^{-1}(y_1,\, y_2),\, x_2 = h_2^{-1}(y_1,\, y_2) \end{gathered}$$

とし，$\boldsymbol{Y} = (h_1(X_1,\, X_2),\, h_2(X_1,\, X_2))^\top$, $\boldsymbol{y} = (y_1,\, y_2)^\top$ とする．このとき

$$J_{\boldsymbol{h}^{-1}}(\boldsymbol{y}) = \det\begin{pmatrix} \frac{\partial h_1^{-1}(\boldsymbol{y})}{\partial y_1} & \frac{\partial h_1^{-1}(\boldsymbol{y})}{\partial y_2} \\ \frac{\partial h_2^{-1}(\boldsymbol{y})}{\partial y_1} & \frac{\partial h_2^{-1}(\boldsymbol{y})}{\partial y_2} \end{pmatrix} = \det\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = 1$$

より

$$\begin{aligned} \mathsf{p}^{\boldsymbol{Y}}(\boldsymbol{y}) &= \begin{cases} 1 & (0 < y_1 - y_2 < 1,\, 0 < y_2 < 1) \\ 0 & (\text{その他の場合}) \end{cases} \\ &= \begin{cases} 1 & (y_2 < y_1 < 1 + y_2,\, 0 < y_2 < 1) \\ 0 & (\text{その他の場合}) \end{cases} \end{aligned}$$

を得る．また

$$\int\!\!\int_{\mathbb{R}^2} \mathsf{p}^{\boldsymbol{Y}}(\boldsymbol{y})\, \mathrm{d}y_1 \mathrm{d}y_2 = 1$$

となっていることに注意せよ． □

1.7　正規分布に関連した分布

この節では, 正規分布に関連した分布であるガンマ分布, χ^2 分布, F 分布, t 分布を定義し, これらの基本的な性質を述べる. これらの分布は正規母集団から標本に基づく統計量の分布 (いわゆる標本分布) の議論で重要な役割を担う.

まず, ガンマ関数を導入する. $\alpha > 0$ に対してガンマ関数を

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\,\mathrm{d}x$$

で定義する. $\alpha > 1$ のとき, 部分積分により

$$\begin{aligned}\Gamma(\alpha) &= \int_0^\infty x^{\alpha-1}e^{-x}\,\mathrm{d}x = -x^{\alpha-1}e^{-x}\Big|_0^\infty + (\alpha-1)\int_0^\infty x^{\alpha-2}e^{-x}\,\mathrm{d}x \\ &= \Gamma(\alpha-1)\end{aligned}$$

となる. $\Gamma(1) = \int_0^\infty e^{-x}\,\mathrm{d}x = 1$ なので, 帰納法により

$$\Gamma(n) = (n-1)! \qquad (n = 1, 2, \ldots)$$

となる.

1.7.1　ガンマ分布

定義 1.69. (ガンマ分布) $\alpha, \beta > 0$ とする. 連続型確率変数 X が母数 (α, β) のガンマ分布に従うとは, X の p.d.f. が

$$\mathsf{p}(x|\,\alpha, \beta) = \begin{cases} \dfrac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x} & (x > 0) \\ 0 & (\text{その他の場合}) \end{cases}$$

で与えられるときをいう. これを $X \sim \mathsf{Ga}(\alpha, \beta)$ と記す.

補題 1.70. $n \geq 2$ は自然数とする. $X_1, X_2, \ldots, X_n$ は独立な確率変数, 各 $X_j \sim \mathsf{Ga}(\alpha_j, \beta)\,(j = 1, 2, \ldots, n, \alpha_j > 0, \beta > 0$ としたとき

$$X_1 + X_2 + \cdots + X_n \sim \mathsf{Ga}(\alpha_1 + \alpha_2 + \cdots + \alpha_n, \beta)$$

となる.

Proof. 証明は次の補題を用いてする. □

補題 1.71. 連続型確率変数 X, Y は独立とし，それぞれの p.d.f. を p^X, p^Y とする．このとき $X+Y$ の p.d.f. は

$$\mathsf{p}^{X+Y}(x) = \int_{-\infty}^{\infty} \mathsf{p}^X(x-y)\mathsf{p}^Y(y)\,\mathrm{d}y$$

で与えられる．

Proof. X, Y の同時 p.d.f. は $\mathsf{p}^X(x)\mathsf{p}^Y(y)$ になることに注意する．$t \in \mathbb{R}$ に対して

$$\begin{aligned}
\mathsf{Pr}\bigl(X+Y \le t\bigr) &= \int\int_{x+y\le t} \mathsf{p}^X(x)\mathsf{p}^Y(y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty} \mathsf{p}^Y(y)\left\{\int_{-\infty}^{t-y} \mathsf{p}^X(x)\,\mathrm{d}x\right\}\mathrm{d}y \\
&= \int_{-\infty}^{\infty} \mathsf{p}^Y(y)\left\{\int_{-\infty}^{t} \mathsf{p}^Y(z-y)\,\mathrm{d}z\right\}\mathrm{d}y \\
&\qquad \left(x\,\Big|_{-\infty}^{t-y} = z-y\,\Big|_{-\infty}^{t} \text{ と変換}\right) \\
&= \int_{-\infty}^{t}\left\{\int_{-\infty}^{\infty} \mathsf{p}^X(z-y)\mathsf{p}^Y(y)\,\mathrm{d}y\right\}\mathrm{d}z
\end{aligned}$$

となる．よって微積分の基本定理より

$$\mathsf{p}^{X+Y}(t) = \frac{d}{dt}\mathsf{Pr}\bigl(X+Y\le t\bigr) = \int_{-\infty}^{\infty} \mathsf{p}^X(t-y)\mathsf{p}^Y(y)\,\mathrm{d}y$$

を得る． □

補題 1.70 の証明. $X_1 \sim \mathsf{Ga}(\alpha_1, \beta)$, $X_2 \sim \mathsf{Ga}(\alpha_2, \beta)$ とし，X_1 と X_2 は独立とする．$\mathsf{Ga}(\alpha, \beta)$ の p.d.f. を $\mathsf{p}(\cdot\,|\,\alpha, \beta)$ と書く．補題 1.71 より

$$\begin{aligned}
\mathsf{p}^{X_1+X_2}(x) &= \int_{-\infty}^{\infty} \mathsf{p}(x-y|\,\alpha_1, \beta)\mathsf{p}(y|\,\alpha_2, \beta)\,\mathrm{d}y \\
&= \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} e^{-\beta x}\int_0^x (x-y)^{\alpha_1-1}y^{\alpha_2-1}\,\mathrm{d}y \\
&\qquad \bigl(x-y>0 \text{ かつ } y>0 \text{ より } 0<y<x \text{ となる.}\bigr) \\
&= \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} x^{\alpha_1+\alpha_2-1}e^{-\beta x}\int_0^1 (1-z)^{\alpha_1-1}z^{\alpha_2-1}\,\mathrm{d}z \\
&\qquad (y = xz \text{ と変換})
\end{aligned}$$

を得る．ここで $\mathsf{p}^{X_1+X_2}$ は p.d.f. であることに注意すれば

$$
\begin{aligned}
1 &= \int_0^\infty \mathsf{p}^{X_1+X_2}(x)\,\mathrm{d}x \\
&= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\left(\int_0^1 (1-z)^{\alpha_1-1}z^{\alpha_2-1}\,\mathrm{d}z\right)\beta^{\alpha_1+\alpha_2}\int_0^\infty x^{\alpha_1+\alpha_2-1}e^{-\beta x}\,\mathrm{d}x \\
&= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\left(\int_0^1 (1-z)^{\alpha_1-1}z^{\alpha_2-1}\,\mathrm{d}z\right)\int_0^\infty w^{\alpha_1+\alpha_2-1}e^{-w}\,\mathrm{d}w \\
&\quad (w=\beta z \text{ と変換}) \\
&= \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\int_0^1 (1-z)^{\alpha_1-1}z^{\alpha_2-1}\,\mathrm{d}z \qquad (\text{ガンマ関数の定義より})
\end{aligned}
$$

より

$$
\int_0^1 (1-z)^{\alpha_1-1}z^{\alpha_2-1}\,\mathrm{d}z = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)} \tag{1.18}
$$

を得る．よって

$$
X_1+X_2 \sim \mathsf{Ga}(\alpha_1+\alpha_2,\, \beta)
$$

がわかる．あとはこのことを繰り返せばよい．

注意 1.72. $\alpha_1>0,\ \alpha_2>0$ に対して

$$
B(\alpha_1,\,\alpha_2) := \int_0^1 (1-z)^{\alpha_1-1}z^{\alpha_2-1}\,\mathrm{d}z
$$

を**ベータ関数**とよぶ．すると (1.18) から

$$
B(\alpha_1,\,\alpha_2) = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)}
$$

という関係式が成り立つことがわかる． □

1.7.2　χ^2 分布

定義 1.73. $Z_1,\,Z_2,\,\ldots,\,Z_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}(0,\,1)$ とする．

$$
S = Z_1^2 + Z_2^2 + \cdots + Z_n^2
$$

の分布を**自由度 n の χ^2 分布**といい，$S \sim \chi_n^2$ と記す．

$Z \sim \mathsf{N}(0, 1)$ のとき $Y = Z^2$ の p.d.f. は

$$\mathsf{p}^Y(y) = \begin{cases} \dfrac{1}{\sqrt{2\pi}} y^{(1/2)-1} e^{-y/2} & (y > 0) \\ 0 & (\text{その他の場合}) \end{cases} \tag{1.19}$$

となる．演習問題 1.1 を参照せよ．一方, $\mathsf{Ga}(1/2, 1/2)$ の p.d.f. は

$$\mathsf{p}\left(y \,\middle|\, \frac{1}{2}, \frac{1}{2}\right) = \begin{cases} \dfrac{1}{2\Gamma\left(\frac{1}{2}\right)} y^{(1/2)-1} e^{-y/2} & (y > 0) \\ 0 & (\text{その他の場合}) \end{cases}$$

であった．これらはともに p.d.f. なので, 積分をすれば 1 となる．このことより

$$\sqrt{2}\Gamma\left(\frac{1}{2}\right) = \sqrt{2\pi} \Leftrightarrow \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

を得る．さらに補題 1.70 より

$$\chi_n^2 = \mathsf{Ga}\left(\frac{n}{2}, \frac{1}{2}\right)$$

がわかる．よって χ_n^2 の p.d.f. は

$$\mathsf{p}(x|\, n) = \begin{cases} \dfrac{\left(\dfrac{1}{2}\right)^n}{\Gamma\left(\dfrac{n}{2}\right)} x^{(n/2)-1} e^{-x/2} & (x > 0) \\ 0 & (\text{その他の場合}) \end{cases}$$

である．

1.7.3　F 分布

定義 1.74. $k, m \in \mathbb{N}$ とする．X と Y は独立で $X \sim \chi_k^2$ と $Y \sim \chi_m^2$ とする．このとき

$$F = \frac{X/k}{Y/m}$$

の分布を**自由度** (k, m) **の** F **分布**といい, $\mathsf{F} \sim \mathsf{F}(k, m)$ と記す．

注意 1.75. 上の定義では, F 分布の記号を $\mathsf{F}(k, m)$ と書いた．分布関数の記号と同じ F の字体 F を使用しているが, 文脈から判断できるので記号の乱用をする． □

補題 1.76. $k,\, m \in \mathbb{N}$ とする. $F \sim \mathsf{F}(k,\, m)$ のとき, F の p.d.f. $\mathsf{p}(x|\, k,\, m)$ は

$$\mathsf{p}(x|\, k,\, m) = \begin{cases} \dfrac{\Gamma\left(\dfrac{k+m}{2}\right)}{\Gamma\left(\dfrac{k}{2}\right)\Gamma\left(\dfrac{m}{2}\right)} \dfrac{k^{k/2} m^{m/2} x^{(k/2)-1}}{(m+kx)^{(k+m)/2}} & (x > 0) \\ 0 & (\text{その他の場合}) \end{cases}$$

で与えられる.

Proof. 証明は次の補題を用いて行う. □

補題 1.77. $X,\, Y$ を正値の連続型確率変数とし, それぞれの p.d.f. を $\mathsf{p}^X(x)$ と $\mathsf{p}^Y(x)$ とする. このとき

$$Z = \frac{X}{Y}$$

の p.d.f. は

$$\mathsf{p}^Z(z) = \begin{cases} \displaystyle\int_0^\infty y \mathsf{p}^X(zy) \mathsf{p}^Y(y)\, \mathrm{d}y & (z > 0) \\ 0 & (\text{その他の場合}) \end{cases}$$

で与えられる.

Proof. 仮定から $(X,\, Y)$ の同時 p.d.f. は $\mathsf{p}^X(x)\mathsf{p}^Y(y)$ となる. よって $t > 0$ に対して

$$\begin{aligned} \mathsf{Pr}\big(Z \le t\big) &= \int_0^\infty \left\{ \int_0^{ty} \mathsf{p}^X(x) \mathsf{p}^Y(y)\, \mathrm{d}x \right\} \mathrm{d}y \\ &= \int_0^\infty \left\{ \int_0^{t} \mathsf{p}^X(zy) \mathsf{p}^Y(y) y\, \mathrm{d}z \right\} \mathrm{d}y \quad (x = zy \text{ と変換}) \\ &= \int_0^t \left\{ \int_0^{\infty} \mathsf{p}^X(zy) \mathsf{p}^Y(y) y\, \mathrm{d}y \right\} \mathrm{d}z. \end{aligned}$$

ここで微積分の基本定理を用いると

$$\mathsf{p}^Z(t) = \frac{d}{dt} \mathsf{Pr}\big(Z \le t\big) = \int_0^\infty y \mathsf{p}^X(zy) \mathsf{p}^Y(y)\, \mathrm{d}y$$

となる. □

補題 1.76 の証明　補題 1.77 を用いる. $z > 0$ のとき, $Z = X/Y$ の

p.d.f. は

$$
\begin{aligned}
\mathsf{p}^Z(z) &= \int y\mathsf{p}^X(zy)\mathsf{g}^Y(y)\,\mathrm{d}y \\
&= \frac{\left(\frac{1}{2}\right)^{k/2}\left(\frac{1}{2}\right)^{m/2}}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{m}{2}\right)}\int_0^\infty y(zy)^{(k/2)-1}e^{-(zy)/2}y^{(m/2)-1}e^{-y/2}\,\mathrm{d}y \\
&= \frac{\left(\frac{1}{2}\right)^{(k+m)/2}}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{m}{2}\right)}z^{(k/2)-1}\int_0^\infty y^{(k+m)/2-1}e^{-(w+1)y/2}\,\mathrm{d}y \\
&= \frac{\left(\frac{1}{2}\right)^{(k+m)/2}}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{m}{2}\right)}z^{(k/2)-1}\int_0^\infty \left(\frac{2x}{z+1}\right)^{(k+m)/2-1}e^{-x}\frac{2}{z+1}\,\mathrm{d}x \\
&\qquad \left(x = \tfrac{z+1}{2}y\right) \text{ と変換} \\
&= \frac{1}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{m}{2}\right)}z^{(k/2)-1}(1+z)^{-(k+m)/2}\int_0^\infty x^{(k+m)/2-1}e^{-x}\,\mathrm{d}x \\
&= \frac{\Gamma\left(\frac{k+m}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{m}{2}\right)}z^{(k/2)-1}(1+z)^{-(k+m)/2}
\end{aligned}
$$

となる. $X = \frac{m}{k}Z$ なので

$$
\mathsf{p}(x|\,k,\,m) = \mathsf{p}^Z\left(\frac{k}{m}x\right)\frac{k}{m}
$$

となる. この式を整理すると主張は証明される.

1.7.4　t 分布

定義 1.78. $Z_0,\,Z_1,\,Z_2,\,\ldots,\,Z_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}(0,\,1)$ のとき

$$
T = \frac{Z_0}{\sqrt{\dfrac{1}{n}\left(Z_1^2 + Z_2^2 + \cdots + Z_n^2\right)}}
$$

を自由度 n の t 分布といい, これを $T \sim t_n$ と記す.

補題 1.79. T の p.d.f. は

$$
\mathsf{p}^T(x|\,n) = \begin{cases} \dfrac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)}\dfrac{1}{\sqrt{n}}\left(1+\dfrac{x^2}{n}\right)^{-(n+1)/2} & (x > 0) \\ 0 & (\text{その他の場合}) \end{cases}
$$

で与えられる

Proof. 定義より

$$
T^2 \sim \mathsf{F}(1,\,n)
$$

であることをまず思い出す. T の分布は対称なので

$$\mathsf{p}^T(x|\,n) = \mathsf{p}^T(-x|\,n) \qquad (x \geq 0)$$

となる. よって $x > 0$ に対して

$$2\mathsf{Pr}\big(0 \leq T \leq x\big) = \mathsf{Pr}\big(-x \leq T \leq x\big) = \mathsf{Pr}\big(T^2 \leq x^2\big)$$

となる. 上の式から $\mathsf{F}(1,\,n)$ の p.d.f. を $\mathsf{p}^{1,\,n}$ と書けば

$$2\int_0^x \mathsf{p}^T(t|\,n)\,\mathrm{d}t = \int_0^{x^2} \mathsf{p}^{1,\,n}(t)\,\mathrm{d}t$$

となる. さらに, $x = y^2$ と変換すると

$$\int_0^{x^2} \mathsf{p}^{1,\,n}(t)\,\mathrm{d}t = \int_0^x \mathsf{p}^{1,\,n}(t^2)2t\,\mathrm{d}t \tag{1.20}$$

となる. ここで補題 1.77 から

$$\mathsf{p}^{1,\,n}(x) = \frac{\Gamma\left(\dfrac{n+1}{2}\right)}{\Gamma\left(\dfrac{1}{2}\right)\Gamma\left(\dfrac{n}{2}\right)}\frac{n^{n/2}x^{-1/2}}{(n+x)^{(n+1)/2}}$$

であることを思い出す. この式を (1.20) の左辺に代入すると

$$\mathsf{p}^T(x|\,n) = \mathsf{p}^{1,\,n}(x^2)x = \frac{\Gamma\left(\dfrac{n+1}{2}\right)}{\Gamma\left(\dfrac{1}{2}\right)\Gamma\left(\dfrac{n}{2}\right)}\frac{1}{\sqrt{n}}\left(1+\frac{x^2}{n}\right)^{-(n+1)/2}$$

となる. □

1.8　章末注釈と参考文献

第 1.1 節は [20] を参照した. 第 1.2 節は [7] を参照した. 第 1.3 節から第 1.6 節は [27] を参照した. 第 1.7 節は [18] からの借用である.

1.9　演習問題

演習問題 1.1. 確率空間 $(\Omega,\,\mathcal{A},\,\mathsf{Pr})$ 上の離散型確率変 X は以下の p.m.f. を持つ :

$$\mathsf{p}(x) = \begin{cases} \begin{pmatrix} 3 \\ x \end{pmatrix}\left(\dfrac{1}{4}\right)^x\left(\dfrac{3}{4}\right)^{3-x} & (x = 0,\,1,\,2,\,3) \\ 0 & (\text{その他の場合}) \end{cases}$$

である. ただし

$$\begin{pmatrix} 3 \\ x \end{pmatrix} = \frac{3!}{x! \times (3-x)!}, \qquad 0! = 1$$

である.
(1) $\mathsf{p}(0)$, $\mathsf{p}(1)$, $\mathsf{p}(2)$, $\mathsf{p}(3)$ の値を計算せよ.
(2) $\mathsf{Pr}(0 < X < 3)$ を求めよ.

演習問題 1.2. $Z \sim \mathsf{N}(0, 1)$ とする. Z の p.d.f. を

$$\mathsf{p}(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad (-\infty < z < \infty)$$

としたとき, 以下の問いに答えよ.
(1) 任意の $y > 0$ に対して, $\mathsf{Pr}(Z^2 \leq y)$ を p も用いて表現せよ.
(2) $Y := Z^2$ とし, Y の p.d.f. を p^Y と書く. このとき, $y > 0$ に対して

$$\mathsf{p}^Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} \mathsf{Pr}(Y \leq y)$$

となることを利用して, Y の p.d.f. が (1.19) で与えられることを示せ.

演習問題 1.3. X を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の連続型確率変数とし, X の p.d.f. を

$$\mathsf{p}(x) = \begin{cases} 2x & (0 < x < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

とする. さらに, 離散型確率変数 Y を

$$Y := \mathbb{1}_{(1/2, \infty)}(X) := \begin{cases} 1 & \left(X > \frac{1}{2}\right) \\ 0 & \left(X \leq \frac{1}{2}\right) \end{cases}$$

で定める. このとき, 以下の問に答えよ.
(1) 確率変数 X の c.d.f. を求めよ. なお, X の c.d.f. を F と書くことにする.
(2) 確率変数 Y の p.m.f. を求めよ. なお, X の p.m.f. を p^Y と書くことにする.
(3) 確率 $\mathsf{Pr}\left(0 < X \leq \frac{1}{2}\right)$, $\mathsf{Pr}(Y = 0)$ を求めよ.
(4) 確率 $\mathsf{Pr}\left(0 < X \leq \frac{1}{2}, Y = 0\right)$ を求めよ.
(5) 確率変数 X と Y は独立か従属かを判定せよ.

演習問題 1.4. 大小 2 つのサイコロを投げたとき, それぞれの出る目を X と Y とおく. これらを確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上で定義されている確率変数と考える. この空間上の確率変数 Z と W を

$$Z(\omega) = \min\{X(\omega), Y(\omega)\}, \quad W(\omega) = \max\{X(\omega), Y(\omega)\} \quad (\forall \omega \in \Omega)$$

で定める. このとき, 以下の問いに答えよ.
(1) 確率ベクトル (Z, W) の同時 p.m.f. $\mathsf{p}^{(Z,W)}$ を求めよ.
(2) 確率 $\mathsf{Pr}(Z = W)$ と条件付き確率 $\mathsf{Pr}(Z = 2 | W < 5)$ を求めよ.
(3) 確率変数 Z の周辺 p.m.f. p^Z を求めよ.
(4) 確率変数 W の周辺 p.m.f. p^W を求めよ.
(5) 確率変数 Z と W は独立であるかどうかを判定せよ.

演習問題 1.5. p を自然数とする. 連続型確率変数 T は p.d.f.

$$\mathsf{p}^T(t) = \frac{\Gamma((p+1)/2)}{\Gamma(p/2)} \frac{1}{\sqrt{p\pi}} \frac{1}{(1 + t^2/p)^{(p+1)/2}}, \qquad -\infty < t < \infty$$

を持つとする. ただし, $\Gamma(\cdot)$ の Euler のガンマ関数で

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x}\, \mathrm{d}x \qquad (a > 0)$$

で定義する.
(1) $\displaystyle\int_{-\infty}^{\infty} \mathsf{p}^T(t)\, \mathrm{d}t$ を計算せよ.
(2) 連続型確率変数 U と V は独立で, U は自由度 p の χ^2 分布 χ^2_p に従い, V は $\mathsf{N}(0, 1)$ に従うとする.

$$T = \frac{V}{\sqrt{\dfrac{U}{p}}}$$

とおいたとき, T の p.d.f. p_T を求めよ.

$\boxed{\text{ヒント}}$ 確率ベクトル (U, V) の同時 p.d.f. $\mathsf{p}^{(U,V)}$ は

$$\mathsf{p}^{(U,V)}(u, v) = \begin{cases} \dfrac{1}{\sqrt{2\pi}} e^{-v^2/2} \dfrac{1}{\Gamma(p/2) 2^{p/2}} u^{(p/2)-1} e^{-u/2} & \left(\begin{array}{l} -\infty < v < \infty; \\ 0 < u < \infty \end{array} \right) \\ 0 & (\text{その他の場合}) \end{cases}$$

となる. さらに

$$T = \frac{V}{\sqrt{\dfrac{U}{p}}}; \quad W = U$$

として, (1.14) を利用して, 同時 p.d.f. $\mathsf{p}^{(T,W)}$ を求め

$$\mathsf{p}^{T}(t) = \int_0^\infty \mathsf{p}^{(T,W)}(t,\, w)\,\mathrm{d}w$$

を計算すればよい.

第2章　期待値の基礎事項

この章では, 確率変数の期待値と関連するものを導入し, その基本的な性質を述べる. 節 2.1 では, 確率変数に対する期待値を定義する. 節 2.2 では, 期待値を確率ベクトルに対して定義する. 節 2.3 では, 分布のばらつきを測る量である分散と共分散を導入し, 基本的な性質を説明する. 節 2.4 では, 条件付き分布に対する期待値を導入する. 節 2.5 では, 分布を特定する量である積率母関数を導入し, その基本的な性質を説明する.

2.1　期待値

X を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率変数とする. X が離散型のとき, その p.m.f. を p とし, X の取り得る値を $x_1, x_2, \ldots$ とする. 連続型のとき, その p.d.f. も p と書くことにする.

定義 2.1. $g : \mathbb{R} \to \mathbb{R}$ を可測[1]関数とする. 確率変数 $g(X)$ の期待値 $\mathsf{E}[g(X)]$ を次のように定義する.

(1) $g(x) \geq 0\,(x \in \mathbb{R})$ のとき

$$\mathsf{E}[g(X)] = \begin{cases} \displaystyle\sum_{n=1}^{\infty} g(x_n)\mathsf{p}(x_n) & (\text{離散型の場合}) \\ \displaystyle\int_{-\infty}^{\infty} g(x)\mathsf{p}(x)\,\mathrm{d}x & (\text{連続型の場合}) \end{cases}$$

と定義する. 右辺は ∞ を許せば, 必ず存在する.

(2) 一般の可測関数 g に対して

$$g^+(x) = \max\{g(x), 0\}, \quad g^-(x) = \max\{-g(x), 0\}$$

と定義すれば, $g^+(x) \geq 0$, $g^-(x) \geq 0\,(x \in \mathbb{R})$ となる. $\mathsf{E}[g^+(X)]$ または $\mathsf{E}[g^-(X)]$ のいずれかが有限ならば

$$\mathsf{E}[g(X)] := \mathsf{E}[g^+(X)] - \mathsf{E}[g^-(X)]$$

[1]可測関数の定義は定義 C.39 を参照.

と定義する．$\mathsf{E}[g^+(X)] = \mathsf{E}[g^-(X)] = \infty$ のときは, $g(X)$ の期待値は定義されない．$\mathsf{E}[g^+(X)] < \infty$ かつ $\mathsf{E}[g^-(X)] < \infty$ のとき, $\mathsf{E}[g(X)]$ は有限となる.

補題 2.2. X を確率変数とする．可測関数 $g, h : \mathbb{R} \to \mathbb{R}$ に対して, $\mathsf{E}[|g(X)|] < \infty$, $\mathsf{E}[|h(X)|] < \infty$ を仮定する.

(1) $a, b \in \mathbb{R}$ に対して

$$\mathsf{E}[ag(X) + bh(X)] = a\,\mathsf{E}[g(X)] + b\,\mathsf{E}[h(X)]$$

となる.

(2) $g(x) \le h(x)\,(\forall x \in \mathbb{R})$ ならば

$$\mathsf{E}[g(X)] \le \mathsf{E}[h(X)]$$

となる.

Proof. 積分の性質に帰着される. □

補題 2.3. $0 < q < r$ に対して

$$\mathsf{E}[|X|^r] < \infty \Rightarrow \mathsf{E}[|X|^q] < \infty$$

となる.

Proof. 積分の性質に帰着される. □

問 2.1. 連続確率変数 X が p.d.f. p をもつとき, 補題 2.3 の証明を具体的に書け．$q < r$ なので, $|x|^{q-r} \le 1\,(|x| > 1)$ となることを使えばよい.

定義 2.4. (1) $k = 1, 2, \ldots$ に対して, $\mathsf{E}[|X|^k] < \infty$ のとき, $\mathsf{E}[X^k]$ を X の k **次モーメント** (または**積率**) という.

(2) $\mathsf{E}[|X|] < \infty$ のとき $\mathsf{E}[X]$ を X **の平均値**[2]という.

(3) $\mathsf{E}[X^2] < \infty$ のとき X **の分散**を

$$\mathsf{Var}[X] := \mathsf{E}[\{X - \mathsf{E}[X]\}^2]$$

で定義する.

[2]簡単に「平均」ともいう.

(4) 部分集合 $A \subset \mathbb{R}$ に対して

$$\mathbb{1}_A(x) = \begin{cases} 1 & (x \in A) \\ 0 & (x \notin A) \end{cases}$$

を A の**指示関数**[3]という．$A \in \mathcal{B}(\mathbb{R})$ のとき

$$\mathsf{E}[\mathbb{1}_A(X)] = \mathsf{Pr}(X \in A)$$

となる．

注意 2.5. 分散 $\mathsf{Var}[X]$ は X の分布の平均 μ まわりの散らばりを測る量である．分散がおおきいほど分布は広がっていることになる． □

注意 2.6. (1) X は確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の離散型確率変数とする．p^X を X の p.m.f. とし

$$S_X := \{x \in \mathbb{R};\ \mathsf{p}^X(x) > 0\}$$

とする．関数 $g : \mathbb{R} \to \mathbb{R}$ に対して

$$Y = g(X)$$

とおく．$y \in \mathbb{R}$ に対して

$$A_y := \{x \in S_X;\ g(x) = y\}$$

とおき

$$S_Y := \{y \in \mathbb{R};\ A_y \neq \varnothing\}$$

とする．$y \in S_Y$ に対して

$$\mathsf{p}^Y(y) := \mathsf{Pr}(Y = y) = \mathsf{Pr}(X \in A_y) = \sum_{x \in A_y} \mathsf{Pr}(X = x) = \sum_{x \in A_y} \mathsf{p}^X(x)$$

と書ける．正項級数は項の順番を並び替えてもその値は変わらないので

$$\begin{aligned} \sum_{y \in S_Y} |y| \mathsf{p}^Y(y) &= \sum_{y \in S_Y} |y| \sum_{x \in A_y} \mathsf{p}^X(x) = \sum_{y \in S_Y} \sum_{x \in A_y} |g(x)| \mathsf{p}^X(x) \\ &= \sum_{x \in S_X} |g(x)| \mathsf{p}^X(x) \end{aligned}$$

となる．さらにいずれかの和が有限ならば

$$\sum_{y \in S_Y} y \mathsf{p}^Y(y) = \sum_{y \in S_Y} y \sum_{x \in A_y} \mathsf{p}^X(x) = \sum_{y \in S_Y} \sum_{x \in A_y} g(x) \mathsf{p}^X(x) = \sum_{x \in S_X} g(x) \mathsf{p}^X(x)$$

[3]指示関数は $\mathbb{R}$ の任意の部分集合に定義ができることに注意をせよ．

となる．よって

$$\mathsf{E}[Y] = \mathsf{E}[g(X)]$$

となる．

(2) 確率変数 X は連続型とする．p^X を X の p.d.f. とし

$$S_X := \{x \in \mathbb{R};\ \mathsf{p}^X(x) > 0\}$$

とする．関数 $g\colon S_X \to \mathbb{R}$ は狭義単調増加かつ C^1 級とする．このとき

$$Y = g(X), \quad S_Y := \{y \in \mathbb{R};\ y = g(x)\,(\exists x \in S_X)\}$$

とする．ここで $g\colon S_X \to S_Y$ と制限[4]すれば g の逆関数 $g^{-1}\colon S_Y \to S_X$ が存在し

$$\begin{aligned}\mathsf{F}^Y(y) &:= \mathsf{Pr}(Y \le y) = \mathsf{Pr}\bigl(g^{-1}(Y) \le g^{-1}(y)\bigr) = \mathsf{Pr}(X \le g^{-1}(y)) \\ &= \mathsf{F}^X(g^{-1}(y))\end{aligned}$$

となる．さらに $g(g^{-1}(y)) = y$ より

$$\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y) = \frac{1}{\dot{g}(g^{-1}(y))}$$

となる．ただし $\dot{g}(y) = \dfrac{\mathrm{d}g}{\mathrm{d}y}(y)$ である．これらより, $y \in S_Y$ に対して

$$\begin{aligned}\mathsf{p}^Y(y) &= \frac{\mathrm{d}}{\mathrm{d}y}\mathsf{F}^Y(y) = \frac{\mathrm{d}}{\mathrm{d}y}\mathsf{F}^X(g^{-1}(y)) = \dot{\mathsf{F}}^X(g^{-1}(y))\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y) \\ &= \mathsf{p}^X(g^{-1}(y))\frac{1}{\dot{g}\bigl(g^{-1}(y)\bigr)}\end{aligned}$$

となる．よって

$$\mathsf{p}^Y(y) = \mathsf{p}^X\bigl(g^{-1}(y)\bigr)\frac{1}{\dot{g}(g^{-1}(y))}$$

となる．g が狭義単調減少の場合もふくめると

$$\mathsf{p}^Y(y) = \mathsf{p}^X\bigl(g^{-1}(y)\bigr)\frac{1}{\bigl|\dot{g}\bigl(g^{-1}(y)\bigr)\bigr|}$$

となる．　□

[4] 制限したものを同じ g を用いて表現するのは, 記号の乱用である．記号が煩雑になるので, 記号を乱用した．

2.2　確率ベクトルの期待値

確率変数 X, Y を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上で定義された確率変数とする．(X, Y) を確率ベクトルという．

確率ベクトル (X, Y) が離散型のときその同時 p.m.f. を $\mathsf{p}(x, y)$ とし，連続型のときその同時 p.d.f. も $\mathsf{p}(x, y)$ と書くことにする．

定義 2.7. 可測関数 $g : \mathbb{R}^2 \to \mathbb{R}$ に対して $g(X, Y)$ の期待値を次のように定義する．
(1) $g \geq 0$ のとき

$$\mathsf{E}[g(X, Y)] = \begin{cases} \displaystyle\sum_{x, y} g(x, y)\mathsf{p}(x, y) & (\text{離散型の場合}) \\ \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y)\mathsf{p}(x, y)\,\mathrm{d}x\mathrm{d}y & (\text{連続型の場合}) \end{cases}$$

で定義する．
(2) 一般の g に対して

$$g^+(x, y) := \max\{g(x, y), 0\}, \quad g^-(x, y) := \max\{-g(x, y), 0\}$$

と定義すれば $g^+, g^- \geq 0$ となる．$\mathsf{E}[g^+(X, Y)]$ または $\mathsf{E}[g^-(X, Y)]$ のいずれかが有限ならば

$$\mathsf{E}[g(X, Y)] := \mathsf{E}[g^+(X, Y)] - \mathsf{E}[g^-(X, Y)]$$

で定義する．$\mathsf{E}[g^+(X, Y)] = \mathsf{E}[g^-(X, Y)] = \infty$ のときは，$g(X, Y)$ の期待値は定義されない．$\mathsf{E}[g^+(X, Y)] < \infty$ かつ $\mathsf{E}[g^-(X, Y)] < \infty$ のとき，$\mathsf{E}[g(X, Y)]$ は有限の値を取る．

注意 2.8. 3 つ以上の確率変数 $X_1, X_2, \ldots, X_n$ に対しても期待値を定義 2.7 と同様に定義する． □

定理 2.9. $X_1, X_2, \ldots, X_n$ を確率変数とし，各 $X_j\,(j = 1, 2, \ldots, n)$ の期待値は有限とする．$a_1, a_2, \ldots, a_n$ を定数としたとき

$$\mathsf{E}\left[\sum_{j=1}^{n} a_j X_j\right] = \sum_{j=1}^{n} a_j \mathsf{E}[X_j]$$

となる．

Proof. 期待値の線型性 (補題 2.2(1)) よりわかる． □

例 2.10. $0 < p < 1$ と $j = 1, 2, \ldots, n$ に対して, $X_j \sim \mathsf{Bino}(p)$ とする. このとき

$$\mathsf{E}[X_j] = \sum_{x=0}^{1} x\mathsf{Pr}(X = x) = 0 \times (1-p) + 1 \times p = p$$

である. $S = \sum_{j=1}^{n} X_j$ としたとき

$$\mathsf{E}[S] = \sum_{j=1}^{n} \mathsf{E}[X_j] = np$$

がわかる. □

2.3　分散と共分散

X を確率変数とし, $\mathsf{E}[X^2] < \infty$ とする. X **の分散**を

$$\mathsf{Var}[X] := \mathsf{E}[(X-\mu)^2]$$

で定義した. ただし $\mu = \mathsf{E}[X]$ と書いた. さらに $\sqrt{\mathsf{Var}[X]}$ を X **の標準偏差**という.

定理 2.11. 以下の確率変数は有限の 2 次の積率を持つとする. このとき, 次が成立する.

(1) $\mathsf{Var}[X] = \mathsf{E}[X^2] - \{\mathsf{E}[X]\}^2$ となる.

(2) 定数 $a, b \in \mathbb{R}\,(a \neq 0)$ に対して

$$\mathsf{Var}[aX + b] = a^2\mathsf{Var}[X]$$

となる.

(3) X と Y は独立で $\mathsf{E}[|XY|] < \infty$ とする. このとき

$$\mathsf{E}[XY] = \mathsf{E}[X]\mathsf{E}[Y]$$

となる.

(4) $X_1, X_2, \ldots, X_n$ は独立とし, $\mathsf{E}[X_j^2] < \infty\,(j = 1, 2, \ldots, n)$ とする. $a_1, a_2, \ldots, a_n$ は定数としたとき

$$\mathsf{Var}[a_1X_1 + a_2X_2 + \cdots + a_nX_n] = a_1^2\mathsf{Var}[X_1] + a_2^2\mathsf{Var}[X_2] + \cdots + a_n^2\mathsf{Var}[X_n]$$

となる.

Proof. (1), (2) は分散を期待値の記号を用いて表現し, 期待値の線型性を用いて計算すればよい. (3) については, 連続型の場合を示す. (X, Y) の同時 p.d.f. は, X と Y の周辺 p.d.f. p^X と p^Y を用いて $\mathsf{p}^X(x)\mathsf{p}^Y(y)$ という形でかけるので

$$\begin{aligned}\mathsf{E}[XY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\mathsf{p}^X(x)\mathsf{p}^Y(y)\,\mathrm{d}x\mathrm{d}y = \int_{-\infty}^{\infty} x\mathsf{p}^X(x)\,\mathrm{d}y \int_{-\infty}^{\infty} y\mathsf{p}^Y(y)\,\mathrm{d}y \\ &= \mathsf{E}[X]\mathsf{E}[Y]\end{aligned}$$

となること[5]がわかる. (4) は, 分散を期待値の記号を用いて表現し, (3) に注意する. 期待値の線型性を用いて計算すればよい. □

例 2.12. (例 2.10 の続き) 例 2.10 の設定に加えて $X_1, X_2, \ldots, X_n$ は独立とする. 定理 2.11(1) に注意すれば

$$\begin{aligned}\mathsf{Var}[X_j] &= \mathsf{E}[X_j^2] - \{\mathsf{E}[X_j]\}^2 = \sum_{x=0}^{1} x^2\mathsf{Pr}(X = x) - p^2 \\ &= 0 \times (1-p) + 1 \times p - p^2 = p(1-p)\end{aligned}$$

となる. これと定理 2.11(4) から

$$\mathsf{Var}[S] = \sum_{j=1}^{n} \mathsf{Var}[X_j] = np(1-p)$$

がわかる. □

定理 2.13. $n \geq 2$ とし, $X_1, X_2, \ldots, X_n$ は i.i.d. 確率変数列とし

$$\mathsf{E}[X_1] = \mu, \quad \mathsf{Var}[X_1] = \sigma^2$$

とする. ただし, $\mu \in \mathbb{R}$, $0 < \sigma < \infty$ である. $X_1, X_2, \ldots, X_n$ に基づく**標本平均**を

$$\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$$

で定義し, **標本** (不偏) **分散**を

$$S_n^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(X_j - \overline{X}_n\right)^2$$

で定義する. このとき

$$(1)\ \mathsf{E}\left[\overline{X}_n\right] = \mu; \qquad (2)\ \mathsf{Var}\left[\overline{X}_n\right] = \frac{\sigma^2}{n}; \qquad (3)\ \mathsf{E}\left[S_n^2\right] = \sigma^2$$

となる.

[5] $\mathsf{Var}[X] < \infty$, $\mathsf{Var}[Y] < \infty$ なので, Cauchy-Schwarz の不等式から $\mathsf{E}[|XY|] < \infty$ が確認できるので, Fubini 定理から積分順序の入れ替えが保証されることがわかる.

Proof. (1), (2) は定理 2.9, 2.11(4) よりわかる. (3) を証明するために

$$\sum_{j=1}^{n}\bigl(X_j-\overline{X}_n\bigr)^2=\sum_{j=1}^{n}(X_j-\mu)^2-n\bigl(\overline{X}_n-\mu\bigr)^2$$

に注意する. 定理 2.11(1) より

$$\mathsf{E}\Bigl[\sum_{j=1}^{n}(X_j-\mu)^2\Bigr]=n\sigma^2$$

がわかる. さらに, (1) と (2) より

$$\mathsf{E}\bigl[\bigl(\overline{X}_n-\mu\bigr)^2\bigr]=\mathsf{Var}\bigl[\overline{X}_n\bigr]=\frac{\sigma^2}{n}$$

が示される. したがって

$$\begin{aligned}\mathsf{E}[S_n^2]&=\frac{1}{n-1}\Bigl\{\sum_{j=1}^{n}\mathsf{E}\bigl[(X_j-\mu)^2\bigr]-n\mathsf{E}\bigl[\bigl(\overline{X}_n-\mu\bigr)^2\bigr]\Bigr\}\\&=\frac{1}{n-1}\bigl\{n\sigma^2-\sigma^2\bigr\}=\sigma^2\end{aligned}$$

がわかる. □

定義 2.14. X と Y は確率変数とし

$$\mathsf{E}[X]=\mu_X,\quad \mathsf{Var}[X]=\sigma_X^2,\quad \mathsf{E}[Y]=\mu_Y,\quad \mathsf{Var}[Y]=\sigma_Y^2$$

とする. ただし $\mu_X,\,\mu_Y\in\mathbb{R}$, $0<\sigma_X,\,\sigma_Y<\infty$ とする. このとき **X と Y の共分散**を

$$\mathsf{Cov}[X,\,Y]=\mathsf{E}\bigl[(X-\mu_X)(Y-\mu_Y)\bigr]$$

で定義し, **X と Y の (Pearson) の相関係数**を

$$\rho:=\rho[X,\,Y]:=\frac{\mathsf{Cov}[X,\,Y]}{\sigma_X\,\sigma_Y}$$

で定義する.

定理 2.15. $X,\,Y,\,Z$ は 2 次の積率が有限な確率変数とする.
(1) 共分散は

$$\mathsf{Cov}[X,\,Y]=\mathsf{E}[XY]-\mathsf{E}[X]\mathsf{E}[Y]$$

と書き直せる.
(2) $\mathsf{Cov}[X,\,Y]=\mathsf{Cov}[Y,\,X]$ である.
(3) 定数 $a,\,b\neq 0$ に対して

$$\mathsf{Cov}[aX+bY,\,Z]=a\mathsf{Cov}[X,\,Z]+b\mathsf{Cov}[Y,\,Z]$$

となる.
(4) 相関係数は

$$-1 \le \rho[X, Y] \le 1$$

をみたす.
(5) ある定数 a, b が存在して $Y = aX + b$ となったとき

$$a > 0 \Rightarrow \rho[X, Y] = 1,$$
$$a < 0 \Rightarrow \rho[X, Y] = -1$$

である.
(6) X と Y が独立のとき

$$\mathsf{Cov}[X, Y] = 0$$

となる.

Proof. (1) は期待値の線型性からわかる．(2)(3) は共分散の定義と期待値の線型性よりわかる．(4) は $\mathsf{Var}[X] = \mathsf{Var}[Y] = 0$ のときは明らかである．よって $\mathsf{Var}[X] \neq 0$ として証明する．補題 2.2(2) より

$$\mathsf{Var}[X]t^2 - 2\mathsf{Cov}[X, Y]t + \mathsf{Var}[Y] = \mathsf{E}[\{t(X - \mathsf{E}[X]) + Y - \mathsf{E}[Y]\}^2] \ge 0$$

となる．ここで, 判別式をとれば

$$\{\mathsf{Cov}[X, Y]\}^2 - \mathsf{Var}[X]\mathsf{Var}[Y] \le 0 \tag{2.1}$$

がわかる．(5) は (2) からわかる．(6) は定理 2.11(2) からわかる． □

注意 2.16. (2.1) において, $\mathsf{E}[X] = \mathsf{E}[Y] = 0$ の場合を考えると Cauchy-Schwarz の不等式

$$\left|\mathsf{E}[XY]\right| \le \sqrt{\mathsf{E}[X^2]}\sqrt{\mathsf{E}[Y^2]}$$

と呼ばれる重要なものを得る.

注意 2.17. (6) の逆は一般に正しくない．反例は下の問いをみよ. □

問 2.2. Z を $[0, 2\pi]$ 上の一様分布とし

$$X = \cos Z, \quad Y = \sin Z$$

とおく.

(1) $\mathsf{E}[X] = 0$, $\mathsf{E}[Y] = 0$, $\mathsf{E}[XY] = 0$ を確かめよ.

(2) $\mathsf{Pr}(X \le 1/2,\, Y \le 1/2) = \mathsf{Pr}((5/4)\pi \le Z \le (6/4)\pi)$ を確かめよ.

(3) $\mathsf{Pr}(X \le 1/2) = \mathsf{Pr}((1/4)\pi \le Z \le (7/4)\pi)$ を確かめよ[6].

定理 2.18. (1) X, Y は有限な 2 次の積率を持つ確率変数とする. このとき $\mathsf{Var}[X+Y] = \mathsf{Var}[X] + \mathsf{Var}[Y] + 2\mathsf{Cov}[X, Y]$ となる.
(2) $d \ge 2$ とする. $X_1, X_2, , \ldots, X_d$ は有限な 2 次の積率を持つ確率変数とする. このとき

$$\mathsf{Var}\left[\sum_{j=1}^{d} a_j X_j\right] = \sum_{j=1}^{d} a_j \mathsf{Var}[X_j] + 2\sum_{j=1}^{d}\sum_{\ell=j+1}^{d} a_j a_\ell \mathsf{Cov}[X_j, X_\ell]$$

となる.

Proof. 共分散を期待値で表現し, 展開して期待値の線型性を用いればよい. □

問 2.3. 定理 2.18(1) の証明を具体的に書け.

定義 2.19. $X_1, X_2, \ldots, X_d$ を有限な 2 次の積率を持つ確率変数とし

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_d \end{pmatrix}$$

と書く. このとき確率ベクトル $\boldsymbol{X}$ の期待値を

$$\mathsf{E}[\boldsymbol{X}] := \begin{pmatrix} \mathsf{E}[X_1] \\ \mathsf{E}[X_2] \\ \vdots \\ \mathsf{E}[X_d] \end{pmatrix}$$

で定義する. $\boldsymbol{X}$ の共分散を

$$\mathsf{Var}[\boldsymbol{X}] := \mathsf{E}\big[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^\top\big]$$

で定義する[7]. ただし $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_d)^\top := \mathsf{E}[\boldsymbol{X}]$ である. これは

$$\mathsf{Var}[\boldsymbol{X}] = \begin{pmatrix} \mathsf{Var}[X_1] & \mathsf{Cov}[X_1, X_2] & \cdots & \mathsf{Cov}[X_1, X_d] \\ \mathsf{Cov}[X_2, X_1] & \mathsf{Var}[X_2] & \cdots & \mathsf{Cov}[X_2, X_d] \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{Cov}[X_d, X_1] & \mathsf{Cov}[X_d, X_2] & \cdots & \mathsf{Var}[X_d] \end{pmatrix}$$

である.

[6] これらより, $\mathsf{Pr}(X \le 1/2, Y \le 1/2) \ne \mathsf{Pr}(X \le 1/2)\mathsf{Pr}(Y \le 1/2)$ となり, $\mathsf{Cov}[X, Y] = 0$ だが, X と Y は従属であることがわかる.

[7] 確率ベクトルと同様に確率変数を成分とする行列を確率行列という. 確率行列の期待値はそれぞれの成分の期待値を取ったものを配置した行列と定義している.

注意 2.20. 定義から $\mathsf{Var}[\boldsymbol{X}]$ は半正定値対称行列となる．なぜならば，期待値の線型性と補題 2.2(2) から，任意の $\boldsymbol{a} \in \mathbb{R}^d$ に対して

$$\begin{aligned}\boldsymbol{a}^\top \mathsf{Var}[\boldsymbol{X}]\boldsymbol{a} &= \mathsf{E}\big[\boldsymbol{a}^\top(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^\top\boldsymbol{a}\big] \\ &= \mathsf{E}\big[\big\{\boldsymbol{a}^\top(\boldsymbol{X}-\boldsymbol{\mu})\big\}^2\big] \geq 0\end{aligned}$$

となる．最後の不等号は補題 2.2(2) からわかる．よって $\mathsf{Var}[\boldsymbol{X}]$ は半正定値であることが示せた． □

補題 2.21. $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)^\top$ は確率ベクトルで各成分は有限な 2 次の積率を持つとし

$$\mathsf{E}[\boldsymbol{X}] = \boldsymbol{\mu}, \quad \mathsf{Var}[\boldsymbol{X}] = \boldsymbol{\Sigma}$$

とする．ただし $\boldsymbol{\mu} \in \mathbb{R}^d$, $\boldsymbol{\Sigma}$ は $d \times d$ の半正定値行列[8]である．
(1) 任意の定数ベクトル $\boldsymbol{a} \in \mathbb{R}^d$ に対して

$$\mathsf{E}[\boldsymbol{a}^\top\boldsymbol{X}] = \boldsymbol{a}^\top\boldsymbol{\mu}, \quad \mathsf{Var}[\boldsymbol{a}^\top\boldsymbol{X}] = \boldsymbol{a}^\top\boldsymbol{\Sigma}\boldsymbol{a}$$

となる．
(2) $k \in \mathbb{N}$ とする．任意の定数の $k \times d$ 行列 $\mathbf{A}$ に対して

$$\mathsf{E}[\mathbf{A}\boldsymbol{X}] = \mathbf{A}\mathsf{E}[\boldsymbol{X}], \quad \mathsf{Var}[\mathbf{A}\boldsymbol{X}] = \mathbf{A}\mathsf{Var}[\boldsymbol{X}]\mathbf{A}^\top$$

となる．

Proof. 期待値の線型性を用いて計算すればよい． □

問 2.4. $d = 2$ として，補題 2.21 を確認せよ．

系 2.22. 確率ベクトル $\boldsymbol{X}$ は，任意の定数ベクトル $\boldsymbol{a}(\neq \boldsymbol{0})$ に対して

$$\mathsf{Pr}(\boldsymbol{a}^\top\boldsymbol{X} = 0) = 0$$

をみたす．このとき，$\boldsymbol{\Sigma}$ は正定値である．

注意 2.23. $\boldsymbol{X}$ の共分散行列 $\boldsymbol{X}$ は，補題 2.21(2) から半正定値であることがわかる．系 2.22 の仮定は，確率ベクトル $\boldsymbol{X}$ が d 次元空間より次元の低い空間に集中しないことを意味している．したがって，確率ベクトル $\boldsymbol{X}$ の値を取る空間が退化していなければ，その共分散行列は正定値となることがわかる． □

[8] $d \times d$ の対称行列 $\mathbf{A}$ が半正定値であるとは，$\forall \boldsymbol{a} \in \mathbb{R}^d$ に対して $\boldsymbol{a}^\top\mathbf{A}\boldsymbol{a} \geq 0$ が成立するときをいう．また $d \times d$ の対称行列 $\mathbf{A}$ が正定値であるとは，$\forall \boldsymbol{a} \in \mathbb{R}^d$ $(\boldsymbol{a} \neq \boldsymbol{0})$ に対して $\boldsymbol{a}^\top\mathbf{A}\boldsymbol{a} > 0$ が成立するときをいう．

2.4　条件付き期待値

定義 2.24. (1) X と Y を確率変数とし, 条件付き p.d.f.(または条件付き p.m.f.) を $\mathsf{p}^{X|Y}$(または $\mathsf{p}^{X|Y}$) とする．$Y = y$ を与えたときの X の条件付き期待値を

$$\mathsf{E}[X|\,Y = y] = \begin{cases} \displaystyle\sum_x x\mathsf{p}^{X|Y}(x|\,y) & (\text{離散型の場合}) \\ \displaystyle\int_{-\infty}^{\infty} x\mathsf{p}^{X|Y}(x|\,y)\,\mathrm{d}x & (\text{連続型の場合}) \end{cases}$$

で定義する．ただし考えている $Y = y$ で条件付き p.d.f. または p.m.f. は定義され, さらに $\mathsf{E}[|X|] < \infty$ とする．
(2) (Borel 可測) 関数 $g : \mathbb{R}^2 \to \mathbb{R}$ に対して, $g(X, Y)$ の条件付き期待値を

$$\mathsf{E}[g(X, Y)|\,Y = y] = \begin{cases} \displaystyle\sum_x g(x, y)\mathsf{p}^{X|Y}(x|\,y) & (\text{離散型の場合}) \\ \displaystyle\int_{-\infty}^{\infty} g(x, y)\mathsf{p}^{X|Y}(x|\,y)\,\mathrm{d}x & (\text{連続型の場合}) \end{cases}$$

で定義する．ただし, 考えている $Y = y$ での条件付き p.d.f. または p.m.f. は定義され, $\mathsf{E}[|g(X, Y)|] < \infty$ とする．

注意 2.25. $\mathsf{E}[X]$ は定数であるが, $\mathsf{E}[X|\,Y = y]$ は一般に y の関数である．このことから

$$h(y) := \mathsf{E}[X|\,Y = y]$$

とおいたときに $h(y)$ に Y を代入したものは確率変数になる．これを

$$\mathsf{E}[X|\,Y] := h(Y)$$

と記すことにする．したがって $\omega \in \Omega$ に対して, $y = Y(\omega)$ と書けば

$$\mathsf{E}[X|\,Y] : \Omega \ni \omega \mapsto \mathsf{E}[X|\,Y(\omega)] = \mathsf{E}[X|\,Y = y] \in \mathbb{R}$$

は可測となる．
　測度論的確率論の教科書では, Radon-Nikodym の定理から条件付き期待値を定義する．これから条件付き p.d.f. を定義することになる．この点については [7, pp.181 − 182] を参照のこと．　□

例 2.26. 連続型確率変数 Y は p.d.f.

$$\mathsf{p}^Y(y) = \begin{cases} 1 & (0 < y < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

を持つとする．$Y = y\,(0 < y < 1)$ を観測したとき

$$X|\,Y = y \sim \mathsf{Unif}(y,\,1)$$

とする．すなわち $0 < y < 1$ のとき

$$\mathsf{p}^{X|\,Y}(x|\,y) = \begin{cases} \dfrac{1}{1-y} & (y < x < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

である．よって

$$\mathsf{E}[X|\,Y = y] = \int_y^1 y\mathsf{p}^{X|\,Y}(x|\,y)\,\mathrm{d}x = \frac{1}{1-y}\int_y^1 \mathrm{d}x = \frac{1+y}{2}$$

となる．これより

$$\mathsf{E}[X|\,Y] = \frac{1+Y}{2}$$

となる． □

定理 2.27. (1) 有限な期待値を持つ確率変数 $X,\,Y$ と確率変数 Z に対して

$$\mathsf{E}[X+Y|\,Z] = \mathsf{E}[X|\,Z] + \mathsf{E}[Y|\,Z]$$

となる．
(2) 有限な期待値を持つ確率変数 $X,\,Y$ に対して

$$\mathsf{E}[\mathsf{E}[Y|\,X]] = \mathsf{E}[Y], \quad \mathsf{E}[\mathsf{E}[X|\,Y]] = \mathsf{E}[X]$$

となる．
(3) 一般の (Borel 可測) 関数 $g : \mathbb{R}^2 \to \mathbb{R}$ を考える．$\mathsf{E}[|g(X,\,Y)|] < \infty$ のとき

$$\mathsf{E}[\mathsf{E}[g(X,\,Y)|\,Y]] = \mathsf{E}[g(X,\,Y)]$$

となる．
(4) $\mathsf{E}[XY|\,Y] = Y\mathsf{E}[X|\,Y]$ となる．

Proof. (1) は期待値の線型性よりわかる．連続型の場合について (2) の第 1 番目の等式を示す．他の場合もほとんど同じように証明できる．p を $(X,\,Y)$ の同時 p.d.f. とする．このとき

$$\mathsf{p}(x,\,y) = \mathsf{p}^Y(y)\mathsf{p}^{X|\,Y}(x|\,y)$$

となる．ただし p^X は X の周辺 p.d.f. で $\mathsf{p}^{Y|\,X}$ は $X = x$ を与えたときの Y の条件付き p.d.f. である．$\mathsf{E}[X|\,Y = y] =: h(y)$ とおいたとき,

$\mathsf{E}[X|\,Y] := h(Y)$ と書いたことを思い出す．すると

$$\begin{aligned}\mathsf{E}[\mathsf{E}[X|\,Y]] = \mathsf{E}[h(Y)] &= \int_{-\infty}^{\infty} h(y)\mathsf{p}^Y(y)\,\mathrm{d}y = \int_{-\infty}^{\infty} \mathsf{E}[X|\,Y=y]\mathsf{p}^Y(y)\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} x\mathsf{p}^{X|\,Y}(x|\,y)\,\mathrm{d}x\right\}\mathsf{p}^Y(y)\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\mathsf{p}^{X|\,Y}(x|\,y)\mathsf{p}^Y(y)\,\mathrm{d}x\mathrm{d}y \\ &= \int_{-\infty}^{\infty} x\left\{\int_{-\infty}^{\infty}\mathsf{p}(x,\,y)\,\mathrm{d}y\right\}\mathrm{d}x \\ &= \int_{-\infty}^{\infty} x\mathsf{p}^X(x)\,\mathrm{d}x \\ &= \mathsf{E}[X]\end{aligned}$$

となる．積分の順序交換は $g(X,\,Y)$ が有限の期待値を持つことから保証されること[9]が知られている． □

例 2.28. (例 2.26 の続き)

$$\mathsf{E}[X] = \mathsf{E}\left[\frac{1+Y}{2}\right] = \frac{3}{4}$$

となる．一方 $0 < x < y < 1$ に対して

$$\mathsf{p}(x,\,y) = \mathsf{p}^{X|\,Y}(x|\,y)\mathsf{p}^Y(y) = \frac{1}{1-y}$$

となる．よって

$$\begin{aligned}\mathsf{E}[X] &= \int_0^1\left\{\int_y^1 x\,\frac{1}{1-y}\,\mathrm{d}x\right\}\mathrm{d}y \\ &= \int \frac{1}{1-y}\left[\frac{x^2}{2}\right]_y^1\mathrm{d}y \\ &= \int_0^1 \frac{1}{1-y}\frac{1-y^2}{2}\,\mathrm{d}y = \left[\frac{(1+y)^2}{4}\right]_0^1 = \frac{3}{4}\end{aligned}$$

となる． □

定義 2.29. X は有限の 2 次の積率を持つとする．$Y=y$ を与えたときの X の条件付き p.d.f. $\mathsf{p}^{X|Y}$(p.m.f. $\mathsf{p}^{X|Y}$) が定義できる y を考える．このとき，$Y=y$ を与えたときの条件付き分散を

$$\mathsf{Var}[X|\,Y=y] = \begin{cases}\displaystyle\sum_x \{x-\mu(y)\}^2\mathsf{p}^{X|\,Y}(x|\,y) & \text{(離散型の場合)} \\ \displaystyle\int_{-\infty}^{\infty}\{x-\mu(y)\}^2\mathsf{p}^{X|\,Y}(x|\,y)\,\mathrm{d}x & \text{(連続型の場合)}\end{cases}$$

[9] Fubini の定理からわかる．

で定義する．ただし $\mu(y) = \mathsf{E}[X|\,Y = y]$ である．これは

$$\mathsf{Var}[X|\,Y] = \mathsf{E}[X^2|\,Y] - \{\mathsf{E}[X|\,Y]\}^2$$

とも書ける．

定理 2.30. X, Y を確率変数とし $\mathsf{E}[X^2] < \infty$ とする．このとき

$$\mathsf{Var}[X] = \mathsf{E}[\mathsf{Var}[X|\,Y]] + \mathsf{Var}[\mathsf{E}[X|\,Y]]$$

が成立する．ただし $h(y) := \mathsf{Var}[X|\,Y = y]$ としたとき $\mathsf{Var}[X|\,Y] := h(Y)$ と定義した．

Proof. まず

$$\begin{aligned}
\mathsf{Var}[X] &= \mathsf{E}\big[\{X - \mathsf{E}[X]\}^2\big] \\
&= \mathsf{E}\Big[\big\{X - \mathsf{E}[X|\,Y] + \mathsf{E}[X|\,Y] - \mathsf{E}[X]\big\}^2\Big] \\
&= \mathsf{E}\big[\{X - \mathsf{E}[X|\,Y]\}^2\big] + \mathsf{E}\big[\{\mathsf{E}[X|\,Y] - \mathsf{E}[X]\}^2\big] \\
&\quad + 2\mathsf{E}\big[\{X - \mathsf{E}[X|\,Y]\}\{\mathsf{E}[X|\,Y] - \mathsf{E}[X]\}\big] \qquad (2.2)
\end{aligned}$$

と書き直す．しかし, (2.2) の最右辺の各項は以下のように評価できる．

$$\begin{aligned}
\mathsf{E}\big[\{X - \mathsf{E}[X|\,Y]\}^2\big] &= \mathsf{E}\big[X^2 - 2X\mathsf{E}[X|\,Y] + \{\mathsf{E}[X|\,Y]\}^2\big] \\
&= \mathsf{E}\Big[\mathsf{E}\big[X^2 - 2X\mathsf{E}[X|\,Y] + \{\mathsf{E}[X|\,Y]\}^2\big|\,Y\big]\Big] \\
&\qquad (\because \text{ 定理 } 2.27(3)) \\
&= \mathsf{E}\Big[\mathsf{E}\big[X^2\big|\,Y\big] - 2\mathsf{E}[X|\,Y]\,\mathsf{E}[X|\,Y] + \{\mathsf{E}[X|\,Y]\}^2\Big] \\
&= \mathsf{E}\Big[\mathsf{E}[X^2|\,Y] - \{\mathsf{E}[X|\,Y]\}^2\Big] \\
&= \mathsf{E}\big[\mathsf{Var}[X|\,Y]\big], \\
\mathsf{E}\Big[\{\mathsf{E}[X|\,Y] - \mathsf{E}[X]\}^2\Big] &= \mathsf{E}\Big[\big\{\mathsf{E}[X|\,Y] - \underbrace{\mathsf{E}[\mathsf{E}[X|\,Y]]}_{=\mathsf{E}[X]}\big\}^2\Big] \\
&\qquad (\because \text{ 定理 } 2.27(3)) \\
&= \mathsf{Var}\Big[\mathsf{E}[X|\,Y]\Big], \\
\mathsf{E}\big[\{X - \mathsf{E}[X|\,Y]\}\{\mathsf{E}[X|\,Y] - \mathsf{E}[X]\}\big] &= \mathsf{E}\Big[\mathsf{E}\big[\{X - \mathsf{E}[X|\,Y]\}\{\mathsf{E}[X|\,Y] - \mathsf{E}[X]\}\big|\,Y\big]\Big] \\
&\qquad (\because \text{ 定理 } 2.27(3)) \\
&= \mathsf{E}\Big[\{\mathsf{E}[X|\,Y] - \mathsf{E}[X]\}\,\underbrace{\mathsf{E}\big[X - \mathsf{E}[X|\,Y]\big|Y\big]}_{=0}\Big] \\
&= 0
\end{aligned}$$

となる．最後から 2 番目の等号は定理 2.27(3) を用いた．これらの結果を (2.2) の最右辺の各項に代入すれば, 定理は証明される． □

2.5 積率母関数

定義 2.31. X を確率変数とし, ある $t_0 > 0$ が存在して, $\mathsf{E}[e^{tX}] < \infty$ ($\forall |t| < t_0$) とする．このとき, X **の積率母関数** (Moment Generating Function (MGF)) を

$$\mathsf{m}^X(t) := \mathsf{E}[e^{tX}] \quad (-t_0 < t < t_0)$$

と定義する．

注意 2.32. 確率変数 X の積率母関数 $\mathsf{m}^X(t)$ が存在するとき, 期待値と微分の記号の入れ替えが保証されること[10]が知られている．このことから

$$\begin{aligned}\dot{\mathsf{m}}^X(0) &= \dot{\mathsf{m}}^X(t)\Big|_{t=0} = \left[\frac{\mathrm{d}}{\mathrm{d}t} m^X(t)\right]\Bigg|_{t=0} = \left[\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{E}\big[e^{tX}\big]\right]\Bigg|_{t=0} = \mathsf{E}\left[\frac{\mathrm{d}}{\mathrm{d}t} e^{tX}\right]\Bigg|_{t=0} \\ &= \mathsf{E}\big[Xe^{tX}\big]\Big|_{t=0} = \mathsf{E}[X]\end{aligned}$$

となる．この議論を繰り返せば $k = 2, 3, \ldots$ に対して

$$\{\mathsf{m}^X\}^{(k)}(0) = \mathsf{E}\big[X^k\big]$$

がわかる． □

例 2.33. $X \sim \mathsf{Exp}(1)$ とする．$t < 1$ に対して

$$\mathsf{m}^X(t) = \mathsf{E}\big[e^{tX}\big] = \int_0^\infty e^{tx} e^{-x}\,\mathrm{d}x = \int_0^\infty e^{-(1-t)x}\,\mathrm{d}x = \frac{1}{1-t}$$

となる．$t \geq 1$ のときは, e^{tX} の期待値は発散する．したがって

$$\mathsf{m}^X(t) = \frac{1}{1-t} \qquad (t < 1)$$

となる．簡単な計算から

$$\dot{\mathsf{m}}^X(0) = 1, \quad \ddot{\mathsf{m}}^X(0) = 2$$

なので

$$\mathsf{E}[X] = 1, \quad \mathsf{E}[X^2] = 2, \quad \mathsf{Var}[X] = \mathsf{E}[X^2] - \{\mathsf{E}[X]\}^2 = 1$$

となる． □

[10] たとえば, [30, pp.75-76] を参照のこと．

補題 2.34. (1) $a,\,b \in \mathbb{R}\,(a \neq 0)$ とする．$Y = aX + b$ としたとき

$$\mathsf{m}^Y(t) = e^{tb}\mathsf{m}^X(at)$$

となる．
(2) $X_1,\,X_2,\,\ldots,\,X_d$ は独立とし $Y = \sum_{j=1}^d X_j$ とする．このとき

$$\mathsf{m}^Y(t) = \prod_{j=1}^d \mathsf{m}^{X_j}(t)$$

となる．

Proof. (1) は指数関数と期待値の性質よりわかる．(2) は指数関数の性質と定理 2.11(3) よりわかる． □

定理 2.35. X と Y を確率変数とする．ある数 $t_0 > 0$ が存在して

$$\mathsf{m}^X(t) = \mathsf{m}^Y(t) \quad (|t| < t_0)$$

ならば

$$X \overset{d}{=} Y$$

となる．ただし X と Y の c.d.f. を F^X と F^Y としたとき

$$X \overset{d}{=} Y \;\Leftrightarrow\; \mathsf{F}_X(x) = \mathsf{F}_Y(x) \qquad (\forall x \in \mathbb{R})$$

である．

Proof. これは信じることにする． □

注意 2.36. 定理 2.35 の証明は, 積率母関数が存在する範囲に対する $\mathbb{C}$ の帯領域に解析接続し, それに対して, Fourier 逆変換の公式を適応して証明するのが標準的であろう．

例 2.37. $n_1,\,n_2 \in \mathbb{N},\,0 < p < 1$ とし $X_1 \sim \mathsf{Bino}(n_1,\,p),\,X_2 \sim \mathsf{Bino}(n_2,\,p)$ は独立とする．$Y = X_1 + X_2$ としたとき

$$\mathsf{m}^Y(t) = \mathsf{m}^{X_1}(t)\mathsf{m}^{X_2}(t) = \left(pe^t + q\right)^{n_1}\left(pe^t + q\right)^{n_2} = (pe^t + q)^{n_1+n_2}$$

となる．ただし $q = 1 - p$ である．よって $Y \sim \mathsf{Bino}(n_1 + n_2,\,p)$ となる．
□

例 2.38. $\lambda_1,\,\lambda_2 > 0$ とし, $X_1 \sim \mathsf{Po}(\lambda_1),\,X_2 \sim \mathsf{Po}(\lambda_2)$ は独立とする．$Y = X_1 + X_2$ としたとき

$$\mathsf{m}^Y(t) = \mathsf{m}^{X_1}(t)\mathsf{m}^{X_2}(t) = e^{\lambda_1(e^t-1)}e^{\lambda_2(e^t-1)} = e^{(\lambda_1+\lambda_2)(e^t-1)}$$

となる．したがって, $Y \sim \mathsf{Po}(\lambda_1 + \lambda_2)$ がわかる． □

2.6　章末注釈と参考文献

この章は [27] を参考にした.

2.7　演習問題

演習問題 2.1. 確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率変数 X は

$$\mathsf{Pr}(X = c) = 1; \qquad c \text{ は定数}$$

をみたすとする. このとき, X の平均 $\mathsf{E}[X]$ と分散 $\mathsf{Var}[X]$ を求めよ.

演習問題 2.2. (1) $X \sim \mathsf{Ber}(\theta)\,(0 < \theta < 1)$ のとき

$$\mathsf{E}[X] = \theta, \quad \mathsf{Var}[X] = \theta(1-\theta)$$

を示せ.
(2) $X_1, X_2, \ldots, X_n \sim \text{i.i.d.}\mathsf{Ber}(\theta)\,(0 < \theta < 1)$ とする. $S = \sum_{j=1}^{n} X_j$ とおいたとき, $S \sim \mathsf{Bino}(n, \theta)$ となることを示せ.
(3) $S \sim \mathsf{Bino}(n, \theta)\,(n \in \mathbb{N}, 0 < \theta < 1)$ のとき

$$\mathsf{E}[S] = n\theta, \quad \mathsf{Var}[S] = n\theta(1-\theta)$$

を示せ.
(4) $X \sim \mathsf{Po}(\theta)\,(\theta > 0)$ とする. このとき, X の平均 $\mathsf{E}[X]$ と分散 $\mathsf{Var}[X]$ を求めよ.
(5) $U \sim \mathsf{U}(0, 1)$ のとき

$$\mathsf{E}[X] = \frac{1}{2}, \quad \mathsf{Var}[X] = \frac{1}{12}$$

を示せ.
(6) $X \sim \mathsf{Ex}(\theta)\,(\theta > 0)$ とする. このとき

$$\mathsf{E}[X] = \frac{1}{\theta}, \quad \mathsf{Var}[X] = \frac{1}{\theta^2}$$

を示せ.
(6) $X \sim \mathsf{N}(\mu, \sigma^2)\,(\mu \in \mathbb{R}, \sigma > 0)$ とする. このとき

$$\mathsf{E}[X] = \mu, \quad \mathsf{Var}[X] = \sigma^2$$

を示せ.
(7) $X \sim \chi^2_n$ のとき

$$\mathsf{E}[X] = n, \quad \mathsf{E}[X] = 2n$$

を示せ.

演習問題 2.3. 確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率変数 X は $\mathsf{Var}[X] = \sigma^2$ $(0 < \sigma < \infty)$ とする．c を定数としたとき, $\mathsf{Var}[X + c]$ と $\mathsf{Var}[cX]$ を σ と c を用いて表せ.

演習問題 2.4. (1) $X \sim \mathsf{Ga}(\alpha, \beta)$ とする．X の k 次の積率は以下で与えられることを示せ.

$$\mathsf{E}[X^k] = \frac{(a+k-1)(a+k-2)\cdots\alpha}{\beta^k}.$$

(2) 上の問いの結果から

$$\mathsf{E}[X] = \frac{\alpha}{\beta}, \quad \mathsf{E}[X^2] = \frac{(\alpha+1)\alpha}{\beta^2}, \quad \mathsf{Var}[X] = \frac{\alpha}{\beta^2}$$

を示せ.

(3) X の積率母関数は以下で与えられることを示せ.

$$\mathsf{m}^X(t) = \left(\frac{1}{1-\beta t}\right)^{\alpha} \quad (t < 1/\beta).$$

演習問題 2.5. X を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の連続型確率変数とし, X の p.d.f. を

$$\mathsf{p}(x) = \begin{cases} 2x & (0 < x < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

とする．さらに, 離散型確率変数 Y を

$$Y := \mathbb{1}_{(1/2,\,\infty)}(X) := \begin{cases} 1 & \left(X > \frac{1}{2}\right) \\ 0 & \left(X \le \frac{1}{2}\right) \end{cases}$$

で定める．このとき, 以下の問に答えよ.

(1) 確率変数 X の平均 $\mathsf{E}[X]$ と分散 $\mathsf{Var}[X]$ を求めよ.

(2) 確率変数 Y の平均 $\mathsf{E}[Y]$ と分散 $\mathsf{Var}[Y]$ を求めよ.

(3) 確率変数 X と Y の共分散 $\mathsf{Cov}[X, Y]$ を求めよ.

演習問題 2.6. X_1, X_2 を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の連続型確率変数とし,

$$X_1 \sim \mathsf{N}(x_0 + \mu, \sigma^2), \qquad X_2 |\, X_1 = x_1 \sim \mathsf{N}(x_1 + \mu, \sigma^2)$$

とする．ただし, $x_0, x_1, \mu \in \mathbb{R}$ である．このとき, 以下の問いに答えよ．ただし, 正規分布 $\mathsf{N}(\mu, \sigma^2)$ に従う確率変数 Z の期待値と分散が $\mathbb{E}[Z] = \mu$, $\mathsf{Var}[Z] = \sigma^2$ となることは証明なしで用いてよい．また, 期待

値, 分散, 共分散, 条件付き期待値に係る資料に書いてある性質も証明なしで用いてよい. なお, どの性質を用いたかは明示すること.
(1) 確率変数 $X_2 - X_1$ の期待値 $\mathbb{E}[X_2 - X_1]$ を求めよ.
(2) X_2 の分散 $\mathsf{Var}[X_2]$ を求めよ.
(3) X_1 と X_2 の共分散 $\mathsf{Cov}[X_1, X_2]$ を求めよ.
(4) $X_2 - X_1$ の分散 $\mathsf{Var}[X_2 - X_1]$ を $\mathsf{Var}[X_1]$, $\mathsf{Var}[X_2]$, $\mathsf{Cov}[X_1, X_2]$ で表現せよ.
(5) $X_2 - X_1$ の分散 $\mathsf{Var}[X_2 - X_1]$ を求めよ.

演習問題 2.7. $(\Omega, \mathcal{A}, \mathsf{Pr})$ を確率空間とし, X をこの空間上で定義された確率変数とする. X は開区間 $(0, 1)$ 上の一様分布に従うとし

$$\boldsymbol{X} = \begin{bmatrix} X \\ X^2 \end{bmatrix}$$

とおく. このとき, 以下の問いに答えよ.
(1) 確率ベクトル $\boldsymbol{X}$ の期待値 $\mathsf{E}[\boldsymbol{X}]$ を求めよ.
(2)

$$\mathsf{Var}[\boldsymbol{X}] = \mathsf{E}[\boldsymbol{X}\boldsymbol{X}^\top] - \mathsf{E}[\boldsymbol{X}]\big\{\mathsf{E}[\boldsymbol{X}]\big\}^\top$$

を示せ.
(3) 確率ベクトル $\boldsymbol{X}$ の分散共分散行列 $\mathsf{Var}[\boldsymbol{X}]$ を求めよ.
(4) $\det\big[\mathsf{Var}[\boldsymbol{X}]\big]$ を求めよ.

演習問題 2.8. 確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上で定義された連続型確率ベクトル (X, Y) は同時 p.d.f.

$$\mathsf{p}^{(X,Y)}(x, y) = \begin{cases} \dfrac{3}{8} \times y & (0 \le x \le y \le 2) \\ 0 & (\text{その他の場合}) \end{cases}$$

を持つとする.
(1) X の周辺 p.d.f. $\mathsf{p}^X(x)$ を求め,

$$\int_{-\infty}^{\infty} \mathsf{p}^X(x)\,\mathrm{d}x = 1$$

を確認せよ.
(2) Y の周辺 p.d.f. $\mathsf{p}^Y(y)$ を求め,

$$\int_{-\infty}^{\infty} \mathsf{p}^Y(y)\,\mathrm{d}y = 1$$

を確認せよ.

(3) X と Y の期待値 $\mathsf{E}[X]$ と $\mathsf{E}[Y]$ を求めよ．
(4) XY の期待値 $\mathsf{E}[XY]$ を求め, X と Y の共分散 $\mathsf{Cov}[X,\, Y]$ を求めよ．
(5) $\mathsf{p}^{Y}(y) > 0$ なる y に対して, $Y = y$ を与えたときの X の条件付き p.d.f. $\mathsf{p}^{X|Y}(x|\, y)$ を求めよ．$\mathsf{p}^{X|Y}(x|\, y) > 0$ となる x の範囲を書いたうえで,

$$\int_{-\infty}^{\infty} \mathsf{p}^{X|Y}(x|\, y)\,\mathrm{d}x = 1$$

を確認せよ．

演習問題 2.9. 確率空間 $(\Omega,\, \mathcal{A},\, \mathsf{Pr})$ 上で定義された連続型確率ベクトル $(X,\, Y)$ は同時 p.d.f.

$$\mathsf{p}^{(X,Y)}(x,\, y) = \begin{cases} e^{-(x+y)} & (0 < x < \infty,\, 0 < y < \infty) \\ 0 & (\text{その他の場合}) \end{cases}$$

を持つ．

$$U = 2X, \qquad V = X + Y$$

とおく．
(1) 確率ベクトル $(U,\, V)$ の同時 p.d.f. $\mathsf{p}^{(U,V)}(u,\, v)$ を求めよ．さらに

$$T = \{(u,\, v) \in \mathbb{R}^2 : \mathsf{p}^{(U,V)}(u,\, v) > 0\}$$

を図示せよ．
(2) U の周辺 p.d.f. $\mathsf{p}^{U}(u)$ を求めよ．

演習問題 2.10. $\lambda > 0$ とし, W は確率空間 $(\Omega,\, \mathcal{A},\, \mathsf{Pr})$ 上で定義された確率変数で, 母数 λ の指数分布 $\mathsf{Exp}(\lambda)$ に従うとする．すなわち, W は連続型確率変数で, p.d.f.

$$\mathsf{p}^{W}(w) = \begin{cases} \lambda e^{-\lambda w} & (w > 0) \\ 0 & (\text{その他の場合}) \end{cases}$$

をもつ．同じ確率空間 $(\Omega,\, \mathcal{A},\, \mathsf{Pr})$ 上で定義された確率変数 X と Y を

$$X = \lfloor W \rfloor, \quad Y = W - X$$

で定める．ただし, $\lfloor x \rfloor$ は実数 x 以下の最大の整数とする．たとえば, $\lfloor 1.5 \rfloor = 1$ となる．このとき, 以下の問いに答えよ．
(1) 非負の整数 x に対して

$$\mathsf{Pr}(X = x)$$

を求めることで X の p.m.f. p^{X} を求めよ．

(2) $\displaystyle\sum_{x=0}^{\infty}\mathsf{p}^X(x)$ を計算せよ.
(3) x を非負の整数とし, $0 < y < 1$ とする. $X = x$ を与えたときの $\{Y \leq y\}$ の条件付き確率

$$\mathsf{Pr}(Y \leq y|\, X = x)$$

を求めよ.
(4) x を非負の整数とする. $y \in \mathbb{R}$ に対して, $\mathsf{F}^{Y|X}(y|\, x) = \mathsf{Pr}(Y \leq y|\, X = x)$ と定める. $\mathsf{F}^{Y|X}(y|\, x)$ を実数上で定義された関数として書き表せ.
(5) x を非負の整数とする. $X = x$ を与えたときの Y の条件付き p.d.f. $\mathsf{p}^{Y|X}(y|\, x)$ を求めよ.
(6) 期待値 $\mathsf{E}[Y]$ を求めよ.

補足 $\mathsf{Pr}(0 < Y < 1) = 1$ から $0 < \mathsf{E}[Y] < 1$ となる. 得られた結果について, 不等式 $0 < \mathsf{E}[Y] < 1$ となっているかを確認するとよい.

演習問題 2.11. X を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の連続型確率変数とし, X の p.d.f. を

$$\mathsf{p}(x) = \begin{cases} 2x & (0 < x < 1) \\ 0 & \text{(その他の場合)} \end{cases}$$

とする. さらに, 離散型確率変数 Y を

$$Y := \mathbb{1}_{(1/2,\,\infty)}(X) := \begin{cases} 1 & \left(X > \frac{1}{2}\right) \\ 0 & \left(X \leq \frac{1}{2}\right) \end{cases}$$

で定める. このとき, 以下の問に答えよ.
(1) 確率変数 X の平均 $\mathsf{E}[X]$ と分散 $\mathsf{Var}[X]$ を求めよ.
(2) 確率変数 Y の平均 $\mathsf{E}[Y]$ と分散 $\mathsf{Var}[Y]$ を求めよ.
(3) 確率変数 X と Y の共分散 $\mathsf{Cov}[X, Y]$ を求めよ.

演習問題 2.12. (U, Y) の同時分布を以下のように定める. $U \sim \mathsf{Unif}(0, 1)$ とし, $U = u\,(0 < u < 1)$ が与えられたときの Y の条件付き分布は $\mathsf{N}(\theta, \sigma^2)$ とする. ただし,

$$\theta = 2 - 6u, \quad \sigma^2 = 25$$

とする.
(1) U の期待値 $\mathsf{E}[U]$ を計算せよ.
(2) U の分散 $\mathsf{Var}[U]$ を計算せよ.
(3) $U = u\,(0 < u < 1)$ を与えたときの Y の条件付き期待値 $\mathsf{E}[Y|\,U = u]$ を答えよ.

(4) Y の期待値 $\mathsf{E}[Y]$ を計算せよ.
(5) Y の分散 $\mathsf{Var}[Y]$ を求めよ.
(6) $U = u\,(0 < u < 1)$ を与えたときの条件付き分散 $\mathsf{Var}[Y\,|\,U = u]$ を答えよ.
(7) $\mathsf{E}\big[\mathsf{Var}[Y\,|\,U]\big] + \mathsf{Var}\big[\mathsf{E}[Y\,|\,U]\big]$ を計算せよ.
(8) 連続関数 $g\colon [0,\,1] \to \mathbb{R}$ に対して

$$\mathsf{E}\big[\big(Y - g(U)\big)^2\big] \geq 25$$

となること示せ.

第3章　確率と期待値の不等式

　直接計算するのが困難な確率や期待値に対して上限ないしは下限を与える不等式は有効である.

3.1　確率に対する不等式

定理 3.1. (Markov の不等式) 確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の非負値確率変数を X とし, $\mathsf{E}[X] < \infty$ とする. このとき, $\forall t > 0$ に対して

$$\mathsf{Pr}\big(X \geq t\big) \leq \frac{\mathsf{E}[X]}{t}$$

が成り立つ.

Proof. X は連続型で p.d.f. p を持つ場合を示す.

$$\begin{aligned}\mathsf{E}[X] &= \int_0^\infty x\mathsf{p}(x)\,\mathrm{d}x = \int_0^t x\mathsf{p}(x)\,\mathrm{d}x + \int_t^\infty x\mathsf{p}(x)\,\mathrm{d}x \geq \int_t^\infty x\mathsf{p}(x)\,\mathrm{d}x \\ &\geq t\int_t^\infty \mathsf{p}(x)\,\mathrm{d}x = t\mathsf{Pr}(X \geq t)\end{aligned}$$

からわかる. □

系 3.2. $\lambda > 0$ とする. X を確率変数とし $\mathsf{E}\big[e^{\lambda X}\big] < \infty$ とする. このとき $\forall t > 0$ に対して

$$\mathsf{Pr}\big(X \geq t\big) \leq e^{-\lambda t}\mathsf{E}\big[e^{\lambda X}\big]$$

が成り立つ.

Proof. $e^{\lambda X}$ は非負値確率変数なので定理 3.1 を適用すれば

$$\mathsf{Pr}\big(X \geq t\big) = \mathsf{Pr}\big(e^{\lambda X} \geq e^{\lambda t}\big) \leq e^{-\lambda t}\mathsf{E}\big[e^{\lambda X}\big]$$

となる. □

系 3.3. (Chebyshev の不等式) X を確率変数とし $\mu = \mathsf{E}[X]$, $\sigma^2 = \mathsf{Var}[X] < \infty\,(0 < \sigma < \infty)$ とする．このとき $\forall t > 0$ に対して

$$\mathsf{Pr}\bigl(|X - \mu| \geq t\bigr) \leq \frac{\sigma^2}{t^2}$$

が成り立つ．

Proof. $(X - \mu)^2$ に対して Markov の不等式 (定理 3.1) を適用する．すると

$$\mathsf{Pr}\bigl(|X - \mu| \geq t\bigr) = \mathsf{Pr}\bigl((X - \mu)^2 \geq t^2\bigr) \leq \frac{\mathsf{E}[(X - \mu)^2]}{t^2}$$

がわかる． □

次に Hoeffding の不等式を証明するための補題を与える．

補題 3.4. $a, b \in \mathbb{R}$ は $a < 0 < b$ なる定数とする．確率変数 X は

$$\mathsf{E}[X] = 0, \quad a < X < b$$

とみたすとき $\forall \lambda > 0$ に対して

$$\mathsf{E}\bigl[e^{\lambda X}\bigr] \leq e^{\lambda^2 (b-a)^2/8}$$

となる．

Proof. X を以下のように書き直す．

$$X = \gamma b + (1 - \gamma) a, \quad \gamma = \frac{X - a}{b - a}$$

となる．すると $e^{\lambda x}$ の凸性より

$$e^{\lambda X} \leq \gamma e^{\lambda b} + (1 - \gamma) e^{\lambda a} = \frac{X - a}{b - a} e^{\lambda b} + \frac{b - X}{b - a} e^{\lambda a}$$

となる．$\mathsf{E}[X] = 0$ に注意して上の式の両辺の期待値を計算すると

$$\begin{aligned}
\mathsf{E}\bigl[e^{\lambda X}\bigr] &\leq -\frac{a}{b - a} e^{\lambda b} + \frac{b}{b - a} e^{\lambda a} \\
&= c e^{\lambda b} + (1 - c) e^{\lambda a} \quad \left(c := -\frac{a}{b - a},\ 1 - c = \frac{b}{b - a}\right) \\
&= c e^{\lambda (1 - c)(b - a)} + (1 - c) e^{-\lambda c (b - a)} \\
&= e^{-\lambda c (b - a)} \bigl\{1 - c + c e^{\lambda (b - a)}\bigr\} \\
&= e^{-cu} \bigl\{1 - c + c e^{u}\bigr\} \quad (u := \lambda (b - a)) \\
&= \exp\bigl\{-cu + \log\bigl(1 - c + c e^{u}\bigr)\bigr\} =: \exp\bigl\{g(u)\bigr\}
\end{aligned}$$

がわかる．いま

$$g(0) = \dot{g}(0) = 0, \quad \ddot{g}(u) \leq \frac{1}{4}\,(u > 0) \tag{3.1}$$

に注意する．実際

$$\begin{aligned}
\dot{g}(u) &= \frac{\mathrm{d}g}{\mathrm{d}u} = -\lambda + \frac{ee^u}{1 - c + ce^u}, \\
\ddot{g}(u) &:= \frac{\mathrm{d}^2 g}{\mathrm{d}u^2} = \frac{ce^u}{1 - c + ce^u} - \frac{c^2 e^{2u}}{(1 - c + ce^u)^2} \\
&= \frac{ce^u(1 - c)}{(1 - c + ce^u)^2} \leq \frac{1}{4} \qquad \left(\text{相加相乗平均を } 1 - c \text{ と } ce^u \text{ に適用する．}\right)
\end{aligned}$$

からわかる．(3.1) に注意して, Taylor 展開をすれば

$$\begin{aligned}
g(u) &= g(0) + u\dot{g}(0) + \frac{u^2}{2}\ddot{g}(\xi) \quad (\xi \in (0,\, u)) \\
&= \frac{u^2}{2}\ddot{g}(\xi) \leq \frac{u^2}{8} = \frac{\lambda^2(b - a)^2}{8}
\end{aligned}$$

がわかる．よって

$$\mathsf{E}\big[e^{\lambda X}\big] \leq e^{g(u)} \leq e^{\lambda^2(b-a)^2/8}$$

を得る． □

定理 3.5. (Hoeffding の不等式) $a_j,\, b_j\,(j = 1,\, 2,\, \ldots,\, n)$ は $a_j < 0 < b_j$ なる定数とする．確率空間 $(\Omega,\, \mathcal{A},\, \mathsf{Pr})$ 上の確率変数 $X_1,\, X_2,\, \ldots,\, X_n$ は独立で

$$\mathsf{E}[X_j] = 0, \quad a_j \leq X_j \leq b_j$$

をみたすとする．このとき, $\forall t > 0$ に対して

$$\mathsf{Pr}\left(\sum_{j=1}^n X_j \geq t\right) \leq \exp\left\{-\frac{t^2}{\sum_{j=1}^n (b_j - a_j)^2}\right\}$$

が成立する．

Proof. 系 3.2 を用いる. $\forall\lambda > 0$ に対して

$$
\begin{aligned}
\mathsf{Pr}\left(\sum_{j=1}^{n} X_j \geq t\right) &\leq e^{-\lambda t}\mathsf{E}\left[\exp\left(\lambda\sum_{j=1}^{n} X_j\right)\right] \\
&= e^{-\lambda t}\prod_{j=1}^{n}\mathsf{E}\bigl[\exp\{\lambda X_j\}\bigr] \quad (\because \text{独立性}) \\
&\leq e^{-\lambda t}\prod_{j=1}^{n}\exp\left\{\frac{\lambda^2(b_j-a_j)^2}{8}\right\} \quad (\because \text{補題 } 3.4) \\
&= \exp\left\{-\lambda t + \frac{\lambda^2}{8}\sum_{j=1}^{n}(b_j-a_j)^2\right\} \\
&= \exp\left[\frac{\sum_{j=1}^{n}(b_j-a_j)^2}{8}\left\{\lambda - \frac{4t}{\sum_{j=1}^{n}(b_j-a_j)^2}\right\}^2\right. \\
&\qquad\left. - \frac{2t^2}{\sum_{j=1}^{n}(b_j-a_j)^2}\right]
\end{aligned}
$$

より

$$
\mathsf{Pr}\left(\sum_{j=1}^{n} X_j \geq t\right) \leq \exp\left\{-\frac{2t^2}{\sum_{j=1}^{n}(b_j-a_j)^2}\right\}
$$

がわかる. □

定理 3.6. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Ber}(\theta)$ とする. ただし $0 < \theta < 1$. このとき $\forall t > 0$ に対して

$$
\mathsf{Pr}\bigl(\bigl|\overline{X}_n - \theta\bigr| \geq t\bigr) \leq 2e^{-2nt^2}
$$

となる. ただし $\overline{X}_n = (1/n)\sum_{j=1}^{n} X_j$ である.

Proof. $j = 1, 2, \ldots, n$ に対して

$$
Y_j = \frac{1}{n}(X_j - \theta)
$$

とおけば

$$
\mathsf{E}[Y_j] = 0, \quad a \leq Y_j \leq b, \quad a = -\frac{\theta}{n}, \quad b = \frac{1-\theta}{n}, \quad (b-a)^2 = \frac{1}{n^2}
$$

となる. よって定理 3.5 より

$$
\mathsf{Pr}\bigl(\overline{X}_n - \theta \geq t\bigr) = \mathsf{Pr}\left(\sum_{j=1}^{n} Y_j \geq t\right) \leq e^{-2nt^2} \tag{3.2}
$$

となる．同様に

$$\mathsf{Pr}\big(\overline{X}_n - \theta \le -t\big) = \mathsf{Pr}\left(\sum_{j=1}^{n} Y_j \le -t\right) = \mathsf{Pr}\left(\sum_{j=1}^{n}(-Y_j) \ge t\right) \le e^{-2nt^2} \tag{3.3}$$

がわかる．よって, (3.2) と (3.3) を合わせると

$$\begin{aligned}\mathsf{Pr}\big(\big|\overline{X}_n - \theta\big| \ge t\big) &= \mathsf{Pr}\big(\big\{\overline{X}_n - \theta \ge t\big\} \cup \big\{\overline{X}_n - \theta \le -t\big\}\big) \\ &\le \mathsf{Pr}\big(\overline{X}_n - \theta \ge t\big) + \mathsf{Pr}\big(\overline{X}_n - \theta \le -t\big) \\ &\le 2e^{-2nt^2}\end{aligned}$$

がわかる．　□

例 3.7. (1) $X_1, X_2, \ldots, X_{100} \overset{\text{i.i.d.}}{\sim} \mathsf{Ber}(\theta)$ とする．ただし $0 < \theta < 1$ である．事象 $\big|\overline{X}_n - \theta\big| \ge 0.2$ の確率を Chebyshev の不等式を用いて θ に関して一様に上から評価すると

$$\mathsf{Pr}\big(\big|\overline{X}_n - \theta\big| \ge 0.2\big) \le 0.0625$$

がわかる．

一方 Hoeffding の不等式を用いて θ に関して一様に評価すると

$$\mathsf{Pr}\big(\big|\overline{X}_n - \theta\big| \ge 0.2\big) \le 2e^{-2(100)(0.2)^2} = 0.00067$$

となる．

(2) $0 < \alpha < 1$ を固定する．いま

$$t = \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}$$

とおく．すると Hoeffding の不等式より

$$\mathsf{Pr}\left(\big|\overline{X}_n - \theta\big| \ge \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}\right) \le \alpha$$

となる．これより

$$\mathsf{Pr}\left(\big|\overline{X}_n - \theta\big| \le \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}\right) > 1 - \alpha$$

を得る．よって

$$C = \left[\overline{X}_n - \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}, \overline{X}_n + \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}\right]$$

とおけば

$$\mathsf{Pr}(\theta \in C) > 1 - \alpha$$

を得る．すなわち信頼係数 $(1-\alpha)$ の p の信頼区間

$$\left[\overline{X}_n - \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}, \overline{X}_n + \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}\right]$$

を得る． □

注意 3.8. 信頼区間については節 8.5 を参照のこと．

定理 3.9. (Mill の不等式) $Z \sim \mathsf{N}(0, 1)$ とする．このとき $\forall t > 0$ に対して

$$\mathsf{Pr}(|Z| \geq t) \leq \sqrt{\frac{2}{\pi}}\frac{e^{-t^2/2}}{t}$$

が成り立つ．

Proof.

$$\begin{aligned}
\mathsf{Pr}(Z \geq t) &= \frac{1}{\sqrt{2\pi}}\int_t^\infty \exp\left(-\frac{z^2}{2}\right) \mathrm{d}z \\
&\leq \frac{1}{\sqrt{2\pi}}\int_t^\infty \frac{z}{t}\exp\left(-\frac{z^2}{2}\right) \mathrm{d}z \\
&= \frac{1}{t\sqrt{2\pi}}\int_t^\infty \left\{-\frac{\partial}{\partial z}\exp\left(-\frac{z^2}{2}\right)\right\} \mathrm{d}z \\
&= \frac{1}{t\sqrt{2\pi}}\left(-\frac{t^2}{2}\right).
\end{aligned}$$

2 番目の不等式は $\mathsf{N}(0, 1)$ の p.d.f. が偶関数であることからわかる．3 番目の不等式は

$$\begin{aligned}
\mathsf{Pr}(|Z| \geq t) &= \mathsf{Pr}(\{Z \geq t\} \cup \{Z \leq -t\}) \\
&\leq \mathsf{Pr}(Z \geq t) + \mathsf{Pr}(Z \leq -t) \\
&= 2\mathsf{Pr}(Z \geq t)
\end{aligned}$$

からわかる． □

3.2　期待値に対する不等式

定理 3.10. (Cauchy-Schwarz の不等式) 確率変数 X と Y は 2 次の有限な期待値を持つとき

$$\mathsf{E}\big[|XY|\big] \leq \sqrt{\mathsf{E}[X^2]\,\mathsf{E}[Y^2]}$$

となる.

Proof. $\mathsf{E}[X^2]=\mathsf{E}[Y^2]=0$ のとき不等式は自明である. $\mathsf{E}[X^2]\neq 0$ として証明を進める. いま, $g(t)=\mathsf{E}\big[(tX-Y)^2\big]$ とおく. 期待値の中を展開して期待値の線型性を用いると

$$
\begin{aligned}
0\leq g(t)&=\mathsf{E}\big[t^2X^2-2tXY+Y^2\big]\\
&=\mathsf{E}[X^2]\left\{t-\frac{\mathsf{E}[XY]}{\mathsf{E}[X^2]}\right\}^2+\frac{\mathsf{E}[X^2]\mathsf{E}[Y^2]-\mathsf{E}[XY]}{\mathsf{E}[X^2]}
\end{aligned}
$$

となる.

$$
g\left(\frac{\mathsf{E}[XY]}{\mathsf{E}[X^2]}\right)=\frac{\mathsf{E}[X^2]\mathsf{E}[Y^2]-\mathsf{E}[XY]}{\mathsf{E}[X^2]}\geq 0\;\Leftrightarrow\;\mathsf{E}[X^2]\mathsf{E}[Y^2]-\mathsf{E}[XY]\geq 0
$$

より, 不等式は示された. 等号が成立するのは $g(t)=0$ が重解を持つときである. 重解を c とおけば

$$
g(t)=\mathsf{E}[(cX-Y)^2]=0\;\Leftrightarrow\;\mathsf{Pr}(Y=cX)=1
$$

となる[1]. □

定義 3.11. 関数 $g:\mathbb{R}\to\mathbb{R}$ が凸であるとは各 $x,y\in\mathbb{R}$ と $0\leq\forall t\leq 1$ に対して

$$
g(tx+(1-t)y)\leq tg(x)+(1-t)g(y)
$$

が成立するときをいう. さらに $-g$ が凸のとき g は concave であるという.

定理 3.12. (Jensen の不等式) X を有限な期待値を持つ確率変数とする.
(1) 関数 $g:\mathbb{R}\to\mathbb{R}$ は凸で $g(X)$ の期待値は有限のとき

$$
\mathsf{E}[g(X)]\geq g(\mathsf{E}[X])
$$

となる.
(2) g が concave のとき

$$
\mathsf{E}[g(X)]\leq g(\mathsf{E}[X])
$$

となる.

[1]非負値確率変数 X に対して

$$
\mathsf{E}[X]=0\;\Leftrightarrow\;\mathsf{Pr}(X=0)=1
$$

であることに注意せよ. 実際 $\mathsf{Pr}(X>0)>0$ と仮定する. するとあるの $\epsilon>0$ が存在して $\mathsf{Pr}(X>\epsilon)>0$ となる. しかし $X\geq\epsilon\mathbb{1}\{X>\epsilon\}$ より

$$
0=\mathsf{E}[X]\geq\epsilon\mathsf{E}[\mathbb{1}\{X>\epsilon\}]=\epsilon\mathsf{Pr}(X>\epsilon)>0
$$

となり矛盾する. よって $\mathsf{E}[X]=0\Leftrightarrow\mathsf{Pr}(X=0)=1$ がわかる.

Proof. $\forall x \in \mathbb{R}$ に対してある定数 $r \in \mathbb{R}$ が存在[2]して

$$g(\mathsf{E}[X]) + r\{x - \mathsf{E}[X]\} \leq g(x)$$

となる. x に X を代入して上の不等式の両辺の期待値を取れば

$$g(\mathsf{E}[X]) \leq \mathsf{E}[g(X)]$$

がわかる. □

系 3.13. (Young の不等式) $p,\, q > 1$ とし

$$\frac{1}{p} + \frac{1}{q} = 1$$

をみたすとする. このとき $\forall a,\, b > 0$ に対して

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}$$

となる.

Proof. 関数 g を凸とし, $Y \sim \mathsf{Unif}(0,\, 1)$ とする. 可積分関数 $h : \mathbb{R} \longrightarrow \mathbb{R}$ に対して $X = h(Y)$ とすれば

$$g\left(\int_0^1 h(y)\,\mathrm{d}y\right) = g(\mathsf{E}[X]) \leq \mathsf{E}[g(X)] = \int_0^1 g(h(y))\,\mathrm{d}y$$

を得る. ここで

$$g(x) = e^x, \quad h(y) = \begin{cases} p\log a & \left(0 \leq y < \frac{1}{p}\right) \\ q\log b & \left(\frac{1}{p} \leq y \leq 1\right) \end{cases}$$

とおけば g は凸なので

$$\begin{aligned} ab &= \exp\left\{\frac{p\log a}{p} + \frac{q\log b}{q}\right\} = \exp\left\{\int_0^1 h(y)\,\mathrm{d}y\right\} \leq \int_0^1 \exp\bigl\{h(y)\bigr\}\,\mathrm{d}y \\ &= \frac{1}{p}\exp\bigl\{p\log a\bigr\} + \frac{1}{q}\exp\bigl\{q\log b\bigr\} = \frac{a^p}{p} + \frac{b^q}{q} \end{aligned}$$

がわかる. □

3.3　章末注釈と参考文献

この章で扱った凸関数は統計的推測理論では重要な役割を果たす. Hoeffding の不等式では, 統計的機械学習理論で用いられる基本不等式である. この章は [23, 4 章] を参考にした.

[2]定理 D.9 を参照.

3.4　演習問題

演習問題 3.1. $X \sim \mathsf{Po}(\theta)\,(\theta > 0)$ とする．このとき，以下の問いに答えよ．
(1) X の平均 $\mathsf{E}[X]$ と分散 $\mathsf{Var}[X]$ を求めよ．
(2) Chebyshev の不等式を用いて

$$\mathsf{Pr}\big(X \geq 2\theta\big) \leq \frac{1}{\theta}$$

を示せ．

演習問題 3.2. $n \in \mathbb{N}$ とする．$1 < p < \infty$, $p^{-1} + q^{-1} = 1$ のとき，任意の実数 $x_j,\, y_j\,(j = 1,\, 2,\, \ldots,\, n)$ に対して

$$\sum_{j=1}^{n} |x_j y_j| \leq \left(\sum_{j=1}^{n} |x_j|^p\right)^{1/p} \left(\sum_{j=1}^{n} |y_j|^q\right)^{1/q}$$

が成り立つことを Young の不等式を利用して示せ．この不等式を Hölder の不等式という．

演習問題 3.3.

第4章　確率変数列と分布の収束

確率論の最も重要な側面のひとつは確率変数列の挙動に関することである．確率論のこの部分のことを「大標本論」,「極限論」,「漸近論」と数理統計学では呼んでいる．大標本論の基本的な問いは次である．確率変数列 $X_1, X_2, \ldots$ の極限の振る舞いについて言えることは何だろうか?統計学はデータの収集にかかわる学問である．したがってデータを集めれば集めるほど何が起こるかを調べることは重要である．

すこし実解析の話題を復習する．実数列 $\{x_n\}$ が点 x に収束するとは,任意の $\epsilon > 0$ に対してある正の整数 $N \in \mathbb{N}$ が存在して

$$\forall n > N \Rightarrow |x_n - x| < \epsilon$$

が成り立つことである．このことを $\lim_{n\to\infty} x_n = x$ と書いた．たとえば $x_n = x\,(\forall n \in \mathbb{N})$ ならば, 明らかに $\lim_{n\to\infty} x_n = x$ となる．これと同じことを確率空間 $\big(\Omega, \mathcal{A}, \mathsf{Pr}\big)$ 上の確率変数列について考えてみる．

$$X_1, X_2, \ldots \overset{\text{i.i.d.}}{\sim} \mathsf{N}(0, 1)$$

とする．すなわちこの確率変数列の任意の有限個の確率変数列は独立同一に $\mathsf{N}(0, 1)$ に従う．さらに別の確率変数 X も $\mathsf{N}(0, 1)$ に従うとする．このとき X_n は X に「収束」するとしたい．しかし

$$\mathsf{Pr}\big(X_n \neq X\big) = 1 \qquad (\forall n \in \mathbb{N})$$

である．

別の例をあげる．$X_1, X_2, \ldots$ は独立で

$$X_n \sim \mathsf{N}\left(0, \frac{1}{n}\right) \qquad (n = 1, 2, \ldots)$$

とする．直観的には, 十分大きな n に対して X_n は 0 の近辺に集中すると予想するだろう．しかし X_n は連続型確率変数なので, すべての n に対して

$$\mathsf{Pr}\big(X_n = 0\big) = 0$$

である．これらの例から, 確率変数列の収束については, 実数列の収束とは異なる道具立てが必要になることがわかる．この章では確率変数列の収束の定義を述べ, これに関わる基本的な事項をまとめたうえで次の重要な事項を説明する．

(1) **大数の法則**. $X_1, X_2, \ldots$ を i.i.d. 確率変数列とし

$$\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j, \qquad \mu = \mathsf{E}[X_1]\,(-\infty < \mu < \infty)$$

とする．このとき $\overline{X}_n$ は高い確率で μ の近くにいることを保証する定理である．

(2) **中心極限定理**. $\mathsf{Var}[X_1] = \sigma^2\,(0 < \sigma < \infty)$ とする．n が十分大きいとき $\sqrt{n}\big(\overline{X}_n - \mu\big)$ の分布は正規分布で近似できることを保証する定理である．

以上の事項について証明なしで主張まずを紹介する．証明については補遺 G にまとめることにする．

4.1　確率変数列の収束のタイプ

定義 4.1. 確率空間 $\big(\Omega, \mathcal{A}, \mathsf{Pr}\big)$ 上の $X, X_1, X_2, \ldots$ を確率変数列とする．さらに, 各 $X_n\,(n = 1, 2, \ldots)$ の c.d.f. を F_n, X の c.d.f. を F とする．

(1) 確率変数列 $\{X_n\}$ は確率変数 X に**確率収束**するとは, $\forall\epsilon > 0$ に対して

$$\lim_{n\to\infty} \mathsf{Pr}\big(\big|X_n - X\big| \geq \epsilon\big) = 0$$

が成り立つときをいう．このことを $X_n \xrightarrow{\mathrm{P}} X$ と書く．

(2) 確率変数列 $\{X_n\}$ は確率変数 X に**概収束**するとは

$$\mathsf{Pr}\left(\omega \in \Omega;\ \lim_{n\to\infty} X_n(\omega) = X(\omega)\right) = 0$$

が成り立つときをいう．このことを $X_n \xrightarrow{\mathrm{a.s.}} X$ と書く．

(3) 確率変数列 $\{X_n\}$ は確率変数 X に**分布収束**するとは, F のすべての連続点 x において

$$\lim_{n\to\infty} \mathsf{F}_n(x) = \mathsf{F}(x)$$

が成り立つときをいう．このことを $X_n \rightsquigarrow X$ と書く．また, $X_n \rightsquigarrow X$ かつ $X \sim \mathsf{N}(0, 1)$ のときには $X_n \rightsquigarrow \mathsf{N}(0, 1)$ とも書くことがある．

注意 4.2. (1) c を定数とする．$\mathsf{Pr}(X = c) = 1$ かつ $X_n \xrightarrow{\mathrm{P}} X$ とき $X_n \xrightarrow{\mathrm{P}} c$ と記す．
(2) $\mathsf{Pr}(X = c) = 1$ かつ $X_n \rightsquigarrow X$ とき $X_n \rightsquigarrow c$ と記す．　□

定義 4.3. $X_1, X_2, \ldots$ を確率変数列とし X を別の確率変数とする．さらに $\mathsf{E}\left[X_n^2\right] < \infty$, $\mathsf{E}[X^2] < \infty\,(n = 1, 2, \ldots)$ とする．確率変数列 $\{X_n\}$ が確率変数 X に**平均 2 乗の意味で収束**するとは

$$\lim_{n\to\infty} \mathsf{E}\left[(X_n - X)^2\right] = 0$$

が成り立つときをいう．このことを $X_n \overset{\text{q.m.}}{\to} c$ と記す

注意 4.4. 確率収束と概収束の違いを説明する．$X, X_1, X_2, \ldots$ を同じ確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上で定義された確率変数列とし

$$M_n(\omega) := \sup_{k\geq n} |X_k(\omega) - X(\omega)| \qquad (\omega \in \Omega)$$

とおく．すると

$$X_n \overset{\text{a.s.}}{\to} X \Leftrightarrow M_n \overset{\text{P}}{\to} 0$$

となることにが以下の議論からわかる．

まず, $\forall \ell \in \mathbb{N}$ に対して, ある十分大きな $n_0 \in \mathbb{N}$ があって

$$\forall n \geq n_0 \Rightarrow M_n(\omega) < \frac{1}{\ell}$$

ならば

$$M_n(\omega) \overset{n\to\infty}{\longrightarrow} 0$$

となることを示す．ここで, $\ell \in \mathbb{N}$ に対して

$$B_\ell := \bigcup_{n=n_0}^{\infty} \left\{\omega \in \Omega;\, M_n(\omega) < \frac{1}{\ell}\right\}$$

と定めると

$$\left\{\omega \in \Omega;\, \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\} = \bigcap_{\ell=1}^{\infty} B_\ell$$

と書けるので

$$X_n \overset{\text{a.s.}}{\to} X \Leftrightarrow \mathsf{Pr}\left(\bigcap_{\ell=1}^{\infty} B_\ell\right) = 1$$

がわかる．

次に, $\ell_1 < \ell_2\,(\ell_1, \ell_2 \in \mathbb{N})$ に対し

$$B_{\ell_1} = \bigcup_{n=1}^{\infty} \left\{\omega \in \Omega;\, M_n(\omega) < \frac{1}{\ell_1}\right\} \supset \bigcup_{n=1}^{\infty} \left\{\omega \in \Omega;\, M_n(\omega) < \frac{1}{\ell_2}\right\} = B_{\ell_2}$$

なので

$$B_1 \supset B_2 \supset \cdots \supset B_\ell \supset \cdots \qquad \left(\Rightarrow \mathsf{Pr}(B_1) \geq \mathsf{Pr}(B_2) \geq \cdots \geq \mathsf{Pr}(B_\ell) \geq \cdots\right)$$

となる．このことと補題 1.7(7) から

$$\mathsf{Pr}\left(\bigcap_{\ell=1}^{\infty} B_\ell\right) = \lim_{\ell\to\infty} \mathsf{Pr}(B_\ell) \leq \mathsf{Pr}(B_\ell) \leq 1\,(\forall \ell \in \mathbb{N})$$

となる．よって

$$1 = \mathsf{Pr}\left(\bigcap_{\ell=1}^{\infty} B_\ell\right) \Leftrightarrow \mathsf{Pr}(B_\ell) = 1\,(\forall \ell \in \mathbb{N})$$

がわかる．さらに，$n_1 < n_2\,(n_1,\, n_2 \in \mathbb{N})$ に対して

$$M_{n_1}(\omega) := \sup_{k \geq n_1} |X_k(\omega) - X(\omega)| \geq \sup_{k \geq n_2} |X_k(\omega) - X(\omega)|$$

から

$$\left\{\omega \in \Omega;\, M_{n_1}(\omega) < \frac{1}{\ell}\right\} \subset \left\{\omega \in \Omega;\, M_{n_2}(\omega) < \frac{1}{\ell}\right\}$$

がわかる． よって，$\left\{\omega \in \Omega;\, M_n(\omega) < \dfrac{1}{\ell}\right\}_{n=1}^{\infty}$ は増大列となるで，補題 1.7(6) から

$$\begin{aligned}\mathsf{Pr}(B_\ell) &= \mathsf{Pr}\left(\bigcup_{n=1}^{\infty}\left\{\omega \in \Omega;\, M_n(\omega) < \frac{1}{\ell}\right\}\right)\\ &= \lim_{n\to\infty} \mathsf{Pr}\left(\left\{\omega \in \Omega;\, M_n(\omega) < \frac{1}{\ell}\right\}\right)\end{aligned}$$

となる．よって

$$\begin{aligned}1 = \mathsf{Pr}\left(\bigcap_{\ell=1}^{\infty} B_\ell\right) &\Leftrightarrow \mathsf{Pr}(B_\ell) = 1\,(\forall \ell \in \mathbb{N})\\ &\Leftrightarrow \lim_{n\to\infty} \mathsf{Pr}\left(\left\{\omega \in \Omega;\, M_n(\omega) < \frac{1}{\ell}\right\}\right) = 1\,(\forall \ell \in \mathbb{N})\\ &\Leftrightarrow \lim_{n\to\infty} \mathsf{Pr}\left(\left\{\omega \in \Omega;\, M_n(\omega) \geq \frac{1}{\ell}\right\}\right) = 0\,(\forall \ell \in \mathbb{N})\end{aligned}$$

がわかる．したがって

$$X_n \xrightarrow{\text{a.s.}} X \Leftrightarrow M_n \xrightarrow{\text{P}} X \tag{4.1}$$

となる．以上の議論から，概収束は n ステップ後の $|X_k - X|\,(k \geq n)$ が最大になるものが 0 に確率収束することと同値である． □

例 4.5. $X_1, X_2, \ldots$ は同じ確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上で定義された独立な確率変数列で

$$X_n \sim \mathsf{Ber}(p_n)\,(n \in \mathbb{N})$$

とする．ただし, $0 < p_n < 1$ とする．すると

$$X_n \xrightarrow{\mathrm{P}} 0 \Leftrightarrow p_n \xrightarrow{n\to\infty} 0$$

である．実際, 十分小さな任意の $\epsilon > 0$ に対して, $\mathsf{Pr}(X_n \leq -\epsilon) = 0$ と $X_n(\omega) \geq \epsilon \Leftrightarrow X_n(\omega) = 1\,(\omega \in \Omega)$ であることに注意すると

$$\begin{aligned}\mathsf{Pr}(|X_n| \geq \epsilon) &= \mathsf{Pr}\big(\{X_n \leq -\epsilon\} \cup \{X_n \geq \epsilon\}\big) = \mathsf{Pr}(X_n \geq \epsilon) = \mathsf{Pr}(X_n = 1)\\ &= p_n\end{aligned}$$

がわかる．

次に

$$M_n = \sup_{k \geq n} |X_k - X|$$

とし, $\pi_n = \mathsf{Pr}(M_n = 1)$ とおく．すると $M_n(\omega) \in \{0, 1\}\,(\omega \in \Omega)$ なので

$$\begin{aligned}1 - \pi_n = \mathsf{Pr}(M_n = 0) &= \mathsf{Pr}\left(\bigcap_{k=n}^{\infty}\{X_k = 0\}\right)\\ &= \prod_{k=n}^{\infty}\mathsf{Pr}(X_k = 0) \qquad (\because \{X_n\}_{n=1}^{\infty} \text{ は独立})\\ &= \prod_{k=n}^{\infty}(1 - p_k)\end{aligned}$$

を得る．すると (4.1) と $X_n \xrightarrow{\mathrm{a.s.}} 0$ から $\pi_n \xrightarrow{n\to\infty} 0$ がわかる．ここで

$$\prod_{k=n}^{\infty}(1 - p_k) \to 1\,(n \to \infty) \Leftrightarrow \sum_{k=n}^{\infty} p_k < \infty\,(n \to \infty)$$

であること[1]に注意すると

$$X_n \xrightarrow{\mathrm{a.s.}} 0 \Leftrightarrow \sum_{k=n}^{\infty} p_k < \infty\,(n \to \infty)$$

がわかる．

たとえば, $p_n = \dfrac{1}{n}\,(n \in \mathbb{N})$ とすると

$$X_n \xrightarrow{\mathrm{P}} 0$$

[1]たとえば, [25, pp.91 − 92] を参照のこと．

だが

$$\mathsf{Pr}\left(\left\{\omega\in\Omega;\ \lim_{n\to\infty}X_n(\omega)=0\right\}\right)<1\ \Leftrightarrow\ X_n\overset{\text{a.s.}}{\to}0\ \text{は偽}$$

がわかる. □

注意 4.6. $n=1,2,\ldots$ に対して $X_n\sim\mathsf{N}(0,1/n)$ とする. $\{X_n\}_{n=1}^{\infty}$ は 0 に収束することが期待される.

まず分布収束について確認する. そのために次のような c.d.f. を考える.

$$\mathsf{F}(x)=\begin{cases}1 & (x\geq 0)\\ 0 & (x<0)\end{cases}$$

と定義する. すなわち $\mathsf{Pr}(X=0)=1$ をみたす確率変数 X の c.d.f. である. $\mathsf{F}(x)$ は $x=0$ で不連続であることに注意する. したがって $x=0$ 以外での各点収束を言えばよい.

そのために $\sqrt{n}X_n\sim\mathsf{N}(0,1)\,(n=1,2,\ldots)$ であることに注意する. ただし, Z は標準正規分布 $\mathsf{N}(0,1)$ に従う確率変数とする. $x<0$ のとき

$$\begin{aligned}\mathsf{F}_n(x)&=\mathsf{Pr}\big(X_n\leq x\big)=\mathsf{Pr}\big(\sqrt{n}X_n\leq\sqrt{n}x\big)=\mathsf{Pr}(Z\leq\sqrt{n}x)\\&=\int_{-\infty}^{\sqrt{n}x}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}\,\mathrm{d}z\to 0\quad(n\to\infty)\end{aligned}$$

となる. 最後の極限は $\sqrt{n}x\overset{n\to\infty}{\longrightarrow}-\infty$ より $\mathsf{F}(x)=0\,(x<0)$ がわかる.

つぎに $x>0$ のとき

$$\begin{aligned}\mathsf{F}_n(x)&=\mathsf{Pr}\big(X_n\leq x\big)=\mathsf{Pr}\big(\sqrt{n}X_n\leq\sqrt{n}x\big)=\mathsf{Pr}(Z\leq\sqrt{n}x)\\&=\int_{-\infty}^{\sqrt{n}x}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}\,\mathrm{d}z\to 1\quad(n\to\infty)\end{aligned}$$

となる. 最後の極限は $\sqrt{n}x\overset{n\to\infty}{\longrightarrow}\infty$ よりわかる. よって, $\mathsf{F}(x)=1\,(x>0)$ である.

以上の議論から $x\neq 0$ のとき

$$\lim_{n\to\infty}\mathsf{F}_n(x)=\mathsf{F}(x)$$

がわかる.

しかし, $x=0$ は $\mathsf{F}(x)$ の不連続点なので, $\mathsf{F}_n(0)=1/2\neq 1=F(0)$ であることは問題ない.

最後に, 確率収束を示そう. Markov の不等式 (定理 3.1) より, $\forall\epsilon>0$ に対して

$$\mathsf{Pr}\big(|X_n|\geq\epsilon\big)=\mathsf{Pr}\big(X_n^2\geq\epsilon^2\big)\leq\frac{\mathsf{E}\big[X_n^2\big]}{\epsilon^2}=\frac{1}{n\epsilon^2}\overset{n\to\infty}{\longrightarrow}0$$

となる．最後の等号は，$X_n \sim \mathsf{N}(0,\,1/2)$ なので，$\mathsf{E}[X_n^2] = \dfrac{1}{n}$ となることを用いた．よって $X_n \xrightarrow{\mathrm{P}} 0$ である． □

次に収束のタイプ間の関係について述べる．

定理 4.7. 次に関係が成立する．

(1) $X_n \xrightarrow{qm} X \;\Rightarrow\; X_n \xrightarrow{\mathrm{P}} X$ である．

(2) $X_n \xrightarrow{\mathrm{P}} X \;\Rightarrow\; X_n \rightsquigarrow X$ である．

(3) $X_n \rightsquigarrow X$ かつある定数 c があって $\mathsf{Pr}(X = c) = 1$ のとき $X_n \xrightarrow{\mathrm{P}} c$ である．

Proof. 節 4.4.5.2 で示す． □

注意 4.8. 定理 4.7(1)(2) の逆は一般に成立しない．
(1) **定理** 4.7(1) **の逆の反例**．$U \sim \mathsf{Unif}(0,\,1)$, $X_n = \sqrt{n}\mathbb{1}_{(0,\,1/n)}(U)\;(n = 1,\,2,\,\ldots)$, $X = 0$ とおく．このとき $\forall\epsilon > 0$ に対して

$$\begin{aligned}\mathsf{Pr}\big(\big|X_n\big| \geq \epsilon\big) &= \mathsf{Pr}\big(\sqrt{n}\mathbb{1}_{(0,\,1/n)}(U) \geq \epsilon\big) = \mathsf{Pr}\left(0 < U < \frac{1}{n}\right)\\ &= \mathsf{Pr}\left(0 \leq U \leq \frac{1}{n}\right) = \frac{1}{n} \xrightarrow{n\to\infty} 0\end{aligned}$$

となる．よって $X_n \xrightarrow{\mathrm{P}} X$ である．しかし

$$\mathsf{E}\big[X_n^2\big] = n\int_0^{1/n} \mathrm{d}u = 1$$

なので，$X_n \xrightarrow{qm} X$ は成立しない．
(2) **定理** 4.7(2) **の逆の反例**．$X \sim \mathsf{N}(0,\,1)$ とし，$X_n = -X\;(n = 1,\,2,\,\ldots)$ とする．したがって，$X_n \sim \mathsf{N}(0,\,1)$ である．明らかに

$$\lim_{n\to\infty} \mathsf{F}_n(x) = \mathsf{F}(x)$$

である．よって $X_n \rightsquigarrow X$ となる．しかし $\forall\epsilon > 0$ に対して

$$\mathsf{Pr}\big(\big|X_n - X\big| > \epsilon\big) = \mathsf{Pr}(|2X| > \epsilon) = \mathsf{Pr}\left(|X| > \frac{\epsilon}{2}\right) \neq 0$$

である．よって $X_n \xrightarrow{\mathrm{P}} X$ は成立しない． □

定理 4.9 (Portmanteau の補題)**.** 確率変数列 $X_n\,(n=1,2,\ldots)$, X に対し, X_n と X の分布関数を F_n と F とそれぞれ書く. このとき, 以下の (1) ∼ (7) は同値である.
(1) $X_n \rightsquigarrow X$ である.
(2) $\mathbb{R}$ 上の任意の有界連続関数 g に対して

$$\lim_{n\to\infty} \mathsf{E}[g(X_n)] = \mathsf{E}[g(X)]$$

である.
(3) $\mathbb{R}$ 上の任意の有界 Lipshitz 連続関数 g に対して

$$\lim_{n\to\infty} \mathsf{E}[g(X_n)] = \mathsf{E}[g(X)]$$

である.
(4) $\mathbb{R}$ 上の任意の非負値連続関数 g に対して

$$\liminf_{n\to\infty} \mathsf{E}[g(X_n)] \geq \mathsf{E}[g(X)]$$

である.
(5) $\mathbb{R}$ の任意の開集合 O に対して

$$\liminf_{n\to\infty} \mathsf{Pr}(X_n \in O) \geq \mathsf{Pr}(X \in O)$$

である.
(6) $\mathbb{R}$ の任意の閉集合 C に対して

$$\limsup_{n\to\infty} \mathsf{Pr}(X_n \in C) \leq \mathsf{Pr}(X \in C)$$

である.
(7) $\mathbb{R}$ の任意の Borel 集合 B が $\mathsf{Pr}(X \in \partial B) = 0$ ならば

$$\lim_{n\to\infty} \mathsf{Pr}(X_n \in B) = \mathsf{Pr}(X \in B)$$

である. ただし, ∂B は B の境界である.

Proof. 節 4.4.5.3 で示す. ただし, 証明には積分の収束定理の知識が必要である. □

注意 4.10. 定理 4.9 を有限次元の確率ベクトルに拡張することができる. 記号が煩雑にはなるが, 証明は本質的に同じである. □

定理 4.11. X_n, X, Y_n, Y $(n=1,2,\ldots)$ は確率変数列とする. $g:\mathbb{R}\to\mathbb{R}$ を連続関数とし, c を定数とする.

(1) $X_n \xrightarrow{\mathrm{P}} X$ かつ $Y_n \xrightarrow{\mathrm{P}} Y \;\Rightarrow\; X_n + Y_n \xrightarrow{\mathrm{P}} X + Y$ である.

(2) $X_n \xrightarrow{qm} X$ かつ $Y_n \xrightarrow{qm} Y \;\Rightarrow\; X_n + Y_n \xrightarrow{qm} X + Y$ である.

(3) $X_n \rightsquigarrow X$ かつ $Y_n \xrightarrow{\mathrm{P}} c \;\Rightarrow\; X_n + Y_n \rightsquigarrow X + c$ である.

(4) $X_n \xrightarrow{\mathrm{P}} X$ かつ $Y_n \xrightarrow{\mathrm{P}} Y \;\Rightarrow\; X_n Y_n \xrightarrow{\mathrm{P}} XY$ である.

(5) $X_n \rightsquigarrow X$ かつ $Y_n \xrightarrow{\mathrm{P}} c \;\Rightarrow\; X_n Y_n \rightsquigarrow cX$ である.

(6) $X_n \xrightarrow{\mathrm{P}} X \;\Rightarrow\; g(Y_n) \xrightarrow{\mathrm{P}} g(Y)$ である.

(7) $X_n \rightsquigarrow X \;\Rightarrow\; g(X_n) \rightsquigarrow g(X)$ である.

Proof. 節 4.4.5.4 で示す. □

4.2 大数の法則

定理 4.12. (大数の弱法則) $X_1, X_2, \ldots$ は確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の i.i.d. 確率変数列とする. $\mathsf{E}\big[|X_1|\big] < \infty$ のとき

$$\overline{X}_n := \frac{1}{n}\big(X_1 + X_2 + \cdots + X_n\big) \xrightarrow{\mathrm{P}} \mu, \qquad \mu = \mathsf{E}[X_1]$$

が成立する.

Proof. より強い条件 $\mathsf{E}\big[X_1^2\big] < \infty$ のもとで定理の主張を証明する. $\mathsf{Var}[X_1] = \sigma^2$ とおく. Markov の不等式 (定理 3.1) より

$$\mathsf{Pr}\big(|\overline{X}_n - \mu| \ge \epsilon\big) \le \frac{\mathsf{Var}\big[\overline{X}_n\big]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \xrightarrow{n\to\infty} 0$$

よりわかる. □

大数の強法則を述べる前に, 定理の証明に必要な補題を述べる.

補題 4.13. $X_1, X_2, \ldots$ は非負値確率変数列で, 条件

$$\sum_{n=1}^{\infty} \mathsf{E}[X_n] < \infty \tag{4.2}$$

をみたすとする. このとき

$$\mathsf{Pr}\left(\omega \in \Omega;\, \sum_{n=1}^{\infty} X_n(\omega) < \infty\right) = 1$$

が成り立つ.

Proof. 証明は節 4.5.1 で行う. □

定理 4.14. (大数の強法則) $X_1, X_2, \ldots$ は確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の i.i.d. 確率変数列とし, $\mathsf{E}\big[|X_1|^4\big] < \infty$ とする. このとき

$$\overline{X}_n := \frac{1}{n}\big(X_1 + X_2 + \cdots + X_n\big) \overset{\text{a.s.}}{\to} \mu, \qquad \mu = \mathsf{E}[X_1] \tag{4.3}$$

が成立する.

Proof. まず

$$\mathsf{E}\big[|X_1|^4\big] =: K < \infty, \quad T_n = X_1 + X_2 + \cdots + X_n, \quad \overline{X}_n = \frac{T_n}{n}$$

とおく.

最初に, $\mu = 0$ として (4.3) を示す. 以下の事象の包含関係に注意する. すなわち

$$\left\{\omega \in \Omega;\ \sum_{n=1}^{\infty}\big(\overline{X}_n(\omega)\big)^4 < \infty\right\} \subset \left\{\omega \in \Omega;\ \lim_{n\to\infty} \overline{X}_n(\omega) = 0\right\}$$

が成立する. このことより

$$\mathsf{Pr}\left(\sum_{n=1}^{\infty} \overline{X}_n^4 < \infty\right) = 1 \tag{4.4}$$

がわかると

$$1 = \mathsf{Pr}\left(\sum_{n=1}^{\infty} \overline{X}_n^4 < \infty\right) \le \mathsf{Pr}\left(\lim_{n\to\infty} \overline{X}_n = 0\right)$$

となり, $\mu = 0$ のときに (4.3) がわかる.

以下では $\mu = 0$ として (4.4) を示す. そのために多項定理を用いて, T_n を展開する. すると

$$\big(X_1 + \cdots + X_n\big)^4 = \sum_{\substack{\ell_1 + \cdots + \ell_n = 4 \\ \ell_j \ge 0\,(j=1,\ldots,n)}} \frac{4!}{\ell_1! \times \ell_2! \times \cdots \times \ell_n!} X_1^{\ell_1} \times X_2^{\ell_2} \times \cdots \times X_n^{\ell_n}$$

となる. ただし $\ell_1, \ell_2, \ldots, \ell_n$ は 0 以上の整数である. ここで $X_1^{\ell_1}, X_2^{\ell_2}, \ldots, X_n^{\ell_n}$ は独立あることに注意すると

$$\mathsf{E}\big[T_n^4\big] = \sum_{\substack{\ell_1 + \ell_2 + \cdots + \ell_n = 4 \\ \ell_j \ge 0\,(j=1,2,\ldots,n)}} \frac{4!}{\ell_1! \ell_2! \times \cdots \times \ell_n!} \mathsf{E}\big[X_1^{\ell_1}\big] \times \mathsf{E}\big[X_2^{\ell_2}\big] \times \cdots \times \mathsf{E}\big[X_n^{\ell_n}\big]$$

がわかる．さらに, $\mathsf{E}\big[X_k\big] = \mu = 0\,(k = 1,\,2,\,\ldots,\,n)$ と $\dfrac{4!}{2!2!} = 6$ を用いると

$$\mathsf{E}\big[T_n^4\big] = \sum_{k=1}^{n} \mathsf{E}\big[X_k^4\big] + 6 \sum_{1\le k<\ell\le n} \mathsf{E}\big[X_k^2\big]\,\mathsf{E}\big[X_\ell^2\big]$$

を得る．ここで $\mathsf{E}\big[X_k^4\big] = K\,(k = 1,\,2,\ldots,\,n)$ と Cauchy-Schwarz の不等式から

$$\mathsf{E}\big[X_k^2\big] \le \sqrt{\mathsf{E}\big[X_k^4\big]} = \sqrt{K}$$

となることに注意すると

$$\mathsf{E}\big[T_n^4\big] \le nK + 6 \sum_{1\le k<\ell\le n} \sqrt{K}\,\sqrt{K} = nK + 3n(n-1)K \le 3Kn^2$$

を得る．よって

$$\sum_{n=1}^{\infty} \mathsf{E}\big[\overline{X}_n^4\big] = \sum_{n=1}^{\infty} \frac{1}{n^4}\mathsf{E}\big[T_n^4\big] \le 3K \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$$

となる．したがって 補題 4.13 から (4.4) が成立することがわかる．

つぎに $\mu \ne 0$ の場合を示す．$Y_k = X_k - \mu\,(k = 1,\,2,\,\ldots,\,n)$ とおくと

$$\mathsf{E}\big[Y_k^4\big] = \mathsf{E}\Big[\big\{|X_k| + |\mu|\big\}^4\Big] \le 8\mathsf{E}\big[|X_k|^4 + |\mu|^4\big] \le 8\big(K + \mu^4\big) < \infty$$

が成り立つ．したがってこの定理の証明の前半部で得られた結果から

$$\mathsf{Pr}\left(\lim_{n\to\infty} \frac{1}{n}\big(Y_1 + Y_2 + \cdots + Y_n\big) = 0\right) = 1$$

となる．最後に, $\forall\omega \in \Omega$ に対して

$$\begin{aligned}&\lim_{n\to\infty} \frac{1}{n}\big(Y_1(\omega) + Y_2(\omega) + \cdots + Y_n(\omega)\big) = 0\\ &\qquad \Leftrightarrow \lim_{n\to\infty} \frac{1}{n}\big(X_1(\omega) + X_2(\omega) + \cdots + X_n(\omega)\big) = \mu\end{aligned}$$

に注意すればよい．　□

注意 4.15. 大数の法則は $\mathsf{E}\big[|X_1|\big] < \infty$ で成立する．この条件下での大数の強法則の証明は定理 G.4 で記した．　□

系 4.16 (Weierstrass の近似定理)**.** 閉区間 $[0, 1]$ 上の任意の連続関数 f は多項式の極限として表すことができる．特に，$f(0) = f(1) = 0$ のとき[2], $n \in \mathbb{N}$ に対して

$$B_n(x) = \sum_{j=1}^{n} f\left(\frac{j}{n}\right)\left(\begin{array}{c} n \\ j \end{array}\right) x^j (1-x)^{n-j}$$

とおいたとき

$$f(x) = \lim_{n\to\infty} B_n(x) \tag{4.5}$$

が成立する．

Proof. $x = 0, 1$ は (4.5) は明らかなので, $0 < x < 1$ に対して, (4.5) が成立すること示す．$X_j\,(j = 1, 2, \ldots)$ は i.i.d. 確率変数列で

$$\mathsf{Pr}(X_j = 1) = x = 1 - \mathsf{Pr}(X_j = 0)$$

をみたすとする．いま

$$\mathsf{E}\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right] = x$$

に注意して, Chebyshev の不等式 (系 3.3) を用いると, 任意の $\epsilon > 0$ に対して

$$\begin{aligned}
\mathsf{Pr}&\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - x\right| \geq \epsilon\right) \\
&\leq \frac{\mathsf{Var}\left(\dfrac{X_1 + X_2 + \cdots + X_n}{n}\right)}{\epsilon^2} \\
&= \frac{\dfrac{1}{n^2}\bigl(\mathsf{Var}(X_1) + \mathsf{Var}(X_2) + \cdots + \mathsf{Var}(X_n)\bigr)}{\epsilon^2} \\
&= \frac{x(1-x)}{n\epsilon^2}
\end{aligned}$$

となる．したがって

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \xrightarrow{P} x$$

がわかる．さらに, 定理 4.7(2) より

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \rightsquigarrow x$$

[2] $f(0) = a,\ f(1) = b$ に対して, $\widetilde{f}(x) = f(x) + \bigl\{f(b) - f(a)\bigr\}x + f(a)$ とおくと $\widetilde{f}$ は $[0, 1]$ 上の連続関数で $\widetilde{f}(0) = \widetilde{f}(1) = 0$ となるので, 左記の設定は一般性を失わない仮定である．

を得る．f は有界連続関数なので, 定理 4.9(2) から

$$\lim_{n\to\infty} \mathsf{E}\left[f\left(\frac{X_1+X_2+\cdots+X_n}{n}\right)\right] = f(x)$$

となる．しかし, $S := X_1+X_2+\cdots+X_n \sim \mathsf{Bino}(n,\, x)$ なので

$$\begin{aligned}
\mathsf{E}\left[f\left(\frac{X_1+X_2+\cdots+X_n}{n}\right)\right] &= \mathsf{E}\left[f\left(\frac{S}{n}\right)\right] \\
&= \sum_{j=0}^{n} f\left(\frac{j}{n}\right)\binom{n}{j} x^j(1-x)^{n-j} \\
&= B_n(x)
\end{aligned}$$

がわかる．よって, 主張は証明された．□

4.3　中心極限定理

定理 4.17. $X_1,\, X_2,\, \ldots$ を i.i.d. 確率変数列とし, $\mathsf{E}[X_1] = \mu$, $\mathsf{Var}[X_1] = \sigma^2\,(0<\sigma<\infty)$ とする．このとき

$$Z_n := \frac{\overline{X}_n - \mu}{\sqrt{\mathsf{Var}\left[\overline{X}_n\right]}} = \frac{\sqrt{n}\left(\overline{X}_n - \mu\right)}{\sigma} \rightsquigarrow Z$$

が成り立つ．ただし $\overline{X}_n = \dfrac{1}{n}\left(X_1+X_2+\cdots+X_n\right)$, $Z \sim \mathsf{N}(0,\, 1)$ である．

Proof. 中心極限定理の証明は第 G 章でする．□

例 4.18. $X_1,\, X_2,\, \ldots$ を i.i.d. 確率変数列とし, $\mathsf{E}\left[X_1\right] = \mu$, $\mathsf{Var}\left[X_1\right] = \sigma^2\,(0<\sigma<\infty)$ とする．このとき中心極限定理 (定理 4.17) より

$$Z_n = \frac{\sqrt{n}\left(\overline{X}_n - \mu\right)}{\sigma} \rightsquigarrow \mathsf{N}(0,\, 1) \tag{4.6}$$

が成立する．ここで

$$S_n^2 := \frac{1}{n-1}\sum_{j=1}^{n}\left(X_j - \overline{X}_n\right)^2, \quad S_n = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}\left(X_j - \overline{X}_n\right)^2}$$

とおく．すると

$$S_n^2 = \frac{n}{n-1}\left\{\frac{1}{n}\sum_{j=1}^{n}\left(X_j-\mu\right)^2 - \left(\overline{X}_n - \mu\right)^2\right\}$$

と書き直せる. 設定から $\mathsf{E}[(X_1-\mu)^2]=\sigma^2<\infty$ なので, 大数の法則 (定理 4.12) より

$$\frac{1}{n}\sum_{j=1}^{n}(X_j-\mu)^2\xrightarrow{\mathrm{P}}\mathsf{E}\left[(X_1-\mu)^2\right]=\sigma^2, \tag{4.7}$$

$$\overline{X}_n\xrightarrow{\mathrm{P}}\mu \tag{4.8}$$

である. (4.8) に対して, 定理 4.11(6)($g(x)=(x-\mu)^2$) を用いると

$$\left(\overline{X}_n-\mu\right)^2\xrightarrow{\mathrm{P}}0 \tag{4.9}$$

となる. さらに, (4.7), (4.9) と定理 4.11(3)(5) より

$$S_n^2\xrightarrow{\mathrm{P}}\sigma^2$$

である. 再度, 定理 4.11(6) を用いると

$$\frac{S_n}{\sigma}\xrightarrow{\mathrm{P}}1 \tag{4.10}$$

がわかる. 最後に (4.6), (4.10) と定理 4.11(5) より

$$\frac{\sqrt{n}\left(\overline{X}_n-\mu\right)}{S_n}\rightsquigarrow\mathsf{N}(0,\,1)$$

を得る.

□

定理 4.19. (Berry-Essèen の不等式) $X_1,\,X_2,\,\ldots$ を i.i.d. 確率変数列とする. $\mathsf{E}\left[|X_1|^3\right]<\infty$ のとき

$$\sup_z\left|\mathsf{Pr}\left(Z_n\le z\right)-\Phi(z)\right|\le\frac{33}{4}\frac{\mathsf{E}\left[|X_1-\mu|^3\right]}{\sqrt{n}\sigma^2}$$

が成立する. ただし $Z_n=\frac{\sqrt{n}\left(\overline{X}_n-\mu\right)}{\sigma}$, $\Phi(z)=\int_{-\infty}^{z}e^{-t^2/2}\,dt\,(z\in\mathbb{R})$ である.

Proof. 信じることにする. □

注意 4.20. 証明についてのコメントを書くこと. □

定理 4.21. $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ を i.i.d. 確率ベクトル列とする．ただし $j = 1, 2, \ldots$ に対して

$$\boldsymbol{X}_j = \begin{pmatrix} X_{1j} \\ X_{2j} \\ \vdots \\ X_{dj} \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_d \end{pmatrix} = \begin{pmatrix} \mathsf{E}[X_{1j}] \\ \mathsf{E}[X_{2j}] \\ \vdots \\ \mathsf{E}[X_{dj}] \end{pmatrix},$$

$$\mathsf{Var}[\boldsymbol{X}_1] = \boldsymbol{\Sigma} = \begin{pmatrix} \mathsf{Var}[X_{11}] & \mathsf{Cov}[X_{11}, X_{21}] & \cdots & \mathsf{Cov}[X_{11}, X_{d1}] \\ \mathsf{Cov}[X_{21}, X_{11}] & \mathsf{Var}[X_{21}] & \cdots & \mathsf{Cov}[X_{21}, X_{d1}] \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{Cov}[X_{d1}, X_{11}] & \mathsf{Cov}[X_{d1}, X_{21}] & \cdots & \mathsf{Var}[X_{d1}] \end{pmatrix}$$

は正定値となる．このとき任意の $\boldsymbol{c} \in \mathbb{R}^d$ に対して

$$\sqrt{n}\boldsymbol{c}^\top\bigl(\overline{\boldsymbol{X}}_n - \boldsymbol{\mu}\bigr) \rightsquigarrow \mathsf{N}\bigl(0, \boldsymbol{c}^\top\boldsymbol{\Sigma}\boldsymbol{c}\bigr)$$

が成立する．このことを

$$\sqrt{n}\bigl(\overline{\boldsymbol{X}}_n - \boldsymbol{\mu}\bigr) \rightsquigarrow \mathsf{N}_d\bigl(\boldsymbol{0}, \boldsymbol{\Sigma}\bigr)$$

と書くことにする．

Proof. 証明は後で行う． □

4.4 デルタ法

$\{Y_n\}$ を確率変数列とする．Y_n の極限分布が正規分布のとき滑らかな実数値関数 $g\colon \mathbb{R} \to \mathbb{R}$ に対して $g(Y_n)$ の極限分布を求めよう．

定理 4.22. (デルタ法) $Y_1, Y_2, \ldots$ を確率変数列とし

$$\frac{\sqrt{n}\bigl(Y_n - \mu\bigr)}{\sigma} \rightsquigarrow \mathsf{N}(0, 1),$$

とし，g は $x = \mu$ の近傍で連続微分可能な関数で $\dot{g}(\mu) \neq 0$ とする．ただし，$\mu \in \mathbb{R}$, $0 < \sigma < \infty$, $\dot{g}(t) = \dfrac{\mathrm{d}g}{\mathrm{d}t}(t)$ である．このとき

$$\frac{\sqrt{n}\bigl(g(Y_n) - g(\mu)\bigr)}{\bigl|\dot{g}(\mu)\bigr|\sigma} \rightsquigarrow \mathsf{N}(0, 1)$$

が成立する．すなわち

$$Y_n \approx \mathsf{N}\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow g(Y_n) \approx \mathsf{N}\left(g(\mu), \bigl(\dot{g}(\mu)\bigr)^2\frac{\sigma^2}{n}\right)$$

である．ただし「≈」は「分布が近似できる」の意味である．

Proof. 節 4.4.5.4 で証明をする． □

例 4.23. $X_1, X_2, \ldots$ を i.i.d. 確率変数列とし，$\mathsf{E}\big[X_1\big] = \mu$, $\mathsf{Var}\big[X_1\big] = \sigma^2\,(0 < \sigma < \infty)$ とする．このとき中心極限定理 (定理 4.17) より

$$\frac{\sqrt{n}\big(\overline{X}_n - \mu\big)}{\sigma} \rightsquigarrow \mathsf{N}(0,\,1)$$

となる．いま

$$W_n = e^{\overline{X}_n}$$

とおく．したがって $g(x) = e^x\,(x \in \mathbb{R})$ とすれば $\dot{g}(x) = e^x$ となる．よってデルタ法 (定理 4.22) より

$$\frac{\sqrt{n}\big(W_n - e^{\mu}\big)}{e^{\mu}\sigma} \rightsquigarrow \mathsf{N}(0,\,1)$$

となる．よって

$$W_n \approx \mathsf{N}\left(e^{\mu},\,\frac{e^{2\mu}\sigma^2}{n}\right)$$

がわかる． □

定理 4.24. (多次元デルタ法) $\boldsymbol{Y}_n = (Y_{1n}, Y_{2n}, \ldots, Y_{dn})^\top\,(n = 1, 2, \ldots)$ を確率ベクトル列とし

$$\sqrt{n}\big(\boldsymbol{Y}_n - \boldsymbol{\mu}\big) \rightsquigarrow \mathsf{N}_d\big(\boldsymbol{0}_d,\,\boldsymbol{\Sigma}\big)$$

とする．ただし $\boldsymbol{\mu} \in \mathbb{R}^d$ で $\boldsymbol{\Sigma}$ は $d \times d$ の正値対称行列とする．関数 $g\colon \mathbb{R}^d \to \mathbb{R}$ は滑らかで

$$\nabla g(\boldsymbol{y}) = \begin{pmatrix} \frac{\partial g}{\partial y_1} \\ \frac{\partial g}{\partial y_2} \\ \vdots \\ \frac{\partial g}{\partial y_d} \end{pmatrix}, \quad \boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_d \end{pmatrix}, \quad \nabla_{\boldsymbol{\mu}} := \nabla g(\boldsymbol{y})\Big|_{\boldsymbol{y}=\boldsymbol{\mu}}$$

とする．このとき

$$\sqrt{n}\big(g\big(\boldsymbol{Y}_n\big) - g\big(\boldsymbol{\mu}\big)\big) \rightsquigarrow \mathsf{N}\big(0,\,\nabla_{\boldsymbol{\mu}}^\top \boldsymbol{\Sigma} \nabla_{\boldsymbol{\mu}}\big)$$

となる．

例 4.25.

$$\begin{pmatrix} X_{11} \\ X_{21} \end{pmatrix}, \begin{pmatrix} X_{12} \\ X_{22} \end{pmatrix}, \ldots, \begin{pmatrix} X_{1n} \\ X_{2n} \end{pmatrix}$$

は i.i.d. 確率ベクトル列で

$$\mathsf{E}\begin{bmatrix} X_{11} \\ X_{21} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \boldsymbol{\mu}, \quad \mathsf{Var}\begin{bmatrix} X_{11} \\ X_{21} \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} = \boldsymbol{\Sigma}$$

とする．ただし $\boldsymbol{\Sigma}$ は 2×2 の正値対称行列である．

$$\overline{X}_1 = \frac{1}{n}\sum_{j=1}^{n} X_{1j}, \quad \overline{X}_2 = \frac{1}{n}\sum_{j=1}^{n} X_{2j}$$

とし

$$Y_n = \overline{X}_1\overline{X}_2, \quad g(s_1,\, s_2) = s_1 s_2\,(s_1,\, s_2 \in \mathbb{R})$$

とおく．中心極限定理 (定理 4.21) より

$$\sqrt{n}\begin{pmatrix} \overline{X}_1 - \mu_1 \\ \overline{X}_2 - \mu_2 \end{pmatrix} \rightsquigarrow \mathsf{N}_2\bigl(\mathbf{0}_2,\, \boldsymbol{\Sigma}\bigr)$$

となる．いま

$$\nabla g(\boldsymbol{s}) = \begin{pmatrix} \dfrac{\partial g}{\partial s_1} \\ \dfrac{\partial g}{\partial s_2} \end{pmatrix} = \begin{pmatrix} s_2 \\ s_1 \end{pmatrix}, \quad \boldsymbol{s} = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}$$

であるので

$$\nabla_{\boldsymbol{\mu}}^{\top}\boldsymbol{\Sigma}\nabla_{\boldsymbol{\mu}} = \mu_2^2\sigma_{11} + 2\mu_1\mu_2\sigma_{12} + \mu_1^2\sigma_{22}$$

である．したがって

$$\sqrt{n}\bigl(\overline{X}_1\overline{X}_2 - \mu_1\mu_2\bigr) \rightsquigarrow \mathsf{N}\bigl(0,\, \mu_2^2\sigma_{11} + 2\mu_1\mu_2\sigma_{12} + \mu_1^2\sigma_{22}\bigr)$$

がわかる． □

4.5 補題 4.13, 定理 4.7, 4.9, 4.11, 4.22 の証明

4.5.1 補題 4.13 の証明

背理法で証明する．そのために

$$\mathsf{Pr}\left(\sum_{k=1}^{\infty} X_k < \infty\right) < 1$$

を仮定する. さらに事象 F と N を

$$F := \left\{\omega \in \Omega;\ \sum_{k=1}^{\infty} X_k(\omega) < \infty\right\}, \qquad N := \left\{\omega \in \Omega;\ \sum_{k=1}^{\infty} X_k(\omega) = \infty\right\}$$

で定める. このとき

$$\Omega = F \cup N \quad \text{かつ} \quad F \cap N = \varnothing$$

が成り立つ. したがって

$$1 = \mathsf{Pr}(\Omega) = \mathsf{Pr}\big(F \cup N\big) = \mathsf{Pr}(F) + \mathsf{Pr}(N) \qquad (\because \mathsf{Pr}\ \text{の加法性})$$

がわかる. この関係式と背理法の仮定 $\mathsf{Pr}(F) < 1$ から $\mathsf{Pr}(N) > 0$ となる. 一方, $\omega \in N$ のとき

$$\sum_{k=1}^{\infty} X_k(\omega) = \infty$$

であり, $\omega \in N^c$ のとき

$$0 \leq \sum_{k=1}^{\infty} X_k(\omega) < \infty$$

である. したがって, 任意の正の実数 $r \geq 0$ に対して

$$r\mathbb{1}_N(\omega) \leq \sum_{k=1}^{\infty} X_k(\omega) \qquad (\forall \omega \in \Omega) \tag{4.11}$$

が成り立つ. (4.11) の両辺の期待値をとる. すると (4.11) と (4.17) から

$$r \times \mathsf{Pr}(N) = \mathsf{E}[r\mathbb{1}_N] \leq \mathsf{E}\left[\sum_{k=1}^{\infty} X_k\right] = \sum_{k=1}^{\infty} \mathsf{E}\big[X_k\big] < \infty \tag{4.12}$$

を得る. (4.12) の最後の等号は単調収束定理からわかる. ここで $\mathsf{Pr}(N) > 0$ であり, $r \geq 0$ は任意の実数だったので, $r \to \infty$ とすれば, (4.12) の最左辺は $+\infty$ となるので, 矛盾が生じる. したがって, $\mathsf{Pr}(N) = 1$ が成り立つ. □

4.5.2　定理 4.7 の証明

(1) $X_n \overset{qm}{\to} X$ とする. $\epsilon > 0$ を固定する. このとき Markov の不等式 (定理 3.1) より

$$\mathsf{Pr}\big(\big|X_n - X\big| \geq \epsilon\big) = \mathsf{Pr}\big(\big|X_n - X\big|^2 \geq \epsilon^2\big) \leq \frac{\mathsf{E}\big[\big|X_n - X\big|^2\big]}{\epsilon^2} \overset{n\to\infty}{\longrightarrow} 0$$

がわかる.
(2) $X_n \xrightarrow{\mathrm{P}} X$ とする. $\epsilon > 0$ を固定し, x を X の c.d.f. $\mathsf{F}(x)$ の連続点とする. このとき

$$\begin{aligned}\mathsf{F}_n(x) &= \mathsf{Pr}\big(X_n \le x\big)\\ &= \mathsf{Pr}\big(X_n \le x,\, X < x+\epsilon\big) + \mathsf{Pr}\big(X_n \le x,\, X \ge x+\epsilon\big)\\ &\le \mathsf{Pr}\big(X \le x+\epsilon\big) + \mathsf{Pr}\big(\big|X_n - X\big| \ge \epsilon\big) \qquad (4.13)\\ &= \mathsf{F}(x+\epsilon) + \mathsf{Pr}\big(\big|X_n - X\big| \ge \epsilon\big) \qquad (4.14)\end{aligned}$$

となる. 同様に

$$\begin{aligned}\mathsf{F}(x-\epsilon) &= \mathsf{Pr}\big(X \le x-\epsilon\big)\\ &= \mathsf{Pr}\big(X_n \le x-\epsilon,\, X_n \le x\big) + \mathsf{Pr}\big(X \le x-\epsilon,\, X_n \ge x\big)\\ &\le \mathsf{Pr}\big(X_n \le x\big) + \mathsf{Pr}\big(\big|X_n - X\big| \ge \epsilon\big) \qquad (4.15)\\ &= \mathsf{F}_n(x) + \mathsf{Pr}\big(\big|X_n - X\big| \ge \epsilon\big) \qquad (4.16)\end{aligned}$$

となる. (4.14) と (4.16) を合わせると

$$\mathsf{F}(x-\epsilon) - \mathsf{Pr}\big(\big|X_n - X\big| \ge \epsilon\big) \le \mathsf{F}_n(x) \le \mathsf{F}(x+\epsilon) + \mathsf{Pr}\big(\big|X_n - X\big| \ge \epsilon\big)$$

がわかる. 上の式の辺々で $n \to \infty$ とすれば

$$\mathsf{F}(x-\epsilon) \le \liminf_{n\to\infty} \mathsf{F}_n(x) \le \limsup_{n\to\infty} \mathsf{F}_n(x) \le \mathsf{F}(x+\epsilon)$$

となる. ここで $\epsilon \to 0$ とし x を $\mathsf{F}(x)$ の連続点とすれば

$$\lim_{n\to\infty} \mathsf{F}_n(x) = \mathsf{F}(x)$$

がわかる.
(3) $X_n \rightsquigarrow X$ かつある定数 c があって $\mathsf{Pr}(X = c) = 1$ とする. $c \pm \epsilon$ を $\mathsf{F}(x)$ の連続点になるようにして $\epsilon > 0$ を固定する[3]. このとき

$$\begin{aligned}\mathsf{Pr}\big(\big|X_n - c\big| \ge \epsilon\big) &= \mathsf{Pr}\big(X_n \le c-\epsilon\big) + \mathsf{Pr}\big(X_n \ge c+\epsilon\big)\\ &= \mathsf{Pr}\big(X_n \le c-\epsilon\big) + \mathsf{Pr}\big(X_n \ge c+\epsilon\big)\\ &= \mathsf{F}_n(c-\epsilon) + 1 - \mathsf{F}_n(c+\epsilon)\\ &\to \mathsf{F}(c-\epsilon) + 1 - \mathsf{F}(c-\epsilon) \quad (n \to \infty)\\ &= 0 + 1 - 1 = 0\end{aligned}$$

となる. □

[3] $\mathsf{F}(x)$ は有界非減少関数なので, $\mathsf{F}(x)$ の不連続点は高々可算個である. よってこのように ϵ を取ることとができる.

