数理統計　講義ノート[1]

加藤　賢悟[2]

[1]First version: 2015 年 3 月. This version: 平成 29 年 4 月 17 日. ちゃんと校正していないので，誤植・間違いがあると思います．何かコメントがあればメール下さい．

[2]東京大学大学院経済学研究科．〒 113-0033 東京都文京区本郷 7-3-1. E-mail: kkato@e.u-tokyo.ac.jp.

はじめに

この講義ノートは東京大学経済学部講義「数理統計」のために用意されたものである．「数理統計」は学部 3・4 年生向けの講義であり，2 年生向けの講義「統計」に続いて，中級レベルの数理統計学をカバーすることになっている．授業時間としては，週 2 回の 105 分授業が 13 週間分あるので，ある程度の分量の内容をカバーできる．本講義は竹村 (1991) を教科書として指定していて，講義ノートもそれに準拠している．その他に，講義のレベルに適合する教科書として，久保川 (2015), Bickel and Doksum (2015), Knight (2000) がある．ちょっと古いが，竹内 (1963) も味わい深い．Wasserman (2003) はユニークな教科書であり，中級レベルの数理統計学の教科書では扱われることの少なかったノンパラメトリック回帰やカーネル密度推定などもカバーしている．統計学の幅広いトピックを概観するにはよい本といえる [3].

数理統計学は測度論的な確率論に立脚していて，数理統計学を数学的に厳密に理解しようとするならば，測度論的な確率論を理解することが必要になる．しかし，本講義では，そこまで厳密性にはこだわらず，教養レベルの解析学と線形代数のみを前提として，数理統計学の基本的な考え方を理解することを目標にしている (線形回帰は「計量経済学」でカバーされるので「数理統計」ではカバーしない).

このような事情から，本講義ノートでは，厳密性を犠牲にしている箇所が多くある．そのような箇所には適宜注意を加えている．いずれにせよ，数学的な細部が気になる場合，講義ノートだけをもとにあれこれ悩むよりは，早い段階で測度論的確率論を勉強してしまったほうがすっきりする．測度論的確率論に関するある程度平易な教科書として，舟木 (2004), Resnick (1998), Williams (1991) を推薦しておこう．Resnick (1998) と Williams (1991) は測度論の一般論もカバーしている．数学的なバックグラウンドに不安がある場合は，Resnick (1998) が最適であろう．測度論そのものについては，吉田伸 (2006) が最近の標準的な教科書である．

測度論にもとづく (より厳密な) 数理統計学の教科書として，Lehmann and Casella (1998), Lehmann and Romano (2005), 鍋谷 (1978), 吉田朋 (2006) などがある．Lehmann の 2 巻本と鍋谷 (1978) は統計的決定理論に重点をおいた教科書である．内容は古典的ではあるが，勉強しておいて損はないであろう．ただし，Lehmann の 2 巻本は分量もあり，通読は難しい．鍋谷 (1978) はコンパクトにまとまっているが，具体例が少なく，抽象度が高い．鍋谷 (1978) を勉強しつつ，Lehmann を適宜参照するのがよいのかもしれない．吉田朋 (2006) はもっとバランスのとれた教科書であり，漸近理論にも詳しい．

講義の後半に扱う漸近理論に関しては，Ferguson (1996), Serfling (1980), Neywey and McFadden (1994), van der Vaart (1998) などが標準的な文献である．この中では Ferguson (1996) がコンパクトにまとまっていて読みやすい (本文は 170 ページで，演習問題の解答がついている)．van der Vaart (1998) は漸近決定理論をすっきりとまとめていて，研究レベルにおいても頻繁に引用される教科書であるが，証明の詳細を省略している箇所が多く，

[3] 同じ著者による Wasserman (2006) もノンパラメトリック統計を概観するにはよい本である．

通読するには相当な数学的成熟さが必要である．

その他，各トピックに関する参考文献を講義ノートのあちこちで紹介しているので，適宜参照されたい．

この講義に引き続いて統計学 (や計量経済学) を専門的に勉強したい，という学生は,「数理統計」の講義内容をちゃんと復習して，(1) 回帰分析，(2) 測度論と測度論的確率論，(3) Ferguson (1996) レベルの漸近理論，(4) R などのプログラミング言語，を勉強しておくと，そのあとの選択肢が広がる (ような気がする)．

目 次

1　初等確率論

1.1　確率空間と確率変数

数理統計は (測度論的) 確率論にもとづいている．測度論的確率論では，確率が定義される事象の集まりを σ 加法族として公理的に決めて,「確率」とはそのような集合族上の関数として定義される．この確率論の公理的な基礎づけは Kolmogorov (1933) による．

Ω を空でない集合とし，$\mathcal{F}$ を Ω の部分集合族とする．$\mathcal{F}$ が次の 3 条件をみたすとき，それを σ 加法族 (σ-field) と呼ぶ [4]：

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

(2) $A \in \mathcal{F} \Rightarrow A^c := \Omega \setminus A := \{\omega \in \Omega : \omega \notin A\} \in \mathcal{F}$.

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

Ω と σ 加法族 $\mathcal{F}$ のペア $(\Omega, \mathcal{F})$ を 可測空間 (measurable space) と呼ぶ.

Lemma 1.1. $(\Omega, \mathcal{F})$ を可測空間とする．

(a) $\varnothing \in \mathcal{F}$.

(b) $A_n \in \mathcal{F}, n = 1, 2, \cdots \Rightarrow \bigcap_{n=1}^\infty A_n \in \mathcal{F}$.

Proof. (a). $\varnothing = \Omega^c \in \mathcal{F}$.

(b). 各 n に対して，$A_n^c \in \mathcal{F}$ である．また，ド・モルガンの法則より，

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

である．□

なお，有限個の $A_1, \ldots, A_N \in \mathcal{F}$ に対しても，$A_n = \varnothing, n \geq N+1$ とすれば，

$$\bigcup_{n=1}^N A_n = \bigcup_{n=1}^\infty A_n$$

だから，

$$A_n \in \mathcal{F}, n = 1, \ldots, N \Rightarrow \bigcup_{n=1}^N A_n \in \mathcal{F}$$

である．また，

$$A_n \in \mathcal{F}, n = 1, \ldots, N \Rightarrow \bigcap_{n=1}^N A_n \in \mathcal{F}$$

[4] (左辺) := (右辺) と書いたら，(左辺) は (右辺) で定義されるという意味である．

である．

集合列 $A_n \subset \Omega, n = 1, 2, \dots$ が 排反 (disjoint) であるとは，任意の相異なる $m, n = 1, 2, \dots$ に対して，$A_m \cap A_n = \varnothing$ となることをいう．$\mathcal{F}$ 上の関数 $P : \mathcal{F} \to [0, 1]$ が次の 2 条件をみたすとき，P を $(\Omega, \mathcal{F})$ 上の 確率測度 (probability measure) と呼ぶ：

(1) $P(\Omega) = 1$.

(2) $A_n \in \mathcal{F}, n = 1, 2, \dots$ が排反なら，

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

$\Omega, \mathcal{F}, P$ の 3 つ組 $(\Omega, \mathcal{F}, P)$ を 確率空間 (probability space) と呼ぶ．また，$\mathcal{F}$ に属する集合を 事象 (event) と呼ぶ．

Remark 1.1. (2) において，$P(A_n) \geq 0$ より，$a_N = \sum_{n=1}^{N} P(A_n)$ とおくと，$a_N, N = 1, 2, \dots$ は非減少な数列だから，$N \to \infty$ のときの極限は $+\infty$ を許せば必ず存在する．(2) の意味は，その極限が $P(\bigcup_{n=1}^{\infty} A_n)$ に等しいということである．さらに，$P(\bigcup_{n=1}^{\infty} A_n)$ は有限だから，$\sum_{n=1}^{\infty} P(A_n)$ は絶対収束する．

Lemma 1.2. $(\Omega, \mathcal{F}, P)$ を確率空間とする．

(a) $P(\varnothing) = 0$.

(b) $A_n \in \mathcal{F}, n = 1, \dots, N$ が排反なら，$P(\bigcup_{n=1}^{N} A_n) = \sum_{n=1}^{N} P(A_n)$.

(c) $A, B \in \mathcal{F}, A \subset B \Rightarrow P(B \setminus A) = P(B) - P(A)$. 特に，$P(A) \leq P(B)$.

(d) $A_n \in \mathcal{F}, n = 1, 2, \dots$ に対して，$P(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} P(A_n)$.

(e) $A_n \in \mathcal{F}, n = 1, 2, \dots$ が

$$A_n \subset A_{n+1}, n = 1, 2, \dots$$

をみたすなら，$P(A_n) \uparrow P(\bigcup_{n=1}^{\infty} A_n)$.

(f) $A_n \in \mathcal{F}, n = 1, 2, \dots$ が

$$A_n \supset A_{n+1}, n = 1, 2, \dots$$

をみたすなら，$P(A_n) \downarrow P(\bigcap_{n=1}^{\infty} A_n)$.

Remark 1.2. (d) において，$\sum_{n=1}^{\infty} P(A_n)$ は非負数列の和だから，$+\infty$ を許せば必ず存在する．

Proof. (a). $A_n = \varnothing, n = 1, 2, \ldots$ とすれば，$A_n, n = 1, 2, \ldots$ は排反である．ここで，$\bigcup_{n=1}^{\infty} A_n = \varnothing$ だから，$P(\varnothing) = \sum_{n=1}^{\infty} P(A_n) \geq P(A_1) + P(A_2) = P(\varnothing) + P(\varnothing)$ であり，これを解いて $P(\varnothing) \leq 0$ を得る．$P \geq 0$ だから，$P(\varnothing) = 0$ である．

(b). $A_n = \varnothing, n \geq N + 1$ とすればよい．

(c). $C = B \setminus A = B \cap A^c \in \mathcal{F}$ とおくと，$B = A \cup C$ であって，A, C は排反だから，$P(B) = P(A) + P(C)$. よって，$P(C) = P(B) - P(A)$.

(d). $B_n, n = 1, 2, \ldots$ を

$$B_1 = A_1, B_n = A_n \setminus \bigcup_{m=1}^{n-1} A_m = A \cap \left(\bigcup_{m=1}^{n-1} A_m \right)^c, n = 2, 3, \ldots$$

と定義すれば，$B_n, n = 1, 2, \ldots$ は排反であって，それぞれ $\mathcal{F}$ に属する．さらに，$\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n$ だから，$P(\bigcup_{n=1}^{\infty} A_n) = P(\bigcup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} P(B_n)$. ところで，各 n に対して $B_n \subset A_n$ だから，(c) より，$P(B_n) \leq P(A_n)$ である．よって，$\sum_{n=1}^{\infty} P(B_n) \leq \sum_{n=1}^{\infty} P(A_n)$ を得る．

(e). (c) より，$P(A_n) \leq P(A_{n+1})$. 次に，$B_n, n = 1, 2, \ldots$ を (d) の証明と同様に定義すると，$\bigcup_{m=1}^{n-1} A_m = A_{n-1}$ だから，$B_n = A_n \setminus A_{n-1}$ である．よって，

$$\sum_{n=1}^{N} P(B_n) = P(A_N) - P(A_{N-1}) + \cdots + P(A_2) - P(A_1) + P(A_1) = P(A_N)$$

だから，

$$P\left(\bigcup_{n=1}^{\infty} A_n \right) = P\left(\bigcup_{n=1}^{\infty} B_n \right) = \lim_{N \to \infty} \sum_{n=1}^{N} P(B_n) = \lim_{N \to \infty} P(A_N)$$

を得る．

(d). $A_n^c, n = 1, 2, \ldots$ に対して (e) を適用すればよい． □

事象 $A_1, \ldots, A_n \in \mathcal{F}$ が独立 (independent) であるとは，任意の $1 \leq i_1 < \cdots < i_k \leq n$ $(k = 1, \ldots, n)$ に対して，

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k})$$

が成り立つことをいう．例えば，$n = 3$ なら，A_1, A_2, A_3 が独立であるとは，

$$\begin{cases} P(A_1 \cap A_2) = P(A_1)P(A_2) \\ P(A_2 \cap A_3) = P(A_2)P(A_3) \\ P(A_3 \cap A_1) = P(A_3)P(A_1) \\ P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3) \end{cases}$$

がすべて成り立つことである．

Lemma 1.3. $A_1,\ldots,A_n \in \mathcal{F}$ を独立とし，各 $i=1,\ldots,n$ に対して，B_i を A_i か A_i^c とする．このとき，$B_1,\ldots,B_n$ も独立である．

Proof. 略証のみ与える．詳細は演習問題とする．$B_i = A_i^c$ となる i の個数に関する帰納法により，$P(B_1\cap\cdots\cap B_n) = P(B_1)\cdots P(B_n)$ を得る．これから，

$$A_1,\ldots,A_n\text{が独立} \Rightarrow P(B_1\cap\cdots\cap B_n) = P(B_1)\cdots P(B_n) \text{ for } B_i = A_i \text{ or } A_i^c, i=1,\ldots,n$$

を得る．さらに，$A_1,\ldots,A_n$ が独立なら，任意の $1 \le i_1 < \cdots < i_k \le n$ に対して，$A_{i_1},\ldots,A_{i_k}$ も独立だから，$P(B_{i_1}\cap\cdots\cap B_{i_k}) = P(B_{i_1})\cdots P(B_{i_k})$ を得る．以上より，$B_1,\ldots,B_n$ は独立である． □

$A, B \in \mathcal{F}, P(B) > 0$ に対して，

$$P(A \mid B) = \frac{P(A\cap B)}{P(B)}$$

を B を与えたときの A の 条件付き確率 (conditional probability) と呼ぶ．このとき，

$$P(A\cap B) = P(A \mid B)P(B)$$

が成り立つ．よって，

$$A, B\text{ が独立} \Leftrightarrow P(A \mid B) = P(A)$$

である．

次に確率変数を定義しよう．Ω から $\mathbb{R} := (-\infty,\infty)$ への関数 $X : \Omega \to \mathbb{R}$ が 確率変数 (random variable, r.v.) であるとは，

$$\forall x \in \mathbb{R},\ \{X \le x\} := X^{-1}((-\infty, x]) := \{\omega \in \Omega : X(\omega) \le x\} \in \mathcal{F}$$

となることをいう．ここで，$a < b$ に対して，

$$\{X \in (a,b]\} := \{\omega \in \Omega : X(\omega) \in (a,b]\} = \{X \le b\} \setminus \{X \le a\} \in \mathcal{F}$$

であって，$A := (a,b) = \bigcup_{n=1}^{\infty}(a, b-1/n] =: \bigcup_{n=1}^{\infty} A_n$ より，

$$\{X \in (a,b)\} = X^{-1}(A) = X^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right) = \bigcup_{n=1}^{\infty} X^{-1}(A_n) = \bigcup_{n=1}^{\infty}\{X \in (a, b-1/n]\} \in \mathcal{F}$$

である [5]．さらに次の補題が成り立つ．

Lemma 1.4. X を r.v. とし，$A \subset \mathbb{R}$ を開集合か閉集合とする．このとき，$\{X \in A\} := X^{-1}(A) := \{\omega \in \Omega : X(\omega) \in A\} \in \mathcal{F}$ である．

[5] 一般に次の性質が成り立つ．(1) 任意の集合族 $A_i \subset \mathbb{R}, i \in \mathcal{I}$ に対して，$X^{-1}(\bigcup_{i\in\mathcal{I}} A_i) = \bigcup_{i\in\mathcal{I}} X^{-1}(A_i), X^{-1}(\bigcap_{i\in\mathcal{I}} A_i) = \bigcap_{i\in\mathcal{I}} X^{-1}(A_i)$．(2) 任意の $A \subset \mathbb{R}$ に対して，$X^{-1}(A^c) = (X^{-1}(A))^c$．

Proof. A を開集合とする．このとき，A は可算無限個の開区間の和で表せる．すなわち，ある $a_n < b_n, n = 1, 2, \ldots$ が存在して，$A = \bigcup_{n=1}^{\infty}(a_n, b_n)$ と表せる．ここで，$A_n = (a_n, b_n)$ とおくと，

$$\{X \in A\} = X^{-1}(A) = X^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right) = \bigcup_{n=1}^{\infty} X^{-1}(A_n) = \bigcup_{n=1}^{\infty}\{X \in (a_n, b_n)\} \in \mathcal{F}$$

である．A が閉集合のときは，$A^c = \mathbb{R} \setminus A$ は開集合であるから，$\{X \in A^c\} \in \mathcal{F}$ である．よって，$\{X \in A\} = X^{-1}(A) = (X^{-1}(A^c))^c \in \mathcal{F}$ を得る． □

r.v. X に対して，

$$F(x) := P(X \leq x) := P(\{\omega \in \Omega : X(\omega) \leq x\}), \quad x \in \mathbb{R}$$

を X の <u>分布関数</u> (distribution function, d.f.) と呼ぶ．定義より，

$$0 \leq F(x) \leq 1 \ \forall x \in \mathbb{R}$$

である．

Theorem 1.1. F を r.v. X の d.f. とする．このとき，F は次の (a)–(c) をみたす．

(a) F は単調非減少：$x < y \Rightarrow F(x) \leq F(y)$.

(b) $\lim_{x\to\infty} F(x) = 1, \lim_{x\to-\infty} F(x) = 0$.

(c) F は右連続：$x_n \downarrow x \Rightarrow \lim_{n\to\infty} F(x_n) = F(x)$.

Remark 1.3. 逆に，$\mathbb{R}$ 上の関数 F が (a)–(c) をみたせば，F を d.f. にもつ r.v. が存在することが知られている．

Proof. (a). $x < y$ なら，$\{X \leq x\} \subset \{X \leq y\}$ であるから，$F(x) = P(X \leq y) \leq P(Y \leq y) = F(y)$ である．

(b). $\Omega = \{X < \infty\} = \bigcup_{n=1}^{\infty}\{X \leq n\}$ であって，

$$\{X \leq n\} \subset \{X \leq n+1\}$$

であるから，$F(n) = P(X \leq n) \uparrow 1$. すなわち，任意の $\varepsilon > 0$ に対して，ある $N \in \mathbb{N}$ が存在して，$F(N) \geq 1 - \varepsilon$ となる．ここで，(a) より，任意の $x \geq N$ に対して，$F(x) \geq F(N) \geq 1 - \varepsilon$ であるから，$\lim_{x\to\infty} F(x) = 1$ を得る．$\lim_{x\to-\infty} F(x) = 0$ の証明も同様である．

(c). $x_n \downarrow x$ に対して，

$$\{X \leq x_n\} \supset \{X \leq x_{n+1}\}, \quad \bigcap_{n=1}^{\infty}\{X \leq x_n\} = \{X \leq x\}$$

だから，$F(x) = P(X \leq x) = \lim_{n\to\infty} P(X \leq x_n) = \lim_{n\to\infty} F(x_n)$ を得る． □

d.f. F に対して，

$$F(\infty) := \lim_{x\to\infty} F(x) = 1, \quad F(-\infty) := \lim_{x\to-\infty} F(x) = 0$$

と定義しておく．

Remark 1.4 (分布)**.** 以下に述べる内容の証明は本講義の範囲を超えるが，重要なので述べておく．Ω の部分集合族 $\mathcal{A}$ が与えられたとき，

$$\sigma(\mathcal{A}) := \bigcap\{\mathcal{C} : \mathcal{C} \supset \mathcal{A},\ \mathcal{C} \text{ は } \sigma \text{ 加法族 }\}$$

と定義すると，$\sigma(\mathcal{A})$ は $\mathcal{A}$ を含む最小の σ 加法族である．$\sigma(\mathcal{A})$ を $\mathcal{A}$ によって 生成される σ 加法族と呼ぶ．$\mathbb{R}$ の部分集合族 $\{(-\infty, x] : x \in \mathbb{R}\}$ によって生成される σ 加法族を，$\mathbb{R}$ の Borel σ 加法族 と呼び，$\mathcal{B}$ と書く：

$$\mathcal{B} := \sigma(\{(-\infty, x] : x \in \mathbb{R}\}).$$

$\mathcal{B}$ に属する集合を $\mathbb{R}$ の Borel 集合と呼ぶ．$\mathcal{B}$ は $\mathbb{R}$ の開集合と閉集合をすべて含むかなり大きな集合族である．証明は省略するが，X が r.v. なら，任意の $A \in \mathcal{B}$ に対して $X^{-1}(A) \in \mathcal{F}$ であって，

$$\mu(A) := P(X^{-1}(A)),\ A \in \mathcal{B}$$

と定めると，μ は $(\mathbb{R}, \mathcal{B})$ 上の確率測度になる．μ のことを X の 分布 (distribution) と呼ぶ．分布は d.f. から一意に決まるので，本講義では分布と d.f. を同一視する．以下，r.v. X が d.f. F をもつことを，

$$X \sim F$$

と書くことにする．このとき，X は F に従うともいう．また，X, Y が同じ d.f. をもつとき，$X \overset{d}{=} Y$ と書く．

F を d.f. とする．このとき，

$$F(x-) := \sup\{F(y) : y < x\}$$

と定義する．この定義から，

$$x_n < x, x_n \uparrow x \Rightarrow \lim_{n\to\infty} F(x_n) = F(x-)$$

が成り立つ．さらに，$x_n < x, x_n \uparrow x$ に対して，

$$\{X \le x_n\} \subset \{X \le x_{n+1}\},\ \bigcup_{n=1}^{\infty}\{X \le x_n\} = \{X < x\}$$

であるから，

$$F(x-) = \lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} P(X \le x_n) = P(X < x)$$

を得る．また，

$$P(X=x)=P(\{X\le x\}\setminus\{X<x\})=P(X\le x)-P(X<x)=F(x)-F(x-)$$

である．

Theorem 1.2. $x\in\mathbb{R}$ を所与とする．このとき，次の (a)–(c) は同値である．

(a) F は x で連続である．

(b) $F(x)=F(x-)$.

(c) $P(X=x)=0$.

Proof. (b) と (c) の同値性は明らか．また，(a) $\Rightarrow$ (b) も明らか．(b) $\Rightarrow$ (a) を示す必要があるが，その証明は各自に任せる． □

Remark 1.5. F が連続なら，任意の $x\in\mathbb{R}$ に対して $P(X=x)=0$ だから，$a<b$ に対して，$P(a\le X\le b)=P(a<X\le b)=P(a\le X<b)=P(a<X<b)$ である．

r.v. X が有限個か可算無限個の値 $\{x_1,x_2,\dots\}$ しかとらないとき，X を 離散型 (discrete) の r.v. と呼び，

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

を 確率関数 (probability mass function, p.m.f.) と呼ぶ．このとき，d.f. は

$$F(x)=\sum_{n:x_n\le x}p(x_n) \tag{*}$$

と表せる．(*) の形に表せる d.f. を 離散分布 (discrete distribution) と呼ぶ．

Example 1.1 (コイン投げ)**.** $\Omega=\{0,1\},\mathcal{F}=2^\Omega,P(\{0\})=P(\{1\})=1/2$ として，$X:\Omega\to\mathbb{R}$ を $X(\omega)=\omega$ と定義すれば，

$$P(X=0)=P(X=1)=1/2$$

である．このとき，d.f. は

$$F(x)=\begin{cases}0 & x<0\\ 1/2 & 0\le x<1\\ 1 & x\ge 1\end{cases}$$

である．

密度関数

広義 Riemann 積分を復習する．関数 $f:\mathbb{R}\to\mathbb{R}$ は各有界区間上で有界かつ Riemann 積分可能であるとする (例えば，f が連続ならこの仮定はみたされる)．このとき，$\int_{-\infty}^{\infty} f(x)dx$ を定義しよう．$f\geq 0$ のときとそうでない場合で場合分けする．

$f\geq 0$ のときは，$\int_{|x|\leq R} f(x)dx$ が R について非減少であるから，

$$\lim_{R\to\infty}\int_{|x|\leq R} f(x)dx$$

は $+\infty$ を許せば必ず存在する．そこで，このとき，

$$\int_{-\infty}^{\infty} f(x)dx := \lim_{R\to\infty}\int_{|x|\leq R} f(x)dx$$

と定義する．

f が負の値もとりうる場合，

$$f^{+}=\max\{f,0\},\quad f^{-}=\max\{-f,0\}$$

とおくと，$f^{+}, f^{-}\geq 0, f=f^{+}-f^{-}$ である．そこで，$\int_{-\infty}^{\infty} f^{+}(x)dx$ と $\int_{-\infty}^{\infty} f^{-}(x)dx$ のうちどちらかが有限なとき，

$$\int_{-\infty}^{\infty} f(x)dx := \int_{-\infty}^{\infty} f^{+}(x)dx - \int_{-\infty}^{\infty} f^{-}(x)dx$$

と定義する．右辺は $\pm\infty$ になりうる．$\int_{-\infty}^{\infty} f^{+}(x)dx<\infty$ & $\int_{-\infty}^{\infty} f^{-}(x)dx<\infty$ のとき，f は <u>可積分</u> (integrable) であるという．$|f|=f^{+}+f^{-}$ より，f が可積分であるためには，

$$\int_{-\infty}^{\infty} |f(x)|dx<\infty$$

であることが必要十分である．f が可積分であれば，$\int_{-\infty}^{\infty} f(x)dx$ は有限である．

Remark 1.6. (積分に関する注意).

- $f:\mathbb{R}\to\mathbb{R}_{+}:=[0,\infty)$ が連続で，$\int_{-\infty}^{\infty} f(x)dx=0$ なら，$f(x)=0\ \forall x\in\mathbb{R}$ となる．
- ここでは，$\mathbb{R}$ 上の広義積分を考察したが，そのほかの集合上の広義積分も同様に定義する．ただし，ここでは，広義積分を Lebesgue 積分と整合的になるように定義しているので，初等解析の教科書に現れる広義積分の定義と少し異なっている．なお，これ以降の議論において，Riemann 積分で不都合が生じる場合は，積分を Lebesgue 積分とみなす．
- 積分範囲が文脈から明らかなときは，積分範囲を省略する場合がある．

$f:\mathbb{R}\to\mathbb{R}_+$ を可積分な関数とし，

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

をみたすとする．このような関数 f を (確率) 密度関数 (probability density function, p.d.f.) と呼ぶ．確率密度関数 f に対して，

$$F(x) := \int_{-\infty}^{x} f(y)dy = \lim_{R\to\infty}\int_{-R}^{x} f(y)dy \tag{**}$$

は連続な d.f. になる．(**) の形の d.f. を 絶対連続 (absolutely continuous) な d.f. という．また，絶対連続な d.f. をもつ r.v. を 連続型 (continuous) の r.v. と呼ぶ．f が連続なら，F は C^1 級であって，$F' = f$ である．逆に，次の補題が成り立つ．

Lemma 1.5. 与えられた d.f. F が C^1 級なら，F は絶対連続であって，$f = F'$ を密度関数にもつ．

Proof. f は連続であって，

$$F(R) - F(-R) = \int_{|x|\le R} f(x)dx$$

をみたす．さらに，F は非減少であるから，$f \ge 0$ であって，$R\to\infty$ のとき，$F(R) - F(-R) \to 1$ となるから，$\int_{-\infty}^{\infty} f(x)dx = 1$ である．よって，f は確率密度関数である．さらに，

$$F(x) = \lim_{R\to\infty}\int_{-R}^{x} f(y)dy = \int_{-\infty}^{x} f(y)dy$$

を得る． □

Remark 1.7. もっと一般に，F が連続な d.f. で，ある $-\infty \le a < b \le \infty$ に対して，$F(b) = 1, F(a) = 0$ であって，(a,b) 上で C^1 級なら，F は絶対連続であって，

$$f(x) = \begin{cases} 0 & x \ge b \\ F'(x) & a < x < b \\ 0 & x \le a \end{cases}$$

は F の密度関数になる．

X が密度関数 f をもてば，区間か区間の有限和 $A\subset\mathbb{R}$ に対して，

$$P(X\in A) = \int_A f(x)dx \tag{*3}$$

が成り立つ．密度関数に関する積分を Lebesgue 積分とみなせば，Borel 集合 $A\subset\mathbb{R}$ に対して，(*3) が成り立つ (本講義ではこの関係は認める)．

なお，d.f. F が密度関数 f をもつとき，f を 1 点での値だけ変更した関数 f^* も同じ d.f. を導く．しかし，f, g が F の密度関数なら，“ほとんどすべての”$x \in \mathbb{R}$ に対して，$f(x) = g(x)$ が成り立つ．あとで密度関数に関する積分を考察するが，積分を Lebesgue 積分とみなせば，f に関する積分と g に関する積分は (積分が定義できる限り) 一致するので，この意味で d.f. は一意な密度関数をもつといえる．

なお，これ以降の議論において，確率関数や密度関数を分布と同一視する場合がある．例えば，r.v. X が密度関数 f をもつとき，$X \sim f$ と書く場合がある．

Example 1.2 (一様分布)**.** $a < b$ に対して，

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

は確率密度関数である．この f を密度関数にもつ分布を (a, b) 上の 一様分布 (uniform distribution) と呼ぶ．その d.f. は

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

である．(a, b) 上の一様分布を $U(a, b)$ と表し，r.v. X が $U(a, b)$ に従うことを，

$$X \sim U(a, b)$$

と書く．

また，

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

も密度関数であって，$U(a, b)$ と同じ d.f. を導くが，便宜的に $U[a, b]$ と書いたらその密度関数は g と約束しておく．

分位点関数

次に d.f. の分位点関数を定義しよう．d.f. F に対して，

$$F^{\leftarrow}(u) := \inf\{x \in \mathbb{R} : F(x) \geq u\},\ u \in (0, 1)$$

を F の 分位点関数 (quantile function) と呼ぶ．また，$u \in (0, 1)$ に対して，$F^{\leftarrow}(u)$ を F の u 分位点と呼ぶ．1/2 分位点のことを メディアン (median) と呼ぶ．F が連続かつ狭義

単調増加なら，$F^{\leftarrow}$ は F の逆関数に他ならない．F が $(0,1)$ 上に定義された逆関数をもつときは，$F^{\leftarrow}$ の代わりに F^{-1} と書く場合がある．しかし，一般には d.f. F は $(0,1)$ 上に定義された逆関数をもつとは限らない．例えば，コイン投げの例に現れた d.f. は $0, 1/2, 1$ にしか値をとらないので，$(0,1)$ 上に定義された逆関数をもたない．

ここで，$F^{\leftarrow}$ の定義より，$F(x_n) \geq u, x_n \downarrow F^{\leftarrow}(u)$ をみたす数列 x_n が存在する．このとき，F の右連続性より，

$$F(F^{\leftarrow}(u)) = \lim_{n\to\infty} F(x_n) \geq u$$

となるから，$F^{\leftarrow}$ の定義において inf は達成される．F が連続なら，$F(\infty) = 1, F(-\infty) = 0$ と中間値の定理より，任意の $u \in (0,1)$ に対して，$F(x) = u$ をみたす点 x が存在する．ここで，定義より $F^{\leftarrow}(u) \leq x$ だから，$F(F^{\leftarrow}(u)) \leq F(x) = u$ となる．よって，F が連続なら，$F(F^{\leftarrow}(u)) = u \ (\forall u \in (0,1))$ となる．しかし，F が不連続なら，$F(F^{\leftarrow}(u)) = u$ とは限らない．

Example 1.3. $X \sim U(a,b), a < b$ に対して，X の d.f. は

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

だから，分位点関数は

$$F^{\leftarrow}(u) = F^{-1}(u) = a + (b-a)u,\ u \in (0,1)$$

である．

Example 1.4. $P(X=0) = P(X=1) = 1/2$ に対して，X の d.f. は

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

だから，分位点関数は

$$F^{\leftarrow}(u) = \begin{cases} 0 & 0 < u \leq 1/2 \\ 1 & 1/2 < u < 1 \end{cases}$$

である．この場合，$u \neq 1/2$ のとき $F(F^{\leftarrow}(u)) \neq u$ である．

Theorem 1.3. 分位点関数について，次の (a)–(c) が成り立つ．

(a) $F^{\leftarrow}$ は非減少．

(b) $F^{\leftarrow}$ は左連続：$u_n \uparrow u \Rightarrow \lim_{n\to\infty} F^{\leftarrow}(u_n) = F^{\leftarrow}(u)$.

(c) $F^{\leftarrow}(u) \le x \Leftrightarrow u \le F(x)$.

Proof. (a). $u < v$ に対して，$\{x : F(x) \ge u\} \supset \{x : F(x) \ge v\}$ より明らか．

(b). $x_n := F^{\leftarrow}(u_n)$ とおくと，x_n は非減少かつ $x_n \le F^{\leftarrow}(u) =: x_0$ であるから，$x_n \uparrow x \le x_0$ となる．$x < x_0$ と仮定して矛盾を導く．$\varepsilon = (x_0 - x)/2$ とおくと，

$$u_n \le F(x_n + \varepsilon) \le F(x_0 - \varepsilon)$$

であるから，$n \to \infty$ として，$u \le F(x_0 - \varepsilon)$ となる．しかし，$F^{\leftarrow}$ の定義から，$F(x_0 - \varepsilon) < u$ であるから，矛盾が生じる．

(c). $u \le F(x) \Rightarrow F^{\leftarrow}(u) \le x$ は $F^{\leftarrow}$ の定義から明らか．逆に，$F^{\leftarrow}(u) \le x$ なら，

$$u \le F(F^{\leftarrow}(u)) \le F(x)$$

である． □

次の系は与えられた d.f. に従う r.v. は一様乱数を用いて発生させることができることを示している．

Corollary 1.1. F を d.f. とする．このとき，r.v. $U \sim U(0,1)$ に対して，$X = F^{\leftarrow}(U) \sim F$ となる．

Proof. (c) より，$\{X \le x\} = \{U \le F(x)\}$ であるから，

$$P(X \le x) = P(U \le F(x)) = F(x)$$

を得る． □

逆に，F が連続なら，$X \sim F$ に対して，$F(X) \sim U(0,1)$ となる (なぜか)．

1.2 期待値

期待値を定義しよう．X は離散型か連続型とし，離散型のときはその確率関数を $p(x)$ とし，連続型のときはその密度関数を $f(x)$ とする．このとき，関数 $g : \mathbb{R} \to \mathbb{R}$ に対して，$g(X)$ の <u>期待値</u> (expectation) $E[g(X)]$ を次のように定義する．ただし，X が連続型のときは，$g(x)f(x)$ は各有界区間上で有界かつ Riemann 積分可能であると仮定しておく．

$g \ge 0$ のとき：

$$E[g(X)] := \begin{cases} \sum_n g(x_n)p(x_n) & X \text{ が離散型のとき} \\ \int_{-\infty}^{\infty} g(x)f(x)dx & X \text{ が連続型のとき} \end{cases}$$

と定義する．右辺は $+\infty$ を許せば必ず存在する．

g が負の値もとるとき：$E[g^+(X)]$ と $E[g^-(X)]$ のいずれかが有限なら，

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

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．$E[|g(X)|] = E[g(X)^+] + E[g(X)^-] < \infty$ のとき，$g(X)$ は可積分であるという．$g(X)$ が可積分なら，$E[g(X)]$ が定義できて，$E[g(X)]$ は有限である．

例えば，定数 $c \in \mathbb{R}$ に対しては，$g(X) = c$ は可積分であって，

$$E[c] = \int_{-\infty}^{\infty} cf(x)dx = c\int_{-\infty}^{\infty} f(x)dx = c$$

となる．

なお，X が離散型のときは，形式的に $p(x) = 0$ $x \notin \{x_1, x_2, \dots\}$ と定義しておけば，

$$E[g(X)] = \sum_{x:p(x)>0} g(x)p(x)$$

と表すことができる．また，しばしば，右辺を $\sum_x g(x)p(x)$ と省略する．

次の補題の証明は難しくないので省略する．

Lemma 1.6. $g, h : \mathbb{R} \to \mathbb{R}$ に対して，次のいずれかを仮定する：(i) $g \geq 0, h \geq 0$. (ii) $E[|g(X)|] < \infty, E[|h(X)|] < \infty$.

(a) (期待値の線形性). (i) なら $a, b \geq 0$ に対して，(ii) なら $a, b \in \mathbb{R}$ に対して，

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

が成り立つ．

(b) (期待値の単調性). $g(x) \leq h(x)$ $\forall x \in \mathbb{R}$ なら，

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

が成り立つ．

Lemma 1.7. $0 < q < r$ に対して，$E[|X|^r] < \infty$ なら，$E[|X|^q] < \infty$ である．

Proof. $|x|^q \leq 1 + |x|^r$ より，

$$E[|X|^q] \leq E[1 + |X|^r] = 1 + E[|X|^r] < \infty$$

を得る． □

$k=1,2,\dots$ に対して，$E[X^k]$ が存在するとき，それを X の k 次 モーメント (moment) と呼ぶ．モーメントの値は $\pm\infty$ でもよいが，$E[|X|^k]<\infty$ なら，$E[X^k]$ は存在して有限である．そこで，$E[|X|^k]<\infty$ のとき，X は有限な k 次モーメントをもつという．特に，$E[X]$ を X の 平均 (mean) と呼ぶ．また，$E[|X|]<\infty$ のとき，

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

を X の 分散 (variance) と呼ぶ．なお，

$$\mathrm{Var}(X) = E[X^2-2XE[X]+(E[X])^2] = E[X^2]-(E[X])^2$$

である．$\mathrm{Var}(X)<\infty$ になるのは $E[X^2]<\infty$ のとき，またそのときに限る．

Remark 1.8. 平均，分散，k 次モーメントは，r.v. X というよりもその分布によって決まるので，$X\sim F$ に対して $E[X]$ のことを F の平均といったりもする．

集合 $A\subset\mathbb{R}$ に対して，

$$I_A(x) := I(x\in A) := \begin{cases} 1 & x\in A \\ 0 & x\notin A \end{cases}$$

と定義する．I_A を A の 指示関数 (indicator function) と呼ぶ．また，例えば，A が区間か区間の有限和なら，

$$E[I_A(X)] = P(X\in A) \tag{*}$$

が成り立つ．もっと一般に，密度関数に関する積分を Lebesgue 積分とみなせば，Borel 集合 $A\subset\mathbb{R}$ に対して，(*) が成り立つ．

Remark 1.9. X を離散型とし，$g:\mathbb{R}\to\mathbb{R}$ に対して，$Y=g(X)$ とおくと，Y も離散型である．X,Y の確率関数を p,q とおくと，$Y=g(X)$ の期待値は，

$$\sum_x g(x)p(x), \quad \sum_y yq(y)$$

と 2 通りに計算できる．このとき，$g\geq 0$ か，どちらか一方の和が絶対収束していれば (このときもう片方も絶対収束する)，両者は等しいことを示そう．次の解析の基本的な結果を使う (証明は標準的な解析の教科書を参照せよ)．

Lemma 1.8. $a_n\in\mathbb{R}, n=1,2,\dots$ に対して，その項の順序を並べ替えた数列を $a'_n, n=1,2,\dots$ とする．このとき，$a_n\geq 0\ \forall n$ or $\sum_n|a_n|<\infty$ なら，$\sum_n a_n=\sum_n a'_n$ となる．

各 y に対して，$A_y=\{x: g(x)=y, p(x)>0\}$ とおくと，$q(y)=P(Y=y)=P(X\in A_y)=\sum_{x\in A_y}P(X=x)=\sum_{x\in A_y}p(x)$ であるから，

$$\sum_y |y|q(y) = \sum_y |y|\sum_{x\in A_y}p(x) = \sum_y\sum_{x\in A_y}|g(x)|p(x).$$

ここで，$|g(x)|p(x) \geq 0$ より，項の順序を並べ替えても右辺の値は変わらない．よって，$\sum_y |y|q(y) = \sum_x |g(x)|p(x)$ を得る．さらに，いずれかの和が有限なら，

$$\sum_y yq(y) = \sum_y y \sum_{x \in A_y} p(x) = \sum_y \sum_{x \in A_y} g(x)p(x) = \sum_x g(x)p(x)$$

となる．

Remark 1.10. X が離散型でも連続型でもない場合でも期待値を定義することはできる．例えば，$f(x)$ を確率密度関数，$p(x)$ を確率関数とし，$0 < \alpha < 1$ に対して，X の d.f. が

$$F(x) = \alpha \int_{-\infty}^{x} f(y)dy + (1-\alpha) \sum_{z \leq x} p(z),\ x \in \mathbb{R} \tag{*}$$

と表される場合，$g : \mathbb{R} \to \mathbb{R}_+$ に対して，$g(X)$ の期待値を

$$E[g(X)] = \alpha \int_{-\infty}^{\infty} g(x)f(x)dx + (1-\alpha) \sum_x g(x)p(x)$$

と定義する．$g : \mathbb{R} \to \mathbb{R}$ に対しては，$E[g^+(X)]$ と $E[g^-(X)]$ のいずれかが有限のとき，$E[g(X)] = E[g^+(X)] - E[g^-(X)]$ と定義する．この期待値に対しても線形性や単調性が成り立つのは明らかである．

厳密にいうと，すべての d.f. が (*) の形に表せるわけではない．それでも，一般の r.v. に対して期待値を定義することは可能である．詳細は測度論的確率論の教科書を参照せよ．

次の Markov の不等式は最も基本的な確率不等式の 1 つである．

Theorem 1.4 (Markov の不等式)**.** $Y \geq 0$ なる r.v. Y に対して，

$$P(Y \geq c) \leq \frac{E[Y]}{c}, \quad \forall c > 0.$$

Proof. $Y \geq cI(Y \geq c)$ だから，両辺の期待値をとって，

$$E[Y] \geq cE[I(Y \geq c)] = cP(Y \geq c)$$

を得る． □

Corollary 1.2 (Chebyshev の不等式)**.** $E[X^2] < \infty$ なる r.v. X に対して，

$$P(|X - E[X]| \geq c) \leq \frac{\mathrm{Var}(X)}{c^2},\ \forall c > 0.$$

Proof. $|X - E[X]| \geq c \Leftrightarrow |X - E[X]|^2 \geq c^2$ より，

$$P(|X - E[X]| \geq c) = P(|X - E[X]|^2 \geq c^2).$$

$Y = |X - E[X]|^2$ として Markov の不等式を適用すれば，

$$P(|X - E[X]|^2 \geq c^2) \leq \frac{E[|X - E[X]|^2]}{c^2} = \frac{\mathrm{Var}(X)}{c^2}.$$

□

Corollary 1.3. r.v. $Y \geq 0$ に対して，$E[Y] = 0$ なら $P(Y = 0) = 1$ である[6].

Proof. Markov の不等式より，任意の $c > 0$ に対して，

$$P(Y \geq c) \leq \frac{E[Y]}{c} = 0.$$

よって，

$$P(Y > 0) = P\left(\bigcup_{n=1}^{\infty}\{Y \geq 1/n\}\right) \leq \sum_{n=1}^{\infty} P(Y \geq 1/n) = 0.$$

□

Example 1.5. r.v. X が有限な平均をもち，$\mathrm{Var}(X) = 0$ なら，$E[(X - E[X])^2] = 0$ だから，$P(X = E[X]) = 1$ となる．

モーメントと d.f. の間には次のような関係がある．

Theorem 1.5. 任意の $0 < p < \infty$ (p は整数である必要はない) に対して，

$$E[|X|^p] = \int_0^{\infty} px^{p-1} P(|X| > x)dx.$$

この関係はどちらかが $+\infty$ ならもう一方も $+\infty$ になるという意味で可積分性の条件なしに成り立つ．

Proof.

$$|X|^p = \int_0^{|X|} px^{p-1}dx = \int_0^{\infty} px^{p-1} I(x < |X|)dx$$

であって，両辺の期待値をとって，

$$E[|X|^p] = \int_0^{\infty} px^{p-1} \underbrace{E[I(x < |X|)]}_{=P(|X|>x)} dx$$

を得る．積分と期待値の順序交換は Fubini の定理から保証される． □

この定理から，$P(|X| > x)$ が $x \to \infty$ のとき十分速く減衰すれば ($P(|X| > x) = O\{x^{-p}(\log x)^{-2}\}$ $(x \to \infty)$ であればよい)，$E[|X|^p] < \infty$ となることがわかる．十分大きな x に対して，確率 $P(|X| > x)$ のことを X (or X の分布) の (両側) 裾確率 (tail probability) と呼ぶ[7]．高次の有限モーメントをもたない分布のことを，裾の重い分布といったりもする．例えば，Cauchy 分布は，

$$f(x) = \frac{1}{\pi(1 + x^2)},\ x \in \mathbb{R}$$

[6] この系では，(離散型や連続型とは限らない) 一般の r.v. に対して期待値が定義できることを認める．

[7] 十分大きな $x > 0$ に対して，$P(X > x)$ を X の上側裾確率，$P(X < -x)$ を X の下側裾確率といったりもする．しかしながら，裾確率という用語は数学的にちゃんと定義されているわけではない．

を密度関数にもつ分布であるが，Cauchy 分布は 1 次の有限モーメントをもたない．すなわち，Cauchy 分布に従う r.v. X に対して，$E[|X|] = \infty$ となる．Cauchy 分布は裾の重い分布の代表例である．

平均や分散は分布の "中心" や "散らばり具合" を表す指標と言われるが，そもそも有限な平均や分散が存在しない分布も存在する．そのような裾の重い分布に対しては，メディアン

$$F^{\leftarrow}(1/2)$$

や <u>四分位範囲</u> (interquantile range)

$$F^{\leftarrow}(3/4) - F^{\leftarrow}(1/4)$$

が分布の "中心" や "散らばり具合" の指標として適切であるといえる．もちろんこれらの指標も絶対的ではない．

変数変換

$-\infty \le a < b \le \infty$ とする．X を (a, b) に値をとる r.v. とし，$g : (a, b) \to \mathbb{R}$ に対して，$Y = g(X)$ とおく．このとき，Y の d.f. は

$$F_Y(y) = P(Y \le y) = P(g(X) \le y)$$

である．X は (a, b) 上に連続な密度関数 f_X をもつとする．g が狭義単調増加かつ C^1 級なら，g の逆関数 g^{-1} が存在して，$F_Y(y) = F_X(g^{-1}(y))$ となる．さらに，$g(g^{-1}(y)) = y$ より，

$$\{g^{-1}(y)\}' g'(g^{-1}(y)) = 1,\ \{g^{-1}(y)\}' = 1/g'(g^{-1}(y))$$

だから，Y は

$$f_Y(y) = F_Y'(y) = f_X(g^{-1}(y)) \frac{1}{g'(g^{-1}(y))},\ y \in (g(a), g(b))$$

を密度関数にもつ．

g が C^1 級かつ狭義単調減少なら，Y の密度関数は

$$f_Y(y) = f_X(g^{-1}(y)) \frac{1}{|g'(g^{-1}(y))|},\ y \in (g(b), g(a))$$

となる (演習問題)．

1.3 母関数

本節では分布を特徴づける関数として，確率母関数，モーメント母関数，および特性関数を考察する．こうした関数は分布の性質を調べるときに便利である．

確率母関数

X を $\{0,1,2,\dots\}$ に値をとる r.v. とし，$p(k)=P(X=k), k=0,1,2,\dots$ とおく．このとき，$|s|\le 1$ に対して，

$$G(s)=\sum_{k=0}^{\infty} s^k p(k)$$

を X の <u>確率母関数</u> (probability generating function) と呼ぶ $(0^0=1)$．$\sum_k p(k)=1$ より，$|s|\le 1$ に対して，$\sum_{k=0}^{\infty} s^k p(k)$ は一様に絶対収束する (Weierstrass の M テスト)．さらに，$\sum_k s^k p(k)$ は s の整級数なので，$|s|<1$ では項別微分可能である．すなわち，$m=1,2,\dots$ に対して，

$$G^{(m)}(s)=\sum_{k=m}^{\infty} k(k-1)\cdots(k-m+1)s^{k-m}p(k)=E[X(X-1)\cdots(X-m+1)s^{X-m}].$$

ここで，$s=0$ を代入すれば，$G^{(m)}(0)=m!p(m)$ であるから，

$$p(m)=\frac{G^{(m)}(0)}{m!}$$

を得る．これと $p(0)=G(0)$ より，G と $\{p(k)\}_{k=0}^{\infty}$ は 1 対 1 に対応することがわかる．

また，$G(s)$ の収束半径が 1 より大きいなら，$s=1$ を代入して，

$$G^{(m)}(1)=E[X(X-1)\cdots(X-m+1)]$$

を得る．

モーメント母関数

r.v. X に対して，ある $a>0$ が存在して，$E[e^{\theta X}]<\infty\ \forall|\theta|<a$ のとき，

$$\psi(\theta):=E[e^{\theta X}],\ |\theta|<a$$

を X の <u>モーメント母関数</u> (moment generating function) と呼ぶ [8]．モーメント母関数は常に存在するわけではない．任意の $k=1,2,\dots$ に対して，$|x|\to+\infty$ のとき，

$$\frac{|x|^k}{e^{|x|}}\to 0$$

であるから，十分大きい $L>0$ に対して，$|x|^k\le e^{|x|}\le e^x+e^{-x}\ \forall|x|>L$ である．さらに，$|x|\le L$ では $|x|^k\le L^k$ であるから，

$$|x|^k\le L^k+e^x+e^{-x}\ \forall x\in\mathbb{R}$$

[8] $e^{\theta X}\ge 0$ であるから，$E[e^{\theta X}]$ は $+\infty$ を許せば必ず存在する．しかし，X のモーメント母関数が存在するといったら，ある $a>0$ が存在して，$E[e^{\theta X}]<\infty\ \forall|\theta|<a$ となることを要求している．

を得る．よって，モーメント母関数 $\psi(\theta)$ が存在するなら，十分小さな $\theta \neq 0$ に対して，

$$E[|\theta X|^k] \leq L^k + \psi(\theta) + \psi(-\theta) < \infty$$

となる．つまり，モーメント母関数が存在すれば，X は任意次の有限なモーメントをもつ．逆に，ある正整数 k に対して，$E[|X|^k] = +\infty$ なら，X のモーメント母関数は存在しない．

モーメント母関数 $\psi(\theta)$ が $|\theta| < a$ において存在することを仮定する．ここで，

$$\frac{\partial}{\partial\theta} e^{\theta X} = X e^{\theta X}$$

であって，$E[|X|e^{\theta X}] < \infty \ \forall |\theta| < a$ となることが示せる．これから期待値と微分の交換が正当化できて (Lebesgue の優収束定理による)，$\psi(\theta)$ は $|\theta| < a$ で微分可能であって，

$$\psi'(0) = E[X]$$

となる．この操作を繰り返せば，$\psi(\theta)$ は $|\theta| < a$ で無限回微分可能であって，$k = 1, 2, \ldots$ に対して，

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

となることが示せる．

モーメント母関数は存在すれば分布と 1 対 1 に対応する．

Theorem 1.6. $X \sim F, Y \sim G$ に対して，それぞれモーメント母関数 ψ_F, ψ_G が存在するとする．このとき，十分小さい $\varepsilon > 0$ に対して，

$$\psi_F(\theta) = \psi_G(\theta) \ \forall |\theta| < \varepsilon$$

ならば $F \equiv G$ である．

この定理の証明は 1.10 節を参照せよ．

特性関数

モーメント母関数は常に存在するわけではないが，似たような役割をもつ特性関数は常に存在する．$i = \sqrt{-1}$ として，

$$\varphi(t) := E[e^{itX}] := E[\cos(tX)] + iE[\sin(tX)], \ t \in \mathbb{R}$$

を X の <u>特性関数</u> (characteristic function) と呼ぶ．$|\sin x| \leq 1, |\cos x| \leq 1$ より，右辺の期待値は存在しかつ有限である．さらに，$k = 1, 2, \ldots$ に対して，

$$\frac{\partial^k}{\partial t^k} e^{itX} = (iX)^k e^{itX}$$

であるから，$E[|X|^k] < \infty$ であれば，期待値と微分の交換が正当化できて (Lebesgue の優収束定理による)，$\varphi(t)$ は k 回微分可能であって，

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

となる．

X が整数値のとき，$p(k) = P(X = k), k = \cdots, -1, 0, 1, \ldots$ とおくと，

$$\varphi(t) = \sum_k e^{itk} p(k)$$

である．ここで，

$$\varphi_n(t) = \sum_{|k| \le n} e^{itk} p(k)$$

とおくと，$h = \cdots, -1, 0, 1, \ldots$ に対して，

$$\int_{-\pi}^{\pi} e^{ith} dt = \begin{cases} 2\pi & h = 0 \\ 0 & h \neq 0 \end{cases}$$

であるから，

$$\int_{-\pi}^{\pi} \varphi_n(t) e^{-itk} dt = \sum_{|j| \le n} p(j) \int_{-\pi}^{\pi} e^{it(j-k)} dt = \begin{cases} (2\pi) p(k) & |k| \le n \\ 0 & |k| > n \end{cases}$$

を得る．いま，

$$|\varphi(t) - \varphi_n(t)| \le \sum_{|k| > n} |e^{itk}| p(k) \le \sum_{|k| > n} p(k)$$

より，$n \to \infty$ のとき，$\sup_{|t| \le \pi} |\varphi(t) - \varphi_n(t)| \to 0$ となる．これから，確率関数に対する反転公式 (inversion formula)

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi(t) e^{-itk} dt = \frac{1}{2\pi} \lim_n \int_{-\pi}^{\pi} \varphi_n(t) e^{-itk} dt = p(k)$$

を得る．よって，このとき，φ と $\{p(k)\}_{k=-\infty}^{\infty}$ は 1 対 1 に対応する．

また，X が連続な密度関数 f をもつとき，f を特性関数 φ を使って表現してみよう．以下の議論は直感的なものであり厳密でない．$h > 0$ を所与とし，

$$p_h(k) = P(kh < X \le (k+1)h),\ k = \cdots, -1, 0, 1, \ldots$$

とおく．このとき，$\sum_k p_h(k) = 1$ であって，

$$\varphi_h(t) := \sum_k e^{itkh} p_h(k)$$

とおくと，$h \to 0$ のとき，

$$\varphi_h(t) = \sum_k e^{itkh} \int_{kh}^{(k+1)h} f(x)dx \approx \int_{-\infty}^{\infty} e^{itx} f(x)dx = \varphi(t)$$

である．さらに前述の議論より，

$$(2\pi)p_h(k) = \int_{-\pi}^{\pi} \varphi_h(t/h)e^{-itk}dt = h\int_{-\pi/h}^{\pi/h} \varphi_h(t)e^{-itkh}dt \approx h\int_{-\infty}^{\infty} \varphi(t)e^{-itkh}dt$$

である．ただし，$\int_{-\infty}^{\infty}|\varphi(t)|dt < \infty$ を仮定している．よって，

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \varphi(t)e^{-itkh}dt \approx \frac{1}{h}\int_{kh}^{(k+1)h} f(y)dy$$

を得る．ここで，$h \to 0, kh \to x$ として，

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \varphi(t)e^{-itx}dt \approx f(x)$$

を得る．以上の議論は全く直感的なものであったが，$\int_{-\infty}^{\infty}|\varphi(t)|dt < \infty$ であれば，密度関数に対する反転公式

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \varphi(t)e^{-itx}dt,\ x \in \mathbb{R}$$

が成り立つことが知られている (φ の可積分性から X が連続な密度関数をもつことも従う)．

もっと一般に次の定理が成り立つ．その証明は 1.10 節を参照せよ．

Theorem 1.7. $X \sim F, Y \sim G$ に対して，特性関数をそれぞれ φ_F, φ_G とする．このとき，$\varphi_F \equiv \varphi_G$ ならば $F \equiv G$ である．

1.4 主な 1 次元分布

Bernoulli 試行と 2 項分布

$0 \le p \le 1$ に対して，r.v. X が成功確率 p の <u>Bernoulli 試行</u> であるとは，

$$P(X=1) = p,\ P(X=0) = 1-p$$

をみたすことをいう．$X_1, \ldots, X_n$ を成功確率 p の Bernoulli 試行とし，$x_1, \ldots, x_n \in \{0,1\}$ に対して，$\{X_1 = x_1\}, \ldots, \{X_n = x_n\}$ は独立とする．このとき，

$$Y = X_1 + \cdots + X_n$$

の従う分布を2項分布 (binomial distribution) と呼び，$Bin(n,p)$ と表す．Y は $\{0,1,\dots,n\}$ にしか値をとらないから，2項分布は離散分布である．その確率関数を求めてみよう．$k=0,1,\dots,n$ に対して，$x_1+\cdots+x_n=k$ となる $(x_1,\dots,x_n)\in\{0,1\}^n$ 全体を $\mathcal{S}_{n,k}$ とおくと，

$$
\begin{aligned}
p(k)=P(Y=k)&=P\left(\bigcup_{(x_1,\dots,x_n)\in\mathcal{S}_{n,k}}\{X_1=x_1,\dots,X_n=x_n\}\right)\\
&=\sum_{(x_1,\dots,x_n)\in\mathcal{S}_{n,k}}P(X_1=x_1,\dots,X_n=x_n)=\sum_{(x_1,\dots,x_n)\in\mathcal{S}_{n,k}}P(X_1=x_1)\cdots P(X_n=x_n)
\end{aligned}
$$

である．$(x_1,\dots,x_n)\in\mathcal{S}_{n,k}$ のとき，$x_1,\dots,x_n$ のうち k 個が 1 で $(n-k)$ 個が 0 だから，$P(X_1=x_1)\cdots P(X_n=x_n)=p^k(1-p)^{n-k}$ である．さらに，

$$|\mathcal{S}_{n,k}|=\binom{n}{k}=\frac{n!}{k!(n-k)!}$$

であるから $(0!=1)$，

$$p(k)=\binom{n}{k}p^k(1-p)^{n-k},\ k=0,1,\dots,n$$

を得る．

次に $Bin(n,p)$ の確率母関数とモーメントを求めてみよう．確率母関数は

$$G(s)=\sum_{k=0}^{n}s^kp(k)=\sum_{k=0}^{n}\binom{n}{k}(sp)^k(1-p)^{n-k}=(sp+1-p)^n.$$

である．ここで，2項定理を使った：$a,b\in\mathbb{R},n=1,2,\dots$ に対して，

$$(a+b)^n=\sum_{k=0}^{n}\binom{n}{k}a^kb^{n-k}.$$

また，$m=1,\dots,n$ に対して，

$$E[Y(Y-1)\cdots(Y-m+1)]=G^{(m)}(1)=n(n-1)\cdots(n-m+1)p^m$$

である．特に，

$$E[Y]=np,\ E[Y(Y-1)]=n(n-1)p^2$$

より，

$$\mathrm{Var}(Y)=E[Y^2]-(E[Y])^2=n(n-1)p^2+np-(np)^2=np(1-p)$$

を得る．

Poisson 分布

$\lambda > 0$ に対して，

$$p(k) = \frac{\lambda^k}{k!}e^{-\lambda},\ k = 0, 1, 2, \ldots$$

を確率関数にもつ分布を Poisson 分布 と呼び，$Po(\lambda)$ と表す．Maclaurin 展開より，

$$e^{\lambda} = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}$$

であるから，

$$\sum_{k=0}^{\infty} p(k) = 1$$

であることが確認できる．$X \sim Po(\lambda)$ は $0, 1, 2, \ldots$ にしか値をとらない．

$Po(\lambda)$ の確率母関数とモーメントを求めてみよう．

$$G(s) = \sum_{k=0}^{\infty} s^k p(k) = \sum_{k=0}^{\infty} \frac{(s\lambda)^k}{k!}e^{-\lambda} = e^{(s-1)\lambda}.$$

よって，$X \sim Po(\lambda)$ に対して，

$$E[X(X-1)\cdots(X-m+1)] = G^{(m)}(1) = \lambda^m.$$

特に，$E[X] = \lambda, E[X^2] = E[X(X-1)] + E[X] = \lambda^2 + \lambda$ より，$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = \lambda$ である．すなわち，Poisson 分布の平均と分散は等しい．

正規分布

いままでは離散分布を扱ってきたが，以降は絶対連続分布を紹介する．密度関数

$$\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2},\ x \in \mathbb{R}$$

をもつ分布を 標準正規分布 (standard normal distribution) と呼び，$N(0,1)$ と表す．ϕ が確率密度関数であることは，Gauss 積分

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$$

より，

$$\int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2}\int_{-\infty}^{\infty} e^{-y^2} dy = \sqrt{2\pi}$$

となることから確かめられる．$N(0,1)$ の d.f. を

$$\Phi(x)=\int_{-\infty}^{x}\phi(y)dy$$

と表すことにする．

$N(0,1)$ のモーメント母関数を求めてみよう．所与の $\theta\in\mathbb{R}$ に対して，

$$e^{\theta x}e^{-x^2/4}\to 0,\ |x|\to\infty$$

であるから，$M>0$ を十分大きくとれば，

$$e^{\theta x}\phi(x)=(2\pi)^{-1/2}e^{\theta x}e^{-x^2/4}e^{-x^2/4}\le e^{-x^2/4},\ \forall|x|>M$$

であって，$|x|\le M$ では $e^{\theta x}\phi(x)$ は有界である．そこで，$e^{\theta x}\phi(x)\le M'\ \forall|x|\le M$ として，

$$g(x)=\begin{cases}M' & |x|\le M\\ e^{-x^2/4} & |x|>M\end{cases}$$

とおくと，g は可積分であって，$e^{\theta x}\phi(x)\le g(x)\ \forall x\in\mathbb{R}$ をみたす．よって，$X\sim N(0,1)$ に対して，$E[e^{\theta X}]<\infty\ \forall\theta\in\mathbb{R}$ である．$\psi(\theta)=E[e^{\theta X}]$ の値を求めよう．いま，

$$e^{\theta x}e^{-x^2/2}=e^{-(x-\theta)^2/2}e^{\theta^2/2}$$

であって，$e^{-(x-\theta)^2/2}$ の積分は $\sqrt{2\pi}$ なので，

$$\psi(\theta)=e^{\theta^2/2}$$

である．また，

$$E[X]=\psi'(0)=0,\ E[X^2]=\psi''(0)=1$$

である (直接計算してもよい)．次に，特性関数を求めてみよう．

$$\varphi(t)=E[e^{itX}]=\int_{-\infty}^{\infty}\cos(tx)\phi(x)dx+i\int_{-\infty}^{\infty}\sin(tx)\phi(x)dx.$$

ここで，$\sin(-tx)\phi(-x)=\sin(tx)\phi(x)$ より，右辺第 2 項は 0 である．さらに，両辺を t について微分して，

$$\begin{aligned}\varphi'(t)&=\int_{-\infty}^{\infty}\sin(tx)\{-x\phi(x)\}dx=\int_{-\infty}^{\infty}\sin(tx)\phi'(x)dx\\&=[\sin(tx)\phi(x)]_{-\infty}^{\infty}-t\int_{-\infty}^{\infty}\cos(tx)\phi(x)dx=-t\varphi(t).\end{aligned}$$

よって，$\{\varphi(t)e^{t^2/2}\}'=e^{-t^2/2}\{\varphi'(t)+t\varphi(t)\}=0$ より，$\varphi(t)e^{t^2/2}=1$ であるから，

$$\varphi(t)=e^{-t^2/2}$$

を得る．

$X \sim N(0,1)$ のとき，$\mu \in \mathbb{R}, \sigma \geq 0$ に対して，$Y = \mu + \sigma X$ とおくと，

$$E[Y] = \mu, \ \mathrm{Var}(Y) = \sigma^2$$

である．Y の従う分布を平均 μ と分散 σ^2 をもつ 正規分布 (normal distribution) と呼び，$N(\mu, \sigma^2)$ と表す [9]．$Y \sim N(\mu, 0)$ なら，$P(Y = \mu) = 1$ である．$\sigma > 0$ のとき，

$$P(Y \leq y) = P(X \leq (y-\mu)/\sigma) = \Phi((y-\mu)/\sigma)$$

であって，右辺は y について連続微分可能である．よって，Y は

$$\frac{d}{dy} P(Y \leq y) = \frac{1}{\sigma}\phi((y-\mu)/\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2/(2\sigma^2)}$$

を密度関数にもつ．さらに，Y のモーメント母関数は

$$E[e^{\theta Y}] = E[e^{\theta\mu + \theta\sigma X}] = e^{\theta\mu} E[e^{\theta\sigma X}] = e^{\theta\mu + \theta^2\sigma^2/2}$$

であって，特性関数は

$$E[e^{itY}] = E[e^{it\mu + it\sigma X}] = e^{it\mu - t^2\sigma^2/2}$$

である．

ガンマ分布とベータ分布

$\alpha > 0$ に対して，

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

を ガンマ関数 と呼ぶ．ここで，$\Gamma(\alpha) < \infty$ である (なぜか)．$\alpha = 1$ のときは，

$$\Gamma(1) = \int_0^\infty e^{-x} dx = [-e^{-x}]_0^\infty = 1$$

であって，$\alpha = 1/2$ のときは，$x^{1/2} = y$ と変数変換して，

$$\Gamma(1/2) = \int_0^\infty x^{-1/2} e^{-x} dx = \int_0^\infty y^{-1} e^{-y^2} 2y dy = \sqrt{\pi}$$

を得る．さらに，$\alpha > 0$ に対して，$(x^\alpha e^{-x})' = \alpha x^{\alpha-1} e^{-x} - x^\alpha e^{-x}$ より，$x^\alpha e^{-x} = (x^\alpha e^{-x})' + \alpha x^{\alpha-1} e^{-x}$ であるから，

$$\Gamma(\alpha+1) = \int_0^\infty x^\alpha e^{-x} dx = [x^\alpha e^{-x}]_0^\infty + \alpha \int_0^\infty x^{\alpha-1} e^{-x} dx = \alpha\Gamma(\alpha)$$

[9] 正規分布のことを Gauss 分布と呼ぶ場合もある．

である．よって，$n = 1, 2, \ldots$ に対して，

$$\Gamma(n+1) = n\Gamma(n) = n(n-1)\Gamma(n-1) = \cdots = n(n-1)\cdots 1 = n!$$

となる．

$\alpha > 0, \beta > 0$ に対して，

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} & x > 0 \\ 0 & x \le 0 \end{cases}$$

を密度関数にもつ分布をシェイプパラメータ α とスケールパラメータ β をもつ<u>ガンマ分布</u>と呼び，$Ga(\alpha, \beta)$ と表す．$f(x)$ がちゃんと確率密度関数になっていることは，

$$\begin{aligned} \int_0^\infty f(x)dx &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty x^{\alpha-1} e^{-x/\beta} dx \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty (\beta x)^{\alpha-1} e^{-y} \beta dy = 1 \end{aligned}$$

からわかる．さらに，

$$Y \sim Ga(\alpha, 1) \Rightarrow \beta Y \sim Ga(\alpha, \beta)$$

となることもわかる．ガンマ分布は $(0, \infty)$ に集中した分布である：$X \sim Ga(\alpha, \beta)$ に対して，$P(X > 0) = 1$ となる．

$Ga(\alpha, \beta)$ のモーメント母関数を求めてみよう．$\theta \in \mathbb{R}$ に対して，

$$x^{\alpha-1} e^{-x/\beta} e^{\theta x} = x^{\alpha-1} e^{-x(1/\beta - \theta)}.$$

$\theta < 1/\beta$ なら，右辺の 0 から $+\infty$ までの積分は有限になる．$\theta < 1/\beta$ のとき，

$$\psi(\theta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty x^{\alpha-1} e^{-x(1/\beta-\theta)} dx = (1 - \beta\theta)^{-\alpha}.$$

これから，$X \sim Ga(\alpha, \beta)$ に対して，

$$E[X] = \psi'(0) = \alpha\beta,\ E[X^2] = \psi''(0) = \alpha(\alpha+1)\beta^2,\ \mathrm{Var}(X) = E[X^2] - (E[X])^2 = \alpha\beta^2$$

を得る．

$\alpha = 1$ のとき，$\lambda > 0$ に対して，$Ga(1, 1/\lambda)$ を<u>指数分布</u> (exponential distribution) と呼び，$Ex(\lambda)$ と表す．すなわち，$Ex(\lambda)$ は

$$f(x) = \lambda e^{-\lambda x} I(x > 0)$$

を密度関数にもつ分布である．$Ex(\lambda)$ の分布関数は陽に計算できて，

$$F(x) = \int_{-\infty}^x f(y)dy = \begin{cases} 1 - e^{-\lambda x} & x > 0 \\ 0 & x \le 0 \end{cases}$$

である．よって，分位点関数は

$$F^{\leftarrow}(u) = F^{-1}(u) = -\frac{1}{\lambda}\log(1-u),\ u \in (0,1)$$

だから，$U \sim U(0,1)$ に対して，$-\lambda^{-1}\log(1-U) \sim Ex(\lambda)$ となる．ここで，$1-U \overset{d}{=} U$ だから，$-\lambda^{-1}\log U \sim Ex(\lambda)$ でもある．

$\alpha > 0, \beta > 0$ に対して，

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx \tag{*}$$

を ベータ関数 という．$\alpha < 1$ のとき，$x^{\alpha-1} \to +\infty\ (x \downarrow 0)$ であって，$\beta < 1$ のとき，$(1-x)^{\beta-1} \to +\infty\ (x \uparrow 1)$ であるから，その場合は，(*) の積分は広義積分である．しかし，いずれの場合も，$B(\alpha,\beta) < \infty$ である (なぜか)．

$\alpha > 0, \beta > 0$ に対して，

$$f(x) = \frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}I(0<x<1)$$

を密度関数にもつ分布を ベータ分布 と呼び，$Be(\alpha,\beta)$ と表す．ベータ分布は $(0,1)$ 上に集中した分布である：$X \sim Be(\alpha,\beta)$ に対して，$P(X \in (0,1)) = 1$ となる．あとで示すように，

