

# 目次

|       |                            |    |
|-------|----------------------------|----|
| 第 1 章 | 古典大標本理論                    | 4  |
| 1.1   | 最尤推定法                      | 4  |
| 1.1.1 | モデルの定義                     | 4  |
| 1.1.2 | 一致性・漸近正規性を持つ十分条件           | 5  |
| 1.1.3 | 収束レート                      | 6  |
| 1.1.4 | Neyman-Scott の問題           | 6  |
| 1.2   | 漸近決定理論の枠組み                 | 7  |
| 1.2.1 | 一致推定量の定義                   | 7  |
| 1.2.2 | 一致推定量の存在                   | 8  |
| 1.3   | $M$ -推定量                   | 8  |
| 1.3.1 | 一様大数の法則                    | 9  |
| 1.3.2 | 一致推定量の収束の速度                | 10 |
| 1.3.3 | $M$ -推定量の影響関数              | 10 |
| 1.3.4 | モーメント法                     | 10 |
| 1.3.5 | 一般化モーメント法                  | 11 |
| 1.4   | 最小コントラスト推定量                | 11 |
| 1.4.1 | 定義と例                       | 11 |
| 1.4.2 | 独立観測を仮定しない強一致性定理           | 12 |
| 1.4.3 | 一致性を持つ推定量の例                | 12 |
| 1.4.4 | 独立観測の構造を仮定した場合             | 13 |
| 1.4.5 | コンパクト集合に終局するための条件          | 13 |
| 1.5   | $Z$ -推定量                   | 13 |
| 1.5.1 | 定義と例                       | 13 |
| 1.5.2 | 漸近 $Z$ -推定量の一致性            | 14 |
| 1.5.3 | コンパクト集合に終局するための条件          | 14 |
| 1.5.4 | 一致性を持つ $Z$ -推定量の漸近分布       | 14 |
| 1.5.5 | LeCam のワンステップ改善による最適推定量の構成 | 15 |
| 1.5.6 | $Z$ -推定量の影響関数              | 16 |
| 1.6   | 尤度比検定とその同値な検定              | 16 |
| 1.6.1 | 定義と漸近分布                    | 16 |
| 1.6.2 | 複合仮説の検定                    | 17 |
| 1.6.3 | Rao 検定と Wald 検定            | 17 |
| 1.6.4 | 多項分布の検定                    | 18 |
| 第 2 章 | 非径数的手法                     | 19 |
| 2.1   | 情報量規準                      | 19 |
| 2.1.1 | KL-変分によるモデル評価              | 19 |
| 2.1.2 | 交差エントロピーの推定                | 20 |
| 2.1.3 | AIC の含蓄                    | 21 |

|       |                          |    |
|-------|--------------------------|----|
| 2.2   | 密度推定                     | 24 |
| 2.2.1 | 核型推定量                    | 24 |
| 2.2.2 | その他の手法                   | 25 |
| 2.3   | 射影と条件付き期待値               | 25 |
| 2.3.1 | Hajek の射影                | 25 |
| 2.3.2 | Hoeffding の分解            | 26 |
| 2.4   | $U$ -統計量                 | 26 |
| 2.4.1 | 定義と例                     | 26 |
| 2.4.2 | 漸近正規性                    | 28 |
| 2.4.3 | $U$ -統計量の変種              | 28 |
| 2.4.4 | 2 標本 $U$ -統計量            | 29 |
| 2.4.5 | 退化 $U$ -統計量              | 29 |
| 第 3 章 | 半径数的手法の理論                | 30 |
| 3.1   | 尤度比確率場の局所漸近構造            | 30 |
| 3.1.1 | 接触性                      | 30 |
| 3.1.2 | 尤度比の挙動                   | 31 |
| 3.1.3 | 局所漸近正規なモデル               | 31 |
| 3.1.4 | 検定への応用                   | 32 |
| 3.1.5 | 推定への応用                   | 32 |
| 3.1.6 | Ibragimov-Has'minskii 理論 | 33 |
| 3.2   | 接触構造                     | 33 |
| 3.2.1 | 尤度比と測度変換                 | 33 |
| 3.2.2 | 接触性                      | 34 |
| 3.3   | 局所漸近正規性                  | 35 |
| 3.3.1 | 尤度の拡張                    | 35 |
| 3.3.2 | 正規実験への収束                 | 36 |
| 3.4   | 漸近決定理論                   | 36 |
| 3.4.1 | 実験列の収束の定義                | 36 |
| 3.4.2 | 漸近表現定理                   | 37 |
| 3.4.3 | 漸近正規性                    | 37 |
| 3.4.4 | 一様分布                     | 37 |
| 3.4.5 | Pareto 分布                | 37 |
| 3.4.6 | 漸近混合正規性                  | 38 |
| 3.5   | 尤度比確率場の局所漸近構造            | 38 |
| 3.5.1 | 漸近決定理論                   | 38 |
| 3.5.2 | 統計的実験                    | 38 |
| 3.5.3 | 局所漸近正規性                  | 39 |
| 3.5.4 | 局所漸近正規な例                 | 39 |
| 3.5.5 | 漸近有効性                    | 40 |
| 3.6   | 漸近有効性                    | 40 |
| 3.6.1 | 実験の下界                    | 40 |
| 3.6.2 | Gauss モデルの平均の推定          | 40 |
| 3.6.3 | 畳み込み定理                   | 40 |
| 3.6.4 | 局所漸近 minimax 定理          | 41 |
| 3.6.5 | 下界を達成する推定量               | 41 |
| 第 4 章 | 漸近展開                     | 42 |

|       |                       |    |
|-------|-----------------------|----|
| 4.1   | 漸近展開 . . . . .        | 42 |
| 4.2   | 平滑化補題 . . . . .       | 42 |
| 4.3   | 特性関数の展開 . . . . .     | 43 |
| 4.4   | 漸近展開の正当性の証明 . . . . . | 43 |
| 4.5   | 漸近展開の変換 . . . . .     | 43 |
| 4.6   | 最尤推定量の漸近展開 . . . . .  | 43 |
| 4.7   | 歴史 . . . . .          | 43 |
| 第 5 章 | 情報幾何学                 | 44 |
| 5.1   | 漸近展開と情報幾何 . . . . .   | 44 |
| 5.2   | 変分 . . . . .          | 44 |
|       | 5.2.1 定義 . . . . .    | 44 |
|       | 5.2.2 例 . . . . .     | 44 |
| 第 6 章 | 再標本化法                 | 46 |
| 6.1   | ブートストラップ法 . . . . .   | 46 |
| 第 7 章 | 参考文献                  | 48 |
| 参考文献  |                       | 49 |
| 参考文献  |                       | 50 |

## 第 1 章

# 古典大標本理論

非線型モデル・非ガウスモデルに対しては、統計量の精密な分布を導出することが困難なため（それが非線型性である）、統計的決定理論を適用するのは非現実的であるが、これに替わって漸近的方法に基づく大標本理論が有力な解析手段となる。ここで、正規近似に基づく 1 次の漸近理論を扱う。続いて、大標本理論の一番最初の応用先として統計量の漸近分布を求めることがあるので、統計的検定問題を例に、漸近決定理論を考える。しかし、実際の統計推測では標本数を意識した理論が要請される。そこで統計研究は独自の発達をしてしまった。その原点回帰が高次の漸近論である。

- (1)  $M$ -推定量：一般化された対数尤度関数  $\bar{U}_n(\mathbf{x}; \theta) : \bar{\mathcal{X}} \times \Theta \rightarrow (-\infty, \infty]$  を、ある関数  $U : \mathcal{X} \times \Theta \rightarrow (-\infty, \infty]$  の経験平均とする。こうして得られた目的関数  $\bar{U}_n(\mathbf{x}; \theta)$  は、独立観測の仮定の下で、独立確率変数列の和の構造を持つことにより、 $\bar{U}_n(\mathbf{x}_n; \theta)$  は  $\mathbf{x} \in \bar{\mathcal{X}}$  に依らず  $\bar{U}(\theta) := E_{P_\theta}[U(X; \theta)]$  にコンパクト一様収束するための簡明な十分条件が得られる 1.3.4. なお、 $M$ -推定量は  $Z$ -推定量である。
- (2) 最小コントラスト推定量：一般化された対数尤度関数  $\bar{\Psi}_n(\mathbf{x}; \theta) : \bar{\mathcal{X}} \times \Theta \rightarrow (-\infty, \infty]$  を考え、これの真値からの差  $\bar{U}(\mathbf{x}, \theta) := \bar{\Psi}(\mathbf{x}; \theta) - \bar{\Psi}(\mathbf{x}; \theta_0)$  の最大化という付随する問題を考える場面は多い。この枠組の中で、一様対数の法則が一方に成り立つ下では、いくつかの追加の条件の下で強一貫性が従う 1.4.4.  $M$ -推定量は最小コントラスト推定量であるが、この場合はさらに明瞭である 1.4.8.
- (3)  $Z$ -推定量： $M$ -推定量は  $Z$ -推定量である。一般化されたスコア関数  $\psi$  を推定関数といい、この経験平均  $\bar{\psi}_n$  の零点として得られる推定量をいう。このクラスは漸近正規性を持つ。

### 記法 1.0.1.

- (1)  $\mathcal{X}$  上のディラック測度の全体は、 $P(\mathcal{X})$  の極点の全体  $\text{Ex}(P(\mathcal{X}))$  に等しい。この凸結合の全体は、経験分布の全体に等しく、正しくは  $\text{Conv}(\text{Ex}(P(\mathcal{X})))$  と表すべきだが、 $\text{Em}(P(\mathcal{X}))$  と略記することとする。

## 1.1 最尤推定法

統計推測においては、Bayes 法と最尤法とが双璧をなす。最尤推定量がごく自然な条件の下で一貫性を持つのは、i.i.d. の場合は Wald、これを一般化して Huber が示した。また、正則な場合は有効な推定量にもなるが、これは「尤度最大」であることは関係がなく、「尤度方程式の根である」ことによる。そこで、 $Z$ -推定量なる対象に議論が移る。

### 1.1.1 モデルの定義

尤度とはその標本  $X_1, \dots, X_n$  の  $P_\theta$  の下でのもっともらしさである。 $X_1, \dots, X_n$  が独立であるとき、 $L_n(\theta; \mathbf{x}_n)$  は積としての簡単な表示を持つ。尤度が最大な  $P_\theta$  をファイナルアンサーとするのは自然な発想である。

**模型 1.1.1** (最尤推定法が適用可能なモデル (Fisher)). モデル  $\{P_\theta\}_{\theta \in \Theta} \subset P(\mathcal{X})$  はある  $\sigma$ -有限測度  $\nu$  に支配されており、密度  $p_\theta(x)$  を持つとする。

- (1)  $L_n(\theta; x_1, \dots, x_n) := \prod_{j \in [n]} p_\theta(x_j)$  によって定まる母数空間上の関数  $L_n : \Theta \rightarrow \mathbb{R}_+$  を、観測  $x_1, \dots, x_n$  が定める**尤度関数**と

いう.

(2) 合成  $l_n(\theta; x_1, \dots, x_n) := \log L_n(\theta; x_1, \dots, x_n)$  を対数尤度関数という.

(3)  $(p_\theta)$  が微分可能ならば  $l_n$  も微分可能で,  $S(\theta; x_1, \dots, x_n) := \frac{\partial l}{\partial \theta}(\theta; x_1, \dots, x_n)$  が定まる. これをスコア関数という.

(4)  $\hat{\theta}_n(x_1, \dots, x_n) := \arg \max_{\theta \in \Theta} L_n(\theta; x_1, \dots, x_n)$  によって定まる統計量  $\hat{\theta}_n: \mathcal{X}^n \rightarrow \Theta$  を最尤推定量という.

**補題 1.1.2 (最尤推定量の特徴付け).** 観測  $\mathbf{x}_n = (x_1, \dots, x_n)$  について, 次は同値:

- (1)  $\hat{\theta}_n$  は尤度関数  $L_n(\theta; \mathbf{x}_n)$  を最大にする.
- (2)  $\hat{\theta}_n$  は対数尤度関数  $l_n(\theta; \mathbf{x}_n)$  を最小にする.
- (3) 真値  $\theta_0$  について, 対数尤度関数の差  $l_n(\theta_0) - l_n(\theta)$  を最小にする.
- (4)  $\hat{\theta}_n$  はスコア関数  $S(\theta; \mathbf{x}_n)$  の零点である.

**補題 1.1.3 (尤度関数の情報理論的な意味).**

- (1) 対数尤度関数の差は, KL-変分に概収束する:

$$\frac{1}{n}(l_n(\theta) - l_n(\theta_0)) \xrightarrow{P_{\theta_0}\text{-a.e.}} -D(P_{\theta_0}|P_\theta) < 0.$$

差を取らないことには, それぞれの  $l_n(\theta)/n$  の極限は発散する, はず.

- (2) 尤度関数の比  $L_n(\theta)/L_n(\theta_0)$  は,  $e^{-nD(P_{\theta_0}|P_\theta)}$  のレートで収束する.

### 1.1.2 一貫性・漸近正規性を持つ十分条件

**系 1.1.4 (最尤推定量は一貫推定量である [7]Th'm17).** 次を仮定する:

- (M1) 母数空間  $\Theta$  はコンパクト.
- (M2) 第二引数下半連続性:  $\forall x \in \mathcal{X} \ \theta \mapsto p_\theta(x)$  は  $\theta$  について下半連続.
- (M3)  $L^1$  有界性:  $\exists M \in L^1(\mathcal{X}, P_{\theta_0}) \ \forall (x, \theta) \in \mathcal{X} \times \Theta \ |U(x, \theta)| \leq M(x)$ .
- (M4) 第一引数の弱い可測性:  $\forall \theta \in \Theta \ \exists \rho_0 > 0 \ \forall 0 < \rho < \rho_0 \ x \mapsto \sup_{|\theta' - \theta| < \rho} p_{\theta'}(x) \in L(\mathcal{X})$ .
- (M5) 識別可能性:  $(P_\theta)$  は単射.

このとき, 任意の独立観測列  $\{x_n\}_{n \in \mathbb{N}} \subset \mathcal{X}$  について,

$$P_{\theta_0}[\lim_{n \rightarrow \infty} \hat{\theta}_n = \theta_0] = 1.$$

**[証明].**  $M$ -推定量が一貫性を持つための条件 1.4.8 の系である. ■

**要諦 1.1.5.**

- (1)  $\hat{\theta}_n: \mathcal{X}^n \rightarrow \Theta$  の可測性はなくても成り立つ.
- (2)

**系 1.1.6.** 模型 1.1.1 において,

$$Z_n(\theta) := \sqrt{n} E_{\mathbb{P}_n}[S(\theta; \mathbf{x})] = \frac{1}{\sqrt{n}} \sum_{j \in [n]} \partial_\theta \log p_\theta(x_j)$$

とする.  $\theta_0 \in V \stackrel{\text{open}}{\subset} \Theta$  を開近傍とし, 対数尤度  $l_1(\theta; x) = \log p_\theta(x): \mathcal{X} \times \Theta \rightarrow \mathbb{R}_{\leq 0}$  とその導関数  $S(\theta; x) := \partial_\theta \log p_\theta(x)$  について次を仮定する:

- (F1) 第二引数は真値の近傍で  $C^2$ -級: 対数尤度  $\log p_\theta(x): \mathcal{X} \times \Theta \rightarrow \mathbb{R}^p$  は  $V$  上  $C^2$  級.
- (F2) 第一引数可測性:  $\log p_\theta(-)$  は可測.
- (F3)  $S$  は真値において中心化された  $L^2$  関数である:  $S(\theta_0; x) := \partial_\theta \log p_{\theta_0}(x) \in \mathcal{L}^2(P_{\theta_0}; \mathbb{R}^p)$  で,  $E_{P_{\theta_0}}[\psi(x, \theta_0)] = 0$ . さらに, Fisher 情報行列

$$\Gamma = E_{P_{\theta_0}}[\partial_\theta S(x, \theta_0)] = \int_{\mathcal{X}} (\partial_\theta \log p)^\top (\partial_\theta \log p)(x, \theta_0) P_{\theta_0}(dx) = I(\theta_0)$$

は正則.

