

目次

第 1 章	線型推測論	4
1.1	独立標準正規分布が定める 2 次形式の分布	4
1.1.1	Fisher-Cochran の定理	4
1.1.2	二次形式確率変数	5
1.1.3	二次形式確率変数の構成	5
1.2	統計的推測の枠組み	5
1.2.1	不偏推定量と推定可能関数	5
1.2.2	信頼区間	6
1.3	Gauss-Markov モデル	6
1.3.1	正規方程式の解	7
1.3.2	不偏推定量となるための条件	7
1.3.3	誤差の分散の推定	8
1.4	正規性を仮定した Gauss-Markov モデルによる仮説検定	8
1.4.1	検定統計量の分布	9
1.4.2	母数に線型制約がある場合の検定統計量の分布	9
1.4.3	母数に線型制約がある場合の検定統計量の表示	9
1.4.4	検定の構成	9
1.5	切片項を持つ正規誤差 Gauss-Markov モデル	10
1.5.1	正規方程式の解	10
1.5.2	正規誤差仮定の下での検定の構成	11
1.5.3	正規誤差仮定の下での予測・信頼区間の構成	12
1.5.4	正規誤差について	12
1.6	その他の分散分析の例	13
1.6.1	一元配置	13
1.6.2	二元配置	14
1.7	主成分分析と因子分析	14
第 2 章	統計的決定理論	15
2.1	統計的決定問題	15
2.1.1	Wald の枠組み	15
2.1.2	最良な決定関数	15
2.1.3	不偏推定問題	16
2.1.4	Bayes 推定問題	18
2.1.5	一様最強力検定問題	18
2.1.6	不偏検定問題	20
2.1.7	不変検定問題	20
2.1.8	区間推定問題	21
2.2	統計量の基礎理論	21
2.2.1	十分統計量の定義と例	21

2.2.2	最小十分統計量	24
2.2.3	十分統計量の下での条件付き期待値	24
2.2.4	十分統計量の密度の型による特徴付け：因子分解定理	24
2.2.5	十分統計量による決定関数の一様改善：Rao-Blackwell の定理	25
2.2.6	完備統計量の定義と例	26
2.2.7	不偏推定量の一様最小分散性の特徴付け	28
2.2.8	Basu の定理	28
2.3	推定論	29
2.3.1	一般の最良不偏推定量	29
2.3.2	正則な統計的実験	30
2.3.3	正則なモデルでの不偏推定量の分散行列の下界	31
2.3.4	ベイズ推定	32
2.3.5	James-Stein 推定量	32
2.4	検定論	33
2.4.1	単純仮説同士の検定における尤度比検定の特徴付け	33
2.4.2	複合仮説同士の検定と単調尤度比	33
2.4.3	一般化された Neyman-Pearson の補題	34
2.4.4	不偏検定の特徴付け	34
2.4.5	帰無仮説の補集合となる対立仮説の検定	35
2.4.6	両側 t -検定	35
2.4.7	不変検定	36
第 3 章	種々の統計手法とその漸近論的解釈	37
3.1	モーメント法による母数推定	37
3.1.1	標本平均	37
3.1.2	標本積率	38
3.1.3	一般化モーメント法	38
3.2	χ^2 -検定	38
3.2.1	概要	38
3.2.2	Yates の補正	39
3.2.3	G-検定	39
3.2.4	Fisher の正確検定	39
3.2.5	種々の標本統計量	39
3.3	順位検定	39
3.3.1	順序統計量の分布	39
3.3.2	順位相関係数	40
第 4 章	種々の模型の統計化	41
4.1	多変量に対する模型と変数選択	41
4.1.1	回帰模型の概観	41
4.1.2	その後の研究の歴史	42
4.1.3	主成分抽出の手法	42
4.1.4	因子分析と正準相関分析	42
4.1.5	発展：独立成分分析	43
4.1.6	共通成分抽出の手法	43
4.1.7	個体類別の手法	45
4.1.8	識別可能性の問題	45
4.2	判別分析と機械学習	46

4.3	線型回帰模型	46
4.3.1	標準線型回帰模型に対する Gauss-Markov の定理	47
4.3.2	Rao-Blackwell の定理	47
4.3.3	線型正規回帰模型	48
4.3.4	線型漸近正規回帰模型	48
4.3.5	半径数的線型回帰モデル	48
4.3.6	自己回帰モデル	49
4.4	非線型回帰模型	49
4.4.1	Probit 回帰模型	49
4.5	回帰模型の検定	50
4.5.1	誤差分散の推定と検定	50
4.5.2	推定量の分布	50
4.6	回帰模型の診断	50
4.6.1	決定係数	50
4.6.2	AIC	50
4.6.3	一般化最小二乗法	50
4.6.4	Bayes 回帰	51
第 5 章	参考文献	52
参考文献	53
参考文献	54

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.

Gauss-Markov モデル $Y = \beta X + \epsilon$ は次に従う [竹内啓, 2016]:

- (1) 分布の仮定を置かないとき, 正規方程式の解 $\hat{\beta}$ が定める線型汎函数が, 線型不偏推定量の中で最良である.
- (2) 分布に正規性の仮定をおけば, 一次元のとき, 不偏推定量の中で最良のものが存在するが, $n \geq 2$ のときは存在しない (正規分布族が完備でないため). しかし F -検定は一樣最強力な不変検定である.
- (3) 一般には不偏推定量で最良のものは見つからないため, Pitman 推定量, または, 漸近最適な不偏推定量である最尤推定量を用いる. なお, 正規性の仮定を置いたとき, OLS 推定量が最尤推定量である.

記法 1.0.1.

- (1) $M_n(\mathbb{R})$ は転置 $^\top$ を $*$ -準同型として, $*$ -代数の構造を持つ.
- (2) しかし $^\top$ は $*$ -準同型 (反線形で反変的な対合群準同型 $M_n(\mathbb{R}) \rightarrow M_n(\mathbb{R})$) であるだけでなく, $\text{Tr}(A^\top B) = (A|B)$ として内積も定める.
- (3) すると, 標準的な内積空間 \mathbb{R}^n はこの (退化した) 例と見れる. そこで, $c^\top \beta$ などの代わりに $(c|\beta)$ とも表し得る.
- (4) \mathbb{R}^k における標準基底の和を 1_k で表す.

$$1_k := \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^k$$

1.1 独立標準正規分布が定める 2 次形式の分布

1.1.1 Fisher-Cochran の定理

定理 1.1.1. $\{A_i\}_{i \in [k]} \subset M_n(\mathbb{R})$ は I_n の分解とする. このとき, 次の 4 条件は同値:

- (1) 各 A_i は射影である.
- (2) $\forall i, j \in [k] \ i \neq j \Rightarrow A_i A_j = O$.
- (3) $\mathbb{R}^n = \bigoplus_{i \in [k]} \text{Im } A_i$.
- (4) $\sum_{i \in [k]} \text{rank } A_i = n$.

定理 1.1.2 (Fisher-Cochran). $\{A_i\}_{i \in [k]} \subset M_n(\mathbb{R})$ は対称行列からなる I_n の分解, $Y \sim N_n(\mu, I_n)$ ($\mu \in \mathbb{R}^n$) とする. 係数行列

A_i の定める二次形式 $Q_i := Y^\top A_i Y$, $n_i := \text{rank } A_i$ について, 次の5条件は同値:

- (1) $\exists \delta_i \in \mathbb{R}_+ \quad Q_i \sim \chi^2(n_i, \delta_i)$ かつ Q_1, \dots, Q_k は独立.
- (2) $\sum_{i \in [k]} n_i = n$.
- (3) 各 A_i は射影である.
- (4) $\forall i, j \in [k] \quad i \neq j \Rightarrow A_i A_j = O$.
- (5) $\mathbb{R}^n = \oplus_{i \in [k]} \text{Im } A_i$.

このとき, $\delta_i = \mu^* A_i \mu$ が成り立ち, 特に $\sum_{j \in [n]} \mu_j^2 = \sum_{i \in [k]} \delta_i$ が成り立つ.

1.1.2 二次形式確率変数

系 1.1.3 (二次形式が χ^2 -分布に従う必要十分条件). 対称行列 $A \in M_n(\mathbb{R})$ について, 次の2条件は同値:

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

このとき, $k = \text{rank } A$, $\delta = \mu^* A \mu$ として得られる.

系 1.1.4 (χ^2 -分布に従う二次形式が独立であることの特徴付け). 対称行列 $A_1, A_2 \in M_n(\mathbb{R})$ について, $Q_i = Y^\top A_i Y$ は非心 χ^2 -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) $A_1 A_2 = O$.

1.1.3 二次形式確率変数の構成

χ^2 -二次形式確率変数の差が再び χ^2 -分布に従う十分条件を求める.

補題 1.1.5. 対称な射影行列 $A, B \in M_n(\mathbb{R})$ について, $A - B \geq O$ ならば, 次の2条件が成り立つ.

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

定理 1.1.6. $A_i \in M_n(\mathbb{R})$ ($i \in 3$) を $A_0 = A_1 + A_2$ を満たす対称行列, $Q_i := Y^\top A_i Y$ とする. $Q_0 \sim \chi^2(n_0, \delta_0)$, $Q_1 \sim \chi^2(n_1, \delta_1)$, $Q_2 \geq 0$ a.s. のとき, 次が成り立つ:

- (1) $n_2 := n_0 - n_1$, $\delta_2 := \delta_0 - \delta_1$ について, $Q_2 \sim \chi^2(n_2, \delta_2)$.
- (2) $Q_1 \perp\!\!\!\perp Q_2$.

1.2 統計的推測の枠組み

1.2.1 不偏推定量と推定可能関数

推定可能な汎関数のことを, 統計的汎関数ともいう.

定義 1.2.1 (unbiased estimator).

- (1) 統計的関数 $g(\beta, \sigma^2) : \mathbb{R}^p \times \mathbb{R}_+ \rightarrow \mathbb{R}$ の推定量 $\delta(Y) : \mathbb{R}^n \rightarrow \mathbb{R}$ が不偏であるとは,

$$\forall (\beta, \sigma^2) \in \mathbb{R}^p \times \mathbb{R}_+ \quad E_{\beta, \sigma^2}[\delta(Y)] = g(\beta, \sigma^2)$$

を満たすことをいう.

(2) モデル $\{P_\theta\}_{\theta \in \Theta} \subset P(\mathcal{X})$ と写像 $g: \Theta \rightarrow \mathbb{R}^k$ について、推定量 $\delta: \mathcal{X} \rightarrow \mathbb{R}^k$ が

$$\forall \theta \in \Theta \quad \int_{\mathcal{X}} \delta(x) P_\theta(dx) = g(\theta)$$

を満たすとき、**不偏**であるという。

定義 1.2.2 (estimable linear functional, statistical functional).

- (1) $\beta \in \mathbb{R}^p$ 上の線型汎関数 $c^\top: \mathbb{R}^p \rightarrow \mathbb{R}$ ($c \in \mathbb{R}^p$) が (**U**-) **推定可能**であるとは、 $c \in \text{Im}(X^\top) = (\text{Ker } X)^\perp$ を満たすことをいう。
- (2) 一般のモデル $\mathcal{P} \subset P(\mathbb{R})$ において**推定可能**な汎関数とは、分布の関数 $P(\mathbb{R}) \rightarrow \mathbb{R}$ とみなせる関数をいう。これを**統計的汎関数**ともいう。
- (3) 一般のモデル $\mathcal{P} \subset P(\mathbb{R})$ において **U-推定可能**な汎関数とは、これに対する不偏推定量が存在することをいう。

1.2.2 信頼区間

予測区間は、将来の観測に対して構成する観測の確率変数、信頼区間は、母数の推定のために構成する観測の確率変数である。確率変数と言っても、実数値ではなく、閉区間値であるが。

定義 1.2.3 (prediction interval, confidence interval).

- (1) 将来の観測 Y_{n+1} について予測を行うために構成された、観測の空間 \mathcal{X}^n 上の閉区間値の確率変数を**予測区間**という。
- (2) モデルの母数を推定するために構成された、パラメータ $g: \mathcal{P} \rightarrow \mathbb{R}^p$ に関する信頼係数 $1 - \alpha$ ($\alpha \in (0, 1)$) の**信頼区間**とは、観測の確率変数の組 $(u, v): \mathcal{X}^n \rightarrow \mathbb{R}^2$ であって、任意の真の分布 $P \in \mathcal{P}$ に対して $P[g(P) \in [u, v]] \geq 1 - \alpha$ を満たす確率区間 $I(Y_1, \dots, Y_n) = [u(Y_1, \dots, Y_n), v(Y_1, \dots, Y_n)] \subset \mathbb{R}$ をいう。

要諦 1.2.4 (Bayesian credible interval). 頻度主義では、測定を繰り返した際にノイズによって変動するのは信頼区間の方であって、あくまでも真値は決定されていると考える。もし「 $1 - \alpha$ の確率で、真の母数は、ある実現値 $I(Y_1, \dots, Y_n)$ のどこにある」という知識がほしい場合は、ベイズ主義の枠組み信頼区間を構成し直す必要がある。頻度主義的にいえば、任意の実現値 $I(Y_1, \dots, Y_n)$ に対して、真の母数がこれに含まれる「確率」という語用は意味をなさない。

1.3 Gauss-Markov 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 ϵ について、 $E[\epsilon] = 0$ かつ $\text{Var}[\epsilon] = \sigma^2 I_n$ なる平均-分散構造の仮定 (Gauss-Markov assumption) をおいた線型回帰モデルを **Gauss-Markov モデル**という。したがって、推定量は観測の 2 次以下の多項式で構成するしかない。特に、2 次形式を用いて構成された検定量を分散分析という。このとき、

- (1) β の最良線型不偏推定量は、正規方程式の任意の解が与える。これは M -推定量の例でもある。
- (2) これは最小二乗誤差の最小化による M -推定量とみなせ、OLS estimator とも呼ばれる。

模型 1.3.1 (Gauss-Markov モデル). n 種の観測 $Y_i \in \mathbb{R}$ を、それぞれに対応した要因 $x_i \in \mathbb{R}^p$ で説明することを考える： $Y_i = x_i^\top \beta + \epsilon_i$ 。誤差項は中心性 $E[\epsilon_i] = 0$ と無相関性 $E[\epsilon_i \epsilon_j] = \sigma^2 \delta_{ij}$ を仮定する。すると仮定を $Y = X\beta + \epsilon$ 、ただし、

$$Y := \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \in \mathbb{R}^n, X := \begin{bmatrix} x_1^\top \\ \vdots \\ x_n^\top \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} \in M_{np}(\mathbb{R}), \epsilon := \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}.$$

とした。このとき、 $\epsilon \sim (0, \sigma^2 I_n)$ であり、モデルの母数は $(\beta, \sigma^2) \in \mathbb{R}^p \times \mathbb{R}_+$ となる。

要諦 1.3.2.

- (1) デザイン X は根源的な要因の非線形関数とすれば良いので、 Y の X に関する線形性は応用上は強い制約ではなく、ある種の第一近似とみれる。 X は**計画行列** (design matrix) という。
- (2) モデルの母数は (β, σ^2) である。特に、3 次以上の積率について過程を置いていないので、 $E_{\beta, \sigma^2}[Y_1^3]$ などは定まらない。したがって、推定量は Y の 2 次以下の多項式で構成する必要がある。
- (3) 計量経済学では、さらに仮定 $E[\epsilon_i|X] = 0$ 、すなわち、 $\forall_{i,j \in [n]} E[x_j \cdot \epsilon_i] = 0$ をおくことがあり、これを**狭義外生性** (strict exogeneity) という。逆に、説明変数 x_i と誤差 ϵ_i の間に相関があるとき、**内生的** (endogeneity) であるという。これは正しい説明変数を選び切れていないとも考えられるため、操作変数法によって解決され得る。
- (4) 残差について置かれた仮定 $\sigma^2 \in \mathbb{R}_+$ は、分散の均一性 (Homoscedasticity / homogeneity of variance) と呼ばれる。ギリシャ語 skedastikos が「分散」を意味する形容詞。球状誤差 (spherical error) ともいう。誤差が「自己相関」を持つとき、この仮定は破られ、これは時系列データでよく見られる。

歴史 1.3.3. 定理 1.3.9 に当たる内容を、Gauss が最初に、正規性と独立性の仮定の下に導いた。これを後に、Markov が上のより一般的なモデルについて一般化した。

手法 1.3.4 (最小二乗法 (OLS)). β のパラメータ空間 \mathbb{R}^p 上の二次形式 $Q: \mathbb{R}^p \rightarrow \mathbb{R}$

$$Q(\beta) := \|Y - X\beta\|_2^2 = (Y - X\beta)^\top (Y - X\beta) = \sum_{i \in [n]} (Y_i - x_i^\top \beta)^2$$

を最小化する $\beta \in \mathbb{R}^p$ を推定量とする。 Q を二次剰余和 (SSR: Sum of Squared Residuals) ともいう。停留点であるための条件

$$X^\top X\beta = X^\top Y$$

は必要条件となり、これを**正規方程式** (normal equation) という。 $\text{Im}(X^\top X) = \text{Im}(X^\top)$ より、これは常に解 $\hat{\beta}$ を持つが、一意性を持つのは $X^\top X$ が正則である場合に限る。このとき、 $\text{Var}[\hat{\beta}] = \sigma^2(X^\top X)^{-1}$ である。実は、正規方程式の任意の解が最小二乗推定量になり 1.3.6、さらに最良の線型不偏推定量でもある 1.3.9。この解空間を過程とみると、共分散は表せる 1.3.12。

要諦 1.3.5 (一般化最小二乗法 (GLS)). Gauss-Markov モデルでは、分散の均一性を仮定しているが、これを外し、 ϵ の行分散行列がスカラーでない場合でも、同様に BLUE である推定量 Aitken 推定量を構成できる (Aitken 1935)。このように、残差 $\epsilon \in L^2(\Omega)$ にこれ以上の分布の仮定を置かないものを**一般化線型モデル (GLM)** (Nelder and Wedderburn 1972)、多変量正規分布という仮定を置いたものを**一般線形モデル** (general linear model) という。一般化線型モデルでは、 Y の分布は指数型になる。

1.3.1 正規方程式の解

最小二乗法により β を推定してから、 σ^2 を推定することを考える。

定理 1.3.6. 任意の $Y \in \mathbb{R}^n$ に対する正規方程式の任意の解 $\hat{\beta} \in \mathbb{R}^p$ について、次が成り立つ。

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

定義 1.3.7.

- (1) $\hat{Y} := X\hat{\beta} = P_X Y$ は $\hat{\beta}$ の取り方によらずに定まる。
- (2) $e := Y - \hat{Y} \in \mathbb{R}^n$ を**残差ベクトル**という。

1.3.2 不偏推定量となるための条件

統計的関数 $g(\beta): \mathbb{R}^p \rightarrow \mathbb{R}$ が線型汎関数 $g = c^\top$ ($c \in \mathbb{R}^p$) である場合を考える。このとき、正規方程式の任意の解が、任意の線型な不偏推定量の中で分散が最小のものになる。

定理 1.3.8 (線型な不偏推定量の性質)。

- (1) 線型関数 $Y \mapsto a^\top Y$ ($a \in \mathbb{R}^n$) が $g(\beta)$ の不偏推定量であることは、 $X^\top a = c$ に同値。
 (2) $g(\beta)$ の線型な不偏推定量が存在するための必要十分条件は、 $g(\beta)$ が推定可能であることである： $c \in \text{Im}(X^\top) = (\text{Ker } X)^\perp$ 。

定理 1.3.9 (正規方程式の解は BLUE である (Gauss-Markov)). $g(\beta)$ は推定可能とする。このとき、次が成り立つ：

- (1) ある $b \in \mathbb{R}^p$ を用いて、 $c = X^\top X b, c^\top \hat{\beta} = b^\top X^\top Y$ と表せる。
 (2) $c^\top \hat{\beta}$ は正規方程式の解 $\hat{\beta}$ の取り方に依らず、 $g(\beta)$ の不偏推定量となる。
 (3) 正規方程式の任意の解 $\hat{\beta}$ について、 $c^\top \hat{\beta}$ が**最良線型不偏推定量 (BLUE)** である。すなわち、任意の $(\beta, \sigma^2) \in \mathbb{R}^p \times \mathbb{R}_+$ について、

$$\text{Var}[c^\top \hat{\beta}] = \min \{ \text{Var}[a^\top Y] \mid a^\top Y \text{ は不偏推定量}, a \in \mathbb{R}^n \}$$

が成り立つ。ただし、 Var は $P_{(\beta, \sigma^2)}$ について考える。

要諦 1.3.10 (Stein's paradox). バイアスがある推定量を許すと、より分散の小さい推定量が存在する (James-Stein 推定量など)。3つ以上の母数を同時に推定する際は、それぞれの母数を個別に最適に推定する推定量の組よりも、より最小二乗誤差が小さいものが存在する。これは「たしからしさは得られた観測に引っ張られる」「あえてバイアスを入れることで効率性は改善する」という Bayes 統計的な立場から構成された統計量だともみなせる。同様な手法でリッジ回帰の「0に向けたひっぱり方」も、Lasso 回帰に改善される。

系 1.3.11 . $a \in \mathbb{R}^n, g(\beta) = c^\top \beta$ ($c \in \mathbb{R}^p$) について、次の3条件は同値：

- (1) 線型統計量 $a^\top Y$ は $g(\beta)$ の BLUE である。
 (2) $c^\top \beta$ は推定可能で、任意の真値 $(\beta, \sigma^2) \in \mathbb{R}^p \times \mathbb{R}_+$ について $c^\top \hat{\beta} = a^\top Y$ a.s.
 (3) $X^\top a = c$ かつ $a \in \text{Im}(X)$ 。

命題 1.3.12 (最小二乗推定量系の共分散).

- (1) $c_i \in \text{Im}(X^\top)$ のとき、 $c_1^\top (X^\top X)^- c_2$ は一般可逆行列 $(X^\top X)^-$ の取り方によらない。
 (2) $c_i^\top \beta$ が推定可能ならば、それぞれの最良線型不偏推定量 $c_i^\top \hat{\beta}$ の共分散は、 $\text{Cov}_{(\beta, \sigma^2)}[c_1^\top \hat{\beta}, c_2^\top \hat{\beta}] = \sigma^2 c_1^\top (X^\top X)^- c_2$ 。

1.3.3 誤差の分散の推定

残差 $\|Y - X\beta\|_2^2$ の最小値 R_0^2 をスケールすることで、 σ^2 の不偏推定量が構成出来る。

命題 1.3.13 (分散の 2step 最小二乗推定量の表現). $e = Y - \hat{Y}$ を残差ベクトルとする。

- (1) $e = (I_n - P_X)Y = (I_n - P_X)\epsilon$ と表せる。特に、 $e \sim (0, \sigma^2(I_n - P_X))$ 。
 (2) $g(\beta) = c^\top \beta$ ($c \in \mathbb{R}^p$) は推定可能とする。このとき、 $\text{Cov}[c^\top \hat{\beta}, e] = 0$ 。
 (3) $R_0^2 := \min_{\beta \in \mathbb{R}^p} |Y - X\beta|^2$ は $R_0^2 = e^\top e$ と表せる。

定理 1.3.14 (不偏推定量の構成). 任意の $(\beta, \sigma^2) \in \mathbb{R}^p \times \mathbb{R}_+$ について、

- (1) $E_{\beta, \sigma^2}[R_0^2] = \sigma^2(n - r)$ 。ただし、 $r := \text{rank } X$ 。
 (2) 特に、 $n > r$ のとき、 $\hat{\sigma}^2 := \frac{R_0^2}{n - r}$ と定めれば、これは σ^2 の不偏推定量である。

1.4 正規性を仮定した Gauss-Markov モデルによる仮説検定

Gauss-Markov モデルに対して、さらに誤差分布の正規性 $\epsilon \sim N_n(0, \sigma^2 I_n)$ を仮定した下で、パラメータ空間 Θ 上の検定を構成することを考える。正則性の仮定は、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。

1.4.1 検定統計量の分布

残差 e による変動は、仮定に依らずに $\chi^2(n-r)$ に従う。

定理 1.4.1. $R_0^2 := \min_{\beta \in \mathbb{R}^p} |Y - X\beta|^2$, $r := \text{rank } X < n$ とする. このとき, $\frac{R_0^2}{\sigma^2}$ は真値 $(\beta, \sigma^2) \in \mathbb{R}^p \times \mathbb{R}_+$ の下では $\chi^2(n-r)$ に従う.

