1 Introduction

1.1 Definition and properties of probability space

Definition 1.1 (Probability space)

$(\Omega, B, P)$ is called a probability space if the following conditions (i) and (ii) hold:

(i) $B$ is a $\sigma$-field of $\Omega$, that is, $B$ is a family of subsets of $\Omega$, and (B1) $\sim$ (B3) are satisfied.

(B1) $\Omega \in B$.

(B2) $A \in B \Rightarrow A^c \in B$.

(B3) $A_1, A_2, \ldots \in B \Rightarrow \bigcup_{i=1}^{\infty} A_i \in B$.

Ω and $(\Omega, B)$ are called a sample space and a measurable space, respectively. Elements of $B$ (subsets of $\Omega$) are called events.

(ii) $P$ is a probability measure on $(\Omega, B)$, that is, $P$ is a real valued function on $B$, and (P1 $\sim$ P3) holds.

(P1) $0 \leq P(A) \leq 1$ for any $A \in B$.

(P2) $P(\Omega) = 1$.

(P3) If $A_1, A_2, \ldots \in B$ and $A_i \cap A_j = \emptyset$ ($i \neq j$) $\Rightarrow$ $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$.

Theorem 1.1 (Properties of probability spaces)

$(\Omega, B, P)$ : a probability space

$\{A_n\}_{n=1,2,\ldots}$ : a sequence of events in $B$

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

2. $A_n \subset A_{n+1}$ ($n = 1, 2, \ldots$) $\Rightarrow P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n)$. 
(3) $A_n \supset A_{n+1}$ ($n = 1, 2, \ldots$) $\Rightarrow$ $P\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n \to \infty} P(A_n)$.

(4) $\overline{\lim}_{n \to \infty} A_n = \liminf_{n \to \infty} A_n$ $\Rightarrow$ $P(\lim_{n \to \infty} A_n) = \lim_{n \to \infty} P(A_n)$.

Problem 1.1
Prove Theorem 1.1.

Thoerem 1.2 (The first Borel–Cantelli lemma)
\{A_n\}_{n=1,2,\ldots} : a sequence of events.
\[\sum_{n=1}^{\infty} P(A_n) < \infty \Rightarrow P(\overline{\lim}_{n \to \infty} A_n) = 0 \quad (P(\lim_{n \to \infty} A_n^c) = 1).\]

Thoerem 1.3 (The second Borel–Cantelli lemma)
\{A_n\}_{n=1,2,\ldots} : a sequence of independent events.
\[\sum_{n=1}^{\infty} P(A_n) = \infty \Rightarrow P(\overline{\lim}_{n \to \infty} A_n) = 1 \quad (P(\lim_{n \to \infty} A_n^c) = 0).\]

Definition 1.2 (Borel field)
The minimum $\sigma$-field which includes all open subsets of $\mathbb{R}$ is called the (one dimensional) Borel field, which is denoted by $\mathcal{B}$. The measurable space $(\mathbb{R}, \mathcal{B})$ is called the Borel space.

Definition 1.3 (Random variable)
$(\Omega, \mathcal{B}, P)$ : a probability space
X : $\Omega \mapsto \mathbb{R}$
If $X$ is Borel measurable, $X$ is called a random variable on $(\Omega, \mathcal{B}, P)$, which means
\[\forall B \in \mathcal{B}, X^{-1}(B) := \{\omega \in \Omega \mid X(\omega) \in B\} \in \mathcal{B}.\]
Here $P_X : B \in \mathcal{B} \mapsto P(X^{-1}(B))$ becomes a probability measure on $(\mathbb{R}, \mathcal{B})$, which is called the distribution of $X$.

Definition 1.4 (Expectation)
X : a random variable
f : a Borel measurable function ($\forall B \in \mathcal{B}, f^{-1}(B) \in \mathcal{B}$)
\[E[f(X)] = \int f(x) dP_X(x)\] is called the expectation of $f(X)$.
\[E[X] : \text{the mean of } X, \quad \text{Var}[X] = E[(X - E[X])^2] : \text{the variance of } X.\]

Lemma 1.1 (Chebyshev’s inequality)
X : a random variable such that $E[X]$ exists.
\[\forall a > 0, P(|X - E[X]| \geq a) \leq \frac{\text{Var}[X]}{a^2}\]

Definition 1.5 (Convergence in probability)
\{X_n\}_{n=1,2,\ldots} : a sequence of random variables defined on $(\Omega, \mathcal{B}, P)$.
X : a random variable defined on $(\Omega, \mathcal{B}, P)$.
We say “$X_n$ converges to $X$ in probability”, and denote “$X_n \xrightarrow{P} X$ ($n \to \infty$)” if
\[\forall \delta > 0, \lim_{n \to \infty} P(|X_n - X| > \delta) = 0.\]
Example 1.1 (Weak law of large numbers)
\{X_n\}_{n=1,2,\ldots} : a sequence of independent random variables on (\Omega, \mathcal{B}, P)
E[X_i] = \mu, \text{Var}[X_i] = \sigma^2 (i = 1, 2, \ldots; \mu, \sigma^2 \in \mathbb{R})

\Rightarrow \bar{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i \overset{p}{\rightarrow} \mu (n \to \infty).