問 4.1. (4.13) と (4.15) を確認せよ．すなわち

$$
\begin{aligned}
&\mathsf{Pr}\bigl(X_n \le x,\, X > x+\epsilon\bigr) \le \mathsf{Pr}\bigl(\bigl|X_n - X\bigr| > \epsilon\bigr),\\
&\mathsf{Pr}\bigl(X \le x-\epsilon,\, X_n > x\bigr) \le \mathsf{Pr}\bigl(\bigl|X_n - X\bigr| > \epsilon\bigr)
\end{aligned}
$$

を示せ．

4.5.3　定理 4.9 の証明

(1) ⇒ (2) の証明: g を任意の有界連続関数とする．一般性を失わずに $|g(x)| \le 1/2\,(x \in \mathbb{R})$ と仮定してよい．任意の $\epsilon > 0$ に対して，十分大きな有界閉区間 K をとると

$$
\mathsf{Pr}(X \in K^c) \le \epsilon \tag{4.17}
$$

とできる．そこで，この K を $m\,(m \in \mathbb{N})$ 個の互いに素な区間 $K_j = (a_j,\, b_j]\,(j = 1,\, 2,\, \ldots,\, m); a_j,\, b_j \in \mathbb{R})$ に直和分解する．すなわち，$K = \bigcup_{j=1}^m K_j$ となっている．このとき，m を十分大きく取ると

$$
\max_{j \in \{1,\,2,\,\ldots,\,m\}} \max_{x,\,y \in K_j} |g(x) - g(y)| \le \epsilon \tag{4.18}
$$

となる．さらに，$\{a_j,\, b_j\}_{j=1}^m$ は X の分布関数 F の連続点[4]に取る．このとき，(1) より

$$
\mathsf{Pr}(X_n \in K) \overset{n\to\infty}{\longrightarrow} \mathsf{Pr}(X \in K) \tag{4.19}
$$

となる．つぎに，各 K_j から 1 点 x_j を任意に選び，関数 g_ϵ を

$$
g_\epsilon(x) = \sum_{j=1}^m g(x_j)\mathbb{1}_{K_j}(x) \qquad (x \in \mathbb{R})
$$

と定める．すると

$$
\max_{x \in K} \bigl|g(x) - g_\epsilon(x)\bigr| \le \epsilon \tag{4.20}
$$

[4] F の不連続点は高々可算個なので，このような取れることがわかる．

となる．したがって, (4.17) と (4.20) に注意すると

$$
\begin{aligned}
&\left|\mathsf{E}[g(X)] - \mathsf{E}[g_\epsilon(X)]\right| \\
&\quad = \Big|\mathsf{E}[\mathbb{1}_K(X)g(X)] + \mathsf{E}[\mathbb{1}_{K^c}(X)g(X)] - \mathsf{E}[\mathbb{1}_K(X)g_\epsilon(X)] \\
&\qquad\quad - \mathsf{E}[\mathbb{1}_{K^c}(X)g_\epsilon(X)]\Big| \\
&\quad \le \left|\mathsf{E}[\mathbb{1}_K(X)g(X)] - \mathsf{E}[\mathbb{1}_K(X)g_\epsilon(X)]\right| + \Big|\mathsf{E}[\mathbb{1}_{K^c}(X)g(X)] \\
&\qquad\quad - \mathsf{E}[\mathbb{1}_{K^c}(X)g_\epsilon(X)]\Big| \\
&\quad \le \mathsf{E}\big[\mathbb{1}_K(X)\left|g(X) - g_\epsilon(X)\right|\big] + \mathsf{E}\big[\mathbb{1}_{K^c}(X)\left|g(X) - g_\epsilon(X)\right|\big] \\
&\quad \le \underbrace{\max_{x \in K}\left|g(x) - g_\epsilon(x)\right|}_{\le \epsilon \quad \because (4.20)} \mathsf{E}\big[\mathbb{1}_K(X)\big] + \mathsf{E}\big[\mathbb{1}_{K^c}(X)\left|g(X) - g_\epsilon(X)\right|\big] \\
&\quad \le \epsilon \mathsf{Pr}(X \in K) + \mathsf{Pr}(X \in K^c) \\
&\quad \le 2\epsilon
\end{aligned}
\tag{4.21}
$$

となる．(4.19) から, n を十分大きく取ると

$$
\Big|\underbrace{\mathsf{Pr}(X_n \in K^c)}_{=1-\mathsf{Pr}(X_n \in K)} - \underbrace{\mathsf{Pr}(X \in K^c)}_{=1-\mathsf{Pr}(X \in K)}\Big| = \left|\mathsf{Pr}(X \in K) - \mathsf{Pr}(X \in K)\right| \le \epsilon
$$

とできるので

$$
\mathsf{Pr}(X_n \in K^c) \le \mathsf{Pr}(X \in K^c) + \left|\mathsf{Pr}(X_n \in K^c) - \mathsf{Pr}(X \in K^c)\right| \le 2\epsilon
$$

がわかる．同様に

$$
\left|\mathsf{E}[g(X_n)] - \mathsf{E}[g_\epsilon(X_n)]\right| \le \epsilon \mathsf{Pr}(X_n \in K) + \mathsf{Pr}(X_n \in K^c) \le 3\epsilon \tag{4.22}
$$

を得る．g_ϵ の作りかたから n を十分大きくとると

$$
\begin{aligned}
&\left|\mathsf{E}[g_\epsilon(X)] - \mathsf{E}[g_\epsilon(X_n)\right| \\
&\qquad \le \sum_{j=1}^{n} \left|\mathsf{Pr}(X \in K_j) - \mathsf{Pr}(X_n \in K_j)\right| \max_{j=\{1,2,\dots,m\}} |g(x_j)| \\
&\qquad \le \frac{\epsilon}{2} < \epsilon
\end{aligned}
\tag{4.23}
$$

となる．よって, (4.21) – (4.23)

$$
\begin{aligned}
\left|\mathsf{E}[g(X_n)] - \mathsf{E}[g(X)]\right| &\le \Big|\underbrace{\mathsf{E}[g(X_n)] - \mathsf{E}[g_\epsilon(X_n)]\Big|}_{\le 2\epsilon} + \underbrace{\left|\mathsf{E}[g_\epsilon(X_n)] - \mathsf{E}[g_\epsilon(X)]\right|}_{\le \epsilon} \\
&\quad + \underbrace{\left|\mathsf{E}[g_\epsilon(X)] - \mathsf{E}[g(X)]\right|}_{\le 3\epsilon} \\
&\le 6\epsilon
\end{aligned}
$$

がわかるので, 主張は証明された.
(2) $\Rightarrow$ (3) の証明: 明らか.
(2) $\Rightarrow$ (4) の証明: 非負値連続関数 g と $M>0$ に対して, $g_M(x)=\min\{g(x),\,M\}\geq 0\,(x\in\mathbb{R})$ と定める g_M は有界連続となる. よって

$$\liminf_{n\to\infty}\mathsf{E}[g(X_n)]\leq\liminf_{n\to\infty}\mathsf{E}[g_M(X_n)]=\mathsf{E}[g_M(X)]\qquad\big(\because (2)\big)$$

となる. 両辺で $M\to\infty$ とすることにより, 有界単調収束定理から (4) を得る.
(4) $\Rightarrow$ (2) の証明: g を有界連続関数とする. するとある $M>0$ が存在して, $|g|\leq M$ とできる. すると $M\pm g\geq 0$ は非負値連続関数となるので, (4) を用いると (2) を得る.
(3) $\Rightarrow$ (5) の証明: 開集合 O と $x\in\mathbb{R}$ に対して $g_M(x)\uparrow \mathbb{1}_O(x)\,(M\to\infty)$ となるような Lipschitz 連続関数列 $g_M\geq 0$ を取ることができるので, (3) から

$$\liminf_{n\to\infty}\mathsf{Pr}(X_n\in)\geq\liminf_{n\to\infty}\mathsf{E}[g_M(X_n)]=\mathsf{E}[g_M(X)]$$

となる. この両辺で $M\to\infty$ とすると, 単調収束定理により (5) を得る.
(5) $\Leftrightarrow$ (6) の証明: お互いの補集合をとればよい.
(5) + (6) $\Rightarrow$ (7) の証明: B° を B の内部, $\mathrm{cl}(B)$ を B の閉包とすると $\partial B=\mathrm{cl}(B)\setminus B^\circ$ である. このことに注意すると

$$\begin{aligned}\mathsf{Pr}\big(X\in B^\circ\big)&\leq\liminf_{n\to\infty}\mathsf{Pr}\big(X_n\in B^\circ\big)\leq\limsup_{n\to\infty}\mathsf{Pr}\big(X_n\in\mathrm{cl}(B)\big)\\&\leq\mathsf{Pr}\big(X_n\in\mathrm{cl}(B)\big)\end{aligned}$$

となる. $\mathsf{Pr}(X\in\partial B)=0$ より, 上式の右辺と左辺は等しいので, (7) が得られる.
(7) $\Rightarrow$ (1) の証明: $C=(-\infty,\,x]$ ととればよい. □

4.5.4　定理 4.11 の証明

(6) の証明: 関数 g は $x_0\,(\in\mathbb{R})$ で連続なので, 任意の $\epsilon>0$ に対して, 存在する $\exists\delta>0$ があって

$$\forall x\in\mathbb{R}:\,|x-x_0|<\delta\;\Rightarrow\;|g(x)-g(x_0)|<\epsilon$$

である. よって

$$\{\omega\in\Omega:\,|X(\omega)-X_n(\omega)|<\delta\}\subset\{\omega\in\Omega:\,|g\big(X(\omega)\big)-g\big(X_n(\omega)\big)|<\delta\}$$

となる．上式の補事象を取り，補題 1.7(4) の後半の主張を適用すると

$$\mathsf{Pr}\big(|X - X_n| \geq \delta\big) \geq \mathsf{Pr}\big(|g(X) - g(X_n)| \geq \epsilon)\big)$$

がわかる．よって，$g(X_n) \xrightarrow{P} g(X)$ が示せた．
(7) の証明: 有界連続関数 f を取る．$g \circ f$ も有界連続関数なので，$X_n \rightsquigarrow X$ と定理 4.9 から

$$\lim_{n\to\infty} \mathsf{E}[f\big(g(X_n)\big)] = \lim_{n\to\infty} \mathsf{E}[(f \circ g)(X_n)] = \mathsf{E}[(f \circ g)(X)] = \mathsf{E}[f\big(g(X)\big)]$$

から (7) は示される．
(1) の証明: $X_n \xrightarrow{P} X$ かつ $Y_n \xrightarrow{P} Y$ から $(X_n, Y_n) \xrightarrow{P} (X, Y)$ がわかるので，(6) の関数を $g(x, y) = x + y$ ととればよい．
(2) の証明: 三角不等式を使えばよい．
(3) の証明: $Y_n \xrightarrow{P} c$ なので，注意 4.2(2) から $Y_n \rightsquigarrow c$ となる．このことから $(X_n, Y_n) \rightsquigarrow (X, Y)$ となることが証明[5]できる．あとは (7) において，$g(x, y) = x + y$ とすればよい．
(4) の証明: (6) を用いればよい．
(5) の証明: (7) を用いればよい．□

4.5.5　定理 4.22 の証明

$g(x)$ は $x = \mu$ の近傍で連続微分可能である．このことからある $\delta_1 > 0$ が存在して $|x - \mu| < \delta_1$ なる任意の x に対して

$$g(x) - g(\mu) = (x - \mu) \int_0^1 \dot{g}\big(\mu + t(x - \mu)\big)\, \mathrm{d}t \tag{4.24}$$

が成立する．また $\dot{g}(x)$ は $x = \mu$ で連続なので，任意の $\epsilon > 0$ に対してある $\delta_2 > 0$ があって

$$|x - \mu| < \delta_2 \;\Rightarrow\; \big|\dot{g}(x) - \dot{g}(\mu)\big| < \epsilon$$

となる．よって $|x - \mu| < \min(\delta_1, \delta_2) =: \delta$ なる任意の x に対して

$$\begin{aligned}\left| \int_0^1 \dot{g}\big(\mu + t(x - \mu)\big)\, dt - \dot{g}(\mu) \right| &\leq \int_0^1 \big|\dot{g}\big(\mu + t(x - \mu)\big) - \dot{g}(\mu)\big|\, dt \\ &\leq \epsilon \int_0^1 dt = \epsilon \end{aligned} \tag{4.25}$$

[5] これは特性関数の収束と分布収束が同値である事実からわかる．あとは特性関数に対して三角不等式を用いればよい．詳しくは [30, p.135] を参照のこと．

となる．(4.25) より

$$\mathsf{Pr}\left(\left|\int_0^1 \dot{g}\big(\mu + t(Y_n-\mu)\big)\,dt - \dot{g}(\mu)\right| < \epsilon\right) > \mathsf{Pr}\big(|Y_n-\mu| < \delta\big) \tag{4.26}$$

が成立する．$\sqrt{n}\big(Y_n-\mu\big)/\sigma \rightsquigarrow \mathsf{N}(0,\,1)$ なので $Y_n \xrightarrow{\mathrm{P}} \mu$ である．(4.26) から

$$\int_0^1 \dot{g}\big(\mu + t(Y_n-\mu)\big)\,\mathrm{d}t \xrightarrow{\mathrm{P}} \dot{g}(\mu) \tag{4.27}$$

が成立する．(4.24) に $x = Y_n$ を代入すれば

$$\frac{\sqrt{n}\big(g(Y_n)-g(\mu)\big)}{\sigma} = \frac{\sqrt{n}\big(Y_n-\mu\big)}{\sigma}\int_0^1 \dot{g}\big(\mu + t(Y_n-\mu)\big)\,\mathrm{d}t \tag{4.28}$$

を得る． さらに，$\sqrt{n}(Y_n-\mu)/\sigma \rightsquigarrow \mathsf{N}(0,\,1)$ と (4.27) に注意して定理 4.11(5) を (4.28) に適用すれば

$$\frac{\sqrt{n}\big(g(Y_n)-g(\mu)\big)}{\sigma} \rightsquigarrow \dot{g}(\mu)\mathsf{N}(0,\,1)$$

を得る．よって定理は示された． □

4.6　章末注釈と参考文献

第 4.1 節 は [14, 30] を参考にした．注意 4.4 は [14] からの借用である．第 4.2 節 は [26] を借用した．第 4.4 節 は [14, 30] を借用した．定理 4.22 の証明は [8] からの借用である．

4.7　演習問題

演習問題 4.1. $n \in \mathbb{N}$ とする．離散型確率変数 X_n は $\left\{\frac{1}{n},\,\frac{2}{n},\,\ldots,\,1\right\}$ 上の一様分布に従うとする．すなわち

$$\mathsf{Pr}\left(X_n = \frac{k}{n}\right) = \begin{cases} \dfrac{1}{n} & (k = 1,\,2,\,\ldots,\,n) \\ 0 & (\text{その他の場合}) \end{cases}$$

である．連続型確率変数 X は $(0,\,1)$ 上の一様分布に従うとする．すなわち，X は p.d.f.

$$\mathsf{p}(x) = \begin{cases} 1 & (0 < x < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

をもつ．このとき，以下の問いに答えよ．
(1) $\dfrac{k}{n} \le x < \dfrac{(k+1)}{n}$ $(k = 1, \ldots, n)$ に対して

$$\mathsf{Pr}(X_n \le x)$$

を求めることで，$x \in \mathbb{R}$ に対して $\mathsf{Pr}(X_n \le x)$ を求めよ．
(2) $x \in \mathbb{R}$ に対して

$$\mathsf{Pr}(X \le x)$$

を求めよ．
(3)

$$\lim_{n\to\infty} \big|\mathsf{Pr}(X_n \le x) - \mathsf{Pr}(X \le x)\big|$$

を評価することで $X_n \rightsquigarrow X$ を示せ．

演習問題 4.2. $\{X_n\}_{n=1}^{\infty}$ を確率変数列とし，$\mathsf{E}[X_n^2] < \infty\,(n \in \mathbb{N})$ とする．
(1) $\mu \in \mathtt{R}$ を定数とする．

$$\mathsf{E}[(X_n - \mu)^2] = \mathsf{Var}[X_n] + \big\{\mathsf{E}[X_n] - \mu\big\}^2$$

が成り立つことを示せ．
(2) 次の条件は同値であることを示せ．

(a) $\lim_{n\to\infty} \mathsf{E}[(X_n - \mu)^2] = 0.$

(b) $\lim_{n\to\infty} \mathsf{E}[X_n] = \mu$ かつ $\lim_{n\to\infty} \mathsf{Var}[X_n] = 0$

演習問題 4.3. $\{X_n\}_{n=1}^{\infty}$, $\{Y_n\}_{n=1}^{\infty}$, X, Y を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ で定義された確率変数列とする．確率ベクトル (X_n, Y_n) が確率ベクトル (X, Y) に確率収束するとは，任意の $\epsilon > 0$ に対して

$$\lim_{n\to\infty} \mathsf{Pr}\Big(\sqrt{(X_n - X)^2 + (Y_n - Y)^2} \ge \epsilon\Big) = 0$$

が成り立つこととする．これを $(X_n, Y_n) \xrightarrow{\mathrm{P}} (X, Y)$ と記すことにする．
このとき，以下の問いを答えよ．
(1) $X_n \xrightarrow{\mathrm{P}} X$ かつ $Y_n \xrightarrow{\mathrm{P}} Y$ ならば，$(X_n, Y_n) \xrightarrow{\mathrm{P}} (X, Y)$ が成り立つことを示せ．
(2) $(X_n, Y_n) \xrightarrow{\mathrm{P}} (X, Y)$ ならば，$X_n \xrightarrow{\mathrm{P}} X$ かつ $Y_n \xrightarrow{\mathrm{P}} Y$ が成り立つことを示せ．

第II部

宴編: 統計的推測

第5章　統計的推測論の枠組み

この章では統計的推測の枠組みについて述べる．第 5.1 節では, 統計的推測理論で重要な役割を果たす統計的モデルを説明する．5.2 節では, 統計的推測理論を統一的な枠組みで議論する統計的決定理論の簡単な説明をする．

5.1　統計的実験と母数モデル

$\mathbb{X}$ は距離空間とし, $\mathcal{B}(\mathbb{X})$ は $\mathbb{X}$ 上の Borel 集合族とする．ただし $\mathcal{B}(\mathbb{X})$ は $\mathbb{X}$ 上の Borel 集合族[1]である．すなわち, $\big(\mathbb{X}, \mathcal{B}(\mathbb{X})\big)$ は可測空間である．

いま, $X_1, X_2, \ldots, X_n$ は確率空間 $\big(\Omega, \mathcal{A}, \mathsf{Pr}\big)$ 上の $\mathbb{X}$ 値確率要素[2]とする．さらに, $\mathbf{X} = (X_1, X_2, \ldots, X_n)^\top$ と書くことする．典型的な例は $\mathbb{X} = \mathbb{R}^d\,(d \in \mathbb{N})$ 等である．

この講義では $X_1, X_2, \ldots, X_n$ は独立同一の分布に従うものとする．このような場合, $X_1, X_2, \ldots, X_n$ を**ランダム標本**と呼ぶことにする．また, $\mathbb{X}^n = \underbrace{\mathbb{X} \times \mathbb{X} \times \cdots \times \mathbb{X}}_{n\text{ 個}}$ を**標本空間**と呼ぶことにする．

可測空間 $\big(\mathbb{X}, \mathcal{B}(\mathbb{X})\big)$ 上の確率測度 P を

$$\mathsf{P}(B) := \mathsf{Pr}(X_1 \in B) \qquad (B \in \mathcal{B}(\mathbb{X}))$$

で定義する．確率測度 P を X_1 の**確率分布** (簡単に分布ともいうことがある) といい, $X_1 \sim \mathsf{P}$ と表記することにする．さらに, $\boldsymbol{X}$ の分布を $\mathsf{P}^{\otimes n}$ または $\mathsf{P}^{\mathbf{X}}$ と書くことにする．したがって, $B_j \in \mathcal{B}(\mathbb{X})\,(\,j = 1, 2, \ldots, n)$ に対して

$$\mathsf{P}^{\otimes n}(B_1 \times B_2 \times \cdots \times B_n) = \mathsf{Pr}\big(\mathbf{X} \in B_1 \times B_2 \times \cdots \times B_n\big)$$

となる[3]．

[1]開集合族を含む最小の σ 加法族

[2]確率変数, 確率ベクトル, 確率行列の総称を確率要素という．

[3]この記法 $\mathsf{P}^{\otimes n}$ にも測度論的な背景がある．すなわち, $\mathsf{P}^{\otimes n}$ は P の n 個の測度の直接測度になっている．

注意 5.1. $B \in \mathcal{B}(\mathbb{X})$ に対して, $\mathsf{Pr}(X_1 \in B)$ と書いたとき

$$\mathsf{Pr}(X_1 \in B) = \mathsf{Pr}\bigl(\{\omega \in \Omega;\, X_1(\omega) \in B\}\bigr)$$

の意味である． □

確率論的アプローチでは観測データを生成するメカニズムを表現する確率測度 P は既知であり, 確率要素 $\boldsymbol{X}$ の分布論的な特徴を調べる．一方統計的推測のアプローチでは観測データを生成するメカニズムを表現する確率測度 P は観測者には未知であり, 観測データ $\boldsymbol{X}$ に基づいて未知の確率測度 P を回復することを目指す．統計的推測は確率論的アプローチの逆問題と言える．

統計的推測の考え方は統計的実験という概念を基礎に組み立てられる．$X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{P}^*$ とする．ここで, P^* はデータを生成する**真の確率分布**とした．すなわち, $\boldsymbol{X} \sim \bigl(\mathsf{P}^*\bigr)^{\otimes n}$ である．統計的推測では観測データを生成するメカニズムを表現する未知の確率測度 P^* を観測データから回復することを目的とする．$\bigl(\mathbb{X}, \mathcal{B}(\mathbb{X})\bigr)$ 上の確率分布全体から真の分布を探すのは一般に多くの困難を伴うので, 候補となる確率測度の適当な集まりを設定する．この確率分布族の特定の集まりを $\mathcal{P}$ と書いたとき, この $\mathcal{P}$ を**統計的モデル**という．そして真の確率測度 P^* は設定した統計的モデル $\mathcal{P}$ に含まれるとこの講義では仮定[4]する．

さらに統計的モデルの各要素はある集合 Θ の要素 θ で添え字付けれると仮定する．すなわち, 添え字集合全体と統計的モデルとの対応である写像

$$\Theta \ni \theta \mapsto \mathsf{P}_\theta \in \mathcal{P}$$

を想定[5]する．これを統計的モデルの**母数化**といい, Θ を**母数空間**, その要素 θ を**母数**という．そして真の確率測度 P^* に対応するある母数 $\theta^* \in \Theta$ が存在[6]すると仮定する．この θ^* を**真の母数**ということにする．すなわち,

$$\mathsf{P}^* = \mathsf{P}_{\theta^*}$$

である．もちろん $\theta^* \in \Theta$ ではあるが, 真の母数 θ^* は観測者には未知である．したがって, 統計的推測では未知の真の母数 θ^* を観測データ $\boldsymbol{X}$ から回復することが目的[7]となる．

[4]もちろん, 真の確率分布 P^* が想定した統計的モデルに含まれないことを想定した議論もある．しかし, 統計的モデルに真の分布が含まれない場合には, より高度な議論が必要となるので, この講義録ではより基本的な設定を考えている．

[5]この写像がどのような性質を持つかを数学的に定義することはできる．しかし, 確率分布の集まりの空間に位相をいれることになり, 数学的に議論が高度になるので, 漠然とした写像と理解しておくおくことにしよう．

[6]あとで, 一意性も仮定することになる．

[7]統計的機械学習では, 真の母数 θ^* の回復を目標とするより, 真の確率測度と同等のメカニズムから生成される未来の観測 $\widetilde{X}_{n+1}$ を観測データ $\boldsymbol{X}$ に基づいて予測することを目標にしている．統計的機械学習では, 統計的モデルの役割が相対的に低くなる．

次に統計的モデルのなかでも最も基本的な母数モデルを定義する.

定義 5.2. 統計的モデル $\mathcal{P} = \{\mathsf{P}_\theta; \theta \in \Theta\}$ が**正則母数モデル** であるとは, 次の条件をみたすときをいう.

(1) 母数空間 Θ は 有限次元 Euclid 空間 $\mathbb{R}^d$ の「よい」部分集合である. ただし $d \in \mathbb{N}$ である.

(2) 写像
$$\Theta \ni \theta \mapsto \mathsf{P}_\theta \in \mathcal{P}$$
は「滑らか」である[8].

条件 (2) を母数化の正則性という. 正則母数モデルを簡単に母数モデルということもある.

さらにこの講義では次の条件も仮定する.

(3) $\forall \theta_1, \theta_2 \in \Theta$ に対して
$$\theta_1 \neq \theta_2 \Rightarrow \mathsf{P}_{\theta_1} \neq \mathsf{P}_{\theta_2}$$
をみたす.

すなわち母数化を与える写像は単射である. このような母数化を**識別可能**であるという.

注意 5.3. 定義 5.2 は数学的にはすぼらな表現である. 数学的により厳密な母数モデルの定義については Bickel *et al.* (1993, pp.11-13) を参照のこと.

注意 5.4. 母数空間 Θ が有限次元ではないような統計的モデルを考えることも重要である. Θ が無限次元の統計的モデルのことを「ノンパラメトリック・モデル」と統計学では慣例的に呼んでいることが多い. 母数空間が無限次元とはいえ, 統計的モデルは母数化されているので, 「非母数モデル」と呼ぶのは奇異である. しかし歴史的にこの用語が使用されてきたので, この講義録では統計学の歴史的慣例に従うことにする. この言葉使いを嫌う数理統計学者は「ノンパラメトリック・モデル」のことを「無限次元統計的モデル」と呼んでいる. さすがに「無限次元母数モデル」とは言わないようである. さらに母数空間が有限次元の母数空間と無限次元の母数空間の直積で表現され, 有限次元の母数を回復の対象とするような統計的モデルを「セミパラメトリック・モデル[9]」という.

[8] $\mathcal{P}$ は測度の集合なので位相をどのようにいれるかはすこし難しい議論になる. $\mathcal{P}$ が Radon 測度の集まりならば, weak-star 位相を入れることができる. この議論は Tojo and Yoshino (2021) を参照のこと.

[9] 英語読みをすれば, 「セマイパラメトリックモデル」という. 「semi-parametric model」の最初の「i」は長母音であることに注意が必要である.

これも言葉の意味のしては奇異であるが，統計学の習慣に従うことにする．生存データ解析で広く使用される Cox の比例ハザード・モデルはセミパラメトリック・モデルの最高傑作であろう．20 世紀の数理統計学の到達点のひとつであるセミパラメトリック・モデルに対する統計的推測理論については Bickel *et al.* (1993), van der Vaart and Wellner (1996), van der Vaart (1998) , Kosorok (2007), 久保木・鈴木 (2015) を参照のこと． □

定義 5.5. (1) 可測空間 $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ と母数モデル

$$\mathcal{P} = \{\mathsf{P}_\theta;\, \theta \in \Theta,\, \mathsf{P}_\theta \text{ は } (\mathbb{X}, \mathcal{B}(\mathbb{X})) \text{ 上の確率測度}\}$$

の組を**統計的実験**といい

$$\big(\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n), \{\mathsf{P}_\theta;\, \theta \in \Theta\}\big)$$

と書く．
(2) 観測データを生成するメカニズムを表現する真の確率測度 P^* に対応する母数を $\theta^* \in \Theta$ を書くことにする．θ^* を真の母数である．すなわち

$$\mathsf{P}^* = \mathsf{P}_{\theta^*}$$

である．

この講義では $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ と書いたとき，$X_1, X_2, \ldots, X_n$ は独立に同一分布 P^* に従うことを仮定する．上記の仮定をおいた観測データ $\boldsymbol{X}$ のことを**標本の大きさが** n **のランダム標本**という．

以上の議論から，この講義で扱う統計的実験をまとめると下記のようにある．

この講義で仮定する統計的実験

(1) 観測データを $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top$ と書き，可測空間 $\bigl(\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n)\bigr)$ に値をとる．

(2) 観測データは真の分布 P^* からの標本である．すなわち

$$\boldsymbol{X} \sim \bigl(\mathsf{P}^*\bigr)^{\otimes n}$$

である．ただし P^* は $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ 上の確率測度である．

(3) 統計的モデル $\mathcal{P} = \bigl\{\mathsf{P}_\theta; \theta \in \Theta\bigr\}$ を設定する．ただし P_θ も P^* と同じ可測空間 $\bigl(\mathbb{X}, \mathcal{B}(\mathbb{X})\bigr)$ 上の確率測度である．

(4) 母数空間 Θ は Euclid 空間 $\mathbb{R}^d$ の「よい」部分集合である．ただし $d \in \mathbb{N}$ である．

(5) $\Theta \ni \theta$ から $\mathsf{P}_\theta \in \mathcal{P}$ への写像は「滑らか」かつ単射 (母数化の識別可能性を仮定).

(6) ある $\theta^* \in \Theta$ が唯一あって $\mathsf{P}^* = \mathsf{P}_{\theta^*}$ と書ける．

これらをまとめて統計的実験といい

$$\bigl(\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n), \{\mathsf{P}_\theta; \theta \in \Theta\}\bigr)$$

と書き，$\mathcal{P}$ を正則母数モデルという．以後は単に母数モデルということにする．そして統計的推測の目標は真の分布 P^* からのランダム標本 $(X_1, X_2, \ldots, X_n)$ に基づき θ^* を回復することである．

例 5.6. (1) $X_1, X_2, \ldots, X_n$ は正規分布 $\mathsf{N}(\mu, \sigma^2)$ から標本の大きさが n のランダム標本とする．ただし $\mu, \sigma\,(0 < \sigma < \infty)$ が共に未知とする．このとき統計的実験

$$\left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \left\{\mathsf{p}(x|\,\theta) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) (-\infty < x < \infty); \right.\right.$$
$$\left.\left. \theta := (\mu, \sigma) \in \Theta := \mathbb{R} \times (0, \infty)\right\}\right)$$

を想定していることになる．統計的モデルは分布が特定できる表現でよいので，この場合には p.d.f. で表現している．

(2) $X_1, X_2, \ldots, X_n$ は Bernoulli 分布 $\mathsf{Ber}(\theta)\,(0 < \theta < 1)$ から標本の大きさが n のランダム標本とする．ただし θ が未知のときには，統計的実験

$$\left(\{0, 1\}^n, 2^{\{0,1\}^n}, \left\{\mathsf{p}(\boldsymbol{x}|\,\theta) = \theta^x(1-\theta)^{1-x}\,(x = 0, 1); \theta \in \Theta = (0, 1)\right\}\right)$$

を想定していることになる． □

注意 5.7. $X_1, X_2, \dots, X_n$ は分布 P から標本の大きさが n のランダム標本といったときには，

$$\left(\mathbb{R}^n,\, \mathcal{B}(\mathbb{R}^n),\, \{\mathsf{P}_\theta;\, \theta \in \Theta\}\right)$$

のような統計的実験を仮定し，ある $\theta^* \in \Theta$ があって，$\mathsf{P}_{\theta^*} = \mathsf{P}^*$ であることを想定している．ことさらに統計的実験という用語を今後は用いないことにする． □

5.2 統計的決定問題

統計的推測論には多くのアプローチがある．その中で代表的なアプローチが二つある．一つは頻度論的なもので，もう一方はベイズ論的のものである．以下では頻度論的推測論の枠組みを説明することにする．Bayes 論的推測論は第 9 章で説明する．以下では，頻度論的推測論の枠組みを統計的決定理論[10]の言葉を使って説明する．

標本空間を $\mathbb{X}^n$ とし，観測データを $\boldsymbol{X}$ とする．

(1) まず観測データに基づき行う行動のすべてを集めた集合を**行動空間**といい $\mathbb{A}$ で記す．この講義では $\mathbb{A} = \mathbb{R}$ や $\mathbb{A} = \{0, 1\}$ などである．観測者が観測データに基づき行動 $\mathbb{A}$ の要素を選択するルールを決定関数といい

$$d : \mathbb{X} \ni \boldsymbol{x} \mapsto d(\boldsymbol{x})$$

で記す．行動関数の集まりを行動空間といい，$\mathbb{D}$ と記す．したがって観測者は合理的な行動 d が存在すればありがたいわけである．

(2) 次に行動を評価するための道具として直積空間 $\mathbb{A} \times \Theta$ 上の非負値実数値関数

$$\mathsf{L} : \mathbb{A} \times \Theta \mapsto [0, \infty) \cup \{\infty\}$$

を用意する[11]．この関数を**損失関数**といい，$\mathsf{L}(a, \theta)$ の値が小さいほど望ましい行動であるとする．決定関数の「よさ」を評価するには観測データの実現値 $\boldsymbol{X} = \boldsymbol{x}$ と真の母数 θ^* における損失関数の値 $\mathsf{L}\bigl(d(\boldsymbol{x}), \theta^*\bigr)$ がわかればよい．したがって決定関数 d の「よさ」の

[10]統計的決定理論はゲーム論の概念を借用して，統計的推測論の枠組みと最適理論を定式化 (言語化) したものである．

[11]ただし，区間推定のばあいには，$\mathsf{L} : \mathbb{A} \times \Theta \mapsto [-1, \infty) \cup \{\infty\}$ とすることもある．

評価に $\mathsf{L}\big(d(\boldsymbol{X}),\,\theta^*\big)$ を使えばよいのだが, これは用いることができない. これはランダムな量であり, 未知の母数 θ^* がわからないと知ることができない量であるからである. そこで母数 θ に対して観測データが $\mathsf{P}_\theta^{\otimes n}$ によって生成されたと仮定し, 損失関数 $\mathsf{L}\big(d(\boldsymbol{X}),\,\theta\big)$ を $\mathsf{P}_\theta^{\otimes n}$ に関して期待値を取ったもの

$$\mathsf{R}(d,\,\theta) := \mathsf{E}_\theta\big[\mathsf{L}\big(d(\boldsymbol{X}),\,\theta\big)\big]$$

を考える. ただし, $\mathsf{E}_\theta[\,\cdot\,]$ は $\boldsymbol{X} \sim \mathsf{P}_\theta^{\otimes n}\,(\theta \in \Theta)$ のもとでの期待値である. これを決定関数 d の母数 θ に対する**危険関数**という.

(3) 決定関数 d の「よさ」の評価は真の母数 θ^* のもとで行たいところである. しかしこれは未知である. 危険関数の $\theta \in \Theta$ に関するなんらかの一様な評価が必要になってくる. このことから危険関数の母数空間に関する一様な評価が統計的推測論の深みと困難の淵源である. またこれが統計的推測論のわかりにくさの原因でもあろう. 前節で説明した統計的実験に行動空間, 決定空間, そして損失関数を加えた組

$$\big(\mathbb{X}^n,\,\{\mathsf{P}_\theta;\,\theta \in \Theta\},\,\mathbb{A},\,\mathbb{D},\,\mathsf{L}\big)$$

を統計的決定問題[12]という.

(4) 決定空間 $\mathbb{D}$ は $\mathbb{X}^n$ から $\mathbb{A}$ への可測関数全体とすることもできる. しかし目標は危険関数の Θ に関する一様な評価であるので, $\mathbb{D}$ には可測性以外の合理的な制限[13]を設けるのが一般的である.

点推定問題

真の母数 θ^* を観測データ $\boldsymbol{X}$ に基づいて 1 点で回復するのが点推定である. したがって $\mathbb{A} = \Theta$ となる. 点推定の場合には決定関数 $d(\boldsymbol{X})$ を**推定量**といい, 観測データの実現値 $\boldsymbol{X} = \boldsymbol{x}$ における推定量の値 $d(\boldsymbol{x})$ を**推定値**という. $\Theta = \mathbb{R}$ ならば損失関数として

$$\mathsf{L}\big(a,\,\theta\big) = \big(a-\theta\big)^2, \quad \mathsf{L}\big(a,\,\theta\big) = \big|a-\theta\big|$$

を取るのが代表なアプローチである. 上記の損失関数それぞれに対応する危険関数

$$\mathsf{R}(d,\,\theta) = \mathsf{E}_\theta\big[\mathsf{L}\big(d(\boldsymbol{X}),\,\theta\big)\big]$$

[12]本来であれば, どこで可測であるかを考える必要があるので,

$$\Big(\big(\mathbb{X}^n,\mathcal{B}(\mathbb{X}^n)\big),\,\{\mathsf{P}_\theta;\,\theta \in \Theta\},\,\big(\mathbb{A},\,\mathcal{B}(\mathbb{A})\big),\,\big(\mathbb{D},\,\mathcal{B}(\mathbb{D})\big),\,\mathsf{L}\Big)$$

と書くべきである.

[13]合理的な制限の概念として不変性や不偏性などがある. また尤度に基づく方法に限定するといった考え方もある.

を**平均 2 乗誤差** と **平均絶対誤差**という．したがって平均 2 乗誤差を Θ に関してなんらかの意味で一様に評価することで考えている推定量の族 $\mathbb{D}$ の中から「最適」な推定量ないしは合理的な観点から正当化される推定量をみつけたいわけである．

検定問題

母数空間を 2 つの排反な部分集合に分ける．すなわち

$$\Theta = \Theta_0 \cup \Theta_1, \quad \Theta_0 \cap \Theta_1 = \varnothing.$$

行動は真の母数 θ^* が Θ_0 に属するか, Θ_1 の属するかを判断する．したがって行動空間は $\mathbb{A} = \{0, 1\}$ と書ける．決定関数は標本空間 $\mathbb{X}^n$ の部分集合 W に対して

$$d(\boldsymbol{x}) = \begin{cases} 1 & (\boldsymbol{x} \in W) \\ 0 & (\boldsymbol{x} \notin W) \end{cases}$$

で定めること[14]ができる．検定問題では d のことを**検定関数**という．損失関数としては

$$\mathsf{L}\big(0, \theta\big) = \begin{cases} 0 & (\theta \in \Theta_0) \\ 1 & (\theta \in \Theta_1) \end{cases} ; \qquad \mathsf{L}\big(1, \theta\big) = \begin{cases} 1 & (\theta \in \Theta_0) \\ 0 & (\theta \in \Theta_1) \end{cases}$$

と取る．

	$\theta \in \Theta_0$	$\theta \in \Theta_1$
$d = 0$	0	1
$d = 1$	1	0

通常 $H_0 : \theta \in \Theta_0$ のことを**帰無仮説**とよぶ．さらに, $d(\boldsymbol{x})$ に形式的に $\boldsymbol{X}$ を代入した $d(\boldsymbol{X})$ を**検定統計量**という．$H_1 : \theta \in \Theta_1$ のことを**対立仮説**という．危険関数 $\mathsf{R}\big(d, \theta\big) = \mathsf{E}_\theta\big[\mathsf{L}\big(d(\boldsymbol{X}), \theta\big)\big]$ は以下のようになる．こ

	$\theta \in \Theta_0$	$\theta \in \Theta_1$
$d = 0$	0	第 2 種の誤り
$d = 1$	第 1 種の誤り	0

こで第 1 種の誤りの確率と第 2 種の誤りの確率はトレード・オフの関係になっていることが鍵である．すなわち, 同時には二つの確率を小さくできない．実は

$$(\text{第 1 種の誤りの確率}) + (\text{第 2 種の誤りの確率}) \geq \text{下限}$$

[14]正確には確率化決定関数を考える必要があるが, 議論を簡単にするためにこれは考えないことにする．

ということになっているのである[15]．そこで $\theta \in \Theta_0$ のとき

$$\beta(\theta) := \mathsf{R}\bigl(d,\, \theta\bigr)$$

とし，$\theta \in \Theta_1$ のとき

$$\beta(\theta) := 1 - \mathsf{R}\bigl(d,\, \theta\bigr)$$

と定義したものを**検出力関数**という．仮説検定では与えられた数 $\alpha\,(0 < \alpha < 1)$ に対して

$$\sup_{\theta\in\Theta_0} \mathsf{R}\bigl(d,\, \theta\bigr) \leq \alpha$$

をみたす検定関数の中から $\theta \in \Theta_1$ において $\beta(\theta)$ を大きくするもの，すなわち $\mathsf{R}\bigl(d,\, \theta\bigr)$ を小さくするものと選ぶことを目指す．ちなみに α のことを**有意水準**という．$\sup_{\theta\in\Theta_0} \mathsf{R}\bigl(d,\, \theta\bigr)$ を検定関数 d の**サイズ**という．したがってサイズが有意水準より小さい検定関数の中から $\theta \in \Theta_1$ において検出力関数の値が一様に大きなもの[16]を探したいわけである．

区間推定

議論を簡単にするために $\Theta = \mathbb{R}$ とする．区間推定において行動は $\mathbb{R}$ の区間となる．したがって行動空間は観測データから区間への対応となる．観測データの実現値 $\boldsymbol{X} = \boldsymbol{x}$ に基づく母数 θ の推定区間 $[\ell(\boldsymbol{x}),\, u(\boldsymbol{x})]$ に対して損失関数として

$$\mathsf{L}\bigl([\ell,\, u],\, \theta\bigr) = (d_2 - \ell) - \mathbb{1}\bigl\{\theta \in [\ell,\, u]\bigr\}$$

などが考えられる．この場合には，L は負の値を取ることもある．決定関数

$$d(\boldsymbol{X}) = [\ell(\boldsymbol{X}),\, u(\boldsymbol{X})]$$

に対して危険関数は

$$\mathsf{R}\bigl(d,\, \theta\bigr) = \mathsf{E}_\theta\bigl[u(\boldsymbol{X}) - \ell(\boldsymbol{X})\bigr] - \mathsf{Pr}_\theta\bigl(\theta \in [\ell(\boldsymbol{X}),\, u(\boldsymbol{X})]\bigr)$$

となる．

実数 $\alpha\,(0 < \alpha < 1)$ が与えられたとき

$$\mathsf{Pr}_\theta\bigl(\theta \in [\ell(\boldsymbol{X}),\, u(\boldsymbol{X})]\bigr) \geq 1 - \alpha$$

のもとで区間の長さの期待値 $\mathsf{E}_\theta\bigl[u(\boldsymbol{X}) - \ell(\boldsymbol{X})\bigr]$ を短くする区間が望ましい区間といえよう．α を**信頼係数**とよぶ．

[15] このことは第 7 章で説明する Neyman-Pearson の補題からわかる．

[16] これは究極の目標であり，一様に検出力関数の値を最大にする検定統計量は存在しないかもしれない．

以上のように統計的決定問題の枠組みで統計的推測の問題である点推定, 区間推定, および検定を統一的に扱うことができる.

次に決定空間の元の間に順序 $\prec$ を導入しよう. 決定関数 $\forall d_1, d_2 \in \mathbb{D}$ に対して

$$
\begin{aligned}
&d_1 \prec d_2 \\
&\quad \Leftrightarrow \mathsf{R}(d_1, \theta) \le \mathsf{R}(d_2, \theta)\,(\forall\theta \in \Theta)\text{ かつ }\mathsf{R}(d_1, \theta_0) < \mathsf{R}(d_2, \theta_0)\,(\exists\theta_0 \in \Theta)
\end{aligned}
$$

定める. すると決定空間 $\mathbb{D}$ を標本空間 $\mathbb{X}$ から行動空間 $\mathbb{A}$ への可測関数すべてから成る集合とすれば順序 $\prec$ は半順序になる. すなわち順序 $\prec$ の意味で一番よいものは存在しない.

たとえば $X \sim \mathsf{N}(\mu, 1)$ によって μ を推定する問題を損失関数 $\mathsf{L}\bigl(\widehat{\mu}, \mu\bigr) = \bigl(\widehat{\mu} - \mu\bigr)^2$ のもとで考える. ただし $\widehat{\mu}$ は μ の推定量である. このとき

$$\widehat{\mu}_0 = 0$$

なる推定量は許容的になる. なぜならば $\mu = 0$ において $\widehat{\mu}_0$ の危険関数の値は 0 となるので, $\widehat{\mu}_0$ よりよい推定量は存在しないわけである.

最小の決定関数が存在しない場合には決定関数を比較するための別の観点の導入が必要となる. 主なもので次の二つがある.

(1) 決定関数の最適性について別の概念を導入する. 代表的なものとしてミニマックス基準と Bayes 基準がある.

(2) 考察する決定関数を制限し, その中で危険関数を母数 Θ に関して一様に小さくする決定関数をみつける. たとえば不偏性, 不変性などを導入して, 考察する決定関数を制限する方法がある. また Neyman-Pearson の補題による議論がある.

さらに決定空間のなかからよい決定関数をみつけるのではなく, 一定の原理のよって導かれる決定関数を考えて, それについてなんらかの合理性を証明する方針がある. 統計的決定問題の枠組みからははずれるが, ある原理に基づきなんらかのかたちで合理的な正当化ができる決定関数を導出することが考えられてきた. 導出の原理として推定ではモーメント法, 最尤法 (第 7 章) などが知られている. 検定法では尤度比検定, スコア検定, Wald 検定, Rao 検定 (第 8 章) がある. 区間推定では検定統計量の反転, ピボット法 (第 8 章) などがある.

5.3　章末注釈と参考文献

この章の話題を数学的に厳密に議論をするためには, 数学の進んだ道具立てが必要となるので, 直観的な説明に留めた.

5.4　演習問題

演習問題 5.1.

演習問題 5.2.

第6章　母数モデル

第 6.1 では, 統計的モデルのなかでも最も基本的な正則母数モデルとそのもとので Fisher 情報量を導入する. 第 6.2 では, 母数モデルのなかでも非常によい性質と数学的な美しさをもつ指数型分布族とその基本的な性質を説明する. 第 6.3 では, 十分統計量を定義し, 十分性の利便性の高い判定条件を述べる. 第 6.4 では, 最小十分統計量と完備十分統計量を定義し, それらに関わる重要な定理を述べる. いくつかの主張の証明で Lebeague 積分の積分の収束定理を用いた. この点について慣れていない読者は証明を読み飛ばしても全体の流れを理解するのには困難が伴わないようにした.

6.1　正則母数モデルと Fisher 情報量

定義 6.1. $d, n \in \mathbb{N}$ とし, $\mathbb{X}^n \left(\subset \mathbb{R}^n\right)$ を標本空間とする. $\mathbb{X}$ 上の母数モデル $\mathcal{P} = \{\mathsf{P}_{\boldsymbol{\theta}};\, \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d\}$ が与えられているとする. 分布 $\mathsf{P}_{\boldsymbol{\theta}}$ の p.d.f.(または p.m.f.) $\mathsf{p}(x|\,\boldsymbol{\theta})$ と表記する. このとき母数モデルは正則であるとは次の条件をみたすときをいう.

(1) $\Theta \subset \mathbb{R}^d$ は開集合.

(2) $\mathcal{P}$ に属する分布の p.d.f.(または p.m.f.) は同じ台をもつ. すなわち, 集合 $\{x \in \mathbb{R};\, \mathsf{p}(x|\,\boldsymbol{\theta}) > 0\}$ は $\boldsymbol{\theta} \in \Theta$ に依存しない.

(3) $\forall \boldsymbol{\theta} \in \Theta$ とする. $\mathsf{p}(x|\,\boldsymbol{\theta})$ の $\boldsymbol{\theta}$ の 1 次と 2 次の偏導関数は $x(\in \mathbb{X})$ に関して連続である.

(4) $\mathsf{p}(x|\,\boldsymbol{\theta})$ の $\boldsymbol{\theta}$ に関する 1 次と 2 次の偏導関数は $x(\in \mathbb{X})$ の関数として可積分である.

(5) $\mathsf{p}(x|\,\boldsymbol{\theta})$ の 1 次と 2 次の偏導関数は $\boldsymbol{\theta}$ の微分記号と x の積分記号と交換が可能である.

$X_1 = x_1,\, X_2 = x_2,\, \ldots,\, X_n = x_n$ を観測したとき Θ 上の実数値関数 ℓ_n を

$$\ell_n(\boldsymbol{\theta}|\,\boldsymbol{x}) = \log \prod_{j=1}^{n} \mathsf{p}(x_j|\,\boldsymbol{\theta}) \qquad (\boldsymbol{x} = (x_1,\, x_2,\, \ldots,\, x_n))$$

で定義[1]する.

$\boldsymbol{X} \sim \mathsf{P}_{\boldsymbol{\theta}}^{\otimes n}$ のときこの分布に関する期待値, 分散および共分散を

$$\mathsf{E}_{\boldsymbol{\theta}}[\,\cdot\,], \qquad \mathsf{Var}_{\boldsymbol{\theta}}[\,\cdot\,], \qquad \mathsf{Cov}_{\boldsymbol{\theta}}[\,\cdot, \cdot\,]$$

と表記する. すなわち, 可積分な関数 $h : \mathbb{X}^n\, rightarrow\, \mathbb{R}$ に対して

$$\mathsf{E}_{\boldsymbol{\theta}}[h(\boldsymbol{X})] := \begin{cases} \sum_{\boldsymbol{x}\in\mathbb{X}} h(\boldsymbol{x}) \prod_{j=1}^{n} \mathsf{p}(x_j | \boldsymbol{\theta}) & \text{(離散型の場合)} \\ \int_{\mathbb{X}} h(\boldsymbol{x}) \prod_{j=1}^{n} \mathsf{p}(x_j | \boldsymbol{\theta})\, \mathrm{d}\boldsymbol{x} & \text{(連続型の場合)} \end{cases}$$

と定める. ただし $\mathrm{d}\boldsymbol{x} = \mathrm{d}x_1 \mathrm{d}x_2 \times \cdots \times \mathrm{d}x_n$ である. さらに, $\boldsymbol{X}$ を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率ベクトルとしたとき, 任意の $B \in \mathcal{B}(\mathbb{X}^n)$ に対して, 測度空間 $(\Omega, \mathcal{A})$ 上の確率測度 $\mathsf{Pr}_{\boldsymbol{\theta}}$ を

$$\mathsf{Pr}_{\boldsymbol{\theta}}\big(\boldsymbol{X}^{-1}(B)\big) := \mathsf{Pr}_{\boldsymbol{\theta}}(\boldsymbol{X} \in B) := \mathsf{E}_{\boldsymbol{\theta}}\big[\mathbb{1}_B(\boldsymbol{X})\big]$$

で定める. ただし, $\boldsymbol{X}^{-1}(B) := \{\omega \in \Omega;\, \boldsymbol{X}(\omega) \in B\} \in \mathcal{A}$ である.

さらに, $i, j = 1, 2, \ldots, p$ に対して

$$S_j := S_j(\boldsymbol{x}| \boldsymbol{\theta}) = \frac{\partial \ell_n(\boldsymbol{\theta}| \boldsymbol{x})}{\partial \theta_j}, \quad \dot{S}_{jk} := \dot{S}_{jk}(\boldsymbol{x}| \boldsymbol{\theta}) = \frac{\partial^2 \ell_n(\boldsymbol{\theta}| \boldsymbol{x})}{\partial \theta_j \partial \theta_k},$$
$$\boldsymbol{S} = (S_1, S_2, \ldots, S_d)^\top, \quad \dot{\boldsymbol{S}} = \big(\dot{S}_{jk}\big), \quad \boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d)^\top$$

と定義する. $\boldsymbol{X} \sim \mathsf{P}_{\boldsymbol{\theta}}^{\otimes n}$ のとき Fisher **情報量** $\mathcal{F}_{\boldsymbol{X}}(\boldsymbol{\theta})$ は $d \times d$ 行列で $\mathcal{F}_{\boldsymbol{X}}(\boldsymbol{\theta})$ の (j, k) 成分 $\{\mathcal{F}_{\boldsymbol{X}}(\boldsymbol{\theta})\}_{jk}$ は

$$\{\mathcal{F}_{\boldsymbol{X}}(\boldsymbol{\theta})\}_{jk} = \mathsf{E}_{\boldsymbol{\theta}}\big[S_j(\boldsymbol{X}| \boldsymbol{\theta})\, S_k(\boldsymbol{X}| \boldsymbol{\theta})\big] \qquad (1 \le j, k \le d)$$

で定義される.

定理 6.2. 標本空間 $\mathbb{X}$ 上の母数モデル $\mathcal{P} = \{\mathsf{P}_{\boldsymbol{\theta}};\, \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d\}$ は正則であるとする. $\boldsymbol{X} \sim \mathsf{P}_{\boldsymbol{\theta}}^{\otimes n}$ としたとき以下が成立する.

(1) $\forall \boldsymbol{\theta} \in \Theta$ と $j = 1, 2, \ldots, \mathrm{d}$ に対して $\mathsf{E}_{\boldsymbol{\theta}}\big[S_j(\boldsymbol{X}| \boldsymbol{\theta})\big] = 0$ となる.

(2) $\mathcal{F}_{\boldsymbol{X}}(\boldsymbol{\theta}) = \mathsf{Cov}_{\boldsymbol{\theta}}\big[\boldsymbol{S}(\boldsymbol{X}| \boldsymbol{\theta})\big] = \mathsf{E}_{\boldsymbol{\theta}}\big[\boldsymbol{S}(\boldsymbol{X}| \boldsymbol{\theta}) \boldsymbol{S}^\top(\boldsymbol{X}| \boldsymbol{\theta})\big]$ が成り立つ.

(3) $\mathcal{F}_{\boldsymbol{X}}(\boldsymbol{\theta}) = -\mathsf{E}_{\boldsymbol{\theta}}\big[\dot{\boldsymbol{S}}(\boldsymbol{X}| \boldsymbol{\theta})\big]$ が成り立つ.

Proof. 証明は連続型分布の場合を示す. 離散型の場合は積分を和の記号に置き替えればよい. $\mathsf{P}_{\boldsymbol{\theta}}$ の p.d.f. を $\mathsf{p}(x| \boldsymbol{\theta})$ と表記する. 母数モデルの

[1]対数 尤度(ゆうど) 関数 と呼ぶ.

正則性から微分記号と積分記号の交換が保証されているので

$$\begin{aligned}
\mathsf{E}_{\boldsymbol{\theta}}\big[S_j(\boldsymbol{X}|\,\boldsymbol{\theta})\big] &= \sum_{k=1}^{n}\int_{\mathbb{X}}\left(\frac{\partial}{\partial\theta_j}\log\mathsf{p}(x_k|\,\boldsymbol{\theta})\right)\mathsf{p}(x_k|\,\boldsymbol{\theta})\,\mathrm{d}x_k \\
&= \sum_{k=1}^{n}\int_{\mathbb{X}}\frac{1}{\mathsf{p}(x_k|\,\boldsymbol{\theta})}\left(\frac{\partial}{\partial\theta_j}\mathsf{p}(x_k|\,\boldsymbol{\theta})\right)\mathsf{p}(x_k|\,\boldsymbol{\theta})\,\mathrm{d}x_k \\
&= \sum_{k=1}^{n}\int_{\mathbb{X}}\frac{\partial}{\partial\theta_j}\mathsf{p}(x_k|\,\boldsymbol{\theta})\,\mathrm{d}x_k \\
&= \sum_{k=1}^{n}\frac{\partial}{\partial\theta_j}\int_{\mathbb{X}}\mathsf{p}(x_k|\,\boldsymbol{\theta})\,\mathrm{d}x_k \\
&= \sum_{k=1}^{n}\frac{\partial}{\partial\theta_j}1 = 0
\end{aligned}$$

となる．よって (1) は示された.
(2) は明らか.
(3) を示すために次に注意する．$x \in \mathbb{X}$ に対して

$$\frac{\partial^2\log\mathsf{p}(x|\,\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k} = \frac{1}{\mathsf{p}(x|\,\boldsymbol{\theta})}\frac{\partial^2\mathsf{p}(x|\,\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k} - \left(\frac{\partial}{\partial\theta_j}\log\mathsf{p}(x|\,\boldsymbol{\theta})\right)\left(\frac{\partial}{\partial\theta_k}\log\mathsf{p}(x|\,\boldsymbol{\theta})\right)$$

となる．このとき

$$\begin{aligned}
\mathsf{E}_{\boldsymbol{\theta}}\left[\frac{1}{\mathsf{p}(X_1|\,\boldsymbol{\theta})}\frac{\partial^2\mathsf{p}(X_1|\,\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k}\right] &= \int_{\mathbb{X}}\frac{1}{\mathsf{p}(x|\,\boldsymbol{\theta})}\left(\frac{\partial^2\mathsf{p}(x|\,\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k}\right)\mathsf{p}(x|\,\boldsymbol{\theta})\,\mathrm{d}x \\
&= \frac{\partial^2}{\partial\theta_j\partial\theta_k}\int_{\mathbb{X}}\mathsf{p}(x|\,\boldsymbol{\theta})\,\mathrm{d}x \\
&= \frac{\partial^2}{\partial\theta_j\partial\theta_k}1 = 0
\end{aligned}$$

となる．このことより

$$\begin{aligned}
\mathsf{E}_{\boldsymbol{\theta}}\left[\frac{\partial^2\log\mathsf{p}(X_1|\,\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k}\right] &= \int_{\mathbb{X}}\frac{1}{\mathsf{p}(x|\,\boldsymbol{\theta})}\left(\frac{\partial^2\log\mathsf{p}(x|\,\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k}\right)\mathsf{p}(x|\,\boldsymbol{\theta})\,\mathrm{d}x \\
&\quad - \mathsf{E}_{\boldsymbol{\theta}}\left[\left(\frac{\partial}{\partial\theta_j}\log\mathsf{p}(X_1|\,\boldsymbol{\theta})\right)\left(\frac{\partial}{\partial\theta_k}\log\mathsf{p}(X_1|\,\boldsymbol{\theta})\right)\right] \\
&= -\mathsf{E}_{\boldsymbol{\theta}}\left[\left(\frac{\partial}{\partial\theta_j}\log\mathsf{p}(X_1|\,\boldsymbol{\theta})\right)\left(\frac{\partial}{\partial\theta_k}\log\mathsf{p}(X_1|\,\boldsymbol{\theta})\right)\right] \qquad (6.1)
\end{aligned}$$

がわかる．よって

$$
\begin{aligned}
\{\mathcal{F}_{\boldsymbol{X}}(\boldsymbol{\theta})\}_{ij} &= \mathsf{E}\left[\left(\sum_{k_1=1}^{n}\frac{\partial\log\mathsf{p}(X_{k_1}|\,\boldsymbol{\theta})}{\partial\theta_i}\right)\left(\sum_{k_1=2}^{n}\frac{\partial\log\mathsf{p}(X_{k_2}|\,\boldsymbol{\theta})}{\partial\theta_j}\right)\right] \\
&= \sum_{k=1}^{n}\mathsf{E}\left[\frac{\partial\log\mathsf{p}(X_k|\,\boldsymbol{\theta})}{\partial\theta_i}\frac{\partial\log\mathsf{p}(X_k|\,\boldsymbol{\theta})}{\partial\theta_j}\right] \\
&\qquad + \sum_{k_1\neq k_2}\mathsf{E}\left[\frac{\partial\log\mathsf{p}(X_{k_1}|\,\boldsymbol{\theta})}{\partial\theta_i}\frac{\partial\log\mathsf{p}(X_{k_2}|\,\boldsymbol{\theta})}{\partial\theta_j}\right] \\
&= \sum_{k=1}^{n}\mathsf{E}\left[\frac{\partial\log\mathsf{p}(X_k|\,\boldsymbol{\theta})}{\partial\theta_i}\frac{\partial\log\mathsf{p}(X_k|\,\boldsymbol{\theta})}{\partial\theta_j}\right] \\
&\qquad + \sum_{k_1\neq k_2}\underbrace{\mathsf{E}\left[\frac{\partial\log\mathsf{p}(X_{k_1}|\,\boldsymbol{\theta})}{\partial\theta_i}\right]}_{=0}\mathsf{E}\left[\frac{\partial\log\mathsf{p}(X_{k_2}|\,\boldsymbol{\theta})}{\partial\theta_j}\right] \\
&= -\sum_{k=1}^{n}\mathsf{E}_{\boldsymbol{\theta}}\left[\frac{\partial^2\log\mathsf{p}(X_k|\,\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k}\right] \qquad (\because (6.1)) \\
&= -\mathsf{E}\left[\frac{\partial^2\ell_n(\boldsymbol{X}|\,\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right] = -\mathsf{E}\big[\dot{\mathbf{S}}(\boldsymbol{X}|\,\boldsymbol{\theta})\big]
\end{aligned}
$$

がわかる． □

例 6.3. $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ で $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}(\mu, \sigma^2)$ とする．このとき, 統計的モデルは $\mathcal{P} = \{\mathsf{N}(\mu, \sigma^2);\, \mu\in\mathbb{R},\, \sigma>0\}$ で与えられる．したがって

$$
\Theta = \{\boldsymbol{\theta} = (\mu, \sigma)\in\mathbb{R}^2;\, \mu\in\mathbb{R},\, \sigma>0\} = \mathbb{R}\times(0, \infty)
$$

となる．このとき

$$
\ell_n(\boldsymbol{\theta}|\,\boldsymbol{x}) = -\sum_{j=1}^{n}\frac{(x_j-\mu)^2}{2\sigma^2} - n\log\sigma - n\log(\sqrt{2\pi}); \qquad \boldsymbol{x} = (x_1, x_2, \ldots, x_n)
$$

となる．よって

$$
\begin{aligned}
S_1(\boldsymbol{x}|\,\boldsymbol{\theta}) &= \frac{\partial\ell_n(\boldsymbol{\theta}|\,\boldsymbol{x})}{\partial\mu} = \sum_{j=1}^{n}\frac{x_j-\mu}{\sigma^2}, \\
S_2(\boldsymbol{x}|\,\sigma) &= \frac{\partial\ell_n(\boldsymbol{\theta}|\,\boldsymbol{x})}{\partial\sigma} = \sum_{j=1}^{n}\frac{(x_j-\mu)^2}{\sigma^3} - \frac{n}{\sigma}
\end{aligned}
$$

となる．これらの期待値を取ると

$$
\begin{aligned}
\mathsf{E}_{\boldsymbol{\theta}}\big[S_1^2(\boldsymbol{X}|\,\boldsymbol{\theta})\big] &= \frac{n}{\sigma^2}, \\
\mathsf{E}_{\boldsymbol{\theta}}\big[S_2^2(\boldsymbol{X}|\,\boldsymbol{\theta})\big] &= \frac{3n}{\sigma^2} - \frac{2n}{\sigma^2} + \frac{n}{\sigma^2} = \frac{2n}{\sigma^2}, \\
\mathsf{E}_{\boldsymbol{\theta}}\big[S_1(\boldsymbol{X}|\,\boldsymbol{\theta})S_2(\boldsymbol{X}|\,\boldsymbol{\theta})\big] &= 0
\end{aligned}
$$

となるので

$$\mathcal{F}_X(\boldsymbol{\theta}) = \frac{n}{\sigma^2}\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

を得る．

一方

$$\dot{S}_{11} = \frac{\partial^2 \ell(\boldsymbol{\theta}|\,\boldsymbol{x})}{\partial \mu^2} = \frac{\partial}{\partial \mu}\left(\sum_{j=1}^{n} \frac{x_j - \mu}{\sigma^2}\right) = -\frac{n}{\sigma^2},$$

$$\dot{S}_{22} = \frac{\partial^2 \ell(\boldsymbol{\theta}|\,\boldsymbol{x})}{\partial \sigma^2} = \frac{\partial}{\partial \sigma}\left(\sum_{j=1}^{n} \frac{(x_j - \mu)^2}{\sigma^2} - \frac{n}{\sigma}\right) = -3\sum_{j=1}^{n} \frac{(x_j - \mu)^2}{\sigma^2} + \frac{n}{\sigma^2},$$

$$\dot{S}_{12} = \frac{\partial^2 \ell(\boldsymbol{\theta}|\,\boldsymbol{x})}{\partial \mu \partial \sigma} = \frac{\partial}{\partial \sigma}\left[\sum_{j=1}^{n} \frac{x_j - \mu}{\sigma^2}\right] = -2\sum_{j=1}^{n} \frac{x_j - \mu}{\sigma^3}$$

となる．よって

$$\mathsf{E}_{\boldsymbol{\theta}}\big[\dot{S}_{11}(\boldsymbol{X}|\,\boldsymbol{\theta})\big] = -\frac{n}{\sigma^2},\ \mathsf{E}_{\boldsymbol{\theta}}\big[\dot{S}_{22}(\boldsymbol{X}|\,\boldsymbol{\theta})\big] = -\frac{2n}{\sigma^2},\ \mathsf{E}_{\boldsymbol{\theta}}\big[\dot{S}_{12}(\boldsymbol{X}|\,\boldsymbol{\theta})\big] = 0$$

なので

$$\begin{bmatrix} \mathsf{E}_{\boldsymbol{\theta}}\big[\dot{S}_{11}(\boldsymbol{X}|\,\boldsymbol{\theta})\big] & \mathsf{E}_{\boldsymbol{\theta}}\big[\dot{S}_{12}(\boldsymbol{X}|\,\boldsymbol{\theta})\big] \\ \mathsf{E}_{\boldsymbol{\theta}}\big[\dot{S}_{12}(\boldsymbol{X}|\,\boldsymbol{\theta})\big] & \mathsf{E}_{\boldsymbol{\theta}}\big[\dot{S}_{22}(\boldsymbol{X}|\,\boldsymbol{\theta})\big] \end{bmatrix} = -\mathcal{F}_X(\boldsymbol{\theta})$$

となる． □

6.2 指数型分布族

この節では, 統計的モデルでもっとも重要なものを導入し, そのモデルの性質を説明する．

6.2.1 1 変数の場合

定義 6.4. $d \in \mathbb{N}$ とし, $\mathbb{X} \subset \mathbb{R}^d$ を空でない部分集合とする．さらに $\Theta \subset \mathbb{R}$ も空でない部分集合とする．$\mathbb{X}$ 上の統計的モデル $\{\mathsf{P}_\theta;\, \theta \in \Theta\}$ は 1 **母数指数型分布族**であるとは, Θ 上の実数値関数 $A(\theta)$ と $\kappa(\theta)$, $\mathbb{X}$ 上の実数値関数 T と h が存在して[2], P_θ の p.d.f. または p.m.f. $\mathsf{p}(\boldsymbol{x}|\,\theta)$ が

$$\mathsf{p}(\boldsymbol{x}|\,\theta) = \begin{cases} h(\boldsymbol{x}) \exp\big\{A(\theta)T(\boldsymbol{x}) - \kappa(\theta)\big\} & (\boldsymbol{x} \in \mathbb{X} \subset \mathbb{R}^d) \\ 0 & (\boldsymbol{x} \notin \mathbb{X}) \end{cases} \tag{6.2}$$

の形で書けるときをいう．

[2] A と B は Greek letters 大文字の α と β である．

注意 6.5. 関数 A, h, T の表現は一意ではない. □

例 6.6. (Poisson 分布族) $X \sim \mathsf{Po}(\theta)\,(\theta > 0)$ とする. その p.m.f. は

$$\mathsf{p}(x|\,\theta) = \frac{\theta^x e^{-\theta}}{x!} = \frac{1}{x!}\exp\bigl\{x\log\theta - \theta\bigr\} \qquad (x = 0,\,1,\,\ldots)$$

である. したがって $\{\mathsf{Po}(\theta);\,\theta \in (0,\,\infty)\}$ は

$$d = 1,\; A(\theta) = \log\theta,\; \kappa(\theta) = \theta,\; T(x) = x,\; h(x) = \frac{1}{x!}$$

によって生成される 1 母数指数型分布族である. □

例 6.7. (2 項分布族) $X \sim \mathsf{Bino}(n,\,\theta)$ とする. ただし $n \in \mathbb{N},\, 0 < \theta < 1$ である. このとき X の p.m.f. は

$$\begin{aligned}\mathsf{p}(x|\,\theta) &= \begin{pmatrix} n \\ x \end{pmatrix}\theta^x(1-\theta)^{n-x} \\ &= \begin{pmatrix} n \\ x \end{pmatrix}\exp\left\{x\log\left(\frac{\theta}{1-\theta}\right) + n\log(1-\theta)\right\} \qquad (x = 0,\,1,\,\ldots,\,n)\end{aligned}$$

と書ける. ただし

$$\begin{pmatrix} n \\ x \end{pmatrix} = \frac{n!}{x! \times (n-x)!}, \qquad 0! = 1$$

である. したがって $\{\mathsf{Bino}(n,\,\theta);\,\theta \in (0,\,1)\}$ は

$$d = 1,\; A(\theta) = \log\left(\frac{\theta}{1-\theta}\right),\; \kappa(\theta) = -n\log(1-\theta),\; T(x) = x,$$
$$h(x) = \begin{pmatrix} n \\ x \end{pmatrix}$$

によって生成される 1 母数指数型分布族である. □

例 6.8. $\boldsymbol{X} = (Z,\,Y)^\top,\, Z,\, W \overset{\text{i.i.d.}}{\sim} \mathsf{N}(0,\,1)$ とし, $Y = Z + \theta W\,(\theta > 0)$ とする. このとき $\boldsymbol{X}$ の同時 p.d.f. は次で与えられる. $\boldsymbol{x} = (y,\,z)$ としたとき

$$\begin{aligned}\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta) = \mathsf{p}^{\boldsymbol{X}}(y,\,z|\,\theta) &= \mathsf{p}^{Z}(z)\mathsf{p}^{Y|\,Z}(y|\,z) = \varphi(z)\theta^{-1}\varphi\left(\frac{y-z}{\theta}\right) \\ &= \frac{1}{2\pi\theta}\exp\left\{-\frac{1}{2}\left(z^2 + \frac{(y-z)^2}{\theta^2}\right)\right\} \\ &= \frac{1}{2\pi}\exp\left\{-\frac{z^2}{2}\right\}\exp\left\{-\frac{(y-z)^2}{2\theta^2} - \log\theta\right\}\end{aligned}$$

と書ける．ただし p_Z は Z の p.d.f. とし，$\mathsf{p}_{Y|Z}$ は $Z=z$ を与えたときの Y の条件付き p.d.f.，φ は $\mathsf{N}(0,1)$ の p.d.f. で

$$\varphi(z)=\frac{1}{\sqrt{2\pi}}e^{-z^2/2}\qquad(z\in\mathbb{R})$$

である．したがって $\boldsymbol{X}$ の分布は

$$d=2,\ A(\theta)=-\frac{1}{2\theta^2},\ \kappa(\theta)=\log\theta,\ T(\boldsymbol{x})=(y-z)^2,$$
$$h(\boldsymbol{x})=\frac{1}{2\pi}\exp\left\{-\frac{z^2}{2}\right\}$$

で生成される 1 母数指数型分布族に属する． □

1 母数指数型分布族に属する分布からのランダム標本の分布も 1 母数指数型分布族に属する．特に，$n\in\mathbb{N}$ とし，$X_1,\,X_2,\,\ldots,\,X_n\overset{\text{i.i.d.}}{\sim}\mathsf{P}_\theta$ とする．ただし，$\{\mathsf{P}_\theta;\,\theta\in\Theta\}$ は (6.2) で与えられる指数型分布族とする．すると $\{\mathsf{P}_\theta^{\otimes n};\,\theta\in\Theta\}$ は $\boldsymbol{X}^{(n)}=(\boldsymbol{X}_1,\,\boldsymbol{X}_2,\,\ldots,\,\boldsymbol{X}_m)$ に対する指数型分布族となる．ただし，$\mathbb{R}^{dn}$ 上の確率測度を

$$\mathsf{P}_\theta^{\otimes n}=\underbrace{\mathsf{P}_\theta\times\mathsf{P}_\theta\times\cdots\times\mathsf{P}_\theta}_{n\,個}$$

と定めてた．このとき，$\mathsf{P}_\theta^{\otimes n}$ の同時 p.d.f. $\prod_{j=1}^n\mathsf{p}(\,\cdot\,|\,\theta)$ は次で与えられる．すると

$$\begin{aligned}\prod_{j=1}^n\mathsf{p}(\boldsymbol{x}_j|\,\theta)&=\prod_{j=1}^n h(\boldsymbol{x}_j)\exp\Bigl\{A(\theta)T(\boldsymbol{x}_j)-\kappa(\theta)\Bigr\}\\&=\left[\prod_{j=1}^n h(\boldsymbol{x}_j)\right]\exp\left\{A(\theta)\sum_{j=1}^n T(\boldsymbol{x}_j)-n\kappa(\theta)\right\}\end{aligned}$$

となる．したがって $\{\mathsf{P}_\theta;\,\theta\in\Theta\}$ は

$$A(\theta),\quad\sum_{j=1}^n T(\boldsymbol{x}_j),\quad n\kappa(\theta),\quad\prod_{j=1}^n h(\boldsymbol{x}_j)$$

で生成される $\mathbb{R}^{dn}$ 上の 1 母数指数型分布族となる．

統計量 $\sum_{j=1}^n T(\boldsymbol{X}_j)$ は θ の自然十分統計量となる．

指数型分布族の正準表示: (6.2) で表現された指数型分布族を θ でなく $\eta\bigl(=A(\theta)\bigr)$ で添え字付けることを考える．記号の乱用[3]すると 1 母数指数

[3]ここでは，記号の乱用をしている．本来であれば，領域を制限すると A の逆関数は存在するので，$\mathsf{p}(\boldsymbol{x}|\,\eta)$ は $\mathsf{p}(\boldsymbol{x}|\,A^{-1}(\eta))$ と書くべきであろう．

型分布族は

$$\mathsf{p}(\boldsymbol{x}|\,\eta) = \begin{cases} h(\boldsymbol{x})\exp\bigl\{\eta T(\boldsymbol{x}) - \kappa^{\vee}(\eta)\bigr\} & (\boldsymbol{x} \in \mathbb{X} \subset \mathbb{R}^d) \\ 0 & (\boldsymbol{x} \notin \mathbb{X}) \end{cases} \tag{6.3}$$

と書くことができる．ただし

$$\kappa^{\vee}(\eta) = \begin{cases} \log\left(\sum_{\boldsymbol{x}\in\mathbb{X}} h(\boldsymbol{x})\exp\bigl\{\eta T(\boldsymbol{x})\bigr\}\right) & (\text{離散型の場合}) \\ \log\left(\int_{\mathbb{X}} h(\boldsymbol{x})\exp\bigl\{\eta T(\boldsymbol{x})\bigr\}\,\mathrm{d}\boldsymbol{x}\right) & (\text{連続型の場合}) \end{cases}$$

である．$\kappa^{\vee}$ の定義より，直ちに $\kappa^{\vee}(\eta) = \kappa^{\vee}(A(\theta)) = \kappa(\theta)$ がわかる．さらに

$$\mathcal{E} := \{\eta \in \mathbb{R};\, \kappa^{\vee}(\eta) < \infty\}$$

とおく．$\mathcal{E}$ は $\mathbb{R}$ の区間となる．すると確率分布族

$$\Bigl\{\mathsf{p}(\boldsymbol{x}|\,\eta) = h(\boldsymbol{x})\exp\bigl\{\eta T(\boldsymbol{x}) - \kappa^{\vee}(\eta)\bigr\};\, \eta \in \mathcal{E}\Bigr\}$$

は確率分布族

$$\Bigl\{\mathsf{p}(\boldsymbol{x}|\,\theta) = h(\boldsymbol{x})\exp\bigl\{A(\theta)T(\boldsymbol{x}) - \kappa(\theta)\bigr\};\, \theta \in \Theta\Bigr\}$$

を含む．確率分布族 $\{\mathsf{p}(\boldsymbol{x}|\,\eta);\, \eta \in \mathcal{E}\}$ は T, h によって生成された **正準**(せいじゅん)**指数型分布族**または**自然指数型分布族**といい，$\mathcal{E}$ を**自然母数空間**といい，T を自然十分統計量という．

例 6.9. (例 6.6 の続き) Poisson 分布族は

$$\begin{aligned} \mathsf{p}(x|\,\eta) &= \frac{1}{x!}\exp\bigl\{\eta x - e^{\eta}\bigr\} \qquad (x = 0,\, 1,\, 2, \ldots), \\ \eta &= \log\theta,\, h(x) = \frac{1}{x!},\, T(x) = x, \\ \exp\bigl(\kappa^{\vee}(\eta)\bigr) &= \sum_{x=0}^{\infty} h(x)\exp\bigl(\eta T(x)\bigr) = \sum_{x=0}^{\infty}\frac{\exp(\eta x)}{x!} = \sum_{x=0}^{\infty}\frac{\bigl(e^{\eta}\bigr)^x}{x!} = \exp\bigl(e^{\eta}\bigr), \\ \mathcal{E} &= \mathbb{R} \end{aligned}$$

となる．□

補題 6.10. [4]

$$\mathcal{E} := \Bigl\{\eta \in \mathbb{R};\, \kappa^{\vee}(\eta) := \mathsf{E}[\exp\{\eta T(\boldsymbol{X})\}] < \infty\Bigr\}$$

とする．このとき，関数 $\kappa^{\vee}(\eta)$ は $\mathcal{E}^{\circ}$ 上の無限回微分可能である．ただし $\mathcal{E}^{\circ}$ は $\mathcal{E}$ の内部である．さらに積分記号と微分記号の交換は可能である．

[4] 証明には，優収束的理を用いている．

Proof. $\eta \in \mathcal{E}^\circ$ なので, ある $\epsilon > 0$ が存在して $[\eta - 2\epsilon,\, \eta + 2\epsilon] \subset \mathcal{E}^\circ$ とできる.

$$
\begin{aligned}
\frac{d}{d\eta} \exp\bigl\{\kappa^\vee(\eta)\bigr\} &= \lim_{n\to\infty} \frac{\exp\left\{\kappa^\vee\left(\eta + \dfrac{\epsilon}{n}\right)\right\} - \exp\bigl\{\kappa^\vee(\eta)\bigr\}}{\epsilon/n} \\
&= \lim_{n\to\infty} \int \frac{\exp\left\{\left(\eta + \dfrac{\epsilon}{n}\right) T(\boldsymbol{x})\right\} - \exp\bigl\{\kappa^\vee(\eta)\bigr\}}{\epsilon/n} h(\boldsymbol{x}) \,\mathrm{d}x \\
&= \lim_{n\to\infty} \int \frac{\exp\bigl\{\eta T(\boldsymbol{x})\bigr\} \left[\exp\left\{\dfrac{\epsilon T(\boldsymbol{x})}{n}\right\} - 1\right]}{\epsilon/n} h(\boldsymbol{x}) \,\mathrm{d}x \\
&=: \lim_{n\to\infty} \int \exp\bigl\{\eta T(\boldsymbol{x})\bigr\} f_n(\boldsymbol{x}) \,\mathrm{d}x
\end{aligned}
$$

と書ける. ただし

$$
f_n(\boldsymbol{x}) = \frac{\exp\bigl\{\epsilon T(\boldsymbol{x})/n\bigr\} - 1}{\epsilon/n} h(\boldsymbol{x})
$$

である. すると

$$
\lim_{n\to\infty} f_n(\boldsymbol{x}) =: f(\boldsymbol{x}) = T(\boldsymbol{x}) h(\boldsymbol{x})
$$

となる. ここで

$$
|e^t - 1| \le |t| e^{|t|}; \qquad |t| \le e^{|t|} \qquad (t \in \mathbb{R})
$$

に注意する. これらの不等式を用いると

$$
\begin{aligned}
\left| \frac{\exp\bigl\{\epsilon T(\boldsymbol{x})/n\bigr\} - 1}{\epsilon/n} \right| &\le \left\{ \left| \frac{\epsilon T(\boldsymbol{x})}{n} \Big/ \frac{\epsilon}{n} \right| \right\} \exp\left\{ \left| \frac{\epsilon T(\boldsymbol{x})}{n} \right| \right\} \\
&\le \frac{1}{\epsilon} \bigl|\epsilon T(\boldsymbol{x})\bigr| \exp\left\{ \left| \frac{\epsilon T(\boldsymbol{x})}{n} \right| \right\} \\
&\le \frac{1}{\epsilon} \exp\bigl\{\bigl|\epsilon T(\boldsymbol{x})\bigr|\bigr\} \exp\bigl\{\bigl|\epsilon T(\boldsymbol{x})\bigr|\bigr\} \\
&= \frac{1}{\epsilon} \exp\bigl\{\bigl|2\epsilon T(\boldsymbol{x})\bigr|\bigr\} \\
&\le \frac{1}{\epsilon} \Bigl\{ \exp\bigl(2\epsilon T(\boldsymbol{x})\bigr) + \exp\bigl(-2\epsilon T(\boldsymbol{x})\bigr) \Bigr\}
\end{aligned}
$$

を得る. よって

$$
|f_n(\boldsymbol{x})| \le \frac{1}{\epsilon} \Bigl\{ \exp\bigl(2\epsilon T(\boldsymbol{x})\bigr) + \exp\bigl(-2\epsilon T(\boldsymbol{x})\bigr) \Bigr\} h(\boldsymbol{x})
$$

となるので

$$\begin{aligned}\Big|\exp\{\eta T(\boldsymbol{x})\} f_n(\boldsymbol{x})\Big| &\le \frac{1}{\epsilon}\Big\{\exp\big((\eta+2\epsilon)T(\boldsymbol{x})\big)+\exp\big((\eta-2\epsilon)T(\boldsymbol{x})\big)\Big\}h(\boldsymbol{x})\\ &=: g(\boldsymbol{x})\end{aligned}$$

を得る．$[\eta-2\epsilon, \eta+2\epsilon] \subset \mathcal{E}^\circ$ から

$$\int g(\boldsymbol{x})\,\mathrm{d}x < \infty$$

となる．よって優収束定理から

$$\begin{aligned}\frac{d}{d\eta}\exp\{\kappa^\vee(\eta)\} &= \lim_{n\to\infty}\int \exp\{\eta T(\boldsymbol{x})\} f_n(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\\ &= \int \exp\{\eta T(x)\} T(\boldsymbol{x}) h(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\end{aligned}$$

がわかる．この操作を繰り返して行けば，$\kappa^\vee(\eta)$ は無限回微分可能なことがわかる． □

定理 6.11. X は (6.3) で与えられた自然指数型分布族に属する分布 $\mathsf{p}(\boldsymbol{x}|\,\eta)$ に従うとする．η は $\mathcal{E}$ の内点としたとき，$T(\boldsymbol{X})$ の積率母関数は原点の近傍で存在し

$$M_T(s) = \exp\big[\kappa^\vee(s+\eta) - \kappa^\vee(\eta)\big]$$

で与えられる．ただし，s s は 0 のある近傍に含まれるとする．さらに

$$\mathsf{E}[T(\boldsymbol{X})] = \dot\kappa^\vee(\eta); \qquad \mathsf{Var}[T(\boldsymbol{X})] = \ddot\kappa^\vee(\eta) \tag{6.4}$$

である．ただし

$$\dot\kappa^\vee(\eta) = \frac{\mathrm{d}\kappa^\vee}{\mathrm{d}\eta}(\eta),\ \ddot\kappa^\vee(\eta) = \frac{\mathrm{d}^2\kappa^\vee}{\mathrm{d}\eta^2}(\eta)$$

である．

Proof. 連続型分布の場合について証明する．離散型の場合には積分を和に変更するばよい．$\kappa^\vee$ の定義に注意すると

$$\begin{aligned}M(s) &= \mathsf{E}\big[\exp(sT(\boldsymbol{X}))\big]\\ &= \int_{\mathbb{X}} h(\boldsymbol{x})\exp\big[(s+\eta)T(\boldsymbol{x}) - \kappa^\vee(\eta)\big]\,\mathrm{d}\boldsymbol{x}\\ &= \exp\big[\kappa^\vee(s+\eta)-\kappa^\vee(\eta)\big]\int_{\mathbb{X}} h(x)\exp\big[(s+\eta)T(\boldsymbol{x}) - \kappa^\vee(s+\eta)\big]\,\mathrm{d}\boldsymbol{x}\\ &= \exp\big[\kappa^\vee(s+\eta)-\kappa^\vee(\eta)\big]\end{aligned}$$

から 1 番目の主張はわかる. 最後の等号は $s+\eta \in \mathcal{E}^\circ$ となるように s をとると

$$\int_{\mathbb{X}} h(\boldsymbol{x}) \exp\bigl[(s+\eta)T(\boldsymbol{x}) - \kappa^\vee(s+\eta)\bigr]\,\mathrm{d}\boldsymbol{x} = 1$$

となることよりわかる. 次に, 補題 6.10 から

$$\int_{\mathbb{X}} T(\boldsymbol{x}) \exp\bigl\{\eta T(\boldsymbol{x})\bigr\} h(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \frac{\mathrm{d}}{\mathrm{d}\eta} \exp\bigl\{\kappa^\vee(\eta)\bigr\} = \dot{\kappa}^\vee(\eta) \exp\bigl\{\kappa^\vee(\eta)\bigr\}$$

となるので

$$\dot{\kappa}^\vee(\eta) = \int T(\boldsymbol{x}) \exp\bigl\{\eta T(\boldsymbol{x}) - \kappa^\vee(\eta)\bigr\} h(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \mathsf{E}\bigl[T(\boldsymbol{X})\bigr]$$

がわかる. □

6.3　十分統計量

$\mathbb{X} \subset \mathbb{R}$ を空でない部分集合とする. $n \in \mathbb{N}$ とし, $X_1, X_2, \ldots, X_n$ は確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の $\mathbb{X}$ 値確率変数列で, 互いに独立で各々同一の確率分布を持つとする. $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ とおき $B \in \mathcal{B}(\mathbb{X}^n)$ に対し

$$\begin{aligned}\mathsf{P}^{\boldsymbol{X}}(B) &= \mathsf{Pr}\bigl\{\boldsymbol{X}^{-1}(B)\bigr\} \\ &= \mathsf{Pr}\Bigl(\bigl\{\omega \in \Omega;\ \boldsymbol{X}(\omega) = (X_1(\omega), X_2(\omega), \ldots, X_n(\omega)) \in B\bigr\}\Bigr)\end{aligned}$$

とおく. さらに $B \in \mathcal{B}(\mathbb{X})$ に対し

$$\mathsf{P}^{X_j}(B) = \mathsf{Pr}\Bigl(\bigl\{\omega \in \Omega;\ X_j(\omega) \in B\bigr\}\Bigr) \qquad (j = 1, 2, \ldots, n)$$

と書いたとき

$$\mathsf{P}^{\boldsymbol{X}} = \mathsf{P}^{X_1} \times \mathsf{P}^{X_2} \times \cdots \times \mathsf{P}^{X_n}$$

と書ける. $\boldsymbol{X}$ により誘導された確率空間を $(\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n), \mathsf{P}^{\boldsymbol{X}})$ とかく.

$(\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n))$ 上の統計的モデルを $\mathcal{P} = \{\mathsf{P}_\theta;\ \theta \in \Theta\}$ と書く. ただし θ は母数で Θ は母数空間である. さらにある $\theta^* \in \Theta$ が存在して

$$\mathsf{P}^{X_1} = \mathsf{P}_{\theta^*}$$

とする.

次に統計量 $\boldsymbol{T}(\boldsymbol{X})$ を

$$\boldsymbol{T} \colon (\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n)) \to (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$$

なる可測関数で θ に依存しないものとする. ただし $k \in \mathbb{N}$ である.

定義 6.12. $\boldsymbol{X} \sim \mathsf{P}_\theta^{\otimes n}\,(\theta \in \Theta)$ とする．$\forall B \in \mathcal{B}(\mathbb{R}^n)$ に対し統計量 $\boldsymbol{T}(\boldsymbol{X})$ を与えたときの条件付き確率

$$\mathsf{P}_\theta(B|\,\boldsymbol{T})$$

が $\theta \in \Theta$ に無関係であるとき, $\boldsymbol{T}$ は $\mathcal{P} = \{\mathsf{P}_\theta;\, \theta \in \Theta\}$ に対する**十分統計量** (sufficient statistic) であるという．

定理 6.13. (因子分解定理) 統計的モデルを $\mathcal{P} = \{\mathsf{P}_\theta;\, \theta \in \Theta\}$ とする．ただし θ は母数で Θ は母数空間とする．$\boldsymbol{X}$ は確率空間 $\big(\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n), \mathsf{P}_\theta^{\otimes n}\big)$ 上の大きさが n の標本とし，同時 p.d.f. $\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)$ [5]を持つとする．このとき統計量 $\boldsymbol{T} : \mathbb{R}^n \longrightarrow \mathbb{R}^k$ が統計的モデル $\mathcal{P}$ に対する十分統計量であるための必要十分条件は, $\boldsymbol{X}$ の同時 p.d.f.(または同時 p.m.f.) $\mathsf{p}(\boldsymbol{x}|\,\theta)$ が

$$\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta) = g_\theta\{\boldsymbol{T}(\boldsymbol{x})\}h(\boldsymbol{x}) \qquad (\boldsymbol{x} \in \mathbb{X}^n) \tag{6.5}$$

の形で表されるときである．ここで $g_\theta\{\boldsymbol{T}(\boldsymbol{x})\}$ と $h(\boldsymbol{x})$ は非負値関数で $h(\boldsymbol{x})$ は θ に無関係な関数で, g_θ は $\boldsymbol{T}$ を通してのみ $\boldsymbol{x}$ に依存する．

Proof. $\boldsymbol{X}$ が離散型のときのみの証明を与える．$\boldsymbol{T}$ は十分統計量とする．$\boldsymbol{T}$ の p.m.f. を $\mathsf{p}^{\boldsymbol{T}}(\boldsymbol{t}|\,\theta)\,(\boldsymbol{t} \in \mathbb{R}^k)$ と書く．$\boldsymbol{T} = \boldsymbol{t}$ が与えられたときの $\boldsymbol{X}$ の条件付き p.m.f. を $\mathsf{p}^{\boldsymbol{X}|\boldsymbol{T}}(\boldsymbol{x}|\,\boldsymbol{T} = \boldsymbol{t})$ とする．仮定から $\boldsymbol{T}$ は十分統計量なので $\mathsf{p}^{\boldsymbol{X}|\boldsymbol{T}}(\boldsymbol{x}|\,\boldsymbol{T} = \boldsymbol{t})$ は θ に依存しない．条件付き p.m.f. の定義から

$$\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta) = \mathsf{p}^T(\boldsymbol{t}|\,\theta)\mathsf{p}^{\boldsymbol{X}|\boldsymbol{T}}(\boldsymbol{x}|\,\boldsymbol{T} = \boldsymbol{t}) \tag{6.6}$$

となる．(6.6) において

$$h(\boldsymbol{x}) = \mathsf{p}^{\boldsymbol{X}|\boldsymbol{T}}(\boldsymbol{x}|\,\boldsymbol{T} = \boldsymbol{t}), \quad g_\theta\{\boldsymbol{T}(\boldsymbol{x})\} = \mathsf{p}^T(\boldsymbol{t}|\,\theta)$$

とおけば, (6.5) の形になる．

次に (6.5) と書けたときに $\boldsymbol{T}$ は十分統計量であることを示す．まず

$$\begin{aligned}
\mathsf{p}^T(\boldsymbol{t}|\,\theta) &= \sum_{\boldsymbol{y} \in \mathbb{R}^n;\, \boldsymbol{T}(\boldsymbol{y}) = \boldsymbol{t}} \mathsf{p}^{\boldsymbol{X}}(\boldsymbol{y}|\,\theta) \\
&= \sum_{\boldsymbol{y} \in \mathbb{R}^n;\, \boldsymbol{T}(\boldsymbol{y}) = \boldsymbol{t}} g_\theta\big\{\boldsymbol{T}(\boldsymbol{y})\big\}h(\boldsymbol{y}) \\
&= g_\theta(\boldsymbol{t}) \sum_{\boldsymbol{y} \in \mathbb{R}^n;\, \boldsymbol{T}(\boldsymbol{y}) = \boldsymbol{t}} h(\boldsymbol{y})
\end{aligned}$$

に注意する．$\boldsymbol{T}(\boldsymbol{x}) = \boldsymbol{t}$ とする．(6.5) から $\boldsymbol{T} = \boldsymbol{t}$ を与えたときの $\boldsymbol{X}$ の条件付き p.m.f. は

$$\begin{aligned}
\frac{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)}{\mathsf{p}^T(\boldsymbol{t}|\,\theta)} &= \frac{g_\theta(\boldsymbol{t})h(\boldsymbol{x})}{g_\theta(\boldsymbol{t}) \sum_{\boldsymbol{y} \in \mathbb{R}^n;\, \boldsymbol{T}(\boldsymbol{y}) = \boldsymbol{t}} h(\boldsymbol{y})} \\
&= \frac{h(\boldsymbol{x})}{\sum_{\boldsymbol{y} \in \mathbb{R}^n;\, \boldsymbol{T}(\boldsymbol{y}) = \boldsymbol{t}} h(\boldsymbol{y})}
\end{aligned}$$

[5]離散型のときは同時 p.m.f. $\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)$) をもつ．

となり, $\mathsf{p}^{\boldsymbol{X}|\boldsymbol{T}}$ は θ に依存しない．よって $\boldsymbol{T}$ は $\mathcal{P} = \{\mathsf{P}_\theta;\, \theta \in \Theta\}$ に対する十分統計量である．

□

注意 6.14. Fisher-Neyman の因子分解定理の測度論的な証明には Radon-Nikodym の微分に関わる知識が必要になる．これについては定理 C.70 を参照のこと．

例 6.15. $n \in \mathbb{N}$ とし, $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Ber}(\theta)\,(0 \leq \theta \leq 1)$ とする．$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ の同時 p.m.f. は

$$\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta) = \theta^{\sum_{j=1}^n x_j}(1-\theta)^{n-\sum_{j=1}^n x_j}\mathbb{1}_{\{0,\,1\}^n}(\boldsymbol{x}),$$
$$\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \{0, 1\}^n,$$
$$\mathbb{1}_{\{0,\,1\}^n}(\boldsymbol{x}) = \begin{cases} 1 & (\boldsymbol{x} \in \{0, 1\}^n) \\ 0 & (\text{その他の場合}) \end{cases}$$

となる．したがって

$$T(\boldsymbol{x}) = \sum_{j=1}^n x_j, \quad g_\theta\{T(\boldsymbol{x})\} = \theta^{T(\boldsymbol{x})}(1-\theta)^{n-T(\boldsymbol{x})}, \quad h(\boldsymbol{x}) = \mathbb{1}_{\{0,\,1\}^n}(\boldsymbol{x})$$

とおけば, 定理 6.13 から $T(\boldsymbol{X}) = X_1 + X_2 + \cdots + X_n$ は $\mathcal{P} = \{p(\boldsymbol{x}|\,\theta);\, 0 \leq \theta \leq 1\}$ の十分統計量であることがわかる．□

例 6.16. $\theta \in \mathbb{R}$ とし, $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}(\theta, 1)$ とする．$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ の同時 p.d.f. は

$$\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta) = \frac{1}{(2\pi)^{n/2}}\exp\left\{-\frac{1}{2}\sum_{j=1}^n (x_j - \theta)^2\right\} \qquad \boldsymbol{x} = (x_1, x_2, \ldots, x_n)$$

である．$T(\boldsymbol{x}) = n^{-1}\sum_{j=1}^n x_j$ とすると

$$\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta) = \exp\left\{nT(\boldsymbol{x})\theta - \frac{n\theta^2}{2}\right\} \times \left(\frac{1}{2\pi}\right)^{n/2} \times \exp\left\{-\frac{1}{2}\sum_{j=1}^n x_j^2\right\}$$

となる．よって (6.5) で

$$g_\theta\bigl\{T(\boldsymbol{x})\bigr\} = \exp\left\{nT(\boldsymbol{x})\theta - \frac{n\theta^2}{2}\right\}, \quad h(\boldsymbol{x}) = \left(\frac{1}{2\pi}\right)^{n/2}\exp\left\{-\frac{1}{2}\sum_{j=1}^n x_j^2\right\}$$

とおけば, 定理 6.13 から $T(\boldsymbol{X}) = (X_1 + X_2 + \cdots + X_n)/n$ は $\mathcal{P} = \{\mathsf{N}^{\otimes n}(\theta, 1);\, \theta \in \mathbb{R}\}$ の十分統計量であることがわかる．ただし, $\mathsf{N}^{\otimes n}(\theta, 1)$ は $\mathsf{N}(\mu, 1)$ の n 個の直積分布である．□

例 6.17. $\nu \in \mathbb{R}$, $0 < \sigma < \infty$ とする．$X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}(\mu, \sigma^2)$ とする．$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ の同時 p.d.f. は

$$\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}(x_j-\mu)^2\right\}$$

となる．ただし $\theta = (\mu, \sigma^2)$ である．ここで

$$\overline{x}_n = \frac{1}{n}\sum_{j=1}^{n} x_j, \qquad s_n^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j - \overline{x})^2$$

とおけば

$$\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{n}{2\sigma^2}\left(s_n^2 + (\overline{x}_n - \mu)^2\right)\right\}$$

と書き直せる．(6.5) において

$$T(\boldsymbol{x}) = (\overline{x}_n,\, s_n^2), \quad g_\theta\{T(\boldsymbol{x})\} = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{n}{2\sigma^2}\left(s_n^2 + (\overline{x}_n - \mu)^2\right)\right\},$$
$$h(\boldsymbol{x}) = 1$$

とすれば, 定理 6.13 から

$$\boldsymbol{T}(\boldsymbol{X}) = \left(\frac{1}{n}\sum_{j=1}^{n} X_j (=: \overline{X}),\ \frac{1}{n}\sum_{j=1}^{n}(X_j - \overline{X})^2\right)$$

は $\mathcal{P} = \{\mathsf{N}^{\otimes n}(\mu, \sigma^2);\ \theta = (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)\}$ の十分統計量である．ただし, $\mathsf{N}^{\otimes n}(\mu, \sigma^2)$ は $\mathsf{N}(\mu, \sigma)$ の n 個の直積分布である． □

6.4 統計量の最小十分性と完備性

与えられた統計的モデル $\mathcal{P}$ に対して, たくさんの十分統計量が存在する．$\boldsymbol{S}$ と $\boldsymbol{T}$ が $\mathsf{P} \in \mathcal{P}$ に対する十分統計量で, S の値域上で定義された可測関数 ψ が存在して

$$\boldsymbol{T} = \psi(\boldsymbol{S})$$

と表現できたとする．すると

$$\sigma(\boldsymbol{T}) \subset \sigma(\boldsymbol{S})$$

となる．このことから $\boldsymbol{T}$ は統計的モデル $\mathcal{P}$ の情報を失くことなく, データの情報を S より縮約していることがわかる．したがって T は S よりも有用な統計量となる．この観点を定式化してみよう．

定義 6.18. すべての $\mathsf{P} \in \mathcal{P}$ に対して, $\mathsf{P}(A) = 0$ となる事象 A を除いて, ある命題 M が成立しているとき

$$M \qquad \text{a.s.}\,\mathcal{P}$$

と記すことにする.

定義 6.19. (最小十分性) $\mathcal{P}$ を統計的モデルとし, 統計量 $\boldsymbol{T}$ を $\mathsf{P} \in \mathcal{P}$ に対する十分統計量とする. 十分統計量 T は $\mathsf{P} \in \mathcal{P}$ に関して最小であるとは, 任意の他の十分統計量 S に対して可測関数 ψ が存在して

$$\boldsymbol{T} = \psi(\boldsymbol{S}), \qquad \text{a.s.}\,\mathcal{P}$$

と表現できるときをいう. もちろん ψ は $\boldsymbol{S}$ ごとに定まればよい.

例 6.20. $\mathcal{P} = \big\{\mathsf{P}$ は開区間 $(\theta, \theta+1)$ 上の一様分布; $\theta \in \mathbb{R}\big\}$ とする. $n \geq 2\,(\in \mathbb{N})$ とし

$$X_1,\, X_2,\, \ldots,\, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{P} \in \mathcal{P}$$

とする. Lebesgue 測度に関する $\boldsymbol{X} := (X_1,\, X_2,\, \ldots,\, X_n)$ の同時 p.d.f. は

$$\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta) = \Pi_{j=1}^{n} \mathbb{1}_{(\theta,\,\theta+1)}(x_j) = \mathbb{1}_{(x_{(n)}-1,\,x_{(1)})}(\theta),$$
$$\boldsymbol{x} = (x_1,\, x_2,\, \ldots,\, x_n) \in \mathbb{R}^n$$

となる. ただし $x_{(1)}$, $x_{(n)}$ は $x_1,\, x_2,\, \ldots,\, x_n$ の最小値と最大値である.

$$X_{(1)} = \min\{X_1,\, X_2,\, \ldots,\, X_n\}, \quad X_{(n)} = \max\{X_1,\, X_2,\, \ldots,\, X_n\}$$

とおくと Fisher-Neyman の因子分解定理から $\boldsymbol{T} = (X_{(1)},\, X_{(n)})$ は $\mathsf{P} \in \mathcal{P}$ に対する十分統計量である.

$$\begin{aligned} &\theta < x_j < \theta + 1 \quad (\forall j \in \{1,\, 2, \ldots,\, n\}) \\ &\quad \Leftrightarrow\ \theta < x_{(1)} < x_{(n)} < \theta + 1 \end{aligned} \tag{6.7}$$

となることに注意する. $\boldsymbol{X} = \boldsymbol{x}$ を観測したとき

$$x_{(1)} = \sup\{\theta \in \mathbb{R};\, \mathsf{p}(\boldsymbol{x}|\,\theta) > 0\}, \quad x_{(n)} = 1 + \inf\{\theta \in \mathbb{R};\, \mathsf{p}(\boldsymbol{x}|\,\theta) > 0\}$$

となる. ここで統計量 $\boldsymbol{S}$ は $\mathsf{P} \in \mathcal{P}$ に対して十分であると仮定する. Fisher-Neyman の因子分解定理から可測関数 g_θ と h が存在して

$$\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta) = g_\theta\big(\boldsymbol{S}(\boldsymbol{x})\big)h(\boldsymbol{x})$$

と書ける. $\{\boldsymbol{x} \in \mathbb{R}^n;\, h(\boldsymbol{x}) > 0\}$ 上で (6.7) が成立するので, ある可測関数 ψ が存在して A 上で

$$T := (x_{(1)},\, x_{(n)}) = \psi\big(S(\boldsymbol{x})\big)$$

と表現できる. さらに $\mathsf{Pr}(\boldsymbol{X} \in A) = 1$ なので $\boldsymbol{T}$ は最小十分統計量となる. □

定義 6.21. $k \in \mathbb{R}$ とし, $\mathbb{T} \subset \mathbb{R}^k$ とする. $\mathcal{P}$ を統計的モデルとし, $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{P} \in \mathcal{P}$ とする. $\boldsymbol{T}(\boldsymbol{X})\,(\in \mathbb{T})$ を $\mathsf{P} \in \mathcal{P}$ に対する十分統計量とする. ただし, $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top$ と書いた. 統計量 $\boldsymbol{T}$ は $\mathsf{P} \in \mathcal{P}$ に対して完備であるとは, 任意の可測関数 $f : \mathbb{T} \to \mathbb{R}$ に対して

$$\mathsf{E}\big[f(\boldsymbol{T})\big] = 0\,(\forall \mathsf{P} \in \mathcal{P}) \;\Rightarrow\; f(\boldsymbol{T}) = 0, \text{ a.s.}\,\mathcal{P}$$

が成り立つときをいう.

次の命題は, 指数型分布族の十分統計量は完備であることを主張している. 証明では, 本講義録では証明していない Fubini の定理と Laplace 変換の一意性を用いている.

命題 6.22. $k, n \in \mathbb{N}$ かつ $n > k$ とする. $\boldsymbol{\eta} \in \mathbb{R}^k$, $\mathbb{X} \subset \mathbb{R}$ を空でない部分集合とし

$$\mathsf{p}(x|\,\boldsymbol{\eta}) = \exp\big\{\boldsymbol{\eta}^\top \boldsymbol{T}\big(x\big)\big\}h(x), \qquad x \in \mathbb{X}$$

とする. ただし

$$\mathcal{E} := \left\{\boldsymbol{\eta} \in \mathbb{R}^k;\, \int_{\mathbb{X}} \exp\{\boldsymbol{\eta}^\top \boldsymbol{T}(x)\}h(x)\,\mathrm{d}\boldsymbol{x} < \infty\right\}$$

は $\mathbb{R}^n$ の開集合を含み

$$\boldsymbol{T} : \mathbb{X}^n \to \mathbb{R}^k, \quad h : \mathbb{X}^k \to \mathbb{R}, \quad K : \mathcal{E} \to \mathbb{R}$$

は可測関数とする. $\boldsymbol{X} \sim \mathsf{P}^{\otimes n}\,(\mathsf{P} \in \mathcal{P} := \{\mathsf{p}(x|\,\boldsymbol{\eta});\, \boldsymbol{\eta} \in \mathcal{E}\})$ としたとき, $\boldsymbol{T}(\boldsymbol{X})$ は $\mathsf{P} \in \mathcal{P}$ に対する完備かつ十分統計量である.

Proof. Fisher-Neyman の因子分解定理から $\boldsymbol{T}$ は $\mathsf{P} \in \mathcal{P}$ に対する十分統計量であることは明らか. つぎに $f : \mathbb{R}^k \to \mathbb{R}$ を可測関数とし

$$\mathsf{E}\big[f(\boldsymbol{T})\big] = 0 \quad (\forall \boldsymbol{\eta} \in \mathcal{E})$$

とする. $\mathsf{E}\big[|f(\boldsymbol{T})|\big] < \infty$ なので Fubini の定理から

$$\begin{aligned} 0 &= \int_{\mathbb{X}} f\big(\boldsymbol{T}(\boldsymbol{x})\big)\exp\big\{\boldsymbol{\eta}^\top \boldsymbol{T}(\boldsymbol{x}) - K(\boldsymbol{\eta})\big\}h(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\ &= \int_{\boldsymbol{x}:\,\boldsymbol{T}(\boldsymbol{x})=\boldsymbol{t}} h(\boldsymbol{x})\left\{\int_{\mathbb{T}} f(\boldsymbol{t})\exp\big\{\boldsymbol{\eta}^\top \boldsymbol{t} - K(\boldsymbol{\eta})\big\}\,\mathrm{d}\boldsymbol{t}\right\}\mathrm{d}\mathsf{m}_{n-k} \end{aligned}$$

と書ける. ただし $\mathbb{T} = \{\boldsymbol{T}(\boldsymbol{x}) \in \mathbb{R}^k;\, \boldsymbol{x} \in \mathbb{X}\}$ で, m_{n-k} は $\mathbb{R}^{n-k}$ 上の Lebeague 測度である. よって

$$\int_{\mathcal{T}} f(\boldsymbol{t})\exp\big\{\boldsymbol{\eta}^\top \boldsymbol{t} - K(\boldsymbol{\eta})\big\}\,\mathrm{d}\boldsymbol{t} = 0 \qquad \text{a.s. } \mathsf{m}_k$$

となる．$\boldsymbol{\eta}_0 \in \mathcal{E}$ の内点とし，$\epsilon > 0$ を十分小さく取ると

$$\int f_+(\boldsymbol{t}) \exp\{\boldsymbol{\eta}^\top \boldsymbol{t}\}\,\mathrm{d}\boldsymbol{t} = \int f_-(\boldsymbol{t}) \exp\{\boldsymbol{\eta}^\top \boldsymbol{t}\}\,\mathrm{d}\boldsymbol{t}$$
$$\big(\forall \boldsymbol{\eta} \in N(\boldsymbol{\eta}_0) := \{\boldsymbol{\eta} \in \mathcal{R}^k;\, |\boldsymbol{\eta} - \boldsymbol{\eta}_0|_{2,k} < \epsilon\big)$$

となる．ここで

$$f_+(\boldsymbol{t}) := \max\{f(\boldsymbol{t}), 0\}, \quad f_-(\boldsymbol{t}) := \max\{-f(\boldsymbol{t}), 0\},$$

で，$|\cdot|_{2,k}$ は $\mathbb{R}^k$ 上の Euclid の距離である．特に

$$\int f_+(\boldsymbol{t}) \exp\{\boldsymbol{\eta}_0^\top \boldsymbol{t}\}\,\mathrm{d}\boldsymbol{t} = \int f_-(\boldsymbol{t}) \exp\{\boldsymbol{\eta}_0^\top \boldsymbol{t}\}\,\mathrm{d}\boldsymbol{t} =: c$$

となる．$c = 0$ のとき，$f(\boldsymbol{t}) = 0\,(\mathrm{a.s.})$ となる．$c > 0$ のとき，Laplace 変換の一意性[6]から

$$f_+(\boldsymbol{t}) = f_-(\boldsymbol{t}), \quad \mathrm{a.s.}$$

となる．よって $f(\boldsymbol{t}) = f_+(\boldsymbol{t}) - f_-(\boldsymbol{t}) = 0$ がわかる． □

注意 6.23. 完備十分統計量は最小十分統計量であることが知られている．しかし最小十分統計量で完備でないものの存在が知られている． □

6.5　章末注釈と参考文献

第 6.1 節は [1, pp.69 − 71] を借用した．補題 6.10 は [14, pp.28 − 30] を借用した．

6.6　演習問題

演習問題 6.1.

演習問題 6.2.

[6]Laplace 変換の一意性の証明は [39, pp.348-349] を参照のこと．

第7章　推定

この章では, 点推定法における推定量の導出原理と推定量の精度評価についての考え方の説明を行う. 第 7.1 節では, モーメント法について説明する. 第 7.2 節では, 最尤法について説明する. 最尤推定量の漸近分布は, 統計的モデルが指数型分布族の場合について示している. 第 7.3 では, 推定量のひとつの性質である不偏推定の概念を説明し, それに関係する情報不等式を与える.

7.1　モーメント法

この方法で得られた推定量は最適性を持たないことがある. しかし求めるのが容易であるので, 陽には式が与えられない別の推定量の値を繰り返し計算で求めるときの初期値として用いることができる.

$\mathbb{X} \subset \mathbb{R}$ を空でない部分集合とする. $d, n \in \mathbb{N}$ とし, 統計的実験を $\big(\mathbb{X}^n, \{\mathsf{P}_{\boldsymbol{\theta}} = \mathsf{P}_{\boldsymbol{\theta}};\, \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d\}\big)$ を考え

$$X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{P}_{\boldsymbol{\theta}^*} \quad (\boldsymbol{\theta}^* \in \Theta)$$

とする. ただし, $\mathsf{P}_{\boldsymbol{\theta}}^{\otimes n}$ と $\mathsf{P}_{\boldsymbol{\theta}}$ は, それぞれ測度空間 $\big(\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n)\big)$ と $\big(\mathbb{X}, \mathcal{B}(\mathbb{X})\big)$ 上の分布である.

$n \geq d$ とする. $1 \leq j \leq d$ に対して

$$\alpha_j := \alpha_j(\boldsymbol{\theta}) := \mathsf{E}\big[X_1^j\big], \quad \widehat{\alpha}_j := \frac{1}{n}\sum_{\ell=1}^n X_\ell^j$$

と定める.

定義 7.1. $n \geq d$ とし $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{P}_{\boldsymbol{\theta}^*}\ \big(\boldsymbol{\theta}^* \in \Theta\big)$ とする. このとき $\boldsymbol{\theta}^*$ の**モーメント法推定量** $\widehat{\boldsymbol{\theta}}_n$ とは次をみたす解 (存在すれば) である.

$$\alpha_1(\widehat{\boldsymbol{\theta}}_n) = \widehat{\alpha}_1,\ \alpha_2(\widehat{\boldsymbol{\theta}}_n) = \widehat{\alpha}_2, \ldots,\ \alpha_d(\widehat{\boldsymbol{\theta}}_n) = \widehat{\alpha}_d \tag{7.1}$$

注意 7.2. $n \geq 2$ とする. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}\big(\mu^*, (\sigma^*)^2\big)$ とする. ただし $\boldsymbol{\theta}^* := (\mu^*, \sigma^*) \in \mathbb{R} \times (0, \infty) =: \Theta$ で (μ^*, σ^*) は未知とする. このとき

$$\alpha_1 = \mathsf{E}\big[X_1\big] = \mu^*, \quad \alpha_2 = \mathsf{E}\big[X_1^2\big] = \mathsf{Var}\big[X_1\big] + \big\{\mathsf{E}\big[X_1\big]\big\}^2 = (\sigma^*)^2 + (\mu^*)^2$$

である．したがって $\widehat{\boldsymbol{\theta}}_n = (\widehat{\mu}_n, \widehat{\sigma}_n)$ とすれば

$$\widehat{\mu}_n = \frac{1}{n}\sum_{\ell=1}^{n} X_\ell, \quad \widehat{\sigma}_n^2 + \{\widehat{\mu}_n\}^2 = \frac{1}{n}\sum_{\ell=1}^{n} X_\ell^2$$

を解くと (μ^*, σ^*) のモーメント法推定量は

$$\widehat{\mu}_n = \frac{1}{n}\sum_{\ell=1}^{n} X_\ell, \quad \widehat{\sigma}_n = \sqrt{\frac{1}{n}\sum_{\ell=1}^{n}(X_\ell - \widehat{\mu}_n)}$$

となる．□

定理 7.3. $d, n \in \mathbb{N}$ は $n \geq d$ とする．

$$X_1, X_2, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{P}_{\boldsymbol{\theta}^*}\,(\boldsymbol{\theta}^* \in \Theta \subset \mathbb{R}^d), \quad \mathsf{E}\big[|X_1|^{2d}\big] < \infty$$

とする．(7.1) で定義した $\alpha_j\,(j = 1, 2, \dots, d)$ の逆写像 α_j^{-1} が存在して，$\boldsymbol{\theta}^*$ の近傍で全微分可能とする．このとき適当な条件のもとで以下が成立する．

(1) $\mathsf{Pr}_{\boldsymbol{\theta}^*}\big(\, \widehat{\boldsymbol{\theta}}_n$ は存在$\big) \overset{n\to\infty}{\longrightarrow} 0$ である．

(2) $\forall\epsilon > 0$ に対して

$$\mathsf{Pr}_{\boldsymbol{\theta}^*}\big(\big|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*\big|_{2,\,d} \geq \epsilon\big) \overset{n\to\infty}{\longrightarrow} 0$$

である．ただし $|\cdot|_{2,\,d}$ は $\mathbb{R}^d$ の Euclid ノルムである．

(3) $\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*\big) \rightsquigarrow \mathsf{N}_d\big(\mathbf{0}, \boldsymbol{\Sigma}\big)$ である．ただし

$$\begin{aligned}
\boldsymbol{\Sigma} &= \boldsymbol{G}\big(\boldsymbol{\theta}^*\big)\mathsf{E}_{\boldsymbol{\theta}^*}\big[\boldsymbol{Y}\boldsymbol{Y}^\top\big]\boldsymbol{G}^\top\big(\boldsymbol{\theta}^*\big), \\
\boldsymbol{G}\big(\boldsymbol{\theta}\big) &= \big(\dot{\boldsymbol{g}}_1(\boldsymbol{\theta}), \dot{\boldsymbol{g}}_2(\boldsymbol{\theta}), \dots, \dot{\boldsymbol{g}}_d(\boldsymbol{\theta})\big), \\
\dot{\boldsymbol{g}}_j(\boldsymbol{\theta}) &= \left(\frac{\partial\alpha_j^{-1}}{\partial\theta_1}(\boldsymbol{\theta}), \frac{\partial\alpha_j^{-1}}{\partial\theta_2}(\boldsymbol{\theta}), \dots, \frac{\partial\alpha_j^{-1}}{\partial\theta_d}(\boldsymbol{\theta})\right)^\top (j = 1, 2, \dots, d), \\
\boldsymbol{Y} &= \big(X_1, X_1^2, \dots, X_1^d\big)^\top
\end{aligned}$$

が成立する．

Proof. 証明は加筆予定である．□

7.2 最尤法

$\mathbb{X}$ を $\mathbb{R}$ の空でない部分集合とし，$d, n \in \mathbb{N}$ とする．統計的実験 $\left(\mathbb{X}^n, \left\{\mathsf{p}(x|\,\boldsymbol{\theta});\, \boldsymbol{\theta} \in \Theta\right\}\right)$ を考える．ただし $\mathsf{p}(x|\,\boldsymbol{\theta})$ は p.d.f.（または p.m.f.）で $\Theta \subset \mathbb{R}^d$ である．$\boldsymbol{\theta}^* \in \Theta$ とし

$$X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{p}(x|\,\boldsymbol{\theta}^*)$$

とする．

定義 7.4. $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ を観測したとき $\boldsymbol{\theta}$ の **尤**(ゆう) **度関数** $\mathsf{lik}_n(\boldsymbol{\theta}|\,\boldsymbol{x})$ を

$$\mathsf{lik}_n\big(\boldsymbol{\theta}|\,\boldsymbol{x}\big) = \prod_{j=1}^{n} \mathsf{p}(x_j|\,\boldsymbol{\theta})$$

で定義し，**対数尤度** $\ell_n(\boldsymbol{\theta}|\,\boldsymbol{x})$ を

$$\ell_n(\boldsymbol{\theta}|\,\boldsymbol{x}) = \log \mathsf{lik}_n(\boldsymbol{\theta}|\,\boldsymbol{x})$$

で定義する．ただし，$\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ である．

注意 7.5. 尤度関数は

$$\mathsf{lik}_n(\,\cdot\,|\,\boldsymbol{x}) :\ \Theta \ni \boldsymbol{\theta} \mapsto \mathsf{lik}_n(\boldsymbol{\theta}|\,\boldsymbol{x}) \in [0, \infty)$$

である． □

関数 $g(x)$ の最大値を取る点を表す集合を

$$\arg\max_{x \in \mathbb{R}} g(x)$$

と書く．たとえば $g(x) = -(x-1)^2$ のとき

$$\arg\max_{0 \le x \le 4\pi} g(x) = \{1\}$$

となる．$g(x) = \sin x$ のとき

$$\arg\max_{0 \le x \le 4\pi} g(x) = \{\pi/2, 5\pi/2\}$$

となる．

定義 7.6. $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ を観測したとき $\boldsymbol{\theta}^*$ の**最尤推定値** (maximum likelihood estimate) を $\mathsf{lik}_n(\boldsymbol{\theta}|\,\boldsymbol{x})$ を最大にする値 $\widehat{\boldsymbol{\theta}}_n(\boldsymbol{x})$ で定義する．すなわち

$$\widehat{\boldsymbol{\theta}}_n(\boldsymbol{x}) \in \arg\max_{\boldsymbol{\theta} \in \Theta} \mathsf{lik}_n(\boldsymbol{\theta}|\,\boldsymbol{x})$$

である．$\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ に $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ を代入したもの $\widehat{\boldsymbol{\theta}}_n(\boldsymbol{X})$ を $\boldsymbol{\theta}^*$ の**最尤推定量** (maximum likelihood estimator=m.l.e.) という．

注意 7.7. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Ber}(\theta^*)$ とする．ただし $(0, 1) =: \Theta \ni \theta^*$ は未知とする．すなわち

$$\mathsf{p}(x|\,\theta^*) = \begin{cases} (\theta^*)^x \bigl(1-\theta^*\bigr)^{1-x} & (x = 0, 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

である．$X_j = x_j\,(j = 1, 2, \ldots, n)$ を観測したとき

$$\mathsf{lik}_n\bigl(\theta|\,\boldsymbol{x}\bigr) = \prod_{j=1}^n \mathsf{p}(x_j|\,\theta) = \prod_{j=1}^n \theta^{x_j}(1-\theta)^{1-x_j} = \theta^{t_n}(1-\theta)^{n-t_n}$$

となる．ただし，$\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ と $t_n = \sum_{j=1}^n x_j$ である．よって対数尤度は

$$\ell_n(\theta|\,\boldsymbol{x}) = t_n \log\theta + (n - t_n)\log(1-\theta) \qquad (0 < \theta < 1)$$

となる．このことから，$0 < t_n < n$ のとき

$$\frac{t_n}{n} \in \arg\max_{\theta\in(0,\,1)} \ell_n(\theta|\,\boldsymbol{x})$$

がわかる．したがって，$0 < t_n < n$ のとき θ^* の最尤推定量は $\widehat{\theta}_n = \dfrac{\sum_{j=1}^n X_j}{n}$ となり，$t_n = 0$ または $t_n = n$ のとき，最尤推定量は存在しない． □

注意 7.8. $n \geq 2\,(n \in \mathbb{N})$ とする．$\boldsymbol{\theta}^* = (\mu^*, \sigma^*) \in \mathbb{R} \times (0, \infty)$ とし，

$$X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}\bigl(\mu^*, (\sigma^*)^2\bigr)$$

とする．$X_j = x_j\,(j = 1, 2, \ldots, n)$ を観測したとき，尤度関数は

$$\begin{aligned}
\mathsf{lik}_n(\mu, \sigma|\,\boldsymbol{x}) &= \prod_{j=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x_j-\mu)^2}{\sigma^2}\right\} \\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^n (x_j-\mu)^2\right\} \\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sigma^n} \exp\left\{-\frac{ns_n^2}{2\sigma^2}\right\} \exp\left\{-\frac{n(\overline{x}_n-\mu)^2}{2\sigma^2}\right\} \qquad (7.2)
\end{aligned}$$

となる．ただし

$$\overline{x}_n = \frac{1}{n}\sum_{j=1}^n x_j, \qquad s_n^2 = \frac{1}{n}\sum_{j=1}^n (x_j - \overline{x}_n)^2, \qquad \sum_{j=1}^n (x_j - \overline{x}_n)^2 \neq 0$$

である. (7.2) の最後の等号は

$$\sum_{j=1}^{n}(x_j-\mu)^2 = ns_n^2 + n(\overline{x}_n-\mu)^2 \tag{7.3}$$

からわかる. 対数尤度は

$$\ell_n(\mu,\,\sigma|\,\boldsymbol{x}) = -n\log\sigma - \frac{ns_n^2}{2\sigma^2} - \frac{n(\overline{x}_n-\mu)^2}{2\sigma^2} + (\text{定数項})$$

となる. よって

$$\begin{cases} \dfrac{\partial\ell_n(\mu,\,\sigma|\,\boldsymbol{x})}{\partial\mu}(\mu,\,\sigma) = -\dfrac{n(\overline{x}_n-\mu)}{2\sigma^2} = 0 \\ \dfrac{\partial\ell_n}{\partial\sigma}(\mu,\,\sigma|\,\boldsymbol{x}) = -\dfrac{n}{\sigma} + \dfrac{ns_n^2}{\sigma^3} + \dfrac{n(\overline{x}_n-\overline{x}_n)^2}{\sigma^3} = 0 \end{cases}$$

を解くと

$$\mu = \overline{x}_n, \qquad \sigma = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\bigl(x_j-\overline{x}_n\bigr)^2}$$

となる. $\ell_n(\mu,\,\sigma)$ の Hessian を求める.

$$\boldsymbol{H} := \begin{pmatrix} \dfrac{\partial\ell_n^2}{\partial\mu^2}\bigl(\overline{x}_n,\,s_n|\,\boldsymbol{x}\bigr) & \dfrac{\partial\ell_n^2}{\partial\mu\partial\sigma}\bigl(\overline{x}_n,\,s_n|\,\boldsymbol{x}\bigr) \\ \dfrac{\partial\ell_n^2}{\partial\partial\sigma\mu}\bigl(\overline{x}_n,\,s_n|\,\boldsymbol{x}\bigr) & \dfrac{\partial\ell_n^2}{\partial\sigma^2}\bigl(\overline{x}_n,\,s_n|\,\boldsymbol{x}\bigr) \end{pmatrix} = \begin{pmatrix} -\dfrac{1}{2s_n^2} & 0 \\ 0 & -\dfrac{2n}{s_n^2} \end{pmatrix} \tag{7.4}$$

より, $-\boldsymbol{H}$ は正定値行列となる. したがって, 関数

$$\Theta \ni (\mu,\,\sigma) \mapsto \ell_n(\mu,\,\sigma)$$

は $(\mu,\,\sigma) = (\overline{x}_n,\,s_n)$ で最大となる. 以上の議論から $(\mu^*,\,\sigma^*)$ の最尤推定量は

$$\widehat{\mu}_n = \frac{1}{n}\sum_{j=1}^{n}X_j =: \overline{X}_n, \quad \widehat{s}_n = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\bigl(X_j-\overline{X}_n\bigr)^2}$$

となる. □

問 7.1. (7.3) と (7.4) を確認せよ.

定理 7.9.

$$X_1,\,X_2,\,\ldots,\,X_n \overset{\text{i.i.d.}}{\sim} \mathsf{p}(x|\,\eta) = h(x)\exp\{\eta^*T(x)-\kappa^\vee(\eta^*)\}\,(\eta^*\in\mathcal{E}^o)$$

とする. ただし $\mathcal{E}^o$ は自然母数空間 $\mathcal{E} \subset \mathbb{R}$ の内部である. さらに $\ddot{\kappa}^\vee(\eta^*) > 0\,(\eta \in \mathcal{E}^o)$ を仮定する. このとき η^* の最尤推定量 $\widehat{\eta}$ は確率 1 で一意的に存在して

$$
\begin{gathered}
\widehat{\eta} \xrightarrow{\text{a.s.}} \eta^*, \\
\sqrt{n}\big(\widehat{\eta} - \eta^*\big) \rightsquigarrow \mathsf{N}\big(0,\, \big(\ddot{\kappa}^\vee\big)^{-1}(\eta^*)\big)
\end{gathered}
$$

が成立する.

Proof. $X_1 = x_1,\, X_2,\, \ldots,\, X_n = x_n$ を観測したときの尤度関数 lik_n は

$$
\mathsf{lik}_n(\eta|\,\boldsymbol{x}) = \left\{\prod_{j=1}^n h(x_j)\right\} \exp\Big[\eta \sum_{j=1}^n T(x_j) - n\kappa^\vee(\eta)\Big]
$$

となる. このことから対数尤度関数は

$$
\ell_n(\eta|\,\boldsymbol{x}) = \log\big(\mathsf{lik}_n(\eta|\,\boldsymbol{x})\big) = n\big\{\eta\overline{T} - \kappa^\vee(\eta)\big\} + (\text{constant})
$$

となる. ただし $\overline{T} = n^{-1}\sum_{j=1}^n T(x_j)$ である. したがって

$$
\dot{\ell}_n(\eta|\,\boldsymbol{x}) = \frac{d\ell_n}{d\eta}(\eta|\,\boldsymbol{x}) = n\big(\overline{T} - \dot{\kappa}^\vee\big) = 0 \;\Leftrightarrow\; \overline{T} = \dot{\kappa}^\vee(\eta) = \mathsf{E}[T(X_1)]
$$

となる. 最後の等号は (6.4) からわかる.

以後, 対数尤度関数の $x_j\,(j = 1, 2, \ldots, n)$ のところに X_j を代入して, 確率変数にしたものを考える. $T(X_1)$ は有限の期待値をもつので, 大数の法則 (定理 4.12) から

$$
\overline{T} = \frac{1}{n}\sum_{j=1}^n T(X_j) \xrightarrow{\text{a,s.}} \mathsf{E}[T(X_1)] = \dot{\kappa}^\vee(\eta^*) \quad (n \to \infty)
$$

がわかる. したがって $\overline{T}$ は関数 $\mathcal{E}^o \ni \eta \mapsto \dot{\kappa}^\vee(\eta)$ の値域に含まれる. さらに $\ddot{\kappa}(\eta) > 0\,(\eta \in \mathcal{E}^o)$ なので関数 $\mathcal{E}^o \ni \eta \mapsto \dot{\kappa}^\vee(\eta)$ は狭義単調増加関数である. このことより

$$
\overline{T} = \dot{\kappa}^\vee(\eta)
$$

をみたす $\eta \in \mathcal{E}^o$ が存在する. したがって

$$
\widehat{\eta} = \big(\dot{\kappa}^\vee\big)^{-1}\big(\overline{T}\big)
$$

は一意的に定まる. さらに

$$
\overline{T} \xrightarrow{\text{a.s.}} \dot{\kappa}^\vee(\eta^*) \quad \text{かつ} \quad \big(\dot{\kappa}^\vee\big)^{-1}(\eta)\ \text{は}\ \eta^*\ \text{の近傍で連続}
$$

なので定理 4.11(6) から

$$
\widehat{\eta} = \big(\dot{\kappa}^\vee\big)^{-1}\big(\overline{T}\big) \xrightarrow{\text{a.s.}} \big(\dot{\kappa}^\vee\big)^{-1}\big(\dot{\kappa}^\vee(\eta^*)\big) = \eta^*
$$

がわかる．一方, (6.4) から $\mathsf{E}[T(X_1)] = \dot{\kappa}^\vee(\eta^*)$, $\mathsf{Var}[T(X_1)] = \ddot{\kappa}^\vee(\eta^*)$ となることに注意して, 中心極限定理 (定理 4.17) を用いると

$$\sqrt{n}\big(\overline{T} - \dot{\kappa}^\vee(\eta^*)\big) \rightsquigarrow \mathsf{N}\big(0,\, \ddot{\kappa}^\vee(\eta^*)\big)$$

がわかる．以上のことを踏まえて, $\widehat{\eta} = \big(\dot{\kappa}^\vee\big)^{-1}\big(\overline{T}\big)$ に対して, 陰関数の定理とデルタ法 (定理 4.22) を適用すると

$$\sqrt{n}\big(\widehat{\eta} - \eta^*\big) \rightsquigarrow \mathsf{N}\big(0,\, \big(\ddot{\kappa}^\vee\big)^{-1}(\eta^*)\big)$$

がわかる． □

7.3　不偏推定と情報不等式

$\mathbb{X} \subset \mathbb{R}$ を空でない部分集合とする．$(\Omega, \mathcal{A}, \mathsf{Pr})$ を確率空間とし, $X_1, X_2, \ldots, X_n$ をこの空間上の独立同一分布に従う $\mathbb{X}$ 値確率変数列 (ランダム標本) とする．また $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top$ と記し, $\boldsymbol{X}$ の値域 (標本空間) を $\mathbb{X}^n \subset \mathbb{R}^n$ と表記する．さらに

$$\mathsf{P} = \mathsf{Pr} \circ X_1^{-1}, \quad \mathsf{P}^{\otimes n} = \mathsf{Pr} \circ \boldsymbol{X}^{-1}$$

とおく．すると $\mathsf{P}^{\otimes n} = \underbrace{\mathsf{P} \times \mathsf{P} \times \cdots \times \mathsf{P}}_{n\ 個}$ となっている．

$\mathcal{P} = \{\mathsf{P}_\theta;\, \theta \in \Theta \subset \mathbb{R}\}$ を $\mathbb{X}$ 上の正則母数モデルとする．以下では, $\Theta \subset \mathbb{R}$ とし, P_θ に関する期待値と分散を $\mathsf{E}_\theta[\,\cdot\,]$, $\mathsf{Var}_\theta[\,\cdot\,]$ と表記する．

定義 7.10. $\boldsymbol{X} := (X_1, X_2, \ldots, X_n)^\top \sim \mathsf{P}_\theta^{\otimes n}\,(\theta \in \Theta \subset \mathbb{R})$ とする．統計量 $T(\boldsymbol{X})$ が $\forall\theta \in \Theta$ に対して

$$\mathsf{E}_\theta\big[T(\boldsymbol{X})\big] = \theta$$

をみたすとき $T(\boldsymbol{X})$ は θ の**不偏推定量** (unbiased estimator) という．

例 7.11. $n \geq 2$ とし, $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}(\mu, 1)\,(\mu \in \mathbb{R})$ とする．このとき $\overline{X}_n = \dfrac{1}{n}\sum_{j=1}^n X_j$ は μ の不偏推定量である． □

定理 7.12 (Rao-Blackwell の定理)**.** $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top \sim \mathsf{P}_\theta^{\otimes n}\,(\theta \in \Theta \subset \mathbb{R})$ で $T = T(\boldsymbol{X})$ は θ の十分統計量とする．$\widehat{\theta}(\boldsymbol{X})$ は θ の任意の不偏推定量で, 有限の分散 $\mathsf{Var}_\theta\big[\widehat{\theta}(\boldsymbol{X})\big]$ をもつものとする．推定量 $\widetilde{\theta}(T)$ を

$$\widetilde{\theta}(T) := \mathsf{E}_\theta\big[\widehat{\theta}(\boldsymbol{X}) | T\big]$$

により定めたとき, 次の (1), (2) が成立する．

(1) $\widetilde{\theta}(T)$ は θ の不偏推定量である.

(2) $\mathsf{Var}_\theta\big[\widetilde{\theta}(T)\big] \le \mathsf{Var}_\theta\big[\widehat{\theta}(\boldsymbol{X})\big]$ $(\forall\theta\in\Theta)$ が成立する.

Proof. (1) T は十分統計量なので, $\mathsf{E}_\theta\big[\widehat{\theta}\big|\,T\big]$ は θ に依存しないので $\widetilde{\theta}(T)$ は推定量となる. また定理 2.27(2) より

$$\mathsf{E}_\theta\big[\widetilde{\theta}(T)\big] = \mathsf{E}_\theta\big[\mathsf{E}_\theta\big[\widehat{\theta}(\boldsymbol{X})\big|\,T\big]\big] = \mathsf{E}_\theta\big[\widehat{\theta}(\boldsymbol{X})\big] = \theta \qquad (\theta\in\Theta)$$

となり $\widetilde{\theta}(T)$ は θ の不偏推定量であることが示せた.
(2) 定理 2.27(4) より

$$\begin{aligned}
\mathsf{E}_\theta\big[\big(\widehat{\theta}(\boldsymbol{X})-\widetilde{\theta}(T)\big)\widetilde{\theta}(T)\big|T\big] &= \widetilde{\theta}(T)\mathsf{E}_\theta\big[\widehat{\theta}(\boldsymbol{X})-\widetilde{\theta}(T)|T\big]\\
&= \widetilde{\theta}(T)\Big(\mathsf{E}_\theta\big[\widehat{\theta}(\boldsymbol{X})|T\big]-\widetilde{\theta}(T)\Big)\\
&= \widetilde{\theta}(T)\big(\widetilde{\theta}(T)-\widetilde{\theta}(T)\big) = 0
\end{aligned}$$

となり

$$\mathsf{E}_\theta\big[\big(\widehat{\theta}(\boldsymbol{X})-\widetilde{\theta}(T)\big)\widetilde{\theta}(T)\big] = \mathsf{E}_\theta\big[\mathsf{E}_\theta\big[\big(\widehat{\theta}(\boldsymbol{X})-\widetilde{\theta}(T)\big)\widetilde{\theta}(T)\big|T\big]\big] = 0 \tag{7.5}$$

を得る. 一方

$$\begin{aligned}
\mathsf{Var}_\theta\big[\widehat{\theta}(\boldsymbol{X})\big] &= \mathsf{E}_\theta\big[\big\{\widehat{\theta}(\boldsymbol{X})-\mathsf{E}_\theta\big[\widehat{\theta}(\boldsymbol{X})\big]\big\}^2\big]\\
&= \mathsf{E}_\theta\big[\big\{\widehat{\theta}(\boldsymbol{X})-\theta\big\}^2\big]\\
&= \mathsf{E}_\theta\big[\big\{\big(\widehat{\theta}(\boldsymbol{X})-\widetilde{\theta}(T)\big)+\big(\widetilde{\theta}(T)-\theta\big)\big\}^2\big]\\
&= \mathsf{E}_\theta\big[\big\{\widehat{\theta}(\boldsymbol{X})-\widetilde{\theta}(T)\big\}^2\big]+\mathsf{E}_\theta\big[\big\{\widetilde{\theta}(T)-\theta\big\}^2\big]\\
&\quad+2\mathsf{E}_\theta\big[\big(\widehat{\theta}(\boldsymbol{X})-\widetilde{\theta}(T)\big)\big(\widetilde{\theta}(T)-\theta\big)\big]
\end{aligned}$$

である. しかし, (7.5) および $\widehat{\theta}(\boldsymbol{X})$ と $\widetilde{\theta}(T)$ は θ の不偏推定量であることに注意すると

$$\begin{aligned}
&\mathsf{E}_\theta\big[\big(\widehat{\theta}(\boldsymbol{X})-\widetilde{\theta}(T)\big)\big(\widetilde{\theta}(T)-\theta\big)\big]\\
&\quad= \underbrace{\mathsf{E}_\theta\big[\big(\widehat{\theta}(\boldsymbol{X})-\widetilde{\theta}(T)\big)\widetilde{\theta}(T)\big]}_{=0 \quad \because(7.5)} - \theta\underbrace{\big\{\mathsf{E}_\theta\big[\widehat{\theta}(\boldsymbol{X})\big]-\mathsf{E}_\theta\big[\widetilde{\theta}(T)\big]\big\}}_{=0}\\
&\quad= 0
\end{aligned}$$

がわかる. 以上から

$$\begin{aligned}
\mathsf{Var}_\theta\big[\widehat{\theta}(\boldsymbol{X})\big] &= \mathsf{E}_\theta\big[\big\{\widehat{\theta}(\boldsymbol{X})-\widetilde{\theta}(T)\big\}^2\big]+\mathsf{Var}_\theta\big[\widetilde{\theta}(T)\big]\\
&\ge \mathsf{Var}_\theta\big[\widetilde{\theta}(T)\big] \qquad (\theta\in\Theta)
\end{aligned}$$

を得る. □

定義 7.13. $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top \sim \mathsf{P}_\theta^{\otimes n}\,(\theta \in \Theta \subset \mathbb{R})$ とする. θ の任意の不偏推定量 $\widehat{\theta}(\boldsymbol{X})$ に対して

$$\mathsf{Var}_\theta\big[\widehat{\theta}_0(\boldsymbol{X})\big] \le \mathsf{Var}_\theta\big[\widehat{\theta}(\boldsymbol{X})\big] \qquad (\theta \in \Theta)$$

をみたす θ の不偏推定量 $\widehat{\theta}_0(\boldsymbol{X})$ が存在するとき $T_0(\boldsymbol{X})$ を θ の**一様最小分散不偏推定量** (uniformly minimum variance unbiased estimator: UMVUE) という.

可測関数 $g : \mathbb{R} \to \mathbb{R}$ は連続微分可能とする. $\boldsymbol{X} \sim \mathsf{P}_\theta^{\otimes n}$ に基づき $g(\theta)$ の推定問題を考える.

定理 7.14. $\{\mathsf{P}_\theta; \theta \in \Theta \subset R\}$ は正則母数モデルとする. $\boldsymbol{X} \sim \mathsf{P}_\theta^{\otimes n}\,(\theta \in \Theta \subset \mathbb{R})$ とし $\boldsymbol{X}$ は同時 p.d.f. または p.m.f. $\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)$ をもつとする. $\widehat{\theta} = \widehat{\theta}(\boldsymbol{X})$ を $g(\theta)$ の任意の不偏推定量とし

$$\begin{aligned} A(\widetilde{\theta}, \theta) &:= \mathsf{Var}_\theta\left[\frac{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{X}|\,\widetilde{\theta})}{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{X}|\,\theta)}\right] \qquad (\widetilde{\theta}, \theta \in \Theta) \qquad (7.6)\\ A(\widetilde{\theta}, \theta) &> 0 \qquad (\widetilde{\theta} \in \Theta, \widetilde{\theta} \ne \theta) \end{aligned}$$

とおく. このとき

$$\mathsf{Var}_\theta\big[\widehat{\theta}(\boldsymbol{X})\big] \ge \sup_{\widetilde{\theta} \in \Theta} \frac{\{g(\widetilde{\theta}) - g(\theta)\}^2}{A(\widetilde{\theta}, \theta)}$$

が成り立つ.

Proof. $\boldsymbol{X}$ は連続型確率変数の場合の証明を与える. $\boldsymbol{X}$ の値域を $\mathbb{X}^n$ と書く. また, 離散型の場合は, 積分記号を和の記号に変更すればよい. 推定量 $\widehat{\theta}$ は $g(\theta)$ に対して不偏なので

$$\mathsf{E}_\theta\big[\widehat{\theta}(\boldsymbol{X})\big] = \int_{\mathbb{R}^n} \widehat{\theta}(\boldsymbol{x})\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)\,\mathrm{d}\boldsymbol{x} = g(\theta) \qquad (\theta \in \Theta)$$

となる. よって

$$\begin{aligned} \mathsf{E}_\theta\left[\widehat{\theta}(\boldsymbol{X})\left\{\frac{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{X}|\,\widetilde{\theta})}{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{X}|\,\theta)} - 1\right\}\right] &= \int_{\mathbb{X}^n} \widehat{\theta}(\boldsymbol{x})\left\{\frac{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\widetilde{\theta})}{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)} - 1\right\}\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)\,\mathrm{d}\boldsymbol{x}\\ &= \int_{\mathbb{X}^n} \widehat{\theta}(\boldsymbol{x})\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\widetilde{\theta})\,d\boldsymbol{x} - \int_{\mathbb{X}^n} \widehat{\theta}(\boldsymbol{x})\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)\,\mathrm{d}\boldsymbol{x}\\ &= g(\widetilde{\theta}) - g(\theta) \qquad (7.7) \end{aligned}$$

となる. 一方

$$\mathsf{E}_\theta\left[\frac{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{X}|\,\widetilde{\theta})}{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{X}|\,\theta)}\right] = \int_{\mathbb{X}^n} \frac{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\widetilde{\theta})}{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)}\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)\,\mathrm{d}\boldsymbol{x} = \int_{\mathbb{X}^n} \mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\widetilde{\theta})\,\mathrm{d}\boldsymbol{x} = 1 \quad (7.8)$$

と (7.7) に注意すると

$$
\begin{aligned}
g(\widetilde{\theta})-g(\theta) &= \mathsf{E}_\theta\left[\widehat{\theta}(\boldsymbol{X})\left\{\frac{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{X}|\,\widetilde{\theta})}{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{X}|\,\theta)}-1\right\}\right] \\
&= \mathsf{E}_\theta\left[\left\{\widehat{\theta}(\boldsymbol{X})-\mathsf{E}_\theta\big[\widehat{\theta}(\boldsymbol{X})\big]+\mathsf{E}_\theta\big[\widehat{\theta}(\boldsymbol{X})\big]\right\}\left\{\frac{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{X}|\,\widetilde{\theta})}{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{X}|\,\theta)}-1\right\}\right] \\
&= \mathsf{E}_\theta\left[\left\{\widehat{\theta}(\boldsymbol{X})-\mathsf{E}_\theta\big[\widehat{\theta}(\boldsymbol{X})\big]\right\}\left\{\frac{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{X}|\,\widetilde{\theta})}{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{X}|\,\theta)}-1\right\}\right] \\
&\qquad + \mathsf{E}_\theta\big[\widehat{\theta}(\boldsymbol{X})\big]\underbrace{\mathsf{E}_\theta\left[\frac{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{X}|\,\widetilde{\theta})}{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{X}|\,\theta)}-1\right]}_{=0} \\
&= \mathsf{E}_\theta\left[\big\{\widehat{\theta}(\boldsymbol{X})-\mathsf{E}_\theta\big[\widehat{\theta}(\boldsymbol{X})\big]\big\}\left\{\frac{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{X}|\,\widetilde{\theta})}{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{X}|\,\theta)}-1\right\}\right]
\end{aligned}
$$

を得る．Cauchy-Schwarz の不等式 (定理 3.10) を上の式の最右辺に適用すると

$$
\begin{aligned}
\big|g(\widetilde{\theta})-g(\theta)\big| &= \left|\mathsf{E}_\theta\left[\widehat{\theta}(\boldsymbol{X})\left\{\frac{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{X}|\,\widetilde{\theta})}{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)}-1\right\}\right]\right| \\
&\le \sqrt{\big\{\mathsf{E}_\theta\big[\widehat{\theta}(\boldsymbol{X})-\mathsf{E}_\theta\big[\widehat{\theta}(\boldsymbol{X})\big]\big\}^2\big]}\sqrt{\mathsf{E}_\theta\left[\left\{\frac{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{X}|\,\widetilde{\theta})}{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)}-1\right\}^2\right]} \\
&= \sqrt{\mathsf{Var}_\theta\big[\widehat{\theta}(\boldsymbol{X})\big]}\sqrt{\mathsf{Var}_\theta\left[\frac{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{X}|\,\widetilde{\theta})}{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)}\right]} \qquad (7.9)
\end{aligned}
$$

を得る．最後の等号は (7.8) を用いた．したがって

$$
\mathsf{Var}_\theta\big[\widehat{\theta}(\boldsymbol{X})\big]\mathsf{Var}_\theta\left[\frac{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{X}|\,\widetilde{\theta})}{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)}\right] \ge \{g(\widetilde{\theta})-g(\theta)\}^2 \qquad (7.10)
$$

を得る．したがって (7.6) と (7.10) から $\forall\theta\in\Theta$ に対して

$$
\mathsf{Var}_\theta\big[T(\boldsymbol{X})\big] \ge \frac{\{g(\widetilde{\theta})-g(\theta)\}^2}{A\big(\widetilde{\theta},\,\theta\big)}
$$

となる．上式の $\widetilde{\theta}$ は任意だったので, 上式の左辺において $\widetilde{\theta}$ に関して sup を取ると定理の主張はわかる．□

定理 7.15. (Cramér-Rao の不等式) 次の条件を仮定する．

(1) $\{\mathsf{P}_\theta;\,\theta\in\Theta\subset R\}$ は正則母数モデルとする．

(2) $g:\mathbb{R}\longrightarrow\mathbb{R}$ は微分可能でその導関数を $\dot{g}(\theta)$ としたとき

$$
\lim_{\widetilde{\theta}\to\theta}\frac{A\big(\widetilde{\theta},\,\theta\big)}{(\widetilde{\theta}-\theta)^2} = J(\theta) > 0
$$

が存在する. ただし, $A\bigl(\widetilde{\theta},\theta\bigr)$ は (7.6) で与えたものである.

(3) $\forall\theta\in\Theta$ に対して十分小さな $\epsilon>0$ をとると任意の $\widetilde{\theta}:\ \bigl|\widetilde{\theta}-\theta\bigr|<\epsilon$ に対して, ある関数 $G(\boldsymbol{x}|\,\theta)$ が存在して

$$\left|\frac{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\widetilde{\theta})-\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)}{(\widetilde{\theta}-\theta)\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)}\right|<G(\boldsymbol{x}|\,\theta)\quad\text{かつ}\quad\mathsf{E}_\theta\bigl[G^2(\boldsymbol{X}|\,\theta)\bigr]<\infty$$

をみたす.

このとき $g(\theta)$ の任意の不偏推定量 $T(\boldsymbol{X})$ に対して

$$\mathsf{Var}_\theta\bigl[\widehat{\theta}(\boldsymbol{X})\bigr]\geq\frac{\dot{g}(\theta)}{\mathcal{F}_{\boldsymbol{X}}(\theta)}\qquad(\theta\in\Theta)\tag{7.11}$$

が成り立つ. ただし

$$\mathcal{F}_{\boldsymbol{X}}(\theta)=\mathsf{E}_\theta\left[\left\{\frac{\partial}{\partial\theta}\log\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{X}|\,\theta)\right\}^2\right]$$

である.

注意 7.16. (7.11) を Cramér-Rao の不等式といい, その右辺を Cramér-Rao の下限という. □

定理 7.15 **の証明**: $\boldsymbol{X}$ の値域を $\mathbb{X}^n$ と書く. 定理 7.14 と仮定 (2) から

$$\mathsf{Var}_\theta\bigl[\widehat{\theta}(\boldsymbol{X})\bigr]\geq\lim_{\widetilde{\theta}\to\theta}\frac{\left\{\dfrac{g(\widetilde{\theta})-g(\theta)}{\widetilde{\theta}-\theta}\right\}^2}{\dfrac{A\bigl(\widetilde{\theta},\theta\bigr)}{(\widetilde{\theta}-\theta)^2}}=\frac{\{\dot{g}(\theta)\}^2}{J(\theta)}\tag{7.12}$$

を得る. 次に仮定 (3) に注意して Lebeague の優収束定理を用いると

$$\begin{aligned}J(\theta)&=\lim_{\widetilde{\theta}\to\theta}\frac{A\bigl(\widetilde{\theta},\theta\bigr)}{\bigl(\widetilde{\theta}-\theta\bigr)^2}\\&=\lim_{\widetilde{\theta}\to\theta}\int_{\mathbb{X}^n}\left\{\frac{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\widetilde{\theta})}{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)}-1\right\}^2\frac{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)}{\bigl(\widetilde{\theta}-\theta\bigr)^2}\,\mathrm{d}\boldsymbol{x}\\&=\int_{\mathbb{X}^n}\left\{\lim_{\widetilde{\theta}\to\theta}\frac{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\widetilde{\theta})-\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)}{\bigl(\widetilde{\theta}-\theta\bigr)\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)}\right\}^2\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)\,\mathrm{d}\boldsymbol{x}\\&=\int_{\mathbb{X}^n}\left\{\frac{\partial}{\partial\theta}\log\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)\right\}^2\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)\,\mathrm{d}\boldsymbol{x}=\mathcal{F}_{\boldsymbol{X}}(\theta)\end{aligned}$$

となる．これと (7.12) を合わせると

$$\mathsf{Var}_\theta\big[\widehat{\theta}(\boldsymbol{X})\big] \geq \frac{\{\dot{g}(\theta)\}^2}{\mathcal{F}_{\boldsymbol{X}}(\theta)} \tag{7.13}$$

を得る．

□

注意 7.17. 指数型分布族が定理 7.15 の正則条件 (2) と (3) をみたしているかを確認する．

以下, 加筆をすること．

□

定義 7.18. $\boldsymbol{X} \sim \mathsf{P}_\theta^{\otimes n}\,(\theta \in \Theta \subset \mathbb{R})$ とする．θ の不偏推定量 $\widehat{\theta}$ は R 有効[1]であるとは

$$\mathsf{Var}_\theta\big[\widehat{\theta}\big] = \big[\mathcal{F}_{\boldsymbol{X}}(\theta)\big]^{-1} \qquad (\theta \in \Theta)$$

をみたすときをいう．

定理 7.19. $\boldsymbol{X} = (X_1,\, X_2,\, \ldots,\, X_n) \sim \mathsf{P}_\theta^{\otimes n}\,(\theta \in \Theta \subset \mathbb{R})$ とする．θ の不偏推定量 $\widehat{\theta}$ が R 有効であるための必要十分条件は $\mathbb{R}$ 上の $\mathbb{R}$ 値関数 $T(x)$, $A(\theta)$, $\kappa(\theta)$, $g(x)$ があって

$$\widehat{\theta}_n(\boldsymbol{X}) = \frac{1}{n}\sum_{j=1}^{n} T(X_j) \tag{7.14}$$

$$\log \mathsf{p}(x|\,\theta) = A(\theta)T(x) - \kappa(\theta) + g(x), \tag{7.15}$$

$$\int_{\mathbb{X}} T(x)\mathsf{p}(x|\,\theta)\,\mathrm{d}x = \theta \tag{7.16}$$

をみたすときである．ただし X_1 の p.d.f.(または p.m.f.) と値域をそれぞれ $\mathsf{p}(x|\,\theta)$ と $\mathbb{X}(\subset \mathbb{R})$ と書いた．

Proof. 連続型分布に対する証明を与える．離散型分布に対しては, 積分記号を和の記号に替えればよい．表現 (7.14) – (7.16) が成立したとする．$\ell_n(\theta|\,\boldsymbol{x}) = \sum_{j=1}^n \log \mathsf{p}(x_j|\theta)\,\big(\boldsymbol{x} = (x_1,\, x_2,\, \ldots,\, x_n)\big)$ とおく．すると

$$\dot{\ell}_n(\theta|\,\boldsymbol{x}) = \frac{\partial}{\partial\theta}\ell_n(\theta|\,\boldsymbol{x}) = \dot{A}(\theta)\sum_{j=1}^{n} T(x_j) - n\dot{\kappa}(\theta),$$

$$\dot{A}(\theta) = \frac{\partial A}{\partial\theta}(\theta), \quad \dot{\kappa}(\theta) = \frac{\partial B}{\partial\theta}(\theta)$$

[1]Rao 有効の意味であろう．

となる．よって

$$\mathsf{E}_\theta\big[\dot{\ell}_n(\theta|\,\boldsymbol{X})\big] = 0 \Leftrightarrow \dot{A}(\theta)\mathsf{E}_\theta\big[T(X_1)\big] = \dot{\kappa}(\theta)$$

と (7.16) とあわせると

$$\theta\dot{A}(\theta) = \dot{\kappa}(\theta) \tag{7.17}$$

となる．さらに

$$0 = \int_{\mathbb{X}^n}\{T(x)-\theta\}\mathsf{p}(x|\,\theta)\,\mathrm{d}x = \int_{\mathbb{X}^n}\{T(x)-\theta\}\exp\{\log\mathsf{p}(x|\,\theta)\}\,\mathrm{d}x$$

を (7.15) と 7.17) に注意して, θ に関して微分すると

$$\begin{aligned}
0 &= -\int_{\mathbb{X}}\mathsf{p}_1(x|\,\theta)\,\mathrm{d}x + \int_{\mathbb{X}}\{T(x)-\theta\}\big\{\dot{A}(\theta)T(x) - \underbrace{\dot{\kappa}(\theta)}_{=\theta\dot{A}(\theta)}\big\}\mathsf{p}(x|\,\theta)\,\mathrm{d}x \\
&= -\int_{\mathbb{X}}\mathsf{p}(x|\,\theta)\,\mathrm{d}x + \int_{\mathbb{X}}\{T(x)-\theta\}\big\{\dot{A}(\theta)T(x) - \theta\dot{A}(\theta)\big\}\mathsf{p}(x|\,\theta)\,\mathrm{d}x
\end{aligned}$$

となることがわかる．よって, $\int_{\mathbb{X}}\mathsf{p}(x|\,\theta)\,\mathrm{d}x = 1$ に注意して, 上式を整理すると

$$1 = \dot{A}(\theta)\mathsf{E}_\theta\big[\{T(X_1)-\theta\}^2\big]$$

を得る．さらに $\mathsf{E}_\theta[T(X_1)] = \theta$ に注意すれば

$$\mathsf{Var}_\theta\big[T(X_1)\big] = \frac{1}{\dot{A}(\theta)} \tag{7.18}$$

となる．したがって

$$\begin{aligned}
\mathcal{F}_{\boldsymbol{X}}(\theta) &= \mathsf{Var}_\theta\big[\dot{\ell}_n(\theta|\,\boldsymbol{X})\big] \qquad (7.19)\\
&= \mathsf{Var}_\theta\Bigg[\sum_{j=1}^n\{\dot{A}(\theta)T(X_j) - \dot{\kappa}(\theta)\}\Bigg] \\
&= \mathsf{Var}_\theta\Bigg[\sum_{j=1}^n\dot{A}(\theta)\big\{T(X_j)\big\}\Bigg] \qquad (\because \text{分散は平行移動に関して不変}) \\
&= \sum_{j=1}^n\{\dot{A}(\theta)\}^2\mathsf{Var}\big[T(X_j)\big] \qquad (\because X_1,\,X_2,\,\ldots,\,X_n \text{ は独立}) \\
&= n\{\dot{A}(\theta)\}^2\mathsf{Var}\big[T(X_1)\big] \qquad (7.20)\\
&= n\dot{A}(\theta) \qquad (\because (7.18))
\end{aligned}$$

を得る．また, $\widehat{\theta}_n(\boldsymbol{X}) = \dfrac{1}{n}\sum_{j=1}^n T(X_j)$ なので

$$\mathsf{E}_\theta\big[\widehat{\theta}_n(\boldsymbol{X})\big] = \theta$$

となり $\widehat{\theta}_n$ は θ の不偏推定量である．再度, (7.18) に注意すると

$$\mathsf{Var}_\theta\big[\widehat{\theta}_n\big] = \mathsf{Var}_\theta\left[\frac{1}{n}\sum_{j=1}^n T(X_j)\right] = \frac{1}{n^2}\sum_{j=1}^n \mathsf{Var}_\theta\big[T(X_j)\big] = \frac{1}{n\dot{A}(\theta)} = \frac{1}{\mathcal{F}_{\boldsymbol{X}}(\theta)}$$

となり $\widehat{\theta}_n$ は R 有効推定量となる．

次に逆を示す．P_θ の同時 p.d.f. $\prod_{j=1}^n \mathsf{p}(x_j|\,\theta)$ に対して, その対数尤度関数を $\ell_n(\theta|\,\boldsymbol{x}) = \log\Big(\prod_{j=1}^n \mathsf{p}(x_j|\,\theta)\Big)$ と書くことにする．すると

$$\int_{\mathbb{X}^n} \exp\{\ell_n(\theta|\,\boldsymbol{X})\}\, d\boldsymbol{x} = 1$$

である．この式を θ に関して微分すると

$$0 = \int_{\mathbb{X}^n} \dot{\ell}_n(\theta|\,\boldsymbol{x}) \exp\{\ell_n(\theta|\,\boldsymbol{x})\}\, \mathrm{d}\boldsymbol{x} = \mathsf{E}_\theta\big[\dot{\ell}_n(\boldsymbol{x}|\,\boldsymbol{X})\big] \tag{7.21}$$

となる．同様に $\mathsf{E}_\theta\big[\widehat{\theta}_n(\boldsymbol{X})\big] = \theta$ より

$$\int_{\mathbb{X}^n} \widehat{\theta}_n(\boldsymbol{x}) \exp\{\ell_n(\theta|\,\boldsymbol{x})\}\, \mathrm{d}\boldsymbol{x} = \theta$$

となる．この式を θ に関して微分すると

$$1 = \int_{\mathbb{X}^n} \widehat{\theta}_n(\boldsymbol{x})\dot{\ell}_n(\theta|\,\boldsymbol{x}) \exp\{\ell_n(\theta|\,\boldsymbol{x})\}\, d\boldsymbol{x} = \mathsf{E}_\theta\big[\widehat{\theta}_n(\boldsymbol{X})\dot{\ell}_n(\theta|\,\boldsymbol{X})\big] \tag{7.22}$$

を得る．(7.21) と (7.22) を合わせると

$$\begin{aligned}\mathsf{E}_\theta\big[(\widehat{\theta}_n(\boldsymbol{X}) - \theta)\dot{\ell}_n(\theta|\,\boldsymbol{X})\big] &= \underbrace{\mathsf{E}_\theta\big[\widehat{\theta}_n(\boldsymbol{X})\dot{\ell}_n(\theta|\,\boldsymbol{X})\big]}_{=1} - \theta\,\underbrace{\mathsf{E}_\theta\big[\dot{\ell}_n(\theta|\,\boldsymbol{X})\big]}_{=0} \\ &= 1 \end{aligned} \tag{7.23}$$

を得る．いま

$$Y = \frac{\dot{\ell}_n(\theta|\,\boldsymbol{X})}{\mathcal{F}_{\boldsymbol{X}}(\theta)}$$

とおく．(7.19) に注意すると

$$\mathsf{E}_\theta\big[Y\dot{\ell}_n(\theta|\,\boldsymbol{X})\big] = \mathsf{E}_\theta\left[\frac{\big\{\dot{\ell}_n(\theta|\,\boldsymbol{X})\big\}^2}{\mathcal{F}_{\boldsymbol{X}}(\theta)}\right] = \frac{1}{\mathcal{F}_{\boldsymbol{X}}(\theta)}\mathsf{Var}[\dot{\ell}_n(\theta|\,\boldsymbol{X})] = 1$$

となる．これと (7.23) を合わせると

$$\begin{aligned}&\mathsf{E}_\theta\big[(\widehat{\theta}_n(\boldsymbol{X}) - \theta - Y)Y\big] \\ &\quad = \frac{1}{\mathcal{F}_{\boldsymbol{X}}(\theta)}\underbrace{\mathsf{E}_\theta\big[(\widehat{\theta}_n(\boldsymbol{X}) - \theta)\dot{\ell}_n(\theta|\,\boldsymbol{X})\big]}_{=1} - \frac{1}{\mathcal{F}_{\boldsymbol{X}}(\theta)}\underbrace{\mathsf{E}_\theta\big[Y\dot{\ell}_n(\theta|\,\boldsymbol{X})\big]}_{=1} \\ &\quad = 0\end{aligned}$$

を得る．上の式を用いると

$$\begin{aligned}\mathsf{E}_\theta\big[\big(\widehat{\theta}_n(\boldsymbol{X})-\theta\big)^2\big] &= \mathsf{E}_\theta\big[\big(\widehat{\theta}_n(\boldsymbol{X})-\theta-Y+Y\big)^2\big]\\ &= \mathsf{E}_\theta\big[\big(\widehat{\theta}_n(\boldsymbol{X})-\theta-Y\big)^2\big]+\mathsf{E}_\theta\big[Y^2\big]\\ &\geq \mathsf{E}_\theta\big[Y^2\big]\end{aligned}$$

となる．よって等号成立は

$$\begin{aligned}\widehat{\theta}_n(\boldsymbol{X})-\theta-Y=0 &\Leftrightarrow \widehat{\theta}_n(\boldsymbol{X})-\theta=\frac{\dot{\ell}(\theta|\,\boldsymbol{X})}{\mathcal{F}_{\boldsymbol{X}}(\theta)}\\ &\Leftrightarrow \mathcal{F}_{\boldsymbol{X}}(\theta)\big(\widehat{\theta}_n(\boldsymbol{X})-\theta\big)=\dot{\ell}_n(\theta|\,\boldsymbol{X})\end{aligned}$$

となる．このことより，$\overline{\theta},\ \underline{\theta}\in\mathbb{X}$ とすると

$$\begin{aligned}\ell_n(\overline{\theta}|\,\boldsymbol{X})-\ell_n(\underline{\theta}|\,\boldsymbol{X}) &= \int_{\underline{\theta}}^{\overline{\theta}}\dot{\ell}_n(\boldsymbol{X}|\,\theta)\,\mathrm{d}\theta=\int_{\underline{\theta}}^{\overline{\theta}}\mathcal{F}_{\boldsymbol{X}}(\theta)\big(\widehat{\theta}_n(\boldsymbol{X})-\theta\big)\,\mathrm{d}\theta\\ &= \widetilde{A}(\overline{\theta},\,\underline{\theta})\widehat{\theta}_n(\boldsymbol{X})-\widetilde{B}(\overline{\theta},\,\underline{\theta})\end{aligned}$$

と書ける．ただし

$$\widetilde{A}(\overline{\theta},\,\underline{\theta}):=\int_{\underline{\theta}}^{\overline{\theta}}\mathcal{F}_{\boldsymbol{X}}(\theta)\,\mathrm{d}\theta;\qquad \widetilde{B}(\overline{\theta},\,\underline{\theta}):=\int_{\underline{\theta}}^{\overline{\theta}}\theta\mathcal{F}_{\boldsymbol{X}}(\theta)\,\mathrm{d}\theta$$

である．ここで $n=1,\ \overline{\theta}=\theta,\ \underline{\theta}=1,\ X_1=x$ とおくと

$$\ell_1(\theta|\,x)=T(x)A(\theta)-B(\theta)+g(x)$$

と書ける．ただし

$$T(x)=\widehat{\theta}_1(x),\quad g(x)=\ell_1(1|\,x),\quad A(\theta)=\widetilde{A}(\theta,\,1),\quad B(\theta)=\widetilde{B}(\theta,\,1)$$

と書ける． □

注意 7.20. 定理 7.19 から，R 有効な推定量が存在する統計的モデルは指数型分布族となることがわかる． □

7.4 章末注釈と参考文献

定理 7.9 の証明は [6, pp.241 − 242] を借用した．定理 7.12 の証明は [32, pp.35 − 41] を借用した．定理 7.19 は [21, pp.39 − 40] を借用した．

7.5 演習問題

演習問題 7.1. 離散型確率変数 X は, p.m.f.

$$\mathsf{p}(x|\,\theta) = \begin{cases} \left(\dfrac{\theta}{2}\right)^{|x|}(1-\theta)^{1-|x|} & (x=-1,\,0,\,1) \\ 0 & (\text{その他の場合}) \end{cases}$$

からの標本の大きさが 1 のランダム標本とする．ただし, $0<\theta<1$ で, X は確率空間 $(\Omega,\,\mathcal{A},\,\mathsf{Pr})$ 上で定義された確率変数とする．ふたつの統計量

$$S = S(X) = |X|, \qquad T = T(X) = \begin{cases} 2 & (X=1) \\ 0 & (\text{その他の場合}) \end{cases}$$

を考える.
(1) S の確率分布を求めよ[2].
(2) T の確率分布を求めよ.
(3) S が与えられたときの X の条件付確率分布を求め[3], S は θ の十分統計量かどうかを調べよ.
(4) S は θ の不偏推定量かどうかを調べよ.
(5) T は θ の不偏推定量かどうかを調べよ.
(6) S と T の平均 2 乗誤差 $\mathsf{MSE}_S(\theta)=\mathsf{E}[(S-\theta)^2]$ と $\mathsf{MSE}_T(\theta)=\mathsf{E}[(T-\theta)^2]$ を求めよ．S と T の平均 2 乗誤差 $\mathsf{MSE}_S(\theta)$ と $\mathsf{MSE}_T(\theta)$ の大小の比較をせよ. (横軸を θ とし, 縦軸を MSE の値として, S と T の平均 *2* 乗誤差 $\mathsf{MSE}_S(\theta)$ と $\mathsf{MSE}_T(\theta)$ のグラフを描き比較すること.)