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

という関係が成り立つから，

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

と計算できる．

1.5 確率ベクトル

行列 A に対して，A' は A の転置を表す．特に断らない限り，これ以降ベクトルはすべて列ベクトルとする．ただし，$x, y \in \mathbb{R}$ に対して，$(x,y)'$ といちいち書くのはわずらわしいので，(x,y) と書く．

r.v.'s $X_1,\ldots,X_n$ に対して，それらを並べたベクトル $X = (X_1,\ldots,X_n)'$ を n 次元の 確率ベクトル (random vector) と呼ぶ．このとき，$x_1,\ldots,x_n \in \mathbb{R}$ に対して，

$$F(x_1,\ldots,x_n) = P(X_1 \le x_1,\cdots,X_n \le x_n) = P(\{X_1 \le x_1\} \cap \cdots \cap \{X_n \le x_n\})$$

を X の 同時分布関数 (joint distribution function) と呼ぶ．また，各 i に対して，X_i の分布を 周辺分布 (marginal distribution) と呼ぶ [10]．

[10] "同時" とか "周辺" とかは省略する場合も多い．

Remark 1.11. $\mathbb{R}^n$ の部分集合族 $\{(-\infty,x_1]\times\cdots\times(-\infty,x_n]:-\infty<x_1,\dots,x_n<\infty\}$ の生成する σ 加法族のことを $\mathbb{R}^n$ の Borel σ 加法族と呼び，$\mathcal{B}^n$ と書く．$\mathcal{B}^n$ に属する集合を $\mathbb{R}^n$ の Borel 集合と呼ぶ．X が n 次元の確率ベクトルなら，任意の $A\in\mathcal{B}^n$ に対して $X^{-1}(A)\in\mathcal{F}$ であって，

$$\mu(A):=P(X^{-1}(A)),\ A\in\mathcal{B}^n$$

と定めると，μ は $(\mathbb{R}^n,\mathcal{B}^n)$ 上の確率測度になる．μ のことを X の分布と呼ぶ．X の分布は d.f. から一意に決まるので，1 次元のときと同様にして，分布と d.f. を同一視する．また，X が d.f. F をもつことを，$X\sim F$ と書く．

以下，$n=2$ の場合を主に考えて，(X,Y) を 2 次元の確率ベクトルとする．

まず，X,Y が離散型の場合を考える．X,Y がそれぞれ，$\{x_1,x_2,\dots\}$ と $\{y_1,y_2,\dots\}$ に値をとるとき，(X,Y) は $\{(x_m,y_n):m,n=1,2,\dots\}$ に値をとる．このとき，(X,Y) を離散型の確率ベクトルといって，

$$p(x_m,y_n)=P(X=x_m,Y=y_n)$$

を 同時確率関数 と呼ぶ．形式的に，

$$p(x,y)=0,\ (x,y)\notin\{(x_m,y_n):m,n=1,2,\dots\}$$

と定義しておけば，

$$F(x,y)=\sum_{u:u\le x}\sum_{v:v\le y}p(u,v)$$

と表せる．このように表せる d.f. F を離散分布関数と呼ぶ．また，

$$p_X(x):=P(X=x)=\sum_y P(X=x,Y=x)=\sum_y p(x,y)$$

を X の 周辺確率関数 と呼ぶ．

$p_X(x)>0$ なる x に対して，

$$p_{Y|X}(y\mid x)=P(Y=y\mid X=x)=\frac{P(X=x,Y=y)}{P(X=x)}=\frac{p(x,y)}{p_X(x)}$$

を $X=x$ を与えたときの Y の 条件付き確率関数 と呼ぶ．$\sum_y p_{Y|X}(y\mid x)=1$ だから，$p_{Y|X}(y\mid x)$ は y の関数として確率関数になる．$p_{Y|X}(\cdot\mid x)$ を確率関数にもつ分布のことを $X=x$ を与えたときの Y の 条件付き分布 (conditional distribution) と呼ぶ．なお，$p_X(x)=0$ なら $p(x,y)=0$ だから，そのような x に対しては $p_{Y|X}(y\mid x)$ をどう選んでも，

$$p(x,y)=p_{Y|X}(y\mid x)p_X(x)\ \forall(x,y)$$

が成り立つ．

次に，(X,Y) の同時 d.f. が，適当な (可積分) 関数 $f:\mathbb{R}^2\to\mathbb{R}_+$ を用いて，

$$F(x,y)=\int_{-\infty}^{y}\int_{-\infty}^{x}f(u,v)dudv,\ (x,y)\in\mathbb{R}^2$$

と表せるとき，(X,Y) を連続型の確率ベクトルといって，f を <u>同時 (確率) 密度関数</u> と呼ぶ [11]．このとき，F は <u>絶対連続</u> であるという．

同時密度関数 f は

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)dxdy=1 \tag{*}$$

をみたす．逆に，与えられた関数 $f:\mathbb{R}^2\to\mathbb{R}_+$ が条件 (*) をみたすなら，f を同時密度にもつ確率ベクトル (X,Y) が存在することが知られている．

(X,Y) が連続型のとき，X の周辺分布は

$$F_X(x)=P(X\le x)=P(X\le x,Y<\infty)=F(x,\infty)=\int_{-\infty}^{x}\left\{\int_{-\infty}^{\infty}f(u,y)dy\right\}du$$

と表せるから，X は連続型であって，その密度関数は

$$f_X(x)=\int_{-\infty}^{\infty}f(x,y)dy$$

と表せる．f_X を X の <u>周辺密度関数</u> と呼ぶ．

$f_X(x)>0$ なる x に対して，

$$f_{Y|X}(y\mid x)=\frac{f(x,y)}{f_X(x)}$$

を $X=x$ を与えたときの Y の <u>条件付き密度関数</u> と呼ぶ．$f_{Y|X}(y\mid x)$ は y の関数として確率密度関数であって，対応する分布を $X=x$ を与えたときの Y の条件付き分布と呼ぶ．

Remark 1.12. (X,Y) が同時密度 $f(x,y)$ をもつとき，密度関数に関する積分を Lebesgue 積分とみなせば，$A\in\mathcal{B}^2$ に対して，

$$P\{(X,Y)\in A\}=\iint_A f(x,y)dxdy$$

が成り立つ．本講義ではこの関係は認める．

r.v.'s の独立性を定義しよう．n 個の r.v.'s $X_1,\dots,X_n$ に対して，$(X_1,\dots,X_n)'$ の同時 d.f. を $F(x_1,\dots,x_n)$ とおいて，各 X_i の周辺 d.f. を $F_{X_i}(x_i)$ とおくと，

$$F(x_1,\dots,x_n)=F_{X_1}(x_1)\cdots F_{X_n}(x_n),\ \forall x_1,\dots,x_n\in\mathbb{R} \tag{**}$$

[11] 重積分を Riemann 積分の範囲でちゃんと扱おうとすると煩雑になるので，これ以降の議論においては，積分がちゃんと定義できることは暗に仮定してしまっている．

が成り立つとき，$X_1, \ldots, X_n$ は独立であるという．明らかに，$X_1, \ldots, X_n$ が独立なら，任意の $1 \leq i_1 < \cdots < i_m \leq n$ に対して，$X_{i_1}, \ldots, X_{i_m}$ も独立になる．

$X_1, \ldots, X_n$ が離散型の場合，それらが独立であることは，同時確率関数が周辺確率関数の積で表されることと同値である．すなわち，$(X_1, \ldots, X_n)'$ の同時確率関数を $p(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n)$ とおいて，各 X_i の周辺確率関数を $p_{X_i}(x_i) = P(X_i = x_i)$ とおくと，$X_1, \ldots, X_n$ が独立であることは，

$$p(x_1, \ldots, x_n) = p_{X_1}(x_1) \cdots p_{X_n}(x_n), \ \forall x_1, \ldots, x_n. \tag{*3}$$

となることと同値である．これを証明してみよう．(*3) が成り立つなら，$X_1, \ldots, X_n$ が独立なことは明らかである．そこで，逆を $n = 2$ の場合に示そう (一般の n でも同様である)．X_1, X_2 が独立なら，$P(X_1 \leq x_1, X_2 \leq x_2) = P(X_1 \leq x_1)P(X_2 \leq x_2)$ である．ここで，所与の x_1 に対して，$x_1^m \uparrow x_1$ となる数列 x_1^m をとると，$P(X_1 \leq x_1^m, X_2 \leq x_2) = P(X_1 \leq x_1^m)P(X_2 \leq x_2)$ より，$P(X_1 < x_1, X_2 \leq x_2) = P(X_1 < x_1)P(X_2 \leq x_2)$ となる．よって，

$$\begin{aligned} P(X_1 = x_1, X_2 \leq x_2) &= P(X_1 \leq x_1, X_2 \leq x_2) - P(X_1 < x_1, X_2 \leq x_2) \\ &= \{P(X_1 \leq x_1) - P(X_1 < x_1)\}P(X_2 \leq x_2) \\ &= P(X_1 = x_1)P(X_2 \leq x_2) \end{aligned}$$

となる．同様の操作を続けて，

$$p(x_1, x_2) = P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1)P(X_2 = x_2) = p_{X_1}(x_1)p_{X_2}(x_2)$$

を得る．

次に各 X_i が連続型の場合を考察する．一般に，各 X_i が連続型であっても，それらを並べたベクトル $(X_1, \ldots, X_n)'$ は連続型とは限らない．例えば，$X \sim N(0,1), Y = -X$ とすると，各 X, Y は連続型だが，(X, Y) は集合 $S = \{(x, y) : y = -x\}$ に集中していて，S の面積は 0 だから，(X, Y) は同時密度をもちえない．しかしながら，各 X_i が連続型のとき，その周辺密度関数を f_{X_i} とおくと，$X_1, \ldots, X_n$ が独立であれば，$(X_1, \ldots, X_n)'$ は連続型であって，同時密度関数

$$f(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n), \ \forall x_1, \ldots, x_n \in \mathbb{R} \tag{*4}$$

をもつ．これは，

$$\begin{aligned} F_{X_1}(x_1) \cdots F_{X_n}(x_n) &= \int_{-\infty}^{x_1} f_{X_1}(y_1) dy_1 \cdots \int_{-\infty}^{x_n} f_{X_n}(y_n) dy_n \\ &= \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f_{X_1}(y_1) \cdots f_{X_n}(y_n) dy_1 \cdots dy_n \end{aligned}$$

と表せることから従う．逆に，$(X_1, \ldots, X_n)'$ が (*4) を同時密度関数にもてば，$X_1, \ldots, X_n$ が独立になることは明らかである．

なお，確率ベクトルたちに対しても，独立性を同様に定義する．例えば，2 次元の確率ベクトル (X_1, X_2) と r.v. X_3 が独立であることは，

$$F(x_1, x_2, x_3) = \underbrace{F_{(X_1,X_2)}(x_1, x_2)}_{=P(X_1 \le x_1, X_2 \le x_2)} F_{X_3}(x_3),\ \forall x_1, x_2, x_3 \in \mathbb{R}$$

が成り立つことと定義される．

$X_1, \ldots, X_n$ を独立な r.v.'s とすると，それらをグループ分けしたものも独立になる．例えば，X_1, X_2, X_3 が独立なら，(X_1, X_2) と X_3 も独立である．これは，X_1, X_2, X_3 の独立性から，X_1, X_2 が独立なので，$F_{(X_1,X_2)}(x_1, x_2) = F_{X_1}(x_1)F_{X_2}(x_2)$ となるから，

$$F(x_1, x_2, x_3) = F_{X_1}(x_1)F_{X_2}(x_2)F_{X_3}(x_3) = F_{(X_1,X_2)}(x_1, x_2)F_{X_3}(x_3)$$

となるためである．

1.6 変数変換と Jacobian

$X = (X_1, \ldots, X_n)'$ を n 次元の連続型の確率ベクトルとし，その密度関数を

$$f_X(x) = f_X(x_1, \ldots, x_n),\ x = (x_1, \ldots, x_n)'$$

とする．$f_X(x)$ は $\mathbb{R}^n$ の開集合 A の外側では 0 になるとする : $f_X(x) = 0\ \forall x \notin A^c$. $g : A \to \mathbb{R}^n$ を連続微分可能な 1 対 1 関数とし，連続微分可能な逆関数 $g^{-1}(y)$ をもつとする．また，

$$g(x) = \begin{pmatrix} g_1(x) \\ \vdots \\ g_n(x) \end{pmatrix},\ h(y) = \begin{pmatrix} h_1(y) \\ \vdots \\ h_n(y) \end{pmatrix} = g^{-1}(y)$$

と表す．

このとき，$Y = g(X)$ の同時密度関数 $f_Y(y)$ を求めてみよう．$x = h(y)$ とおくと，x の各座標を y の各座標で偏微分した係数を要素とする行列

$$J = J(\partial x/\partial y) = \begin{pmatrix} \partial h_1(y)/\partial y_1 & \cdots & \partial h_1(y)/\partial y_n \\ \vdots & \ddots & \vdots \\ \partial h_n(y)/\partial y_1 & \cdots & \partial h_n(y)/\partial y_n \end{pmatrix}$$

を h の <u>Jacobi 行列</u>という．J の行列式 $|J|$ を h の <u>Jacobian</u> と呼ぶ (変換 $y \mapsto x$ の Jacobian ともいう)．このとき，Y の密度関数 $f_Y(y)$ は

$$f_Y(y) = f_X(g^{-1}(y)) \cdot ||J(\partial x/\partial y)||,\ y \in \{g(x) : x \in A\}$$

で与えられる．ただし，$||J||$ は $|J|$ の絶対値である．

Remark 1.13. $g(g^{-1}(x)) = x$ より，$I = J(\partial x/\partial y)J(\partial y/\partial x)$ であるから，Jacobi 行列と Jacobian は

$$J(\partial x/\partial y) = J(\partial y/\partial x)^{-1},\ |J(\partial x/\partial y)| = 1/|J(\partial y/\partial x)|$$

と計算できる．ただし，正確には，

$$J(\partial x/\partial y) = \left(\begin{pmatrix} \partial g_1(x)/\partial x_1 & \cdots & \partial g_1(x)/\partial x_n \\ \vdots & \ddots & \vdots \\ \partial g_n(x)/\partial x_1 & \cdots & \partial g_n(x)/\partial x_n \end{pmatrix} \Bigg|_{x=g^{-1}(y)} \right)^{-1}$$

である．

Example 1.6 (アファイン変換)**.** $a \in \mathbb{R}^n$ と $n \times n$ 正則行列 B に対して，$g(x) = a + Bx$ とおくと，$y = g(x)$ を解いて，$x = h(y) = B^{-1}(y - a)$ を得る．このとき，Jacobi 行列は

$$J(\partial x/\partial y) = B^{-1}$$

であるから，Jacobian は

$$|J(\partial y/\partial x)| = |B^{-1}| = 1/|B|$$

となる．

Example 1.7. X, Y を独立とし，X は正の値をとる r.v. とする．いま，X, Y がそれぞれ f_X, f_Y を密度関数にもつとき，$Z = XY$ の密度関数を求めてみよう．$x = x, z = xy$ を解くと $x = x, y = z/x$ であるから，変換 $(x, z) \mapsto (x, y)$ の Jacobian は

$$\begin{vmatrix} 1 & 0 \\ -z/x^2 & 1/x \end{vmatrix} = \frac{1}{x}$$

である．よって，(X, Z) の同時密度は

$$f_{X,Z}(x, z) = \frac{1}{x} f_X(x) f_Y(z/x),\ x > 0$$

であるから，Z の密度関数は

$$f_Z(z) = \int_0^\infty \frac{1}{x} f_Y(z/x) f_X(x) dx$$

である．

　ところで，X の周辺密度は f_X であるから，$X = x$ を与えたときの Z の条件付き密度関数は

$$f_{Z|X}(z \mid x) = \frac{f_{X,Z}(x, z)}{f_X(x)} = \frac{1}{x} f_Y(z/x)$$

である．

Example 1.8 (極座標変換). $X, Y \sim N(0,1)$ とし，X と Y は独立とする．同時密度は

$$f(x,y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$$

である．このとき，$X = r\cos\theta, Y = r\sin\theta$ $(r > 0, 0 < \theta < 2\pi)$ とおくと，変換 $(r,\theta) \mapsto (x,y)$ の Jacobian は

$$\begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r(\cos^2\theta + \sin^2\theta) = r$$

であるから，(r,θ) の密度関数は

$$f(r\cos\theta, r\sin\theta)r = \frac{1}{2\pi} \times re^{-r^2/2}, \quad r > 0, 0 < \theta < 2\pi$$

である．よって，r と θ は独立であって，r^2 $(= X^2 + Y^2) \sim Ex(1/2), \theta \sim U(0, 2\pi)$ である．

この関係から，逆に，$r > 0, \theta \in (0, 2\pi)$ が $r^2 \sim Ex(1/2), \theta \sim U(0, 2\pi)$ なる独立な r.v.'s なら，$r\cos\theta, r\sin\theta$ は独立に $N(0,1)$ に従う．さらに，独立な $U_1, U_2 \sim U(0,1)$ に対して，$-2\log U_1 \sim Ex(1/2), 2\pi U_2 \sim U(0, 2\pi)$ だから，

$$X = \sqrt{-2\log U_1}\cos(2\pi U_2),\ Y = \sqrt{-2\log U_1}\sin(2\pi U_2)$$

は独立に $N(0,1)$ に従う．このようにして一様乱数から $N(0,1)$ に従う r.v.'s を発生させる方法は <u>Box-Muller 法</u> と呼ばれる．

Example 1.9. $X \sim Ga(\alpha, 1), Y \sim Ga(\beta, 1)$ とし，X と Y は独立とする．$Z = X + Y, W = X/(X+Y)$ とおくと，Z, W は独立であって，$X \sim Ga(\alpha+\beta, 1), W \sim Be(\alpha, \beta)$ となることを示す．(X, Y) の同時密度は

$$f(x,y) = \frac{1}{\Gamma(a)\Gamma(b)} x^{\alpha-1} y^{\beta-1} e^{-x-y}, \quad x, y > 0$$

である．$z = x + y, w = x/(x+y)$ を解くと，$x = zw, y = z(1-w)$ であるから，変換 $(z,w) \mapsto (x,y)$ の Jacobian は

$$\begin{vmatrix} w & z \\ 1-w & -z \end{vmatrix} = -z$$

である．よって，(Z, W) の同時密度は

$$\begin{aligned} f_{Z,W}(z,w) &= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} (zw)^{\alpha-1} \{z(1-w)\}^{\beta-1} e^{-z} \\ &= \frac{1}{\Gamma(\alpha+\beta)} z^{\alpha+\beta-1} e^{-z} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} w^{\alpha-1}(1-w)^{\beta-1}, \quad z > 0, 0 < w < 1 \end{aligned}$$

であるから，Z, W は独立であって，$Z \sim Ga(\alpha+\beta, 1), W \sim Be(\alpha, \beta)$ となる．また，この計算から，

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

を得る．

1.7　確率ベクトルに関する期待値

(X,Y) を離散型か連続型とし，$g:\mathbb{R}^2\to\mathbb{R}$ に対して，$E[g(X,Y)]$ を定義する．1 次元のときと同様に，$g\geq 0$ のときは，

$$E[g(X,Y)]=\begin{cases}\sum_{x,y}g(x,y)p(x,y) & \text{離散型のとき}\\ \iint g(x,y)f(x,y)dxdy & \text{連続型のとき}\end{cases}$$

と定義し，一般の g に対しては，$g^+=\max\{g,0\},g^-=\max\{-g,0\}$ として，$E[g^+(X,Y)]$ と $E[g^-(X,Y)]$ のどちらかが有限のとき，

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

と定義する．定義から明らかなように，確率ベクトルに関する期待値についても線形性や単調性が成り立つ．

Remark 1.14. ここで，g が x のみの関数 $g(x)$ なら，$E[g(X)]$ は次の 2 通りに計算できる：

$$\int g(x)f_X(x)dx,\quad \iint g(x)f(x,y)dxdy. \tag{*}$$

形式的には，

$$\iint g(x)f(x,y)dxdy=\int g(x)\left\{\int f(x,y)dy\right\}dx=\int g(x)f_X(x)dx$$

であるから，(*) の 2 つの積分は一致するはずである．この操作は，$g\geq 0$ であるか，(*) のどちらかの積分が絶対収束していれば (このときもう一方も絶対収束する)，正当化できる (Fubini の定理による)．

X,Y が独立で，$E[|g(X)|]<\infty,E[|h(Y)|]<\infty$ なら，

$$\begin{aligned}E[g(X)h(Y)]&=\iint g(x)h(y)f_X(x)f_Y(y)dxdy\\&=\int g(x)f_X(x)dx\int h(y)f_Y(y)dy=E[g(X)]E[h(Y)]\end{aligned}$$

となる．

また，X,Y が独立なら，(可測) 関数 $g:\mathbb{R}\to\mathbb{R},h:\mathbb{R}\to\mathbb{R}$ と (Borel 集合) $A,B\subset\mathbb{R}$ に対して，

$$\begin{aligned}&P(g(X)\in A,h(Y)\in B)=E[I(g(X)\in A)I(h(X)\in B)]\\&\quad=E[I(g(X)\in A)]E[I(h(X)\in B)]=P(g(X)\in A)P(h(X)\in B)\end{aligned}$$

となる．特に，

$$X, Y \text{ が独立なら } g(X), h(Y) \text{ も独立}$$

である．

より一般に，$X_1, \dots, X_n$ が独立な r.v.'s で，$E[|g_i(X_i)|] < \infty$ $(i = 1, \dots, n)$ なら，

$$E[g(X_1) \cdots g(X_n)] = E[g(X_1)] \cdots E[g(X_n)]$$

となる．また，(可測) 関数 $g_i : \mathbb{R} \to \mathbb{R}$ $(i = 1, \dots, n)$ に対して，$g_1(X_1), \dots, g_n(X_n)$ も独立になる．以上の結果は，$X_1, \dots, X_n$ が確率ベクトルであっても成り立つ．

Example 1.10 (混合分布). $\alpha_1, \dots, \alpha_n \geq 0$ を $\sum_{i=1}^n \alpha_i = 1$ をみたす定数とし，U を $P(U = i) = \alpha_i$ $(i = 1, \dots, n)$ をみたす r.v. とする．さらに，$F_1, \dots, F_n$ を $\mathbb{R}$ 上の d.f. とし，各 i に対して $X_i \sim F_i$ とし，U は $(X_1, \dots, X_n)'$ と独立とする．このとき，

$$Y = \begin{cases} X_1 & \text{if} \quad U = 1 \\ X_2 & \text{if} \quad U = 2 \\ \quad \vdots & \\ X_n & \text{if} \quad U = n \end{cases}$$

とおくと，

$$\begin{aligned} P(Y \leq y) &= \sum_{i=1}^n P(Y \leq y, U = i) = \sum_{i=1}^n P(X_i \leq x, U = i) \\ &= \sum_{i=1}^n P(X_i \leq y) P(U = i) = \sum_{i=1}^n \alpha_i F_i(y) \end{aligned}$$

となる．Y の分布を $F_1, \dots, F_n$ の <u>混合分布</u> (mixture distribution) と呼び，$\alpha_1, \dots, \alpha_n$ を <u>混合比</u> (mixture weight) と呼ぶ．

各 F_i が密度関数 f_i をもつときは，Y は密度関数

$$f_Y(y) = \sum_{i=1}^n \alpha_i f_i(y)$$

をもつ．

共分散と相関

X, Y が $E[X^2] < \infty, E[Y^2] < \infty$ となる r,v,'s なら，

$$|XY| \leq (X^2 + Y^2)/2$$

より，$E[|XY|] < \infty$ である．このとき，

$$\mathrm{Cov}(X,Y) = E[(X-E[X])(Y-E[Y])] = E[XY] - E[X]E[Y]$$

を X と Y の 共分散 (covariance) と呼ぶ．$X=Y$ のときは，$\mathrm{Cov}(X,X) = \mathrm{Var}(X)$ である．X と Y が独立なときは，

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

であるから，$\mathrm{Cov}(X,Y) = 0$ である．ただし，$\mathrm{Cov}(X,Y) = 0$ でも，X と Y が独立とは限らない (演習問題).

Lemma 1.9 (Cauchy-Schwarz の不等式). $E[X^2] < \infty, E[Y^2] < \infty$ なら，

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

Proof. $X, Y \ge 0$ の場合を考えればよい．定義より，

$$0 \le E[(X-tY)^2] = E[Y^2]t^2 - 2E[XY]t + E[X^2]$$

であって，この不等式がすべての t に対して成り立つ．$E[Y^2] = 0$ なら，$-2E[XY]t + E[X^2] \ge 0$ であって，仮に $E[XY] > 0$ なら，$t < E[X^2]/(2E[XY])$ に対してこの不等式が成り立たない．よって，$E[Y^2] = 0$ なら $E[XY] = 0$ である [12]．さらに，$E[Y^2] > 0$ なら，

$$E[Y^2]t^2 - 2E[XY]t + E[X^2] = E[Y^2]\left(t - \frac{E[XY]}{E[Y^2]}\right)^2 + E[X^2] - \left(\frac{(E[XY])^2}{E[Y^2]}\right)$$

であって，これがすべての t に対して ≥ 0 であるから，

$$(E[XY])^2 \le E[X^2]E[Y^2]$$

である． □

X, Y を $E[X^2] < \infty, E[Y^2] < \infty$ となる r.v.'s とし，$\mathrm{Var}(X) > 0, \mathrm{Var}(Y) > 0$ とする．このとき，

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}$$

を X と Y の 相関 (correlation) と呼ぶ．Schwarz の不等式より，

$$|\mathrm{Cov}(X,Y)| \le \sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}$$

だから，

$$-1 \le \mathrm{Corr}(X,Y) \le 1$$

である．

[12] $E[Y^2] = 0$ なら $P(Y=0) = 1$ だから，$P(XY=0) = 1$．よって，$E[XY] = 0$ としてもよい．

$\mathrm{Corr}(X,Y)$ が定義できるとき，$\mu_X = E[X], \mu_Y = E[Y], \sigma_X^2 = \mathrm{Var}(X), \sigma_Y^2 = \mathrm{Var}(Y), \widetilde{X} = (X - \mu_X)/\sigma_X, \widetilde{Y} = (Y - \mu_Y)/\sigma_Y$ とおくと，$\mathrm{Corr}(X,Y) = 1$ なら，$\mathrm{Cov}(\widetilde{X}, \widetilde{Y}) = 1$ だから，$\mathrm{Var}(\widetilde{X} - \widetilde{Y}) = \mathrm{Var}(\widetilde{X}) - 2\,\mathrm{Cov}(\widetilde{X}, \widetilde{Y}) + \mathrm{Var}(\widetilde{Y}) = 0$. よって，$P(\widetilde{X} = \widetilde{Y}) = 1$ を得る．同様にして，$\mathrm{Corr}(X,Y) = -1$ なら，$P(\widetilde{X} = -\widetilde{Y}) = 1$ である．

$X_1, \ldots, X_n$ を $E[X_i^2] < \infty$ $(1 \le \forall i \le n)$ なる r.v.'s とすると，$b_1, \ldots, b_n \in \mathbb{R}$ に対して，

$$\mathrm{Var}\left(\sum_{i=1}^n b_i X_i\right) = \sum_{i=1}^n b_i^2 \,\mathrm{Var}(X_i) + 2\sum_{i<j} b_i b_j \,\mathrm{Cov}(X_i, X_j)$$

である．$X_1, \ldots, X_n$ が独立なら，

$$\mathrm{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{Var}(X_i)$$

となる．

$X = (X_1, \ldots, X_n)'$ を n 次元の確率ベクトルとする．このとき，

$$E[X] = \begin{pmatrix} E[X_1] \\ \vdots \\ E[X_n] \end{pmatrix}$$

と定義する (右辺の期待値の存在は仮定する)．行列の期待値も同様に定義する．また，$E[X_i^2] < \infty$ $1 \le \forall i \le n$ のとき，

$$\Sigma := \mathrm{Var}(X) := E[(X - E[X])(X - E[X])'] = (\mathrm{Cov}(X_i, X_j))_{1 \le i,j \le n}$$

を X の共分散行列と呼ぶ．共分散行列は明らかに対称である：$\Sigma = \Sigma'$．次の補題の証明は演習問題とする．

Lemma 1.10. $a = (a_1, \ldots, a_m)' \in \mathbb{R}^m$ と $m \times n$ 行列 B に対して，$Y = a + BX$ とおく．このとき，

$$E[Y] = a + BE[X], \quad \mathrm{Var}(Y) = B\,\mathrm{Var}(X)B'$$

である (それぞれの場合において，有限な $E[X]$ と $\mathrm{Var}(X)$ の存在は仮定する)．

$b = (b_1, \ldots, b_n)'$ に対して，$b'X$ の分散は

$$\mathrm{Var}(b'X) = b'\Sigma b$$

であって，$\mathrm{Var}(b'X) \ge 0$ より，Σ は半正定値対称行列である．また，$a, b \in \mathbb{R}^n$ に対して，

$$\mathrm{Cov}(a'X, b'X) = a'\Sigma b$$

である．

特性関数

n 次元確率ベクトル $X=(X_1,\ldots,X_n)'$ に対して，その特性関数を

$$\varphi(t)=E\left[e^{i\sum_{j=1}^n t_jX_j}\right]=E[e^{it'X}],\ t=(t_1,\ldots,t_n)'\in\mathbb{R}^n$$

と定義する．1 次元のときと同様に，特性関数と分布は 1 対 1 に対応する．

Theorem 1.8. X,Y を n 次元確率ベクトルとし，$X\sim F,Y\sim G$ とする．また，X,Y の特性関数をそれぞれ φ_F,φ_G とおく．このとき，$\varphi_F\equiv\varphi_G$ ならば $F\equiv G$ である．

r.v.'s $X_1,\ldots,X_n$ が独立なら，$X=(X_1,\ldots,X_n)'$ の特性関数 φ は

$$\varphi(t)=E\left[e^{i\sum_{j=1}^n t_jX_j}\right]=E\left[\prod_{j=1}^n e^{it_jX_j}\right]=\prod_{j=1}^n \underbrace{E[e^{it_jX_j}]}_{=:\varphi_{X_j}(t_j)}=\prod_{j=1}^n\varphi_{X_j}(t_j),\ t=(t_1,\ldots,t_n)'\in\mathbb{R}^n$$

をみたす．逆に，$X=(X_1,\ldots,X_n)'$ が

$$\varphi(t)=\prod_{j=1}^n\varphi_{X_j}(t_j),\ \forall t=(t_1,\ldots,t_n)'\in\mathbb{R}^n \tag{**}$$

をみたすなら，右辺は $X_1,\ldots,X_n$ が独立のときの X の特性関数に一致して，特性関数が分布を一意に決めることから，$X_1,\ldots,X_n$ は独立であることがわかる．よって，次の系を得る．

Corollary 1.4. $X=(X_1,\ldots,X_n)'$ の特性関数を φ とおき，各 X_j の特性関数を φ_{X_j} とおく．このとき，$X_1,\ldots,X_n$ が独立であるためには，(**) が成り立つことが必要十分である．

条件付き期待値

$E[|g(X,Y)|]<\infty$ のとき，$X=x$ を与えたときの $g(X,Y)$ の条件付き期待値 (conditional expectation) を

$$E[g(X,Y)\mid X=x]=\begin{cases}\sum_y g(x,y)p_{Y|X}(y\mid x) & \text{離散型のとき}\\ \int g(x,y)f_{Y|X}(y\mid x)dy & \text{連続型のとき}\end{cases}$$

と定義する．ここで，$p_X(x)>0$ or $f_X(x)>0$ を仮定している．$p_X(x)=0$ or $f_X(x)=0$ のときは，$E[g(X,Y)\mid X=x]$ の値は任意としておく．

$E[g(X,Y)\mid X=x]$ の x に X を代入した値を $E[g(X,Y)\mid X]$ と書く：

$$E[g(X,Y)\mid X]=E[g(X,Y)\mid X=x]|_{x=X}.$$

$E[g(X,Y) \mid X]$ は r.v. である．このとき，

$$
\begin{aligned}
E[E[g(X,Y) \mid X]] &= \int_{\{x:f_X(x)>0\}} E[g(X,Y) \mid X=x] f_X(x)dx \\
&= \int_{\{x:f_X(x)>0\}} \int g(x,y) \underbrace{f_{Y|X}(y \mid x) f_X(x)}_{=f(x,y)} dydx
\end{aligned}
$$

となる．ここで，$f_X(x) = 0$ なら “ほとんどすべての”y に対して $f(x,y) = 0$ になるから (f が連続ならすべての y に対して $f(x,y) = 0$ になる)，そのような x に対しては $\int g(x,y)f(x,y)dy = 0$ になる．よって，

$$
\int_{\{x:f_X(x)>0\}} \int g(x,y)f(x,y)dydx = \iint g(x,y)f(x,y)dydx = E[g(X,Y)]
$$

であるから，最終的に，

$$
E[E[g(X,Y) \mid X]] = E[g(X,Y)]
$$

を得る．これを 期待値の繰り返し法則 (law of iterated expectations) と呼ぶ．

同様にして，(Borel 集合) $A \subset \mathbb{R}$ に対して，

$$
E[E[g(X,Y) \mid X]I(X \in A)] = E[g(X,Y)I(X \in A)]
$$

が成り立つ．X, Y が独立なら，$f_{Y|X}(y \mid x) = f_Y(y)$ より，

$$
E[g(X,Y) \mid X=x] = E[g(x,Y)], \quad f_X(x) > 0
$$

である．つまり，X, Y が独立な場合は，$X = x$ を与えたときの (X,Y) に関する条件付き期待値は，$X = x$ を固定して Y の周辺分布に関して期待値をとったものに等しい．

次に，(Borel 集合) $A \subset \mathbb{R}$ に対して，

$$
P(Y \in A \mid X=x) := E[I(Y \in A) \mid X=x]
$$

と定義する．

また，$p_X(x) > 0$ or $f_X(x) > 0$ なる x に対して，y の関数

$$
F_{Y|X}(y \mid x) := P(Y \le y \mid X=x)
$$

を $X = x$ を与えたときの Y の 条件付き分布関数 と呼ぶ．$p_X(x) = 0$ or $f_X(x) = 0$ なら，G を $\mathbb{R}$ 上の任意の d.f. として，$F_{Y|X}(y \mid x) = G(y)$ としておく．このように決めておくと，任意の $x \in \mathbb{R}$ に対して，$F_{Y|X}(y \mid x)$ は y の関数として d.f. になる．そこで，

$$
F^{\leftarrow}_{Y|X}(u \mid x) =: \inf\{y \in \mathbb{R} : F_{Y|X}(y \mid x) \ge u\},\ u \in (0,1)
$$

を $X = x$ を与えたときの Y の 条件付き分位点関数 と呼ぶ．X が多次元の場合も，条件付き分布関数と条件付き分位点関数を同様に定義する．

条件付き分位点関数を使うと，多次元分布に従う確率ベクトルを独立な一様確率変数列から発生させることができる．$(X_1,\dots,X_n)'$ を離散型か連続型の確率ベクトルとし，同時 d.f. を $F(x_1,\dots,x_n)=P(X_1\le x_1,\dots,X_n\le x_n)$ とおく．ここで，$F_1(x_1)=F_{X_1}(x_1)$ とし，$k=2,\dots,n$ に対して，$F_k(x_k\mid x_1,\dots,x_{k-1})$ を $(X_1,\dots,X_{k-1})'=(x_1,\dots,x_{k-1})'$ を与えたときの X_k の条件付き分布関数とする．いま，$U_1,\dots,U_n$ を独立に $U(0,1)$ に従う r.v.'s とし，$Y_1,\dots,Y_n$ を

$$\begin{aligned}&Y_1=F_1^{\leftarrow}(U_1),\\&Y_k=F_k^{\leftarrow}(U_k\mid Y_1,\dots,Y_{k-1}),\ k=2,\dots,n\end{aligned}$$

とおく．このとき，$Y_1\sim F_{X_1}$ であって，$Y_1=y_1$ を与えたとき，Y_2 は定義から $F_2(\cdot\mid y_1)$ に従う．よって，(Y_1,Y_2) の同時分布は

$$\begin{aligned}&P(Y_1\le y_1,Y_2\le y_2)=E[I(Y_1\le y_1)I(Y_2\le y_2)]=E[E[I(Y_2\le y_2)\mid Y_1]I(Y_1\le y)]\\&=E[F_2(y_2\mid Y_1)I(Y_1\le y_1)]=E[F_2(y_2\mid X_1)I(X_1\le y_1)]\\&=E[E[I(X_2\le y_2)\mid X_1]I(X_1\le y)]=P(X_1\le y_1,X_2\le y_2)\end{aligned}$$

だから，(X_1,X_2) のそれと等しい．同様の操作を続けて，$(Y_1,\dots,Y_n)'\sim F$ を得る．

1.8 独立な確率変数の和の分布

$X\sim F, Y\sim G$ を r.v.'s とし，X と Y は独立とする．このとき，$Z=X+Y$ の d.f. を $F*G$ と表し，F と G の たたみ込み (convolution) と呼ぶ．たたみ込みは明らかに可換である：$F*G=G*F$．Z の分布の求め方として，(1) Z の確率 (密度) 関数を直接求める方法と，(2) 母関数を用いる方法がある．

確率 (密度) 関数を直接求める方法

X,Y が離散型のときは，

$$P(Z=z\mid Y=y)=P(X+Y=z\mid Y=y)=P(X=z-y\mid Y=y)=p_X(z-y)$$

であるから，Z の確率関数は

$$p_Z(z)=P(Z=z)=\sum_y P(Z=z\mid Y=y)p_Y(y)=\sum_y p_X(z-y)p_Y(y)$$

となる．p_Z のことを p_X と p_Y の たたみ込み と呼び，$p_Z=p_X*p_Y$ と書く．

次に，X,Y が連続型の場合を考える．このとき，

$$\begin{pmatrix}x\\y\end{pmatrix}\mapsto\begin{pmatrix}x+y\\y\end{pmatrix}=\begin{pmatrix}z\\y\end{pmatrix}$$

という変換を考えると，逆変換は

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} z-y \\ y \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} z \\ y \end{pmatrix}$$

であるから，Jacobian は 1 である．よって，(Z, Y) の同時密度関数 $f_{Z,Y}(z, y)$ は

$$f_{Z,Y}(z, y) = f_X(z-y) f_Y(y)$$

であるから，Z の周辺密度関数は

$$f_Z(z) = \int_{-\infty}^{\infty} f_{Z,Y}(z, y) dy = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y) dy$$

である．f_Z のことを f_X と f_Y の たたみ込み と呼び，$f_Z = f_X * f_Y$ と書く．

Example 1.11. $X, Y \sim U(-1/2, 1/2)$ を独立とし，$Z = X + Y$ とおく．このとき，

$$f_X(x) = I(-1/2 < x < 1/2), \ f_Y(y) = I(-1/2 < y < 1/2)$$

であるから，

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} I(-1/2 < z-y < 1/2) I(-1/2 < y < 1/2) dy \\ &= \int_{(-1/2,1/2)\cap(z-1/2,z+1/2)} dy \\ &= \begin{cases} 1-z & 0 \leq z < 1 \\ 1+z & -1 < z < 0 \\ 0 & |z| \geq 1 \end{cases} \\ &= (1-|z|)^+. \end{aligned}$$

Z の分布を 三角分布 (triangular distribution) と呼ぶ．

母関数を用いる方法

例として，モーメント母関数を用いる方法を説明しよう．確率母関数や特性関数を用いる場合でも議論は同様である．Z のモーメント母関数は

$$\psi_Z(\theta) = E[e^{\theta Z}] = E[e^{\theta(X+Y)}] = E[e^{\theta X}] E[e^{\theta Y}] = \psi_X(\theta) \psi_Y(\theta)$$

である (存在は仮定する)．モーメント母関数と分布の一意性から，ψ_Z から Z の分布が一意に決まる．例えば，

$$X \sim N(\mu_1, \sigma_1^2), \quad Y \sim N(\mu_2, \sigma_2^2)$$

のとき，$\psi_X(\theta) = e^{\mu_1\theta+\sigma_1^2\theta^2/2}, \psi_Y(\theta) = e^{\mu_2\theta+\sigma_2^2\theta^2/2}$ であるから，

$$\psi_Z(\theta) = e^{(\mu_1+\mu_2)\theta+(\sigma_1^2+\sigma_2^2)\theta^2/2}$$

であって，右辺は $N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$ のモーメント母関数であるから，

$$Z \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$$

を得る．これは

$$N(\mu_1, \sigma_1^2) * N(\mu_2, \sigma_2^2) = N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$$

と表すことができる．すなわち，正規分布のたたみ込みは再び正規分布になる．この性質を正規分布の 再生性 (regeneration property) と呼ぶ．同様にして，

$$\begin{aligned}
&Bin(n,p) * Bin(m,p) = Bin(n+m,p),\\
&Po(\lambda) * Po(\kappa) = Po(\lambda+\kappa),\\
&Ga(\alpha_1,\beta) * Ga(\alpha_2,\beta) = Ga(\alpha_1+\alpha_2,\beta)
\end{aligned}$$

も示すことができる．

3 回以上のたたみ込み

$\mathbb{R}$ 上の d.f.'s F, G, H に対して，F, G, H のたたみ込みを

$$F * G * H = (F * G) * H$$

と定義する．このとき，$F*G*H$ は独立な r.v.'s $X \sim F, Y \sim G, Z \sim H$ の和 $X+Y+Z$ の d.f. に一致する．なぜなら，X, Y, Z の独立性から，(X, Y) と Z は独立なので，$X+Y$ と Z は独立である．$X+Y$ の d.f. は $F*G$ だから，$X+Y+Z = (X+Y)+Z$ の d.f. は $(F*G)*H$ になる．この議論から，明らかに $F*G*H = F*(G*H)$ でもある．

もっと一般に，$F_1, \ldots, F_n$ $(n \geq 3)$ を $\mathbb{R}$ 上の d.f.'s すると，それらのたたみ込みを帰納的に

$$F_1 * \cdots * F_{k-1} * F_k = (F_1 * \cdots * F_{k-1}) * F_k,\ k = 3, \ldots, n$$

と定義する．このとき，$F_1*\cdots*F_n$ は独立な r.v.'s $X_1 \sim F_1, \ldots, X_n \sim F_n$ の和 $X_1+\cdots+X_n$ の d.f. に一致する．これから，$\{1, \ldots, n\}$ の置換 σ に対して，$F_{\sigma(1)}*\cdots*F_{\sigma(n)} = F_1*\cdots*F_n$ であることもわかる．

確率関数や密度関数の 3 回以上のたたみ込みも同様に定義する．例えば，$f_1, \ldots, f_n$ $(n \geq 3)$ を $\mathbb{R}$ 上の密度関数とすると，それらのたたみ込みは，帰納的に，

$$f_1 * \cdots * f_{k-1} * f_k = (f_1 * \cdots * f_{k-1}) * f_k,\ k = 3, \ldots, n$$

と定義される．独立な r.v.'s $X_1 \sim f_1, \ldots, X_n \sim f_n$ に対して，$X_1+\cdots+X_n$ は $f_1*\cdots*f_n$ を密度関数にもつ．

Example 1.12. f が $U(0,1)$ の密度関数 $f(x)=I(0<x<1)$ なら，$n\geq 2$ に対して，f の n 回のたたみ込み $f^{n*}=f*\cdots*f$ が

$$f^{*n}(x)=\frac{1}{(n-1)!}\sum_{k=0}^{n}(-1)^k\binom{n}{k}(x-k)^{n-1}I(x\geq k)$$

と表せることを帰納法で示そう．$n=2$ ならこの表示は正しい．次に，ある $n\geq 2$ でこの表示が成り立つとすると，

$$f^{*(n+1)}(x)=(f^{n*}*f)(x)=\int_{-\infty}^{\infty}f^{*n}(x-y)f(y)dy=\int_0^1 f^{*n}(x-y)dy.$$

ここで，

$$\begin{aligned}
&\int_0^1(x-y-k)^{n-1}I(x-y\geq k)dy\\
&=\begin{cases}0 & x-k\leq 0\\ \int_0^{x-k}(x-y-k)^{n-1}dy=\frac{1}{n}(x-k)^n & 0<x-k<1\\ \int_0^1(x-y-k)^{n-1}dy=\frac{1}{n}\{(x-k)^n-(x-1-k)^n\} & x-k\geq 1\end{cases}\\
&=\frac{1}{n}\{(x-k)^nI(x\geq k)-(x-k-1)^nI(x\geq k+1)\}.
\end{aligned}$$

よって，

$$f^{*(n+1)}(x)=\frac{1}{n!}\sum_{k=0}^{n}(-1)^k\binom{n}{k}\{(x-k)^nI(x\geq k)-(x-k-1)^nI(x\geq k+1)\}.$$

ここで，$\binom{n}{k-1}+\binom{n}{k}=\binom{n+1}{k}$ という関係を使って整理して，

$$f^{*(n+1)}(x)=\frac{1}{n!}\sum_{k=0}^{n+1}(-1)^k\binom{n+1}{k}(x-k)^nI(x\geq k)$$

を得る．

1.9　多次元分布

多項分布

$p_1,\dots,p_k\geq 0$ を $\sum_{j=1}^k p_j=1$ をみたすとする．$e_j\in\mathbb{R}^k$ を j 番目の座標が 1 でその他の座標が 0 のベクトルとすると，

$$P\{(X_1,\dots,X_k)'=e_j\}=p_j$$

なる確率ベクトル $(X_1,\dots,X_k)'$ をパラメータ $p_1,\dots,p_k$ をもつ <u>多次元 Bernoulli 試行</u> と呼ぶ．各 j に対して，X_j は成功確率 p_j の Bernoulli 試行である．

次に，$(X_{i,1},\dots,X_{i,k})', i=1,\dots,n$ をパラメータ $p_1,\dots,p_k$ をもつ独立な多次元 Bernoulli 試行とすると，

$$(Y_1,\dots,Y_k)' = \left(\sum_{i=1}^n X_{i,1},\dots,\sum_{i=1}^n X_{i,k}\right)'$$

の従う分布をパラメータ $n,p_1,\dots,p_k$ をもつ <u>多項分布</u> (multinomial distribution) と呼び，

$$(Y_1,\dots,Y_k)' \sim Mn(n,p_1,\dots,p_k)$$

と書く．2 項分布の確率関数の導出と同様にして，$y_1+\cdots+y_k=n$ なる整数 $y_1,\dots,y_k \geq 0$ に対して，

$$P(Y_1=y_1,\dots,Y_k=y_k) = \frac{n!}{y_1!\cdots y_k!}p_1^{y_1}\cdots p_k^{y_k}$$

となる．ここで，各 j に対して，$Y_j=\sum_{i=1}^n X_{i,j}$ だから，

$$Y_j \sim Bin(n,p_j)$$

である．よって，

$$E[Y_j]=np_j, \quad \mathrm{Var}(Y_j)=np_j(1-p_j)$$

となる．また，$j\neq\ell$ に対しては，

$$\mathrm{Cov}(Y_j,Y_\ell) = -np_jp_\ell$$

だから (演習問題)，

$$\mathrm{Var}(Y) = \begin{pmatrix} np_1(1-p_1) & -np_1p_2 & \cdots & -np_1p_k \\ -np_2p_1 & p_2(1-p_2) & \cdots & -np_2p_k \\ \vdots & \vdots & \ddots & \vdots \\ -np_kp_1 & -np_kp_2 & \cdots & p_k(1-p_k) \end{pmatrix}$$

である．

さらに，$Y_1,\dots,Y_k$ を m 個のグループに分けて，

$$Z_1 = Y_1+\cdots+Y_{j_1},\dots,Z_m = Y_{j_{m-1}+1}+\cdots+Y_k,$$
$$q_1 = p_1+\cdots+p_{j_1},\dots,q_m = p_{j_{m-1}+1}+\cdots+p_k$$

とおくと $(j_0=1, j_m=k)$，各 $\ell=1,\dots,m$ に対して，$Z_\ell=\sum_{i=1}^n\sum_{j=j_{\ell-1}+1}^{j_\ell}X_{i,j}$ であって，

$$\left(\sum_{j=1}^{j_1}X_{i,j},\dots,\sum_{j=j_{m-1}+1}^{k}X_{i,j}\right)' \sim Mn(1,q_1,\dots,q_m)$$

だから，

$$(Z_1,\dots,Z_m)' \sim Mn(n,q_1,\dots,q_m)$$

である．

多変量正規分布

$X_1,\ldots,X_n$ を独立な r.v.'s とし，$X_j \sim N(0,1)$ とする．このとき，$X=(X_1,\ldots,X_n)'$ の分布を n 次元標準正規分布と呼び，$X \sim N(0,I_n)$ と書く．ここで，I_n は $n\times n$ の単位行列である：

$$I_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

X の密度関数は

$$f(x) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-x_i^2/2} = \frac{1}{(2\pi)^{n/2}} e^{-\sum_{i=1}^{n} x_i^2/2} = \frac{1}{(2\pi)^{n/2}} e^{-x'x/2},\ x \in \mathbb{R}^n$$

である．また，$\mu \in \mathbb{R}^n$ と $n\times n$ 行列 B に対して，$Y=\mu+BX$ とおくと，

$$E[Y]=\mu, \quad \mathrm{Var}(Y) = B\,\mathrm{Var}(X)B' = BB'$$

である．このとき，$\Sigma = BB'$ とおいて，Y の分布を平均ベクトル μ, 共分散行列 Σ をもつ多変量正規分布 (multivariate normal distribution) と呼び，$Y \sim N(\mu,\Sigma)$ と書く．

(1). 与えられた半正定値対称行列 Σ に対して，$\Sigma = BB'$ をみたす $n\times n$ 行列 B が存在する．従って，多変量正規分布 $N(\mu,\Sigma)$ はあらゆる $\mu \in \mathbb{R}^n$ と $n\times n$ 半正定値対称行列 Σ に対して定義される．また，$|\Sigma| = |B|^2$ より，

$$\Sigma \text{ が正則} \Leftrightarrow B \text{ が正則}$$

である．ここで注意すべきなのは，$\Sigma = BB'$ をみたす $n\times n$ 行列 B は一意でないことである．$N(\mu,\Sigma)$ の定義が B の選び方によらないことを確認しよう．$X \sim N(0,I_n)$ の特性関数は

$$\varphi_X(t) = E\left[e^{i\sum_{j=1}^{n} t_j X_j}\right] = \prod_{j=1}^{n} E[e^{it_jX_j}] = \prod_{j=1}^{n} e^{-t_j^2/2} = e^{-t't/2},\ t=(t_1,\ldots,t_n)' \in \mathbb{R}^n$$

だから，$Y=\mu+BX$ の特性関数は

$$\begin{aligned} \varphi_Y(t) &= E[e^{it'Y}] = E[e^{it'(\mu+BX)}] = e^{it'\mu}E[e^{i(B't)'X}] \\ &= \varphi_X(B't) = e^{it'\mu}e^{-t'BB't/2} = e^{it'\mu - t'\Sigma t/2} \end{aligned}$$

であって，B の選び方によらない．特性関数と分布は 1 対 1 に対応していることから，Y の分布は $\Sigma = BB'$ をみたす $n\times n$ 行列 B の選び方によらないことがわかる．

(2). B が正則のとき，Y の密度関数を求めてみよう．$y=\mu+Bx$ より，$x=B^{-1}(y-\mu)$ であって，変換 $y\mapsto x$ の Jacobian は $1/|B|$ である．ここで，$|\Sigma|=|B|^2$ より，$||B||=|\Sigma|^{1/2}$ だから，

$$\begin{aligned} f_Y(y) &= \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}(B^{-1}(y-\mu))'(B^{-1}(y-\mu))\right\} \\ &= \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu)\right\} \end{aligned}$$

となる．

$$(B^{-1})'B^{-1}=(B')^{-1}B^{-1}=(BB')^{-1}=\Sigma^{-1}$$

という関係を使った．

B が特異な場合，あるベクトル $a\in\mathbb{R}^n, a\neq 0$ が存在して，$a'B=0$ になる．このとき，$a'Y=a'\mu$ だから，Y は密度関数をもちえない (仮に Y が密度関数をもつなら，集合 $\{y:a'y=a'\mu\}$ の n 次元体積は 0 だから，$P(a'Y=a'\mu)=0$ になってしまう)．特異な共分散行列をもつ場合，多変量正規分布は退化している (degenerate) といわれる．

(3). $\mathrm{Cov}(Y_j,Y_k)=0\ \forall j\neq k$ なら，$Y_1,\dots,Y_n$ は独立になる．実際，$\mathrm{Cov}(Y_j,Y_k)=0\ \forall j\neq k$ なら，Σ は対角行列になる：

$$\Sigma=\begin{pmatrix}\sigma_1^2 & 0 & \cdots & 0\\ 0 & \sigma_2^2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \sigma_n^2\end{pmatrix}.$$

ここで，$\sigma_j^2=\mathrm{Var}(Y_j)$ である．そこで，

$$\widetilde{B}=\begin{pmatrix}\sigma_1 & 0 & \cdots & 0\\ 0 & \sigma_2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \sigma_n\end{pmatrix}$$

とおくと，$\widetilde{B}$ は $\Sigma=\widetilde{B}\widetilde{B}'$ をみたすから，

$$Y\overset{d}{=}\widetilde{B}X=(\sigma_1X_1,\dots,\sigma_nX_n)'$$

を得る．よって，$Y_1,\dots,Y_n$ は独立であって，$Y_j\sim N(0,\sigma_j^2)$ となることが示された．

(4). 任意の $m\times n$ 行列 A に対して，$AY\sim N(A\mu,A\Sigma A')$ となる．実際，Y の特性関数は $\varphi_Y(t)=E[e^{it'Y}]=e^{it'\mu-t'\Sigma t/2}$ だから，AY の特性関数は

$$E[e^{it'AY}]=E[e^{i(A't)'Y}]=\varphi_Y(A't)=e^{it'A\mu-t'A\Sigma A't/2}$$

である．これは $N(A\mu,A\Sigma A')$ の特性関数だから，$AY\sim N(A\mu,A\Sigma A')$ を得る．

(5). (4) より，Σ の第 (j,j) 成分を σ_j^2 とおくと，各 Y_j の周辺分布は $N(\mu_j, \sigma_j^2)$ である．すなわち，多変量正規分布に従う確率ベクトルの各成分の周辺分布は正規分布になる．この逆は成り立たない．すなわち，周辺分布が正規分布である r.v.'s を並べたベクトルが多変量正規分布に従うとは限らない．例えば，$U, V \sim N(0,1)$ を独立とし，

$$W = \begin{cases} U & \text{if } UV \geq 0 \\ -U & \text{if } UV < 0 \end{cases}$$

と定めると，$-U \stackrel{d}{=} U$ であって，$P(UV=0)=0$ だから，

$$P(W \leq x) = P(U \leq x, UV \geq 0) + P(-U \leq x, (-U)V > 0) = 2P(U \leq x, UV > 0).$$

ここで，$\{UV > 0\} = \{U > 0, V > 0\} \cup \{U < 0, V < 0\}$ だから，

$$P(W \leq x) = 2\{P(0 < U \leq x)P(V > 0) + P(U < \min\{x, 0\})P(V < 0)\} = P(U \leq x).$$

よって，$W \sim N(0,1)$ である．しかし，(U, W) は集合 $S = \{(u,w) : w = u \text{ or } w = -u\}$ に集中していて，S は面積 0 なので，(U,W) は同時密度をもたない．仮に (U,W) が多変量正規分布に従う場合，(U,W) が同時密度をもたないのは，(U,W) の共分散行列が特異な場合のみであって，それは $\mathrm{Corr}(U,W) = 1$ か $\mathrm{Corr}(U,W) = -1$, i.e., $P(W = U) = 1$ か $P(W = -U) = 1$ のいずれかの場合である．しかし，W の定義からそのいずれも起こり得ないので，(U,W) は多変量正規分布に従わないことが示された．

以上より，同時分布が多変量正規分布であることは，周辺分布が正規分布であることよりもずっと強い制約であるといえる．

(6). $\mathbb{R}^n$ の標準ノルムを $\|x\| = \sqrt{x'x}, x \in \mathbb{R}^n$ とし，$\mathbb{R}^n$ の単位球面を $\mathbb{S}^{n-1} = \{x \in \mathbb{R}^n : \|x\| = 1\}$ とおく．このとき，

$$U = (U_1, \ldots, U_n)' = \frac{X}{\|X\|}$$

は $\mathbb{S}^{n-1}$ に値をとる確率ベクトルである ($\|X\| = 0$ となる確率は 0 だから，そのような事象は無視している)．U は密度関数をもたない．しかし，

$$h(\theta) = h(\theta_1, \theta_2, \ldots, \theta_{n-1}) = \begin{pmatrix} \cos\theta_1 \\ \sin\theta_1 \cos\theta_2 \\ \vdots \\ \sin\theta_1 \cdots \sin\theta_{n-2} \cos\theta_{n-1} \\ \sin\theta_1 \cdots \sin\theta_{n-2} \sin\theta_{n-1} \end{pmatrix}$$

とおくと，極座標変換から，$\mathbb{S}^{n-1}$ 上の有界 (可測) 関数 g に対して，

$$E[g(U)] = E[g(X/\|X\|)] = \frac{1}{(2\pi)^{n/2}} \int g(x/\|x\|) e^{-\|x\|^2/2} dx$$
$$= \frac{1}{(2\pi)^{n/2}} \int_0^{2\pi} \int_0^{\pi} \cdots \int_0^{\pi} \underbrace{\left\{ \int_0^{\infty} r^{n-1} e^{-r^2/2} dr \right\}}_{=2^{n/2-1}\Gamma(n/2)} g(h(\theta)) \left\{ \prod_{i=1}^{n-2} (\sin\theta_i)^{n-i-1} \right\} d\theta_1 \cdots d\theta_{n-2} d\theta_{n-1}$$
$$= \frac{\Gamma(n/2)}{2\pi^{n/2}} \int_0^{2\pi} \int_0^{\pi} \cdots \int_0^{\pi} g(h(\theta)) \left\{ \prod_{i=1}^{n-2} (\sin\theta_i)^{n-i-1} \right\} d\theta_1 \cdots d\theta_{n-2} d\theta_{n-1}$$

が成り立つ．ここで，

$$\frac{2\pi^{n/2}}{\Gamma(n/2)}$$

は $\mathbb{S}^{n-1}$ の表面積だから，(Borel 集合) $A \subset \mathbb{S}^{n-1}$ に対して，

$$P(U \in A) = E[I_A(U)] = \frac{A\text{ の表面積}}{\mathbb{S}^{n-1}\text{ の表面積}}$$

となる．よって，U の分布を $\mathbb{S}^{n-1}$ 上の一様分布 と呼ぶ．

Dirichlet 分布

$\alpha_1, \dots, \alpha_k$ を正の実数とし，$Z_1, \dots, Z_k$ を独立な r.v.'s であって，各 $j = 1, \dots, k$ に対して，$Z_j \sim Ga(\alpha_j, 1)$ とする．このとき，

$$(Y_1, \dots, Y_k)' = \left(\frac{Z_1}{\sum_{j=1}^k Z_j}, \dots, \frac{Z_k}{\sum_{j=1}^k Z_j} \right)'$$

の従う分布をパラメータ $\alpha_1, \dots, \alpha_k$ をもつ Dirichlet 分布 と呼び，

$$(Y_1, \dots, Y_k)' \sim Di(\alpha_1, \dots, \alpha_k)$$

と書く．ここで，$Y_1 + \cdots + Y_k = 1$ なので，$(Y_1, \dots, Y_k)'$ は k 次元体積が 0 の集合 $\{(y_1, \dots, y_k)' : y_j > 0 \ (j = 1, \dots, k),\ \sum_{j=1}^k y_j = 1\}$ に集中しているため，密度関数をもたないが，$(Y_1, \dots, Y_{k-1})'$ は密度関数をもつ．それを求めてみよう．いま，$W = \sum_{j=1}^k Z_j$ とおいて，

$$y_1 = \frac{z_1}{\sum_{j=1}^k z_j}, \dots, y_{k-1} = \frac{z_{k-1}}{\sum_{j=1}^k z_j}, w = \sum_{j=1}^k z_j$$

を解くと，$z_1 = wy_1, \dots, z_{k-1} = wy_{k-1}, z_k = w(1 - \sum_{j=1}^{k-1} y_j)$ であるから，変換

$$(y_1, \dots, y_{k-1}, w)' \mapsto (z_1, \dots, z_{k-1}, z_k)'$$

の Jacobian は

$$\begin{vmatrix} w & 0 & \cdots & 0 & y_1 \\ 0 & w & \cdots & 0 & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & w & y_{k-1} \\ -w & -w & \cdots & -w & 1-\sum_{j=1}^{k-1} y_j \end{vmatrix} = w^{k-1}$$

である．ここで，分割行列の行列式の公式

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |A||D - CA^{-1}B|$$

を使った．いま，$(Z_1, \ldots, Z_k)'$ の密度関数は，

$$g(z_1, \ldots, z_k) = \frac{1}{\prod_{j=1}^k \Gamma(\alpha_j)} \left(\prod_{j=1}^k z^{\alpha_j - 1} \right) e^{-\sum_{j=1}^k z_j}$$

であるから，$(Y_1, \ldots, Y_{k-1}, W)'$ の密度関数は

$$\begin{aligned} g & \left(wy_1, \ldots, wy_{k-1}, w \left(1 - \sum_{j=1}^{k-1} y_j \right) \right) w^{k-1} \\ &= \frac{\Gamma(\sum_{j=1}^k \alpha_j)}{\prod_{j=1}^k \Gamma(\alpha_j)} \left(\prod_{j=1}^{k-1} y_j^{\alpha_j - 1} \right) \left(1 - \sum_{j=1}^{k-1} y_j \right)^{\alpha_k - 1} \times \frac{1}{\Gamma(\sum_{j=1}^k \alpha_j)} w^{\sum_{j=1}^k \alpha_j - 1} e^{-w} \end{aligned}$$

である．ここで，$y_1, \ldots, y_{k-1}, w$ は $y_j > 0\ (j = 1, \ldots, k-1)$，$\sum_{j=1}^{k-1} y_j < 1, w > 0$ と制約されている．従って，$(Y_1, \ldots, Y_{k-1})'$ と W は独立であって $(Y_1, \ldots, Y_{k-1}$ は $\sum_{j-1}^{k-1} y_j < 1$ という制約のため独立にはならない)，$(Y_1, \ldots, Y_{k-1})'$ は密度関数

$$f(y_1, \ldots, y_{k-1}) = \frac{\Gamma(\sum_{j=1}^k \alpha_j)}{\prod_{j=1}^k \Gamma(\alpha_j)} \left(\prod_{j=1}^{k-1} y_j^{\alpha_j - 1} \right) \left(\prod_{j=1}^{k-1} y_j^{\alpha_j - 1} \right) I_S(y_1, \ldots, y_{k-1})$$

をもつ．ここで，$S = \{(y_1, \ldots, y_{k-1})' : y_j > 0\ (j = 1, \ldots, k-1),\ \sum_{j=1}^{k-1} y_j < 1\}$ である．密度関数の形からもわかる通り，Dirichlet 分布はベータ分布の多変量への拡張である．

Y_j の周辺分布は，ガンマ分布の再生性から，$\sum_{i \neq j} Z_i \sim Ga(\sum_{i \neq j} \alpha_i, 1)$ だから，

$$Y_j = \frac{Z_j}{Z_j + \sum_{i \neq j} Z_i} \sim Be \left(\alpha_j, \sum_{i \neq j} \alpha_i \right)$$

である．よって，

$$E[Y_j] = \frac{\alpha_j}{\sum_{i=1}^k \alpha_i},\ \mathrm{Var}(Y_j) = \frac{\alpha_j \sum_{i \neq j} \alpha_i}{(\sum_{i=1}^k \alpha)^2 (\sum_{i=1}^k \alpha_i + 1)}$$

である．また，$j \neq \ell$ に対して，

$$\mathrm{Cov}(Y_j, Y_\ell) = -\frac{\alpha_j \alpha_\ell}{(\sum_{i=1}^k \alpha)^2 (\sum_{i=1}^k \alpha_i + 1)}$$

である (演習問題).