1.4.2 母数に線型制約がある場合の検定統計量の分布

仮説による変動 $R_1^2 - R_0^2$ は $\chi^2(n-r)$ に従う. このとき, 2つの変動の比は F -分布に従う.

記法 1.4.2.

- (1) $H \in M_{pq}(\mathbb{R})$, $q < p$ は, $\text{Im}(H) \subset \text{Im}(X^\top) = (\text{Ker } X)^\perp$ かつ $\text{rank } H = q$ を満たすとする.
- (2) ある $d \in \mathbb{R}^q$ について, 超平面を $\Theta_0 := \{\beta \in \mathbb{R}^p \mid H^\top \beta = d\}$ とする.
- (3) この範囲での最小残差を $R_1^2 := \min_{\beta \in \Theta_0} |Y - X\beta|^2$ とする.

定理 1.4.3.

- (1) 任意の真値 $(\beta, \sigma^2) \in \mathbb{R}^p \times \mathbb{R}_+$ について,
 - (a) $R_0^2 \perp R_1^2 - R_0^2$.
 - (b) $\exists \delta \in \mathbb{R}_+ \frac{R_1^2 - R_0^2}{\sigma^2} \sim \chi^2(q, \delta)$.
 - (c) 特に, $\frac{(R_1^2 - R_0^2)/q}{R_0^2/(n-r)} \sim F(q, n-r, \delta)$.
- (2) 真値 $(\beta, \sigma^2) \in \Theta_0 \times \mathbb{R}_+$ について,
 - (a) $\frac{R_1^2 - R_0^2}{\sigma^2} \sim \chi^2(q)$.
 - (b) 特に, $\frac{(R_1^2 - R_0^2)/q}{R_0^2/(n-r)} \sim F(q, n-r)$.

1.4.3 母数に線型制約がある場合の検定統計量の表示

命題 1.4.4. 正規方程式の任意の解 $\hat{\beta}$ について,

$$R_1^2 - R_0^2 = (H^\top \hat{\beta} - d)^\top (H^\top (X^\top X)^{-1} H)^{-1} (H^\top \hat{\beta} - d).$$

1.4.4 検定の構成

問題 1.4.5 (平均ベクトルの affine 集合への所属の決定問題). affine 超平面 $\Theta_0 \subset \mathbb{R}^p$ に関する検定問題 $H_0: \beta \in \Theta_0$ vs. $H_1: \beta \in \mathbb{R}^p \setminus \Theta_0$ を考える.

検定 1.4.6 (F -検定).

$$F := \frac{(R_1^2 - R_0^2)/q}{R_0^2/(n-r)}$$

はデータから計算出来る. この値が大きいくに H_0 を棄却する検定が考えられる. F は H_1 に対しては H_0 の下よりも大きな値が出やすい傾向がある. 棄却域の設定には, F の分布である F -分布を用いることになる 1.4.3.

要諦 1.4.7 (ANOVA: Analysis Of Variance). これは分散分析の例である. 一般に, データのなすベクトル $Y \in \mathbb{R}^n$ の二次形式 R_0^2, R_1^2 を用いて, データの変動を \mathbb{R}^p 上大域的における場合と Θ_0 に制限した場合のものに分解し, これを基に商として検定量を構成する手法を分散分析という. 今回は次のように分解した:

- (1) $R_1^2 - R_0^2$: 仮説による変動. 自由度は q .
 (2) R_0^2 : 残差による変動で, 仮説には依らずに $\sim \chi^2(n-r)$.

問題 1.4.8. $(Y_j)_{j \in [n]}$ を $N(\mu, \sigma^2)$ からの無作為標本とし, $H_0: \mu = \mu_0$ を検定する. これは Gauss-Markov モデルの退化した例で, $X = (1, \dots, 1)^\top, \beta = \mu, H = 1, d = \mu_0$ としたものである.

検定 1.4.9 (両側 t -検定). 付随する F -検定は

$$F = \frac{n(\bar{Y} - \mu_0)^2}{\frac{\sum_{j \in [n]} (Y_j - \bar{Y})^2}{n-1}}$$

となる. なお, 同値な検定として

$$T := \sqrt{F} = \frac{\sqrt{n}(\bar{Y} - \mu_0)}{\sqrt{\frac{\sum_{j \in [n]} (Y_j - \bar{Y})^2}{n-1}}}$$

の絶対値が大きいときに H_0 を棄却することが考えられる. これは $T \sim t(n-1)$ under H_0 だから, 両側 t -検定という.

問題 1.4.10. $(Y_{1j})_{j \in [n_1]}, (Y_{2j})_{j \in [n_2]}$ をそれぞれ $N(\mu_1, \sigma^2), N(\mu_2, \sigma^2)$ からの独立な標本とする. $H_0: \mu_1 = \mu_2$ を検定する.

検定 1.4.11. Gauss-Markov モデルによる F 検定量と, 同値な t -検定量

$$T = \frac{\frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{1/n_1 + 1/n_2}}}{\sqrt{\frac{\sum_{j \in [n_1]} (Y_{1j} - \bar{Y}_1)^2 + \sum_{j \in [n_2]} (Y_{2j} - \bar{Y}_2)^2}{n_1 + n_2 - 2}}}$$

が得られる. $T \sim t(n_1 + n_2 - 2)$ under H_0 である.

1.5 切片項を持つ正規誤差 Gauss-Markov モデル

すなわち, 線型重回帰モデルのことをいう

1つの要因の線型関係ではなく, 複数の要因からの多重線型汎関数として説明したいときの Gauss-Markov モデルを, **重回帰分析**という. 重回帰分析においては, 切片項 β_0 を含む **Gauss-Markov モデル**が用いられる. これは, $\beta \in \mathbb{R}^p$ の各成分をそれぞれの観測 Y_i 毎のパラメータとしていたのに対して, 単一パラメータ $\beta \in \mathbb{R}^p$ の反復観測と考えられる設定である.

模型 1.5.1 (線型重回帰モデル). n 回の独立な観測 $Y_i \in \mathbb{R}^n$ を, $x_{i1}, \dots, x_{ip} \in \mathbb{R}$ と定数 $\beta_0 \in \mathbb{R}$ によって説明することを考える:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \epsilon_i \quad (i \in [n])$$

これは, $Y = X\beta + \epsilon$, ただし,

$$\beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_p \end{bmatrix} \in \mathbb{R}^{p+1}, Y, \epsilon \in \mathbb{R}^n, X_1 := (x_{ia}) \in M_{np}(\mathbb{R}), X := \begin{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} & X_1 \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \ddots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix} \in M_{n, (p+1)}(\mathbb{R})$$

とした. さらに $\epsilon \sim (0, \sigma^2 I_n)$ を仮定すると, モデルの母数は $(\beta, \sigma^2) \in \mathbb{R}^p \times \mathbb{R}_+$ からなる. $p = 1$ のときを**単回帰モデル** (straight line regression) という.

1.5.1 正規方程式の解

命題 1.5.2. $\hat{\beta} := (\beta_1, \dots, \beta_p)^\top$ と表す.

- (1) Gram 行列 $X^\top X \in M_{p+1}(\mathbb{R})$ は

$$X^\top X = \begin{bmatrix} n & 1^\top X_1 \\ X_1^\top 1 & X_1^\top X_1 \end{bmatrix}$$

(2) 正規方程式 $X^\top X\beta = X^\top Y$ は係数を

$$s_{ab} := \frac{1}{n} \sum_{i \in [n]} (x_{ia} - \bar{x}_a)(x_{ib} - \bar{x}_b) \quad (a, b \in [p]), \quad s_{ya} := \frac{1}{n} \sum_{i \in [n]} (Y_i - \bar{Y})(x_{ia} - \bar{x}_a) \quad (a \in [p]).$$

とおくと、次に同値：

$$\begin{cases} \sum_{b \in [p]} s_{ab} \beta_b = s_{ya} & a \in [p] \\ \beta_0 = \bar{Y} - \sum_{a \in [p]} \bar{x}_a \beta_a \end{cases}$$

(3) $s := (s_{ab}) \in \text{GL}_p(\mathbb{R})$ と仮定し、逆行列を $s^{-1} = (s^{ab}) \in \text{GL}_p(\mathbb{R})$ で表す。このとき、正規方程式のただ一つの解 $\hat{\beta}$ は次のように表せる：

$$\begin{cases} \hat{\beta}_a = \sum_{b \in [p]} s^{ab} s_{yb} & a \in [p] \\ \hat{\beta}_0 = \bar{Y} - \sum_{a \in [p]} \bar{x}_a \hat{\beta}_a \end{cases}$$

(4) $\hat{\sigma}^2 = \frac{R_0^2}{n-p-1}$ は σ^2 の不偏推定量である。

1.5.2 正規誤差仮定の下での検定の構成

$\epsilon \sim N_n(0, \sigma^2 I_n)$ を仮定する。

命題 1.5.3.

$$(X^\top X)^{-1} = \frac{1}{n} \begin{bmatrix} 1 + \frac{1}{n} 1^\top X_1 s^{-1} \frac{1}{n} X_1^\top 1 & \frac{1}{n} 1^\top X_1 s^{-1} \\ -s^{-1} \frac{1}{n} X_1^\top 1 & s^{-1} \end{bmatrix}$$

問題 1.5.4 (特定の独立変数の有意性検定). 任意の $a \in [p]$ に関して、 $H_0 : \beta_a = 0$ を検定する。すなわち、変量 x_a が y の説明に必要なかどうかを決定する。

検定 1.5.5. a に対応する成分のみ 1 のベクトル $H = (0 \cdots 0 1 0 \cdots 0)^\top \in \mathbb{R}^{p+1}$ と $d = 0$ について Gauss-Markov モデルの F -検定を考えると、 $R_1^2 - R_0^2 = \frac{\hat{\beta}_a^2}{s^{aa}/n}$ より、定理 1.4.3 より、

$$\frac{\hat{\beta}_a^2}{\hat{\sigma}^2 s^{aa}/n} \sim F(1, n-p-1) \text{ under } H_0.$$

一方で、 H_1 の下では非心 F -分布に従い、値が大きくなる傾向がある。

問題 1.5.6 (定数項の優位性検定). $H_0 : \beta_0 = 0$ を検定する。定数項が必要かどうかの決定である。

検定 1.5.7. $H = (1 0 \cdots 0) \in \mathbb{R}^{p+1}$, $d = 0$ についての F -検定を考えると、

$$\frac{\hat{\beta}_0^2}{\hat{\sigma}^2 \left(1 + \sum_{a,b \in [p]} \bar{x}_a \bar{x}_b s^{ab}\right)/n} \sim F(1, n-p-1) \text{ under } H_0.$$

問題 1.5.8 (回帰モデル全体の優位性検定). $H_0 : \beta_1 = \cdots = \beta_p = 0$ を検定する。

検定 1.5.9 (回帰モデルの適合度検定量 (の例) を決定係数という). $H^\top := [O \ I_p] \in M_{p,p+1}(\mathbb{R})$, $d = 0$ についての F -検定を考えると、

$$F = \frac{|X\hat{\beta} - \bar{Y}1|^2/p}{\hat{\sigma}^2} \sim F(p, n-p-1) \text{ under } H_0.$$

なお、

$$R^2 := \frac{abs X\hat{\beta} - \bar{Y}1^2}{|Y - \bar{Y}1|^2}$$

を決定係数といい、データへの回帰モデルの fitting の度合いを表すとされる。

1.5.3 正規誤差仮定の下での予測・信頼区間の構成

真の分布 $P \in \mathcal{P}$ に対して $P[g(P) \in I] \geq 1 - \alpha$ を満たす閉区間 $I \subset \mathbb{R}$ を、信頼係数 $1 - \alpha$ ($\alpha \in (0, 1)$) の信頼区間という。頻度主義では、測定を繰り返した際にノイズによって変動するのは信頼区間の方であって、あくまでも真値は決定されていると考える。もし「 $1 - \alpha$ の確率で真の分布は I のどこかにある」という知識がほしい場合は、ベイズ主義の枠組み信頼区間を構成し直す必要がある。

問題 1.5.10. 観測 Y_1, \dots, Y_n から、説明変数を $(x_1, \dots, x_p) = (\xi_1, \dots, \xi_p) =: \xi \in \mathbb{R}^p$ に設定したとする。

(1) 将来の観測

$$Y_\xi := \beta_0 + \beta_1 \xi_1 + \dots + \beta_p \xi_p + \epsilon_*$$

を予測し、予測区間を構成せよ。

(2) 母数 μ_ϵ に対する信頼区間を構成せよ。

解法 1.5.11. $c_\xi := (1 \ \xi^\top)^\top$ が定める線型汎関数 $g: \mathbb{R}^{p+1} \rightarrow \mathbb{R}$ に、正規方程式の解を代入したもの $\hat{Y}_\xi := g(\hat{\beta}) = c_\xi^\top \hat{\beta}$ が与える。

(1) この推定量の分布は

$$\hat{Y}_\epsilon \sim N(\mu_\xi, \sigma^2 c_\xi^\top (X^\top X)^{-1} c_\xi) \text{ under } N_n(\beta, \sigma^2).$$

(2) $\hat{Y}_\epsilon \perp \hat{\sigma}^2 = \frac{\mathbf{e}^\top \mathbf{e}}{n - p - 1}$ で、さらにこの2つと $Y_\xi \sim N(\mu_\xi, \sigma^2)$ は独立。

(3) 次が成り立つ：

$$\frac{Y_\xi - \hat{Y}_\xi}{\sqrt{\hat{\sigma}^2 \left(1 + \frac{1}{n} (1 + (\xi - \bar{x})^\top s^{-1} (\xi - \bar{x}))\right)}} \sim t(n - p - 1).$$

(4) これを用いて、信頼係数 $1 - \alpha$ に対応した予測域を構成できる。 t -分布の上側 $100\alpha/2\%$ 点を $t_\alpha(n - p - 1)$ で表すと、

$$I := \left[\hat{Y}_\epsilon - t_\alpha(n - p - 1) \hat{\sigma} \sqrt{1 + \frac{1}{n} (1 + (\xi - \bar{x})^\top s^{-1} (\xi - \bar{x}))}, \hat{Y}_\epsilon + t_\alpha(n - p - 1) \hat{\sigma} \sqrt{1 + \frac{1}{n} (1 + (\xi - \bar{x})^\top s^{-1} (\xi - \bar{x}))} \right]$$

とおける。

(5) 同様にして

$$\frac{\hat{Y}_\xi - \mu_\xi}{\sqrt{\hat{\sigma}^2 1 + \frac{1}{n} (1 + (\xi - \bar{x})^\top s^{-1} (\xi - \bar{x}))}} \sim t(n - p - 1).$$

(6) よって、真値 μ_ξ に対して $\mu_\xi \in I$ が成り立つ確率は $1 - \alpha$ になる： $P_{\mu_\xi}[\mu_\xi \in I] = 1 - \alpha$ 。

(7) このような区間を、信頼係数 $1 - \alpha$ の信頼区間という。

1.5.4 正規誤差について

計画された実験に基づくデータならば、正規誤差の仮定は(事前分布としても)極めて論理的であるが、「実験が十分に統制されていないこと」が誤差の要因になる。

要諦 1.5.12. 位置母数についての、正規性を仮定した推測法は、あまりに misspecification が激しくない限りロバストで、また misspecification が激しい場合は、標本数が著しく少なくない限り検出可能である。そしてそういう場合は実験や観測の条件に重大な異常があるのが普通であるので、それを十分チェックしモデルを再検討するのが良い。それゆえ、この場合に、ただ形式的にロバストな手法を導入することは適当でない。

1.6 その他の分散分析の例

まとめると、分散分析とは仮説検定において検定統計量を構成する手法の1つで、観測誤差 ϵ 、各要因 x_i 、それらの相互関係による変動への寄与を分解することを基本とする。主な枠組みは Fisher により、したがって用語は実験計画法に根ざしている。各要因 x_i は所与のものとする点で因子分析とは志向が違う。

要諦 1.6.1. 特に初源的な Fisher の設定では、複数の土地で得られた標本 X_1^i, \dots, X_n^i が、 i に依らず平均が等しいかどうかを検定する際に、その分散の級内変動 R_0^2 と級間変動 $R_1^2 - R_0^2$ の和としての構造を用いる手法をいう。名前に分散とあるが、本質的には平均母数の検定である。よって、Gauss-Markov の定理 1.3.9 の通り、 ϵ への3次以上の仮定は必要がないという点で、ノンパラメトリックな手法である。Fisher の精神の通り、実験は計画され、変量 x_i は既に考え抜かれており、標本は十分な管理下で得られている。Pearson の生物学や、行動科学において因子分析が考えられる点と対照的である。

用語 1.6.2 (factor, levels, n -way, layout).

- (1) A を **要因** または **因子** といい、後者の用語を採用する際は、操作可能性・選択の余地が強調される。
- (2) A_1, \dots, A_n を **n -水準** という。
- (3) 因子 A, B, \dots の数を一元、二元と表現する。
- (4) p 要因 n 水準実験の場合、 pn 個の標本が必要になる。これを **配置** という。

1.6.1 一元配置

one-way ANOVA

模型 1.6.3. 因子 A に p -水準 A_1, \dots, A_p を考える。各水準 A_i による効果を変量 μ_i で表し、水準 A_i の $j \in [n_i]$ 番目の観測 Y_{ij} を

$$Y_{ij} = \mu_i + \epsilon_{ij}$$

によって説明することを考える。すると、これは Gauss-Markov モデル $Y = X\beta + \epsilon$ とみなせる、ただし、 $n := \sum_{i \in [p]} n_i$ として

$$Y = \begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ \vdots \\ Y_{pn_p} \end{bmatrix} \in \mathbb{R}^n, \epsilon \in \mathbb{R}^n, X = \text{diag}(1_{n_1}, \dots, 1_{n_p}) \in M_{np}(\mathbb{R}), \beta = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_p \end{bmatrix}$$

実験がうまく管理されているとき、多くの場合は $\epsilon_{ij} \sim N(0, \sigma^2)$ という正規性と一様分散の仮定をおく。

問題 1.6.4. 検定問題 $H_0: \mu_1 = \dots = \mu_p$ を考える。

検定 1.6.5. affine 制約を

$$H^\top = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & -1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & -1 \end{bmatrix}$$

$d = 0$ としたときの F -検定を用いる。このとき、 R_0^2 を **級内変動**、 $R_1^2 - R_0^2$ を **級間変動** という。

1.6.2 二元配置

1.7 主成分分析と因子分析

元来、変量 x_i 、すなわち計画行列 X は所与としたが、このデザインを問題にすると、因子分析・主成分分析が問題となる。

要諦 1.7.1 (主成分分析). デザイン X を決めてしまって、その成分の間に直交条件を課す. $p = 2$ のときは、行列 $X \in M_{n2}(\mathbb{R})$ は二次元プロットで表せ、これを **biplot** という. 各独立成分を**主成分**という. X は縦長行列となるが、モデル $Y = X\beta + \epsilon$ で β を推定するには、まず X の特異値分解を考えると良い.

命題 1.7.2 (特異値分解 (Autonne 1915)). 任意の縦長行列 $A \in M_{m,n}(\mathbb{R})$ について、直交行列の縦長剪断 $U \in O_{mn}(\mathbb{R})$ と通常の直交行列 $V \in O_n(\mathbb{R})$ が存在して、 $r := \text{rank}(A)$ とすると

$$A = U\Sigma V^\top, \quad \Sigma = \begin{bmatrix} D & O_{r,n-r} \\ O_{m-r,r} & O_{m-r,n-r} \end{bmatrix}, \quad D = \text{diag}(\sigma_1, \dots, \sigma_r), \quad (\sigma_1 \geq \dots \geq \sigma_r > 0)$$

と表せる.

歴史 1.7.3 (因子分析). 因子分析法は Pearson から始まるが、心理学分野の Spearman が [Spearman, 1904] で創始し、大きく育てた. 観測量について、説明変数の選択・決定問題と剪定を狙うのが因子分析であり、既存の説明変数を最も「線型独立」に近いようなものへの変換を構成するのが主成分分析である.

模型 1.7.4. 各個体 $i \in [n]$ の変量 $j \in [m]$ の実現値を z_j^i とし、これを

- (1) (その個体に関する) 共通因子 f_1^i, \dots, f_p^i
- (2) (その変量に関する) 独自因子 u_j^i

によって説明しようとする. 特に、各個体における共通因子 $f^i \in \mathbb{R}^p$ の構造を決めることを目的とする.

- (1) 共通因子の係数行列 $\beta = (\beta_a^i) \in M_{np}(\mathbb{R})$ を**因子パターン**と呼ぶ.
- (2) 因子の間の計量 $(\text{Cov}[z_j^i | f_a^i]) \in M_{mp}(\mathbb{R})$ を**因子構造**という. 外生性・内生性に関する概念である.

共通因子を取り直す基準を探すには、相関計数の行列 $(\text{Corr}[z_{j_1}^i | z_{j_2}^i])_{j_1, j_2 \in [m]} \in M_m(\mathbb{R})$ を実現するベクトルとして、各 z_j^i は把握できる.

第 2 章

統計的決定理論

(意志) 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，特定の決定関数を無条件に是とするものではない．

- (1) Neyman-Pearson の理論はある指数分布族モデルに対して一様最強力検定の構成を可能にする．
- (2) Rao-Blackwell の理論は不偏推定量のクラスの中で分散が最小のものが存在することを明らかにする．
- (3) 一方でこのような理論が使えないときは，漸近最適理論と呼ばれる．

2.1 統計的決定問題

2.1.1 Wald の枠組み

定義 2.1.1 (decision problem). $(\mathcal{G}, \mathcal{B})$ を決定空間とする．

- (1) 可測関数 $\delta: \mathcal{X} \rightarrow \mathcal{G}$ を (非確率的) **決定関数** という．
- (2) 関数 $\delta: \mathcal{X} \rightarrow P(\mathcal{G})$ が $\delta(-|B) \in L(\mathcal{X}; [0, 1])$ ($B \in \mathcal{B}$) を定めるとき **確率的決定関数** という．
- (3) 一点をデルタ測度によって同一視 $\mathcal{G} \hookrightarrow P(\mathcal{G}); d \mapsto 1_{\{d\}}$ すれば，一般の決定関数を確率的決定関数とみなせる．
- (4) 決定空間と確率的決定関数の集合 $\Delta \subset \text{Map}(\mathcal{G}, P(\mathcal{G}))$ との組を **決定問題** という．

模型 2.1.2 (statistical decision problem (Abraham Wald 1939)). 次の統計的実験と決定空間に加え，決定関数の空間と損失関数を併せた 4 組 $((\mathcal{X}, \mathcal{A}, \mathcal{P} = (P_\theta)_{\theta \in \Theta}), (\mathcal{G}, \mathcal{B}), \Delta, W)$ を **統計的決定問題** という．

- (1) 可測空間 $(\mathcal{X}, \mathcal{A})$ を観測が値を取る空間とし，**標本空間** という．モデル $\mathcal{P} = \{P_\theta\}_{\theta \in \Theta} \subset P(\mathcal{X})$ によって統計的実験をなすとする．
- (2) 可測空間 $(\mathcal{G}, \mathcal{B})$ を，可能な決定の全体のなす空間とし，**決定空間** または **行動空間** という．
- (3) $\Delta \subset \text{Map}(\mathcal{X}; P(\mathcal{G}))$ は確率的決定関数の族とする．
- (4) 関数 $W: \mathcal{P} \rightarrow L(\mathcal{G})_+$ を **損失関数** といい，真値が $\theta \in \Theta$ であるときに決定 $d \in \mathcal{G}$ を下したことによって起こる損失を表す．

2.1.2 最良な決定関数

損失関数を大域的に積分して得る対応 $R: \mathcal{P} \times \Delta \rightarrow \mathbb{R}$ を危険関数といい， $\Delta \hookrightarrow \text{Map}(\mathcal{P}; \mathbb{R})$ は像の一致ノルムから引き起こされた順序構造を Δ に定める． Δ の極小元を許容的決定，最小元を一様に良い決定という．

定義 2.1.3 (risk function). 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 $\delta: \mathcal{B} \times \mathcal{X} \rightarrow [0, 1]$ 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．

定義 2.1.4 (uniformly good, best, admissible). 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, δ_1 と δ_2 は同程度に良いまたは一様に悪くはない決定関数であるという.
- (2) 一様に悪くはな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, δ_1 は δ_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. この順序関係に関する Δ の最大元が存在するならば, これを最良の決定関数という.
- (4) Δ に最大元が存在せず, 極大元 δ_* のみが存在するとき, 許容的であるという.