演習問題 7.2. $\theta>0$ とする．連続型確率変数 X は p.d.f.

$$\mathsf{p}(x|\,\theta) = \begin{cases} \theta x^{\theta-1} & (0<x<1) \\ 0 & (\text{その他}) \end{cases}$$

からの大きさ 1 のランダム標本とする. ただし, X は確率空間 $(\Omega,\,\mathcal{A},\,\mathsf{Pr})$ 上で定義された確率変数とする.
(1) 連続型確率変数 X の期待値 $\mathsf{E}[X]$ を求めよ.
(2) 連続型確率変数 X の分布関数 $\mathsf{F}(x)=\mathsf{Pr}(X\le x)\,(x\in\mathbb{R})$ を求めよ.
(3) $X=x\,(0<x<1)$ を観測したのときの尤度関数 $\mathsf{lik}(\theta|\,x)$ を述べよ *(*答えのみでよい*)*.
(4) θ の最尤推定値を求めよ.

[2] 確率関数または確率分布表を求めること.
[3] 条件付き p.m.f. を求めるか, X が与えられた値ごとの確率分布表を求めればよい.

演習問題 7.3. $m,\, n \geq 2$ を整数とし

$$\begin{aligned} X_1,\, X_2,\, \ldots,\, X_m &\overset{\text{i.i.d.}}{\sim} N(\mu_1,\, \sigma^2) \\ Y_1,\, Y_2,\, \ldots,\, Y_n &\overset{\text{i.i.d.}}{\sim} N(\mu_2,\, \sigma^2) \end{aligned}$$

とする．さらに

$$\begin{aligned} \overline{X}_m &= \frac{1}{m}\sum_{i=1}^{m} X_i \\ \overline{Y}_n &= \frac{1}{n}\sum_{i=1}^{n} Y_i \\ \widehat{\sigma}^2 &= \frac{1}{m+n-2}\left\{\sum_{i=1}^{m}(\overline{X}_m - X_i)^2 + \sum_{i=1}^{n}(\overline{Y}_n - Y_i)^2\right\} \end{aligned}$$

とおく．これらの確率変数は同じ確率空間 $(\Omega,\, \mathcal{A},\, \mathsf{Pr})$ 上で定義されたものとする．このとき, 以下の問いに答えよ．

(1)

$$\overline{X}_m - \overline{Y}_n \sim \mathsf{N}\left(\mu_1 - \mu_2,\, \frac{m+n}{mn}\sigma^2\right)$$

となることを説明 *(*証明*)* せよ．

(2)

$$\mathsf{E}[\widehat{\sigma}^2] = \sigma^2$$

となることを説明（証明）せよ．

演習問題 7.4. $n \geq 2$ を整数とし, $X_1,\, X_2,\, \ldots,\, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Po}(\lambda)$ とする[4]．ただし, $\lambda > 0$ で, これらの確率変数は確率空間 $(\Omega,\, \mathcal{A},\, \mathsf{Pr})$ 上で定義されているとする．

(1) $\sum_{i=1}^{n} X_i$ の分布は $\mathsf{Po}(n\lambda)$ となることを積率母関数を計算することで示せ．

(2) λ の任意の不偏推定量の分散についてのその下限（Cramér-Rao の下限）を求めよ．

(3) λ の最尤推定量を求め, その分散が Cramér-Rao の下限に到達することを確認せよ．

---

[4]

$$\mathsf{Pr}(X_1 = x) = \frac{\lambda^x}{x!}e^{-\lambda} \qquad (x = 0,\, 1,\, \ldots)$$

である．ただし, これらの確率変数は確率空間 $(\Omega,\, \mathcal{A},\, \mathsf{Pr})$ 上で定義されたものとする．その積率母関数は

$$\mathsf{M}_{X_1}(t) = \mathsf{E}[e^{tX_1}] = e^{(e^t-1)\lambda}$$

である．

演習問題 7.5. $X_1, X_2, \ldots, X_n$ を母集団分布 (平均は θ, 分散は 1^2 の正規分布)

$$\mathsf{p}(x|\,\theta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-\theta)^2}{2}\right)$$

からの大きさ n のランダム標本とする. ただし, 母数 $\theta\,(-\infty < \theta < \infty)$ は未知とする. このとき, 以下の問いに答えよ.
(1) $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ を観測したときの θ の尤度関数 $\mathsf{lik}_n(\theta|\,x_1, x_2, \ldots, x_n)$ と対数尤度関数 $\mathsf{lik}_n(\theta|\,x_1, x_2, \ldots, x_n)$ を書け.
(2) θ の最尤推定値 $\widehat{\theta}_n(x_1, x_2, \ldots, x_n)$ を求めよ.
(3) θ の最尤推定量 $\widehat{\theta}_n(X_1, X_2, \ldots, X_n)$ の平均 $\mathsf{E}[\widehat{\theta}_n(X_1, X_2, \ldots, X_n)]$ を求めよ.
(4) θ の最尤推定量 $\widehat{\theta}_n(X_1, X_2, \ldots, X_n)$ の分散 $\mathsf{Var}[\widehat{\theta}_n(X_1, X_2, \ldots, X_n)]$ を求めよ.
(5) 任意の正数 ϵ に対して,

$$\lim_{n\to\infty} \mathsf{Pr}\big(|\widehat{\theta}_n(X_1, X_2, \ldots, X_n) - \theta| \geq \epsilon\big) = 0$$

を示せ.

演習問題 7.6.

演習問題 7.7.

第8章　検定と信頼区間

第 8.1 では仮説検定問題の枠組みと考え方を説明する．第 8.2 では, 帰無仮説も対立仮説も単純であるとき, 検定統計量の最適定理である Neyman-Pearson の補題を説明する．第 8.3 では, 検定統計量の導出原理を説明する．第 8.4 では様々な検定方法をまとめる．第 8.5 では, 区間推定量の考え方を説明する．第 8.6 では, 区間推定量の構成法の代表的なものを説明する．

8.1　仮説検定の考え方

母集団分布を特徴付ける母数について想定したある仮説の真偽を標本に基づいて調べることを**仮説検定** hypothsis test という．

いま

$$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top \sim \mathsf{P}_{\boldsymbol{\theta}^*}^{\otimes n}\,(\boldsymbol{\theta}^* \in \Theta \subset \mathbb{R}^d)$$

とする．ただし $d, n \in \mathbb{N}$ で, Θ は母数空間である．また, $\mathbb{X}^n\,(\subset \mathbb{R}^n)$ を $\boldsymbol{X}$ の値域としたとき, $\mathsf{P}_{\boldsymbol{\theta}^*}^{\otimes n}$ は $\big(\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n)\big)$ 上の確率測度 ($\boldsymbol{X}$ の分布) である．$\Theta_0 \subset \Theta$ は空でない Θ の真部分集合とし, $\boldsymbol{\theta}^*$ が Θ_0 に入るか否かを調べたいとき, 仮説

$$H_0 : \boldsymbol{\theta}^* \in \Theta_0 \quad \text{vs.} \quad H_1 : \boldsymbol{\theta}^* \in \Theta_1 := \Theta \setminus \Theta_0 \tag{8.1}$$

を考える．H_0 または H_1 のいずれかが正しいかを判断することを「H_0 を H_1 に対して検定 (test) する」という．H_0 を**帰無仮説** (null hypothsis) といい, H_1 を**対立仮説** (alternative hypothsis) という．Θ_0 が Θ の 1 つの元から成るとき H_0 を**単純仮説** (simple hypothsis) という．そうでないとき H_0 を**複合仮説** (composite hypothsis) という．言葉を乱用して, 「Θ_0 は単純仮説である」等ということもある．

仮説 (8.1) に対して, 以下のように検定方式を定めることができる．$\boldsymbol{X}$ の取り得るすべての値の集合を $\mathbb{X}^n(\subset \mathbb{R}^n)$ と表す．$\mathbb{X}^n$ を 2 つの排反で空でない部分集合 W と W^c に分割する．すなわち $W \neq \varnothing, W^c \neq \varnothing$ で $W \cup W^c = \mathbb{X}^n$ かつ $W \cap W^c = \varnothing$ である．$\boldsymbol{X}$ の実現値を $\boldsymbol{x}$ と書いたと

き, 検定方式は

$$\begin{aligned} \boldsymbol{x} \in W &\Rightarrow \text{帰無仮説 } H_0 \text{ を棄却し, 対立仮説 } H_1 \text{ を採択,} \\ \boldsymbol{x} \in W^c &\Rightarrow \text{帰無仮説 } H_0 \text{ を受容} \end{aligned}$$

と表現できる. このとき W を**棄却域** (critical region) といい, W^c を**受容域** (acceptance region) という.

上のように定めた検定方式には 2 つのタイプの誤りが起こる可能性がある. (1) 帰無仮説 H_0 が正しいにもかかわらず標本の実現値 $\boldsymbol{x}$ に基づいて検定した結果, H_0 を棄却してしまうこと. 逆に, (2) 対立仮説 H_1 が正しいにもかかわらず標本の実現値 $\boldsymbol{x}$ に基づいて検定した結果, H_0 を受容してしまうことである. (1) の誤判断を**第 1 種の誤り**といい, (2) の誤判断を**第 2 種の誤り**とそれぞれ呼ぶ. 一般に一方の誤りが起こる確率を小さくする検定方式は, 他方の誤りを起こす確率を大きくする. すなわち, 両者の誤りが起こる確率を同時に小さくする検定方式はないことが知られている.

以下では「よい」検定方式を一般的な形で定式化することを考える. 関数 $\phi\colon \mathbb{X}^n \to [0, 1]$ は可測関数とする. この関数 ϕ を用いて次のように検定方式を定める. $\boldsymbol{X} = \boldsymbol{x}$ を観測したとき, 確率 $\phi(\boldsymbol{x})$ で帰無仮説 H_0 を棄却する検定方式を考える. この ϕ を**検定関数** (test function) という. 関数 ϕ が $\mathbb{X}$ の空でない部分集合の定義関数のとき, この検定関数 ϕ で定まる検定方式を**非確率化検定** (nonrandomized test) という. すなわち

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & (\boldsymbol{x} \in W) \\ 0 & (\boldsymbol{x} \in W^c) \end{cases}$$

とすると棄却域 W をもつ非確率化検定が定まる. そうでない検定方式を**確率化検定** (randomized test) という.

以後 $\phi(\boldsymbol{X})$ を検定統計量 (test statistic) ということにする. さらに検定統計量 ϕ によって定まる検定方式を単に検定ということにする. 検定統計量 $\phi(\boldsymbol{X})$ の第 1 種の誤りの確率は

$$\mathsf{E}_{\boldsymbol{\theta}}\big[\phi(\boldsymbol{X})\big] \qquad (\boldsymbol{\theta} \in \Theta_0)$$

となり, 第 2 種の誤りの確率は

$$1 - \mathsf{E}_{\boldsymbol{\theta}}\big[\phi(\boldsymbol{X})\big] \qquad (\boldsymbol{\theta} \in \Theta_1)$$

となる. ただし

$$\mathsf{E}_{\boldsymbol{\theta}}[\phi(\boldsymbol{X})] = \int_{\mathbb{X}^n} \phi(\boldsymbol{x}) \mathsf{p}(\boldsymbol{x} | \, \boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{x}$$

と定めた．「よい」検定として, まず第 1 種の誤りの確率の Θ_0 上の上限を $\alpha\,(0<\alpha<1)$ 以下にするような検定を考える．すなわち

$$\sup_{\boldsymbol{\theta}\in\Theta_0} \mathsf{E}_{\boldsymbol{\theta}}\big[\phi(\boldsymbol{X})\big] \le \alpha \tag{8.2}$$

である．(8.1) をみたす検定または ϕ を**有意水準 α の検定** (level α test) という．つぎに有意水準 α のある検定 ϕ で, 有意水準 α の検定の中で第 2 種の誤りの確率

$$1-\mathsf{E}_{\boldsymbol{\theta}}\big[\phi(\boldsymbol{X})\big] \qquad (\boldsymbol{\theta}\in\Theta_1) \tag{8.3}$$

を Θ_1 上で最小にするするものを見つけることを目指す．すなわち

$$\mathsf{E}_{\boldsymbol{\theta}}\big[\phi(\boldsymbol{X})\big] \qquad (\boldsymbol{\theta}\in\Theta_1)$$

を最大にするものである．この確率を $\boldsymbol{\theta}\in\Theta_1$ の関数とみて

$$\beta\big(\boldsymbol{\theta}\big) := \mathsf{E}_{\boldsymbol{\theta}}\big[\phi(\boldsymbol{X})\big] \qquad (\boldsymbol{\theta}\in\Theta_1)$$

と表記する．これを**検出力関数** (power function) または**検出力**という．したがって「よい」検定は次のように定義される．

定義 8.1. 有意水準 $\alpha\,(0<\alpha<1)$ の検定 ϕ で検出力を任意の $\boldsymbol{\theta}\in\Theta_1$ に対して最大にするものを有意水準 α の**一様最強力検定** (uniformly most powerful test= u.m.p. 検定) という．特に帰無仮説と対立仮説が単純仮説であるとき, u.m.p. 検定を単に有意水準 α の**最強力検定** (m.p. 検定) という．

8.2　Neyman-Pearson **の定理**

まず m.p. 検定を求める最も基本的定理を述べる．以下では, 簡単のために $\boldsymbol{X}$ は同時 p.d.f. $\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\boldsymbol{\theta})$ をもつとして議論を進めていく．離散型確率変数のときは同時 p.m.f. を考え, 積分を和の記号に替えればよい．

定理 8.2. (Neyman-Pearson の定理) 母数空間は $\mathbb{R}^d$ の異なる 2 点から成るとする．$\Theta=\{\boldsymbol{\theta}_0,\boldsymbol{\theta}_1\}$ である．$\boldsymbol{X}=(X_1,X_2,\ldots,X_n)^\top \sim \mathsf{P}_{\boldsymbol{\theta}}^{\otimes n}\,(\boldsymbol{\theta}\in\Theta\subset\mathbb{R}^d)$ とし, $\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\boldsymbol{\theta})\,(\boldsymbol{\theta}\in\Theta,\boldsymbol{x}\in\mathbb{R}^n)$ を $\boldsymbol{X}$ の同時 p.d.f. または p.m.f. とする．検定問題

$$H_0:\boldsymbol{\theta}=\boldsymbol{\theta}_0 \quad \text{vs.} \quad H_1:\boldsymbol{\theta}=\boldsymbol{\theta}_1$$

に対する有意水準 $\alpha\,(0<\alpha<1)$ の m.p. 検定 ϕ_0 は以下で与えられる．

$$\phi_0(\boldsymbol{x}) = \begin{cases} 1 & (\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\boldsymbol{\theta}_1) > c\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\boldsymbol{\theta}_0)) \\ \gamma & (\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\boldsymbol{\theta}_1) = c\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\boldsymbol{\theta}_0)) \\ 0 & (\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\boldsymbol{\theta}_1) < c\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\boldsymbol{\theta}_0)) \end{cases} \tag{8.4}$$

である．ただし $\gamma,\,c\,(0\le\gamma\le 1,\,c>0)$ は

$$\mathsf{E}_{\boldsymbol{\theta}_0}\big[\phi_0(\boldsymbol{X})\big] = \alpha \tag{8.5}$$

から定まる定数である．

Proof. まず

$$\begin{aligned} B_1 &= \{\boldsymbol{x}\in\mathbb{R}^n;\ \mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\boldsymbol{\theta}_1) > c\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\boldsymbol{\theta}_0)\}, \\ B_2 &= \{\boldsymbol{x}\in\mathbb{R}^n;\ \mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\boldsymbol{\theta}_1) = c\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\boldsymbol{\theta}_0)\}, \\ B_3 &= \{\boldsymbol{x}\in\mathbb{R}^n;\ \mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\boldsymbol{\theta}_1) < c\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\boldsymbol{\theta}_0)\} \end{aligned}$$

とする．ϕ は有意水準 α の任意の検定とする．すなわち

$$\mathsf{E}_{\boldsymbol{\theta}_0}\big[\phi(\boldsymbol{X})\big] \le \alpha \tag{8.6}$$

をみたす．一方 (8.4) より

$$
\begin{aligned}
&\mathsf{E}_{\boldsymbol{\theta}_1}\big[\phi_0(\boldsymbol{X})\big]-\mathsf{E}_{\boldsymbol{\theta}_1}\big[\phi(\boldsymbol{X})\big]\\
&\quad=\int_{\mathbb{X}}\phi_0(\boldsymbol{x})\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta_1)\,\mathrm{d}\boldsymbol{x}-\int_{\mathbb{X}}\phi(\boldsymbol{x})\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta_1)\,\mathrm{d}\boldsymbol{x}\\
&\quad=\int_{B_1}\underbrace{\phi_0(\boldsymbol{x})}_{=1}\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta_1)\,\mathrm{d}\boldsymbol{x}+\int_{B_2}\underbrace{\phi_0(\boldsymbol{x})}_{=\gamma}\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta_1)\,\mathrm{d}\boldsymbol{x}+\int_{B_3}\underbrace{\phi_0(\boldsymbol{x})}_{=0}\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta_1)\,\mathrm{d}\boldsymbol{x}\\
&\qquad-\int_{B_1}\phi(\boldsymbol{x})\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta_1)\,\mathrm{d}\boldsymbol{x}-\int_{B_2}\phi(\boldsymbol{x})\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta_1)\,\mathrm{d}\boldsymbol{x}-\int_{B_3}\phi(\boldsymbol{x})\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta_1)\,\mathrm{d}\boldsymbol{x}\\
&\quad=\int_{B_1}\{1-\phi(\boldsymbol{x})\}\underbrace{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\boldsymbol{\theta}_1)\,\mathrm{d}\boldsymbol{x}}_{>c\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\boldsymbol{\theta}_0)\,(\boldsymbol{x}\in B_1)}+\int_{B_2}\{\gamma-\phi(\boldsymbol{x})\}\underbrace{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\boldsymbol{\theta}_1)\,\mathrm{d}\boldsymbol{x}}_{=c\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\boldsymbol{\theta}_0)\,(\boldsymbol{x}\in B_2)}\\
&\qquad+\int_{B_3}\{-\phi(\boldsymbol{x})\}\underbrace{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\boldsymbol{\theta}_1)}_{<c\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\boldsymbol{\theta}_0)\,(\boldsymbol{x}\in B_3)}\,\mathrm{d}\boldsymbol{x}\\
&\quad\geq\int_{B_1}\{1-\phi(\boldsymbol{x})\}c\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\boldsymbol{\theta}_0)\,\mathrm{d}\boldsymbol{x}+\int_{B_2}\{\gamma-\phi(\boldsymbol{x})\}c\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\boldsymbol{\theta}_0)\,\mathrm{d}\boldsymbol{x}\\
&\qquad+\int_{B_3}\{-\phi(\boldsymbol{x})\}c\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\boldsymbol{\theta}_0)\,\mathrm{d}\boldsymbol{x}\\
&\quad=c\int_{B_1}\{1-\phi(\boldsymbol{x})\}\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\boldsymbol{\theta}_0)\,\mathrm{d}\boldsymbol{x}+c\int_{B_2}\{\gamma-\phi(\boldsymbol{x})\}\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\boldsymbol{\theta}_0)\,\mathrm{d}\boldsymbol{x}\\
&\qquad+c\int_{B_3}\{-\phi(\boldsymbol{x})\}\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\boldsymbol{\theta}_0)\,\mathrm{d}\boldsymbol{x}\\
&\quad=c\int_{\mathbb{X}^n}\{\phi_0(\boldsymbol{x})-\phi(\boldsymbol{x})\}\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\boldsymbol{\theta}_0)\,\mathrm{d}\boldsymbol{x}\\
&\quad=c\Big\{\mathsf{E}_{\theta_0}[\phi_0(\boldsymbol{X})]-\mathsf{E}_{\theta_0}[\phi(\boldsymbol{X})]\Big\}\qquad(\because (8.4))\\
&\quad=c\Big\{\alpha-\mathsf{E}_{\boldsymbol{\theta}_0}\big[\phi(\boldsymbol{X})\big]\Big\}\geq 0\qquad(\because (8.6)\text{より})
\end{aligned}
$$

を得る．したがって

$$
\mathsf{E}_{\boldsymbol{\theta}_1}\big[\phi_0(\boldsymbol{X})\big]\geq\mathsf{E}_{\boldsymbol{\theta}_1}\big[\phi(\boldsymbol{X})\big]
$$

となるので, ϕ_0 は有意水準 α の m.p. 検定となる． □

例 8.3. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}(\theta, \sigma^2)$ で $\sigma^2\,(\sigma > 0)$ は既知とする．このとき検定問題

$$
H_0: \theta=\theta_0 \quad \text{vs.} \quad H_1: \theta=\theta_1 \qquad (\theta_1>\theta_0)
$$

に対する m.p. 検定を求める．まず $\boldsymbol{X}=(X_1, X_2, \ldots, X_n)^\top$ の同時 p.d.f.

は

$$\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left[-\frac{1}{2\sigma^2}\sum_{j=1}^{n}(x_j-\theta)^2\right], \qquad \boldsymbol{x} = (x_1,\, x_2,\, \ldots,\, x_n)^\top$$

で与えられることに注意をする. 簡単な計算から

$$\begin{aligned}\log\left[\frac{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta_1)}{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta_0)}\right] &= -\frac{1}{2\sigma^2}\left[\sum_{j=1}^{n}(x_j-\theta_1)^2 - \sum_{j=1}^{n}(x_j-\theta_0)^2\right] \\ &= \frac{(\theta_1-\theta_0)^2}{\sigma^2}\left(\overline{x}_n - \frac{\theta_0+\theta_1}{2}\right) \end{aligned} \tag{8.7}$$

となる. ただし

$$\overline{x}_n = \frac{1}{n}\sum_{j=1}^{n} x_j$$

である. ここで

$$\begin{aligned}\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta_0) > c\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta_1) &\Leftrightarrow \log\left[\frac{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta_1)}{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta_0)}\right] > \log c \\ &\Leftrightarrow \overline{x}_n > c' \qquad (\because (8.7) \text{ より})\end{aligned}$$

である. $\boldsymbol{X}$ は連続型確率変数なので, $\overline{X}_n = n^{-1}\sum_{j=1}^{n} X_j = c'$ である確率は 0 となるので, m.p. 検定の形は

$$\phi_0(\boldsymbol{x}) = \begin{cases} 1 & (\overline{x}_n > c') \\ 0 & (\overline{x}_n \le c') \end{cases}$$

となる. 定数 c' は

$$\alpha = \mathsf{Pr}_{\theta_0}\{\overline{X}_n > c'\} \tag{8.8}$$

から定まる. ただし, $B \in \mathcal{B}(\mathbb{R}^n)$ に対して

$$\mathsf{Pr}_\theta(\boldsymbol{X} \in B) = \mathsf{E}_\theta[\mathbb{1}_B(\boldsymbol{X})] = \int_B \mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)\,\mathrm{d}\boldsymbol{x} \qquad (\theta \in \Theta)$$

と定めた. (8.8) は

$$\alpha = \mathsf{Pr}_{\theta_0}\left\{\frac{\sqrt{n}(\overline{X}_n-\theta_0)}{\sigma} > \frac{\sqrt{n}(c'-\theta_0)}{\sigma}\right\} \tag{8.9}$$

と書き直せ, $\theta = \theta_0$ のもとで $\sqrt{n}(\overline{X}_n-\theta_0)/\sigma \sim \mathsf{N}(0,\,1)$ であるので

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,\mathrm{d}x$$

とすると (8.9) は

$$\alpha = 1 - \Phi\left[\frac{\sqrt{n}(c' - \theta_0)}{\sigma}\right]$$

となる．標準正規分布の上側 $100 \times \alpha\%$ を z_α とすると

$$\frac{\sqrt{n}(c' - \theta_0)}{\sigma} = z_\alpha \Leftrightarrow c' = \theta_0 + \frac{z_\alpha \sigma}{\sqrt{n}}$$

を得る．よって有意水準 α の m.p. 検定は

$$\phi_0(\boldsymbol{x}) = \begin{cases} 1 & \left(\overline{x}_n > \theta_0 + \frac{z_\alpha\sigma}{\sqrt{n}}\right) \\ 0 & \left(\overline{x}_n \le \theta_0 + \frac{z_\alpha\sigma}{\sqrt{n}}\right) \end{cases}$$

となる．次に ϕ_0 の検出力は次のようになる．

$$\begin{aligned} \beta_{\phi_0}(\theta_1) &= \mathsf{E}_{\theta_1}\big[\phi_0(\boldsymbol{X})\big] = \mathsf{Pr}_{\theta_1}\left\{\overline{X}_n > \theta_0 + \frac{z_\alpha\sigma}{\sqrt{n}}\right\} \\ &= \mathsf{Pr}_{\theta_1}\left\{\frac{\sqrt{n}\big(\overline{X}_n - \theta_1\big)}{\sigma} > z_\alpha - \frac{\sqrt{n}\big(\theta_1 - \theta_0\big)}{\sigma}\right\} \end{aligned} \tag{8.10}$$

となる．$\theta = \theta_1$ のとき $\sqrt{n}\big(\overline{X}_n - \theta_1\big)/\sigma \sim \mathsf{N}(0, 1)$ なので，(8.10) より仮説間の平均の差 $\theta_0 - \theta_1 (> 0)$ が大きいほど検出力は大きくなる．また標本 n が大きくなっても検出力が大きくなることがわかる． □

例 8.4. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Ber}(\theta)\,(0 < \theta < 1)$ とする．このとき検定問題

$$H_0 : \theta = \theta_0, \quad H_1 : \theta = \theta_1 \qquad (\theta_1 > \theta_0) \tag{8.11}$$

に対する m.p. 検定を求める．まず $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top$ の同時確率関数 p.m.f. は

$$\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta) = \prod_{j=1}^{n} \theta^{x_j}(1-\theta)^{1-x_j} \qquad \big(\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^\top\big) \tag{8.12}$$

で与えられるので

$$\log\frac{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta_1)}{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta_0)} = \left\{\sum_{j=1}^{n} x_j\right\}\log\left\{\frac{\theta_1(1-\theta_0)}{(1-\theta_1)\theta_0}\right\} + n\log\left\{\frac{1-\theta_1}{1-\theta_0}\right\}$$

となる．$\theta_1 > \theta_0$ としたので，$\dfrac{\theta_1(1-\theta_0)}{(1-\theta_1)\theta_0} > 1$ であることに注意すると

$$\begin{aligned} \mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta_1) > c\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta_0) &\Leftrightarrow \log\left[\frac{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta_1)}{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta_0)}\right] > c' \\ &\Leftrightarrow \sum_{j=1}^{n} x_j > c'' \end{aligned}$$

と書きかえることができる．よって有意水準 α の m.p. 検定は

$$\phi_0(\boldsymbol{x}) = \begin{cases} 1 & \left(\sum_{j=1}^n x_j > c''\right) \\ \gamma & \left(\sum_{j=1}^n x_j = c''\right) \\ 0 & \left(\sum_{j=1}^n x_j < c''\right) \end{cases} \tag{8.13}$$

の形になる．ここで γ と c'' を

$$\begin{aligned} \alpha &= \mathsf{E}_{\theta_0}\big[\phi_0(\boldsymbol{X})\big] \\ &= \mathsf{Pr}_{\theta_0}\left\{\sum_{j=1}^n X_j > c''\right\} + \gamma \mathsf{Pr}_{\theta_0}\left\{\sum_{j=1}^n X_j = c''\right\} \end{aligned} \tag{8.14}$$

から定まる．$\sum_{j=1}^n X_j$ は $\theta = \theta_0$ のとき二項分布 $\mathsf{Bino}(n,\, \theta_0)$ に従うので，(8.14) は

$$\alpha = \sum_{j=c''+1}^{n} \binom{n}{j} \theta_0^j (1-\theta_0)^{n-j} + \gamma \binom{n}{c''} \theta_0^{c''} (1-\theta)^{n-c''}$$

となる．まずは c'' を

$$\sum_{j=c''+1}^{n} \binom{n}{j} \theta_0^j (1-\theta_0)^{n-j} \le \alpha < \sum_{j=c''}^{n} \binom{n}{j} \theta_0^j (1-\theta_0)^{n-j}$$

をみたす整数を定める．これから c_0 と書くことにする．すると γ は

$$\gamma = \left[\alpha - \sum_{j=c_0+1}^{n} \binom{n}{j} \theta_0^j (1-\theta_0)^{n-j}\right] \Bigg/ \binom{n}{c_0} \theta_0^{c_0} (1-\theta_0)^{n-c_0}$$

で定められる．　□

8.3　検定統計量の導出方法

$\mathbb{X}^n$ を標本空間とし，$\boldsymbol{X} = (X_1,\, X_2,\, \ldots,\, X_n)^\top \sim \mathsf{P}_{\boldsymbol{\theta}}^{\otimes n}\,(\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d)$ とする．検定関数 $\phi : \mathbb{X}^n \to [0,\, 1]$ によって定まる検定方式は，以下のように定まることがある．ある統計量 $S : \mathbb{X}^n \longrightarrow \mathbb{R}$ と定数 c が存在して

$$S(\boldsymbol{x}) \le c \Rightarrow \phi(\boldsymbol{x}) = 1$$
$$S(\boldsymbol{x}) > c \Rightarrow \phi(\boldsymbol{x}) = 0$$

となる．この場合，$S(\boldsymbol{X})$ のことも検定統計量と呼ぶことにする．

$\boldsymbol{X} = (X_1,\, X_2,\, \ldots,\, X_n)^\top \sim \mathsf{P}_{\boldsymbol{\theta}}^{\otimes n}\,(\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d)$ とする．ただし，$\mathsf{P}_{\boldsymbol{\theta}}^{\otimes n}$ は $\mathbb{R}^n$ 上の確率測度である．$\mathsf{P}_{\boldsymbol{\theta}}^{\otimes n}$ は同時 p.d.f. $\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\boldsymbol{\theta})$ をもつとする．また，母数空間 Θ は Θ_0 と Θ_0 に分割されたとする．すなわち $\Theta_0 \cup \Theta_1 = \Theta$, $\Theta_0 \cap \Theta_1 = \varnothing$, $\Theta_0 \neq \varnothing$, $\Theta_1 \neq \varnothing$ である．

定義 8.5. 検定問題

$$H_0 : \boldsymbol{\theta} \in \Theta_0 \quad \text{vs.} \quad H_1 : \boldsymbol{\theta} \in \Theta_1$$

を検定するための**尤度比検定統計量** (likelihood ratio statistic=l.r. 統計量) は

$$\lambda(\boldsymbol{X}) = \frac{\sup_{\boldsymbol{\theta} \in \Theta_0} \mathsf{p}^{\boldsymbol{X}}(\boldsymbol{X} | \boldsymbol{\theta})}{\sup_{\boldsymbol{\theta} \in \Theta} \mathsf{p}^{\boldsymbol{X}}(\boldsymbol{X} | \boldsymbol{\theta})}$$

で与えらる．このとき正の定数 C が存在して H_0 の棄却域が

$$W = \{\boldsymbol{x} \in \mathbb{R}^n ; \lambda(\boldsymbol{x}) \le C\}$$

で与えられる検定を**尤度比検定** (likelihood ratio test=l.r.t.) という．すなわち

$$\phi(\boldsymbol{x}) = \mathbb{1}_W(\boldsymbol{x}) = \begin{cases} 1 & (\boldsymbol{x} \in W) \\ 0 & (\boldsymbol{x} \notin W) \end{cases}$$

となる．

注意 8.6. $\widehat{\boldsymbol{\theta}}(\boldsymbol{X})$ を母数空間 Θ での $\boldsymbol{\theta}$ の最尤推定量とし, $\widehat{\boldsymbol{\theta}}_0(\boldsymbol{X})$ を母数空間を Θ_0 に制限したときの $\boldsymbol{\theta}$ の最尤推定量とする．このとき

$$\lambda(\boldsymbol{X}) = \frac{\mathsf{p}^{\boldsymbol{X}}\big(\boldsymbol{X} | \widehat{\boldsymbol{\theta}}_0(\boldsymbol{X})\big)}{\mathsf{p}^{\boldsymbol{X}}\big(\boldsymbol{X} | \widehat{\boldsymbol{\theta}}(\boldsymbol{X})\big)}$$

と表現できる．

定理 8.7. $\boldsymbol{X} \sim \mathsf{P}_{\boldsymbol{\theta}^*}^{\otimes n}$ $(\boldsymbol{\theta}^* \in \Theta \subset \mathbb{R}^d)$ とする．Θ の次元を d, Θ_0 の次元を r $(r < d)$ とする．検定問題

$$H_0 : \boldsymbol{\theta}^* \in \Theta_0 \quad \text{vs.} \quad \boldsymbol{\theta}^* \in H_1 : \Theta_1 := \Theta \setminus \Theta_0$$

に対する尤度比検定統計量を $\lambda(\boldsymbol{X})$ とする．このとき H_0 のもとで次が成り立つ．

$$-2 \log \lambda(\boldsymbol{X}) \rightsquigarrow \chi^2_{d-r}$$

が成立する．

Proof. $d = 1$ の場合について, 証明の概略を与える． □

注意 8.8. 定理 8.7 の結果を用いると尤度比検定の棄却域は

$$W = \left\{\boldsymbol{x} \in \mathbb{R}^n ; -2 \log \lambda(\boldsymbol{x}) > \chi^2_{d-r,\alpha}\right\}$$

で与えらる．ただし $\chi^2_{d-r,\alpha}$ は自由度 $d-r$ の χ^2 分布の上側 $100 \times \alpha\%$ 点である．したがって, 検定手続きは

$$\boldsymbol{x} \in W \Rightarrow H_0 \text{ は棄却}$$

となる． □

8.4 様々な検定

8.5 区間推定の考え方

$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top \sim \mathsf{P}_\theta^{\otimes n}\,(\theta \in \Theta \subset R)$ とする．$0 < \alpha < 1$ を固定する．母数 θ に依存しない区間 $\big[L(\boldsymbol{X}), U(\boldsymbol{X})\big] \subset \Theta$ が $\forall \theta \in \Theta$ に対して

$$\mathsf{Pr}_\theta\Big\{L(\boldsymbol{X}) \leq \theta \leq U(\boldsymbol{X})\Big\} \geq 1 - \alpha \tag{8.15}$$

をみたすとき区間 $\big[L(\boldsymbol{X}), U(\boldsymbol{X})\big]$ を信頼係数 $(1-\alpha)$ の θ の**信頼区間** (confidence interval) という．$L(\boldsymbol{X}), U(\boldsymbol{X})$ を**信頼限界** (confident limit) という．通常 α として 0.05, 0.01, 0.1 等が用いられる．(8.15) の関係式は, たとえば 100 組の実現値を発生させると 100α 回程度は信頼区間に真の母数 θ は含まれないと考える．

例 8.9. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}(\theta, \sigma^2)$ とする．ただし $\theta \in \mathbb{R}$ で $\sigma^2\,(\sigma > 0)$ の値は既知である．このとき標本平均 $\overline{X}_n = n^{-1}\sum_{j=1}^n X_j$ は

$$\overline{X}_n \sim \mathsf{N}\bigg(\theta, \frac{\sigma^2}{n}\bigg)$$

となる．このことより

$$S_\theta(\boldsymbol{X}) = \frac{\sqrt{n}\big(\overline{X}_n - \theta\big)}{\sigma} \sim \mathsf{N}(0, 1)$$

となる．ここで $z_{\alpha/2}$ を標準正規分布 $\mathsf{N}(0, 1)$ の上側 $100(\alpha/2)\%$ 点ととると

$$\begin{aligned} 1 - \alpha &= \mathsf{Pr}_\theta\Big\{-z_{\alpha/2} \leq S_\theta(\boldsymbol{X}) \leq z_{\alpha/2}\Big\} \\ &= \mathsf{Pr}_\theta\bigg\{\overline{X}_n - \frac{\sigma}{\sqrt{n}}z_{\alpha/2} \leq \theta \leq \overline{X}_n + \frac{\sigma}{\sqrt{n}}z_{\alpha/2}\bigg\} \end{aligned}$$

なることがわかる．したがって

$$\big[L(\boldsymbol{X}), U(\boldsymbol{X})\big] = \bigg[\overline{X}_n - \frac{\sigma}{\sqrt{n}}z_{\alpha/2}, \overline{X}_n + \frac{\sigma}{\sqrt{n}}z_{\alpha/2}\bigg]$$

は信頼係数 $(1-\alpha)$ の θ の信頼区間となる．□

8.6 信頼区間の構成法

8.6.1 検定方式の反転

$\Theta \subset \mathbb{R}$ とし, $\theta_0 \in \Theta$ を取る[1]. 検定問題

$$H_0 : \theta = \theta_0, \quad H_1 : \theta \neq \theta_0$$

を考える. 有意水準 $\alpha\,(0 < \alpha < 1)$ の検定の受容域を $A(\theta_0)$ とおく. すなわち

$$\mathsf{Pr}_{\theta_0}\big\{\boldsymbol{X} \in A(\theta_0)\big\} \geq 1 - \alpha$$

が成り立っている. そこで $\boldsymbol{X} \in A(\theta_0)$ を θ_0 に関して解くことによって

$$C(\boldsymbol{X}) = \{\theta \in \Theta; \boldsymbol{x} \in A(\theta)\}$$

が得られる. 一般に $C(\boldsymbol{X})$ は連結区間になるという保証はない. $C(\boldsymbol{X})$ が連結区間となれば

$$\mathsf{Pr}_{\theta}\big\{\theta \in A(\theta)\big\} \geq 1 - \alpha$$

となるので, $C(\boldsymbol{X})$ は信頼係数 $(1-\alpha)$ の信頼区間となる.

例 8.10. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Ber}(\theta)\,(0 < \theta < 1)$ とする. $\theta_0 \in (0, 1)$ と固定し, 検定問題

$$H_0 : \theta = \theta_0, \quad H_1 : \theta \neq \theta_0$$

を考える. この検定問題に対する尤度比検定統計量は

$$\begin{aligned}\lambda(\boldsymbol{X}) &= \frac{\prod_{j=1}^n \theta_0^{X_j}(1-\theta_0)^{1-X_j}}{\prod_{j=1}^n \overline{X}_n^{X_j}(1-\overline{X}_n)^{1-X_j}} \\ &= \prod_{j=1}^n \left(\frac{\theta_0}{\overline{X}_n}\right)^{X_j}\left(\frac{1-\theta_0}{1-\overline{X}_n}\right)^{1-X_j}\end{aligned}$$

となる. これより受容域 $A(\theta_0)$ は

$$\begin{aligned}&A(\theta_0) = \Bigg\{\boldsymbol{x} \in \{0, 1\}^n; \\ &-2\log\lambda(\boldsymbol{x}) = 2n\overline{x}_n \log\left(\frac{\overline{x}_n}{\theta_0}\right) + 2n(1-\overline{x}_n)\log\left(\frac{1-\overline{x}_n}{1-\theta_0}\right) \leq \chi^2_{1,\alpha}\Bigg\}\end{aligned}$$

となる. ただし $\overline{x}_n = n^{-1}\sum_{j=1}^n x_j$, $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ である. よって

$$C(\boldsymbol{X}) := \left\{\theta; \overline{X}_n \log\left(\frac{\overline{X}_n}{\theta}\right) + (1-\overline{X}_n)\log\left(\frac{1-\overline{X}_n}{1-\theta}\right) \leq \frac{\chi^2_{1,\alpha}}{2n}\right\}$$

となる. ただし信頼限界 $L(\boldsymbol{X})$, $U(\boldsymbol{X})$ を陽に求めることはできない. □

[1] θ_0 を θ^* と書くべきであろうが, 後の表記上の都合でこの記号を採用した.

8.6.2　枢軸量 (pivotal quantity)

一般に $Q(\boldsymbol{X},\,\theta)$ の分布が θ に依存しないとき, $Q(\boldsymbol{X},\,\theta)$ を**枢軸量** (pivotal quantity) という．このとき

$$\mathsf{Pr}_\theta\Big(a \le Q(\boldsymbol{X},\,\theta) \le b\Big) = 1-\alpha$$

をみたす $a,\,b$ を定めて, $a \le Q(\boldsymbol{X},\,\theta) \le b$ を θ に関して解くことにより, 信頼係数 $(1-\alpha)$ の θ の信頼区間

$$C(\boldsymbol{X}) = \big\{\theta \in \Theta;\, a \le Q(\boldsymbol{X},\,\theta) \le b\big\}$$

が得られる．もちろん $C(\boldsymbol{X})$ が連結区間になる保証は一般的にはないが, うまく連結区間になれば, 信頼区間として使用できる.

例 8.11. $X_1,\,X_2,\,\ldots,\,X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Exp}(\theta)\,(\theta > 0)$ とする．すなわち, p.d.f. は

$$\mathsf{p}(x|\,\theta) = \begin{cases} \theta e^{-\theta x} & (x > 0) \\ 0 & (\text{その他の場合}) \end{cases}$$

で与えられる．すると $2\theta\sum_{j=1}^n X_j$ は自由度 $2n$ の χ^2 分布に従うので

$$Q(\boldsymbol{X},\,\theta) = 2\theta\sum_{j=1}^n X_j$$

とおく．よって $\chi^2_{2n,\,\alpha}$ を自由度 $2n$ の χ^2 分布の上側 $100\alpha\%$ 点とすると

$$\mathsf{Pr}_\theta\Big(\chi^2_{2n,\,1-\alpha/2} \le Q(\boldsymbol{X},\,\theta) \le \chi^2_{2n,\,\alpha/2}\Big) = 1-\alpha$$

となるので,

$$\big[L(\boldsymbol{X}),\,U(\boldsymbol{X})\big] = \left[\frac{\chi^2_{2n,\,1-\alpha/2}}{2\sum_{j=1}^n X_j},\,\frac{\chi^2_{2n,\,\alpha/2}}{2\sum_{j=1}^n X_j}\right]$$

は信頼係数 $(1-\alpha)$ の θ の信頼区間となる．　□

問 8.1. $X_1,\,X_2,\,\ldots,\,X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Exp}(\theta)\,(\theta > 0)$ のとき, $2\theta\sum_{j=1}^n X_j \sim \chi^2_{2n}$ となることを示せ.

8.7　章末注釈と参考文献

節 8.4 は [32, pp.62 – 67] を借用した.

8.8 演習問題

演習問題 8.1. 確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上で定義された確率変数 X_1, X_2, X_3 は独立同一に母数 $\theta\,(0<\theta<1)$ の Benoulli 分布 $\mathsf{Ber}(\theta)$ に従っているとする．検定問題

$$\mathrm{H}_0 : \theta = \frac{1}{2} \quad \text{vs} \quad \mathrm{H}_1 : \theta \neq \frac{1}{2}$$

の検定問題に対して, 検定方式

$$\begin{aligned} W_1 &= \{(x_1, x_2, x_3) \in S; \sum_{i=1}^{3} x_i \in \{0, 3\}\} \\ W_2 &= \{(x_1, x_2, x_3) \in S; \sum_{i=1}^{3} x_i \in \{2, 3\}\} \end{aligned}$$

を考える．ただし, $S = \{(x_1, x_2, x_3); x_i \in \{0, 1\}\,(i = 1, 2, 3)\}$ である．検定方式 W_1, W_2 のサイズを求めよ．すなわち

$$\mathsf{Pr}_{\theta=1/2}((X_1, X_2, X_3) \in R_1) \quad \text{および} \quad \mathsf{Pr}_{\theta=1/2}((X_1, X_2, X_3) \in R_2)$$

の確率である．

$\boxed{\text{記号}}$ $\mathsf{p}(x|\,\theta)$ を $\mathsf{Ber}(\theta)$ の p.m.f. とする．Borel 集合 $A \subset \mathbb{R}^3$ に対して

$$\mathsf{Pr}_{\theta=1/2}\big((X_1, X_2, X_3) \in A\big) = \sum_{(x_1, x_2, x_3) \in A \cap S} \mathsf{p}\big(x_1|\,\frac{1}{2}\big)\mathsf{p}\big(x_2|\,\frac{1}{2}\big)\mathsf{p}\big(x_3|\,\frac{1}{2}\big)$$

と定めている．

演習問題 8.2. $\theta \in \mathbb{R}$ とし, 連続型確率変数 X は p.d.f.

$$\mathsf{p}(x|\,\theta) = \frac{1}{\pi\{1 + (x-\theta)^2\}} \qquad (-\infty < x < \infty)$$

を持つとする．仮説検定問題

$$\text{帰無仮説 } H_0 : \theta = 0 \quad \text{vs.} \quad \text{対立仮説 } H_0 : \theta = 1$$

を X に基づく検定する．この検定問題に対して, 棄却域

$$W := \{x \in \mathbb{R} : 1 < x < 3\}$$

を考える．以下の問いに答えよ．ただし, $\arctan 2 = 1.107$, $\arctan 3 = 1.249$, $\pi = 3.1416$ として計算せよ．

(1) 積分

$$\int_{-\infty}^{\infty} \mathsf{p}(x|\,\theta)\,\mathrm{d}x$$

を計算せよ.
(2) W で定まる検定のサイズ (第 1 種の誤りの確率) α の値を小数第 3 位まで求めよ.
(3) W で定まる検定の検出力 $1-\beta$ (第 1 種の誤りの確率) を小数第 3 位まで求めよ.
(4) 尤度比

$$\Lambda(x) = \frac{\mathsf{p}(x|\,1)}{\mathsf{p}(x|\,0)}$$

の $x=0$ と $x=1$ における値を求めよ.
(5) 不等式 $\Lambda(x) > 2$ をみたす領域を求めることにより, W で与えらるは (1) で与えられた α を有意水準とする検定の中で最強力検定となることを Neyman-Peason の補題を用いて証明せよ.

演習問題 8.3. 連続型確率変数列 $X_1, X_2, \ldots, X_{16}$ は正規分布 $\mathsf{N}(\mu, 3^2)$ からのランダム標本とする. ただし, $-\infty < \mu < \infty$ で, これらの確率変数は確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上で定義されたものとする.
(1) 標本平均 $\overline{X}_{16} = (1/16)(X_1 + \cdots + X_{16})$ の期待値と分散を計算することにより, $\overline{X}_{16}$ の分布を述べよ.
(2) $a > 0$, b を定数とし, $Z = a\overline{X}_{16} + b$ としたとき, Z の分布が標準正規分布になるように定数 a, b を定めよ.
(3) 信頼係数 90% の μ の信頼区間を構成せよ.

演習問題 8.4. 連続型確率変数列 X_1, X_2, X_3 は正規母集団 $\mathsf{N}(\mu, 3)$ からの標本の大きさ 3 のランダム標本とする. 次の仮説検定問題を考える:

$$H_0 : \mu = 0 \quad \text{vs.} \quad H_1 : \mu = 2$$

このとき, 以下の問いに答えよ.
(1) H_0 と H_1 は単純仮説か複合仮説かを答えよ.
(2) $\overline{X}_3 = (1/3)(X_1 + X_2 + X_3)$ とする. $\bar{X}_3$ は帰無仮説 $H_0 : \mu = 0$ のもとでどのようなものになるかを答えよ. 理由も述べること.
(3) H_0 の棄却域を

$$C(t) = \{(x_1, x_2, x_3) \in \mathsf{R}^3; x_1 + x_2 + x_3 > t\}$$

としたとき, $C(t)$ が有意水準 0.1 の検定の棄却域になるように t をひとつ定めよ. t を求めるときには小数第 *3* 位を四捨五入せよ.

(4) 棄却域 $C(t)$ の検出力を 対立仮説 $H_1 : \mu = 2$ のもとで求めよ. ただし, 解答は関数 $\Phi(x) = \int_{-\infty}^{x} (1/\sqrt{2\pi}) e^{-t^2/2}\,\mathrm{d}t$ および t を用いて表現せよ. すなわち, 具体的な値を計算しなくともよい.

第9章　Bayes的推測

9.1　Bayes 的推測の考え方

前節までの統計的手法は頻度論的手法 (frequentist methods) と呼ばれるものである．頻度論的アプローチの考え方は以下のようにまとめられる．

【F1】確率は大量に観測されたデータの度数の極限と考える．

【F2】真の母数は未知だか, 固定した値であると考える．したがって真の母数は変動しないし, 真の母数に対する意味のある確率的な主張はない．

【F3】推測手法は大量にデータが観測されれば, うまく機能する保証があるように設計されている．たとえば信頼係数 0.95 の信頼区間は, 大量に観測されたデータの極限的な頻度において, 0.95 の信頼度が保証されている．

この考え方とは異なるアプローチの 1 つが Bayes 的推測である．Bayes 的アプローチは以下のような考え方に基づいている．

【B1】確率は極限的な頻度ではなく, 信頼の程度を表現するものであると考える．データはある確率変数の実現値と考えるだけでなく, それ以外の色々なものも確率変数の実現値と考える．

【B2】真の母数は固定された定数にも関わらず, 真の母数の確率的な主張をする．

【B3】真の母数に対して確率分布を想定し, 未知の母数の推測を行う．点推定値や区間推定値もこの分布からの実現値と考える．

確率に主観的な見方を導入する Bayes 的アプローチには, 頻度論的立場の主流派 (古典的な) 統計学者からの大きな批判がある．しかし, 統計学の隣接分野である機械学習やデータマイニングの分野では Bayes 的なアプローチが広く採用されている．

哲学的な議論は脇において, Bayes 的推測がどのような形式で行われるかをこの節ではみていく．

9.2 Bayes 的推測手法

Bayes 的推測は以下のステップに従い行われる. 母数モデルを $\{\mathsf{P}_\theta;\, \theta \in \Theta\}$ とする.

1. 事前分布 (prior distribution) と呼ばれる母数 θ についての分布 $\boldsymbol{\pi}(\theta)$ を想定する.

2. 母数 θ に与えられたときにデータの分布を表現する条件付きの分布 (頻度論的な立場では真の分布) $\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)$ を想定する.

3. 観測されたデータ $X_1 = x_1,\, X_2 = x_2,\, \ldots,\, X_n = x_n$ に基づき, θ の信頼度をアップデートする. すなわち, **事後分布** (posterior distribution) $\boldsymbol{\pi}_{\theta|\,\boldsymbol{X}}(\theta|\,\boldsymbol{x})$ を求める. ただし $\boldsymbol{x} = (x_1,\, x_2,\, \ldots,\, x_n)$ とした.

そして, 事後分布に基づき, 母数の推測を行う. 3 番目のステップがどのように行われるかを θ の事前分布とデータ $\boldsymbol{X}$ の分布が共に離散型の場合で説明する.

真の母数 θ は確率変数 $\boldsymbol{\Theta}$ の実現値と考える. いまは離散型確率変数の設定なので

$$
\begin{aligned}
\mathsf{Pr}\big(\boldsymbol{\Theta} = \theta|\, \boldsymbol{X} = \boldsymbol{x}\big) &= \frac{\mathsf{Pr}\big(\boldsymbol{\Theta} = \theta,\, \boldsymbol{X} = \boldsymbol{x}\big)}{\mathsf{Pr}(\boldsymbol{X} = \boldsymbol{x})} \\
&= \frac{\mathsf{Pr}\big(\boldsymbol{\Theta} = \theta,\, \boldsymbol{X} = \boldsymbol{x}\big)}{\sum_\theta \mathsf{Pr}\big(\boldsymbol{X} = \boldsymbol{x}\big|\, \boldsymbol{\Theta} = \theta\big)\mathsf{Pr}\big(\boldsymbol{\Theta} = \theta\big)}
\end{aligned}
$$

となる. 最後の等号は全確率の法則 (補題 1.11) を用いた. これを連続型分布の場合に形式的に書き直すと

$$
\boldsymbol{\pi}_{\boldsymbol{\Theta}|\,\boldsymbol{X}}(\theta|\,\boldsymbol{x}) = \frac{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)\boldsymbol{\pi}(\theta)}{\int \mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)\boldsymbol{\pi}(\theta)\,\mathrm{d}\theta}
$$

と書ける. ただし $\boldsymbol{\pi}_{\boldsymbol{\Theta}|\,\boldsymbol{X}}(\,\cdot\,|\,\boldsymbol{x})$ は $\boldsymbol{X} = \boldsymbol{x}$ が与えられたときの $\boldsymbol{\Theta}$ の条件付き p.d.f. であり, θ の事後分布である. $\mathsf{p}(\boldsymbol{x}|\,\theta)$ は真の母数が θ のときのデータの分布の p.d.f. で, $\boldsymbol{\pi}(\theta)$ は $\boldsymbol{\Theta}$ の分布の p.d.f. である.

n 個のランダム標本 $\boldsymbol{X} = (X_1,\, X_2,\, \ldots,\, X_n)$ に対して, 真の母数が θ のときの $\boldsymbol{X}$ の同時 p.d.f. $\mathsf{p}^{\boldsymbol{X}}(\,\cdot\,|\,\theta)$ とし, $X_1 = x_1,\, X_2 = x_2,\, \ldots,\, X_n = x_n$ を観測したときの尤度関数 $\mathsf{lik}_n(\theta)$ を

$$
\mathsf{lik}_n(\theta) = \mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)
$$

と書くことにする. よって

$$
\boldsymbol{\pi}_{\boldsymbol{\Theta}|\,\boldsymbol{X}}(\theta|\,\boldsymbol{x}) = \frac{\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)\boldsymbol{\pi}(\theta)}{\int \mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)\boldsymbol{\pi}(\theta)\,\mathrm{d}\theta} =: \frac{\mathsf{lik}_n(\theta)\boldsymbol{\pi}(\theta)}{c_n} \propto \mathsf{lik}_n(\theta)\boldsymbol{\pi}(\theta)
$$

と表すことができる．ただし

$$c_n := \int \mathsf{lik}_n(\theta)\boldsymbol{\pi}(\theta)\,\mathrm{d}\theta$$

は正規化定数と呼ばれる値である．c_n は θ に依存せず, データの実現値に依存した値である．よって事後分布の条件付き p.d.f. は事前分布と尤度関数の積の定数倍である．

$$\boldsymbol{\pi}_{\boldsymbol{\Theta}|\,\boldsymbol{X}}(\theta|\,\boldsymbol{x}) \propto \mathsf{lik}_n(\theta)\boldsymbol{\pi}(\theta).$$

c_n を無視してもよいのだろうか? 必要なときは求めることができるので問題ない．事後分布の平均やモード (最頻値) を推測に用いることが多い．たとえば,

$$\int \{y-\theta\}^2 \boldsymbol{\pi}_{\boldsymbol{\Theta}|\,\boldsymbol{X}}(\theta|\,\boldsymbol{x})\,\mathrm{d}\theta$$

を最小にする y の値を $\overline{\theta}$ と書くと

$$\overline{\theta} = \frac{\int \theta \boldsymbol{\pi}_{\boldsymbol{\Theta}|\,\boldsymbol{X}}(\theta|\,\theta)\,\mathrm{d}\theta}{\int \theta^2\, \boldsymbol{\pi}_{\boldsymbol{\Theta}|\,\boldsymbol{X}}(\theta|\,\theta)\,\mathrm{d}\theta} = \frac{\int \theta\, \mathsf{lik}_n(\theta)\boldsymbol{\pi}(\theta)\,\mathrm{d}\theta}{\int \theta^2\, \mathsf{lik}_n(\theta)\boldsymbol{\pi}(\theta)\,\mathrm{d}\theta}$$

と表現できる．また信頼区間であれば, $0 < \alpha < 1/2$ に対して

$$\int_{-\infty}^{a} \boldsymbol{\pi}_{\boldsymbol{\Theta}|\,\boldsymbol{X}}(\theta|\,\boldsymbol{x})\,\mathrm{d}\theta = \int_{b}^{\infty} \boldsymbol{\pi}_{\boldsymbol{\Theta}|\,\boldsymbol{X}}(\theta|\,\boldsymbol{x})\,\mathrm{d}\theta$$

をみたす $a,\, b$ を求めると開区間 $C = (a,\, b)$ は

$$\mathsf{Pr}\big(\theta \in C\big) = 1-\alpha$$

をみたす事後信頼区間となる．

例 9.1. $X_1,\, X_2,\, \ldots,\, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Ber}(\theta)\,(0 < \theta < 1)$ とし, 事前分布は $(0,\, 1)$ 上の一様分布とする．すなわち

$$\boldsymbol{\pi}(\theta) = \begin{cases} 1 & (0 < \theta < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

である．すると

$$\begin{aligned} \boldsymbol{\pi}_{\boldsymbol{\Theta}|\,\boldsymbol{X}}(\theta|\,\boldsymbol{x}) \propto& \boldsymbol{\pi}(\theta)\mathsf{lik}_n(\theta) = \prod_{j=1}^{n} \theta^{x_j}(1-\theta)^{1-x_j}\mathbb{1}_{\{0,\,1\}}(x_j) \\ &= \theta^{s}(1-\theta)^{n-s}\mathbb{1}_{\{0,\,1,\,\ldots,\,n\}}(s) \end{aligned}$$

となる．ただし $s = \sum_{j=1}^{n} x_j$ である．

一方 $\alpha > 0,\, \beta > 0$ とし

$$\boldsymbol{\pi}(\theta|\,\alpha,\,\beta) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} & (x > 0) \\ 0 & (\text{その他の場合}) \end{cases}$$

とする．ただし $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\,\mathrm{d}x$ である．すなわち θ の事前分布は $\mathsf{B}(\alpha,\,\beta)$ である．このとき

$$\boldsymbol{\pi}_{\boldsymbol{\Theta}|\,\boldsymbol{X}}(\theta|\,\boldsymbol{x}) = \frac{\Gamma(n+2)}{\Gamma(s+1)\Gamma(n-s+1)}\theta^{(s+1)-1}(1-\theta)^{(n-s+1)-1} \tag{9.1}$$

となる．したがって

$$\boldsymbol{\Theta}|\,\boldsymbol{X} = \boldsymbol{x} \sim \mathsf{B}(s+1,\,n-s+1)$$

がわかる．このとき

$$\overline{\theta} = \int \theta\boldsymbol{\pi}_{\boldsymbol{\Theta}|\,\boldsymbol{X}}(\theta|\,\boldsymbol{x})\,\mathrm{d}\theta$$

とおくと

$$\overline{\theta} = \frac{s+1}{n+2}$$

を得る．$\widehat{\theta} = \dfrac{s}{n},\, \widetilde{\theta} = \dfrac{1}{2}$ とおくと

$$\overline{\theta} = \lambda_n\widehat{\theta} + (1-\lambda)\widetilde{\theta}, \quad \lambda_n = \frac{n}{n+2}$$

と書ける．

信頼係数 0.95 の θ の事後信頼区間 $C = (a,\,b)$ は

$$\int_a^b \boldsymbol{\pi}_{\boldsymbol{\Theta}|\,\boldsymbol{X}}(\theta|\,\boldsymbol{x})\,\mathrm{d}\theta = 0.95$$

をみたす $a,\,b$ を数値計算で求めればよい．□

問 9.1. $X_1,\,X_2,\,\ldots,\,X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Ber}(\theta)\,(0 < \theta < 1)$ とし，

$$\boldsymbol{\pi}(\theta|\,\alpha,\,\beta) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} & (x > 0) \\ 0 & (\text{その他の場合}) \end{cases}$$

とする．このとき，$X_1 = x_1,\, X_2 = x_2,\, \ldots,\, X_n = x_n$ が与えられたときの $\boldsymbol{\Theta}$ の条件付き p.d.f. が (9.1) で与えられることを示せ．

問 9.2. $X \sim \mathsf{B}(\alpha,\,\beta)$ のとき，$\mathsf{E}[X] = \dfrac{\alpha}{\alpha+\beta}$ となることを示せ．

例 9.2. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}(\mu, \sigma^2)$ とする．ただし $\mu \in \mathbb{R}$ で σ^2 は既知とする．事前分布として

$$\boldsymbol{\Theta} \sim \mathbf{N}(a, b^2)$$

を仮定する．ただし $a \in \mathbb{R}, 0 < b < \infty$ である．すると

$$\begin{aligned}
&\boldsymbol{\Theta} | \boldsymbol{X} = \boldsymbol{x} \sim \mathsf{N}(\overline{\theta}, \tau^2), \qquad (9.2)\\
&\overline{\theta} = w\overline{x} + (1-w)a, \quad \overline{x} = \frac{1}{n}\sum_{j=1}^{n} x_j,\\
&w = \frac{\frac{1}{\mathsf{se}^2}}{\frac{1}{\mathsf{se}^2} + \frac{1}{b^2}}, \quad \frac{1}{\tau^2} = \frac{1}{\mathsf{se}^2} + \frac{1}{b^2}, \quad \mathsf{se} = \frac{\sigma}{\sqrt{n}}
\end{aligned}$$

となる．問 9.3 を参照のこと．ただし，$\boldsymbol{X} = (X_1, X_2, \ldots, X_n), \boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ とした．

このように事前分布と事後分布が同じ母数モデルに属するとき，事前分布はこのモデルに**随伴する** (conjugate) という．または，このような事前分布を随伴事前分布という．

$n \to \infty$ のとき $w \to 1$ かつ $\dfrac{\tau}{\mathsf{se}} \to 1$ となる．標本数が大きいとき

$$\boldsymbol{\Theta} | \boldsymbol{X} = \boldsymbol{x} \approx \mathsf{N}\big(\widehat{\theta}, \mathsf{se}^2\big)$$

となる．また n を固定する．$b \to \infty$ としたとき

$$\boldsymbol{\Theta} | \boldsymbol{X} = \boldsymbol{x} \approx \mathsf{N}\big(\widehat{\theta}, \mathsf{se}^2\big)$$

となる．これは一様な事前分布に対応するものである．

区間 $C = (c, d)$ は

$$\mathsf{Pr}\big(\theta \in C \big| \boldsymbol{X} = \boldsymbol{x}\big) = 0.95$$

をみたすものとする．したがって

$$\mathsf{Pr}\big(\theta < c \big| \boldsymbol{X} = \boldsymbol{x}\big) = 0.025, \quad \mathsf{Pr}\big(\theta > d \big| \boldsymbol{X} = \boldsymbol{x}\big) = 0.025$$

となるように c, d を選べばよい．このことから c を以下をみたすように選べばよい．

$$\mathsf{Pr}\big(\theta < c \big| \boldsymbol{X} = \boldsymbol{x}\big) = \mathsf{Pr}\left(\frac{\theta - \overline{\theta}}{\tau} < \frac{c - \overline{\theta}}{\tau} \middle| \boldsymbol{X} = \boldsymbol{x}\right) = \mathsf{Pr}\left(Z < \frac{c - \overline{\theta}}{\tau}\right).$$

ただし $Z \sim \mathsf{N}(0, 1)$ である．さらに

$$\mathsf{Pr}(Z < -1.96) = 0.025$$

なので

$$\frac{c-\overline{\theta}}{\tau} = -1.96$$

とすればよい．よって

$$c = \overline{\theta} - 1.96\tau$$

とする．同様に

$$d = \overline{\theta} + 1.96\tau$$

を得る．これらのことから信頼係数 0.95 の Bayes 的信頼区間は $(\overline{\theta} - 1.96\tau, \overline{\theta} + 1.96\tau)$ となる．さらに $\widehat{\theta} \approx \widetilde{\theta}$ かつ $\tau \approx \mathsf{se}$ なので信頼係数 0.95 の Bayes 的信頼区間は近似的に $(\widehat{\theta} - 1.96\mathsf{se}, \widehat{\theta} + 1.96\mathsf{se})$ となり，頻度論的信頼区間と同じになる．

問 9.3. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}(\mu, \sigma^2)$ とする．ただし $\mu \in \mathbb{R}$ で σ^2 は既知とする．さらに，事前分布として

$$\mathbf{\Theta} \sim \mathbf{N}(a, b^2)$$

を仮定する．ただし $a \in \mathbb{R}, 0 < b < \infty$ である．このとき，$\boldsymbol{X} = \boldsymbol{x}$ が与えられたときの $\mathbf{\Theta}$ の条件付き分布が (9.2) で与えられることを確かめよ．

9.3　事前分布の選択について

9.4　章末注釈と参考文献

この章は [23, pp.175 − 192] を借用した．

9.5　演習問題

演習問題 9.1.

演習問題 9.2.

第10章　大標本理論

この章では, 正則母数モデルにおける最尤推定量の一致性と漸近正規性について説明する. 節 10.1 では, 準備として, ランダム関数の確率収束について説明する. 尤度関数ないし対数尤度関数を未知母数の関数とみたして議論をするためのものである. 節 10.2 では, 前節の結果を利用して, 最尤推定量の一致性を証明する. 節 10.3 では, 最尤推定量の漸近正規性の証明を行う. 節 10.4 では, 帰無仮説のもとで尤度比統計量の対数の -2 倍が χ^2 分布に分布収束することを主張する Wilks の定理を証明する. 節 10.5 では, 指数分布族の仮定のもとで不完全データに基づく最尤推定値を求めるアルゴリズムである EM アルゴリズムを説明する.

10.1　ランダム関数の確率収束

$K \subset \mathbb{R}^d$ をコンパクト集合とし, $\{X_n\}_{n=1}^{\infty}$ を i.i.d. 確率変数列とする. $h\colon K \times \mathbb{R} \to \mathbb{R}$ を各 $x \in \mathbb{R}$ に対して

$$h(\,\cdot\,, x)\colon K \ni \boldsymbol{t} \mapsto h(\boldsymbol{t}, x) \in \mathbb{R}$$

が連続となるような関数とする. さらに, $n = 1, 2, \ldots$ に対して

$$W_n(\boldsymbol{t}) := h(\boldsymbol{t}, X_n) \qquad (\boldsymbol{t} \in K)$$

と定めると $\{W_n\}_{n=1}^{\infty}$ は K 上の i.i.d. ランダム関数列となる. また, $C(K)$ を K 上の連続関数全体がなす集合とする. したがって, $W_n\,(n \in \mathbb{N})$ は $C(K)$ 値ランダム要素の列となる.

集合 $C(K)$ に値を取るランダム要素はベクトルと同じように和, 差とスカラー倍が定義される. このような性質をもつ集合を線型空間という. さらに, 収束の概念も $C(K)$ に導入できる. $w \in C(K)$ に対して, w のノルム $\|w\|_\infty$ を

$$\|w\|_\infty := \sup_{\boldsymbol{t} \in K} |w(\boldsymbol{t})|$$

で定める. これを w の sup **ノルム**とよぶ. 関数列 $w, w_n \in C(K)\,(n = 1, 2, \ldots)$ に対して, $\{w_n\}_{n=1}^{\infty}$ が w に収束することを

$$\|w_n - w\|_\infty \xrightarrow{n\to\infty} 0$$

で定める. 線型空間 $C(K)$ はこのノルムに関して完備[1]となる. 完備な線型ノルム空間は Banach **空間**とよばれる. Banach 空間 $C(K)$ の最後のよい性質は可分性である. すなわち, $C(K)$ は稠密な可算部分集合をもつ.

次は, $C(K)$ に値をとる i.i.d. 関数列に対する大数の弱法則を証明するために必要な補題である.

補題 10.1. $K \subset \mathbb{R}^d$ をコンパクト部分集合とし, W を $C(K)$ 値ランダム要素とし

$$\mu(\boldsymbol{t}) = \mathsf{E}\big[W(\boldsymbol{t})\big] \qquad (\boldsymbol{t} \in C(K))$$

と定める.
(1) $\mathsf{E}\big[\|W\|_\infty\big] < \infty$ のとき, μ は $\boldsymbol{t} \in K$ に関して連続となる.
(2) $\epsilon > 0$ と $\mathsf{E}\big[\|W\|_\infty\big] < \infty$ とする. このとき

$$\sup_{\boldsymbol{t}\in K} \sup_{\boldsymbol{s}\in K:\, |\boldsymbol{s}-\boldsymbol{t}|_{2,\,d}<\epsilon} |W(\boldsymbol{s}) - W(\boldsymbol{t})| \xrightarrow{\epsilon\to 0} 0$$

となる.

Proof. (1) の証明: $\boldsymbol{t}_n, \boldsymbol{t} \in K\ (n = 1, 2, \ldots)$ とし

$$\lim_{n\to\infty} \boldsymbol{t}_n = \boldsymbol{t}$$

とする. ランダム関数 W は連続なので

$$W(\boldsymbol{t}_n) \overset{\text{a.s.}}{\to} W(\boldsymbol{t})$$

となる. さらに, $|W(\boldsymbol{t}_n)| \le \|W\|_\infty$ かつ $\mathsf{E}\big[\|W\|_\infty\big] < \infty$ であることに注意して, 優収束定理を用いると

$$\mu(\boldsymbol{t}_n) = \mathsf{E}\big[W(\boldsymbol{t}_n)\big] \xrightarrow{n\to\infty} \mathsf{E}\big[W(\boldsymbol{t})\big] = \mu(\boldsymbol{t})$$

となる. 点列 $\{\boldsymbol{t}_n\}_{n=1}^\infty$ は任意だったので, μ は連続であることがわかる.
(2) の証明: $\epsilon > 0$ に対して

$$M_\epsilon(\boldsymbol{t}) := \sup_{\boldsymbol{s}\in K:\, |\boldsymbol{s}-\boldsymbol{t}|_{2,\,d}<\epsilon} |W(\boldsymbol{s}) - W(\boldsymbol{t})|,$$
$$\lambda_\epsilon(\boldsymbol{t}) := \mathsf{E}\big[M_\epsilon(\boldsymbol{t})\big] \qquad (\boldsymbol{t} \in K)$$

と定める. W は $\boldsymbol{t} \in K$ の連続関数だったので, M_ϵ も $\boldsymbol{t} \in K$ の連続関数となる. また, $|M_\epsilon(\boldsymbol{t})| \le \|M_\epsilon\|_\infty$ かつ

$$\mathsf{E}\big[\|M_\epsilon\|_\infty\big] = \mathsf{E}\Big[\sup_{\boldsymbol{t}\in K} |M_\epsilon(\boldsymbol{t})|\Big] \le \mathsf{E}\Big[\sup_{\boldsymbol{t}\in K} \sup_{\boldsymbol{s}\in K:\, |\boldsymbol{s}-\boldsymbol{t}|_{2,\,d}<\epsilon} \big\{|W(\boldsymbol{s})| + |W(\boldsymbol{t})|\big\}\Big]$$
$$\le 2\mathsf{E}\big[\|W\|_\infty\big] < \infty$$

[1] $C(K)$ の任意の Cauchy 列は $C(K)$ のある元に必ず収束する.

であることに注意する. $\boldsymbol{t}\,(\in K)$ に収束する任意の K の点列 $\{\boldsymbol{t}_n\}_{n=1}^{\infty}$ を取る. 再度, 優収束定理を用いると

$$\begin{aligned}\lim_{n\to\infty}\lambda_\epsilon(\boldsymbol{t}_n) &= \lim_{n\to\infty}\mathsf{E}\big[M_\epsilon(\boldsymbol{t}_n)\big] = \mathsf{E}\big[\lim_{n\to\infty}M_\epsilon(\boldsymbol{t}_n)\big] \qquad (\because \text{優収束定理})\\ &= \mathsf{E}\big[M_\epsilon(\lim_{n\to\infty}\boldsymbol{t}_n)\big] \qquad (M_\epsilon \text{ の連続性})\\ &= \mathsf{E}\big[M_\epsilon(\boldsymbol{t})\big]\end{aligned}$$

となるので, λ_ϵ は $\boldsymbol{t}\,(\in K)$ の連続関数である. M_ϵ の定義から $\epsilon\to 0$ のとき, $\lambda_\epsilon \downarrow 0$ となる. したがって, Dini (定理 A.25) の定理から

$$\sup_{\boldsymbol{t}\in K}\lambda_\epsilon(\boldsymbol{t}) \downarrow 0 \qquad (\epsilon\to 0)$$

がわかる. □

以下の定理では, i.i.d. ランダム関数の平均が, その期待値に収束することを主張する. 各点での収束は大数の弱法則から直ちにわかる. しかし, 次の定理は, その収束がコンパクト集合 K に関して一様に確率収束することを主張するものである.

定理 10.2. $K\subset\mathbb{R}^d$ をコンパクト集合とする. $W, W_1, W_2,\ldots$ を i.i.d. $C(K)$ 値ランダム要素とする. さらに

$$\mu(\boldsymbol{t}) = \mathsf{E}\big[W(\boldsymbol{t})\big] \qquad (\boldsymbol{t}\in K)$$

とし

$$\mathsf{E}\big[\|W\|_\infty\big] < \infty$$

を仮定する. $n\in\mathbb{N}$ に対して

$$\overline{W}_n := \frac{1}{n}\big\{W_1+W_2+\cdots+W_n\big\}$$

とおいたとき

$$\|\overline{W}_n-\mu\|_\infty \xrightarrow{\mathrm{P}} 0 \tag{10.1}$$

が成り立つ.

Proof. $\epsilon>0$ を固定する. $\delta>0$ と $i=1,2,\ldots,n$ に対して

$$\begin{aligned}M_{\delta,i}(\boldsymbol{t}) &:= \sup_{\boldsymbol{s}\in K:\,|\boldsymbol{s}-\boldsymbol{t}|_{2,d}<\delta}|W_i(\boldsymbol{s})-W_i(\boldsymbol{t})|,\\ \lambda_\delta(\boldsymbol{t}) &:= \mathsf{E}\big[M_{\delta,i}(\boldsymbol{t})\big]\end{aligned}$$

と定める．補題 10.1(2) から δ を十分小さくとると

$$\lambda_\delta(\boldsymbol{t}) = \mathsf{E}\left[\sup_{\boldsymbol{s}\in K:\,|\boldsymbol{s}-\boldsymbol{t}|_{2,\,d}<\delta} |W(\boldsymbol{s}) - W(\boldsymbol{t})|\right] \le \epsilon \qquad (\forall \boldsymbol{t} \in K)$$

とできる．ここで

$$B(\boldsymbol{t};\,\delta) = \{\boldsymbol{s} \in K : |\boldsymbol{s}-\boldsymbol{t}|_{2,\,d} < \delta\}$$

とおく．すると K はコンパクトなので, K の開被覆 $\{B(\boldsymbol{t};\,\delta)\}_{\boldsymbol{t}\in K}$ から有限被覆を取ることができる．すなわち, $m \in \mathbb{N}$ と $\boldsymbol{t}_1,\,\boldsymbol{t}_2,\,\ldots,\,\boldsymbol{t}_n \in K$ が存在して

$$K \supset \bigcup_{i\in\{1,\,2,\,\ldots,\,m\}} B(\boldsymbol{t}_i;\,\delta)$$

となる．$O_i = B(\boldsymbol{t}_i;\,\delta)\,(i = 1,\,2,\,\ldots,\,m)$ とおく．このとき

$$\begin{aligned}
\|\overline{W}_n - \mu\|_\infty &= \max_{i\in\{1,2,\ldots,m\}} \sup_{\boldsymbol{t}\in O_i} |\overline{W}_n(\boldsymbol{t}) - \mu(\boldsymbol{t})| \\
&\le \max_{i\in\{1,2,\ldots,m\}} \sup_{\boldsymbol{t}\in O_i} \Bigl[|\overline{W}_n(\boldsymbol{t}) - \overline{W}_n(\boldsymbol{t}_i)| + |\overline{W}_n(\boldsymbol{t}_i) - \mu(\boldsymbol{t}_i)| \\
&\qquad\qquad + |\mu(\boldsymbol{t}_i) - \mu(\boldsymbol{t})|\Bigr] \\
&\le \max_{i\in\{1,2,\ldots,m\}} \sup_{\boldsymbol{t}\in O_i} |\overline{W}_n(\boldsymbol{t}) - \overline{W}_n(\boldsymbol{t}_i)| \\
&\qquad + \max_{i\in\{1,2,\ldots,m\}} |\overline{W}_n(\boldsymbol{t}_i) - \mu(\boldsymbol{t}_i)| + \epsilon \qquad (\because \mu \text{ の連続性})
\end{aligned}$$

となる．さらに

$$\begin{aligned}
\sup_{\boldsymbol{t}\in O_i} |\overline{W}_n(\boldsymbol{t}) - \overline{W}_n(\boldsymbol{t}_i)| &= \frac{1}{n} \sup_{\boldsymbol{t}\in O_i} \left| \sum_{j=1}^n \bigl(W_j(\boldsymbol{t}) - W_j(\boldsymbol{t}_i)\bigr) \right| \\
&\le \frac{1}{n} \sum_{j=1}^n \sup_{\boldsymbol{t}\in O_i} |W_j(\boldsymbol{t}) - W_j(\boldsymbol{t}_i)| = \frac{1}{n} \sum_{j=1}^n M_{\delta,\,j}(\boldsymbol{t}_i) \\
&=: \overline{M}_{\delta,\,n}(\boldsymbol{t}_i)
\end{aligned}$$

となる．大数の弱法則から

$$\overline{M}_{\delta,\,n}(\boldsymbol{t}_i) \xrightarrow{\mathrm{P}} \lambda_\delta(\boldsymbol{t}_i) < \epsilon \qquad (i = 1,\,2,\,\ldots,\,m)$$

が成立する．これらから

$$
\begin{aligned}
\|\overline{W}_n - \mu\|_\infty &\le \max_{i\in\{1,2,\dots,m\}} \underbrace{\sup_{\boldsymbol{t}\in O_i} |\overline{W}_n(\boldsymbol{t}) - \overline{W}_n(\boldsymbol{t}_i)|}_{\le \overline{M}_{\delta,n}(\boldsymbol{t}_i)} \\
&\quad + \max_{i\in\{1,2,\dots,m\}} |\overline{W}_n(\boldsymbol{t}_i) - \mu(\boldsymbol{t}_i)| + \epsilon \\
&\le \max_{i\in\{1,2,\dots,m\}} \overline{M}_{\delta,n}(\boldsymbol{t}_i) + \max_{i\in\{1,2,\dots,m\}} |\overline{W}_n(\boldsymbol{t}_i) - \mu(\boldsymbol{t}_i)| + \epsilon \\
&\le \max_{i\in\{1,2,\dots,m\}} |\overline{M}_{\delta,n}(\boldsymbol{t}_i) - \lambda_\delta(\boldsymbol{t}_i)| + \max_{i\in\{1,2,\dots,m\}} \underbrace{\lambda_\delta(\boldsymbol{t}_i)}_{\le\epsilon} \\
&\quad + \max_{i\in\{1,2,\dots,m\}} |\overline{W}_n(\boldsymbol{t}_i) - \mu(\boldsymbol{t}_i)| + \epsilon \\
&\le 2\epsilon + \max_{i\in\{1,2,\dots,m\}} |\overline{M}_{\delta,n}(\boldsymbol{t}_i) - \lambda_\delta(\boldsymbol{t}_i)| \\
&\quad + \max_{i\in\{1,2,\dots,m\}} |\overline{W}_n(\boldsymbol{t}_i) - \mu(\boldsymbol{t}_i)|
\end{aligned}
$$

となる．上の不等式の最右辺の最後の 2 項はともに 0 に確率収束するので

$$
\mathsf{Pr}\big(\|\overline{W}_n - \mu\|_\infty > 3\epsilon\big) \overset{n\to\infty}{\longrightarrow} 0
$$

がわかる． □

注意 10.3. 大数の強法則から

$$
\|\overline{W}_n - \mu\|_\infty \overset{\text{a.s.}}{\to} 0
$$

も同様に証明できる． □

次の結果は一様確率収束の有効性を示す．以下の結果は各点の確率収束では成立しない．

定理 10.4. $K \subset \mathbb{R}^d$ をコンパクト集合とし，$\{G_n\}_{n=1}^\infty$ を $C(K)$ 値ランダム要素とし，ある非ランダム関数 $g\,(\in C(K))$ が存在して

$$
\|G_n - g\|_\infty \overset{\text{P}}{\to} 0
$$

が成立すると仮定する．このとき，以下が成立する．
(1) 確率ベクトル列 $\{\boldsymbol{t}_n\}_{n=1}^\infty \subset K$ はある定数ベクトル $\boldsymbol{t}^* \in K$ に確率収束するとする．このとき

$$
G_n(\boldsymbol{t}_n) \overset{\text{P}}{\to} g(\boldsymbol{t}^*)
$$

となる．

(2) 非ランダムな関数 g は $\boldsymbol{t}^* \in K$ で唯一最大値を取るとし, 確率ベクトル $\boldsymbol{t}_n\,(n = 1, 2, \ldots)$ で G_n は最大値を取るとする. すなわち

$$G_n(\boldsymbol{t}_n) = \sup_{\boldsymbol{t}\in K} G_n(\boldsymbol{t})$$

である. このとき

$$\boldsymbol{t}_n \xrightarrow{\mathrm{P}} \boldsymbol{t}$$

が成立する.
(3) $K \subset \mathbb{R}$ とし, $t^* \in K$ は $g(t) = 0$ の唯一の解とし, 確率変数 t_n は $G_n(t) = 0$ の解とする. このとき

$$t_n \xrightarrow{\mathrm{P}} t^*$$

が成立する.

Proof. (1) の証明: まず

$$\begin{aligned}|G_n(\boldsymbol{t}_n) - g(\boldsymbol{t}^*)| &\le |G_n(\boldsymbol{t}_n) - g(\boldsymbol{t}_n)| + |g(\boldsymbol{t}_n) - g(\boldsymbol{t}^*)| \\ &\le \|G_n - g\|_\infty + |g(\boldsymbol{t}_n) - g(\boldsymbol{t}^*)|\end{aligned}$$

となることに注意する. $\boldsymbol{t}_n \xrightarrow{\mathrm{P}} \boldsymbol{t}^*$ と g の連続性から

$$g(\boldsymbol{t}_n) \xrightarrow{\mathrm{P}} g(\boldsymbol{t}^*)$$

となる. したがって, $\forall\epsilon > 0$ に対して

$$\begin{aligned}\mathsf{Pr}\big(|G_n(\boldsymbol{t}_n) - g(\boldsymbol{t}^*)| > \epsilon\big) &\le \mathsf{Pr}\big(\|G_n - g\|_\infty + |g(\boldsymbol{t}_n) - g(\boldsymbol{t}^*)| > \epsilon\big) \\ &\le \mathsf{Pr}\left(\|G_n - g\|_\infty > \frac{\epsilon}{2}\right) + \mathsf{Pr}\left(|g(\boldsymbol{t}_n) - g(\boldsymbol{t}^*)| > \frac{\epsilon}{2}\right) \\ &\xrightarrow{n\to\infty} 0\end{aligned}$$

がわかる. よって, (1) は示された.
(2) の証明: $\epsilon > 0$ を固定し

$$K_\epsilon := K \setminus B(\boldsymbol{t}^*; \epsilon) = K \cap \{B(\boldsymbol{t}^*; \epsilon)\}^c$$

とおく. K は有界なので, K_ϵ も有界である. また, K_ϵ は 2 つの閉集合の共通部分なので, 閉集合となる. これらのことから K_ϵ もコンパクトとなる. ここで

$$M := g(\boldsymbol{t}^*) = \sup_{\boldsymbol{t}\in K} g(\boldsymbol{t}); \qquad M_\epsilon := \sup_{\boldsymbol{t}\in K_\epsilon} g(\boldsymbol{t})$$

とおく. K_ϵ はコンパクトなので, ある $\boldsymbol{t}_\epsilon^* \in K_\epsilon$ が存在して

$$M_\epsilon = g(\boldsymbol{t}_\epsilon^*)$$

となる. g は K 上で唯一の点 $\boldsymbol{t}^*$ で最大値を取るので, $\boldsymbol{t}_\epsilon^* \neq \boldsymbol{t}^*$ となる. よって

$$M_\epsilon < M$$

となる. ここで

$$\delta := M - M_\epsilon > 0$$

とおく. $\|G_n - g\| < \dfrac{\delta}{2}$ が成り立てば

$$\begin{aligned} \sup_{\boldsymbol{t} \in K_\epsilon} G_n(\boldsymbol{t}) &= \sup_{\boldsymbol{t} \in K_\epsilon} \bigl\{ g(\boldsymbol{t}) + G_n(\boldsymbol{t}) - g(\boldsymbol{t}) \bigr\} \\ &\leq \sup_{\boldsymbol{t} \in K_\epsilon} g(\boldsymbol{t}) + \|G_n - g\|_\infty \leq M_\epsilon + \frac{\delta}{2} = M - \frac{\delta}{2} \end{aligned} \tag{10.2}$$

となる. 一方

$$\begin{aligned} \sup_{\boldsymbol{t} \in K} G_n(\boldsymbol{t}) &\geq G_n(\boldsymbol{t}^*) = g(\boldsymbol{t}^*) + G_n(\boldsymbol{t}^*) - g(\boldsymbol{t}^*) \geq g(\boldsymbol{t}^*) - \sup_{\boldsymbol{t} \in K} |G(\boldsymbol{t}) - g(\boldsymbol{t})| \\ &= g(\boldsymbol{t}^*) - \|G_n - g\|_\infty > g(\boldsymbol{t}^*) - \frac{\delta}{2} \end{aligned} \tag{10.3}$$

となる. (10.2) と (10.3) から

$$\sup_{\boldsymbol{t} \in K} G_n(\boldsymbol{t}) > \sup_{\boldsymbol{t} \in K_\epsilon} G_n(\boldsymbol{t})$$

となるので

$$\boldsymbol{t}_n \in B(\boldsymbol{t}^*;\, \epsilon)$$

がわかる. したがって

$$\|G_n - g\|_\infty < \frac{\delta}{2} \;\Rightarrow\; |\boldsymbol{t}_n - \boldsymbol{t}^*|_{2,\,d} < \epsilon$$

となるので

$$\mathsf{Pr}\left(\|G_n - g\|_\infty < \frac{\delta}{2} \right) \leq \mathsf{Pr}\bigl(|\boldsymbol{t}_n - \boldsymbol{t}^*|_{2,\,d} < \epsilon \bigr)$$

がわかる. それぞれの事象の補事象を取ると

$$\mathsf{Pr}\bigl(|\boldsymbol{t}_n - \boldsymbol{t}^*|_{2,\,d} \geq \epsilon \bigr) \leq \mathsf{Pr}\left(\|G_n - g\|_\infty \geq \frac{\delta}{2} \right) \xrightarrow{n \to \infty} 0$$

となる．よって

$$\boldsymbol{t}_n \xrightarrow{\mathrm{P}} \boldsymbol{t}^*$$

が示せた．

(3) の証明: 一般性を失わず, G_n は t_n の近傍で非減少とし, g は t^* の近傍で非減少と仮定できる．$\epsilon > 0$ と $\delta > 0$ を

$$g(t^* - \epsilon) < -2\delta \quad \text{かつ} \quad g(t^* + \epsilon) > 2\delta$$

となるように取る．G_n の非減少性から

$$G_n(t^* - \epsilon) < -\delta \quad \text{かつ} \quad t_n \leq t^* - \epsilon \Rightarrow G_n(t_n) < -\delta$$

となる．同様に

$$G_n(t^* + \epsilon) > \delta \,\text{かつ}\, t_n \geq t^* + \epsilon \Rightarrow G_n(t_n) > \delta$$

となる．また

$$\begin{aligned}
&\mathsf{Pr}\big(G_n(t^* - \epsilon) < -\delta),\, G_n(t^* + \epsilon) > \delta\big) \\
&\qquad \leq \mathsf{Pr}(t^* - \epsilon < t_n < t^* + \epsilon) + \mathsf{Pr}\big(\{G_n(t^* - \epsilon) < -\delta\} \cap \{t_n \leq t^* - \epsilon\}\big) \\
&\qquad + \mathsf{Pr}\big(\{G_n(t^* + \epsilon) > \delta\} \cap \{t_n \geq t^* + \epsilon\}\big) \\
&\qquad \leq \mathsf{Pr}(t^* - \epsilon < t_n < t^* + \epsilon) + \mathsf{Pr}\big(G_n(t_n) < -\delta\big) + \mathsf{Pr}\big(G_n(t_n) > \delta\big)
\end{aligned} \tag{10.4}$$

となる．しかし, $\|G_n - g\|_\infty < \delta$ が起こると

$$G_n(t^* - \epsilon) \leq g(t^* - \epsilon) + |G_n(t^* - \epsilon) - g(t^* - \epsilon)| < -2\delta + \|G_n - g\|_\infty = -\delta$$

となる．よって

$$\mathsf{Pr}(\|G_n - g\|_\infty < \delta) \leq \mathsf{Pr}(G_n(t^* - \epsilon) < -\delta)$$

となる．また

$$G_n(t^* + \epsilon) \geq g(t^* + \epsilon) + |G_n(t^* + \epsilon) - g(t^* + \epsilon)| > 2\delta - \|G_n - g\|_\infty = \delta$$

となる．よって

$$\mathsf{Pr}(G_n(t^* - \epsilon) < -\delta) \overset{n\to\infty}{\longrightarrow} 1 \qquad \mathsf{Pr}(G_n(t^* + \epsilon) > \delta) \overset{n\to\infty}{\longrightarrow} 1$$

となるので

$$\mathsf{Pr}\big(G_n(t^* - \epsilon) < -\delta); \qquad G_n(t^* + \epsilon) > \delta\big) \overset{n\to\infty}{\longrightarrow} 1$$

となる．さらに

$$\mathsf{Pr}\big(G_n(t_n) < -\delta\big) \overset{n\to\infty}{\longrightarrow} 0; \qquad \mathsf{Pr}\big(G_n(t_n) > \delta\big) \overset{n\to\infty}{\longrightarrow} 0$$

であることと注意すると (10.4 から

$$\mathsf{Pr}(t^* - \epsilon < t_n < t^* + \epsilon) \overset{n\to\infty}{\longrightarrow} 1$$

がわかる． □

10.2 最尤推定量の一致性

$\Theta \subset \mathbb{R}^d$ をコンパクト集合とし, 正則統計的モデル $\mathcal{P} = \{\mathsf{p}(\,\cdot\,|\,\boldsymbol{\theta});\, \boldsymbol{\theta} \in \Theta\}$ を考える. ただし, $\mathsf{p}(\,\cdot\,|\,\boldsymbol{\theta})$ は $\mathbb{R}$ 上の p.d.f. または p.m.f. とする. また, $\mathsf{p}(\,\cdot\,|\,\boldsymbol{\theta})$ に対応する $\mathbb{R}$ 上の確率測度を $\mathsf{P}_{\boldsymbol{\theta}}$ と記す. $\boldsymbol{\theta}^* \in \Theta$ を固定して

$$X_1,\, X_2,\, \ldots \overset{\text{i.i.d.}}{\sim} \mathsf{p}(\,\cdot\,|\,\boldsymbol{\theta}^*)$$

とし, $\boldsymbol{X} = (X_1,\, X_2,\, \ldots,\, X_n)\,(n \in \mathbb{N})$ に基づく対数尤度 $\ell_n(\boldsymbol{\theta})$ を

$$\ell_n(\boldsymbol{\theta}) = \log \prod_{j=1}^{n} \mathsf{p}(X_j|\,\boldsymbol{\theta}) = \sum_{j=1}^{n} \log \mathsf{p}(X_j|\,\boldsymbol{\theta})$$

で定める. このとき, $\boldsymbol{\theta}^*$ の最尤推定量 $\widehat{\boldsymbol{\theta}}_n = \widehat{\boldsymbol{\theta}}_n(X_1,\, X_2,\, \ldots,\, X_n)$ は対数尤度 ℓ_n を最大化する. 統計的モデル $\mathcal{P}$ の正則性から, ほとんど至ところの $x \in \mathbb{R}$ に対して, 写像

$$\Theta \ni \boldsymbol{\theta} \mapsto \mathsf{p}(x|\,\boldsymbol{\theta})$$

は連続である.

定義 10.5. Kullback-Leibler 情報量を

$$\mathsf{KL}(\boldsymbol{\theta}^*,\, \boldsymbol{\theta}) := \mathsf{E}_{\boldsymbol{\theta}^*}\left[\log\left\{\frac{\mathsf{p}(X|\,\boldsymbol{\theta}^*)}{\mathsf{p}(X|\,\boldsymbol{\theta})}\right\}\right]$$

で定める.

補題 10.6. $\forall \boldsymbol{\theta} \in \Theta$ に対して, $\mathsf{P}_{\boldsymbol{\theta}^*} \neq \mathsf{P}_{\boldsymbol{\theta}}$ ならば, $\mathsf{KL}(\boldsymbol{\theta}^*,\, \boldsymbol{\theta}) > 0$ となる.

Proof. Jensen の不等式から

$$\begin{aligned}
-\mathsf{KL}(\boldsymbol{\theta}^*,\, \boldsymbol{\theta}) &= \mathsf{E}_{\boldsymbol{\theta}^*}\left[\log\left\{\frac{\mathsf{p}(X|\,\boldsymbol{\theta})}{\mathsf{p}(X|\,\boldsymbol{\theta}^*)}\right\}\right] \leq \log \mathsf{E}_{\boldsymbol{\theta}^*}\left[\frac{\mathsf{p}(X|\,\boldsymbol{\theta})}{\mathsf{p}(X|\,\boldsymbol{\theta}^*)}\right] \\
&= \log \int_{x \in \mathbb{R};\, \mathsf{p}(x|\,\boldsymbol{\theta}^*) > 0} \frac{\mathsf{p}(x|\,\boldsymbol{\theta})}{\mathsf{p}(x|\,\boldsymbol{\theta}^*)} \mathsf{p}(x|\,\boldsymbol{\theta}^*)\, \mathrm{d}x \\
&= \log \int_{x \in \mathbb{R};\, \mathsf{p}(x|\,\boldsymbol{\theta}^*) > 0} \mathsf{p}(x|\,\boldsymbol{\theta}^*)\, \mathrm{d}x \leq \log 1 = 0
\end{aligned}$$

となる. 上の不等式の等号成立条件は $\dfrac{\mathsf{p}(x|\,\boldsymbol{\theta})}{\mathsf{p}(x|\,\boldsymbol{\theta}^*)}$ がほとんど至ところの $x \in \mathbb{R}$ において定数のときである. この場合には, $\mathsf{P}_{\boldsymbol{\theta}} = \mathsf{P}_{\boldsymbol{\theta}^*}$ となるので

$$\mathsf{KL}(\boldsymbol{\theta}^*,\, \boldsymbol{\theta}) > 0$$

がわかる. □

$\boldsymbol{\theta} \in \Theta$ と $X \sim \mathsf{p}(\cdot\,|\,\boldsymbol{\theta}^*)$ に対して

$$W(\boldsymbol{\theta}) = \log\left[\frac{\mathsf{p}(X|\,\boldsymbol{\theta})}{\mathsf{p}(X|\,\boldsymbol{\theta}^*)}\right]$$

と定める.

定理 10.7. 以下を仮定する.
(a) $\Theta \subset \mathbb{R}^d$ はコンパクト集合.
(b) $\mathsf{E}_{\boldsymbol{\theta}^*}\left[\|W\|_\infty\right] < \infty$.
(c) ほとんど至るところの $x \in \mathbb{R}$ で写像

$$\Theta \ni \boldsymbol{\theta} \mapsto \mathsf{p}(x|\,\boldsymbol{\theta}) \in \mathbb{R}$$

は連続である.
(d) $\forall \boldsymbol{\theta} \in \Theta$ に対して

$$\boldsymbol{\theta} \neq \boldsymbol{\theta}^* \Rightarrow \mathsf{P}_{\boldsymbol{\theta}} \neq \mathsf{P}_{\boldsymbol{\theta}^*}$$

である.
このとき, $\boldsymbol{\theta}^*$ の最尤推定量 $\widehat{\boldsymbol{\theta}}$ に対して

$$\widehat{\boldsymbol{\theta}} \xrightarrow{\mathrm{P}} \boldsymbol{\theta}^*$$

が成り立つ.

Proof. $n \in \mathbb{N}$ とし, $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{p}(\cdot\,|\,\boldsymbol{\theta}^*)$ とする. $j = 1, 2, \ldots, n$ に対して

$$W_j(\boldsymbol{\theta}) = \log\left[\frac{\mathsf{p}(X_j|\,\boldsymbol{\theta})}{\mathsf{p}(X_j|\,\boldsymbol{\theta}^*)}\right] \qquad (\boldsymbol{\theta} \in \Theta)$$

とおく. このとき, $\{W_j(\boldsymbol{\theta})\}_{j=1}^n$ は i.i.d. の $C(\Theta)$ 値ランダム要素で

$$\mu(\boldsymbol{\theta}) = \mathsf{E}[W_j(\boldsymbol{\theta})] = -\mathsf{KL}(\boldsymbol{\theta}^*, \boldsymbol{\theta}) \qquad (\boldsymbol{\theta} \in \Theta)$$

と記す. 仮定 (d) と補題 10.6 から

$$\boldsymbol{\theta} \neq \boldsymbol{\theta}^* \Rightarrow \mu(\boldsymbol{\theta}) < 0$$

であり, $\mu(\boldsymbol{\theta}^*) = 0$ である. よって, μ は $\boldsymbol{\theta}^*$ で唯一の最大値を取る. また

$$\overline{W}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{j=1}^n W_j(\boldsymbol{\theta}) = \frac{\ell_n(\boldsymbol{\theta}) - \ell_n(\boldsymbol{\theta}^*)}{n}$$

なので, $\overline{W}_n$ は $\widehat{\boldsymbol{\theta}}_n$ で最大値を取る. 定理 10.2 から

$$\|\overline{W}_n - \mu\|_\infty \xrightarrow{\mathrm{P}} 0$$

となる. よって, 定理 10.4(2) から

$$\widehat{\boldsymbol{\theta}}_n \xrightarrow{\mathrm{P}} \boldsymbol{\theta}^*$$

がわかる. □

定理 10.8. 統計的モデル $\mathcal{P} = \{\mathsf{p}(\cdot\,|\,\boldsymbol{\theta});\, \boldsymbol{\theta} \in \Theta = \mathbb{R}^d\}$ を考える．ただし，$\mathsf{p}(\cdot\,|\,\boldsymbol{\theta})$ は p.d.f. または p.m.f. とする．$\mathsf{p}(\cdot\,|\,\boldsymbol{\theta})$ 対応する測度空間 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ と $\left(\Omega, \mathcal{A}\right)$ 上の確率測度をそれぞれ $\mathsf{P}_{\boldsymbol{\theta}}$ と $\mathsf{Pr}_{\boldsymbol{\theta}}$ と記す．$\boldsymbol{\theta}^* \in \Theta$ とし

$$
\begin{aligned}
&X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{p}(\cdot\,|\,\boldsymbol{\theta}^*),\\
&W_j(\boldsymbol{\theta}) = \log\left[\frac{\mathsf{p}(X_j|\,\boldsymbol{\theta})}{\mathsf{p}(X_j|\,\boldsymbol{\theta}^*)}\right] \qquad (j = 1, 2, \ldots, n),\\
&W(\boldsymbol{\theta}) = \log\left[\frac{\mathsf{p}(X|\,\boldsymbol{\theta})}{\mathsf{p}(X|\,\boldsymbol{\theta}^*)}\right] \qquad (\boldsymbol{\theta} \in \Theta)
\end{aligned}
$$

とおく．　たたし，X は generic な確率変数で $X \sim \mathsf{p}(\cdot\,|\,\boldsymbol{\theta}^*)$ である．以下を仮定する．

(a) ほとんどいたるところの $x \in \mathbb{R}$ に関して，写像

$$\Theta \ni \boldsymbol{\theta} \mapsto \mathsf{p}(x|\,\boldsymbol{\theta}) \in [0, \infty)$$

は連続関数である．

(b) $\forall \boldsymbol{\theta} \neq \boldsymbol{\theta}^*\, (\boldsymbol{\theta} \in \Theta)$ に対して，$\mathsf{P}_{\boldsymbol{\theta}} \neq \mathsf{P}_{\boldsymbol{\theta}^*}$ である．

(c) $|\boldsymbol{\theta}|_{2,\,d} \to \infty$ のとき，$\mathsf{p}(x|\,\boldsymbol{\theta}) \to 0$ となる．

(d) 任意の $\mathbb{R}^d$ のコンパクト集合 K に対して

$$\mathsf{E}_{\boldsymbol{\theta}^*}\left[\|\mathbb{1}_K \cdot W\|_\infty\right] < \infty$$

となる．ただし，$g \in C(\mathbb{R}^d)$ に対して $\|g\|_\infty = \sup_{\boldsymbol{x}\in\mathbb{R}^d} |g(\boldsymbol{x})|$ である．

(e) ある $a > 0$ が存在して

$$\mathsf{E}_{\boldsymbol{\theta}^*}\left[\sup_{|\boldsymbol{\theta}|_{2,\,d} > a} W(\boldsymbol{\theta})\right] < \infty$$

となる．

このとき，$\boldsymbol{\theta}^*$ の最尤推定量 $\widehat{\boldsymbol{\theta}}_n$ は，$\mathsf{P}_{\boldsymbol{\theta}^*}$ のもと

$$\widehat{\boldsymbol{\theta}}_n \xrightarrow{\mathrm{P}} \boldsymbol{\theta}^*$$

となる．

Proof. $|\boldsymbol{\theta}|_{2,\,d} \to \infty$ のとき，$\mathsf{p}(x|\,\boldsymbol{\theta}) \to 0$ なので，$\mathsf{p}(X|\,\boldsymbol{\theta}^*) > 0$ のとき

$$\lim_{b\to\infty} \sup_{|\boldsymbol{\theta}|_{2,\,d} > b} W(\boldsymbol{\theta}) = -\infty$$

となる．仮定 (e) から優収束定理を用いると

$$\lim_{b\to\infty} \mathsf{E}_{\boldsymbol{\theta}^*}\left[\sup_{|\boldsymbol{\theta}|_{2,\,d} > b} W(\boldsymbol{\theta})\right] = -\infty$$

がわかる．よって, 十分大きな $b_0 > 0$ に対して

$$\mathsf{E}_{\boldsymbol{\theta}^*}\left[\sup_{|\boldsymbol{\theta}|_{2,\,d} > b_0} W(\boldsymbol{\theta})\right] < 0$$

とできる．$W(\boldsymbol{\theta}^*) = 0$ なので, $b_0 > |\boldsymbol{\theta}^*|_{2,\,d}$ である．さらに

$$\sup_{|\boldsymbol{\theta}|_{2,\,d} > b_0} \overline{W}_n(\boldsymbol{\theta}) \leq \frac{1}{n}\sum_{j=1}^{n} \sup_{|\boldsymbol{\theta}|_{2,\,d} > b_0} W_j(\boldsymbol{\theta}) \xrightarrow{\mathrm{P}} \mathsf{E}_{\boldsymbol{\theta}^*}\left[\sup_{|\boldsymbol{\theta}|_{2,\,d} > b_0} W(\boldsymbol{\theta})\right]$$

となるので

$$\mathsf{Pr}_{\boldsymbol{\theta}^*}\left(\sup_{|\boldsymbol{\theta}|_{2,\,d} > b_0} \overline{W}_n(\boldsymbol{\theta})\right) \xrightarrow{n\to\infty} 0$$

となる．いま, K を $\boldsymbol{\theta}^*$ を中心とした半径 $\delta\,(>0)$ の閉球とし, $\widetilde{\boldsymbol{\theta}}_n$ を K 上で $\overline{W}_n$ を最大にする確率ベクトルとする．定理 10.7 から

$$\widetilde{\boldsymbol{\theta}}_n \xrightarrow{\mathrm{P}} \boldsymbol{\theta}^*$$

となる．さらに

$$\sup_{|\boldsymbol{\theta}|_{2,\,d} > b_0} \overline{W}_n(\boldsymbol{\theta}) < \overline{W}_n(\boldsymbol{\theta}^*) = 0$$

のとき, $\widehat{\boldsymbol{\theta}}_n \in K$ となる．よって

$$\mathsf{Pr}_{\boldsymbol{\theta}^*}\big(\widehat{\boldsymbol{\theta}}_n = \widetilde{\boldsymbol{\theta}}_n\big) \xrightarrow{n\to\infty} 1$$

となる．よって

$$\widehat{\boldsymbol{\theta}}_n \xrightarrow{\mathrm{P}} \boldsymbol{\theta}^*$$

がわかる． □

10.3　最尤推定量の漸近正規性

定理 10.9. 統計的モデル $\mathcal{P} = \{\mathsf{p}(\,\cdot\,|\,\theta);\, \theta \in \Theta)\}$ を考える．ただし, $\Theta \subset \mathbb{R}$ で p は $\mathbb{R}$ 上の p.d.f. または p.m.f. である．以下を仮定する．
(a) $X_1, X_2, \ldots \overset{\text{i.i.d.}}{\sim} \mathsf{p}(\,\cdot\,|\,\theta^*)$ とする．ただし, $\theta^* \in \Theta \subset \mathbb{R}$ である．
(b) 集合 $A := \{x \in \mathbb{R};\, \mathsf{p}(x|\,\theta) > 0\}$ は θ に依存しない．
(c) すべての $x \in A$ に対して, $\mathsf{p}(x|\,\theta)$ は θ の 2 回連続微分可能な関数とする．
(d) $X \sim \mathsf{p}(\,\cdot\,|\,\theta^*)$ に対して

$$W(\theta) = \log \mathsf{p}(X|\,\theta) \qquad (\theta \in \Theta)$$

と定める. 1 つの標本に基づく Fisher 情報量 $I(\theta^*)$ は有界で

$$I(\theta^*) = \mathsf{E}_{\theta^*}\big[\overset{\bullet}{W}(\theta^*)\big] \quad \text{または} \quad I(\theta^*) = -\mathbb{E}_{\theta^*}\big[\overset{\bullet\bullet}{W}(\theta^*)\big]$$

と表現できる. ただし

$$\overset{\bullet}{W}(\theta) = \frac{\mathrm{d}W}{\mathrm{d}\theta}(\theta); \qquad \overset{\bullet\bullet}{W}(\theta) = \frac{\mathrm{d}^2W}{\mathrm{d}\theta^2}(\theta)$$

と定めた.
(e) Θ のすべての内点 θ^* に対して, ある ϵ が存在して

$$\mathsf{E}_{\theta^*}\big[\|\mathbb{1}_{[\theta^*-\epsilon,\,\theta^*+\epsilon]}\overset{\bullet\bullet}{W}\|_\infty\big] < \infty$$

とする.
(f) θ^* の最尤推定量 $\widehat{\theta}_n$ は一致性をもつ.
このとき, Θ の任意の内点 θ^* に対して, P_{θ^*} のもとで

$$\sqrt{n}\big(\widehat{\theta}_n - \theta^*\big) \rightsquigarrow \mathsf{N}\bigg(0,\, \frac{1}{I(\theta)}\bigg)$$

が成立する.

注意 10.10. □

定理を証明するために, 次の補題が使う.

補題 10.11. Y, $\{Y_n\}_{n=1}^\infty$ を確率変数列とし, $Y_n \rightsquigarrow Y$ とする. $\{B_n\}_{n=1}^\infty$ を事象列とし $\mathsf{Pr}(B_n) \overset{n\to\infty}{\longrightarrow} 1$ とする. このとき, 任意の確率変数列 $\{Z_n\}_{n=1}^\infty$ に対して

$$Y_n\mathbb{1}_{B_n} + Z_n\mathbb{1}_{B_n^c} \rightsquigarrow Y$$

となる.

Proof. $\forall\epsilon > 0$ に対して

$$\mathsf{Pr}\big(|Z_n\mathbb{1}_{B_n^c}| > \epsilon\big) \le \mathsf{Pr}(B_n^c) = 1 - \mathsf{Pr}(B_n) \overset{n\to\infty}{\longrightarrow} 0$$

となる. したがって

$$Z_n\mathbb{1}_{B_n^c} \,\mathrm{P}\!\!\rightarrow 0 \tag{10.5}$$

がわかる. さらに

$$\mathsf{Pr}\big(|\mathbb{1}_{B_n} - 1| > \epsilon\big) \le \mathsf{Pr}(B_n^c) = 1 - \mathsf{Pr}(B_n) \overset{n\to\infty}{\longrightarrow} 0$$

となるので

$$\mathbb{1}_{B_n} \xrightarrow{\mathrm{P}} 1 \tag{10.6}$$

がわかる. (10.5) と (10.6) 注意して, Slutsky の定理を用いると

$$Y_n \mathbb{1}_{B_n} + Z_n \mathbb{1}_{B_n^c} \rightsquigarrow Y$$

がわかる. □

定理 10.9 **の証明**: 仮定 (e) を用いて, $\epsilon > 0$ をうまく取って

$$[\theta^* - \epsilon,\, \theta^* + \epsilon] \subset \Theta \quad \text{かつ} \quad \mathsf{E}_{\theta^*}\big[\| \mathbb{1}_{[\theta^*-\epsilon,\, \theta^*+\epsilon]} \ddot{W} \|_\infty\big] < \infty$$

となるようにする. 事象 B_n を

$$B_n := \{\widehat{\theta}_n \in [\theta^* - \epsilon,\, \theta^* + \epsilon]\}$$

と定める. $\widehat{\theta}_n$ は一致性をもつので

$$\mathsf{Pr}_{\theta^*}(B_n) \xrightarrow{n\to\infty} 1 \tag{10.7}$$

となる. さらに, B_n が起こったとき, $\widehat{\theta}_n$ は $n\overline{W}_n(\,\cdot\,) = \ell_n(\,\cdot\,)$ を最大にするので

$$\left.\frac{\mathrm{d}\overline{W}_n}{\mathrm{d}\theta}(\theta)\right|_{\theta = \widehat{\theta}_n} = 0$$

となる. ここで θ^* の回りで $\dot{\overline{W}}$ を Taylor 展開すると

$$\dot{\overline{W}}_n\,(\widehat{\theta}) = \dot{\overline{W}}_n\,(\theta^*) + \ddot{\overline{W}}_n\,(\widetilde{\theta})\big(\widehat{\theta}_n - \theta^*\big) \tag{10.8}$$

を得る. ただし, $\widetilde{\theta}_n$ は θ^* と $\widehat{\theta}_n$ を結ぶ直線上の確率変数である. (10.8) の左辺の項は 0 に等しいので, B_n が起こったとき

$$\sqrt{n}\big(\widehat{\theta}_n - \theta^*\big) = \frac{\sqrt{n}\,\dot{\overline{W}}_n\,(\theta^*)}{-\,\ddot{\overline{W}}_n\,(\widetilde{\theta}_n)}$$

となる. 仮定 (d) より, $\dot{\overline{W}}_n\,(\theta^*)$ は平均が 0 で分散が $I(\theta^*)$ の i.i.d. 確率変数の平均であることに注意して, 中心極限定理を用いると

$$\sqrt{n}\,\dot{\overline{W}}_n\,(\theta^*) \rightsquigarrow Z \sim \mathsf{N}\big(0,\, I(\theta^*)\big) \tag{10.9}$$

となる．B_n が起こったとき

$$\left|\widetilde{\theta}_n - \theta^*\right| \leq \left|\widehat{\theta}_n - \theta^*\right|$$

であることと $\widehat{\theta}_n$ が一致性をもつことから

$$\mathsf{Pr}\left(\left|\widetilde{\theta}_n - \theta^*\right| \leq \epsilon\right) \geq \mathsf{Pr}\left(\left|\widehat{\theta}_n - \theta^*\right| \leq \epsilon\right) \ {}^{n\to\infty}\ 1$$

となることがわかる．よって，

$$\widetilde{\theta}_n \xrightarrow{\mathrm{P}} \theta^* \tag{10.10}$$

となる．定理 10.2 から

$$\|\mathbb{1}_{[\theta^*-\epsilon,\,\theta^*+\epsilon]}\big(\ddot{\overline{W}}_n - \mu\big)\|_\infty \xrightarrow{\mathrm{P}} 0 \tag{10.11}$$

となる．ただし，$\mu(\theta) = \mathsf{E}_{\theta^*}\big[\ddot{W}(\theta)\big]$ である．さらに，(10.10) と (10.11) に注意して，定理 10.4(2) を用いると

$$\ddot{\overline{W}}_n(\widetilde{\theta}_n) \xrightarrow{\mathrm{P}} \mu(\theta^*) = -I(\theta^*) \tag{10.12}$$

となる．B_n^c が起こったときの $\widehat{\theta}_n$ の振る舞いには分布収束は依存しない．最後に，(10.7), (10.9) および (10.12) に注意して補題 10.11 を用いると

$$\begin{aligned}
\sqrt{n}\big(\widehat{\theta}_n - \theta^*\big) &= \sqrt{n}\big(\widehat{\theta}_n - \theta^*\big)\mathbb{1}_{B_n} + \sqrt{n}\big(\widehat{\theta}_n - \theta^*\big)\mathbb{1}_{B_n^c} \\
&= \frac{\sqrt{n}\,\dot{\overline{W}}_n}{-\ddot{\overline{W}}_n(\widetilde{\theta}_n)}\mathbb{1}_{B_n} + \sqrt{n}\big(\widehat{\theta}_n - \theta^*\big)\mathbb{1}_{B_n^c} \\
&\rightsquigarrow \frac{Z}{I(\theta^*)} \sim \mathsf{N}\left(0,\, \frac{1}{I(\theta^*)}\right)
\end{aligned}$$

がわかる．　□

定理 10.12 (Cramér の定理)**.** $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{p}(\cdot\,|\,\boldsymbol{\theta}^*)$ とする．ただし，$\boldsymbol{\theta}^* \in \Theta \subset \Theta^d$ とする．さらに以下の条件 (a) ～ (e) を仮定する．
(a) Θ は $\mathbb{R}^d$ の開集合である．
(b) $A = \{x \in \mathbb{R};\, \mathsf{p}(x|\,\boldsymbol{\theta})\}$ は $\boldsymbol{\theta}$ に依存しない．さらに，ほとんどいたるところの $x \in A$ に関して，関数 $\Theta \ni \boldsymbol{\theta} \mapsto \mathsf{p}(x|\,\boldsymbol{\theta})$ は 2 回連続微分可能である．また，その 2 回偏微分演算と積分記号の交換ができる．
(c) 関数 $K(x)$ が存在して，$\mathsf{E}_{\boldsymbol{\theta}^*}[K(X)] < \infty$ であり

$$\dot{\mathbf{W}}(\boldsymbol{\theta}) = \left(\frac{\partial^2}{\partial\theta_j\partial\theta_k}\log\mathsf{p}(X|\,\boldsymbol{\theta})\right)_{j,k=1,2,\ldots,d}$$

のどの成分の絶対値も $\boldsymbol{\theta}^*$ の近傍で一様に $K(X)$ で上からおさえられる.
(d) $\mathbf{I}(\boldsymbol{\theta}^*) = -\mathsf{E}_{\boldsymbol{\theta}^*}[\dot{\mathbf{W}}(\boldsymbol{\theta}^*)]$ は正定値である.
(e) $\mathsf{p}(x|\,\boldsymbol{\theta}) = \mathsf{p}(x|\,\boldsymbol{\theta}^*) \Leftrightarrow \boldsymbol{\theta} = \boldsymbol{\theta}^*$ である.
このとき, 最尤推定量 $\widehat{\boldsymbol{\theta}}_n$ は

$$\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*\big) \rightsquigarrow \mathsf{N}_d\big(\mathbf{0}_1, \mathsf{I}(\boldsymbol{\theta}^*)^{-1}\big)$$

をみたす.

Proof. 証明は Ferguson (2017, pp.135-136) を参照のこと. □

10.4　尤度比検定統計量の漸近分布

統計的モデル $\mathcal{P} = \{\mathsf{p}(\cdot\,|\,\boldsymbol{\theta});\, \boldsymbol{\theta} \in \Theta \subset^d\}$ を考える. ただし, $\mathsf{p}(\cdot\,|\,\boldsymbol{\theta})$ は $\mathbb{R}^d$ 上の p.d.f. または p.m.f. である. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{p}(\cdot\,|\,\boldsymbol{\theta}^*)$ と仮定する. ただし, $\boldsymbol{\theta}^* \in \Theta$ である. 標本 $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ に基づき, 仮説

$$H_0 : \boldsymbol{\theta}^* \in \Theta_0 \quad \text{vs.} \quad H_1 : \boldsymbol{\theta}^* \in \Theta \setminus \Theta_0$$

を検定する. ただし, Θ_0 は Θ の真部分集合である. 汎用的な検定統計量である尤度比検定統計量

$$\lambda_n(\boldsymbol{X}) = \frac{\sup_{\boldsymbol{\theta}\in\Theta_0} \prod_{j=1}^n \mathsf{p}(X_j|\,\boldsymbol{\theta})}{\sup_{\boldsymbol{\theta}\in\Theta} \prod_{j=1}^n \mathsf{p}(X_j|\,\boldsymbol{\theta})} = \frac{\mathsf{lik}_n(\widehat{\boldsymbol{\theta}}_{n,0})}{\mathsf{lik}_n(\widehat{\boldsymbol{\theta}}_n)} \tag{10.13}$$

を考える. すなわち, $\lambda(\boldsymbol{X})$ の実現値が非常に小さいときに, H_0 は棄却されるという検定方式である. ただし, $\widehat{\boldsymbol{\theta}}_{n,0}$ は Θ_0 上の最尤推定量で, $\widehat{\boldsymbol{\theta}}_n$ は Θ 上の最尤推定量である. 帰無仮説のもとでの尤度比検定統計量の漸近分布を求めるために, $\dim\Theta_0$ を含む最小の affine 平面の次元が $d - r\,(1 \le r \le k)$ であると仮定し, $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d)^\top$ としたとき, 帰無仮説が

$$H_0 : \theta_1 = \theta_2 = \cdots = \theta_r = 0$$

と書けると仮定する.

定理 10.13 (Wilks の定理)**.** 定理 10.12 の仮定のもとで真の母数 $\boldsymbol{\theta}^*$ が帰無仮説 H_0 を満足しているとき, 次が成立する.

$$-2\log\lambda_n(\boldsymbol{X}) \rightsquigarrow \chi^2_r.$$

Proof. 10.13) から

$$-2\log\lambda_n(\boldsymbol{X}) = 2\bigl\{\ell_n(\widehat{\boldsymbol{\theta}}_n) - \ell(\widehat{\boldsymbol{\theta}}_{n,0})\bigr\}$$

と書けることに注意する．ただし, $\ell_n(\boldsymbol{\theta}) = \log \mathsf{lik}(\boldsymbol{\theta})$ である．$\ell_n(\widehat{\boldsymbol{\theta}}_{n,0})$ を $\widehat{\boldsymbol{\theta}}_n$ の周りで 2 次の項まで展開すると

$$\ell_n(\widehat{\boldsymbol{\theta}}_{n,0}) = \ell_n(\widehat{\boldsymbol{\theta}}) + \overset{\bullet}{\ell}_n(\widehat{\boldsymbol{\theta}})\bigl(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\bigr) - n\bigl(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\bigr)^\top \widetilde{\mathbf{I}}_n(\widehat{\boldsymbol{\theta}}_{n,0})\bigl(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\bigr)$$

と書ける．ただし

$$\widetilde{\mathbf{I}}_n(\widehat{\boldsymbol{\theta}}_{n,0}) = -\frac{1}{n}\int_0^1\int_0^1 v\, \overset{\bullet\bullet}{\ell}_n\left\{\widehat{\boldsymbol{\theta}}_n + uv(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n)\right\} \mathrm{d}u\mathrm{d}v$$

である．さらに

$$\widetilde{\mathbf{I}}_n(\widehat{\boldsymbol{\theta}}_{n,0}) \overset{\text{a.s.}}{\to} \frac{1}{2}\mathbf{I}(\boldsymbol{\theta}^*) \tag{10.14}$$

となること[2]がわかる．ただし, $\mathbf{I}(\cdot)$ は定理 10.12(e) で与えられた Fisher 情報量である．$\ell_n(\widehat{\boldsymbol{\theta}}_n) = 0$ が成り立つので, 十分大きな n に対して

$$\begin{aligned}
-2\log\lambda_n(\boldsymbol{X}) &= 2n\bigl(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\bigr)^\top \mathbf{I}_n\bigl(\widehat{\boldsymbol{\theta}}_{n,0}\bigr)\bigl(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\bigr) \\
&= n\bigl(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\bigr)^\top \mathbf{I}\bigl(\boldsymbol{\theta}^*\bigr)\bigl(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\bigr) + o_P(1)
\end{aligned}$$

が成り立つ．ただし, 確率変数 A と B に対して, $A = B + o_P(1)$ と書いたとき, $A - B \overset{\mathrm{P}}{\to} 0$ を意味する．

H_0 が単純仮説ならば, $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}^*$ のとき, $\widehat{\boldsymbol{\theta}}_{n,0} = \boldsymbol{\theta}^*$ である．このとき

$$\sqrt{n}\bigl(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*\bigr) \rightsquigarrow \mathsf{N}_d\bigl(\mathbf{0}_d, \mathbf{I}(\boldsymbol{\theta}^*)^{-1}\bigr)$$

なので, 証明はおわり．

複合帰無仮説のときに, $\sqrt{n}\bigl(\widehat{\boldsymbol{\theta}}_n - \widehat{\boldsymbol{\theta}}_{n,0}\bigr)$ の漸近分布を求めるために, $\overset{\bullet}{\ell}_n(\widehat{\boldsymbol{\theta}}_{n,0})$ を $\widehat{\boldsymbol{\theta}}_n$ の周りで展開する．

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\overset{\bullet}{\ell}_n(\widehat{\boldsymbol{\theta}}_{n,0}) \\
&\qquad = \frac{1}{\sqrt{n}}\overset{\bullet}{\ell}_n(\widehat{\boldsymbol{\theta}}_{n,0}) + \frac{1}{n}\left(\int_0^1 \overset{\bullet\bullet}{\ell}_n\left\{\widehat{\boldsymbol{\theta}}_n + v(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n)\right\}\mathrm{d}v\right)\sqrt{n}\bigl(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\bigr) \\
&\qquad = -\mathbf{I}(\boldsymbol{\theta}^*)\sqrt{n}\bigl(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\bigr) + o_P(1)
\end{aligned}$$

となる．ただし, $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d)^\top$ と書いたときに

$$\overset{\bullet}{\ell}_n(\boldsymbol{\theta}) = \left(\frac{\partial\ell_n}{\partial\theta_1}, \frac{\partial\ell_n}{\partial\theta_2}, \ldots, \frac{\partial\ell_n}{\partial\theta_d}\right)^\top; \qquad \overset{\bullet\bullet}{\ell}_n(\boldsymbol{\theta}) = \left(\frac{\partial^2\ell_n}{\partial\theta_j\partial\theta_k}\ell_n(\boldsymbol{\theta})\right)_{j,k=1,2,\ldots,d}$$

[2]証明はこの証明の後の補遺を参照のこと．

と定めた．したがって

$$\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_{n,0}-\widehat{\boldsymbol{\theta}}_n\big) \rightsquigarrow -\mathrm{I}(\boldsymbol{\theta}^*)^{-1}\,\dot{\ell}_n\,(\widehat{\boldsymbol{\theta}}_{n,0})+o_P(1)$$

を得る．このことから

$$-2\log\lambda_n=\frac{1}{\sqrt{n}}\,\dot{\ell}_n\,(\widehat{\boldsymbol{\theta}}_{n,0})^\top\mathbf{I}(\boldsymbol{\theta}^*)^{-1}\,\dot{\ell}_n\,(\widehat{\boldsymbol{\theta}}_{n,0})+o_P(1)$$

と表現できる．$\dot{\ell}_n\,(\widehat{\boldsymbol{\theta}}_{n,0})$ の漸近分布を求めるために $\boldsymbol{\theta}^*$ の周りでもこれを展開する．

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\,\dot{\ell}_n\,(\widehat{\boldsymbol{\theta}}_{n,0})\\
&\quad=\frac{1}{\sqrt{n}}\,\dot{\ell}_n\,(\boldsymbol{\theta}^*)+\frac{1}{n}\left(\int_0^1\ddot{\ell}_n\,\big(\boldsymbol{\theta}^*+v(\widehat{\boldsymbol{\theta}}_{n,0}-\boldsymbol{\theta}^*)\big)\,\mathrm{d}v\right)\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_{n,0}-\boldsymbol{\theta}^*\big)
\end{aligned}\tag{10.15}$$

を得る．ここで，$d\times d$ 行列 $\mathbf{I}(\boldsymbol{\theta}^*)$ を 4 つのブロック行列に分解する．

$$\mathbf{I}(\boldsymbol{\theta}^*)=\begin{pmatrix}\mathbf{G}_1 & \mathbf{G}_2\\ \mathbf{G}_2^\top & \mathbf{G}_3\end{pmatrix}$$

と分解する．ただし，$\mathbf{G}_1$ は $r\times r$ 行列である．さらに

$$\mathbf{H}:=\begin{pmatrix}\mathbf{0}_{r\times r} & \mathbf{0}_{r\times(d-r)}\\ \mathbf{0}_{r\times(d-r)}^\top & \mathbf{G}_3^{-1}\end{pmatrix}$$

とする．$\dot{\ell}_n\,(\widehat{\boldsymbol{\theta}}_{n,0})$ の後半の $d-r$ 個の要素は 0 でなので

$$\mathbf{H}\,\dot{\ell}_n\,(\widehat{\boldsymbol{\theta}}_{n,0})=\mathbf{0}_d$$

となるので，(10.15) から

$$\begin{aligned}
\mathbf{H}\frac{1}{\sqrt{n}}\,\dot{\ell}_n\,(\boldsymbol{\theta}^*)\ &\sim\ \mathbf{H}\mathbf{I}(\boldsymbol{\theta}^*)\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_{n,0}-\boldsymbol{\theta}^*\big)\\
&=\begin{pmatrix}\mathbf{0}_{r\times r} & \mathbf{0}_{r\times(d-r)}\\ \mathbf{0}_{r\times(d-r)}^\top & \mathbf{G}_3^{-1}\end{pmatrix}\begin{pmatrix}\mathbf{G}_1 & \mathbf{G}_2\\ \mathbf{G}_2^\top & \mathbf{G}_3\end{pmatrix}\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_{n,0}-\boldsymbol{\theta}^*\big)\\
&=\begin{pmatrix}\mathbf{0}_{r\times r} & \mathbf{0}_{r\times(d-r)}\\ \mathbf{G}_3^{-1}\mathbf{G}_2^\top & \mathbf{I}_{d-r}\end{pmatrix}\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_{n,0}-\boldsymbol{\theta}^*\big)\\
&=\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_{n,0}-\boldsymbol{\theta}^*\big)
\end{aligned}\tag{10.16}$$

となる．最後の等号は $\widehat{\boldsymbol{\theta}}_{n,0}$ と $\boldsymbol{\theta}^*$ の最初の r 個の成分は 0 であることからわかる．(10.16) を (10.15) に代入すると

$$\frac{1}{\sqrt{n}}\,\dot{\ell}_n\,(\widehat{\boldsymbol{\theta}}_{n,0})=\big(\mathbf{I}_d-\mathbf{I}(\boldsymbol{\theta}^*)\mathbf{H}\big)\frac{1}{\sqrt{n}}\,\dot{\ell}_n\,(\boldsymbol{\theta}^*)+o_P(1)$$

を得る．中心極限定理から

$$\frac{1}{\sqrt{n}} \overset{\bullet}{\ell}_n(\boldsymbol{\theta}^*) \rightsquigarrow \mathsf{N}_d\big(\mathbf{0}_d, \mathbf{I}(\boldsymbol{\theta}^*)\big)$$

が成り立つので

$$\frac{1}{\sqrt{n}} \overset{\bullet}{\ell}_n(\widehat{\boldsymbol{\theta}}_{n,0}) \rightsquigarrow (\mathbf{I}_d - \mathbf{I}(\boldsymbol{\theta}^*)\mathbf{H})\boldsymbol{Y}, \qquad \mathbf{Y} \sim \mathsf{N}_d\big(\mathbf{0}_d, \mathbf{I}(\boldsymbol{\theta}^*)\big)$$

である．これより

$$\begin{aligned} -2\log\lambda_n &\rightsquigarrow \boldsymbol{Y}^\top\big(\mathbf{I}_d - \mathbf{I}(\boldsymbol{\theta}^*)\mathbf{H}\big)^\top \mathbf{I}(\boldsymbol{\theta}^*)^{-1}\big(\mathbf{I}_d - \mathbf{I}(\boldsymbol{\theta}^*)\mathbf{H}\big)\boldsymbol{Y} \\ &= \boldsymbol{Y}^\top\big(\mathbf{I}_d^{-1} - \mathbf{H}\big)^\top \mathbf{I}(\boldsymbol{\theta}^*)^{-1}\big(\mathbf{I}_d - \mathbf{I}(\boldsymbol{\theta}^*)\mathbf{H}\big)\boldsymbol{Y} \\ &= \boldsymbol{Z}^\top \mathbf{I}(\boldsymbol{\theta}^*)^{1/2}\big(\mathbf{I}_d - \mathbf{I}(\boldsymbol{\theta}^*)\mathbf{H}\big)\mathbf{I}(\boldsymbol{\theta}^*)^{1/2} \end{aligned}$$

を得る．ただし, $\boldsymbol{Z} \sim \mathsf{N}_d(\mathbf{0}_d, \mathbf{I}_d)$ である．ここで

$$\mathbf{P} := \mathbf{I}(\boldsymbol{\theta}^*)^{1/2}\big\{\mathbf{I}_d - \mathbf{I}(\boldsymbol{\theta}^*)\mathbf{H}\big\}\mathbf{I}(\boldsymbol{\theta}^*)^{1/2}$$

は

$$\mathbf{P}^2 = \mathbf{P}; \quad \mathbf{P}^\top = \mathbf{P}; \, \mathrm{rank}(\mathbf{P}) = \mathrm{tr}\,(\mathbf{P}) = \mathrm{tr}\,\big\{\mathbf{I} - \mathbf{I}(\boldsymbol{\theta}^*)\mathbf{H}\big\} = r$$

をみたす．すなわち, $\mathbf{P}$ はランク r の正射影である．したがって

$$-2\log\lambda_n(\boldsymbol{X}) \rightsquigarrow \boldsymbol{Z}^\top\mathbf{P}\boldsymbol{Z} \sim \chi^2_r$$

を得る．□

補遺: (10.14) の証明: Ferguson (2017, p.136-137) を参照して, 加筆すること．□

10.5 EM アルゴリズム: 不完全データに基づく最尤推定値の計算

EM アルゴリズムは不完全データから最尤推定値を求めるための再帰的繰り返し計算法である．この節では, 観測されない完全データ X は指数型分布族に含まれる p.d.f. から生成されたものと仮定する．X は未観測であるが, その代わりに X の関数である Y を観測すると仮定する．すなわち, 単射ではない関数 g があって, $Y = g(X)$ と表現できると仮定する．以下では, X は測度 μ に関する p.d.f.

$$\mathsf{p}(x) = h(x)\exp\big\{\eta T(x) - \kappa^\vee(\eta)\big\}$$

をもつとする. ここで, $\mathbb{X}$ を X の標本空間とし, $g : \mathbb{X} \to [0, \infty)$, $T : \mathbb{X} \to \mathbb{R}$ とし, $\eta \in E \subset \mathbb{R}$ とし, $\kappa^\vee : E \to \mathbb{R}$ とする. もちろん, アルゴリズムは高次元母数モデルでも X が離散型の場合でもうまく機能することを注意しておく.

EM アルゴリズムは, 意味データが不完全なとき, 有効である. たとえば, $X_1, X_2, \ldots, X_n$ は i.i.d. ランダム標本とし, $Y_j\,(j = 1, 2, \ldots, n)$ は X_j の小数第 1 位を四捨五入したもの等である. また, EM アルゴリズムは, 欠損データに対してもうまく機能する. たとえば, ある調査で設問が 2 つあり, おのおのの設問では, いくつかの選択肢から回答を選ぶものとする. X は 2 つの設問に対する回答としたとき, Y は 1 つ目の質問をスキップして, 2 つ目だけ回答をしたデータ等である.

いま, $\mathbb{X}$ と $\mathbb{Y}$ を X と Y それぞれの標本空間とする. $\forall y \in \mathbb{Y}$ に対して, 切断面 $\mathbb{X}(y)$ を

$$\mathbb{X}(y) = \bigl\{x \in \mathbb{X} : g(x) = y\bigr\}$$

で定める. このとき

$$Y = y \Leftrightarrow X \in \mathbb{X}(y)$$

となる.

命題 10.14. *μ を $\mathbb{X}$ 上の測度とし, ν を $\mathbb{Y}$ 上の計数測度とする. $\mu \times \nu$ に関する (X, Y) の同時 p.d.f. は*

$$\mathsf{p}^{(X,Y)}(x, y) = \mathbb{1}_{\mathbb{X}(y)}(x) h(x) \exp\bigl\{\eta T(x) - \kappa^\vee\bigr\}$$

で与えられる.

Proof. $f : \mathbb{X} \times \mathbb{Y} \to [0, \infty)$ を任意の可積分関数とする. このとき

$$f(X, Y) = \sum_{y \in \mathbb{Y}} f(X, y) \mathbb{1}\{Y = y\}$$

と書ける. 期待値作用素の線型性から

$$\begin{aligned}
\mathsf{E}\bigl[f(X, Y)\bigr] &= \sum_{y \in \mathbb{Y}} \mathsf{E}\bigl[f(X, y)\mathbb{1}\{g(X) = y\}\bigr] \\
&= \sum_{y \in \mathbb{Y}} \int_{\mathbb{X}} f(x, y) h(x) \exp\bigl\{\eta T(x) - \kappa^\vee\bigr\}\,\mathrm{d}\mu(x)
\end{aligned}$$

となる. $\mathbb{1}\{g(x) = y\} = \mathbb{1}_{\mathbb{X}(y)}(x)$ に注意する. このことから命題の主張は証明された. □

アルゴリズムを定義するために, 完全データに基づく η の最尤推定量は $\psi(T)$ であることに注意する. ただし, ψ は $\overset{\bullet}{\kappa}{}^{\vee} = \dfrac{\mathrm{d}\kappa^{\vee}}{\mathrm{d}\eta}$ の逆関数である. さらに

$$e(y,\, \eta) = \mathsf{E}_{\eta}\big[T(X) |\, Y = y\big]$$

とおく. これは, $Y = y$ を与えたときの X の条件付き p.d.f. に関する積分で計算できる. $(X,\, Y)$ の p.d.f. を Y の周辺 p.m.f. で割る. すなわち

$$\frac{\mathbb{1}_{\mathbb{X}(y)}(x)h(x)\exp\big\{\eta T(x) - \kappa^{\vee}(\eta)\big\}}{\mathsf{p}^{Y}(y|\, \eta)}$$

である. ただし

$$\mathsf{p}^{Y}(y|\, \eta) = \mathsf{Pr}_{\eta}\big(Y = y\big) = \mathsf{Pr}_{\eta}\big(X \in \mathbb{X}(y)\big) \tag{10.17}$$

$$= \int_{\mathbb{X}(y)} h(x)\exp\big\{\eta T(x) - \kappa^{\vee}(\eta)\big\}\, \mathrm{d}\mu(x) \tag{10.18}$$

である.

アルゴリズムは, 最尤推定値 $\widehat{\eta}$ に対する初期推定値 $\widehat{\eta}_0$ から始まる. この初期値とデータ Y を用いて, $T_1(X)$ の値を

$$T_1 = e\big(Y,\, \widehat{\eta}_0\big)$$

と定める. これを **E-Step** とよぶ. 更新された値を $\widehat{\eta}_1 = \psi(T)$ と定める. これを **M-Step** という. E-Step と M-Step を交互に繰り返して行き, 更新された値が収束するまでこの 2 つの Step を交互に繰り返す.

指数型分布族が正準型ではなく

$$h(x)\exp\big\{\eta(\theta)T(x) - \kappa(\theta)\big\}$$

で表現されるとする. すると E-Step は前の場合と同じで, M-Step は

$$\Theta \ni \theta \mapsto \eta(\eta)T(x) - \kappa(\theta)$$

を最大にする値 $\theta_k\, (k = 1,\, 2, \ldots)$ である. ただし, $\Theta = \overset{\bullet}{\kappa}{}^{\vee}(E)$ である.

EM アルゴリズムの収束を様子を調べてみる. そのために, EM アルゴリズムが $\widetilde{\eta}$ に収束したとき, $\widetilde{\eta}$ は関係式

$$\widehat{\eta} = \psi\big(e(Y,\, \widetilde{\eta})\big) \tag{10.19}$$

をみたす. $\psi = \overset{\bullet}{\kappa}{}^{\vee}$ に注意すると, (10.19) は

$$\overset{\bullet}{\kappa}{}^{\vee}(\widetilde{\eta}) = e(Y,\, \widetilde{\eta}) \tag{10.20}$$

と同値となる．(10.20) は以下の議論からわかる．(10.18) より

$$
\begin{aligned}
\frac{\partial}{\partial\eta}\log \mathsf{p}^{Y}(Y|\,\eta) &= \frac{\dfrac{\partial}{\partial\eta}\displaystyle\int_{\mathbb{X}(Y)} h(x)\exp\bigl\{\eta T(x)-\kappa^{\vee}(\eta)\bigr\}\,\mathrm{d}\mu(x)}{\mathsf{p}^{Y}(Y|\,\eta)} \\
&= \frac{\displaystyle\int_{\mathbb{X}(Y)} \frac{\partial}{\partial\eta} h(x)\exp\bigl\{\eta T(x)-\kappa^{\vee}(\eta)\bigr\}\,\mathrm{d}\mu(x)}{\mathsf{p}^{Y}(Y|\,\eta)} \\
&= \frac{\displaystyle\int_{\mathbb{X}(Y)} \bigl\{T(x)-\overset{\bullet}{\kappa}{}^{\vee}(\eta)\bigr\} h(x)\exp\bigl\{\eta T(x)-\kappa^{\vee}(\eta)\bigr\}\,\mathrm{d}\mu(x)}{\mathsf{p}^{Y}(Y|\,\eta)} \\
&= \int_{\mathbb{X}(Y)} T(x)\,\underbrace{\frac{h(x)\exp\bigl\{\eta T(x)-\kappa^{\vee}(\eta)\bigr\}}{\mathsf{p}^{Y}(Y|\,\eta)}\,\mathrm{d}\mu(x)}_{=\mathsf{p}^{X|\,Y}(X|\,Y=y)} \\
&\qquad - \overset{\bullet}{\kappa}{}^{\vee}(\eta)\frac{\displaystyle\int_{\mathbb{X}(Y)} h(x)\exp\bigl\{\eta T(x)-\kappa^{\vee}(\eta)\bigr\}\,\mathrm{d}\mu(x)}{\mathsf{p}^{Y}(Y|\,\eta)} \\
&= \mathsf{E}_{\eta}\bigl[T(X)|\,Y\bigr]-\overset{\bullet}{\kappa}{}^{\vee}(\eta) \\
&= e(Y,\,\eta)-\overset{\bullet}{\kappa}{}^{\vee}(\eta)
\end{aligned}
$$

となることに注意する．対数尤度 $\eta \mapsto \log \mathsf{p}^{Y}(Y|\,\eta)$ は $\eta=\widetilde{\eta}$ で最大化されるので, その傾きは 0 となる．よって

$$
\left.\frac{\partial}{\partial\eta}\log \mathsf{p}^{Y}(Y|\,\eta)\right|_{\eta=\widetilde{\eta}} = 0 \;\Leftrightarrow\; e(Y,\,\widetilde{\eta}) = \overset{\bullet}{\kappa}{}^{\vee}(\widetilde{\eta})
$$

となることがわかる．

例 10.15. $X_1,\,X_2,\,\ldots,\,X_n \overset{\text{i.i.d.}}{\sim} \mathsf{p}(x|\,\eta)$ とする．ただし

$$
\mathsf{p}(x|\,\eta) = \begin{cases} \eta e^{-\eta x} & (x>0) \\ 0 & (x\le 0) \end{cases}
$$

である．いま

$$
Y_j = \lfloor X_j \rfloor \qquad (j=1,\,2,\,\ldots,\,n)
$$

とする．すなわち, Y_j は X_j を越えない整数である．$\boldsymbol{X}=(X_1,\,X_2,\,\ldots,\,X_n)$ の同時 p.d.f. は正準母数 η と十分統計量 $T=-(X_1+X_2+\cdots+X_n)$ の指数型分布族を作る．$\boldsymbol{X}$ に基づく η の最尤推定量は $\psi(T)=-\dfrac{n}{T}$ である．ここで ψ はこの指数型分布族のキュムラント関数 $\kappa^{\vee}(\eta)=-n\log\eta$ の導

関数の逆関数であることに注意する．命題 10.14 から，$j = 1, 2, \ldots, n$ に対して

$$\begin{aligned}
\mathsf{E}_\eta\big[X_j\big| Y_1, Y_2, \ldots, Y_n\big] &= \mathsf{E}_\eta\big[X_j\big| y_j \leq X_j < y_j + 1\big] \\
&= \frac{\displaystyle\int_{y_j}^{y_j+1} x\eta e^{-\eta x}\,\mathrm{d}x}{\displaystyle\int_{y_j}^{y_j+1} \eta e^{-\eta x}\,\mathrm{d}x} \\
&= \frac{\Big[-xe^{-\eta x}\Big]_{y_j}^{y_j+1} + \displaystyle\int_{y_j}^{y_j+1} e^{-\eta x}\,\mathrm{d}x}{\Big[-e^{-\eta x}\Big]_{y_j}^{y_j+1}} \\
&= \frac{e^{-\eta y_j}\big(y_j - (1+y_j)e^{-\eta}\big) - \dfrac{e^{-\eta y_j}}{\eta}\big(e^{-\eta} - 1\big)}{e^{-\eta y_j}\big(1 - e^{-\eta}\big)} \\
&= \frac{y_j(1 - e^{-\eta}) - e^{-\eta} - \dfrac{e^{-\eta} - 1}{\eta}}{1 - e^{-\eta}} \\
&= y_j + \frac{1 - e^{-\eta} - \eta e^{-\eta}}{\eta\big(1 - e^{-\eta}\big)} \\
&= y_j + \frac{e^{\eta} - 1 - \eta}{\eta\big(e^{\eta} - 1\big)}
\end{aligned}$$

となる．さらに，独立性から

$$\mathsf{E}_\eta\big[X_j\big| Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n\big] = \mathsf{E}_\eta\big[X_j\big| Y_j = y_j\big]$$

となるので，$\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ に対して

$$\begin{aligned}
e(\boldsymbol{y}, \eta) &= \mathsf{E}_\eta\big[T\big| Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n\big] \\
&= -\sum_{j=1}^{n} \mathsf{E}_\eta\big[X_j\big| Y_j = y_j\big] \\
&= -\sum_{j=1}^{n} \left\{ y_j + \frac{e^{\eta} - 1 - \eta}{\eta\big(e^{\eta} - 1\big)} \right\} \\
&= -n\left\{ \overline{y} + \frac{e^{\eta} - 1 - \eta}{\eta\big(e^{\eta} - 1\big)} \right\}
\end{aligned}$$

となる．よって，$k-\mathrm{Step}\,(k = 1, 2, \ldots)$ を $\widehat{\eta}_k$ と書いたとき，EM アルゴリズムは

$$\widehat{\eta}_{k+1} = -\frac{n}{T_{k+1}} \quad \text{かつ} \quad T_{k+1} = -n\left\{ \overline{Y} + \frac{e^{\widehat{\eta}_k} - 1 - \widehat{\eta}_k}{\eta_k\big(e^{\widehat{\eta}_k} - 1\big)} \right\} \tag{10.21}$$

で与えられる．ただし, $\overline{Y} = \dfrac{1}{n}(Y_1 + Y_2 + \cdots + Y_n)$ である．□

注意 10.16. 例 10.15 では, Y_j の p.m.f. は陽に計算できる．実際, $y = 0, 1, \ldots$ に対して

$$\begin{aligned}
\mathsf{Pr}_\eta\bigl(Y_j = y\bigr) &= \mathsf{Pr}_\eta\bigl(y \le X_j < y+1\bigr) \\
&= \int_y^{y+1} \eta e^{-\eta x}\,\mathrm{d}x = \Bigl[-e^{-\eta x}\Bigr]_y^{y+1} = -e^{-\eta(y+1)} + e^{-\eta y} \\
&= \bigl(1 - e^{-\eta}\bigr)\bigl(e^{-\eta}\bigr)^y
\end{aligned}$$

となる．したがって, $Y_1, Y_2, \ldots, Y_n$ は母数 $\pi = 1 - e^{-\eta}$ の幾何分布からのランダム標本となる．よって, π の最尤推定量は

$$\widehat{\pi} = \frac{1}{1 + \overline{Y}}$$

で与えられる．$\eta = -\log(1-\pi)$ なので, η の最尤推定量は

$$\widehat{\eta} = -\log\bigl(1 - \widehat{\pi}\bigr) = \log\left(1 + \frac{1}{\overline{Y}}\right)$$

である．

いま, EM アルゴリズムの繰り返し $\{\widehat{\eta}_k\}_{k=0}^{\infty}$ の $\widehat{\eta}$ への収束を調べてみる．そのために, $\widehat{\eta}_k = \widehat{\eta} + \epsilon$ と書く．すると $\widehat{\eta}$ は繰り返し計算の固定点なので

$$-\frac{n}{\widehat{\eta}} = -n\left\{\overline{y} + \frac{1}{\widehat{\eta}} - \frac{1}{e^{\widehat{\eta}} - 1}\right\}$$

をみたしていること[3]に注意する．ここで, 次の Taylor 展開を考える．

$$\begin{aligned}
\frac{1}{\widehat{\eta} + 1} &= \frac{1}{\widehat{\eta}} - \frac{\epsilon}{\widehat{\eta}^2} + O(\epsilon^2), \\
\frac{1}{e^{\widehat{\eta}+\epsilon} - 1} &= \frac{1}{e^{\widehat{\eta}} - 1} - \frac{e^{\widehat{\eta}}\epsilon}{\bigl(e^{\widehat{\eta}} - 1\bigr)^2} + O(\epsilon^2).
\end{aligned}$$

[3] $\lim_{k\to\infty} T_k = T_\infty$ とおくと, $\widehat{\eta} = -\frac{n}{T_\infty}$ となる．したがって, (10.21) から

$$-\frac{n}{\widehat{\eta}} = -n\left\{\overline{Y} + \frac{e^{\widehat{\eta}} - 1 - \widehat{\eta}}{\widehat{\eta}\bigl(e^{\widehat{\eta}} - 1\bigr)}\right\}$$

がわかる．

これらより

$$
\begin{aligned}
T_{k+1} &= -n\left\{\overline{Y} + \frac{e^{\widehat{\eta}_k} - 1 - \widehat{\eta}_k}{\eta_k\left(e^{\widehat{\eta}_k} - 1\right)}\right\} \\
&= -n\left\{\overline{Y} + \frac{1}{\widehat{\eta}_k} - \frac{1}{e^{\widehat{\eta}_k} - 1}\right\} \\
&= -n\left\{\overline{Y} + \frac{1}{\widehat{\eta} + \epsilon} - \frac{1}{e^{\widehat{\eta}_k + \epsilon} - 1}\right\} \\
&= -n\left\{\underbrace{\overline{Y} + \frac{1}{\widehat{\eta}} - \frac{1}{e^{\widehat{\eta}} - 1}}_{=-\frac{n}{\widehat{\eta}}} - \frac{\epsilon}{\widehat{\eta}^2} + \frac{e^{\widehat{\eta}}\epsilon}{\left(e^{\widehat{\eta}} - 1\right)^2} + O(\epsilon^2)\right\} \\
&= -\frac{n}{\widehat{\eta}}\left\{1 - \left(\frac{\epsilon}{\widehat{\eta}} - \frac{\epsilon\widehat{\eta}e^{\widehat{\eta}}}{\left(e^{\widehat{\eta}} - 1\right)^2}\right) + O(\epsilon^2)\right\}
\end{aligned}
$$

を得る．この式から

$$
\begin{aligned}
\widehat{\eta}_{k+1} &= -\frac{n}{T_{k+1}} \\
&= \frac{\widehat{\eta}}{1 - \epsilon\left(\frac{1}{\widehat{\eta}} - \frac{\widehat{\eta}e^{\widehat{\eta}}}{\left(e^{\widehat{\eta}} - 1\right)^2}\right) + O(\epsilon^2)} \\
&= \widehat{\eta}\left\{1 + \epsilon\left(\frac{1}{\widehat{\eta}} - \frac{\widehat{\eta}e^{\widehat{\eta}}}{\left(e^{\widehat{\eta}} - 1\right)^2}\right) + O(\epsilon^2)\right\} \\
&= \widehat{\eta} + \epsilon\left\{1 - \frac{\widehat{\eta}e^{\widehat{\eta}}}{\left(e^{\widehat{\eta}} - 1\right)^2}\right\} + O(\epsilon^2)
\end{aligned}
$$

を得る．特に, $\widehat{\eta}_k = \widehat{\eta}$ のとき, $\epsilon = 0$ なので, $\widehat{\eta}_{k+1} = \eta$ となる．したがって, $\widehat{\eta}$ は繰り返しの固定点であることがわかる．□

注意 10.17. 一般的に EM アルゴリズムは安定しており信頼できる．このアルゴリズムの良い性質は, 繰り返しをするごとに尤度関数を増加させることである．すなわち, EM アルゴリズムの $k-$Step の推定値を $\widehat{\eta}_k\,(k = 0, 1, \ldots)$ としたとき

$$
\prod_{j=1}^{n} \mathsf{p}^Y(y_j|\,\widehat{\eta}_{k+1}) \geq \prod_{j=1}^{n} \mathsf{p}^Y(y_j|\,\widehat{\eta}_k)
$$

となることである．これは, EM アルゴリズムが MM アルゴリズムの特別な場合であることからわかる．この点については, Lange (2014) を参照のこと．しかし, 収束は一般には保証されない．たとえば, 尤度関数のモードが複数ある場合には, アルゴリズムは局所最大に収束することが

ある. アルゴリズムの収束のための十分条件は Wu (1983) を参照のこと. EM アルゴリズムは安定しているが, 収束は遅いことが知られている. □

10.6　章末注釈と参考文献

10.7　演習問題

演習問題 10.1.

演習問題 10.2.

第III部

宴の始末編：付録

第A章　集合と位相の復習

A.1　集合の復習

To be written

A.2　ベクトル空間と線型写像の復習

定義 A.1. (1) $n \in \mathbb{N}$ とする．空でない集合 $\mathbb{X} \subset \mathbb{R}^n$ が**実ベクトル空間**であるとは $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{X}$ と $\forall c \in \mathbb{R}$ に対して

$$\boldsymbol{x} + \boldsymbol{y} \in \mathbb{X} \quad \text{かつ} \quad c\boldsymbol{x} \in \mathbb{X}$$

をみたすときである．
(2) $k \in \mathbb{N}$ とする．$\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k \in \mathbb{X}$ と $c_1, c_2, \ldots, c_k \in \mathbb{R}$ を用いて定義されるベクトル

$$c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 + \cdots + c_k\boldsymbol{x}_k$$

を $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$ **の線型結合**という．$S \subset \mathbb{R}^n$ とし，E を S の元の線型結合のすべてから成る集合としたとき S は E **を張る**といい，$E = \mathrm{span}(S)$ と書く．
(3) ベクトル $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k \in \mathbb{R}^n$ から成る集合が**独立**であるとは，$\forall c_1, c_2, \ldots, c_k \in \mathbb{R}$ に対して

$$c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 + \cdots + c_k\boldsymbol{x}_k = \mathbf{0}_n \Rightarrow c_1 = c_2 = \cdots = c_k = 0$$

をみたすときをいう．そうでないとき $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k\}$ は**従属**であるという．ただし $\mathbf{0}_n^\top = \underbrace{(0, 0, \ldots, 0)}_{k\text{ 個}}$ である．
(4) 実ベクトル空間 $\mathbb{X}$ が r 個の独立なベクトルからなる集合を部分集合として持ち，$(r+1)$ 個の独立なベクトルの集合を部分集合として持たないとき $\mathbb{X}$ の**次元**は r であるといい，$\dim \mathbb{X} = r$ と書く．
(5) 実ベクトル空間 $\mathbb{X}$ の独立な部分集合 S が $\mathbb{X}$ を張るとき集合 S は $\mathbb{X}$ の**基底**という．

注意 A.2. 実ベクトル空間 $\mathbb{R}^n$ を考える．

$$\boldsymbol{e}_i = (\underbrace{0, \ldots, 0}_{i-1\text{ 個}}, 1, \underbrace{0, \ldots, 0}_{n-i\text{ 個}})^\top \qquad (i = 1, 2, \ldots, n)$$

とおく．$S := \{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ は $\mathbb{R}^n$ の基底となる．この集合 S を $\mathbb{R}^n$ の**標準基底**とよぶ． □

定理 A.3. $r \in \mathbb{N}$ とし, $\mathbb{X}$ を実ベクトル空間とする．ベクトル空間 $\mathbb{X}$ が $\mathbb{X}$ の r 個のベクトルで張られるとき $\dim \mathbb{X} \leq r$ となる．

Proof. Rudin (1976, p.205) を参照のこと． □

系 A.4. $\dim \mathbb{R}^n = n$.

