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 [39, 10, 17, 38] 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

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

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

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

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

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

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 $f : X \to Y$ 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) $f^{-1}(B_1 \setminus B_2) = f^{-1}(B_1) \setminus f^{-1}(B_2)$.
(9) $f^{-1}\bigl(f(A)\bigr) \supset A$.
(10) $f\bigl(f^{-1}(B)\bigr) \subset B$.

Proof. 証明は [39, 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) ∼ (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 [39, 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 [39, 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 [39, 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 [39, 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 $d : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ 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, $d(x, y) = 0 \Leftrightarrow x = y$ がある.
(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 A が**開集合**であるとは, 任意の $x \in A$ に対して, $\exists r > 0$ を選んで, $B(x;\, r) \subset A$ とできること[1]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.

[1] $C,\, D$ を命題とする. そのとき, C が成立しないならば, 命題「$C \Rightarrow D$」は真と約束する. したがって, $A = \varnothing$ の場合には, 任意の $x \in A$ が取れないので, 空集合も開集合である. 本によっては, 上記のような論理を踏まずに空集合は開集合であることを約束する場合もあるようだ. この論法を使うと空集合は閉集合でもあることがわかる. Euclid 位相空間において, 空集合と全体集合は常に開集合でもあり, 閉集合である.

(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) $\sim$ (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 [38] を借用した. 第 0.2 節は [39, pp.26 − 29] を借用した. 第 0.3 節は [38] を借用した. 第 0.4 節は [38] と [39] を参照した. 第 0.5 節は [38, 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

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

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

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

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

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

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) $A_n \in \mathcal{A}$ 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.

(6) $A_n \in \mathcal{A}$ 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.

(7) (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.

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) $F_1 := A_1$, $F_2 = A_2 \setminus A_1$, $\ldots$, $F_{n+1} := A_{n+1} \setminus A_n$ 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．
(6) $A = \bigcap_{n=1}^{\infty} A_n$ 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

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

がわかる．
(7) まず

$$
\mathsf{Pr}(A_1 \cup A_2) \leq \mathsf{Pr}(A_1) + \mathsf{Pr}(A_2) \tag{1.1}
$$

を示せす. そのために

$$A_1 \cup A_2 = A_1 \cup \Big(A_2 \setminus (A_1 \cap A_2) \Big) \quad \text{かつ} \quad A_1 \bigcap \Big(A_2 \setminus (A_1 \cap A_2) \Big) = \varnothing$$

であることに注意する. (3) から

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

がわかる. (1.1) の操作を繰り返せば, $N \in \mathbb{N}$ に対して

$$\mathsf{Pr}\left(\bigcup_{n=1}^{N} A_n \right) \le \sum_{n=1}^{N} \mathsf{Pr}(A_n) \tag{1.2}$$

がわかる. $\{\cup_{n=1}^{N} A_n\}_{N=1}^{\infty}$ は増加列なので, (5) と (1.2) から

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

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):

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

(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) − (2) 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 + の人が本当に 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. 写像 $X : \Omega \to \mathbb{X}$ 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. (1) $d \geq 2\,(d\in\mathbb{N})$ とする. $\bigl(\mathbb{X}, \mathcal{B}\bigr) = \bigl(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\bigr)$ のとき, X は確率空間 $\bigl(\Omega, \mathcal{A}, \mathsf{Pr}\bigr)$ 上の**確率ベクトル**と呼ばれる. 定義されている確率空間に誤解がないときには, 簡単に確率ベクトルということもある.
(2) $d = 1$ のとき, X は確率空間 $\bigl(\Omega, \mathcal{A}, \mathsf{Pr}\bigr)$ 上の **確率変数**と呼ばれる. 定義されている確率空間に誤解がないときには, 簡単に確率変数と簡単にいうこともある.
(3) □

定理 1.16. $\bigl(\Omega, \mathcal{A}\bigr)$ と $\bigl(\mathbb{X}, \mathcal{B}\bigr)$ 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, $\bigl\{\omega\in\Omega :\ X(\omega)\in B\bigr\}$ 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

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

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

となる．したがって, 集合族 $\mathcal{D} := \{B \in \mathcal{B};\ \{X \in B\} \in \mathcal{A}\}$ は σ 加法族 (必要ならば, $\mathbb{X}$ も加える) となる．よって, $\mathcal{C} \subset \mathcal{D}$ であり, $\mathcal{C}$ は $\mathcal{B}$ を生成するので, 最小性から $\mathcal{B} \subset \mathcal{D}$ となることから X は確率変数であることがわかる． □

注意 1.17. (1). $\big(\mathbb{X}, \mathcal{B}\big) = \big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ 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．
(2). $d \geq 2\,(d \in \mathbb{N})$ とする．$\big(\mathbb{X}, \mathcal{B}\big) = \big(\mathbb{R}^d, \mathcal{B}(\mathbb{R})\big)$ 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. $\big(\Omega, \mathcal{A}\big)$, $\big(\mathbb{X}, \mathcal{B}\big)$, $\big(\mathbb{Y}, \mathcal{C}\big)$ 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．

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．

[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．

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. □

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

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

から $\inf_n X_n$ 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. □

[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.

注意 1.23. (1). $X_1, X_2, \ldots\ldots$ を確率変数列とする. 定理 1.22 と 定理 1.19 から

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

は可測集合となる. $\mathsf{Pr}(\Omega_0)=1$ 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

$$X_\infty(\omega) := \lim_{n\to\infty} X_n(\omega) \qquad (\text{a.s.})\,\omega\text{in}\Omega$$

と書ける.
(2). 任意の $a, b\in\mathbb{R}\,(a<b)$ とし, $\{-\infty\}\cup(-\infty, a), (a, b), (b, \infty)\cup\{+\infty\}$ によって生成される σ 加法族を$\mathcal{B}(\mathbb{R}\cup\{\pm\infty\})$ と書く. 集合 $D\subset\Omega$ が定義域で, $\mathbb{R}\cup\{\pm\infty\}$ を終域ともつ関数 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

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

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.

[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.

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.

(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, $\mathsf{F}(x-) := \sup_y\{\mathsf{F}(y);\, y < x\}$ とした.

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

Proof. (1) ⇒ (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) ⇒ (3) の証明:

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

であるので, 補題 1.7(6) より

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

を得る. さらに補題 1.7(4) より

$$\begin{aligned}\mathsf{Pr}(X = x) &= \mathsf{Pr}\Big(\{X \le x\} \setminus \{X < x\}\Big) = \mathsf{Pr}(X \le x) - \mathsf{Pr}(X < x) \\ &= F(x) - F(x-) \qquad (1.3)\end{aligned}$$

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: (1.3) 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

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

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.

[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.

[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.

(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

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

となる.

(2) $\Omega = [0, 1]$, $\mathcal{A} = \mathcal{B}(\mathbb{R}) \cap [0, 1] := \{B \cap [0, 1]; B \in \mathcal{B}(\mathbb{R})\}$ とし

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

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

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

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

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

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

$$\mathsf{F}\big(\mathsf{F}^{-}(y)\big) = \mathsf{F}\big(\lim_{n\to\infty} x_n\big) = \lim_{n\to\infty} \mathsf{F}(x_n) \geq y \tag{1.4}$$

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

$$\mathsf{F}^{-}(y) \leq x \tag{1.5}$$

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. (1.5) と F の非減少性から

$$\mathsf{F}\big(\mathsf{F}^{-}(y)\big) \leq \mathsf{F}(x) = y \tag{1.6}$$

となる. 一方, F^{-} の定義から

$$\mathsf{F}\big(\mathsf{F}^{-}(y)\big) \geq y \tag{1.7}$$

となる. (1.6) と (1.7) 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

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

となる. 両辺の inf をとると

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

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

$$x_n = \mathsf{F}^{-}(y_n) \leq \mathsf{F}^{-}(y) =: x_0 \tag{1.8}$$

となる. よって, (1.8) の両辺の極限を取ると

$$x := \lim_{n\to\infty} x_n \leq x_0 \tag{1.9}$$

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 (1.7) に注意すると

$$y_n \leq \mathsf{F}\bigl(\mathsf{F}^{-}(y_n)\bigr) \leq \mathsf{F}(x_n) \leq \mathsf{F}(x + \epsilon) = \mathsf{F}(x_0 - \epsilon)$$

となる. すると

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

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 $\mathsf{F}(z) \geq y$ をみたす z の inf である. $x_0 > x_0 - \epsilon$ だから, $x_0 - \epsilon \not\in \{z \in \mathbb{R} :\ \mathsf{F}(z) \geq y\}$ 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 $y \leq \mathsf{F}(x_0 - \epsilon)$ 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

$$y \leq \mathsf{F}(x) \ \Rightarrow\ \mathsf{F}^{-}(y) \leq \mathsf{F}^{-}\bigl(\mathsf{F}(x)\bigr) \leq x$$

となる．最後の不等号は

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

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

逆に，$\mathsf{F}^{-}(y) \leq x$ ならば，(1.7) と F の非減少性から

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

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

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

となる．よって

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

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

$0 \leq \theta \leq 1$ 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．

注意 1.42. $\theta = 0$ のとき，$0^0 = 1$ と定めていることに注意せよ．

2 **項分布**

$n \in \mathbb{N},\, 0 \le \theta \le 1$ とする．確率変数 X は**母数** $(n,\, \theta)$ **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．

注意 1.43. 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 1.44. 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, $z = (x-\mu)/\sigma$ 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** (α, β) **のガンマ分布**に従うとは, X の p.d.f. $\mathsf{p}(\,\cdot\,|\,\alpha, \beta)$ が

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

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

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^2_p 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, **同時確率関数** (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 1.45. 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 1.46. (X, Y) の同時累積分布関数 (同時 c.d.f.) F を

$$\begin{aligned}\mathsf{F}(x, y) &:= \mathsf{Pr}(X \leq x, Y \leq y) &\quad (1.10)\\ &:= \mathsf{Pr}(\{\omega \in \Omega;\ X(\omega) \leq x, Y(\omega) \leq y\}) \qquad (\forall(x, y) \in \mathbb{R}^2)\end{aligned}$$

で定義する. (X, Y) の同時分布 P を

$$\mathsf{P}(A) = \mathsf{Pr}\big((X, Y) \in A\big) \quad (\forall A \in \mathcal{B}(\mathbb{R}^2)) \tag{1.11}$$

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, (1.10) 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 1.47. 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 1.48. (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 1.49. 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 1.50. 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 1.50 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 1.51. 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 1.52. 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 1.53. (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 1.52 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. □

例 1.54. 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 1.55. 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 1.56. 定理 1.55 の主張も「ほとんど至る所」でよい. □

[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.
[14] 有界でなくともよい.

例 1.57. 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

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

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

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 1.58. (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 1.59. (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 1.60. 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 $\mathsf{Pr}(X \le 1/4 | Y = 1/3)$ を求めてみよう．まず $0 < y < 1$ に対して

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

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 1.59 より

$$\mathsf{Pr}\left(X \le \frac{1}{4} \middle| Y = \frac{1}{3}\right) = \int_0^{1/4} p^{X|Y}\left(x \middle| \frac{1}{3}\right)\mathrm{d}x = \int_0^{1/4} \frac{x + \frac{1}{3}}{\frac{1}{3} + \frac{1}{2}}\,\mathrm{d}x = \frac{11}{80}$$

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 1.61. 本講義録では, ベクトルは縦ベクトルとする． □

定義 1.62. $X_1,\, X_2,\, \ldots,\, X_n$ 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

$$\mathsf{Pr}(X_1 \in A_1,\, X_2 \in A_2,\, \ldots,\, X_n \in A_n) = \prod_{j=1}^{n} \mathsf{Pr}(X_j \in A_j) \tag{1.12}$$

が成立するときである．

注意 1.63. $\boldsymbol{X}$ 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. (1.12) 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 1.64. $X_1, X_2, \ldots, X_n$ 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 からの標本の大きさが 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, $\sum_{j=1}^{d} x_j = n$ で

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

[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.