(F4)  $S$  の導関数は  $L^1$ -有界:  $\exists_{M \in L^1(\mathcal{X}, \mathbb{R})} \forall_{x \in \mathcal{X}, \theta \in V} |\partial_\theta \psi(x, \theta)| = |\partial_\theta^2 \log p_\theta(x)| \leq M(x)$  で, その積分は  $\theta_0$  においては

$$\int_{\mathcal{X}} \partial_\theta^2 \log p_{\theta_0}(x) P_{\theta_0}(dx) = -I(\theta_0).$$

(F5)  $\hat{\theta}_n$  は漸近  $Z$ -推定量である: 確率変数列  $\tilde{\theta}_n: \mathcal{X}^n \rightarrow \Theta$  が  $\hat{\theta}_n \xrightarrow{P} \theta_0$  かつ  $S_n(\hat{\theta}_n) = o_p(\sqrt{n})$ , すなわち,  $Z_n(\hat{\theta}_n) = o_p(1)$  を満足する.

このとき,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) - I(\theta_0)^{-1} Z_n(\theta_0) = o_p(1) \quad (n \rightarrow \infty) \text{ under } \bar{P}.$$

特に,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \Rightarrow N_p(0, I(\theta_0)^{-1}) \quad (n \rightarrow \infty).$$

**要諦 1.1.7.** 条件 (F3) の, スコア関数  $S$  が中心化されているという条件は, 正則性の仮定 (F1),(F4) の下で常に成立する. この系により, 「正則」な統計モデルの中で, 最尤推定量は漸近有効であることが, 漸近決定理論の枠組みから分かる.

### 1.1.3 収束レート

結局, 滑らかなパラメトリックモデルにおける最尤推定量は次の性質を持つ:

- (1) 漸近的一致性を持つ.
- (2) 真値への収束レートが最速である:  $n^{-1/2}$ .
- (3) 極限分布の分散が最小である. 実際, Cramer-Rao を漸近的に満たす.

### 1.1.4 Neyman-Scott の問題

正規分布族  $(N_p(\theta, \sigma^2 I_p))_{\theta \in \mathbb{R}^p, \sigma \in \mathbb{R}^+}$  について,

- (1)  $\sigma^2$  を既知とするならば, 位置母数  $\theta$  に対して, 標本平均  $\bar{x}$  は UMV 不偏推定量である.
- (2) 不偏推定量に限らなければ, James-Stein 推定量はこれを一様に改善する.

正規性の仮定をおかずとも, 一般化線型モデルには次の不偏推定量が存在する:

- (1)  $\theta$  の線型関数は, 推定可能ならば, BLUE が存在する.
- (2)  $R_0^2 := \min_{\theta \in \mathbb{R}^p} |Y - X\theta|^2$  について,  $\hat{\sigma}^2 := \frac{R_0^2}{n-r}$  ( $\text{rank } X =: r$ ) は不偏推定量である.

**定理 1.1.8 (Neyman and Scott (48)).** すべての観測  $Y = (Y_{ij})$  は独立であるとする. このとき, 標本のレプリカの個数  $n$  を固定したとき,  $\sigma^2$  の最尤推定量  $\hat{\sigma}^2$  は  $p \rightarrow \infty$  の極限において一致性を持たない.

**例 1.1.9.**

- (1) 標本のコピーの数を  $n = 2$  として考える. 2つの独立な標本  $\{x_j\}_{j \in \mathbb{N}} \perp \{y_j\}_{j \in \mathbb{N}}$  はそれぞれ  $x_j, y_j \sim N(\mu_j, \sigma^2)$  を満たすとする. このとき,

$$\hat{\mu}_j = \frac{x_j + y_j}{2}, \quad \hat{\sigma}^2 = \frac{1}{4n} \sum_{j \in [n]} (x_j - y_j)^2$$

はいずれも最尤推定量である. しかし,  $\hat{\sigma}^2 \rightarrow \frac{\sigma^2}{2}$  a.s. で一致性を満たさない.

- (2) 混合正規分布

$$p(x, \theta) := (1-t)\phi(x; \mu_1, \sigma_1^2) + t\phi(x; \mu_2, \sigma_2^2) \quad \theta := (t, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2), t \in (0, 1)$$

において、その尤度関数は

$$\forall_{n \geq 2} L_n(\theta; \mathbf{x}_n) \geq (1-t)t^{n-1} \phi(x_1; \mu_1, \sigma_1^2) \prod_{j=2}^n \phi(x_j; \mu_2, \sigma_2^2)$$

と評価出来るが、これは  $(\mu_1, \sigma_1^2) \in \{x_1\} \times (0, \infty)$  なる非コンパクト集合上では  $\sigma_1^2 \searrow 0$  で発散する。このように、母数空間が非コンパクトだと最尤推定量は構成できない。しかしワンステップ推定量は構成できて、これは漸近正規性を有する。

□

## 1.2 漸近決定理論の枠組み

### 漸近的に不偏な推定量

一般のモデルについて、性能評価は漸近論において行う。その際、漸近決定論に依らずとも指標として適当であろうと思われる性質が、「漸近的に不偏推定量」、すなわち、一貫性である。

### 1.2.1 一致推定量の定義

**定義 1.2.1 (consistency).** パラメータの推定量  $\hat{\theta}_n: \mathcal{X}^n \rightarrow \Theta$  について、

- (1) 任意の  $\theta \in \Theta$  について、これが真値であったとき、 $\tilde{\theta}_n \rightarrow \theta$  a.s. ( $n \rightarrow \infty$ ) が成り立つとき、この推定量は**強一貫性**を持つという。
- (2) 推定量  $\tilde{\theta}_n$  が可測で、任意の  $\theta \in \Theta$  について、これが真値であったとき、 $\tilde{\theta}_n \xrightarrow{P} \theta$  ( $n \rightarrow \infty$ ) が成り立つとき、すなわち、 $\tilde{\theta}_n - \theta = o_p(1)$  が成り立つとき、この推定量は**弱一貫性**を持つという。
- (3) 弱一貫な推定量が特に、 $\hat{\theta}_n - \theta = O_p(1/\sqrt{n})$  を満たすとき、これは  $\sqrt{n}$ -一貫性を持つという。
- (4) 推定量  $\hat{\theta}_n: \mathcal{X}^n \rightarrow \Theta$  は自然に経験分布上の関数  $\text{Em}(P(\mathcal{X})) \rightarrow \Theta$  と同一視出来る。この延長  $T: P(\mathcal{X}) \rightarrow \Theta$  が **Fisher の一貫性**を持つとは、 $\forall_{\theta \in \Theta} T(P_\theta) = \theta$  を満たすことをいう。

可測な推定量について、強一致ならば弱一致。

**要諦 1.2.2.** 殆どの場合、最小コントラスト推定量は可測に選べるので、技術的な違いしかない [1.4.3](#)

**補題 1.2.3 (分散が消える不偏推定量は一致推定量である).** 統計量  $U_n: \mathcal{X}^n \rightarrow \Theta$  について、

- (1)  $U_n$  が  $\theta$  にある  $r > 0$  について  $r$ -次平均収束するならば、弱一致推定量である。
- (2) 特に、 $U_n$  が  $\theta$  の不偏推定量かつ  $\text{Var}[U_n] \rightarrow 0$  が成り立つならば、 $\theta$  の弱一致推定量である。

**例 1.2.4 (一致推定量).**

- (1)  $U_n(X) := \bar{X} = \frac{X_1 + \cdots + X_n}{n}$  は母集団平均の強一致推定量である（大数の強法則による）。
- (2) 全く同様のことは、一般の標本モーメント  $\hat{m}_r := \frac{1}{n} \sum_{j=1}^n X_j^r$  が強一致推定量であると言える。
- (3) 一般の標本中心化モーメントも強一致推定量であるが、そのままでは不偏推定量にはならない。
- (4) セミパラメトリックな線型回帰モデル  $Y_j = \alpha + \beta X_j + \epsilon_j$  ( $j \in [n]$ ) の OLS (ordinary least squares estimator) はいくらかの条件の下一致推定量である。
- (5) 自己回帰モデル (AR: Autoregressive Model)  $Y_j = \alpha + \beta Y_{j-1} + \epsilon_j$  の最小二乗推定量

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

は一致推定量である。ただし、 $\bar{Y}_{i,j}$  を  $Y_i$  から  $Y_j$  までのデータを使った標本平均とした。

□

## 1.2.2 一致推定量の存在

これにより、 $\Theta$  全域で  $L^2$ -微分可能で非退化な Fisher 情報行列を持つ識別可能なパラメトリックモデル  $(P_\theta)_{\theta \in \Theta}$  には正則で漸近最適な  $\theta$  の統計量が存在することが示せる。または、極限実験  $\{P_\theta\}$  において、 $\delta_n$  が離散位相を定めていれば良い。

**定理 1.2.5 (Cramer).** 可微分模型 1.5.2 において、 $\theta_0 \in \Theta^\circ, \theta_0 \in B \stackrel{\text{open}}{\subset} \Theta$  を開近傍とする。推定関数  $\psi: \mathcal{X} \times \Theta \rightarrow \mathbb{R}$  について (E1) (E4) 1.5.5 を仮定する。このとき、次が成り立つ：

- (1) 任意の十分小さい開球  $B_0 := B(\theta_0, r)$  と  $\beta \in (0, 1/2)$  について、ある事象列  $\mathcal{X}_n^* \subset \mathcal{X}^n$  が存在して、次を満たす：
  - (a) 十分大きな  $n$  に対して、 $\exists! \hat{\theta}_n \in B_0, \psi_n(\mathcal{X}_n^*; \hat{\theta}_n) = \{0\} \wedge |\hat{\theta}_n - \theta_0| \leq n^{-\beta}$ .
  - (b)  $P^{\otimes n}[\mathcal{X}_n^*] = 1 - o(1) \ (n \rightarrow \infty)$ .
  - (c) (a) が定める対応  $\hat{\theta}_n: \mathcal{X}_n^* \rightarrow \Theta$  は  $\mathcal{X}^n$  全域に可測な延長を持つ。
- (2) (1) で定まる推定量の列  $(\hat{\theta}_n)$  について、次が成り立つ：

$$\sqrt{n}(\hat{\theta}_n - \theta_0) - \Gamma^{-1} \frac{1}{\sqrt{n}} \psi_n(\theta_0) = o_p(1) \ (n \rightarrow \infty), \quad \Gamma := - \int_{\mathcal{X}} \partial_\theta \psi(x, \theta_0) P(dx) \in \text{GL}_p(\mathbb{R}).$$

特に、 $\sqrt{n}(\hat{\theta}_n - \theta_0) \Rightarrow N_p(0, \Sigma)$ .

**要諦 1.2.6.**  $\hat{\theta}_n$  の構成の仕方は真値  $\theta_0$  に依存してしまう。したがって、推定方程式の解が唯一しかない場合のみ、実用的な意味を持つ。

**記法 1.2.7.**  $(\mathcal{X}_n, \mathcal{A}_n, \{P_{n,\theta}\}_{\theta \in \Omega}) \ (\Omega \stackrel{\text{open}}{\subset} \mathbb{R}^p)$  を統計的実験の列とする。

**定義 1.2.8.** 弱一致推定量  $(\hat{\theta}_n)$  が局所一様であるとは、

$$\forall \theta \in \Omega \ \forall \epsilon > 0 \ \exists \delta > 0 \ \lim_{n \rightarrow \infty} \sup_{\|\theta' - \theta\| < \delta} P_{n,\theta'}[\|\hat{\theta}_n - \theta'\| > \epsilon] = 0$$

が成り立つことをいう。

**命題 1.2.9.**

- (1)  $\theta_1, \theta_2 \in \Omega$  に対して、 $P_{n,\theta_1}, P_{n,\theta_2}$  を優越する  $\sigma$ -有限測度  $\mu_n$  が取れる。
- (2)  $L^1(\mu_n)$ -ノルムが、 $\{P_{n,\theta}\}$  上に距離  $\delta_n$  を定める。
- (3) この距離は全変動ノルムの2倍  $\delta_n(\theta_1, \theta_2) = 2 \sup_{A \in \mathcal{A}_n} |P_{n,\theta_1}(A) - P_{n,\theta_2}(A)|$  と特徴付けられる。特に、 $\text{Im } \delta_n \subset [0, 2]$ .

**定理 1.2.10.** 統計的実験列  $(\mathcal{X}_n, \mathcal{A}_n, \{P_{n,\theta}\}_{\theta \in \Omega}) \ (\Omega \stackrel{\text{open}}{\subset} \mathbb{R}^p)$  について、

- (1)  $\theta$  の一致推定量が存在するならば、 $\theta_1 \neq \theta_2 \Rightarrow \lim_{n \rightarrow \infty} \delta_n(\theta_1, \theta_2) = 2$ .
- (2) 局所一様一致推定量が存在するならば、 $\forall \epsilon > 0 \ \delta_{\epsilon} > 0 \ \forall \theta \in \Omega \ \lim_{n \rightarrow \infty} \sup_{\epsilon < \|\theta' - \theta\| < \delta} \delta_n(\theta, \theta') = 2$ .

[証明].

(1)

■

## 1.3 M-推定量

最尤推定量は多くの場合一致性を持つが、この現象はより広いクラスについて成り立つ。M-推定量は、ある目的関数 (一般化対数尤度関数)  $U: \mathcal{X} \times \Theta \rightarrow (-\infty, \infty]$  の経験平均  $\bar{U}_n$  を最大化する  $\theta$  として定義される。すると独立観測の仮定の下で、 $\bar{U}_n$  は独立同分布列の和となるから、一様大数の法則を通じて強一致性を持つための十分条件が得られる。



## 1.3.1 一様大数の法則

真値  $\theta_0 \in \Theta$  はわからないため、あらゆる  $\theta \in \Theta$  について一様な一致性を持ってほしい。そこで、統計推測への応用には、大数の法則を一様ノルムで考えたい。最尤推定量はこの性質を持つ。これは、対数尤度関数が独立確率変数の和の構造を持っていたためである。これを一般化すると、任意の可測-連続関数  $U(x, \theta) : \mathcal{X} \times \Theta \rightarrow \mathbb{R}$  が定める一般化対数尤度関数  $\bar{U}_n(x_1, \dots, x_n; \theta)$  は、母数空間の任意のコンパクト部分集合上で一様に真の対数尤度  $\bar{U}(\theta) = E_{p_\theta}[U(X; \theta)]$  に収束する。

**模型 1.3.1** ( $M$ -推定量が定義可能なモデル). 推定量は  $n \in \mathbb{N}$  に依ってそれぞれ定義域が違うが、これを帰納極限  $\bar{\mathcal{X}}$  上に引き戻して、一元的に議論する。以後、自然な同一視  $\mathcal{X}^n \hookrightarrow \bar{\mathcal{X}}$  を採用する。

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

$$\bar{U}_n(\mathbf{x}_n, \theta) := E_{\bar{P}_n}[U(\mathbf{x}_n, \theta)] = \frac{1}{n} \sum_{j=1}^n U(x_j, \theta)$$

と表す。この  $\bar{U}_n : \mathcal{X}^n \times \Theta \rightarrow \mathbb{R} \cup \{\infty\}$  を、 $U \in L(\mathcal{X} \times \Theta)$  が定める一般化対数尤度関数ということにしよう。ただし、 $\mathbf{x}_n = (x_1, \dots, x_n) \in \mathcal{X}^n$  とした。

- (3) 関数  $U$  の真の平均を  $\bar{U}(\theta) := E_{\bar{P}}[U(\mathbf{x}; \theta)]$  とする。

$U : \mathcal{X} \times \Theta \rightarrow (-\infty, \infty]$  の平均  $\alpha_1(U(\theta)) := E[U(X, \theta)]$  について、明らかに  $\forall_{\theta \in \Theta} P_\theta[\bar{U}_n(\mathbf{x}_n, \theta) \rightarrow \alpha_1(U(\theta))] = 1$ 。これを、 $\theta \in \Theta$  に依らない方法で評価する。

**補題 1.3.2** (単一方向に関する一様対数の法則).  $(\Theta, d)$  を距離空間、 $\Theta_1 \subset \Theta$  をコンパクト集合、 $U : \mathcal{X} \times \Theta \rightarrow [-\infty, \infty]$  を次を満たす関数とする：

- (A1) 第2引数について  $\forall_{x \in \mathcal{X}} \theta \mapsto U(x, \theta)$  は下半連続。  
 (A2)  $\forall_{\theta \in \Theta} r$  が十分に小さければ、 $x \mapsto \inf_{\theta' \in \Theta, d(\theta', \theta) < r} U(x, \theta')$  は可測。  
 (A3) ある可積分関数  $M : \mathcal{X} \rightarrow \mathbb{R}$  が存在して、 $\forall_{(x, \theta) \in \mathcal{X} \times \Theta} U(x, \theta) \geq M(x)$ 。

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

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

- (2)  $\bar{U}$  も  $\theta$  について下半連続。

**要諦 1.3.3.**  $\Theta$  が可分で  $U(-, \theta)$  が可測ならば (A2) のために十分。また、 $U$  が  $\theta_0$  で下半連続とは、次に同値：

- (1)  $\lim_{r \searrow 0} \inf_{\theta \in \Theta, d(\theta, \theta_0) < r} U(\theta) = U(\theta_0)$ .  
 (2)  $\theta_0$  の任意の収束列  $(\theta_n)$  について、 $\liminf_{n \rightarrow \infty} U(\theta_n) \geq U(\theta_0)$ 。

すなわちこの補題は、関数  $\bar{U}$  は下半連続で、関数  $\bar{U}_n$  はこれに下半連続性の位相について収束することを言っている ( $\Theta_1$  を一点とすれば良い)。

**定理 1.3.4** (一様大数の法則 (LeCam)).  $(\Theta, d)$  を距離空間、 $\Theta_1 \subset \Theta$  をコンパクト集合、 $U : \mathcal{X} \times \Theta \rightarrow [-\infty, \infty]$  を関数とし、以下の条件を仮定する：

- (B1) 第二引数連続性： $\forall_{x \in \mathcal{X}} U(x, -) : \Theta \rightarrow \mathbb{R}$  は連続。  
 (B2) 第一引数可測性： $\forall_{\theta \in \Theta} U(-, \theta) : \mathcal{X} \rightarrow \mathbb{R}$  は可測。  
 (B3)  $L^1$ -有界性：ある可積分関数  $M \in L^1(\mathcal{X})$  が存在して、 $\forall_{(x, \theta) \in \mathcal{X} \times \Theta} |U(x, \theta)| \leq M(x)$ 。

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

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

### 1.3.2 一致推定量の収束の速度

$\Theta$  上の関数  $\theta \mapsto P(\Psi_\theta - \Psi_{\theta_0})$  と  $\theta \mapsto E^*[|G_n(\Psi_\theta - \Psi_{\theta_0})|]$  の  $\theta_0$  近くの Holder 指数によって,  $d(\hat{\theta}_n, \theta_0)$  の収束レートが定まる.