さらに，ガンマ分布の再生性より，$r_1, \ldots, r_\ell$ を $0 < r_1 < \cdots < r_\ell = k$ なる整数とすると，

$$\left(\sum_{i=1}^{r_1} Y_i, \ldots, \sum_{i=r_{\ell-1}+1}^{r_\ell} Y_i \right)' \sim Di \left(\sum_{i=1}^{r_1} \alpha_i, \ldots, \sum_{i=r_{\ell-1}+1}^{r_\ell} \alpha_i \right)$$

となる．

1.10　特性関数に関する補足

本節では，特性関数が分布を一意に決めることと (Theorem 1.7)，その系として，モーメント母関数が存在すれば分布を一意に決めること (Theorem 1.6) を証明する．簡単のために 1 次元の場合を考える．次の技術的な補題が重要である．

Lemma 1.11. X を r.v. とし，その特性関数を φ とする．また，$\sigma > 0$ に対して $Z^\sigma \sim N(0, \sigma^2)$ であって，X と独立とする．このとき，有界区間の外側では 0 になる連続関数 $g : \mathbb{R} \to \mathbb{R}$ に対して，

$$E[g(X + Z^\sigma)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} \widehat{g}(t) e^{-\sigma^2 t^2/2} \varphi(-t) dt$$

が成り立つ．ここで，

$$\widehat{g}(t) = \int_{-\infty}^{\infty} g(x) e^{itx} dx$$

である．

Proof. $x \in \mathbb{R}$ に対して，

$$E[g(x + Z^\sigma)] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} g(x + z) e^{-z^2/(2\sigma^2)} dz = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} g(u) e^{-(u-x)^2/(2\sigma^2)} du.$$

ここで，$N(0,1)$ の特性関数が $e^{-t^2/2}$ であることから，

$$e^{-t^2/2} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2 + itx} dx$$

が成り立つ．変数を入れ替えて，

$$e^{-(u-x)^2/(2\sigma^2)} = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\sigma^2 t^2/2 + i(u-x)t} dt$$

を得る．よって，

$$E[g(x+Z^\sigma)] = \frac{1}{2\pi}\iint g(u)e^{itu}e^{-\sigma^2t^2/2}e^{-itx}dtdu = \frac{1}{2\pi}\int_{-\infty}^{\infty}\widehat{g}(t)e^{-\sigma^2t^2/2}e^{-itx}dt$$

を得る．積分順序の交換は Fubini の定理から保証される．従って，再び Fubini の定理より，

$$E[g(X+Z^\sigma)] = E[E[g(x+Z^\sigma)]|_{x=X}] = \frac{1}{2\pi}\int_{-\infty}^{\infty}\widehat{g}(t)e^{-\sigma^2t^2/2}\underbrace{E[e^{-itX}]}_{=\varphi(-t)}dt$$

を得る． □

Theorem 1.7 を証明しよう．そのステートメントを再掲する．

Theorem 1.9. $X \sim F, Y \sim G$ に対して，特性関数をそれぞれ φ_F, φ_G とする．このとき，$\varphi_F \equiv \varphi_G$ ならば $F \equiv G$ である．

Proof. $\sigma > 0$ に対して $Z^\sigma \sim N(0,\sigma^2)$ として，(X,Y) と独立とする．任意の $a < b, \varepsilon > 0$ に対して，

$$g_{a,b,\varepsilon}(x) = \begin{cases} 0 & x < a-\varepsilon \\ \text{linear} & a-\varepsilon \le x < a \\ 1 & a \le x \le b \\ \text{linear} & b < x \le b+\varepsilon \\ 0 & x > b+\varepsilon \end{cases}$$

とおくと，

$$I_{(a,b]}(x) \le g_{a,b,\varepsilon}(x) \le I_{(a-\varepsilon,b+\varepsilon]}(x),\ \forall x \in \mathbb{R}, \quad |g_{a,b,\varepsilon}(x) - g_{a,b,\varepsilon}(y)| \le \frac{|x-y|}{\varepsilon},\ \forall x,y \in \mathbb{R}$$

をみたす． よって，

$$0 \le E[g_{a,b,\varepsilon}(X)] - P(a < X \le b) \le P(a-\varepsilon < X \le a) + P(b < X \le b+\varepsilon)$$

であって，

$$|E[g_{a,b,\varepsilon}(X)] - E[g_{a,b,\varepsilon}(X+Z^\sigma)]| \le E[|g_{a,b,\varepsilon}(X) - g_{a,b,\varepsilon}(X+Z^\sigma)|] \le \frac{E[|Z^\sigma|]}{\varepsilon} \le \frac{\sigma}{\varepsilon}.$$

同様の評価が X を Y に替えても成り立つ．ここで，$\varphi_F \equiv \varphi_G$ と (*) より，$E[g_{a,b,\varepsilon}(X+Z^\sigma)] = E[g_{a,b,\varepsilon}(Y+Z^\sigma)]$ だから，

$$\begin{aligned}|P(a < X \le b) - P(a < Y \le b)| &\le P(a-\varepsilon < X \le a) + P(b < X \le b+\varepsilon) \\ &\quad + P(a-\varepsilon < Y \le a) + P(b < Y \le b+\varepsilon) + \frac{2\sigma}{\varepsilon}\end{aligned}$$

を得る．さらに，$a \to -\infty, \sigma \downarrow 0, \varepsilon \downarrow 0$ の順に極限をとって，$F(b) = G(b)$ を得る． □

Theorem 1.6 の証明に行く前に，特性関数に関する重要な結果である，Riemann-Lebesgue の補題を証明しよう．

Theorem 1.10 (Riemann-Lebesgue の補題)**.** *X は連続な密度関数をもつとし，その特性関数を φ とおく．このとき，$\varphi(t) \to 0$ $(|t| \to \infty)$ となる．*

Riemann-Lebesgue の補題において，密度関数の連続性は必要ないのであるが，ここでは簡単のために連続性を仮定しておく．

Proof. f は密度関数だから，$M \to \infty$ のとき，

$$\int_{|x|>M} f(x)dx \to 0$$

となる．そこで，任意の $\varepsilon > 0$ に対して，$M = M_\varepsilon$ を十分大きく選んで，

$$\int_{|x|>M} f(x)dx \le \varepsilon$$

としておく．f は連続なので，$[-M, M]$ 上で一様連続である．すなわち，ある $\delta > 0$ が存在して，$x, y \in [-M, M], |x - y| \le \delta$ ならば $|f(x) - f(y)| \le \varepsilon/(2M)$ となる．そこで，分点

$$-M = a_0 < a_1 < \cdots < a_k = M$$

を $a_{j-1} - a_j \le \delta$ $(j = 1, \ldots, k)$ をみたすように選び，

$$f_\varepsilon(x) = \sum_{j=1}^{k} b_j I_{[a_{j-1}, a_j)}(x),\ b_j = f(a_{j-1})\ (j = 1, \ldots, k)$$

とおくと，

$$\int_{\mathbb{R}} |f(x) - f_\varepsilon(x)|dx \le \int_{|x|>M} f(x)dx + \int_{-M}^{M} |f(x) - f_\varepsilon(x)|dx \le 2\varepsilon$$

となる．よって，

$$\left|\int_{\mathbb{R}} \cos(tx)f(x)dx - \int_{\mathbb{R}} \cos(tx)f_\varepsilon(x)dx\right| \le \int_{\mathbb{R}} |f(x) - f_\varepsilon(x)|dx \le 2\varepsilon,$$
$$\left|\int_{\mathbb{R}} \sin(tx)f(x)dx - \int_{\mathbb{R}} \sin(tx)f_\varepsilon(x)dx\right| \le 2\varepsilon$$

であるから，

$$\lim_{|t|\to\infty} \int_{\mathbb{R}} \cos(tx)f_\varepsilon(x)dx = 0,\ \lim_{|t|\to\infty} \int_{\mathbb{R}} \sin(tx)f_\varepsilon(x)dx = 0$$

を示せばよい．ここで，$t \ne 0$ に対して，

$$\int_{\mathbb{R}} \cos(tx)f_\varepsilon(x)dx = \sum_{j=1}^{k} b_j \int_{a_{j-1}}^{a_j} \cos(tx)dx = \frac{1}{t}\sum_{j=1}^{k} b_j\{\sin(ta_j) - \sin(ta_{j-1})\}$$

であって，右辺は $|t| \to \infty$ のとき 0 に収束する．残りの場合も同様である． □

Riemann-Lebesgue の補題の結論は X が離散の場合には成り立たない．例えば，X が整数値なら，$P(X=k)=p(k)$ とおくと，$\varphi(t)=\sum_k p(k)e^{itk}$ となって，φ は周期 2π の周期関数になる．このとき，$\varphi(2\pi\ell)=1$ $(\ell\in\mathbb{Z})$ だから，$\varphi(t)$ は $|t|\to\infty$ のとき 0 に収束しない．もっと一般に，X が離散分布に従っている場合，

$$\limsup_{|t|\to\infty}|\varphi(t)|=1$$

となることが知られている．

次に，Theorem 1.6 を証明する．そのステートメントを再掲する．

Theorem 1.11. $X\sim F, Y\sim G$ に対して，それぞれモーメント母関数 ψ_F,ψ_G が存在するとする．このとき，十分小さい $\varepsilon>0$ に対して，

$$\psi_F(\theta)=\psi_G(\theta)\ \forall|\theta|<\varepsilon$$

ならば $F\equiv G$ である．

Proof. ψ_F を複素平面の領域 $D=\{\theta+it:|\theta|<\varepsilon,-\infty<t<\infty\}$ に拡張する：

$$\widetilde{\psi}_F(z)=\widetilde{\psi}_F(\theta,t)=E[e^{(\theta+it)X}],\ z=\theta+it,\quad |\theta|<\varepsilon,-\infty<t<\infty.$$

ここで，$E[e^{\theta X}]<\infty\ \forall|\theta|<\varepsilon$ より，$\widetilde{\psi}_F$ は複素数値関数としてちゃんと定義されている．$\widetilde{\psi}_F$ が D 上で正則であることを確認しよう．そのためには $\widetilde{\psi}_F(\theta,t)$ が (θ,t) について連続微分可能であって，Cauchy-Riemann の方程式

$$\frac{\partial\widetilde{\psi}_F}{\partial\theta}(\theta,t)+i\frac{\partial\widetilde{\psi}_F}{\partial t}(\theta,t)=0$$

をみたすことを確認すればよい．いま，

$$\frac{\partial}{\partial\theta}e^{(\theta+it)X}=Xe^{(\theta+it)X},\quad \frac{\partial}{\partial t}e^{(\theta+it)X}=iXe^{(\theta+it)X}$$

であって，$|\theta|<\varepsilon$ において期待値と偏微分の交換が正当化できて，

$$\frac{\partial\widetilde{\psi}_F}{\partial\theta}(\theta,t)=E[Xe^{(\theta+it)X}],\ \frac{\partial\widetilde{\psi}_F}{\partial t}(\theta,t)=iE[Xe^{(\theta+it)X}]$$

となるから (Lebesgue の優収束定理による)，Cauchy-Riemann の方程式がみたされる．さらに，再び Lebesgue の優収束定理より，$(\theta,t)\mapsto E[Xe^{(\theta+it)X}]$ が $|\theta|<\varepsilon$ において連続であることが示せるから，$\widetilde{\psi}_F$ は D 上で正則である．同様に，ψ_G も D 上の正則関数 $\widetilde{\psi}_G$ に拡張できる．ここで仮定より，$\widetilde{\psi}_F$ と $\widetilde{\psi}_G$ は $\{\theta\in\mathbb{R}:|\theta|<\varepsilon\}$ 上で一致しているので，正則関数に対する一致の定理より，$\widetilde{\psi}_F(z)=\widetilde{\psi}_G(z)\ \forall z\in D$ を得る．これから $E[e^{itX}]=E[e^{itY}]\ \forall t\in\mathbb{R}$, i.e., $F\equiv G$ を得る． □

モーメント母関数は存在すれば分布を一意に決めることから，r.v. X に対して，有限な k 次モーメント $m_k = E[X^k], k = 1, 2, \ldots$ がすべて存在するなら，モーメント列 $\{m_k\}_{k=1}^{\infty}$ から X の分布が一意に決まるであろうか．実はそうでないことが次の例からわかる．

Example 1.13 (Heyde (1963) の例)**.** $Z \sim N(0,1)$ に対して，$X = e^Z$ とおく．ここで，$x > 0$ に対して，$P(X \le x) = P(\log X \le \log x) = \Phi(\log x)$ だから，両辺を x で微分して，X は密度関数

$$f_X(x) = \frac{1}{\sqrt{2\pi}} x^{-1} e^{-(\log x)^2/2},\ x > 0$$

をもつことがわかる．X の分布のことを 対数正規分布 (log-normal distribution) と呼ぶ．いま，Z のモーメント母関数は $\psi_Z(\theta) = e^{\theta^2/2}$ だから，$k = 1, 2, \ldots$ に対して，

$$E[X^k] = E[e^{kZ}] = \psi(k) = e^{k^2/2}$$

である．一方，Y を密度関数

$$f_Y(y) = f_X(y)(1 + \sin(2\pi \log y)),\ y > 0$$

をもつ r.v. とする．f_Y がちゃんと確率密度関数になっていることは，

$$\int_0^{\infty} f_X(y) \sin(2\pi \log y) dy = E[\sin(2\pi \log X)] = E[\sin(2\pi Z)] = 0$$

から確認できる．さらに，$k = 1, 2, \ldots$ に対して，

$$\int_0^{\infty} y^k f_X(y) \sin(2\pi \log y) dy = E[e^{kZ} \sin(2\pi Z)] = \frac{e^{k^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sin(2\pi(z-k)) e^{-(z-k)^2/2} dz$$
$$= \frac{e^{k^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sin(2\pi z) e^{-z^2/2} dz = 0$$

だから，$E[Y^k] = E[X^k]$ である．

1.11 エントロピーと KL ダイバージェンス

本節の内容は，これ以降の内容と独立しているが，それ自身面白いトピックである．なお本節の内容は，Cover and Thomas (2006) の Chapter 8 を参考にしている．f を $\mathbb{R}^k$ 上の密度関数とし，$X \sim f$ とする．このとき，X の エントロピー (entropy) を

$$H(X) = -\int f(x) \log f(x) dx = -E[\log f(X)]$$

と定義する．積分範囲は $\{x : f(x) > 0\}$ と理解する．f が有限集合上の確率関数の場合は log を底 2 の対数関数 $\log_2$ に置き替えてエントロピーを同様に定義し，その場合，エントロピーの情報理論的な解釈が知られている．詳しくは，Cover and Thomas (2006) を参照せよ．これ以降の議論において，有限な積分の存在を常に仮定する．

Example 1.14. $\mu \in \mathbb{R}^k$ とし，Σ を $k \times k$ 正定値対称行列とする．このとき，$X \sim N(\mu, \Sigma)$ に対して，エントロピーを計算してみると，

$$H(X) = \frac{1}{2} \log\{(2\pi e)^k |\Sigma|\}$$

である．

次に，X, Y を確率ベクトルとし，同時密度 $f(x, y)$ をもつとする．このとき，X を与えたときの X の 条件付きエントロピー を

$$H(Y \mid X) = -\iint f(x, y) \log f_{Y|X}(y \mid x) dydx$$

と定義する．ここで，$\log f(x, y) = \log f_{Y|X}(y \mid x) + \log f_X(x)$ なので，

$$\begin{aligned}
&\iint f(x, y) \log f_{Y|X}(y \mid x) dydx \\
&= \iint f(x, y) \log f(x, y) dydx - \int f_X(x) \log f_X(x) dx \\
&= -H(X, Y) + H(X)
\end{aligned}$$

である．

確率ベクトルの列 $X_1, \ldots, X_n$ に対して，その同時密度 $f(x_1, \ldots, x_n)$ の存在を仮定する．$(X_1, \ldots, X_{i-1})$ が与えられたときの X_i の条件付き密度を $f(x_i \mid x_1, \ldots, x_{i-1})$ と書くことにすれば，

$$f(x_1, \ldots, x_n) = f(x_1) f(x_2 \mid x_1) \cdots f(x_n \mid x_1, \ldots, x_{n-1})$$

と分解できるから，

$$H(X_1, \ldots, X_n) = H(X_1) + H(X_2 \mid X_1) + \cdots + H(X_n \mid X_1, \ldots, X_{n-1})$$

を得る．

次に，エントロピーに関連する概念として，Kullback-Leibler ダイバージェンス (KL ダイバージェンス) を定義する．f, g を $\mathbb{R}^k$ 上の密度関数とし，$\{x : g(x) > 0\} \supset \{x : f(x) > 0\}$ とする．このとき，f の g に対する KL ダイバージェンスを

$$D(f \| g) = \int f(x) \log \frac{f(x)}{g(x)} dx$$

と定義する．積分範囲は $\{x : f(x) > 0\}$ と理解する．f, g が確率関数の場合は，積分を和を取り替える．以下では，密度関数の場合を考える．KL ダイバージェンスは分布間のある種の距離と解釈できる．その理由は次の定理による．

Theorem 1.12. $D(f||g) \geq 0$ であって，等号が成立するのは，“ほとんどすべての $x \in \mathbb{R}^k$” に対して，$f(x) = g(x)$ となるときのみである．

Proof. $y > 0$ に対して，$(\log y)'' = -1/y^2$ だから，Taylor の定理より，$\log y \leq y - 1$ であって，等号が成立するのは $y = 1$ のときのみである．$y = \log \frac{g(x)}{f(x)}$ を代入して，$f(x)$ について積分をとると，

$$-D(f||g) \leq \int_{\{f>0\}} \left\{ \frac{g(x)}{f(x)} - 1 \right\} f(x)dx = \int g(x)dx - \int f(x)dx = 1 - 1 = 0$$

である．ここで，等号が成立するのは，ほとんどすべての $x \in \mathbb{R}^k$ に対して，$f(x) = g(x)$ となるときのみである． □

従って，f と g が異なる分布であれば $D(f||g) > 0$ であって，同じ分布であれば $D(f||g) = 0$ となる．しかし，KL ダイバージェンスは対称性と三角不等式をみたさないので，数学的な意味で距離になっているわけではない．なお，$D(f||g) \geq 0$ を言い換えると，

$$\int f(x) \log f(x)dx \geq \int f(x) \log g(x)dx$$

であって，右辺を最大化する g は $g = f$ で与えられる，というようにも解釈できる．

KL ダイバージェンスを使って，確率ベクトル間の従属関係の尺度を与えよう．X, Y を確率ベクトルとし，同時密度 $f(x, y)$ をもつとする．このとき，X と Y の間の 相互情報量 (mutual information) を

$$I(X;Y) = D(f(x,y)||f_X(x)f_Y(y)) = \iint f(x,y) \log \frac{f(x,y)}{f_X(x)f_Y(y)} dxdy$$

と定義する．$f_X(x)f_Y(y)$ は仮に X, Y が独立だった場合の同時密度であって，$I(X;Y)$ は真の同時密度 $f(x, y)$ からの，独立な場合の同時密度 $f_X(x)f_Y(y)$ の乖離を測っていると解釈できる．ここで，KL ダイバージェンスの性質から，次の系を得る．

Corollary 1.5. $I(X;Y) \geq 0$ であって，等号が成立するのは X と Y が独立なときのみである．

さらに，

$$\log \frac{f(x,y)}{f_X(x)f_Y(y)} = \log \frac{f_{Y|X}(y \mid x)}{f_Y(y)} = \log f_{Y|X}(y \mid x) - \log f_Y(y)$$

であるから，

$$I(X;Y) = H(Y) - H(Y \mid X)$$

とも表せる．よって，次の系を得る．

Corollary 1.6. $H(Y \mid X) \le H(Y)$ であって，等号が成立するのは X と Y が独立なときのみである．

ここで，エントロピーを使って，次の正規分布の面白い特徴づけを証明しよう．

Theorem 1.13. Σ を $k \times k$ 正定値対称行列とし，X を平均ベクトルが 0，共分散行列が Σ の確率ベクトルとする (X は密度関数をもつとする)．このとき，

$$H(X) \le \frac{1}{2} \log\{(2\pi e)^k |\Sigma|\}$$

であって，等号が成立するのは $X \sim N(0, \Sigma)$ のときのみである．

Proof. X の密度関数を g とし，$N(0, \Sigma)$ の密度関数を ϕ_Σ とおく：

$$\phi_\Sigma(x) = \frac{1}{(2\pi)^{k/2} |\Sigma|^{1/2}} e^{-x'\Sigma^{-1}x/2}.$$

このとき，

$$0 \le D(g || \phi_\Sigma) = \int g(x) \log \frac{g(x)}{\phi_\Sigma(x)} dx = -H(X) - \int g(x) \log \phi_\Sigma(x) dx.$$

ここで，$\log \phi_\Sigma(x) = -(1/2) \log\{(2\pi)^k |\Sigma|\} - x'\Sigma^{-1}x/2$ であって，

$$\begin{aligned}
\int g(x)(x'\Sigma^{-1}x) dx &= E[X'\Sigma^{-1}X] = E[\mathrm{tr}(X'\Sigma^{-1}X)] = E[\mathrm{tr}(\Sigma^{-1}XX')] \\
&= \mathrm{tr}\, E[\Sigma^{-1}XX'] = \mathrm{tr}\, \Sigma^{-1}E[XX'] = \mathrm{tr}\, \Sigma^{-1}\Sigma = \mathrm{tr}\, I_k = k.
\end{aligned}$$

よって，

$$0 \le D(g || \phi_\Sigma) = -H(X) + \frac{1}{2} \log\{(2\pi e)^k |\Sigma|\}$$

であって，右辺が 0 になるのは，$X \sim N(0, \Sigma)$ のときのみである． □

これまでに証明したエントロピーに対する不等式を利用して，いくつかの行列不等式を証明できる．最初に Hadamard の不等式を証明しよう．

Theorem 1.14 (Hadamard の不等式)**.** Σ を $k \times k$ 正定値対称行列とし，その対角成分を $\sigma_{11}, \ldots, \sigma_{kk}$ とする．このとき，

$$|\Sigma| \le \prod_{j=1}^{k} \sigma_{jj}. \tag{*}$$

Proof. $X = (X_1, \ldots, X_k)' \sim N(0, \Sigma)$ とすると，

$$H(X) \le H(X_1) + H(X_2 \mid X_1) + \cdots + H(X_k \mid X_1, \ldots, X_{k-1}) \le H(X_1) + \cdots + H(X_k)$$

であって，左辺と右辺を比較して (*) を得る． □

Theorem 1.15. Σ_1, Σ_2 を $k \times k$ 正定値対称行列とし，$\lambda \in [0,1]$ とする．このとき，

$$|\lambda\Sigma_1 + (1-\lambda)\Sigma_2| \geq |\Sigma_1|^{\lambda}|\Sigma_2|^{1-\lambda}. \tag{**}$$

言い換えると，$\log|\Sigma|$ は $k \times k$ 正定値対称行列のなす空間の上で凹関数になる．

Proof. $\theta \sim Bin(1,\lambda), X \sim N(0,\Sigma_1), Y \sim N(0,\Sigma_2)$ とし，θ, X, Y は独立とする．さらに，

$$Z = \begin{cases} X & \theta = 1 \\ Y & \theta = 0 \end{cases}$$

とおく．このとき，Z は密度関数

$$g(x) = \lambda\phi_{\Sigma_1}(x) + (1-\lambda)\phi_{\Sigma_2}(x)$$

をもつから，$E[Z] = 0, E[ZZ'] = \lambda\Sigma_1 + (1-\lambda)\Sigma_2$ である．よって，Theorem 1.13 より，

$$\frac{1}{2}\log\{(2\pi e)^k|\lambda\Sigma_1 + (1-\lambda)\Sigma_2|\} \geq H(Z).$$

ここで，KL ダイバージェンスの非負性より，

$$\begin{aligned} H(Z) &= -\lambda\int \phi_{\Sigma_1}(x)\log g(x)dx - (1-\lambda)\int \phi_{\Sigma_2}(x)\log g(x)dx \\ &\geq -\lambda\int \phi_{\Sigma_1}(x)\log\phi_{\Sigma_1}(x)dx - (1-\lambda)\int \phi_{\Sigma_2}(x)\log\phi_{\Sigma_2}(x)dx \\ &= \frac{\lambda}{2}\log\{(2\pi e)^k|\Sigma_1|\} + \frac{1-\lambda}{2}\log\{(2\pi e)^k|\Sigma_2|\}. \end{aligned}$$

以上より，(**) を得る． □

2　標本分布論

推測統計では，標本はある分布に従う確率変数列とみなし，背後にある分布 (母集団分布) のパラメータに関して，推定，検定，区間推定を行う．

$X_1, \ldots, X_n$ を独立な r.v.'s とし，各 X_i は d.f. F に従うとする．このとき，

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

と書く．F が母集団分布である．$X = (X_1, \ldots, X_n)'$ の関数

$$T(X) = T(X_1, \ldots, X_n) \in \mathbb{R}$$

を X の 統計量 (statistic) と呼ぶ．統計量の分布を 標本分布 (sampling distribution) と呼ぶ．$T_1(X), \ldots, T_m(X)$ を統計量とすると，確率ベクトル $T(X) = (T_1(X), \ldots, T_m(X))'$ を m 次元の統計量と呼ぶ．各 X_i が多次元のときも同様に，$X = (X_1', \ldots, X_n')'$ の関数を X の統計量と呼ぶ．

2.1　正規分布のもとでの標本分布

$X_1, \ldots, X_n \sim N(0, 1)$ i.i.d. とし，

$$Y = X_1^2 + \cdots + X_n^2$$

とおく．Y の従う分布を自由度 n の χ^2 分布 と呼び，$Y \sim \chi^2(n)$ と書く．Y の密度関数を求めてみよう．いま，$x > 0$ に対して，

$$P(X_1^2 \le x) = P(-\sqrt{x} \le X_1 \le \sqrt{x}) = \int_{-\sqrt{x}}^{\sqrt{x}} \phi(y)dy = 2\int_0^{\sqrt{x}} \phi(y)dy$$

であって，両辺を x について微分して，

$$\frac{d}{dx}P(X_1^2 \le x) = x^{-1/2}\phi(\sqrt{x}) = \frac{1}{\sqrt{2\pi}}x^{-1/2}e^{-x/2}$$

となる．右辺は $Ga(1/2, 2)$ の密度関数であるから，$X_1^2 \sim Ga(1/2, 2)$ である．よって，ガンマ分布の再生性より，

$$\chi^2(n) = Ga(n/2, 2)$$

であって，その密度関数は

$$f_Y(y) = \frac{1}{\Gamma(n/2)2^{n/2}}y^{n/2-1}e^{-y/2}I(y > 0)$$

である．$n = 2$ のときは $\chi^2(2) = Ga(1, 2) = Ex(1/2)$ である．ガンマ分布に対する平均・分散の公式から，$Y \sim \chi^2(n)$ に対して，

$$E[Y] = n,\ \mathrm{Var}(Y) = 2n$$

である．これは χ^2 分布の定義から直接計算することもできる．
　いま，$n \geq 2$ として，

$$X_1, \ldots, X_n \sim N(\mu, \sigma^2) \ \text{ i.i.d.}, \quad \mu \in \mathbb{R}, \sigma^2 > 0$$

に対して，

$$\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i,\ S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$$

とおく．$\overline{X}$ は <u>標本平均</u> (sample mean)，S^2 は <u>標本分散</u> (sample variance) と呼ばれる．

Theorem 2.1. $\overline{X}$ と S^2 は独立であって，$\overline{X} \sim N(\mu, \sigma^2/n), (n-1)S^2/\sigma^2 \sim \chi^2(n-1)$.

Proof. $\mu = 0, \sigma^2 = 1$ と仮定する．$X = (X_1, \ldots, X_n)'$ とおく．このとき，$X \sim N(0, I_n)$ である．$n \times n$ 行列 G を 1 行目が

$$(1/\sqrt{n}, 1/\sqrt{n}, \ldots, 1/\sqrt{n})$$

であって，$k = 2, 3, \ldots, n$ に対して，k 行目が

$$(\underbrace{1, \ldots, 1}_{k-1}, -k+1, 0, \ldots, 0)/\sqrt{k(k-1)}$$

となる行列とする．例えば，$n = 3$ なら，

$$G = \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} \end{pmatrix}$$

である．G の各行は直交しているので，

$$G'G = GG' = I_n$$

をみたす．すなわち，G は直交行列である．G は <u>Helmert 変換</u> と呼ばれる．
　そこで，$Y = GX$ とおくと，$GG' = I_n$ より，$Y \sim N(0, I_n)$ である．いま，

$$\sum_{i=1}^n X_i^2 = X'X = (GX)'(GX) = Y'Y = \sum_{i=1}^n Y_i^2$$

であって，$Y_1 = \sum_{i=1}^n X_i/\sqrt{n} = \sqrt{n}\overline{X}$ である．これより，

$$\sum_{i=1}^n (X_i - \overline{X})^2 = \sum_{i=1}^n X_i^2 - n\overline{X}^2 = \sum_{i=2}^n Y_i^2$$

を得る．よって，

$$\overline{X} = Y_1/\sqrt{n} \sim N(0, 1/n),\ (n-1)S^2 = \sum_{i=2}^{n} Y_i^2 \sim \chi^2(n-1)$$

であって，$Y_1, Y_2, \ldots, Y_n$ の独立性より，$\overline{X}$ と S^2 は独立である[13]． □

次に，$U \sim N(0,1), V \sim \chi^2(m)$ とし，U と V は独立とする．このとき，

$$T = \frac{U}{\sqrt{V/m}}$$

の分布を自由度 m の <u>t 分布</u> と呼び，$T \sim t(m)$ と書く．

Example 2.1. $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ i.i.d. に対して，$S = \sqrt{S^2}$ として，

$$T = \frac{\sqrt{n}(\overline{X} - \mu)}{S} \tag{*}$$

とおくと，

$$T = \frac{\sqrt{n}(\overline{X} - \mu)/\sigma}{\sqrt{S^2/\sigma^2}}$$

であるから，$T \sim t(n-1)$ である．(*) の T を <u>t 統計量</u> と呼ぶ．

Theorem 2.2. $t(m)$ の密度関数は

$$f_T(t) = \frac{\Gamma((m+1)/2)}{\sqrt{\pi m}\Gamma(m/2)} \left(1 + \frac{t^2}{m}\right)^{-(m+1)/2}, \quad t \in \mathbb{R}$$

である．

Proof. (U, V) の同時密度は

$$f(u, v) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \frac{v^{m/2-1} e^{-v/2}}{2^{m/2}\Gamma(m/2)},\ u \in \mathbb{R}, v > 0$$

である．ここで，

$$T = \frac{U}{\sqrt{V/m}},\ V = V$$

という変換を考える．$t = u/\sqrt{v/m}, v = v$ を解くと，$u = t\sqrt{v/m}, v = v$ であるから，変換 $(t, v) \mapsto (u, v)$ の Jacobian は

$$\begin{vmatrix} \sqrt{v/m} & * \\ 0 & 1 \end{vmatrix} = \sqrt{v/m}$$

[13] もっとちゃんというと，$Y_1, \ldots, Y_n$ の独立性から，Y_1 と $(Y_2, \ldots, Y_n)'$ は独立である．$\overline{X}$ は Y_1 のみの関数であって，S^2 は $Y_2, \ldots, Y_n$ のみの関数だから，$\overline{X}$ と S^2 の独立性が従う．

である．よって，(T,V) の同時密度は

$$g(t,v)=\frac{1}{\sqrt{2\pi}}e^{-t^2v/(2m)}\frac{v^{m/2-1}e^{-v/2}}{2^{m/2}\Gamma(m/2)}\sqrt{v/m}=\frac{v^{(m+1)/2-1}e^{-v(1+t^2/m)/2}}{2^{m/2}\Gamma(m/2)\sqrt{2\pi m}}$$

である．これを v について 0 から ∞ まで積分すると，$\alpha=(m+1)/2,\beta^{-1}=(1+t^2/m)/2$ とおくと，

$$\int_0^\infty v^{\alpha-1}e^{-v/\beta}dv=\beta^\alpha\Gamma(\alpha)=2^{(m+1)/2}\Gamma((m+1)/2)\left(1+\frac{t^2}{m}\right)^{-(m+1)/2}$$

となるから，求める結論を得る． □

t 分布の性質をまとめておこう．$t(m)$ の密度関数を f_m とおく．

- $m\to\infty$ のとき，Stirling の公式より，各 $t\in\mathbb{R}$ に対して $f_m(t)\to\phi(t)$ となる (演習問題)．
- 一方，$m=1$ のときは

  $$f_1(t)=\frac{1}{\pi(1+t^2)}$$

  だから，$t(1)$ は Cauchy 分布である．
- $Y\sim t(m)$ とすると，

  $$E[|Y|^r]\begin{cases}<\infty & 0<r<m\\ =\infty & r\geq m\end{cases}$$

  である (演習問題)．すなわち，t 分布は裾の重さについて，任意次の有限モーメントをもつ正規分布と，1 次の有限モーメントももたない Cauchy 分布との間を補間する分布といえる．ただし，t 分布はモーメント母関数をもたない．

$U\sim\chi^2(\ell),V\sim\chi^2(m)$ とし，U,V は独立とする．このとき，

$$Y=\frac{U/\ell}{V/m}$$

の分布を自由度 (ℓ,m) の <u>F 分布</u> と呼び，$Y\sim F(\ell,m)$ と書く．

Theorem 2.3. $Y\sim F(\ell,m)$ の密度関数は

$$f_Y(y)=\frac{\ell^{\ell/2}m^{m/2}}{B(\ell/2,m/2)}\frac{y^{\ell/2-1}}{(m+\ell y)^{(\ell+m)/2}},\quad y>0$$

である．

Proof. 最初に $\widetilde{Y} = U/V$ の密度関数を求める．

$$Z = \frac{\widetilde{Y}}{1+\widetilde{Y}} = \frac{U}{U+V} = \frac{U/2}{U/2+V/2}$$

とおくと，$U/2 \sim Ga(\ell/2, 1), V/2 \sim Ga(m/2, 1)$ であって，U と V は独立であるから，$Z \sim Be(\ell/2, m/2)$ である．

$$f_Z(z) = \frac{1}{B(\ell/2, m/2)} z^{\ell/2-1}(1-z)^{m/2-1}, \quad 0 < z < 1.$$

いま，$z = \widetilde{y}/(1+\widetilde{y})$ に対して，$dz = (1+\widetilde{y})^{-2}d\widetilde{y}$ であるから，$\widetilde{Y}$ の密度関数は

$$f_{\widetilde{Y}}(\widetilde{y}) = \frac{1}{(1+\widetilde{y})^2} f_Z(\widetilde{y}/(1+\widetilde{y})) = \frac{1}{B(\ell/2, m/2)} \frac{\widetilde{y}^{\ell/2-1}}{(1+\widetilde{y})^{(\ell+m)/2}}$$

となる．これから，$Y = \widetilde{Y}m/\ell$ の密度関数が導かれる． □

2.2 基本的な極限定理

統計量の厳密な標本分布の導出は難しいことが多い．しかし，$n \to \infty$ とすれば，標本分布の近似分布を導出できる場合がある．

必要な定義を紹介する．$X, X_n, n = 1, 2, \ldots$ を r.v.'s とし，X の d.f. を F とおき，X_n の d.f. を F_n とおく．

X_n が X に 確率収束する (converge in probability) とは，任意の $\varepsilon > 0$ に対して，

$$\lim_n P(|X_n - X| > \varepsilon) = 0$$

となることである．このとき，

$$X_n \xrightarrow{P} X$$

と書く．$X_n \xrightarrow{P} c$ なら，$x = c$ において連続な関数 $f(x)$ に対して，

$$f(X_n) \xrightarrow{P} f(c)$$

となる．なぜなら，$f(x)$ は $x = c$ で連続であるから，

$$\forall \varepsilon > 0, \exists \delta > 0 \text{ s.t. } |x - c| < \delta \Rightarrow |f(x) - f(c)| < \varepsilon$$

であって，

$$P(|f(X_n) - f(c)| \geq \varepsilon) \leq P(|X_n - c| \geq \delta) \to 0$$

となるからである．例えば，$c \neq 0$ なら，$f(x) = 1/x \ (x \neq 0), f(0) = 0$ とおくと，$f(x)$ は $x = c$ で連続であるから，

$$\frac{1}{X_n} \xrightarrow{P} \frac{1}{c}$$

となる．厳密には，左辺は $X_n = 0$ のとき定義されないが，$P(X_n = 0) \to 0$ より，そのような確率は漸近的には無視できる．

また，X_n が X に分布収束する (converge in distribution) とは，F の任意の連続点 $x \in \mathbb{R}$ に対して，

$$\lim_n F_n(x) = F(x)$$

となることである．F_n が F に分布収束するともいう．このとき，

$$X_n \xrightarrow{d} X \quad \text{or} \quad X_n \xrightarrow{d} F \quad \text{or} \quad F_n \xrightarrow{d} F$$

と書く．

Example 2.2. $X_n = X + 1/n$ なら，$F_n(x) = P(X \le x - 1/n) = F(x - 1/n)$．よって，$n \to \infty$ のとき，$F_n(x) \to F(x-)$ となる．x が F の連続点なら，$F_n(x) \to F(x)$ となるから，$X_n \xrightarrow{d} X$ となる．

Remark 2.1. 分布収束や確率収束はモーメントの収束を含意しない．例えば，X_n を $P(X_n = 0) = 1 - n^{-1}, P(X_n = n) = n^{-1}$ なる r.v. とすると，$P(X_n = 0) \to 1$ だから，$X_n \xrightarrow{P} 0$ であるが，$E[X_n] = 1$ である．

分布収束と確率収束の関係を考察していこう．まず，確率収束していれば，分布収束する：

$$X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X.$$

実際，$X_n \xrightarrow{P} X$ なら，任意の $\varepsilon > 0$ に対して，$P(|X_n - X| > \varepsilon) \to 0$ であって，

$$\begin{aligned}
F_n(x) &= P(X_n \le x) \\
&= P(\{X_n \le x\} \cap \{|X_n - X| \le \varepsilon\}) + P(\{X_n \le x\} \cap \{|X_n - X| > \varepsilon\}) \\
&\le P(X \le x + \varepsilon) + P(|X_n - X| > \varepsilon) \\
&= F(x + \varepsilon) + P(|X_n - X| > \varepsilon).
\end{aligned}$$

同様にして，

$$F_n(x) \ge F(x - \varepsilon) - P(|X_n - X| > \varepsilon)$$

であるから，

$$\limsup_n F_n(x) \le F(x + \varepsilon), \quad \liminf_n F_n(x) \ge F(x - \varepsilon)$$

となる．x が F の連続点なら，$\varepsilon \downarrow 0$ として，$F(x + \varepsilon), F(x - \varepsilon) \to F(x)$ となるから，$\lim_n F_n(x) = F(x)$ を得る．逆は一般に成り立たない．

Example 2.3. $X, Y \sim N(0,1)$ i.i.d. とし，$X_n = Y$ とおくと，$X_n \xrightarrow{d} X$ だが，$P(|X_n - X| > \varepsilon) = P(|Y - X| > \varepsilon)$ であって，右辺は n によらず正であるから，X_n は X に確率収束しない．

ただし，X が定数 $X \equiv c$ なら，逆も成り立つ (演習問題)：

$$X_n \xrightarrow{d} c \Rightarrow X_n \xrightarrow{P} c.$$

Lemma 2.1 (Slutsky の補題)**.** $X_n \xrightarrow{d} X, Y_n \xrightarrow{P} c$ とする．

(a) $X_n + Y_n \xrightarrow{d} X + c$.

(b) $Y_n X_n \xrightarrow{d} cX$.

Proof. (a). $X + c$ の d.f. は $F_c(z) := P(X + c \leq z) = F(z - c)$ である．$z \in \mathbb{R}$ を F_c の連続点とする．このとき，任意の $\varepsilon > 0$ に対して，

$$\begin{aligned} P(X_n + Y_n \leq z) &\leq P(\{X_n + Y_n \leq z\} \cap \{|Y_n - c| \leq \varepsilon\}) + P(|Y_n - c| > \varepsilon) \\ &\leq P(X_n \leq z - c + \varepsilon) + P(|Y_n - c| > \varepsilon). \end{aligned}$$

同様にして，

$$P(X_n + Y_n \leq z) \geq P(X_n \leq z - c - \varepsilon) - P(|Y_n - c| > \varepsilon).$$

よって，

$$\begin{aligned} &\limsup_n P(X_n + Y_n \leq z) \leq F(z - c + \varepsilon) = F_c(z + \varepsilon), \\ &\liminf_n P(X_n + Y_n \leq z) \geq F_c(z - \varepsilon). \end{aligned}$$

z は F_c の連続点だから，$\varepsilon \downarrow 0$ として，$\lim_n P(X_n + Y_n \leq z) = F_c(z)$ を得る．
(b) の証明は演習問題とする． □

定数への分布収束は確率収束と同値だから，次の系を得る (直接証明することも難しくない)．

Corollary 2.1. $X_n \xrightarrow{P} c_1, Y_n \xrightarrow{P} c_2$ なら，$X_n + Y_n \xrightarrow{P} c_1 + c_2, X_n Y_n \xrightarrow{P} c_1 c_2$ となる．

$F_n \xrightarrow{d} F$ であって，F が連続なら，各 $x \in \mathbb{R}$ に対して，$\lim_n F_n(x) = F(x)$ となるが，実は F_n は F に一様収束する．

Theorem 2.4 (Pólya)**.** $F_n \xrightarrow{d} F$ とし，F は連続とする．このとき，$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \to 0$ となる．

Proof. $k = 1, 2, \ldots$ に対して，$x_{k,0} = -\infty, x_{k,j} = F^{\leftarrow}(j/k), j = 1, \ldots, k-1, x_{k,k} = \infty$ とおくと，$F(x_{k,j}) = j/k$ $(\forall j = 0, 1, \ldots, k)$ である．ここで，

$$\Delta_{n,k} = \max_{1 \leq j \leq k-1} |F_n(x_{k,j}) - F(x_{k,j})|$$

とおくと，$\lim_n \Delta_{n,k} = 0$ である．また，$x \in (x_{j-1,k}, x_{j,k})$ に対して，

$$F_n(x) - F(x) \le F_n(x_{j,k}) - F(x_{j,k}) + F(x_{j,k}) - F(x_{k,j-1}) \le \Delta_{n,k} + 1/k$$

である．同様にして，

$$F_n(x) - F(x) \ge -\Delta_{n,k} - 1/k$$

を得る．よって，

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \le \Delta_{n,k} + 1/k$$

であって，$n \to \infty, k \to \infty$ として，定理の結論を得る． □

Remark 2.2. Pólya の定理より，$F_n \xrightarrow{d} F$ であって，F が連続なら，

$$P(X_n < x) = F_n(x-) \to F(x-) = F(x)$$

となる．実際，$x_m \uparrow x, x_m < x$ に対して，$|F_n(x-) - F(x)| = \lim_m |F_n(x_m) - F(x_m)|$ となるから，$|F_n(x-) - F(x)| \le \sup_{y\in\mathbb{R}} |F_n(y) - F(y)| \to 0$ を得る．従って，このとき，

$$P(a \le X_n \le b) = P(X_n \le b) - P(X_n < a) \to F(b) - F(a)$$

となる．

また，分布収束は特性関数の各点収束と同値である．

Theorem 2.5 (連続性定理 (continuity theorem))**.** X の特性関数を φ とし，X_n の特性関数を φ_n とする．このとき，

$$X_n \xrightarrow{d} X \Leftrightarrow \lim_n \varphi_n(t) = \varphi(t)\ \forall t \in \mathbb{R}.$$

連続性定理の証明は 6.1 節で与える．

Example 2.4. $X_n \sim N(\mu_n, \sigma_n^2)$ として，$n \to \infty$ のとき，$X_n \xrightarrow{d} X$ としよう．このとき，μ_n, σ_n^2 はそれぞれ有限な値に収束して，$\mu_n \to \mu, \sigma_n^2 \to \sigma^2$ とすると，$X \sim N(\mu, \sigma^2)$ となることを証明しよう．まず，$\varphi_n(t) = e^{it\mu_n - t^2\sigma_n^2/2}$ であって，連続性定理より $\lim_n \varphi_n(t) = \varphi(t)$ だから，$\lim_n |\varphi_n(t)| = \lim_n e^{-t^2\sigma_n^2/2} = |\varphi(t)|$ である．特に，$t = \sqrt{2}$ を代入すると，$e^{-\sigma_n^2} \to |\varphi(\sqrt{2})|$ だから，$\sigma_n^2 \to -\log|\varphi(\sqrt{2})| =: \sigma^2$ である．この結果から，

$$e^{it\mu_n} \to \varphi(t) e^{t^2\sigma^2/2}$$

を得る．μ_n が Cauchy 列であることを示そう．∞ に発散する任意の増加列 n_k, m_k に対して，$\mu_{n_k} - \mu_{m_k} \to 0$ $(k \to \infty)$ を示せばよい．いま，

$$e^{it(\mu_{n_k} - \mu_{m_k})} \to 1$$

である．U_k を $P(U_k = \mu_{n_k} - \mu_{m_k}) = 1$ なる.r.v.'s とすると，これは $E[e^{itU_k}] \to 1$ とみなせる．よって，連続性定理より，$U_k \xrightarrow{d} 0$ を得る．収束先が定数なので，$U_k \xrightarrow{P} 0$ となるが，U_k の定義からこれは $\mu_{n_k} - \mu_{m_k} \to 0$ を意味する．

以上より，μ_n が収束列であることが示された．$\lim_n \mu_n = \mu$ とすると，

$$\varphi_n(t) = e^{it\mu_n - t^2\sigma_n^2/2} \to e^{it\mu - t^2\sigma^2/2}$$

だから，再び連続性定理より $X \sim N(\mu, \sigma^2)$ を得る．

これ以降，

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

として，$E[X_1^2] < \infty$ を仮定する．また，

$$E[X_1] = \mu, \quad \mathrm{Var}(X_1) = \sigma^2 > 0, \quad \sigma = \sqrt{\sigma^2}$$

とする．

$$\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$$

とおく．

Theorem 2.6 (大数の弱法則 (weak law of large numbers))**.** $\overline{X} \xrightarrow{P} \mu$.

Proof. $E[\overline{X}] = \mu, \mathrm{Var}(\overline{X}) = n^{-2}\sum_{i=1}^n \mathrm{Var}(X_i) = \sigma^2/n$ であるから，Chebyshev の不等式より，

$$P(|\overline{X} - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$$

となる．$n \to \infty$ のとき，右辺 $\to 0$ となるから，$\overline{X} \xrightarrow{P} \mu$ を得る． □

Example 2.5 (Weierstrass の近似定理)**.** やや脱線になるが，大数の弱法則の証明に関連した話題として，Weierstrass の近似定理に対する確率論的な証明を与える．

Theorem 2.7. $f : [0,1] \to \mathbb{R}$ を連続な関数とすると，任意の $\varepsilon > 0$ に対して，次をみたす多項式 $p_\varepsilon(x)$ が存在する：$\sup_{x\in[0,1]} |f(x) - p_\varepsilon(x)| < \varepsilon$.

Proof. $n = 1, 2, \ldots$ に対して，

$$f_n(x) = \sum_{m=0}^n \binom{n}{m} x^m (1-x)^{n-m} f(m/n)$$

とおく．f_n を f の n 次 Bernstein 多項式と呼ぶ．このとき，

$$\sup_{x\in[0,1]} |f_n(x) - f(x)| \to 0, \ n \to \infty$$

を示す．定理の結論はこれから直ちに従う．

$X_1, \ldots, X_n \sim Bin(1,p)$ i.i.d. とすると，$n\overline{X} \sim Bin(n,p)$ だから，$E[f(\overline{X})] = f_n(p)$ である．$M = \sup_{x\in[0,1]} |f(x)|$ とおいて，$\varepsilon > 0$ を任意に固定する．f は $[0,1]$ 上で一様連続であるから，$\exists\delta > 0$ s.t. $|x-y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon$. 一方，Chebyshev の不等式より，

$$P(|\overline{X} - p| \geq \delta) \leq n^{-1}\delta^{-2}p(1-p) \leq \frac{1}{4n\delta^2}$$

である．$Y = f(\overline{X}) - f(p)$ とおくと，

$$\begin{aligned}
|\underbrace{E[f(\overline{X})]}_{=f_n(p)} - f(p)| &= |E[Y]| \leq E[|Y|] \\
&\leq E[|Y|I(|\overline{X} - p| < \delta)] + E[|Y|I(|\overline{X} - p| \geq \delta)] \\
&\leq \varepsilon + 2MP(|\overline{X} - p| \geq \delta) \leq \varepsilon + \frac{M}{2n\delta^2}.
\end{aligned}$$

最右辺は p に依存しないので，$\limsup_n \sup_{p\in[0,1]} |f_n(p) - f(p)| \leq \varepsilon$ を得る． □

Example 2.6 (モンテカルロ法と重点サンプリング法). f を $\mathbb{R}^k$ 上の密度関数とし，$h : \mathbb{R}^k \to \mathbb{R}$ を所与の関数とする．ここで，少なくとも $\int h(x)^2 f(x)dx < \infty$ は仮定しておく．このとき，積分の値

$$J = \int h(x)f(x)dx$$

を計算することを考える．f に従う独立な確率ベクトル $X_1, \ldots, X_n \sim f$ i.i.d. を発生させることができれば，$h(X_1), \ldots, h(X_n)$ は i.i.d. であって，その期待値は $E[h(X_i)] = \int h(x)f(x)dx = J$ だから，大数の法則より，$n \to \infty$ のとき，

$$\frac{1}{n}\sum_{i=1}^{n} h(X_i) \xrightarrow{P} J$$

となる．このように乱数発生を用いて積分を近似する方法を (直接) モンテカルロ法と呼ぶ．

一様分布や標準正規分布などの標準的な分布に従う (疑似) 乱数を発生させるオプションは，標準的な統計解析ソフトウェアに備わっているはずである．しかし，f がよく知られた密度関数でもなく，その関数型が複雑な場合，f から直接乱数を発生させるのは困難である．そのような場合は次の重点サンプリング法 (importance sampling) が有効である．g を $\mathbb{R}^k$ 上の密度関数とし，g からの乱数発生は可能とする．ここで，$\{x : g(x) > 0\} \supset \{x : f(x) > 0\}$ ならば，形式的に，

$$\int h(x)f(x)dx = \int_{\{g>0\}} h(x)f(x)dx = \int_{\{g>0\}} \frac{h(x)f(x)}{g(x)} g(x)dx$$

と表すことができる．いま，

$$\int_{\{g>0\}} \left(\frac{h(x)f(x)}{g(x)}\right)^2 g(x)dx < \infty \tag{*}$$

と仮定すれば，$X_1, \dots, X_n \sim g$ i.i.d. に対して，大数の法則より，

$$\frac{1}{n}\sum_{i=1}^n \frac{h(X_i)f(X_i)}{g(X_i)} \xrightarrow{P} E\left[\frac{h(X_1)f(X_1)}{g(X_1)}\right] = \int_{\{g>0\}} \frac{h(x)f(x)}{g(x)} g(x)dx = J$$

となる．g を <u>重点関数</u> (importance function) と呼ぶ．重点サンプリング法のパフォーマンスは，重点関数の選択に依存する．そもそも，$\{x : g(x) > 0\} \supset \{x : f(x) > 0\}$ である必要があり，さらに，g の形状があまりにも f と異なっている場合，(*) の条件がみたされないかもしれない．従って，g は f と似た形状をもつように選ぶべきであるとされる．

重点サンプリングは f の正規化定数の計算が難しい場合にも適用できる．すなわち，f が $f(x) = Cf_0(x)$ の形で与えられているとする．ここで，f_0 は $0 < \int f_0(x)dx < \infty$ をみたす非負関数であって，$C = 1/\int f_0(x)dx$ は正規化定数である．C の計算が難しい場合でも，$w(x) = \frac{f_0(x)}{g(x)}$ とおくと，$X_1, \dots, X_n \sim g$ i.i.d. に対して，

$$\frac{\sum_{i=1}^n h(X_i)w(X_i)}{\sum_{i=1}^n w(X_i)} = \frac{n^{-1}\sum_{i=1}^n h(X_i)w(X_i)}{n^{-1}\sum_{i=1}^n w(X_i)} \xrightarrow{P} \frac{\int h(x)w(x)g(x)dx}{\int w(x)g(x)dx} = \frac{\int h(x)f_0(x)dx}{\int f_0(x)dx} = J$$

となる．左辺は C に依存しないので，左辺を J の近似値として利用すればよい．

次に，中心極限定理 (central limit theorem, CLT) を証明しよう [14]．その前に，複素指数関数に関する技術的な補題を証明する．

Lemma 2.2. z_n を複素数列とし，$z_n \to z$ とする．このとき，

$$\lim_n \left(1 + \frac{z_n}{n}\right)^n = e^z.$$

Proof. 次の 2 つの評価を使う．

- 絶対値が θ 以下の複素数 $z_1, \dots, z_n, w_1, \dots, w_n$ に対して，

$$\left|\prod_{j=1}^n z_j - \prod_{j=1}^n w_j\right| \le \theta^{n-1}\sum_{j=1}^n |z_j - w_j|.$$

- 絶対値が 1 以下の複素数 z に対して，$|e^z - (1+z)| \le |z|^2$.

[14] CLT の歴史については，Le Cam (1986) が面白い．

最初の評価は，

$$\left|\prod_{j=1}^n z_j - \prod_{j=1}^n w_j\right| \le \left|z_1\prod_{j=2}^n z_j - z_1\prod_{j=2}^n w_j\right| + \left|z_1\prod_{j=2}^n w_j - w_1\prod_{j=2}^n w_j\right|$$
$$\le \theta\left|\prod_{j=2}^n z_j - \prod_{j=2}^n w_j\right| + \theta^{n-1}|z_1 - w_1|$$
$$\vdots$$
$$\le \theta^{n-1}\sum_{j=1}^n |z_j - w_j|$$

から従う．2 番目の評価は，$e^z = 1 + z + z^2\sum_{j=2}^\infty z^{j-2}/j!$ より，$|z| \le 1$ のとき，

$$|e^z - (1+z)| \le |z|^2\sum_{j=2}^\infty \frac{1}{j!} = |z|^2(e-2) \le |z|^2$$

となることから従う．

いま，$\gamma > |z|$ を 1 つ固定すると，十分大きな n に対して，$|z_n| \le \gamma$ である．ここで，$|1 + z_n/n| \le 1 + |z_n|/n \le 1 + \gamma/n \le e^{\gamma/n}, |e^{z/n}| \le e^{|z|/n} \le e^{\gamma/n}$ より，

$$\left|\left(1 + \frac{z_n}{n}\right)^n - (e^{z_n/n})^n\right| \le (e^{\gamma/n})^{n-1}n|1 + z_n/n - e^{z_n/n}| \le e^{(n-1)\gamma/n}|z_n|^2/n \to 0.$$

一方，$e^{z_n} \to e^z$ だから，補題の結論を得る． $\square$

Theorem 2.8 (CLT). $\sqrt{n}(\overline{X} - \mu)/\sigma \xrightarrow{d} N(0,1)$.

Proof. $Z_j = (X_j - \mu)/\sigma$ とおくと，$E[Z_j] = 0, \mathrm{Var}(Z_j) = 1$ であって，$\sqrt{n}(\overline{X} - \mu)/\sigma = \sqrt{n}\overline{Z}$ であるから，はじめから $\mu = 0, \sigma^2 = 1$ と仮定してよい．X_1 の特性関数を φ とおくと，$\sqrt{n}\overline{X} = \sum_{j=1}^n X_j/\sqrt{n}$ の特性関数は

$$\varphi_n(t) = E[e^{it\sum_{j=1}^n X_j/\sqrt{n}}] = \prod_{j=1}^n E[e^{itX_j/\sqrt{n}}] = \{\varphi(t/\sqrt{n})\}^n.$$

ここで，$\varphi(0) = 1, \varphi'(0) = iE[X_1] = 0, \varphi''(0) = i^2E[X_1^2] = -1$ であるから，$\varphi(t)$ は

$$\varphi(t) = 1 - \frac{t^2}{2} + t^2R(t), \quad \lim_{t\to 0} R(t) = 0$$

と展開できる．よって，

$$\varphi(t/\sqrt{n}) = 1 - \frac{t^2}{2n} + \frac{t^2}{n}R(t/\sqrt{n})$$

と展開できて，$\lim_n R(t/\sqrt{n}) = 0$ となるから，

$$\varphi_n(t) = \left(1 - \frac{t^2}{2n} + \frac{t^2}{n} R(t/\sqrt{n})\right)^n \to e^{-t^2/2}$$

となる．$e^{-t^2}/2$ は $N(0,1)$ の特性関数であるから，連続性定理より，

$$\sqrt{n}\overline{X} \xrightarrow{d} N(0,1)$$

を得る． □

Remark 2.3. Slutsky の補題より，

$$\sqrt{n}(\overline{X} - \mu) \xrightarrow{d} N(0, \sigma^2) \tag{**}$$

である．$\sigma = 0$ の場合，$P(\overline{X} = \mu) = 1$ だから，$\sqrt{n}(\overline{X} - \mu)$ は確率 1 で 0 である．一方，$N(0,0)$ は 0 に集中した分布だから，$\sigma = 0$ の場合も含めて，(**) は正しい．

Example 2.7. $Y_n \sim Bin(n,p), 0 < p < 1$ とすると，$X_1, \ldots, X_n \sim Bin(1,p)$ i.i.d. に対して，

$$Y_n \stackrel{d}{=} X_1 + \cdots + X_n$$

と表せるから，

$$\frac{\sqrt{n}(Y_n/n - p)}{\sqrt{p(1-p)}} \xrightarrow{d} N(0,1)$$

となる．

Example 2.8. $X_1, \ldots, X_n \sim F$ i.i.d. に対して，

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \le x),\ x \in \mathbb{R}$$

を <u>経験分布関数</u> (empirical distribution function) と呼ぶ．$\widehat{F}_n(x)$ は x の関数として d.f. であって，確率 $1/n$ で X_i に値をとる分布に対応している．ここで，$x \in \mathbb{R}$ を固定すると，$I(X_i \le x), i = 1, \ldots, n$ は i.i.d. であって，その平均と分散は

$$\begin{aligned}
&E[I(X_1 \le x)] = P(X_1 \le x) = F(x), \\
&\mathrm{Var}(I(X_1 \le x)) = E[I(X_1 \le x)] - (E[I(X_1 \le x)])^2 = F(x) - F(x)^2 = F(x)(1 - F(x))
\end{aligned}$$

である．よって，$n \to \infty$ のとき，大数の弱法則と CLT より，$\widehat{F}_n(x) \xrightarrow{P} F(x)$, $\sqrt{n}(\widehat{F}_n(x) - F(x)) \xrightarrow{d} N(0, F(x)(1 - F(x)))$ が成り立つ．

Remark 2.4. 有限な平均や分散が存在しない場合，CLT は成り立たない．例えば，$X_1,\dots,X_n$ を Cauchy 分布に従う i.i.d. r.v.'s とすれば，X_1 の特性関数は $\varphi(t)=E[e^{itX_1}]=e^{-|t|}$ である．よって，$\overline{X}$ の特性関数は $\varphi_n(t)=E[e^{it\overline{X}}]=(e^{-|t|/n})^n=e^{-|t|}$ となり，$\overline{X}$ も Cauchy 分布に従う．もっと一般に，i.i.d. r.v.'s $X_1,\dots,X_n\sim F$ に対して，ある $a\in\mathbb{R},b>0$ が存在して，$\sqrt{n}(\overline{X}-a)/b\xrightarrow{d}N(0,1)$ が成り立つなら，必ず $E[X_1^2]<\infty$ であって，$a=E[X_1],b^2=\mathrm{Var}(X_1)$ でなくてはならないことが知られている．

さて，追加的に，

$$E[X_1^4]<\infty$$

を仮定して，t 統計量

$$T_n=\frac{\sqrt{n}(\overline{X}-\mu)}{S}$$

の極限分布を求めてみよう．$F=N(\mu,\sigma^2)$ なら $T_n\sim t(n-1)$ であったが，F が正規分布でないなら，$T_n\sim t(n-1)$ ではない．$\mu=0,\sigma^2=1$ と仮定してよい．このとき，大数の弱法則と Slutsky の補題より，

$$S^2=\frac{n}{n-1}\frac{1}{n}\sum_{i=1}^n X_i^2-\frac{n}{n-1}(\overline{X})^2\xrightarrow{P}\sigma^2-0=1$$

となるから，

$$\frac{1}{S}=\frac{1}{\sqrt{S^2}}\xrightarrow{P}1$$

となる．さらに，CLT より，

$$\sqrt{n}\overline{X}\xrightarrow{d}N(0,1)$$

であるから，Slutsky の補題より，

$$T_n\xrightarrow{d}N(0,1)$$

を得る．つまり，$E[X_1^4]<\infty$ なら，F がどうであれ，T_n の分布は $N(0,1)$ で近似できる．

2.3　順序統計量

$X_1,\dots,X_n\sim F$ i.i.d. が与えられたとき，それを小さい順に並べ替えた

$$X_{(1)}\le X_{(2)}\le\cdots\le X_{(n)}$$

を 順序統計量 (order statistics) と呼ぶ．ここで，$x\in\mathbb{R}$ に対して，

$$X_{(n)}\le x\Leftrightarrow X_i\le x\ 1\le\forall i\le n$$

であるから，

$$P(X_{(n)}\le x)=F(x)^n$$

となる．また，

$$X_{(1)} > x \Leftrightarrow X_i > x \ 1 \leq \forall i \leq n$$

であるから，

$$P(X_{(1)} \leq x) = 1 - P(X_{(1)} > x) = 1 - \{1 - F(x)\}^n$$

となる．

もっと一般に，$x \in \mathbb{R}$ を固定して，$Y = \sum_{i=1}^{n} I(X_i \leq x)$ とおくと，$Y \sim Bin(n, F(x))$ であって，

$$X_{(i)} \leq x \Leftrightarrow Y \geq i$$

であるから，

$$P(X_{(i)} \leq x) = P(Y \geq i) = \sum_{k=i}^{n} \binom{n}{k} F(x)^k \{1 - F(x)\}^{n-k}$$

となる．

F が連続で，ある $-\infty \leq a < b \leq \infty$ に対して，$F(b) = 1, F(a) = 0$ であって，(a, b) 上で C^1 級としよう．このとき，$X_{(i)}$ の密度関数を求めてみる．

$$p(k, m) = \binom{m}{k} p^k (1-p)^{m-k}$$

とおくと，

$$\begin{aligned}
&\frac{d}{dp}\binom{n}{k} p^k (1-p)^{n-k} \\
&= \frac{n!}{(k-1)!(n-k)!} p^{k-1}(1-p)^{n-k} - \frac{n!}{k!(n-k-1)!} p^k (1-p)^{n-k-1} \\
&= n\{p(k-1, n-1) - p(k, n-1)\}
\end{aligned}$$

となるから，$p = F(x)$ とおいて，$x \in (a, b)$ に対して，

$$\begin{aligned}
f_{X_{(i)}}(x) &= \frac{d}{dx} P(X_{(i)} \leq x) = n f(x) \sum_{k=i}^{n} \{p(k-1, n-1) - p(k, n-1)\} \\
&= n f(x) p(i-1, n-1) \\
&= \frac{n!}{(i-1)!(n-i)!} f(x) F(x)^{i-1} \{1 - F(x)\}^{n-i}
\end{aligned}$$

となる．

Example 2.9. $F = U(0, 1)$ なら，$0 < x < 1$ に対して $f(x) = 1, F(x) = x$ であるから，

$$X_{(i)} \sim Be(i, n-i+1)$$

となる．従って，$X_{(i)}$ の平均と分散は

$$E[X_{(i)}] = \frac{i}{n+1},\ \mathrm{Var}(X_{(i)}) = \frac{i(n-i+1)}{(n+1)^2(n+2)}$$

となる．

$X_{(1)}, \ldots, X_{(n)}$ の同時密度は

$$f_{X_{(1)},\ldots,X_{(n)}}(x_1, \ldots, x_n) = n! f(x_1) \cdots f(x_n) I(x_1 < \cdots < x_n)$$

となる (演習問題)．よって，$X_{(1)}, \ldots, X_{(n)}$ は独立でない．また，x_n を $f(x_n) > 0$ となる点とすると，$X_{(n)} = x_n$ を与えたときの $X_{(1)}, \ldots, X_{(n-1)}$ の条件付き密度は，

$$\begin{aligned} f_{X_{(1)},\ldots,X_{(n-1)}|X_{(n)}}(x_1, \ldots, x_{n-1} \mid x_n) &= \frac{f_{X_{(1)},\ldots,X_{(n)}}(x_1, \ldots, x_n)}{f_{X_{(n)}}(x_n)} \\ &= (n-1)! \prod_{i=1}^{n-1} \frac{f(x_i)}{F(x_n)} I(x_1 < \cdots < x_{n-1} < x_n) \end{aligned}$$

となる．ここで，

$$x \mapsto \frac{f(x)}{F(x_n)} I(x < x_n)$$

は $\mathbb{R}$ 上の確率密度関数である．そこで，

$$F_{x_n}(x) = \int_{-\infty}^{x} \frac{f(u)}{F(x_n)} I(u < x_n) du = \frac{F(\min\{x, x_n\})}{F(x_n)},\ x \in \mathbb{R}$$

とおくと，$X_{(n)} = x_n$ を与えたときの $X_{(1)}, \ldots, X_{(n-1)}$ の条件付き分布は，F_{x_n} からのサイズ $(n-1)$ の独立標本の順序統計量の同時分布に等しい．

Example 2.10. $F = U(0,1)$ なら，$X_{(n)}$ を与えたときの $X_{(1)}, \ldots, X_{(n-1)}$ の条件付き分布は，$U(0, X_{(n)})$ からのサイズ $(n-1)$ の独立標本の順序統計量の同時分布に等しい．

極値分布

次に，$X_1, \ldots, X_n \sim F$ i.i.d. に対して，最大値 $X_{(n)}$ の極限分布を考察する．

Example 2.11. $F = U(0,1)$ のとき，$0 < u < 1$ に対して $P(X_{(n)} \geq u) = 1 - P(X_{(n)} < u) = 1 - u^n$ だから，$x \in \mathbb{R}$ に対して，

$$P\{n(1 - X_{(n)}) \leq x\} = P\{X_{(n)} \geq 1 - x/n\} = 1 - \left(1 - \frac{x}{n}\right)^n \to 1 - e^{-x}$$

となる．よって，

$$n(1 - X_{(n)}) \xrightarrow{d} Ex(1)$$

となる．

Example 2.12. $F = N(0,1)$ のときは，次のようになる．

Theorem 2.9. $X_1, \ldots, X_n \sim N(0,1)$ i.i.d. とする．このとき，

$$a_n = (2\log n)^{-1/2},\ b_n = (2\log n)^{1/2} - \frac{1}{2}(2\log n)^{-1/2}(\log\log n + \log 4\pi)$$

とおくと，$(X_{(n)} - b_n)/a_n \xrightarrow{d} \Lambda$ となる．ここで，

$$\Lambda(x) = e^{-e^{-x}},\ x \in \mathbb{R}$$

である．Λ は Gumbel 分布 と呼ばれる．

証明は次の補題による．

Lemma 2.3. $X_1, \ldots, X_n \sim F$ i.i.d. とする．与えられた定数 $\tau \geq 0$ と数列 u_n に対して，

$$n(1 - F(u_n)) \to \tau \Leftrightarrow P(X_{(n)} \leq u_n) \to e^{-\tau}.$$

Proof. $\Rightarrow$.

$$P(X_{(n)} \leq u_n) = F^n(u_n) = \{1 - (1 - F(u_n))\}^n = (1 - \tau/n + o(n^{-1}))^n \to e^{-\tau}.$$

$\Leftarrow$.

$$n\log\{1 - (1 - F(u_n))\} = \log P(X_{(n)} \leq u_n) \to -\tau.$$

$1 - F(u_n) \to 0$ だから，左辺 $= -n(1 - F(u_n))(1 + o(1))$ より，$n(1 - F(u_n)) \to \tau$. $\square$

Proof of Theorem 2.12. u_n を $n(1 - \Phi(u_n)) = e^{-x}$ により定義すれば，$P(X_{(n)} \leq u_n) \to \Lambda(x)$ となる．u_n を評価していく．

$$\frac{1 - \Phi(u)}{\phi(u)/u} \to 1,\ u \to \infty$$

より，$n^{-1}e^{-x}u_n/\phi(u_n) \to 1$ となる．両辺の対数をとって整理すると，

$$-\log n - x + \log u_n + \frac{1}{2}\log 2\pi + \frac{u_n^2}{2} \to 0.$$

$u_n \to \infty, u_n^2/(2\log n) \to 1$ だから，$2\log u_n - \log 2 - \log\log n \to 0$, i.e.,

$$\log u_n = \frac{1}{2}(\log 2 + \log\log n) + o(1).$$

この評価を使うと，

$$u_n^2 = 2\log n + 2x - \log 2 - \log\log n - \log 2\pi + o(1).$$

これを書き直すと，

$$u_n^2 = (2\log n)\left\{1 + \frac{x - \frac{1}{2}\log 4\pi - \frac{1}{2}\log\log n}{\log n} + o((\log n)^{-1})\right\}.$$

$\sqrt{x} = 1 + x/2 + O(x^2)$ $(x \to 0)$ なる評価を使うと，

$$\begin{aligned} u_n &= (2\log n)^{1/2}\left\{1 + \frac{x - \frac{1}{2}\log 4\pi - \frac{1}{2}\log\log n}{2\log n} + o((\log n)^{-1})\right\} \\ &= a_n x + b_n + o(a_n). \end{aligned}$$

従って，

$$P((X_{(n)} - b_n)/a_n \le x + o(1)) = P(X_{(n)} \le u_n) \to \Lambda(x)$$

を得る． □

もっと一般に，次のことが知られている．2 つの d.f.'s F, G に対して，ある $\alpha > 0, \beta \in \mathbb{R}$ が存在して，

$$G(x) = F(\alpha x + \beta),\ \forall x \in \mathbb{R}$$

となるとき，F と G は同じタイプをもつという．また，サポートが 1 点集合でない d.f. を非退化な d.f. と呼ぶ．

Theorem 2.10 (Fischer-Tippett-Gnedenko)**.** $X_1, \ldots, X_n \sim F$ i.i.d. に対して，ある数列 $a_n > 0, b_n \in \mathbb{R}$ が存在して，$(X_{(n)} - b_n)/a_n$ が非退化な d.f. G に分布収束するならば，G は次の 3 つの d.f.'s のどれかと同じタイプである：

$$(1)\ \Phi_\alpha(x) = \begin{cases} 0 & \text{if } x \le 0 \\ e^{-x^{-\alpha}} & \text{if } x > 0 \end{cases},\ \alpha > 0.$$

$$(2)\ \Psi_\alpha(x) = \begin{cases} e^{-(-x)^\alpha} & \text{if } x < 0 \\ 1 & \text{if } x \ge 0 \end{cases},\ \alpha > 0.$$

$$(3)\ \Lambda(x) = e^{-e^{-x}}, x \in \mathbb{R}.$$

(1)–(3) の分布はまとめて 極値分布 (extreme value distribution) と呼ばれる．個別には，Φ_α は Fréchet 分布，Ψ_α は Weibull 分布，Λ は Gumbel 分布と呼ばれる．

この定理は $X_{(n)}$ の適当に正規化したあとでの極限分布は 3 種類しかないことを示している．Fisher-Tippett-Gnedenko の定理の証明は Resnick (1998) を参照せよ．極値分布は稀にしか起こらない事象の統計解析において現れる (Coles, 2001)．その他に，順序統計量や極値理論に関する発展的な文献として，Reiss (1989), Resnick (1987), Leadbetter et al. (1983) をあげておく．

3　点推定

$\mathcal{X}$ を有限次元ユークリッド空間とし[15]，$\varnothing \neq \Theta \subset \mathbb{R}^k$ として，各 $\theta \in \Theta$ に対して p_θ を $\mathcal{X}$ 上の確率 (密度) 関数とする．このとき，$\{p_\theta : \theta \in \Theta\}$ は分布の族に対応している．θ をパラメータ (parameter)，Θ をパラメータ空間 (parameter space) と呼び，$\{p_\theta : \theta \in \Theta\}$ をパラメトリックな分布族とかパラメトリックモデルと呼ぶ．Θ として関数空間の部分集合を考える場合があり，そのような場合は Θ によって添え字付けられた分布族をノンパラメトリックモデルと呼ぶ．講義ノートでは基本的にはパラメトリックモデルを考察する．

Example 3.1 (Bernoulli 試行)**.** $\mathcal{X} = \mathbb{R}, \Theta = (0,1)$ とし，$\theta \in \Theta$ に対して，$p_\theta(x)$ を

$$p_\theta(x) = \begin{cases} \theta & x = 1 \\ 1-\theta & x = 0 \\ 0 & x \notin \{0,1\} \end{cases}$$

とすれば，$\{p_\theta : \theta \in \Theta\}$ は，分布の族 $\{Bin(1,\theta) : \theta \in (0,1)\}$ に対応している．

Example 3.2 (正規分布)**.** $\mathcal{X} = \mathbb{R}, \Theta = \{(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$ として，各 $\mu \in \mathbb{R}, \sigma^2 > 0$ に対して，$p_{(\mu,\sigma^2)}$ を $N(\mu,\sigma^2)$ の密度関数とすれば，$\{p_\theta : \theta \in \Theta\}$ は分布の族 $\{N(\mu,\sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$ に対応している．

いま，ある $\theta \in \Theta$ に対して，p_θ に従う i.i.d. 確率ベクトルたち

$$X_1, \ldots, X_n \sim p_\theta \text{ i.i.d.} \tag{*}$$

が得られているとする．(*) の意味は，$X_1, \ldots, X_n$ は独立であって，各 X_i は p_θ を確率 (密度) 関数にもつ分布に従うということである．このとき，$X = (X_1', \ldots, X_n')'$ にもとづいて，パラメータ θ に関する何らかの決定を行うとする．この決定の取り得る値を含む集合を D とおく．D を決定空間 (decision space) と呼ぶ．パラメータが θ のとき，$d \in D$ という決定をとることから生じる損失を

$$L(\theta, d) \geq 0$$

とし，$\Theta \times D$ から $\mathbb{R}_+$ への関数

$$L : \Theta \times D \to \mathbb{R}_+$$

を損失関数 (loss function) と呼ぶ．さらに，$\mathcal{X}^n$ から D への関数 $\delta : \mathcal{X}^n \to D$ を決定関数 (decision function) と呼び，$L(\theta, \delta(X))$ を X について期待値をとった

$$R(\theta, \delta) = E_\theta[L(\theta, \delta(X))]$$

[15] 講義ノートに現れるほとんどの例では $\mathcal{X} = \mathbb{R}$ である．

を リスク関数 (risk function) と呼ぶ．ただし，$E_\theta[\cdot]$ とは，(*) に対して期待値をとることを意味する．$P_\theta, \mathrm{Var}_\theta, \mathrm{Cov}_\theta$ なども同様に定義する．ここで，重要な注意として，決定関数は θ には依存してはいけない．

