

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. 詳しくは [45, 10, 20, 44] を参照のこと. 第 0.1 節では, 集合の言葉使いと数の集合の基礎事項をまとめている. 第 0.2 節では, 写像の性質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. これらのことを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 \emptyset と記し, 集合 X のすべての部分集合をの族を 2^X と記す. すなわち

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

である.

\mathcal{E} を集合 X の部分集合族としたとき

$$\bigcup_{E \in \mathcal{E}} E := \{x \in X; \text{ある } E \in \mathcal{E} \text{ が存在して, } x \in E\},$$

$$\bigcap_{E \in \mathcal{E}} E := \{x \in X; \text{すべての } E \in \mathcal{E} \text{ に対して, } x \in E\}$$

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

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

と表す. $\alpha \neq \beta (\alpha, \beta \in A)$ に対して, $E_\alpha \cap E_\beta = \emptyset$ 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

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

と定める. すると

$$\limsup E_n = \{x \in X; \text{可算個の } n \text{ に対して, } x \in E_n\}$$

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

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

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

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

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

が成立する.

集合 X と Y の直積を $X \times Y$ と書く. すなわち

$$X \times Y = \{(x, y); x \in X, y \in Y\}$$

である. 有限個の集合 X_1, X_2, \dots, 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, $X_1 \times X_2 \times \cdots \times X_n$ 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

$$f: X \rightarrow Y$$

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 $f: X \rightarrow Y$ を

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

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 $\mathbb{1}_A: X \rightarrow \{0, 1\}$ 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 $\iota: X \rightarrow Y$ を

$$\iota(x) = x \quad (x \in X)$$

により定める. ι を**包含写像**という. 特に, $X = Y$ のとき, ι を id_X と表し, **恒等写像**という.

(4) X, Y を空でない集合, $f: X \rightarrow Y$ を写像とし, $A \subset X, A \neq \emptyset$ とする. このとき, 写像 $f|_A: A \rightarrow Y$ を

$$f|_A(x) = f(x) \quad (x \in A)$$

により定める. この $f|_A$ 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, $f: X \rightarrow Y$ を写像とする. このとき, $G(f) \subset X \times Y$ を

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

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, $f: X \rightarrow Y$ を写像とする.

(1) X の部分集合 A に対して, $\{f(x); x \in A\}$ 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, $f: X \rightarrow Y$ 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) $f^{-1}(f(A)) \supset A$.

(10) $f(f^{-1}(B)) \subset B$.

Proof. 証明は [45, 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 $f: \mathbb{R} \rightarrow \mathbb{R}$ を

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

により定める.

(1) $A_1 = (-1, 0]$, $A_2 = [0, 1)$ とすると

$$\begin{aligned} f(A_1 \cap A_2) &= \{0\}; & f(A_1) \cap f(A_2) &= [0, 1); \\ f(A_1 \setminus A_2) &= (0, 1); & f(A_1) \setminus f(A_2) &= \emptyset \end{aligned}$$

となる.

(2) $A = [0, 1)$, $B = (-1, 1)$ とすると

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

となる. □

命題 0.8. X, Y を空でない集合とし, $f: X \rightarrow Y$ 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) ~ (4) 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, $f^{-1}(f(A)) = A$.
- (4) f が全射ならば, $f(f^{-1}(B)) = B$.

Proof. 証明は [45, 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, $f: X \rightarrow Y$, $g: Y \rightarrow Z$ を写像とする. このとき, X から Z への写像 $g \circ f$ を

$$(g \circ f)(x) = g(f(x)) \quad (x \in X)$$

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, $f: X \rightarrow Y$, $g: Y \rightarrow Z$, $h: Z \rightarrow W$ 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 [45, p.40] を参照のこと. □

命題 0.11. X, Y, Z を空でない集合, $f: X \rightarrow Y, g: Y \rightarrow Z$ を写像とする. このとき, (1), \sim (3) 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 [45, 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 [45, 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 $d: \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}$ が次の 3 条件をみたすとき, d を \mathbb{X} 上の距離関数といい, 組 (\mathbb{X}, d) を距離空間という.

- (1) 任意の $x, y \in \mathbb{X}$ に対して, $d(x, y) \geq 0$ である. 特に, $d(x, y) = 0 \Leftrightarrow x = y$ である.
- (2) 任意の $x, y \in \mathbb{X}$ に対して, $d(x, y) = d(y, x)$.
- (3) 任意の $x, y, z \in \mathbb{X}$ に対して, $d(x, z) \leq d(x, y) + d(y, z)$.

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

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

から定まる Euclid の距離 $d(\mathbf{x}, \mathbf{y})$ を

$$d(\mathbf{x}, \mathbf{y}) = |\mathbf{x} - \mathbf{y}|_{2,d} \quad (\mathbf{x}, \mathbf{y} \in \mathbb{R}^d)$$

で定める. □

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

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

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 $f: \mathbb{X} \rightarrow \mathbb{Y}$ 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

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

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, $A = \emptyset$ 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.

をみたすときをいう. 任意の $x \in \mathbb{X}$ で f が連続ならば, 単に f は連続という.

(2) 写像 f が一様連続であるとは, 任意の $\epsilon > 0$ に対して, ある $\delta > 0$ が存在して, 任意の $x, x' \in \mathbb{X}$ に対して

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

がなりたつときをいう.

定義 0.19. X を集合, (Y, d) を距離空間とする. $f, f_1, f_2, \dots, f_n \dots$ を X から 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

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

であることをいう.

(2) X の部分集合 $A (\neq \emptyset)$ に対して, $\{f_n\}_{n \in \mathbb{N}}$ が f に A 上で一様収束するとは

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

がなりたつときをいう.

0.4.2 位相空間

定義 0.20. \mathbb{X} を空でない集合とする. \mathbb{X} 上の集合族 $\mathcal{O} \subset 2^{\mathbb{X}}$ が次の条件 (1) ~ (3) をみたすとき, \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) $\emptyset, \mathbb{X} \in \mathcal{O}$.
- (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 $\text{cl}(A)$ を

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

により定める. また, $\text{cl}(A)$ の元を A の触点という.

(3) $A \subset \mathbb{X}$ とする. $\partial A = \text{cl}(A) \setminus A^\circ$ 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, $\text{cl}(A) = \mathbb{X}$ 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 $f: \mathbb{X} \rightarrow \mathbb{Y}$ 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 $f: \mathbb{X} \rightarrow \mathbb{Y}$ 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, $f: \mathbb{X} \rightarrow \mathbb{Y}$ 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 $\bar{\mathbb{R}}$ と書くことにする. $\bar{\mathbb{R}}$ 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, $a \times (\pm\infty) = \mp\infty, \pm\infty \times a = \mp\infty.$
 - $0 \times (\pm\infty) = 0, \pm\infty \times 0 = 0.$
 - $0/(\pm\infty) = 0.$
- $(\pm\infty) + \pm(\infty) = \pm\infty, (\pm\infty) - (\mp\infty) = \pm\infty.$
- $(\pm\infty) \times (\pm\infty) = +\infty, (\pm\infty) \times (\mp\infty) = -\infty.$
- $(\pm\infty)/(\pm\infty), (\pm\infty)/(\mp\infty)$ は定義されないものとする.
- $|\pm\infty| = +\infty.$
- $\overline{\mathbb{R}}$ の部分集合 $A (\neq \emptyset)$ 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 $[a, b]$ と开区間 (a, b) を

$$[a, b] = \{x \in \overline{\mathbb{R}}; a \leq x \leq b\}, \quad (a, b) = \{x \in \overline{\mathbb{R}}; a < x < b\}$$
 により定める.

0.6 章末注釈と参考文献

第 0.1 節は [44] を借用した. 第 0.2 節は [45, pp.26 – 29] を借用した. 第 0.3 節は [44] を借用した. 第 0.4 節は [44] と [45] を参照した. 第 0.5 節は [44, 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 \emptyset (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) $A_n \in \mathcal{A} (n = 1, 2, \dots) \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

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 (Ω, \mathcal{A}) を可測空間と呼ぶ

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

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

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

Proof. (1) $\emptyset = \Omega^c \in \mathcal{A}$.

(2) $A_n^c \in \mathcal{A} (n = 1, 2, \dots)$ 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. 集合族 \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

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

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

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

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 **冪集合** という. □

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

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

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

- (1) $\text{Pr}(\Omega) = 1$ である.
- (2) 互いに排反⁴な事象列 $A_n \in \mathcal{A} (n = 1, 2, \dots)$ に対して

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

をみたす.

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

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

- (1) $\text{Pr}(\emptyset) = 0$ である.
- (2) $A \in \mathcal{A}$ に対して, $\text{Pr}(A^c) = 1 - \text{Pr}(A)$ 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

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

となる.

- (4) $A, B \in \mathcal{A}, A \subset B \Rightarrow \text{Pr}(B \setminus A) = \text{Pr}(B) - \text{Pr}(A)$ となる. よって, $A \subset B \Rightarrow \text{Pr}(A) \leq \text{Pr}(B)$ 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, $A_m \cap A_n = \emptyset$ が成立していること.

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

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

となる.

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

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

となる.

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

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

となる.

Proof. (1) $F_1 := \Omega$, $F_n := \emptyset$ ($n \geq 2$) 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

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

を得る. よって

$$\Pr(\emptyset) = 0$$

がわかる.

(2) $F_1 := A$, $F_2 := A^c$, $F_n = \emptyset$ ($n \geq 3$) 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

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

がわかる. よって, (2) は示せた.

(3) $F_i := A_i$ ($i = 1, 2, \dots, N$) と $F_i = \emptyset$ ($i \geq N + 1$) 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

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

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, $F_1 := B \cap A^c, F_2 = A, F_n = \emptyset (n \geq 3)$ 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

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

がわかる. よって, (4) は示せた.

(5) $F_1 := A_1, F_2 = A_2 \setminus A_1, \dots, F_{n+1} := A_{n+1} \setminus A_n$ 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

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

を得る. よって, (5) 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 $\emptyset = (A_1 \setminus A_1) \subset (A_1 \setminus A_2) \subset (A_1 \setminus A_3) \subset \dots$ となる. このことに注意して分配法則を用いると

$$\bigcup_{n=1}^\infty (A_1 \setminus A_n) = \bigcup_{n=1}^\infty (A_1 \cap A_n^c) = A_1 \cap \left(\bigcap_{n=1}^\infty A_n\right)^c = A_1 \cap A^c = A_1 \setminus A \tag{1.1}$$

となる. 次に (4) と (6) を用いると

$$\begin{aligned} \Pr(A_1) - \Pr(A) &= \Pr(A_1 \setminus A) \quad (A_1 \supset A \text{ なので, (4) を用いた}) \\ &= \lim_{n \rightarrow \infty} \Pr(A_1 \setminus A_n) \quad (\because (1.1) \text{ から (5) を用いた}) \\ &= \lim_{n \rightarrow \infty} \left(\Pr(A_1) - \Pr(A_n)\right) \quad (\because (4)) \\ &= \Pr(A_1) - \lim_{n \rightarrow \infty} \Pr(A_n) \end{aligned}$$

がわかる.

(7) まず

$$\Pr(A_1 \cup A_2) \leq \Pr(A_1) + \Pr(A_2) \quad (1.2)$$

を示す. そのために

$$A_1 \cup A_2 = A_1 \cup \left(A_2 \setminus (A_1 \cap A_2) \right) \quad \text{かつ} \quad A_1 \cap \left(A_2 \setminus (A_1 \cap A_2) \right) = \emptyset$$

であることに注意する. (3) から

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

がわかる. (1.2) の操作を繰り返せば, $N \in \mathbb{N}$ に対して

$$\Pr\left(\bigcup_{n=1}^N A_n\right) \leq \sum_{n=1}^N \Pr(A_n) \quad (1.3)$$

がわかる. $\{\bigcup_{n=1}^N A_n\}_{N=1}^{\infty}$ は増加列なので, (5) と (1.3) から

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

がわかる. □

注意 1.8. 補題 1.7(6)(7) において, $\{\Pr(A_n)\}_{n=1}^{\infty}$ 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. $(\Omega, \mathcal{A}, \Pr)$ を確率空間とし, $A, B \in \mathcal{A}$ とする.

(1) (独立性):

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

(2) (条件付き確率) $\Pr(B) > 0$ のとき, B を与えたときの A の条件付き確率 $\Pr(A|B)$ を

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

で定める.

- (3) $m \geq 3$ とし, $A_j \in \mathcal{A} (j = 1, 2, \dots, m)$ とする. 事象 A_1, A_2, \dots, A_m は互いに独立であるとは, $2 \leq \forall k \leq m$ と $\forall 1 \leq j_1 < j_2 < \dots < j_k \leq m$ に対して

$$\Pr(A_{j_1} \cap A_{j_2} \cap \dots \cap A_{j_k}) = \Pr(A_{j_1})\Pr(A_{j_2}) \cdots \Pr(A_{j_k})$$

が成り立つときをいう.

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

$$\Pr(A|B) = \Pr(A)$$

が成立する.

- (2) $\Pr(B) > 0$ のとき

$$\Pr(A \cap B) = \Pr(B)\Pr(A|B)$$

となる. これを**乗法の公式**という.

- (3) $\Pr(B) > 0$ のとき, 関数 $\Pr(\cdot|B) : \mathcal{A} \rightarrow [0, 1]$ は定義 1.6(1) – (2) をみたす. すなわち (Ω, \mathcal{A}) 上の確率測度である.

- (4) $\Omega = \{b_1, b_2, b_3, b_4\}$ とし

$$\Pr(\{b_j\}) = \frac{1}{4} \quad (j = 1, 2, 3, 4)$$

とする. いま

$$A_j = \{b_j, b_4\} \quad (j = 1, 2, 3)$$

とする. すると

$$A_1 \cap A_2 = A_1 \cap A_3 = A_2 \cap A_3 = A_1 \cap A_2 \cap A_3 = \{b_4\}$$

である. したがって

$$\begin{aligned} \Pr(A_1 \cap A_2) &= \Pr(A_1)\Pr(A_2); \Pr(A_1 \cap A_3) = \Pr(A_1)\Pr(A_3); \\ \Pr(A_2 \cap A_3) &= \Pr(A_2)\Pr(A_3) \end{aligned}$$

である. しかし

$$\frac{1}{4} = \Pr(A_1 \cap A_2 \cap A_3) \neq \Pr(A_1)\Pr(A_2)\Pr(A_3) = \frac{1}{8}$$

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. $A_1, A_2, \dots, A_N \in \mathcal{A}$ は Ω の分割とする. すなわち, すべての $m \neq n (m, n \in \{1, 2, \dots, N\})$ に対し, $A_m \cap A_n = \emptyset$ 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

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

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を注意して, 定義 1.6(2) を用いればよい. \square

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

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

が成立する.

(2) $A, B \in \mathcal{A}$ とし, $\Pr(A) > 0, \Pr(B) > 0$ と仮定する. このとき

$$\Pr(A | B)\Pr(B) = \Pr(B | A)\Pr(A)$$

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. \square

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.

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

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

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

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. (Ω, \mathcal{A}) と $(\mathbb{X}, \mathcal{B})$ を可測空間とする。写像 $X : \Omega \rightarrow \mathbb{X}$ は (Ω, \mathcal{A}) から $(\mathbb{X}, \mathcal{B})$ への可測写像であるとは

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

をみたすときをいう。

注意 1.15. (1) $d \geq 2 (d \in \mathbb{N})$ とする。 $(\mathbb{X}, \mathcal{B}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ のとき、 X は確率空間 $(\Omega, \mathcal{A}, \text{Pr})$ 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 $(\Omega, \mathcal{A}, \text{Pr})$ 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。 \square

定理 1.16. (Ω, \mathcal{A}) と $(\mathbb{X}, \mathcal{B})$ を可測空間とし、 $X : \Omega \rightarrow \mathbb{X}$ 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、 $\{\omega \in \Omega; X(\omega) \in C\} \in \mathcal{A}$ 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、 $\{\omega \in \Omega : X(\omega) \in B\}$ 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

$$\begin{aligned} \left\{ X \in \bigcup_{n=1}^{\infty} B_n \right\} &= \bigcup_{n=1}^{\infty} \{X \in B_n\} \\ \{X \in B^c\} &= \{X \in B\}^c \end{aligned}$$

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。 \square

注意 1.17. (1). $(\mathbb{X}, \mathcal{B}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 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。 $(\mathbb{X}, \mathcal{B}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ に対して、定理 1.16 における \mathcal{C} の選択として

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

⁵ \mathcal{B} は \mathcal{C} を含む最小の σ 集合族。

がある. □

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

Proof. $C \in \mathcal{C}$ に対して

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

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

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

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

である. さらに, 集合族 $\{A_1 \times A_2 \times \dots \times A_n; A_i \in \mathcal{B}(\mathbb{R}) (i = 1, 2, \dots, n)\}$ は $\mathcal{B}(\mathbb{R}^n)$ を生成⁶する. したがって, 定理 1.16 から $(\mathbb{X}, \mathcal{B}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ に対して, (X_1, X_2, \dots, 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, \dots, X_n は確率変数のとき, $X_1 + X_2 + \dots + X_n$ も確率変数となる.

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

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

は \mathbb{R}^n の開集合となるので

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

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. □

⁶この事実の証明には, 測度論の議論が必要なので, 信じることにしよう.

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

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

にある.

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. \square

定理 1.22. X_1, X_2, \dots 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

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

から $\inf_n X_n$ は確率変数であることがわかる. 同様に

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

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. \square

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

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

は可測集合となる. $\Pr(\Omega_0) = 1$ 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ひ, 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

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

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 $(\Omega, \mathcal{A}, \Pr)$ 上の確率変数 X に対して

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

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

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

を X の**分布**という. したがって, $P^X((-\infty, x]) = F^X(x)$ である.

(2) 確率変数 X, Y を $(\Omega, \mathcal{A}, \Pr)$ の確率変数とし, それぞれの c.d.f. を F^X, F^Y とする. このとき

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

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 $X \stackrel{d}{=} Y$ と書く.

(3) 確率変数 X が c.d.f. F を持つとき, $X \sim F$ と書く.

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

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

(2) c.d.f. F が与えられると

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

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. このことにより X の c.d.f. と分布を同一視する. さらに $X \sim F$ とし, c.d.f. F から定まる確率測度を P としたとき, $X \sim P$ とも書く.

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

(3) 確率変数 X と Y の分布が同じとき

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

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 $(\Omega, \mathcal{A}, \Pr)$ 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; $x < y \Rightarrow F(x) \leq F(y)$.
- (2) $\lim_{x \rightarrow \infty} F(x) = 1$; $\lim_{x \rightarrow -\infty} F(x) = 0$.
- (3) F は右連続関数; $\lim_{y \rightarrow x+0} F(y) = F(x)$.
- (4) F の不連続点は高々可算個.

Proof. 確率変数 X の分布を P とする. すなわち $P(B) := \Pr(X \in B)$ ($\forall B \in \mathcal{B}(\mathbb{R})$) 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) $A_n = \left(-\infty, x + \frac{1}{n}\right]$ として, P に補題 1.7(7) を適用すればよい.
- (4) $n \in \mathbb{N}$ に対して

$$A_n := \left\{ r \in \mathbb{R}; \Pr(X = r) > \frac{1}{n} \right\}$$

とおく. $\#A_n \leq n$ となる. なぜならば, n 個を越えてあると仮定すると, その中の n 個を $x_1 < x_2 < \dots < x_n$ となるように選ぶことができる. すると

$$\begin{aligned} F^X(x_n) &= \Pr(X = x_n) + \lim_{x \rightarrow x_n-0} F^X(x) > \frac{1}{n} + F^X(x_{n-1}) > \dots \\ &> \frac{n-1}{n} + F^X(x_1) > 1 \end{aligned}$$

となり, $0 \leq F^X(x) \leq 1$ ($x \in \mathbb{R}$) に矛盾する.

$$A := \{r \in \mathbb{R}; \Pr(X = r) > 0\} = \bigcup_{n=1}^{\infty} A_n$$

なので, $\#A$ は高々加算となる. □

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 $(\Omega, \mathcal{A}, \Pr)$ 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) $F(x) = F(x-) (\forall x \in \mathbb{R})$. ただし, $F(x-) := \sup_y \{F(y); y < x\}$ とした.
- (3) $\Pr(X = x) = 0 (\forall x \in \mathbb{R})$.

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

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

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

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

であるので, 補題 1.7(6) より

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

を得る. さらに補題 1.7(4) より

$$\begin{aligned} \Pr(X = x) &= \Pr\left(\{X \leq x\} \setminus \{X < x\}\right) = \Pr(X \leq x) - \Pr(X < x) \\ &= F(x) - F(x-) \end{aligned} \tag{1.4}$$

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

(3) \Rightarrow (1) の証明: (1.4) から

$$\Pr(X = x) \Rightarrow F(x) = F(x-)$$

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. \square

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

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

を示せ.

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

$$\Pr(a \leq X \leq b) = \Pr(a < X < b) = \Pr(a < X \leq b) = \Pr(a \leq X < b)$$

である. □

定義 1.30. 確率変数 X を確率空間 $(\Omega, \mathcal{A}, \Pr)$ 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 $\{x_1, x_2, \dots\}$ 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

$$p(x) := \Pr(X = x) = F(x) - F(x-) \quad (x \in \{x_1, x_2, \dots\})$$

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) $\Pr(X = x) = 0 (\forall x \in \mathbb{R})$ 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

$$F(x) = \int_{-\infty}^x p(t) dt \quad (\forall x \in \mathbb{R})$$

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

$$\dot{F}(x) = \frac{dF}{dx}(x) = p(x)$$

となる.

注意 1.31. 確率変数 X の分布を P とする.

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

(1) X を離散型とし, $S = \{x \in \mathbb{R}; p(x) > 0\}$ とおく¹¹. すると任意の Borel 集合 $B \subset \mathbb{R}$ に対して

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

が成り立つことを示すことができる.

(2) X を連続型とし, その p.d.f. p が定義されるとする. このとき, 任意の Borel 集合 $B \subset \mathbb{R}$ に対して

$$P(B) = \Pr(X \in B) = \int_B p(x) dx$$

が成り立つことを示すことができる. また,

$$\Pr(B) = \int_B p_1(x) dx = \int_B p_2(x) dx \quad (\forall B \in \mathcal{B}(\mathbb{R}))$$

ならば, 高々可算個の点を除いて, $p_1 = p_2$ となることが知られている. したがって, 1.30(2) の p.d.f. は存在するならば, 高々可算個の点を除いて, 同じ値を取るものになる. これらの主張の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) $\Omega = \{0, 1\}$, $\mathcal{A} = 2^\Omega$, $\Pr(\{0\}) = \Pr(\{1\}) = 1/2$ とし

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

と定義すれば

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

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

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

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

$$\Pr((a, b]) = b - a \quad (0 \leq a < b \leq 1)$$

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

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

となる.

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.

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) $F^- : (0, 1) \rightarrow \mathbb{R}$ を

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

で定義する. また, $y \in (0, 1)$ に対して, $F^{-1}(y)$ 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 $\text{supp } F = \{x \in \mathbb{R}; 0 < F(x) < 1\}$ 上で狭義単調増加のとき, F^- は F の逆関数になる. $\text{supp } F$ は累積分布関数が $F(x) > 0$ かつ $F(x) < 1$ となる点 x の全体である. 関数の台はこの集合の閉包で定義するので, 違いに注意せよ. \square

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

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

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

$$F(F^-(y)) = F\left(\lim_{n \rightarrow \infty} x_n\right) = \lim_{n \rightarrow \infty} F(x_n) \geq y \quad (1.5)$$

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. \square

注意 1.36. c.d.f. F は連続とする. c.d.f. の性質

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

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

$$F(x) = y$$

をみたす. ここで, F^- の定義から

$$F^-(y) \leq x \quad (1.6)$$

となる. なぜならば, $F^-(y)$ は条件 $F(z) \geq y$ をみたす $z \in \mathbb{R}$ の \inf である. 一方, $F(x) = y$ だから $x \in \{z \in \mathbb{R} : F(z) \geq y\}$ なので, $x \geq F^-(y)$ がわかる. (1.6) と F の非減少性から

$$F(F^-(y)) \leq F(x) = y \quad (1.7)$$

となる. 一方, F^- の定義から

$$F(F^-(y)) \geq y \quad (1.8)$$

となる. (1.7) と (1.8) を合わせると

$$F(F^-(y)) = y$$

を得る. したがって

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

となる.

注意 1.37. F が不連続ならば

$$F(F^-(y)) = y$$

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, $X \sim U(a, b)$ とする. すると X の c.d.f. は

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

である. よって, F の分位点関数は

$$F^-(y) = a + (b-a)y \quad (y \in (0, 1))$$

となる. □

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

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

となる. よって, F の分位点関数は

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

となる. この場合, $y \neq \frac{1}{2}$ のとき

$$F(F^-(y)) \neq y$$

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 $\lim_{n \rightarrow \infty} y_n = y$ としたとき

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

が成立する.

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

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

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

となる. 両辺の \inf をとると

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

となる. よって, F^- の非減少性が証明できた.

(2) $x_n := F^-(y_n)$ ($n \in \mathbb{N}$) 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

$$x_n = F^-(y_n) \leq F^-(y) =: x_0 \tag{1.9}$$

となる. よって, (1.9) の両辺の極限を取ると

$$x := \lim_{n \rightarrow \infty} x_n \leq x_0 \tag{1.10}$$

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 (1.8) に注意すると

$$y_n \leq F(F^-(y_n)) \leq F(x_n) \leq F(x + \epsilon) = F(x_0 - \epsilon)$$

となる. すると

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

となる. しかし, F^- の定義から

$$F(x_0 - \epsilon) < y$$

となる. なぜならば, $F^-(y) = x_0$ なので, x_0 は条件 $F(z) \geq y$ をみたす z の \inf である. $x_0 > x_0 - \epsilon$ だから, $x_0 - \epsilon \notin \{z \in \mathbb{R} : F(z) \geq y\}$ となる. よって, $F(x_0 - \epsilon) < y$ となることがわかる. 以上から $y \leq F(x_0 - \epsilon)$ かつ $y > F(x_0 - \epsilon)$ となるので, 矛盾. よって

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

となるので, F^- の左連続性がわかる.

(3) F^- の非減少性から

$$y \leq F(x) \Rightarrow F^-(y) \leq F^-(F(x)) \leq x$$

となる. 最後の不等号は

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

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

逆に, $F^-(y) \leq x$ ならば, (1.8) と F の非減少性から

$$y \leq F(F^-(y)) = F(x)$$

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, $U \sim U(0, 1)$ に対して

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

となる.

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

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

となる. よって

$$\Pr(X \leq x) = \Pr(U \leq F(x)) = F(x)$$

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. $p_\theta(\cdot)$ が

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

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

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

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

のときをいう. ただし

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}, \quad 0! = 1$$

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

問 1.7. $p_{\theta,n}(x)$ を 2 項分布 $\text{Bino}(n, \theta)$ ($0 < \theta < 1$, $n \in \mathbb{N}$) p.m.f. としたとき, $\sum_{x=0}^n p_{\theta,n}(x) = 1$ 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. $p(\cdot|\theta)$ が

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

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

問 1.8. $p_\theta(x)$ を幾何分布 $\text{Geo}(\theta)$ ($\theta > 0$) の p.m.f. としたとき, $\sum_{x=0}^{\infty} p_\theta(x) = 1$ 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. $p_\theta(\cdot)$ が

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

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

問 1.9. $p_\theta(x)$ を Poisson 分布 $\text{Po}(\theta)$ ($\theta > 0$) の p.m.f. としたとき, $\sum_{x=0}^{\infty} p_\theta(x) = 1$ 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. $p_{\mu, \sigma^2}(\cdot)$ が

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

のときをいう. ただし $\exp(x) = e^x$ である. このことを $X \sim \text{N}(\mu, \sigma^2)$ 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. $p_{\mu, \sigma^2}(x)$ 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. $p_{\mu, \sigma^2}(x)$ 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. $p_{\mu, \sigma^2}(x)$ 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

$$\int_0^{\infty} e^{-t^2} dt = \sqrt{\pi}$$

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

$$\int_{-\infty}^{\infty} e^{-z^2/2} dz = \sqrt{2\pi}$$

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.

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

$$\int_{-\infty}^{\infty} p_{\mu, \sigma^2}(x) dx = 1$$

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. $p_{\alpha, \beta}(\cdot)$ が

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

のときをいう. ただし

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx \quad (\alpha > 0)$$

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

問 1.10. $p(x|\alpha, \beta)$ をガンマ分布 $\text{Ga}(\alpha, \beta)$ ($\alpha > 0, \beta > 0$) の p.d.f. としたとき, $\int_0^{\infty} p_{\alpha, \beta}(x) dx = 1$ 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 $(\Omega, \mathcal{A}, \text{Pr})$ 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

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

である.

定義 1.45. 確率空間 $(\Omega, \mathcal{A}, \Pr)$ 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 \mathbf{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) $\mathbf{p}(x, y) \geq 0 (\forall (x, y) \in \mathbb{R}^2)$.
- (2) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{p}(x, y) dx dy = 1$.
- (3) $\forall A \in \mathcal{B}(\mathbb{R}^2)$ に対して

$$\Pr((X, Y) \in A) = \int \int_A \mathbf{p}(x, y) dx dy.$$

注意 1.46. (X, Y) の同時累積分布関数 (同時 c.d.f.) F を

$$\begin{aligned} F(x, y) &:= \Pr(X \leq x, Y \leq y) && (1.11) \\ &:= \Pr\{\omega \in \Omega; X(\omega) \leq x, Y(\omega) \leq y\} && (\forall (x, y) \in \mathbb{R}^2) \end{aligned}$$

で定義する. (X, Y) の同時分布 P を

$$P(A) = \Pr((X, Y) \in A) \quad (\forall A \in \mathcal{B}(\mathbb{R}^2)) \quad (1.12)$$

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ことを明示したいときには, $F^{(X, Y)}$ または $P^{(X, Y)}$ と記す.

細かなことであるが, (1.11) 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 ?? 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.

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

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

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

となる. □

1.5.2 周辺分布

定義 1.48. (X, Y) を確率空間 $(\Omega, \mathcal{A}, \Pr)$ 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. ρ を持つとする. X の周辺確率関数 (周辺 p.m.f.) を

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

で定義する. ただし

$$S_X := \{x \in \mathbb{R}; \text{ある } y \in \mathbb{R} \text{ に対して, } \rho(x, y) > 0\}$$

$$S_Y := \{y \in \mathbb{R}; \text{ある } x \in \mathbb{R} \text{ に対して, } \rho(x, y) > 0\}$$

である.

(2) (X, Y) は連続型とし, 同時 p.d.f. ρ 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

$$p^X(x) = \int_{-\infty}^{\infty} \rho(x, y) dy \quad (\forall x \in \mathbb{R})$$

で定義し, Y の周辺確率密度関数 (周辺 p.d.f.) を

$$p^Y(y) = \int_{-\infty}^{\infty} \rho(x, y) dx \quad (\forall y \in \mathbb{R})$$

で定義する.

注意 1.49. 連続型確率ベクトル (X, Y) に対して

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

となるので, X の周辺 p.d.f. と X の p.d.f. は高々可算個の点を除いて同じ値を取ることがわかる. \square

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

$$F^X(x) = \lim_{y \rightarrow \infty} F(x, y); \quad F^Y(y) = \lim_{x \rightarrow \infty} F(x, y)$$

が成立する. \square

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 $(\Omega, \mathcal{A}, \Pr)$ 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

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

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

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

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. \square

注意 1.53. (1) 連続型の場合には, 測度 0 の集合を除いて¹³p.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. \square

例 1.54. 連続型確率変数 X と Y は独立で同時 p.d.f.

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

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

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

となる. したがって

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

となる. \square

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.

定理 1.55. 連続型確率ベクトル (X, Y) は同時 p.d.f. $p(x, y)$ を持つとする. ただし $\{(x, y) \in \mathbb{R}^2; p(x, y) > 0\}$ 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

$$p(x, y) = q(x)r(y)$$

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. □

例 1.57. 連続型確率ベクトル (X, Y) は同時 p.d.f.

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

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

$$p(x, y) = q(x)r(y)$$

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. $p(x, y)$ を持つとする. $p_Y(y) > 0$ 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

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

で定義する.

(2) 連続型確率ベクトル (X, Y) は同時 p.d.f. $p(x, y)$ を持つとする. $p^Y(y) > 0$ 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

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

で定義する.

¹⁴有界でなくともよい.

注意 1.59. (X, Y) が連続型確率ベクトルのとき, $p^Y(y) > 0$ 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

$$\Pr(X \in A | Y = y) := \int_A p^{X|Y}(x|y) dx \quad (\forall A \in \mathcal{B}(\mathbb{R}))$$

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.

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

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

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

となる. したがって

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

となる. よって $0 < y < 1$ と $0 < x < 1$ に対して

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

となる. これより $0 < y < 1$ のとき

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

となる. 注意 1.59 より

$$\Pr\left(X \leq \frac{1}{4} \mid Y = \frac{1}{3}\right) = \int_0^{1/4} p^{X|Y}\left(x \mid \frac{1}{3}\right) dx = \int_0^{1/4} \frac{x + \frac{1}{3}}{\frac{1}{3} + \frac{1}{2}} dx = \frac{11}{80}$$

となる. □

1.6 多次元分布と i.i.d. 標本

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

$$\mathbf{X} = (X_1, X_2, \dots, X_n)^\top$$

と書く. \mathbf{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, \dots, X_n が独立であるとは、すべての Borel 集合 $A_1, A_2, \dots, A_n \in \mathcal{B}(\mathbb{R})$ に対して

$$\Pr(X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n) = \prod_{j=1}^n \Pr(X_j \in A_j) \quad (1.13)$$

が成立するときである。

注意 1.63. \mathbf{X} の同時 p.d.f. を $p(x_1, x_2, \dots, x_n)$ と書き、各 $X_j (j = 1, 2, \dots, n)$ の周辺 p.d.f. を p^{X_j} と書くことにする。(1.13) を示すためには

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

を示せばよい。 □

定義 1.64. X_1, X_2, \dots, X_n は独立で、各 $X_j (j = 1, 2, \dots, n)$ は同じ c.d.f. F を持つとき、 X_1, X_2, \dots, X_n は**独立同一分布に従う** (i.i.d. = identically and independently distributed) といひ

$$X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} F$$

と書く¹⁵。 X_1, X_2, \dots, 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、 $\mathbf{p} = (p_1, p_2, \dots, p_d)$, $p_j (0 \leq p_j \leq 1; j = 1, 2, \dots, d)$, $\sum_{j=1}^d p_j = 1$ とす

¹⁵p.d.f. p を使い

$$X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} p$$

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

る. $\mathbf{X} = (X_1, X_2, \dots, X_d)$ は母数 (d, \mathbf{p}) の多項分布に従うとは, \mathbf{X} の p.m.f. が

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

で与えられたときをいう. ただし, $\sum_{j=1}^d x_j = n$ で

$$\binom{n}{x_1, x_2, \dots, x_d} = \frac{n!}{x_1! x_2! \dots x_d!}, \quad 0! = 1$$

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

補題 1.65. $\mathbf{X} \sim \text{Multi}_d(n, \mathbf{p})$ とする. ただし, $\mathbf{p} = (p_1, p_2, \dots, p_d)$ と $\mathbf{X} = (X_1, X_2, \dots, X_d)$ とする. このとき X_j ($j = 1, 2, \dots, d$) の周辺分布は $\text{Bino}(n, p_j)$ 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

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

$$\mathbf{Z} := \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_d \end{pmatrix} \tag{1.14}$$

と定める¹⁶. すると \mathbf{Z} の同時 p.d.f. は

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

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せて表現している. 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.

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

$$\int \cdots \int_{\mathbb{R}^d} \mathbf{p}^{\mathbf{Z}}(z_1, z_2, \dots, z_d) dz_1 dz_2 \cdots dz_d = 1$$

となっていることもわかる.

次に

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

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

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

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

$$\mathbf{X} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2} \mathbf{Z} \tag{1.15}$$

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

$$\mathbf{p}_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}^{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} \det(\boldsymbol{\Sigma})^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\} \quad (\mathbf{x} \in \mathbb{R}^d) \tag{1.16}$$

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 $\boldsymbol{\Sigma}$ の d 変量正規分布という.

注意 1.66. (1.16) の導出は以下のように行うことができる. 一般に

$$\mathbf{Z} = (Z_1, Z_2, \dots, Z_d)^\top$$

を d 次元確率ベクトルとし, その同時 p.d.f. を $\mathbf{p}^{\mathbf{Z}}$ と書くことにする. さらに $\mathbb{Z} := \{\mathbf{z} \in \mathbb{R}^d; \mathbf{p}^{\mathbf{Z}}(\mathbf{z}) > 0\}$ とし, 関数

$$\mathbf{h}(\cdot) = (h_1(\cdot), h_2(\cdot), \dots, h_d(\cdot))^\top : \mathbb{Z} \rightarrow \mathbf{h}(\mathbb{Z})$$

は 1 対 1 とし, $j = 1, 2, \dots, d$ に対して

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

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

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

$$\mathbf{h}^{-1} = (h_1^{-1}, h_2^{-1}, \dots, h_d^{-1})^\top : \quad \mathbf{h}(\mathbb{Z}) \rightarrow \mathbb{Z}$$

が存在すると仮定する. よって, $\mathbf{h}^{-1}(\mathbf{h}(z)) = z$ なので, $\mathbf{Z} = \mathbf{h}^{-1}(\mathbf{X})$ 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 \mathbf{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. 変換 \mathbf{h} の Jacobian を

$$J_{\mathbf{h}}(z) = \det \begin{pmatrix} \frac{\partial h_1(z)}{\partial z_1} & \dots & \frac{\partial h_1(z)}{\partial z_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_d(z)}{\partial z_1} & \dots & \frac{\partial h_d(z)}{\partial z_d} \end{pmatrix}$$

とおく. さらに, \mathbf{X} の同時 p.d.f. を求めるために, $\mathbf{h}^{-1}(\mathbf{x})$ の Jacobian $\mathbf{J}_{\mathbf{h}^{-1}}(\mathbf{x})$ を

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

で定める.

$\mathbf{J}_{\mathbf{h}}(z) \neq 0 (z \in \mathbb{Z})$ ¹⁸が成立するとき

$$p^{\mathbf{X}}(\mathbf{x}) = \frac{p^{\mathbf{Z}}(\mathbf{h}^{-1}(\mathbf{x}))}{J_{\mathbf{h}}(\mathbf{h}^{-1}(\mathbf{x}))} = |\mathbf{J}_{\mathbf{h}^{-1}}(\mathbf{x})| p^{\mathbf{Z}}(\mathbf{h}^{-1}(\mathbf{x})) \quad (1.17)$$

となることが知られている. これは, 標準的な多変数の積分の変換公式からわかる.

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

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

を (1.17) に代入すれば, (1.16) はわかる. 以上の議論から

$$\int \int \dots \int_{\mathbb{R}^d} p^{\mathbf{X}}(\mathbf{x}) dx_1 dx_2 \dots dx_d = \int \int \dots \int_{\mathbb{R}^d} p^{\mathbf{Z}}(z) dz_1 dz_2 \dots dz_d = 1$$

となっていることもわかる. □

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

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

を示せ.

¹⁸実は, a.e \mathbb{X} でよい.

注意 1.67. Σ を半正定値としたときに, 多変量正規分布を (1.14) と変換 (1.15) を用いて定義することが出来る. しかし, $\det(\Sigma) = 0$ のときは, $\forall B \in \mathcal{B}(\mathbb{R}^d)$ に対して, 確率 $\Pr(\mathbf{X} \in B)$ は定義できるが, \mathbf{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. \square

定理 1.68. $\mathbf{X} = (X_1, X_2, \dots, X_d)^\top \sim N_d(\boldsymbol{\mu}, \Sigma)$ とする. ただし $\boldsymbol{\mu} \in \mathbb{R}^d$ で, Σ は $d \times d$ の正定値行列である. このとき次が成立する.

- (1) $X_j (j = 1, 2, \dots, d) \sim N(\mu_j, \sigma_{jj})$. ただし μ_j は $\boldsymbol{\mu}$ の第 j 成分, σ_{jj} は Σ の第 j 対角成分である.
- (2) $X_\ell = x_\ell (\ell = 1, 2, \dots, d)$ を与えたときの $X_j (j \neq \ell)$ の条件付き分布は

$$X_j | X_\ell = x_\ell \sim N\left(\mu_j + \frac{\sigma_{j\ell}}{\sigma_{\ell\ell}}(x_\ell - \mu_\ell), \sigma_{jj} - \frac{\sigma_{j\ell}^2}{\sigma_{\ell\ell}}\right)$$

となる.

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

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

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

- (4) $V := (\mathbf{X} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{X} - \boldsymbol{\mu}) \sim \chi_d^2$ である. なお, 自由度 d の χ^2 分布 χ_d^2 は定義 1.74 で導入する.

Proof. (1) $j = 1$ として示す. Σ の $(i, j) (i, j = 1, 2, \dots, d)$ 成分を σ_{ij} と書き, $(d-1) \times (d-1)$ 行列 Σ_1 と $(d-1) \times 1$ ベクトル $\boldsymbol{\sigma}_1$ をそれぞれ

$$\Sigma_1 = (\sigma_{ij})_{i,j=2,3,\dots,d}, \quad \boldsymbol{\sigma}_1 = (\sigma_{21}, \sigma_{31}, \dots, \sigma_{d1})^\top$$

で定める. すると

$$\Sigma = \left(\begin{array}{c|c} \sigma_{11} & \boldsymbol{\sigma}_1^\top \\ \hline \boldsymbol{\sigma}_1 & \Sigma_1 \end{array} \right)$$

となる. このとき

$$\Sigma = \left(\begin{array}{c|c} 1 & \mathbf{0}_{d-1}^\top \\ \hline \sigma_{11}^{-1} \boldsymbol{\sigma}_1 & \mathbf{I}_{d-1} \end{array} \right) \left(\begin{array}{c|c} \sigma_{11} & \mathbf{0}_{d-1}^\top \\ \hline \mathbf{0}_{d-1} & \Sigma_1 - \sigma_{11}^{-1} \boldsymbol{\sigma}_1 \boldsymbol{\sigma}_1^\top \end{array} \right) \left(\begin{array}{c|c} 1 & \sigma_{11}^{-1} \boldsymbol{\sigma}_1^\top \\ \hline \mathbf{0}_{d-1} & \mathbf{I}_{d-1} \end{array} \right)$$

と書ける. ただし \mathbf{I}_{d-1} は $(d-1)$ 次の単位行列で, $\mathbf{0}_{d-1}$ は $(d-1) \times 1$ の零ベクトルである. Σ が正定値ならば, $\Sigma_2 - \sigma_{11}^{-1} \boldsymbol{\sigma}_1 \boldsymbol{\sigma}_1^\top$ も正定値であるので

$$\begin{aligned} \Sigma^{-1} &= \left(\begin{array}{c|c} 1 & -\sigma_{11}^{-1} \boldsymbol{\sigma}_1^\top \\ \hline \mathbf{0}_{d-1} & \mathbf{I}_{d-1} \end{array} \right) \left(\begin{array}{c|c} \sigma_{11}^{-1} & \mathbf{0}_{d-1}^\top \\ \hline \mathbf{0}_{d-1} & \{\Sigma_2 - \sigma_{11}^{-1} \boldsymbol{\sigma}_1 \boldsymbol{\sigma}_1^\top\}^{-1} \end{array} \right) \\ &\quad \times \left(\begin{array}{c|c} 1 & \mathbf{0}_{d-1}^\top \\ \hline -\sigma_{11}^{-1} \boldsymbol{\sigma}_1 & \mathbf{I}_{d-1} \end{array} \right) \end{aligned} \quad (1.18)$$

となる. $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_d)^\top$ と $\mathbf{x} = (x_1, x_2, \dots, x_d)^\top$ と書いたときに, $\boldsymbol{\mu}_1 = (\mu_2, \mu_3, \dots, \mu_d)^\top$ と $\mathbf{x}_1 = (x_2, x_3, \dots, x_d)^\top$ とおく. (1.18) から

$$\begin{aligned} \text{tr}(\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) &= \frac{1}{\sigma_{11}} (x_1 - \mu_1)^2 \\ &\quad + \text{tr} \left[\left(\mathbf{x}_1 - \boldsymbol{\mu}_1 - \frac{x_1 - \mu_1}{\sigma_{11}} \boldsymbol{\sigma}_1 \right)^\top \left\{ \Sigma_2 - \sigma_{11}^{-1} \boldsymbol{\sigma}_1 \boldsymbol{\sigma}_1^\top \right\}^{-1} \right. \\ &\quad \left. \times \left(\mathbf{x}_1 - \boldsymbol{\mu}_1 - \frac{x_1 - \mu_1}{\sigma_{11}} \boldsymbol{\sigma}_1 \right) \right] \end{aligned} \quad (1.19)$$

$$\begin{aligned} &=: \frac{1}{\sigma_{11}} (x_1 - \mu_1)^2 + g_{d-1}(\mathbf{x}_1) \end{aligned} \quad (1.20)$$

となる. ただし

$$\begin{aligned} g_{d-1}(\mathbf{x}_1) &= \text{tr} \left[\left(\mathbf{x}_1 - \boldsymbol{\mu}_1 - \frac{x_1 - \mu_1}{\sigma_{11}} \boldsymbol{\sigma}_1 \right)^\top \left\{ \Sigma_2 - \sigma_{11}^{-1} \boldsymbol{\sigma}_1 \boldsymbol{\sigma}_1^\top \right\}^{-1} \right. \\ &\quad \left. \times \left(\mathbf{x}_1 - \boldsymbol{\mu}_1 - \frac{x_1 - \mu_1}{\sigma_{11}} \boldsymbol{\sigma}_1 \right) \right] \end{aligned}$$

である. このことから

$$\begin{aligned} &\int_{\mathbb{R}^{d-1}} \exp \left[-\frac{1}{2} \text{tr}(\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right] dx_2 dx_3 \times dx_d \\ &= \exp \left[-\frac{(x_1 - \mu_1)^2}{2\sigma_{11}} \right] \int_{\mathbb{R}^{d-1}} \exp \left[-\frac{1}{2} g_{d-1} \right] dx_2 dx_3 \times dx_d \\ &= \exp \left[-\frac{(x_1 - \mu_1)^2}{2\sigma_{11}} \right] \det \left(2\pi (\Sigma_2 - \sigma_{11}^{-1} \boldsymbol{\sigma}_1 \boldsymbol{\sigma}_1^\top) \right)^{1/2} \end{aligned}$$

がわかる. よって

$$\int_{\mathbb{R}^{d-1}} p^{\mathbf{X}}(\mathbf{x} | \boldsymbol{\mu}, \Sigma) dx_2 dx_3, \dots, dx_d = \frac{1}{\sqrt{2\pi\sigma_{11}}} \exp \left[-\frac{(x_1 - \mu_1)^2}{2\sigma_{11}} \right]$$

がわかる.

- (2) (1.20) からわかる.
 (3) $\mathbf{c}^\top \mathbf{X}$ の第 2 章で定義する積率を計算すればよい.
 (4) (3) から

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

となる. あとは χ^2 分布の定義からわかる. □

(1.17) 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.

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

を持つとする. さらに

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

とし, $\mathbf{Y} = (h_1(X_1, X_2), h_2(X_1, X_2))^\top$, $\mathbf{y} = (y_1, y_2)^\top$ 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 \mathbf{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

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

より

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

を得る. また

$$\int \int_{\mathbb{R}^2} p^{\mathbf{Y}}(\mathbf{y}) dy_1 dy_2 = 1$$

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

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

で定義する。 $\alpha > 1$ のとき、部分積分により

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

となる。 $\Gamma(1) = \int_0^{\infty} e^{-x} dx = 1$ なので、帰納法により

$$\Gamma(n) = (n-1)! \quad (n = 1, 2, \dots)$$

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 (α, β) のガンマ分布に従うとは、 X の p.d.f. が

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

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

補題 1.71. $n \geq 2$ は自然数とする。 X_1, X_2, \dots, X_n は独立な確率変数、各 $X_j \sim \text{Ga}(\alpha_j, \beta)$ ($j = 1, 2, \dots, n, \alpha_j > 0, \beta > 0$ としたとき

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

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

$$p^{X+Y}(x) = \int_{-\infty}^{\infty} p^X(x-y)p^Y(y) dy$$

で与えられる.

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

$$\begin{aligned} \Pr(X + Y \leq t) &= \int \int_{x+y \leq t} p^X(x)p^Y(y) dx dy \\ &= \int_{-\infty}^{\infty} p^Y(y) \left\{ \int_{-\infty}^{t-y} p^X(x) dx \right\} dy \\ &= \int_{-\infty}^{\infty} p^Y(y) \left\{ \int_{-\infty}^t p^X(z-y) dz \right\} dy \\ &\quad (x|_{-\infty}^{t-y} = z-y|_{-\infty}^t \text{ と変換}) \\ &= \int_{-\infty}^t \left\{ \int_{-\infty}^{\infty} p^X(z-y)p^Y(y) dy \right\} dz \end{aligned}$$

となる. よって微積分の基本定理より

$$p^{X+Y}(t) = \frac{d}{dt} \Pr(X + Y \leq t) = \int_{-\infty}^{\infty} p^X(t-y)p^Y(y) dy$$

を得る. □

補題 1.71 の証明. $X_1 \sim \text{Ga}(\alpha_1, \beta), X_2 \sim \text{Ga}(\alpha_2, \beta)$ とし, X_1 と X_2 は独立とする. $\text{Ga}(\alpha, \beta)$ の p.d.f. を $p(\cdot | \alpha, \beta)$ と書く. 補題 1.72 より

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

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

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

より

$$\int_0^1 (1-z)^{\alpha_1-1} z^{\alpha_2-1} dz = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)} \quad (1.21)$$

を得る. よって

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

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

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

をベータ関数とよぶ. すると (1.21) 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. $Z_1, Z_2, \dots, Z_n \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$ とする.

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

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

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

$$p^Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}} y^{(1/2)-1} e^{-y/2} & (y > 0) \\ 0 & (\text{その他の場合}) \end{cases} \quad (1.22)$$

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

$$p_{1/2, 1/2}(y) = \begin{cases} \frac{1}{\sqrt{2}\Gamma(\frac{1}{2})} y^{(1/2)-1} e^{-y/2} & (y > 0) \\ 0 & (\text{その他の場合}) \end{cases}$$

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

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

がわかる. よって χ_n^2 の p.d.f. は

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

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, $F \sim F(k, m)$ と記す.

注意 1.76. 上の定義では, F 分布の記号を $F(k, m)$ 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. $F \sim F(k, m)$ のとき, F の p.d.f. $p(x|k, m)$ は

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

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 を正值の連続型確率変数とし, それぞれの p.d.f. を $p^X(x)$ と $p^Y(x)$ とする. このとき

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

の p.d.f. は

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

で与えられる.

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

$$\begin{aligned} \Pr(Z \leq t) &= \int_0^\infty \left\{ \int_0^{ty} p^X(x)p^Y(y) dx \right\} dy \\ &= \int_0^\infty \left\{ \int_0^t p^X(zy)p^Y(y)y dz \right\} dy \quad (x = zy \text{ と変換}) \\ &= \int_0^t \left\{ \int_0^\infty p^X(zy)p^Y(y)y dy \right\} dz \end{aligned}$$

となる. ここで微積分の基本定理を用いると

$$p^Z(t) = \frac{d}{dt} \Pr(Z \leq t) = \int_0^\infty yp^X(ty)p^Y(y) dy$$

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

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

となる. $F = \frac{m}{k}Z$ なので

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

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. $Z_0, Z_1, Z_2, \dots, Z_n \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$ のとき

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

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

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

で与えられる

Proof. 定義より

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

であることをまず思い出す. T の分布は対称なので

$$p_n^T(x) = p_n^T(-x) \quad (x \geq 0)$$

となる. よって $x > 0$ に対して

$$2\Pr(0 \leq T \leq x) = \Pr(-x \leq T \leq x) = \Pr(T^2 \leq x^2)$$

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

$$2 \int_0^x p^T(t|n) dt = \int_0^{x^2} p_{1,n}(t) dt$$

となる. さらに, $t \mapsto t^2$ と変換すると

$$\int_0^{x^2} p_{1,n}(t) dt = \int_0^x p_{1,n}(t^2) 2t dt \quad (1.23)$$

となる. ここで補題 1.77 から

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

であることを思い出す. この式を (1.23) の右辺に代入すると

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

となる. □

1.8 章末注釈と参考文献

第 1.1 節は [28] 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 [36] を参照した. 第 1.7 節は [23] からの借用である.

1.9 演習問題

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

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

である. ただし

$$\binom{3}{x} = \frac{3!}{x! \times (3-x)!}, \quad 0! = 1$$

である.

- (1) $p(0), p(1), p(2), p(3)$ の値を計算せよ.
- (2) $\Pr(0 < X < 3)$ を求めよ.

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

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

としたとき, 以下の問いに答えよ.

- (1) 任意の $y > 0$ に対して, $\Pr(Z^2 \leq y)$ 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

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

となることを利用して, Y の p.d.f. が (1.22) で与えられることを示せ.

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

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

とする. さらに, 離散型確率変数 Y を

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

で定める. このとき, 以下の問いに答えよ.

- (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 $\Pr\left(0 < X \leq \frac{1}{2}\right)$, $\Pr(Y = 0)$ を求めよ.
- (4) 確率 $\Pr\left(0 < X \leq \frac{1}{2}, Y = 0\right)$ 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 $(\Omega, \mathcal{A}, \Pr)$ 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. $p^{(Z, W)}$ を求めよ.
- (2) 確率 $\Pr(Z = W)$ と条件付き確率 $\Pr(Z = 2 | W < 5)$ 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. 標準正規分布に独立に従う 2 つの確率変数 X と Y に対して

$$Z = \min\{X, Y\}, \quad W = \begin{cases} 1 & (Z = X) \\ 0 & (Z = Y) \end{cases}$$

と定める. また, 標準正規分布の p.d.f と c.d.f. を p と F と書くことにする.

- (1) $z \in \mathbb{R}$, $w = 0, 1$ に対して

$$\Pr(Z \leq z, W = w)$$

を求めよ.

- (2) Z の周辺 c.d.f. を求めよ.
- (3) 事象 $\{Z \leq z\}$ が与えられたときの $W = 1$ となる条件付き確率

$$\Pr(W = 1 | Z \leq z)$$

を求めよ.

- (4) Z と W は独立かどうかを示せ.