定義 2.1.5 (minimax, admissible). Δ には「一様に良い」という順序関係だけでなく, リスク関数の上限 $\sup_{\theta \in \Theta} R(\theta, \delta)$ によるノルムが入る. これに関する最小元となる決定関数 $\delta \in \Delta$ を, ミニマックス決定関数という.

要諦 2.1.6. ミニマックス解は事前分布が全くわかっていない時に利用される決定方式である. 逆に事前分布がわかっているときは Bayes 解が用いられる. この両極端の中間が Bayes 決定理論になる.

定義 2.1.7. 部分集合 $\Delta_0 \subset \Delta$ について,

- (1) $\forall \delta \in \Delta \setminus \Delta_0 \ \exists \delta_0 \in \Delta_0 \ \delta < \delta_0$ のとき, 完備類であるという.
- (2) 完備類 Δ_0 のどの真部分集合も完備類ではないとき, 最小完備類という.
- (3) Δ_0 が本質的完備類であるとは, 弱めた条件 $\forall \delta \in \Delta \setminus \Delta_0 \ \exists \delta_0 \in \Delta_0 \ \delta \leq \delta_0$ が成り立つことをいう.

要諦 2.1.8. 実は, 最小完備類は, 存在するならば, 許容的な決定関数の全体に等しい.

例 2.1.9 (平均値の推定量の決定). $(\mathcal{X}, \mathcal{A}) := (\mathbb{R}^n, \mathcal{B}_n)$,

$$\mathcal{P}_1 := \left\{ P = P_*^{\otimes n} \in P(\mathbb{R}^n) \mid P_* \in P(\mathbb{R}), \int_{\mathbb{R}} x^2 P_*(dx) < \infty \right\}$$

$\Theta := \mathcal{P}_1$ 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

$$T_1 := X_1, \quad T_2 := \sum_{j=1}^n \frac{X_j}{n}, \quad T_3 := X_n$$

という非確率的決定関数の族で, 損失関数は二乗誤差 $W(P, a) := (a - \mu_1(P))^2$ 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. \square

2.1.3 不偏推定問題

不偏推定量は複数考え得るから, その間の効率性を比べる問題は, 統計的決定問題である. しかし一般には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, 多くの場合は分散を次の「最適性」の指標とし, この場合の最良不偏推定量を UMVUE という.

例 2.1.10 (不偏推定量の例). 完備十分統計量を利用して作ることになる.

- (1) $\{B(1, \theta)^n\}_{\theta \in [0,1]}$ について, $\delta(x) := \frac{T(x)}{n}$ は θ の不偏推定量である. さらに, 統計的関数 $g : [0, 1] \rightarrow [0, 1]$ を $g(\theta) = \theta^2$ とすると, $\delta_1 = \frac{T(T-1)}{n(n-1)}$ ($n \geq 2$) はこの不偏推定量である. 一般に, $g(\theta) = \theta^k$ の不偏推定量は

$$\delta = \frac{T(T-1) \cdots (T-k+1)}{n(n-1) \cdots (n-k+1)}$$

が与える。しかし、 $1/\theta$ は U -推定可能でない。不偏推定量の期待値は必ず有界であるためである。

(2) $(G(1/\theta, \nu))_{\theta \in \mathbb{R}^+}$ ($\nu \in \mathbb{R}^+$) について、 $g(\theta) = \theta^r$ ($r \in \mathbb{N}^+$) の不偏推定量は

$$\delta(x) = \frac{\Gamma(\nu)}{\Gamma(\nu + r)} x^r$$

が与える。

(3) $(\text{Pois}(\theta))_{\theta \in \mathbb{R}^+}$ について、 $g(\theta) = \theta^r$ ($r \in \mathbb{N}^+$) の不偏推定量は

$$\delta(x) = x(x-1) \cdots (x-r+1)$$

が与える。

(4) $(\text{Mult}(n; \theta_1, \dots, \theta_k))_{\theta \in \Theta}$, $\Theta := \left\{ \theta \in \mathbb{R}_+^k \mid \sum_{i \in [k]} \theta_i = 1 \right\}$ について、 $\theta_i, \theta_i \theta_j$ の不偏推定量は

$$\delta_i(x) = \frac{x_i}{n}, \quad \delta_{ij}(x) = \frac{x_i x_j}{n(n-1)} \quad (n \geq 2, i \neq j)$$

が与える。

また、これらはいずれも UMV 不偏推定量でもある (Lehmann-Scheffé 2.3.1). □

例 2.1.11 (不偏分散). モデル \mathcal{P} を、2 次の積率が有限な確率分布 P の n -直積全体とする：

$$\mathcal{P} := \left\{ P = P_*^{\otimes n} \in P(\mathbb{R}^n) \mid P_* \in P(\mathbb{R}), \int_{\mathbb{R}} x^2 P_*(dx) < \infty \right\}$$

(1) $\hat{\mu} = \bar{x} := \frac{1}{n} \sum_{j \in [n]} x_j$ は平均 $\alpha_1(P)$ の不偏推定量である。

(2) $U^2 := \frac{1}{n-1} \sum_{j \in [n]} (x_j - \bar{x})^2$ は分散 $\mu_2(P)$ の不偏推定量である。これを**不偏分散**という。

また、これらの推定量は、モデルを $\{N(\theta_1, \theta_2)^{\otimes n}\}_{\theta \in \mathbb{R} \times \mathbb{R}^+}$ に限ったときは UMV 不偏推定量でもある。

$$\text{Var}[\bar{x}] = \frac{\theta_2}{n}, \quad \text{Var}[U^2] = \frac{2\theta_2^2}{n-1}.$$

□

問題 2.1.12 (不偏推定問題). 統計的決定理論の枠組みで、ある U -推定可能な統計的関数 $g: \mathcal{X} \rightarrow \mathbb{R}^p$ についての不偏推定問題を考えると、次の通りになる。

- (1) 統計的実験を $(\mathcal{X}, \mathcal{A}, (P_\theta))$ とする。
- (2) 決定空間は $\mathcal{D} \subset \mathbb{R}^p$ なる凸 Borel 集合とする。
- (3) 損失関数 $W: \Theta \rightarrow \mathcal{L}(\mathcal{D})_+$ の像は凸関数の空間とする。
- (4) 決定関数全体の空間を $\Delta_g \subset L(\mathcal{X}; \mathcal{D})$ と書き、 g の不偏推定量の全体とする： $\Delta_g \neq \emptyset$ 。

定義 2.1.13 (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 δ を**一様最小分散不偏推定量**という。

要諦 2.1.14. 一般には二乗損失を考えるため、損失関数は $W(\theta, a) = [a - g(\theta)]^2$ 、危険関数は $R(\theta, \delta) = \text{Var}_\theta[\delta]$ となる。この観点で、分散が最も小さい推定量を**有効 (efficient)**であるという。これを古典的には、**一様最小分散不偏推定量**という。

定義 2.1.15 (BLUE: best linear unbiased estimator). Δ_g をさらに線型関数に限る。

2.1.4 Bayes 推定問題

不偏推定問題では、最適な不偏推定量のクラスとして、一様最小分散 (UMV) 不偏推定量が考えられた。Bayes 推定問題では、ミニマックス Bayes 推定量が考えられる。

ミニマックス解は事前分布が全くわかっていない時に利用される決定方式である。逆に事前分布がわかっているときは Bayes 解が用いられる。この両極端の中間が Bayes 決定理論になる。そのような状況では、事前分布に対する仮定の下で、最悪のリスクを最小にすることを考えるため、むしろ不偏推定量になることは少ない。意図的にバイアスを加えているとも言える。

模型 2.1.16. まず母数の空間 Θ を可測とする。これに伴って損失関数 $W: \Theta \times \mathcal{D} \rightarrow \mathbb{R}_+$ を可測、危険関数 $R(-, \delta): \Theta \rightarrow \mathbb{R}$ を可測とする。さらに Θ 上の確率分布 $\pi \in P(\Theta)$ を用意し、**事前分布**という。これが定める $\delta \in \Delta$ の関数

$$R(\pi; \delta) := \int_{\Theta} R(\theta, \delta) \pi(d\theta)$$

を **Bayes リスク**といい、これを最小にする決定関数 $\delta \in \Delta$ を **Bayes 決定関数**という。特に、Bayes 決定関数が推定量であるとき、Bayes 推定量という。

観察 2.1.17. 確率分布 $P_{\pi}: \mathcal{A} \times \mathcal{B}(\Theta) \rightarrow [0, 1]$ を

$$P_{\pi}(A \times C) := \int_C P_{\theta}(A) \pi(d\theta)$$

によって定められたとする。 $X: \mathcal{X} \times \Theta \rightarrow \mathcal{X}$ を第一射影とし、これについての第二射影 $\vartheta: \mathcal{X} \times \Theta \rightarrow \Theta$ の条件付き分布 $p(d\theta|X=x)$ が考えら得るとすると、これを**事後分布**という。このとき、Bayes リスクは

$$\begin{aligned} R(\pi; \delta) &= \int_{\mathcal{X} \times \Theta} \left(\int_{\mathcal{D}} W(\theta, a) \delta(da|X=x) \right) P_{\theta}(dx) \pi(d\theta) \\ &= \int_{\mathcal{X}} \int_{\mathcal{D}} \left(\int_{\Theta} W(\theta, a) p(d\theta|x) \right) \delta(da|x) P_{\pi}^X(dx) \end{aligned}$$

と表せるから、決定関数 $\delta: \mathcal{X} \rightarrow P(\mathcal{D})$ は $\delta(x)$ を、 $\int_{\Theta} W(\theta, a) p(d\theta|x)$ を最小にする決定 $a \in \mathcal{D}$ 上のデルタ関数とすれば、これは Bayes 決定関数になる。よって、Bayes 決定関数は本質的に非確率的であるはずである。 \square

2.1.5 一様最強力検定問題

検定とは、態度がそもそも非常に科学的な存在である。

検定問題は、発見 H_1 に対立して、帰無的な反駁 H_0 を棄却して科学的発見を定立出来るか否かを焦点に構成されるため、通常の統計的決定問題の枠組みを変形して使うことが多い (**Neyman-Pearson の基準**)。有意水準は十分小さくし、第1種の過誤 (偽陽性、早とちり) の最小化を念頭に検定が構成することを前提とし、危険関数は $R(\theta, \delta_{\varphi}) 1_{\Theta_1}$ として使用する。すると、 Θ_1 上に台を持ち、 H_1 が正しいのに H_0 を棄却できない (第二種の過誤) 確率 $1 - \beta_{\varphi}(\theta)$ (の小ささ) を検定の良さの基準とし、これを一様に最小にするものを **UMP 検定**という。 β_{φ} を検出力関数というが、これは「発見 H_1 の可能性の検出」を意味する。

模型 2.1.18. 統計的実験 $(\mathcal{X}, \mathcal{A}, (P_{\theta})_{\theta \in \Theta})$ において、

- (1) **検定問題**とは、母数空間の非自明な分割 $\Theta = \Theta_0 + \Theta_1$ をいい、各成分を**仮説**という。
- (2) 単元集合で表される仮説を**単純仮説**、そうでない場合を**複合仮説**という。
- (3) 可測関数 $\varphi \in L(\mathcal{X}; [0, 1])$ を**検定 (関数)**といい、 $\beta_{\varphi}(\theta) := E_{\theta}[\varphi]$ を Θ_1 上の関数とみたものを**検出力関数**、 Θ_0 上での上限 $\sup_{\theta \in \Theta_0} \beta_{\varphi}(\theta)$ を**サイズ**という。
- (4) 大きさ $\alpha \in [0, 1]$ 以下の検定を**水準 α -検定**といい、その全体を $\Phi_{\alpha} := \alpha B \subset L^{\infty}(\mathcal{X}; [0, 1])$ で表す。

これを統計的決定問題の枠組みに載せるには,

(1) 決定空間 $\mathcal{D} = 2$ は採択する仮説の番号の空間とする.

(2) 損失関数は $W(\theta, \alpha) = \begin{cases} 1_{\{1\}}(\alpha) & \theta \in \Theta_0, \\ 1_{\{0\}}(\alpha) & \theta \in \Theta_1. \end{cases}$ とすると, 危険関数は

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

となる.

(3) $\Phi_1 \rightarrow \Delta$ を,

$$\delta_\varphi(dz|x) = \varphi(x)\delta_1(dz) + (1 - \varphi(x))\delta_0(dz)$$

と同一視すると, $\Phi_\alpha = \{\alpha \in \Phi_1 \mid \sup_{\theta \in \Theta_0} R(\theta, \delta_\alpha) \leq \alpha\}$ と表せる.

定義 2.1.19 (UMP test (uniformly most powerful)). Θ_1 上で一様に危険関数を最小にする検定 φ_0

$$\forall \varphi \in \Phi_\alpha \quad \forall \theta \in \Theta_1 \quad \beta_{\varphi_0}(\theta) \geq \beta_\varphi(\theta)$$

を一様最強力検定という.

例 2.1.20 (ランダム化検定). H_0 の立場に立って, そこでの誤謬を導き出し, これを棄却するのが仮説検定の意味論である. 従って検定関数は, 基本的には帰無仮説 H_0 の下で計算した確率が十分小さい時に棄却する, という算譜に従い, 第一種の過誤 (H_0 が正しいのにこれを棄却してしまうという思い上がり) の最小化を優先する. このときの棄却域 $[0, \alpha]$ の右端を**有意水準**といい, 大抵 $\alpha \leq \frac{1}{20}$ である. すると検定関数は, 計算して得られた分布 P_{H_0} に基づいて $\varphi: \mathcal{X} \rightarrow \partial[0, 1] = \mathcal{D}; x \mapsto P_{H_0}[X = x]$ のように思えるが, $[0, 1]$ 値の可測関数であるのは, ランダム化検定を表すためである. これは, 棄却域を \mathcal{X} に対して $\partial[0, 1] = \mathcal{D}$ に対応させるのではなく, 確率的に採択することとする. $\varphi(x) \in [0, 1]$ は, H_0 を棄却する確率を表す (1 に近いほど H_1 を採択しやすい).

(1) $\alpha = 0.05$ 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」と解釈の方が選択される.

(2) $x_0 = 15$ 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 φ で H_0 を棄却し, 確率 $1 - \varphi$ で H_0 を採択することとする.

□

例 2.1.21 (goodness of fit).

(1) Θ_0 として特定の分布族の全体を取るとき, 例えば帰無仮説 H_0 を「観測値 Y は正規分布に従う」とするとき, これを**モデルの適合度検定**と呼ぶ.

(2) さらに一般に, 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.

□

2.1.6 不偏検定問題

一般に一樣最強力検定は存在しないが、これは Neyman-Pearson の基準に合わない検定問題にもこれを適用していることにより生じる。そこでより普遍的な基準とみなせる枠組みを採用する。

例 2.1.22 (一樣最強力検定が存在しない形の検定問題). $\{N(\theta, 1)\}_{\theta \in \mathbb{R}}$ における検定問題 $H_0: \theta = 0$ vs. $H_1: \theta \neq 0$ に一樣最強力検定は存在しない。実際、対立仮説が $\theta > 0$ ならば、一樣最強力検定の形は $x \geq x_0$ というものである (Neyman-Pearson の補題 2.4.2) が、これは $\theta < 0$ に対する最適な検定ではありえない。しかし問題の本質は、決定関数の空間を大きくしすぎていることである。検定 $1_{\{x \geq x_0\}}$ などは、 $\theta' < 0$ のときの検出力は極めて低い。 \square

定義 2.1.23 (level- α unbiased test, uniformly most powerful unbiased test). 非自明な検定 $\Theta = \Theta_0 + \Theta_1$ の問題 $H_0: \theta \in \Theta_0$ vs. $H_1: \theta \in \Theta_1$ について、

(1) 検定 $\varphi \in L(\mathcal{X}; [0, 1])$ が水準 α の不偏検定であるとは、次が成り立つことをいう：

$$[\forall \theta \in \Theta_0 \ E_\theta[\varphi] \leq \alpha] \wedge [\forall \theta \in \Theta_1 \ E_\theta[\varphi] \geq \alpha].$$

水準 α の不偏検定全体の集合を $\Phi_\alpha^u \subset \Phi_\alpha \subset L(\mathcal{X}; [0, 1]) = \Phi_1$ で表す。

(2) $\varphi_0 \in \Phi_\alpha^u$ が水準 α の一樣最強力不偏検定であるとは、次を満たすことをいう：

$$\forall \theta \in \Theta_1 \ \forall \varphi \in \Phi_\alpha^u \ \beta_{\varphi_0}(\theta) \geq \beta_\varphi(\theta).$$

要諦 2.1.24. この制限は本質的ではない、なぜなら、「 φ は何も判断しない定値検定 α よりも悪くない」ことを課しているのみである。

2.1.7 不変検定問題

不偏検定への制限が自然であったように、不変検定への制限も自然である。これは十分統計量の形式化にも思える。Guass-Markov モデルや 3 元配置分散分析などは、不変検定の立場から明瞭に数学的に一樣最強性を示せる。

模型 2.1.25. $G \subset L(\mathcal{X}; \mathcal{X})$ を可測変換の部分群、 $\{P_\theta\} \subset P(\mathcal{X})$ を識別可能なモデル ($\theta \mapsto P_\theta$ は単射) とし、 \mathcal{P} は G -不変とする：
 $\forall g \in G \ \forall P \in \mathcal{P} \ P^g \in \mathcal{P}$.

- (1) このとき $P^g =: P_{\bar{g}\theta}$ とすると、 $\bar{g}: \Theta \rightarrow \Theta$ は全単射で、 $G \simeq \bar{G} \subset \text{Aut}_{\text{Set}}(\Theta)$ と同一視できる。
- (2) 検定問題 $\Theta = \Theta_0 + \Theta_1$ が G -不変であるとする。すなわち、 $\forall g \in G \ \bar{g}(\Theta_i) \subset \Theta_i$ 。このとき実は等号が成り立つ。
- (3) $T \in L(\mathcal{X})$ が G -不変であるとは、 $\forall g \in G \ T(g-) = T(-)$ をいう。
- (4) 写像 $S: \mathcal{X} \rightarrow \mathbb{R}$ が最大不変量であるとは、 \mathcal{X} の G -軌道上の単射となっていることをいう： G -不変かつ $\forall x_1, x_2 \in X \ S(x_1) = S(x_2) \Rightarrow (\exists g \in G \ x_1 = gx_2)$ をいう。

例 2.1.26.

- (1) 両側 t -検定は affine 変換 $g: (x_j) \mapsto (a(x_j - \mu_0) + \mu_0)$ ($a \in \mathbb{R}, a \neq 0$) からなる群 G について不変で、検定統計量 T による検定 $\varphi \in L(\mathcal{X})$ も G -不変である。
- (2) 正規性を仮定した Gauss-Markov モデル $Y = X\beta + \epsilon \in \sim N_n(0, \sigma^2 I_n)$ について、 F -検定 $F \geq c$ は一樣最強力不変検定になる。

\square

定義 2.1.27 (UMP 不変検定). G -不変な分布族 \mathcal{P} に関する G -不変検定問題に対して、検定統計量を G -不変なものに限定することも自然である。その中で検出力を一樣に最大にするものを一樣最強力不変検定という。実は任意の G -不変検定は最大不変量の関数となるため、さらにこのクラスで探せば良い。

2.1.8 区間推定問題

パラメータ $\theta \in \Theta$ に対して、集合値統計量 $S: \mathcal{X} \rightarrow P(\Theta)$ によって推定する問題を、領域推定問題という。すると、なるべく区間の幅を小さくしたいが、信頼係数 $1 - \alpha$ も高くしたいというトレードオフが起きる。

模型 2.1.28. 集合値写像 $S: \mathcal{X} \rightarrow P(\Theta)$ が信頼係数 $1 - \alpha$ ($\alpha \in (0, 1)$) の信頼域であるとは、次の2条件を満たすことをいう：

- (1) 真値が領域 $S(x)$ 入っているかの事象は可測： $\forall \theta \in \Theta \{x \in \mathcal{X} \mid \theta \in S(x)\} \in \mathcal{A}$ を満たす。
- (2) $\forall \theta \in \Theta P_\theta[\theta \in S(x)] \geq 1 - \alpha$.

2.2 統計量の基礎理論

決定関数（推定量・検定統計量）としては、十分統計量の関数のみを考えれば良く、十分統計量は分布族の密度関数の形から特徴付けられる。小標本理論唯一の望みであり、実用的には分布族は十分統計量の特徴付け (Fisher-Neyman 2.2.13) が与えるものよりもさらに単純なサブクラスを考える必要がある。

歴史 2.2.1. UC Berkeley の Blackwell, Lehmann, Scheffé らが達成した数理統計学の華と目される理論である。Rao-Blackwell は、十分統計量で条件付ければ不偏推定量を一樣に改善することが出来ること、Lehmann-Scheffé はさらにその十分統計量が完備でもあるならば最良不偏推定量であることを保証する。

問題 2.2.2. $(\mathcal{X}, \mathcal{A}, (P_\theta))$ を統計的実験、 $(\mathcal{T}, \mathcal{B})$ を統計量の終域の可測空間とし、統計量 $T: \mathcal{X} \rightarrow \mathcal{T}$ を考える。

- (1) T の分布は、各 P_θ に対して、 P_θ^T に従うので、新たな実験 $(\mathcal{T}, \mathcal{B}, (P_\theta^T))$ を得る。
- (2) $(\mathcal{X}, \mathcal{A})$ は完備距離空間上の可測空間だとすれば、 T を与えた下での X の正則条件付き確率 $\mu_\theta^{X|T}(t, A): \mathcal{T} \times \mathcal{A} \rightarrow [0, 1]$ が本質的に一意に定まる。すなわち、
 - (a) 任意の $A \in \mathcal{A}$ について、 $\mu_\theta^{X|T}(-, A): \mathcal{T} \rightarrow [0, 1]$ は \mathcal{B} -可測関数。
 - (b) 任意の $A \in \mathcal{A}$ について、

$$P_\theta[A \cap T^{-1}(B)] = \int_B \mu_\theta^{X|T}(t, A) P_\theta^T(dt), \quad (B \in \mathcal{B}).$$

- (c) 任意の $t \in \mathcal{T}$ について、 $\mu_\theta^{X|T}(t, -): \mathcal{A} \rightarrow [0, 1]$ は $(\mathcal{X}, \mathcal{A})$ 上の確率測度。

要諦 2.2.3. 確率変数 T を与えた下での正則条件付き確率は、 σ -代数 $\sigma[T]$ を与えた下での正則条件付き確率 $P[A|\sigma[T]] := E[1_A|\sigma[T]]$ と考えても良い。すると $P[-|\sigma[T]]$ とは、 $\mathcal{A} \rightarrow L^1(\mathcal{T}; [0, 1])$ という写像であり、 $A \in \mathcal{A}$ を与える毎に可積分な確率変数を返し、

$$E\left[1_A 1_{T^{-1}(B)}\right] = E\left[P[A|\sigma[T]] 1_{T^{-1}(B)}\right], \quad B \in \mathcal{B}$$

を満たす。

2.2.1 十分統計量の定義と例

未知パラメータの除去

データが押し出す分布族 (P_θ^X) から真の分布を推定したい。統計量 $T: \mathcal{X} \rightarrow \mathcal{T}$ で条件付けて得る条件付き分布の族 $\{P_\theta^X[-|T]\}$ 、すなわち、 $X|T$ の分布が θ に依らない T に添字付けられた分布族になるとき、 T を十分統計量という。このとき、更なる押し出し $(P_\theta^{T(X)})_{t \in \text{Im } T}$ の問題に帰着される。

定義 2.2.4 (sufficient statistics). 統計的実験 $(\mathcal{X}, \mathcal{A}, (P_\theta)_{\theta \in \Theta})$ に対して、統計量 $T: \mathcal{X} \rightarrow \mathcal{T}$ が十分統計量であるとは、

- (1) T を与えた下での P_θ に関する X の正則条件付き分布 $P_\theta^X[-|T]$ が $\theta \in \Theta$ に依らないことをいう。

- (2) すなわち, $A \in \mathcal{A}$ を与える毎に, $\theta \in \Theta$ に依らない可積分確率変数 $P^X[A|T] \in L^1(\mathcal{T})$ が存在して, 任意の $B \in \mathcal{B}, \theta \in \Theta$ に対して,

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

が成り立つことをいう.