点推定では θ の関数 $g(\theta) \in \mathbb{R}$ の値を X にもとづいて "あてる" (guess) ことを考える．$g(\theta)$ は多次元でもよいが，以下では 1 次元の場合を考える．このとき，$D = \mathbb{R}$ としておけばよくて，決定関数 $\delta : \mathcal{X}^n \to \mathbb{R}$ のことを $g(\theta)$ の 推定量 (estimator) と呼ぶ．多くの場合，$\delta(X)$ のことも推定量と呼ぶ．$\delta(X)$ の実現値 $\delta(x)$ を 推定値 (estimate) と呼ぶ．損失関数の選択は任意性があるが，2 乗損失関数 (quadratic loss function)

$$L(\theta, d) = (d - g(\theta))^2$$

は代表的な損失関数である．もっと一般に，$0 < q < \infty$ に対して，ℓ^q 損失関数

$$L(\theta, d) = |d - g(\theta)|^q$$

というのもある．

点推定の目標は，よりリスクの小さい推定量を構成することである．しかし，あらゆる推定量のなかでリスクを一様に最小にする推定量は一般に存在しない．

Example 3.3. $0 < \theta_1 < \theta_2 < 1$ とし，$\theta \in \{\theta_1, \theta_2\}$ に対して，$X \sim Bin(1, \theta)$ とする．このとき，$L(\theta, d) = (d - \theta)^2$ に対して，

$$R(\theta, \delta^*) \le R(\theta, \delta), \quad \forall \delta : \theta \text{ の推定量} \tag{**}$$

をみたす θ の推定量 $\delta^*(X)$ は存在 しない．仮に (**) をみたす推定量 $\delta^*(X)$ が存在したとする．このとき，任意に固定した $\theta_0 \in \{\theta_1, \theta_2\}$ に対して，$\delta(X) = \theta_0$ をとると，(**) より，

$$R(\theta_0, \delta^*) \le R(\theta_0, \delta) = 0$$

となる．θ_0 は任意だったから，$R(\theta, \delta^*) = 0 \ \forall \theta \in \{\theta_1, \theta_2\}$ となるが，これはありえない．

ところで，これ以降の議論において，有限標本における性質 (固定した n に対して成り立つ性質) を考察するときは，"$X_1, \ldots, X_n$ が i.i.d." という仮定は本質的ではなくて，X が何らかのパラメトリックモデルに従っている，という仮定が本質的である．例えば，$Y_1, \ldots, Y_m$ が独立な r.v.'s であって，$Y_i \sim N(\alpha + \beta z_i, \sigma^2)$ $(\alpha, \beta \in \mathbb{R}, \sigma^2 > 0)$ という回帰モデルを考える．ここで，$z_1, \ldots, z_m$ は確定的とする．このとき，$Y_1, \ldots, Y_m$ は同一分布に従ってはいないが，$z = (z_1, \ldots, z_m)', 1_m = (1, \ldots, 1)' \in \mathbb{R}^m$ とおくと，$Y = (Y_1, \ldots, Y_m)' \sim N(\alpha 1_m + \beta z, \sigma^2 I_m)$ だから，形式的に $\mathcal{X} = \mathbb{R}^m, X_1 = Y, n = 1$ とすれば，これ以降の議論を適用できる．2 標本問題を扱うときも同様に考える [16]．もちろん，漸近理論にもとづく結果は，i.i.d. という仮定に本質的に依存している．

[16] とはいえ講義ノートでは回帰モデルや 2 標本問題は扱わない．

3.1 十分統計量

パラメータ空間を $\varnothing \neq \Theta \subset \mathbb{R}^k$ とし，$X_1, \dots, X_n \sim p_\theta$ i.i.d. とする．このとき，

$$p_\theta^n(x) = p_\theta(x_1) \cdots p_\theta(x_n), \quad x = (x_1', \dots, x_n')' \in \mathcal{X}^n$$

とおくと，$X = (X_1', \dots, X_n')' \sim p_\theta^n$ である．パラメータ θ に対する統計的推測は X の統計量にもとづくが，X の統計量は無数にある．しかし，多くの場合，十分統計量と呼ばれる統計量の関数だけ考えればよい．

十分統計量の定義を与える前に，統計量を与えたときの X の条件付き期待値を定義する必要がある．$\mathbb{R}^m$ の 長方形 (rectangle) とは，

$$\prod_{j=1}^m (a_j, b_j], \quad -\infty \leq a_j \leq b_j \leq \infty, 1 \leq j \leq m$$

という形の集合のことをいう．ただし，$b = \infty$ のとき，$(a, b] = (a, \infty)$ と理解する．X を確率ベクトルとし，$T = T(X) = (T_1(X), \dots, T_m(X))'$ を X の統計量として，$E[|g(X)|] < \infty$ をみたす関数 g に対して，ある関数 $\eta : \mathbb{R}^m \to \mathbb{R}$ が存在して，

$$E[g(X)I(T \in A)] = E[\eta(T)I(T \in A)], \quad \forall A \subset \mathbb{R}^m : \text{長方形} \tag{*}$$

が成り立つとする．このとき，$\eta(t)$ を $T = t$ を与えたときの $g(X)$ の 条件付き期待値 と呼び，$E[g(X) \mid T = t]$ と書く．X が離散なら，X, T の同時確率関数を $p(x, t)$ とおくと，

$$E[g(X)I(T \in A)] = \sum_{t \in A} \sum_x g(x)p(x, t) = \sum_{t \in A} \left\{ \sum_x g(x) p_{X|T}(x \mid t) \right\} p_T(t)$$

であるから，

$$E[g(X) \mid T = t] = \sum_x g(x) p_{X|T}(x \mid t), \quad p_T(t) > 0$$

である ($p_T(t) = 0$ なる t に対して $E[g(X) \mid T = t]$ の値は任意)．これは以前の条件付き期待値の定義と整合的である．しかし，X が連続のとき，(X, T) は密度関数をもたないので，一般化された条件付き期待値が必要になる．条件付き期待値 $E[g(X) \mid T = t]$ は必ず存在し，次の意味で一意であることが知られている：$\widetilde{\eta}$ も (*) をみたすなら，$P(\eta(T) = \widetilde{\eta}(T)) = 1$ となる．注意として，条件付き期待値 $E[g(X) \mid T = t]$ は特定の t に対してではなく，t の関数として一意に決まる．

また，$E[g(X) \mid T = t]$ に $t = T$ を代入したものを，$E[g(X) \mid T]$ と書く：

$$E[g(X) \mid T] = E[g(X) \mid T = t]|_{t=T}.$$

このとき，定義より，

$$E[E[g(X) \mid T]] = E[g(X)]$$

が成り立つ．

十分統計量の定義を与える．$X \sim p_\theta^n$ とし，$T = T(X)$ を (ベクトル値の) 統計量とする．$E_\theta[|g(X)|] < \infty$ をみたす関数 $g : \mathcal{X}^n \to \mathbb{R}$ に対して，$T = t$ を与えたときの $g(X)$ の条件付き期待値を $E_\theta[g(X) \mid T = t]$ と書く．

Definition 1 (十分統計量)**.** T が θ に対する 十分統計量 (sufficient statistic) であるとは，$E_\theta[|g(X)|] < \infty\ \forall\theta \in \Theta$ をみたす任意の関数 $g : \mathcal{X}^n \to \mathbb{R}$ に対して，条件付き期待値 $E_\theta[g(X) \mid T = t]$ を t の関数として θ に依存しないように選べることをいう．

T が十分統計量のとき，

$$E_\theta[g(X) \mid T = t] = E[g(X) \mid T = t]$$

と書くことにする．

Remark 3.1. 十分統計量とはパラメータの特定の値に対して定義されるのではなく，分布の族 $\mathcal{P} = \{p_\theta^n : \theta \in \Theta\}$ に対して定義される．正確には，T は $\mathcal{P}$ に対する十分統計量と呼ぶべきであるが，慣例として，θ に対する十分統計量と呼んでいる．

Example 3.4. $\Theta = (0, 1)$ とし，$\theta \in \Theta$ に対して，$X_1, \ldots, X_n \sim Bin(1, \theta)$ i.i.d. とする．このとき，

$$p_\theta^n(x) = \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}, \quad x = (x_1, \ldots, x_n)' \in \{0, 1\}^n$$

である．$T = \sum_{i=1}^n X_i$ が θ に対する十分統計量であることを示そう．$X = (X_1, \ldots, X_n)'$ と T の同時確率関数を $q_\theta(x, t)$ とおくと，$\sum_{i=1}^n x_i = t$ なる (x, t) に対して，

$$q_\theta(x, t) = \theta^t(1-\theta)^{n-t}$$

である．一方，$T \sim Bin(n, \theta)$ であるから，T の確率関数は

$$q_\theta^T(t) = \binom{n}{t}\theta^t(1-\theta)^{n-t}$$

である．よって，

$$q_\theta^{X|T}(x \mid t) = \frac{q_\theta(x, t)}{q_\theta^T(t)} = \frac{1}{\binom{n}{t}}, \quad t = \sum_{i=1}^n x_i$$

である．$q_\theta^{X|T}(x \mid t)$ は θ に依存しないから，T は θ に対する十分統計量である．

Example 3.5. $\Theta = \mathbb{R}$ とし，$\theta \in \Theta$ に対して，$X_1, \ldots, X_n \sim N(\theta, 1)$ i.i.d. とする．このとき，

$$p_\theta^n(x) = \frac{1}{(2\pi)^{n/2}}e^{-\frac{1}{2}\sum_{i=1}^n(x_i-\theta)^2}, \quad x = (x_1, \ldots, x_n)' \in \mathbb{R}^n$$

である．ここで，$\overline{x} = n^{-1}\sum_{i=1}^n x_i$ とおくと，

$$\sum_{i=1}^n (x_i - \theta)^2 = \sum_{i=1}^n \{x_i - \overline{x} + (\overline{x} - \theta)\}^2 = \sum_{i=1}^n (x_i - \overline{x})^2 + n(\overline{x} - \theta)^2.$$

G を Helmert 変換とし，$x \in \mathbb{R}^n$ に対して $y = Gx$ とおくと，

$$\sqrt{n}\overline{x} = y_1, \quad \sum_{i=1}^n (x_i - \overline{x})^2 = \sum_{i=2}^n y_i^2$$

である．よって，$T = \sqrt{n}\overline{X}$ とおくと，$E_\theta[|g(X)|] < \infty$ をみたす関数 $g : \mathbb{R}^n \to \mathbb{R}$ と区間 $A \subset \mathbb{R}$ に対して，

$$\begin{aligned} E_\theta[g(X)I(T \in A)] &= \int g(x)I(\sqrt{n}\overline{x} \in A)p_\theta^n(x)dx \\ &= \int_A \left\{ \int \cdots \int g(G^{-1}y)\frac{1}{(2\pi)^{(n-1)/2}}e^{-\frac{1}{2}\sum_{i=2}^n y_i^2}dy_2\cdots dy_n \right\} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(y_1 - \sqrt{n}\theta)^2}dy_1 \end{aligned}$$

となる．従って，

$$E_\theta[g(X) \mid T = y_1] = \int \cdots \int g(G^{-1}y)\frac{1}{(2\pi)^{(n-1)/2}}e^{-\frac{1}{2}\sum_{i=2}^n y_i^2}dy_2\cdots dy_n$$

であって，右辺は θ に依存しないので，T は θ に対する十分統計量である．

十分統計量を見つけるには，次の因子分解定理 (factorization theorem) が便利である．

Theorem 3.1 (因子分解定理)**.** 統計量 T が θ に対する十分統計量であるためには，各 $\theta \in \Theta$ に対して，p_θ^n が

$$p_\theta^n(x) = g_\theta(T(x))h(x), \quad x \in \mathcal{X}^n \tag{**}$$

の形に分解できることが必要十分である．

因子分解定理の最初のバージョンは J. Neyman の 1935 年の論文によって与えられた．測度論にもとづく，一般的な場合の因子分解定理の証明は Halmos and Savage (1949) と Bahadur (1954) によって与えられた．

Proof. X が離散の場合に定理を証明する．$q_\theta(x,t)$ を X, T の同時確率関数とし，$q_\theta^T(t)$ を T の確率関数とする．また，$q_\theta^{X|T}(x \mid t)$ を T を与えたときの X の条件付き確率関数とする．

必要性．T が十分統計量なら，$q_\theta^{X|T}(x \mid t)$ を θ に依存しないように選べる．そこで，$q_\theta^{X|T}(x \mid t) = q^{X|T}(x \mid t)$ と書くと，

$$q_\theta(x,t) = q^{X|T}(x \mid t)q_\theta^T(t)$$

となる．$h(x) = q^{X|T}(x \mid T(x))$ とおくと，$t = T(x)$ なる (x,t) に対して，

$$p_\theta^n(x) = q_\theta(x,t) = q_\theta^T(T(x))h(x)$$

となる．よって，$g_\theta(t) = q_\theta^T(t)$ とすればよい．

十分性．p_θ^n が (**) の形に分解できているとすると，

$$q_\theta^T(t) = \sum_{x:T(x)=t} q_\theta(x,t) = \sum_{x:T(x)=t} p_\theta^n(x) = g_\theta(t) \sum_{x:T(x)=t} h(x).$$

$T(x) = t, \sum_{z:T(z)=t} h(z) > 0$ なる (x,t) に対して，

$$q_\theta(x,t) = p_\theta^n(x) = g_\theta(t)h(x) = \frac{h(x)}{\sum_{z:h(z)=t} h(z)} q_\theta^T(t)$$

となる．よって，

$$q_\theta^{X|T}(x \mid t) = \frac{h(x)}{\sum_{z:h(z)=t} h(z)}, \quad T(x) = t, \sum_{z:T(z)=t} h(z) > 0$$

と選べるから T は十分統計量である． □

Remark 3.2. 因子分解定理から明らかなように，T が十分統計量なら，T の 1 対 1 変換も十分統計量である．

Example 3.6. $\Theta = \{(\mu,\sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$ とし，$(\mu,\sigma^2) \in \Theta$ に対して，$X_1,\ldots,X_n \sim N(\mu,\sigma^2)$ i.i.d. とする $(n \geq 2)$．このとき，

$$p_\theta^n(x) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \overline{x})^2 - \frac{n}{2\sigma^2}(\overline{x} - \mu)^2\right\}$$

であるから，$T(X) = (\overline{X}, S^2)$ は θ に対する十分統計量である．

Example 3.7. $\theta \in \Theta = (0,\infty)$ に対して，$X_1,\ldots,X_n \sim U(0,\theta)$ とする．このとき，

$$p_\theta^n(x) = \frac{1}{\theta^n} I(X_{(1)} > 0) I(X_{(n)} < \theta)$$

である．よって，$X_{(n)}$ は θ に対する十分統計量である．

十分統計量が存在するとき，推定量として十分統計量の関数だけ考えても一般性を失わない．これは Rao-Blackwell の定理の帰結である．Rao-Blackwell 定理の証明の前に，Jensen の不等式を証明しよう．区間 $I \subset \mathbb{R}$ に対して，関数 $\varphi : I \to \mathbb{R}$ が <u>凸関数</u> であるとは，

$$x, y \in I, \lambda \in [0,1] \Rightarrow \varphi(\lambda x + (1-\lambda)y)) \leq \lambda\varphi(x) + (1-\lambda)\varphi(y)$$

をみたすことをいう．I が開区間で，φ が 2 回微分可能なら，φ が凸関数であるためには，$\varphi'' \geq 0$ となることが必要十分である．

Lemma 3.1 (Jensen の不等式). $I \subset \mathbb{R}$ を開区間とし，$\varphi : I \to \mathbb{R}$ を凸関数とする．また，Y を I に値をとる r.v. とし，$E[|Y|] < \infty$ を仮定する．このとき，$E[\varphi(Y)]$ が定義できて，$\varphi(E[Y]) \leq E[\varphi(Y)]$ となる．

Proof. φ が 2 回微分可能と仮定する．$c = E[Y]$ とおくと，Taylor の定理より，各 $y \in I$ に対して，ある $\lambda = \lambda_y \in [0,1]$ が存在して，

$$\varphi(y) = \varphi(c) + \varphi'(c) + \frac{1}{2}\varphi''(\lambda y + (1-\lambda)c)(y-c)^2$$

と展開できる．ここで，$\varphi'' \geq 0$ より，

$$\varphi(y) \geq \varphi(c) + \varphi'(c)(y-c)$$

であるから，$E[\varphi^-(Y)] < \infty$ である．よって，$E[\varphi(Y)]$ は定義できる．さらに，

$$E[\varphi(Y)] \geq \varphi(c) + \varphi'(c)(E[Y]-c) = \varphi(E[Y])$$

を得る． □

Theorem 3.2 (Rao-Blackwell). $g : \Theta \to \mathbb{R}$ を所与とし，損失関数 $L(\theta, d)$ は d の凸関数とする $(D = \mathbb{R})$．また，$\delta(X)$ を $g(\theta)$ の推定量とし，$\forall \theta \in \Theta$ に対して，$E_\theta[|\delta(X)|] < \infty$ とする．さらに，$T = T(X)$ を θ に対する十分統計量とする．このとき，

$$\delta^*(T) = E[\delta(X) \mid T]$$

とおくと，$R(\theta, \delta^*) \leq R(\theta, \delta)\ \forall \theta \in \Theta$ となる．

Remark 3.3. T は十分統計量だから，$E[\delta(X) \mid T]$ は θ に依存しない．

Proof. $\theta \in \Theta$ を任意に固定する．$R(\theta, \delta) = \infty$ なら何も示すことはない．$R(\theta, \delta) < \infty$ なら，Jensen の不等式より，

$$E[L(\theta, \delta(X)) \mid T] \geq L(\theta, E[\delta(X) \mid T]) = L(\theta, \delta^*(T)).$$

よって，

$$R(\theta, \delta) = E_\theta[E[L(\theta, \delta(X)) \mid T]] \geq E_\theta[L(\theta, \delta^*(T))] = R(\theta, \delta^*)$$

を得る． □

厳密にいうと，Rao-Blackwell の定理の証明において，条件付き期待値に対して Jensen の不等式を適用している．X が離散の場合は問題ないが，X が連続のときは測度論の議論が必要になるので，ここでは詳細は省略する．

3.2　不偏推定

あらゆる推定量のなかでリスクを一様に最小化する推定量は一般に存在しないことを注意したが，候補となる推定量のクラスを制限すればそのクラスのなかでリスクを一様に最小化する推定量が存在する場合がある．そのような制約として，不偏性を考察する．ここでは，$g(\theta) \in \mathbb{R}$ の推定を考えて，2 乗損失関数 $L(\theta, d) = (d - g(\theta))^2$ を採用する．

Definition 2 (不偏推定量)**.** $g(\theta)$ の推定量 $\delta(X)$ が <u>不偏</u> (unbiased) であるとは，$E_\theta[\delta(X)] = g(\theta)\ \forall\theta \in \Theta$ となることをいう．

Example 3.8 (不偏推定量が存在しない場合)**.** もちろん不偏推定量が存在しない場合もある．$\theta \in (0, 1)$ に対して，$X \sim Bin(n, \theta)$ とし，$g(\theta)$ の推定を考える．このとき，$\delta(X)$ が $g(\theta)$ の不偏推定量であるためには，

$$\sum_{x=0}^{n} \delta(x) \binom{n}{x} \theta^x (1-\theta)^{n-x} = g(\theta),\ \forall\theta \in (0,1)$$

をみたすことが必要十分であるが，左辺は θ の n 次多項式なので，$g(\theta)$ の不偏推定量が存在するためには，$g(\theta)$ は θ の n 次以下の多項式でなくてはならない．従って，例えば，$\sqrt{\theta}$ の不偏推定量は存在しない．

$\delta(X)$ が $g(\theta)$ の不偏推定量なら，リスクは

$$R(\theta, \delta) = E_\theta[(\delta(X) - g(\theta))^2] = \mathrm{Var}_\theta(\delta(X))$$

となる．

Definition 3 (UMVU)**.** $g(\theta)$ の推定量 $\delta^*(X)$ が <u>一様最小分散不偏</u> (uniformly minimum variance unbiased, UMVU) 推定量であるとは，$E_\theta[\delta^*(T)^2] < \infty\ \forall\theta \in \Theta$ をみたし，$\delta^*(X)$ が $g(\theta)$ の不偏推定量であって，$g(\theta)$ のあらゆる不偏推定量 $\delta(X)$ に対して $\mathrm{Var}_\theta(\delta^*(X)) \leq \mathrm{Var}_\theta(\delta(X))\ \forall\theta \in \Theta$ となることをいう．

UMVU 推定量を求めるには，完備十分統計量の関数で不偏な推定量をとってくればよい．

Definition 4 (完備十分統計量)**.** θ に対する (m 次元の) 十分統計量 T が <u>完備</u> (complete) であるとは，$E_\theta[\varphi(T)] = 0\ \forall\theta \in \Theta$ をみたす任意の関数 $\varphi : \mathbb{R}^m \to \mathbb{R}$ に対して，$P_\theta(\varphi(T) = 0) = 1\ \forall\theta \in \Theta$ となることをいう．

Theorem 3.3 (Lehmann and Scheffé (1950))**.** T を θ に対する完備十分統計量とし，$\delta^*(T)$ を T の関数であって，$E_\theta[\delta^*(T)^2] < \infty\ \forall\theta \in \Theta$ をみたし，かつ $g(\theta)$ の不偏推定量とする．このとき，次の (a) と (b) が成り立つ．

(a) $\delta^*(T)$ は UMVU である．

(b) $g(\theta)$ の任意の UMVU 推定量 $\delta(X)$ に対して，$P_\theta(\delta(X) = \delta^*(T)) = 1\ \forall\theta \in \Theta$ となる．

(b) の意味で，(a) の $\delta^*(T)$ を 一意なUMVU 推定量と呼ぶ．

Proof. (a). $\delta(X)$ を $g(\theta)$ の不偏推定量とする．このとき，$\mathrm{Var}_\theta(\delta^*(T)) \le \mathrm{Var}_\theta(\delta(X))\ \forall\theta \in \Theta$ を示せばよい．$\eta(T) = E[\delta(X) \mid T]$ とおくと，Rao-Blackwell の定理より，$\mathrm{Var}_\theta(\eta(T)) \le \mathrm{Var}_\theta(\delta(X))$ である．ここで，$\forall\theta \in \Theta$ に対して，

$$E_\theta[\eta(T) - \delta^*(T)] = E_\theta[\eta(T)] - E_\theta[\delta^*(T)] = g(\theta) - g(\theta) = 0$$

であるから，T の完備性より，$P_\theta(\eta(T) = \delta^*(T)) = 1\ \forall\theta \in \Theta$ である．以上より，

$$\mathrm{Var}_\theta(\delta^*(T)) = \mathrm{Var}_\theta(\eta(T)) \le \mathrm{Var}_\theta(\delta(X))$$

を得る．

(b). (a) の証明において，$\delta(X)$ を $g(\theta)$ の UMVU 推定量とする．$P_\theta(\eta(T) = \delta^*(T)) = 1$ であったから，$P_\theta(\delta(X) = \eta(T)) = 1$ を示せばよい．ここで，

$$\begin{aligned}&\{\delta(X) - g(\theta)\}^2 = \{\delta(X) - \eta(T) + \eta(T) - g(\theta)\}^2\\&= \{\eta(T) - g(\theta)\}^2 + 2\{\eta(T) - g(\theta)\}\{\delta(X) - \eta(T)\} + \{\delta(X) - \eta(T)\}^2.\end{aligned}$$

両辺の期待値をとると，$\eta(T) = E[\delta(X) \mid T]$ だったから，条件付き期待値の性質より [17]，

$$E_\theta[\{\eta(T) - g(\theta)\}\{\delta(X) - \eta(T)\}] = E_\theta[\{\eta(T) - g(\theta)\}E[\{\delta(X) - \eta(T)\} \mid T]] = 0$$

となる．よって，

$$\mathrm{Var}_\theta(\delta(X)) = \mathrm{Var}_\theta(\eta(T)) + E_\theta[\{\delta(X) - \eta(T)\}^2].$$

$\delta(X)$ は UMVU だったから，$\mathrm{Var}_\theta(\delta(X)) = \mathrm{Var}_\theta(\delta^*(T)) = \mathrm{Var}_\theta(\eta(T))$ であり，従って，$E_\theta[\{\delta(X) - \eta(T)\}^2] = 0$ を得る．以上より，$P_\theta(\delta(X) = \eta(T)) = 1$ が示された． □

Example 3.9. $\theta \in \Theta = (0, 1)$ に対して，$X_1, \ldots, X_n \sim Bin(1, \theta)$ i.i.d. とする．このとき，$T = \sum_{i=1}^n X_i \sim Bin(n, \theta)$ は十分統計量であった．T の完備性を示す．$\varphi : \mathbb{R} \to \mathbb{R}$ を

$$0 = E_\theta[\varphi(T)] = \sum_{t=0}^n \varphi(t)\binom{n}{t}\theta^t(1-\theta)^{n-t}, \quad \forall\theta \in (0, 1)$$

をみたす関数とする．両辺を $(1-\theta)^n$ で割って，$\psi(t) = \varphi(t)\binom{n}{t}, r = \theta/(1-\theta)$ とおくと，

$$\sum_{t=0}^n \psi(t)r^t = 0,\ \forall r > 0$$

となる．これから，$\psi(0) = \cdots = \psi(n) = 0$, i.e., $\varphi(0) = \cdots = \varphi(n) = 0$ を得る．

さらに，$\overline{X} = T/n$ は θ の不偏推定量であるから，$\overline{X}$ は θ の一意な UMVU 推定量である．

[17] X が連続のときは測度論の議論が必要になる．

Example 3.10. $\theta \in \Theta = \mathbb{R}$ に対して，$X_1, \ldots, X_n \sim N(\theta, 1)$ とする．このとき，$T = \sqrt{n}\overline{X} \sim N(\sqrt{n}\theta, 1)$ は十分統計量であった．T の完備性を示す．$\varphi : \mathbb{R} \to \mathbb{R}$ を

$$0 = E_\theta[\varphi(T)] = \frac{1}{\sqrt{2\pi}} \int \varphi(t) e^{-(t-\sqrt{n}\theta)^2/2} dt, \quad \forall \theta \in \mathbb{R} \tag{*}$$

をみたす関数とする．φ は連続と仮定しておこう．

$$(*) \Leftrightarrow \int \{\varphi(t) e^{-t^2/2}\} e^{\theta t} dt = 0, \quad \forall \theta \in \mathbb{R}.$$

さらに，$\psi(t) = \varphi(t) e^{-t^2/2}$ とおくと，

$$\int \psi^+(t) e^{\theta t} dt = \int \psi^-(t) e^{\theta t} dt, \quad \forall \theta \in \mathbb{R} \tag{**}$$

となる．$\theta = 0$ として，

$$c := \int \psi^+(t) dt = \int \psi^-(t) dt \geq 0.$$

$c = 0$ なら，$\psi^+(t) = \psi^-(t) = 0 \; \forall t \in \mathbb{R}$ となるから，$\varphi \equiv 0$ を得る．$c > 0$ なら，$\psi^+/c, \psi^-/c$ は確率密度関数であって，(**) はそれらのモーメント母関数が一致していることを意味する．モーメント母関数と分布は 1 対 1 に対応しているから，

$$\int_{-\infty}^t \psi^+(s) ds = \int_{-\infty}^t \psi^-(s) ds, \; \forall t \in \mathbb{R}$$

を得る．両辺を t について微分して，$\psi^+(t) = \psi^-(t) \; \forall t \in \mathbb{R}$, i.e., $\varphi \equiv 0$ を得る．φ が連続でなくても，"ほとんどすべての"$t \in \mathbb{R}$ に対して $\varphi(t) = 0$ となることが示せる (Lebesgue の微分定理による)．これから，T の完備性が従う．

さらに，$\overline{X} = n^{-1/2} T$ は θ の不偏推定量であるから，一意な UMVU 推定量である．

Remark 3.4. 十分統計量は余計な統計量を加えても十分統計量のままだが (これは因子分解定理より明らか)，完備性は崩れる．上の正規分布の例だと，$(\overline{X}, X_1 - \overline{X})$ は十分統計量だが，完備でない．完備でないことは，$E_\theta[X_1 - \overline{X}] = 0 \; \forall \theta \in \mathbb{R}$ となることからわかる．

もっと一般に指数型分布族に対して，完備十分統計量の存在が示される．p_θ が

$$p_\theta(u) = h(u) \exp\left\{ \sum_{j=1}^m \xi_j(\theta) S_j(u) - c(\theta) \right\}, \quad u \in \mathcal{X}$$

の形に表せるとき，$\{p_\theta : \theta \in \Theta\}$ は m パラメータの 指数型分布族 (exponential family) をなすという．2 項分布，Poisson 分布，正規分布，ガンマ分布，ベータ分布などは指数型分布族をなす．なお，$e^{-c(\theta)}$ は $\int p_\theta(u) du = 1$ をみたすための正規化定数であって，

$$e^{c(\theta)} = \int h(u) e^{\sum_{j=1}^m \xi_j(\theta) S_j(u)} du$$

である．また，指数部分は正であるから，

$$\{u \in \mathcal{X} : p_\theta(u) > 0\} = \{u \in \mathcal{X} : h(u) > 0\}$$

であって，右辺は θ に依存しない．

このとき，$X = (X_1', \dots, X_n')'$ の同時確率 (密度) 関数は

$$p_\theta^n(x) = \left(\prod_{i=1}^n h(x_i)\right) \exp\left\{\sum_{j=1}^m \xi_j(\theta)\left(\sum_{i=1}^n S_j(x_i)\right) - nc(\theta)\right\}, \quad x = (x_1', \dots, x_n')' \in \mathcal{X}^n$$

である．ここで，

$$H(x) = \prod_{i=1}^n h(x_i),\ C(\theta) = nc(\theta),\ \xi(\theta) = (\xi_1(\theta), \dots, \xi_m(\theta))',$$
$$T(x) = (T_1(x), \dots, T_m(x))' = \left(\sum_{i=1}^n S_1(x_1), \dots, \sum_{i=1}^n S_m(x_n)\right)'$$

とおくと，

$$p_\theta^n(x) = H(x)\exp\left\{\xi(\theta)'T(x) - C(\theta)\right\}, \quad x \in \mathcal{X}^n \tag{*3}$$

と表せる．よって，$\{p_\theta^n : \theta \in \Theta\}$ も指数型分布族をなす．また，因子分解定理より，$T = T(X)$ は十分統計量である．

Theorem 3.4. p_θ^n を (*3) の形の確率 (密度) 関数とし，$\Xi = \{\xi(\theta) : \theta \in \Theta\} \subset \mathbb{R}^m$ の内部は空でないとする．このとき，T は θ に対する完備十分統計量である．

Remark 3.5. $\xi \in \Xi$ を <u>自然パラメータ</u> (natural parameter) と呼ぶ．

Proof. $m = 1$ かつ p_θ^n が確率関数のときに定理を証明する．

$$\{x : H(x) > 0\} = \{x(1), x(2), \dots\}$$

とおくと，$T = T(X)$ の確率関数は，

$$p_\theta^T(t) = \sum_{\nu : T(x(\nu)) = t} p_\theta^n(x(\nu)) = e^{\xi(\theta)t - C(\theta)} \underbrace{\sum_{\nu : T(x(\nu)) = t} H(x(\nu))}_{=G(t)}$$

と表せる．ここで，T のとりうる値を

$$\{T(x(\nu)) : \nu = 1, 2, \dots\} = \{t(\nu) : \nu = 1, 2, \dots\}$$

とおき，$\varphi : \mathbb{R} \to \mathbb{R}$ を

$$\sum_\nu \varphi(t(\nu))p_\theta^T(t(\nu)) = 0,\ \forall\theta \in \Theta \tag{*3}$$

をみたす関数とする．このとき，$\varphi(t(\nu))=0\ \forall\nu=1,2,\dots$ を示せばよい．Ξ の内部は空でないから，必要なら平行移動させて，Ξ は原点の開近傍を含むと仮定してよい．ここで，

$$\begin{aligned}(*3) &\Leftrightarrow \sum_{\nu}\varphi(t(\nu))e^{\xi t(\nu)}G(t(\nu))=0,\ \forall\xi\in\Xi\\ &\Leftrightarrow \sum_{\nu}\underbrace{\varphi^{+}(t(\nu))G(t(\nu))}_{\psi^{+}(t(\nu))}e^{\xi t(\nu)}=\sum_{\nu}\underbrace{\varphi^{-}(t(\nu))G(t(\nu))}_{=\psi^{-}(t(\nu))}e^{\xi t(\nu)},\ \forall\xi\in\Xi\end{aligned}$$

であり，$\xi=0$ として，

$$\sum_{\nu}\psi^{+}(t(\nu))=\sum_{\nu}\psi^{-}(t(\nu))=:c$$

を得る．$c=0$ なら，$\psi^{+}(t(\nu))=\psi^{-}(t(\nu))=0\ \forall\nu=1,2,\dots$ であって，これから，$\varphi(t(\nu))=0\ \forall\nu=1,2,\dots$ を得る．$c>0$ なら，$\psi^{+}(t(\nu))/c,\psi^{-}(t(\nu))/c,\nu=1,2,\dots$ は確率関数であって，それらのモーメント母関数が一致している．よって，$\psi^{+}(t(\nu))=\psi^{-}(t(\nu))\ \forall m=1,2,\dots$ であって，これは $\varphi(t(\nu))=0\ \forall\nu=1,2,\dots$ を意味する． □

Example 3.11. $\mu\in\mathbb{R},\sigma^2>0$ に対して，$X_1,\dots,X_n\sim N(\mu,\sigma^2)$ i.i.d. とすると $(n\geq 2)$，$X=(X_1,\dots,X_n)'$ の同時密度は，

$$p_\theta^n(x)=\exp\left\{\frac{\mu}{\sigma^2}\sum_{i=1}^{n}x_i-\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2-\frac{n}{2\sigma^2}\mu^2-\frac{n}{2}\log(2\pi\sigma^2)\right\}$$

であって，指数型分布族をなす．ここで，自然パラメータのパラメータ空間は

$$\{(\mu/\sigma^2,-1/2\sigma^2):\mu\in\mathbb{R},\sigma^2>0\}=\mathbb{R}\times(-\infty,0)$$

であって，その内部は空でないから，$(\sum_{i=1}^n X_i,\sum_{i=1}^n X_i^2)$ は完備十分統計量であって，その 1 対 1 変換 $(\overline{X},S^2)$ も完備十分統計量である．よって，$\overline{X}$ は μ の一意な UMVU 推定量であり，S^2 は σ^2 の一意な UMVU 推定量である．

これらは自然な推定量といえる．次に，μ^2 の推定を考えてみよう．このとき，$E_{(\mu,\sigma^2)}[\overline{X}^2]=\mu^2+\sigma^2/n$ であるから，

$$\delta^*(\overline{X},S^2)=\overline{X}^2-S^2/n$$

が μ^2 の一意な UMVU 推定量である．しかし，$\delta^*(\overline{X},S^2)$ は正の確率で負になるので，μ^2 の推定量として不合理である．このように，不偏性にこだわると，不合理な推定量が得られてしまうこともある．

正規分布の例において，μ と σ^2 の間になんらかの関係がある場合は，$(\overline{X},S^2)$ は完備にならない．

Example 3.12. $X_1,\dots,X_n\sim N(\theta,\theta^2)$ i.i.d., $\theta>0$ とする．このとき，因子分解定理より，$(\overline{X},S^2)$ は十分統計量であるが，完備でない．実際，

$$E_\theta\left[\overline{X}^2-\frac{n+1}{n}S^2\right]=0$$

だが，$\overline{X}^2 - \{(n+1)/n\}S^2$ は確率 1 で 0 でない．

Poisson 分布やガンマ分布に対しても，Theorem 3.4 が適用できる．しかし，Theorem 3.4 は指数型でない分布族に対しては適用できない．そのような場合でも，十分統計量の完備性を直接確認できる場合がある．

Example 3.13. $\theta > 0$ に対して，$X_1, \dots, X_n \sim U(0,\theta)$ i.i.d. とすると，$X_{(n)}$ は θ に対する十分統計量であった．$X_{(n)}$ の完備性を示そう．$\{U(0,\theta) : \theta > 0\}$ は指数型分布族でないので，Theorem 3.4 は適用できない．$0 < t < \theta$ に対して，

$$P_\theta(X_{(n)} \le t) = P_\theta(X_i \le t,\ 1 \le \forall i \le n) = (t/\theta)^n$$

であるから，$X_{(n)}$ の密度関数は，

$$f_\theta(t) = \frac{nt^{n-1}}{\theta^n} I(0 < t < \theta)$$

である．いま，$\varphi : \mathbb{R} \to \mathbb{R}$ を

$$E_\theta[\varphi(T)] = 0,\ \forall\theta > 0 \Leftrightarrow \int_0^\theta \varphi(t)t^{n-1}dt = 0,\ \forall\theta > 0$$

をみたす関数とする．φ が連続なら，両辺を θ について微分して，$\varphi(\theta) = 0\ \forall\theta > 0$ を得る．φ が連続でなくても，“ほとんどすべての”$t \in (0,\infty)$ に対して $\varphi(t) = 0$ となることが示せるので，$X_{(n)}$ の完備性が従う．

さらに，

$$E_\theta[X_{(n)}] = \frac{n}{\theta^n}\int_0^\theta t^n dt = \frac{n}{n+1}\theta$$

であるから，

$$\delta^*(X_{(n)}) = \frac{n+1}{n}X_{(n)}$$

が θ の一意な UMVU 推定量である．

3.3 Cramér-Rao の不等式

不偏推定量の分散の下界を導出してみよう．本節では，$p_\theta(\cdot) = p(\cdot;\theta)$ として，$X = (X_1', \dots, X_n')'$ の同時確率 (密度) 関数を $p_n(x;\theta)$ とおく：

$$p_n(x;\theta) = \prod_{i=1}^n p(x_i;\theta), \quad x = (x_1', \dots, x_n')' \in \mathcal{X}^n.$$

また，θ に関する偏微分を上付きのドットで表すことにする．例えば，十分なめらかな関数 $g : \Theta \to \mathbb{R}$ に対して，

$$\dot{g}_a(\theta) = \frac{\partial g(\theta)}{\partial\theta_a},\ \ddot{g}_{ab}(\theta) = \frac{\partial^2 g(\theta)}{\partial\theta_a\partial\theta_b},\ a, b = 1, \dots, k$$

である．以下，説明のために，$p(u;\theta)$ は密度関数とするが，確率関数の場合は積分を和に取り替えればよい．

このとき，次の仮定をおく：

- パラメータ空間 Θ は $\mathbb{R}^k$ の空でない開集合である．
- 集合 $\{u \in \mathcal{X} : p(u;\theta) > 0\}$ は θ に依存しない．$A = \{u \in \mathcal{X} : p(u;\theta) > 0\}$ とおく．
- 各 $u \in A$ に対して，$p(u;\theta)$ は θ について偏微分可能である．
- $\ell(u;\theta) = \log p(u;\theta)$ とおいて [18]，各 $a = 1, \ldots, k$ に対して，
$$\int |\dot{\ell}_a(u;\theta)|^2 p(u;\theta) du < \infty,\ \forall \theta \in \Theta.$$
- 各 $a = 1, \ldots, k$ に対して，
$$\int \frac{\partial p(u;\theta)}{\partial \theta_a} du = \frac{\partial}{\partial \theta_a} \underbrace{\int p(u;\theta) du}_{=1} = 0. \tag{*}$$

Remark 3.6. (*) は微分と積分の順序交換が成り立つことを仮定している．Lebesgue 積分の理論を使えば，(*) が成り立つための初等的な条件を与えることが難しくないが，ここでは詳細は省略する．

以上の仮定のもとで，各 $\theta \in \Theta$ に対して，$k \times k$ 行列 $I(\theta) = (I_{ab}(\theta))_{1 \le a,b \le k}$ を

$$I_{ab}(\theta) = E_\theta[\dot{\ell}_a(X_1;\theta)\dot{\ell}_b(X_1;\theta)] = \int \dot{\ell}_a(u;\theta)\dot{\ell}_b(u;\theta)p(u;\theta)du$$

と定義して，$I(\theta)$ を $p(\cdot;\theta)$ の Fisher 情報行列 (Fisher information matrix) と呼ぶ．$k=1$ のときは，$I(\theta)$ を Fisher 情報量 (Fisher information) と呼ぶ．

(1). (*) より，

$$0 = \int \frac{\partial p(u;\theta)}{\partial \theta_a} du = \int \dot{\ell}_a(u;\theta)p(u;\theta)du = E_\theta[\dot{\ell}_a(X_1;\theta)]$$

が成り立つから，

$$I_{ab}(\theta) = \mathrm{Cov}_\theta(\dot{\ell}_a(X_1;\theta), \dot{\ell}_b(X_1;\theta))$$

を得る．これから，

$$I(\theta) = \mathrm{Var}_\theta(\dot{\ell}(X_1;\theta))$$

を得る．特に，$I(\theta)$ は半正定値対称行列である．

[18] $\ell(u;\theta)$ は $u \in A$ に対して定義される．$p(u;\theta)$ に関する積分は積分範囲を A に制限しているとみなす．

(2). さらに，$p(u;\theta)$ が θ について 2 回偏微分可能なら，

$$\ddot{\ell}_{ab}(u;\theta) = \frac{\partial^2 \ell(u;\theta)}{\partial\theta_a \partial\theta_b} = \frac{\partial^2 p(u;\theta)/\partial\theta_a\partial\theta_b}{p(u;\theta)} - \underbrace{\frac{\{\partial p(u;\theta)/\partial\theta_a\}\{\partial p(u;\theta)/\partial\theta_b\}}{p(u;\theta)^2}}_{=\dot{\ell}_a(u;\theta)\dot{\ell}_b(u;\theta)}$$

であり，微分と積分の順序交換

$$E_\theta\left[\frac{\partial^2 p(X_1;\theta)/\partial\theta_a\partial\theta_b}{p(X_1;\theta)}\right] = \int \frac{\partial^2 p(u;\theta)}{\partial\theta_a\partial\theta_b} du = \frac{\partial^2}{\partial\theta_a\partial\theta_b}\int p(u;\theta)du = 0 \qquad (**)$$

を認めれば，

$$I_{ab}(\theta) = E_\theta[-\ddot{\ell}_{ab}(X_1;\theta)] \qquad (*3)$$

を得る．多くの場合，(*3) の方が計算しやすい．(*3) の等式を 情報量等式 (information identity) と呼ぶ．

(3). 同様にして，$p_n(\cdot;\theta)$ の Fisher 情報行列 $I^n(\theta)$ は，$\ell^n(x;\theta) = \log p_n(x;\theta)$ とおくと，

$$I^n_{ab}(\theta) = E_\theta[\dot{\ell}^n_a(X;\theta)\dot{\ell}^n_b(X;\theta)] = \int \dot{\ell}_a(x;\theta)\dot{\ell}_b(x;\theta)p_n(x;\theta)dx$$

と定義される．ここで，$p_n(x;\theta) = \prod_{i=1}^n p(x_i;\theta)$ より，$\ell^n(x;\theta) = \sum_{i=1}^n \ell(x_i;\theta)$ であるから，$I^n(\theta)$ の存在はよい．さらに，$E_\theta[\dot{\ell}^n(X;\theta)] = 0$ と $X_1,\dots,X_n \sim p(\cdot;\theta)$ i.i.d. より，

$$I^n(\theta) = \mathrm{Var}_\theta(\dot{\ell}^n(X;\theta)) = \sum_{i=1}^n \underbrace{\mathrm{Var}_\theta(\dot{\ell}(X_i;\theta))}_{=I(\theta)} = nI(\theta)$$

を得る．

Theorem 3.5 (Cramér-Rao の不等式). 次の条件を仮定する．

- $g : \Theta \to \mathbb{R}$ は偏微分可能な関数である．
- 各 $\theta \in \Theta$ に対して Fisher 情報行列 $I(\theta)$ は正則である．
- $\delta(X)$ は $E_\theta[\delta(X)^2] < \infty\ \forall\theta \in \Theta$ をみたし，$g(\theta)$ の不偏推定量であって，

  $$\dot{g}_a(\theta) = \frac{\partial}{\partial\theta_a}\int \delta(x)p_n(x;\theta)dx = \int \delta(x)\frac{\partial p_n(x;\theta)}{\partial\theta_a}dx,\ a = 1,\dots,k$$

  が成り立つ．

このとき，

$$\mathrm{Var}_\theta(\delta(X)) \ge \dot{g}(\theta)' I(\theta)^{-1}\dot{g}(\theta)/n,\ \forall\theta \in \Theta \qquad (*4)$$

が成り立つ．

(*4) の右辺を $g(\theta)$ に対する Cramér-Rao の下界 (lower bound) と呼ぶ. 例えば, $g(\theta)=\theta_a$ なら,

$$n^{-1}\dot{g}(\theta)'\{I(\theta)\}^{-1}\dot{g}(\theta)=n^{-1}\underbrace{(I(\theta)^{-1})_{aa}}_{I(\theta)^{-1}\text{ の第 }(a,a)\text{ 成分}}$$

である. また, $k=1$ なら,

$$\mathrm{Var}_\theta(\delta(X))\geq\{g'(\theta)\}^2/\{nI(\theta)\}$$

である.

Proof. $\delta(X)$ は $g(\theta)$ の不偏推定量であるから,

$$g(\theta)=E_\theta[\delta(X)]=\int\delta(x)p_n(x;\theta)dx$$

が成り立つ. 両辺を θ_a について偏微分して,

$$\dot{g}_a(\theta)=\int\delta(x)\frac{\partial p_n(x;\theta)/\partial\theta_a}{p_n(x;\theta)}p_n(x;\theta)dx=E_\theta[\delta(X)\dot{\ell}_a^n(X;\theta)]=\mathrm{Cov}_\theta(\delta(X),\dot{\ell}_a^n(X;\theta)).$$

最後の等号は $E_\theta[\dot{\ell}_a^n(X;\theta)]=0$ から従う. 従って, $\forall z\in\mathbb{R}^k$ に対して,

$$z'\dot{g}(\theta)=\mathrm{Cov}_\theta(\delta(X),z'\dot{\ell}^n(X;\theta))$$

を得る. Schwarz の不等式より,

$$\{z'\dot{g}(\theta)\}^2\leq\mathrm{Var}_\theta(\delta(X))\underbrace{\mathrm{Var}_\theta(z'\dot{\ell}^n(X;\theta))}_{=z'I^n(\theta)z}$$

であるから, $\forall z\in\mathbb{R}^k\setminus\{0\}$ に対して,

$$\mathrm{Var}_\theta(\delta(X))\geq\frac{\{z'\dot{g}(\theta)\}^2}{z'I^n(\theta)z}$$

を得る. 右辺を z について最大化する. $k\times k$ 行列 B を $I^n(\theta)=BB'$ をみたすように選び (B は正則), $w=B'z$ とおくと,

$$\frac{\{z'\dot{g}(\theta)\}^2}{z'I^n(\theta)z}=\frac{\{w'B^{-1}\dot{g}(\theta)\}^2}{w'B^{-1}I^n(\theta)(B')^{-1}w}=\frac{\{w'B^{-1}\dot{g}(\theta)\}^2}{w'w}$$

となる. 右辺は $w=B^{-1}\dot{g}(\theta)$ のとき最大値

$$\begin{aligned}&\dot{g}(\theta)'(B^{-1})'B^{-1}\dot{g}(\theta)=\dot{g}(\theta)'(BB')^{-1}\dot{g}(\theta)\\&=\dot{g}(\theta)'I^n(\theta)^{-1}\dot{g}(\theta)=n^{-1}\dot{g}(\theta)'I(\theta)^{-1}\dot{g}(\theta)\end{aligned}$$

をとる [19]. 以上より定理が示された. □

[19] $a=B^{-1}\dot{g}(\theta)$ とおくと, Schwarz の不等式より, $(w'a)^2\leq(w'w)(a'a)$ であって, 等号は $w=a$ のとき成立する.

Cramér-Rao の下界を達成する不偏推定量は UMVU である．例えば，$X_1, \ldots, X_n \sim N(\theta, 1)$ i.i.d. とすると，$N(\theta, 1)$ の Fisher 情報量は 1 である．よって，Cramér-Rao の下界は $1/n$ である．いま，$\overline{X}$ の分散は $1/n$ であるから，Cramér-Rao の下界を達成する．よって，$\overline{X}$ は θ の UMVU 推定量である (もちろん，$\overline{X}$ は θ に対する完備十分統計量なので，Lehmann-Scheffé の定理から $\overline{X}$ が UMVU 推定量であることもわかる)．

なお，次の例が示すように，Cramér-Rao の下界は達成可能とは限らない．

Example 3.14. $\mu \in \mathbb{R}, \sigma^2 > 0$ に対して，$X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ i.i.d. とする ($n \geq 2$)．このとき，$N(\mu, \sigma^2)$ の密度関数は

$$p(u; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(u-\mu)^2}{2\sigma^2}}$$

である．$\tau = \sigma^2$ とおくと，

$$\ell(u; \mu, \tau) = \log p(u; \mu, \tau) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\tau - \frac{(u-\mu)^2}{2\tau}$$

であるから，$N(\mu, \tau)$ の Fisher 情報行列は

$$I(\mu, \tau) = \begin{pmatrix} \frac{1}{\tau} & 0 \\ 0 & \frac{1}{2\tau^2} \end{pmatrix}$$

であって (演習問題)，その逆行列は

$$I(\mu, \tau)^{-1} = \begin{pmatrix} \tau & 0 \\ 0 & 2\tau^2 \end{pmatrix}$$

となる．よって，μ に対する Cramér-Rao の下界は，$\tau/n = \sigma^2/n$ であって，これは $\overline{X}$ によって達成可能である．一方，σ^2 に対する Cramér-Rao の下界は，$2\tau^2/n = 2\sigma^4/n$ である．しかし，σ^2 の一意な UMVU 推定量は S^2 であって，その分散は

$$\mathrm{Var}_{(\mu,\sigma^2)}(S^2) = \frac{2\sigma^4}{n-1}$$

と，Cramér-Rao の下界より大きい [20]．S^2 が UMVU なことから，σ^2 の不偏推定量が $2\sigma^4/(n-1)$ より小さい分散をもつことはありえないので，Cramér-Rao の下界は達成不可能であることがわかる．

Example 3.15. $\{u \in \mathcal{X} : p(u; \theta) > 0\}$ が θ に依存しないという仮定は Cramér-Rao の不等式において本質的である．いま，$X \sim U(0, \theta), \theta > 0$ としよう．このとき，$p(x; \theta) =$

---

[20] $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$ と $\chi^2(k)$ の分散が $2k$ であることから，$(n-1)^2 \mathrm{Var}(S^2)/\sigma^4 = \mathrm{Var}((n-1)S^2/\sigma^2) = 2(n-1)$ だから，$\mathrm{Var}(S^2) = 2\sigma^4/(n-1)$ を得る．

$\theta^{-1}I(0 < x < \theta)$ であって，$p(x;\theta) > 0$ なる x の集合は θ に依存する．$0 < x < \theta$ という制約を無視して形式的に Fisher 情報量を計算すると，

$$E_\theta[\dot{\ell}(X;\theta)^2] = \theta^{-2}$$

となる．しかし，θ の UMVU 推定量は $2X$ であって，その分散は $\theta^2/3$ であって，Cramér-Rao の下界 θ^2 より小さい．また，この例だと，$\ddot{\ell}(x;\theta) = \theta^{-2}$ になるので，情報量等式が成り立っていない．

このように，Cramér-Rao の不等式は UMVU 推定量を求めることに関しては Lehmann-Scheffé の定理より便利な方法とは言えない．しかし，Fisher 情報行列は漸近理論において重要な役割を果たす．

Remark 3.7. Cramér-Rao の不等式は，H.L. Cramér (1946 年) と C.R. Rao (1945 年) が独立に導いたことから彼らの名前がついているが，それより前の 1943 年に M. Fréchet によってすでに導出されていて，さらにその拡張が 1945 年に G. Darmois によってなされていることが判明している (Lehmann and Casella, 1998, p. 143)．従って，(**) の不等式を単に "情報量不等式" と呼ぶ場合もある．

3.4 最尤推定

いままで，不偏推定を考察してきたが，そもそも不偏推定量が存在しない場合も多くある．また，不偏性が推定量の制約として適切かという問題もある．最尤推定法はパラメトリックな分布族に対して汎用的に適用可能な推定手法であり，広く使われている．というよりも，パラメトリックモデルに対しては，最尤推定か次節で考察する Bayes 推定がほぼデフォルトとして使用されている．

以下，パラメータ空間を $\varnothing \neq \Theta \subset \mathbb{R}^k$ とし，$\{p(\cdot;\theta) : \theta \in \Theta\}$ を $\mathcal{X}$ 上のパラメトリックな分布族とする．$X_1, \dots, X_n \sim p(\cdot;\theta)$ i.i.d. として．$X = (X_1', \dots, X_n')'$ の確率 (密度) 関数を $p_n(x;\theta) = \prod_{i=1}^n p(x_i;\theta)$ とおく．このとき，$p_n(X;\theta)$ を θ の関数とみなした

$$L_n(\theta) = p_n(X;\theta)$$

を θ の <u>尤度関数</u> (likelihood function) と呼び，尤度関数を最大化する点 $\widehat{\theta} = \widehat{\theta}(X)$

$$L_n(\widehat{\theta}) = \max_{\theta\in\Theta} L_n(\theta)$$

を θ の <u>最尤推定量</u> (maximum likelihood estimator, MLE) と呼ぶ (MLE の存在は仮定する)．$L_n(\theta)$ の最大化はその対数 $\ell^n(\theta) = \log L_n(\theta)$ の最大化と等価であるから，MLE $\widehat{\theta}$ は

$$\ell^n(\widehat{\theta}) = \max_{\theta\in\Theta} \ell^n(\theta)$$

とも定義できる．$\ell^n(\theta)$ は <u>対数尤度関数</u> (log likelihood function) と呼ばれる．

$\ell^n(\theta)$ が θ のなめらかな関数なら，$\widehat{\theta}$ が Θ の内点である限り，

$$\dot{\ell}^n(\theta) = \left(\frac{\partial \ell^n(\theta)}{\partial \theta_1}, \ldots, \frac{\partial \ell^n(\theta)}{\partial \theta_k}\right)' = 0 \tag{*}$$

をみたす．(*) は 尤度方程式 (likelihood equation) と呼ばれる．

十分統計量 $T = T(X)$ が存在するとき，因子分解定理より，$p_n(x;\theta)$ は $p_n(x;\theta) = g_\theta(T(x))h(x)$ の形に分解できるから，対数尤度は

$$\ell^n(\theta) = \log g_\theta(T(X)) + \log h(X)$$

となる．$h(X)$ は θ に依存しないので，対数尤度の最大化は $\log g_\theta(T(X))$ の最大化と等価である．よって，$\widehat{\theta}$ は T のみの関数である．

また，MLE はパラメータの 1 対 1 変換に関して不変である．すなわち，$g(\theta)$ が θ が 1 対 1 変換なら，$\xi = g(\theta)$ の MLE は $\widehat{\xi} = g(\widehat{\theta})$ になる．これは，ξ の尤度が

$$\xi \mapsto L_n(g^{-1}(\xi)) =$$

であることから明らかである．例えば，$X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ i.i.d. に対して，簡単な計算から (μ, σ^2) の MLE は

$$\widehat{\mu} = \overline{X},\ \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2$$

である．$\widehat{\sigma}^2 \neq S^2$ であることに注意する．ここで，$\sigma = \sqrt{\sigma^2}$ の MLE は $\widehat{\sigma} = \sqrt{\widehat{\sigma}^2}$ になる．この不変性は不偏推定にはない性質である．

Remark 3.8. MLE が陽に求められる場合はむしろまれであり，ほとんどの場合，数値的な最適化によって計算する．

MLE は $n \to \infty$ のとき，よい漸近的な性質をもつ．簡単のため，$k = 1$ としよう．$\theta = \theta_0$ を真値とすると，いくつかの正則条件のもとで，MLE $\widehat{\theta}$ は 一致性 (consistency)

$$\widehat{\theta} \xrightarrow{P} \theta_0$$

と 漸近正規性 (asymptotic normality)

$$\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, 1/I(\theta_0))$$

をみたす．$I(\theta)$ は $p(\cdot;\theta)$ の Fisher 情報量である．すなわち，$\widehat{\theta}$ は近似的に

$$N(\theta_0, 1/(nI(\theta_0)))$$

に従うから，近似的に Cramér-Rao の下界を達成する．

ここでは，MLE の一致性を認めて漸近正規性を導出してみよう．θ_0 を Θ の内点とし，$\ell^n(\theta)$ は θ についてなめらかとしよう．このとき，$\widehat{\theta}$ が Θ の内点である確率は 1 に近づく．そこで，尤度方程式を展開して，

$$0 = \dot{\ell}^n(\theta_0) + \ddot{\ell}^n(\widetilde{\theta})(\widehat{\theta} - \theta_0)$$

を得る．ここで，$\overline{\theta}$ は $\widehat{\theta}$ と θ_0 の間の点である．いま，$E_{\theta_0}[\dot{\ell}(X_i;\theta_0)] = 0$ と CLT より，

$$n^{-1/2}\dot{\ell}^n(\theta_0) = n^{-1/2}\sum_{i=1}^n \dot{\ell}(X_i;\theta_0) \xrightarrow{d} N(0, I(\theta_0))$$

である．一方，$\widehat{\theta} \xrightarrow{P} \theta_0$ より，いくつかの正則条件のもとで，

$$n^{-1}\ddot{\ell}^n(\overline{\theta}) = n^{-1}\sum_{i=1}^n \ddot{\ell}(X_i;\overline{\theta}) \xrightarrow{P} E_{\theta_0}[\ddot{\ell}(X_i;\theta_0)]$$

となる．情報量等式を認めれば，右辺は $-I(\theta_0)$ に等しい．よって，Slutsky の補題より，

$$\sqrt{n}(\widehat{\theta} - \theta_0) = \left(-n^{-1}\ddot{\ell}^n(\overline{\theta})\right)^{-1}\left(n^{-1/2}\dot{\ell}^n(\theta_0)\right) \xrightarrow{d} N(0, 1/I(\theta_0))$$

を得る．

Example 3.16. $\theta \in (0,1)$ に対して，$X_1, \ldots, X_n \sim Bin(1,\theta)$ とする．このとき，

$$p_n(x;\theta) = \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}$$

であるから，対数尤度は

$$\ell^n(\theta) = n\overline{X}\log\theta + n(1-\overline{X})\log(1-\theta)$$

である．これを θ について微分して，

$$0 = \frac{n\overline{X}}{\theta} - \frac{n(1-\overline{X})}{1-\theta}$$

を解くと，MLE は $\widehat{\theta} = \overline{X}$ である．$\theta = \theta_0$ を真値とすると，CLT より，

$$\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, \theta_0(1-\theta_0))$$

となる．ここで，$Bin(1,\theta)$ の Fisher 情報量 $I(\theta)$ は，$\ell(u;\theta) = u\log\theta + (1-u)\log(1-\theta)$ とおくと，

$$I(\theta) = E_\theta[-\partial^2\ell(X_i,\theta)/\partial\theta^2] = E_\theta\left[\frac{X_i}{\theta^2} + \frac{1-X_i}{(1-\theta)^2}\right] = \frac{1}{\theta} + \frac{1}{1-\theta} = \frac{1}{\theta(1-\theta)}$$

であるから，$1/I(\theta_0) = \theta_0(1-\theta_0)$ である．

3.5 Bayes 推定

Bayes 推定は最尤推定と並ぶ汎用的な推定法である．Bayes 推定では，パラメータ空間 Θ の上に確率 (密度) 関数 $\pi(\theta)$ を用意して，$\theta \sim \pi$ とみなす．π を 事前分布 (prior distribution) と呼ぶ．$p_\theta^n(x)$ を θ を与えたときの X の条件付き確率 (密度) 関数とみなす．これを

$$\begin{cases} X_1, \dots, X_n \mid \theta \sim p_\theta \quad \text{i.i.d.} \\ \theta \sim \pi \end{cases} \tag{*}$$

と表す．ここで，$g(\theta)$ の推定を考えて，$g(\theta)$ の推定量 $\delta(X)$ に対して，そのリスク関数を $R(\theta, \delta)$ とする．このとき，$R(\theta, \delta)$ を θ について期待値をとった

$$r(\pi, \delta) = \int_\Theta R(\theta, \delta) \pi(\theta) d\theta$$

を Bayes リスク と呼ぶ (π が確率関数なら積分を和に取り替える)．Bayes リスクを最小にする推定量 $\delta^\pi(X)$ を事前分布 π に対する Bayes 推定量 と呼ぶ：

$$r(\pi, \delta^\pi) = \min_{\delta:\ g(\theta)\ \text{の推定量}} r(\delta, \pi).$$

説明のために，π は密度関数とし，分布と確率 (密度) 関数を同一視する．Bayes 推定量の計算は事後分布にもとづく．$X \mid \theta \sim p_\theta^n, \theta \sim \pi$ のとき，(X, θ) の同時分布は

$$p_\theta^n(x)\pi(\theta)$$

であるから，X の周辺分布は

$$m_\pi(x) = \int p_\theta^n(x) \pi(\theta) d\theta$$

であって，$X = x$ を与えたときの θ の条件付き分布は

$$\pi(\theta \mid x) = \frac{p_\theta^n(x)\pi(\theta)}{m_\pi(x)}, \quad m_\pi(x) > 0$$

である．これを $X = x$ を与えたときの θ の 事後分布 (posterior distribution) と呼ぶ．このとき，Bayes リスクは

$$r(\pi, \delta) = \int_\Theta \int L(\theta, \delta(x)) p_\theta^n(x) \pi(\theta) dx d\theta = \int \left\{ \int_\Theta L(\theta, \delta(x)) \pi(\theta \mid x) d\theta \right\} m_\pi(x) dx$$

と表せる．各 $d \in D(= \mathbb{R})$ に対して，$X = x$ を与えたときの 条件付きリスク を

$$r_x(\pi, d) = \int_\Theta L(\theta, d) \pi(\theta \mid x) d\theta$$

と定義して，各 x に対して，$\delta^\pi(x)$ を $r_x(\pi, d)$ を最小化する d にとる：

$$r_x(\pi, \delta^\pi(x)) = \min_{d \in D} r_x(\pi, d).$$

このとき，$\delta^\pi(X)$ は Bayes 推定量になっている．十分統計量 T が存在する場合，T を与えたときの θ の事後分布にもとづいて Bayes 推定量を計算すればよい．

Theorem 3.6. $T = T(X)$ を θ に対する (ベクトル値の) 十分統計量とする．このとき，X を与えたときの θ の事後分布は，T を与えたときの θ の事後分布に等しい．

Proof. p_θ^n が確率関数の場合に定理を証明する．因子分解定理より，$p_\theta^n(x) = g_\theta(T(x))h(x)$ と表せるから，

$$\widetilde{m}_\pi(t) = \int g_\theta(t)\pi(\theta)d\theta$$

とおくと，X を与えたときの θ の事後分布は

$$\pi(\theta \mid X) = \frac{g_\theta(T)h(X)\pi(\theta)}{h(X)\widetilde{m}_\pi(T)} = \frac{g_\theta(T)\pi(\theta)}{\widetilde{m}_\pi(T)}$$

となる．一方，

$$P(T = t \mid \theta) = g_\theta(t) \underbrace{\sum_{x:T(x)=t} h(x)}_{=\widetilde{h}(t)}$$

であるから，T を与えたときの θ の事後分布は

$$\frac{g_\theta(T)\widetilde{h}(T)\pi(\theta)}{\int g_\vartheta(T)\widetilde{h}(T)\pi(\vartheta)d\vartheta} = \frac{g_\theta(T)\pi(T)}{\widetilde{m}_\pi(T)}$$

である． □

以下，2 乗損失関数 $L(\theta, d) = (d - g(\theta))^2$ を考える．ここで，$g : \Theta \to \mathbb{R}$ を $E[g(\theta)^2] < \infty$ をみたす関数とする．このとき，条件付きリスクは

$$r_x(\pi, d) = \int_\Theta (d - g(\theta))^2 \pi(\theta \mid x)d\theta$$

であって，これを最小化する d は条件付き期待値

$$\delta^\pi(x) = \int g(\theta)\pi(\theta \mid x)d\theta$$

である．よって，Bayes 推定量は

$$\delta^\pi(X) = \int g(\theta)\pi(\theta \mid X)d\theta \tag{**}$$

である．(**) を $g(\theta)$ の 事後平均 (posterior mean) と呼ぶ．十分統計量 T が存在すれば，

$$\delta^\pi(X) = E[g(\theta) \mid T]$$

である．

Example 3.17. $\theta \in (0,1), \alpha, \beta > 0$ に対して，

$$\begin{aligned}
&X_1, \dots, X_n \mid \theta \sim Bin(1, \theta) \quad \text{i.i.d.} \\
&\theta \sim Be(\alpha, \beta)
\end{aligned}$$

とする．$T = \sum_{i=1}^n X_i$ は θ に対する十分統計量である．このとき，(T, θ) の同時分布は

$$\pi(t, \theta) = \binom{n}{t} \theta^t (1-\theta)^{n-t} \times \frac{1}{Be(\alpha, \beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

であるから，

$$\pi(\theta \mid t) \propto \theta^{t+\alpha-1} (1-\theta)^{n-t+\beta-1}$$

である．よって，

$$\theta \mid T \sim Be(T+\alpha, n-T+\beta)$$

であるから，θ の事後平均は

$$\widehat{\theta}^{\alpha,\beta} = \frac{T+\alpha}{n+\alpha+\beta}$$

である．

Example 3.18. $\mu, \xi \in \mathbb{R}, \sigma^2, \tau^2 > 0$ に対して，

$$\begin{aligned}
&X_1, \dots, X_n \mid \theta \sim N(\mu, \sigma^2) \quad \text{i.i.d.} \\
&\mu \sim N(\xi, \tau^2)
\end{aligned}$$

とし，σ^2 は既知とする．$T = \overline{X}$ は μ に対する十分統計量である．(T, μ) の同時密度は

$$\pi(t, \mu) = \frac{\sqrt{n}}{2\pi\sigma\tau} \exp\left\{ -\frac{1}{2} \{ n(t-\mu)^2/\sigma^2 + (\mu-\xi)^2/\tau^2 \} \right\}$$

である．ここで，

$$n(t-\mu)^2/\sigma^2 + (\mu-\xi)^2/\tau^2 = \left(n/\sigma^2 + 1/\tau^2\right) \left(\mu - \frac{nt/\sigma^2 + \xi/\tau^2}{n/\sigma^2 + 1/\tau^2} \right)^2 + (t \text{ の関数})$$

であるから，

$$\mu \mid T \sim N\left(\frac{nT/\sigma^2 + \xi/\tau^2}{n/\sigma^2 + 1/\tau^2}, \left(n/\sigma^2 + 1/\tau^2\right)^{-1} \right).$$

従って，事後平均は

$$\widehat{\mu}^{\xi,\tau^2} = \frac{n\overline{X}/\sigma^2 + \xi/\tau^2}{n/\sigma^2 + 1/\tau^2}$$

である．

Example 3.19. $\sum_{j=1}^k \theta_j = 1$ をみたす $\theta_1, \ldots, \theta_k > 0$ と $\alpha_1, \ldots, \alpha_k > 0$ に対して，

$$\begin{aligned} &X = (X_1, \ldots, X_k)' \mid (\theta_1, \ldots, \theta_k)' \sim Mn(n, \theta_1, \ldots . \theta_k) \\ &(\theta_1, \ldots, \theta_k)' \sim Di(\alpha_1, \ldots, \alpha_k) \end{aligned}$$

とする．ここで，$Di(\alpha_1, \ldots, \alpha_k)$ はパラメータ $\alpha_1, \ldots, \alpha_k$ をもつ Dirichlet 分布である．Dirichlet 分布の定義から，$(\theta_1, \ldots, \theta_k)' \sim Di(\alpha_1, \ldots, \alpha_k)$ のとき，$\sum_{j=1}^k \theta_j = 1$ という制約は自動的にみたされる．いま，$(\theta_1, \ldots, \theta_k)'$ は密度関数をもたないが，$(\theta_1, \ldots, \theta_{k-1})'$ は密度関数をもっていて，それは

$$\pi(\theta_1., \ldots, \theta_{k-1}) \propto \left(\prod_{j=1}^{k-1} \theta_j^{\alpha_j - 1} \right) \left(1 - \sum_{j=1}^{k-1} \theta_j \right)^{\alpha_k - 1}$$

で与えられる．$\theta_1, \ldots, \theta_{k-1}$ は $\theta_j > 0$ $(j = 1, \ldots, k-1), \sum_{j=1}^{k-1} \theta_j < 1$ に制約されている．よって，$X, \theta_1, \ldots, \theta_{k-1}$ の同時分布は

$$\pi(x, \theta_1, \ldots, \theta_{k-1}) \propto \left(\prod_{j=1}^{k-1} \theta_j^{x_j + \alpha_j - 1} \right) \left(1 - \sum_{j=1}^{k-1} \theta_j \right)^{x_k + \alpha_k - 1}$$

であって，これと $\theta_k = 1 - \sum_{j=1}^{k-1} \theta_j$ という制約から，

$$(\theta_1, \ldots, \theta_k)' \mid X \sim Di(X_1 + \alpha_1, \ldots, X_k + \alpha_k)$$

を得る．よって，各 θ_j の事後平均は

$$\widehat{\theta}_j = \frac{X_j + \alpha_j}{n + \sum_{i=1}^k \alpha_i}$$

である．

上の例のように，事前分布と事後分布が同じ分布族に入る事前分布を，共役事前分布 (conjugate prior distribution) という．また，事前分布に現れるパラメータをハイパーパラメータ (hyper parameter) と呼ぶ．上の 2 つの例で，事後平均は不偏になってない．一般に事後平均は不偏にならない．

Theorem 3.7. $X \mid \theta \sim p_\theta^n, \theta \sim \pi$ とし，$g : \Theta \to \mathbb{R}$ を $E[g(\theta)^2] < \infty$ をみたす関数とする．さらに，$g(\theta)$ の事後平均 $\delta^\pi(X) = E[g(\theta) \mid X]$ は不偏になっているとする：

$$E[\delta^\pi(X) \mid \theta] = g(\theta) \ \forall \theta \in \Theta.$$

このとき，(X, θ) の同時分布について，$P(\delta^\pi(X) = g(\theta)) = 1$ となる．

Proof. δ^π の Bayes リスクは X を条件付けして，

$$r(\pi,\delta^\pi) = E[(\delta^\pi(X)-g(\theta))^2] = E[(E[g(\theta)\mid X]-g(\theta))^2] = E[g(\theta)^2]-E[\delta^\pi(X)^2]$$

と計算できる．一方，$\delta^\pi(X)$ が不偏なら，θ を条件付けして，

$$r(\pi,\delta^\pi) = E[E[\delta^\pi(X)^2\mid\theta]-g(\theta)^2] = E[\delta^\pi(X)^2]-E[g(\theta)^2]$$

とも計算できる．よって，$r(\pi,\delta^\pi)=-r(\pi,\delta^\pi)$ であるから，$r(\pi,\delta^\pi)=0$ を得る．これから，$P(\delta^\pi(X)=g(\theta))=1$ を得る． □

なお，Bayes 推定において，パラメータが本当に確率的であることを信じる必要はなくて，本当はパラメータには真値があるが，単に推定量を得るための手段として (*) という設定を考えている，と解釈するほうが生産的である [21]．

Bayes 推定量は MLE と同様によい漸近的な性質をもつ．簡単のため，$k=1$ として，$\theta=\theta_0$ を真値とする．このとき，いくつかの正則条件のもとで，θ の事後平均 $\widehat{\theta}^\pi$ は，MLE $\widehat{\theta}$ と次の意味で漸近的に同等であることが示せる：

$$\sqrt{n}(\widehat{\theta}^\pi-\widehat{\theta}) \xrightarrow{P} 0, \quad n\to\infty.$$

従って，$\sqrt{n}(\widehat{\theta}^\pi-\theta_0) \xrightarrow{d} N(0,1/I(\theta_0))$ となる．この結果は，事前密度 π が $\int\theta^2\pi(\theta)d\theta<\infty$ をみたしていて，真値 $\theta=\theta_0$ の近傍で連続かつ正の確率をもつ限り，その選び方にはよらず成り立つ．そのほかに，Bayes 推定量は決定理論的な意味からも望ましい性質をもつ (後述)．

Remark 3.9. 事後平均が陽に求められる場合はむしろまれであり，多くの場合，事後分布に近似的に従う乱数を発生させて，積分を数値的に近似する．事後分布に近似的に従う乱数を発生させる有効な手段として，マルコフチェイン・モンテカルロ法 (MCMC) と呼ばれる手法がある．MCMC とその Bayes 統計への応用に関しては，Gamerman and Lopes (2006) や Robert and Casella (2004) が詳しい (前者の方が入門的である)．Bayes 統計そのものに関しては，Robert (2007) が詳しい．

3.6 許容性とミニマクス性

推定量の最適性を再考する．$X\sim p_\theta^n, \theta\in\Theta\subset\mathbb{R}^k$ とし，与えられた $g:\Theta\to\mathbb{R}$ に対して，$g(\theta)$ の推定を考える．決定空間は $D=\mathbb{R}$ とし，損失関数を $L(\theta,d)\geq 0$ とおく．このとき，$g(\theta)$ の推定量 $\delta(X)$ に対して，リスク関数は

$$R(\theta,\delta) = E_\theta[L(\theta,\delta(X))]$$

と定義されるのであった．

[21] 個人的な意見として，Bayes 統計の考え方が正しく，従ってそれ以外の考え方を認めない (し勉強しない)，というのは不健全だと思う．

Definition 5. (優越性と許容性).