Problem 1.2
Prove Example 1.1 with using Chebyshev’s inequality.

Definition 1.6 (Almost sure convergence)
\{X_n\}_{n=1,2,\ldots} : a sequence of random variables defined on (\Omega, \mathcal{B}, P).
X : a random variable defined on (\Omega, \mathcal{B}, P).

We say “\(X_n\) converges to \(X\) almost surely”, and denote \(\overline{X}_n \overset{a.s.}{\rightarrow} X (n \to \infty)\)” if
\[ \exists \Omega_0 \in \Omega \ s.t. \ P(\Omega_0) = 1, \ and \ \forall \omega \in \Omega_0, \ \lim_{n \to \infty} X_n(\omega) = X(\omega) \]

1.2 Strong law of large numbers

Theorem 1.4 (Kolmogorov’s inequality)
\{X_n\}_{n=1,2,\ldots} : a sequence of independent random variables on (\Omega, \mathcal{B}, P).
\[ S_j = X_1 + \cdots + X_j.\]
E[\(X_j\)] = 0, \(j = 1, \ldots, n \Rightarrow \)
\[ P(\max_{i=1,\ldots,n} |S_i| \geq a) \leq \frac{E(S_n^2)}{a^2}, \quad a > 0. \tag{1.1} \]

Theorem 1.5 (Convergence of the partial sum)
\{X_n\}_{n=1,2,\ldots} : a sequence of independent random variables on (\Omega, \mathcal{B}, P).
\[ S_j = X_1 + \cdots + X_j.\]
E[\(X_n\)] = 0, \(n = 1, 2, \ldots\) and \[ \sum_{n=1}^{\infty} \text{Var}[X_n] < \infty \]

\Rightarrow \(S_n = \sum_{k=1}^{n} X_k\) converges to a certain random variable almost surely.

Lemma 1.2 (Kronecker’s lemma)
\{m_n\}_{n=1,2,\ldots} : a sequence of positive numbers such that \(m_n \leq m_{n+1}\) \((n = 1, 2, \ldots)\) and \[ \lim_{n \to \infty} m_n = \infty. \]
\{x_n\} : a sequence of real numbers.
\[ \lim_{n \to \infty} \sum_{k=1}^{n} \frac{x_k}{m_k} < \infty \Rightarrow \lim_{n \to \infty} \frac{1}{m_n} \sum_{k=1}^{n} x_k = 0. \]

Theorem 1.6
\{X_n\}_{n=1,2,\ldots} : a sequence of independent random variables on (\Omega, \mathcal{B}, P).
E[\(X_n\)] = 0 and \[ \sum_{n=1}^{\infty} \frac{E[X_n^2]}{n^2} < \infty \Rightarrow \frac{1}{n} \sum_{k=1}^{n} X_n \overset{a.s.}{\rightarrow} 0 (n \to \infty). \]
**Theorem 1.7 (Kolmogorov’s strong law)**

\( \{X_n\}_{n=1,2,...} \) : a sequence of independent random variables on \((\Omega, \mathcal{B}, P)\).

\[
E[X_n] = a, \ Var[X_n] \leq \nu, \ n = 1, 2, \ldots \Rightarrow \frac{1}{n} \sum_{k=1}^{n} X_n \overset{a.s.}{\rightarrow} a \ (n \to \infty).
\]

**Problem 1.3**

Prove Kolmogorov’s strong law. (You can use Theorem 1.6 without proof.)

**Theorem 1.8 (Khinchin)**

\( X_1, X_2, \ldots \) : independent identically distributed random variables on \((\Omega, \mathcal{B}, P)\)

\( p \) : real number such that \( 2 > p \geq 1 \)

\[
n^{-1/p} \sum_{k=1}^{n} (X_k - m) \overset{a.s.}{\rightarrow} 0 \iff E[|X_1|^p] < \infty \text{ and } m = E[X_1]
\]

**Lemma 1.3**

\( X_1, X_2, \ldots \) : independent identically distributed random variables on \((\Omega, \mathcal{B}, P)\)

\( p \) : real number such that \( 2 > p \geq 1 \)

Assume

\[
E[|X_1|^p] < \infty \text{ and } E[X_1] = 0.
\]

Define

\[
Y_n(\omega) = \begin{cases} 
X_n(\omega), & |X_n(\omega)| \leq n^{1/p} \\
0, & |X_n(\omega)| > n^{1/p}
\end{cases}
\]

Then

\[
\sum_{n=1}^{\infty} P(X_n \neq Y_n) < \infty, \quad (1.2)
\]

\[
\sum_{n=1}^{\infty} n^{-2/p} E[Y_n^2] < \infty, \quad (1.3)
\]

\[
\lim_{n \to \infty} n^{-1/p} \sum_{k=1}^{n} E[Y_k] = 0. \quad (1.4)
\]

**1.3 Central limit theorem**

**Definition 1.7 (Characteristic function)**

\( X \) : a random variable

\( \varphi_X(t) = E[e^{itX}] \) is called the *characteristic function* of \( X \).

