

# 目次

第 1 章	線型推測論	4
1.1	射影行列と逆行列	4
1.1.1	射影	4
1.1.2	一般化逆行列	4
1.1.3	ブロック行列の逆	5
1.2	カイ 2 乗分布	5
1.3	フィッシャー・コ克蘭の定理	6
1.3.1	独立性	7
1.4	$t$ 分布と $F$ 分布	7
1.4.1	$t$ 分布	7
1.4.2	$F$ 分布	7
1.5	ガウス・マルコフモデル	7
1.5.1	正規方程式	8
1.5.2	最良線型不偏推定量	8
1.5.3	分散共分散行列	9
1.5.4	分散の推定	9
1.6	仮説検定	9
1.6.1	理論	9
1.6.2	応用	10
1.7	平均の検定	10
1.8	重回帰分析	10
1.9	一元配置	10
1.10	二元配置	10
第 2 章	統計的決定理論	11
2.1	統計推論の枠組み	11
2.1.1	Wald の定義	11
2.1.2	危険関数	12
2.2	統計量の十分性と完備性	12
2.2.1	十分統計量	12
2.2.2	因子分解定理	12
2.2.3	Rao-Blackwell の定理	12
2.2.4	完備性	12
2.3	指数型分布族	13
2.4	統計的推定	13
2.4.1	不偏推定	13
2.4.2	最適性・有効性	13
2.4.3	Cramer-Rao の不等式	13
2.4.4	ベイズ推定	14

2.4.5	非許容性	14
2.5	統計的仮説検定	14
2.5.1	仮説検定の枠組み	14
2.5.2	ランダム化検定	15
2.5.3	形式化	15
2.5.4	Neyman-Pearson の補題	16
2.5.5	単調尤度比と複合仮説の検定	16
2.5.6	一般化された Neyman-Pearson の補題	16
2.5.7	不偏検定	16
2.5.8	両側 $t$ -検定	16
2.5.9	不変検定	16
2.6	区間推定	16
第 3 章	大標本理論	17
3.1	最尤推定	17
3.1.1	MLE	17
3.1.2	一致性	18
3.2	一様対数の法則	18
3.3	最小コントラスト推定	19
3.3.1	定理と応用	19
3.3.2	一致性を持つ推定量の例	20
3.3.3	独立観測の構造を仮定した場合	20
3.3.4	可測な最小コントラスト推定量の存在	20
3.4	$M$ -推定量	20
3.5	推定量の漸近正規性	21
3.5.1	一致性を持つ $M$ -推定量の漸近分布	21
3.5.2	最尤推定量での例	21
3.5.3	一般化モーメント法での例	21
3.6	LeCam のワンステップ推定量	22
3.7	$M$ -推定量の存在	22
3.8	頑健推定	22
3.8.1	影響関数	22
3.8.2	バイアスロバスト推定量	23
3.8.3	Fisher の一致性	23
3.9	尤度比検定	23
3.9.1	尤度比検定	23
3.9.2	棄却域の定め方	24
3.9.3	複合仮説の検定	24
3.9.4	Rao 検定	24
3.9.5	Wald 検定	24
3.10	多項分布の検定	24
3.11	尤度比確率場の局所漸近構造	25
3.11.1	漸近決定理論	25
3.11.2	尤度比確率場の接触構造	25
3.11.3	統計的実験	25
3.11.4	局所漸近正規性	26
3.11.5		26
3.12	情報量規準	27

3.12.1	枠組みと KL 距離 . . . . .	27
3.12.2	バイアス補正 . . . . .	27
3.13	密度推定 . . . . .	28
3.14	$U$ -統計量 . . . . .	28
第 4 章	漸近展開とその応用 . . . . .	29
4.1	漸近展開 . . . . .	29
4.2	平滑化補題 . . . . .	29
4.3	特性関数の展開 . . . . .	30
4.4	漸近展開の正当性の証明 . . . . .	30
4.5	漸近展開の変換 . . . . .	30
4.6	最尤推定量の漸近展開 . . . . .	30
4.7	漸近展開と情報幾何 . . . . .	30
4.8	ブートストラップ法 . . . . .	30
参考文献		31

# 第 1 章

## 線型推測論

線型推測論が統計モデルの線形代数である。

Anderson による標準的教科書 "An Introduction to Multivariate Statistical Analysis" (1958) が多変量解析の、George Box と Jenkins による "Time Series Analysis, Forecasting and Control" (1976) が時系列解析の最初の金字塔であるが、これらはいずれも線形モデルを扱っている。大規模計算機時代を迎えて以降、非線形モデルが発展の中心となったが、応用における一次近似と、その後の非線形解析の方向性を見るための試金石としての意味で、線形モデルの意義は不動である。例えば、数学的に整理された解析力学を差し置いて、工学の場面では Newton 的力学がもっとも重要になることに等しい。実際、 $X$  のデザインを非線形にすることが可能であるから、線形性の仮定は応用上は強い制約ではない。

また、正規近似の手法で用いられる、正規分布から派生する確率分布 (カイ 2 乗分布、 $t$  分布、 $F$  分布) についてまとめる。

### 1.1 射影行列と逆行列

#### 1.1.1 射影

Euclid 空間  $\mathbb{R}^n$  では、任意の自己作用素は有界である。よって、冪等律を満たす行列を射影行列といい、さらに自己共役であるものを直交射影という。

補題 1.1.1 (射影の特徴付け)。線型写像  $P: \mathbb{R}^n \rightarrow \mathbb{R}^n$  について、

- (1)  $P \circ P = P$  .
- (2)  $\mathbb{R}^n = \text{Im } P \oplus \text{Ker } P$  と表せる .

また、次の 2 条件とも同値 .

- (3)  $Q \in \text{GL}_n(\mathbb{R})$  が存在して、 $Q^{-1}PQ = \text{diag}[1, \dots, 1, 0, \dots, 0]$  .
- (4)  $\text{rank } P + \text{rank}(I_n - P) = n$  .

なお、このとき  $\text{rank } P = \text{Tr } P = r$  である。

補題 1.1.2. 射影  $P \in M_n(\mathbb{R})$  について、

- (1)  $P$  は直交射影である .
- (2)  $P^\top = P$  .

#### 1.1.2 一般化逆行列

記法 1.1.3.

- (1) 行列  $A \in M_{mn}(\mathbb{R})$  の列ベクトルではられる線型空間を  $L[A] \subset \mathbb{R}^n$  と表す .
- (2)  $X \in M_{np}(\mathbb{R})$  に対して、 $L[X]$  が定める直交射影を  $P_X$  で表す .

(3)  $H \in M_{pq}(\mathbb{R})$  は  $L[H] \subset L[X^\top]$  かつ  $\text{rank} H = q$  を満たすとする .

$$L[X|H] := X((\text{Im } H)^\perp) = \{X\gamma \in \mathbb{R}^n \mid \gamma \in \text{Ker } H^\top = (\text{Im } H)^\perp\}$$

への直交射影を  $P_{X|H}$  で表す .  $r := \text{rank} X$  とする . なお ,  $\forall T \in B(H) \text{ Ker } T^* = (\text{Im } T)^\perp$  なので ,  $\text{Ker } H^\top = (\text{Im } H)^\perp$  .

定義 1.1.4 (generalized inverse matrix). 行列  $A \in M_{mn}(\mathbb{R})$  に対して ,  $AA^-A = A$  を満足する  $A^- \in M_{nm}(\mathbb{R})$  を ,  $A$  の一般化逆行列という .

補題 1.1.5. 行列  $A \in M_{mn}(\mathbb{R})$  について ,

- (1) 一般逆  $A^-$  は存在する .
- (2) 一意とは限らない .

[ 証明 ] .

- (1) ある  $B \in \text{GL}_m(\mathbb{R}), C \in \text{GL}_n(\mathbb{R})$  について ,  $D := BAC = \text{diag}[1, \dots, 1, 0, \dots, 0]$  .  $A^- := CD^\top B$  とおくと ,  $A = B^{-1}DC^{-1}$  より ,  $AA^-A = ACD^\top BA = B^{-1}DC^{-1}CD^\top BB^{-1}DC^{-1} = B^{-1}DD^\top DC^{-1} = B^{-1}DC^{-1} = A$  .
- (2)  $A := (1 \ 0; 0 \ 0)$  について ,  $A$  自身も ,  $B := (1 \ 1; 0 \ 0)$  も一般逆である .

■

定理 1.1.6.  $X \in M_{np}(\mathbb{R})$  とする .  $X^\top X$  の任意の一般化逆行列  $(X^\top X)^-$  に対して ,  $P_X = X(X^\top X)^-X^\top$  .

定理 1.1.7.  $X \in M_{np}(\mathbb{R})$  とする .  $X^\top X$  の任意の一般化逆行列  $(X^\top X)^-$  に対して ,  $G := (X^\top X)^-H$  と定める . このとき ,  $G^\top X^\top XG$  は正則であり ,

$$P_{X|H} = P_X - XG(G^\top X^\top XG)^{-1}G^\top X^\top .$$

また ,  $\text{rank} P_{X|H} = r - q$  である .

### 1.1.3 ブロック行列の逆

記法 1.1.8.  $A \in M_p(\mathbb{R}), B \in M_{pq}(\mathbb{R}), C \in M_{qp}(\mathbb{R}), D \in M_q(\mathbb{R})$  について ,  $M := [A \ B; C \ D] \in M_{p+q}(\mathbb{R})$  とする .  $D \in \text{GL}_q(\mathbb{R})$  のとき  $F := A - BD^{-1}C$  ,  $A \in \text{GL}_p(\mathbb{R})$  のとき  $G := D - CA^{-1}B$  とおく .

命題 1.1.9. (1)  $D$  が正則であるとき ,  $|M| = |F||D|$  . 特に ,  $D, M$  が正則であることと ,  $D, F$  が正則であることは同値 .

(2)  $D, M$  が正則であるとき ,

$$M^{-1} = \begin{bmatrix} F^{-1} & -F^{-1}BD^{-1} \\ -D^{-1}CF^{-1} & D^{-1} + D^{-1}CF^{-1}BD^{-1} \end{bmatrix} .$$

(3)  $A$  が正則であるとき ,  $|M| = |A||G|$  . 特に ,  $A, M$  が正則であることと ,  $A, G$  が正則であることは同値 .

(4)  $A, M$  が正則であるとき ,

$$M^{-1} = \begin{bmatrix} A^{-1} + A^{-1}BG^{-1}CA^{-1} & -A^{-1}BG^{-1} \\ -G^{-1}CA^{-1} & G^{-1} \end{bmatrix} .$$

## 1.2 カイ 2 乗分布

記法 1.2.1. Gamma 分布  $G(\alpha, \nu) : \mathbb{R}_{>0}^2 \rightarrow P(\mathbb{R}, \mathcal{B}_1)$  の確率密度関数は

$$g(x; \alpha, \nu) = \frac{1}{\Gamma(\nu)} \alpha^\nu x^{\nu-1} e^{-\alpha x} 1_{x>0}, \quad \Gamma(\nu) = \int_0^\infty t^{\nu-1} e^{-t} dt .$$

と表せる .

定義 1.2.2 (noncentral chi-square distribution with  $k$  degree of freedom and noncentrality parameter  $\delta$ ).  $(X_j)_{j \in [k]}$  を  $X_1 \sim N(\nu, 1), X_j \sim N(0, 1) (j = 2, \dots, k)$  を満たす独立な実確率変数列とする . このとき , 二乗和で定まる確率変数  $Y := \sum_{j=1}^k X_j^2$  が定める分布を自由度  $k$  , 非心率  $\delta = \mu^2$  の非心カイ 2 乗分布といい ,  $\chi^2(k, \delta)$  で表す .

注 1.2.3.  $\delta = 0$  のとき,  $\chi^2(k) = \chi^2(k, 0)$  をカイ 2 乗分布という. 自由度  $k = 0$  の非心カイ 2 乗分布は  $x = 0$  上のデルタ測度とし, 非心率は  $\delta = 0$  とする.

命題 1.2.4.  $\chi^2(k, \delta = \mu^2)$  について,

$$(1) \text{ 確率密度関数は } f(x) := e^{-\frac{\mu^2}{2}} \sum_{r=0}^{\infty} \frac{1}{r!} \left(\frac{\mu^2}{2}\right)^r g\left(x; \frac{1}{2}, r + \frac{k}{2}\right).$$

(2) 特性関数は

$$\varphi(u) = e^{-\frac{\mu^2}{2}} \sum_{r=0}^{\infty} \frac{1}{r!} \left(\frac{\mu^2}{2}\right)^r (1 - 2iu)^{-r - \frac{k}{2}} = (1 - 2iu)^{-\frac{k}{2}} \exp\left(\frac{\mu^2 iu}{1 - 2iu}\right)$$

系 1.2.5.  $X_j \sim N(\mu_j, 1)$  ( $j \in [k]$ ) を独立な実確率変数列とする. このとき,  $\delta := \sum_{j=1}^k \mu_j^2$  とすると,  $\sum_{j=1}^k X_j^2 \sim \chi^2(k, \delta)$ .

系 1.2.6.  $(X_j)$  ( $j \in [k]$ ) を,  $\chi^2(k_j, \delta_j)$  に従う独立同分布確率変数列とする. このとき,  $\sum_{j=1}^k X_j \sim \chi^2\left(\sum_{j=1}^k k_j, \sum_{j=1}^k \delta_j\right)$ .