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

補題 1.65. $\boldsymbol{X} \sim \mathsf{Multi}_d(n, \boldsymbol{p})$ とする．ただし，$\boldsymbol{p} = (p_1, p_2, \dots, p_d)$ と $\boldsymbol{X} = (X_1, X_2, \dots, X_d)$ とする．このとき $X_j\ (j = 1, 2, \dots, d)$ 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． □

多変量正規分布

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

$$\boldsymbol{Z} := \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_d \end{pmatrix} \tag{1.13}$$

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

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

と書ける．ただし $\boldsymbol{z} = (z_1, z_2, \dots, z_d)^\top$ である．これを $\boldsymbol{Z} \sim \mathsf{N}_d(\boldsymbol{0}, \boldsymbol{I}_d)$ 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

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

となっていることもわかる．

次に

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

[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，$f((z_1, z_2, \dots, z_d)^\top)$ などと書くのは煩わしい．

とする. ただし $\mathbf{\Sigma}$ 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

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

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

$$\boldsymbol{X}=\boldsymbol{\mu}+\mathbf{\Sigma}^{1/2}\boldsymbol{Z} \tag{1.14}$$

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

$$\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\boldsymbol{\mu},\,\mathbf{\Sigma})=\frac{1}{(2\pi)^{d/2}\det(\mathbf{\Sigma})^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top\mathbf{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}\quad(\boldsymbol{x}\in\mathbb{R}^d) \tag{1.15}$$

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 1.66. (1.15) の導出は以下のように行うことができる. 一般に $\boldsymbol{Z}=(Z_1,\,Z_2,\,\dots,\,Z_d)^\top$ 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

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

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

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

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

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

が存在して, $\boldsymbol{Z}=\boldsymbol{h}^{-1}(\boldsymbol{X})$ となる. ここで, すこし記号の乱用をしている. すなわち, h_j^{-1} は $\boldsymbol{h}^{-1}$ の第 j 成分の表記であって, h_1 の逆関数ではないことに注意せよ. $\boldsymbol{X}$ 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

$$\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x})=\left|\mathbf{J}_{\boldsymbol{h}^{-1}}(\boldsymbol{x})\right|\mathsf{p}^{\boldsymbol{Z}}(\boldsymbol{h}^{-1}(\boldsymbol{x})) \tag{1.16}$$

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

[18] 実は, a.e $\mathbb{X}$ でよい.

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 (1.16) に代入すれば, (1.15) はわかる. 以上の議論から

$$\int\int\cdots\int_{\mathbb{R}^d} \mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_d = \int\int\cdots\int_{\mathbb{R}^d} \mathsf{p}^{\boldsymbol{Z}}(\boldsymbol{z})\,\mathrm{d}z_1\mathrm{d}z_2\cdots\mathrm{d}z_d = 1$$

となっていることもわかる. □

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

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

を示せ.

注意 1.67. $\boldsymbol{\Sigma}$ を半正定値としたときに, 多変量正規分布を (1.13) と変換 (1.14) 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 1.68. $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)^\top \sim \mathsf{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 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

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

となる.

(3) 定数ベクトル $\boldsymbol{c} \in \mathbb{R}^d\,(\boldsymbol{c} \neq \mathbf{0}_d)$ 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, $\mathbf{0}_d$ 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. なお, 自由度 d の χ^2 分布 χ^2_d は定義 1.74 で導入する.

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

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

と書ける．ただし $\boldsymbol{I}_{d-1}$ は $(d-1)$ 次の単位行列で，$\boldsymbol{0}_{d-1}$ 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

$$\begin{aligned}\boldsymbol{\Sigma}^{-1}&=\left(\begin{array}{c|c}1 & -\sigma_{11}^{-1}\boldsymbol{\sigma}_1^\top \\ \hline \boldsymbol{0}_{d-1} & \boldsymbol{I}_{d-1}\end{array}\right)\left(\begin{array}{c|c}\sigma_{11}^{-1} & \boldsymbol{0}_{d-1}^\top \\ \hline \boldsymbol{0}_{d-1} & \left\{\boldsymbol{\Sigma}_2-\sigma_{11}^{-1}\boldsymbol{\sigma}_1\boldsymbol{\sigma}_1^\top\right\}^{-1}\end{array}\right)\\&\quad\times\left(\begin{array}{c|c}1 & \boldsymbol{0}_{d-1}^\top \\ \hline -\sigma_{11}^{-1}\boldsymbol{\sigma}_1 & \boldsymbol{I}_{d-1}\end{array}\right)\end{aligned}\tag{1.17}$$

となる．$\boldsymbol{\mu}=(\mu_1,\,\mu_2,\,\ldots,\,\mu_d)^\top$ と $\boldsymbol{x}=(x_1,\,x_2,\,\ldots,\,x_d)^\top$ と書いたときに，$\boldsymbol{\mu}_1=(\mu_2,\,\mu_3,\,\ldots,\,\mu_d)^\top$ と $\boldsymbol{x}_1=(x_2,\,x_3,\,\ldots,\,x_d)^\top$ とおく．(1.17) から

$$\begin{aligned}\mathrm{tr}\,(\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})&=\frac{1}{\sigma_{11}}(x_1-\mu_1)^2\\&\quad+\mathrm{tr}\left[\left(\boldsymbol{x}_1-\boldsymbol{\mu}_1-\frac{x_1-\mu_1}{\sigma_{11}}\boldsymbol{\sigma}_1\right)^\top\left\{\boldsymbol{\Sigma}_2-\sigma_{11}^{-1}\boldsymbol{\sigma}_1\boldsymbol{\sigma}_1^\top\right\}^{-1}\right.\qquad(1.18)\\&\quad\left.\times\left(\boldsymbol{x}_1-\boldsymbol{\mu}_1-\frac{x_1-\mu_1}{\sigma_{11}}\boldsymbol{\sigma}_1\right)\right]\\&=:\frac{1}{\sigma_{11}}(x_1-\mu_1)^2+g_{d-1}(\boldsymbol{x}_1)\qquad(1.19)\end{aligned}$$

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) (1.19) からわかる．

(3) $\boldsymbol{c}^\top \boldsymbol{X}$ の 第 2 章で定義する積率を計算すればよい．

(4) (3) から

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

となる．あとは χ^2 分布の定義からわかる． □

(1.16) を用いた例題を述べておく

例 1.69. 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．なお, 上式では記号の乱用をしている．h_1^{-1}, h_2^{-1} は単に $\boldsymbol{h}^{-1}$ の成分を表し, h_1, h_2 の逆写像ではないことに注意せよ．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

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

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 1.70. (ガンマ分布) $\alpha,\, \beta > 0$ とする．連続型確率変数 X が母数 $(\alpha,\, \beta)$ 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 1.71. $n \geq 2$ 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 1.72. 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

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

となる. よって微積分の基本定理より

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

を得る. □

補題 1.71 **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 1.72 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

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

より

$$\int_0^1 (1-z)^{\alpha_1-1}z^{\alpha_2-1}\,\mathrm{d}z = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)} \tag{1.20}$$

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 1.73. $\alpha_1>0,\ \alpha_2>0$ 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 (1.20) 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 1.74. $Z_1,\,Z_2,\,\ldots,\,Z_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}(0,\,1)$ 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

$$\mathsf{p}^Y(y) = \begin{cases} \dfrac{1}{\sqrt{2\pi}} y^{(1/2)-1} e^{-y/2} & (y > 0) \\ 0 & (\text{その他の場合}) \end{cases} \tag{1.21}$$

となる．演習問題 1.2 を参照せよ．一方, $\mathsf{Ga}(1/2, 1/2)$ の p.d.f. は

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

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 1.71 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

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

である．

1.7.3 F 分布

定義 1.75. $k, m \in \mathbb{N}$ 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 1.76. 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 1.77. $k,\, m \in \mathbb{N}$ 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 1.78. $X,\, Y$ 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

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

となる．ここで微積分の基本定理を用いると

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

となる． □

補題 1.77 の証明　補題 1.78 を用いる．$z > 0$ のとき, $Z = X/Y$ の

p.d.f. は

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

となる. $F = \frac{m}{k}Z$ 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 1.79. $Z_0,\,Z_1,\,Z_2,\,\ldots,\,Z_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}(0,\,1)$ のとき

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

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

補題 1.80. 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, $t \mapsto t^2$ と変換すると

$$\int_0^{x^2} \mathsf{p}^{1,\,n}(t)\,\mathrm{d}t = \int_0^x \mathsf{p}^{1,\,n}(t^2)2t\,\mathrm{d}t \tag{1.22}$$

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 (1.22) の右辺に代入すると

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

となる. □

1.8 章末注釈と参考文献

第 1.1 節は [22] を参照した. 第 1.2 節は [8] を参照した. 第 1.3 節から第 1.6 節は [30] を参照した. 第 1.7 節は [19] 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 (1.21) 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, (1.16) 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．関数 $g,\, h : \mathbb{R} \to \mathbb{R}$ を可測とする．

(1) $h(X)$ は有限の期待値を持つとする．$g(x) \leq h(x)\,(\forall x \in \mathbb{R})$ ならば

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

となる．

(2) $\mathsf{E}[|g(X)|] < \infty$, $\mathsf{E}[|h(X)|] < \infty$ を仮定する．$a,\, b \in \mathbb{R}$ に対して

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

となる．

Proof. (1) の証明: $h(x) - g(x) \geq 0\,(x \in \mathbb{R})$ であるので, 定義 2.1(1) と積分の性質から

$$\mathsf{E}\big[h(X) - g(X)\big] = \int_{-\infty}^{\infty} \underbrace{\big\{h(x) - g(x)\big\}\mathsf{p}^X(x)}_{\geq 0}\,\mathrm{d}x \geq 0$$

となる．さらに, (1) を用いると $\mathsf{E}[h(X)] \geq \mathsf{E}[g(X)]$ がわかる．
(2) の証明: 確率変数 X が p.d.f. p^X を持つ場合を示す．$\big|ag(x) + bh(x)\big| \leq |a||g(x)| + |b||h(x)|$ と $g(X)$ と $h(X)$ の期待値が有限なので, (1) から $ag(X) + bh(X)$ は有限な期待値をもつ．このことから, 定義 2.1(1) のステップは省略できる．あとは, 積分の線型性から

$$\begin{aligned}
\mathsf{E}\big[ag(X) + bh(X)\big] &= \int_{-\infty}^{\infty} \big\{ag(x) + bh(x)\big\}\mathsf{p}^X(x)\,\mathrm{d}x \\
&= a\int_{-\infty}^{\infty} g(x)\mathsf{p}^X(x)\,\mathrm{d}x + b\int_{-\infty}^{\infty} h(x)\mathsf{p}^X(x)\,\mathrm{d}x \\
&= a\mathsf{E}[g(X)] + b\mathsf{E}[h(X)]
\end{aligned}$$

がわかる． □

注意 2.3. $0 < q < r$ に対して

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

となる．

Proof. Young の不等式 (系 3.13)

$$ab \le \frac{a^s}{s} + \frac{b^t}{t}; \quad a,\, b > 0; \quad s,\, t > 0; \quad \frac{1}{t} + \frac{1}{s} = 1$$

において

$$a = |X|^q;\, s = \frac{r}{q}\,(0 < s < 1) \quad b = 1;\, t = \frac{1}{1 - s/2}\,(0 < t < 1)$$

とおくと

$$|X|^q \le \frac{|X|^r}{s} + \frac{1}{t}$$

を得る. 両辺の期待値を取ればよい. □

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.

(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. □

[2] 簡単に「平均」ともいう.

[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.

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

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

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 $g : S_X \to \mathbb{R}$ 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

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

となる．さらに $g(g^{-1}(y)) = y$ より

$$\frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y) = \frac{1}{\overset{\bullet}{g}\,(g^{-1}(y))}$$

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

$$\begin{aligned}\mathsf{p}^Y(y) &= \frac{\mathrm{d}}{\mathrm{d}y}\mathsf{F}^Y(y) = \frac{\mathrm{d}}{\mathrm{d}y}\mathsf{F}^X(g^{-1}(y)) = \overset{\bullet}{\mathsf{F}}{}^X\,(g^{-1}(y))\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y) \\ &= \mathsf{p}^X(g^{-1}(y))\frac{1}{\overset{\bullet}{g}\,\big(g^{-1}(y)\big)}\end{aligned}$$

となる．よって

$$\mathsf{p}^Y(y) = \mathsf{p}^X\big(g^{-1}(y)\big)\frac{1}{\overset{\bullet}{g}\,(g^{-1}(y))}$$

となる．g が狭義単調減少の場合もふくめると

$$\mathsf{p}^Y(y) = \mathsf{p}^X\big(g^{-1}(y)\big)\frac{1}{\big|\,\overset{\bullet}{g}\,\big(g^{-1}(y)\big)\big|}$$

となる．

以上のことから

$$\begin{aligned}\mathsf{E}[Y] &= \int_{S_Y} y\mathsf{p}^Y(y)\,\mathrm{d}y \\ &= \int_{S_Y} y\mathsf{p}^X(g^{-1}(y))\frac{1}{\overset{\bullet}{g}\,\big(g^{-1}(y)\big)}\,\mathrm{d}y \\ &= \int_{S_X} g(x)\mathsf{p}^X(x)\,\mathrm{d}x \qquad \big(\because x = g^{-1}(y) \Leftrightarrow y = g(x);\ \mathrm{d}y = \overset{\bullet}{g}\,(x)\mathrm{d}x\big) \\ &= \mathsf{E}[g(X)]\end{aligned}$$

がわかる． □

[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. まず, $\left|\sum_{j=1}^{n} a_j x_j\right| \leq \sum_{j=1}^{n} a_j |x_j|$ なので, 補題 2.2(1) から $\sum_{j=1}^{n} a_j X_j$ は有限な期待値をもつ. $\boldsymbol{X} := (X_1, X_2, \ldots, X_n)^\top$ とおき, $\boldsymbol{X}$ は同時 p.d.f.