Proof. $\mathrm{span}(\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}) = \mathbb{R}^n$ なので $\dim \mathbb{R}^n \leq n$ となる．しかし $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ は独立なので $\dim \mathbb{R}^n \geq n$ となる．よって定理は証明された． □

定理 A.5. $\mathbb{X}$ を実ベクトル空間とし, $\dim \mathbb{X} = n$ とする．以下が成立する．
(1) E を $\mathbb{X}$ の n 個の異なる元から成る集合とする．このとき

$$\mathrm{span} E = \mathbb{X} \Rightarrow E \text{ は独立.}$$

(2) $\mathbb{X}$ は基底を持ち, すべての基底は $\mathbb{X}$ の n 個の異なる元から成る．
(3) $r \in \mathbb{N}$ を $1 \leq r \leq n$ とし, $\{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_r\} \subset \mathbb{X}$ は独立とする．このとき $\mathbb{X}$ は $\{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_r\}$ を含む $\mathbb{X}$ のベクトルから成る基底を持つ．

Proof. Rudin (1976, p.206) を参照． □

定義 A.6. $\mathbb{X}$ と $\mathbb{Y}$ を実ベクトル空間とする．写像 $T : \mathbb{X} \to \mathbb{Y}$ は**線型**であるとは

$$T(\boldsymbol{x}_1 + \boldsymbol{x}_2) = A(\boldsymbol{x}_1) + A(\boldsymbol{x}_2), \quad T(c\boldsymbol{x}) = cT(\boldsymbol{x}), \quad \big(\forall \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x} \in \mathbb{X}, \forall c \in \mathbb{R}\big)$$

をみたすときをいう．

定理 A.7. $\mathbb{X}$ を有限次元ベクトル空間とする．線型写像 $T : \mathbb{X} \to \mathbb{X}$ が 1 対 1 であるための必要十分条件は

$$\mathbb{X} = \mathrm{range}(T) := \{T(\boldsymbol{x}); \boldsymbol{x} \in \mathbb{X}\}$$

が成り立つときである．

Proof. Rudin (1976, p.207) を参照． □

定義 A.8. $\mathbb{X}$, $\mathbb{Y}$, $\mathbb{W}$ を実ベクトル空間とする．
(1) $\mathcal{L}(\mathbb{X},\,\mathbb{Y})$ を $\mathbb{X}$ から $\mathbb{Y}$ への線型写像すべてから成る集合とする．特に $\mathcal{L}(\mathbb{X},\,\mathbb{X})$ を $\mathcal{L}(\mathbb{X})$ と書くことにする．$\forall T_1,\,T_2 \in \mathcal{L}(\mathbb{X},\,\mathbb{Y})$ と $\forall c_1,\,c_2 \in \mathbb{R}$ に対して

$$\big(c_1T_1 + c_2T_2\big)(\boldsymbol{x}) = c_1T_1(\boldsymbol{x}) + c_2T_2(\boldsymbol{x}) \qquad (\forall \boldsymbol{x} \in \mathbb{X})$$

と定める．すると $c_1T_1 + c_2T_2 \in \mathcal{L}(\mathbb{X},\,\mathbb{Y})$ となる．
(2) $T \in \mathcal{L}(\mathbb{X},\,\mathbb{Y})$ と $S \in \mathcal{L}(\mathbb{Y},\,\mathbb{W})$ とする．この時 T と S の積 ST を T と S の合成写像で定める．すなわち

$$\big(ST\big)(\boldsymbol{x}) := S\big(T(\boldsymbol{x})\big) \qquad (\forall \boldsymbol{x} \in \mathbb{X})$$

である．すると $ST \in \mathcal{L}(\mathbb{X},\,\mathbb{W})$ となる．$\mathbb{X} = \mathbb{Y} = \mathbb{W}$ のときでも ST と TS は一般には一致しないことに注意せよ．
(3) $m,\,n \in \mathbb{N}$ とする．$T \in \mathcal{L}(\mathbb{R}^n,\,\mathbb{R}^m)$ に対して T の**ノルム** $\|T\|$ を

$$\|T\| := \sup_{\boldsymbol{x}\in\mathbb{R}^n;\,|\boldsymbol{x}|_{2,\,n}\le 1} \big|T(\boldsymbol{x})\big|_{2,\,m}$$

で定める．ただし $|\cdot|_{2,\,n}$ は $\mathbb{R}^n$ 上の Euclid ノルムである．

定理 A.9. (1) $T \in \mathcal{L}(\mathbb{R}^n,\,\mathbb{R}^m)$ のとき $\|T\| < \infty$ で

$$T : \mathbb{R}^n \to \mathbb{R}^m$$

は一様連続である．
(2) $\forall T,\,S \in \mathcal{L}(\mathbb{R}^n,\,\mathbb{R}^m)$ と $\forall c \in \mathbb{R}$ に対して

$$\|T + S\| \le \|T\| + \|S\|, \quad \|cT\| = |c|\,\|T\|$$

である．$\forall T,\,S,\,R \in \mathcal{L}(\mathbb{R}^n,\,\mathbb{R}^m)$ に対して関数

$$d : \mathcal{L}(\mathbb{R}^n,\,\mathbb{R}^m) \times \mathcal{L}(\mathbb{R}^n,\,\mathbb{R}^m) \ni (T,\,S) \longrightarrow d(T,\,S) = \|T - S\|$$

と定めると d は距離関数となる．すなわち, 任意の $S,\,T,\,R \in \mathcal{L}\big(\mathbb{R}^n,\,\mathbb{R}^m\big)$ に対して

(2a) $d(T,\,S) \ge 0;\quad d(T,\,S) = 0 \Leftrightarrow T = S,$

(2b) $d(T,\,S) = d(S,\,T),$

(2c) $d(T,\,R) \le d(T,\,S) + d(S,\,R)$

をみたす．
(3) $T \in \mathcal{L}(\mathbb{R}^n,\,\mathbb{R}^m)$ と $S \in \mathcal{L}(\mathbb{R}^m,\,\mathbb{R}^k)$ としたとき

$$\|ST\| \le \|S\| \cdot \|T\|$$

がなりたつ．

Proof. Rudin (1976, p.208-209) を参照. □

定理 A.10. $\mathcal{K}$ を $\mathbb{R}^n$ から $\mathbb{R}^n$ への可逆な線型写像すべての成す集合とする．このとき以下が成立する．
(1) $S \in \mathcal{K}$ と $T \in \mathcal{L}\big(\mathbb{R}^n\big)$ が

$$\|S - T\| \cdot \|S^{-1}\| < 1 \tag{A.1}$$

をみたすとき

$$T \in \mathcal{K}$$

である．
(2) $\mathcal{K}$ は $\mathcal{L}\big(\mathbb{R}^n\big)$ の開部分集合で写像

$$\mathcal{K} \ni S \mapsto S^{-1} \in \mathcal{K}$$

は連続である．

Proof. (i) $\mathrm{id}_n : \mathbb{R}^n \to \mathbb{R}^n$ を恒等写像とすると, 定理 $A.9$(iii) から

$$1 = \|\mathrm{id}_n\| = \|SS^{-1}\| \leq \|S\| \cdot \|S^{-1}\|$$

となる．よって, から $\|S^{-1}\| \neq 0$ であるので

$$\frac{1}{\|S^{-1}\|} \leq \|S\| < \infty$$

となる．このことを踏まえ

$$\|S^{-1}\| =: \frac{1}{\alpha} > 0$$

とおく．さらに $\|S - T\| =: \beta \geq 0$ とおく．仮定 (A.1) から

$$\beta = \|S - T\| < \frac{1}{\|S^{-1}\|} = \alpha$$

となる．よって

$$\beta < \alpha$$

である．一方, $\forall \boldsymbol{x} \in \mathbb{R}^n$ に対して

$$\begin{aligned}\alpha|\boldsymbol{x}|_{2,n} &= \alpha\big|S^{-1}S(\boldsymbol{x})\big|_{2,n} \leq \alpha\|S^{-1}\| \cdot |S(\boldsymbol{x})|_{2,n} = |S(\boldsymbol{x})|_{2,n} \\ &\leq \big|\big(S - T\big)(\boldsymbol{x})\big|_{2,n} + |T(\boldsymbol{x})|_{2,n} \leq \beta|\boldsymbol{x}|_{2,n} + |T(\boldsymbol{x})|_{2,n}\end{aligned}$$

となる．この不等式から

$$(\alpha-\beta)|\boldsymbol{x}|_{2,n} \le \bigl|T(\boldsymbol{x})\bigr|_{2,n} \qquad \bigl(\forall \boldsymbol{x}\in\mathbb{R}^n\bigr) \tag{A.2}$$

を得る．$\alpha-\beta>0$ なので

$$\boldsymbol{x}\neq \mathbf{0}_n \;\Rightarrow\; T\boldsymbol{x}\neq \mathbf{0}_n$$

を得る．よって T は 1 対 1 となる．定理 A.7 から

$$\mathbb{R}^n = \mathrm{range}T$$

となるので $T\in\mathcal{K}$ がわかる．すべての $S\in\mathcal{K}$ に対して T は $\|S-T\|<\alpha$ をみたせば $T\in\mathcal{K}$ となる．このことから $\mathcal{K}$ は　$\mathcal{L}\bigl(\mathbb{R}^n\bigr)$ の開集合であることがわかる．

(ii) $S, T\in\mathcal{K}$ とする．(A.2) において $\boldsymbol{x}=S^{-1}\boldsymbol{y}$ と置き換えると

$$(\alpha-\beta)\bigl|S^{-1}(\boldsymbol{y})\bigr|_{2,n} \le \bigl|SS^{-1}(\boldsymbol{y})\bigr|_{2,n} = |\boldsymbol{y}|_{2,n} \qquad (\forall \boldsymbol{y}\in\mathbb{R}^n)$$

となる．よって $|\boldsymbol{y}|_{2,n}\neq 0$ のとき

$$\frac{\bigl|S^{-1}(\boldsymbol{y})\bigr|_{2,n}}{|\boldsymbol{y}|_{2,n}} \le \frac{1}{\alpha-\beta}$$

となるので

$$\|S^{-1}\| \le \frac{1}{\alpha-\beta}$$

を得る．等式

$$S^{-1}-T^{-1} = S^{-1}\bigl(T-S\bigr)T^{-1}$$

と定理 A.9(ii) を合わせると

$$\|S^{-1}-T^{-1}\| \le \|S^{-1}\|\cdot\|T-S\| \bullet \|T^{-1}\| \le \frac{\beta}{\alpha(\alpha-\beta)}$$

を得る．$\beta\to 0$ とすれば $T\to S$ となるので写像 $\mathcal{K}\ni S\mapsto S^{-1}\in\mathcal{K}$ は連続となる．　□

A.3　多変数関数の微分と逆写像定理

定義 A.11. $E\subset R^n$ を開集合とし, $\boldsymbol{x}\in E$ とし, $T\colon E\to\mathbb{R}^m$ を写像とする．線型写像 $A\colon\mathbb{R}^n\to\mathbb{R}^m$ が存在して

$$\lim_{\boldsymbol{h}\to\mathbf{0}_n}\frac{\bigl|T(\boldsymbol{x}+\boldsymbol{h})-T(\boldsymbol{x})-A(\boldsymbol{h})\bigr|_{2,m}}{|\boldsymbol{h}|_{2,n}} = 0 \tag{A.3}$$

が成り立つとき T は点 $\boldsymbol{x}$ で微分可能といい

$$\overset{\bullet}{T}(\boldsymbol{x}) := A$$

と書く. T がすべての点 $\boldsymbol{x} \in E$ で微分可能なとき T は E 上で微分可能という.

注意 A.12. (A.3) をみたす $A \in \mathcal{L}\big(\mathbb{R}^n, \mathbb{R}^m\big)$ は一意である. これを示すために $A_1, A_2 \in \mathcal{L}\big(\mathbb{R}^n, \mathbb{R}^m\big)$ があって A_1 と A_2 はともに (A.3) をみたすと仮定する. $B := A_1 - A_2$ とおくと $\forall \boldsymbol{h} \in \mathbb{R}^n$ に対して

$$\begin{aligned}\big|B(\boldsymbol{h})\big|_{2,m} &= \Big|T(\boldsymbol{x}+\boldsymbol{h}) - T(\boldsymbol{x}) - A_2(\boldsymbol{h}) - \big\{T(\boldsymbol{x}+\boldsymbol{h}) - T(\boldsymbol{x}) - A_1(\boldsymbol{h})\big\}\Big|_{2,m} \\ &\le \big|T(\boldsymbol{x}+\boldsymbol{h}) - T(\boldsymbol{x}) - A_2(\boldsymbol{h})\big|_{2,m} + \big|T(\boldsymbol{x}+\boldsymbol{h}) - T(\boldsymbol{x}) - A_1(\boldsymbol{h})\big|_{2,m}\end{aligned}$$

となるので

$$\lim_{\boldsymbol{x} \to \boldsymbol{0}_n} \frac{\big|B(\boldsymbol{x})\big|_{2,m}}{|\boldsymbol{h}|_{2,n}} = 0$$

となる. $\boldsymbol{h} \neq \boldsymbol{0}_n$ を固定すると

$$\lim_{t \to 0} \frac{\big|B(t\boldsymbol{h})\big|_{2,m}}{|t\boldsymbol{h}|_{2,n}} = 0 \tag{A.4}$$

を得る. B の線型性から (A.5) の左辺は t と独立なので $B(\boldsymbol{h}) = \boldsymbol{0}_m$ $(\forall \boldsymbol{h} \in \mathbb{R}^n)$ となる. よって $B = \boldsymbol{0}_{m\times n}$ がわかる. □

定理 A.13. $E \subset \mathbb{R}^n$ を開集合とする. 写像 $T : E \to \mathbb{R}^m$ は点 $\boldsymbol{x}_0 \in E$ で微分可能とし, $S : T(E) \to \mathbb{R}^k$ は点 $T(\boldsymbol{x}_0)$ で微分可能とする. さらに写像 $R : E \to \mathbb{R}^k$ を

$$R(\boldsymbol{x}) := S\big(T(\boldsymbol{x})\big)$$

で定義する. このとき R は点 $\boldsymbol{x}_0$ で微分可能で

$$\overset{\bullet}{R}(\boldsymbol{x}_0) = \overset{\bullet}{S}\big(T(\boldsymbol{x}_0)\big)\, \overset{\bullet}{T}(\boldsymbol{x}_0)$$

となる.

Proof. $\boldsymbol{y}_0 = T(\boldsymbol{x}_0)$, $A = \overset{\bullet}{T}(\boldsymbol{x}_0)$, $B = \overset{\bullet}{S}(\boldsymbol{y}_0)$ とおき

$$\begin{aligned} U(\boldsymbol{h}) &:= T(\boldsymbol{x}_0 + \boldsymbol{h}) - T(\boldsymbol{x}_0) - A(\boldsymbol{h}) \qquad (\boldsymbol{h} \in \mathbb{R}^n), \\ V(\boldsymbol{k}) &:= S(\boldsymbol{y}_0 + \boldsymbol{k}) - S(\boldsymbol{y}_0) - B(\boldsymbol{k}) \qquad (\boldsymbol{k} \in \mathbb{R}^m) \end{aligned}$$

と定める．ただし $\boldsymbol{x}_0+\boldsymbol{h}\in E$, $\boldsymbol{y}_0+\boldsymbol{k}\in T(E)$ なるように $\boldsymbol{h}$, $\boldsymbol{k}$ を定めた．このとき

$$\bigl|U(\boldsymbol{h})\bigr|_{2,\,n}=\epsilon(\boldsymbol{h})|\boldsymbol{h}|_{2,\,n},\quad \bigl|V(\boldsymbol{k})\bigr|_{2,\,m}=\eta(\boldsymbol{k})|\boldsymbol{k}|_{2,\,m} \tag{A.5}$$

$$\epsilon(\boldsymbol{h})\to 0\,(\boldsymbol{h}\to\boldsymbol{0}_n),\quad \eta(\boldsymbol{k})\to 0\,(\boldsymbol{k}\to\boldsymbol{0}_m) \tag{A.6}$$

と書く．与えられた $\boldsymbol{h}\in\mathbb{R}^n$ に対して

$$\boldsymbol{k}:=T(\boldsymbol{x}_0+\boldsymbol{h})-T(\boldsymbol{x}_0)$$

とすると

$$|\boldsymbol{k}|_{2,\,m}=\bigl|A(\boldsymbol{h})+U(\boldsymbol{h})\bigr|_{2,\,m}\le\bigl[\|A\|+\epsilon(\boldsymbol{h})\bigr]\times|\boldsymbol{h}|_{2,\,n} \tag{A.7}$$

と

$$\begin{aligned}
R(\boldsymbol{x}_0+\boldsymbol{h})-R(\boldsymbol{x}_0)-BA(\boldsymbol{h})&=S\bigl(T(\boldsymbol{x}_0+\boldsymbol{h})\bigr)-S\bigl(T(\boldsymbol{x}_0)\bigr)-BA(\boldsymbol{h})\\
&=S\Bigl(T(\boldsymbol{x}_0)+\bigl(T(\boldsymbol{x}_0+\boldsymbol{h})-T(\boldsymbol{x}_0)\bigr)\Bigr)\\
&\qquad-S(\boldsymbol{y}_0)-BA(\boldsymbol{h})\\
&=S(\boldsymbol{y}_0+\boldsymbol{k})-S(\boldsymbol{y}_0)-BA(\boldsymbol{h})\\
&=V(\boldsymbol{k})+B(\boldsymbol{k})-BA(\boldsymbol{h})\\
&=V(\boldsymbol{k})+B\Bigl(T(\boldsymbol{x}_0+\boldsymbol{h})-T(\boldsymbol{x}_0)-A(\boldsymbol{h})\Bigr)\\
&=V(\boldsymbol{k})+B\bigl(U(\boldsymbol{h})\bigr)
\end{aligned} \tag{A.8}$$

となる．(A.5) と (A.7) から $\boldsymbol{h}\ne\boldsymbol{0}_n$ のとき

$$\begin{aligned}
\frac{\bigl|R(\boldsymbol{x}_0+\boldsymbol{h})-R(\boldsymbol{x}_0)-BA(\boldsymbol{h})\bigr|_{2,\,k}}{|\boldsymbol{h}|_{2,\,n}}&=\frac{\bigl|BU(\boldsymbol{h})+V(\boldsymbol{k})\bigr|_{2,\,k}}{|\boldsymbol{h}|_{2,\,n}}\qquad\bigl(\because\text{(A.8)}\bigr)\\
&\le\frac{\|B\|\epsilon(\boldsymbol{h})|\boldsymbol{h}|_{2,\,n}+\eta(\boldsymbol{k})|\boldsymbol{k}|_{2,\,m}}{|\boldsymbol{h}|_{2,\,n}}\\
&\le\|B\|\epsilon(\boldsymbol{h})+\frac{\eta(\boldsymbol{k})\bigl[\|A\|+\epsilon(\boldsymbol{h})\bigr]|\boldsymbol{h}|_{2,\,n}}{|\boldsymbol{h}|_{2,\,n}}\\
&\qquad\bigl(\because\text{(A.7)}\bigr)\\
&\le\|B\|\epsilon(\boldsymbol{h})+\eta(\boldsymbol{k})\bigl[\|A\|+\epsilon(\boldsymbol{h})\bigr]
\end{aligned}$$

となる．$\boldsymbol{h}\to\boldsymbol{0}_n$ のとき $\epsilon(\boldsymbol{h})\to 0$ かつ $\boldsymbol{k}\to\boldsymbol{0}_m$ なので $\eta(\boldsymbol{k})\to 0$ となる．よって

$$\overset{\bullet}{R}\,(\boldsymbol{x}_0)=BA$$

がわかる. □

$E \subset \mathbb{R}^n$ を開集合とし, $T: E \to \mathbb{R}^m$ を写像とする. $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ と $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m\}$ を $\mathbb{R}^n$ と $\mathbb{R}^m$ の標準基底とする. 写像 T の成分は実数値関数 $T_i: E \to \mathbb{R}\,(i = 1, 2, \ldots, m)$ を用いて

$$T(\boldsymbol{x}) = \sum_{i=1}^{m} T_i(\boldsymbol{x})\boldsymbol{u}_i \qquad (\boldsymbol{x} \in E)$$

と表現できる. $\boldsymbol{x} \in E$ と $i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$ に対して

$$\big(D_jT_i\big)(\boldsymbol{x}) := \lim_{t\to 0} \frac{T_i(\boldsymbol{x} + t\boldsymbol{e}_j) - T_i(\boldsymbol{x})}{t}$$

と定める. これを

$$D_jT_i = \frac{\partial T_i}{\partial x_j} \qquad \boldsymbol{x} = (x_1, x_2, \ldots, x_n)^\top$$

と書くことにする.

定理 A.14. $E \subset \mathbb{R}^n$ を開集合とし, 写像 $T: E \to \mathbb{R}^m$ は点 $\boldsymbol{x} \in E$ で微分可能とする. このとき偏導関数 $\big(D_jT_i\big)(\boldsymbol{x})\,(i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n)$ は存在し

$$\overset{\bullet}{T}(\boldsymbol{x})(\boldsymbol{e}_j) = \sum_{i=1}^{m} \big(D_jT_i\big)(\boldsymbol{x})\boldsymbol{u}_i \qquad (j = 1, 2, \ldots, n)$$

と書ける. ただし $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ と $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m\}$ を $\mathbb{R}^n$ と $\mathbb{R}^m$ の標準基底である.

Proof. $j\,(1 \le j \le n;\ j \in \mathbb{N})$ を固定する. 写像 T は点 $\boldsymbol{x} \in E$ で微分可能なので

$$\begin{aligned} T(\boldsymbol{x} + t\boldsymbol{e}_j) - T(\boldsymbol{x}) &= \overset{\bullet}{T}(\boldsymbol{x})(t\boldsymbol{e}_j) + \boldsymbol{r}(t\boldsymbol{e}_j), \\ \lim_{t\to 0} \frac{\big|\boldsymbol{r}(t\boldsymbol{e}_j)\big|_{2,m}}{t} &= 0 \qquad (\boldsymbol{x} + t\boldsymbol{e}_j \in E) \end{aligned}$$

と書ける. $\overset{\bullet}{T}(\boldsymbol{x})$ の線型性から

$$\begin{aligned} \lim_{t\to 0} \frac{T(\boldsymbol{x} + t\boldsymbol{e}_j) - T(\boldsymbol{x})}{t} &= \lim_{t\to 0} \frac{\overset{\bullet}{T}(\boldsymbol{x})(t\boldsymbol{e}_j) + \boldsymbol{r}(t\boldsymbol{e}_j)}{t} \\ &= \lim_{t\to 0} \frac{t\,\overset{\bullet}{T}(\boldsymbol{x})(\boldsymbol{e}_j) + \boldsymbol{r}(t\boldsymbol{e}_j)}{t} \\ &= \overset{\bullet}{T}(\boldsymbol{x})(\boldsymbol{e}_j) + \lim_{t\to 0} \frac{\boldsymbol{r}(t\boldsymbol{e}_j)}{t} \\ &= \overset{\bullet}{T}(\boldsymbol{x})(\boldsymbol{e}_j) \end{aligned}$$

となる．これを T の成分で表現すると

$$\lim_{t\to 0}\sum_{i=1}^{m}\frac{T_i(\boldsymbol{x}+t\boldsymbol{e}_j)-T_i(\boldsymbol{x})}{t}\boldsymbol{u}_i = \overset{\bullet}{T}(\boldsymbol{x})\boldsymbol{e}_j$$

となる．よって $\big(D_jT_i\big)(\boldsymbol{x})$ は存在する．さらに

$$\sum_{i=1}^{m}\big(D_jT_i\big)(\boldsymbol{x})\boldsymbol{u}_i = \overset{\bullet}{T}(\boldsymbol{x})\boldsymbol{e}_j$$

から

$$\begin{aligned}\overset{\bullet}{T}(\boldsymbol{x})\boldsymbol{e}_j &= \lim_{t\to 0}\frac{T(\boldsymbol{x}+t\boldsymbol{e}_j)-T(\boldsymbol{x})}{t} = \lim_{t\to 0}\sum_{i=1}^{m}\frac{T_i(\boldsymbol{x}+t\boldsymbol{e}_j)-T_i(\boldsymbol{x})}{t}\boldsymbol{u}_i \\ &= \sum_{i=1}^{m}\big(D_jT_i\big)(\boldsymbol{x})\boldsymbol{u}_i\end{aligned}$$

を得る．□

このことより

$$\overset{\bullet}{T}(\boldsymbol{x}) = \begin{pmatrix}\big(D_1T_1\big)(\boldsymbol{x}) & \cdots & \big(D_nT_1\big)(\boldsymbol{x}) \\ \vdots & \ddots & \vdots \\ \big(D_1T_m\big)(\boldsymbol{x}) & \cdots & \big(D_nT_m\big)(\boldsymbol{x})\end{pmatrix}$$

と書ける．$\boldsymbol{h} = \sum_{j=1}^{n} h_j\boldsymbol{e}_j \in \mathbb{R}^n$ $(h_j \in \mathbb{R};\ j = 1,\,2,\,\ldots,\,n)$ に対して

$$\overset{\bullet}{T}(\boldsymbol{h}) = \sum_{i=1}^{m}\left\{\sum_{j=1}^{n}\big(D_jT_i\big)(\boldsymbol{x})h_j\right\}\boldsymbol{u}_i$$

となる．

定理 A.15. $E \subset \mathbb{R}^n$ を凸開集合とする．写像 $T : E \to \mathbb{R}^m$ は E 上で微分可能とし，ある定数 $M > 0$ が存在して

$$\big\|\overset{\bullet}{T}(\boldsymbol{x})\big\| \le M \qquad (\forall \boldsymbol{x} \in E)$$

とする．このとき

$$\big|T(\boldsymbol{a}) - T(\boldsymbol{b})\big|_{2,\,n} \le M\big|\boldsymbol{a} - \boldsymbol{a}\big|_{2,\,n} \qquad (\forall \boldsymbol{a},\,\boldsymbol{b} \in E)$$

が成立する．

Proof. $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$ を固定し

$$\boldsymbol{r}(t) := (1-t)\boldsymbol{a} + t\boldsymbol{b} \qquad (t \in \mathbb{R})$$

と定める．E は凸集合なので

$$\boldsymbol{r}(t) \in E \qquad (0 \leq t \leq 1)$$

となる．いま

$$S(t) = T\big(\boldsymbol{h}(t)\big)$$

とおく．このとき定理 A.13 を用いると

$$\overset{\bullet}{S}(t) = \overset{\bullet}{T}\big(\boldsymbol{r}(t)\big)\,\overset{\bullet}{\boldsymbol{r}}(t) = \overset{\bullet}{T}\big(\boldsymbol{r}(t)\big)(\boldsymbol{b}-\boldsymbol{a}) \in \mathbb{R}^m$$

を得る．よって

$$\big|\overset{\bullet}{S}(t)\big|_{2,m} \leq \big\|\overset{\bullet}{T}\big(\boldsymbol{r}(t)\big)\big\| \cdot |\boldsymbol{b}-\boldsymbol{a}|_{2,n} \leq M\,|\boldsymbol{b}-\boldsymbol{a}|_{2,n} \qquad (\forall t \in [0,\,1])$$

となる．しかし $S(1) = T(\boldsymbol{a})$ と $S(0) = T(\boldsymbol{b})$ なので

$$\begin{aligned}\big|T(\boldsymbol{a}) - T(\boldsymbol{b})\big|_{2,m} &= \left|\int_0^1 \overset{\bullet}{S}(t)\,\mathrm{d}t\right| \leq \int_0^1 \big|\overset{\bullet}{S}(t)\big|\,\mathrm{d}t \\ &=\leq M|\boldsymbol{b}-\boldsymbol{a}|_{2,n}\int_0^1 \mathrm{d}t = M|\boldsymbol{b}-\boldsymbol{a}|_{2,n}\end{aligned}$$

がわかる． □

定義 A.16. $E \subset \mathbb{R}^n$ を開集合とする．写像 $T:\, E \to \mathbb{R}^m$ が E 上で連続微分可能 (C^1 級) であるとは

$$\overset{\bullet}{T}:\, E \to \mathcal{L}\big(\mathbb{R}^n,\, \mathbb{R}^m\big)$$

が連続写像となることである．すなわち $\forall \boldsymbol{x} \in E$ と $\forall \epsilon > 0$ に対してある $\delta > 0$ があって，$\forall \boldsymbol{y} \in E$ に対して

$$|\boldsymbol{x}-\boldsymbol{y}|_{2,n} \Rightarrow \|\overset{\bullet}{T}(\boldsymbol{x}) - \overset{\bullet}{T}(\boldsymbol{y})\| < \epsilon$$

をみたすことである．

定理 A.17. $E \subset \mathbb{R}^n$ を開集合とし，$T:\, E \to \mathbb{R}^m$ を写像とする．このとき T が E 上で C^1 級であるための必要十分条件は，すべての $i = 1,\, 2,\, \ldots,\, m;\ j = 1,\, 2, \ldots,\, n$ に対して偏導関数 D_jT_i が存在し，E 上で連続であることである．

Proof. 写像 T は C^1 級とする. すると

$$\overset{\bullet}{T}(\boldsymbol{x})\boldsymbol{e}_j = \sum_{i=1}^{m} \big(D_j T_i\big)(\boldsymbol{x})\boldsymbol{u}_i$$

から

$$\big(\overset{\bullet}{T}(\boldsymbol{x})\boldsymbol{e}_j \,\big|\, \boldsymbol{u}_i\big)_m = \big(D_j T_i\big)(\boldsymbol{x}) \qquad (i = 1,\, 2,\, \ldots,\, m;\, j = 1,\, 2, \ldots,\, n)$$

となる. ただし $(\cdot|\cdot)_m$ は $\mathbb{R}^m$ 上の標準内積である. しがたって

$$\big(D_j T_i\big)(\boldsymbol{x}) - \big(D_j T_i\big)(\boldsymbol{y}) = \big(\{\overset{\bullet}{T}(\boldsymbol{x}) - \overset{\bullet}{T}(\boldsymbol{y})\}\boldsymbol{e}_j \,\big|\, \boldsymbol{u}_i\big)_m$$

となる. $|\boldsymbol{u}_i|_m = |\boldsymbol{e}_j|_n = 1$ から

$$\begin{aligned}\Big|\big(\{\overset{\bullet}{T}(\boldsymbol{x}) - \overset{\bullet}{T}(\boldsymbol{y})\}\boldsymbol{e}_j \,\big|\, \boldsymbol{u}_i\big)_m\Big| &\le \big|\{\overset{\bullet}{T}(\boldsymbol{x}) - \overset{\bullet}{T}(\boldsymbol{y})\}\boldsymbol{e}_j\big|_{2,\,n} \cdot |\boldsymbol{u}_i|_{2,\,m} \\ &\le \|\,\overset{\bullet}{T}(\boldsymbol{x}) - \overset{\bullet}{T}(\boldsymbol{y})\|\end{aligned}$$

となり, $D_j T_i$ は連続であることがわかる.

逆を示す. $m = 1$ の場合を考えれば十分なことに注意する. $\boldsymbol{x} \in E$ と $\epsilon > 0$ を取る. E は開集合なので $\boldsymbol{x}$ を中心とした半径 $r > 0$ の開球 B を

$$B \subset E$$

とすることができる. $D_j T_i$ の連続性から r を必要ならばさらに小さく取り

$$\big|\big(D_j T_i\big)(\boldsymbol{y}) - \big(D_j T_i\big)(\boldsymbol{x})\big| \le \frac{\epsilon}{n} \qquad (\boldsymbol{y} \in B,\, 1 \le j \le n)$$

とできる. ここで, $k = 1,\, 2,\, \ldots,\, n$ に対して

$$\boldsymbol{h} := \sum_{j=1}^{n} h_j \boldsymbol{e}_j, \quad |\boldsymbol{h}|_{2,\,n} \le r, \quad \boldsymbol{v}_0 = \boldsymbol{0}_{2,\,n}, \quad \boldsymbol{e}_j = \boldsymbol{v}_j$$

とおく. このとき

$$T(\boldsymbol{x} + \boldsymbol{h}) - T(\boldsymbol{x}) = \sum_{j=1}^{m} \big[T(\boldsymbol{x} + \boldsymbol{v}_j) - T(\boldsymbol{x} + \boldsymbol{v}_{j-1})\big]$$

となる. $|\boldsymbol{v}_k|_{2,\,n} < r\,(k = 1,\, 2,\, \ldots,\, n)$ で B は凸開集合なので

$$\boldsymbol{x} + \boldsymbol{v}_{j-1},\, \boldsymbol{x} + \boldsymbol{v}_j \in B$$

となる．$\boldsymbol{v}_j = \boldsymbol{v}_{j-1} + h_j \boldsymbol{e}_j$ $(j = 1,\, 2,\, \ldots,\, n)$ なので中間値の定理から

$$\begin{aligned} T(\boldsymbol{x}+\boldsymbol{v}_j) - T(\boldsymbol{x}+\boldsymbol{v}_{j-1}) &= T(\boldsymbol{x}+\boldsymbol{v}_{j-1}+h_j\boldsymbol{e}_j) - T(\boldsymbol{x}+\boldsymbol{v}_{j-1}) \\ &= h_j \big(D_j T\big)(\boldsymbol{x}+\boldsymbol{v}_{j-1}+\theta_j h_j \boldsymbol{e}_j) \\ &\quad (0 < \theta_j < 1;\, j = 1,\, 2,\, \ldots,\, n) \end{aligned}$$

となる．ただし $D_j T = (D_1 T_1,\, D_j T_2,\, \ldots,\, D_j T_m)^\top$ である．以上のことから

$$\begin{aligned} &\left| T(\boldsymbol{x}+\boldsymbol{h}) - T(\boldsymbol{x}) - \sum_{j=1}^n h_j \big(D_j T\big)(\boldsymbol{x}) \right|_{2,\, m} \\ &\quad = \left| \sum_{j=1}^n h_j \Big\{ \big(D_j T\big)(\boldsymbol{x}+\boldsymbol{v}_{j-1}+\theta_j h_j \boldsymbol{e}_j) - \big(D_j T\big)(\boldsymbol{x}) \Big\} \right|_{2,\, m} \\ &\quad \leq \frac{1}{n} \sum_{j=1}^n |h_j| \epsilon = |\boldsymbol{h}|_{2,\, n} \epsilon \qquad (\forall |\boldsymbol{h}|_{2,\, n} < \epsilon) \end{aligned}$$

となる．よって写像 T は点 $\boldsymbol{x}$ で微分可能で $\overset{\bullet}{T}(\boldsymbol{x})$ は線型で $\boldsymbol{h} = \sum_{j=1}^n h_j \boldsymbol{e}_j$ に対して

$$\overset{\bullet}{T}(\boldsymbol{x})(\boldsymbol{h}) = \sum_{j=1}^n h_j \big(D_j T\big)(\boldsymbol{x})$$

となる．さらに

$$\overset{\bullet}{T}(\boldsymbol{x}) = \begin{bmatrix} \big\{\big(D_1 T\big)(\boldsymbol{x})\big\}^\top \\ \big\{\big(D_2 T\big)(\boldsymbol{x})\big\}^\top \\ \vdots \\ \big\{\big(D_n T\big)(\boldsymbol{x})\big\}^\top \end{bmatrix}$$

となり，$D_1 T,\, D_2 T,\, \ldots,\, D_n T$ は連続なので $\overset{\bullet}{T}$ も連続であることがわかる． □

定義 A.18. 写像 $T : \mathbb{R}^n \to \mathbb{R}^n$ は次の条件をみたすとする．ある $c < 1$ があって

$$\big|T(\boldsymbol{x}) - T(\boldsymbol{y})\big|_{2,\, n} \leq c |\boldsymbol{x} - \boldsymbol{y}|_{2,\, n} \qquad (\forall \boldsymbol{x},\, \boldsymbol{y} \in \mathbb{R}^n)$$

となる．このとき T を $\mathbb{R}^n$ **から** $\mathbb{R}^n$ **への縮小写像**という．

定理 A.19. $C \subset \mathbb{R}^n$ を閉集合とし, $T : C \to C$ を縮小写像とする．このときある $\boldsymbol{x}_0 \in C$ が唯一存在して

$$T(\boldsymbol{x}_0) = \boldsymbol{x}_0$$

をみたす．

Proof. $\boldsymbol{x}_0 \in C$ を任意に取り点列 $\{\boldsymbol{x}_k\}_{k=1}^{\infty} \subset C$ を

$$\boldsymbol{k}_{k+1} = T(\boldsymbol{x}_k) \qquad (k = 1,\, 2, \ldots)$$

と定める．$k \geq 1$ に対して

$$|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k|_{2,\,n} = |T(\boldsymbol{x}_k) - T(\boldsymbol{x}_{k-1})|_{2,\,n} \leq c|\boldsymbol{x}_k - \boldsymbol{x}_{k-1}|_{2,\,n}$$

となる．したがって

$$|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k|_{2,\,n} \leq c\,|\boldsymbol{x}_1 - \boldsymbol{x}_0|_{2,\,n} \qquad (k = 1,\, 2,\, \ldots)$$

を得る．$k < \ell$ に対して

$$\begin{aligned}|\boldsymbol{x}_k - \boldsymbol{x}_\ell|_{2,\,n} &\leq \sum_{i=k+1}^{\ell} |\boldsymbol{x}_i - \boldsymbol{x}_{i-1}|_{2,\,n} \leq \left\{c^k + c^{k-1} + \cdots + c^{\ell-1}\right\}|\boldsymbol{x}_1 - \boldsymbol{x}_0|_{2,\,n} \\ &\leq \left[\frac{|\boldsymbol{x}_1 - \boldsymbol{x}_0|_n}{1-c}\right]c^k\end{aligned}$$

となる．よって　$\{\boldsymbol{x}_k\}_{k=1}^{\infty}$ は Cauchy 列であることがわかる．よって C は閉集合であるのである $\boldsymbol{x} \in C$ があって

$$\lim_{k\to\infty} \boldsymbol{x}_k = \boldsymbol{x}$$

となる．さらに T は縮小写像なので連続である．よって

$$T(\boldsymbol{x}) = T\left(\lim_{k\to\infty} \boldsymbol{x}_k\right) = \lim_{k\to\infty} T(\boldsymbol{x}_k) = \lim_{k\to\infty} \boldsymbol{x}_{k+1} = \boldsymbol{x}$$

がわかる．よって定理は証明された． □

定理 A.20. $E \subset \mathbb{R}^n$ を開集合とし，$T :\, E \to \mathbb{R}^n$ を C^1 級とする．すなわち $\dot{T}$ は連続で

$$\begin{aligned}&T(\boldsymbol{x}+\boldsymbol{h}) - T(\boldsymbol{x}) = \dot{T}\,(\boldsymbol{x})(\boldsymbol{h}) + \boldsymbol{r}(\boldsymbol{h}) \qquad (\forall \boldsymbol{x} \in E,\, \boldsymbol{x}+\boldsymbol{h} \in E), \\ &\lim_{\boldsymbol{h}\to\boldsymbol{0}_n} \frac{|\boldsymbol{r}(\boldsymbol{h})|_{2,\,n}}{|\boldsymbol{h}|_{2,\,n}} = 0\end{aligned}$$

をみたすことである．さらにある $\boldsymbol{a} \in E$ に対して $\dot{T}\,(\boldsymbol{a})$ は可逆とし，$\boldsymbol{b} = T(\boldsymbol{a})$ とおく．このとき以下が成立する．

(1) $\mathbb{R}^n$ のある開集合 U と V で $\boldsymbol{a} \in U$ と $\boldsymbol{b} \in V$ なるものがあって

$$T :\, U \to V$$

は 1 対 1 で $T(U) = V$ となる．

(2) V 上で定義さらた T の逆写像を S とする．すなわち

$$S\big(T(\boldsymbol{x})\big) = \boldsymbol{x} \qquad (\forall \boldsymbol{x} \in U)$$

である．このとき S は C^1 級である．

Proof. (1) $\overset{\bullet}{T}(\boldsymbol{a}) = A$ とおき, $\lambda > 0$ を

$$2\lambda \|A\| = 1 \tag{A.9}$$

となるように取る．$\boldsymbol{a}$ を中心とした開球 U が存在して

$$U \subset E$$

とできる．$\overset{\bullet}{T}$ は点 $\boldsymbol{a}$ で連続なので必要ならばU 半径をさらに小さくとると

$$\| \overset{\bullet}{T}(\boldsymbol{x}) - A \| < \lambda \qquad (\boldsymbol{x} \in U) \tag{A.10}$$

とできる．

各 $\boldsymbol{k} \in \mathbb{R}^m$ に対応して関数 φ を

$$\varphi(\boldsymbol{x}) := \boldsymbol{x} + A^{-1}\big(\boldsymbol{y} - T(\boldsymbol{x})\big) \qquad (\boldsymbol{x} \in E) \tag{A.11}$$

と定める．このとき

$$T(\boldsymbol{x}) = \boldsymbol{y} \Leftrightarrow \boldsymbol{x} \text{ は関数 } \varphi \text{ の不動点}$$

に注意する．簡単な計算から

$$\overset{\bullet}{\varphi}(\boldsymbol{x}) = I_n - A^{-1} \overset{\bullet}{T}(\boldsymbol{x}) = A^{-1}\big(A - \overset{\bullet}{T}(\boldsymbol{x})\big)$$

なので (A.10) と (A.11) から

$$\| \overset{\bullet}{\varphi}(\boldsymbol{x}) \| \le \|A\| \cdot \|A - \overset{\bullet}{T}(\boldsymbol{x})\| < \frac{1}{2\lambda} \cdot \lambda = \frac{1}{2} \qquad (\boldsymbol{x} \in U)$$

を得る．いま $\boldsymbol{x}_1, \boldsymbol{x}_2 \in U$ を取る．中間値の定理から $\boldsymbol{x}_1$ と $\boldsymbol{x}_2$ を結ぶ直線上の端点以外の点 $\widetilde{\boldsymbol{x}} \in U$ が存在して

$$\varphi(\boldsymbol{x}_1) - \varphi(\boldsymbol{x}_2) = \overset{\bullet}{\varphi}(\widetilde{\boldsymbol{x}})(\boldsymbol{x}_1 - \boldsymbol{x}_2)$$

と書ける．したがって

$$|\varphi(\boldsymbol{x}_1) - \varphi(\boldsymbol{x}_2)|_{2,n} \le \| \overset{\bullet}{\varphi}(\widetilde{\boldsymbol{x}}) \| \cdot |\boldsymbol{x}_1 - \boldsymbol{x}_2|_{2,n} < \frac{1}{2} |\boldsymbol{x}_1 - \boldsymbol{x}_2|_{2,n} \tag{A.12}$$

となる．よって定理 A.19 から $\boldsymbol{x} \in \overline{U}$ が唯一存在して

$$\varphi(\boldsymbol{x}) = \boldsymbol{x} \ \Leftrightarrow \ \boldsymbol{y} = T(\boldsymbol{x})$$

となる．よって T は $\overline{U}$ 上で 1 対 1 である．

次に $V = T(U)$ とおき $\boldsymbol{y}_0 \in V$ を取る．このときある点 $\boldsymbol{x}_0 \in U$ が唯一存在して

$$\boldsymbol{y}_0 = T(\boldsymbol{x}_0)$$

となる．B を $\boldsymbol{x}_0$ を中心とした半径 $r > 0$ の開球とする．r を十分小さくすると

$$\overline{B} := \{\boldsymbol{z} \in \mathbb{R}^n;\ |\boldsymbol{z} - \boldsymbol{x}_0|_{2,n} < r\} \subset U$$

とできる．このとき

$$\boldsymbol{y} \in V \ \Rightarrow \ |\boldsymbol{y} - \boldsymbol{y}_0|_{2,n} < \lambda r$$

を示すことができると V は開集合であることがわかる．そのために $\boldsymbol{y}$ を固定する．すると

$$\begin{aligned} |\varphi(\boldsymbol{x}_0) - \varphi(\boldsymbol{x})|_{2,n} &= \Big|\big\{\boldsymbol{x}_0 + A^{-1}\big(\boldsymbol{y} - T(\boldsymbol{x}_0)\big)\big\} - \boldsymbol{x}_0\Big|_{2,n} \\ &= \Big|A^{-1}\big(\boldsymbol{y} - T(\boldsymbol{x}_0)\big)\Big|_{2,n} \\ &\leq \|A^{-1}\| \cdot |\boldsymbol{y} - T(\boldsymbol{x}_0)|_{2,n} = \|A^{-1}\| \cdot |\boldsymbol{y} - \boldsymbol{y}_0|_{2,n} \\ &< \frac{1}{2\lambda} \cdot \lambda r = \frac{r}{2} \end{aligned} \tag{A.13}$$

を得る．$\boldsymbol{x} \in \overline{B}$ のとき

$$\begin{aligned} |\varphi(\boldsymbol{x}) - \boldsymbol{x}_0|_{2,n} &= |\varphi(\boldsymbol{x}) - \varphi(\boldsymbol{x}_0)|_n + |\varphi(\boldsymbol{x}_0) - \boldsymbol{x}_0|_{2,n} \\ &\leq \frac{1}{2}|\boldsymbol{x} - \boldsymbol{x}_0|_{2,n} + \frac{r}{2} \qquad \big(\because \text{(A.12) と (A.13)}\big) \\ &\leq \frac{r}{2} + \frac{r}{2} = r \end{aligned}$$

がわかる．これより $\varphi(\boldsymbol{x}) \in B$ となる．さらに $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \overline{B} \subset U$ のとき

$$|\varphi(\boldsymbol{x}_1) - \varphi(\boldsymbol{x}_2)|_{2,n} \leq \frac{1}{2N}|\boldsymbol{x}_1 - \boldsymbol{x}_2|_{2,n}$$

が成立する．よって $\varphi : \overline{B} \to \overline{B}$ は縮小写像となる．$\overline{B}$ は $\mathbb{R}^n$ の閉集合なので点 $\boldsymbol{x} \in \overline{B}$ が唯一存在hして φ の不動点となる．すると

$$T(\boldsymbol{x}) = \boldsymbol{y}$$

をみたすので

$$\boldsymbol{y} \in T\big(\overline{B}\big) \subset T(U) = V$$

がわかる．よって (1) が示せた．
(2) $\boldsymbol{y} \in V,\ \boldsymbol{y}+\boldsymbol{k} \in V$ を取る．すると $\boldsymbol{x} \in U$ と $\boldsymbol{x}+\boldsymbol{h} \in U$ が存在して

$$\boldsymbol{y} = T(\boldsymbol{x}), \quad \boldsymbol{y}+\boldsymbol{k} = T(\boldsymbol{x}+\boldsymbol{h})$$

となる．すると

$$\begin{aligned}
\varphi(\boldsymbol{x}+\boldsymbol{h}) - \varphi(\boldsymbol{x}) &= \bigl\{\boldsymbol{x}+\boldsymbol{h} - A^{-1}\bigl(\boldsymbol{y}+\boldsymbol{k} - T(\boldsymbol{x}+\boldsymbol{h})\bigr)\bigr\} \\
&\qquad - \bigl\{\boldsymbol{x} - A^{-1}\bigl(\boldsymbol{y}+\boldsymbol{k} - T(\boldsymbol{x})\bigr)\bigr\} \\
&= \boldsymbol{h} + A^{-1}\bigl\{T(\boldsymbol{x}) - T(\boldsymbol{x}+\boldsymbol{h})\bigr\} \\
&= \boldsymbol{h} - A^{-1}\boldsymbol{k}
\end{aligned}$$

となる．よって (A.13) から

$$|\boldsymbol{h} - A^{-1}\boldsymbol{k}|_{2,n} = |\varphi(\boldsymbol{x}+\boldsymbol{h}) - \varphi(\boldsymbol{x})|_{2,n} \leq \frac{1}{2}|\boldsymbol{h}|_{2,n}$$

となる．したがって

$$|\boldsymbol{h} - A^{-1}\boldsymbol{k}|_{2,n} \geq |\boldsymbol{h}|_{2,n} \times |A^{-1}\boldsymbol{k}|_{2,m}$$

から

$$|A^{-1}\boldsymbol{k}|_{2,n} \geq \frac{1}{2}|\boldsymbol{h}|_{2,n}$$

と

$$|\boldsymbol{h}|_{2,n} \leq \|A^{-1}\| \cdot |\boldsymbol{k}|_{2,m}\frac{|\boldsymbol{k}|_{2,m}}{\lambda} \tag{A.14}$$

を得る．よって (A.13) と (A.14) から

$$\|\overset{\bullet}{T}(\boldsymbol{x}) - A\| \cdot \|A^{-1}\| < \lambda\frac{1}{2\lambda} = \frac{1}{2}$$

となるので定理 A.10 から $\overset{\bullet}{T}$ は逆写像を持つことがわかる．そこで $\overset{\bullet}{T}(\boldsymbol{x})$ の逆写像を W と書くと

$$\begin{aligned}
S(\boldsymbol{y}+\boldsymbol{k}) - S(\boldsymbol{y}) - W(\boldsymbol{k}) &= S\bigl(T(\boldsymbol{x}+\boldsymbol{h})\bigr) - S\bigl(T(\boldsymbol{x})\bigr) - W(\boldsymbol{k}) \\
&= \boldsymbol{x}+\boldsymbol{h} - \boldsymbol{x} - W(\boldsymbol{k}) \\
&= \boldsymbol{h} - W(\boldsymbol{k}) \\
&= \boldsymbol{h} - W\bigl\{T(\boldsymbol{x}+\boldsymbol{h}) - T(\boldsymbol{x})\bigr\} \\
&= -W\bigl\{T(\boldsymbol{x}+\boldsymbol{h}) - T(\boldsymbol{x}) - \overset{\bullet}{T}(\boldsymbol{x})\boldsymbol{h}\bigr\} \qquad \text{(A.15)} \\
&\qquad \bigl(\because W\overset{\bullet}{T}(\boldsymbol{x}) = I_n\bigr)
\end{aligned}$$

となる．(A.14) から

$$
\begin{aligned}
\frac{\bigl|S(\boldsymbol{y}+\boldsymbol{k})-S(\boldsymbol{y})-W(\boldsymbol{k})\bigr|_{2,m}}{|\boldsymbol{h}|_{2,n}} &= \frac{\bigl|-W\bigl[T(\boldsymbol{x}+\boldsymbol{h})-T(\boldsymbol{x})-\overset{\bullet}{T}(\boldsymbol{x})\boldsymbol{h}\bigr]\bigr|_{2,n}}{|\boldsymbol{h}|_{2,n}} \\
&\leq \frac{\|W\|}{\lambda}\frac{\bigl|T(\boldsymbol{x}+\boldsymbol{h})-T(\boldsymbol{x})-\overset{\bullet}{T}(\boldsymbol{x})\boldsymbol{h}\bigr|_{2,n}}{|\boldsymbol{h}|_{2,n}} \qquad (\because \text{(A.15)})
\end{aligned}
$$

を得る．$\boldsymbol{k}\to\boldsymbol{0}_m$ のとき $\boldsymbol{h}\to\boldsymbol{0}_n$ となり

$$
\begin{aligned}
&\lim_{\boldsymbol{k}\to\boldsymbol{0}_m}\frac{\bigl|S(\boldsymbol{y}+\boldsymbol{k})-S(\boldsymbol{y})-W(\boldsymbol{k})\bigr|_{2,m}}{|\boldsymbol{h}|_{2,n}} \\
&\qquad\leq \lim_{\boldsymbol{h}\to\boldsymbol{0}_n}\frac{\|W\|}{\lambda}\frac{\bigl|T(\boldsymbol{x}+\boldsymbol{h})-T(\boldsymbol{x})-\overset{\bullet}{T}(\boldsymbol{x})(\boldsymbol{h})\bigr|_{2,n}}{|\boldsymbol{h}|_{2,n}} = 0
\end{aligned}
$$

となる．よって

$$
\overset{\bullet}{S}(\boldsymbol{y}) = W
$$

がわかる．W は $\overset{\bullet}{T}(\boldsymbol{x}) = \overset{\bullet}{T}\bigl(S(\boldsymbol{y})\bigr)$ の逆写像になうように選んだ．したがって

$$
\overset{\bullet}{S}(\boldsymbol{y}) = \Bigl\{\overset{\bullet}{T}\bigl(S(\boldsymbol{y})\bigr)\Bigr\}^{-1} \qquad (\boldsymbol{y}\in V)
$$

を得る． □

A.4　連続関数の性質

定理 A.21 (Heine-Borel)**.** $K\subset\mathbb{R}^d$ を部分集合とする．K がコンパクトであるために必要十分条件は，K が閉かつ有界集合であることである．

Proof. □

命題 A.22. $K\subset\mathbb{R}^d$ をコンパクト集合とし，$f\in C(K)$ とする．ただし，$C(K)$ は K 上の実数値連続関数全体のなす集合である．さらに，$M := \sup_{\boldsymbol{x}\in K} f(\boldsymbol{x})$ とおく．このとき，ある $\boldsymbol{x}\in K$ があって，$f(\boldsymbol{x}) = M$ が成立する．

Proof. □

定義 A.23. $D\subset\mathbb{R}^d$ を部分集合とする．関数 $f: D\to\mathbb{R}$ が**一様連続**であるとは

$$
\sup_{\boldsymbol{x},\,\boldsymbol{y}\in D;\,|\boldsymbol{x}-\boldsymbol{y}|_{2,\,d}<\epsilon}|f(\boldsymbol{x})-f(\boldsymbol{y})| \xrightarrow{\epsilon\to 0} 0
$$

が成り立つときをいう．

命題 A.24. $K subset \mathbb{R}^d$ をコンパクト集合とし, $f \in C(K)$ とする. このとき, f は一様連続である.

Proof. □

定理 A.25 (Dini)**.** $K \subset \mathbb{R}^d$ をコンパクト集合とｓる. $\{f_n\}_{n=1}^{\infty} \subset C(K)$ は $f_n \geq 0\,(n \in \mathbb{N})$ かつ $f_1 \geq f_2 \geq \cdots$ で $\lim_{n\to\infty} f_n(\boldsymbol{x}) \to 0\,(\forall \boldsymbol{x} \in K)$ をみたすとする. このとき, $\sup_{\boldsymbol{x}\in K} f_n(\boldsymbol{x}) \overset{n\to\infty}{\longrightarrow} 0$ となる.

Proof. 固定した $\epsilon > 0$ に対して

$$O_n := f_n^{-1}\big((-\infty, \epsilon)\big) = \{\boldsymbol{x} \in K;\, f_n(\boldsymbol{x}) < \epsilon\} \qquad (n \in \mathbb{N})$$

とおく. 関数 f は連続なので, 集合 $O_n\,(n \in \mathbb{N})$ は開であり, 関数列 $f_1, f_2, \ldots$ は減少列なので, $O_1 \subset O_2 \subset \cdots$ である. $f_n(\boldsymbol{x}) \overset{n\to\infty}{\longrightarrow} 0$ なので, n を十分おおくすると $\boldsymbol{x} \in O_n$ とできる. したがって, $\bigcup_{n=1}^{\infty} \supset K$ となる. K はコンパクト部分集合であるので, ある $m \in \mathbb{N}$ と $n_j \in \mathbb{N}\,(j = 1, 2, \ldots, m$ が存在して, $\bigcup_{j=1}^{m} O_{n_j} \supset K$ となる. ここで, $N = \max\{n_1, n_2, \ldots, n_m\}$ とおくと $f_N(\boldsymbol{x}) < \epsilon\,(\forall \boldsymbol{x} \in K)$ となる. 関数列 $\{f_n\}_{n=1}^{\infty}$ は減少列なので, $\forall n \geq N$ と $\forall \boldsymbol{x} \in K$ に対して, $f(\boldsymbol{x}) < \epsilon$ となる. したがって

$$\limsup_{n\to\infty} \sup_{\boldsymbol{x}\in K} f_n(x) \leq \epsilon$$

となる. $\epsilon > 0$ は任意だったので, $\sup_{\boldsymbol{x}\in K} f_n(x) \overset{n\to\infty}{\longrightarrow} 0$ がわかる. □

A.5　章末注釈と参考文献

第B章　数列と級数の収束と関数の性質

B.1　数列と級数

定義 B.1. $\{a_n\}_{n=1}^{\infty}$ を数列とし, a を実数とする.
(1) 任意の実数 $\epsilon > 0$ に対してある自然数 N が存在し, 条件

$$\forall n \geq N \Rightarrow |a_n - a| < \epsilon$$

をみたすとき数列 $\{a_n\}_{n=1}^{\infty}$ は a に**収束**するといい, a を数列 $\{a_n\}_{n=1}^{\infty}$ の**極限**または**極限値**という. 数列 $\{a_n\}_{n=1}^{\infty}$ が a に収束するとき $\lim_{n\to\infty} a_n = a$ あるいは $a_n \overset{n\to\infty}{\longrightarrow} a$ と記す.
(2) 任意の $\epsilon > 0$ に対してある $N \in \mathbb{N}$ が存在し

$$\forall k, \ell \geq N \Rightarrow |a_k - a_\ell| < \epsilon$$

をみたすとき $\{a_n\}_{n=1}^{\infty}$ は **Cauchy 列**であるという.

命題 B.2. (数列の重要な性質) 以下が成立する.

(1) 収束する数列の極限値はただ一つ.

(2) 収束する数列は有界.

(3) $\mathbb{R}$ 内の有界な単調数列は収束する.

(4) 数列 $\{a_n\}_{n=1}^{\infty}$ が $a \in \mathbb{R}$ に収束するための必要十分条件は, すべての部分列が a に収束することである.

(5) (Bolzano-Weierstrass) 有界な数列は収束する部分列を持つ.

(6) 有界な数列 $\{a_n\}_{n=1}^{\infty}$ が収束するための必要十分条件は, その上極限と下極限が同じ有限値であることである.

(7) 数列が収束するための必要十分条件は, それが Cauchy 列であることである.

Proof. 証明は後ほど書く. □

命題 B.3. $\{a_n\}_{n=1}^{\infty}$, a を実数列とする. 数列 $\{a_n\}$ の任意の部分列 $\{a_{n(k)}\}_{k=1}^{\infty}$ が a に収束するさらなる部分列 $\{a_{n(k(\ell))}\}_{\ell=1}^{\infty}$ を持つならば, $a_n \overset{m\to\infty}{\longrightarrow} a$ となる.

Proof. 背理法を用いて証明する. $a_n \not\to a$ と仮定する. あると開区間 G が存在して, $a \in G$ であり, $a_{n(m)}$ の部分列をうまくとれば $a_{n(m)} \notin G\ (\forall m \in \mathbb{N})$ とできる. 明らかに $a_{n(m)} \overset{n\to\infty}{\not\to} a$ となり矛盾. □

注意 B.4. 命題 $B.2(4)$, (5) から, 数列 $\{a_n\}_{n=1}^{\infty}$ が有界のとき, $\{a_n\}_{n=1}^{\infty}$ の収束する任意の部分列 $\{a_{n(k)}\}_{k=1}^{\infty}$ の収束先が同じ $a \in \mathbb{R}$ のとき, $\lim_{n\to\infty} a_n = a$ となる. □

数列 $\{a_n\}_{n=1}^{\infty}$ に対し, その n 項部分列 s_n を

$$s_n = a_1 + a_2 + \cdots + a_n$$

で定義する. $\sum_{n=1}^{\infty} a_n$ の収束と発散を部分和の列 $\{s_n\}_{n=1}^{\infty}$ の収束と発散で定義する.

命題 B.5. (級数の重要な性質) 以下が成立する.

(1) 級数 $\sum_{n=1}^{\infty} a_n$ が収束すれば $\lim_{n\to\infty} a_n = 0$ となる.

(2) 級数 $\sum_{n=1}^{\infty} |a_n|$ が収束すれば $\sum_{n=1}^{\infty} a_n$ も収束する.

(3) 正項級数 $\sum_{n=1}^{\infty} a_n$ が収束するための必要十分条件は, 部分和の列 $\{s_n\}_{n=1}^{\infty}$ が有界であることである.

(4) (d'Alambert の収束判定法)

- $\limsup_{n\to\infty} \dfrac{a_{n+1}}{a_n} < 1 \Rightarrow \sum_{n=1}^{\infty} A_n$ は収束.
- $\liminf_{n\to\infty} \dfrac{a_{n+1}}{a_n} > 1 \Rightarrow \sum_{n=1}^{\infty} A_n$ は発散.

(5) (Cauchy の収束判定法)

- $\limsup_{n\to\infty} \sqrt[n]{a_n} < 1 \Rightarrow \sum_{n=1}^{\infty} A_n$ は収束.
- $\liminf_{n\to\infty} \sqrt[n]{a_n} > 1 \Rightarrow \sum_{n=1}^{\infty} A_n$ は発散.

(6) $\zeta(s) = \sum_{n=1}^{n} \dfrac{1}{n^s}\ (s > 0)$ とする.

- $s > 1 \Rightarrow \zeta(s) < \infty$.
- $0 < s \leq 1 \Rightarrow \zeta(s) = \infty$.

Proof. 証明は後ほど書く. □

B.2 Starling の公式

定理 B.6. (Wallis の公式) 次式が成り立つ.

$$\lim_{n\to\infty} \frac{2^{2n}(n!)^2}{\sqrt{n}(2n)!} = \sqrt{\pi}.$$

Proof. 石谷 (2021, p.189) から借用すること. Stroock (2011, p.32) から借用か. □

定理 B.7. (Starling の公式) 次式が成り立つ.

$$\lim_{n\to\infty} \frac{n!}{\sqrt{2\pi n}}\left(\frac{e}{n}\right)^n = 1.$$

Proof. 石谷 (2021, p.190-191) から借用すること. □

B.3 下半連続関数

定義 B.8. $d \in \mathbb{N}$ とし, $\boldsymbol{x} \in \mathbb{R}^d$ とする. 関数 $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ は $c < f(\boldsymbol{x})$ となる任意の $c \in \mathbb{R}$ に対して $\boldsymbol{x}$ の近傍 U が存在して, $\forall \boldsymbol{y} \in U$ に対して $c < f(\boldsymbol{y})$ となるとき, f は**下半連続**という. 関数 f がすべての $\boldsymbol{x} \in \mathbb{R}^d$ で下半連続のとき, f は $\mathbb{R}^d$ で下半連続とという.

補題 B.9. 関数 $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ は $\boldsymbol{x} \in \mathbb{R}^d$ で下半連続であるための必要十分条件は $\lim_{n\to\infty} \boldsymbol{x}_n = \boldsymbol{x}$ となる任意の点列 $\{\boldsymbol{x}_n\}_{n=1}^\infty \subset \mathbb{R}^d$ で

$$\liminf_{n\to\infty} f(\boldsymbol{x}_n) \geq f(\boldsymbol{x}) \tag{B.1}$$

となることである.

Proof. (⇒ の証明): f は $\boldsymbol{x} = \lim_{n\to\infty} \boldsymbol{x}_n$ で下半連続で $c < f(\boldsymbol{x})$ とする. このとき $\boldsymbol{x}$ の近傍 U が存在して, $\forall \boldsymbol{y} \in U$ に対して

$$c < f(\boldsymbol{y})$$

である. 一方 $\boldsymbol{x}_n \overset{n\to\infty}{\longrightarrow} \boldsymbol{x}$ なので, ある $N \in \mathbb{N}$ が存在して

$$\forall n \geq N \Rightarrow \boldsymbol{x}_n \in U$$

となる. したがって $c < f(\boldsymbol{x})$ なる任意の c に対して

$$\forall n \geq N \ \Rightarrow\ f(\boldsymbol{x}_n) > c$$

となるので

$$\liminf_{n\to\infty} f(\boldsymbol{x}_n) \geq f(\boldsymbol{x})$$

を得る．

(⇐ の証明): $\lim_{n\to\infty} \boldsymbol{x}_n = \boldsymbol{x}$ となる任意の点列 $\{\boldsymbol{x}_n\}_{n=1}^{\infty} \subset \mathbb{R}^d$ で (B.1) をみたすとする．$c < f(\boldsymbol{x})$ とする．すべての $\boldsymbol{y} \in U$ で $c < f(\boldsymbol{y})$ となるような $\boldsymbol{x}$ の近傍 U が存在しないとすると仮定する．すると $\forall n \in \mathbb{N}$ で

$$|\boldsymbol{x} - \boldsymbol{x}_n|_{2,n} < \frac{1}{n}, \quad f(\boldsymbol{x}_n) \le c$$

となる．このとき $\boldsymbol{x}_n \overset{n\to\infty}{\longrightarrow} \boldsymbol{x}$ で

$$\liminf_{n\to\infty} f(\boldsymbol{x}_n) \le c < f(\boldsymbol{x})$$

となり, (B.1) と矛盾する． □

補題 B.10. $\mathbb{R}^d$ のコンパクト部分集合[1] K 上の下半連続関数 f は最小値を K 上で持つ．

Proof. $\mu := \inf_{\boldsymbol{x}\in K} f(\boldsymbol{x})$ とおく．$f(\boldsymbol{x}_n) \overset{n\to\infty}{\longrightarrow} \mu$ となる点列 $\{\boldsymbol{x}_n\}_{n=1}^{\infty} \subset K$ をとると K はコンパクトなので, ある点 $\boldsymbol{x} \in K$ に収束する部分列 $\{\boldsymbol{x}_{n(k)}\}_{k=1}^{\infty}$ が存在する．補題 (B.9) から

$$f(\boldsymbol{x}) \le \liminf_{k\to\infty} f(\boldsymbol{x}_{n(k)}) = \mu$$

となる．一方 μ の定義から $f(\boldsymbol{x}) \ge \mu$ なので $f(\boldsymbol{x}) = \mu$ となる．よって f は最小値 μ を $\boldsymbol{x}$ で取る． □

B.4　章末注釈と参考文献

[1] この場合は有界閉集合と同値.

第C章　測度論

C.1　測度の導入

$\mathbb{X}$ を空でない集合とし, $\mathcal{C}$ を $\mathbb{X}$ の部分集合の集まりで $\varnothing \in \mathcal{C}$ とする. $\mathbb{X}$ の部分集合列 $\{A_n\}_{n=1}^{\infty}$ は**互いに排反**であるとは, $i \neq j$ のとき $A_i \cap A_j = \varnothing$ となるときをいう.

定義 C.1. (1) 関数 $\mu : \mathcal{C} \to \mathbb{R} \cup \{\pm\infty\}$ は**有限加法的**であるとは次の (1a) と (1b) をみたすときをいう.

(1a) $\mu(\varnothing) = 0$.

(1b) $n \in \mathbb{N}$ とする. 互いに排反な集合 $A_j \in \mathcal{C}\,(j = 1, 2, \ldots, n)$ に対して

$$A := \bigcup_{j=1}^{n} A_j \in \mathcal{C} \quad \text{かつ} \quad \mu(A) = \sum_{j=1}^{n} \mu(A_j).$$

(2) $A_j \in \mathcal{C}\,(j = 1, 2, \ldots)$ は互いに排反で

$$B := \bigcup_{j=1}^{\infty} A_j \in \mathcal{C}$$

とする. このとき

$$\mu(B) = \sum_{j=1}^{\infty} \mu(A_j)$$

のとき μ は σ **加法的**という.

注意 C.2. 上の定義において和が定まっているので, ある $i \neq j$ があって $\mu(A_i) = +\infty$, $\mu(A_j) = -\infty$ のようにはならない. □

定義 C.3. (1) 与えられた空でない集合 $\mathbb{X}$ に対して集合族 $\mathcal{A} \subset 2^{\mathbb{X}}$ は**環**であるとは次の (1a) と (1b) をみたすときをいう.

(1a) $\varnothing \in \mathcal{A}$.

(1b) $A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$ かつ $B \setminus A \in \mathcal{A}$.

ただし $2^{\mathbb{X}}$ は $\mathbb{X}$ **の冪集合** を表し, $\mathbb{X}$ のすべての部分集合からなる集合族である.
(2) 環 $\mathcal{A}$ が**代数**であるとは

(2a) $\mathbb{X} \in \mathcal{A}$.

をみたすときをいう.
(3) 代数 $\mathcal{A}$ が σ **代数** であるとは

(3a) 任意の列 $A_n \in \mathcal{A}\,(n = 1,\,2,\,\ldots)$ に対して

$$\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$$

をみたすときをいう.

注意 C.4. (1) $\mathcal{R}$ を環とする. すると $A,\,B \in \mathcal{R}$ のとき

$$A \cap B = A \setminus (A \setminus B) \in \mathcal{R}$$

となる.
(2) $2^{\mathbb{X}}$ は σ 代数である.
(3) 任意の集合族 $\mathcal{C} \subset 2^{\mathbb{X}}$ に対して $\mathcal{C}$ を含む最小の σ 代数が唯一存在する. すなわち $\mathcal{C}$ を含むすべての σ 代数の共通部分である. これを $\mathcal{C}$ **によって生成される** σ **代数**といい, $\sigma[\mathcal{C}]$ と表すことにする.

問 C.1. $\mathcal{C}_1$ と $\mathcal{C}_2$ を σ 代数としたとき $\mathcal{C}_1 \cap \mathcal{C}_2$ も σ 代数であることを示せ.

定理 C.5. $\mathcal{A}$ を代数とする. 有限加法的関数 $\mu : \mathcal{A} \to \mathbb{R}$ が σ 加法的であるための必要十分条件は, 任意の集合列 $A_n \in \mathcal{A}$ で $A_n \supset A_{n+1}\,(n = 1,\,2,\,\ldots)$ で $\bigcap_{n=1}^{\infty} A_n = \varnothing$ なるものに対して

$$\mu(A_n) \overset{n\to\infty}{\longrightarrow} 0$$

をみたすことである.

Proof. ($\Rightarrow$): μ は σ 加法的と仮定し, $A_n \in \mathcal{A},\, A_n \supset A_{n+1}\,(n = 1,\,2,\,\ldots)$ かつ $\bigcap_{n=1}^{\infty} A_n = \varnothing$ とする. このとき $A_n \setminus A_{n+1}$ は互いに排反で

$$A_m = \bigcup_{n=m}^{\infty} \bigl(A_n \setminus A_{n+1}\bigr) \qquad (m \in \mathbb{N})$$

なので

$$\mu(A_m) = \mu\biggl(\bigcup_{n=m}^{\infty} (A_n \setminus A_{n+1})\biggr) = \sum_{n=m}^{\infty} \mu(A_n \setminus A_{n+1})$$

となる. 級数 $\sum_{n=1}^{\infty}\mu(A_n\setminus A_{n+1})$ は収束するので

$$\sum_{n=m}^{\infty}\mu(A_n\setminus A_{n+1})\overset{m\to\infty}{\longrightarrow}0$$

となる. よって $\mu(A_m)\overset{m\to\infty}{\longrightarrow}0$ となる.
($\Leftarrow$): $B_j\in\mathcal{A}\,(j=1,2,\ldots)$ で $B_1,B_2,\ldots,B_n\,(n\in\mathbb{N})$ は互いに排反とする. $B:=\bigcup_{j=1}^{\infty}B_j\in\mathcal{A}$ とし, $A_n:=B\setminus\bigcup_{j=1}^{n-1}B_j$ とおく. すると $A_n\in\mathcal{A}$ となり, $A_n\supset A_{n+1}$ かつ $\bigcap_{n=1}^{\infty}A_n=\varnothing$ となる. μ の有限加法性から各 $n\in\mathbb{N}$ に対して

$$\mu(B)=\mu\left(A_n\cup\left(\bigcup_{j=1}^{n-1}B_j\right)\right)=\mu(A_n)+\mu\left(\bigcup_{j=1}^{n-1}B_j\right)=\mu(A_n)+\sum_{j=1}^{n-1}\mu(B_j)$$

となる. ここで $n\to\infty$ とすると仮定より $\mu(A_n)\to 0$ なので

$$\mu(B)=\sum_{j=1}^{\infty}\mu(B_j)$$

を得る. □

定義 C.6. $\mathbb{X}$ を空でない集合とし, $\mathcal{A}$ を $\mathbb{X}$ の部分集合の σ 代数とする. $\mathcal{A}$ 上の σ 加法測度が $\mu:\mathcal{A}\to[0,\infty)\bigcup\{+\infty\}$ を**測度**という. さらに $(\mathbb{X},\mathcal{A},\mu)$ を**測度空間**という.

注意 C.7. 任意の $A\subset\mathbb{X}$ に対して, A が有限集合のとき, $\mu(A)$ を A の元の個数とし, A が無限集合のとき, $\mu(A)=+\infty$ と定める. すると μ は測度となる. この測度を**計数測度** ($\mathbb{X}$ 上の計数測度) という. □

定理 C.8. 任意の空でない集合 $\mathbb{X}$ と $\mathbb{X}$ の部分集合の環 $\mathcal{A}$ 上の σ 加法的な関数 $\mu:\mathcal{A}\to[0,\infty)\cup\{+\infty\}$ は $\mathcal{A}$ によって生成される σ 加法族 $\sigma[\mathcal{A}]$ 上の測度に拡張できる.

Proof. 任意の $E\subset\mathbb{X}$ に対して

$$\mu^*(E):=\inf\left\{\sum_{n=1}^{\infty}\mu(A_n):\,A_n\in\mathcal{A},\,E\subset\bigcup_{n=1}^{\infty}A_n\right\}$$

と定める. $E\subset\bigcup_{n=1}^{\infty}A_n$ をみたす $\{A_n\}_{n=1}^{\infty}$ が存在しないとき $\mu^*(E)=+\infty$ と約束する. 証明は次の 4 つの補題に分割される.

補題 C.9. 任意の E と $E_n\subset\mathbb{X}\,(n=1,2,\ldots)$ に対して $E\subset\bigcup_{n=1}^{\infty}E_n$ のとき

$$\mu^*(E)\le\sum_{n=1}^{\infty}\mu^*(E_n)$$

となる.

Proof. $\sum_{n=1}^{\infty} \mu^*(E_n) = +\infty$ のときは自明なので, $\sum_{n=1}^{\infty} \mu^*(E_n) < +\infty$ と仮定する. $\forall\epsilon > 0$ をとる. 各 $n \in \mathbb{N}$ に対して $A_{mn} \in \mathcal{A}\,(m = 1, 2, \ldots)$ を

$$E_n \subset \bigcup_{m=1}^{\infty} A_{mn} \quad \text{かつ} \quad \sum_{m=1}^{\infty} \mu(A_{mn}) < \mu^*(E_n) + \frac{\epsilon}{2^n}$$

となるように取る. すると

$$E \subset \bigcup_{n=1}^{\infty}\left(\bigcup_{m=1}^{\infty} A_{mn}\right)$$

なので

$$\mu^*(E) = \sum_{n=1}^{\infty}\left\{\sum_{m=1}^{\infty} \mu(A_{mn})\right\} \leq \sum_{n=1}^{\infty}\left\{\mu^*(E_n) + \frac{\epsilon}{2^n}\right\} \leq \sum_{n=1}^{\infty} \mu^*(E_n) + \epsilon$$

を得る. ϵ は任意だったので, $\epsilon \downarrow 0$ とすれば, 補題の主張は得られる. □

補題 C.10. $\forall A \in \mathcal{A}$ に対して

$$\mu^*(A) = \mu(A).$$

Proof. $A_n \in \mathcal{A}$ かつ $A \subset \bigcup_{n=1}^{\infty} A_n$ とし

$$B_n := A \bigcap\left(A_n \setminus \bigcup_{j=1}^{n-1} A_j\right)$$

とおく. $B_1, B_2, \ldots$ は互いに排反で $B_n \in \mathcal{A}$ かつ $\bigcup_{n=1}^{\infty} B_n = A$ となる. μ は $\mathcal{A}$ 上で σ 加法的なので

$$\mu(A) = \sum_{n=1}^{\infty} \mu(B_n) \leq \sum_{n=1}^{\infty} \mu(A_n)$$

となる. したがって

$$\mu(A) \leq \mu^*(A)$$

を得る.

逆に $A_1 = A$, $A_n = \varnothing\,(n = 2, 3, \ldots)$ と取る. すると

$$\mu^*(A) = \inf\left\{\sum_{n=1}^{\infty} \mu(A_n);\, A \subset \bigcup_{n=1}^{\infty} A_n,\, A_n \in \mathcal{A}\right\}$$

なので

$$\mu^*(A) \leq \mu(A)$$

を得る. よって

$$\mu^*(A) = \mu(A)$$

が示せた. □

定義 C.11. $F \subset \mathbb{X}$ は μ^* 可測であるとは $\forall E \subset \mathbb{X}$ に対して

$$\mu^*(E) = \mu^*(E \cap F) + \mu^*(E \setminus F) \tag{C.1}$$

をみたすときをいう．(C.1) をみたす集合全体を $\mathcal{M}(\mu^*)$ と記すことにする．

補題 C.12. $\mathcal{A} \subset \mathcal{M}(\mu^*)$.

Proof. $A \in \mathcal{A}$ とし, $E \subset \mathbb{X}$ は $\mu^*(E) < +\infty$ をみたすとする．与えられた $\epsilon > 0$ に対して $A_n \in \mathcal{A}\,(n = 1,\, 2,\, \ldots)$ をうまく取ると

$$E \subset \bigcup_{n=1}^{\infty} A_n \quad \text{かつ} \quad \sum_{n=1}^{\infty} \mu(A_n) \leq \mu^*(A) + \epsilon$$

となる．このとき

$$E \cap A = \bigcup_{n=1}^{\infty} (A \cap A_n), \quad E \setminus A \subset \bigcup_{n=1}^{\infty} (A_n \setminus A)$$

なので

$$\begin{aligned} \mu^*(E \cap A) + \mu^*(E \setminus A) &\leq \sum_{n=1}^{n} \Big\{ \mu(A \cap A_n) + \mu(A_n \setminus) \Big\} \\ &= \sum_{n=1}^{\infty} \mu(A_n) \end{aligned}$$

となる．$\epsilon \downarrow 0$ とすれば

$$\mu^*(E \cap A) + \mu^*(E \setminus A) \leq \mu^*(E)$$

を得る．よって $A \in \mathcal{M}(\mu^*)$ が示せた．　□

補題 C.13. $\mathcal{M}(\mu^*)$ は σ 加法族で μ^* は $\mathcal{M}(\mu^*)$ 上の測度となる．

Proof. $\underline{\mathcal{M}(\mu^*)\text{ は代数であることの証明}}$: 明らかに

$$E \in \mathcal{M}(\mu^*) \ \Leftrightarrow \ \mathbb{X} \setminus F \in \mathbb{M}(\mu^*)$$

である．$A,\, B \in \mathcal{M}(\mu^*)$ のとき

$$A \cup B = \mathbb{X} \setminus \big[(\mathbb{X} \setminus A) \cap (\mathbb{X} \setminus B) \big]$$

である. $\forall E \subset \mathbb{X}$ に対して

$$\begin{aligned}
\mu^*(E) &= \mu^*(E \cap A) + \mu^*(E \setminus A) \qquad \big(\because A \in \mathcal{M}(\mu^*)\big) \\
&= \mu^*(E \cap A \cap B) + \mu^*\big((E \cap A)\setminus\big) + \mu^*(E \setminus A) \\
&\qquad \big(\because B \in \mathcal{M}(\mu^*)\big) \\
&= \mu^*\big(E \cap (A \cap B)\big) + \mu^*\big(E \setminus (A \cap B)\big)
\end{aligned}$$

となるので

$$A \cap B \in \mathcal{M}(\mu^*)$$

が示せた. $A \cup B = (A^c \cap B^c)^c$ より $A \cup B \in \mathcal{M}(\mu^*)$ となるので, $\mathcal{M}(\mu^*)$ は代数であることがわかる.

<u>$\mathcal{M}(\mu^*)$ は σ 代数であることの証明</u>: いま $E_n \in \mathcal{M}(\mu^*)\,(n = 1, 2, \ldots)$ とし

$$E := \bigcup_{j=1}^{\infty} E_j, \quad F_n := \bigcup_{j=1}^{n} E_j$$

とおく. $\forall n \in \mathbb{N}$ に対して

$$E_n \setminus \bigcup_{j=1}^{n-1} E_j \in \mathcal{M}(\mu^*) \qquad \big(\because \mathcal{M}(\mu^*) \text{ は代数}\big)$$

となる. このことから $E_1, E_2, \ldots, E_n$ は互いに排反としてよい. $\forall \in \mathbb{X}$ に対して

$$\begin{aligned}
\mu^*(E) &= \mu^*(E \setminus F_n) + \mu^*(E \cap F_n) \qquad \big(\because F_n \in \mathcal{M}(\mu^*)\big) \\
&= \mu^*(E \setminus F_n) + \mu^*(E \cap F_n \cap E_n) + \mu^*\big((E \cap F_n) \setminus E_n\big) \\
&\qquad \big(\because E_n \in \mathcal{M}(\mu^*)\big) \\
&= \mu^*(E \setminus F_n) + \mu^*(E \cap E_n) + \mu^*\left(E \bigcap \left(\bigcup_{j=1}^{n-1} E_j\right)\right) \\
&\qquad (\because E_n \subset F_n)
\end{aligned}$$

となる. n に対する帰納法により

$$\begin{aligned}
\mu^*(E) &\geq \mu^*(E \setminus F_n) + \sum_{j=1}^{n} \mu^*(E \cap E_j) \\
&\geq \mu^*(E \setminus F) + \sum_{j=1}^{n} \mu^*(E \cap E_j)
\end{aligned}$$

を得る. $n \to \infty$ とすると

$$\begin{aligned}\mu^*(E) &\geq \mu^*(E \setminus F) + \sum_{j=1}^{\infty} \mu^*(E \cap E_j) \\ &\geq \mu^*(E \setminus F) + \mu^*(E \cap F)\end{aligned}$$

を得る. なぜならば $\bigcup_{j=1}^{\infty}(E \cap E_j) = E \cup \left(\bigcup_{j=1}^{\infty} E_j\right) = E \cap F$ なので補題 C.9 から $\mu^*(E \cap F) \leq \sum_{n=1}^{\infty} \mu^*(E \cap E_j)$ となることよりわかる. したがって $F \in \mathcal{M}(\mu^*)$ が示せた.
μ^* は $\mathcal{M}(\mu^*)$ 上の σ 加法的測度であることの証明: 再び補題 C.9 から

$$\mu^*(E) = \mu^*(E \setminus F) + \sum_{j=1}^{\infty} \mu^*(E \cap E_j)$$

となる. $E = F$ とおくと μ^* は $\mathcal{M}(\mu^*)$ 上の測度となる.
　$\mathcal{M}(\mu^*)$ は $\mathcal{A}$ を含む σ 代数であるので, $\sigma[\mathcal{A}]$ の定義から

$$\sigma[\mathcal{A}] \subset \mathcal{M}(\mu^*)$$

となる. したがって μ^* は $\sigma[\mathcal{A}]$ 上の測度となる. □

命題 C.14. $E \subset \mathbb{X}$ に対して

$$\mu^*(E) = 0 \Rightarrow E \in \mathcal{M}(\mu^*).$$

となる.

Proof. $\forall A \mathbb{X}$ に対して

$$\begin{aligned}\mu^*(A) &\geq \mu^*(A \setminus F) = \mu^*(A \setminus E) + \mu^*(E) \bigl(\because \mu^*(E) = 0\bigr) \\ &\geq \mu^*(A \setminus E) + \mu^*(A \cap E)\end{aligned}$$

なので $E \in \mathcal{M}(\mu^*)$. □

定義 C.15. $\mathbb{X}$ の部分集合列 $\{A_n\}_{n=1}^{\infty} \subset \mathcal{A}$ が存在して, すべての $n \in \mathbb{N}$ に対して $|\mu(A_n)| < \infty$ かつ $\mathbb{X} = \bigcup_{n=1}^{\infty} A_n$ となるとき, μ は σ 有限測度という.

定理 C.16. μ は代数 $\mathcal{A}$ 上での σ 加法的な非負値関数とする. α を $\sigma[\mathcal{A}]$ 上の測度で $A \in \mathcal{A}$ に対して $\alpha(A) = \mu(A)$ とする. このとき $\forall A \in \sigma[\mathcal{A}]$ で $\mu^*(A) < \infty$ なるものに対して

$$\alpha(A) = \mu^*(A)$$

となる. μ が σ 有限のとき代数 $\mathcal{A}$ からの σ 代数 $\sigma[\mathcal{A}]$ への μ の拡張は一意的である.

Proof. $A \in \sigma[\mathcal{A}]$ で $A_n \in \mathcal{A}\,(n = 1, 2, \ldots)$ で $A \subset \bigcup_{n=1}^{\infty} A_n$ とする．補題 C.9 の証明より

$$\alpha(A) \leq \sum_{n=1}^{\infty} \alpha(A_n) = \sum_{n=1}^{\infty} (A_n) \qquad \bigl(\because A_n \in \mathcal{A}\bigr)$$

となる．すると

$$\alpha(A) \leq \mu^*(A) = \inf\biggl\{\sum_{n=1}^{\infty} \mu(A_n);\, A \subset \bigcup_{n=1}^{\infty} A_n,\, A_n \in \mathcal{A}\biggr\} \tag{C.2}$$

となる．$\mu^*(A) < +\infty$ のとき, 与えられた $\epsilon > 0$ に対して

$$\sum_{n=1}^{\infty} \mu(A_n) \leq \mu^*(A) + \frac{\epsilon}{2}$$

とできる．$k \in \mathbb{N}$ に対して $B_k := \bigcup_{m=1}^{k-1} A_m$ とおくと $B_k \in \mathcal{A}$ で $B_\infty := \bigcup_{m=1}^{\infty} A_m \in \sigma[\mathcal{A}]$ となる．$A \subset B_\infty$ から

$$\mu^*(A) \leq \mu^*(B_\infty) < \mu^*(A) + \frac{\epsilon}{2}$$

となる．補題 C.12 と補題 C.13 から

$$\sigma[\mathcal{A}] \subset \mathcal{M}(\mu^*)$$

となる．定理 C.5 から十分大きな k に対して

$$\mu^*(B_\infty \setminus B_k) < \frac{\epsilon}{2} \tag{C.3}$$

となる．$B_\infty \in \sigma[\mathcal{A}] \subset \mathcal{M}(\mu^*)$ から

$$\mu^*(B_\infty \setminus A) = \mu^*(B_\infty) - \mu^*(A) < \frac{\epsilon}{2}$$

となる．(C.11) から

$$\alpha(B_\infty) \leq \mu^*(B_\infty) < \infty$$

なので

$$\begin{aligned}
\alpha(A) &= \alpha(B_\infty) - \alpha(B_\infty \setminus) \\
&\geq \alpha(B_k) - \mu^*(B_\infty \setminus A) \geq \mu^*(B_k) - \frac{\epsilon}{2} \qquad \bigl(\because \text{(C.12)}\bigr) \\
&\geq \mu^*(B_\infty) - \epsilon \geq \mu^*(A) - \epsilon
\end{aligned}$$

を得る．$\epsilon \downarrow 0$ とすると

$$\alpha(A) \geq \mu^*(A).$$

したがって

$$\alpha(A) = \mu^*(A).$$

さらに μ が σ 加法的なとき, $A_1, A_2, \ldots, A_n$ は互いに排反で $\mu(A_n) < +\infty$ [1]とすると

$$\alpha(A) = \sum_{n=1}^{\infty} \alpha(A \cap A_n) = \sum_{n=1}^{\infty} \mu^*(A \cap A_n) = \mu^*(A)$$

となるので, α の一意性は示せた. □

C.2　半環と環

節 C.1 において, 環 $\mathcal{A}$ 上の σ 有限な関数を σ 代数 $\sigma[\mathcal{A}]$ 上の関数に拡張できることをしました. $\mathbb{X} = \mathbb{R}$ とし, $\mathcal{C} = \{x \in \mathbb{R};\, a < x \leq b\,(a,\, b \in \mathbb{R}\,(a < b)\}$ とし, $C := (a,\, b]$ に対して, $\mathcal{C}$ 上の関数 μ を $\mu(C) = b - a$ で定める. すると μ は $\mathcal{C}$ 上の σ 有限な測度となること[2]がわかる. しかし, $\mathcal{C}$ は環ではない. $\mathcal{C}$ がある性質をみたすと節 C.1 で述べた拡張は有効に働くことがわかる. そのための議論を以下で行う.

定義 C.17. 空でない集合 $\mathbb{X}$ に対して集合族 $\mathcal{D} \subset 2^{\mathbb{X}}$ が**半環**であるとは次の条件をみたすときをいう.

(i) $\varnothing \in \mathcal{D}$.

(ii) $A,\, B \in \mathcal{D} \Rightarrow A \cap B \in \mathcal{D}$ かつ有限個の互いに排反な $C_j \in \mathcal{D}\,(j = 1,\, 2,\, \ldots,\, n;\, n \in \mathbb{N})$ が存在して

$$A \setminus B = \bigcup_{j=1}^{n} C_j$$

と書ける.

命題 C.18. $\mathcal{C} = \{(a,\, b];\, a,\, b \in \mathbb{R}\}$ は半環である.

Proof. $\varnothing \in (a,\, a] \in \mathcal{C}$ となる. $a < c < d < b$ に対して $(a,\, b] \setminus (c,\, d]$ は互いに排反な $\mathcal{C}$ の区間の和となる. □

---

[1]σ 有限なので, このようにできる.
[2]Dudley (2002, pp.87-88) を参照のこと.

命題 C.19. $\mathbb{Y}$, $\mathbb{Z}$ を空でない集合とし, $\mathbb{X} = \mathbb{Y} \times \mathbb{Z}$ とする. $\mathcal{A}$ と $\mathcal{B}$ をそれぞれ $\mathbb{Y}$ と $\mathbb{Z}$ の部分集合族の半環とし

$$\mathcal{D} := \{A \times B;\, A \in \mathcal{A},\, B \in \mathcal{B}\}$$

とする. このとき $\mathcal{D}$ は半環である.

Proof. まず $\varnothing \times \varnothing = \varnothing \in \mathcal{D}$ に注意する. 次に $A, E \in \mathcal{A}$ と $B, F \in \mathcal{B}$ に対して

$$(A \times B) \cap (E \times F) = (A \cap E) \times (B \cap F) \in \mathcal{D}$$

となる. さらに

$$D := (A \times B) \setminus (E \times F) = (A \times B) \setminus \big((A \cap E) \times (b \cap F)\big)$$

なので $E \subset A$ と $F \subset B$ を仮定してよい. A, E は半環 $\mathcal{A}$ の元なので, $m \in \mathbb{N}$ と互いに排反な $G_j \in \mathcal{A}\,(j = 1, 2, \ldots)$ があって, $A \setminus E = \bigcup_{j=1}^m G_j$ と書ける. 同様に, $n \in \mathbb{N}$ と互いに排反な $H_k \in \mathcal{B}\,(k = 1, 2, \ldots, n)$ があって, $B \setminus F = \bigcup_{k=1}^n H_k$ と書ける. このことから

$$\begin{aligned} D &= \big((A \setminus E) \times B\big) \bigcup \big(E \times (B \setminus F)\big) \qquad \big(\because E \subset A,\, F \subset B\big) \\ &= \left\{\bigcup_{j=1}^m (G_j \times B)\right\} \bigcup \left\{\bigcup_{k=1}^n (E \cap H_k)\right\} \end{aligned}$$

となる. $G_j \subset A \setminus E$ なので, $G_j \cap E = \varnothing$ となるので, $\{(G_j \cap B)\}_{j=1}^m$, $\{E \times H_k\}_{k=1}^n$ は互いに排反で $G_j \times B$, $E \times H_k \in \mathcal{D}$ なので, $\mathcal{D}$ は半環であることが示せた. □

命題 C.20. $\mathcal{D}$ を任意の半環とし, $\mathcal{R}$ を $\mathcal{D}$ の有限個の排反な元の和がなす集合全体の集まりとする. このとき $\mathcal{R}$ は環である.

Proof. 明らかに $\varnothing \in \mathcal{R}$ である. 互いに排反な $A_j \in \mathcal{D}\,(j = 1, 2, \ldots, m;\, m \in \mathbb{N})$ と互いに排反な $B_k \in \mathcal{D}\,(k = 1, 2, \ldots, n;\, n \in \mathbb{N})$ に対して

$$A := \bigcup_{j=1}^m A_j, \qquad B := \bigcup_{k=1}^n B_k$$

とおいたとき

$$A \cap B = \bigcup \left\{A_j \cap B_k;\, j = 1, 2, \ldots, m;\, k = 1, 2, \ldots, n\right\}$$

であり, $A_j \cap B_k \in \mathcal{D}$ は互いに排反である. よって $A \cap B \in \mathcal{R}$ となる.

$$A \cup B = A \cup \big(B \setminus A\big)$$

なので, $B \setminus A \in \mathcal{R}$ を示せばよい.

$$B \setminus A = \bigcup_{k=1}^{n} \left\{ B_k \setminus \left(\bigcup_{j=1}^{m} A_j \right) \right\}$$

と書ける. $B_k \setminus \left(\bigcup_{j=1}^{m} A_j \right)$ $(k = 1, 2, \ldots, n)$ は互いに排反なので, あとは

$$B_k \setminus \left(\bigcup_{j=1}^{m} A_j \right) = \bigcap_{j=1}^{m} (B_k \setminus A_j) \in \mathcal{R}$$

を示せばよい. $B_k, A_j \in \mathcal{D}$ なので, 半環の定義から $B_k \setminus A_j \in \mathcal{D} \subset \mathcal{R}$ である. $\mathcal{R}$ の任意の有限個の元の和も $\mathcal{R}$ に含まれるので, $B_k \setminus \bigcup_{j=1}^{m} A_j \in \mathcal{R}$ となる. よって $B \setminus A \in \mathcal{R}$ となる. 以上の議論から $\mathcal{R}$ は環となる. □

定義 C.21. 任意の集合族 $\mathcal{A} \subset 2^{\mathbb{X}}$ に対して $\mathcal{A}$ を含むすべての環の共通部分を $\mathcal{A}$ **によって生成される環**という.

注意 C.22. $\mathcal{A}$ が半環のとき $\mathcal{A}$ によって生成される環は命題 C.20 で与えられたものである. □

命題 C.23. $\mathcal{D}$ を半環とし, α を $\mathcal{D}$ から $[0, \infty)$ への有限加法的な関数とする. 有限個の互いに排反な $D_j \in \mathcal{D}$ $(j = 1, 2, \ldots, m;\ m \in \mathbb{N})$ に対して

$$\mu\left(\bigcup_{j=1}^{m} D_j \right) = \sum_{j=1}^{m} \alpha(D_j) \tag{C.4}$$

とおく. このとき μ は well-defined で $\mathcal{D}$ で生成される環 $\mathcal{R}$ 上で有限加法的である. さらに α が $\mathcal{D}$ 上で σ 加法的ならば μ は環 $\mathcal{R}$ によって生成された σ 加法族 $\sigma[\mathcal{R}]$ 上の測度に拡張できる.

Proof. $C_j \in \mathcal{D}$ $(j = 1, 2, \ldots, m;\ m \in \mathbb{N})$ と $D_k \in \mathcal{D}$ $(k = 1, 2, \ldots, n;\ n \in \mathbb{N}$ はそれぞれ排反とし

$$\bigcup_{j=1}^{m} C_j = \bigcup_{k=1}^{n} D_k$$

とする. 各 $j = 1, 2, \ldots, m$ に対して

$$C_j = \bigcup_{k=1}^{n} (C_j \cap D_k), \quad \{C_j \cap D_k\}_{k=1}^{n} \text{ は互いに排反で } C_j \cap D_k \in \mathcal{D}$$

となる. α の有限加法性から

$$\sum_{j=1}^{m} \alpha(C_j) = \sum_{j=1}^{m} \sum_{k=1}^{n} \alpha(C_j \cap D_k) = \sum_{k=1}^{n} \alpha\left(\left(\bigcup_{j=1}^{m} C_j \right) \bigcap D_k \right) = \sum_{k=1}^{n} \alpha(D_k)$$

とんる．よって α は well-defined である．

次に $n \in \mathbb{N}$ とし, $B_1, B_2, \ldots, B_n \in \mathcal{R}$ は互いに排反とする．すると各 $j = 1, 2, \ldots, n$ に対して $m(j) \in \mathbb{N}$ と $B_{j1}, B_{j2}, \ldots, B_{jm(j)} \in \mathcal{D}$ があって

$$B_j = \bigcup_{k=1}^{m(k)} B_{jk}$$

と書ける．さらに

$$B := \bigcup_{j=1}^{n} B_j$$

とおく．$B = \bigcup_{j=1}^{n} \bigcup_{k=1}^{m(j)} B_{jk}$, $B_{jk} \in \mathcal{D}$ であり, α は well-defined ので

$$\begin{aligned}
\mu(B) &= \sum_{j=1}^{n} \sum_{k=1}^{m(j)} \alpha(B_{jk}) \qquad \bigl(\because (\mathrm{C}.4)\bigr) \\
&= \sum_{j=1}^{n} \left(\sum_{k=1}^{m(j)} \alpha(B_{jk} \right) = \sum_{j=1}^{n} \mu\left(\bigcup_{k=1}^{m(j)} B_{jk} \right) \qquad \bigl(\because (\mathrm{C}.4)\bigr) \\
&= \sum_{j=1}^{n} \mu(B_j)
\end{aligned}$$

となる．よって μ は有限加法的であることがわかる．

次に α は σ 加法的と仮定する．$B \in \mathcal{R}$ とすると $m \in \mathbb{N}$ と互いに排反な $C_j \in \mathcal{D}\,(j = 1, 2, \ldots, n)$ があって,

$$B = \bigcup_{j=1}^{n} C_j$$

と書ける．さらに $A_k \in \mathcal{R}\,(k = 1, 2, \ldots)$ があって

$$B = \bigcup_{k=1}^{\infty} A_k$$

と書けたとする．$A_k \in \mathcal{R}$ なので, $\ell(k) \in \mathbb{N}$ と互いに排反な $A_{k1}, A_{k2}, \ldots, A_{k\ell(k)} \in \mathcal{D}$ があって

$$A_k = \cup_{m=1}^{\ell(k)} A_{km}$$

と書ける．これらから

$$B = \bigcup_{j=1}^{n} \bigcup_{k=1}^{\infty} \bigcup_{m=1}^{\ell(k)} C_j \cap A_{km}$$

となり, $\{C_j \cap A_{km};\ j = 1, 2, \ldots;\ k = 1, 2, \ldots;\ m = 1, 2, \ldots, \ell(k)\}$ の任意の有限個は互いに排反になる. さらに各 A_k は $\mathcal{D}$ の有限個の互いに排反な元の和である. α は $\mathcal{D}$ 上で σ 加法的なので

$$\begin{aligned}
\mu(B) &= \mu\Bigl(\bigcup_{j=1}^{n} C_j\Bigr) = \sum_{j=1}^{n} \alpha(C_j) \qquad \bigl(\because \text{(C.4)}\bigr) \\
&= \sum_{j=1}^{n} \alpha\Bigl(C_j \cap \Bigl(\bigcup_{k=1}^{\infty} A_k\Bigr)\Bigr) = \sum_{j=1}^{n} \alpha\Bigl(\bigcup_{k=1}^{\infty}\bigcup_{m=1}^{\ell(k)} (C_j \cap A_{km})\Bigr) \\
&= \sum_{j=1}^{n}\sum_{k=1}^{\infty}\sum_{m=1}^{\ell(k)} \alpha(C_j \cap A_{km}) \qquad \bigl(\alpha \text{ の } \sigma \text{ 加法性}\bigr) \\
&= \sum_{k=1}^{\infty}\sum_{j=1}^{n}\sum_{m=1}^{\ell(k)} \alpha(C_j \cap A_{km}) \\
&= \sum_{k=1}^{\infty} \mu\Bigl(\Bigl(\bigcup_{j=1}^{n} C_j\Bigr) \bigcap \Bigl(\bigcup_{m=1}^{\ell(k)} A_{km}\Bigr)\Bigr) \qquad \bigl(\because \text{(C.4)}\bigr) \\
&= \sum_{k=1}^{\infty} \mu(B \cap A_k) = \sum_{k=1}^{\infty} \mu(A_k) \qquad \bigl(\because A_k \subset B\bigr)
\end{aligned}$$

となり, μ は $\mathcal{R}$ 上で σ 加法的になる. よって定理 C.8 から μ は $\sigma[\mathcal{R}]$ 上の測度に拡張できる. □

定義 C.24. $\mathbb{X}$ を空でない集合とする. $\mathbb{X}$ の部分集合族 $\mathcal{R} \subset 2^{\mathbb{X}}$ が**σ 環**であるとは次の条件 (i)-(iii) をみたすときをいう.

(i) $\varnothing \in \mathcal{R}$.

(ii) $\forall A, B \in \mathcal{R} \Rightarrow A \setminus B \in \mathcal{R}$.

(iii) $A_j \in \mathcal{R}\,(j = 1, 2, \ldots) \Rightarrow \bigcup_{j=1}^{\infty} A_j \in \mathcal{R}$.

注意 C.25. (1) 任意の σ 代数は σ 環である. 逆に $\mathcal{R}$ が σ 環のとき

$$\mathcal{R} \text{ が } \sigma \text{ 環} \iff \mathbb{X} \in \mathcal{R}$$

である.
(2) $\mathcal{R}$ の任意の可算部分集合全体のなす族は σ 環であるが, σ 代数ではない.
(3) f を $\mathbb{X}$ 上の実数値関数とし, $\mathcal{R}$ を $\mathbb{X}$ の部分集合の σ 環とする. このとき f は可測であるとは, 0 を含まない任意の $B \in \mathcal{B}(\mathbb{R})$ に対して

$$f^{-1}(B) = \{x \in \mathbb{X};\ f(x) \in B\} \in \mathcal{R}$$

である.
(4) $\mathcal{C}$ を $\mathbb{X}$ の部分集合族とする. $\mathcal{C}$ を含む最小の σ 環を $\mathcal{C}$ **によって生成される** σ **環**という. □

定理 C.26. $(\mathbb{X}, \mathcal{A})$ と $(\mathbb{Y}, \mathcal{B})$ を可測空間とする. $\mathcal{C}$ を $\mathbb{Y}$ の部分集合族とし, $\mathcal{B}$ は $\mathcal{C}$ によって生成されるとする. 関数 $f : \mathbb{X} \to \mathbb{Y}$ が可測であるための必要十分条件は

$$f^{-1}(C) \in \mathcal{A} \quad (\forall C \in \mathcal{C})$$

である. $\mathbb{Y} = \mathbb{R}$ とし, $\mathcal{A}$ を $\mathbb{X}$ 上の σ 環とする. このとき f が可測であるための必要十分条件は

$$f^{-1}(B) \in \mathcal{A} \quad (\forall B \in \mathcal{B}(\mathbb{R}), 0 \notin B)$$

である.

Proof. $(\Rightarrow)$ は明らかである.
$(\Leftarrow)$ の証明:

$$\mathcal{D} := \{D \in \mathcal{B};\ f^{-1} \in \mathcal{A}\}$$

とおく. $\mathcal{C} \subset \mathcal{D}$ となる. 各 $n \in \mathbb{N}$ に対して $D_n \in \mathcal{D}$ のとき

$$f^{-1}\left(\bigcup_{n=1}^{\infty} D_n\right) = \bigcup_{n=1}^{\infty} f^{-1}(D_n) \in \mathcal{A}$$

なので, $\bigcup_{n=1}^{\infty} D_n \in \mathcal{D}$ となる. $D, E \in \mathcal{D}$ に対して

$$f^{-1}(D \setminus E) = f^{-1}(D) \setminus f^{-1}(E) \in \mathcal{A}$$

より $D \setminus E \in \mathcal{D}$ がわかる. したがって $\mathcal{D}$ は σ 環である. $\mathcal{A}$ が σ 代数のとき $f^{-1}(\mathbb{Y}) = \mathbb{X} \in \mathcal{S}$ なので $\mathbb{Y} \in \mathcal{D}$ となる. よって 注意 C.25(1) から $\mathcal{D}$ は σ 代数. $\mathcal{D}$ の作り方から $\mathcal{C} \subset \mathcal{D} \subset \mathcal{B}$ である. しかし $\mathcal{B}$ は $\mathcal{C}$ を含む最小の σ 代数なので, $\mathcal{B} \subset \mathcal{D}$ となる. したがって $\mathcal{B} = \mathcal{D}$ がわかる. □

注意 C.27. (1) 定理 C.26 の記号を踏襲する. $\mathbb{R}$ 上の Borel 集合族 $\mathbb{B}(\mathbb{R})$ を生成する $\mathbb{R}$ の部分集合族として $\{(t, \infty);\ t \in \mathbb{R}\}$ がある. よって $f : \mathbb{X} \to \mathbb{R}$ の可測性を示すために

$$\{x \in \mathbb{X};\ f(x) > t\} \in \mathbb{A}$$

を示せばよい.

(2) $(\mathbb{Y}, \mathcal{B}), (\mathbb{Z}, \mathcal{C})$ を可測空間とし，$\mathbb{Y} \times \mathbb{Z}$ の部分集合族 $\{B \times C;\, B \in \mathcal{B}, C \in \mathcal{C}\}$ によって生成される σ 代数を $\mathcal{B} \otimes \mathcal{C}$ と記す．この $\mathcal{B} \otimes \mathcal{C}$ を直積空間 $\mathbb{Y} \times \mathbb{Z}$ 上の直積 σ 代数という．
(3) $(\mathbb{X}, \mathcal{A})$ を可測空間とし，$f : \mathbb{X} \to \mathbb{Y}$ と $g; \mathbb{X} \to \mathbb{Z}$ を関数とする．関数 $h : \mathbb{X} \to \mathbb{Y} \times \mathbb{Z}$ を

$$h(x) := (f(x), g(x))$$

と定める．命題 C.23 から

$$h : \mathbb{X} \to \mathbb{Y} \times \mathbb{Z} \text{ は可測} \quad \Longleftrightarrow \quad f \text{ と } g \text{ はぞれぞれ可測}$$

となる． □

C.3 Lebeague-Stieltjes 測度

$\mathbb{R}$ の部分集合族 $\mathcal{D}$ を

$$\mathcal{D} = \{(a, b];\, a, b \in \mathbb{R}\}$$

で定めたとき, $\mathcal{D}$ は半環である．$\mathcal{D}$ の有限個の互いに排反な元の和で書ける集合全体の集まりを $\mathcal{R}$ とすると $\mathcal{R}$ は環となる．さらに $\mathbb{R} \in \mathcal{R}$ であれば $\mathcal{R}$ は代数となる．$F : \mathbb{R} \to \mathbb{R}$ を非増加右連続関数とし, F に対応する $\mathcal{D}$ 上の集合関数 μ_F を

$$\mu_F((a, b]) = F(b) - F(a) \qquad ((a, b] \in \mathcal{D})$$

で定める．

定理 C.28. 環 $\mathcal{R}$ 上の集合関数 μ_F は環 $\mathcal{R}$ 上の σ 加法的測度である．

Proof. σ 加法性を示すために

$$A = (a, b],\, A_j = (a_j, b_j] \in \mathcal{R}\, (j = 1, 2, \ldots)$$

とする．ただし $A_1, A_2, \ldots$ は互いに排反とする．さらに

$$A = \bigcup_{j=1}^{\infty} A_j$$

とする．$\forall n \in \mathbb{N}$ に対して $\bigcup_{j=1}^{n} A_j \subset A$ なので

$$\sum_{j=1}^{n} \mu_F(A_j) \le \mu_F(A)$$

となる．したがって $n \to \infty$ とすると

$$\sum_{j=1}^{\infty} \mu_F(A_j) \le \mu_F(A) \tag{C.5}$$

を得る．反対向きの不等号を示すために $\epsilon > 0$ を固定する．関数 F の右連続性から $\delta > 0$ と $\delta_j > 0\,(j = 1,\, 2,\, \ldots$ をうまく取ると

$$\mu_F((a,\, a+\delta]) = F(a+\delta) - F(a) \le \epsilon$$

と

$$\mu_F((b_j,\, b_j+\delta_j]) = F(b_j+\delta) - F(b_j) \le \frac{\epsilon}{2^j}$$

とできる．このとき

$$[a+\delta,\, b] \subset \bigcup_{j=1}^{\infty} (a_j,\, b_j+\delta_j)$$

となる．$[a+\delta,\, b]$ はコンパクトなので, $k \in \mathbb{N}$ が存在して

$$[a+\delta,\, b] \subset \bigcup_{j=1}^{k} (a_j,\, b_j+\delta_j)$$

とできる．必要ならば順番を入れ替えればよい．μ_F の非負性と有限加法性から

$$\begin{aligned}
\mu_F\big((a+\delta,\, b]\big) &= \mu_F\left(\bigcup_{j=1}^{k} (a_j,\, b_j+\delta_j)\right) = \sum_{j=1}^{k} \mu_F\big((a_j,\, b_j+\delta_j]\big) \\
&\le \sum_{j=1}^{\infty} \mu\big((a_j,\, b_j+\delta_j]\big) \\
&\le \sum_{j=1}^{\infty} \mu_F\big((a_j,\, b_j]\big) + \sum_{j=1}^{\infty} \mu\big((b_j,\, b_j+\delta_j]\big)
\end{aligned}$$

を得る．δ と ϵ の取り方から

$$\begin{aligned}
\mu_F\big((a,\, b]\big) &\le \mu_F\big((a,\, a+\delta]\big) + \mu_F\big((a+\delta,\, b]\big) \\
&\le \epsilon + \sum_{j=1}^{\infty} \mu_F\big((a_j,\, b_j]\big) + \sum_{j=1}^{\infty} \mu_F\big((b_j,\, b_j+\delta_j]\big) \\
&\le \epsilon + \sum_{j=1}^{\infty} \mu_F\big((a_j,\, b_j]\big) + \sum_{j=1}^{\infty} \frac{\epsilon}{2^j} \\
&\le \sum_{j=1}^{\infty} \mu_F\big((a_j,\, b_j]\big) + 2\epsilon
\end{aligned}$$

となる. $\epsilon \to 0$ とすると

$$\mu_F\big((a,\,b]\big) \le \sum_{j=1}^{\infty} \mu_F\big((a_j,\,b_j]\big) \tag{C.6}$$

を得る. (C.5) と (C.6) から

$$\mu_F(A) = \sum_{j=1}^{\infty} \mu_F\big((a_j,\,b_j]\big)$$

を得る. $\mathcal{R}$ は $\mathcal{D}$ の排反な有限個の元の和で書けるので, 上記の議論から μ_F は $\mathcal{R}$ 上の σ 加法的速度となることがわかる. □

定理 C.28 と C.16 から μ_F は $\sigma[\mathcal{R}]$ 上の測度 $\tilde{\mu}_F$ に一意的に拡張できる. すなわち $\tilde{\mu}_F$ は $\mathcal{R}$ を含む最小の σ 加法族 $\sigma[\mathcal{R}]$ 上の測度で

$$\tilde{\mu}_F(R) = \mu_F(R) \qquad R \in \mathcal{R}$$

である. この拡張された測度 $\tilde{\mu}_F$ のことを非増加右連続関数 F に対応する Lebeague-Stieltjes 測度という.

つぎの命題は定理 C.28 の逆を主張するものである.

命題 C.29. $\mathcal{D} = \{(a,\,b];\, a,\,b \in \mathbb{R}\}$ とし, $\mu : \mathcal{D} \to [0,\,\infty)$ は σ 加法的集合関数とする. このとき非増加右連続関数 $F : \mathbb{R} \to \mathbb{R}$ が存在して

$$\mu\big((a,\,b]\big) = F(b) - F(a) \qquad (\forall a,\,b \in \mathbb{R},\, a < b)$$

と書ける.

Proof.

$$F(x) := \begin{cases} \mu\big((0,\,x]\big) & (x > 0) \\ 0 & (x = 0) \\ -\mu\big((x,\,0]\big) & (x < 0) \end{cases}$$

と定める. $b \ge a \ge 0$ のとき μ は加法的なので

$$\mu\big((a,\,b]\big) = \mu\big((0,\,b]\big) - \mu\big((0,\,a]\big) = F(b) - F(a)$$

となる. $a \le b \le 0$ のときも同様の式が得られる. つぎに F は非減少関数であることを示す.

$$a \le b \;\Rightarrow\; F(b) - F(a) = \mu\big((a,\,b]\big) \ge 0 \;\Rightarrow\; F(a) \le F(b)$$

よりわかる．最後に F の右連続性を示す．$\{a_n\}_{n=1}^{\infty} \subset \mathbb{R}$ は非増加列で $a_n \downarrow (n \to \infty)$ とする．このとき

$$(b,\, a_] = \sum_{j=1}^{\infty} (a_{j+1},\, a_j)$$

と書けるので, μ の σ 加法性から

$$\begin{aligned} F(a_1) - \mathsf{F}(b) = \mu\bigl((0,\, a_1]\bigr) - \mu\bigl((0,\, b]\bigr) = \mu\bigl(b,\, a_1]\bigr) = \sum_{j=1}^{\infty} \mu\bigl((a_{j+1},\, a_j]\bigr) \\ = \sum_{j=1}^{\infty} \bigl\{F(a_j) - F(a_{j+1})\bigr\} \\ = F(a_1) - \lim_{j\to\infty} F(a_j) \\ \bigl(\because \text{上の式の級数は絶対収束するので, 順序を入れ替えてよい}\bigr) \end{aligned}$$

となる．よって

$$F(b) = \lim_{j\to\infty} F(a_j)$$

となるので, F の右連続性が示せた．□

以上の議論から次の主張を得る．

命題 C.30. $\mu :, \mathcal{B}\bigl(\mathbb{R}\bigr) \to [0,\, \infty) \cup \{+\infty\}$ を測度とし, $\mathbb{R}$ の有界区間に対して有限な値を取るものとする．このとき非増加右連続関数 $F : \mathbb{R} \to \mathbb{R}$ が存在して

$$\mu\bigl((a,\, b]\bigr) = F(b) - F(a) \qquad \bigl(\forall a,\, b \in \mathbb{R},\, a < b\bigr)$$

をみたす．

C.4　Dynkin システムと Lebeague 測度の一意性

定義 C.31. $\mathbb{X}$ を空でない集合とし, $\mathcal{D} \subset 2^{\mathbb{X}}$ とする．$\mathcal{D}$ は Dynkin **システム**であるとは次の条件 (D_1) – (D_3) をみたすときをいう．

(D_1) $\mathbb{X} \in \mathcal{D}$,

(D_2) $\forall D \in \mathcal{D} \ \Rightarrow\ D^c \in \mathcal{D}$,

(D_3) $\{D_n\}_{n=1}^{\infty} \subset \mathcal{D}$ は互いに排反なとき

$$\bigcup_{n=1}^{\infty} D_n \in \mathcal{D}.$$

注意 C.32. $\mathcal{D}$ が σ 代数ならば $\mathcal{D}$ は Dynkin システムである．しかし逆は一般には真ではない．たとえば $k \in \mathbb{N}$ と $\mathbb{X} = \{1, 2, \ldots, 4k-1, 4k\}$ とする．このとき

$$\mathcal{D} := \{D \subset \mathbb{X};\, \#(D) = \text{偶数}\}$$

と定めると $\mathcal{D}$ は Dynkin システムであることが確認できる．しかし σ 代数ではない．たとえば $D_1 := \{1, 3\}$, $D_2 := 3, 4$ とすると $D_1 \cap D_2 = \{3\} \notin \mathcal{D}$ となることに注意せよ． □

命題 C.33. $\mathbb{X}$ を空でない集合とし, $\mathcal{G} \subset 2^{\mathbb{X}}$ を $\mathbb{X}$ の任意の部分集合族とする．このとき $\mathcal{G}$ を含む最小の Dynkin システムが存在する．これを $\mathcal{G}$ **によって生成された Dynkin システム**といい, $\delta[\mathcal{G}]$ と記す．さらに

$$\mathcal{G} \subset \delta[\mathcal{G}] \subset \sigma[\mathcal{G}]$$

となる．

Proof.

$$\mathcal{D} := \bigcap_{\mathcal{F}\text{ は Dynkin システムで }\mathcal{F} \subset \mathcal{G}} \mathcal{F}$$

とおく． $2^{\mathbb{X}}$ は Dynkin システムなので $\mathcal{D} \neq \varnothing$ である． Dynkin システムである部分集合族の共通部分は Dynkin システムになるので $\mathcal{D}$ は Dynkin システムである．つぎに $\mathcal{D}$ の最小性を示すために $\mathcal{G}$ を含む任意の Dynkin システム $\widetilde{\mathcal{D}}$ を取る．$\mathcal{D}$ の定義から $\mathcal{D} \subset \widetilde{\mathcal{D}}$ が直ちにわかる．よって $\mathcal{D}$ の最小性が示せた．最後の主張を示す．$\mathcal{D}$ が Dynkin システムのとき $\mathcal{D} = \delta[\mathcal{D}]$ である．このことと $\sigma[\mathcal{D}]$ も Dynkin システムであることから

$$\delta\big[\sigma[\mathcal{D}]\big] = \sigma[\mathcal{G}]$$

となる．よって $\mathcal{G} \subset \sigma[\mathcal{G}]$ かつ $\sigma[\mathcal{G}]$ は Dynkin システムである．$\delta[\mathcal{G}]$ の最小性から

$$\delta[\mathcal{G}] \subset \sigma[\mathcal{G}]$$

がわかる． □

補題 C.34. Dynkin システム $\mathcal{D}$ が σ 代数であるための必要十分条件は

$$\forall D,\, E \in \mathcal{D} \;\Rightarrow\; D \cap E \in \mathcal{D} \tag{C.7}$$

である.

Proof. (1) をみたす Dynkin システムは σ 代数であることを示せばよい. そのために $\{D_n\}_{n=1}^{\infty} \subset \mathcal{D}$ に対して

$$D := \bigcup_{n=1}^{\infty} D_n \in \mathcal{D}$$

を示せばよい. $E_1 := D_1 \in \mathcal{D}$ とおき $n \geq 1\,(n \in \mathbb{N})$ に対して

$$\begin{aligned} E_{n+1} &:= \underbrace{\Bigl(\cdots\Bigl(}_{n\text{ の括弧}} \bigl(D_{n+1} \setminus D_n\bigr) \setminus D_{n-1}\Bigr) \cdots \Bigr) \setminus D_1 \\ &= D_{n+1} \bigcup D_n^c \bigcup D_{n-1}^c \bigcup \cdots \bigcup D_1^c \in \mathcal{D} \end{aligned}$$

となる. 上の式の最後の等号は (C.7) を用いた. すると $\{E_n\}_{n=1}^{\infty} \subset \mathcal{D}$ は互いに排反となるので Dynkin システムの条件 D_3 から

$$D = \bigcup_{n=1}^{\infty} E_n \in \mathcal{D}$$

となる. よって $\mathcal{D}$ は σ 代数となる. □

定理 C.35. $\mathcal{G} \subset 2^{\mathbb{X}}$ を $\mathbb{X}$ の任意の部分集合族とし,

$$\forall D,\, E \in \mathcal{G}, \Rightarrow\; D \cap D \in \mathcal{G}$$

をみたすとする.

Proof. 命題 C.33 から $\delta[\mathcal{G}] \subset \sigma[\mathcal{G}]$ である. したがって $\delta[\mathcal{G}] \supset \sigma[\mathcal{G}]$ を示せばよい. $\delta[\mathcal{G}]$ が σ 代数であることがわかれば $\sigma[\mathcal{G}]$ の最小性から $\delta[\mathcal{G}] \supset \sigma[\mathcal{G}]$ がわかる.

$\delta[\mathcal{G}]$ が σ 代数であることを確認するためには

$$\forall D,\, E \in \delta[\mathcal{G}] \;\Rightarrow\; D \cap E \in \delta[\mathcal{G}] \tag{C.8}$$

がわかえれば補題 C.34 から $\delta[\mathcal{G}]$ は σ 代数であることがわかる. そのために $D \in \delta[\mathcal{G}]$ を固定する. $\mathbb{X}$ の部分集合族 $\mathcal{D}_D$ を

$$\mathcal{D}_D := \bigl\{Q \in \mathbb{X};\, Q \cap D \in \delta[\mathcal{G}]\bigr\}$$

と定めたとき, $\mathcal{D}_D$ は Dynkin システムであることを示す．$\mathbb{X} \in \mathcal{D}_D$ は明らかなので $\mathcal{D}_D$ は Dynkin システムの定義 C.31 の条件 D_1 をみたす．つぎに $Q \in \mathcal{D}_D$ を取る．すると

$$Q^c \cap D := \bigl(Q^c \cap D\bigr)^c \cap D = \Bigl\{\underbrace{(Q \cap D)}_{\in\delta[\mathcal{G}]} \bigcup \underbrace{D^c}_{\in\delta[\mathcal{G}]}\Bigr\}^c$$

となる．$Q \cap D$ と D^c は排反なので

$$(Q \cap D) \cup D^c \in \delta[] \ \Rightarrow\ Q^c \cap D \in \delta[\mathcal{G}]$$

である．よって $Q^c \in \mathcal{D}_D$ となる．$\{D_n\}_{n=1}^{\infty} \subset \mathcal{D}_D$ は互いに排反とする．すると $\{Q_n \cap D\}_{n=1}^{\infty}$ は互いに排反かつ $Q_n \cap D \in \delta[\mathcal{G}]\,(n = 1, 2, \ldots)$ となることから

$$\left(\bigcup_{n=1}^{\infty} Q_n\right) \bigcap D = \bigcup_{n=1}^{\infty} (Q_n \cap D) \in \delta[\mathcal{G}]$$

となる．よって $\bigcup_{n=1}^{\infty} Q_n \in \mathcal{D}_D$ がわかるので $\mathcal{D}_D$ は Dynkin システムであることが示せた．$\delta[\mathcal{G}]$ の最小性から

$$\delta[\mathcal{G}] \subset \mathcal{D}_G$$

となる．$\mathcal{D}_G$ の定義から

$$\forall G \in \mathcal{G}, \forall D \in \delta[\mathcal{G}] \ \Rightarrow\ D \cap G \in \delta[\mathcal{G}]$$

となる．このことから

$$\forall G \in \mathcal{G}, \forall D \in \delta[\mathcal{G}] \ \Rightarrow\ G \in \mathcal{D}_D$$

なので $\forall D \in \delta[\mathcal{G}]$ に対して

$$\mathcal{G} \subset \mathcal{D}_D$$

がわかる．よって $\mathcal{D}_D$ は $\mathcal{G}$ を含む Dynkin システムなので

$$\delta[\mathcal{G}] \subset \mathcal{D}_D \qquad (\forall D \in \delta[\mathcal{G}])$$

となる．Dynkin システム $\delta[\mathcal{G}]$ が

$$D, E \in \delta[\mathcal{G}] \ \Rightarrow\ D \cap E \in \delta[\mathcal{G}]$$

をみたしていることを意味する．よって補題 C.34 から $\delta[\mathcal{G}]$ は σ 代数であることがわかる．以上から補題は証明された． □

定理 C.36. $(\mathbb{X}, \mathcal{A})$ を測度空間とする．$\mathbb{X}$ の部分集合族 $\mathcal{G}$ は以下の条件 (1) – (3) をみたすとする．

(1) $\forall G, H \in \mathcal{G} \Rightarrow G \cap H \in \mathcal{G}$,

(2) $\{G_n\}_{n=1}^{\infty} \subset \mathcal{G}$ があって $G_n \subset G_{n+1}\,(n = 1, 2, \ldots)$ かつ $\bigcup_{n=1}^{\infty} G_n = \mathbb{X}$,

(3) $\mathcal{A} = \sigma[\mathcal{G}]$.

このとき $(\mathbb{X}, \mathcal{A})$ 上の 2 つの測度 μ と ν が

$$\mu(G) = \nu(G) \quad (\forall \in \mathcal{G}) \qquad \text{かつ} \qquad \mu(G_n) = \nu(G_n) < \infty\,(n = 1, 2, \ldots)$$

ならば

$$\mu(A) = \nu(A) \qquad (\forall A \in \mathcal{A})$$

となる．

Proof. $n \in \mathbb{N}$ に対して

$$\mathcal{D}_n := \big\{A \in \mathcal{A};\, \mu(G_n \cap A)\nu(G_n \cap A) < \infty\big\}$$

とおく．すると $\forall n \in \mathbb{N}$ に対して $\mathcal{D}_n$ は Dynkin システムとなることを示す．まず $\mathbb{X} \in \mathcal{D}_n$ は明らか．$A \in \mathcal{D}_n$ を取る．すると

$$\begin{aligned}
\mu(G_n \cap A^c) &= \mu(G_n \setminus A) = \mu(G_n) - \mu(G_n \cap A) \\
&= \nu(G_n) - \nu(G_n \cap A) \quad \big(\because G_n \in \mathcal{G}\,\text{かつ}\,G_n \cap A \in \mathcal{D}_n\big) \\
&= \nu(G_n \setminus A) = \nu(G_n \cap A^c)
\end{aligned}$$

となるので, $A^c \in \mathcal{D}_n$ となる．最後に $\{A_k\}_{k=1}^{\infty} \subset \mathcal{D}_n$ は互いに排反とする．このとき

$$\begin{aligned}
\mu\bigg(G_n \bigcap \bigg(\bigcup_{k=1}^{\infty} A_k\bigg)\bigg) &= \nu\bigg(\bigcup_{k=1}^{\infty} G_n \cap A_k\bigg) = \sum_{k=1}^{\infty} \mu(G_n \cap A_k) \\
&= \sum_{k=1}^{\infty} \nu(G_n \cap A_k) \quad \big(\because A_k \in \mathcal{D}_n\big) \\
&= \nu\bigg(\bigcup_{k=1}^{\infty} G_n \cap A_k\bigg) = \nu\bigg(G_n \bigcap \bigg(\bigcup_{k=1}^{\infty} A_k\bigg)\bigg)
\end{aligned}$$

となるので, $\bigcup_{k=1}^{\infty} A_k \in \mathcal{D}_n$ となる．

さらに $\mathcal{G}$ は条件 (i) をみたしているので定理 C.35 から

$$\delta[\mathcal{G}] = \sigma[\mathcal{G}]$$

となる．したがって

$$\mathcal{D}_n \subset \mathcal{G}, \quad \text{かつ} \quad \mathcal{D}_n \text{ は Dynkin システム } \Rightarrow \mathcal{D}_n \subset \delta[\mathcal{G}] = \sigma[\mathcal{G}] \quad (\forall n \in \mathbb{N})$$

となる．一方

$$\mathcal{A} = \sigma[\mathcal{G}] \subset \mathcal{D}_n = \mathcal{A}$$

なので $\mathcal{A} = \mathcal{D}_n\ (\forall n \in \mathbb{N})$ である．このことより

$$\mu(G_n \cap A) = \nu(G_n \cap A) \qquad (\forall n \in \mathbb{N},\ \forall A \in \mathcal{A})$$

となる．よって $\forall A \in \mathcal{A}$ に対して

$$\begin{aligned}
\mu(A) &= \mu(A \cap \mathbb{X}) = \mu\left(A \cap \left(\bigcup_{n=1}^{\infty} G_n\right)\right) = \mu\left(\bigcup_{n=1}^{\infty} A \cap G_n\right) \\
&= \lim_{n\to\infty} \mu(A \cap G_n) \quad \bigl(\because A \cap G_n \subset A \cap G_{n+1}\bigr) \\
&= \lim_{n\to\infty} \nu(A \cap G_n) = \nu\left(\bigcup_{n=1}^{\infty} A \cap G_n\right) = \nu\left(A \cap \left(\bigcup_{n=1}^{\infty} G_n\right)\right) \\
&= \nu(A \cap \mathbb{X}) = \nu(A)
\end{aligned}$$

となり, 定理は証明された. □

定理 C.37. (1) n 次元 Lebeague 測度 λ_n は

$$\lambda_n(\boldsymbol{x} + B) = \lambda_n(B) \qquad \bigl(\forall \boldsymbol{x} \in \mathbb{R}^n,\ \forall B \in \mathcal{B}(\mathbb{R}^n)\bigr)$$

をみたす．ただし $\boldsymbol{x} + B := \{\boldsymbol{x} + \boldsymbol{y} \in \mathbb{R}^n;\ \boldsymbol{y} \in B\}$ である．
(2) 可測空間 $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ 上の測度 μ が

$$\mu(\boldsymbol{x} + B) = \mu(B) \qquad \bigl(\forall \boldsymbol{x} \in \mathbb{R}^n,\ \forall B \in \mathcal{B}(\mathbb{R}^n)\bigr)$$

をみたすならば

$$\mu = \kappa\lambda_n, \qquad \kappa = \mu\bigl([0,\,1)^n\bigr) < \infty$$

と書ける．ただし $(0,\,1]^n = \underbrace{(0,\,1] \times (0,\,1] \times \cdots (0,\,1]}_{n\text{ 個の }(0,\,1]}$ である．

Proof. まず

$$B \in \mathcal{B}\bigl(\mathbb{R}^n\bigr) \Rightarrow \boldsymbol{x} + B \in \mathcal{B}\bigl(\mathbb{R}^n\bigr) \tag{C.9}$$

を確認する．ただし $\boldsymbol{x} + B = \{\boldsymbol{x} + \boldsymbol{b};\ \boldsymbol{b} \in B\}$ である．そのために

$$\mathcal{A}_{\boldsymbol{x}} := \bigl\{B \in \mathcal{B}\bigl(\mathbb{R}^n\bigr);\ \boldsymbol{x} + B \in \mathcal{B}\bigl(\mathbb{R}^n\bigr)\bigr\}$$

と定める. 明らかに $\mathcal{A}_{\boldsymbol{x}}$ は σ 代数で

$$\begin{aligned}\mathcal{D} &= \bigl\{[a_1,\,b_1)\times[a_2,\,b_2)\times\cdots\times[a_n,\,b_n) :\, a_i,\,b_i\in\mathbb{R}\,(i=1,\,2,\,\ldots,\,n)\bigr\}\\ &\subset \mathcal{A}_{\boldsymbol{x}}\end{aligned}$$

となる. よって

$$\mathcal{B}\bigl(\mathbb{R}^n\bigr)=\sigma[\mathcal{D}]=\mathcal{A}_{\boldsymbol{x}}\subset\mathcal{B}\bigl(\mathbb{R}^n\bigr)$$

となり, (C.9) は示せた.

(1) $\boldsymbol{x}=(x_1,\,x_2,\,\ldots,\,x_n)^\top\in\mathbb{R}^n$ を固定し

$$\nu(B):=\lambda_n(\boldsymbol{x}+B)\qquad\bigl(\forall B\in\mathcal{B}\bigl(\mathbb{R}^n\bigr)\bigr)$$

と定める. いま

$$D=[a_1,\,b_1)\times[a_2,\,b_2)\times\cdots\times[a_n,\,b_n)\in\mathcal{D}\qquad\bigl(a_i,\,b_i\in\mathcal{R};\,i=1,\,2,\,\ldots,\,n\bigr)$$

を取る. ある $i\in\{1,\,2,\ldots,\,n\}$ に対して $a_i\geq b_i$ の場合は $D=\varnothing$ と約束する. すると

$$\boldsymbol{x}+D=[a_1+x_1,\,b_1+x_1)\times\cdots\times[a_n+x_1,\,b_n+x_n)\in\mathcal{D}$$

となる. したがって

$$\begin{aligned}\nu(D)&=\lambda_n(\boldsymbol{x}+D)=\prod_{i=1}^n\bigl\{(b_i+x_i)-(a_i+x_i)\bigr\}=\prod_{i=1}^n(b_i-a_i)\\&=\lambda_n(D)\end{aligned}$$

となる. よって

$$\nu(D)=\mu(D)\qquad(\forall D\in\mathcal{D})$$

となる. $a_i,\,b_i,\,c_i,\,d_i\in\mathcal{R}\,(i=1,\,2,\ldots,\,n)$ に対して

$$\begin{aligned}&\Bigl([a_1,\,b_1)\times\cdots\times[a_n,\,b_n)\Bigr)\bigcap\Bigl([c_1,\,d_1)\times\cdots\times[c_n,\,d_n)\Bigr)\\&\quad=[a_1\vee c_1,\,b_1\wedge d_1)\times\cdots\times[a_n\vee c_n,\,b_n\wedge d_n)\end{aligned}$$

となる. よって $\mathcal{D}$ は共通部分を取る演算に関して閉じているので定理 C.36 から

$$\nu(B)=\lambda_n(B)\qquad\bigl(\forall B\in\mathcal{B}\bigl(\mathbb{R}^n\bigr)=\sigma[\mathcal{D}]\bigr)$$

となる.

(2) $D = [a_1, b_1) \times [a_2, b_2) \times \cdots \times [a_n, b_n)$ で $a_i, b_i \in \mathbb{Q}\,(i = 1, 2, \ldots,)$ とする．するとある $M \in \mathbb{N}$, $k(D) \in \mathbb{N}$ と $\boldsymbol{x}^{(i)} \in \mathbb{R}^n\,(i = 1, 2, \ldots, n)$ があって

$$D = \bigsqcup_{i=1}^{k(D)} \left\{ \boldsymbol{x}^{(i)} + \left[0,\, \frac{1}{M}\right] \times \cdots \times \left[0,\, \frac{1}{M}\right] \right\}$$

と書ける．だだし $\bigsqcup$ は直和を表す．μ と λ_n は移動不変なので

$$\begin{aligned}
\mu(D) &= k(D)\mu\left(\left[0,\, \frac{1}{M}\right] \times \cdots \times \left[0,\, \frac{1}{M}\right]\right), \\
\lambda_n(D) &= k(D)\lambda_n\left(\left[0,\, \frac{1}{M}\right] \times \cdots \times \left[0,\, \frac{1}{M}\right]\right), \\
\mu\bigl([0,\, 1) \times \cdots \times [0,\, 1)\bigr) &= M^n\mu\left(\left[0,\, \frac{1}{M}\right] \times \cdots \times \left[0,\, \frac{1}{M}\right]\right) \\
\underbrace{\lambda_n\bigl([0,\, 1) \times \cdots \times [0,\, 1)\bigr)}_{=1} &= M^n\mu\left(\left[0,\, \frac{1}{M}\right] \times \cdots \times \left[0,\, \frac{1}{M}\right]\right)
\end{aligned}$$

となる．よって

$$\begin{aligned}
\mu(D) &= k(D)\mu\left(\left[0,\, \frac{1}{M}\right] \times \cdots \times \left[0,\, \frac{1}{M}\right]\right) \\
&= k(D)M^n\mu\left(\left[0,\, \frac{1}{M}\right] \times \cdots \times \left[0,\, \frac{1}{M}\right]\right), \\
\lambda_n(D) &= k(D)\lambda_n\left(\left[0,\, \frac{1}{M}\right] \times \cdots \times \left[0,\, \frac{1}{M}\right]\right) = k(D)M^n
\end{aligned}$$

なので

$$\mu(D) = \mu\bigl([0,\, 1) \times \cdots \times [0,\, 1)\bigr)\lambda_n(D)$$

を得る． □

定理 C.38. $A \in \mathsf{M}(n,\, \mathbb{R})$ は正則行列とする．このとき

$$A(\lambda_n) = |\det A|^{-1}\lambda_n$$

となる．ただし λ_n は $\mathbb{R}^n$ 上の Lebeague 測度で

$$A(\lambda_n)(B) = \lambda_n\bigl(A^{-1}B\bigr) := \lambda_n\Bigl(\bigl\{A^{-1}\boldsymbol{x};\, \boldsymbol{x} \in B\bigr\}\Bigr) \qquad (B \in \mathcal{B}(\mathbb{R}^n)$$

である．

Proof. $\forall B \in \mathcal{B}(\mathbb{R}^n)$ に対して

$$\mu(B) := \lambda_n\bigl(A^{-1}B\bigr)$$

とおく．すると $\forall \boldsymbol{x} \in \mathbb{R}^n$ に対して

$$\begin{aligned}\mu(\boldsymbol{x}+B) &= \lambda_n\bigl(A^{-1}(\boldsymbol{x}+B)\bigr) = \lambda_n\bigl(A^{-1}B\bigr) \qquad \bigl(\because \text{定理 C.37(i)}\bigr)\\ &= \mu(B)\end{aligned}$$

となる．よって定理 C.37(ii) から

$$\mu(B) = \mu\bigl([0,\,1)\times\cdots\times[0,\,1)\bigr)\lambda_n(B) = \mu\bigl(A^{-1}\bigl([0,\,1)\times\cdots\times[0,\,1)\bigr)\bigr)\lambda_n(B)$$

となる．$A^{-1}\bigl([0,\,1)\times\cdots\times[0,\,1)\bigr)$ は平行四辺形となり

$$\mathsf{vol}^{(n)} = |\det A^{-1}| = |\det A|^{-1}$$

となるので

$$\mu(B) = |\det A|^{-1}\lambda_n(B)$$

がわかる． □

C.5　積分の定義

定義 C.39. (1) $d \in \mathbb{N}$) とし，$A \subset \mathcal{B}\bigl(\mathbb{R}^d\bigr)$ とする．集合 A の**特性関数** $\mathbb{1}_A(\boldsymbol{x}) : \mathbb{R}^d \to \{0,\,1\}$ を

$$\mathbb{1}_A(\boldsymbol{x}) = \begin{cases} 1 & (\boldsymbol{x} \in A)\\ 0 & (\boldsymbol{x} \notin A)\end{cases}$$

で定める．
(2) 関数 $f:\ \mathbb{R}^d \to \mathbb{R}\cup\{\pm\infty\}$ が Borel 可測であるとは各 $r \in \mathbb{R}$ に対して

$$\bigl\{\boldsymbol{x} \in \mathbb{R}^d;\ f(\boldsymbol{x}) > r\bigr\} = f^{-1}\bigl((r,\,\infty)\bigr) \in \mathcal{B}\bigl(\mathbb{R}^d\bigr)$$

となることである．
(3) $N \in \mathbb{N}$ とする．有限個の $a_1,\,a_2,\,\ldots,\,a_N \in \mathbb{R}$ とそれらと同数の Borel 可測集合の有限個の列 $\{A_n\}_{n=1}^N \subset \mathcal{B}\bigl(\mathbb{R}^d\bigr)$ を用いて

$$f(\boldsymbol{x}) = \sum_{n=1}^{N} a_n \mathbb{1}_{A_n}(\boldsymbol{x})$$

と表せる関数 $f : \mathbb{R}^d \to \mathbb{R}$ を Borel 単関数 という．

定理 C.40. 連続関数は Borel 可測である.

Proof. $\forall r \in \mathbb{R}$ に対して (r, ∞) は $\mathbb{R}$ の開集合である. f は連続であるので, $f^{-1}\big((r, \infty)\big)$ は $\mathbb{R}^d$ の開集合となるので,

$$f^{-1}\big((r, \infty)\big) \in \mathcal{B}\big(\mathbb{R}^d\big)$$

がわかる. □

補題 C.41. Borel 集合 $E \subset \mathbb{R}^d$ から $\mathbb{R} \cup \{\pm\}$ への関数について次の性質が成り立つ.

(1) f が Borel 可測のとき $\forall r \in \mathbb{R}$ に対して

$$\{f < r\} = f^{-1}\big(-\infty, r)\big) \subset \mathcal{B}\big(\mathbb{R}^d\big).$$

(2) f, g を Borel 可測とするとき和 $f+g$, 積 $f \cdot g$ は Borel 可測である. ただし $f(x) + g(x) = \infty + (-\infty)$ または $(-\infty) + \infty$, $f(x)g(x) = 0 \cdot (\pm\infty)$ または $f(x)g(x) = \pm\infty \cdot 0$ の場合を除く.

(3) $\{f_n\}_{n=1}^{\infty}$ を Borel 可測関数列とするとき

$$\sup_{n \in \mathbb{N}} f_n,\ \inf_{n \in \mathbb{N}} f_n,\ \limsup_{n \to \infty} f_n,\ \liminf_{n \to \infty} f_n$$

も Borel 可測となる. さらに $\lim_{n\to\infty} f_n$ が存在すれば $\lim_{n\to\infty} f_n$ も Borel 可測である.

Proof. (1)

$$(-\infty, r) = \bigcup_{n=1}^{\infty} \left(-\infty, r - \frac{1}{n}\right]$$

を用いると

$$\{f < r\} = \bigcup_{n=1}^{\infty} \left\{f \leq r - \frac{1}{n}\right] = \bigcup_{n=1}^{\infty} \left\{f > r - \frac{1}{n}\right\} \in \mathcal{B}\big(\mathbb{R}^d\big)$$

よりわかる.
(2) $\forall r \in \mathbb{R}$ とすると

$$\big\{f + g > r\big\} = \bigcup_{q \in \mathbb{Q}} \big\{f > q\big\} \cup \big\{g > r - q\big\}$$

となる. よって $f+g$ も Borel 可測である. また

$$\big\{f^2 > r\big\} = \begin{cases} \big\{g > \sqrt{r}\big\} \cup \big\{g < -\sqrt{r}\big\} & (r \geq 0), \\ \mathbb{R} & (r < 0) \end{cases} \in \mathcal{B}\big(\mathbb{R}^d\big)$$

となるので f^2 も Borel 可測となる．このことから

$$f \cdot g = \frac{1}{2}(f+g)^2 - \frac{1}{2}f^2 - \frac{1}{2}g^2$$

と書けるので $f \cdot g$ も Borel 可測となる．
(3) $\forall r \in \mathbb{R}$ に対して

$$\begin{aligned}
\left\{\sup_{n\in\mathbb{N}} f_n > r\right\} &= \bigcup_{n=1}^{\infty}\{f_n > r\}, \\
\left\{\inf_{n\in\mathbb{N}} f_n > r\right\} &= \bigcup_{m=1}^{\infty}\bigcap_{n=1}^{\infty}\left\{f_n > r + \frac{1}{m}\right\}, \\
\limsup_{n\to\infty} &= \inf_{n\in\mathbb{N}}\left(\sup_{k\ge n} f_k\right), \\
\liminf_{n\to\infty} f_n &= \sup_{n\in\mathbb{N}}\left(\inf_{k\ge n} f_k\right)
\end{aligned}$$

よりわかる． □

定理 C.42. $\mathbb{R}^d$ で定義される正値 Borel 可測函数は正値 Borel 単関数 $\{f_n\}_{n=1}^{\infty}$ の単調増大列の極限として表される．

Proof. $f : \mathbb{R}^d \to [0, \infty) \cup \{+\infty\}$ に対して単関数 f_n を

$$f_n(\boldsymbol{x}) = \begin{cases} \dfrac{j-1}{2^n} & \left(\dfrac{j-1}{2^n} \le f(\boldsymbol{x}) < \dfrac{j}{2^n}\ \big(j = 1, 2, \ldots, n2^n\big)\right), \\ n & (f(\boldsymbol{x}) \ge n) \end{cases}$$

で定義すると f_n は Borel 可測となる．また

$$\begin{aligned}
&0 \le f_1(\boldsymbol{x}) \le f_2(\boldsymbol{x}) \le \cdots \le f(\boldsymbol{x}), \\
&\lim_{n\to\infty} f_n(\boldsymbol{x}) = f(\boldsymbol{x})
\end{aligned}$$

となる． □

注意 C.43. $\displaystyle\sup_{\boldsymbol{x}\in\mathbb{R}^d} f(\boldsymbol{x}) < \infty$ のとき

$$\lim_{n\to\infty}\sup_{\boldsymbol{x}\in\mathbb{R}^d}\big|f(\boldsymbol{x}) - f_n(\boldsymbol{x})\big| = 0$$

となる． □

定義 C.44. $E \subset \mathbb{R}^d$ は Borel 可測集合とする．$f\colon E \to [0, \infty)$ を正値 Borel 単関数とする．$N \in \mathbb{N}$ とし，$a_1, a_2, \ldots, a_N \in (0, \infty)$ と互いに交わらない $A_1, A_2, \ldots, A_n \in \mathcal{B}(\mathbb{R}^d)$ を用いた表現として

$$f = \sum_{j=1}^{N} a_j \mathbb{1}_{A_j}$$

と表されるとき Borel 単関数 f の E 上の積分を

$$\int_E f(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) := \sum_{j=1}^{N} a_j \mu\bigl(E \cap A_j\bigr)$$

と定める．

注意 C.45. 正値単関数 $f\colon E \to [0, \infty)$ の積分 $\int_E f(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x})$ は f の表現のとり方によらないことがわかる． □

定義 C.46. E を $\mathbb{R}^d$ の Borel 可測集合とする．$f\colon E \to [0, \infty) \cup \{+\infty\}$ が正値 Borel 可測関数のとき f の積分 $\int_E f\, d\mu(\boldsymbol{x})$ を

$$\int_E f(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) = \sup\left\{\int_E g(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x});\ g \text{ は } 0 \le g \le f \text{ なる単関数}\right\}$$

と定める．

注意 C.47. 上記の定義において $\int_E f(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) = 0,\ \infty$ の可能性も許している． □

命題 C.48. $g\colon E \to [0, \infty)$ を非負値単関数とする．$\{f_n\}_{n=1}^{\infty}$ が $\forall \boldsymbol{x} \in E$ に対して

$$\lim_{n\to\infty} f_n(\boldsymbol{x}) \ge g(\boldsymbol{x})$$

をみたす正値 Borel 関数の増大列のとき

$$\int_E g(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) = \lim_{n\to\infty} \int_E \min\{f_n(\boldsymbol{x}),\ g(\boldsymbol{x})\}\,\mathrm{d}\mu(\boldsymbol{x})$$

が成り立つ．

Proof. 単関数の積分について加法性が成り立つので，一般性を失わず

$$g(\boldsymbol{x}) = a\mathbb{1}_A(\boldsymbol{x}), \quad a > 0,\ A \in \mathcal{B}\bigl(\mathbb{R}^d\bigr)$$

と仮定して

$$a\mu(A) = \lim_{n\to\infty} \int_E \min\bigl\{a\mathbb{1}_A(\boldsymbol{x}),\ f_n(\boldsymbol{x})\bigr\}\,\mathrm{d}\mu(\boldsymbol{x}) \tag{C.10}$$

を示せばよい．$\epsilon > 0$ を固定する．$n = 1, 2, \ldots$ として

$$A_n := \{f_n > a - \epsilon\} \cap A$$

とおくと $\lim_{n\to\infty} f_n(\boldsymbol{x}) \geq a\mathbb{1}_A(\boldsymbol{x})\,(\boldsymbol{x} \in \mathbb{R}^d)$ から

$$A_1 \subset A_2 \subset \cdots \quad \text{かつ} \quad A = \bigcup_{n=1}^{\infty} A_n$$

となる．すると

$$a\mathbb{1}_A(\boldsymbol{x}) \geq \min\bigl\{a\mathbb{1}_A(\boldsymbol{x}),\, f_n(\boldsymbol{x})\bigr\}$$

から

$$\begin{aligned}
a\mu(A) &\geq \lim_{n\to\infty} \int_E \min\bigl\{a\mathbb{1}_A(\boldsymbol{x}),\, f_n(\boldsymbol{x})\bigr\}\,\mathrm{d}\mu(\boldsymbol{x}) \\
&\geq \lim_{n\to\infty} \int_E (a-\epsilon)\mathbb{1}_{A_n}(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) \\
&= (a-\epsilon) \lim_{n\to\infty} \mu(A_n) \\
&= (a-\epsilon)\mu\biggl(\bigcup_{n=1}^{\infty} A_n\biggr) \\
&= (a-\epsilon)\mu(A)
\end{aligned}$$

となる．ϵ は任意だったので

$$a\mu(A) = \lim_{n\to\infty} \int_E \min\bigl\{a\mathbb{1}_A(\boldsymbol{x}),\, f_n(\boldsymbol{x})\bigr\}\, d\mu(\boldsymbol{x})$$

を得る． □

定理 C.49. $0 \leq f_n \leq f_{n+1}\,(n \in \mathbb{N})$ をみたす Borel 可測集合 $E \subset \mathbb{R}^d$ 上で定義された Borel 可測関数列 $\{f_n\}_{n=1}^{\infty}$ が与えられているとする．関数 $f\colon E \to \mathbb{R}$ を各点 $\boldsymbol{x} \in E$ に対して $\lim_{n\to\infty} f_n(\boldsymbol{x}) = f(\boldsymbol{x})$ で定めると

$$\lim_{n\to\infty} \int_E f_n(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) = \int_E f(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x})$$

が成り立つ．

Proof. f_n の単調増加性より

$$\lim_{n\to\infty} \int_E f_n(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) \leq \int_E f(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x})$$

は明らかである. 逆向きの不等式を示すために単関数 g を $0 \leq g \leq f$ となるようにとる. 命題 C.48 から

$$\int_E g(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) = \lim_{n\to\infty}\int_E \min\bigl\{g(\boldsymbol{x}),\, f_n(\boldsymbol{x})\bigr\}\,\mathrm{d}\mu(\boldsymbol{x}) \leq \lim_{n\to\infty}\int_E f_n(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x})$$

となる. さらに積分の定義から

$$\int_E f(\boldsymbol{x})\,d\mu(\boldsymbol{x}) = \sup_{0\leq g\leq f}\int_E g(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) \leq \lim_{n\to\infty}\int_E f_n(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x})$$

を得る. よって定理は示された. □

命題 C.50. f, g は Borel 集合 $E \subset \mathbb{R}^d$ 上で定義された正値 Borel 可測関数ならば

$$\int_E \bigl\{f(\boldsymbol{x}) + g(\boldsymbol{x})\bigr\}\,\mathrm{d}\mu(\boldsymbol{x}) = \int_E f(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) + \int_E g(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x})$$

となる.

Proof. 単関数の増大列 $\{f_n\}_{n=1}^\infty$, $\{g_n\}_{n=1}^\infty$ をとり $\lim\limits_{n\to\infty} f_n(\boldsymbol{x}) = f(\boldsymbol{x})$, $\lim\limits_{n\to\infty} g_n(\boldsymbol{x}) = g(\boldsymbol{x})$ とする. 定理 C.49 より

$$\begin{aligned}
\int_E \bigl\{f(\boldsymbol{x}) + g(\boldsymbol{x})\bigr\}\,\mathrm{d}\mu(\boldsymbol{x}) &= \lim_{n\to\infty}\int_E \bigl\{f_n(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) + \int_E g_n(\boldsymbol{x})\bigr\}\,\mathrm{d}\mu(\boldsymbol{x}) \\
&= \lim_{n\to\infty}\left\{\int_E f_n(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) + \int_E g_n(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x})\right\} \\
&= \lim_{n\to\infty}\int_E f_n(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) + \lim_{n\to\infty}\int_E g_n(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) \\
&= \int_E f(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) + \int_E g(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x})
\end{aligned}$$

よりわかる. □

定理 C.51. (Fatou の補題) Borel 集合 $E \subset \mathbb{R}^d$ 上で定義さらた正値 Borel 可測関数列 $\{f_n\}_{n=1}^\infty$ に対して

$$\int_E \liminf_{n\to\infty} f_n(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) \leq \liminf_{n\to\infty}\int_e f_n(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x})$$

が成り立つ.

Proof. $\left\{\inf\limits_{\ell\in\mathbb{N}\cap[k,\infty)} f_\ell\right\}_{k=1}^\infty$ は単調増加列であることに注意して定理 C.49

を用いると

$$\begin{aligned}\int_E \liminf_{\ell\to\infty} f_\ell(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) &= \int_E \lim_{k\to\infty}\left\{\inf_{\ell\in\mathbb{N}\cap[k,\,\infty)} f_\ell(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x})\right\}\\ &= \lim_{k\to\infty}\int_E\left\{\inf_{\ell\in[k,\,\infty)} f_\ell(\boldsymbol{x})\right\}\mathrm{d}\mu(\boldsymbol{x})\\ &= \liminf_{k\to\infty}\int_E\left\{\inf_{\ell\in[k,\,\infty)} f_\ell(\boldsymbol{x})\right\}\mathrm{d}\mu(\boldsymbol{x})\\ &= \liminf_{k\to\infty}\int_E f_k(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x})\end{aligned}$$

となることから定理は証明された. □

定義 C.52. $E\subset\mathbb{R}^d$ を Borel 可測集合とする. $f\colon E\to\mathbb{R}\cup\{\pm\infty\}$ に対して

$$f^+ := \max\{f,\,0\},\quad f^- := \max\{-f,\,0\}$$

とおく. f が可積分であるとは

$$\int_E f^+(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x})<\infty,\quad \int_E f^{-1}(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x})<\infty$$

のときをいう. さらにこのとき

$$\int_E f(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) = \int_E f^+(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) + \int_E f^{-1}(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x})$$

と定める.

注意 C.53. Borel 可測関数 $f\colon E\to\mathbb{R}\cup\{\pm\infty\}$ が可積分でることと $|f|$ が可積分であることは同値である.

定理 C.54. (Lebeague の収束定理) $E\subset\mathbb{R}^d$ を Borel 可測集合とする. $f_n\colon E\to\mathbb{R}\cup\{\pm\infty\}\,(n=1,\,2,\,\ldots)$ を関数 f に各点で収束する Borel 可測関数とする. 非負値可積分関数 $g\colon E\to[0,\,\infty)\cup\{+\infty\}$ が存在し, $|f_n(\boldsymbol{x})|\le g(\boldsymbol{x})\,(\forall\boldsymbol{x}\in E,\,\forall n\in\mathbb{N})$ をみたしていれば

$$\lim_{n\to\infty}\int_E f_n(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) = \int_E f(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x})$$

となる.