**Theorem 1.9**

Let \( \varphi_X(t) \) be a characteristic function. Then we have

(1) \( \varphi_X(0) = 1 \).

(2) \( |\varphi_X(t)| \leq 1 \).

(3) \( \varphi_X(t) \) is uniformly continuous
(4) \( \varphi_{X+a}(t) = e^{itd}\varphi_X(ct) \). where \( c, d \) are constants.

(5) If \( \text{E}(|X|^n) < \infty \), then \( \varphi_X(t) \) is \( C^n \)-class and

\[
\left. \frac{d^n}{dt^n} \varphi_X(t) \right|_{t=0} = i^n\text{E}(X^n).
\]

**Problem 1.4**
Prove Theorem 1.9.

**Lemma 1.4**
(1) \( \gamma_y \geq 0 \),

\[
0 \leq (\text{sgn } \alpha) \int_0^y \frac{\sin \alpha x}{x} \, dx \leq \int_0^\pi \frac{\sin x}{x} \, dx.
\]

(2) \( \int_0^\infty \frac{\sin \alpha x}{x} \, dx = \frac{\pi}{2} \text{sgn } \alpha. \)

where

\[
\text{sgn } \alpha = \begin{cases} 
1, & \alpha > 0 \\
0, & \alpha = 0 \\
-1, & \alpha < 0.
\end{cases}
\]

**Theorem 1.10 (Inversion formulae)**

\( X \): a random variable
\( F_X \): the distribution function of \( X \)
\( \varphi_X \): the characteristic function of \( X \)

If \( F_X \) is continuous at \( a \) and \( b \) \((a < b)\), it holds that

\[
F_X(b) - F_X(a) = \lim_{T \to \infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \varphi_X(t) \, dt.
\]

**Theorem 1.11 (one-to-one correspondence)**

\( X_1, X_2 \): random variables
\( \mu_k \): the distribution of \( X_k \) \((k = 1, 2)\)
\( \varphi_k \): the characteristic function of \( X_k \) \((k = 1, 2)\)

Then it holds that

\( \varphi_1 = \varphi_2 \iff \mu_1 = \mu_2. \)

**Problem 1.5**

Cauchy distribution with the location parameter \( \mu \) and the scale parameter \( \sigma \) is denoted as \( Ca(\mu, \sigma) \), and defined by the following probability density function:

\[
f(x; \mu, \sigma) = \frac{\sigma}{\pi \{\sigma^2 + (x - \mu)^2\}}
\]

(1) Derive the characteristic function of \( Ca(0, 1) \). (Hint: residue theorem)

(2) Show that if \( X_1, \ldots, X_n \overset{i.i.d.}{\sim} Ca(\mu, 1), \bar{X}_n \sim Ca(\mu, 1). \)
**Definition 1.8** (分布収束)

\[ \{X_n\}_{n=1,2,\ldots} : \text{a sequence of random variables} \]

\[ X : \text{a random variables} \]

\[ F_n : \text{the distribution function of } X_n \ (n = 1, 2, \ldots) \]

\[ F : \text{the distribution function of } X \]

We say “\( X_n \) converges to \( X \) in distribution”, and denote “\( X_n \overset{d}{\to} X \ (n \to \infty) \)” if

\[ \lim_{n \to \infty} F_n(x) = x \]

for any continuous point \( x \) of \( F \).

**Lemma 1.5**

If \( X_n \overset{d}{\to} X \ (n \to \infty) \), \( \lim_{n \to \infty} E[f(X_n)] = E[f(X)] \) for any continuous and bounded function \( f \).

**Theorem 1.12 (Lévy’s continuity theorem)**

\( \varphi_n(t) : \text{the characteristic function of a random variable } X_n \ (n = 1, 2, \ldots) \)

If \( \varphi_n(t) \) converges to \( \varphi(t) \) at each \( t \) and \( \varphi(t) \) is continuous at \( t = 0 \), \( \varphi(t) \) is the characteristic function of certain distribution \( P_X \), and \( X_n \overset{d}{\to} P_X \ (n \to \infty) \).