(1) 推定量 δ_1 が別の推定量 δ_2 を 優越する (dominate) とは，$\forall\theta \in \Theta$ に対して，$R(\theta, \delta_1) \le R(\theta, \delta_2)$ であって，ある $\theta_0 \in \Theta$ に対して，$R(\theta_0, \delta_1) < R(\theta_0, \delta_2)$ となることをいう．

(2) 推定量 δ^* が 許容的 (admissible) であるとは，δ^* を優越する推定量が存在しないことをいう．許容的でない推定量を 非許容的 (inadimissible) という．

許容性は推定量がもつべき弱い要請である．例えば，多くの場合，固定されたパラメータ $\theta_0 \in \Theta$ への "決めうち" $\delta(X) = g(\theta_0)$ は許容的になる．

Theorem 3.8. 任意の相異なる $\theta_1, \theta_2 \in \Theta$ に対して，$P_{\theta_1}(X \in A) = 0 \Leftrightarrow P_{\theta_2}(X \in A) = 0$ と仮定する．さらに，$L(\theta, d) = 0 \Leftrightarrow d = g(\theta)$ と仮定する．このとき，任意に固定した $\theta_0 \in \Theta$ に対して，$\delta_0(X) = g(\theta_0)$ という推定量は許容的である．

Remark 3.10. $\theta_1 \ne \theta_2$ に対して，$P_{\theta_1}(X \in A) = 0 \Leftrightarrow P_{\theta_2}(X \in A) = 0$ となる条件は，$\{p_\theta : \theta \in \Theta\}$ が指数型分布族なら成り立つ．

Proof. δ を $R(\theta, \delta) \le R(\theta, \delta_0)\ \forall\theta \in \Theta$ をみたす推定量とする．このとき，$R(\theta_0, \delta_0) = 0$ より，$R(\theta_0, \delta) = 0$ であって，これから，$P_{\theta_0}\{L(\theta_0, \delta(X)) = 0\} = 1$ を得る．よって，$P_{\theta_0}(\delta(X) = g(\theta_0)) = 1$．さらに，仮定より，$P_\theta(\delta(X) = g(\theta_0)) = 1\ \forall\theta \in \Theta$ となるから，$R(\theta, \delta) = R(\theta, \delta_0)\ \forall\theta \in \Theta$ となる．従って，δ_0 を優越する推定量は存在しない． □

とはいえ，許容性の要請から，いつかの不合理と思われる推定量を排除できる．

Example 3.20. $\mu \in \mathbb{R}, \sigma^2 > 0$ を未知とし，$X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ i.i.d. とすると，μ^2 の UMVU 推定量は $\delta(X) = \overline{X}^2 - S^2/n$ であった．このとき，$\delta^+(X) = \max\{\overline{X}^2 - S^2/n, 0\}$ とおくと，$E_{(\mu,\sigma^2)}[(\delta^+(X) - \mu^2)^2] < E_{(\mu,\sigma^2)}[(\delta(X) - \mu^2)^2]\ \forall(\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$ となる．よって，δ は (UMVU であるが) 非許容的である．

Bayes 推定量は多くの場合，許容的になる．

Theorem 3.9. π を Θ 上の事前分布とし，π に対して，δ^π を一意な Bayes 推定量とする．すなわち，η^π がもう 1 つの π に対する Bayes 推定量なら，$P_\theta(\eta^\pi(X) = \delta^\pi(X)) = 1\ \forall\theta \in \Theta$ になるとする．このとき，δ^π は許容的である．

Proof. δ を $R(\theta, \delta) \le R(\theta, \delta^\pi)\ \forall\theta \in \Theta$ をみたす推定量とする．このとき，$r(\pi, \delta) \le r(\pi, \delta^\pi)$ であるから，δ も Bayes 推定量である．しかし，δ^π の一意性より，$P_\theta(\delta(X) = \delta^\pi(X)) = 1\ \forall\theta \in \Theta$ となるから，$R(\theta, \delta) = R(\theta, \delta^\pi)\ \forall\theta \in \Theta$ となる．よって，δ^π を優越する推定量は存在しない． □

Remark 3.11. $L(\theta, d) = (d - g(\theta))^2$ のときは，事前分布 π に対して，$E[g(\theta)^2] < \infty$ なら，事後平均 $\delta^\pi = E[\theta \mid X]$ は X の周辺分布について一意である．すなわち，η^π がもう 1 つの Bayes 推定量なら，X の周辺分布について，$P(\eta^\pi(X) = \delta^\pi(X)) = 1$ になる．よって，

$P(X \in A) = 0 \Rightarrow P_\theta(X \in A) = 0\ \forall\theta \in \Theta$ なら，δ^π は許容的である．十分統計量 T が存在するときは，T の関数からなる推定量のクラスのなかで許容的なら，すべての推定量のクラスのなかで許容的である (Rao-Blackwell の定理)．よって，$P(T \in A) = 0 \Rightarrow P_\theta(T \in A) = 0\ \forall\theta \in \Theta$ なら，$\delta^\pi(X) = E[\theta \mid T]$ は許容的である．

次に，もう 1 つの最適性の基準として，ミニマクス性を考察する．

Definition 6 (ミニマクス性)**.** 推定量 δ^* が

$$\sup_{\theta\in\Theta} R(\theta, \delta^*) = \inf_\delta \sup_{\theta\in\Theta} R(\theta, \delta)$$

をみたすとき，δ^* は ミニマクス (minimax) であるという．

ミニマクスな推定量は最悪のケースのリスクを最小化するものである．ミニマクス性の要請から，多くの場合，“決めうち” 推定量を排除できる．統計的決定理論の 1 つのゴールは，ミニマクスかつ許容的な推定量を構成することである．そこで，Bayes 推定量のミニマクス性を考察してみよう．

Theorem 3.10. π を Θ 上の事前分布とし，π に対して，δ^π を定数リスクをもつ Bayes 推定量とする．すなわち，$R(\theta, \delta^\pi)$ は θ によらず一定であるとする．このとき，δ^π はミニマクスである．

Proof. δ を任意の推定量とすると，$\forall\theta \in \Theta$ に対して，$R(\theta, \delta^\pi) = r(\pi, \delta^\pi) \le r(\pi, \delta) \le \sup_{\vartheta\in\Theta} R(\vartheta, \delta)$ である． □

Example 3.21 (Example 3.17 の続き)**.** 損失関数を $L(\theta, d) = (d - \theta)^2$ とする．Bayes 推定量 $\widehat{\theta}^{\alpha,\beta} = (T + \alpha)/(n + \alpha + \beta)$ のリスクは

$$R(\theta, \widehat{\theta}^{\alpha,\beta}) = E_\theta[(\widehat{\theta}^{\alpha,\beta} - \theta)^2] = \frac{n\theta(1-\theta) + \alpha^2 - 2\theta\alpha(\alpha+\beta) + \theta^2(\alpha+\beta)^2}{(n+\alpha+\beta)^2}$$

であって，右辺は $\alpha = \beta = \sqrt{n}/2$ のときに θ に依存しない．よって，$(T + \sqrt{n}/2)/(n + \sqrt{n})$ はミニマクスかつ許容的である．

Bayes 推定量でない推定量のミニマクス性を示すには，次の定理が便利である．

Theorem 3.11. π_N を Θ 上の事前分布の列とし，δ^{π_N} を π_N に対する Bayes 推定量とする．いま，推定量 δ^* が

$$\sup_{\theta\in\Theta} R(\theta, \delta^*) \le \limsup_N r(\pi_N, \delta^{\pi_N})$$

をみたすなら，δ^* はミニマクスである．

Proof. δ を任意の推定量とすると，

$$\sup_{\theta\in\Theta} R(\theta,\delta^*) \ge r(\pi_N,\delta) \ge r(\pi_N,\delta^{\pi_N})$$

であって，ここで，$\limsup_N r(\pi_N,\delta^{\pi_N}) \ge \sup_{\theta\in\Theta} R(\theta,\delta^*)$ より，定理の結論を得る． □

Example 3.22 (Example 3.18 の続き)．$\sigma^2 = 1$ として，損失関数を $L(\mu,d) = (d-\mu)^2$ とする．Bayes 推定量 $\widehat{\mu}^{\xi,\tau^2} = (n\overline{X} + \xi/\tau^2)/(n+1/\tau^2)$ は定数リスクをもちえない．標本平均 $\widehat{\mu} = \overline{X}$ のミニマクス性を示そう．$R(\theta,\widehat{\mu}) = 1/n$ であって，$\widehat{\mu}^{\xi,\tau^2}$ の Bayes リスクは

$$r(N(\xi,\tau^2),\widehat{\mu}^{\xi,\tau^2}) = \frac{1}{n+1/\tau^2}$$

である．そこで，$\tau^2 = \tau_N^2 \to \infty$ とすれば，右辺 $\to 1/n$ なので，$\widehat{\mu}$ のミニマクス性が示された．さらに，$\widehat{\mu}$ は許容的であることも知られている．

James-Stein 推定量

$X \sim N(\mu, I_k)$ に対して，平均ベクトル $\mu \in \mathbb{R}^k$ の推定を考える．損失関数は 2 乗損失 $L(\mu,d) = \|d-\mu\|^2, d \in \mathbb{R}^k$ を採用する．ここで，$\|x\| = \sqrt{x'x}$ である．このとき，X はミニマクスであることが，Example 3.22 とほぼ同様の証明からわかる．さらに，X は完備十分統計量なので，Lehmann-Scheffé の定理より，X は最良不偏推定量である[22]．しかし，$k \ge 3$ のとき，X は非許容的である ($k = 1, 2$ では許容的であることが知られている)．この結果は Stein (1956) による．そのあとに，James and Stein (1961) は

$$\widehat{\mu}^{JS} = \left(1 - \frac{k-2}{\|X\|^2}\right) X$$

という推定量が X を優越することを示した．$\widehat{\mu}^{JS}$ は James-Stein 推定量 と呼ばれる．

Remark 3.12. $X_1,\dots,X_n \sim N(\mu, I_k)$ i.i.d. に対して，$\sqrt{n}\overline{X} \sim N(\sqrt{n}\mu, I_d)$ は μ の十分統計量である．$\sqrt{n}\mu$ を μ におきなおせば，μ の推定については，$n = 1$ の場合を考察すれば十分である．

$\widehat{\mu}^{JS}$ が X を優越することを示そう．

Theorem 3.12 (James and Stein (1961))．$k \ge 3$ に対して，

$$E_\mu[\|\widehat{\mu}^{JS} - \mu\|^2] < k = E_\mu[\|X-\mu\|^2],\ \forall \mu \in \mathbb{R}^k.$$

この定理の証明は，次の Stein の等式が本質的である．

[22] 厳密には Lehmann-Scheffé の定理を $g(\theta)$ が 1 次元のときにしか証明していなかったが，$g(\theta)$ が多次元の場合でも Lehmann-Scheffé の定理の結論が成り立つことは証明から明らかである．

Lemma 3.2 (Stein の等式)**.** $Z \sim N(\mu, 1)$ とし，$g : \mathbb{R} \to \mathbb{R}$ を C^1 級関数であって，$E[|g'(Z)|] < \infty$ をみたすとする．このとき，$E[(Z-\mu)g(Z)] = E[g'(Z)]$ が成り立つ (左辺の期待値の存在も主張の一部である)．

Proof. $\mu = 0, g(0) = 0$ の場合に補題を示せば十分である．まず，$E[|Zg(Z)|] < \infty$ を示す．

$$E[|Zg(Z)|] = \int_0^\infty z|g(z)|\phi(z)dz + \int_0^\infty z|g(-z)|\phi(z)dz.$$

ここで，$z \geq 0$ に対して，

$$g(z) = \int_0^z g'(w)dw = \int_0^\infty g'(w)I(w \leq z)dw$$

であるから，

$$\begin{aligned}\int_0^\infty z|g(z)|\phi(z)dz &\leq \int_0^\infty \int_0^\infty z|g'(w)|I(w \leq z)\phi(z)dwdz \\ &= \int_0^\infty \underbrace{\left\{\int_w^\infty z\phi(z)dz\right\}}_{=\phi(w)} |g'(w)|dw = \int_0^\infty |g'(w)|\phi(w)dw.\end{aligned}$$

積分順序の交換は Fubini の定理から保証される．同様にして，

$$\int_0^\infty z|g(-z)|\phi(z)dz \leq \int_0^\infty |g'(-w)|\phi(w)dw = \int_{-\infty}^0 |g'(w)|\phi(w)dw$$

であるから，$E[|Zg(Z)|] \leq E[|g'(Z)|] < \infty$ を得る．また，同様の操作から，$E[Zg(Z)] = E[g'(Z)]$ も示される． □

Proof of Theorem 3.12. $g(X) = (k-2)/\|X\|^2$ とおくと，

$$\|\widehat{\mu}^{JS} - \mu\|^2 = \|(X-\mu) - g(X)X\|^2 = \|X-\mu\|^2 - 2g(X)(X-\mu)'X + (k-2)g(X)$$

と分解できる．ここで，極座標変換より，$E_\mu[1/\|X\|^2] < \infty$ が確かめられる．次に，$g(X)(X-\mu)'X = \sum_{j=1}^k (X_j - \mu_j)\{g(X)X_j\}$ であって，各 j に対して，$X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_k$ を条件付けて Stein の補題を適用する．ここで，

$$\frac{\partial}{\partial x_j}\{g(x)x_j\} = \frac{\partial g(x)}{\partial x_j}x_j + g(x) = \frac{-2(k-2)x_j^2}{\|x\|^4} + g(x) = g(x)\left\{1 - \frac{2x_j^2}{\|x\|^2}\right\}$$

であって，右辺に $x = X$ を代入したものは可積分である．よって，

$$E_\mu[g(X)(X-\mu)'X] = (k-2)E_\mu[g(X)]$$

を得る．以上より，

$$E_\mu[\|\widehat{\mu}^{JS} - \mu\|^2] = k - (k-2)E_\mu[g(X)] < k$$

が示された． □

$k \geq 3$ のとき，$\widehat{\mu}^{JS}$ はミニマクスな推定量 X を優越するから，$\widehat{\mu}^{JS}$ もミニマクスである．では $\widehat{\mu}^{JS}$ は許容的であるだろうか．実は $\widehat{\mu}^{JS}$ は非許容的である．$\widehat{\mu}^{JS}$ は X を原点方向に縮小することによって得られるが，$\|X\|^2 < k-2$ のときは X の符号まで逆転させてしまう．これは不合理だと考えられる．実際，James-Stein 推定量において $\|X\|^2 < k-2$ のときは 0 になるように修正した推定量

$$\widehat{\mu}^{JS+} = \left(1 - \frac{k-2}{\|X\|^2}\right)^+ X = \begin{cases} \left(1 - \frac{k-2}{\|X\|^2}\right) X & \text{if } \|X\|^2 > k-2 \\ 0 & \text{otherwise} \end{cases}$$

が $\widehat{\mu}^{JS}$ を優越することが知られている．$\widehat{\mu}^{JS+}$ は正部分 James-Stein 推定量 (positive part James-Stein estimator) と呼ばれる．実は正部分 James-Stein 推定量も非許容的であることが知られている．正部分 James-Stein 推定量を優越する許容的な推定量が存在するかどうかはいまのところ未解決問題である [23]．

[23] James-Stein 推定量を優越する許容的な推定量は Kubokawa (1991) によって与えられた．

4　検定

$\mathcal{X}$ を有限次元ユークリッド空間とし，$\varnothing \neq \Theta \subset \mathbb{R}^k$ をパラメータ空間として，$\{p(\cdot;\theta) : \theta \in \Theta\}$ を $\mathcal{X}$ 上のパラメトリックな分布族とする．$X_1, \ldots, X_n \sim p(\cdot;\theta)$ i.i.d. が与えられたとき，θ が既知の集合 $\Theta_0 \subsetneq \Theta$ に属しているかを決定したいとする．これを，$\theta \in \Theta_0$ という仮説に対して，$\theta \in \Theta_1 =: \Theta \setminus \Theta_0$ という対立する仮説に対して検定する問題とみなして，

$$H_0 : \theta \in \Theta_0 \text{ vs. } H_1 : \theta \in \Theta_1 \tag{*}$$

と記述する．H_0 を 帰無仮説 (null hypothesis) と呼び，H_1 を 対立仮説 (alternative hypothesis) と呼ぶ．さらに，Θ_0 が 1 点集合のとき ($\Theta_0 = \{\theta_0\}$)，H_0 を 単純帰無仮説 (simple null hypothesis) と呼び，そうでないとき，H_0 を 複合帰無仮説 (composite null hypothesis) と呼ぶ．単純対立仮説，複合対立仮説も同様に定義する．

$k = 1$ のときは，

$$H_0 : \theta = \theta_0 \text{ vs. } H_1 : \theta \neq \theta_0$$

という形の検定問題を 両側検定問題 (two-sided testing problem)，

$$H_0 : \theta \leq \theta_0 \text{ vs. } H_1 : \theta > \theta_0$$

という形の検定問題を 片側検定問題 (one-sided testing problem) と呼ぶ．

検定問題 (*) が与えられたとき，データ $X = (X_1', \ldots, X_n')'$ にもとづいて，H_0 を 棄却する (reject) かしないかを決める．ここで，H_0 を棄却する決定を $d = 1$ と表し，H_0 を棄却しない決定を $d = 0$ を表すとすると，検定とは，$\{0, 1\}$ に値をとる X の関数に他ならない．ただし，理論的には，次の確率的な決定も考えた方が都合がよい : $\gamma \in [0, 1]$ に対して，H_0 を確率 γ で棄却する．この決定を $d = \gamma$ と表すことにすれば，検定問題では，決定空間は $D = [0, 1]$ であって，関数 $\delta : \mathcal{X}^n \to [0, 1]$ を 検定関数 (test function) or 検定と呼ぶ．$\delta(X) = \gamma$ とは，X を与えたとき，確率 γ で H_0 を棄却する事象に対応している．$\{0, 1\}$ にしか値ととらない検定を 非確率化検定 (non-randomized test) と呼び，$(0, 1)$ にも値をとる検定を 確率化検定 (randomized test) と呼ぶ．

H_0 が正しいのに，H_0 を棄却してしまう誤りを タイプ I エラー と呼び，H_1 が正しいのに，H_0 を棄却しない誤りを タイプ II エラー と呼ぶ．標準的な検定理論では，タイプ I エラーとタイプ II エラーを対称に扱わず，タイプ I エラーをより重視して，タイプ I エラーの確率を与えられた確率 $\alpha \in [0, 1]$ 以下に押さえつつ，タイプ II エラーの確率をなるべく小さくすることを考える．

パラメータが θ のとき，検定 δ が H_0 を棄却する確率は

$$\beta_\delta(\theta) = E_\theta[\delta(X)] = E_\theta[P(\{H_0 \text{ が棄却される }\} \mid X)]$$

であるから，

$$\begin{aligned} P_\theta(\{\text{ タイプ I エラーが起こる }\}) &= \beta_\delta(\theta), \quad \theta \in \Theta_0, \\ P_\theta(\{\text{ タイプ II エラーが起こる }\}) &= 1 - \beta_\delta(\theta), \quad \theta \in \Theta_1 \end{aligned}$$

と表せる．$\beta_\delta(\theta)$ を 検出力関数 (power function) と呼ぶ．δ が非確率的なら，

$$\beta_\delta(\theta) = P_\theta\{\delta(X) = 1\}$$

である．検定問題のゴールは，

$$\beta_\delta(\theta) \le \alpha, \ \forall \theta \in \Theta_0 \tag{*3}$$

をみたしつつ，$\theta \in \Theta_1$ に対して $\beta_\delta(\theta)$ がなるべく 1 に近い検定 δ を構成することである．(*3) をみたす検定 δ を 水準 α (level α) の検定と呼ぶ．α の値は，0.05 や 0.01 が使われることが多い．また，

$$\sup_{\theta \in \Theta_0} \beta_\delta(\theta)$$

の値を δ の サイズ (size) と呼ぶ．検定が水準 α をもつためには，そのサイズが α 以下であればよい．サイズを厳密に α に等しくとる必要はないが，多くの場合，サイズが小さくなると H_1 のもとでの検出力が小さくなるので，サイズを α に一致させるように検定を選ぶのが合理的である．

Definition 7 (UMP 検定)**.** $\alpha \in (0,1)$ を所与とする．水準 α の検定 δ^* が一様最強力 (uniformly most powerful, UMP) であるとは，水準 α の任意の検定 δ に対して，

$$\beta_{\delta^*}(\theta) \ge \beta_\delta(\theta), \ \forall \theta \in \Theta_1$$

となることである．$\Theta_1 = \{\theta_1\}$ のときは，UMP 検定を単に最強力 (most powerful, MP) 検定という．

Remark 4.1. UMP 検定は望ましい検定であるが，制約がきついので，いくつかの簡単な場合に対しては存在するが，存在しない場合も多い．

なお，多くの場合，非確率化検定 δ は，ある 1 次元の統計量 $T = T(X)$ を用いて，

$$\delta(X) = I(T > c)$$

と表せる．このとき，T を 検定統計量 (text statistic) と呼び，c を 棄却点 (critical point) と呼ぶ．この検定を

$$T > c \Rightarrow \text{reject}$$

と記述する．棄却点 c は，サイズが α 以下になるように選ぶ：

$$\sup_{\theta \in \Theta_0} P_\theta(T > c) \le \alpha.$$

H_0 が単純仮説，すなわち，$\Theta_0 = \{\theta_0\}$ なら，T の H_0 のもとでの分布 (帰無分布と呼ぶ) は既知なので，その $(1-\alpha)$ 分位点を c に選べばよい．すなわち，$F(t) = P_{\theta_0}(T(X) \le t)$ として，

$$c = F^{\leftarrow}(1-\alpha)$$

とおくと，

$$P_{\theta_0}(T > c) = 1 - F(F^{\leftarrow}(1-\alpha)) \leq 1 - (1-\alpha) = \alpha$$

となる．さらに，F が連続なら，$P_{\theta_0}(T > c) = \alpha$ になる．

p 値

検定統計量 T が与えられたとき，

$$p(t) = \sup_{\theta \in \Theta_0} P_\theta(T \geq t)$$

とおく．このとき，$p(T)$ を T の $\underline{p\text{ 値}}$ (p-value) と呼ぶ．$p(T)$ は r.v. である．

Theorem 4.1. 各 $\theta \in \Theta_0$ のもとで T の d.f. が連続なら，所与の $\alpha \in (0,1)$ に対して，

$$p(T) \leq \alpha \Rightarrow \text{reject}$$

という検定は水準 α をもつ．

Proof. $\theta \in \Theta_0$ を任意に固定し，$S = -T, F(s) = P_\theta(S \leq s)$ とおくと，F は連続であって，$p(T) \geq F(S)$ である．このとき，θ のもとで，$F(S) \sim U(0,1)$ であるから，$P_\theta(p(T) \leq \alpha) \leq P_\theta(F(S) \leq \alpha) = \alpha$ を得る． □

Remark 4.2. 証明より，H_0 が単純なら ($\Theta_0 = \{\theta_0\}$)，$p(T) \sim U(0,1)$ であることがわかる．

水準 α の検定は自動的に水準 $\alpha' > \alpha$ の検定でもあるから，設定する水準を小さくしていくと，帰無仮説は棄却されにくくなる．従って，p 値が小さい場合，それは帰無仮説が正しくないことに対する強い証拠になっているとされる．

4.1 Neyman-Pearson の補題

まず，H_0 と H_1 がともに単純な場合を考える．この場合には，MP 検定が存在する．いま，$\Theta_0 = \{\theta_0\}, \Theta_1 = \{\theta_1\}, \theta_0 \neq \theta_1$ とし，検定問題

$$H_0 : \theta = \theta_0 \text{ vs. } H_1 : \theta = \theta_1 \tag{*}$$

を考える．$X = (X_1', \ldots, X_n')'$ の同時確率 (密度) 関数を $p_n(x;\theta) = \prod_{i=1}^n p(x_i;\theta)$ とおく．

Theorem 4.2. (Neyman-Pearson の補題). 任意の $\alpha \in (0,1)$ に対して，ある定数 $c \geq 0, \gamma \in [0,1]$ が存在して，

$$\delta_{c,\gamma}(x) = \begin{cases} 1 & \text{if } p_n(x;\theta_1) > cp_n(x;\theta_0) \\ \gamma & \text{if } p_n(x;\theta_1) = cp_n(x;\theta_0) \\ 0 & \text{if } p_n(x;\theta_1) < cp_n(x;\theta_0) \end{cases}$$

はサイズ α の検定になる．さらに，$\delta_{c,\gamma}$ は水準 α の MP 検定である．

Proof. $T = T(X) = \frac{p_n(X;\theta_1)}{p_n(X;\theta_0)}$ とおく．ここで，$0/0 = 0, a/0 = \infty$ $(a > 0)$ とみなす．$P_{\theta_0}(p_n(X;\theta_0) = 0) = 0$ より，$P_{\theta_0}(T < \infty) = 1$ である．そこで，θ_0 のもとでの T の $(1-\alpha)$ 分位点を c とおくと，

$$\alpha - P_{\theta_0}(T = c) \leq P_{\theta_0}(T > c) \leq \alpha$$

となる．$P_{\theta_0}(T > c) = \alpha$ なら $\gamma = 0$ とおき，$P_{\theta_0}(T > c) < \alpha$ なら，

$$\gamma = \frac{\alpha - P_{\theta_0}(T > c)}{P_{\theta_0}(T = c)}$$

とおくと，$\gamma \in [0, 1]$ であって，$\delta_{c,\gamma}$ のサイズは，

$$E_{\theta_0}[\delta_{c,\gamma}(X)] = P_{\theta_0}(T > c) + \gamma P_{\theta_0}(T = c) = \alpha$$

になる．

次に，$\delta_{c,\gamma}$ が MP 検定であることを示す．δ を水準 α の任意の検定とすると，

$$\{\delta_{c,\gamma}(x) - \delta(x)\}\{p_n(x;\theta_1) - cp_n(x;\theta_0)\} \geq 0.$$

よって，これを展開して，

$$E_{\theta_1}[\delta_{c,\gamma}(X) - \delta(X)] \geq cE_{\theta_0}[\delta_{c,\gamma}(X) - \delta(X)] \geq c(\alpha - \alpha) = 0$$

を得る．これから，$\beta_{\delta_{c,\gamma}}(\theta_1) \geq \beta_\delta(\theta_1)$ を得る． □

Remark 4.3. $P_{\theta_0}(p_n(X;\theta_1) = cp_n(X;\theta_0)) = 0$ なら，確率化の必要はなく，

$$\delta_c(x) = \begin{cases} 1 & \text{if } p_n(x;\theta_1) > cp_n(x;\theta_0) \\ 0 & \text{if } p_n(x;\theta_1) < cp_n(x;\theta_0) \end{cases}$$

という形の検定を考えれば十分である．このとき，c は

$$T(X) = \frac{p_n(X;\theta_1)}{p_n(X;\theta_0)}$$

の θ_0 のもとでの $(1-\alpha)$ 分位点をとればよい．$T(X)$ を <u>尤度比</u> (likelihood ratio, LR) と呼ぶ．

ところで，データに関係なく確率 α で $\theta = \theta_0$ を棄却する検定 $(\delta(x) \equiv \alpha)$ はサイズ α をもつが，明らかに不合理である．よって，合理的な検定は $\theta = \theta_1$ のもとで α より大きい検出力をもつべきである．もちろん，$p_n(\cdot;\theta_0)$ と $p_n(\cdot;\theta_1)$ が同じ分布なら，データにもとづいて $\theta = \theta_0$ と $\theta = \theta_1$ を区別するすべはないが，そうでなければ，MP 検定は $\theta = \theta_1$ のとき α より大きい検出力をもつ．

Corollary 4.1. (*) に対する水準 α の MP 検定の $\theta = \theta_1$ のときの検出力を β とおく．このとき，$p_n(\cdot;\theta_0)$ と $p_n(\cdot;\theta_1)$ が相異なる分布なら，$\beta > \alpha$ である．

Remark 4.4. 言い換えると，$p_n(\cdot;\theta_0)$ と $p_n(\cdot;\theta_1)$ が相異なる分布なら，

$$\sup\{\beta_\delta(\theta_1) : \beta_\delta(\theta_0) = \alpha\} > \alpha$$

が成り立つ．

Proof. $\delta_{c,\gamma}$ を Neyman-Pearson の補題に現れる検定とすると，$\beta = E_{\theta_1}[\delta_{c,\gamma}(X)]$ である．このとき，

$$\{\delta_{c,\gamma}(x) - \alpha\}\{p_n(x;\theta_1) - cp_n(c;\theta_0)\} \geq \min\{1-\alpha, \alpha\}|p_n(x;\theta_1) - cp_n(x;\theta_0)|$$

であって，$p_n(\cdot;\theta_0)$ と $p_n(\cdot;\theta_1)$ は相異なる分布であるから，右辺の x に関する積分 (和) は 0 にならない．よって，

$$\beta - \alpha = E_{\theta_1}[\delta_{c,\gamma}(X) - \alpha] > cE_{\theta_0}[\delta_{c,\gamma}(X) - \alpha] = 0$$

を得る． □

Remark 4.5 (MP 検定の一意性)**.** $S = \{x : p_n(x;\theta_1) \neq cp_n(x;\theta_0)\}$ とおくと，δ が水準 α の MP 検定なら，"ほとんどすべての"$x \in S$ に対して，$\delta(x) = \delta_{c,\gamma}(x)$ になる ($p_n(x;\theta)$ が確率関数なら，すべての $x \in S$ に対して，$\delta(x) = \delta_{c,\gamma}(x)$ になる)．

Example 4.1 (正規分布)**.** $\theta_0 < \theta_1, \theta \in \{\theta_0, \theta_1\}$ に対して，$X_1, \ldots, X_n \sim N(\theta, 1)$ i.i.d. として，検定問題

$$H_0 : \theta = \theta_0 \text{ vs. } H_1 : \theta = \theta_1$$

を考える．このとき，尤度比は

$$\frac{p_n(X;\theta_1)}{p_n(X;\theta_0)} = \exp\{(\theta_1 - \theta_0)n\overline{X} - (n/2)(\theta_1^2 - \theta_0^2)\}$$

であって，右辺は $\overline{X}$ について狭義単調増加である．ここで，$\overline{X}$ は連続型なので，

$$\overline{X} > c \Rightarrow \text{reject}$$

という形の検定が MP 検定になる．このとき，$\theta = \theta_0$ のもとで，$\overline{X} \sim N(\theta_0, 1/n)$, i.e., $\sqrt{n}(\overline{X} - \theta_0) \sim N(0,1)$ であるから，

$$P_{\theta_0}(\overline{X} > c) = P_{\theta_0}(\sqrt{n}(\overline{X} - \theta_0) > \sqrt{n}(c - \theta_0)) = 1 - \Phi(\sqrt{n}(c - \theta_0))$$

となる (Φ は $N(0,1)$ の d.f. である)．右辺が α に等しくなるのは，

$$1 - \Phi(\sqrt{n}(c - \theta_0)) = \alpha \Leftrightarrow c = \theta_0 + \frac{\Phi^{-1}(1-\alpha)}{\sqrt{n}}$$

のときである[24]．よって，$z_\alpha = \Phi^{-1}(1-\alpha)$ とおくと，

$$\overline{X} > \theta_0 + \frac{z_\alpha}{\sqrt{n}} \Rightarrow \text{reject}$$

が水準 α の MP 検定である．

Example 4.2 (2 項分布)**.** $X \sim Bin(n,\theta), \theta \in \{\theta_0, \theta_1\}, \theta_0 < \theta_1$ に対して，

$$H_0 : \theta = \theta_0 \text{ vs. } H_1 : \theta = \theta_1$$

という検定問題を考える．このとき，尤度比は

$$\frac{p(X;\theta_1)}{p(X;\theta_0)} = \frac{\theta_1^X(1-\theta_1)^{n-X}}{\theta_0^X(1-\theta_0)^{n-X}} \propto \left(\frac{r_1}{r_0}\right)^X$$

である．ここで，$r_i = \theta_i/(1-\theta_i), i = 0, 1$．右辺は X について狭義単調増加であるから，

$$\delta_{k,\gamma}(X) = \begin{cases} 1 & \text{if } X > k \\ \gamma & \text{if } X = k \\ 0 & \text{if } X < k \end{cases}$$

という形の検定が MP 検定になる．$Bin(n,\theta)$ の d.f. を $F_{n,\theta}$ とおくと，k, γ を

$$k = F_{n,\theta_0}^{\leftarrow}(1-\alpha),\ \gamma = \frac{F_{n,\theta_0}(k) - (1-\alpha)}{F_{n,\theta_0}(k) - F_{n,\theta_0}(k-)}$$

と選べば，$\delta_{k,\gamma}$ はサイズ α になる．

Remark 4.6. 上の 2 つの例において，MP 検定は θ_1 に依存しないので，

$$H_0 : \theta = \theta_0 \text{ vs. } H_1 : \theta > \theta_0$$

という検定問題に対する水準 α の UMP 検定にもなっている．

次に，Θ を $\mathbb{R}$ の区間とし，右端点でない与えられた $\theta_0 \in \Theta$ に対して，

$$H_0 : \theta \leq \theta_0 \text{ vs. } H_1 : \theta > \theta_0 \tag{**}$$

という検定問題を考える．この検定問題に対しては，あるクラスの分布族に対しては UMP 検定が存在する．

Definition 8 (単調尤度比)**.** 分布族 $\{p_n(\cdot;\theta) : \theta \in \Theta\}$ が 1 次元の統計量 T について<u>単調尤度比</u> (monotone likelihood ratio, MLR) をもつとは，任意の $\theta_1 < \theta_2$ に対して，ある非減少関数 $g_{\theta_1,\theta_2} : \mathbb{R} \to \mathbb{R}_+$ が存在して，

$$p_n(x;\theta_2) = g_{\theta_1,\theta_2}(T(x)) p_n(x;\theta_1)$$

と表せることをいう．

[24] Φ は連続かつ狭義単調増加なので，$(0,1)$ 上に定義された逆関数 Φ^{-1} が存在する．

Example 4.3. $\{p(\cdot;\theta):\theta\in\Theta\}$ が 1 パラメータの指数型分布族をなすとき，

$$p_n(x;\theta)=H(x)\exp\{\xi(\theta)T(x)-C(\theta)\},\ x\in\mathcal{X}^n$$

と表せる $(\xi(\theta),T(x)\in\mathbb{R}$ である$)$．$\xi(\theta)$ が θ について非減少なら，$\theta_1<\theta_2$ に対して，

$$\frac{p_n(x;\theta_2)}{p_n(x;\theta_1)}\propto\exp\{(\xi(\theta_2)-\xi(\theta_1))T(x)\}$$

となるから，$\{p_n(\cdot;\theta):\theta\in\Theta\}$ は T について MLR をもつ．

Theorem 4.3. $\{p_n(\cdot;\theta):\theta\in\Theta\}$ は T について MLR をもつとする．このとき，

$$\delta_{c,\gamma}(x)=\begin{cases}1 & \text{if } T(x)>c\\ \gamma & \text{if } T(x)=c\\ 0 & \text{if } T(x)<c\end{cases}$$

という形の検定は水準 α の UMP 検定である．ここで，$c\in\mathbb{R},\gamma\in[0,1]$ は

$$E_{\theta_0}[\delta_{c,\gamma}(X)]=\alpha$$

をみたすように選ぶ．

Proof. $E_{\theta_0}[\delta_{c,\gamma}(X)]=\alpha$ をみたす $c\in\mathbb{R},\gamma\in[0,1]$ が存在することの証明は，Neyman-Pearson の補題の証明と同様である．

次に，$\delta_{c,\gamma}$ が (**) に対する UMP 検定であることを示す．そのために，$\beta_{\delta_{c,\gamma}}(\theta)=E_\theta[\delta_{c,\gamma}(X)]$ が θ について非減少であることを示す．$\theta_1<\theta_2$ なる $\theta_1,\theta_2\in\Theta$ を任意に固定して，

$$H_0:\theta=\theta_1\ \text{vs.}\ H_1:\theta=\theta_2 \tag{*3}$$

という検定問題を考える．$\alpha_1=\beta_{\delta_{c,\gamma}}(\theta_1),\kappa=g_{\theta_1,\theta_2}(c)$ とおくと，(*3) に対する水準 α_1 の任意の検定 δ に対して，

$$\{\delta_{c,\gamma}(x)-\delta(x)\}\{p_n(x;\theta_2)-\kappa p_n(x;\theta_1)\}\geq 0$$

となるから，

$$E_{\theta_2}[\delta_{c,\gamma}(X)-\delta(X)]\geq\kappa E_{\theta_1}[\delta_{c,\gamma}(X)-\delta(X)]\geq\kappa(\alpha_1-\alpha_1)=0$$

を得る．よって，$\delta_{c,\gamma}$ は (*3) に対する水準 α_1 の MP 検定である．$\delta(x)\equiv\alpha_1$ という検定と比較して，$\beta_{\delta_{c,\gamma}}(\theta_1)=\alpha_1\leq\beta_{\delta_{c,\gamma}}(\theta_2)$ を得る．

さらに，(*3) において $\theta_1=\theta_0$ とすれば，$\delta_{c,\gamma}$ は水準 α の MP 検定である．$\delta_{c,\gamma}$ は θ_2 に依存しないので，

$$H_0:\theta=\theta_0\ \text{vs.}\ H_1:\theta>\theta_0 \tag{*4}$$

という検定問題に対する UMP 検定になっていることもわかる．$\theta\leq\theta_0$ に対しては，$\beta_{\delta_{c,\gamma}}(\theta)\leq\beta_{\delta_{c,\gamma}}(\theta_0)=\alpha$ になるから，$\delta_{c,\gamma}$ は (**) に対する水準 α の検定である．(**) に対する水準 α の検定は (*4) に対しても水準 α をもつので，$\delta_{c,\gamma}$ は (**) に対する水準 α の UMP 検定である．□

証明からわかるように，Theorem 4.3 の $\delta_{c,\gamma}$ に対して，$\beta_{\delta_{c,\gamma}}(\theta)$ は θ について非減少になる．Neyman-Pearson の補題の系より，さらに次のことがわかる．分布族 $\{p_n(\cdot;\theta):\theta\in\Theta\}$ が 識別可能 (identifiable) であるとは，$\theta_1\neq\theta_2$ なる $\theta_1,\theta_2\in\Theta$ に対して，$p_n(\cdot;\theta_1)$ と $p_n(\cdot;\theta_2)$ が相異なる分布であることをいう．

Corollary 4.2. $\{p_n(\cdot;\theta):\theta\in\Theta\}$ が識別可能なら，Theorem 4.3 の $\delta_{c,\gamma}$ に対して，$\beta_{\delta_{c,\gamma}}(\theta)$ は $0<\beta_{\delta_{c,\gamma}}(\theta)<1$ の範囲で θ について狭義単調増加である．

Example 4.4. Examples 4.1 & 4.2 の検定は，それぞれ (*) に対する UMP 検定でもある．

4.2 不偏検定

本節では，両側検定問題

$$H_0:\theta=\theta_0 \text{ vs. } H_1:\theta\neq\theta_0$$

を考える．両側検定問題に対しては，ほとんどの場合，UMP 検定は存在しない．

Example 4.5. $X\sim N(\theta,1)$ に対して，

$$H_0:\theta=0 \text{ vs. } H_1:\theta\neq 0 \tag{*}$$

という検定問題を考えよう．このとき，仮に δ^* が水準 α の UMP 検定なら，任意の $\theta_1\neq 0$ に対して，δ^* は

$$H_0:\theta=0 \text{ vs. } H_1:\theta=\theta_1 \tag{**}$$

という検定問題に対する MP 検定になっている．しかし，Neyman-Pearson の補題より，$\theta_1>0$ のときは，$I(x>c)$ なる検定が (**) に対する MP 検定であって，$\theta_1<0$ のときは，$I(x<c)$ なる検定が (**) に対する MP 検定である．MP 検定の一意性を認めると，(*) に対する UMP 検定が存在しないことがわかる．

そこで両側検定問題については，検定のクラスを制限する．

Definition 9 (不偏検定と UMPU 検定)**.** 検定問題 $H_0:\theta\in\Theta_0$ vs. $H_1:\theta\in\Theta_1$ に対する水準 α の検定 δ が 不偏 (unbiased) であるとは，

$$\beta_\delta(\theta)=E_\theta[\delta(X)]\geq\alpha,\ \forall\theta\in\Theta_1$$

となることをいう．

さらに，水準 α の不偏検定 δ^* が 一様最強力不偏検定 (uniformly most powerful unbiased text, UMPU 検定) であるとは，水準 α の任意の不偏検定 δ に対して，

$$\beta_{\delta^*}(\theta)\geq\beta_\delta(\theta),\ \forall\theta\in\Theta$$

となることをいう．

いま，Θ は $\mathbb{R}$ の開区間とし，$X = (X_1', \ldots, X_n')'$ の同時確率 (密度) 関数は

$$p_n(x;\theta) = H(x)\exp\{\theta T(x) - C(\theta)\}$$

の形に表せるとしよう．

Theorem 4.4. 次の形の検定は $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$ に対する水準 α の UMPU 検定になる：

$$\delta^*(x) = \begin{cases} 1 & \text{if } T(x) < c_1 \text{ or } T(x) > c_2 \\ \gamma_i & \text{if } T(x) = c_i, i = 1, 2 \\ 0 & \text{otherwise} \end{cases}.$$

ただし，$c_i, \gamma_i, i = 1, 2$ は

$$E_{\theta_0}[\delta^*(X)] = \alpha,\ E_{\theta_0}[T(X)\delta^*(X)] = E_{\theta_0}[T(X)]\alpha \tag{*3}$$

をみたすように選ぶ．

Proof. 略証のみ与える．この証明では，$p_n(x;\theta)$ は密度関数とし，微分と積分の順序交換を自由に行う (ちゃんと正当化できる)．さらに，(*3) をみたす $c_i, \gamma_i, i = 1, 2$ の存在は認めて，δ^* が UMPU 検定であることを示す．δ を水準 α の任意の不偏検定とすると，

$$\beta_\delta(\theta_0) \leq \alpha, \quad \beta_\delta(\theta) \geq \alpha,\ \forall\theta \neq \theta_0$$

となる．ここで，$\beta_\delta(\theta) = E_\theta[\delta(X)]$ は θ について微分可能であることが示せる．$\beta_\delta(\theta)$ の連続性より，$\beta_\delta(\theta_0) = \alpha$ であって，$\beta_\delta(\theta)$ は $\theta = \theta_0$ で最小になる．よって，

$$\beta_\delta(\theta_0) = \alpha,\ 0 = \beta_\delta'(\theta_0) = \int \delta(x)\frac{\partial}{\partial\theta}p_n(x;\theta_0)dx = E_{\theta_0}[T(X)\delta(X)] - C'(\theta_0)\alpha$$

となる．ここで，$e^{C(\theta)} = \int e^{\theta T(x)}H(x)dx$ の両辺を θ について微分して，$C'(\theta) = E_\theta[T(X)]$ を得る．従って，

$$\int \delta(x)\frac{\partial}{\partial\theta}p_n(x;\theta_0)dx = 0 \Leftrightarrow E_{\theta_0}[T(X)\delta(X)] = E_{\theta_0}[T(X)]\alpha$$

である．

次に，$\theta_1 \neq \theta_0$ を任意に固定して，$a_1, a_2 \in \mathbb{R}$ に対して，

$$\begin{aligned} r_{a_1,a_2}(x) &= p_n(x;\theta_1) - a_1 p_n(x;\theta_0) - a_2\frac{\partial}{\partial\theta}p_n(x;\theta_0) \\ &= p_n(x;\theta_0)e^{C(\theta_0)-C(\theta_1)}\{e^{(\theta_1-\theta_0)T(x)} - \widetilde{a}_1 - \widetilde{a}_2 T(x)\} \end{aligned}$$

とおく．ここで，$\widetilde{a}_1 = e^{C(\theta_1)-C(\theta_0)}\{a_1 - C'(\theta_0)\}, \widetilde{a}_2 = e^{C(\theta_1)-C(\theta_0)}a_2$ である．c_1, c_2 に対して，a_1, a_2 を $c_1, c_2, \theta_0, \theta_1$ に依存させて適当に選べば，$H(x) > 0$ なる x に対して，

$$\delta^*(x) = 1 \Leftrightarrow r_{a_1,a_2}(x) > 0, \quad \delta^*(x) = 0 \Leftrightarrow r_{a_1,a_2}(x) < 0$$

が成り立つ．よって，

$$\{\delta^*(x)-\delta(x)\}r_{a_1,a_2}(x)\geq 0$$

となるから，両辺を積分して，

$$\begin{aligned}&\beta_{\delta^*}(\theta_1)-\beta_\delta(\theta_1)=\int\{\delta^*(x)-\delta(x)\}p_n(x;\theta_1)dx\\&\geq a_1\int\{\delta^*(x)-\delta(x)\}p_n(x;\theta_0)dx+a_2\int\{\delta^*(x)-\delta(x)\}\frac{\partial}{\partial\theta}p_n(x;\theta_0)dx=0\end{aligned}$$

を得る．さらに，$\delta(x)\equiv\alpha$ なる検定と比較して，$\beta_{\delta^*}(\theta_1)\geq\alpha$ を得る．$\theta_1\neq\theta_0$ は任意だったから，δ^* が UMPU 検定であることが示された． □

Example 4.6. $X_1,\ldots,X_n\sim N(\theta,1)$ i.i.d. に対して，

$$H_0:\theta=\theta_0 \text{ vs. } H_1:\theta\neq\theta_0$$

という検定問題を考える．このとき，$T(X)=n\overline{X}$ であって，$\overline{X}$ は連続型なので，

$$\delta(X)=I(\overline{X}<c_1 \text{ or } \overline{X}>c_2)$$

という形の検定が UMPU 検定になる．$\theta=\theta_0$ のもとで $\overline{X}-\theta_0\sim N(0,1/n)$ であるから，$z_{\alpha/2}=\Phi^{-1}(1-\alpha/2)$ とおいて，$c_1=\theta_0-z_{\alpha/2}/\sqrt{n},c_2=\theta_0+z_{\alpha/2}/\sqrt{n}$ にとれば，

$$\begin{aligned}&E_{\theta_0}[\delta(X)]=P_{\theta_0}(|\overline{X}-\theta_0|>z_{\alpha/2}/\sqrt{n})=\alpha,\\&E_{\theta_0}[T(X)\delta(X)]=n\underbrace{E_{\theta_0}[(\overline{X}-\theta_0)I(|\overline{X}-\theta_0|>z_{\alpha/2}/\sqrt{n})]}_{=0}+n\theta_0E_{\theta_0}[\delta(X)]=n\theta_0\alpha\end{aligned}$$

となる．以上より，

$$\sqrt{n}|\overline{X}-\theta_0|>z_{\alpha/2}\Rightarrow\text{reject}$$

という検定が水準 α の UMPU 検定になる．

Example 4.7. $X_1,\ldots,X_n\sim N(0,\sigma^2)$ i.i.d. に対して，

$$H_0:\sigma^2=\sigma_0^2 \text{ vs. } H_1:\sigma^2\neq\sigma_0^2$$

という検定問題を考える．このとき，

$$\begin{aligned}&p_n(x;\sigma^2)=(2\pi\psi^{-1})^{-n/2}e^{-\psi\sum_{i=1}^n x_i^2/2}=e^{\psi T(x)-(n/2)\log(2\pi\psi^{-1})},\\&\psi=1/\sigma^2,T(x)=-\sum_{i=1}^n x_i^2/2\end{aligned}$$

であって，検定問題は $\psi_0=1/\sigma_0^2$ とおくと，

$$H_0:\psi=\psi_0 \text{ vs. } H_1:\psi\neq\psi_0$$

と等価である．よって，$W=\sum_{i=1}^n X_i^2/\sigma_0^2$ とおくと，W は連続型なので，

$$\delta(X)=I(W<c_1 \text{ or } W>c_2)$$

という形の検定が UMPU 検定になる．c_1, c_2 は

$$E_{\psi=\psi_0}[\delta(X)]=\alpha,\ E_{\psi=\psi_0}[T(X)\delta(X)]=E_{\psi=\psi_0}[T(X)]\alpha$$

をみたすように選ぶ．この条件は

$$P_{\psi=\psi_0}(c_1\le W\le c_2)=1-\alpha,\ E_{\psi=\psi_0}[WI(c_1\le W\le c_2)]=n(1-\alpha)$$

と等価である．$\psi=\psi_0$ のとき，$W\sim\chi^2(n)$ であるから，その密度関数は

$$f_n(w)=\frac{1}{\Gamma(n/2)2^{n/2}}w^{n/2-1}e^{-w/2}$$

であって，

$$\begin{aligned}E_{\psi=\psi_0}[WI(c_1\le W\le c_2)]&=\frac{1}{\Gamma(n/2)2^{n/2}}\int_{c_1}^{c_2}w^{n/2}\underbrace{e^{-w/2}}_{=(-2e^{-w/2})'}dw\\&=-2\{c_2f_n(c_2)-c_1f_n(c_1)\}+n\int_{c_1}^{c_2}f_n(w)dw\end{aligned}$$

となる．以上より，c_1, c_2 は

$$\int_{c_1}^{c_2}f_n(w)dw=1-\alpha,\ c_1f_n(c_1)=c_2f_n(c_2)$$

で与えられる．

この検定は理論的には望ましいが，ちょっと面倒である．より簡便な検定は，$\chi^2(n)$ の $(1-\alpha)$ 分位点を $\chi^2_\alpha(n)$ とおくと，

$$W<\chi^2_{(1-\alpha)/2}(n) \text{ or } W>\chi^2_{\alpha/2}(n)\Rightarrow \text{reject}$$

という検定である．この検定はサイズ α であるが，不偏ではない．

Example 4.8. $X_n\sim Bin(n,\theta)$ に対して，

$$H_0:\theta=\theta_0 \text{ vs. } H_1:\theta\ne\theta_0$$

という検定問題を考える．X_n は離散型なので，UMPU 検定は確率化検定になる．これは面倒なので，通常は次のような簡便な検定が用いられる．$\theta=\theta_0$ のもとで，$n\to\infty$ のとき，CLT より，

$$\frac{\sqrt{n}(X_n/n-\theta_0)}{\sqrt{\theta_0(1-\theta_0)}}\xrightarrow{d}N(0,1)$$

となるから，Pólya の定理より，

$$\frac{\sqrt{n}|X_n/n-\theta_0|}{\sqrt{\theta_0(1-\theta_0)}} > z_{\alpha/2} \Rightarrow \text{reject}$$

は近似的にサイズ α の検定になる．この検定は不偏でもないし，UMPU でもないが，合理的な検定といえる．

Example 4.9. $(\mu,\sigma^2) \in \mathbb{R}\times(0,\infty)$ を未知として，$X_1,\ldots,X_n \sim N(\mu,\sigma^2)$ i.i.d. とする．このとき，

$$H_0 : \mu = \mu_0 \text{ vs. } H_1 : \mu \neq \mu_0$$

という検定問題を考える．この検定問題は，正確には，

$$H_0 : \mu = \mu_0, \sigma^2 > 0 \text{ vs. } H_1 : \mu \neq \mu_0, \sigma^2 > 0$$

であるから，H_0 は複合仮説である．σ^2 のように，未知だが検定問題にとってさしあたり興味のないパラメータを局外パラメータ (nuisance parameter) と呼ぶ．この検定問題に対しては，t 統計量を

$$T = T(X) = \frac{\sqrt{n}(\overline{X}-\mu_0)}{S}$$

とおくと，

$$|T| > c \Rightarrow \text{reject}$$

という検定が UMPU 検定になる．$\mu=\mu_0$ のもとでは，$T\sim t(n-1)$ であるから，t 分布の対称性より，$t(n-1)$ の $(1-\alpha/2)$ 分位点を $t_{\alpha/2}(n-1)$ とおくと，

$$|T| > t_{\alpha/2}(n-1) \Rightarrow \text{reject}$$

が水準 α の UMPU 検定になる．この証明は入り組んでいるので，省略する．

4.3 最尤法にもとづく検定

まず，$\Theta\subset\mathbb{R}$ として，

$$H_0 : \theta = \theta_0 \text{ vs. } H_1 : \theta \neq \theta_0.$$

という検定問題を考える．$\widehat{\theta}$ を MLE とし，$p(\cdot;\theta)$ の Fisher 情報量を $I(\theta)$ とおく．このとき，$\theta=\theta_0$ が正しければ，いくつかの正則条件のもとで，

$$\sqrt{n}(\widehat{\theta}-\theta_0) \xrightarrow{d} N(0,1/I(\theta_0))$$

となり，$\theta>\theta_0$ であれば，

$$\sqrt{n}(\widehat{\theta}-\theta_0) = \underbrace{\sqrt{n}(\widehat{\theta}-\theta)}_{\xrightarrow{d} N(0,1/I(\theta))} + \sqrt{n}(\theta-\theta_0) \xrightarrow{P} \infty$$

となる[25]．同様に，$\theta < \theta_0$ であれば，$\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{P} -\infty$ となる．よって，$z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2)$ とおくと，

$$\sqrt{nI(\theta_0)}|\widehat{\theta} - \theta_0| > z_{\alpha/2} \Rightarrow \text{reject}$$

なる検定は近似的にサイズ α をもち，$\theta \neq \theta_0$ なら検出力は $n \to \infty$ のとき 1 に近づく．従って，この検定は合理的な検定といえる．対立仮説が，

$$H_1 : \theta > \theta_0$$

ならば，

$$\sqrt{nI(\theta_0)}(\widehat{\theta} - \theta_0) > z_\alpha \Rightarrow \text{reject}$$

とすればよい．この方法はややアドホックであるが，汎用的である．

もっと一般に，Θ が多次元のとき，

$$H_0 : \theta \in \Theta_0 \text{ vs. } H_1 : \theta \in \Theta_1$$

という検定問題を考える．このとき，

$$\Lambda_n = \frac{\sup_{\theta \in \Theta} p_n(X; \theta)}{\sup_{\theta \in \Theta_0} p_n(X; \theta)}$$

を 尤度比検定統計量 (likelihood ratio test statistic, LRT 統計量) と呼び，

$$\Lambda_n > c \Rightarrow \text{reject}$$

なる検定を 尤度比検定 (likelihood ratio test, LR 検定) と呼ぶ．棄却点 c は次のように選べばよい．Θ を $\mathbb{R}^k$ の開集合とし，Θ_0 において "自由に動けるパラメータの数" を $r < k$ とする．このとき，いくつかの正則条件を仮定すると，H_0 のもとで，

$$2 \log \Lambda_n \xrightarrow{d} \chi^2(k - r)$$

が成り立つ．この結果は Wilks の定理と総称される．よって，$\chi^2_\alpha(k - r)$ を $\chi^2(k - r)$ 分布の $(1 - \alpha)$ 分位点とおくと，

$$2 \log \Lambda_n > \chi^2_\alpha(k - r) \Rightarrow \text{reject}$$

は近似的に水準 α をもつ．

[25] $Y_n \xrightarrow{P} \infty$ とは，任意の $M > 0$ に対して，$P(Y_n > M) \to 1$ となることをいう．

4.4 多項分布に対する検定

$Y = (Y_1, \dots, Y_k)' \sim Mn(n, p_1, \dots, p_k)$ とし，$(p_1, \dots, p_{k-1})'$ のパラメータ空間を $\Delta = \{(p_1, \dots, p_{k-1})' : \sum_{j=1}^{k-1} p_j < 1, p_j > 0, j = 1, \dots, k-1\}$ とする．ただし，$p_k = 1 - \sum_{j=1}^{k-1} p_j$ である．$(p_{0,1}, \dots, p_{0,k-1})' \in \Delta$ が与えられたとき，

$$H_0 : p_j = p_{0,j} \ 1 \le \forall j \le k-1 \text{ vs. } H_1 : p_j \ne p_{0,j} \ 1 \le \exists j \le k-1$$

という検定問題を考える．Y の確率関数は

$$P(Y_1 = y_1, \dots, Y_k = y_k) = \frac{n!}{y_1! \cdots y_k!} p_1^{y_1} \cdots p_k^{y_k}, \quad \sum_{j=1}^{k} y_j = n$$

である．よって，$(p_1, \dots, p_{k-1})'$ の MLE は

$$\widehat{p}_1 = Y_1/n, \dots, \widehat{p}_{k-1} = Y_{k-1}/n$$

であって，$p_{0,k} = 1 - \sum_{j=1}^{k-1} p_{0,j}, \widehat{p}_k = 1 - \sum_{j=1}^{k-1} p_j = Y_k/n$ とおくと，LRT 統計量は

$$\Lambda_n = (\widehat{p}_1/p_{0,1})^{Y_1} \cdots (\widehat{p}_k/p_{0,k})^{Y_k}$$

である．ここで，Λ_n は $X_1, \dots, X_n \sim Mn(1, p_1, \dots, p_k)$ i.i.d. にもとづく LRT 統計量とみなせる (なせか) ので，H_0 のもとで，

$$2 \log \Lambda_n \xrightarrow{d} \chi^2(k-1)$$

となる．よって，

$$2 \log \Lambda_n > \chi^2_\alpha(k-1) \Rightarrow \text{reject}$$

という検定は近似的にサイズ α をもつ．

LRT 統計量のほかに，次の Pearson の χ^2 検定統計量

$$\chi^2_n = \sum_{j=1}^{k} \frac{(Y_j - np_{0,j})^2}{np_{0,j}}$$

もよく使われる．ここで，

$$\chi^2_n > c \Rightarrow \text{reject}$$

とする．H_0 のもとで，χ^2_n と $2 \log \Lambda_n$ が漸近的に同等であることを示そう．この結果より，H_0 のもとで，$\chi^2_n \xrightarrow{d} \chi^2(k-1)$ となることがわかる．

Theorem 4.5. H_0 のもとで，

$$2 \log \Lambda_n - \chi^2_n \xrightarrow{P} 0.$$

Proof. H_0 を仮定する．

$$\widehat{d}_j = \frac{\widehat{p}_j - p_{0,j}}{p_{0,j}} = \frac{Y_j - np_{0,j}}{np_{0,j}},\ j = 1, \ldots, k$$

とおくと，

$$2\log\Lambda_n = \sum_{j=1}^{j} Y_j \log(\widehat{p}_j/p_{0,j}) = \sum_{j=1}^{k} np_{0,j}(1+\widehat{d}_j)\log(1+\widehat{d}_j)$$

と表せる．ここで，

$$(1+t)\log(1+t) = t + t^2/2 + t^2 R(t),\ \lim_{t\to 0} R(t) = 0$$

と展開できる．CLT より，

$$\sqrt{n}\widehat{d}_j \xrightarrow{d} N(0, (1-p_{0,j})/p_{0,j})$$

となるから，

$$np_{0,j}(1+\widehat{d}_j)\log(1+\widehat{d}_j) = np_{0,j}(\widehat{d}_j + \widehat{d}_j^2/2) + \underbrace{p_{0,j}(\sqrt{n}\widehat{d}_j)^2 R(\widehat{d}_j)}_{=\widehat{r}_j}$$

と展開できて，$R(\widehat{d}_j) \xrightarrow{P} 0$ より，$\widehat{r}_j \xrightarrow{P} 0$ となる．よって，

$$2\log\Lambda_n = \underbrace{2\sum_{j=1}^{k} np_{0,j}\widehat{d}_j}_{=0} + \underbrace{\sum_{j=1}^{k} np_{0,j}\widehat{d}_j^2}_{=\chi_n^2} + \underbrace{2\sum_{j=1}^{k}\widehat{r}_j}_{\xrightarrow{P}0}$$

となるから，

$$2\log\Lambda_n - \chi_n^2 \xrightarrow{P} 0$$

をえる． □

分割表の独立性検定

$Z \in \{1, \ldots, I\}, W \in \{1, \ldots, J\}$ を離散 r.v.'s とし，(Z, W) と同じ分布に従う独立な確率ベクトル $(Z_1, W_1), \ldots, (Z_n, W_n)$ が得られているとする．このとき，

$$p_{i,j} = P(Z = i, W = j), \quad Y_{i,j} = \sum_{m=1}^{n} I(Z_m = i, W_m = j),\ 1 \le i \le I, 1 \le j \le J$$

とおくと，

$$Y = (Y_{1,1}, \ldots, Y_{1,J}, Y_{2,1}, \ldots, Y_{I,J})' \sim Mn(n, p_{1,1}, \ldots, p_{1,J}, p_{2,1}, \ldots, p_{I,J})$$

である．ここで，

$$H_0 : Z \text{ と } W \text{ は独立 vs. } H_1 : Z \text{ と } W \text{ は独立でない}$$

という検定問題を考える．

$$p_{i+} = P(Z=i) = \sum_{j=1}^{J} p_{i,j}, \quad p_{+j} = P(W=j) = \sum_{i=1}^{I} p_{i,j}$$

とおくと，この検定問題は

$$H_0 : p_{i,j} = p_{i+}p_{+j},\ \forall(i,j) \quad \text{vs.} \quad H_1 : p_{i,j} \neq p_{i+}p_{+j},\ \exists(i,j)$$

と等価である．

この検定問題に対する LRT 統計量を求めてみよう．無制約のときの $p_{i,j}$ の MLE は

$$\widehat{p}_{i,j} = Y_{i,j}/n$$

である．一方，H_0 のもとで，Y の確率関数は

$$P(Y_{i,j} = y_{i,j}\ \forall(i,j)) = \frac{n!}{\prod_{i,j} y_{i,j}!} \prod_{i=1}^{I} p_{i+}^{\sum_{\ell=1}^{J} Y_{i,\ell}} \prod_{j=1}^{J} p_{+j}^{\sum_{k=1}^{I} Y_{k,j}},\ \sum_{i,j} y_{i,j} = n$$

であるから，p_{i+}, p_{+j} の MLE は

$$\widehat{p}_{i+} = \sum_{j=1}^{J} Y_{i,j}/n, \quad \widehat{p}_{+j} = \sum_{i=1}^{I} Y_{i,j}/n$$

である．以上より，LRT 統計量は

$$\Lambda_n = \prod_{i,j} \left(\frac{\widehat{p}_{i,j}}{\widehat{p}_{i+}\widehat{p}_{+j}} \right)^{Y_{i,j}}$$

である．無制約のとき，自由に動けるパラメータの数は $IJ-1$ であって，H_1 のもとで自由に動けるパラメータの数は $I+J-2$ であるから，その差は

$$IJ - 1 - (I+J-2) = (I-1)(J-1)$$

である．よって，H_0 のもとで，

$$2\log\Lambda_n \xrightarrow{d} \chi^2((I-1)(J-1))$$

となるから，

$$2\log\Lambda_n > \chi^2_\alpha((I-1)(J-1)) \Rightarrow \text{reject}$$

という検定は近似的に水準 α をもつ．

LRT 統計量のほかに，次の χ^2 検定統計量

$$\chi_n^2 = \sum_{i,j} \frac{(Y_{i,j} - n\widehat{p}_{i+}\widehat{p}_{+j})^2}{n\widehat{p}_{i+}\widehat{p}_{+j}}$$

もよく使われる．LRT 統計量と同様に，H_0 のもとで，

$$\chi_n^2 \xrightarrow{d} \chi^2((I-1)(J-1))$$

となることが示せるから，

$$\chi_n^2 > \chi_\alpha^2((I-1)(J-1)) \Rightarrow \text{reject}$$

という検定は近似的に水準 α をもつ．

5 区間推定

本節ではパラメータに対する信頼域の構成を考察する．推定量はパラメータへのあてはめ値を返すが，推定量は確率的なので，真値に等しくなるわけではない (推定量が連続型なら，真値に等しい確率は 0 である) し，推定量の "精度" についても何も情報をもたらさない．区間推定は真値を含む確率があらかじめ決められた値を達成するような，データにもとづく (確率的な) 集合を構成することによって，点推定を補完する．

$\mathcal{X}$ を有限次元ユークリッド空間とし，$\varnothing \neq \Theta \subset \mathbb{R}^k$ をパラメータ空間として，$\{p_\theta : \theta \in \Theta\}$ を $\mathcal{X}$ 上のパラメトリックな分布族とする．$\theta \in \Theta$ に対して，$X_1, \ldots, X_n \sim p_\theta$ i.i.d. が与えられたとして，$X = (X_1', \ldots, X_n')'$ とおく．

Definition 10 (信頼域・信頼区間・被覆確率)**.** 与えられた $\alpha \in (0,1)$ に対して，X にもとづく集合 $S(X) \subset \mathbb{R}^k$ が

$$P_\theta\{\theta \in S(X)\} \geq 1 - \alpha, \ \forall \theta \in \Theta$$

をみたすとき，$S(X)$ を水準 $(1-\alpha)$ の <u>信頼域</u> (confidence region, CR) と呼ぶ．$k = 1$ であって，$S(X)$ が区間のとき，$S(X)$ を <u>信頼区間</u> (confidence interval, CI) と呼ぶ．また，$P_\theta\{\theta \in S(X)\}$ を θ における $S(X)$ の <u>被覆確率</u> (coverage probability) と呼ぶ.