**定義 1.3.5.** 一致推定量  $(\hat{\theta}_n)$  が, 無限大に発散する正数列  $\{c_n\} \subset l_\infty(\mathbb{N}; \mathbb{R}_+)$  に関して **オーダー**  $(c_n)$  を持つとは,  $c_n(\hat{\theta}_n - \theta)$  が極限分布を持つこと, すなわち裾が無大へ逃げて分布が潰れてしまうことがないことをいう:

$$\forall \theta_0 \in \Omega \quad \exists \delta > 0 \quad \lim_{L \rightarrow \infty} \limsup_{n \rightarrow \infty} \sup_{\|\theta - \theta_0\| < \delta} P_{n, \theta}[c_n \|\hat{\theta}_n - \theta\| > L] = 0.$$

**記法 1.3.6.** コントラスト関数  $\Psi_\theta(x)$  について, 目的関数  $\theta \mapsto \mathbb{P}_n \Psi_\theta(x)$  を漸近的に最大化する推定量列を  $(\hat{\theta}_n)$  とする.

**定理 1.3.7 (rate of convergence).** ある  $C \in \mathbb{R}, \alpha > \beta \in \mathbb{R}$  と任意の  $n \in \mathbb{N}$  と十分小さい任意の  $\delta > 0$  について, 次が成り立つとする:

$$\sup_{d(\theta, \theta_0) < \delta} P(\Psi_\theta - \Psi_{\theta_0}) \leq -C\delta^\alpha, \quad E^* \left[ \sup_{d(\theta, \theta_0) < \delta} |G_n(\Psi_\theta - \Psi_{\theta_0})| \right] \leq C\delta^\beta.$$

$$\mathbb{P}_n \Psi_{\hat{\theta}_n} \geq \mathbb{P}_n \Psi_{\theta_0} - O_P \left( n^{\frac{\alpha}{2\beta - 2\alpha}} \right), \quad \hat{\theta}_n \xrightarrow{P^*} \theta_0$$

このとき,  $n^{\frac{1}{2\alpha - 2\beta}} d(\hat{\theta}_n, \theta_0) = O_P^*(1)$ .

### 1.3.3 $M$ -推定量の影響関数

Huber and Ronchetti (2009)[Huber, Peter J., and Ronchetti, E. M., 2009] p.47. Hampel et al. (1986) p.85[Hampel, Frank R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel W. A., 2005]  $M$ -推定量の影響関数は, 推定関数  $\psi$  の定数倍になる. 特に最尤推定量について言えば, スコア関数の定数倍となる. また,  $\text{Var}[\text{IF}^2]$  が漸近分散となる. 最尤推定量について, これは Fisher 情報量である.

**定理 1.3.8.**

$$\text{IF}(x; P, \theta) =$$

### 1.3.4 モーメント法

モーメント法はしばしば  $Z$ -推定量でもあるが, ここで扱うこととする. 良い条件の下でモーメント推定量は  $\sqrt{n}$ -一致性と漸近正規性を持つが, 最尤推定量のものより漸近分散は大きい. これはデルタ法の結果である.

**記法 1.3.9.**  $\{P_\theta\}_{\theta \in \Theta} \subset P(\mathcal{X})$  を分布族,  $g: \mathcal{X} \rightarrow \mathbb{R}^p$  を所与の関数とする.  $\psi(x, \theta) := g(x) - \int_{\mathcal{X}} g(y) P_\theta(dy)$  とすると, 真値  $\theta_0 \in \Theta$  について  $E[\psi(x, \theta_0)] = 0$  である. このとき, 次の識別可能条件がなりたつとする

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

このとき,  $\psi(x, \theta)$  の  $P_{\theta_0}$ -平均  $E_{\theta_0}[\psi(x, \theta)]$  の零点は  $\theta_0$  に限る. たとえば  $g(x) = x^n$  と取ると, この条件を満たす.

**定義 1.3.10 (method of moments).** この  $\psi$  を推定関数とした  $M$ -推定量  $\hat{\theta}_n^i$ , すなわち次の  $\theta$  についての方程式の解を  $g$  に関するモーメント推定量という:

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

あるいは、あるノルム  $\|\cdot\|$  について、 $\|\bar{\psi}_n(x, \theta)\|$  を最小にする  $\hat{\theta}$  を探す (これを近似的モーメント推定量 (approximate moment estimator) という). このとき、推定関数  $\psi$  を **モーメント関数** という.

**定理 1.3.11**. 次の条件を仮定する.

(M1)  $e(\theta) : \theta \mapsto P_\theta f \in \mathbb{R}^k$  は  $\Theta \subset \mathbb{R}^k$  上の単射で、 $\theta_0$  上で連続微分可能で、非特異な Jacobi 行列  $e'_{\theta_0} \in \text{GL}_k(\mathbb{R})$  を持つとする.

(M2)  $P_{\theta_0} \|f\|^2 < \infty$ .

このとき、モーメント推定量  $\hat{\theta}_n$  は任意に 1 に近い確率で存在し、次を満たす:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\theta_0} N\left(0, e'_{\theta_0}{}^{-1} P_{\theta_0} f f^\top (e'_{\theta_0}{}^{-1})^\top\right).$$

### 1.3.5 一般化モーメント法

Lars Hansen は Pearson 1894 によるモーメント法を拡張してファイナンスに応用し、2013 年のノーベル経済学賞を受賞した. モーメントに限らず、真の分布を特徴付ける別の母数  $\psi$  で計算しやすいものがあつたら、これに変換することで  $M$ -推定量を構成する. 最尤法は一般化モーメント法の特殊な場合とみなせる.

## 1.4 最小コントラスト推定量

$M$ -推定量の目的関数のような、独立観測とその和としての構造を仮定せず、一般の目的関数  $\bar{U}_n$  を最小化する推定量の性質を考える.

### 1.4.1 定義と例

最尤推定量は、真の分布との対数尤度の差、すなわち KL-分離度を最小化するものとして得られた. 経験分布の真の分布との KL-分離度は、 $n \in \mathbb{N}$  に対応して関数を与える. これを  $n \rightarrow \infty$  の極限で漸近的に最小にするクラスを最小コントラスト推定量という.

**模型 1.4.1** (最小コントラスト推定量が構成可能な模型). 母数空間  $(\Theta, d)$  は距離空間、 $(\bar{X}, \bar{\mathcal{A}}, \bar{P})$  を帰納極限として得られる確率空間とし、

- (1)  $\bar{\Psi} : \bar{\mathcal{X}} \times \Theta \rightarrow (-\infty, \infty]$  を **コントラスト関数** という.
- (2) これは  $\bar{U}_n$  とは別の意味「最小化すべき目的関数」という意味で、「一般化対数尤度関数」の一般化である. 一方で、付随する目的関数  $\bar{U}_n$  に今後は独立観測の構造も、その和である構造も仮定しない.
- (3)  $\hat{\theta} := \arg \min_{\theta \in \Theta} \bar{\Psi}(x; \theta)$  として得られる対応  $\hat{\theta} : \bar{\mathcal{X}} \rightarrow \Theta$  を **最小コントラスト推定量** という.

**例 1.4.2** (最尤推定). 最尤推定量は、第一義的には、 $\bar{\Psi}_n(\theta) := -E_{\mathbb{P}_n}[\log p(x, \theta)]$  を最小化した.

が、一致性の議論の中で、真の分布との対数尤度の差  $|l_n(\theta) - l_n(\tilde{\theta})|$  の最小化を考えた. これは損失関数のようでより直感的である. すなわち、

$$\bar{U}_n(\theta) = E_{\mathbb{P}_n} \left[ \log \frac{p(x, \theta_0)}{p(x, \theta)} \right] \xrightarrow{P_{\theta_0}\text{-a.e.}} D(P_{\theta_0} | P_\theta)$$

と取ることも出来る. この考え方は、 $Z$ -推定量では出来ない、最小コントラスト推定量独自の議論である. □

**定理 1.4.3** (可測選択定理 (von Neumann 49)).  $\bar{\Psi} : \bar{\mathcal{X}} \times \Theta \rightarrow \mathbb{R}$  が次を満たすとき、最小コントラスト推定量  $\hat{\theta} : \bar{\mathcal{X}} \rightarrow \Theta$  は可測に選べる.

- (1)  $(\Theta, d)$  は完備可分.
- (2)  $\bar{\Psi}$  は第一引数について可測で第二引数について連続.
- (3) 下限の達成:  $\forall x \in \bar{\mathcal{X}} \quad \inf_{\theta \in \Theta} \bar{\Psi}(x, \theta) \in \bar{\Psi}(x, \Theta)$ .

## 1.4.2 独立観測を仮定しない強一貫性定理

対数尤度関数  $\bar{\Phi}_n(\mathbf{x}_n; \theta) := \sum_{i \in [n]} \log p_\theta(x_i)$  に付随する KL-変分  $\bar{U}_n(\mathbf{x}; \theta) := \frac{1}{n}(l_n(\theta_0) - l_n(\theta))$  を最小化しても良いのであった 1.1.2. この部分を抽象化し、 $M$ -推定量としての構造を全く仮定しないこととする。このときでも、単一方向に対する一様大数の法則と、もう一方に関する到達可能性 ( $\bar{U}_n, \bar{U}$  をうまく設定していること) を持つならば、強一貫性を持つ。

**定理 1.4.4 (最小コントラスト推定量の強一貫性の十分条件).** 関数列  $\bar{U}_n$ , 関数  $\bar{U} : \Theta \rightarrow (-\infty, \infty]$ , 空でない部分集合  $\Theta' \subset \Theta$ , 写像列  $\hat{\theta}_n : \mathcal{X} \rightarrow \Theta$  は次の3条件を満たすとする。

(C1) 単一方向に関する一様大数の法則:  $\forall \epsilon > 0 \liminf_{n \rightarrow \infty} \inf_{\theta \in \Theta: d(\theta, \Theta') \geq \epsilon} \bar{U}_n(\theta) \geq \inf_{\theta \in \Theta: d(\theta, \Theta') \geq \epsilon} \bar{U}(\theta) \text{ a.s.}$

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

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

(C4) 到達可能性:  $\limsup_{n \rightarrow \infty} \inf_{\theta \in \Theta'} \bar{U}_n(\theta) \leq \inf_{\theta \in \Theta'} \bar{U}(\theta) \text{ a.s.}$

このとき、次が成り立つ:  $\lim_{n \rightarrow \infty} d(\hat{\theta}_n, \Theta') = 0 \text{ a.s.}$

**要諦 1.4.5.** (C4) は,

$$\exists \theta_0 \in \Theta' \limsup_{n \rightarrow \infty} \bar{U}_n(\theta_0) \leq \bar{U}(\theta_0)$$

が十分条件としてあり、 $\bar{U}$  が  $\bar{U}_n$  の極限関数として得られる保証である。

## 1.4.3 一貫性を持つ推定量の例

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

**例 1.4.6 (最尤推定量).**  $\{P_\theta\} \ll \nu$  (模型 1.1.1) に対する最尤推定量  $\hat{\theta}_0$  は、定理から一貫性を持つことが認められる。これは独立観測の構造によるので、一般に  $M$ -推定量について議論出来る。  $\square$

**例 1.4.7 (多項分布に対する最尤推定量).**  $k$  個の事象への分割  $\Omega = \sum_{i=1}^k E_i$  を考え、 $P_\theta[E_i] =: \theta^i$  を  $k$  個のパラメータとする。この試行を  $n$  回繰り返し、 $E_i$  が観測された回数を  $y_i$  とすると、これは多項分布に従う確率変数である:  $(y_1, \dots, y_k) \sim M(n; \theta^1, \dots, \theta^k)$ . パラメータ空間を

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

とする。コントラスト関数を

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

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

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

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

### 1.4.4 独立観測の構造を仮定した場合

$M$ -推定量は最小コントラスト推定量の特別な場合である．特に最尤推定量は  $U(x, \theta) := \Psi(x, \theta) - \Psi(x, \theta_0)$  とした場合である．

**定理 1.4.8** ( $M$ -推定量の強一致性の十分条件). 模型 1.3.1 において,  $(\Theta, d)$  をコンパクト,  $\Theta' := \left\{ \theta \in \Theta \mid \bar{U}(\theta) = \inf_{\hat{\theta} \in \Theta} \bar{U}(\hat{\theta}) \right\}$  とし,  $\hat{\theta}_n : \mathcal{X}^n \rightarrow \Theta$  は写像列で (A1),(A2),(A3)1.3.2 と (C3)1.4.4 を満たすとする．このとき,

$$\lim_{n \rightarrow \infty} d(\hat{\theta}_n, \Theta') = 0 \text{ a.s.}$$

### 1.4.5 コンパクト集合に終局するための条件

## 1.5 $Z$ -推定量

#### 最尤法のノンパラメトリック化

最尤推定量は対数尤度関数を最大化する推定量で, 最小コントラスト推定量は真の分布との分離度を最小化する推定量であった．このように, 損失関数または効用関数 (2つの概念を併せて目的関数) の極値点として与えられる推定量は, モデルが微分可能であるとき, これらは導関数の零点として得られる推定量として一般化できる．これを  $M$ -推定量という.<sup>a</sup> この一般化は, 分布関数  $f$  への言及を除去しており, ノンパラメトリック化と言える．

Huber が頑健推定の動機で 1964 年に最尤推定を一般化し, "maximum likelihood-type" の頭文字から  $M$ -推定量のクラスを定めた.<sup>b</sup> 実はこれは可微分構造を持つ正則なモデルに関して最適でもある．このセミパラメトリックな一般化を経験尤度法 (Owen 1988) という．

<sup>a</sup> この意味での推定量を  $Z$ -推定量とし,  $M$ -推定量は効用関数の最大化で得るものと狭義に解釈することもある．

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

### 1.5.1 定義と例

$M$ -推定量はロバスト推定の代表である．

**模型 1.5.1** ( $Z$ -推定量が構成可能なモデル).  $M$ -推定量の模型 1.3.1 と同様, 自然な同一視  $\mathcal{X}^n \hookrightarrow \overline{\mathcal{X}}$  の下で考える．

(1) 関数  $\psi : \mathcal{X} \times \Theta \rightarrow \mathbb{R}^p$  の経験平均

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

を**推定関数**という．

(2) 推定方程式の解  $\hat{\theta}_n \in \{\theta \in \Theta \mid \bar{\psi}_n(\theta) = 0\}$  として構成される推定量  $\hat{\theta}_n : \mathcal{X}^n \rightarrow \Theta$  を  **$Z$ -推定量**という．