演習問題 1.6. p を自然数とする. 連続型確率変数 T は p.d.f.

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

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

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx \quad (a > 0)$$

で定義する.

(1) $\int_{-\infty}^\infty p^T(t) dt$ 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_p^2 に従い, V は $N(0, 1)$ に従うとする.

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

とおいたとき, T の p.d.f. p^T を求めよ.

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

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

となる. さらに

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

として, (1.17) を利用して, 同時 p.d.f. $p^{(T,W)}$ を求め

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

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

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

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

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

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。 $E[g^+(X)]$ または $E[g^-(X)]$ のいずれかが有限ならば

$$E[g(X)] := E[g^+(X)] - E[g^-(X)]$$

¹可測関数の定義は定義 ?? を参照。

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

補題 2.2. X を確率変数とする. 関数 $g, h: \mathbb{R} \rightarrow \mathbb{R}$ を可測とする.

- (1) $h(X)$ は有限の期待値を持つ²とする. $0 \leq g(x) \leq h(x) (\forall x \in \mathbb{R})$ ならば

$$E[g(X)] \leq E[h(X)]$$

となる.

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

$$E[ag(X) + bh(X)] = aE[g(X)] + bE[h(X)]$$

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

$$E[h(X) - g(X)] = \int_{-\infty}^{\infty} \underbrace{\{h(x) - g(x)\}}_{\geq 0} p^X(x) dx \geq 0$$

となる. さらに, (1) を用いると $E[h(X)] \geq E[g(X)]$ 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. $|ag(x) + bh(x)| \leq |a||g(x)| + |b||h(x)|$ 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 のステップは省略³できる. あとは, 積分の線型性から

$$\begin{aligned} E[ag(X) + bh(X)] &= \int_{-\infty}^{\infty} \{ag(x) + bh(x)\} p^X(x) dx \\ &= a \int_{-\infty}^{\infty} g(x) p^X(x) dx + b \int_{-\infty}^{\infty} h(x) p^X(x) dx \\ &= aE[g(X)] + bE[h(X)] \end{aligned}$$

がわかる. □

² g, h を共に非負値関数としているので, $E[|h(X)|] < \infty$ だけを仮定すればよい. もし, g, h を可測関数とすると $E[|g(X)|] < \infty$ の仮定も必要になる. なぜならば, $g(x) \leq h(x) (x \in \mathbb{R})$ と $E[|h(X)|] < \infty$ だけだと $E[h(X)]$ が定義できない場合があるからである. なので, 非負値関数に関する主張とした方が実用的である.

³関数を非負と非正の部分に分けて, それぞれが可積であることは自明になるので, これらのステップを飛ばして, 積分を考えてよいことになる.

補題 2.3. X を非負値確率変数とする. このとき,

$$E[X] = 0 \Rightarrow \Pr(X = 0) = 1$$

となる.

Proof. まず

$$A := \{\omega \in \Omega; X(\omega) > 0\} = \{X > 0\},$$

$$A_n := \left\{ \omega \in \Omega; X(\omega) > \frac{1}{n} \right\} \quad (n = 1, 2, \dots)$$

とおく. すると $A_1 \subset A_2 \subset \dots$ かつ $\bigcup_{n=1}^{\infty} A_n = A$ となる. よって, 補題 1.7(5) から

$$\lim_{n \rightarrow \infty} \Pr(A_n) = \Pr\left(\bigcup_{n=1}^{\infty} A_n\right) = \Pr(A) \quad (2.1)$$

となる. 一方, $A_n \subset A$ なので, $0 \leq \mathbb{1}_{A_n}(\omega) \leq \mathbb{1}_A(\omega)$ ($\omega \in \Omega$) で⁴ある. このことに注意すると

$$0 = E[X] = E[\underbrace{X\mathbb{1}_{A^c}}_{\geq 0} + X\mathbb{1}_A] \geq E[X\mathbb{1}_A] \geq E[X\mathbb{1}_{A_n}]$$

$$\geq \frac{1}{n}E[\mathbb{1}_{A_n}] = \frac{1}{n}\Pr(A_n)$$

を得る⁵. したがって, $\Pr(A_n) = 0$ ($\forall n \in \mathbb{N}$) がわかる. (2.1) と合わせると $\Pr(A) = 0$ となる. よって

$$1 = \Pr(A^c) = \Pr\left(\left(\{X \leq 0\} \cap \{X \geq 0\}\right) \cup \left(\{X \leq 0\} \cap \{X < 0\}\right)\right)$$

$$= \Pr(X = 0) + \underbrace{\Pr(X < 0)}_{=0} = \Pr(X = 0)$$

が示せた. □

補題 2.4. $0 < q < r$ に対して

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

となる.

⁴ $\mathbb{1}_A$ は $\mathbb{1}_A : \Omega \ni \omega \mapsto \mathbb{1}_A(\omega) = \begin{cases} 1 & (\omega \in A) \\ 0 & (\omega \in A^c) \end{cases}$ なる確率変数である.

⁵

Proof. Young の不等式 (系 3.13)

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

において

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

とおくと

$$|x|^q \leq \frac{|x|^r}{s} + \frac{1}{t}$$

を得る. よって, 補題 2.2(1) からわかる. □

問 2.1. 連続確率変数 X が p.d.f. p をもつとき, 補題 2.4 の証明を具体的に書け. $q < r$ なので, $|x|^{q-r} \leq 1$ ($|x| > 1$) となることを使えばよい.

定理 2.5. $t > 0$ とする.

$$E[|X|^t] < \infty$$

ならば

$$\lim_{n \rightarrow \infty} n^t \Pr(|X| > n) = 0$$

となる.

Proof. 連続型の場合について証明を与えておく. X の p.d.f. を p と書いたとき

$$\infty > \int_{-\infty}^{\infty} |x|^t p(x) dx = \lim_{n \rightarrow \infty} \int_{-n}^n |x|^t p(x) dx$$

から

$$\lim_{n \rightarrow \infty} \int_{|x|>n} |x|^t p(x) dx = 0$$

となる. しかし

$$\begin{aligned} \lim_{n \rightarrow \infty} \int_{|x|>n} |x|^t p(x) dx &\geq \lim_{n \rightarrow \infty} n^t \int_{|x|>n} p(x) dx = \lim_{n \rightarrow \infty} n^t \Pr(|X| > n) \\ &= 0 \end{aligned}$$

がわかる. □

定理 2.6. X を確率変数とし, ある $t > 0$ に対して

$$\lim_{n \rightarrow \infty} n^t \Pr(|X| > n) = 0$$

をみたすとする. このとき, $0 < \forall s < t$ に対して

$$E[|X|^s] < \infty$$

となる.

注意 2.7. 定理 2.6 は $E[|X|^t] < \infty$ までは主張していないことに注意せよ. □

この定理の証明のために, 以下の補題と系を準備する.

補題 2.8. X を非負値確率変数で分布関数 F をもつとする. このとき, $E[X] < \infty$ または $\int_0^\infty \{1 - F(x)\} dx < \infty$ のいずれかが成り立つならば, もう片方も有限で

$$E[X] = \int_0^\infty \{1 - F(x)\} dx \tag{2.2}$$

が成り立つ.

Proof. X は連続型確率変数で p.d.f. p をもつ場合のみの証明を与えておく. さらに, 分布関数を F と記すことにする.

① $E[X] < \infty$ の場合: まず

$$E[X] = \int_0^\infty xp(x) dx = \lim_{n \rightarrow \infty} \int_0^n xp(x) dx$$

に注意する. 上式の右辺を部分積分を用いて書き換えると

$$\begin{aligned} \int_0^n xp(x) dx &= \{xF(x)\} \Big|_{x=0}^{x=n} - \int_0^n F(x) dx \\ &= nF(n) - 0 \cdot F(0) - \int_0^n F(x) dx \\ &= nF(n) - n + n \int_0^n F(x) dx \\ &= -n[1 - F(n)] + \int_0^n [1 - F(x)] dx \\ &= -n \Pr(X > n) + \int_0^n [1 - F(x)] dx \end{aligned}$$

となる. よって

$$E[X] = \lim_{n \rightarrow \infty} \left(-n\Pr(X > n) + \int_0^n [1 - F(x)] dx \right) \quad (2.3)$$

$$\begin{aligned} &\stackrel{\text{定理 2.6}}{=} 0 + \lim_{n \rightarrow \infty} \int_0^n [1 - F(x)] dx \\ &= \int_0^\infty [1 - F(x)] dx \end{aligned} \quad (2.4)$$

$$= \int_0^\infty [1 - F(x)] dx \quad (2.5)$$

がわかる. よって

$$E[X] < \infty \Rightarrow \int_0^\infty [1 - F(x)] dx < \infty$$

となり, (2.2) が成り立つことがわかる.

② $\int_0^\infty [1 - F(x)] dx < \infty$ の場合: $\forall n \in \mathbb{N}$ に対して, 部分積分を用いると

$$\int_0^n xp(x) dx = \int_0^n |x|p(x) dx = -n[1 - F(n)] + \int_0^n [1 - F(x)] dx$$

となることに注意する. $-n[1 - F(n)] \leq 0$ なので

$$\int_0^n |x|p(x) dx \leq \int_0^n [1 - F(x)] dx \leq \int_0^\infty [1 - F(x)] dx \quad (\forall n \in \mathbb{N})$$

を得る. したがって

$$\lim_{n \rightarrow \infty} \int_0^n |x|p(x) dx = \int_0^\infty |x|p(x) dx \leq \int_0^\infty [1 - F(x)] dx < \infty$$

から $E[X] < \infty$ がわかる. さらに, (2.5) から (2.2) もわかる. □

系 2.9. $s > 0$ とする. $E[|X|^s] < \infty$ または $s \int_0^\infty y^{s-1} \Pr(|X| > y) dy < \infty$ のいずれかが成り立つとき, もう片方も有限で

$$E[|X|^s] = s \int_0^\infty y^{s-1} \Pr(|X| > y) dy$$

が成り立つ.

Proof. $|X|^2$ の分布関数を $F^{|X|^s}$ と記す. 補題 2.8 から

$$E[|X|^s] \stackrel{\text{補題 2.8}}{=} \int_0^\infty [1 - F^{|X|^s}(z)] dz = \int_0^\infty \Pr(|X|^s > z) dz$$

となる. ここで $z = y^s$ と変換すると $\frac{dz}{dy} = sy^{s-1}$ と $dz = sy^{s-1}dy$ となる. したがって

$$\begin{aligned} \int_0^\infty \Pr(|X|^s > z) dz &= \int_0^\infty \Pr(|X|^s > y^s) sy^{s-1} dy \\ &= s \int_0^\infty y^{s-1} \Pr(|X|^s > y^s) dy \\ &= s \int_0^\infty y^{s-1} \Pr(|X| > y) dy \end{aligned}$$

となることがわかる. よって, 系 2.9 から, $E[|X|^s] < \infty$ となる. □

定理 2.6 の証明: 任意の $\epsilon > 0$ に対して, N をうまく選らぶと

$$\Pr(|X| > n) < \frac{\epsilon}{n^t} \quad (\forall n \geq N) \tag{2.6}$$

が成立する. さらに, $c < -1$ のとき

$$\int_N^\infty y^c dy = \left[-\frac{1}{c+1} y^{c+1} \right]_N^\infty = -\frac{1}{c+1} N^{c+1} < \infty \tag{2.7}$$

となること注意をする. すると

$$\begin{aligned} s \int_0^\infty y^{s-1} \Pr(|X| > y) dy &= s \int_0^N y^{s-1} \underbrace{\Pr(|X| > y)}_{\leq 1} dy + s \int_N^\infty y^{s-1} \underbrace{\Pr(|X| > y)}_{< \epsilon/y^t \text{ : (2.6)}} dy \\ &\leq \int_0^N sy^{s-1} dy + s\epsilon \int_N^\infty y^{s-1} \frac{1}{y^t} dy \\ &= N^s + s\epsilon \int_N^\infty y^{s-1-t} dy \\ &< \infty \end{aligned}$$

がわかる. 最後の不等号は $s - 1 - t < -1$ と (2.7) からわかる. よって, $0 < \forall s < t$ に対して

$$E[|X|^s] < \infty$$

がわかる. □

定義 2.10. (1) $k = 1, 2, \dots$ に対して, $E[|X|^k] < \infty$ のとき, $E[X^k]$ を X の k 次モーメント (または積率) という.

(2) $E[|X|] < \infty$ のとき $E[X]$ を X の平均値⁶という.

⁶簡単に「平均」ともいう.

(3) $E[X^2] < \infty$ のとき X の分散を

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

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⁷という. $A \in \mathcal{B}(\mathbb{R})$ のとき

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

となる.

注意 2.11. 分散 $\text{Var}[X]$ 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. \square

注意 2.12. (1) X は確率空間 $(\Omega, \mathcal{A}, \Pr)$ 上の離散型確率変数とする. \mathbf{p}^X を X の p.m.f. とし

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

とする. 関数 $g: \mathbb{R} \rightarrow \mathbb{R}$ 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

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

とする. $y \in S_Y$ に対して

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

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

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

⁷指示関数は \mathbb{R} の任意の部分集合に定義ができることに注意をせよ.

となる. さらにいずれかの和が有限ならば

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

となる. よって

$$E[Y] = E[g(X)]$$

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

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

とする. 関数 $g: S_X \rightarrow \mathbb{R}$ 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 $g: S_X \rightarrow S_Y$ と制限⁸すれば g の逆関数 $g^{-1}: S_Y \rightarrow S_X$ が存在し

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

となる. さらに $g(g^{-1}(y)) = y$ より

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

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

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

となる. よって

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

となる. g が狭義単調減少の場合もふくめると

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

⁸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.

となる.

以上のことから

$$\begin{aligned} E[Y] &= \int_{S_Y} y p^Y(y) dy \\ &= \int_{S_Y} y p^X(g^{-1}(y)) \frac{1}{\dot{g}(g^{-1}(y))} dy \\ &= \int_{S_X} g(x) p^X(x) dx \quad (\because x = g^{-1}(y) \Leftrightarrow y = g(x); dy = \dot{g}(x) dx) \\ &= E[g(X)] \end{aligned}$$

がわかる. □

2.2 確率ベクトルの期待値

確率変数 X, Y を確率空間 $(\Omega, \mathcal{A}, \Pr)$ 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 $p(x, y)$ とし, 連続型のときその同時 p.d.f. も $p(x, y)$ と書くことにする.

定義 2.13. 可測関数 $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ 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

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

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. $E[g^+(X, Y)]$ または $E[g^-(X, Y)]$ のいずれかが有限ならば

$$E[g(X, Y)] := E[g^+(X, Y)] - E[g^-(X, Y)]$$

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

注意 2.14. 3 つ以上の確率変数 X_1, X_2, \dots, X_n に対しても期待値を定義 2.13 と同様に定義する. \square

定理 2.15. X_1, X_2, \dots, X_n を確率変数とし, 各 $X_j (j = 1, 2, \dots, n)$ の期待値は有限とする. a_1, a_2, \dots, a_n を定数としたとき

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

となる.

Proof. まず, $|\sum_{j=1}^n a_j x_j| \leq \sum_{j=1}^n |a_j x_j|$ 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. $\mathbf{X} := (X_1, X_2, \dots, X_n)^\top$ とおき, \mathbf{X} は同時 p.d.f. $p^{\mathbf{X}}$ を持つとする. すると定義 2.13 から 補題 2.2(2) と同様の期待値の線型性から

$$\begin{aligned} E\left[\sum_{j=1}^n a_j X_j\right] &= \int_{\mathbb{R}^n} \left\{ \sum_{j=1}^n a_j x_j \right\} p^{\mathbf{X}}(x_1, x_2, \dots, x_n) dx_1 dx_2 \cdots dx_n \\ &= \sum_{j=1}^n a_j \int_{\mathbb{R}^n} x_j p^{\mathbf{X}}(x_1, x_2, \dots, x_n) dx_1 dx_2 \cdots dx_n \\ &= \sum_{j=1}^n a_j E[X_j] \end{aligned}$$

がわかる. \square

例 2.16. $0 < p < 1$ と $j = 1, 2, \dots, n$ に対して, $X_j \sim \text{Ber}(p)$ とする. このとき

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

である. $S = \sum_{j=1}^n X_j$ としたとき

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

がわかる. \square

2.3 分散と共分散

X を確率変数とし, $E[X^2] < \infty$ とする. X の分散を

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

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

定理 2.17. 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) $\text{Var}[X] = E[X^2] - \{E[X]\}^2$ となる.

(2) 定数 $a, b \in \mathbb{R}$ に対して

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

となる.

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

$$E[XY] = E[X]E[Y]$$

となる.

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

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

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 $p^X(x)p^Y(y)$ という形でかけるので

$$\begin{aligned} E[XY] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyp^X(x)p^Y(y) dx dy = \int_{-\infty}^{\infty} xp^X(x) dx \int_{-\infty}^{\infty} yp^Y(y) dy \\ &= E[X]E[Y] \end{aligned}$$

となること⁹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. \square

例 2.18. (例 2.16 の続き) 例 2.16 の設定に加えて X_1, X_2, \dots, X_n は互いに独立とする. 定理 2.15 に注意すれば

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

⁹ $\text{Var}[X] < \infty, \text{Var}[Y] < \infty$ なので, Cauchy-Schwarz の不等式から $E[|XY|] < \infty$ 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.

となる. これと定理 2.17(4) から

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

がわかる. □

定理 2.19. $n \geq 2$ とし, X_1, X_2, \dots, X_n は i.i.d. 確率変数列とし

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

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

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

で定義し, 標本 (不偏) 分散を

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

で定義する. このとき

$$(1) \mathbf{E}[\bar{X}_n] = \mu; \quad (2) \text{Var}[\bar{X}_n] = \frac{\sigma^2}{n}; \quad (3) \mathbf{E}[S_n^2] = \sigma^2$$

となる.

Proof. (1), (2) は定理 2.15, 2.17(4) よりわかる. (3) を証明するために

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

に注意する. 定理 2.17(1) より

$$\mathbf{E}\left[\sum_{j=1}^n (X_j - \mu)^2\right] = n\sigma^2$$

がわかる. さらに, (1) と (2) より

$$\mathbf{E}[(\bar{X}_n - \mu)^2] = \text{Var}[\bar{X}_n] = \frac{\sigma^2}{n}$$

が示される. したがって

$$\begin{aligned} \mathbf{E}[S_n^2] &= \frac{1}{n-1} \left\{ \sum_{j=1}^n \mathbf{E}[(X_j - \mu)^2] - n\mathbf{E}[(\bar{X}_n - \mu)^2] \right\} \\ &= \frac{1}{n-1} \{n\sigma^2 - \sigma^2\} = \sigma^2 \end{aligned}$$

がわかる. □

定義 2.20. X と Y は確率変数とし

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

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

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

で定義し, X と Y の (Pearson) の相関係数を

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

で定義する.

定理 2.21. 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

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

と書き直せる.

(2) $\text{Cov}[X, Y] = \text{Cov}[Y, X]$ である.

(3) 定数 a, b に対して

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

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

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

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 $\text{Var}[X] = \text{Var}[Y] = 0$ のときは明らかである. よって $\text{Var}[X] \neq 0$ として証明する. 補題 2.2(2) より

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

となる. ここで, 判別式をとれば

$$\{\text{Cov}[X, Y]\}^2 - \text{Var}[X]\text{Var}[Y] \leq 0 \quad (2.8)$$

がわかる. (5) は (2), (3) および定理 2.17(2) からわかる. (6) は定理 2.17(3) からわかる. \square

注意 2.22. (2.8) において, $E[X] = E[Y] = 0$ の場合を考えると Cauchy-Schwarz の不等式

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

と呼ばれる重要なものを得る.

注意 2.23. (6) の逆は一般に正しくない. 反例は下の問いをみよ. \square

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) $E[X] = 0, E[Y] = 0, E[XY] = 0$ を確かめよ.
- (2) $\Pr(X \leq 1/2, Y \leq 1/2) = \Pr((\pi/3) \leq Z \leq (5/6)\pi)$ を確かめよ.
- (3) $\Pr(X \leq 1/2) = \Pr((\pi/3) \leq Z \leq (7/3)\pi)$ を確かめよ¹⁰.

定理 2.24. (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 $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}[X, Y]$ となる.

(2) $d \geq 2$ とする. X_1, X_2, \dots, 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

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

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. \square

問 2.3. 定理 2.24(1) の証明を具体的に書け.

¹⁰これらより, $\Pr(X \leq 1/2, Y \leq 1/2) \neq \Pr(X \leq 1/2)\Pr(Y \leq 1/2)$ となり, $\text{Cov}[X, Y] = 0$ だが, X と Y は従属であることがわかる.

定義 2.25. X_1, X_2, \dots, X_d を有限な 2 次の積率を持つ確率変数とし

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

と書く. このとき確率ベクトル \mathbf{X} の期待値を

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

で定義する. \mathbf{X} の共分散を

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

で定義する¹¹. ただし $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_d)^\top := E[\mathbf{X}]$ である. これは

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

である.

注意 2.26. 定義から $\text{Var}[\mathbf{X}]$ 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 $\mathbf{a} \in \mathbb{R}^d$ に対して

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

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

補題 2.27. $\mathbf{X} = (X_1, X_2, \dots, X_d)^\top$ 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

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

¹¹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.

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¹²である.

(1) 任意の定数ベクトル $\boldsymbol{a} \in \mathbb{R}^d$ に対して

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

となる.

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

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

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:

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

とおくと

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

がわかる. つぎに, 共分散の定義と上の結果から

$$\begin{aligned} \text{Var}[\boldsymbol{A}\boldsymbol{X}] &= E[\{\boldsymbol{A}\boldsymbol{X} - E[\boldsymbol{A}\boldsymbol{X}]\}\{\boldsymbol{A}\boldsymbol{X} - E[\boldsymbol{A}\boldsymbol{X}]\}^\top] \\ &= E[\{\boldsymbol{A}\boldsymbol{X} - \boldsymbol{A}E[\boldsymbol{X}]\}\{\boldsymbol{A}\boldsymbol{X} - \boldsymbol{A}E[\boldsymbol{X}]\}^\top] \\ &= E[\boldsymbol{A}\{\boldsymbol{X} - E[\boldsymbol{X}]\}\{\boldsymbol{X} - E[\boldsymbol{X}]\}^\top \boldsymbol{A}^\top] \\ &= \boldsymbol{A}E[\{\boldsymbol{X} - E[\boldsymbol{X}]\}\{\boldsymbol{X} - E[\boldsymbol{X}]\}^\top] \boldsymbol{A}^\top \\ &= \boldsymbol{A}\text{Var}[\boldsymbol{X}]\boldsymbol{A}^\top \end{aligned}$$

からわかる. □

問 2.4. $d = 2$ として, 補題 2.27 を確認せよ.

¹² $d \times d$ の対称行列 \boldsymbol{A} が半正定値であるとは, $\forall \boldsymbol{a} \in \mathbb{R}^d$ に対して $\boldsymbol{a}^\top \boldsymbol{A} \boldsymbol{a} \geq 0$ が成立するときをいう. また $d \times d$ の対称行列 \boldsymbol{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 $\boldsymbol{a}^\top \boldsymbol{A} \boldsymbol{a} > 0$ が成立するときをいう.

系 2.28. 確率ベクトル \mathbf{X} は, 任意の定数ベクトル $\mathbf{a}(\neq \mathbf{0})$ に対して

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

をみます. このとき, Σ は正定値である.

注意 2.29. \mathbf{X} の共分散行列 Σ は, 補題 2.27(1) から半正定値であることがわかる. 系 2.28 の仮定は, 確率ベクトル \mathbf{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 \mathbf{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. \square

2.4 条件付き期待値

定義 2.30. (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 $p^{X|Y}$ (または $p^{X|Y}$) とする. $Y = y$ を与えたときの X の条件付き期待値を

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

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 $E[|X|] < \infty$ とする.

(2) (Borel 可測) 関数 $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ に対して, $g(X, Y)$ の条件付き期待値を

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

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, $E[|g(X, Y)|] < \infty$ とする.

注意 2.31. 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, $E[|g(X)|] < \infty$ の仮定のもとでは, $E[g^+(X)|Y = y] < \infty$, $E[g^-(X)|Y = y] < \infty$ が保証されるので, 上記のように簡単に定義をした.

注意 2.32. $E[X]$ は定数であるが, $E[X|Y = y]$ は一般に y の関数である. このことから

$$h(y) := E[X|Y = y]$$

とおいたときに $h(y)$ に Y を代入したものの $h(Y)$ は確率変数¹³になる. これを

$$E[X|Y] := h(Y)$$

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

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

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, pp.181 – 182] を参照のこと. \square

例 2.33. 連続型確率変数 Y は p.d.f.

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

を持つとする. $Y = y (0 < y < 1)$ を観測したとき

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

とする. すなわち $0 < y < 1$ のとき

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

である. よって

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

となる. これより

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

となる. \square

定理 2.34. (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

$$E[X + Y|Z] = E[X|Z] + E[Y|Z]$$

となる.

¹³ $h \mapsto h(y)$ の可測性は測度論の知識が必要となる. このことは定理 ?? から保証される.

(2) 有限な期待値を持つ確率変数 X, Y に対して

$$E[E[Y|X]] = E[Y], \quad E[E[X|Y]] = E[X]$$

となる.

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

$$E[E[g(X, Y)|Y]] = E[g(X, Y)]$$

となる.

(4) $E[XY|Y] = YE[X|Y]$ 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

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

となる. ただし p^X は X の周辺 p.d.f. で, $p^{Y|X}$ は $X = x$ を与えたときの Y の条件付き p.d.f. である. $E[Y|X = x] =: h(x)$ とおいたとき, $E[Y|X] := h(X)$ と書いたことを思い出す. また, $S_X = \{x \in \mathbb{R}; p^X(x) > 0\}$ とおいたとき,

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

となる¹⁴. 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¹⁵が知られている. (3) と (4) も同様に証明できる. \square

問 2.5. 定理 2.34(3)(4) の証明を書け.

例 2.35. (例 2.33 の続き)

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

となる. 一方 $0 < y < x < 1$ に対して

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

となる. よって

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

となる. \square

定義 2.36. 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. $p^{X|Y}$ (p.m.f. $p^{X|Y}$) 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

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

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

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

とも書ける.

定理 2.37. X, Y を確率変数とし $E[X^2] < \infty$ とする. このとき

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

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

¹⁴ h は $S_Y := \{y \in \mathbb{R}; p^Y(y) > 0\}$ 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.

¹⁵Fubini の定理からわかる.

Proof. まず

$$\begin{aligned} \text{Var}[X] &= E[\{X - E[X]\}^2] \\ &= E[\{X - E[X|Y] + E[X|Y] - E[X]\}^2] \\ &= E[\{X - E[X|Y]\}^2] + E[\{E[X|Y] - E[X]\}^2] \\ &\quad + 2E[\{X - E[X|Y]\}\{E[X|Y] - E[X]\}] \end{aligned} \quad (2.9)$$

と書き直す. しかし, (2.9) の最右辺の各項は以下のように評価できる.

$$\begin{aligned} E[\{X - E[X|Y]\}^2] &= E[X^2 - 2XE[X|Y] + \{E[X|Y]\}^2] \\ &= E\left[E[X^2 - 2XE[X|Y] + \{E[X|Y]\}^2] \middle| Y\right] \\ &\quad (\because \text{定理 2.34(3)}) \\ &= E\left[E[X^2|Y] - 2E[X|Y]E[X|Y] + \{E[X|Y]\}^2\right] \\ &= E\left[E[X^2|Y] - \{E[X|Y]\}^2\right] \\ &= E[\text{Var}[X|Y]], \\ E[\{E[X|Y] - E[X]\}^2] &= E\left[\{E[X|Y] - \underbrace{E[E[X|Y]]}_{=E[X]}\}^2\right] \\ &\quad (\because \text{定理 2.34(3)}) \\ &= \text{Var}\left[E[X|Y]\right], \\ E[\{X - E[X|Y]\}\{E[X|Y] - E[X]\}] &= E\left[E[\{X - E[X|Y]\}\{E[X|Y] - E[X]\}] \middle| Y\right] \\ &\quad (\because \text{定理 2.34(3)}) \\ &= E\left[\{E[X|Y] - E[X]\} \underbrace{E[X - E[X|Y]|Y]}_{=0}\right] \\ &= 0 \end{aligned}$$

となる. 最後から 2 番目の等号は定理 2.34(3) を用いた. これらの結果を (2.9) の最右辺の各項に代入すれば, 定理は証明される. \square

2.5 積率母関数

定義 2.38. X を確率変数とし, ある $t_0 > 0$ が存在して, $E[e^{tX}] < \infty$ ($\forall |t| < t_0$) とする. このとき, X の積率母関数 (Moment Generating Function (m.g.f.))

を

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

と定義する.

注意 2.39. 確率変数 X の積率母関数 $m^X(t)$ 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¹⁶が知られている. このことから

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

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

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

がわかる. □

例 2.40. $X \sim \text{Exp}(1)$ とする. $t < 1$ に対して

$$m^X(t) = E[e^{tX}] = \int_0^\infty e^{tx} e^{-x} dx = \int_0^\infty e^{-(1-t)x} dx = \frac{1}{1-t}$$

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

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

となる. 簡単な計算から

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

なので

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

となる. □

補題 2.41. (1) $a, b \in \mathbb{R}$ ($a \neq 0$) とする. $Y = aX + b$ としたとき

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

となる.

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

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

となる.

¹⁶たとえば, [39, 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 ($j = 1, 2, \dots, d$) の独立性と指数関数の性質に注意して, 定理 2.17(3) を適用するとわかる. \square

定理 2.42. X と Y を確率変数とする. ある数 $t_0 > 0$ が存在して

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

ならば

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

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

$$X \stackrel{d}{=} Y \Leftrightarrow F^X(x) = F^Y(x) \quad (\forall x \in \mathbb{R})$$

である.

Proof. これは信じることにする. \square

注意 2.43. 定理 2.42 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 2.44. $n_1, n_2 \in \mathbb{N}, 0 < p < 1$ とし $X_1 \sim \text{Bino}(n_1, p), X_2 \sim \text{Bino}(n_2, p)$ は独立とする. $Y = X_1 + X_2$ としたとき

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

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

例 2.45. $\lambda_1, \lambda_2 > 0$ とし, $X_1 \sim \text{Po}(\lambda_1), X_2 \sim \text{Po}(\lambda_2)$ は独立とする. $Y = X_1 + X_2$ としたとき

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

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

2.6 章末注釈と参考文献

この章は [36] を参考にした.

2.7 演習問題

演習問題 2.1. (1) 非負の正数 n に対して

$$(1+x)^n = \sum_{k=0}^n \binom{n}{k} x^k$$

を帰納法を用いて証明せよ. ただし

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}, \quad 0! = 1$$

である.

(2)

$$\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n} = 2^n \quad (n \geq 0)$$

を示せ.

(3)

$$\binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \binom{n}{3} + \cdots + (-1)^n \binom{n}{n} = 0 \quad (n \geq 1)$$

を示せ.

(4)

$$\sum_{k=1}^n k \cdot \binom{n}{k} = n2^{n-1} \quad (n \geq 1)$$

を示せ.

(5)

$$\sum_{k=2}^n k(k-1) \cdot \binom{n}{k} = n(n-1)2^{n-2} \quad (n \geq 2)$$

を示せ.

演習問題 2.2. 確率空間 $(\Omega, \mathcal{A}, \Pr)$ 上の確率変数 X は

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

をみたすとする. このとき, X の平均 $E[X]$ と分散 $\text{Var}[X]$ を求めよ.

演習問題 2.3. (1) $X \sim \text{Ber}(\theta)$ ($0 < \theta < 1$) のとき

$$E[X] = \theta, \quad \text{Var}[X] = \theta(1 - \theta)$$

を示せ.

(2) $X_1, X_2, \dots, X_n \sim \text{i.i.d. Ber}(\theta)$ ($0 < \theta < 1$) とする. $S = \sum_{j=1}^n X_j$ とおいたとき, $S \sim \text{Bino}(n, \theta)$ となることを示せ.

(3) $S \sim \text{Bino}(n, \theta)$ ($n \in \mathbb{N}, 0 < \theta < 1$) のとき

$$E[S] = n\theta, \quad \text{Var}[S] = n\theta(1 - \theta)$$

を示せ.

(4) $X \sim \text{Po}(\theta)$ ($\theta > 0$) とする. このとき, X の平均 $E[X]$ と分散 $\text{Var}[X]$ を求めよ.

(5) $U \sim U(0, 1)$ のとき

$$E[X] = \frac{1}{2}, \quad \text{Var}[X] = \frac{1}{12}$$

を示せ.

(6) $X \sim \text{Ex}(\theta)$ ($\theta > 0$) とする. このとき

$$E[X] = \frac{1}{\theta}, \quad \text{Var}[X] = \frac{1}{\theta^2}$$

を示せ.

(6) $X \sim N(\mu, \sigma^2)$ ($\mu \in \mathbb{R}, \sigma > 0$) とする. このとき

$$E[X] = \mu, \quad \text{Var}[X] = \sigma^2$$

を示せ.

(7) $X \sim \chi_n^2$ のとき

$$E[X] = n, \quad \text{Var}[X] = 2n$$

を示せ.

演習問題 2.4. 確率空間 $(\Omega, \mathcal{A}, \text{Pr})$ 上の確率変数 X は $\text{Var}[X] = \sigma^2$ ($0 < \sigma < \infty$) とする. $c (\neq 0)$ を定数としたとき, $\text{Var}[X + c]$ と $\text{Var}[cX]$ を σ と c を用いて表せ.

演習問題 2.5. (1) $X \sim \text{Ga}(\alpha, \beta)$ とする. X の k 次の積率は以下で与えられることを示せ.

$$E[X^k] = \frac{(a + k - 1)(a + k - 2) \cdots \alpha}{\beta^k}.$$

(2) 上の問いの結果から

$$E[X] = \frac{\alpha}{\beta}, \quad E[X^2] = \frac{(\alpha + 1)\alpha}{\beta^2}, \quad \text{Var}[X] = \frac{\alpha}{\beta^2}$$

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.

$$m^X(t) = \left(\frac{1}{1 - \beta t} \right)^\alpha \quad (t < 1/\beta).$$

演習問題 2.6. X を確率空間 $(\Omega, \mathcal{A}, \text{Pr})$ 上の連続型確率変数とし, X の p.d.f. を

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

とする. さらに, 離散型確率変数 Y を

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

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 $E[X]$ と分散 $\text{Var}[X]$ を求めよ.
- (2) 確率変数 Y の平均 $E[Y]$ と分散 $\text{Var}[Y]$ を求めよ.
- (3) 確率変数 X と Y の共分散 $\text{Cov}[X, Y]$ を求めよ.

演習問題 2.7. X_1, X_2 を確率空間 $(\Omega, \mathcal{A}, \text{Pr})$ 上の連続型確率変数とし,

$$X_1 \sim N(x_0 + \mu, \sigma^2), \quad X_2 | X_1 = x_1 \sim N(x_1 + \mu, \sigma^2)$$

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 $N(\mu, \sigma^2)$ に従う確率変数 Z の期待値と分散が $E[Z] = \mu, \text{Var}[Z] = \sigma^2$ 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 $E[X_2 - X_1]$ を求めよ.
- (2) X_2 の分散 $\text{Var}[X_2]$ を求めよ.
- (3) X_1 と X_2 の共分散 $\text{Cov}[X_1, X_2]$ を求めよ.
- (4) $X_2 - X_1$ の分散 $\text{Var}[X_2 - X_1]$ を $\text{Var}[X_1], \text{Var}[X_2], \text{Cov}[X_1, X_2]$ で表現せよ.
- (5) $X_2 - X_1$ の分散 $\text{Var}[X_2 - X_1]$ を求めよ.

演習問題 2.8. $(\Omega, \mathcal{A}, \Pr)$ 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 は开区間 $(0, 1)$ 上の一様分布に従うとし

$$\mathbf{X} = \begin{bmatrix} X \\ X^2 \end{bmatrix}$$

とおく. このとき, 以下の問いに答えよ.

- (1) 確率ベクトル \mathbf{X} の期待値 $E[\mathbf{X}]$ を求めよ.
- (2)

$$\text{Var}[\mathbf{X}] = E[\mathbf{X}\mathbf{X}^\top] - E[\mathbf{X}]\{E[\mathbf{X}]\}^\top$$

を示せ.

- (3) 確率ベクトル \mathbf{X} の分散共分散行列 $\text{Var}[\mathbf{X}]$ を求めよ.
- (4) $\det[\text{Var}[\mathbf{X}]]$ を求めよ.

演習問題 2.9. 確率空間 $(\Omega, \mathcal{A}, \Pr)$ 上で定義された連続型確率ベクトル (X, Y) は同時 p.d.f.

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

を持つとする.

- (1) X の周辺 p.d.f. $p^X(x)$ を求め,

$$\int_{-\infty}^{\infty} p^X(x) dx = 1$$

を確認せよ.

- (2) Y の周辺 p.d.f. $p^Y(y)$ を求め,

$$\int_{-\infty}^{\infty} p^Y(y) dy = 1$$

を確認せよ.

- (3) X と Y の期待値 $E[X]$ と $E[Y]$ を求めよ.
- (4) XY の期待値 $E[XY]$ を求め, X と Y の共分散 $\text{Cov}[X, Y]$ を求めよ.
- (5) $p^Y(y) > 0$ 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. $p^{X|Y}(x|y)$ を求めよ. $p^{X|Y}(x|y) > 0$ となる x の範囲を書いたうえで,

$$\int_{-\infty}^{\infty} p^{X|Y}(x|y) dx = 1$$

を確認せよ.

演習問題 2.10. 確率空間 $(\Omega, \mathcal{A}, \Pr)$ 上で定義された連続型確率ベクトル (X, Y) は同時 p.d.f.

$$p^{(X,Y)}(x, y) = \begin{cases} e^{-(x+y)} & (0 < x < \infty, 0 < y < \infty) \\ 0 & (\text{その他の場合}) \end{cases}$$

を持つ.

$$U = 2X, \quad V = X + Y$$

とおく.

(1) 確率ベクトル (U, V) の同時 p.d.f. $p^{(U,V)}(u, v)$ を求めよ. さらに

$$T = \{(u, v) \in \mathbb{R}^2 : p^{(U,V)}(u, v) > 0\}$$

を図示せよ.

(2) U の周辺 p.d.f. $p^U(u)$ を求めよ.

演習問題 2.11. $\lambda > 0$ とし, W は確率空間 $(\Omega, \mathcal{A}, \Pr)$ 上で定義された確率変数で, 母数 λ の指数分布 $\text{Exp}(\lambda)$ に従うとする. すなわち, W は連続型確率変数で, p.d.f.

$$p^W(w) = \begin{cases} \lambda e^{-\lambda w} & (w > 0) \\ 0 & (\text{その他の場合}) \end{cases}$$

をもつ. 同じ確率空間 $(\Omega, \mathcal{A}, \Pr)$ 上で定義された確率変数 X と Y を

$$X = [W], \quad Y = W - X$$

で定める. ただし, $[x]$ は実数 x 以下の最大の整数とする. たとえば, $[1.5] = 1$ 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

$$\Pr(X = x)$$

を求めることで X の p.m.f. p^X を求めよ.

(2) $\sum_{x=0}^{\infty} p^X(x)$ 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

$$\Pr(Y \leq y | X = x)$$

を求めよ.

(4) x を非負の整数とする. $y \in \mathbb{R}$ に対して, $F^{Y|X}(y|x) = \Pr(Y \leq y | X = x)$ と定める. $F^{Y|X}(y|x)$ 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. $p^{Y|X}(y|x)$ を求めよ.

(6) 期待値 $E[Y]$ を求めよ.

補足 $\Pr(0 < Y < 1) = 1$ から $0 < E[Y] < 1$ となる. 得られた結果について, 不等式 $0 < E[Y] < 1$ となっているかを確認するとよい.

演習問題 2.12. X を確率空間 $(\Omega, \mathcal{A}, \Pr)$ 上の連続型確率変数とし, X の p.d.f. を

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

とする. さらに, 離散型確率変数 Y を

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

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 $E[X]$ と分散 $\text{Var}[X]$ を求めよ.
- (2) 確率変数 Y の平均 $E[Y]$ と分散 $\text{Var}[Y]$ を求めよ.
- (3) 確率変数 X と Y の共分散 $\text{Cov}[X, Y]$ を求めよ.

演習問題 2.13. (U, Y) の同時分布を以下のように定める. $U \sim \text{Unif}(0, 1)$ 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 $N(\theta, \sigma^2)$ とする. ただし,

$$\theta = 2 - 6u, \quad \sigma^2 = 25$$

とする.

- (1) U の期待値 $E[U]$ を計算せよ.
- (2) U の分散 $\text{Var}[U]$ 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 $E[Y|U = u]$ を答えよ.
- (4) Y の期待値 $E[Y]$ を計算せよ.
- (5) Y の分散 $\text{Var}[Y]$ を求めよ.
- (6) $U = u$ ($0 < u < 1$) を与えたときの条件付き分散 $\text{Var}[Y|U = u]$ を答えよ.
- (7) $E[\text{Var}[Y|U]] + \text{Var}[E[Y|U]]$ を計算せよ.
- (8) 連続関数 $g: [0, 1] \rightarrow \mathbb{R}$ に対して

$$E[(Y - g(U))^2] \geq 25$$

となること示せ.

演習問題 2.14. $X \sim N(0, 1)$ とする. $\text{Var}[g(X)] < \infty$ なる 絶対連続関数 $g: \mathbb{R} \rightarrow \mathbb{R}$ に対して

$$\text{Var}[g(X)] \leq E[\{\dot{g}(X)\}^2]$$

が成立することを示せ.

演習問題 2.15. 確率変数 X と Y は独立同一に共通の分布関数 F と p.d.f. p を持つ分布に従うとする.

(1) $V = \max X, Y$ は分布関数 $F^V(v) = \{F(v)\}^2 (v \in \mathbb{R})$ と p.d.f. $p^V(v) = 2p(v)F(v)$ を持つことを示せ.

(2) $U = \min\{X, Y\}$ の分布関数 F^U と p.d.f. p^U を求めよ.

(3) X と Y は独立同一に $[0, 1]$ 上の一様分布に従うとき, $E[U]$ と $\text{Cov}[U, V]$ を求めよ.

(4) X と Y は独立同一に指数分布

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

に従うとき, U の分布を求めよ. さらに, $E[V]$ と $\text{Var}[V]$ を求めよ.

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 $(\Omega, \mathcal{A}, \Pr)$ 上の非負値確率変数を X とし, $E[X] < \infty$ とする. このとき, $\forall t > 0$ に対して

$$\Pr(X \geq t) \leq \frac{E[X]}{t}$$

が成り立つ.

Proof. X は連続型で p.d.f. p を持つ場合を示す.

$$\begin{aligned} E[X] &= \int_0^{\infty} xp(x) dx = \int_0^t xp(x) dx + \int_t^{\infty} xp(x) dx \geq \int_t^{\infty} xp(x) dx \\ &\geq t \int_t^{\infty} p(x) dx = t\Pr(X \geq t) \end{aligned}$$

からわかる. □

系 3.2. $\lambda > 0$ とする. X を確率変数とし $E[e^{\lambda X}] < \infty$ とする. このとき $\forall t > 0$ に対して

$$\Pr(X \geq t) \leq e^{-\lambda t} E[e^{\lambda X}]$$

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

$$\Pr(X \geq t) = \Pr(e^{\lambda X} \geq e^{\lambda t}) \leq e^{-\lambda t} E[e^{\lambda X}]$$

となる. □

系 3.3. (Chebyshev の不等式) X を確率変数とし $\mu = E[X]$, $\sigma^2 = \text{Var}[X] < \infty$ ($0 < \sigma < \infty$) とする. このとき $\forall t > 0$ に対して

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

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

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

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

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

とみたすとき $\forall \lambda > 0$ に対して

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

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. $E[X] = 0$ に注意して上の式の両辺の期待値を計算すると

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

がわかる. いま

$$g(0) = \dot{g}(0) = 0, \quad \ddot{g}(u) \leq \frac{1}{4} (u > 0) \quad (3.1)$$

に注意する. 実際

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

からわかる. (3.1) に注意して, Taylor 展開をすれば

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

がわかる. よって

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

を得る. □

定理 3.5. (Hoeffding の不等式) $a_j, b_j (j = 1, 2, \dots, n)$ は $a_j < 0 < b_j$ なる定数とする. 確率空間 $(\Omega, \mathcal{A}, \Pr)$ 上の確率変数 X_1, X_2, \dots, X_n は独立で

$$E[X_j] = 0, \quad a_j \leq X_j \leq b_j$$

をみたすとする. このとき, $\forall t > 0$ に対して

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

が成立する.

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

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

より

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

がわかる. □

定理 3.6. $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Ber}(\theta)$ とする. ただし $0 < \theta < 1$. このとき $\forall t > 0$ に対して

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

となる. ただし $\bar{X}_n = (1/n) \sum_{j=1}^n X_j$ である.

Proof. $j = 1, 2, \dots, n$ に対して

$$Y_j = \frac{1}{n}(X_j - \theta)$$

とおけば

$$\mathbf{E}[Y_j] = 0, \quad a \leq Y_j \leq b, \quad a = -\frac{\theta}{n}, \quad b = \frac{1-\theta}{n}, \quad (b-a)^2 = \frac{1}{n^2}$$

となる. よって定理 3.5 より

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

となる. 同様に

$$\Pr(\bar{X}_n - \theta \leq -t) = \Pr\left(\sum_{j=1}^n Y_j \leq -t\right) = \Pr\left(\sum_{j=1}^n (-Y_j) \geq t\right) \leq e^{-2nt^2} \quad (3.3)$$

がわかる. よって, (3.2) と (3.3) を合わせると

$$\begin{aligned} \Pr(|\bar{X}_n - \theta| \geq t) &= \Pr(\{\bar{X}_n - \theta \geq t\} \cup \{\bar{X}_n - \theta \leq -t\}) \\ &\leq \Pr(\bar{X}_n - \theta \geq t) + \Pr(\bar{X}_n - \theta \leq -t) \\ &\leq 2e^{-2nt^2} \end{aligned}$$

がわかる. □

例 3.7. $X_1, X_2, \dots, X_{100} \stackrel{\text{i.i.d.}}{\sim} \text{Ber}(\theta)$ とする. ただし $0 < \theta < 1$ である.
 (1) $n = 100$ とし, 事象 $|\bar{X}_{100} - \theta| \geq 0.2$ 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

$$\Pr(|\bar{X}_{100} - \theta| \geq 0.2) \leq 0.0625$$

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

$$\Pr(|\bar{X}_{100} - \theta| \geq 0.2) \leq 2e^{-2(100)(0.2)^2} = 0.00067$$

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

$$\Pr\left(|\bar{X}_n - \theta| \geq \sqrt{\frac{1}{2n} \log\left(\frac{2}{\alpha}\right)}\right) \leq \alpha$$

となる. これより

$$\Pr\left(|\bar{X}_n - \theta| \leq \sqrt{\frac{1}{2n} \log\left(\frac{2}{\alpha}\right)}\right) \geq 1 - \alpha$$

を得る. よって

$$C = \left[\bar{X}_n - \sqrt{\frac{1}{2n} \log\left(\frac{2}{\alpha}\right)}, \bar{X}_n + \sqrt{\frac{1}{2n} \log\left(\frac{2}{\alpha}\right)} \right]$$

とおけば

$$\Pr(\theta \in C) \geq 1 - \alpha$$

を得る. すなわち信頼係数 $(1 - \alpha)$ の θ の信頼区間

$$\left[\bar{X}_n - \sqrt{\frac{1}{2n} \log \left(\frac{2}{\alpha} \right)}, \bar{X}_n + \sqrt{\frac{1}{2n} \log \left(\frac{2}{\alpha} \right)} \right]$$

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) $Z \sim N(0, 1)$ とする. このとき $\forall t > 0$ に対して

$$\Pr(|Z| \geq t) \leq \sqrt{\frac{2}{\pi}} \frac{e^{-t^2/2}}{t}$$

が成り立つ.

Proof. まず

$$\begin{aligned} \Pr(Z \geq t) &= \frac{1}{\sqrt{2\pi}} \int_t^\infty \exp\left(-\frac{z^2}{2}\right) dz \\ &\leq \frac{1}{\sqrt{2\pi}} \int_t^\infty \frac{z}{t} \exp\left(-\frac{z^2}{2}\right) dz \\ &= \frac{1}{t\sqrt{2\pi}} \int_t^\infty \left\{ -\frac{\partial}{\partial z} \exp\left(-\frac{z^2}{2}\right) \right\} dz \\ &= \frac{1}{t\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) \end{aligned} \tag{3.4}$$

が成り立つことに注意する. さらに, $N(0, 1)$ の p.d.f. が偶関数であることに注意すると

$$\begin{aligned} \Pr(|Z| \geq t) &= \Pr(\{Z \geq t\} \cup \{Z \leq -t\}) \\ &\leq \Pr(Z \geq t) + \Pr(Z \leq -t) \\ &= 2\Pr(Z \geq t) \end{aligned}$$

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

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

となる.

Proof. $E[X^2] = E[Y^2] = 0$ のとき, 補題 2.3 から $\Pr(X = 0) = \Pr(Y = 0) = 1$ となるので, 不等式は自明である. このことを踏まえ, $E[X^2] \neq 0$ として証明を進める.

いま, $g(t) := E[(tX - Y)^2]$ とおく. 期待値の中を展開して期待値の線型性を用いると

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

となる. これから

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

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

$$g(t) = E[(cX - Y)^2] = 0 \Leftrightarrow \Pr(Y = cX) = 1$$

となる¹. □

¹非負値確率変数 X に対して

$$E[X] = 0 \Leftrightarrow \Pr(X = 0) = 1$$

であることに注意せよ. 実際 $\Pr(X > 0) > 0$ と仮定する. するとあるの $\epsilon > 0$ が存在して $\Pr(X > \epsilon) > 0$ となる. しかし $X \geq \epsilon \mathbb{1}\{X > \epsilon\}$ より

$$0 = E[X] \geq \epsilon E[\mathbb{1}\{X > \epsilon\}] = \epsilon \Pr(X > \epsilon) > 0$$

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

定義 3.11. 関数 $g: \mathbb{R} \rightarrow \mathbb{R}$ が凸であるとは各 $x, y \in \mathbb{R}$ と $0 \leq t \leq 1$ に対して

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

が成立するときをいう. さらに $-g$ が凸のとき g は concave であるという.

定理 3.12. (Jensen の不等式) X を有限な期待値を持つ確率変数とする.
(1) 関数 $g: \mathbb{R} \rightarrow \mathbb{R}$ は凸で $g(X)$ の期待値は有限のとき

$$E[g(X)] \geq g(E[X])$$

となる.

(2) g が concave のとき

$$E[g(X)] \leq g(E[X])$$

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

$$g(E[X]) + r\{x - E[X]\} \leq g(x)$$

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

$$g(E[X]) \leq E[g(X)]$$

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.

²このような直線を下から支える直線という. 関数が凸であることとその関数上の任意の点において下から支える直線が存在することが同値になることが知られている. 詳しくは補遺の A.12 章を参照.

Proof. 関数 g を凸とし, $Y \sim \text{Unif}(0, 1)$ とする. 可積分関数 $h: \mathbb{R} \rightarrow \mathbb{R}$ に対して $X = h(Y)$ とすれば

$$g\left(\int_0^1 h(y) dy\right) = g(\mathbb{E}[X]) \leq \mathbb{E}[g(X)] = \int_0^1 g(h(y)) dy$$

を得る. ここで

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

とおけば g は凸なので

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

がわかる. □

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

$$\mathbb{E}[|XY|] \leq \{\mathbb{E}[|X|^p]\}^{1/p} \{\mathbb{E}[|Y|^q]\}^{1/q}$$

となる.

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

$$\{\mathbb{E}[|X + Y|^p]\}^{1/p} \leq \{\mathbb{E}[|X|^p]\}^{1/p} + \{\mathbb{E}[|Y|^p]\}^{1/p}$$

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

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

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

$$E[|XY|] \leq \frac{E[|X|^p]}{p} + \frac{E[|Y|^q]}{q} = \frac{1}{p} + \frac{1}{q} = 1 \quad (3.7)$$

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

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

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. $E[|X + Y|^p] = 0$ のときは, 不等式は自明なので, $E[|X + Y|^p] \neq 0$ を仮定しても一般性を失わない. このとき

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

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

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 [32, 4 章] を参考にした.

3.4 演習問題

演習問題 3.1. $X \sim \text{Po}(\theta)$ ($\theta > 0$) とする. このとき, 以下の問いに答えよ.

(1) X の平均 $E[X]$ と分散 $\text{Var}[X]$ を求めよ.

(2) Chebyshev の不等式を用いて

$$\Pr(X \geq 2\theta) \leq \frac{1}{\theta}$$

を示せ.

演習問題 3.2. $n \in \mathbb{N}$ とする. $1 < p < \infty$, $p^{-1} + q^{-1} = 1$ のとき³, 任意の実数 x_j, y_j ($j = 1, 2, \dots, n$) に対して

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

が成り立つことを Young の不等式を利用して示せ. この不等式を Hölder の不等式という.

演習問題 3.3. $s > t > 0$ とする. 確率変数 X が $E[|X|^s] < \infty$ ならば, $E[|X|^t] < \infty$ となることを Young の不等式を用いて示せ.

演習問題 3.4. $X \sim N(0, 1)$ とする. $\text{Var}[g(X)] < \infty$ なる可微分関数 $g: \mathbb{R} \rightarrow \mathbb{R}$ に対して

$$\text{Var}[g(X)] \leq E[\{\dot{g}(X)\}^2]$$

が成立することを示せ. ただし, $\dot{g}(x) = \frac{dg}{dx}(x)$ である.

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.