$\mathsf{p}^{\boldsymbol{X}}$ を持つとする．すると定義 2.7 から 補題 2.2(2) と同様の期待値の線型性から

$$\begin{aligned}\mathsf{E}\left[\sum_{j=1}^{n} a_j X_j\right] &= \int_{\mathbb{R}^n} \left\{\sum_{j=1}^{n} a_j x_j\right\} \mathsf{p}^{\boldsymbol{X}}(x_1, x_2, \ldots, x_n)\, \mathrm{d}x_1 \mathrm{d}x_2 \cdots \mathrm{d}x_n \\ &= \sum_{j=1}^{n} a_j \int_{\mathbb{R}^n} x_j \mathsf{p}^{\boldsymbol{X}}(x_1, x_2, \ldots, x_n)\, \mathrm{d}x_1 \mathrm{d}x_2 \cdots \mathrm{d}x_n \\ &= \sum_{j=1}^{n} a_j \mathsf{E}[X_j]\end{aligned}$$

がわかる． □

例 2.10. $0 < p < 1$ と $j = 1, 2, \ldots, n$ に対して，$X_j \sim \mathsf{Ber}(p)$ 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 $a, b \in \mathbb{R}$ 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

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

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 $X_1, X_2, \ldots, X_n$ は互いに独立とする. 定理 2.9 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

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

[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.

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

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

で定義する．このとき

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

となる．

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

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

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

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

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 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

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

をみたす.
(5) ある定数 $a, b \, (a \neq 0)$ 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

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

となる. ここで, 判別式をとれば

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

がわかる. (5) は (2), (3) および定理 2.11(2) からわかる. (6) は定理 2.11(3) 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

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

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) $\mathsf{Pr}(X\leq 1/2,\, Y\leq 1/2)=\mathsf{Pr}((\pi/3)\leq Z\leq (5/6)\pi)$ を確かめよ.

(3) $\mathsf{Pr}(X\leq 1/2)=\mathsf{Pr}((\pi/3)\leq Z\leq (7/3)\pi)$ 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) $d\geq 2$ 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

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

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

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

[6] これらより, $\mathsf{Pr}(X\leq 1/2, Y\leq 1/2)\neq \mathsf{Pr}(X\leq 1/2)\mathsf{Pr}(Y\leq 1/2)$ 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.

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．

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

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

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．

[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．

[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, $\forall \boldsymbol{a} \in \mathbb{R}^d$ $(\boldsymbol{a} \neq \mathbf{0})$ に対して $\boldsymbol{a}^\top \mathbf{A} \boldsymbol{a} > 0$ が成立するときをいう．

(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. (1) の証明: 期待値の線型性を用いて計算すればよい.
(2) の証明:

$$\mathbf{A} = \begin{bmatrix} \boldsymbol{a}_1^\top \\ \boldsymbol{a}_2^\top \\ \vdots \\ \boldsymbol{a}_k^\top \end{bmatrix}; \qquad \boldsymbol{a}_j \in \mathbb{R}^d \, (j = 1, 2, \ldots, k)$$

とおくと

$$\mathsf{E}\bigl[\mathbf{A}\boldsymbol{X}\bigr] = \mathsf{E}\begin{bmatrix} \boldsymbol{a}_1^\top \boldsymbol{X} \\ \boldsymbol{a}_2^\top \boldsymbol{X} \\ \vdots \\ \boldsymbol{a}_k^\top \boldsymbol{X} \end{bmatrix} = \begin{bmatrix} \mathsf{E}\bigl[\boldsymbol{a}_1^\top \boldsymbol{X}\bigr] \\ \mathsf{E}\bigl[\boldsymbol{a}_2^\top \boldsymbol{X}\bigr] \\ \vdots \\ \mathsf{E}\bigl[\boldsymbol{a}_k^\top \boldsymbol{X}\bigr] \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_1^\top \boldsymbol{\mu} \\ \boldsymbol{a}_2^\top \boldsymbol{\mu} \\ \vdots \\ \boldsymbol{a}_k^\top \boldsymbol{\mu} \end{bmatrix} = \mathbf{A}\boldsymbol{\mu}$$

がわかる. つぎに, 共分散の定義と上の結果から

$$\begin{aligned}
\mathsf{Var}\bigl[\mathbf{A}\boldsymbol{X}\bigr] &= \mathsf{E}\Bigl[\bigl\{\mathbf{A}\boldsymbol{X} - \mathsf{E}\bigl[\mathbf{A}\boldsymbol{X}\bigr]\bigr\}\bigl\{\mathbf{A}\boldsymbol{X} - \mathsf{E}\bigl[\mathbf{A}\boldsymbol{X}\bigr]\bigr\}^\top\Bigr] \\
&= \mathsf{E}\Bigl[\bigl\{\mathbf{A}\boldsymbol{X} - \mathbf{A}\mathsf{E}\bigl[\boldsymbol{X}\bigr]\bigr\}\bigl\{\mathbf{A}\boldsymbol{X} - \mathbf{A}\mathsf{E}\bigl[\boldsymbol{X}\bigr]\bigr\}^\top\Bigr] \\
&= \mathsf{E}\Bigl[\mathbf{A}\bigl\{\boldsymbol{X} - \mathsf{E}\bigl[\boldsymbol{X}\bigr]\bigr\}\bigl\{\boldsymbol{X} - \mathsf{E}\bigl[\boldsymbol{X}\bigr]\bigr\}^\top\mathbf{A}^\top\Bigr] \\
&= \mathbf{A}\mathsf{E}\left[\bigl\{\boldsymbol{X} - \mathsf{E}\bigl[\boldsymbol{X}\bigr]\bigr\}\bigl\{\boldsymbol{X} - \mathsf{E}\bigl[\boldsymbol{X}\bigr]\bigr\}^\top\right]\mathbf{A}^\top \\
&= \mathbf{A}\mathsf{Var}[\boldsymbol{X}]\mathbf{A}^\top
\end{aligned}$$

からわかる. □

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 $\boldsymbol{a}(\neq \mathbf{0})$ 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 $\boldsymbol{\Sigma}$ は, 補題 2.21(1) 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. □

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. 定義 2.1 と同様に関数 g を g^+ と g^- の部分に分けて条件付き期待値を定義をするべきであろうが, $\mathsf{E}[|g(X)|]<\infty$ の仮定のもとでは, $\mathsf{E}[g^+(X)|\,Y=y]<\infty$, $\mathsf{E}[g^-(X)|\,Y=y]<\infty$ が保証されるので, 上記のように簡単に定義をした.

注意 2.26. $\mathsf{E}[X]$ 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 $h(Y)$ は確率変数[9]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 $[8,\mathrm{pp}.181-182]$ を参照のこと. □

[9] $h\mapsto h(y)$ の可測性は測度論の知識が必要となる. このことは定理 $C.66$ から保証される.

例 2.27. 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

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

となる．これより

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

となる． □

定理 2.28. (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

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

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．$\mathsf{E}[Y|\,X=x]=:h(x)$ とおいたとき, $\mathsf{E}[Y|\,X]:=h(X)$ と書いたことを思い出す．また, $S_X=\{x\in\mathbb{R};\, \mathsf{p}^X(x)>0\}$ とおいたとき,

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

となる[10]．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[11]が知られている．(3) と (4) も同様に証明できる． □

問 2.5. 定理 2.28(3)(4) の証明を書け．

例 2.29. (例 2.27 の続き)

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

となる．一方 $0<y<x<1$ に対して

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

[10] h は $S_Y:=\{y\in\mathbb{R};\,\mathsf{p}^Y(y)>0\}$ においてのみ定義されるが, 形式的に $h(y)=0\,(y\in S_Y)$ と考えて上の式を理解すればよい．

[11] Fubini の定理からわかる．

となる．よって

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

となる． □

定義 2.30. 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

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

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 2.31. 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.

$$\begin{aligned}
\mathsf{E}\big[\{X-\mathsf{E}[X|Y]\}^2\big] &= \mathsf{E}\big[X^2-2X\mathsf{E}[X|Y]+\{\mathsf{E}[X|Y]\}^2\big] \\
&= \mathsf{E}\Big[\mathsf{E}\big[X^2-2X\mathsf{E}[X|Y]+\{\mathsf{E}[X|Y]\}^2\big]\Big|\,Y\Big] \\
&\quad (\because \text{ 定理 } 2.28(3)) \\
&= \mathsf{E}\Big[\mathsf{E}\big[X^2\big|\,Y\big]-2\mathsf{E}[X|Y]\,\mathsf{E}[X|Y]+\{\mathsf{E}[X|Y]\}^2\Big] \\
&= \mathsf{E}\Big[\mathsf{E}[X^2|Y]-\{\mathsf{E}[X|Y]\}^2\Big] \\
&= \mathsf{E}\big[\mathsf{Var}[X|Y]\big], \\
\mathsf{E}\Big[\{\mathsf{E}[X|Y]-\mathsf{E}[X]\}^2\Big] &= \mathsf{E}\Big[\big\{\mathsf{E}[X|Y]-\underbrace{\mathsf{E}[\mathsf{E}[X|Y]]}_{=\mathsf{E}[X]}\big\}^2\Big] \\
&\quad (\because \text{ 定理 } 2.28(3)) \\
&= \mathsf{Var}\Big[\mathsf{E}[X|Y]\Big], \\
\mathsf{E}\big[\{X-\mathsf{E}[X|Y]\}\{\mathsf{E}[X|Y]-\mathsf{E}[X]\}\big] &= \mathsf{E}\Big[\mathsf{E}\big[\{X-\mathsf{E}[X|Y]\}\{\mathsf{E}[X|Y]-\mathsf{E}[X]\}\Big|\,Y\Big] \\
&\quad (\because \text{ 定理 } 2.28(3)) \\
&= \mathsf{E}\Big[\{\mathsf{E}[X|Y]-\mathsf{E}[X]\}\,\underbrace{\mathsf{E}\big[X-\mathsf{E}[X|Y]\big|Y\big]}_{=0}\Big] \\
&= 0
\end{aligned}$$

となる. 最後から 2 番目の等号は定理 2.28(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 2.32. 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 (m.g.f.)) を

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

と定義する.

注意 2.33. 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[12]が知られている．このことから

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

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 2.34. $X \sim \mathsf{Exp}(1)$ 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． □

補題 2.35. (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．

[12] たとえば, [33, pp.75-76] を参照のこと．

Proof. (1) は指数関数と期待値の性質よりわかる. (2) は $X_j\,(j=1,\,2,\,\ldots,\,d)$ の独立性と指数関数の性質に注意して, 定理 2.11(3) を適用するとわかる. □

定理 2.36. 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

$$X\overset{d}{=}Y\ \Leftrightarrow\ \mathsf{F}^X(x)=\mathsf{F}^Y(x)\,(\forall x\in\mathbb{R})$$

である.