### 1.3 フィッシャー・コ克蘭の定理

命題 1.3.1.  $A_i \in M_n(\mathbb{R})$  ( $i \in [k]$ ) は  $\sum_{i=1}^k A_i = I_n$  を満たすとする. このとき, 次の 4 条件は同値.

- (1)  $\forall i \in [k] \ A_i^2 = A_i$ .
- (2)  $\forall i, j \in [k] \ i \neq j \Rightarrow A_i A_j = O$ .
- (3)  $\mathbb{R}^n = \bigoplus_{i=1}^k \text{Im } A_i$  と直和分解される.
- (4)  $\sum_{i=1}^k \text{rank } A_i = n$ .

記法 1.3.2.  $\mu := (\mu_1, \dots, \mu_n)^\top \in \mathbb{R}^n$ ,  $Y = (Y_1, \dots, Y_n)^\top \sim N_n(v, I_n)$  とする.

定理 1.3.3 (Fisher-Cochran).  $A_i \in M_n(\mathbb{R})$  ( $i \in [k]$ ) は  $\sum_{i=1}^k A_i = I_n$  を満たす対称行列の列とする.  $Q_i := Y^\top A_i Y$ ,  $\text{rank } A_i =: n_i$  とおく. このとき, 次の 5 条件は同値.

- (1)  $\exists \delta_i \geq 0 \ Q_i \sim \chi^2(n_i, \delta_i)$  かつ  $Q_1, \dots, Q_k$  は独立.
- (2)  $\sum_{i=1}^k n_i = n$ .
- (3)  $\forall i \in [k] \ A_i^2 = A_i$ .
- (4)  $\forall i, j \in [k] \ i \neq j \Rightarrow A_i A_j = O$ .
- (5)  $\mathbb{R}^n = \bigoplus_{i \in [k]} \text{Im } A_i$ .

さらにこのとき,  $\delta_i = \mu^\top A_i \mu$  と表わせ, 特に  $\sum_{j=1}^n \mu_j^2 = \sum_{i=1}^k \delta_i$ .

系 1.3.4.  $A \in M_n(\mathbb{R})$  を対称行列とする. 次の 2 条件は同値.

- (1)  $\exists \delta \geq 0 \ \exists k \in \mathbb{Z}_+ \ Q := Y^\top A Y \sim \chi^2(k, \delta)$ .
- (2)  $A^2 = A$ .

このとき,  $k = \text{rank } A$ ,  $\delta = \mu^\top A \mu$  が成り立つ.

### 1.3.1 独立性

系 1.3.5.  $A_i \in M_n(\mathbb{R})$  ( $i = 1, 2$ ) を対称行列とし,  $Q_i := Y^\top A_i Y$  は非心カイ 2 乗分布に従うとする. このとき, 次の 2 条件は同値.

- (1)  $Q_1 \perp\!\!\!\perp Q_2$ .
- (2)  $A_1 A_2 = O$ .

補題 1.3.6.  $A, B \in M_n(\mathbb{R})$  を対称な射影行列とする:  $A^2 = A, B^2 = B$ . 差は非負値行列とする:  $A - B \geq 0$ . このとき,

- (1)  $AB = BA = B$ .
- (2)  $(A - B)^2 = A - B$ .

定理 1.3.7.  $A_i \in M_n(\mathbb{R})$  ( $i = 0, 1, 2$ ) を  $A_0 = A_1 + A_2$  を満たす対称行列とし,  $Q_i := Y^\top A_i Y$  は  $Q_j \sim \chi^2(n_j, \delta_j)$  ( $j = 0, 1$ ),  $Q_2 \geq 0$  a.s. とする. このとき,  $n_2 := n_0 - n_1, \delta_2 = \delta_0 - \delta_1$  について

- (1)  $Q_2 \sim \chi^2(n_2, \delta_2)$ .
- (2)  $Q_1 \perp\!\!\!\perp Q_2$ .

## 1.4 $t$ 分布と $F$ 分布

### 1.4.1 $t$ 分布

定義 1.4.1 (noncentral  $t$ -distribution).  $Y \sim \chi^2(n), X \sim N(0, 1)$  は独立であるとする. このとき,  $X := \frac{Z}{\sqrt{\frac{Y}{n}}}$  が定める分布を, 自由度  $n$ , 非心率  $\delta$  の非心  $t$  分布といい,  $t(n, \delta)$  で表す. 特に  $\delta = 0$  のとき  $t(n) := t(n, 0)$  を自由度  $n$  の  $t$  分布という.

命題 1.4.2 (確率密度関数).

### 1.4.2 $F$ 分布

定義 1.4.3 (noncentral  $F$ -distribution).  $Y_1 \sim \chi^2(m, \delta), Y_2 \sim \chi^2(n)$  は独立であるとする. このとき,  $X := \frac{Y_1/m}{Y_2/n}$  が定める分布を, 自由度  $m, n$ , 非心率  $\delta$  の非心  $F$  分布といい,  $F(m, n, \delta)$  で表す. 特に  $\delta = 0$  のとき  $F(m, n) := F(m, n, 0)$  を自由度  $m, n$  の  $F$  分布という.

命題 1.4.4 (確率密度関数).

## 1.5 ガウス・マルコフモデル

$Y = X\beta + \epsilon$  なる形で表せるモデルを線型回帰モデルといい,  $X \in \mathbb{R}^p$  の次元が  $p \geq 2$  を満たすとき, 重回帰という. この誤差項  $\epsilon$  について,  $E[\epsilon] = 0$  かつ  $\text{Var}[\epsilon] = \sigma^2 I_n$  なる仮定 (Gauss-Markov assumption) をおいた線型回帰モデルを Gauss-Markov モデルという.

記法 1.5.1 (Gauss-Markov model).  $n$  個の実確率変数  $Y_i$  ( $i \in [n]$ ) が, 定数関数  $x_i, \beta: \Omega \rightarrow \mathbb{R}^p$  と, 期待値 0 かつ互いに無相関で等分散を持つ  $E[\epsilon_i] = 0, E[\epsilon_i \epsilon_j] = \sigma^2 \delta_{ij}$  を満たす確率変数列  $\epsilon_i: \Omega \rightarrow \mathbb{R}^{\dagger 1}$  について

$$Y_i = x_i^\top \beta + \epsilon_i$$

なる関係を満たすとする. 添字  $i \in [n]$  は観測の単位を表し,  $x_i$  を,  $E[Y_i]$  を説明する  $p$  個の要因を並べた変数,  $\sigma^2$  は観測の分散

<sup>†1</sup> 独立ならば無相関である.

を表すパラメータである．

$$Y := \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \quad \epsilon := \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}, \quad X := \begin{bmatrix} x_1^\top \\ \vdots \\ x_n^\top \end{bmatrix}$$

をそれぞれ  $n$  次元ベクトル値確率変数と  $n \times p$  次元定行列とすると，モデルは

$$Y = X\beta + \epsilon, \quad E[\epsilon] = 0, \quad \text{Var}[\epsilon] = \sigma^2 I_n.$$

と表せる．すなわち，

$$\mathcal{G} := \{v \in P(\mathbb{R}^n, \mathcal{B}_n) \mid \exists \beta \in \mathbb{R}^p E[v] = X\beta, \exists \sigma \in (0, \infty) \text{Var}[v] = \sigma^2 I_n\}$$

なるモデルである（正規分布よりも遥かに広いクラスである）．

### 1.5.1 正規方程式

結局，線型汎関数  $\beta \mapsto c^\top \beta$  が推定可能であるとき，正規方程式の解から最良線形不偏推定量が構成できる．これが  $M$ -推定量の雛形となる．

定義 1.5.2 (sum of squared residuals, normal equation). 未知パラメータ  $\beta \in \mathbb{R}^p$  の推定を考える．

$$Q(\beta) = \text{SSR}(\beta) := (Y - X\beta)^\top (Y - X\beta) = \sum_{i=1}^n (Y_i - x_i^\top \beta)^2$$

について， $\hat{\beta} := \arg \min_{\beta \in \mathbb{R}^p} Q(\beta)$  とすればよい．特にこれは，正規方程式

$$\frac{dQ}{d\beta} = 0 \Leftrightarrow X^\top X\beta = X^\top Y$$

の解である．この解は  $L[X^\top X] = L[X^\top]$  より，この解は常に存在し， $X^\top X$  が正則であるとき一意に存在する．

定理 1.5.3.  $Y \in \mathbb{R}^n$  とする．正規方程式の任意の解  $\hat{\beta}$  に対して，

- (1)  $X\hat{\beta} = P_X Y$  .
- (2)  $|Y - X\hat{\beta}| = \min_{\beta \in \mathbb{R}^p} |Y - X\beta|$  .

### 1.5.2 最良線型不偏推定量

未知パラメータ  $\beta, \sigma^2$  の関数（主に線型汎関数） $g(\beta, \sigma^2)$  の推定を考える．どの  $(\beta, \sigma^2) \in \Pi \subset \mathbb{R}^p \times (0, \infty)$  が真の値であろうと，良い結果がほしいとなると，これを不偏性という．

定義 1.5.4 (unbiased estimator). 統計量  $\delta(Y) : \text{Meas}(\mathcal{Y}, \mathbb{R}) \rightarrow \mathbb{R}$  について，

$$\forall (\beta, \sigma^2) \in \Pi \quad E_{\beta, \sigma^2}[\delta(Y)] = g(\beta, \sigma^2)$$

が成り立つとき，これを不偏推定量であるという．

定義 1.5.5 (estimable).  $c \in \mathbb{R}^p$  について， $c \in L[X^\top]$  のとき，線型汎関数  $\mathbb{R}^p \ni \beta \mapsto c^\top \beta \in \mathbb{R}$  を推定可能であるという．

定理 1.5.6.  $a \in \mathbb{R}^n, c \in \mathbb{R}^p, \beta \in \mathbb{R}^p$  について，

- (1) 次の2条件は同値．
  - (a) 確率変数  $Y$  の線型関数  $a^\top Y$  が  $\beta$  の線型関数  $c^\top \beta$  の不偏推定量である．
  - (b)  $X^\top a = c$  .
- (2) 次の2条件は同値．



- (a) 線型関数  $c^\top \beta$  の線形不偏推定量が存在する .
- (b)  $c^\top \beta$  は推定可能である .

定理 1.5.7. 線型関数  $c^\top \beta$  は推定可能であるとする . このとき ,

- (1)  $c^\top$  と  $c^\top \hat{\beta}$  は , ある  $b \in \mathbb{R}^p$  を用いて  $c = X^\top X b, c^\top \hat{\beta} = b^\top X^\top Y$  と表される .
- (2)  $c^\top \hat{\beta}$  は正規方程式の解  $\hat{\beta}$  の取り方に依らず ,  $c^\top \beta$  の線形不偏推定量である .

定理 1.5.8 (BLUE: best linear unbiased estimator).  $\beta$  の線型関数  $c^\top \beta$  は推定可能であるとする . このとき , 正規方程式の任意の解  $\hat{\beta}$  に対して ,  $c^\top \hat{\beta}$  は  $c^\top \beta$  の最良線型不偏推定量である :

$$\forall_{(\beta, \sigma^2) \in \Pi} \quad \text{Var}[c^\top \hat{\beta}] = \min \{ \text{Var}[a^\top Y] \geq 0 \mid a^\top Y \text{ は } c^\top \beta \text{ の線形不偏推定量} \} .$$

ただし ,  $\text{Var}$  とは , 真値が  $(\beta, \sigma^2)$  であるときの分散を表す .

定理 1.5.9 (BLUE の特徴付け). 次の 3 条件は同値である .

- (1) 線型統計量  $a^\top Y$  は線型関数  $c^\top \beta$  の最良線型不偏推定量である .
- (2)  $c^\top \beta$  は推定可能で , 任意の真値  $(\beta, \sigma^2) \in \Pi$  に対して ,  $c^\top \hat{\beta} = a^\top Y$  a.s.
- (3)  $X^\top a = c$  かつ  $a \in L[X]$  .

### 1.5.3 分散共分散行列

### 1.5.4 分散の推定

## 1.6 仮説検定

ある部分集合  $\Theta_0 \subset \Theta := \mathbb{R}^p$  の部分集合について , 問題

$$H_0 : H^\top \beta = d \quad \text{vs.} \quad H_1 : H^\top \beta \neq d$$

を検定することを考える . Gauss-Markov モデルに対して , さらに誤差項の分布の正規性  $\epsilon \sim N_n(0, \sigma^2 I_n)$  を仮定する . この仮定は , 誤差分布を楕円形分布に拡張することや , ロバスト推測論によって乗り越えられる .

### 1.6.1 理論

記法 1.6.1.  $R_0^2 := \min_{\beta \in \mathbb{R}^p} |Y - X\beta|^2$  とし ,  $r := \text{rank} X < n$  を仮定する .

定理 1.6.2. 任意の真値  $(\beta, \sigma^2) \in \Pi$  について ,  $\frac{R_0^2}{\sigma^2} \sim \chi^2(n-r)$  .

定理 1.6.3. 次を仮定する .

- (A1)  $H \in M_{p,q}(\mathbb{R})$  ( $q < p$ ) は  $L[H] \subset L[X^\top]$  かつ  $\text{rank} H = q$  を満たすとする .
- (A2)  $d \in \mathbb{R}^q$  について ,  $\Theta_0 = \{\beta \in \mathbb{R}^p \mid H^\top \beta = d\}$  と表せるとする .