検定のときと同様に，α には 0.05 や 0.01 が使われることが多い．CR を構成する一般的な方法は，検定の受容域を反転させることである．ここで，$\theta_0 \in \Theta$ を任意に固定して，

$$H_0 : \theta = \theta_0 \text{ vs. } H_1 : \theta \neq \theta_0 \tag{*}$$

という検定を考える．この検定問題に対する非確率化検定 δ_{θ_0} が与えられたとき，その <u>受容域</u> (acceptance region) を

$$A(\theta_0) = \{x \in \mathcal{X}^n : \delta_{\theta_0}(x) = 0\}$$

と定義する．δ_{θ_0} は水準 α をもつとする：

$$P_{\theta_0}\{\delta_{\theta_0}(X) = 1\} \leq \alpha.$$

このとき，θ_0 を自由に動かして，

$$S(x) = \{\theta \in \Theta : x \in A(\theta)\}, \ x \in \mathcal{X}^n$$

とおく．

Theorem 5.1. このように構成した $S(X)$ は水準 $(1-\alpha)$ の信頼域である．

Proof. $P_\theta\{\theta \in S(X)\} \geq 1-\alpha \ \forall\theta \in \Theta$ を示せばよい．ここで，

$$\theta \in S(X) \Leftrightarrow X \in A(\theta) \Leftrightarrow \delta_\theta(X) = 0$$

であるから，

$$P_\theta\{\theta \in S(X)\} = P_\theta\{\delta_\theta(X) = 0\} = 1 - P_{\theta_0}\{\delta_\theta(X) = 1\} \geq 1-\alpha$$

を得る． □

逆に，$S(X)$ が水準 $(1-\alpha)$ の CR なら，$A(\theta) = \{x : \theta \in S(x)\}$ とおくと，

$$X \notin A(\theta_0) \Rightarrow \text{reject}$$

という検定は (*) に対する水準 α の検定である．従って，(*) に対する水準 α の非確率化検定と θ に対する水準 $(1-\alpha)$ の CR は 1 対 1 に対応する．この関係は検定と CR の 双対性 (duality) と呼ばれる．

Remark 5.1. 局外パラメータ η がある場合でも，各 $\theta_0 \in \Theta_0$ に対して，

$$H_0 : \theta = \theta_0, \eta : \text{free vs. } H_1 : \theta \neq \theta_0, \eta : \text{free}$$

という検定問題に対する水準 α の検定 δ_{θ_0} があれば，同じ操作によって構成された信頼域 $S(X)$ は水準 $(1-\alpha)$ をもつ．

Remark 5.2. θ が 1 次元のときは，(*) の対立仮説として，$H_1 : \theta > \theta_0$ や $H_1 : \theta < \theta_0$ も考えることができる．このとき，受容域を反転させた CR は $[L(X), \infty)$ や $(-\infty, U(X)]$ ような半開区間になることが多い．

$$P_\theta\{L(X) \leq \theta\} \geq 1-\alpha, \ \forall\theta \in \Theta$$

をみたす統計量 $L(X)$ を水準 $(1-\alpha)$ の 信頼下界 (lower confidence bound, LCB) と呼び，

$$P_\theta\{\theta \leq U(X)\} \geq 1-\alpha, \ \forall\theta \in \Theta$$

をみたす統計量 $U(X)$ を水準 $(1-\alpha)$ の 信頼上界 (upper confidence bound, UCB) と呼ぶ．LCB や UCB は θ の推定量とも解釈できる．例えば，水準 $(1-\alpha)$ の UCB は，θ を過小推定する確率が α より小さい推定量と解釈できる．

Example 5.1. $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ i.i.d. に対して，$H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$ という検定問題に対する水準 α の UMPU 検定の受容域は

$$A(\mu_0) = \{x : \sqrt{n}|\overline{x} - \mu_0| \leq t_{\alpha/2}(n-1)s\}$$

である．これを反転して水準 $(1-\alpha)$ の CI

$$\{\mu : |\mu - \overline{X}| \leq t_{\alpha/2}(n-1)S/\sqrt{n}\} = \left[\overline{X} - t_{\alpha/2}(n-1)S/\sqrt{n}, \overline{X} + t_{\alpha/2}(n-1)S/\sqrt{n}\right]$$

を得る．

Example 5.2. $X_n \sim Bin(n,\theta)$ として，$H_0 : \theta = \theta_0$ vs. $H_0 : \theta \neq \theta_0$ という検定問題に対して，

$$\frac{\sqrt{n}|X_n/n - \theta_0|}{\sqrt{\theta_0(1-\theta_0)}} > z_{\alpha/2} \Rightarrow \text{reject}$$

という検定は近似的に水準 α の検定であった．この検定の受容域を反転して，θ に対する CI を得る．この CI は，$\widehat{\theta} = X_n/n$ とおくと，

$$\frac{n\widehat{\theta} + z_{\alpha/2}^2/2}{n + z_{\alpha/2}^2} \pm \frac{z_{\alpha/2}\sqrt{n}}{n + z_{\alpha/2}^2}\sqrt{\widehat{\theta}(1-\widehat{\theta}) + z_{\alpha/2}^2/(4n)}$$

と計算できる．この CI は Wilson の CI と呼ばれる．

しかし，$\theta = \theta_0$ のとき，

$$\frac{\sqrt{n}(\widehat{\theta} - \theta_0)}{\sqrt{\widehat{\theta}(1-\widehat{\theta})}} \xrightarrow{d} N(0,1)$$

となることを利用して，

$$\left[\widehat{\theta} - z_{\alpha/2}\sqrt{\widehat{\theta}(1-\widehat{\theta})}/\sqrt{n}, \widehat{\theta} + z_{\alpha/2}\sqrt{\widehat{\theta}(1-\widehat{\theta})}/\sqrt{n}\right] \tag{**}$$

というより簡便な CI を利用することの方が標準的である．(**) の CI も近似的に水準 $(1-\alpha)$ をもつ．

Remark 5.3. (**) の CI は標準的であるが，(n,θ) の組み合わせによっては過小な被覆確率をもつことが指摘されている．Brown et al. (2001) を参照のこと．

この例において，CI は標本サイズの決め方にも応用できる．例えば，ある政党の支持率 $\theta \in [0,1]$ を調査したいとして，n 人に対してアンケートを行い，X_n 人がその政党を支持すると答えたとする．このとき，$X_n \sim Bin(n,\theta)$ とみなして，θ に対する水準 0.95 の CI の長さが 0.02 以下になるように標本サイズ n を決めるとする．ここで，$z_{0.025} \approx 1.96$ という近似を使うと，(**) の CI の長さは

$$2 \times 1.96 \underbrace{\sqrt{\widehat{\theta}(1-\widehat{\theta})}}_{\leq 1/2}/\sqrt{n} \leq 1.96/\sqrt{n}$$

だから，これが 0.02 以下であるためには，

$$1.96/\sqrt{n} \leq 0.02 \Leftrightarrow n \geq \left(\frac{1.96}{0.02}\right)^2 = 9604$$

であればよい．

水準をみたすだけなら，CR を Θ にとってしまえばよいが，それでは役に立たない．従って，水準をみたしつつ，より "小さい"CR がよい CR といえる (よって，なるべく被覆確率が水準に等しくなるように CR を選ぶべきである)．ここでは，非確率化 UMPU 検定の受容域を反転して得られる CR が，あるクラスの CR の中で最良であることを示そう．θ に対する水準 $(1-\alpha)$ の CR $S(X)$ が <u>不偏</u> (unbiased) であるとは，

$$P_\theta\{\widetilde{\theta} \in S(X)\} \le 1-\alpha,\ \forall\theta \neq \widetilde{\theta}$$

となることをいう．

Lemma 5.1. (a) 各 $\theta_0 \in \Theta$ に対して，(*) に対する水準 α の不偏な非確率化検定の受容域 $A(\theta_0)$ が与えられたとき，$S(X) := \{\theta : X \in A(\theta)\}$ は θ に対する水準 $(1-\alpha)$ の不偏な CR である．

(b) 逆に，$S(X)$ が θ に対する水準 $(1-\alpha)$ の不偏な CR なら，各 $\theta_0 \in \Theta$ に対して，$A(\theta_0) := \{x : \theta \in S(x)\}$ は (*) に対する水準 α の不偏な非確率化検定の受容域になる．

Proof. (a). $\theta_0 \in S(x) \Leftrightarrow x \in A(\theta_0)$ より，任意の $\theta \neq \theta_0$ に対して，

$$P_\theta\{\theta_0 \in S(X)\} = P_\theta\{X \in A(\theta_0)\} = 1 - P_\theta\{X \notin A(\theta_0)\} \le 1-\alpha$$

である．よって，$S(X)$ は不偏な CR である．

(b). 同様にして，$\theta \neq \theta_0$ に対して，

$$P_\theta\{X \notin A(\theta_0)\} = 1 - P_\theta\{X \in A(\theta_0)\} = 1 - P_\theta\{\theta_0 \in S(X)\} \ge \alpha$$

となるから，$X \notin A(\theta_0) \Rightarrow$ reject という検定は不偏である． □

$S \subset \mathbb{R}^k$ に対して，$\mathrm{Vol}(S)$ を S の体積とする (体積がちゃんと定義できることは仮定する)．

Theorem 5.2 (Pratt (1961))**.** 各 $\theta_0 \in \Theta$ に対して，(*) に対する水準 α の非確率化 UMPU 検定の存在を仮定し，その受容域を $A^*(\theta_0)$ とおいて，対応する CR を $S^*(X) := \{\theta : X \in A^*(\theta)\}$ とおく．このとき，任意の水準 $(1-\alpha)$ の不偏な CR $S(X)$ に対して，

$$E_\theta[\mathrm{Vol}(S^*(X))] \le E_\theta[\mathrm{Vol}(S(X))],\ \forall\theta \in \Theta$$

が成り立つ．

Proof. $\theta_0 \in \Theta$ を任意に固定する．$A(\theta_0) := \{x : \theta_0 \in S(x)\}$ は (*) に対する水準 α の不偏な非確率化検定の受容域である．このとき，$A^*(\theta_0)$ の定義より，任意の $\theta \neq \theta_0$ に対して，

$$P_\theta\{X \notin A(\theta_0)\} \le P_\theta\{X \notin A^*(\theta_0)\}$$

であるから，

$$P_\theta\{\theta_0 \in S(X)\} \ge P_\theta\{\theta_0 \in S^*(X)\}$$

を得る．θ と θ_0 を入れ替えて，

$$P_{\theta_0}\{\theta \in S(X)\} \geq P_{\theta_0}\{\theta \in S^*(X)\},\ \forall\theta \neq \theta_0$$

を得る．ここで，

$$\begin{aligned}
E_{\theta_0}[\mathrm{Vol}(S(X))] &= E_{\theta_0}\left[\int_{\mathbb{R}^k} I(\theta \in S(X))d\theta\right] \\
&= \int_{\mathbb{R}^k} E_{\theta_0}[I(\theta \in S(X))]d\theta = \int_{\theta\neq\theta_0} P_{\theta_0}\{\theta \in S(X)\}d\theta \\
&\geq \int_{\theta\neq\theta_0} P_{\theta_0}\{\theta \in S^*(X)\}d\theta = E_{\theta_0}[\mathrm{Vol}(S^*(X))]
\end{aligned}$$

より，定理の結論を得る． □

5.1 最尤法にもとづく方法

$X_1, \ldots, X_n \sim p_\theta$ i.i.d. とする．さらに，$k = 1$ とし，$\Theta \subset \mathbb{R}$ を開区間とする．$\widehat{\theta}$ を θ の MLE とすると，いくつかの正則条件のもとで，

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} N(0, 1/I(\theta))$$

が成り立つ．ここで，$I(\theta)$ は Fisher 情報量である．よって，

$$S(X) = \{\theta : \sqrt{nI(\theta)}|\widehat{\theta} - \theta| \leq z_{\alpha/2}\}$$

とおくと，$S(X)$ は近似的に水準 $(1-\alpha)$ の CR になる．

もっと簡便な方法として，$I(\theta)$ が θ について連続なら，

$$I(\widehat{\theta}) \xrightarrow{P} I(\theta)$$

となるから，Slutsky の補題より，

$$\sqrt{nI(\widehat{\theta})}(\widehat{\theta} - \theta) \xrightarrow{d} N(0, 1)$$

を得る．よって，

$$\left[\widehat{\theta} - z_{\alpha/2}/\sqrt{nI(\widehat{\theta})}, \widehat{\theta} + z_{\alpha/2}/\sqrt{nI(\widehat{\theta})}\right]$$

は近似的に水準 $(1-\alpha)$ の CI になる．

もう 1 つの方法として，<u>分散安定化変換</u> (variance stabilizing transformation) を利用する方法がある．$g : \mathbb{R} \to \mathbb{R}$ を各 $\theta \in \Theta$ で微分可能な関数とすると，デルタ法より，

$$\sqrt{n}(g(\widehat{\theta}) - g(\theta)) \xrightarrow{d} N(0, \{g'(\theta)\}^2/I(\theta))$$

となる．よって，g が

$$\{g'(\theta)\}^2 = I(\theta) \tag{*}$$

をみたせば，

$$\sqrt{n}(g(\widehat{\theta}) - g(\theta)) \xrightarrow{d} N(0,1)$$

となる．このように，極限の正規分布の分散が既知になるような変換 g のことを分散安定化変換という．このとき，

$$\{\theta : |\sqrt{n}(g(\widehat{\theta}) - g(\theta))| \le z_{\alpha/2}\}$$

は近似的に水準 $(1-\alpha)$ の CR になる．さらに，g が Θ 上で狭義単調増加なら，この CR は

$$\left[g^{-1}\left(g(\widehat{\theta}) - z_{\alpha/2}/\sqrt{n}\right), g^{-1}\left(g(\widehat{\theta}) + z_{\alpha/2}/\sqrt{n}\right)\right]$$

という形になる．また，このとき，(*) $\Leftrightarrow g'(\theta) = \sqrt{I(\theta)}$ だから，

$$g(\theta) = \underbrace{\int \sqrt{I(\theta)} d\theta}_{\text{不定積分}} + \text{定数}$$

の形になる．

Example 5.3. $X_1, \dots, X_n \sim N(0, \tau)$ i.i.d. $(\tau > 0)$ とすると，τ の MLE は

$$\widehat{\tau} = \frac{1}{n}\sum_{i=1}^n X_i^2$$

であって，$\sqrt{n}(\widehat{\tau} - \tau) \xrightarrow{d} N(0, 2\tau^2)$ である．ここで，$\sqrt{n/2}(\log\widehat{\tau} - \log\tau) \xrightarrow{d} N(0,1)$ だから，

$$\left[\exp\left(\log\widehat{\tau} - \sqrt{2}z_{\alpha/2}/\sqrt{n}\right), \exp\left(\log\widehat{\tau} + \sqrt{2}z_{\alpha/2}/\sqrt{n}\right)\right]$$

は τ に対する近似的に水準 $(1-\alpha)$ の CI になる．

Θ が多次元のときは，

$$H_0 : \theta = \theta_0 \text{ vs. } H_1 : \theta \neq \theta_0$$

という検定問題に対する LRT 統計量を

$$\Lambda_n(\theta_0) = \frac{\sup_{\theta\in\Theta} p_\theta^n(X)}{p_{\theta_0}^n(X)}$$

とおくと，

$$S(X) = \{\theta \in \Theta : 2\log\Lambda_n(\theta) \le \chi_\alpha^2(k)\}$$

は近似的に水準 $(1-\alpha)$ の CR になる．

Example 5.4. $X_1, \dots, X_n \sim U[0, \theta]$ i.i.d. として，θ に対して $(-\infty, U(X)]$ という形の CI を考える．θ は分布のサポートの右端なので，このような CI を考えるのは自然である．このとき，MLE は $X_{(n)}$ であって，

$$n(\theta - X_{(n)}) \xrightarrow{d} Ex(1/\theta)$$

となる．よって，MLE は漸近正規性をみたさないが，

$$\frac{n(\theta - X_{(n)})}{\theta} \xrightarrow{d} Ex(1)$$

となるから，この結果を利用して θ の UCB を構成することができる．$c < n$ に対して

$$\frac{n(\theta - X_{(n)})}{\theta} \le c \Leftrightarrow \theta \le \frac{X_{(n)}}{1 - c/n}$$

であって，$Ex(1)$ の $(1-\alpha)$ 分位点 $= -\log \alpha$ だから，

$$P_\theta \left\{ \theta \le \frac{X_{(n)}}{1 + (1/n) \log \alpha} \right\} = 1 - \alpha + o(1)$$

を得る．ところで，

$$\frac{X_{(n)}}{\theta} \sim Be(n, 1)$$

であって，$Be(n, 1)$ の α 分位点は $\alpha^{1/n}$ だから，

$$P_\theta \left\{ \theta \le \frac{X_{(n)}}{\alpha^{1/n}} \right\} = P_\theta \left\{ \frac{X_{(n)}}{\theta} \ge \alpha^{1/n} \right\} = 1 - \alpha$$

も得る．ここで，

$$\alpha^{1/n} = e^{(1/n) \log \alpha} = 1 + (1/n) \log \alpha + O(n^{-2})$$

であるから，2 つの CI に大きな違いはない．

5.2 Bayes 信用区間

Θ を $\mathbb{R}$ の区間とし，π を Θ 上の事前密度とする：

$$\begin{aligned} &X_1, \dots, X_n \mid \theta \sim p_\theta \text{ i.i.d.} \\ &\theta \sim \pi. \end{aligned}$$

このとき，θ の事後密度は

$$\pi(\theta \mid X) \propto p_\theta^n(X) \pi(\theta)$$

である．

X にもとづく区間 $C = C(X)$ が

$$P(\theta \in C \mid X) = \int_C \pi(\theta \mid X) d\theta \geq 1 - \alpha$$

をみたすとき，$C(X)$ を水準 $(1-\alpha)$ の <u>信用区間</u> (credible interval) と呼ぶ．信用区間の選び方には任意性があるが，例えば次のようにすればよい．$\alpha \in (0,1)$ に対して $\widehat{\zeta}_\alpha = \widehat{\zeta}_\alpha(X)$ を θ の事後分布の $(1-\alpha)$ 分位点とおく：

$$\int_{\widehat{\zeta}_\alpha}^{\infty} \pi(\theta \mid X) d\theta = \alpha.$$

このとき，$\alpha, \beta > 0, \alpha + \beta < 1$ に対して，

$$\left[\widehat{\zeta}_{1-\alpha}, \widehat{\zeta}_\beta\right]$$

は水準 $(1-\alpha-\beta)$ の信用区間になる．さらに，パラメータの真値 $\theta = \theta_0$ の存在を仮定すると，$n \to \infty$ のとき，いくつかの正則条件のもとで，

$$P_{\theta_0}\left\{\theta_0 \in \left[\widehat{\zeta}_{1-\alpha}, \widehat{\zeta}_\beta\right]\right\} = 1 - \alpha - \beta + o(1)$$

となることが示せる．

5.3　ブートストラップ

F を $\mathbb{R}$ 上の d.f. とし，$X_1, \ldots, X_n \sim F$ i.i.d. とする．このとき，F の汎関数 $\theta = \theta(F)$ に対して CI を構成することを考える．$\theta(F)$ の例として，F の平均，分散，分位点などがある．さらに，θ に対して，推定量

$$\widehat{\theta} = \widehat{\theta}(X_1, \ldots, X_n)$$

と何らかの統計量 $\widehat{\sigma} = \widehat{\sigma}(X_1, \ldots, X_n) > 0$ が存在して，$n \to \infty$ のとき，

$$T_n := \frac{\sqrt{n}(\widehat{\theta} - \theta)}{\widehat{\sigma}} \xrightarrow{d} T$$

とする．$G(t) = P(T \leq t), t \in \mathbb{R}$ とおき，$\alpha \in (0,1)$ に対して，ξ_α を G の $(1-\alpha)$ 分位点とする：$\xi_\alpha = G^{\leftarrow}(1-\alpha)$．$G$ が連続なら，$\alpha, \beta > 0, \alpha + \beta < 1$ に対して，

$$\left[\widehat{\theta} - \xi_\beta \widehat{\sigma}/\sqrt{n}, \widehat{\theta} - \xi_{1-\alpha}\widehat{\sigma}/\sqrt{n}\right]$$

は θ に対する近似的に水準 $(1-\alpha-\beta)$ の CI になる [26]．これは，

$$\theta \in \left[\widehat{\theta} - \xi_\beta \widehat{\sigma}/\sqrt{n}, \widehat{\theta} - \xi_{1-\alpha}\widehat{\sigma}/\sqrt{n}\right] \Leftrightarrow \xi_{1-\alpha} \leq \frac{\sqrt{n}(\widehat{\theta} - \theta)}{\widehat{\sigma}} \leq \xi_\beta$$

という同値関係からわかる．例えば，水準 0.95 の CI を構成したいなら，$\alpha = \beta = 0.025$ とすればよい．

[26] ここでは “パラメータ” は F であって，その “関数” $\theta(F)$ に対して CI を構成することを考えている．よって，ここでの設定はこれまでの設定と少し異なっている．

Example 5.5. $\widehat{\sigma}$ と T の分布の組み合わせには任意性がある．ある $\tau = \tau(F) > 0$ が存在して，$n \to \infty$ のとき，

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} N(0, \tau^2)$$

としよう．このとき，$\widehat{\sigma} = 1$ とすると，$T \sim N(0, \tau^2)$ となり，$\widehat{\sigma}$ を τ の一致推定量とすると，$T \sim N(0, 1)$ となる．

<u>ブートストラップ</u> (bootstrap) とは統計量の標本分布を推定する汎用的な手法であり，Efron (1979) によって提案された．$\widehat{F}_n(x) = n^{-1} \sum_{i=1}^n I(X_i \le x)$ を経験分布関数とし，$X_1, \ldots, X_n$ を所与として，

$$X_1^*, \ldots, X_n^* \sim \widehat{F}_n$$

を人工的に発生させる．$X_1^*, \ldots, X_n^*$ を<u>ブートストラップ標本</u> (bootstrap sample) と呼ぶ．P^* をブートストラップ標本に関する確率とする．例えば，

$$P^*(X_i^* \le x) = \widehat{F}_n(x)$$

である．ここで，

$$\widehat{\theta}^* = \widehat{\theta}(X_1^*, \ldots, X_n^*),\ \widehat{\sigma}^* = \widehat{\sigma}(X_1^*, \ldots, X_n^*),\ T_n^* = \frac{\sqrt{n}(\widehat{\theta}^* - \widehat{\theta})}{\widehat{\sigma}^*}$$

とおいて，T_n^* の P^* のもとでの d.f. を $\widehat{G}_n$ とおく：

$$\widehat{G}_n(t) = P^*(T_n^* \le t),\ t \in \mathbb{R}.$$

$\widehat{G}_n$ は $X_1, \ldots, X_n$ に依存するから，確率的な d.f. である．また，$\widehat{\xi}_\alpha$ を $\widehat{G}_n$ の $(1-\alpha)$ 分位点とおく：

$$\widehat{\xi}_\alpha = \widehat{G}_n^{\leftarrow}(1-\alpha).$$

$\widehat{G}_n$ は G を近似していると考えられるから，

$$P(T_n \le \widehat{\xi}_\alpha) \approx 1 - \alpha$$

となることが予想される．そこで，

$$\left[\widehat{\theta} - \widehat{\xi}_\beta \widehat{\sigma}/\sqrt{n}, \widehat{\theta} - \widehat{\xi}_{1-\alpha} \widehat{\sigma}/\sqrt{n}\right] \tag{*}$$

という CI を考える．

Remark 5.4. ほとんどの場合，$\widehat{\xi}_\alpha$ は陽には計算できないので，その計算はシミュレーションによる．ブートストラップ標本を独立に B 回発生させる：

$$X_{1,b}^*, \ldots, X_{n,b}^* \sim \widehat{F}_n,\ b = 1, \ldots, B,\ \text{i.i.d.}$$

このとき，$\widehat{\theta}_b^* = \widehat{\theta}(X_{1,b}^*, \ldots, X_{n,b}^*), \widehat{\sigma}_b^* = \widehat{\sigma}(X_{1,b}^*, \ldots, X_{n,b}^*), T_{n,b}^* = \sqrt{n}(\widehat{\theta}_b^* - \widehat{\theta})/\widehat{\sigma}_b^*$ とおくと，$X_1, \ldots, X_n$ を与えたとき，

$$T_{n,1}^*, \ldots, T_{n,B}^* \sim \widehat{G}_n \text{ i.i.d.}$$

であるから，$\widehat{G}_n(t)$ は

$$\widehat{G}_n(t) \approx \frac{1}{B}\sum_{b=1}^{B} I(T_{n,b}^* \leq t)$$

と近似できる．以上より，$\widehat{\xi}_\alpha$ は

$$\widehat{\xi}_\alpha \approx \inf\left\{t \in \mathbb{R} : \frac{1}{B}\sum_{b=1}^{B} I(T_{n,b}^* \leq t) \geq 1-\alpha\right\}$$

と近似すればよい．

- $\widehat{\sigma} = 1$ のとき，(*) の CI の構成法を パーセンタイル法 (percentile method) と呼ぶ．$\widehat{\zeta}_\alpha$ を $\widehat{\theta}^*$ の P^* のもとでの d.f. の $(1-\alpha)$ 分位点とする：

  $$\widehat{\zeta}_\alpha = \inf\{t \in \mathbb{R} : P^*(\widehat{\theta}^* \leq t) \geq 1-\alpha\}.$$

  このとき，$T_n^* = \sqrt{n}(\widehat{\theta}^* - \widehat{\theta})$ より，

  $$\widehat{\xi}_\alpha = \sqrt{n}(\widehat{\zeta}_\alpha - \widehat{\theta})$$

  となる．よって，パーセンタイル法による CI は

  $$\left[2\widehat{\theta} - \widehat{\zeta}_\beta, 2\widehat{\theta} - \widehat{\zeta}_{1-\alpha}\right]$$

  とも表せる．$\widehat{\zeta}_\alpha$ もブートスラップ標本を多数発生させることによって，近似計算できる．

- $\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} N(0, \tau^2), \tau > 0$ であって，$\widehat{\sigma}$ が τ の一致推定量のとき，(*) の CI の構成法を パーセンタイル t 法 (percentile t-method) と呼ぶ．

$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} N(0, \tau^2)$ であって，τ の一致推定量が容易に構成できる場合，パーセンタイル t 法は，パーセンタイル法や正規近似にもとづく CI $[\widehat{\theta} - z_\beta\widehat{\sigma}/\sqrt{n}, \widehat{\theta} - z_{1-\alpha}\widehat{\sigma}/\sqrt{n}]$ と比べて (ここで $\widehat{\sigma}$ は τ の一致推定量とする)，より小さい被覆確率の誤差をもつといわれる [27]．しかし，パーセンタイル法は τ の推定を必要としない分，τ の一致推定が難しい問題に対しても有効である．そのような問題として，分位点の推定を考察してみよう．

[27] こうした被覆確率の比較は Edgeworth 展開 と呼ばれる正規近似の精密評価にもとづく．詳細は Hall (1993) を参照せよ．

Example 5.6. $u \in (0,1)$ とし，F の u 分位点 $\theta_u = F^{\leftarrow}(u)$ の推定を考える．経験分布関数を $\widehat{F}_n(x) = n^{-1}\sum_{i=1}^n I(X_i \le x)$ とおくと，θ_u の標準的な推定量は標本 u 分位点 $\widehat{\theta}_u = \widehat{F}_n^{\leftarrow}(u)$ である．いま，F は密度関数 f をもち，f は θ_u で正かつ連続と仮定する．このとき，

$$\sqrt{n}(\widehat{\theta}_u - \theta_u) \xrightarrow{d} N\left(0, \frac{u(1-u)}{f(\theta_u)^2}\right)$$

となる (後述)．ここで，漸近分散

$$\frac{u(1-u)}{f(\theta_u)^2}$$

は未知の密度関数に依存していて，その推定はそれほど明らかではない．密度関数の値 $f(\theta_u)$ を一致推定する手法はいくつかあるが，その場合，バンド幅と呼ばれるパラメータをユーザーが決めなければならない．

u 分位点 θ_u に対してパーセンタイル法を使った CI を構成してみよう．$\widehat{\theta}_u^*$ をブートストラップ標本 $X_1^*, \ldots, X_n^*$ にもとづく標本 u 分位点とする．すなわち，

$$\widehat{F}_n^*(x) = \frac{1}{n}\sum_{i=1}^n I(X_i^* \le x),\ x \in \mathbb{R}$$

とおくと，

$$\widehat{\theta}_u^* = \widehat{F}_n^{*\leftarrow}(u) = \inf\{x \in \mathbb{R} : \widehat{F}_n^*(x) \ge u\}$$

である．そこで，$\widehat{\theta}_u^*$ の P^* のもとでの $1-\alpha$ 分位点を $\widehat{\zeta}_{u,\alpha}$ とおく：

$$\widehat{\zeta}_{u,\alpha} = \inf\{x \in \mathbb{R} : P^*(\widehat{\theta}_u^* \le x) \ge 1-\alpha\}.$$

このとき，パーセンタイル法にもとづく CI は

$$\left[2\widehat{\theta}_u - \widehat{\zeta}_{u,\beta}, 2\widehat{\theta}_u - \widehat{\zeta}_{u,1-\alpha}\right]$$

で与えられる．この CI の利点は複雑な漸近分散の推定を省略できる点にある．この CI は，前述の仮定のもとで近似的に水準 $1-\alpha-\beta$ をもつことが示される (6.2 節を参照せよ)．

(*) の CI の漸近的な正当性を保証する十分条件を与えよう．

Theorem 5.3. G は連続であって，さらに次の条件が成り立つことを仮定する：

$$\sup_{t\in\mathbb{R}} |\widehat{G}_n(t) - G(t)| \xrightarrow{P} 0. \tag{**}$$

このとき，

$$P\left\{\theta \in \left[\widehat{\theta} - \widehat{\xi}_\beta \widehat{\sigma}/\sqrt{n}, \widehat{\theta} - \widehat{\xi}_{1-\alpha}\widehat{\sigma}/\sqrt{n}\right]\right\} \to 1-\alpha-\beta.$$

Proof. $G_n(t) = P(T_n \le t), t \in \mathbb{R}$ とおく．$G_n \xrightarrow{d} G$ と G が連続なことから，Pólya の定理より，

$$\sup_{t\in\mathbb{R}} |G_n(t) - G(t)| \to 0$$

となる．$Y_n \xrightarrow{P} 0$ なら，ある数列 $\varepsilon_n \to 0$ が存在して $P(|Y_n| > \varepsilon_n) \le \varepsilon_n$ となるから，十分遅い $\varepsilon_n \to 0$ に対して，

$$\sup_{t\in\mathbb{R}} |G_n(t) - G(t)| \le \varepsilon_n,\ P\left\{\sup_{t\in\mathbb{R}} |\widehat{G}_n(t) - G(t)| \le \varepsilon_n\right\} > 1 - \varepsilon_n$$

となる．そこで，

$$E_n = \left\{\sup_{t\in\mathbb{R}} |\widehat{G}_n(t) - G(t)| \le \varepsilon_n\right\} \subset \Omega$$

とおくと，E_n 上で，

$$\widehat{G}_n(\xi_{\alpha-\varepsilon_n}) \ge G(\xi_{\alpha-\varepsilon_n}) - \varepsilon_n = 1 - (\alpha - \varepsilon_n) - \varepsilon_n = 1 - \alpha$$

となることから，

$$\widehat{\xi}_\alpha \le \xi_{\alpha-\varepsilon_n} \quad \text{on } E_n$$

を得る．これから，

$$P(T_n \le \widehat{\xi}_\alpha) \le \underbrace{P(T_n \le \xi_{\alpha-\varepsilon_n})}_{=G_n(\xi_{\alpha-\varepsilon_n})} + P(E_n^c) \le G(\xi_{\alpha-\varepsilon_n}) + 2\varepsilon_n = 1 - \alpha + 3\varepsilon_n$$

を得る．次に，E_n 上で，

$$G(\widehat{\xi}_\alpha) \ge \widehat{G}_n(\widehat{\xi}_\alpha) - \varepsilon_n \ge 1 - \alpha - \varepsilon_n$$

となることから，

$$\xi_{\alpha+\varepsilon_n} \le \widehat{\xi}_\alpha \quad \text{on } E_n$$

を得る．よって，

$$P(T_n \le \widehat{\xi}_\alpha) \ge \underbrace{P(T_n \le \xi_{\alpha+\varepsilon_n})}_{=G_n(\xi_{\alpha+\varepsilon_n})} - P(E_n^c) \ge G(\xi_{\alpha+\varepsilon_n}) - 2\varepsilon_n = 1 - \alpha - 3\varepsilon_n.$$

以上より，$P(T_n \le \widehat{\xi}_\alpha) \to 1 - \alpha$ を得る．同様に，$P(T_n < \widehat{\xi}_\alpha) \to 1 - \alpha$ も従う．よって，

$$\begin{aligned}
&P\left\{\theta \in \left[\widehat{\theta} - \widehat{\xi}_\beta \widehat{\sigma}/\sqrt{n}, \widehat{\theta} - \widehat{\xi}_{1-\alpha}\widehat{\sigma}/\sqrt{n}\right]\right\} \\
&\quad = P(T_n \le \widehat{\xi}_\beta) - P(T_n < \widehat{\xi}_{1-\alpha}) \\
&\quad \to 1 - \alpha - \beta
\end{aligned}$$

となるから定理が示された． □

(**) の条件は多くの例に対して成り立つが，ここではもっとも単純な例を考察しよう．

Example 5.7. $E[X_1^2] < \infty$ と仮定して，$\theta = E[X_1], \widehat{\theta} = \overline{X}, \widehat{\sigma} = 1$ とする．このとき，

$$T_n = \sqrt{n}(\overline{X} - \theta),\ T_n^* = \sqrt{n}(\overline{X}^* - \overline{X})$$

である．また，$\tau^2 = \mathrm{Var}(X_1)$ とおくと，CLT より，

$$T_n \xrightarrow{d} N(0, \tau^2)$$

となる．$\tau > 0$ と仮定すると，$N(0, \tau^2)$ の d.f. は $\Phi(\cdot/\tau)$ である．このとき，

$$\sup_{t \in \mathbb{R}} |\widehat{G}_n(t) - \Phi(t/\tau)| \xrightarrow{P} 0 \tag{*3}$$

となる．直観的には，P^* のもとで，$X_i^*, i = 1, \ldots, n$ は i.i.d. であって，その平均と分散はそれぞれ $\overline{X}, n^{-1}\sum_{i=1}^n (X_i - \overline{X})^2 =: \widehat{\tau}^2$ であるから，CLT より，

$$\widehat{G}_n(t) \approx \Phi(t/\widehat{\tau}) \tag{*4}$$

となることが予想される．さらに，$\widehat{\tau}^2 \xrightarrow{P} \tau^2$ であるから，

$$\Phi(t/\widehat{\tau}) \xrightarrow{P} \Phi(t/\tau)$$

であるので，(*3) が従うことが予想される．以上の議論は直観的なものであって，厳密ではない．厳密には，(*4) の近似の意味を明確にする必要があるし，(*3) を示すためには，各 $t \in \mathbb{R}$ に対して確率収束 $\widehat{G}_n(t) \xrightarrow{P} \Phi(t/\tau)$ を示すだけでは不十分であって，$t \in \mathbb{R}$ に関して一様に確率収束を示さなくてはならない．(*3) のフォーマルな証明は次節を参照せよ．

Example 5.8. ブートストラップは常にうまく働くわけではない．例えば，$X_1, \ldots, X_n \sim U[0, \theta]$ i.i.d. として，θ に対して CI を構成することを考える．このとき，θ の MLE は $X_{(n)}$ であって，$n(\theta - X_{(n)}) \xrightarrow{d} Ex(1/\theta)$ となる．しかし，$X_1^*, \ldots, X_n^*$ のなかに $X_{(n)}$ が含まれる確率は $1-(1-(1/n))^n = 1-e^{-1}+o(1)$ だから，$(X_1, \ldots, X_n)$ を与えたとき，$n(X_{(n)} - X_{(n)}^*)$ は $1 - e^{-1} + o(1)$ の確率で 0 になってしまって，P^* のもとでの分布が $Ex(1/\theta)$ を近似しない．よって，この場合，パーセンタイル法による CI は誤った被覆確率をもつ．

5.4 Hoeffding の不等式

F を $\mathbb{R}$ 上の d.f. とし，そのサポートは $[a, b]$ に含まれているとする (i.e., $F(a-) = 0, F(b) = 1$)．$X_1, \ldots, X_n \sim F$ i.i.d. が与えられたとき，その平均 $\mu = \mu(F) = E[X_1]$ に対して CI を構成する問題を考える．ここでは，水準 $(1 - \alpha)$ の UCB を構成することを考えよう：

$$P\{\mu \leq U(X)\} \geq 1 - \alpha. \tag{*}$$

分布についてこれ以上の情報はないとき，CLT やブートストラップはこのような UCB を構成する方法を与えるが，水準に近似誤差が生じる．有限標本において (*) を 厳密に みたすような UCB は作れないだろうか．もちろん，$U(X) = b$ としてしまえば，(*) が必ず成り立つが，それでは意味がない．F の分散 σ^2 を既知とすると，CLT にもとづく UCB は $U(X) = \overline{X} + z_\alpha \sigma/\sqrt{n}$ であって，μ との差は $|U(X) - \mu| \le |\overline{X} - \mu| + |z_\alpha|\sigma/\sqrt{n}$ である．ここで，$|\overline{X} - \mu|$ は "確率的に" $O(n^{-1/2})$ であって [28]，$z_\alpha = O(\sqrt{\log(1/\alpha)})$ $(\alpha \to 0)$ だから，μ との差が $O(\sqrt{\log(1/\alpha)}/\sqrt{n})$ $(n \to \infty, \alpha \to 0)$ であって，かつ (*) を厳密にみたすような UCB を作りたい．その 1 つの方法は次の Hoeffding の不等式を用いるものである．

Theorem 5.4 (Hoeffding (1963))**.** $X_1, \dots, X_n$ を独立な r.v.'s とし，各 X_i は $P(X_i \in [a_i, b_i]) = 1$ をみたすとする $(-\infty < a_i < b_i < \infty)$．このとき，任意の $x > 0$ に対して，

$$P\left\{\sum_{i=1}^n (X_i - E[X_i]) > x\right\} \le e^{-\frac{2x^2}{\sum_{i=1}^n (b_i - a_i)^2}} \tag{**}$$

が成り立つ．

Proof. $E[X_i] = 0$ と仮定してよい (このとき，$a_i \le 0, b_i \ge 0$ である)．Markov の不等式より，$x, t > 0$ に対して，

$$P\left(\sum_{i=1}^n X_i > x\right) = P\left(e^{t\sum_{i=1}^n X_i} > e^{tx}\right) \le e^{-tx} E[e^{t\sum_{i=1}^n X_i}] = e^{-tx} \prod_{i=1}^n E[e^{tX_i}].$$

i を固定する．$X_i \in [a_i, b_i]$ だから，$\alpha = (X_i - a_i)/(b_i - a_i)$ とおくと，$\alpha \in [0, 1]$ であって，

$$X_i = \alpha b_i + (1 - \alpha) a_i$$

と表せる．ここで，$x \mapsto e^{tx}$ は凸関数だから，

$$e^{tX_i} \le \alpha e^{tb_i} + (1 - \alpha) e^{ta_i} = \frac{X_i - a_i}{b_i - a_i} e^{tb_i} + \frac{b_i - X_i}{b_i - a_i} e^{ta_i}$$

であって，両辺の期待値をとって，

$$E[e^{tX_i}] \le \underbrace{\frac{-a_i}{b_i - a_i}}_{=\gamma} e^{tb_i} + \frac{b_i}{b_i - a_i} e^{ta_i} = e^{-\gamma t(b_i - a_i)} \left\{\gamma e^{t(b_i - a_i)} + (1 - \gamma)\right\}$$

を得る．そこで，$u = t(b_i - a_i), g(y) = -\gamma y + \log(1 - \gamma + \gamma e^y)$ とおくと，上式の最右辺は $e^{g(u)}$ と表せる．$a_i \le 0$ より $\gamma \ge 0$ だから，g は $\mathbb{R}$ 上で定義されている．ここで，$g(0) = g'(0) = 0$ であって，

$$g''(y) = \frac{\gamma e^y (1 - \gamma)}{(1 - \gamma + \gamma e^y)^2} = \frac{\gamma e^y/(1 - \gamma)}{\{1 + \gamma e^y/(1 - \gamma)\}^2} \le \frac{1}{4}$$

[28] 次節を参照.

だから，Taylor の定理より，$g(u) \leq u^2/8 = t^2(b_i - a_i)^2/8$ を得る．以上より，

$$P\left(\sum_{i=1}^{n} X_i > x\right) \leq e^{-tx + t^2 \sum_{i=1}^{n}(b_i - a_i)^2/8}$$

を得る．あとは右辺を t について最適化して定理の結論を得る． □

(**) において，X_i を $-X_i$ に取り替えると，任意の $x > 0$ に対して，

$$P\left\{\sum_{i=1}^{n}(X_i - E[X_i]) < -x\right\} \leq e^{-\frac{2x^2}{\sum_{i=1}^{n}(b_i - a_i)^2}}$$

が成り立つことをわかる．これと (**) を合わせて，

$$P\left\{\left|\sum_{i=1}^{n}(X_i - E[X_i])\right| > x\right\} \leq 2e^{-\frac{2x^2}{\sum_{i=1}^{n}(b_i - a_i)^2}},\ \forall x > 0$$

を得る．

Chebyshev の不等式から導かれる評価

$$P\left\{\left|\sum_{i=1}^{n}(X_i - E[X_i])\right| > x\right\} \leq x^{-2}\sum_{i=1}^{n}\mathrm{Var}(X_i)$$

と比較すると，Hoeffding の不等式は $x \to \infty$ のとき，バウンドが $e^{-\mathrm{const}.x^2}$ のオーダーで減衰していくのに比べて，Chebyshev の不等式のバウンドは x^{-2} のオーダーでしか減衰しない．この意味で，Hoeffding の不等式はよりシャープなバウンドと導くといえる (ただし，Chebyshev の不等式は 2 次モーメントが有限であれば適用できるのに対して，Hoeffding の不等式は有界な r.v.'s に対してしか適用できないことに注意する)．

Hoeffding の不等式 (Chebyshev 不等式もであるが) は，集中不等式 (concentration inequality) と呼ばれるものの代表的な例である [29]．近年，集中不等式は数理統計学や関連分野において極めて重要な役割を果たすことが認識されてきている．集中不等式については，Boucheron et al. (2013) が優れた文献である．

もともとの問題に戻ると，$X_1, \ldots, X_n$ を i.i.d. とし，$P(X_i \in [a, b]) = 1$ とする．$\mu = E[X_i]$ とおくと，Hoeffding の不等式より，任意の $\alpha \in (0, 1)$ に対して，

$$\overline{X} + (b - a)\sqrt{\frac{\log(1/\alpha)}{2n}}$$

は μ に対する水準 $(1 - \alpha)$ の UCB になる．

[29] 集中不等式とは r.v. が適当な定数 (平均やメディアン) から乖離する確率をバウンドする不等式のことをいう．

Example 5.9 (モンテカルロ近似に必要な標本サイズ)**.** モンテカルロ近似に必要な標本サイズを考察してみる．f を $\mathbb{R}^k$ 上の密度関数とし，$h : \mathbb{R}^k \to [a, b]$ を所与の関数とする．このとき，積分

$$J = \int h(x)f(x)dx$$

をモンテカルロ法によって近似する．$X_1, \ldots, X_N \sim f$ i.i.d. とし，

$$J_N = \frac{1}{N}\sum_{i=1}^{N} h(X_i)$$

とする．所与の $\varepsilon > 0$ と $\alpha \in (0, 1)$ (いずれも十分小さい値とする) に対して，確率 $(1 - \alpha)$ 以上で J_N が J の ε 近傍に入るように標本サイズ N を決めたい．いま，$E[J_N] = J$ であって，Hoeffding の不等式より，

$$P(|J_N - J| \le \varepsilon) \ge 1 - 2e^{-\frac{2N\varepsilon^2}{(b-a)^2}}$$

となる．右辺が $(1 - \alpha)$ に等しくなるような N は，

$$N = N(\varepsilon, \alpha) = \frac{(b-a)^2}{2\varepsilon^2}\log(2/\alpha)$$

である．

6　漸近理論

推定量の “良さ” を評価したり，検定統計量の棄却点を決めたり，信頼区間を構成するときに，統計量の標本分布を求める必要があるが，有限標本において標本分布の厳密分布を求めるのは難しいことが多い．また，そもそもパラメトリックモデルを仮定しない場合，統計量の厳密分布の評価は (ほとんどの場合) 不可能である．従って，そのような場合，漸近理論に頼ることになる [30]．本節は漸近理論に関するごく基本的な内容を扱う．

6.1　基本的な極限定理 (補足)

本節では，基本的な極限定理について，いくつかの補足的な結果を紹介する．最初に d.f. に関する次の技術的な補題を証明する．

Lemma 6.1. F を $\mathbb{R}$ 上の d.f. とすると，F の不連続点は高々可算個である．

Proof. $x \in \mathbb{R}$ が F の不連続点であることは，$F(x) - F(x-) > 0$ と同値である．よって，F の不連続点の全体を D_F とおくと，

$$D_F = \{x : F(x) - F(x-) > 0\} = \bigcup_{n=1}^{\infty} \{x : F(x) - F(x-) > n^{-1}\}$$

と表せる．ここで，$F(+\infty) - F(-\infty) = 1$ より，任意の $\varepsilon > 0$ に対して，$F(x) - F(x-) > \varepsilon$ となる x の個数は $1/\varepsilon$ 以下である．よって，D_F は高々可算である． □

$C_b(\mathbb{R})$ を $\mathbb{R}$ 上の有界連続関数の全体とする．

Theorem 6.1. X, X_n を r.v.'s とする．このとき，

$$X_n \stackrel{d}{\to} X \Leftrightarrow \lim_n E[g(X_n)] = E[g(X)] \ \forall g \in C_b(\mathbb{R}).$$

Proof. X, X_n の d.f. をそれぞれ F, F_n とおく．

($\Rightarrow$). $\varepsilon > 0, g \in C_b(\mathbb{R})$ を任意に固定する．$|g| \le 1$ と仮定してよい．F の連続点 $a < b$ を

$$\varepsilon \ge P(X \in (a, b]^c) = 1 - (F(b) - F(a))$$

となるように選ぶと，$F_n(a) \to F(a), F_n(b) \to F(b)$ であるので，十分大きな n に対して，

$$2\varepsilon \ge 1 - (F_n(b) - F_n(a)) = P(X_n \in (a, b]^c)$$

となる．ここで，点列 $a = a_1 < a_2 < \cdots < a_N < a_{N+1} = b$ を，(1) $a_2, \dots, a_N$ は F の連続点であって，(2) $|g(x) - g(y)| \le \varepsilon$ ($\forall x, y \in [a_i, a_{i+1}], \forall i = 1, \dots, N$) となるように選ぶ．$F$

[30] 計量経済学においてはパラメトリックな仮定をなるべく避ける傾向にあるので，かなりの部分を漸近理論に頼ることになる．

の不連続点は高々可算個であって，g は区間 $[a,b]$ 上で一様連続であるから，$a_1,\dots,a_{N+1}$ をこのように選ぶことは可能である．このとき，

$$g_\varepsilon(x)=\sum_{i=1}^N g(a_i)I_{(a_i,a_{i+1}]}(x)$$

とおくと，

$$|g(x)-g_\varepsilon(x)|\le\sum_{i=1}^N|g(x)-g(a_i)|I_{(a_i,a_{i+1}]}(x)\le\varepsilon,\ \forall x\in(a,b]$$

となる．従って，積分区間を $(a,b]$ と $(a,b]^c$ に分けて，

$$\begin{aligned}&|E[g(X_n)]-E[g_\varepsilon(X_n)]|\\&\quad\le E[|g(X_n)-g_\varepsilon(X_n)|I_{(a,b]}(X_n)]+E[|g(X_n)-g_\varepsilon(X_n)|I_{(a,b]^c}(X_n)]\\&\quad\le\varepsilon+2E[I_{(a,b]^c}(X_n)]=\varepsilon+2P(X_n\in(a,b]^c)\le5\varepsilon\end{aligned}$$

を得る．同様にして，

$$|E[g(X)]-E[g_\varepsilon(X)]|\le3\varepsilon$$

を得る．さらに，$F_n(a_i)\to F(a_i)$ $(\forall i=1,\dots,N+1)$ だから，

$$\begin{aligned}E[g_\varepsilon(X_n)]&=\sum_{i=1}^N g(a_i)E[I_{(a_i,a_{i+1}]}(X_n)]=\sum_{i=1}^N g(a_i)P(X_n\in(a_i,a_{i+1}])\\&=\sum_{i=1}^N g(a_i)\{F_n(a_{i+1})-F_n(a_i)\}\to\sum_{i=1}^N g(a_i)\{F(a_{i+1})-F(a_i)\}=E[f_\varepsilon(X)]\end{aligned}$$

となる．以上より

$$\limsup_n|E[g(X_n)]-E[g(X)]|\le8\varepsilon$$

を得る．

$(\Leftarrow)$．$g:\mathbb{R}\to\mathbb{R}$ を

$$g(x)=\begin{cases}1 & x<0\\1-x & 0\le x\le1\\0 & x>1\end{cases}$$

として，$y\in\mathbb{R},\varepsilon>0$ に対して，$g_{y,\varepsilon}(x)=g((x-y)/\varepsilon)$ とおくと，$I(x\le y)\le g_{y,\varepsilon}(x)\le I(x\le y+\varepsilon)$ である．ここで，$g_{y,\varepsilon}\in C_b(\mathbb{R})$ より，

$$\limsup_n F_n(y)\le\lim_n E[g_{y,\varepsilon}(X_n)]=E[g_{y,\varepsilon}(X)]\le F(y+\varepsilon)$$

であって，$\varepsilon \downarrow 0$ として，$\limsup_n F_n(y) \le F(y)$ を得る．同様にして，

$$\liminf_n F_n(y) \ge F(y-\varepsilon)$$

であって，y が F の連続点なら，$\varepsilon \downarrow 0$ として，$\liminf_n F_n(y) \ge F(y)$ を得る． □

分布収束のこの特徴づけから，次の連続写像定理が従う．

Corollary 6.1 (連続写像定理)**.** $X_n \xrightarrow{d} X$ なら，任意の連続関数 $g : \mathbb{R} \to \mathbb{R}$ に対して，$g(X_n) \xrightarrow{d} g(X)$ となる．

証明を保留していた連続性定理を証明しよう．そのステートメントを再掲する．

Theorem 6.2 (連続性定理)**.** X の特性関数を φ とし，X_n の特性関数を φ_n とする．このとき，

$$X_n \xrightarrow{d} X \Leftrightarrow \lim_n \varphi_n(t) = \varphi(t)\ \forall t \in \mathbb{R}.$$

Proof. $\Rightarrow$ は $\sin x$ と $\cos x$ がともに有界連続関数であることから従う．$\Leftarrow$ を証明する．証明を 2 つのステップに分割する．$g : \mathbb{R} \to \mathbb{R}$ に対して，$\|g\|_\infty = \sup_{x\in\mathbb{R}} |g(x)|$ と定める．

ステップ 1． 有界区間の外側では 0 になる連続関数 $g : \mathbb{R} \to \mathbb{R}$ に対して，

$$\lim_n E[g(X_n)] = E[g(X)]$$

を示す．Lemma 1.11 より，$\sigma > 0$ に対して $Z^\sigma \sim N(0, \sigma^2)$ を X と独立とすると，

$$E[g(X + Z^\sigma)] = \frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{g}(t) e^{-\sigma^2 t^2/2} \varphi(-t) dt$$

となる．ここで，

$$\widehat{g}(t) = \int_{-\infty}^{\infty} g(x) e^{itx} dx$$

である．さらに，g は一様連続だから，

$$\forall \varepsilon > 0, \exists \delta > 0 \text{ s.t. } |x-y| < \delta \Rightarrow |g(x) - g(y)| < \varepsilon$$

より，

$$\begin{aligned}
&|E[g(X)] - E[g(X+Z^\sigma)]| \le E[|g(X) - g(X+Z^\sigma)|] \\
&\quad \le E[|g(X) - g(X+Z^\sigma)| I(|Z^\sigma| < \delta)] + E[|g(X) - g(X+Z^\sigma)| I(|Z^\sigma| \ge \delta)] \\
&\quad \le \varepsilon + 2\|g\|_\infty P(|Z^\sigma| \ge \delta)
\end{aligned}$$

となる．よって，

$$|E[g(X_n)] - E[g(X)]| \le \frac{1}{2\pi}\int_{-\infty}^{\infty} |\widehat{g}(t)| e^{-\sigma^2 t^2/2} |\varphi_n(-t) - \varphi(-t)| dt + 2\varepsilon + 4\|g\|_\infty P(|Z^\sigma| \ge \delta)$$

を得る．$\widehat{g}$ は有界であって，$\lim_n \varphi_n(t) = \varphi(t)\ \forall t \in \mathbb{R}$ だから，Lebesgue の優収束定理より，

$$\limsup_n |E[g(X_n)] - E[g(X)]| \le 2\varepsilon + 4\|g\|_\infty \underbrace{P(|Z^\sigma| \ge \delta)}_{=2\{1-\Phi(\delta/\sigma)\}}$$

を得る．あとは $\sigma \downarrow 0, \varepsilon \downarrow 0$ の順に極限をとって，$\lim_n E[g(X_n)] = E[g(X)]$ を得る．

ステップ 2. 有界連続関数 $g : \mathbb{R} \to \mathbb{R}$ に対して，$\lim_n E[g(X_n)] = E[g(X)]$ を示そう．これから $X_n \stackrel{d}{\to} X$ が従う．任意の $\varepsilon > 0$ に対して，$M > 1$ を十分大きくとって，$P(X \in [-M+1, M-1]) \ge 1 - \varepsilon$ とする．$\eta : \mathbb{R} \to \mathbb{R}$ を

$$\eta(x) = \begin{cases} 0 & x < -M \\ \text{linear} & -M \le x < -M+1 \\ 1 & -M+1 \le x \le M-1 \\ \text{linear} & M-1 < x \le M \\ 0 & x > M \end{cases}$$

と定めると，

$$I_{[-M+1,M-1]}(x) \le \eta(x) \le I_{[-M,M]}(x),\ \forall x \in \mathbb{R}$$

だから，ステップ 1 の結果より，

$$\lim_n E[\eta(X_n)] = E[\eta(X)] \ge P(X \in [-M+1, M-1]) \ge 1 - \varepsilon$$

となる．よって，

$$\begin{aligned} |E[g(X_n)] - E[g(X)]| &\le |E[g(X_n)\eta(X_n)] - E[g(X)\eta(X)]| \\ &\quad + 2\|g\|_\infty \underbrace{\{E[1-\eta(X_n)] + E[1-\eta(X)]\}}_{\le 2\varepsilon + o(1)} \end{aligned}$$

を得る．ここで，$g(x)\eta(x)$ は $[-M, M]$ の外側では 0 になる連続関数だから，ステップ 1 の結果より，$\lim_n E[g(X_n)] = E[g(X)]$ を得る． □

以前，2 次モーメントが有限な場合に大数の弱法則を証明したが，大数の弱法則は 1 次モーメントが有限なら成り立つ．

Theorem 6.3 (大数の弱法則)**.** F を $\mathbb{R}$ 上の d.f. とし，$X_1, \ldots, X_n \sim F$ i.i.d. として，$E[|X_1|] < \infty$ とする．このとき，$E[X_1] = \mu$ とおくと，$\overline{X} = n^{-1}\sum_{j=1}^n X_j \stackrel{P}{\to} \mu$.

Proof. $i = \sqrt{-1}$ とする．X_1 の特性関数を $\varphi(t)$ とおくと，$E[|X_1|] < \infty$ より，$\varphi(t)$ は微分可能であって，$\varphi'(0) = iE[X_1] = i\mu$ となる．よって，

$$\varphi(t) = 1 + i\mu t + tR(t),\ \lim_{t \to 0} R(t) = 0$$

と展開できる．ここで，$\overline{X}$ の特性関数を $\varphi_n(t)$ とおくと，$n \to \infty$ のとき，

$$\varphi_n(t) = E[e^{it\overline{X}}] = \prod_{j=1}^{n} E[e^{itX_j/n}] = \{\varphi(t/n)\}^n = \left(1 + \frac{i\mu t}{n} + \frac{t}{n}R(t/n)\right)^n \to e^{i\mu t}$$

となる．右辺は $X \equiv \mu$ の特性関数だから，連続性定理より，$\overline{X} \xrightarrow{d} \mu$ であって，μ は定数だから，$\overline{X} \xrightarrow{P} \mu$ を得る． □

Example 6.1 (Neyman-Pearson 検定の一致性)**.** $\theta \in \{\theta_0, \theta_1\}$ に対して，p_θ をある有限次元ユークリッド空間上の確率 (密度) 関数とし，$X_1, \ldots, X_n \sim p_\theta$ i.i.d. が得られているとする．ここで，$\theta_0 \neq \theta_1$ であって，p_{θ_0} と p_{θ_1} は相異なる分布として，

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta = \theta_1$$

という検定問題を考える．いま，

$$T_n = \frac{1}{n}\sum_{i=1}^{n} \log \frac{p_{\theta_1}(X_i)}{p_{\theta_0}(X_i)}$$

とおくと，Neyman-Pearson の補題より，与えられた $\alpha \in (0,1)$ に対して，水準 α の MP 検定は，

$$\delta_n(T_n) = \begin{cases} 1 & T_n > c_n \\ \gamma_n & T_n = c_n \\ 0 & T_n < c_n \end{cases}$$

で与えられるのであった．ここで，c_n は T_n の θ_0 のもとでの $(1-\alpha)$ 分位点であって，$P_{\theta_0}(T_n = c_n) = 0$ なら $\gamma_n = 0$ であって，$P_{\theta_0}(T_n = c_n) > 0$ なら，

$$\gamma_n = \frac{\alpha - P_{\theta_0}(T_n > c_n)}{P_{\theta_0}(T_n = c_n)}$$

である．次の 2 つの条件のもとで，Neyman-Pearson 検定 $\delta_n(T_n)$ の $\theta = \theta_1$ のもとでの検出力が $n \to \infty$ のとき 1 に収束することを示そう．

- $\{x : p_{\theta_0}(x) > 0\} = \{x : p_{\theta_1}(x) > 0\}$.
- KL ダイバージェンス $D(p_{\theta_0} || p_{\theta_1}), D(p_{\theta_1} || p_{\theta_0})$ は有限.

ここで，KL ダイバージェンス $D(p_{\theta_0} || p_{\theta_1})$ は

$$D(p_{\theta_0} || p_{\theta_1}) = \int_{\{p_{\theta_0} > 0\}} p_{\theta_0}(x) \log \frac{p_{\theta_0}(x)}{p_{\theta_1}(x)} dx$$

と定義されるのであった (確率関数の場合は，積分を和に取り替える)．p_{θ_0} と p_{θ_1} は相異なる分布だったから，$D(p_{\theta_0} || p_{\theta_1}) > 0, D(p_{\theta_1} || p_{\theta_0}) > 0$ である．いま，大数の弱法則より，

$\theta = \theta_0$ のもとで，$T_n \xrightarrow{P} -D(p_{\theta_0}||p_{\theta_1}) < 0$ となるから，ある $\varepsilon > 0$ が存在して，十分大きな n に対して，$c_n \le -\varepsilon$ となる．一方，$\theta = \theta_1$ のもとでは，$T_n \xrightarrow{P} D(p_{\theta_1}||p_{\theta_0}) > 0$ となるから，$n \to \infty$ のとき，

$$\beta_{\delta_n}(\theta_1) = E_{\theta_1}[\delta_n(T_n)] \ge P_{\theta_1}(T_n > c_n) \ge P_{\theta_1}(T_n > -\varepsilon) \to 1$$

となる．定義より，$\beta_{\delta_n}(\theta_1) \le 1$ だから，$\lim_n \beta_{\delta_n}(\theta_1) = 1$ を得る．

一般に，対立仮説をみたすパラメータの各点で，検出力が $n \to \infty$ のとき 1 に収束するとき，その検定は一致性をもつといわれる．上の議論は適当な仮定のもとで，Neyman-Pearson 検定が一致性をもつことを示している．

Example 6.2. F を $\mathbb{R}$ 上の d.f. とし，$X_1, \ldots, X_n \sim F$ i.i.d. とする．いま，$k \ge 2$ を正整数として，$E[|X_1|^k] < \infty$ を仮定する．このとき，F の k 次中心化モーメント $\mu_k = E[(X_1 - \mu)^k]$ の推定を考える．ここで，$\mu = E[X_1]$ である．$k = 2$ なら，$\mu_2 = \mathrm{Var}(X_1)$ である．

$$\widehat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k$$

という推定量を考える．$\widehat{\mu}_k$ が μ_k の一致推定量であることを示そう．$\ell = 1, \ldots, k$ に対して，$\widetilde{\mu}_\ell = n^{-1}\sum_{i=1}^{n}(X_i - \mu)^\ell$ とおくと，大数の弱法則より，$\widetilde{\mu}_\ell \xrightarrow{P} \mu_\ell$ である．ここで，2 項定理より，$\widehat{\mu}_k$ は

$$\widehat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu + \mu - \overline{X})^k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu - \widetilde{\mu}_1)^k = \sum_{\ell=0}^{k}\binom{k}{\ell}(-1)^\ell \widetilde{\mu}_{k-\ell}\widetilde{\mu}_1^\ell$$

と表せる．$\mu_1 = 0$ だから，Slutsky の補題より，$\sum_{\ell=1}^{k}\binom{k}{\ell}(-1)^\ell \widetilde{\mu}_{k-\ell}\widetilde{\mu}_1^\ell \xrightarrow{P} 0$ である．よって，再び Slutsky の補題より，$\widehat{\mu}_k \xrightarrow{P} \mu_k$ を得る．

F を $\mathbb{R}$ 上の d.f. として，$X_1, \ldots, X_n \sim F$ i.i.d. とする．$E[X_1^2] < \infty$ と仮定して，$\mu = E[X_1], \sigma^2 = \mathrm{Var}(X_1)$ とおき，$\sigma = \sqrt{\sigma^2} > 0$ とする．このとき，CLT より，$\sqrt{n}(\overline{X} - \mu)/\sigma \xrightarrow{d} N(0,1)$ となる．$N(0,1)$ の d.f. Φ は連続だから，Pólya の定理より，

$$\sup_{x \in \mathbb{R}}\left|P\left\{\sqrt{n}(\overline{X} - \mu)/\sigma \le x\right\} - \Phi(x)\right| \to 0$$

となる．$E[|X_1|^3] < \infty$ なら収束のスピードは $O(n^{-1/2})$ である．これは次の Berry-Esseen の定理から従う．

Theorem 6.4 (Berry-Esseen)**.** F を $\mathbb{R}$ 上の d.f. とし，$X_1, \ldots, X_n \sim F$ i.i.d. として，$E[|X_1|^3] < \infty$ を仮定する．また，$\mu = E[X_1], \sigma^2 = \mathrm{Var}(X_1)$ とおき，$\sigma = \sqrt{\sigma^2} > 0$ とする．このとき，

$$\sup_{x \in \mathbb{R}}\left|P\left\{\sqrt{n}(\overline{X} - \mu)/\sigma \le x\right\} - \Phi(x)\right| \le \frac{AE[|X_1 - \mu|^3]}{\sqrt{n}\sigma^3}$$

が成り立つ．ここで，A は絶対定数である．

Remark 6.1. $E[|X_1-\mu|^3] \le 4(E[|X_1|^3]+|\mu|^3)$ であって，$x \mapsto |x|^3$ は凸関数だから，Jensen の不等式より，$|\mu|^3 \le E[|X_1|^3]$ である．よって，$E[|X_1-\mu|^3] \le 8E[|X_1|^3]$ を得る．

Berry-Esseen の定理の証明は相当の労力を要するので，講義ノートでは省略する．Chung (2001, Section 7.4) か Stroock (2011, Section 2.2) を参照せよ．しかし，より粗いバウンド

$$\sup_{x\in\mathbb{R}} \left| P\left\{\sqrt{n}(\overline{X}-\mu)/\sigma \le x\right\} - \Phi(x)\right| \le A\left(\frac{E[|X_1-\mu|^3]}{\sqrt{n}\sigma^3}\right)^{1/4} \tag{*}$$

を証明するのはそれほど大変ではない．6.5 節を参照せよ．なお，以降の議論で Berry-Esseen の定理を使う箇所があるが，(*) のバウンドでも十分である．

Example 6.3. Berry-Esseen の定理のバウンドのオーダーは一様には改善できない．すなわち，

$$\liminf_n \sqrt{n}\left| P\left\{\sqrt{n}(\overline{X}-\mu)/\sigma \le x\right\} - \Phi(x)\right| > 0$$

となるような分布 F が存在する．例えば，$X_1,\dots,X_{2n}$ を i.i.d. であって，$P(X_i=-1) = P(X_i=1)=1/2$ とし，$S_{2n}=\sum_{i=1}^{2n} X_i$ とおくと，$S_{2n}=0$ となるのは $X_1,\dots,X_{2n}$ のうち n 個が 1 で残りの n 個が -1 のときだから，

$$P(S_{2n}=0) = \binom{2n}{n} 2^{-2n}$$

である．ここで，$+\infty$ に発散する数列 a_n, b_n に対して，

$$a_n \sim b_n \Leftrightarrow \lim_n \frac{a_n}{b_n} = 1$$

と書くと，Stirling の公式より，

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

だから，

$$P(S_{2n}=0) \sim \frac{1}{\sqrt{2\pi n}}$$

である．いま，$P(S_{2n}=0) = P(S_{2n}\le 0) - P(S_{2n}<0) = P(S_{2n}\le 0) - \{1-P(S_{2n}\ge 0)\}$ であって，$S_{2n} \overset{d}{=} -S_{2n}$ だから，$P(S_{2n}=0) = 2P(S_{2n}\le 0)-1 = 2\{P(S_{2n}/\sqrt{2n}\le 0)-\Phi(0)\}$ である．以上より，

$$\lim_n \sqrt{2n}\{P(S_{2n}/\sqrt{2n}\le 0)-\Phi(0)\} = \frac{1}{2\sqrt{\pi}}$$

を得る．

Example 6.4 (ブートストラップ CLT)**.** $\widehat{F}_n(x) = n^{-1}\sum_{i=1}^n I(X_i \le x)$ として，$X_1,\dots,X_n$ を与えたとき，

$$X_1^*,\dots,X_n^* \sim \widehat{F}_n \text{ i.i.d.}$$

とする．P^* をブートストラップ標本に関する確率とする．$E[|X_1|^3] < \infty, \sigma > 0$ のとき，

$$\sup_{x \in \mathbb{R}} |P^*\{\sqrt{n}(\overline{X}^* - \overline{X}) \le x\} - \Phi(x/\sigma)| \xrightarrow{P} 0 \tag{**}$$

を示そう．ここで，

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2$$

とおく．$E^*[\cdot]$ をブートストラップ標本に関する期待値とすると，

$$E^*[X_i^*] = \overline{X},\ E^*[(X_i^* - \overline{X})^2] = \widehat{\sigma}^2,\ E^*[|X_i^*|^3] = n^{-1}\sum_{j=1}^n |X_j|^3$$

だから，Berry-Esseen の定理より，$\widehat{\sigma} = \sqrt{\widehat{\sigma}^2} > 0$ のとき，

$$\sup_{x \in \mathbb{R}} |P^*\{\sqrt{n}(\overline{X}^* - \overline{X}) \le x\} - \Phi(x/\widehat{\sigma})| \le \frac{8An^{-1}\sum_{i=1}^n |X_i|^3}{\sqrt{n}\widehat{\sigma}^3}$$

となる．ここで，$\widehat{\sigma} \xrightarrow{P} \sigma, n^{-1}\sum_{i=1}^n |X_i|^3 \xrightarrow{P} E[|X_1|^3]$ より，右辺 $\xrightarrow{P} 0$ である．さらに，

$$B = \frac{1}{\sqrt{2\pi}}\sup_{y \ge 0} ye^{-y^2/2} = \frac{e^{-1/2}}{\sqrt{2\pi}}$$

とおくと，

$$\left|\frac{\partial \Phi(x/\sigma)}{\partial \sigma}\right| \le \frac{|x|}{\sqrt{2\pi}\sigma^2} e^{-\frac{x^2}{2\sigma^2}} \le \frac{B}{\sigma}$$

であることから，

$$\sup_{x \in \mathbb{R}} |\Phi(x/\widehat{\sigma}) - \Phi(x/\sigma)| \le \frac{B}{\min\{\sigma, \widehat{\sigma}\}}|\widehat{\sigma} - \sigma| \xrightarrow{P} 0.$$

以上より，(**) が示された．

確率ベクトルの収束

$\mathbb{R}^k$ の標準ノルムを $\|x\| = \sqrt{x'x}, x \in \mathbb{R}^k$ と書く．$X, X^n, n = 1, 2, \ldots$ を k 次元の確率ベクトルとする．任意の $\varepsilon > 0$ に対して，$\lim_n P(\|X^n - X\| > \varepsilon) = 0$ となるとき，X^n は X に確率収束するといって，$X^n \xrightarrow{P} X$ と書く．明らかに，

$$X^n \xrightarrow{d} X \Leftrightarrow \|X^n - X\| \xrightarrow{P} 0 \Leftrightarrow X_j^n \xrightarrow{P} X_j\ \forall j = 1, \ldots, k$$

である．よって，多次元の確率ベクトルの確率収束を示すには，各座標の確率収束を示せばよい．従って，大数の弱法則は平均が有限な i.i.d. 確率ベクトル列に対しても成り立つ．

次に，k 次元の確率ベクトル $X=(X_1,\ldots,X_k)'$ に対して，その (同時) 分布関数は

$$F(x)=F(x_1,\ldots,x_k)=P(X_1\le x_1,\ldots,X_k\le x_k),\ x=(x_1,\ldots,x_k)'\in\mathbb{R}^k$$

であった．X^n を k 次元確率ベクトル列とし，その d.f. を F_n とおく．ここで，F の任意の連続点 $x\in\mathbb{R}^k$ に対して，$\lim_n F_n(x)=F(x)$ となるとき，X^n は X に分布収束するといって，$X^n\xrightarrow{d}X$ or $X^n\xrightarrow{d}F$ と書く．

ここで，注意として，各 X_j^n が分布収束していても，ベクトルとして X^n が分布収束するとは限らない (逆は後述する連続写像定理から成り立つ)．例えば，$k=2$ の場合に，$U\sim U(0,1)$ として，

$$X_1^n=U,\quad X_2^n=\begin{cases}U & n\text{ が奇数}\\ 1-U & n\text{ が偶数}\end{cases}$$

とおくと，$X_1^n\sim U(0,1), X_2^n\sim U(0,1)$ であるが，(X_1^n,X_2^n) は明らかに分布収束しない．

Example 6.5. $X_1,\ldots,X_n\sim U(0,1)$ i.i.d. とすると，$0<u<v<1$ に対して，

$$P(X_{(1)}\le u,X_{(n)}\ge v)=1-(1-u)^n-v^n+(v-u)^n$$

となる (演習問題)．このとき，$x,y>0$ に対して，$u=u_n=x/n, v_n=1-y/n$ とおくと，n が十分大きいとき，$0<u_n<v_n<1$ だから，

$$\begin{aligned}&P(X_{(1)}\le x/n,X_{(n)}\ge 1-y/n)=1-(1-x/n)^n-(1-y/n)^n+(1-(x+y)/n)^n\\&\quad\to 1-e^{-x}-e^{-y}+e^{-x-y}=(1-e^{-x})(1-e^{-y}).\end{aligned}$$

ここで，

$$X_{(1)}\le x/n\ \&\ X_{(n)}\ge 1-y/n\Leftrightarrow nX_{(1)}\le x\ \&\ n(1-X_{(n)})\le y$$

だから，独立に $Ex(1)$ に従う r.v.'s V,W に対して，

$$n(X_{(1)},1-X_{(n)})\xrightarrow{d}(V,W)$$

を得る．従って，$X_{(1)}$ と $X_{(n)}$ は有限の n では独立でないが，漸近的には独立になる．

1 次元の r.v.'s の分布収束について成り立つ多くの結果は，多次元の確率ベクトルに対しても成り立つ．例えば，X,X^n,Y^n を k 次元の確率ベクトルとし，$c\in\mathbb{R}^k$ を定数ベクトルとする．このとき，次が成り立つ (証明は 1 次元の場合と同様である)．

- 確率収束と分布収束の関係．

$$X^n\xrightarrow{P}X\Rightarrow X^n\xrightarrow{d}X,\quad X^n\xrightarrow{d}c\Rightarrow X^n\xrightarrow{P}c.$$

- Slutsky の補題．

$$X^n\xrightarrow{d}X\ \&\ Y^n\xrightarrow{P}c\Rightarrow X^n+Y^n\xrightarrow{d}X+c.$$

- 連続写像定理．$C(\mathbb{R}^k) = \{g : g : \mathbb{R}^k \to \mathbb{R}, g \text{ は連続 }\}$ とおく．このとき，

$$X^n \xrightarrow{d} X \Rightarrow \forall g \in C(\mathbb{R}^k), g(X^n) \xrightarrow{d} g(X).$$

- 連続性定理．X, X_n の特性関数をそれぞれ $\varphi(t), \varphi_n(t)$ とおくと[31],

$$X^n \xrightarrow{d} X \Leftrightarrow \lim_n \varphi_n(t) = \varphi(t)\ \forall t \in \mathbb{R}^k.$$

連続性定理より，次の Cramér-Wold 法を得る．

Lemma 6.2 (Cramér-Wold 法)**.** $X^n \xrightarrow{d} X \Leftrightarrow t'X^n \xrightarrow{d} t'X\ \forall t \in \mathbb{R}^k$.