要諦 2.2.5 (十分統計量の意味). 上式は各 $B \in \mathcal{G}$ 上で, 積分を通じれば, $P^X[A|T]$ が $\theta \in \Theta$ に依らずに, 素朴な意味での条件付き確率 $P_\theta[A|\sigma[T]]$ を模倣することを要請している. よって, $T(X_1, \dots, X_n)$ を考察すれば, これ以上データ X_1, \dots, X_n を観察しても θ についての追加情報を得られないことを意味する.

例 2.2.6 (独立和が与える十分統計量). 代表的な分布族に対して, 基本的に, 観測量の和 (平均) が十分統計量になる.

- (1) 空間 $(2^n, P(2^n))$ 上の二項分布族 $\{\text{Ber}(\theta)^{\otimes n}\}_{\theta \in [0,1]}$ に対して, 総成功回数を表す統計量 $T(x) := \sum_{j \in [n]} x_j$ は十分である.
- (2) 2 回の試行により得る標本空間 \mathbb{N}^2 上の Poisson 分布族 $\{\text{Pois}(\lambda)^{\otimes 2}\}_{\lambda \in \mathbb{R}^+}$ に対して, 事故の総発生回数を表す統計量 $T(x_1, x_2) := x_1 + x_2$ は十分である.
- (3) \mathbb{R}^n 上の正規分布族 $\{P_\theta := N(\theta, 1)^{\otimes n}\}_{\theta \in \mathbb{R}}$ に対して, 統計量 $T(x) := \sum_{j \in [n]} x_j$ は十分である. 実際,

$$\frac{dP_\theta}{dx}(x) = \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{1}{2} \sum_{j \in [n]} x_j^2} e^{\theta T(x) - \frac{n\theta^2}{2}}.$$

□

[証明].

- (1) 実際, 成功回数 $T = t$ が与えられたときの条件付き確率は, $\left\{ x \in 2^n \mid \sum_{i \in [n]} x_i = t \right\}$ 上に台を持ち, 並べ替えのみを違いとしてそれぞれ重さ ${}_nC_t$ を持つ. したがって, 成功確率 $\theta \in [0, 1]$ には依らない.
- (2) 母数を $\theta \in \mathbb{R}^+$ で表す. Poisson 分布の再生性 $\text{Pois}(\theta) * \text{Pois}(\theta) = \text{Pois}(2\theta)$ に注意すれば, $T \sim \text{Pois}(2\theta)$ であり, $P_\theta[\{T = t\}]$ は $e^{-2\theta} \frac{(2\theta)^t}{t!}$ と計算出来る. よって, 条件付き確率は素朴に

$$\begin{aligned} P_\theta[\{(x_1, x_2)\} | T = t] &= \frac{P_\theta[\{(x_1, x_2)\} \cap \{T = t\}]}{P_\theta[\{T = t\}]} \\ &= \frac{e^{-\theta} \frac{\theta^{x_1}}{x_1!} e^{-\theta} \frac{\theta^{t-x_1}}{(t-x_1)!}}{e^{-2\theta} \frac{(2\theta)^t}{t!}} \\ &= \frac{\frac{1}{x_1!(t-x_1)!}}{2^t \frac{1}{t!}} = \frac{1}{2^t} \frac{t!}{x_1!(t-x_1)!} = \frac{1}{2^t} \binom{t}{x_1}. \end{aligned}$$

これは θ に依らない. 特に, $X_1|T \sim B(t, 1/2)$ を得た.

- (3) 直交行列 $C \in O_n(\mathbb{R})$ の第一行は $\frac{1}{\sqrt{n}} \mathbf{1}_n$ であるとする.

$$y := Cx =: (y^1, y^*) \in \mathbb{R} \otimes \mathbb{R}^{n-1}$$

と定めると, $T(x) := \sqrt{n}y^1$. よって, 残りの成分によって積分して得る周辺確率密度関数が

$$q^*(A|y^1) := \int_{\mathbb{R}^{n-1}} \mathbf{1}_A(C^\top y) \phi(y^*; 0, I_{n-1}) dy^*, \quad (A \in \mathcal{B}(\mathbb{R}^n))$$

になることを示せば, これは θ に依らないことから従う.

実際, 任意の $A \in \mathcal{B}(\mathbb{R}^n), B \in \mathcal{B}(\mathbb{R})$ について, Fubini の定理より,

$$\begin{aligned} P_\theta[A \cap T^{-1}(B)] &= \int_{\mathbb{R}^n} \mathbf{1}_A(x) \mathbf{1}_B(T(x)) \phi(x; \theta \mathbf{1}, I_n) dx \\ &= \int_{\mathbb{R}^n} \mathbf{1}_A(C^\top y) \mathbf{1}_B(\sqrt{n}y^1) \phi(y^*; 0, I_{n-1}) \phi(y^1; \sqrt{n}\theta, 1) dy^1 dy^* \\ &= \int_{\mathbb{R}} q^*(A|y^1) \mathbf{1}_B(\sqrt{n}y^1) \phi(y^1; \sqrt{n}\theta, 1) dy^1 \end{aligned}$$

$$= \int_B q^*(A|t/\sqrt{n})\phi(t;n\theta,n)dt.$$

で、たしかに $\phi(t;n\theta,n)$ は T によって押し出された $N(\theta,1)^{\otimes n}$ の密度である。なお、残った $n-1$ 成分の平均ベクトルが 0 であるのは、 C の第 2 列から第 n 列の列ベクトルが、第 1 列すなわち $\mathbf{1}$ に直交することによる。

例 2.2.7 (対称な分布の族に対して順序統計量は十分). \mathbb{R}^n 上の対称な分布の族 \mathcal{P} を考える。すなわち、 $P \in \mathcal{P}$ は任意の置換 $\sigma \in S_n$ に対して、 $P^\sigma = P$ を満たす。このとき、 n 個のデータを降順に並べ替える写像 $T: \mathbb{R}^n \rightarrow \mathbb{R}^n; x \mapsto t(x)$ を **順序統計量** という。これは \mathcal{P} に対して十分である。 \square

[証明].

$$q(A|t) := \frac{1}{n!} \sum_{\sigma \in S_n} 1_A(\sigma(t)), \quad (t \in \mathbb{R}^n, A \in \mathcal{G}(\mathbb{R}^n))$$

が条件付き確率を与え、たしかに $P \in \mathcal{P}$ に依らない。実際、任意の $A, B \in \mathcal{G}(\mathbb{R}^n)$ に対して、

$$\begin{aligned} P[A \cap T^{-1}(B)] &= \int_{\mathbb{R}^n} 1_A(x) 1_B(T(x)) P(dx) \\ &= \int_{\mathbb{R}^n} 1_A(\sigma(x)) 1_B(T(x)) P(dx) \\ &= \int_{\mathbb{R}^n} q(A|T(x)) 1_B(T(x)) P(dx) \\ &= \int_B q(A|t) P^T(dt). \end{aligned}$$

例 2.2.8 (順序統計量の射影として得る十分統計量).

- (1) 一様分布族 $\{U(0, \theta)^{\otimes n}\}_{\theta \in \mathbb{R}^+}$ に対して、順序統計量の射影 $T(x) := x_{(n)} = \max_{j \in [n]} x_j$ は十分である。
- (2) 一様分布族 $\{U(\theta - 1/2, \theta + 1/2)^{\otimes n}\}_{\theta \in \mathbb{R}}$ に対しても、順序統計量の射影 $T(x) = (x_{(1)}, x_{(n)})$ は十分である。

[証明].

- (1) $\tilde{1}_A := \sum_{\sigma \in S_n} \frac{1_A(\sigma(x))}{n!}$ ($A \in \mathcal{G}(\mathbb{R}^n)$) について、

$$q(A|t) := \int_{\mathbb{R}^{n-1}} \tilde{1}_A(x_1, \dots, x_{n-1}, t) \frac{1}{t^{n-1}} \sum_{j=1}^{n-1} 1_{(0,t)}(x_j) dx_1 \cdots dx_{n-1}, \quad (A \in \mathcal{G}(\mathbb{R}^n))$$

と定めると、これが T を与えた下での x の条件付き確率分布を与えることを示せば良い。 $U(0, \theta)^{\otimes n}$ の密度は

$$P_\theta(dx) = \frac{1}{\theta^n} \prod_{j \in [n]} 1_{(0,\theta)}(x_j) dx, \quad (x \in \mathbb{R}^n)$$

であるから、任意の $A \in \mathcal{G}(\mathbb{R}^n), B \in \mathcal{G}(\mathbb{R})$ に対して、

$$\begin{aligned} P_\theta[A \cap T^{-1}(B)] &= \int_{\mathbb{R}^n} \tilde{1}_A(x) 1_B(T(x)) \frac{1}{\theta^n} \prod_{j \in [n]} 1_{(0,\theta)}(x_j) dx \\ &= n \int_{\mathbb{R}^n} \tilde{1}_A(x) 1_B(x_n) \frac{1}{\theta^n} \prod_{j=1}^{n-1} 1_{(0,x_n)}(x_j) 1_{(0,\theta)}(x_n) dx \\ &= \int_B q(A|x_n) P_\theta^T(dx_n). \end{aligned}$$

2.2.2 最小十分統計量

定義 2.2.9 (minimal sufficient statistic). ある十分統計量 T^* が任意の他の十分統計量 T に対して、 T の関数であるとき、これを最小十分統計量という。

2.2.3 十分統計量の下での条件付き期待値

T が十分統計量であるとき、任意の P_θ について可積分な任意の確率変数 $f: \mathcal{X} \rightarrow \mathbb{R}$ の T についての条件付き期待値は、やはりパラメータ $\theta \in \Theta$ に依らない。

命題 2.2.10. 統計的実験 $(\mathcal{X}, \mathcal{A}, (P_\theta))$ 上の確率変数 $f \in \cap_{\theta \in \Theta} L^1(P_\theta)$ について、

- (1) 十分統計量 $T: \mathcal{X} \rightarrow \mathcal{T}$ に関する任意の P_θ の条件付き期待値に殆ど至る所で等しいような、 \mathcal{T} -可測なバージョン $g \in L(\mathcal{T})$ を持つ：

$$g(t) = E_\theta[f|T=t] \quad P_\theta^T\text{-a.e.} \quad (\theta \in \Theta).$$

- (2) 十分統計量 $T: \mathcal{X} \rightarrow \mathcal{T}$ に関する任意の P_θ の条件付き期待値に殆ど至る所で等しいような、 \mathcal{A} -可測なバージョン $f' \in L(\mathcal{X})$ を持つ：

$$f'(x) = E_\theta[f|T](x) \quad P_\theta\text{-a.e.} \quad (\theta \in \Theta).$$

2.2.4 十分統計量の密度の型による特徴付け：因子分解定理

統計量がモデル \mathcal{P} に対して十分であることは、分布族 \mathcal{P} の密度関数の構造で特徴付けられる。それは、密度関数が $g(x)H_\theta(T(x), \theta)$ と、ある θ に依らない関数 g と、 T を通してのみ x に依存する関数 H とに分解できることが特徴付けになる。

仮定 2.2.11. 統計的実験 $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ について、ある σ -有限測度 μ が存在して $\mathcal{P} \ll \mu$ を満たすとする。

補題 2.2.12 (凸結合による優測度の構成). 上述の仮定の下では、ある $\{P_n\} \subset \mathcal{P}$ が存在して、 $P_* := \sum_{n=1}^{\infty} \frac{2^n}{P_n}$ とすれば $\mathcal{P} \ll P_*$ を満たす。

定理 2.2.13 (Fisher-Neyman factorization theorem). 仮定 2.2.11 を満たす統計的実験 $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ 上の可測関数 $T: \mathcal{X} \rightarrow \mathcal{T}$ について、次の2条件は同値：

- (1) T は \mathcal{P} に対して十分。
 (2) ある $H_\theta \in L(\mathbb{T})_+$ と $g \in L(\mathcal{X})_+$ が存在して、各 P_θ の密度は

$$\frac{dP_\theta}{d\mu}(x) = g(x)H_\theta(T(x)) \quad \mu\text{-a.e.} \quad (\theta \in \Theta).$$

と表せる。

[証明].

(1) \Rightarrow (2) T が \mathcal{P} に対して十分ならば、ある条件付き確率 $q(A|t)$ が存在して、

$$\int_B q(A|t)P_\theta^T(dt) = \int_{\mathcal{X}} 1_A(x)1_B(T(x))P_\theta(dx), \quad (B \in \mathcal{B}, \theta \in \Theta)$$

が成り立つ。補題により得られる優越測度 $P_* := \sum_{n \in \mathbb{N}^*} 2^{-n} P_{\theta_n}$ を取ると、これについて

$$\int_B q(A|t)P_*^T(dt) = \int_{\mathcal{X}} 1_A(x)1_B(T(x))P_*(dx) \quad (B \in \mathcal{B})$$

が成り立つ. $B \in \mathcal{B}$ は任意だったから, Lebesgue の優収束定理より, 任意の p_*^T -可積分な \mathcal{B} -可測関数 H に対して,

$$\int_{\mathcal{T}} q(A|t)H(t)P_*^T(dt) = \int_{\mathcal{X}} 1_A(x)H(T(x))P_*(dx).$$

この式に, 各 P_θ^T の Radon-Nikodym 微分 $H_\theta(t) := \frac{dP_\theta^T}{dP_*^T}(t)$ を代入すると,

$$P_\theta[A] = \int_{\mathcal{T}} q(A|t)P_\theta^T(dt) = \int_{\mathcal{X}} 1_A(x)H_\theta(T(x))\frac{dP_*}{d\mu}(x)\mu(dx)$$

を得る. $A \in \mathcal{A}$ は任意であるから, 被積分関数は μ -a.e. で一致する.

(2)⇒(1) Step1 各 $P_\theta \in \mathcal{P}$ に対して, ある非負 \mathcal{B} -可測関数 $\tilde{H}_\theta \in L(\mathcal{T})_+$ が存在して,

$$\frac{dP_\theta}{dP_*}(x) = \tilde{H}_\theta(T(x)) \text{ } P_*\text{-a.e.}$$

が成り立つ.

実際, まず

$$H_*(t) := \sum_{n \in \mathbb{N}^+} 2^{-n} H_{\theta_n}(t)$$

と定めると,

$$\frac{dP_*}{d\mu}(x) = g(x)H_*(T(x)) \text{ } \mu\text{-a.e.}$$

が成り立っている. これを,

$$\tilde{H}_\theta(t) := \frac{H_\theta(t)}{H_*(t)} 1_{B_*}, \quad B_* := \{t \in \mathcal{T} \mid H_*(t) > 0\}$$

と修正すれば良い.

Step2 P_* に関する条件付き期待値を用いて,

$$q(A|t) := E^{P_*}[1_A|T=t]$$

と定めると, これは θ に依らず P_θ に関する T を与えた下での条件付き確率になっている.

実際, 条件付き期待値の定義から, 任意の P_*^T -可積分関数 $f: \mathcal{T} \rightarrow \mathbb{R}$ について,

$$\int_{\mathcal{T}} q(A|t)f(t)P_*^T(dt) = \int_{\mathcal{X}} 1_A(x)f(T(x))P_*(dx)$$

が成り立つから, $f := 1_B \tilde{H}_\theta$ ($B \in \mathcal{B}, \theta \in \Theta$) の場合を特に考えることより,

$$\int_B q(A|t)P_\theta^T(dt) = \int_{T^{-1}(B)} 1_A(x)P_\theta(dx) = P_\theta[A \cap T^{-1}(B)].$$

■

例 2.2.14. $P_\theta \sim N_n(\theta \mathbf{1}_n, I_n)$ のとき (例 2.2.6), 密度は

$$\begin{aligned} \frac{dP_\theta}{dx}(x) &= \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{1}{2}(x - \theta \mathbf{1}_n)^\top (x - \theta \mathbf{1}_n)} \\ &= \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{1}{2} \sum_{i \in [n]} (x_i - \theta)^2} \\ &= \underbrace{\left(\frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{1}{2} \sum_{i \in [n]} x_i^2}}_{=g(x)} \underbrace{e^{\theta \sum_{i \in [n]} x_i - \frac{n\theta^2}{2}}}_{=H_\theta(T(x))}. \end{aligned}$$

と, 指数型に表せているから, Fisher-Neyman の因子分解定理より, $T(x) := \sum_{i \in [n]} x_i$ は $(N(\theta, 1)^{\otimes n})$ の十分統計量である. \square

2.2.5 十分統計量による決定関数の一様改善: Rao-Blackwell の定理

決定関数＝「推定量の候補」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．すなわち、適当な推定量 δ に対して、その十分統計量に関する条件付き期待値 $\delta_0 := E[\delta|T]$ は δ よりも改善はすれど悪くはならない決定関数である．

定理 2.2.15 (Rao-Blackwell). $T: \mathcal{X} \rightarrow \mathcal{T}$ は \mathcal{P} についての十分統計量、 $\delta \in \cap_{\theta \in \Theta} L^1(P_\theta)$ を非確率的決定関数とし、次を仮定する：

[A1] 決定空間 $\mathcal{D} \subset \mathbb{R}^p$ は凸な Borel 可測集合である．

[A2] 損失関数 $W: \Theta \rightarrow L(\mathcal{D})_+$ の像は凸関数の空間になる．すなわち、 $W(\theta, -)$ は凸．

このとき、次が成り立つ：

- (1) $\delta_0 := E[\delta|T]$ は \mathcal{X} 上の関数として θ に依らずに定まる．
- (2) δ よりも一様に悪くない：

$$\forall_{\theta \in \Theta} R(\theta, \delta_0) \leq R(\theta, \delta).$$

[証明].

- (1) 命題 2.2.10 による．
- (2) 任意の $\theta \in \Theta$ を取る．仮定 [A2] より、 δ の $\sigma[T]$ の下での正則条件付き分布 $p_\theta(-|x)$ に対する Jensen の不等式から

$$W\left(\theta, \int_{\mathcal{D}} \alpha p_\theta(d\alpha|x)\right) \leq \int_{\mathcal{D}} W(\theta, \alpha) p_\theta(d\alpha|x)$$

すなわち、 $W(\theta, \delta_0) = W(\theta, E_\theta[\delta|T]) \leq E_\theta[W(\theta, \delta)|T]$ P_θ -a.e. を得る．両辺の P_θ に関する期待値を取って、結論を得る．

■

命題 2.2.16 (Rao-Blackwell の不偏推定量に関する特別な場合：十分統計量による不偏推定量の一様改善). $T: \mathcal{X} \rightarrow \mathcal{T}$ は \mathcal{P} についての十分統計量、 $S \in \cap_{\theta \in \Theta} L^2_{P_\theta}(\mathcal{X})$ を $g \in L(\Theta)$ の不偏推定量とする．このとき、

- (1) 統計量である： $S^* := E[S|T]: \mathcal{X} \rightarrow \mathbb{R}$ は再び $g(\theta)$ の不偏推定量である．
- (2) 一様改善である： $\forall_{\theta \in \Theta} \text{Var}_\theta[S] \geq \text{Var}_\theta[S^*]$ ．等号成立条件は、 $\forall_{\theta \in \Theta} P_\theta[S = S^*] = 1$ のとき、すなわち、ある可測関数 $\psi \in L(\mathcal{T}; \mathbb{R})$ について $S = \psi(T)$ と表せるときに限る．

[証明].

- (1) (a) T は十分統計量であるから、 $S^* = E[S|T]$ は再び \mathcal{G} -可測、すなわち、統計量である．
 (b) S の不偏性から、条件付き期待値の性質より、

$$g(\theta) = E_\theta[S] = E_\theta[E[S|T]] = E_\theta[S^*].$$

- (2) 全分散の公式より、任意の $\theta \in \Theta$ について、

$$\text{Var}_\theta[S] = \text{Var}_\theta[S^*] + E_\theta[\text{Var}_\theta[S|T]] \geq \text{Var}_\theta[S^*].$$

等号は $E_\theta[\text{Var}_\theta[S|T]] = 0 \Leftrightarrow E_\theta[(S - S^*)^2]$ の場合に限る．

■

2.2.6 完備統計量の定義と例

分布族 \mathcal{P} が定める期待値の族 $\{E_P\}_{P \in \mathcal{P}} \subset L(\mathcal{X})^*$ は適切な $\text{Map}(\mathcal{X}; \mathbb{R})$ の部分空間上の線型汎関数の族とみなせる．これが $L(\mathcal{X})$ を分離するとき完備であるといい、 $L^\infty(\mathcal{X})$ を分離するとき有界完備であるという．

定義 2.2.17 (complete, boundedly complete). 統計的実験 $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ について、

- (1) 分布族 \mathcal{P} が完備であるとは、任意のデータの可測関数 $f \in L(\mathcal{G})$ について、 $\forall_{\theta \in \Theta} E_{\theta}[f] = 0$ が $\forall_{\theta \in \Theta} f = 0$ P_{θ} -a.e. を含意することをいう。
- (2) 分布族 \mathcal{P} が有界完備であるとは、同様の性質が任意の有界可測関数 $f \in L^{\infty}(\mathcal{G})$ には保証されることをいう：

$$\forall_{f \in L^{\infty}(\mathcal{G})} (\forall_{\theta \in \Theta} E_{\theta}[f] = 0) \Rightarrow (\forall_{\theta \in \Theta} f = 0 \text{ } P_{\theta}\text{-a.e.}).$$

- (3) 統計量 $T: \mathcal{G} \rightarrow \mathcal{T}$ が分布族 \mathcal{P} を完備に押し出すとは、押し出しの族 $\{P_{\theta}^T\}$ が完備であることをいう。

例 2.2.18 (完備に押し出す統計量).

- (1) 二項分布族 $\{B(1, \theta)^{\otimes n}\}_{\theta \in [0,1]}$ 上の十分統計量である標本平均 $T(x) := \sum_{j \in [n]} x_j$ は、分布族を完備に押し出す。すなわち、完備十分統計量である。
- (2) 一様分布族 $\{U(0, \theta)^{\otimes n}\}_{\theta \in \mathbb{R}^+}$ 上の十分統計量である最大元 $T(x) := x_{(n)}$ は、分布族を完備に押し出す。すなわち、完備十分統計量である。
- (3) 一方で、一様分布族 $\{U(\theta - 1/2, \theta + 1/2)^{\otimes n}\}_{\theta \in \mathbb{R}}$ を、その十分統計量 $T(x) = (x_{(1)}, x_{(n)})$ は完備には押し出さない。

□

[証明].

- (1) 任意の可測関数（すなわち任意の関数） $h: \{0, \dots, n\} \rightarrow \mathbb{R}$ に対して、 T の押し出す分布族は二項分布族であることに注意して、

$$E_{\theta}^T[h] = \sum_{t=0}^n h(t) \binom{n}{t} \theta^t (1-\theta)^{n-t} = 0, \quad (\theta \in [0, 1])$$

と仮定する。両辺を $(1-\theta)^n$ で割り、 $z := \frac{\theta}{1-\theta}$ とおくと、

$$\sum_{t=0}^n h(t) \binom{n}{t} z = 0, \quad (z \in (0, \infty))$$

が必要。多項式が $(0, \infty)$ 上で恒等的に零であるためには、係数がすべて零であることが必要であるから、 $h = 0$ である。

- (2) 任意の可測関数 $h: \mathbb{R}_+ \rightarrow \mathbb{R}$ について、 $E_{\theta}^T[h] = 0 \Leftrightarrow E_{\theta}[h(T)] = 0$ とする。一様分布 $U(0, 1)$ からの標本の順序統計量 T は $\text{Beta}(n, 1)$ に従うから、今回の T の密度は

$$p_{\theta}(t) = \frac{1}{\theta^n} n t^{n-1} 1_{(0, \theta)}(t)$$

であることに注意すれば、

$$\int_0^{\theta} h^+(t) t^{n-1} dt = \int_0^{\theta} h^-(t) t^{n-1} dt \quad (\theta > 0)$$

が必要なことが判る。これは $(0, \infty)$ 上殆ど至る所 $h^+ = h^-$ であることを要請しており、 \mathbb{R}_+ 上殆ど至る所 $h = 0$ であることが解った。

■

例 2.2.19 (正規分布族の完備性). 正規分布族自体は、平均を母数に与えただけでは完備ではないが、標本平均は統計量としては完備である！

- (1) 一次元正規分布族 $\{N(\theta, 1)\}_{\theta \in \mathbb{R}}$ や $\{N(\mu, \nu)\}_{\mu \in \mathbb{R}, \nu \in \mathbb{R}^+}$ は完備である。
- (2) $\{N(\theta, 1)^{\otimes n}\}_{\theta \in \mathbb{R}}$ は $n \geq 2$ について完備ではない。
- (3) 統計量 $T(x) := \sum_{j \in [n]} x_j$ は、分布族 $\{N(\theta, \sigma^2)^{\otimes n}\}_{\theta \in \mathbb{R}}$ を完備に押し出す。

□

[証明].