このとき ,  $R_1^2 := \{\beta \in \Theta_0\} | Y - X\beta|^2$  について ,

- (1) 任意の真値  $(\beta, \sigma^2) \in \Pi$  について ,  $R_0^2$  と  $R_1^2 - R_0^2$  は独立であり ,  $\frac{R_0^2}{\sigma^2} \sim \chi^2(n-r)$  , かつ , ある  $\sigma \in \mathbb{R}_+$  に対して ,  $\frac{R_1^2 - R_0^2}{\sigma^2} \sim \chi^2(q, \delta)$  . 特に , 
$$\frac{(R_1^2 - R_0^2)/q}{R_0^2/(n-r)} \sim F(q, n-r, \delta) .$$

(2) 真値  $\beta$  が  $\beta \in \Theta_0$  を満たすならば,  $\frac{R_1^2 - R_0^2}{\sigma^2} \sim \chi^2(q)$  であり, 特に

$$\frac{(R_1^2 - R_0^2)/q}{R_0^2/(n-r)} \sim F(q, n-r).$$

命題 1.6.4.

## 1.6.2 応用

このように,  $F$  分布に基づく検定を  $F$ -検定という. このように, データのなすベクトルの 2 次形式による, データの変動の分解を基に検定を構成する手法を分散分析 (ANOVA: ANalysis Of VAriance) と呼ぶ.

## 1.7 平均の検定

## 1.8 重回帰分析

重回帰分析においては, 切片項  $\beta_0$  を含む **Gauss-Markov** モデルが用いられる.

議論 1.8.1.

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \epsilon$$

なる関係で説明したい.  $n$  個の観測  $Y_i$  ( $i \in [n]$ ) を説明したい場合, ベクトル表記で  $Y = X\beta + \epsilon$  とまとめられるが,  $i \in [n]$  が個体番号を表すとき, 誤差項の独立性は成り立つであろう. こうして, Gauss-Markov モデルにたどり着く.

## 1.9 一元配置

one-way ANOVA

## 1.10 二元配置

## 第 2 章

# 統計的決定理論

(意訳) 決定理論の立場から推定と検定の小標本理論を考察する。統計的な推定を行うことは、特定の損失関数を採用して、これについて統計的決定問題を解くことに等しい。

統計的決定問題が設定されたとき、決定問題の下で起こる確率現象を解明するのが数理統計学の課題であって、特定の決定関数無条件に是とするものではない。

## 2.1 統計推論の枠組み

決定すべきは関数 = 射である。その文脈は 8-組で表現できる。

### 2.1.1 Wald の定義

定義 2.1.1 (statistical decision problem, sample space, decision / action space, loss function, nonrandomized, randomized decision function (Abraham Wald)). 次の 8-組  $(\mathcal{X}, \mathcal{A}, \mathcal{P}, \Theta, \mathcal{D}, \mathcal{B}, W, \Delta)$  を統計的決定問題という。

- (1)  $(\mathcal{X}, \mathcal{A})$  は可測空間で、標本空間という。
- (2)  $\Theta$  はパラメータの空間で、 $\mathcal{P} := (P_\theta)_{\theta \in \Theta}$  は  $(\mathcal{X}, \mathcal{A})$  の確率分布の族。
- (3)  $(\mathcal{D}, \mathcal{B})$  は可測空間で、決定空間または行動空間という。
- (4) 第二変数について可測な関数  $W: \Theta \times \mathcal{D} \rightarrow \mathbb{R}_+$  を損失関数という。
- (5)  $\Delta$  は決定関数の族とする。
  - (a) 可測写像  $\delta: \mathcal{X} \rightarrow \mathcal{D}$  を非確率的決定関数という。
  - (b) 第一変数についての currying  $\mathcal{X} \rightarrow \text{Map}(\mathcal{B}, [0, 1])$  は  $(\mathcal{D}, \mathcal{B})$  上の確率測度を与え、第二変数についての currying  $\mathcal{B} \rightarrow \text{Map}(\mathcal{X}, [0, 1])$  は  $\mathcal{X}$  上の  $\mathcal{A}$ -可測関数となる関数  $\delta: \mathcal{B} \times \mathcal{X} \rightarrow [0, 1]$  を確率的決定関数という。

注 2.1.2. 非確率決定関数  $\delta: \mathcal{X} \rightarrow \mathcal{D}$  が与えられたとき、任意の  $B \in \mathcal{B}$  に対して  $\tilde{\delta}(B|x) := 1_{\{\delta(x) \in B\}}$  で  $\tilde{\delta}: \mathcal{B} \times \mathcal{X} \rightarrow [0, 1]$  が定まるから、確率的決定関数の方がより一般的な設定となる。

例 2.1.3 (平均値の推定量の決定).  $(\mathcal{X}, \mathcal{A}) := (\mathbb{R}^n, \mathcal{B}_n)$ ,  $\mathcal{P}_1 := \left\{ P = P_*^n \in \mathcal{M}(\mathbb{R}^n) \mid P_* \text{ は } \mathbb{R} \text{ 上の確率分布で } \int_{\mathbb{R}} x^2 P_*(dx) < \infty \right\}$ ,  $\Theta := \mathcal{P}_1$  とする。決定空間は平均値の全体としたいから  $(\mathcal{D}, \mathcal{B}) := (\mathbb{R}, \mathcal{B}_1)$ 。推定量  $\Delta := \{T_1, T_2, T_3\}$  はそれぞれ  $T_1 := X_1, T_2 := \sum_{j=1}^n \frac{X_j}{n}, T_3 := X_n$  という非確率的決定関数の族で、損失関数は  $W(P, a) := (a - \mu_1(P))^2$  とする。

すると、損失関数  $W$  から定まる危険関数は

$$\begin{aligned} R(P, T_i) &= \int W(P, T_i(x_1, \dots, x_n)) P(dx_1, \dots, dx_n) \quad (P \in \mathcal{P}_1) \\ &= \text{Var}_P[T_i] \end{aligned}$$

となる。こうして、 $R(P, T_1) = R(P, T_3) = \text{Var}_P[X_1], R(P, T_2) = \frac{\text{Var}_P[X_1]}{n}$  となり、 $T_2$  が危険関数  $R$  を最小にするとわかる。

### 2.1.2 危険関数

定義 2.1.4 (risk function, admissible). 危険関数  $R: \Theta \times \Delta \rightarrow \mathbb{R}$  を次のように定める.

- (1) 確率的な決定関数  $\delta: \mathcal{G} \times \mathcal{X} \rightarrow [0, 1]$  については,  $R(\theta, \delta) := \int_{\mathcal{X}} \int_{\mathcal{G}} W(\theta, a) \delta(da|x) P_{\theta}(dx)$  と定める.
- (2) 非確率的な決定関数  $\delta: \mathcal{X} \rightarrow \mathcal{G}$  については,  $R(\theta, \delta) := \int_{\mathcal{X}} W(\theta, \delta(x)) P_{\theta}(dx)$  と定める.

2つの決定関数  $\delta_1, \delta_2 \in \Delta$  について,

- (1)  $\forall \theta \in \Theta R(\theta, \delta_1) \leq R(\theta, \delta_2)$  が成り立つとき,  $\delta_1$  と  $\delta_2$  は同程度に良い決定関数であるという.
- (2) 同程度に良い決定関数がさらに  $\exists \theta_0 \in \Theta R(\theta_0, \delta_1) < R(\theta_0, \delta_2)$  を満たすとき,  $\delta_1$  は  $\delta_2$  より一様に良い決定関数であるという.
- (3) 「一様に良い」という関係は順序を定める.  $\Delta$  に最大元が存在せず, 極大元  $\delta_*$  のみが存在するとき, 許容的であるという.

## 2.2 統計量の十分性と完備性

統計量とは可測関数  $T: (\mathcal{X}, \mathcal{A}) \rightarrow (\mathcal{T}, \mathcal{B})$  をいう. 考えている確率分布族  $\{P_{\theta}\}_{\theta \in \Theta} \subset P(\mathcal{X}, \mathcal{A})$  に応じて,  $T$  がその構造を保つかどうかが変わる. が, 一般的に好ましい性質が定義できる.

### 2.2.1 十分統計量

記法 2.2.1. 可測関数  $T: (\mathcal{X}, \mathcal{A}) \rightarrow (\mathcal{T}, \mathcal{B})$  を考える. 確率分布族  $\{P_{\theta}\}_{\theta \in \Theta} \subset P(\mathcal{X}, \mathcal{A})$  についての条件付き確率  $P_{\theta}[A|T=t]$  が, 母数  $\theta \in \Theta$  に依存しないとき,  $T$  を得てしまえば, これはデータの持つ情報のすべてを持っていると言える.

定義 2.2.2 (sufficient statistic).  $T: \mathcal{X} \rightarrow \mathcal{T}$  が  $\{P_{\theta}\}$  に関して十分であるとは, 次が成り立つことをいう:

任意の  $A \in \mathcal{A}$  に対して,  $\mathcal{B}$ -可測関数  $q(A|t): \mathcal{B} \rightarrow \mathbb{R}$  が存在して,

$$\forall B \in \mathcal{B} \quad \forall \theta \in \Theta \quad \int_B q(A|t) P_{\theta}^{\perp}(dt) = P_{\theta}[A \cap T^{-1}(B)].$$

例 2.2.3 (order statistic). 確率分布族

$$\mathcal{P} := \{P \in P(\mathbb{R}^n) \mid \forall \sigma \in S_n \quad P^{\sigma} = P\}$$

を考えているとき,  $x = (x_1, \dots, x_n)$  の各成分を小さい順に並べて出来る列  $t(x) = (x_{(1)}, \dots, x_{(n)})$  へ写す可測関数  $T: \mathbb{R}^n \rightarrow \mathbb{R}^n; x \mapsto t(x)$  を順序統計量という.

### 2.2.2 因子分解定理

### 2.2.3 Rao-Blackwell の定理

決定関数としては, 十分統計量の関数のみを考えれば十分であることを定立する.

### 2.2.4 完備性

定義 2.2.4 (complete, boundedly complete).

- (1) 分布族  $\mathcal{P} \subset P(\mathcal{X}, \mathcal{A})$  が [有界的] 完備であるとは, 任意の [有界] 可測関数  $f: \mathcal{X} \rightarrow \mathbb{R}$  に関して  $\forall \theta \in \Theta \quad E_{\theta}[f] = 0 \Rightarrow f = 0$   $\mathcal{P}$ -a.s. が成り立つことという. ただし,  $\mathcal{P}$ -a.s. とは, 任意の  $P \in \mathcal{P}$  について  $P$ -a.s. であることをいう.
- (2) 可測関数  $T: \mathcal{X} \rightarrow \mathcal{T}$  が [有界的] 完備であるとは,  $(\mathcal{T}, \mathcal{B})$  上に引き起こされる分布族  $(P_{\theta}^T)_{\theta \in \Theta}$  が [有界的] 完備であることをいう.

## 2.3 指数型分布族

## 2.4 統計的推定

### 2.4.1 不偏推定

定義 2.4.1 (unbiased estimator,  $U$ -estimable). 確率分布族  $(P_\theta)$  に対して,

- (1)  $\forall \theta \in \Theta \ E_\theta[\delta] = \theta$  を満たす推定量  $\delta$  を不偏推定量という.
- (2) パラメータの関数  $g: \Theta \rightarrow \Theta$  に対して不偏推定量が存在するとき,  $g$  を  $U$ -推定可能という.

### 2.4.2 最適性・有効性

一般に不偏推定量は複数ある. そこで, 設定した損失関数に応じて, 「最適」な推定量を選ぶ必要がある. 一般には二乗損失を考えるため, 分散を最適性の指標とし, 分散が最も小さい推定量を有効 (efficient) であるという.

記法 2.4.2. 統計的決定理論の枠組みで不偏推定問題を考えると, 次の通りになる.

- (1) 標本空間  $(\mathcal{X}, \mathcal{A})$  上の確率分布族  $(P_\theta)_{\theta \in \Theta}$  を考える.
- (2) 決定空間は,  $\mathcal{D} \subset \mathbb{R}^p$  を凸ボレル集合,  $\mathcal{B}$  はその上のボレル集合族とする.
- (3) 損失関数  $W$  は凸とする.
- (4)  $\Delta := \Delta_g$  は  $p$  次元関数  $g(\theta)$  の不偏推定量を非確率論的決定関数  $\delta: \mathcal{X} \rightarrow \mathcal{D}$  と観て, その全体

$\Delta_g \neq \emptyset$ , すなわち,  $g: \Theta \rightarrow \Theta^p$  は  $U$ -推定可能とする.

定理 2.4.3 (Lehmann-Scheffe).  $T$  を分布族  $\mathcal{P}$  に対する完備十分統計量とする. 任意の不偏推定量  $\delta \in \Delta_g$  に対して,  $\delta_0 := E[\delta|T]$  とすると,  $\delta_0$  は  $g(\theta)$  の最良の不偏推定量である:  $\forall g' \in \Delta_g \ \forall \theta \in \Theta \ R(\theta, \delta_0) \leq R(\theta, g')$ .

要諦 2.4.4. 一般には,  $p = 1$  次元とし, 損失関数は二乗損失  $W(\theta, a) = [a - g(\theta)]^2$  を考えることが多い. このとき, 危険関数は  $R(\theta, \delta) = \text{Var}_\theta[\delta]$  となる. したがって, 分散を次の「最適性」の指標とする.