**Lemma 1.6**

Under the assumptions of Theorem 1.12, for any positive number \( \varepsilon \), there exists \( A \) such that

\[ P(|X_n| \leq A) > 1 - \varepsilon, \quad n = 1, 2, \ldots \] (1.5)

**Lemma 1.7**

For any real number \( x \), it holds that

\[ \left| e^{ix} - \sum_{k=0}^{n-1} \frac{(ix)^k}{k!} \right| \leq \frac{|x|^n}{n!} \]

**Theorem 1.13 (Central limit theorem (Lindeberg))**

\( X_{n,1}, X_{n,2}, \ldots, X_{n,n} : \text{independent random variables for each } n \)

Assume \( E[X_{n,k}] = 0, \sum_{k=1}^{n} \text{Var}[X_{n,k}] = 1 \) and

\[ \forall \tau > 0, \quad \lim_{n \to \infty} \sum_{k=1}^{n} E[X_{n,k}^2 1_{|X_{k,n}| > \tau}] = 0, \] (1.6)

Then \( Z_n := \sum_{k=1}^{n} X_{n,k} \overset{d}{\to} N(0, 1) \).

**Theorem 1.14 (Central limit theorem)**

\( \{X_n\}_{n=1,2,\ldots} : \text{a sequence of independent and identically distributed random variables} \)

Suppose \( E(X_j) = \mu, \text{Var}(X_j) = \sigma^2, \quad j = 1, 2, \ldots. \)

Then

\[ Z_n = \frac{1}{\sqrt{n} \sigma} \sum_{j=1}^{n} (X_j - \mu) = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \overset{d}{\to} N(0, 1). \]

**Problem 1.6**

Prove Theorem 1.14.
2 最尤推定量の分布

Definition 2.1 (標本と母集団分布)
確率変数 $X_1, \ldots, X_n$ が互いに独立で、同一の確率分布 $P$ に従うとき、$X_1, \ldots, X_n$ は、分布 $P$ から、大きさ $n$ 無作為標本といい、$P$ を母集団分布という。

Definition 2.2 (統計モデル)
母集団分布が、ある確率分布の集まり

$$P = \{P_\theta; \theta \in \Theta\}, \Theta \subset \mathbb{R}^p \ (p \text{ ある自然数})$$

に属すると想定されるとき、$P$ を母集団分布の統計モデルという。このとき、$\theta$ を母数、$\Theta$ を母数空間と呼ぶ。

Example 2.1
小学校6年男児 $n$ 人の身長 $X_1, \ldots, X_n \ (\text{cm})$ を測定するとする。このとき、$X_1, \ldots, X_n$ の母集団分布は正規分布 $N(\mu, \sigma^2)$ からの無作為標本と考えられる。このとき、母数は、$\theta = (\mu, \sigma^2)$、母数空間は $\Theta = \mathbb{R} \times (0, \infty)$ となる。平均 $\mu$、分散 $\sigma^2$ の値がわかれば、例えば、小学6年男児の内、身長が130cm以下の人数の比率は

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{130} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx$$

によって近似することができる。

Definition 2.3 (カルバック-ライブラーの擬距離)
$X$ の確率密度関数を $f(x)$ とする。確率密度関数 $g(x)$ に対して

$$\text{KL}(f; g) = E\left[ \frac{f(X)}{g(X)} \right]$$

をカルバック-ライブラーの擬距離と呼ぶ。ただし、$P(f(X) > 0, g(X) = 0) > 0$ のときは、$\text{KL}(f; g) = \infty$ と定義する。

Lemma 2.1
任意の $g$ に対して、$\text{KL}(f; g) \geq 0$ が成り立ち、等号は確率1で $f(X) = g(X)$ のときに限る。

$$X_1, \ldots, X_n \overset{i.i.d.}{\sim} f(x; \theta_0), \theta_0 \in \Theta \text{ とし、次の仮定 (A0) が成り立つとする。}

(A0) $\theta \neq \theta'$ ならば $P(f(X; \theta) \neq f(X; \theta')) > 0$。

大数の法則から

$$\frac{1}{n} \sum_{i=1}^{n} \log \frac{f(X_i; \theta)}{f(X_i; \theta_0)} \xrightarrow{a.s.} -\text{KL}(f(\cdot; \theta_0); f(\cdot; \theta)) \ (n \to \infty)$$

である。仮定 (A0) が成り立つとき、右辺は $\theta = \theta_0$ のときのみ最大値をとることから、左辺、すなわち $\prod_{i=1}^{n} f(X_i, \theta)$ を最大とする $\theta$ を $\hat{\theta}_n = \hat{\theta}_n(X)$ とするとき、$\hat{\theta}_n$ は、$n \to \infty$ のとき $\theta_0$ に収束することが期待される。
標本変量 \( \mathbf{X} = (X_1, \ldots, X_n)' \) に対する統計モデルを

\[
P(\mathbf{X} \in A) = \begin{cases} 
\sum_{x \in A} f(x; \theta), & \mathbf{X} \text{が離散型の場合} \\
\int_{x \in A} f(x; \theta) dx, & \mathbf{X} \text{が連続型の場合}
\end{cases}
\]