(3)  $\lim_{n \rightarrow \infty} \bar{\psi}_n(\hat{\theta}_n) \rightarrow 0$  a.s. を満たすとき, **漸近  $Z$ -推定量**という．

**模型 1.5.2** (可微分モデル).  $M$ -推定量の模型 1.3.1 と同様, 自然な同一視  $\mathcal{X}^n \hookrightarrow \overline{\mathcal{X}}$  の下で考える．さらに,  $\{P_\theta\} \subset P(\mathcal{X})$  はある  $\sigma$ -有限測度に支配され, 密度  $(p_\theta)$  は  $\Theta$  上微分可能であるとする．

(1)  $\Theta$  上可微分なコントラスト関数  $\Phi_n : \mathcal{X}^n \times \Theta \rightarrow (-\infty, \infty]$  について,  $\bar{\psi}_n := \partial_\theta \Psi_n$  を推定関数とした  $Z$ -推定量を考えれば良いから, この場合に最小コントラスト推定量 1.4.1 は  $Z$ -推定量でもある．

(2) 以降,  $\psi_n(\mathbf{x}_n; \theta) := n\bar{\psi}_n(\theta) = \sum_{j \in [n]} \psi(x_j; \theta)$  なる記法も採用する．

(3)  $\Theta \subset \mathbb{R}^p$  は可測集合,  $\psi$  は  $\Theta$  上連続とする．このとき,  $\forall_{\hat{\theta}_n \in L(\mathcal{X}^n; \Theta)} \psi(\mathbf{x}_n; \hat{\theta}(\mathbf{x}_n)) \in \mathcal{L}(\mathcal{X}^n)$ ．

**例 1.5.3.**

- (1) 最尤推定量は、スコア関数を推定関数とする  $Z$ -推定量であり、尤度方程式  $S_n(x, \theta) = \partial_\theta l(\theta, x) = 0$  の零点として特徴付けられる。
- (2) 次を推定関数に持つ  $Z$ -推定量のクラスを  $g$  の定めるモーメント推定量という：

$$\psi(x; \theta) := g(x) - \int_{\mathcal{X}} g(x') P_\theta(dx') \quad g \in L^1(\mathcal{X}; \mathbb{R}^p).$$

その簡明さから古くから用いられているが、適用可能なケースは殆ど最尤推定量と同じであるが、漸近分散は最尤推定量のものより一般に大きい。さらに、ロバスト性の見地からいまいちである。

- (3) 既知の分散  $\sigma^2 > 0$  について、モデル  $(N(\theta, \sigma^2))_{\theta \in \mathbb{R}}$  を考える。関数  $\psi(x, \theta) = g(x - \theta)$  が定める  $Z$ -推定量は自然な対象である。なお、分布族の affine 変換不変性から、 $\theta_0 = 0$  の場合を考えれば十分である。最尤推定量の影響関数は、 $\theta_0 = 0$  のとき  $\text{IF}(x) = x$  となり、有界でない。一方で、この  $\psi$  の代わりに、剪断する Huber の  $\psi(x) = (-K) \vee x \wedge K$  を用いたものを、頑健回帰という。これが定める  $Z$ -推定量は一致性、漸近正規性を満たし、さらに影響関数が有界になる、 $B$ -ロバスト推定量である。ただし、漸近有効ではなく、定数  $K$  が有効性とロバスト性のトレードオフになっている。この  $K$  を調整定数 (tuning constant) という。

□

### 1.5.2 漸近 $Z$ -推定量の一致性

**定理 1.5.4** ( $Z$ -推定量が一致性を持つための条件).  $\Theta \subset \mathbb{R}^p$  をコンパクト集合、 $\psi := (\psi^a)_{a \in [p]} : \mathcal{X} \times \Theta \rightarrow \mathbb{R}^p$  は次を満たすとする：

- (D1) 第二引数連続性： $\forall_{x \in \mathcal{X}} \psi(x, -) : \Theta \rightarrow \mathbb{R}^p$  は連続。
- (D2) 第一引数可測性： $\forall_{\theta \in \Theta} \psi(-, \theta) : \mathcal{X} \rightarrow \mathbb{R}^p$  は可測。
- (D3)  $L^1$ -有界性： $\exists_{M \in L^1(\mathcal{X})} \forall_{(x, \theta) \in \mathcal{X} \times \Theta} |\psi(x, \theta)| \leq M(x)$ 。
- (D4) 極限解の存在： $\Theta' := \left\{ \theta \in \Theta \mid \int_{\mathcal{X}} \psi(x, \theta) P(dx) = 0 \in \mathbb{R}^p \right\} \neq \emptyset$  とする。

このとき、漸近  $Z$ -推定量  $\hat{\theta}_n$  は強一致性を持つ： $d(\hat{\theta}_n, \Theta') \rightarrow 0$  a.s.

### 1.5.3 コンパクト集合に終局するための条件

#### 1.5.4 一致性を持つ $Z$ -推定量の漸近分布

一致性は大数の法則に当たるならば、その真値への収束の速度と分布が同様に気になる。これが漸近正規性である。これを、可微分模型 1.5.2 に於ける  $Z$ -推定量について議論する。

**定理 1.5.5** (漸近  $Z$ -推定量が漸近正規であるための十分条件). 可微分模型 1.5.2 において、 $\theta_0 \in \Theta^\circ$  とし、 $B \stackrel{\text{open}}{\subset} \Theta$  を  $\theta_0$  の開近傍とする。推定関数  $\psi : \mathcal{X} \times \Theta \rightarrow \mathbb{R}$  は次の 5 条件を満たすとする：

- (E1) 第二引数は真値の近傍で  $C^1$ -級：推定関数  $\psi : \mathcal{X} \times \Theta \rightarrow \mathbb{R}^p$  は  $B$  上  $C^1$  級。
- (E2) 第一引数可測性： $\psi(-, \theta)$  は可測。
- (E3) 真値において中心化された  $L^2$  関数である： $\psi(x, \theta_0) \in \mathcal{L}^2(P; \mathbb{R}^p)$  で、 $E_P[\psi(x, \theta_0)] = 0$ 。
- (E4) 導関数は  $L^1$ -有界： $\exists_{M \in \mathcal{L}^1(\mathcal{X}, \mathbb{R})} \forall_{x \in \mathcal{X}, \theta \in B} |\partial_\theta \psi(x, \theta)| \leq M(x)$ 。さらに、 $\Gamma := -P \partial_\theta \psi(x, \theta_0)$  は正則。
- (E5)  $\hat{\theta}_n$  は漸近  $Z$ -推定量である：確率変数列  $\tilde{\theta}_n : \mathcal{X}^n \rightarrow \Theta$  が  $\hat{\theta}_n \xrightarrow{P} \theta_0$  かつ  $\psi_n(\hat{\theta}_n) = o_p(\sqrt{n})$  を満足する。

このとき、

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

特に、

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N_p(0, \Sigma) \quad (n \rightarrow \infty), \quad \Sigma := \Gamma^{-1} \Phi (\Gamma^\top)^{-1}, \Phi := \int_{\mathcal{X}} \phi \phi^\top(x, \theta_0) P(dx)$$

**要諦 1.5.6.**  $I$  が



## 1.5.5 LeCam のワンステップ改善による最適推定量の構成

$Z$ -推定量が、方程式を解くことの困難さのために明示的に得られないとしよう。すなわち、漸近  $Z$ -推定量  $\hat{\theta}_n$  は直接得られない、1.5.5(E5) を満たすものが見つからないとする。このとき、 $\sqrt{n}$ -一致性を持つ推定量さえ見つければ、これを  $(\psi_n$  をスコア関数とするときは最尤推定量と同じ漸近分散を持つまで) 改善する算譜が存在する。実は、この算譜は極めて自由度が高い。

- (1)  $\hat{\theta}_n^0$  が  $\sqrt{n}$ -一致性を持つとき、 $\hat{\theta}_n := \hat{\theta}_n^0 - (\partial_\theta \psi_n(\hat{\theta}_n^0))^{-1} \psi_n(\hat{\theta}_n^0)$  とすると、再び  $Z$ -推定量と同じ漸近分布を持つ。 $-\partial_\theta \psi_n(\theta)$  を観測情報行列という。
- (2)  $\hat{\theta}_n^* := \hat{\theta}_n^0 - (n\eta^*(\hat{\theta}_n^0))^{-1}(\hat{\theta}_n^0)$  も同じ漸近分散を持つ。 $-n\eta^*(\theta)$  はスコア行列の一般化である。

**定義 1.5.7 (one-step estimator).**  $\Theta$ -値確率変数列  $(\hat{\theta}_n^0)$  について、

- (1) 新たな推定量の系列  $(\hat{\theta}_n)$  を

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

で定めると、これは

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

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

- (2) 新たな推定量の系列  $(\hat{\theta}_n^*)$  を

$$\hat{\theta}_n^* := \hat{\theta}_n^0 - (n\eta^*(\hat{\theta}_n^0))^{-1}(\hat{\theta}_n^0), \quad \eta^*(\theta) := E_{P_\theta}[\partial_\theta \psi(X; \theta)] = \int_{\mathcal{X}} \partial_\theta \psi(x; \theta) P_\theta(dx).$$

$-n\eta^*(\theta)$  は Fisher 情報行列に当たる。

**要諦 1.5.8** ( $\hat{\theta}_n^*$  は  $\hat{\theta}_n$  と漸近同等).  $\eta^*$  が真値  $\theta_0$  の近傍で連続ならば、 $\hat{\theta}_n^*$  は  $\hat{\theta}_n$  と漸近同等になる。実は、推定関数の標本平均  $\bar{\psi}_n$  が、ある正則行列  $\dot{\psi}_0 \in \mathbb{R}^p$  について次を満たすとき、

$$\sup_{\sqrt{n}\|\theta - \theta_0\| < M} \left\| \sqrt{n}(\bar{\psi}_n(\theta) - \bar{\psi}_n(\theta_0)) - \dot{\psi}_0 \sqrt{n}(\theta - \theta_0) \right\| \xrightarrow{P} 0.$$

係数  $(\partial_\theta \psi_n(\hat{\theta}_n^0))^{-1}$  は任意の推定量で良い。このとき、任意のランダムな行列  $\dot{\psi}_{n,0}$  はある  $\dot{\psi}_0$  に収束するならば、

$$\hat{\theta}_n := \hat{\theta}_n^0 - \dot{\psi}_{n,0}^{-1} \bar{\psi}_n(\hat{\theta}_n^0)$$

で定まる推定量も **ワンステップ推定量** といい、次の定理を満たす。

**定理 1.5.9 (ワンステップ推定量が漸近正規であるための十分条件).**  $\theta_0 \in \Theta^\circ$ ,  $B \subset \Theta$  を  $\theta_0$  の開近傍とする。推定関数  $\psi : \mathcal{X} \times \Theta \rightarrow \mathbb{R}$  は条件 (E1)~(E4) 1.5.5 を満足し、初期推定量  $(\hat{\theta}_n^0)$  は次の条件を満足するとする：

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

このとき、ワンステップ推定量  $\hat{\theta}_n$  に対して、通常の漸近  $Z$ -推定量の漸近正規性 1.5.5 と同様の結果が成り立つ：

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -\Gamma^{-1} \frac{1}{\sqrt{n}} \psi_n(\theta_0) + o_p(1).$$

特に、 $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N_p(0, \Sigma)$ .

<sup>†1</sup> 条件  $\hat{\theta}_n^0 - \theta_0 = O_p(1/\sqrt{n})$  は、弱一一致性  $\hat{\theta}_n^0 - \theta_0 = o_p(1)$  を含意することに注意。これに加えて、 $n^{-1/2}$  倍の範囲で真値  $\theta_0$  を捕まえていることは要求する。

## 1.5.6 Z-推定量の影響関数

1964年にHuberによって提唱された。推定量の、新たな観測点  $x_{n+1} \in \mathcal{X}$  に対する感度を分析することで、推定算譜の頑健性を評価できるだろう。

**模型 1.5.10** (推定量に対する感度分析). Z-推定量の模型 1.5.1 を考える。母数値関数  $T: P(\mathcal{X}) \rightarrow \Theta$  は

$$\forall P \in P(\mathcal{X}) \quad \int_{\mathcal{X}} \psi(x, T(P)) P(dx) = 0$$

を満たすとする。これは、任意の経験分布  $\mathbb{P}_n$  について  $\hat{\theta}_n = T(\mathbb{P}_n)$  を満たす対応  $T: \text{Em}(P(\mathcal{X})) \rightarrow \Theta$  の延長である。一般に  $T: P(\mathcal{X}) \rightarrow \Theta$  を微分して感度を分析することが考えられるが、特に極点  $\mathbb{P}_n \in \text{Em}(P(\mathcal{X}))$  の  $\delta_x$  方向への摂動を考えると、**新たな観測に対する Z-推定量  $\hat{\theta}_n$  の感度が考察できるだろう。**

- (1)  $P \in P(\mathcal{X}), x \in \mathcal{X}$  について、 $P$  を通る  $x$ -道とは、曲線  $(P_\epsilon^x := (1 - \epsilon)P + \epsilon\delta_x)_{\epsilon \in [0,1]}: [0,1] \rightarrow P(\mathcal{X})$  をいう。
- (2)  $\Theta \overset{\text{open}}{\subset} \mathbb{R}^p$  とし、適切なモデルの正則性を仮定する。このとき、

$$\text{IF}(x; T, P) := \frac{dT(P_\epsilon^x)}{d\epsilon}(0) := \Gamma(P)^{-1} \psi(x; T(P)), \quad \Gamma(P) := - \int_{\mathcal{X}} \partial_\theta \psi(y; T(P)) P(dy).$$

で定まる関数  $\text{IF}_{T,P}: \mathcal{X} \rightarrow \mathbb{R}^p$  を **母数  $T$  の  $P$  における影響関数** という。

- (3) 影響関数が  $\mathcal{X}$  上有界になるような推定量  $T: P(\mathcal{X}) \rightarrow \Theta$  は、頑健性の観点から好ましいと考えられる、これを **B-ロバスト推定量 (bias robust estimator)** という:  $\exists M \in \mathbb{R} \quad |\text{IF}(x; T, P)| = |\Gamma(P)^{-1} \psi(x; \theta)| \leq M$ .
- (4)

**例 1.5.11.**

- (1)  $(N(\theta, 1))_{\theta \in \mathbb{R}}$  の位置母数最尤推定量  $\hat{\theta}_n$  の影響関数は  $\text{IF}(x; T, N(\theta, 1)) = x - \theta$  である。真の分布  $\theta$  から観測が離れるにつれて、線型に影響がましていく。このように、最尤推定量は分散の意味で最適であるが、一般に影響関数は有界にはならない。

□

## 1.6 尤度比検定とその同値な検定

尤度比検定統計量  $\Lambda_n^*$ , Rao 検定量  $S_n$ , Wald 検定量  $W_n$  はいずれも漸近同等である。

## 1.6.1 定義と漸近分布

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

**模型 1.6.1** (単純仮説の尤度比検定).  $\Theta \subset \mathbb{R}^p, \theta_0 \in \Theta^\circ$  とし、 $\sigma$ -有限測度  $\nu$  に支配されているモデル  $\{P_\theta\}_{\theta \in \Theta} \subset P(\mathcal{X})$  上で問題  $H_0: \theta = \theta_0$  を考える。

- (1) 尤度比

$$\Lambda_n := \frac{L_n(\theta_0)}{\sup_{\theta \in \Theta} L_n(\theta)} = \frac{L_n(\theta_0)}{L_n(\hat{\theta}_n)}$$

の値が小さい時に帰無仮説  $H_0$  を棄却する検定を **尤度比検定** という。ただし、 $\hat{\theta}_n$  は  $\theta$  の最尤推定量とした。

- (2) 尤度というのは確らしさで、これが大きいほどたしからしい。最尤推定量の尤度に比べて  $L_n(\theta_0)$  が十分小さい場合は  $H_0$  は疑わしいことになる。
- (3) 一方で、尤度比  $\Lambda_n$  の精密分布は複雑になるため、極限分布で代用することになる。



(4) 次を基準化されたスコア関数  $Z_n : \mathcal{X}^n \times \Theta \rightarrow \mathbb{R}^p$  という：

$$Z_n(\theta) := \frac{1}{\sqrt{n}} \sum_{j \in [n]} \partial_\theta \log p_\theta(x_j).$$

$Z := Z(\theta_0) \in \mathbb{R}^p$  とすると,  $\theta_0$  が最尤推定量に一致するならば  $Z = 0$  である.