定義 2.4.5 (UMVUE: uniformly minimum variance unbiased estimator).  $g(\theta)$  の任意の不偏推定量  $\delta' \in \Delta_g$  に対して,  $\forall \theta \in \Theta \ \text{Var}_\theta[\delta] \leq \text{Var}_\theta[\delta']$  を満たす不偏推定量  $\delta$  を一様最小分散不偏推定量という.

### 2.4.3 Cramer-Rao の不等式

Fisher 情報量は, スコア関数の 2 次のモーメントと定義される. また, KL 情報量の二次の項としても登場する. これを用いて算出される, 不偏推定量の「最適性」の情報理論的境界を, 不偏推定量の分散行列の下界を与えることで示す定理を, Cramer-Rao の不等式という. したがって, Cramer-Rao の不等式の下界を達成する不偏推定量が存在するならば, それは UMV 不偏推定量である.

定義 2.4.6. パラメータ  $\theta = (\theta_i) \in \mathbb{R}^q$  に依存する確率密度関数  $p_\theta(x)$  に対して,

$$I_{ij}(\theta) := \int_{\mathcal{X}} \frac{\partial}{\partial \theta_i} \log p_\theta(x) \cdot \frac{\partial}{\partial \theta_j} \log p_\theta(x) \cdot p_\theta(x) \mu(dx)$$

として定まる行列  $I(\theta)$  を Fisher 情報行列という. 特に  $q = 1$  の場合, Fisher 情報量という.

記法 2.4.7. パラメータの関数  $g: \Theta \rightarrow \Theta^p$  について,  $J(\theta) := \frac{\partial g}{\partial \theta}$  を  $p \times q$ -行列とする.

定理 2.4.8 (Cramer-Rao).

### 2.4.4 ベイズ推定

ベイズ決定関数とミニマックス決定関数の関係を考える。

記法 2.4.9. 次の統計的決定問題を考える。

- (1) 標本空間  $(\mathcal{X}, \mathcal{A})$  上の確率分布族  $(P_\theta)_{\theta \in \Theta}$  について,  $\Theta$  が可測であるとする。
- (2) 損失関数  $W: \Theta \times \mathcal{D} \rightarrow \mathbb{R}_+$  は可測, 危険関数  $R(-, \delta): \Theta \rightarrow \mathbb{R}$  を可測とする。

定義 2.4.10 (prior distribution, Bayes risk, Bayes estimator, minimax decision function).

- (1) 先験的な確率分布  $\pi \in P(\Theta)$  を事前分布という。
- (2)  $R(\pi, \delta) := \int_{\Theta} R(\theta, \delta) \pi(d\theta)$  を,  $\pi$  に関する  $\delta$  のベイズリスクという。
- (3) ベイズリスクを最小にする決定関数  $\delta$  をベイズ決定関数という。これに対応するものが推定量であるとき, ベイズ推定量という。
- (4) 一方で, リスク関数の上限  $\sup_{\theta \in \Theta} R(\theta, \delta)$  を最小にする決定関数をミニマックス決定関数という。

### 2.4.5 非許容性

$p$  変量正規分布族  $(N_p(\theta, \sigma^2 I_p))_{\theta \in \mathbb{R}^p}$  の位置母数の推定量  $\bar{x} = n^{-1} \sum_{j=1}^n x_j$  は UMV 不偏推定量であるが,  $p \geq 3$  のとき非許容である。

## 2.5 統計的仮説検定

### 2.5.1 仮説検定の枠組み

定義 2.5.1 (hypothesis testing, null hypothesis, alternative hypothesis, significance level, critical region). 確率空間  $(\mathcal{X}, P)$  に対して, 事前に与えられた設定  $\alpha \in (0, 1)$ ,  $H_0, H_1$  から定められる部分集合  $\mathcal{X}_1 \subset \mathcal{X}$  を検定という。

- (1) 仮説検定において, 仮説とは, 母数または確率分布に関する (メタ) 条件をいう。
- (2) 疑いの目が向けられており, 検定される仮説  $H_0$  を帰無仮説といい, それに対立する仮説  $H_1$  を対立仮説という。
- (3) 文脈としては,  $H_0$  を棄却し  $H_1$  の成立を示すことで社会的な意味を見出そうという文脈に当てはまるように  $H_0, H_1$  を選ぶ。
- (4) 仮説  $H_0$  の下で計算した確率が「小さい」とはどういう意味かを形式化した指標を有意水準という。  $\alpha = 0.05, 0.01$  などが用いられる。
- (5)  $\mathcal{X}_1 \subset \mathcal{X}$  を  $\alpha$  と  $H_0, H_1$  の意義から事前に定めて棄却域といい, 観測値  $x$  が  $x \in \mathcal{X}_1$  のとき棄却し,  $x \notin \mathcal{X}_1$  のとき採択する。  $\mathcal{X}_1$  は,  $H_0$  の下では非常に稀だが,  $H_1$  の下ではそうでないような観測値の集合である。

例 2.5.2.  $\alpha = 0.05$  とする。20 回中事象  $E$  が 13 回起こったとする。帰無仮説  $H_0$  を  $P(E) = 1/2$  とし, 対立仮説  $H_1$  を  $P(E) > 1/2$  とする。すると, 棄却域は  $x_0 \in \mathbb{N}$  を用いて  $\mathcal{X}_1 = \{x \in [20] \mid x \geq x_0\}$  と定めるのが良いと考えられる。いま,

$$\sum_{k=14}^{20} \binom{20}{k} \left(\frac{1}{2}\right)^{20} \approx 0.0577$$

より,  $13 < x_0$  であって, 観測値 13 は  $\mathcal{X}_1$  には入らないから, 棄却できない。「確率 0.05 程度の事象が偶々起こった」と解釈する方が選択される。

定義 2.5.3 (type I error, type II error).  $H_0$  が正しいのに  $H_0$  を棄却してしまう誤りを第一種の誤りまたは偽陽性といい,  $H_1$  が正しいのに  $H_0$  を採択してしまう誤りを第二種の誤りまたは偽陰性という。

## 2.5.2 ランダム化検定

検定の構成法の例をあげる．検定 = 棄却域の良さは，第一種の誤りの確率を有意水準  $\alpha$  で制限した後に，その範囲内で第二種の誤りの確率を最小化することで得られる．

例 2.5.4. 例 2.5.2 について， $x_0 = 15$  とすると，

$$\sum_{k=15}^{20} \binom{20}{k} \left(\frac{1}{2}\right)^{20} \approx 0.0207$$

より，棄却域のサイズは 5% を大きく切って 2% 程度となってしまう．そこで， $14 < x_0 < 15$  にあたる棄却域を構成したいが， $x_0 \in \mathbb{Z}$  の値は変えられない．そこで，棄却域をランダム化する．

$$\binom{20}{k} \left(\frac{1}{2}\right)^{20} \varphi + \sum_{k=15}^{20} \binom{20}{k} \left(\frac{1}{2}\right)^{20} = \alpha = 0.05$$

を満たす  $\varphi \in (0, 1)$  を用いて，観測値  $x = 14$  に対しては確率  $\varphi$  で  $H_0$  を棄却し，確率  $1 - \varphi$  で  $H_0$  を採択することとする．

## 2.5.3 形式化

定義 2.5.5 (test / critical (function), size, level- $\alpha$ -test, power function, uniformly most powerful (UMP) test). 確率分布族  $\mathcal{P} = (P_\theta)_{\theta \in \Theta} \subset \mathcal{M}(\mathcal{X}, \mathcal{A})$  を考える．

- (1) 棄却域  $\mathcal{X}_1 \subset \mathcal{X}$  の定めるパラメータ空間の分割  $\Theta = \Theta_0 \cup \Theta_1$  を考え，帰無仮説を  $H_0 : \theta \in \Theta_0$  で表し，対立仮説を  $H_1 : \theta \in \Theta_1$  と表す．
- (2)  $|\Theta_i| = 1$  の場合を単純仮説といい，そうでない場合を複合仮説という。<sup>†1</sup>
- (3) 可測関数  $\varphi : \mathcal{X} \rightarrow [0, 1]$  を検定 (関数) という． $\exists A \in \mathcal{A} \ \varphi = 1_A$  であるとき検定  $\varphi$  は非確率的である。<sup>†2</sup>
- (4)  $\sup_{\theta \in \Theta_0} E_\theta[\varphi]$  を  $\varphi$  の大きさという．大きさが  $\alpha$  以下の検定を水準  $\alpha$  検定という．水準  $\alpha$  検定の全体を  $\Phi_\alpha \subset \text{Meas}(\mathcal{X}, [0, 1])$  で表す．
- (5)  $\beta_\varphi : \Theta_1 \rightarrow [0, 1]; \theta \mapsto E_\theta[\varphi]$  を検定出力関数という． $1 - \beta_\varphi(\theta)$  は第二種の誤りの確率を表す．
- (6)  $\forall \varphi \in \Phi_\alpha \ \forall \theta \in \Theta_1 \ \beta_{\varphi_0}(\theta) \geq \beta_\varphi(\theta)$  を満たす検定  $\varphi_0 \in \Phi_\alpha$  を一様最強力検定という．

要諦 2.5.6 (統計的決定理論の枠組みでの解釈). 決定空間は，採択する仮説の添字からなる空間  $\mathcal{D} := 2 = \{0, 1\}$  である (0 が受容，1 が棄却)．損失関数は

$$W(\theta, a) = \begin{cases} 1_{\{1\}}(a), & \theta \in \Theta_0, \\ 1_{\{0\}}(a), & \theta \in \Theta_1. \end{cases}$$

で，検定関数  $\varphi$  は確率的決定関数

$$\delta_\varphi(dz|x) = \varphi(x)\epsilon_1(dz) + (1 - \varphi(x))\epsilon_0(dz)$$

に対応する．ただし， $\epsilon_\alpha$  はデルタ測度とした．したがって危険関数は

$$\begin{aligned} R(\theta, \delta_\varphi) &= \int_{\mathcal{X}} \int_{\mathcal{D}} W(\theta, z) \delta_\varphi(dz|x) \epsilon_0(dz) \\ &= \begin{cases} E_\theta[\varphi], & \theta \in \Theta_0, \\ 1 - E_\theta[\varphi], & \theta \in \Theta_1. \end{cases} \end{aligned}$$

となり， $\Phi_\alpha = \left\{ \varphi \in \text{Meas}(\mathcal{X}, [0, 1]) \mid \sup_{\theta \in \Theta_0} R(\theta, \delta_\varphi) \leq \alpha \right\}$  と表せる．決定関数の全体は  $\Delta \subset \Phi_\alpha$  となる．

$\Delta$  の中に入る一番自然な順序は一様順序で，これによる最大元が存在するときこれを UMP と呼ぶ．

<sup>†1</sup>  $\Theta$  が 1 次元で， $H_0 : \theta = \theta_0, H_1 : \theta \neq \theta_0$  の形のときを両側検定， $H_0 : \theta \leq \theta_0, H_1 : \theta > \theta_0$  の形のときを片側検定という．

<sup>†2</sup> 単に棄却域とそうでない域に分割するため．そうでない場合は，ランダム化検定の方法論を取り入れたことになる．



定義 2.5.7 (nuisance parameter, goodness of fit test).

- (1)  $\Theta$  が2次元以上で、特定のパラメータにしか興味がないとき、分割  $\Theta_1 + \Theta_2$  は商空間となる。この時の興味のない母数を局外母数または攪乱母数という。
- (2) 帰無仮説を「確率変数が正規分布に従う」という命題にする場合、特に分布の適合度検定と呼ぶ。適合度検定や「確率分布が独立である」などの数学的な仮定が現実のデータと矛盾がないかチェックするための検定のクラスを統計的モデルの診断という。

#### 2.5.4 Neyman-Pearson の補題

サイズ  $\alpha$  の検定の中で、検出力  $\beta$  が最も大きい検定の存在を保証し、尤度比検定によって達成されることを主張する補題。  
「 $\alpha$  を決めておき、その中で検出力が最も大きい検定法を選択する」という方針をネイマン・ピアソンの基準という。

#### 2.5.5 単調尤度比と複合仮説の検定

#### 2.5.6 一般化された Neyman-Pearson の補題

#### 2.5.7 不偏検定

一様最強力検定は一般には存在しない。

#### 2.5.8 両側 $t$ -検定

#### 2.5.9 不変検定

### 2.6 区間推定

パラメータ  $\theta \in \Theta$  に対して、区間  $S(x)$  を考えることを、区間推定という。

定義 2.6.1 (confidence coefficient, confidence region). 写像  $S: \mathcal{X} \rightarrow P(\Theta)$  は、 $\alpha \in (0, 1)$  を所与の定数として、次の2条件をみたすとする：

- (1)  $\forall_{\theta \in \Theta} \{x \in \mathcal{X} \mid \theta \in S(x)\} \in \mathcal{A}$ .
- (2)  $\forall_{\theta \in \Theta} P_{\theta}[\theta \in S(x)] \geq 1 - \alpha$

このとき、 $1 - \alpha$  を信頼係数、 $S$  を信頼域という。



## 第 3 章

# 大標本理論

非線形モデル・非ガウスモデルに対しては、統計量の分布が複雑なため、標準的な統計的決定理論を適用するのは困難であるが、これに替わって漸近的方法に基づく大標本理論が有力な解析手段となる。ここで、正規近似に基づく 1 次の漸近理論を扱う。