とする。ここで, \( \theta \in \Theta \)。確率密度関数 \( f(x; \theta) \) を \( x \) を固定して \( \theta \) の関数とみなすとき

\[
L(\theta; x) \quad (= f(x; \theta))
\]

(2.7)

と表し, 尤度関数とよぶ。

Definition 2.4
尤度関数 \( L(\theta; x) \) の最大を実現する \( \theta \) を \( \hat{\theta} = \hat{\theta}(x) \) と表し, \( \theta \) の最尤推定値という。すなわち

\[
L(\hat{\theta}; x) = \sup_{\theta \in \Theta} L(\theta; x).
\]

また, \( \hat{\theta}(X) \) を \( \theta \) の最尤推定値という。

対数尤度 \( \log L(\theta; x) \) を \( \ell(\theta; x) \) と表す。多くの場合, 尤度推定値は尤度方程式

\[
\frac{\partial}{\partial \theta} \ell(\theta; x) = 0
\]

の解として与えられる。

Problem 2.1
(1) \( X_1, \ldots, X_n \overset{i.i.d.}{\sim} B(1, \theta), \theta \in (0, 1) \) すなわち \( P(X_i = 1) = 1 - P(X_i) = \theta, \, i = 1, \ldots, n \) とする。\( \theta \) の最尤推定値を求めよ。

(2) \( X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2), \theta = (\mu, \sigma^2) \) とする。\( \mu, \sigma^2 \) の最尤推定値を求めよ。

(3) \( X_1, \ldots, X_n \overset{i.i.d.}{\sim} U(0, \theta) \) (区間 \( (0, \theta) \) 上の一様分布) とする。\( \theta \) の最尤推定値を求めよ。

Lemma 2.2
\( \{X_n \} \) を \( (\Omega, B, P) \) 上の確率変数列とするとき, \( X_n \overset{a.s.}{\Rightarrow} X \) であるための必要十分条件は

\[
A_n(\varepsilon) = \{ \omega; |X_n - X| \leq \varepsilon \}
\]

とするとき,

\[
\forall \varepsilon > 0, \quad P(\lim_{n \to \infty} A_n(\varepsilon)) = 1
\]

である。

Theroem 2.1 (最尤推定値の一致性)
\( X_1, \ldots, X_n \overset{i.i.d.}{\sim} P_{\theta_0}, P_{\theta} \) は確率密度関数, あるいは確率関数 \( f(x; \theta), \theta \in \Theta \subset \mathbb{R} \) を持つ。次の (A1, A2) を仮定する。

(A1) 任意の \( \theta_*(\theta_* \neq \theta_0) \) に対して正数 \( \delta = \delta(\theta_*) \) が存在して

\[
\eta(\theta_*) := E\left[ \sup_{\theta \in U_{\theta_*, \delta}} \log \frac{f(X_1; \theta)}{f(X_1; \theta_0)} \right] < 0.
\]

ただし, \( U_{\theta_*, \delta} = \{ \theta \in \Theta; \|\theta - \theta_*\| < \delta \} \).
(A2) Θのコンパクト部分集合 \( K \) が存在して
\[
\eta_K := E \left[ \sup_{\theta \in K} \log \frac{f(X_1; \theta)}{f(X_1; \theta_0)} \right] < 0.
\]
このとき，各 \( n \) に対して確率 1 で尤度推定量 \( \hat{\theta}_n \) が存在するならば \( \hat{\theta}_n \overset{a.s.}{\to} \theta_0 \) \((n \to \infty)\).

Theorem 2.2 (漸近正規性；クラメルの十分条件)
\( X_1, \ldots, X_n \) を確率密度関数 \( f(x; \theta_0) \) を持つ連続型分布からのランダム標本とする。ただし，\( \theta_0 \in \Theta \subseteq \mathbb{R} \)，\( \Theta \) は開集合である。また，\( \theta_0 \) の近傍 \( U \)，関数 \( F_1(x) \)，\( F_2(x) \)，\( H(x) \) と定数 \( M \) が存在して次の条件が成り立つと仮定する。

(A1) \( f(x; \theta) \) を \( \theta \) に関して 3 回微分可能である。

(A2) \( \theta \in U \) で，\( \left| \frac{\partial f(x; \theta)}{\partial \theta} \right| < F_1(x) \)，\( \left| \frac{\partial^2 f(x; \theta)}{\partial \theta^2} \right| < F_2(x) \) かつ，\( \left| \frac{\partial^3 \log f(x; \theta)}{\partial \theta^3} \right| < H(x) \) である。また，\( F_1(x) \)，\( F_2(x) \) は可積分で，\( H \)，\( M \) は
\[
E_{\theta_0} \{ H(X) \} \leq M
\]
を満たす。