(1) $E_\theta[f] = 0$ ($\theta \in \mathbb{R}$), すなわち,

$$F(\theta) := \int_{\mathbb{R}} f(x) e^{-\frac{x^2}{2}} e^{\theta x} dx = 0 \quad (\theta \in \mathbb{R})$$

を仮定する. 右辺は任意の $\theta \in \mathbb{C}$ について絶対収束し, 正則であることが判る. 一致の定理より, 特に虚軸上でも $F(iu) = 0$ ($u \in \mathbb{R}$) が必要. 可積分関数の Fourier 変換の単射性より, $f(x) = 0$ a.e. である. 分布族 $\{N(\mu, \nu)\}_{\mu \in \mathbb{R}, \nu \in \mathbb{R}^+}$ についても全く同様の論法が通る.

(2) 統計量 $f(x) := x_1 - x_2$ を考えると,

$$E_\theta[f] = E_\theta[x_1] - E_\theta[x_2] = \theta - \theta_0$$

であるが, 当然あらゆる $N_n(\theta \mathbf{1}_n, I_n)$ について $f = 0$ a.e. ではない. $f(x) := (x_1 - x_2)^2 - 2$ としても, $x_1 - x_2 \sim N(0, 2)$ より, 同様の議論が出来る.

(3) 標本平均 T は, \mathbb{R} 上に 1 次元の正規分布 $\{N(n\theta, n\sigma^2)\}_{\theta > 0}$ を押し出すため, (1) から従う.

■

2.2.7 不偏推定量の一致最小分散性の特徴付け

Rao-Blackwell の定理により, 不偏推定量は十分統計量の関数に粗くすることで分散を改善出来る. が, 不偏推定量は他にも存在するかもしれない, それについての改善はもっと効率が良いかもしれない. この点の解決には, 一致最小分散な不偏推定量を特徴付ける他ない.

定理 2.2.20 (Lehmann-Scheffé). $T: \mathcal{X} \rightarrow \mathcal{T}$ は \mathcal{P} についての完備十分統計量, $S \in \cap_{\theta \in \Theta} L_{P_\theta}^2(\mathcal{X})$ を $g \in L(\Theta)$ の不偏推定量とする. このとき, $g(\theta)$ の任意の不偏推定量 $U: \mathcal{X} \rightarrow \mathbb{R}$ に対して,

$$\text{Var}_\theta[U] \geq \text{Var}_\theta[S^*], \quad (\theta \in \Theta).$$

等号成立条件は $\forall_{\theta \in \Theta} P_\theta[S^* = U] = 1$.

[証明].

方針 U の一致改善 $U^* := E[U|T]$ に対して, Rao-Blackwell の定理から, $\text{Var}_\theta[U] \geq \text{Var}_\theta[U^*]$ ($\theta \in \Theta$) が成り立ち, 等号は $P_\theta[U = U^*] = 1$ ($\theta \in \Theta$) のときに限って成り立つ. これについて, $P_\theta[S^* = U^*] = 1$ ($\theta \in \Theta$) を示せば良い.

証明 いずれも T の可測関数であるから, ある可測関数 $g_1, g_2 \in L(\mathcal{T})$ について, $U^* = g_1(T), S^* = g_2(T)$ と表せる. U^*, S^* の不偏性から, 任意の $\theta \in \Theta$ について $E_\theta[g_1(T) - g_2(T)] = 0$ が必要であるが, T の完備性より $P_\theta[g_1(T) = g_2(T)] = 1$ ($\theta \in \Theta$) が必要である.

■

要諦 2.2.21. 完備十分統計量の関数でもある不偏推定量は本質的に 1 つしかなく, 見つけた時点でそれが一致最小分散である.

2.2.8 Basu の定理

完備十分統計量 (多次元たりえる) と任意の補助統計量は分布族の任意の元の下で独立である, という強い主張である. 「完備十分性」がもっているほしい最終的な性質という意味で, 真に取りを飾るべき定理でもあろう. 無限分解可能性とも関連する定理である.

定義 2.2.22 (ancillary statistic (Fisher)). 統計的実験 $(\mathcal{X}, \mathcal{A}, (P_\theta))$ 上の統計量 $V: \mathcal{X} \rightarrow \mathcal{V}$ が補助統計量であるとは, 押し出して得る分布族 $\{P_\theta^V\}$ が一点集合となることをいう.

例 2.2.23 (Student の t -統計量は分散の補助統計量である). 独立標本 $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ について, 標本平均 \bar{X} と不偏分散 S^2 について,

$$t(X_1, \dots, X_n | \mu) := \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

であり, σ^2 に依らない. □

定理 2.2.24 ([Basu, 1955]). $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ を統計的実験, $T: \mathcal{X} \rightarrow \mathcal{T}$ をその上の完備十分統計量, $V: \mathcal{X} \rightarrow \mathcal{V}$ の分布 P_θ^V は θ に依らないとする. このとき, 任意の $\theta \in \Theta$ に対して, T と V は独立である:

$$P_\theta[T \in A, V \in B] = P_\theta[T \in A]P_\theta[V \in B] \quad (A \in \mathcal{G}, B \in \mathcal{C}, \theta \in \Theta)$$

[証明].

Step1 仮定より, $p_B := P_\theta[V \in B], q_B(T) := P_\theta[V \in B|T]$ は $\theta \in \Theta$ に依らない. これに対して, 条件付き期待値の性質から

$$p_B = E_\theta[1_B(V)] = E_\theta[E_\theta[1_B(V)|T]] = E_\theta[q_B(T)]$$

であるから, $E_\theta[p_B - q_B(T)] = 0$ が従う. 完備性から, $P_\theta[p_B = q_B(T)] = 1$.

Step2 よって, 任意の $\theta \in \Theta$ について,

$$\begin{aligned} P_\theta[T \in A, V \in B] &= E_\theta[1_A(T)1_B(V)] \\ &= E_\theta[1_A(T)E_\theta[1_B(V)|T]] \\ &= E_\theta[1_A(T)q_B(T)] \\ &= E_\theta[1_A(T)p_B] \\ &= E_\theta[1_A(T)]p_B \\ &= P_\theta[T \in A]P_\theta[V \in B]. \end{aligned}$$

注 2.2.25 (Basu の定理の一般化). この定理は, より一般的な, 有界完備な十分統計量について成り立つ.

系 2.2.26. 正規分布 $N(\mu, \sigma^2)$ からの無作為標本の, 標本平均と標本分散は独立である.

[証明]. 次の2点より, Basu の定理から, 標本平均と標本分散は独立である: $\bar{X} \perp\!\!\!\perp S^2$. 同様に, 標本平均と不偏分散も独立である.

標本平均は平均の完備十分統計量である 分布族 $\{N(\mu, \sigma^2)^{\otimes n}\}_{\mu \in \mathbb{R}}$ は指数型であり, 統計量 $T_1(x) := \sum_{i \in [n]} x_i = n\bar{X}$ は θ の完備十分統計量である.

標本分散は補助統計量である 標本分散の分布は

$$S^2 := \frac{1}{n} \sum_{i \in [n]} (X_i - \bar{X})^2 \sim \chi^2(n-1).$$

より, パラメータ $\mu \in \mathbb{R}$ に依らない. ■

注 2.2.27. 実はこの逆も言える [Kawata and Sakamoto, 1949].

2.3 推定論

2.3.1 一般の最良不偏推定量

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. これを古典的には, 一様最小分散不偏推定量という.

定理 2.3.1 (Lehmann-Scheffé (55)). 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, T に関する条件付き期待値 $\delta_0 := E[\delta|T]$ は, $g(\theta)$ の最良の不偏推定量 $\text{BUE} : \forall_{g' \in \Delta_g} \forall_{\theta \in \Theta} R(\theta, \delta_0) \leq R(\theta, g')$.

要諦 2.3.2 (UMVUE). 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」の指標とし、この場合の最良不偏推定量とは UMVUE を意味することとなる。

系 2.3.3. 完備十分統計量の関数であるような不偏統計量が存在するならば、それは一意的であり、また UMV 不偏推定量である。

例 2.3.4 (繰り返しのある二元配置分散分析). 独立な観測 $X_{ijk} \sim N(\mu_{ij}, \sigma^2)$ ($i \in [I], j \in [J], k \in [K]$) について、

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}, \quad \sum_{i \in [I]} \alpha_i = \sum_{j \in [J]} \beta_j = \sum_{i \in [I]} \gamma_{ij} = \sum_{j \in [J]} \gamma_{ij} = 0.$$

とする。 □

2.3.2 正則な統計的実験

- 確率空間 $(\mathcal{X}, \mathcal{A}, P_\theta)$ のエントロピー＝平均情報量とは、対数尤度の平均 $-E[\log L(\theta|x)|\theta]$ であり、系の乱雑さを表す。一様分布で、どの事象も起こり得て予測が難しいとき、値が大きくなる。
- 対数尤度の θ に関する微分 $V(x; \theta)$ の平均は $E[V(x; \theta)|\theta] = 0$ を満たすが (θ がいくら変わろうと確率分布である限り総重量は 1 なので)、この分散に興味がある。もし至る所で変わるなら、分散は極めて大きくなるはずである。Fisher 情報量とはこれであり、分布 P_θ 自身が母数 θ に関してもつ情報の量を表す。

定義 2.3.5 (Fisher's finite information). $L_2(\mathcal{X}, \nu)$ について、

$$I(\theta) := 4 \int_{\mathcal{X}} \left(\frac{\partial}{\partial \theta} p^{1/2}(x; \theta) \right)^2 d\nu = 4 \left\| \frac{\partial}{\partial \theta} p^{1/2} \right\|_\nu^2$$

を Fisher 情報、

$$I(\theta) = 4 \int_{\mathcal{X}} \frac{\partial}{\partial \theta} p^{1/2}(x; \theta) \left(\frac{\partial}{\partial \theta} p^{1/2}(x; \theta) \right)^\top d\nu$$

を Fisher 情報行列という。ここで、関数 $p^{1/2}(-; u)$ が $u = \theta$ で微分可能で導関数が $L_2(\nu)$ 級であることは必ずしも仮定しない。

$$\exists \psi \in L^2(\mathcal{X}, \text{Meas}(\Theta, \mathbb{R}^K)) \int_{\mathcal{X}} |g(x; \theta + h) - g(x; \theta) - (\psi(x; \theta), h)|^2 d\nu = o(|h|^2) \quad (h \rightarrow 0)$$

が満たされるとき、統計的実験 E は $\theta \in \Theta$ で有限な Fisher 情報を持つという。 $\frac{\partial p}{\partial \theta}(x; \theta)$ はこの L^2 -微分の意味で表している。

要諦 2.3.6. 対数尤度関数 $\log L(\theta|x)$ の θ による微分をスコア関数という。

$$V(x; \theta) = \frac{\partial}{\partial \theta} \log L(\theta|x) = \frac{1}{L} \frac{\partial L}{\partial \theta}.$$

θ を増やしたら尤度が上がる場合、スコア関数は正で、さらに尤度が低ければ低いほど $1/L$ により拡大される。これはたしかに情報量の考え方である。実は $E[V(x; \theta)|\theta] = 0$ が成り立つ。そこで、スコア関数の 2 次のモーメント $I(\theta) := E[V(x; \theta)^2|\theta]$ を Fisher 情報量という、これはスコア関数の分散である： $I(\theta) = \text{Var}[V(x; \theta)]$ 。なお、2 つの定義の等価性

$$\begin{aligned} \left(\frac{\partial}{\partial \theta} p^{1/2}(x; \theta) \right)^2 &= \left(\frac{1}{2} p^{-1/2} \frac{\partial}{\partial \theta} p(x; \theta) \right)^2 \\ &= \frac{1}{4} p^{-1}(x; \theta) \left(\frac{\partial p(x; \theta)}{\partial \theta} \right)^2 \end{aligned}$$

による。

定義 2.3.7 (regular statistical experiment). 組 $E := (\mathcal{X}, \mathcal{A}, \{P_\theta\}_{\theta \in \Theta}, \Theta \overset{\text{open}}{\subset} \mathbb{R}^K)$ が、 P_θ が ν -絶対連続であるとき、統計的実験という。 E が Θ 上で正則であるとは、次の 3 条件を満たすことをいう：

- (1) Radon-Nikodym 微分の族 $p(x, -) : \Theta \rightarrow [0, 1]$ は ν -a.e. x に関して連続である。
- (2) 任意の点 $\theta \in \Theta$ において有限な Fisher 情報が存在する。
- (3) 関数 $\psi(-, \theta) \in L^2(\nu)$ は連続。

2.3.3 正則なモデルでの不偏推定量の分散行列の下界

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」の情報理論的限界を、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。

模型 2.3.8. $U \subset \mathbb{R}^q$ で添字づけられたモデル $\{(P_\theta)_{\theta \in U} \subset P(\mathcal{X})$ は μ に支配されており、Radon-Nikodym 微分 p_θ への対応 $p : \Theta \rightarrow L^1(\mathcal{X}; [0, 1])$ は θ について滑らかであるとする。

(1) 関数 $\log p_\theta(x) : \Theta \times \mathcal{X} \rightarrow \mathbb{R}_-$ の微分 $V : \Theta \times \mathcal{X} \rightarrow \mathbb{R}$

$$V(\theta; x) := \begin{bmatrix} \frac{\partial \log p_\theta(x)}{\partial \theta_1} \\ \vdots \\ \frac{\partial \log p_\theta(x)}{\partial \theta_q} \end{bmatrix}$$

をスコアベクトルという。この成分を ψ_i とかく。

(2) スコアベクトルの2次の積率(実は分散である)

$$I(\theta) = E[V(\theta)V(\theta)^\top] = \begin{bmatrix} E[V_1^2] & \cdots & E[V_1 V_q] \\ \vdots & \ddots & \vdots \\ E[V_q V_1] & \cdots & E[V_q^2] \end{bmatrix}$$

を **Fisher 情報行列** という。その成分は

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

と表示される。 $q = 1$ の場合は Fisher 情報量と呼ばれる。

ここで、推定したいパラメータを $g : U \rightarrow \mathbb{R}^p$ とし、偏微分可能と仮定し、Jacobi 行列を $J(\theta) := \frac{\partial g}{\partial \theta} \in M_{pq}(\mathbb{R})$ とする。

命題 2.3.9.

- (1) スコア関数は中心化されている： $E_{p_\theta}[V(\theta)] = 0$ 。
- (2) Fisher 行列は半正定値な対称行列になる。また、対角成分は $I_{ii} \in \mathbb{R}_+$ を満たす。
- (3) (独立観測に対する加法性) $\forall \theta \in U \quad V(\theta) \in L^2(P_\theta; \mathbb{R}^q), E_{p_\theta}[V(\theta)] = 0$ を仮定し、全く同様の条件を満たすモデルを $(\mathcal{X}', \mathcal{H}', (P'_\theta))$ とする。それぞれの Fisher 情報行列を $I(\theta), I'(\theta)$ とすると、直積分布族 $\{P_\theta \times P'_\theta\}_{\theta \in U}$ の Fisher 情報行列は $I(\theta) + I'(\theta)$ となる。

定理 2.3.10 (Cramer-Rao). 上述のモデル $(P_\theta)_{\theta \in U}$ と、偏微分可能なパラメータ $g : U \rightarrow \mathbb{R}^p$ の不偏推定量 $\delta \in L^2(\mathcal{X}; \mathbb{R}^p)$ と、固定された真値 $\theta_0 \in U$ について、次の4条件を仮定する。

- (CR1) モデル $(p_\theta(x))_{\theta \in U}$ は P_{θ_0} -a.e. x について偏微分可能： $P_{\theta_0}[\{p_\theta(x) \text{ は } U \text{ 上偏微分可能}\}] = 1$ 。すなわち、スコアは $p_{\theta_0}(x)$ の台の上では定義されているため、零集合の上での値は自由に定めて良い。
- (CR2) スコアベクトルは中心化されていて $E_{p_{\theta_0}} V(\theta_0) = 0$ 、二乗可積分 $V(\theta_0) \in L^2(P_{\theta_0}; \mathbb{R}^q)$ である。
- (CR3) Fisher 情報行列は θ_0 上正値： $I(\theta_0) > 0$ 。特に可逆である。
- (CR4) 次の微分と積分の可換性が成り立つ：

$$\int_{\mathcal{X}} \delta(x) \psi_i(x; \theta_0) p_{\theta_0}(x) \mu(dx) = \frac{\partial}{\partial \theta_i} \int_{\mathcal{X}} \delta(x) p_{\theta_0}(x) \mu(dx).$$

このとき、パラメータ $g : U \rightarrow \mathbb{R}^p$ の Jacobi 行列 $J(\theta_0)$ について、次の行列不等式が成り立つ：

$$\text{Var}_{\theta_0}[\delta] \geq J(\theta_0) I(\theta_0)^{-1} J(\theta_0)^\top.$$

[証明] 任意に $\theta \in U$ を取り, $I := I(\theta), J := J(\theta)$ と略記する.

共分散への翻訳 仮定 (A4) より,

$$J = \frac{\partial g(\theta)}{\partial \theta} = \frac{\partial}{\partial \theta} E_{\theta}[\delta] = E_{\theta}[\delta \psi^{\top}].$$

これと $E_{\theta}[\psi] = 0$ より,

$$\text{Cov}_{\theta}[\psi, \delta] = E_{\theta}[\psi \delta^{\top}] = J^{\top}.$$

$$I = E_{\theta}[\psi \psi^{\top}] = \text{Var}_{\theta}[\psi].$$

証明 すると, Cauchy-Schwarz の不等式同様, $\text{Var}_{\theta}[\delta - JI^{-1}\psi] \geq 0$ であることから, Cov の双線型性のみから従う. また, 対称行列 S に対して, $S^{-1} = (S^{-1})^{\top} = (S^{\top})^{-1}$ であることと Fisher 情報行列が対称であることに注意すると,

$$\begin{aligned} 0 &\leq \text{Var}_{\theta}[\delta - JI^{-1}\psi] = \text{Cov}[\delta - JI^{-1}\psi, \delta - JI^{-1}\psi] \\ &= \text{Var}_{\theta}[\delta] - JI^{-1}\text{Cov}_{\theta}[\psi, \delta] - \text{Cov}_{\theta}[\delta, \psi]I^{-1}J^{\top} + JI^{-1}\text{Var}[\psi]I^{-1}\psi^{\top} \\ &= \text{Var}_{\theta}[\delta] - JI^{-1}J^{\top}. \end{aligned}$$

■

要諦 2.3.11. 本質的には Cauchy-Schwarz の不等式である. \mathbb{R}^q の内積を $(-|-)$, $\mathcal{L}^2(P_{\theta})$ の内積を $\langle -|-\rangle$ で表すと, $\|\langle \psi|\delta \rangle\|^2 \leq (\|\psi\|^2\|\delta\|^2)$.

2.3.4 バイズ推定

独自のバイズリスクと呼ばれる目的関数を最小化する決定関数をバイズ決定関数という 2.1.16. バイズ決定関数を修正すると, ミニマックスの意味で最適/最良な決定関数を得る.

定理 2.3.12. $\delta \in \Delta$ を Bayes 決定関数とする. $R(\theta, \delta)$ が $\theta \in \Theta$ に関して定値ならば, これはミニマックス決定関数でもある.

定理 2.3.13. δ_n を事前分布 π_n に関する Bayes 決定関数とする. 決定関数 δ が

$$\forall \theta \in \Theta \quad R(\theta, \delta) \leq \limsup_{n \rightarrow \infty} R(\pi_n, \delta_n)$$

を満たすならば, δ はミニマックス決定関数でもある.

2.3.5 James-Stein 推定量

σ^2 を既知として, 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 不偏推定量であった. しかし, 決定関数の空間 Δ_g を不偏推定量の空間から拡張し, 一般の推定量とすると, $p \geq 3$ のとき非許容である. このときの \bar{x} よりも一様に良い推定量には, James-Stein 推定量がある.

補題 2.3.14 (Stein's equation). $x, \theta \in \mathbb{R}^p, F: \mathbb{R}^p \rightarrow \mathbb{R}$ は微分可能で, F, F' は緩増加であるとする. このとき,

$$\int_{\mathbb{R}^p} \frac{\partial F}{\partial x^i}(x) \phi(x; \theta, \sigma^2 I_p) dx = \int_{\mathbb{R}^p} F(x) \frac{x^i - \theta^i}{\sigma^2} \phi(x; \theta, \sigma^2 I_p) dx.$$

要諦 2.3.15 (部分積分という手法). 証明は部分積分によるが, $p = \infty$ のときの同様の公式は Malliavin 解析が与える. この形の等式は, 他に, 漸近展開を導く際の, 対象となる確率変数 (あるいは確率微分方程式に関連した汎関数) の特性関数の評価に用いられる.

定理 2.3.16 (James-Stein (55)).

$$\hat{\theta}_{JS} = \left(1 - \frac{\sigma^2(p-2)}{n|\bar{x}|^2}\right) \bar{x}$$

は \bar{x} を二乗誤差損失関数について一様に改善する.

2.4 検定論

最も自然な検定として、確率の比 $\frac{P[\theta \in \Theta_1]}{P[\theta \in \Theta_0]}$ が十分大きいときに H_0 を棄却するタイプの検定が考えられる。

2.4.1 単純仮説同士の検定における尤度比検定の特徴付け

2つの単純仮説の間の検定では、尤度比検定であることと水準 $\alpha \in (0, 1)$ の最強力検定であることが同値になる。これは Fisher と代わって「対立仮説」の概念を顕在化させ、危険関数を用いるのではなく「検出力」で検定の良さを評価する枠組みを採用したことで浮き彫りになった消息である。

模型 2.4.1 (単純仮説同士の検定). モデル $\mathcal{P} = \{P_\theta\}_{\theta \in \Theta} \subset P(\mathcal{X})$ は、ある σ -有限測度 μ に関する微分 p_0, p_1 を持つとし、 $H_0 : \theta = 0$ を有意水準 $\alpha \in (0, 1)$ で検定する。

定理 2.4.2 (Neyman-Pearson). 上述の模型について、

- (1) サイズ α で、尤度比 p_1/p_0 がある $k \in \mathbb{R}_+$ よりも大きいときに H_0 を棄却する、非確率的な検定 $\varphi_k \in L(\mathcal{X}; [0, 1])$ が存在する：

$$E_0[\varphi] = \alpha, \quad \varphi_k(x) = \begin{cases} 1 & p_1(x) > k p_0(x), \\ 0 & p_1(x) < k p_0(x). \end{cases}$$

- (2) 逆に、ある $k \in \mathbb{R}_+$ に対して上の条件を満たす検定 $\varphi_k \in L(\mathcal{X}; [0, 1])$ は水準 α の最強力検定である。これは $\alpha = 0, 1$ の場合も成り立つ。

- (3) $\varphi \in L(\mathcal{X}; [0, 1])$ を水準 α の最強力検定であるとする。このとき、 $\exists k \in \mathbb{R}_+ \varphi = \varphi_k \mu$ -a.e.