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, \dots 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 $\lim_{n \rightarrow \infty} x_n = x$ と書いた。たとえば $x_n = x (\forall n \in \mathbb{N})$ ならば、明らかに $\lim_{n \rightarrow \infty} x_n = x$ となる。これと同じことを確率空間 $(\Omega, \mathcal{A}, \Pr)$ 上の確率変数列について考えてみる。

$$X_1, X_2, \dots \stackrel{\text{i.i.d.}}{\sim} \mathbf{N}(0, 1)$$

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 $\mathbf{N}(0, 1)$ に従う。さらに別の確率変数 X も $\mathbf{N}(0, 1)$ に従うとする。このとき X_n は X に「収束」するをしたい。しかし

$$\Pr(X_n \neq X) = 1 \quad (\forall n \in \mathbb{N})$$

である。

別の例をあげる。 X_1, X_2, \dots は独立で

$$X_n \sim \mathbf{N}\left(0, \frac{1}{n}\right) \quad (n = 1, 2, \dots)$$

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

$$\Pr(X_n = 0) = 0$$

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, \dots を i.i.d. 確率変数列とし

$$\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j, \quad \mu = E[X_1] \quad (-\infty < \mu < \infty)$$

とする。このとき \bar{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**. $\text{Var}[X_1] = \sigma^2 \quad (0 < \sigma < \infty)$ とする。 n が十分大きいとき $\sqrt{n}(\bar{X}_n - \mu)$ 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 ?? 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 $(\Omega, \mathcal{A}, \text{Pr})$ 上の X, X_1, X_2, \dots を確率変数列とする。さらに、各 $X_n (n = 1, 2, \dots)$ 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

$$\lim_{n \rightarrow \infty} \text{Pr}(|X_n - X| \geq \epsilon) = 0$$

が成り立つときをいう。このことを $X_n \xrightarrow{P} X (n \rightarrow \infty)$ 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

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

が成り立つときをいう。このことを $X_n \xrightarrow{\text{a.s.}} X (n \rightarrow \infty)$ 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

$$\lim_{n \rightarrow \infty} F_n(x) = F(x)$$

が成り立つときをいう。このことを $X_n \rightsquigarrow X (n \rightarrow \infty)$ と書く。また、 $X_n \rightsquigarrow X (n \rightarrow \infty)$ かつ $X \sim N(0, 1)$ のときには $X_n \rightsquigarrow N(0, 1) (n \rightarrow \infty)$ とも書くことがある。

注意 4.2. (1) c を定数とする. $\Pr(X = c) = 1$ かつ $X_n \xrightarrow{P} X$ とき $X_n \xrightarrow{P} c$ と記す.

(2) $\Pr(X = c) = 1$ かつ $X_n \rightsquigarrow X (n \rightarrow \infty)$ とき $X_n \rightsquigarrow c (n \rightarrow \infty)$ と記す. □

定義 4.3. X_1, X_2, \dots を確率変数列とし X を別の確率変数とする. さらに $E[X_n^2] < \infty, E[X^2] < \infty (n = 1, 2, \dots)$ 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

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

が成り立つときをいう. このことを $X_n \xrightarrow{q.m.} X (n \rightarrow \infty)$ と記す

定理 4.4. (1) $X_n \xrightarrow{a.s.} X (n \rightarrow \infty)$ ならば, $X_n \xrightarrow{P} X (n \rightarrow \infty)$ である.

(2) $X_n \xrightarrow{q.m.} X (n \rightarrow \infty)$ ならば, $X_n \xrightarrow{P} X (n \rightarrow \infty)$ である.

Proof. (1) 集合 $A := \{\omega \in \Omega; \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}$ は以下のように表現できる. $m, j \in \mathbb{N}$ に対して

$$A_{m,j} = \left\{ \omega \in \Omega; |X_m(\omega) - X(\omega)| \leq \frac{1}{j} \right\} \tag{4.1}$$

とおくと

$$A = \bigcap_{j=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_{m,j}$$

と表現できる. $X_n \xrightarrow{a.s.} X (n \rightarrow \infty)$ のとき, $\Pr(A) = 1$ である. $j_0 \in \mathbb{N}$ に対して

$$\bigcap_{j=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_{m,j} \supset \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_{m,j_0}$$

であるので

$$1 = \Pr(A) \leq \Pr\left(\bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_{m,j_0}\right) \Rightarrow \Pr\left(\bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_{m,j_0}\right) = 1$$

がわかる. $B_{n,j_0} = \bigcap_{m=n}^{\infty} A_{m,j_0}$ とおくと, $B_1 \supset B_2 \supset \dots$ である. よって, 確率測度の連続性から

$$\lim_{n \rightarrow \infty} \Pr(B_{n,j_0}) = \Pr\left(\bigcup_{n=1}^{\infty} B_{n,j_0}\right) = \Pr\left(\bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_{m,j_0}\right) = 1$$

となる. さらに, $B_{n, j_0} \supset A_{n, j_0}$ だから, 任意の $j_0 = 1, 2, \dots$ に対して

$$\lim_{n \rightarrow \infty} \Pr(A_{n, j_0}) = \lim_{n \rightarrow \infty} \Pr\left(|X_n - X| \leq \frac{1}{j_0}\right) = 1$$

であることがわかった. さて, 任意の $\epsilon > 0$ が与えられたとき, j_0 を十分大きくとり, $\frac{1}{j_0} < \epsilon$ とすると

$$\left\{|X_n - X| \leq \frac{1}{j_0}\right\} \supset \{|X_n - X| \leq \epsilon\}$$

となる. したがって

$$\lim_{n \rightarrow \infty} \Pr(|X_n - X| \leq \epsilon) = 1$$

が得られる. これの補集合をとると (1) の結論が証明できた.

(2) Chebyshev の不等式 (系 3.3) から

$$\Pr(|X_n - X| > \epsilon) \leq \frac{1}{\epsilon^2} E[|X_n - X|^2]$$

であることに注意すればよい. □

注意 4.5 (定理 4.4(1) の別証明). 任意の $\epsilon > 0$ を取る. $\mathbb{1}_{(\epsilon, \infty)}(X_n - X) \leq 1$ と $X_n \xrightarrow{\text{a.s.}} X (n \rightarrow \infty)$ であるので, $\lim_{n \rightarrow \infty} \mathbb{1}_{(\epsilon, \infty)}(X_n - X) = 0$ a.s. であることに注意して, Lebesgue の有界収束定理を用いると

$$\begin{aligned} \lim_{n \rightarrow \infty} \Pr(|X_n - X| > \epsilon) &= \lim_{n \rightarrow \infty} E[\mathbb{1}_{(\epsilon, \infty)}(X_n - X)] = E\left[\lim_{n \rightarrow \infty} \mathbb{1}_{(\epsilon, \infty)}(X_n - X)\right] \\ &= 0 \end{aligned}$$

がわかる. □

注意 4.6. 確率収束と概収束の違いを説明する. X, X_1, X_2, \dots を同じ確率空間 $(\Omega, \mathcal{A}, \Pr)$ 上で定義された確率変数列とし

$$M_n(\omega) := \sup_{k \geq n} |X_k(\omega) - X(\omega)| \quad (\omega \in \Omega)$$

とおく. すると

$$X_n \xrightarrow{\text{a.s.}} X \Leftrightarrow M_n \xrightarrow{\text{P}} 0 (n \rightarrow \infty)$$

となることに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

$$M_n(\omega) \xrightarrow{n \rightarrow \infty} 0$$

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

$$\left\{ \omega \in \Omega; \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \right\} = \bigcap_{\ell=1}^{\infty} B_\ell$$

と書けるので

$$X_n \xrightarrow{\text{a.s.}} X \Leftrightarrow \Pr\left(\bigcap_{\ell=1}^{\infty} B_\ell\right) = 1$$

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

$$B_1 \supset B_2 \supset \cdots \supset B_\ell \supset \cdots \quad \left(\Rightarrow \Pr(B_1) \geq \Pr(B_2) \geq \cdots \geq \Pr(B_\ell) \geq \cdots \right)$$

となる. このことと補題 1.7(7) から

$$\Pr\left(\bigcap_{\ell=1}^{\infty} B_\ell\right) = \lim_{\ell \rightarrow \infty} \Pr(B_\ell) \leq \Pr(B_\ell) \leq 1 \quad (\forall \ell \in \mathbb{N})$$

となる. よって

$$1 = \Pr\left(\bigcap_{\ell=1}^{\infty} B_\ell\right) \Leftrightarrow \Pr(B_\ell) = 1 \quad (\forall \ell \in \mathbb{N})$$

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

$$\begin{aligned} \Pr(B_\ell) &= \Pr\left(\bigcup_{n=1}^{\infty} \left\{\omega \in \Omega; M_n(\omega) < \frac{1}{\ell}\right\}\right) \\ &= \lim_{n \rightarrow \infty} \Pr\left(\left\{\omega \in \Omega; M_n(\omega) < \frac{1}{\ell}\right\}\right) \end{aligned}$$

となる. よって

$$\begin{aligned} 1 &= \Pr\left(\bigcap_{\ell=1}^{\infty} B_\ell\right) \Leftrightarrow \Pr(B_\ell) = 1 \ (\forall \ell \in \mathbb{N}) \\ &\Leftrightarrow \lim_{n \rightarrow \infty} \Pr\left(\left\{\omega \in \Omega; M_n(\omega) < \frac{1}{\ell}\right\}\right) = 1 \ (\forall \ell \in \mathbb{N}) \\ &\Leftrightarrow \lim_{n \rightarrow \infty} \Pr\left(\left\{\omega \in \Omega; M_n(\omega) \geq \frac{1}{\ell}\right\}\right) = 0 \ (\forall \ell \in \mathbb{N}) \end{aligned}$$

がわかる. したがって

$$X_n \xrightarrow{\text{a.s.}} X \Leftrightarrow M_n \xrightarrow{\text{P}} 0 \ (n \rightarrow \infty) \quad (4.2)$$

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. \square

例 4.7. X_1, X_2, \dots は同じ確率空間 $(\Omega, \mathcal{A}, \Pr)$ 上で定義された独立な確率変数列で

$$X_n \sim \text{Ber}(p_n) \ (n \in \mathbb{N})$$

とする. ただし, $0 < p_n < 1$ とする. すると

$$X_n \xrightarrow{\text{P}} 0 \ (n \rightarrow \infty) \Leftrightarrow p_n \xrightarrow{n \rightarrow \infty} 0$$

である. 実際, 十分小さな任意の $\epsilon > 0$ に対して, $\Pr(X_n \leq -\epsilon) = 0$ 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

$$\begin{aligned} \Pr(|X_n| \geq \epsilon) &= \Pr(\{X_n \leq -\epsilon\} \cup \{X_n \geq \epsilon\}) = \Pr(X_n \geq \epsilon) = \Pr(X_n = 1) \\ &= p_n \end{aligned}$$

がわかる.

次に

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

とし, $\pi_n = \Pr(M_n = 1)$ とおく. すると $M_n(\omega) \in \{0, 1\}$ ($\omega \in \Omega$) なので

$$\begin{aligned} 1 - \pi_n &= \Pr(M_n = 0) = \Pr\left(\bigcap_{k=n}^{\infty} \{X_k = 0\}\right) \\ &= \prod_{k=n}^{\infty} \Pr(X_k = 0) \quad (\because \{X_n\}_{n=1}^{\infty} \text{ は独立}) \\ &= \prod_{k=n}^{\infty} (1 - p_k) \end{aligned}$$

を得る. すると (4.2) と $X_n \xrightarrow{\text{a.s.}} 0$ ($n \rightarrow \infty$) から $\pi_n \xrightarrow{n \rightarrow \infty} 0$ がわかる. ここで

$$\prod_{k=n}^{\infty} (1 - p_k) \xrightarrow{n \rightarrow \infty} 1 \Leftrightarrow \sum_{n=1}^{\infty} p_n < \infty \quad (n \rightarrow \infty)$$

であること¹に注意すると

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

がわかる.

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

$$X_n \xrightarrow{\text{P}} 0 \quad (n \rightarrow \infty)$$

だが

$$\Pr\left(\left\{\omega \in \Omega; \lim_{n \rightarrow \infty} X_n(\omega) = 0\right\}\right) < 1 \Leftrightarrow X_n \xrightarrow{\text{a.s.}} 0 \quad (n \rightarrow \infty) \text{ は偽}$$

がわかる. □

注意 4.8. $n = 1, 2, \dots$ に対して $X_n \sim N(0, 1/n)$ 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.

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

と定義する. すなわち $\Pr(X = 0) = 1$ をみたす確率変数 X の c.d.f. である. $F(x)$ 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.

¹たとえば, [34, pp.91 – 92] を参照のこと.

そのために $\sqrt{n}X_n \sim N(0, 1) (n = 1, 2, \dots)$ であることに注意する. ただし, Z は標準正規分布 $N(0, 1)$ に従う確率変数とする. $x < 0$ のとき

$$\begin{aligned} F_n(x) &= \Pr(X_n \leq x) = \Pr(\sqrt{n}X_n \leq \sqrt{nx}) = \Pr(Z \leq \sqrt{nx}) \\ &= \int_{-\infty}^{\sqrt{nx}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \rightarrow 0 \quad (n \rightarrow \infty) \end{aligned}$$

となる. 最後の極限は $\sqrt{nx} \xrightarrow{n \rightarrow \infty} -\infty$ より $F(x) = 0 (x < 0)$ がわかる. つぎに $x > 0$ のとき

$$\begin{aligned} F_n(x) &= \Pr(X_n \leq x) = \Pr(\sqrt{n}X_n \leq \sqrt{nx}) = \Pr(Z \leq \sqrt{nx}) \\ &= \int_{-\infty}^{\sqrt{nx}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \rightarrow 1 \quad (n \rightarrow \infty) \end{aligned}$$

となる. 最後の極限は $\sqrt{nx} \xrightarrow{n \rightarrow \infty} \infty$ よりわかる. よって, $F(x) = 1 (x > 0)$ である.

以上の議論から $x \neq 0$ のとき

$$\lim_{n \rightarrow \infty} F_n(x) = F(x)$$

がわかる.

しかし, $x = 0$ は $F(x)$ の不連続点なので, $F_n(0) = 1/2 \neq 1 = F(0)$ 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

$$\Pr(|X_n| \geq \epsilon) = \Pr(X_n^2 \geq \epsilon^2) \leq \frac{E[X_n^2]}{\epsilon^2} = \frac{1}{n\epsilon^2} \xrightarrow{n \rightarrow \infty} 0$$

となる. 最後の等号は, $X_n \sim N(0, 1/n)$ なので, $E[X_n^2] = \frac{1}{n}$ となることを用いた. よって $X_n \xrightarrow{P} 0 (n \rightarrow \infty)$ である. \square

次に収束のタイプ間の関係について述べる.

定理 4.9. 次に関係が成立する.

- (1) $X_n \xrightarrow{q.m.} X \Rightarrow X_n \xrightarrow{P} X (n \rightarrow \infty)$ である.
- (2) $X_n \xrightarrow{P} X \Rightarrow X_n \rightsquigarrow X (n \rightarrow \infty)$ である.
- (3) $X_n \rightsquigarrow X$ かつある定数 c があって $\Pr(X = c) = 1$ のとき $X_n \xrightarrow{P} c (n \rightarrow \infty)$ である.

Proof. 節 4.5.3 で示す. \square

注意 4.10. 定理 4.9(1)(2) の逆は一般に成立しない.

(1) **定理 4.9(1) の逆の反例.** $U \sim \text{Unif}(0, 1)$, $X_n = \sqrt{n}\mathbb{1}_{(0,1/n)}(U)$ ($n = 1, 2, \dots$), $X = 0$ とおく. このとき $\forall \epsilon > 0$ に対して

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

となる. よって $X_n \xrightarrow{P} X$ ($n \rightarrow \infty$) である. しかし

$$E[X_n^2] = n \int_0^{1/n} du = 1$$

なので, $X_n \xrightarrow{q.m.} X$ ($n \rightarrow \infty$) は成立しない.

(2) **定理 4.9(2) の逆の反例.** $X \sim N(0, 1)$ とし, $X_n = -X$ ($n = 1, 2, \dots$) とする. したがって, $X_n \sim N(0, 1)$ である. 明らかに

$$\lim_{n \rightarrow \infty} F_n(x) = F(x)$$

である. よって $X_n \rightsquigarrow X$ ($n \rightarrow \infty$) となる. しかし $\forall \epsilon > 0$ に対して

$$\Pr(|X_n - X| > \epsilon) = \Pr(|2X| > \epsilon) = \Pr\left(|X| > \frac{\epsilon}{2}\right) \neq 0$$

である. よって $X_n \xrightarrow{P} X$ ($n \rightarrow \infty$) は成立しない. □

注意 4.11. 以上の結果をまとめると

$$\begin{aligned} X_n \xrightarrow{a.s.} X (n \rightarrow \infty) &\Rightarrow X_n \xrightarrow{P} X (n \rightarrow \infty) \Rightarrow X_n \rightsquigarrow X (n \rightarrow \infty) \\ &\quad \uparrow \\ &X_n \xrightarrow{q.m.} X (n \rightarrow \infty) \end{aligned}$$

となる.

「 $X_n \rightsquigarrow X (n \rightarrow \infty) \Rightarrow X_n \xrightarrow{q.m.} X (n \rightarrow \infty)$ 」も「 $X_n \xrightarrow{q.m.} X (n \rightarrow \infty) \Rightarrow X_n \rightsquigarrow X (n \rightarrow \infty)$ 」も一般には真ではない反例は演習問題 4.4 と 4.5 を参照のこと.

Proof.

定理 4.12 (Portmanteau の補題). 確率変数列 X_n ($n = 1, 2, \dots$), 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

$$\lim_{n \rightarrow \infty} E[g(X_n)] = E[g(X)]$$

である.

(3) \mathbb{R} 上の任意の有界 Lipschitz 連続関数 g に対して

$$\lim_{n \rightarrow \infty} E[g(X_n)] = E[g(X)]$$

である.

(4) \mathbb{R} 上の任意の非負値連続関数 g に対して

$$\liminf_{n \rightarrow \infty} E[g(X_n)] \geq E[g(X)]$$

である.

(5) \mathbb{R} の任意の開集合 O に対して

$$\liminf_{n \rightarrow \infty} \Pr(X_n \in O) \geq \Pr(X \in O)$$

である.

(6) \mathbb{R} の任意の閉集合 C に対して

$$\limsup_{n \rightarrow \infty} \Pr(X_n \in C) \leq \Pr(X \in C)$$

である.

(7) \mathbb{R} の任意の Borel 集合 B が $\Pr(X \in \partial B) = 0$ ならば

$$\lim_{n \rightarrow \infty} \Pr(X_n \in B) = \Pr(X \in B)$$

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 4.13. 定理 4.12 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 4.14. $X_n, X, Y_n, Y (n = 1, 2, \dots)$ は確率変数列とする. $g: \mathbb{R} \rightarrow \mathbb{R}$ を連続関数とし, c を定数とする.

(1) $X_n \xrightarrow{P} X$ かつ $Y_n \xrightarrow{P} Y \Rightarrow X_n + Y_n \xrightarrow{P} X + Y (n \rightarrow \infty)$ である.

(2) $X_n \xrightarrow{q.m.} X$ かつ $Y_n \xrightarrow{q.m.} Y \Rightarrow X_n + Y_n \xrightarrow{q.m.} X + Y (n \rightarrow \infty)$ である.

- (3) $X_n \rightsquigarrow X$ かつ $Y_n \xrightarrow{P} c \Rightarrow X_n + Y_n \rightsquigarrow X + c (n \rightarrow \infty)$ である.
- (4) $X_n \xrightarrow{P} X$ かつ $Y_n \xrightarrow{P} Y \Rightarrow X_n Y_n \xrightarrow{P} XY (n \rightarrow \infty)$ である.
- (5) $X_n \rightsquigarrow X$ かつ $Y_n \xrightarrow{P} c \Rightarrow X_n Y_n \rightsquigarrow cX (n \rightarrow \infty)$ である.
- (6) $X_n \xrightarrow{P} X \Rightarrow g(X_n) \xrightarrow{P} g(X) (n \rightarrow \infty)$ である.
- (7) $X_n \rightsquigarrow X \Rightarrow g(X_n) \rightsquigarrow g(X) (n \rightarrow \infty)$ である.

Proof. 節 4.5.5 で示す. □

定義 4.15. (1) $\{X_n\}_{n=1}^\infty, \{Y_n\}_{n=1}^\infty$ を確率変数列とし, すべての $n \geq 1$ に対して, $\Pr(Y_n > 0) = 1$ とする. $X_n = o_P(Y_n)$ であるとは

$$\frac{X_n}{Y_n} \xrightarrow{P} 0 \quad (n \rightarrow \infty)$$

が成り立つときをいう.

(2) $X_n = O_P(Y_n)$ であるとは, 任意の $\epsilon > 0$ に対して, ある $K_\epsilon > 0$ が存在して

$$\Pr\left(\left|\frac{X_n}{Y_n}\right| > K_\epsilon\right) < \epsilon$$

が成り立つときをいう.

注意 4.16. 次の関係式が成り立つことが証明できる.

- (1) $o_P(1) + O_P(1) = O_P(1)$,
- (2) $O_P(1) + O_P(1) = O_P(1)$,
- (3) $o_P(1) + o_P(1) = o_P(1)$,
- (4) $O_P(1) \cdot O_P(1) = O_P(1)$,
- (5) $o_P(1) \cdot O_P(1) = o_P(1)$.

問 4.1. 注意 4.16(1) – (5) を証明せよ.

4.2 大数の法則

定理 4.17. (大数の弱法則) X_1, X_2, \dots は確率空間 $(\Omega, \mathcal{A}, \Pr)$ 上の i.i.d. 確率変数列とする. $E[|X_1|] < \infty$ のとき

$$\bar{X}_n := \frac{1}{n}(X_1 + X_2 + \dots + X_n) \xrightarrow{P} \mu, \quad (n \rightarrow \infty)$$

が成立する. ただし, $\mu = E[X_1]$ である.

Proof. より強い条件 $E[X_1^2] < \infty$ 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. $\text{Var}[X_1] = \sigma^2$ とおく. Chebyshev の不等式 (系 3.3) より

$$\Pr(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\text{Var}[\bar{X}_n]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \xrightarrow{n \rightarrow \infty} 0$$

よりわかる.

仮定のもとの証明は, 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 4.18. X_1, X_2, \dots は非負値確率変数列で, 条件

$$\sum_{n=1}^{\infty} E[X_n] < \infty \tag{4.3}$$

をみたすとする. このとき

$$\Pr\left(\omega \in \Omega; \sum_{n=1}^{\infty} X_n(\omega) < \infty\right) = 1$$

が成り立つ.

Proof. 証明は節 4.5.2 で行う. □

定理 4.19. (大数の強法則) X_1, X_2, \dots は確率空間 $(\Omega, \mathcal{A}, \Pr)$ 上の i.i.d. 確率変数列とし, $E[|X_1|^4] < \infty$ とする. このとき

$$\bar{X}_n := \frac{1}{n}(X_1 + X_2 + \dots + X_n) \xrightarrow{\text{a.s.}} \mu, \quad (n \rightarrow \infty) \tag{4.4}$$

が成立する. ただし, $\mu = E[X_1]$ である.

Proof. まず

$$E[|X_1|^4] =: K < \infty, \quad T_n = X_1 + X_2 + \dots + X_n, \quad \bar{X}_n = \frac{T_n}{n}$$

とおく.

最初に, $\mu = 0$ として (4.4) を示す. 以下の事象の包含関係に注意する. すなわち

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

が成立する. このことより

$$\Pr\left(\sum_{n=1}^{\infty} \bar{X}_n^4 < \infty\right) = 1 \tag{4.5}$$

がわかると

$$1 = \Pr\left(\sum_{n=1}^{\infty} \bar{X}_n^4 < \infty\right) \leq \Pr\left(\lim_{n \rightarrow \infty} \bar{X}_n = 0\right)$$

となり, $\mu = 0$ のときに (4.4) がわかる.

以下では $\mu = 0$ として (4.5) 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

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

となる. ただし $\ell_1, \ell_2, \dots, \ell_n$ は 0 以上の整数である. ここで $X_1^{\ell_1}, X_2^{\ell_2}, \dots, X_n^{\ell_n}$ は独立あることに注意すると

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

がわかる. さらに, $E[X_k] = \mu = 0 (k = 1, 2, \dots, n)$ と $\frac{4!}{2!2!} = 6$ を用いると

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

を得る. ここで $E[X_k^4] = K (k = 1, 2, \dots, n)$ と Cauchy-Schwarz の不等式から

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

となることに注意すると

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

を得る². よって

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

²上式の最初の等号は $\sum_{1 \leq k < \ell \leq n} (\cdot)$ で足し合わせる項数が $\frac{n(n-1)}{2}$ 項あることからわかる.

となる. したがって 補題 4.18 から (4.5) が成立することがわかる.

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

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

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

$$\Pr\left(\lim_{n \rightarrow \infty} \frac{1}{n}(Y_1 + Y_2 + \dots + Y_n) = 0\right) = 1$$

となる. 最後に, $\forall \omega \in \Omega$ に対して

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

に注意すればよい. □

注意 4.20. 大数の法則は $E[|X_1|] < \infty$ で成立する. この条件下での大数の強法則の証明は節 A.1.2 に記した. □

系 4.21 (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³, $n \in \mathbb{N}$ に対して

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

とおいたとき

$$f(x) = \lim_{n \rightarrow \infty} B_n(x) \tag{4.6}$$

が成立する.

Proof. $x = 0, 1$ のとき (A.8) は明らかなので, $0 < x < 1$ に対して, (A.8) が成立すること示す. X_j ($j = 1, 2, \dots$) は i.i.d. 確率変数列で

$$\Pr(X_j = 1) = x = 1 - \Pr(X_j = 0)$$

をみたすとする. いま

$$E\left[\frac{X_1 + X_2 + \dots + X_n}{n}\right] = x$$

³ $f(0) = a, f(1) = b$ に対して, $\tilde{f}(x) = f(x) + \{f(b) - f(a)\}x + f(a)$ とおくと \tilde{f} は $[0, 1]$ 上の連続関数で $\tilde{f}(0) = \tilde{f}(1) = 0$ となるので, 左記の設定は一般性を失わない仮定である.

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

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

となる. したがって

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \xrightarrow{P} x \quad (n \rightarrow \infty)$$

がわかる. さらに, 定理 4.9(2) より

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \rightsquigarrow x \quad (n \rightarrow \infty)$$

を得る. f は有界連続関数なので, 定理 4.12(2) から

$$\lim_{n \rightarrow \infty} \mathbb{E}\left[f\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right)\right] = f(x)$$

となる. しかし, $S := X_1 + X_2 + \cdots + X_n \sim \text{Bino}(n, x)$ なので

$$\begin{aligned} \mathbb{E}\left[f\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right)\right] &= \mathbb{E}\left[f\left(\frac{S}{n}\right)\right] \\ &= \sum_{j=0}^n f\left(\frac{j}{n}\right) \binom{n}{j} x^j (1-x)^{n-j} \\ &= B_n(x) \end{aligned}$$

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 4.22. X_1, X_2, \dots を i.i.d. 確率変数列とし, $\mathbb{E}[X_1] = \mu$, $\text{Var}[X_1] = \sigma^2$ ($0 < \sigma < \infty$) とする. このとき

$$Z_n := \frac{\bar{X}_n - \mu}{\sqrt{\text{Var}[\bar{X}_n]}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \rightsquigarrow Z \quad (n \rightarrow \infty)$$

が成り立つ. ただし $\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$, $Z \sim \mathbf{N}(0, 1)$ である.

Proof. 中心極限定理の証明は第 A.2 章です. □

例 4.23. X_1, X_2, \dots を i.i.d. 確率変数列とし, $E[X_1] = \mu$, $\text{Var}[X_1] = \sigma^2$ ($0 < \sigma < \infty$) とする. このとき中心極限定理 (定理 4.22) より

$$Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \rightsquigarrow \mathbf{N}(0, 1) \quad (n \rightarrow \infty) \quad (4.7)$$

が成立する. ここで

$$S_n^2 := \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X}_n)^2, \quad S_n = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X}_n)^2}$$

とおく. すると

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

と書き直せる. 設定から $E[(X_1 - \mu)^2] = \sigma^2 < \infty$ なので, 大数の法則 (定理 4.17) より

$$\frac{1}{n} \sum_{j=1}^n (X_j - \mu)^2 \xrightarrow{\text{P}} E[(X_1 - \mu)^2] = \sigma^2, \quad (4.8)$$

$$\bar{X}_n \xrightarrow{\text{P}} \mu \quad (n \rightarrow \infty) \quad (4.9)$$

である. (4.9) に対して, 定理 4.14(6) ($g(x) = (x - \mu)^2$) を用いると

$$(\bar{X}_n - \mu)^2 \xrightarrow{\text{P}} 0 \quad (n \rightarrow \infty) \quad (4.10)$$

となる. さらに, (4.8), (4.10) と定理 4.14(3)(5) より

$$S_n^2 \xrightarrow{\text{P}} \sigma^2 \quad (n \rightarrow \infty)$$

である. 再度, 定理 4.14(6) を用いると

$$\frac{S_n}{\sigma} \xrightarrow{\text{P}} 1 \quad (n \rightarrow \infty) \quad (4.11)$$

がわかる. 最後に (4.7), (4.11) と定理 4.14(5) より

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \rightsquigarrow \mathbf{N}(0, 1) \quad (n \rightarrow \infty)$$

を得る.

□

定理 4.24. $\mathbf{X}_1, \mathbf{X}_2, \dots$ を i.i.d. 確率ベクトル列とする. ただし $j = 1, 2, \dots$ に対して

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

とし,

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

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

$$\sqrt{n}\mathbf{c}^\top(\bar{\mathbf{X}}_n - \boldsymbol{\mu}) \rightsquigarrow N(0, \mathbf{c}^\top \boldsymbol{\Sigma} \mathbf{c}) \quad (n \rightarrow \infty)$$

が成立する. このことを

$$\sqrt{n}(\bar{\mathbf{X}}_n - \boldsymbol{\mu}) \rightsquigarrow N_d(\mathbf{0}, \boldsymbol{\Sigma}) \quad (n \rightarrow \infty)$$

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

定理 4.25. (デルタ法) Y_1, Y_2, \dots を確率変数列とし

$$\frac{\sqrt{n}(Y_n - \mu)}{\sigma} \rightsquigarrow N(0, 1) \quad (n \rightarrow \infty)$$

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

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

が成立する. すなわち

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

である. ただし「 \approx 」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 4.26. X_1, X_2, \dots を i.i.d. 確率変数列とし, $E[X_1] = \mu, \text{Var}[X_1] = \sigma^2 (0 < \sigma < \infty)$ とする. このとき中心極限定理 (定理 4.22) より

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \rightsquigarrow \mathbf{N}(0, 1) \quad (n \rightarrow \infty)$$

となる. いま

$$W_n = e^{\bar{X}_n}$$

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 4.25) より

$$\frac{\sqrt{n}(W_n - e^\mu)}{e^\mu \sigma} \rightsquigarrow \mathbf{N}(0, 1) \quad (n \rightarrow \infty)$$

となる. よって

$$W_n \approx \mathbf{N}\left(e^\mu, \frac{e^{2\mu}\sigma^2}{n}\right)$$

がわかる. □

定理 4.27. (多次元デルタ法) $\mathbf{Y}_n = (Y_{1n}, Y_{2n}, \dots, Y_{dn})^\top (n = 1, 2, \dots)$ を確率ベクトル列とし

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

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 $g: \mathbb{R}^d \rightarrow \mathbb{R}$ は滑らかで

$$\nabla g(\mathbf{y}) = \begin{pmatrix} \frac{\partial g}{\partial y_1} \\ \frac{\partial g}{\partial y_2} \\ \vdots \\ \frac{\partial g}{\partial y_d} \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_d \end{pmatrix}, \quad \nabla_{\boldsymbol{\mu}} := \nabla g(\mathbf{y}) \Big|_{\mathbf{y}=\boldsymbol{\mu}}$$

とする. このとき

$$\sqrt{n}(g(\mathbf{Y}_n) - g(\boldsymbol{\mu})) \rightsquigarrow \mathbf{N}(0, \nabla_{\boldsymbol{\mu}}^\top \boldsymbol{\Sigma} \nabla_{\boldsymbol{\mu}}) \quad (n \rightarrow \infty)$$

となる.

例 4.28.

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

は i.i.d. 確率ベクトル列で

$$\begin{aligned} \mathbb{E} \begin{bmatrix} X_{11} \\ X_{21} \end{bmatrix} &= \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \boldsymbol{\mu}, \\ \text{Var} \begin{bmatrix} X_{11} \\ X_{21} \end{bmatrix} &= \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} = \boldsymbol{\Sigma} \end{aligned}$$

とする. ただし $\boldsymbol{\Sigma}$ は 2×2 の正値対称行列である.

$$\bar{X}_1 = \frac{1}{n} \sum_{j=1}^n X_{1j}, \quad \bar{X}_2 = \frac{1}{n} \sum_{j=1}^n X_{2j}$$

とし

$$Y_n = \bar{X}_1 \bar{X}_2, \quad g(s_1, s_2) = s_1 s_2 \quad (s_1, s_2 \in \mathbb{R})$$

とおく. 中心極限定理 (定理 4.24) より

$$\sqrt{n} \begin{pmatrix} \bar{X}_1 - \mu_1 \\ \bar{X}_2 - \mu_2 \end{pmatrix} \rightsquigarrow \mathbf{N}_2(\mathbf{0}_2, \boldsymbol{\Sigma}) \quad (n \rightarrow \infty)$$

となる. いま

$$\nabla g(\mathbf{s}) = \begin{pmatrix} \frac{\partial g}{\partial s_1} \\ \frac{\partial g}{\partial s_2} \end{pmatrix} = \begin{pmatrix} s_2 \\ s_1 \end{pmatrix}, \quad \mathbf{s} = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}$$

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

$$\sqrt{n}(\bar{X}_1 \bar{X}_2 - \mu_1 \mu_2) \rightsquigarrow \mathbf{N}(0, \mu_2^2 \sigma_{11} + 2\mu_1 \mu_2 \sigma_{12} + \mu_1^2 \sigma_{22}) \quad (n \rightarrow \infty)$$

がわかる. □

4.5 定理 4.17, 補題 4.18, 定理 4.9, 4.12, 4.14, 4.25 の証明

4.5.1 定理 4.17 の証明

任意の $\delta > 0$ と n に対して

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

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

と定める. 明らかに

$$X_j = Y_j + Z_j, \quad (j = 1, 2, \dots, n)$$

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

$$\begin{aligned} \text{Var}[Y_j] &= \text{Var}[Y_1] = \text{E}[Y_1^2] - (\text{E}[Y_1])^2 \leq \text{E}[Y_1^2] \\ &= \text{E}[X_1^2 \mathbf{1}_{[-\delta n, \delta n]}(X_1)] \\ &= \int_{-\infty}^{\infty} x^2 \mathbf{1}_{[-\delta n, \delta n]}(x) \mathbf{p}(x) \, dx = \int_{-\delta n}^{\delta n} x^2 \mathbf{p}(x) \, dx \\ &\leq \delta n \int_{-\delta n}^{\delta n} |x| \mathbf{p}(x) \, dx \leq \delta n \int_{-\infty}^{\infty} |x| \mathbf{p}(x) \, dx \\ &= \delta n \text{E}[|X_1|] \end{aligned}$$

となる. すなわち

$$\text{Var}[Y_j] \leq \delta n \text{E}[|X_1|] \quad (j = 1, 2, \dots, n) \quad (4.12)$$

である. 次に

$$\text{E}[Y_j] = \text{E}[Y_1] = \text{E}[X_1 \mathbf{1}_{[-\delta n, \delta n]}(X_1)] = \int_{-\infty}^{\infty} x \mathbf{1}_{[-\delta n, \delta n]}(x) \mathbf{p}(x) \, dx$$

と書ける. ここで

$$\left| x \mathbf{1}_{[-\delta n, \delta n]}(x) \mathbf{p}(x) \right| \leq |x| \mathbf{p}(x), \quad x \mathbf{1}_{[-\delta n, \delta n]}(x) \mathbf{p}(x) \xrightarrow{n \rightarrow \infty} x \mathbf{p}(x)$$

かつ

$$\int_{-\infty}^{\infty} |x| \mathbf{p}(x) \, dx < \infty$$

であることに注意して, 優収束定理を用いると

$$E[Y_1] = \int_{-\infty}^{\infty} x \mathbb{1}_{[-\delta n, \delta n]}(x) p(x) dx \xrightarrow{n \rightarrow \infty} \int_{-\infty}^{\infty} x p(x) dx$$

となることがわかる. したがって

$$E[Y_1] \xrightarrow{n \rightarrow \infty} \mu \quad (4.13)$$

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

$$\begin{aligned} \Pr\left(\left|\frac{1}{n} \sum_{j=1}^n Y_j - E[Y_1]\right| \geq \epsilon\right) &= \Pr\left(\left|\sum_{j=1}^n Y_j - E\left[\sum_{j=1}^n Y_j\right]\right| \geq n\epsilon\right) \\ &\leq \frac{1}{n^2 \epsilon^2} \text{Var}\left[\sum_{j=1}^n Y_j\right] \quad (\because \text{系 3.3}) \\ &= \frac{n}{n^2 \epsilon^2} \text{Var}[Y_1] \leq \frac{n \delta n E[|X_1|]}{n^2 \epsilon^2} \quad (\because (4.12)) \\ &= \frac{\delta}{\epsilon^2} E[|X_1|] \end{aligned}$$

となる. したがって

$$\Pr\left(\left|\frac{1}{n} \sum_{j=1}^n Y_j - E[Y_1]\right| \geq \epsilon\right) \leq \frac{\delta}{\epsilon^2} E[|X_1|] \quad (4.14)$$

となる. よって, (4.13) と (4.14) から, 十分大きな n に対して

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

となる⁴. したがって

$$\Pr\left(\left|\frac{1}{n} \sum_{j=1}^n Y_j - \mu\right| \geq 2\epsilon\right) \leq \frac{\delta}{\epsilon^2} E[|X_1|] \quad (4.15)$$

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

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

$$\int_{|x|>\delta n} |x|p(x) dx < \delta^2$$

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

$$\begin{aligned} \Pr(Z_j \neq 0) &= \Pr(|Z_j| > \delta n) = \Pr(|X_j| > \delta n) = \int_{-\infty}^{-\delta n} p(x) dx + \int_{\delta n}^{\infty} p(x) dx \\ &= \int_{|x|>\delta n} p(x) dx < \int_{|x|>\delta n} \frac{|x|}{\delta n} p(x) dx = \frac{1}{\delta n} \int_{|x|>\delta n} |x|p(x) dx \\ &< \frac{\delta^2}{\delta n} = \frac{\delta}{n} \end{aligned} \quad (4.16)$$

となる. したがって, (4.16) から

$$\Pr\left(\sum_{j=1}^n Z_j \neq 0\right) = \Pr\left(\bigcup_{j=1}^n \{Z_j \neq 0\}\right) = n\Pr(Z_1 \neq 0) \leq \delta \quad (4.17)$$

を得る. よって, (4.15) と (4.17) から, 十分大きな n に対して

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

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

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

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 4.18 の証明

背理法で証明する. そのために

$$\Pr\left(\sum_{k=1}^{\infty} X_k < \infty\right) < 1$$

を仮定する. さらに事象 F と N を

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

で定める. このとき

$$\Omega = F \cup N \quad \text{かつ} \quad F \cap N = \emptyset$$

が成り立つ. したがって

$$1 = \Pr(\Omega) = \Pr(F \cup N) = \Pr(F) + \Pr(N) \quad (\because \Pr \text{ の加法性})$$

がわかる. この関係式と背理法の仮定 $\Pr(F) < 1$ から $\Pr(N) > 0$ となる. 一方, $\omega \in N$ のとき

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

であり, $\omega \in N^c$ のとき

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

である. したがって, 任意の正の実数 $r \geq 0$ に対して

$$r \mathbf{1}_N(\omega) \leq \sum_{k=1}^{\infty} X_k(\omega) \quad (\forall \omega \in \Omega) \quad (4.18)$$

が成り立つ. (4.18) の両辺の期待値をとる. すると 定義 2.10(4), (4.24) と (4.18) から

$$r \times \Pr(N) = E[r \mathbf{1}_N] \leq E\left[\sum_{k=1}^{\infty} X_k\right] = \sum_{k=1}^{\infty} E[X_k] < \infty \quad (4.19)$$

を得る. (4.19) の最後の等号は単調収束定理からわかる. ここで $\Pr(N) > 0$ であり, $r \geq 0$ は任意の実数だったので, $r \rightarrow \infty$ とすれば, (4.19) の最左辺は $+\infty$ となるので, 矛盾が生じる. したがって, $\Pr(F) = 1$ が成り立つ. \square

4.5.3 定理 4.9 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

$$\Pr(|X_n - X| \geq \epsilon) = \Pr(|X_n - X|^2 \geq \epsilon^2) \leq \frac{\mathbf{E}[|X_n - X|^2]}{\epsilon^2} \xrightarrow{n \rightarrow \infty} 0$$

がわかる.

(2) $X_n \xrightarrow{P} X$ とする. $\epsilon > 0$ を固定し, x と $x \pm \epsilon$ を X の c.d.f. $F(x)$ の連続点⁵ とする. このとき

$$\begin{aligned} F_n(x) &= \Pr(X_n \leq x) \\ &= \Pr(X_n \leq x, X < x + \epsilon) + \Pr(X_n \leq x, X \geq x + \epsilon) \\ &\leq \Pr(X \leq x + \epsilon) + \Pr(|X_n - X| \geq \epsilon) \end{aligned} \tag{4.20}$$

$$= F(x + \epsilon) + \Pr(|X_n - X| \geq \epsilon) \tag{4.21}$$

となる. 同様に

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

$$= F_n(x) + \Pr(|X_n - X| \geq \epsilon) \tag{4.23}$$

となる. (4.21) と (4.23) を合わせると

$$F(x - \epsilon) - \Pr(|X_n - X| \geq \epsilon) \leq F_n(x) \leq F(x + \epsilon) + \Pr(|X_n - X| \geq \epsilon)$$

がわかる. 上の式の辺々で $n \rightarrow \infty$ とすれば

$$F(x - \epsilon) \leq \liminf_{n \rightarrow \infty} F_n(x) \leq \limsup_{n \rightarrow \infty} F_n(x) \leq F(x + \epsilon)$$

となる. ここで $\epsilon \rightarrow 0$ とし x を $F(x)$ の連続点とすれば

$$\lim_{n \rightarrow \infty} F_n(x) = F(x)$$

がわかる.

⁵ X の分布関数の不連続点は高々可算個なので, 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 $\Pr(X = c) = 1$ とする. $c \pm \epsilon$ を $F(x)$ の連続点になるようにして $\epsilon > 0$ を固定する⁶. このとき

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

となる. □

問 4.2. (4.20) と (4.22) を確認せよ. すなわち

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

を示せ.

4.5.4 定理 4.12 の証明

(1) \Rightarrow (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

$$\Pr(X \in K^c) \leq \epsilon \tag{4.24}$$

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

$$\max_{j \in \{1, 2, \dots, m\}} \max_{x, y \in K_j} |g(x) - g(y)| \leq \epsilon \tag{4.25}$$

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⁷に取る. このとき, (1) より

$$\Pr(X_n \in K) \xrightarrow{n \rightarrow \infty} \Pr(X \in K) \tag{4.26}$$

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

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

⁶ $F(x)$ は有界非減少関数なので, $F(x)$ の不連続点は高々可算個である. よってこのように ϵ を取ることもできる.

⁷ 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

$$\max_{x \in K} |g(x) - g_\epsilon(x)| \leq \epsilon \quad (4.27)$$

となる. したがって, (4.24) と (4.27) に注意すると

$$\begin{aligned} & |E[g(X)] - E[g_\epsilon(X)]| \\ &= |E[\mathbf{1}_K(X)g(X)] + E[\mathbf{1}_{K^c}(X)g(X)] - E[\mathbf{1}_K(X)g_\epsilon(X)] \\ &\quad - E[\mathbf{1}_{K^c}(X)g_\epsilon(X)]| \\ &\leq |E[\mathbf{1}_K(X)g(X)] - E[\mathbf{1}_K(X)g_\epsilon(X)]| + |E[\mathbf{1}_{K^c}(X)g(X)] \\ &\quad - E[\mathbf{1}_{K^c}(X)g_\epsilon(X)]| \\ &\leq E[\mathbf{1}_K(X)|g(X) - g_\epsilon(X)|] + E[\mathbf{1}_{K^c}(X)|g(X) - g_\epsilon(X)|] \\ &\leq \underbrace{\max_{x \in K} |g(x) - g_\epsilon(x)| E[\mathbf{1}_K(X)]}_{\leq \epsilon \quad \because (4.27)} + E[\mathbf{1}_{K^c}(X)|g(X) - g_\epsilon(X)|] \\ &\leq \epsilon \Pr(X \in K) + \Pr(X \in K^c) \\ &\leq 2\epsilon \end{aligned} \quad (4.28)$$

となる. (4.26) から, n を十分大きく取ると

$$\left| \underbrace{\Pr(X_n \in K^c)}_{=1-\Pr(X_n \in K)} - \underbrace{\Pr(X \in K^c)}_{=1-\Pr(X \in K)} \right| = |\Pr(X \in K) - \Pr(X_n \in K)| \leq \epsilon$$

とできるので

$$\Pr(X_n \in K^c) \leq \Pr(X \in K^c) + |\Pr(X_n \in K^c) - \Pr(X \in K^c)| \leq 2\epsilon$$

がわかる. 同様に

$$|E[g(X_n)] - E[g_\epsilon(X_n)]| \leq \epsilon \Pr(X_n \in K) + \Pr(X_n \in K^c) \leq 3\epsilon \quad (4.29)$$

を得る. g_ϵ の作りかたから n を十分大きくとると

$$\begin{aligned} & |E[g_\epsilon(X)] - E[g_\epsilon(X_n)]| \\ &\leq \sum_{j=1}^n |\Pr(X \in K_j) - \Pr(X_n \in K_j)| \max_{j=\{1,2,\dots,m\}} |g(x_j)| \\ &\leq \frac{\epsilon}{2} < \epsilon \end{aligned} \quad (4.30)$$

となる. よって, (4.28) - (4.30) を合わせると

$$\begin{aligned} |E[g(X_n)] - E[g(X)]| &\leq \underbrace{|E[g(X_n)] - E[g_\epsilon(X_n)]|}_{\leq 3\epsilon} + \underbrace{|E[g_\epsilon(X_n)] - E[g_\epsilon(X)]|}_{\leq \epsilon} \\ &\quad + \underbrace{|E[g_\epsilon(X)] - E[g(X)]|}_{\leq 2\epsilon} \\ &\leq 6\epsilon \end{aligned}$$

がわかるので, 主張は証明された.

(2) \Rightarrow (3) の証明: 明らか.

(2) \Rightarrow (4) の証明: 非負値連続関数 g と $M > 0$ に対して, $g_M(x) = \min\{g(x), M\} \geq 0 (x \in \mathbb{R})$ と定める g_M は有界連続となる. よって

$$\liminf_{n \rightarrow \infty} \mathbf{E}[g(X_n)] \geq \liminf_{n \rightarrow \infty} \mathbf{E}[g_M(X_n)] = \mathbf{E}[g_M(X)] \quad (\because (2))$$

となる. 両辺で $M \rightarrow \infty$ 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 $g_M(x) \uparrow \mathbb{1}_O(x) (M \rightarrow \infty)$ 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

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

となる. この両辺で $M \rightarrow \infty$ 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, $\text{cl}(B)$ を B の閉包とすると $\partial B = \text{cl}(B) \setminus B^\circ$ である. このことに注意すると

$$\begin{aligned} \Pr(X \in B^\circ) &\leq \liminf_{n \rightarrow \infty} \Pr(X_n \in B^\circ) \leq \limsup_{n \rightarrow \infty} \Pr(X_n \in \text{cl}(B)) \\ &\leq \Pr(X_n \in \text{cl}(B)) \end{aligned}$$

となる. $\Pr(X \in \partial B) = 0$ 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 4.14 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

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

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

$$\Pr(|X - X_n| \geq \delta) \geq \Pr(|g(X) - g(X_n)| \geq \epsilon)$$

がわかる. よって, $g(X_n) \xrightarrow{P} g(X) (n \rightarrow \infty)$ 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. $X_n \rightsquigarrow X (n \rightarrow \infty)$ と定理 4.12 から

$$\lim_{n \rightarrow \infty} E[f(g(X_n))] = \lim_{n \rightarrow \infty} E[(f \circ g)(X_n)] = E[(f \circ g)(X)] = E[f(g(X))]$$

から (7) は示される.

(1) の証明: $X_n \xrightarrow{P} X (n \rightarrow \infty)$ かつ $Y_n \xrightarrow{P} Y (n \rightarrow \infty)$ から $(X_n, Y_n) \xrightarrow{P} (X, Y) (n \rightarrow \infty)$ 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 $\{(x + y) - (a + b)\}^2 \leq 2\{(x - a)^2 + (y - b)^2\}$ を使えばよい.

(3) の証明: $Y_n \xrightarrow{P} c (n \rightarrow \infty)$ なので, 注意 4.2(2) から $Y_n \rightsquigarrow c (n \rightarrow \infty)$ となる. このことから $(X_n, Y_n) \rightsquigarrow (X, Y) (n \rightarrow \infty)$ となることが証明⁸ 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 4.25 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

$$g(x) - g(\mu) = (x - \mu) \int_0^1 \dot{g}(\mu + t(x - \mu)) dt \quad (4.31)$$

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

$$|x - \mu| < \delta_2 \Rightarrow |\dot{g}(x) - \dot{g}(\mu)| < \epsilon$$

となる. よって $|x - \mu| < \min(\delta_1, \delta_2) =: \delta$ なる任意の x に対して

$$\begin{aligned} \left| \int_0^1 \dot{g}(\mu + t(x - \mu)) dt - \dot{g}(\mu) \right| &\leq \int_0^1 |\dot{g}(\mu + t(x - \mu)) - \dot{g}(\mu)| dt \\ &< \epsilon \int_0^1 dt = \epsilon \end{aligned} \quad (4.32)$$

⁸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 [39, p.135] を参照のこと.

となる. (4.32) より

$$\Pr\left(\left|\int_0^1 \dot{g}(\mu + t(Y_n - \mu)) dt - \dot{g}(\mu)\right| \geq \epsilon\right) \leq \Pr(|Y_n - \mu| \geq \delta) \quad (4.33)$$

が成立する. $\sqrt{n}(Y_n - \mu)/\sigma \rightsquigarrow N(0, 1)$ なので $Y_n \xrightarrow{P} \mu$ である. (4.33) から

$$\int_0^1 \dot{g}(\mu + t(Y_n - \mu)) dt \xrightarrow{P} \dot{g}(\mu) \quad (4.34)$$

が成立する. $|Y_n - \mu| < \delta$ が起こったとき, (4.31) に $x = Y_n$ を代入すれば

$$\frac{\sqrt{n}(g(Y_n) - g(\mu))}{\sigma} = \frac{\sqrt{n}(Y_n - \mu)}{\sigma} \int_0^1 \dot{g}(\mu + t(Y_n - \mu)) dt \quad (4.35)$$

を得る. さらに, $\sqrt{n}(Y_n - \mu)/\sigma \rightsquigarrow N(0, 1)$ と (4.34) に注意して定理 4.14(5) を (4.35) に適用すれば

$$\begin{aligned} & \frac{\sqrt{n}(g(Y_n) - g(\mu))}{\sigma} \\ &= \frac{\sqrt{n}(g(Y_n) - g(\mu))}{\sigma} \mathbb{1}\{|Y_n - \mu| < \delta\} + \frac{\sqrt{n}(g(Y_n) - g(\mu))}{\sigma} \mathbb{1}\{|Y_n - \mu| \geq \delta\} \\ &= \frac{\sqrt{n}(Y_n - \mu)}{\sigma} \underbrace{\int_0^1 \dot{g}(\mu + t(Y_n - \mu)) dt \mathbb{1}\{|Y_n - \mu| < \delta\}}_{\dot{g}(\mu) + o_P(1)} \\ & \quad + \underbrace{\frac{\sqrt{n}(g(Y_n) - g(\mu))}{\sigma} \mathbb{1}\{|Y_n - \mu| \geq \delta\}}_{=o_P(1)} \\ & \rightsquigarrow \dot{g}(\mu) N(0, 1) \end{aligned}$$

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 [17, 39] を参考にした. 注意 4.6 は [17] からの借用である. 第 4.2 節は [35] を借用した. 第 4.4 節は [17, 39] を借用した. 定理 4.17 の証明は [27] からの借用である. 定理 4.25 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 $\left\{ \frac{1}{n}, \frac{2}{n}, \dots, 1 \right\}$ 上の一様分布に従うとする. すなわち

$$\Pr\left(X_n = \frac{k}{n}\right) = \begin{cases} \frac{1}{n} & (k = 1, 2, \dots, n) \\ 0 & (\text{その他の場合}) \end{cases}$$

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.

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

をもつ. このとき, 以下の問いに答えよ.

(1) $\frac{k}{n} \leq x < \frac{(k+1)}{n}$ ($k = 1, \dots, n$) に対して

$$\Pr(X_n \leq x)$$

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

(2) $x \in \mathbb{R}$ に対して

$$\Pr(X \leq x)$$

を求めよ.

(3)

$$\lim_{n \rightarrow \infty} |\Pr(X_n \leq x) - \Pr(X \leq x)|$$

を評価することで $X_n \rightsquigarrow X$ ($n \rightarrow \infty$) を示せ.

演習問題 4.2. $\{X_n\}_{n=1}^{\infty}$ を確率変数列とし, $E[X_n^2] < \infty$ ($n \in \mathbb{N}$) とする.

(1) $\mu \in \mathbb{R}$ を定数とする.

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

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) $\lim_{n \rightarrow \infty} E[(X_n - \mu)^2] = 0.$

(b) $\lim_{n \rightarrow \infty} E[X_n] = \mu$ かつ $\lim_{n \rightarrow \infty} \text{Var}[X_n] = 0$

演習問題 4.3. $\{X_n\}_{n=1}^\infty, \{Y_n\}_{n=1}^\infty, X, Y$ を確率空間 $(\Omega, \mathcal{A}, \Pr)$ 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

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

が成り立つこととする. これを $(X_n, Y_n) \xrightarrow{P} (X, Y) (n \rightarrow \infty)$ と記すことにする. このとき, 以下の問いを答えよ.

- (1) $X_n \xrightarrow{P} X (n \rightarrow \infty)$ かつ $Y_n \xrightarrow{P} Y (n \rightarrow \infty)$ ならば, $(X_n, Y_n) \xrightarrow{P} (X, Y) (n \rightarrow \infty)$ が成り立つことを示せ.
 (2) $(X_n, Y_n) \xrightarrow{P} (X, Y) (n \rightarrow \infty)$ ならば, $X_n \xrightarrow{P} X (n \rightarrow \infty)$ かつ $Y_n \xrightarrow{P} Y (n \rightarrow \infty)$ が成り立つことを示せ.