(A3) フィッシャー情報量 \( J(\theta_0) \) は有限かつ正である。
このとき，任意の正数 \( \varepsilon \) に対して \( N \) が存在して，\( n \geq N \) ならば \( 1 - \varepsilon \) より大きい確率で，尤度方程式 \( \partial l(\theta, X) / \partial \theta = 0 \) は，\( (\theta_0 - \delta, \theta_0 + \delta) \) 内に解を持つ。この解で \( \hat{\theta}_n \) とすると
\[
\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{d}{\to} N(0, J(\theta_0)^{-1}) \quad (2.8)
\]
が成り立つ。

Example 2.2
\( X_1, \ldots, X_n \overset{i.i.d.}{\sim} Ca(\theta_0, 1) \) とする。以下，\( \theta_0 = 0 \) とする。
\[
\log \frac{f(x; \theta)}{f(x; 0)} = \log \frac{1 + x^2}{1 + (x - \theta)^2}
\]

\( 0 < \delta_0 < \theta_* \) とする。\( |\theta - \theta_*| < \delta \) のとき
\[
0 \leq (x - \theta)^2 \leq \max\{(\theta_* + \delta - x)^2, (x - \theta_* + \delta)^2\}
\]
であるから
\[
\sup_{\theta \in U_{\delta_* + \delta_0}} \log \frac{1 + x^2}{1 + (x - \theta)^2} \\
\leq \log(1 + x^2) + \log(1 + (x - \theta_* + \delta)^2) \\
+ \log(1 + (x - \theta_* - \delta)^2), \\
E_{\theta_0} \{ \log(1 + (X_1 - \alpha)^2) \} < \infty
\]
であるからルベーグの収束定理より
\[
\lim_{\delta \to 0} E_{\theta_0} \left[ \sup_{\theta \in U_{\delta_* + \delta}} \log \frac{f(X; \theta)}{f(X; \theta_0)} \right] = E_{\theta_0} \left[ \log \frac{f(X; \theta_*)}{f(X; \theta_0)} \right]
\]
また、補題2.1より、\( \theta_s \neq \theta_0 \) ならば \( E_{\theta_0} [ \log \frac{f(x; \theta_s)}{f(x; \theta_0)} ] < 0 \)。したがって、定理2.1の(A1)を満たす。

\[
\sup_{|\theta| > k} \log \frac{f(x; \theta)}{f(x; \theta_0)} = \begin{cases} 
\log(1 + x^2), & |x| > k \\
\log \frac{1 + x^2}{1 + (x - k)^2}, & 0 \leq x \leq k \\
\log \frac{1 + x^2}{1 + (x + k)^2}, & 0 > x \geq -k 
\end{cases}
\]

\[
E[\sup_{|\theta| > k} \log \frac{f(X_1; \theta)}{f(X_1; \theta_0)}] \\
= E[\log(1 + X_1^2)] - E[1_{0 \leq X_1 \leq k} \log \{1 + (k - X_1)^2\}] \\
- E[1_{0 > X_1 > -k} \log \{1 + (k + X_1)^2\}] \\
= E[\log(1 + X_1^2)] \\
- 2E[1_{0 \leq X_1 \leq k} \log \{1 + (k - X_1)^2\}] \\
\geq 1 \text{ とす}
\]

であるから定理2.1の(A2)が成り立つ。
対数尤度関数は

\[
l(\theta, X) = - \sum_{i=1}^{n} \log \{1 + (X_i - \theta)^2\} - n \log \pi
\]

であり、\( \lim_{\theta \to \pm \infty} l(\theta, X) = -\infty, l(\theta, X) \leq -n \log \pi \)であるから、確率1で尤度推定量 \( \hat{\theta}_n \) は存在し、定理2.1より \( \hat{\theta}_n \overset{a.s.}{\to} \theta_0 \) である。

**Example 2.3**

\( X_1, \ldots, X_n \overset{i.i.d.}{\sim} Ca(\theta_0, 1), \theta_0 = 0 \) とする。

\[
\frac{\partial f(x; \theta)}{\partial \theta} = \frac{2(x - \theta)}{\pi \{1 + (x - \theta)^2\}^2}, \\
\frac{\partial^2 f(x; \theta)}{\partial \theta^2} = \frac{-8}{\pi \{1 + (x - \theta)^2\}^3} + \frac{6}{\pi \{1 + (x - \theta)^2\}^2}, \\
\frac{\partial \log f(x; \theta)}{\partial \theta} = \frac{2(x - \theta)}{\pi \{1 + (x - \theta)^2\}}, \\
\frac{\partial^3 \log f(x; \theta)}{\partial \theta^3} = \frac{-16(x - \theta)}{\pi \{1 + (x - \theta)^2\}^3} + \frac{4(x - \theta)}{\pi \{1 + (x - \theta)^2\}^2}
\]
\[ \sup_{|\theta| < \delta} \left| \frac{x - \theta}{1 + (x - \theta)^2} \right| \leq \sup_{|\theta| < \delta} \frac{1}{1 + (x - \theta)^2}^{k-1/2} \]