Proof. これは信じることにする. □

注意 2.37. 定理 2.36 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 2.38. $n_1,\,n_2\in\mathbb{N},\,0<p<1$ 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 2.39. $\lambda_1,\,\lambda_2>0$ 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 [30] 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) $X \sim \chi_n^2$ 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 $\mathsf{E}[X_2 - X_1]$ 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)

$$\mathsf{Var}[\boldsymbol{X}] = \mathsf{E}[\boldsymbol{X}\boldsymbol{X}^\top] - \mathsf{E}[\boldsymbol{X}]\bigl\{\mathsf{E}[\boldsymbol{X}]\bigr\}^\top$$

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) $\det\bigl[\mathsf{Var}[\boldsymbol{X}]\bigr]$ 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.

$$\mathsf{p}^{(X,\,Y)}(x,\,y) = \begin{cases} \dfrac{3}{8} \times y & (0 \leq x \leq y \leq 2) \\ 0 & (\text{その他の場合}) \end{cases}$$

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．

$\boxed{\text{補足}}$ $\mathsf{Pr}(0 < Y < 1) = 1$ 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

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

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 2.12. $(U,\, Y)$ 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

$$\mathsf{E}\Big[\big(Y - g(U)\big)^2\Big] \ge 25$$

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

$$\mathsf{Pr}\big(|X-\mu| \geq t\big) \leq \frac{\sigma^2}{t^2}$$

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

$$\mathsf{Pr}\big(|X-\mu| \geq t\big) = \mathsf{Pr}\big((X-\mu)^2 \geq t^2\big) \leq \frac{\mathsf{E}[(X-\mu)^2]}{t^2}$$

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

$$\mathsf{E}[X] = 0, \quad a \leq X \leq b$$

とみたすとき $\forall \lambda > 0$ に対して

$$\mathsf{E}\big[e^{\lambda X}\big] \leq e^{\lambda^2(b-a)^2/8}$$

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

$$\begin{aligned}
\mathsf{E}\big[e^{\lambda X}\big] &\leq -\frac{a}{b-a}e^{\lambda b} + \frac{b}{b-a}e^{\lambda a} \\
&= ce^{\lambda b} + (1-c)e^{\lambda a} \quad \left(c := -\frac{a}{b-a},\ 1-c = \frac{b}{b-a}\right) \\
&= ce^{\lambda(1-c)(b-a)} + (1-c)e^{-\lambda c(b-a)} \\
&= e^{-\lambda c(b-a)}\big\{1-c+ce^{\lambda(b-a)}\big\} \\
&= e^{-cu}\big\{1-c+ce^{u}\big\} \quad (u := \lambda(b-a)) \\
&= \exp\big\{-cu + \log\big(1-c+ce^{u}\big)\big\} =: \exp\big\{g(u)\big\}
\end{aligned}$$

がわかる．いま

$$g(0) = \overset{\bullet}{g}(0) = 0, \quad \overset{\bullet\bullet}{g}(u) \leq \frac{1}{4}\,(u > 0) \tag{3.1}$$

に注意する．実際

$$\begin{aligned}
\overset{\bullet}{g}(u) &= \frac{\mathrm{d}g}{\mathrm{d}u} = -c + \frac{ce^u}{1-c+ce^u}, \\
\overset{\bullet\bullet}{g}(u) &:= \frac{\mathrm{d}^2 g}{\mathrm{d}u^2} = \frac{ce^u}{1-c+ce^u} - \frac{c^2 e^{2u}}{(1-c+ce^u)^2} \\
&= \frac{ce^u(1-c)}{(1-c+ce^u)^2} \leq \frac{1}{4} \qquad \text{（相加相乗平均を } 1-c \text{ と } ce^u \text{ に適用する．）}
\end{aligned}$$

からわかる．(3.1) に注意して, Taylor 展開をすれば

$$\begin{aligned}
g(u) &= g(0) + u\,\overset{\bullet}{g}(0) + \frac{u^2}{2}\ddot{g}(\xi) \quad (\xi \in (0,\,u)) \\
&= \frac{u^2}{2}\,\overset{\bullet\bullet}{g}(\xi) \leq \frac{u^2}{8} = \frac{\lambda^2(b-a)^2}{8}
\end{aligned}$$

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

$$\mathsf{Pr}\left(\sum_{j=1}^n X_j \geq t\right) \leq \exp\left\{-\frac{2t^2}{\sum_{j=1}^n (b_j - a_j)^2}\right\}$$

が成立する．

Proof. 系 3.2 を用いる．$\forall\lambda > 0$ に対して

$$\begin{aligned}
\mathsf{Pr}\left(\sum_{j=1}^{n} X_j \geq t\right) &\leq e^{-\lambda t}\mathsf{E}\left[\exp\left(\lambda\sum_{j=1}^{n} X_j\right)\right] \\
&= e^{-\lambda t}\prod_{j=1}^{n}\mathsf{E}\big[\exp\{\lambda X_j\}\big] \quad (\because \text{独立性}) \\
&\leq e^{-\lambda t}\prod_{j=1}^{n}\exp\left\{\frac{\lambda^2(b_j-a_j)^2}{8}\right\} \quad (\because \text{補題 } 3.4) \\
&= \exp\left\{-\lambda t + \frac{\lambda^2}{8}\sum_{j=1}^{n}(b_j-a_j)^2\right\} \\
&= \exp\left[\frac{\sum_{j=1}^{n}(b_j-a_j)^2}{8}\left\{\lambda - \frac{4t}{\sum_{j=1}^{n}(b_j-a_j)^2}\right\}^2\right. \\
&\qquad \left. - \frac{2t^2}{\sum_{j=1}^{n}(b_j-a_j)^2}\right]
\end{aligned}$$

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

$$\mathsf{Pr}\big(|\overline{X}_n - \theta| \geq t\big) \leq 2e^{-2nt^2}$$

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

$$\mathsf{Pr}\big(\overline{X}_n - \theta \geq t\big) = \mathsf{Pr}\left(\sum_{j=1}^{n} Y_j \geq t\right) \leq e^{-2nt^2} \tag{3.2}$$

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. $X_1, X_2, \ldots, X_{100} \overset{\text{i.i.d.}}{\sim} \mathsf{Ber}(\theta)$ とする．ただし $0 < \theta < 1$ である．
(1) $n = 100$ とし, 事象 $\big|\overline{X}_{100} - \theta\big| \ge 0.2$ 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

$$\mathsf{Pr}\big(\big|\overline{X}_{100} - \theta\big| \ge 0.2\big) \le 0.0625$$

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

$$\mathsf{Pr}\big(\big|\overline{X}_{100} - \theta\big| \ge 0.2\big) \le 2e^{-2(100)(0.2)^2} = 0.00067$$

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

$$\mathsf{Pr}\left(\big|\overline{X}_n - \theta\big| \le \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}\right) \ge 1 - \alpha$$

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
$$\mathsf{Pr}(\theta \in C) \geq 1-\alpha$$
を得る．すなわち信頼係数 $(1-\alpha)$ の θ 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．

$\alpha = 0.05$, $\theta = 1/2$ として, $n = 10$, 20, 50, 100 の信頼区間の比較の計算を行う．CLT の方は数値計算で．

□

注意 3.8. 信頼区間については節 8.6 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. まず
$$\begin{aligned}
\mathsf{Pr}(Z \geq t) &= \frac{1}{\sqrt{2\pi}}\int_t^\infty \exp\left(-\frac{z^2}{2}\right)\mathrm{d}z \\
&\leq \frac{1}{\sqrt{2\pi}}\int_t^\infty \frac{z}{t}\exp\left(-\frac{z^2}{2}\right)\mathrm{d}z \\
&= \frac{1}{t\sqrt{2\pi}}\int_t^\infty \left\{-\frac{\partial}{\partial z}\exp\left(-\frac{z^2}{2}\right)\right\}\mathrm{d}z \\
&= \frac{1}{t\sqrt{2\pi}}\left(-\frac{t^2}{2}\right) \qquad (3.4)
\end{aligned}$$
が成り立つことに注意する．さらに, $\mathsf{N}(0,1)$ の p.d.f. が偶関数であることに注意すると
$$\begin{aligned}
\mathsf{Pr}(|Z| \geq t) &= \mathsf{Pr}(\{Z \geq t\} \cup \{Z \leq -t\}) \\
&\leq \mathsf{Pr}(Z \geq t) + \mathsf{Pr}(Z \leq -t) \\
&= 2\mathsf{Pr}(Z \geq t)
\end{aligned}$$
からわかる．この不等式と (3.4) を合わせると定理の不等式が得られる．
□

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

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

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

$$\begin{aligned}0 \le g(t) &= \mathsf{E}\big[t^2X^2 - 2tXY + Y^2\big] \\ &= \mathsf{E}[X^2]\left\{t - \frac{\mathsf{E}[XY]}{\mathsf{E}[X^2]}\right\}^2 + \frac{\mathsf{E}[X^2]\mathsf{E}[Y^2] - \{\mathsf{E}[XY]\}^2}{\mathsf{E}[X^2]}\end{aligned}$$

となる.

$$\begin{aligned}g\left(\frac{\mathsf{E}[XY]}{\mathsf{E}[X^2]}\right) = \frac{\mathsf{E}[X^2]\mathsf{E}[Y^2] - \{\mathsf{E}[XY]\}^2}{\mathsf{E}[X^2]} \ge 0 \;&\Leftrightarrow\; \mathsf{E}[X^2]\mathsf{E}[Y^2] - \{\mathsf{E}[XY]\}^2 \ge 0 \\ &\Leftrightarrow\; \Big|\mathsf{E}[XY]\Big| \le \sqrt{\mathsf{E}[X^2]\mathsf{E}[Y^2]}\end{aligned} \tag{3.5}$$

を得る. (3.5) において, X, Y を $|X|, |Y|$ と置き換えると定理の不等式は示される. 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 $0 \le \forall t \le 1$ に対して

$$g(tx + (1-t)y) \le tg(x) + (1-t)g(y)$$

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.

[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 $X \ge \epsilon 1\!\!1\{X > \epsilon\}$ より

$$0 = \mathsf{E}[X] \ge \epsilon\mathsf{E}[1\!\!1\{X > \epsilon\}] = \epsilon\mathsf{Pr}(X > \epsilon) > 0$$

となり矛盾する. よって $\mathsf{E}[X] = 0 \Leftrightarrow \mathsf{Pr}(X = 0) = 1$ がわかる.

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.

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

$$g(x) = e^x, \quad h(y) = \begin{cases} p\log a & \left(0 \leq y < \frac{1}{p}\right) \\ q\log b & \left(\frac{1}{p} \leq y \leq 1\right) \end{cases}$$

[2]定理 D.9 を参照.

とおけば g は凸なので

$$\begin{aligned}ab &= \exp\left\{\frac{p\log a}{p} + \frac{q\log b}{q}\right\} = \exp\left\{\int_0^1 h(y)\,\mathrm{d}y\right\} \le \int_0^1 \exp\{h(y)\}\,\mathrm{d}y \\ &= \frac{1}{p}\exp\{p\log a\} + \frac{1}{q}\exp\{q\log b\} = \frac{a^p}{p} + \frac{b^q}{q}\end{aligned}$$

がわかる. □

定理 3.14. (1)(Hölder の不等式) p, q は $1 \le p \le +\infty$, $1 \le q \le +\infty$ と $\dfrac{1}{p} + \dfrac{1}{q} = 1$ とし, 確率変数 X, Y は $\mathsf{E}\big[|X|^p\big] < \infty$, $\mathsf{E}\big[|Y|^q\big] < \infty$ をみたすとする. このとき, $\mathbb{E}\big[|XY|\big] < \infty$ で

$$\mathsf{E}\big[|XY|\big] \le \big\{\mathsf{E}\big[|X|^p\big]\big\}^{1/p}\big\{\mathsf{E}\big[|Y|^q\big]\big\}^{1/q}$$

となる.
(2)(Minkowski の不等式) $1 \le p \le +\infty$ で $\dfrac{1}{p} + \dfrac{1}{q} = 1$ とし, 確率変数 X, Y は $\mathsf{E}\big[|X|^p\big] < \infty$, $\mathsf{E}\big[|Y|^p\big] < \infty$ をみたすとする. このとき, $\mathbb{E}\big[|X+Y|^p\big] < \infty$ で

$$\big\{\mathsf{E}\big[|X+Y|^p\big]\big\}^{1/p} \le \big\{\mathsf{E}\big[|X|^p\big]\big\}^{1/p} + \big\{\mathsf{E}\big[|Y|^p\big]\big\}^{1/p}$$

となる.

Proof. (1) の証明: $1 < p < \infty$, $1 < q < \infty$ の場合についてのみ証明を与える. $p = 1$, $q = \infty$ の場合は Cohen (2010, pp.93-94) を参照のこと.
　系 3.13 から

$$|XY| \le \frac{|X|^p}{p} + \frac{|Y|^q}{q} \tag{3.6}$$

である. まず, $\mathsf{E}[|X|^p] = 1$, $\mathsf{E}[|Y|^q] = 1$ とする. (3.6) の両辺の期待値を取ると

$$\mathsf{E}[|XY|] \le \frac{\mathsf{E}[|X|^p]}{p} + \frac{\mathsf{E}[|Y|^q]}{q} = \frac{1}{p} + \frac{1}{q} = 1 \tag{3.7}$$

を得る. $\mathsf{E}[|X|^p] \ne 1$ または $\mathsf{E}[|Y|^q] \ne 1$ のとき, (3.7) において, X, Y を $X/\{\mathsf{E}[|X|^p]\}^{1/p}$, $Y/\{\mathsf{E}[|Y|^q]\}^{1/q}$ と置き換えると

$$\frac{\mathsf{E}[|\mathsf{XY}|]}{\{\mathsf{E}[|X|^p]\}^{1/p}\{\mathsf{E}[|Y|^q]\}^{1/q}} \le 1$$

を得る. よって, $1 < p < \infty$, $1 < q < \infty$ の場合について (1) は証明された.