演習問題 4.4. X_1, X_2, \dots は互いに独立な確率変数列で

$$\Pr(X_n = 1) = \frac{1}{n}, \quad \Pr(X_n = 0) = 1 - \frac{1}{n} \quad (n = 1, 2, \dots)$$

とする. このとき, $X_n \xrightarrow{q.m.} 0 (n \rightarrow \infty)$ だが, $X_n \not\xrightarrow{a.s.} 0 (n \rightarrow \infty)$ であることを示せ.

ヒント Borel-Cantelli の第 2 補題を用いるとよい.

演習問題 4.5. $\Omega = (0, 1], \mathcal{A} = \mathcal{B}((0, 1])$ とし

$$\Pr((a, b]) = b - a \quad (a, b \in \Omega)$$

とする.

$$X_n(\omega) = \sqrt{n} \mathbf{1}_{(0, 1/n]}(\omega) \quad (n = 1, 2, \dots)$$

と定義したとき, $X_n(\omega) \xrightarrow{n \rightarrow \infty} 0$ だが, $\lim_{n \rightarrow \infty} E[X_n^2] \neq 0$ を示せ.

第II部

宴編: 統計的推測

第5章 統計的推測論の枠組み

この章では統計的推測の枠組みについて述べる。第 5.1 節では、統計的推測理論で重要な役割を果たす統計的モデルを説明する。5.2 節では、統計的推測理論を統一的な枠組みで議論する統計的決定理論の簡単な説明をする。

5.1 統計的実験と母数モデル

\mathbb{X} は距離空間とし、 $\mathcal{B}(\mathbb{X})$ は \mathbb{X} 上の Borel 集合族とする。ただし $\mathcal{B}(\mathbb{X})$ は \mathbb{X} 上の Borel 集合族¹である。すなわち、 $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ は可測空間である。

いま、 X_1, X_2, \dots, X_n は確率空間 $(\Omega, \mathcal{A}, \Pr)$ 上の \mathbb{X} 値確率要素²とする。さらに、 $\mathbf{X} = (X_1, X_2, \dots, X_n)^\top$ と書くこととする。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, \dots, 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, \dots, X_n を**ランダム標本**と呼ぶことにする。また、 $\mathbb{X}^n = \underbrace{\mathbb{X} \times \mathbb{X} \times \dots \times \mathbb{X}}_{n \text{ 個}}$ を**標本空間**と呼ぶことにする。

可測空間 $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ 上の確率測度 P を

$$P(B) := \Pr(X_1 \in B) \quad (B \in \mathcal{B}(\mathbb{X}))$$

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、 $X_1 \sim P$ と表記することにする。さらに、 \mathbf{X} の分布を $P^{\otimes n}$ または $P^{\mathbf{X}}$ と書くことにする。したがって、 $B_j \in \mathcal{B}(\mathbb{X})$ ($j = 1, 2, \dots, n$) に対して

$$P^{\otimes n}(B_1 \times B_2 \times \dots \times B_n) = \Pr(\mathbf{X} \in B_1 \times B_2 \times \dots \times B_n)$$

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 $P^{\otimes n}$ にも測度論的な背景がある。すなわち、 $P^{\otimes n}$ は P の n 個の測度の直積測度になっている。

注意 5.1. $B \in \mathcal{B}(\mathbb{X})$ に対して, $\Pr(X_1 \in B)$ と書いたとき

$$\Pr(X_1 \in B) = \Pr(\{\omega \in \Omega; X_1(\omega) \in B\})$$

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 \mathbf{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 \mathbf{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, \dots, X_n ^{i.i.d.} 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, $\mathbf{X} \sim (P^*)^{\otimes n}$ 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. $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ 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 で添え字付けられると仮定する. 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

$$\Theta \ni \theta \mapsto P_\theta \in \mathcal{P}$$

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

$$P^* = P_{\theta^*}$$

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 \mathbf{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 \tilde{X}_{n+1} を観測データ \mathbf{X} に基づいて予測することを目指している. 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 $\mathcal{P} = \{P_\theta; \theta \in \Theta\}$ 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

$$\Theta \ni \theta \mapsto P_\theta \in \mathcal{P}$$

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

$$\theta_1 \neq \theta_2 \Rightarrow P_{\theta_1} \neq P_{\theta_2}$$

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 は数学的にはすばらな表現である. 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

$$\mathcal{P} = \{P_\theta; \theta \in \Theta, P_\theta \text{ は } (\mathbb{X}, \mathcal{B}(\mathbb{X})) \text{ 上の確率測度}\}$$

の組を**統計的実験**といい

$$\left(\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n), \{P_\theta; \theta \in \Theta\} \right)$$

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

$$P^* = P_{\theta^*}$$

である.

この講義では $\mathbf{X} = (X_1, X_2, \dots, X_n)$ と書いたとき, X_1, X_2, \dots, 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 \mathbf{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 $\mathbf{X} = (X_1, X_2, \dots, X_n)^\top$ と書き, 可測空間 $(\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n))$ 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

$$\mathbf{X} \sim (P^*)^{\otimes n}$$

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 $\mathcal{P} = \{P_\theta; \theta \in \Theta\}$ を設定する. ただし P_θ も P^* と同じ可測空間 $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ 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 $P_\theta \in \mathcal{P}$ 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 $P^* = P_{\theta^*}$ と書ける.

これらをまとめて統計的実験といい

$$(\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n), \{P_\theta; \theta \in \Theta\})$$

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, \dots, 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, \dots, X_n は正規分布 $N(\mu, \sigma^2)$ から標本の大きさが n のランダム標本とする. ただし μ, σ ($0 < \sigma < \infty$) が共に未知とする. このとき統計的実験

$$\left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \left\{ p_\theta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (-\infty < x < \infty); \right. \right. \\ \left. \left. \theta := (\mu, \sigma) \in \Theta := \mathbb{R} \times (0, \infty) \right\} \right)$$

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, \dots, X_n は Bernoulli 分布 $\text{Ber}(\theta)$ ($0 < \theta < 1$) 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

$$\left(\{0, 1\}^n, 2^{\{0, 1\}^n}, \left\{ p_\theta(\mathbf{x}) = \theta^x (1-\theta)^{1-x} \quad (x = 0, 1); \theta \in \Theta = (0, 1) \right\} \right)$$

を想定していることになる. □

注意 5.7. X_1, X_2, \dots, X_n は分布 P から標本の大きさが n のランダム標本といったときには,

$$\left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{P_\theta; \theta \in \Theta\} \right)$$

のような統計的実験を仮定し, ある $\theta^* \in \Theta$ があって, $P_{\theta^*} = P^*$ 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なものである. 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 \mathbf{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

$$d: \mathbb{X} \ni \mathbf{x} \mapsto d(\mathbf{x})$$

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

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

を用意する¹¹. この関数を**損失関数**といい, $L(\theta, a)$ 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 $\mathbf{X} = \mathbf{x}$ と真の母数 θ^* における損失関数の値 $L(\theta^*, d(\mathbf{x}))$ がわかればよい. したがって決定関数 d の「よさ」の評

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, $L: \mathbb{A} \times \Theta \mapsto [-1, \infty) \cup \{\infty\}$ とすることもあ

価に $L(\theta^*, d(\mathbf{X}))$ 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 $P_\theta^{\otimes n}$ によって生成されたと仮定し、損失関数 $L(\theta, d(\mathbf{X}))$ を $P_\theta^{\otimes n}$ に関して期待値を取ったもの

$$R(\theta, d) := E_\theta[L(\theta, d(\mathbf{X}))]$$

を考える。ただし、 $E_\theta[\cdot]$ は $\mathbf{X} \sim P_\theta^{\otimes n}$ ($\theta \in \Theta$) 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^* のもとで行いたいところである。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

$$(\mathbb{X}^n, \{P_\theta; \theta \in \Theta\}, \mathbb{A}, \mathbb{D}, L)$$

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 \mathbf{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 $d(\mathbf{X})$ を**推定量**といい、観測データの実現値 $\mathbf{X} = \mathbf{x}$ における推定量の値 $d(\mathbf{x})$ を**推定値**という。 $\Theta = \mathbb{R}$ ならば損失関数として

$$L(a, \theta) = (a - \theta)^2, \quad L(a, \theta) = |a - \theta|$$

を取るのが代表的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

$$R(\theta, d) = E_\theta[L(\theta, d(\mathbf{X}))]$$

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、

$$\left((\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n)), \{P_\theta; \theta \in \Theta\}, (\mathbb{A}, \mathcal{B}(\mathbb{A})), (\mathbb{D}, \mathcal{B}(\mathbb{D})), L \right)$$

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きたいわけである。

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

$$\Theta = \Theta_0 \cup \Theta_1, \quad \Theta_0 \cap \Theta_1 = \emptyset.$$

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 $A = \{0, 1\}$ 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

$$d(\mathbf{x}) = \begin{cases} 1 & (\mathbf{x} \in W) \\ 0 & (\mathbf{x} \notin W) \end{cases}$$

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

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

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、 $d(\mathbf{x})$ に形式的に \mathbf{X} を代入した $d(\mathbf{X})$ 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 $R(\theta, d) = E_\theta[L(\theta, d(\mathbf{X}))]$ は以下ようになる。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

$$\beta(\theta) := R(d, \theta)$$

とし, $\theta \in \Theta_1$ のとき

$$\beta(\theta) := 1 - R(\theta, d)$$

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 α ($0 < \alpha < 1$) に対して

$$\sup_{\theta \in \Theta_0} R(\theta, d) \leq \alpha$$

をみたす検定関数の中から $\theta \in \Theta_1$ において $\beta(\theta)$ を大きくするもの, すなわち $R(d, \theta)$ を小さくするものを選ぶことを目指す. ちなみに α のことを**有意水準**という. $\sup_{\theta \in \Theta_0} R(\theta, d)$ 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 $\mathbf{X} = \mathbf{x}$ に基づく母数 θ の推定区間 $[\ell(\mathbf{x}), u(\mathbf{x})]$ に対して損失関数として

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

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

$$d(\mathbf{X}) = [\ell(\mathbf{X}), u(\mathbf{X})]$$

に対して危険関数は

$$R(\theta, d) = E_\theta[u(\mathbf{X}) - \ell(\mathbf{X})] - \Pr_\theta(\theta \in [\ell(\mathbf{X}), u(\mathbf{X})])$$

となる.

実数 α ($0 < \alpha < 1$) が与えられたとき

$$\Pr_\theta(\theta \in [\ell(\mathbf{X}), u(\mathbf{X})]) \geq 1 - \alpha$$

のもとで区間の長さの期待値 $E_\theta[u(\mathbf{X}) - \ell(\mathbf{X})]$ 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

$$d_1 \prec d_2$$

$$\Leftrightarrow R(\theta, d_1) \leq R(\theta, d_2) (\forall \theta \in \Theta) \text{ かつ } R(\theta_0, d_1) < R(\theta_0, d_2) (\exists \theta_0 \in \Theta)$$

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 $X \sim N(\mu, 1)$ によって μ を推定する問題を損失関数 $L(\hat{\mu}, \mu) = (\hat{\mu} - \mu)^2$ のもとで考える. ただし $\hat{\mu}$ は μ の推定量である. このとき

$$\hat{\mu}_0 = 0$$

なる推定量は許容的になる. なぜならば $\mu = 0$ において $\hat{\mu}_0$ の危険関数の値は 0 となるので, $\hat{\mu}_0$ 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見つける. 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見つけるのではなく, 一定の原理のよって導かれる決定関数を考えて, それについてなんらかの合理性を証明する方針がある. 統計的決定問題の枠組みからははずれるが, ある原理に基づきなんらかのかたちで合理的な正当化ができる決定関数を導出することが考えられてきた. 導出の原理として推定ではモーメント法, 最尤法 (第 7 章) などが知られている. 検定法では尤度比検定, スコア検定, Wald 検定, Rao 検定 (第 8 章) がある. 区間推定では検定統計量の反転, ピボット法 (第 8 章) などがある.

5.3 章末注釈と参考文献

この章の話題を数学的に厳密に議論をするためには, 数学の進んだ道具立てが必要となるので, 直観的な説明に留めた.

5.4 演習問題

演習問題 5.1.

演習問題 5.2.

第6章 母数モデル

第 6.1 では, 統計的モデルのなかでも最も基本的な正則母数モデルとそのもとの Fisher 情報量を導入する. 第 6.2 では, 母数モデルのなかでも非常によい性質と数学的な美しさをもつ指数型分布族とその基本的な性質を説明する. 第 6.4 では, 十分統計量を定義し, 十分性の利便性の高い判定条件を述べる. 第 6.5 では, 最小十分統計量と完備十分統計量を定義し, それらに関わる重要な定理を述べる. いくつかの主張の証明で Lebeague 積分の積分の収束定理を用いた. この点について慣れていない読者は証明を読み飛ばしても全体の流れを理解するには困難が伴わないようにした.

6.1 正則母数モデルと Fisher 情報量

定義 6.1. $d, n \in \mathbb{N}$ とし, $\mathbb{X}^n (\subset \mathbb{R}^n)$ を標本空間¹とする. \mathbb{X} 上の母数モデル $\mathcal{P} = \{P_\theta; \theta \in \Theta \subset \mathbb{R}^d\}$ が与えられているとする. 分布 P_θ の p.d.f.(または p.m.f.) $p_\theta(x)$ と表記する. このとき母数モデルは正則であるとは次の条件をみたすときをいう.

- (1) $\Theta \subset \mathbb{R}^d$ は開集合.
- (2) \mathcal{P} に属する分布の p.d.f.(または p.m.f.) は同じ台をもつ. すなわち, 集合 $\{x \in \mathbb{R}; p_\theta(x) > 0\}$ は $\theta \in \Theta$ に依存しない.
- (3) $\forall \theta \in \Theta$ とする. $p_\theta(x)$ の θ の 1 次と 2 次の偏導関数は $x(\in \mathbb{X})$ に関して連続である.
- (4) $p_\theta(x)$ の θ に関する 1 次と 2 次の偏導関数は $x(\in \mathbb{X})$ の関数として可積分である.
- (5) $p_\theta(x)$ の 1 次と 2 次の偏導関数は θ の微分記号と x の積分記号と交換が可能である.

¹すこし言葉を乱用して, \mathbb{X} のことも標本空間と呼ぶことにする.

$X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ を観測したとき Θ 上の実数値関数 ℓ_n を

$$\ell_n(\boldsymbol{\theta} | \mathbf{x}) = \log \prod_{j=1}^n p_{\boldsymbol{\theta}}(x_j) \quad (\mathbf{x} = (x_1, x_2, \dots, x_n))$$

で定義²する.

$\mathbf{X} \sim P_{\boldsymbol{\theta}}^{\otimes n}$ のときこの分布に関する期待値, 分散および共分散を

$$E_{\boldsymbol{\theta}}[\cdot], \quad \text{Var}_{\boldsymbol{\theta}}[\cdot], \quad \text{Cov}_{\boldsymbol{\theta}}[\cdot, \cdot]$$

と表記する. すなわち, 可積分な関数 $h: \mathbb{X}^n \rightarrow \mathbb{R}$ に対して

$$E_{\boldsymbol{\theta}}[h(\mathbf{X})] := \begin{cases} \sum_{\mathbf{x} \in \mathbb{X}^n} h(\mathbf{x}) \prod_{j=1}^n p_{\boldsymbol{\theta}}(x_j) & (\text{離散型の場合}) \\ \int_{\mathbb{X}^n} h(\mathbf{x}) \prod_{j=1}^n p_{\boldsymbol{\theta}}(x_j) d\mathbf{x} & (\text{連続型の場合}) \end{cases}$$

と定める. ただし $d\mathbf{x} = dx_1 dx_2 \times \dots \times dx_n$ である. さらに, \mathbf{X} を確率空間 $(\Omega, \mathcal{A}, \text{Pr})$ 上の確率ベクトルとしたとき, 任意の $B \in \mathcal{B}(\mathbb{X}^n)$ に対して, 可測空間 (Ω, \mathcal{A}) 上の確率測度 $\text{Pr}_{\boldsymbol{\theta}}$ を

$$\text{Pr}_{\boldsymbol{\theta}}(\mathbf{X}^{-1}(B)) := \text{Pr}_{\boldsymbol{\theta}}(\mathbf{X} \in B) := E_{\boldsymbol{\theta}}[\mathbb{1}_B(\mathbf{X})]$$

で定める. ただし, $\mathbf{X}^{-1}(B) := \{\omega \in \Omega; \mathbf{X}(\omega) \in B\} \in \mathcal{A}$ である.

さらに, $i, j = 1, 2, \dots, d$ に対して

$$\begin{aligned} \dot{\ell}_{n,i} &:= \dot{\ell}_{n,i}(\boldsymbol{\theta} | \mathbf{x}) = \frac{\partial \ell_n(\boldsymbol{\theta} | \mathbf{x})}{\partial \theta_i}, & \ddot{\ell}_{n,ij} &:= \ddot{\ell}_{n,ij}(\boldsymbol{\theta} | \mathbf{x}) = \frac{\partial^2 \ell_n(\boldsymbol{\theta} | \mathbf{x})}{\partial \theta_j \partial \theta_i}, \\ \dot{\boldsymbol{\ell}}_n &= (\dot{\ell}_{n,1}, \dot{\ell}_{n,2}, \dots, \dot{\ell}_{n,d})^\top, & \ddot{\boldsymbol{\ell}}_n &= (\ddot{\ell}_{n,ij}), & \boldsymbol{\theta} &= (\theta_1, \theta_2, \dots, \theta_d)^\top \end{aligned}$$

と定義する. $\mathbf{X} \sim P_{\boldsymbol{\theta}}^{\otimes n}$ のとき Fisher 情報量 $\mathcal{F}_{\mathbf{X}}(\boldsymbol{\theta})$ は $d \times d$ 行列で $\mathcal{F}_{\mathbf{X}}(\boldsymbol{\theta})$ の (i, j) 成分 $\{\mathcal{F}_{\mathbf{X}}(\boldsymbol{\theta})\}_{jk}$ は

$$\{\mathcal{F}_{\mathbf{X}}(\boldsymbol{\theta})\}_{ij} = E_{\boldsymbol{\theta}}[\dot{\ell}_{n,i}(\boldsymbol{\theta} | \mathbf{X}) \dot{\ell}_{n,j}(\boldsymbol{\theta} | \mathbf{X})] \quad (1 \leq i, j \leq d)$$

で定義される.

定理 6.2. 標本空間 \mathbb{X} 上の母数モデル $\mathcal{P} = \{P_{\boldsymbol{\theta}}; \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d\}$ は正則であるとする. $\mathbf{X} \sim P_{\boldsymbol{\theta}}^{\otimes n}$ としたとき以下が成立する.

- (1) $\forall \boldsymbol{\theta} \in \Theta$ に対して $E_{\boldsymbol{\theta}}[\dot{\boldsymbol{\ell}}_n(\boldsymbol{\theta} | \mathbf{X})] = \mathbf{0}_d$ となる.
- (2) $\mathcal{F}_{\mathbf{X}}(\boldsymbol{\theta}) = \text{Cov}_{\boldsymbol{\theta}}[\dot{\boldsymbol{\ell}}_n(\boldsymbol{\theta} | \mathbf{X})] = E_{\boldsymbol{\theta}}[\dot{\boldsymbol{\ell}}_n(\boldsymbol{\theta} | \mathbf{X}) \dot{\boldsymbol{\ell}}_n^\top(\boldsymbol{\theta} | \mathbf{X})]$ が成り立つ.

²対数尤度関数と呼ぶ.

(3) $\mathcal{F}_{\mathbf{X}}(\boldsymbol{\theta}) = -E_{\boldsymbol{\theta}}[\ddot{\ell}_n(\boldsymbol{\theta} | \mathbf{X})]$ が成り立つ.

Proof. 証明は連続型分布の場合を示す. 離散型の場合は積分を和の記号に置き替えればよい. $P_{\boldsymbol{\theta}}$ の p.d.f. を $p_{\boldsymbol{\theta}}(x)$ と表記する. さらに, $\tilde{\mathbb{X}} = \{x \in \mathbb{X}; p_{\boldsymbol{\theta}}(x) > 0\}$ とおく. 母数モデルの正則性から微分記号と積分記号の交換が保証されているので, $i = 1, 2, \dots, d$ に対して

$$\begin{aligned} E_{\boldsymbol{\theta}}[\dot{\ell}_{n,i}(\boldsymbol{\theta} | \mathbf{X})] &= \sum_{k=1}^n \int_{\mathbb{X}} \left(\frac{\partial}{\partial \theta_i} \log p_{\boldsymbol{\theta}}(x_k) \right) p_{\boldsymbol{\theta}}(x_k) dx_k \\ &= \sum_{k=1}^n \left\{ \int_{\tilde{\mathbb{X}}} \left(\frac{\partial}{\partial \theta_i} \log p_{\boldsymbol{\theta}}(x_k) \right) p_{\boldsymbol{\theta}}(x_k) dx_k \right. \\ &\quad \left. + \int_{\mathbb{X} \setminus \tilde{\mathbb{X}}} \left(\frac{\partial}{\partial \theta_i} \log p_{\boldsymbol{\theta}}(x_k) \right) \underbrace{p_{\boldsymbol{\theta}}(x_k)}_{=0} dx_k \right\} \\ &= \sum_{k=1}^n \int_{\tilde{\mathbb{X}}} \frac{1}{p_{\boldsymbol{\theta}}(x_k)} \left(\frac{\partial}{\partial \theta_i} p_{\boldsymbol{\theta}}(x_k) \right) p_{\boldsymbol{\theta}}(x_k) dx_k \\ &= \sum_{k=1}^n \int_{\tilde{\mathbb{X}}} \frac{\partial}{\partial \theta_i} p_{\boldsymbol{\theta}}(x_k) dx_k \\ &= \sum_{k=1}^n \frac{\partial}{\partial \theta_i} \int_{\tilde{\mathbb{X}}} p_{\boldsymbol{\theta}}(x_k) dx_k \\ &= \sum_{k=1}^n \frac{\partial}{\partial \theta_i} \left\{ \int_{\tilde{\mathbb{X}}} p_{\boldsymbol{\theta}}(x_k) dx_k + \int_{\mathbb{X} \setminus \tilde{\mathbb{X}}} \underbrace{p_{\boldsymbol{\theta}}(x_k)}_{=0} dx_k \right\} \\ &= \sum_{k=1}^n \frac{\partial}{\partial \theta_i} \underbrace{\int_{\mathbb{X}} p_{\boldsymbol{\theta}}(x_k) dx_k}_{=1} \\ &= 0 \end{aligned}$$

となる. よって (1) は示された.

(2) は明らか.

(3) を示すために次に注意する. $i, j = 1, 2, \dots, d$ と $x \in \tilde{\mathbb{X}}$ に対して

$$\frac{\partial^2 \log p_{\boldsymbol{\theta}}(x)}{\partial \theta_i \partial \theta_j} = \frac{1}{p_{\boldsymbol{\theta}}(x)} \frac{\partial^2 p(x | \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} - \left(\frac{\partial}{\partial \theta_i} \log p_{\boldsymbol{\theta}}(x) \right) \left(\frac{\partial}{\partial \theta_j} \log p_{\boldsymbol{\theta}}(x) \right)$$

となる. このとき

$$\begin{aligned} \int_{\tilde{\mathbb{X}}} \frac{1}{p_{\boldsymbol{\theta}}(x)} \left(\frac{\partial^2 p_{\boldsymbol{\theta}}(x)}{\partial \theta_i \partial \theta_j} \right) p_{\boldsymbol{\theta}}(x) dx &= \frac{\partial^2}{\partial \theta_i \partial \theta_j} \left\{ \int_{\tilde{\mathbb{X}}} p_{\boldsymbol{\theta}}(x) dx + \int_{\mathbb{X} \setminus \tilde{\mathbb{X}}} p_{\boldsymbol{\theta}}(x) dx \right\} \\ &= \frac{\partial^2}{\partial \theta_i \partial \theta_j} \underbrace{\int_{\mathbb{X}} p_{\boldsymbol{\theta}}(x) dx}_{=1} = 0 \end{aligned}$$

と

$$\begin{aligned} & \int_{\tilde{\mathcal{X}}} \left(\frac{\partial}{\partial \theta_i} \log \mathbf{p}_{\boldsymbol{\theta}}(x) \right) \left(\frac{\partial}{\partial \theta_j} \log \mathbf{p}_{\boldsymbol{\theta}}(x) \right) \mathbf{p}_{\boldsymbol{\theta}}(x) dx \\ &= \mathbf{E}_{\boldsymbol{\theta}} \left[\left(\frac{\partial}{\partial \theta_i} \log \mathbf{p}_{\boldsymbol{\theta}}(X_1) \right) \left(\frac{\partial}{\partial \theta_j} \log \mathbf{p}_{\boldsymbol{\theta}}(X_1) \right) \right] \end{aligned}$$

となる. このことより

$$\begin{aligned} \mathbf{E}_{\boldsymbol{\theta}} \left[\frac{\partial^2 \log \mathbf{p}_{\boldsymbol{\theta}}(X_1)}{\partial \theta_i \partial \theta_j} \right] &= \int_{\tilde{\mathcal{X}}} \frac{\partial^2 \log \mathbf{p}_{\boldsymbol{\theta}}(x)}{\partial \theta_i \partial \theta_j} \mathbf{p}_{\boldsymbol{\theta}}(x) dx \\ &\quad + \int_{\mathcal{X} \setminus \tilde{\mathcal{X}}} \frac{\partial^2 \log \mathbf{p}_{\boldsymbol{\theta}}(x)}{\partial \theta_i \partial \theta_j} \mathbf{p}_{\boldsymbol{\theta}}(x) dx \\ &= \int_{\tilde{\mathcal{X}}} \frac{\partial^2 \log \mathbf{p}_{\boldsymbol{\theta}}(x)}{\partial \theta_i \partial \theta_j} \mathbf{p}_{\boldsymbol{\theta}}(x) dx \\ &= \underbrace{\int_{\tilde{\mathcal{X}}} \frac{1}{\mathbf{p}_{\boldsymbol{\theta}}(x)} \left(\frac{\partial^2 \mathbf{p}_{\boldsymbol{\theta}}(x)}{\partial \theta_i \partial \theta_j} \right) \mathbf{p}_{\boldsymbol{\theta}}(x) dx}_{=0} \\ &\quad - \mathbf{E}_{\boldsymbol{\theta}} \left[\left(\frac{\partial}{\partial \theta_i} \log \mathbf{p}_{\boldsymbol{\theta}}(X_1) \right) \left(\frac{\partial}{\partial \theta_j} \log \mathbf{p}_{\boldsymbol{\theta}}(X_1) \right) \right] \\ &= -\mathbf{E}_{\boldsymbol{\theta}} \left[\left(\frac{\partial}{\partial \theta_i} \log \mathbf{p}_{\boldsymbol{\theta}}(X_1) \right) \left(\frac{\partial}{\partial \theta_j} \log \mathbf{p}_{\boldsymbol{\theta}}(X_1) \right) \right] \quad (6.1) \end{aligned}$$

がわかる. よって

$$\begin{aligned} \{\mathcal{F}_{\mathbf{X}}(\boldsymbol{\theta})\}_{ij} &= \mathbf{E} \left[\left(\sum_{k_1=1}^n \frac{\partial \log \mathbf{p}_{\boldsymbol{\theta}}(X_{k_1})}{\partial \theta_i} \right) \left(\sum_{k_2=1}^n \frac{\partial \log \mathbf{p}_{\boldsymbol{\theta}}(X_{k_2})}{\partial \theta_j} \right) \right] \\ &= \sum_{k=1}^n \mathbf{E} \left[\frac{\partial \log \mathbf{p}_{\boldsymbol{\theta}}(X_k)}{\partial \theta_i} \frac{\partial \log \mathbf{p}_{\boldsymbol{\theta}}(X_k)}{\partial \theta_j} \right] \\ &\quad + \sum_{k_1 \neq k_2} \mathbf{E} \left[\frac{\partial \log \mathbf{p}_{\boldsymbol{\theta}}(X_{k_1})}{\partial \theta_i} \frac{\partial \log \mathbf{p}_{\boldsymbol{\theta}}(X_{k_2})}{\partial \theta_j} \right] \\ &= \sum_{k=1}^n \mathbf{E} \left[\frac{\partial \log \mathbf{p}_{\boldsymbol{\theta}}(X_k)}{\partial \theta_i} \frac{\partial \log \mathbf{p}_{\boldsymbol{\theta}}(X_k)}{\partial \theta_j} \right] \\ &\quad + \sum_{k_1 \neq k_2} \underbrace{\mathbf{E} \left[\frac{\partial \log \mathbf{p}_{\boldsymbol{\theta}}(X_{k_1})}{\partial \theta_i} \right]}_{=0} \mathbf{E} \left[\frac{\partial \log \mathbf{p}_{\boldsymbol{\theta}}(X_{k_2})}{\partial \theta_j} \right] \\ &= -\sum_{k=1}^n \mathbf{E}_{\boldsymbol{\theta}} \left[\frac{\partial^2 \log \mathbf{p}_{\boldsymbol{\theta}}(X_k)}{\partial \theta_i \partial \theta_j} \right] \quad (\because (6.1)) \\ &= -\mathbf{E} \left[\frac{\partial^2 \ell_n(\boldsymbol{\theta} | \mathbf{X})}{\partial \theta_i \partial \theta_j} \right] = -\mathbf{E}[\ddot{\ell}_{n,ij}(\boldsymbol{\theta} | \mathbf{X})] \end{aligned}$$

がわかる. □

例 6.3. $\mathbf{X} = (X_1, X_2, \dots, X_n)$ で $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$ とする. このとき, 統計的モデルは $\mathcal{P} = \{N(\mu, \sigma^2); \mu \in \mathbb{R}, \sigma > 0\}$ で与えられる. したがって

$$\Theta = \{\boldsymbol{\theta} = (\mu, \sigma) \in \mathbb{R}^2; \mu \in \mathbb{R}, \sigma > 0\} = \mathbb{R} \times (0, \infty)$$

となる. このとき

$$\ell_n(\boldsymbol{\theta} | \mathbf{x}) = -\sum_{j=1}^n \frac{(x_j - \mu)^2}{2\sigma^2} - n \log \sigma - n \log(\sqrt{2\pi}); \quad \mathbf{x} = (x_1, x_2, \dots, x_n)$$

となる. よって

$$\begin{aligned} \dot{\ell}_{n,1}(\boldsymbol{\theta} | \mathbf{x}) &= \frac{\partial \ell_n(\boldsymbol{\theta} | \mathbf{x})}{\partial \mu} = \sum_{j=1}^n \frac{x_j - \mu}{\sigma^2}, \\ \dot{\ell}_{n,2}(\boldsymbol{\theta} | \mathbf{x}) &= \frac{\partial \ell_n(\boldsymbol{\theta} | \mathbf{x})}{\partial \sigma} = \sum_{j=1}^n \frac{(x_j - \mu)^2}{\sigma^3} - \frac{n}{\sigma} \end{aligned}$$

となる. これらの期待値を取ると

$$\begin{aligned} \mathbb{E}_{\boldsymbol{\theta}}[\{\dot{\ell}_{n,1}(\boldsymbol{\theta} | \mathbf{X})\}^2] &= \frac{n}{\sigma^2}, \\ \mathbb{E}_{\boldsymbol{\theta}}[\{\dot{\ell}_{n,2}(\boldsymbol{\theta} | \mathbf{X})\}^2] &= \frac{3n}{\sigma^2} - \frac{2n}{\sigma^2} + \frac{n}{\sigma^2} = \frac{2n}{\sigma^2}, \\ \mathbb{E}_{\boldsymbol{\theta}}[\dot{\ell}_{n,1}(\boldsymbol{\theta} | \mathbf{X}) \dot{\ell}_{n,2}(\boldsymbol{\theta} | \mathbf{X})] &= 0 \end{aligned}$$

となるので

$$\mathcal{F}_X(\boldsymbol{\theta}) = \frac{n}{\sigma^2} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

を得る.

一方

$$\begin{aligned} \ddot{\ell}_{n,11} &= \frac{\partial^2 \ell(\boldsymbol{\theta} | \mathbf{x})}{\partial \mu^2} = \frac{\partial}{\partial \mu} \left(\sum_{j=1}^n \frac{x_j - \mu}{\sigma^2} \right) = -\frac{n}{\sigma^2}, \\ \ddot{\ell}_{n,22} &= \frac{\partial^2 \ell(\boldsymbol{\theta} | \mathbf{x})}{\partial \sigma^2} = \frac{\partial}{\partial \sigma} \left(\sum_{j=1}^n \frac{(x_j - \mu)^2}{\sigma^2} - \frac{n}{\sigma} \right) = -3 \sum_{j=1}^n \frac{(x_j - \mu)^2}{\sigma^2} + \frac{n}{\sigma^2}, \\ \ddot{\ell}_{n,12} &= \frac{\partial^2 \ell(\boldsymbol{\theta} | \mathbf{x})}{\partial \mu \partial \sigma} = \frac{\partial}{\partial \sigma} \left[\sum_{j=1}^n \frac{x_j - \mu}{\sigma^2} \right] = -2 \sum_{j=1}^n \frac{x_j - \mu}{\sigma^3} \end{aligned}$$

となる. よって

$$\mathbb{E}_{\boldsymbol{\theta}}[\ddot{\ell}_{n,11}(\boldsymbol{\theta} | \mathbf{X})] = -\frac{n}{\sigma^2}, \mathbb{E}_{\boldsymbol{\theta}}[\ddot{\ell}_{n,22}(\boldsymbol{\theta} | \mathbf{X})] = -\frac{2n}{\sigma^2}, \mathbb{E}_{\boldsymbol{\theta}}[\ddot{\ell}_{n,12}(\boldsymbol{\theta} | \mathbf{X})] = 0$$

なので

$$\begin{bmatrix} E_{\theta}[\ddot{\ell}_{n,11}(\boldsymbol{\theta}|\mathbf{X})] & E_{\theta}[\ddot{\ell}_{n,12}(\boldsymbol{\theta}|\mathbf{X})] \\ E_{\theta}[\ddot{\ell}_{n,12}(\boldsymbol{\theta}|\mathbf{X})] & E_{\theta}[\ddot{\ell}_{n,22}(\boldsymbol{\theta}|\mathbf{X})] \end{bmatrix} = -\mathcal{F}_X(\boldsymbol{\theta})$$

となる.

□

6.2 指数型分布族

この節では, 統計的モデルでもっとも重要なものを導入し, そのモデルの性質を説明する.

6.2.1 1 変数の場合

定義 6.4. $d \in \mathbb{N}$ とし, $\mathbb{X} \subset \mathbb{R}^d$ を空でない部分集合とする. さらに $\Theta \subset \mathbb{R}$ も空でない部分集合とする. \mathbb{X} 上の統計的モデル $\{P_{\theta}; \theta \in \Theta\}$ は 1 母数指数型分布族であるとは, Θ 上の実数値関数 A と κ^{\vee} , \mathbb{X} 上の実数値関数 T と h が存在して³, P_{θ} の p.d.f. または p.m.f. $p_{\theta}^{\vee}(\mathbf{x})$ ⁴が

$$p_{\theta}^{\vee}(\mathbf{x}) = \begin{cases} h(\mathbf{x}) \exp\{A(\theta)T(\mathbf{x}) - \kappa^{\vee}(\theta)\} & (\mathbf{x} \in \mathbb{X} \subset \mathbb{R}^d) \\ 0 & (\mathbf{x} \notin \mathbb{X}) \end{cases} \quad (6.2)$$

の形で書けるときをいう.

注意 6.5. 関数 A, h, T の表現は一意ではない.

□

例 6.6. (Poisson 分布族) $X \sim \text{Po}(\theta)$ ($\theta > 0$) とする. その p.m.f. は

$$p_{\theta}^{\vee}(x) = \frac{\theta^x e^{-\theta}}{x!} = \frac{1}{x!} \exp\{x \log \theta - \theta\} \quad (x = 0, 1, \dots)$$

である. したがって $\{\text{Po}(\theta); \theta \in (0, \infty)\}$ は

$$d = 1, A(\theta) = \log \theta, \kappa^{\vee}(\theta) = \theta, T(x) = x, h(x) = \frac{1}{x!}$$

によって生成される 1 母数指数型分布族である.

□

³ A と B は Greek letters 大文字の α と β である.

⁴後に別の母数化である自然指数型分布族を導入するときに記号を簡単にするために, ここでは「 p^{\vee} と κ^{\vee} 」という少し不自然な記号を用いた.

例 6.7. (2 項分布族) $X \sim \text{Bino}(n, \theta)$ とする. ただし $n \in \mathbb{N}$, $0 < \theta < 1$ である. このとき X の p.m.f. は

$$\begin{aligned} p_{\theta}^{\vee}(x) &= \binom{n}{x} \theta^x (1 - \theta)^{n-x} \\ &= \binom{n}{x} \exp\left\{x \log\left(\frac{\theta}{1 - \theta}\right) + n \log(1 - \theta)\right\} \quad (x = 0, 1, \dots, n) \end{aligned}$$

と書ける. ただし

$$\binom{n}{x} = \frac{n!}{x! \times (n - x)!}, \quad 0! = 1$$

である. したがって $\{\text{Bino}(n, \theta); \theta \in (0, 1)\}$ は

$$d = 1, A(\theta) = \log\left(\frac{\theta}{1 - \theta}\right), \kappa^{\vee}(\theta) = -n \log(1 - \theta), T(x) = x,$$

$$h(x) = \binom{n}{x}$$

によって生成される 1 母数指数型分布族である. □

例 6.8. $\mathbf{X} = (Y, Z)^{\top}$, $Z, W \stackrel{\text{i.i.d.}}{\sim} \text{N}(0, 1)$ とし, $Y = Z + \theta W$ ($\theta > 0$) とする. このとき \mathbf{X} の同時 p.d.f. は次で与えられる. $\mathbf{x} = (y, z)$ としたとき

$$\begin{aligned} p_{\theta}^{\mathbf{X}}(\mathbf{x}) &= p_{\theta}^{\mathbf{X}}(y, z) = p^Z(z) p^{Y|Z}(y|z) = \varphi(z) \theta^{-1} \varphi\left(\frac{y - z}{\theta}\right) \\ &= \frac{1}{2\pi\theta} \exp\left\{-\frac{1}{2}\left(z^2 + \frac{(y - z)^2}{\theta^2}\right)\right\} \\ &= \frac{1}{2\pi} \exp\left\{-\frac{z^2}{2}\right\} \exp\left\{-\frac{(y - z)^2}{2\theta^2} - \log \theta\right\} \end{aligned}$$

と書ける. ただし p_Z は Z の p.d.f. とし, $p_{Y|Z}$ は $Z = z$ を与えたときの Y の条件付き p.d.f., φ は $\text{N}(0, 1)$ の p.d.f. で

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad (z \in \mathbb{R})$$

である. したがって \mathbf{X} の分布は

$$\begin{aligned} d &= 2, A(\theta) = -\frac{1}{2\theta^2}, \kappa^{\vee}(\theta) = \log \theta, T(\mathbf{x}) = (y - z)^2, \\ h(\mathbf{x}) &= \frac{1}{2\pi} \exp\left\{-\frac{z^2}{2}\right\} \end{aligned}$$

で生成される 1 母数指数型分布族に属する. □

1 母数指数型分布族に属する分布からのランダム標本の分布も 1 母数指数型分布族に属する. 特に, $n \in \mathbb{N}$ とし, $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} P_\theta$ とする. ただし, $\{P_\theta; \theta \in \Theta\}$ は (6.2) で与えられる指数型分布族とする. すると $\{P_\theta^{\otimes n}; \theta \in \Theta\}$ は $\mathbf{X}^{(n)} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$ に対する指数型分布族となる. ただし, \mathbb{R}^{dn} 上の確率測度を

$$P_\theta^{\otimes n} = \underbrace{P_\theta \times P_\theta \times \dots \times P_\theta}_{n \text{ 個}}$$

と定めてた. このとき, $P_\theta^{\otimes n}$ の同時 p.d.f. $\prod_{j=1}^n p_\theta^\vee(\cdot)$ は次で与えられる. すると

$$\begin{aligned} \prod_{j=1}^n p_\theta^\vee(\mathbf{x}_j) &= \prod_{j=1}^n h(\mathbf{x}_j) \exp\left\{A(\theta)T(\mathbf{x}_j) - \kappa^\vee(\theta)\right\} \\ &= \left[\prod_{j=1}^n h(\mathbf{x}_j)\right] \exp\left\{A(\theta) \sum_{j=1}^n T(\mathbf{x}_j) - n\kappa^\vee(\theta)\right\} \end{aligned}$$

となる. したがって $\{P_\theta^{\otimes n}; \theta \in \Theta\}$ は

$$A(\theta), \quad \sum_{j=1}^n T(\mathbf{x}_j), \quad n\kappa^\vee(\theta), \quad \prod_{j=1}^n h(\mathbf{x}_j)$$

で生成される \mathbb{R}^{dn} 上の 1 母数指数型分布族となる.

統計量 $\sum_{j=1}^n T(\mathbf{X}_j)$ は θ の自然十分統計量⁵となる.

6.3 指数型分布族の正準表示

(6.2) で表現された指数型分布族を θ でなく $\eta (= A(\theta))$ で添え字付けることを考える. 記号の乱用⁶すると 1 母数指数型分布族は

$$p_\eta(\mathbf{x}) = \begin{cases} h(\mathbf{x}) \exp\{\eta T(\mathbf{x}) - \kappa(\eta)\} & (\mathbf{x} \in \mathbb{X} \subset \mathbb{R}^d) \\ 0 & (\mathbf{x} \notin \mathbb{X}) \end{cases} \quad (6.3)$$

と書くことができる. ただし

$$\kappa(\eta) = \begin{cases} \log\left(\sum_{\mathbf{x} \in \mathbb{X}} h(\mathbf{x}) \exp\{\eta T(\mathbf{x})\}\right) & (\text{離散型の場合}) \\ \log\left(\int_{\mathbb{X}} h(\mathbf{x}) \exp\{\eta T(\mathbf{x})\} d\mathbf{x}\right) & (\text{連続型の場合}) \end{cases}$$

⁵十分統計量の定義は定義 6.12 を参照のこと.

⁶ここでは, 記号の乱用をしている. 本来であれば, 領域を制限すると A の逆関数は存在するので, $p(\mathbf{x}|\eta)$ は $p(\mathbf{x}|A^{-1}(\eta))$ と書くべきであろう.

である. κ の定義より, 直ちに $\kappa(\eta) = \kappa(A(\theta)) = \kappa^\vee(\theta)$ がわかる. さらに

$$\mathcal{E} := \{\eta \in \mathbb{R}; \kappa(\eta) < \infty\}$$

とおく. \mathcal{E} は \mathbb{R} の区間となる. すると確率分布族

$$\left\{ p(\mathbf{x}|\eta) = h(\mathbf{x}) \exp\{\eta T(\mathbf{x}) - \kappa(\eta)\}; \eta \in \mathcal{E} \right\}$$

は確率分布族

$$\left\{ p_\theta(\mathbf{x}) = h(\mathbf{x}) \exp\{A(\theta)T(\mathbf{x}) - \kappa(\theta)\}; \theta \in \Theta \right\}$$

を含む. 確率分布族 $\{p(\mathbf{x}|\eta); \eta \in \mathcal{E}\}$ は T, h によって生成された せいじゅん正準指数型分布族または自然指数型分布族といい, \mathcal{E} を自然母数空間といい, T を自然十分統計量という.

例 6.9. (例 6.6 の続き) Poisson 分布族は

$$\begin{aligned} p_\eta(x) &= \frac{1}{x!} \exp\{\eta x - e^\eta\} \quad (x = 0, 1, 2, \dots), \\ \eta &= \log \theta, \quad h(x) = \frac{1}{x!}, \quad T(x) = x, \\ \exp(\kappa(\eta)) &= \sum_{x=0}^{\infty} h(x) \exp(\eta T(x)) = \sum_{x=0}^{\infty} \frac{\exp(\eta x)}{x!} = \sum_{x=0}^{\infty} \frac{(e^\eta)^x}{x!} = \exp(e^\eta), \\ \mathcal{E} &= \mathbb{R} \end{aligned}$$

となる. □

補題 6.10. ⁷

$$\mathcal{E} := \left\{ \eta \in \mathbb{R}; \kappa(\eta) := \mathbb{E}[\exp\{\eta T(\mathbf{X})\}] < \infty \right\}$$

とする. このとき, 関数 $\kappa(\eta)$ は \mathcal{E}° 上で無限回微分可能である. ただし \mathcal{E}° は \mathcal{E} の内部である. さらに積分記号と微分記号の交換は可能である.

Proof. $\eta \in \mathcal{E}^\circ$ なので, ある $\epsilon > 0$ が存在して $[\eta - 2\epsilon, \eta + 2\epsilon] \subset \mathcal{E}^\circ$ とで

⁷証明には, 優収束的理を用いている.

きる.

$$\begin{aligned}
 \frac{d}{d\eta} \exp\{\kappa(\eta)\} &= \lim_{n \rightarrow \infty} \frac{\exp\left\{\kappa\left(\eta + \frac{\epsilon}{n}\right)\right\} - \exp\{\kappa(\eta)\}}{\epsilon/n} \\
 &= \lim_{n \rightarrow \infty} \int \frac{\exp\left\{\left(\eta + \frac{\epsilon}{n}\right)T(\mathbf{x})\right\} - \exp\{\kappa(\eta)\}}{\epsilon/n} h(\mathbf{x}) \, d\mathbf{x} \\
 &= \lim_{n \rightarrow \infty} \int \frac{\exp\{\eta T(\mathbf{x})\} \left[\exp\left\{\frac{\epsilon T(\mathbf{x})}{n}\right\} - 1\right]}{\epsilon/n} h(\mathbf{x}) \, d\mathbf{x} \\
 &=: \lim_{n \rightarrow \infty} \int \exp\{\eta T(\mathbf{x})\} f_n(\mathbf{x}) \, d\mathbf{x}
 \end{aligned}$$

と書ける. ただし

$$f_n(\mathbf{x}) = \frac{\exp\{\epsilon T(\mathbf{x})/n\} - 1}{\epsilon/n} h(\mathbf{x})$$

である. すると

$$\lim_{n \rightarrow \infty} f_n(\mathbf{x}) =: f(\mathbf{x}) = T(\mathbf{x})h(\mathbf{x})$$

となる. ここで

$$|e^t - 1| \leq |t|e^{|t|}; \quad |t| \leq e^{|t|} \quad (t \in \mathbb{R})$$

に注意する. これらの不等式を用いると

$$\begin{aligned}
 \left| \frac{\exp\{\epsilon T(\mathbf{x})/n\} - 1}{\epsilon/n} \right| &\leq \left\{ \left| \frac{\epsilon T(\mathbf{x})}{n} \right| / \frac{\epsilon}{n} \right\} \exp\left\{ \left| \frac{\epsilon T(\mathbf{x})}{n} \right| \right\} \\
 &\leq \frac{1}{\epsilon} |\epsilon T(\mathbf{x})| \exp\left\{ \left| \frac{\epsilon T(\mathbf{x})}{n} \right| \right\} \\
 &\leq \frac{1}{\epsilon} \exp\{|\epsilon T(\mathbf{x})|\} \exp\{|\epsilon T(\mathbf{x})|\} \\
 &= \frac{1}{\epsilon} \exp\{|2\epsilon T(\mathbf{x})|\} \\
 &\leq \frac{1}{\epsilon} \left\{ \exp(2\epsilon T(\mathbf{x})) + \exp(-2\epsilon T(\mathbf{x})) \right\}
 \end{aligned}$$

を得る. よって

$$|f_n(\mathbf{x})| \leq \frac{1}{\epsilon} \left\{ \exp(2\epsilon T(\mathbf{x})) + \exp(-2\epsilon T(\mathbf{x})) \right\} h(\mathbf{x})$$

となるので

$$\left| \exp\{\eta T(\mathbf{x})\} f_n(\mathbf{x}) \right| \leq \frac{1}{\epsilon} \left\{ \exp((\eta + 2\epsilon)T(\mathbf{x})) + \exp((\eta - 2\epsilon)T(\mathbf{x})) \right\} h(\mathbf{x})$$

$$=: g(\mathbf{x})$$

を得る. $[\eta - 2\epsilon, \eta + 2\epsilon] \subset \mathcal{E}^\circ$ から

$$\int g(\mathbf{x}) \, d\mathbf{x} < \infty$$

となる. よって優収束定理から

$$\begin{aligned} \frac{d}{d\eta} \exp\{\kappa(\eta)\} &= \lim_{n \rightarrow \infty} \int \exp\{\eta T(\mathbf{x})\} f_n(\mathbf{x}) \, d\mathbf{x} \\ &= \int \exp\{\eta T(\mathbf{x})\} T(\mathbf{x}) h(\mathbf{x}) \, d\mathbf{x} \end{aligned}$$

がわかる. この操作を繰り返して行けば, $\kappa(\eta)$ は無限回微分可能なことがわかる. \square

定理 6.11. X は (6.3) で与えられた自然指数型分布族に属する分布 $\mathbf{p}(\mathbf{x}|\eta)$ に従うとする. η は \mathcal{E} の内点としたとき, $T(\mathbf{X})$ の積率母関数は原点の近傍で存在し

$$M_T(s) = \exp[\kappa(s + \eta) - \kappa(\eta)]$$

で与えられる. ただし, s は 0 のある近傍に含まれるとする. さらに

$$E[T(\mathbf{X})] = \dot{\kappa}(\eta); \quad \text{Var}[T(\mathbf{X})] = \ddot{\kappa}(\eta) \quad (6.4)$$

である. ただし

$$\dot{\kappa}(\eta) = \frac{d\kappa}{d\eta}(\eta), \quad \ddot{\kappa}(\eta) = \frac{d^2\kappa}{d\eta^2}(\eta)$$

である.

Proof. 連続型分布の場合について証明する. 離散型の場合には積分を和に変更すればよい. κ の定義に注意すると

$$\begin{aligned} M(s) &= E[\exp(sT(\mathbf{X}))] \\ &= \int_{\mathbf{X}} h(\mathbf{x}) \exp[(s + \eta)T(\mathbf{x}) - \kappa(\eta)] \, d\mathbf{x} \\ &= \exp[\kappa(s + \eta) - \kappa(\eta)] \int_{\mathbf{X}} h(\mathbf{x}) \exp[(s + \eta)T(\mathbf{x}) - \kappa(s + \eta)] \, d\mathbf{x} \\ &= \exp[\kappa(s + \eta) - \kappa(\eta)] \end{aligned}$$

から 1 番目の主張はわかる. 最後の等号は $s + \eta \in \mathcal{E}^\circ$ となるように s をとると

$$\int_{\mathbf{x}} h(\mathbf{x}) \exp[(s + \eta)T(\mathbf{x}) - \kappa(s + \eta)] d\mathbf{x} = 1$$

となることよりわかる. 次に, $\kappa(\cdot)$ の定義と補題 6.10 から

$$\begin{aligned} \frac{d\kappa}{d\eta}(\eta) &= \frac{d}{d\eta} \log \left(\int_{\mathbf{x}} \exp\{\eta T(\mathbf{x})\} h(\mathbf{x}) d\mathbf{x} \right) \\ &= \frac{\frac{d}{d\eta} \int_{\mathbf{x}} \exp\{\eta T(\mathbf{x})\} h(\mathbf{x}) d\mathbf{x}}{\int_{\mathbf{x}} \exp\{\eta T(\mathbf{x})\} h(\mathbf{x}) d\mathbf{x}} \\ &= \frac{\int_{\mathbf{x}} \frac{d}{d\eta} \exp\{\eta T(\mathbf{x})\} h(\mathbf{x}) d\mathbf{x}}{\exp\{\kappa(\eta)\}} \\ &= \frac{\int_{\mathbf{x}} T(\mathbf{x}) \exp\{\eta T(\mathbf{x})\} h(\mathbf{x}) d\mathbf{x}}{\exp\{\kappa(\eta)\}} \\ &= \int_{\mathbf{x}} T(\mathbf{x}) \exp\{\eta T(\mathbf{x}) - \kappa(\eta)\} h(\mathbf{x}) d\mathbf{x} \\ &= E[T(\mathbf{X})] \end{aligned}$$

から (6.4) の 1 番目の等号が示せた. 同様に

$$\begin{aligned} \frac{d^2\kappa}{d\eta^2}(\eta) &= \frac{d}{d\eta} \int_{\mathbf{x}} T(\mathbf{x}) \exp\{\eta T(\mathbf{x}) - \kappa(\eta)\} h(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbf{x}} T(\mathbf{x}) h(\mathbf{x}) \frac{d}{d\eta} \exp\{\eta T(\mathbf{x}) - \kappa(\eta)\} d\mathbf{x} \\ &= \int_{\mathbf{x}} T(\mathbf{x}) h(\mathbf{x}) \exp\{\eta T(\mathbf{x}) - \kappa(\eta)\} \frac{d}{d\eta} (\eta T(\mathbf{x}) - \kappa(\eta)) d\mathbf{x} \\ &= \int_{\mathbf{x}} T^2(\mathbf{x}) h(\mathbf{x}) \exp\{\eta T(\mathbf{x}) - \kappa(\eta)\} d\mathbf{x} \\ &\quad - \underbrace{\frac{d}{d\eta} \kappa(\eta)}_{=E[T(\mathbf{X})]} \int_{\mathbf{x}} T(\mathbf{x}) h(\mathbf{x}) \exp\{\eta T(\mathbf{x}) - \kappa(\eta)\} d\mathbf{x} \\ &= E[T^2(\mathbf{X})] - \{E[T(\mathbf{X})]\}^2 \\ &= \text{Var}[T(\mathbf{X})] \end{aligned}$$

から (6.4) の 2 番目の等号が示せた. □

6.4 十分統計量

$\mathbb{X} \subset \mathbb{R}$ を空でない部分集合とする. $n \in \mathbb{N}$ とし, X_1, X_2, \dots, X_n は確率空間 $(\Omega, \mathcal{A}, \Pr)$ 上の \mathbb{X} 値確率変数数列で, 互いに独立で各々同一の確率分布を持つとする. $\mathbf{X} = (X_1, X_2, \dots, X_n)^\top$ とおき $B \in \mathcal{B}(\mathbb{X}^n)$ に対し

$$\begin{aligned} \Pr^{\mathbf{X}}(B) &= \Pr\{\mathbf{X}^{-1}(B)\} \\ &= \Pr\left(\{\omega \in \Omega; \mathbf{X}(\omega) = (X_1(\omega), X_2(\omega), \dots, X_n(\omega)) \in B\}\right) \end{aligned}$$

とおく. さらに $B \in \mathcal{B}(\mathbb{X})$ に対し

$$\Pr^{X_j}(B) = \Pr\left(\{\omega \in \Omega; X_j(\omega) \in B\}\right) \quad (j = 1, 2, \dots, n)$$

と書いたとき

$$\Pr^{\mathbf{X}} = \Pr^{X_1} \times \Pr^{X_2} \times \dots \times \Pr^{X_n}$$

と書ける. \mathbf{X} により誘導された確率空間を $(\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n), \Pr^{\mathbf{X}})$ と書く.