例 2.4.3 (二項分布族における検定). モデル $\mathcal{P} = \{B(n, \theta)\}_{\theta \in (0, 1)}$ を考える。

- (1) 問題 $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$ ($\theta_0 < \theta_1$) を検定する。 $n+1$ 上の計数測度を μ とすると、 \mathcal{P} はこれに支配され、尤度比は

$$\log \frac{p_1(x)}{p_2(x)} = ax + n \log \frac{1 - \theta_1}{1 - \theta_0}, \quad a := \log \frac{\frac{\theta_1}{1 - \theta_1}}{\frac{\theta_0}{1 - \theta_0}} > 0$$

と表せるから、

$$\varphi(x) = \begin{cases} 1 & x > x_0, \\ \gamma & x = x_0, \\ 0 & x < x_0. \end{cases} \quad (x_0, \gamma \text{ は } E_{\theta_0}[\varphi] = \alpha \text{ を満たすように定める})$$

と定めると、これは水準 α の最強力検定である。

- (2) ここで、 φ は θ_1 に依存していないから、 $H_1 : \theta > \theta_0$ という複合仮説を対立させても φ_1 は一様最強力であることがわかる。これは、Neyman-Pearson の基準が、 Θ_0 上での危険関数の値を考慮に入れていないために起こる。

□

2.4.2 複合仮説同士の検定と単調尤度比

模型 2.4.4 (T -単調な尤度比を持つモデル). ある σ -有限測度 μ に支配されるモデル $\mathcal{P} = (P_\theta)_{\theta \in \Theta} \subset P(\mathcal{X})$ ($\theta \in \mathbb{R}$) において、検定問題 $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$ ($\theta_0 \in \Theta$) を考える。 $\Theta_1 := \{\theta \in \Theta \mid \theta > \theta_0\} \neq \emptyset$ を仮定し、さらに次を仮定する：

モデル \mathcal{P} が可測関数 $T \in L(\mathcal{X})$ に関して単調な尤度比を持つ (monotone likelihood ratio) とは、 $\frac{dP_\theta}{d\mu}$ のバージョン $p_\theta : \mathcal{X} \rightarrow [0, 1]$ が存在して、

$$\forall \theta_1, \theta_2 \in \Theta \quad \theta_1 < \theta_2 \Rightarrow \left(\exists_{H_{\theta_1, \theta_2} : \text{Im } T \rightarrow \overline{\mathbb{R}_+} : \text{単調増加}} \frac{p_{\theta_2}}{p_{\theta_1}} = H_{\theta_1, \theta_2}(T) \text{ on } \mathcal{X} \setminus \{p_{\theta_1} = p_{\theta_2} = 0\} \right)$$

を満たすことをいう。

例 2.4.5 (指数型分布族が T -単調な尤度比を持つための十分条件). 1-次の指数型分布族 $(P_\theta)_{\theta \in \Theta}$ ($\Theta \subset \mathbb{R}$) は

$$p_\theta(x) = g(x)e^{a(\theta)T(x) - \psi(\theta)}$$

と表せるとする. 関数 $a: \Theta \rightarrow \mathbb{R}$ が単調増加ならば, \mathcal{P} は T -単調な尤度比を持つ分布族である. \square

定理 2.4.6 (単調尤度比検定の特徴付け). 分布族 $\{P_\theta\}_{\theta \in \Theta}$ は $T \in L(\mathcal{G})$ に関して単調な尤度比を持ち, $\Theta_1 \neq \emptyset$ を満たす検定問題 $H_0: \theta \leq \theta_0$ vs. $H_1: \theta > \theta_0$ ($\theta_0 \in \Theta$) を考える.

(1) 次の検定 φ_0 が, 水準 $E_{\theta_0}[\varphi_0]$ の一様最強力検定を与える:

$$\varphi_0(x) = \begin{cases} 1 & T(x) > c, \\ \gamma & T(x) = c, \\ 0 & T(x) < c. \end{cases} \quad c \in \mathbb{R}, \gamma \in [0, 1].$$

(2) 任意の水準 $\alpha \in (0, 1)$ に対して, ある $c \in \mathbb{R}, \gamma \in [0, 1]$ が存在して, これが定める φ_0 が水準 α の一様最強力検定を与える.

要諦 2.4.7. 双対問題 $H_0: \theta \geq \theta_0$ vs. $H_1: \theta < \theta_0$ に対しても, 分布族のパラメータを $\{P_{-\theta}\}_{\theta \in \Theta}$ と取り替えて, $-T$ に関する単調尤度比を持つモデルとみなせばよい.

2.4.3 一般化された Neyman-Pearson の補題

定理 2.4.8 (generalized Neyman-Pearson). $(\mathcal{X}, \mathcal{A}, \mu)$ を σ -有限測度空間, $f_1, \dots, f_m, g \in L^1(\mathcal{X}, \mathcal{A}, \mu)$ とし,

$$\Phi_c := \left\{ \varphi \in \Phi_1 \mid \forall i \in [m] \int_{\mathcal{X}} \varphi(x) f_i(x) d\mu = c_i \right\} \neq \emptyset \quad (c \in \mathbb{R}^m)$$

と表す.

(1) 検定 $\varphi_0 \in \Phi_c$ がある $k \in \mathbb{R}^m$ に対して

$$\varphi_0(x) = \begin{cases} 1 & g(x) > \sum_{i \in [m]} k_i f_i(x), \\ 0 & g(x) < \sum_{i \in [m]} k_i f_i(x). \end{cases} \quad \mu \text{ a.e. } x$$

を満たすならば, 次が成り立つ:

$$\int_{\mathcal{X}} \varphi_0(x) g(x) d\mu = \sup_{\varphi \in \Phi_c} \int_{\mathcal{X}} \varphi(x) g(x) d\mu.$$

(2) さらに, $k \in \mathbb{R}_+^m$ と選べるならば, 次が成り立つ:

$$\int_{\mathcal{X}} \varphi_0(x) g(x) d\mu = \sup \left\{ \int_{\mathcal{X}} \varphi(x) g(x) d\mu \in \mathbb{R} \mid \varphi \in \Phi, \forall i \in [m] \int_{\mathcal{X}} \varphi(x) f_i(x) d\mu \leq c_i \right\}.$$

2.4.4 不偏検定の特徴付け

最適な不偏検定を見つけるには, 相似検定の中から最適なものを探せば十分.

定義 2.4.9 (similar). 検定 $\varphi \in \Phi_1$ が部分集合 $\Theta' \subset \Theta$ に対して相似であるとは, $E_\theta[\varphi]$ が Θ' 上定値であることをいう.

補題 2.4.10 (相似検定は不偏検定であるための必要条件). $\Theta' := \partial\Theta_0 \cap \partial\Theta_1 \subset \Theta$ とし, 検定問題 $H_0: \theta \in \Theta_0$ vs. $H_1: \theta \in \Theta_1$ を考える. 任意の不偏検定 $\varphi \in \Phi^u$ について, 関数 $E_-[\varphi]: \Theta \rightarrow \mathbb{R}$ が連続ならば, φ は Θ' -相似である.

補題 2.4.11 (相似検定が不偏であるための条件). $\emptyset \subsetneq \Theta' \subset \Theta, \alpha \in (0, 1)$ について, Θ' 上の相似検定の全体を

$$\Phi'_\alpha := \{ \varphi \in \Phi \mid \forall \theta \in \Theta' E_\theta[\varphi] = \alpha \}$$

とおく. $\varphi_0 \in \Phi'_\alpha$ がさらに次の2条件を満たせば, 水準 α -不偏検定である:

- (1) $\forall \theta \in \Theta_0 \quad \forall \varphi \in \Phi'_\alpha \quad E_\theta[\varphi_0] \leq E_\theta[\varphi].$
 (2) $\forall \theta \in \Theta_1 \quad \forall \varphi \in \Phi'_\alpha \quad E_\theta[\varphi_0] \geq E_\theta[\varphi].$

2.4.5 帰無仮説の補集合となる対立仮説の検定

模型 2.4.12. $\Theta \subset \mathbb{R}^m$ ($m \geq 2$) を開矩形 (開区間の積) 集合とする. m -次の自然母数 $\theta \in \Theta$ を持つ指数型分布族

$$\frac{dP_\theta}{d\mu}(x) = g(x) \exp \left(\sum_{i \in [m]} \theta_i T_i(x) - \psi(\theta) \right) \mu\text{-a.e.} x$$

上での検定問題 $H_0 : \theta_1 = b$ vs. $H_1 : \theta_1 \neq b$ を考える. これに伴って, 記法

$$t^* := (t_2, \dots, t_m), \quad T^* := (T_2, \dots, T_m), \quad \theta^* := (\theta_2, \dots, \theta_m), \quad \Theta^* := \{\theta^* \in \mathbb{R}^{m-1} \mid \theta \in \Theta\}$$

を導入する. Θ^* は局外母数の空間である.

定理 2.4.13 (一様最強力不偏検定の構成). 任意の $\alpha \in (0, 1)$ に対して, $\gamma_1, \gamma_2, u_1, u_2 \in L(\text{Im } t^*)$ が存在して, 次の $\varphi_0 \in \Phi'_\alpha$ は水準 α -の一様最強力不偏検定である:

$$\varphi_0(x) = \begin{cases} 1 & T_1(x) < u_1(T^*(x)) \vee T_1(x) > u_2(T^*(x)), \\ \gamma_1(T^*(x)) & T_1(x) = u_1(T^*(x)), \\ \gamma_2(T^*(x)) & T_1(x) = u_2(T^*(x)), \\ 0 & u_1(T^*(x)) < T_1(x) < u_2(T^*(x)). \end{cases}$$

例 2.4.14. モデル $\{N(\theta, \sigma^2)\}_{\theta \in \mathbb{R}}$ に対して, 検定 $H_0 : \theta = b$ vs. $H_1 : \theta \neq b$ を考える. n 個の標本の分布は局外母数のない 1-パラメータ指数型分布族

$$g(x) \exp \left(\theta T - \frac{n\theta^2}{2\sigma^2} \right), \quad T = \frac{1}{\sigma^2} \sum_{j \in [n]} x_j.$$

となる. このとき, 定理から

$$\varphi_0 = 1_{\{T \leq u_1\} \cup \{T \geq u_2\}}$$

の形の一様最強力不偏検定が存在するから, あとは u_1, u_2 の値を決めれば良い. □

補題 2.4.15 (不偏にならない場合). $P_0, P_1 \in P(\mathbb{R})$ は σ -有限測度 μ に支配され, 微分のあるバージョン f_0, f_1 は $f_0 > 0$ μ -a.e. かつ f_1/f_0 は狭義単調増加であるとする. このとき, $E_0[\varphi] \in (0, 1)$ なる検定関数 $\varphi \in L(\mathbb{R}; [0, 1])$ に対して, φ が片側であり $\exists \alpha \in \mathbb{R} \quad \varphi((-\infty, \alpha)) = \{0\} \wedge \varphi(\alpha, \infty) = 1$, P_0 は δ_α ではないならば, 次が成り立つ:

$$\int \varphi dP_0 < \int \varphi dP_1.$$

2.4.6 両側 t -検定

正規分布モデルの (分散が未知の下での) 平均の検定は t -検定になるのであった 1.4.9 が, これは一様最強力不偏検定になる.

問題 2.4.16. モデル $\{N(\mu, \sigma^2)^{\otimes n}\}_{\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+}$ に対して, 平均母数の検定 $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$ を考える. 確率密度関数は, パラメータの取替について

$$p_\theta(x) = e^{\theta_1 T_1(x) + \theta_2 T_2(x) - \psi(\theta)}, \quad \left(\theta_1 := \frac{\mu - \mu_0}{\sigma^2}, \theta_2 := -\frac{1}{2\sigma^2}, \theta := (\theta_1, \theta_2) \right)$$

$$T_1(x) = \sum_{j \in [n]} (x_j - \mu_0), \quad x = (x_1, \dots, x_n),$$

$$T_2(x) = \sum_{j \in [n]} (x_j - \mu_0)^2, \quad \psi(\theta) := -n \frac{\theta_1^2}{4\theta_2} + \frac{n}{2} \log \left(-\frac{\pi}{\theta_2} \right).$$

と表せ, 問題は $H_0 : \theta_1 = 0$ vs. $H_1 : \theta_1 \neq 0$ に変換される. こうして, 定理から一様最強力検定が構成できる.

命題 2.4.17.

$$T := \frac{\sqrt{n-1}(\bar{x} - \mu_0)}{S}, \quad \left(\bar{x} = \frac{1}{n} \sum_{i \in [n]} x_i, S^2 = \frac{1}{n} \sum_{j \in [n]} (x_j - \bar{x})^2 \right)$$

は H_0 の下で $t(n-1)$ に従い, t -検定 $1_{\{|T| \geq t_{n-1}(\alpha)\}}$ は一様最強力不偏検定になる.

2.4.7 不変検定

前項の両側 t -検定も不変検定であった [2.1.26](#).

命題 2.4.18. 任意の最大不変量 $S: \mathcal{X} \rightarrow \mathcal{S}$ と G -不変統計量 $T \in L(\mathcal{X}; \mathcal{T})$ について,

- (1) ある可測関数 $f: \text{Im } S \rightarrow \mathcal{T}$ が存在して $T = f \circ S$ を満たす. すなわち, T は S の可測関数である.
- (2) T の分布は \overline{G} に対する最大不変量 $\tau: \Theta \rightarrow \mathcal{S}$ のみに依存する.

例 2.4.19.

- (1) $G := \text{Aut}_{\text{Set}}([n])$ を置換群とする. 順序統計量 $S = (x_{(1)}, \dots, x_{(n)})$ は最大不変量であるから, 任意の G -不変統計量 T は S の関数として表される.
- (2) $\mathcal{X} := \mathbb{R}^n \setminus \{x \in \mathbb{R}^n \mid \exists i \neq j \in [n] x_i = x_j\}$ とし, G を狭義単調増加連続関数 $g: \mathbb{R} \rightarrow \mathbb{R}$ の全体のなす群とすると, 順位統計量 $S(x) = (r_1, \dots, r_n)$ は最大不変量である. ただし, $r_i := |j \in [n] \mid x_j \geq x_i|$ を順位とした. よって, 任意の G -普遍統計量 $T: \mathcal{X} \rightarrow \mathbb{R}$ は S の関数である.

□

第 3 章

種々の統計手法とその漸近論的解釈

Bhattacharya [Bhattacharya et al., 2016] に従って、種々の統計手法の漸近的性質を調べる。これを用いれば、検定や信頼領域を構成出来る。精密標本理論と違い、種々の統計的汎関数に対する統一的な議論が可能という点で見通し通い。漸近的不偏性たる一致性が指導原理となる。

データ $X: \Omega \rightarrow \mathcal{X}$ がどのような規則 $X \sim P$ に従って生成されているかを推定し、将来のデータを予測する問題を考える。

- (1) モーメント法は一般のモデルに対して使える普遍的な手法で、さらに局外母数を置いたままの推定を可能にする。
- (2) χ^2 -検定はデータの訊問などにおいて、標本から検定を構成する普遍的な手法である。

データ $X: \Omega \rightarrow \mathcal{X}$ の従う規則 $X \sim P$ を、前兆となる観測 Y から説明し、将来のデータを予測する問題を考える。

- (1) あるクラスの間数関係にノイズがかかったものとしてフィッティングする手法を、回帰分析という。
- (2) ノイズを想定せずに関数関係を推定する手法を、統計的学習という。

3.1 モーメント法による母数推定

モーメント法

パラメータが積率の言葉で特徴付けられるモデルにおいては、標本モーメントを通じて、母集団の積率を推定し、そこからパラメータを推定する 2 段階化が考え得る。これをモーメント法という。

- (1) バイアスがあるが、一致推定量で、計算が簡単である。
- (2) そこでモーメント法推定量から逐次改善を通じて、最尤推定の代わりとすることは十分ありえる。
- (3) 特にモデルが未確定な場合である。未確定な局外母数の空間を持つセミパラメトリックモデルでは、これを一般化した一般化モーメント法 (GMM) が用いられる。
- (4) 十分統計量ではないことがある。

3.1.1 標本平均

模型 3.1.1. $X_1, \dots, X_n \in \mathcal{X}$ を $P \in \mathcal{P}^2(\mathcal{X})$ の観測とする。

$$\bar{X}_n := \frac{\sum_{i \in [n]} X_i}{n}$$

を標本平均という。このとき、

$$E[\bar{X}] = \alpha_1(P), \quad \text{Var}[\bar{X}] = \frac{\mu_2(P)}{n}.$$

命題 3.1.2. 上の模型について、

- (1) \bar{X} は $\alpha_1(P)$ の一致推定量である。

[証明].

- (1) $\bar{X} \rightarrow \alpha_1 \in L^2(\mathcal{X}^\infty)$ が示せる。

3.1.2 標本積率

模型 3.1.3. $X_1, \dots, X_n \in \mathcal{X}$ を $P \in P^k(\mathcal{X})$ ($k \in \mathbb{N}^+$) の観測とする.

$$\hat{m}_r := \mathbb{P}_n[X^r] = \frac{1}{n} \sum_{j=1}^n X_j^r \quad (r \in [k])$$

を r -標本積率という.

命題 3.1.4. 上の積率について,

- (1) \hat{m}_r は $\beta_r(P)$ の不偏な強一致推定量である.
- (2) $\hat{\mu}_r$ は $\mu_r(P)$ の強一致推定量であるが, 不偏推定量ではない.

3.1.3 一般化モーメント法

一般化モーメント法は Z -推定量の例であり, 一致性・漸近正規性を持ち, また漸近最適である.

要諦 3.1.5. moment condition, orthogonality condition (Sargan, 1958, 1959), unbiased estimating equation (Huber, 1967; Wang et al., 1997) などと呼ばれる条件は, ある関数 $g: \mathcal{X} \times \Theta \rightarrow \mathbb{R}^p$ を用いて $E[g(Y_t, \theta_0)] = 0$ と表せる p 次元の条件と, 識別可能性の組である. この値 $m(\theta) := E[g(Y_t, \theta)]$ を一般化モーメント, g をモーメント関数というのである. そして, この値を 0 にする (等価であるが, 何らかのノルムで測り最小化する) $\hat{\theta}$ を推定値とする.

注 3.1.6. これは, 任意の関数 $g: \mathcal{X} \rightarrow \mathbb{R}^p$ を用いて

$$\psi(x, \theta) := g(x) - \int_{\mathcal{X}} g(y) P_{\theta}(dy)$$

という形の推定関数が定める Z -推定量である, ということになる.

3.2 χ^2 -検定

漸近最適な検定統計量の殆どは χ^2 -分布に従うことに基づく, 広く一般化した統計手続きの呼び名である.

3.2.1 概要

模型 3.2.1. ノンパラメトリックな実験 $(\mathcal{X}, \mathcal{A}, P(\mathcal{X}))$ において, 標本空間 \mathcal{X} の有限な分割 $\Omega = \sum_{i \in [k]} A_i$ と標本 X_1, \dots, X_n を考える. このとき, $P(\mathcal{X})$ は多項分布の族 $\{(P[A_i])_{i \in [k]}\}_{P \in P(\mathcal{X})} \subset P([k])$ を定める. なお,

$$\Delta := P([k]) := \left\{ p \in \mathbb{R}_+^k \mid \sum_{i \in [k]} p_i = 1 \right\}$$

である. この多項分布族 Δ に関して, 真の分布 $P_0 \in P(\mathcal{X})$ が定めるものを $p_0 := \text{Mult}(p, (P_0[A_i])_{i \in [k]})$ とする.

問題 3.2.2. 上述のモデル $([k], \Delta)$ と真値 $p \in \Delta$ において,

- (1) 単純仮説の検定

$$H_0: p = (p_i^*)_{i \in [k]} \text{ vs. } H_1: p_0 \neq (p_i^*)_{i \in [k]}$$

を適合度検定という．この場合の検定統計量

$$Q_n := \sum_{i \in [k]} \frac{(O_i - n \cdot P[A_i])^2}{n \cdot P[A_i]}, \quad O_i := \# \{j \in [n] \mid X_j \in A_i\}.$$

は漸近的に $\chi^2(k-1)$ に従う．

(2) $k = k_1 k_2, p = (p_{ij})_{(i,j) \in [k_1] \times [k_2]}$ とし，複合仮説の検定

$$H_0 : p \in \mathcal{P} \text{ vs. } H_1 : p \notin \mathcal{P}, \quad \mathcal{P} := \{p \in \Delta \mid (p_{i-})_{i \in [k_1]} \perp (p_{-j})_{j \in [k_2]}\}$$

を独立性の検定という．この場合の検定統計量は漸近的に $\chi^2((k_1-1)(k_2-1))$ に従う．

いずれの場合も，多項分布については，この統計量 Q_n は尤度比検定と漸近同等である．

要諦 3.2.3. Cramer 1999 によると， χ_n^2 が χ^2 -分布に従うと十分近似出来るには，条件 $\forall i \in [p] \ nP[A_i] \geq 10$ を満たす必要があるとしている．これが満たされない場合は尤度比検定の 1 つである G -検定を用いることが適切になる．また，そもそも標本数 n が小さい場合は，二項検定，さらに $(k_1, k_2) = (2, 2)$ であるとき，Fisher の正確確率検定を用いることが適切になる．

応用例 3.2.4. Q_n をカイ自乗値という．行列 $(np_{ij})_{(i,j) \in [k_1] \times [k_2]} \in M_k(\mathbb{R})$ を分割表 (cross table) という．

3.2.2 Yates の補正

$(k_1, k_2) = (2, 2)$ における独立性の検定において，Yates の補正 (1934, [Yates, 1934])

$$Q_n^* := \sum_{i \in 4} \frac{(|O_i - n \cdot P[A_i]| - 0.5)^2}{n \cdot P[A_i]}$$

は， $\exists i \in [4] \ np_i \leq 5$ を満たす際に，第一種の過誤の確率を下げ，結果として近似精度を上げることがある．

3.2.3 G -検定

3.2.4 Fisher の正確検定

3.2.5 種々の標本統計量

記法 3.2.5. X_1, \dots, X_n を $P \in \mathcal{P}(\mathbb{R})$ からの標本とする．

$$(1) \ U^2 := \frac{1}{n-1} \sum_{i \in [n]} (X_i - \bar{X}) \text{ とする.}$$

命題 3.2.6. $\beta_2(P) < \infty$ とする．このとき， S^2 の漸近分布は $N\left(\sigma^2, \frac{\mu_4 - \sigma^4}{n}\right)$ ．

3.3 順位検定

3.3.1 順序統計量の分布

定理 3.3.1. X_1, \dots, X_n を独立同分布列，その分布関数を F とすると， r 番目の順序統計量 $X_{n:r}$ の分布関数 F_r は次のように表せる：

$$F_r(x) = \sum_{i=r}^n \binom{n}{i} F(x)^i (1-F(x))^{n-i}.$$

[証明]. $X_{n:n+1} = \infty$ と定めると，事象 $\{X_{n:r} \leq x\} = \{x \in [X_{n:r}, \infty)\}$ を，どの区間 $[X_{n:i}, X_{n:i+1}]$ ($i = r, \dots, n$) に入るかで場合分けして考えることより，

$$F_r(x) = P[X_{n:r} \leq x] = \sum_{i=r}^n P[X_{n:i} \leq x < X_{n:i+1}]$$

であるが,

$$\{X_{n:i} \leq x < X_{n:i+1}\} = \{x \text{ 以下の観測が } i \text{ 個}\}$$

に注意すれば, 右辺は $Y := 1_{\{X \leq x\}} \sim B(n, p)$ に関する事象と見て, その確率を計算出来る:

$$P[X_{n:i} \leq x < X_{n:i+1}] = \binom{n}{i} F(x)^i (1 - F(x))^{n-i}.$$

系 3.3.2. X_1 の分布は絶対連続とする. このとき, $X_{n:r}$ も絶対連続で,

$$f_r(x) = r \binom{n}{r} F(x)^{r-1} (1 - F(x))^{n-r} f(x).$$

[証明]. F_r の微分による.

3.3.2 順位相関係数

相関係数に順位統計量を用いることで, 分布に対する仮定を取り去ることができる.

歴史 3.3.3 (Charles Spearman 1863-1945). 陸軍士官を辞めて実験心理学に転向した英国の心理学者で, 当時哲学の一分野からやっと経験主義的な兆しを見せ始めていたことに興味を持ち, 独の Wundt に学んだ. この文脈から因子分析なる手法を初めて考察した [Spearman, 1904] 人となり, University College of London の哲学教授から, 心理学部が創設されるとその初代教授となる. Galton の研究もよく見ており, 相関係数の考え方を発展させて開発した順位相関係数に名を残すことになる. Pearson には認められないが, 王立協会会員となる.

歴史 3.3.4 (Maurice Kendall 07-83).

定義 3.3.5. X, Y を確率変数とする.