**定理 1.6.2.** 模型 1.6.1 には検定点  $\theta_0$  での弱一致性  $\hat{\theta}_n \xrightarrow{P_{\theta_0}} \theta_0$  を満たす最尤推定量  $\hat{\theta}_n$  が存在し, これ 1.1.1 が漸近正規になるための条件 (F1) (F4) 1.1.6 を満たすとする. このとき,

$$-2\text{Log } \Lambda_n \Rightarrow \chi^2(p) \quad (n \rightarrow \infty) \quad \text{under } P_{\theta_0}.$$

### 1.6.2 複合仮説の検定

殆ど変わらず,  $p$  を  $r$  に変えるのみである.

**模型 1.6.3** (複合仮説の尤度比検定).  $\Theta \overset{\text{open}}{\subset} \mathbb{R}^p, \Theta_0 \subset \Theta$  とする.  $\theta_0 \in \Theta_0$  の開近傍上の  $C^2$ -級関数  $R : V \rightarrow \mathbb{R}^r$  が存在して,  $\Theta_0 \cap V = R^{-1}(0)$  が成り立つとする. ただし,  $1 \leq r \leq p-1$  かつ  $\partial_\theta R(\theta_0) \in M_r(\mathbb{R})$  は非退化である. 検定問題  $H_0 : \theta \in \Theta_0$  を考える.

(1) 次による検定が尤度比検定を与える：

$$\Lambda_n^* := \frac{\sup_{\theta \in \Theta_0} L_n(\theta)}{\sup_{\theta \in \Theta} L_n(\theta)}.$$

(2) これが小さいとき, 帰無仮説  $H_0$  を棄却する.

(3) 制約のない最尤推定量を  $\hat{\theta}_n$ ,  $\Theta_0$  の下での最尤推定量を  $\hat{\tau}_n$  とする.

**定理 1.6.4.** 上の模型には検定点  $\theta_0$  での弱一致性  $\hat{\theta}_n \xrightarrow{P_{\theta_0}} \theta_0$  を満たす最尤推定量  $\hat{\theta}_n$  と  $\hat{\tau}_n$  が存在し, これ 1.1.1 が漸近正規になるための条件 (F1) (F4) 1.1.6 を満たすとする. このとき,

$$-2\text{Log } \Lambda_n^* \Rightarrow \chi^2(r) \quad (n \rightarrow \infty) \quad \text{under } P_{\theta_0}.$$

### 1.6.3 Rao 検定と Wald 検定

複合仮説 1.6.3 の検定に関して, 尤度比検定に変わって, ラオ検定 (Lagrange 未定乗数法検定) や Wald 検定を用いることが出来る. これらは  $H_0$  の下で尤度比検定統計量  $\Lambda_n^*$  と同じ漸近分布  $\chi^2(r)$  を持つ. 特に単純仮説であったとき, Rao 検定においては最尤推定量を計算する必要はない.

**定義 1.6.5.** 複合仮説 1.6.3 について,

(1) 次が大きいときに  $H_0$  を棄却する検定が考えられる：

$$S_n := Z_n(\hat{\tau}_n)^\top I(\hat{\tau}_n)^{-1} Z_n(\hat{\tau}_n).$$

これをラオ検定 (Lagrange 未定乗数法検定) という.

(2) 次が大きいときに  $H_0$  を棄却する検定が考えられる：

$$W_n := nR(\hat{\theta}_n)^\top C(\hat{\theta}_n)^{-1} R(\hat{\theta}_n), \quad C(\theta) := \partial_\theta R(\theta) I(\theta)^{-1} (\partial_\theta R(\theta))^\top.$$

これをワルド検定という.

**定理 1.6.6.**  $S_n, W_n$  はいずれも  $-2\log \Lambda_n^*$  と漸近同等である.

## 1.6.4 多項分布の検定

正則性を仮定すれば、多項分布の検定においては、尤度比検定  $-2\log \Lambda_n^*$  と漸近同等な統計量  $Q_n, Q_n^*$  が見つかる。

**模型 1.6.7** (Fisher の  $\chi^2$ -検定).  $\mathbb{R}^k$  の標準基底  $\Omega := \{e_1, \dots, e_k\}$  を確率空間とし、分布の全体を

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

と表す.  $1 \leq m < l \leq k-1$  とする. パラメトリックモデル  $p = (p_\theta)_{\theta \in \Theta} : \mathbb{R}^l \xrightarrow{\text{open}} \Theta \rightarrow \Delta^\circ$  は  $C^1$  級の単射で,  $\theta : \mathbb{R}^m \xrightarrow{\text{open}} U \rightarrow \Theta \subset \mathbb{R}^l$  も  $C^1$ -級の単射として, これによって定まる部分パラメトリックモデル  $(p_\theta)_{\theta \in \theta(U)}$  についての問題  $H_0 : \theta \in \theta(U)$  を考える.

(1) 最尤推定量と同様, スコア関数と漸近同等な  $\theta$  の推定量  $\hat{\theta}_n$  と  $u$  の推定量  $\hat{u}_n$  が得られているとする:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) - I^{-1}Z_n = o_p(1), \quad I := \sum_{i \in [k]} \frac{1}{p_i(\theta_0)} ((\partial_\theta p_i)^\top \partial_\theta p_i)(\theta_0), \quad Z_n = \frac{1}{\sqrt{n}} \sum_{j \in [n]} (\partial_\theta \log p_{\theta_0}(x_j))^\top.$$

$$\sqrt{n}(\hat{u}_n - u_0) - \mathcal{G}^{-1}J^\top Z_n = o_p(1), \quad \mathcal{G} := J^\top I J, \quad J := \frac{\partial \theta}{\partial u}(u_0).$$

ただし,  $I$  は  $\theta_0 \in \Theta$  における Fisher 情報行列,  $\mathcal{G}$  は  $u_0 \in \theta(U)$  におけるそれである.

(2) それぞれの検定統計量の値が大きい時に  $H_0$  を棄却する:

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

**定理 1.6.8.**  $H_0$  の下での  $Q_n, Q_n^*$  の漸近分布は  $\chi^2(l-m)$  である.

**要諦 1.6.9.** すなわち, 尤度比検定

$$-2\log \Lambda_n^* = -2\log \frac{\prod_{j \in [n]} p \circ \theta(\hat{u}_n)^{x_j}}{\prod_{j \in [n]} p(\hat{\theta}_n)^{x_j}}$$

と漸近同等である.

**例 1.6.10.**

(1) モデル  $\Delta$  内の適合度検定  $H_0 : (p_i)_{i \in [k]} = (p_i^*)_{i \in [k]}$  は, 単純仮説の検定であるが, 再び  $Q_n, Q_n^*$  の漸近分布は  $\chi^2(k-1)$  である.

(2) モデル  $\Delta$  の独立性の検定  $H_0 : (p_{i-})_{i \in [r]} \perp (p_{-j})_{j \in [c]}$  の検定統計量の漸近分布は  $\chi^2((r-1)(c-1))$  である.

□

## 第 2 章

# 非径数的手法

### 2.1 情報量規準

情報量規準という、モデルを比較する形式を手に入れたことで、検定と推定を統合する視座が得られる。これは標本空間上の分布の空間  $P(\mathcal{X})$  上の測度  $AIC : \mathcal{B}(P(\mathcal{X})) \rightarrow \mathbb{R}$  であるか？幾何学的に言えば、モデル  $\mathcal{M} \subset P(\mathcal{X})$  の、真の分布  $P_0 \in P(\mathcal{X})$  との距離の遠さと、大きさ（次元）の和である。

#### 2.1.1 KL-変分によるモデル評価

**模型 2.1.1** (パラメトリックモデルの評価).  $\sigma$ -有限測度  $\nu$  に支配されたモデル  $\{P_\theta\} \subset P(\mathcal{X})$  と真の分布  $\mu \gg Q \in P(\mathcal{X})$  との関係を考える。  $\hat{\theta}_n : \mathcal{X}^n \rightarrow \Theta$  を推定量とすると、代入  $P_{\hat{\theta}_n}$  が分布に対する推定量となる。この、分布に対する推定の精度を、Banach 空間  $P(\mathcal{X})$  上の距離や変分を用いて評価したい。

**例 2.1.2** (最尤法とは真の分布からの KL-情報量を最小にする分布を採用する手法). KL 情報量の最小化と交差エントロピーの最大化は同値である。そして交差エントロピーに経験分布を代入したものは対数尤度関数に他ならない：

$$h(p, \hat{\theta}_n(x); \mathbb{P}_n) = E_{\mathbb{P}_n}[\log p_{\hat{\theta}_n}(X)] = l_n(\mathbf{x}_n; \hat{\theta}_n).$$

これが情報論 (エントロピー理論) からみた最尤法の意味である。

(1) 真の分布  $Q$  と予測分布  $P_{\hat{\theta}_n}$  の KL-分離度は

$$D(Q|P_{\hat{\theta}_n}) = \int_{\mathcal{X}} q(z) \log \frac{q(z)}{p_{\hat{\theta}_n}(z)} \nu(dz) = \int_{\mathcal{X}} q(z) \log q(z) \nu(dz) - \int_{\mathcal{X}} q(z) \log p(z, \hat{\theta}_n) \nu(dz).$$

ここで、モデル  $p$  と推定量  $\hat{\theta}_n$  の選択に依るのは第二項の交差 (連続) エントロピー  $h_{p_{\hat{\theta}_n}, Q}$

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

である。

(2) 真の分布  $Q$  からの交差エントロピーは、 $Q$  には何の仮定も置いていないため、経験分布  $\mathbb{P}_n$  から推定することになる：プラグイン推定量  $h(p, \hat{\theta}_n(x); \mathbb{P}_n)$  を用いる。実はこれが対数尤度関数  $l_n$  である！

□

**定義 2.1.3 (information criterion).** 尤度関数にペナルティ項  $\mathbf{p}$  を付けた

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

を、推定量  $\hat{\theta}_n : \mathcal{X}^n \rightarrow \Theta$  に対する情報量規準という。一般に  $\mathbf{p}$  は真の分布  $Q$  に依存するので、一致推定量で置き換える。

**例 2.1.4.** バイアス補正の観点から、 $\mathbf{p}(n) = \text{Tr}(C)$  とおけば、IC は交差エントロピー (の逆符号) の一致推定量である。従って IC の値が小さいほどモデルによる推定量  $P_{\hat{\theta}_*}$  と真の分布  $Q$  との KL-変分が小さいため、良いモデルだと考えられる。

(1) 推定量  $\hat{\theta}_n$  が  $\Psi$  の定める最小コントラスト推定量であるとき,  $E_Q[\partial_\theta \Psi(x; \theta_*)] = 0$  であるから,

$$\mathbf{p}(n) = -\text{Tr} \left( \left( \int_{\mathcal{X}} \partial_\theta^2 \Psi(x; \theta_*) Q(dx) \right)^{-1} \int_{\mathcal{X}} (\partial_\theta \Psi(x; \theta_*))^\top \partial_\theta \log p(x; \theta_*) Q(dx) \right)$$

となる. この場合を小西-北川の一般化情報量規準 (の特殊な場合) という.

(2) さらに  $\hat{\theta}_n$  が最尤推定量であるとき,  $E_Q[s(X; \theta_*)] = 0$  より,

$$C = J_Q^{-1} I_Q \quad I_Q := \int_{\mathcal{X}} (\partial_\theta \log p)^\top \partial_\theta \log p(x; \theta_*) Q(dx), J_Q := - \int_{\mathcal{X}} \partial_\theta^2 \log p(x; \theta_*) Q(dx).$$

となるから, 補正項はその一致推定量を取る. この場合を竹内の情報量規準という.

(3)  $\mathbf{p}(n) = \mathbf{k}$  を  $\Theta \subset \mathbb{R}^k$  の次元とすると, これを赤池情報量規準という. これは,  $\hat{\theta}_n$  が最尤推定量で,  $Q \in \{P_\theta\}_{\theta \in \Theta}$  が成り立つ場合の, 竹内の情報量規準の退化した場合でもある.

歴史的に最初に提案されたのは赤池情報量規準で, これはモデルの misspecification に脆弱である. しかし一般に, 情報量規準はパラメータ数を過大評価することが証明出来る.  $\square$

### 2.1.2 交差エントロピーの推定

真の分布と, そのモデルがはじき出す予測分布との分離度を, そのモデルの misspecification の度合いを測る尺度に用いるという発想である. すると, 「モデル  $f$  の平均対数尤度が大きいほど良いモデル」ということになる. しかし, 分離度を経験分布から求めても, そのバイアスの推定の仕事が残っている. これはバイアスの漸近値 (の一致推定量 plug-in) を用いるということである. このバイアスは, 同じデータ  $X$  を, パラメータの推定と, 推定されたモデルの平均対数尤度の推定とに二度用いたことによって生じる. こうして得られる種々の分離度の推定量を, 情報量規準という.

**模型 2.1.5** (KL-変分によるパラメトリックモデルの評価). KL-変分においてモデル  $\{P_\theta\} \subset P(\mathcal{X})$  とその推定量  $\hat{\theta}_n$  を評価する事を考える. これは真の分布  $Q$  からの交差エントロピーを (ノンパラメトリック) 推定することに帰着するが, その際のバイアス

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

を評価する. ある  $\theta_* \in \Theta$  と  $\varphi_Q \in L^2(\mathcal{X}, Q; \mathbb{R}^k)$  が存在して, 次が成り立つことを仮定する:

(H1) スコア関数の中心性:  $\int_{\mathcal{X}} \varphi_Q(x) Q(dx) = 0$

(H2)

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

が成り立つと仮定する.

(H3) その他の正則性条件…….

さらに, 次の記号を用意する.

(1)  $\hat{u}_n := \sqrt{n}(\hat{\theta}_n(x_n) - \theta_*)$  とおく.

(2) スコアを  $s(x, \theta) = \partial_\theta \log p(x, \theta)$  とし,  $Q$  からの差を

$$s_0(x; \theta) := s(x; \theta) - \int_{\mathcal{X}} s(z; \theta) Q(dz)$$

とする.

(3)  $g(x, v) := -2 \int_0^1 (1-t) \partial_\theta^2 \log p(x; \theta^* + tv) dt$ ,

$$g_0(x, v) := g(x, v) - \int_{\mathcal{X}} g(z, v) Q(dz).$$

(4)  $C := \text{Cov}_Q[s(-, \theta_*)^\top, \varphi_Q]$  とする

$Q \notin \{P_\theta\}$  のとき, 一般に  $E_Q[s(x; \theta_*)] \neq 0$  に注意. このとき, Taylor 展開は

$$\log p(z; \hat{\theta}_n(\mathbf{x}_n)) - \log p(z; \theta_*) = \frac{1}{\sqrt{n}} s(z; \theta_*) \hat{\mathbf{u}}_n - \frac{1}{2n} \hat{\mathbf{u}}_n^\top \mathbf{g}(z, n^{-1/2} \hat{\mathbf{u}}_n) \hat{\mathbf{u}}_n$$

を与える.

**要諦 2.1.6.** この仮定は,  $\theta_* \in \Theta$  を唯一の最小点に持つ関数  $\theta \mapsto \int_{\mathcal{X}} \Psi(x; \theta) Q(dx)$  が定める最小コントラスト推定量が, スコア関数

$$\varphi_Q(x) := - \left( \int_{\mathcal{X}} \partial_\theta^2 \Psi(z; \theta_*) Q(dz) \right)^{-1} (\partial_\theta \Psi(x; \theta_*))^\top$$

について満たす.

**定理 2.1.7 (バイアスの漸近分布).** 仮定 (H1) (H3) の下で, 次が成り立つ:

$$nb_n = \int_{\mathcal{X}^n} Q^n(d\mathbf{x}_n) \left( 1_{A_n} \sum_{j \in [n]} \left( \log p(x_j; \hat{\theta}_n(\mathbf{x}_n)) - \int_{\mathcal{X}} \log p(z; \hat{\theta}_n(\mathbf{x}_n)) Q(dz) \right) \right) + o(1) = \Delta_1 + \Delta_2 + \Delta_3 + o(1).$$

かつ

$$\Delta_1 = o(1), \quad \Delta_2 \rightarrow \text{Tr} C, \quad \Delta_3 = o(1).$$

特に,  $nb_n \xrightarrow{n \rightarrow \infty} \text{Tr} C$ .