$(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ 上の統計的モデルを $\mathcal{P} = \{P_\theta; \theta \in \Theta\}$ と書く. ただし θ は母数で Θ は母数空間である. さらにある $\theta^* \in \Theta$ が存在して

$$\Pr^{X_1} = P_{\theta^*}$$

とする.

次に統計量 $\mathbf{T}(\mathbf{X})$ を

$$\mathbf{T} : (\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n)) \rightarrow (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$$

なる可測関数で θ に依存しないものとする. ただし $k \in \mathbb{N}$ である.

定義 6.12. $\mathbf{X} \sim P_\theta^{\otimes n}$ ($\theta \in \Theta$) とする. $\forall B \in \mathcal{B}(\mathbb{X}^n)$ に対して, 統計量 $\mathbf{T}(\mathbf{X})$ を与えたときの \mathbf{X} の条件付き確率

$$P_\theta^{\otimes n}(B | \mathbf{T})$$

が $\theta \in \Theta$ に無関係であるとき, \mathbf{T} は $\mathcal{P} = \{P_\theta; \theta \in \Theta\}$ に対する**十分統計量** (sufficient statistic) であるという.

定理 6.13. (Fisher-Neyman の因子分解定理) 統計的モデルを $\mathcal{P} = \{P_\theta; \theta \in \Theta\}$ とする. ただし θ は母数で Θ は母数空間とする. \mathbf{X} は確率空間 $(\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n), P_\theta^{\otimes n})$ 上の大きさが n の標本とし, 同時 p.d.f.(または p.m.f.) $p^{\mathbf{X}}(\mathbf{x} | \theta)$ ⁸を持つとする. このとき統計量 $\mathbf{T} : \mathbb{R}^n \rightarrow \mathbb{R}^k$ が統計的モデル

⁸離散型のときは同時 p.m.f. $p_\theta^{\mathbf{X}}(\mathbf{x} | \theta)$ をもつ.

\mathcal{P} に対する十分統計量であるための必要十分条件は, \mathbf{X} の同時 p.d.f.(または同時 p.m.f.) $p^{\mathbf{X}}(\mathbf{x}|\theta)$ が

$$p_{\theta}^{\mathbf{X}}(\mathbf{x}) = g_{\theta}\{\mathbf{T}(\mathbf{x})\}h(\mathbf{x}) \quad (\mathbf{x} \in \mathbb{X}^n) \quad (6.5)$$

の形で表されるときである. ここで $g_{\theta}\{\mathbf{T}(\mathbf{x})\}$ と $h(\mathbf{x})$ は非負値関数で $h(\mathbf{x})$ は θ に無関係な関数で, g_{θ} は \mathbf{T} を通してのみ \mathbf{x} に依存する.

Proof. \mathbf{X} が離散型の際のみの証明を与える. \mathbf{T} は十分統計量とする. \mathbf{T} の p.m.f. を $p_{\theta}^{\mathbf{T}}(\mathbf{t})$ ($\mathbf{t} \in \mathbb{R}^k$) と書く. $\mathbf{T} = \mathbf{t}$ が与えられたときの \mathbf{X} の条件付き p.m.f. を $p^{\mathbf{X}|\mathbf{T}}(\mathbf{x}|\mathbf{T} = \mathbf{t})$ とする. 仮定から \mathbf{T} は十分統計量なので $p^{\mathbf{X}|\mathbf{T}}(\mathbf{x}|\mathbf{T} = \mathbf{t})$ は θ に依存しない. 条件付き p.m.f. の定義から

$$p_{\theta}^{\mathbf{X}}(\mathbf{x}) = p_{\theta}^{\mathbf{T}}(\mathbf{t})p^{\mathbf{X}|\mathbf{T}}(\mathbf{x}|\mathbf{T} = \mathbf{t}) \quad (6.6)$$

となる. (6.6) において

$$h(\mathbf{x}) = p^{\mathbf{X}|\mathbf{T}}(\mathbf{x}|\mathbf{T} = \mathbf{t}), \quad g_{\theta}\{\mathbf{T}(\mathbf{x})\} = p_{\theta}^{\mathbf{T}}(\mathbf{t})$$

とおけば, (6.5) の形になる.

次に (6.5) と書けたときに \mathbf{T} は十分統計量であることを示す. まず

$$\begin{aligned} p_{\theta}^{\mathbf{T}}(\mathbf{t}) &= \sum_{\mathbf{y} \in \mathbb{X}^n; \mathbf{T}(\mathbf{y})=\mathbf{t}} p^{\mathbf{X}}(\mathbf{y}|\theta) \\ &= \sum_{\mathbf{y} \in \mathbb{X}^n; \mathbf{T}(\mathbf{y})=\mathbf{t}} g_{\theta}\{\mathbf{T}(\mathbf{y})\}h(\mathbf{y}) \\ &= g_{\theta}(\mathbf{t}) \sum_{\mathbf{y} \in \mathbb{X}^n; \mathbf{T}(\mathbf{y})=\mathbf{t}} h(\mathbf{y}) \end{aligned}$$

に注意する. (6.5) から $\mathbf{T} = \mathbf{t}$ を与えたときの \mathbf{X} の条件付き p.m.f. は

$$\begin{aligned} \frac{p^{\mathbf{X}}(\mathbf{x}|\theta)}{p^{\mathbf{T}}(\mathbf{t}|\theta)} &= \frac{g_{\theta}(\mathbf{t})h(\mathbf{x})}{g_{\theta}(\mathbf{t}) \sum_{\mathbf{y} \in \mathbb{X}^n; \mathbf{T}(\mathbf{y})=\mathbf{t}} h(\mathbf{y})} \\ &= \frac{h(\mathbf{x})}{\sum_{\mathbf{y} \in \mathbb{X}^n; \mathbf{T}(\mathbf{y})=\mathbf{t}} h(\mathbf{y})} \end{aligned}$$

となり, $p^{\mathbf{X}|\mathbf{T}}$ は θ に依存しない. よって \mathbf{T} は $\mathcal{P} = \{P_{\theta}; \theta \in \Theta\}$ に対する十分統計量である.

□

注意 6.14. Fisher-Neyman の因子分解定理の測度論的な証明には Radon-Nikodym の微分に関わる知識が必要になる. これについては定理 ?? を参照のこと.

例 6.15. $n \in \mathbb{N}$ とし, $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Ber}(\theta)$ ($0 \leq \theta \leq 1$) とする. $\mathbf{X} = (X_1, X_2, \dots, X_n)$ の同時 p.m.f. は

$$\begin{aligned} p_{\theta}^{\mathbf{X}}(\mathbf{x}) &= \theta^{\sum_{j=1}^n x_j} (1 - \theta)^{n - \sum_{j=1}^n x_j} \mathbb{1}_{\{0,1\}^n}(\mathbf{x}), \\ \mathbf{x} &= (x_1, x_2, \dots, x_n) \in \{0, 1\}^n, \\ \mathbb{1}_{\{0,1\}^n}(\mathbf{x}) &= \begin{cases} 1 & (\mathbf{x} \in \{0, 1\}^n) \\ 0 & (\text{その他の場合}) \end{cases} \end{aligned}$$

となる. したがって

$$T(\mathbf{x}) = \sum_{j=1}^n x_j, \quad g_{\theta}\{T(\mathbf{x})\} = \theta^{T(\mathbf{x})} (1 - \theta)^{n - T(\mathbf{x})}, \quad h(\mathbf{x}) = \mathbb{1}_{\{0,1\}^n}(\mathbf{x})$$

とおけば, 定理 6.13 から $T(\mathbf{X}) = X_1 + X_2 + \dots + X_n$ は

$$\mathcal{P} = \{p_{\theta}(x) = \theta^x (1 - \theta)^{1-x} \mathbb{1}_{\{0,1\}}(x); 0 \leq \theta \leq 1\}$$

の十分統計量であることがわかる. □

例 6.16. $\theta \in \mathbb{R}$ とし, $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{N}(\theta, 1)$ とする. $\mathbf{X} = (X_1, X_2, \dots, X_n)$ の同時 p.d.f. は

$$p_{\theta}^{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2} \sum_{j=1}^n (x_j - \theta)^2\right\} \quad \mathbf{x} = (x_1, x_2, \dots, x_n)$$

である. $T(\mathbf{x}) = n^{-1} \sum_{j=1}^n x_j$ とすると

$$p_{\theta}^{\mathbf{X}}(\mathbf{x}) = \exp\left\{nT(\mathbf{x})\theta - \frac{n\theta^2}{2}\right\} \times \left(\frac{1}{2\pi}\right)^{n/2} \times \exp\left\{-\frac{1}{2} \sum_{j=1}^n x_j^2\right\}$$

となる. よって (6.5) で

$$g_{\theta}\{T(\mathbf{x})\} = \exp\left\{nT(\mathbf{x})\theta - \frac{n\theta^2}{2}\right\}, \quad h(\mathbf{x}) = \left(\frac{1}{2\pi}\right)^{n/2} \exp\left\{-\frac{1}{2} \sum_{j=1}^n x_j^2\right\}$$

とおけば, 定理 6.13 から $T(\mathbf{X}) = (X_1 + X_2 + \dots + X_n)/n$ は $\mathcal{P} = \{\text{N}(\theta, 1); \theta \in \mathbb{R}\}$ の十分統計量であることがわかる. □

例 6.17. $n \in \mathbb{N}$ ($n \geq 2$) とし, $\mu \in \mathbb{R}$, $0 < \sigma < \infty$ とする. $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{N}(\mu, \sigma^2)$ とする. $\mathbf{X} = (X_1, X_2, \dots, X_n)$ の同時 p.d.f. は

$$p_{\theta}^{\mathbf{X}}(\mathbf{x}) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{j=1}^n (x_j - \mu)^2\right\}$$

となる. ただし $\theta = (\mu, \sigma^2)$ である. ここで

$$\bar{x}_n = \frac{1}{n} \sum_{j=1}^n x_j, \quad s_n^2 = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2$$

とおけば

$$p_{\theta}^{\mathbf{X}}(\mathbf{x}) = \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left\{ -\frac{n}{2\sigma^2} (s_n^2 + (\bar{x}_n - \mu)^2) \right\}$$

と書き直せる. (6.5) において

$$T(\mathbf{x}) = (\bar{x}_n, s_n^2), \quad g_{\theta}\{T(\mathbf{x})\} = \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left\{ -\frac{n}{2\sigma^2} (s_n^2 + (\bar{x}_n - \mu)^2) \right\},$$

$$h(\mathbf{x}) = 1$$

とすれば, 定理 6.13 から

$$\mathbf{T}(\mathbf{X}) = \left(\frac{1}{n} \sum_{j=1}^n X_j (=:\bar{X}), \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^2 \right)$$

は $\mathcal{P} = \{N^{\otimes n}(\mu, \sigma^2); \theta = (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)\}$ の十分統計量である. ただし, $N^{\otimes n}(\mu, \sigma^2)$ は $N(\mu, \sigma)$ の n 個の直積分布である. \square

6.5 統計量の最小十分性と完備性

6.5.1 十分統計量の最小性

与えられた統計的モデル \mathcal{P} に対して, たくさんの十分統計量が存在する. \mathbf{S} と \mathbf{T} が $P \in \mathcal{P}$ に対する十分統計量で, S の値域上で定義された可測関数 ψ が存在して

$$\mathbf{T} = \psi(\mathbf{S})$$

と表現できたとする. すると

$$\sigma(\mathbf{T}) \subset \sigma(\mathbf{S})$$

となる. このことから \mathbf{T} は統計的モデル \mathcal{P} の情報を失くことなく, データの情報を S より縮約していることがわかる. したがって T は S よりも有用な統計量となる. この観点を定式化してみよう.

定義 6.18. すべての $P \in \mathcal{P}$ に対して, $P(A) = 0$ となる事象 A を除いて, ある命題 M が成立しているとき

$$M \quad \text{a.s. } \mathcal{P}$$

と記すことにする.

定義 6.19. (最小十分性) \mathcal{P} を統計的モデルとし, 統計量 \mathbf{T} を $P \in \mathcal{P}$ に対する十分統計量とする. 十分統計量 T は $P \in \mathcal{P}$ に関して最小であるとは, 任意の他の十分統計量 S に対して可測関数 ψ が存在して

$$\mathbf{T} = \psi(\mathbf{S}), \quad \text{a.s. } \mathcal{P}$$

と表現できるときをいう. もちろん ψ は \mathbf{S} ごとに定まればよい.

注意 6.20. $\mathcal{P} = \{P_\theta; \theta \in \Theta\}$ を統計的モデルとする. 2つの母数 $\theta_1, \theta_2 (\in \Theta)$ の疑似測度を

$$\rho(\theta_1, \theta_2) = \sup_{B \in \mathcal{B}(\mathbb{X})} |P_{\theta_1}(B) - P_{\theta_2}(B)|$$

に関して Θ が可分⁹であるとき, \mathcal{P} に対する最小十分統計量が存在することが知られている. 証明は [51, pp.78-81] を参照のこと. \square

例 6.21. $\mathcal{P} = \{P \text{ は開区間 } (\theta, \theta + 1) \text{ 上の一様分布; } \theta \in \mathbb{R}\}$ とする. $n \geq 2 (\in \mathbb{N})$ とし

$$X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} P \in \mathcal{P}$$

とする. Lebesgue 測度に関する $\mathbf{X} := (X_1, X_2, \dots, X_n)$ の同時 p.d.f. は

$$\begin{aligned} p_\theta^{\mathbf{X}}(\mathbf{x}) &= \prod_{j=1}^n \mathbb{1}_{(\theta, \theta+1)}(x_j) = \mathbb{1}_{(x_{(n)}-1, x_{(1)})}(\theta), \\ \mathbf{x} &= (x_1, x_2, \dots, x_n) \in \mathbb{R}^n \end{aligned}$$

となる. ただし $x_{(1)}, x_{(n)}$ は x_1, x_2, \dots, x_n の最小値と最大値である.

$$X_{(1)} = \min\{X_1, X_2, \dots, X_n\}, \quad X_{(n)} = \max\{X_1, X_2, \dots, X_n\}$$

とおくと Fisher-Neyman の因子分解定理から $\mathbf{T} = (X_{(1)}, X_{(n)})$ は $P \in \mathcal{P}$ に対する十分統計量である.

次に, 十分統計量 \mathbf{T} が最小性もみたすことを示す.

$$\begin{aligned} \theta < x_j < \theta + 1 \quad (\forall j \in \{1, 2, \dots, n\}) \\ \Leftrightarrow \theta < x_{(1)} < x_{(n)} < \theta + 1 \end{aligned} \tag{6.7}$$

となることに注意する. $\mathbf{X} = \mathbf{x}$ を観測したとき

$$x_{(1)} = \sup\{\theta \in \mathbb{R}; p(\mathbf{x}|\theta) > 0\}, \quad x_{(n)} = 1 + \inf\{\theta \in \mathbb{R}; p(\mathbf{x}|\theta) > 0\}$$

となる. ここで統計量 \mathbf{S} は $P \in \mathcal{P}$ に対して十分であると仮定する. Fisher-Neyman の因子分解定理から可測関数 g_θ と h が存在して

$$p_\theta^{\mathbf{X}}(\mathbf{x}) = g_\theta(\mathbf{S}(\mathbf{x}))h(\mathbf{x})$$

⁹ Θ の可算無限集合 C が存在して, 任意の θ と ϵ に対して, $\theta_c \in C$ をうまく選ぶと $\rho(\theta, \theta_c) < \epsilon$ とできることである.

と書ける. $A^c := \{\mathbf{x} \in \mathbb{R}^n; h(\mathbf{x}) > 0\}$ 上で (6.7) が成立するので, ある可測関数 ψ が存在して A 上で

$$\mathbf{T} := (x_{(1)}, x_{(n)}) = \psi(\mathbf{S}(\mathbf{x}))$$

と表現できる. さらに $\Pr(\mathbf{X} \in A) = 0$ なので \mathbf{T} は最小十分統計量となる. ここでは, 最小十分統計量であることを直接的に確認したが, 次の定理 6.22 の判定条件を用いる簡単に最小性の確認をできる. \square

定理 6.22. $(\mathbb{X}, \mathcal{B})$ を標本空間とし, それ上の統計的モデル $\mathcal{P} = \{P_\theta; \theta \in \Theta\}$ は測度 μ で支配されており, p.d.f. $p(x|\theta) = g_\theta(T(x))h(x)$ を持つとする. 任意の x, y に対して, 関数

$$\theta \mapsto \frac{p_\theta(x)}{p_\theta(y)}$$

が定数ならば, $T(x) = T(y)$ が成り立つとき, T は最小十分統計量となる.

Proof. 厳密な証明は測度論の知識を用いることになる. ここでは証明の基本的な考え方を述べるだけにする. \tilde{T} を任意の十分統計量とする. すると Fisher-Neyman の因子分解定理から

$$p_\theta(x) = \tilde{g}_\theta(\tilde{T}(x))\tilde{h}(x)$$

と書ける. かりに T は \tilde{T} の関数でないとすると, あるデータ x と y があって,

$$\tilde{T}(x) = \tilde{T}(y) \quad \text{かつ} \quad T(x) \neq T(y) \tag{6.8}$$

とできること¹⁰になる. しかし

$$\frac{p_\theta(x)}{p_\theta(y)} = \frac{\tilde{g}_\theta(\tilde{T}(x))\tilde{h}(x)}{\tilde{g}_\theta(\tilde{T}(y))\tilde{h}(y)} = \frac{\tilde{h}(x)}{\tilde{h}(y)}$$

となり, 関数 $\theta \mapsto \frac{p(x|\theta)}{p(y|\theta)}$ は定数となる. よって, T が最小十分統計量であるという仮定より $T(x) = T(y)$ となり, (6.8) と矛盾する. したがって, ある関数 f があって, $\tilde{T} = f(T)$ とかける. \square

¹⁰ $\tilde{T} = f(T)$ と書けるとき,

$$T(x) = T(y) \implies \tilde{T}(x) = \tilde{T}(y)$$

なので, $\tilde{T} = f(T)$ と書けない場合は, ある x, y があって

$$T(x) = T(y) \quad \text{かつ} \quad \tilde{T}(x) \neq \tilde{T}(y)$$

となる.

例 6.23. $(\mathbb{X}, \mathcal{B})$ を標本空間とし, $\mathcal{P} = \{P_\theta; \theta \in \Theta\}$ を $(\mathbb{X}, \mathcal{B})$ 上の d 母数指数型分布族とする. すなわち p.d.f. $p_\theta(x) (\theta \in \Theta)$ が

$$p_\theta(x) = \exp\{\eta(\theta)T(x) - \kappa(\theta)\}h(x)$$

と書ける. ただし $\kappa: \Theta \rightarrow \mathbb{R}^d$ であり, $T \in \mathbb{R}^d$ で, $(\cdot|\cdot)$ は \mathbb{R}^d の Euclid 内積である. Fisher-Neyman の因子分解定理から T は統計的モデル \mathcal{P} に対する十分統計量となる. データ x と y は, 関数

$$\theta \mapsto \frac{p_\theta(x)}{p_\theta(y)}$$

が定数となるものとする. するとある θ に依存しない定数 c (x と y に依存してよい) があって

$$(\eta(\theta)|T(x)) = (\eta(\theta)|T(y)) + c$$

と書けることになる. 任意の点 $\theta_0, \theta_1 \in \Theta (\theta_0 \neq \theta_1)$ に対して

$$\begin{aligned} ([\eta(\theta_0) - \eta(\theta_1)]|T(x)) &= ([\eta(\theta_0) - \eta(\theta_1)]|T(y)) \\ \Leftrightarrow ([\eta(\theta_0) - \eta(\theta_1)]|T(x) - T(y)) &= 0 \end{aligned}$$

を得る. $\text{span}\left\{[\eta(\theta_0) - \eta(\theta_1)] \in \mathbb{R}^d : \theta_0, \theta_1 \in \Theta\right\} = \mathbb{R}^d$ のとき, $T(x) = T(y)$ となり, T は最小十分統計量となる. \square

例 6.24. $n \geq 2 (n \in \mathbb{N})$ と $\theta \in \mathbb{R}$ とし, X_1, X_2, \dots, X_n は独立同一に分布する確率変数列で, 共通の p.d.f.

$$p_\theta(x) = \frac{1}{2}e^{-|x-\theta|} \quad (x \in \mathbb{R})$$

を持つとする. $\mathbf{X} := (X_1, X_2, \dots, X_n)$ の同時 p.d.f. $p_{\mathbf{X}}(\mathbf{x}|\theta)$ は

$$p_{\mathbf{X}}(\mathbf{x}|\theta) = \frac{1}{2^n} \exp\left\{-\sum_{i=1}^n |x_i - \theta|\right\} \quad (\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n)$$

となる. $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ を順序統計量とする. Fisher-Neyman の因子分解定理から $\mathbf{T} = (X_{(1)}, X_{(2)}, \dots, X_{(n)})$ は十分統計量となる. データを $x = (x_1, x_2, \dots, x_n)$ と $y = (y_1, y_2, \dots, y_n)$ としたとき, $\sum_{i=1}^n |x_i - \theta|$ と $\sum_{i=1}^n |y_i - \theta|$ の差は θ に依存しない. この二つの θ の関数は区分的に連続で, 不連続点は $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ と $y_{(1)}, y_{(2)}, \dots, y_{(n)}$ である. 関数

$$\theta \mapsto \sum_{i=1}^n |x_i - \theta| - \sum_{i=1}^n |y_i - \theta|$$

が θ に関して定数であるときは, ジャンプの点が一致することである. したがって, $x_{(i)} = y_{(i)} (i = 1, 2, \dots, n)$ となるので, \mathbf{T} は最小十分であることがわかる.

6.5.2 十分統計量の完備性

定義 6.25. $k \in \mathbb{R}$ とし, $\mathbb{T} \subset \mathbb{R}^k$ とする. \mathcal{P} を統計的モデルとし, $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} P \in \mathcal{P}$ とする. $\mathbf{T}(\mathbf{X}) (\in \mathbb{T})$ を $P \in \mathcal{P}$ に対する十分統計量とする. ただし, $\mathbf{X} = (X_1, X_2, \dots, X_n)^\top$ と書いた. 統計量 \mathbf{T} は $P \in \mathcal{P}$ に対して完備であるとは, 任意の可測関数 $f: \mathbb{T} \rightarrow \mathbb{R}$ に対して

$$E[f(\mathbf{T})] = 0 (\forall P \in \mathcal{P}) \Rightarrow f(\mathbf{T}) = 0, \text{ a.s. } \mathcal{P}$$

が成り立つときをいう.

次の命題は, 指数型分布族の十分統計量は完備であることを主張している. 本講義録では証明していない Fubini の定理と Laplace 変換の一意性を用いてこの命題を証明することができる.

命題 6.26. $k, n \in \mathbb{N}$ かつ $n > k$ とする. $\boldsymbol{\eta} \in \mathbb{R}^k, \mathbb{X} \subset \mathbb{R}^n$ を空でない部分集合とし

$$p_{\boldsymbol{\eta}}(\mathbf{x}) = \exp\{\boldsymbol{\eta}^\top \mathbf{T}(\mathbf{x})\} h(\mathbf{x}) \quad (\mathbf{x} \in \mathbb{X}^n)$$

とする. ただし

$$\mathcal{E} := \left\{ \boldsymbol{\eta} \in \mathbb{R}^k; \kappa(\boldsymbol{\eta}) := \int_{\mathbb{X}^n} \exp\{\boldsymbol{\eta}^\top \mathbf{T}(\mathbf{x})\} h(\mathbf{x}) d\mathbf{x} < \infty \right\}$$

は \mathbb{R}^n の開集合を含み

$$\mathbf{T}: \mathbb{X}^n \rightarrow \mathbb{R}^k, \quad h: \mathbb{X}^n \rightarrow \mathbb{R}, \quad \kappa: \mathcal{E} \rightarrow \mathbb{R}$$

は可測関数とする. $\mathbf{X} \sim P (P \in \mathcal{P} := \{p_{\boldsymbol{\eta}}(\mathbf{x}); \boldsymbol{\eta} \in \mathcal{E}\})$ としたとき, $\mathbf{T}(\mathbf{X})$ は $P \in \mathcal{P}$ に対する完備かつ十分統計量である.

Proof. Fisher-Neyman の因子分解定理から \mathbf{T} は $P \in \mathcal{P}$ に対する十分統計量であることは明らか. つぎに $f: \mathbb{R}^k \rightarrow \mathbb{R}$ を可測関数とし

$$E[f(\mathbf{T})] = 0 \quad (\forall \boldsymbol{\eta} \in \mathcal{E})$$

とする. $E[|f(\mathbf{T})|] < \infty$ なので Fubini の定理から

$$\begin{aligned} 0 &= \int_{\mathbb{X}} f(\mathbf{T}(\mathbf{x})) \exp\{\boldsymbol{\eta}^\top \mathbf{T}(\mathbf{x}) - \kappa(\boldsymbol{\eta})\} h(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbf{x}: \mathbf{T}(\mathbf{x})=\mathbf{t}} h(\mathbf{x}) \left\{ \int_{\mathbb{T}} f(\mathbf{t}) \exp\{\boldsymbol{\eta}^\top \mathbf{t} - \kappa(\boldsymbol{\eta})\} d\mathbf{t} \right\} d\mathbf{m}_{n-k} \end{aligned}$$

と書ける. ただし $\mathbb{T} = \{\mathbf{T}(\mathbf{x}) \in \mathbb{R}^k; \mathbf{x} \in \mathbb{X}\}$ で, \mathbf{m}_{n-k} は \mathbb{R}^{n-k} 上の Lebeague 測度である. よって

$$\int_{\mathbb{T}} f(\mathbf{t}) \exp\{\boldsymbol{\eta}^\top \mathbf{t} - \kappa(\boldsymbol{\eta})\} d\mathbf{t} = 0 \quad \text{a.s. } \mathbf{m}_{n-k}$$

となる. $\boldsymbol{\eta}_0 \in \mathcal{E}$ の内点とし, $\epsilon > 0$ を十分小さく取ると

$$\begin{aligned} \int f_+(\mathbf{t}) \exp\{\boldsymbol{\eta}^\top \mathbf{t}\} d\mathbf{t} &= \int f_-(\mathbf{t}) \exp\{\boldsymbol{\eta}^\top \mathbf{t}\} d\mathbf{t} \\ (\forall \boldsymbol{\eta} \in N(\boldsymbol{\eta}_0) &:= \{\boldsymbol{\eta} \in \mathcal{R}^k; |\boldsymbol{\eta} - \boldsymbol{\eta}_0|_{2,k} < \epsilon\}) \end{aligned}$$

となる. ここで

$$f_+(\mathbf{t}) := \max\{f(\mathbf{t}), 0\}, \quad f_-(\mathbf{t}) := \max\{-f(\mathbf{t}), 0\}$$

で, $|\cdot|_{2,k}$ は \mathbb{R}^k 上の Euclid の距離である. 特に

$$\int f_+(\mathbf{t}) \exp\{\boldsymbol{\eta}_0^\top \mathbf{t}\} d\mathbf{t} = \int f_-(\mathbf{t}) \exp\{\boldsymbol{\eta}_0^\top \mathbf{t}\} d\mathbf{t} =: c$$

となる. $c = 0$ のとき, $f(\mathbf{t}) = 0$ (a.s.) となる. $c > 0$ のとき, Laplace 変換の一意性¹¹から

$$f_+(\mathbf{t}) = f_-(\mathbf{t}), \quad \text{a.s.}$$

となる. よって $f(\mathbf{t}) = f_+(\mathbf{t}) - f_-(\mathbf{t}) = 0$ がわかる. □

注意 6.27. 完備十分統計量は最小十分統計量であることが知られている. しかし最小十分統計量で完備でないものの存在が知られている. □

6.6 補遺: Laplace 変換の一意性

定義 6.28. $[0, \infty)$ 上に集中した分布 P について

$$f(s) = \int_{[0, \infty)} e^{-sx} dP(x)$$

を P の Laplace **変換**という. また, 非負値確率変数 X の Laplace 変換とは

$$f(s) = E[e^{-sX}] = \int_{[0, \infty)} e^{-sx} dP(x) \quad (s \geq 0)$$

のことである.

¹¹Laplace 変換の一意性の証明は次章の定理 A.8 を参照のこと.

補題 6.29. 区間 $[0, 1]$ に集中した分布の分布関数 G_1, G_2 に対して

$$\int_0^1 x^n dG_1(x) = \int_0^1 x^n dG_2(x) \quad (n = 0, 1, 2, \dots) \quad (6.9)$$

が成り立てば

$$G_1 = G_2$$

である.

Proof. Weierstrass の多項式近似定理により, (A.15) が成り立てば, 任意の $[0, 1]$ 上の連続関数 f に対して

$$\int_0^1 f(x) dG_1(x) = \int_0^1 f(x) dG_2(x)$$

が成り立つ.

いま, $0 \leq a < 1$ と $n \geq 1$ に対して

$$f_n(x) = \begin{cases} 1 & (0 \leq x \leq a), \\ n(x-a) & (a \leq x \leq a+1/n), \\ 0 & (n+1/n \leq x \leq 1) \end{cases}$$

と定める. すると $G_1(0) = G_2(0) = 0$ より

$$\begin{aligned} G_1(a) &= \int_0^a dG_1 = \lim_{n \rightarrow \infty} \int_0^1 f_n(x) dG_1(x) = \lim_{n \rightarrow \infty} \int_0^1 f_n(x) dG_2(x) \\ &= \int_0^a dG_2 = G_2(a) \end{aligned}$$

を得る. また $G_1(1) = G_2(1)$ だから, 全ての $a \in [0, 1]$ に対して $G_1(a) = G_2(a)$ がわかる. \square

定理 6.30. $[0, \infty)$ に集中した確率をもつ 2 つの分布関数 F_1, F_2 に対して, それぞれの Laplace 変換を f_1, f_2 とする. このとき, f_1 と f_2 が一致すれば, $F_1 = F_2$ である.

Proof. $F : [0, \infty) \rightarrow \mathbb{R}$ を確率分布関数とする. $x = e^{-t}$, $G(x) = 1 - F(t)$ とすると

$$\int_0^\infty e^{-\lambda t} dF(t) = \int_0^1 x^\lambda dG(x) \quad (\lambda \geq 0)$$

となる. よって $f_1 = f_2$ ならば,

$$\int_0^1 x^n dG_1(x) = \int_0^1 x^n dG_2(x) \quad (n = 0, 1, 2, \dots)$$

となる. あとは補題 A.7 を用いれば, 定理は証明される. \square

6.7 章末注釈と参考文献

第 6.1 節は [1, pp.69 – 71] を借用した. 補題 6.10 は [17, pp.28 – 30] を借用した.

6.8 演習問題

演習問題 6.1. $n \in \mathbb{N}$, $0 < \theta < 1$ とし,

$$p_\theta(x) = \Pr(X = x) = \begin{cases} \binom{n}{x} \theta^x (1 - \theta)^{n-x} & (x = 0, 1, \dots, n) \\ 0 & (\text{その他の場合}) \end{cases},$$

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}, \quad 0! = 1$$

とする. 母数モデル

$$\mathcal{P} = \{p_\theta; 0 < \theta < 1\}$$

を考える.

(1) $X \sim \text{Bino}(n, \theta)$ のとき

$$m^X(t) = E[e^{tX}] = \{\theta e^t + 1 - \theta\}^n \quad (t \in \mathbb{R})$$

を示せ.

(2) $\left. \frac{d}{dt} m^X(t) \right|_{t=0}$ を計算することで $E[X]$ を求めよ.

(3) $\left. \frac{d^2}{dt^2} m^X(t) \right|_{t=0}$ を計算することで $E[X^2]$ を求めよ.

(4) $\text{Var}[X]$ を計算せよ.

(5) $X \sim p_\theta$ とする. $0 < \forall \theta < 1$ に対して, $E_\theta[\dot{\ell}_1(X|\theta)]$ を求めよ. ただし, $X = x$ を観測したとき, $\ell_1(\theta|x) = \log p_\theta(x)$ とし, $\dot{\ell}_1(\theta|x) = \frac{d}{d\theta} \ell_1(\theta|x)$ と書いた.

(6) $E_\theta[\{\dot{\ell}_1(\theta|X)\}^2]$ を計算せよ.

(7) $-E_\theta[\ddot{\ell}_1(\theta|X)]$ を計算せよ. ただし, $\ddot{\ell}_1(\theta|x) = \frac{d^2}{d\theta^2} \ell_1(\theta|x)$ と書いた.

演習問題 6.2. $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Po}(\theta)$ ($\theta > 0$) とする. $S := X_1 + X_2 + \dots + X_n$ は $\mathcal{P} = \{\text{Po}(\theta); \theta > 0\}$ の十分統計量であることを示せ.

演習問題 6.3.

第7章 推定

この章では、点推定法における推定量の導出原理と推定量の精度評価についての考え方の説明を行う。第7.1節では、モーメント法について説明する。第7.2節では、最尤法について説明する。最尤推定量の漸近分布は、統計的モデルが指数型分布族の場合について示している。第7.3では、推定量のひとつの性質である不偏推定概念を説明し、それに関する情報不等式を与える。

7.1 モーメント法

この方法で得られた推定量は最適性を持たないことがある。しかし求めるのが容易であるので、陽には式が与えられない別の推定量の値を繰り返し計算で求めるときの初期値として用いることができる。

$\mathbb{X} \subset \mathbb{R}$ を空でない部分集合とする。 $d, n \in \mathbb{N}$ とし、統計的実験を $(\mathbb{X}^n, \{P_\theta = P_\theta; \theta \in \Theta \subset \mathbb{R}^d\})$ を考え

$$X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} P_{\theta^*} \quad (\theta^* \in \Theta)$$

とする。ただし、 P_{θ^*} と P_θ は、それぞれ測度空間 $(\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n))$ と $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ 上の分布である。

$n \geq d$ とする。 $1 \leq j \leq d$ に対して

$$\alpha_j := \alpha_j(\theta) := E[X_1^j], \quad \hat{\alpha}_j := \frac{1}{n} \sum_{\ell=1}^n X_\ell^j$$

と定める。

定義 7.1. $n \geq d$ とし $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} P_{\theta^*} (\theta^* \in \Theta)$ とする。このとき θ^* のモーメント法推定量 $\hat{\theta}_n$ とは次をみたす解 (存在すれば) である。

$$\alpha_1(\hat{\theta}_n) = \hat{\alpha}_1, \alpha_2(\hat{\theta}_n) = \hat{\alpha}_2, \dots, \alpha_d(\hat{\theta}_n) = \hat{\alpha}_d \quad (7.1)$$

注意 7.2. $n \geq 2$ とする。 $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N(\mu^*, (\sigma^*)^2)$ とする。ただし $\theta^* := (\mu^*, \sigma^*) \in \mathbb{R} \times (0, \infty) =: \Theta$ で (μ^*, σ^*) は未知とする。このとき

$$\alpha_1 = E[X_1] = \mu^*, \quad \alpha_2 = E[X_1^2] = \text{Var}[X_1] + \{E[X_1]\}^2 = (\sigma^*)^2 + (\mu^*)^2$$

である. したがって $\hat{\theta}_n = (\hat{\mu}_n, \hat{\sigma}_n)$ とすれば

$$\hat{\mu}_n = \frac{1}{n} \sum_{\ell=1}^n X_\ell, \quad \hat{\sigma}_n^2 + \{\hat{\mu}_n\}^2 = \frac{1}{n} \sum_{\ell=1}^n X_\ell^2$$

を解くと (μ^*, σ^*) のモーメント法推定量は

$$\hat{\mu}_n = \frac{1}{n} \sum_{\ell=1}^n X_\ell, \quad \hat{\sigma}_n = \sqrt{\frac{1}{n} \sum_{\ell=1}^n (X_\ell - \hat{\mu}_n)^2}$$

となる. □

定理 7.3. $d, n \in \mathbb{N}$ は $n \geq d$ とする.

$$X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} P_{\theta^*}(\theta^* \in \Theta \subset \mathbb{R}^d), \quad E[|X_1|^{2d}] < \infty$$

とする. (7.1) で定義した $\alpha_j (j = 1, 2, \dots, d)$ の逆写像 α_j^{-1} が存在して, θ^* の近傍で全微分可能とする. このとき適当な条件のもとで以下が成立する.

(1) $\Pr_{\theta^*}(\hat{\theta}_n \text{ は存在}) \xrightarrow{n \rightarrow \infty} 0$ である.

(2) $\forall \epsilon > 0$ に対して

$$\Pr_{\theta^*}(|\hat{\theta}_n - \theta^*|_{2,d} \geq \epsilon) \xrightarrow{n \rightarrow \infty} 0$$

である. ただし $|\cdot|_{2,d}$ は \mathbb{R}^d の Euclid ノルムである.

(3) $\sqrt{n}(\hat{\theta}_n - \theta^*) \rightsquigarrow N_d(\mathbf{0}, \Sigma)$ である. ただし

$$\begin{aligned} \Sigma &= \mathbf{G}(\theta^*) \mathbf{E}_{\theta^*}[\mathbf{Y}\mathbf{Y}^\top] \mathbf{G}^\top(\theta^*), \\ \mathbf{G}(\theta) &= (\dot{\mathbf{g}}_1(\theta), \dot{\mathbf{g}}_2(\theta), \dots, \dot{\mathbf{g}}_d(\theta)), \\ \dot{\mathbf{g}}_j(\theta) &= \left(\frac{\partial \alpha_j^{-1}}{\partial \theta_1}(\theta), \frac{\partial \alpha_j^{-1}}{\partial \theta_2}(\theta), \dots, \frac{\partial \alpha_j^{-1}}{\partial \theta_d}(\theta) \right)^\top \quad (j = 1, 2, \dots, d), \\ \mathbf{Y} &= (X_1, X_1^2, \dots, X_1^d)^\top \end{aligned}$$

が成立する.

Proof. 証明は加筆予定である. □

7.2 最尤法

\mathbb{X} を \mathbb{R} の空でない部分集合とし, $d, n \in \mathbb{N}$ とする. 統計的実験 $(\mathbb{X}^n, \{p_{\theta}(x); \theta \in \Theta\})$ を考える. ただし $p(x|\theta)$ は p.d.f. (または p.m.f.) で $\Theta \subset \mathbb{R}^d$ である. $\theta^* \in \Theta$ とし

$$X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} p_{\theta^*}(x)$$

とする.

定義 7.4. $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ を観測したとき θ の **尤度関数** $\text{lik}_n(\theta|\mathbf{x})$ を

$$\text{lik}_n(\theta|\mathbf{x}) = \prod_{j=1}^n p_{\theta}(x_j)$$

で定義し, **対数尤度** $\ell_n(\theta|\mathbf{x})$ を

$$\ell_n(\theta|\mathbf{x}) = \log \text{lik}_n(\theta|\mathbf{x})$$

で定義する. ただし, $\mathbf{x} = (x_1, x_2, \dots, x_n)$ である.

注意 7.5. 尤度関数は

$$\text{lik}_n(\cdot|\mathbf{x}) : \Theta \ni \theta \mapsto \text{lik}_n(\theta|\mathbf{x}) \in [0, \infty)$$

である. □

記法: 関数 $g(x)$ の最大値を取る点を表す集合を

$$\arg \max_{x \in \mathbb{R}} g(x)$$

と書く. たとえば $g(x) = -(x-1)^2$ のとき

$$\arg \max_{x \in \mathbb{R}} g(x) = \{1\}$$

となる. $g(x) = \sin x$ のとき

$$\arg \max_{0 \leq x \leq 4\pi} g(x) = \{\pi/2, 5\pi/2\}$$

となる.

定義 7.6. $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ を観測したとき θ^* の**最尤推定値** (maximum likelihood estimate) を $\text{lik}_n(\theta|\mathbf{x})$ を最大にする値 $\hat{\theta}_n(\mathbf{x})$ で定義する. すなわち

$$\hat{\theta}_n(\mathbf{x}) \in \arg \max_{\theta \in \Theta} \text{lik}_n(\theta|\mathbf{x})$$

である. $\mathbf{x} = (x_1, x_2, \dots, x_n)$ に $\mathbf{X} = (X_1, X_2, \dots, X_n)$ を代入したものの $\hat{\theta}_n(\mathbf{X})$ を θ^* の**最尤推定量** (maximum likelihood estimator=m.l.e.) という.

例 7.7. $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Ber}(\theta^*)$ とする. ただし $(0, 1) =: \Theta \ni \theta^*$ は未知とする. すなわち

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

である. $X_j = x_j (j = 1, 2, \dots, n)$ を観測したとき

$$\text{lik}_n(\theta|\mathbf{x}) = \prod_{j=1}^n p(x_j|\theta) = \prod_{j=1}^n \theta^{x_j} (1 - \theta)^{1-x_j} = \theta^{t_n} (1 - \theta)^{n-t_n}$$

となる. ただし, $\mathbf{x} = (x_1, x_2, \dots, x_n)$ と $t_n = \sum_{j=1}^n x_j$ である. よって対数尤度は

$$\ell_n(\theta|\mathbf{x}) = t_n \log \theta + (n - t_n) \log(1 - \theta) \quad (0 < \theta < 1)$$

となる. このことから, $0 < t_n < n$ のとき

$$\frac{t_n}{n} \in \arg \max_{\theta \in (0,1)} \ell_n(\theta|\mathbf{x})$$

がわかる. したがって, $0 < t_n < n$ のとき θ^* の最尤推定値は $\hat{\theta}_n = \frac{\sum_{j=1}^n x_j}{n}$ となり, $t_n = 0$ または $t_n = n$ のとき, 最尤推定値は存在しない. \square

注意 7.8. 定義 7.6 は [3] の流儀に従った. 一方, 母数空間 Θ の閉包 $\text{cl}(\Theta)$ を考え, $\text{cl}(\Theta)$ に尤度関数 lik_n の定義域を拡張して

$$\hat{\boldsymbol{\theta}}_n(\mathbf{x}) \in \arg \min_{\boldsymbol{\theta} \in \text{cl}(\Theta)} \text{lik}_n(\boldsymbol{\theta}|\mathbf{x})$$

と最尤推定値を定義する流儀もある. [29, 41] を参照のこと. 例 ?? を踏まえるとこちらの定義のが数学的には扱いやすくなるようにみえる.

例 7.9. $n \geq 2 (n \in \mathbb{N})$ とする. $\boldsymbol{\theta}^* = (\mu^*, \sigma^*) \in \mathbb{R} \times (0, \infty)$ とし

$$X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{N}(\mu^*, (\sigma^*)^2)$$

とする. $X_j = x_j (j = 1, 2, \dots, n)$ を観測したとき, 尤度関数は

$$\begin{aligned} \text{lik}_n(\mu, \sigma|\mathbf{x}) &= \prod_{j=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x_j - \mu)^2}{\sigma^2}\right\} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{j=1}^n (x_j - \mu)^2\right\} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sigma^n} \exp\left\{-\frac{ns_n^2}{2\sigma^2}\right\} \exp\left\{-\frac{n(\bar{x}_n - \mu)^2}{2\sigma^2}\right\} \end{aligned} \quad (7.2)$$

となる. ただし

$$\bar{x}_n = \frac{1}{n} \sum_{j=1}^n x_j, \quad s_n^2 = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x}_n)^2, \quad \sum_{j=1}^n (x_j - \bar{x}_n)^2 \neq 0$$

である. (7.2) の最後の等号は

$$\sum_{j=1}^n (x_j - \mu)^2 = ns_n^2 + n(\bar{x}_n - \mu)^2 \quad (7.3)$$

からわかる. 対数尤度は

$$\ell_n(\mu, \sigma | \mathbf{x}) = -n \log \sigma - \frac{ns_n^2}{2\sigma^2} - \frac{n(\bar{x}_n - \mu)^2}{2\sigma^2} + (\text{定数項})$$

となる. よって

$$\begin{cases} \frac{\partial \ell_n}{\partial \mu}(\mu, \sigma | \mathbf{x}) = -\frac{n(\bar{x}_n - \mu)}{2\sigma^2} = 0 \\ \frac{\partial \ell_n}{\partial \sigma}(\mu, \sigma | \mathbf{x}) = -\frac{n}{\sigma} + \frac{ns_n^2}{\sigma^3} + \frac{n(\bar{x}_n - \mu)^2}{\sigma^3} = 0 \end{cases}$$

を解くと

$$\mu = \bar{x}_n, \quad \sigma = \sqrt{\frac{1}{n} \sum_{j=1}^n (x_j - \bar{x}_n)^2}$$

となる. $\ell_n(\mu, \sigma | \mathbf{x})$ の Hessian を求める.

$$\mathbf{H} := \begin{pmatrix} \frac{\partial^2 \ell_n}{\partial \mu^2}(\bar{x}_n, s_n | \mathbf{x}) & \frac{\partial^2 \ell_n}{\partial \mu \partial \sigma}(\bar{x}_n, s_n | \mathbf{x}) \\ \frac{\partial^2 \ell_n}{\partial \sigma \partial \mu}(\bar{x}_n, s_n | \mathbf{x}) & \frac{\partial^2 \ell_n}{\partial \sigma^2}(\bar{x}_n, s_n | \mathbf{x}) \end{pmatrix} = \begin{pmatrix} -\frac{1}{2s_n^2} & 0 \\ 0 & -\frac{2n}{s_n^2} \end{pmatrix} \quad (7.4)$$

より, $-\mathbf{H}$ は正定値行列となる. したがって, 関数

$$\Theta \ni (\mu, \sigma) \mapsto \ell_n(\mu, \sigma | \mathbf{x})$$

は $(\mu, \sigma) = (\bar{x}_n, s_n)$ で最大となる. 以上の議論から (μ^*, σ^*) の最尤推定量は

$$\hat{\mu}_n = \frac{1}{n} \sum_{j=1}^n X_j =: \bar{X}_n, \quad \hat{s}_n = \sqrt{\frac{1}{n} \sum_{j=1}^n (X_j - \bar{X}_n)^2}$$

となる. □

注意 7.10. $g : \mathbb{R} \times [0, \infty) \rightarrow [0, \infty)$ は 1 対 1 とする. すると $\hat{\theta}_n$ が θ^* の最尤推定量ならば, $g(\hat{\theta}_n)$ は $g(\theta^*)$ の最尤推定量であることがわかる. すなわち, 最尤推定量は母数の変換に関して不変である. このことに注意すれば, 例 7.9 の結果から $(\hat{\mu}_n, s_n^2)$ は $(\mu^*, (\sigma^*)^2)$ の最尤推定量になることがわかる. \square

問 7.1. (7.3) と (7.4) を確認せよ.

定理 7.11.

$$X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} p(x|\eta) = h(x) \exp\{\eta^* T(x) - \kappa(\eta^*)\} (\eta^* \in \mathcal{E}^o)$$

とする. ただし \mathcal{E}^o は自然母数空間 $\mathcal{E} \subset \mathbb{R}$ の内部である. さらに $\ddot{\kappa}(\eta^*) > 0$ ($\eta^* \in \mathcal{E}^o$) を仮定する. このとき η^* の最尤推定量 $\hat{\eta}_n$ は十分大きな n に対して確率 1 で一意的に存在¹して

$$\begin{aligned} \hat{\eta}_n &\xrightarrow{\text{a.s.}} \eta^*, \\ \sqrt{n}(\hat{\eta}_n - \eta^*) &\rightsquigarrow \mathbf{N}\left(0, \frac{1}{\ddot{\kappa}(\eta^*)}\right) \end{aligned}$$

が成立する.

Proof. $X_1 = x_1, X_2, \dots, X_n = x_n$ を観測したときの尤度関数 lik_n は

$$\text{lik}_n(\eta | \mathbf{x}) = \left\{ \prod_{j=1}^n h(x_j) \right\} \exp\left[\eta \sum_{j=1}^n T(x_j) - n\kappa(\eta)\right]$$

となる. このことから対数尤度関数は

$$\ell_n(\eta | \mathbf{x}) = \log(\text{lik}_n(\eta | \mathbf{x})) = n\{\eta \bar{T}_n - \kappa(\eta)\} + (\text{constant})$$

となる. ただし $\bar{T}_n = n^{-1} \sum_{j=1}^n T(x_j)$ である. したがって

$$\dot{\ell}_n(\eta | \mathbf{x}) = \frac{d\ell_n}{d\eta}(\eta | \mathbf{x}) = n(\bar{T}_n - \dot{\kappa}) = 0 \Leftrightarrow \bar{T}_n = \dot{\kappa}(\eta) = \mathbf{E}[T(X_1)]$$

となる. 最後の等号は (6.4) からわかる.

¹これは不明瞭な表現である. 意味するところは, $A_n = \{\hat{\eta}_n \text{ は存在}\}$ としたとき

$$\lim_{n \rightarrow \infty} \Pr\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) = 1$$

が成立することである.

以後, 対数尤度関数の $x_j (j = 1, 2, \dots, n)$ のところに X_j を代入して, 確率変数にしたものを考える. $T(X_1)$ は有限の期待値をもつので, 大数の法則 (定理 4.17) から

$$\bar{T}_n = \frac{1}{n} \sum_{j=1}^n T(X_j) \xrightarrow{\text{a.s.}} \mathbb{E}[T(X_1)] = \dot{\kappa}(\eta^*) \quad (n \rightarrow \infty) \quad (7.5)$$

がわかる. η^* の適当な近傍を $U(\eta^*)$ と書き, $\dot{\kappa}(U(\eta^*)) := \{\dot{\kappa}(\eta); \eta \in U(\eta^*)\}$ とおく. (7.5) から

$$\Pr\left(\lim_{n \rightarrow \infty} \bar{T}_n \in \dot{\kappa}(U(\eta^*))\right) = 1$$

となる. さらに, $\ddot{\kappa}(\eta^*) > 0$ なので, $\ddot{\kappa}(\eta) > 0 (\eta \in U(\eta^*))$ となる. したがって, 関数 $\mathcal{E}^\circ \ni \eta \mapsto \dot{\kappa}(\eta)$ は近傍 $U(\eta^*)$ 上で狭義単調増加関数である. このことより

$$\bar{T}_n = \dot{\kappa}(\eta)$$

をみたく $\eta \in U(\eta^*) \subset \mathcal{E}^\circ$ が存在する. したがって, 近傍 $\kappa(U(\eta^*))$ 上で $\dot{\kappa}$ の逆写像 $\dot{\kappa}^{-1}$ が存在する. このことを踏まえて

$$\hat{\eta}_n = \dot{\kappa}^{-1}(\bar{T}_n)$$

とおく. さらに

$$\bar{T}_n \xrightarrow{\text{a.s.}} \dot{\kappa}(\eta^*) \quad \text{かつ} \quad \dot{\kappa}^{-1}(t) \text{ は近傍 } \dot{\kappa}(U(\eta^*)) \text{ 上で連続}$$

なので定理 4.14(6) から

$$\hat{\eta}_n = \dot{\kappa}^{-1}(\bar{T}_n) \xrightarrow{\text{a.s.}} \dot{\kappa}^{-1}(\dot{\kappa}(\eta^*)) = \eta^*$$

がわかる. 一方, (6.4) から $\mathbb{E}[T(X_1)] = \dot{\kappa}(\eta^*)$, $\text{Var}[T(X_1)] = \dot{\kappa}(\eta^*)$ となることに注意して, 中心極限定理 (定理 4.22) を用いると

$$\sqrt{n}(\bar{T}_n - \dot{\kappa}(\eta^*)) \rightsquigarrow \mathbf{N}(0, \ddot{\kappa}(\eta^*))$$

がわかる. 以上のことを踏まえて, $\hat{\eta}_n = \dot{\kappa}^{-1}(\bar{T}_n)$ に対して, デルタ法 (定理 4.25) を適用すると

$$\sqrt{n}(\hat{\eta}_n - \eta^*) \rightsquigarrow \mathbf{N}\left(0, \frac{1}{\ddot{\kappa}(\eta^*)}\right)$$

がわかる. □

7.3 不偏推定と情報不等式

$\mathbb{X} \subset \mathbb{R}$ を空でない部分集合とする. $(\Omega, \mathcal{A}, \Pr)$ を確率空間とし, X_1, X_2, \dots, X_n をこの空間上の独立同一分布に従う \mathbb{X} 値確率変数列 (ランダム標本) とする. また $\mathbf{X} = (X_1, X_2, \dots, X_n)^\top$ と記し, \mathbf{X} の値域 (標本空間) を $\mathbb{X}^n \subset \mathbb{R}^n$ と表記する. さらに

$$P = \Pr \circ X_1^{-1}, \quad P^{\otimes n} = \Pr \circ \mathbf{X}^{-1}$$

とおく. すると $P^{\otimes n} = \underbrace{P \times P \times \dots \times P}_{n \text{ 個}}$ となっている.

$\mathcal{P} = \{P_\theta; \theta \in \Theta \subset \mathbb{R}\}$ を \mathbb{X} 上の正則母数モデルとする. 以下では, $\Theta \subset \mathbb{R}$ とし, P_θ に関する期待値と分散を $E_\theta[\cdot], \text{Var}_\theta[\cdot]$ と表記する.

定義 7.12. $\mathbf{X} := (X_1, X_2, \dots, X_n)^\top \sim P_\theta^{\otimes n}$ ($\theta \in \Theta \subset \mathbb{R}$) とする. 統計量 $T(\mathbf{X})$ が $\forall \theta \in \Theta$ に対して

$$E_\theta[T(\mathbf{X})] = \theta$$

をみたすとき $T(\mathbf{X})$ は θ の**不偏推定量** (unbiased estimator) という.

例 7.13. $n \geq 2$ とし, $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N(\theta, 1)$ ($\theta \in \mathbb{R}$) とする. このとき $\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j$ は θ の不偏推定量である. \square

定理 7.14 (Rao-Blackwell の定理). $\mathbf{X} = (X_1, X_2, \dots, X_n)^\top \sim P_\theta^{\otimes n}$ ($\theta \in \Theta \subset \mathbb{R}$) で $T = T(\mathbf{X})$ は θ の十分統計量とする. $\hat{\theta}_n(\mathbf{X})$ は θ の任意の不偏推定量で, 有限の分散 $\text{Var}_\theta[\hat{\theta}_n(\mathbf{X})]$ をもつものとする. 推定量 $\tilde{\theta}_n(T)$ を

$$\tilde{\theta}_n(T) := E_\theta[\hat{\theta}_n(\mathbf{X}) | T]$$

により定めたとき, 次の (1), (2) が成立する.