Proof. ($\Rightarrow$). $x \mapsto t'x$ は連続なので，連続写像定理より，$t'X^n \xrightarrow{d} t'X$ を得る．

($\Leftarrow$). 逆に，任意の $t \in \mathbb{R}^k$ に対して，$t'X^n \xrightarrow{d} t'X$ を仮定する．このとき，連続性定理より，$E[e^{it'X^n}] \to E[e^{it'X}]$ となる．$t \in \mathbb{R}^k$ は任意だから，これは $\varphi_n(t) \to \varphi(t)\ \forall t \in \mathbb{R}^k$ を意味する．再び連続性定理より，$X^n \xrightarrow{d} X$ を得る． □

Example 6.6. X^n を k 次元確率ベクトルとし，$X^n \xrightarrow{d} X \sim N(\mu, \Sigma)$ とする．このとき，$m \times k$ 行列 A に対して，

$$AX^n \xrightarrow{d} N(A\mu, A\Sigma A')$$

となる．なぜなら，任意の $t \in \mathbb{R}^m$ に対して，Cramér-Wold 法より，$t'AX^n = (A't)'X^n \xrightarrow{d} (A't)'X = t'AX$ となる．よって，再び Cramér-Wold 法より，$AX^n \xrightarrow{d} AX \sim N(A\mu, A\Sigma A')$ を得る．

Cramér-Wold 法より，多次元の確率ベクトルの分布収束の証明は 1 次元の r.v.'s の分布収束の証明に帰着させることができる．Cramér-Wold 法より，次の多変量 CLT が直ちに従う．

Theorem 6.5 (多変量 CLT)**.** F を $\mathbb{R}^k$ 上の d.f. とし，$X_1, \dots, X_n \sim F$ i.i.d. とする．また，$E[\|X_1\|^2] < \infty$ と仮定して，$\mu = E[X_1], \Sigma = \mathrm{Var}(X_1)$ とおく．このとき，

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu) \xrightarrow{d} N(0, \Sigma).$$

Example 6.7. $Y^n = (Y_1^n, \dots, Y_k^n)' \sim Mn(n, p_1, \dots, p_k)$ とすると，$X_i = (X_{i,1}, \dots, X_{i,k})' \sim Mn(1, p_1, \dots, p_k)$ i.i.d. に対して，

$$Y^n \stackrel{d}{=} \sum_{i=1}^n X_i$$

[31] k 次元確率ベクトル X の特性関数は $\varphi(t) := E[e^{it'X}], t \in \mathbb{R}^k, i = \sqrt{-1}$ と定義されるのであった．

であるから，$p=(p_1,\dots,p_k)'$ とおくと，多変量 CLT より，

$$\frac{1}{\sqrt{n}}(Y^n-np)\xrightarrow{d}N(0,\Sigma)$$

を得る．ここで，

$$\Sigma=\begin{pmatrix} p_1(1-p_1) & -p_1p_2 & \cdots & -p_1p_k \\ -p_2p_1 & p_2(1-p_2) & \cdots & -p_2p_k \\ \vdots & & \ddots & \vdots \\ -p_kp_1 & -p_kp_2 & \cdots & p_k(1-p_k) \end{pmatrix}$$

である．

この結果と連続写像定理を使って，Pearson の χ^2 検定統計量

$$\chi_n^2=\sum_{j=1}^k\frac{(Y_j^n-np_j)^2}{np_j}$$

が $\chi^2(k-1)$ に分布収束することを (直接) 示そう．$\widetilde{Y}_j^n=Y_j^n/\sqrt{p_j}, \widetilde{Y}^n=(\widetilde{Y}_1^n,\dots,\widetilde{Y}_k^n)', q=(\sqrt{p}_1,\dots,\sqrt{p}_k)'$ とおくと，

$$\chi_n^2=\{n^{-1/2}(\widetilde{Y}^n-nq)\}'\{n^{-1/2}(\widetilde{Y}^n-nq)\}.$$

ここで，

$$\frac{1}{\sqrt{n}}(\widetilde{Y}^n-nq)\xrightarrow{d}\widetilde{Y}\sim N(0,\widetilde{\Sigma}),$$

$$\widetilde{\Sigma}=\begin{pmatrix} 1-p_1 & -\sqrt{p_1}\sqrt{p_2} & \cdots & -\sqrt{p_1}\sqrt{p_k} \\ -\sqrt{p_2}\sqrt{p_1} & 1-p_2 & \cdots & -\sqrt{p_2}\sqrt{p_k} \\ \vdots & & \ddots & \vdots \\ -\sqrt{p_k}\sqrt{p_1} & -\sqrt{p_k}\sqrt{p_2} & \cdots & 1-p_k \end{pmatrix}=I_k-qq'.$$

$\mathbb{R}^k\ni y\mapsto y'y$ は連続だから，連続写像定理より，

$$\chi_n\xrightarrow{d}\widetilde{Y}'\widetilde{Y}$$

となる．さらに，$q'q=\sum_{j=1}^k p_j=1$ より，$k\times(k-1)$ 行列 R を (R,q) が直交行列になるように選ぶと，

$$I_k=(R,q)\begin{pmatrix} R' \\ q' \end{pmatrix}=RR'+qq'$$

より，$I_k-qq'=RR'$．従って，$Z\sim N(0,I_{k-1})$ に対して，$\widetilde{Y}\overset{d}{=}RZ$ だから，

$$\widetilde{Y}'\widetilde{Y}\overset{d}{=}Z'\underbrace{R'R}_{=I_{k-1}}Z=Z'Z\sim\chi^2(k-1)$$

を得る．

Example 6.8. $X_1,\dots,X_n\sim U(0,1)$ i.i.d. とする．このとき，独立な $V\sim N(0,1/12), W\sim Ex(1)$ に対して，

$$\left(\sqrt{n}(\overline{X}-1/2), n(1-X_{(n)})\right)\xrightarrow{d}(V,W)$$

となる．このことを示そう．$V_n=n^{-1/2}\sum_{i=1}^{n-1}X_{(i)}-\sqrt{n}/2, W_n=n(X_{(n)}-1)$ とおくと，

$$|\sqrt{n}(\overline{X}-1/2)-V_n|\le X_{(n)}/\sqrt{n}\le 1/\sqrt{n}\to 0$$

だから，

$$(V_n,W_n)\xrightarrow{d}(V,W)$$

を示せばよい．ここで，$v\in\mathbb{R}, w>0$ に対して，

$$P(V_n\le v, W_n\le w)=E[\underbrace{E[I(V_n\le v)\mid X_{(n)}]}_{=P(V_n\le v|X_{(n)})}I(W_n\le w)].$$

いま，$X_{(n)}$ を与えたときの $X_{(1)},\dots,X_{(n-1)}$ の条件付き分布は，$U(0,X_{(n)})$ からのサイズ $(n-1)$ の独立標本の順序統計量の同時分布に等しい．ここで，$U(0,X_{(n)})$ の平均，分散，3 次モーメントはそれぞれ，$X_{(n)}/2, X_{(n)}^2/12, X_{(n)}^3/4$ であって，

$$c_n=\sqrt{n/(n-1)},\ \Delta_n=\sqrt{n}/2-(n-1)X_{(n)}/(2\sqrt{n})$$

とおくと，

$$V_n\le v\Leftrightarrow (n-1)^{-1/2}\sum_{i=1}^{n-1}(X_{(i)}-X_{(n)}/2)\le c_n(v+\Delta_n)$$

だから，Berry-Esseen の定理より，ある絶対定数 $B>0$ が存在して，

$$\left|P(V_n\le v\mid X_{(n)})-\Phi\left(\frac{c_n(v+\Delta_n)}{\sqrt{X_{(n)}^2/12}}\right)\right|\le\frac{B}{\sqrt{n-1}}$$

が成り立つ．ここで，$\sqrt{n}(1-X_{(n)})\xrightarrow{P}0$ より，

$$\Delta_n=\frac{\sqrt{n}}{2}(1-X_{(n)})+\frac{X_{(n)}}{2\sqrt{n}}\xrightarrow{P}0$$

だから，

$$\zeta_n:=\left|\Phi\left(\frac{c_n(v+\Delta_n)}{\sqrt{X_{(n)}^2/12}}\right)-\Phi(\sqrt{12}v)\right|\xrightarrow{P}0$$

である．以上より，

$$|E[P(V_n\le v\mid X_{(n)})I(W_n\le w)]-\Phi(\sqrt{12}v)P(W_n\le w)|\le E[\zeta_n]+\frac{B}{\sqrt{n}}$$

を得る．ここで，Φ は有界だから，$E[\zeta_n] \to 0$ となる[32]．従って，右辺 $\to 0$ である．あとは $W_n \xrightarrow{d} Ex(1)$ より，$P(W_n \le w) \to 1 - e^{-w}$ だから，

$$P(V_n \le v, W_n \le w) \to \Phi(\sqrt{12}v)(1 - e^{-w})$$

を得る．右辺は (V, W) の d.f. だから，求める結論を得る．

確率オーダー

X^n を k 次元確率ベクトルとする．$X^n \xrightarrow{P} 0$ のとき，$X^n = o_P(1)$ と書く．次に，任意の $\varepsilon > 0$ に対して，ある $M = M_\varepsilon > 0$ が存在して，$P(\|X^n\| > M) \le \varepsilon\ \forall n \ge 1$ となるとき，X^n は確率有界 (stochastically bounded) であるといって，$X^n = O_P(1)$ と書く．さらに，(1 次元の) r.v.'s R_n に対して，

$$\begin{aligned} X^n &= O_P(R_n) \quad \text{if} \quad X^n = R_n Y^n \ \&\ Y^n = O_P(1), \\ X^n &= o_P(R_n) \quad \text{if} \quad X^n = R_n Y^n \ \&\ Y^n = o_P(1) \end{aligned}$$

と定義する．

Lemma 6.3. $X^n \xrightarrow{d} X$ なら，$X^n = O_P(1)$ である．

Proof. $\varepsilon > 0$ を任意に固定する．$M_1 > 0$ を $\|X\|$ の d.f. の連続点であって，$P(\|X\| > M_1) < \varepsilon$ となるように選ぶ．$X^n \xrightarrow{d} X$ だったから，連続写像定理より，$\|X^n\| \xrightarrow{d} \|X\|$ であって，よって，

$$\lim_{n\to\infty} P(\|X^n\| > M_1) = P(\|X\| > M_1) < \varepsilon.$$

を得る．従って，$\exists N \in \mathbb{N}$ s.t. $P(\|X^n\| > M_1) < \varepsilon\ \forall n > N$. あとは，$M_2 > 0$ を

$$P(\|X^n\| > M_2) < \varepsilon\ \forall n = 1, \dots, N$$

となるように選んで，$M = \max\{M_1, M_2\}$ とすればよい． □

Remark 6.2. 逆は明らかに正しくないが，$X^n = O_P(1)$ なら X^n は分布収束する部分列をもつ (Prohorov の定理)．

確率オーダーは (通常の) オーダー記号と同様の規則に従う．例えば，

$$\begin{aligned} &o_P(1) + o_P(1) = o_P(1), \\ &O_P(1) + o_P(1) = O_P(1), \\ &\underbrace{o_P(1)}_{1\text{ 次元}}\underbrace{O_P(1)}_{k\text{ 次元}} = \underbrace{o_P(1)}_{k\text{ 次元}}, \end{aligned}$$

などが成り立つ．

[32] $\zeta_n \le 1$ だから，$E[\zeta_n] = E[\zeta_n I(\zeta_n \le \varepsilon)] + E[\zeta_n I(\zeta_n > \varepsilon)] \le \varepsilon + P(\zeta_n > \varepsilon)$ より，$\limsup_n E[\zeta_n] \le \varepsilon$. $\varepsilon > 0$ は任意だから，$\varepsilon \downarrow 0$ として，$\lim_n E[\zeta_n] = 0$ を得る．

Example 6.9 (Example 6.2 の続き)．Example 6.2 において，$E[X_1^{2k}] < \infty$ と仮定して，$\sqrt{n}(\widehat{\mu}_k - \mu_k)$ が正規分布に分布収束することを示す．CLT より，$\sqrt{n}\widetilde{\mu}_1$ は正規分布に分布収束するから，$\widetilde{\mu}_1 = O_P(n^{-1/2})$ である．よって，

$$\widehat{\mu}_k = \sum_{\ell=0}^{k} \binom{k}{\ell} (-1)^\ell \widetilde{\mu}_{k-\ell}\widetilde{\mu}_1^\ell = \widetilde{\mu}_k - k\widetilde{\mu}_{k-1}\widetilde{\mu}_1 + o_P(n^{-1/2})$$

と展開できる．さらに，大数の弱法則より，$\widetilde{\mu}_{k-1} = \mu_{k-1} + o_P(1)$ だから，$\widetilde{\mu}_{k-1}\widetilde{\mu}_1 = (\mu_{k-1} + o_P(1))\widetilde{\mu}_1 = \mu_{k-1}\widetilde{\mu}_1 + o_P(n^{-1/2})$ と展開できる．よって，

$$\widehat{\mu}_k - \mu_k = \widetilde{\mu}_k - \mu_k - k\mu_{k-1}\widetilde{\mu}_1 + o_P(n^{-1/2}) = (1, -k\mu_{k-1}) \begin{pmatrix} \widetilde{\mu}_k - \mu_k \\ \widetilde{\mu}_1 \end{pmatrix} + o_P(n^{-1/2})$$

を得る．多変量 CLT より，

$$\sqrt{n} \begin{pmatrix} \widetilde{\mu}_k - \mu_k \\ \widetilde{\mu}_1 \end{pmatrix} \xrightarrow{d} N(0, \Sigma)$$

であるから，Slutsky の補題より，

$$\sqrt{n}(\widehat{\mu}_k - \mu_k) \xrightarrow{d} N\left(0, (1, -k\mu_{k-1})\Sigma \begin{pmatrix} 1 \\ -k\mu_{k-1} \end{pmatrix}\right)$$

を得る．ここで，2×2 行列 Σ は

$$\Sigma = \begin{pmatrix} \mu_{2k} - \mu_k^2 & \mu_{k+1} \\ \mu_{k+1} & \mu_2 \end{pmatrix}$$

と計算できるから，

$$(1, -k\mu_{k-1})\Sigma \begin{pmatrix} 1 \\ -k\mu_{k-1} \end{pmatrix} = \mu_{2k} - \mu_k^2 - 2k\mu_{k-1}\mu_{k+1} + k^2\mu_2\mu_{k-1}^2$$

である．

デルタ法

写像 $g = (g_1, \ldots, g_m)' : \mathbb{R}^k \to \mathbb{R}^m$ が $\theta \in \mathbb{R}^k$ において微分可能であるとは，$m \times k$ 行列 $\dot{g}(\theta)$ が存在して，

$$g(\theta + h) - g(\theta) = \dot{g}(\theta)h + \|h\|R(h), \ \lim_{h \to 0} R(h) = 0 \tag{*}$$

と展開できることをいう．各 g_j が θ の近傍で連続微分可能なら，g は θ において微分可能である．このとき，

$$\dot{g}(\theta) = \begin{pmatrix} \frac{\partial g_1(\theta)}{\partial \theta_1} & \cdots & \frac{\partial g_1(\theta)}{\partial \theta_k} \\ \vdots & & \vdots \\ \frac{\partial g_m(\theta)}{\partial \theta_1} & \cdots & \frac{\partial g_m(\theta)}{\partial \theta_k} \end{pmatrix}$$

である．

Theorem 6.6. $Y^n, n = 1, 2, \ldots$ を k 次元の確率ベクトルとし，ある $\theta \in \mathbb{R}^k$ と $k \times k$ 半正定値対称行列 Σ に対して，$n \to \infty$ のとき $\sqrt{n}(Y^n - \theta) \xrightarrow{d} N(0, \Sigma)$ とする．このとき，θ において微分可能な写像 $g : \mathbb{R}^k \to \mathbb{R}^m$ に対して，

$$\sqrt{n}(g(Y^n) - g(\theta)) \xrightarrow{d} N(0, \dot{g}(\theta)\Sigma\dot{g}(\theta)')$$

となる．

Proof. (*) において $h = Y^n - \theta$ を代入して，

$$g(Y^n) - g(\theta) = \dot{g}(\theta)(Y^n - \theta) + \|Y^n - \theta\| R(Y^n - \theta).$$

$\sqrt{n}(Y^n - \theta) \xrightarrow{d} Z \sim N(0, \Sigma)$ であって，$y \mapsto \|y\|$ は連続だから，連続写像定理より，$\sqrt{n}\|Y^n - \theta\| \xrightarrow{d} \|Z\|$ となる．よって，$\|Y^n - \theta\| = n^{-1/2}(\sqrt{n}\|Y^n - \theta\|) = n^{-1/2}O_P(1)$ である．さらに，$\lim_{h \to 0} R(h) = 0$ より，

$$\forall \varepsilon > 0, \exists \delta > 0 \text{ s.t. } \|h\| < \delta \Rightarrow \|R(h)\| < \varepsilon$$

だから，

$$P\{\|R(Y_n - \theta)\| \geq \varepsilon\} \leq P\{\|Y^n - \theta\| \geq \delta\} \to 0$$

となる．以上より，

$$\|Y^n - \theta\| R(Y^n - \theta) = n^{-1/2}O_P(1)o_P(1) = n^{-1/2}o_P(1)$$

となるから，

$$\sqrt{n}(g(Y^n) - g(\theta)) = \dot{g}(\theta)\sqrt{n}(Y^n - \theta) + o_P(1) \xrightarrow{d} N(0, \dot{g}(\theta)\Sigma\dot{g}(\theta)')$$

を得る． □

Example 6.10. F を $\mathbb{R}$ 上の d.f. とし，$X_1, \ldots, X_n \sim F$ i.i.d. とする．$E[X_1^4] < \infty$ と仮定して，$E[X_1] = \mu, \sigma^2 = \mathrm{Var}(X_1)$ とおく．

$$\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i,\ \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X}^2$$

とおいて，$\sqrt{n}(\overline{X} - \mu, \widehat{\sigma}^2 - \sigma^2)'$ の極限分布を求めよう．いま，$Y_i = X_i^2, g(x, y) = (x, y - x^2)'$ とおくと，

$$\sqrt{n}\begin{pmatrix} \overline{X} - \mu \\ \widehat{\sigma}^2 - \sigma^2 \end{pmatrix} = \sqrt{n}(g(\overline{X}, \overline{Y}) - g(\mu, E[X_1^2])).$$

ここで，

$$\sqrt{n}\begin{pmatrix} \overline{X} - \mu \\ \overline{Y} - E[X_1^2] \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & E[X_1^3] - \mu E[X_1^2] \\ E[X_1^3] - \mu E[X_1^2] & E[X_1^4] - \sigma^4 \end{pmatrix}\right),$$

$$\dot{g}(\mu, E[X_1^2]) = \begin{pmatrix} 1 & 0 \\ -2\mu & 1 \end{pmatrix}.$$

よって，デルタ法より，

$$\sqrt{n}\begin{pmatrix}\overline{X}-\mu\\ \widehat{\sigma}^2-\sigma^2\end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\sigma^2 & E[(X_1-\mu)^3]\\ E[(X_1-\mu)^3] & \mathrm{Var}((X_1-\mu)^2)\end{pmatrix}\right)$$

を得る．

Example 6.11. g が微分可能でない場合は，$\sqrt{n}(g(Y^n)-g(\theta))$ が正規分布に分布収束するとは限らない．例えば，F を $\mathbb{R}$ 上の d.f. とし，$X_1,\ldots,X_n \sim F$ i.i.d. として，$E[X_1^2]<\infty$ とする．いま，$E[X_1]=\theta$ とし，$g(\theta)=\max\{\theta,0\}$ とする．このとき，$\sigma^2=\mathrm{Var}(X_1)$ とおくと，$\theta\neq 0$ なら，$\sqrt{n}(g(\overline{X})-g(\theta)) \xrightarrow{d} N(0,\sigma^2)$ となるが，$\theta=0$ なら，$\sqrt{n}(g(\overline{X})-g(\theta)) = \max\{\sqrt{n}\overline{X},0\}$ であって，$x\mapsto \max\{x,0\}$ は連続だから，連続写像定理より，

$$\max\{\sqrt{n}\overline{X},0\} \xrightarrow{d} \max\{Z,0\}, \quad Z\sim N(0,\sigma^2)$$

となる．

6.2 分位点の推定

F を $\mathbb{R}$ 上の d.f. とし，$X_1,\ldots,X_n\sim F$ i.i.d. とする．$u\in(0,1)$ に対して，F の u 分位点 $\theta_u = F^{\leftarrow}(u)$ の推定を考える．経験分布関数を $\widehat{F}_n(x)=n^{-1}\sum_{i=1}^n I(X_i\le x)$ とおくと，θ_u の標準的な推定量は標本 u 分位点

$$\widehat{\theta}_u = \widehat{F}_n^{\leftarrow}(u)$$

である．本節の目標は次の定理の証明である．

Theorem 6.7. F は密度関数 f をもち，f は θ_u で正かつ連続とする．このとき，

$$\sqrt{n}(\widehat{\theta}_u-\theta_u) \xrightarrow{d} N\left(0, \frac{u(1-u)}{f(\theta_u)^2}\right)$$

となる．

Proof. $\Delta(x)=\sqrt{F(x)\{1-F(x)\}}$ とおく．次の 2 つの事実を使う．

- $\widehat{\theta}_u\le x \Leftrightarrow u\le \widehat{F}_n(x)$.
- Berry-Esseen の定理より，$0<F(x)<1$ となる $x\in\mathbb{R}$ に対して，

$$\sup_{y\in\mathbb{R}}\left|P\{\sqrt{n}(\widehat{F}_n(x)-F(x))\le y\}-\Phi(y/\Delta(x))\right| \le \frac{A}{\sqrt{n}\Delta(x)^3}$$

  が成り立つ．ここで，$|I(X_i\le x)-F(x)|\le 1$ を使った．

$x \in \mathbb{R}$ に対して，

$$\begin{aligned}\sqrt{n}(\widehat{\theta}_u - \theta_u) \le x &\Leftrightarrow u \le \widehat{F}_n(\theta_u + x/\sqrt{n}) \\ &\Leftrightarrow \sqrt{n}\{\widehat{F}_n(\theta_u + x/\sqrt{n}) - F(\theta_u + x/\sqrt{n})\} \ge \sqrt{n}\{u - F(\theta_u + x/\sqrt{n})\}.\end{aligned}$$

よって，$y_n = \sqrt{n}\{u - F(\theta_u + x/\sqrt{n})\}$ とおくと，

$$P\{\sqrt{n}(\widehat{\theta}_u - \theta_u) \le x\} = P\left\{\sqrt{n}\{\widehat{F}_n(\theta_u + x/\sqrt{n}) - F(\theta_u + x/\sqrt{n})\} \ge y_n\right\}.$$

ここで，

$$\begin{aligned}&\left|P\left\{\sqrt{n}\{\widehat{F}_n(\theta_u + x/\sqrt{n}) - F(\theta_u + x/\sqrt{n})\} \ge y_n\right\} - \{1 - \Phi(y_n/\Delta(\theta_u + x/\sqrt{n}))\}\right| \\ &\quad \le \frac{A}{\sqrt{n}\Delta(\theta_u + x/\sqrt{n})^3}.\end{aligned}$$

いま，$n \to \infty$ のとき，

$$\Delta(\theta_u + x/\sqrt{n}) \to \Delta(\theta_u) = \sqrt{u(1-u)},\ y_n \to -f(\theta_u)x$$

であるから，$\Phi(-y) = 1 - \Phi(y)$ という事実を使って，

$$P\{\sqrt{n}(\widehat{\theta}_u - \theta_u) \le x\} \to \Phi\left(\frac{f(\theta_u)x}{\sqrt{u(1-u)}}\right)$$

を得る．右辺は

$$N\left(0, \frac{u(1-u)}{f(\theta_u)^2}\right)$$

の d.f. だから，定理の結論を得る． $\square$

Remark 6.3. 分位点は有限な平均や分散をもたない分布に対しても定義できる．さらに，標本分位点の漸近正規性はモーメント条件なしに成り立つ．

Example 6.12. f を Laplace 分布の密度関数とする：

$$f(u) = \frac{1}{2}e^{-|u|},\ u \in \mathbb{R}.$$

$\theta \in \mathbb{R}$ に対して，$X_1, \dots, X_n \sim f(\bullet - \theta)$ i.i.d. とすると，対数尤度は

$$\ell_n(\theta) = -n \log 2 - \sum_{i=1}^n |X_i - \theta|$$

であって，標本メディアンは $\ell_n(\theta)$ を最大化する．n が偶数のときは対数尤度の最大化点は一意でないが，いずれにせよ $\widehat{\theta} = \widehat{F}_n^{\leftarrow}(1/2)$ を MLE に選ぶと，θ は $f(\bullet - \theta)$ のメディアンであるから，

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} N(0,1)$$

となる．

さて，ブートスラップの節において，分位点に対するパーセンタイル・ブートストラップ CI を考察したが，その漸近的な正当化は保留していた．ブートストラップ標本

$$X_1^*, \ldots, X_n^* \sim \widehat{F}_n \text{ i.i.d.}$$

に対して，$\widehat{\theta}_u^*$ をブートスラップ標本にもとづく標本 u 分位点とする．すなわち，

$$\widehat{F}_n^*(x) = \frac{1}{n}\sum_{i=1}^n I(X_i^* \le x),\ x \in \mathbb{R}$$

とおくと，

$$\widehat{\theta}_u^* = \widehat{F}_n^{*\leftarrow}(u) = \inf\{x \in \mathbb{R} : \widehat{F}_n^*(x) \ge u\}$$

である．$u \in (0,1)$ は固定して，

$$\widehat{G}_n(t) = P^*\{\sqrt{n}(\widehat{\theta}_u^* - \widehat{\theta}_u) \le t\}$$

とおく．いま，

$$\sigma_u^2 = \frac{u(1-u)}{f(\theta_u)^2}$$

とおくと，

$$\sqrt{n}(\widehat{\theta}_u - \theta_u) \xrightarrow{d} N(0, \sigma_u^2)$$

であって，$N(0,\sigma_u^2)$ の d.f. は $\Phi(t/\sigma_u)$ であるから，Theorem 5.3 より，パーセンタイル・ブートストラップ CI の漸近的な正当性を示すには，次の定理を示せば十分である．

Theorem 6.8. F は密度関数 f をもち，f は θ_u で正かつ連続とする．このとき，

$$\sup_{t \in \mathbb{R}} |\widehat{G}_n(t) - \Phi(t/\sigma_u)| \xrightarrow{P} 0.$$

従って，Example 5.6 に現れたパーセンタイル・ブートストラップ CI

$$\left[2\widehat{\theta}_u - \widehat{\zeta}_{u,\beta}, 2\widehat{\theta}_u - \widehat{\zeta}_{u,1-\alpha}\right]$$

は近似的に水準 $1-\alpha-\beta$ をもつ：

$$P\left\{\theta_u \in \left[2\widehat{\theta}_u - \widehat{\zeta}_{u,\beta}, 2\widehat{\theta}_u - \widehat{\zeta}_{u,1-\alpha}\right]\right\} \to 1-\alpha-\beta.$$

Proof. 前半のみ示せばよい．完全な証明は講義ノートのレベルを超えるので，略証のみ与える．標本 u 分位点 $\widehat{\theta}_u$ に対して，次の展開が成り立つ．

$$\sqrt{n}(\widehat{\theta}_u - \theta_u) = \frac{1}{f(\theta_u)}\sqrt{n}\{u - \widehat{F}_n(\theta_u)\} + o_P(1).$$

この展開は標本分位点に対する Bahadur 表現 とも呼ばれる．$\widehat{\theta}_u^*$ に対しても同様にして，

$$\sqrt{n}(\widehat{\theta}_u^* - \theta_u) = \frac{1}{f(\theta_u)}\sqrt{n}\{u - \widehat{F}_n^*(\theta_u)\} + o_P(1)$$

という展開が成り立つ．これらから，

$$\sqrt{n}(\widehat{\theta}_u^* - \widehat{\theta}_u) = \underbrace{\frac{1}{f(\theta_u)}\sqrt{n}\{\widehat{F}_n(\theta_u) - \widehat{F}_n^*(\theta_u)\}}_{=U_n^*} + o_P(1)$$

を得る．ここで，P^* のもとで，$I(X_i^* \le \theta_u), i = 1, \ldots, n$ が i.i.d. で平均 $\widehat{F}_n(\theta_u)$ と分散 $\widehat{F}_n(\theta_u)(1 - \widehat{F}_n(\theta_u)) = u(1-u) + o_P(1)$ をもつことと，Berry-Esseen の定理から，

$$\sup_{t \in \mathbb{R}} |P^*(U_n^* \le t) - \Phi(t/\sigma_u)| \xrightarrow{P} 0$$

を得る．$\Phi(t/\sigma_u)$ が連続なことを使って，求める結論を得る． □

Remark 6.4. 標本分位点に対する漸近正規性は，汎関数デルタ法 (functional delta method) によっても示すことができる．こちらのほうがよりモダンな証明方法である．ブートストラップの正当性も汎関数デルタ法を使って示すことができる．汎関数デルタ法の厳密な説明は講義ノートのレベルを超えるので，関心がある場合は，van der Vaart (1998) を参照せよ．

6.3 MLE の漸近理論

$\mathcal{X}$ を有限次元ユークリッド空間とし，$\varnothing \ne \Theta \subset \mathbb{R}^k$ をパラメータ空間，$\{p(\cdot;\theta) : \theta \in \Theta\}$ をパラメトリックな分布族とする．ここで，

$$X_1, \ldots, X_n \sim p(\cdot;\theta) \text{ i.i.d.}$$

が与えられているとし，$X = (X_1', \ldots, X_n')'$ とおく．X の同時確率 (密度) 関数を

$$p_n(x;\theta) = \prod_{i=1}^n p(x_i;\theta),\ x = (x_1', \ldots, x_n')' \in \mathcal{X}^n$$

とおくと，対数尤度関数は

$$\ell^n(\theta) = \ell^n(X;\theta) = \log p_n(X;\theta) = \sum_{i=1}^n \underbrace{\log p(X_i;\theta)}_{=\ell(X_i;\theta)}$$

である．このとき，θ の MLE $\widehat{\theta}$ は

$$\ell^n(\widehat{\theta}) = \max_{\theta \in \Theta} \ell^n(\theta)$$

と定義される．$\ell^n(\theta)$ が θ についてなめらかで，$\widehat{\theta}$ が Θ の内点なら，$\widehat{\theta}$ は尤度方程式

$$\dot{\ell}^n(\theta) = \left(\frac{\partial \ell^n(\theta)}{\partial \theta_1}, \dots, \frac{\partial \ell^n(\theta)}{\partial \theta_k}\right)' = 0$$

をみたす．

以下では，$k=1$ のとき，MLE の漸近正規性を示そう．説明のために，$p(\cdot;\theta)$ は密度関数とする．$\Theta \subset \mathbb{R}$ を開区間，$\theta = \theta_0$ を真値として，次の条件を仮定する．

- $A := \{u \in \mathcal{X} : p(u;\theta) > 0\}$ は θ によらない．
- $p(u;\theta)$ は θ について 3 回微分可能である．
- θ_0 の開近傍 $B \subset \Theta$ と関数 $g : A \to \mathbb{R}_+, H : A \to \mathbb{R}_+$ が存在して，

  $$\left|\frac{\partial p(u;\theta)}{\partial \theta}\right| \le g(u),\ \left|\frac{\partial^2 p(u;\theta)}{\partial \theta^2}\right| \le g(u),\ \left|\frac{\partial^3 \ell(u;\theta)}{\partial \theta^3}\right| \le H(u),\ \forall u \in A, \forall \theta \in B,$$
  $$\int_A g(u)du < \infty,\ E_{\theta_0}[H(X_1)] < \infty$$

  をみたす．
- $0 < I(\theta_0) := E_{\theta_0}[\{\dot{\ell}(X_1;\theta_0)\}^2] < \infty$.

以上の仮定のもとで，Lebesgue の優収束定理より，次の微分と積分の交換が成り立つ：$\theta \in B$ に対して，

$$\int \frac{\partial p(u;\theta)}{\partial \theta} du = \frac{d}{d\theta}\int p(u;\theta)du = 0,\ \int \frac{\partial^2 p(u;\theta)}{\partial \theta^2} du = \frac{d^2}{d\theta^2}\int p(u;\theta)du = 0.$$

よって，情報量等式

$$I(\theta_0) = E_{\theta_0}[-\ddot{\ell}(X_1;\theta_0)]$$

が成り立つ．

Theorem 6.9 (Cramér (1946)). (a) 次をみたす推定量 $\widehat{\theta}$ が存在する：$n \to \infty$ のとき，

$$P_{\theta_0}\{\dot{\ell}^n(\widehat{\theta}) = 0\} \to 1, \quad \widehat{\theta} \xrightarrow{P} \theta. \tag{*}$$

(b) (*) をみたす <u>任意の</u> 推定量 $\widehat{\theta}$ に対して，$\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, 1/I(\theta_0))$.

Proof. (a). $\theta \in B$ に対して，Taylor の定理より，θ と θ_0 の間の点 $\overline{\theta}$ が存在して，

$$\frac{1}{n}\dot{\ell}^n(\theta) = \underbrace{\frac{1}{n}\dot{\ell}^n(\theta_0)}_{=S_n} + \underbrace{\frac{1}{n}\ddot{\ell}^n(\theta_0)}_{=J_n}(\theta - \theta_0) + \frac{1}{2n}\dddot{\ell}^n(\overline{\theta})(\theta - \theta_0)^2$$

と展開できる．ここで，大数の弱法則より，

$$S_n = \frac{1}{n}\sum_{i=1}^n \dot{\ell}(X_i;\theta_0) \xrightarrow{P} 0,\ J_n = \frac{1}{n}\sum_{i=1}^n \ddot{\ell}(X_i;\theta_0) \xrightarrow{P} -I(\theta_0) < 0,$$

であって，さらに，

$$\frac{1}{n}|\dddot{\ell}^n(\overline{\theta})| \le \frac{1}{n}\sum_{i=1}^n H(X_i) \xrightarrow{P} E_{\theta_0}[H(X_1)] =: C$$

となる．従って，

$$\left|\frac{1}{n}\dot{\ell}^n(\theta) + I(\theta_0)(\theta-\theta_0)\right| \le o_P(1) + o_P(1)|\theta-\theta_0| + \frac{1}{2}\{C + o_P(1)\}(\theta-\theta_0)^2$$

を得る．ここで，$o_P(1)$ の項は θ に依存しない．$Y_n = o_P(1)$ なら，$\exists\varepsilon_n \downarrow 0$ s.t. $P(|Y_n| > \varepsilon_n) \le \varepsilon_n$ となるから，十分遅い $\varepsilon_n \to 0$ に対して，

$$P_{\theta_0}\left\{\exists\theta \in [\theta_0-\varepsilon_n, \theta_0+\varepsilon_n] \text{ s.t. } \dot{\ell}^n(\theta) = 0\right\} \to 1$$

を得る．そこで，

$$\widehat{\theta} = \begin{cases} \min\{\theta \in [\theta_0-\varepsilon_n, \theta_0+\varepsilon_n] : \dot{\ell}^n(\theta) = 0\} & \text{右辺の集合が空でないとき} \\ 0 & \text{それ以外のとき} \end{cases}$$

とおくと，$P_{\theta_0}\{\dot{\ell}^n(\widehat{\theta}) = 0\} \to 1$ であって，$\widehat{\theta} \xrightarrow{P} \theta_0$ をみたす．

(b). $\widehat{\theta}$ を (*) をみたす推定量とする．このとき，Taylor の定理より，$\widehat{\theta}$ と θ_0 の間の点 $\overline{\theta}$ が存在して，

$$\frac{1}{n}\dot{\ell}^n(\widehat{\theta}) = S_n + J_n(\widehat{\theta}-\theta_0) + \frac{1}{2n}\dddot{\ell}^n(\overline{\theta})(\widehat{\theta}-\theta_0)^2$$

と展開できる．ここで，$P_{\theta_0}\{\dot{\ell}^n(\widehat{\theta}) = 0\} \to 0$ だから，左辺は $o_P(n^{-1/2})$ である．また，$|\overline{\theta}-\theta_0| \le |\widehat{\theta}-\theta_0| = o_P(1)$ より，$P(\overline{\theta} \in B) \to 1$ であって，$\overline{\theta} \in B$ のとき $|n^{-1}\dddot{\ell}^n(\overline{\theta})| \le n^{-1}\sum_{i=1}^n H(X_i)$ が成り立つ．いま，$n^{-1}\sum_{i=1}^n H(X_i) = O_P(1)$ だから，$|n^{-1}\dddot{\ell}^n(\overline{\theta})| = O_P(1)$ であって，よって，

$$\left|\frac{1}{n}\dddot{\ell}^n(\overline{\theta})(\widehat{\theta}-\theta_0)\right| = O_P(1)o_P(1) = o_P(1)$$

を得る．以上の評価と，$J_n = -I(\theta_0) + o_P(1)$ より，

$$o_P(n^{-1/2}) = S_n + \{-I(\theta_0) + o_P(1)\}(\widehat{\theta}-\theta_0)$$

を得る．あとは，CLT と Slutsky の補題より，

$$\sqrt{n}(\widehat{\theta}-\theta_0) = \{I(\theta_0) + o_P(1)\}^{-1}\sqrt{n}S_n + o_P(1) \xrightarrow{d} N(0, 1/I(\theta_0))$$

を得る． □

この定理の言っていることは，尤度方程式の根のなかには一致性をみたすものが存在し，そのような根は漸近正規性をみたす，ということである．尤度方程式が複数の根をもつ場合，MLE が一致性をみたす尤度方程式の根に一致しているとは限らないので，Theorem 6.9 は必ずしも MLE の漸近正規性を保証するものではない．しかし，尤度方程式が一意な根をもつ場合は，それは MLE に一致しかつ一致性をみたすので，その場合は MLE の漸近正規性が従う [33]．

後述するように，いくつかの観点から，$N(0,1/I(\theta_0))$ は最良の極限分布である．しかし，尤度方程式が複数の根をもつ場合などは MLE の計算が難しいし，MLE が一致性をもつ尤度方程式の根に一致している保証はない．そのような場合でも，適当な初期推定量から $N(0,1/I(\theta_0))$ を極限分布にもつような推定量を構成できる．

$$\widehat{I}(\theta) = -n^{-1}\ddot{\ell}^n(\theta)$$

とおく．$\widehat{I}(\theta)$ は <u>観測 Fisher 情報量</u> (observed Fisher information) と呼ばれる．初期推定量 $\widetilde{\theta}$ に対して，

$$\check{\theta} = \widetilde{\theta} + \widehat{I}(\widetilde{\theta})^{-1}\{n^{-1}\dot{\ell}^n(\widetilde{\theta})\}$$

とおく．$\check{\theta}$ は <u>ワンステップ推定量</u> (one-step estimator) と呼ばれる．$\check{\theta}$ をワンステップ推定量と呼ぶ理由は，尤度方程式をニュートン・ラフソン法によって解くときに，$\check{\theta}$ が初期値 $\widetilde{\theta}$ を 1 回更新した値になっているためである．

Theorem 6.10. $\widetilde{\theta}$ が $n^{1/4}(\widetilde{\theta}-\theta_0) = o_P(1)$ をみたせば，

$$\sqrt{n}(\check{\theta}-\theta_0) \xrightarrow{d} N(0,1/I(\theta_0))$$

となる．

Proof. Taylor の定理より，$\widetilde{\theta}$ と θ_0 の間の点 $\overline{\theta}$ が存在して，

$$n^{-1}\ddot{\ell}^n(\widetilde{\theta}) = \underbrace{n^{-1}\ddot{\ell}^n(\theta_0)}_{=J_n} + n^{-1}\dddot{\ell}^n(\overline{\theta})(\widetilde{\theta}-\theta_0)$$

と展開できる．ここで，$n^{-1}\dddot{\ell}^n(\overline{\theta}) = O_P(1)$ であって，$J_n \xrightarrow{P} -I(\theta_0) \neq 0$ より，

$$\widehat{I}(\widetilde{\theta})^{-1} = \{-n^{-1}\ddot{\ell}^n(\widetilde{\theta})\}^{-1} = -J_n^{-1} + O_P(1)|\widehat{\theta}-\theta_0| = J_n^{-1} + o_P(n^{-1/4})$$

を得る．一方，

$$\begin{aligned}\frac{1}{\sqrt{n}}\dot{\ell}^n(\widetilde{\theta}) &= \underbrace{\frac{1}{\sqrt{n}}\dot{\ell}^n(\theta_0)}_{\sqrt{n}S_n} + \underbrace{\frac{1}{n}\ddot{\ell}^n(\theta_0)}_{=J_n}\sqrt{n}(\widetilde{\theta}-\theta_0) + \underbrace{\frac{1}{2}\{n^{1/4}(\widetilde{\theta}-\theta_0)\}^2\cdot\frac{1}{n}\dddot{\ell}^n(\overline{\theta})}_{=o_P(1)} \\ &= \sqrt{n}S_n + J_n\sqrt{n}(\widetilde{\theta}-\theta_0) + o_P(1)\end{aligned}$$

[33] とはいえもっと一般的な条件のもとで MLE の一致性を証明することをできる．これは Wald (1949) による．van der Vaart (1998, Section 5.2) を参照せよ．

であって，$\sqrt{n}S_n = O_P(1), J_n^{-1} = O_P(1), \sqrt{n}(\widetilde{\theta} - \theta_0) = o_P(n^{1/4})$ より，

$$\begin{aligned}\sqrt{n}(\check{\theta} - \theta_0) &= \sqrt{n}(\widetilde{\theta} - \theta_0) + \{-J_n^{-1} + o_P(n^{-1/4})\}\{\sqrt{n}S_n + J_n\sqrt{n}(\widetilde{\theta} - \theta_0) + o_P(1)\} \\ &= -J_n^{-1}\sqrt{n}S_n + o_P(1)\end{aligned}$$

を得る．あとは $J_n \xrightarrow{P} -I(\theta_0), \sqrt{n}S_n \xrightarrow{d} N(0, 1/I(\theta_0))$ と Slutsky の補題より，

$$\sqrt{n}(\check{\theta} - \theta_0) \xrightarrow{d} N(0, 1/I(\theta_0))$$

を得る． □

初期推定量が $\sqrt{n}(\widetilde{\theta} - \theta_0) = O_P(1)$ であって，$I(\theta)$ が $\theta = \theta_0$ において連続なら，$\widehat{I}(\widetilde{\theta})$ を $I(\widetilde{\theta})$ に取り換えてよい．

Corollary 6.2. $\sqrt{n}(\widetilde{\theta} - \theta_0) = O_P(1)$ とし，$I(\theta)$ は $\theta = \theta_0$ において連続とする．このとき，$\check{\theta} = \widetilde{\theta} + I(\widetilde{\theta})^{-1}\{n^{-1}\dot{\ell}^n(\widetilde{\theta})\}$ に対して，

$$\sqrt{n}(\check{\theta} - \theta_0) \xrightarrow{d} N(0, 1/I(\theta_0))$$

となる．この推定量 $\check{\theta}$ のこともワンステップ推定量と呼ぶ．

Proof. Theorem 6.10 の証明と $J_n = -I(\theta_0) + o_P(1), \sqrt{n}(\widetilde{\theta} - \theta_0) = O_P(1)$ より，

$$\frac{1}{\sqrt{n}}\dot{\ell}^n(\widetilde{\theta}) = \sqrt{n}S_n + J_n\sqrt{n}(\widetilde{\theta} - \theta_0) + o_P(1) = \sqrt{n}S_n - I(\theta_0)\sqrt{n}(\widetilde{\theta} - \theta_0) + o_P(1).$$

一方，$I(\theta)$ は $\theta = \theta_0$ で連続だから，$I(\widetilde{\theta}) = I(\theta_0) + o_P(1)$ となる．よって，

$$\begin{aligned}\sqrt{n}(\check{\theta} - \theta_0) &= \sqrt{n}(\widetilde{\theta} - \theta_0) + \{I(\theta_0)^{-1} + o_P(1)\}\{\sqrt{n}S_n - I(\theta_0)\sqrt{n}(\widetilde{\theta} - \theta_0) + o_P(1)\} \\ &= I(\theta_0)^{-1}\sqrt{n}S_n + o_P(1) \xrightarrow{d} N(0, 1/I(\theta_0)).\end{aligned}$$

□

Example 6.13. f を Cauchy 分布の密度関数とする：

$$f(u) = \frac{1}{\pi(1 + u^2)},\ u \in \mathbb{R}.$$

$\theta \in \mathbb{R}$ に対して，$X_1, \ldots, X_n \sim f(\bullet - \theta)$ i.i.d. とすると，尤度方程式は

$$\sum_{i=1}^{n} \frac{2(X_i - \theta)}{1 + (X_i - \theta)^2} = 0$$

である．両辺に $\prod_{i=1}^{n}\{1 + (X_i - \theta)^2\}$ をかけると，尤度方程式は

$$\sum_{i=1}^{n}(X_i - \theta)\prod_{j \neq i}\{1 + (X_j - \theta)^2\} = 0$$

と等価である．これは θ の $(2n-1)$ 次多項式だから，尤度方程式は一般に $(2n-1)$ 個の根をもつ．ところで，f は原点対称なので，θ は X_i の d.f. のメディアンでもある．よって，標本メディアンを $\widetilde{\theta}$ とおくと，$f(0) = 1/\pi$ より，

$$\sqrt{n}(\widetilde{\theta} - \theta) \xrightarrow{d} N(0, \pi^2/4)$$

となる．しかし，$f(\bullet - \theta)$ の Fisher 情報量は

$$I(\theta) = \int \frac{\{f'(u)\}^2}{f(u)} du = \cdots = \frac{1}{2}$$

だから，$1/I(\theta) = 2 < \pi^2/4$ である．

そこで，ワンステップ推定を使って漸近分散を改善する．$\sqrt{n}(\widetilde{\theta} - \theta) = O_P(1)$ であって，$I(\theta) = 1/2$ だから，

$$\check{\theta} = \widetilde{\theta} + \frac{4}{n} \sum_{i=1}^{n} \frac{X_i - \widetilde{\theta}}{1 + (X_i - \widetilde{\theta})^2}$$

とおくと，$\sqrt{n}(\check{\theta} - \theta) \xrightarrow{d} N(0, 2)$ となる．

Remark 6.5 (多次元の場合). 多次元の場合も，1 次元の場合と同様に，いくつかの正則条件のもとで，$I(\theta)$ を $p(\cdot;\theta)$ の Fisher 情報行列として，$\theta = \theta_0$ を真値とすると，$n \to \infty$ のとき，

$$P_{\theta_0}\{\dot{\ell}^n(\widehat{\theta}) = 0\} \to 1, \quad \sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, I(\theta_0)^{-1})$$

をみたす推定量 $\widehat{\theta}$ が存在する．

MLE の漸近最適性

MLE の漸近最適性を考察する．厳密な考察は講義のレベルを超えるので，van der Vaart (1998, Chapter 8) に譲る．以下では，簡単のために，$k = 1$ とし，Θ を $\mathbb{R}$ の開区間とする．

$\theta \in \Theta$ に対する 2 つの推定量 $\widehat{\theta}_n = \widehat{\theta}_n(X_1, \ldots, X_n), \widetilde{\theta}_n = \widetilde{\theta}_n(X_1, \ldots, X_n)$ が与えられていて，$n \to \infty$ のとき，

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} N(0, \sigma^2(\theta)), \quad \sqrt{n}(\widetilde{\theta}_n - \theta) \xrightarrow{d} N(0, \tau^2(\theta))$$

とする．$\sigma^2(\theta) > 0, \tau^2(\theta) > 0$ とする．このとき，θ における，$\widehat{\theta}_n$ の $\widetilde{\theta}_n$ に対する漸近相対有効性 (asymptotic relative efficiency, ARE) を

$$\mathrm{ARE}_\theta(\widehat{\theta}_n, \widetilde{\theta}_n) = \frac{\tau^2(\theta)}{\sigma^2(\theta)}$$

と定義する．$\mathrm{ARE}_\theta(\widehat{\theta}_n, \widetilde{\theta}_n)$ が 1 より大きいとき，θ において $\widehat{\theta}_n$ は $\widetilde{\theta}_n$ より漸近有効であるという．

Example 6.14. $X_1, \ldots, X_n \sim N(\theta, 1)$ i.i.d. のとき，θ の MLE は $\widehat{\theta}_n = \overline{X}$ であって，$\sqrt{n}(\widehat{\theta} - \theta) \sim N(0,1)$．一方，$\theta$ は $N(\theta, 1)$ のメディアンであるから，$\widetilde{\theta}_n$ を標本メディアンとすると，

$$\sqrt{n}(\widetilde{\theta}_n - \theta) \xrightarrow{d} N(0, \pi/2)$$

となる．よって，ARE は，

$$\mathrm{ARE}_\theta(\widehat{\theta}_n, \widetilde{\theta}_n) = \frac{\pi}{2} > 1$$

である．

Example 6.15. $X_1, \ldots, X_n \sim Po(\lambda)$ i.i.d. $(\lambda > 0)$ とし，$\theta = e^{-\lambda} = P_\lambda(X_i = 0)$ の推定を考える．このとき，λ の MLE は $\overline{X}$ だから，θ の MLE は $\widehat{\theta}_n = e^{-\overline{X}}$ である．その他に $\widetilde{\theta}_n = n^{-1} \sum_{i=1}^n I(X_i = 0)$ も自然な推定量である．ここで，CLT とデルタ法より，

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} N(0, \lambda e^{-2\lambda}),\ \sqrt{n}(\widetilde{\theta}_n - \theta) \xrightarrow{d} N(0, e^{-\lambda} - e^{-2\lambda})$$

であるから，ARE は

$$\mathrm{ARE}_\lambda(\widehat{\theta}_n, \widetilde{\theta}_n) = \frac{e^\lambda - 1}{\lambda} > 1$$

である．

ARE には次の意味がある．$\delta > 0$ とし，推定量が θ の δ 近傍に入る確率を考える．$P_\theta(|\widehat{\theta}_n - \theta| \leq \delta)$ に対して，同じ確率を $\widetilde{\theta}_m$ が達成するために必要な標本サイズを m とする．このとき，$Z \sim N(0,1)$ に対して，

$$P_\theta(|\widehat{\theta}_n - \theta| \leq \delta) \approx P(|Z| \leq \delta\sigma(\theta)/\sqrt{n}), \quad P_\theta(|\widetilde{\theta}_m - \theta| \leq \delta) \approx P(|Z| \leq \delta\tau(\theta)/\sqrt{m})$$

であるから，

$$\frac{\sigma(\theta)}{\sqrt{n}} \approx \frac{\tau(\theta)}{\sqrt{m}}, \quad \text{i.e.,} \quad \frac{m}{n} \approx \frac{\tau^2(\theta)}{\sigma^2(\theta)}$$

である．すなわち，ARE は同じ精度を達成するために必要な標本サイズの比の近似になっている．

推定量 $\widehat{\theta}_n = \widehat{\theta}_n(X_1, \ldots, X_n)$ が，各 $\theta_0 \in \Theta, h \in \mathbb{R}$ に対して，$\theta = \theta_n = \theta_0 + h/\sqrt{n}$ が真値のときに，

$$\sqrt{n}(\widehat{\theta}_n - \theta_n) \xrightarrow{d} \underbrace{L_{\theta_0}}_{\text{d.f.}}$$

をみたし，L_{θ_0} が h に依存しないとき，$\widehat{\theta}$ は <u>正則</u> (regular) な推定量であるという．ここで，$h = 0$ のとき $\theta_n = \theta_0$ だから，L_{θ_0} は $\sqrt{n}(\widehat{\theta} - \theta_0)$ の $\theta = \theta_0$ のもとでの極限分布である．いくつかの条件のもとで，MLE は正則になる．さらに，任意の正則な推定量 $\widehat{\theta}_n$ に対して，

$$(L_{\theta_0}\ \text{の分散}) \geq 1/I(\theta_0)$$

となる (分散が存在しないときは，左辺は $+\infty$ とみなす)．よって，MLE は正則な推定量のなかで最小な極限分散 (正確には極限分布の分散) をもつ．この結果は Hájek-Le Cam のたたみ込み定理 (convolution theorem) から従う．

では，正則 でない 推定量であって，すべての $\theta \in \Theta$ に対して，極限分散が $1/I(\theta)$ 以下で，かつある $\theta_0 \in \Theta$ において極限分散が $1/I(\theta_0)$ より小さくなるものは存在するであろうか．答えは YES である．

Example 6.16. $X_1, \dots, X_n \sim N(\theta, 1)$ i,i.d. とすると，MLE は $\widehat{\theta}_n = \overline{X}$ であって，$\sqrt{n}(\widehat{\theta}_n - \theta) \sim N(0,1)$. ここで，次のような推定量を考える：

$$\widetilde{\theta}_n = \begin{cases} 0 & \text{if } |\overline{X}| \le n^{-1/4} \\ \overline{X} & \text{otherwise} \end{cases}.$$

$\widetilde{\theta}_n$ を Hodges の推定量 と呼ぶ．$\sqrt{n}\overline{X} \sim N(\sqrt{n}\theta, 1)$ だから，

$$\begin{aligned} &P_\theta(|\overline{X}| \le n^{-1/4}) = P_\theta(|\sqrt{n}\overline{X}| \le n^{1/4}) = \Phi(n^{1/4} - \sqrt{n}\theta) - \Phi(-n^{1/4} - \sqrt{n}\theta) \\ &\to \begin{cases} 1 & \theta = 0 \\ 0 & \theta \ne 0 \end{cases}. \end{aligned}$$

よって，

$$\sqrt{n}(\widetilde{\theta}_n - \theta) \xrightarrow{d} \begin{cases} 0 & \theta = 0 \\ N(0,1) & \theta \ne 0 \end{cases}$$

となる．

Hodges の推定量は正則でない．実際，$\theta = \theta_n = h/\sqrt{n}$ が真値のとき，$P_{\theta_n}(|\overline{X}| \le n^{-1/4}) \to 1$ だから，

$$\sqrt{n}(\widetilde{\theta}_n - \theta_n) \xrightarrow{P} -h$$

となる．

しかしながら，この結果から Hodges の推定量は MLE よりよいと結論づけるのは早計である．リスク $E_\theta[\{\sqrt{n}(\widetilde{\theta}_n - \theta)\}^2]$ をシミュレーションしてみればわかるように，有限の n において，Hodges の推定量は $\theta = 0$ でのリスクを改善する代わりに，$\theta = 0$ の近傍でのリスクを増大させている．実際，$\theta = \theta_n = h/\sqrt{n}$ のとき，Fatou の補題より，

$$\liminf_n E_{h/\sqrt{n}}[\{\sqrt{n}(\widetilde{\theta}_n - h/\sqrt{n})\}^2] \ge h^2$$

となって，右辺は $|h| \to \infty$ のとき ∞ に発散する．一方，MLE のリスクは $E_\theta[\{\sqrt{n}(\widehat{\theta} - \theta)\}^2] = 1$ であって，n にも θ に依存しない．

正則とは限らない推定量と比較したときの MLE の漸近最適性に関しては次の 2 つの結果が知られている．

- $\widehat{\theta}_n$ を各 $\theta \in \Theta$ に対して $\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} L_\theta$ をみたす任意の推定量とする ($\widehat{\theta}$ は正則である必要はない)．このとき，いくつかの条件のもとで，“ほとんどすべての” $\theta \in \Theta$ に対して，

$$(L_\theta \text{ の分散}) \geq 1/I(\theta)$$

が成り立つ．つまり，極限分散が $1/I(\theta)$ より小さくなるような θ の集合は Lebesgue 測度 0 である (概たたみ込み定理[34])．

- $\ell : \mathbb{R} \to \mathbb{R}_+$ を次をみたす関数とする：各 $c \geq 0$ に対して $\{x \in \mathbb{R} : \ell(x) \leq c\}$ が凸集合かつ原点対称となる．このとき，いくつかの条件のもとで，任意の (正則とは限らない) 推定量 $\widehat{\theta}_n$ と各 $\theta \in \Theta$ に対して，

$$\begin{aligned}&\sup_{I \subset \mathbb{R}:\text{finite}} \liminf_n \sup_{h \in I} E_{\theta + h/\sqrt{n}} \left[\ell \left(\sqrt{n}(\widehat{\theta}_n - (\theta + h/\sqrt{n})) \right) \right] \\ &\quad \geq E[\ell(Z)],\ Z \sim N(0, 1/I(\theta))\end{aligned}$$

が成り立つ (Hájek-Le Cam の局所漸近ミニマクス定理)．

Neyman-Scott 問題[35]

$(X_1, Y_1), \ldots, (X_n, Y_n)$ を独立な 2 次元確率ベクトルとし，

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_i \\ \mu_i \end{pmatrix}, \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \right),\ \mu_i \in \mathbb{R}, \sigma^2 > 0$$

とする．σ^2 の推定を考える．$X_i - Y_i \sim N(0, 2\sigma^2)$ だから，

$$\widetilde{\sigma}^2 = \frac{1}{2n} \sum_{i=1}^n (X_i - Y_i)^2$$

とおくと，$\widetilde{\sigma}^2 \xrightarrow{P} \sigma^2$ となる．しかし σ^2 の MLE は一致性をもたない．$(X_1, Y_1), \ldots, (X_n, Y_n)$ の同時密度は，

$$\frac{1}{(2\pi\sigma^2)^n} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_i)^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu_i)^2 \right\}$$

だから，$\mu_1, \ldots, \mu_n, \sigma^2$ の MLE は，

$$\widehat{\mu}_i = \frac{1}{2}(X_i + Y_i),\ \widehat{\sigma}^2 = \frac{1}{2n} \left\{ \sum_{i=1}^n (X_i - \widehat{\mu}_i)^2 + \sum_{i=1}^n (Y_i - \widehat{\mu}_i)^2 \right\}$$

[34] “Almost everywhere convolution theorem” の訳.
[35] Neyman and Scott (1948).

である．$\{(X_i-\widehat{\mu}_i)^2+(Y_i-\widehat{\mu}_i)^2\}/\sigma^2$ は独立に $\chi^2(1)$ に従うから，

$$(2n)\widehat{\sigma}^2/\sigma^2 \sim \chi^2(n)$$

である．従って，

$$\widehat{\sigma}^2 \xrightarrow{P} \frac{\sigma^2}{2}$$

となって，一致性をもたない．

$\widehat{\sigma}^2$ が一致性をもたない理由は，μ_i の推定に使える標本サイズが 2 しかないため，$\widehat{\mu}_i$ が一致性をもたず，そのバイアスが σ^2 の推定に影響を及ぼすためである．このように局外パラメータの数が標本サイズとともに増えていくために，関心のある共通パラメータの推定量が一致性をもたなくなる問題は，局外パラメータ問題 (incidental parameters problem) と呼ばれる．局外パラメータ問題は計量経済学において固定効果をもつパネルデータモデルの推定に現れる (Lancaster, 2000)．

なお，μ_i の推定に使える標本サイズが 2 ではなく n とともに増える場合，σ^2 の MLE は一致性をもつ．すなわち，$X_{i,j}, i=1,\dots,n, j=1,\dots,m$ を独立な r.v.'s とし，各 i に対して，

$$X_{i,j} \sim N(\mu_i, \sigma^2)$$

とする．このとき，σ^2 の MLE は，

$$\widehat{\sigma}^2 = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}(X_{i,j}-\widehat{\mu}_i)^2,\ \widehat{\mu}_i = \frac{1}{m}\sum_{j=1}^{m}X_{i,j}$$

である．ここで，

$$(nm)\widehat{\sigma}^2/\sigma^2 \sim \chi^2(n(m-1))$$

だから，平均と分散を評価すると，

$$E[\widehat{\sigma}^2] = \frac{(m-1)}{m}\sigma^2,\ \mathrm{Var}(\widehat{\sigma}^2) = \frac{2(m-1)}{nm^2}\sigma^4$$

である．よって，$m=m_n\to\infty$ なら，$E[\widehat{\sigma}^2]\to\sigma^2, \mathrm{Var}(\widehat{\sigma}^2)\to 0$ だから，Chebyshev の不等式より，

$$\widehat{\sigma}^2 \xrightarrow{P} \sigma^2$$

を得る．

6.4 U 統計量

F を $\mathbb{R}$ 上の d.f. とし，$X_1,\dots,X_n \sim F$ i.i.d. $(n\ge 2)$ とする．また，$h:\mathbb{R}^2\to\mathbb{R}$ を対称な関数とする：$h(x,y)=h(y,x)\ \forall x,y\in\mathbb{R}$. ここで，

$$\theta = E[h(X_1,X_2)]$$

の推定を考える ($E[|h(X_1,X_2)|] < \infty$ を仮定する)．このとき，次の推定量を考える：

$$U_n = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} h(X_i, X_j).$$

U_n のことをカーネル h に対する U 統計量 と呼ぶ．ここで，$i \neq j$ に対して $E[h(X_i,X_j)] = \theta$ だから，

$$E[U_n] = \theta$$

が成り立つ．すなわち，U_n は θ の不偏推定量になっている (この場合，F をパラメータとみなす)．U 統計量は Halmos (1946) と Hoeffding (1948) によって導入された．

Example 6.17 (標本分散)**.** $h(x,y) = (x-y)^2/2$ とおくと，$\theta = E[h(X_1,X_2)] = \mathrm{Var}(X_1)$ であって，

$$U_n = \frac{1}{n(n-1)} \sum_{1 \le i < j \le n} (X_i - X_j)^2 = \cdots = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X})^2.$$

Example 6.18. $\theta = P(X_1 + X_2 \le 0)$ とすると，対応するカーネルは $h(x,y) = I(x+y \le 0)$ である．このとき，

$$U_n = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} I(X_i + X_j \le 0).$$

この統計量は，F が連続かつ原点対称 (i.e., $F(x) = 1 - F(-x)\ \forall x \in \mathbb{R}$) という帰無仮説に対する検定統計量として用いられる．この帰無仮説のもとで，$\theta = E[F(-X_2)] = 1/2$ だから，

$$|U_n - 1/2| > c \Rightarrow \text{reject}$$

という検定を考えることができる．

以下，$E[h(X_1,X_2)^2] < \infty$ という仮定のもとで，$n \to \infty$ のとき $\sqrt{n}(U_n - \theta)$ が正規分布に分布収束することを示そう．ここで，

$$g_1(x) = E[h(x, X_2)] - \theta,\ g_2(x,y) = h(x,y) - g_1(x) - g_2(y) - \theta$$

とおくと，

$$U_n - \theta = \underbrace{\frac{2}{n}\sum_{i=1}^n g_1(X_i)}_{=U_n^{(1)}} + \underbrace{\frac{2}{n(n-1)} \sum_{1 \le i < j \le n} g_2(X_i, X_j)}_{=U_n^{(2)}}$$

と分解できる．この分解を Hoeffding 分解 と呼ぶ．

Theorem 6.11 (Hoeffding (1948))**.** $E[h(X_1,X_2)^2] < \infty$ と仮定する．このとき，$\sqrt{n}(U_n - \theta - U_n^{(1)}) \xrightarrow{P} 0$ となる．よって，

$$\sqrt{n}(U_n - \theta) \xrightarrow{d} N(0, 4E[g_1(X_1)^2])$$

が成り立つ．

Proof. $E[g_1(X_1)^2] \le E[\{E[h(x,X_2)]|_{x=X_1}\}^2] \le E[h(X_1,X_2)^2] < \infty$ に注意する．いま，

$$\zeta_1 = \mathrm{Cov}(h(X_1,X_2),h(X_1,X_3)),\ \zeta_2 = \mathrm{Var}(h(X_1,X_2))$$

とおくと，h の対称性と i.i.d. の仮定から，

$$\begin{aligned}\mathrm{Var}(U_n^{(2)}) &= \frac{4}{n^2(n-1)^2}\sum_{1\le i<j\le n}\sum_{1\le k<\ell\le n}\underbrace{\mathrm{Cov}(g_2(X_i,X_j),g_2(X_k,X_\ell))}_{=\mathrm{Cov}(h(X_i,X_j),h(X_k,X_\ell))} \\ &= \frac{4}{n^2(n-1)^2}\left\{\frac{n(n-1)}{2}\zeta_2 + n(n-1)(n-2)\zeta_1\right\}\end{aligned}$$

を得る．よって，$E[U_n^{(2)}] = 0, \mathrm{Var}(U_n^{(2)}) = o(n^{-1})$ だから，Chebyshev の不等式より，$U_n^{(2)} = o_P(n^{-1/2})$ を得る． □

Example 6.19 (Example 6.18 の続き)．$h(x,y) = I(x+y\le 0)$ とすると，

$$U_n^{(1)} = \frac{2}{n}\sum_{i=1}^n (F(-X_i) - E[F(-X_1)])$$

となる．F が連続かつ原点対称 $(F(x) = 1-F(-x))$ なら，$F(-X_2) \sim U(0,1)$ だから，

$$\sqrt{n}(U_n - 1/2) \xrightarrow{d} N(0,1/3)$$

となる．

もっと一般に，$h : \mathbb{R}^r \to \mathbb{R}$ を対称な関数とすると (i.e., $\{1,\dots,r\}$ の任意の置換 σ に対して，$h(x_1,\dots,x_r) = h(x_{\sigma(1)},\dots,x_{\sigma(r)})$)，

$$U_n = \frac{1}{\binom{n}{r}}\sum_{1\le i_1<\cdots<i_r\le n} h(X_{i_1},\dots,X_{i_r})$$

を r 次の U 統計量と呼ぶ．本節では簡単のため $r=2$ の場合を考察したが，一般の r に対しても，$E[h(X_1,\dots,X_r)^2] < \infty$ なら $\sqrt{n}(U_n - E[U_n])$ が正規分布に分布収束することが示せる．U 統計量については Serfling (1980, Chapter 5) が詳しい．

6.5 Berry-Esseen の定理の初等的なバージョンの証明

本節の目標は次の定理の証明である．

Theorem 6.12. F を $\mathbb{R}$ 上の d.f. とし，$X_1,\dots,X_n \sim F$ i.i.d. として，$E[|X_1|^3] < \infty$ を仮定する．また，$\mu = E[X_1], \sigma^2 = \mathrm{Var}(X_1)$ とおき，$\sigma = \sqrt{\sigma^2} > 0$ とする．このとき，

$$\sup_{x\in\mathbb{R}}\left|P\left\{\sqrt{n}(\overline{X}-\mu)/\sigma \le x\right\} - \Phi(x)\right| \le A\left(\frac{E[|X_1-\mu|^3]}{\sqrt{n}\sigma^3}\right)^{1/4}$$

が成り立つ．ここで，A は絶対定数である．

一般性を失うことなく，

$$\mu = 0,\ \sigma^2 = 1$$

を仮定してよい．

$$\check{S}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$$

とおく．また，$\gamma = E[|X_1|^3]$ とおく．次の補題が本質的である．関数 $g : \mathbb{R} \to \mathbb{R}$ に対して，$\|g\|_\infty = \sup_{x \in \mathbb{R}} |g(x)|$ と定める．

Lemma 6.4. $g : \mathbb{R} \to \mathbb{R}$ を有界な C^3 級関数であって，3 階までの導関数はすべて有界とする．このとき，$Z \sim N(0,1)$ に対して，

$$|E[g(\check{S}_n)] - E[g(Z)]| \le \frac{\|g'''\|_\infty \gamma}{2\sqrt{n}}$$

が成り立つ．

Proof of Lemma 6.4. $Z_1, \ldots, Z_n \sim N(0,1)$ i.i.d. とし，$(Z_1, \ldots, Z_n)'$ と $(X_1, \ldots, X_n)'$ は独立とする．$k = 1, \ldots, n$ に対して，

$$\check{S}_n^{(k)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{k-1} Z_i + \frac{1}{\sqrt{n}} \sum_{i=k+1}^{n} X_i$$

とおく．このとき，

$$E[g(\check{S}_n)] - E[g(Z)] = \sum_{k=1}^n \{E[g(\check{S}_n^{(k)} + n^{-1/2} X_k)] - E[g(\check{S}_n^{(k)} + n^{-1/2} Z_k)]\}$$

と表せる．Taylor の定理より，

$$g(\check{S}_n^{(k)} + n^{-1/2} X_k) = g(\check{S}_n^{(k)}) + \frac{X_k}{\sqrt{n}} g'(\check{S}_n^{(k)}) + \frac{X_k^2}{2n} g''(S_n^{(k)}) + R_k^X,\ |R_k^X| \le \frac{|X_k|^3}{6n^{3/2}} \|g'''\|_\infty,$$

$$g(\check{S}_n^{(k)} + n^{-1/2} Z_k) = g(\check{S}_n^{(k)}) + \frac{Z_k}{\sqrt{n}} g'(\check{S}_n^{(k)}) + \frac{Z_k^2}{2n} g''(S_n^{(k)}) + R_k^Z,\ |R_k^Z| \le \frac{|Z_k|^3}{6n^{3/2}} \|g'''\|_\infty$$

と展開できる．ここで，X_k と $\check{S}_n^{(k)}$ は独立だから，

$$E[X_k g'(\check{S}_n^{(k)})] = E[X_k] E[g'(\check{S}_n^{(k)})] = 0,\ E[X_k^2 g''(\check{S}_n^{(k)})] = E[X_k^2] E[g''(\check{S}_n^{(k)})] = E[g''(\check{S}_n^{(k)})]$$

となる．同様にして，$E[Z_k g'(\check{S}_n^k)] = 0, E[Z_k^2 g''(\check{S}_n^{(k)})] = E[g''(\check{S}_n^{(k)})]$ となる．よって，

$$|E[g(\check{S}_n)] - E[g(Z)]| \le \sum_{k=1}^n \{E[|R_k^X|] + E[|R_k^Z|]\} \le \frac{\|g'''\|_\infty}{6n^{3/2}} \sum_{k=1}^n (E[|X_k|^3] + E[|Z_k|^3]) \quad (*)$$

を得る．ところで，$E[|Z_k|^3] = 2\sqrt{2}/\sqrt{\pi}$ であって，$x \mapsto |x|^{3/2}$ の凸性と Jensen の不等式より，$\gamma = E[(X_1^2)^{3/2}] \geq (E[X_1^2])^{3/2} = 1$ だから，

$$(*) \leq \left(\frac{1}{6} + \frac{\sqrt{2}}{3\sqrt{\pi}}\right) \frac{\|g'''\|_\infty \gamma}{\sqrt{n}} \leq \frac{\|g'''\|_\infty \gamma}{2\sqrt{n}}$$

を得る． □

Proof of Theorem 6.12. $\mu = 0, \sigma^2 = 1$ と仮定していることを思い出す．

$$g(x) = \begin{cases} 0 & x \geq 1 \\ \frac{1}{B(4,4)} \int_x^1 y^3(1-y)^3 dy & 0 < x < 1 \\ 1 & x \leq 0 \end{cases}$$

とおくと，g は C^3 級であって，3 階までの導関数はすべて有界である．そこで，$x \in \mathbb{R}, \varepsilon > 0$ に対して，$g_{x,\varepsilon}(y) = g((y-x)/\varepsilon)$ とおくと，$I(y \leq x) \leq g_{x,\varepsilon}(y) \leq I(y \leq x + \varepsilon)$ であって，$\|g'''_{x,\varepsilon}\|_\infty \leq \varepsilon^{-3}\|g'''\|_\infty$ である．よって，

$$\begin{aligned} P(\check{S}_n \leq x) - P(Z \leq x) &\leq E[g_{x,\varepsilon}(\check{S}_n)] - E[g(Z)] + E[g(Z)] - P(Z \leq x) \\ &\leq \frac{\|g'''\|_\infty \gamma}{2\varepsilon^3\sqrt{n}} + \underbrace{P(x < Z \leq x + \varepsilon)}_{\leq \frac{\varepsilon}{\sqrt{2\pi}}}. \end{aligned}$$

同様にして，

$$P(\check{S}_n \leq x) - P(Z \leq x) \geq -\frac{\|g'''\|_\infty \gamma}{2\varepsilon^3\sqrt{n}} - \frac{\varepsilon}{\sqrt{2\pi}}$$

であるから，

$$|P(\check{S}_n \leq x) - \Phi(x)| \leq \frac{\|g'''\|_\infty \gamma}{2\varepsilon^3\sqrt{n}} + \frac{\varepsilon}{\sqrt{2\pi}}$$

を得る．あとは $\varepsilon = (\gamma/\sqrt{n})^{1/4}$ として定理の結論を得る． □

A　宿題

2016.4.14. 出題：加藤賢悟

宿題 1

- 提出期限：4/21 の講義終了時.