Proof. $g\pm f_n\ge 0$ であるから, Fatou の補題から

$$\begin{aligned}\int_E\{g(\boldsymbol{x})+f(\boldsymbol{x})\}\,\mathrm{d}\mu(\boldsymbol{x}) &\le \int_E g(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) + \liminf_{n\to\infty}\int_E f_n(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x})\\ \int_E\{g(\boldsymbol{x})-f(\boldsymbol{x})\}\,\mathrm{d}\mu(\boldsymbol{x}) &\le \int_E g(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) + \liminf_{n\to\infty}\int_E\{-f_n(\boldsymbol{x})\}\,\mathrm{d}\mu(\boldsymbol{x})\\ &= \int_E g(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) - \limsup_{n\to\infty}\int_E f_n(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x})\end{aligned}$$

となる．各項は有限であるので

$$\begin{aligned}\limsup_{n\to\infty}\int_E f_n(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) \le \int_E f(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) \le \liminf_{n\to\infty}\int_E f_n(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x})\\ \le \limsup_{n\to\infty}\int_E f_n(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x})\end{aligned}$$

となり, 定理は証明された. □

系 C.55. Borel 可測集合 $E \subset \mathbb{R}^d$ 上の可積分 Borel 可測関数列 $\{f_n\}_{n=1}^{\infty}$ が $\displaystyle\sum_{n=1}^{\infty}\int_E |f_n(\boldsymbol{x})|\,\mathrm{d}\mu(\boldsymbol{x}) < \infty$ をみたすとする．

$$F = \left\{x \in E;\ \textstyle\sum_{n=1}^{\infty}|f_n(\boldsymbol{x})| \text{ は収束}\right\}$$

とおいたとき

$$\mu\bigl(E \setminus F\bigr) = 0$$

で

$$\sum_{n=1}^{\infty}\int_E f_n(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) = \int_F \sum_{n=1}^{\infty} f_n(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x})$$

が成り立つ．

Proof.

$$t_n(\boldsymbol{x}) := \sum_{\ell=1}^{n}|f_\ell(\boldsymbol{x})|$$

とおく．単調収束定理から

$$\begin{aligned}\sum_{n=1}^{\infty}\int_E |f_n(\boldsymbol{x})|\,\mathrm{d}\mu(\boldsymbol{x}) &= \lim_{n\to\infty}\sum_{\ell=1}^{n}\int_E |f_n(\boldsymbol{x})|\,\mathrm{d}\mu(\boldsymbol{x})\\ &= \lim_{n\to\infty}\int_E \sum_{\ell=1}^{n}|f_n(\boldsymbol{x})|\,\mathrm{d}\mu(\boldsymbol{x})\\ &= \lim_{n\to\infty}\int_E t_n(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x})\\ &= \int_E \lim_{n\to\infty} t_n(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x})\\ &= \int_E \sum_{n=1}^{\infty}|f_n(\boldsymbol{x})|\,\mathrm{d}\mu(\boldsymbol{x})\end{aligned}$$

となるので, $\displaystyle\sum_{n=1}^{\infty}|f_n(\boldsymbol{x})|$ は可積分である．あとは $\displaystyle\sum_{n=1}^{\infty}f_n(\boldsymbol{x})$ に Lebeague の収束定理を適用すればよい． □

定理 C.56. $E \subset \mathbb{R}^d$ を Borel 可測集合とする．$a, b \in \mathbb{R}(a < b)$ とし，関数 $f : (a, b) \times E \to \mathbb{R} \cup \{\pm\infty\}$ は次の条件をみたしているとする．

(1) $\boldsymbol{x} \in \mathbb{R}^d$ を止めるごとに $f(\cdot, \boldsymbol{x})$ は連続関数となる．

(2) 非負値可積分関数 $g : E \to [0, \infty) \cup \{+\infty\}$ が存在して $|f(t \cdot)| \leq g$ となる．

このとき $t_0 \in (a, b)$ に対して

$$\lim_{t \to t_0} \int_E f(t, \boldsymbol{x}) \, \mathrm{d}\mu(\boldsymbol{x}) = \int_E f(t_0, \boldsymbol{x}) \, \mathrm{d}\mu(\boldsymbol{x})$$

が成り立つ．

Proof. $\{t_n\}_{n=1}^{\infty} \subset (a, b)$ で $\lim_{n \to \infty} t_n = t_0$ なる点列を取る．Lebeague の収束定理から

$$\lim_{n \to \infty} \int_E f(t_n, \boldsymbol{x}) \, \mathrm{d}\mu(\boldsymbol{x}) = \int_E \lim_{n \to \infty} f(t_n, \boldsymbol{x}) \, \mathrm{d}\mu(\boldsymbol{x}) = \int_E f(t_0, \boldsymbol{x}) \, \mathrm{d}\mu(\boldsymbol{x})$$

よりわかる． □

定理 C.57. $E \subset \mathbb{R}^d$ を Borel 可測集合とする．関数 $f : (a, b) \times E \to \mathbb{R}$ は次の条件をみたすとする．

(1) $\forall t \in (a, b)$ と $\forall \boldsymbol{x} \in \mathbb{R}^d$ に対して $\dfrac{\partial}{\partial t} f(t, \boldsymbol{x})$ は存在する．

(2) $f(t, \cdot)$ は可積分である．

(3) 非負値可積分 $g : E \to (0, \infty)$ が存在して

$$\left| \frac{\partial}{\partial t} f(t, \boldsymbol{x}) \right| \leq g(\boldsymbol{x}) \qquad (\boldsymbol{x} \in E)$$

となる．

このとき

$$\frac{d}{dt} \int_E f(t, \boldsymbol{x}) \, \mathrm{d}\mu(\boldsymbol{x}) = \int_E \frac{\partial}{\partial t} f(t, \boldsymbol{x}) \, \mathrm{d}\mu(\boldsymbol{x})$$

が成り立つ．

Proof. $t_0 \in (a, b)$ を固定する．$t \in (a, b)$ と $\boldsymbol{x} \in \mathbb{R}^d$ に対して関数 $G(t, \boldsymbol{x})$ を

$$G(t, \boldsymbol{x}) = \begin{cases} \dfrac{f(t, \boldsymbol{x}) - f(t_0, \boldsymbol{x})}{t - t_0} & (t \neq t_0) \\ \dfrac{\partial}{\partial t} f(t_0, \boldsymbol{x}) & (t = t_0) \end{cases}$$

と定める．定理 C.56 を関数 G に適用すればよい． □

C.6　Fubini の定理

命題 C.58. $(\mathbb{X}, \mathcal{T})$ と $(\mathbb{Y}, \mathcal{U})$ 位相空間とする．任意の移送空間 $(\mathbb{Z}, \mathcal{V})$ に対して $\mathbb{B}(\mathbb{Z}, \mathcal{V})$ を $\mathcal{V}$ の位相によって生成される $\mathbb{Z}$ の Borel 集合族とする．このとき $\mathbb{X} \times \mathbb{Y}$ の積位相によって生成される $\mathbb{X} \times \mathbb{Y}$ 上の集合族 $\mathcal{C}$ は Borel 直積 σ 集合族 $\mathcal{B}(\mathbb{X}, \mathcal{T})$ と $\mathcal{B}(\mathbb{Y}, \mathcal{U})$ を含む．$(\mathbb{X}, \mathcal{T})$ と $(\mathbb{Y}, \mathcal{U})$ がともに第 2 可算公理をみたすとき

$$\mathcal{C} = \mathcal{B}(\mathbb{X}, \mathcal{T}) \otimes \mathcal{B}(\mathbb{Y}, \mathcal{U})$$

となる．

Proof. 証明はどこかをみること． □

$(\mathbb{X}, \mathcal{A}, \mu)$ と $(\mathbb{Y}, \mathcal{B}, \nu)$ を測度空間とし，

$$\mathcal{D} := \{A \times B;\ A \in \mathcal{A},\ B \in \mathcal{B}\} \tag{C.11}$$

とする．すると命題 C.19 より $\mathcal{D}$ は半環となる．$\forall A \times B \in \mathcal{R}$ に対して

$$\rho(A \times B) := \mu(A) \times \nu(B) \tag{C.12}$$

と定める．ただし $0 \cdot \infty = \infty \cdot 0 = 0$ と約束する．

定理 C.59. (C.12) で定義された ρ は $\mathcal{D}$ 上で σ 加法的となる．

Proof. $n \in \mathbb{N}$ とし，$B_n \in \mathcal{B}$, $C_n \in \mathcal{C}$ とする．$\{B_n \times C_n\}_{n=1}^{\infty}$ の任意の有限部分集合は互いに排反とし

$$B \times C := \bigcup_{n=1}^{\infty} (B_n \times C_n)$$

とおく．すると $x \in \mathbb{X}$ と $y \in \mathbb{Y}$ に対して

$$\mathbb{1}_B(x)\mathbb{1}_C(y) = \sum_{n=1}^{\infty} \mathbb{1}_{B_n}(x)\mathbb{1}_{C_n}(y)$$

と書ける．各 x に対して，ν で積分をすると

$$\begin{aligned}
\mathbb{1}_B(x)\nu(C) &= \int \mathbb{1}_B(x)\mathbb{1}_C(y)\,\mathrm{d}\nu(y) = \int \left(\sum_{n=1}^{\infty} \mathbb{1}_{B_n}(x)\mathbb{1}_{C_n}(y)\right) \mathrm{d}\nu(y) \\
&= \sum_{n=1}^{\infty} \mathbb{1}_{B_n}(x) \int \mathbb{1}_{C_n}(y)\,\mathrm{d}\nu(y) \qquad (\because \text{単調収束定理 (定理 C.49)}) \\
&= \sum_{n=1}^{\infty} \mathbb{1}_{B_n}(x)\nu(C_n)
\end{aligned}$$

となる．さらに μ で積分すると

$$
\begin{aligned}
\mu(B)\nu(C) &= \int \mathbb{1}_B(x)\nu(C)\,\mathrm{d}\mu(x) = \int \sum_{n=1}^{\infty} \mathbb{1}_{B_n}(x)\nu(C_n)\,\mathrm{d}\mu(x) \\
&= \sum_{n=1}^{\infty} \nu(C_n) \int \mathbb{1}_{B_n}(x)\,\mathrm{d}\mu(x) \qquad \text{(∵ 単調収束定理 (定理 C.49))} \\
&= \sum_{n=1}^{\infty} \mu(B_n)\nu(C_n)
\end{aligned}
$$

を得る．よって ρ は $\mathcal{D}$ 上で σ 加法的である．□

$\mathcal{R}$ を (C.11) で定義さらた半環 $\mathcal{D}$ によって生成される環[3]とする．注意 C.22 から $\mathcal{R}$ は $\mathcal{D}$ の有限個の排反な元の和が成す部分集合族である．$\mathbb{X} \times \mathbb{Y} \in \mathcal{R}$ なので注意 C.25(1) から σ 代数となる．

注意 C.60. (1) $n \in \mathbb{N}$ とする．排反な $R_1, R_2, \ldots, R_n \in \mathcal{R}$ に対して

$$
\rho\Bigl(\bigcup_{j=1}^{n} R_j\Bigr) = \sum_{j=1}^{n} \rho(R_j)
$$

となるので命題 C.23 と定理 C.59 から ρ は well-defined で $\mathcal{R}$ 上で σ 加法的である．
(2) すると ρ は直積 σ 集合族 $\mathcal{B} \otimes \mathcal{C}$ 上の σ 加法的測度に拡張できる．しかし拡張が一意的であるとは一般的に限らない．次のステップで拡張が一意的であるための十分条件を求めよう．□

定義 C.61. 集合族 $\mathcal{M}$ は**単調族**であるとは，$M_n \in \mathcal{M}\,(n = 1, 2, \ldots)$ かつ $M_1 \supset M_2 \supset \cdots$ で $\bigcup_{n=1}^{\infty} M_n =: M$ または $M_1 \subset M_2 \subset \cdots$ で $\bigcup_{n=1}^{\infty} M_n =: M$ に対して $M \in \mathcal{M}$ のときをいう．

注意 C.62. 任意の σ 代数は単調族であるが，位相は一般にはそうではない．なぜならば無限個の開集合の共通部分は開集合でないことからわかる．またベキ集合は単調俗である．したがって任意の集合族 $\mathcal{D}$ を含む最小の単調族が存在する．□

定理 C.63. $\mathbb{X}$ を空でない集合とする．$\mathcal{A}$ を $\mathbb{X}$ の部分集合の代数とする．このとき $\mathcal{A}$ を含む最小の単調族 $\mathcal{M}$ は σ 代数である．

Proof. $\mathcal{N} := \{E \in \mathcal{M}; \mathbb{X} \setminus E \in \mathcal{M}\}$ とおく．このとき $\mathcal{M}$ は単調族なので，$\mathcal{N}$ も単調族で $\mathcal{A} \subset \mathcal{N}$ となる．よって $\mathcal{M} \subset \mathcal{N}$ となる．各 $A \subset \mathbb{X}$ に

[3] 作り方から $\mathcal{R}$ は σ 環になっていることに注意せよ．

対して

$$\mathcal{M}_A := \{E :\ E \cap A \in \mathcal{M}\}$$

とおく．このとき $A \in \mathcal{A}$ に対して $\mathcal{A} \subset \mathcal{M}_A$ であり, $\mathcal{M}_A$ も単調族なので, $\mathcal{M} \subset \mathcal{M}_A$ となる．さらに $E \in \mathcal{M}$ に対して

$$\mathcal{M}_E := \{F;\ F \cap E \in \mathcal{M}\}$$

とおくと $\mathcal{A} \subset \mathcal{M}_E$ なので $\mathcal{M} \subset \mathcal{M}_E$ となる．したがって $\mathcal{M}$ は代数となる．さらに $\mathcal{M}$ は単調族なので $\mathcal{M}$ は σ 代数となる． □

定理 C.64. $(\mathbb{X}, \mathcal{A}, \mu)$ と $(\mathbb{Y}, \mathcal{B}, \nu)$ を測度空間とし, $\mu(\mathbb{X}) < \infty$ かつ $\nu(\mathbb{Y}) < \infty$ とする．

$$\mathcal{F} := \left\{E \subset \mathbb{X} \times \mathbb{Y};\right.$$
$$\left.\int \left[\int \mathbb{1}_E(x, y)\,\mathrm{d}\mu(x)\right] \mathrm{d}\nu(y) = \int \left[\int \mathbb{1}_E(x, y)\,\mathrm{d}\nu(y)\right] \mathrm{d}\mu(x)\right\}$$

とおく．このとき

$$\mathcal{A} \otimes \mathcal{B} \subset \mathcal{F}$$

となる．

Proof. $B \in \mathcal{B}$ と $C \in \mathcal{C}$ に対して $E := B \times C$ とすると

$$\begin{aligned}\iint \mathbb{1}_E(x, y)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y) &= \mu(B) \int \mathbb{1}_C(y)\,\mathrm{d}\nu(y) = \mu(B)\nu(C)\\ &= \nu(C) \int \mathbb{1}_B(x)\,\mathrm{d}\mu(x)\\ &= \iint 1_E(x, y)\,\mathrm{d}\nu(y)\,\mathrm{d}\mu(x)\end{aligned}$$

となる．したがって

$$\mathcal{R} := \{B \times C;\ B \in \mathcal{B}, C \in \mathcal{C}\}$$

としたとき

$$\mathcal{R} \subset \mathcal{F}$$

となる．$E_n \in \mathcal{F}\,(n = 1, 2, \ldots)$ で $E_1 \supset E_2 \supset \cdots$ かつ $\bigcap_{n=1}^{\infty} E_n =: E$ または $E_1 \subset E_2 \subset \cdots$ かつ $\bigcup_{n=1}^{\infty} E_n =: E$ としたとき, $\mu(\mathbb{X}) < \infty$ かつ $\nu(\mathbb{Y}) < \infty$ なので, 有界収束定理 (定理 C.49) から $E \in \mathcal{F}$ となる．したがって $\mathcal{F}$ は単調族. $\mathcal{A}$ を $\mathcal{R}$ の有限個の互いに排反な元の和全体からなる集合族とすると $\mathcal{A} \subset \mathcal{F}$ となる．したがって定理 C.63 から $\mathcal{F}$ は σ 代数で $\mathcal{B} \otimes \mathcal{C} \subset \mathcal{F}$ となる． □

定理 C.65. $(\mathbb{X}, \mathcal{A}, \mu)$ と $(\mathbb{Y}, \mathcal{B}, \nu)$ を σ 有限測度空間とする．このとき

$$\rho(A \times B) = \mu(A)\nu(B) \qquad (A \in \mathcal{A},\, B \in \mathcal{B})$$

と定めると ρ は $\mathcal{A} \otimes \mathcal{B}$ 上の測度に一意的に拡張され，$E \in \mathcal{A} \otimes \mathcal{B}$ に対して

$$\rho(E) = \iint \mathbb{1}_E(x,\, y)\, \mathrm{d}\mu(x)\, \mathrm{d}\nu(y) = \iint \mathbb{1}_E(x,\, y)\, \mathrm{d}\nu(y)\, \mathrm{d}\mu(x)$$

となる．

Proof. まず $\mu(\mathbb{X}) < \infty$ かつ $\nu(\mathbb{Y}) < \infty$ と仮定する．$E \in \mathcal{B} \otimes \mathcal{C}$ に対して

$$\alpha(E) := \iint \mathbb{1}_E(x,\, y)\, \mathrm{d}\mu(x)\, \mathrm{d}\nu(y)$$

とおく．定理 C.64 より積分の順序の交換ができる．α は有限加法的であり，単調収束定理（定理 C.49）から σ 加法性もわかる．

さらに $\mathcal{B} \otimes \mathcal{C}$ の測度 β に拡張できたとする．すると

$$\mathcal{G} := \{E \in \mathcal{B} \otimes \mathcal{C};\, \alpha(E) = \beta(E) = 1\}$$

とおくと $\mathcal{G} \subset \mathcal{A}$ かつ $\mathcal{G}$ は単調族なので，$\mathcal{G} \subset \mathcal{B} \otimes \mathcal{C}$ となる．よって定理は有限測度に対して成立する．

次に $B_m \subset \mathbb{X}\, (m = 1,\, 2,\, \ldots)$ と $C_n\, (n = 1,\, 2,\, \ldots)$ は排反な集合で $\mathbb{X} = \bigcup_{m=1}^{\infty} B_m$, $\mathbb{Y} = \bigcup_{n=1}^{\infty} C_n$ とする．$E \in \mathcal{B} \otimes \mathcal{C}$ とし，$E_{mn} := E \bigcup (B_m \times C_n)$ とおく．各 $m,\, n \in \mathbb{N}$ に対して

$$\iint \mathbb{1}_{E_{mn}}(x,\, y)\, \mathrm{d}\mu(x)\, \mathrm{d}\nu(y) = \iint \mathbb{1}_{E_{mn}}(x,\, y)\, \mathrm{d}\nu(y)\, \mathrm{d}\mu(x)$$

となる．上の式について m と n に関して和を取る．単調収束定理（定理 C.49）より，$\forall E \in \mathcal{B} \otimes \mathcal{C}$ に対して

$$\begin{aligned}
\alpha(E) &:= \iint \mathbb{1}_E(x,\, y)\, \mathrm{d}\mu(x)\, \mathrm{d}\nu(y) \\
&= \iint \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \mathbb{1}_{E_{mn}}(x,\, y)\, \mathrm{d}\mu(x)\, \mathrm{d}\nu(y) \\
&= \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \iint \mathbb{1}_{E_{mn}}(x,\, y)\, \mathrm{d}\mu(x)\, \mathrm{d}\nu(y) \\
&= \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \iint \mathbb{1}_{E_{mn}}(x,\, y)\, \mathrm{d}\nu(y)\, \mathrm{d}\mu(x) \\
&= \iint \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \mathbb{1}_{E_{mn}}(x,\, y)\, \mathrm{d}\nu(y)\, \mathrm{d}\mu(x) \\
&= \iint \mathbb{1}_E(x,\, y)\, \mathrm{d}\nu(y)\, \mathrm{d}\mu(x)
\end{aligned}$$

となる. α は有限加法的であり, 単調収束定理（定理 C.49）より, σ 加法的になり, α は $\mathcal{B}\otimes\mathcal{C}$ 上の測度で

$$\alpha(E)=\rho(E)\qquad(\forall E\in\mathcal{A})$$

となる. β を $\mathcal{B}\otimes\mathcal{C}$ 上の測度で

$$\beta(E)=\rho(E)\qquad(\forall E\in\mathcal{A})$$

とすると $\forall E\in\mathcal{B}\otimes\mathcal{C}$ に対して

$$\begin{aligned}\beta(E)=\beta\left(\bigcup_{m=1}^{\infty}\bigcup_{n=1}^{\infty}E_{mn}\right)&=\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\beta(E_{mn})=\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\alpha(E_{mn})\\&=\alpha\left(\bigcup_{m=1}^{\infty}\bigcup_{n=1}^{\infty}E_{mn}\right)=\alpha(E)\qquad(E_{mn}\in\mathcal{A})\end{aligned}$$

となり, 拡張は一意的であることもわかる. □

定理 C.66. $(\mathbb{X},\,\mathcal{A},\,\mu)$ と $(\mathbb{Y},\,\mathcal{B},\,\nu)$ を σ 有限測度空間とし, $f:\mathbb{X}\times\mathbb{Y}\to[0,\,\infty)\cup\{+\infty\}$ は $\mathcal{A}\otimes\mathcal{B}$ 可測関数または $f\in\mathcal{L}^1(\mathbb{X}\times\mathbb{Y},\,\mathcal{A}\otimes\mathcal{B},\,\mu\times\nu)$ とする. このとき

$$\begin{aligned}\int f(x,\,y)\,\mathrm{d}(\mu\times\nu)(x,\,y)&=\iint f(x,\,y)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y)\\&=\iint f(x,\,y)\,\mathrm{d}\nu(y)\,\mathrm{d}\mu(x)\end{aligned}$$

が成り立つ. ほとんど至るところの $y\in\mathbb{Y}$ に対して $\int f(x,\,y)\,\mathrm{d}\mu(x)$ は定義され, ほとんど至るところの $x\in\mathbb{X}$ に対して $\int f(x,\,y)\,\mathrm{d}\nu(y)$ は定義される.

Proof. 定理 C.65 より, 非負値単関数 f に対して定理は成立する. 非負値関数 f に対して, 定理 C.42 と単調収束定理（定理 C.49）より, 定理は成り立つ. $f\in\mathcal{L}^1(\mathbb{X}\times\mathbb{Y},\,\mathcal{B}\otimes\mathcal{C},\,\mu\times\nu)$ のとき, $f^+=\max\{f,\,0\}$, $f^-=\max\{-f,\,0\}$ に対して, 定理は成立する. したがってほとんど至るところの $y\in\mathbb{Y}$ で $\int f^+(x,\,y)\,\mu(x)<\infty$ は定義される. 同様に f^- も同様の性質を持つ. よって ν に関してほとんど至るところの $y\in\mathbb{Y}$ に対して $\int|f(x,\,y)|\,\mathrm{d}\mu(x)<+\infty$ となる. よって

$$\int f(x,\,y)\,\mathrm{d}\mu(x)=\int f^+(x,\,y)\,\mathrm{d}\mu(x)-\int f^-(x,\,y)\,\mathrm{d}\mu(x)$$

となり, それぞれの積分は有限値となる. よって $\iint f(x,\,y)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y)$ は定義され, 有限値を取るので

$$\begin{aligned}&\iint f(x,\,y)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y)\\&\quad=\iint f^+(x,\,y)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y)-\iint f^-(x,\,y)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y)\end{aligned}$$

となる．よって $f \in \mathcal{L}^1(\mathbb{X}\times\mathbb{Y}, \mathcal{B}\otimes\mathcal{C}, \mu\times\nu)$ に対しても定理は成立する．
□

C.7　絶対連続性と Radon-Nikodym の定理

定義 C.67. $(\mathbb{X}, \mathcal{A})$ を可測空間とし, μ と ν を $(\mathbb{X}, \mathcal{A})$ 上の測度とする．$\forall A \in \mathcal{A}$ に対して

$$\mu(A) = 0 \quad \Rightarrow \quad \nu(A) = 0$$

のとき ν は μ に関して**絶対連続**であるといい, $\nu \ll \mu$ と書く．

注意 C.68. f を可測関数とし, μ を測度とし

$$\int f(x)\,\mathrm{d}\mu(x)$$

が定義されているとする．このとき

$$\nu(A) := \int_A f(x)\,\mathrm{d}\mu(x) \qquad (A \in \mathcal{A})$$

と定めると $\mu\{x \in \mathbb{X};\, f(x) < 0\} > 0$ のとき ν は負値も取るので**符号付測度**となる．□

補題 C.69. μ と ν を可測空間 $(\mathbb{X}, \mathcal{A})$ 上の測度とする．次の条件を考える．

　任意の $\epsilon > 0$ に対してある $\delta > 0$ が存在して

$$\mu(A) < \delta \Rightarrow \nu(A) < \epsilon. \tag{C.13}$$

このとき, 以下が成立する．

- 条件 (C.13) が成立 $\Rightarrow$ $\nu \ll \mu$.
- $\nu \ll \mu$ かつ ν は有限測度 $\Rightarrow$ (C.13) が成立．

Proof. 第 1 の主張の証明: $\forall\epsilon > 0$ に対して $\mu(A) = 0$ とする．このとき $\nu(A) < \epsilon$ となる．ϵ は任意だったので, $\nu(A) = 0$. よって $\nu \ll \mu$.
第 2 の主張の証明: $\nu \ll \mu$ とする．ν は有限測度として, (1) は成立しないと仮定する．すると $\exists\epsilon > 0$ が存在して, $\forall n \in \mathbb{N}$ に対して $A_n \in \mathcal{A}$ が存在して

$$\mu(A_n) < \frac{1}{n^2} \quad \text{かつ} \quad \nu(A_n) \geq \epsilon$$

とできる.

$$A = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n$$

とおくと, Borel-Cantelli の第 1 の補題より

$$\mu(A) = 0 \qquad \left(\because \sum_{n=1}^{\infty} \mu(A_n) < \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty \right)$$

である. $B_k = \bigcup_{n=k}^{\infty} A_n$, $B = \bigcap_{k=1}^{\infty} B_k$ とおくと各 $k \in \mathbb{N}$ に対して, $B \subset B_k$ なので,

$$\mu(B) \leq \mu(B_k).$$

さらに

$$\mu(B_k) \leq \sum_{n=k}^{\infty} \mu(A_n) < \infty$$

なので,

$$\lim_{k \to \infty} \mu(B_k) = 0$$

となるので, $\mu(B) = 0$ となる. ν は有限測度なので

$$\nu(A) = \mu\left(\bigcap_{k=1}^{\infty} B_k \right) = \lim_{k \to \infty} \mu(B_k) = \lim_{k \to \infty} \mu\left(\bigcup_{n=k}^{\infty} A_n \right) \geq \epsilon$$

となり, $\nu(A) = 0$ と矛盾. よって第 2 の主張も証明された. □

定理 C.70. (Radon-Nikodym の定理) μ と ν を可測空間 $(\mathbb{X}, \mathcal{A})$ 上の測度とし, μ は σ 有限とする. このとき可測関数 $f : \mathbb{X} \to [0, \infty) \cup \{+\infty\}$ が存在して, $\forall A \in \mathcal{A}$ に対して

$$\nu(A) = \int_A f(x) \, \mathrm{d}\mu(x) \tag{C.14}$$

と書ける. さらに $g : \mathbb{X} \to \mathbb{R}$ は ν 可積分とする. このとき

$$\int g(x) \, \mathrm{d}\nu(x) = \int g(x) f(x) \, \mathrm{d}\mu(x)$$

が成立する. 関数 f を μ に関する ν の Radon-Nikodym **の微分**といい, $\mu - a.e.$ で一意的である. f のことを $\dfrac{\mathrm{d}\nu}{\mathrm{d}\mu}(x)$ と書く. ν が σ 有限のとき $\mu - a.e.$ で f は有限となる.

Proof. (1) 一意性の証明: (C.14) をみたす f が存在したとする．g は別の関数で

$$\mu\{x \in \mathbb{X};\, f(x) \neq g(x)\} > 0 \tag{C.15}$$

とする．$n \in \mathbb{N}$ に対して

$$A_n := \left\{x \in \mathbb{X};\, f(x) > g(x) + \frac{1}{n}\right\} \quad B_n := \left\{x \in \mathbb{X};\, f(x) < g(x) - \frac{1}{n}\right\}.$$

とする．(C.15) からある $n \in \mathbb{N}$ が存在して

$$\mu(A_n) > 0 \quad \text{または} \quad \mu(B_n) > 0$$

となる．$A \subset A_n$ または $A \subset B_N$ とし

$$0 < \mu(A_n) < +\infty$$

とする．このとき

$$\int_A f(x)\,\mathrm{d}\mu(x) \neq \int_A g(x)\,\mathrm{d}\mu(x)$$

となるので, $g \neq \dfrac{\mathrm{d}\nu}{\mathrm{d}\mu}$ となる．
(2) f の存在の証明: 以下を順に証明していく．

(2-1) μ が有限測度のとき (C.14) が成立すれば, μ が σ 有限測度でも (C.14) が成立する．

(2-2) μ が有限測度のとき (C.14) をみたす f を構成し, f は可測である．

(2 − 1) の証明: μ は σ 有限とする．$A_n \in \mathcal{A}\,(n = 1, 2, \ldots)$ は互いに排反で, $\mu(A_n) < \infty$ かつ $\bigcup_{n=1}^{\infty} A_n = \mathbb{X}$ とする．各 $k \in \mathbb{N}$ に対して

$$\mu_k(A) := \mu(A \cap A_k) \quad (A \in \mathcal{A}), \qquad \nu_k(A) := \nu(A \cap A_k) \quad (A \in \mathcal{A})$$

とする．明らかに各 $k \in \mathbb{N}$ について $\nu_k \ll \mu_k$ となる．すると $k \in \mathbb{N}$ についてある可測関数 $f_k : \mathbb{X} \to [0, +\infty) \cup \{+\infty\}$ が存在して

$$\nu_k(A) = \int_A f_k(x)\,\mathrm{d}\mu(x) \qquad (A \in \mathcal{A})$$

とできる．ここで

$$f(x) := \sum_{k=1}^{\infty} \mathbb{1}_{A_k} f_k(x)$$

とおけば, f は (C.14) をみたすことがわかる.
(2 − 2) の証明: μ は有限測度とし, 非負の有理数 $q \in \mathbb{Q}$ に対して

$$\nu_q := q\mu - \nu$$

と定める. すると $\nu_q(A) < \infty\,(\forall A \in \mathcal{A})$ となる. 各 $q \in \mathbb{Q}$ に対して

$$\begin{aligned}&\mathcal{P}_q := \{A \in \mathcal{A};\, \nu_q(B) \geq 0,\, \forall B \subset A\},\\ &\lambda_q := \sup_{A \in \mathcal{P}_q} \nu_q(A)\end{aligned}$$

と定める. 明らかに $\varnothing \in \mathcal{P}_q$ であるので, $\lambda_q \geq 0$ である. $\{A_n\}_{n=1}^{\infty} \subset \mathcal{A}$ かつ $A_n \in \mathcal{P}_q$ をうまくとり

$$\lambda_q = \lim_{n\to\infty} \nu_q(A_n)$$

となるようにできる. さらに

$$A^q := \bigcup_{n=1}^{\infty} A_n$$

とおく. A^q の部分集合は A_n の部分集合の和で表現できるので, $A^q \in \mathcal{P}_q$ となる[4]. したがって

$$\lambda_q \geq \nu_q(A^q) \tag{C.16}$$

がわかる. 一方, 各 $n \in \mathbb{N}$ について $A^q \setminus A_n \subset A^q$ となる. $A^q \in \mathcal{P}_q$ なので, 各 $n \in \mathbb{N}$ について

$$\nu_q(A^q \setminus A_n) \geq 0$$

となる. よって $\{A_n\}_{n=1}^{\infty}$ の取り方から

$$\nu_q(A^q) = \nu_q(A^q \setminus A_n) + \nu(A_n) \geq \nu_q(A_n) \overset{n\to\infty}{\longrightarrow} \lambda_q$$

となる. よって $\nu_q(A^q) \geq \lambda_q$ と (C.16) より

$$\lambda_q = \nu_q(A^q)$$

がわかる.
　次に $B^q = (A^q)^c$ とおいたとき

$$\nu_q(B) \leq 0 \qquad (\forall B \subset B^q) \tag{C.17}$$

[4] $\forall B \subset A^q$ に対して $B \cap A_n \subset A_n$. よって $0 \leq \nu_q(B \cap A_n) \leq \nu_q(B)$ なので, $A^q \in \mathcal{P}_q$ がわかる.

であることを背理法で示す.

そのために

$$\text{ある } B \subset B^q \text{ があって } \nu_q(B) > 0 \tag{C.18}$$

と仮定する. B が負の測度をもつ部分集合を持たなければ, $B \in \mathcal{P}_q$ となる. B と A^q は排反なので

$$\nu_q(A^q \cup B) = \nu_q(B^q) + \nu_q(B) > \lambda_q$$

となる. したがって

$$B \cup A^q \in \mathcal{P}_q \quad \text{かつ} \quad \nu_q(A^q \cup B) > \lambda$$

となるので, これは A^q の定義と矛盾する. よって

$$\exists \widetilde{B} \subset B \text{ で } \nu_q(\widetilde{B}) < 0$$

となるものが存在する.

このことを踏まえて自然数列 $\{n_k\}_{k=1}^{\infty}$ を以下のように定める. 各 $k \in \mathbb{N}$ と $k = 2, 3, \ldots$ に対して

$$\begin{aligned}
n_1 &:= \min\left\{n \in \mathbb{N};\, \nu_q(B_1) < -\frac{1}{n},\, \exists B_1 \subset B\right\}, \\
n_2 &:= \min\left\{n \in \mathbb{N} \cap [n_1, \infty);\, \nu_q(B_2) < -\frac{1}{n},\, \exists B_2 \subset B \setminus B_1\right\}, \\
&\ \vdots \\
n_k &:= \min\left\{n \in \mathbb{N} \cap [n_{k-1}, \infty);\, \nu_q(B_k) < -\frac{1}{n},\, \exists B_k \subset B \setminus \left(\bigcup_{\ell=1}^{k-1} B_\ell\right)\right\}
\end{aligned}$$

とする. いま

$$C := B \setminus \left(\bigcup_{k=1}^{\infty} B_k\right)$$

とおく. 明らかに $\nu_q(C) > 0$ である. なぜならば

$$\nu_q(C) = \nu_q(B) - \underbrace{\nu_q\left(\bigcup_{k=1}^{\infty} B_k\right)}_{<} > 0$$

よりわかる.

もし C が負の測度を持つ部分集合を持たなければ, $C \in \mathcal{P}_q$ となり,

$$C \subset B \subset B^q = (A^q)^c$$

と矛盾し, (C.17) が示せる．このことを踏まえ

$$\nu_q(D) \geq 0 \qquad (\forall D \subset C) \tag{C.19}$$

を背理法で示す．そのためにある $D \subset C$ が存在して

$$\nu_q(D) =: -\epsilon < 0$$

を仮定する．(C.18) で $\nu_q(B) > 0$ と仮定していたので

$$\sum_{k=1}^{\infty} \nu_q(B_k) > -\infty$$

となる．なぜならば $\sum_{k=1}^{\infty} \nu_q(B_k) = -\infty$ ならば

$$C = B \setminus \left(\bigcup_{k=1}^{\infty} B_k \right)$$

なので $\nu_q(B) > 0$ とはならない．よって

$$\lim_{k\to\infty} \nu_q(B_k) = 0 \quad \Rightarrow \quad \lim_{k\to\infty} n_k = \infty$$

となることがわかる．このことからある $k \in \mathbb{N}$ が存在して

$$\frac{1}{n_{k+1} - 1} < \epsilon \tag{C.20}$$

とできる．したがって

$$\nu_q(D) = -\epsilon < -\frac{1}{n_{k+1} - 1} \quad かつ \quad D \subset B \setminus \left(\bigcup_{\ell=1}^{k} B_\ell \right)$$

であることと n_{k+1} の定義に注意すると

$$\nu_q(D) \geq -\frac{1}{n_{k+1}}$$

となり, これと (C.20) は矛盾する．よって (C.19) が確認できた．これで (C.17) が示せた．

$\widetilde{q} > q$ なる $\widetilde{q} \in \mathbb{Q}$ をとる．すると

$$A^{\widetilde{q}} \supset A^q \Leftrightarrow B^q \supset B^{\widetilde{q}}$$

となること[5]に注意する．

$$A^q \cap B^{\widetilde{q}} \subset B^{\widetilde{q}} \quad かつ \quad \left(B^{\widetilde{q}}\right)^c = A^{\widetilde{q}} \in \mathcal{P}_{\widetilde{q}}$$

[5]なぜならば $\nu_{\widetilde{q}}(A) = \widetilde{q}\mu(A) - \nu(A) > q\mu(A) - \nu(A) = \nu_q(A)$ からわかる．

なので

$$\widetilde{q}\mu\bigl(A^q \cap B^{\widetilde{q}}\bigr) - \nu\bigl(A^q \cap B^{\widetilde{q}}\bigr) = \nu_{\widetilde{q}}\bigl(A^q \cap B^{\widetilde{q}}\bigr) \le 0 \tag{C.21}$$

となる．一方

$$A^q \cap B^{\widetilde{q}} \subset A^q \quad \text{かつ} \quad A^q \in \mathcal{P}_q$$

なので

$$q\mu\bigl(A^q \cap B^{\widetilde{q}}\bigr) - \nu\bigl(A^q \cap B^{\widetilde{q}}\bigr) = \nu_q\bigl(A^q \cap B^{\widetilde{q}}\bigr) \ge 0 \tag{C.22}$$

となる．(C.21) − (C.22) から

$$\bigl(\widetilde{q} - q\bigr)\mu\bigl(A^q \cap B^{\widetilde{q}}\bigr) \le 0$$

を得る．$\bigl(\widetilde{q} - q\bigr) > 0$ なので

$$\mu\bigl(A^q \cap B^{\widetilde{q}}\bigr) = 0$$

を得る．

以上のことを踏まえて

$$f(x) := \sup\{q \in \mathbb{Q};\, x \in B^q\} \tag{C.23}$$

と定まる．

まず f の可測性を示す．$B^0 = \mathbb{X}$ なので

$$f(x) \ge 0 \qquad (x \in \mathbb{X})$$

となる．明らかに

$$x \in B^q \quad \Rightarrow \quad f(x) \ge 0; \qquad x \in A^q \quad \Rightarrow \quad f(x) \le 0;$$

である．さらに

$$\{x \in \mathbb{X};\, f(x) \ge q\} = \bigcup_{\widetilde{q} \in \mathbb{Q};\, \widetilde{q} \ge q} B^{\widetilde{q}}$$

となる．よって補題 C.41 より f は可測となる．

つぎに (C.23) で定めた関数 f は任意の $A \in \mathcal{A}$ に対して (C.14) をみたすことを示す．そのために $A \in \mathcal{A}$ とし，$\epsilon > 0$ を取る．N を正の整数とし

$$N > \frac{\mu(A)}{\epsilon} \tag{C.24}$$

をみたすものとする．$k = 0,\,1,\,2,\,\ldots$ に対して

$$E_k := A \cap B^{k/N} \cap A^{(k+1)/N}; \quad E_\infty := A \setminus \left(\bigcup_{k=1}^{\infty} A^{k/N} \right)$$

と定める．すると

$$\begin{aligned}\left(\bigcup_{k=1}^{\infty} E_k \right) \cup E_\infty &= A \cap \left(B^{k/N} \right) \cap \left(\bigcup_{k=1}^{\infty} A^{(k+1)/N} \right) \cup \left\{ A \setminus \left(\bigcup_{k=1}^{\infty} A^{k/N} \right) \right\} \\ &= A\end{aligned}$$

となる．さらに $k < \widetilde{k}$ のとき

$$\left(B^{k/N} \cap A^{(k+1)/N} \right) \cap \left(B^{\widetilde{k}/N} \cap A^{(\widetilde{k}+1)/N} \right) = \varnothing$$

となるので，$\{E_k\}_{k=1}^{\infty}$ は排反になる．したがって

$$\nu(A) = \nu(E_\infty) + \sum_{k=1}^{\infty} \nu(E_k)$$

となる．f の作り方と

$$x \in E_k \;\Leftrightarrow\; x \in B^{k/N} \quad \text{かつ} \quad x \in A^{(k+1)/N}$$

から

$$\begin{aligned} x \in E_k \;&\Rightarrow\; \frac{k}{N} \leq f(x) \leq \frac{k+1}{N}; \\ x \in E_\infty \;&\Rightarrow\; f(x) = \infty \end{aligned}$$

となる．さらに

$$\begin{aligned} E_k \subset B^{k/N} \;&\Rightarrow\; \quad \nu_{k/N}(E_k) \leq 0; \\ E_k \subset A^{(k+1)/N} \;&\Rightarrow\; \nu_{(k+1)/N}(E_k) \geq 0 \end{aligned}$$

となる．こららのことから有限の k に対して

$$\begin{gathered} \frac{k}{N}\mu(E_k) \leq \int_{E_k} f(x)\,\mathrm{d}\mu(x) \leq \frac{k+1}{N}\mu(E_k) \\ \Rightarrow \\ \frac{k}{N}\mu(E_k) - \nu(E_k) \leq \int_{E_k} f(x)\,\mathrm{d}\mu(x) - \nu(E_k) \leq \frac{k+1}{N}\mu(E_k) - \nu(E_k) \end{gathered}$$

から

$$\left| \nu(E_k) - \int_{E_k} f(x)\,\mathrm{d}\mu(x) \right| \leq \frac{1}{N}\mu(E_k) \tag{C.25}$$

がわかる．任意の $q \in \mathbb{Q}\,(q \geq 0)$ に対して

$$E_\infty \subset \left(\bigcup_{k=1}^{\infty} A^{k/N}\right)^c$$

なので $0 > \nu_q(E_\infty) = q\mu(E_\infty) - \nu(E_\infty)$ となる．$\mu(E_\infty) > 0$ ならば

$$\nu(E_\infty) = \infty$$

となる．もし $\mu(E_\infty) =$ ならば, ν は μ に関して絶対連続なので, $\nu(E_\infty) = 0$ となる．いずれの場合でも

$$\nu(E_\infty) = \int_{E_\infty} f(x)\,\mathrm{d}\mu(x)$$

となる．これと (C.24) – (C.25) を合わせると

$$\left|\nu(A) - \int_A f(x)\,\mathrm{d}\mu(x)\right| \leq \frac{1}{N}\mu(A) < \epsilon$$

を得る．ϵ は任意だったので, 任意の $A \in \mathcal{A}$ に対して (C.14) をみたすことを示せた． □

注意 C.71. 上記の証明は Hahn 分解を経由したものである．Hilbert 空間と Riesz の定理を用いた関数解析的な証明は Axler (2020, pp.272-274) を参照のこと．Schilling(2017, pp.230-234) により関数解析で初等的な証明がある． □

C.7.1 定理 C.70 の証明の主張 $(2-2)$ の Bradley (1989) による別証明

ここでは Bradley (1989) による Radon-Nikodym の微分の存在の初等的な証明を書く．Hahn 分解を経由することなく $\mathbb{X}$ の分割を用いて Radon-Nikodym の微分を構成している．

まず特別な場合について Radon-Nikodym の微分を構成する．

定理 C.72. $(\mathbb{X}, \mathcal{A})$ を可測空間とし, ν と μ を $(\mathbb{X}, \mathcal{A})$ 上の非負値測度で

(i) $\mu(\mathbb{X}) = 1$,

(ii) $\forall A \in \mathcal{A}$ に対して $\nu(A) \leq \mu(A)$

をみたすとする．このとき $\mathbb{X}$ 上の $\mathcal{A}$ 可測関数 h で $0 \leq h(x) \leq 1\,(x \in \mathbb{X})$ なるものが存在して任意の $A \in \mathcal{A}$ に対して

$$\nu(A) = \int_A h(x)\,\mathrm{d}\mu(x) \tag{C.26}$$

が成り立つ．すなわち $h = \dfrac{\mathrm{d}\nu}{\mathrm{d}\mu}$ である．

以下では定理 C.72 を証明するための準備をする．

各 $n \in \mathbb{N}$ に対して $\Pi = \{A_1, A_2, \ldots, A_n\}$ を $\mathbb{X}$ **の** $\mathcal{A}$ **可測な分割**とする．すなわち

$$\mathbb{X} = \bigcup_{i=1}^{n} A_i \quad \text{かつ} \quad A_i \in \mathcal{A},\, A_i \cap A_j = \varnothing\,(i \neq j;\, i,\, j = 1,\, 2,\, \ldots,\, n)$$

である．$\mathbb{X}$ 上の関数 h_Π をつぎのように定める．$\forall x \in A_i\,(i = 1,\, 2,\, \ldots,\, n)$ に対して

$$h_\Pi(x) = \begin{cases} \dfrac{\nu(A_i)}{\mu(A_i)} & (\mu(A_i) > 0) \\ 0 & (\mu(A_i) = 0) \end{cases}$$

と定める．すると定理 C.72 の条件 (ii) より

$$0 \le h_\Pi(x) \le 1 \qquad (x \in \mathbb{X}) \tag{C.27}$$

となることに注意する．

補題 C.73. $n \in \mathbb{N}$ とし，$\Pi = \{A_j\}_{j=1}^{n}$ を $\mathbb{X}$ の $\mathcal{A}$ 可測な分割とする．ある添え字集合 $I \subset \{1,\, 2,\, \ldots,\, n\}$ に対して

$$A = \bigcup_{i \in I} A_i$$

と書けたとき

$$\nu(A) = \int_A h_\Pi(x)\,\mathrm{d}\mu(x)$$

となる．特に

$$\nu(\mathbb{X}) = \int_{\mathbb{X}} h_\Pi(x)\,\mathrm{d}\mu(x)$$

である．

Proof.

$$\begin{aligned} \int_A h_\Pi(x)\,\mathrm{d}\mu(x) &= \int_{\bigcup_{i \in I} A_i} h_\Pi(x)\,\mathrm{d}\mu(x) = \sum_{i \in I} \int_{A_i} h_\Pi(x)\,\mathrm{d}\mu(x) \\ &= \sum_{i \in I} \int_{A_i} \frac{\nu(A_i)}{\mu(A_i)}\,\mathrm{d}\mu(x) = \sum_{i \in I} \frac{\nu(A_i)}{\mu(A_i)} \int_{A_i} \mathrm{d}\mu(x) \\ &= \sum_{i \in I} \nu(A_i) = \nu\left(\bigcup_{i \in I} A_i\right) = \nu(A). \end{aligned}$$

□

補題 C.74. Π_1 と Π_2 は $\mathbb{X}$ の $\mathcal{A}$ 可測な分割で Π_2 は Π_1 の細分とする. すなわち別の分割 Π_0 があって

$$\Pi_2 = \{A \cap B;\, A \in \Pi_1,\, B \in \Pi_0\}$$

と書ける. このとき

$$\begin{aligned}\int_{\mathbb{X}} \{h_{\Pi_2}(x)\}^2 \,\mathrm{d}\mu(x) &= \int_{\mathbb{X}} \{h_{\Pi_1}(x)\}^2 \,\mathrm{d}\mu(x) + \int_{\mathbb{X}} \{h_{\Pi_2}(x) - h_{\Pi_1}(x)\}^2 \,\mathrm{d}\mu(x) \\ &\geq \int_{\mathbb{X}} \{h_{\Pi_1}(x)\}^2 \,\mathrm{d}\mu(x)\end{aligned}$$

が成り立つ.

Proof. $A \in \Pi_1$ に対して

$$\int_A h_{\Pi_2}(x) \,\mathrm{d}\mu(x) = \nu(A) = \int_A h_{\Pi_1}(x) \,\mathrm{d}\mu(x)$$

となる. 2 番目の等号は h_{Π_1} の定義と $A \in \Pi_1$ から明らかである. 1 番目の等号を示す. Π_2 は Π_1 の細分なので, ある添え字集合 $I \subset \{1, 2, \dots, \#(\Pi_2)\}$ があって

$$A = \bigcup_{i \in I} B_i, \qquad B_i \in \Pi_2 =: \{B_1, B_2, \dots, B_{\#(\Pi_2)}\}$$

と書ける. このことから

$$\begin{aligned}\int_A h_{\Pi_2}(x) \,\mathrm{d}\mu(x) &= \sum_{i \in I} \int_{B_i} h_{\Pi_2}(x) \,\mathrm{d}\mu(x) = \sum_{i \in I} \frac{\nu(B_i)}{\mu(B_i)} \mu(B_i) = \sum_{i \in I} \nu(B_i) \\ &= \nu\left(\bigcup_{i \in I} B_i\right) = \nu(A)\end{aligned}$$

からわかる. さらに $A \in \Pi_1$ に対して

$$\int_A h_{\Pi_1}(x) h_{\Pi_2}(x) \,\mathrm{d}\mu(x) = \int_A \{h_{\Pi_1}(x)\}^2 \,\mathrm{d}\mu(x) \tag{C.28}$$

となる．なぜならば

$$
\begin{aligned}
\int_A h_{\Pi_1}(x) h_{\Pi_2}(x)\,\mathrm{d}\mu(x) &= \int_{\bigcup_{i\in I} B_i} h_{\Pi_1}(x) h_{\Pi_2}(x)\,\mathrm{d}\mu(x) \\
&= \sum_{i\in I} \int_{B_i} h_{\Pi_1}(x) h_{\Pi_2}(x)\,\mathrm{d}\mu(x) \\
&= \sum_{i\in I} \int_{B_i} \frac{\nu(B_i)}{\mu(B_i)} \frac{\nu(A)}{\mu(A)}\,\mathrm{d}\mu(x) \\
&= \sum_{i\in I} \frac{\nu(B_i)}{\mu(B_i)} \frac{\nu(A)}{\mu(A)} \int_{B_i} \mathrm{d}\mu(x) \\
&= \sum_{i\in I} \frac{\nu(B_i)}{\mu(B_i)} \frac{\nu(A)}{\mu(A)} \mu(B_i) \\
&= \sum_{i\in I} \nu(B_i) \frac{\nu(A)}{\mu(A)} \\
&= \nu\Bigg(\bigcup_{i\in I} B_i\Bigg) \frac{\nu(A)}{\mu(A)} \\
&= \nu(A) \frac{\nu(A)}{\mu(A)} \\
&= \frac{\nu(A)}{\mu(A)} \frac{\nu(A)}{\mu(A)} \int_A \mathrm{d}\mu(x) \\
&= \int_A \left\{\frac{\nu(A)}{\mu(A)}\right\}^2 \mathrm{d}\mu(x) \\
&= \int_A \{h_{\Pi_1}(x)\}^2\,\mathrm{d}\mu(x)
\end{aligned}
$$

からわかる．(C.28) から

$$
\begin{aligned}
\int_{\mathbb{X}} \{h_{\Pi_2}(x)\}^2\,\mathrm{d}\mu(x) &= \int_{\mathbb{X}} \Big\{ h_{\Pi_1}(x) + \big(h_{\Pi_2}(x) - h_{\Pi_1}(x)\big) \Big\}^2 \mathrm{d}\mu(x) \\
&= \int_{\mathbb{X}} \{h_{\Pi_1}(x)\}^2\,\mathrm{d}\mu(x) + \int_{\mathbb{X}} \big(h_{\Pi_2}(x) - h_{\Pi_1}(x)\big)^2\,\mathrm{d}\mu(x) \\
&\quad + 2 \underbrace{\int_{\mathbb{X}} h_{\Pi_1}(x) \big\{h_{\Pi_2}(x) - h_{\Pi_1}(x)\big\}\,\mathrm{d}\mu(x)}_{=0 \quad (\because \text{(C.28)})} \\
&= \int_{\mathbb{X}} \{h_{\Pi_1}(x)\}^2\,\mathrm{d}\mu(x) + \int_{\mathbb{X}} \big(h_{\Pi_2}(x) - h_{\Pi_1}(x)\big)^2\,\mathrm{d}\mu(x) \\
&\geq \int_{\mathbb{X}} \{h_{\Pi_1}(x)\}^2\,\mathrm{d}\mu(x)
\end{aligned}
$$

を得る．よって補題は証明された．　□

定理 C.72 **の証明** ((C.26) における h の構成):

$$c := \sup \int_{\mathbb{X}} \{h_{\Pi}(x)\}^2 \, \mathrm{d}\mu(x)$$

とおく．ただし上の式の上限は $\mathbb{X}$ の $\mathcal{A}$ 可測な有限分割すべてに関するものである．(C.27) と $\mu(\mathbb{X}) = 1$ より

$$0 \leq \sup \int_{\mathbb{X}} \{h_{\Pi}(x)\}^2 \, \mathrm{d}\mu(x) \leq \sup \int_{\mathbb{X}} \mathrm{d}\mu(x) = \mu(\mathbb{X}) = 1$$

となるので

$$c \leq 1 \tag{C.29}$$

となる．

各 $n = 1, 2, \ldots$ に対して Π_n を $\mathbb{X}$ の $\mathcal{A}$ 可測な有限分割で

$$\int_{\mathbb{X}} \{h_{\Pi_n}(x)\}^2 \, \mathrm{d}\mu(x) \geq c - \left(\frac{1}{4}\right)^n$$

をみたすように取る．

$$\Pi_n^c := \left\{A_1 \cap A_2 \cap \cdots \cap A_n ;\, A_i \in \Pi_i \, (i = 1, 2, \ldots, n)\right\} \tag{C.30}$$

とする．すなわち $\Pi_1, \Pi_2, \ldots, \Pi_n$ の共通の細分で一番おおきいものである．補題 C.74 より $\forall n \in \mathbb{N}$ に対して

$$\begin{aligned} c - \left(\frac{1}{4}\right)^n &\leq \int_{\mathbb{X}} \{h_{\Pi_n}(x)\}^2 \, \mathrm{d}\mu(x) \leq \int_{\mathbb{X}} \{h_{\Pi_n^c}(x)\}^2 \, \mathrm{d}\mu(x) \\ &\leq \int_{\mathbb{X}} \{h_{\Pi_{n+1}^c}(x)\}^2 \, \mathrm{d}\mu(x) \end{aligned}$$

となる．よって各 $n \in \mathbb{N}$ に対して

$$\begin{aligned} &\int_{\mathbb{X}} \left\{h_{\Pi_{n+1}^c}(x) - h_{\Pi_n^c}(x)\right\}^2 \mathrm{d}\mu(x) \\ &\quad = \int_{\mathbb{X}} \{h_{\Pi_{n+1}^c}(x)\}^2 \, \mathrm{d}\mu(x) - 2 \underbrace{\int_{\mathbb{X}} h_{\Pi_{n+1}^c}(x) h_{\Pi_n^c}(x) \, \mathrm{d}\mu(x)}_{= \int_{\mathbb{X}} \{h_{\Pi_n^c}(x)\}^2 \, \mathrm{d}\mu(x) \quad (\because \text{(C.28)})} \\ &\qquad + \int_{\mathbb{X}} \{h_{\Pi_n^c}(x)\}^2 \, \mathrm{d}\mu(x) \\ &\quad = \int_{\mathbb{X}} \{h_{\Pi_{n+1}^c}(x)\}^2 \, \mathrm{d}\mu(x) - \int_{\mathbb{X}} \{h_{\Pi_n^c}(x)\}^2 \, \mathrm{d}\mu(x) \\ &\quad \leq c - \left\{c - \left(\frac{1}{4}\right)^2\right\} = \left(\frac{1}{4}\right)^2 \end{aligned} \tag{C.31}$$

を得る．Cauchy-Schwarz の不等式から

$$\begin{aligned}
&\int_{\mathbb{X}} \left| h_{\Pi^c_{n+1}}(x) - h_{\Pi^c_n}(x) \right| \mathrm{d}\mu(x) \\
&\quad \le \sqrt{\int_{\mathbb{X}} \left\{ h_{\Pi^c_{n+1}}(x) - h_{\Pi^c_n}(x) \right\}^2 \mathrm{d}\mu(x)} \sqrt{\int_{\mathbb{X}} 1^2 \, \mathrm{d}\mu(x)} \\
&\quad \le \sqrt{\int_{\mathbb{X}} \left\{ h_{\Pi^c_{n+1}}(x) - h_{\Pi^c_n}(x) \right\}^2 \mathrm{d}\mu(x)} \sqrt{\mu(\mathbb{X})} \\
&\quad \le \left(\frac{1}{2} \right)^2 \qquad (\because \text{(C.31) と } \mu(\mathbb{X}) = 1)
\end{aligned}$$

を得る．単調収束定理から

$$\begin{aligned}
\int_{\mathbb{X}} \left\{ \sum_{n=1}^{\infty} \left| h_{\Pi^c_{n+1}}(x) - h_{\Pi^c_n}(x) \right| \right\} \mathrm{d}\mu(x) &= \sum_{n=1}^{\infty} \int_{\mathbb{X}} \left| h_{\Pi^c_{n+1}}(x) - h_{\Pi^c_n}(x) \right| \mathrm{d}\mu(x) \\
&\le \sum_{n=1}^{\infty} \left(\frac{1}{2} \right)^n < \infty
\end{aligned}$$

となる．よって積分の性質から

$$\mu\left(\sum_{n=1}^{\infty} \left| h_{\Pi^c_{n+1}}(x) - h_{\Pi^c_n}(x) \right| < \infty \right) = 1$$

となる．すなわち μ に関してほとんど至るところで

$$\sum_{n=1}^{\infty} \left| h_{\Pi^c_{n+1}}(x) - h_{\Pi^c_n}(x) \right| < \infty$$

である．このことより μ に関してほとんどいたるところで

$$\lim_{n \to \infty} h_{\Pi^c_n}(x)$$

は収束する．$\mathbb{X}$ 上の関数 h をつぎのように定める．$x \in \mathbb{X}$ に対して

$$h(x) := \begin{cases} \lim_{n \to \infty} h_{\Pi^c_n}(x) & (\text{右辺の極限が収束するとき}) \\ 0 & (\text{その他}) \end{cases}$$

とする．作り方から関数 h は $\mathcal{A}$ 可測であり, (C.27) から

$$0 \le h(x) \le 1 \qquad (x \in \mathbb{X})$$

となる．あとは $\forall A \in \mathbb{A}$ に対して (C.26) となることを示せばよい．そのために (C.30) で定めた分割 Π^c_n に対して

$$\Pi^A_n := \{ B \cap A, \, B \cap (\mathbb{X} \setminus A); \, B \in \Pi^c_n \}$$

とおく. (C.28) と (C.31) から

$$\begin{aligned}\int_{\mathbb{X}}\left\{h_{\Pi_n^A}(x)-h_{\Pi_n^c}(x)\right\}^2\mathrm{d}\mu(x)&=\int_{\mathbb{X}}\{h_{\Pi_n^A}(x)\}^2\,\mathrm{d}\mu(x)-\int_{\mathbb{X}}\{h_{\Pi_n^c}(x)\}^2\,\mathrm{d}\mu(x)\\&\le\left(\frac{1}{4}\right)^4\end{aligned}\tag{C.32}$$

となる. 再度 Cauchy-Schwarz の不等式と (C.32) を用いると

$$\left|\int_{\mathbb{A}}\{h_{\Pi_n^A}(x)-h_{\Pi_n^c}(x)\}\,\mathrm{d}\mu(x)\right|\tag{C.33}$$

$$\begin{aligned}&\le\int_{\mathbb{A}}\left|h_{\Pi_n^A}(x)-h_{\Pi_n^c}(x)\right|\mathrm{d}\mu(x)\\&\le\sqrt{\int_A\left\{h_{\Pi_n^A}(x)-h_{\Pi_n^c}(x)\right\}^2\mathrm{d}\mu(x)}\sqrt{\int_A 1\,\mathrm{d}\mu(x)}\\&\le\left(\frac{1}{2}\right)^2\underbrace{\sqrt{\mu(\mathbb{X})}}_{=1}\le\left(\frac{1}{2}\right)^2\end{aligned}\tag{C.34}$$

となる. 補題 C.73 から $n\in\mathbb{N}$ に対して

$$\nu(A)=\int_A h_{\Pi_n^A}(x)\,\mathrm{d}\mu(x)=\int_A\left\{h_{\Pi_n^A}(x)-h_{\Pi_n^c}(x)\right\}\mathrm{d}\mu(x)+\int_A h_{\Pi_n^c}(x)\,\mathrm{d}\mu(x)$$

となる. (C.34) より

$$\lim_{n\to\infty}\int_A\left\{h_{\Pi_n^A}(x)-h_{\Pi_n^c}(x)\right\}\ \mathrm{d}\mu(x)=0$$

となる. このことと有界収束定理から

$$\begin{aligned}\nu(A)&=\lim_{n\to\infty}\int_A h_{\Pi_n^A}(x)\,\mathrm{d}\mu\\&=\underbrace{\lim_{n\to\infty}\int_A\left\{h_{\Pi_n^A}(x)-h_{\Pi_n^c}(x)\right\}\ \mathrm{d}\mu(x)}_{=0}+\lim_{n\to\infty}\int_A h_{\Pi_n^c}(x)\,\mathrm{d}\mu(x)\\&=\int_A\lim_{n\to\infty}\quad h_{\Pi_n^c}(x)\,\mathrm{d}\mu(x)\\&\qquad(\because |h_{\Pi_n^c}(x)|\le 1\text{ と }\mu(\mathbb{X})=1\text{ から有界収束定理})\\&=\int_A h(x)\,\mathrm{d}\mu(x)\end{aligned}$$

を得る. よって定理 C.72 は証明された. □

系 C.75. μ と ν は測度空間 $(\mathbb{X},\mathcal{A})$ 上の非負値測度で

(i) $0 < \mu(\mathbb{X}) < \infty$,

(ii) $\forall A \in \mathcal{A}$ に対して $\nu(A) \leq \mu(A)$

とする．このとき $\mathbb{X}$ 上の $\mathcal{A}$ 可測関数 h で $0 \leq h(x) \leq 1\,(x \in \mathbb{X})$ なるものが存在して

$$\nu(A) = \int_A h(x)\,\mathrm{d}\mu(x) \qquad (A \in \mathcal{A}) \tag{C.35}$$

が成り立つ．すなわち $h = \dfrac{\mathrm{d}\nu}{\mathrm{d}\mu}$ である．

Proof.

$$\widetilde{\mu}(A) = \frac{\mu(A)}{\mu(\mathbb{X})}, \quad \widetilde{\nu}(A) = \frac{\nu(A)}{\mu(\mathbb{X})}$$

とおくと

$$\widetilde{\nu}(A) \leq \widetilde{\mu}(A)\,(\forall A \in \mathcal{A}) \quad \text{かつ} \quad \widetilde{\mu}(\mathbb{X}) = 1$$

となるので定理 C.72 の仮定を $\widetilde{\mu}$ と $\widetilde{\nu}$ はみたす．したがって $\mathcal{A}$ 可測関数で $0 \leq h(x) \leq 1\,(x \in \mathbb{X})$ なるものが存在して

$$\widetilde{\nu}(A) = \int_A h(x)\,\mathrm{d}\widetilde{\mu}(x) \qquad (A \in \mathcal{A})$$

となる．よって

$$\begin{aligned}\nu(A) &= \mu(\mathbb{X})\widetilde{\nu}(A) = \mu(\mathbb{X})\int_A h(x)\,\mathrm{d}\widetilde{\mu}(x) = \int_A h(x)\widetilde{\mu}(\mathbb{X})\,\mathrm{d}\widetilde{\mu}(x)\\ &= \int_A h(x)\,\mathrm{d}\mu(x)\end{aligned}$$

がわかる．□

系 C.76. μ と ν は可測空間 $(\mathbb{X}, \mathcal{A})$ 上の非負値測度で

(1) $0 < \mu(\mathbb{X}) < \infty$,

(2) $\forall A \in \mathcal{A} \Rightarrow \nu(A) \leq \mu(A)$

をみたすとする．任意の非負値 $\mathcal{A}$ 可測関数 $g : \mathbb{X} \to \mathbb{R}$ に対して

$$\int_{\mathbb{X}} g(x)\,\mathrm{d}\nu(x) = \int_{\mathbb{X}} g(x)h(x)\,\mathrm{d}\mu(x) \tag{C.36}$$

が成り立つ．ただし $h =: \dfrac{\mathrm{d}\nu}{\mathrm{d}\mu}$ は (C.35) をみたす関数である．

Proof. $B \in \mathcal{A}$ に対して $g(x) = \mathbb{1}_B(x)$ のとき (C.36) を示す．これは系 C.75 を用いると

$$\int_{\mathbb{X}} g(x)\,\mathrm{d}\nu(x) = \int_{\mathbb{X}} \mathbb{1}_B(x)\,\mathrm{d}\nu(x) = \nu(B) = \int_B h(x)\,\mathrm{d}\mu(x)$$
$$= \int_{\mathbb{X}} \mathbb{1}_B(x)h(x)\,\mathrm{d}\mu(x) \tag{C.37}$$

となることからわかる．つぎに $\{g_n\}_{n=1}^{\infty}$ を非負値単関数の増加列で

$$g_n \uparrow g\,(n \to \infty)$$

とする．すると単調収束定理から

$$\begin{aligned}
\int_{\mathbb{X}} g(x)\,\mathrm{d}\nu(x) &= \lim_{n\to\infty} \int_{\mathbb{X}} g_n(x)\,\mathrm{d}\nu(x) && (\because \text{単調収束定理}) \\
&= \lim_{n\to\infty} \int_{\mathbb{X}} g_n(x)h(x)\,\mathrm{d}\nu(x) && (\because \text{(C.37) から}) \\
&= \int_{\mathbb{X}} \lim_{n\to\infty} g_n(x)h(x)\,\mathrm{d}\nu(x) && (\because \text{単調収束定理}) \\
&= \int_A g(x)h(x)\,\mathrm{d}\mu(x)
\end{aligned}$$

となり系は証明された．□

定理 C.77. μ と ν は測度空間 $(\mathbb{X}, \mathcal{A})$ 上の非負値測度で

(1) $0 < \nu(\mathbb{X}) < \infty$ かつ $0 < \mu(\mathbb{X}) < \infty$,

(2) $\nu \ll \mu$

とする．このとき $\mathbb{X}$ 上の $\mathcal{A}$ 可測関数 h が存在して

$$\nu(A) = \int_A h(x)\,\mathrm{d}\mu(x) \qquad (\forall A \in \mathcal{A})$$

をみたす．すなわち $h = \dfrac{\mathrm{d}\nu}{\mathrm{d}\mu}$ である．

Proof. $\forall A \in \mathcal{A}$ に対して $\nu(A) \le (\nu+\mu)(A)$ かつ $\mu(A) \le (\nu+\mu)(A)$ なので, 系 C.75 から $\mathbb{X}$ 上の $\mathcal{A}$ 可測関数 h_ν と h_μ で $0 \le h_\nu(x) \le 1$, $0 \le h_\mu(x) \le 1\,(x \in \mathbb{X})$ なるものが存在して

$$\nu(A) = \int_A h_\nu(x)\,\mathrm{d}(\nu+\mu)(x), \qquad \nu(A) = \int_A h_\mu(x)\,\mathrm{d}(\nu+\mu)(x),$$

となる．ここで

$$B := \{x \in \mathbb{X};\, h_\mu(x) > 0\}, \quad C := B^c = \{x \in \mathbb{X};\, h_\mu(x) = 0\}$$

とおく．すると

$$\mu(C) = \int_C h_\mu(x)\,\mathrm{d}(\nu+\mu)(x) = 0$$

となるので, 条件 (ii) から

$$\nu(C) = 0$$

となる．このことから $\mathbb{X}$ 上の関数 h を

$$h(x) := \begin{cases} \dfrac{h_\nu(x)}{h_\mu(x)} & (x \in B) \\ 0 & (x \in C) \end{cases} \tag{C.38}$$

と定める．$A \in \mathcal{A}$ かつ $A \subset B$ に対して

$$\begin{aligned} \nu(A) &= \int_A h_\nu(x)\,\mathrm{d}(\nu+\mu)(x) = \int_A \frac{h_\nu(x)}{h_\mu(x)}\,h_\mu(x)\,\mathrm{d}(\nu+\mu)(x) \\ &= \int_A h(x)h_\mu(x)\,\mathrm{d}(\nu+\mu)(x) = \int_A h(x)\,\mathrm{d}\mu(x) \quad (\because \text{系 C.76}) \end{aligned} \tag{C.39}$$

となる．$\nu(C) = \mu(C) = 0$ かつ $A \cap C \in \mathcal{A}$ である．このこと B と C の定義に注意すると $\nu(A \cap C) = 0$, $A \cap B \subset B$, $B \cup C = \mathbb{X}$ となる．上記を踏まえて (C.39) を用いると

$$\begin{aligned} \nu(A) &= \nu(A \cap B) + \underbrace{\nu(A \cap C)}_{=0} \qquad (\because B \cap C = \varnothing,\ B \cup C = \mathbb{X}) \\ &= \nu(A \cap B) \\ &= \int_{A \cap B} h(x)\,\mathrm{d}\mu(x) \quad (\because \text{(C.39)}) \\ &= \int_{A \cap B} h(x)\,\mathrm{d}\mu(x) + \int_{A \cap C} h(x)\,\mathrm{d}\mu(x) \\ &\qquad (\text{(C.38) から } A \cap C \text{ 上では } h(x) = 0\ ) \\ &= \int_A h(x)\,\mathrm{d}\mu(x) \end{aligned}$$

を得る．よって定理は証明された．　□

C.8　Jacobi 変換定理

補題 C.78. $(\mathbb{X}, \mathcal{A})$ と $(\mathbb{X}', \mathcal{A}')$ を測度空間とし, $T : \mathbb{X} \to \mathbb{X}'$ を可測関数とする．$(\mathbb{X}, \mathcal{A})$ 上の測度 μ に対して

$$\mu'(A') := \mu\bigl(T^{-1}(A')\bigr) \qquad (\forall A' \in \mathcal{A}')$$

と定めたとき, μ' は $(\mathbb{X}', \mathcal{A}')$ 上の測度となる．

Proof. $A' = \varnothing$ のとき, $T^{-1}(\varnothing) = \varnothing$ から

$$\mu'(\varnothing) = \mu(\varnothing) = 0$$

となる. $\{A'_n\}_{n=1}^{\infty} \subset \mathcal{A}'$ 互いに排反な部分集合列としたとき

$$\begin{aligned}
\mu'\Bigl(\bigcup_{n=1}^{\infty} A'_n\Bigr) = \mu\Bigl(T^{-1}\Bigl(\bigcup_{n=1}^{\infty} A'_n\Bigr)\Bigr) &= \mu\Bigl(\bigcup_{n=1}^{\infty} T^{-1}(A'_n)\Bigr) \\
&\quad (\because \text{逆像の性質}) \\
&= \sum_{n=1}^{\infty} \mu\bigl(T^{-1}(A'_n)\bigr) \qquad (\because \mu \text{ の } \sigma \text{ 加法性}) \\
&= \sum_{n=1}^{\infty} \mu'(A'_n)
\end{aligned}$$

となる. よって μ' の σ 加法性は証明された. □

注意 C.79. μ' のことを $T(\mu)$ と書くことにする. すなわち

$$T(\mu)(A') := \mu\bigl(T^{-1}(A')\bigr) \qquad (A' \in \mathcal{A}')$$

である. □

補題 C.80. $(\mathbb{X}, \mathcal{A}, \mu)$ を測度空間とし, $(\mathbb{X}', \mathcal{A}')$ を可測空間とし, $T : \mathbb{X} \to \mathbb{X}'$ を可測写像とする. $T(\mu)$ 可積関数 $f : (\mathbb{X}', \mathcal{A}') \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ に対して

$$\int_{\mathbb{X}} f(\boldsymbol{x}')\,\mathrm{d}T(\mu)(\boldsymbol{x}') = \int_{\mathbb{X}'} f\bigl(T(\boldsymbol{x})\bigr)\,\mathrm{d}\mu(\boldsymbol{x}) \tag{C.40}$$

が成り立つ.

Proof. f と T は可測なので $f \circ T$ も可測. 標準機械 (standard machine) を用いて証明する.
<u>第 1 段階</u>: $f(\boldsymbol{x}') = \mathbb{1}_{A'}(\boldsymbol{x}')\,(A' \in \mathcal{A}')$ とする. すると $\mathbb{1}_{T^{-1}(A')}(\boldsymbol{x}) = \mathbb{1}_{A'}\bigl(T(\boldsymbol{x})\bigr)$ となるので

$$\begin{aligned}
\int_{\mathbb{X}'} f(\boldsymbol{x}')\,\mathrm{d}T(\mu)(\boldsymbol{x}') &= \int_{\mathbb{X}'} \mathbb{1}_{A'}(\boldsymbol{x}')\,\mathrm{d}T(\mu)(\boldsymbol{x}') \\
&= T(\mu)(A') = \mu\bigl(T^{-1}(A')\bigr) = \int_{\mathbb{X}} \mathbb{1}_{T^{-1}(A')}(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) \\
&= \int_{\mathbb{X}} \mathbb{1}_{T^{-1}(A')}\bigl(T(\boldsymbol{x})\bigr)\,\mathrm{d}\mu(\boldsymbol{x}) = \int_{\mathbb{X}} f\bigl(T(\boldsymbol{x})\bigr)\,\mathrm{d}\mu(\boldsymbol{x})
\end{aligned}$$

となるので $f = \mathbb{1}_{A'}$ のとき (C.40) は成立する.

第 2 段階: $N \in \mathbb{N}$, $\alpha_n \geq 0$, $A'_n \in \mathcal{A}'$ $(n = 1, 2, \ldots, N)$ とし

$$f(\boldsymbol{x}') = \sum_{n=1}^{N} \alpha_n \mathbb{1}_{A'_n}(\boldsymbol{x}')$$

とすると

$$\begin{aligned}
\int_{\mathbb{X}'} f(\boldsymbol{x}')\,\mathrm{d}T(\mu)(\boldsymbol{x}') &= \int_{\mathbb{X}'} \left\{ \sum_{n=1}^{N} \alpha_n \mathbb{1}_{A'_n}(\boldsymbol{x}') \right\} \mathrm{d}T(\mu)(\boldsymbol{x}') \\
&= \sum_{n=1}^{N} \alpha_n \int_{\mathbb{X}'} \mathbb{1}_{A'_n}(\boldsymbol{x}')\,\mathrm{d}T(\mu)(\boldsymbol{x}') \\
&= \sum_{n=1}^{N} \alpha_n \int_{\mathbb{X}'} \mathbb{1}_{T^{-1}(A'_n)}\big(T(\boldsymbol{x})\big)\,\mathrm{d}\mu(\boldsymbol{x}) \\
&\qquad \text{(}\because \text{第 1 段階の結果)} \\
&= \int_{\mathbb{X}} \left\{ \sum_{n=1}^{N} \alpha_n \mathbb{1}_{T^{-1}(A'_n)}\big(T(\boldsymbol{x})\big) \right\} \mathrm{d}\mu(\boldsymbol{x}) \\
&= \int_{\mathbb{X}} \left\{ \sum_{n=1}^{N} \alpha_n f\big(T(\boldsymbol{x})\big) \right\} \mathrm{d}\mu(\boldsymbol{x})
\end{aligned}$$

となり, 単関数 f に対して (C.40) は成立する.
第 3 段階: $f \geq 0$ とし $\int_{\mathbb{X}} f(\boldsymbol{x}')\,\mathrm{d}\mu(\boldsymbol{x}') < \infty$ とする. すると非負値単関数列 $\{f_n\}_{n=1}^{\infty}$ で $f_n \uparrow f$ $(n \to \infty)$ となるものが存在する. このことより

$$\begin{aligned}
\int_{\mathbb{X}'} f(\boldsymbol{x}')\,\mathrm{d}T(\mu)(\boldsymbol{x}') &= \int_{\mathbb{X}'} \lim_{n\to\infty} f_n(\boldsymbol{x}')\,\mathrm{d}T(\mu)(\boldsymbol{x}') \\
&= \lim_{n\to\infty} \int_{\mathbb{X}'} f_n(\boldsymbol{x}')\,\mathrm{d}T(\mu)(\boldsymbol{x}') \qquad (\because \text{単調収束定理}) \\
&= \lim_{n\to\infty} \int_{\mathbb{X}} f_n\big(T(\boldsymbol{x})\big)\,\mathrm{d}\mu(\boldsymbol{x}) \\
&= \int_{\mathbb{X}} \lim_{n\to\infty} f_n\big(T(\boldsymbol{x})\big)\,\mathrm{d}\mu(\boldsymbol{x}) \qquad (\because \text{単調収束定理}) \\
&= \int_{\mathbb{X}} f\big(T(\boldsymbol{x})\big)\,\mathrm{d}\mu(\boldsymbol{x})
\end{aligned}$$

となる. よって $f \geq 0$ のとき (C.40) は成立する.
第 4 段階: $\int_{\mathbb{X}'} |f(\boldsymbol{x}')|\,\mathrm{d}T(\mu)(\boldsymbol{x}') < \infty$ のとき

$$f^{+}(\boldsymbol{x}') = \max\big\{f(\boldsymbol{x}'), 0\big\}, \quad f^{-}(\boldsymbol{x}') = \max\big\{-f(\boldsymbol{x}'), 0\big\}$$

とおく. すると

$$\begin{aligned}\int_{\mathbb{X}'} f(\boldsymbol{x}')\,\mathrm{d}T(\mu)(\boldsymbol{x}') &= \int_{\mathbb{X}'} f^+(\boldsymbol{x}')\,\mathrm{d}T(\mu)(\boldsymbol{x}') - \int_{\mathbb{X}'} f^-(\boldsymbol{x}')\,\mathrm{d}T(\mu)(\boldsymbol{x}') \\ &= \int_{\mathbb{X}} f^+\big(T(\boldsymbol{x})\big)\,\mathrm{d}\mu(\boldsymbol{x}) - \int_{\mathbb{X}} f^-\big(T(\boldsymbol{x})\big)\,\mathrm{d}\mu(\boldsymbol{x}) \\ &\qquad \big(\because \text{第 4 段階の結果}\big)\end{aligned}$$

となり, 補題は証明された. □

補題 C.81. $T(\boldsymbol{x}) = A\boldsymbol{x} + \boldsymbol{b}\,(\boldsymbol{x} \in \mathbb{R}^n)$ とする. ただし A は $n \times n$ の行列で $\det(A) \neq 0$, $\boldsymbol{b} \in \mathbb{R}^n$ である. このとき可測関数 $f : \mathbb{R}^n \to \mathbb{R}$ に対して

$$\int_{T(U)} f(\boldsymbol{y})\,\mathrm{d}\lambda_n(\boldsymbol{y}) = \int_U f\big(A\boldsymbol{x} + \boldsymbol{b}\big)|\det(A)|\,\mathrm{d}\lambda_n(\boldsymbol{x}) \qquad \big(U \subset \mathbb{R}^n \text{ は開集合}\big)$$

が成立する. ただし $\mathbb{R}^n$ 上の Lebeague 速度を λ_n である.

Proof. 補題 C.80 から

$$\begin{aligned}\int_U f\big(A\boldsymbol{x} + \boldsymbol{b}\big)|\det(A)|\,\mathrm{d}\lambda_n(\boldsymbol{x}) &= \int_U f\big(T(\boldsymbol{x})\big)|\det(A)|\,\mathrm{d}\lambda_n(\boldsymbol{x}) \\ &= \int_U f\big(T(\boldsymbol{x})\big)\mathbb{1}_U(\boldsymbol{x})|\det(A)|\,\mathrm{d}\lambda_n(\boldsymbol{x}) \\ &= \int_{\mathbb{R}^n} f\big(T(\boldsymbol{x})\big)\mathbb{1}_{T(U)}\big(T(\boldsymbol{x})\big)|\det A|\,\mathrm{d}\lambda_n(\boldsymbol{x}) \\ &= \int_{\mathbb{R}^n} f(\boldsymbol{y})\mathbb{1}_{T(U)}(\boldsymbol{y})\,\mathrm{d}T(|\det A|\lambda_n)(\boldsymbol{y}) \\ &= \int_{T(U)} f(\boldsymbol{y})\,\mathrm{d}T(|\det A|\lambda_n)(\boldsymbol{y})\end{aligned}$$

を得る. あとは

$$T(|\det A|\lambda_n) = \lambda_n$$

を示せばよい. $T^{-1}(\boldsymbol{y}) = A^{-1}(\boldsymbol{y}) - A^{-1}\boldsymbol{b}$ なので $B \in \mathcal{B}(\mathbb{R}^n)$ に対して

$$\begin{aligned}T(|\det A|\lambda_n)(B) &= |\det A|\lambda_n\big(T^{-1}(B)\big) = |\det A|\lambda_n\big(A^{-1}(B) - A^{-1}(\boldsymbol{b})\big) \\ &= |\det A|\lambda_n\big(A^{-1}(B)\big) \\ &= |\det A| \cdot |\det A^{-1}|\lambda_n(B) \quad \big(\because \text{定理 C.38}\big) = \lambda_n(B)\end{aligned}$$

より補題は証明された. □

定義 C.82. $U, V \subset \mathbb{R}^n$ を開集合とする. 写像 $T : U \to V$ は C^1 級微分同相であるとは次の条件をみたすときをいう.

(1) T は全単射.

(2) T と T^{-1} は連続微分可能.

点 $\boldsymbol{x} \in U$ における微分

$$DT(\boldsymbol{x}) := \left(\frac{\partial}{\partial x_j} T_j(\boldsymbol{x})\right)_{i,j=1}^{n}$$

を**ヤコビアン** (**Jacobian**) という．ただし $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, $T = (T_1, T_2, \ldots, T_n)$ である.

定理 C.83. (Rudin (1976, pp.221-228)) $U \subset \mathbb{R}^n$ を開集合とする．写像

$$T : U \to V = T(U)$$

が C^1 級微分同相であるための必要十分条件は次の (a), (b), (c) すべてが成立することである.

(a) $T : U \to \mathbb{R}^n$ は単射.

(b) $T : U \to \mathbb{R}^n$ は連続微分可能.

(c) $DT(\boldsymbol{x})$ は各 $\boldsymbol{x} \in U$ で可逆.

さらに (a), (b), (c) が成り立つとき

$$D(T^{-1})(\boldsymbol{y}) = (DT)^{-1}\bigl(T^{-1}(\boldsymbol{y})\bigr) \qquad (\forall \boldsymbol{y} \in V)$$

が成り立つ.

Proof. 証明は定理 A.20 のところをみよ. □

定理 C.84. U, V を $\mathbb{R}^n$ の開集合とし, $T : U \to V$ は C^1 級微分同相写像とする．$W \subset \mathbb{R}^n$ と可測空間 $(W, \mathcal{B}(W))$ 上の Lebeague 測度を λ_W を

$$\lambda_W(A) = \lambda_n(A \cap W) \qquad (A \in \mathcal{B}(\mathbb{R}^n)) \tag{C.41}$$

で定める．ただし λ_n は $\mathbb{R}^n$ 上の Lebeague 速度である．このとき

$$\lambda_V = T\bigl(|\det DT(\,\cdot\,)|\lambda_U\bigr) \tag{C.42}$$

となる．すなわち $f : V \to [0, \infty)$ に対して

$$\int_V f(\boldsymbol{y})\,\mathrm{d}\lambda_V(\boldsymbol{y}) = \int_U f\bigl(T(\boldsymbol{x})\bigr)|\det DT(\boldsymbol{x})|\,\mathrm{d}\lambda_U(\boldsymbol{x}) \tag{C.43}$$

である.

定理 C.84 の証明はいくつかの段階を必要とする. $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ に対して

$$|\boldsymbol{x}|_{\ell^\infty(n)} := \max\bigl\{|x_1|, |x_2|, \ldots, |x_n|\bigr\}$$

と定める. さらに

$$\mathcal{D} := \{(a_1, b_1] \times (a_2, b_2] \times \cdots (a_n, b_n] \subset \mathbb{R}^n;\, a_i < b_i \in \mathbb{R}\,(i = 1, 2, \ldots, n)\}$$

とする.

補題 C.85. *$d \in \mathbb{N}$ とし, $T = (T_1, T_2, \ldots, T_d) : \mathbb{R}^n \to \mathbb{R}^d$ を Lipschitz 写像とする. すなわち $\exists L \geq 0$ があって*

$$|T(\boldsymbol{x}) - T(\boldsymbol{y})|_{\ell^\infty(d)} \leq L|\boldsymbol{x} - \boldsymbol{y}|_{\ell^\infty(n)} \qquad (\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n)$$

である. さらに $\forall \boldsymbol{c} = (c_1, c_2, \ldots, c_n) \in \mathbb{R}^n$ と $s > 0$ に対して

$$D = [c_1 - s, c_1 + s] \times [c_2 - s, c_2 + s] \times \cdots \times [c_n - s, c_n + s]$$

とする. このとき

$$T(D) \subset E = [\tilde{a}_1, \tilde{b}_1] \times [\tilde{a}_2, \tilde{b}_2] \times \cdots \times [\tilde{a}_d, \tilde{b}_d]$$

となる. ただし

$$\tilde{a}_i = T_i(\boldsymbol{c}) - Ls, \quad \tilde{b}_i = T_i(\boldsymbol{c}) + Ls\,(i = 1, 2, \ldots, \mathrm{d})$$

である. 特に

$$\lambda_d(T(D)) \leq L^d\bigl(\lambda_n(D)\bigr)^{d/n}$$

となる. ただし λ_n は $\mathbb{R}^n$ 上の Lebeague 測度である.

Proof. $D = \{\boldsymbol{y} \in \mathbb{R}^n;\, |\boldsymbol{c} - \boldsymbol{y}|_{\ell^\infty(n)} \leq s\}$ と書けることに注意する. Lipschitz 連続性から

$$T\bigl(\{\boldsymbol{y} \in \mathbb{R}^n;\, |\boldsymbol{c} - \boldsymbol{y}|_{\ell^\infty(n)} \leq s\}\bigr) \subset \{T(\boldsymbol{y});\, |T(\boldsymbol{c}) - T(\boldsymbol{y})|_{\ell^\infty(d)} \leq Ls\} = E$$

となる. D の一辺は $2s$, E の一辺は $2Ls$ であるので

$$\lambda_d\bigl(T(D)\bigr) \leq \lambda_d(E) = (2Ls)^d = L^d\bigl((2s)^n\bigr)^{d/n} = L^d\bigl(\lambda_n(D)\bigr)^{d/n}$$

を得る. □

補題 C.86. $\mathbb{R}$ 上の半環 $\mathcal{D}$ を

$\mathcal{D} := \{(a_1,\, b_1] \times (a_2,\, b_2] \times \cdots \times (a_n,\, b_n] \subset \mathbb{R}^n;\, a_i < b_i \in \mathbb{R}\,(i = 1,\, 2,\, \ldots,\, n)\}$

で定義し, $\mathcal{A} = \sigma[\mathcal{D}]$ とする. さらに μ と ν を可測空間 $(\mathbb{R},\, \mathcal{A})$ 上の測度とし

$$\mu(D) \leq \nu(D) < \infty \qquad (\forall D \in \mathcal{D})$$

で $\{D_n\}_{n=1}^{\infty} \subset \mathcal{D}$ があって $\bigcup_{n=1}^{\infty} D_n = \mathbb{R}$ とする. このとき

$$\mu(A) \leq \nu(A) \qquad (\forall A \in \mathcal{A})$$

が成り立つ.

Proof. μ と ν の性質から

$$\rho := \mu - \nu :\, \mathcal{D} \to [0,\, \infty)$$

となる. ρ は $\mathcal{D}$ 上で σ 加法的なので, ρ は $\mathcal{A}$ 上の測度 $\tilde{\rho}$ に一意的に拡張できる. さらに

$$\nu(D) = (\rho + \mu)(D) = \widetilde{\rho + \mu}(D) \qquad (\forall D \in \mathcal{D})$$

となる. ただし $\widetilde{\rho + \mu}$ は $\rho + \mu$ の $\mathcal{A} = \sigma[\mathcal{D}]$ 上への拡張である. しかし

$$\nu(D) = \widetilde{\rho + \mu}(D) = (\rho + \mu)(D) = \tilde{\rho}(D) + \mu(D) \qquad (\forall D \in \mathcal{D})$$

となる. $\mathcal{A}$ 上への拡張は一意なので

$$\begin{aligned} &\nu(A) = \tilde{\rho}(A) + \mu(A) \quad (\forall A \in \mathcal{A}) \\ &\qquad \Leftrightarrow l\nu(A) - \mu(A) = \tilde{\rho}(A) \geq 0 \quad (\forall A \in \mathcal{A}) \end{aligned}$$

となり

$$\nu(A) \geq \mu(A) \qquad (\forall A \in \mathcal{A})$$

がわかる. □

補題 C.87. $U,\, V$ を $\mathbb{R}^n$ の開集合とし, $T : U \to V$ を C^1 級微分同相写像とする. このとき

$$\lambda_V\bigl(T(D)\bigr) \leq \int_D |\det DT(\boldsymbol{x})|\, \mathrm{d}\lambda_V(\boldsymbol{x})$$

である. ただし $D \in \mathcal{D}$ で $\overline{D} \in V$ であり, λ_U と λ_V は (C.41) により定まる U と V 上の Lebeague 測度である.

Proof. $a_i,\, b_i \in \mathbb{R}\,(i = 1,\, 2,\, \ldots,\, n)$ に対して $D = [a_1,\, b_1) \times [a_2,\, b_2) \times \cdots \times [a_n,\, b_n)$ と書いたとき $\overline{D} = [a_1,\, b_1] \times [a_2,\, b_2] \times \cdots \times [a_n,\, b_n] \subset V$ である．また $A \in \mathsf{M}(n,\, \mathbb{R})$ に対して

$$\|A\| := \sup_{|\boldsymbol{x}|_{\ell^\infty(n)} \le 1} |A\boldsymbol{x}|_{\ell^\infty(n)}$$

と定める．逆写像定理 (定理 C.83) を用いると

$$\begin{aligned} L := \sup_{\boldsymbol{x} \in \overline{D}} \|(DT)^{-1}(\boldsymbol{x})\| &= \sup_{\boldsymbol{y} \in T(\overline{D})} \|(DT)^{-1}\big(T^{-1}(\boldsymbol{y})\big)\| \\ &= \sup_{\boldsymbol{y} \in T(\overline{D})} \|D\big(T^{-1}\big)(\boldsymbol{y})\| \end{aligned} \tag{C.44}$$

となる．DT は $\overline{D} \subset U$ 上で一様連続なので $\forall \epsilon > 0$ に対してある $\delta > 0$ があって

$$\sup_{\boldsymbol{x},\, \boldsymbol{x}' \in \overline{D};\, |\boldsymbol{x} - \boldsymbol{x}'|_{\ell^\infty(n)} \le \delta} \|DT(\boldsymbol{x}) - DT(\boldsymbol{x}')\| \le \frac{\epsilon}{L} \tag{C.45}$$

とできる．いま $D_i\,(i = 1,\, 2,\, \ldots,\, N)$ は一辺の長さが $\delta > 0$ 以下の $\mathcal{D}$ のとする．すなわち

$$\begin{gathered} D_j = [a_{i1},\, a_{i1} + \delta] \times [a_{i2},\, a_{i2} + \delta] \times \cdots \times [a_{in},\, a_{in} + \delta], \\ a_{ij} \in \mathbb{R}\,(i = 1,\, 2,\, \ldots,\, N;\, j = 1,\, 2,\, \ldots,\, n), \end{gathered}$$

とし，$D_k \bigcap D_\ell = \varnothing\,(k \ne \ell)$ とする．さらに D は $\{D \cap D_j\}_{j=1}^N$ で分割されるとする．

$$D \subset \bigcup_{n=1}^N D_n$$

とする．DT と $\det DT$ は連続関数で各 $i = 1,\, 2,\, \ldots,\, N$ に対して $\boldsymbol{x}_i \in \overline{D}_i$ を

$$\|\det DT(\boldsymbol{x}_i)\| = \inf_{\boldsymbol{y} \in D_i} \|DT(\boldsymbol{y})\| \tag{C.46}$$

と定める．$A_i := DT(\boldsymbol{x}_i)\,(i = 1,\, 2,\, \ldots,\, N)$ は可逆で写像 $\boldsymbol{x} \mapsto A_i\boldsymbol{x}$ は

$$\begin{aligned} D\big(A_i^{-1} \circ T\big)(\boldsymbol{x}) &= A_i^{-1} \circ \big(DT\big)(\boldsymbol{x}) \quad (\because D \text{ の線型性}) \\ &= \mathsf{id}_n + A_i^{-1} \circ \big[DT(\boldsymbol{x}) - DT(\boldsymbol{x}_i)\big] \end{aligned} \tag{C.47}$$

をみたす．ただし id_n は $\mathbb{R}^n$ から $\mathbb{R}^n$ への恒等写像である．(C.44) – (C.47) から

$$\begin{aligned} \sup_{\boldsymbol{x} \in \overline{D}} \|D\big(A_i^{-1} \circ T\big)(\boldsymbol{x})\| &\le \|\mathsf{id}_n\| + \sup_{\boldsymbol{x} \in \overline{D}_i} \Big\|A_i^{-1} \circ \big[DT(\boldsymbol{x}) - DT(\boldsymbol{x}_i)\big]\Big\| \\ &\le 1 + L\frac{\epsilon}{L} = 1 + \epsilon \end{aligned} \tag{C.48}$$

となる．なぜならば

$$
\begin{aligned}
&\|A_i^{-1} \circ \big[DT(\boldsymbol{x}) - DT(\boldsymbol{x}_i)\big]\| \\
&\quad = \|(DT)^{-1}(\boldsymbol{x}_i)\big[DT(\boldsymbol{x}) - DT(\boldsymbol{x}_i)\big]\| \\
&\quad \le \|(DT)^{-1}(\boldsymbol{x}_i)\| \times \|DT(\boldsymbol{x}) - DT(\boldsymbol{x}_i)\| \\
&\quad \le \sup_{\boldsymbol{x}\in\overline{D}} \|(DT)^{-1}(\boldsymbol{x})\| \times \sup_{\boldsymbol{x},\boldsymbol{x}'\in\overline{D};\, |\boldsymbol{x}-\boldsymbol{x}'|_{\ell^\infty(n)}\le\delta} \|DT(\boldsymbol{x}) - DT(\boldsymbol{x}')\| \\
&\quad \le L \times \frac{\epsilon}{L}
\end{aligned}
$$

からわかる．さらに中間値の定理から $\boldsymbol{\xi} \in \overline{D}$ があって

$$
\begin{aligned}
\big|A_i^{-1} \circ T(\boldsymbol{x}) - A_i^{-1} \circ T(\boldsymbol{x}')\big|_{\ell^\infty(n)} &= \big|\big(D(A_i^{-1} \circ T)(\boldsymbol{\xi})\big)(\boldsymbol{x} - \boldsymbol{x}')\big|_{\ell^\infty(n)} \\
&\le \big\|D\big(A_i^{-1} \circ T)(\boldsymbol{\xi})\big\| \times |\boldsymbol{x} - \boldsymbol{x}'|_{\ell^\infty(n)} \\
&\le \sup_{\boldsymbol{x}\in\overline{D}} \big\|D\big(A_i^{-1} \circ T)(\boldsymbol{\xi})\big\| \times |\boldsymbol{x} - \boldsymbol{x}'|_{\ell^\infty(n)} \\
&\le (1+\epsilon)|\boldsymbol{x} - \boldsymbol{x}'|_{\ell^\infty(n)}
\end{aligned}
$$

となるので写像 $\boldsymbol{x} \mapsto A_i^{-1} \circ T$ は Lipschitz 連続である．よって補題 C.81 と補題 C.85 から

$$
\begin{aligned}
\lambda_V\big(T(D_i)\big) &= \lambda_V\big(A_i \circ A_i^{-1} \circ T(D_i)\big) \quad \big(\because \text{補題 C.81}\big) \\
&= |\det A_i| \lambda_{A_i^{-1}V}\big(A_i^{-1} \circ T(D_i)\big) \quad \big(\because \text{補題 C.85}\big) \\
&\le |\det A_i| (1+\epsilon)^n \lambda_U(D_i)
\end{aligned}
$$

となる．$D = \bigcup_{i=1}^N D_i$ と $|\det A_i| \le |\det DT(\boldsymbol{x}_i)|$ $(\boldsymbol{x} \in D_i)$ なので

$$
\begin{aligned}
\lambda_V\big(T(D)\big) &\le \sum_{i=1}^N \lambda_V\big(T(D_i)\big) \\
&\le (1+\epsilon)^n \sum_{i=1}^N \int_{D_i} |\det A_i| \, \mathrm{d}\lambda_U(\boldsymbol{x}) \\
&\le (1+\epsilon)^n \int_D |\det DT(\boldsymbol{x})| \, \mathrm{d}\lambda_U(\boldsymbol{x})
\end{aligned}
$$

となる．$\epsilon \searrow 0$ とすれば補題は証明される． □

定理 C.84 の証明: まず (C.42) $\Leftrightarrow$ (C.43) を証明する．$B \in \mathcal{B}(\mathbb{R}^n \cap V)$ と $f = \mathbb{1}_B$ とおく．このとき (C.43) は (C.42) となる．一方 (C.43) は標準機械 (standard machine) を用いると (C.42) から導くことができる．
第 2 段階: $S := T^{-1}$ とおくと

$$\mu := \lambda_V \circ T = \lambda_V \circ S^{-1} = S(\lambda_V)$$

となる．一方 $B \in \mathcal{B}(\mathbb{R}^n \cap U)$ に対して

$$\nu(B) := \int_{B\cap U} |\det DT(\boldsymbol{x})| \,\mathrm{d}\lambda_U(\boldsymbol{x})$$

は密度 $|\det DT(\boldsymbol{x})|$ を持つ測度となる．補題 C.87 から

$$\mu(D) \le \nu(D) \qquad (\forall D \in \mathcal{D})$$

である．よって補題 C.86 から

$$\mu(A) \le \nu(A) \qquad (A \in \mathcal{B}(\mathbb{R}^n))$$

となる．第 1 段階の結果から

$$\int f(\boldsymbol{y}) \,\mathrm{d}\lambda_V(\boldsymbol{y}) \le \int f\bigl(T(\boldsymbol{x})\bigr) |\det DT(\boldsymbol{x})| \,\mathrm{d}\lambda_U(\boldsymbol{x}) \tag{C.49}$$

となる．
<u>第 3 段階</u>: 逆向きの不等号を示すために $S = T^{-1}$ は C^1 級微分同相であることを使う．(C.49) において $U \longleftrightarrow V$ と $\boldsymbol{x} \leftrightarrow \boldsymbol{y}$ と交換すると

$$\int f(\boldsymbol{x}) \,\mathrm{d}\lambda_U(\boldsymbol{x}) \le \int f\bigl(S(\boldsymbol{y})\bigr) |\det DS(\boldsymbol{y})| \,\mathrm{d}\lambda_V(\boldsymbol{y}) \tag{C.50}$$

となる．ここで $A \in \mathcal{B}(\mathbb{R}^n \cap U)$ を取り

$$f(\boldsymbol{x}) = \mathbb{1}_{T(A)} \circ T(\boldsymbol{x}) |\det DT(\boldsymbol{x})|$$

とおく．このとき

$$\begin{aligned}
\int f(\boldsymbol{x}) \,\mathrm{d}\lambda_U(\boldsymbol{x}) &= \int \mathbb{1}_{T(A)}\bigl(T(\boldsymbol{x})\bigr) |\det DT(\boldsymbol{x})| \,\mathrm{d}\lambda_U(\boldsymbol{x}) \\
&\le \int \bigl(\mathbb{1}_{T(A)} \circ T(\boldsymbol{x}) |\det DT(\boldsymbol{x})|\bigr) \circ S(\boldsymbol{y}) |\det DS(\boldsymbol{y})| \,\mathrm{d}\lambda_V(\boldsymbol{y}) \\
&\quad \bigl(\because \text{(C.50)}\bigr) \\
&\le \int \bigl(\mathbb{1}_{T(A)}\bigl(T \circ S(\boldsymbol{y})\bigr) |\det(DT) \circ S(\boldsymbol{y})| \times |\det DS(\boldsymbol{y})| \,\mathrm{d}\lambda_V(\boldsymbol{y}) \\
&\quad \bigl(\because \boldsymbol{x} = S(\boldsymbol{y}) \text{ を代入}\bigr) \\
&= \int \mathbb{1}_{T(A)}(\boldsymbol{y}) \Bigl|\det\Bigl[\underbrace{(DT) \circ S(\boldsymbol{y}) \times DS(\boldsymbol{y})}_{=\mathsf{id}_n = D\left(T\circ S\right)(\boldsymbol{y}) = (DT)\left(S(\boldsymbol{y})\right) \times DS}\Bigr]\Bigr| \,\mathrm{d}\lambda_V(\boldsymbol{y}) \\
&= \int \mathbb{1}_{T(A)}(\boldsymbol{y}) \,\mathrm{d}\lambda_V(\boldsymbol{y}) \\
&= \lambda_V(T(A))
\end{aligned}$$

となる．あとは標準機械 (standard machine) を用いると

$$\int f\bigl(T(\boldsymbol{x})\bigr)|\det DT(\boldsymbol{x})|\,\mathrm{d}\lambda_V(\boldsymbol{x}) \le \int f(\boldsymbol{y})\,\mathrm{d}\lambda_V(\boldsymbol{y}) \tag{C.51}$$

を得る．(C.49) と (C.51) から

$$\int f(\boldsymbol{y})\,\mathrm{d}\lambda_V(\boldsymbol{y}) = \int f\bigl(T(\boldsymbol{x})\bigr)|\det DT(\boldsymbol{x})|\,\mathrm{d}\lambda_U(\boldsymbol{x})$$

がわかる． □

C.9　条件付き期待値: 再訪問

$(\Omega, \mathcal{A}, \mathsf{Pr})$ を確率空間とし，$\mathcal{F} \subset \mathcal{A}$ も σ 加法族とする．X は $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率変数で $\mathsf{E}\bigl[|X|\bigr] < \infty$ とする．

定義 C.88. $\mathcal{F}$ を与えたときの X の**条件付き期待値**を次の条件をみたす任意の確率変数 Y とする．

(i) Y は $\mathcal{F}$ 可測．

(ii) 任意の $A \in \mathcal{F}$ に対して

$$\int_A X(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) = \int_A Y(\omega)\,\mathrm{d}\mathsf{Pr}(\omega).$$

まず条件付き期待値の存在と一意性を確認する．

補題 C.89. Y を定義 C.88 の条件 (i)(ii) をみたすとき Y は可積である．

Proof. $A := \{\omega \in \Omega; Y(\omega) > 0\}$ とおくと，A, $A^c \in \mathcal{F}$ となる．(ii) を用いると

$$\int_A Y(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) = \int_A X(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) \le \int_A |X(\omega)|\,\mathrm{d}\mathsf{Pr}(\omega),$$
$$\int_{A^c} \bigl(-Y(\omega)\bigr)\,\mathrm{d}\mathsf{Pr}(\omega) = \int_{A^c} \bigl(-X(\omega)\bigr)\,\mathrm{d}\mathsf{Pr}(\omega) \le \int_{A^c} |X(\omega)|\,\mathrm{d}\mathsf{Pr}(\omega).$$

よって

$$\begin{aligned}\int |Y(\omega)|\,\mathrm{d}\mathsf{Pr}(\omega) &= \int_A Y(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) + \int_{A^c} \bigl(-Y(\omega)\bigr)\,\mathrm{d}\mathsf{Pr}(\omega)\\ &\le \int_A |X(\omega)|\,\mathrm{d}\mathsf{Pr}(\omega) + \int_{A^c} |X(\omega)|\,\mathrm{d}\mathsf{Pr}(\omega)\\ &= \int |X(\omega)|\,\mathrm{d}\mathsf{Pr}(\omega)\end{aligned}$$

からわかる． □

補題 C.90. Y は一意.

Proof. Y' も定義 C.88 の条件 (i)(ii) をみたすとする. すると (ii) より

$$\int_A Y(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) = \int_A X(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) = \int_A Y'(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) \qquad (\forall A \in \mathcal{A}).$$

$\epsilon > 0$ を取り, $A := \{\omega \in \Omega;\, Y(\omega) - Y'(\omega) \geq \epsilon\}$ とする. $A \in \mathcal{F}$ であることから (ii) から

$$\begin{aligned} 0 &= \int_A X(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) - \int_A X(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) \\ &= \int_A Y(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) - \int_A Y'(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) \\ &= \int_A \big\{Y(\omega) - Y'(\omega)\big\}\,\mathrm{d}\mathsf{Pr}(\omega) \\ &\geq \epsilon \int_A \mathrm{d}\mathsf{Pr}(\omega) = \epsilon \mathsf{Pr}(A) \end{aligned}$$

となり, $\mathsf{Pr}(A) = 0$ がわかる.

$$\Big\{\omega \in \Omega;\, Y(\omega) - Y'(\omega) > 0\Big\} = \bigcup_{n=1}^{\infty} \Big\{\omega \in \Omega;\, Y(\omega) - Y'(\omega) \geq \frac{1}{n}\Big\}$$

と表現できることとすべての $\epsilon > 0$ に対して $\mathsf{Pr}(A) = 0$ なので

$$\begin{aligned} \mathsf{Pr}\bigg(\Big\{\omega \in \Omega;\, Y(\omega) - Y'(\omega) > 0\Big\}\bigg) &= \mathsf{Pr}\bigg(\bigcup_{n=1}^{\infty} \Big\{\omega \in \Omega;\, Y(\omega) - Y'(\omega) \geq \frac{1}{n}\Big\}\bigg) \\ &= \lim_{n\to\infty} \mathsf{Pr}\bigg(\Big\{\omega \in \Omega;\, Y(\omega) - Y'(\omega) \geq \frac{1}{n}\Big\}\bigg) \\ &= 0 \end{aligned}$$

がわかる. よって

$$\mathsf{Pr}\big(\omega \in \Omega;\, Y(\omega) \leq Y'(\omega)\big) = 1$$

となる. 次に $A := \{\omega \in \Omega;\, Y'(\omega) - Y(\omega) \geq \epsilon\}$ として, 同じ議論を行えば

$$\mathsf{Pr}\big(\omega \in \Omega;\, Y(\omega) \geq Y'(\omega)\big) = 1.$$

よって

$$\mathsf{Pr}\big(\omega \in \Omega;\, Y(\omega) = Y'(\omega)\big) = 1$$

となる. □

定義 C.91. 定義 C.88 で定めた Y を

$$Y(\omega) = \mathsf{E}\big[X|\,\mathcal{F}\big](\omega)$$

と記すことにする.

補題 C.92. $\mathsf{E}\big[X|\,\mathcal{F}\big](\omega)$ は存在する.

Proof. $\mu = \mathsf{Pr}$ と書くことにする. まず $X \geq 0$ とする. $\mathcal{F}$ 上の測度 ν を

$$\nu(A) = \int_A X(\omega)\,\mathrm{d}\mu(\omega) \qquad (A \in \mathcal{F})$$

と定める. 単調収束定理を用いると ν は $\mathcal{F}$ 上の測度となる. さらに $\nu \ll \mu$ である. よって Radon-Nikodym の定理より, $\mathcal{F}$ 可測関数 $\dfrac{\mathrm{d}\nu}{\mathrm{d}\mu}$ が存在し

$$\begin{aligned}\int_A \mathsf{E}\big[X|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mu(\omega) &= \int_A X(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) = \nu(A)\\ &= \int_A \frac{\mathrm{d}\nu}{\mathrm{d}\mu}(\omega)\,\mathrm{d}\mu(\omega) \qquad (A \in \mathcal{F})\end{aligned} \tag{C.52}$$

となる. $A = \Omega$ とすると $\dfrac{\mathrm{d}\nu}{\mathrm{d}\mu} \geq 0$ は可積となる. Radon-Nikodym の一意性から

$$\mathsf{E}\big[X|\,\mathcal{F}\big](\omega) = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}(\omega)$$

となる.

一般の場合については

$$\begin{aligned}&X(\omega) = X^+(\omega) - X^-(\omega),\\ &X^+(\omega) = \max\{X(\omega),\,0\}, \quad X^-(\omega) = \max\{-X(\omega),\,0\}\end{aligned}$$

とする. $Y_1(\omega) = \mathsf{E}\big[X^+|\,\mathcal{F}\big](\omega)$ と $Y_2(\omega) = \mathsf{E}\big[X^{-1}|\,\mathcal{F}\big](\omega)$ とおくと $Y_1 - Y_2$ は $\mathcal{F}$ 可測である. さらに $\forall A \in \mathcal{F}$ に対して

$$\begin{aligned}\int_A X(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) &= \int_A X^+(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) - \int_A X^-(\omega)\,\mathrm{d}\mathsf{Pr}(\omega)\\ &= \int_A Y_1(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) - \int_A Y_2(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) \qquad \big(\because \text{(C.52)}\big)\\ &= \int_A \left\{Y_1(\omega) - \int_A Y_2(\omega)\right\}\mathrm{d}\mathsf{Pr}(\omega)\\ &= \int_A \big(Y_1 - Y_2\big)(\omega)\,\mathrm{d}\mathsf{Pr}(\omega)\end{aligned}$$

となる. よって $\big(Y_1 - Y_2\big)(\omega) = \mathsf{E}\big[X|\,\mathcal{F}\big](\omega)$ がわかる. □

注意 C.93. 以下の記法を導入する．
(1) $A \in \mathcal{A}$ に対して

$$\mathsf{Pr}\big(A|\,\mathcal{F}\big)(\omega) := \mathsf{E}\big[\mathbb{1}_A|\,\mathcal{F}\big](\omega)$$

と定める．
(2) $A,\, B \in \mathcal{A}$ に対して

$$\mathsf{Pr}\big(A|\,B\big) := \frac{\mathsf{Pr}\big(A \cap B\big)}{\mathsf{Pr}\big(B\big)}$$

と定める．
(3) X と Y を $(\Omega,\, \mathcal{A},\, \mathsf{Pr})$ 上の確率変数としたとき

$$\mathsf{E}\big[X|\,Y\big](\omega) := \mathsf{E}\big[X|\,\sigma(Y)\big](\omega)$$

とする．ただし $\sigma(Y)$ は Y によって生成された σ 加法族である．すなわち

$$\sigma(Y) := \sigma\Big[\big\{\omega \in \Omega;\, Y(\omega) \le r\,(\forall r \in \mathbb{R})\big\}\Big]$$

である．
(4) X と Y を $(\Omega,\, \mathcal{A},\, \mathsf{Pr})$ 上の連続型確率変数とする．さらに $(X,\, Y)$ は同時 p.d.f. $\mathsf{p}(x,\, y)$ を持つとする．議論を簡単にするために $\forall y \in \mathbb{R}$ に対して

$$\int \mathsf{p}(x,\, y)\,\mathrm{d}x > 0$$

とする．
　いま関数 $g : \mathbb{R} \to \mathbb{R}$ で $\mathsf{E}\big[|g(X)|\big] < \infty$ なるものと考える．このとき

$$\mathsf{E}\big[g(X)|\,Y\big](\omega) = h(Y(\omega)), \qquad h(y) = \frac{\int g(x)\mathsf{p}(x,\, y)\,\mathrm{d}x}{\int \mathsf{p}(x,\, y)\,\mathrm{d}x}$$

となる．これを示すために $A \in \sigma(Y)$ を取る．このときある $B \in \mathcal{B}(\mathbb{R})$ が存在して

$$A := \big\{\omega \in \Omega;\, Y(\omega) \in B\big\}$$

と書ける．したがって Fubini の定理から

$$
\begin{aligned}
\mathsf{E}\big[h(Y)\mathbb{1}_A\big] &= \mathsf{E}\big[h(Y)\mathbb{1}_B(Y)\big] \qquad \big(\because Y \in B \Leftrightarrow \omega \in A\big) \\
&= \int_B \int h(y)p(x,\,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_B \int \left\{ \frac{\int g(x)p(x,\,y)\,\mathrm{d}x}{\int p(x,\,y)\,\mathrm{d}x} \right\} p(x,\,y)\,\mathrm{d}x\,\mathrm{d}y \qquad \big(\because h \text{ の定義を代入}\big) \\
&= \int_B \left\{ \int \frac{\int g(x)p(x,\,y)\,\mathrm{d}x}{\int p(x,\,y)\,\mathrm{d}x} p(x,\,y)\,\mathrm{d}x \right\} \mathrm{d}y \qquad \big(\because \text{Fubini の定理}\big) \\
&= \int_B \frac{\int g(x)p(x,\,y)\,\mathrm{d}x}{\int p(x,\,y)\,\mathrm{d}x} \left\{ \int p(x,\,y)\,\mathrm{d}x \right\} \mathrm{d}y \\
&= \int_B \int g(x)p(x,\,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \mathsf{E}\big[g(X)\mathbb{1}_A\big] \qquad \big(\because Y \in B \Leftrightarrow \omega \in A\big)
\end{aligned}
$$

となる．よって

$$
\int_A h\big(Y(\omega)\big)\,\mathrm{d}\mathsf{Pr}(\omega) = \int_A g\big(X(\omega)\big)\,\mathrm{d}\mathsf{Pr}(\omega)
$$

なので

$$
\mathsf{E}\big[g(X)\,|\,\mathcal{F}\big](\omega) = h\big(Y(\omega)\big)
$$

がわかる． □

C.10　条件付き期待値の性質

定理 C.94. X, $\{X_n\}_{n=1}^{\infty}$, Y を確率空間 $(\Omega,\,\mathcal{A},\,\mathsf{Pr})$ 上の確率変数列とする．
(1) $\mathsf{E}\big[|X|\big] < \infty$ かつ $\mathsf{E}\big[|Y|\big] < \infty$ とする．このとき $a,\,b \in \mathbb{R}$ に対して

$$
\mathsf{E}\big[aX + bY\,|\,\mathcal{F}\big](\omega) = a\mathsf{E}\big[X\,|\,\mathcal{F}\big](\omega) + b\mathsf{E}\big[Y\,|\,\mathcal{F}\big](\omega)
$$

と成り立つ．
(2) さらに $X \leq Y$ のとき

$$
\mathsf{E}\big[X\,|\,\mathcal{F}\big](\omega) \leq \mathsf{E}\big[Y\,|\,\mathcal{F}\big](\omega)
$$

が成り立つ．
(3) $X_n \geq 0\,(n = 1,\,2,\,\ldots)$ かつ $X_n \uparrow X\,(n \to \infty)$ で $\mathsf{E}\big[X\big] < \infty$ のとき

$$
\mathsf{E}\big[X_n\,|\,\mathcal{F}\big](\omega) \uparrow \mathsf{E}\big[Y\,|\,\mathcal{F}\big](\omega) \quad (n \to \infty)
$$

が成り立つ．

Proof. (1) あきらかに $a\mathsf{E}\big[X|\,\mathcal{F}\big] + b\mathsf{E}\big[Y|\,\mathcal{F}\big]$ は $\mathcal{F}$ 可測である. $\forall A \in \mathcal{F}$ に対して

$$\begin{aligned}
&\int_A \big\{a\mathsf{E}\big[X|\,\mathcal{F}\big](\omega) + b\mathsf{E}\big[Y|\,\mathcal{F}\big](\omega)\big\}\,\mathrm{d}\mathsf{Pr}(\omega)\\
&\qquad = a\int_A \mathsf{E}\big[X|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) + b\int_A \mathsf{E}\big[Y|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega)\\
&\qquad = a\int_A X(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) + b\int_A Y(\omega)\,\mathrm{d}\mathsf{Pr}(\omega)\\
&\quad = \int_A \big(aX + bY\big)(\omega)\,\mathrm{d}\mathsf{Pr}(\omega)
\end{aligned}$$

を得る.

(2) 条件付き期待値の定義より $\forall A \in \mathcal{F}$ に対して

$$\begin{aligned}
\int_A \mathsf{E}\big[X|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) &= \int_A X(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) \le \int_A Y(\omega)\,\mathrm{d}\mathsf{Pr}(\omega)\\
&= \int_A \mathsf{E}\big[Y|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega)
\end{aligned}$$

を得る. $\forall \epsilon > 0$ に対して

$$A := \big\{\omega \in \Omega;\ \mathsf{E}\big[X|\,\mathcal{F}\big](\omega) - \mathsf{E}\big[Y|\,\mathcal{F}\big](\omega) \ge \epsilon\big\}$$

とおくと $\mathsf{Pr}(A) = 0$ となる. ϵ は任意だったので

$$\begin{aligned}
&\Big\{\omega \in \Omega;\ \mathsf{E}\big[X|\,\mathcal{F}\big](\omega) > \mathsf{E}\big[Y|\,\mathcal{F}\big](\omega)\Big\}\\
&\qquad = \bigcup_{n=1}^{\infty}\left\{\omega \in \Omega;\ \mathsf{E}\big[X|\,\mathcal{F}\big](\omega) - \mathsf{E}\big[Y|\,\mathcal{F}\big](\omega) \ge \frac{1}{n}\right\}
\end{aligned}$$

となる. よって

$$\begin{aligned}
&\mathsf{Pr}\Big(\Big\{\omega \in \Omega;\ \mathsf{E}\big[X|\,\mathcal{F}\big](\omega) > \mathsf{E}\big[Y|\,\mathcal{F}\big](\omega)\Big\}\Big)\\
&\qquad = \lim_{n\to\infty}\mathsf{Pr}\left(\left\{\omega \in \Omega;\ \mathsf{E}\big[X|\,\mathcal{F}\big](\omega) - \mathsf{E}\big[Y|\,\mathcal{F}\big](\omega) \ge \frac{1}{n}\right\}\right) = 0
\end{aligned}$$

となるから

$$\mathsf{Pr}\Big(\omega \in \Omega;\ \mathsf{E}\big[X|\,\mathcal{F}\big](\omega) \le \mathsf{E}\big[Y|\,\mathcal{F}\big](\omega)\Big) = 1$$

となる.

(3) $Y_n := X - X_n\,(n \in \mathbb{N})$ とおくと $\mathsf{E}\big[Y_n|\,\mathcal{F}\big] \downarrow 0\,(n \to \infty)$ を示せばよい. $Y_n \downarrow 0\,(n \to \infty)$ なので (2) の結果から

$$Z_n(\omega) := \mathsf{E}\big[Y_n|\,\mathcal{F}\big](\omega)$$

はほとんど確実に単調減少列なので

$$Z_\infty(\omega) := \lim_{n\to\infty} \mathsf{E}\big[Y_n \big| \mathcal{F}\big](\omega)$$

はほとんど確実に存在する. $A \in \mathcal{F}$ に対して

$$\int_A Z_n(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) = \int_A Y_n(\omega)\,\mathrm{d}\mathsf{Pr}(\omega)$$

である. $n \to \infty$ のとき $Y_n(\omega) \downarrow 0\,(n \to \infty)$ なので, 単調収束定理から

$$\int_A Z_\infty(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) = 0 \qquad (\forall A \in \mathcal{F})$$

となる. よって $Z_\infty \equiv 0$ となる. □

定理 C.95. $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{A}$ とする. このとい

(1) $\mathsf{E}\big[\mathsf{E}\big[X \big| \mathcal{F}_1\big] \big| \mathcal{F}_2\big](\omega) = \mathsf{E}\big[X \big| \mathcal{F}_1\big](\omega)$.

(2) $\mathsf{E}\big[\mathsf{E}\big[X \big| \mathcal{F}_2\big] \big| \mathcal{F}_1\big](\omega) = \mathsf{E}\big[X \big| \mathcal{F}_1\big](\omega)$.

Proof. (1) $\mathcal{F}_1 \subset \mathcal{F}_2$ なので, $\mathsf{E}\big[X \big| \mathcal{F}_1\big]$ は $\mathcal{F}_2$ 可測でもある. よって $A \in \mathcal{F}_2$ に対して

$$\int_A \mathsf{E}\big[X \big| \mathcal{F}_1\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) = \int_A \mathsf{E}\Big[\mathsf{E}\big[X \big| \mathcal{F}_1\big] \Big| \mathcal{F}_2\Big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega)$$

がわかる.
(2) $\mathsf{E}\big[X \big| \mathcal{F}_1\big]$ は $\mathcal{F}_1$ 可測である. $A \in \mathcal{F}_1 \subset \mathcal{F}_2$ に対して

$$\begin{aligned}
\int_A \mathsf{E}\big[X \big| \mathcal{F}_1\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) &= \int_A X(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) \qquad \big(\because A \in \mathcal{F}_1\big) \\
&= \int_A \mathsf{E}\big[X \big| \mathcal{F}_2\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) \qquad \big(\because A \in \mathcal{F}_2\big) \\
&= \int_A \Big\{\mathsf{E}\Big[\mathsf{E}\big[X \big| \mathcal{F}_2\big] \Big| \mathcal{F}_1\Big](\omega)\Big\}\,\mathrm{d}\mathsf{Pr}(\omega) \qquad \big(\because A \in \mathcal{F}_1\big)
\end{aligned}$$

より

$$\mathsf{E}\big[X \big| \mathcal{F}_1\big](\omega) = \mathsf{E}\Big[\mathsf{E}\big[X \big| \mathcal{F}_2\big] \Big| \mathcal{F}_1\Big](\omega)$$

がわかる.

定理 C.96. X は $\mathcal{F}$ 可測とし, $\mathsf{E}\big[|X|\big] < \infty$ かつ $\mathsf{E}\big[|XY|\big] < \infty$ とする. このとき

$$\mathsf{E}\big[XY \big| \mathcal{F}\big](\omega) = X(\omega)\mathsf{E}\big[Y \big| \mathcal{F}\big](\omega) \tag{C.53}$$

が成立する.

Proof. $X(\omega) = \mathbb{1}_B(\omega)\,(\forall B \in \mathcal{F})$ のとき (C.53) が成り立つことを示す. $\forall A \in \mathcal{F}$ に対して

$$\begin{aligned}\int_A \mathbb{1}_B(\omega)\mathsf{E}\big[Y|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) &= \int_{A\cap B} \mathsf{E}\big[Y|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega)\\ &= \int_{A\cap B} Y(\omega)\,\mathrm{d}\mathsf{Pr}(\omega)\\ &= \int_A \mathbb{1}_B(\omega)Y(\omega)\,\mathrm{d}\mathsf{Pr}(\omega)\end{aligned}$$

となる. よって $X(\omega) = \mathbb{1}_B(\omega)$ のとき (C.53) は 成立する.

次に $X, Y \geq 0$ とし, X_n は階段関数とし, $X_n \uparrow X\,(n \to \infty)$ とする. 単調収束定理から

$$\begin{aligned}\int_A X(\omega)\mathsf{E}\big[Y|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) &= \lim_{n\to\infty}\int_A X_n(\omega)\mathsf{E}\big[Y|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega)\\ &= \lim_{n\to\infty}\int_A \mathsf{E}\big[X_nY|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) \qquad \big(\because \text{(1) の結果}\big)\\ &= \int_A \lim_{n\to\infty}\mathsf{E}\big[X_nY|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) \qquad \big(\because \text{単調収束定理}\big)\\ &= \int_A \lim_{n\to\infty}\mathsf{E}\big[XY|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) \qquad \big(\because \text{定理 C.94(3)}\big)\end{aligned}$$

がわかり, $X, Y \geq 0$ のとき (C.53) は成立する.

最後に一般の X, Y に対して

$$X^+ = \max\{X, 0\}, \quad X^{-1} = \max\{-X, 0\}, \quad Y^+ = \max\{Y, 0\}, \quad Y^{-1} = \max\{-Y, 0\}$$

として, 上の結果を用いればよい. □

C.11　Radon-Nikodym の定理と十分統計量

Radon-Nikodym の定理から導かれることを用いて Fisher-Neyman の因子分解定理の測度論的な証明を与える.

補題 C.97. μ を可測空間 $(\mathbb{X}, \mathcal{A})$ 上の σ 有限測度とし, $\mathcal{N}$ を $(\mathbb{X}, \mathcal{A})$ 上の測度の集合で, $\forall \nu \in \mathcal{N}$ に対して, $\nu \ll \mu$ をみたしているとする. このとき非負の数列 $\{c_n\}_{n=1}^\infty$ で $\sum_{n=1}^\infty c_n = 1$ なるものと $\{\nu_n\}_{n=1}^\infty \subset \mathcal{N}$ が存在して $\forall \nu \in \mathcal{N}$ に対して $\nu \ll \sum_{n=1}^\infty c_n\nu_n$ となる.

Proof. $\mathcal{N}$ が可算集合ならば定理の主張は自明であるので, $\mathcal{N}$ は非可算とする.

μ は有限測度の場合は

$$\lambda = \mu$$

とおく．μ が有限測度でない場合には $\mathbb{X}$ の可算個の部分集合 $\{\mathbb{X}_n\}_{n=1}^{\infty}$ をうまくとると

$$\mathbb{X} = \bigcup_{n=1}^{\infty} \mathbb{X}_n; \qquad 0 < \mu(\mathbb{X}_n) = d_n < \infty$$

とできる．$\forall B \in \mathcal{A}$ に対して

$$\lambda(B) := \sum_{n=1}^{\infty} \frac{\mu(B \cap \mathbb{X}_n)}{2^n d_n}$$

とおく．いずれの場合でも λ は有限測度で $\nu \ll \lambda\,(\forall \nu \in \mathcal{N})$ となる[6]．ここで

$$\mathcal{Q} := \left\{ \sum_{n=1}^{\infty} c_n \nu_n;\ \sum_{n=1}^{\infty} c_n = 1\ c_n \in \mathbb{R},\ \text{かつ}\ \{\nu_n\}_{n=1}^{\infty} \subset \mathcal{N} \right\}$$

とする．明らかに

$$\beta \in \mathcal{Q}\ \Rightarrow\ \beta \ll \lambda$$

である．

つぎに

$$\mathcal{D} := \left\{ C \in \mathcal{A};\ \exists \mathsf{Q} \in \mathcal{Q}\ \text{s.t.}\ \lambda\left(\left\{ x \in C;\ \frac{\mathrm{d}\mathsf{Q}}{\mathrm{d}\lambda}(x) = 0 \right\} \right) = 0\ \text{かつ}\ \mathsf{Q}(C) > 0 \right\}$$

とする．まず $\mathcal{D} \neq \varnothing$ を確認する．そのために $\nu \in \mathcal{N}$ で $\nu \neq 0$ なるものを取り

$$C := \left\{ x \in \mathbb{X};\ \frac{\mathrm{d}\nu}{\mathrm{d}\lambda}(x) > 0 \right\}$$

とする．$\mathsf{Q} = \nu$ としたとき

$$\left\{ x \in C;\ \frac{\mathrm{d}\nu}{\mathrm{d}\lambda}(x) = 0 \right\} = \varnothing$$

となり

$$Q(C) = \nu(C) = \mu(\mathbb{X}) > 0$$

[6] $\lambda(A) = 0$ なる $A \in \mathcal{A}$ をとる．すると $\mu(A \cap \mathbb{X}_n) = 0\,(\forall n \in \mathbb{N})$ となる．したがって

$$\mu(A) = \sum_{n=1}^{\infty} \mu(A \cap \mathbb{X}_n) = 0$$

となる．$\nu \ll \mu$ から $\nu(A) = 0$ となる．よって $\lambda \ll \mu$ がわかる．

となる．よって $C \in \mathcal{D}$ となる．λ は有限測度なので

$$\sup_{C \in \mathcal{D}} \lambda(C) =: c < \infty$$

となる．よって $\{C_n\}_{n=1}^{\infty} \subset \mathcal{D}$ かつ $C_1 \subset C_2 \subset \cdots$ をうまくとると

$$\lim_{n \to \infty} \lambda(C_n) = c$$

とできる．ここで

$$C_0 := \bigcup_{n=1}^{\infty} C_n$$

とおき，$\mathsf{Q}_n \in \mathcal{Q}\,(n \in \mathbb{N})$ で

$$\mathsf{Q}_n(C_n) > 0 \text{ かつ } \lambda\left(\left\{x \in C_n;\, \frac{\mathrm{d}\mathsf{Q}_n}{\mathrm{d}\lambda}(x) = 0\right\}\right) = 0$$

なるもの[7]を取る．さらに

$$\mathsf{Q}_0 := \sum_{n=1}^{\infty} \frac{\mathsf{Q}_n}{2^n} \in \mathcal{Q} \qquad \left(\because \sum_{n=1}^{\infty} \frac{1}{2^n} = 1\right)$$

とおく．すると

$$\frac{\mathrm{d}\mathsf{Q}_0}{\mathrm{d}\lambda} = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{\mathrm{d}\mathsf{Q}_n}{\mathrm{d}\lambda}$$

かつ

$$\left\{x \in C_0;\, \frac{\mathrm{d}\mathsf{Q}_0}{\mathrm{d}\lambda}(x) = 0\right\} \subset \bigcup_{n=1}^{\infty} \left\{x \in C_n;\, \frac{\mathrm{d}\mathsf{Q}_n}{\mathrm{d}\lambda}(x) = 0\right\}$$

となる．このことより

$$\lambda\left(\left\{x \in C_0;\, \frac{\mathrm{d}\mathsf{Q}_0}{\mathrm{d}\lambda}(x) = 0\right\}\right) \leq \sum_{n=1}^{\infty} \lambda\left(\left\{x \in C_n;\, \frac{\mathrm{d}\mathsf{Q}_n}{\mathrm{d}\lambda}(x) = 0\right\}\right)$$

となるで，$C_0 \in \mathcal{D}$ がわかる．さらに

$$\lambda(C_0) = \lambda\left(\bigcup_{n=1}^{\infty} C_n\right) = \lim_{n \to \infty} \lambda(C_n) = c$$

がわかる．

[7] $C_n \in \mathcal{D}$ なのでこのような $\mathsf{Q}_n \in \mathcal{Q}$ が取れる．

あとは $\mathsf{Q}_0 \in \mathcal{Q}$ なので $\forall \nu \in \mathcal{N}$ に対して $\nu \ll Q_0$ を示せばよい．そのために $A \in \mathcal{A}$ で $\mathsf{Q}_0(A) = 0$ なるものと $\forall \nu \in \mathcal{N}$ を取る．$x \in C_0$ に対して

$$\mathsf{Q}_0(A \cap C_0) \leq \mathsf{Q}_0(A) = 0 \text{ かつ } \frac{\mathrm{d}\mathsf{Q}_0}{\mathrm{d}\lambda}(x) > 0 \, (\forall x \in C_0)$$

となる[8]ので

$$\lambda(A \cap C_0) = 0 \;\Rightarrow\; \nu(A \cap C_0) = 0 \qquad (\because \nu \ll \lambda)$$

となる．ここで

$$C := \left\{ x \in \mathbb{X};\, \frac{\mathrm{d}\nu}{\mathrm{d}\lambda}(x) > 0 \right\}$$

とおく．すると

$$x \in C^c \;\Rightarrow\; \frac{\mathrm{d}\nu}{\mathrm{d}\lambda}(x) = 0$$

なので

$$\nu(A \cap C_0^c \cap C^c) \leq \int_{C^c} \frac{\mathrm{d}\nu}{\mathrm{d}\lambda}(x) \, \mathrm{d}\lambda(x) = 0$$

となる．さらに $D := A \cap C_0^c \cap C$ とおくと $D \cap C_0 = \varnothing$ となる．$\lambda(D) > 0$ のとき

$$\lambda(C_0 \cap D) > \lambda(D)$$

であり

$$Q_0\left(\left\{ x \in D;\, \frac{\mathrm{d}Q_0}{\mathrm{d}\lambda}(x) = 0 \right\}\right) = 0$$

となる[9]となるので，$D \in \mathcal{D}$ となる．明らかに $C_0 \cup D \in \mathcal{D}$ であり

$$\lambda(C_0 \cup D) > \lambda(C_0) = c$$

---

[8] $x \in C_0$ のときある $n \in \mathbb{N}$ があって $x \in C_n$ となる．すると

$$\lambda\left(\left\{ x \in C_n;\, \frac{\mathrm{d}\mathsf{Q}_n}{\mathrm{d}\lambda}(x) = 0 \right\}\right) = 0$$

なので

$$\frac{\mathrm{d}\mathsf{Q}_0}{\mathrm{d}\lambda}(x) > 0$$

となる．

[9]

$$\mathsf{Q}_0 \in \mathcal{Q} \;\Rightarrow\; \mathsf{Q}_0 \ll \lambda$$

であることからわかる．

となるので, c の定義[10]と矛盾する．しがたって

$$\lambda(D)=0 \text{ かつ } \nu(D)=0$$

となる．このことより

$$\begin{aligned}\nu(A)&=\nu(A\cap C_0)+\nu(A\cap C_0^c)\\&=\nu(A\cap C_0)+\nu(A\cap C_0^c\cap C^c)+\nu(A\cap C_0^c\cap C)\\&=\nu(A\cap C_0)+\nu(A\cap C_0^c\cap C^c)+\nu(D)=0\end{aligned}$$

がわかる．よって定理が証明された． □

定理 C.98. ν を可測空間 $(\mathbb{X},\mathcal{A})$ 上の σ 有限測度とし,

$$\mathcal{P}=\{\mathsf{P};\ \mathsf{P} \text{ は } (\mathbb{X},\mathcal{A}) \text{ 上の確率分布で}, \mathsf{P}\ll\nu\}$$

を統計的モデルとする．X を母集団分布 $\mathsf{P}\in\mathcal{P}$ からの標本とし, $T(X)$ を統計量とする．このとき $T(X)$ は $\mathcal{P}$ に対する十分統計量であるための必要十分条件は, $(\mathbb{X},\mathcal{A})$ 上の可測関数 h (これは $\mathsf{P}\in\mathcal{P}$ に依存しない) と T の値域上の可測関数 g_P (これは $\mathsf{P}\in\mathcal{P}$ に依存する) が存在して

$$\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\nu}(x)=g_{\mathsf{P}}(T(x))h(x),\qquad \mathsf{P}-a.s. \tag{C.54}$$

と書けることである．

Proof. (⇒) の証明: T は $\mathcal{P}$ に対する十分統計量とする．補題 C.97 から, $\{\mathsf{P}_n\}_{n=1}^{\infty}\subset\mathcal{P}$ と数列 $\{c_n\}_{n=1}^{\infty}$ ($\sum_{n=1}^{\infty}c_n=1$) をうまく取れば

$$\mathsf{P}\ll\mathsf{Q}=\sum_{n=1}^{\infty}c_n\mathsf{P}_n$$

とできる． ここで $\dfrac{d\mathsf{Q}}{d\nu}$ は $\mathsf{P}\in\mathcal{P}$ に依存しないことに注意する． $\mathsf{E}[\cdot]$ と $\mathsf{E}[\cdot|T]$ を P に関する期待値と条件付き期待値とし, $\mathsf{E}_{\mathsf{Q}}[\cdot]$, $\mathsf{E}_{\mathsf{P}_n}[\cdot]$ と $\mathsf{E}_{\mathsf{Q}}[\cdot|T]$, $\mathsf{E}_{\mathsf{P}_n}[\cdot|T]$ を Q, P_n それぞれにに関する期待値と条件付き期待値とする．T は十分統計量であるので T を与えたときの X の条件付き分布は $\mathsf{P}\in\mathcal{P}$ に依存しないので

$$\mathsf{E}[\mathbb{1}_A(X)|T]=\mathsf{E}_{\mathsf{P}_n}[\mathbb{1}_A(X)|T]\qquad(\forall A\in\mathcal{A}) \tag{C.55}$$

[10] $c=\sup_{C\in\mathcal{D}}\lambda(C)$ である．

が成立することに注意する．$\forall A \in \mathcal{A}$ と $B \in \sigma(T)$ に対して

$$
\begin{aligned}
\mathsf{Q}(A \cap B) &= \sum_{n=1}^{\infty} c_n \mathsf{P}_n(A \cap B) = \sum_{n=1}^{\infty} c_n \mathsf{E}_{\mathsf{P}_n}[\mathbb{1}_{A\cap B}(X)] \\
&= \sum_{n=1}^{\infty} c_n \mathsf{E}_{\mathsf{P}_n}\big[\mathsf{E}_{\mathsf{P}_n}[\mathbb{1}_{A\cap B}(X)|\, T]\big] \qquad (\because \text{条件付き期待値の性質}) \\
&= \sum_{n=1}^{\infty} c_n \mathsf{E}_{\mathsf{P}_n}\big[\mathsf{E}_{\mathsf{P}_n}[\mathbb{1}_{A}(X)\mathbb{1}_{B}(X)|\, T]\big] \\
&= \sum_{n=1}^{\infty} c_n \mathsf{E}_{\mathsf{P}_n}\big[\mathbb{1}_{B}(X)\mathsf{E}_{\mathsf{P}_n}[\mathbb{1}_{A}(X)|\, T]\big] \qquad \big(\because \mathbb{1}_B(X) \text{ は } \sigma(T) \text{ 可測}\big) \\
&= \sum_{n=1}^{\infty} c_n \mathsf{E}_{\mathsf{P}_n}\big[1_{B}(X)\mathsf{E}[\mathbb{1}_{A}(X)|\, T]\big] \qquad \big(\because \text{(C.55)}\big) \\
&= \sum_{n=1}^{\infty} c_n \int_B \Big\{\mathsf{E}[\mathbb{1}_{A}(X)|\, T](x)\Big\}\, \mathrm{d}\mathsf{P}_n \\
&= \int_B \Big\{\mathsf{E}[\mathbb{1}_{A}(X)|\, T](x)\Big\} \sum_{n=1}^{\infty} c_n\, \mathrm{d}\mathsf{P}_n \qquad (\because \text{有界収束定理}) \\
&= \int_B \Big\{\mathsf{E}[\mathbb{1}_{A}(X)|\, T](x)\Big\}\, \mathrm{d}\mathsf{Q}(x)
\end{aligned}
$$

となる．一方

$$
\mathsf{Q}(A \cap B) = \int_B \mathbb{1}_A(x)\, \mathrm{d}\mathsf{Q}(x) = \int_B \Big\{\mathsf{E}_Q[\mathbb{1}_{A}(X)|\, T](x)\Big\}\, \mathrm{d}\mathsf{Q}(x)
$$

となるので

$$
\mathsf{E}[\mathbb{1}_{A}(X)|\, T](x) = \mathsf{E}_{\mathsf{Q}}[\mathbb{1}_{A}(X)|\, T](x), \qquad \mathsf{Q} - a.s.
$$

を得る．いま

$$
\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{Q}}(x) =: g_P\big(T(x)\big)
$$

とおくと $\forall A \in \mathcal{A}$ に対して

$$\begin{aligned}
\mathsf{P}(A) &= \mathsf{E}[\mathbb{1}_A(X)] = \mathsf{E}\big[\mathsf{E}[\mathbb{1}_A(X)|\,T]\big] = \int \left\{ \mathsf{E}[\mathbb{1}_A(X)|\,T](x) \right\} \mathrm{d}\mathsf{P}(x) \\
&= \int \left\{ \mathsf{E}_Q[\mathbb{1}_A(X)|\,T](x)\, \frac{\mathrm{d}mathsfP}{\mathrm{d}\mathsf{Q}}(x) \right\} \mathrm{d}\mathsf{Q}(x) \\
&= \int \left\{ \mathsf{E}_\mathsf{Q}[\mathbb{1}_A(X)|\,T](x)\, g_\mathsf{P}(T(x)) \right\} \mathrm{d}\mathsf{Q}(x) \\
&= \int \left\{ \mathsf{E}_\mathsf{Q}[\mathbb{1}_A(X) g_\mathsf{P}(T(X))|\,T](x) \right\} \mathrm{d}\mathsf{Q}(x) \qquad \big(\because g_\mathsf{P}(T(x)) \text{ は } \sigma(T) \text{ 可測}\big) \\
&= \mathsf{E}_\mathsf{Q}\big[\mathbb{1}_A(X) g_\mathsf{P}(T(X))\big] \qquad \big(\because \text{条件付き期待値の定義}\big) \\
&= \int \left\{ \mathbb{1}_A(x) g_\mathsf{P}(T(x)) \right\} \mathrm{d}\mathsf{Q}(x) \\
&= \int_A \left\{ g_\mathsf{P}(T(x)) \frac{\mathrm{d}\mathsf{Q}}{\mathrm{d}\nu}(x) \right\} \mathrm{d}\nu(x)
\end{aligned}$$

となる．よって

$$h = \frac{\mathrm{d}\mathsf{Q}}{\mathrm{d}\nu}$$

ととればよいことがわかる．

(⇐) の証明: (C.54) が成り立つとする．T が $\mathcal{P}$ に対する十分統計量であることを示すために

$$\begin{aligned}
&\mathsf{P}(A|\,T) = \mathsf{E}_\mathsf{Q}[\mathbb{1}_A(X)|\,T] \qquad \mathsf{P}-a.s. \\
\Leftrightarrow\ &\int_B \left\{ \mathsf{P}(A|\,T)(x) \right\} \mathrm{d}\mathsf{P}(x) = \int_B \left\{ \mathsf{E}_\mathsf{Q}[\mathbb{1}_A(X)|\,T](x) \right\} \mathrm{d}\mathsf{P}(x) \quad \big(\forall B \in \sigma(T)\big)
\end{aligned} \tag{C.56}$$

を示せばよい．Radon-Nikodym の微分の性質と (C.54) から

$$\begin{aligned}
\frac{d\mathsf{P}}{d\mathsf{Q}}(x) &= \frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\nu}(x) \Big/ \frac{\mathrm{d}\mathsf{Q}}{\mathrm{d}\nu}(x) = \frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\nu}(x) \Big/ \sum_{n=1}^{\infty} c_n \frac{\mathrm{d}\mathsf{P}_n}{\mathrm{d}\nu}(x) \\
&= g_\mathsf{P}(T(x))h(x) \Big/ \sum_{n=1}^{\infty} c_n g_{\mathsf{P}_n}(T(x))h(x) \\
&= g_\mathsf{P}(T(x)) \Big/ \sum_{n=1}^{\infty} c_n g_{\mathsf{P}_n}(T(x))
\end{aligned}$$

となることに注意する．このことより $\dfrac{d\mathsf{P}}{d\mathsf{Q}}(x)$ は $\sigma(T)$ 可測であることが

わかる. $\forall B \in \sigma(T)$ に対して

$$\begin{aligned}
\int_B \left\{ \mathsf{E}_{\mathsf{Q}}[\mathbb{1}_A(X)|\,T](x) \right\} \mathrm{d}\mathsf{P}(x) &= \int_B \left\{ \mathsf{E}_{\mathsf{Q}}[\mathbb{1}_A(X)|\,T](x) \right\} \mathrm{d}\mathsf{Q}(x) \\
&= \int_B \left\{ \mathsf{E}_{\mathsf{Q}}\left[\mathbb{1}_A(X) \frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{Q}}(X) \middle| T \right](x) \right\} \mathrm{d}\mathsf{Q}(x) \\
&= \int_B \left\{ \mathbb{1}_A(x) \frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{Q}}(x) \right\} \mathrm{d}\mathsf{Q}(x) \qquad \text{(条件付き期待値の定義)} \\
&= \int_B \mathbb{1}_A(x)\, \mathrm{d}\mathsf{P}(x) = \int_B \left\{ \mathsf{E}[\mathbb{1}_A|\,T](x) \right\} \mathrm{d}\mathsf{P}(x) \\
&= \int_A \left\{ \mathsf{P}(A|\,T)(x) \right\} \mathrm{d}\mathsf{P}(x)
\end{aligned}$$

となる. よって (C.56) が示されたので, T は $\mathcal{P}$ に対する十分統計量となることがわかる. □

C.12　章末注釈と参考文献

第D章　凸解析と最適化アルゴリズム

D.1　凸関数とその性質

この節では, 凸関数の定義をまず述べる. 凸関数のこの定義は幾何的なものであり, 凸関数がどのようなものであるかを直観的に理解するが用意である. しかし, 解析的な議論をするためには, この定義では不十分[1]である. 凸関数であるための必要十分条件が重要になってくる. まず, ① 閉区間上の凸関数はその内部で絶対連続であることを示す. さらに, 凸関数であるための必要十分条件を 2 つ述べる. ひとつは, ② 凸関数の平均変動は単調増加であることであり, もうひとつは, ③下から支える直線の存在することである. これらは, 凸関数を特徴付けていることに注意する.

定義 D.1. $a, b \in \mathbb{R}, a < b$ とする. 関数 $f : [a, b] \to \mathbb{R}$ が $x, y \in [a, b], 0 \leq \lambda \leq 1$ に対して

$$f\bigl(\lambda x + (1-\lambda)y\bigr) \leq \lambda f(x) + (1-\lambda)f(y)$$

をみたすとき, 関数 f は**凸関数**であるという. とくに $x \neq y, 0 < \lambda < 1$ のとき

$$f\bigl(\lambda x + (1-\lambda)y\bigr) < \lambda f(x) + (1-\lambda)f(y)$$

が成立する場合は, 関数 f は**狭義凸関数**であるという.

注意 D.2. $-f$ が凸関数 (狭義凸関数) のとき, 関数 f は**凹関数** (**狭義凹関数**) であるという. □

定理 D.3. 区間 $[a, b]$ 上の凸関数は開区間 (a, b) 上の各点で連続である.

Proof. まず, $\forall x \in [a, b]$ は

$$x = \lambda a + (1-\lambda)b \quad (0 \leq \lambda \leq 1)$$

と表せるから, 凸関数の定義より

$$f(x) \leq \lambda f(a) + (1-\lambda)f(b) \leq \max\{f(a), f(b)\} =: K$$

[1]たとえば, この定義の直接的に利用では Jensen の不等式の証明できない.

となることに注意する．よって, 凸関数 f は上に有界である．$\forall x_0 \in (a,\, b)$ とし, $f(x)$ は $x = x_0$ で連続であることを示すために関数 g を

$$g(t) := f(x_0 + t) - f(x_0) \qquad (a - x_0 < t < b - x_0)$$

と定める．このとき $g(t)$ は $(a - x_0,\, b - x_0) \ni \{0\}$ 上で凸関数であり, $g(0) = 0$ である．いま $\forall \epsilon > 0$ に対して正数 ϵ' を

$$0 < \epsilon' < \min\left\{\frac{\epsilon}{K + |f(x_0)|},\, 1\right\} \tag{D.1}$$

をみたすように選ぶ．さらに正数 ρ と δ を

$$0 < \rho < \min\{x_0 - a,\, b - x_0\}, \quad 0 < \delta < \rho\epsilon' \tag{D.2}$$

をみたすように定める．$0 < \epsilon' \leq 1$ だから

$$|t| < \delta \;\Rightarrow\; |t| < \rho$$

となる．したがって $|t| < \delta$ に対して, $g(t)$ は定義されている．しかも

$$|t| < \delta \;\Rightarrow\; \left|\frac{t}{\epsilon'}\right| < \rho \;\Rightarrow\; x_0 + \frac{t}{\epsilon'} \in (a,\, b)$$

となること[2]がわかる．よって, (D.3) から

$$\begin{aligned} g\left(\frac{t}{\epsilon'}\right) &= f\left(x_0 + \frac{t}{\epsilon'}\right) - f(x_0) \leq \left|f\left(x_0 + \frac{t}{\epsilon'}\right)\right| + |f(x_0)| \\ &\leq K + |f(x_0)| =: K_1 \end{aligned}$$

を得る．ところで

$$t = (1 - \epsilon') \times 0 + \epsilon' \left(\frac{t}{\epsilon'}\right)$$

と表されることと (D.1) から

$$g(t) \leq (1 - \epsilon')g(0) + \epsilon' g\left(\frac{t}{\epsilon'}\right) = \epsilon' g\left(\frac{t}{\epsilon'}\right) \leq \epsilon' K_1 < \epsilon \tag{D.3}$$

となることに注意をする．他方

$$0 = \frac{\epsilon'}{1 + \epsilon'}\left(-\frac{t}{\epsilon'}\right) + \frac{1}{1 + \epsilon'} \cdot t$$

[2] $\left|\dfrac{t}{\epsilon'}\right| < \rho$ と (D.2) から $-x_0 + a < -\rho < \dfrac{t}{\epsilon'} < \rho < b - x_0$ となるので $a < \dfrac{t}{\epsilon'} + x_0 < b$ がわかる．

と表されるので

$$
\begin{aligned}
0 \le g(0) &\le \frac{\epsilon'}{1+\epsilon'} g\biggl(-\frac{t}{\epsilon'}\biggr) + \frac{1}{1+\epsilon'} g(t) \le \frac{\epsilon' K_1}{1+\epsilon'} + \frac{1}{1+\epsilon'} g(t) \\
&< \frac{\epsilon}{1+\epsilon'} + \frac{1}{1+\epsilon'} g(t)
\end{aligned}
$$

となり

$$
-\epsilon < g(t) \tag{D.4}
$$

を得る．よって (D.3) と (D.4) より

$$
|t| < \delta \;\Rightarrow\; |g(t)| < \epsilon
$$

である．$t = x - x_0$ と $g(t)$ の定義より

$$
|x - x_0| < \delta \;\Rightarrow\; |f(x) - f(x_0)| < \epsilon
$$

となり, $f(x)$ は点 x_0 で連続となる． □

定理 D.4. 関数 $\varphi\colon [a,\, b] \to \mathbb{R}$ が単調増加ならば

$$
f(x) := C + \int_a^x \varphi(t)\,\mathrm{d}t \qquad (a \le x \le b)
$$

は凸関数である．ただし C は定数である．

Proof. まず, $x,\, y \in [a,\, b]$, $x < y$, $0 \le \lambda \le 1$ のとき

$$
x \le \lambda x + (1-\lambda)y \le y
$$

であることに注意する．ここで, 関数 φ の単調性に注意すれば

$$
\int_{\lambda x+(1-\lambda)y}^{y} \varphi(t)\,\mathrm{d}t \ge \{y - \lambda x - (1-\lambda)y\}\varphi\bigl(\lambda x + (1-\lambda)y\bigr), \tag{D.5}
$$

$$
\int_{x}^{\lambda x+(1-\lambda)y} \varphi(t)\,\mathrm{d}t \le \{\lambda x + (1-\lambda)y - x\}\varphi\bigl(\lambda x + (1-\lambda)y\bigr) \tag{D.6}
$$

がわかる．f の凸性を示すために, 以下の式が非負であることがわかればよい．

$$
\begin{aligned}
&\lambda f(x) + (1-\lambda)f(y) - f\bigl(\lambda x + (1-\lambda)y\bigr) \\
&\quad = (1-\lambda)\bigl\{f(y) - f\bigl(\lambda x + (1-\lambda)y\bigr)\bigr\} - \lambda\bigl\{f\bigl(\lambda x + (1-\lambda)y\bigr) - f(x)\bigr\} \\
&\quad = (1-\lambda)\int_{\lambda x+(1-\lambda)y}^{y} \varphi(t)\,\mathrm{d}t - \lambda\int_{x}^{\lambda x+(1-\lambda)y} \varphi(t)\,\mathrm{d}t \\
&\quad \ge (1-\lambda)\{y - \lambda x - (1-\lambda)y\}\varphi\bigl(\lambda x + (1-\lambda)y\bigr) \qquad (\because \text{(D.5)}) \\
&\qquad - \lambda\{\lambda x + (1-\lambda)y - x\}\varphi\bigl(\lambda x + (1-\lambda)y\bigr) \qquad (\because \text{(D.6)}) \\
&\quad = 0
\end{aligned}
$$

となる. 以上の議論から

$$\lambda f(x)+(1-\lambda)f(y)\geq f\bigl(\lambda x+(1-\lambda)y\bigr)$$

が示せた. □

定理 D.5. 関数　f が区間 $[a, b]$ 上の凸のとき, 開区間 (a, b) の各点 x で右微分係数 $\overset{\bullet}{f}_+(x)$ が存在し, 関数

$$\overset{\bullet}{f}_+\colon (a, b)\ni x\mapsto \overset{\bullet}{f}_+(x)\in\mathbb{R}$$

は単調増加である.

　同様に左微分係数 $\overset{\bullet}{f}_-(x)$ が存在し, 関数

$$\overset{\bullet}{f}_-\colon (a, b)\ni\mapsto \overset{\bullet}{f}_-(x)\in\mathbb{R}$$

も単調増加である. さらに各点 $x\in(a, b)$ において

$$\overset{\bullet}{f}_-(x)\leq \overset{\bullet}{f}_+(x)$$

が成立する.