(2) の証明: $1 < p < \infty,\, 1 < q < \infty$ の場合についてのみ証明を与える. $p = 1,\, q = \infty$ の場合は Cohen (2010, pp.94-95) を参照のこと. $\mathsf{E}[|X+Y|^p] = 0$ のときは, 不等式は自明なので, $\mathsf{E}[|X+Y|^p] \neq 0$ を仮定しても一般性を失わない. このとき

$$
\begin{aligned}
\mathsf{E}[|X+Y|^p] &\le \mathsf{E}\big[\big(|X|+|Y|\big)|X+Y|^{p-1}\big] \\
&= \mathsf{E}\big[|X| \times |X+Y|^{p-1}\big] + \mathsf{E}\big[|Y| \times |X+Y|^{p-1}\big] \\
&\le \big\{\mathsf{E}[|X|^p]\big\}^{1/p} \{\mathsf{E}[|X+Y|^{q(p-1)}]\}^{1/q} \\
&\quad + \big\{\mathsf{E}[|Y|^p]\big\}^{1/p} \{\mathsf{E}[|X+Y|^{q(p-1)}]\}^{1/q} \qquad (\because (1) \text{ を用いた}) \\
&= \big\{\mathsf{E}[|X|^p]\big\}^{1/p} \{\mathsf{E}[|X+Y|^p]\}^{1/q} + \big\{\mathsf{E}[|Y|^p]\big\}^{1/p} \{\mathsf{E}[|X+Y|^p]\}^{1/q} \\
&= \Big\{\big\{\mathsf{E}[|X|^p]\big\}^{1/p} + \big\{\mathsf{E}[|Y|^p]\big\}^{1/p}\Big\} \{\mathsf{E}[|X+Y|^p]\}^{1/q}
\end{aligned}
$$

を得る. 上の不等式の最左辺と最右辺を $\{\mathsf{E}[|X+Y|^p]\}^{1/q}$ で割れば, $1 < p < \infty,\, 1 < q < \infty$ の場合の (2) が示せた. □

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 [26, 4 章] を参考にした.

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

$$
\mathsf{Pr}\big(X \ge 2\theta\big) \le \frac{1}{\theta}
$$

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[3], 任意の実数 $x_j,\, y_j\ (j = 1, 2, \ldots, n)$ に対して

$$
\sum_{j=1}^{n} |x_j y_j| \le \left(\sum_{j=1}^{n} |x_j|^p\right)^{1/p} \left(\sum_{j=1}^{n} |y_j|^q\right)^{1/q}
$$

[3]問題を易しくするために, $p = 1$ の場合を除いていることに注意せよ.

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. $s > t > 0$ とする．確率変数 X が $\mathsf{E}\big[|X|^s\big] < \infty$ ならば，$\mathbb{E}\big[|X|^t\big] < \infty$ となることを Young の不等式を用いて示せ．

第 4 章　確率変数列と分布列の収束

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

$$\mathsf{Pr}\left(\omega \in \Omega;\ \lim_{n\to\infty} X_n(\omega) = X(\omega)\right) = 1$$

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 $\mathsf{E}\big[X_n^2\big] < \infty$, $\mathsf{E}[X^2] < \infty\,(n=1,2,\ldots)$ 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

$$\lim_{n\to\infty} \mathsf{E}\big[(X_n - X)^2\big] = 0$$

が成り立つときをいう．このことを $X_n \overset{\text{q.m.}}{\to} X$ 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

$$X_n \overset{\text{a.s.}}{\to} X \;\Leftrightarrow\; M_n \overset{\mathrm{P}}{\to} 0$$

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．そのために, $\ell \in \mathbb{N}$ 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

$$B_1 \supset B_2 \supset \cdots \supset B_\ell \supset \cdots \qquad \Bigl(\Rightarrow \mathsf{Pr}(B_1) \geq \mathsf{Pr}(B_2) \geq \cdots \geq \mathsf{Pr}(B_\ell) \geq \cdots\Bigr)$$

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, $\left\{\omega \in \Omega;\, M_n(\omega) < \frac{1}{\ell}\right\}_{n=1}^{\infty}$ 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

$$X_n \xrightarrow{\text{a.s.}} X \;\Leftrightarrow\; M_n \xrightarrow{\text{P}} 0 \tag{4.1}$$

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 (極大) 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

$$X_n \xrightarrow{\text{P}} 0 \;\Leftrightarrow\; p_n \xrightarrow{n\to\infty} 0$$

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

$$M_n = \sup_{k \geq n} |X_k - 0|$$

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 $X_n \xrightarrow{\text{a.s.}} 0$ から $\pi_n \xrightarrow{n\to\infty} 0$ がわかる．ここで

$$\prod_{k=n}^{\infty}(1 - p_k) \to 1\,(n \to \infty) \Leftrightarrow \sum_{n=1}^{\infty} p_n < \infty\,(n \to \infty)$$

であること[1]に注意すると

$$X_n \xrightarrow{\text{a.s.}} 0 \;\Leftrightarrow\; \sum_{n=1}^{\infty} p_n < \infty\,(n \to \infty)$$

[1]たとえば, [28, pp.91 − 92] を参照のこと．

がわかる．

たとえば, $p_n = \dfrac{1}{n}\,(n \in \mathbb{N})$ とすると

$$X_n \overset{\mathrm{P}}{\to} 0$$

だが

$$\mathsf{Pr}\left(\left\{\omega \in \Omega;\ \lim_{n\to\infty} X_n(\omega) = 0\right\}\right) < 1 \ \Leftrightarrow\ X_n \overset{\mathrm{a.s.}}{\to} 0 \text{ は偽}$$

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. Chebyshev の不等式 (系 3.3) より, $\forall \epsilon > 0$ に対して

$$\mathsf{Pr}\big(\big|X_n\big| \geq \epsilon\big) = \mathsf{Pr}\big(X_n^2 \geq \epsilon^2\big) \leq \frac{\mathsf{E}\big[X_n^2\big]}{\epsilon^2} = \frac{1}{n\epsilon^2} \overset{n\to\infty}{\longrightarrow} 0$$

となる. 最後の等号は, $X_n \sim \mathsf{N}(0, 1/n)$ なので, $\mathsf{E}[X_n^2] = \dfrac{1}{n}$ となることを用いた. よって $X_n \overset{\mathrm{P}}{\to} 0$ 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) $X_n \overset{\mathrm{q.m.}}{\to} X \ \Rightarrow\ X_n \overset{\mathrm{P}}{\to} X$ である.

(2) $X_n \overset{\mathrm{P}}{\to} X \ \Rightarrow\ X_n \rightsquigarrow X$ 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 $X_n \overset{\mathrm{P}}{\to} c$ である.

Proof. 節 4.5.3 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

$$\begin{aligned}\mathsf{Pr}\big(\big|X_n\big| \geq \epsilon\big) &= \mathsf{Pr}\big(\sqrt{n}\mathbb{1}_{(0,\,1/n)}(U) \geq \epsilon\big) = \mathsf{Pr}\left(0 < U < \frac{1}{n}\right)\\ &= \mathsf{Pr}\left(0 \leq U \leq \frac{1}{n}\right) = \frac{1}{n} \overset{n\to\infty}{\longrightarrow} 0\end{aligned}$$

となる. よって $X_n \overset{\mathrm{P}}{\to} X$ 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, $X_n \overset{\mathrm{q.m.}}{\to} X$ 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 4.4.5.4 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) $X_n \xrightarrow{\text{q.m.}} X$ かつ $Y_n \xrightarrow{\text{q.m.}} Y \ \Rightarrow\ X_n + Y_n \xrightarrow{\text{q.m.}} X + Y$ 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) $X_n \xrightarrow{\mathrm{P}} X \ \Rightarrow\ g(X_n) \xrightarrow{\mathrm{P}} g(X)$ である.

(7) $X_n \rightsquigarrow X \ \Rightarrow\ g(X_n) \rightsquigarrow g(X)$ である.

Proof. 節 4.5.5 で示す. □

定義 4.12. o_P と O_P の記号をここで導入すること (2024/05/29 記).

4.2 大数の法則

定理 4.13. (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

$$\overline{X}_n := \frac{1}{n}(X_1 + X_2 + \cdots + X_n) \xrightarrow{\mathrm{P}} \mu, \qquad \mu = \mathsf{E}[X_1]$$

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. 定理の仮定のもとの証明は節 4.5.1 で与える. $\mathsf{Var}[X_1] = \sigma^2$ とおく. Chebyshev の不等式 (系 3.3) より

$$\mathsf{Pr}\big(|\overline{X}_n - \mu| \geq \epsilon\big) \leq \frac{\mathsf{Var}\big[\overline{X}_n\big]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \xrightarrow{n\to\infty} 0$$

よりわかる.

仮定のもとでの証明は, 節 4.5.1 を参照せよ. □

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 4.14. X_1, X_2, ... 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 4.5.2 で行う. □

定理 4.15. (大数の強法則) X_1, X_2, ... は確率空間 $\bigl(\Omega, \mathcal{A}, \mathsf{Pr}\bigr)$ 上の i.i.d. 確率変数列とし, $\mathsf{E}\bigl[|X_1|^4\bigr] < \infty$ とする. このとき

$$\overline{X}_n := \frac{1}{n}\bigl(X_1 + X_2 + \cdots + X_n\bigr) \overset{\text{a.s.}}{\to} \mu, \qquad \mu = \mathsf{E}[X_1] \tag{4.3}$$

が成立する.

Proof. まず

$$\mathsf{E}\bigl[|X_1|^4\bigr] =: K < \infty, \quad T_n = X_1 + X_2 + \cdots + X_n, \quad \overline{X}_n = \frac{T_n}{n}$$

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

$$\left\{\omega \in \Omega;\ \sum_{n=1}^{\infty} \bigl(\overline{X}_n(\omega)\bigr)^4 < \infty\right\} \subset \left\{\omega \in \Omega;\ \lim_{n\to\infty} \overline{X}_n(\omega) = 0\right\}$$

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

$$\big(X_1 + \cdots + X_n\big)^4 = \sum_{\substack{\ell_1+\cdots+\ell_n=4 \\ \ell_j \geq 0\,(j=1,\ldots,n)}} \frac{4!}{\ell_1! \times \ell_2! \times \cdots \times \ell_n!} X_1^{\ell_1} \times X_2^{\ell_2} \times \cdots \times X_n^{\ell_n}$$

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

$$\mathsf{E}\big[T_n^4\big] = \sum_{\substack{\ell_1+\ell_2+\cdots+\ell_n=4 \\ \ell_j \geq 0\,(j=1,2,\ldots,n)}} \frac{4!}{\ell_1!\ell_2! \times \cdots \times \ell_n!} \mathsf{E}\big[X_1^{\ell_1}\big] \times \mathsf{E}\big[X_2^{\ell_2}\big] \times \cdots \times \mathsf{E}\big[X_n^{\ell_n}\big]$$

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

$$\mathsf{E}\big[T_n^4\big] = \sum_{k=1}^{n} \mathsf{E}\big[X_k^4\big] + 6 \sum_{1 \leq k < \ell \leq n} \mathsf{E}\big[X_k^2\big]\,\mathsf{E}\big[X_\ell^2\big]$$

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

$$\mathsf{E}\big[X_k^2\big] \leq \sqrt{\mathsf{E}\big[X_k^4\big]} = \sqrt{K}$$

となることに注意すると

$$\mathsf{E}\big[T_n^4\big] \leq nK + 6 \sum_{1 \leq k < \ell \leq n} \sqrt{K}\,\sqrt{K} = nK + 3n(n-1)K \leq 3Kn^2$$

を得る．よって

$$\sum_{n=1}^{\infty} \mathsf{E}\big[\overline{X}_n^4\big] = \sum_{n=1}^{\infty} \frac{1}{n^4} \mathsf{E}\big[T_n^4\big] \leq 3K \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$$

となる．したがって 補題 4.14 から (4.4) が成立することがわかる．

つぎに $\mu \neq 0$ の場合を示す．$Y_k = X_k - \mu\,(k = 1, 2, \ldots, n)$ とおくと

$$\mathsf{E}\big[Y_k^4\big] \leq \mathsf{E}\Big[\big\{|X_k| + |\mu|\big\}^4\Big] \leq 8\mathsf{E}\big[|X_k|^4 + |\mu|^4\big] \leq 8\big(K + \mu^4\big) < \infty$$

が成り立つ．したがってこの定理の証明の前半部分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

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{n}\big(Y_1(\omega) + Y_2(\omega) + \cdots + Y_n(\omega)\big) = 0 \\ &\Leftrightarrow \lim_{n\to\infty} \frac{1}{n}\big(X_1(\omega) + X_2(\omega) + \cdots + X_n(\omega)\big) = \mu \end{aligned}$$