大標本理論の一番最初の応用先として、統計的検定問題を考える。漸近論が準備できたことにより、尤度比検定はより理解できるようになった。

### 3.1 最尤推定

Ronald Fisher 1912-22 が開発。最尤推定量は一致性を満たすが、この現象はより広いクラスについて成り立つ。

#### 3.1.1 MLE

問題 3.1.1. 可測空間  $(\mathcal{X}, \mathcal{A})$  上の分布族  $(P_\theta)_{\theta \in \Theta}$  について、独立な観測値  $x_1, \dots, x_n$  から  $\theta$  を推定する問題を考える。

定義 3.1.2 (MLE: maximum likelihood estimator).  $(\mathcal{X}, \mathcal{A})$  上の  $\sigma$ -有限測度  $\nu$  に関して  $P_\theta$  が絶対連続であるとし、微分を  $p(x, \theta) := dP_\theta/d\nu(x)$  と表す。

- (1)  $L_n(\theta) := \prod_{j=1}^n p(x_j, \theta)$  によって定まる関数  $\mathcal{X}^n \times \Theta \rightarrow \mathbb{R}$  を尤度関数という。これは、ある  $\theta \in \Theta$  について、これが定める分布  $P_\theta$  から観測値  $x_1, \dots, x_n$  が得られる条件付き確率に当たる。
- (2) 各  $(x_1, \dots, x_n) \in \mathcal{X}^n$  に対して、尤度関数を最大にする  $\widetilde{\theta}_n(x_1, \dots, x_n) \in \Theta$  を対応させる関数  $\widetilde{\theta}_n: \mathcal{X}^n \rightarrow \Theta$  を最尤推定量という。<sup>†1</sup>

注 3.1.3 ( $M$ -推定量としての最尤推定量). 単調増加関数との合成を考えて、対数尤度関数  $l_n := \log \circ L_n$  の最大化を考えるにあたって、パラメータの空間  $\Theta$  が可微分構造を持つとき、ランダムな方程式  $\partial_\theta l_n(\theta) = 0$  の解としての最尤推定量を特徴付けることが出来る。これを尤度方程式という。このとき出現した対数尤度の微分を（標準化されていない）スコア関数  $S(x, \theta) := \frac{\partial}{\partial \theta} l(\theta, x)$  という。

$$\begin{aligned}
 L_n(\theta) &:= \prod_{j=1}^n p(x_j, \theta) \text{ の最大化} \Leftrightarrow l_n(\theta) := \log L_n(\theta) = \sum_{j=1}^n \log p(x_j, \theta) \text{ の最大化} \\
 &\Leftrightarrow S_n(x, \theta) := \frac{\partial}{\partial \theta} l_n(\theta) = \sum_{j=1}^n \frac{\partial}{\partial \theta} \log p(x_j, \theta) \text{ の零点を探す}
 \end{aligned}$$

<sup>†1</sup> ここでは推定量は可測という仮定は置かない。

## 3.1.2 一致性

要諦 3.1.4 (最尤推定量の一致性). 大数尤度関数  $l_n(\theta) := \log L_n(\theta) = \sum_{j=1}^n \log p(x_j, \theta)$  は, 独立同分布に従う確率変数の和である .  
 よって, 大数の法則により, 真値  $\theta_0 \in \Theta$  について,

$$\forall_{\theta \in \Theta} \quad \frac{1}{n} [l_n(\theta) - l_n(\theta_0)] \rightarrow \int \log \left( \frac{p(x, \theta)}{p(x, \theta_0)} \right) P_{\theta_0}(dx) \quad P_{\theta_0}\text{-a.s.}$$

が成り立つ .  $P : \Theta \rightarrow \{P_\theta\}$  が単射であるという識別可能性条件を仮定すると, 相対エントロピーの非負性より, 真値  $\theta_0$  はこの右辺を最大にする唯一の点となる . そこで, 左辺を最大にする推定量が  $\tilde{\theta}_n$  なのであったから,  $n \rightarrow \infty$  のとき,  $\tilde{\theta}_n \rightarrow \theta_0$  が予想される . あとは, 任意の  $\theta \in \Theta$  に対する一様性を確認すれば良い .

補題 3.1.5. 可測空間  $(\Omega, \mathcal{F})$  上の  $\sigma$ -有限測度  $\nu$  に関して絶対連続な確率測度  $P, Q$  について, その微分を  $p(\omega) := dP/d\nu(\omega), q(\omega) := dQ/d\nu(\omega)$  とおく . このとき,

$$\int_{\Omega} \log \left( \frac{p(\omega)}{q(\omega)} \right) P(d\omega) \leq 0$$

で, 等号成立条件は  $P = Q$  のとき .

定義 3.1.6 (consistency). パラメータの推定量  $\theta_n : \mathcal{X}^n \rightarrow \Theta$  について,

- (1) 任意の  $\theta \in \Theta$  について, これが真値であったとき,  $\tilde{\theta}_n \rightarrow \theta$  a.s. ( $n \rightarrow \infty$ ) が成り立つとき, この推定量は強一致性を持つという .
- (2) 推定量  $\tilde{\theta}_n$  が可測で, 任意の  $\theta \in \Theta$  について, これが真値であったとき,  $\tilde{\theta}_n \xrightarrow{P} \theta$  ( $n \rightarrow \infty$ ) が成り立つとき, すなわち,  $\tilde{\theta}_n - \theta = o_p(1)$  が成り立つとき, この推定量は弱一致性を持つという .

たったこれだけの議論で, 相対エントロピーなる確率測度間の距離概念と, 実数の乗法群と加法群の間の準同型  $\mathbb{R}_{>0} \rightarrow \mathbb{R}$  としての対数関数とのいずれもが出てくるとは思わなかった . 「尤度」なる意味論と「相対エントロピー」という意味論とが繋がるのが一番の驚き .

## 3.2 一様対数の法則

最尤推定量の性質を確認するためには,  $\theta \in \Theta$  に依らずに一様に収束することを確認すれば良い .

記法 3.2.1.

- (1)  $(\mathcal{X}, \mathcal{A}, P)$  の可算無限直積  $(\overline{\mathcal{X}}, \overline{\mathcal{A}}, \overline{P})$  上に,  $\pi_n := (\text{pr}_1, \dots, \text{pr}_n) : \overline{\mathcal{X}} \rightarrow \mathcal{X}^n$  で  $\mathcal{X}^n$  上の可測関数を引き戻して議論する .
- (2)  $U : \mathcal{X} \times \Theta \rightarrow (-\infty, \infty]$  に対して,

$$\overline{U}_n(\theta) = \overline{U}_n(x, \theta) = \frac{1}{n} \sum_{j=1}^n U(x_j, \theta)$$

と表す . ただし,  $x = (x_1, \dots, x_n) \in \mathcal{X}^n$  とした .

補題 3.2.2.  $(\Theta, d)$  を距離空間,  $\Theta_1 \subset \Theta$  をコンパクト集合,  $U : \mathcal{X} \times \Theta \rightarrow [-\infty, \infty]$  を関数とし, 以下の条件を仮定する :

- (A1)  $\forall_{x \in \mathcal{X}} \theta \mapsto U(x, \theta)$  は下半連続 .
- (A2)  $\forall_{\theta \in \Theta} r$  が十分に小さければ,  $x \mapsto \inf_{\theta' \in \Theta, d(\theta', \theta) < r} U(x, \theta')$  は可測 .
- (A3) ある可積分関数  $M : \mathcal{X} \rightarrow \mathbb{R}$  が存在して,  $\forall_{(x, \theta) \in \mathcal{X} \times \Theta} U(x, \theta) \geq M(x)$  .

このとき,  $\overline{U}(\theta) := \int_{\mathcal{X}} U(x, \theta) P(dx) \in (-\infty, \infty]$  に対して,

$$(1) \quad \liminf_{n \rightarrow \infty} \inf_{\theta \in \Theta_1} \overline{U}_n(\theta) \geq \inf_{\theta \in \Theta_1} \overline{U}(\theta) \quad \text{a.s.}$$

(2)  $\bar{U}$  は下半連続 .

定理 3.2.3 (一様大数の法則).  $(\Theta, d)$  を距離空間,  $\Theta_1 \subset \Theta$  をコンパクト集合,  $U: \mathcal{X} \times \Theta \rightarrow [-\infty, \infty]$  を関数とし, 以下の条件を仮定する:

(A1)  $\forall x \in \mathcal{X} \ U(x, -): \Theta \rightarrow \mathbb{R}$  は連続 .

(A2)  $\forall \theta \in \Theta \ U(-, \theta): \mathcal{X} \rightarrow \mathbb{R}$  は可測 .

(A3) ある可積分関数  $M \in \mathcal{L}^1(\mathcal{X})$  が存在して,  $\forall (x, \theta) \in \mathcal{X} \times \Theta \ |U(x, \theta)| \leq M(x)$  .

このとき,  $\bar{U}(\theta) := \int_{\mathcal{X}} U(x, \theta) P(dx) \in \mathbb{R}$  は連続で,

$$\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta_1} |\bar{U}_n(\theta) - \bar{U}(\theta)| = 0 \quad \text{a.s.}$$

### 3.3 最小コントラスト推定

最尤推定の枠組みを一般化して, まずは一様大数の法則を準備した. 最小コントラスト推定なる一般的な枠組みを用意して, 一様大数の法則と一致性との間の十分条件を示す.  $U, U_n$  に関わる仮定は, ここまで削ぎ落とすことが出来る. 基本的に, 最尤推定のときのように, 尤度関数  $\Phi_n$  を変形して  $\bar{U}_n$  を作る.

#### 3.3.1 定理と応用

定義 3.3.1 (contrast function, minimum contrast estimator).  $(\Theta, d)$  を距離空間,  $(\mathcal{X}, \mathcal{A}, P)$  を確率空間とする.

(1) コントラスト関数とは, 関数  $\bar{\Psi}: \mathcal{X} \times \Theta \rightarrow (-\infty, \infty]$  をいう.

(2) これを最小にする写像  $\hat{\theta}: \mathcal{X} \rightarrow \Theta$  を最小コントラスト推定量という.

定理 3.3.2 (最小コントラスト推定量が強一致性を持つための十分条件). 関数列  $\bar{U}_n$ , 関数  $\bar{U}: \Theta \rightarrow (-\infty, \infty]$ , 空でない部分集合  $\Theta' \subset \Theta$ , 写像列  $\hat{\theta}_n: \mathcal{X} \rightarrow \Theta$  は次の3条件を満たすとする.

(C1) 一様性:  $\forall \epsilon > 0 \ \liminf_{n \rightarrow \infty} \inf_{\theta \in \Theta: d(\theta, \Theta') \geq \epsilon} \bar{U}_n(\theta) \geq \inf_{\theta \in \Theta: d(\theta, \Theta') \geq \epsilon} \bar{U}(\theta) \quad \text{a.s.}$

(C2) 識別可能性:  $\forall \epsilon > 0 \ \inf_{\theta \in \Theta: d(\theta, \Theta') \geq \epsilon} \bar{U}(\theta) > \inf_{\theta \in \Theta'} \bar{U}(\theta)$ .

(C3) 最小コントラスト推定量である:  $\limsup_{n \rightarrow \infty} \left[ \bar{U}_n(\hat{\theta}_n) - \inf_{\theta \in \Theta'} \bar{U}_n(\theta) \right] \leq 0 \quad \text{a.s.}$

このとき, 一様大数の法則が成り立つ:  $\limsup_{n \rightarrow \infty} \inf_{\theta \in \Theta'} \bar{U}_n(\theta) \leq \inf_{\theta \in \Theta'} \bar{U}(\theta) \quad \text{a.s.}$  ならば,  $\lim_{n \rightarrow \infty} d(\hat{\theta}_n, \Theta') = 0 \quad \text{a.s.}$

例 3.3.3 (最尤推定). 尤度関数は  $L_n: \mathcal{X}^n \times \Theta \rightarrow [0, 1]$  と定め, これを最大化したが, 一致性の議論の中で, 真の分布との対数尤度の差  $|l_n(\theta) - l_n(\hat{\theta})|$  の最小化を考えた.

すなわち, 確率密度関数を  $p(x, \theta)$ , 独立観測を  $x_1, \dots, x_n$  とすると, コントラスト関数は対数尤度関数の  $-\frac{1}{n}$  倍

$$\bar{\Psi}_n(\theta) = -\frac{1}{n} \sum_{j=1}^n \log p(x_j, \theta)$$

で,  $\bar{U}_n$  を真値との対数尤度の差

$$\bar{U}_n(\theta) = \frac{1}{n} \sum_{j=1}^n [\log p(x_j, \theta_0) - \log p(x_j, \theta)]$$

と取ることが出来る.

この大数の法則と定理とを併せた議論は完全に 3.1.4 の議論の厳密化となっている.