- (1) $\tilde{\theta}_n(T)$ は θ の不偏推定量である.
- (2) $\text{Var}_\theta[\tilde{\theta}_n(T)] \leq \text{Var}_\theta[\hat{\theta}_n(\mathbf{X})]$ ($\forall \theta \in \Theta$) が成立する.

Proof. (1) T は十分統計量なので, $E_\theta[\hat{\theta}_n(\mathbf{X}) | T]$ は θ に依存しないので $\tilde{\theta}_n(T)$ は推定量となる. また定理 2.34(2) より

$$E_\theta[\tilde{\theta}_n(T)] = E_\theta[E_\theta[\hat{\theta}_n(\mathbf{X}) | T]] = E_\theta[\hat{\theta}_n(\mathbf{X})] = \theta \quad (\theta \in \Theta)$$

となり $\tilde{\theta}_n(T)$ は θ の不偏推定量であることが示せた.

(2) 定理 2.34(4) より

$$\begin{aligned} \mathbf{E}_\theta[(\widehat{\theta}_n(\mathbf{X}) - \widetilde{\theta}_n(T))\widetilde{\theta}_n(T)|T] &= \widetilde{\theta}_n(T)\mathbf{E}_\theta[\widehat{\theta}_n(\mathbf{X}) - \widetilde{\theta}_n(T)|T] \\ &= \widetilde{\theta}_n(T)\left(\mathbf{E}_\theta[\widehat{\theta}_n(\mathbf{X})|T] - \widetilde{\theta}_n(T)\right) \\ &= \widetilde{\theta}_n(T)(\widetilde{\theta}_n(T) - \widetilde{\theta}_n(T)) = 0 \end{aligned}$$

となり

$$\mathbf{E}_\theta[(\widehat{\theta}_n(\mathbf{X}) - \widetilde{\theta}_n(T))\widetilde{\theta}_n(T)] = \mathbf{E}_\theta[\mathbf{E}_\theta[(\widehat{\theta}_n(\mathbf{X}) - \widetilde{\theta}_n(T))\widetilde{\theta}_n(T)|T]] = 0 \quad (7.6)$$

を得る. 一方

$$\begin{aligned} \text{Var}_\theta[\widehat{\theta}_n(\mathbf{X})] &= \mathbf{E}_\theta[\{\widehat{\theta}_n(\mathbf{X}) - \mathbf{E}_\theta[\widehat{\theta}_n(\mathbf{X})]\}^2] \\ &= \mathbf{E}_\theta[\{\widehat{\theta}_n(\mathbf{X}) - \theta\}^2] \\ &= \mathbf{E}_\theta[\{(\widehat{\theta}_n(\mathbf{X}) - \widetilde{\theta}_n(T)) + (\widetilde{\theta}_n(T) - \theta)\}^2] \\ &= \mathbf{E}_\theta[\{\widehat{\theta}_n(\mathbf{X}) - \widetilde{\theta}_n(T)\}^2] + \mathbf{E}_\theta[\{\widetilde{\theta}_n(T) - \theta\}^2] \\ &\quad + 2\mathbf{E}_\theta[(\widehat{\theta}_n(\mathbf{X}) - \widetilde{\theta}_n(T))(\widetilde{\theta}_n(T) - \theta)] \end{aligned}$$

である. しかし, (7.6) および $\widehat{\theta}_n(\mathbf{X})$ と $\widetilde{\theta}_n(T)$ は θ の不偏推定量であることに注意すると

$$\begin{aligned} &\mathbf{E}_\theta[(\widehat{\theta}_n(\mathbf{X}) - \widetilde{\theta}_n(T))(\widetilde{\theta}_n(T) - \theta)] \\ &= \underbrace{\mathbf{E}_\theta[(\widehat{\theta}_n(\mathbf{X}) - \widetilde{\theta}_n(T))\widetilde{\theta}_n(T)]}_{=0 \quad \because (7.6)} - \underbrace{\theta\{\mathbf{E}_\theta[\widehat{\theta}_n(\mathbf{X})] - \mathbf{E}_\theta[\widetilde{\theta}_n(T)]\}}_{=0} \\ &= 0 \end{aligned}$$

がわかる. 以上から

$$\begin{aligned} \text{Var}_\theta[\widehat{\theta}_n(\mathbf{X})] &= \mathbf{E}_\theta[\{\widehat{\theta}_n(\mathbf{X}) - \widetilde{\theta}_n(T)\}^2] + \text{Var}_\theta[\widetilde{\theta}_n(T)] \\ &\geq \text{Var}_\theta[\widetilde{\theta}_n(T)] \quad (\theta \in \Theta) \end{aligned}$$

を得る. □

定義 7.15. $\mathbf{X} = (X_1, X_2, \dots, X_n)^\top \sim P_\theta^{\otimes n}$ ($\theta \in \Theta \subset \mathbb{R}$) とする. θ の任意の不偏推定量 $\widetilde{\theta}_n(\mathbf{X})$ に対して

$$\text{Var}_\theta[\widehat{\theta}_n(\mathbf{X})] \leq \text{Var}_\theta[\widetilde{\theta}_n(\mathbf{X})] \quad (\theta \in \Theta)$$

をみたす θ の不偏推定量 $\widehat{\theta}_n(\mathbf{X})$ が存在するとき, $\widehat{\theta}_n(\mathbf{X})$ を θ の一様最小分散不偏推定量 (uniformly minimum variance unbiased estimator: UMVUE) という.

注意 7.16. 存在すれば、一意的である。演習問題 7.7 を参照のこと。

可測関数 $g: \mathbb{R} \rightarrow \mathbb{R}$ は連続微分可能とする。 $\mathbf{X} \sim P_\theta^{\otimes n}$ に基づき $g(\theta)$ の推定問題を考える。

定理 7.17. $\{P_\theta; \theta \in \Theta \subset R\}$ は正則母数モデルとする。 $\mathbf{X} \sim P_\theta^{\otimes n}$ ($\theta \in \Theta \subset \mathbb{R}$) とし \mathbf{X} は同時 p.d.f. または p.m.f. $p_\theta^{\mathbf{X}}(\mathbf{x})$ をもつとする。 $\hat{\theta}_n = \hat{\theta}_n(\mathbf{X})$ を $g(\theta)$ の任意の不偏推定量とし

$$A(\tilde{\theta}, \theta) := \text{Var}_\theta \left[\frac{p_{\tilde{\theta}}^{\mathbf{X}}(\mathbf{X})}{p_\theta^{\mathbf{X}}(\mathbf{X})} \right] \quad (\tilde{\theta}, \theta \in \Theta) \quad (7.7)$$

$$A(\tilde{\theta}, \theta) > 0 \quad (\tilde{\theta} \in \Theta, \tilde{\theta} \neq \theta)$$

とおく。このとき

$$\text{Var}_\theta [\hat{\theta}_n(\mathbf{X})] \geq \sup_{\tilde{\theta} \in \Theta} \frac{\{g(\tilde{\theta}) - g(\theta)\}^2}{A(\tilde{\theta}, \theta)}$$

が成り立つ。

Proof. \mathbf{X} は連続型確率変数の場合の証明を与える。 \mathbf{X} の値域を \mathbb{X}^n と書く。さらに、 $\tilde{\mathbb{X}}^n = \{\mathbf{x} \in \mathbb{X}^n; p_\theta^{\mathbf{X}}(\mathbf{x}) > 0\}$ とおく。また、離散型の場合は、積分記号を和の記号に変更すればよい。推定量 $\hat{\theta}_n$ は $g(\theta)$ に対して不偏なので

$$E_\theta [\hat{\theta}_n(\mathbf{X})] = \int_{\mathbb{X}^n} \hat{\theta}_n(\mathbf{x}) p_\theta^{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = g(\theta) \quad (\theta \in \Theta)$$

となる。よって

$$\begin{aligned} E_\theta \left[\hat{\theta}_n(\mathbf{X}) \left\{ \frac{p_{\tilde{\theta}}^{\mathbf{X}}(\mathbf{X})}{p_\theta^{\mathbf{X}}(\mathbf{X})} - 1 \right\} \right] &= \int_{\tilde{\mathbb{X}}^n} \hat{\theta}_n(\mathbf{x}) \left\{ \frac{p_{\tilde{\theta}}^{\mathbf{X}}(\mathbf{x})}{p_\theta^{\mathbf{X}}(\mathbf{x})} - 1 \right\} p_\theta^{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbb{X}^n} \hat{\theta}_n(\mathbf{x}) p_{\tilde{\theta}}^{\mathbf{X}}(\mathbf{x}) d\mathbf{x} - \int_{\mathbb{X}^n} \hat{\theta}_n(\mathbf{x}) p_\theta^{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\ &= g(\tilde{\theta}) - g(\theta) \end{aligned} \quad (7.8)$$

となる。一方

$$E_\theta \left[\frac{p_{\tilde{\theta}}^{\mathbf{X}}(\mathbf{X})}{p_\theta^{\mathbf{X}}(\mathbf{X})} \right] = \int_{\tilde{\mathbb{X}}^n} \frac{p_{\tilde{\theta}}^{\mathbf{X}}(\mathbf{x})}{p_\theta^{\mathbf{X}}(\mathbf{x})} p_\theta^{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = \int_{\mathbb{X}^n} p_{\tilde{\theta}}^{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = 1 \quad (7.9)$$

と (7.8) に注意すると

$$\begin{aligned}
 g(\tilde{\theta}) - g(\theta) &= E_{\theta} \left[\hat{\theta}_n(\mathbf{X}) \left\{ \frac{p_{,\tilde{\theta}}^{\mathbf{X}}(\mathbf{X})}{p_{\theta}^{\mathbf{X}}(\mathbf{X})} - 1 \right\} \right] \\
 &= E_{\theta} \left[\left\{ \hat{\theta}_n(\mathbf{X}) - E_{\theta}[\hat{\theta}_n(\mathbf{X})] + E_{\theta}[\hat{\theta}_n(\mathbf{X})] \right\} \left\{ \frac{p_{,\tilde{\theta}}^{\mathbf{X}}(\mathbf{X})}{p_{\theta}^{\mathbf{X}}(\mathbf{X})} - 1 \right\} \right] \\
 &= E_{\theta} \left[\left\{ \hat{\theta}_n(\mathbf{X}) - E_{\theta}[\hat{\theta}_n(\mathbf{X})] \right\} \left\{ \frac{p_{,\tilde{\theta}}^{\mathbf{X}}(\mathbf{X})}{p_{\theta}^{\mathbf{X}}(\mathbf{X})} - 1 \right\} \right] \\
 &\quad + E_{\theta}[\hat{\theta}_n(\mathbf{X})] \underbrace{E_{\theta} \left[\frac{p_{,\tilde{\theta}}^{\mathbf{X}}(\mathbf{X})}{p_{\theta}^{\mathbf{X}}(\mathbf{X})} - 1 \right]}_{=0} \\
 &= E_{\theta} \left[\left\{ \hat{\theta}_n(\mathbf{X}) - E_{\theta}[\hat{\theta}_n(\mathbf{X})] \right\} \left\{ \frac{p_{,\tilde{\theta}}^{\mathbf{X}}(\mathbf{X})}{p_{\theta}^{\mathbf{X}}(\mathbf{X})} - 1 \right\} \right]
 \end{aligned}$$

を得る. Cauchy-Schwarz の不等式 (定理 3.10) を上の式の最右辺に適用すると

$$\begin{aligned}
 |g(\tilde{\theta}) - g(\theta)| &= \left| E_{\theta} \left[\hat{\theta}_n(\mathbf{X}) \left\{ \frac{p_{,\tilde{\theta}}^{\mathbf{X}}(\mathbf{X})}{p_{\theta}^{\mathbf{X}}(\mathbf{x})} - 1 \right\} \right] \right| \\
 &\leq \sqrt{E_{\theta} \left[\left\{ \hat{\theta}_n(\mathbf{X}) - E_{\theta}[\hat{\theta}_n(\mathbf{X})] \right\}^2 \right]} \sqrt{E_{\theta} \left[\left\{ \frac{p_{,\tilde{\theta}}^{\mathbf{X}}(\mathbf{X})}{p_{\theta}^{\mathbf{X}}(\mathbf{x})} - 1 \right\}^2 \right]} \\
 &= \sqrt{\text{Var}_{\theta}[\hat{\theta}_n(\mathbf{X})]} \sqrt{\text{Var}_{\theta} \left[\frac{p_{,\tilde{\theta}}^{\mathbf{X}}(\mathbf{X})}{p_{\theta}^{\mathbf{X}}(\mathbf{x})} \right]} \tag{7.10}
 \end{aligned}$$

を得る. 最後の等号は (7.9) を用いた. したがって

$$\text{Var}_{\theta}[\hat{\theta}_n(\mathbf{X})] \text{Var}_{\theta} \left[\frac{p_{,\tilde{\theta}}^{\mathbf{X}}(\mathbf{X})}{p_{\theta}^{\mathbf{X}}(\mathbf{x})} \right] \geq \{g(\tilde{\theta}) - g(\theta)\}^2 \tag{7.11}$$

を得る. したがって (7.7) と (7.11) から $\forall \theta \in \Theta$ に対して

$$\text{Var}_{\theta}[T(\mathbf{X})] \geq \frac{\{g(\tilde{\theta}) - g(\theta)\}^2}{A(\tilde{\theta}, \theta)}$$

となる. 上式の $\tilde{\theta}$ は任意だったので, 上式の左辺において $\tilde{\theta}$ に関して \sup を取ると定理の主張はわかる. \square

定理 7.18. (Cramér-Rao の不等式) 次の条件を仮定する.

- (1) $\{P_{\theta}; \theta \in \Theta \subset R\}$ は正則母数モデルとする.

- (2) $g: \mathbb{R} \rightarrow \mathbb{R}$ は連続微分可能でその導関数は $\dot{g}(\theta) \neq 0 (\theta \in \Theta)$ をみたす. さらに

$$\lim_{\tilde{\theta} \rightarrow \theta} \frac{A(\tilde{\theta}, \theta)}{(\tilde{\theta} - \theta)^2} = J(\theta) > 0$$

が存在する. ただし, $A(\tilde{\theta}, \theta)$ は (7.7) で与えたものである.

- (3) $\forall \theta \in \Theta$ に対して十分小さな $\epsilon > 0$ をとると任意の $\tilde{\theta} \in \{\psi \in \Theta; |\psi - \theta| < \epsilon\}$ に対して, ある関数 $G(\mathbf{x}|\theta)$ が存在して

$$\left| \frac{p_{\tilde{\theta}}^{\mathbf{X}}(\mathbf{x}) - p_{\theta}^{\mathbf{X}}(\mathbf{x})}{(\tilde{\theta} - \theta)p^{\mathbf{X}}(\mathbf{x}|\theta)} \right| < G(\mathbf{x}|\theta) \quad \text{かつ} \quad E_{\theta}[G^2(\mathbf{X}|\theta)] < \infty$$

をみたす.

このとき $g(\theta)$ の任意の不偏推定量 $T(\mathbf{X})$ に対して

$$\text{Var}_{\theta}[\hat{\theta}_n(\mathbf{X})] \geq \frac{\{\dot{g}(\theta)\}^2}{\mathcal{F}_{\mathbf{X}}(\theta)} \quad (\theta \in \Theta) \quad (7.12)$$

が成り立つ. ただし

$$\mathcal{F}_{\mathbf{X}}(\theta) = E_{\theta} \left[\left\{ \frac{\partial}{\partial \theta} \log p_{\theta}^{\mathbf{X}}(\mathbf{X}) \right\}^2 \right]$$

である.

注意 7.19. (7.12) を Cramér-Rao の不等式といい, その右辺を Cramér-Rao の下限という. \square

定理 7.18 の証明: \mathbf{X} の値域を \mathbb{X}^n とし, $\tilde{\mathbb{X}}^n = \{\mathbf{x} \in \mathbb{X}^n; p_{\theta}^{\mathbf{X}}(\mathbf{x}) > 0\}$ とおく. 定理 7.17 と仮定 (2) から

$$\text{Var}_{\theta}[\hat{\theta}_n(\mathbf{X})] \geq \lim_{\tilde{\theta} \rightarrow \theta} \frac{\left\{ \frac{g(\tilde{\theta}) - g(\theta)}{\tilde{\theta} - \theta} \right\}^2}{\frac{A(\tilde{\theta}, \theta)}{(\tilde{\theta} - \theta)^2}} = \frac{\{\dot{g}(\theta)\}^2}{J(\theta)} \quad (7.13)$$

を得る. 次に仮定 (3) に注意して Lebesgue の優収束定理を用いると

$$\begin{aligned} J(\theta) &= \lim_{\tilde{\theta} \rightarrow \theta} \frac{A(\tilde{\theta}, \theta)}{(\tilde{\theta} - \theta)^2} \\ &= \lim_{\tilde{\theta} \rightarrow \theta} \int_{\tilde{\mathcal{X}}^n} \left\{ \frac{p_{\tilde{\theta}}^{\mathbf{X}}(\mathbf{x})}{p_{\theta}^{\mathbf{X}}(\mathbf{x})} - 1 \right\}^2 \frac{p_{\theta}^{\mathbf{X}}(\mathbf{x})}{(\tilde{\theta} - \theta)^2} d\mathbf{x} \\ &= \int_{\tilde{\mathcal{X}}^n} \left\{ \lim_{\tilde{\theta} \rightarrow \theta} \frac{p_{\tilde{\theta}}^{\mathbf{X}}(\mathbf{x}) - p_{\theta}^{\mathbf{X}}(\mathbf{x})}{(\tilde{\theta} - \theta)p_{\theta}^{\mathbf{X}}(\mathbf{x})} \right\}^2 p_{\theta}^{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\ &= \int_{\tilde{\mathcal{X}}^n} \left\{ \frac{\partial}{\partial \theta} \log p_{\theta}^{\mathbf{X}}(\mathbf{x}) \right\}^2 p_{\theta}^{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = \mathcal{F}_{\mathbf{X}}(\theta) \end{aligned}$$

となる. これと (7.13) を合わせると

$$\text{Var}_{\theta}[\hat{\theta}_n(\mathbf{X})] \geq \frac{\{\dot{g}(\theta)\}^2}{\mathcal{F}_{\mathbf{X}}(\theta)} \quad (7.14)$$

を得る. □

注意 7.20. 指数型分布族が定理 7.18 の正則条件 (2) と (3) をみたしているかを確認する.

以下, 加筆をすること.

□

定義 7.21. $\mathbf{X} \sim P_{\theta}^{\otimes n} (\theta \in \Theta \subset \mathbb{R})$ とする. θ の不偏推定量 $\hat{\theta}_n(\mathbf{X})$ は R 有効²であるとは

$$\text{Var}_{\theta}[\hat{\theta}_n(\mathbf{X})] = \frac{1}{\mathcal{F}_{\mathbf{X}}(\theta)} \quad (\theta \in \Theta)$$

をみたすときをいう.

定理 7.22. $\mathbf{X} = (X_1, X_2, \dots, X_n) \sim P_{\theta}^{\otimes n} (\theta \in \Theta \subset \mathbb{R})$ とする. θ の不偏推定量 $\hat{\theta}_n$ が R 有効であるための必要十分条件は \mathbb{R} 上の \mathbb{R} 値関数 $T(x), A(\theta), \kappa(\theta), g(x)$ があって

$$\hat{\theta}_n(\mathbf{X}) = \frac{1}{n} \sum_{j=1}^n T(X_j) \quad (7.15)$$

$$\log p(x|\theta) = A(\theta)T(x) - \kappa(\theta) + g(x), \quad (7.16)$$

$$\int_{\mathcal{X}} T(x)p(x|\theta) dx = \theta \quad (7.17)$$

²Rao 有効の意味であらう.

をみたとときである. ただし X_1 の p.d.f.(または p.m.f.) と値域をそれぞれ $p(x|\theta)$ と $\mathbb{X}(\subset \mathbb{R})$ と書いた.

Proof. 連続型分布に対する証明を与える. 離散型分布に対しては, 積分記号を和の記号に替えればよい. 表現 (7.15) – (7.17) が成立したとする.

$\ell_n(\theta|\mathbf{x}) = \sum_{j=1}^n \log p(x_j|\theta)$ ($\mathbf{x} = (x_1, x_2, \dots, x_n)$) とおく. すると

$$\begin{aligned} \dot{\ell}_n(\theta|\mathbf{x}) &= \frac{\partial}{\partial \theta} \ell_n(\theta|\mathbf{x}) = \dot{A}(\theta) \sum_{j=1}^n T(x_j) - n \dot{\kappa}(\theta), \\ \dot{A}(\theta) &= \frac{\partial A}{\partial \theta}(\theta), \quad \dot{\kappa}(\theta) = \frac{\partial B}{\partial \theta}(\theta) \end{aligned}$$

となる. よって

$$E_\theta[\dot{\ell}_n(\theta|\mathbf{X})] = 0 \Leftrightarrow \dot{A}(\theta) E_\theta[T(X_1)] = \dot{\kappa}(\theta)$$

と (7.17) とあわせると

$$\theta \dot{A}(\theta) = \dot{\kappa}(\theta) \tag{7.18}$$

となる. さらに

$$0 = \int_{\mathbb{X}^n} \{T(x) - \theta\} p(x|\theta) dx = \int_{\mathbb{X}^n} \{T(x) - \theta\} \exp\{\log p(x|\theta)\} dx$$

を (7.16) と 7.18) に注意して, θ に関して微分すると

$$\begin{aligned} 0 &= - \int_{\mathbb{X}} p_\theta(x) dx + \int_{\mathbb{X}} \{T(x) - \theta\} \{ \dot{A}(\theta) T(x) - \underbrace{\dot{\kappa}(\theta)}_{=\theta \dot{A}(\theta)} \} p_\theta(x) dx \\ &= - \int_{\mathbb{X}} p_\theta(x) dx + \int_{\mathbb{X}} \{T(x) - \theta\} \{ \dot{A}(\theta) T(x) - \theta \dot{A}(\theta) \} p_\theta(x) dx \end{aligned}$$

となることがわかる. よって, $\int_{\mathbb{X}} p_\theta(x) dx = 1$ に注意して, 上式を整理すると

$$1 = \dot{A}(\theta) E_\theta[\{T(X_1) - \theta\}^2]$$

を得る. さらに $E_\theta[T(X_1)] = \theta$ に注意すれば

$$\text{Var}_\theta[T(X_1)] = \frac{1}{\dot{A}(\theta)} \tag{7.19}$$

となる. したがって

$$\begin{aligned}
 \mathcal{F}_{\mathbf{X}}(\theta) &= \text{Var}_{\theta}[\dot{\ell}_n(\theta | \mathbf{X})] & (7.20) \\
 &= \text{Var}_{\theta} \left[\sum_{j=1}^n \{ \dot{A}(\theta) T(X_j) - \dot{\kappa}(\theta) \} \right] \\
 &= \text{Var}_{\theta} \left[\sum_{j=1}^n \dot{A}(\theta) \{ T(X_j) \} \right] \quad (\because \text{分散は平行移動に関して不変}) \\
 &= \sum_{j=1}^n \{ \dot{A}(\theta) \}^2 \text{Var}[T(X_j)] \quad (\because X_1, X_2, \dots, X_n \text{ は独立}) \\
 &= n \{ \dot{A}(\theta) \}^2 \text{Var}[T(X_1)] & (7.21) \\
 &= n \dot{A}(\theta) \quad (\because (7.19))
 \end{aligned}$$

を得る. また, $\hat{\theta}_n(\mathbf{X}) = \frac{1}{n} \sum_{j=1}^n T(X_j)$ なので

$$E_{\theta}[\hat{\theta}_n(\mathbf{X})] = \theta$$

となり $\hat{\theta}_n$ は θ の不偏推定量である. 再度, (7.19) に注意すると

$$\text{Var}_{\theta}[\hat{\theta}_n] = \text{Var}_{\theta} \left[\frac{1}{n} \sum_{j=1}^n T(X_j) \right] = \frac{1}{n^2} \sum_{j=1}^n \text{Var}_{\theta}[T(X_j)] = \frac{1}{n \dot{A}(\theta)} = \frac{1}{\mathcal{F}_{\mathbf{X}}(\theta)}$$

となり $\hat{\theta}_n$ は R 有効推定量となる.

次に逆を示す. P_{θ} の同時 p.d.f. $\prod_{j=1}^n p_{\theta}(x_j)$ に対して, その対数尤度関数を $\ell_n(\theta | \mathbf{x}) = \log \left(\prod_{j=1}^n p_{\theta}(x_j) \right)$ と書くことにする. すると

$$\int_{\mathbb{X}^n} \exp\{\ell_n(\theta | \mathbf{x})\} d\mathbf{x} = 1$$

である. この式を θ に関して微分すると

$$0 = \int_{\mathbb{X}^n} \dot{\ell}_n(\theta | \mathbf{x}) \exp\{\ell_n(\theta | \mathbf{x})\} d\mathbf{x} = E_{\theta}[\dot{\ell}_n(\mathbf{x} | \mathbf{X})] \quad (7.22)$$

となる. 同様に $E_{\theta}[\hat{\theta}_n(\mathbf{X})] = \theta$ より

$$\int_{\mathbb{X}^n} \hat{\theta}_n(\mathbf{x}) \exp\{\ell_n(\theta | \mathbf{x})\} d\mathbf{x} = \theta$$

となる. この式を θ に関して微分すると

$$1 = \int_{\mathbb{X}^n} \hat{\theta}_n(\mathbf{x}) \dot{\ell}_n(\theta | \mathbf{x}) \exp\{\ell_n(\theta | \mathbf{x})\} d\mathbf{x} = E_{\theta}[\hat{\theta}_n(\mathbf{X}) \dot{\ell}_n(\theta | \mathbf{X})] \quad (7.23)$$

を得る. (7.22) と (7.23) を合わせると

$$\begin{aligned} \mathbb{E}_\theta[(\hat{\theta}_n(\mathbf{X}) - \theta) \dot{\ell}_n(\theta | \mathbf{X})] &= \underbrace{\mathbb{E}_\theta[\hat{\theta}_n(\mathbf{X}) \dot{\ell}_n(\theta | \mathbf{X})]}_{=1} - \theta \underbrace{\mathbb{E}_\theta[\dot{\ell}_n(\theta | \mathbf{X})]}_{=0} \\ &= 1 \end{aligned} \quad (7.24)$$

を得る. いま

$$Y = \frac{\dot{\ell}_n(\theta | \mathbf{X})}{\mathcal{F}_\mathbf{X}(\theta)}$$

とおく. (7.20) に注意すると

$$\mathbb{E}_\theta[Y \dot{\ell}_n(\theta | \mathbf{X})] = \mathbb{E}_\theta\left[\frac{\{\dot{\ell}_n(\theta | \mathbf{X})\}^2}{\mathcal{F}_\mathbf{X}(\theta)}\right] = \frac{1}{\mathcal{F}_\mathbf{X}(\theta)} \text{Var}[\dot{\ell}_n(\theta | \mathbf{X})] = 1$$

となる. これと (7.24) を合わせると

$$\begin{aligned} \mathbb{E}_\theta[(\hat{\theta}_n(\mathbf{X}) - \theta - Y)Y] &= \frac{1}{\mathcal{F}_\mathbf{X}(\theta)} \underbrace{\mathbb{E}_\theta[(\hat{\theta}_n(\mathbf{X}) - \theta) \dot{\ell}_n(\theta | \mathbf{X})]}_{=1} - \frac{1}{\mathcal{F}_\mathbf{X}(\theta)} \underbrace{\mathbb{E}_\theta[Y \dot{\ell}_n(\theta | \mathbf{X})]}_{=1} \\ &= 0 \end{aligned}$$

を得る. 上の式を用いると

$$\begin{aligned} \mathbb{E}_\theta[(\hat{\theta}_n(\mathbf{X}) - \theta)^2] &= \mathbb{E}_\theta[(\hat{\theta}_n(\mathbf{X}) - \theta - Y + Y)^2] \\ &= \mathbb{E}_\theta[(\hat{\theta}_n(\mathbf{X}) - \theta - Y)^2] + \mathbb{E}_\theta[Y^2] \\ &\geq \mathbb{E}_\theta[Y^2] \end{aligned}$$

となる. よって等号成立は

$$\begin{aligned} \hat{\theta}_n(\mathbf{X}) - \theta - Y = 0 &\Leftrightarrow \hat{\theta}_n(\mathbf{X}) - \theta = \frac{\dot{\ell}_n(\theta | \mathbf{X})}{\mathcal{F}_\mathbf{X}(\theta)} \\ &\Leftrightarrow \mathcal{F}_\mathbf{X}(\theta)(\hat{\theta}_n(\mathbf{X}) - \theta) = \dot{\ell}_n(\theta | \mathbf{X}) \end{aligned}$$

となる. このことより, $\bar{\theta}, \underline{\theta} \in \mathbb{X}$ とすると

$$\begin{aligned} \ell_n(\bar{\theta} | \mathbf{X}) - \ell_n(\underline{\theta} | \mathbf{X}) &= \int_{\underline{\theta}}^{\bar{\theta}} \dot{\ell}_n(\mathbf{X} | \theta) d\theta = \int_{\underline{\theta}}^{\bar{\theta}} \mathcal{F}_\mathbf{X}(\theta)(\hat{\theta}_n(\mathbf{X}) - \theta) d\theta \\ &= \tilde{A}(\bar{\theta}, \underline{\theta})\hat{\theta}_n(\mathbf{X}) - \tilde{B}(\bar{\theta}, \underline{\theta}) \end{aligned}$$

と書ける. ただし

$$\tilde{A}(\bar{\theta}, \underline{\theta}) := \int_{\underline{\theta}}^{\bar{\theta}} \mathcal{F}_\mathbf{X}(\theta) d\theta; \quad \tilde{B}(\bar{\theta}, \underline{\theta}) := \int_{\underline{\theta}}^{\bar{\theta}} \theta \mathcal{F}_\mathbf{X}(\theta) d\theta$$

である. ここで $n = 1, \bar{\theta} = \theta, \underline{\theta} = 1, X_1 = x$ とおくと

$$\ell_1(\theta|x) = T(x)A(\theta) - B(\theta) + g(x)$$

と書ける. ただし

$$T(x) = \hat{\theta}_1(x), \quad g(x) = \ell_1(1|x), \quad A(\theta) = \tilde{A}(\theta, 1), \quad B(\theta) = \tilde{B}(\theta, 1)$$

と書ける. □

注意 7.23. 定理 7.22 から, R 有効な推定量が存在する統計的モデルは指数型分布族となることがわかる. □

7.4 章末注釈と参考文献

定理 7.11 の証明は [7, pp.241 – 242] を借用した. 定理 7.14 の証明は [42, pp.35 – 41] を借用した. 定理 7.22 は [30, pp.39 – 40] を借用した.

7.5 演習問題

演習問題 7.1. 離散型確率変数 X は, p.m.f.

$$p_\theta(x) = \begin{cases} \left(\frac{\theta}{2}\right)^{|x|} (1-\theta)^{1-|x|} & (x = -1, 0, 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

からの標本の大きさが 1 のランダム標本とする. ただし, $0 < \theta < 1$ で, X は確率空間 $(\Omega, \mathcal{A}, \text{Pr})$ 上で定義された確率変数とする. ふたつの統計量

$$S = S(X) = |X|, \quad T = T(X) = \begin{cases} 2 & (X = 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

を考える.

- (1) S の確率分布を求めよ³.
- (2) T の確率分布を求めよ.
- (3) S が与えられたときの X の条件付確率分布を求め⁴, S は θ の十分統計量かどうかを調べよ.
- (4) S は θ の不偏推定量かどうかを調べよ.
- (5) T は θ の不偏推定量かどうかを調べよ.

³確率関数または確率分布表を求めること.

⁴条件付き p.m.f. を求めるか, X が与えられた値ごとの確率分布表を求めればよい.

(6) S と T の平均 2 乗誤差 $\text{MSE}_S(\theta) = E[(S - \theta)^2]$ と $\text{MSE}_T(\theta) = E[(T - \theta)^2]$ を求めよ. S と T の平均 2 乗誤差 $\text{MSE}_S(\theta)$ と $\text{MSE}_T(\theta)$ の大小の比較をせよ. (横軸を θ とし, 縦軸を MSE の値として, S と T の平均 2 乗誤差 $\text{MSE}_S(\theta)$ と $\text{MSE}_T(\theta)$ のグラフを描き比較すること.)

演習問題 7.2. $\theta > 0$ とする. 連続型確率変数 X は p.d.f.

$$p_\theta(x) = \begin{cases} \theta x^{\theta-1} & (0 < x < 1) \\ 0 & (\text{その他}) \end{cases}$$

からの大きさ 1 のランダム標本とする. ただし, X は確率空間 $(\Omega, \mathcal{A}, \text{Pr})$ 上で定義された確率変数とする.

- (1) 連続型確率変数 X の期待値 $E[X]$ を求めよ.
- (2) 連続型確率変数 X の分布関数 $F(x) = \text{Pr}(X \leq x)$ ($x \in \mathbb{R}$) を求めよ.
- (3) $X = x$ ($0 < x < 1$) を観測したときの尤度関数 $\text{lik}(\theta|x)$ を述べよ (答えのみでよい).
- (4) θ の最尤推定値を求めよ.

演習問題 7.3. $m, n \geq 2$ を整数とし

$$\begin{aligned} X_1, X_2, \dots, X_m & \stackrel{\text{i.i.d.}}{\sim} N(\mu_1, \sigma^2) \\ Y_1, Y_2, \dots, Y_n & \stackrel{\text{i.i.d.}}{\sim} N(\mu_2, \sigma^2) \end{aligned}$$

とする. さらに

$$\begin{aligned} \bar{X}_m &= \frac{1}{m} \sum_{i=1}^m X_i \\ \bar{Y}_n &= \frac{1}{n} \sum_{i=1}^n Y_i \\ \hat{\sigma}^2 &= \frac{1}{m+n-2} \left\{ \sum_{i=1}^m (\bar{X}_m - X_i)^2 + \sum_{i=1}^n (\bar{Y}_n - Y_i)^2 \right\} \end{aligned}$$

とおく. これらの確率変数は同じ確率空間 $(\Omega, \mathcal{A}, \text{Pr})$ 上で定義されたものとする. このとき, 以下の問いに答えよ.

(1)

$$\bar{X}_m - \bar{Y}_n \sim N\left(\mu_1 - \mu_2, \frac{m+n}{mn} \sigma^2\right)$$

となることを説明 (証明) せよ.

(2)

$$E[\hat{\sigma}^2] = \sigma^2$$

となることを説明 (証明) せよ.

演習問題 7.4. $n \geq 2$ を整数とし, $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Po}(\lambda)$ とする⁵.
ただし, $\lambda > 0$ で, これらの確率変数は確率空間 $(\Omega, \mathcal{A}, \text{Pr})$ 上で定義されているとする.

- (1) $\sum_{i=1}^n X_i$ の分布は $\text{Po}(n\lambda)$ となることを積率母関数を計算することで示せ.
- (2) λ の任意の不偏推定量の分散についてのその下限 (Cramér-Rao の下限) を求めよ.
- (3) λ の最尤推定量を求め, その分散が Cramér-Rao の下限に到達することを確認せよ.

演習問題 7.5. X_1, X_2, \dots, X_n を母集団分布 (平均は θ , 分散は 1^2 の正規分布)

$$p(x|\theta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-\theta)^2}{2}\right)$$

からの大きさ n のランダム標本とする. ただし, 母数 $\theta (-\infty < \theta < \infty)$ は未知とする. このとき, 以下の問いに答えよ.

- (1) $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ を観測したときの θ の尤度関数 $\text{lik}_n(\theta|x_1, x_2, \dots, x_n)$ と対数尤度関数 $\text{lik}_n(\theta|x_1, x_2, \dots, x_n)$ を書け.
- (2) θ の最尤推定値 $\hat{\theta}_n(x_1, x_2, \dots, x_n)$ を求めよ.
- (3) θ の最尤推定量 $\hat{\theta}_n(X_1, X_2, \dots, X_n)$ の平均 $E[\hat{\theta}_n(X_1, X_2, \dots, X_n)]$ を求めよ.
- (4) θ の最尤推定量 $\hat{\theta}_n(X_1, X_2, \dots, X_n)$ の分散 $\text{Var}[\hat{\theta}_n(X_1, X_2, \dots, X_n)]$ を求めよ.
- (5) 任意の正数 ϵ に対して,

$$\lim_{n \rightarrow \infty} \Pr(|\hat{\theta}_n(X_1, X_2, \dots, X_n) - \theta| \geq \epsilon) = 0$$

を示せ.

演習問題 7.6. 確率空間 (Ω, \mathcal{A}, P) 上の確率変数 X_1, X_2, X_3 は独立同一に平均が μ , 分散が σ^2 の正規分布 $N(\mu, \sigma^2)$ に従うとする. ただし, $\mu \in \mathbb{R}, 0 < \sigma < \infty$ である. このとき, 以下の問に答えよ.

- (1) X_1 の積率母関数 $M(t) = E[e^{tX_1}]$ が

$$M(t) = \exp\left(\mu t + \frac{\sigma^2}{2} t^2\right)$$

5

$$\Pr(X_1 = x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad (x = 0, 1, \dots)$$

である. ただし, これらの確率変数は確率空間 $(\Omega, \mathcal{A}, \text{Pr})$ 上で定義されたものとする. その積率母関数は

$$M_{X_1}(t) = E[e^{tX_1}] = e^{(e^t - 1)\lambda}$$

である.

で与えられることを示せ. ただし, 任意の $\mu \in \mathbb{R}$ と $\sigma > 0$ に対して

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = 1$$

を証明なしで用いてよい.

(2) X_1 の期待値 $E[X_1]$ と 2 次の積率 $E[X_1^2]$ を (1) で求めた積率母関数を利用して求めよ.

(3) $a, b, c, d \in \mathbb{R}$ に対して, σ^2 の推定量の族

$$T := g(X_1, X_2, X_3) = a\{X_1^2 + X_2^2 + X_3^2\} + b\{X_1X_2 + X_1X_3 + X_2X_3\} \\ + c\{X_1 + X_2 + X_3\} + d$$

を考える.

$$E[T] = \sigma^2$$

をみたす a, b, c, d を求めよ.

(4) (3) で求めた a, b, c, d に対する T は

$$T = \frac{1}{2} \left((X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + (X_3 - \bar{X})^2 \right); \quad \bar{X} = \frac{1}{3}(X_1 + X_2 + X_3)$$

の形に変形できることを確認せよ.

演習問題 7.7. T_1 と T_2 を母数 θ の最小分散不偏推定量とする. このとき, 以下の問いに答えよ.

- (1) $E\left[\frac{1}{2}(T_1 + T_2)\right]$ を計算せよ.
- (2) $\text{Var}\left[\frac{1}{2}(T_1 + T_2)\right]$ を $\text{Var}[T_1], \text{Cov}[T_1, T_2]$ で表現せよ.
- (3) $\text{Cov}[T_1, T_2] = \text{Var}[T_1]$ を証明せよ.
- (4) $\text{Var}[T_1 - T_2]$ の値を求めよ.
- (5) $\Pr(T_1 = T_2) = 1$ を示せ.

ヒント (3) T_1 は最小分散不偏推定量であることと Cauchy-Scharz の不等式を $\text{Cov}[T_1, T_2]$ に適用するとよい.

(5) 非負値確率変数 W に対して

$$E[W] = 0 \Rightarrow \Pr(W = 0) = 1$$

を証明なしで用いてよい.

演習問題 7.8.

演習問題 7.9.

第8章 検定と信頼区間

第8.1では仮説検定問題の枠組みと考え方を説明する. 第8.2では, 帰無仮説も対立仮説も単純であるとき, 検定統計量の最適定理である Neyman-Pearson の補題を説明する. 第8.3では, 検定統計量の導出原理を説明する. 第??では様々な検定方法をまとめる. 第8.6では, 区間推定量の考え方を説明する. 第8.7では, 区間推定量の構成法の代表的なものを説明する.

8.1 仮説検定の考え方

母集団分布を特徴付ける母数について想定したある仮説の真偽を標本に基づいて調べることを**仮説検定** hypothesis test という.

いま

$$\mathbf{X} = (X_1, X_2, \dots, X_n)^\top \sim P_{\boldsymbol{\theta}^*}^{\otimes n} (\boldsymbol{\theta}^* \in \Theta \subset \mathbb{R}^d)$$

とする. ただし $d, n \in \mathbb{N}$ で, Θ は母数空間である. また, $\mathbb{X}^n (\subset \mathbb{R}^n)$ を \mathbf{X} の値域としたとき, $P_{\boldsymbol{\theta}^*}^{\otimes n}$ は $(\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n))$ 上の確率測度 (\mathbf{X} の分布) である. $\Theta_0 \subset \Theta$ は空でない Θ の真部分集合とし, $\boldsymbol{\theta}^*$ が Θ_0 に入るか否かを調べたいとき, 仮説

$$H_0 : \boldsymbol{\theta}^* \in \Theta_0 \quad \text{vs.} \quad H_1 : \boldsymbol{\theta}^* \in \Theta_1 := \Theta \setminus \Theta_0 \quad (8.1)$$

を考える. H_0 または H_1 のいずれかが正しいかを判断することを「 H_0 を H_1 に対して検定 (test) する」という. H_0 を**帰無仮説** (null hypothesis) といい, H_1 を**対立仮説** (alternative hypothesis) という. Θ_0 が Θ の1つの元から成るとき H_0 を**単純仮説** (simple hypothesis) という. そうでないとき H_0 を**複合仮説** (composite hypothesis) という. 言葉を乱用して, 「 Θ_0 は単純仮説である」等ということもある.

仮説 (8.1) に対して, 以下のように検定方式を定めることができる. \mathbf{X} の取り得るすべての値の集合を $\mathbb{X}^n (\subset \mathbb{R}^n)$ と表す. \mathbb{X}^n を2つの排反で空でない部分集合 W と W^c に分割する. すなわち $W \neq \emptyset, W^c \neq \emptyset$ で $W \cup W^c = \mathbb{X}^n$ かつ $W \cap W^c = \emptyset$ である. \mathbf{X} の実現値を \boldsymbol{x} と書いたと

き, 検定方式は

$$\boldsymbol{x} \in W \Rightarrow \text{帰無仮説 } H_0 \text{ を棄却し, 対立仮説 } H_1 \text{ を採択,}$$

$$\boldsymbol{x} \in W^c \Rightarrow \text{帰無仮説 } H_0 \text{ を受容}$$

と表現できる. このとき W を**棄却域** (critical region) といい, W^c を**受容域** (acceptance region) という.

上のように定めた検定方式には 2 つのタイプの誤りが起こる可能性がある. (1) 帰無仮説 H_0 が正しいにもかかわらず標本の実現値 \boldsymbol{x} に基づいて検定した結果, H_0 を棄却してしまうこと. 逆に, (2) 対立仮説 H_1 が正しいにもかかわらず標本の実現値 \boldsymbol{x} に基づいて検定した結果, H_0 を受容してしまうことである. (1) の誤判断を**第 1 種の誤り**といい, (2) の誤判断を**第 2 種の誤り**とそれぞれ呼ぶ. 一般に一方の誤りが起こる確率を小さくする検定方式は, 他方の誤りを起こす確率を大きくする. すなわち, 両者の誤りが起こる確率を同時に小さくする検定方式はないことが知られている.

以下では「よい」検定方式を一般的な形で定式化することを考える. 関数 $\phi: \mathbb{X}^n \rightarrow [0, 1]$ は可測関数とする. この関数 ϕ を用いて次のように検定方式を定める. $\boldsymbol{X} = \boldsymbol{x}$ を観測したとき, 確率 $\phi(\boldsymbol{x})$ で帰無仮説 H_0 を棄却する検定方式を考える. この ϕ を**検定関数** (test function) という. 関数 ϕ が \mathbb{X} の空でない部分集合の定義関数のとき, この検定関数 ϕ で定まる検定方式を**非確率化検定** (nonrandomized test) という. すなわち

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & (\boldsymbol{x} \in W) \\ 0 & (\boldsymbol{x} \in W^c) \end{cases}$$

とすると棄却域 W をもつ非確率化検定が定まる. そうでない検定方式を**確率化検定** (randomized test) という.

以後 $\phi(\boldsymbol{X})$ を検定統計量 (test statistic) ということにする. さらに検定統計量 ϕ によって定まる検定方式を単に検定ということにする. 検定統計量 $\phi(\boldsymbol{X})$ の第 1 種の誤りの確率は

$$E_{\theta}[\phi(\boldsymbol{X})] \quad (\theta \in \Theta_0)$$

となり, 第 2 種の誤りの確率は

$$1 - E_{\theta}[\phi(\boldsymbol{X})] \quad (\theta \in \Theta_1)$$

となる. ただし

$$E_{\theta}[\phi(\boldsymbol{X})] = \int_{\mathbb{X}^n} \phi(\boldsymbol{x}) p_{\theta}(\boldsymbol{x}) d\boldsymbol{x}$$

と定めた。「よい」検定として、まず第 1 種の誤りの確率の Θ_0 上の上限を α ($0 < \alpha < 1$) 以下にするような検定を考える。すなわち

$$\sup_{\theta \in \Theta_0} E_{\theta}[\phi(\mathbf{X})] \leq \alpha \quad (8.2)$$

である。(8.1) をみたく検定関数または検定 ϕ を**有意水準 α の検定** (level α test) という。つぎに有意水準 α の検定 ϕ の中で、第 2 種の誤りの確率

$$1 - E_{\theta}[\phi(\mathbf{X})] \quad (\theta \in \Theta_1) \quad (8.3)$$

を Θ_1 上で最小にするものを見つけることを目指す。すなわち

$$E_{\theta}[\phi(\mathbf{X})] \quad (\theta \in \Theta_1)$$

を最大にするものである。この確率を $\theta \in \Theta_1$ の関数とみて

$$\beta(\theta) := E_{\theta}[\phi(\mathbf{X})] \quad (\theta \in \Theta_1)$$

と表記する。これを**検出力関数** (power function) または**検出力**という。したがって「よい」検定は次のように定義される。

定義 8.1. 有意水準 α ($0 < \alpha < 1$) の検定 ϕ で検出力を任意の $\theta \in \Theta_1$ に対して最大にするものを有意水準 α の**一様最強力検定** (uniformly most powerful test = u.m.p. 検定) という。特に帰無仮説と対立仮説が単純仮説であるとき、u.m.p. 検定を単に有意水準 α の**最強力検定** (m.p. 検定) という。

最後に、 P 値を定義する。検定問題 (8.1) において、関数 $T: \mathbb{X}^n \rightarrow \mathbb{R}$ と定数 $c \in \mathbb{R}$ により棄却域 $W \subset \mathbb{X}^n$ が

$$W := \{\mathbf{x} \in \mathbb{X}^n; T(\mathbf{x}) > c\}$$

で定まる検定関数 ϕ を考える。すなわち、 $\phi(\mathbf{x}) = \mathbb{1}_W(\mathbf{x})$ ($\mathbf{x} \in \mathbb{X}^n$) である。ただし、定数 c は

$$\sup_{\theta \in \Theta_0} E_{\theta}[\phi(\mathbf{X})] \leq \alpha$$

をみたすものである。

定義 8.2.

$$p(\mathbf{x}) := \sup_{\theta \in \Theta_0} \Pr_{\theta}(T(\mathbf{X}) \geq T(\mathbf{x})) = \sup_{\theta \in \Theta_0} E_{\theta}[\mathbb{1}_{[W(\mathbf{x}), \infty)}(W(\mathbf{X}))]$$

を実現値 $\mathbf{X} = \mathbf{x}$ に対する P 値 (単に P 値) という。

定理 8.3. すべての $\theta \in \Theta_0$ と α ($0 < \alpha < 1$) に対して

$$\Pr_{\theta}(p(\mathbf{X}) \leq \alpha) \leq \alpha$$

が成り立つ.

Proof. $\theta_0 \in \Theta_0$ を固定して

$$p_{\theta_0}(\mathbf{x}) := \Pr_{\theta_0}(T(\mathbf{X}) \geq T(\mathbf{x}))$$

とおき, $-T(\mathbf{X})$ の分布関数を F_{θ_0} と書く. すなわち

$$F_{\theta_0}(t) = \Pr_{\theta_0}(-T(\mathbf{X}) \leq t) \quad (t \in \mathbb{R})$$

である. すると

$$p_{\theta_0}(\mathbf{x}) = \Pr_{\theta_0}(T(\mathbf{X}) \geq T(\mathbf{x})) = \Pr_{\theta_0}(-T(\mathbf{X}) \leq -T(\mathbf{x})) = F_{\theta_0}(-T(\mathbf{x}))$$

と書き直せる. このことから, $F_{\theta_0}(-T(\mathbf{X}))$ は $-T(\mathbf{X})$ の分布関数 F_{θ_0} に $-T(\mathbf{X})$ を代入したものであるに注意する. すると命題 1.41 から

$$F_{\theta_0}(-T(\mathbf{X})) \sim U(0, 1) \tag{8.4}$$

となることがわかる. $p_{\theta_0}(\mathbf{X}) = F_{\theta_0}(-T(\mathbf{X}))$ と (8.4) から, $0 < \alpha < 1$ に対して

$$\Pr_{\theta_0}(p(\mathbf{X} | \theta_0) \leq \alpha) = \Pr_{\theta_0}(F_{\theta_0}(-T(\mathbf{X})) \leq \alpha) = \alpha$$

を得る. さらに

$$p(\mathbf{x}) = \sup_{\theta \in \Theta_0} p_{\theta}(\mathbf{x}) \geq p_{\theta_0}(\mathbf{x})$$

であるから

$$\Pr_{\theta_0}(p(\mathbf{X} \leq \alpha) \leq \Pr_{\theta_0}(p_{\theta_0}(\mathbf{X}) \leq \alpha) = \alpha$$

がわかる. よって, 定理の主張は証明された. □

8.2 Neyman-Pearson の定理

まず m.p. 検定を求める最も基本的定理を述べる. 以下では, 簡単のために \mathbf{X} は同時 p.d.f. $p_{\theta}^{\mathbf{X}}(\mathbf{x})$ をもつとして議論を進めていく. 離散型確率変数のときは同時 p.m.f. を考え, 積分を和の記号に替えればよい.

定理 8.4. (Neyman-Pearson の定理) 母数空間は \mathbb{R}^d の異なる 2 点から成るとする. $\Theta = \{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1\}$ である. $\mathbf{X} = (X_1, X_2, \dots, X_n)^\top \sim P_{\boldsymbol{\theta}}^{\otimes n}$ ($\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$) とし, $p_{\boldsymbol{\theta}}^{\mathbf{X}}(\mathbf{x})$ ($\boldsymbol{\theta} \in \Theta, \mathbf{x} \in \mathbb{R}^n$) を \mathbf{X} の同時 p.d.f. または p.m.f. とする. 検定問題

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \quad \text{vs.} \quad H_1 : \boldsymbol{\theta} = \boldsymbol{\theta}_1$$

に対する有意水準 α ($0 < \alpha < 1$) の m.p. 検定 ϕ_0 は以下で与えられる.

$$\phi_0(\mathbf{x}) = \begin{cases} 1 & (p_{\boldsymbol{\theta}_1}^{\mathbf{X}}(\mathbf{x}) > cp_{\boldsymbol{\theta}_0}^{\mathbf{X}}(\mathbf{x})) \\ \gamma & (p_{\boldsymbol{\theta}_1}^{\mathbf{X}}(\mathbf{x}) = cp_{\boldsymbol{\theta}_0}^{\mathbf{X}}(\mathbf{x})) \\ 0 & (p_{\boldsymbol{\theta}_1}^{\mathbf{X}}(\mathbf{x}) < cp_{\boldsymbol{\theta}_0}^{\mathbf{X}}(\mathbf{x})) \end{cases} \quad (8.5)$$

である. ただし γ, c ($0 \leq \gamma \leq 1, c > 0$) は

$$E_{\boldsymbol{\theta}_0}[\phi_0(\mathbf{X})] = \alpha \quad (8.6)$$

から定まる定数である.