- (1) $\rho(X, Y) := \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X]\text{Var}[Y]}}$ を Pearson の (中心) 積率相関係数という.
- (2) 標本相関係数とは, 各データの組 (X_i, Y_i) について平均を取ったものをいう.
- (3) 標本 $((X_i, Y_i))_{i \in [n]}$ が定める順位付け関数 $R: [n] \rightarrow [n]$ について, $\rho(R(X_i), R(Y_i))$ を Spearman の順位相関係数という.

一方で, 次を Kendall の順位相関係数という:

$$\tau := \frac{K - L}{\binom{n}{2}}, \quad K := \#\{ \{i, j\} \in \binom{[n]}{2} \mid (X_i > X_j \wedge Y_i > Y_j) \vee (X_i < X_j \wedge Y_i < Y_j) \}, L := \#\{ \{i, j\} \in \binom{[n]}{2} \mid (X_i - X_j)(Y_i - Y_j) < 0 \}.$$

補題 3.3.6. $((X_i, Y_i))_{i \in [n]}$ を標本とする.

- (1) 各 $\{X_i\}$ と各 $\{Y_i\}$ が重複のない離散値であった場合, Spearman の順位相関係数は次に等しい:

$$r_s = 1 - \frac{6 \sum_{i \in [n]} d_i^2}{n(n^2 - 1)}, \quad d_i := R(X_i) - R(Y_i).$$

第 4 章

種々の模型の統計化

古くから一方から他方を説明する $f(X) = Y$ などという模型 (方程式) は物理学・経済学の分野で考えられてきた。これに確率論的な不確定性を許して **fitting** する行為も、統計学の枠組みで議論出来る。代表的な営みが回帰分析と呼ばれており、事実、殆どのパラメトリックモデルは回帰模型である。事実、不確定性と計算コストがトレードオフであるから、「より厳密な把握」より、「不確定性に対する正しい知識」の方が筋が良い場面が多い。すると、方程式 $f(X) = Y$ での説明と、誤差項 ϵ へのしわ寄せとのトレードオフの問題になる。例えば方程式形における最良の知識は、条件付き期待値 $E[Y|X]$ にほかならない。そして、 (X, Y) の確率分布族を指定することは、 $E[Y|X]$ の関数形を指定することに同値であるから、結局回帰分析は統計的決定問題に等価である。あるいは、 Y_1, \dots, Y_n が独立同分布の仮定に違反する際に、共変量 X と抱き合わせて独立同分布性を回復する技術だともみれる。

4.1 多変量に対する模型と変数選択

多変量分析の殆どは、明示的・暗示的な方向性を持った確率ベクトルの組 $(X, Y) \in \mathbb{R}^{n_1+n_2}$ に対して、線型な回帰関係 $Y = \beta X + \epsilon$ を考える、という枠組みの中に入る。ここで種々多様な問題が生じる。

4.1.1 回帰模型の概観

確率的なデータ (X, Y) に対して、条件付き期待値 $E[Y|X = x]$ と残差 ϵ とに分解する模型を回帰模型といい、前者を回帰関数という。特に、 (X, Y) が多変量正規分布に従うとき、回帰関数は線型になる。このときの計画行列 $X \in M_{np}(\mathbb{R})$ の設計を変数選択という。

歴史 4.1.1 (Galton 「平均への回帰」). 父の身長が高い場合、息子の身長は低い確率が高く、父親の身長が低い場合、息子の身長はそれよりは高い場合が多い。これは「人間の体格は世代を経る毎に平均人に回帰していく」ことを意味するか？は非自明であるが、人間の身長が正規分布すると仮定したとしても、これを「発見」として帰無仮説を棄却することは出来ない、ある意味で自然な観測である。

例 4.1.2 (線型模型の重要性). 相関係数は、線型関係の模型からの乖離を測る。相関係数が低いからと行って関係がないとは言えない。しかし、非線型への対応として、データに変換を施し、その後の相関係数を測るということは可能であるから、やはり線型模型の議論が肝心部分になる。□

要諦 4.1.3. 一般に条件付き期待値 $E[Y|X = x]$ の関数形はわからないから、まずは線型回帰模型を考えることとなる。そしてその結果を見て、残差を精査したりして、検定にかかる (残差系列によるモデル診断) ことでデータの訊問を進めるのである。

定理 4.1.4 (Gauss-Markov). 確率変数 Y は、ある確率変数 $\epsilon \sim (0, \sigma^2 I_n)$ と $X \in M_{np}(\mathbb{R}), \beta \in \mathbb{R}^p$ を用いて $Y = X\beta + \epsilon$ の関係を持つと仮定し、 Y_1, \dots, Y_n をその独立同分布列とする。

- (1) β の最小自乗推定量は最良線型不偏推定量になる。
- (2) 任意の β の推定可能な汎関数 $g(\beta)$ についても同様。

(3) $\epsilon \sim N_n(0, \sigma^2 I_n)$ ならば、最小自乗推定量が全ての不偏推定量の中で最適であり、最尤推定量に一致する。

4.1.2 その後の研究の歴史

歴史 4.1.5 (経済学).

(1) Harold Hotelling 95-73 米

- (a) 元はジャーナリズムを学んでいたが、数学修士 (Washington), 数学博士 (Princeton) をとり、数理経済学者の先駆けとなる。初めは Stanford 大学で数学を教えるが、Columbia 大学経済学部教授に移り、Milton Friedman や Kenneth Arrow を育てる。
- (b) Fisher の著作 [Fisher, 1925] に莫大な影響を受け、統計の分野で主成分分析と正準相関分析とを開発 [Hotelling, 1935], [Hotelling, 1936] する。また、多くの大学で統計学部を創設する流れを巻き起こすリーダーシップを執った。
- (c) なお、Fisher と学術的な交流を根気強く続け、キュムラントの語は Thiele がデンマーク後で用いていた半不変量 (semi-invariant) に当たることを指摘した。
- (d) Henry Mann や Abraham Wald らがナチ政権から亡命する際に Columbia 大学在学期間中 31-42 に迎え入れ、そこで Wald は逐次解析 (sequential analysis) と統計的決定理論の研究成果を上げる。Hotelling はこれを "pragmatism in action" と呼んで評価した。

4.1.3 主成分抽出の手法

問題 4.1.6 (低階数近似の問題). 主成分分析、因子分析を実行するのに必要なデータは次の通り：

- (1) 説明変数 $X \in M_{np}(\mathbb{R})$ は $1_n^\top X = 0_p^\top$ を満たす列平均が零な行列とする。
- (2) この変数の数 (行数) を $m \leq p$ まで減らしたいとする。
- (3) 既存の説明変数 X の線型結合をいくつか取ることで、有効な説明変数を取り直したい状況がほとんどの PCA である。
- (4) 一方で、 X に潜在変数を取り、これによって観測 Y を説明し切るための潜在変数選択の文脈を帯びるのが FA である。

定義 4.1.7 (PCA, FA).

- (1) 新たな計画行列 $F \in M_{nm}(\mathbb{R})$ と変換行列 $A \in M_{pm}(\mathbb{R})$ とを用いて、 $X = FA^\top + E$ と分解することを **低階数近似** という。
- (2) 誤差行列 E の分散を最小にすることを、**主成分分析** という。 m 次元と p 次元の関係 A は **負荷量行列** という。
- (3) 誤差行列 E が

$$\Psi := \frac{1}{n} E^\top E = \text{diag}(\psi_1, \dots, \psi_p), \quad 1_n^\top E = 0_p^\top, 1_n^\top F = 0_m^\top, F^\top E = O$$

を満たすように調整することを、**因子分析** という。これは、低階数近似の残差が平均 0 で無相関になることを要請している。

- (4) 一般に、低階数近似の解 (F, A) は無数に存在する。が、 $F^\top F = nI_m$ の仮定を置けば、主な不定性は直交行列 $T \in O_m(\mathbb{R})$ に関する不定性になるため、付加条件をおいて求めた解 (F, A) に対して、より便利な AT を探す解析を **回転** という。varimax 回転はおそらく分散を最大化する。

歴史 4.1.8. 主成分分析は [Pearson, 1901] が力学の主軸定理に着想を得て発表した。が、経済学者 [Hotelling, 1935], [Hotelling, 1936] が再発見し、PCA の名前をつけた。

4.1.4 因子分析と正準相関分析

歴史 4.1.9 (因子分析の歴史).

- (1) 因子分析は心理学者 Spearman [Spearman, 1904] の知能の研究から始まり、その後「多変量相関分析」と呼べるような手法の総体が、心理学とその雑誌を中心に発展した。
- (2) 説明変数 X は「知能」なる構成概念が定める潜在変数で、これらの線型結合によって試験の点数などのデータ Y を説明する際に、 X の各相関構造が説明し切れ、残差 E が独立になるような線型結合の組を得たい。こうして、「流動知性」Big-5 など、種々の知能概念が提案される。

- (3) ここで突然潜在変数の概念が登場したが、潜在変数による観測量への回帰を取り扱うモデルは経済学では測定方程式モデルとも呼ばれる [豊田秀樹, 2009]。そして、潜在変数・観測量の区別なく回帰をしあえるモデルの全体を構造方程式モデルというのである。
- (4) しかし残る新たな因子 F の間の相関はどうするのであろうか。こうして相関分析、共分散構造分析の発想が要請される。したがって、構成概念が互いに独立である場合の探索的因子分析モデルの更なる特殊化である。
- (5) 正準相関分析は重回帰分析と因子分析を包含するデータ縮約方法とも見れる ([Levine, 1977], X を因子分析にかける変数群, Y を共通因子得点とする) が、特にマルチモーダルなデータに対して変数を要約する方法ともみれる。
- (6) 主成分分析との比較で言えば、これは1つの変数ベクトルの分散を見て座標系を選択するが、正準相関分析は2つの変数ベクトルの共分散を見る。

模型 4.1.10 (探索的因子分析模型 (EFA)). 上述の条件を満たす低階数近似の結果得る回帰模型

$$Y = \beta X + e, \quad e \sim (0, \sigma^2 I_n), E[Xe^\top] = O$$

が因子分析模型になる。

- (1) X の成分である構成概念を表す潜在変数を**因子**という。 β は**因子パタン行列**という [豊田秀樹, 2007]。
- (2) ここで、構成概念 X も潜在変数であるから、これも確率変数になる。 $\text{Cov}[X]$ の対角成分は1とすることが多いようだ [豊田秀樹, 2007]。なお、 $\text{Cov}[X]$ は**因子間相関行列**という。
- (3) 観測 Y の共分散構造は

$$\text{Cov}[Y] = \beta \text{Cov}[X] \beta^\top + \text{Cov}[e]$$

で与えられる。

- (4) $\text{Cov}[X] = I$ を満たすようにとったモデルを**直交模型**という [豊田秀樹, 2009]。

模型 4.1.11 (検証的因子分析模型 (CFA)). EFA があらゆる因子があらゆる可観測量に影響するとしているのとは違って、特定の因子が特定の測定変数に影響・付加することを指定している。よって、いくらか SEM に近い描像になる [Grimm and Yarnold, 2016]。

要諦 4.1.12 (高次積率の利用と独立成分分析への展望). データが非正規な分布に従う際には、高次の積率を扱うことが有効になる [Mooijaart, 1985]。どこでも、非正規性へのアプローチは高次理論とセットである。これと情報幾何を応用した独立成分分析 [甘利俊一 and 村田昇, 2002] はアイデアは同じである。狩野裕と清水昌平による [Shimizu and Y., 2003] もそうである。これは高次理論を構造方程式モデルに応用する流れ [甘利俊一 et al., 2002] の一つである。そして、[Shimizu et al., 2005] は LiNGAM を用いてパスを探索して因果を発見する。因果推論と両輪な時代の流れであるかもしれない。

4.1.5 発展：独立成分分析

主成分分析と因子分析は(日本の)信号処理の分野で独自の受容を受けた。「カクテルパーティー効果の工学的実現」を指す。

問題 4.1.13 (BSS: Blind Source Separation). 複数の未知の信号系列を未知の線形混合系で混合した複数の測定値系列から、それぞれの信号系列を分離する問題を考えたい。これを**ブラインド信号源分離**という。

注 4.1.14. 独立成分分析は主成分分析の拡張である。主成分分析は算譜も含めて成熟しているが、評価関数や最適化法が未確定で、場合によって結果が人によって変わってしまい得る点で独立成分分析は未熟であるといえる [村田昇, 2004]。

4.1.6 共通成分抽出の手法

転移学習、データ融合、情報推薦などの文脈で言われる「マルチビューデータ」「マルチモーダル（五感から得た情報を組み合わせる一つの意思決定を行うこと）」「マルチドメイン」「センサーフュージョン」なるデータに対する場合の PCA に値するような、古典的な解析手法の1つ。

問題 4.1.15 (変数要約の問題). 正準相関分析、正準因子分析を実行するためのデータは次の通り：

- (1) 説明変数・被説明変数を全て含めたデータ $X \in M_{np}(\mathbb{R})$ は $1_n^\top X = 0_p^\top$ を満たす列平均が零な行列とする．変数の分割 $X = [X_1, \dots, X_J]$ を考える，どれがどれを説明しているか，どのような回帰構造を持つかは問わない．
- (2) 各 X_j の列数を p_j とし， $C_j \in M_{p_j m}(\mathbb{R})$ との積 $X_j C_j \in M_{nm}(\mathbb{R})$ を考える．このとき， $m \leq \min(n, p)$ とする．すなわち， $C_j, X_j C_j$ はいずれも縦長である．
- (3) X_i の成分の線型結合 $X_i C_i$ と， X_j の成分の線型結合 $X_j C_j$ との組であって，互いに (何らかの意味で) 高い相関を持つような線型結合係数 C_i, C_j を定める問題を正準相関分析という．
- (4) このときの相関係数を**正準相関係数**という．相関係数は Pearson のもの，Spearman のものなど，いずれもありえる．
- (5) 分析の結果得られた， X_i, X_j の線型変換 $X_i C_i, X_j C_j$ を新たな変量と見るとき，**正準変量**という．

定義 4.1.16 (CCA: canonical correlation analysis, GCCA).

- (1) $J = 2$ のとき，

$$\arg \min \mu := \|X_1 C_1 - X_2 C_2\|^2, \quad \text{under } \frac{1}{n} C_j^\top X_j^\top X_j C_j = I_m \quad (j = 1, 2).$$

を満たす C_1, C_2 を求める手法を**正準相関分析**という．すなわち，2つの変数群 X_1, X_2 をどのように足し合わせれば，最も合致するかを最適化する．このとき， $\text{Im } C_1, \text{Im } C_2$ として高い相関を持つ線型部分空間の組を発見している．

- (2) $J \geq 3$ のときを**一般化正準相関分析**といい，最小化するところの目的関数に種々の提案がなされている．初等的には，新たな変数 $F \in M_{nm}(\mathbb{R})$ を追加した最小化問題

$$\eta(F, C) := \sum_{j \in [J]} \|F - X_j C_j\|^2, \quad \frac{1}{n} F^\top F = I_m.$$

を解くことを考える．また，深層学習を用いて拡張することも提案されている．

例 4.1.17 (GCCA の例)．一般に，マルチタスク学習の観点において各タスクが類似している時には，対応する各学習器が共通要因を獲得することで性能が上がるのが期待される．これは，GCCA において入力の変数数の数を増やすことにより，それぞれの相関がより大きくなるような，各変数対の射影が求まることに対応する．^{t1} □

要諦 4.1.18. 主成分分析は，多くの変数から「強い」ものを順に抽出し，「弱い」信号はノイズとして処理される．一方で正準相関分析は，多くの変数が「同じ対象を複数の観点から見たもの」と解して，「共通する信号」が弱かろうとこれを抽出し，「独立信号」をノイズと解する．^{t2}

定義 4.1.19 (CA: Correspondence Analysis). 与えられたデータ行列 $X = (X_1, \dots, X_n)$ が質的データ (categorical data) であるとする．

- (1) 所与のデータを次のように，行列の列 $(G_i)_{i \in [n]}$ に変換する．各 $G_i \in M_{np_i}(2)$ は，各個体が i 番目の p_i 個のカテゴリー分けについて，どれに当てはまるかを $\{0, 1\}$ でコードした行列で，**メンバーシップ行列**または**ダミー変数行列**という．
- (2) この G_1, \dots, G_n に対する，新たに付加を追加した一般化正準相関分析

$$\eta(F, C) := \sum_{j \in [J]} \|F - X_j C_j\|^2, \quad \frac{1}{n} F^\top F = I_m, 1_n^\top F = 0_m^\top.$$

を**多重対応分析**または歴史的には**数量化第 III 類**という．

- (3) これによって求まる全 $p := \sum_{i \in [n]} p_i$ カテゴリーに対する $C \in M_{pm}(\mathbb{R})$ 行列を**カテゴリ得点**， $F \in M_{nm}(\mathbb{R})$ を**得点行列**という．

要諦 4.1.20.

- (1) 「個体 i の得点と，その個体が属するカテゴリの得点とは，極めて似通うはずである」という仮定を**等質性仮定**という．この下で設定される目的関数も， $\eta(F, C)$ に等しくなるため，この観点から多重対応分析を**等質性分析**ともいう．^{t3}

^{t1} 岩瀬智亮，中山英樹『深層一般化正準相関分析』

^{t2} 赤穂昭太郎『正準相関分析入門』

^{t3} Gifi, A. (1990). *Nonlinear Multivariate Analysis*. Wiley.

(2) 多重対応分析の解には、回転の分の不定性があるが、これは

$$\frac{1}{Jn} C^\top D_G C = \frac{1}{Jn} \sum_{j=1}^J C_j^\top G_j^\top G_j C_j$$

が対角行列になる、という制約をおくと一意に定まる。

(3) 一般に対応分析は「分割表の視覚化」と説明される。というのも、地図のようにプロットして、互いに近い点は関係が深いことが推察できるためである。このときに、上の対角行列の制限を置けば、プロットの各軸にある種の指標と解することにも意味が出てくる。

4.1.7 個体類別の手法

クラスタ分析は類別の方法を見出す手法であり、判別分析は所与の類別に対してどこに最もフィットするかを見出す手法である。

問題 4.1.21 (類別の問題). クラスター分析、判別分析を行うために必要なデータは次の通りである：

- (1) まず、各データは有限集合 $[K]$ の元のいずれかに帰属すると発想する。興味のある変数は $[K]$ 上の確率変数となる。
- (2) データ $X \in M_{np}(\mathbb{R})$ から、各個体の所属カテゴリを表す $y: [n] \rightarrow [K]$ と、メンバーシップ行列 G とを (K などの決め方も含めて) 推定するのがクラスタ分析である。
- (3) Gram 行列 $D := G^\top G$ は対角行列で、その対角成分 n_k は、カテゴリ $k \in [K]$ に属する個体の数を表す。

命題 4.1.22. $J_n := I_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top$ を中心化行列とする。次の等式が成り立つ：

$$\text{Cov}[X] = \frac{1}{n} X^\top J X = \frac{1}{n} X^\top (I_p - G^\top D^{-1} G^\top) X + \frac{1}{n} X^\top (G D^{-1} G^\top - n^{-1} \mathbf{1}_n \mathbf{1}_n^\top) X =: S_W + S_B.$$

定義 4.1.23 (cluster analysis).

- (1) S_W を群内共分散行列、 S_B を群間共分散行列という。
- (2) 個体の分類を樹形図で出力する手法を階層的クラスター分析という。重心法や Ward 法がある。
- (3) 最小二乗法に基づく場合に k -平均法がある。

定義 4.1.24 (DA: discriminant analysis).

- (1) 判別分析という際には $K = 2$ の場合を主に指し、3つ以上のグループの判別は重判別分析や正準判別分析と呼ばれる。
- (2)

歴史 4.1.25.

- (1) 線型判別分析を LDA、二次曲線による判別分析を QDA と言う。
- (2) [Hastie and Tibshirani, 1996] が混合判別分析 (MDA) を発表した。

4.1.8 識別可能性の問題

与えられたデータに関して、2つの全く関係のない解が見つかってしまう恐れがある。また識別可能でも、殆どのデータに対して適合度のスコアが一致してしまうことがある。これを「同値模型」という。この取り扱いの問題になり得る。

4.2 判別分析と機械学習

実は判別分析問題には、ロジスティック回帰やサポートベクトルマシンなど、様々な解き方が存在する。 [wikipedia page](#) の次の記述を理解したい：

等分散性の仮定を外した物が二次判別分析である。それぞれのグループで異なる共分散行列を使用してマハラノビス距離を計算して、等距離になる場所を判別曲面とする方法である。この方法は二次関数となり、正規分布が成立している場合は正しい結果になる。

線形判別分析において、グループ間の確率のロジットは線形関数となるが、ここで線形関数という仮定を残したまま、正規分布や等分散性の仮定を外すとロジスティック回帰や単純パーセプトロンになる [14]。

さらに別な方法としては、線形判別関数を使用したい場合は、線形サポートベクターマシンで線形判別関数を求めるという方法もある。

4.3 線型回帰模型

パラメトリック模型のうち、 $y_n | \alpha, \beta \sim (\alpha + \beta x_n, \sigma^2)$ と表せるものを線型回帰模型という。すなわち、これは平均を表すパラメータの空間 $\Theta \subset \mathbb{R}^p$ を affine 部分空間に取った模型である。

- (1) 標準線型回帰模型 $y \sim (X\beta, \sigma^2 I)$, $\text{rank } X = p$ について、最小二乗推定量 $\hat{\beta} = (X^\top X)^{-1} X^\top y$ は、全ての線型な不偏推定量の中で分散が最小である (Gauss-Markov の定理)。
- (2) 線型正規回帰模型 $y \sim N(X\beta, \sigma^2 I)$, $\text{rank } X = p$ について、最小二乗推定量は最尤推定量になり、従って漸近有効であるが、それだけでなく全ての不偏推定量の中で分散が最小である (Rao-Blackwell の理論による)。
- (3) 誤差の独立同分布性と計画行列について2つ仮定を置くと、標準線型回帰模型の最小二乗推定量は漸近正規になることが示せる。

模型 4.3.1 (線型回帰模型総覧)。

- (1) Y_1, \dots, Y_n を均一な誤差を持った非同分布な過程で、それぞれの平均 $E[Y_i]$ の構造についても何の知識もないならば、模型を素朴に $E[Y_i] = \eta_i$ として、 $\text{Var}[Y_i] = \sigma^2$ と仮定する外ない。
- (2) しかし、平均 η_i を予測するのに、いくつかの説明変数 x_1, \dots, x_p が見つかるならば、これらを追加して η_i に線型な流動性を持たせることを考えられる。この結果得るモデルは、

$$Y = X\beta + \eta_1 + \epsilon$$

と表せる。位置パラメータ $\beta \in \mathbb{R}^p$ について線型であるときに、線型モデルといい、 X, η_1 は問わない。

- (3) ここで、 $\eta = \eta_0 + \eta_1$ は直交分解 $\text{Im } X \oplus \text{Ker } X$ とした。具体的には非斉次項 $\eta_1 = (I - X(X^\top X)^{-1} X^\top) \eta$ を模型のバイアスという。説明変数の数をデータよりも大きくして $p \geq n$ としない限り、ここは消えない。

- (4) 標準線型回帰模型は、ここから次のことを仮定する：

- (a) $\eta_1 = 0$.
- (b) $\epsilon \sim (0, \sigma^2 I)$ ($\sigma > 0$).
- (c) $\text{rank } X = p \leq n$.

この下で、 $\hat{\beta} := (X^\top X)^{-1} X^\top y$ を最小二乗推定量という (仮定 (c) より存在する)。 $\|\epsilon\|^2$ の値が停留するための必要条件 $X^\top X \beta = X^\top y$ を正規方程式という。この際の $\|\epsilon\|^2$ の最小値を残差平方和 (RSS) という。最小二乗推定量は、幾何学的には y の $\text{Im } X$ への射影で得られる p 次元ベクトルである。

- (5) 仮定 (b) を $\epsilon \sim N(0, \sigma^2 I)$ に取り替えたものを、線型正規回帰模型という。
- (6) 仮定 (a) (c) に加えて次を仮定したものを線型漸近正規回帰模型と呼ぼう：
 - (a) $\epsilon = (\epsilon^1, \dots, \epsilon^n)^\top$ は独立同分布に従う。

$$(b) \forall_{i \in [p]} \frac{\max_{j \in [p]} x_{ij}^2}{a_{ii}^{(n)}} \xrightarrow{n \rightarrow \infty} 0.$$

(c) 正則な極限 $\lim_{n \rightarrow \infty} R_n = R_\infty \in \text{GL}_p(\mathbb{R})$ が存在する.