### 3.3.2 一貫性を持つ推定量の例

続いて、この枠組みを用いてより一般的なクラスについて一貫性を得よう。この枠組みの強みは、最小コントラスト推定量が陽に（解析的に）計算できない場合でも、一貫性を示すことが出来る点である。

例 3.3.4 (多項分布).  $k$  個の事象への分割  $\Omega = \sum_{i=1}^k E_i$  を考え、 $P_\theta[E_i] =: \theta^i$  を  $k$  個のパラメータとする。この試行を  $n$  回繰り返し、 $E_i$  が観測された回数を  $y_i$  とすると、これは多項分布に従う確率変数である： $(y_1, \dots, y_k) \sim M(n; \theta^1, \dots, \theta^k)$ 。パラメータ空間を

$$\Theta := \left\{ (\theta^1, \dots, \theta^k) \in [0, 1]^k \mid \sum_{i=1}^k \theta^i = 1 \right\}$$

とする。コントラスト関数を

$$\bar{\Psi}_n(\theta) := - \sum_{i=1}^k \frac{y_i}{n} \log \theta^i$$

と定めると、これに対応する最小コントラスト推定量とは最尤推定量に他ならない。これに対して、関数を次のように定めれば、定理が用いることが出来て、一貫性がわかる：

$$\begin{aligned} \bar{U}(\theta) &:= - \sum_{i=1}^k \theta_0^i [\log \theta^i - \log \theta_0^i] \\ \bar{U}_n(\theta) &:= - \sum_{i=1}^k \frac{y_i}{n} [\log \theta^i - \log \theta_0^i] \end{aligned}$$

今回は実は最尤推定量  $\widehat{\theta}_n = (y_1/n, \dots, y_k/n)$  は陽に計算できるが、一般に尤度方程式も、コントラスト関数の最小点も求めることが出来るとは限らない。この枠組みは、そのような場合でも通用するように出来ている。

### 3.3.3 独立観測の構造を仮定した場合

### 3.3.4 可測な最小コントラスト推定量の存在

## 3.4 $M$ -推定量

Huber が 1964 年に最尤推定を一般化し、"maximum likelihood-type" の頭文字から  $M$ -推定量のクラスを定めた。<sup>a</sup> このセミパラメトリックな一般化を経験尤度法 (Owen 1988) という。

<sup>a</sup> <https://en.wikipedia.org/wiki/M-estimator>

定義 3.4.1 (estimating function / equation).  $\Theta \subset \mathbb{R}^p$  とする。関数  $\psi: \mathcal{X} \times \Theta \rightarrow \mathbb{R}^p$  に対して、 $x_1, \dots, x_n$  が定める経験分布についての平均を

$$\bar{\psi}_n(\theta) = \bar{\psi}_n(x, \theta) = \frac{1}{n} \sum_{j=1}^n \psi(x_j, \theta)$$

とおき、推定方程式  $\bar{\psi}_n(\widehat{\theta}_n) = 0$  の零点となるような、未知パラメータの推定量  $\widehat{\theta}_n$  を  $M$ -推定量という。 $\bar{\psi}_n$  を目的関数または効用関数、 $\psi$  を推定関数という。

例 3.4.2.

- (1) 最尤推定量は、スコア関数を推定関数とする  $M$ -推定量であり、尤度方程式  $S_n(x, \theta) = \partial_\theta l(\theta, x) = 0$  の零点として特徴付けられる。
- (2) 最小コントラスト推定量は、 $\partial_\theta \bar{\Psi}_n(\theta) = 0$  なる推定関数に対応する  $M$ -推定量である。

### 3.5 推定量の漸近正規性

一致性は大数の法則に当たるならば、その真値への収束の速度が同様に気になる。これが漸近正規性である。これを、 $M$ -推定量について議論する。

#### 3.5.1 一致性を持つ $M$ -推定量の漸近分布

記法 3.5.1. モデルの可微分構造を仮定： $\Theta \subset \mathbb{R}^p$  を可測集合とする。推定関数  $\psi : \mathcal{X} \times \Theta \rightarrow \mathbb{R}^p$  とし、 $\psi(x, -)$  は連続とする。  
 $\psi_n(\theta) := \mathbb{P}_n \psi(x, \theta) = \sum_{j=1}^n \psi(x_j, \theta)$  と定める。 $\bar{\psi}_n(\theta) := \mathbb{E}_n[\psi(\theta)] = \frac{1}{n} \psi_n(\theta)$ ,  $\bar{\psi}(\theta) := \int_{\mathcal{X}} \psi(x, \theta) P(dx)$  と定める。

定理 3.5.2 (漸近正規性の十分条件).  $\theta_0 \in \Theta^\circ$  とし、 $B \subset \Theta$  を  $\theta_0$  の開近傍とする。次の5条件を仮定する。

- (E1) 推定関数  $\psi : \mathcal{X} \times \Theta \rightarrow \mathbb{R}^p$  は  $B$  上  $C^1$  級。
- (E2)  $\psi(-, \theta)$  は可測。
- (E3)  $\psi(x, \theta_0) \in \mathcal{L}^2(P; \mathbb{R}^p)$  で、 $P\psi(x, \theta_0) = 0$ 。
- (E4)  $\exists M \in \mathcal{L}^1(\mathcal{X}, \mathbb{R}) \forall x \in \mathcal{X}, \theta \in B |\partial_\theta \psi(x, \theta)| \leq M(x)$ 。さらに、 $\Gamma := -P\partial_\theta \psi(x, \theta_0)$  は正則。
- (E5) 確率変数列  $\tilde{\theta}_n : \mathcal{X}^n \rightarrow \Theta$  が  $\tilde{\theta}_n \xrightarrow{P} \theta_0$  かつ  $\psi_n(\tilde{\theta}_n) = o_p(\sqrt{n})$  を満足する。

このとき、

$$\sqrt{n}(\tilde{\theta}_n - \theta_0) - \Gamma^{-1} \frac{1}{\sqrt{n}} \psi_n(\theta_0) = o_p(1) \quad (n \rightarrow \infty).$$

特に、 $\Sigma := \Gamma^{-1} \Phi(\Gamma^\top)^{-1}$ ,  $\Phi := \int_{\mathcal{X}} \phi \phi^\top(x, \theta_0) P(dx)$  について、

$$\sqrt{n}(\tilde{\theta}_n - \theta_0) \xrightarrow{d} N_p(0, \Sigma) \quad (n \rightarrow \infty).$$

#### 3.5.2 最尤推定量での例

定理 3.5.3 (最尤推定量の漸近正規性の十分条件).

#### 3.5.3 一般化モーメント法での例

Lars Hansen は Pearson 1894 によるモーメント法を拡張してファイナンスに応用し、2013 年のノーベル経済学賞を受賞した。モーメントに限らず、真の分布を特徴付ける別の母数  $\psi$  で計算しやすいものがあつたら、これに変換することで  $M$ -推定量を構成する。

定義 3.5.4.  $\{P_\theta\}_{\theta \in \Theta} \subset P(\mathcal{X})$  を分布族、 $g : \mathcal{X} \rightarrow \mathbb{R}^p$  を所与の関数とする。 $\psi(x, \theta) := g(x) - \int_{\mathcal{X}} g(y) P_\theta(dy)$  とすると、真値  $\theta_0 \in \Theta$  について  $E[\psi(x, \theta_0)] = 0$  である。このとき、次の識別可能条件がなりたつとする

$$\theta_1 \neq \theta_2 \Rightarrow \int_{\mathcal{X}} g(y) P_{\theta_1}(dy) \neq \int_{\mathcal{X}} g(y) P_{\theta_2}(dy).$$

このとき、 $\psi(x, \theta)$  の  $x$  に関する  $P_{\theta_0}$ -平均の零点は  $\theta_0$  に限る。そこで、この  $\psi$  を推定関数とした  $M$ -推定量  $\hat{\theta}_n^i$ 、すなわち次の  $\theta$  についての方程式の解を  $g$  に関するモーメント推定量という：

$$\bar{\psi}_n(x, \theta) := \mathbb{P}_n[\psi(x, \theta)] = \frac{1}{n} \sum_{j=1}^n \psi(x_j, \theta) = 0$$

あるいは、あるノルム  $\|\cdot\|$  について、 $\|\bar{\psi}_n(x, \theta)\|$  を最小にする  $\hat{\theta}$  を探す。このとき、推定関数  $\psi$  をモーメント関数という。

このとき、次の条件が成り立てば、モーメント推定量は一致性と漸近正規性を持つ。一方で、最尤推定量のものより漸近分散は大きい。最尤法は一般化モーメント法の特殊な場合とみなせる。

### 3.6 LeCam のワンステップ推定量

一致推定量さえ見つければ、漸近分散は逆 Fisher 情報行列まで改善できる標準的算譜がある

$M$ -推定量は推定関数の零点として定まるが、実際に方程式を解くことは容易でない場合も多い．一方で、有効でないが、一致性をもつ推定量は容易に見つかることがしばしばある．このとき、それを用いて、最尤推定量と同じ漸近分散を持つ推定量を構成できる．

推定量の漸近有効性は、数値解析のような Newton 反復法を採用すれば良いことは初等的に思いつくが、実はこの手続きは 1 回で最尤推定量と同等にまで改善される．<sup>a</sup>

<sup>a</sup> <https://www.ma.imperial.ac.uk/~das01/MyWeb/M3S3/Handouts/OneStep.pdf>

定義 3.6.1 (one-step estimator).  $\Theta$ -値確率変数列  $(\hat{\theta}_n^0)$  から、新たな推定量の系列  $(\hat{\theta}_n)$  を

$$\hat{\theta}_n := \hat{\theta}_n^0 - (\partial_{\theta} \psi_n(\hat{\theta}_n^0))^{-1} \psi_n(\hat{\theta}_n^0)$$

で定めると、これは

$$\mathcal{X}_n^* := \left\{ x \in \mathcal{X}^n \mid \hat{\theta}_n^0 \text{ において } \psi_n \text{ は微分可能で } \partial_{\theta} \psi_n(\hat{\theta}_n^0) \text{ は正則で } \hat{\theta}_n \in \Theta \right\}$$

上で定まるから、これを  $\mathcal{X}$  上に可測に延長する．延長の仕方は任意で良い．こうして得た  $(\hat{\theta}_n)$  を  $(\hat{\theta}_n^0)$  を初期推定量とするワンステップ推定量という．

定理 3.6.2 (ワンステップ推定量が漸近正規であるための十分条件).  $\theta_0 \in \Theta^\circ$ ,  $B \subset \Theta$  を  $\theta_0$  の開近傍とする．条件 (E1) (E4) を仮定し、次の条件を仮定する：

(E6) 初期推定量の列  $\hat{\theta}_n^0 : \mathcal{X}^n \rightarrow \Theta$  が  $\sqrt{n}$ -一致性を持つ： $\hat{\theta}_n^0 - \theta_0 = O_p(1/\sqrt{n})$ .<sup>t2</sup>

このとき、ワンステップ推定量  $\hat{\theta}_n$  に対して、定理 3.5.2 と同様の結果が成り立つ．

### 3.7 $M$ -推定量の存在

Cramer による．

### 3.8 頑健推定

#### 3.8.1 影響関数

影響関数は、有界線型作用素  $T : P(\mathcal{X}) \rightarrow \Theta$  の特別な微分とみなせる．

記法 3.8.1. 推定関数  $\Psi : \mathcal{X} \times \Theta \rightarrow \mathbb{R}^p$  に対して、写像  $T : P(\mathcal{X}) \rightarrow \Theta$  は  $\int_{\mathcal{X}} \psi(x, T(P)) P(dx) = 0$  を満足すると仮定する．このとき、 $P$  が経験測度（デルタ測度の凸結合）であるとき、 $\hat{\theta}_n := T(P_n)$  は推定関数  $\Psi$  に関する  $M$ -推定量である．写像  $T : P(\mathcal{X}) \rightarrow \Theta$  はその延長である．デルタ測度は凸集合  $P(\mathcal{X})$  の極点であり、Krein-Milman の定理より、その凸包は  $P(\mathcal{X})$  上弱稠密であることに注意．もし  $T$  が有界線型作用素ならば、一意的な連続延長がある．

定義 3.8.2.

- (1) 確率測度  $P \in P(\mathcal{X})$  に対して、 $x \in \mathcal{X}$  が定めるデルタ測度との  $\epsilon \in [0, 1]$  が定める凸結合を  $P_\epsilon^x := (1 - \epsilon)P + \epsilon\delta_x \in P(\mathcal{X})$  ( $x \in \mathcal{X}, \epsilon \in [0, 1]$ ) と定める．

<sup>t2</sup> 条件  $\hat{\theta}_n^0 - \theta_0 = O_p(1/\sqrt{n})$  は、弱一致性  $\hat{\theta}_n^0 - \theta_0 = o_p(1)$  を含意することに注意．これに加えて、収束レートが  $n^{-1/2}$  であることをいう．

(2)  $T$  について,  $P \in P(\mathcal{X})$  における  $\epsilon \in [0, 1]$  に関する微分係数  $\text{IF}(x; T, P) := \frac{d}{d\epsilon} T(P_\epsilon^x)|_{\epsilon=0} = \Gamma(P)^{-1} \psi(x, T(P))$  を関数  $\mathcal{X} \rightarrow \Theta$  とみたとき, これを  $T$  の  $P$  における影響関数という. これは, 各点  $x \in \mathcal{X}$  に観測を得た場合に,  $M$ -推定量を与える写像  $T$  がどれほど影響を受けるかを表す指標とみなせる.

例 3.8.3. 正規母集団  $N(\theta, 1)$  の位置母数  $\theta$  に対する最尤推定量  $\hat{\theta}_n$  の影響関数は  $\text{IF}(x; T, N(\theta, 1)) = x - \theta$  となる. したがって,  $x \in \mathbb{R}$  の絶対値が大きいほど, 最尤推定量への影響が大きいので, ここに異常値が来ると, 大きく推論に影響する危険がある.

### 3.8.2 バイアスロバスト推定量

頑健統計では, 外れ値による影響がなるべく小さい  $M$ -推定量を構成することを考える.