に注意すればよい． □

注意 4.16. 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 4.17 (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

$$B_n(x) = \sum_{j=1}^{n} f\left(\frac{j}{n}\right)\begin{pmatrix} n \\ j \end{pmatrix} x^j(1-x)^{n-j}$$

とおいたとき

$$f(x) = \lim_{n\to\infty} B_n(x) \tag{4.5}$$

が成立する．

Proof. $x = 0, 1$ のとき (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

$$\begin{aligned}
&\mathsf{Pr}\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - x\right| \geq \epsilon\right) \\
&\quad \leq \frac{\mathsf{Var}\left(\dfrac{X_1 + X_2 + \cdots + X_n}{n}\right)}{\epsilon^2} \\
&\quad = \frac{\dfrac{1}{n^2}\big(\mathsf{Var}(X_1) + \mathsf{Var}(X_2) + \cdots + \mathsf{Var}(X_n)\big)}{\epsilon^2} \\
&\quad = \frac{x(1-x)}{n\epsilon^2}
\end{aligned}$$

となる．したがって

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \xrightarrow{P} x$$

[2] $f(0) = a$, $f(1) = b$ に対して，$\widetilde{f}(x) = f(x) + \big\{f(b) - f(a)\big\}x + f(a)$ 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．

がわかる. さらに, 定理 4.7(2) より

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \rightsquigarrow x$$

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 4.18. $X_1, X_2, \ldots$ 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 4.19. $X_1, X_2, \ldots$ 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 4.18) 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 4.13) より

$$\frac{1}{n}\sum_{j=1}^{n}\left(X_j-\mu\right)^2 \xrightarrow{\mathrm{P}} \mathsf{E}\left[\left(X_1-\mu\right)^2\right] = \sigma^2, \tag{4.7}$$

$$\overline{X}_n \xrightarrow{\mathrm{P}} \mu \tag{4.8}$$

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 4.20. (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 4.21. 証明についてのコメントを書くこと． □

定理 4.22. $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ 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

$$\boldsymbol{X}_j = \begin{pmatrix} X_{1j} \\ X_{2j} \\ \vdots \\ X_{dj} \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_d \end{pmatrix} = \begin{pmatrix} \mathsf{E}[X_{1j}] \\ \mathsf{E}[X_{2j}] \\ \vdots \\ \mathsf{E}[X_{dj}] \end{pmatrix}$$

とし，

$$\mathsf{Var}[\boldsymbol{X}_1] = \boldsymbol{\Sigma} = \begin{pmatrix} \mathsf{Var}[X_{11}] & \mathsf{Cov}[X_{11}, X_{21}] & \cdots & \mathsf{Cov}[X_{11}, X_{d1}] \\ \mathsf{Cov}[X_{21}, X_{11}] & \mathsf{Var}[X_{21}] & \cdots & \mathsf{Cov}[X_{21}, X_{d1}] \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{Cov}[X_{d1}, X_{11}] & \mathsf{Cov}[X_{d1}, X_{21}] & \cdots & \mathsf{Var}[X_{d1}] \end{pmatrix}$$

は正定値とする．このとき任意の $\boldsymbol{c} \in \mathbb{R}^d$ に対して

$$\sqrt{n}\boldsymbol{c}^\top\big(\overline{\boldsymbol{X}}_n - \boldsymbol{\mu}\big) \rightsquigarrow \mathsf{N}\big(0, \boldsymbol{c}^\top\boldsymbol{\Sigma}\boldsymbol{c}\big)$$

が成立する．このことを

$$\sqrt{n}\big(\overline{\boldsymbol{X}}_n - \boldsymbol{\mu}\big) \rightsquigarrow \mathsf{N}_d\big(\boldsymbol{0}, \boldsymbol{\Sigma}\big)$$

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 $g : \mathbb{R} \to \mathbb{R}$ に対して $g(Y_n)$ の極限分布を求めよう．

定理 4.23. (デルタ法) $Y_1, Y_2, \ldots$ を確率変数列とし

$$\frac{\sqrt{n}\big(Y_n - \mu\big)}{\sigma} \rightsquigarrow \mathsf{N}(0, 1),$$

とし，g は $x = \mu$ の近傍で連続微分可能な関数で $\overset{\bullet}{g}(\mu) \neq 0$ とする．ただし，$\mu \in \mathbb{R}$, $0 < \sigma < \infty$, $\overset{\bullet}{g}(t) = \dfrac{\mathrm{d}g}{\mathrm{d}t}(t)$ である．このとき

$$\frac{\sqrt{n}\big(g(Y_n) - g(\mu)\big)}{\big|\overset{\bullet}{g}(\mu)\big|\sigma} \rightsquigarrow \mathsf{N}(0, 1)$$

が成立する．すなわち

$$Y_n \approx \mathsf{N}\left(\mu,\, \frac{\sigma^2}{n}\right) \Rightarrow g(Y_n) \approx \mathsf{N}\left(g(\mu),\, \big(\dot{g}(\mu)\big)^2 \frac{\sigma^2}{n}\right)$$

である．ただし「≈」は「分布が近似できる」の意味である．

Proof. 節 4.4.5.5 で証明をする． □

例 4.24. $X_1, X_2, \ldots$ 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 4.18) 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 4.23) 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 4.25. (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

$$\sqrt{n}\big(\boldsymbol{Y}_n - \boldsymbol{\mu}\big) \rightsquigarrow \mathsf{N}_d\big(\mathbf{0}_d,\, \boldsymbol{\Sigma}\big)$$

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 $g : \mathbb{R}^d \to \mathbb{R}$ 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 4.26.

$$\begin{pmatrix} X_{11} \\ X_{21} \end{pmatrix}, \begin{pmatrix} X_{12} \\ X_{22} \end{pmatrix}, \ldots, \begin{pmatrix} X_{1n} \\ X_{2n} \end{pmatrix}$$

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 4.22) より

$$\sqrt{n}\begin{pmatrix} \overline{X}_1 - \mu_1 \\ \overline{X}_2 - \mu_2 \end{pmatrix} \rightsquigarrow \mathsf{N}_2\big(\mathbf{0}_2,\, \boldsymbol{\Sigma}\big)$$

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

$$\sqrt{n}\big(\overline{X}_1\overline{X}_2 - \mu_1\mu_2\big) \rightsquigarrow \mathsf{N}\big(0,\, \mu_2^2\sigma_{11} + 2\mu_1\mu_2\sigma_{12} + \mu_1^2\sigma_{22}\big)$$

がわかる. □

4.5 定理 4.13, 補題 4.14, 定理 4.7, 4.9, 4.11, 4.23 の証明

4.5.1 定理 4.13 の証明

任意の $\delta > 0$ と n に対して

$$Y_j := Y_j(n) = \begin{cases} X_j & (|X_j| \leq \delta n) \\ 0 & (|X_j| > \delta n) \end{cases}$$

$$Z_j := Z_j(n) = \begin{cases} 0 & (|X_j| \leq \delta n) \\ X_j & (|X_j| > \delta n) \end{cases} ; \quad (j = 1, 2, \ldots, n)$$

と定める. 明らかに

$$X_j = Y_j + Z_j, \qquad (j = 1, 2, \ldots, n)$$

である. 以下では, $X_j\,(j = 1, 2, \ldots, n)$ は連続型確率変数とし, 共通の p.d.f. p を持つとして議論を進めていく. このとき

$$\begin{aligned} \mathsf{Var}[Y_j] &= \mathsf{Var}[Y_1] = \mathsf{E}[Y_1^2] - \bigl(\mathsf{E}[Y_1]\bigr)^2 \leq \mathsf{E}[Y_1^2] \\ &= \mathsf{E}\bigl[X_1^2 \mathbb{1}_{[-\delta n,\, \delta n]}(X_1)\bigr] \\ &= \int_{-\infty}^{\infty} x^2 \mathbb{1}_{[-\delta n,\, \delta n]}(x)\mathsf{p}(x)\,\mathrm{d}x = \int_{-\delta n}^{\delta n} x^2 \mathsf{p}(x)\,\mathrm{d}x \\ &\leq \delta n \int_{-\delta n}^{\delta n} |x|\mathsf{p}(x)\,\mathrm{d}x \leq \delta n \int_{-\infty}^{\infty} |x|\mathsf{p}(x)\,\mathrm{d}x \\ &= \delta n \mathsf{E}[X_1] \end{aligned}$$

となる. すなわち

$$\mathsf{Var}[Y_j] \leq \delta n \mathsf{E}[|X_1|] \qquad (j = 1, 2, \ldots, n) \tag{4.11}$$

である. 次に

$$\mathsf{E}[Y_j] = \mathsf{E}[Y_1] = \mathsf{E}[X_1 \mathbb{1}_{[-\delta n,\, \delta n]}(X_1)] = \int_{-\infty}^{\infty} x \mathbb{1}_{[-\delta n,\, \delta n]}(x)\mathsf{p}(x)\,\mathrm{d}x$$

と書ける. ここで

$$\Bigl| x \mathbb{1}_{[-\delta n,\, \delta n]}(x)\mathsf{p}(x) \Bigr| \leq |x|\mathsf{p}(x), \qquad x \mathbb{1}_{[-\delta n,\, \delta n]}(x)\mathsf{p}(x) \xrightarrow{n\to\infty} x\mathsf{p}(x)$$

かつ

$$\int_{-\infty}^{\infty} |x|\mathsf{p}(x)\,\mathrm{d}x < \infty$$

であることに注意して, 優収束定理 (定理 C.51) を用いると

$$\mathsf{E}[Y_1] = \int_{-\infty}^{\infty} x\mathbb{1}_{[-\delta n,\, \delta n]}(x)\mathsf{p}(x)\,\mathrm{d}x \overset{n\to\infty}{\longrightarrow} \int_{-\infty}^{\infty} x\mathsf{p}(x)\,\mathrm{d}x$$

となることがわかる. したがって

$$\mathsf{E}[Y_1] \overset{n\to\infty}{\longrightarrow} \mu \tag{4.12}$$

である. 以上のことを踏まえると任意の $\epsilon > 0$ に対して

$$\begin{aligned}\mathsf{Pr}\left(\left|\frac{1}{n}\sum_{j=1}^{n} Y_j - \mathsf{E}[Y_1]\right| \geq \epsilon\right) &= \mathsf{Pr}\left(\left|\sum_{j=1}^{n} Y_j - \mathsf{E}\left[\sum_{j=1}^{n} Y_j\right]\right| \geq n\epsilon\right)\\ &\leq \frac{1}{n^2\epsilon^2}\mathsf{Var}\left[\sum_{j=1}^{n} Y_j\right] \qquad (\because \text{系 } 3.3)\\ &= \frac{n}{n^2\epsilon^2}\mathsf{Var}[Y_1] \leq \frac{n\delta n\mathsf{E}[|X_1|]}{n^2\epsilon^2} \qquad (\because (4.11))\\ &= \frac{\delta}{\epsilon^2}\mathsf{E}[|X_1|]\end{aligned}$$

となる. したがって

$$\mathsf{Pr}\left(\left|\frac{1}{n}\sum_{j=1}^{n} Y_j - \mathsf{E}[Y_1]\right| \geq \epsilon\right) \leq \frac{\delta}{\epsilon^2}\mathsf{E}[|X_1|] \tag{4.13}$$

となる. よって, (4.12) と (4.13) から, 十分大きな n に対して

$$\begin{aligned}\mathsf{Pr}\left(\left|\frac{1}{n}\sum_{j=1}^{n} Y_j - \mu\right| \geq 2\epsilon\right) &= \mathsf{Pr}\left(\left|\frac{1}{n}\sum_{j=1}^{n} Y_j - \mathsf{E}[Y_1] + \mathsf{E}[Y_1] - \mu\right| \geq 2\epsilon\right)\\ &\leq \mathsf{Pr}\left(\left|\frac{1}{n}\sum_{j=1}^{n} Y_j - \mathsf{E}[Y_1]\right| + \left|\mathsf{E}[Y_1] - \mu\right| \geq 2\epsilon\right)\\ &\leq \mathsf{Pr}\left(\left|\frac{1}{n}\sum_{j=1}^{n} Y_j - \mathsf{E}[Y_1]\right| \geq \epsilon\right) + \mathsf{Pr}\left(\left|\mathsf{E}[Y_1] - \mu\right| \geq \epsilon\right)\\ &\leq \frac{\delta}{\epsilon^2}\mathsf{E}[|X_1|]\end{aligned}$$

となる[3]. したがって

$$\mathsf{Pr}\left(\left|\frac{1}{n}\sum_{j=1}^{n} Y_j - \mu\right| \geq 2\epsilon\right) \leq \frac{\delta}{\epsilon^2}\mathsf{E}[|X_1|] \tag{4.14}$$

[3] (4.12) から n を十分大きくとると $|\mathsf{E}[Y_1]-\mu| \geq \epsilon$ となるので, $\mathsf{Pr}(|\mathsf{E}[Y_1]-\mu| \geq \epsilon) = 0$ となる.