### 2.1.3 AIC の含蓄

$(-2) \log_e(\text{最大尤度})$  の第1項は変わらず, 第2項のペナルティ項をいろいろ変えることで, 情報量規準が作れる. ペナルティ項をモデルの次元としたものが AIC で, 交差エントロピーの一致推定量としては misspecification に脆弱であるが, 仮定が少ないものを採用する「オッカムの剃刀」の方針に則ったものとしては非常にシンプルな存在である.

**議論 2.1.8** (独立性・有意性検定は最小 AIC 推定で置き換えられる (あるいは近似的には同じ行為である)). 二次関数に正規誤差  $N(0, \sigma^2)$  を持たせて発生させたデータに, 1次, 2次, 5次で fitting すると, 実は5次の方が最小二乗誤差は小さいが, 真のデータ生成規則と異なる. これを AIC が解決することが最初の例として挙げられている (1976[14]). これは「雑音で乱された信号の復元にも AIC が使えることを表している」と論じている.

続いて, 独立性の検定に応用する. 独立性を仮定した場合と仮定しなかった場合のどちらのモデルを採用するかを AIC で判定できるといふ. Fisher の  $\chi^2$ -検定と結論が同じになると論じている. さらに  $\chi^2$ -検定に対する有意性として「大量データの統計解析の自動化の可能性があることを示している」と提示している.

**歴史 2.1.9** (AIC の誕生 (1971) [14] 参照). AIC は予測問題と密接に関係があることが, 従来の検定・推定の立場と最も異なる点であるという.

従来の推定論は, 単に未知パラメータの推定という形で展開され, 推定結果の利用目的が明確でないために, 推定精度の評価基準, 例えば2乗平均誤差などの選定上の指導理念を与えることができなかった.

- (1) 定常時系列  $(\mathbf{y}_n)$  を, 過去  $k$  段階前までの観測値の線型結合で予測することを考える. それぞれの線型結合の係数は, 最小二乗誤差で捉え, 最終予測誤差 (FPE) を

$$FPE := E[y_n - \hat{a}_{k0} - \hat{a}_{k1}y_{n-1} \cdots - \hat{a}_{kk}y_{n-k}]$$

で定めて, 最良の次数  $k \in \mathbb{N}$  を決めたい. なお, ここに  $\mathbf{z}_n := y_n - \hat{a}_{k0} - \hat{a}_{k1}y_{n-1} \cdots - \hat{a}_{kk}y_{n-k}$  が互いに独立であると仮定できるとき, これをこれを自己回帰モデルという. この次数  $k \in \mathbb{N}$  の決定は従来検定を用いることが考えられていたが, 実用的なものは知られていなかった. しかし FPE の最小化と考えると, 推定論の問題である. FPE を推定して, その最小化を考えて  $k$  を選択すれば良い, とした (Akaike 1969,70).

この最小 FPE 法と呼ばれる方法の実用上の有効性は著しく, 現在まで多くの成功的な適用例, 特に各種研究分野におけるスペクトル解析への応用例, が報告されている. この方法を  $\{\mathbf{y}_n\}$  がベクトル過程である場合に拡張したものは, セメント焼成工程の計算機制御の実現に用いられ, またそのほかの複雑な統計的システムの解析に利用されつつある.



(2) しかし実は、 $\{y_n\}$  がベクトル過程である場合への拡張には困難がある。  $X \in \mathbb{R}^q$  を共変量として、  $y \in \mathbb{R}^p$  に対する回帰  $Y = FX + Z$  を因子分析モデルという。  $F \in M_{pq}, Z$  の各成分は  $X$  の各成分と独立であるとする。これは、  $Y$  の成分間の相関関係を、低次元因子ベクトル  $X$  で説明しようとするモデルで、  $Y$  の分散共分散行列  $C$  に、  $C = FF^T + V$  ( $V$  は対角行列) なるパラメトリックな仮定を置いていることになる。  $Z$  はノイズである。従来の決定法は、  $Z$  にパラメトリックな仮定を置いて、  $q \in \mathbb{N}$  を増加させながら最尤法によって「 $q$  に伴う最大尤度の増加分の検定結果」が有意でなくなるような最初の  $q$  を因子数として決定するものであった。するとこれは、一体何を基準に考えているのかが明らかでない。FPE のように、「何を最小化しようとしているのか？」が明らかでない。

(3) 役者が揃った。因子分析モデルから最尤法が採用され、AR モデルから FPE という「統計量にまとめること」が採用されることになる。

MLE は Fisher によって初めて詳細に論じられてから極めて広く用いられているが、その有効性の根拠が何にあるのかは明らかではなく、その合理性の説明は通常直感に訴えてなされてきた。前項で論じた因子分析法には最尤法が用いられている。そこで、最尤法が何を最適化しようとしているのかが明らかになりさえすれば、先の問題は解決されることになる。

この後すぐに、漸近論的な立場から、 $M$ -推定量のような発想「最尤法は  $E[\log_e f(Y|\theta)]$  を最大にする  $\theta$  の推定量である」さらに言い換えると、「KL-情報量を規準として、真の分布を最もよく近似する  $f(y|\theta)$  を求めようとしている」と捉え直している。

最後にこの観点から、FPE の場合の平均二乗誤差の代わりに、 $\mathbb{P}_n[I(g; f(\hat{\theta}))]$  を取ることを考える。この値の推定は、第1項を無視して、 $\mathbb{P}_n[E[\log_e f(Y|\hat{\theta})]]$  の推定と等価。そして実は  $N \cdot \mathbb{P}_n[E[\log_e f(Y|\hat{\theta})]]$  の推定値は  $l(\hat{\theta}) - k$  となる。KL 情報量または FPE の場合に合致するように符号を反転させると、これが AIC である。

**要諦 2.1.10.** 赤池さんはすごく統計的決定理論の霊性を持っていることがわかる。因子分析モデルの最尤法はどのような意味で最適化を明らかにした。さらに、FPE は将来の値の予測であったが、AIC は将来の値の分布の予測、という質的な進化も経ている。この結果として、AIC はほぼ無際限の広い適用分野を見出すことが可能になった。そして赤池さん多分 Fisher 大好きだ。そしてパラメトリックな仮定というのは「当たれば最強」という点において統計学者の腕の見せ所であり、人類の希望でもあるのだろうな。我々はセミパラメトリックな Fisher になろう。

我々はここに適当なモデルの系列によってデータの外の情報あるいは知識（事前情報）を適切に表現することにより、しばしば飛躍的な予測精度の向上をもたらす高度に実用的な推定法が実現されているのを見ることができる。機械的盲目的な推定・検定の分類は、まさに逆立ちした有意性検定の虚像に基づくものといえよう。Fisher によって書かれた書物が現在に至るまで多くの応用分野の研究者によって利用されているのは、そこに与えられている各種のモデルの系列の持つ有効性の魅力によるものといえる。有意性検定を支えているものは、その一見客観的に見える論理構成ではなく、Fisher 自身実際のデータの解析を通じてその有効性を信ずるに至った、極めて経験的なその手続きの実用性である。この特性はある程度そのまま AIC に受け継がれ、それが AIC の特徴と同時にまた限界を形成している。

**例 2.1.11** (共分散構造分析 (SEM)). 赤池 [15] に AIC の応用として次の議論がある。因子分析は共分散行列  $F$  の推定の営みとみなせる。が、事前の仮定を組み込む余地のない、純粋に形式的なモデルである。そこで、共分散構造分析は、実験や観測を適切に計画し、観測データのベクトル  $Y$  に対してあらかじめ想定した構造によって定まる係数行列  $L$  を用いて、 $Y = LX + U$  という関係が成立するようにする。  $Y$  は正規分布に従い、  $X \sim (0, \Phi), U \sim (M, \Psi)$  とし、種々の制約条件（成分間の相互関係）を  $\Phi, \Psi$  の言葉で書く（ $\Phi$  は対角形とする、など）。モデルの有意性の検定は  $\chi^2$ -検定などでなされるが、代わりに AIC を使うことが考えられる。AIC を使うと、直接比較ができなかったモデルの間の比較も機械的に可能となる。□

**歴史 2.1.12** (「バイアス補正法」として拡張の歴史を辿った)。モデルの評価基準を「その予測能力としよう」として初めて提案したのは赤池 (1973,74)。-2 の係数はこのときの名残である。モデルが真の分布を含まない場合にも拡張したのが竹内啓 (76) である。小西・北川 (96) は、バイアス補正の方法を一般の統計量に拡張した。石黒 (97) はブートストラップ法によってバイアス補正を行うことを提案し、拡張情報量規準 EIC(Extended IC) と呼ばれる。

**注 2.1.13.** AIC の発明のときから取っている基本的な考え方として次の3点がある。

- (1) モデルの良さはその予測能力でみる。
- (2) 予測は分布に対して行う。

(3) 分布の近さは(二乗誤差などではなく)KL 情報量で測る。

**要諦 2.1.14.** 本懐はバイアスの補正であるとするなら、ロバスト統計と考えていることの方向性は同じではないか？

[15]

AIC 最小化の考え方は、因子分析モデルのように、構造的意味を直接持たない計測用モデルと呼ぶべきモデルの当てはめ(fitting)の実現に極めて有効である。しかし AIC 最小化の真の意味での実現は、上述の通り適切なモデルあるいは仮説の導入によってのみ可能となる。積極的に種々の仮説を展開し、それに対応する統計的モデルを構成し、必要な実験観測を計画し、その結果の解析を通じて再びモデルの発展を図っていく。こういう研究の過程によって初めて AIC 最小化が本来目指すものが実現される。

### 2.1.3.1 検定とは何か

検定とは、モデルの良さ・悪さの比較を行っているのである。情報量規準という、モデル同士の比較の形式を手に入れたために視座が高くなり、検定に対するより深い理解と認識論的な転換がある。

**議論 2.1.15** (最小 AIC 推定の哲学から見た有意性検定). Fisher の有意性検定は、例えば「接種と発病とが独立である」という仮説を、データによって反駁することを目的とするものとされてきた。しかし赤池 [14] はこう論じる。

しかしこうした見方が論理的矛盾を含むものであることは明らかである。非常に厳密に見れば、接種は必ず大なり小なりの影響を人体に与えるはずで、その結果は発病と完全に無関係ではあり得ない。あるサイコロの正しさを検定するという問題も全く同様で、現実のサイコロで完全に対称なものが存在し得ないことは明らかである。このように、仮説は常に否定される立場にあり(帰無仮説)、データによる検定結果を待つまでもなく結論は見えている。この論理的矛盾に加え、有意性検定における判定基準選定上の曖昧さは著しく、いわゆる神聖な 5 パーセント水準の信仰の発生など議論の絶え間がなく、遂に有意性検定の有用性を疑う人まで現れるようになった。

これは、認識論と Bayes 学派とを折衷する立場じゃないのか？ 帰無仮説を棄却できるというのは、「実は無効であるという制約された人工モデルの方が予測精度が良い」という我々の無知が生んだ奇異な状況を脱却できるかの試練である。

有意性検定を、「帰無仮説は AIC を最小化するか？」という問題を解いている行為とほとんど等しいとみなせることを思い出すと、有効モデル＝対立仮説は常に何らかの意味では正しいが知識不足であり(いわばノンパラメトリックモデル？)、無向モデル＝帰無仮説は人工的に制約を加えたモデルであると考えられる。実験データが十分に大きいならば、機械学習なり、 $n(i, j)/N$  を  $p(i, j)$  の推定量として使えば良い。しかし、データは限られてるので、無効モデルの方が有効な近似を与える可能性があるし、何より推定すべきパラメータは減少するので我々にも feasible である。この時、制約を加えている無効モデルは現実の構造からあまりにも隔たっているかもしれない。こうして、トレードオフの関係にある。

- (1) 有意性検定における無効モデルは、相関の非存在を主張するから、考慮すべきパラメータ数は減少する。従って、人類の計算能力からすると予測精度は上がる。
- (2) しかし、現実とは異なる仮定をおいているため、これを原因としたバイアスがあるはずである。

この2つのどちらのモデルが良いかを、「無効モデルの方が正しいと想定して比較検討」して、どちらかのより予測精度が高いモデルを採用してこれを「科学的知識」とするのが有意性検定である。

### 2.1.3.2 尤度とは何か

たぶん、セミパラメトリック化以外の精神は全て赤池先生がやってくれているような気がする。

**議論 2.1.16** (尤度の客観性・主観性の二面性). AIC の導入は、平均対数尤度  $l(\theta)/N = \mathbb{P}_n[\log f(Y|\theta)]$  が、 $E[\log_e f(Y|\theta)]$  の推定値であるという着想の下に成立した。この見方に従えば、 $\{f(y|\theta)\}_{\theta \in \Theta}$  の族が  $Y$  の真の分布を与える  $g(y)$  を含まない場合においても、最尤法が統計的モデルのパラメータ決定に有効なものでありうる事が容易にわかる。直交射影ではないが、集合  $\{f(y|\theta)\}_{\theta \in \Theta}$  の中で  $g$  に最も KL-距離の意味で近い点を選出する算譜なのだろう。つまり、赤池さんの言葉で言えば、最尤推定量  $\hat{\theta}$  はエントロ

ピーに関して  $g$  を最もよく近似する  $f(y|\theta)$  を与える  $\theta$  の、観測値  $y_1, \dots, y_N$  に基づく推定値を与える。ただし、適用条件は、平均対数尤度が  $E[\log_e f(Y|\theta)]$  の推定値として有効な限りであり、またモデル  $\{f(Y|\theta)\}$  も人間が勝手に自分の責任において取捨選択するものである。「Fisher が明らかにし得なかった尤度の本質的に実験的帰納的な性格と、統計理論における尤度利用の必然性が、ここに客観的に描き出されている」。つまり、 $B(g; f_\theta) = E[\log f_\theta(X)] - E[\log g(X)]$  の最大化は真の分布を知らなくても  $E[\log f_\theta(X)]$  さえ与えられれば実現できる、という意味ではノンパラメトリックな方法である。この特性から、 $E[\log f_\theta(X)]$  を標本平均で置き換えることで実用性を獲得したのが最尤法である。

**歴史 2.1.17** (稲垣宣生). 赤池 [14] は  $g(y)$  がある  $\theta_0$  によって  $g = f(-|\theta_0)$  と与えられる状態を想定することは一つの**主観的な行動原理**であるという。つまり、「想定したモデルが正しいという仮定の下に、エントロピーの期待値を評価する」という行動指針はどういう意味を持つのか？という疑問を呈している。そして有意性検定も、AIC も、この行動原理に貫かれている。そして、この主観性を明確に表現する損失関数の下での最小 AIC 推定量の漸近最適性を証明したのが稲垣氏の仕事である。

### 2.1.3.3 エントロピー最大化原理

AIC (対数尤度を用いたモデル評価) を用いることは、次の指針に従って統計的推測を行うことと等価である [14]. 「予測分布」を作りたい。これを、観測データの関数として表すためのモデル  $\{f_\theta\}$  を適切に定義し、エントロピーを評価基準としてその最適化を図り、モデル  $\{f_\theta\}$  を選択し (AIC), パラメータ  $\theta$  を選択する (MLE). なお、エントロピーとは KL-情報量  $-I(g; f_\theta)$  である。

**要諦 2.1.18** (Boltzman のエントロピーの確率論的解釈). エントロピーが増大するとは乱雑になるということであるが、これは確率論から翻訳することもできる。エントロピー  $B(g; f_\theta) := -I(g; f_\theta)$  について、 $f_\theta$  が定める確率分布に従う確率変数を独立に  $N$  個観測した時に得られる標本分布として、 $g$  に近いものが得られる確率の対数が、漸近的に  $N \cdot B(g; f_\theta)$  に一致する [?]. 従って、 $B$  が大きいほど、 $f_\theta$  は  $g$  をよく近似していると考えられる。

## 2.2 密度推定

密度推定には積分変換の理論が応用出来る。

### 2.2.1 核型推定量

バンド幅は平滑化のためのパラメータである。いずれにしろ可測関数  $K_h: \mathbb{R} \rightarrow \mathbb{R}$  を用いて、 $x - x_j$  の重み  $K_h(x)$  付き標本平均を  $f(x)$  の推定値とするのである。これは、 $K_h(x)$  なる波形を各観測点  $x_j$  において合成波を取ったものと見れる。実際、熱核を各  $x_j$  において熱の総量を求める操作に等しい。なお、ヒストグラムもカーネル推定の例と見れる。カーネル推定は合成波の手法で、基本波形を求めようとする逆問題と思える。このため、時系列解析でも頻出する問題である。