ただし, $A_n = (a_{ij}^{(n)}) = X^\top X \in M_p(\mathbb{R})$ を計画行列が定める Gram 行列とすると最小二乗推定量は $\hat{\beta}_n = A_n^{-1} c_n$ ($c_n := X^\top y$) と表せ, $R_n := D_n^{-1} A_n D_n^{-1} \in M_p(\mathbb{R})$ ($D_n := \text{diag}(\sqrt{a_{11}^{(n)}}, \dots, \sqrt{a_{pp}^{(n)}})$) を説明変数間の標本相関係数行列とした.

4.3.1 標準線型回帰模型に対する Gauss-Markov の定理

例 4.3.2 (種々の線型推定量). b が β の線型な推定量 $b = Cy$ ならば,

$$E[b] = CE[u] = CX\beta, \quad \text{Var}[b] = \sigma CC^\top.$$

よって, $CX = I$ を満たす $C \in M_{pn}(\mathbb{R})$ に対して, $b = Cy$ は不偏になる. 特に $\text{rank } Z = p, Z \in M_{np}(\mathbb{R})$ を用いて $C := (Z^\top X)^{-1} Z^\top$ とすると, これは不偏推定量である. これを**操作変数推定量**という. 最小二乗推定量は, 説明変数を操作変数とする操作変数推定量とみれる. \square

例 4.3.3. $\|\epsilon\|_1$ を最小にする推定量 b は, y の線型像にはならず, さらに不偏とも限らないが, ϵ の分布の対称性を仮定すれば不偏になる. \square

4.3.2 Rao-Blackwell の定理

完備十分統計量 t が見つければ, この関数であって不偏な関数 $\hat{\theta}(t)$ が見つければ, これが最小分散不偏推定量である.

定義 4.3.4 (sufficient statistics). 模型 $(\mathcal{X}, \mathcal{A}, (P_\theta)_{\theta \in \Theta})$ のパラメータ $\theta \in \Theta \subset \mathbb{R}^p$ と統計量 $t: \mathcal{X} \rightarrow \mathbb{R}^p$ について,

(1) 観測 $y \in \mathcal{X}$ の条件付き密度関数が

$$f(y|\theta) = g(t|\theta)h(y)$$

と「 θ は t を通じてのみ密度関数に現れる」という形に因数分解できるとき, t を**十分統計量**という.

(2) 十分統計量 t が**完備**でもあるとは, 任意の f の可測関数 $f(t)$ について, 次が成り立つことをいう:

$$[\forall_{\theta \in \Theta} E_\theta[f(t)] = 0] \Leftrightarrow [\forall_{\theta \in \Theta} f \equiv 0 \text{ } P_\theta\text{-a.e.}]$$

補題 4.3.5. 完備十分統計量に関する不偏推定量は一意的である. 特に, Rao-Blackwell の定理と併せれば, これが不偏推定量の中で分散を最小にする.

[証明]. $\hat{\theta}_1, \hat{\theta}_2$ をいずれも不偏推定量とすると,

$$\forall_{\theta \in \Theta} E_\theta[\hat{\theta}_1(t) - \hat{\theta}_2(t)] = 0.$$

t の完備性より, これは $\hat{\theta}_1 \equiv \hat{\theta}_2$ a.s. を意味する. \blacksquare

定理 4.3.6 (Rao-Blackwell). $\hat{\theta}$ を θ の不偏推定量, T を十分統計量とする. $\hat{\theta}^* := E[\hat{\theta}|\sigma[T]]$ と定めると,

- (1) $\hat{\theta}^*$ も θ の不偏推定量である.
- (2) $\text{Var}[\hat{\theta}^*] \leq \text{Var}[\hat{\theta}]$.

例 4.3.7 (指数型分布族の完備十分統計量). $\Theta \subset \mathbb{R}^p$ 上の指数分布族とは,

$$f(y|\theta) = \exp(a(\theta)^\top b(y) + c(\theta) + d(y)) \quad a, b \in \mathbb{R}^p, c, d \in \mathbb{R}.$$

と表せるもののことをいい, この分布族については $b: \mathcal{Y} \rightarrow \mathbb{R}^p$ が完備十分統計量になる. \square

4.3.3 線型正規回帰模型

模型 4.3.8. 模型 $y \sim N(X\beta, \sigma^2 I)$, $\text{rank } X = p$ の密度関数は

$$f(y|\beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2} \exp\left(-\frac{1}{2\sigma^2}(y - X\beta)^\top(y - X\beta)\right)}.$$

であるから, β の最尤推定量は二次形式 $(y - X\beta)^\top(y - X\beta)$ を最小化するものとして得られる. これを最小二乗推定量である.

定理 4.3.9.

- (1) 最小二乗推定量は最尤推定量である. 従って特に, 一致性と漸近有効性を持つ.
- (2) $X^\top y, y^\top y$ が β, σ^2 の完備十分統計量になる.
- (3) 最小二乗推定量は最小分散不偏推定量である.
- (4) $\hat{\beta} \sim N(\beta, \sigma^2(X^\top X)^{-1})$.

[証明].

- (1) 上の議論から明らか.
- (2) 正規分布 $N(X\beta, \sigma^2 I)$ を指数型分布族と見ればわかる.
- (3) 最小二乗推定量は完備十分統計量 $X^\top y$ の関数となっているから, Rao-Blackwell の定理から. また, Cramer-Rao の不等式を達成することからも示せる.
- (4) $\hat{\beta} = (X^\top X)^{-1}X^\top y = (X^\top X)^{-1}X^\top(X\beta + \epsilon) = \beta + (X^\top X)^{-1}X^\top \epsilon$ と変形でき, これと $\epsilon \sim N(0, \sigma^2 I)$ の仮定による.

■

4.3.4 線型漸近正規回帰模型

標準線型回帰模型の弱みは, 分布の仮定が弱いために, $\hat{\beta}$ の分布に関して標本分布は勿論, 漸近的にも何も言えないことである. そこで, この点を改良する.

定理 4.3.10. 線型漸近正規回帰模型について, $D_n(\hat{\beta}_n - \beta)$ は漸近的に $N(0, \sigma^2 R_\infty^{-1})$ に従う. 特に, 最小二乗推定量 $\hat{\beta}_n$ は漸近正規である.

[証明]. Lindeberg-Feller の中心極限定理による.

■

系 4.3.11. 線型漸近正規回帰模型の, 誤差の同分布性の仮定 (d) は緩めて, 代わりに次の一様可積分性を採用しても, $D_n(\hat{\beta}_n - \beta)$ は漸近的に $N(0, \sigma^2 R_\infty^{-1})$ に従う.

- (1) $\epsilon_j \sim (0, \sigma^2)$ は互いに独立に分布する.
- (2) 確率変数族 (ϵ^2) は一様可積分である. すなわち, ϵ_j の分布を F_j と置くと,

$$\sup_{j \in \mathbb{N}^+} \int_{|\epsilon| > c} \epsilon^2 dF_j(s) \xrightarrow{c \rightarrow \infty} 0.$$

4.3.5 半径数的線型回帰モデル

模型 4.3.12 (semiparametric regression model). $(X_1, Y_1), \dots, (X_n, Y_n)$ を観測として, 一般化線型回帰モデル

$$Y_j = \alpha + \beta X_j + \epsilon_j \quad (j \in [n], \epsilon_j \in L_0^2(\Omega))$$

を考える. ϵ_j には, 追加で次のような仮定を置くことが多い.

- (A1) ϵ_j は独立同分布に従う.
- (A2) 均一な分散を持ち, 互いに無相関である: $\exists \sigma^2 \in \mathbb{R}^+ \quad \forall i, j \in [n] \quad E[\epsilon_i \epsilon_j] = \sigma^2 \delta_{ij}$.

(A3) 不均一な分散を持つが、互いに無相関である： $\exists_{\{\sigma_i^2\} \subset \mathbb{R}_+} \forall_{i,j \in [n]} E[\epsilon_i \epsilon_j] = \sigma_i^2 \delta_{ij}$. さらに、分散は有界である： $\lim_{n \rightarrow \infty} \max_{i \in [n]} \sigma_i^2 < \infty$.

(A4) 不均一な分散を持つが、互いに無相関であり、分散は次を満たす：

$$\frac{\gamma_n}{n} \rightarrow 0, \frac{\hat{m}_2 \gamma_n}{\sum_{j \in [n]} (X_j - \bar{X})^2} \rightarrow 0 \quad \gamma_n := \max_{j \in [n]} \sigma_j^2.$$

推定量には次が考えられる.

(1) $(\hat{\alpha}_n, \hat{\beta}_n) := \arg \min_{(\alpha, \beta) \in \mathbb{R}^2} \sum_{i \in [n]} (Y_i - \alpha - \beta X_i)^2$ を通常の**最小自乗推定量** (OLS estimator) という.

(2) $\hat{\sigma}^2 := \frac{\sum_{j \in [n]} (Y_j - \hat{\alpha} - \hat{\beta} X_j)^2}{n - 2}$.

命題 4.3.13. 上の模型について、仮定 (A1),(A2),(A3) のいずれかをおけば、次が成り立つ：

(1) 変数変換 $\delta := \alpha + \beta \bar{X}$ について、等価な模型 $Y_j = \delta + \beta(X_j - \bar{X}) + \epsilon_j$ を考える. このとき、最小自乗推定量は

$$\hat{\delta}_n = \bar{Y}, \quad \hat{\beta} = \beta + \frac{\sum_{j \in [n]} \epsilon_j (X_j - \bar{X})}{\sum_{j \in [n]} (X_j - \bar{X})^2}.$$

と表せ、 $\hat{\delta}$ は Y の不偏な一致推定量であり、 $\sum_{j \in [n]} (X_j - \bar{X})^2 \rightarrow \infty$ が成り立つならば、 $\hat{\beta}$ も β の不偏な一致推定量である.

(2) 次が成り立つならば、 $\hat{\alpha}, \hat{\beta}$ は α, β の不偏な一致推定量である.

(a) $\sum_{j \in [n]} (X_j - \bar{X})^2 \rightarrow \infty$.

(b) $\frac{\bar{X}^2}{\sum_{j \in [n]} (X_j - \bar{X})^2} \rightarrow 0$.

4.3.6 自己回帰モデル

模型 4.3.14 (線型自己回帰模型). $k \in \mathbb{N}^+$ を次数とし、 Y_1, \dots, Y_n を観測とする.

$$Y_j = \alpha + \beta Y_{j-1} + \epsilon_j \quad (j \in [n], \epsilon_j \in L_0^2(\Omega), E[\epsilon_i \epsilon_j] = \sigma^2 \delta_{ij} > 0).$$

さらに、次の3つを仮定することが多い.

(A1) 安定性条件： $|\beta| < 1$.

(A2) $\epsilon_j \in L^4(\Omega)$, $Y_0 \in L^2(\Omega)$.

(A3) $\forall_{j \in [n]} \text{Cov}[Y_0, \epsilon_j] = 0$.

また、分散の等質性を、 $(\text{Var}[\epsilon_j])_{j \in \mathbb{N}}$ は有界列を定めることに緩めても良い.

命題 4.3.15. 上の模型について、

(1) 最小自乗推定量は、

$$\hat{\alpha} = \bar{Y}_{1,n} - \hat{\beta} \bar{Y}_{0,n-1}, \quad \hat{\beta} = \beta + \frac{\sum_{r \in [n]} \epsilon_r (Y_{r-1} - \bar{Y}_{0,n-1})}{\sum_{r \in [n]} (Y_{r-1} - \bar{Y}_{0,n-1})^2}. \quad \bar{Y}_{i,j} := \frac{1}{j-i} \sum_{r=i}^j Y_r.$$

(2) 最小自乗推定量は一致性を持つ.

4.4 非線型回帰模型

4.4.1 Probit 回帰模型

模型 4.4.1. $\mathcal{Y} = \{\pm 1\}$ とし、この上の模型

$$P[y = 1 | \theta, x] = \Phi(\alpha + \beta x), \quad \Phi : \mathbb{N}(0, 1) \text{ の累積分布関数}$$

を **probit 回帰模型** という.

4.5 回帰模型の検定

模型は回帰係数 β の推定を目的として作られるが、その残差 e を用いて誤差 ϵ の分散 σ^2 が推定できる。これは模型の診断にも使える。

4.5.1 誤差分散の推定と検定

4.5.2 推定量の分布

4.6 回帰模型の診断

4.6.1 決定係数

モデル $(\beta_0, \dots, \beta_p) \in \mathbb{R}^{p+1}$ の損失関数を、観測 y_1, \dots, y_n に対して、

$$R^2 := 1 - \frac{\sum_{i \in [n]} (y_i - \hat{y}_i)^2}{\sum_{i \in [n]} (y_i - \bar{y})^2}$$

と定める。これを**決定係数**という。最小二乗法は決定係数が最大になるようなパラメータの選択である。

4.6.2 AIC

説明変数を用意しすぎると、分散が大きくなる。そこで AIC の原理が入る。シンプルかつ本質を捉えた基準は、理論と実践の交差点でのみ生まれる。

要諦 4.6.1. 赤池情報量規準 (AIC) やベイズ情報量規準 (BIC) などの情報量規準 (英語版) は、交差検証法よりも計算が高速であり、小さいサンプルでもパフォーマンスが変動しにくい、交差検証よりも好ましい場合がある [6]。情報量規準は、モデルのサンプル内精度を最大化することによって推定器の正則化パラメータを選択すると同時に、その有効なパラメーターの数/自由度にペナルティを課す。^{†4}

4.6.3 一般化最小二乗法

f を非線形化するといっても、明示的に用いられるのは二次項のみで、殆ど場合は変数変換と線型汎函数の合成とみなす。また、相関がある場合に基底を取り直すことで、線型変換と線型汎函数の合成とみなすことも多い。

例 4.6.2 (log, logit モデル). 非線形モデル

$$y' = \beta_0 + \beta_1 x'_1 + \dots + \beta_p x'_p + \epsilon, \quad y' = \phi_0(y), x'_1 = \phi_1(x_1), \dots, x'_p = \phi_p(x_p)$$

について、

- (1) 全ての変数が正であるとき、 $\phi_i = \log$ と取ることが多い。
- (2) $y \in n+1$ であるとき、ロジット変換 $\phi_0(y) = \log \frac{y}{n-y}$ を用いることが多い (Cox 58)。
- (3) ϕ_0 を**連結関数**ともいう。

□

^{†4} <https://ja.wikipedia.org/wiki/ラッソ回帰>

要諦 4.6.3. モデルは同じく 1958 年に発表された単純パーセプトロンと等価であるが、scikit-learn などでは、パラメータを決める最適化問題で確率的勾配降下法を使用する物をパーセプトロンと呼び、座標降下法や準ニュートン法などを使用する物をロジスティック回帰と呼んでいる。

例 4.6.4 (GLS). 一般化線型モデルにおいて、 $\epsilon: \Omega \rightarrow \mathbb{R}^n$ の共分散行列 Σ を対角化し、標準的な線型回帰モデルに戻して最小二乗推定量を計算してから、もう一度基底を戻すことが考えられる。これを**一般化最小二乗推定量** $\hat{\beta}^*$ という。□

例 4.6.5 (RIDGE 正則化 (1970)). $x_1, \dots, x_p \in \mathbb{R}^n$ が強い相関を持つ場合を考える。これは $X\beta$ が多重線形性を持ってしまっていることを意味する。このとき、通常の最小二乗推定量

$$\hat{\beta} = (X^\top X)^{-1} X^\top y$$

よりも、次の推定量が一様に改善する：

$$\hat{\beta}_R := (X^\top X + kI_p)^{-1} X^\top y.$$

この改善のメカニズムは、大きな回帰係数を縮小して、過適合を抑制していること (**正則化**) にある。実際、 l^2 -ノルムに制約を与えた問題

$$\min_{\beta_0, \beta} \left(\sum_{i \in [n]} (y_i - \beta_0 - x_i^\top \beta)^2 \right) \quad \text{subject to} \quad \sum_{j \in [p]} \beta_j^2 \leq t$$

に等価な次の問題

$$\min_{\beta \in \mathbb{R}^p} \left(\frac{1}{n} \|y - X\beta\|_2^2 + \lambda \|\beta\|_2 \right).$$

が定める推定量に同値である。 λ を乗数として、次とも同値らしい。

$$(\hat{\beta}_R)_j = (1 + n\lambda)^{-1} \hat{\beta}_j$$

$1 + n\lambda$ を係数として一様に収縮する。□

例 4.6.6 (LASSO 変数選択: least absolute shrinkage and selection operator). 「大きな回帰係数を縮小して、過適合を抑制する」に加えて、共変量を取捨選択するメカニズム (**best subset regression**) も用いる。すなわち、回帰係数の l^1 -ノルムに制約を与える。線型回帰モデルにおいて、OLS の代わりに、 $(\beta_1, \dots, \beta_p) \in \mathbb{R}^p$ をある半空間に制限して、次の問題を解く：

$$\min_{\beta_0, \beta} \left(\sum_{i \in [n]} (y_i - \beta_0 - x_i^\top \beta)^2 \right) \quad \text{subject to} \quad \sum_{j \in [p]} |\beta_j| \leq t$$

Lagrange の未定乗数法より、次に同値：

$$\min_{\beta \in \mathbb{R}^p} \left(\frac{1}{n} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1 \right).$$

□

要諦 4.6.7. λ は t の変数変換であるから、正則化パラメータとも呼ぶ。この選択には、交差検証法や情報量規準が用いられる。また、係数の事前分布として正規分布を仮定した場合の MAP 推定値がリッジ回帰に相当するのと同様に、係数の事前分布としてラプラス分布を仮定した場合の MAP 推定値がラッソ回帰に相当する。

4.6.4 Bayes 回帰

RIDGE 回帰も LASSO 回帰も、 $\beta \in \mathbb{R}^{p+1}$ に関する異なる事前分布に関する最大事後確率推定量 (MAP) に他ならない。

第 5 章

参考文献

参考文献

- [Basu, 1955] Basu, D. (1955). On statistics independent of a complete sufficient statistic. Sankhyā: The Indian Journal of Statistics, 15(4):377–380.
- [Bhattacharya et al., 2016] Bhattacharya, R., Lin, L., and Patrangenaru, V. (2016). A Course in Mathematical Statistics and Large Sample Theory. Springer Texts in Statistics. Springer New York.
- [Fisher, 1925] Fisher, R. A. (1925). Statistical Methods for Research Workers. Oliver and Boyd (Edinburgh).
- [Grimm and Yarnold, 2016] Grimm, L. G. and Yarnold, P. R. (2016). 研究論文を読み解くための多変量解析入門 応用篇. 北大路書房.
- [Hastie and Tibshirani, 1996] Hastie, T. and Tibshirani, R. (1996). Discriminant analysis by gaussian mixtures. Journal of the Royal Statistical Society, 1(Series B 58):155–176.
- [Hotelling, 1935] Hotelling, H. (1935). The most predictable criterion. Journal of Educational Psychology, 26(2):139–142.
- [Hotelling, 1936] Hotelling, H. (1936). Relations between two sets of variates. Biometrika, 27:321–377.
- [Kawata and Sakamoto, 1949] Kawata, T. and Sakamoto, H. (1949). On the characterisation of the normal population by the independence of the sample mean and the sample variance. Journal of Mathematical Society of Japan, 1(2):111–115.
- [Levine, 1977] Levine, M. S. (1977). Canonical Analysis and Factor Comparison. Quantitative Applications in the Social Sciences. SAGE Publications, Inc.
- [Mooijart, 1985] Mooijart, A. (1985). Factor analysis for non-normal variables. Psychometrika, 50(3):323–342.
- [Pearson, 1901] Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. Philosophical Magazine, 2(11):559–572. 主成分分析の初出論文.
- [Shimizu et al., 2005] Shimizu, S., Hyvarinen, A., Hoyer, P., and Kano, Y. (2005). Finding a causal ordering via independent component analysis. Computational Statistics and Data Analysis, 50(11):665–672.
- [Shimizu and Y., 2003] Shimizu, S. and Y., K. (2003). Examination of independence in independent component analysis. In H. Yanai, A. Okada, K. Shigemasu, Y. Kano, J. J. Meulman, editor, New Developments in Psychometrics.
- [Spearman, 1904] Spearman, C. (1904). 'general intelligence,' objectively determined and measured. the American Journal of Psychology, 15(2):201–293.
- [Yates, 1934] Yates, F. (1934). Contingency tables involving small numbers and the χ^2 test. Supplement to the Journal of the Royal Statistical Society, 2(1):217–235.
- [甘利俊一 et al., 2002] 甘利俊一, 狩野裕, 佐藤俊哉, 松山裕, 竹内啓, and 石黒真木夫 (2002). 多変量解析の展開: 隠れた構造と因果を推理する, volume 5 of 統計科学のフロンティア. 岩波書店.
- [甘利俊一 and 村田昇, 2002] 甘利俊一 and 村田昇 (2002). 独立成分分析: 多変量データ解析の新しい方法, volume 18 of 臨時別冊・数理科学 SGC ライブラリ. サイエンス社.
- [村田昇, 2004] 村田昇 (2004). 入門 独立成分分析. 東京電機大学出版局.
- [竹内啓, 2016] 竹内啓 (2016). 数理統計学の考え方: 推測理論の基礎. 岩波書店.
- [豊田秀樹, 2007] 豊田秀樹 (2007). 共分散構造分析 [理論編]. 統計ライブラリー. 朝倉書店.
- [豊田秀樹, 2009] 豊田秀樹 (2009). 共分散構造分析 [実践編]. 統計ライブラリー. 朝倉書店.

参考文献

- [1] 吉田吉田朋広 (2006) 『数理統計学』(朝倉書店)
- [2] 鍋谷鍋谷清治 (1978) 『数理統計学』(共立出版)
- [3] 竹村竹村彰道 (2020) 『現代数理統計学』(学術図書)
- [4] 久保川久保川達也 (2017) 『現代数理統計学の基礎』(共立出版)
- [5] 西山陽一西山陽一 (2011) 『マルチンゲール理論による統計解析』(近代科学社)
- [6] 稲垣宣生稲垣宣生 (1990) 『数理統計学』(裳華房)
- [7] 高松高松俊朗 (1977) 『数理統計学入門』(学術図書)
- [8] 竹内 16 竹内啓 (2016) 『数理統計学の考え方』(岩波書店).
- [9] Hogg Hogg, R. V., McKean, J. W., and Craig, A. T. (2005) *Introduction to Mathematical Statistics*. 6th Ed. Prentice Hall.
- [10] 蓑谷蓑谷千風彦 (2012) 『正規分布ハンドブック』(朝倉書店).
- [11] 佐和佐和隆光 (1979). 『回帰分析』(朝倉書店, 統計ライブラリー).
- [12] 統計教育大学間ネットワーク JINSE 編集. (2017). 『現代統計学』(日本評論社).
- [13] 松原松原望 (2013) 『統計学』(東京図書)
- [14] Bhattacharya Rabi Bhattacharya, Lizhen Lin, Victor Patrangenaru. (2016). *Course in Mathematical Statistics and Large Sample Theory*. Springer.
- [15] Ibragimov HasMinskii Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical Estimation*. Springer.
- [16] van der Vaart van der Vaart. (1998). *Asymptotic Statistics*. Cambridge University Press.
- [17] Hampel Hampel - Robust Statistics - 2005
- [18] Huber Huber and Ronchetti - Robust Statistics - 2009
- [19] Lehmann E. L. Lehmann - Nonparametrics: Statistical Methods Based on Ranks
- [20] 赤池 76 赤池広次 (1976) 「情報量規準 AIC とは何か」数理科学 153 号, 5-11 ページ.
- [21] 赤池 79 赤池広次 (1979) 「統計的検定の新しい考え方」数理科学 198 号, 51-57 ページ.
- [22] Sanov61 Sanov, I. N. (1961). IMS and AMS Selected Translations in Mathematical Statistics and Probability, Vol. 1, 213-244.
- [23] Yates Yates, F. (1934). Contingency Tables Involving Small Numbers and the χ^2 Test. *Supplement to the Journal of the Royal Statistical Society*, Wiley. 1(2): 217-235.