Proof. まず

$$\begin{aligned} B_1 &= \{\mathbf{x} \in \mathbb{X}^n; p_{\boldsymbol{\theta}_1}^{\mathbf{X}}(\mathbf{x}) > cp_{\boldsymbol{\theta}_0}^{\mathbf{X}}(\mathbf{x})\}, \\ B_2 &= \{\mathbf{x} \in \mathbb{X}^n; p_{\boldsymbol{\theta}_1}^{\mathbf{X}}(\mathbf{x}) = cp_{\boldsymbol{\theta}_0}^{\mathbf{X}}(\mathbf{x})\}, \\ B_3 &= \{\mathbf{x} \in \mathbb{X}^n; p_{\boldsymbol{\theta}_1}^{\mathbf{X}}(\mathbf{x}) < cp_{\boldsymbol{\theta}_0}^{\mathbf{X}}(\mathbf{x})\} \end{aligned}$$

とする. ϕ は有意水準 α の任意の検定とする. すなわち

$$E_{\boldsymbol{\theta}_0}[\phi(\mathbf{X})] \leq \alpha \quad (8.7)$$

をみます. 一方 (8.5) より

$$\begin{aligned}
 & \mathbf{E}_{\theta_1}[\phi_0(\mathbf{X})] - \mathbf{E}_{\theta_1}[\phi(\mathbf{X})] \\
 &= \int_{\mathbb{X}^n} \phi_0(\mathbf{x}) p_{\theta_1}^{\mathbf{X}}(\mathbf{x}) d\mathbf{x} - \int_{\mathbb{X}^n} \phi(\mathbf{x}) p_{\theta_1}^{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\
 &= \int_{B_1} \underbrace{\phi_0(\mathbf{x}) p_{\theta_1}^{\mathbf{X}}(\mathbf{x})}_{=1} d\mathbf{x} + \int_{B_2} \underbrace{\phi_0(\mathbf{x}) p_{\theta_1}^{\mathbf{X}}(\mathbf{x})}_{=\gamma} d\mathbf{x} + \int_{B_3} \underbrace{\phi_0(\mathbf{x}) p_{\theta_1}^{\mathbf{X}}(\mathbf{x})}_{=0} d\mathbf{x} \\
 &\quad - \int_{B_1} \phi(\mathbf{x}) p_{\theta_1}^{\mathbf{X}}(\mathbf{x}) d\mathbf{x} - \int_{B_2} \phi(\mathbf{x}) p_{\theta_1}^{\mathbf{X}}(\mathbf{x}) d\mathbf{x} - \int_{B_3} \phi(\mathbf{x}) p_{\theta_1}^{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\
 &= \int_{B_1} \{1 - \phi(\mathbf{x})\} \underbrace{p_{\theta_1}^{\mathbf{X}}(\mathbf{x})}_{>cp_{\theta_0}^{\mathbf{X}}(\mathbf{x}) (x \in B_1)} d\mathbf{x} + \int_{B_2} \{\gamma - \phi(\mathbf{x})\} \underbrace{p_{\theta_1}^{\mathbf{X}}(\mathbf{x})}_{=cp_{\theta_0}^{\mathbf{X}}(\mathbf{x}) (x \in B_2)} d\mathbf{x} \\
 &\quad + \int_{B_3} \{-\phi(\mathbf{x})\} \underbrace{p_{\theta_1}^{\mathbf{X}}(\mathbf{x})}_{<cp_{\theta_0}^{\mathbf{X}}(\mathbf{x}) (x \in B_3)} d\mathbf{x} \\
 &\geq \int_{B_1} \{1 - \phi(\mathbf{x})\} cp_{\theta_0}^{\mathbf{X}}(\mathbf{x}) d\mathbf{x} + \int_{B_2} \{\gamma - \phi(\mathbf{x})\} cp_{\theta_0}^{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\
 &\quad + \int_{B_3} \{-\phi(\mathbf{x})\} cp_{\theta_0}^{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\
 &= c \int_{B_1} \{1 - \phi(\mathbf{x})\} p_{\theta_0}^{\mathbf{X}}(\mathbf{x}) d\mathbf{x} + c \int_{B_2} \{\gamma - \phi(\mathbf{x})\} p_{\theta_0}^{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\
 &\quad + c \int_{B_3} \{-\phi(\mathbf{x})\} p_{\theta_0}^{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\
 &= c \int_{\mathbb{X}^n} \{\phi_0(\mathbf{x}) - \phi(\mathbf{x})\} p_{\theta_0}^{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\
 &= c \left\{ \mathbf{E}_{\theta_0}[\phi_0(\mathbf{X})] - \mathbf{E}_{\theta_0}[\phi(\mathbf{X})] \right\} \quad (\because (8.5)) \\
 &= c \left\{ \alpha - \mathbf{E}_{\theta_0}[\phi(\mathbf{X})] \right\} \geq 0 \quad (\because (8.7) \text{ より})
 \end{aligned}$$

を得る. したがって

$$\mathbf{E}_{\theta_1}[\phi_0(\mathbf{X})] \geq \mathbf{E}_{\theta_1}[\phi(\mathbf{X})]$$

となるので, ϕ_0 は有意水準 α の m.p. 検定となる. □

例 8.5. $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N(\theta, \sigma^2)$ で $\sigma^2 (\sigma > 0)$ は既知とする. このとき検定問題

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta = \theta_1 \quad (\theta_1 > \theta_0)$$

に対する m.p. 検定を求め. まず $\mathbf{X} = (X_1, X_2, \dots, X_n)^\top$ の同時 p.d.f. は

$$p_{\theta}^{\mathbf{X}}(\mathbf{x}) = \left(\frac{1}{\sqrt{2\pi\sigma}} \right)^n \exp \left[-\frac{1}{2\sigma^2} \sum_{j=1}^n (x_j - \theta)^2 \right], \quad \mathbf{x} = (x_1, x_2, \dots, x_n)^\top$$

で与えられることに注意をする. 簡単な計算から

$$\begin{aligned} \log \left[\frac{p_{\theta_1}^{\mathbf{X}}(\mathbf{x})}{p_{\theta_0}^{\mathbf{X}}(\mathbf{x})} \right] &= -\frac{1}{2\sigma^2} \left[\sum_{j=1}^n (x_j - \theta_1)^2 - \sum_{j=1}^n (x_j - \theta_0)^2 \right] \\ &= \frac{n(\theta_1 - \theta_0)}{\sigma^2} \left(\bar{x}_n - \frac{\theta_0 + \theta_1}{2} \right) \end{aligned} \quad (8.8)$$

となる. ただし

$$\bar{x}_n = \frac{1}{n} \sum_{j=1}^n x_j$$

である. ここで

$$\begin{aligned} p_{\theta_0}^{\mathbf{X}}(\mathbf{x}) > c p_{\theta_1}^{\mathbf{X}}(\mathbf{x}) &\Leftrightarrow \log \left[\frac{p_{\theta_1}^{\mathbf{X}}(\mathbf{x})}{p_{\theta_0}^{\mathbf{X}}(\mathbf{x})} \right] > \log c \\ &\Leftrightarrow \bar{x}_n > c' \quad (\because (8.8) \text{ より}) \end{aligned}$$

である. \mathbf{X} は連続型確率変数なので, $\bar{X}_n = n^{-1} \sum_{j=1}^n X_j = c'$ である確率は 0 となるので, m.p. 検定の形は

$$\phi_0(\mathbf{x}) = \begin{cases} 1 & (\bar{x}_n > c') \\ 0 & (\bar{x}_n \leq c') \end{cases}$$

となる. 定数 c' は

$$\alpha = \Pr_{\theta_0} \{ \bar{X}_n > c' \} \quad (8.9)$$

から定まる. ただし, $B \in \mathcal{B}(\mathbb{R}^n)$ に対して

$$\Pr_{\theta}(\mathbf{X} \in B) = E_{\theta}[\mathbb{1}_B(\mathbf{X})] = \int_B p_{\theta}^{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \quad (\theta \in \Theta)$$

と定めた. (8.9) は

$$\alpha = \Pr_{\theta_0} \left\{ \frac{\sqrt{n}(\bar{X}_n - \theta_0)}{\sigma} > \frac{\sqrt{n}(c' - \theta_0)}{\sigma} \right\} \quad (8.10)$$

と書き直せ, $\theta = \theta_0$ のもとで $\sqrt{n}(\bar{X}_n - \theta_0)/\sigma \sim \mathbf{N}(0, 1)$ であるので

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

とすると (8.10) は

$$\alpha = 1 - \Phi \left[\frac{\sqrt{n}(c' - \theta_0)}{\sigma} \right]$$

となる. 標準正規分布の上側 $100 \times \alpha\%$ を z_α とすると

$$\frac{\sqrt{n}(c' - \theta_0)}{\sigma} = z_\alpha \Leftrightarrow c' = \theta_0 + \frac{z_\alpha \sigma}{\sqrt{n}}$$

を得る. よって有意水準 α の m.p. 検定は

$$\phi_0(\mathbf{x}) = \begin{cases} 1 & \left(\bar{x}_n > \theta_0 + \frac{z_\alpha \sigma}{\sqrt{n}} \right) \\ 0 & \left(\bar{x}_n \leq \theta_0 + \frac{z_\alpha \sigma}{\sqrt{n}} \right) \end{cases}$$

となる. 次に ϕ_0 の検出力は次のようになる.

$$\begin{aligned} \beta_{\phi_0}(\theta_1) &= \mathbf{E}_{\theta_1}[\phi_0(\mathbf{X})] = \Pr_{\theta_1} \left\{ \bar{X}_n > \theta_0 + \frac{z_\alpha \sigma}{\sqrt{n}} \right\} \\ &= \Pr_{\theta_1} \left\{ \frac{\sqrt{n}(\bar{X}_n - \theta_1)}{\sigma} > z_\alpha - \frac{\sqrt{n}(\theta_1 - \theta_0)}{\sigma} \right\} \end{aligned} \quad (8.11)$$

となる. $\theta = \theta_1$ のとき $\sqrt{n}(\bar{X}_n - \theta_1)/\sigma \sim N(0, 1)$ なので, (8.11) より仮説間の平均の差 $\theta_1 - \theta_0 (> 0)$ が大きいほど検出力は大きくなる. また標本 n が大きくなっても検出力が大きくなることがわかる. \square

例 8.6. $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Ber}(\theta)$ ($0 < \theta < 1$) とする. このとき検定問題

$$H_0 : \theta = \theta_0, \quad H_1 : \theta = \theta_1 \quad (\theta_1 > \theta_0) \quad (8.12)$$

に対する m.p. 検定を求める. まず $\mathbf{X} = (X_1, X_2, \dots, X_n)^\top$ の同時 p.m.f. は

$$p_\theta^{\mathbf{X}}(\mathbf{x}) = \prod_{j=1}^n \theta^{x_j} (1 - \theta)^{1-x_j} \quad (\mathbf{x} = (x_1, x_2, \dots, x_n)^\top) \quad (8.13)$$

で与えられるので

$$\log \frac{p_{\theta_1}^{\mathbf{X}}(\mathbf{x})}{p_{\theta_0}^{\mathbf{X}}(\mathbf{x})} = \left\{ \sum_{j=1}^n x_j \right\} \log \left\{ \frac{\theta_1(1 - \theta_0)}{(1 - \theta_1)\theta_0} \right\} + n \log \left\{ \frac{1 - \theta_1}{1 - \theta_0} \right\}$$

となる. $\theta_1 > \theta_0$ としたので, $\frac{\theta_1(1 - \theta_0)}{(1 - \theta_1)\theta_0} > 1$ であることに注意すると

$$\begin{aligned} p_{\theta_1}^{\mathbf{X}}(\mathbf{x}) > c p_{\theta_0}^{\mathbf{X}}(\mathbf{x}) &\Leftrightarrow \log \left[\frac{p_{\theta_1}^{\mathbf{X}}(\mathbf{x})}{p_{\theta_0}^{\mathbf{X}}(\mathbf{x})} \right] > c' \\ &\Leftrightarrow \sum_{j=1}^n x_j > c'' \end{aligned}$$

と書きかえることができる. よって有意水準 α の m.p. 検定は

$$\phi_0(\mathbf{x}) = \begin{cases} 1 & (\sum_{j=1}^n x_j > c'') \\ \gamma & (\sum_{j=1}^n x_j = c'') \\ 0 & (\sum_{j=1}^n x_j < c'') \end{cases} \quad (8.14)$$

の形になる. ここで γ と c'' を

$$\begin{aligned} \alpha &= E_{\theta_0}[\phi_0(\mathbf{X})] \\ &= \Pr_{\theta_0}\left\{\sum_{j=1}^n X_j > c''\right\} + \gamma \Pr_{\theta_0}\left\{\sum_{j=1}^n X_j = c''\right\} \end{aligned} \quad (8.15)$$

から定まる. $\sum_{j=1}^n X_j$ は $\theta = \theta_0$ のとき二項分布 $\text{Bino}(n, \theta_0)$ に従うので, (8.15) は

$$\alpha = \sum_{j=c''+1}^n \binom{n}{j} \theta_0^j (1-\theta_0)^{n-j} + \gamma \binom{n}{c''} \theta_0^{c''} (1-\theta_0)^{n-c''}$$

となる. まずは c'' を

$$\sum_{j=c''+1}^n \binom{n}{j} \theta_0^j (1-\theta_0)^{n-j} \leq \alpha < \sum_{j=c''}^n \binom{n}{j} \theta_0^j (1-\theta_0)^{n-j}$$

をみたす整数を定める. これから c_0 と書くことにする. すると γ は

$$\gamma = \left[\alpha - \sum_{j=c_0+1}^n \binom{n}{j} \theta_0^j (1-\theta_0)^{n-j} \right] / \left[\binom{n}{c_0} \theta_0^{c_0} (1-\theta_0)^{n-c_0} \right]$$

で定められる. □

8.3 検定統計量の導出方法

\mathbb{X}^n を標本空間とし, $\mathbf{X} = (X_1, X_2, \dots, X_n)^\top \sim \mathbf{P}_\theta^{\otimes n}$ ($\theta \in \Theta \subset \mathbb{R}^d$) とする. 検定関数 $\phi: \mathbb{X}^n \rightarrow [0, 1]$ によって定まる検定方式は, 以下のよう
に定まることがある. ある統計量 $S: \mathbb{X}^n \rightarrow \mathbb{R}$ と定数 c が存在して

$$\begin{aligned} S(\mathbf{x}) \leq c &\Rightarrow \phi(\mathbf{x}) = 1 \\ S(\mathbf{x}) > c &\Rightarrow \phi(\mathbf{x}) = 0 \end{aligned}$$

となる. この場合, $S(\mathbf{X})$ のことも検定統計量と呼ぶことにする.

$\mathbf{X} = (X_1, X_2, \dots, X_n)^\top \sim \mathbf{P}_\theta^{\otimes n}$ ($\theta \in \Theta \subset \mathbb{R}^d$) とする. ただし, $\mathbf{P}_\theta^{\otimes n}$ は \mathbb{R}^n 上の確率測度である. $\mathbf{P}_\theta^{\otimes n}$ は同時 p.d.f. $p_\theta^{\mathbf{X}}(\mathbf{x})$ をもつとする. また, 母数空間 Θ は Θ_0 と Θ_1 に分割されたとする. すなわち $\Theta_0 \cup \Theta_1 = \Theta$, $\Theta_0 \cap \Theta_1 = \emptyset$, $\Theta_0 \neq \emptyset$, $\Theta_1 \neq \emptyset$ である.

8.3.1 尤度比検定統計量

定義 8.7. 検定問題

$$H_0 : \boldsymbol{\theta} \in \Theta_0 \quad \text{vs.} \quad H_1 : \boldsymbol{\theta} \in \Theta_1$$

を検定するための**尤度比検定統計量** (likelihood ratio statistic=l.r. 統計量) は

$$\lambda(\mathbf{X}) = \frac{\sup_{\boldsymbol{\theta} \in \Theta_0} p_{\boldsymbol{\theta}}^{\mathbf{X}}(\mathbf{X})}{\sup_{\boldsymbol{\theta} \in \Theta} p_{\boldsymbol{\theta}}^{\mathbf{X}}(\mathbf{X})}$$

で与えられる. このとき正の定数 C が存在して H_0 の棄却域が

$$W = \{\mathbf{x} \in \mathbb{X}^n; \lambda(\mathbf{x}) \leq C\}$$

で与えられる検定を**尤度比検定** (likelihood ratio test=l.r.t.) という. すなわち

$$\phi(\mathbf{x}) = \mathbf{1}_W(\mathbf{x}) = \begin{cases} 1 & (\mathbf{x} \in W) \\ 0 & (\mathbf{x} \notin W) \end{cases}$$

となる.

注意 8.8. $\hat{\boldsymbol{\theta}}(\mathbf{X})$ を母数空間 Θ での $\boldsymbol{\theta}$ の最尤推定量とし, $\hat{\boldsymbol{\theta}}_0(\mathbf{X})$ を母数空間を Θ_0 に制限したときの $\boldsymbol{\theta}$ の最尤推定量とする. このとき

$$\lambda(\mathbf{X}) = \frac{p_{\hat{\boldsymbol{\theta}}_0(\mathbf{X})}^{\mathbf{X}}(\mathbf{X})}{p_{\hat{\boldsymbol{\theta}}(\mathbf{X})}^{\mathbf{X}}(\mathbf{X})}$$

と表現できる.

定理 8.9. $\mathbf{X} \sim P_{\boldsymbol{\theta}^*}^{\otimes n} (\boldsymbol{\theta}^* \in \Theta \subset \mathbb{R}^d)$ とする. Θ の次元を d , Θ_0 の次元を $r (r < d)$ とする. 検定問題

$$H_0 : \boldsymbol{\theta}^* \in \Theta_0 \quad \text{vs.} \quad \boldsymbol{\theta}^* \in H_1 : \Theta_1 := \Theta \setminus \Theta_0$$

に対する尤度比検定統計量を $\lambda(\mathbf{X})$ とする. このとき H_0 のもとで次が成り立つ.

$$-2 \log \lambda(\mathbf{X}) \rightsquigarrow \chi_{d-r}^2$$

が成立する.

Proof. 節 12.4 の定理 12.12 で証明を与える. □

注意 8.10. 定理 8.9 の結果を用いると尤度比検定の棄却域は

$$W = \{\mathbf{x} \in \mathbb{X}^n; -2 \log \lambda(\mathbf{x}) > \chi_{d-r, \alpha}^2\}$$

で与えられる. ただし $\chi_{d-r, \alpha}^2$ は自由度 $d-r$ の χ^2 分布の上側 $100 \times \alpha\%$ 点である. したがって, 検定手続きは

$$\mathbf{x} \in W \Rightarrow H_0 \text{ は棄却}$$

となる. □

8.3.2 スコア検定

分布 P_θ の p.d.f./p.m.f. を p_θ と書いたとき, スコア関数を

$$\dot{\ell}_\theta(x) := \nabla_\theta \log p_\theta(x)$$

で定める. ただし, $\nabla_\theta = \left(\frac{\partial}{\partial \theta_1}, \frac{\partial}{\partial \theta_2}, \dots, \frac{\partial}{\partial \theta_d} \right)^\top$ である. $X \sim P_\theta$ のとき, 適当な正則条件のもとで

$$E_\theta[\dot{\ell}_\theta(X)] = 0$$

となる. 実現値 $X = x$ の対するスコア関数の実現値 $\dot{\ell}_{\theta_0}(x)$ が 0 からある程度離れているとき, θ_0 は真の母数でないことを示唆することになる.

つぎに, 「ある程度 0 から離れている」を定量化する. $\mathbf{X} = (X_1, X_2, \dots, X_n)$ をランダム標本とする. すると \mathbf{X} の同時 p.d.f./p.m.f. は

$$p_\theta^{\mathbf{X}}(\mathbf{x}) = \prod_{j=1}^n p_\theta(x_j)$$

となる. このことから, 帰無仮説 $H_0: \theta = \theta_0$ に対するスコア関数は

$$\nabla_{\theta_0} \log p_{\theta_0}^{\mathbf{X}}(\mathbf{X}) = \sum_{j=1}^n \dot{\ell}_{\theta_0}(X_j)$$

の形で表現できることがわかる. ここで, $\dot{\ell}_{\theta_0}$ はひとつのデータに対するスコア関数である. 帰無仮説 H_0 のもとでは, $E_{\theta_0}[\dot{\ell}_{\theta_0}(X_j)] = 0$ ($j = 1, 2, \dots, n$) である. したがって, 中心極限定理 4.24 から, 帰無仮説 H_0 のもと

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n \dot{\ell}_{\theta_0}(X_j) \rightsquigarrow N_d(\mathbf{0}_d, \mathcal{F}(\theta_0)) \quad (n \rightarrow \infty)$$

となる. ただし

$$\mathcal{F}(\theta_0) = E_{\theta_0}[\dot{\ell}_{\theta_0}(X_1) \dot{\ell}_{\theta_0}(X_1)^\top]$$

である. よって

$$\left| \mathcal{F}(\theta_0)^{-1} \sum_{j=1}^n \dot{\ell}_{\theta_0}(X_j) \right|_{2,d}^2 > \chi_d^2(\alpha) \Rightarrow H_0 \text{ を棄却} \quad (8.16)$$

とすればよい, ただし, $\chi_d^2(\alpha)$ は自由度 d の χ 自乗分布の上側 α 点で $|\cdot|_{2,d}$ は \mathbb{R}^d の Euclid ノルムである. (8.16) で与えられる検定方式をスコア検定という.

次に、複合帰無仮説に対するスコア検定統計量を導出しよう。一般性を失くことなく、 Θ_0 は

$$\Theta_0 = \{\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \in \mathbb{R}^d; \boldsymbol{\theta}_1 \in \mathbb{R}^r, \boldsymbol{\theta}_2 = \mathbf{0}_{d-r}\}$$

と表現されるとする。ただし、 $0 < r < d$ である。この分割に対応させてスコア関数

$$\dot{\ell}_{\boldsymbol{\theta}}^{\top} = (\dot{\ell}_{\boldsymbol{\theta},1}^{\top}, \dot{\ell}_{\boldsymbol{\theta},2}^{\top})$$

と分割する。 H_0 のもとの $\boldsymbol{\theta}_0$ の最尤推定量 $\hat{\boldsymbol{\theta}}_0^{\top} = (\hat{\boldsymbol{\theta}}_{0,1}^{\top}, \mathbf{0}_{d-r}^{\top})$ は尤度方程式

$$\sum_{j=1}^n \dot{\ell}_{\hat{\boldsymbol{\theta}}_0,1}^{\top}(X_j) = \mathbf{0}_r$$

をみたら。このことから

$$\sum_{j=1}^n \dot{\ell}_{\hat{\boldsymbol{\theta}}_0}^{\top}(X_j) = \left(\mathbf{0}_r^{\top}, \sum_{j=1}^n \dot{\ell}_{\hat{\boldsymbol{\theta}}_0,2}^{\top}(X_j) \right)$$

と表現できる。したがって、(8.16) は

$$\frac{1}{n} \left(\sum_{j=1}^n \dot{\ell}_{\hat{\boldsymbol{\theta}}_0,2}^{\top}(X_j) \right)^{\top} \left(\mathcal{F}_{\hat{\boldsymbol{\theta}}_0}^{-1} \right)_{2,2} \left(\sum_{j=1}^n \dot{\ell}_{\hat{\boldsymbol{\theta}}_0,2}^{\top}(X_j) \right) > \chi_{d-r}^2(\alpha)$$

となる。ここで、 $d \times d$ の正値対称行列 \mathbf{A} に対して、 $(\mathbf{A}^{-1})_{2,2}$ は \mathbf{A}^{-1} の $(d-r) \times (d-r)$ の下側左ブロック行列である。

8.3.3 Wald 検定

$\hat{\boldsymbol{\theta}}$ をフルモデルの最尤推定量とし、 $\hat{\boldsymbol{\theta}}_0$ を帰無仮説 H_0 のもとでの最尤推定量としたとき、帰無仮説 H_0 のもとで

$$n(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_0)^{\top} \widehat{\mathcal{F}}_{\hat{\boldsymbol{\theta}}_0} (\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_0) > \chi_{d-r}^2(\alpha) \Rightarrow H_0 \text{ を棄却}$$

とする検定方式を Wald 検定という。

8.4 検定の一様最適性

8.4.1 単調尤度比

$n \in \mathbb{N}$ とし、 \mathbb{X}^n を標本空間とする。 $\mathbf{X} = (X_1, X_2, \dots, X_n)$ をランダム標本とし、その同時 p.d.f./p.m.f. を $p_{\boldsymbol{\theta}}^{\mathbf{X}}$ とする。ただし、 $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}$ である。以下では、集合 $\{\mathbf{x} \in \mathbb{X}^n; p_{\boldsymbol{\theta}}^{\mathbf{X}}(\mathbf{x}) > 0\}$ は $\boldsymbol{\theta}$ に依存しないと仮定する。

定義 8.11. 関数 $U : \mathbb{X}^n \rightarrow \mathbb{R}$ があって, $\theta < \theta'$ なる任意の $\theta, \theta' \in \Theta$ に対して

$$\frac{p_{\theta'}^{\mathbf{X}}(\mathbf{x})}{p_{\theta}^{\mathbf{X}}(\mathbf{x})}$$

は $U(\mathbf{x})$ の関数であり, $U(\mathbf{x})$ に関して単調増加であるとき, すなわち, 任意の $\forall \mathbf{x}, \tilde{\mathbf{x}} \in \mathbb{X}^n$ に対して

$$U(\mathbf{x}) < U(\tilde{\mathbf{x}}) \Rightarrow \frac{p_{\theta'}^{\mathbf{X}}(\mathbf{x})}{p_{\theta}^{\mathbf{X}}(\mathbf{x})} \leq \frac{p_{\theta'}^{\mathbf{X}}(\tilde{\mathbf{x}})}{p_{\theta}^{\mathbf{X}}(\tilde{\mathbf{x}})}$$

が成り立つとき, $\{p_{\theta}; \theta \in \Theta\}$ は U に関して**単調尤度比** (monotone likelihood ratio=m.l.r.) を持つという. U に関して強い意味の単調であるとき, U に関して強い意味での単調尤度比を持つという.

定理 8.12. 標本空間 \mathbb{X}^n 上の確率変数列 X_1, X_2, \dots, X_n は指数型分布族

$$p_{\theta}(x) = h(x) \exp\{A(\theta)T(\theta) - \kappa^{\vee}(\theta)\} \mathbb{1}_{\mathbb{X}}(x)$$

からのランダム標本とする. ただし, $\theta \in \Theta \subset \mathbb{R}$ で, $A : \Theta \rightarrow [0, \infty)$, $T : \mathbb{X} \rightarrow [0, \infty)$, $\kappa^{\vee} : \Theta \rightarrow [0, \infty)$ である. さらに, A は θ の単調増加関数とし, $U(\mathbf{x}) = \sum_{j=1}^n T(x_j)$ ($\mathbf{x} = (x_1, x_2, \dots, x_n)$) とおく. このとき, $\mathbf{X} = (X_1, X_2, \dots, X_n)$ の同時 p.d.f./p.m.f. の分布族 $\{p_{\theta}^{\mathbf{X}}; \theta \in \Theta\}$ は U に関して単調尤度比を持つ.

Proof. まず

$$p_{\theta}^{\mathbf{X}}(\mathbf{x}) = \prod_{j=1}^n p_{\theta}(x_j) = \left(\prod_{j=1}^n h(x_j) \right) \exp\left\{ A(\theta) \sum_{j=1}^n T(x_j) - n\kappa^{\vee}(\theta) \right\}$$

と書けることに注意する. $\theta < \theta'$ とすると

$$\frac{p_{\theta'}^{\mathbf{X}}(\mathbf{x})}{p_{\theta}^{\mathbf{X}}(\mathbf{x})} = \exp\left\{ -n\{\kappa(\theta') - \kappa(\theta)\} \right\} \exp\left\{ (A(\theta') - A(\theta)) \sum_{j=1}^n T(x_j) \right\}$$

となる. A は単調増加と仮定しているので, $A(\theta') - A(\theta) \geq 0$ である. よって, $p_{\theta'}^{\mathbf{X}}(\mathbf{x})/p_{\theta}^{\mathbf{X}}(\mathbf{x})$ は $\sum_{j=1}^n T(x_j)$ に関して単調増加であることがわかるので, 定理の主張は示された. \square

定理 8.13. 確率ベクトル $\mathbf{X} = (X_1, X_2, \dots, X_n)$ の同時 p.d.f./p.m.f. を $p_{\theta}^{\mathbf{X}}(\mathbf{x})$ ($\theta \in \Theta \subset \mathbb{R}$) とする. 分布族 $\{p_{\theta}^{\mathbf{X}}; \theta \in \Theta\}$ は U に関して強い意味で単調尤度比を持つとする. $\theta_0 \in \Theta$ を定めて, $\Theta_0 = \Theta \cap (-\infty, \theta_0]$ とする. 検定問題

$$H_0 : \theta \in \Theta_0 \quad \text{v.s.} \quad H_1 : \theta \in \Theta \setminus \Theta_0$$

に対する有意水準 α ($0 < \alpha < 1$) の検定関数

$$\phi^*(\mathbf{x}) = \begin{cases} 1 & (U(\mathbf{x}) > c) \\ \gamma & (U(\mathbf{x}) = c) \\ 0 & (U(\mathbf{x}) < c) \end{cases}$$

は u.m.p. 検定となる. ただし, c, γ は $E_{\theta_0}[\phi^*(\mathbf{X})] = \alpha$ から定める定数である.

Proof. $\theta' \in \Theta \setminus \Theta_0$ とする. まず, 検定問題

$$H_0 : \theta = \theta_0 \quad \text{v.s.} \quad H_1 : \theta = \theta'$$

を考える. この場合, 帰無仮説 H_0 と対立仮説 H_1 はともに単純なので, 定理 8.4 より

$$\phi^*(\mathbf{x}) = \begin{cases} 1 & (p_{\theta'}^{\mathbf{X}}(\mathbf{x}) > c_0 p_{\theta}^{\mathbf{X}}(\mathbf{x})) \\ \gamma_0 & (p_{\theta'}^{\mathbf{X}}(\mathbf{x}) = c_0 p_{\theta}^{\mathbf{X}}(\mathbf{x})) \\ 0 & (p_{\theta'}^{\mathbf{X}}(\mathbf{x}) < c_0 p_{\theta}^{\mathbf{X}}(\mathbf{x})) \end{cases}$$

は有意水準 α の m.p. 検定である. ただし, c_0 と γ_0 は

$$E_{\theta_0}[\phi^*(\mathbf{X})] = \alpha$$

となるように定められた定数である. 仮定から, 強い意味での単調増加関数 $\tau : \mathbb{R} \rightarrow \mathbb{R}$ が存在して

$$\frac{p_{\theta'}^{\mathbf{X}}(\mathbf{x})}{p_{\theta}^{\mathbf{X}}(\mathbf{x})} = \tau(U(\mathbf{x}))$$

と書ける. このとき

$$\frac{p_{\theta'}^{\mathbf{X}}(\mathbf{x})}{p_{\theta}^{\mathbf{X}}(\mathbf{x})} > c_0 \Leftrightarrow \tau(U(\mathbf{x})) > c_0 \Leftrightarrow U(\mathbf{x}) > \tau^{-1}(c_0)$$

が成立する. また, 上の同値関係は $>$ を $=$ と $<$ としても成り立つ. ここで $c := \tau^{-1}(c_0)$ と $\gamma_0 = \gamma$ とおくと

$$\phi^*(\mathbf{x}) = \begin{cases} 1 & (U(\mathbf{x}) > c) \\ \gamma & (U(\mathbf{x}) = c) \\ 0 & (U(\mathbf{x}) < c) \end{cases}$$

と表現できる. さらに

$$\begin{aligned} \alpha &= E_{\theta_0}[\phi^*(\mathbf{X})] \\ &= \Pr_{\theta_0}(p_{\theta'}^{\mathbf{X}} > c_0 p_{\theta}^{\mathbf{X}}) + \gamma_0 \Pr_{\theta_0}(p_{\theta'}^{\mathbf{X}} = c_0 p_{\theta}^{\mathbf{X}}) \\ &= \Pr_{\theta_0}(U(\mathbf{X}) > c) + \gamma \Pr_{\theta_0}(U(\mathbf{X}) = c) \end{aligned}$$

から, c と γ は θ_0 にのみ依存する. すなわち, θ' には無関係である. したがって, あらゆる $\theta' \in \Theta \setminus \Theta_0$ に対する検定問題

$$H_0 : \theta = \theta_0 \quad \text{v.s.} \quad H_1 : \theta = \theta'$$

に対して, 検定関数 ϕ^* は m.p. 検定となる. すなわち, 任意の検定関数 ϕ に対して

$$E_{\theta_0}[\phi(\mathbf{X})] \leq \alpha \Rightarrow E_{\theta'}[\phi^*(\mathbf{X})] \geq E_{\theta'}[\phi(\mathbf{X})] \quad (\theta' \in \Theta \setminus \Theta_0)$$

が成り立つ.

以上のことから, すべての $\theta \in \Theta_0$ に対して

$$E_{\theta}[\phi^*(\mathbf{X})] \leq \alpha \tag{8.17}$$

が成り立つことを示せば, 定理の主張は証明される. ここで, $\theta' \in \Theta_0 \setminus \{\theta_0\}$ に対して

$$\alpha(\theta') := E_{\theta'}[\phi^*(\mathbf{X})] \tag{8.18}$$

とおく. 検定問題

$$H_0 : \theta = \theta' \quad \text{v.s.} \quad H_1 : \theta = \theta_0$$

に対する有意水準 $\alpha(\theta')$ の m.p. 検定は

$$\tilde{\phi}(\mathbf{x}) = \begin{cases} 1 & (\mathbf{p}_{\theta'}^{\mathbf{X}}(\mathbf{x}) > \tilde{c}\mathbf{p}_{\theta}^{\mathbf{X}}(\mathbf{x})) \\ \tilde{\gamma} & (\mathbf{p}_{\theta'}^{\mathbf{X}}(\mathbf{x}) = \tilde{c}\mathbf{p}_{\theta}^{\mathbf{X}}(\mathbf{x})) \\ 0 & (\mathbf{p}_{\theta'}^{\mathbf{X}}(\mathbf{x}) < \tilde{c}\mathbf{p}_{\theta}^{\mathbf{X}}(\mathbf{x})) \end{cases}$$

で与えられることが定理 8.4 からわかる. ただし, \tilde{c} と $\tilde{\gamma}$ は

$$E_{\theta'}[\tilde{\phi}(\mathbf{X})] = \alpha(\theta')$$

となるように定めた. さらに, これは

$$\tilde{\phi}(\mathbf{x}) = \begin{cases} 1 & (U(\mathbf{x}) > c') \\ \tilde{\gamma} & (U(\mathbf{x}) = c') \\ 0 & (U(\mathbf{x}) < c') \end{cases}$$

とも表すことができる. ただし, $c' = \tau^{-1}(\tilde{x})$ である. $\tilde{\phi}$ の定め方から

$$\alpha(\theta') = E_{\theta'}[\tilde{\phi}(\mathbf{X})] = \Pr_{\theta'}(U(\mathbf{X}) > c') + \tilde{\gamma}\Pr_{\theta'}(U(\mathbf{X}) = c') \tag{8.19}$$

が成立している. 一方, (8.18) から

$$\alpha(\theta') = E_{\theta'}[\phi^*(\mathbf{X})] = \Pr_{\theta'}(U(\mathbf{X}) > c) + \gamma\Pr_{\theta'}(U(\mathbf{X}) = c) \tag{8.20}$$

が成立している. (8.19) と (8.20) を比較すると

$$c = c', \quad \gamma = \tilde{\gamma}$$

と c' と $\tilde{\gamma}$ を取ることができる. したがって, $\tilde{\phi}(\mathbf{X}) = \phi^*(\mathbf{X})$ である. いま, $\alpha(\theta') \equiv \phi(\mathbf{x})$ なる検定関数を考えると

$$E_{\theta'}[\phi(\mathbf{X})] = E_{\theta'}[\alpha(\theta')] = \alpha(\theta')$$

であるから, 検定関数 ϕ は検定問題

$$H_0 : \theta = \theta' \quad \text{v.s.} \quad H_1 : \theta = \theta_0$$

の有意水準 $\alpha(\theta')$ 検定である. 検定関数 $\tilde{\phi}$ はこの検定問題に対する m.p. 検定であることと (8.18) から, $\forall \theta' \in \Theta_0 \setminus \{\theta_0\}$ に対して

$$\alpha = E_{\theta_0}[\phi^*(\mathbf{X})] = E_{\theta_0}[\tilde{\phi}(\mathbf{X})] \geq E_{\theta_0}[\phi(\mathbf{X})] = \alpha(\theta') = E_{\theta'}[\phi^*(\mathbf{X})]$$

が成立することがわかる. 以上から (8.17) が確認できた. 以上の議論から, $\forall \theta \in \Theta_0$ に対して

$$E_{\theta}[\phi^*(\mathbf{X})] \leq \alpha$$

が成り立つので, ϕ^* は有意水準 α の検定となる. さらに, 任意の有意水準 ϕ の検定と $\forall \theta \in \Theta \setminus \Theta_0$ に対して

$$E_{\theta}[\phi^*(\mathbf{X})] \geq E_{\theta}[\phi(\mathbf{X})]$$

となることが示せたので, 検定関数 ϕ^* は u.m.p. 検定であることが示せた. □

例 8.14. $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Ber}(\theta)$ ($0 < \theta < 1$) とする. $\theta_0 \in (0, 1)$ を固定して, 検定問題

$$H_0 : \theta \leq \theta_0 \quad \text{v.s.} \quad H_1 : \theta > \theta_0 \tag{8.21}$$

に対する有意水準 α ($0 < \alpha < 1$) の u.m.p. 検定を求めよう. $\mathbf{X} = (X_1, X_2, \dots, X_n)$ の同時 p.m.f. は

$$p_{\theta}^{\mathbf{X}}(\mathbf{x}) = \theta^{\sum_{j=1}^n x_j} (1 - \theta)^{n - \sum_{j=1}^n x_j} \prod_{j=1}^n \mathbb{1}_{\{0,1\}}(x_j) \quad (\mathbf{x} = (x_1, x_2, \dots, x_n))$$

であるから

$$p_{\theta}^{\mathbf{X}}(\mathbf{x}) = \exp \left\{ n \log(1 - \theta) + \sum_{j=1}^n x_j \log \left(\frac{\theta}{1 - \theta} \right) \right\} \prod_{j=1}^n \mathbb{1}_{\{0,1\}}(x_j)$$

と書き直せるので, 1 母数指数型分布族で

$$A(\theta) = \log\left(\frac{\theta}{1-\theta}\right)$$

は θ の強い意味での単調増加関数である. したがって, 定理 8.12 から $U(\mathbf{x}) := \sum_{j=1}^n x_j$ に関して強い意味での単調尤度比を持つ. したがって, 検定問題 (8.21) において, 有意水準 α の u.m.p. 検定が存在し

$$\phi^*(\mathbf{x}) = \begin{cases} 1 & (\sum_{j=1}^n x_j > c) \\ \gamma & (\sum_{j=1}^n x_j = c) \\ 0 & (\sum_{j=1}^n x_j < c) \end{cases}$$

が u.m.p. 検定となる. ただし, c と γ は

$$\alpha = \mathbf{E}_{\theta_0}[\phi^*(\mathbf{X})]$$

から定まる定数である. □

例 8.15. $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N(0, \theta)$ ($0 < \theta < \infty$) とする. $\theta_0 \in (0, \infty)$ を固定して, 検定問題

$$H_0 : \theta \leq \theta_0 \quad \text{v.s.} \quad H_1 : \theta > \theta_0 \tag{8.22}$$

に対する有意水準 α ($0 < \alpha < 1$) の u.m.p. 検定を求めよう. $\mathbf{X} = (X_1, X_2, \dots, X_n)$ の同時 p.d.f. は

$$p_{\theta}^{\mathbf{X}}(\mathbf{x}) = \left(\frac{1}{\sqrt{2\pi\theta}}\right)^n \exp\left\{-\frac{1}{2\theta} \sum_{j=1}^n x_j^2\right\} \quad (\mathbf{x} = (x_1, x_2, \dots, x_n))$$

であるから, 1 母数指数型分布族で

$$A(\theta) = -\frac{1}{2\theta}$$

は θ の強い意味での単調増加関数である. したがって, 定理 8.12 から $U(\mathbf{x}) := \sum_{j=1}^n x_j^2$ に関して強い意味での単調尤度比を持つ. したがって, 検定問題 (8.21) において, 有意水準 α の u.m.p. 検定が存在し

$$\phi^*(\mathbf{x}) = \begin{cases} 1 & (\sum_{j=1}^n x_j^2 > c) \\ \gamma & (\sum_{j=1}^n x_j^2 = c) \\ 0 & (\sum_{j=1}^n x_j^2 < c) \end{cases}$$

が u.m.p. 検定となる. ただし, c と γ は

$$\alpha = \mathbf{E}_{\theta_0}[\phi^*(\mathbf{X})]$$

から定まる定数である. □

8.4.2 不偏検定

節 8.4.1 では, 1 次元の母数の片側検定問題において, u.m.p. 検定が構成できる場合があることを示した. しかし, 単調尤度比の手法は両側検定に対しては適用できない. そこで, 両側検定問題の場合において, 自然と思われる検定方式の正当化のための概念を導入する.

まず, 簡単な例をみてみよう.

例 8.16. $\mu^* \in \mathbb{R}$ とし, $X \sim N(\mu, 1)$ とする. 観測 X に基づいて, 有意水準 α ($0 < \alpha < 1$) の検定問題

$$H_0: \mu = 0 \quad \text{v.s.} \quad H_1: \mu \neq 0 \quad (8.23)$$

を考える.

u.m.p. 検定が存在しないことを確認しよう. そのために, $\mu_1 \in \mathbb{R} \setminus \{0\}$ を固定して, 検定問題

$$H_0: \mu = 0 \quad \text{v.s.} \quad H_1: \mu = \mu_1$$

を考える. 帰無仮説も対立仮説も単純であることので, 定理 8.4 を用いて m.p. 検定が構成できることに注意する. $\mu_1 > 0$ の場合には

$$\phi_1(x) = \begin{cases} 1 & (X > c') \\ 0 & (X \leq c') \end{cases}$$

は m.p. 検定となる. ただし, c' は $E_{\mu=0}[\phi_1(X)] = \alpha$ となるように定めた定数である. 一方, $\mu_1 < 0$ の場合には

$$\phi_2(x) = \begin{cases} 1 & (X < -c') \\ 0 & (X \geq -c') \end{cases}$$

は m.p. 検定となる. u.m.p. 検定はどんな対立仮説 $H_1: \mu = \mu_1$ に対しても同時に m.p. 検定でなければならない. しかし, $\mu_1 > 0$ と $\mu_1 < 0$ の場合では, m.p. 検定の形が異なるので, u.m.p. 検定は存在しないことがわかる.

検定問題 8.22 において, 自然な検定関数は

$$\phi_3(x) = \begin{cases} 1 & (|X| > c) \\ 0 & (|X| \leq c) \end{cases}$$

である. ただし, c は $E_{\mu=0}[\phi_3(X)] = \alpha$ をみたすように定めた定数である. この検定関数は, 以下で定める不偏検定の族の中で m.p. となることがわかる. \square

定義 8.17. Θ を母数空間とし, $\Theta_1, \Theta_2 \subset \Theta$ は $\Theta_1 \neq \emptyset, \Theta_2 = \emptyset, \Theta_1 \cap \Theta_2 = \emptyset, \Theta = \Theta_1 \cup \Theta_2$ をみたすとする. 母数 $\theta \in \Theta$ に対する有意水準 α ($0 < \alpha < 1$) の検定問題

$$H_0 : \theta \in \Theta_1 \quad \text{v.s.} \quad H_1 : \theta \in \Theta_1$$

を考える. さらに, X を母数 θ の母集団分布から観測とする. 有意水準 α の検定関数 ϕ は**不偏** (unbiased) であるとは

$$\beta_\phi(\theta) = E_\theta[\phi(X)] \geq \alpha \quad (\forall \theta \in \Theta_1)$$

が成り立つことである.

定義 8.18. 定義 8.17 の設定を踏襲する. 有意水準 α の不偏検定の族の中で, すべての対立仮説について検出力を最大にする検定関数 ϕ^* が存在すれば, ϕ^* を**一様最強力不偏検定** (uniformly most powerful unbiased = u.m.p.u.) という. すなわち, 有意水準 α の任意の不偏検定関数 ϕ に対して

$$\beta_{\phi^*}(\theta) \geq \beta_\phi(\theta) \quad (\forall \theta \in \Theta_1)$$

が成り立つことである.

定理 8.19. 有意水準 α ($0 < \alpha < 1$) の検定問題において, u.m.p. 検定は u.m.p.u. 検定である.

Proof. 定義 8.17 の設定を踏襲して, 証明を与える. 有意水準 α の u.m.p. 検定関数を ϕ^* とおく. すると

$$E_\theta[\phi^*(X)] \leq \alpha \quad (\theta \in \Theta_0)$$

である.

いま, $\tilde{\phi}(x) \equiv \alpha$ なる検定関数を考える. すると

$$E_\theta[\tilde{\phi}(X)] = E_\theta[\alpha] = \alpha \quad (\theta \in \Theta)$$

であることがわかる. 一方, ϕ^* は u.m.p. 検定関数なので

$$E_\theta[\phi^*(X)] \geq E_\theta[\tilde{\phi}(X)] = \alpha \quad (\theta \in \Theta_1)$$

である. したがって, ϕ^* は不偏検定である. ϕ^* は u.m.p. なので, 有意水準 α の任意の検定関数 ϕ に対して

$$\beta_{\phi^*}(\theta) \geq \beta_\phi(\theta) \quad (\forall \theta \in \Theta_1)$$

が成立する. したがって, ϕ^* は u.m.p.u. 検定である. □

定理 8.20. 標本空間 \mathbb{X}^n 上の確率変数列 X_1, X_2, \dots, X_n は指数型分布族

$$p_\theta(x) = h(x) \exp\{A(\theta)T(x) - \kappa^\vee(\theta)\} \mathbb{1}_{\mathbb{X}}(x)$$

からのランダム標本とする. ただし, $\theta \in \Theta \subset \mathbb{R}$ で, $A : \Theta \rightarrow [0, \infty)$, $T : \mathbb{X} \rightarrow [0, \infty)$, $\kappa^\vee : \Theta \rightarrow [0, \infty)$ である. $\theta_1, \theta_2 \in \Theta$, $\theta_1 < \theta_2$ に対して,

$$\Theta_0 = \Theta \cap [\theta_1, \theta_2]$$

と定め, 有意水準 α ($0 < \alpha < 1$) の検定問題

$$H_0 : \theta \in \Theta_0 \quad \text{v.s.} \quad H_1 : \theta \in \Theta \setminus \Theta_0$$

を考える. さらに, $\mathbf{x} = (x_1, x_2, \dots, x_n)$, $\mathbf{X} = (X_1, X_2, \dots, X_n)$ とおく. 検定関数 ϕ^* を

$$\phi^*(\mathbf{x}) = \begin{cases} 1 & (\sum_{j=1}^n T(x_j) < c_1 \text{ または } \sum_{j=1}^n T(x_j) > c_2) \\ \gamma & (\sum_{j=1}^n T(x_j) = c_1 \text{ または } \sum_{j=1}^n T(x_j) = c_2) \\ 0 & (\text{その他の場合}) \end{cases}$$

で定めたとき, 検定関数 ϕ^* は u.m.p.u. 検定となる. ただし, γ, c_1, c_2 は $E_{\theta_j}[\phi^*(\mathbf{X})] = \alpha$ ($j = 1, 2$) から定まる定数である.

Proof. 草間 (1975, pp.107-108) では証明は省略されている. Lehmann-Romano (2022, pp.126-128) を参照すること. \square

8.5 多重比較と FDR

8.6 区間推定の考え方

$\mathbf{X} = (X_1, X_2, \dots, X_n)^\top \sim P_\theta^{\otimes n}$ ($\theta \in \Theta \subset \mathbb{R}$) とする. $0 < \alpha < 1$ を固定する. 母数 θ に依存しない区間 $[L(\mathbf{X}), U(\mathbf{X})] \subset \Theta$ が $\forall \theta \in \Theta$ に対して

$$\Pr_\theta \left\{ L(\mathbf{X}) \leq \theta \leq U(\mathbf{X}) \right\} \geq 1 - \alpha \tag{8.24}$$

をみたすとき区間 $[L(\mathbf{X}), U(\mathbf{X})]$ を信頼係数 $(1 - \alpha)$ の θ の信頼区間 (confidence interval) という. $L(\mathbf{X}), U(\mathbf{X})$ を信頼限界 (confident limit) という. 通常 α として 0.05, 0.01, 0.1 等が用いられる. (8.24) の関係式は, たとえば 100 組の実現値を発生させると 100α 回程度は信頼区間に真の母数 θ は含まれないと考える.

例 8.21. $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N(\theta, \sigma^2)$ とする. ただし $\theta \in \mathbb{R}$ で σ^2 ($\sigma > 0$) の値は既知である. このとき標本平均 $\bar{X}_n = n^{-1} \sum_{j=1}^n X_j$ は

$$\bar{X}_n \sim N\left(\theta, \frac{\sigma^2}{n}\right)$$

となる. このことより

$$S_\theta(\mathbf{X}) = \frac{\sqrt{n}(\bar{X}_n - \theta)}{\sigma} \sim N(0, 1)$$

となる. ここで $z_{\alpha/2}$ を標準正規分布 $N(0, 1)$ の上側 $100(\alpha/2)\%$ 点とすると

$$\begin{aligned} 1 - \alpha &= \Pr_\theta \left\{ -z_{\alpha/2} \leq S_\theta(\mathbf{X}) \leq z_{\alpha/2} \right\} \\ &= \Pr_\theta \left\{ \bar{X}_n - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \leq \theta \leq \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \right\} \end{aligned}$$

なることがわかる. したがって

$$[L(\mathbf{X}), U(\mathbf{X})] = \left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}, \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \right]$$

は信頼係数 $(1 - \alpha)$ の θ の信頼区間となる. □

8.7 信頼区間の構成法

8.7.1 検定方式の反転

$\Theta \subset \mathbb{R}$ とし, $\theta_0 \in \Theta$ を取る¹. 検定問題

$$H_0 : \theta = \theta_0, \quad H_1 : \theta \neq \theta_0$$

を考える. 有意水準 α ($0 < \alpha < 1$) の検定の受容域を $A(\theta_0)$ とおく. すなわち

$$\Pr_{\theta_0} \{ \mathbf{X} \in A(\theta_0) \} \geq 1 - \alpha$$

が成り立っている. そこで $\mathbf{X} \in A(\theta_0)$ を θ_0 に関して解くことによって

$$C(\mathbf{X}) = \{ \theta \in \Theta; \mathbf{x} \in A(\theta) \}$$

¹ θ_0 を θ^* と書くべきであろうが, 後の表記上の都合でこの記号を採用した.

が得られる. 一般に $C(\mathbf{X})$ は連結区間になるという保証はない. $C(\mathbf{X})$ が連結区間となれば

$$\Pr_{\theta}\{\theta \in A(\theta)\} \geq 1 - \alpha$$

となるので, $C(\mathbf{X})$ は信頼係数 $(1 - \alpha)$ の信頼区間となる.

例 8.22. $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Ber}(\theta)$ ($0 < \theta < 1$) とする. $\theta_0 \in (0, 1)$ と固定し, 検定問題

$$H_0 : \theta = \theta_0, \quad H_1 : \theta \neq \theta_0$$

を考える. この検定問題に対する尤度比検定統計量は

$$\begin{aligned} \lambda(\mathbf{X}) &= \frac{\prod_{j=1}^n \theta_0^{X_j} (1 - \theta_0)^{1-X_j}}{\prod_{j=1}^n \bar{X}_n^{X_j} (1 - \bar{X}_n)^{1-X_j}} \\ &= \prod_{j=1}^n \left(\frac{\theta_0}{\bar{X}_n} \right)^{X_j} \left(\frac{1 - \theta_0}{1 - \bar{X}_n} \right)^{1-X_j} \end{aligned}$$

となる. これより受容域 $A(\theta_0)$ は

$$\begin{aligned} A(\theta_0) &= \left\{ \mathbf{x} \in \{0, 1\}^n; \right. \\ &\quad \left. -2 \log \lambda(\mathbf{x}) = 2n\bar{x}_n \log\left(\frac{\bar{x}_n}{\theta_0}\right) + 2n(1 - \bar{x}_n) \log\left(\frac{1 - \bar{x}_n}{1 - \theta_0}\right) \leq \chi_{1, \alpha}^2 \right\} \end{aligned}$$

となる. ただし $\bar{x}_n = n^{-1} \sum_{j=1}^n x_j$, $\mathbf{x} = (x_1, x_2, \dots, x_n)$ である. よって

$$C(\mathbf{X}) := \left\{ \theta; \bar{X}_n \log\left(\frac{\bar{X}_n}{\theta}\right) + (1 - \bar{X}_n) \log\left(\frac{1 - \bar{X}_n}{1 - \theta}\right) \leq \frac{\chi_{1, \alpha}^2}{2n} \right\}$$

となる. ただし信頼限界 $L(\mathbf{X}), U(\mathbf{X})$ を陽に求めることはできない. \square

8.7.2 枢軸量 (pivotal quantity)

一般に $Q(\mathbf{X}, \theta)$ の分布が θ に依存しないとき, $Q(\mathbf{X}, \theta)$ を枢軸量 (pivotal quantity) という. このとき

$$\Pr_{\theta}\left(a \leq Q(\mathbf{X}, \theta) \leq b\right) = 1 - \alpha$$

をみたま a, b を定めて, $a \leq Q(\mathbf{X}, \theta) \leq b$ を θ に関して解くことにより, 信頼係数 $(1 - \alpha)$ の θ の信頼区間

$$C(\mathbf{X}) = \{\theta \in \Theta; a \leq Q(\mathbf{X}, \theta) \leq b\}$$

が得られる. もちろん $C(\mathbf{X})$ が連結区間になる保証は一般的にはないが, うまく連結区間になれば, 信頼区間として使用できる.

例 8.23. $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Exp}(\theta) (\theta > 0)$ とする. すなわち, p.d.f. は

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

で与えられる. すると $2\theta \sum_{j=1}^n X_j$ は自由度 $2n$ の χ^2 分布に従うので

$$Q(\mathbf{X}, \theta) = 2\theta \sum_{j=1}^n X_j$$

とおく. よって $\chi_{2n, \alpha}^2$ を自由度 $2n$ の χ^2 分布の上側 $100\alpha\%$ 点とすると

$$\Pr_\theta \left(\chi_{2n, 1-\alpha/2}^2 \leq Q(\mathbf{X}, \theta) \leq \chi_{2n, \alpha/2}^2 \right) = 1 - \alpha$$

となるので,

$$[L(\mathbf{X}), U(\mathbf{X})] = \left[\frac{\chi_{2n, 1-\alpha/2}^2}{2 \sum_{j=1}^n X_j}, \frac{\chi_{2n, \alpha/2}^2}{2 \sum_{j=1}^n X_j} \right]$$

は信頼係数 $(1 - \alpha)$ の θ の信頼区間となる. □

問 8.1. $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Exp}(\theta) (\theta > 0)$ のとき, $2\theta \sum_{j=1}^n X_j \sim \chi_{2n}^2$ となることを示せ.

8.8 章末注釈と参考文献

節 ?? は [42, pp.62 – 67] を借用した.

8.9 演習問題

演習問題 8.1. 確率空間 $(\Omega, \mathcal{A}, \Pr)$ 上で定義された確率変数 X_1, X_2, X_3 は独立同一に母数 $\theta (0 < \theta < 1)$ の Benoulli 分布 $\text{Ber}(\theta)$ に従っているとす. 検定問題

$$H_0 : \theta = \frac{1}{2} \quad \text{vs} \quad H_1 : \theta \neq \frac{1}{2}$$

の検定問題に対して, 検定方式

$$W_1 = \left\{ (x_1, x_2, x_3) \in S; \sum_{i=1}^3 x_i \in \{0, 3\} \right\}$$

$$W_2 = \left\{ (x_1, x_2, x_3) \in S; \sum_{i=1}^3 x_i \in \{2, 3\} \right\}$$

を考える. ただし, $S = \{(x_1, x_2, x_3); x_i \in \{0, 1\} (i = 1, 2, 3)\}$ である. 検定方式 W_1, W_2 のサイズを求めよ. すなわち

$$\Pr_{\theta=1/2}((X_1, X_2, X_3) \in R_1) \quad \text{および} \quad \Pr_{\theta=1/2}((X_1, X_2, X_3) \in R_2)$$

の確率である.