定義 3.8.4 (bias robust estimation).  $(P_\theta)_{\theta \in \Theta}$  をモデルとする. ある定数  $M \in \mathbb{R}$  が存在して,  $\Gamma_\theta := \Gamma(P_\theta)$  について  $|\Gamma_\theta^{-1} \psi(x, \theta)| \leq M$  を満たす  $\psi$  が定める  $M$ -推定量を  $B$ -ロバスト推定量という.

例 3.8.5. 最尤推定量は, 漸近分散が最小という意味では最適であるが, 一般に影響関数是有界にはならない.

### 3.8.3 Fisher の一致性

Fisher の一致性を仮定すると, 最適な不偏ロバスト推定量を構成できる.

定義 3.8.6 (Fisher consistency). 写像  $T: P(\mathcal{X}) \rightarrow \Theta$  が, モデル  $(P_\theta)$  について,  $T(P_\theta) = \theta$  を満たすとき, **Fisher** 一致性を満たすという.

## 3.9 尤度比検定

初等統計で扱われるカイ 2 乗検定の根拠は, この漸近理論によって明らかになる.

### 3.9.1 尤度比検定

記法 3.9.1.  $\Theta \subset \mathbb{R}^p$ ,  $\theta_0 \in \Theta^\circ$  とする.  $(\mathcal{X}, \mathcal{A})$  上の  $\sigma$ -有限な確率分布族  $\{P_\theta\}_{\theta \in \Theta} \subset P(X)$  に対する検定

$$H_0: \theta = \theta_0 \quad \text{vs.} \quad H_2: \theta \neq \theta_0$$

を考える.  $x = (x_1, \dots, x_n) \in \mathcal{X}^n$  を無作為標本とする.

定義 3.9.2. 尤度関数  $L_n(\theta) := \prod_{j=1}^n p(x_j, \theta)$  の比

$$\Lambda_n := \frac{L_n(\theta)}{\sup_{\theta \in \Theta} L_n(\theta)} \in [0, 1]$$

の値を考え,  $\Lambda_n: \mathcal{X}^n \rightarrow [0, 1]$  の値が小さい時に帰無仮説を棄却する検定を尤度比検定という.

注 3.9.3.  $\hat{\theta}_n$  が  $\theta$  の最尤推定量であるとき, 定義上これが尤度関数を最大にする点であるから,  $\Lambda_n = \frac{L_n(\theta_0)}{L_n(\hat{\theta}_n)}$  となる.



### 3.9.2 棄却域の定め方

では、棄却域はどう定めるのが「良い」か？  $\Lambda_n : \mathcal{X}^n \rightarrow [0, 1]$  の分布は極めて複雑であるため、近似を用いる。ここで漸近論が登場し、漸近正規性が、ミニマックスの意味で「妥当」であることが言える。正しくは、漸近決定理論の枠組みで捉えるのが良い。

定義 3.9.4.

- (1)  $\theta$  に関する、規準化されたスコア関数を  $Z_n(\theta) := \frac{1}{\sqrt{n}} \sum_{j=1}^n \partial_\theta \log p(x_j, \theta)$  とする。
- (2)  $Z_n := Z_n(\theta_0)$  とする。

定理 3.9.5.  $\hat{\theta}_n \xrightarrow{P_{\theta_0}} \theta_0$  を満たす最尤推定量が存在し、漸近正規性を持つための十分条件 [F1] [F4] を満たすとする 3.5.3。このとき、 $P_{\theta_0}$  の下で

$$-2 \log \Lambda_n \xrightarrow{d} \chi^2(p) \quad (n \rightarrow \infty)$$

### 3.9.3 複合仮説の検定

#### 3.9.4 Rao 検定

複合仮説の検定に関しては、尤度比検定に変わって、ラオ検定 (Lagrange 未定乗数法検定) を用いることが出来る。これは  $H_0$  の下で尤度比検定と同じ漸近分布  $\chi^2(r)$  を持つ。

#### 3.9.5 Wald 検定

Wald 検定も尤度比検定と漸近同等であって、特に  $H_0$  の下で漸近的に分布  $\chi^2(r)$  に従う。

## 3.10 多項分布の検定

記法 3.10.1 (モデルの設定).

$$\Delta := \left\{ p = (p_1, \dots, p_k) \in \mathbb{R}_+^k \mid \sum_{i=1}^k p_i = 1 \right\}$$

とするとこれは多項分布の全体とみなせる。開集合  $\Theta \subset \mathbb{R}^l$  から  $\text{Int}(\Delta)$  への単射な  $C^1$  級写像  $p \mapsto \Theta \rightarrow \text{Int}(\Delta)$  で定まるパラメトリックモデル  $\Theta$  を考える 3.3.4。さらに  $U \subset \mathbb{R}^m$  ( $1 \leq m < l \leq k-1$ ) 上の  $C^1$  級部分モデル  $\theta : U \hookrightarrow \Theta$  を考える。

問題 3.10.2. 仮説検定問題

$$H_0 : \theta \in \theta(U) \quad \text{vs.} \quad H_1 : \theta \notin \theta(U)$$

を考える。スコア関数と漸近同等な  $\theta, u$  の推定量  $\hat{\theta}_n, \hat{u}_n$  がモデル  $\Theta, U$  のそれぞれに対して得られているとする。このとき、検定統計量  $Q_n, Q_n^*$  を

$$Q_n = \sum_{i=1}^k \frac{[np_i(\hat{\theta}_n) - np_i \circ \theta(\hat{u}_n)]^2}{np_i \circ \theta(\hat{u}_n)}$$

$$Q_n^* = \sum_{i=1}^k \frac{[np_i(\hat{\theta}_n) - np_i \circ \theta(\hat{u}_n)]^2}{np_i(\hat{\theta}_n)}$$

とし、 $Q_n, Q_n^*$  の値が大きいに  $H_0$  を棄却することを考える。



定理 3.10.3.  $H_0$  の下で,  $Q_n, Q_n^*$  の漸近分布はカイ 2 乗分布  $\chi^2(l-m)$  である.

要諦 3.10.4. 証明抽出により, 尤度比検定と漸近同等であることがわかる.

例 3.10.5 (goodness-of-fit test). 各事象  $E_i$  が起こる確率  $p_i$  は, 与えられた値  $p_i^*$  に等しい, という単純仮説の検定を適合度検定という. このとき, 漸近分布は  $\chi^2(k-1)$  である.

### 3.11 尤度比確率場の局所漸近構造

統計推測の最適性の問題は, 尤度解析に帰着される.

#### 3.11.1 漸近決定理論

漸近決定理論の方法は, その普遍性ゆえに, 確率過程の統計推測論などの新しい分野と合流し発展を続けている. 統計量の漸近有効性を議論するためには, 統計的実験の列に対する漸近決定理論の立場で一般論を展開する方が好ましい. これは, 小標本理論における十分性・不偏性・不変性などの精緻な理論を, 非線形な場合にも拡張する企てである.

#### 3.11.2 尤度比確率場の接触構造

定義 3.11.1 (contiguous). 可測空間の列  $(\mathcal{X}^n, \mathcal{A}^n)$  と, その上の確率測度の列が 2 つ  $(P_n), (Q_n)$  と与えられているとする.  $(Q_n)$  は  $(P_n)$  に接触しているとは, 任意の可測集合の列  $(A_n), A_n \in \mathcal{A}^n$  に対して,  $P_n(A_n) \rightarrow 0 \Rightarrow Q_n(A_n) \rightarrow 0$  が成り立つことをいう.

補題 3.11.2 (接触性の特徴付け). 次の 2 条件は同値.

- (1)  $(Q_n)$  は  $(P_n)$  に接触している.
- (2) 任意の確率変数列  $X_n: \mathcal{X} \rightarrow \mathbb{R}$  に対して,  $X_n \xrightarrow{P_n} 0 \Rightarrow X_n \xrightarrow{Q_n} 0$ .

定理 3.11.3 (接触性の十分条件).  $P_n, Q_n$  はある  $\sigma$ -有限測度  $\nu_n$  に対して絶対連続であるとし, Radon-Nikodym 微分を  $p_n, q_n$  とおく.

$$\Lambda_n := \begin{cases} \log(q_n/p_n), & p_n, q_n > 0, \\ 0, & p_n = q_n = 0, \\ \infty, & p_n = 0, q_n > 0, \\ -\infty, & p_n > 0, q_n = 0. \end{cases}$$

ある確率変数  $\Lambda$  が存在して  $P_n$  の下で  $\Lambda_n \xrightarrow{d} \Lambda$  かつ  $E[e^\Lambda] = 1$  であるならば,  $(Q_n)$  は  $(P_n)$  に接触している.

#### 3.11.3 統計的実験

漸近正規性を定義する枠組みを整備する. Frechet 微分のような考え方をするが,  $L^2$  ノルムにより条件が弱められている.

定義 3.11.4. 可測空間とその上の確率分布族  $(P_\theta)_{\theta \in \Theta}$  の組  $(\mathcal{X}, \mathcal{A}, \{P_\theta\}_{\theta \in \Theta})$  を統計的実験という. 大標本理論では, 直積による列  $(\mathcal{X}^n, \mathcal{A}^n, \{P_\theta^n\}_{\theta \in \Theta})$  を考える.

記法 3.11.5.  $\Theta \subset^{\text{open}} \mathbb{R}^p$  とし,  $P_\theta$  の真値  $P_{\theta_0}$  に対する Lebesgue 分解における, 絶対連続部分の Radon-Nikodym 微分を  $dP_\theta/dP_{\theta_0}$ , 特異部分を  $\sigma_{\theta/\theta_0}$  で表すと,

$$P_\theta[A] = \int_A \frac{dP_\theta}{dP_{\theta_0}} dP_{\theta_0} + \sigma_{\theta/\theta_0}(A) \quad (A \in \mathcal{A})$$

関数  $D_u : \mathcal{X} \rightarrow \mathbb{R}$  ( $u \in \mathbb{R}^p$ ) を

$$D_u(x) := \sqrt{\frac{dP_{\theta_0+u}}{dP_{\theta_0}}(x)} - 1$$

で定める .

(G1) 原点の近傍の  $u \in \mathbb{R}^p$  (十分小さな変位) に対して  $D_u \in L^2(P_{\theta_0})$  であって ,  $u \mapsto D_u$  は  $u = 0$  において  $L^2(P_{\theta_0})$  の意味で微分可能である :

$$\exists_{\varphi=\varphi_{\theta_0} \in L^2(P_{\theta_0}; \mathbb{R}^p)} \int_{\mathcal{X}} (D_u(x) - u' \varphi(x))^2 P_{\theta_0}(dx) = o(|u|^2) \quad (|u| \rightarrow 0).$$

(G2)  $\sigma_{\theta_0+u/\theta_0}(\mathcal{X}) = o(|u|^2)$  ( $|u| \rightarrow 0$ ) .

補題 3.11.6. G1,G2 を仮定し ,  $P_\theta$  は確率密度関数族  $p(x, \theta)$  を持ち , 対応  $\theta \mapsto p(x, \theta)$  は  $\theta_0$  において微分可能とする . このとき ,

$$\varphi(x) = \frac{1}{2p(x, \theta_0)} \frac{\partial p}{\partial \theta}(x, \theta_0).$$

また , Fisher 乗法行列に当たる作用素  $\Theta \rightarrow \mathbb{R}$  を

$$I(\theta_0) := 4 \int_{\mathcal{X}} \varphi(x) \varphi(x)^\top P_{\theta_0}(dx)$$

とおく .

補題 3.11.7.

### 3.11.4 局所漸近正規性

定理の  $Z_n(u; \theta_0)$  の展開が成り立つとき , すなわち , G1,G2 が成り立つとき , 統計的実験  $(\mathcal{X}^n, \mathcal{A}^n, \{P_\theta^n\}_{\theta \in \Theta})$  は  $\theta_0$  において局所漸近正規であるという . 対応  $\mathbb{R}^p \ni u \mapsto Z_n(u; \theta_0) \in C(\mathbb{R}^p)$  を尤度比確率場という . Ibragimov HasMinskii は , この関数の  $C(\mathbb{R}^p)$  上での弱収束を示した :  $Z_n(-; \theta_0) \xrightarrow{d} Z(-; \theta_0)$  .

記法 3.11.8.

$$Z_n(u; \theta_0) := \prod_{j=1}^n \frac{dP_{\theta_0+n^{-1/2}u}}{dP_{\theta_0}}(x_j), \quad (u \in \mathbb{R}^p, x \in \mathcal{X}^n)$$

とおくと ,  $Z_n$  は十分大きな  $n \in \mathbb{N}$  に対して定義されている .

定理 3.11.9.

定義 3.11.10 (LAN: Locally Asymptotically Normal). 定理の  $Z_n(u; \theta_0)$  の展開が成り立つとき , すなわち , G1,G2 が成り立つとき , 統計的実験  $(\mathcal{X}^n, \mathcal{A}^n, \{P_\theta^n\}_{\theta \in \Theta})$  は  $\theta_0$  において局所漸近正規であるという .

### 3.11.5

sup によって , パラメータの 1 点でのみ良いパフォーマンスを示す病的な推定量 (超有効推定量) が競争から排除されている . これが漸近ミニマックスリスク関数である .