**模型 2.2.1** (核型密度推定モデル).  $\{X_j\}$  を確率密度関数  $f: \mathbb{R} \rightarrow [0, 1]$  の独立同分布列とし、 $f$  を推定する。

(1) 関数  $K \in L(\mathbb{R})$  について、

$$\hat{f}_n(x) := \frac{1}{nh} \sum_{j \in [n]} K\left(\frac{x - X_j}{h}\right) = \frac{1}{h} E_{\mathbb{P}_n} \left[ K\left(\frac{x - X}{h}\right) \right] = E_{\mathbb{P}_n} [K_h(x - X)] \quad K_h(x) := \frac{1}{h} K(x/h).$$

を核  $K$  幅  $h$  の核型推定量 (kernel-type estimator) またはパルツェン窓 (Parzen-Rosenblatt window) という。

(2) 核型推定量の危険関数としては、 $L^2$ -危険関数、すなわち、積分自乗誤差が用いられる：

$$\text{MISE}(\hat{f}) := \int_{\mathbb{R}} E[(\hat{f}_n(x) - f(x))^2] dx.$$

**定義 2.2.2** (super kernel). 核  $K \in L(\mathbb{R})$  と  $s \geq 2$  について、



(1) 次を満たすとき、**クラス  $s$  の核**という：

$$\int_{\mathbb{R}} |x|^s |K(x)| dx < \infty, \quad \forall i \in s \quad \int_{\mathbb{R}} x^i K(x) dx = 0$$

(2) 任意の  $s \geq 2$  についてクラス  $s$  の核であることを、**超核**であるという。

**注 2.2.3.** 積分核の中でも統計分野で**カーネル**というとき、対称分布の確率密度関数、すなわち、次の2条件を満たすものをいう：

$$(1) \int_{\mathbb{R}} K(u) du = 1.$$

$$(2) \text{偶関数} : K(-u) = K(u) \text{ a.s..}$$

なお、時系列解析では**窓関数**という。特に、ガウス関数

$$K(x) = \frac{1}{2\pi} e^{-x^2/2}$$

をカーネル関数として採用することが多い。

**定理 2.2.4.**  $f \in C^s(\mathbb{R})$  かつ  $f^{(s)} \in L^2(\mathbb{R})$  とする。クラス  $s$  の核  $K \in L^2(\mathbb{R})$ ,  $\int_{\mathbb{R}} K(x) dx = 1$  について、ある定数  $C \in \mathbb{R}$  が存在して、

$$\forall n \in \mathbb{N} \quad \forall h > 0 \quad \text{MISE}(\hat{f}) \leq C \left( \frac{1}{nh} + h^{2s} \right).$$

**要諦 2.2.5.** 誤差  $\frac{1}{nh} + h^{2s}$  を最小にする  $h$  のレートは  $n^{-1/(2s+1)}$  で、そのときの積分2乗誤差のオーダーは  $O(n^{-2s/(2s+1)})$  となる。これは必ずパラメトリックモデルの  $n^{-1}$  よりも遅くなる。

## 2.2.2 その他の手法

カーネル法は1950s、カーネル回帰法がNadaraya (1964) と Watson (1964) による。その後ハザード関数、分布関数、密度比などの推定に拡張され、漸近正規性も示されている。カーネル法は滑らかな推測結果を与える手法として評価され、パラメトリックな第二段階との結合により、半径数的な推測法に繋がる。

## 2.3 射影と条件付き期待値

統計量の漸近分布を求める方法の1つに、既に漸近分布が知れている確率変数と漸近同等であることを示す方法がある。このとき、Slutsky の補題の簡単な系から、極限法則が一致することが従う。

すると、何で近似するかという問題が生じるが、条件付き期待値は射影であり、これは  $L^2$ -近似であることを考えると、この言葉を使うのが良からう。この手法は  $U$ -統計量に対して普遍的な効力を持ち、漸近正規性を導く。

### 2.3.1 Hajek の射影

Hajek の射影は、ある確率変数  $X_1, \dots, X_n$  と、それぞれの一変数関数  $g_i$  との合成  $g_1(X_1), \dots, g_n(X_n)$  の形で  $T$  を最もよく近似する元である。

**記法 2.3.1.**

$$S := \left\{ \sum_{i=1}^n g_i(X_i) \in L^2(\Omega) \mid \forall i \in [n] \quad g_i(X_i) \in L^2(\Omega) \right\}$$

とおく。

**補題 2.3.2 (Hajek projection).**  $X_1, \dots, X_n$  は独立とする。このとき、任意の  $T \in L^2(\Omega)$  の  $S$  への射影は

$$\text{pr}_S(T) = \sum_{i=1}^n E[T|X_i] - (n-1)E[T]$$

と表せる.

**要諦 2.3.3.** 特に  $X_1, \dots, X_n$  が同分布に従うとし,  $T = T(X_1, \dots, X_n) : \Omega \rightarrow \mathcal{X}^n \rightarrow \mathbb{R}$  を対称な可測関数とする. このとき, 条件付き期待値は

$$E[T|X_i = x] = E[T(x, X_2, \dots, X_n)]$$

と簡略化されるから,

$$S' := \left\{ \sum_{i=1}^n g(X_i) \in \mathcal{L}(\Omega) \mid g \in \mathcal{L}(\mathcal{X}) \right\}$$

上への射影と一致する:  $\text{pr}_S = \text{pr}_{S'}$ .

## 2.3.2 Hoeffding の分解

一変数関数  $g_i : \mathcal{X} \rightarrow \mathbb{R}$  を用いるのではなく, より変数を多くして,  $X_1, \dots, X_n$  の間の複雑な依存関係も表現できるようにすることで, 近似の精度を上げることを考える. これは, 初めは  $(H_i)_{i \in [n]}$  の線型結合で表せる元の空間だったものを, どんどん互いに直交する部分空間を追加することで大きくしていく算譜である.

**記法 2.3.4.** 任意の  $A \in P([n])$  に対して,

$$H_A := \left\{ g_A(X_i : i \in A) \in L^2(\Omega) \mid g_A \in \mathcal{L}(\mathcal{X}^{|A|}), \forall B \in P([n]) \mid |B| < |A| \Rightarrow E[g_A(X_i : i \in A) | X_j : j \in B] = 0 \right\}$$

と定める.

**補題 2.3.5.** 各  $\{H_A\}_{A \in P(n)} \subset L^2(\Omega)$  は組ごとに直交している.

**定理 2.3.6.**  $X_1, \dots, X_n$  を独立な確率変数,  $T \in L^2(\Omega)$  とする.

- (1)  $\text{pr}_{H_A}(T) = \sum_{B \in P(A)} (-1)^{|A|-|B|} E[T | X_i : i \in B]$ .
- (2)  $\forall B \in P(A) \ T \perp H_B$  ならば,  $E[T | X_i : i \in A] = 0$ .
- (3)  $\bigoplus_{B \in P(A)} H_B$  は,  $(X_i)_{i \in A}$  の関数で 2 乗可積分であるものの全体を含む部分空間となる.

**要諦 2.3.7.** 特に  $X_1, \dots, X_n$  が同分布に従うとし,  $T = T(X_1, \dots, X_n) : \Omega \rightarrow \mathcal{X}^n \rightarrow \mathbb{R}$  を対称な可測関数とする. このとき, Hoeffding 分解は

$$T = \sum_{r=0}^n \sum_{|A|=r} g_r(X_i : i \in A), \quad g_r(x_1, \dots, x_r) = \sum_{B \in P[r]} (-1)^{r-|B|} E[T(x_i \in B, X_i \notin B)]$$

となり, 各項  $\sum_{|A|=r} g_r(X_i : i \in A)$  は退化した核を持った  $r$  次の  $U$ -統計量で, 全ての項は互いに直交だから, 分散は容易に計算できる:

$$\text{Var}[T] = \sum_{r=1}^n \binom{n}{r} E[g^2(X_1, \dots, X_r)].$$

## 2.4 $U$ -統計量

### 2.4.1 定義と例

**歴史 2.4.1.**  $U$ -統計量の理論は Hoeffding 1948 による. まず Hoeffding (61) はマルチンゲールへの分解という基礎を与えた. Hajek 射影の方法による漸近正規性の証明から, 中心極限定理を  $U$ -統計量に一般化する研究で, Berry-Esseen 限界が求められた. ここからさらに Edgeworth 展開の研究へと跳躍し, 現在では漸近  $U$ -統計量の漸近展開も得られている (Lai and Wang 93). これから, 一般のスチューデント化統計量について, 漸近  $U$ -統計量とバイアスとの和への分解が得られるから, この Edgeworth 展開が得られることになる (Maesono 97).

**模型 2.4.2** ( $U$ -統計量が構成可能な模型).  $\{X_j\}_{j \in [n]}$  を同分布の独立観測とする. 以下, 標本の部分集合  $A = \{i_1, \dots, i_r\} \subset [n]$  に対して,  $X_A := (X_{i_1}, \dots, X_{i_r})$  と略記する. 便宜的に,  $X_\emptyset = 0$  とすると,  $E[h(X_A)|X_\emptyset] = E[h(X_A)]$  に注意. 標本上の対称な可測関数  $h \in L(\mathcal{G}^n)$  は  $[\mathcal{G}]^n$  を誘導するため,  $h(X_A) : [\mathcal{G}]^r \rightarrow \mathbb{R}$  は **well-defined** である.

- (1)  $h(X_-) \in L^1([\mathcal{G}]^r)$  を仮定する.
- (2) 標本の更なる部分集合  $B \subset A \subset [n]$  に対して, 射影の和を

$$h(X_A)_B := \sum_{C \subset B} (-1)^{|B|-|C|} E[h(X_A)|X_C]$$

と定める. これを  $h(X_A)$  の **Hajek の射影** という. このとき, 反転公式

$$h(X_A) = \sum_{B \subset A} h(X_A)_B$$

が成り立つ. 「 $U$ -統計量はより退化次元の大きい  $U$ -統計量の和として表せる」という **Hoeffding の分解** の最も単純な場合である.

- (3) 次の形を持つ統計量を **核  $h$  に関する  $r$  次の  $U$ -統計量** という:

$$U_{r,n} := \binom{n}{r}^{-1} \sum_{A \in [[n]]^r} h(X_A) = E_{U([\mathcal{G}]^r)}[h(X_A)].$$

ただし,  $U([\mathcal{G}]^r)$  とは, 標本の大きさ  $r$  の部分集合  $X_A \in [\mathcal{G}]^r$  全体の集合上の一様分布である. この意味で,  $U$ -統計量は, 経験平均統計量  $E_{\mathbb{P}_n}[h(X)]$  の一般化でもある. ただし, この場合と違って, 各項は互いに独立にはならない.

- (4) 前  $k$  変数に関する条件付き期待値を

$$h_k(x_1, \dots, x_k) := E[h(x_1, \dots, x_k, X_{k+1}, \dots, X_r)] = E[h(X_{[r]})|X_1 = x_1, \dots, X_k = x_k], \quad h_0 := E[h(X_1, \dots, X_r)]$$

と表す.  $U$ -統計量  $U_{r,n}$  が **退化次元  $i-1$  ( $2 \leq i \leq r$ ) を持つ** とは, その核が

$$h_{i-1} = 0 \in L([\mathcal{G}]^{i-1}) \wedge h_i \neq 0 \in L([\mathcal{G}]^i)$$

を満たすことをいう. すなわち,  $h$  の部分空間  $L_{\mathcal{G}^{i-1}}(\mathcal{G}^n)$  への射影が零になることをいう. これが  $i=1$  としても成り立つとき, **非退化** であるという.

- (5) 推定したいパラメータを  $\theta := h_0$  と表そう. 核の射影  $h_k$  の中心化を  $\tilde{h}_k := h_k - \theta$  ( $k \in r+1$ ) とする. 中心化された核  $\tilde{h}_r : [\mathcal{G}]^r \rightarrow \mathbb{R}$  に関する  $[k] \subset [n]$  への Hajek の射影を, 再び関数  $g_k : \mathcal{G}^k \rightarrow \mathbb{R}$  で表す:

$$g_k(x_k) := \sum_{C \subset [k]} (-1)^{k-|C|} \tilde{h}_{|C|}(x_C).$$

すなわち,  $\hat{h}_r(X_A)_B = g_{|B|}(X_B)$  である.

- (6) このとき, 反転公式を繰り返すことで, 中心化された  $U$ -統計量  $U_{r,n} - \theta$  に関する一般の **Hoeffding の分解**

$$U_n - \theta = \sum_{k=1}^r \binom{n}{k}^{-1} \binom{r}{k} S_k(n), \quad S_k(n) := \sum_{B \in [[n]]^k} g_k(X_B)$$

を得る.

### 例 2.4.3.

- (1) 標本平均  $\bar{X}$  と不偏分散  $U^2$  は  $U$ -統計量である.
- (2) 2次元データの無作為標本  $(X_j, Y_j)_{j \in [n]}$  について,

$$\tau := \frac{4}{n(n-1)} \sum_{i < j \in [n]} 1_{\{(X_i - X_j)(Y_i - Y_j) > 0\}} - 1$$

で定まる  $\tau : [\mathcal{G}]^2 \rightarrow \mathbb{R}$  は 2 次の  $U$ -統計量である. これを **Kendall の順位相関係数** という. また,

$$K := |\{ \{i, j\} \in [[n]]^2 \mid (x_i > x_j \wedge y_i > y_j) \vee (x_i < x_j \wedge y_i < y_j) \}|,$$

$$L := |\{ \{i, j\} \in [[n]]^2 \mid \neg(x_i > x_j \wedge y_i > y_j) \wedge \neg(x_i < x_j \wedge y_i < y_j) \}|$$

として  $\tau = \binom{n}{2}^{-1} (K - L)$  と表せる.

(3) Wilcoxon の順位検定統計量が漸近正規であることも、 $U$ -統計量に関係する。

□

## 2.4.2 漸近正規性

$\theta := E[h(X_1, \dots, X_r)]$  の推定を考えると、当然  $h$  自身も  $\theta$  の不偏推定量であるが、 $h$  を核とする  $U$ -統計量  $U_{r,n}$  は (混ぜているだけなので) 再び不偏推定量で、より小さい漸近分散を持つ (また漸近正規である)。実際、 $U_{r,n}$  は  $h$  の射影 (条件付き期待値) となっており、射影はノルム減少的であるから分散を減らすのである。

**定理 2.4.4.** 上の模型について、可積分性  $h \in L^1([\mathcal{X}]^r)$  を仮定する： $h_0 = E[h(X_1, \dots, X_r)] < \infty$ 。

- (1)  $U_{r,n}$  は  $\theta := h_0$  の不偏推定量である。
- (2) さらに  $h \in L^2([\mathcal{X}]^r)$  ならば、 $v_1 := E[g_1(X_1)^2] > 0$  として、

$$\sqrt{n}(U_{r,n} - \theta) \Rightarrow N(0, r^2 v_1) \quad (n \rightarrow \infty)$$

**[証明].**

- (1)  $U_{r,n}$  の定義より、各項  $h(X_A)$  の平均は常に  $h_0$  になる。また、 $U$ -統計量は逆マルチンゲールになっていることから従う
- (2) 補題から、 $U_{r,n}$  は  $n^{-1/2} r S_1(n)$  に漸近同等である。 $n^{-1/2} r S_1(n)$  は、Hajek の射影  $g_k(X_B)$  の和であり、中心極限定理から  $N(0, v_1)$  に分布収束する。

■

**要諦 2.4.5.**  $U_{r,n}$  は Hoeffding の分解により、射影の組み合わせのような構造で組み上がっていき、 $n \rightarrow \infty$  の極限にて  $\theta$  を中心とした正規分布に弱収束する。このとき、リサンプリングの大きさ  $r$  は小さいほどよい。

**補題 2.4.6.**

- (1)  $h \in L^1([\mathcal{X}]^r)$  とする。任意の  $B \in [[n]]^k, b \in B$  について、 $E[g_k(X_B) | X_{[n] \setminus \{b\}}] = 0$ 。
- (2)  $h \in L^2([\mathcal{X}]^r)$  とする。 $v_k := E[g_k(X_1, \dots, X_k)^2]$  とすると、 $B, C \subset [n]$  について、 $E[g_{|B|}(X_B) g_{|C|}(X_C)] = \delta_{B,C} v_{|B|}$ 。特に、

$$\text{Var}[U_n] = \sum_{k=1}^r \binom{n}{k}^{-1} \binom{r}{k}^2 v_k.$$

また、

$$E[(U_n - \theta - n^{-1} r S_1(n))^2] = \sum_{k=2}^r \binom{n}{k}^{-1} \binom{r}{k}^2 v_k = O(n^{-2}).$$

## 2.4.3 $U$ -統計量の変種

**例 2.4.7 (K-statistic).** キュムラントの最小分散不偏推定量を Fisher の  $K$ -統計量と言い、John Tukey はこれから一般化  $K$ -統計量 (polykay) を導いたが、これは斉次多項式が定める  $U$ -統計量に他ならない。

□

**定義 2.4.8 (V-statistics).** 対称な関数  $h \in L([\mathcal{X}]^r)$  に関して、

$$V_{r,n} := \frac{1}{n^r} \sum_{(i_1, \dots, i_r) \in [n]^r} h(X_{i_1}, \dots, X_{i_r}) = E_{U([n]^r)}[h(X_A)]$$

を  $V$ -統計量という。 $U$ -統計量は標本  $[n]$  について、重複を許さない  $[[n]]^r$  上の一様分布をとったが、 $V$ -統計量では重複を許す  $[n]^r$  上の一様分布を取る。 $V$ -統計量は Hoeffding の一年前である 1947 年に Richard von Mises が導入した。<sup>†1</sup>  $v_1 > 0$  のとき、 $U$ -統計量と漸近同等である： $\sqrt{n}(V_n - U_n) \xrightarrow{P} 0$ 。

<sup>†1</sup> von Mises, R. (1947). "On the asymptotic distribution of differentiable statistical functions". *Annals of Mathematical Statistics*. 18 (2): 309 – 348. doi:10.1214/aoms/1177730385. JSTOR 2235734.