問題

記号は講義で用いたものに従うとする．以下，$(\Omega, \mathcal{F}, P)$ を確率空間とする．

1. $A_1, \dots, A_n \in \mathcal{F}$ を独立な事象とし，各 $i = 1, \dots, n$ に対して，$B_i = A_i$ or A_i^c とする．このとき，$B_1, \dots, B_n$ も独立であることを示せ．ヒント：帰納法．難しい場合は，$n = 3$ の場合を示せば部分点を与える．

2. X を $\{x_1, \dots, x_k\}$ に値をとる確率変数とし，$p_j = P(X = x_j)$ とおく．このとき，X の分布関数と分位点関数をなるべく明示的に表現せよ．$x_1 < \cdots < x_k$ と仮定してよい．

3. F を $\mathbb{R}$ 上の分布関数とし，$x \in \mathbb{R}$ とする．このとき，$F(x) = F(x-)$ なら，F は x で連続であることを示せ．

4. $\mathbb{R}$ 上の分布関数 F に対して，その サポート (support) を
$$S = \{x \in \mathbb{R} : F(x + \varepsilon) - F(x - \varepsilon) > 0 \ \forall \varepsilon > 0\}$$
と定義する．S が閉集合であることを示せ．ヒント：S の補集合 S^c が開集合であることを示せばよい．

5. X を 0 以上の整数に値をとる確率変数とすると，
$$E[X] = \sum_{k=0}^{\infty} P(X > k)$$
と表せることを示せ．

6. $C > 0$ を定数とし，
$$f(x) = \frac{C}{1 + x^2}, \ x \in \mathbb{R}$$
と定義する．このとき，f が確率密度関数になるような C の値を求めよ．また，X がこの密度関数をもつとき，$E[|X|] = +\infty$ を示せ．この密度関数をもつ分布を Cauchy 分布 と呼ぶ．

2016.4.21. 出題：加藤賢悟

宿題 2

- 提出期限：4/28 の講義終了時.

問題

記号は講義で用いたものに従うとする．以下，$(\Omega, \mathcal{F}, P)$ を確率空間とする．

1. $x \in \mathbb{R}$ に対して，$\lfloor x \rfloor$ を x を超えない最大の整数とする．$X \sim Ex(\lambda)$ に対して，$\lfloor X \rfloor$ の確率関数を求めよ．

2. 関数
$$f(x) = \frac{1}{2} e^{-|x|}, \ x \in \mathbb{R} \tag{*}$$
が確率密度関数になることを確かめよ．さらに，X がこの密度関数をもつとき，モーメント母関数を求め，$E[X]$ と $\mathrm{Var}(X)$ を計算せよ．
コメント：(*) の密度関数をもつ分布を Laplace 分布 と呼ぶ．

3. (a). $t \in \mathbb{R}$ に対して，積分
$$\int_0^\infty e^{-x} \cos(tx) dx$$
を計算せよ．この結果を使って，Laplace 分布の特性関数を求めよ．
(b). (a) の結果と密度関数に対する反転 (逆転) 公式を認めて，Cauchy 分布 (宿題 1 問題 6 を参照) の特性関数を求めよ．

4. ベータ関数 $B(\alpha, \beta)$ に対して，
$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$
という関係は認めて，$X \sim Be(\alpha, \beta)$ に対して，$E[X]$ と $\mathrm{Var}(X)$ を求めよ．

5. 確率ベクトル (X, Y) の同時分布関数を $F(x, y) = P(X \le x, Y \le y)$ とおいて，$x, y \in \mathbb{R}, \Delta x, \Delta y > 0$ に対して，
$$\begin{aligned} &P(x < X \le x + \Delta x, y < Y \le y + \Delta y) \\ &\quad = F(x + \Delta x, y + \Delta y) - F(x, y + \Delta y) - F(x + \Delta x, y) + F(x, y) \end{aligned}$$
を示せ．

2016.4.28. 出題：加藤賢悟

宿題 3

- 提出期限：5/9 の講義終了時.

問題

記号は講義で用いたものに従うとする．以下，$(\Omega, \mathcal{F}, P)$ を確率空間とする．

1. 次をみたす離散型の確率変数 X, Y の例を構成せよ:X, Y は独立でないが，$\mathrm{Cov}(X, Y) = 0$ となる.

2. (Schwarz の不等式). $E[X^2] < \infty, E[Y^2] < \infty$ なら，
$$E[|XY|] \le \sqrt{E[X^2]}\sqrt{E[Y^2]}$$
となることを示せ.
ヒント：$X, Y \ge 0$ と仮定してよい．$E[(X - tY)^2] \ge 0$ なる不等式がすべての $t \in \mathbb{R}$ に対して成り立つことを使う．$E[Y^2] = 0$ の場合にも注意を払うこと.

3. $X = (X_1, \ldots, X_n)'$ を n 次元の確率ベクトルとし，$a = (a_1, \ldots, a_m)' \in \mathbb{R}^m$ と $m \times n$ 行列 B に対して，$Y = a + BX$ とおく．このとき，
$$E[Y] = a + BE[X], \quad \mathrm{Var}(Y) = B\,\mathrm{Var}(X)B'$$
を示せ (それぞれの場合において，有限な $E[X]$ と $\mathrm{Var}(X)$ の存在は仮定する).

4. 次の関係を示せ.
$$\text{(a) } Bin(n, p) * Bin(m, p) = Bin(n + m, p),$$
$$\text{(b) } Po(\lambda) * Po(\kappa) = Po(\lambda + \kappa).$$

5. X_1, X_2, X_3 を独立とし，$Ex(\lambda)$ に従うとする．ここで，$\lambda > 0$ とする．いま，
$$Z_1 = \frac{X_1}{X_1 + X_2},\ Z_2 = \frac{X_1 + X_2}{X_1 + X_2 + X_3},\ Z_3 = X_1 + X_2 + X_3$$
とおくと，Z_1, Z_2, Z_3 が独立であることを示せ．また，それぞれの周辺分布を求めよ.

6. $p_1, \ldots, p_k \ge 0$ は $\sum_{j=1}^k p_j = 1$ をみたすとし，$(Y_1, \ldots, Y_k)' \sim Mn(n, p_1, \ldots, p_k)$ とする．このとき，$j \ne \ell$ に対して，$\mathrm{Cov}(Y_j, Y_\ell) = -np_jp_\ell$ を示せ.

2016.5.12. 出題：加藤賢悟

宿題 4

- 提出期限：5/19 の講義終了時.

問題

記号は講義で用いたものに従うとする．以下，$(\Omega, \mathcal{F}, P)$ を確率空間とする．

1. (X, Y) を 2 次元の連続型確率ベクトルとする．また，$g: \mathbb{R}^2 \to \mathbb{R}$ を $E[g(X,Y)^2] < \infty$ をみたす関数とし，$h: \mathbb{R} \to \mathbb{R}$ を $E[h(X)^2] < \infty$ をみたす関数とする．このとき，
$$E[h(X)\{g(X,Y) - E[g(X,Y) \mid X]\}] = 0$$
を示せ.

2. (a) $X \sim t(m)$ とする．このとき，$0 < r < m$ に対して，$E[|X|^r] < \infty$ であって，$r \geq m$ に対して，$E[|X|^r] = \infty$ となることを示せ．(b) f_n を自由度 n の t 分布の密度関数とし，ϕ を $N(0,1)$ の密度関数とする．Stirling の公式は認めて，$n \to \infty$ のとき，各 $x \in \mathbb{R}$ に対して，$f_n(x) \to \phi(x)$ を示せ.

3. f を $\mathbb{R}$ 上の密度関数とし，$X_1, \ldots, X_n \sim f$ i.i.d. とする．また，$X_{(1)} \leq \cdots \leq X_{(n)}$ を順序統計量とする．このとき，$(X_{(1)}, \ldots, X_{(n)})$ は同時密度
$$g(x_1, \ldots, x_n) = n! f(x_1) \cdots f(x_n), \quad x_1 < \cdots < x_n$$
をもつことを示せ.

4. X_n, Y_n を r.v.'s とし，c を定数とする．
   (a) $X_n \xrightarrow{d} c$ なら $X_n \xrightarrow{P} c$ となることを示せ.
   (b) $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{P} c$ とする．このとき，$Y_n X_n \xrightarrow{d} cX$ を示せ.

5. $u \in (0,1)$ に対して，
$$\rho_u(x) = \{u - I(x \leq 0)\}x$$
をチェック関数と呼ぶ．例えば，$u = 1/2$ なら，$\rho_{1/2}(x) = |x|/2$ である．このとき，$x, y \in \mathbb{R}$ に対して，
$$\rho_u(x-y) - \rho_u(x) = -y\{u - I(x \leq 0)\} + y \int_0^1 \{I(x \leq ys) - I(x \leq 0)\} ds$$
を示せ.

6. $X_1, \dots, X_n$ を r.v.'s とし，$F_n(x) = n^{-1} \sum_{i=1}^n I(X_i \le x)$ とおく．F_n は 経験分布関数 empirical distribution function) と呼ばれる．F_n の分位点関数を $F_n^{\leftarrow}(u)$ とする：

$$F_n^{\leftarrow}(u) = \inf\{x \in \mathbb{R} : F_n(x) \ge u\}, \ u \in (0,1).$$

$F_n^{\leftarrow}(u)$ を 標本 u 分位点 と呼ぶ．

(a) $X_1, \dots, X_n$ にタイはないとし，$X_{(1)} < \cdots < X_{(n)}$ を順序統計量とする．このとき，$F_n^{\leftarrow}(u)$ を順序統計量を用いて表現せよ．

(b) $u \in (0,1)$ に対して，$F_n^{\leftarrow}(u)$ は最小化問題

$$\min_{x \in \mathbb{R}} \sum_{i=1}^n \rho_u(X_i - x)$$

の最適解であることを示せ．

2016.5.19. 出題：加藤賢悟

宿題 5

- 提出期限：5/26 の講義終了時.

問題

記号は講義で用いたものに従うとする．以下，$(\Omega, \mathcal{F}, P)$ を確率空間とする．

1. $(\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$ を未知とし，$X \sim N(\mu, \sigma^2)$ とする $(n = 1)$．このとき，σ^2 の不偏推定量は存在しないことを示せ．

   ヒント：仮に σ^2 の不偏推定量 δ が存在したとする．このとき，$Y, Z \sim N(0, 1)$ i.i.d. に対して，$E[\delta(Y + Z) \mid Z]$ と $E[\delta(Y + Z)]$ を比較せよ．

2. $0 < \theta_1 < \theta_2 < \infty$ とし，$X_1, \dots, X_n \sim U(\theta_1, \theta_2)$ i.i.d. とする．このとき，$(X_{(1)}, X_{(n)})$ が (θ_1, θ_2) に対する十分統計量であることを示せ．

3. $X \sim Bin(n, \theta), \theta \in (0, 1)$ に対して，$\theta(1 - \theta)$ の UMVU 推定量を求めよ．

4. $\mu \in \mathbb{R}, \sigma^2 > 0$ に対して，$X_1, \dots, X_n \sim N(\mu, \sigma^2)$ i.i.d. とする．また，$\overline{X} = n^{-1} \sum_{i=1}^n X_i, S^2 = (n-1)^{-1} \sum_{i=1}^n (X_i - \overline{X})^2$ とおく．このとき，$S = \sqrt{S^2}$ に対して，$E_{(\mu,\sigma^2)}[S]$ を求めよ．また，この結果を使って，$\sigma = \sqrt{\sigma^2}$ の UMVU 推定量を導出せよ．

5. (問題 4 の続き)．μ^2 の推定を考えると，その UMVU 推定量は $\delta(\overline{X}, S^2) = \overline{X} - S^2/n$ であった．このとき，各 (μ, σ^2) に対して，$P_{(\mu,\sigma^2)}(\delta < 0) > 0$ を示せ．さらに，$\delta^+ = \max\{\delta, 0\}$ とおくと，$E_{(\mu,\sigma^2)}[(\delta^+ - \mu^2)^2] < E_{(\mu,\sigma^2)}[(\delta - \mu^2)^2]$ となることを示せ．

6. $\lambda > 0$ に対して，$X_1, \dots, X_n \sim Ex(\lambda)$ i.i.d. とする $(n \geq 2)$．このとき，$T = \sum_{i=1}^n X_i$ が λ に対する完備十分統計量であることを示せ．また，λ の UMVU 推定量を求め，その分散を評価せよ．

   ヒント：$Ex(\lambda) = Ga(1, 1/\lambda)$ より，$T \sim Ga(n, 1/\lambda)$．$E_\lambda[1/T]$ を計算してみよ．

2016.5.19. 出題：加藤賢悟

宿題 6 (提出不要)

問題

記号は講義で用いたものに従うとする．以下，$(\Omega, \mathcal{F}, P)$ を確率空間とする．

1. $\mu \in \mathbb{R}, \tau > 0$ に対して，$N(\mu, \tau)$ の Fisher 情報行列 $I(\mu, \tau)$ を計算せよ．

2. f を Cauchy 分布の密度関数とする．このとき，分布族 $\{\sigma^{-1} f((\cdot - \mu)/\sigma) : \mu \in \mathbb{R}, \sigma > 0\}$ に対して，Fisher 情報行列 $I(\mu, \sigma)$ を計算せよ．

3. (デルタ法). Y_n を r.v.'s とし，ある $\theta \in \mathbb{R}$ と $\sigma^2 > 0$ に対して，$\sqrt{n}(Y_n - \theta) \xrightarrow{d} N(0, \sigma^2)$ とする．また，$g : \mathbb{R} \to \mathbb{R}$ を θ で微分可能な関数とする．このとき，$\sqrt{n}(g(Y_n) - g(\theta)) \xrightarrow{d} N(0, g'(\theta)^2 \sigma^2)$ を示せ．

4. $\mu \in \mathbb{R}, \sigma^2 > 0$ に対して，$X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ i.i.d. とする．このとき，$\sigma = \sqrt{\sigma^2}$ の MLE $\widehat{\sigma}$ に対して，$\sqrt{n}(\widehat{\sigma} - \sigma)$ の $n \to \infty$ のときの漸近分布を導出せよ．

5. $\theta > 0$ に対して，$[0, \theta]$ 上の一様分布を $U[0, \theta]$ と表す．$X_1, \ldots, X_n \sim U[0, \theta]$ i.i.d. に対して，MLE が $\widehat{\theta} = X_{(n)}$ で与えられることを確認せよ．また，$n \to \infty$ のとき，

$$n(\theta - X_{(n)}) \xrightarrow{d} Ex(1/\theta)$$

を示せ．従って，この例に対しては，MLE の漸近正規性は成り立たない．

2016.6.16. 出題：加藤賢悟

宿題 7

- 提出期限：6/23 の講義終了時.

問題

1. (Hardy-Weinberg 平衡モデル). ある座位 (locus) に 2 つのアレル (allele) A, a があって，それぞれ確率 $\theta, 1-\theta$ で出現するとすると，遺伝子型 (genotype) AA, Aa, aa が出現する確率はそれぞれ $\theta^2, 2\theta(1-\theta), (1-\theta)^2$ である．いま，n 個の個体のうち，遺伝子型 AA, Aa, aa をもつ個体の個数をそれぞれ Y_1, Y_2, Y_3 とおくと，$(Y_1, Y_2, Y_3)'$ は多項分布 $Mn(n, \theta^2, 2\theta(1-\theta), (1-\theta)^2)$ に従う．
   (a) θ の MLE $\widehat{\theta}$ を導出せよ．
   (b) $n \to \infty$ のときの $\sqrt{n}(\widehat{\theta} - \theta)$ の極限分布を求めよ．
   (c) $\alpha, \beta > 0$ に対して，$\theta \sim Be(\alpha, \beta)$ という事前分布を考える．このとき，θ の事後平均 $\widetilde{\theta}^{\alpha,\beta}$ を求めよ．
   (d) θ の真値を 1 つ固定したとき，$n \to \infty$ のとき，$\sqrt{n}(\widetilde{\theta}^{\alpha,\beta} - \widehat{\theta}) \xrightarrow{P} 0$ を示せ．

2. $X_1, \ldots, X_n \mid \lambda \sim Po(\lambda)$ i.i.d., $\lambda \sim Ga(\alpha, \beta)$ とする $(\alpha > 0, \beta > 0)$. このとき，λ の事後平均を求めよ．

3. $X \sim N(\mu, I_k), \mu \in \mathbb{R}^k$ とする．
   (a) $k = 1, 2$ のとき，$E_0[1/\|X\|^2] = \infty$ となることを示せ．ヒント：$\mu = 0$ のとき，$\|X\|^2 \sim \chi^2(k)$ である．
   (b) $k \geq 3$ のとき，$E_\mu[1/\|X\|^2] < \infty \ \forall \mu \in \mathbb{R}^k$ であることを示せ．
   ヒント：$\|x - \mu\|^2 = \|x\|^2 - 2x'\mu + \|\mu\|^2$ と不等式 $2ab \leq a^2/2 + 2b^2$ より，$\|x - \mu\|^2 \geq \|x\|^2/2 - \|\mu\|^2$．あとは極座標変換を使う．

4. $X \sim N(\mu, 1), \mu \in \mathbb{R}$ とし，μ の推定を考える．$a, b \in \mathbb{R}$ に対して，$\delta_{a,b}(X) = aX + b$ という推定量を考える．このとき，2 乗損失関数 $L(\mu, d) = (d - \mu)^2$ の下で，次の場合に $\delta_{a,b}$ が非許容的になることを示せ．
   (1) $a > 1$, または $a = 1$ かつ $b \neq 0$. ヒント：$\delta_{a,b}(X)$ と X を比較せよ．
   (2) $a < 0$.

5. F を連続な分布関数とし，$X \sim F$ とする．このとき，$F(X) \sim U(0, 1)$ を示せ．

2016.6.23. 出題：加藤賢悟

宿題 8

- 提出期限：6/30 の講義終了時.

問題

$\alpha \in (0,1)$ は所与とする.

1. 検定問題
$$H_0 : \theta = \theta_0 \text{ vs. } H_1 : \theta \neq \theta_0$$
に対して，独立な検定統計量 $T_1, \ldots, T_k$ が与えられたとして，各 T_j に対してその p 値を P_j とおく．各 T_j が θ_0 のもとで連続な d.f. をもつとき，
$$-2\sum_{j=1}^{k} \log P_j > \chi^2_\alpha(2k) \Rightarrow \text{reject}$$
という検定はサイズ α をもつことを示せ．ここで，$\chi^2_\alpha(2k)$ は $\chi^2(2k)$ の $(1-\alpha)$ 分位点である．
ヒント：$U \sim U(0,1)$ に対して，$-2\log U \sim \chi^2(2)$ である．

2. Neyman-Pearson の補題において，$p_n(x;\theta), \theta \in \{\theta_0, \theta_1\}$ を確率関数とする．いま，$S = \{x : p_n(x;\theta_1) \neq cp_n(x;\theta_0)\}$ とおくと，δ が水準 α の MP 検定なら，すべての $x \in S$ に対して $\delta(x) = \delta_{c,\gamma}(x)$ となることを示せ．

3. $X_1, \ldots, X_n \sim N(0,\sigma^2)$ i.i.d. とする．検定問題
$$H_0 : \sigma^2 \leq 1 \text{ vs. } H_1 : \sigma^2 > 1$$
に対する水準 α の UMP 検定を求めよ．

4. $X_1, \ldots, X_n \sim Po(\lambda)$ i.i.d. として，
$$H_0 : \lambda \leq 1 \text{ vs. } H_1 : \lambda > 1$$
に対する水準 α の UMP 検定を求めよ．

5. $\alpha > 0, \beta > 0$ に対して，

$$f(t) = \alpha^\beta \beta t^{-(1+\beta)} I(t > \alpha)$$

という密度関数をもつ分布を Pareto 分布 と呼び，$Pa(\alpha, \beta)$ と表す．Pareto 分布は所得の分布のモデリングに用いられ，β は Pareto 指数と呼ばれる．いま，$c > 0$ は既知，$\theta > 1$ は未知として，

$$X_1, \ldots, X_n \sim Pa(c, \theta) \text{ i.i.d.}$$

とする．

(a) $Pa(c, \theta)$ の平均 μ を θ を用いて表せ．

(b) $\mu_0 > c$ を所与として，$H_0 : \mu = \mu_0$ vs. $H_0 : \mu > \mu_0$ という検定問題に対して，

$$\sum_{i=1}^n \log X_i > \kappa \Rightarrow \text{reject}$$

という形の検定が UMP 検定になることを示せ．

ヒント：$\mu_1 > \mu_0$ として，$H_0 : \mu = \mu_0$ vs. $H_1 : \mu = \mu_1$ という検定問題に対して MP 検定を構成せよ．

(c) $2\theta \log(X_i/c) \sim \chi^2(2)$ を示し，これを利用して (b) の検定がサイズ α になるような κ の値を求めよ．

(d) 中心極限定理を使って (b) の検定が近似的にサイズ α をもつような κ の値を求めよ．

6. 与えられた定数 $\mu \neq 0, c_1 < c_2$ に対して，

$$e^{\mu t} > a + bt \Leftrightarrow t < c_1 \text{ or } t > c_2, \quad e^{\mu t} < a + bt \Leftrightarrow c_1 < t < c_2$$

をみたす a, b が

$$a = \frac{c_2 e^{\mu c_1} - c_1 e^{\mu c_2}}{c_2 - c_1}, \ b = \frac{e^{\mu c_2} - e^{\mu c_1}}{c_2 - c_1}$$

で与えられることを示せ．

2016.6.30. 出題：加藤賢悟

宿題 9

- 提出期限：7/7 の講義終了時．

問題

1. $X_1, \dots, X_n \sim Ex(\lambda)$ i.i.d. として，
$$H_0 : \lambda = 1 \text{ vs. } H_1 : \lambda \neq 1$$
に対する水準 α の UMPU 検定を求めよ．
ヒント：必要なら，$2\lambda \sum_{i=1}^n X_i \sim \chi^2(2n)$ という事実を使え．

2. $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ i.i.d. とし $(n \geq 2)$，$\mu \in \mathbb{R}$ と $\sigma^2 > 0$ は未知とする．このとき，
$$H_0 : \mu = \mu_0 \text{ vs. } H_1 : \mu \neq \mu_0$$
に対する尤度比検定が t 検定と等価であることを示せ．ただし，t 検定とは，
$$T = \frac{\sqrt{n}(\overline{X} - \mu_0)}{S}$$
とおくと，
$$|T| > c \Rightarrow \text{reject}$$
という形の検定である．

3. $X_1, \dots, X_n \sim Po(\lambda)$ i.i.d. とする $(\lambda > 0)$．
   (a) λ の最尤推定量 $\widehat{\lambda}$ を求め，$\sqrt{n}(\widehat{\lambda} - \lambda)$ の極限分布を導出せよ．
   (b) 検定問題
   $$H_0 : \lambda = \lambda_0 \text{ vs. } H_1 : \lambda \neq \lambda_0$$
   を考える $(\lambda_0 > 0)$．(a) の結果を使って，近似的にサイズ α をもつ検定を構成せよ．

4. $(Y_1, Y_2, Y_3)' \sim Mn(n, p_1, p_2, p_3)$ に対して，$(Y_1, Y_2, Y_3)'$ が Hardy-Weinberg 平衡モデルに従っているかどうか検定したいとする．すなわち，
$$H_0 : (p_1, p_2, p_3) = (\theta^2, 2\theta(1-\theta), (1-\theta)^2) \ \exists \theta \in (0,1) \text{ vs. } H_1 : H_0 \text{ の否定}$$
という検定問題を考える．この検定問題に対する尤度比検定を説明せよ．

5. $(\Omega, \mathcal{F}, P)$ を確率空間とする．Ω 上の r.v.'s X, Y に対して，

$$\alpha(X,Y) = \inf\{\varepsilon \geq 0 : P(|X-Y| > \varepsilon) \leq \varepsilon\} \tag{*}$$

と定義する．

(a) (*) において inf は達成されることを示せ．すなわち，$\alpha = \alpha(X,Y)$ に対して，$P(|X-Y| > \alpha) \leq \alpha$ を示せ．ヒント：$|X-Y|$ の d.f. を F とおくと，$\alpha(X,Y) = \inf\{\varepsilon \geq 0 : 1 - F(\varepsilon) \leq \varepsilon\}$ である．あとは F の右連続性を使う．

(b) L^0 を Ω 上の r.v.'s の全体とする．このとき，α は次の性質をみたすことを示せ：(1) $X, Y \in L^0$ に対して，$\alpha(X,Y) = 0 \Leftrightarrow P(X = Y) = 1$. (2) $X, Y, Z \in L^0$ に対して，$\alpha(X,Z) \leq \alpha(X,Y) + \alpha(Y,Z)$.

コメント：α は明らかに非負性と対称性をみたすので，確率 1 で等しい r.v.'s を同一視すれば，α は L^0 上の距離になる．この距離 α は Ky Fan 距離 と呼ばれる．

6. (問題 5 の続き). r.v.'s $X, X_n, n = 1, 2, \ldots$ に対して，

$$X_n \xrightarrow{P} X \Leftrightarrow \alpha(X_n, X) \to 0$$

を示せ．

コメント：よって，$X_n \xrightarrow{P} 0$ なら，$\varepsilon_n = \alpha(X_n, X)$ とおくと，$\varepsilon_n \to 0$ & $P(|X_n - X| > \varepsilon_n) \leq \varepsilon_n$ となる．

2016.7.7. 出題：加藤賢悟

宿題 10

- 提出期限：7/14 の講義終了時.

問題

1. $X_1, \ldots, X_n \sim Po(\lambda)$ i.i.d. $(\lambda > 0)$ とする．$\widehat{\lambda} = \overline{X}$ とおく．このとき，

$$H_0 : \lambda = \lambda_0 \text{ vs. } H_1 : \lambda \neq \lambda_0$$

という検定問題に対して，

$$\frac{\sqrt{n}|\widehat{\lambda} - \lambda_0|}{\sqrt{\lambda_0}} > z_{\alpha/2} \Rightarrow \text{reject}$$

という検定は近似的にサイズ α をもつのであった $(z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2))$．この検定の受容域を反転させて λ に対して近似的に水準 $(1 - \alpha)$ の CI を構成せよ．

2. (問題 1 の続き)．デルタ法を認めて，$\sqrt{n}(g(\widehat{\lambda}) - g(\lambda)) \xrightarrow{d} N(0, 1)$ となるような関数 g を 1 つ求めよ．この結果を利用して，λ に対する CI を構成せよ．

3. $X \sim Ga(k, \nu)$ とする．ただし，k は正の整数であって，$\nu > 0$ とする．このとき，$2X/\nu \sim \chi^2(2k)$ を示せ．この関係を利用して，$Ga(k, \nu)$ の $(1 - \alpha)$ 分位点を $\chi^2(2k)$ の $(1 - \alpha)$ 分位点 $\chi^2_\alpha(2k)$ を用いて表せ．

4. (問題 1 の続き)．$\lambda \sim Ga(k, \nu)$ という事前分布を入れる．ここで，k は正の整数であって，$\nu > 0$ とする．$\widehat{\zeta}_\alpha$ を λ の事後分布の $(1 - \alpha)$ 分位点とすると，$\widehat{\zeta}_\alpha$ を χ^2 の分位点を用いて表せ．

5. $X_n \sim Bin(n, \theta), \theta \in (0, 1)$ とし，$g(\theta) = \theta(1 - \theta)$ に対して水準 $(1 - \alpha)$ の CI を構成することを考える．θ の MLE を $\widehat{\theta} = X_n/n$ とおき，デルタ法を認めて，$\sqrt{n}(g(\widehat{\theta}) - g(\theta))$ の極限分布を求めよ．これにもとづいて $g(\theta)$ に対する CI を構成せよ．

2016.7.14. 出題：加藤賢悟

宿題 11

- 提出期限：7/21 の講義終了時.

問題

1. $X_1, \dots, X_n \sim U(0,1)$ i.i.d. とすると，$0 < u < v < 1$ に対して，

$$P(X_{(1)} \le u, X_{(n)} \ge v) = 1 - (1-u)^n - v^n + (v-u)^n$$

を示せ.

2. $X = (X_1, \dots, X_k)', X^n = (X_1^n, \dots, X_k^n)'$ を k 次元の確率ベクトルとする．このとき，

$$X^n \xrightarrow{P} X \Leftrightarrow X_j^n \xrightarrow{P} X_j \ \forall j = 1, \dots, k$$

を示せ．ヒント：$x = (x_1, \dots, x_k)' \in \mathbb{R}^k$ に対して，$\|x\| \ge |x_j|$ $(j = 1, \dots, k)$, $\|x\| \le \sum_{j=1}^k |x_j|$ である.

3. k 次元の確率ベクトルの列 $X^n = (X_1^n, \dots, X_k^n)'$ に対して，$X^n = O_P(1)$ であることと，各 $j = 1, \dots, k$ に対して $X_j^n = O_P(1)$ であることは同値であることを示せ.

4. X_n, Y_n を 1 次元の r.v.'s とし，$X_n = O_P(1), Y_n = o_P(1)$ とする．このとき，次の関係を示せ.

   (a) $X_n + Y_n = O_P(1)$.

   (b) $Y_n X_n = o_P(1)$.

5. F を $\mathbb{R}$ 上の d.f. とし，$X_1, \dots, X_n \sim F$ i.i.d. とする．$\widehat{F}_n(x) = n^{-1} \sum_{i=1}^n I(X_i \le x)$ とおいて，$x < y$ に対して，

$$\sqrt{n} \begin{pmatrix} \widehat{F}_n(x) - F(x) \\ \widehat{F}_n(y) - F(y) \end{pmatrix}$$

の極限分布を求めよ.

2016.7.14. 出題：加藤賢悟

宿題 12 (提出不要)

問題

1. $\theta_1 < \theta_2$ とし，$X_1, \dots, X_n \sim U[\theta_1, \theta_2]$ i.i.d. とする．

   (a) (θ_1, θ_2) の MLE が $(X_{(1)}, X_{(n)})$ で与えられることを確認せよ．

   (b) $n(X_{(1)} - \theta_1, \theta_2 - X_{(n)})$ の極限分布を求めよ．

2. 密度関数
$$f(x) = \frac{e^{-x}}{(1+e^{-x})^2},\ x \in \mathbb{R}$$
をもつ分布をロジスティック分布という．いま，$\theta \in \mathbb{R}$ に対して，$f(x;\theta) = f(x-\theta)$ とおき，$f(\cdot;\theta)$ を密度関数にもつ d.f. を F_θ とおく．

   (a) F_θ のメディアンが θ に一致することを示せ．

   (b) $X_1, \dots, X_n \sim F_\theta$ i.i.d. として，$\widehat{\theta}$ を標本メディアンとする．このとき，$\sqrt{n}(\widehat{\theta} - \theta)$ の極限分布を求めよ．

   (c) F_θ の Fisher 情報量 $I(\theta)$ を求めよ．

   (d) (b) で求めた極限分布の分散が $1/I(\theta)$ より大きいことを確認せよ．標本メディアンにワンステップ推定を適用して，極限分布の分散が $1/I(\theta)$ を達成する推定量を構成せよ．

3. $X_1, \dots, X_n \sim U[0, \theta]$ i.i.d. $(\theta > 0)$ として，MLE $\widehat{\theta} = X_{(n)}$ と UMVU 推定量 $\widetilde{\theta} = \{(n+1)/n\}X_{(n)}$ を考える．このとき，
$$\lim_{n\to\infty} E_\theta[\{n(\widehat{\theta} - \theta)\}^2], \quad \lim_{n\to\infty} E_\theta[\{n(\widetilde{\theta} - \theta)\}^2]$$
を比較せよ．

B　その他の演習問題

第 1 節

1. 任意の関数 $X:\Omega\to\mathbb{R}$ に対して，次の (a)–(d) を示せ．

   (a) 任意の集合族 $A_i\subset\mathbb{R}, i\in\mathcal{I}$ に対して，

   $$X^{-1}\left(\bigcup_{i\in\mathcal{I}}A_i\right)=\bigcup_{i\in\mathcal{I}}X^{-1}(A_i),\ X^{-1}\left(\bigcap_{i\in\mathcal{I}}A_i\right)=\bigcap_{i\in\mathcal{I}}X^{-1}(A_i).$$

   (b) 任意の $A\subset\mathbb{R}$ に対して，$X^{-1}(A^c)=(X^{-1}(A))^c$.

   (c) $x_n\downarrow x$ に対して，$\bigcap_{n=1}^{\infty}\{X\le x_n\}=\{X\le x\}$.

   (d) $x_n\uparrow x, x_n<x$ に対して，$\bigcup_{n=1}^{\infty}\{X\le x_n\}=\{X<x\}$.

2. F を $\mathbb{R}$ 上の d.f. とし，$g:\mathbb{R}\to\mathbb{R}$ を狭義単調増加とする．このとき，$X\sim F$ に対して，$Y=g(X)$ の d.f. を $G(y)=P(Y\le y)$ とおくと，G の分位点関数が $G^{\leftarrow}(u)=g(F^{\leftarrow}(u))$ で与えられることを示せ．

3. F,G を $\mathbb{R}$ 上の d.f.'s とする．F が G に対して確率優位にあるとは，$F(x)\le G(x)\ \forall x\in\mathbb{R}$ となることをいう．このとき，$F^{\leftarrow}(u)\ge G^{\leftarrow}(u)\ \forall u\in(0,1)$ を示せ．

4. $X\sim F, Y\sim G$ とし，$g:\mathbb{R}\to\mathbb{R}$ を単調非減少関数であって，$E[|g(X)|]<\infty, E[|g(Y)|]<\infty$ を仮定する．F が G に対して確率優位にあるなら，$E[g(X)]\ge E[g(Y)]$ となることを示せ．ヒント：$U\sim U(0,1)$ に対して，$F^{\leftarrow}(U)\sim F$.

5. r.v. X が連続な密度関数 f をもつなら，X の d.f. のサポートは $\{x\in\mathbb{R}:f(x)\ne 0\}$ の閉包に一致することを示せ．

6. X を (a,b) に値をとる r.v. とし，(a,b) 上に連続な密度関数 f_X をもつとする．$g:(a,b)\to\mathbb{R}$ が狭義単調減少かつ C^1 級なら，$Y=g(X)$ の密度関数が

   $$f_Y(y)=f_X(g^{-1}(y))\frac{1}{|g'(g^{-1}(y))|},\ y\in(g(b),g(a))$$

   で与えられることを示せ．

7. X を r.v. とし，$0<p<\infty$ とする．このとき，$P(|X|>x)=O\{x^{-p}(\log x)^{-2}\}\ (x\to\infty)$ なら $E[|X|^p]<\infty$ であることを示せ．

8. X を r.v. とし，ある $a>0$ が存在して，$E[e^{\theta X}]<\infty\ \forall|\theta|<a$ とする．このとき，任意の $k=1,2,\ldots$ と $|\theta|<a$ に対して，$E[|X|^k e^{\theta X}]<\infty$ を示せ．

9. $X \sim N(0,1), k = 1, 2, \ldots$ に対して，

$$E[X^k] = \begin{cases} 0 & k\,\text{が奇数のとき} \\ \frac{2^{k/2}\Gamma((k+1)/2)}{\Gamma(1/2)} & k\,\text{が偶数のとき} \end{cases}$$

を示せ．

10. (Chernoff の不等式) X を r.v. とすると，任意の $x > 0$ に対して，$P(X \geq x) \leq \inf_{t>0} e^{-tx} E[e^{tX}]$ を示せ．

11. $X \sim N(0,1)$ とし，$x > 0$ に対して，$P(X > x) \leq e^{-x^2/2}, P(|X| > x) \leq 2e^{-x^2/2}$ を示せ．

12. X を $E[X^2] < \infty$ なる r.v. とする．このとき，$g(t) = E[(X-t)^2]$ は $t = E[X]$ で最小化されることを示せ．

13. X を $E[X^2] < \infty$ なる r.v. とし，X のメディアンを $M(X)$ とおく．このとき，$|M(X) - E[X]| \leq \sqrt{2\,\mathrm{Var}(X)}$ を示せ．

14. X, Y を $E[|X|] < \infty, E[|Y|] < \infty$ なる r.v.'s とする．このとき，$E[\max\{X, Y\}] \geq \max\{E[X], E[Y]\}$ を示せ．

15. $g, h : \mathbb{R} \to \mathbb{R}$ をともに単調非減少関数とし，X を $E[|g(X)|] < \infty, E[|h(X)|] < \infty, E[|g(X)h(X)|] < \infty$ をみたす r.v. とする．このとき，

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

を示せ．g が単調非減少，h が単調非増加なら，逆の不等式が成り立つことを示せ．

ヒント：Y を X と同じ分布に従う独立な r.v. とすると，

$$\{g(X) - g(Y)\}\{h(X) - h(Y)\}$$

の符号を考察せよ．

16. Cauchy 分布の分布関数と分位点関数を求めよ．

17. Laplace 分布の分布関数と分位点関数を求めよ．

18. $f(x) = (1 - |x|)^+$ を密度関数にもつ分布を三角分布と呼ぶのであった．三角分布の特性関数が

$$\varphi(t) = \frac{2(1 - \cos t)}{t^2}$$

で与えられることを示せ．

19. 関数

$$f(x) = \frac{1-\cos x}{\pi x^2}$$

が確率密度関数になることを確認せよ．この密度関数をもつ分布を Pólya 分布と呼ぶ．Pólya 分布の特性関数が

$$\varphi(t) = (1-|t|)^+$$

で与えられることを示せ．

20. f を Cauchy 分布の密度関数として，$f * \cdots * f$ (n 回のたたみ込み) を求めよ．

21. $X, Y \sim N(0,1)$ とし，X, Y は独立とする．このとき，X/Y の分布を求めよ．

22. $0 < \alpha < 1, U \sim U(0,1), X \sim Ga(\alpha+1, 1)$ として，$Y = U^{1/\alpha}X$ とおく．このとき，$Y \sim Ga(\alpha, 1)$ を示せ．

23. $X_1, \ldots, X_n$ を (独立とは限らない) r.v.'s とし，$X = (X_1, \ldots, X_n)'$ とおく．このとき，X が多変量正規分布に従うことは，任意の $a \in \mathbb{R}^n$ に対して $a'X$ が正規分布に従うことと同値であることを示せ．

24. $(Y_1, \ldots, Y_k)' \sim Di(\alpha_1, \ldots, \alpha_k)$ とすると，$j \neq \ell$ に対して，

$$\mathrm{Cov}(Y_j, Y_\ell) = -\frac{\alpha_j \alpha_\ell}{(\sum_{i=1}^k \alpha)^2 (\sum_{i=1}^k \alpha_i + 1)}$$

であることを示せ．

25. X を連続な密度関数をもつ r.v. とすると，$n \to \infty$ のとき，$E[\cos^2(nX)] \to 1/2$ となることを示せ．

第 2 節

1. $X_n, n = 1, 2, \ldots$ を r.v.'s とし，各 X_n は確率 1 で 1 点 x_n をとるとする：$P(X_n = x_n) = 1$. このとき，X_n が分布収束することは数列 x_n が収束することと同値なことを示せ．

2. $X, X_n, n = 1, 2, \ldots$ を整数に値をとる r.v.'s とすると，$X_n \xrightarrow{d} X \Leftrightarrow \lim_n P(X_n = k) = P(X = k)\ \forall k \in \mathbb{Z}$ を示せ．

3. X_n を r.v.'s とし，$E[X_n] \to \mu \in \mathbb{R}, \mathrm{Var}(X_n) \to 0$ とする．このとき，$X_n \xrightarrow{P} \mu$ を示せ．

4. $X_n \sim Bin(n, p_n)$ とし，$np_n \to \lambda > 0$ ($n \to \infty$) とする．このとき，特性関数の収束を示すことにより，$X_n \xrightarrow{d} Po(\lambda)$ を示せ．

5. $X_1, \ldots, X_n \sim Ex(1)$ i.i.d. に対して，

$$Y_1 = nX_{(1)},\ Y_k = (n-k+1)(X_{(k)} - X_{(k-1)}),\ k = 2, \ldots, n$$

とおく．

(a) 次の関係を示せ．

$$(X_{(1)}, \ldots, X_{(i)}, \ldots, X_{(n)}) = \left(\frac{Y_1}{n}, \ldots, \sum_{k=1}^{i} \frac{Y_k}{n-k+1}, \ldots, \sum_{k=1}^{n} \frac{Y_k}{n-k+1} \right)$$

(b) $Y_1, \ldots, Y_n \sim Ex(1)$ i.i.d. となることを示せ．

この問題から，$E_1, \ldots, E_n \sim Ex(0,1)$ i.i.d. に対して，$(Y_1, \ldots, Y_n) \stackrel{d}{=} (E_n, \ldots, E_1)$ より，

$$(X_{(1)}, \ldots, X_{(i)}, \ldots, X_{(n)}) \stackrel{d}{=} \left(\frac{E_n}{n}, \ldots, \sum_{k=n-i+1}^{n} \frac{E_k}{k}, \ldots, \sum_{k=1}^{n} \frac{E_k}{k} \right) \qquad (*)$$

という関係が成り立つ．(*) を Rényi 表現 と呼ぶ．

6. $X_1, \ldots, X_n \sim F$ i.i.d. とし，$U(t) = F^{\leftarrow}(1-1/t), t > 1$ とおく．このとき，$E_1, \ldots, E_n \sim Ex(1)$ i.i.d. に対して，

$$(X_{(1)}, \ldots, X_{(n)}) \stackrel{d}{=} (U(e^{E_{(1)}}), \ldots, U(e^{E_{(n)}}))$$

を示せ．

7. $X_1, \ldots, X_n \sim U(0,1)$ i.i.d. とし，$n \geq 2$ とする．$X_1, \ldots, X_n$ を $(0,1)$ 上にランダムに散らばった点とみなす．このとき，$0 < d < 1/(n-1)$ に対して，$X_1, \ldots, X_n$ のどの 2 点も d より離れている確率を具体的に計算せよ．ヒント：求める確率は $P(X_{(i+1)} > X_{(i)} + d\ \forall i = 1, \ldots, n-1)$ である．

第 3 節

1. $X \sim N(0,1)$ とし，$T = |X|$ とおく．このとき，$E[|g(X)|] < \infty$ なる関数 g に対して，$E[g(X) \mid T]$ を求めよ．

2. $(X, Y) \sim N(0, I_2)$ とし，$T = X^2 + Y^2$ とおく．このとき，$E[|g(X,Y)|] < \infty$ なる関数 g に対して，$E[g(X,Y) \mid T]$ を求めよ．ヒント：極座標変換．

3. F を密度関数 f をもつ $\mathbb{R}$ 上の d.f. とし，$X_1, \ldots, X_n \sim F$ i.i.d. とする．このとき，$E[|g(X_1, \ldots, X_n)|] < \infty$ なる関数 g に対して，

$$E[g(X_1, \ldots, X_n) \mid X_{(1)}, \ldots, X_{(n)}] = \frac{1}{n!} \sum_{\sigma} g(X_{(\sigma(1))}, \ldots, X_{(\sigma(n))})$$

を示せ．ここで，$\sum_\sigma$ は $\{1, \ldots, n\}$ の置換 σ 全体にわたってとる．

4. 2 項分布，Poisson 分布，ガンマ分布，ベータ分布がそれぞれ指数型分布族をなすことを確認せよ．

5. f を $\mathbb{R}$ 上の確率密度関数とし，$f(x) > 0 \ \forall x \in \mathbb{R}$ とする．さらに，f は微分可能であって，

$$\int_{-\infty}^{\infty} \frac{\{f'(x)\}^2}{f(x)} dx < \infty$$

とする．このとき，分布族 $\{f(\bullet - \theta) : \theta \in \mathbb{R}\}$ に対して Fisher 情報量が

$$I(\theta) = \int_{-\infty}^{\infty} \frac{\{f'(x)\}^2}{f(x)} dx$$

で与えられることを示せ．

6. $X_1, \ldots, X_n \sim Po(\lambda)$ i.i.d. $(\lambda > 0)$ とする．

   (a) $T = \sum_{i=1}^n X_i$ が λ に対する完備十分統計量であることを示せ．

   (b) $g(\lambda) = e^{-\lambda}$ に対する UMVU 推定量を求めよ．ヒント：$e^{-\lambda} = P_\lambda(X_1 = 0)$ に注意して，$P(X_1 = 0 \mid T = t)$ を評価せよ．

   (c) $Po(\lambda)$ の Fisher 情報量を求めよ．

   (d) $g(\lambda) = e^{-\lambda}$ に対して Cramér-Rao の下界を求め，(b) で求めた UMVU 推定量の分散と比較せよ．

7. $(X_1, X_2, X_3)' \sim Mn(n, p_1, p_2, 1 - p_1 - p_2)$ $(p_1 > 0, p_2 > 0, p_1 + p_2 < 1)$ とする．

   (a) (X_1, X_2) が完備十分統計量であることを示せ．

   (b) Fisher 情報行列 $I(p_1, p_2)$ を求めよ．

   (c) $g(p_1, p_2) = p_1 p_2$ に対して UMVU 推定量を求めよ．また，求めた UMVU 推定量の分散と Cramér-Rao の下界を比較せよ．

8. Cramér-Rao の不等式において，$k = 1$ とし，$\delta(X)$ を θ の不偏推定量とする．このとき，$\delta(X)$ が Cramér-Rao の下界を達成するための必要十分条件について考察せよ．

9. $X_1, \ldots, X_n \sim Ex(\lambda)$ i.i.d. $(\lambda > 0)$ とする．

   (a) λ の MLE $\widehat{\lambda}$ を求めて，$n \to \infty$ のときの $\sqrt{n}(\widehat{\lambda} - \lambda)$ の極限分布を求めよ．

   (b) いま，$X_1, \ldots, X_n$ は直接観測できないが，与えられた $\delta > 0$ に対して，

   $$Y_i = k\delta \quad \text{if } k\delta \le X_i < (k+1)\delta, k = 0, 1, 2, \ldots$$

   が観測されているとする．このとき，$Y_1, \ldots, Y_n$ にもとづく λ の MLE $\widetilde{\lambda}$ を求め，$n \to \infty$ のときの $\sqrt{n}(\widetilde{\lambda} - \lambda)$ の極限分布を求めよ．

ヒント：$y \in \{k\delta : k = 0, 1, 2, \dots\}$ に対して，

$$P_\lambda(Y_i = y) = P_\lambda(y \le X_i < y + \delta)$$

である．

(c) (b) で求めた極限分布の分散を $\sigma^2(\lambda, \delta)$ とおいて，$\lim_{\delta \to 0} \sigma^2(\lambda, \delta)$ を求めよ．

10. $X \sim Ga(\alpha, 1/\beta)$ とし $(\alpha > 0, \beta > 0)$ に対して，$Y = 1/X$ の分布をパラメータ $(\alpha.\beta)$ の逆ガンマ分布 (inverse Gamma distribution) と呼び，$IG(\alpha, \beta)$ と表す．Y の密度関数を導出せよ．

11. $X_1, \dots, X_n \mid \eta \sim N(0, \eta)$ i.i.d., $\eta \sim IG(\alpha/2, \beta/2)$ とする $(\alpha > 0, \beta > 0)$．このとき，η の事後分布と事後平均を求めよ．

12. $X \sim N(\theta, 1)$ とし，θ は区間 $[-m, m]$ に制約されていることがわかっているとする $(0 < m < 1)$．

    (a) π を確率関数とし，$\pi(m) = \pi(-m) = 1/2$ とする．$\theta \sim \pi$ という事前分布を入れたとき，θ の事後分布は $\{m, -m\}$ にのみ正の確率をもつ離散分布である．θ の事後分布を求め，事後平均 $\widehat{\theta}^\pi$ を計算せよ．

    (b) 損失関数を $L(\theta, d) = (d - \theta)^2$ とすると，$R(\theta, \widehat{\theta}^\pi) \le r(\pi, \widehat{\theta}^\pi)\ \forall \theta \in [-m, m]$ を示せ．これから $\widehat{\theta}^\pi$ がミニマクスであることを導け．

第 4 節

1. F を $\mathbb{R}$ 上の (連続とは限らない) d.f. として，$X \sim F$ とする．このとき，X と独立な $U \sim U(0, 1)$ に対して，

$$Y = F(X-) + U(F(X) - F(X-))$$

とおくと，$Y \sim U(0, 1)$ となることを示せ．

2. 検定や区間推定において，小さい $\alpha \in (0, 1)$ に対する $N(0, 1)$ の $(1 - \alpha)$ 分位点 $\Phi^{-1}(1 - \alpha)$ がよく現れる．そこで，$\alpha \downarrow 0$ のときの $\Phi^{-1}(1 - \alpha)$ の挙動を調べてみる．

$$\lim_{\alpha \downarrow 0} \frac{\Phi^{-1}(1 - \alpha)}{\sqrt{2 \log(1/\alpha)}} = 1$$

を示せ．

3. f_0, f_1 を $\mathbb{R}^k$ 上の確率 (密度) 関数とし，$X \sim f_i\ (i = 1, 2)$ とする．また，ある関数 $T : \mathbb{R}^k \to \mathbb{R}$ と単調非減少関数 $g : \mathbb{R} \to \mathbb{R}_+$ が存在して，

$$f_1(x) = g(T(x)) f_0(x)$$

と表せるとする．$X \sim f_i$ のときの $T = T(X)$ の d.f. を F_i^T とおくと，F_1^T は F_0^T に対して確率優位にあることを示せ．

ヒント：$t \in \mathbb{R}$ を任意に固定して，$h(y) = I(y \le t)$ とおくと，$F_1^T(t) = E_1[h(T)] = E_0[g(T)h(T)]$ と表せる．

4. $Y \sim Mn(n, p_1, \ldots, p_k)$ として，与えられた $(p_{0,1} \ldots, p_{0,k-1})' \in \Delta$ に対して，H_0 : $p_j = p_{0,j}$ $(1 \le \forall j \le k-1)$ vs. H_0 の否定，という検定問題を考える．このとき，Y にもとづく LRT 検定統計量は，$X_1., \ldots, X_n \sim Mn(1, p_1, \ldots, p_k)$ i.i.d. にもとづく LRT 検定統計量に分布の意味で等しいことを示せ．

5. （続き)．独立性の検定問題 $H_0 : p_{i,j} = p_{i+}p_{+j}$ $(\forall(i,j))$ vs. H_0 の否定，に対しても同様の結果を示せ．

6. $(Y_1, Y_2, Y_3, Y_4)' \sim Mn(n, p_1, p_2, p_3, p_4)$ に対して，

$$\begin{aligned} H_0 : (p_1, p_2, p_3, p_4) &= \left(\frac{1}{4}(2+\theta), \frac{1}{4}(1-\theta), \frac{1}{4}(1-\theta), \frac{1}{4}\theta\right) \ (\exists \theta \in (0,1)) \\ \text{vs. } H_1 &: H_0 \text{ の否定} \end{aligned}$$

という検定問題を考える．帰無仮説のモデルは遺伝学において Fisher の遺伝的連鎖モデル (genetic linkage model) と呼ばれるものである．

(a) H_0 のもとでの θ の MLE $\widehat{\theta}$ を求めよ．

(b) この検定問題に対して，尤度比検定を説明せよ．

第 5 節

1. $X_1, \ldots, X_n \sim Ex(\lambda)$ i.i.d. $(\lambda > 0)$ として，$\widehat{\lambda}$ を λ の MLE とする．このとき，$\sqrt{n}(\widehat{\lambda} - \lambda)$ の極限分布を求め，これを利用して λ に対する CI を構成せよ．

2. (続き)．$\sqrt{n}(g(\widehat{\lambda}) - g(\lambda)) \xrightarrow{d} N(0,1)$ となるような関数 g を 1 つ求めよ．この結果を利用して，λ にする CI を構成せよ．

3. r, s を正整数とし，$X \sim Be(r, s)$ とする．このとき，

$$Y = \frac{sX}{r(1-X)} \sim F(2r, 2s)$$

を示せ．

4. $X \mid \theta \sim Bin(n, \theta), \theta \sim Be(r, s)$ とする．ただし，r, s は正整数とする．このとき，θ の事後分布の $(1-\alpha)$ 分位点を F 分布の分位点を用いて表せ．

5. $Y \sim Mn(n, \theta^2, 2\theta(1-\theta), (1-\theta)^2), \theta \in (0,1)$ とする．

   (a) 最尤推定にもとづいて，θ に対する CI を求めよ．

   (b) r, s を正整数とし，$\theta \sim Be(r,s)$ という事前分布を入れる．このとき，θ に対する水準 $(1-\alpha)$ の信用区間を，F 分布の分位点を用いて表せ．

6. $\widehat{F}_n$ を $\mathbb{R}$ 上の確率的な d.f. とし，F を $\mathbb{R}$ 上の (確率的でない) d.f. であって，F は連続とする．このとき，次の 2 条件が同値であることを示せ．

   (1) 各 $x \in \mathbb{R}$ に対して，$\widehat{F}_n(x) \xrightarrow{P} F(x)$.

   (2) $\sup_{x\in\mathbb{R}} |\widehat{F}_n(x) - F(x)| \xrightarrow{P} 0$.

   ヒント：Pólya の定理の証明．

第 6 節

1. $U^n = (U_1^n, \ldots, U_n^n)'$ を $\mathbb{S}^{n-1}$ 上の一様分布に従う確率ベクトルとする．このとき，固定した正整数 k に対して，$n \to \infty$ のとき，

$$\sqrt{n}(U_1^n, \ldots, U_k^n)' \xrightarrow{d} N(0, I_k)$$

   となることを示せ．

2. F を $\mathbb{R}$ 上の d.f. とし，$X_1, \ldots, X_n \sim F$ i.i.d. とする．F が Poisson 分布に等しいかどうか検定したいとする：$H_0 : F = Po(\lambda)$ $(\exists\lambda > 0)$. Poisson 分布の平均と分散が等しいという事実から，

$$\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i,\ \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2$$

   とおいて，$|\overline{X}/\widehat{\sigma}^2 - 1|$ の値が十分大きいとき H_0 を棄却するという検定を考える．ここで，$F = Po(\lambda)$ $(\lambda > 0)$ のとき，

$$\sqrt{n}\left(\frac{\overline{X}}{\widehat{\sigma}^2} - 1\right)$$

   の極限分布を求めよ．

   ヒント：$Y_i = X_i^2$ とおいて，$\overline{X}/\widehat{\sigma}^2$ を $(\overline{X}, \overline{Y})$ の関数に表してデルタ法を適用する．

3. F を $\mathbb{R}$ 上の d.f. とし，密度関数 f をもつとする．また，f は $x_0 \in \mathbb{R}$ において正かつ連続と仮定する．$X_1, \ldots, X_n \sim F$ i.i.d. とし，$\widehat{F}_n(x) = n^{-1}\sum_{i=1}^n I(X_i \le x)$ を経験分布関数として，$b_n > 0, b_n \to 0$ に対して，

$$\widehat{f}_n(x_0) = \frac{\widehat{F}_n(x_0 + b_n) - \widehat{F}_n(x_0 - b_n)}{2b_n}$$

   とおく．$nb_n \to \infty$ を仮定する．

(a) $\widehat{f}_n(x_0) \xrightarrow{P} f(x_0)$ となることを示せ．ヒント：$E[\widehat{f}_n(x_0)] \to f(x_0), \mathrm{Var}(\widehat{f}_n(x_0)) \to 0$ を示せ．

(b) Berry-Esseen の定理を利用して，

$$\sqrt{nb_n}(\widehat{f}_n(x_0) - E[\widehat{f}_n(x_0)]) \xrightarrow{d} N(0, f(x_0)/2)$$

を示せ．

コメント：$\widehat{f}$ はカーネル密度推定量と呼ばれるものの一例である．

4. $Y \sim Mn(n, (2+\theta)/4, (1-\theta)/4, (1-\theta)/4, \theta/4), \theta \in (0,1)$ として，θ の MLE $\widehat{\theta}$ に対して，$\sqrt{n}(\widehat{\theta} - \theta)$ の極限分布を求めよ．

5. $\theta \in (0,1)$ に対して，

$$p_\theta(x) = \{\theta e^{-x} + (1-\theta)xe^{-x}\}I(x > 0)$$

とおいて，$X_1, \ldots, X_n \sim p_\theta$ i.i.d. とする．

(a) $\mu = E_\theta[X_1]$ とおいて，θ を μ の関数 $\theta = g(\mu)$ として表せ．

(b) $\widehat{\theta} = g(\overline{X})$ という推定量を考える．このとき，$\sqrt{n}(\widehat{\theta} - \theta)$ の極限分布を求めよ．

(c) p_θ の Fisher 情報量 $I(\theta)$ を求めよ．

(d) (b) の推定量 $\widehat{\theta}$ にワンステップ推定を適用して，極限分布の分散が $1/I(\theta)$ を達成するような推定量 $\check{\theta}$ を構成せよ．

6. $X_1, \ldots, X_n \sim Ex(1)$ i.i.d. とする．このとき，独立な $V \sim N(0,1), W \sim Ex(1)$ に対して，$\left(\sqrt{n}(\overline{X} - 1), nX_{(1)}\right) \xrightarrow{d} (V, W)$ を示せ．ヒント：$V_n = \sqrt{n}(\overline{X} - 1), W_n = nX_{(1)}$ とおく．Rényi 表現を使って，$E[e^{isV_n + itW_n}] = E[E[e^{isV_n} \mid X_{(1)}]e^{itW_n}]$ を評価する．

7. $\alpha, \beta > 0$ を未知として，$X_1, \ldots, X_n \sim Pa(\alpha, \beta)$ i.i.d. とする．ただし，$Pa(\alpha, \beta)$ の密度関数を

$$f(x; \alpha, \beta) = \alpha^\beta \beta x^{-\beta-1} I(x \geq \alpha)$$

と修正する．

(a) (α, β) の MLE $(\widehat{\alpha}, \widehat{\beta})$ を求めよ．

(b) $\left(n(\widehat{\alpha} - \alpha), \sqrt{n}(\widehat{\beta} - \beta)\right)$ の極限分布を求めよ．

8. $\mu \in \mathbb{R}$ に対して，$X_1, \ldots, X_n \sim N(\mu, 1)$ i.i.d. とする．$x \in \mathbb{R}$ を任意に固定して，$\theta = P_\mu(X_1 \leq x) = \Phi(x - \mu)$ の推定を考える．このとき，$\widehat{\theta} = \Phi(x - \overline{X}), \widetilde{\theta} = n^{-1}\sum_{i=1}^n I(X_i \leq x)$ とおいて，$\widehat{\theta}$ の $\widetilde{\theta}$ に対する漸近相対効率を計算せよ．

9. F を $\mathbb{R}$ 上の d.f. とし，$X_1, \ldots, X_n \sim F$ i.i.d. とする．$E[X_1^2] < \infty$ と仮定して，$\mu = E[X_1]$ とおく．このとき，

$$U_n = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} X_i X_j$$

に対して，$\sqrt{n}(U_n - \mu^2)$ の極限分布を求めよ．

参考文献

Bahadur, R.R. (1954). Sufficiency and statistical decision functions. *Ann. Math. Statist.* **25** 423-462.

Bickel, P.J. and Doksum, K.A. (2015). *Mathematical Statistics: Basic Ideas and Selected Topics, Volume I, Second Edition*. Chapman & Hall.

Boucheron, S., Lugosi, G., and Massart, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press.

Brown, L.D., Cai, T.T., and DasGupta, A. (2001). Interval estimation for a binomial proportion. **16** 101-133.

Chung, K.-L. (2001). *A Course in Probability Theory (3rd Edition)*. Academic Press.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer.

Cover, T.M. and Thomas, J.A. (2006). *Elements of Information Theory (2nd Edition)*. Wiley.

Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press.

Efron, B. (1979). Bootstrap: another look at the jackknife. *Ann. Statist.* **7** 1-26.

Ferguson, T.S. (1996). *A Course in Large Sample Theory*. Chapman & Hall.

Gamerman, D. and Lopes, H.F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference (2nd Edition)*. Chapman & Hall.

Hall, P. (1993). *The Bootstrap and Edgeworth Expansion*. Springer.

Hájek, J. (1970). A characterization of limiting distributions of regular estimates. Z. Warscheinlichkeistheorie Verw. Gabiete. **14** 323-330.

Hájek, J. (1972). Local asymptotic minimax and admissibility in estimation. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **1** 175-194.

Halmos, P.R. (1946). The theory of unbiased estimation. *Ann. Marh. Statist.* **11** 34-43.

Halmos, P.R. and Savage, L.J. (1949). Application of the Radon-Nikodym theorem to the theory of sufficient statistics. *Ann. Math. Statist.* **20** 225-241.

Heyde, C. (1963). On a property of the log normal distribution. *J. Roy. Stat. Soc. Ser. B* **25** 392-393.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Statist.* **19** 293-325.

Hoeffiding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13-30.

James, W. and Stein, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 361-379.

Knight, K. (2000). *Mathematical Statistics.* Chapman & Hall.

Kolmogorov, A.N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Springer-Verlag.

Kubokawa, T. (1991). An approach to improving the James-Stein estimator. *J. Multivariate Anal.* **36** 121-126.

Lancaster, T. (2000). The incidental parameter problem since 1948. *J. Econometrics* **95** 391-413.

Leadbetter, M.R., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes.* Springer.

Le Cam, L. (1972). Limits of experiments. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **1** 245-261.

Le Cam, L. (1986). Central limit theorem around 1935. *Statist. Science* **1** 78-96.

Le Cam, L. and Yang, G. (2000). *Asymptotics in Statistics (2nd Edition).* Springer.

Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation (2nd Edition).* Springer.

Lehmann, E.L. and Romano, J. (2005). *Testing Statistical Hypotheses (3rd edition).* Springer.

Lehmann, E.L. and Scheffé, H. (1950). Completeness, similar regions, and unbiased estimation: Part I. *Sankhya* **10** 305-340.

Newey, W.K. and McFadden, D.L. (1994). Large sample estimation and hypothesis testing. In: *Handbook of Econometrics Vol. IV* (ed. by R.F. Engle and D.L. McFadden).

Neyman, J. and Scott, E.L. (1948). Consistent estimation from partially consistent observations. *Econometrica* **16** 1-32.

Pratt, J. (1961). Length of confidence intervals. *J. Amer. Statist. Assoc.* **56** 549-567.

Reiss, R.-D. (1989). *Approximate Distributions of Order Statistics: With Applications to Nonparametric Statistics.* Springer.

Resnick, S. (1987). *Extreme Values, Regular Variations, and Point Processes.* Springer.

Resnick, S. (1998). *A Probability Path.* Birkhauser.

Robert, C.P. (2007). *Bayesian Choice (2nd Edition).* Springer.

Robert, C.P. and Casella, G. (2004). *Monte Carlo Statistical Methods.* Spinger.

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate distribution. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 197-206.

Stroock, D.W. (2011). *Probability Theory: An Analytic View (2nd Edition).* Cambridge University Press.

van der Vaart, A. (1998). *Asymptotic Statistics.* Cambridge University Press.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20** 595-601.

Wasserman, L. (2003). *All of Statistics: A Concise Course in Statistical Inference.* Springer.

Wasserman, L. (2006). *All of Nonparametric Statistics.* Springer.

Williams, D. (1991). *Probability with Martingales.* Cambridge University Press.

久保川達也. (2015).「現代数理統計学の基礎」近刊.

竹内啓. (1963).「数理統計学」東洋経済.

竹村彰通. (1991). 「現代数理統計学」創文社.

鍋谷清治. (1978).「数理統計学」共立出版.

舟木直久. (2004).「確率論」朝倉書店.

吉田朋広. (2006).「数理統計学」朝倉書店.

吉田伸生. (2006). 「ルベーグ積分入門」遊星社.