記号 $p_\theta(x)$ を $\text{Ber}(\theta)$ の p.m.f. とする. Borel 集合 $A \subset \mathbb{R}^3$ に対して

$$\Pr_{\theta=1/2}((X_1, X_2, X_3) \in A) = \sum_{(x_1, x_2, x_3) \in A \cap S} p_{1/2}(x_1)p_{1/2}(x_2)p_{1/2}(x_3)$$

と定めている.

演習問題 8.2. $\theta \in \mathbb{R}$ とし, 連続型確率変数 X は p.d.f.

$$p_\theta(x) = \frac{1}{\pi\{1 + (x - \theta)^2\}} \quad (-\infty < x < \infty)$$

を持つとする. 仮説検定問題

$$\text{帰無仮説 } H_0 : \theta = 0 \quad \text{vs.} \quad \text{対立仮説 } H_1 : \theta = 1$$

を X に基づく検定する. この検定問題に対して, 棄却域

$$W := \{x \in \mathbb{R} : 1 < x < 3\}$$

を考える. 以下の問いに答えよ. ただし, $\arctan 2 = 1.107$, $\arctan 3 = 1.249$, $\pi = 3.1416$ として計算せよ.

(1) 積分

$$\int_{-\infty}^{\infty} p_\theta(x) dx$$

を計算せよ.

(2) W で定まる検定のサイズ (第 1 種の誤りの確率) α の値を小数第 3 位まで求めよ.

(3) W で定まる検定の検出力 $1 - \beta$ (第 1 種の誤りの確率) を小数第 3 位まで求めよ.

(4) 尤度比

$$\Lambda(x) = \frac{p_1(x)}{p_0(x)}$$

の $x = 0$ と $x = 1$ における値を求めよ.

(5) 不等式 $\Lambda(x) > 2$ をみたす領域を求めることにより, W で与えられる (1) で与えられた α を有意水準とする検定の中で最強力検定となることを Neyman-Pearson の補題を用いて証明せよ.

演習問題 8.3. 連続型確率変数列 X_1, X_2, \dots, X_{16} は正規分布 $N(\mu, 3^2)$ からのランダム標本とする. ただし, $-\infty < \mu < \infty$ で, これらの確率変数は確率空間 $(\Omega, \mathcal{A}, \Pr)$ 上で定義されたものとする.

- (1) 標本平均 $\bar{X}_{16} = (1/16)(X_1 + \dots + X_{16})$ の期待値と分散を計算することにより, \bar{X}_{16} の分布を述べよ.
- (2) $a > 0, b$ を定数とし, $Z = a\bar{X}_{16} + b$ としたとき, Z の分布が標準正規分布になるように定数 a, b を定めよ.
- (3) 信頼係数 90% の μ の信頼区間を構成せよ.

演習問題 8.4. 連続型確率変数列 X_1, X_2, X_3 は正規母集団 $N(\mu, 3)$ からの標本の大きさ 3 のランダム標本とする. 次の仮説検定問題を考える:

$$H_0: \mu = 0 \quad \text{vs.} \quad H_1: \mu = 2$$

このとき, 以下の問いに答えよ.

- (1) H_0 と H_1 は単純仮説か複合仮説かを答えよ.
- (2) $\bar{X}_3 = (1/3)(X_1 + X_2 + X_3)$ とする. \bar{X}_3 は帰無仮説 $H_0: \mu = 0$ のもとでどのようなものになるかを答えよ. 理由も述べること.
- (3) H_0 の棄却域を

$$C(t) = \{(x_1, x_2, x_3) \in \mathbb{R}^3; x_1 + x_2 + x_3 > t\}$$

としたとき, $C(t)$ が有意水準 0.1 の検定の棄却域になるように t をひとつ定めよ. t を求めるときには小数第 3 位を四捨五入せよ.

- (4) 棄却域 $C(t)$ の検出力を対立仮説 $H_1: \mu = 2$ のもとで求めよ. ただし, 解答は関数 $\Phi(x) = \int_{-\infty}^x (1/\sqrt{2\pi})e^{-t^2/2} dt$ および t を用いて表現せよ. すなわち, 具体的な値を計算しなくともよい.

第9章 Bayes 的推測

9.1 Bayes 的推測の考え方

前節までの統計的手法は頻度論的手法 (frequentist methods) と呼ばれるものである。頻度論的アプローチの考え方は以下のようにまとめられる。

- 【F1】 確率は大量に観測されたデータの度数の極限と考える。
- 【F2】 真の母数は未知だが、固定した値であると考え。したがって真の母数は変動しないし、真の母数に対する意味のある確率的な主張はない。
- 【F3】 推測手法は大量にデータが観測されれば、うまく機能する保証があるように設計されている。たとえば信頼係数 0.95 の信頼区間は、大量に観測されたデータの極限的な頻度において、0.95 の信頼度が保証されている。

この考え方とは異なるアプローチの 1 つが Bayes 的推測である。Bayes 的アプローチは以下のような考え方に基づいている。

- 【B1】 確率は極限的な頻度ではなく、信頼の程度を表現するものであると考える。データはある確率変数の実現値と考えるだけでなく、それ以外の色々なものも確率変数の実現値と考える。
- 【B2】 真の母数は固定された定数にも関わらず、真の母数の確率的な主張をする。
- 【B3】 真の母数に対して確率分布を想定し、未知の母数の推測を行う。点推定値や区間推定値もこの分布からの実現値と考える。

確率に主観的な見方を導入する Bayes 的アプローチには、頻度論的立場の主流派 (古典的な) 統計学者からの大きな批判がある。しかし、統計学の隣接分野である機械学習やデータマイニングの分野では Bayes 的なアプローチが広く採用されている。

哲学的な議論は脇において、Bayes 的推測がどのような形式で行われるかをこの節ではみていく。

記法に関する注意 母数 θ で添え字付けられた p.d.f./p.m.f. を $p_\theta(\cdot)$ と前に章まで記してきた. この章では, 条件付きのあることを強調するために, $p(\cdot | \theta)$ のように記すことにする. □

9.2 Bayes 的推測手法

Bayes 的推測は以下のステップに従い行われる. 母数モデルを $\{P_\theta; \theta \in \Theta\}$ とする.

1. 事前分布 (prior distribution) と呼ばれる母数 θ についての分布 $p^\Theta(\theta)$ を想定する.
2. 母数 θ に与えられたときにデータの分布を表現する条件付きの分布 (頻度論的な立場では真の分布) $p^{\mathbf{X}|\Theta}(\mathbf{x}|\theta)$ を想定する.
3. 観測されたデータ $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ に基づき, θ の信頼度をアップデートする. すなわち, **事後分布** (posterior distribution) $p^{\Theta|\mathbf{X}}(\theta|\mathbf{x})$ を求める. ただし $\mathbf{x} = (x_1, x_2, \dots, x_n)$ とした.

そして, 事後分布に基づき, 母数の推測を行う. 3 番目のステップがどのように行われるかを θ の事前分布とデータ \mathbf{X} の分布が共に離散型の場合で説明する.

真の母数 θ は確率変数 Θ の実現値と考える. いまは離散型確率変数の設定なので

$$\begin{aligned} \Pr(\Theta = \theta | \mathbf{X} = \mathbf{x}) &= \frac{\Pr(\Theta = \theta, \mathbf{X} = \mathbf{x})}{\Pr(\mathbf{X} = \mathbf{x})} \\ &= \frac{\Pr(\Theta = \theta, \mathbf{X} = \mathbf{x})}{\sum_{\theta} \Pr(\mathbf{X} = \mathbf{x} | \Theta = \theta) \Pr(\Theta = \theta)} \end{aligned}$$

となる. 最後の等号は全確率の法則 (補題 1.11) を用いた. これを連続型分布の場合に形式的に書き直すと

$$p^{\Theta|\mathbf{X}}(\theta|\mathbf{x}) = \frac{p^{\mathbf{X}|\Theta}(\mathbf{x}|\theta)p^\Theta(\theta)}{\int p^{\mathbf{X}|\Theta}(\mathbf{x}|\theta)p^\Theta(\theta) d\theta}$$

と書ける. ただし $p^{\Theta|\mathbf{X}}(\cdot|\mathbf{x})$ は $\mathbf{X} = \mathbf{x}$ が与えられたときの Θ の条件付き p.d.f. であり, θ の事後分布である. $p^{\mathbf{X}|\Theta}(\mathbf{x}|\theta)$ は真の母数が θ のときのデータの分布の p.d.f. で, $p^\Theta(\theta)$ は Θ の分布の p.d.f. である.

n 個のランダム標本 $\mathbf{X} = (X_1, X_2, \dots, X_n)$ に対して, 真の母数が θ のときの \mathbf{X} の同時 p.d.f. $p^{\mathbf{X}|\Theta}(\cdot|\theta)$ とし, $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ を観測したときの尤度関数 $\text{lik}_n(\theta)$ を

$$\text{lik}_n(\theta) = p^{\mathbf{X}|\Theta}(\mathbf{x}|\theta)$$

と書くことにする. よって

$$p^{\Theta|\mathbf{X}}(\theta|\mathbf{x}) = \frac{p^{\mathbf{X}|\Theta}(\mathbf{x}|\theta)p^{\Theta}(\theta)}{\int p^{\mathbf{X}|\Theta}(\mathbf{x}|\theta)p^{\Theta}(\theta) d\theta} =: \frac{\text{lik}_n(\theta)p^{\Theta}(\theta)}{c_n} \propto \text{lik}_n(\theta)p^{\Theta}(\theta)$$

と表すことができる. ただし

$$c_n := \int \text{lik}_n(\theta)p^{\Theta}(\theta) d\theta$$

は正規化定数と呼ばれる値である. c_n は θ に依存せず, データの実現値に依存した値である. よって事後分布の条件付き p.d.f. は事前分布と尤度関数の積の定数倍である.

$$p^{\Theta|\mathbf{X}}(\theta|\mathbf{x}) \propto \text{lik}_n(\theta)p^{\Theta}(\theta).$$

c_n を無視してもよいのだろうか? 必要なときは求めることができるので問題ない. 事後分布の平均やモード (最頻値) を推測に用いることが多い. たとえば,

$$\int \{y - \theta\}^2 p^{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta$$

を最小にする y の値を $\bar{\theta}$ と書くと

$$\bar{\theta} = \frac{\int \theta p^{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta}{\int \theta^2 p^{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta} = \frac{\int \theta \text{lik}_n(\theta)p^{\Theta}(\theta) d\theta}{\int \theta^2 \text{lik}_n(\theta)p^{\Theta}(\theta) d\theta}$$

と表現できる. また信頼区間であれば, $0 < \alpha < 1/2$ に対して

$$\int_{-\infty}^a p^{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta = \int_b^{\infty} p^{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta = \frac{\alpha}{2}$$

をみたま a, b を求めると開区間 $C = (a, b)$ は

$$\Pr(\theta \in C) = 1 - \alpha$$

をみたま事後信頼区間となる.

例 9.1. $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Ber}(\theta)$ ($0 < \theta < 1$) とし, 事前分布は $(0, 1)$ 上の一様分布とする. すなわち

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

である. すると

$$\begin{aligned} p^{\Theta|\mathbf{X}}(\theta|\mathbf{x}) &\propto p^{\Theta}(\theta)\text{lik}_n(\theta) = \prod_{j=1}^n \theta^{x_j}(1-\theta)^{1-x_j} \mathbb{1}_{\{0,1\}}(x_j) \\ &= \theta^{s_n}(1-\theta)^{n-s_n} \mathbb{1}_{\{0,1,\dots,n\}}(s_n) \end{aligned}$$

となる. ただし $s_n = \sum_{j=1}^n x_j$ である. したがって

$$\Theta|\mathbf{X} = \mathbf{x} \sim \text{Beta}(s_n + 1, n - s_n + 1)$$

となることがわかる.

一方 $\alpha > 0, \beta > 0$ とし

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

とする. ただし $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x} dx$ である. すなわち θ の事前分布は $B(\alpha, \beta)$ である. このとき

$$\begin{aligned} p^{\Theta|\mathbf{X}}(\theta|\mathbf{x}) &= \frac{\Gamma(n+\alpha+\beta)}{\Gamma(s_n+\alpha)\Gamma(n-s_n+\beta)} \theta^{(s_n+\alpha)-1}(1-\theta)^{(n-s_n+\beta)-1} \mathbb{1}_{\{0,1,\dots,n\}}(s_n) \end{aligned} \tag{9.1}$$

となる. したがって

$$\Theta|\mathbf{X} = \mathbf{x} \sim B(s_n + \alpha, n - s_n + \beta)$$

がわかる. このとき

$$\bar{\theta} = \int \theta p^{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta$$

とおくと

$$\bar{\theta} = \frac{s_n + \alpha}{n + \alpha + \beta}$$

を得る. $\hat{\theta} = \frac{s_n}{n}, \tilde{\theta} = \frac{1}{2}$ とおくと

$$\bar{\theta} = \lambda_n \hat{\theta} + (1 - \lambda_n) \tilde{\theta}, \quad \lambda_n = \frac{n}{n + \alpha + \beta}$$

と書ける.

信頼係数 0.95 の θ の事後信頼区間 $C = (a, b)$ は

$$\int_a^b p^{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta = 0.95$$

をみたく a, b を数値計算で求めればよい. □

問 9.1. $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Ber}(\theta)$ ($0 < \theta < 1$) とし,

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

とする. このとき, $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ が与えられたときの Θ の条件付き p.d.f. が (9.1) で与えられることを示せ.

問 9.2. $X \sim \text{B}(\alpha, \beta)$ のとき, $E[X] = \frac{\alpha}{\alpha + \beta}$ となることを示せ.

例 9.2. $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{N}(\mu, \sigma^2)$ とする. ただし $\mu \in \mathbb{R}$ で σ^2 は既知とする. 事前分布として

$$\Theta \sim \text{N}(a, b^2)$$

を仮定する. ただし $a \in \mathbb{R}, 0 < b < \infty$ である. すると

$$\Theta | \mathbf{X} = \mathbf{x} \sim \text{N}(\bar{\theta}, \tau^2), \tag{9.2}$$

$$\bar{\theta} = w\bar{x}_n + (1-w)a, \quad \bar{x}_n = \frac{1}{n} \sum_{j=1}^n x_j,$$

$$w = \frac{\frac{1}{\text{se}^2}}{\frac{1}{\text{se}^2} + \frac{1}{b^2}}, \quad \frac{1}{\tau^2} = \frac{1}{\text{se}^2} + \frac{1}{b^2}, \quad \text{se} = \frac{\sigma}{\sqrt{n}}$$

となる. 問 9.3 を参照のこと. ただし, $\mathbf{X} = (X_1, X_2, \dots, X_n), \mathbf{x} = (x_1, x_2, \dots, x_n)$ とした.

このように事前分布と事後分布が同じ母数モデルに属するとき, 事前分布はこのモデルに**随伴する** (conjugate) という. または, このような事前分布を**随伴事前分布**という.

$n \rightarrow \infty$ のとき $w \rightarrow 1$ かつ $\frac{\tau}{\text{se}} \rightarrow 1$ となる. 標本数が大きいとき

$$\Theta | \mathbf{X} = \mathbf{x} \approx \text{N}(\bar{x}_n, \text{se}^2)$$

となる. また n を固定する. $b \rightarrow \infty$ としたとき

$$\Theta | \mathbf{X} = \mathbf{x} \approx \text{N}(\bar{x}_n, \text{se}^2)$$

となる. これは一様な事前分布に対応するものである.

区間 $C = (c, d)$ は

$$\Pr(\theta \in C | \mathbf{X} = \mathbf{x}) = 0.95$$

をみたすものとする. したがって

$$\Pr(\theta < c | \mathbf{X} = \mathbf{x}) = 0.025, \quad \Pr(\theta > d | \mathbf{X} = \mathbf{x}) = 0.025$$

となるように c, d を選べばよい. このことから c を以下をみたすように選べばよい.

$$\Pr(\theta < c | \mathbf{X} = \mathbf{x}) = \Pr\left(\frac{\theta - \bar{\theta}}{\tau} < \frac{c - \bar{\theta}}{\tau} \middle| \mathbf{X} = \mathbf{x}\right) = \Pr\left(Z < \frac{c - \bar{\theta}}{\tau}\right).$$

ただし $Z \sim N(0, 1)$ である. さらに

$$\Pr(Z < -1.96) = 0.025$$

なので

$$\frac{c - \bar{\theta}}{\tau} = -1.96$$

とすればよい. よって

$$c = \bar{\theta} - 1.96\tau$$

とする. 同様に

$$d = \bar{\theta} + 1.96\tau$$

を得る. これらのことから信頼係数 0.95 の Bayes 的信頼区間は $(\bar{\theta} - 1.96\tau, \bar{\theta} + 1.96\tau)$ となる. さらに $\hat{\theta} \approx \bar{\theta}$ かつ $\tau \approx \text{se}$ なので信頼係数 0.95 の Bayes 的信頼区間は近似的に $(\hat{\theta} - 1.96\text{se}, \hat{\theta} + 1.96\text{se})$ となり, 頻度論的信頼区間と同じになる.

問 9.3. $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$ とする. ただし $\mu \in \mathbb{R}$ で σ^2 は既知とする. さらに, 事前分布として

$$\Theta \sim N(a, b^2)$$

を仮定する. ただし $a \in \mathbb{R}, 0 < b < \infty$ である. このとき, $\mathbf{X} = \mathbf{x}$ が与えられたときの Θ の条件付き分布が (9.2) で与えられることを確かめよ.

9.3 事前分布の選択について

Bayes 的推測における重要な疑問は, 事前分布 π の選択である. ある学派は, 母数 θ に関する主観的な意見を反映させて事前分布を選択すればよいと主張している. ある場合には, このやり方は実行可能かもしれない. しかし, 多変数の母数を含む複雑な問題に対しては, 実行可能性に欠けることもあるだろう. さらに, 客観的かつ科学的な決定を目標とするデータ解析において, 主観的な意見を入り込ませることは相容れないという見方ができるであろう.

別の方策として, ある種の無情報事前分布を用いるやり方がある. たとえば, Bernoulli 試行において, 母数 $\theta (0 < \theta < 1)$ に対する事前分布 π を

$$p^\Theta(\theta) \equiv 1$$

と定まる. このとき, 形式的な計算から

$$\Theta | \mathcal{D}_n \sim \text{Beta}(s_n + 1, n - s_n + 1)$$

を得る. ただし, $\mathcal{D}_n = \{x_1, x_2, \dots, x_n\}$ で $s_n = \sum_{j=1}^n x_j$ である.

9.3.1 一様事前分布

$X \sim N(\theta, \sigma^2)$ とする. ただし, σ は既知とする. いま, \mathcal{D}_n を $N(\theta, \sigma^2)$ からの標本の大きさが n 個のランダム標本の実現値とする. θ の事前分布 p^Θ を

$$p^\Theta(\theta) \propto c$$

を採用する. ただし, $c > 0$ は定数である. このとき

$$\int p^\Theta(\theta) d\theta = \infty$$

となるので, p^Θ は p.d.f. ではない. しかし, 形式的な計算により

$$p^{\Theta|X}(\theta | \mathcal{D}_n) \propto \text{lik}_n(\theta) p^\Theta(\theta) \propto \text{lik}_n(\theta)$$

となることがわかる. $\bar{x}_n = (1/n) \sum_{j=1}^n x_j$ とおいたとき

$$\Theta | \mathcal{D}_n \sim N\left(\bar{x}_n, \frac{\sigma^2}{n}\right)$$

となることがわかる. なぜならば

$$\begin{aligned} \text{lik}_n(\theta) p^\Theta(\theta) &= c \prod_{j=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_j - \theta)^2}{2\sigma^2}\right\} \\ &= c \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\frac{\sum_{j=1}^n (x_j - \bar{x}_n)^2}{2\sigma^2} - \frac{(\bar{x}_n - \theta)^2}{2\sigma^2/n}\right\} \\ &\propto \exp\left\{-\frac{(\bar{x}_n - \theta)^2}{2\sigma^2/n}\right\} \end{aligned}$$

であるからである. 以上の議論から Bayes 的推定量は頻度論的な推定量である標本平均と同じになる.

しかし, 一様事前分布は母数の変換に関して不変ではない. たとえば, $X \sim \text{Ber}(\theta)$ ($0 < \theta < 1$) とし, 事前分布 p^Θ として

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

を採用する. いま, 母数の変換 Ψ を

$$\Psi := \Psi(\theta) = \log\left(\frac{\theta}{1-\theta}\right) \Leftrightarrow \Psi := \log\left(\frac{\Theta}{1-\Theta}\right)$$

と定める. すると

$$\begin{aligned} p^\Psi(\psi) &= \frac{d}{d\psi} \Pr^\Psi(\Psi \leq \psi) = \frac{d}{d\psi} \Pr^\Psi\left(\log\left(\frac{\Theta}{1-\Theta}\right) \leq \psi\right) \\ &= \frac{d}{d\psi} \Pr\left(\Theta \leq \frac{e^\psi}{1+e^\psi}\right) = \frac{d}{d\psi} \left(\frac{e^\psi}{1+e^\psi}\right) = \frac{e^\psi}{(1+e^\psi)^2} \quad (\psi \in \mathbb{R}) \end{aligned}$$

となるので, Θ の分布は一様分布であるので, Θ についての事前情報はな
い. Θ に関して無知であれば, Θ の変換である Ψ に関して事前情報は
ないはずである. しかし, Ψ の分布は一様分布でないので, Ψ についての
事前情報があることになる. したがって, Ψ についても事前情報がないこ
とがないことと p^Ψ の事前分布の間には齟齬が生じることになる.

9.3.2 Jeffreys 事前分布

母数の変換に関して不変な事前分布を定めるために, Jeffreys を

$$p^\Theta(\theta) \propto \sqrt{|\mathcal{F}(\theta)|}$$

で定める. ここで, $\mathcal{F}(\theta)$ は Fisher 情報行列で

$$\mathcal{F}(\theta) = \mathbb{E} \left[\left(\frac{\partial}{\partial \theta} \log \text{lik}_n(\theta | \mathbf{X}) \right) \left(\frac{\partial}{\partial \theta} \log \text{lik}_n(\theta | \mathbf{X}) \right)^\top \middle| \Theta = \theta \right]$$

で定義され, $|\cdot|$ は行列式を表す. また, 期待値は \mathbf{X} の分布に関して取っ
ていることに注意せよ.

定理 9.3. Jeffreys 事前分布は母数の変換に対して不変である.

Proof.

□

例 9.4. Benoulli 分布を考える. すると

$$\begin{aligned} \mathcal{F}(\theta) &= \mathbb{E} \left[\left(\frac{d}{d\theta} \log(\theta^X (1-\theta)^{1-X}) \right)^2 \middle| \Theta = \theta \right] \\ &= \mathbb{E} \left[\left(\frac{X}{\theta} - \frac{1-X}{1-\theta} \right)^2 \middle| \Theta = \theta \right] \\ &= \mathbb{E} \left[\left(\frac{X-\theta}{\theta(1-\theta)} \right)^2 \middle| \Theta = \theta \right] = \frac{1}{\theta(1-\theta)} \end{aligned}$$

となる. よって, Jeffreys 事前分布 p^\ominus は

$$p^\ominus(\theta) \propto \sqrt{|\mathcal{F}(\theta)|} = \theta^{-1/2}(1-\theta)^{-1/2}$$

となる. これは Beta(1/2, 1/2) の p.d.f. であり, 一様な分布に非常に近い. □

9.3.3 共役事前分布

定義 9.5. 事前分布は共役であるとは, 事後分布が事前分布の分布族に属するときをいう. すなわち, 事前分布の分布族を \mathcal{P} としたとき

$$p^{\ominus|\mathbf{X}}(\cdot|\mathbf{x}) \in \mathcal{P}$$

が成立することである. ただし, $p_\theta^{\mathbf{X}}$ を標本分布の p.d.f. で

$$p^{\ominus|\mathbf{X}}(\theta|\mathbf{x}) = \frac{p_\theta^{\mathbf{X}|\ominus}(\mathbf{x})p^\ominus(\theta)}{\int p_\theta^{\mathbf{X}|\ominus}(\mathbf{x})p^\ominus(\theta) d\theta}$$

である.

指数型分布族に対する共役事前分布を導出してみよう. σ 有限な測度 μ に関する p.d.f. $p^{\mathbf{X}|\ominus}(\cdot|\theta)$ は

$$p^{\mathbf{X}|\ominus}(\mathbf{x}|\theta) = \exp\{\boldsymbol{\theta}^\top \mathbf{x} - \kappa(\boldsymbol{\theta})\} \tag{9.3}$$

で与えられたとする. ただし, $\boldsymbol{\theta} \in \mathbb{R}^d$ で, 母数空間 $\Theta \subset \mathbb{R}^d$ は部分集合

$$\left\{ \boldsymbol{\theta} \in \mathbb{R}^d; \kappa(\boldsymbol{\theta}) = \log\left(\int \exp\{\boldsymbol{\theta}^\top \mathbf{x} - \kappa(\mathbf{x})\} d\mu(\mathbf{x})\right) < \infty \right\}$$

の内部で定義する.

定理 9.6. (9.3) で与えられる指数型分布族に対する共役事前分布の p.d.f. は

$$p_{\mathbf{x}_0, n_0}^\ominus(\boldsymbol{\theta}) = \frac{\exp\left(n_0 \mathbf{x}_0^\top \boldsymbol{\theta} - n_0 \kappa(\boldsymbol{\theta})\right)}{\int \exp\left(n_0 \mathbf{x}_0^\top \boldsymbol{\theta} - n_0 \kappa(\boldsymbol{\theta})\right) d\mu(\boldsymbol{\theta})}$$

で与えられる. ただし, $\mathbf{x}_0 \in \mathbb{R}^d$, $n_0 \in \mathbb{R}$ である.

Proof. 以下の計算からわかる.

$$\begin{aligned} p^{\Theta|\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta})p_{\mathbf{x}_0, n_0}^{\Theta}(\boldsymbol{\theta}) &\propto \exp\left(\boldsymbol{\theta}^\top \mathbf{x} - \kappa(\boldsymbol{\theta})\right) \exp\left(n_0 \mathbf{x}_0^\top \boldsymbol{\theta} - n_0 \kappa(\boldsymbol{\theta})\right) \\ &= \exp\left((\mathbf{x} + \mathbf{x}_0)^\top \boldsymbol{\theta} - (1 + n_0)\kappa(\boldsymbol{\theta})\right) \\ &= \exp\left\{(1 + n_0)\left(\frac{1}{1 + n_0}\mathbf{x} + \frac{n_0}{1 + n_0}\mathbf{x}_0\right)^\top - (1 + n_0)\kappa(\boldsymbol{\theta})\right\} \\ &\propto p_{\frac{1}{1+n_0}\mathbf{x} + \frac{n_0}{1+n_0}\mathbf{x}_0, 1+n_0}^{\Theta}(\boldsymbol{\theta}). \end{aligned}$$

□

$\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ を (9.3) からの標本の大きさが n のランダム標本とする. すると $(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$ の同時 p.d.f. は

$$\prod_{j=1}^n p^{\mathbf{X}|\Theta}(\mathbf{x}_j|\boldsymbol{\theta}) = \exp\left(n\boldsymbol{\theta}^\top \bar{\mathbf{x}}_n - n\kappa(\boldsymbol{\theta})\right), \quad \bar{\mathbf{x}}_n = \frac{1}{n} \sum_{j=1}^n \mathbf{x}_j$$

となるので

$$\begin{aligned} p^{\Theta|(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)}(\boldsymbol{\theta}|\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) &\propto \exp\left(n\boldsymbol{\theta}^\top \bar{\mathbf{x}}_n - n\kappa(\boldsymbol{\theta})\right) \exp\left(n_0 \mathbf{x}_0^\top \boldsymbol{\theta} - n_0 \kappa(\boldsymbol{\theta})\right) \\ &= \exp\left((n + n_0)\left(\frac{n}{1 + n_0}\bar{\mathbf{x}}_n + \frac{n_0}{1 + n_0}\mathbf{x}_0\right)^\top - (n + n_0)\kappa(\boldsymbol{\theta})\right) \end{aligned}$$

となる. よって, 事後分布の p.d.f. は $p_{\frac{n}{n+n_0}\bar{\mathbf{x}}_n + \frac{n_0}{n+n_0}\mathbf{x}_0, n+n_0}^{\Theta}(\boldsymbol{\theta})$ で与えられることがわかる.

いま, $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_d)^\top$ と $\nabla = \left(\frac{\partial}{\partial \theta_1}, \frac{\partial}{\partial \theta_2}, \dots, \frac{\partial}{\partial \theta_d}\right)^\top$ としたとき

$$\nabla p_{\mathbf{x}_0, n_0}^{\Theta}(\boldsymbol{\theta}) = n_0(\mathbf{x}_0 - \nabla \kappa(\boldsymbol{\theta}))p_{\mathbf{x}_0, n_0}^{\Theta}(\boldsymbol{\theta})$$

となる. 一方

$$\int \nabla p_{\mathbf{x}_0, n_0}^{\Theta}(\boldsymbol{\theta}) d\mu(\boldsymbol{\theta}) = \nabla \underbrace{\int p_{\mathbf{x}_0, n_0}^{\Theta}(\boldsymbol{\theta}) d\mu(\boldsymbol{\theta})}_{=1} = \mathbf{0}_d$$

となる. これらの 2 つの等式から

$$\begin{aligned} E[\nabla \kappa(\boldsymbol{\theta})] &= E\left[\mathbf{x}_0 - \frac{1}{n_0} \nabla p_{\mathbf{x}_0, n_0}^{\Theta}(\boldsymbol{\theta})\right] \\ &= \mathbf{x}_0 - \frac{1}{n_0} \int \nabla p_{\mathbf{x}_0, n_0}^{\Theta}(\boldsymbol{\theta}) d\mu(\boldsymbol{\theta}) \\ &= \mathbf{x}_0 \end{aligned}$$

を得る. ただし, 期待値は $p_{x_0, n_0}^\theta(\theta)$ に関してである. もっと一般的には

$$E[\nabla \kappa(\theta) | \mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n] = \frac{n}{n+n_0} \bar{\mathbf{X}}_n + \frac{n_0}{n+n_0} \mathbf{x}_0$$

となる.

9.4 1次元の van Trees 不等式

母数空間 Θ を \mathbb{R} の開部分集合とする状態空間 $(\mathbb{X}, \mathcal{B})$ 上の統計的モデル

$$\mathcal{P} := \{P_\theta; \theta \in \Theta, P_\theta \text{ は } (\mathbb{X}, \mathcal{B}) \text{ 上の確率分布で, } P_\theta \ll \mu\}$$

を考える. ただし, μ は $(\mathbb{X}, \mathcal{B})$ 上の σ 有限な測度である. さらに

$$p_\theta := \frac{dP_\theta}{d\mu}$$

とし, 写像

$$\Theta \times \mathbb{X} \ni (\theta, x) \mapsto p_\theta(x)$$

は可測¹と仮定する. さらに

$$\sqrt{p_\theta} \in L^2(\mu)$$

と仮定する. 以降, $\|\cdot\|_\mu$ は $L^2(\mu)$ における L^2 ノルムとする. すなわち, $L^2(\mu)$ の元 $g: \mathbb{X} \rightarrow \mathbb{R}$ に対して

$$\|g\|_\mu = \sqrt{\int_{\mathbb{X}} g^2(x) d\mu(x)}$$

と定める. また, \mathbb{X} のノルムを $\|\cdot\|$ と記すことにする.

定義 9.7 (L^2 における可微分性). σ 有限な測度 μ によって支配される統計的モデル \mathcal{P} は $\theta_0 \in \Theta$ で $L^2(\mu)$ に関して微分可能であるとは, 任意の $\sqrt{p_\theta} \in L^2(\mu)$ に対して, ある $\dot{\sqrt{p}}_{\theta_0} \in L^2(\mu)$ が存在して,

$$\|\sqrt{p_\theta} - \sqrt{p_{\theta_0}} - (\theta - \theta_0) \dot{\sqrt{p}}_{\theta_0}\|_\mu = o(|\theta - \theta_0|) \quad (\theta \rightarrow \theta_0)$$

をみたすときをいう. ここで, $\dot{\sqrt{p}}_{\theta_0}$ を θ_0 における $L^2(\mu)$ 微分とよぶ.

¹ $\|\cdot\|_\mu$ によって誘導される位相における開集合族の Borel 集合族に関して可測と考えるのが自然であろう.

注意 9.8. 統計的モデル \mathcal{P} が $L^2(\mu)$ に関して微分可能のとき, $\theta_0 (\in \Theta)$ における Fisher 情報量 $\mathcal{F}_{\mathcal{P}}(\theta_0)$ は

$$\begin{aligned} \mathcal{F}_{\mathcal{P}}(\theta_0) &= \int_{\mathbb{X}} \left\{ \frac{d}{d\theta} \log p_{\theta}(x) \Big|_{\theta=\theta_0} \right\}^2 p_{\theta_0}(x) d\mu(x) \\ &= 4 \int_{\mathbb{X}} \left\{ \dot{\sqrt{p}}_{\theta_0}(x) \right\}^2 d\mu(x) = \int_{\mathbb{X}} \left\{ \frac{\dot{p}_{\theta}(x)}{p_{\theta}(x)} \right\}^2 d\mu(x) \end{aligned}$$

と表現できる.

定義 9.9. 開集合 $\Theta (\subset \mathbb{R})$ 上に集中する確率分布 Q はよい振る舞いの事前分布であるとは, Q は \mathbb{R} 上の Lebesgue 測度に関して絶対連続な密度関数 q を持ち, さらにほとんど至る所存在する導関数 \dot{q} を持ち

$$\mathcal{F}_Q := \int_{\Theta} \{ \dot{q}(\theta) \}^2 \frac{\mathbb{1}\{q(\theta) > 0\}}{q(\theta)} dm(\theta) < \infty$$

をみたすときをいう. ただし, m は \mathbb{R} 上の Lebesgue 測度である.

統計量 $T : \mathbb{X} \rightarrow \mathbb{R}$ は関数 $\psi(\theta)$ の推定量とする. ここで, ψ は絶対連続で, ほとんど至る所導関数 $\dot{\psi}$ をもつとする. E_{θ} は P_{θ} に関する期待値としたとき, van Trees 不等式は

$$\int_{\Theta} E_{\theta}[\{T - \psi(\theta)\}^2] dQ(\theta) \geq \frac{\left(\int_{\Theta} \dot{\psi}(\theta) dQ(\theta) \right)^2}{\mathcal{F}_Q + \int_{\Theta} \mathcal{F}_{\mathcal{P}}(\theta) dQ(\theta)} \quad (9.4)$$

と表現できる. この不等式のために以下の仮定をおく.

仮定 9.10. (1) Θ は \mathbb{R} の開集合.

(2) 統計的モデル $\mathcal{P} = \{P_{\theta}; \theta \in \Theta\}$ は $(\mathbb{X}, \mathcal{B})$ 上の σ 有限測度 μ によって支配される.

(3) μ に関する P_{θ} の密度を $p_{\theta} = \frac{dP_{\theta}}{d\mu}$ と書いたとき, 写像

$$\Theta \times \mathbb{X} \ni (\theta, x) \mapsto p_{\theta}(x)$$

は可測である.

(4) \mathcal{P} は $\Theta \cap \text{supp}(q)$ 上のほとんど至る所で可微分である. ただし, $\text{supp}(q) := \left\{ \theta \in \Theta; \frac{dQ}{dm} = q(\theta) > 0 \right\}$ であり, m は \mathbb{R} 上の Lebesgue 測度である.

(5) 関数 $\psi : \Theta \rightarrow \mathbb{R}$ は絶対連続である.

(6) 関数 ψ^2 と $\dot{\psi}$ は Q 可積である.

(7)

$$\int_{\Theta} E_{\theta}[T^2] dQ(\theta) < \infty; \quad \int_{\Theta} \mathcal{F}_{\mathcal{P}}(\theta) dQ(\theta) < \infty$$

である.

(8) 任意の $B \in \mathcal{B}$ に対して, 関数

$$\Theta \cap \text{supp}(\mathbf{q}) \ni \theta \mapsto P_{\theta}(B)$$

は絶対連続である.

(9) θ が Θ の有限な境界に近づくととき, $\mathbf{q}(\theta) \rightarrow 0$ が成立する.

定理 9.11. 仮定 9.10 が成立し, $\mathcal{F}_{\mathcal{Q}} > 0$ とする. このとき, van Trees 不等式 (9.4) が成立する.

証明のために次の補題を準備する.

補題 9.12. 統計的モデル $\mathcal{P} = \{P_{\theta}; \theta \in \Theta, P_{\theta} \ll \mu\}$ は $\theta_0 \in \Theta$ において, $L^2(\mu)$ に関して可微分とする. 有界な統計量 $T: \mathbb{X} \rightarrow \mathbb{R}$ を考える. すなわち, ある $M > 0$ が存在して, $\mu\{|T| > M\} = 0$ である. このとき

$$\gamma_T: \Theta \ni \theta \mapsto E_{\theta}[T]$$

は θ_0 において微分可能で, 導関数

$$\dot{\gamma}_T(\theta_0) = 2 \int_{\mathbb{X}} \dot{\sqrt{p}}_{\theta_0}(x) \sqrt{p}_{\theta_0}(x) T(x) d\mu(x) \quad (9.5)$$

をもつ.

Proof. 導関数を求めるために

$$p_{\theta} - p_{\theta_0} - 2(\theta - \theta_0) \dot{\sqrt{p}}_{\theta_0} \sqrt{p}_{\theta_0}$$

を評価する. そのために, $r_{\theta} = \sqrt{p}_{\theta} - \sqrt{p}_{\theta_0} - (\theta - \theta_0) \dot{\sqrt{p}}_{\theta_0}$ とおく. すると

$$\begin{aligned} p_{\theta} - p_{\theta_0} - 2(\theta - \theta_0) \dot{\sqrt{p}}_{\theta_0} \sqrt{p}_{\theta_0} &= \{\sqrt{p}_{\theta}\}^2 - \{\sqrt{p}_{\theta_0}\}^2 - 2(\theta - \theta_0) \dot{\sqrt{p}}_{\theta_0} \sqrt{p}_{\theta_0} \\ &= \{\sqrt{p}_{\theta_0} + (\theta - \theta_0) \dot{\sqrt{p}}_{\theta_0} + r_{\theta}\}^2 - \{\sqrt{p}_{\theta_0}\}^2 \\ &\quad - 2(\theta - \theta_0) \dot{\sqrt{p}}_{\theta_0} \sqrt{p}_{\theta_0} \\ &= (\theta - \theta_0)^2 \{\dot{\sqrt{p}}_{\theta_0}\}^2 + r_{\theta}^2 + 2\sqrt{p}_{\theta_0} r_{\theta} \\ &\quad + 2(\theta - \theta_0) \dot{\sqrt{p}}_{\theta_0} r_{\theta} \end{aligned} \quad (9.6)$$

となる. 最右辺を各項を積分したものを評価する.

$$\begin{aligned} \int_{\mathbb{X}} (\theta - \theta_0)^2 \{\dot{\sqrt{p_{\theta_0}}}\}^2 d\mu &= O(|\theta - \theta_0|^2), & (\because \mathcal{P} \text{ の可微分性}) \\ \int_{\mathbb{X}} r_{\theta}^2 d\mu &= o(|\theta - \theta_0|^2), & (\because \mathcal{P} \text{ の可微分性}) \\ \left| \int_{\mathbb{X}} \sqrt{p_{\theta_0}} r_{\theta} d\mu \right| &\leq \underbrace{\sqrt{\int_{\mathbb{X}} p_{\theta_0} d\mu}}_{=1} \sqrt{\int_{\mathbb{X}} r_{\theta}^2 d\mu} = o(|\theta - \theta_0|) \\ \left| \int_{\mathbb{X}} (\theta - \theta_0) \dot{\sqrt{p_{\theta_0}}} r_{\theta} d\mu \right| &\leq |\theta - \theta_0| \sqrt{\int_{\mathbb{X}} \{\dot{\sqrt{p_{\theta_0}}}\}^2 d\mu} \sqrt{\int_{\mathbb{X}} r_{\theta}^2 d\mu} \\ &= o(|\theta - \theta_0|^2) \end{aligned}$$

となる. 以上から

$$\begin{aligned} &\left| \gamma_S(\theta) - \gamma_S(\theta_0) - 2(\theta - \theta_0) \int_{\mathbb{X}} \dot{\sqrt{p_{\theta_0}}}(x) \sqrt{p_{\theta_0}}(x) S(x) d\mu(x) \right| \\ &= \left| \int_{\mathbb{X}} S(x) p_{\theta}(x) d\mu(x) - \int_{\mathbb{X}} S(x) p_{\theta_0}(x) d\mu(x) \right. \\ &\quad \left. - 2(\theta - \theta_0) \int_{\mathbb{X}} \dot{\sqrt{p_{\theta_0}}}(x) r_{\theta}(x) S(x) d\mu(x) \right| \\ &\leq M \left| \int_{\mathbb{X}} p_{\theta}(x) d\mu(x) - \int_{\mathbb{X}} p_{\theta_0}(x) d\mu(x) \right. \\ &\quad \left. - 2(\theta - \theta_0) \int_{\mathbb{X}} \dot{\sqrt{p_{\theta_0}}}(x) r_{\theta}(x) d\mu(x) \right| \\ &= o(|\theta - \theta_0|) \end{aligned}$$

がわかる. 以上から, $\gamma_S(\theta_0)$ は θ_0 で可微分で, (9.5) で与えられる導関数 $\dot{\gamma}_S(\theta_0)$ をもつことがわかる. \square

定理 9.11 の証明: いま

$$\begin{aligned} \Delta : \mathbb{X} \times \Theta \ni (x, \theta) &\mapsto \dot{\mathbf{q}}(\theta) \frac{\mathbb{1}\{\mathbf{q}(\theta) > 0\}}{2\sqrt{\mathbf{q}(\theta)}} \sqrt{p_{\theta}(x)} \\ &\quad + \sqrt{\mathbf{q}(\theta)} \dot{\sqrt{p_{\theta}}}(x) \end{aligned} \quad (9.7)$$

とおく. ただし, ほとんど至所の $\theta \in \Theta_q := \Theta \cap \text{supp}(\mathbf{q})$ 上では, Δ は well-defined であり, $\theta \notin \text{supp}(\mathbf{q})$ 上では, $\Delta = 0$ となる. m を \mathbb{R} 上の Lebesgue 測度としたとき, 以下の主張を証明する.

主張

$$2 \int_{\Theta \times \mathbb{X}} \Delta(x, \theta) \sqrt{\mathbf{q}(\theta)} \sqrt{p_{\theta}(x)} \{T(x) - \psi(\theta)\} dm(\theta) d\mu(x) = \int_{\Theta} \dot{\psi}(\theta) dQ(\theta) \quad (9.8)$$

が成り立つ.

主張の証明: (9.4) もしくは (9.8) を T が指示関数の有限和である場合について証明し, あとは標準機械の議論を経由して, 河積な T に対しても成り立つことを示せばよい. さらに, $n \in \mathbb{N}$ に対して, $\psi_n = \max\{-n, \min\{\psi, n\}\}$ とおくと, ほとんど至るところで, $\psi_n \xrightarrow{n \rightarrow \infty} \psi$ かつ $\dot{\psi}_n \xrightarrow{n \rightarrow \infty} \dot{\psi}$ である. 有界な ψ_n と $\dot{\psi}_n$ に対して (9.4) もしくは (9.8) が成立すれば, 優収束定理により, 一般の ψ と $\dot{\psi}$ に対しても成立することがわかる. 以上の議論から, 一般性を失うことなく, $T, \psi, \dot{\psi}$ は有界と仮定してよい.

(9.8) の左辺の項ごとに評価していく:

$$\begin{aligned} & 2 \int_{\Theta_q} \dot{q}(\theta) \frac{\mathbb{1}\{q(\theta) > 0\}}{2\sqrt{q(\theta)}} \sqrt{q(\theta)} T(x) p_\theta(x) d\mu(x) dm(\theta) \\ &= \int_{\Theta_q \times \mathbb{X}} \dot{q}(\theta) \left\{ \int_{\mathbb{X}} T(x) p_\theta(x) d\mu(x) \right\} dm(\theta) \\ &= - \int_{\Theta_q} \dot{q}(\theta) \gamma_T(\theta) dm(\theta), \end{aligned} \tag{9.9}$$

$$\begin{aligned} & -2 \int_{\Theta_q \times \mathbb{X}} \dot{q}(\theta) \frac{\mathbb{1}\{q(\theta) > 0\}}{2\sqrt{q(\theta)}} \sqrt{q(\theta)} \psi(\theta) p_\theta(x) d\mu(x) dm(\theta) \\ &= - \int_{\Theta_q} \dot{q}(\theta) \psi(\theta) \left\{ \int_{\mathbb{X}} p_\theta(x) d\mu(x) \right\} dm(\theta) \\ &= \int_{\Theta_q} \dot{q}(\theta) \gamma_T(\theta) dm(\theta), \end{aligned} \tag{9.10}$$

$$\begin{aligned} & 2 \int_{\Theta_q \times \mathbb{X}} \sqrt{q(\theta)} \sqrt{\dot{p}_\theta}(x) \sqrt{q(\theta)} \sqrt{p_\theta(x)} T(x) d\mu(x) dm(\theta) \\ &= 2 \int_{\Theta_q} q(\theta) \left\{ \int_{\mathbb{X}} \sqrt{\dot{p}_\theta}(x) \sqrt{p_\theta(x)} T(x) d\mu(x) \right\} dm(\theta) \\ &= \int_{\Theta_q} q(\theta) \dot{\gamma}_T(\theta) dm(\theta), \end{aligned} \tag{9.11}$$

$$\begin{aligned} & 2 \int_{\Theta_q \times \mathbb{X}} \sqrt{q(\theta)} \sqrt{\dot{p}_\theta}(x) \sqrt{q(\theta)} \sqrt{p_\theta(x)} \psi(\theta) d\mu(x) dm(\theta) \\ &= -2 \int_{\Theta_q} \psi(\theta) q(\theta) \left\{ \int_{\mathbb{X}} \sqrt{\dot{p}_\theta}(x) \sqrt{p_\theta(x)} d\mu(x) \right\} dm(\theta) \\ &= 0 \end{aligned}$$

となる. 4番目の式の最後の等号は, 補題 9.12 において, $T = 1$ とおくと $\gamma_T(\theta) = 1$ となるので, $\dot{\gamma}_T(\theta) = 2 \int_{\mathbb{X}} \sqrt{\dot{p}_\theta} \sqrt{p_\theta} d\mu = 0$ となることからわかる.

次に, Θ_q は可算個の互いに排反な区間 (a_r, b_r) ($r \in R$) の和集合として表現できることに注意する. Θ_q の有限な各境界は, Θ の有限な境界, Θ の

内点, $\text{supp}(\mathbf{q})$ の有限な境界のいずれかである. ψ の連続性から, $\text{supp}(\mathbf{q})$ の境界に θ が近づくと ψ と $\dot{\psi}$ の値は 0 に収束する. γ_T と ψ の有界性から, θ が a_r または b_r に近づくと, $\gamma_T(\theta)\mathbf{q}(\theta)$ と $\psi(\theta)\mathbf{q}(\theta)$ の値は 0 に近づく. さらに, $\lim_{\theta \rightarrow \pm\infty} \mathbf{q}(\theta) = 0$ である.

したがって, 閉区間 $[c, d] \subset \Theta_q$ に対して, (9.9) + (9.11) は

$$\int_{[c, d]} \{ \dot{\mathbf{q}}(\theta)\gamma_T(\theta) + \mathbf{q}(\theta)\dot{\gamma}_T(\theta) \} d\mathbf{m}(\theta) = [\mathbf{q}(\theta)\gamma_T(\theta)]_c^d,$$

$$\int_{[c, d]} \psi(\theta)\dot{\mathbf{q}}(\theta) d\mathbf{m}(\theta) = [\psi(\theta)\mathbf{q}(\theta)]_c^d - \int_{[c, d]} \dot{\psi}(\theta)\mathbf{q}(\theta) d\mathbf{m}(\theta)$$

となる. よって, $c \rightarrow a_r$ と $d \rightarrow b_r$ とすれば

$$\int_{a_r}^{b_r} \{ \dot{\mathbf{q}}(\theta)\gamma_T(\theta) + \mathbf{q}(\theta)\dot{\gamma}_T(\theta) \} d\mathbf{m}(\theta) = 0,$$

$$\int_{a_r}^{b_r} \psi(\theta)\dot{\mathbf{q}}(\theta) d\mathbf{m}(\theta) = - \int_{a_r}^{b_r} \dot{\psi}(\theta)\mathbf{q}(\theta) d\mathbf{m}(\theta)$$

がわかる. 最後に, $r \in R$ について和を取れば, (9.8) は示せた.

(9.4) の証明: Cauchy-Schwarz の不等式を (9.8) に適用すると

$$4 \int_{\Theta \times \mathbb{X}} \Delta(x, \theta)^2 d\mathbf{m}(\theta) d\mu(x) \int_{\Theta \times \mathbb{X}} \{T(x) - \psi(\theta)\}^2 dP_\theta(x) dQ(\theta)$$

$$\geq \left(\int_{\Theta} \dot{\psi}(\theta) dQ(\theta) \right)^2$$

となる. さらに

$$4 \int_{\Theta \times \mathbb{X}} \Delta(x, \theta)^2 d\mathbf{m}(\theta) d\mu(x)$$

$$= 4 \int_{\Theta \times \mathbb{X}} \{ \dot{\mathbf{q}}(\theta) \}^2 \frac{\mathbb{1}\{\mathbf{q}(\theta) > 0\}}{4\mathbf{q}(\theta)} \mathbf{p}_\theta(x) d\mathbf{m}(\theta) d\mu(x)$$

$$+ 4 \int_{\Theta \times \mathbb{X}} \{ \sqrt{\dot{\mathbf{p}}_\theta}(x) \}^2 \mathbf{q}(\theta) d\mathbf{m}(\theta) d\mu(x)$$

$$+ 4 \int_{\Theta \times \mathbb{X}} \dot{\mathbf{q}}(\theta) \mathbb{1}\{\mathbf{q}(\theta) > 0\} \sqrt{\dot{\mathbf{p}}_\theta}(x) \sqrt{\mathbf{p}_\theta}(x) d\mathbf{m}(\theta) d\mu(x)$$

$$= \mathcal{F}_Q + \int_{\Theta} \mathcal{F}_P(\theta) dQ(\theta)$$

$$+ \int_{\Theta} \dot{\mathbf{q}}(\theta) \mathbb{1}\{\mathbf{q}(\theta) > 0\} \underbrace{\left\{ \int_{\mathbb{X}} \sqrt{\dot{\mathbf{p}}_\theta}(x) \sqrt{\mathbf{p}_\theta}(x) d\mu(x) \right\}}_{=0 \text{ (補題 9.12 で } S=1 \text{)}} d\mu(x)$$

$$= \mathcal{F}_Q + \int_{\Theta} \mathcal{F}_P(\theta) dQ(\theta)$$

から

$$\int_{\Theta \times \mathbb{X}} \{T(x) - \psi(\theta)\}^2 dP_{\theta}(x) dQ(\theta) \geq \frac{\left(\int_{\Theta} \dot{\psi}(\theta) dQ(\theta)\right)^2}{\mathcal{F}_Q + \int_{\Theta} \mathcal{F}_P(\theta) dQ(\theta)}$$

がわかる. よって, 定理の不等式は証明された. □

9.5 統計的決定理論からみた Bayes 法

竹村 (2020, pp.326-334) と草間 (1975, pp.115-119) から借用すること.

9.6 章末注釈と参考文献

この章は [32, pp.175 – 192] を借用した.

9.7 演習問題

演習問題 9.1. $0 < \theta < 1$ と $n \in \mathbb{N}$ に対して, $X \sim \text{Bino}(n, \theta)$ とする. θ の事前分布を $\Theta \sim B(\alpha, \beta)$ とする. ただし, $\alpha > 0, \beta > 0$ である. すなわち, Θ の p.d.f. と $\Theta = \theta$ が与えられたときの X の条件付き p.m.f. はそれぞれ

$$p^{\Theta}(\theta) = \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} \mathbb{1}_{(0,1)}(\theta),$$

$$p^{X|\Theta}(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \mathbb{1}_{\{0,1,\dots,n\}}(x)$$

である. ただし

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx$$

である. このとき, 以下の問いに答えよ.

- (1) X の周辺 p.m.f. $p^X(\cdot)$ を求めよ.
- (2) $x = 0, 1, \dots, n$ に対して, $X = x$ を与えたときの Θ の条件付き分布 (事後分布) は $B(\alpha + x, \beta + n - x)$ となることを示せ.
- (3) $d^{\text{Bayes}}(X) = E[\Theta|X]$ とおいたとき

$$d^{\text{Bayes}}(X) = \frac{\alpha + \beta}{\alpha + \beta + x} \cdot \frac{\alpha}{\alpha + \beta} + \frac{n}{\alpha + \beta + n} \cdot \frac{X}{n}$$

であることを示せ.

(4)

$$R_{d^{\text{Bayes}}}(\theta) := E[\{d^{\text{Bayes}}(X) - \theta\}^2]$$

を計算し, $R_{d^{\text{Bayes}}}(\theta)$ が θ に依存しない定数関数になるための条件は

$$\alpha = \beta = \frac{\sqrt{n}}{2}$$

であることを示せ.

(5) $\alpha = \sqrt{n}/2$, $\beta = \sqrt{n}/2$ のときの d^{Bayes} を d^{M} とする. すなわち

$$d^{\text{M}} = \frac{1}{1 + \sqrt{n}} \cdot \frac{1}{2} + \frac{\sqrt{n}}{1 + \sqrt{n}} \cdot \frac{X}{n}$$

である. このとき

$$R_{d^{\text{M}}}(\theta) = \frac{1}{4(1 + \sqrt{n})^2} < \sup_{\theta \in (0, 1)} R_{X/n}(\theta)$$

を示せ. ただし

$$R_{X/n}(\theta) := E\left[\left(\frac{X}{n} - \theta\right)^2\right]$$

である.

ヒント

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}, \quad \Gamma\alpha = \int_0^\infty x^{\alpha-1}e^{-x} dx, \quad \Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$$

である.

演習問題 9.2.