Proof. 任意の $x\in(a, b)$ に対して実数 η と ξ を

$$0<\eta<\min\{x-a, b-x\},\quad 0<\xi<\eta$$

をみたすように選ぶ. このとき

$$0<\frac{\xi}{\eta}<1,\quad x+\xi=\frac{\xi}{\eta}(x+\eta)+\left(1-\frac{\xi}{\eta}\right)x$$

と書けることに注意する. 関数 f は凸であることから

$$f(x+\xi)\leq\frac{\xi}{\eta}f(x+\eta)+\left(1-\frac{\xi}{\eta}\right)f(x)=\frac{\xi}{\eta}\bigl\{f(x+\eta)-f(x)\bigr\}+f(x)$$

となる. つまり $0<\xi<\eta$ のとき

$$\frac{f(x+\xi)-f(x)}{\xi}\leq\frac{f(x+\eta)-f(x)}{\eta}$$

となり, $\xi\searrow 0$ とすれば

$$\overset{\bullet}{f}_+(x)\leq\frac{f(x+\eta)-f(x)}{\eta}\tag{D.7}$$

がわかる．同様に

$$x-\xi=\frac{\xi}{\eta}(x-\eta)+\left(1-\frac{\xi}{\eta}\right)x$$

と表せるから

$$f(x-\xi)\le\frac{\xi}{\eta}f(x-\eta)+\left(1-\frac{\xi}{\eta}\right)f(x)=\frac{\xi}{\eta}\big\{f(x-\eta)-f(x)\big\}+f(x)$$

である．したがって $0<\xi<\eta$ のとき

$$\frac{f(x)-f(x-\eta)}{\eta}\le\frac{f(x)-f(x-\xi)}{\xi}$$

となり, $\xi\searrow 0$ とすれば

$$\overset{\bullet}{f}_-(x)\ge\frac{f(x)-f(x-\eta)}{\eta} \tag{D.8}$$

を得る．さらに

$$f(x)=f\left(\frac{(x+\xi)+(x-\xi)}{2}\right)\le\frac{f(x+\xi)+f(x-\xi)}{2}$$

となり

$$f(x)-f(x-\xi)\le f(x+\xi)-f(x)$$

である．だから

$$\frac{f(x)-f(x-\xi)}{\xi}\le\frac{f(x+\xi)-f(x)}{\xi}$$

ががわかる．ここで, $\xi\searrow 0$ とすると

$$\overset{\bullet}{f}_-(x)\le\overset{\bullet}{f}_+(x) \tag{D.9}$$

がわかる．したがって, (D.7) – (D.9) を考慮すれば

$$\frac{f(x)-f(x-\eta)}{\eta}\le\overset{\bullet}{f}_-(x)\le\overset{\bullet}{f}_+(x)\le\frac{f(x+\eta)-f(x)}{\eta} \tag{D.10}$$

がわかる．よって $\overset{\bullet}{f}_-(x)$ と $\overset{\bullet}{f}_+(x)$ は有界であることがわかる．

次に, 単調性を示す．そのために $a<x_1<x_2<b$ とする．(D.10) の

$$f_-(x)\le f_+(x)\le\frac{f(x+\eta)-f(x)}{\eta}$$

において, $x=x_1$, $x+\eta=x_2$ とおくと

$$f_-(x_1)\le f_+(x_1)\le\frac{f(x_2)-f(x_1)}{x_2-x_1}$$

を得る．同様に (D.10) の

$$\frac{f(x)-f(x-\eta)}{\eta} \leq f_{-}(x) \leq f_{+}(x)$$

において, $x = x_2$, $x - \eta = x_1$ とおくと

$$\frac{f(x_2)-f(x_1)}{x_2 - x_1} \leq f_{-}(x_2) \leq f_{+}(x_2)$$

を得る．これらの式を合わせると

$$\dot{f}_{-}(x_1) \leq \dot{f}_{+}(x_1) \leq \frac{f(x_2)-f(x_1)}{x_2 - x_1} \leq \dot{f}_{-}(x_2) \leq \dot{f}_{+}(x_2) \tag{D.11}$$

となる．よって

$$\dot{f}_{-}(x_1) \leq \dot{f}_{-}(x_2), \quad \dot{f}_{+}(x_1) \leq \dot{f}_{+}(x_2)$$

が得られる． □

定理 D.6. 区間 (a, b) 上で定義された凸関数 f は, 単調増加で右連続関数 g と点 $c \in (a, b)$ によって

$$f(x) = f(c) + \int_c^x g(t)\,\mathrm{d}t \qquad (a < x < b)$$

と表せる．

Proof. $a < c < x < b$ とし, 閉区間 $[c, x]$ の分割を

$$c = x_0 < x_1 < \cdots < x_n = x$$

とする．このとき (D.11) より, $k = 1, 2, \ldots, n$ に対し

$$\dot{f}_{-}(x_{k-1}) \leq \dot{f}_{+}(x_{k-1}) \leq \frac{f(x_k)-f(x_{k-1})}{x_k - x_{k-1}} \leq \dot{f}_{-}(x_k) \leq \dot{f}_{+}(x_k)$$

と書ける．これらに $x_k - x_{k-1}$ を掛ければ

$$\begin{aligned}\dot{f}_{-}(x_{k-1})(x_k - x_{k-1}) \leq \dot{f}_{+}(x_{k-1})(x_k - x_{k-1}) &\leq f(x_k) - f(x_{k-1}) \\ &\leq \dot{f}_{-}(x_k)(x_k - x_{k-1}) \leq \dot{f}_{+}(x_k)(x_k - x_{k-1})\end{aligned}$$

が成立し, 上の不等式の辺々について, 1 から n まで k について和をとると

$$\begin{aligned}\sum_{k=1}^{n} \dot{f}_{-}(x_{k-1})(x_k - x_{k-1}) \leq \sum_{k=1}^{n} \dot{f}_{+}(x_{k-1})(x_k - x_{k-1}) &\leq f(x) - f(c) \\ &\leq \sum_{k=1}^{n} \dot{f}_{-}(x_k)(x_k - x_{k-1}) \leq \sum_{k=1}^{n} \dot{f}_{+}(x_k)(x_k - x_{k-1})\end{aligned}$$

が成立する．$\overset{\bullet}{f}_-$ と $\overset{\bullet}{f}_+$ は単調増加だから Riemann 積分可能である．よって $\max_{1\le k\le n}\{x_k - x_{k-1}\} \to 0\,(n\to\infty)$ とすれば

$$
\begin{aligned}
\sum_{k=1}^{n} \overset{\bullet}{f}_-\,(x_{k-1})(x_k - x_{k-1}) &\to \int_c^x \overset{\bullet}{f}_-\,(t)\,\mathrm{d}t \\
\sum_{k=1}^{n} \overset{\bullet}{f}_+\,(x_{k-1})(x_k - x_{k-1}) &\to \int_c^x \overset{\bullet}{f}_+\,(t)\,\mathrm{d}t \\
\sum_{k=1}^{n} \overset{\bullet}{f}_-\,(x_{k})(x_k - x_{k-1}) &\to \int_c^x \overset{\bullet}{f}_-\,(t)\,\mathrm{d}t \\
\sum_{k=1}^{n} \overset{\bullet}{f}_+\,(x_{k})(x_k - x_{k-1}) &\to \int_c^x \overset{\bullet}{f}_+\,(t)\,\mathrm{d}t
\end{aligned}
$$

となり

$$
f(x) - f(c) = \int_c^x \overset{\bullet}{f}_-\,(t)\,\mathrm{d}t = \int_c^x \overset{\bullet}{f}_+\,(t)\,\mathrm{d}t
$$

がわかる．他方 $a < x < b$ のときも閉区間 $[x,\,c]$ について同様の論理を展開して

$$
f(c) - f(x) = \int_x^c \overset{\bullet}{f}_-\,(t)\,\mathrm{d}t = \int_x^c \overset{\bullet}{f}_+\,(t)\,\mathrm{d}t
$$

を示すことができる．したがって

$$
g(x) = \overset{\bullet}{f}_+\,(x+0)
$$

とおけば, g は右連続で単調増加で

$$
f(x) = f(c) + \int_c^x g(t)\,\mathrm{d}t
$$

となる． □

定理 D.7. 関数 f が区間 $(a,\,b)$ 上の凸であるための必要十分条件は, $\forall x_0 \in (a,\,b)$ について関数

$$
(a,\,b)\setminus\{x_0\} \ni x \mapsto \frac{f(x) - f(x_0)}{x - x_0} \in \mathbb{R} \tag{D.12}
$$

が単調増加となることである．

Proof. 必要性: 関数 f が開区間 $(a,\,b)$ 上で凸ならば, 定理 D.6 により, 右連続な単調増加関数 g が存在して

$$
f(x) - f(x_0) = \int_{x_0}^x g(t)\,\mathrm{d}t
$$

と表せる．よって $x \neq x_0$ のとき

$$\frac{f(x) - f(x_0)}{x - x_0} = \frac{1}{x - x_0} \int_{x_0}^{x} g(t)\,\mathrm{d}t$$

は $(a, b) \setminus \{x_0\}$ 上で単調増加である．

十分性: (D.12) が成立したと仮定する．$\forall x, y \in (a, b)$, $x < y$, $0 < \lambda < 1$ に対し

$$x_0 = \lambda x + (1 - \lambda) y$$

とおく．(D.12) より $x < x_0 < y$ のとき

$$\frac{f(x_0) - f(x)}{x_0 - x} \leq \frac{f(y) - f(x_0)}{y - x_0}.$$

この式に $x_0 = \lambda x + (1 - \lambda) y$ を代入すると

$$\frac{f\bigl(\lambda x + (1 - \lambda) y\bigr) - f(x)}{(1 - \lambda)(y - x)} \leq \frac{f(y) - f\bigl(\lambda x + (1 - \lambda) y\bigr)}{\lambda (y - x)}$$

を得る．上の不等式の辺々に $\lambda(1 - \lambda)(y - x)$ を掛けと

$$\lambda \Bigl\{ f\bigl(\lambda x + (1 - \lambda) y\bigr) - f(x) \Bigr\} \leq (1 - \lambda) \Bigl\{ f(y) - f\bigl(\lambda x + (1 - \lambda) y\bigr) \Bigr\}$$

を得る．これを整理すれば

$$f\bigl(\lambda x + (1 - \lambda) y\bigr) \leq \lambda f(x) + (1 - \lambda) f(y)$$

がわかる．よって, $\lambda = 0, 1$ のときは自明だから, 関数 f は (a, b) 上で凸であることがわかる．□

定義 D.8. 関数 f を (a, b) 上の凸とする．点 $x_0 \in (a, b)$ において適当に実数 m を定めると $\forall x \in (a, b)$ に対して

$$S(x) := m(x - x_0) + f(x_0) \leq f(x) \tag{D.13}$$

が成立するとき, 直線 $y = S(x)$ を**点 x_0 において f を下から支える直線**という．

定理 D.9. 関数 f が区間 (a, b) 上の凸であるための必要十分条件は, (a, b) の各点 x_0 において f を下から支える直線 $y = S(x)$ が存在することである．

Proof. 必要性: 関数 f は開区間 (a, b) 上の凸で $x_0 \in (a, b)$ なので, 定理 D.6 より, ある右連続な単調増加関数 g が存在して

$$f(x) = f(x_0) + \int_{x_0}^{x} g(t)\,\mathrm{d}t$$

と書ける．このとき

$$f(x) \geq f(x_0) + g(x_0) \int_{x_0}^{x} \mathrm{d}t = f(x_0) + g(x_0)(x - x_0)$$

となることがわかる．したがって, 直線

$$S(x) := f(x_0) + g(x_0)(x - x_0)$$

は点 x_0 で $f(x)$ を下から支えることがわかる．

十分性: $x, y \in (a, b)$, $x < y$, $0 \leq \lambda \leq 1$ とし

$$x_0 = \lambda x + (1 - \lambda) y$$

とおく．仮定より x_0 において凸関数 f を下から支える直線 $S(x)$ が存在する．すると (D.13) から $S(x) \leq f(x)$ とである．さらに, S の線型性から

$$\begin{aligned} f\bigl(\lambda x + (1-\lambda) y\bigr) &= f(x_0) = S(x_0) = \lambda S(x) + (1-\lambda) S(y) \\ &\leq \lambda f(x) + (1-\lambda) f(y) \end{aligned}$$

となる．よって, 関数 f は (a, b) 上の凸である．□

D.2　凸集合とその性質

定義 D.10. C を $\mathbb{R}^n$ の空でない部分集合とする．部分集合 C は凸であるとは次の条件

$$\forall x, y \in C \text{ と } 0 \leq \lambda \leq 1 \Rightarrow \lambda x + (1 - \lambda) y \in C$$

をみたすときをいう．

定義 D.11. $p \in \mathbb{N}$ とし, $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_p\} \subset \mathbb{R}^n$ とする．

(1) $\lambda_k \geq 0\,(k = 1, 2, \ldots, p)$ かつ $\sum_{k=1}^{p} \lambda_k = 1$ のとき

$$\sum_{k=1}^{p} \lambda_k \boldsymbol{a}_k$$

を $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_p$ の**凸結合**という．

(2) $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_p$ によって**生成された凸多面体**を

$$\bigl\langle \boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_p \bigr\rangle_{2,n} = \left\{ \sum_{k=1}^{p} \lambda_k \boldsymbol{a}_k ;\, \lambda_k \geq 0, \sum_{k=1}^{p} \lambda_k = 1 \right\} \subset \mathbb{R}^n$$

で定義する．

定理 D.12. (Carathéodory の定理) $p \in \mathbb{N}$ とし, $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_p\} \subset \mathbb{R}^n$ とする. $\boldsymbol{y} \in \left\langle \boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_p \right\rangle$ は以下のように表せる.

$$
\begin{aligned}
&\boldsymbol{y} = \sum_{k=1}^{n+1} \lambda_k \boldsymbol{a}_{j_k}, \\
&\quad 1 \leq j_1 < j_2 < \cdots < j_{n+1} \leq p, \quad \lambda_k \geq 0, \quad \sum_{k=1}^{n+1} \lambda_k = 1.
\end{aligned}
$$

Proof. $\boldsymbol{y} \in\in \left\langle \boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_p \right\rangle =: A$ なので, ある $r \in \mathbb{N}$ と $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_r \in A$ があって

$$
\boldsymbol{y} = \sum_{i=1}^{r} \lambda_i \boldsymbol{x}_i, \quad \lambda_i > 0\,(i = 1, 2, \ldots, r), \quad \sum_{i=1}^{r} \lambda_i = 1
$$

と書ける. r は最小とする. すなわち, $r-1$ 個以下の A の点では $\boldsymbol{y}$ をそれらの凸結合として表せないとする. $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_r$ はアフアィン従属と仮定する. すると, ある $a_i \in \mathbb{R}\,(i = 1, 2, \ldots, r)$ で $a_1 a_2 \cdots a_r \neq 0$ なるものが存在して

$$
\sum_{i=1}^{r} a_i \boldsymbol{x}_i = \boldsymbol{0}_n \quad \text{かつ} \quad \sum_{i=1}^{r} a_i = 0
$$

とできる. ただし, $\boldsymbol{0}_n$ は $\mathbb{R}^n$ の零ベクトルである. 集合 $\{a_1, a_2, \ldots, a_r\}$ の中の少なくとも 1 つは正である. 正であるもののなかで $\dfrac{\lambda_i}{a_i}$ が最小となるもの $a_i\,(i \in \{1, 2, \ldots, r\})$ を選ぶ. 一般性を失わずにそれを a_1 と書くことができる. すると

$$
\boldsymbol{y} = \sum_{i=1}^{r} \lambda_i \boldsymbol{x}_i = \sum_{i=1}^{r} \lambda_i \boldsymbol{x}_i - \frac{\lambda_1}{a_1} \underbrace{\sum_{i=1}^{r} a_i \boldsymbol{x}_i}_{=0} = \sum_{i=1}^{r} \left(\lambda_i - \frac{\lambda_1}{a_1} a_i \right) \boldsymbol{x}_i
$$

と書ける. すると $a_i > 0\,(i \in \{2, 3, \ldots, r\})$ のとき, $\dfrac{\lambda_1}{a_1}$ の最小性から, $\dfrac{\lambda_i}{a_i} > \dfrac{\lambda_1}{a_1}$ となる. したがって

$$
\lambda_i - \frac{\lambda_1}{a_1} a_i \geq \lambda_i - \frac{\lambda_i}{a_i} a_i = 0
$$

となる. $a_i \leq 0\,(i \in \{2, 3, \ldots, r\})$ のときは, 明らかのに $\lambda_i - \dfrac{\lambda_1}{a_1} a_i > 0$ となる. 以上から任意の $i \in \{1, 2, \ldots, r\}$ に対して

$$
\lambda_i - \frac{\lambda_1}{a_1} a_i \geq 0
$$

がわかる．さらに $\left\{\lambda_i - \dfrac{\lambda_1}{a_1}a_i\right\}_{i=1,\,2,\,\dots,\,r}$ の中の少なくとも 1 は零ではない．よって, r の最小性と矛盾する．よって, $x_1,\,x_2,\,\dots,x_r$ はアフファィン独立となるので, $r \le k+1$ がわかる． □

補題 D.13. C を空でない $\mathbb{R}^n$ 凸閉部分集合とし, $\mathbf{0} \notin C$ とする．$\langle\cdot,\,\cdot\rangle_{2,n}$ を $\mathbb{R}^n$ の Euclid 内積とし, $|\cdot|_{2,n}$ を $\mathbb{R}^n$ の Euclid ノルムとする．

(1) $\boldsymbol{x}_0 \in C$ があって

$$|\boldsymbol{x}_0|_{2,\,n} = \min\bigl\{|\boldsymbol{x}|_{2,\,n};\,\boldsymbol{x}\in C\bigr\} > 0$$

となる．

(2) さらに

$$0 \le \langle\boldsymbol{x}_0,\,\boldsymbol{x}-\boldsymbol{x}_0\rangle_{2,\,n}\,(\forall\boldsymbol{x}\in C) \tag{D.14}$$

が成立する．

Proof.

$$\delta = \inf\bigl\{|\boldsymbol{x}|_{2,\,n};\,\boldsymbol{x}\in C\quad\bigr\} \ge 0$$

とおく．点列 $\{\boldsymbol{x}_k\}_{k=1}^{\infty} \subset C$ を選んで

$$\delta = \lim_{k\to\infty}|\boldsymbol{x}_k|_{2,\,n}$$

となるようにする．$\{\boldsymbol{x}_k\}_{k=1}^{\infty}$ は有界なので Bolzano-Weirstrass の定理より収束する部分列 $\{\boldsymbol{x}_{k(j)}\}_{j=1}^{\infty}$ が存在するので, この部分列の収束先を

$$\lim_{j\to\infty}\boldsymbol{x}_{k(j)} =: \boldsymbol{x}_0$$

とおく．結果

$$|\boldsymbol{x}_0|_{2,\,n} = \lim_{j\to\infty}|\boldsymbol{x}_{k(j)}|_{2,\,n} = \delta$$

となる．部分集合 C は閉なので $\boldsymbol{x}_0 \in C$ となる．$\mathbf{0}_n \notin C$ なので $\delta > 0$ がわかる．

(2) $\boldsymbol{x}_0 \in C$ は $|\boldsymbol{x}_0|_{2,\,n} = \delta$ をみたす唯一の点であることを示すために $\boldsymbol{x}_1 \in C$ で $|\boldsymbol{x}_1|_{2,\,n} = \delta$ とする．部分集合 C の凸性から

$$\frac{1}{2}(\boldsymbol{x}_0+\boldsymbol{x}_1) \in C$$

となる．したがって

$$\left|\frac{1}{2}(\boldsymbol{x}_0+\boldsymbol{x}_1)\right|_{2,\,n} \ge \delta$$

である．しかし

$$\begin{aligned}4\delta^2 &\le |\boldsymbol{x}_0+\boldsymbol{x}_1|_{2,n}^2 \le |\boldsymbol{x}_1-\boldsymbol{x}_0|_{2,n}^2+|\boldsymbol{x}_1+\boldsymbol{x}_0|_{2,n}^2 \le 2\big\{|\boldsymbol{x}_1|_{2,n}^2+|\boldsymbol{x}_0|_{2,n}^2\big\}\\ &= 4\delta^2\end{aligned}$$

より

$$|\boldsymbol{x}_1-\boldsymbol{x}_0|_{2,n}^2=0 \ \Rightarrow\ \boldsymbol{x}_1=\boldsymbol{x}_0$$

である．(D.14) を示すために $\forall\boldsymbol{x}\in C$ を取る．このとき $(1-\lambda)\boldsymbol{x}_0+\lambda\boldsymbol{x}\in C\,(0\le\lambda\le 1)$ である．したがって

$$|\boldsymbol{x}_0|_{2,n}^2 \le |(1-\lambda)\boldsymbol{x}_0+\lambda\boldsymbol{x}|_{2,n}^2 = |\boldsymbol{x}_0|_{2,n}^2+\lambda^2|\boldsymbol{x}-\boldsymbol{x}_0|_{2,n}^2+2\lambda\langle\boldsymbol{x}_0,\,\boldsymbol{x}-\boldsymbol{x}_0\rangle_{2,n}$$

である．結果

$$0\le\lambda|\boldsymbol{x}-\boldsymbol{x}_0|_{2,n}^2+2\langle\boldsymbol{x}_0,\,\boldsymbol{x}-\boldsymbol{x}_0\rangle_{2,n}$$

がわかる．ここで, $\lambda\to 0$ とすれば (D.14) がわかる．　□

定義 D.14. S_1 と S_2 を $\mathbb{R}^n$ の空でない部分集合とする．$\boldsymbol{a}\in\mathbb{R}^n$, $b\in\mathbb{R}$ とする．

(1) 超平面 $\langle\boldsymbol{a},\,\boldsymbol{x}\rangle_{2,n}+b$ は S_1 と S_2 を**分離**するとは次の条件をみたすときをいう．

$$\begin{aligned}\forall\boldsymbol{x}\in S_1 \ &\Rightarrow\ \langle\boldsymbol{a},\,\boldsymbol{x}\rangle_{2,n}+b\ge 0,\\ \forall\boldsymbol{x}\in S_2 \ &\Rightarrow\ \langle\boldsymbol{a},\,\boldsymbol{x}\rangle_{2,n}+b\le 0.\end{aligned}$$

(2) 超平面 $\langle\boldsymbol{a},\,\boldsymbol{x}\rangle_{2,n}+b$ は S_1 と S_2 を**厳密に分離**するとは次の条件をみたすときをいう．

$$\begin{aligned}\forall\boldsymbol{x}\in S_1 \ &\Rightarrow\ \langle\boldsymbol{a},\,\boldsymbol{x}\rangle_{2,n}+b> 0,\\ \forall\boldsymbol{x}\in S_2 \ &\Rightarrow\ \langle\boldsymbol{a},\,\boldsymbol{x}\rangle_{2,n}+b< 0.\end{aligned}$$

定理 D.15. $C\subset\mathbb{R}^n$ を空でない凸閉部分集合とし, $\boldsymbol{y}\notin C$ とする．このとき C と $\boldsymbol{y}$ を厳密に分離する超平面 $\langle\boldsymbol{a},\,\boldsymbol{x}\rangle_{2,n}+b\,(\boldsymbol{a}\in\mathbb{R}^n,\,b\in\mathbb{R})$ が存在する．

Proof. 一般性を失うことなく $\boldsymbol{y}=\boldsymbol{0}$ としてよい．左の仮定から $\boldsymbol{0}\notin C$ に注意して補題 D.13 を用いると $\boldsymbol{x}_0\in C$ があって

$$0<\delta=|\boldsymbol{x}_0|_{2,n}=\min\{|\boldsymbol{x}|_{2,n};\,\boldsymbol{x}\in C\}$$

となる．ここで $\boldsymbol{z}=(1/2)\boldsymbol{x}_0$, $\boldsymbol{a}=\boldsymbol{x}_0/\delta$ とおく．超平面 $\langle\boldsymbol{a},\,\boldsymbol{x}-\boldsymbol{z}\rangle_{2,n}=0$ を考える．すなわち

$$\langle\boldsymbol{a},\,\boldsymbol{x}\rangle_{2,n}+b=0,\qquad b=-\langle\boldsymbol{a},\,\boldsymbol{z}\rangle_{2,n}$$

である．$\boldsymbol{y}=\boldsymbol{0}_n$ に対して

$$\langle \boldsymbol{a},\, \boldsymbol{y}\rangle_{2,n}+b=b=-\langle \boldsymbol{a},\, \boldsymbol{z}\rangle_{2,n}=-\left\langle \frac{\boldsymbol{x}_0}{\delta},\, \frac{\boldsymbol{x}_0}{2}\right\rangle_{2,n}=-\frac{|\boldsymbol{x}_0|_{2,n}^2}{2\delta}=-\frac{\delta}{2}<0. \tag{D.15}$$

$\boldsymbol{x}\in C$ に対して (D.14) から

$$0\le \langle \boldsymbol{x}_0,\, \boldsymbol{x}-\boldsymbol{x}_0\rangle_{2,n}$$

である．よって

$$0\le \langle \delta\boldsymbol{a},\, \boldsymbol{x}-2\boldsymbol{z}\rangle_{2,n}$$

となる．上の式の両辺を $\delta>0$ で割れば

$$\begin{aligned}
0&\le \langle \boldsymbol{a},\, \boldsymbol{x}-\boldsymbol{z}\rangle_{2,n}-\langle \boldsymbol{a},\, \boldsymbol{z}\rangle_{2,n}=\langle \boldsymbol{a},\, \boldsymbol{x}-\boldsymbol{z}\rangle_{2,n}-\left\langle \frac{\boldsymbol{x}_0}{\delta},\, \frac{\boldsymbol{x}_0}{2}\right\rangle_{2,n}\\
&=\langle \boldsymbol{a},\, \boldsymbol{x}-\boldsymbol{z}\rangle_{2,n}-\frac{|\boldsymbol{x}_0|_{2,n}^2}{2\delta}\\
&=\langle \boldsymbol{a},\, \boldsymbol{x}-\boldsymbol{z}\rangle_{2,n}-\frac{\delta}{2}
\end{aligned}$$

となる．よって

$$\langle \boldsymbol{a},\, \boldsymbol{x}-\boldsymbol{z}\rangle_{2,n}=\langle \boldsymbol{a},\, \boldsymbol{x}\rangle_{2,n}-\langle \boldsymbol{a},\, \boldsymbol{z}\rangle_{2,n}=\langle \boldsymbol{a},\, \boldsymbol{x}\rangle_{2,n}+b\ge\frac{\delta}{2}>0 \tag{D.16}$$

となる．(D.15) と (D.16) より $\langle \boldsymbol{a},\, \boldsymbol{x}\rangle_{2,n}+b$ は $\boldsymbol{y}=\boldsymbol{0}_n$ と C を厳密に分離することが示せた． □

定義 D.16. $\{\boldsymbol{a}_1,\, \boldsymbol{a}_2,\, \dots,\, \boldsymbol{a}_p\}\subset\mathbb{R}^n$ とする．集合

$$C=\left\{\sum_{j=1}^{p}\lambda_j\boldsymbol{a}_j;\, \lambda_j\ge 0\,(j=1,\, 2,\, \dots,\, p)\right\}$$

を $\{\boldsymbol{a}_1,\, \boldsymbol{a}_2,\, \dots,\, \boldsymbol{a}_p\}$ によって生成された**有限錐**という．

定義 D.17. $\{\boldsymbol{a}_1,\, \boldsymbol{a}_2,\, \dots,\, \boldsymbol{a}_p\}$ が線型独立のとき，このベクトルで生成された有限錐を**基本錐**という．

補題 D.18. $\{\boldsymbol{a}_1,\, \boldsymbol{a}_2,\, \dots,\, \boldsymbol{a}_p\}\subset\mathbb{R}^n$ は有限錐 C を生成しているとする．$C_1,\, C_2,\, \dots,\, C_q$ を $\{\boldsymbol{a}_1,\, \boldsymbol{a}_2,\, \dots,\, \boldsymbol{a}_p\}$ の部分集合で生成されるすべての基本錐とする．このとき

$$C=\bigcup_{j=1}^{q}C_j$$

となる．

Proof. $C_j \subset C\,(j = 1, 2, \ldots, q)$ なので

$$\bigcup_{j=1}^{q} C_j \subset C$$

となる．いま $\boldsymbol{b} \in C = \{\mathbf{A}\boldsymbol{\beta};\, \boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^\top, \beta_j \geq 0\,(j = 1, 2, \ldots, p)\}$ を選ぶ．ただし $\mathbf{A} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_p)$ は $n \times p$ 行列である．このとき

$$\mathbf{A}\boldsymbol{\beta} = \boldsymbol{b}, \quad \boldsymbol{\beta} \geq \mathbf{0}_p$$

なので, 基本解 $\boldsymbol{\beta}^* = (\beta_1^*, \beta_2^*, \ldots, \beta_p^*)^\top \in \mathbb{R}^p$ が存在して

$$\mathbf{A}\boldsymbol{\beta}^* = \boldsymbol{b}$$

となることがわかる．すなわち $\exists m \in \mathbb{N}, (1 \leq m \leq n)$ があって, $1 \leq j_1 < \cdots < j_m \leq p$ は

$$\beta_{j_\ell} > 0\,(j = 1, 2, \ldots, m)$$

で

$$\{\boldsymbol{a}_{j_1}, \boldsymbol{a}_{j_2}, \ldots, \boldsymbol{a}_{j_m}\}$$

は線型独立である．よって $\boldsymbol{b}$ は $\{\boldsymbol{a}_{j_1}, \boldsymbol{a}_{j_2}, \ldots, \boldsymbol{a}_{j_m}\}$ で生成された基本錐に含まれるので

$$\boldsymbol{b} \in \cup_{j=1}^{q} C_j$$

となる．よって

$$C \subset \bigcup_{j=1}^{q} C_j$$

となる． □

定理 D.19. $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_p\} \subset \mathbb{R}^n$ とし, C はこのベクトルで生成された有限錐とする．このとき C は凸かつ閉部分集合である．

Proof. 凸性: $\boldsymbol{c}_1, \boldsymbol{c}_2 \in C$, $0 \leq \lambda \leq 1$ とする．このとき $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in \mathbb{R}^p$ が存在して

$$\boldsymbol{c}_1 = \mathbf{A}\boldsymbol{\beta}_1, \quad \boldsymbol{c}_2 = \mathbf{A}\boldsymbol{\beta}_2$$

となる．ただし $\mathbf{A} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_p)$, $\boldsymbol{\beta}_1 \geq \mathbf{0}_p$, $\boldsymbol{\beta}_2 \geq \mathbf{0}_p$ である．したがって

$$(1-\lambda)\boldsymbol{c}_1 + \lambda\boldsymbol{c}_2 = \mathbf{A}\big((1-\lambda)\boldsymbol{\beta}_1 + \lambda\boldsymbol{\beta}_2\big)$$

となる．さらに $(1-\lambda)\boldsymbol{\beta}_1 + \lambda\boldsymbol{\beta}_2 \geq \mathbf{0}_p$ である．よって,

$$(1-\lambda)\boldsymbol{c}_1 + \lambda\boldsymbol{c}_2 \in C$$

より C は凸集合であることがわかる．

<u>C は閉集合</u>: $C_j\,(j=1,\,2,\,\ldots,\,q)$ を独立なベクトル

$$\{\boldsymbol{a}_{j_1},\,\boldsymbol{a}_{j_2},\,\ldots,\,\boldsymbol{a}_{j_\ell}\}\subset\{\boldsymbol{a}_1,\,\boldsymbol{a}_2,\,\ldots,\,\boldsymbol{a}_p\}$$

によって生成される基本錐とする．$\{\boldsymbol{c}_k\}_{k=1}^\infty$ は C_j の収束する点列とし, $\boldsymbol{c}_0\in\mathbb{R}^n$ に収束するとする．

$$\lim_{k\to\infty}\boldsymbol{c}_k=\boldsymbol{c}_0.$$

$\boldsymbol{c}_0\in C_j$ を示そう．$\{\boldsymbol{a}_{j_1},\,\boldsymbol{a}_{j_2},\,\ldots,\,\boldsymbol{a}_{j_\ell}\}$ は線型独立なので, 任意の $\boldsymbol{u}=(u_1,\,u_2,\,\ldots,\,u_\ell)^\top\in\mathbb{R}^\ell$ に対し

$$\sum_{k=1}^{\ell}u_k\boldsymbol{a}_{j_k}=\boldsymbol{0}_n\ \Rightarrow\ \boldsymbol{u}=\boldsymbol{0}_n$$

となる．よって上の主張の対偶を取ると $\boldsymbol{u}\in\mathbb{R}^\ell$ で $|\boldsymbol{u}|_{2,\,\ell}=1$ なるものに対して

$$\sum_{k=1}^{\ell}u_k\boldsymbol{a}_{j_k}\neq\boldsymbol{0}_\ell$$

となる．よって C_j は凸集合なので, 補題 D.18 より

$$\min\left\{\left|\sum_{k=1}^{\ell}u_k\boldsymbol{a}_{j_k}\right|_{2,\,\ell};\ |\boldsymbol{u}|_{2,\,\ell}=1\right\}=:\epsilon>0$$

がわかる．したがって ϵ の定義に注意すると $\boldsymbol{v}=(v_1,\,v_2,\,\ldots,\,v_\ell)^\top\in\mathbb{R}^\ell$ に対し

$$\left|\sum_{k=1}^{\ell}v_k\boldsymbol{a}_{j_k}\right|_{2,\,\ell}=|\boldsymbol{v}|_{2,\,\ell}\left|\sum_{j=1}^{\ell}\frac{v_k}{|\boldsymbol{v}|_{2\,\ell}}\boldsymbol{a}_{jk}\right|_{2,\,\ell}\geq\epsilon|\boldsymbol{v}|_{2,\,\ell}\tag{D.17}$$

となる．$\boldsymbol{c}_k\in C_j\,(k=1,\,2,\,\ldots)$ に対して

$$\boldsymbol{c}_k=\mathbf{A}\boldsymbol{\beta}_k$$

とおけば

$$\{j_1,\,j_2,\,\ldots,\,j_\ell\}=\{j\in\{1,\,2,\,\ldots,\,p\};\,\beta_j>0,\,\boldsymbol{\beta}_k=(\beta_1,\,\beta_2,\,\ldots,\,\beta_p)^\top\}.$$

このとき

$$\boldsymbol{c}_k=\sum_{k=1}^{\ell}\beta_{j_k}\boldsymbol{a}_{j_k}.$$

(D.17) より

$$\frac{1}{\epsilon}|\boldsymbol{c}_k|_{2,\,p}\geq|\boldsymbol{\beta}_k|_{2,\,p}\quad(k=1,\,2,\,\ldots)$$

となる. よって $\{\boldsymbol{c}_k\}_{k=1}^{\infty}$ は収束するので, $\{\boldsymbol{\beta}_k\}_{k=1}^{\infty}$ は有界. よって Bolzano-Weierstrass の定理より収束する部分列 $\{\boldsymbol{\beta}_{k_j}\}_{j=1}^{\infty}$ が取れ

$$\lim_{j\to\infty}\boldsymbol{\beta}_{k_j} =: \boldsymbol{\beta}_0$$

とする. このとき $\boldsymbol{\beta}_0 \geq \mathbf{0}$ で

$$\mathbf{A}\boldsymbol{\beta}_0 = \lim_{j\to\infty}\mathbf{A}\boldsymbol{\beta}_{k_j} = \lim_{j\to\infty}\boldsymbol{c}_{k_j} = \boldsymbol{c}_0$$

となる. $\boldsymbol{\beta}_0$ の j_k 成分 $(k = 1, 2, \ldots, \ell)$ 以外は零なので, $\boldsymbol{c}_0 \in C_j$ となる. よって C_j は閉集合である. だから

$$C = \bigcup_{j=1}^{q} C_j$$

も閉集合となることがわかる. □

定理 D.20. (Farkas の交代定理) $\mathbf{A}$ を $m \times n$ 行列とし, $\boldsymbol{b} \in \mathbb{R}^m$ とする. このとき次のいずれかが成立し, 同時には成立しない.

(1) $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$, $\boldsymbol{x} \geq \mathbf{0}_n$ は解 $\boldsymbol{x}^* \in \mathbb{R}^n$ を持つ.

(2) $\mathbf{A}^\top\boldsymbol{y} \geq \mathbf{0}_m$, $\langle \boldsymbol{y}, \boldsymbol{b}\rangle_{2,m} < 0$ は解 $\boldsymbol{y}^* \in \mathbb{R}^m$ を持つ.

ただし $\boldsymbol{w} = (w_1, w_2, \ldots, w_n)^\top$, $\boldsymbol{z} = (z_1, z_2, \ldots, z_n)^\top \in \mathbb{R}^n$ のとき

$$\boldsymbol{w} \geq \boldsymbol{z} \Leftrightarrow w_j \geq z_j\,(j = 1, 2, \ldots, n)$$

である.

Proof. $\boldsymbol{x}^* \in \mathbb{R}^m$ は (1) の解, $\boldsymbol{y}^* \in \mathbb{R}^n$ は (2) の解とする. このとき

$$0 \leq \langle \boldsymbol{x}^*, \mathbf{A}^\top\boldsymbol{y}^*\rangle_{2,n} = \langle \mathbf{A}\boldsymbol{x}^*, \boldsymbol{y}^*\rangle_{2,m} = \langle \boldsymbol{b}, \boldsymbol{y}\rangle_{2,m} < 0$$

となる. よって矛盾. (1) と (2) は同時に起こらないことがわかる.

(1) が起こらないとき, (2) が起こることを示す. 仮定より

$$\boldsymbol{b} \notin C := \{\mathbf{A}\boldsymbol{x};\, \boldsymbol{x} \geq \mathbf{0}_n\}$$

となる. C は閉凸集合なので超平面が存在して C と $\boldsymbol{b}$ を厳密に分離する. すなわち, $\boldsymbol{a} \in \mathbb{R}^m$ と $c \in \mathbb{R}$ があって

$$\boldsymbol{z} \in C \Rightarrow \langle \boldsymbol{a}, \boldsymbol{z}\rangle_{2,m} + c > 0 \tag{D.18}$$

かつ

$$\langle \boldsymbol{a}, \boldsymbol{b}\rangle_{2,m} + c < 0. \tag{D.19}$$

$\boldsymbol{x} \geq \mathbf{0}_n$, $\lambda > 0$ とし

$$\boldsymbol{z} := \lambda \mathbf{A}\boldsymbol{x} = \mathbf{A}(\lambda \boldsymbol{x}) \in C$$

となる. (D.18) より

$$\langle \boldsymbol{a},\, \lambda \mathbf{A}\boldsymbol{x} \rangle_{2,n} + c > 0.$$

$\lambda > 0$ で割って $\lambda \to \infty$ とすれば

$$0 \leq \langle \boldsymbol{a},\, \mathbf{A}\boldsymbol{x} \rangle_{2,n} = \langle \mathbf{A}^\top \boldsymbol{a},\, \boldsymbol{x} \rangle_{2,n} \quad (\forall \boldsymbol{x} \geq \mathbf{0}_n)$$

となる. したがって

$$\mathbf{A}^\top \boldsymbol{a} \geq \mathbf{0}_n$$

となる. いま $\boldsymbol{y} = \boldsymbol{a}$ とおけば, $\mathbf{A}^\top \boldsymbol{y} \geq \mathbf{0}_n$ となる. (D.18) において $\boldsymbol{z} = \mathbf{0}_n \in C$ とおけば

$$c > 0$$

となる. さらに, (D.19) より

$$\langle \boldsymbol{b},\, \boldsymbol{y} \rangle_{2,m} = \langle \boldsymbol{b},\, \boldsymbol{a} \rangle_{2,m} < -c < 0$$

となる. よって $\langle \boldsymbol{b},\, \boldsymbol{y} \rangle_{2,n} < 0$　となる. よって, (1) が起こらなければ (2) が起こる. 対偶をとれば (2) が起こらなければ (1) が起こる. □

記法: $\mathbf{A}$ を $m \times n$ 行列とする.

$$\begin{aligned}
\ker(\mathbf{A}) &:= \{\boldsymbol{x};\, \mathbf{A}\boldsymbol{x} = \mathbf{0}_m\} \subset \mathbb{R}^n, \\
\mathrm{range}(\mathbf{A}) &:= \{\mathbf{A}\boldsymbol{x};\, \boldsymbol{x} \in \mathbb{R}^m\} \subset \mathbb{R}^m.
\end{aligned}$$

$S \subset \mathbb{R}^m$ としたとき

$$S^\perp := \{\boldsymbol{x} \subset \mathbb{R}^m;\, \langle \boldsymbol{x},\, \boldsymbol{s} \rangle_{2,m} = 0\, (\forall \boldsymbol{s} \in S)\}$$

とする.

定理 D.21. $\mathbf{A}$ を $m \times n$ 行列, $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{b} \in \mathbb{R}^m$ とする. このとき次は同値である.

(1) $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$ は解 $\boldsymbol{x}^*$ を持つ.

(2) $\boldsymbol{b} \in \ker\left(\mathbf{A}^\top\right)^\perp$ である.

Proof. $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$ が解を持つための必要十分条件は

$$\widetilde{\mathbf{A}} = [\mathbf{A},\, -\mathbf{A}] \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix} = \boldsymbol{b}, \quad \boldsymbol{u} \geq \mathbf{0}_n,\, \boldsymbol{v} \geq \mathbf{0}_n$$

は解を持つ．これは

$$\widetilde{\mathbf{A}} = \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix} = \boldsymbol{b}, \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix} \geq \mathbf{0}_{2n}$$

とした Farkas(1) である．よって

$$\text{「}\mathbf{A}\boldsymbol{x} = \boldsymbol{b} \text{ は解 } \boldsymbol{x}^* \text{ を持つ」} \Leftrightarrow \text{「Farkas (1) が成立」} \tag{D.20}$$

となる．

一方 Farkas(2) が成立するとき

$$\widetilde{\mathbf{A}}^\top \boldsymbol{y} \geq \mathbf{0}_n, \quad \langle \boldsymbol{y}, \boldsymbol{b} \rangle_{2,n} < 0$$

が解を持つ．これは

$$\begin{bmatrix} \mathbf{A}^\top \\ -\mathbf{A}^\top \end{bmatrix} \boldsymbol{y} \geq \begin{bmatrix} \mathbf{0}_n \\ \mathbf{0}_n \end{bmatrix}$$

である．よって $\mathbf{A}^\top \boldsymbol{y} \geq \mathbf{0}_n$ かつ $-\mathbf{A}^\top \boldsymbol{y} \geq \mathbf{0}_n$ より $\mathbf{A}^\top \boldsymbol{y} = \mathbf{0}_n$ なので

$$\mathbf{A}^\top \boldsymbol{y} = \mathbf{0}_n, \quad \langle \boldsymbol{y}, \boldsymbol{b} \rangle_{2,n} < 0$$

は解をもつ．

以上の議論から

1. 「$\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$ は解 $\boldsymbol{x}^*$ を持つ」が成立する．
2. $\widetilde{\mathbf{A}}^\top \boldsymbol{y} \geq \mathbf{0}_n, \quad \langle \boldsymbol{y}, \boldsymbol{b} \rangle_{2,n} < 0$ は解をもつ．

のいずれかが成立 (同時には起こらない) することがわかる．

さらに

$$\text{Farkas (2) が起こらない } \Leftrightarrow \boldsymbol{b} \in \ker\left(\mathbf{A}^\top\right)^\perp$$

である．なぜならば $\boldsymbol{b} \in \ker\left(\mathbf{A}^\top\right)^\perp$ ならば

$$\mathbf{A}^\top \boldsymbol{y} = \mathbf{0}_n \quad \text{かつ} \quad \langle \boldsymbol{y}, \boldsymbol{b} \rangle_{2,n} = 0$$

なので (2) は成立しない．逆に (2) が成立しないならば Farkas(1) が成立する．(D.20) より

$$\text{「Farkas (1) が起こる」} \Leftrightarrow \text{「}\mathbf{A}\boldsymbol{x} = \boldsymbol{b} \text{ は解 } \boldsymbol{x}^* \text{ を持つ」}$$

なので $\boldsymbol{b} = \mathbf{A}\boldsymbol{x}^*$ と書ける．よって $\forall \boldsymbol{y} \in \ker\left(\mathbf{A}^\top\right)$ に対して

$$\langle \boldsymbol{y}, \boldsymbol{b} \rangle_{2,n} = \langle \boldsymbol{y}, \mathbf{A}\boldsymbol{x}^* \rangle_{2,n} = \langle \mathbf{A}^\top \boldsymbol{y}, \boldsymbol{x}^* \rangle_{2,n} = 0$$

から $\boldsymbol{b} \in \ker\bigl(\mathbf{A}^\top\bigr)^\perp$ がわかる.
　以上の議論から

$$
\begin{aligned}
\text{「}\mathbf{A}\boldsymbol{x} = \boldsymbol{b} \text{ は解 } \boldsymbol{x}^* \text{ を持つ」} &\Leftrightarrow \text{「Farkas(1) が成立」} \\
&\Leftrightarrow \text{ Farkas(2) が成立しない」} \\
&\Leftrightarrow \boldsymbol{b} \in \ker\bigl(\mathbf{A}^\top\bigr)^\perp
\end{aligned}
$$

より定理は証明された. □

D.3　共役凸関数と Young の不等式

定義 D.22. 関数 $f : \mathbb{R}^n \to \mathbb{R}$ が凸であるとは, $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ と $0 \leq \lambda \leq 1$ に対して

$$
f\bigl(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}\bigr) \leq \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y})
$$

が成り立つときをいう.

注意 D.23. 関数 $f : \mathbb{R}^n \to \mathbb{R}$ が凸であるための必要十分条件はそのエピグラフ

$$
E = \left\{ \begin{bmatrix} \boldsymbol{x} \\ y \end{bmatrix} : y \geq f(\boldsymbol{x}),\, \boldsymbol{x} \in \mathbb{R}^n \right\} \subset \mathbb{R}^{n+1}
$$

が凸集合となることである.
　$m \in \mathbb{N}$ とする. 関数 f が凸のとき, 帰納法により $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m \in \mathbb{R}^n$ と $\lambda_1, \lambda_2, \ldots, \lambda_m \geq 0$ で $\sum_{k=1}^m \lambda_k = 1$ に対して

$$
f\left(\sum_{k=1}^m \lambda_k \boldsymbol{x}_k\right) \leq \sum_{k=1}^m \lambda_k f(\boldsymbol{x}_k)
$$

が成り立つ. □

定理 D.24. (1) 関数 $f : \mathbb{R}^n \to \mathbb{R}$ は連続微分可能とする. 関数 f が凸となるための必要十分条件は $\forall \boldsymbol{x}, \boldsymbol{x}_0 \in \mathbb{R}^n$ に対して

$$
f(\boldsymbol{x}_0) + \bigl\langle \nabla f(\boldsymbol{x}_0), \boldsymbol{x} - \boldsymbol{x}_0 \bigr\rangle_{2,n} \leq f(\boldsymbol{x})
$$

が成り立つことである. ただし

$$
\begin{aligned}
&\nabla f(\boldsymbol{x}) = \left(\frac{\partial f}{\partial x_1}(\boldsymbol{x}), \frac{\partial f}{\partial x_2}(\boldsymbol{x}), \ldots, \frac{\partial f}{\partial x_n}(\boldsymbol{x}) \right)^\top, \\
&\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^\top
\end{aligned}
$$

である.

(2) 関数 $f:\mathbb{R}^n\to\mathbb{R}$ は 2 回連続微分可能とする. 関数 f が凸であるための必要十分条件は $\forall\boldsymbol{x}\in\mathbb{R}^n$ に対して

$$\nabla^2 f(\boldsymbol{x})\succeq\boldsymbol{0}_{n\times n}$$

が成り立つことである. ただし, $\nabla^2 f(\boldsymbol{x})\succeq\boldsymbol{0}_{n\times n}$ は $\forall\boldsymbol{y}=(y_1,\,y_2,\,\ldots,\,y_n)^\top\in\mathbb{R}^n$ に対して

$$\boldsymbol{y}^\top\nabla^2 f(\boldsymbol{x})\boldsymbol{y}=\sum_{i,\,j=1}^{n}\frac{\partial f^2}{\partial x_i\partial x_j}(\boldsymbol{x})y_iy_j\geq 0$$

が成り立つことである.

Proof. まず, 関数 $f:\mathbb{R}^n\to\mathbb{R}$ は凸であるための必要十分条件は $\forall\boldsymbol{x},\,\boldsymbol{y}\in\mathbb{R}^n$ に対して, 関数

$$\phi(t):=f(\boldsymbol{x}+t\boldsymbol{y})\qquad(t\in\mathbb{R})\tag{D.21}$$

が凸であることである.

まず, 関数 f は凸とし, 関数 ϕ は凸であることを示す. そのために, $\boldsymbol{x},\,\boldsymbol{y}\in\mathbb{R}^n$, $0\leq\lambda\leq 1$, $s,\,t\in\mathbb{R}$ を取り,

$$\boldsymbol{w}:=\boldsymbol{x}+s\boldsymbol{y},\quad \boldsymbol{z}:=\boldsymbol{x}+t\boldsymbol{y}$$

とおく. このとき

$$f\bigl(\lambda\boldsymbol{w}+(1-\lambda)\boldsymbol{z}\bigr)\leq\lambda f(\boldsymbol{w})+(1-\lambda)f(\boldsymbol{z})=\lambda\phi(s)+(1-\lambda)\phi(t)\tag{D.22}$$

となる. また

$$f\bigl(\lambda\boldsymbol{w}+(1-\lambda)\boldsymbol{z}\bigr)=f\bigl(\boldsymbol{x}+\{\lambda t+(1-\lambda)s\}\boldsymbol{y}\bigr)=\phi(\lambda t+(1-\lambda)s)\tag{D.23}$$

となる. よって, (D.22) と (D.23) を考慮すると ϕ は凸関数であることがわかる.

逆を示す. $0\leq\lambda\leq 1, \boldsymbol{x},\,\boldsymbol{y}\in\mathbb{R}^n$ とする. (D.21) において, $\boldsymbol{x},\,\boldsymbol{y}$ は任意だったので, $\boldsymbol{y}$ と $\boldsymbol{x}-\boldsymbol{y}$ とする. すると

$$\phi(t)=f\bigl(\boldsymbol{y}+t(\boldsymbol{x}-\boldsymbol{y})\bigr)$$

とおく. 関数 ϕ は凸だから

$$\begin{aligned}f\bigl(\lambda\boldsymbol{x}+(1-\lambda)\boldsymbol{y}\bigr)&=f\bigl(\boldsymbol{y}+\lambda(\boldsymbol{x}-\boldsymbol{y})\bigr)=\phi(\lambda)\\&=\phi\bigl((1-\lambda)\times 0+\lambda\times 1\bigr)\\&\leq(1-\lambda)\phi(0)+\lambda\phi(1)\\&=(1-\lambda)f(\boldsymbol{y})+\lambda f(\boldsymbol{x})\end{aligned}$$

となる．よって $\forall \boldsymbol{x},\, \boldsymbol{y} \in \mathbb{R}^n$ に対して関数

$$\mathbb{R} \ni t \mapsto \phi(t) = f(\boldsymbol{x} + t\boldsymbol{y})$$

が凸関数ならば f も凸関数であることがわかる．
2．凸関数 $\phi(t) = f\big(\boldsymbol{x}_0 + t\boldsymbol{y}\big)$ に対して

$$\phi(t) + \dot{\phi}(t)(s-t) \leq \phi(s) \quad (\forall s,\, t \in \mathbb{R})$$

である．ただし $\dot{\phi}(t) = \dfrac{d\phi}{dt} = \langle \nabla f\big(\boldsymbol{x}_0 + t\boldsymbol{y}\big),\, \boldsymbol{y} \rangle_{2,n}$ である．ここで $s = 1,\, t = 0$ とおくと

$$\phi(0) + \dot{\phi}(0) \leq \phi(1) \iff f(\boldsymbol{x}_0) + \big\langle \nabla f(\boldsymbol{x}_0),\, \boldsymbol{y} \big\rangle_{2,n} \leq f\big(\boldsymbol{x}_0 + \boldsymbol{y}\big)$$

となる．上式に $\boldsymbol{y} = \boldsymbol{x} - \boldsymbol{x}_0$ を代入すると

$$f(\boldsymbol{x}_0) + \big\langle \nabla f(\boldsymbol{x}_0),\, \boldsymbol{x} - \boldsymbol{x}_0 \big\rangle_{2,n} \leq f(\boldsymbol{x})$$

を得る．
3．凸関数 $\phi(t) = f\big(\boldsymbol{x}_0 + t\boldsymbol{y}\big)$ は

$$\ddot{\phi}(t) \geq 0$$

をみたす．このとき

$$\dot{\phi}(t) = \big\langle \nabla f\big(\boldsymbol{x}_0 + t\boldsymbol{y}\big),\, \boldsymbol{y} \big\rangle_{2,n}, \quad \ddot{\phi}(t) = \boldsymbol{y}^\top \nabla^2 f\big(\boldsymbol{x}_0 + t\boldsymbol{y}\big)\boldsymbol{y}$$

となる．ただし $\nabla^2 f(\boldsymbol{x}) = \left(\dfrac{\partial^2 f}{\partial x_i \partial x_j}\right)_{i,j=1,2,\ldots,n}$ である．$t = 0$ とおくと

$$\boldsymbol{y}^\top \nabla^2 f(\boldsymbol{x}_0)\boldsymbol{y} \geq 0 \quad (\forall y \in \mathbb{R}^n)$$

がわかる．　□

定義 D.25. 関数 $f : \mathbb{R}^n \to \mathbb{R}$ はとする．各 $\boldsymbol{x} \in \mathbb{R}^n$ に対して

$$\partial f(\boldsymbol{x}) = \big\{ \boldsymbol{r} \in \mathbb{R}^n;\, f(\boldsymbol{x}) + \langle \boldsymbol{r},\, \boldsymbol{y} - \boldsymbol{x} \rangle_{2,n} \leq f(\boldsymbol{y})\, (\forall \boldsymbol{y} \in \mathbb{R}^n) \big\}$$

とおく．この集合を点 $\boldsymbol{x}$ における凸関数 f の**劣微分**という．

注意 D.26. 凸関数 $f : \mathbb{R}^n \to \mathbb{R}$ は点 $\boldsymbol{x}$ で連続微分可能なとき

$$\partial f(\boldsymbol{x}) = \{\nabla f(\boldsymbol{x})\}$$

となる．一般に，$\partial f(\boldsymbol{x})$ は 1 つの元よりも多い元をもつことがある．　□

例 D.27. (1) $n=1$, $f(x)=|x|$ とする．このとき

$$\partial f(x)=\begin{cases}\{1\} & (x>0)\\ [-1,\,1] & (x=0)\\ \{-1\} & (x<0)\end{cases}$$

となる．
(2) $n>1$, $f(\boldsymbol{x})=|\boldsymbol{x}|_{2,\,n}$ とする．このとき

$$\partial f(\boldsymbol{x})=\begin{cases}\left\{\dfrac{\boldsymbol{x}}{|\boldsymbol{x}|_{2,\,n}}\right\} & (\boldsymbol{x}\neq\boldsymbol{0}_n)\\ B(1) & (\boldsymbol{x}=\boldsymbol{0}_n)\end{cases}$$

となる．ただし $B(1):=\{\boldsymbol{x}\in\mathbb{R}^n;\,|\boldsymbol{x}|_{2,\,n}=1\}$ である．
　実際, $\forall\boldsymbol{r}\in\partial f(\boldsymbol{0}_n)$ は $\forall\boldsymbol{x}\in\mathbb{R}^n$ に対して

$$|\boldsymbol{0}_n|_{2,\,n}+\langle\boldsymbol{r},\,\boldsymbol{x}-\boldsymbol{0}_n\rangle_{2,\,n}\leq|\boldsymbol{x}|_{2,\,n}$$

を意味する．$|\langle\boldsymbol{r},\,\boldsymbol{x}\rangle_{2,\,n}|\leq|\boldsymbol{r}|_{2,\,n}\times|\boldsymbol{x}|_{2,\,n}$ から

$$\langle\boldsymbol{r},\,\boldsymbol{x}\rangle_{2,\,n}\leq|\boldsymbol{x}|_{2,\,n}\ \Leftrightarrow\ |\boldsymbol{r}|_{2,\,n}\leq 1$$

となる．　□

定理 D.28. 関数 $f:\mathbb{R}^n\to\mathbb{R}$ は凸とする．各 $\boldsymbol{x}\in\mathbb{R}^n$ に対して

$$\partial f(\boldsymbol{x})\text{ は凸閉集合}$$

である．

Proof. 明らかに $\partial f(\boldsymbol{x})$ は凸集合である．これが閉集合であることを示すために

$$\{\boldsymbol{r}_k\}_{k=1}^{\infty}\subset\partial f(\boldsymbol{x})\quad\text{かつ}\quad\boldsymbol{r}_0=\lim_{k\to\infty}\boldsymbol{r}_k$$

を取る．$\boldsymbol{r}_k\in\partial f(\boldsymbol{x})\,(k=1,\,2,\,\ldots)$ だから $\forall\boldsymbol{y}\in\mathbb{R}^n$ に対して

$$f(\boldsymbol{x})+\langle\boldsymbol{r}_k,\,\boldsymbol{y}-\boldsymbol{x}\rangle_{2,\,n}\leq f(\boldsymbol{x})$$

となる．$k\to\infty$ とすれば, $\forall\boldsymbol{y}\in\mathbb{R}^n$ に対して

$$f(\boldsymbol{x})+\langle\boldsymbol{r}_0,\,\boldsymbol{y}-\boldsymbol{x}\rangle_{2,\,n}\leq f(\boldsymbol{y})$$

となるから $\boldsymbol{r}_0\in\partial f(\boldsymbol{x})$ となる．よって $\partial f(\boldsymbol{x})$ は閉集合となる．　□

定理 D.29. 関数 $f:\mathbb{R}^n\to\mathbb{R}$ は凸とする．このとき, 各 $\boldsymbol{x}\in\mathbb{R}^n$ に対して $\partial f(\boldsymbol{x})\neq\varnothing$ となる．

Proof. 凸関数 f のエピグラフ[3]E は凸閉集合なので正の整数 k に対して

$$\boldsymbol{e}_k := \begin{bmatrix} \boldsymbol{x} \\ f(\boldsymbol{x}) - \dfrac{1}{k} \end{bmatrix} \notin E$$

である. 分離定理 (定理 D.15) よりある超平面

$$\langle \boldsymbol{a}_k,\, \boldsymbol{x} \rangle_{2,\,n} + d_k \,(\boldsymbol{a}_k \in \mathbb{R}^{n+1},\, \mathrm{d}_k \in \mathbb{R})$$

は $\boldsymbol{e}_k$ と E を厳密に分離する.

$$\boldsymbol{a}_k = \begin{bmatrix} \boldsymbol{b}_k \\ c_k \end{bmatrix}, \quad \boldsymbol{b}_k \in \mathbb{R}^n, \quad c_k \in \mathbb{R}$$

と書くと

$$\langle \boldsymbol{a}_k,\, \boldsymbol{e}_k \rangle_{2,\,n} + d_k < 0, \tag{D.24}$$

$$\langle \boldsymbol{a}_k,\, \boldsymbol{z} \rangle_{2,\,n} + d_k > 0 \qquad (\forall \boldsymbol{z} \in E) \tag{D.25}$$

となる. (D.24) は

$$\langle \boldsymbol{b}_k,\, \boldsymbol{x} \rangle_{2,\,n} + c_k \left(f(\boldsymbol{x}) - \frac{1}{k} \right) + d_k < 0 \tag{D.26}$$

となる. $\forall \boldsymbol{y} \in \mathbb{R}^n$ に対して

$$\boldsymbol{z} = \begin{bmatrix} \boldsymbol{y} \\ f(\boldsymbol{y}) \end{bmatrix} \in E$$

なので

$$\langle \boldsymbol{b}_k,\, \boldsymbol{y} \rangle_{2,\,n} + c_k f(\boldsymbol{y}) + d_k > 0 \tag{D.27}$$

となる. (D.26) を -1 倍して (D.27) に加わえると $\forall \boldsymbol{y} \in \mathbb{R}^n$ に対して

$$\langle \boldsymbol{b}_k,\, \boldsymbol{y} - \boldsymbol{x} \rangle_{2,\,n} + c_k \left(f(\boldsymbol{y}) - f(\boldsymbol{x}) + \frac{1}{k} \right) > 0 \tag{D.28}$$

となる. (D.28) に $\boldsymbol{y} = \boldsymbol{x}$ を代入すると

$$\frac{c_k}{k} > 0 \quad \Rightarrow c_k > 0$$

となる. したがって (D.28) は $\forall \boldsymbol{y} \in \mathbb{R}^n$ に対して

$$f(\boldsymbol{y}) + \frac{1}{k} \geq -\left\langle \frac{\boldsymbol{b}_k}{c_k},\, \boldsymbol{y} - \boldsymbol{x} \right\rangle_{2,\,n} + f(\boldsymbol{x}) \tag{D.29}$$

[3]epigraph の epi は「上の」という意味の接頭語.

を意味する．ここで

$$\boldsymbol{r}_k = -\frac{1}{c_k}\boldsymbol{b}_k$$

とおく．

次に $\{\boldsymbol{r}_k\}_{k=1}^{\infty} \subset \mathbb{R}^n$ は有界であることを示す．$|\boldsymbol{r}_k|_{2,\,n} \neq 0$ ならば (D.29) において

$$\boldsymbol{y} = \boldsymbol{x} + \frac{\boldsymbol{r}_k}{|\boldsymbol{r}_k|_{2,\,n}}$$

とおく[4]と

$$f\left(\boldsymbol{x} + \frac{\boldsymbol{r}_k}{|\boldsymbol{r}_k|_{2,\,n}}\right) + \frac{1}{k} \geq |\boldsymbol{r}_k|_{2,\,n} + f(\boldsymbol{x})$$

となる．よって $k = 1,\, 2,\, \ldots$ に対して

$$|\boldsymbol{r}_k|_{2,\,n} \leq 1 + \max_{|\boldsymbol{u}|_{2,\,n}=1} \bigl|f(\boldsymbol{x}+\boldsymbol{u})\bigr| + f(\boldsymbol{x}) =: M$$

となる．よって $\{\boldsymbol{r}_k\}_{k=1}^{\infty}$ は有界であることがわかる．

Bolzano-Weirstrass の定理より収束する部分列 $\{\boldsymbol{r}_{k(j)}\}_{j=1}^{\infty}$ が取れる．

$$\boldsymbol{r}_0 := \lim_{j\to\infty} \boldsymbol{r}_{k(j)}$$

とおく．(D.29) において $k = k_j \xrightarrow{j\to\infty} \infty$ とすれば

$$f(\boldsymbol{y}) \geq \langle \boldsymbol{r}_0,\, \boldsymbol{y}-\boldsymbol{x}\rangle_{2,\,n} + f(\boldsymbol{x}) \quad (\forall \boldsymbol{y} \in \mathbb{R}^n)$$

を得る．よって $\boldsymbol{r}_0 \in \partial f(\boldsymbol{x})$ がわかる．したがって $\partial f(\boldsymbol{x}) \neq \varnothing$ が示せた．
□

以降, この節では関数 $f : \mathbb{R}^n \to \mathbb{R}$ は凸で

$$\lim_{|\boldsymbol{x}|_{2,\,n}\to\infty} \frac{f(\boldsymbol{x})}{|\boldsymbol{x}|_{2,\,n}} = +\infty \tag{D.30}$$

をみたすとする．これを**超線型成長条件**という．

定義 D.30. $\boldsymbol{y} \in \mathbb{R}^n$ に対し

$$f^*(\boldsymbol{y}) = \max_{\boldsymbol{x}\in\mathbb{R}^n}\bigl\{\langle \boldsymbol{x},\, \boldsymbol{y}\rangle_{2,\,n} - f(\boldsymbol{x})\bigr\}$$

とする．f^* を f **の凸共役関数**という．

[4]定義から $|\boldsymbol{b}_k|^2 = c_k^2|\boldsymbol{r}_k|^2$ である．

$$\left\langle \frac{\boldsymbol{b}_k}{c_k},\, \frac{\boldsymbol{r}_k}{|\boldsymbol{r}_k|_{2,\,n}} \right\rangle_{2,\,n} = \left\langle \frac{\boldsymbol{b}_k}{c_k},\, -\frac{\boldsymbol{b}_k}{c_k|\boldsymbol{r}_k|_{2,\,n}} \right\rangle_{2,\,n} = -\frac{|\boldsymbol{b}_k|_{2,\,n}^2}{c_k^2|\boldsymbol{r}_k|_{2,\,n}} = -|\boldsymbol{r}_k|_{2,\,n}$$

になることに注意する．

注意 D.31. $n=1,\ f(x)=x^2/2\,(x\in\mathbb{R})$ とする．このとき $y\in\mathbb{R}$ に対し

$$f^*(y)=\max_x\left(xy-\frac{x^2}{2}\right)=\frac{y^2}{2}$$

となる．
(2) $n=1,\ 1<p<\infty,\ f(x)=|x|^p/p$ とする．このとき $y\in\mathbb{R}$ に対し

$$f^*(y)=\max_x\left(xy-\frac{|x|^p}{p}\right)=\frac{|y|^q}{q},\qquad \frac{1}{p}+\frac{1}{q}=1$$

となる．
∵) y を固定し

$$g(x)=xy-\frac{|x|^p}{p}$$

とする．すると $\dot{g}(x)=y-|x|^{p-1}\mathrm{sgn}(x)$ となる．ただし

$$\mathrm{sgn}(x)=\begin{cases}1 & (x>0)\\ -1 & (x<0)\\ 0 & (x=0)\end{cases}$$

とした．したがって

$$0=\dot{g}(x)\ \Leftrightarrow\ y=|x|^{p-1}\mathrm{sgn}(x)\ \Rightarrow\ x=|y|^{1/(p-1)}\mathrm{sgn}(y)$$

となる．結果

$$\begin{aligned}f^*(y)&=\left(|y|^{1/(p-1)}\mathrm{sgn}(y)\right)y-\frac{\left||y|^{1/(p-1)}\mathrm{sgn}(y)\right|^p}{p}=|y|^{\frac{1}{1-p}+1}-\frac{|y|^{\frac{p}{p-1}}}{p}\\&=\left(1-\frac{1}{p}\right)|y|^{\frac{p}{p-1}}=\frac{|y|^q}{q}\end{aligned}$$

がわかる． □

補題 D.32. $\boldsymbol{x},\,\boldsymbol{y}\in\mathbb{R}^n$ に対し

$$\langle\boldsymbol{x},\,\boldsymbol{y}\rangle_{2,n}\le f(\boldsymbol{x})+f^*(\boldsymbol{y})\tag{D.31}$$

である．これを Fenchel-Young **双対**という．

Proof. 双対関数の定義 (定義 D.30) よりわかる． □

定理 D.33. (1) $f^*:\mathbb{R}^n\to\mathbb{R}$ は凸関数である．
(2) $\displaystyle\lim_{|\boldsymbol{y}|_{2,n}\to\infty}\frac{|f^*(\boldsymbol{y})|_{2,n}}{|\boldsymbol{y}|_{2,n}}=+\infty$ となる．
(3) $f^{**}=f$ である．

Proof. 任意の $\boldsymbol{y}$, $\boldsymbol{w} \in \mathbb{R}^n$, $0 \leq \lambda \leq 1$ に対し

$$
\begin{aligned}
f^*\bigl(\lambda \boldsymbol{y} + (1-\lambda)\boldsymbol{w}\bigr) &= \max_{\boldsymbol{x} \in \mathbb{R}^n} \Bigl(\langle \lambda \boldsymbol{y} + (1-\lambda)\boldsymbol{w},\, \boldsymbol{x} \rangle_{2,\,n} - f(\boldsymbol{x}) \Bigr) \\
&= \max_{\boldsymbol{x} \in \mathbb{R}^n} \Bigl(\lambda \bigl\{ \langle \boldsymbol{y},\, \boldsymbol{x}[rangle_{2,\,n} - f(\boldsymbol{x}) \bigr\} \\
&\qquad + (1-\lambda) \bigl\{ \lambda \langle \boldsymbol{w},\, \boldsymbol{x} \rangle_{2,\,n} - f(\boldsymbol{x}) \bigr\} \Bigr) \\
&\leq \lambda \max_{\boldsymbol{x} \in \mathbb{R}^n} \Bigl(\langle \boldsymbol{y},\, \boldsymbol{x} \rangle_{2,\,n} - f(\boldsymbol{x}) \Bigr) \\
&\qquad + (1-\lambda) \max_{\boldsymbol{x} \in \mathbb{R}^n} \Bigl(\langle \boldsymbol{w},\, \boldsymbol{x} \rangle_{2,\,n} - f(\boldsymbol{x}) \Bigr) \\
&= \lambda f^*(\boldsymbol{y}) + (1-\lambda) f^*(\boldsymbol{w})
\end{aligned}
$$

がわかる．したがって 関数 f^* は凸である．
(2)．任意の $\boldsymbol{x}$, $\boldsymbol{y} \in \mathbb{R}^n$ に対して

$$
f(\boldsymbol{x}) + f^*(\boldsymbol{y}^*) \geq \langle \boldsymbol{x},\, \boldsymbol{y} \rangle_{2,\,n}
$$

であった．ここで, $\boldsymbol{y} \neq \boldsymbol{0}$ と $\lambda > 0$ に対し

$$
\boldsymbol{x} = \frac{\lambda \boldsymbol{y}}{|\boldsymbol{y}|_{2,\,n}}
$$

とおく．このとき

$$
f^*(\boldsymbol{y}) \geq \left\langle \frac{\lambda \boldsymbol{y}}{|\boldsymbol{y}|_{2,\,n}},\, \boldsymbol{y} \right\rangle_{2,\,n} - f\left(\frac{\lambda \boldsymbol{y}}{|\boldsymbol{y}|_{2,\,n}} \right) \geq \lambda |\boldsymbol{y}|_{2,\,n} - \max_{|\boldsymbol{u}|_{2,\,n} = \lambda} f(\boldsymbol{u}).
$$

となることがわかる．したがって

$$
\frac{f^*(\boldsymbol{y})}{|\boldsymbol{y}|_{2,\,n}} \geq \lambda - \frac{1}{|\boldsymbol{y}|_{2,\,n}} \max_{|\boldsymbol{u}|_{2,\,n} = \lambda} f(\boldsymbol{u})
$$

となる．ここで

$$
\liminf_{|\boldsymbol{y}|_{2,\,n} \to \infty} \frac{f^*(\boldsymbol{y})}{|\boldsymbol{y}|_{2,\,n}} \geq \lambda > 0
$$

を得る．最後に $\lambda \to \infty$ とすれば, (2) がわかる．
(3) まず, $f(\boldsymbol{x}) + f^*(\boldsymbol{y}) \geq \langle \boldsymbol{x},\, \boldsymbol{y} \rangle_{2,\,n}$ から

$$
f(\boldsymbol{x}) \geq \max_{\boldsymbol{y} \in \mathbb{R}^n} \Bigl(\langle \boldsymbol{x},\, \boldsymbol{y} \rangle_{2,\,n} - f^*(\boldsymbol{y}) \Bigr) = f^{**}(\boldsymbol{x}) \tag{D.32}
$$

となることに注意する．逆に, 定理 D.29 から $\partial f(\boldsymbol{x}) \neq \varnothing$ であることがわかる．このことから $\boldsymbol{r} \in \partial f(\boldsymbol{x})$ を取る．すると $\boldsymbol{z} \in \mathbb{R}^n$ に対して

$$
f(\boldsymbol{z}) \geq f(\boldsymbol{x}) + \langle \boldsymbol{r},\, \boldsymbol{z} - \boldsymbol{x} \rangle_{2,\,n}
$$

となる．上の式を書きかえると

$$\langle \boldsymbol{r},\, \boldsymbol{x}\rangle_{2,n} - f(\boldsymbol{x}) \geq \langle \boldsymbol{r},\, \boldsymbol{z}\rangle_{2,n} - f(\boldsymbol{z})$$

を得る．よって

$$\langle \boldsymbol{r},\, \boldsymbol{x}\rangle_{2,n} - f(\boldsymbol{x}) = \max_{\boldsymbol{z}\in\mathbb{R}^n}\Bigl(\langle \boldsymbol{r},\, \boldsymbol{z}\rangle_{2,n} - f(\boldsymbol{z})\Bigr) = f^*(\boldsymbol{r}) \tag{D.33}$$

がわかる．さらに, (D.33) から

$$f^{**}(\boldsymbol{x}) = \max_{\boldsymbol{y}\in\mathbb{R}^n}\Bigl(\langle \boldsymbol{x},\, \boldsymbol{y}\rangle_{2,n} - f^*(\boldsymbol{y})\Bigr) \geq \langle \boldsymbol{x},\, \boldsymbol{r}\rangle_{2,n} - f^*(\boldsymbol{r}) = f(\boldsymbol{x}) \tag{D.34}$$

がわかる．よって, (D.32) と (D.34) から $f^{**} = f$ がわかる． □

定理 D.34. $\boldsymbol{x}$, $\boldsymbol{y} \in \mathbb{R}^n$ に対して次の条件は同値である．

(1) $\langle \boldsymbol{x},\, \boldsymbol{y}\rangle_{2,n} = f(\boldsymbol{x}) + f^*(\boldsymbol{y})$.

(2) $\boldsymbol{y} \in \partial f(\boldsymbol{x})$.

(3) $\boldsymbol{x} \in \partial f^*(\boldsymbol{y})$.

Proof. (1) $\Rightarrow$ (2) **の証明**．任意の $\boldsymbol{z} \in \mathbb{R}^n$ に対し

$$\langle \boldsymbol{x},\, \boldsymbol{y}\rangle_{2,n} - f(\boldsymbol{x}) = f^*(\boldsymbol{y}) \geq \langle \boldsymbol{y},\, \boldsymbol{z}\rangle_{2,n} - f(\boldsymbol{z})$$

となる．上の式の最左辺と最右辺から $\forall \boldsymbol{z} \in \mathbb{R}^n$ に対し

$$f(\boldsymbol{z}) \geq \langle \boldsymbol{y},\, \boldsymbol{z} - \boldsymbol{x}\rangle_{2,n} + f(\boldsymbol{x})$$

を得る．よって

$$\boldsymbol{y} \in \partial f(\boldsymbol{x})$$

が示せた．

(2) $\Rightarrow$ (1) **の証明**．$\boldsymbol{y} \in \partial f(\boldsymbol{x})$ のとき, $\forall \boldsymbol{z} \in \mathbb{R}^n$ に対し

$$f(\boldsymbol{z}) \geq \langle \boldsymbol{y},\, \boldsymbol{z} - \boldsymbol{x}\rangle_{2,n} + f(\boldsymbol{z}) \;\Rightarrow\; \langle \boldsymbol{x},\, \boldsymbol{y}\rangle_{2,n} - f(\boldsymbol{x}) \geq \langle \boldsymbol{y},\, \boldsymbol{z}\rangle_{2,n} - f(\boldsymbol{z})$$

となる．よって

$$\langle \boldsymbol{x},\, \boldsymbol{y}\rangle_{2,n} - f(\boldsymbol{x}) \geq \max_{\boldsymbol{z}\in\mathbb{R}^n}\Bigl(\langle \boldsymbol{y},\, \boldsymbol{z}\rangle_{2,n} - f(\boldsymbol{z})\Bigr) = f^*(\boldsymbol{y}) \tag{D.35}$$

から

$$\langle \boldsymbol{x},\, \boldsymbol{y}\rangle_{2,n} \geq f(\boldsymbol{x}) + f^*(\boldsymbol{y})$$

を得る. Fenchel-Young の不等式 (D.31) から

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{2,n} \leq f(\boldsymbol{x}) + f^*(\boldsymbol{y}) \tag{D.36}$$

である. よって, (D.35) と (D.36) を合わせると

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{2,n} = f(\boldsymbol{x}) + f^*(\boldsymbol{y})$$

が示せた.

(1) ⇔ (3) も同様に示すことができる. □

D.4　非線形最適化と KKT 条件

$n, p \in \mathbb{N}$ とし, $f, h_1, h_2, \ldots, h_p : \mathbb{R}^n \to \mathbb{R}$ は連続微分可能とする. 次の記号を導入する.

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \boldsymbol{h} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_p \end{bmatrix},$$

$$\nabla \boldsymbol{h} = \begin{bmatrix} (\nabla h_1)^\top \\ (\nabla h_2)^\top \\ \vdots \\ (\nabla h_p)^\top \end{bmatrix} = \begin{bmatrix} \dfrac{\partial h_1}{\partial x_1} & \dfrac{\partial h_1}{\partial x_2} & \cdots & \dfrac{\partial h_1}{\partial x_n} \\ \dfrac{\partial h_2}{\partial x_1} & \dfrac{\partial h_2}{\partial x_2} & \cdots & \dfrac{\partial h_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial h_p}{\partial x_1} & \dfrac{\partial h_p}{\partial x_2} & \cdots & \dfrac{\partial h_p}{\partial x_n} \end{bmatrix}.$$

この節では, 制約付く最適化問題

$$\text{minimize}\, f(\boldsymbol{x}) \qquad \text{subject to} \quad \boldsymbol{h} \leq \boldsymbol{0}_p \tag{D.37}$$

の解 $\boldsymbol{x}^*$ をみつけたい. ただし $\boldsymbol{h}(\boldsymbol{x}) \leq \boldsymbol{0}_p$ は $\forall j \in \{1, 2, \ldots, p\}$ に対して $h_j(\boldsymbol{x}) \leq 0$ が成立していることである.

$\boldsymbol{h}(\boldsymbol{x}) \leq \boldsymbol{0}_p$ をみたす点 $\boldsymbol{x}^*$ を**実行可能解**という. 実行可能解 $\boldsymbol{x}^*$ に対して $h_j(\boldsymbol{x}^*) = 0$ となる制約式を**有効制約式**という. 解 $\boldsymbol{x}^*$ に対する有効制約式の添え字集合を J と書く. すなわち

$$J = \{j \in \{1, 2, \ldots, p\};\, h_j(\boldsymbol{x}^*) = 0\}$$

である.

記号 関数 $\boldsymbol{r} : \mathbb{R} \to \mathbb{R}^n$ に対して

$$\boldsymbol{r} = o(t) \Leftrightarrow \lim_{t \to 0+} \frac{|\boldsymbol{r}(t)|_{2,n}}{t} = 0$$

と定める.

定義 D.35. $\boldsymbol{x}^*$ における制約 qualification 条件 (CQ 条件) が成立するとは

$$\left\langle \boldsymbol{y},\, \nabla h_j(\boldsymbol{x}) \right\rangle_{2,\,n} \leq 0 \qquad (j \in J) \tag{D.38}$$

をみたす $\boldsymbol{y} \in \mathbb{R}^n$ にたいしてある正数 t_0 と連続曲線 $\{\boldsymbol{x}(t),\, 0 \leq t < t_0\}$ があって

$$\boldsymbol{h}(\boldsymbol{x}) \leq \boldsymbol{0}_p \qquad (0 \leq t < t_0) \tag{D.39}$$

かつ

$$\boldsymbol{x}(t) = \boldsymbol{x}^* + t\boldsymbol{y} + o(t) \qquad (t \to 0+) \tag{D.40}$$

が成立するときをいう.

注意 D.36. 条件 (D.39) は $\boldsymbol{x}(t)\,(0 \leq t < t_0)$ は実行可能解であることを保証している. (D.40) は $\boldsymbol{x}(t)$ の $t = 0$ における右微分係数ベクトル $\dot{\boldsymbol{x}}\,(0)$ が存在し

$$\dot{\boldsymbol{x}}\,(0) = \boldsymbol{y}$$

であることを保証している. □

定理 D.37. (Karush-Kuhn-Tucker(KKT) 条件) $\boldsymbol{x}^*$ は問題 (D.37) の解とし, $\boldsymbol{x}^*$ において CQ 条件は成立しているとする. このとき実数 $\lambda_1^*,\, \lambda_2^*,\, \ldots,\, \lambda_p^*$ が存在して

$$\nabla f(\boldsymbol{x}^*) + \sum_{j=1}^{p} \lambda_j^* \nabla h_j(\boldsymbol{x}^*) = \boldsymbol{0}_n \tag{D.41}$$

が成立する. さらに $\boldsymbol{\lambda}^* := (\lambda_1^*,\, \lambda_2^*,\, \ldots,\, \lambda_p^*)^\top$ は

$$\boldsymbol{\lambda}^* \geq \boldsymbol{0}_p, \quad \left\langle \boldsymbol{\lambda}^*,\, \boldsymbol{h}(\boldsymbol{x}^*) \right\rangle_{2,\,p} = 0 \tag{D.42}$$

をみたしている;

注意 D.38. 条件 (D.41) と (D.42) を合わせて Karush-Kuhn-Tuckeer 条件 (KKT 条件) という. (D.42) は

$$\nabla f(\boldsymbol{x}^*) + \nabla \boldsymbol{h}(\boldsymbol{x}^*)^\top \boldsymbol{\lambda}^* = \boldsymbol{0}_n$$

とも表すことができる. $\boldsymbol{\lambda}^*$ は Lagrange 乗数という. $\left\langle \boldsymbol{\lambda}^*,\, \boldsymbol{h}(\boldsymbol{x}^*) \right\rangle_{2,\,n} = 0$ を相補性条件という. 定理の証明からわかることであるが $\lambda_j^* = 0$ か $h_j(\boldsymbol{x}^*) = 0$ のいずれかが成立しているのでこのように呼んでいる. □

Proof. **定理** D.37 **の証明**. ベクトル $\boldsymbol{y} \in \mathbb{R}^n$ は (D.40) をみたしているとする. CQ 条件で保証されている $\boldsymbol{y}$ に対応する曲線を $\{\boldsymbol{x}(t);\, 0 \le t < t_0\}$ とする. いま

$$\phi(t) := f(\boldsymbol{x}(t)) \tag{D.43}$$

とおく. このとき $\boldsymbol{x}^*$ は問題 (D.37) の解なので

$$\phi(0) = f(\boldsymbol{x}^*) \le f(\boldsymbol{x}(t)) = \phi(t) \quad (0 \le t < t_0)$$

が成り立つ. したがって ϕ は $t = 0$ で最小値を取るので

$$\overset{\bullet}{\phi}(0) \ge 0$$

となる. さらに, (D.43) を t に関して微分して, $t = 0$ を代入すると

$$\overset{\bullet}{\phi}(0) = \langle \nabla f(\boldsymbol{x}^*),\, \overset{\bullet}{\boldsymbol{x}}(0) \rangle_{2,n} = \langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{y} \rangle_{2,n}$$

を得る. この式から (D.38) をみたす $\boldsymbol{y} \in \mathbb{R}^n$ に対し

$$\langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{y} \rangle_{2,n} \ge 0$$

となるとなることがわかる. よって

$$\langle \boldsymbol{y},\, \nabla h_j(\boldsymbol{x}^*) \rangle_{2,n} \le 0\,(j \in J) \;\Rightarrow\; \langle \boldsymbol{y},\, \nabla f(\boldsymbol{x}^*) \rangle_{2,n} \ge 0 \tag{D.44}$$

を示せた. ここで Farkas の交代定理 (定理 D.20) を思い出す. 以下の (1) か (2) のいずれかは成立する.

(1) $\mathbf{A}\boldsymbol{x} = \boldsymbol{b},\, \boldsymbol{x} \ge \mathbf{0}_n$ は解 $\boldsymbol{x}$ を持つ.

(2) $\mathbf{A}^\top \boldsymbol{y} \ge \mathbf{0}_n,\, \langle \boldsymbol{y},\, \boldsymbol{b} \rangle_{2,n} < 0$ は解を持つ.

さらに (1) と (2) は同時に成立しない. ただし, $\mathbf{A}$ は $n \times n$ 行列である. $J = \{j_1,\, j_2,\, \ldots,\, j_k\}$ と書き

$$\begin{aligned} \mathbf{A} &= -\big[\nabla h_{j_1}(\boldsymbol{x}^*),\, \nabla h_{j_2}(\boldsymbol{x}^*),\, \ldots,\, \nabla h_{j_k}(\boldsymbol{x}^*)\big], \\ \boldsymbol{b} &= \nabla f(\boldsymbol{x}^*) \end{aligned}$$

に対して Farkas の交代定理を適用する. (D.44) は

$$\mathbf{A}^\top \boldsymbol{y} \ge \mathbf{0}_n \;\Rightarrow\; \langle \boldsymbol{y},\, \boldsymbol{b} \rangle_{2,n} \ge 0$$

を主張しているので, Farkas 交代定理の (2) は成立しない. 結果 Farkas 交代定理の (1) が成立する. つまり $\sigma_j \ge 0\,(j \in J)$ が存在して

$$-\sum_{j \in J} \sigma_j \nabla h_j(\boldsymbol{x}^*) = \nabla f(\boldsymbol{x}^*)$$

が成り立つ. そこで $\boldsymbol{\lambda}^* = (\lambda_1^*, \lambda_2^*, \ldots, \lambda_p^*)^\top$ を

$$\lambda_j^* = \begin{cases} \sigma_j & (j \in J) \\ 0 & (j \notin J) \end{cases}$$

と定めると $j \in J$ のとき $h_j(\boldsymbol{x}^*) = 0$, $j \notin J$ のとき, $\lambda_j^* = 0$ なので

$$\langle \boldsymbol{\lambda}^*, \boldsymbol{h}(\boldsymbol{x}^*) \rangle_{2,p} = 0$$

となる. さらに Frakas の交代定理の (1) から

$$\begin{aligned} &\boldsymbol{\lambda}^* \geq \mathbf{0}_p, \\ &\nabla f(\boldsymbol{x}^*) + \sum_{j=1}^{p} \lambda_j^* \nabla h_j(\boldsymbol{x}^*) = \mathbf{0}_n \end{aligned}$$

もわかる. □

注意 D.39. CQ 条件をみたす制約の例を上げておく

(1) $\{h_j\}_{j=1}^p$ は $\boldsymbol{x}$ の線型関数.

(2) $\{\nabla h_j(\boldsymbol{x}^*)\}_{j \in J}^p$ は $\mathbb{R}^n$ の線型独立なベクトル.

これらが CQ 条件をみたすことの証明は Evans の pp.115-117 を参照のこと. □

D.5 Lagrange 乗数についてのさらなる議論

$f, g_1, g_2, \ldots, g_m, h_1, h_2, \ldots, h_p : \mathbb{R}^n \to \mathbb{R}$ 連続微分可能とし

$$\boldsymbol{g} = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_m \end{bmatrix}, \quad \boldsymbol{h} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_p \end{bmatrix}$$

とおく. この節では, 下記の最小化問題

$$\text{minimize} \quad f(\boldsymbol{x}) \qquad \text{subject to} \quad \boldsymbol{g}(\boldsymbol{x}) = \mathbf{0}_m \text{ and } \boldsymbol{h}(\boldsymbol{x}) \leq \mathbf{0}_p. \tag{D.45}$$

を考える.

定理 D.40. $\boldsymbol{x}^* \in \mathbb{R}^n$ は問題 (D.45) の解とする．このとき実数の組

$$\gamma^*, \quad \boldsymbol{\mu}^* = (\mu_1^*, \mu_2^*, \ldots, \mu_m^*)^\top, \quad \boldsymbol{\lambda}^* = (\lambda_1^*, \quad \lambda_2^*, \ldots, \lambda_p^*)^\top$$

が存在し，これはすべては 0 に同時になることはなく，さらに

$$\gamma^* \nabla f(\boldsymbol{x}^*) + \sum_{k=1}^{m} \mu_k^* \nabla g_k(\boldsymbol{x}^*) + \sum_{j=1}^{p} \lambda_j^* \nabla h_j(\boldsymbol{x}^*) = \boldsymbol{0}_n \tag{D.46}$$

と

$$\gamma^* \geq 0, \quad \boldsymbol{\lambda} \geq \boldsymbol{0}_p, \quad \langle \boldsymbol{\lambda}, \boldsymbol{h}(\boldsymbol{x}^*) \rangle_{2,p} = 0 \tag{D.47}$$

をみたす．

注意 D.41. (1) (D.46) は

$$\gamma^* \nabla f(\boldsymbol{x}^*) + \nabla \boldsymbol{g}(\boldsymbol{x}^*)^\top \boldsymbol{\mu}^* + \nabla \boldsymbol{h}(\boldsymbol{x}^*) \boldsymbol{\lambda}^* = \boldsymbol{0}_n$$

と書き直すことができる．ただし

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \boldsymbol{h} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_p \end{bmatrix}, \quad \boldsymbol{g} = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_m \end{bmatrix},$$

$$\nabla \boldsymbol{h} = \begin{bmatrix} (\nabla h_1)^\top \\ (\nabla h_2)^\top \\ \vdots \\ (\nabla h_p)^\top \end{bmatrix} = \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \frac{\partial h_1}{\partial x_2} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \frac{\partial h_2}{\partial x_1} & \frac{\partial h_2}{\partial x_2} & \cdots & \frac{\partial h_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_p}{\partial x_1} & \frac{\partial h_p}{\partial x_2} & \cdots & \frac{\partial h_p}{\partial x_n} \end{bmatrix},$$

$$\nabla \boldsymbol{g} = \begin{bmatrix} (\nabla g_1)^\top \\ (\nabla g_2)^\top \\ \vdots \\ (\nabla g_m)^\top \end{bmatrix} = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1} & \frac{\partial g_m}{\partial x_2} & \cdots & \frac{\partial g_m}{\partial x_n} \end{bmatrix}$$

である．

(2) $\gamma^* = 1$ のとき (D.46) と (D.47) は

$$\nabla f(\boldsymbol{x}^*) + \nabla \boldsymbol{g}(\boldsymbol{x}^*)^\top \boldsymbol{\mu}^* + \nabla \boldsymbol{h}(\boldsymbol{x}^*) \boldsymbol{\lambda}^* = \boldsymbol{0}_n \tag{D.48}$$

と

$$\boldsymbol{\lambda} \geq \mathbf{0}_p, \quad \langle \boldsymbol{\lambda}, \boldsymbol{h}(\boldsymbol{x}^*) \rangle_{2,p} = 0 \tag{D.49}$$

となり KKT 条件となる.

$\gamma^* \neq 0$ のときは γ^* または $-\gamma^*$ で割ることで (D.46) と (D.47) を KKT 条件に直すことができる.

$\gamma^* = 0$ のとき, 異常な乗数とよぶ. □

Proof. **(定理** D.40 **の証明)**. $k = 1, 2, \ldots$ に対して

$$F_k(\boldsymbol{x}) := f(\boldsymbol{x}) + \frac{k}{2}\left(|\boldsymbol{g}(\boldsymbol{x})|_{2,m}^2 + |\boldsymbol{h}^+(\boldsymbol{x})|_{2,p}^2\right)^2 + \frac{1}{2}|\boldsymbol{x} - \boldsymbol{x}^*|_{2,n}$$

とおく. ただし $\ell = 1, 2, \ldots, p$ に対し

$$h_\ell^+(\boldsymbol{x}) = \begin{cases} h_\ell(\boldsymbol{x}) & (h_\ell(\boldsymbol{x}) \geq 0 \text{ の場合}) \\ 0 & (h_\ell(\boldsymbol{x}) = 0 \text{ の場合}) \end{cases}$$

である. ここで $B = \{\boldsymbol{x} \in \mathbb{R}^n;\ |\boldsymbol{x} - \boldsymbol{x}^*|_{2,n} \leq 1\}$ とおく. すると B はコンパクトで F_k は連続関数なので, 点 $\boldsymbol{x}_k \in B$ があって

$$F_k(\boldsymbol{x}_k) = \min_{\boldsymbol{x} \in B} F_k(\boldsymbol{x})$$

が成り立つ. したがって

$$F_k(\boldsymbol{x}_k) \leq F_k(\boldsymbol{x}^*)$$

である. さらに $\boldsymbol{x}^*$ は問題 (D.45) の解なので, $\boldsymbol{g}(\boldsymbol{x}^*) = \mathbf{0}_m$ かつ $\boldsymbol{h}(\boldsymbol{x}^*) \leq \mathbf{0}_p$ となる. よって h_ℓ^+ の定義から $\boldsymbol{h}^+(\boldsymbol{x}^*) = \mathbf{0}_p$ となる. したがって

$$f(\boldsymbol{x}_k) + \frac{k}{2}\left(|\boldsymbol{g}(\boldsymbol{x}_k)|_{2,m}^2 + |\boldsymbol{h}^+(\boldsymbol{x}_k)|_{2,p}^2\right) + \frac{1}{2}|\boldsymbol{x}_k - \boldsymbol{x}^*|_{2,n}^2 = F_k(\boldsymbol{x}_i) \tag{D.50}$$

$$\leq F_k(\boldsymbol{x}^*) = f(\boldsymbol{x}^*) + \frac{k}{2}\left(\underbrace{|\boldsymbol{g}(\boldsymbol{x}^*)|_{2,m}^2}_{=0} + \underbrace{|\boldsymbol{h}^+(\boldsymbol{x}^*)|_{2,p}^2}_{=0}\right) = f(\boldsymbol{x}^*) \tag{D.51}$$

が成り立つ. したがって, $\left\{\dfrac{k}{2}|\boldsymbol{g}(\boldsymbol{x}_k)|_{2,m}\right\}_{k=1}^\infty$ と $\left\{\dfrac{k}{2}|\boldsymbol{h}^+(\boldsymbol{x}_k)|_{2,p}\right\}_{k=1}^\infty$ は有界なので

$$\lim_{k\to\infty} \boldsymbol{g}(\boldsymbol{x}_k) = \mathbf{0}_m, \quad \lim_{k\to\infty} \boldsymbol{h}^+(\boldsymbol{x}_k) = \mathbf{0}_p \tag{D.52}$$

となる．そうでなければ, $\{k/2\}_{k=1}^{\infty}$ は発散するので, $\left\{\dfrac{k}{2}|\boldsymbol{g}(\boldsymbol{x}_k)|_{2,m}\right\}_{k=1}^{\infty}$ と $\left\{\dfrac{k}{2}|\boldsymbol{h}^{+}(\boldsymbol{x}_k)|_{2,p}\right\}_{k=1}^{\infty}$ にはならない．さらに Bolzano-Weirstrass の定理から $\{\boldsymbol{x}_k\}_{k=1}^{\infty} \subset B$ から収束する部分列 $\{\boldsymbol{x}_{k(j)}\}_{j=1}^{\infty}$ を選ぶことができることがわかる．以上から

$$\lim_{j\to\infty} \boldsymbol{x}_{k(j)} = \overline{\boldsymbol{x}} \in B$$

と書くことにする．(D.51) は

$$f(\boldsymbol{x}_k) + \frac{1}{2}|\boldsymbol{x}_k - \boldsymbol{x}^*|_{2,n}^2 \le f(\boldsymbol{x}^*)$$

となる．したがって

$$f(\overline{\boldsymbol{x}}) + \frac{1}{2}|\overline{\boldsymbol{x}} - \boldsymbol{x}^*|_{2,n}^2 \le f(\boldsymbol{x}^*) \tag{D.53}$$

となる．しかし (D.52) から

$$\boldsymbol{h}(\overline{\boldsymbol{x}}) = \boldsymbol{0}_p, \quad \text{かつ} \quad \boldsymbol{g}(\overline{\boldsymbol{x}}) = \boldsymbol{0}_m$$

となるので, $\overline{\boldsymbol{x}}$ も実行可能となる．したがって

$$f(\boldsymbol{x}^*) \le f(\overline{\boldsymbol{x}}) \tag{D.54}$$

を得る．(D.53) と (D.54) を合わせると

$$f(\boldsymbol{x}^*) + \frac{1}{2}|\overline{\boldsymbol{x}} - \boldsymbol{x}^*|_{2,n} \le f(\overline{\boldsymbol{x}}) + \frac{1}{2}|\overline{\boldsymbol{x}} - \boldsymbol{x}^*|_{2,n} \le f(\boldsymbol{x}^*) \Leftrightarrow |\overline{\boldsymbol{x}} - \boldsymbol{x}^*|_{2,n} = 0$$

を得る．したがって

$$\boldsymbol{x}^* = \overline{\boldsymbol{x}}$$

がわかる．収束する部分列 $\{\boldsymbol{x}_{k(j)}\}_{j=1}^{\infty}$ はすべて同じ収束先をもつので

$$\lim_{k\to\infty} \boldsymbol{x}_k = \boldsymbol{x}^* \tag{D.55}$$

がわかる．

$\{\boldsymbol{x}_k\}_{k=1}^{\infty}$ は $\boldsymbol{x}^*$ に収束するので, k を十分大きくとれば $\boldsymbol{x}_k$ は B の境界から離れている．このことより関数

$$B \ni \boldsymbol{x} \mapsto F_k(\boldsymbol{x}) \in \mathbb{R}$$

は制約なしの局所解 $\boldsymbol{x}_k$ を持つ．よって

$$\begin{aligned}
\boldsymbol{0}_n &= \nabla F_k(\boldsymbol{x}_k) \\
&= \nabla f(\boldsymbol{x}_k) + k\Bigg(\big(\nabla \boldsymbol{g}(\boldsymbol{x}_k)\big)^{\top}\boldsymbol{g}(\boldsymbol{x}_k) + \big(\nabla \boldsymbol{h}^{+}(\boldsymbol{x}_k)\big)^{\top}\boldsymbol{h}^{+}(\boldsymbol{x}_k)\Bigg) + \boldsymbol{x}^* - \boldsymbol{x}_k \\
&= \nabla f(\boldsymbol{x}_k) + k\Bigg(\big(\nabla \boldsymbol{g}(\boldsymbol{x}_k)\big)^{\top}\boldsymbol{g}(\boldsymbol{x}_k) + \big(\nabla \boldsymbol{h}^{+}(\boldsymbol{x}_k)\big)^{\top}\boldsymbol{h}(\boldsymbol{x}_k)\Bigg) + \boldsymbol{x}^* - \boldsymbol{x}_k
\end{aligned} \tag{D.56}$$

となる．最後の等号は $h_\ell^+ \neq 0\,(\ell = 1, 2, \ldots, p)$ のとき $\dfrac{\partial}{\partial x_\ell} h_\ell^+ = \dfrac{\partial}{\partial x_\ell} h_\ell$ であることからわかる．つぎに定数

$$\gamma_k := \left(1 + k^2 |\boldsymbol{g}(\boldsymbol{x}_k)|_{2,m}^2 + k^2 |\boldsymbol{h}(\boldsymbol{x}_k)|_{2,p}^2\right)^{-1/2} > 0$$

を (D.56) の最右辺に掛けると

$$\mathbf{0}_n = \gamma_k \nabla f(\boldsymbol{x}_k) + \left(\nabla \boldsymbol{g}(\boldsymbol{x}_k)\right)^\top \boldsymbol{\mu}_k + \left(\nabla \boldsymbol{h}(\boldsymbol{x}_k)\right)^\top \boldsymbol{\lambda}_k + \gamma_k (\boldsymbol{x}_k - \boldsymbol{x}^*) \tag{D.57}$$

を得る．ただし

$$\boldsymbol{\mu}_k := \gamma_k k \boldsymbol{g}(\boldsymbol{x}_k), \quad \boldsymbol{\lambda}_k =: \gamma_k k \boldsymbol{h}^+(\boldsymbol{x}_k) \geq \mathbf{0}_p$$

である．記号の定義から

$$\gamma_k^2 + |\boldsymbol{\mu}_k|_{2,m}^2 + |\boldsymbol{\lambda}_k|_{2,p}^2 = 1$$

なので，$\{(\gamma_k, \boldsymbol{\mu}_k, \boldsymbol{\lambda}_k)\}_{k=1}^\infty$ は有界である．したがって，収束する部分列 $\{(\gamma_{k(j)}, \boldsymbol{\mu}_{k(j)}, \boldsymbol{\lambda}_{k(j)})\}_{j=1}^\infty$ を取ることができる．すなわち $j \to \infty$ のとき

$$\gamma_{k(j)} \to \gamma_0 \geq\in \mathbb{R}, \quad \boldsymbol{\mu}_{k(j)} \to \boldsymbol{\mu}_0 \in \mathbb{R}^m, \quad \boldsymbol{\lambda}_{k(j)} \to \boldsymbol{\lambda}_0 \in \mathbb{R}^p$$

と書くことができる．このとき

$$\gamma_0^2 + |\boldsymbol{\mu}_0|_{2,m}^2 + |\boldsymbol{\lambda}|_{2,p}^2 = 1$$

なので

$$(\gamma_0, , \boldsymbol{\mu}_0, \boldsymbol{\lambda}_0) \neq (0, \mathbf{0}_m, \mathbf{0}_p)$$

である．(D.57) で $k = k(j) \overset{j\to\infty}{\infty}$ とすると (D.55) から (D.46) を得る．

次に $\boldsymbol{h}^*(\boldsymbol{x}^*) \geq \mathbf{0}_p$ と $\boldsymbol{\lambda}^* \geq \mathbf{0}_p$ に注意する．ある $\ell \in \{1, 2, \ldots, p\}$ に対して

$$h_\ell(\boldsymbol{x}^*) < 0$$

ならば十分大きな k に対して

$$h_\ell(\boldsymbol{x}_k) < 0$$

となる．よって $h_\ell^+(\boldsymbol{x}_k) = 0$ なので λ_k の第 ℓ 成分は 0 となるので $\lambda_\ell^* = 0$ となる．よって

$$\langle \boldsymbol{\lambda}^*, \boldsymbol{h}(\boldsymbol{x}^*) \rangle_{2,n} = 0$$

が成り立つ．よって (D.47) が示された．　□

単調増加関数の不連続点は高々可算個であることの証明をのせる．

D.6　変分不等式

$C \subset \mathbb{R}^n$ を空でない凸部分集合とする. 基本的な問題

$$\text{minimize}\, f(\boldsymbol{x}) \qquad \text{subject to } \boldsymbol{x} \in C \tag{D.58}$$

の解 $\boldsymbol{x}^*$ をみつけることを考える.

定理 D.42. (i) 関数 $f : \mathbb{R}^n \to \mathbb{R}$ は連続微分可能で, $\boldsymbol{x}^* \in C$ は問題 (D.58) の解のとき

$$\big\langle \nabla f(\boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^* \big\rangle_{2,n} \geq 0 \qquad (\boldsymbol{x} \in C) \tag{D.59}$$

が成立する.
(ii) 関数 f は凸とする. $\boldsymbol{x}^* \in C$ が (D.59) をみたすとき, $\boldsymbol{x}^*$ は問題 (D.58) の解である.

注意 D.43. (D.59) を**変分不等式**という. これは制約問題 (D.58) の 1 次変分の形になっている. □

定理 D.42 **の証明**: (i) $\boldsymbol{x} \in C$ を任意の点とする. このとき, C の凸から $0 \leq t \leq 1$ に対して

$$\boldsymbol{x}^* + t(\boldsymbol{x} - \boldsymbol{x}^*) = t\boldsymbol{x} + (1-t)\boldsymbol{x}^* \in C$$

である. さらに $\boldsymbol{x}^*$ は問題 (D.58) の解なので, 関数

$$\phi(t) := f\big(\boldsymbol{x}^* + t(\boldsymbol{x} - \boldsymbol{x}^*)\big) \qquad (0 \leq t \leq 1)$$

は $t = 0$ で最小となる. したがって

$$\dot{\phi}(0) = \left.\frac{\mathrm{d}\phi}{\mathrm{d}t}\right|_{t=0} \geq 0$$

である. ただし $\dot{\phi}(0)$ は $t = 0$ における ϕ の右微分係数である. しかし

$$\dot{\phi}(t) = \big\langle \nabla f\big(\boldsymbol{x}^* + t(\boldsymbol{x} - \boldsymbol{x}^*)\big), \boldsymbol{x} - \boldsymbol{x}^* \big\rangle_{2,n}$$

なので

$$\dot{\phi}(0) = \big\langle \nabla f(\boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^* \big\rangle_{2,n} \geq 0$$

となる.
(ii) 関数 f は点 $\boldsymbol{x}^*$ で連続微分可能なので

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) + \big\langle \nabla f(\boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^* \big\rangle_{2,n} \qquad (\boldsymbol{x} \in C)$$

である．(D.59) より

$$\left\langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{x} - \boldsymbol{x}^* \right\rangle_{2,\,n} \quad \geq 0$$

なので

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) \qquad (\boldsymbol{x} \in C)$$

となる．よって $\boldsymbol{x}^*$ は問題 (D.58) の解である．

例 D.44. 関数 $f : \mathbb{R}^n \to \mathbb{R}$ は連続微分可能とし，$\mathbf{A} \in \mathrm{Mat}(p,\, n;\, \mathbb{R})$, $\boldsymbol{b} \in \mathbb{R}^p$ とする．問題

$$\text{minimize}\, f(\boldsymbol{x}) \qquad \text{subject to } \mathbf{A}\boldsymbol{x} = \boldsymbol{b} \tag{D.60}$$

を考える．

$\boldsymbol{x}^*$ は問題 (D.60) の解のとき (D.59) は

$$\left\langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{x} - \boldsymbol{x}^* \right\rangle_{2,\,n} \geq 0 \qquad (\forall \boldsymbol{x} \in \mathbb{R}^n;\, \mathbf{A}\boldsymbol{x} = \boldsymbol{b}) \tag{D.61}$$

である．$\boldsymbol{x} \in \mathbb{R}^n$ は

$$\boldsymbol{x} = \boldsymbol{x}^* + \boldsymbol{w} \qquad (\boldsymbol{x},\, \boldsymbol{w} \in \ker(\mathbf{A}))$$

の形で書く．すると (D.61) から

$$\left\langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{w} \right\rangle_{2,\,n} = \left\langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{x} - \boldsymbol{x}^* \right\rangle_{2,\,n} \geq 0 \qquad (\forall \boldsymbol{w} \in \ker(\mathbf{A}))$$

となる．$\boldsymbol{w}$ を $-\boldsymbol{w}$ に置き換えると

$$\left\langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{w} \right\rangle_{2,\,n} \leq 0 \qquad (\forall \boldsymbol{w} \in \ker(\mathbf{A}))$$

となるので

$$\left\langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{w} \right\rangle_{2,\,n} = 0 \qquad (\forall \boldsymbol{w} \in \ker(\mathbf{A}))$$

がわかる．したがって

$$\nabla f(\boldsymbol{x}^*) \in \bigl(\ker(\mathbf{A})\bigr)^{\perp} = \mathrm{range}(\mathbf{A}^\top)$$

がわかる．したがって，$\nabla f(\boldsymbol{x}^*)$ が $\mathbf{A}^\top$ の像に含まれるので，ある $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ が存在して

$$\nabla f(\boldsymbol{x}^*) = \mathbf{A}^\top(-\boldsymbol{\lambda}^*) \;\Leftrightarrow\; \nabla f(\boldsymbol{x}^*) + \mathbf{A}^\top \boldsymbol{\lambda}^* = \mathbf{0}_n$$

と書ける．このとき $\boldsymbol{\lambda}^* = (\lambda_1^*,\, \lambda_2^*,\, \ldots,\, \lambda_p^*)^\top$ は制約 $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$ の Lagrange 乗数となる． □

D.7　凸性と Lagrange 乗数

問題

$$\text{minimize } f(\boldsymbol{x}) \qquad \text{subject to } \boldsymbol{h} \leq \boldsymbol{0}_p \tag{D.62}$$

の解 $\boldsymbol{x}^*$ をみつけるときの Lagrange 乗数の存在を定理 D.37 は保証している．これは関数の凸性の仮定を用いていない．しかし多くの場合 qualification 条件を確認することが容易でない．この節では強い仮定

$$f,\, h_1,\, h_2,\, \ldots,\, h_p : \mathbb{R}^n \to \mathbb{R} \text{ は凸関数} \tag{D.63}$$

を課す．

まず凸関数にたいして，KKT 条件は最適解の十分条件になることを示す．

定理 D.45. $f,\, h_1,\, h_2,\, \ldots,\, h_p : \mathbb{R}^n \to \mathbb{R}$ は凸関数かつ連続微分可能とする．$\boldsymbol{x}^* \in \mathbb{R}^n$ は

$$\boldsymbol{h}(\boldsymbol{x}^*) \leq \boldsymbol{0}_p \tag{D.64}$$

をみたすとする．さらに，ある $\boldsymbol{\mu}^* = (\mu_1^*,\, \mu_2^*,\, \ldots,\, \mu_p^*)^\top \in \mathbb{R}^p$ が存在して KKT 条件

$$\nabla f(\boldsymbol{x}^*) + \sum_{j=1}^{p} \mu_j^* \nabla h_j(\boldsymbol{x}^*) = \boldsymbol{0}_n \tag{D.65}$$

$$\boldsymbol{\mu}^* \geq \boldsymbol{0}_p, \qquad \langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\boldsymbol{x}^*) \rangle_{2,\,p} \geq 0 \tag{D.66}$$

が成立するとする．このとき $\boldsymbol{x}^*$ は問題 (D.62) の解である．

Proof. $C := \{\boldsymbol{x} \in \mathbb{R}^n;\, \boldsymbol{h}(\boldsymbol{x}) \leq \boldsymbol{0}\}$ は実行可能解の集合とする．各 $h_j\ (j = 1,\, 2,\, \ldots,\, p)$ は凸関数なので

$$C_j := \{\boldsymbol{x} \in \mathbb{R}^n;\, h_j(\boldsymbol{x}) \leq 0\}$$

は凸集合である．したがって

$$C = \bigcap_{j=1}^{p} C_j$$

も凸集合となる．(D.65) は

$$\begin{aligned} 0 &= \Big\langle \underbrace{\nabla f(\boldsymbol{x}^*) + \sum_{j=1}^{p} \mu_j^* \nabla h_j(\boldsymbol{x}^*)}_{=\boldsymbol{0}_n},\, \boldsymbol{x} - \boldsymbol{x}^* \Big\rangle_{2,\,n} \\ &= \big\langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{x} - \boldsymbol{x}^* \big\rangle_{2,\,n} + \left\langle \sum_{j=1}^{p} \mu_j^* \nabla h_j(\boldsymbol{x}^*),\, \boldsymbol{x} - \boldsymbol{x}^* \right\rangle_{2,\,n} \qquad (\forall \boldsymbol{x} \in \mathbb{R}^n) \end{aligned} \tag{D.67}$$

となる. $\boldsymbol{x} \in C$ である. C の定義と $h_j\,(j = 1, 2, \ldots, p)$ の凸性と (D.64) より

$$h_j(\boldsymbol{x}^*) + \big\langle \nabla h_j(\boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^* \big\rangle_{2,n} \le h_j(\boldsymbol{x}) \le 0 \tag{D.68}$$

がわかる. なぜならば, 関数 $\phi(t) := h_j\big(\boldsymbol{x}^* + t(\boldsymbol{x} - \boldsymbol{x}^*)\big)\,(0 \le t \le 1)$ を考える. すると関数 h_j の凸性から関数 ϕ も凸となる. したがって

$$h_j(0) + t\frac{\mathrm{d}}{\mathrm{d}t}\phi(t)\bigg|_{t=0} \le h_j(t) \tag{D.69}$$

を得る. さらに

$$\frac{\mathrm{d}}{\mathrm{d}t}\phi(t)\bigg|_{t=0} = \big\langle \nabla h_j(\boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^* \big\rangle_{2,n}$$

であることに注意して, (D.69) に $t = 1$ を代入すると

$$h_j(\boldsymbol{x}^*) + \big\langle \nabla h_j(\boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^* \big\rangle_{2,n} = \phi(0) + \frac{\mathrm{d}}{\mathrm{d}t}\phi(t)\bigg|_{t=0} \le \phi(1) = h_j(\boldsymbol{x})$$

を得る. (D.68) の両辺を $\mu_j^*\,(\ge 0)$ 倍して和をとれば

$$0 \ge \underbrace{\big\langle \boldsymbol{\mu}^*, \boldsymbol{h}(\boldsymbol{x}^*) \big\rangle_{2,p}}_{\ge 0 \quad \because (\mathrm{D}.66)} + \left\langle \sum_{j=1}^{p} \mu_j^* \nabla h_j(\boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^* \right\rangle_{2,n}$$

がわかる. したがって, 上式と (D.65) から

$$\left\langle \sum_{j=1}^{p} \mu_j^* \nabla h_j(\boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^* \right\rangle_{2,n} \le 0 \Leftrightarrow \left\langle \nabla f(\boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^* \right\rangle_{2,n} \ge 0 \qquad (\forall \boldsymbol{x} \in C) \tag{D.70}$$

がわかる. 関数 f の凸性と (D.70) から

$$f(\boldsymbol{x}) \ge f(\boldsymbol{x}^*) + \big\langle \nabla f(\boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^* \big\rangle_{2,n} \ge f(\boldsymbol{x}^*) \qquad (\forall \boldsymbol{x} \in C)$$

を得る. したがって $\boldsymbol{x}^*$ は問題 (D.62) の解であることが示せた. □

定義 D.46. 問題 (D.62) を考える.

$$\text{ある } \overline{\boldsymbol{x}} \in \mathbb{R}^n \text{ が存在して } \boldsymbol{h}(\overline{\boldsymbol{x}}) < \boldsymbol{0}_p \tag{D.71}$$

が成り立つとき, 問題 (D.62) に対する **Slater 条件**が成立するという. ただし, $\boldsymbol{h}(\overline{\boldsymbol{x}}) < \boldsymbol{0}_p$ は, $h_j(\overline{\boldsymbol{x}}) < 0\,(\forall j = 1, 2, \ldots, p)$ が成り立つことである.

定理 D.47. Slater 条件 (D.71) が成立するとする.

$$f,\, h_1,\, h_2,\, \ldots,\, h_p : \mathbb{R}^n \to \mathbb{R}$$

は凸関数かつ連続微分可能とし, $\boldsymbol{x}^*$ をは問題 (D.62) の解とする.
(i) このとき $\boldsymbol{\mu}^* \in \mathbb{R}^p$ が存在して KKT 条件 (D.65) と (D.66) が成り立つ.
(ii) さらに

$$\boldsymbol{x} \mapsto f(\boldsymbol{x}) + \langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\boldsymbol{x}) \rangle_{2,\,p}$$

は点 $\boldsymbol{x}^*$ で最小となる.

Proof. 定理 D.37 の John の formulation から, ある　非負実数の組 $\gamma^*,\, \mu_1^*,\, \mu_2^*,\, \ldots,\, \mu_p^*$ が存在 (すべてが同時に零になることはない.) して

$$\gamma^* \nabla f(\boldsymbol{x}^*) + \sum_{j=1}^{p} \mu_j^* \nabla h_j(\boldsymbol{x}^*) = \boldsymbol{0}_n \tag{D.72}$$

$$\gamma^* \geq 0, \qquad \boldsymbol{\mu}^* \geq \boldsymbol{0}_p, \qquad \langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\boldsymbol{x}^*) \rangle_{2,\,p} \geq 0 \tag{D.73}$$

が成立する.
　まず

$$\gamma^* \neq 0 \tag{D.74}$$

を示す. そのために $\gamma^* = 0$ を仮定して矛盾を導く. すべては同時に零にならないので $\boldsymbol{\mu}^* \neq \boldsymbol{0}_p$ である. 一方, $\gamma^* = 0$ の仮定と (D.72) から

$$\sum_{j=1}^{p} \mu_j^* \nabla h_j(\boldsymbol{x}^*) = \boldsymbol{0}_n \tag{D.75}$$

となる. したがって, (D.75) と Slater 条件 D.71 から

$$\left\langle \sum_{j=1}^{p} \mu_j^* \nabla h_j(\boldsymbol{x}^*),\, \overline{\boldsymbol{x}} - \boldsymbol{x}^* \right\rangle_{2,\,n} = 0 \quad \text{かつ} \quad \boldsymbol{h}(\overline{\boldsymbol{x}}) < \boldsymbol{0}_p \tag{D.76}$$

である. しかし $h_j\,(j = 1,\, 2,\, \ldots,\, p)$ は凸関数なので

$$h_j(\boldsymbol{x}^*) + \left\langle \nabla h_j(\boldsymbol{x}^*),\, \overline{\boldsymbol{x}} - \boldsymbol{x}^* \right\rangle_{2,\,n} \leq h_j(\overline{\boldsymbol{x}})$$

が成り立つ. 上式の辺々と $\mu_j^*\,(j = 1,\, 2, \ldots,\, p)$ の積を取り, j について和を取ると

$$\begin{aligned}
\left\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\overline{\boldsymbol{x}}) \right\rangle_{2,\,p} &= \sum_{j=1}^{p} \mu_j^* h_j(\overline{\boldsymbol{x}}) = \sum_{j=1}^{p} \mu_j^* h_j(\boldsymbol{x}^*) + \sum_{j=1}^{p} \mu_j^* \left\langle \nabla h_j(\boldsymbol{x}^*),\, \overline{\boldsymbol{x}} - \boldsymbol{x}^* \right\rangle_{2,\,n} \\
&= \left\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\boldsymbol{x}^*) \right\rangle_{2,\,p} + \underbrace{\left\langle \sum_{j=1}^{p} \mu_j^* \nabla h_j(\boldsymbol{x}^*),\, \overline{\boldsymbol{x}} - \boldsymbol{x}^* \right\rangle_{2,\,n}}_{=0 \quad \because (\text{D.76})} \\
&= \left\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\boldsymbol{x}^*) \right\rangle_{2,\,p}
\end{aligned}$$

から

$$\left\langle \boldsymbol{\mu}^*, \boldsymbol{h}(\overline{\boldsymbol{x}}) \right\rangle_{2,p} = \left\langle \boldsymbol{\mu}^*, \boldsymbol{h}(\boldsymbol{x}^*) \right\rangle_{2,p} \tag{D.77}$$

を得る．よって, $\boldsymbol{\mu}^* \geq \mathbf{0}_p$ かつ $\boldsymbol{\mu}^* \neq \mathbf{0}_p$ で $\boldsymbol{h}(\overline{\boldsymbol{x}}) < \mathbf{0}_p \left(\because \text{(D.74)}\right)$ なので, (D.79) と合わせると

$$\left\langle \boldsymbol{\mu}^*, \boldsymbol{h}(\boldsymbol{x}^*) \right\rangle_{2,p} < 0$$

となる．これは (D.73) と矛盾する．背理法の仮定から $\gamma^* \neq 0$ となる．よって, このことと D.73) から $\gamma^* > 0$ が示せた．

次に, (D.72) を $\gamma^* > 0$ で割り

$$\nabla f(\boldsymbol{x}^*) + \sum_{j=1}^{p} \frac{\mu_j^*}{\gamma^*} \nabla h_j(\boldsymbol{x}) = \mathbf{0}_n$$

を考える．$\widetilde{\mu}_j^* = \mu_j^*/\gamma^* \geq 0\,(j = 1, 2, \ldots, p)$ とおけば, (D.73) から

$$\begin{aligned} &\nabla f(\boldsymbol{x}^*) + \sum_{j=1}^{p} \widetilde{\mu}_j^* \nabla h_j(\boldsymbol{x}^*) = \mathbf{0}_n, \\ &\widetilde{\boldsymbol{\mu}} = (\widetilde{\mu}_1^*, \widetilde{\mu}_2^*, \ldots, \widetilde{\mu}_p^*) \geq \mathbf{0}_p, \\ &\left\langle \widetilde{\boldsymbol{\mu}}^*, \boldsymbol{h}(\boldsymbol{x}^*) \right\rangle_{2,p} = 0 \end{aligned} \tag{D.78}$$

を得る．ただし, $\widetilde{\boldsymbol{\mu}}^* = (\widetilde{\mu}_1^*, \widetilde{\mu}_2^*, \ldots, \widetilde{\mu}_p^*)^\top$ とした．関数 $h_j\,(j = 1, 2, \ldots, p)$ の凸性より

$$h_j(\overline{\boldsymbol{x}}) \geq h_j(\boldsymbol{x}^*) + \left\langle \nabla h_j(\boldsymbol{x}^*), \boldsymbol{x}^* - \overline{\boldsymbol{x}} \right\rangle_{2,n}$$

を得る．上式の辺々に $\widetilde{\mu}_j$ を掛けて, j について和ををると

$$\left\langle \widetilde{\boldsymbol{\mu}}^*, \boldsymbol{h}(\overline{\boldsymbol{x}}) \right\rangle_{2,p} \geq \left\langle \widetilde{\boldsymbol{\mu}}^*, \boldsymbol{h}(\boldsymbol{x}^*) \right\rangle_{2,p} + \left\langle \sum_{j=1}^{p} \widetilde{\mu}_j^* \nabla h_j(\boldsymbol{x}^*), \boldsymbol{x}^* - \overline{\boldsymbol{x}} \right\rangle_{2,p} \tag{D.79}$$

を得る．さらに f の凸性から

$$f(\overline{\boldsymbol{x}}) \geq f(\boldsymbol{x}^*) + \left\langle \nabla f(\boldsymbol{x}^*), \overline{\boldsymbol{x}} - \boldsymbol{x}^* \right\rangle_{2,n} \tag{D.80}$$

がわかる．(D.79) と (D.80) を合わせると

$$\begin{aligned} f(\overline{\boldsymbol{x}}) + \left\langle \widetilde{\boldsymbol{\mu}}^*, \boldsymbol{h}(\overline{\boldsymbol{x}}) \right\rangle_{2,p} &\geq f(\boldsymbol{x}^*) + \left\langle \nabla f(\boldsymbol{x}^*), \overline{\boldsymbol{x}} - \boldsymbol{x}^* \right\rangle_{2,n} + \left\langle \widetilde{\boldsymbol{\mu}}^*, \boldsymbol{h}(\boldsymbol{x}^*) \right\rangle_{2,p} \\ &\qquad + \left\langle \sum_{j=1}^{p} \widetilde{\mu}_j^* \nabla h_j(\boldsymbol{x}), \overline{\boldsymbol{x}} - \boldsymbol{x}^* \right\rangle_{2,n} \\ &= f(\overline{\boldsymbol{x}}) + \left\langle \widetilde{\boldsymbol{\mu}}^*, \boldsymbol{h}(\boldsymbol{x}^*) \right\rangle_{2,p} \\ &\qquad + \left\langle \underbrace{\nabla f(\boldsymbol{x}^*) + \sum_{j=1}^{p} \widetilde{\mu}_j^* \nabla h_j(\boldsymbol{x})}_{=\mathbf{0}_n \quad \because \text{(D.78)}}, \overline{\boldsymbol{x}} - \boldsymbol{x}^* \right\rangle_{2,n} \\ &= f(\boldsymbol{x}^*) + \left\langle \widetilde{\boldsymbol{\mu}}^*, \boldsymbol{h}(\boldsymbol{x}^*) \right\rangle_{2,p} \end{aligned}$$

を得る．よって (ii) は示された． □

D.8　凸双対性

D.8.1　双対問題

与えられた関数 $f,\, g_1,\, g_2,\, \ldots,\, g_m,\, h_1,\, h_2,\, \ldots,\, h_p : \mathbb{R}^n \to \mathbb{R}$ に対して

$$\boldsymbol{h} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_p \end{bmatrix}, \quad \boldsymbol{g} = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_m \end{bmatrix}$$

と書くことにする．

問題

$$\text{minimize}\, f(\boldsymbol{x}) \quad \text{subject to} \;\; g(\boldsymbol{x}) = \mathbf{0}_m \text{ and } h(\boldsymbol{x}) \leq \mathbf{0}_p \tag{D.81}$$

の解を $\boldsymbol{x}^*$ をみつけたいとする．

定義 D.48. 問題 (D.81) に対応する Lagrange 問題 $L(\boldsymbol{x},\, \boldsymbol{\lambda},\, \boldsymbol{\mu})\, (\boldsymbol{x} \in \mathbb{R}^n,\, \boldsymbol{\lambda} \in \mathbb{R}^m,\, \boldsymbol{\mu} \in \mathbb{R}^p)$ を

$$L(\boldsymbol{x},\, \boldsymbol{\lambda},\, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \left\langle \boldsymbol{\lambda},\, \boldsymbol{g}(\boldsymbol{x}) \right\rangle_{2,\, m} + \left\langle \boldsymbol{\mu},\, \boldsymbol{h}(\boldsymbol{x}) \right\rangle_{2,\, p}$$

で定める．さらに

$$L^*(\boldsymbol{\lambda},\, \boldsymbol{\mu}) = \inf_{\boldsymbol{x} \in \mathbb{R}^n} L(\boldsymbol{x},\, \boldsymbol{\lambda},\, \boldsymbol{\mu})$$

と定める．

注意 D.49. $\boldsymbol{\lambda} = (\lambda_1,\, \lambda_2,\, \ldots,\, \lambda_m)^\top$, $\boldsymbol{\mu} = (\mu_1,\, \mu_2,\, \ldots,\, \mu_p)^\top$ と書いたとき

$$L(\boldsymbol{x},\, \boldsymbol{\lambda},\, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \sum_{j=1}^{m} \lambda_j g_j(\boldsymbol{x}) + \sum_{j=1}^{p} \mu_j h_j(\boldsymbol{x})$$

とも書ける． □

例 D.50. 関数 f は凸とし，問題

$$\text{minimize}\, f(\boldsymbol{x}) \quad \text{subject to } \mathbf{A}\boldsymbol{x} = \boldsymbol{b} \text{ and } \boldsymbol{x} \geq \mathbf{0}_n \tag{D.82}$$

を考える．ただし $\mathbf{A} \in \mathrm{Mat}(m,\, n;\, \mathbb{R})$, $\boldsymbol{b} \in \mathbb{R}^m$ とする．$\boldsymbol{\lambda} \in \mathbb{R}^m$ と $\boldsymbol{\mu} \leq \mathbf{0}_n\,(\boldsymbol{\mu} \in \mathbb{R}^n)$ を取り，$\boldsymbol{h}(\boldsymbol{x}) = -\boldsymbol{x}$ とおくと

$$
\begin{aligned}
L^*(\boldsymbol{\lambda},\, \boldsymbol{\mu}) &= \inf_{\boldsymbol{x}\in\mathbb{R}^n}\left\{f(\boldsymbol{x}) + \big\langle\boldsymbol{\lambda},\, \boldsymbol{g}(\boldsymbol{x})\big\rangle_{2,\,m} + \big\langle\boldsymbol{\mu},\, \boldsymbol{h}(\boldsymbol{x})\big\rangle_{2,\,n}\right\} \\
&= \inf_{\boldsymbol{x}\in\mathbb{R}^n}\left\{f(\boldsymbol{x}) + \big\langle\boldsymbol{\lambda},\, \mathbf{A}\boldsymbol{x} - \boldsymbol{b}\big\rangle_{2,\,m} - \big\langle\boldsymbol{\mu},\, \boldsymbol{x}\big\rangle_{2,\,n}\right\} \\
&= -\langle\boldsymbol{\lambda},\, \boldsymbol{b}\rangle_{2,\,m} + \inf_{\boldsymbol{x}\in\mathbb{R}^n}\left\{f(\boldsymbol{x}) + \big\langle\mathbf{A}^\top\boldsymbol{\lambda} - \boldsymbol{\mu},\, \boldsymbol{x}\big\rangle_{2,\,n}\right\} \\
&= -\langle\boldsymbol{\lambda},\, \boldsymbol{b}\rangle_{2,\,m} - \sup_{\boldsymbol{x}\in\mathbb{R}^n}\left\{\big\langle\boldsymbol{\mu} - \mathbf{A}^\top\boldsymbol{\lambda},\, \boldsymbol{x}\big\rangle_{2,\,n} - f(\boldsymbol{x})\right\}
\end{aligned}
$$

となる．したがって

$$
L^*(\boldsymbol{\lambda},\, \boldsymbol{\mu}) = -\langle\boldsymbol{\lambda},\, \boldsymbol{b}\rangle_{2,\,n} - f^*\big(\boldsymbol{\mu} - \mathbf{A}^\top\boldsymbol{\lambda}\big)
$$

となる．ただし

$$
f^*(\boldsymbol{x}) = \sup_{\boldsymbol{x}\in\mathbb{R}^n}\left\{\langle\boldsymbol{x},\, \boldsymbol{y}\rangle_{2,\,n} - f(\boldsymbol{x})\right\}
$$

は凸関数 f の双対関数である．□

定義 D.51. 問題 (D.81) に対する双対問題は

$$
\text{minimize}\, L^*(\boldsymbol{\lambda},\, \boldsymbol{\mu}) \quad \text{subject to } \boldsymbol{\mu} \geq \mathbf{0}_p \tag{D.83}
$$

の解 $\boldsymbol{\lambda}^* \in \mathbb{R}^m$, $\boldsymbol{\mu}^* \in \mathbb{R}^p$ をみつけることである．

例 D.52. $f(\boldsymbol{x}) = \langle\boldsymbol{x},\, \boldsymbol{a}\rangle_{2,\,n}$ とする．ただし $\boldsymbol{a} \in \mathbb{R}^n$ は定数ベクトルである．線型計画法

$$
\text{minimize } \langle\boldsymbol{x},\, \boldsymbol{a}\rangle_{2,\,n} \quad \text{subject to } \mathbf{A}\boldsymbol{x} = \boldsymbol{b},\, \boldsymbol{x} \geq \mathbf{0}_n
$$

を考える．ただし $\mathbf{A} \in \mathrm{Mat}(m,\, n;\, \mathbb{R})$, $\boldsymbol{b} \in \mathbb{R}^m$ である．いま

$$
f^*(\boldsymbol{y}) = \max_{\boldsymbol{x}\in\mathbb{R}^n}\big\{\langle\boldsymbol{y} - \boldsymbol{a},\, \boldsymbol{x}\rangle_{2,\,n}\big\} = \begin{cases} 0 & (\boldsymbol{y} = \boldsymbol{a}) \\ \infty & (\text{その他の場合}) \end{cases}
$$

と定める．したがって

$$
\begin{aligned}
L^*(\boldsymbol{\lambda},\, \boldsymbol{\mu}) &= -\langle\boldsymbol{\lambda},\, \boldsymbol{b}\rangle_{2,\,m} - f^*\big(\boldsymbol{\mu} - \mathbf{A}^\top\boldsymbol{\lambda}\big) \\
&= \begin{cases} -\langle\boldsymbol{\lambda},\, \boldsymbol{b}\rangle_{2,\,n} & (\boldsymbol{\mu} - \mathbf{A}^\top\boldsymbol{\lambda} = \boldsymbol{a}) \\ -\infty & (\text{その他の場合}) \end{cases}
\end{aligned}
$$

となる．このとき双対問題は

$$
\text{minimize } -\langle\boldsymbol{\lambda},\, \boldsymbol{b}\rangle_{2,\,m} \quad \text{subject to } \boldsymbol{\mu} - \mathbf{A}^\top\boldsymbol{\lambda} = \boldsymbol{a} = \mathbf{0}_p \text{ and } \boldsymbol{\mu} \geq \mathbf{0}_p
$$

となる．□

定理 D.53. (弱双対不等式) 不等式

$$\sup_{\boldsymbol{\mu}\in\mathbb{R}^p,\,\boldsymbol{\lambda}\in\mathbb{R}^m} L^*(\boldsymbol{\lambda},\,\boldsymbol{\mu}) \le \inf_{\boldsymbol{x}\in\mathbb{R}^n;\,\boldsymbol{g}(\boldsymbol{x})=\mathbf{0}_m,\,\boldsymbol{h}(\boldsymbol{x})\le\mathbf{0}_p} f(\boldsymbol{x}) \tag{D.84}$$

が成立する．この不等式が厳密に成立するとき，**双対ギャップ**があるという．

Proof. $\boldsymbol{\mu}\ge\mathbf{0}_p$, $\boldsymbol{\lambda}\in\mathbb{R}^m$, , $\boldsymbol{g}=\mathbf{0}_m$, $\boldsymbol{h}\le\mathbf{0}_p$ とする．このとき

$$L(\boldsymbol{x},\,\boldsymbol{\lambda},\,\boldsymbol{\mu}) = f(\boldsymbol{x}) + \underbrace{\langle\boldsymbol{\mu},\,\boldsymbol{h}(\boldsymbol{x})\rangle_{2,\,p}}_{\le 0} + \langle\boldsymbol{\lambda},\,\underbrace{\boldsymbol{g}(\boldsymbol{x})}_{=0}\rangle_{2,\,m} \le f(\boldsymbol{x})$$

となる．さらに

$$L^*(\boldsymbol{\lambda},\,\boldsymbol{\mu}) = \inf_{\boldsymbol{x}\in\mathbb{R}^n} L(\boldsymbol{x},\,\boldsymbol{\lambda},\,\boldsymbol{\mu}) \le L(\boldsymbol{x},\,\boldsymbol{\lambda},\,\boldsymbol{\mu})$$

より

$$L^*(\boldsymbol{\lambda},\,\boldsymbol{\mu}) \le f(\boldsymbol{x}).$$

よって (D.84) が示された． □

D.8.2　Slaker 条件 (再訪問)

$\mathbf{A}\in\mathrm{Mat}(m,\,n;\,\mathbb{R})$, $m\le n$, $\boldsymbol{b}\in\mathbb{R}^m$ とし

$$\boldsymbol{g}(\boldsymbol{x}) = \mathbf{A}\boldsymbol{x} - \boldsymbol{b}$$

とおく．さらに

$$\mathrm{rank}\,(\mathbf{A}) = m$$

とする．

主問題

$$\text{minimize } f(\boldsymbol{x}) \quad \text{subject to } \mathbf{A}\boldsymbol{x} = \boldsymbol{b} \text{ and } \boldsymbol{h}(\boldsymbol{x}) \le \mathbf{0}_p \tag{D.85}$$

の解 $\boldsymbol{x}^*\in\mathbb{R}^n$ をみつけた．対応する双対問題

$$\text{maximize } L^*(\boldsymbol{\lambda},\,\boldsymbol{\mu}) \quad \text{subject to } \boldsymbol{\mu} \ge \mathbf{0}_p \tag{D.86}$$

の解 $\boldsymbol{\mu}^*\in\mathbb{R}^p$, $\boldsymbol{\lambda}\in\mathbb{R}^m$ をみつけたい．ただし

$$L^*(\boldsymbol{\lambda},\,\boldsymbol{\mu}) = \inf_{\boldsymbol{x}\in\mathbb{R}^n} L(\boldsymbol{x},\,\boldsymbol{\lambda},\,\boldsymbol{\mu}),$$
$$L(\boldsymbol{x},\,\boldsymbol{\lambda},\,\boldsymbol{\mu}) = f(\boldsymbol{x}) + \left\langle\boldsymbol{\lambda},\,\mathbf{A}\boldsymbol{x}+\boldsymbol{b}\right\rangle_{2,\,m} + \left\langle\boldsymbol{\mu},\,\boldsymbol{h}(\boldsymbol{x})\right\rangle_{2,\,p}$$

である．

定理 D.54. 関数 $h_1, h_2, \ldots, h_p : \mathbb{R}^n \to \mathbb{R}$ は凸かつ連続微分可能とする．さらに修正 Slaker 条件が成立するとする．

$$\text{ある } \overline{\boldsymbol{x}} \in \mathbb{R}^n \text{ があって } \boldsymbol{h}(\overline{\boldsymbol{x}}) \leq \mathbf{0}_p \text{ かつ } \mathbf{A}\overline{\boldsymbol{x}} = \boldsymbol{b} \tag{D.87}$$

をみたす．このとき以下が成り立つ．
(i) $\boldsymbol{x}^*$ は主問題 (D.85) の解のとき

$$\text{組 } (\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \text{ は双対問題 (D.86) の解である．} \tag{D.88}$$

(ii) さらに

$$L^*(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = f(\boldsymbol{x}^*) \tag{D.89}$$

が成立する．したがって, 強双対性

$$\sup_{\boldsymbol{\mu} \in \mathbb{R}^p, \boldsymbol{\lambda} \in \mathbb{R}^m, \boldsymbol{\mu} \geq \mathbf{0}_p} L^*(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\boldsymbol{g}(\boldsymbol{x}) = \mathbf{0}_m, \boldsymbol{h}(\boldsymbol{x}) \geq \mathbf{0}_p} f(\boldsymbol{x})$$

が成立する．

Proof. 定理 D.40 から $(\gamma^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \neq (0, \mathbf{0}_p, \mathbf{0}_m)$ があって F.John の条件をみたす．すなわち

$$\gamma^* \nabla f(\boldsymbol{x}^*) + \mathbf{A}^\top \boldsymbol{\lambda}^* + \nabla \boldsymbol{h}(\boldsymbol{x}^*)^\top \boldsymbol{\mu}^* = \mathbf{0}_n, \tag{D.90}$$

$$\gamma^* \geq 0, \boldsymbol{\mu} \geq \mathbf{0}_m, \boldsymbol{\lambda} \geq \mathbf{0}_p, \left\langle \boldsymbol{\mu}^*, \boldsymbol{h})(\boldsymbol{x}^*) \right\rangle_{2,n} = 0 \tag{D.91}$$

が成り立つ．
　まず背理法を用いて

$$\gamma^* \neq 0 \tag{D.92}$$

を示す．そのために $\gamma = 0$ と仮定すると (D.90) から

$$\mathbf{A}^\top \boldsymbol{\lambda}^* + \nabla \boldsymbol{h}(\boldsymbol{x}^*)^\top \boldsymbol{\mu}^* = \mathbf{0}_n \tag{D.93}$$

となる．さらに (D.87) をみたす $\overline{\boldsymbol{x}}$ を用いて $\boldsymbol{x}^* - \overline{\boldsymbol{x}}$ と上の式の内積をとると

$$\begin{aligned} 0 &= \left\langle \mathbf{A}^\top \boldsymbol{\lambda}^*, \overline{\boldsymbol{x}} - \boldsymbol{x}^* \right\rangle_{2,n} + \left\langle \nabla \boldsymbol{h}(\boldsymbol{x}^*)^\top \boldsymbol{\mu}^*, \overline{\boldsymbol{x}} - \boldsymbol{x}^* \right\rangle_{2,n} \\ &= \left\langle \boldsymbol{\lambda}^*, \mathbf{A}\overline{\boldsymbol{x}} - \mathbf{A}\boldsymbol{x}^* \right\rangle_{2,m} + \left\langle \boldsymbol{\mu}^*, \nabla \boldsymbol{h}(\boldsymbol{x}^*)(\overline{\boldsymbol{x}} - \boldsymbol{x}^*) \right\rangle_{2,p} \end{aligned}$$

となる．しかし

$$\mathbf{A}\overline{\boldsymbol{x}} - \mathbf{A}\boldsymbol{x}^* = \boldsymbol{b} - \boldsymbol{b} = \mathbf{0}_m$$

なので

$$\left\langle \boldsymbol{\mu}^*,\, \nabla \boldsymbol{h}(\boldsymbol{x}^*)(\overline{\boldsymbol{x}} - \boldsymbol{x}^*) \right\rangle_{2,\,p} = 0 \tag{D.94}$$

を得る．背理法の仮定から $\gamma \neq 0$ なので KKT 条件は

$$\nabla \boldsymbol{h}(\boldsymbol{x}^*) + \mathbf{A}^\top \boldsymbol{\lambda}^* + \nabla \boldsymbol{h}(\boldsymbol{x}^*)^\top \boldsymbol{\mu}^* = \mathbf{0}_n \tag{D.95}$$

となる．ただし，$\boldsymbol{\mu}^* = \gamma^{-1}\boldsymbol{\mu}$ と $\boldsymbol{\lambda}^* = \gamma^{-1}\boldsymbol{\lambda}$ とした．この等式が (D.92) を導くことを示そう．$f,\, h_j\,(j = 1,\, 2,\, \ldots,\, p)$ は凸関数であることと (D.94) を用いると

$$\begin{aligned}
f(\boldsymbol{x}^*) &= f(\boldsymbol{x}^*) + \langle \boldsymbol{\lambda}^*,\, \underbrace{\mathbf{A}\boldsymbol{x}^* - \boldsymbol{b}}_{=\mathbf{0}_m} \rangle_{2,\,m} + \underbrace{\left\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\boldsymbol{x}^*) \right\rangle_{2,\,p}}_{=0} \\
&\le \underbrace{f(\overline{\boldsymbol{x}}) + \left\langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{x}^* - \overline{\boldsymbol{x}} \right\rangle_{2,\,n}}_{f\ \text{の凸性から}} + \left\langle \boldsymbol{\lambda}^*,\, (\mathbf{A}\boldsymbol{x}^* - \mathbf{A}\overline{\boldsymbol{x}}) + (\mathbf{A}\overline{\boldsymbol{x}} - \boldsymbol{b}) \right\rangle_{2,\,m} \\
&\quad + \underbrace{\left\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\overline{\boldsymbol{x}}) + \nabla \boldsymbol{h}(\boldsymbol{x}^*)(\boldsymbol{x}^* - \overline{\boldsymbol{x}}) \right\rangle_{2,\,p}}_{h_j\ \text{の凸性と}\ \boldsymbol{\mu}^* \ge \mathbf{0}_p} \\
&= f(\boldsymbol{x}) + \left\langle \boldsymbol{\lambda}^*,\, \mathbf{A}\overline{\boldsymbol{x}} - \boldsymbol{b} \right\rangle_{2,\,m} \\
&\quad + \left\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\overline{\boldsymbol{x}}) \right\rangle_{2,\,p} + \langle \underbrace{\nabla f(\boldsymbol{x}^*) + \mathbf{A}^\top \boldsymbol{\lambda}^* + \nabla \boldsymbol{h}(\boldsymbol{x}^*)^\top \boldsymbol{\mu}^*,\, \boldsymbol{x}^* - \overline{\boldsymbol{x}}}_{=\mathbf{0}_n \because \text{(D.95)}} \rangle_{2,\,n} \\
&= f(\boldsymbol{x}) + \left\langle \boldsymbol{\lambda}^*,\, \mathbf{A}\overline{\boldsymbol{x}} - \boldsymbol{b} \right\rangle_{2,\,p} + \left\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\overline{\boldsymbol{x}}) \right\rangle_{2,\,p}
\end{aligned}$$

となる．さらに $h_j\,(j = 1,\, 2,\, \ldots,\, p)$ の凸性から

$$h_j(\boldsymbol{x}^*) + \left\langle \nabla h_j(\boldsymbol{x}^*),\, \overline{\boldsymbol{x}} - \boldsymbol{x}^* \right\rangle_{2,\,n} \le h_j(\overline{\boldsymbol{x}})$$

を得る．上式の辺々に $\mu_j^*\,(j = 1,\, 2,\, \ldots,\, p)$ を掛けて，j について和をとると

$$\begin{aligned}
&\sum_{j=1}^{p} \mu_j^* h_j(\boldsymbol{x}^*) + \sum_{j=1}^{p} \mu_j^* \left\langle \nabla h_j(\boldsymbol{x}^*),\, \overline{\boldsymbol{x}} - \boldsymbol{x}^* \right\rangle_{2,\,n} \le \sum_{j=1}^{p} \mu_j^* h_j(\overline{\boldsymbol{x}}) \\
&\Leftrightarrow \sum_{j=1}^{p} \mu_j^* h_j(\boldsymbol{x}^*) + \left\langle \sum_{j=1}^{p} \mu_j^* \nabla h_j(\boldsymbol{x}^*),\, \overline{\boldsymbol{x}} - \boldsymbol{x}^* \right\rangle_{2,\,n} \le \sum_{j=1}^{p} \mu_j^* h_j(\overline{\boldsymbol{x}}) \\
&\Leftrightarrow \left\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\boldsymbol{x}^*) \right\rangle_{2,\,p} + \left\langle \left(\nabla \boldsymbol{h}(\boldsymbol{x}^*)\right)^\top \boldsymbol{\mu}^*,\, \overline{\boldsymbol{x}} - \boldsymbol{x}^* \right\rangle_{2,\,n} \le \left\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\overline{\boldsymbol{x}}) \right\rangle_{2,\,p} \\
&\Leftrightarrow \left\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\boldsymbol{x}^*) \right\rangle_{2,\,p} + \underbrace{\left\langle \boldsymbol{\mu}^*,\, \nabla \boldsymbol{h}(\boldsymbol{x}^*)(\overline{\boldsymbol{x}} - \boldsymbol{x}^*) \right\rangle_{2,\,p}}_{=0 \quad \because \text{(D.94))}} \le \left\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\overline{\boldsymbol{x}}) \right\rangle_{2,\,p}
\end{aligned}$$

がわかる．したがって

$$\left\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}^*(\boldsymbol{x}^*) \right\rangle_{2,\,p} \le \left\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\overline{\boldsymbol{x}}) \right\rangle_{2,\,p} \tag{D.96}$$

となる．しかし, $\boldsymbol{\mu}^* \neq \mathbf{0}_p$ のとき, 修正 Slater 条件より

$$\left\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\overline{\boldsymbol{x}}) \right\rangle_{2,\,p} < 0 \tag{D.97}$$

となる．(D.96) と (D.97) から

$$\left\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\boldsymbol{x}^*) \right\rangle_{2,\,p} < 0$$

となり, (D.91) と矛盾する．よって $\gamma^* \neq 0$ ならば $\boldsymbol{\mu} \neq \mathbf{0}_m$ とはならない．
　次に $\boldsymbol{\mu} = \mathbf{0}_m$ の場合を考える．すると (D.95) から

$$\mathbf{A}\boldsymbol{\lambda}^* = \mathbf{0}_m$$

となる．$\mathbf{A}$ は非退化なので $\boldsymbol{\lambda}^* = \mathbf{0}_p$ となり, $(\lambda^*,\, \boldsymbol{\lambda},\, \boldsymbol{\mu}) = (0,\, \mathbf{0}_m,\, \mathbf{0}_p)$ となるので矛盾する．よって $\gamma^* \neq 0$ となる．したがって

$$f(\boldsymbol{x}^*) \leq \inf_{\boldsymbol{x}\in\mathbb{R}^n} \left\{ f(\boldsymbol{x}) + \left\langle \boldsymbol{\lambda}^*,\, \mathbf{A}\boldsymbol{x} - \boldsymbol{b} \right\rangle_{2,\,m} + \left\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\boldsymbol{x}) \right\rangle_{2,\,p} \right\} = L^*(\boldsymbol{\lambda}^*,\, \boldsymbol{\mu}^*)$$

となる．逆向きの不等式は定理 D.53 よりわかる．□

D.9　凸双対性（その２）

D.9.1　Fenchel 双対性

　2 つの凸かつ下半連続関数[5]

$$f:\mathbb{R}^n \to (-\infty,\, \infty) \cup \{+\infty\}, \quad g:\mathbb{R}^m \to (-\infty,\, \infty) \cup \{+\infty\}$$

と $m \times n$ 行列 $\mathbf{A}$ が与えられたとする．

定義 D.55. 関数 f と g の**領域**をそれぞれ

$$\mathrm{dom}\ f = \{\boldsymbol{x} \in \mathbb{R}^n;\, f(\boldsymbol{x}) < \infty\}$$
$$\mathrm{dom}\ g = \{\boldsymbol{y} \in \mathbb{R}^m;\, g(\boldsymbol{y}) < \infty\}$$

で定義する．

---

[5]関数 f は点 $\boldsymbol{x}$ で下半(かはん)連続であるとは, 任意の点列 $\{\boldsymbol{x}_k\}_{k=1}^{\infty} \subset \mathbb{R}^n$ で $\lim_{k\to\infty} \boldsymbol{x}_k = \boldsymbol{x}$ なるものに対して

$$\liminf_{k\to\infty} f(\boldsymbol{x}_k) = f(\boldsymbol{x})$$

をみたすときをいう．

仮定 D.56. 関数 f と g は $-\infty$ の値は取らず

$$\mathrm{dom}\, f \neq \varnothing, \quad \mathrm{dom}\, g \neq \varnothing$$

とする．このことが成り立つとき，これらの関数は proper であるという．

関数 f, g は凸なので，Young の不等式 (補題 D.32) を利用すると，任意の $\boldsymbol{x} \in \mathbb{R}^n$ と $\boldsymbol{y} \in \mathbb{R}^m$ に対して

$$f(\boldsymbol{x}) + f^*(\mathbf{A}^\top \boldsymbol{y}) \geq \left\langle \boldsymbol{x},\, \mathbf{A}^\top \boldsymbol{y} \right\rangle_{2,\, n},$$
$$g\bigl(\mathbf{A}\boldsymbol{x}\bigr) + g^*(-\boldsymbol{y}) \geq -\left\langle \mathbf{A}\boldsymbol{x},\, \boldsymbol{y} \right\rangle_{2,\, m}$$

が成立するとする．

これらの不等式の辺々を加えると

$$f(\boldsymbol{x}) + g\bigl(\mathbf{A}\boldsymbol{x}\bigr) \geq -f\bigl(\mathbf{A}^\top \boldsymbol{y}\bigr) - g^*(-\boldsymbol{y}) \quad (\boldsymbol{x} \in \mathbb{R}^n,\, \boldsymbol{y} \in \mathbb{R}^m) \tag{D.98}$$

となる．このことを踏まえて以下の主問題と双対問題を導入することにする．

$$\text{minimize} \left\{ f(\boldsymbol{x}) + g\bigl(\mathbf{A}\boldsymbol{x}\bigr) \right\} \quad (\boldsymbol{x} \in \mathbb{R}^n) \tag{D.99}$$

と双対問題

$$\text{maximize} \left\{ -f^*\bigl(\mathbf{A}^\top \boldsymbol{y}\bigr) - g^*(-\boldsymbol{y}) \right\} \quad (\boldsymbol{y} \in \mathbb{R}^m). \tag{D.100}$$

このとき (D.98) は弱双対性を保証する．

以上の議論から以下の定理を得る．

定理 D.57. 不等式

$$\sup_{\boldsymbol{y} \in \mathbb{R}^m} \left\{ -f^*\bigl(\mathbf{A}^\top \boldsymbol{y}\bigr) - g^*(-\boldsymbol{y}) \right\} \leq \inf_{\boldsymbol{x} \in \mathbb{R}^n} \left\{ f(\boldsymbol{x}) + g\bigl(\mathbf{A}\boldsymbol{x}\bigr) \right\}$$

が成立する．

定義 D.58. 関数 $v \colon \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ を

$$v(\boldsymbol{a}) := \inf_{\boldsymbol{x} \in \mathbb{R}^n} \bigl\{ f(\boldsymbol{x}) + g\bigl(\mathbf{A}\boldsymbol{x} + \boldsymbol{a}\bigr) \bigr\} \quad (\boldsymbol{a} \in \mathbb{R}^m)$$

で定める．

補題 D.59. 関数 $v \colon \mathbb{R}^m \to (-\infty,\, \infty) \cup \{+\infty\}$ は凸かつ下半連続である．

Proof. <u>凸性の証明</u>: 任意の $\boldsymbol{a}, \widetilde{\boldsymbol{a}} \in \mathbb{R}^m$ と $0 \le \lambda \le 1$ に対して, 関数 g の凸性から

$$g\bigl(\mathbf{A}\boldsymbol{x} + (\lambda\boldsymbol{a} + (1-\lambda)\widetilde{\boldsymbol{a}})\bigr) \le \lambda g\bigl(\mathbf{A}\boldsymbol{x} + \boldsymbol{a}\bigr) + (1-\lambda) g\bigl(\mathbf{A}\boldsymbol{x} + \widetilde{\boldsymbol{a}}\bigr)$$

となる. 上の不等式と inf の性質から

$$\begin{aligned}
v\bigl(\lambda\boldsymbol{a} + (1-\lambda)\widetilde{\boldsymbol{a}}\bigr) &\le \inf_{\boldsymbol{x}\in\mathbb{R}^n}\Bigl\{\mathbf{A}\boldsymbol{x} + \lambda g\bigl(\mathbf{A}\boldsymbol{x}+\boldsymbol{a}\bigr) + (1-\lambda) g\bigl(\mathbf{A}\boldsymbol{x}+\widetilde{\boldsymbol{a}}\bigr)\Bigr\} \\
&= \inf_{\boldsymbol{x}\in\mathbb{R}^n}\Bigl\{\lambda\bigl(\mathbf{A}\boldsymbol{x} g\bigl(\mathbf{A}\boldsymbol{x}+\boldsymbol{a}\bigr)\bigr) + (1-\lambda)\bigl(\mathbf{A}\boldsymbol{x} + g\bigl(\mathbf{A}\boldsymbol{x}+\widetilde{\boldsymbol{a}}\bigr)\bigr)\Bigr\} \\
&\le \lambda \inf_{\boldsymbol{x}\in\mathbb{R}^n}\Bigl\{\mathbf{A}\boldsymbol{x} + g\bigl(\mathbf{A}\boldsymbol{x}+\boldsymbol{a}\bigr)\Bigr\} \\
&\quad + (1-\lambda)\inf_{\boldsymbol{x}\in\mathbb{R}^n}\Bigl\{\mathbf{A}\boldsymbol{x} + g\bigl(\mathbf{A}\boldsymbol{x}+\widetilde{\boldsymbol{a}}\bigr)\Bigr\} \\
&= \lambda v(\boldsymbol{a}) + (1-\lambda) v(\widetilde{\boldsymbol{a}})
\end{aligned}$$

を得る. よって, 関数 v は凸であることがわかる.
<u>下半連続性の証明</u>: 任意の $\boldsymbol{a} \in \mathbb{R}^m$ に対して, 任意の数列 $\{\boldsymbol{a}_k\}_{k=1}^{\infty} \subset \mathbb{R}^m$ で $\lim_{k\to\infty}\boldsymbol{a}_k = \boldsymbol{a}$ となるものを取る. このとき, 関数 g の下半連続性から

$$\begin{aligned}
\liminf_{k\to\infty} v(\boldsymbol{a}_k) &= \liminf_{k\to\infty}\Bigl\{f(\boldsymbol{x}) + g\bigl(\mathbf{A}\boldsymbol{x}+\boldsymbol{a}_k\bigr)\Bigr\} \\
&= f(\boldsymbol{x}) + \liminf_{k\to\infty} g\bigl(\mathbf{A}\boldsymbol{x}+\boldsymbol{a}_k\bigr) \\
&= f(\boldsymbol{x}) + g\bigl(\mathbf{A}\boldsymbol{x}+\boldsymbol{a}\bigr) \quad (\because g \text{ は下半連続})
\end{aligned}$$

を得る. よって 関数 v も下半連続である. □

定義 D.60. 関数 f, g と行列 $\mathbf{A}$ の組が**双対条件**をみたすとは

$$\text{値関数 } v \text{ は原点の近くで } +\infty \text{ の値を取らない} \tag{D.101}$$

をみたすときをいう.

注意 D.61. 条件 (D.101) は次のように書きかえることができる. ある小さな正数 $\delta > 0$ があって $\boldsymbol{a} \in \mathbb{R}^m$ が $|\boldsymbol{a}|_{2,m} \le \delta$ をみたすとき, ある $\boldsymbol{y} \in \mathrm{dom}\ g$, $\boldsymbol{x} \in \mathrm{dom}\ f$ があって

$$\boldsymbol{a} = \boldsymbol{y} - \mathbf{A}\boldsymbol{x} \tag{D.102}$$

とかけることである. したがって

$$v(\boldsymbol{a}) \le f(\boldsymbol{x}) + g\bigl(\mathbf{A}\boldsymbol{x}+\boldsymbol{a}\bigr) = f(\boldsymbol{x}) + g(\boldsymbol{y}) < \infty$$

となる. □

定理 D.62. $\boldsymbol{x}^* \in \mathbb{R}^n$ が主問題 (D.99) の解とし, 関数 f, g と行列 $\mathbf{A}$ は双対条件 (D.101) をみたすとする. このとき以下が成立する.
(1) $\boldsymbol{y}^* \in \mathbb{R}^m$ が存在して

$$f(\boldsymbol{x}^*) + g\big(\mathbf{A}\boldsymbol{x}\big) = -f^*\big(\mathbf{A}^\top \boldsymbol{y}^*\big) - g^*(-\boldsymbol{y}^*)$$

をみたす.
(2) 強双対性

$$\inf_{\boldsymbol{x}\in\mathbb{R}^n}\Big\{f(\boldsymbol{x}) + g\big(\mathbf{A}\boldsymbol{x}\big)\Big\} = \sup_{\boldsymbol{y}\in\mathbb{R}^m}\Big\{-f^*\big(\mathbf{A}^\top \boldsymbol{y}^*\big) - g^*(-\boldsymbol{y}^*)\Big\} \tag{D.103}$$

が成立する.

Proof. 注意 D.61 より, 双対条件 (D.101) は原点の近くで有限の値を取ることがわかる. よって, 関数 $v:\mathbb{R}^m \to \mathbb{R}$ は凸であることを合わせて考えると $\partial v(\mathbf{0}) \neq \varnothing$ となることがわかる. そこで

$$-\boldsymbol{y}^* \in \partial v(\mathbf{0}_m)$$

を取る. このとき $\boldsymbol{x} \in \mathbb{R}^n$ と $\boldsymbol{a} \in \mathbb{R}^m$ に対して

$$\begin{aligned}
f(\boldsymbol{x}^*) + g\big(\mathbf{A}\boldsymbol{x}\big) &= v(\mathbf{0}_m) \leq v(\boldsymbol{a}) + \langle \boldsymbol{a},\, \boldsymbol{y}^*\rangle_{2,m} \quad (\because -\boldsymbol{y}^* \in \partial v(\mathbf{0}_m)) \\
&\leq f(\boldsymbol{x}) + g\big(\mathbf{A}\boldsymbol{x} + \boldsymbol{a}\big) + \langle \boldsymbol{a},\, \boldsymbol{y}^*\rangle_{2,m} \\
&= -\Big(\big\langle \mathbf{A}^\top \boldsymbol{y}^*,\, \boldsymbol{x}\big\rangle_{2,n} - f(\boldsymbol{x})\Big) \\
&\qquad - \Big(-\big\langle \boldsymbol{y}^*,\, \mathbf{A}\boldsymbol{x} + \boldsymbol{a}\big\rangle_{2,n} - g\big(\mathbf{A}\boldsymbol{x} + \boldsymbol{a}\big)\Big)
\end{aligned}$$

となる. $\boldsymbol{a}$ に関して infimum を取ると

$$f(\boldsymbol{x}^*) + g\big(\mathbf{A}\boldsymbol{x}\big) \leq -\Big(\big\langle \mathbf{A}^\top \boldsymbol{y}^*,\, \boldsymbol{x}\big\rangle_{2,n} - f(\boldsymbol{x})\Big) - g^*(-\boldsymbol{y}^*)$$

となる. さらに $\boldsymbol{x}$ に関して infimum を取ると

$$f(\boldsymbol{x}^*) + g\big(\mathbf{A}\boldsymbol{x}^*\big) \leq -f\big(\mathbf{A}^\top \boldsymbol{y}^*\big) - g^*(-\boldsymbol{y}^*)$$

を得る.
　逆向きの不等式は定理 D.57 からわかる. □

D.10　半正定値計画法

$\boxed{\text{記号}}$(1) $\mathrm{Sym}(n;\,\mathbb{R})$ を $n\times n$ 対称行列全体の成す線型部分空間とし

$$\mathrm{Sym}^+(n;\,\mathbb{R}) = \{\mathbf{A}\in\mathrm{Sym}(n;\,\mathbb{R});\ \mathbf{A}\succeq\mathbf{0}_{n\times n}\}$$

とする. ただし

$$\mathbf{A}\succeq\mathbf{0}_{n\times n}\ \Leftrightarrow\ \boldsymbol{x}^\top\mathbf{A}\boldsymbol{x}\geq 0\,(\forall\boldsymbol{x}\in\mathbb{R}^n)$$

である.

(2) 行列 $\mathbf{A},\,\mathbf{B}\in\mathrm{Sym}(n;\,\mathbb{R})$ の内積を

$$\big\langle\mathbf{A},\,\mathbf{B}\big\rangle_{\mathrm{F}} = \sum_{i,\,j=1}^{n} a_{ij}b_{ij}\qquad(\mathbf{A}=(a_{ij}),\,\mathbf{B}=(b_{ij}))$$

で定める.

与えられた $\mathbf{A}_k,\,\mathbf{C}\in\mathrm{Sym}(n;\,\mathbb{R})\,(k=1,\,2,\,\ldots,\,m)$ と $b_k\in\mathbb{R}\,(k=1,\,2,\,\ldots,\,k)$ に対して半正定値計画問題

$$\begin{aligned}&\text{minimize }\big\langle\mathbf{C},\,\mathbf{X}\big\rangle_{\mathrm{F}} \\ &\text{subject to }\big\langle\mathbf{A}_1,\,\mathbf{X}\big\rangle_{\mathrm{F}}=b_1,\,\big\langle\mathbf{A}_2,\,\mathbf{X}\big\rangle_{\mathrm{F}}=b_2,\,\ldots,\,\big\langle\mathbf{A}_m,\,\mathbf{X}\big\rangle_{\mathrm{F}}=b_m \\ &\qquad\mathbf{X}\succeq\mathbf{0}_{n\times n}\end{aligned}\tag{D.104}$$

を考える.

<u>**双対性**</u>. まず写像 $\boldsymbol{\alpha}:\,\mathrm{Sym}(n;\,\mathbb{R})\,\to\,\mathbb{R}^m$ を

$$\boldsymbol{\alpha}(\mathbf{X}) = \big[\big\langle\mathbf{A}_1,\,\mathbf{X}\big\rangle_{\mathrm{F}},\,\big\langle\mathbf{A}_2,\,\mathbf{X}\big\rangle_{\mathrm{F}},\,\ldots,\,\big\langle\mathbf{A}_m,\,\mathbf{X}\big\rangle_{\mathrm{F}}\big]^\top$$

と定める. さらに写像 $\boldsymbol{\alpha}^\top:\,\mathbb{R}^m\,\to\,\mathrm{Sym}(n;\,\mathbb{R})$ を

$$\big\langle\boldsymbol{\alpha}^\top(\boldsymbol{y}),\,\mathbf{X}\big\rangle_{\mathrm{F}} = \big\langle\boldsymbol{y},\,\boldsymbol{\alpha}(\mathbf{X})\big\rangle_{2,\,m} = \sum_{k=1}^{m} y_k\big\langle\mathbf{A}_k,\,\mathbf{X}\big\rangle_{\mathbf{F}}\qquad(\forall\mathbf{X}\in\mathrm{Sym}(n;\,\mathbb{R}^n))$$

で定める. 主問題 (D.104) を

$$\text{minimize }f(\mathbf{X})+g\big(\boldsymbol{\alpha}(\mathbf{X})\big)\tag{D.105}$$

と書きかえる. ただし, $\boldsymbol{b}=(b_1,\,b_2,\,\ldots,\,b_m)^\top$ で

$$f(\mathbf{X}) = \begin{cases}\big\langle\mathbf{C},\,\mathbf{X}\big\rangle_{\mathrm{F}} & (\mathbf{X}\succeq\mathbf{0}) \\ +\infty & (\text{その他の場合})\end{cases};\tag{D.106}$$

$$g(\boldsymbol{y}) = \begin{cases}0 & (\boldsymbol{y}=\boldsymbol{b}) \\ +\infty & (\text{その他の場合})\end{cases}\tag{D.107}$$

である．このとき $\boldsymbol{z} \in \mathbb{R}^m$ に対し

$$g^*(\boldsymbol{z}) = \sup_{\boldsymbol{y} \in \mathbb{R}^m} \left\{ \langle \boldsymbol{z},\, \boldsymbol{y} \rangle_{2,\,m} - g(\boldsymbol{y}) \right\} = \langle \boldsymbol{z},\, \boldsymbol{b} \rangle_{2,\,m} \tag{D.108}$$

$$(\because\ \boldsymbol{y} \neq \boldsymbol{b} \text{ のとき } g(\boldsymbol{y}) = +\infty\ )$$

となる．さらに $\mathbf{W} \in \mathrm{Sym}(n;\, \mathbb{R})$ に対して

$$f^*(\mathbf{W}) = \sup_{\mathbf{X} \in \mathrm{Sym}(n;\,\mathbb{R})} \left\{ \left\langle \mathbf{W},\, \mathbf{X} \right\rangle_{\mathrm{F}} - f(\mathbf{W}) \right\} \tag{D.109}$$

$$= \sup_{\mathbf{X} \in \mathrm{Sym}(n;\,\mathbb{R})} \left\{ \left\langle \mathbf{W} - \mathbf{C},\, \mathbf{X} \right\rangle_{\mathrm{F}} \right\} \tag{D.110}$$

$$= \begin{cases} 0 & (\mathbf{C} \preceq \mathbf{W}) \\ +\infty & (\text{その他の場合}) \end{cases} \tag{D.111}$$

$$(\because \mathbf{X} \succeq \mathbf{0}_n \text{ でなければ } f(\mathbf{X}) = +\infty)$$

となる．ただし

$$\mathbf{C} \preceq \mathbf{W} \ \Leftrightarrow\ \mathbf{W} - \mathbf{C} \succeq \mathbf{0}$$

である．さらに, 任意の $\mathbf{X} \in \mathrm{Sym}(n;\, \mathbb{R})$ と $\boldsymbol{y} = (y_1,\, y_2,\, \ldots,\, y_m)^\top \in \mathbb{R}^m$ に対し

$$\left\langle \boldsymbol{\alpha}^\top \boldsymbol{y},\, \mathbf{X} \right\rangle_{\mathrm{F}} = \left\langle \boldsymbol{y},\, \boldsymbol{\alpha}\mathbf{X} \right\rangle_{2,\,n} = \left\langle \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix},\, \begin{bmatrix} \left\langle \mathbf{A}_1,\, \mathbf{X} \right\rangle_{\mathrm{F}} \\ \left\langle \mathbf{A}_2,\, \mathbf{X} \right\rangle_{\mathrm{F}} \\ \vdots \\ \left\langle \mathbf{A}_m,\, \mathbf{X} \right\rangle_{\mathrm{F}} \end{bmatrix} \right\rangle_{2,\,m}$$

$$= \sum_{k=1}^{m} y_k \left\langle \mathbf{A}_k,\, \mathbf{X} \right\rangle_{\mathrm{F}} = \left\langle \sum_{k=1}^{m} y_k \mathbf{A}_k,\, \mathbf{X} \right\rangle_{\mathrm{F}}$$

から

$$\boldsymbol{\alpha}^\top \boldsymbol{y} = \sum_{k=1}^{m} y_k \mathbf{A}_k$$

がわかる．

Fenchel **双対問題**.

$$f^*(\boldsymbol{\alpha}^\top \boldsymbol{y}) - g^*(-\boldsymbol{y})$$

を最大化することは半正定値双対問題

$$\text{minimize } \left\langle \boldsymbol{y},\, \boldsymbol{b} \right\rangle_{2,\,m} \qquad \text{subject to } \sum_{k=1}^{m} y_k \mathbf{A}_k \preceq \mathbf{C} (\Leftrightarrow\ f^*(\boldsymbol{\alpha}^\top \boldsymbol{y}) = 0) \tag{D.112}$$

となる．これは線型計画法の双対問題のアプローチとなる．ただし制約は対称行列に対する不等式制約である．さらに主問題 (D.104) は対称行列 $\mathbf{X}$ が未知であり, 双対問題 (D.105) はベクトル $\boldsymbol{y} \in \mathbb{R}^m$ が未知である．

定理 D.63. 主問題 (D.104) の実行可能解 $\overline{\mathbf{X}}$ で

$$\overline{\mathbf{X}} \succeq \mathbf{0}_{n\times n} \tag{D.113}$$

をみたすものが存在したとする. さらに写像

$$\boldsymbol{\alpha}\colon \mathrm{Sym}(n;\,\mathbb{R}) \to \mathbb{R}^m$$

は全射とする. このとき主問題 (D.104) の解を $\mathbf{X}^*$ としたとき双対問題 (D.105) の実行解 $\boldsymbol{y}^*$ が存在し

$$\left\langle \mathbf{C},\,\mathbf{X}^* \right\rangle_{\mathrm{F}} = \left\langle \boldsymbol{b},\,\boldsymbol{y}^* \right\rangle_{2,\,m}$$

をみたす.

Proof. (D.112) を考慮すると, 十分小さな正数 $\lambda > 0$ が存在して

$$\left\langle \mathbf{Y},\,\mathbf{Y} \right\rangle_{\mathrm{F}} \le \lambda^2 \;\Rightarrow\; \mathbf{X} = \overline{\mathbf{X}} + \mathbf{Y} \in \mathrm{Sym}^+(n;\,\mathbb{R})$$

とできる. 写像

$$\boldsymbol{\alpha}\colon \mathrm{Sym}(n;\,\mathbb{R}) \to \mathbb{R}^m$$

は全射なので, ある正数 $\eta > 0$ が存在して

$$\boldsymbol{\alpha}(\mathbb{B}_{\mathrm{Sym}}(\mathbf{0}_m,\,1)) \supset \mathbb{B}_{\mathbb{R}^m}(\mathbf{0}_m,\,\lambda) = \{\boldsymbol{y} \in \mathbb{R}^m;\,|\boldsymbol{y}|_{2,\,m} \le \lambda\}$$

とできる. ただし $\mathbb{B}_{\mathrm{Sym}}(\mathbf{0}_m,\,1) = \{\mathbf{A} \in \mathrm{Sym}(n;\,\mathbb{R});\,\left\langle \mathbf{A},\,\mathbf{A} \right\rangle_{\mathrm{F}} \le 1\}$ である. よって, 各 $\boldsymbol{a} \in \mathbb{B}_{\mathbb{R}^m}(\mathbf{0}_m,\,\lambda\eta)$ に対して $\mathbf{X} = \overline{\mathbf{X}} + \mathbf{Y} \in \mathrm{Sym}^+(n;\,\mathbb{R})$ かつ $\mathbf{Y} \in \mathbb{B}_{\mathrm{Sym}}(\mathbf{0}_m,\,\lambda)$ が存在して

$$\boldsymbol{\alpha}(\mathbf{X}) = \boldsymbol{\alpha}(\overline{\mathbf{X}}) + \boldsymbol{\alpha}(\mathbf{Y}) = \boldsymbol{b} - \boldsymbol{a}$$

となる.

関数 f, g の定義 (D.106) と (D.107) から

$$\mathrm{dom}\, f = \mathrm{Sym}^+(n;\,\mathbb{R}), \quad \mathrm{dom}\, g = \{\boldsymbol{b}\}$$

である. したがって $\lambda = \delta\eta$ かつ $|\boldsymbol{a}|_{2,\,m} \le \delta$ のとき (D.102) から

$$\boldsymbol{a} = \boldsymbol{b} - \boldsymbol{\alpha}(\mathbf{X})$$

と書ける. 結局双対条件 (D.101) をみたす. すると値関数

$$v(\boldsymbol{a}) = \inf_{\mathbf{X} \in \mathrm{Sym}(n;\,\mathbb{R})} \bigl\{ f(\mathbf{X}) + g\bigl(\boldsymbol{\alpha}(\mathbf{X}) + \boldsymbol{a}\bigr) \bigr\}$$

は原点近くで有限の値を取るので, $\partial v(\mathbf{0}_m) \neq \varnothing$ となる. そこで

$$-\boldsymbol{y}^* \in \partial v(\mathbf{0}_m)$$

を取る. このとき, 任意の $\mathbf{X} \in \mathrm{Sym}(n; \mathbb{R})$ と $\boldsymbol{a} \in \mathbb{R}^m$ に対して

$$\begin{aligned} f(\mathbf{X}^*) + g\big(\mathbf{A}\mathbf{X}\big) &= v(\mathbf{0}) \leq v(\boldsymbol{a}) + \langle \boldsymbol{a},\, \boldsymbol{y}^* \rangle_{2,\,m} \\ &\leq f(\mathbf{X}) + g\big(\boldsymbol{\alpha}(\mathbf{X}) + \boldsymbol{a}\big) + \langle \boldsymbol{a},\, \boldsymbol{y}^* \rangle_{2,\,m} \\ &= -\left(\big\langle \boldsymbol{\alpha}^\top \boldsymbol{y},\, \mathbf{X} \big\rangle_{\mathrm{F}} - f(\mathbf{X})\right) \\ &\qquad - \left(-\big\langle \boldsymbol{y},\, \boldsymbol{\alpha}(\mathbf{X}) + \boldsymbol{a} \big\rangle_{2,\,m} - g\big(\boldsymbol{\alpha}(\mathbf{X}) + \boldsymbol{a}\big)\right) \end{aligned}$$

を得る. 上の式の最右辺を $\boldsymbol{a}$ に関して infimum を取ると

$$f(\mathbf{X}^*) + g\big(\boldsymbol{\alpha}(\mathbf{X}^*)\big) \leq -\left(\big\langle \boldsymbol{\alpha}^\top(\boldsymbol{y}),\, \mathbf{X} \big\rangle_{\mathrm{F}} - f(\mathbf{X})\right) - g^*(-\boldsymbol{y}^*)$$

を得る. さらに上の式の右辺を $\mathbf{X}$ に関して infimum を取ると

$$f(\mathbf{X}^*) + g\big(\boldsymbol{\alpha}(\mathbf{X}^*)\big) \leq -f^*\big(\boldsymbol{\alpha}^\top(\boldsymbol{y}^*)\big) - g^*(-\boldsymbol{y}^*)$$

を得る. 逆向きの不等式は常に成立するので,

$$f(\mathbf{X}^*) + g\big(\boldsymbol{\alpha}(\mathbf{X}^*)\big) = -f^*\big(\boldsymbol{\alpha}^\top(\boldsymbol{y}^*)\big) - g^*(-\boldsymbol{y}^*) \tag{D.114}$$

を得る. (**??**) – (D.111) から

$$\begin{aligned} f(\mathbf{X}^*) &= \big\langle \mathbf{C},\, \mathbf{X}^* \big\rangle_{\mathrm{F}}, \\ g\big(\boldsymbol{\alpha}(\mathbf{X}^*)\big) &= 0 (\Leftrightarrow\ \boldsymbol{\alpha}(\mathbf{X}^*) = \boldsymbol{b}), \\ f^*\big(\boldsymbol{\alpha}^\top(\boldsymbol{y}^*)\big) &= 0\,(\Leftrightarrow\ \boldsymbol{\alpha}^\top(\boldsymbol{y}) \preceq \mathbf{C}), \\ g^*(-\boldsymbol{y}^*) &= -\langle \boldsymbol{y}^*,\, \boldsymbol{b} \rangle_{2,\,m} \end{aligned}$$

となるので, これらを (D.114) に代入すると

$$\big\langle \mathbf{C},\, \mathbf{X}^* \big\rangle_{\mathrm{F}} = \langle \boldsymbol{y}^*,\, \boldsymbol{b} \rangle_{2,\,m}$$

がわかる. □

D.11　最適化アルゴリズム

この節では, 3 つの基本的なアルゴリズムである最急降下法 (steepest-descent method), Newton 法, 共役勾配法 (conjugate-gradient method) について説明をする.　最初の 2 つのアルゴリズムに基づいて多くの最適化がなされている. 最急降下法は収束のスピードは遅く実用的ではないが, 簡単で信頼性が高い. Newton 法は非線型方程式の解をみつける一般的な方法である. しかし解の近傍に初期値がないと停留点に収束しない場合がある. 共役勾配法は凸 2 次計画法における最適化に用いると効果を発揮する.