\[ H_k(x; \delta) := \begin{cases} 
1, & |x| < \delta \\
\frac{1}{(1 + (x - \delta)^2)^{k-1/2}}, & x > \delta \\
\frac{1}{(1 + (x + \delta)^2)^{k-1/2}}, & x < -\delta 
\end{cases} \]

\[ \int_{-\infty}^{\infty} H_k(x; \delta) dx = 2\delta + 2 \int_{0}^{\infty} \frac{dx}{(1 + x^2)^{k-1/2}} \]

よ り \( k > 1 \) ならば \( \int_{-\infty}^{\infty} H_k(x; \delta) dx < \infty \). \( \frac{\partial^2 \log f(x; \theta)}{\partial \theta^2} \) は有界であるから, 定理 2.2 の (A1), (A2) が成り立つ. また, \( J(\theta_0) = \frac{1}{2} \) となるので (A3) も成り立つ.

尤度方程式の解は一意ではないが, 例 2.2 で, 最尤推定量の収束を示しており, 最尤推定量也尤度方程式の解となっているので, 定理 2.2 で存在が保証されている解が最尤推定量であり, したがって最尤推定量の漸近正規性が示される.
3 検定計量の分布

3.1 仮説検定の基礎

\[ X = (X_1, \ldots, X_n)', \quad X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} P_\theta \]

\[ P_\theta \in \mathcal{P} = \{P_\theta; \theta \in \Theta\}, \Theta \subset \mathbb{R}^n \]

とする。\( X \) の実現値 \( x = (x_1, \ldots, x_n)' \) から、仮説

\[ H_0 : \theta \in \Theta_0 (\Theta_0 \subset \Theta) \]

が正しいかどうか判定する問題を仮説検定問題と呼ぶ。\( H_0 \) を帰無仮説、また、

\[ H_1 : \theta \in \Theta \setminus \Theta_0 \]

を対立仮説と呼ぶ。

仮説検定法は、\( X \) の値域 (\( \mathbb{R}^n \)) の部分集合 \( \mathcal{W} \) を指定することによって与えられる。実際、\( \mathcal{W} \) に対して

\[ x \in \mathcal{W} \Rightarrow H_0 \] を棄却する

\[ x \notin \mathcal{W} \Rightarrow H_0 \] を棄却しない

とすれば良い。このとき、\( \mathcal{W} \) を、これによって与えられる検定の棄却域と呼ぶ。

\( H_0 \) が正しいのに \( H_0 \) を棄却する誤りを、第 1 種の過誤、\( H_1 \) が正しいのに \( H_0 \) を棄却しない誤りを、第 2 種の過誤と呼ぶ。

\[ \beta(\theta; \mathcal{W}) = P_\theta (X \in \mathcal{W}) \]

とすると、第 1 種、第 2 種の過誤の起こる確率は、それぞれ

\[ \beta(\theta; \mathcal{W}) \quad (\theta \in \Theta_0), \]

\[ 1 - \beta(\theta; \mathcal{W}) \quad (\theta \notin \Theta_0) \]

と表される。\( \beta(\theta; \mathcal{W}) (\theta \notin \Theta_0) \) を検出力と呼ぶ。実用上は第 1 種の過誤は、それによって被る被害が大きい場合が多いので、通常は、第 1 種の過誤確率の上限として 0.05 や 0.01 など小さな値を設定する。\( \sup_{\theta \in \Theta_0} \beta(\theta; \mathcal{W}) \leq \alpha \) となるような検定を有意水準 \( \alpha \) の検定と呼ぶ。

最適化問題 0 < \alpha < 1 とし,

\[ \sup_{\theta \in \Theta_0} \beta(\theta; \mathcal{W}) \leq \alpha \]

の条件下で、\( \theta \in \Theta \setminus \Theta_0 \) に対して、検出力 \( \beta(\theta; \mathcal{W}) \) を最大とする \( \mathcal{W} \) を求める。そのような検定を最強力検定と呼ぶ。最強力検定は一般に \( \theta \notin \Theta_0 \) に依存するが、任意の \( \theta \notin \Theta_0 \) に対して検出力を最大にする検定があれば、それを一样最強力検定と呼ぶ。
単純仮説の検定 \( \Theta_0 = \{ \theta_0 \} \) のとき, \( H_0 \) を単純仮説と呼ぶ。
\( \Theta = \{ \theta_0, \theta_1 \} (\theta_0 \neq \theta_1), \Theta_0 = \{ \theta_0 \} \) とする. このとき, 帰無仮説, 対立仮説はともに単純仮説で,

\[
H_0 : \theta = \theta_0, \quad H_1 : \theta = \theta_1
\] となる.