となる．さらに, $\mathsf{E}[|X_1|] < \infty$ なので, 任意の $\delta > 0$ と十分大きな n に対して

$$\int_{|x|>\delta n} |x|\mathsf{p}(x)\,\mathrm{d}x < \delta^2$$

とできることに注意する．このことより, $j = 1, 2, \ldots, n$ に対して

$$\begin{aligned}
\mathsf{Pr}(Z_j \neq 0) = \mathsf{Pr}(|Z_j| > \delta n) = \mathsf{Pr}(|X_j| > \delta n) &= \int_{-\infty}^{-\delta n} \mathsf{p}(x)\,\mathrm{d}x + \int_{\delta n}^{\infty} \mathsf{p}(x)\,\mathrm{d}x \\
&= \int_{|x|>\delta n} \mathsf{p}(x)\,\mathrm{d}x < \int_{|x|>\delta n} \frac{|x|}{\delta n}\mathsf{p}(x)\,\mathrm{d}x = \frac{1}{\delta n}\int_{|x|>\delta n} |x|\mathsf{p}(x)\,\mathrm{d}x \\
&< \frac{\delta^2}{\delta n} = \frac{\delta}{n}
\end{aligned} \tag{4.15}$$

となる．したがって, (4.15) から

$$\mathsf{Pr}\Bigl(\sum_{j=1}^{n} Z_j \neq 0\Bigr) = \mathsf{Pr}\Bigl(\bigcup_{j=1}^{n}\{Z_j \neq 0\}\Bigr) = n\mathsf{Pr}\bigl(Z_1 \neq 0\bigr) \leq \delta \tag{4.16}$$

を得る．よって, (4.14) と (4.16) から, 十分大きな n に対して

$$\begin{aligned}
\mathsf{Pr}\Biggl(\biggl|\frac{1}{n}\sum_{j=1}^{n} X_j - \mu\biggr| \geq 4\epsilon\Biggr) &= \mathsf{Pr}\Biggl(\biggl|\frac{1}{n}\sum_{j=1}^{n} Y_j + \frac{1}{n}\sum_{j=1}^{n} Z_j - \mu\biggr| \geq 4\epsilon\Biggr) \\
&\leq \mathsf{Pr}\Biggl(\biggl|\frac{1}{n}\sum_{j=1}^{n} Y_j - \mu\biggr| + \biggl|\frac{1}{n}\sum_{j=1}^{n} Z_j\biggr| \geq 4\epsilon\Biggr) \\
&\leq \mathsf{Pr}\Biggl(\biggl|\frac{1}{n}\sum_{j=1}^{n} Y_j - \mu\biggr| \geq 2\epsilon\Biggr) + \mathsf{Pr}\Biggl(\biggl|\frac{1}{n}\sum_{j=1}^{n} Z_j\biggr| \geq 2\epsilon\Biggr) \\
&\leq \mathsf{Pr}\Biggl(\biggl|\frac{1}{n}\sum_{j=1}^{n} Y_j - \mu\biggr| \geq 2\epsilon\Biggr) + \mathsf{Pr}\Biggl(\sum_{j=1}^{n} Z_j \neq 0\Biggr) \\
&\leq \frac{\delta}{\epsilon^2}\mathsf{E}[|X_1|] + \delta
\end{aligned}$$

となる．ここで, δ を ϵ^3 とおきなおせば, 十分大きな n に対して

$$\mathsf{Pr}\Biggl(\biggl|\frac{1}{n}\sum_{j=1}^{n} X_j - \mu\biggr| \geq 4\epsilon\Biggr) \leq \epsilon\mathsf{E}[|X_1|] + \epsilon^3$$

となる．上式は任意の $\epsilon > 0$ で成立したので, 定理の主張は証明された．
□

4.5.2 補題 4.14 の証明

背理法で証明する．そのために

$$\mathsf{Pr}\Big(\sum_{k=1}^{\infty} X_k < \infty\Big) < 1$$

を仮定する．さらに事象 F と N を

$$F := \Big\{\omega \in \Omega;\ \sum_{k=1}^{\infty} X_k(\omega) < \infty\Big\}, \qquad N := \Big\{\omega \in \Omega;\ \sum_{k=1}^{\infty} X_k(\omega) = \infty\Big\}$$

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

$$0 \le \sum_{k=1}^{\infty} X_k(\omega) < \infty$$

である．したがって, 任意の正の実数 $r \ge 0$ に対して

$$r\mathbb{1}_N(\omega) \le \sum_{k=1}^{\infty} X_k(\omega) \qquad (\forall \omega \in \Omega) \tag{4.17}$$

が成り立つ．(4.17) の両辺の期待値をとる．すると 定義 2.4(4), (4.23) と (4.17) から

$$r \times \mathsf{Pr}(N) = \mathsf{E}[r\mathbb{1}_N] \le \mathsf{E}\Big[\sum_{k=1}^{\infty} X_k\Big] = \sum_{k=1}^{\infty} \mathsf{E}\big[X_k\big] < \infty \tag{4.18}$$

を得る．(4.18) 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, $r \ge 0$ は任意の実数だったので, $r \to \infty$ とすれば, (4.18) の最左辺は $+\infty$ となるので, 矛盾が生じる．したがって, $\mathsf{Pr}(F) = 1$ が成り立つ．　□

4.5.3 定理 4.7 の証明

(1) $X_n \xrightarrow{qm} X$ 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

$$\mathsf{Pr}\big(|X_n - X| \geq \epsilon\big) = \mathsf{Pr}\big(|X_n - X|^2 \geq \epsilon^2\big) \leq \frac{\mathsf{E}\big[|X_n - X|^2\big]}{\epsilon^2} \overset{n\to\infty}{\longrightarrow} 0$$

がわかる．

(2) $X_n \xrightarrow{\mathrm{P}} X$ とする．$\epsilon > 0$ を固定し，x と $x \pm \epsilon$ を X の c.d.f. $\mathsf{F}(x)$ の連続点[4]とする．このとき

$$\begin{aligned}
\mathsf{F}_n(x) &= \mathsf{Pr}\big(X_n \leq x\big) \\
&= \mathsf{Pr}\big(X_n \leq x,\, X < x + \epsilon\big) + \mathsf{Pr}\big(X_n \leq x,\, X \geq x + \epsilon\big) \\
&\leq \mathsf{Pr}\big(X \leq x + \epsilon\big) + \mathsf{Pr}\big(|X_n - X| \geq \epsilon\big) && (4.19) \\
&= \mathsf{F}(x + \epsilon) + \mathsf{Pr}\big(|X_n - X| \geq \epsilon\big) && (4.20)
\end{aligned}$$

となる．同様に

$$\begin{aligned}
\mathsf{F}(x - \epsilon) &= \mathsf{Pr}\big(X \leq x - \epsilon\big) \\
&= \mathsf{Pr}\big(X_n \leq x - \epsilon,\, X_n < x\big) + \mathsf{Pr}\big(X \leq x - \epsilon,\, X_n \geq x\big) \\
&\leq \mathsf{Pr}\big(X_n \leq x\big) + \mathsf{Pr}\big(|X_n - X| \geq \epsilon\big) && (4.21) \\
&= \mathsf{F}_n(x) + \mathsf{Pr}\big(|X_n - X| \geq \epsilon\big) && (4.22)
\end{aligned}$$

となる．(4.20) と (4.22) を合わせると

$$\mathsf{F}(x - \epsilon) - \mathsf{Pr}\big(|X_n - X| \geq \epsilon\big) \leq \mathsf{F}_n(x) \leq \mathsf{F}(x + \epsilon) + \mathsf{Pr}\big(|X_n - X| \geq \epsilon\big)$$

がわかる．上の式の辺々で $n \to \infty$ とすれば

$$\mathsf{F}(x - \epsilon) \leq \liminf_{n\to\infty} \mathsf{F}_n(x) \leq \limsup_{n\to\infty} \mathsf{F}_n(x) \leq \mathsf{F}(x + \epsilon)$$

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．

[4] X の分布関数の不連続点は高々加算個なので，このように点を取れることに注意せよ．

(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[5]．このとき

$$\begin{aligned}\mathsf{Pr}\big(\big|X_n - c\big| \geq \epsilon\big) &= \mathsf{Pr}\big(X_n \leq c - \epsilon\big) + \mathsf{Pr}\big(X_n \geq c + \epsilon\big)\\ &= \mathsf{Pr}\big(X_n \leq c - \epsilon\big) + \mathsf{Pr}\big(X_n \geq c + \epsilon\big)\\ &= \mathsf{F}_n(c-\epsilon) + 1 - \mathsf{F}_n(c+\epsilon)\\ &\to \mathsf{F}(c-\epsilon) + 1 - \mathsf{F}(c+\epsilon) \quad (n \to \infty)\\ &= 0 + 1 - 1 = 0\end{aligned}$$

となる． □

問 4.1. (4.19) と (4.21) を確認せよ．すなわち

$$\begin{aligned}&\mathsf{Pr}\big(X_n \leq x,\, X \geq x + \epsilon\big) \leq \mathsf{Pr}\big(\big|X_n - X\big| \geq \epsilon\big),\\ &\mathsf{Pr}\big(X \leq x - \epsilon,\, X_n \geq x\big) \leq \mathsf{Pr}\big(\big|X_n - X\big| \geq \epsilon\big)\end{aligned}$$

を示せ．

4.5.4 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 $|g(x)| \leq 1/2\,(x \in \mathbb{R})$ 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

$$\mathsf{Pr}(X \in K^c) \leq \epsilon \tag{4.23}$$

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 $K_1 = [a_1,\, b_1],\, K_j = (a_j,\, b_j]\,(j = 2,\, 2,\, \ldots,\, m); a_j,\, b_j \in \mathbb{R})$ 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

$$\max_{j \in \{1,\,2,\,\ldots,\,m\}} \max_{x,\,y \in K_j} |g(x) - g(y)| \leq \epsilon \tag{4.24}$$

となるように $\{K_j\}_{j=1}^m$ をとることができる．さらに, $\{a_j,\, b_j\}_{j=1}^m$ は X の分布関数 F の連続点[6]に取る．このとき, (1) より

$$\mathsf{Pr}(X_n \in K) \overset{n\to\infty}{\longrightarrow} \mathsf{Pr}(X \in K) \tag{4.25}$$

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

$$g_\epsilon(x) = \sum_{j=1}^m g(x_j)\mathbb{1}_{K_j}(x) \qquad (x \in \mathbb{R})$$

[5] $\mathsf{F}(x)$ 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．

[6] F の不連続点は高々可算個なので, このように取れることがわかる．

と定める. すると

$$\max_{x\in K}\bigl|g(x)-g_\epsilon(x)\bigr|\le\epsilon \tag{4.26}$$

となる. したがって, (4.23) と (4.26) に注意すると

$$\begin{aligned}
&\bigl|\mathsf{E}[g(X)]-\mathsf{E}[g_\epsilon(X)]\bigr|\\
&\quad=\bigl|\mathsf{E}[\mathbb{1}_K(X)g(X)]+\mathsf{E}[\mathbb{1}_{K^c}(X)g(X)]-\mathsf{E}[\mathbb{1}_K(X)g_\epsilon(X)]\\
&\qquad-\mathsf{E}[\mathbb{1}_{K^c}(X)g_\epsilon(X)]\bigr|\\
&\quad\le\bigl|\mathsf{E}[\mathbb{1}_K(X)g(X)]-\mathsf{E}[\mathbb{1}_K(X)g_\epsilon(X)]\bigr|+\bigl|\mathsf{E}[\mathbb{1}_{K^c}(X)g(X)]\\
&\qquad-\mathsf{E}[\mathbb{1}_{K^c}(X)g_\epsilon(X)]\bigr|\\
&\quad\le\mathsf{E}\bigl[\mathbb{1}_K(X)\bigl|g(X)-g_\epsilon(X)\bigr|\bigr]+\mathsf{E}\bigl[\mathbb{1}_{K^c}(X)\bigl|g(X)-g_\epsilon(X)\bigr|\bigr]\\
&\quad\le\underbrace{\max_{x\in K}|g(x)-g_\epsilon(x)|}_{\le\epsilon\quad\because(4.26)}\mathsf{E}\bigl[\mathbb{1}_K(X)\bigr]+\mathsf{E}\bigl[\mathbb{1}_{K^c}(X)\bigl|g(X)-g_\epsilon(X)\bigr|\bigr]\\
&\quad\le\epsilon\mathsf{Pr}(X\in K)+\mathsf{Pr}(X\in K^c)\\
&\quad\le2\epsilon
\end{aligned}\tag{4.27}$$