D.11.1　多変数関数の微分

定義 D.64. $U \subset \mathbb{R}^d$ を空でない開部分集合とし，$f : U \to \mathbb{R}$ を関数とする．$\boldsymbol{x} = (x_1, x_2, \ldots, x_d)^\top \in U$ に対し，極限

$$\frac{\partial f}{\partial x_i}(\boldsymbol{x}) := \lim_{t \to 0} \frac{f(x_1, \ldots, x_{i-1}, x_i + t, x_{i+1}, \ldots, x_d) - f(\boldsymbol{x})}{t} \qquad (i = 1, 2, \ldots, \mathrm{d})$$

が存在するとき，$\dfrac{\partial f}{\partial x_i}(\boldsymbol{x})$ 点 $\boldsymbol{x}$ における x_i に関する f の**偏微分係数**という．すべての変数に関して偏微分係数が存在するとき，ベクトル

$$\nabla f(\boldsymbol{x}) := \left(\frac{\partial f}{\partial x_1}(\boldsymbol{x}), \frac{\partial f}{\partial x_2}(\boldsymbol{x}), \ldots, \frac{\partial f}{\partial x_d}(\boldsymbol{x}) \right)^\top$$

は f の**勾配** (gradient) という．

注意 D.65. $\boldsymbol{b} = (b_1, b_2, \ldots, b_d)^\top \in \mathbb{R}^d$ とし

$$\boldsymbol{e}_i := (\underbrace{0, 0, \ldots, 0}_{i-1}, 1, 0, \ldots, 0)^\top \qquad (i = 1, 2, \ldots, \mathrm{d})$$

とする．すると

$$\boldsymbol{b} = b_1 \boldsymbol{e}_1 + b_2 \boldsymbol{e}_2 + \cdots + b_d \boldsymbol{e}_d$$

となる．□

定義 D.66. 方向 $\boldsymbol{d} \in \mathbb{R}^d$ に沿った点 $\boldsymbol{x} \in U$ における f の**方向微分**を

$$\overset{\bullet}{f}(\boldsymbol{x}; \boldsymbol{d}) = \lim_{t \searrow 0} \frac{f(\boldsymbol{x} + t\boldsymbol{d}) - f(\boldsymbol{x})}{t}$$

で定める．ただし $t > 0$ が 0 に近づくとき上の式の右辺の極限が存在するときに方向微分を定める．

明らかに $\alpha \geq 0$ に対して

$$\overset{\bullet}{f}(\boldsymbol{x}; \alpha\boldsymbol{d}) = \alpha \overset{\bullet}{f}(\boldsymbol{x}; \boldsymbol{d})$$

となる．したがって $\overset{\bullet}{f}(\boldsymbol{x}; -\boldsymbol{d}) = -\overset{\bullet}{f}(\boldsymbol{x}; \boldsymbol{d})$ であるので

$$\overset{\bullet}{f}(\boldsymbol{x}; \boldsymbol{d}) = \lim_{t \to 0} \frac{f(\boldsymbol{x} + t\boldsymbol{d}) - f(\boldsymbol{x})}{t}$$

となる．実際

$$\overset{\bullet}{f}(\boldsymbol{x}; \boldsymbol{d}) = -\overset{\bullet}{f}(\boldsymbol{x}; -\boldsymbol{d}) = -\lim_{t \searrow 0} \frac{f(\boldsymbol{x} - t\boldsymbol{d}) - f(\boldsymbol{x})}{t} = \lim_{s \nearrow 0} \frac{f(\boldsymbol{x} + s\boldsymbol{d}) - \boldsymbol{x})}{s}$$

よりわかる．

定義 D.67. 関数 $f : U \to \mathbb{R}$ は点 $\boldsymbol{x} \in U$ で Gâteaux 微分可能であるとは, 任意の方向 $\boldsymbol{d} \in \mathbb{R}^d$ に対して, 方向微分 $\overset{\bullet}{f}(\boldsymbol{x};\, \boldsymbol{d})$ は存在し, それが $\boldsymbol{d}$ の線型関数であるときをいう. f が U の任意の点 $\boldsymbol{x}$ で Gâtaeaux 微分可能のとき, f は U 上で Gâtaeux **微分可能**という.

注意 D.68. 関数 f が点 $\boldsymbol{x}$ において Gâteaux 微分可能とする. このとき, $\boldsymbol{d} = d_1\boldsymbol{e}_1 + d_2\boldsymbol{e}_2 + \cdots + d_d\boldsymbol{e}_d$ とおくと方向微分 $\overset{\bullet}{f}(\boldsymbol{x};\, \boldsymbol{d})$ は $\boldsymbol{d}$ の線型関数なので

$$
\begin{aligned}
\overset{\bullet}{f}(\boldsymbol{x};\, \boldsymbol{d}) &= \overset{\bullet}{f}(\boldsymbol{x};\, d_1\boldsymbol{e}_1 + d_2\boldsymbol{e}_2 + \cdots + d_d\boldsymbol{e}_d) \\
&= d_1 \overset{\bullet}{f}(\boldsymbol{x};\, \boldsymbol{e}_1) + d_2 \overset{\bullet}{f}(\boldsymbol{x};\, \boldsymbol{e}_2) + \cdots + d_d \overset{\bullet}{f}(\boldsymbol{x};\, \boldsymbol{e}_d) \\
&= \langle \boldsymbol{d},\, \nabla f(\boldsymbol{x}) \rangle_{2,\mathrm{d}}
\end{aligned}
$$

となる. ただし, $\langle\cdot,\cdot\rangle_{2,\mathrm{d}}$ は $\mathbb{R}^d$ 上の Euclid 内積である. □

定義 D.69. 関数 $f : U \to \mathbb{R}$ は点 $\boldsymbol{x}$ で Fréchet 微分可能であるとは, ある線型関数 $\ell : \mathbb{R}^d \to \mathbb{R}$ で $\ell(\boldsymbol{x}) = \langle \boldsymbol{\ell},\, \boldsymbol{x} \rangle_{2,\mathrm{d}}$ ($\boldsymbol{\ell} \in \mathbb{R}^d$) なるものが存在して

$$
\lim_{|\boldsymbol{h}|_{2,\mathrm{d}} \to 0} \frac{f(\boldsymbol{x}+\boldsymbol{h}) - f(\boldsymbol{x}) - \langle \boldsymbol{\ell},\, \boldsymbol{h} \rangle_{2,\mathrm{d}}}{|\boldsymbol{h}|_{2,\mathrm{d}}} = 0 \tag{D.115}
$$

が成り立つときをいう. ただし, $|\cdot|_{2,\mathrm{d}} = \sqrt{\langle\cdot,\cdot\rangle_{2,\mathrm{d}}}$ である.

注意 D.70. ベクトル $o(\boldsymbol{h}) \in \mathbb{R}^d$ は

$$
\lim_{|\boldsymbol{h}|_{2,\mathrm{d}} \to 0} \frac{o(\boldsymbol{h})}{|\boldsymbol{h}|_{2,\mathrm{d}}} = 0
$$

をみたすものとする. この記号を用いると f が点 $\boldsymbol{x}$ で Fréchet 微分可能であることは

$$
f(\boldsymbol{x}+\boldsymbol{h}) = f(\boldsymbol{x}) + \langle \boldsymbol{\ell},\, \boldsymbol{h} \rangle_{2,\mathrm{d}} + o(\boldsymbol{h}) \qquad (\boldsymbol{h} \to \boldsymbol{0}_d) \tag{D.116}
$$

と書けることがわかる.

$\boldsymbol{h} = t\boldsymbol{e}_i$ ($i = 1, 2, \ldots, \mathrm{d}$) と $t \neq 0$ と取る. $\boldsymbol{\ell} = (\ell_1, \ell_2, \ldots, \ell_d)^\top$ としたとき, (D.115) は

$$
\lim_{t \to 0} \frac{f(\boldsymbol{x}+t\boldsymbol{e}_i) - f(\boldsymbol{x}) - t\ell_i}{t} = 0
$$

と書きなおすことができる. このとき (D.116) は

$$
f(\boldsymbol{x}+\boldsymbol{h}) = f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}),\, \boldsymbol{h} \rangle_{2,\mathrm{d}} + o(\boldsymbol{h}) \qquad (\boldsymbol{h} \to \boldsymbol{0}_d)
$$

となる. □

定理 D.71. $U \subset \mathbb{R}^d$ は空でない開部分集合とし, 関数 $f : U \to \mathbb{R}$ は点 $\boldsymbol{x} \in U$ で Féchet 微分可能とする. このとき, 関数 f は点 $\boldsymbol{x}$ で Gâteaux 微分可能である.

Proof. Fréchet 微分と Gâteaux 微分の定義から明らか. □

定義 D.72. $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ を点とする. $[\boldsymbol{x}, \boldsymbol{y}]$ は $\boldsymbol{x}$ と $\boldsymbol{y}$ の間の線分とする. すなわち

$$[\boldsymbol{x}, \boldsymbol{y}] = \{\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x}) \in \mathbb{R}^d;\, 0 \leq t \leq 1\}$$

である. 同様に, 両端点を含まない $\boldsymbol{x}$ と $\boldsymbol{y}$ の間の線分を

$$(\boldsymbol{x}, \boldsymbol{y}) = \{\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x}) \in \mathbb{R}^d;\, 0 < t < 1\}$$

とする. さらに, $\boldsymbol{x} = \boldsymbol{y}$ のとき

$$[\boldsymbol{x}, \boldsymbol{y}] = (\boldsymbol{x}, \boldsymbol{y}) = \{\boldsymbol{x}\}$$

と定める.

補題 D.73. $U \subset \mathbb{R}^d$ を空でない開部分集合とし, 関数 $f : U \to \mathbb{R}$ は U 上で Gâteaux 微分可能とする. $\boldsymbol{x}, \boldsymbol{y} \in U$ は異なる点で $[\boldsymbol{x}, \boldsymbol{y}] \subset U$ とする. このとき, 点 $\boldsymbol{z} \in (\boldsymbol{x}, \boldsymbol{y})$ が存在して

$$f(\boldsymbol{y}) = f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{z}), \boldsymbol{y} - \boldsymbol{x} \rangle_{2,\mathrm{d}}$$

とすることができる.

Proof. $\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x}) \in U$ なる任意の $t \in \mathbb{R}$ に対し

$$h(t) := f(\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x}))$$

とおく. f は Gâteaux 微分可能なので, $h(t)$ は微分可能で

$$\overset{\bullet}{h}(t) = \langle \nabla f(\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x})), \boldsymbol{y} - \boldsymbol{x} \rangle_{2,\mathrm{d}}$$

となる. 1 変数実数値関数の中間値の定理から, ある $0 < \bar{t} < 1$ が存在して

$$h(1) - h(0) = \overset{\bullet}{h}(\bar{t}) = f(\boldsymbol{x} + \bar{t}(\boldsymbol{y} - \boldsymbol{x})) = \langle \nabla f(\boldsymbol{x} + \bar{t}(\boldsymbol{y} - \boldsymbol{x})), \boldsymbol{y} - \boldsymbol{x} \rangle_{2,\mathrm{d}}$$

と書ける. ここで, $\boldsymbol{z} := \boldsymbol{x} + \bar{t}(\boldsymbol{y} - \boldsymbol{x})$ とおく. $h(0) = f(\boldsymbol{x})$ かつ $h(1) = f(\boldsymbol{y})$ となるので

$$f(\boldsymbol{x}) - f(\boldsymbol{y}) = \langle \nabla f(\boldsymbol{z}), \boldsymbol{y} - \boldsymbol{x} \rangle_{2,\mathrm{d}}$$

となる. □

定理 D.74. $U \subset \mathbb{R}^d$ を空でない開部分集合とし, $f : U \to \mathbb{R}$ を関数とする．関数 f は点 $\boldsymbol{x}_0 \in U$ で Gâteaux 微分可能で, $\partial f/\partial x_i \, (i = 1, 2, \ldots, \mathrm{d})$ は点 $\boldsymbol{x}_0$ で連続微分可能ならば, f は点 $\boldsymbol{x}_0$ で Fréchet 微分可能である．

Proof. 補題 D.73 より任意の $\boldsymbol{h} \, (\neq \boldsymbol{0}_d) \in \mathbb{R}^d$ に対して，ある $\overline{\boldsymbol{x}} \in (\boldsymbol{x}, \boldsymbol{x} + \boldsymbol{h})$ が存在し

$$f(\boldsymbol{x}_0 + \boldsymbol{h}) - f(\boldsymbol{x}_0) - \langle \nabla f(\boldsymbol{x}_0), \boldsymbol{h} \rangle_{2,\mathrm{d}} = \langle \nabla f(\overline{\boldsymbol{x}}) - \nabla f(\boldsymbol{x}_0), \boldsymbol{h} \rangle_{2,\mathrm{d}}$$

と書ける．さらに, f の勾配 ∇f は点 $\boldsymbol{x}_0$ で連続であるので

$$|\nabla f(\overline{\boldsymbol{x}}) - \nabla f(\boldsymbol{x}_0)|_{2,\mathrm{d}} \to 0 \quad (|\boldsymbol{h}|_{2,\mathrm{d}} \to 0)$$

となる．Cauchy-Schwarz の不等式とこのことから

$$\begin{aligned} \lim_{|\boldsymbol{h}|_{2,\mathrm{d}} \to 0} \frac{|\nabla f(\overline{\boldsymbol{x}}) - \nabla f(\boldsymbol{x}_0)|_{2,\mathrm{d}}}{|\boldsymbol{h}|_{2,\mathrm{d}}} &\leq \lim_{|\boldsymbol{h}|_{2,\mathrm{d}} \to 0} \frac{|\nabla f(\overline{\boldsymbol{x}}) - \nabla f(\boldsymbol{x}_0)|_{2,\mathrm{d}} |\boldsymbol{h}|_{2,\mathrm{d}}}{|\boldsymbol{h}|_{2,\mathrm{d}}} \\ &= \lim_{|\boldsymbol{h}|_{2,\mathrm{d}} \to 0} |\nabla f(\overline{\boldsymbol{x}}) - \nabla f(\boldsymbol{x}_0)|_{2,\mathrm{d}} = 0 \end{aligned}$$

を得る．したがって

$$f(\boldsymbol{x}_0 + \boldsymbol{h}) - f(\boldsymbol{x}_0) - \langle \nabla f(\boldsymbol{x}_0), \boldsymbol{h} \rangle_{2,\mathrm{d}} = o(\boldsymbol{h})$$

となり, f は点 $\boldsymbol{x}_0$ で Fréchet 微分可能であることが示せた． □

D.12　勾配降下法

D.12.1　降下方向

定義 D.75. $\boldsymbol{x} \in M$ は $\nabla f(\boldsymbol{x}) \neq \boldsymbol{0}_n$ とする．したがって点 $\boldsymbol{x}$ は f の**停留点** (critical point) ではない．f の**降下方向** (descent direction) は $\boldsymbol{d} \in \mathbb{R}^d \, (\neq \boldsymbol{0})$ で, ある $\bar{t} > 0$ が存在して

$$f(\boldsymbol{x} + t\boldsymbol{d}) < f(\boldsymbol{x}) \qquad (0 < \forall t < \bar{t})$$

をみたすときをいう．したがって十分小さなステップ幅 $t > 0$ に対して半直線 $\mathbb{R}_{++} := \{\boldsymbol{x} + t\boldsymbol{d};\, t > 0\}$ 上で f は厳密に減少している．

補題 D.76. $\boldsymbol{x} \in U$ は f の停留点ではないとし, $\boldsymbol{d} \in \mathbb{R}^d$ は $\boldsymbol{d} \neq \boldsymbol{0}_n$ とする．$\left\langle \nabla f(\boldsymbol{x}), \boldsymbol{d} \right\rangle_{2,\mathrm{d}} < 0$ のとき $\boldsymbol{d}$ は点 $\boldsymbol{x}$ における f の降下方向となる．

逆に $\boldsymbol{d}$ が点 $\boldsymbol{x}$ における降下方向としたとき $\left\langle \nabla f(\boldsymbol{x}), \boldsymbol{d} \right\rangle_{2,\mathrm{d}} \leq 0$ となる．

Proof. f は Gâteaux 微分可能なので

$$f(\boldsymbol{x}+t\boldsymbol{d}) = f(\boldsymbol{x}) + t\bigl\langle \nabla f(\boldsymbol{x}),\, \boldsymbol{d}\bigr\rangle_{2,\mathrm{d}} + o(t) \tag{D.117}$$

となる. $\boldsymbol{d}$ は

$$\bigl\langle \nabla f(\boldsymbol{x}),\, \boldsymbol{d}\bigr\rangle_{2,\mathrm{d}} < 9$$

をみたすので十分小さな $t>0$ に対して

$$f(\boldsymbol{x}+t\boldsymbol{d}) < f(\boldsymbol{x})$$

となる.

逆に小さな $t>0$ に対して

$$f(\boldsymbol{x}+t\boldsymbol{d}) < f(\boldsymbol{x})$$

のとき, $t \downarrow 0$ とすると

$$\bigl\langle \nabla f(\boldsymbol{x}),\, \boldsymbol{d}\bigr\rangle_{2,\mathrm{d}} \leq 0$$

となる. □

D.12.2　最急降下方向

(D.117) から十分小さな $t>0$ に対して

$$f(\boldsymbol{x}+t\boldsymbol{d}) \approx f(\boldsymbol{x}) + t\bigl\langle \nabla f(\boldsymbol{x}),\, \boldsymbol{d}\bigr\rangle_{2,\mathrm{d}}$$

と書ける. $t>0$ を固定する. $\boldsymbol{d}\in\mathbb{R}^d$ で $|\boldsymbol{d}|_{2,\mathrm{d}}=1$ に対して $\bigl\langle \nabla f(\boldsymbol{x}),\, \boldsymbol{d}\bigr\rangle_{2,\mathrm{d}}$ を最小にするものが f を最小にする方向である.

$\nabla f(\boldsymbol{x})$ と $\boldsymbol{d}$ の角度を θ とする. このとき

$$\bigl\langle \nabla f(\boldsymbol{x}),\, \boldsymbol{d}\bigr\rangle_{2,\mathrm{d}} = \bigl|\nabla f(\boldsymbol{x})\bigr|_{2,\mathrm{d}} \cdot |\boldsymbol{d}|_{2,\mathrm{d}} \cos\theta = \bigl|\nabla f(\boldsymbol{x})\bigr|_{2,\mathrm{d}} \cos\theta$$

となるので $\cos\theta=-1$ と取ると

$$\boldsymbol{d} := -\frac{\nabla f(\boldsymbol{x})}{\bigl|\nabla f(\boldsymbol{x})\bigr|_{2,\mathrm{d}}}$$

となる. このことより方向 $\boldsymbol{d}=-\nabla f(\boldsymbol{x})$ を点 $\boldsymbol{x}$ における f の**最急降下方向** (steepest descent direction) とよぶ.

D.12.3　ステップ幅の選択方法

$k = 1, 2, \ldots$ とする. $\boldsymbol{x}_k$ における降下方向 $\boldsymbol{d}_k$ が選択されると, ステップ幅 t_k を適切に選択して, 次の点 $\boldsymbol{x}_{k+1}$ を

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + t\boldsymbol{d}_k \qquad (t_k > 0)$$

と定める. この方法を**直接降下法**と呼ぶ. t_k を適切に探索しないと $\boldsymbol{x}_k$ が停留点に収束する保証が得られないことがある.

代表的なステップ幅の選択方法である Armijo 法, Goldstein 法, Wolfe 法を説明する.

Armijo 法

$s > 0$, $0 < \beta < 1$, $0 < \sigma < 1$ を固定する. $i = 0, 1, 2, \ldots$ に対して

$$f(\boldsymbol{x}_k) - f(\boldsymbol{x}_k + \beta^i s\boldsymbol{d}_k) \geq -\sigma(\beta^i s)\big\langle \nabla f(\boldsymbol{x}_k), \boldsymbol{d}_k \big\rangle_{2,\mathrm{d}} \tag{D.118}$$

をテストする.

$i = 0$ とすると

$$f(\boldsymbol{x}_k) - f(\boldsymbol{x}_k + s\boldsymbol{d}_k) \geq -\sigma\big\langle \nabla f(\boldsymbol{x}), s\boldsymbol{d}_k \big\rangle_{2,\mathrm{d}} \tag{D.119}$$

をテストすることになる. 不等式 (D.119) が成り立てば

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + s\boldsymbol{d}_k$$

とする. 逆に, (D.119) が成立していなければ, ステップ幅が大きすぎるので, $i = 1$ の (D.118) をテストする. すなわち

$$f(\boldsymbol{x}_k) - f(\boldsymbol{x}_k + \beta s\boldsymbol{d}_k) \geq -\beta\sigma\big\langle \nabla f(\boldsymbol{x}_k), s\boldsymbol{d}_k \big\rangle_{2,\mathrm{d}}$$

をテストする.

以上の操作を繰り返して, i をおおくしていくとある i で (D.118) が必ず成立する.

このことを背理法を用いて証明する. すなわち, (D.118) をみたす $i \in \mathbb{N}$ が存在しないと仮定して, 矛盾を導く. i が十分おおきいとき, $t := \beta^i s > 0$ は十分小さい. したがって (D.118) において

$$f(\boldsymbol{x}_k + t\boldsymbol{d}_k) < f(\boldsymbol{x}_k) + \sigma t\big\langle \nabla f(\boldsymbol{x}_k), \boldsymbol{d}_k \big\rangle_{2,\mathrm{d}}$$

であったとする. ここで, 上式と

$$f(\boldsymbol{x}_k + t\boldsymbol{d}_k) = f(\boldsymbol{x}_k) + t\big\langle \nabla f(\boldsymbol{x}_k), \boldsymbol{d}_k \big\rangle_{2,\mathrm{d}} + o(t)$$

を比較すると

$$(\sigma-1)\big\langle \nabla f(\boldsymbol{x}_k),\, \boldsymbol{d}_k \big\rangle_{2,\,\mathrm{d}} + \frac{o(t)}{t} < 0$$

となる. $t \to 0$ とすると

$$(\sigma-1)\big\langle \nabla f(\boldsymbol{x}_k),\, \boldsymbol{d}_k \big\rangle_{2,\,\mathrm{d}} \le 0$$

となり, $\sigma - 1 < 0$ なので

$$\big\langle \nabla f(\boldsymbol{x}_k),\, \boldsymbol{d}_k \big\rangle_{2,\,\mathrm{d}} > 0$$

ととなる. これは $\boldsymbol{d}_k$ が降下方向であることと矛盾することから (D.118) をみたす i が存在することがわかる.

$\mathbb{R}^d$ の点列 $\{\boldsymbol{x}_k\}_{k=1}^{\infty}$ は最急降下法により求められる点列とし, $\boldsymbol{d}_k$ を $\boldsymbol{x}_k$ における f の降下方向とする. $-\nabla f(\boldsymbol{x}_k)$ と $\boldsymbol{d}_k$ の角度を θ_k と書く. するとある $\epsilon > 0$ が存在して

$$\cos\theta_k := \frac{\big\langle -\nabla f(\boldsymbol{x}_k),\, \boldsymbol{d}_k \big\rangle_{2,\,\mathrm{d}}}{\big|\nabla f(\boldsymbol{x}_k)\big|_{2,\,\mathrm{d}}} \in (\epsilon,\, 1] \qquad |\boldsymbol{d}_k|_{2,\,\mathrm{d}} = 1 \tag{D.120}$$

となる.

Goldstein 法

$c \in (0,\, 1/2)$ を固定する. 降下方向 $\boldsymbol{d}_k\,(k = 1,\, 2,\, \ldots)$ に対して, ステップ幅 t_k を t に関する不等式

$$\begin{aligned} f(\boldsymbol{x}_k) + (1-c)t\big\langle \nabla f(\boldsymbol{x}_k),\, \boldsymbol{d}_k \big\rangle_{2,\,\mathrm{d}} &\le f(\boldsymbol{x}_k + t\boldsymbol{d}_k) \\ &\le f(\boldsymbol{x}_k) + ct\big\langle \nabla f(\boldsymbol{x}_k),\, \boldsymbol{d}_k \big\rangle_{2,\,\mathrm{d}} \end{aligned}$$

をみたすように取る. そして

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + t_k\boldsymbol{d}_k$$

と定める.

Wolfe 法

ふたつの定数 $0 < c_1 < c_2 < 1$ を固定する. 降下方向 $\boldsymbol{d}_k\,(k = 1,\, 2,\, \ldots)$ に対して, ステップ幅 t_k を t に関する不等式

$$\begin{aligned} f(\boldsymbol{x}_k + t\boldsymbol{d}_k) &\le f(\boldsymbol{x}_k) + c_1 t\big\langle \nabla f(\boldsymbol{x}_k),\, \boldsymbol{d}_k \big\rangle_{2,\,\mathrm{d}}, \\ \big\langle \nabla f(\boldsymbol{x}_k + t\boldsymbol{d}_k,\, \boldsymbol{d}_k \big\rangle_{2,\,\mathrm{d}} &\ge c_2\big\langle \nabla f(\boldsymbol{x}_k),\, \boldsymbol{d}_k \big\rangle_{2,\,\mathrm{d}} \end{aligned}$$

をみたすように取る. そして

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + t_k\boldsymbol{d}_k$$

と定める.

D.13　降下法アルゴリズムの収束

Armijo 法でステップ幅を定めた最急降下法の収束を保証する定理を述べる．Goldstein 法と Wolfe 法による降下法の収束は Güller (2010, pp. 367-368) を参照のこと．

定理 D.77. $\{\boldsymbol{x}_k\}_{k=1}^{\infty}$ を降下法によって生成された点列で

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$$

とする．ただし $\boldsymbol{d}_k$ は (D.118) をみたし, α_k はパラメータ s, β, σ に対する Armijo 条件をみたしているとする．$\lim_{k\to\infty} \boldsymbol{x}_k = \boldsymbol{x}^*$ としたとき $\boldsymbol{x}^*$ は f の停留点になる．すなわち

$$\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}_d$$

をみたす．

Proof. 背理法で証明する．すなわち, $\nabla f(\boldsymbol{x}^*) \neq \boldsymbol{0}_d$ を仮定して矛盾を導く．必要であれば $\{\boldsymbol{d}_k\}_{k=1}^{\infty}$ の部分列を取ることで $\{\boldsymbol{d}_{k_i}\}_{i=1}^{\infty}$ はある点 $\boldsymbol{d}^* \in \{\boldsymbol{d} \in \mathbb{R}^d;\, |\boldsymbol{d}|_{2,\mathrm{d}} = 1\}$ に収束する．$|\boldsymbol{d}^*|_2 = 1$ である．i 段階において Armijo 条件が成立していたとする．すなわち

$$\begin{aligned} f(\boldsymbol{x}_{k_i}) - f(\boldsymbol{x}_{k_i} + \alpha_{k_i}\boldsymbol{d}_{k_i}) &\geq -\sigma\alpha_{k_i}\big\langle \nabla f(\boldsymbol{x}_{k_i}),\, \boldsymbol{d}_{k_i}\big\rangle_{2,\mathrm{d}} \\ &= \sigma\alpha_{k_i}\big|\nabla f(\boldsymbol{x}_{k_i})\big|_{2,\mathrm{d}} \cos\theta_{k_i} \\ &\geq \epsilon\sigma\alpha_{k_i}\big|\nabla f(\boldsymbol{x}_{k_i})\big|_{2,\mathrm{d}} \end{aligned} \tag{D.121}$$

となる．ただし, ϵ は (D.120) で定めたものである．α_{k_i} で Armijo 条件が成立しているので, $i-1$ ステップでは Armijo 条件は成立していないことになる．このことから

$$f(\boldsymbol{x}_{k_i}) - f\left(\boldsymbol{x}_{k_i} + \frac{\alpha_{k_i}}{\beta}\boldsymbol{d}_{k_i}\right) < -\sigma\frac{\alpha_{k_i}}{\beta}\big\langle \nabla f(\boldsymbol{x}_{k_i}),\, \boldsymbol{d}_{k_i}\big\rangle_{2,\mathrm{d}} \tag{D.122}$$

が成立していることになる．$k \to \infty$ のとき, 点列 $\{\boldsymbol{x}_k\}_{k=1}^{\infty}$ は $\boldsymbol{x}^*$ に収束するので, その部分列も $\boldsymbol{x}^*$ に収束する．よって, $\boldsymbol{x}_{k_i} \overset{i\to\infty}{\longrightarrow} \boldsymbol{x}^*$ で $\{f(\boldsymbol{x}_{k_i})\}_{i=1}^{\infty}$ は減少列なので, f の連続性から

$$f(\boldsymbol{x}_{k_i}) \searrow f(\boldsymbol{x}^*)$$

となる．(D.121) の左辺は $i \to \infty$ のとき 0 に収束する．なぜならば $f(\boldsymbol{x}_{k_i}) - f(\boldsymbol{x}_{k_{i+1}}) \overset{i\to\infty}{\longrightarrow} 0$ からわかる．さらに $\alpha_{k_i} = |\boldsymbol{x}_{k_i} - \boldsymbol{x}_{k_{i+1}}|_2 \overset{i\to\infty}{\longrightarrow} 0$ なので, (D.121) の他の項も $i \to \infty$ のときに 0 に収束する．$\boldsymbol{d}_k$ の定め方から (D.120) が成立するので,

$$\lim_{i\to\infty}\big\langle \nabla f(\boldsymbol{x}_{k_i}),\, \boldsymbol{d}_{k_i}\big\rangle_{2,\mathrm{d}} = \big\langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{d}^*\big\rangle_{2,\mathrm{d}} \quad \text{かつ} \quad \big\langle \nabla f(\boldsymbol{x}_{k_i}),\, \boldsymbol{d}_{k_i}\big\rangle_{2,\mathrm{d}} < 0$$

である. さらに, 背理法の仮定から $\nabla f(\boldsymbol{x}^*) \neq \boldsymbol{0}_d$ なので

$$\left\langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{d}^* \right\rangle_{2,\mathrm{d}} < 0, \quad \alpha_{k_i} = |\boldsymbol{x}_{k_i} - \boldsymbol{x}_{k_{i+1}}|_{2,\mathrm{d}} \to 0\,(i \to \infty) \tag{D.123}$$

がわかる. i が十分大きいとき $\boldsymbol{x}_{k_i}$ においてバックトラックが必要となることがわかる.

一方, 中間値の定理から $\boldsymbol{z}_{k_i} \in \left(\boldsymbol{x}_{k_i},\, \boldsymbol{x}_{k_i} + \frac{\alpha_{k_i}}{\beta}\boldsymbol{d}_{k_i}\right)$ が存在して

$$f(\boldsymbol{x}_{k_i}) - f\left(\boldsymbol{x}_{k_i} + \frac{\alpha_{k_i}}{\beta}\boldsymbol{d}_{k_i}\right) = -\frac{\alpha_{k_i}}{\beta}\left\langle \nabla f(\boldsymbol{z}_{k_i}),\, \boldsymbol{d}_{k_i} \right\rangle_{2,\mathrm{d}}$$

とできる. (D.122) と比較すると

$$\begin{aligned} &-\frac{\alpha_{k_i}}{\beta}\left\langle \nabla f(\boldsymbol{z}_{k_i}),\, \boldsymbol{d}_{k_i} \right\rangle_{2,\mathrm{d}} < -\sigma\frac{\alpha_{k_i}}{\beta}\left\langle \nabla f(\boldsymbol{x}_{k_i}),\, \boldsymbol{d}_{k_i} \right\rangle_{2,\mathrm{d}} \\ &\Leftrightarrow \left\langle \nabla f(\boldsymbol{z}_{k_i}),\, \boldsymbol{d}_{k_i} \right\rangle_{2,\mathrm{d}} > \sigma\left\langle \nabla f(\boldsymbol{x}_{k_i}),\, \boldsymbol{d}_{k_i} \right\rangle_{2,\mathrm{d}} \end{aligned}$$

を得る. $i \to \infty$ のとき $\lim \boldsymbol{z}_{k_i} = \boldsymbol{x}^*$ となり, $0 < \sigma < 1$ なので, 上の 2 番目の不等式より

$$\left\langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{d}^* \right\rangle_{2,\mathrm{d}} \geq 0$$

を得る. これは (D.123) と矛盾. よって $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}_d$ を示すことができたが示せた. □

D.14　章末注釈と参考文献

第E章　特異値分解

E.1　特異値分解

定義 E.1. $n, p \in \mathbb{N}, n \geq p$ とし, $V \subset \mathbb{R}^n$ をベクトル空間とする. $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_p \in V$ とする.
(1) $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_p$ はベクトル空間 V の基底であるとは, 任意の $\boldsymbol{v} \in V$ に対して, 唯一の $a_1, a_2, \ldots, a_p \in \mathbb{R}$ があって,

$$\boldsymbol{v} = a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + \cdots + a_p\boldsymbol{v}_p$$

とかけるときをいう. この p をベクトル空間 V の**次元**といい, $\dim(V)$ と書く.
(2) $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_p$ は**線型独立** (1 次独立) であるとは,

$$a_1\boldsymbol{v}_1+a_2\boldsymbol{v}_2+\cdots+a_p\boldsymbol{v}_p = \boldsymbol{0}\,(a_1, a_2, \ldots, a_p \in \mathbb{R}) \Rightarrow a_1 = a_2 = \cdots = a_p = 0$$

が成り立つときをいう. ただし, $\boldsymbol{0} = (0, 0, \ldots, 0)^\top \in \mathbb{R}^n$ である.

定義 E.2. (1) $n, p \in \mathbb{N}, n \geq p$ とする. ベクトル $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p \in \mathbb{R}^n$ で張られた $\mathbb{R}^n$ の部分空間 $\mathrm{span}\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p\}$ を

$$\mathrm{span}\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p\} = \big\{a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \cdots + a_p\boldsymbol{x}_p;\, a_1, a_2, \ldots, a_p \in \mathbb{R}\big\}$$

と定義する.
(2) $m, n \in \mathbb{N}$ とし, $\boldsymbol{A} \in \mathrm{Mat}(m, n; \mathbb{R})$ とする. $\boldsymbol{A}$ と $\boldsymbol{A}^\top$ の像と核をそれぞれ

$$\begin{aligned}\mathsf{R}\big(\boldsymbol{A}\big) &= \big\{\boldsymbol{A}\boldsymbol{x};\, \boldsymbol{x} \in \mathbb{R}^n\big\} \subset \mathbb{R}^m,\\ \ker\big(\boldsymbol{A}\big) &= \big\{\boldsymbol{x};\, \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0} \in \mathbb{R}^m\big\} \subset \mathbb{R}^n,\\ \mathsf{R}\big(\boldsymbol{A}^\top\big) &= \big\{\boldsymbol{A}^\top\boldsymbol{y};\, \boldsymbol{y} \in \mathbb{R}^m\big\} \subset \mathbb{R}^n,\\ \ker\big(\boldsymbol{A}^\top\big) &= \big\{\boldsymbol{y};\, \boldsymbol{A}^\top\boldsymbol{y} = \boldsymbol{0} \in \mathbb{R}^n\big\} \subset \mathbb{R}^m\end{aligned}$$

で定める.

定義 E.3. $n \in \mathbb{N}$ とする. 行列式は関数 $\det : \mathrm{Mat}(n; \mathbb{R}) \to \mathbb{R}$ で以下のように定義する.

(1) 1×1 の行列 (a) の行列式を

$$\det(a) = a.$$

(2) $n \in \mathbb{N}$, $n \geq 2$ とき, 行列

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \in \mathrm{Mat}(n; \mathbb{R})$$

の行列式を以下のように定める. $i, j = 1, 2, \ldots, n$ に対し $\boldsymbol{A}_{ij}$ を $\boldsymbol{A}$ から i 行と j 列をとった $(n-1) \times (n-1)$ 行列とする. $n \times n$ 行列 $\boldsymbol{A}$ の行列式を

$$\det(\boldsymbol{A}) = \sum_{k=1}^{n} (-1)^{k+1} a_{1k} \det(\boldsymbol{A}_{1k})$$

で定める.

命題 E.4. $n \times n$ の行列 $\boldsymbol{A} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n) \in \mathrm{Mat}(n; \mathbb{R})$ に対して, 行列式は以下の性質をみたす.

(1) $k = 1, 2, \ldots, n$ と $d \in \mathbb{R}$ に対し,

$$\det(\boldsymbol{a}_1, \ldots, \lambda \boldsymbol{a}_k, \ldots, \boldsymbol{a}_n) = \lambda \det(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k, \ldots, \boldsymbol{a}_n).$$

(2) $j \neq k$ に対し,

$$\det(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k + \lambda \boldsymbol{a}_j, \ldots, \boldsymbol{a}_n) = \lambda \det(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k, \ldots, \boldsymbol{a}_n).$$

(3) $\det(\boldsymbol{I}_n) = 1$.

(4) $\boldsymbol{A}, \boldsymbol{B} \in \mathrm{Mat}(n; \mathbb{R})$ に対し

$$\det(\boldsymbol{A}\boldsymbol{B}) = \det(\boldsymbol{A}) \det(\boldsymbol{B}).$$

定義 E.5. $\boldsymbol{A} \in \mathrm{Mat}(n; \mathbb{R})$ に対し, ある $\boldsymbol{B} \in \mathrm{Mat}(n; \mathbb{R})$ が存在し,

$$\boldsymbol{A}\boldsymbol{B} = \boldsymbol{B}\boldsymbol{A} = \boldsymbol{I}_n$$

が成り立つとき, $\boldsymbol{A}$ は可逆であるといい, $\boldsymbol{B}$ を $\boldsymbol{A}$ の**逆行列**といい, これを $\boldsymbol{A}^{-1}$ と記す.

命題 E.6. $\boldsymbol{A} \in \mathrm{Mat}(n; \mathbb{R})$ とする. このとき

$$\boldsymbol{A} \text{ は可逆} \Leftrightarrow \det(\boldsymbol{A}) \neq 0.$$

定理 E.7. (特異値分解) $m,\, n \in \mathbb{N}$ とする．任意の $\boldsymbol{A} \in \mathrm{Mat}(m,\, n;\, \mathbb{R})$ は次の分解をもつ．

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{\top}.$$

ただし $\boldsymbol{U} \in \mathrm{Mat}(m;\, \mathbb{R})$, $\boldsymbol{V} \in \mathrm{Mat}(n;\, \mathbb{R})$ は直交行列で, $\boldsymbol{D} \in \mathrm{Mat}(m,\, n;\, \mathbb{R})$ は対角行列で，その対角成分は $d_1 \geq d_2 \geq \cdots \geq d_{\min(m,\, n)} \geq 0$ である．$d_j\,(j = 1,\, 2,\, \ldots,\, \min(m,\, n))$ を行列 $\boldsymbol{A}$ の**特異値**といい, $\{(d_j,\, \boldsymbol{u}_j,\, \boldsymbol{v}_j);\, j = 1,\, 2,\, \ldots,\, \min(n,\, n)\}$ を**特異値系**[1]という．

注意 E.8. $n \geq m$ のとき

$$\boldsymbol{D} = \begin{pmatrix} d_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & d_m & 0 & \cdots & 0 \end{pmatrix}$$

である．$m \geq n$ のとき,

$$\boldsymbol{D} = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \\ 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

である．□

<u>**定理** E.7 **の証明**</u> $\boldsymbol{y} \in \mathbb{R}^m$ に対し,

$$|\boldsymbol{y}|_2 = \sqrt{\boldsymbol{y}^{\top}\boldsymbol{y}}$$

とする．ただし, 次元の異なる Euclid 空間のノルムも同じ記号を流用する．行列 $\boldsymbol{A}$ のノルム

$$\|\boldsymbol{A}\| := \sup_{\boldsymbol{x} \in \mathbb{R}^n} \frac{|\boldsymbol{A}\boldsymbol{x}|_2}{|\boldsymbol{x}|_2}$$

で定義すると

$$d_1 := \|\boldsymbol{A}\| := \sup_{\boldsymbol{x} \in \mathbb{R}^n,\, |\boldsymbol{x}|_2 = 1} |\boldsymbol{A}\boldsymbol{x}|_2$$

[1] $j = 1,\, 2,\, \ldots,\, \min(m,\, n)$ に対して

$$\boldsymbol{A}\boldsymbol{v}_j = d_j \boldsymbol{u}_j$$

となっている．

となる. $d_1 \neq 0$ と仮定する. そうでなければ, 定理の証明は自明となる. あるベクトル $\boldsymbol{x} \in \mathbb{R}^n$ ($|\boldsymbol{x}|_2 = 1$) が存在して

$$\left|\boldsymbol{Ax}\right|_2 = d_1$$

をみたすとする.

$$\boldsymbol{y} = \frac{1}{d_1}\boldsymbol{Ax} \in \mathbb{R}^m$$

とおけば

$$|\boldsymbol{y}|_2^2 = \boldsymbol{y}^\top \boldsymbol{y} = \frac{1}{d_1^2}\boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{Ax} = \frac{|\boldsymbol{Ax}|_2^2}{d_1^2} = 1$$

より $|\boldsymbol{y}|_2 = 1$ となる. $\boldsymbol{v}_2, \boldsymbol{v}_3, \ldots, \boldsymbol{v}_n \in \mathbb{R}^n$ と $\boldsymbol{u}_2, \boldsymbol{u}_3, \ldots, \boldsymbol{u}_m \in \mathbb{R}^m$ をうまく選んで, $\{\boldsymbol{x}, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ と $\{\boldsymbol{y}, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m\}$ はそれぞれ $\mathbb{R}^n$ と $\mathbb{R}^m$ の正規直交基底となるようにする.

$\boldsymbol{U}_1 = (\boldsymbol{y}, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m) \in \mathrm{Mat}(m; \mathbb{R})$, $\quad \boldsymbol{V}_1 = (\boldsymbol{x}, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n) \in \mathrm{Mat}(n; \mathbb{R})$,
$\boldsymbol{A}_1 = \boldsymbol{U}_1^\top \boldsymbol{A}\boldsymbol{V}_1$

とおけば,

$$\begin{aligned}
\boldsymbol{A}_1 = \boldsymbol{U}_1^\top \boldsymbol{A}\boldsymbol{V}_1 &= \begin{bmatrix} \boldsymbol{y}^\top \\ \boldsymbol{u}_2^\top \\ \vdots \\ \boldsymbol{u}_m^\top \end{bmatrix} \boldsymbol{A}[\boldsymbol{x}, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n] = \begin{bmatrix} \boldsymbol{y}^\top \\ \boldsymbol{u}_2^\top \\ \vdots \\ \boldsymbol{u}_m^\top \end{bmatrix} [\boldsymbol{Ax}, \boldsymbol{Av}_2, \ldots, \boldsymbol{Av}_n] \\
&= \begin{bmatrix} \boldsymbol{y}^\top \\ \boldsymbol{u}_2^\top \\ \vdots \\ \boldsymbol{u}_m^\top \end{bmatrix} [d_1\boldsymbol{y}, \boldsymbol{Av}_2, \ldots, \boldsymbol{Av}_n] = \begin{bmatrix} d_1\boldsymbol{y}^\top\boldsymbol{y} & \boldsymbol{y}^\top\boldsymbol{Av}_2 & \ldots & \boldsymbol{y}^\top\boldsymbol{Av}_n \\ 0 & \boldsymbol{u}_2^\top\boldsymbol{Av}_2 & \ldots & \boldsymbol{u}_2\boldsymbol{Av}_n \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \boldsymbol{u}_m\boldsymbol{Av}_2 & & \ldots\boldsymbol{u}_m\boldsymbol{Av}_n \end{bmatrix} \\
&=: \begin{bmatrix} d_1 & \boldsymbol{w}^\top \\ \boldsymbol{0} & \boldsymbol{B} \end{bmatrix}
\end{aligned}$$

と書き直せる. ただし

$$\boldsymbol{w} = \begin{bmatrix} \boldsymbol{y}^\top\boldsymbol{Av}_2 \\ \boldsymbol{y}^\top\boldsymbol{Av}_3 \\ \vdots \\ \boldsymbol{y}^\top\boldsymbol{Av}_n \end{bmatrix} \in \mathbb{R}^{n-1},$$

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{u}_2^\top\boldsymbol{Av}_2 & \boldsymbol{u}_2^\top\boldsymbol{Av}_3 & \ldots & \boldsymbol{u}_2^\top\boldsymbol{Av}_n \\ \boldsymbol{u}_3^\top\boldsymbol{Av}_2 & \boldsymbol{u}_3^\top\boldsymbol{Av}3 & \ldots & \boldsymbol{u}_3^\top\boldsymbol{Av}_n \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{u}_m^\top\boldsymbol{Av}_2 & \boldsymbol{u}_m^\top\boldsymbol{Av}_3 & \ldots & \boldsymbol{u}_m^\top\boldsymbol{Av}_n \end{bmatrix} \in \mathrm{Mat}(m-1, n-1; \mathbb{R})$$

である．上記の形に注意して計算すると

$$\boldsymbol{A}_1\begin{bmatrix} d_1 \\ \boldsymbol{w} \end{bmatrix} = \begin{bmatrix} d_1 & \boldsymbol{w}^\top \\ \boldsymbol{0} & \boldsymbol{B} \end{bmatrix}\begin{bmatrix} d_1 \\ \boldsymbol{w} \end{bmatrix} = \begin{bmatrix} d_1^2 + \boldsymbol{w}^\top\boldsymbol{w} \\ \boldsymbol{B}\boldsymbol{w} \end{bmatrix}$$

となる．したがって

$$\begin{aligned} \left|\boldsymbol{A}_1\begin{bmatrix} d_1 \\ \boldsymbol{w} \end{bmatrix}\right|_2^2 = \left|\begin{bmatrix} d_1^2 + \boldsymbol{w}^\top\boldsymbol{w} \\ \boldsymbol{B}\boldsymbol{w} \end{bmatrix}\right|_2^2 &= \{d_1^2 + \boldsymbol{w}^\top\boldsymbol{w}\}^2 + \boldsymbol{w}^\top\boldsymbol{B}^\top\boldsymbol{B}\boldsymbol{w} \\ &\geq \{d_1^2 + |\boldsymbol{w}|_2^2\}^2 \end{aligned}$$

から

$$\left|\boldsymbol{A}_1\begin{bmatrix} d_1 \\ \boldsymbol{w} \end{bmatrix}\right|_2 \geq d_1^2 + |\boldsymbol{w}|_2^2, \quad \left|\frac{1}{\sqrt{d_1^2 + |\boldsymbol{w}|_2}}\begin{bmatrix} d_1 \\ \boldsymbol{w} \end{bmatrix}\right|_2 = 1$$

がわかる．よって

$$\begin{aligned} \|\boldsymbol{A}_1\| = \sup_{\boldsymbol{x}\in\mathbb{R}^n,\, |\boldsymbol{x}|_2=1} \left|\boldsymbol{A}_1\boldsymbol{x}\right|_2 &\geq \frac{1}{\sqrt{d_1^2 + |\boldsymbol{w}|_2^2}}\left|\boldsymbol{A}_1\begin{bmatrix} d_1 \\ \boldsymbol{w} \end{bmatrix}\right|_2 \\ &\geq \sqrt{d_1^2 + |\boldsymbol{w}|_2^2}. \end{aligned} \tag{E.1}$$

直交変換に関して $\|\cdot\|$ は不変[2]なので,

$$d_1 = \|\boldsymbol{A}\| = \|\boldsymbol{A}_1\|.$$

である．これと (E.1) と合わせると

$$d_1 = \|\boldsymbol{A}\| = \|\boldsymbol{A}_1\| \geq \sqrt{d_1^2 + |\boldsymbol{w}|_2^2} \geq d_1$$

となるので,

$$|\boldsymbol{w}|_2 = 0 \Leftrightarrow \boldsymbol{w} = \boldsymbol{0}.$$

よって

$$\boldsymbol{U}_1^\top\boldsymbol{A}\boldsymbol{V}_1 = \boldsymbol{A}_1 = \begin{bmatrix} d_1 & \boldsymbol{0} \\ \boldsymbol{0}^\top & \boldsymbol{B} \end{bmatrix}$$

[2] $\forall\boldsymbol{x}\in\mathbb{R}^n$ ($|\boldsymbol{x}|_2 = 1$) に対して $|\boldsymbol{V}_1\boldsymbol{x}|_2 = 1$ に注意すれば,

$$\begin{aligned} \|\boldsymbol{A}\| = \|\boldsymbol{U}_1\boldsymbol{A}_1\boldsymbol{V}_1\| &= \sup_{\boldsymbol{x}\in\mathbb{R}^n;\, |\boldsymbol{x}|_2=1} |\boldsymbol{U}_1\boldsymbol{A}_1\boldsymbol{V}_1\boldsymbol{x}|_2 = \sup_{\boldsymbol{x}\in\mathbb{R}^n;\, |\boldsymbol{x}|_2=1} |\boldsymbol{U}_1\boldsymbol{A}_1\boldsymbol{x}|_2 \\ &= \sup_{\boldsymbol{x}\in\mathbb{R}^n;\, |\boldsymbol{x}|_2=1} \sqrt{\boldsymbol{x}^\top\boldsymbol{A}_1^\top\boldsymbol{U}_1^\top\boldsymbol{U}_1\boldsymbol{A}_1\boldsymbol{x}} = \sup_{\boldsymbol{x}\in\mathbb{R}^n;\, |\boldsymbol{x}|_2=1} \sqrt{\boldsymbol{x}^\top\boldsymbol{A}_1^\top\boldsymbol{A}_1\boldsymbol{x}} \\ &= \sup_{\boldsymbol{x}\in\mathbb{R}^n;\, |\boldsymbol{x}|_2=1} |\boldsymbol{A}_1\boldsymbol{x}|_2 = \|\boldsymbol{A}_1\| \end{aligned}$$

からわかる．

と書けることがわかった.

次に, $d_2 = \|\boldsymbol{B}\|$ とおけば,

$$\begin{aligned} d_2 &= \sup_{\boldsymbol{x}\in\mathbb{R}^{n-1};\,|\boldsymbol{x}|_2=1} |\boldsymbol{B}\boldsymbol{x}|_2 = \sup_{\boldsymbol{x}\in\mathbb{R}^{n-1};\,|\boldsymbol{x}|_2=1} \left| \begin{bmatrix} d_1 & \boldsymbol{0}^\top \\ \boldsymbol{0} & \boldsymbol{B} \end{bmatrix} \begin{bmatrix} 0 \\ \boldsymbol{x} \end{bmatrix} \right|_2 \\ &\le \sup_{\boldsymbol{z}\in\mathbb{R}^{n};\,|\boldsymbol{z}|_2=1} |\boldsymbol{A}_1\boldsymbol{z}|_2 = \|\boldsymbol{A}_1\| = \|\boldsymbol{A}\| = d_1. \end{aligned}$$

$d_2 = 0$ のとき, $\boldsymbol{B} = \boldsymbol{0}$ より証明はおわり. $d_2 > 0$ と仮定して議論を進める. 前と同じようにすれば,

$$d_2 = \|\boldsymbol{B}\|$$

とおき,

$$|\boldsymbol{B}\boldsymbol{x}|_2 = 1,\ \boldsymbol{x} \in \mathbb{R}^{n-1}\ \text{s.t.} |\boldsymbol{x}|_2 = 1,\ \boldsymbol{y} = \frac{1}{d_2}\boldsymbol{B}\boldsymbol{x} \in \mathbb{R}^{m-1}$$

と取り, さきほどの議論を繰り返すと $\widetilde{\boldsymbol{U}}_2 \in \mathrm{Mat}(m-1;\,\mathbb{R})$, $\widetilde{\boldsymbol{V}}_2 \in \mathrm{Mat}(n-1;\,\mathbb{R})$ で

$$\widetilde{\boldsymbol{U}}_2^\top \boldsymbol{B}\widetilde{\boldsymbol{V}}_2 = \begin{bmatrix} d_2 & \boldsymbol{0}^\top \\ \boldsymbol{0} & \boldsymbol{C} \end{bmatrix},\ \boldsymbol{C} \in \mathrm{Mat}(m-2,\, n-2;\,\mathbb{R})$$

とできて,

$$\boldsymbol{U}_2 = \begin{bmatrix} 1 & \boldsymbol{0}^\top \\ \boldsymbol{0} & \widetilde{\boldsymbol{U}}_2 \end{bmatrix},\ \boldsymbol{V}_2 = \begin{bmatrix} 1 & \boldsymbol{0}^\top \\ \boldsymbol{0} & \widetilde{\boldsymbol{V}}_2 \end{bmatrix}$$

とおけば, $\boldsymbol{U}_2$, $\boldsymbol{V}_2$ は直交行列で

$$\boldsymbol{U}_2^\top \boldsymbol{U}_1^\top \boldsymbol{A}\boldsymbol{V}_1\boldsymbol{V}_2 = \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & \boldsymbol{C} \end{bmatrix}$$

となる. さらに $\boldsymbol{U}_1\boldsymbol{U}_2 \in \mathrm{Mat}(m;\,\mathbb{R})$ と $\boldsymbol{V}_1\boldsymbol{V}_2 \in \mathrm{Mat}(n;\,\mathbb{R})$ も直交行列であることに注意する.

以上の操作を繰り返せば, 定理は証明される. □

定理 E.9. $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top$ と特異値分解されたとする. ただし, $p \le \min(m,\, n)$ に対し

$$d_1 \ge d_2 \ge \cdots \ge d_p > d_{p+1} = d_{p+2} = \cdots = d_{\min(m,\,n)} = 0$$

とする.

$$\boldsymbol{U} = (\boldsymbol{u}_1,\, \boldsymbol{u}_2,\, \ldots,\, \boldsymbol{u}_m),\ \boldsymbol{V} = (\boldsymbol{v}_1,\, \boldsymbol{v}_2,\, \ldots,\, \boldsymbol{v}_n)$$

としたとき,

$$\ker(\boldsymbol{A}) = \mathrm{span}\{\boldsymbol{v}_{p+1}, \boldsymbol{v}_{p+2}, \ldots, \boldsymbol{v}_n\} = \mathsf{R}(\boldsymbol{A})^{\perp},$$
$$\ker(\boldsymbol{A}^\top) = \mathrm{span}\big\{\boldsymbol{u}_{p+1}, \boldsymbol{u}_{p+2}, \ldots, \boldsymbol{u}_m\big\} = \mathsf{R}\big(\boldsymbol{A}^\top\big)^{\perp}.$$

Proof. $\boldsymbol{U}$ の直交性より

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}\,(\boldsymbol{x} \in \mathbb{R}^n) \Leftrightarrow \boldsymbol{D}\boldsymbol{V}^\top\boldsymbol{x} = \begin{bmatrix} d_1\boldsymbol{v}_1^\top\boldsymbol{x} \\ d_2\boldsymbol{v}_2^\top\boldsymbol{x} \\ \vdots \\ d_p\boldsymbol{v}_p^\top\boldsymbol{x} \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \boldsymbol{0}.$$

よって $\ker(\boldsymbol{A})$ の任意の元 $\boldsymbol{x}$ は $\boldsymbol{v}_1, \boldsymbol{v}, \ldots, \boldsymbol{v}_p$ と直交するので,

$$\ker(\boldsymbol{A}) = \mathrm{span}\{\boldsymbol{v}_{p+1}, \boldsymbol{v}_{p+2}, \ldots, \boldsymbol{v}_n\}.$$

さらに $\boldsymbol{A}^\top = \boldsymbol{V}\boldsymbol{D}^\top\boldsymbol{U}^\top$ より

$$\ker\big(\boldsymbol{A}^\top\big) = \mathrm{span}\{\boldsymbol{u}_{p+1}, \boldsymbol{u}_{p+2}, \ldots, \boldsymbol{u}_m\} = \mathsf{R}\big(\boldsymbol{A}\big)^{\perp}.$$

一方, $\forall\widetilde{\boldsymbol{x}} \in \mathsf{R}\big(\boldsymbol{A}^\top\big)$ とすれば, ある $\boldsymbol{y} \in \mathbb{R}^m$ があって, $\widetilde{\boldsymbol{x}} = \boldsymbol{A}^\top\boldsymbol{y}$ とかける. このことより,

$$\boldsymbol{x} \in \ker\big(\boldsymbol{A}\big) \Leftrightarrow \boldsymbol{x}^\top\widetilde{\boldsymbol{x}} = \big(\boldsymbol{A}\boldsymbol{x}\big)^\top\boldsymbol{y} = 0$$

より, $\boldsymbol{x}$ は $\mathsf{R}\big(\boldsymbol{A}^\top\big)$ と直交する. 同様に, $\forall\widetilde{\boldsymbol{y}} \in \mathsf{R}\big(\boldsymbol{A}\big)$ をとる. ある $\boldsymbol{x} \in \mathbb{R}^n$ があって, $\widetilde{\boldsymbol{y}} = \boldsymbol{A}\boldsymbol{x}$ とかける. このことより

$$\boldsymbol{y} \in \ker\big(\boldsymbol{A}^\top\big) \Leftrightarrow \boldsymbol{y}^\top\widetilde{\boldsymbol{y}} = \big(\boldsymbol{A}^\top\boldsymbol{y}\big)^\top\boldsymbol{x} = 0$$

より, $\boldsymbol{y}$ は $\mathsf{R}\big(\boldsymbol{A}\big)$ と直交することもわかる. □

注意 E.10.

$$\boldsymbol{U}_1 = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_p], \boldsymbol{U}_2 = [\boldsymbol{u}_{p+1}, \boldsymbol{u}_{p+2}, \ldots, \boldsymbol{u}_m],$$
$$\boldsymbol{V}_1 = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_p], \boldsymbol{V}_2 = [\boldsymbol{v}_{p+1}, \boldsymbol{v}_{p+2}, \ldots, \boldsymbol{v}_n]$$

とおく. すなわち

$$\boldsymbol{U} = [\boldsymbol{U}_1, \boldsymbol{U}_2], \quad \boldsymbol{V} = [\boldsymbol{V}_1, \boldsymbol{V}_2]$$

である. このとき, $\mathbb{R}^n$ から $\ker\big(\boldsymbol{A}\big)$ への直交射影を $\boldsymbol{P}$, $\mathbb{R}^m$ から $\ker\big(\boldsymbol{A}^\top\big)$ への直交射影を $\boldsymbol{Q}$ とおくと

$$\boldsymbol{P} = \boldsymbol{V}_2\boldsymbol{V}_2^\top, \quad \boldsymbol{Q} = \boldsymbol{U}_2\boldsymbol{U}_2^\top$$

となる. □

E.2　Moore-Penrose の一般化逆行列

$\boldsymbol{A} \in \mathrm{Mat}(m, n; \mathbb{R})$ で $\boldsymbol{A} \neq \boldsymbol{0}$ とする．定理 E.7 から $1 \leq p \leq \min(m, n)$ があって

$$\begin{aligned}
&\boldsymbol{A} = \widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{D}}\widetilde{\boldsymbol{V}}^\top, \\
&\widetilde{\boldsymbol{U}} \in \mathrm{Mat}(m; \mathbb{R}),\ \widetilde{\boldsymbol{V}} \in \mathrm{Mat}(n; \mathbb{R}) \text{ は直交行列}, \\
&\widetilde{\boldsymbol{D}} \in \mathrm{Mat}(m, n; \mathbb{R}) \text{ は対角行列 (非対角成分は 0) で} \\
&\text{その正の対角成分は } d_1 \geq d_2 \geq \cdots \geq d_p > 0
\end{aligned}$$

と書ける．ここで $\widetilde{\boldsymbol{U}}, \widetilde{\boldsymbol{D}}, \widetilde{\boldsymbol{V}}$ を分解し

$$\begin{aligned}
&\widetilde{\boldsymbol{U}} = [\boldsymbol{U}, \boldsymbol{U}^\perp],\ \boldsymbol{U} \in \mathrm{Mat}(m, p; \mathbb{R}), \\
&\widetilde{\boldsymbol{V}} = [\boldsymbol{V}, \boldsymbol{V}^\perp],\ \boldsymbol{V} \in \mathrm{Mat}(n, p; \mathbb{R}), \\
&\boldsymbol{D} = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_p \end{bmatrix} \in \mathrm{Mat}(p; \mathbb{R})
\end{aligned}$$

とする．すると

$$\begin{aligned}
\widetilde{\boldsymbol{U}}^\top\widetilde{\boldsymbol{U}} &= \begin{bmatrix} \boldsymbol{U}^\top\boldsymbol{U} & \boldsymbol{U}^\top\boldsymbol{U}^\perp \\ (\boldsymbol{U}^\perp)^\top\boldsymbol{U} & (\boldsymbol{U}^\perp)^\top\boldsymbol{U}^\perp \end{bmatrix} = \boldsymbol{I}_m, \\
\widetilde{\boldsymbol{V}}^\top\widetilde{\boldsymbol{V}} &= \begin{bmatrix} \boldsymbol{V}^\top\boldsymbol{V} & \boldsymbol{V}^\top\boldsymbol{V}^\perp \\ (\boldsymbol{V}^\perp)^\top\boldsymbol{V} & (\boldsymbol{V}^\perp)^\top\boldsymbol{V}^\perp \end{bmatrix} = \boldsymbol{I}_n
\end{aligned}$$

から

$$\boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}_p,\ \boldsymbol{U}^\top\boldsymbol{U}^\perp = \boldsymbol{0},\ \boldsymbol{V}^\top\boldsymbol{V} = \boldsymbol{I}_p,\ \boldsymbol{V}^\top\boldsymbol{V}^\perp = \boldsymbol{0}$$

となる．よって

$$\boldsymbol{A} = \widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{D}}\widetilde{\boldsymbol{V}}^\top = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top \tag{E.2}$$

となる．ただし

$$\boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{V}^\top\boldsymbol{V} = \boldsymbol{I}_p,\ \boldsymbol{D} = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_p \end{bmatrix}$$

となる．

(E.2) を踏まえて, $\boldsymbol{A}^{\dagger} \in \mathrm{Mat}(n,\, m;\, \mathbb{R})$ を

$$\boldsymbol{A}^{\dagger} := \boldsymbol{V}\boldsymbol{D}^{-1}\boldsymbol{U}^{\top},\ \boldsymbol{D}^{-1} = \begin{bmatrix} d_1^{-1} & 0 & \cdots & 0 \\ 0 & d_2^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_p^{-1} \end{bmatrix} \tag{E.3}$$

と定める. $\boldsymbol{A}^{\dagger}$ を $\boldsymbol{A}$ の Moore-Penrose **の一般逆行列**と呼ぶことにする.
　すると $\boldsymbol{A}^{\dagger}$ は下記の関係式をみたす $n \times m$ の行列となる.

$$\boldsymbol{A}\boldsymbol{A}^{\dagger}\boldsymbol{A} = \boldsymbol{A}, \tag{E.4}$$

$$\boldsymbol{A}^{\dagger}\boldsymbol{A}\boldsymbol{A}^{\dagger} = \boldsymbol{A}^{\dagger}, \tag{E.5}$$

$$\boldsymbol{A}\boldsymbol{A}^{\dagger} = \left(\boldsymbol{A}\boldsymbol{A}^{\dagger}\right)^{\top}, \tag{E.6}$$

$$\boldsymbol{A}^{\dagger}\boldsymbol{A} = \left(\boldsymbol{A}^{\dagger}\boldsymbol{A}\right)^{\top} \tag{E.7}$$

が成立する. 上記 (E.4) – (E.7) は (E.2) と (E.3) よりわかる. 実際

$$\begin{aligned} \boldsymbol{A}\boldsymbol{A}^{\dagger}\boldsymbol{A} &= \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{\top}\boldsymbol{V}\boldsymbol{D}^{-1}\boldsymbol{U}^{\top}\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{\top} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{\top} = \boldsymbol{A}, \\ \boldsymbol{A}^{\dagger}\boldsymbol{A}\boldsymbol{A}^{\dagger} &= \boldsymbol{V}\boldsymbol{D}^{-1}\boldsymbol{U}^{\top}\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{\top}\boldsymbol{V}\boldsymbol{D}^{-1}\boldsymbol{U}^{\top} = \boldsymbol{V}\boldsymbol{D}^{-1}\boldsymbol{U}^{\top} = \boldsymbol{A}^{\dagger}, \\ \boldsymbol{A}\boldsymbol{A}^{\dagger} &= \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{\dagger}\boldsymbol{V}\boldsymbol{D}^{-1}\boldsymbol{U}^{\dagger} = \boldsymbol{U}\boldsymbol{U}^{\top} = \left(\boldsymbol{U}\boldsymbol{U}^{\top}\right)^{\top} = \left(\boldsymbol{A}\boldsymbol{A}^{\dagger}\right)^{\top}, \\ \boldsymbol{A}^{\dagger}\boldsymbol{A} &= \boldsymbol{V}\boldsymbol{D}^{-1}\boldsymbol{U}^{\top}\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{\top} = \boldsymbol{V}\boldsymbol{V}^{\top} = \left(\boldsymbol{V}\boldsymbol{V}^{\top}\right)^{\top} = \left(\boldsymbol{A}^{\dagger}\boldsymbol{A}\right)^{\top} \end{aligned}$$

からわかる. さらに

$$\begin{aligned} \boldsymbol{P} = \boldsymbol{A}\boldsymbol{A}^{\dagger} &: \mathbb{R}^m \ \to\ \mathrm{ran}(\boldsymbol{A} \subset \mathbb{R}^m, \\ \boldsymbol{Q} = \boldsymbol{A}^{\dagger}\boldsymbol{A} &: \mathbb{R}^n \ \to\ \mathrm{ran}\left(\boldsymbol{A}^{\top}\right) \subset \mathbb{R}^n \end{aligned}$$

は射影行列になっている. すなわち $\boldsymbol{P} \in \mathrm{Mat}(m;\, \mathbb{R})$ と $\boldsymbol{Q} \in \mathrm{Mat}(n;\, \mathbb{R})$

$$\boldsymbol{P}^2 = \boldsymbol{P},\ \boldsymbol{P}^{\top} = \boldsymbol{P},\ \boldsymbol{Q}^2 = \boldsymbol{Q},\ \boldsymbol{Q}^{\top} = \boldsymbol{Q}$$

をみたしている.
　最後に $\boldsymbol{A}^{\dagger}$ の一意性を示す. $\boldsymbol{B}^{\dagger}$ も $\boldsymbol{A}$ の Moore-Penrose の一般逆行列とする. すると $\boldsymbol{A}^{\dagger}$ を $\boldsymbol{B}$ と替えた

$$\boldsymbol{A}\boldsymbol{B}\boldsymbol{A} = \boldsymbol{A}, \tag{E.8}$$

$$\boldsymbol{B}\boldsymbol{A}\boldsymbol{B} = \boldsymbol{B}, \tag{E.9}$$

$$\boldsymbol{A}\boldsymbol{B} = \left(\boldsymbol{A}\boldsymbol{B}\right)^{\top}, \tag{E.10}$$

$$\boldsymbol{B}\boldsymbol{A} = \left(\boldsymbol{B}\boldsymbol{A}\right)^{\top} \tag{E.11}$$

も成立する．すると

$$
\begin{aligned}
\boldsymbol{A}^{\dagger} &= \boldsymbol{A}^{\dagger}\boldsymbol{A}\boldsymbol{A}^{\dagger} \qquad (\because \text{(E.4)}) \\
&= \boldsymbol{A}^{\dagger}\bigl(\boldsymbol{A}\boldsymbol{B}\boldsymbol{A}\bigr)\boldsymbol{A}^{\dagger} \qquad (\because \text{(E.8)}) \\
&= \boldsymbol{A}^{\dagger}\bigl(\boldsymbol{A}\boldsymbol{B}\bigr)^{\top}\bigl(\boldsymbol{A}\boldsymbol{A}^{\dagger}\bigr)^{\top} \qquad (\because \text{(E.6) と (E.10)}) \\
&= \boldsymbol{A}^{\dagger}\boldsymbol{B}^{\top}\bigl(\boldsymbol{A}\boldsymbol{A}^{\dagger}\boldsymbol{A}\bigr)^{\top} \\
&= \boldsymbol{A}^{\dagger}\boldsymbol{B}^{\top}\boldsymbol{A}^{\top} \qquad (\because \text{(E.4)}) \\
&= \boldsymbol{A}^{\dagger}\bigl(\boldsymbol{A}\boldsymbol{B}\bigr)^{\top} \\
&= \boldsymbol{A}^{\dagger}\boldsymbol{A}\boldsymbol{B} \qquad (\because \text{(E.10)}) \\
&= \boldsymbol{A}^{\dagger}\boldsymbol{A}\boldsymbol{B}\boldsymbol{A}\boldsymbol{B} \qquad (\because \text{(E.8)} \\
&= \bigl(\boldsymbol{A}^{\dagger}\boldsymbol{A}\bigr)^{\top}\bigl(\boldsymbol{B}\boldsymbol{A}\bigr)^{\top}\boldsymbol{B} \qquad (\because \text{(E.7 と (E.11)}) \\
&= \bigl(\boldsymbol{A}\boldsymbol{A}^{\dagger}\boldsymbol{A}\bigr)^{\top}\boldsymbol{B}^{\top}\boldsymbol{B} \\
&= \boldsymbol{A}^{\top}\boldsymbol{B}^{\top}\boldsymbol{B} \qquad (\because \text{(E.4)}) \\
&= \bigl(\boldsymbol{B}\boldsymbol{A}\bigr)^{\top}\boldsymbol{B} \\
&= \boldsymbol{B}\boldsymbol{A}\boldsymbol{B} \qquad (\because \text{(E.11)}) \\
&= \boldsymbol{B} \qquad (\because \text{(E.8)})
\end{aligned}
$$

からわかる．よって一意性を証明できた．

注意 E.11. 多くの本では，$\boldsymbol{A} \in \mathrm{Mat}(m, n; \mathbb{R})$ に対して (E.4) – (E.7) をみたす $n \times m$ の行列 $\boldsymbol{A}^{\dagger}$ を $\boldsymbol{A}$ の Moore-Penrose の一般逆行列と定義している． □

E.3　章末注釈と参考文献

第F章　特性関数と分布の弱収束

F.1　特性関数の定義と性質

定義 F.1. (1) $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上の分布 P に対して

$$\begin{aligned}\varphi(t) &= \int_{-\infty}^{\infty} e^{\sqrt{-1}tx}\, d\mathsf{P}(x)\\ &= \int_{-\infty}^{\infty} \cos(tx)\, d\mathsf{P}(x) + \sqrt{-1}\int_{-\infty}^{\infty} \sin(tx)\,\mathrm{d}\mathsf{P}(x)\\ &\qquad (-\infty < t < \infty)\end{aligned}$$

を P の**特性関数** (characteristic function) という.
(2) また $X \sim \mathsf{P}$ のとき, X の特性関数 φ_X を

$$\varphi_X(t) = \mathsf{E}\big[e^{\sqrt{-1}tX}\big] \qquad (-\infty < t < \infty)$$

と書く.

注意 F.2. (1) 「確率変数の特性関数」と「分布の特性関数」は実質的に同じであるので, 主として確率変数の特性関数の場合のみを書く.
(2) また $\left|e^{\sqrt{-1}tx}\right| \leq 1$ なので, 特性関数は常に存在する.
(3) この積分の計算の仕方は, X の p.m..f. または p.d.f. を $\mathsf{p}(x)$ と書く. すると離散型の場合は

$$\varphi_X(t) = \sum_x e^{\sqrt{-1}tx}\mathsf{p}(x) \qquad (-\infty < t < \infty)$$

となり, 連続の場合は

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{\sqrt{-1}tx} p(x)\,\mathrm{d}x \qquad (-\infty < t < \infty)$$

となる.

定理 F.3. 確率変数 X の特性関数 φ_X は次の性質をみたす.
(1) $\left|\varphi_X(t)\right| \leq 1$, $\varphi_X(0) = 1$.
(2) $\varphi_X(-t) = \overline{\varphi_X(t)}$.
(3) $\varphi_X(t)$ は t について連続.

Proof. (1), (2) は定義と複素数の性質からわかる. (3) のみ証明を書いておく. $\epsilon \neq 0$ に対して

$$\left|e^{\sqrt{-1}(t+\epsilon)x} - e^{\sqrt{-1}tx}\right| = \left|e^{\sqrt{-1}tx}\right| \cdot \left|e^{\sqrt{-1}\epsilon x} - 1\right| \leq 2$$

となるので, Lebeague の優収束定理より

$$\begin{aligned}\lim_{\epsilon\to 0}\bigl\{\varphi_X(t+\epsilon) - \varphi_X(t)\bigr\} &= \lim_{\epsilon\to 0}\left\{\int_{-\infty}^{\infty} e^{\sqrt{-1}tx}\bigl\{e^{\sqrt{-1}\epsilon} - 1\bigr\}\,\mathrm{dP}(x)\right\} \\ &= \int_{-\infty}^{\infty} e^{\sqrt{-1}tx}\lim_{\epsilon\to 0}\bigl\{e^{\sqrt{-1}\epsilon} - 1\bigr\}\,\mathrm{dP}(x)\Bigr\} \\ &= 0\end{aligned}$$

となるので, $\varphi_X(t)$ の t における連続性は示せた. □

定理 F.4. $a\,(\neq 0)$, b は定数とし, 確率変数 X と $aX+b$ の特性関数をそれぞれ $\varphi_X(t)$, $\varphi_{aX+b}(t)$ とおくと次の関係が成り立つ.

$$\varphi_{aX+b}(t) = e^{\sqrt{-1}tb}\varphi_X(at).$$

Proof. 特性関数の定義からわかる. □

注意 F.5. (1) $X \sim \mathsf{Bino}(n, p)\,(n \in \mathbb{N}, 0 \leq p \leq 1)$ のとき

$$\varphi_X(t) = \bigl(e\ \sqrt{-1}tp + (1-p)\bigr)^n \qquad (-\infty < t < \infty).$$

(2) $X \sim \mathsf{Po}(\lambda)\,(\lambda > 0)$ のとき

$$\varphi_X(t) = \exp\bigl\{\lambda(e^{\sqrt{-1}t} - 1)\bigr\} \qquad (-\infty < t < \infty).$$

(3) $X \sim \mathsf{Exp}(n, p)\,(n \in \mathbb{N}, 0 \leq p \leq 1)$ のとき

$$\varphi_X(t) = \frac{\lambda}{\lambda - \sqrt{-1}t} \qquad (-\infty < t < \infty).$$

(4) $X \sim \mathsf{N}(\mu, \sigma^2)\,(\mu \in \mathbb{R}, 0 < \sigma < \infty)$ のとき

$$\varphi_X(t) = \exp\left\{\sqrt{-1}\mu t - \frac{1}{2}\sigma^2 t^2\right\} \qquad (-\infty < t < \infty).$$

□

定理 F.6. 確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率変数 X の特性関数を φ_X とする. $\mathsf{Pr}(X = a) = \mathsf{Pr}(X = b) = 0$ なる $a, b\,(a < b)$ について次が成りたつ.

$$\mathsf{Pr}\bigl(a < X \leq b\bigr) = \mathsf{P}((a, b]) = \frac{1}{2\pi}\lim_{T\to\infty}\int_{-T}^{T}\frac{e^{-\sqrt{-1}at} - e^{-\sqrt{-1}tb}}{\sqrt{-1}t}\varphi_X(t)\,\mathrm{d}t.$$

Proof. □

系 F.7. 分布はその特性関数から一意的に定まる.

Proof. 定理 *F*.6 からわかる. □

定理 F.8. (1) $\mathsf{E}\big[|X|\big] < \infty$ のとき $\varphi_X(t)$ は連続微分可能であり

$$\dot{\varphi}_X(0) = \frac{\mathrm{d}\varphi_X}{\mathrm{d}t}(t)\bigg|_{t=0} = \sqrt{-1}\mathsf{E}\big[X\big]$$

となる.
(2) $\mathsf{E}\big[|X^2|\big] < \infty$ のとき $\varphi_X(t)$ は 2 回連続微分可能であり

$$\ddot{\varphi}_X(0) = \frac{\mathrm{d}^2\varphi_X}{\mathrm{d}t^2}(t)\bigg|_{t=0} = -\mathsf{E}\big[X^2\big]$$

となる.
(3) したがって

$$\mathsf{E}\big[X\big] = -\sqrt{-1}\varphi_X(0),\ \mathbb{V}\big[X\big] = -\ddot{\varphi}_X(0) + \big\{\dot{\varphi}_X(0)\big\}^2$$

となる.

Proof. □

F.2 分布の収束と Glivenko の定理

定義 F.9. $n \in \mathbb{N}$ とする. $\{\mathsf{P}_n\}$, P を $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上の分布列とし, F_n, F を P_n, P の分布関数とする. P_n が P に**弱収束**するとは

$$\lim_{n\to\infty} \mathsf{F}_n(x) = \mathsf{F}(x)$$

が極限 F のすべての連続点 x において成り立つことをいう. また確率変数 X_n の分布が確率変数 X の分布に弱収束するとき, X_n は X に**法則収束**するという. これらのことを $\mathsf{P}_n \rightsquigarrow \mathsf{P}$, $X_n \rightsquigarrow X$ 等と記すことにする.

注意 F.10. 弱収束の必要十分条件は

$$\mathsf{F}(x-0) \le \liminf_{n\to\infty} \mathsf{F}_n(x) \le \limsup_{n\to\infty} \mathsf{F}_n(x) \le \mathsf{F}(x+0), \qquad (x \in \mathbb{R})$$

である. 証明は笠原 (2010, *p*.135) を参照のこと. □

定理 F.11. $n \in \mathbb{N}$ とする．P_n, P は可測空間 $\big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ 上の分布列とするとき, 以下の 4 条件は同値である．
(1) $\mathsf{P}_n \rightsquigarrow \mathsf{P}$ である．
(2) $\mathsf{P}(\{a\}) = \mathsf{P}(\{b\}) = 0$ なる任意の $a < b$ に対して

$$\mathsf{P}_n\big((a, b]\big) \overset{n\to\infty}{\longrightarrow} \mathsf{P}\big((a, b]\big)$$

となる．
(3) コンパクト台をもつ任意の連続関数 $h(x)$ に対して

$$\int_{-\infty}^{\infty} h(x)\,\mathrm{d}\mathsf{P}_n(x) \overset{n\to\infty}{\longrightarrow} \int_{-\infty}^{\infty} h(x)\,\mathrm{d}\mathsf{P}(x) \tag{F.1}$$

となる．
(4) 任意の有界連続関数 $h(x)$ に対して $(F.1)$ が成り立つ．

Proof. 証明は笠原 (2010, *pp*.136 − 140) を参照のこと． □

定理 F.12. $n \in \mathbb{N}$ とする．X_n, X は確率空間 $\big(\Omega, \mathcal{A}, \mathsf{Pr}\big)$ 上で定義された確率変数列とする．このとき次の 4 条件は同値である．
(1) $X_n \rightsquigarrow X$ となる．
(2) $\mathsf{P}(\{a\}) = \mathsf{P}(\{b\}) = 0$ なる任意の $a < b$ に対して

$$\mathsf{Pr}\big(a < X_n \le b\big) \overset{n\to\infty}{\longrightarrow} \mathsf{Pr}\big(a < X \le b\big)$$

となる．
(3) コンパクト台をもつ任意の連続関数 $h(x)$ に対して

$$\mathsf{E}\big[h(X_n)\big] \overset{n\to\infty}{\longrightarrow} \mathsf{E}\big[h(X)\big] \tag{F.2}$$

となる．
(4) 任意の有界連続関数 $h(x)$ に対して (F.2) が成り立つ．

Proof. 節 E.F.3.1 で行う． □

定理 F.13. (Glivenko の定理) $n \in \mathbb{N}$ とする．測度空間 $\big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ 上の分布 P_n, P の特性関数を φ_n, φ とする．このとき次は同値である．
(1) $\mathsf{P}_n \rightsquigarrow \mathsf{P}\,(n \to \infty)$ である．
(2) $\varphi_n(t) \overset{n\to\infty}{\longrightarrow} \varphi(t)\,(-\infty < t < \infty)$ である．

Proof. 節 E.F.3.1 で行う． □

定理 F.14. 確率変数 X_n が X に法則収束するための必要十分条件は

$$\mathsf{E}\big[e^{\sqrt{-1}tX_n}\big] \overset{n\to\infty}{\longrightarrow} \mathsf{E}\big[e^{\sqrt{-1}tX}\big] \qquad (-\infty < t < \infty)$$

である．

Proof. 定理 F.13 の言いかえである. □

F.3　定理の証明

F.3.1　定理 F.11 の証明

(1) $\Longrightarrow$ (2): $\mathsf{P}_n\big((a,\, b]\big) = \mathsf{F}_n(b) - \mathsf{F}_n(a)$ からわかる.
(2) $\Longrightarrow$ (3): 十分おおきな $M > 0$ をとり, 区間 $[-M,\, M]$ の分割 $-M = a_0 < a_1 < \cdots < a_N = M$ を考える. ただし $N \in \mathbb{N}$ で, すべての点 $a_k\,(k = 0,\, 1,\, \ldots,\, N)$ は $\mathsf{P}\big(\{a_k\}\big) = 0$ となるように取る. P の不連続点は高々可算個しかないのでこのような取り方ができることに注意せよ. このとき

$$h(x) = \sum_{k=1}^{N} c_k \mathbb{1}_{(a_{k-1},\, a_k]}(x)$$

の形の階段関数を考える. ここで $c_1,\, c_2,\, \ldots,\, c_N$ は実数である. この階段関数に対して

$$\int_{-\infty}^{\infty} h(x)\, \mathrm{d}\mathsf{P}_n(x) = \sum_{k=1}^{N} c_k \mathsf{P}_n\big((a_{k-1},\, a_k]\big),$$
$$\int_{-\infty}^{\infty} h(x)\, \mathrm{d}\mathsf{P}(x) = \sum_{k=1}^{N} c_k \mathsf{P}\big((a_{k-1},\, a_k]\big)$$

と書けるから, 上の式の右辺をみると (2) を仮定すると (3) が階段関数 $h(x)$ に成立することがわかる. 次に h がコンパクト台をもつ連続関数の場合を考える. $\forall\epsilon > 0$ に対して階段関数 $h_\epsilon(x)$ が存在して

$$\sup_{x\in\mathbb{R}} \big|h(x) - h_\epsilon(x)\big| < \epsilon$$

とできることが知られている. このことに注意して

$$\begin{aligned}
&\left|\int_{-\infty}^{\infty} h(x)\, \mathrm{d}\mathsf{P}_n(x) - \int_{-\infty}^{\infty} h(x)\, \mathrm{d}\mathsf{P}(x)\right| \\
&\quad\le \left|\int_{-\infty}^{\infty} \{h(x) - h_\epsilon(x)\}\, \mathrm{d}\mathsf{P}_n(x)\right| + \left|\int_{-\infty}^{\infty} h_\epsilon(x)\, \mathrm{d}\mathsf{P}_n(x) - \int_{-\infty}^{\infty} h_\epsilon(x)\, \mathrm{d}\mathsf{P}(x)\right| \\
&\qquad + \left|\int_{-\infty}^{\infty} \{h(x) - h_\epsilon(x)\}\, \mathrm{d}\mathsf{P}(x)\right| \\
&\quad\le 2\sup_{x\in\mathbb{R}} \big|h(x) - h_\epsilon(x)\big| + \left|\int_{-\infty}^{\infty} h_\epsilon(x)\, \mathrm{d}\mathsf{P}_n(x) - \int_{-\infty}^{\infty} h_\epsilon(x)\, \mathrm{d}\mathsf{P}(x)\right|
\end{aligned}$$

より, コンパクト台をもつ $h(x)$ に対して (F.1) が成立することがわかる.

(3) $\Longrightarrow$ (4): $N > 1$ とし

$$g_N(x) = \begin{cases} 1 & (|x| \le N) \\ 0 & (|x| \ge N+1) \end{cases}$$

なる連続関数を考える. (3) を仮定すると $h_N(x) = h(x) g_N(x)$ に対して

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} h(x)\,\mathrm{d}\mathsf{P}_n(x) = \int_{-\infty}^{\infty} h_N(x)\,\mathrm{d}\mathsf{P}(x)$$

が成り立つ. さらに

$$\begin{aligned}\lim_{n\to\infty} \int_{-\infty}^{\infty} h(x)\,\mathrm{d}\mathsf{P}_n(x) &= \lim_{n\to\infty} \int_{-\infty}^{\infty} h_N(x)\,\mathrm{d}\mathsf{P}_n(x) \\ &\quad + \lim_{n\to\infty} \int_{-\infty}^{\infty} h(x)\bigl\{1 - g_N(x)\bigr\}\,\mathrm{d}\mathsf{P}_n(x) \qquad \text{(F.3)} \\ \lim_{n\to\infty} \int_{-\infty}^{\infty} h(x)\,\mathrm{d}\mathsf{P}(x) &= \lim_{n\to\infty} \int_{-\infty}^{\infty} h_N(x)\,\mathrm{d}\mathsf{P}(x) \\ &\quad + \lim_{n\to\infty} \int_{-\infty}^{\infty} h(x)\bigl\{1 - g_N(x)\bigr\}\,\mathrm{d}\mathsf{P}(x) \qquad \text{(F.4)}\end{aligned}$$

と書けることに注意する. (F.3) と (F.4) の右辺の 2 項目を評価すると

$$\begin{aligned}&\limsup_{n\to\infty} \int_{-\infty}^{\infty} h(x)\bigl\{1 - g_N(x)\bigr\}\,\mathrm{d}\mathsf{P}_n(x) \\ &\quad\le \sup_{x\in\mathbb{R}} |h(x)| \limsup_{n\to\infty} \int_{-\infty}^{\infty} \bigl\{1 - g_N(x)\bigr\}\,\mathrm{d}\mathsf{P}_n(x) \\ &\quad\le \sup_{x\in\mathbb{R}} |h(x)| \left\{1 - \liminf_{n\to\infty} \int_{-\infty}^{\infty} g_N(x)\,\mathrm{d}\mathsf{P}_n(x)\right\} \\ &\quad= \sup_{x\in\mathbb{R}} |h(x)| \left\{1 - \int_{-\infty}^{\infty} g_N(x)\,\mathrm{d}\mathsf{P}(x)\right\} \\ &\quad= \sup_{x\in\mathbb{R}} |h(x)| \bigl\{1 - \mathsf{P}\bigl([-N,\, N]\bigr)\bigr\} \\ &\quad= \sup_{x\in\mathbb{R}} |h(x)| \mathsf{P}\bigl([-N,\, N]^c\bigr) \qquad \text{(F.5)}\end{aligned}$$

を得る. 同様にすると

$$\limsup_{n\to\infty} \int_{-\infty}^{\infty} h(x)\bigl\{1 - g_N(x)\bigr\}\,\mathrm{d}\mathsf{P}(x) \le \sup_{x\in\mathbb{R}} \mathsf{P}\bigl([-N,\, N]^c\bigr) \qquad \text{(F.6)}$$

を得る. (F.3) – (F.6) を合わせると

$$
\begin{aligned}
&\limsup_{n\to\infty}\left|\int_{-\infty}^{\infty} h(x)\,\mathrm{d}\mathsf{P}_n(x)-\int_{-\infty}^{\infty} h(x)\,\mathrm{d}\mathsf{P}(x)\right| \\
&\quad\le \limsup_{n\to\infty}\left|\left(\int_{-\infty}^{\infty} h(x)\,\mathrm{d}\mathsf{P}_n(x)-\int_{-\infty}^{\infty} h(x)\,\mathrm{d}\mathsf{P}(x)\right)\right. \\
&\qquad\left.+\int_{-\infty}^{\infty} h(x)\bigl\{1-g_N(x)\bigr\}\,\mathrm{d}\mathsf{P}_n(x)+\int_{-\infty}^{\infty} h(x)\bigl\{1-g_N(x)\bigr\}\,\mathrm{d}\mathsf{P}(x)\right| \\
&\quad\le \limsup_{n\to\infty}\left|\int_{-\infty}^{\infty} h(x)\,\mathrm{d}\mathsf{P}_n(x)-\int_{-\infty}^{\infty} h(x)\,\mathrm{d}\mathsf{P}(x)\right|+2\sup_{x\in\mathbb{R}}|h(x)|\mathsf{P}\bigl([-N,\,N]^c\bigr) \\
&\quad\longrightarrow 0\,(N\to\infty)
\end{aligned}
$$

となるので, (4) を示せた.
(4) $\Longrightarrow$ (1): $a<b$ に対して

$$
h_{a,\,b}=\begin{cases} 1 & (x\le a) \\ \dfrac{b-x}{b-a} & (a<x\le b) \\ 0 & (x\ge b) \end{cases}
$$

とおくと

$$
\mathbb{1}_{(-\infty,\,a]}(x)\le h_{a,\,b}(x)\le \mathbb{1}_{(-\infty,\,b]}(x)
$$

となる. 辺々を P_n と P で積分すると

$$
\begin{aligned}
\mathsf{F}(x)&=\int_{-\infty}^{\infty}\mathbb{1}_{(-\infty,\,a]}(x)\,\mathrm{d}\mathsf{P}(x)\le\int_{-\infty}^{\infty} h_{a,\,b}(x)\,\mathrm{d}\mathsf{P}(x)\le=\int_{-\infty}^{\infty}\mathbb{1}_{(-\infty,\,b]}(x)\,\mathrm{d}\mathsf{P}(x) \\
&=\mathsf{F}(b)
\end{aligned}
\tag{F.7}
$$

を得る. よって

$$
\limsup_{n\to\infty}\mathsf{F}_n(a)\le\lim_{n\to\infty} h_{a,\,b}(x)\,\mathrm{d}\mathsf{P}_n(x)=\int_{-\infty}^{\infty} h_{a,\,b}(x)\,\mathrm{d}\mathsf{P}(x)
$$

となる. 上式と (F.7) と合わせると

$$
\limsup_{n\to\infty}\mathsf{F}_n(a)\le\mathsf{F}(b)
$$

が任意の $a<b$ に成り立つことがわかる. b を右から a に近づけると

$$
\limsup_{n\to\infty}\mathsf{F}_n(a)\le\mathsf{F}(a+0)
$$

を得る. 同様に

$$
\liminf_{n\to\infty}\mathsf{F}_n(b)\ge\lim_{n\to\infty} h_{a,\,b}(x)\,\mathrm{d}\mathsf{P}_n(x)=\int_{-\infty}^{\infty} h_{a,\,b}(x)\,\mathrm{d}\mathsf{P}(x)\ge\mathsf{F}(a)
$$

となるので

$$\liminf_{n\to\infty} \mathsf{F}_n(b) \geq \mathsf{F}(b-0)$$

を得る. $a < b$ は任意だったので, a と b を x とすると

$$\mathsf{F}(x-0) \leq \liminf_{n\to\infty} \mathsf{F}_n(x) \leq \limsup_{n\to\infty} \mathsf{F}_n(x) \leq \mathsf{F}(x+0) \tag{F.8}$$

を得る. F の連続点 x では $\mathsf{F}(x-0) = \mathsf{F}(x+0)$ となるので (F.8) の式は等号で成立することがわかる. よって (1) が示せた.

F.3.2　Glivenko の定理の証明の準備

定義 F.15. $\big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ 上の分布族 $\{\mathsf{P}_\lambda;\, \lambda \in \Lambda\}$ が**タイト** (tight) であるとは, $\forall > 0$ に対して次をみたすような $M > 0$ が存在することである.

$$\mathsf{P}_\lambda\big([-M,\, M]^c\big) < \epsilon \qquad (\forall \lambda \in \Lambda). \tag{F.9}$$

注意 F.16. (F.9) は

$$\lim_{M\to\infty} \sup_{\lambda\in\Lambda} \mathsf{P}_\lambda\big([-M,\, M]^c\big) = 0 \tag{F.10}$$

と書き直すことができる. 特に $\{\mathsf{P}_\lambda\}_{\lambda\in\Lambda}$ が列 $\{\mathsf{P}_n\}_{n\in\mathbb{N}}$ のとき

$$\lim_{M\to\infty} \limsup_{n\to\infty} \mathsf{P}_n\big([-M,\, M]^c\big) = 0 \tag{F.11}$$

と同値となる. 実際 (F.10 から (F.11) は明らかなので, 逆を示す. $\forall \epsilon > 0$ と各 $n = 1, 2, \ldots$ に対して

$$\mathsf{P}_n\big([-M_n,\, M_n]^c\big) < \epsilon$$

となる $M_n > 0$ が存在する. タイトであるとは M_n が n に依存せずに取れることである. したがって $\{M_n\}_{n\in\mathbb{N}}$ が有界であることを示せばよい. (F.10) よりある $M_0 > 0$ が存在して

$$\limsup_{n\to\infty} \mathsf{P}_n\big([-M_0,\, M_0]^c\big) < \frac{\epsilon}{2}$$

とできる. さらに数列 $\{a_n\}_{n\in\mathbb{N}}$ に対して $\limsup_{n\to\infty} a_n \leq \sup_n a_n$ が成立することに注意するとある $n_0 \in \mathbb{N}$ が存在して

$$\sup_{n\geq n_0} \mathsf{P}_n\big([-M_0,\, M_0]\big)$$

となる. よって $n \geq n_0$ に対しては M_n を M_0 と取ると

$$\sup_n M_m = \max\{M_1,\, M_2,\, \ldots,\, M_{n_0},\, M_0\}$$

となり, $\{M_n\}_{n\in\mathbb{N}}$ は有界となることがわかる. □

定理 F.17. (Helly (ヘリー) の定理) 分布列 $\{\mathsf{P}_n\}_{n\in\mathbb{N}}$ はタイトし, P_n の分布関数を F_n とする. するとある部分列 $\{\mathsf{F}_{n(k)}\}_{k\in\mathbb{N}}$ と分布 F が存在して

$$\mathsf{F}_{n(k)} \rightsquigarrow \mathsf{F}\,(k \to \infty)$$

となる.

Proof. Bass (2011, pp.369–370) を参照. □

補題 F.18. φ を $\big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ 上の分布 P の特性関数とする. このとき $\forall M > 0$ に対して $\epsilon = 2/M$ とおくと

$$\mathsf{P}\big([-M,\, M]^c\big) \le \frac{1}{\epsilon}\int_{-\epsilon}^{\epsilon}\{1-\varphi(t)\}\,\mathrm{d}t$$

が成り立つ.

Proof. 特性関数 φ の定義から

$$\frac{1}{2\epsilon}\int_{-\epsilon}^{\epsilon}\varphi(t)\,\mathrm{d}t = \frac{1}{2\epsilon}\int_{-\epsilon}^{\epsilon}\left\{\int_{-\infty}^{\infty} e^{\sqrt{-1}tx}\,\mathrm{d}\mathsf{P}(x)\right\}\mathrm{d}t$$

と書ける. Fubini の定理を適用すると

$$\begin{aligned}
\frac{1}{2\epsilon}\int_{-\epsilon}^{\epsilon}\left\{\int_{-\infty}^{\infty} e^{\sqrt{-1}tx}\,\mathrm{d}\mathsf{P}(x)\right\}\mathrm{d}t &= \frac{1}{2\epsilon}\int_{-\infty}^{\infty}\left\{\int_{-\epsilon}^{\epsilon} e^{\sqrt{-1}tx}\,\mathrm{d}t\right\}\mathrm{d}\mathsf{P}(x)\\
&= \frac{1}{2\epsilon}\int_{-\infty}^{\infty}\left[\frac{e^{\sqrt{-1}tx}}{\sqrt{-1}x}\right]_{-\epsilon}^{\epsilon}\mathrm{d}\mathsf{P}(x)\\
&= \frac{1}{2\epsilon}\int_{-\infty}^{\infty}\frac{e^{\sqrt{-1}\epsilon x}-e^{-\sqrt{-1}\epsilon x}}{\sqrt{-1}x}\,\mathrm{d}\mathsf{P}(x)\\
&= \frac{1}{2\epsilon}\int_{-\infty}^{\infty}\frac{\cos\epsilon x+\sqrt{-1}\sin\epsilon x-\big(\cos\epsilon x-\sqrt{-1}\sin\epsilon x\big)}{\sqrt{-1}x}\,\mathrm{d}\mathsf{P}(x)\\
&= \int_{-\infty}^{\infty}\frac{\sin\epsilon x}{\epsilon x}\,\mathrm{d}\mathsf{P}(x)
\end{aligned}$$

となる. よって

$$\begin{aligned}
\frac{1}{2\epsilon}\int_{-\infty}^{\infty}\{1-\varphi(t)\}\,\mathrm{d}t &= \int_{-\infty}^{\infty}\left(1-\frac{\sin\epsilon x}{\epsilon x}\right)\mathrm{d}\mathsf{P}(x)\\
&\ge \int_{|\epsilon x|>2}\left(1-\frac{\sin\epsilon x}{\epsilon x}\right)\mathrm{d}\mathsf{P}(x)\\
&> \int_{|\epsilon x|>2}\frac{1}{2}\,\mathrm{d}\mathsf{P}(x)\\
&= \frac{1}{2}\mathsf{P}\left(\left[-\frac{2}{\epsilon},\,\frac{2}{\epsilon}\right]^c\right)
\end{aligned}$$

を得る. 最後の不等式は $|y| > 2$ のとき

$$\left|\frac{\sin y}{y}\right| \leq \left|\frac{1}{y}\right| < \frac{1}{2}$$

であることよりわかる. □

補題 F.19. 分布列 $\{\mathsf{P}_n\}_{n\in\mathbb{N}}$ が分布 P に弱収束するための必要十分条件は次の条件 (1) と (2) がともに成立することである.
(1) $\{\mathsf{P}_n\}_{n\in\mathbb{N}}$ は相対コンパクトである.
(2) 任意の部分列 $\{\mathsf{P}_{n(k)}\}_{k\in\mathbb{N}}$ の極限分布は P に限る.

Proof. P_n が P に弱収束していれば, (1) と (2) は成り立つのは明らかなので, 逆を示す. そのために P_n が P に弱収束していても (1) が成り立っていたら, (2) が不成立であることを示せばよい. F_n と F を P_n と P の分布関数とする. P_n が P に弱収束していないとき

$$\left|\mathsf{F}_{n(k)}(x_0) - \mathsf{F}(x_0)\right| > \epsilon \tag{F.12}$$

となる $\epsilon > 0$ と F の連続点 x_0 および部分列 $1 \leq n(1) < n(2) < \cdots$ があることである. このとき (1) が成り立っているので, さらなる部分列 $\{\mathsf{F}_{n(k(\ell))}\}_{\ell\in\mathbb{N}}$ と分布関数 P^* をうまく選ぶと

$$\mathsf{P}_{n(k(\ell))} \rightsquigarrow \mathsf{P}^* \, (\ell \to \infty)$$

となる. この P^* は P と異なる. なぜならば, もし $\mathsf{P} = \mathsf{P}^*$ ならば $\mathsf{F}_{n(k(\ell))}(x_0) \overset{\ell\to\infty}{\longrightarrow} \mathsf{F}(x_0)$ となるので, (F.12) と矛盾するので, $\mathsf{P} \neq \mathsf{P}^*$ であることがわかる. よって (2) は成立しない. これで補題の題意は示せた.
□

F.3.3　Glivenko の定理の証明

補題 F.19(1)(2) を示せばよい. そのために $\{\mathsf{P}_n\}_{n\in\mathbb{N}}$ がタイトであることを示す. すると Helly の定理から (1) が成立することがわかる. また (2) は明らかである.
$\{\mathsf{P}_n\}_{n\in\mathbb{N}}$ **がタイトであることの証明**: 補題 F.18 から

$$\mathsf{P}_n\big([-M,\, M]^c\big) \leq \frac{1}{\epsilon}\int_{-\epsilon}^{\epsilon}\{1-\varphi(t)_n\}\,\mathrm{d}t, \qquad \left(\epsilon = \frac{2}{M}\right)$$

である. 仮定 $\varphi_n(t) \overset{n\to\infty}{\longrightarrow} \varphi(t)$ と Fatou の補題から

$$\limsup_{n\to\infty} \mathsf{P}_n\big([-M,\, M]^c\big) \leq \frac{1}{\epsilon}\int_{-\epsilon}^{\epsilon}\{1-\varphi(t)\}\,\mathrm{d}t$$

が成り立つ. $M \to \infty$ のとき $\epsilon \to 0$ と定理 F.3(1) に注意すると

$$\lim_{M\to\infty} \limsup_{n\to\infty} \mathsf{P}_n\bigl([-M,\ M]^c\bigr) \le \lim_{\epsilon\to 0}\left\{\frac{1}{\epsilon}\int_{-\epsilon}^{\epsilon}\{1-\varphi(t)\}\,\mathrm{d}t\right\} = 2 - 2\varphi(0) = 0$$

となり, $\{\mathsf{P}_n\}_{n\in\mathbb{N}}$ はタイトである. □

F.4　章末注釈と参考文献

第 G 章　大数の強法則と中心極限定理の証明

G.1　大数の強法則の証明

G.1.1　Borel-Cantelli の補題

$(\Omega, \mathcal{A}, \mathsf{Pr})$ を確率空間とする．$\{A_n\}_{n=1}^{\infty}$ を Ω の部分集合列とする．このとき

$$\limsup A_n := \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m = \lim_{n\to\infty} \bigcup_{m=n}^{\infty} A_m,$$
$$\liminf A_n := \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m = \lim_{m\to\infty} \bigcap_{m=n}^{\infty} A_m,$$

と定める．これを

$$\limsup A_n := \{\omega \in \Omega;\, \omega \in A_n,\ \text{i.o.}\} = \{A_n,\ \text{i.o.}\}$$

とも書くことにする．また

$$\liminf A_n \subset \limsup A_n$$

となるとなることに注意する．．なぜならば $\omega \in \liminf A_n$ ならば，ある $n_0 \in \mathbb{N}$ が存在して $\omega \in \bigcap_{m=n_0}^{\infty} A_m$ となるので，$\forall n \in \mathbb{N}$ に対して $\omega \in \bigcup_{m=n}^{\infty} A_m$ である．よって　$\omega \in \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$ がわかる．

また

$$\limsup_{n\to\infty} \mathbb{1}_{A_n} = \mathbb{1}_{\limsup A_n}, \quad \liminf_{n\to\infty} \mathbb{1}_{A_n} = \mathbb{1}_{\liminf A_n}$$

となる．さらに $\{A_n\}_{n=1}^{\infty} \subset \mathcal{A}$ のとき，Fatou の補題より

$$\begin{aligned}
\mathsf{Pr}\Bigl(\liminf A_n\Bigr) &= \int \mathbb{1}_{\liminf A_n}(x)\, \mathrm{d}\mathsf{Pr}(x) \\
&= \int \liminf_{n\to\infty} \mathbb{1}_{A_n}(x)\, \mathrm{d}\mathsf{Pr}(x) \\
&\le \liminf_{n\to\infty} \int \mathbb{1}_{A_n}(x)\, \mathrm{d}\mathsf{Pr}(x) = \liminf_{n\to\infty} \mathsf{Pr}(A_n)
\end{aligned}$$

となる．

補題 G.1. (Borel-Cantelli の第 1 補題) $\{A_n\}_{n=1}^{\infty} \subset \mathcal{A}$ に対して

$$\sum_{n=1}^{\infty} \mathsf{Pr}(A_n) < \infty \;\Rightarrow\; \mathsf{Pr}\big(A_n,\ \text{i.o.}\big) = 0.$$

Proof. $N(\omega) := \sum_{n=1}^{\infty} \mathbb{1}_{A_n}(\omega)\,(\omega \in \Omega)$ と定める．単調収束定理から

$$\mathsf{E}\big[N\big] = \int \sum_{n=1}^{\infty} \mathbb{1}_{A_n}(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) = \sum_{n=1}^{\infty} \int \mathbb{1}_{A_n}(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) = \sum_{n=1}^{\infty} \mathsf{Pr}(A_n) < \infty$$

となる．よって $\mathsf{Pr}\big(N < \infty\big) = 1$ がわかる．したがって $\mathsf{Pr}\big(A_n,\ \text{i.o.}\big) = 0$ がわかる． □

定理 G.2. $\{X_n\}_{n=1}^{\infty}$, X は確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率変数列とする．このとき次の 2 つの条件は同値である．

(1) $X_n \xrightarrow{\mathrm{P}} X$.

(2) $\{X_n\}_{n=1}^{\infty}$ の任意の部分列 $\{X_{n(k)}\}_{k=1}^{\infty}$ に対して，さらなる部分列 $\{X_{n(k(\ell))}\}_{\ell=1}^{\infty}$ が存在して $X_{n(k(\ell))} \xrightarrow{\text{a.s.}} X\ (\ell \to \infty)$ となる．すなわち

$$\mathsf{Pr}\left(\lim_{\ell \to \infty} X_{n(k(\ell))} = X\right) = 1.$$

Proof. <u>(1) ⇒ (2) の証明</u>: $\{\epsilon(\ell)\}_{\ell=1}^{\infty}$ は正の実数列で $\epsilon(\ell) \downarrow 0\,(\ell \to \infty)$ とする．$X_n \xrightarrow{\mathrm{P}} X$ から, 各 $\ell \geq 2$ に対して $n(k(\ell)) > n(k(\ell-1))$ が存在して

$$\mathsf{Pr}\left(|X_{n(k(\ell))} - X| > \epsilon(\ell)\right) \leq \frac{1}{2^{\ell}}$$

とできる．これより

$$\sum_{\ell=1}^{\infty} \mathsf{Pr}\left(|X_{n(k(\ell))} - X| > \epsilon(\ell)\right) < \infty$$

となるので, Borel-Cantelli の第 1 の補題から

$$\mathsf{Pr}\left(|X_{n(k(\ell))} - X| > \epsilon(\ell),\ \text{i.o.}\right) = 0.$$

すなわち

$$\mathsf{Pr}\left(\lim_{\ell \to \infty} X_{n(k(\ell))} = X\right) = 1$$

がわかる (詳細な議論は注意 G.3 を参照のこと).

(2) $\Rightarrow$ (1) の証明: $\forall\delta > 0$ に対して

$$y_n := \mathsf{Pr}\big(|X_n - X| > \delta\big)$$

とおく．すると任意の $\{y_{n(k)}\}_{k=1}^{\infty}$ に対して部分列 $\{y_{n(k(\ell))}\}_{\ell=1}^{\infty}$ が存在して

$$y_{n(k(\ell))} \overset{\ell\to\infty}{\longrightarrow} 0$$

となる[1]．このことと命題 B.5 から $y_n \overset{n\to\infty}{\longrightarrow} 0$ がわかる． □

注意 G.3. 事象 $\{\omega \in \Omega;\ \lim_{n\to\infty} X_n(\omega) = X(\omega)\}$ は以下のように書きかえることができる．$\forall\epsilon > 0$ に対して，ある整数 $n := n(\epsilon, \omega) \geq 1$ が存在して，$\forall m \geq 1\,(m \in \mathbb{N})$ に対して

$$\big|X_{n+m}(\omega) - X(\omega)\big| < \epsilon \tag{G.1}$$

を意味する． さらに (G.1) が成立することは，$\forall\epsilon > 0$ に対して，ある $n \in \mathbb{N}$ が存在して，$\forall m \geq 1\,(m \in \mathbb{N})$ に対して

$$\omega \in \Big\{\omega \in \Omega;\ \big|X_{n+m}(\omega) - X(\omega)\big| < \epsilon\Big\} \tag{G.2}$$

と同値である．これは $\forall\epsilon > 0$ に対して

$$\omega \in \bigcup_{n=1}^{\infty} \bigcap_{m=1}^{\infty} \Big\{\omega \in \Omega;\ \big|X_{n+m}(\omega) - X(\omega)\big| < \epsilon\Big\} \tag{G.3}$$

を意味する．さらに $\forall\epsilon > 0$ に対して

$$\omega \in \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m=1}^{\infty} \left\{\omega \in \Omega;\ \big|X_{n+m}(\omega) - X(\omega)\big| < \frac{1}{k}\right\}$$

と同値である．以上から

$$\begin{aligned}&\{\omega \in \Omega;\ \lim_{n\to\infty} X_n(\omega) = X(\omega)\}\\ &\quad= \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m=1}^{\infty} \left\{\omega \in \Omega;\ \big|X_{n+m}(\omega) - X(\omega)\big| < \frac{1}{k}\right\}\end{aligned}$$

を得る． □

[1] なぜならば $X_{n(k(\ell))} \overset{\text{a.s.}}{\to} X\ (\ell \to \infty)$ なので．

G.1.2　大数の強法則の証明

定理 G.4. $X_1, X_2, \dots$ は対独立な確率変数列で各 $X_n\,(n \in \mathbb{N})$ は同じ分布に従い, $\mathsf{E}\bigl[|X_n|\bigr] < \infty$ とする. $\mathsf{E}\bigl[X_n\bigr] =: \mu$ と $S_n := X_1 + X_2 + \cdots + X_n$ とおく. このとき

$$\frac{S_n}{n} \overset{\text{a.s.}}{\to} \mu \qquad (n \to \infty)$$

が成立する.

大数の強法則を証明するための補題を２つ述べる.

補題 G.5. $n \in \mathbb{N}$ に対して $Y_n = X_n \mathbb{1}\{|X_n| \le n\}$ とし,

$$T_n := Y_1 + Y_2 + \cdots + Y_n \tag{G.4}$$

とする. このとき

$$\frac{T_n}{n} \overset{\text{a.s.}}{\to} \mu \Rightarrow \frac{S_n}{n} \overset{\text{a.s.}}{\to} \mu$$

が成立する.

Proof. Y_n の定義から

$$\sum_{n=1}^{\infty} \mathsf{Pr}\bigl(X_n \ne Y_n\bigr) = \sum_{n=1}^{\infty} \mathsf{Pr}\bigl(|X_n| > n\bigr) \le \int_0^{\infty} \mathsf{Pr}\bigl(|X_1| > t\bigr)\,\mathrm{d}t = \mathsf{E}\bigl[|X_1|\bigr] < \infty$$

となる. よって Borel-Cantelli の第 1 の補題から

$$\mathsf{Pr}\Bigl(X_n \ne Y_n, \text{ i.o.}\Bigr) = 0$$

がわかる. すると $\omega \in \Omega$ に対して有限個の n に対して $X_n \ne Y_n$ なので

$$\lim_{n\to\infty} |S_n(\omega) - T_n(\omega)| = 0$$

となる. よって定理の主張は証明された. □

補題 G.6. 定理 G.4 と補題 G.5 の仮定のもと

$$\sum_{n=1}^{\infty} \frac{\mathsf{Var}\bigl(Y_n\bigr)}{n^2} \le 4\mathsf{E}\bigl[|X_1|\bigr] < \infty$$

が成立する.

Proof. Fubini の定理を用いると

$$\begin{aligned}\mathsf{Var}\big[Y_n\big] \le \mathsf{E}\big[Y_n^2\big] &= \int_0^\infty 2t\,\mathsf{Pr}\big(|Y_n| > t\big)\,\mathrm{d}t = \int_0^n 2t\,\mathsf{Pr}\big(|Y_n| > t\big)\,\mathrm{d}t \\ &= \int_0^\infty \mathbb{1}\{t < n\} 2t\,\mathsf{Pr}\big(|Y_n| > t\big)\,\mathrm{d}t\end{aligned}$$

となる．よって

$$\begin{aligned}\sum_{n=1}^\infty \frac{\mathsf{E}\big[Y_n^2\big]}{n^2} &\le \sum_{n=1}^\infty \frac{1}{n^2} \int_0^\infty \mathbb{1}\{t < n\} 2t\,\mathsf{Pr}\big(|Y_n| > t\big)\,\mathrm{d}t \\ &= \int_0^\infty \left\{ \sum_{n=1}^\infty \frac{\mathbb{1}\{t < n\}}{n^2} \right\} 2t\,\mathsf{Pr}\big(|Y_n| > t\big)\,\mathrm{d}t\end{aligned}$$

となる．最後の等号は, 各項が非負なので, Fubini の定理を用いることで和と積分の記号の交換が保証されることがわかる．よって後は $\forall t \ge 0$ に対して

$$2t \sum_{n>t} \frac{1}{n^2} \le 4 \tag{G.5}$$

を示せばよい．$m \ge 2\,(m \in \mathbb{N})$ に対して

$$\sum_{n \ge m} \frac{1}{n^2} \le \int_{m-1}^\infty \frac{\mathrm{d}x}{x^2} = \frac{1}{m-1}$$

となる．$t \ge 1$ のとき, (G.5) の和は $n = [t] + 1 \ge 2$ で始まるので

$$2t \sum_{n>t} \frac{1}{n^2} \le \frac{2t}{[t]} \le 4 \tag{G.6}$$

となる[2]．$0 \le t \le 1$ の場合

$$2t \sum_{n>t} \frac{1}{n^2} \le 2\left(1 + \sum_{n=2}^\infty \frac{1}{n^2}\right) \le 4 \tag{G.7}$$

となる．よって (G.6), (G.7) から (G.5) がわかる．　□

定理 G.4 **の証明**: $n \in \mathbb{N}$ に対して

$$X_n^+ := \max\{X_n, 0\}, \qquad X_n^- := \max\{-X_n, 0\}$$

[2] なぜならば

$$\frac{t}{[t]} \le 2 \qquad (t \ge 1)$$

であることからわかる．

とおくと X_n^+ と X_n^- は定理 G.4 の仮定をみたし, $X_n = X_n^+ - X_n^-$ である. したがって一般性を失うことなく $X_n \geq 0\,(n \in \mathbb{N})$ と仮定してよい.

$\alpha > 1$ をとり

$$k(n) = [\alpha^n]$$

と定める. $\forall\epsilon > 0$ を取る. Chebyshev の不等式から

$$\begin{aligned}\sum_{n=1}^{\infty} \mathsf{Pr}\Big(\big|T_{k(n)} - \mathsf{E}\big[T_{k(n)}\big]\big| > \epsilon k(n)\Big) &\leq \frac{1}{\epsilon^2}\sum_{n=1}^{\infty} \frac{\mathsf{Var}\big[T_{k(n)}\big]}{\{k(n)\}^2} \\ &= \frac{1}{\epsilon^2}\sum_{n=1}^{\infty} \frac{1}{\{k(n)\}^2}\sum_{m=1}^{k(n)} \mathsf{Var}\big[Y_m\big] \\ &= \frac{1}{\epsilon^2}\sum_{m=1}^{\infty} \mathbb{V}\big[Y_m\big] \sum_{n:\,k(n)\geq m} \frac{1}{\{k(n)\}^2}\end{aligned} \tag{G.8}$$

を得る. 最後の等号は, すべての項が非負なので Fubini の定理を用いて和の順番を交換した. $n \geq 1$ に対して

$$k(n) = [\alpha^n] \quad かつ \quad [\alpha^n] \geq \frac{\alpha^n}{2}$$

なので

$$\sum_{n:\,\alpha^n \geq m} \frac{1}{[\alpha^n]^2} \leq 4 \sum_{n:\,\alpha^n\geq m} \frac{1}{\alpha^{2n}} \leq \frac{4}{(1-\alpha^{-2})m^2} \tag{G.9}$$

となる. (G.8) と (G.9) を合わせると

$$\begin{aligned}\sum_{n=1}^{\infty} \mathsf{Pr}\Big(\big|T_{k(n)} - \mathsf{E}\big[T_{k(n)}\big]\big| > \epsilon k(n)\Big) &\leq \frac{4}{(1-\alpha^{-2})\epsilon^2}\sum_{m=1}^{\infty}\frac{\mathsf{E}[Y_m^2]}{m^2} \\ &< \infty \qquad (\because 補題\ \mathrm{G.6})\end{aligned}$$

となる. ϵ は任意だったので

$$\frac{T_{k(n)} - \mathsf{E}[T_{k(n)}]}{k(n)} \overset{\text{a.s.}}{\to} 0$$

を得る. 有界収束定理から

$$\mathsf{E}[Y_k] \overset{n\to\infty}{\longrightarrow} \mathsf{E}[X_1]$$

である. よって

$$\frac{\mathsf{E}\big[T_{k(n)}\big]}{k(n)} \overset{n\to\infty}{\longrightarrow} \mathsf{E}[X_1]$$

を得る．これより

$$\frac{T_{k(n)}}{k(n)} \overset{\text{a.s.}}{\to} \mathsf{E}[X_1] \qquad (n \to \infty)$$

がわかる．$Y_n \geq 0\,(\forall n \in \mathbb{N})$ なので，$k(n) \leq m < k(n+1)\,(m \in \mathbb{N})$ に対して

$$\frac{T_{k(n)}}{k(n+1)} \leq \frac{T_m}{m} \leq \frac{T_{k(n+1)}}{k(n)}$$

である．$k(n) = [\alpha^n]$ だったので，

$$\frac{k(n+1)}{k(n)} \overset{n\to\infty}{\longrightarrow} \alpha$$

である．よって

$$\frac{1}{\alpha}\mathsf{E}[X_1] \leq \liminf_{m\to\infty} \frac{T_m}{m} \leq \limsup_{m\to\infty} \frac{T_m}{m} \leq \alpha \mathsf{E}[X_1]$$

となる．$\alpha > 1$ は任意だったので

$$\frac{T_m}{m} \overset{\text{a.s.}}{\to} \mathsf{E}[X_1]$$

がわかる．□

G.2　中心極限定理の証明: Lindeberg の証明

まず $\mathsf{E}[X_1^3] < \infty$ を仮定して証明をする．

$$Y_j = \frac{X_j - \mu}{\sigma}, \qquad (j = 1,\, 2,\, \ldots,\, n)$$

とおくと $\mathsf{E}[Y_1] = 0$, $\mathsf{E}[Y_1^2] = 1$, $\mathsf{E}[|Y_1|^3] < \infty$ となることに注意する．$Z_1,\, Z_2,\, \ldots,\, Z_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}(0,\, 1)$ とし

$$S_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} Y_j, \quad S = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} Z_j \sim \mathsf{N}(0,\, 1)$$

とおく．

$\forall x \in \mathbb{R}$ に対して

$$\mathsf{Pr}\bigl(S_n \leq x\bigr) \overset{n\to\infty}{\longrightarrow} \mathsf{Pr}\bigl(S \leq x\bigr) \tag{G.10}$$

を示せばよい．ここで

$$g(y) = \mathbb{1}_{(-\infty,\, x]}(y) \qquad (y \in \mathbb{R})$$

とおくと (G.10) を

$$\mathsf{E}\big[g(S_n)\big] \overset{n\to\infty}{\longrightarrow} \mathsf{E}\big[g(S)\big] \tag{G.11}$$

と書きかえることができる．さらに $\forall\delta > 0$ に対して関数 $h_{\delta,\,x} : \mathbb{R} \longrightarrow [0,\, 1]$ を以下のような性質をもつ関数とする．

- $h_{\delta,\,x}(y) = 1\,(y \le x)$.
- $h_{\delta,\,x}(y) = 0\,(y \le x + \delta)$.
- $h_{\delta,\,x}$ は有界な 3 次の導関数をもつ．3 次の導関数の絶対値の上界を C と書くことにする．

さらに $h_{\delta,\,x-\delta}(y) = h_{\delta,\,x}(y+\delta)$ と定める．すると

$$h_{\delta,\,x-\delta}(y) \le g(y) \le h_{\delta,\,x}(y), \qquad (y \in \mathbb{R})$$

となる．このことから $\forall x \in \mathbb{R}$ に対して

$$\begin{aligned}\mathsf{Pr}\big(S_n \le x\big) &= \mathsf{E}\big[h_{\delta,\,x}(S_n)\big] \\ &= \mathsf{E}\big[h_{\delta,\,x}(S_n)\big] - \mathsf{E}\big[h_{\delta,\,x}(S)\big] + \mathsf{E}\big[h_{\delta,\,x}(S)\big] \\ &\le \mathsf{Pr}\big(Z \le x+\delta\big) + \big|\mathsf{E}\big[h_{\delta,\,x}(S_n)\big] - \mathsf{E}\big[h_{\delta,\,x}(S)\big]\big|\end{aligned}$$

を得る．同様に

$$\mathsf{Pr}\big(S_n \le x\big) \ge \mathsf{Pr}\big(S \le x-\delta\big) - \big|\mathsf{E}\big[h_{\delta,\,x-\delta}(S_n)\big] - \mathsf{E}\big[h_{\delta,\,x-\delta}(S)\big]\big|$$

を得る．これらの不等式を合わせると

$$\begin{aligned}&\mathsf{Pr}\big(S \le x+\delta\big) + \big|\mathsf{E}\big[h_{\delta,\,x}(S_n)\big] - \mathsf{E}\big[h_{\delta,\,x}(S)\big]\big| \\ &\qquad \ge \mathsf{Pr}\big(S_n \le x\big) \ge \mathsf{Pr}\big(S \le x-\delta\big) - \big|\mathsf{E}\big[h_{\delta,\,x-\delta}(S_n)\big] - \mathsf{E}\big[h_{\delta,\,x-\delta}(S)\big]\big|\end{aligned}$$

となる．δ を 0 に近づけると $\mathsf{Pr}\big(S \le x \pm \delta\big)$ は $\mathsf{Pr}\big(S \le x\big)$ に近づく．したがって

$$\mathsf{E}\big[h_{\delta,\,x}(S_n)\big] \overset{n\to\infty}{\longrightarrow} \mathsf{E}\big[h_{\delta,\,x}(S)\big] \quad と \quad \mathsf{E}\big[h_{\delta,\,x-\delta}(S_n)\big] \overset{n\to\infty}{\longrightarrow} \mathsf{E}\big[h_{\delta,\,x-\delta}(S)\big] \tag{G.12}$$

を示せばよい．よって $h : \mathbb{R} \to [0,\, 1]$ を有界な 3 次の導関数をもつ関数として (G.12) を示す．

$1 \le m \le n+1$ に対して

$$T_m = \frac{1}{\sqrt{n}}\bigl(Y_1 + \cdots + Y_{m-1} + Z_m + \cdots + Z_n\bigr)$$

とおく．このとき

$$\begin{aligned}
\bigl|\mathsf{E}\bigl[h(S_n)\bigr] - \mathsf{E}\bigl[h(S)\bigr]\bigr| &= \left|\sum_{m=1}^{n}\bigl\{\mathsf{E}\bigl[h(T_m)\bigr] - \mathsf{E}\bigl[h(T_{m+1})\bigr]\bigr\}\right| \\
&\le \sum_{m=1}^{n}\bigl|\mathsf{E}\bigl[h(T_m) - h(T_{m+1})\bigr]\bigr| \qquad \text{(G.13)}
\end{aligned}$$

と書きかえる．次に $m = 1, 2, \ldots, n$ に対して

$$U_m = \frac{1}{\sqrt{n}}\bigl(Y_1 + \cdots + Y_{m-1} + Z_{m+1} + \cdots + Z_n\bigr)$$

とおくと

$$T_m = U_m + \frac{1}{\sqrt{n}}Z_m, \qquad T_{m+1} = U_m + \frac{1}{\sqrt{n}}Y_m$$

となる．Taylor 展開を用いると

$$\begin{aligned}
h(T_m) - h(U_m) &= (T_m - U_m)\,\dot{h}\,(U_m) + \frac{1}{2}(T_m - U_m)^2\,\ddot{h}\,(U_m) \\
&\quad + \frac{1}{6}(T_m - U_m)^3\dddot{h}(U_m^{*}), \qquad \text{(G.14)} \\
h(T_{m+1}) - h(U_m) &= (T_{m+1} - U_m)\,\dot{h}\,(U_m) + \frac{1}{2}(T_{m+1} - U_m)^2\frac{1}{2}\,\dot{h}\,(U_m) + \ddot{h}\,(U_m) \\
&\quad + \frac{1}{6}(T_{m+1} - U_m)^3\dddot{h}(U_m^{**}), \qquad \text{(G.15)}
\end{aligned}$$

を得る．ただし $\dot{h}$, $\ddot{h}$, $\dddot{h}$ は h の 1 次, 2 次, 3 次の導関数で, U_m^{*}, U_m^{**} は T_m と U_m の間と T_{m+1} と U_m の間の数である．U_m は Y_m と Z_m の両方とも独立であるので

$$\begin{aligned}
&\mathsf{E}\bigl[\dot{h}\,(U_m)Z_m\bigr] = \mathsf{E}\bigl[\dot{h}\,(U_m)Y_m\bigr] = 0, \\
&\mathsf{E}\bigl[\ddot{h}\,(U_m)Z_m^2\bigr] = \mathsf{E}\bigl[\ddot{h}\,(U_m)\bigr] = \mathsf{E}\bigl[\ddot{h}\,(U_m)Y_m^2\bigr], \\
&\bigl|\mathsf{E}\bigl[\dddot{h}(U_m^{*})\bigr]\bigr| \le C,\ \bigl|\mathsf{E}\bigl[\dddot{h}(U_m^{**})\bigr]\bigr| \le C
\end{aligned}$$

となる．上の期待値と (G.14) と (G.15) を合わせると

$$\begin{aligned}
\Bigl|\mathsf{E}\bigl[h(T_m) - h(T_{m+1})\bigr]\Bigr| &= \Bigl|\mathsf{E}\bigl[h(T_m) - h(U_m) - \bigl(h(T_{m+1}) - h(U_m)\bigr)\bigr]\Bigr| \\
&\le \left|\frac{C}{6n^{3/2}}\mathsf{E}\bigl[|Z_m|^3 + |Y_m|^3\bigr]\right|
\end{aligned}$$

を得る．この式を (G.12) に代入すると

$$\left|\mathsf{E}\big[h(S_n)\big] - \mathsf{E}\big[h(S)\big]\right| \leq \frac{C}{6\sqrt{n}}\mathsf{E}\big[|Z_1|^3 + |Y_1|^3\big] \overset{n\to\infty}{\longrightarrow} 0$$

を得る．よって Y_1 が有界な 3 次の積率をもつとき

$$S_n \rightsquigarrow S \sim \mathsf{N}(0,\,1)$$

が示せた．

最後に $\mathsf{E}\big[Y_1^3\big] < \infty$ の仮定を除くための議論を行う．そのために任意の $\epsilon > 0$ とし, $\sup_{x\in\mathbb{R}} |\ddot{h}(y)| \leq C'$, $\sup_{x\in\mathbb{R}} |\dddot{h}(y)| \leq C'$ をみたす定数 C' をとる．$m = 1,\,2,\,\ldots,\,n$ に対して, $|Y_m| \leq \epsilon\sqrt{n}$ のとき

$$\begin{aligned}
&\left|h(T_m) - h(U_m) - \frac{\dot{h}(U_m)Y_m}{\sqrt{n}} - \frac{\ddot{h}(U_m)Y_m^2}{2n}\right|\mathbb{1}\{|Y_m| \leq \epsilon\sqrt{n}\}\\
&\qquad \leq C'\left|\frac{\dddot{h}(U_m^*)Y_m^3}{6n^{3/2}}\right|\mathbb{1}\{|Y_m| \leq \epsilon\sqrt{n}\}\\
&\qquad \leq \frac{C'\epsilon}{6n}Y_m^2\mathbb{1}\{|Y_m| \leq \epsilon\sqrt{n}\}
\end{aligned}$$

となる．一方 $|Y_m| > \epsilon\sqrt{n}$ のとき

$$\begin{aligned}
&\left|h(T_m) - h(U_m) - \frac{\dot{h}(U_m)Y_m}{\sqrt{n}} - \frac{\ddot{h}(U_m)Y_m^2}{2n}\right|\mathbb{1}\{|Y_m| > \epsilon\sqrt{n}\}\\
&\qquad \leq \left\{\left|h(T_m) - h(U_m) - \frac{\dot{h}(U_m)Y_m}{\sqrt{n}}\right| + \left|\frac{\ddot{h}(U_m)Y_m^2}{2n}\right|\right\}\mathbb{1}\{|Y_m| > \epsilon\sqrt{n}\}\\
&\qquad \leq \left\{\left|\frac{\ddot{h}(U_m^{***})Y_m^2}{2n}\right| + \left|\frac{\ddot{h}(U_m)Y_m^2}{2n}\right|\right\}\mathbb{1}\{|Y_m| > \epsilon\sqrt{n}\}\\
&\qquad \leq \frac{C'Y_m^2}{n}\mathbb{1}\{|Y_m| > \epsilon\sqrt{n}\}
\end{aligned}$$

となる．ただし U_m^{***} は T_m と U_m の間にあるものである．これらの不等式を合わせると

$$\begin{aligned}
&\left|h(T_m) - h(U_m) - \frac{\dot{h}(U_m)Y_m}{\sqrt{n}} - \frac{\ddot{h}(U_m)Y_m^2}{2n}\right|\\
&\qquad \leq \frac{C'\epsilon}{6n}Y_m^2\mathbb{1}\{|Y_m| \leq \epsilon\sqrt{n}\} + \frac{C'Y_m^2}{n}\mathbb{1}\{|Y_m| > \epsilon\sqrt{n}\}
\end{aligned}$$

をなる．この式と (G.15) を合わせると

$$\begin{aligned}\Bigl|\mathsf{E}\bigl[h(S_n)\bigr]-\mathsf{E}\bigl[h(S)\bigr]\Bigr| &\le \frac{C}{\sqrt{n}}\mathsf{E}\bigl[Z_1^3\bigr]+\frac{C'}{6}\epsilon\mathsf{E}\bigl[Y_1^2\mathbb{1}\{|Y_1|\le\epsilon\sqrt{n}\}\bigr]\\ &\quad +C'\mathsf{E}\bigl[Y_1^2\{Y_1>\epsilon\sqrt{n}\}\bigr]\\ &\le \frac{C}{\sqrt{n}}\mathsf{E}\bigl[Z_1^3\bigr]+C'\epsilon+\frac{C'}{6}\mathsf{E}\bigl[Y_1^2\{Y_1>\epsilon\sqrt{n}\}\bigr]\\ &\xrightarrow{n\to\infty}\frac{C'}{6}\epsilon\end{aligned}$$

を得る．上の式で $\epsilon\to 0$ とすれば

$$\mathsf{E}\bigl[h(S_n)\bigr]\xrightarrow{n\to\infty}\mathsf{E}\bigl[h(S)\bigr]$$

が証明できた． □

G.3　中心極限定理の証明：特性関数を利用

補題 G.7. $n\in\mathbb{N}$ とする．0 を含む開区間 $I\subset\mathbb{R}$ で関数 h が n 回連続微分可能であるとき

$$h(x)=\sum_{j=1}^{n-1}\frac{h^{(j)}(0)}{j!}x^j+\frac{1}{(n-1)!}\int_0^x(x-t)^{n-1}h^{(n)}(t)\,\mathrm{d}t \tag{G.16}$$

が成り立つ．ただし $h^{(n)}h(x)=\dfrac{\mathrm{d}^n}{\mathrm{d}x^n}h(x)\,(n\ge 1)$, $h^{(0)}(x)=h(x)$ である．

Proof. 帰納法で証明する．微積分の基本定理から

$$h(x)-h(0)=\int_0^x h^{(1)}(t)\,\mathrm{d}t$$

となるので, $n=1$ のときは成立する．$n=k$ のとき (G.16) は成立すると仮定する．すると部分積分の公式から

$$\begin{aligned}\frac{1}{(k-1)!}\int_0^x(x-t)^{k-1}h^{(k)}(t)\,t &= \frac{-1}{k!}\int_0^x\left(\frac{d}{dt}(x-t)^k\right)h^{(k)}(t)\,\mathrm{d}t\\ &= -\frac{1}{k!}(x-t)^kh^{(k)}(t)\Bigr|_{t=0}^{t=x}\\ &\quad +\frac{1}{k!}\int_0^x(x-t)^kh^{(k+1)}(t)\,\mathrm{d}t\\ &= \frac{1}{k!}x^kh^{(k)}(0)+\frac{1}{k!}\int_0^x(x-t)^kh^{(k+1)}(t)\,\mathrm{d}t\end{aligned}$$

となり, $n=k+1$ のときも (G.16) は成立することがわかる．よって補題は証明された． □

補題 G.8. $\forall x \in \mathbb{R}$ に対して次の不等式が成り立つ.
(1) $\left|e^{\sqrt{-1}x} - 1\right| \leq |x|$.
(2) $\left|e^{\sqrt{-1}x} - 1 - \sqrt{-1}x\right| \leq \dfrac{1}{2}x^2$.
(3) $\left|e^{\sqrt{-1}x} - 1 - \sqrt{-1}x - \dfrac{1}{2}x^2\right| \leq \dfrac{1}{6}|x|^3$.

Proof. $h(x) = e^{\sqrt{-1}x}$ として補題 G.16 を適用する．あとは $|e^{\sqrt{-1}x}| \leq 1\,(x \in \mathbb{R})$ に注意すればよい． □

補題 G.9. 確率変数 Y は $\mathsf{E}[Y] = 0$, $\mathsf{E}[Y^2] = 1$ をみたすとする．このとき

$$\lim_{n\to\infty} n\left(\mathsf{E}\left[e^{\sqrt{-1}t\frac{Y}{\sqrt{n}}}\right] - 1\right) = -\frac{t^2}{2} \qquad (-\infty < t < \infty)$$

が成り立つ.

Proof. 補題 G.8(3) から

$$\left|e^{\sqrt{-1}t\frac{Y}{\sqrt{n}}} - 1 - \sqrt{-1}t\frac{Y}{\sqrt{n}} + \frac{t^2Y^2}{2n}\right| \leq \frac{t^3}{6n^{3/2}}|Y|^3 \tag{G.17}$$

である．つぎに $\epsilon > 0$ をとる．$|Y| \leq \epsilon t\sqrt{n}$ のとき, (G.17) の右辺の $|Y|^3$ は

$$|Y|^3 = |Y|^2|Y| \leq \epsilon t\sqrt{n}Y^2$$

と評価できるので

$$\begin{aligned}&\left|e^{\sqrt{-1}t\frac{Y}{\sqrt{n}}} - 1 - \sqrt{-1}t\frac{Y}{\sqrt{n}} + \frac{t^2Y^2}{2n}\right|\mathbb{1}\{|Y| \leq \epsilon t\sqrt{n}\}\\ &\qquad \leq \epsilon\frac{t^4}{6n}|Y|^4\mathbb{1}\{|Y| \leq \epsilon t\sqrt{n}\}\end{aligned} \tag{G.18}$$

を得る．一方 $|Y| > \epsilon t\sqrt{n}$ のとき補題 G.8(2) から

$$\begin{aligned}&\left|e^{\sqrt{-1}t\frac{Y}{\sqrt{n}}} - 1 - \sqrt{-1}t\frac{Y}{\sqrt{n}} + \frac{t^2Y^2}{2n}\right|\mathbb{1}\{|Y| > \epsilon t\sqrt{n}\}\\ &\qquad \leq \left\{\left|e^{\sqrt{-1}t\frac{Y}{\sqrt{n}}} - 1 - \sqrt{-1}t\frac{Y}{\sqrt{n}}\right| + \left|\frac{t^2Y^2}{2n}\right|\right\}\mathbb{1}\{|Y| > \epsilon t\sqrt{n}\}\\ &\quad \leq \frac{t^2Y^2}{n}\mathbb{1}\{|Y| > \epsilon t\sqrt{n}\}\end{aligned} \tag{G.19}$$

を得る．(G.18) と (G.18) を合わせると

$$\left|e^{\sqrt{-1}t\frac{Y}{\sqrt{n}}} - 1 - \sqrt{-1}t\frac{Y}{\sqrt{n}} + \frac{t^2Y^2}{2n}\right| \leq \epsilon\frac{t^2}{6n}Y^4 + \frac{t^2Y^2}{n}\mathbb{1}\{|Y| > \epsilon t\sqrt{n}\}$$

となる．$\mathsf{E}[Y]=0,\ \mathsf{E}[Y^2]=1$ を用いると

$$n\left|\mathsf{E}\left[e^{\sqrt{-1}t\frac{Y}{\sqrt{n}}}-1+\frac{t^2}{2n}\right]\right|\le\epsilon\frac{t^2}{4}+t^2\mathsf{E}\big[Y^2\mathbb{1}\{|Y|>\epsilon t\sqrt{n}\}\big]$$

となる．よって

$$\lim_{n\to\infty}n\left|\mathsf{E}\left[e^{\sqrt{-1}t\frac{Y}{\sqrt{n}}}-1+\frac{t^2}{2n}\right]\right|\le\epsilon\frac{t^2}{4}\searrow 0\,(\epsilon\to 0)$$

を得る． □

補題 G.10. 確率変数 Y は $\mathsf{E}[Y]=0,\ \mathsf{E}[Y^2]=1$ をみたすとする．このとき

$$\lim_{n\to\infty}\left(\mathsf{E}\left[e^{\sqrt{-1}t\frac{Y}{\sqrt{n}}}\right]\right)^n=\lim_{n\to\infty}\left(1-\frac{t^2}{2n}\right)^n=e^{-t^2/2},\qquad(-\infty<t<\infty)$$

が成り立つ．

Proof. 笠原 (2010, pp.162-163) を借用． □

中心極限定理の証明: $j=1,2,\dots,n$ に対して

$$Y_j=\frac{X_j-\mu}{\sigma}$$

とし

$$S_n=\frac{1}{\sqrt{n}}\left(\frac{X_1-\mu}{\sigma}+\frac{X_2-\mu}{\sigma}+\cdots+\frac{X_n-\mu}{\sigma}\right)=\frac{1}{\sqrt{n}}\big(Y_1+Y_2+\cdots+Y_n\big)$$

とおく．すると補題 G.9 から $t\in\mathbb{R}$ に対して

$$\lim_{n\to\infty}\varphi_n(t)=\lim_{n\to\infty}\mathsf{E}\big[e^{\sqrt{-1}tS_n}\big]=\lim_{n\to\infty}\left(\mathsf{E}\left[e^{\sqrt{-1}t\frac{Y_1}{\sqrt{n}}}\right]\right)^2=e^{-t^2/2}$$

を得る．定理 F.13 から

$$S_n\rightsquigarrow\mathsf{N}(0,\,1)$$

がわかる． □

G.4　中心極限定理の証明: Stein の方法による証明

定義 G.11. $Z \sim \mathsf{N}(0,1)$ と関数 $h : \mathbb{R} \to \mathbb{R}$ は $\mathsf{E}\big[|h(Z)|\big] < \infty$ をみたすとする.

$$\begin{aligned} f_h(x) &= \frac{\mathsf{E}\big[h(Z)\mathbb{1}_{(-\infty,\,x]}(Z)\big] - \mathsf{E}\big[h(Z)\big]\mathsf{Pr}\big(Z \le x\big)}{\phi(x)} \\ &= e^{x^2/2}\int_{-\infty}^{x}\big\{h(y) - \mathsf{E}[h(Z)]\big\}e^{-y^2/2}\,\mathrm{d}y \end{aligned} \tag{G.20}$$

で定義される関数 $f_h : \mathbb{R} \to \mathbb{R}$ を h に付随する Stein 関数 (Stein function) とよぶ. ただし $\phi(x) = \dfrac{1}{\sqrt{2\pi}}e^{-x^2/2}\,(-\infty < x < \infty)$ である.

補題 G.12. $Z \sim \mathsf{N}(0,1)$ とし, 関数 $h : \mathbb{R} \longrightarrow \mathbb{R}$ は $\mathsf{E}\big[h(Z)\big] < \infty$ をみたすとする. さらに関数 $f_h : \mathbb{R} \longrightarrow \mathbb{R}$ は h に付随する Stein 関数とする. 関数 h が連続ならば, f_h は連続微分可能で $\forall x \in \mathbb{R}$ に対して

$$\dot{f}_h(x) - xf_h(x) = h(x) - \mathsf{E}\big[h(Z)\big] \tag{G.21}$$

をみたす. ただし $\dot{f}_h(x) = \dfrac{\mathrm{d}}{\mathrm{d}x}f_h(x)$ である.

Proof. Libniz の公式と微積分学の基本定理より, f_h は微分可能である. 実際

$$\begin{aligned} \dot{f}_h(x) &= \frac{\mathrm{d}}{\mathrm{d}x}\left(e^{x^2/2}\int_{-\infty}^{x}\big\{h(y) - \mathsf{E}\big[h(Z)\big\}e^{-y^2/2}\,\mathrm{d}y\right) \\ &= xe^{x^2/2}\int_{-\infty}^{x}\big\{h(y) - \mathsf{E}\big[h(Z)\big\}e^{-y^2/2}\,\mathrm{d}y + \big\{h(x) - \mathsf{E}\big[h(Z)\big]\big\} \end{aligned}$$

となることから (G.21) をみたすことがわかる. よって f_h は連続微分可能となることもわかる. □

X を任意の確率変数とし, 関数 $h : \mathbb{R} \to \mathbb{R}$ は $\mathsf{E}\big[h(Z)\big] < \infty$ をみたすとする. すると補題 G.12 より

$$\dot{f}_h(X) - Xf_h(X) = h(X) - \mathsf{E}\big[h(Z)\big]$$

となる. 両辺の期待値を X に関してとると

$$\mathsf{E}\big[\dot{f}_h(X) - Xf_h(X)\big] = \mathsf{E}\big[h(X)\big] - \mathsf{E}\big[h(Z)\big]$$

となる. 上の式の絶対値を評価するためには

$$\big|\mathsf{E}\big[\dot{f}_h(X) - Xf_h(X)\big]\big|$$

を評価すればよいことがわかる.

補題 G.13. $\phi(x) = \dfrac{1}{\sqrt{2\pi}} e^{-x^2/2}$, $\Phi(x) = \int_{-\infty}^{x} \phi(t)\,\mathrm{d}t\,(x \in \mathbb{R})$ とする．このとき以下が成り立つ．

(1) $x > 0$ に対して

$$\frac{x}{1+x^2} \leq \frac{1-\Phi(x)}{\phi(x)} \leq \frac{1}{x}.$$

(2) $x \in \mathbb{R}$ に対して

$$-\frac{(1+x^2)\Phi(x)}{\phi(x)} \leq x \leq \frac{(1+x^2)\bigl(1-\Phi(x)\bigr)}{\phi(x)}.$$

Proof. (1) の証明:

$$1-\Phi(x) = \int_x^\infty \phi(t)\,\mathrm{d}t = \int_x^\infty \left\{-\frac{\dot{\phi}\,(t)}{t}\right\}\mathrm{d}t \leq -\frac{1}{x}\int_x^\infty \dot{\phi}\,(t)\,\mathrm{d}t = \frac{\phi(x)}{x}$$

より

$$\frac{1-\Phi(x)}{\phi(x)} \leq \frac{1}{x}$$

がわかる．
(2) の証明: $\forall x > 0$ に対して

$$\frac{x}{1+x^2} \leq \frac{1-\Phi(x)}{\phi(x)} \Leftrightarrow \frac{(1+x^2)\bigl(1-\Phi(x)\bigr)}{\phi(x)} \geq x$$

となる．上の式は明らかに $x \leq 0$ のときも成立する．さらに $\Phi(x) = 1-\Phi(-x)$ と $\phi(x) = \phi(-x)$ に注意すると

$$\frac{(1+x^2)\bigl(1-\Phi(-x)\bigr)}{\phi(x)} \geq -x \Leftrightarrow \frac{(1+x^2)\Phi(x)}{\phi(x)} \geq -x \Leftrightarrow \frac{(1+x^2)\Phi(x)}{\phi(x)} \leq x.$$

□

補題 G.14. X を確率変数とする．$f : \mathbb{R} \to \mathbb{R}$ と $g : \mathbb{R} \to \mathbb{R}$ は非減少関数で $\mathsf{E}\bigl[\bigl|f(X)g(X)\bigr|\bigr] < \infty$ とする．このとき

$$\mathsf{E}\bigl[f(X)g(X)\bigr] \geq \mathsf{E}\bigl[f(X)\bigr]\,\mathsf{E}\bigl[g(X)\bigr]$$

が成り立つ．

Proof. X_1 と X_2 は X と同じ分布をもつ独立な確率変数 (X とも独立) とする. $f(X_1)-f(X_2)$ と $g(X_1)-g(X_2)$ は同符号であるので

$$\bigl\{f(X_1)-f(X_2)\bigr\}\bigl\{g(X_1)-g(X_2)\bigr\}\geq 0$$

となる. X_1, X_2 に関して期待値をとると

$$\mathsf{E}\bigl[\bigl\{f(X_1)-f(X_2)\bigr\}\bigl\{g(X_1)-g(X_2)\bigr\}\bigr]\geq 0$$

なので

$$\mathsf{E}\bigl[f(X_1)g(X_1)\bigr]+\mathsf{E}\bigl[f(X_2)g(X_2)\bigr]=\mathsf{E}\bigl[f(X_1)g(X_2)\bigr]+\mathsf{E}\bigl[f(X_2)g(X_1)\bigr]$$

を得る. X_1 と X_2 は独立に X と同じ分布に従うので

$$\mathsf{E}\bigl[f(X_1)g(X_2)\bigr]=\mathsf{E}\bigl[f(X_2)g(X_1)\bigr]=\mathsf{E}\bigl[f(X)\bigr]\mathsf{E}\bigl[g(X)\bigr]$$

となる. よって

$$\mathsf{E}\bigl[f(X)g(X)\bigr]\geq\mathsf{E}\bigl[f(X)\bigr]\mathsf{E}\bigl[g(X)\bigr]$$

がわかる. □

補題 G.15. $Z\sim\mathsf{N}(0,1)$ とする. 関数 $h:\mathbb{R}\to\mathbb{R}$ はほとんどいたるところで連続微分可能で

$$K:=\sup_{x\in\mathbb{R}}\bigl|\overset{\bullet}{h}(x)\bigr| \tag{G.22}$$

をみたすものとする. このとき $\mathsf{E}\bigl[\bigl|h(Z)\bigr|\bigr]<\infty$ である. さらに h に付随する Stein 関数 f_h は C^2 級であり

$$\sup_{x\in\mathbb{R}}\bigl|\overset{\bullet\bullet}{f}_h(x)\bigr|\leq 2K \tag{G.23}$$

となる.

Proof. 平均値の定理より $0<{}^{\exists}C<\infty$ があって

$$\bigl|h(x)\bigr|\leq C+K|x|$$

となるので

$$\mathsf{E}\bigl[\bigl|h(Z)\bigr|\bigr]\leq C+K\mathsf{E}[|Z|]<\infty$$

を得る.

つぎに補題 G.12 より f_h は連続可能微分になることは Lebeague の有界収束定理を (G.20) に適用するとわかる.

最後に (G.23) を 3 つの段階にわけて証明する.

段階 1: $\forall x \in \mathbb{R}$ に対して

$$h(x) - \mathsf{E}[h(Z)] = \int_{-\infty}^{x} \dot{h}(t)\Phi(t)\,\mathrm{d}t - \int_{-\infty}^{x} \dot{h}(t)\bigl\{1-\Phi(t)\bigr\}\,\mathrm{d}t \tag{G.24}$$

が成り立つことを示す. ただし $\Phi(x) = \int_{-\infty}^{x}\phi(t)\,\mathrm{d}t$, $\phi(x) = \dfrac{1}{\sqrt{2\pi}}e^{-x^2/2}$ である. 実際, 部分積分により

$$\begin{aligned}
&\int_{-\infty}^{x} \dot{h}(t)\Phi(t)\,\mathrm{d}t - \int_{x}^{\infty} \dot{h}(t)\bigl\{1-\Phi(t)\bigr\}\,\mathrm{d}t \\
&\quad = h(t)\Phi(t)\Big|_{-\infty}^{x} - \int_{-\infty}^{x} h(t)\phi(t)\,\mathrm{d}t - h(t)\bigl\{1-\Phi(t)\bigr\}\Big|_{x}^{\infty} - \int_{x}^{\infty} h(t)\phi(t)\,\mathrm{d}t \\
&\quad = h(x)\Phi(x) + h(x)\bigl\{1-\Phi(x)\bigr\} - \int_{-\infty}^{\infty} h(t)\phi(t)\,\mathrm{d}t \\
&\quad = h(x) - \int_{-\infty}^{\infty} h(t)\phi(t)\,\mathrm{d}t \\
&\quad = h(x) - \mathsf{E}[h(Z)]
\end{aligned}$$

からわかる.

段階 2: $\forall x \in \mathbb{R}$ に対して

$$f_h(x) = -\frac{1-\Phi(x)}{\phi(x)}\int_{-\infty}^{x} \dot{h}(t)\Phi(t)\,\mathrm{d}t - \frac{\Phi(x)}{\phi(x)}\int_{x}^{\infty} \dot{h}(t)\bigl\{1-\Phi(t)\bigr\}\,\mathrm{d}t \tag{G.25}$$

が成り立つ. 実際, 部分積分の公式から

$$\begin{aligned}
f_h(x) &= e^{x^2/2}\int_{-\infty}^{x}\bigl\{h(t) - \mathsf{E}\bigl[h(Z)\bigr]\bigr\}e^{-t^2/2}\,\mathrm{d}t \\
&= e^{x^2/2}\Biggl[\bigl\{h(t) - \mathsf{E}[h(Z)]\bigr\}\int_{-\infty}^{t} e^{-y^2/2}\,\mathrm{d}y\Big|_{-\infty}^{x} \\
&\qquad - \int_{-\infty}^{x}\dot{h}(t)\biggl\{\int_{-\infty}^{t} e^{-y^2/2}\,\mathrm{d}y\biggr\}\,\mathrm{d}t\Biggr] \\
&= \sqrt{2\pi}e^{x^2/2}\Biggl[\bigl\{h(x) - \mathsf{E}[h(Z)]\bigr\}\Phi(x) - \int_{-\infty}^{x}\dot{h}(t)\Phi(t)\,\mathrm{d}t\Biggr]
\end{aligned} \tag{G.26}$$

を得る. (G.24) を (G.26) に代入すると

$$
\begin{aligned}
f_h(x) &= \sqrt{2\pi}e^{x^2/2}\biggl\{\biggl(\int_{-\infty}^{x} \overset{\bullet}{h}(t)\Phi(t)\,\mathrm{d}t \\
&\qquad - \int_{-\infty}^{x} \overset{\bullet}{h}(t)\bigl\{1-\Phi(t)\bigr\}\,\mathrm{d}t\biggr)\Phi(x) - \int_{-\infty}^{x} \overset{\bullet}{h}(t)\Phi(t)\,\mathrm{d}t\biggr\} \\
&= \frac{1}{\phi(x)}\biggl\{\bigl\{\Phi(x)-1\bigr\}\int_{-\infty}^{x} \overset{\bullet}{h}(t)\Phi(t)\,\mathrm{d}t - \Phi(x)\int_{x}^{\infty} \overset{\bullet}{h}(t)\bigl\{1-\Phi(t)\bigr\}\,\mathrm{d}t\biggr\}
\end{aligned}
$$

を得る. よって (G.25) は示された.
<u>段階 3</u>: (G.21) から

$$
\begin{aligned}
\overset{\bullet\bullet}{f}_h(x) &= f_h(x) + x\,\overset{\bullet}{f}_h(x) + \overset{\bullet}{h}(x) \\
&= (1+x^2)f_h(x) + x\bigl\{h(x)-\mathsf{E}[h(Z)]\bigr\} + \overset{\bullet}{h}(x)
\end{aligned}
$$

を得る. (G.24) – (G.25) から

$$
\begin{aligned}
&(1+x^2)f_h(x) + x\bigl\{h(x)-\mathsf{E}[h(Z)]\bigr\} \\
&\quad = -\frac{(1+x^2)(1-\Phi(x))}{\phi(x)}\int_{-\infty}^{x} \overset{\bullet}{h}(t)\Phi(t)\,\mathrm{d}t - \frac{(1+x^2)\Phi(x)}{\phi(x)}\int_{x}^{\infty} \overset{\bullet}{h}(t)\bigl\{1-\Phi(t)\bigr\}\,\mathrm{d}t \\
&\qquad + x\int_{-\infty}^{x} \overset{\bullet}{h}(t)\Phi(t)\,\mathrm{d}t - x\int_{x}^{\infty} \overset{\bullet}{h}(t)\bigl\{1-\Phi(t)\bigr\}\,\mathrm{d}t
\end{aligned}
$$

と書ける. 補題 G.13(2) から

$$
\frac{(1+x^2)\bigl(1-\Phi(x)\bigr)}{\phi(x)} - x \geq 0, \quad \text{と} \quad \frac{(1+x^2)\Phi(x)}{\phi(x)} + x \geq 0
$$

となるので

$$
\begin{aligned}
&\Bigl|(1+x^2)f_h(x) + x\bigl\{h(x)-\mathsf{E}[h(Z)]\bigr\}\Bigr| \\
&\qquad \leq K\biggl\{\frac{(1+x^2)\bigl(1-\Phi(x)\bigr)}{\phi(x)} - x\biggr\}\int_{-\infty}^{x}\Phi(t)\,\mathrm{d}t \\
&\qquad\quad + K\biggl\{\frac{(1+x^2)\Phi(x)}{\phi(x)} + x\biggr\}\int_{x}^{\infty}\bigl\{1-\Phi(t)\bigr\}\,\mathrm{d}t
\end{aligned}
\tag{G.27}
$$

を得る. 再度, 部分積分の公式から

$$
\begin{aligned}
&\int_{-\infty}^{x}\Phi(t)\,\mathrm{d}t = x\Phi(x) - \int_{-\infty}^{x} t\phi(t)\,\mathrm{d}t = x\Phi(x) + \int_{-\infty}^{x}\overset{\bullet}{\phi}(t)\,\mathrm{d}t = x\Phi(x)+\phi(x), \\
&\int_{x}^{\infty}\bigl\{1-\Phi(t)\bigr\}\,\mathrm{d}t = -x\bigl\{1-\Phi(x)\bigr\} + \int_{x}^{\infty} t\phi(t)\,\mathrm{d}t = -x\bigl\{1-\Phi(x)\bigr\} - \int_{x}^{\infty}\overset{\bullet}{\phi}(t)\,\mathrm{d}t \\
&\qquad = -x\bigl\{1-\Phi(x)\bigr\} + \phi(x)
\end{aligned}
$$

が成り立つ．これらの式を (G.27) に代入して整理すると

$$
\begin{aligned}
&\Bigl|(1+x^2)f_h(x)+x\bigl\{h(x)-\mathsf{E}[h(Z)]\bigr\}\Bigr| \\
&\quad\le K\left\{\frac{(1+x^2)x\bigl(1-\Phi(x)\bigr)}{\phi(x)}-x\right\}\bigl\{x\Phi(x)+\phi(x)\bigr\} \\
&\qquad+K\left\{\frac{(1+x^2)\Phi(x)}{\phi(x)}+x\right\}\bigl\{-x\bigl\{1-\Phi(x)\bigr\}+\phi(x)\bigr\} \\
&\quad=K\left\{\frac{(1+x^2)\bigl(1-\Phi(x)\bigr)\Phi(x)}{\phi(x)}+(1+x^2)\bigl(1-\Phi(x)\bigr)-x^2\Phi(x)-x\phi(x)\right. \\
&\qquad\left.-\frac{(1+x^2)x\bigl(1-\Phi(x)\bigr)\Phi(x)}{\phi(x)}+(1+x^2)\Phi(x)-x^2\bigl(1-\Phi(x)\bigr)+x\phi(x)\right\} \\
&\quad=K
\end{aligned}
$$

を得る．この式と (G.22) から

$$
\bigl|\ddot{f}_h(x)\bigr|\le\bigl|\dot{h}(x)\bigr|+\Bigl|(1+x^2)f_h(x)+x\bigl\{h(x)-\mathsf{E}[h(Z)]\bigr\}\Bigr|\le 2K
$$

を得る．　□

ここで与えられた $\delta>0$ と $\forall z\in\mathbb{R}$ に対して

$$
h_{\delta,z}(x)=\begin{cases}1 & (x\le z),\\ 0 & (x\ge z+\delta)\\ \dfrac{\delta+z-x}{\delta} & (\text{その他})\end{cases}\qquad(x\in\mathbb{R})\tag{G.28}
$$

と定める．すると

$$
\sup_{x\in\mathbb{R}}\bigl|\dot{h}_{\delta,z}(x)\bigr|=\frac{1}{\delta}
$$

となる．したがって補題 G.13 から

$$
\bigl|\sup_{x\in\mathbb{R}}\ddot{h}_{\delta,z}(x)\bigr|\le\frac{2}{\delta}
$$

となる．よって $\forall x,\,y\in\mathbb{R}$ に対して

$$
\bigl|\dot{h}_{\delta,z}(x)-\dot{h}_{\delta,z}(y)\bigr|\le\int_{\min(x,y)}^{\max(x,y)}\bigl|\ddot{h}_{\delta,z}(t)\bigr|\,\mathrm{d}t\le\frac{2}{\delta}|x-y|
$$

を得る．　□

定理 G.16. $Z\sim\mathsf{N}(0,1)$ と $S_n=\dfrac{1}{\sqrt{n}}\sum_{j=1}^{n}Y_j$ とおく．ただし $Y_1,\,Y_2,\,\ldots,\,Y_n$ は互いに独立で $\mathsf{E}[Y_j]=0$, $\mathsf{E}[Y_j^2]=1$, $\mathsf{E}[|Y_j|^3]<\infty\,(j=1,\,2,\,\ldots,\,n)$ で

ある. このとき

$$\sup_{x\in\mathbb{R}}\left|\mathsf{Pr}\big(S_n \le x\big) - \mathsf{Pr}\big(Z \le x\big)\right| \le 2\sqrt{\frac{3}{n^{3/2}}\sum_{j=1}^{n}\mathsf{E}\big[|Y_j|^3\big]} \xrightarrow{n\to\infty} 0$$

が成り立つ.