**Theorem 3.1** (ネイマン-ピアソンの基本定理)
\( X \) の確率密度関数を \( f(x; \theta) \) とし, (3.9) に対する有意水準 \( \alpha \) の検定問題を考える.

\[
W_c = \{ x : f(x; \theta_1) \geq c f(x; \theta_0) \}
\]

とし,

\[
P_{\theta_0}(X \in W_c) = \alpha
\]

を満たす \( c \) が存在すると仮定する. このとき, \( W_c \) を棄却域とする検定は最強力検定である.

対立仮説が複合仮説であるとき, 各 \( \theta \in \Theta \setminus \Theta_0 \) に対して, 定理 3.1 により, 最強力検定が決定されるが, 一般に, 最強力検定は \( \theta \) ごとに異なる. 最強力検定が \( \theta \) に依存しないとき, その検定は, 一様最強力検定となる.

**Example 3.1** (正規母集合の母平均に関する検定)
\( X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, 1) \) とし

\[
H_0 : \mu = 0, \quad H_1 : \mu > 0
\]

の検定を考える.

### 3.2 \( \theta = \theta_0 \) の検定

\( X_1, \ldots, X_n \overset{i.i.d.}{\sim} P_\theta, P_\theta \) は確率密度関数 \( f(x; \theta), \theta \in \Theta \subset \mathbb{R}^p \) を持つとし

帰無仮説 \( H_0 : \theta = \theta_0 \), 対立仮説 \( H_1 : \theta \neq \theta_0 \)

の検定問題を考える.

本節では, 第 2 節で与えた定理のように, 最尤推定量 \( \hat{\theta}_n \) が存在して, 尤度方程式

\[
s_n(\theta) = \sum_{i=1}^{n} \log \frac{\partial f(X_i; \theta)}{\partial \theta} = 0 \tag{3.10}
\]

の解であるとする. このとき, 尤度比基準は

\[
\lambda = \frac{\prod_{i=1}^{n} f(X_i; \theta_0)}{\prod_{i=1}^{n} f(X_i; \hat{\theta}_n)}
\]

となり, 適当な正則条件の下で

\[
-2 \log \lambda \overset{d}{\to} \chi^2_p \quad n \to \infty
\]

13
ワルド検定統計量は

\[ T_W = n(\hat{\theta}_n - \theta_0)'J(\hat{\theta}_n)(\hat{\theta}_n - \theta_0) \]

と定義される。適当な正則条件の下で

\[ T_W \xrightarrow{d} \chi^2_p \quad n \to \infty \]

スコア検定統計量は

\[ T_S = \frac{1}{n} s_n(\theta_0)'J(\theta_0)^{-1}s_n(\theta_0) \]

と定義される。適当な正則条件の下で

\[ T_S \xrightarrow{d} \chi^2_p \quad n \to \infty \]

3.3 関数構造の検定

\[ X_1, \ldots, X_n \overset{i.i.d.}\sim P_\theta, P_\theta \] は確率密度関数 \( f(x; \theta), \theta \in \Theta \subset \mathbb{R}^p \) を持つとする。\( \Theta_0 = \{\theta(\xi); \xi \in \Xi \subset \mathbb{R}^q\}, q < p \) とするとき

帰無仮説 \( H_0: \theta \in \Theta_0 \), 対立仮説 \( H_1: \theta \notin \Theta_0 \)

の検定問題を考える。ここで、\( \theta(\xi) \) は \( \Xi \) 上で定義された \( \Theta \) 内に値をとる \( C^3 \)−級のベクトル値関数とする。\( \xi \) は未知であるとする。

帰無仮説が正しく、真のパラメータが \( \theta_0 = \theta(\xi_0) \) であると仮定する。また、\( \Theta \) 全体での尤度推定量を \( \theta_n \), 帰無仮説の下での \( \xi \) の尤度推定量を \( \xi_n \) とする。このとき、尤度比基準は

\[ \lambda = \frac{\prod_{i=1}^n f(X_i; \theta(\xi_n))}{\prod_{i=1}^n f(X_i; \theta_n)} \]

となる。

ワルド検定統計量、スコア検定統計量はそれぞれ、

\[ T_W = n\{\theta_n - \theta(\xi_n)\}'J(\theta(\xi_n))\{\hat{\theta}_n - \theta(\xi_n)\}, \]

\[ T_S = \frac{1}{n} s_n(\theta(\xi_n))'J(\theta(\xi_n))^{-1}s_n(\theta(\xi_n)) \]

と定義される。このとき、適当な正則条件の下で

\[ -2 \log \lambda \sim T_W \sim T_S \xrightarrow{d} \chi^2_{p-q} \]