となる. (4.25) から, n を十分大きく取ると

$$\bigl|\underbrace{\mathsf{Pr}(X_n\in K^c)}_{=1-\mathsf{Pr}(X_n\in K)}-\underbrace{\mathsf{Pr}(X\in K^c)}_{=1-\mathsf{Pr}(X\in K)}\bigr|=\bigl|\mathsf{Pr}(X\in K)-\mathsf{Pr}(X\in K)\bigr|\le\epsilon$$

とできるので

$$\mathsf{Pr}(X_n\in K^c)\le\mathsf{Pr}(X\in K^c)+\bigl|\mathsf{Pr}(X_n\in K^c)-\mathsf{Pr}(X\in K^c)\bigr|\le2\epsilon$$

がわかる. 同様に

$$\bigl|\mathsf{E}[g(X_n)]-\mathsf{E}[g_\epsilon(X_n)]\bigr|\le\epsilon\mathsf{Pr}(X_n\in K)+\mathsf{Pr}(X_n\in K^c)\le3\epsilon \tag{4.28}$$

を得る. g_ϵ の作りかたから n を十分大きくとると

$$\begin{aligned}
&\bigl|\mathsf{E}[g_\epsilon(X)]-\mathsf{E}[g_\epsilon(X_n)]\bigr|\\
&\quad\le\sum_{j=1}^{n}\bigl|\mathsf{Pr}(X\in K_j)-\mathsf{Pr}(X_n\in K_j)\bigr|\max_{j=\{1,2,\dots,m\}}|g(x_j)|\\
&\quad\le\frac{\epsilon}{2}<\epsilon
\end{aligned}\tag{4.29}$$

となる. よって, (4.27) – (4.29) を合わせると

$$\begin{aligned}
\bigl|\mathsf{E}[g(X_n)]-\mathsf{E}[g(X)]\bigr|&\le\underbrace{\bigl|\mathsf{E}[g(X_n)]-\mathsf{E}[g_\epsilon(X_n)]\bigr|}_{\le3\epsilon}+\underbrace{\bigl|\mathsf{E}[g_\epsilon(X_n)]-\mathsf{E}[g_\epsilon(X)]\bigr|}_{\le\epsilon}\\
&\quad+\underbrace{\bigl|\mathsf{E}[g_\epsilon(X)]-\mathsf{E}[g(X)]\bigr|}_{\le2\epsilon}\\
&\le6\epsilon
\end{aligned}$$

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

$$\liminf_{n\to\infty} \mathsf{E}[g(X_n)] \geq \liminf_{n\to\infty} \mathsf{E}[g_M(X_n)] = \mathsf{E}[g_M(X)] \qquad \big(\because (2)\big)$$

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

$$\liminf_{n\to\infty} \mathsf{Pr}(X_n \in O) \geq \liminf_{n\to\infty} \mathsf{E}[g_M(X_n)] = \mathsf{E}[g_M(X)]$$

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: $B = (-\infty, x]$ ととればよい. □

4.5.5 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, $\exists\delta > 0$ 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

$$\{\omega \in \Omega : |X(\omega) - X_n(\omega)| < \delta\} \subset \{\omega \in \Omega : |g\big(X(\omega)\big) - g\big(X_n(\omega)\big)| < \epsilon\}$$

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: $g \circ f$ が有界連続関数となるように有界連続関数 f を取る. $X_n \rightsquigarrow X$ 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: 実数 $x, y, a, b \in \mathbb{R}$ に対して, 不等式 $\{(x+y) - (a+b)\}^2 \leq 2\{(x-a)^2 + (y-b)\}$ 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[7]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. □

4.5.6 定理 4.23 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

$$g(x) - g(\mu) = (x - \mu) \int_0^1 \dot{g}\,\big(\mu + t(x - \mu)\big)\,\mathrm{d}t \tag{4.30}$$

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

$$\begin{aligned} \left| \int_0^1 \dot{g}\,\big(\mu + t(x-\mu)\big)\,dt - \dot{g}\,(\mu) \right| &\leq \int_0^1 \big|\dot{g}\,\big(\mu + t(x-\mu)\big) - \dot{g}\,(\mu)\big|\,dt \\ &< \epsilon \int_0^1 dt = \epsilon \end{aligned} \tag{4.31}$$

[7] 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 [33, p.135] を参照のこと.

となる. (4.31) より

$$\mathsf{Pr}\left(\left|\int_0^1 \overset{\bullet}{g}\left(\mu+t(Y_n-\mu)\right)dt - \overset{\bullet}{g}(\mu)\right| \geq \epsilon\right) \leq \mathsf{Pr}\left(\left|Y_n-\mu\right| \geq \delta\right) \tag{4.32}$$

が成立する. $\sqrt{n}\bigl(Y_n-\mu\bigr)/\sigma \rightsquigarrow \mathsf{N}(0,\,1)$ なので $Y_n \xrightarrow{\mathrm{P}} \mu$ である. (4.32) から

$$\int_0^1 \overset{\bullet}{g}\left(\mu+t(Y_n-\mu)\right)\mathrm{d}t \xrightarrow{\mathrm{P}} \overset{\bullet}{g}(\mu) \tag{4.33}$$

が成立する. $|Y_n-\mu|<\delta$ が起こったとき, (4.30) に $x=Y_n$ を代入すれば

$$\frac{\sqrt{n}\bigl(g(Y_n)-g(\mu)\bigr)}{\sigma} = \frac{\sqrt{n}\bigl(Y_n-\mu\bigr)}{\sigma}\int_0^1 \overset{\bullet}{g}\left(\mu+t(Y_n-\mu)\right)\mathrm{d}t \tag{4.34}$$

を得る. さらに, $\sqrt{n}(Y_n-\mu)/\sigma \rightsquigarrow \mathsf{N}(0,\,1)$ と (4.33) に注意して定理 4.11(5) を (4.34) に適用すれば

$$\begin{aligned}
&\frac{\sqrt{n}\bigl(g(Y_n)-g(\mu)\bigr)}{\sigma}\\
&= \frac{\sqrt{n}\bigl(g(Y_n)-g(\mu)\bigr)}{\sigma}\mathbb{1}\{|Y_n-\mu|<\delta\} + \frac{\sqrt{n}\bigl(g(Y_n)-g(\mu)\bigr)}{\sigma}\mathbb{1}\{|Y_n-\mu|\geq\delta\}\\
&= \frac{\sqrt{n}(Y_n-\mu)}{\sigma}\underbrace{\int_0^1 \overset{\bullet}{g}\left(\mu+t(Y_n-\mu)\right)\mathrm{d}t\mathbb{1}\{|Y_n-\mu|<\delta\}}_{\overset{\bullet}{g}(\mu)+o_P(1)}\\
&\quad + \underbrace{\frac{\sqrt{n}\bigl(g(Y_n)-g(\mu)\bigr)}{\sigma}\mathbb{1}\{|Y_n-\mu|\geq\delta\}}_{=o_P(1)}\\
&\rightsquigarrow \overset{\bullet}{g}(\mu)\mathsf{N}(0,\,1)
\end{aligned}$$

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 [15, 33] を参考にした. 注意 4.4 は [15] からの借用である. 第 4.2 節 は [29] を借用した. 第 4.4 節 は [15, 33] を借用した. 定理 4.13 の証明は [21] からの借用である. 定理 4.23 の証明は [9] 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) $\dfrac{k}{n} \leq x < \dfrac{(k+1)}{n}$ $(k = 1, \ldots, n)$ に対して

$$\mathsf{Pr}(X_n \leq x)$$

を求めることで, $x \in \mathbb{R}$ に対して $\mathsf{Pr}(X_n \leq x)$ を求めよ.
(2) $x \in \mathbb{R}$ に対して

$$\mathsf{Pr}(X \leq x)$$

を求めよ.
(3)

$$\lim_{n\to\infty} \left|\mathsf{Pr}(X_n \leq x) - \mathsf{Pr}(X \leq x)\right|$$

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) $\mu \in \mathbb{R}$ を定数とする.

$$\mathsf{E}[(X_n - \mu)^2] = \mathsf{Var}[X_n] + \left\{\mathsf{E}[X_n] - \mu\right\}^2$$

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

$$\lim_{n\to\infty} \mathsf{Pr}\Big(\sqrt{(X_n - X)^2 + (Y_n - Y)^2} \geq \epsilon\Big) = 0$$

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．だたし $\mathcal{B}(\mathbb{X})$ 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．こ

[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．

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. $X_1, X_2, \ldots, X_n$ 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,

$$\Big(\mathbb{R}^n,\ \mathcal{B}(\mathbb{R}^n),\ \{\mathsf{P}_\theta;\ \theta \in \Theta\}\Big)$$

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. 決定関数の集まりを**決定空間**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 $\Theta \times \mathbb{A}$ 上の非負値実数値関数

$$\mathsf{L} : \Theta \times \mathbb{A} \mapsto [0, \infty) \cup \{\infty\}$$

を用意する[11]. この関数を**損失関数**といい, $\mathsf{L}(\theta, a)$ 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 $\mathsf{L}\big(\theta^*, d(\boldsymbol{x})\big)$ 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 $\mathsf{L}\big(\theta^*,\, d(\boldsymbol{X})\big)$ 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 $\mathsf{L}\big(\theta,\, d(\boldsymbol{X})\big)$ を $\mathsf{P}_\theta^{\otimes n}$ に関して期待値を取ったもの

$$\mathsf{R}(\theta,\, d) := \mathsf{E}_\theta\big[\mathsf{L}\big(\theta,\, d(\boldsymbol{X})\big)\big]$$

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

$$\mathsf{R}(\theta,\, d) = \mathsf{E}_\theta\big[\mathsf{L}\big(\theta,\, d(\boldsymbol{X})\big)\big]$$

[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

$$\mathsf{L}\big(\theta, 0\big) = \begin{cases} 0 & (\theta \in \Theta_0) \\ 1 & (\theta \in \Theta_1) \end{cases} ; \qquad \mathsf{L}\big(\theta, 1\big) = \begin{cases} 1 & (\theta \in \Theta_0) \\ 0 & (\theta \in \Theta_1) \end{cases}$$

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 $\mathsf{R}\big(\theta, d\big) = \mathsf{E}_\theta\big[\mathsf{L}\big(\theta, d(\boldsymbol{X})\big)\big]$ は以下のようになる. こ

	$\theta \in \Theta_0$	$\theta \in \Theta_1$
$d = 0$	正しい判断	第 2 種の誤り
$d = 1$	第 1 種の誤り	正しい判断

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

$$\beta(\theta) := \mathsf{R}\big(d,\, \theta\big)$$

とし, $\theta \in \Theta_1$ のとき

$$\beta(\theta) := 1 - \mathsf{R}\big(\theta,\, d\big)$$

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

$$\sup_{\theta\in\Theta_0} \mathsf{R}\big(\theta,\, d\big) \le \alpha$$

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 $\mathsf{R}\big(d,\, \theta\big)$ 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. $\sup_{\theta\in\Theta_0} \mathsf{R}\big(\theta,\, d\big)$ 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

$$\mathsf{L}\big(\theta,\, [\ell,\, u]\big) = (u - \ell) - \mathbb{1}\big\{\theta \in [\ell,\, u]\big\}$$

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

$$\mathsf{R}\big(\theta,\, d\big) = \mathsf{E}_\theta\big[u(\boldsymbol{X}) - \ell(\boldsymbol{X})\big] - \mathsf{Pr}_\theta\big(\theta \in [\ell(\boldsymbol{X}),\, u(\boldsymbol{X})]\big)$$

となる.

実数 $\alpha\,(0 < \alpha < 1)$ が与えられたとき

$$\mathsf{Pr}_\theta\big(\theta \in [\ell(\boldsymbol{X}),\, u(\boldsymbol{X})]\big) \ge 1 - \alpha$$

のもとで区間の長さの期待値 $\mathsf{E}_\theta\big[u(\boldsymbol{X}) - \ell(\boldsymbol{X})\big]$ を短くする区間が望ましい区間といえよう. $1 - \alpha$ 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

$$
\begin{aligned}
&d_1 \prec d_2 \\
&\quad \Leftrightarrow \mathsf{R}(\theta, d_1) \le \mathsf{R}(\theta, d_2)\,(\forall\theta \in \Theta) \text{ かつ } \mathsf{R}(\theta_0, d_1) < \mathsf{R}(\theta_0, d_2)\,(\exists\theta_0 \in \Theta)
\end{aligned}
$$

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 $\mathsf{L}\big(\widehat{\mu}, \mu\big) = \big(\widehat{\mu} - \mu\big)^2$ 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.