Proof. 定数 $\delta > 0$ と $\forall z \in \mathbb{R}$ に対して, $h_{\delta,z}$ は (G.28) で与えられ,

$$f_h(x) := e^{x^2/2}\int_{-\infty}^{x}\Big\{h_{\delta,z}(t) - \mathsf{E}[h_{\delta,z}(Z)]\Big\}e^{-t^2/2}\,\mathrm{d}t$$

とおく. このとき $\forall x \in \mathbb{R}$ に対して

$$\begin{aligned}
\mathsf{Pr}\big(S_n \le x\big) - \mathsf{Pr}\big(Z \le x\big) &= \mathsf{E}\big[h_{\delta,x}(S_n)\big] - \mathsf{E}\big[h_{\delta,x}(Z)\big] + \mathsf{E}\big[h_{\delta,x}(Z)\big] - \mathsf{Pr}\big(Z \le x\big)\\
&\le \Big|\mathsf{E}\big[h_{\delta,x}(S_n)\big] - \mathsf{E}\big[h_{\delta,x}(Z)\big]\Big| + \mathsf{Pr}\big(x \le Z \le x+\delta\big)\\
&= \Big|\mathsf{E}\big[h_{\delta,x}(S_n)\big] - \mathsf{E}\big[h_{\delta,x}(Z)\big]\Big| + \int_{x}^{x+\delta}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}\,\mathrm{d}t\\
&\le \Big|\mathsf{E}\big[h_{\delta,x}(S_n)\big] - \mathsf{E}\big[h_{\delta,x}(Z)\big]\Big| + \delta \qquad \text{(G.29)}
\end{aligned}$$

となる. $\forall x \in \mathbb{R}$ に対して

$$\Big|\mathsf{E}\big[h_{\delta,x}(S_n)\big] - \mathsf{E}\big[h_{\delta,x}(Z)\big)\Big| \quad \le \sum_{j=1}^{n}\frac{3}{\delta n^{3/2}}\mathsf{E}\big[|Y_j|^3\big] \tag{G.30}$$

を示すことができれば

$$\delta = \sqrt{\frac{3}{n^{3/2}}\sum_{j=1}^{n}\mathsf{E}\big[|Y_j|^3\big]} \tag{G.31}$$

として, これを (G.29) – (G.30) に代入して整理すると

$$\mathsf{Pr}\big(S_n \le x\big) - \mathsf{Pr}\big(Z \le x\big) \le 2\sqrt{\frac{3}{n^{3/2}}\sum_{j=1}^{n}\mathsf{E}\big[|Y_j|^3\big]} \tag{G.32}$$

を得る. 同様に

$$\begin{aligned}
\mathsf{Pr}\big(Z \le x\big) - \mathsf{Pr}\big(S_n \le x\big) &\le \mathsf{Pr}\big(Z \le x\big) - \mathsf{E}\big[h_{\delta,x}(Z+\delta)\big]\\
&\quad + \mathsf{E}\big[h_{\delta,x}(Z+\delta)\big] - \mathsf{E}\big[h_{\delta,x}(S_n+\delta)\big]\\
&\le \Big|\mathsf{E}\big[h_{\delta,x}(S_n+\delta)\big] - \mathsf{E}\big[h_{\delta,x}(Z+\delta)\big]\Big|\\
&\quad + \int_{x}^{x+\delta}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}\,\mathrm{d}t\\
&\le \Big|\mathsf{E}\big[h_{\delta,x-\delta}(S_n)\big] - \mathsf{E}\big[h_{\delta,x-\delta}(Z)\big]\Big| + \delta
\end{aligned}$$

となるので, (G.30) – (G.31) と合わせると

$$\mathsf{Pr}\big(Z \le x\big) - \mathsf{Pr}\big(S_n \le x\big) \le 2\sqrt{\frac{3}{n^{3/2}}\sum_{j=1}^{n}\mathsf{E}\big[|Y_j|^3\big]} \tag{G.33}$$

を得る. (G.32) – (G.33) を合わせると

$$\Big|\mathsf{Pr}\big(S_n \le x\big) - \mathsf{Pr}\big(Z \le x\big)\Big| \le 2\sqrt{\frac{3}{n^{3/2}}\sum_{j=1}^{n}\mathsf{E}\big[|Y_j|^3\big]} \tag{G.34}$$

を得る. よって (G.30) を示すことができれば, 定理は証明される.

(G.30) を示すために

$$S_n^{(j)} = S_n - \frac{1}{\sqrt{n}}Y_j \qquad (j = 1,\ 2\ \ldots,\ n$$

とおき, W_j は Y_j と同じ分布をもち, 他のすべてと独立な確率変数とする. すると

$$\mathsf{Var}[S_n] = \sum_{j=1}^{n}\frac{1}{n}\mathsf{Var}[W_j] = 1$$

と

$$\mathsf{E}\big[Y_j f_h(S_n^{(j)})\big] = \mathsf{E}[Y_j]\mathsf{E}\big[f_h\big(S_n^{(j)}\big)\big] = 0$$

となる. すると不等式

$$\big|S_n - S_n^{(j)} + t\big| \le |t| + \frac{1}{\sqrt{n}}|Y_j|$$

と補題 G.13 から

$$
\begin{aligned}
&\Big|\mathsf{E}\big[h_{\delta,x}(S_n)\big]-\mathsf{E}\big[h_{\delta,x}(Z)\big]\Big| = \Big|\mathsf{E}\big[\overset{\bullet}{f}_h(S_n)-S_nf_h(S_n)\big]\Big| \\
&\quad= \left|\sum_{j=1}^n\left\{\mathsf{E}\left[\frac{1}{n}W_j^2\overset{\bullet}{f}_h(S_n)-\frac{Y_j}{\sqrt{n}}f_h(S_n)\right]+\underbrace{\mathsf{E}\left[\frac{Y_j}{\sqrt{n}}f_h(S_n^{(j)})\right]}_{=0}\right\}\right| \\
&\quad= \left|\sum_{j=1}^n\mathsf{E}\left[\frac{1}{n}W_j^2\overset{\bullet}{f}_h(S_n)-\frac{Y_j}{\sqrt{n}}\big\{f_h(S_n)-f_h(S_n^{(j)})\big\}\right]\right| \\
&\quad= \left|\sum_{j=1}^n\mathsf{E}\left[\frac{1}{n}W_j^2\overset{\bullet}{f}_h(S_n)-\frac{Y_j}{\sqrt{n}}\int_0^{Y_j/\sqrt{n}}\overset{\bullet}{f}_h\big(S_n^{(j)}+t\big)\,\mathrm{d}t\right]\right| \\
&\quad= \left|\sum_{j=1}^n\mathsf{E}\left[\frac{1}{\sqrt{n}}W_j\int_0^{W_j/\sqrt{n}}\overset{\bullet}{f}_h(S_n)\,\mathrm{d}t-\frac{W_j}{\sqrt{n}}\int_0^{W_j/\sqrt{n}}\overset{\bullet}{f}_h\big(S_n^{(j)}+t\big)\,\mathrm{d}t\right]\right| \\
&\quad= \left|\sum_{j=1}^n\mathsf{E}\left[\frac{1}{\sqrt{n}}W_j\int_0^{W_j/\sqrt{n}}\left\{\overset{\bullet}{f}_h(S_n)-\overset{\bullet}{f}_h\big(S_n^{(j)}+t\big)\,\mathrm{d}t\right\}\right]\right| \\
&\quad\le \sum_{j=1}^n\mathsf{E}\left[\frac{1}{\sqrt{n}}W_j\int_0^{W_j/\sqrt{n}}\left|\overset{\bullet}{f}_h(S_n)-\overset{\bullet}{f}_h\big(S_n^{(j)}+t\big)\right|\mathrm{d}t\right] \\
&\quad\le \sum_{j=1}^n\mathsf{E}\left[\frac{1}{\sqrt{n}}|W_j|\int_0^{W_j/\sqrt{n}}\frac{2}{\delta}\left|S_n-\big(S_n^{(j)}+t\big)\right|\mathrm{d}t\right]\Bigg] \\
&\quad\le \sum_{j=1}^n\mathsf{E}\left[\frac{1}{\sqrt{n}}|W_j|\int_0^{W_j/\sqrt{n}}\frac{2}{\delta}\left\{|t|+\frac{1}{\sqrt{n}}|Y_j|\right)\right\}\mathrm{d}t\right] \\
&\quad= \sum_{j=1}^n\mathsf{E}\left[\frac{|W_j|}{\delta\sqrt{n}}\left\{\frac{|W_j|^2}{n}+\frac{2}{n}|W_j||Y_j|\right\}\right] \\
&\quad= \sum_{j=1}^n\frac{1}{\delta n^{3/2}}\left\{\mathsf{E}\big[|W_j|^3\big]+2\mathsf{E}\big[|W_j|^2\big]\mathsf{E}\big[|Y_j|\big]\right\} \\
&\quad= \sum_{j=1}^n\frac{1}{\delta n^{3/2}}\left\{\mathsf{E}\big[|W_j|^3\big]+2\mathsf{E}\big[|W_j|^2\big]\mathsf{E}\big[|W_j|\big]\right\} \\
&\quad\le \sum_{j=1}^n\frac{1}{\delta n^{3/2}}\left\{\mathsf{E}\big[|W_j|^3\big]+2\mathsf{E}\big[|W_j|^3\big]\right\} \qquad (\because \text{補題 G.14 より}) \\
&\quad= \sum_{j=1}^n\frac{3}{\delta n^{3/2}}\mathsf{E}\big[|W_j|^3\big] \\
&\quad= \sum_{j=1}^n\frac{3}{\delta n^{3/2}}\mathsf{E}\big[|Y_j|^3\big]
\end{aligned}
$$

を得る．よって定理は示された． □

G.5 Fisher 情報量による証明

清水 (1976, pp.141-145) を借用.

G.6 Berry-Esseen 定理

G.7 章末注釈と参考文献

第H章　条件付き期待値, 条件付き確率, Martingale

第I章　不変測度

I.1　位相の復習

定義 I.1. 空でない集合 $\mathbb{X}$ に次の公理 (i)-(iii) をみたす部分集合族 $\mathcal{T}$ が指定されるとき, $\mathbb{X}$ に**位相**が与えられたという. $\mathcal{T}$ を考慮にいれた集合 $\mathbb{X}$ を**位相空間**と呼び, これを $(\mathbb{X}, \mathcal{T})$ と記すことにする.

(i) $\mathbb{X}, \varnothing \in \mathcal{T}$.

(ii) $\mathcal{T}$ に属する有限個の部分集合 $O_1, O_2, \ldots, O_m$ に対して

$$\bigcap_{j=1}^{m} O_j \in \mathcal{T}.$$

(iii) $\mathcal{T}$ に属する有限または無限個の元からなる部分集合族 $\{O_\lambda;\ \lambda \in \Lambda\}$ に対して

$$\bigcup_{\lambda\in\Lambda} O_\lambda \in \mathcal{T}.$$

この場合, $\mathcal{T}$ に属する部分集合を位相空間 $\mathbb{X}$ の**開集合**という. また開集合の補集合を**閉集合**という.

定義 I.2. $\mathbb{X}$ を位相空間とする.

(i) $\mathbb{X}$ は $\mathbf{T}_1$ であるとは, $\forall x \in \mathbb{X}$ に対して $\{x\}$ は $\mathbb{X}$ の閉部分集合であるときをいう.

(ii) $\mathbb{X}$ が **Hausdorff 空間** であるとは, $\forall x, y \in \mathbb{X}\,(x \neq y)$ に対して互いに排反[1]な開集合 U と V が存在して $x \in U$ と $y \in V$ とできるときをいう.

(iii) $\mathbb{X}$ が**正規**であるとは, $\mathbb{X}$ は $\mathbf{T}_1$ であって, さらに互いに排反な閉集合 A と B に対して互いに排反な開集合 U と V が存在して $A \subset U$ と $B \subset V$ とできるときをいう.

定義 I.3. $\mathbb{X}$ を位相空間とする.

[1] $U \cap V = \varnothing$.

(i) $Y \subset \mathbb{X}$ が**コンパクト**であるとは, Y の任意の開被覆[2]$\{O_\lambda;\ \lambda \in \Lambda\}$ に対して, その中から有限個の元から成る部分集合族 $\{O_1,\ O_2,\ \ldots,\ O_m\}$ $(m \in \mathbb{N})$ であってすでに Y の開被覆になっているものがあることである.

(ii) 位相空間 $\mathbb{X}$ は**局所コンパクト**であるとは, $\forall x \in \mathbb{X}$ に対して, その近傍 U でその閉包 $\overline{U}$ がコンパクトとなるものが存在することである.

定理 I.4. $\mathbb{X}$ を局所コンパクト Hausdorff 空間とする. このとき以下が成り立つ.

(i) $\mathbb{X}$ は正規空間である.

(ii) $\mathbb{X}$ がたかだか可算個の閉集合 C_n $(n = 1,\ 2,\ \ldots)$ の和で表されたならば, 少なくとも 1 つの C_n は内点をもつ集合である.

Proof. (i) $x \in \mathbb{X}$ とし, U を x の開近傍とする. x の開近傍 V を $\overline{V} \subset U$ となるようにとれることを示せばよい. $\mathbb{X}$ は局所コンパクトだから U の閉包 $\overline{U}$ はコンパクトであると仮定しても一般性は失わない. $\overline{U}$ はコンパクトだから閉集合 $\overline{U} \setminus U$ もコンパクトである. $x \in U$ だから $x \notin \overline{U} \setminus U$ である. $\mathbb{X}$ は Hausdorff だから $\overline{U} \setminus U$ の各点 y に対して x の開近傍 $V_y \subset U$ と y の開近傍 W_y をうまくとり $V_y \bigcap W_y = \varnothing$ とすることができる. 作り方から $\{W_y;\ y \in \overline{U} \setminus U\}$ は $\overline{U} \setminus U$ の開被覆である. $\overline{U} \setminus U$ がコンパクトだから, その中の有限個の開集合 $W_{y_1},\ W_{y_2},\ \ldots,\ W_{y_m}$ $(m \in \mathbb{N})$ によって

$$\overline{U} \setminus U \subset \bigcup_{j=1}^{m} W_{y_j}$$

とできる. すると

$$V := \bigcup_{j=1}^{m} V_{y_j}$$

は U に含まれる x の開近傍であって

$$\overline{V} \subset \overline{U} \bigcup \left\{ \mathbb{X} \setminus \left(\bigcup_{j=1}^{m} W_{y_j} \right) \right\} \subset U$$

が成り立つ[3]. これで $\mathbb{X}$ が正規であることが証明された.

---

[2] $Y \subset \bigcup_{\lambda \in \Lambda} O_\lambda$.

[3] $V \subset U$ なので $\overline{V} \subset \overline{U}$ となる. $\bigcup_{j=1}^{m} W_{y_j} \subset U$ を示すために, $\forall x \in \bigcup_{j=1}^{m} W_{y_j}$ をとる. すると $x \in \bigcup_{j=1}^{m} W_{y_j}$ かつ $\overline{U} \setminus U \subset \bigcup_{j=1}^{m} W_{y_j}$ から $x \notin \overline{U} \setminus U$ となる. $x \in \overline{U}$ だから $x \in U$ がわかる.

(ii) $\mathbb{X}$ が

$$\mathbb{X}=\bigcup_{n=1}^{\infty} C_n, \qquad C_n\,(n=1,\,2,\,\ldots) \text{ は閉集合}$$

をみたすとする. 背理法で主張を証明するために, どの C_n も内点を含まないと仮定して矛盾を導こう. $\forall x \in \mathbb{X}$ をとり, その開近傍 U_0 で閉包 $\overline{U}_0$ はコンパクトであるものをとる. C_1 は内点を含まないので $U_0 \setminus \bigl(C_1 \cap U_0\bigr)$ は空でない開集合である[4]. (i) から $\mathbb{X}$ は正規であるから, $U_0 \setminus \bigl(C_1 \cap U_0\bigr)$ の中の 1 点の開近傍 U_1 であって

$$\overline{U}_1 \subset U_0 \setminus \bigl(C_1 \cap U_0\bigr)$$

となるものを選ぶことができる. 次に C_2 が内点を持たないから, $U_1 \setminus \bigl(C_2 \cap U_1\bigr) \neq \varnothing$ であって, $U_1 \setminus \bigl(C_2 \cap U_1\bigr)$ の中の 1 点の開近傍 U_2 を

$$\overline{U}_2 \subset U_1 \setminus \bigl(C_2 \cap U_1\bigr)$$

となるように選ぶことができる. 以下, 同様に操作を繰り返して空でない開集合列 $\{U_0,\,U_1,\,U_2,\,\ldots\}$ を

$$\overline{U}_n \subset U_{n-1} \setminus \bigl(C_n \setminus U_{n-1}\bigr) \qquad (n=1,\,2,\,\ldots)$$

が成り立つように選ぶ. すると

$$\overline{U}_0 \supset \overline{U}_1 \supset \cdots \supset \overline{U}_{n-1} \supset \overline{U}_n \supset \cdots$$

はコンパクト集合 $\overline{U}_0$ の中の閉集合の減少列であるから

$$\bigcap_{n=1}^{\infty} \overline{U}_n \neq \varnothing \tag{I.1}$$

であることがわかる. ところで

$$\overline{U}_n \cap C_n = \varnothing$$

であり, 仮定から

$$\mathbb{X}=\bigcup_{n=1}^{\infty} C_n$$

である. よって

$$\bigcup_{n=1}^{\infty} C_n = \varnothing$$

でなければならない. これは $(I.3)$ と矛盾する. よって (ii) は証明された. □

[4]なぜならば $U_0 \setminus \bigl(C_1 \setminus U_0\bigr) = \varnothing$ ならば C_1 は内点を持つことになるからである.

定義 I.5. $\mathbb{X}$ 上の連続関数 f の**台** (support) を

$$\mathrm{supp}(f) := \overline{\{x \in \mathbb{X};\ f(x) \neq 0\}}$$

によって定める．ただし集合 A に対して $\overline{A}$ はその閉包である．

定理 I.6. $K \subset \mathbb{X}$ をコンパクト部分集合とし, 開集合 U は

$$K \subset U \subset \mathbb{X}$$

をみたすものとする．このとき, 連続関数 $f : \mathbb{X} \to [0, 1]$ で

- $x \in K \ \Rightarrow\ f(x) = 1$,
- $U \supset \mathrm{supp}(f)$ はコンパクト

をみたすものが存在する．

Proof. 正規空間の Urysohn の定理は松村 (1968) にある． □

I.2 Radon 測度

補題 I.7. $\mathbb{X}$ を局所コンパクト空間とする．$x \in \mathbb{X}$ と x の任意の近傍 U に対して, $N \subset U$ をみたす x のコンパクトな近傍 N が存在する．

Proof. 野村 (2018, p.257) の定理 B.1..3 を参照. □

補題 I.8. $\mathbb{X}$ を局所コンパクト空間とする．K を $\mathbb{X}$ のコンパクト集合, U を $\mathbb{X}$ の開集合で

$$K \subset U \subset \mathbb{X}$$

とする．このときある開集合 V でその閉包 $\overline{V}$ がコンパクトであるものが存在して

$$K \subset V \subset \overline{V} \subset U$$

となる．

Proof. 野村 (2018, p.257) の定理 B.1..4 を参照. □

補題 I.9. $\mathbb{X}$ は局所コンパクト Hausdorff 空間とする．閉包がコンパクトである開集合 $U_n\,(n = 1, 2, \ldots)$ が存在して

$$\mathbb{X} = \bigcup_{j=n}^{\infty} U_n$$

と表される．

Proof. コンパクト集合列 $\{K_n\}_{n=1}^{\infty}$ があって

$$\mathbb{X} = \bigcup_{n=1}^{\infty} K_n$$

と書ける．各点 $x \in \mathbb{X}$ に対して, x の開近傍 V_x でその閉包がコンパクトであるものをとる．このとき

$$K_n \subset \bigcup_{x \in K_n} V_x \qquad (n = 1, 2, \ldots)$$

となる．K_n はコンパクトだから, $x_{n,1}, x_{n,2}, \ldots, x_{n,k(n)} \in K_n \,(k(n) \in \mathbb{N})$ が存在して

$$K_n \subset \bigcup_{j=1}^{k(n)} V_{x_{n,j}} =: U_n \tag{I.2}$$

とできる．すると各 $U_n \,(n = 1, 2, \ldots)$ は開集合であって

$$\overline{U}_n = \bigcup_{j=1}^{k(n)} \overline{V}_{x_{n,j}}$$

はコンパクト集合である．(I.2) から

$$\mathbb{X} = \bigcup_{n=1}^{\infty} K_n \subset \bigcup_{n=1}^{\infty} U_n \subset \mathbb{X}$$

となるので

$$\mathbb{X} = \bigcup_{n=1}^{\infty} U_n, \qquad U_n \,(n = 1, 2, \ldots) \text{ は開集合}$$

がわかる． □

補題 I.10. $\mathbb{X}$ が第 2 可算 ($\mathcal{T}$ の高々可算個な基底が存在) ならば, 任意の開集合は σ コンパクトである．

Proof. $\{V_n\}_{n=1}^{\infty}$ を $\mathbb{X}$ の位相基底とする．U を $\mathbb{X}$ の開集合とすると各 $x \in U$ に対して x のコンパクト近傍 $_n$ があって

$$N_x \subset U$$

とできる[5]．このとき

$$U = \bigcup_{x \in U} N_x$$

[5]野村 (2018, p.257, 命題 B.3)

である. 各 x に対して番号 $n_x \in \mathbb{N}$ が存在して

$$x \in V_{n_x} \subset N_x^\circ \subset N_x, \qquad \big(N_x^\circ \text{ は } N_x \text{ の内部}\big)$$

となるから $n_x\,(x \in U)$ の内で異なる番号のものをすべて集めたものを M とすると M は高々可算集合であって

$$U = \bigcup_{m \in M} V_m$$

となる. 各 $m \in M$ に対して

$$x_m \subset N_{n_m}$$

となる x_m を一つ選ぶと

$$U = \bigcup_{m \in M} N_{x_m}$$

となるので U は σ コンパクトであることがわかる.

□

注意 I.11. (i) $A \subset \mathbb{X}$ とする. $\overline{A} = \mathbb{X}$ となるとき, A は $\mathbb{X}$ において**稠密**であるという.
(ii) $\mathbb{X}$ が稠密で高々可算個の元からなる部分集合をもつとき, $\mathbb{X}$ は**可分**であるという.
(iii) 可分な距離空間は第 2 可算公理[6]をみたすことが知られている.

位相空間 $\big(\mathbb{X}, \mathcal{T}\big)$ を局所コンパクト Hausdorff 空間で, σ コンパクトとする. $\mathcal{B}(\mathbb{X})$ を Borel σ 加法族とする. すなわち $\mathcal{B}(\mathbb{X}) = \sigma[\mathcal{T}]$ である. $\mathcal{B}(\mathbb{X})$ を定義域とする測度を **Borel 測度**という.

定義 I.12. μ を $\mathbb{X}$ 上の Borel 測度とする. 測度 μ が次の性質をみたすとき, μ は **Radon 測度**と呼ばれる.

(i) すべてのコンパクト集合 $K \subset \mathbb{X}$ に対して

$$\mu(K) < \infty.$$

(ii) 各 $E \subset \mathcal{B}(\mathbb{X})$ に対して

$$\mu(E) = \inf\big\{\mu(U);\, U \text{ は開集合で } E \subset U\big\}.$$

(iii) 各開集合 U に対して

$$\mu(U) = \sup\big\{\mu(K);\, K \text{ はコンパクト集合で } K \subset U\big\}.$$

[6]可算個の元からなる基底をもつこと.

定理 I.13. $\mathbb{X}$ が第 2 可算空間とする．このとき $\mathbb{X}$ 上の Radon 測度 μ で任意のコンパクト集合に対して

$$\mu(K) < \infty$$

となるものは Radon 測度である．

Proof. 野村 (2018, pp.260-263) を参照. □

$\boxed{\text{記号}}$ 台がコンパクトである $\mathbb{X}$ 上の連続関数全体がなすベクトル空間を $C_c(\mathbb{X})$ と記す.

定理 I.14. $\mathbb{X}$ を第 2 可算空間とし, μ を $\mathbb{X}$ 上の Borel 測度で任意のコンパクト集合 K に対して $\mu(K) < \infty$ とする．$1 \leq p < \infty$ に対して $C_c(\mathbb{X})$ は $L^p(\mathbb{X}) := L^p\big(\mathbb{X}, \mathcal{B}(\mathbb{X}), \mu\big)$ において稠密である.

Proof. $\mathbb{X}$ 上の単関数全体のなす集合は $L^p(\mathbb{X})$ で稠密である．したがって $E \subset \mathcal{B}(\mathbb{X})$ に対して指示関数 $\mathbb{1}_E$ を p ノルムで $C_c(\mathbb{X})$ の関数で近似できることを示せたよい.

$E \subset \mathcal{B}(\mathbb{X})$ を Borel 集合とし, $\mu(E) < \infty$ とする．任意の $\epsilon > 0$ をとる．μ は Radon 測度なので

- $E \supset K$ はコンパクトで,
- $E \subset U$ は開集合で,
- $\mu(U \setminus K) < \epsilon$

をみたすものが存在する．Urysohn の定理より $f \in C_c(\mathbb{X})$ で

$$\mathbb{1}_K(x) \leq f(x) \leq \mathbb{1}_U(x) \qquad (x \in \mathbb{X})$$

をみたすものが存在する．これより

$$\|\mathbb{1}_E - f\|_p = \left\{\int_{\mathbb{X}} |\mathbb{1}_E(x) - f(x)|^p \,\mathrm{d}\mu(x)\right\}^{1/p} \leq \big\{\mu(U \setminus K)\big\}^{1/p}$$

となる．よって定理は証明された. □

定義 I.15. $C_c(\mathbb{X})$ 上の線型形式

$$I : C_c(\mathbb{X}) \to \mathbb{C}$$

が**正値**であるとは, $f \geq 0$ である $f \in C_c(\mathbb{X})$ に対して, つねに $I(f) \geq 0$ となることである.

命題 I.16. I を $C_c(\mathbb{X})$ 上の正値線型汎関数とする．各コンパクト集合 $K \subset \mathbb{X}$ に対して，ある定数 $C_K > 0$ が存在して，任意の $f \in C_c(\mathbb{X})$ で $mathrmsupp(f) \subset K$ なるものに対して

$$|I(f)| \leq C_K \|f\|_\infty$$

が成り立つ．ただし

$$\|f\| := \sup_{x \in \mathbb{X}} |f(x)|$$

である．

Proof. 与えられたコンパクト集合 $K \subset \mathbb{X}$ に対して，

$$\phi \in C_c(\mathbb{X}, [0, 1]) := \{f \in C_c(\mathbb{X}); 0 \leq f(x) \leq 1\, (\forall x \in \mathbb{X})\}$$

を選んで

$$\phi(x) = 1 \qquad (\forall x \in K)$$

とできることが Urysohn の補題から保証される．$\mathrm{supp}(f) \subset K$ のとき

$$|f(x)| \leq |f(x)\phi(x)| \leq \sup_{x \in \mathbb{X}} |f(x)|\phi(x) = \|f\|_\infty \phi(x) \qquad (\forall x \in \mathbb{X})$$

となる．したがって

$$\|f\|_\infty \phi(x) - f(x) \geq 0 \quad \text{かつ} \quad \|f\|_\infty \phi(x) + f(x) \geq 0 \qquad (\forall x \in \mathbb{X})$$

となる．よって I は $C_c(\mathbb{X})$ 上で正値なので

$$\begin{aligned} 0 &\leq I\big(\|f\|_\infty \phi - f\big) = \|f\|_\infty I(\phi) - I(f), \\ 0 &\leq I\big(\|f\|_\infty \phi + f\big) = \|f\|_\infty I(\phi) + I(f) \end{aligned}$$

から

$$\big|I(f)\big| \leq I(\phi)\|f\|_\infty$$

を得る． □

$\boxed{\text{記号}}$ U を $\mathbb{X}$ の開集合とし，$f \in C_c(\mathbb{X})$ とする．このとき

$$f \prec U$$

で $0 \leq f(x) \leq 1\, (\forall x \in \mathbb{X})$ かつ $\mathrm{supp}(f) \subset U$ であることを表す．

注意 I.17. $f \prec U$ は

$$0 \le f(x) \le \mathbb{1}_U(x) \qquad (\forall x \in \mathbb{X}) \tag{I.3}$$

より少しだけ強い条件である．なぜならば (I.3) は

$$\mathrm{supp}(f) \subset \overline{U}$$

を意味するからである． □

補題 I.18. $\mathbb{X}$ を局所コンパクト Hausdorff 空間とし，$K \subset U \subset \mathbb{X}$ とする．ただし K はコンパクト集合で U は開集合とする．このときある $f \in C_c(\mathbb{X})$ が存在して，

$$f \prec U \quad \text{かつ} \quad f(x) \ge \mathbb{1}_K(x)\,(\forall x \in \mathbb{X})$$

とできる．

Proof. 他の文献を借用すること． □

命題 I.19. $\mathbb{X}$ を局所コンパクト Hausdorff 空間とし，$K \subset \mathbb{X}$ をコンパクト集合とする．このとき K の有限開被覆 $\{U_i\}_{i=1}^n\,(n \in \mathbb{N})$ に対してある関数列 $g_1, g_2, \ldots, g_n \in C_c(\mathbb{X})$ が存在して，$g_i \prec U_i\,(i = 1, 2, \ldots, n)$ で $\sum_{i=1}^n g_i(x) = 1\,(\forall x \in K)$ とできる．

Proof. $K = \bigcup_{i=1}^n U_i$ なので任意の $x \in K$ に対してある $i \in \{1, 2, \ldots n\}$ が存在して $x \in U_i$ となる．このことと命題 (追記予定) から x のコンパクト近傍 N_x が存在して $N_x \subset U_i$ とできる．

N_x の内部を N_x^o と記すと $\{N_x^o\}_{x \in K}$ は K の開被覆となる．K はコンパクトだから有限集合 $\{x_1, x_2, \ldots, x_m\} \subset K\,(m \in \mathbb{N})$ が存在して

$$K \subset \bigcup_{i=1}^m N_{x_i}^o$$

となる．各 $i \in \{1, 2, \ldots, n\}$ に対して

$$A_i := \{j \in \{1, 2, \ldots, m\};\, N_{x_j} \subset U_i\}, \quad F_i := \bigcup_{j \in A_i} N_{x_j}$$

と定める．すると F_i はコンパクト集合で $F_i \subset U_i$ である．ここで Urysohn の補題を用いるとある $h_i \in C_c(\mathbb{X})$ が存在して

$$h_i \prec U_i \quad \text{かつ} \quad h_i(x) \ge \mathbb{1}_K(x)\,(\forall x \in \mathbb{X})$$

とできる．定め方から

$$\sum_{i=1}^{n} h_i(x) \geq \mathbb{1}_K(x) \qquad (\forall x \in \mathbb{X}) \tag{I.4}$$

となる．なぜならば

$$\begin{aligned} x \in K &\Rightarrow x \in N_{x_j}\,(\text{ある } j \in \{1, 2, \ldots, m\}) \\ &\Rightarrow x \in F_i\,(j \in A_i) \\ &\Rightarrow \mathbb{1}_F(x) \leq h_i(x) \end{aligned}$$

からわかる．ここで

$$U := \{x \in \mathbb{X}; \sum_{i=1}^{n} h_i(x) > 0\}$$

とおく．関数 $h_i\,(i = 1, 2, \ldots, n)$ は連続関数なので U は開集合であり, (I.4) から $K \subset U$ である．再度 Urysohn の補題からある $f \in C_c(\mathbb{X})$ が存在して

$$f \prec U \quad \text{かつ} \quad f(x) \geq \mathbb{1}_K(x)\,(\forall x \in \mathbb{X}) \tag{I.5}$$

とできる．さらに

$$h_{n+1} := 1 - f \in C_c(\mathbb{X})$$

とおくと

$$\sum_{i=1}^{n+1} h_i(x) = \sum_{i=1}^{n} h_i(x) + h_{n+1}(x) = \sum_{i=1}^{n} h_i(x) + 1 - f(x) > 0 \qquad (\forall x \in \mathbb{X})$$

となる．なぜならば $x \in U$ のとき

$$\sum_{i=1}^{n} h_i(x) > 0 \quad \text{かつ} \quad 1 - f(x) > 0$$

であり, $x \notin U$ のとき

$$\sum_{i=1}^{n} h_i(x) = 0 \quad \text{かつ} \quad 1 - f(x) \geq 1 - \mathbb{1}_U(x) > 0$$

からわかる．ここで各 $i \in \{1, 2, \ldots, n\}$ に対して

$$g_i(x) = \frac{h_i(x)}{\sum_{i=1}^{n+1} h_i(x)} \qquad (x \in \mathbb{X}) \tag{I.6}$$

と定める．明らかに $g_i \in C_c(\mathbb{X})$ であり

$$\mathrm{supp}(g_i) \subset \mathrm{supp}(h_i)$$

である．また $h_i \prec U_i$ なので $g_i \prec U_i$ で

$$x \in K \quad \Longrightarrow \quad f(x) = 1 \qquad (\because f(x) \geq 1_K(x)\,(x \in K)\text{ かつ } f \prec U)$$

である．したがって

$$h_{n+1}(x) = 0 \qquad (\forall x \in K)$$

となる．よって

$$\sum_{i=1}^{n} g_i(x) = \frac{\sum_{i=1}^{n} h_i(x)}{\sum_{i=1}^{n+1} h_i(x)} = 1 \qquad (\forall x \in K)$$

がわかる．よって (I.6) で定義した $\{g_i\}_{i=1}^{n}$ は定理で存在を保証している関数となる． □

定理 I.20. $\mathbb{X}$ を局所コンパクト Hausdorff 空間とし, I を $C_c(\mathbb{X})$ 上の正値線型汎関数とする．このとき $C_c(\mathbb{X})$ 上の任意の正値線型形式 I に対して $\mathbb{X}$ 上の Radon 測度 μ が一意的に存在して, 次式が成り立つ.

$$I(f) = \int_{\mathbb{X}} f(x)\,\mathrm{d}\mu(x) \qquad \bigl(\forall f \in C_c(\mathbb{X})\bigr). \tag{I.7}$$

さらに任意の開集合 $U \subset \mathbb{X}$ に対して

$$\mu(U) = \sup\bigl\{I(f);\, f \in C_c(\mathbb{X}),\, f \prec U\bigr\} \tag{I.8}$$

であり, 任意のコンパクト集合 $K \subset \mathbb{X}$ に対して

$$\mu(K) = \inf\bigl\{I(f);\, f \in C_c(\mathbb{X}),\, f(x) \geq \mathbb{1}_K(x)\,(\forall x \in \mathbb{X})\bigr\} \tag{I.9}$$

となる.

Proof. 証明は ① 一意性と ② 存在の証明に分かれる．まず一意性を示す.
① 一意性の証明: μ を Radon 測度とし

$$I(f) = \int_{\mathbb{X}} f(x)\,\mathrm{d}\mu(x) \qquad (\forall f \in C_c(\mathbb{X}))$$

をみたすとする．$U \subset \mathbb{X}$ を開集合とする．このとき U は (I.8) をみたすことを示そう．$f \prec U$ のとき, $\mathbb{1}_U(x) - f(x) \geq 0\,(\forall x \in \mathbb{X})$ と I の正値性から

$$\begin{aligned}0 \leq I\bigl(\mathbb{1}_U - f\bigr) &= \int_{\mathbb{X}} \bigl\{\mathbb{1}_U(x) - f(x)\bigr\}\,\mathrm{d}\mu(x) = \int_{\mathbb{X}} \mathbb{1}_U(x)\,\mathrm{d}\mu(x) - \int_{\mathbb{X}} f(x)\,\mathrm{d}\mu(x)\\ &= \mu(U) - I(f)\end{aligned}$$

となるので

$$\mu(U) \geq I(f) \tag{I.10}$$

となる．いま μ は U で内正則だから, 任意の正数 $\epsilon > 0$ に対してあるコンパクト集合 $K \subset U$ が存在して

$$\mu(K) \geq \mu(U) - \epsilon$$

とできる．Urysohn の補題からある $f \in C_c(\mathbb{X})$ が存在して

$$f \prec U \quad \text{かつ} \quad f(x) \geq \mathbb{1}_K(x)\,(\forall x \in \mathbb{X})$$

とできる．積分の単調性から

$$I(f) = \int_{\mathbb{X}} f(x)\,\mathrm{d}\mu(x) \geq \int_{\mathbb{X}} \mathbb{1}_K(x)\,\mathrm{d}\mu(x) = \mu(K) \geq \mu(U) - \epsilon \tag{I.11}$$

となる．よって (I.10) と (I.11) を合わせると

$$\mu(U) \geq I(f) \geq \mu(U) - \epsilon$$

を得る．ここで ϵ は任意だったので $\epsilon \downarrow 0$ とすると

$$\mu(U) = \inf\bigl\{I(f);\, f \in C_c(\mathbb{X}),\, f \prec U\bigr\}$$

がわかる．よって $\mu(U)$ は I によって決定されることが示せた．E を任意の Borel 集合とすると μ は E で外正則なので

$$\mu(E) = \inf\bigl\{\mu(U);\, U \supset E,\, U \text{ は開集合}\bigr\}$$

と表現できるので, $\mu(E)$ も I によって定まることがわかる．よって μ の一意性を示すことができた．
② 存在の証明: $U \subset \mathbb{X}$ を開集合とし, $E \subset \mathbb{X}$ を任意の集合とする．まず

$$\rho(U) := \sup\bigl\{I(f);\, f \in C_c(\mathbb{X}),\, f \prec U\bigr\}$$

と定める．明らかに $\rho(\varnothing) = 0$ となる．さらに I の正値性と $f \prec U$ から $I(f) \geq 0$ となる．したがって

$$\rho(U) \geq 0 \qquad (U \text{ は開集合})$$

となる．この ρ を用いて

$$\mu^*(E) := \inf\bigl\{\rho(U):\, E \subset U,\, U \text{ は開集合}\bigr\}$$

と定める. U, V は $\mathbb{X}$ の開集合で $U \subset V$ とすると

$$\rho(U) \leq \rho(V)$$

となる. なぜならば

$$f \prec U \ \Rightarrow \ f \prec V$$

なので

$$\rho(U) = \sup\{I(f);\, f \in C_c(\mathbb{X}),\, f \prec U\} \leq \sup\{I(f);\, f \in C_c(\mathbb{X}),\, f \prec V\} = \rho(V)$$

からわかる. したがって任意の開集合 $U \subset \mathbb{X}$ に対して

$$\mu^*(U) = \inf\{\rho(V);\, U \subset V,\, V \text{ は開集合}\} \geq \rho(U)$$

となる. しかし μ^* の定義から

$$\mu^*(U) \leq \rho(U)$$

なので

$$\mu^*(U) = \rho(U) \qquad (U \text{ は開集合})$$

を得る. ここで

$$\mu := \mu^*\Big|_{\mathcal{B}(\mathbb{X})}$$

と定めると各開集合 $U \subset \mathbb{X}$ に対して

$$\mu(U) = \mu^*(U) = \rho(U)$$

となる. 以上を踏まえて次のような段階を経由して μ の存在の証明をする.

- 第 1 段階: μ^* は外測度である.
- 第 2 段階: $\mathbb{X}$ の開集合は μ^* 可測である. すなわち E を任意の集合で $\mu^*(E) < \infty$ なるものとしたとき

$$\mu^*(E) \geq \mu^*(E \cap U) + \mu^*(E \cap U^c) \tag{I.12}$$

  が成立する.
- 第 3 段階: μ は Borel 測度である.
- 第 4 段階: μ は Radon 測度で (I.8) と (I.9) をみたす.

- 第 5 段階:$\forall f \in C_c(\mathbb{X})$ に対して

$$I(f) = \int_{\mathbb{X}} f(x)\,\mathrm{d}\mu(x)$$

  が成り立つ μ は Borel 可測である.

第 1 段階の証明: まず任意の開集合列 $\{U_j\}_{j=1}^{\infty}$ に対して

$$\rho\biggl(\bigcup_{j=1}^{\infty} U_j\biggr) \leq \sum_{j=1}^{\infty} \rho(U_j) \tag{I.13}$$

を示そう. そのために

$$U := \bigcup_{j=1}^{\infty} U_j, \quad f \in C_c(\mathbb{X}) \text{ で } f \prec U, \quad K := \operatorname{supp}(f)$$

とおく. すると

$$\operatorname{supp}(f) \subset U \;\Rightarrow\; K \subset U$$

となることに注意する. $\{U_j\}_{j=1}^{\infty}$ は K の開被覆なので, ある番号 $n \in \mathbb{N}$ と有限部分被覆 $\{U_j\}_{j=1}^{n}$ が存在する. すなわち

$$K \subset \bigcup_{j=1}^{n} U_j$$

となる. 命題 I.19 から $g_1, g_2, \ldots, g_n \in C_c(\mathbb{X})$ が存在して, 各 $j \in \{1, 2, \ldots, n\}$ に対して

$$g_j \prec U_j$$

で

$$\sum_{j=1}^{n} g_j(x) = 1 \qquad (\forall x \in K)$$

とできる. 明らかに

$$f(x) = f(x)\sum_{j=1}^{n} g_j(x) = \sum_{j=1}^{n} f(x)g_j(x) \qquad (\forall x \in K)$$

である. さらに $fg_j \prec U_j\,(j = 1, 2, \ldots, n)$ である. I の線型性と ρ の定義から

$$I(f) = \sum_{j=1}^{n} I(fg_j) \leq \sum_{j=1}^{n} \rho(U_j) \leq \sum_{j=1}^{\infty} \rho(U_j) \tag{I.14}$$

となる. $f \prec U$ なる $f \in C_c(\mathbb{X})$ に対して, (I.14) は成立するので,

$$\rho(U) = \sup\bigl\{I(f);\, f \in C_c(\mathbb{X}),\, f \prec U\bigr\} \leq \sum_{j=1}^{\infty} \rho(U_j)$$

となる. よって (I.13) が示せた. このことを踏まえて任意の部分集合 $E \subset \mathbb{X}$ に対して

$$E \subset \bigcup_{j=1}^{\infty} U_j \qquad \bigl(U_j\,(j = 1,\, 2,\, \ldots) \text{ は開集合}\bigr)$$

とすると

$$\begin{aligned}
\mu^*(E) &= \inf\bigl\{\rho(U);\, E \subset U,\, U \text{ は開集合}\bigr\} \\
&\leq \inf\left\{\rho\left(\bigcup_{j=1}^{\infty} U_j\right);\, E \subset \bigcup_{j=1}^{\infty} U_j,\, U_j\,(j = 1,\, 2,\, \ldots) \text{ は開集合}\right\} \\
&\leq \inf\left\{\sum_{j=1}^{\infty} \rho\bigl(U_j\bigr);\, E \subset \bigcup_{j=1}^{\infty} U_j,\, U_j\,(j = 1,\, 2,\, \ldots) \text{ は開集合}\right\}
\end{aligned} \tag{I.15}$$

となる. $U_1 = U, , U_j = \varnothing\,(j = 2,\, 3,\, \ldots)$ ととると $\mu^*(E) \geq \mu^*(U) = \mu(U)$ で $\bigcup_{j=1}^{\infty} U_j \supset E$ となるので

$$\mu^*(E) \geq \mu(U) \geq \inf\left\{\sum_{j=1}^{\infty} \rho\bigl(U_j\bigr);\, E \subset \bigcup_{j=1}^{\infty} U_j,\, U_j\,(j = 1,\, 2,\, \ldots) \text{ は開集合}\right\}$$

がわかる. 上の式と (I.15) から

$$\mu^*(E) = \inf\left\{\sum_{j=1}^{\infty} \rho\bigl(U_j\bigr);\, E \subset \bigcup_{j=1}^{\infty} U_j,\, U_j\,(j = 1,\, 2,\, \ldots) \text{ は開集合}\right\} \tag{I.16}$$

が示せた. さらに $\rho(U) \geq 0$ かつ $\rho(\varnothing) = 0$ なので

$$\mu^*(E) \geq 0, \quad \mu^*(\varnothing) = 0$$

となる.

最後に μ^* の劣 σ 加法性を示そう. そのために $\mathbb{X}$ の部分集合列 $\{E_j\}_{j=1}^{\infty}$ をとり

$$E := \bigcup_{j=1}^{\infty} E_j$$

とおく．このとき

$$\mu^*(E) \leq \sum_{j=1}^{\infty} \mu^*(E_j) \tag{I.17}$$

を示せばよい．各 $j \in \{1, 2, \ldots\}$ に対して $\mu^*(E_j) < \infty$ と仮定しても一般性を失わない．さもなければ (I.16) は自明な不等式となる．$\forall\epsilon > 0$ をとる．各 $E_j\,(j = 1, 2, \ldots)$ に対して, 開集合族 $\{U_{jk}\}_{k=1}^{\infty}$ をうまくとると (I.16) から

$$\mu^*(E_j) + \frac{\epsilon}{2^j} \geq \sum_{k=1}^{\infty} \rho(U_{jk}) \tag{I.18}$$

とできることがわかる．

$$E \subset \bigcup_{j=1}^{\infty} E_j \subset \bigcup_{j=1}\left(\bigcup_{k=1}^{\infty} U_{jk}\right)$$

なので

$$\begin{aligned}
\sum_{j=1}^{\infty}\left\{\mu^*(E_j) + \frac{\epsilon}{2^j}\right\} &\geq \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \rho(U_{jk}) \qquad (\because \text{(I.18)}) \\
&\geq \rho\left(\bigcup_{j=1}\left(\bigcup_{k=1}^{\infty} U_{jk}\right)\right) \qquad (\because \text{(I.13)}) \\
&\geq \inf\left\{\sum_{j=1}^{\infty} \rho(V_j);\, E \subset \bigcup_{j=1}^{\infty} V_j,\, V_j\,(j = 1, 2, \ldots)\ \text{は開集合}\right\} \\
&= \mu^*(E)
\end{aligned}$$

となる．よって

$$\sum_{j=1}^{\infty} \mu^*(E_j) + \epsilon \geq \mu^*(E)$$

を得る．$\epsilon > 0$ は任意だったので (I.17) がわかる．このことから μ^* は外測度であることが示せた．

第 2 段階の証明: 開集合 U に対して (I.12) を示せばよい．ただし E は $\mathbb{X}$ の部分集合で $\mu^*(E) < \infty$ なるものであった．

　まず初めに $E \subset \mathbb{X}$ が開集合のときに (I.12) が成立することを示そう．仮定から $E \cap U$ も開集合である．任意の $\epsilon > 0$ をとる．ρ の定義からある $f \in C_c(\mathbb{X})$ で $f \prec E \cap U$ なるものが存在して

$$I(f) > \rho(E \cap U) - \epsilon \tag{I.19}$$

とできる．$\mathrm{supp}(f)$ はコンパクトなので閉集合である．よって

$$V := E \setminus \mathrm{supp}(f)$$

は開集合である．再度 ρ の定義からある $g \in C_c(\mathbb{X})$ で $g \prec V$ なるものが存在して

$$I(g) > \rho(V) - \epsilon \tag{I.20}$$

とできる．このとき

$$\begin{aligned}
&f(x)g(x) = 0, \quad 0 \le f(x) \le 1, \quad 0 \le g(x) \le 1 \quad (\forall x \in \mathbb{X}),\\
&\mathrm{supp}(f) \subset E \cap U \subset E, \quad \mathrm{supp}(g) \subset E
\end{aligned}$$

なので

$$f + g \prec E$$

となる．さらに

$$E \cap U^c \subset V$$

であることがわかる．実際 $\mathrm{supp}(f) \subset E \cap U \subset U$ なので

$$V = E \setminus \mathrm{supp}(f) = E \cap \bigl\{\mathrm{supp}(f)\bigr\}^c \supset E \cap U^c$$

となる．よって μ^* の定義から

$$\rho(V) \ge \inf\bigl\{\rho(V);\, V \supset E \cap U^c,\, V \text{ は開集合}\bigr\} = \mu^*(E \cap U^c)$$

となる．以上のことと I の線型性から

$$\begin{aligned}
\mu^*(E) &= \rho(E) \qquad (\because E \text{ は開集合})\\
&\ge I(f+g) \qquad (\because f + g \prec E)\\
&= I(f) + I(g) \qquad (\because I \text{ の線型性})\\
&> \rho(E \cap U) - \epsilon + \rho(V) - \epsilon \qquad (\because \text{(I.19) と (I.20)})\\
&\ge \mu^*(E \cap U) + \mu^*(E \cap U^c) - 2\epsilon \qquad (\because V \subset E \cap U^c)
\end{aligned}$$

をえる．$\epsilon > 0$ は任意だったので E が開集合のとき (I.12) が成立することが示せた．

次に E は $\mathbb{X}$ の任意の部分集合で $\mu^*(E) < \infty$ なるものとする．$\forall \epsilon > 0$ をとる．μ^* の定義から $V \subset E$ なる開集合 V が存在して

$$\rho(V) < \mu^*(E) + \epsilon$$

とできる．すると開集合 U に対して

$$\begin{aligned}\mu^*(E)+\epsilon &> \rho(V)\\ &= \mu^*(V) \qquad (\because V \text{ は開集合})\\ &\geq \mu^*(V\cap U)+\mu^*(V\cap U^c) \qquad (\because \text{前段階の議論})\\ &\geq \mu^*(E\cap U)+\mu^*(E\cap U^c) \qquad (\because V\cap U\supset E\cap V,\, V\cap U^c\supset E\cap U^c)\end{aligned}$$

となる．$\epsilon>0$ は任意だったので (I.12) が成立することを $\mathbb{X}$ の任意の部分集合 E で $\mu^*(E)<\infty$ なるものに対して示せた．よって第 2 段階の証明ができた．

第 3 段階の証明: Carathéodory の拡張定理から

$$\mathcal{M}(\mu^*)\subset\mathcal{B}(\mathbb{X})$$

となる．ただし

$$\begin{aligned}\mathcal{M} &:= \mathcal{M}(\mu^*)\\ &= \left\{A\subset\mathbb{X};\, \mu^*(E)=\mu^*(E\cap A)+\mu^*(E\cap A^c)\,(E \text{ は } \mathbb{X} \text{ の任意の部分集合})\right\}\end{aligned}$$

とした．さらに

$$\mu=\mu^*\big|_{\mathcal{M}}$$

は Borel 測度となる．よって第 3 段階は証明できた．

第 4 段階の証明:

- (4a): μ は (I.8) をみたすこと，
- (4b): μ は　(I.9) をみたすこと，
- (4c): コンパクト集合 K に対して $\mu(K)<\infty$ であること，
- (4d): μ は開集合上で内正則であること

を順に示していく．

(4a) の証明: ρ と μ^* の定義から μ は外正則で (I.8) をみたす．

(4b) の証明: 次に μ は (I.9) をみたすことを示そう．$K\subset\mathbb{X}$ をコンパクト集合とし，$f\in C_c(\mathbb{X})$ は

$$f(x)\geq \mathbb{1}_K(x) \qquad (\forall x\in\mathbb{X})$$

とする．任意の $\epsilon>0$ をとり

$$U:=\{x\in\mathbb{X};\, f(x)>1-\epsilon\}$$

と定める．f は連続関数なので U は開集合であり，$K \subset U$ である．実際 $\forall x \in K$ に対して $f(x) = 1$ なので U の定義から $x \in U$ となることが直ちにわかる．任意の $g \in C_c(\mathbb{X})$ で $g \prec U$ なるものに対して

$$g(x) \leq \mathbb{1}_U(x) \leq \frac{f(x)}{1-\epsilon} \qquad (\forall x \in \mathbb{X})$$

となる．よって I の線型性と正値性から

$$\frac{1}{1-\epsilon} I(f) - I(g) = I\left(\frac{f}{1-\epsilon} - g\right) \geq 0 \;\Rightarrow\; \frac{1}{1-\epsilon} I(f) \geq I(g) \tag{I.21}$$

となる．よって

$$\begin{aligned}
\mu(K) &\leq \mu(U) && (\because K \subset U) \\
&= \rho(V) && (\because U \text{ は開集合}) \\
&= \inf\bigl\{I(g);\, g \in C_c(\mathbb{X}),\, g \prec U\bigr\} \\
&\leq I(g) \\
&\leq \frac{1}{1-\epsilon} I(f) && (\because \text{(I.21)})
\end{aligned}$$

となる．任意の $\epsilon > 0$ に対して上の不等式は成立するので

$$\mu(K) \leq I(g)$$

がわかる．したがって

$$\mu(K) \leq \inf\bigl\{I(f) :\, f \in C_c(\mathbb{X}),\, f(x) \geq \mathbb{1}_K(x)\,(\forall x \in \mathbb{X})\bigr\}$$

を得る．逆向きの不等式を示すために $\epsilon > 0$ をとる．μ は K で外正則なのである開集合 U で $K \subset U$ なるものが存在して

$$\mu(U) \leq \mu(K) + \epsilon$$

とできる．さらに Urysohn の補題から $f \in C_c(\mathbb{X})$ があって

$$f(x) \geq \mathbb{1}_K(x)\,(\forall x \in \mathbb{X}) \quad \text{かつ} \quad f \prec U$$

とできる．したがって

$$\begin{aligned}
I(f) &\leq \sup\bigl\{I(f);\, f \in C_c(\mathbb{X}),\, f \prec U\bigr\} \\
&= \rho(U) \\
&= \mu(U) && (\because U \text{ は開集合}) \\
&\leq \mu(K) + \epsilon
\end{aligned}$$

を得る. ϵ は任意だったので

$$\mu(K) \geq \inf\bigl\{I(f);\, f \in C_c(\mathbb{X}),\, f(x) \geq \mathbb{1}_K(x)\,(\forall x \in \mathbb{X})\bigr\} \leq I(f) < \infty$$

がわかる. よって (I.9) が示せた.
(4c) の証明: 任意のコンパクト集合 K をとる. Urysohn の補題から $f \in C_c(\mathbb{X}$ で $f(x) \geq \mathbb{1}_K(x)\,(\forall x \in \mathbb{X})$ なるものをとる. すると

$$\begin{aligned}\mu(K) &= \inf\bigl\{I(f);\, f \in C_c(\mathbb{X})\ f(x) \geq \mathbb{1}_K(x)\,(\forall x \in \mathbb{X})\bigr\} \\ &\leq I(f) < \infty\end{aligned}$$

よりわかる.
(4d) の証明: 任意の $\epsilon > 0$ をとる. μ の定義から開集合 U に対して $f \in C_c(\mathbb{X})$ で $f \prec U$ なるものが存在して

$$\mu(U) < I(f) + \epsilon$$

とできる.

$$K := \mathrm{supp}(f)$$

とおくと K はコンパクトとなる. $g \in C_c(\mathbb{X})$ が $g(x) \geq \mathbb{1}_K(x) \geq f(x)\,(\forall x \in \mathbb{X})$ をみたすとき

$$I(g) \geq I(f) > \mu(U) - \epsilon \tag{I.22}$$

となる. しかし (I.22) から

$$\mu(K) = \inf\bigl\{I(g);\, g \in C_c(\mathbb{X}),\, g(x) \geq \mathbb{1}_K(x)\,(\forall x \in \mathbb{X})\bigr\} > \mu(U) - \epsilon$$

となる. よって

$$\sup\bigl\{\mu(K);\, K\ \text{はコンパクトで}\ K \subset U\bigr\} > \mu(U) - \epsilon$$

を得る. $\epsilon > 0$ は任意だったので

$$\mu(U) \leq \sup\bigl\{\mu(K);\, K\ \text{はコンパクトで}\ K \subset U\bigr\}$$

となる. 逆向きの不等号は $K \subset U$ より明らかである. よって (I.8) と (I.9) をみたす Borel 測度 μ は Radon 測度であることがわかる. 以上で第 4 段階の証明はおわり.
第 5 段階の証明: $0 \leq f(x) \leq 1\,(\forall x \in \mathbb{X})$ なる $f \in C_c(\mathbb{X})$ に対して

$$I(f) = \int_{\mathbb{X}} f(x)\,\mathrm{d}\mu(x)$$

を示せばよい．なぜならば $\forall \widetilde{f} \in C_c(\mathbb{X})$ に対して

$$f(x) = \frac{\widetilde{f}(x)}{\int_{\mathbb{X}} f(x)\,\mathrm{d}\mu(x)}$$

と表現すればよい．

$n \in \mathbb{N}$ とし

$$K_i := \left\{x \in \mathbb{X};\ f(x) \geq \frac{i}{n}\right\} (n = 1,\, 2,\, \ldots,\, n), \quad K_0 := \mathrm{supp}(f)$$

とおく．すると $K_n \subset K_{n-1} \subset \cdots \subset K_1 \subset K_0$ である．さらに関数列 $f_1,\, f_2,\, \ldots,\, f_n$ を

$$f_i(x) := \min\left\{\max\left(f(x) - \frac{i-1}{n},\, 0\right),\, \frac{1}{n}\right\} \qquad (i = 1,\, 2,\, \ldots,\, n)$$

と定める．すると

$$\begin{aligned} &x \notin K_{i-1} \Rightarrow f_i(x) = 0, \\ &x \in K_i \setminus K_{i-1} \Rightarrow f_i(x) = \frac{1}{n} \qquad (i = 1,\, 2,\, \ldots,\, n) \end{aligned}$$

となる．

まず

$$f(x) = \sum_{i=1}^{n} f_i(x) \qquad (\forall x \in \mathbb{X}) \tag{I.23}$$

を示そう．実際, $i \in \{1,\, 2,\, \ldots,\, n\}$ を固定すると

$$\begin{aligned} &x \notin K_0 \subset \bigcup_{i=1}^{n} K_i \implies f(x) = 0 = \sum_{i=1}^{n} f_i(x), \\ &x \in K_{j-1} \setminus K_j \,(j = 1,\, 2,\, \ldots,\, n) \implies f(x) \geq \frac{j-1}{n} \text{ かつ } f(x) \leq \frac{j}{n} \end{aligned}$$

なので

$$0 \leq f(x) - \frac{j-1}{n} \leq \frac{1}{n}$$

となる．よって

$$f_i(x) = \begin{cases} \dfrac{1}{n} & (i < j), \\ f(x) - \dfrac{j-1}{n} & (i = j) \\ 0 & (i \geq j+1) \end{cases}$$

となることから

$$\sum_{i=1}^{n} f_i(x) = f_j(x) + \sum_{i=1}^{j-1} \frac{1}{n} = f_j(x) + \frac{j-1}{n} = f(x) - \frac{j-1}{n} + \frac{j-1}{n} = f(x)$$

を得る．よって (I.23) が示せた．

明らかに $f_1,\, f_2,\, \ldots,\, f_n$ は連続である．$x \notin K_{i-1}$ のとき

$$\frac{i-1}{n} \geq f(x) \implies f_i(x) = 0$$

となる．よって

$$\mathrm{supp}(f_i) \subset K_{i-1} \subset \mathrm{supp}(f)$$

から $f_i \in C_c(\mathbb{X})\,(i = 1,\, 2,\, \ldots,\, n)$ となる．よって

$$f_i(x) \leq \frac{1}{n}\mathbb{1}_{K_{i-1}}(x) \qquad (\forall x \in \mathbb{X})$$

を得る．さらに

$$x \in K_i \implies f_i(x) = \frac{1}{n}$$

なので

$$\frac{1}{n}\mathbb{1}_{K_i}(x) \leq f_i(x) \leq \frac{1}{n}\mathbb{1}_{K_{i-1}}(x) \qquad (\forall x \in \mathbb{X}) \tag{I.24}$$

を得る．したがって

$$\frac{1}{n}\mu(K_i) \leq \int_{\mathbb{X}} f_i(x)\,\mathrm{d}\mu(x) \leq \frac{1}{n}\mu(K_{i-1}) \tag{I.25}$$

となる．

さらに $U \subset \mathbb{X}$ は開集合で $K_{i-1} \subset U$ のとき

$$nf_i \prec U$$

となるので, (I.8) から

$$I(f_i) \leq \frac{1}{n}\mu(U)$$

である．μ は K_{i-1} で外正則なので

$$I(f_i) \leq \frac{1}{n}\mu(K_{i-1})$$

である. 一方, (I.24) と (I.9) から

$$I(f_i) \geq \frac{1}{n}\mu(K_i)$$

である. これらを合わせると

$$\frac{1}{n}\mu(K_i) \leq I(f_i) \leq \frac{1}{n}\mu(K_{i-1}) \tag{I.26}$$

である. (I.25) と (I.26) をそれぞれ i に関して 1 から n までの和をとると

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\mu(K_i) \leq \sum_{i=1}^{n}\int_{\mathbb{X}} f_i(x)\,\mathrm{d}\mu(x) = \int_{\mathbb{X}} f(x)\,\mathrm{d}\mu(x) \leq \frac{1}{n}\sum_{i=0}^{n-1}\mu(K_i),\\
&\frac{1}{n}\sum_{i=1}^{n}\mu(K_i) \leq \sum_{i=1}^{n} I(f_i) = I(f) \leq \frac{1}{n}\sum_{i=0}^{n-1}\mu(K_i),
\end{aligned}$$

を得る. したがって

$$\begin{aligned}
\left|I(f) - \int_{\mathbb{X}} f(x)\,\mathrm{d}\mu(x)\right| &\leq \frac{1}{n}\left(\sum_{i=0}^{n-1}\mu(K_i) - \sum_{i=1}^{n}\mu(K_i)\right)\\
&= \frac{1}{n}\Big(\mu(K_0) - \mu(K_n)\Big)\\
&\leq \frac{1}{n}\mu(K_0)
\end{aligned}$$

を得る. K_0 はコンパクトなので $\mu(K_0) < \infty$ である. よって $n \to \infty$ とすると

$$I(f) = \int_{\mathbb{X}} f(x)\,\mathrm{d}\mu(x)$$

を得る. 以上で定理の主張は証明できた. □

Radon 測度が与えられたときに, 別の Radon 測度の構成法を述べる. 構成法に関わる命題を述べるまえに下半連続関数の定義と準備のための命題を述べる.

定義 I.21. $\mathbb{X}$ を位相空間とする. 関数 $f : \mathbb{X} \to \mathbb{R}$ は下半連続であるとは, 任意の $a \in \mathbb{R}$ に対して, $f^{-1}\big((a, \infty)\big) = \big\{x \in \mathbb{X};\ f(x) > a\big\}$ が開集合であるときをいう.

注意 I.22. $\sigma[\{(0, \infty);\ a \in \mathbb{R}\}] = \mathcal{B}(\mathbb{R})$ なので, 任意の下半連続な関数は Borel 可測であることがわかる. U を開集合としたとき関数 $x \mapsto \mathbb{1}_U(x)$

は下半連続となることがわかる．なぜならば

$$\begin{aligned} a > 1 \quad &\Longrightarrow \quad \{x;\, 1_U(x) > a\} = \mathbb{R}, \\ 1 \geq a > 0 \quad &\Longrightarrow \quad \{x;\, 1_U(x) > a\} = U, \\ a \leq 0 \quad &\Longrightarrow \quad \{x;\, 1_U(x) < a\} = \varnothing \end{aligned}$$

からわかる．また f, $g : \mathbb{X} \to [0, \infty)$ は下半連続のとき, $fg : \mathbb{X} \to [0, \infty)$ も下半連続となることがわかる．なぜならば, 各 $n \in \mathbb{N}$ に対して

$$C_n := \{x \in \mathbb{X};\, f(x) > na\}, \quad D_n := \left\{x \in \mathbb{X};\, g(x) > \frac{a}{n}\right\}$$

とおくと f, g は下半連続なので C_n と D_n は開集合となる．さらに

$$\{x \in \mathbb{X};\, f(x)g(x) > a\} = \bigcup_{n=1}^{\infty}\bigl(C_n \cap D_n\bigr) = \left\{\bigcup_{n=1}^{\infty} C_n\right\} \bigcap \left\{\bigcup_{n=1}^{\infty} D_n\right\}$$

と書ける．よって開集合の性質より $\{x \in \mathbb{X};\, f(x)g(x) > a\}$ も開集合となる．よって fg は下半連続となる． □

命題 I.23. μ を局所コンパクト Hausdorff 空間 $\mathbb{X}$ 上の Radon 測度とし, 関数 $f : \mathbb{X} \to [0, \infty)$ を下半連続とする．このとき

$$\int_{\mathbb{X}} f(x)\,\mathrm{d}\mu(x) = \sup\left\{\int_{\mathbb{X}} g(x)\,\mathrm{d}\mu(x);\, g \in C_c(\mathbb{X}),\, 0 \leq g(x) \leq f(x)\,(\forall x \in \mathbb{X})\right\}$$

が成立する．

Proof. 明らかに

$$\int_{\mathbb{X}} f(x)\,\mathrm{d}\mu(x) \geq \sup\left\{\int_{\mathbb{X}} g(x)\,\mathrm{d}\mu(x);\, g \in C_c(\mathbb{X}),\, 0 \leq g(x) \leq f(x)\,(\forall x \in \mathbb{X})\right\}$$

である．逆向きの不等号を示すために $\alpha \in \mathbb{R}$ を

$$a < \int_{\mathbb{X}} f(x)\,\mathrm{d}\mu(x)$$

ととり, ある関数 $g \in C_c(\mathbb{X})$ で

$$0 \leq g(x) \leq f(x)\,(\forall x \in \mathbb{X}) \quad \text{かつ} \quad \alpha < \int_{\mathbb{X}} g(x)\,\mathrm{d}\mu(x)$$

をみたつものを見つける．

j, $n \in \mathbb{N}$ に対して

$$U_{n,j} = f^{-1}\bigl((j2^{-n}, \infty)\bigr) = \bigl\{x \in \mathbb{X};\, f(x) > j2^{-n}\bigr\}$$

とおく．f は下半連続なので $U_{n,j}$ $(\forall j,\, n \in \mathbb{N})$ は開集合となる．いま，各 $n \in \mathbb{N}$ に対して関数 s_n を

$$s_n(x) := 2^{-n} \sum_{j=1}^{2^{2n}} \mathbb{1}_{U_{n,j}}(x) \qquad (x \in \mathbb{X})$$

と定める．すると s_n の定め方から

$$s_n(x) = \begin{cases} j2^{-n} & (x \in U_{n,j} \setminus U_{n,j+1},\, 1 \le j \le 2^{2n}), \\ 2^n & (x \in U_{n,2^{2n}}), \\ 0 & (\text{その他}) \end{cases}$$

と表現できる．$n = j+1,\, j+2,\, \dots,\, 2^n$ に対して

$$U_{n-1,j} = U_{n,2j}$$

となるので各 $x \in \mathbb{X}$ で数列 $\{s_n(x)\}_{n=1}^{\infty}$ は単調増加である．さらに各 $x \in \mathbb{X}$ で

$$s_n(x) \nearrow f(x)\,(n \to \infty)$$

となる．実際，$x \in f^{-1}\big((2^{-n},\, \infty)\big)$ のとき

$$0 \le f(x) - s_n(x) \le 2^{-n}$$

からわかる．

単調収束定理から

$$\lim_{n\to\infty} \int_{\mathbb{X}} s_n(x)\,\mathrm{d}\mu(x) = \int_{\mathbb{X}} \lim_{n\to\infty} s_n(x)\,\mathrm{d}\mu(x) = \int_{\mathbb{X}} f(x)\,\mathrm{d}\mu(x)$$

となる．したがって $m \in \mathbb{N}$ を固定して

$$2^{-m} \sum_{j=1}^{2^{2m}} \mu\big(U_{m,j}\big) = \int_{\mathbb{X}} s_m(x)\,\mathrm{d}\mu(x) > \alpha$$

とできる．さらに $\beta \in \mathbb{R}$ を

$$2^{-m} \sum_{j=1}^{2^{2m}} \mu(U_{m,j}) > \beta > \alpha \tag{I.27}$$

となるようにとる．μ は開集合 $U_{m,j}$ 上で内正則なので，各 $1 \le j \le 2^{2m}$ に対してコンパクト集合 $K_j \subset U_{m,j}$ をうまくとると

$$\mu(K_j) > \mu\big(U_{m,j}\big) - \frac{\beta - \alpha}{2^m}$$

とできる．このとき

$$
\begin{aligned}
2^{-m}\sum_{j=1}^{2^{2m}}\mu\bigl(K_j\bigr) &> 2^{-m}\sum_{j=1}^{2^{2m}}\left(\mu\bigl(U_{m,j}\bigr)-\frac{\beta-\alpha}{2^m}\right)\\
&> \beta-(\beta-\alpha) \qquad (\because \text{(I.27)})\\
&= \alpha
\end{aligned}
\tag{I.28}
$$

となる．

いま $1 \leq j \leq 2^{2m}$ に対して Urysohn の補題を用いるとある $g_j \in C_c(\mathbb{X})$ が存在して

$$
g_j \prec U_{m,j} \qquad g(x) \leq 1_{K_j}(x) \quad (x \in \mathbb{X})
$$

とできる．このとき関数 $g\colon \mathbb{X} \to \mathbb{R}$ を

$$
g(x) := 2^{-m}\sum_{j=1}^{2^{2m}} g_j(x) \qquad (x \in \mathbb{X})
$$

と定めると $g_j \prec U_{m,j}\,(j = 1, 2, \ldots, 2^{2m})$ なので $g(x) \leq s_m(x) \leq f(x)\,(\forall x \in \mathbb{X})$ である．さらに

$$
\begin{aligned}
\int_{\mathbb{X}} g(x)\,\mathrm{d}\mu(x) &= \frac{1}{2^m}\sum_{j=1}^{2^{2m}}\int_{\mathbb{X}} g_j(x)\,\mathrm{d}\mu(x)\\
&\geq \frac{1}{2^m}\sum_{j=1}^{2^{2m}}\int_{\mathbb{X}} \mathbb{1}_{K_j}(x)\,\mathrm{d}\mu(x) \qquad \bigl(\because g_j(x) \geq 1_{K_j}(x)\,(x \in \mathbb{X})\bigr)\\
&\geq \frac{1}{2^m}\mu\bigl(K_j\bigr) > \alpha \qquad \bigl(\because \text{(I.28)}\bigr)
\end{aligned}
$$

となる．α は $\int_{\mathbb{X}} f(x)\,\mathrm{d}\mu(x) > \alpha$ をみたす任意のものだったので $\alpha \to \int_{\mathbb{X}} f(x)\,\mathrm{d}\mu(x)$ とすると

$$
\int_{\mathbb{X}} f(x)\,\mathrm{d}\mu(x) = \sup\left\{\int_{\mathbb{X}} g(x)\,\mathrm{d}\mu(x);\, ginC_c(\mathbb{X}),\, 0 \leq g(x) \leq f(x)\,(x \in \mathbb{X})\right\}
$$

を得る．よって命題は証明された． □

命題 I.24. μ を局所コンパクト Hausdorff 空間 $\mathbb{X}$ 上の Radon 測度とし，$\varphi\colon \mathbb{X} \to (0, \infty)$ を連続関数とする．このとき $\mathbb{X}$ 上の Borel 測度 ν を

$$
\nu(E) := \int_E \varphi(x)\,\mathrm{d}\mu(x) \qquad \bigl(E \in \mathcal{B}(\mathbb{X})\bigr) \tag{I.29}
$$

で定義したとき，ν も $\mathbb{X}$ 上の Radon　測度となる．

Proof. 証明の方針は①-④ である.

① ν は外正則であることを示す. すなわち, 任意の $E \in \mathcal{B}(\mathbb{X})$ に対して

$$\nu(E) = \inf\bigl\{\nu(U); U \text{ は開集合で } U \supset E\bigr\} \tag{I.30}$$

が成立することを示せばよい.

② (I.29) の右辺で定義される $C_c(\mathbb{X})$ 上の汎関数を

$$I(f) = \int_{\mathbb{X}} \varphi(x) f(x) \, \mathrm{d}\mu(x)$$

で定義すると I は線型かつ正値である. したがって定理 I.20 から Radon 測度 ν' が一意的に存在して

$$I(f) = \int_{\mathbb{X}} f(x) \, \mathrm{d}\nu'(x)$$

と書ける. さらに

$$\nu'(E) = \nu(E) \qquad \bigl(\forall E \in \mathcal{B}(\mathbb{X})\bigr) \tag{I.31}$$

を示すことができれば, ν は Radon 測度となる. (I.31) を示すためには, ν と ν' はともに外正則であることに注意すると

$$\nu'(U) = \nu(U) \qquad \bigl(U \text{ は任意の開集合}\bigr) \tag{I.32}$$

であることを示せばよい. なぜならば $\forall E \in \mathcal{B}(\mathbb{X})$ に対して

$$\begin{aligned} \nu'(E) &= \sup\bigl\{\nu'(U); U \text{ は開集合で } U \supset E\bigr\} \\ &= \sup\bigl\{\nu'(U); U \text{ は開集合で } U \supset E\bigr\} \qquad \bigl(\because \text{(I.32)}\bigr) \\ &= \nu(E) \end{aligned}$$

からわかる. さらに (I.32) を示すために

③ $\nu'(U) \geq \nu(U)$

④ $\nu'(U) \leq \nu(U)$

を分けて証明する.

① ν は外正則: $E \subset \mathbb{X}$ を任意の Borel 集合とする. $\nu(E) = \infty$ のとき, 明らかに (I.30) をみたすので, $\nu(E) < \infty$ と仮定しても一般性を失わない. 任意の $\epsilon > 0$ に対して, ある開集合 U で $U \supset E$ なるものが存在して

$$\nu(U \setminus E) < \epsilon \tag{I.33}$$

となることを示せばよい. 各 $n \in \mathbb{N}$ に対して $\mathbb{X}$ の部分集合 W_n を

$$W_n := \varphi^{-1}\bigl((n^{-1}, n)\bigr) = \left\{x \in \mathbb{X};\, \frac{1}{n} < \varphi(x) < n\right\}$$

と定める. すると φ の連続性から W_n は開集合となる. 固定した $n \in \mathbb{N}$ に対して $F \subset W_n$ を Borel を集合としたとき

$$\frac{1}{n}\mathbb{1}_F(x) < \varphi(x)\mathbb{1}_F(x) < n\mathbb{1}_F(x) \qquad (\forall x \in \mathbb{X})$$

となるので

$$\begin{aligned} \frac{1}{n}\mu(F) &= \frac{1}{n}\int_{\mathbb{X}} \mathbb{1}_F(x)\,\mathrm{d}\mu(x) < \int_{\mathbb{X}} \varphi(x)\mathbb{1}_F(x)\,\mathrm{d}\mu(x) < n\int_{\mathbb{X}} \mathbb{1}_F(x)\,\mathrm{d}\mu(x) \\ &= n\nu(F) \end{aligned} \tag{I.34}$$

となる. $F = E \cap W_n$ とおくと上の不等式から

$$\mu(E \cap W_n) < n^2\nu(E \cap W_n) \leq n^2\mu(E) < \infty$$

を得る. ここで μ は外正則なので, ある開集合 $U_n \subset \mathbb{X}$ で $U_n \supset E \cap W_n$ なるものが存在して

$$\mu\Bigl(U_n \setminus \bigl(E \cap W_n\bigr)\Bigr) < \frac{\epsilon}{2^n \cdot n} \tag{I.35}$$

とできる. いま

$$U := \bigcup_{n=1}^{\infty}(U_n \cap W_n)$$

とおく. U_n と $W_n\,(n \in \mathbb{N})$ は開集合なので, U も開集合である. さらに $\bigcup_{n=1}^{\infty} W_n = \mathbb{X}$ なので

$$E = E \cap \mathbb{X} = E\bigcap\left(\bigcup_{n=1}^{\infty} W_n\right) = \bigcup_{n=1}^{\infty}(E \cap W_n)$$

となる. 各 $n \in \mathbb{N}$ に対して $U_n \supset E \cap W_n$ なので

$$E \cap W_n \subset U_n \cap W_n$$

である. これと U の定義から

$$E = \bigcup_{n=1}^{\infty}(E \cap W_n) \subset \bigcup_{n=1}^{\infty}(U_n \cap W_n) = U$$

となる．各 $n \in \mathbb{N}$ に対して $E \cap W_n \subset E$ なので

$$\begin{aligned} U \setminus E &= \left(\bigcup_{n=1}^{\infty}(U_n \cap W_n)\right) \setminus E = \bigcup_{n=1}^{\infty}\Big((U_n \cap W_n) \setminus E\Big) \\ &\subset \bigcup_{n=1}^{\infty}\Big((U_n \cap W_n) \setminus (E \cap W_n)\Big) \end{aligned}$$

となる．さらに各 $n \in \mathbb{N}$ に対して

$$(U_n \cap W_n) \setminus (E \cap W_n) \subset W_n$$

なので

$$\begin{aligned} \nu(U \setminus E) &\leq \nu\left(\bigcup_{n=1}^{\infty}\big\{(U_n \cap W_n) \setminus (E \cap W_n)\big\}\right) \\ &\leq \sum_{n=1}^{\infty}\nu\Big(\big\{(U_n \cap W_n) \setminus (E \cap W_n)\big\}\Big) \\ &\leq \sum_{n=1}^{\infty}n\mu\Big(\big\{(U_n \cap W_n) \setminus (E \cap W_n)\big\}\Big) \\ &\qquad \big(\because \text{(I.34) において } F = (U_n \cap W_n) \setminus (E \cap W_n)\big) \\ &\leq \sum_{n=1}^{\infty}\mu\big(U_n \setminus (E \cap W_n)\big) \\ &\leq \sum_{n=1}^{\infty}n \cdot \frac{\epsilon}{2^n \cdot n} \qquad (\because \text{(I.35)}) \\ &= \epsilon \end{aligned}$$

を得る．よって ν は外正則であることが証明された．

③ $\nu(U) \geq \nu'(U)$ **の証明**: 積分の線型性から $0 < \varphi(x) \leq 1\,(x \in \mathbb{X})$ としても一般性を失わないことを注意する．U は開集合なので関数 $x \mapsto \mathbb{1}_U(x)$ は上半連続である．命題 I.23 から

$$\begin{aligned} \nu'(U) &= \int \mathbb{1}_U(x)\,\mathrm{d}\nu'(x) \\ &= \sup\left\{\int_{\mathbb{X}} g(x)\,\mathrm{d}\nu'(x);\, g \in C_c(\mathbb{X}),\, 0 \leq g(x) \leq \mathbb{1}_U(x)\,(\forall x \in \mathbb{X})\right\} \\ &= \sup\left\{\int_{\mathbb{X}} g(x)\varphi(x)\,\mathrm{d}\mu(x) :\, g \in C_c(\mathbb{X}),\, 0 \leq g(x) \leq \mathbb{1}_U(x)\,(\forall x \in \mathbb{X})\right\} \end{aligned}$$

である．$g \in C_c(\mathbb{X})$ に対して

$$\begin{aligned} &0 \leq g(x) \leq \mathbb{1}_U(x)\,(\forall x \in \mathbb{X}) \\ &\qquad \Longrightarrow 0 \leq g(x)\varphi(x) \leq \mathbb{1}_U(x)\varphi(x)\,(\forall x \in \mathbb{X}) \quad \text{かつ} \quad g\varphi \in C_c(\mathbb{X}) \end{aligned}$$

なので

$$
\begin{aligned}
\nu'(U) &= \sup\left\{\int_{\mathbb{X}} g(x)\varphi(x)\,\mathrm{d}\mu(x) :\ g \in C_c(\mathbb{X}),\ 0 \le g(x) \le \mathbb{1}_U(x)\,(\forall x \in \mathbb{X})\right\} \\
&\le \sup\left\{\int_{\mathbb{X}} g(x)\varphi(x)\,\mathrm{d}\mu(x) :\ g\varphi \in C_c(\mathbb{X}),\ 0 \le g(x)\varphi(x) \le \mathbb{1}_U(x)\varphi(x)\,(\forall x \in \mathbb{X})\right\} \\
&\le \sup\left\{\int_{\mathbb{X}} h(x)\,\mathrm{d}\mu(x) :\ h \in C_c(\mathbb{X}),\ 0 \le h(x) \le \mathbb{1}_U(x)\varphi(x)\,(\forall x \in \mathbb{X})\right\} \\
&\le \sup\left\{\int_{\mathbb{X}} h(x)\,\mathrm{d}\mu(x) :\ h \in C_c(\mathbb{X}),\ 0 \le h(x) \le \mathbb{1}_U(x)\,(\forall x \in \mathbb{X})\right\} \\
&= \nu(U)
\end{aligned}
$$

となる．よって

$$
\nu(U) \le \nu'(U)
$$

が示せた．

④ $\nu(U) \le \nu'(U)$ **の証明**: $h \in C_c(\mathbb{X})$ は

$$
0 \le h(x) \le \mathbb{1}_U(x)\varphi(x) \qquad (\forall x \in \mathbb{X})
$$

をみたすとする．各 $n \in \mathbb{N}$ に対して関数 $h_n : \mathbb{X} \to \mathbb{R}$ を

$$
h_n(x) := \frac{h(x)}{(1/n) + \varphi(x)}
$$

と定める．すると $\dfrac{1}{n} + \varphi(x) > 0\,(\forall x \in \mathbb{X})$ なので $h_n \in C_c(\mathbb{X})$ となる．さらに

$$
0 \le h_n(x) \le \mathbb{1}_U(x) \qquad (\forall x \in \mathbb{X})
$$

であり，各 $x \in \mathbb{X}$ で

$$
h_n(x)\varphi(x) \nearrow h(x) \qquad (n \to \infty)
$$

となる．$\int_{\mathbb{X}} h(x)\,\mu(x) < \infty$ なので優収束定理を用いると

$$
\begin{aligned}
\int_{\mathbb{X}} h(x)\,\mathrm{d}\mu(x) &= \int_{\mathbb{X}} \lim_{n\to\infty} h_n(x)\varphi(x)\,\mathrm{d}\mu(x) && \text{(I.36)} \\
&= \lim_{n\to\infty} \int_{\mathbb{X}} h_n(x)\varphi(x)\,\mathrm{d}\mu(x) \quad (\because \text{優収束定理}) && \text{(I.37)} \\
&\le \sup\left\{\int_{\mathbb{X}} g(x)\varphi(x)\,\mu(x);\ g \in C_c(\mathbb{X}),\ 0 \le g(x) \le \mathbb{1}_U(x)\,(\forall x \in \mathbb{X})\right\} && \text{(I.38)} \\
&\quad \big(\because 0 \le h_n(x) \le \mathbb{1}_U(x)\,(\forall x \in \mathbb{X})\big) && \text{(I.39)}
\end{aligned}
$$

となる．したがって

$$\begin{aligned}
\nu(U) &= \int_{\mathbb{X}} \mathbb{1}_U(x)\varphi(x)\,\mathrm{d}\mu(x) \qquad \bigl(\because \text{(I.29)}\bigr)\\
&= \sup\left\{\int_{\mathbb{X}} h(x)\,\mathrm{d}\mu(x);\ h\in C_c(\mathbb{X}),\ 0\le h(x)\le \mathbb{1}_U(x)\varphi(x)\ (\forall x\in\mathbb{X})\right\}\\
&\qquad \bigl(\because \text{命題 I.23}\bigr)\\
&\le \sup\left\{\int_{\mathbb{X}} g(x)\varphi(x)\,\mathrm{d}\mu(x);\ g\in C_c(\mathbb{X}),\ 0\le g(x)\le \mathbb{1}_U(x)\ (\forall x\in\mathbb{X})\right\}\\
&\qquad \bigl(\because \text{(I.39)}\bigr)\\
&= \int_{\mathbb{X}} \mathbb{1}_U(x)\varphi(x)\,\mathrm{d}\mu(x) \qquad \bigl(\because \text{命題 I.23}\bigr)\\
&= \nu'(U)
\end{aligned}$$

から $\nu(U)\le\nu'(U)$ が証明された．

③ と ④ から (I.32) $\Longrightarrow$ (I.31) となるので，ν は Radon 測度となることがわかる．□

I.3　位相群

定義 I.25. 空でない集合 $\mathbb{G}$ があって，$\mathbb{G}$ の 2 元 a と b に対して，その積と呼ばれる $\mathbb{G}$ の元 ab が定義され，次の性質 (i)-(iii) がみたされるとき，$\mathbb{G}$ は**群** (group) であるという．

(i) $\forall a, b\in\mathbb{G}$ に対して

$$(ab)c = a(bc).$$

(ii) 単位元 e が存在する．すなわち元 $e\in\mathbb{G}$ が存在して

$$ea = ae = a \qquad (\forall a\in\mathbb{G}).$$

(iii) 任意の元 $a\in\mathbb{G}$ に対して，a の**逆元**が存在する．すなわち $\forall a\in\mathbb{G}$ に対して $b\in\mathbb{G}$ が存在して

$$ab = ba = e.$$

注意 I.26. (i) $\mathbb{G}$ の単位元はただ一つであることがわかる．
(ii) 元の $a\in\mathbb{G}$ の逆元はただ一つであることがわかる．元 $a\in\mathbb{G}$ の逆元を a^{-1} と記すことにする．□

定義 I.27. $\mathbb{G}$ を群かつ位相空間とする. $\mathbb{G}$ が**位相群**であるとは, 積をとる写像と逆元を対応させる写像

$$\mathbb{G}\times\mathbb{G}\ni(a,\,b)\mapsto ab\in\mathbb{G},$$
$$\mathbb{G}\ni a\mapsto a^{-1}\in\mathbb{G}$$

が連続であるときをいう.

定義 I.28. $\mathbb{G}$ を Hausdorff 位相空間とする. 各 $a\in\mathbb{G}$ に対して

$$L_a(x):=ax,\quad R_a(x):=xa,\quad \varphi(x)=x^{-1}$$

によって写像 La, R_a, φ を定める. L_a を a による**左移動**, R_a を a による**右移動**という.

定義 I.29. $\mathbb{G}$ を群とする. A, $B\subset\mathbb{G}$ と $x\in\mathbb{G}$ に対して

$$\begin{aligned}xA&:=\{xa;\,a\in A\}=\{R_a(x);\,a\in A\},\\Ax&:=\{ax;\,a\in A\}=\{L_a(x);\,a\in A\},\\A^{-1}&:=\{a^{-1};\,a\in A\},\\AB&:=\{ab;\,a\in A,\,b\in B\}.\end{aligned}$$

命題 I.30. $\mathbb{G}$ を位相群とし, e を $\mathbb{G}$ の単位元とする. このとき以下が成立する.

(1) 任意の $g\in\mathbb{G}$ に対して写像

$$x\mapsto gx;\quad x\mapsto xg;\quad x\mapsto x^{-1};$$

は $\mathbb{G}$ の位相同型写像である.

(2) U を $\mathbb{G}$ の開集合 (閉集合) とし, $g\in\mathbb{G}$ としたとき gU, Ug, U^{-1} は開集合 (閉集合) である.

(3) 単位元 e の任意の近傍 U は e の対称近傍 V(すなわち $e\in V=V^{-1}$) を含む.

(4) 単位元 e の任意の近傍 U に対して e の近傍 V で

$$VV\subset U$$

となるものが存在する.

(5) 単位元 e の任意の近傍 U に対して e の対称近傍 V で

$$VV\subset U$$

なるものが存在する.

(6) H を $\mathbb{G}$ の部分群としたとき閉包 $\overline{H}$ も $\mathbb{G}$ の部分群である．さらに H が $\mathbb{G}$ の正規部分群のとき $\overline{H}$ も $\mathbb{G}$ の正規部分群である．

(7) $\mathbb{G}$ の任意の部分群は閉集合である．

(8) $A, B \subset \mathbb{G}$ はコンパクト集合のとき, AB もコンパクト集合である．

Proof. (1) 写像

$$y \mapsto yg^{-1}; \quad y \mapsto y^{-1}$$

は全単射かつ連続で, 逆写像も連続であることからわかる．
(2) (1) から直ちにわかる．
(3) 写像 $y \mapsto y^{-1}$ とその逆写像は連続であるので U^{-1} は開集合で $e \in U^{-1}$ である．集合 $V := U \cap U^{-1}$ も開集合であり, 定め方から対称で $e \in V \subset U$ となる．
(4)

$$U_1 \times U_2 := \{(x, y) \in \mathbb{G} \times y; xy \in U\}$$

と定める．すると $U_1U_2 = U$ となっていることに注意する．写像

$$\mathbb{G} \times \mathbb{G} \to \mathbb{G}, \quad (x, y) \mapsto xy$$

は連続なので $U_1 \times U_2$ は $\mathbb{G} \times \mathbb{G}$ の開集合で $(e, e) \in U_1 \times U_2$ である．積位相の定義から開集合 $V_1, V_2 \subset \mathbb{G}$ が存在して

$$(e, e) \in V_1 \times V_2 \subset U_1 \times U_2$$

とできる．ここで

$$V := V_1V_2 = \{xy \in \mathbb{G}; x \in V_1, y \in V_2\}$$

と定めると V も開集合となる．実際, $\forall x \in V_1$ に対して (2) から xV_2 も開集合である．さらに

$$V_1V_2 = \bigcup_{x \in V_1} xV_2$$

と表現できるので, 開集合の性質から V_1V_2 は開集合であることがわかる．したがって

$$\begin{aligned} VV &= V_1V_2 \qquad \big(V_2 \ni y = e \text{ ととれば } V \subset V_1.\ \text{同様に } V \subset V_2\big) \\ &\subset U_1U_2 = U \end{aligned}$$

がわかる．

(5) (4) を用いて e の近傍 V で

$$VV \subset U$$

なるものをとる．さらに (3) を用いて e の対称近傍 W で

$$W \subset V$$

なるものをとる．すると

$$WW \subset VV \subset U$$

がわかる．
(6) 明らかに $e \in \overline{H}$ である．写像

$$x \mapsto xy; \quad x \mapsto x^{-1}$$

の連続性と逆写像の連続性から $\overline{H}$ は積と逆元をとる作用について閉じている．よって $\overline{H}$ も部分群となる．さらに H が正規のとき, $g \in \mathbb{G}$ に対して

$$H = gHg^{-1} \subset g\overline{H}g^{-1}$$

となる．(2) から $g\overline{H}g^{-1}$ も閉集合なので

$$\overline{H} \subset g\overline{H}g^{-1}$$

となる．このことから

$$g^{-1}\overline{H}g \subset \overline{H}$$

もわかる．g は任意だったので

$$g^{-1}\overline{H}g = \overline{H}$$

となる．よって $\overline{H}$ も正規である．
(7) H は開集合とすれば, (2) より gH も開集合である．よって

$$\bigcup_{g \in \mathbb{G} \setminus H} gH$$

も開集合である．H の商部分集合は $\mathbb{G}$ の分割でるので

$$H^c = \bigcup_{g \in \mathbb{G} \setminus H} gH$$

となる．よって H は閉集合である．
(8) $A \times B$ はコンパクトで写像

$$\mathbb{G} \times \mathbb{G} \to \mathbb{G}, \quad (x, y) \mapsto xy$$

は連続なので AB もコンパクトである．□

定義 I.31. $\mathbb{G}$ を位相群とする．関数 $f \in C_c(\mathbb{G})$ は**左一様連続**であるとは, 任意の $\epsilon > 0$ に対して $\mathbb{G}$ の単位元 e の近傍 V で

$$\|L_y f - f\|_\infty < \epsilon \qquad (\forall y \in V) \tag{I.40}$$

となるものが存在するときをいう．ただし $\|f\| := \sup_{x\in\mathbb{G}} |f(x)|$ と定めた．さらに **右一様連続**を (I.40) において L_y を R_y にかえたもので定める．

命題 I.32. $\mathbb{G}$ を位相群とし, $f \in C_c(\mathbb{G})$ とする．このとき f は左かつ右一様連続である．

Proof. 左一様連続性の証明を与える．右一様連続性については同様にできる．

$\epsilon > 0$ をとり

$$K := \mathrm{supp}(f)$$

とおく．

まず 各 $x \in K$ とする．このとき G の単位元 e の近傍 U_x で

$$\bigl|f(zx) - f(x)\bigr| < \frac{\epsilon}{2} \qquad (\forall z \in U_x) \tag{I.41}$$

となるものが存在する．これは f の連続性と写像 $\mathbb{G} \ni x \mapsto zx \in \mathbb{G}$ の連続性よりわかる．各 x に対して単位元 e の対称近傍 V_x(すなわち $e \in V_x = V_x{-1}$) で

$$V_x V_x \subset U_x \tag{I.42}$$

なるものをとる．これは命題 I.30(5) から存在することがわかる．明らかに $\{V_x x\}_{x\in K}$ は K の開被覆となる．K はコンパクトなので, ある $n \in \mathbb{N}$ とある $x_1, x_2, \ldots, x_n \in K$ が存在して

$$K \subset \bigcup_{j=1}^{n} V_{x_j} x_j \tag{I.43}$$

とできる．ここで

$$V := \bigcap_{j=1}^{n} V_{x_j}$$

とおく．V の定義と V_{x_j} が e の対称近傍であることから V も e の対称近傍となる．$x \in K$ と $y \in V$ に対して, ある $j \in \{1, 2, \ldots, n\}$ が存在して

$$x x_j^{-1} \in V_{x_j} \in U_{x_j} \tag{I.44}$$

となる．よって

$$
\begin{aligned}
y^{-1}xx_j^{-1} &\in VV_{x_j} \subset V_{x_j}V_{x_j} \\
&\subset U_{x_j} \qquad \bigl(\because \text{(I.42)}\bigr)
\end{aligned}
\tag{I.45}
$$

となる．したがって

$$
\begin{aligned}
\bigl|f(y^{-1}x) - f(x)\bigr| &= \bigl|f(y^{-1}xx_j^{-1}x_j) - f(xx_j^{-1}x_j)\bigr| \\
&\leq \bigl|f(y^{-1}xx_j^{-1}x_j) - f(x_j)\bigr| + \bigl|f(x_j) - f(xx_j^{-1}x_j)\bigr| \\
&< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon
\end{aligned}
$$

を得る．最後の不等号は (I.44) と (I.45) から xx_j^{-1}, $y^{-1}xx_j^{-1} \in U_{x_j}$ であるので (I.41) において $z = xx_j^{-1}$, $y^{-1}xx_j^{-1}$ とおくと

$$
\bigl|f(y^{-1}xx_j^{-1}x_j) - f(x_j)\bigr| < \frac{\epsilon}{2}, \quad \bigl|f(x_j) - f(xx_j^{-1}x_j)\bigr| < \frac{\epsilon}{2}
$$

となることよりわかる．

つぎに $x \notin K$ とする．$y \in V$ に対して $y^{-1}x \not{in} K$ のときは明らかなので, $y^{-1}x \in K$ の場合を考える．すると (I.43) から, ある $j \in \{1, 2, \ldots, n\}$ があって

$$
y^{-1}xx_j^{-1} \in V_{x_j}
$$

となる．したがって

$$
xx_j^{-1} = yy^{-1}xx_j^{-1} \in yV_{x_j} \subset U_{x_j}
$$

となる．なぜならば V の定め方と (I.42) から $y \in V \subset V_{x_j}$ かつ $V_{x_j}V_{x_j} \subset U_{x_j}$ となることからわかる．したがって

$$
\begin{aligned}
\bigl|f(y^{-1}x) - f(x)\bigr| &\leq \bigl|f(y^{-1}x) - f(x_j)\bigr| + \bigl|f(x_j) - f(x)\bigr| \\
&= \bigl|f(y^{-1}xx_j^{-1}x_j) - f(x_j)\bigr| + \bigl|f(x_j) - f(xx_j^{-1}x_j)\bigr| \\
&< \epsilon
\end{aligned}
$$

となる．なぜならば

$$
y^{-1}xx_j^{-1} \in U_{x_j} \quad \text{かつ} \quad xx_j^{-1} \in U_{x_j}
$$

であるから (I.41) からわかる．

以上から命題の主張は証明できた．　□

系 I.33. 任意の $f \in C_c(\mathbb{G})$ と $\epsilon > 0$ に対して $\mathbb{G}$ の単位元 e の近傍 V で

$$
y^{-1}x \in V \quad \text{または} \quad yx^{-1} \in V \Longrightarrow \bigl|f(x) - f(y)\bigr| < \epsilon
$$

となるものが存在する．

Proof. $\epsilon > 0$ とする．f は左かつ右一様連続なので, $\mathbb{G}$ の単位元 e の近傍 U_1 が存在して

$$\left|\bigl(L_z f\bigr)(x) - f(x)\right| < \epsilon \qquad (\forall x \in \mathbb{G}, \forall z \in U_1)$$

とできる．さらに e の近傍 U_2 が存在して

$$\left|\bigl(R_z f\bigr)(x) - f(x)\right| < \epsilon \qquad (\forall x \in \mathbb{G}, \forall z \in U_1)$$

とできる．ここで

$$U := U_1 \cap U_2$$

とおく．$y^{-1}x \in U$ ならば, ある $u \in U$ が存在して

$$x = yu$$

と書ける．よって

$$\left|f(x) - f(y)\right| = \left|f(yu) - f(y)\right| = \left|\bigl(R_u f\bigr)(y) - f(y)\right| < \epsilon$$

となる．同様に $yx^{-1} =: u \in U$ のとき

$$\left|f(x) - f(y)\right| = \left|f(u^{-1}y) - f(y)\right| = \left|\bigl(L_u f\bigr)(y) - f(y)\right| < \epsilon$$

となる． □

系 I.34. $\forall f \in C_c(\mathbb{G})$ と $\forall \epsilon > 0$ に対して $\mathbb{G}$ の単位元 e の対称近傍 V が存在して

$$\sup_{x \in \mathbb{G}} \left|f(xy) - f(yx)\right| < \epsilon \qquad (\forall y \in V)$$

となる．

Proof. f は左かつ右一様連続なので

$$\begin{aligned}
\left|f(xy) - f(yx)\right| &\leq \left|f(xy) - f(x)\right| + \left|f(yx) - f(x)\right| \\
&= \left|\bigl(R_y f\bigr)(x) - f(x)\right| + \left|\bigl(L_{y^{-1}} f\bigr)(x) - f(x)\right| \\
&\leq \left\|R_y f - f\right\|_\infty + \left\|L_{y^{-1}} f - f\right\|_\infty
\end{aligned}$$

からわかる． □

命題 I.35. $\mathbb{G}$ を位相群としたとき, 以下が成り立つ．

(i) 各 $a \in \mathbb{G}$ に対して, 写像 L_a, R_a, φ は　$\mathbb{G}$ から $\mathbb{G}$ への位相同型である．

(ii) $U \subset \mathbb{G}$ を開集合とし, $A \subset \mathbb{G}$ を任意の部分集合とする．このとき AU と UA は開集合である．

(iii) $K_1, K_2 \subset \mathbb{G}$ をコンパクト集合とする．このとき K_1K_2 もコンパクト．

(iv) $\mathbb{G}$ の開部分群は閉部分群である．

Proof. 野村 (2018, pp.136-137) を借用予定. □

I.4　Haar 測度

定義 I.36. (i) $\mathbb{G}$ を位相群とし, μ を $\mathbb{G}$ 上の Borel 測度とする．Borel 測度 μ が**左不変** (**右不変**) であるとは

$$\mu(aE) = \mu(E) \quad \big(\mu(Ea) = \mu(E)\big) \qquad \big(\forall E \in \mathcal{B}(\mathbb{G}), \forall \in a \in \mathbb{G}\big)$$

をみたすときをいう．
(ii) $\mu : \mathcal{B}(\mathbb{G}) \to [0, \infty) \cup \{\infty\}$ を $\mathbb{G}$ 上の Radon 測度で零測度[7]でないとする．μ は $\mathbb{G}$ 上の**左** Haar **測度** (**右** Haar **測度**) であるとは, μ が $\mathbb{G}$ 上の左不変測度 (右不変測度) のときである

注意 I.37. (i) μ が左不変測度であることを積分で表現する．f を $\int_{\mathbb{G}} |f(x)| \,\mathrm{d}\mu(x) < \infty$ または $f \geq 0$ とする．このとき

$$\mu \text{ は左不変測度} \iff \int_{\mathbb{G}} f(a^{-1}x)\,\mathrm{d}\mu(x) = \int_{\mathbb{G}} f(x)\,\mathrm{d}\mu(x) \qquad \big(\forall a \in \mathbb{G}\big).$$

(ii) さらに $\mathbb{G}$ 上の左移動 $L_a\,(a \in \mathbb{G})$ を関数空間に持ち上げ, 同じ記号を流用して,

$$\big(L_a f\big)(x) := f(a^{-1}x) \qquad (x \in \mathbb{G})$$

と定める．この記号を用いると左 Haar 測度 μ は

$$\int_{\mathbb{G}} \big(L_a f\big)(x)\,\mathrm{d}\mu(x) = \int_{\mathbb{G}} f(x)\,\mathrm{d}\mu(x)$$

をみたす．
(iii) また $\mathbb{G}$ 上の左移動 $R_a\,(a \in \mathbb{G})$ を関数空間に持ち上げ, 同じ記号を流用して,

$$\big(R_a f\big)(x) := f(xa)$$

と定める． □

[7] μ は零測度であるとは, すべての $E \in \mathcal{B}(\mathbb{G})$ に対して $\mu(E) = 0$ が成り立つときをいう．

命題 I.38. $\mathbb{G}$ を局所コンパクト位相群とする．このとき以下が成立する．
(i) $\mathbb{G}$ 上の Radon 測度 μ が左 Haar 測度であるための必要十分条件は，$\mathbb{G}$ 上の測度 $\widetilde{\mu}$ を

$$\widetilde{\mu}(E) := \mu(E^{-1}) \qquad (\forall E \in \mathbb{B}(\mathbb{G}))$$

で定めたとき，$\widetilde{\mu}$ が $\mathbb{G}$ 上の右 Haar 測度であることである．
(ii) μ を $\mathbb{G}$ 上の左 Haar 測度とする．$\mathbb{G} \supset U$ は開集合で $U \neq \varnothing$ のとき

$$\mu(U) > 0$$

となる．さらに $f \geq 0$ で $f \neq 0$ としたとき

$$\int_{\mathbb{G}} f(x)\,\mathrm{d}\mu(x) > 0$$

である．
(iii) μ を $\mathbb{G}$ 上の左 Haar 測度とする．このとき

$$\mu(\mathbb{G}) < \infty \iff \mathbb{G}\ \text{はコンパクト}$$

である．

Proof. 証明は De Chiffre (2011, pp.23-24) を借用する予定である．□

定理 I.39. 局所コンパクト群上には正の定数倍を除いて一意的に左 Haar 測度で零測度でないものが存在する．右 Haar 測度についても同様な主張が成り立つ．

Proof. De Chiffre (2011, pp.25-31) を借用予定である．□

I.4.1　Haar 測度の存在の証明のための準備

G を局所コンパクト Hausdorff 空間とする．

定義 I.40. (1)

$$C_c^+ := \left\{ f \in C_c(G);\ f(x) \geq 0\,(\forall x \in G),\ \|f\|_\infty > 0 \right\}.$$

ただし $\|f\|_\infty := \sup_{x\in G} |f(x)|$ である．
(2) $f,\ \varphi \in C_c^+(G)$ に対して

$$C_{f,\varphi} := \left\{ \sum_{j=1}^n c_j;\ n \in \mathbb{N},\ c_1,\ c_2,\ \ldots,\ c_n \geq 0,\ x_1,\ x_2,\ \ldots,\ x_n \in G,\right.$$
$$\left. f(y) \leq \sum_{j=1}^n c_j \big(L_{x_j}\varphi\big)(y)\,(\forall y \in G) \right\}.$$

(3) $\big(f : \varphi\big) = \inf C_{f,\varphi}$.

補題 I.41. $f, \varphi \in C_c^+(G)$ とする．このとき $x_1, x_2, \ldots, x_n \in G$ が存在して

$$f(y) \leq 2\|f\|_\infty \cdot \|\varphi\|_\infty^{-1} \sum_{j=1}^{n} \bigl(L_{x_j}\varphi\bigr)(y) \qquad (\forall y \in G)$$

が成立する．

Proof. $U := \{x \in G; \varphi(x) > \|\varphi\|_\infty/2\}$ とおく．すると U は開集合で $U \neq \varnothing$ である．なぜならば $\|\varphi\|_\infty > 0$ だからである．このことから $\bigl\{xU\bigr\}_{x\in G}$ は関数 f の台 $\mathrm{supp}(f)$ の開被覆となる．台 $\mathrm{supp}(f)$ はコンパクトだから, $\exists n \in \mathbb{N}$ と $x_1, x_2, \ldots, x_n \in G$ が存在して

$$\bigcup_{j=1}^{\infty} \bigl(x_j U\bigr) \subset \mathrm{supp}(f)$$

とできる．$y \in U$ と $x \in G$ に対して $f(x) \geq 0$ と $\varphi(y) > 0$ となるので

$$f(x) \leq 2\|f\|_\infty \cdot \|\varphi\|_\infty^{-1} \varphi(y) \tag{I.46}$$

となる．なぜならば

$$y \in U \ \Rightarrow\ \varphi(y) > \frac{\|\varphi\|_\infty}{2}$$

なので

$$2\|f\|_\infty \cdot \|\varphi\|_\infty^{-1} \varphi(y) > 2\|f\|_\infty \cdot \|\varphi\|_\infty^{-1} \cdot \frac{\|\varphi\|_\infty}{2} = \|f\|_\infty > f(x)$$

からわかる．$x \in x_j U\,(i = 1, 2, \ldots, n)$ に対して

$$x_j^{-1} x \in U$$

となる．(I.46) において $y = x_j^{-1} x$ とおくと

$$f(x) \leq 2\|f\|_\infty \cdot \|\varphi\|_\infty^{-1} \varphi\bigl(x_j^{-1} x\bigr) = 2\|f\|_\infty \cdot \|\varphi\|_\infty^{-1} \bigl(L_{x_j}\varphi\bigr)(x)$$

がわかる． □

注意 I.42. (1) 補題 I.41 は $C_{f,\varphi} \neq \varnothing$ を主張している．すなわち

$$2n\|f\|_\infty \cdot \|\varphi\|_\infty^{-1} \in C_{f,\varphi}$$

である．したがって任意の $f, \varphi \in C_c^+(G)$ に対して $\bigl(f : \varphi\bigr) \in \mathbb{R}$ は well-defined である．

(2) Haar 測度の存在を示すために, $C_c(G)$ 上の正値線型汎関数 I を構成する. すると Riesz の表現定理からその汎関数 I に対応する Radon 測度が存在する. さらに I が左不変であれば, それに対応する Radon 測度も左不変になり, Haar になることがわかる. 汎関数 I を直接構成するのではなく, それの近似汎関数を構成することを目指す. その近似汎関数が写像

$$f \mapsto \bigl(f : \varphi\bigr)$$

である. □

命題 I.43. $f,\, g,\, \varphi \in C_c^+(G)$ と $c > 0$ とする. このとき以下が成立する.

(1) $\bigl(f : \varphi\bigr) = \bigl(L_x f : \varphi\bigr) \quad (\forall x \in G)$.

(2) $\bigl(f + g : \varphi\bigr) \leq \bigl(f : \varphi\bigr) + \bigl(g : \varphi\bigr)$.

(3) $\bigl(cf : \varphi\bigr) = c\bigl(f : \varphi\bigr)$.

(4) $f(x) \leq g(x)\,(\forall x \in G)$ のとき, $\bigl(f : \varphi\bigr) \leq \bigl(g : \varphi\bigr)$.

(5) $\bigl(f : \varphi\bigr) \geq \|f\|_\infty \cdot \|\varphi\|_\infty^{-1}$.

(6) $\bigl(f : g\bigr) \leq \bigl(f : g\bigr)\bigl(g : \varphi\bigr)$.

Proof. (1) $\forall x,\, y \in G$ に対して

$$L_x L_y = L_{xy}$$

である. ある $c_1,\, c_2,\, \ldots,\, c_n \geq 0$ と $x_1,\, x_2,\, \ldots,\, x_n \in G$ が存在して

$$f(y) \leq \sum_{j=1}^n c_j \bigl(L_{x_j}\varphi\bigr)(y) \quad (\forall y \in G)$$

のとき

$$\begin{aligned}\bigl(L_x f\bigr)(y) = f(xy) &\leq \sum_{j=1}^n c_j \bigl(L_{x_j}\varphi\bigr)(xy) \leq \sum_{j=1}^n c_j L_x \bigl(L_{x_j}\varphi\bigr)(y) \\ &= \sum_{j=1}^n c_j \bigl(L_{xx_j}\varphi\bigr)(y)\end{aligned}$$

となることから (1) は証明された.

(2) $c_1,\, c_2,\, \ldots,\, c_n,\, \mathrm{d}_1,\, \mathrm{d}_2,\, \ldots,\, \mathrm{d}_m \geq 0$ と $x_1,\, x_2,\, \ldots,\, x_n,\, y_1,\, y_2,\, \ldots,\, y_m inG$ が存在して

$$f(z) \leq \sum_{j=1}^n c_j \bigl(L_{x_j}\varphi\bigr)(z), \quad g(z) \leq \sum_{j=1}^m d_j \bigl(L_{y_j}\varphi\bigr)(z)$$

であったとする．このとき

$$f(z)+g(z) \leq \sum_{j=1}^{n} c_j \bigl(L_{x_j}\varphi\bigr)(z) + \sum_{j=1}^{m} d_j \bigl(L_{y_j}\varphi\bigr)(z)$$

なので

$$\bigl(f+g : \varphi\bigr) \leq \sum_{j=1}^{n} c_j + \sum_{j=1}^{m} d_j$$

となる．上の不等式の右辺の inf をとると

$$\bigl(f+g : \varphi\bigr) \leq \bigl(f : \varphi\bigr) + \bigl(g : \varphi\bigr)$$

がわかる．(3)

$$f(z) \leq \sum_{j=1}^{n} c_j \bigl(L_{x_j}\varphi\bigr)(z)\,(\forall z \in G) \quad\Longleftrightarrow\quad cf(z) \leq \sum_{j=1}^{n} cc_j \bigl(L_{x_j}\varphi\bigr)(z)\,(\forall z \in G)$$

なので

$$\bigl(cf : \varphi\bigr) \leq c\bigl(f : \varphi\bigr)$$

がわかる．
(4) $f(z) \leq g(z)\,(\forall z \in G)$ とし，$\sum_{j=1}^{n} c_j \leq C_{g,\varphi}$ とする．するとある $x_1, x_2, \ldots, x_n \in G$ が存在して

$$f(z) \leq g(z) \leq \sum_{j=1}^{n} c_j \bigl(L_{x_j}\varphi\bigr)(z) \qquad (\forall z \in G)$$

となるので

$$\sum_{j=1}^{n} c_j \in C_{f,\varphi}$$

となる．よって

$$C_{f,\varphi} \subset C_{g,\varphi}$$

がわかる．上の不等式の両辺の inf をとると

$$\bigl(f : \varphi\bigr) \leq \bigl(g : \varphi\bigr)$$

がわかる．

(5) $\sum_{j=1}^{n} c_j \in C_{f,\varphi}$ とする．このとき $x_1,\, x_2,\, \ldots,\, x_n \in G$ が存在して

$$f(z) \leq \sum_{j=1}^{n} c_j \bigl(L_{x_j}\varphi\bigr)(z) \leq \sum_{j=1}^{n} c_j \|\varphi\|_\infty$$

となる．よって

$$\|f\|_\infty \leq \sum_{j=1}^{n} c_j \|\varphi\|_\infty \quad\Longleftrightarrow\quad \|f\|_\infty \cdot \|\varphi\|_\infty^{-1} \leq \sum_{j=1}^{n} c_j$$

がわかる．上の不等式の右辺の inf をとると

$$\|f\|_\infty \cdot \|\varphi\|_\infty^{-1} \leq \bigl(f : \varphi\bigr)$$

がわかる．

(6) $c_1,\, c_2,\, \ldots,\, c_n,\, \mathrm{d}_1,\, \mathrm{d}_2,\, \ldots,\, \mathrm{d}_m \geq 0$ と $x_1,\, x_2,\, \ldots,\, x_n,\, y_1,\, y_2,\, \ldots,\, y_m \in G$ が存在して

$$f(z) \leq \sum_{j=1}^{n} \bigl(L_{x_j} g\bigr)(z), \tag{I.47}$$

$$g(z) \leq \sum_{j=1}^{m} \bigl(L_{y_j}\varphi\bigr)(z) \qquad (\forall z \in G) \tag{I.48}$$

と仮定する．このとき各 $i \in \{1,\, 2,\, \ldots,\, n\}$ に対して

$$\begin{aligned}
\bigl(L_{x_i} g\bigr)(z) = g\bigl(x_i^{-1} z\bigr) &\leq \sum_{j=1}^{m} d_j \bigl(L_{y_j}\varphi\bigr)\bigl(x_i^{-1} z\bigr) \qquad \bigl(\because \text{(I.48)}\bigr) \\
&= \sum_{j=1}^{m} d_j \bigl(L_{x_i} L_{y_j}\varphi\bigr)(z) = \sum_{j=1}^{m} d_j \bigl(L_{x_i y_j}\varphi\bigr)(z)
\end{aligned} \tag{I.49}$$

となる．よって

$$\begin{aligned}
f(z) &\leq \sum_{i=1}^{n} c_i \bigl(L_{x_i} g\bigr)(z) \qquad \bigl(\because \text{(I.47)}\bigr) \\
&\leq \sum_{i=1}^{n} c_i \left(\sum_{j=1}^{m} d_j \bigl(L_{x_i y_j}\varphi\bigr)(z) \right) \qquad \bigl(\because \text{(I.49)}\bigr) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} c_i d_j \bigl(L_{x_i y_j}\varphi\bigr)(z)
\end{aligned}$$

となる．よって

$$\sum_{i=1} \sum_{j=1}^{m} c_i d_j \in C_{f:\varphi}$$

なので

$$\big(f : \varphi\big) \leq \sum_{i=1}^{n}\sum_{j=1}^{m} c_i d_j \leq \left(\sum_{i=1}^{n} c_i\right)\left(\sum_{j=1}^{m} d_j\right)$$

となる．上の不等式の右辺の inf をとると

$$\big(f : \varphi\big) \leq \big(f : g\big)\big(g : \varphi\big)$$

がわかる． □

定義 I.44. $f_0 \in C_c^+(G)$ を固定する． $\varphi \in C_c^+(G)$ に対して写像 $I_\varphi : C_c^+(G) \to (0, \infty)$ を

$$I_\varphi(f) := \frac{\big(f : \varphi\big)}{\big(f_0 : \varphi\big)}$$

で定義する．

補題 I.45. $f, \varphi \in C_c^+(G)$ に対して

$$\big(f_0 : f\big)^{-1} \leq I_\varphi(f) \leq \big(f : f_0\big)$$

が成立する．

Proof. 命題 I.43(6) から

$$\big(f_0 : \varphi\big) \leq \big(f_0 : f\big)\big(f : \varphi\big)$$

なので

$$I_\varphi(f) := \frac{\big(f : \varphi\big)}{\big(f_0 : f\big)} \geq \frac{\big(f : \varphi\big)}{\big(f_0 : f\big)\big(f : \varphi\big)} = \frac{1}{\big(f_0 : f\big)}$$

がわかる．同様に

$$\big(f : \varphi\big) \leq \big(f : f_0\big)\big(f_0 : \varphi\big)$$

から

$$I_\varphi(f) = \frac{\big(f : \varphi\big)}{\big(f_0 : \varphi\big)} \leq \frac{\big(f : f_0\big)\big(f_0 : \varphi\big)}{\big(f_0 : \varphi\big)} = \big(f : f_0\big)$$

がわかる． □

注意 I.46. 補題 I.45 から，写像 $f \mapsto \big(f : \varphi\big)$ より $I_\varphi(f)$ を考えることにする． □

命題 I.47. 汎関数 I_φ は劣線型で左不変である.

Proof. $f,\, g \in C_c^+(G)$ に対して

$$\begin{aligned} I_\varphi\bigl(f+g\bigr) &= \frac{\bigl(f+g : \varphi\bigr)}{\bigl(f_0 : \varphi\bigr)} \\ &\le \frac{\bigl(f : \varphi\bigr) + \bigl(g : \varphi\bigr)}{\bigl(f_0 : \varphi\bigr)} \qquad \bigl(\because \text{命題 I.43(2)}\bigr) \\ &= I_\varphi(f) + I_\varphi(g) \end{aligned}$$

より I_φ の劣線型性は示せた.
<u>左不変性</u>: $x \in G$ と $f \in C_c^+(G)$ に対して

$$\begin{aligned} I_\varphi\bigl(L_x f\bigr) &= \frac{\bigl(L_x f ; \varphi\bigr)}{\bigl(f_0 : \varphi\bigr)} = \frac{\bigl(f ; \varphi\bigr)}{\bigl(f_0 : \varphi\bigr)} \qquad \bigl(\because \text{命題 I.43(1)}\bigr) \\ &= I_\varphi(f) \end{aligned}$$

よりわかる. □

補題 I.48. $f,\, g \in C_c^+(G)$ と $\forall \epsilon > 0$ に対して G の単位元 e の近傍 V が存在して

$$\mathrm{supp}(\varphi) \subset V$$

とする. このとき

$$I_\varphi(f) + I_\varphi(g) \le I_\varphi(f+g) + \epsilon$$

となる.

Proof. Urysohn の補題からある $h_0 \in C_c^+(G)$ が存在して

$$x \in \mathrm{supp}(f+g) \quad \Longrightarrow \quad h_0(x) = 1$$

とできる. $\delta > 0$ をとり

$$h(x) := f(x) + g(x) + \delta h_0(x) \qquad (\forall x \in G)$$

とし

$$h_1(x) := \begin{cases} \dfrac{f(x)}{h(x)} & (x \in \mathrm{supp}(f)), \\ 0 & (\text{その他}), \end{cases} \qquad h_2(x) := \begin{cases} \dfrac{g(x)}{h(x)} & (x \in \mathrm{supp}(f)), \\ 0 & (\text{その他}), \end{cases}$$

とおく．明らかに $h_1,\, h_2 \in C_c^+(G)$ である．系 (加筆予定) から e の近傍 V が存在して $x,\, y \in G$ に対して

$$y^{-1}x \in V \quad \Longrightarrow \quad \bigl|h_1(x) - h_1(y)\bigr| < \delta \quad \text{かつ} \quad \bigl|h_2(x) - h_2(y)\bigr| < \delta \tag{I.50}$$

とできる．$\varphi \in C_c^+(G)$ をとり

$$\mathrm{supp}(\varphi) \subset V \tag{I.51}$$

となるようにする．ある $c_1,\, c_2,\, \ldots,\, c_n \geq 0$ と $x_1,\, x_2,\, \ldots,\, x_n \in G$ が存在して

$$h(y) \leq \sum_{j=1}^{n} c_j \bigl(L_{x_j}\varphi\bigr)(y) \qquad (\forall y \in G)$$

と仮定する．すなわち

$$\sum_{j=1}^{n} c_j \in C_{h,\varphi}$$

である．このとき $y \in G$ で $y^{-1}x \in \mathrm{supp}(\varphi)$ $(j = 1,\, 2,\, \ldots,\, n)$ なるものに対して

$$\begin{aligned} f(y) = h(y)h_1(y) &\leq \sum_{j=1}^{n} c_j \bigl(L_{x_j}\varphi\bigr)(y) h_1(y) \\ &\leq \sum_{j=1}^{n} c_j \bigl(L_{x_j}\varphi\bigr)(y)\bigl\{h_1(x_j) + \delta\bigr\} \\ &\Bigl(\because \text{(I.50) と (I.51) から } x_j^{-1}y \in V \implies |h_1(x) - h_1(x_j)| < \delta\Bigr) \end{aligned} \tag{I.52}$$

となる．同様に

$$\begin{aligned} g(y) = h(y)h_2(y) &\leq \sum_{j=1}^{n} c_j \bigl(L_{x_j}\varphi\bigr)(y) h_2(y) \\ &\leq \sum_{j=1}^{n} c_j \bigl(L_{x_j}\varphi\bigr)(y)\bigl\{h_2(x_j) + \delta\bigr\} \end{aligned} \tag{I.53}$$

を得る．したがって (I.52) と (I.53) から

$$(f : \varphi) \leq \sum_{j=1}^{n} c_j\bigl\{h_1(x_j) + \delta\bigr\}, \qquad (g : \varphi) \leq \sum_{j=1}^{n} c_j\bigl\{h_2(x_j) + \delta\bigr\} \tag{I.54}$$

となる．h_1 と h_2 の定義から

$$
\begin{aligned}
h_1(y)+h_2(y) &= \begin{cases} \dfrac{f(y)}{h(y)}+\dfrac{g(y)}{h(y)} & (y\in V) \\ 0 & (\text{その他}) \end{cases} \\
&= \begin{cases} \dfrac{f(y)+g(y)}{f(y)+g(y)+\delta h_0(y)} & (y\in V) \\ 0 & (\text{その他}) \end{cases} \\
&\le \begin{cases} 1 & (y\in V) \\ 0 & (\text{その他}) \end{cases}
\end{aligned}
$$

となるので

$$
h_1(y)+h_2(y)\le \mathbb{1}_V(y) \qquad (\forall y\in G) \tag{I.55}
$$

となる．よって (I.54) と (I.55) から

$$
\big(f:\varphi\big)+\big(g:\varphi\big)\le \sum_{j=1}^{n} c_j\big\{h_1(x_j)+h_2(x_j)+2\delta\big\}\le \sum_{j=1}^{n} c_j\big\{1+2\delta\big\}
$$

となる．再度, $\sum_{j=1}^{n} c_J\in C_{h,\varphi}$ に注意すると

$$
\big(f:\varphi\big)+\big(g:\varphi\big)\le(1+2\delta)\big(h:\varphi\big)
$$

となる．上の式の両辺を $\big(f_0:\varphi\big)$ で割ると

$$
\begin{aligned}
I_\varphi(f)+I_\varphi(g) &\le I_\varphi(h)(1+2\delta) \\
&\le (1+2\delta)\big\{I_\varphi(f+g)+\delta I_\varphi(h_0)\big\} \qquad (\because I_\varphi \text{ の劣加法性})
\end{aligned} \tag{I.56}
$$

となる．ここで $\delta>0$ は任意であったので, $\epsilon>0$ に対して十分小さくとると

$$
\delta\big\{2\big(f+g:f_0\big)+1+2\delta\big(h_0:f_0\big)\big\}<\epsilon \tag{I.57}
$$

とできることに注意する．これまでの議論を合わせると

$$
\begin{aligned}
I_\varphi(f)+I_\varphi(g) &\le (1+2\delta)\big\{I_\varphi(f+g)+\delta I_\varphi(h_0)\big\} \qquad \big(\because \text{(I.56)}\big) \\
&\le (1+2\delta)\big\{I_\varphi(f+g)+\delta\big(h_0:f_0\big)\big\} \\
&\quad \big(\because \text{補題 I.45 から } I_\varphi(h_0)\le\big(h_0:f_0\big)\big) \\
&= I_\varphi(g+g)+2\delta I_\varphi(g+g)+\delta(1+2\delta)\big(h_0:f_0\big) \\
&\le I_\varphi(f+g)+2\delta\big(f+g:f_0\big)+\delta(1+2\delta)\big(h_0:f_0\big) \\
&\quad \big(\because \text{補題 I.45 から } I_\varphi(f+g)\le\big(f+g:f_0\big)\ \big) \\
&\le I_\varphi(f+g)+\epsilon \qquad \big(\because \text{(I.57)}\big)
\end{aligned}
$$

を得る．よって補題の主張は証明された．　□

I.4.2 Haar 測度の存在の証明

主張 1 局所コンパクト Hausdorff 位相群上の Haar 測度は存在する

Proof. 各 $f \in C_c^+(G)$ に対して閉区間 $X_f \subset \mathbb{R}$ を

$$X_f := \left[\left(f_0 :\, f\right)^{-1}, \left(f :\, f_0\right)\right]$$

と定める. 補題 I.45 から

$$\left(f_0 :\, f\right)^{-1} \leq I_\varphi(f) \leq \left(f :\, f_0\right)$$

なので $X_f \neq \varnothing$ である. さらに

$$X := \prod_{f \in C_c^f(G)} X_f$$

とおく. $\forall \varphi \in C_c^+(G)$ に対して

$$\left\{I_\varphi(f)\right\}_{f \in C_c^+(G)} \subset \prod_{f \in C_c^f(G)} X_f$$

である. Tychonoff の定理から X はコンパクトである.
　G の単位元 e の近傍族を $\mathcal{U}(e)$ とし, 各 $V \in \mathcal{U}(e)$ に対して

$$K_V := \overline{\left\{I_\varphi \in X;\, \varphi \in C_c^+(G),\, \mathrm{supp} \subset V\right\}}$$

とおいたとき $\left\{K_V\right\}_{V \in \mathcal{U}(e)}$ は有限交叉性を持つことを示そう. そのために $V_1, V_2, \ldots, V_n$ を e の近傍とする. 明らかに

$$e \in \bigcap_{j=1}^n V_j$$

である. したがって $\bigcap_{j=1}^n V_j \neq \varnothing$ である. このことに注意して Urysohn の補題を用いるとある $\varphi \in C_c^+(G)$ が存在して

$$\mathrm{supp}(\varphi) \subset \bigcap_{j=1}^n V_j$$

とできる. このことと K_V の定義から

$$\varphi \in K_{\bigcap_{j=1}^n V_j}$$

となる. K_V の定義から

$$K_{\bigcap_{j=1}^n V_j} \subset \bigcap_{j=1}^n K_{V_j}$$

となる．実際

$$\begin{aligned}\forall I_\varphi \in K_{\bigcap_{j=1}^n V_j} \quad &\Longrightarrow \quad \mathrm{supp}(\varphi) \in \bigcap_{j=1}^n V_j \\ &\Longrightarrow \quad \mathrm{supp}(\varphi) \in V_j \, (j = 1, 2, \ldots, n) \\ &\Longrightarrow \quad I_\varphi \in \bigcap_{j=1}^n K_{V_j}\end{aligned}$$

からわかる．よって

$$\bigcap_{j=1}^n K_{V_j} \neq \varnothing$$

である．よって $\{K_V\}_{V\in\mathcal{U}(e)}$ は有限交叉であることがわかる．

$\{K_V\}_{V\in\mathcal{U}(e)}$ は閉集合族で有限交叉であるので

$$\bigcap_{V\in\mathcal{U}(e)} K_V \neq \varnothing$$

となる．このことから

$$I \in \bigcap_{V\in\mathcal{U}(e)} K_V$$

をとることができる．X における I の任意の近傍を $U_I \subset X$ とすると

$$U_I \bigcap \left(\bigcap_{V\in\mathcal{U}(e)} K_V \right) \neq \varnothing$$

である．言いかえると G の単位元 e の任意の近傍 V, 任意の $f_1, f_2, \ldots, f_n \in C_c^+(G)$ と $\forall\epsilon > 0$ に対して

$$\mathrm{supp}(\varphi) \subset V \quad \text{かつ} \quad \bigl|I(f_i) - I_\varphi(f_i)\bigr| < \epsilon \, (i = 1, 2, \ldots, n) \tag{I.58}$$

とできる．

以下では, $\epsilon > 0$ と $f, g \in C_c^+(G)$ を固定する．各 $x \in G$ に対して $\varphi \in C_c^+(G)$ が存在して

$$\bigl|I(f) - I_\varphi(f)\bigr| < \epsilon \quad \text{かつ} \quad \bigl|I(L_x f) - I_\varphi(L_x f)\bigr| < \epsilon) \tag{I.59}$$

とできる．なぜならば (I.58) において $n = 2, f_1 = f, f_2 = L_x f$ とおけばよい．三角不等式と命題 I.43(1) から

$$\begin{aligned}\bigl|I(f) - I(L_x f)\bigr| &\leq \bigl|I(f) - I_\varphi(f)\bigr| + \bigl|I_\varphi(f) - I(L_x f)\bigr| \qquad \bigl(\because \text{三角不等式}\bigr) \\ &= \bigl|I(f) - I_\varphi(f)\bigr| + \bigl|I_\varphi(L_x f) - I(L_x f)\bigr| \qquad \bigl(\because \text{命題 I.43(1)}\bigr) \\ &< \epsilon \qquad \bigl(\because \text{(I.58)}\bigr)\end{aligned}$$

となる. ϵ は任意だったので

$$I(f) = I(L_x f) \qquad (\forall x \in G)$$

がわかった. 同様に $c > 0$ と $\epsilon > 0$ に対して $\varphi \in C_c^+(G)$ が存在して

$$\bigl|I(f) - I_\varphi(f)\bigr| < \frac{\epsilon}{2|c|} \quad \text{かつ} \quad \bigl|I(cf) - cI_\varphi(f)\bigr| < \frac{\epsilon}{2} \tag{I.60}$$

とできる. これらのことから

$$\begin{aligned}\bigl|I(cf) - cI(f)\bigr| &\leq \bigl|I(cf) - I_\varphi(cf)\bigr| + \bigl|I_\varphi(cf) - cI_\varphi(f)\bigr| \\ &= \bigl|I(cf) - cI_\varphi(f)\bigr| + \bigl|I_\varphi(cf) - cI_\varphi(f)\bigr| \\ &= \bigl|I(cf) - cI_\varphi(f)\bigr| + |c|\bigl|I_\varphi(f) - I_\varphi(f)\bigr| \\ &< \epsilon \qquad \bigl(\because \text{(I.60)}\bigr)\end{aligned}$$

となる. よって

$$I(cf) = cI(f)$$

を得る. 補題 I.48 から e の近傍 V が存在して

$$\bigl|I_\varphi(f+g) - I_\varphi(f) - I_\varphi(g)\bigr| < \frac{\epsilon}{3} \quad \text{かつ} \quad \operatorname{supp}(\varphi) \subset V \tag{I.61}$$

とできる. さらに $\varphi \in C_c^+(G)$ が存在して

$$\operatorname{supp}(\varphi) \subset V,\ \bigl|I(f+g) - I_\varphi(f+g)\bigr| < \frac{\epsilon}{3},\ \bigl|I(f) + I(g) - I_\varphi(f) - I_\varphi(g)\bigr| < \frac{\epsilon}{3} \tag{I.62}$$

とできる. よって

$$\begin{aligned}&\bigl|I(f+g) - I(f) - I(g)\bigr| \\ &\quad\leq \bigl|I(f+g) - I_\varphi(f+g)\bigr| + \bigl|I_\varphi(f+g) - I_\varphi(f) - I_\varphi(g)\bigr| \\ &\qquad + \bigl|I_\varphi(f) + I_\varphi(g) - I(f) - I(g)\bigr| \\ &\quad< \epsilon \qquad \bigl(\because \text{(I.60) と (I.61)}\bigr)\end{aligned}$$

となる. ϵ は任意だったので

$$I(f+g) = I(f) + I(g)$$

がわかった.

$f, g \in C_c^+(G)$ は任意に選んだので $I : C_c^+ \to \mathbb{R}$ は左移動不変かつ正値線型であることがわかった.

次に I を $C_c(G)$ 上の汎関数に拡張する．まず $I(0) := 0$ とおく．$\forall f \in C_c(G)$ に対して

$$I(f) := I(f^+) - I(f^{-1})$$

と定まる．ただし

$$f^+(x) = \max\{f(x), 0\}, \quad f^-(x) = \max\{-f(x), 0\} \qquad (\forall x \in G)$$

である．さらに

$$I(f) := I(\mathrm{Re}\, f) + I(\mathrm{Im}\, f)$$

とおけば I は $C_c(G)$ 上の正値線型汎関数となる．ただし $\mathrm{Re}\, f$ は f の実部, $\mathrm{Im}\, f$ は f の虚部である．よって Riesz の表現定理から I に対して G 上の Radon 測度 μ が一意的に存在して

$$I(f) = \int_G f(x)\,\mathrm{d}\mu(x) \qquad (\forall f \in C_c(G))$$

となる．I の定義から μ は零測度ではない．さらに

$$\int_G f(y)\,\mathrm{d}\mu(y) = I(f) = I(L_x f) = \int_G \big(L_x f\big)(y)\,\mathrm{d}\mu(y) \qquad (\forall x \in G,\, f \in C_c(G))$$

となるので μ は左不変 Haar 測度となることがわかった．□

I.4.3　Haar 測度の一意性の証明

$\boxed{\textbf{主張}\ 2}$ $\mathbb{G}$ を局所コンパクト Hausdorff 位相群とし, μ と ν を $\mathbb{G}$ 上の Haar 測度とする．このときある定数 $c > 0$ が存在して

$$\mu = c\nu$$

と書ける．

Proof. ① まず $f \in C_c^+(\mathbb{G})$ に対して

$$\frac{\int_{\mathbb{G}} f(x)\,\mathrm{d}\mu(x)}{\int_{\mathbb{G}} f(x)\,\mathrm{d}\nu(x)} = \frac{\int_{\mathbb{G}} g(x)\,\mathrm{d}\mu(x)}{\int_{\mathbb{G}} g(x)\,\mathrm{d}\nu(x)} \tag{I.63}$$

を示そう．

V_0 を $\mathbb{G}$ の単位元 e の対称コンパクト近傍とし

$$A := \mathrm{supp}(f)V_0 \bigcup V_0\mathrm{supp}(f), \qquad B := \mathrm{supp}(g)V_0 \bigcup V_0\mathrm{supp}(g),$$

とおく．ただし $\mathrm{supp}(f)V_0 = \{xy;\, x \in \mathrm{supp}(f),\, y \in V_0\}$ であり, $V_0\mathrm{supp}(f)$, $\mathrm{supp}(g)V_0$, $V_0\mathrm{supp}(g)$ も同様に定義している．$y \in V_0$ に対して写像 $f_y : \mathbb{G} \to \mathbb{R}$ と $g_y : \mathbb{G} \to \mathbb{R}$ をそれぞれ

$$f_y(x) = f(xy) - f(yx), \quad g_y(x) = g(xy) - g(yx)$$

と定める．明らかに, 各 $y \in V_0$ に対して

$$f_y,\, g_y \in C_c(\mathbb{G}), \quad \mathrm{supp}(f_y) \subset A, \quad \mathrm{supp}(g_y) \subset B$$

である．$\epsilon > 0$ をとる．系 I.34 から $\mathbb{G}$ の単位元 e の対称な近傍 V_1 が存在して

$$\left|f_y(x)\right| < \epsilon \quad \text{かつ} \quad \left|g_y(x)\right| < \epsilon \qquad (\forall y \in V_1)$$

とできる．ここで

$$V := V_0 \bigcap V_1$$

とおくと V は e の対称な近傍となる．

いま $\widetilde{h} \in C_c^+(\mathbb{G})$ をとり

$$\mathrm{supp}(\widetilde{h}) \subset V \tag{I.64}$$

とし

$$h(x) := \widetilde{h}(x) + \widetilde{h}(x^{-1}) \qquad (\forall x \in \mathbb{G})$$

と定める．このとき

$$h \in C_c(\mathbb{G}) \quad \text{かつ} \quad h(x) = h(x^{-1}) \qquad (\forall x \in \mathbb{G})$$

である．(I.64) から

$$\mathrm{supp}(h) \subset V$$

である．f, g, h はコンパクトな台をもつ．よって Tonelli の定理から

$$\begin{aligned}
\left(\int_{\mathbb{G}} h(y)\,\mathrm{d}\nu(y)\right)\left(\int_{\mathbb{G}} f(x)\,\mathrm{d}\nu(x)\right) &= \iint_{\mathbb{G}\times\mathbb{G}} h(y)f(x)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y) \\
&= \iint_{\mathbb{G}\times\mathbb{G}} h(y)f(x)\,\mathrm{d}\mu(L_y x)\,\mathrm{d}\nu(y) \\
&= \iint_{\mathbb{G}\times\mathbb{G}} h(y)f(yx)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y)
\end{aligned} \tag{I.65}$$

となる．さらに h の性質を用いると

$$\begin{aligned}\left(\int_{\mathbb{G}} h(x)\,\mathrm{d}\nu(x)\right)\left(\int_{\mathbb{G}} f(y)\,\mathrm{d}\nu(y)\right) &= \iint_{\mathbb{G}\times\mathbb{G}} h(x)f(y)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y)\\ &= \iint_{\mathbb{G}\times\mathbb{G}} h(x)f(y)\,\mathrm{d}\mu(L_{y^{-1}}x)\,\mathrm{d}\nu(y)\\ &= \iint_{\mathbb{G}\times\mathbb{G}} h(y^{-1}x)f(y)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y)\\ &= \iint_{\mathbb{G}\times\mathbb{G}} h(x^{-1}y)f(y)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y)\\ &= \iint_{\mathbb{G}\times\mathbb{G}} h(x^{-1}y)f(y)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(L_x y)\\ &= \iint_{\mathbb{G}\times\mathbb{G}} h(y)f(xy)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(L_x y)\end{aligned} \tag{I.66}$$

$$\left(\because L_y x\mapsto y;\ y\mapsto xy;\ x^{-1}y\mapsto y\right)$$

を得る．(I.65) と (I.66) を合わせると

$$\begin{aligned}&\left|\left(\int_{\mathbb{G}} h(x)\,\mathrm{d}\mu(x)\right)\left(\int_{\mathbb{G}} f(y)\,\mathrm{d}\nu(y)\right)-\left(\int_{\mathbb{G}} h(y)\,\mathrm{d}\nu(y)\right)\left(\int_{\mathbb{G}} f(x)\,\mathrm{d}\mu(x)\right)\right|\\ &\quad= \left|\iint_{\mathbb{G}\times\mathbb{G}} h(y)f(xy)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y)-\iint_{\mathbb{G}\times\mathbb{G}} h(y)f(yx)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y)\right|\\ &\quad= \left|\iint_{\mathbb{G}\times\mathbb{G}} h(y)\big\{f(xy)-f(yx)\big\}\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y)\right|\\ &\quad= \left|\iint_{\mathbb{G}\times\mathbb{G}} h(y)f_y(x)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y)\right|\\ &\quad\le \epsilon\mu(A)\int_{\mathbb{G}} h(y)\,\mathrm{d}\nu(y)\qquad \left(\because \mathrm{supp}(f_y)\subset A\ \text{かつ}\ |f_y(x)|<\epsilon\,(\forall y\in A)\right)\end{aligned} \tag{I.67}$$

となる．(I.67) の両辺を $\left(\int_{\mathbb{G}} h(y)\,\nu(y)\right)\left(\int_{\mathbb{G}} f(y)\,\mathrm{d}\nu(y)\right)$ で割ると

$$\left|\frac{\int_{\mathbb{G}} h(x)\,\mathrm{d}\mu(x)}{\int_{\mathbb{G}} h(x)\,\mathrm{d}\nu(x)}-\frac{\int_{\mathbb{G}} f(x)\,\mathrm{d}\mu(x)}{\int_{\mathbb{G}} f(x)\,\mathrm{d}\nu(x)}\right|\le\frac{\epsilon\mu(A)}{\int_{\mathbb{G}} f(x)\,\mathrm{d}\nu(x)} \tag{I.68}$$

を得る．f を g におきかえて同様の議論をすると

$$\begin{aligned}&\left|\left(\int_{\mathbb{G}} h(x)\,\mathrm{d}\mu(x)\right)\left(\int_{\mathbb{G}} g(y)\,\mathrm{d}\nu(y)\right)-\left(\int_{\mathbb{G}} h(y)\,\mathrm{d}\nu(y)\right)\left(\int_{\mathbb{G}} g(x)\,\mathrm{d}\mu(x)\right)\right|\\ &\quad\le \epsilon\mu(B)\int_{\mathbb{G}} h(y)\,\mathrm{d}\nu(y)\end{aligned} \tag{I.69}$$

を得る．(I.69) の両辺を $\left(\int_{\mathbb{G}} h(y)\,\nu(y)\right)\left(\int_{\mathbb{G}} g(y)\,\mathrm{d}\nu(y)\right)$ で割ると

$$\left|\frac{\int_{\mathbb{G}} h(x)\,\mathrm{d}\mu(x)}{\int_{\mathbb{G}} h(x)\,\mathrm{d}\nu(x)} - \frac{\int_{\mathbb{G}} g(x)\,\mathrm{d}\mu(x)}{\int_{\mathbb{G}} g(x)\,\mathrm{d}\nu(x)}\right| \le \frac{\epsilon\mu(B)}{\int_{\mathbb{G}} g(x)\,\mathrm{d}\nu(x)} \tag{I.70}$$

を得る．(I.68) と (I.70) を合わせると

$$\begin{aligned}
&\left|\frac{\int_{\mathbb{G}} f(x)\,\mathrm{d}\mu(x)}{\int_{\mathbb{G}} f(x)\,\mathrm{d}\nu(x)} - \frac{\int_{\mathbb{G}} g(x)\,\mathrm{d}\mu(x)}{\int_{\mathbb{G}} g(x)\,\mathrm{d}\nu(x)}\right| \\
&\qquad \le \left|\frac{\int_{\mathbb{G}} f(x)\,\mathrm{d}\mu(x)}{\int_{\mathbb{G}} h(x)\,\mathrm{d}\nu(x)} - \frac{\int_{\mathbb{G}} g(x)\,\mathrm{d}\mu(x)}{\int_{\mathbb{G}} h(x)\,\mathrm{d}\nu(x)}\right| + \left|\frac{\int_{\mathbb{G}} h(x)\,\mathrm{d}\mu(x)}{\int_{\mathbb{G}} h(x)\,\mathrm{d}\nu(x)} - \frac{\int_{\mathbb{G}} g(x)\,\mathrm{d}\mu(x)}{\int_{\mathbb{G}} g(x)\,\mathrm{d}\nu(x)}\right| \\
&\qquad \le \frac{\epsilon\mu(A)}{\int_{\mathbb{G}} f(x)\,\mathrm{d}\nu(x)} + \frac{\epsilon\mu(B)}{\int_{\mathbb{G}} g(x)\,\mathrm{d}\nu(x)}
\end{aligned}$$

を得る．$\epsilon > 0$ は任意だったので

$$\frac{\int_{\mathbb{G}} f(x)\,\mathrm{d}\mu(x)}{\int_{\mathbb{G}} f(x)\,\mathrm{d}\nu(x)} = \frac{\int_{\mathbb{G}} g(x)\,\mathrm{d}\mu(x)}{\int_{\mathbb{G}} g(x)\,\mathrm{d}\nu(x)}$$

を得る．この式は任意の $f,\, g \in C_c^+(\mathbb{G})$ に対して成立したので

$$\int_{\mathbb{G}} f(x)\,\mu(x) = c\int_{\mathbb{G}} f(x)\,\mathrm{d}\nu(x) \qquad (\forall f \in C_c^+(\mathbb{G})) \tag{I.71}$$

を得る．ただし

$$c = \frac{\int_{\mathbb{G}} g(x)\,\mathrm{d}\mu(x)}{\int_{\mathbb{G}} g(x)\,\mathrm{d}\nu(x)}$$

であり，これは g の取り方に依存しない．(I.71) を任意の $f \in C_c(\mathbb{G})$ に対して拡張するのは簡単である．最後に Riesz の表現定理における一意性の主張から $\mu = c\nu$ がわかる． □

I.5　モジュラー関数

μ を局所コンパクト群 $\mathbb{G}$ 上の左 Haar 測度とするとき，各 $a \in \mathbb{G}$ に対して

$$\mu_a(E) := \mu(Ea) \qquad (E \in \mathcal{B}(\mathbb{G}))$$

とすると μ_a も $\mathbb{G}$ 上の左 Haar 測度となる．したがって左 Haar 測度の一意性から $\Delta(a) > 0$ が存在して

$$\mu_a(E) = \Delta(a)\mu(E) \qquad (\forall \in \mathcal{B}(\mathbb{G}))$$

と表現できる．この関数 $\Delta : \mathbb{G} \to (0, \infty)$ を**モジュラー関数**という．

定理 I.49. $\Delta : \mathbb{G} \to \mathbb{R}_{>} := \{a \in \mathbb{R};\, a > 0\}$ は群 $\mathbb{G}$ から乗法群 $\mathbb{R}_{>}$ への連続な準同型写像であり

$$\int_{\mathbb{G}} \big(R_a f\big)(x)\,\mathrm{d}\mu(x) = \Delta(a^{-1}) \int_{\mathbb{G}} f(x)\,\mathrm{d}\mu(x) \qquad (\forall a \in \mathbb{G}) \tag{I.72}$$

が成り立つ.

Proof. $a, b \in \mathbb{G}$ とし, $\mu(E) \neq 0$ をみたす $E \in \mathcal{B}(\mathbb{G})$ をとる. すると

$$\Delta(a,\, b)\mu(E) = \mu(Eab) = \Delta(b)\mu(Ea) = \Delta(b)\Delta(a)\mu(E)$$

となる. ゆえに

$$\Delta(ab) = \Delta(a)\Delta(b)$$

である. 次に

$$xa \in E \iff x \in Ea^{-1}$$

に注意すると

$$\begin{aligned} \int_{\mathbb{G}} \big(R_a \mathbb{1}_E\big)(x)\,\mathrm{d}\mu(x) &= \int_{\mathbb{G}} \mathbb{1}_E(xa)\,\mathrm{d}\mu(x) = \int_{\mathbb{G}} \mathbb{1}_{Ea^{-1}}(x)\,\mathrm{d}\mu(x) \\ &= \Delta\big(Ea^{-1}\big) = \Delta(a^{-1})\mu(E) = \Delta(a^{-1}) \int_{\mathbb{G}} \mathbb{1}_E(x)\,\mathrm{d}\mu(x) \end{aligned}$$

となる. 定理の等式 (I.72) は $f = \mathbb{1}_E\ \big(E \in \mathcal{B}(\mathbb{G})\big)$ のとき成り立つことを示している. あとは標準機械を用いて定理の等式 (I.72) を証明すればよい.

最後に Δ が連続であることこを示そう. まず $\mathbb{G}$ の単位元 e と $\forall a \in \mathbb{G}$ に対して

$$\begin{aligned} \int_{\mathbb{G}} f(x)\,\mathrm{d}\mu(x) &= \int_{\mathbb{G}} \big(R_e f\big)(x)\,\mathrm{d}\mu(x) = \Delta(e^{-1}) \int_{\mathbb{G}} f(x)\,\mathrm{d}\mu(x) \\ &= \Delta(e) \int_{\mathbb{G}} f(x)\,\mathrm{d}\mu(x) \end{aligned}$$

から $\Delta(e) = 1$ となることに注意する. $\forall a,\, a_0 \in \mathbb{G}$ に対して

$$\big|\Delta(a) - \Delta(a_0)\big| = \Delta(a_0)\big|\Delta(a_0^{-1}a) - \Delta(e)\big| = \Delta(a_0)\big|\Delta(a_0^{-1}a) - 1\big|$$

となるので, Δ が点 e において連続であることを示せばよい. 以下では $f \in C_c(\mathbb{G})$ で

$$\int_{\mathbb{G}} f(x)\,\mathrm{d}\mu(x) = 1$$

となるものを一つ固定する. e のコンパクト近傍 W_0 を一つとって固定し

$$S := \Bigl(\bigl(\mathrm{supp}(f)\bigr)W_0\Bigr)\bigcup\bigl(\mathrm{supp}(f)\bigr)$$

とおく. すると命題 I.35(iii) から S はコンパクトでる. 任意の $\epsilon > 0$ が与えられたとき, f は右一様連続なので, e の開近傍 V で $V = V^{-1}$ かつ $V \subset W_0$ なるものが存在して

$$a \in V \iff \bigl|f(xa^{-1}) - f(x)\bigr| < \frac{\epsilon}{\mu(S)} \qquad (x \in \mathbb{G})$$

となる. $a \in V$ とする. すると

$$\Delta(a) = \int_{\mathbb{G}} \bigl(R_{a^{-1}} f\bigr)(x)\,\mathrm{d}\mu(x)$$

であり, $x \notin S$ ならば

$$f(x) = f(xa^{-1}) = 0$$

である. よって

$$\begin{aligned}\bigl|\Delta(a) - 1\bigr| &= \left|\int_{\mathbb{G}} f(xa^{-1})\,\mathrm{d}\mu(x) - \int_{\mathbb{G}} f(x)\,\mathrm{d}\mu(x)\right| \\ &\leq \int_{\mathbb{G}} \bigl|f(xa^{-1}) - f(x)\bigr|\,\mathrm{d}\mu(x)\end{aligned}$$

となり, 定理は証明された. □

定理 I.50. μ を局所コンパクト群 $\mathbb{G}$ 上の左 Haar 測度とし, Δ をそのモジュラー関数とする. このとき

$$\int_{\mathbb{G}} f(x^{-1})\,\mathrm{d}\mu(x) = \int_{\mathbb{G}} f(x)\Delta(x^{-1})\,\mathrm{d}\mu(x) \tag{I.73}$$

が成り立つ.

Proof.

$$I(f) := \int_{\mathbb{G}} f(x)\Delta(x^{-1})\,\mathrm{d}\mu(x)$$

とおくと

$$\begin{aligned}I\bigl(R_a f\bigr) &= \int_{\mathbb{G}} f(xa)\Delta(x^{-1})\,\mathrm{d}\mu(x) = \Delta(a)\int_{\mathbb{G}} f(xa)\Delta\bigl((xa)^{-1}\bigr)\,\mathrm{d}\mu(x) \\ &= \Delta(a)\int_{\mathbb{G}} \bigl\{R_a\bigl(f\Delta^{-1}\bigr)\bigr\}(x)\,\mathrm{d}\mu(x) \\ &= \Delta(a)\Delta(a^{-1})\int_{\mathbb{G}} f(x)\Delta^{-1}(x)\,\mathrm{d}\mu(x) \qquad \bigl(\because \text{(I.72)}\bigr) \\ &= I(f)\end{aligned}$$

となる．すなわち $\Delta(x^{-1})\mu$ は右 Haar 測度．
(2)

$$\rho(E) := \mu(E^{-1})$$

とおくと ρ は右 Haar 不変測度である．よって右 Haar 不変測度の一意性から, ある定数 $c > 0$ が存在して

$$\Delta(x^{-1})\mu(E) = c\rho(E) \qquad \big(\forall E \in \mathcal{B}(\mathbb{G})\big)$$

となる．したがって $f \in C_c(\mathbb{G})$ に対して

$$\int_{\mathbb{G}} f(x)\Delta(x^{-1})\,\mathrm{d}\mu(x) = c\int_{\mathbb{G}} f(x)\,\mathrm{d}\rho(x) = c\int_{\mathbb{G}} f(x^{-1})\,\mathrm{d}\mu(x) \tag{I.74}$$

となる．関数 $g(x) := f(x^{-1})\Delta(x)$ を考えると, $g \in C_c(\mathbb{G})$ であるから, (I.74) に g を適用すると

$$\int_{\mathbb{G}} g(x)\Delta(x^{-1})\,\mathrm{d}\mu(x) = c\int_{\mathbb{G}} g(x^{-1})\,\mathrm{d}\mu(x)$$

となる．すると

$$\begin{aligned}
\int_{\mathbb{G}} f(x^{-1})\,\mathrm{d}\mu(x) &= \int_{\mathbb{G}} f(x^{-1})\Delta(x)\Delta(x^{-1})\,\mathrm{d}\mu(x) = \int_{\mathbb{G}} g(x)\Delta(x^{-1})\,\mathrm{d}\mu(x) \\
&= c\int_{\mathbb{G}} g(x^{-1})\,\mathrm{d}\mu(x) = c\int_{\mathbb{G}} f(x^{-1})\Delta(x)\,\mathrm{d}\mu(x) \\
&= c^2\int_{\mathbb{G}} f(x^{-1})\,\mathrm{d}\mu(x)
\end{aligned}$$

となり, $c = 1$ を得る．よって

$$\int_{\mathbb{G}} f(x)\Delta(x^{-1})\,\mathrm{d}\mu(x) = \int_{\mathbb{G}} f(x^{-1})\,\mathrm{d}\mu(x)$$

がわかる．　□

定義 I.51. 局所コンパクト群 $\mathbb{G}$ が**ユニモジュラー**であるとは, $\Delta(x) = 1\,(\forall x \in \mathbb{G})$ のときをいう．

命題 I.52. $\mathbb{G}$ がコンパクトのとき, $\mathbb{G}$ はユニモジュラー．

$0 < \mu(\mathbb{G}) < \infty$ となる．μ が左 Haar 不変測度とすると

$$\mu(\mathbb{G}) = \mu(\mathbb{G}a) = \Delta(a)\mu(\mathbb{G}) \qquad (\forall a \in \mathbb{G})$$

となる．よって

$$\Delta(a) = 1 \qquad (\forall a \in \mathbb{G}).$$

I.6　章末注釈と参考文献

関連図書

[1] Almudevar, A. (2022). Theory of Statistical Inference. CRC Press.

[2] Axler, S. (2020). Measure, Integration & Real Analysis. *Springer.*

[3] Bisgard, J. (2021). Analysis and Linear algebra: The Singular Value Decomposition and Applications. *Student Mathematical Library* **94**. American Mathematical Society.

[4] Borwein, J.M., Lewis, A.S. (2006). ConvexAnalysis and Nonlinear Optimization Theory and Examples, 2nd. edition. Spriner.

[5] Boyd, S., Vandenberghe, L. (2018). Introduction to Applied Linear algebra. Cambridge University Press.

[6] DasGupta, A. (2008). Asymptotic Theory of Statistics and Probablility. Springer.

[7] Durrett, R. (2019). Probability, 2nd. edition. *Cambridge Series in Statistical and Probalistic Mathematics.*

[8] Ferguson, T. S. (1996). *A Course in Large Sample Theory.* Chapman & Hall.

[9] Folland, G. F. (1999). *Real Analysis*, 2nd edition. John Wiley & Sons, INC.

[10] Gallier, J. (2001). Geometrical Methods and Applications. *Texts in Applied Mathematics* **38**. Springer.

[11] Hansen, E. (2009). Measure Theory. Forth edition. Department of Mathematical Statistics, University of Copenhagen.

[12] Hult, H. (2015). Lecture note on SF3961 Graduate course in Statistical inference. `https://www.math.kth.se/matstat/gru/Statistical\%20inference/literature.html` (accessed at 2023/06/03).

[13] KAIPIO, J., SOMERSLO, E. (2005). Statistical and Computational Inverse Problems. *Applied Mathematical Sciences* **60**. Springer.

[14] KEENER, R.W. (2010). Theoretical Statistics. *Springer Texts in Statistics*. Springer.

[15] LEDERER, J. (2022). Fundamentals of High-Dimensional Statistics with Exercies and R Labs. *Springer Text in Statistics*. Springer.

[16] MUNKRES, J. R. (2000). Topology. Second editon. Pearson.

[17] NYBLOM, J. (2023+). Note on Legendre's Method of Least Squares. To appear in *Stattistical Science*.

[18] PANCHENKO, D. (2019). INTRODUCTION TO PROBABILITY THEORY. ISBN-13: 978-1-9994190-2-8.

[19] RIGOLLET, P., HÜTTER, J.-C. (2017). High Dimensionnal Statistics. Available at `https://math.mit.edu/~rigollet/PDFs/RigNotes17.pdf`(2023/05/17 accessed).

[20] SCHILLING, R. L. (2017). Measures, Integrals and Martingales. *Cambridge*.

[21] SPOKOINY, V. AND DICKHAUS, T. (2015). Basics of Modern Mathematical Statistics. Springer.

[22] STIEGLER, S. M. (1981). Gauss and the invention of least squares. *Annals of Statistics* **9** 465-474.

[23] WASSERMAN, L. (2004). All of Statistics. Springer.

[24] 新井仁之 (2006). 線形代数　基礎と応用 (第 1 刷). 日本評論社.

[25] E. アルティン (2002). ガンマ関数入門 (第 1 版第 1 刷・上野健爾訳). 日本評論社.

[26] 石谷謙介 (2021). ガイダンス 確率統計 (初版).

[27] 加藤賢吾 (2015),『数理統計　講義ノート』, Available at

`https://drive.google.com/file/d/0B7C_CufYq6j6ekFINVhTS09lZzQ/view?resourcekey=0-zaYo9XudjhK_K6kMfpq-XQ(accessed 2022/09/22).`

[28] 笠原勇二 (2013). 明解 確率論入門 (第 1 版第 2 刷). 日本評論社.

[29] 河野敬雄 (2013). 確率概論 (初版第 5 刷). 京都大学学術出版.

[30] 清水康隆 (2021). 統計学への確率論, その先へ (第 2 版). 内田老鶴圃
.

[31] 清水康隆 (2023). 統計学の漸近論, その先は (第 1 版). 内田老鶴圃.

[32] 谷口正信 (2005). 数理統計・時系列・金融工学 (初版第 1 刷). 朝倉書店.

[33] 原啓介 (2017). 『測度・確率・ルベーグ積分』, 講談社.

[34] 日野正訓 (2023). ルベーグ積分の基礎 (初版第 1 刷).　共立出版.

[35] 藤岡敦 (2020). 『集合と位相』, 裳華房 (第 1 版第 1 刷).

[36] 藤岡敦 (2021). 『入門情報幾何』, 共立出版.

[37] 福水健次 (2010). 『カーネル法入門』, 朝倉書店.

[38] 吉田伸生 (2021). 『ルベーグ積分入門』, 日本評論社.

[39] 高橋礼二 (2019). 線型代数講義 (第 1 版第 2 刷). 日本評論社.

[40] 永原正章 (2020). スパースモデリング (第 1 版第 4 刷). コロナ社.

[41] 鍋谷清治 (1989). 数理統計学 (初版版第 3 刷). コロナ社.

[42] 森棟公夫 (2010). 地球の大きさと最小 2 乗法. 社会とマネジメント **8** 111-130.

[43] 柳田英二 (2022). 解析入門 (第 1 版第 1 刷). 裳華房.