定理 3.11.11. 統計的実験は  $\theta_0$  にて局所漸近正規であるとする .  $\theta$  の任意の推定量系列  $T_n$  について ,

$$\forall_{\gamma, \delta > 0} \liminf_{n \rightarrow \infty} \sup_{\theta: |\theta - \theta_0| < \delta} E_\theta [|\sqrt{n}(T_n - \theta)|^\gamma] \geq E[|I(\theta_0)^{-1} \Delta(\theta_0)|^\gamma].$$

## 3.12 情報量規準

### 3.12.1 枠組みと KL 距離

記法 3.12.1.  $\sigma$ -有限測度  $\nu$  に支配されている確率分布族  $\mathcal{P} = \{P_\theta\}_{\theta \in \Theta} \subset P(\mathcal{X}, \mathcal{A})$  を考える．真の分布  $Q$  の確率密度関数も存在して  $q(x) := dQ/d\nu(x)$  と表され， $\mathcal{P}$  に属しているかは不明とする． $n$  個の無作為標本  $x = (x_1, \dots, x_n)$  に基づく未知パラメータ  $\theta$  の推定量を  $\hat{\theta}_n(x)$  とする．新たな観測に対する予測分布として，plug-in 分布  $P_{\hat{\theta}_n}$  で  $Q$  を近似するのは自然である．

定義 3.12.2 (statistical manifold).  $\sigma$ -有限測度  $\nu$  に支配されている確率分布族  $\mathcal{P} = \{P_\theta\}_{\theta \in \Theta}$  が正則統計モデルであるとは，

- (1) パラメータ空間  $\Theta$  はある Euclid 空間  $\mathbb{R}^p$  の開集合と同相．
- (2)  $\theta \mapsto P_\theta$  は微分可能．
- (3) 非特異な Fisher 情報行列を持つ．

このとき， $\mathcal{P}$  は  $p$ -次元多様体をなす．

定義 3.12.3 (divergence).  $(M, (\xi_x))$  を多様体とする．関数  $D: M \times M \rightarrow \mathbb{R}_+$  が分離度であるとは，

- (1) 非退化性:  $D[P, Q] = 0 \Leftrightarrow P = Q$ ．
- (2) 十分近い 2 点  $\xi, \xi + d\xi$  について， $D[\xi, \xi + d\xi] = \frac{1}{2} g_{ij}(\xi) d\xi_i d\xi_j$  が定める行列  $G(\xi) = (g_{ij}(\xi))$  は正定値対称である．
- (3) 点  $P \in M$  の  $r$ -近傍  $N(P, r) := \{Q \in M \mid D[P, Q] < r\}$  は  $r$  について単調に増大する．

要諦 3.12.4. 対称性も三角不等式も成り立たず，距離の概念とは違い，むしろ距離の二乗のような，分散のような概念である．そして，微小な 2 点間の「Riemann 距離」を与えるが，非対称性は受け継がれ，これが双対性という方向を持った新たな構造を授ける．

定義 3.12.5 (Kullback-Leibler information / divergence / relative entropy).  $P, Q \in \mathcal{P}$  について，

$$I[Q, P] := \int_{\mathcal{X}} q(x) \log \frac{q(x)}{p(x)} \nu(dx)$$

によって定まる関数  $I: \mathcal{P} \times \mathcal{P} \rightarrow \mathbb{R}_+$  を **Kullback-Leibler 情報量** という．

例 3.12.6 (正規分布族の KL 分離度).  $M_n(\mathbb{R})$  の中の半正定値行列のなす部分多様体  $GL_n(\mathbb{R})_+$  は  $n(n+1)/2$  次元である．この多様体に

$$D[P, Q] = \text{Tr}(PQ^{-1}) - \log|PQ^{-1}| - n$$

なるダイバージェンスが定まる．これは， $P, Q$  を分散共分散行列とする平均 0 の正規分布の間の KL ダイバージェンスになっている．

### 3.12.2 バイアス補正

真の分布と，そのモデルがはじき出す予測分布との分離度を，そのモデルの misspecification の度合いを測る尺度に用いるという発想である．しかし，分離度を経験分布から求めても，そのバイアスの推定の仕事が残っている．こうして得られる種々の分離度の推定量を，情報量規準という．

議論 3.12.7. 真の分布  $Q$  と予測分布  $P_{\hat{\theta}_n}$  の分離度は

$$I[Q, P_{\hat{\theta}_n}] = \int_{\mathcal{X}} q(z) \log q(z) \nu(dz) - \int_{\mathcal{X}} q(z) \log p(z, \hat{\theta}_n) \nu(dz).$$

ここで，モデルの選択に依るのは第二項

$$h(Q, p, \hat{\theta}_n(x)) := \int_{\mathcal{X}} \log p(z, \hat{\theta}_n(x)) Q(dz)$$

である．

要諦 3.12.8.  $Q$  の経験分布  $P_n$  について,  $h(P_n, p, \hat{\theta}_n(x))$  で推定することになるが, これが最大になるときの  $\hat{\theta}_n$  を最尤推定量というのであった. これが, 情報理論的な観点から再発見される最尤法の意味の一つである.

議論 3.12.9 (バイアスの推定). 第二項  $h$  を経験分布から求めることになるが, そのときのバイアスはどうなるか.

$$b_n := \int_{\mathcal{X}^n} [h(P_n, p, \hat{\theta}_n(x)) - h(Q, p, \hat{\theta}_n(x))] Q^n(dx).$$

このとき, 次が成り立つ.

定理 3.12.10. ある  $\theta_* \in \Theta$  と  $\varphi_Q \in L^2(\mathcal{X}, Q; \mathbb{R}^k)$  が存在して,  $\int_{\mathcal{X}} \varphi_Q(x) Q(dx) = 0$  かつ

$$R := \sqrt{n}(\hat{\theta}_n - \theta_*) - \zeta_n \xrightarrow{Q_n} 0, \quad \zeta_n := n^{-1/2} \sum_{j=1}^n \varphi_Q(x_j)$$

が成り立つと仮定する.<sup>†3</sup>

スコアを  $s(x, \theta) = \partial_{\theta} \log p(x, \theta)$ ,  $C := \text{Cov}_Q[s(-, \theta_*)^{\top}, \varphi_Q]$  とすると,  $nb_n \xrightarrow{n \rightarrow \infty} \text{Tr } C$ .

定義 3.12.11 (information criterion). 尤度関数にペナルティ項  $p$  を付けた

$$IC = -2 \left[ \sum_{j=1}^n \log p(x_j, \hat{\theta}_n(x)) - p(n) \right]$$

を, 推定量  $\hat{\theta}_n$  に対する情報量規準という.

例 3.12.12.

- (1) バイアス修正を考えるならば,  $p(n) = \text{Tr } C$  と取るとよいが, これは  $Q$  に依存するので, その一致推定量を取るとよい.
- (2) 推定量  $\hat{\theta}_n$  が  $\Psi$  が定める最小コントラスト推定量であるときの形は, 小西貞則・北川源四郎による一般化情報量規準 (GIC: Generalized Information Criterion) の特殊な場合である. 特に最尤推定量であるとき, 補正項は  $p(n) = \text{Tr}(J_Q^{-1} I_Q)$  となり, この一致推定量で置き換えた情報量規準を竹内の情報量規準 (TIC) という.
- (3)  $\hat{\theta}_n$  が最尤推定量である上で特にモデルが真の分布  $Q$  を含むとき, パラメータ数を  $k$  として  $p(n) = k$  となる. この場合を赤池情報量規準 (AIC) という.

歴史 3.12.13 (「バイアス補正法」として拡張の歴史を辿った). モデルの評価基準を「その予測能力としよう」として初めて提案したのは赤池 (1973, 74). モデルが真の分布を含まない場合にも拡張したのが竹内啓 (76) である. 小西・北川 (96) は, バイアス補正の方法を一般の統計量に拡張した. 石黒 (97) はブートストラップ法によってバイアス補正を行うことを提案し, 拡張情報量規準 EIC(Extended IC) と呼ばれる.

注 3.12.14. AIC の発明のときから取っている基本的な考え方として次の3点がある.

- (1) モデルの良さはその予測能力でみる.
- (2) 予測は分布に対して行う.
- (3) 分布の近さは (二乗誤差などではなく) KL 情報量で測る.

要諦 3.12.15. 本懐はバイアスの補正であるとするなら, ロバスト統計と考えていることの方角性は同じではないか?

### 3.13 密度推定

ここからはノンパラメトリック解析の手法を概観する.

### 3.14 $U$ -統計量

<sup>†3</sup> 最小コントラスト推定量はこの仮定を満たす.

## 第 4 章

# 漸近展開とその応用

中心極限定理は、正規近似に基づく、分布の一次の漸近理論であった。Edgeworth による展開はその近似をより精密にしたものである。標本数がそれほど大きくない実際的な状況では、分布のより精密な近似を基礎として統計量を構成することが必要になる。

今日、漸近展開法も確率過程にまでその領域を広げ、新しい確率統計学が発展しつつあるが、このような新領域を理解する上でも、独立確率変数列における現象を理解することが重要である。

キュムラント関数をよく使う。その関係で、テンソルがよく出る。また、最尤推定量の漸近展開に出てくる係数の一部は接続係数にほかならない。

### 4.1 漸近展開

平均 0、分散共分散行列  $\Sigma > 0$  の  $d$  次元確率変数列  $\{Z_j\}_{j \in \mathbb{N}}$  は独立同分布に従うとする。中心極限定理によると、

$$S_n = \frac{1}{\sqrt{n}} \sum_{j=1}^n Z_j \xrightarrow{d} N_d(0, \Sigma) \quad (n \rightarrow \infty)$$

であるが、この正規近似はあまり精度は良くないため、標本数  $n$  が小さい時にはより良い近似が必要になる。数学的には、これは  $X$  の特性関数  $\varphi_X(u) = E[e^{iu \cdot X}]$  ( $u \in \mathbb{R}^d$ ) の漸近展開理論に等価である。

定義 4.1.1. 確率変数  $X: \mathcal{X} \rightarrow \mathbb{R}^d$  に対して、

$$\chi_r(u; X) := i^r \kappa_r[u \cdot X] = (\partial_\epsilon)_0^r \log \varphi_{u \cdot X}(\epsilon)$$

をキュムラント関数という。ただし、 $(\partial_\epsilon)_0^r$  とは、 $\epsilon$  で  $r$  回偏微分をして  $\epsilon = 0$  を代入したものをいう。これはキュムラント母関数の  $r$  次項係数（を  $u$  の関数と見たもの）である。

### 4.2 平滑化補題

分布の差と特性関数の差の関係を与える補題を準備する。

## 4.3 特性関数の展開

## 4.4 漸近展開の正当性の証明

## 4.5 漸近展開の変換

## 4.6 最尤推定量の漸近展開

## 4.7 漸近展開と情報幾何

## 4.8 ブートストラップ法

ブートストラップ法では、データから作った分布  $\hat{P}_n$  から、さらに標本を取る。そこから、リサンプリング法とも言われる。パラメトリックモデルではプラグイン推定量  $P_{\hat{\theta}_n}$ 、ノンパラメトリックモデルでは経験分布などを取る。

記法 4.8.1.  $(X_j)$  を確率分布  $P$  に従う独立な実確率変数列とし、 $R_n(\mathbf{X}_n, P) : \mathcal{X} \rightarrow \mathbb{R}$  を確率変数とする。 $R_n(\mathbf{X}_n, P)$  の分布関数を

$$H_n(x; P) := \int_{\mathbb{R}^n} 1_{R_n(\mathbf{x}, P) \leq x} P_n(d\mathbf{x})$$

と定め、 $H_n(c_\alpha; P) = \alpha \in (0, 1)$  を満たす点  $c_\alpha$  を求める問題を考える。

例 4.8.2.  $P$  が未知で、 $\theta = \theta(P)$  の信頼区間を構成したいとき、推定量  $\hat{\theta}_n(\mathbf{X}_n)$  から定まる確率変数  $R_n(\mathbf{X}_n, P) := \sqrt{n}(\hat{\theta}_n(\mathbf{X}_n) - \theta(P))$  の分位点  $c_\alpha$  を求めたい。

定義 4.8.3.  $\hat{P}_n$  を  $P$  の推定量とする。

- (1) 真の分布関数  $H_n(x; P)$  に対する近似  $H_n(x; \hat{P}_n) := \int_{\mathbb{R}^n} 1_{R_n(\mathbf{x}, \hat{P}_n) \leq x} (\hat{P}_n)^n(d\mathbf{x})$  を、ブートストラップ分布という。
- (2) 推定分布  $\hat{P}_n$  に従う大きさ  $n$  の無作為標本  $\mathbf{X}^* = (X_1^*, \dots, X_n^*)$  をブートストラップ標本という。

注 4.8.4 (computer-intensive). 推定分布  $\hat{P}_n$  が複雑であるとき、積分  $H_n(x; \hat{P}_n)$  はモンテカルロ法で近似する。これが、ブートストラップ法が計算機集約的と言われる所以である。

注 4.8.5 (studentization).  $R_n(\mathbf{X}_n, P)$  をステューデント化された確率変数で置き換えることで、ブートストラップ分布の近似精度の改善が可能である。

## 参考文献

- [1] 吉田朋広『数理統計学』(朝倉書店, 2006)
- [2] 竹村彰道『現代数理統計学』(学術図書, 2020)
- [3] 久保川達也『現代数理統計学の基礎』(共立出版, 2017)
- [4] 西山陽一『マルチンゲール理論による統計解析』(近代科学社, 2011)
- [5] Rabi Bhattacharya - Course in Mathematical Statistics and Large Sample Theory
- [6] Ibragimov and Has'minskii - Statistical Estimation