**例 2.4.9** (標本分散と不偏分散).  $h(x, y) = \frac{(x - y)^2}{2}$  が定める 2 次の  $V$ -統計量は

$$V_{2,n} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

で, これは分散の最尤推定量である. しかし同じ核に対する  $U$ -統計量は

$$U_{2,n} = s = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

となる. 漸近同等ではあるが, 不偏推定量となるのは後者である!

□

#### 2.4.4 2 標本 $U$ -統計量

#### 2.4.5 退化 $U$ -統計量

**定義 2.4.10 (degenerated).**  $U$ -統計量の列  $(U_n)$  (の核  $(h_n)$ ) が退化しているとは, 漸近分散が  $r^2 v_1 = 0$  であることをいう.

## 第3章

# 半径数的手法の理論

情報量規準では  $P(\mathcal{X})$  へ視点を広げ、パラメトリックモデルの評価をした。密度推定では、分布に仮定を置かなかった。このような視点は、漸近理論によって展開される。

### 3.1 尤度比確率場の局所漸近構造

$M$ -推定量の漸近正規性の議論では、尤度の概念は全く登場しなかった。これは、最尤推定量に対する全く別のアプローチと言って良い。しかし、一様大数の法則の議論のように、尤度の構造を思い出してみよう。このように、尤度比確率場  $\mathbb{R}^p \ni \mathbf{u} \mapsto Z_n(\mathbf{u}; \theta; \mathbf{x}_n)$  の挙動を記述して仕舞えば、尤度解析で用いられる多くの統計量の漸近分布を、尤度比確率場の理論から自動的に、そして元のモデルの尤度以外の情報に依らず、統一的に導くことが可能である。

#### 3.1.1 接触性

2 元モデル  $\{P, Q\}$  について、互いに特異ならば、特定の範囲でデータが観測されるかを判定すれば良いので、統計的問題としては容易である (エントロピーが発散する)。従って、 $P, Q$  は漸近的に絶対連続な時に推測論としての興味がある。これが、統計的実験の列の漸近的な局所構造の研究として、最初にとるべき対象である。

**模型 3.1.1** (2 元漸近模型). 統計的実験列  $(\mathcal{X}^n, \mathcal{A}^n, \{P_n, Q_n\})_{n \in \mathbb{N}}$  において、ある  $\sigma$ -有限測度  $\nu \in P(\mathcal{X}^n)$  について  $P_n, Q_n \ll \nu_n$  を満たすとする。例えば  $\nu_n := P_n + Q_n$  と取れば良い。このとき、尤度比を

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

とする。

**定義 3.1.2 (contiguity)**. 統計的実験列  $(\mathcal{X}^n, \mathcal{A}^n, \{P_n, Q_n\})_{n \in \mathbb{N}}$  において、 $(Q_n)$  が  $(P_n)$  に接触している  $(Q_n) \ll (P_n)$  とは、任意の可測集合の列  $(A_n \in \mathcal{A}^n)_{n \in \mathbb{N}}$  について、 $P_n[A_n] \rightarrow 0 \Rightarrow Q_n[A_n] \rightarrow 0$  をいう。

**補題 3.1.3**. 次は同値：

- (1)  $Q_n$  は  $P_n$  に接触している： $(Q_n) \ll (P_n)$ .
- (2)  $\forall \{X_n\} \subset L(\mathcal{X})$   $X_n \xrightarrow{P_n} 0 \Rightarrow X_n \xrightarrow{Q_n} 0$ .

**例 3.1.4.**

- (1)  $P_n = U(0, 1), Q_n = U(\theta_n, 1 + \theta_n)$  とする。 $(P_n) \sim (Q_n)$  は  $\theta_n \rightarrow 0$  に同値。
- (2)  $P_n = N(0, 1), Q_n = (\theta_n, 1)$  とする。 $(P_n) \sim (Q_n)$  は  $(\theta_n)$  が有界列であることに同値。

**定理 3.1.5 (接触性の十分条件).** 次が成り立つならば,  $Q_n$  は  $P_n$  に接触している:  $(Q_n) \ll (P_n)$ .

$$\exists_{(\Omega, \mathcal{F}, P) \in \text{Prob}} \exists_{\Lambda \in L(\Omega)} P_n^{\Lambda_n} \Rightarrow P^\Lambda \wedge E[e^\Lambda] = 1.$$

### 3.1.2 尤度比の挙動

**定理 3.1.6.**  $T_n \in L(\mathcal{G}^n)$  を確率変数列であって, ある  $E[\Lambda] = -\frac{\text{Var}[\Lambda]}{2}$  を満たす正規確率変数  $\Lambda \in L(\Omega_\Lambda)$ ,  $T \in L(\Omega_T)$  について  $P_n^{(T_n, \Lambda_n)} \Rightarrow P^{(T, \Lambda)}$  が成り立つとする. ここで,  $P$  は  $\Omega_\Lambda \times \Omega_T$  上のある 2 変量正規分布とした. このとき,

$$Q_n^{(T_n, \Lambda_n)} \Rightarrow P^{(T + \text{Cov}[T, \Lambda], \Lambda + \text{Var}[\Lambda])}.$$

### 3.1.3 局所漸近正規なモデル

Lebesgue 測度について絶対連続とは限らないモデルにおいて, 真値  $P_{\theta_0}$  を中心とした Lebesgue 分解を考えることがあり得る.

**模型 3.1.7 (局所漸近正規なモデル).** 統計的実験  $(\mathcal{G}, \mathcal{A}, (P_\theta)_{\theta \in \Theta})$  の, 直積が定める列  $(\mathcal{G}^n, \mathcal{A}^n, (P_\theta^{\otimes n})_{\theta \in \Theta})_{n \in \mathbb{N}}$  を考える.  $\Theta \subset \mathbb{R}^p$ ,  $\theta_0 \in \Theta$  とする.

(1) 真値  $\theta_0$  を中心とした他の点  $P_\theta \in \mathcal{P}$  の Lebesgue 分解を

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

で表す. すなわち,  $p_{\theta/\theta_0} : \mathcal{G} \rightarrow [0, 1]$  は Radon-Nikodym 微分,  $\sigma_{\theta/\theta_0} : \mathcal{A} \rightarrow \mathbb{R}$  は  $P_{\theta_0}$ -特異測度である.

(2) これについて, 関数  $D_u(x) : \mathcal{G} \times \mathbb{R}^p \rightarrow \mathbb{R}$  を

$$D_u(x) := \sqrt{\frac{dP_{\theta_0+u}}{dP_{\theta_0}}(x)} - 1 = p_{(\theta_0+u)/\theta_0}^{1/2}(x) - 1 = p_{(\theta_0+u)/\theta_0}^{1/2} - p_{\theta_0/\theta_0}^{1/2}(x)$$

と表す.  $(p_{(\theta_0+u)/\theta_0}^{1/2})_{u \in [0,1]}$  を  $P_{\theta_0}$  を通る  $u$ -半道,  $D_u(x)$  を  $P_{\theta_0}$  を通る  $u$ -半変位と呼ぼう.

$\theta_0$  の近傍の性質として, 次を仮定する.

(G1) 変位ベクトル  $u \in \mathbb{R}^p$  の絶対値が十分小さいとき,  $D_u \in L^2(\mathcal{G}, P_{\theta_0})$  であり,  $u = 0$  において  $L^2(P_{\theta_0})$  の意味で微分可能:

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

なお,  $(P_\theta)$  はある  $\sigma$ -有限測度に支配されていて密度  $(p_\theta)$  をもち,  $\theta_0$  にて微分可能で追加の正則性条件を満たせば, 通常の意味での微分

$$\varphi(x) = \frac{1}{2p_{\theta_0}(x)} \frac{\partial p_{\theta_0}}{\partial \theta}(x)$$

に等しい.

(G2) 特異部分の最大値は 2 次のオーダーで減少する:  $\sigma_{\theta_0+u/\theta_0}(\mathcal{G}) = o(|u|^2) \quad (|u| \rightarrow 0)$ .

このとき, おそらく  $\theta_0$  において接触構造が生じている. この仮定は例えば  $(N(\theta, 1))$  や両側指数分布族  $P_\theta(dx) = \frac{e^{-|x-\theta|}}{2} dx$  について成り立つ.  $2\varphi$  は一般化スコア関数で,

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

は一般化 Fisher 情報行列に当たる. さらに, 次で定まる  $Z_n : U \times \Theta \times \mathcal{G}^n \rightarrow \mathbb{R}$  は一般化尤度である:

$$Z_n(u; \theta_0; \mathbf{x}_n) := \prod_{j=1}^n \frac{dP_{\theta_0+n^{-1/2}u}}{dP_{\theta_0}}(x_j) = \prod_{j=1}^n p_{(\theta_0+n^{-1/2}u)/\theta_0}(x_j).$$

$U$  は  $0 \in \mathbb{R}^p$  の開近傍とした. というのも,  $n$  が十分大きい時に,  $n^{-1/2}u$  は十分小さくなるため,  $P_{\theta_0+n^{-1/2}u} \in \mathcal{P}$  となり, well-defined である. 対応  $\mathbb{R}^p \ni u \mapsto Z_n(u; \theta; \mathbf{x}_n)$  を尤度比確率場という.

**補題 3.1.8.** 上の模型において,

- (1) スコア関数は中心化されている:  $E_{\theta_0}[\varphi] = 0$ .
- (2) 変位の平均の一次項は一般化 Fisher 情報行列である:  $E_{\theta_0}[D_u^2] = \frac{1}{4}u^\top I(\theta_0)u + o(|u|^2) \ (u \rightarrow 0)$ .
- (3) 半変位の自乗の平均も同様:  $E_{\theta_0}[D_u] = -\frac{1}{8}u^\top I(\theta_0)u + o(|u|^2) \ (u \rightarrow 0)$ .
- (4)

**定理 3.1.9 (模型は  $\theta_0$  において局所漸近正規である).** 上の模型は  $\theta_0$  において局所漸近正規である. すなわち, 次が成り立つ:

$$Z_n(u; \theta_0; \mathbf{x}_n) = \exp \left( u^\top \Delta_n(\theta_0) - \frac{1}{2}u^\top I(\theta_0)u + \rho_n(u, \theta_0) \right)$$

ただし, 次を満たす:

- (1)  $\Delta_n(\theta_n) \in L(\mathcal{X}^n; \mathbb{R}^p)$  が押し出す  $P_{\theta_0}^n$  は正規分布に弱収束する:  $(P_{\theta_0}^{\otimes n})^{\Delta_n(\theta_0)} \Rightarrow N_p(0, I(\theta_0))$ .
- (2)  $\rho_n(u, \theta_0) \in L(\mathcal{X}^n)$  は 0 に漸近的に確率収束する:  $\rho_n(u, \theta_0) \xrightarrow{P_{\theta_0}^{\otimes n}} 0$ .

**系 3.1.10.** 局所局所漸近ならば,  $P_{\theta_0+n^{-1/2}u}^{\otimes n}$  は  $P_{\theta_0}^{\otimes n}$  に接触している:  $P_{\theta_0+n^{-1/2}u}^{\otimes n} \ll P_{\theta_0}^{\otimes n}$ .

**[証明].**  $E[Z_n] \rightarrow 1$  より,  $\Lambda$  を何らかの意味での  $\log Z_n$  の極限とすれば, 接触性の十分条件 3.1.5 から. ■

### 3.1.4 検定への応用

尤度比検定は一様最強力であった. スコア  $\Delta_n(\theta_0)$  に基づく検定  $\Delta_n(\theta_n) \geq d_\alpha \ (u > 0 \text{ のとき})$  も, (片側対立仮説については) 漸近的に一様最強力である.

**例 3.1.11.**  $\Theta \subset \mathbb{R}$  上での検定問題

$$H_0: \theta = \theta_0 \quad \text{vs.} \quad H_1: \theta = \theta_0 + n^{-1/2}u, \quad \theta_0 \in \Theta, u \neq 0$$

を考える.  $n \rightarrow \infty$  とするとだんだん難しい検定問題になっていく.  $(0, 1)$  の列  $(\alpha_n)$  は  $\alpha \in (0, 1)$  に収束するとする. 任意の  $n \in \mathbb{N}$  について, Neyman-Pearson のサイズ  $\alpha_n$  の最強力検定  $\varphi_n$  は, 尤度比を用いた  $\log Z_n(u; \theta_0) \geq c_{n, \alpha_n}$  という形のものである. この時, 検定  $\varphi_n$  の検出力は

$$\beta_{\varphi_n}(\theta_0 + n^{-1/2}u) = E_{\theta_0+n^{-1/2}u}[\varphi_n] \leq P_{\theta_0+n^{-1/2}u}[\log Z_n(u; \theta_0) \geq c_{n, \alpha_n}]$$

を満たす. さて, 確率変数  $\Delta \min N(0, I(\theta_0))$  とこれについて

$$P \left[ \Delta u - \frac{1}{2}I(\theta_0)u^2 \geq c_\alpha \right] = \alpha$$

を満たす定数  $c_\alpha \in \mathbb{R}$  について, モデルの局所漸近正規性 3.1.9 と対立仮説  $H_1$  における尤度比の挙動の結果 3.1.6 より,

$$\limsup_{n \rightarrow \infty} \beta_{\varphi_n}(\theta_0 + n^{-1/2}u) \leq \lim_{n \rightarrow \infty} P_{\theta_0+n^{-1/2}u}[\log Z_n(u; \theta_0) \geq c_{n, \alpha_n}] = P \left[ \Delta u + \frac{1}{2}I(\theta_0)u^2 \geq c_\alpha \right]$$

が成り立つ. いずれの距離も, ある可測関数すなわち, スコア  $\Delta_n(\theta_0)$  に基づく検定  $\Delta_n(\theta_n) \geq d_\alpha \ (u > 0 \text{ のとき})$  も最大検出力を漸的に達成し得る. この検定統計量は  $u$  に依存しないから, 片側対立仮説については漸的に一様最強力である. □

### 3.1.5 推定への応用

「手法の比較が統計学的に意味のあるのは, それらが何らかの妥当と思われる平均量 (期待値や確率) によって評価される場合に限られる. 従って, 統計推測の最適性の問題が尤度解析に帰着されるのは当然のことである。」

**定理 3.1.12 (Hajek の不等式).** 統計的実験  $(\mathcal{X}_n, \mathcal{A}_n, \mathcal{P}_n)$  は  $\theta_0 \in \Theta$  で局所漸近正規であるとする. 任意の推定量  $T_n$  に対して, 次が成り立つ:

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



**要諦 3.1.13** (漸近決定理論). これは、局所漸近正規なモデルにおける推定量の、漸近的にミニマックスなリスク限界を与える。そこで、この等号を達成する推定量を探すことで、ミニマックスの意味で漸近有効な推定量が探せる。小標本理論では有効性として最小分散性を考えたが、漸近挙動については最大リスクの最小化が危険関数としては標準的である。実は、最尤推定量とベイズ推定量は漸近有効である。

### 3.1.6 Ibragimov-Has'minskii 理論

$P_{\theta_0}$  を中心に見た尤度関数 (= 尤度比)

$$\mathbb{R}^p \ni u \mapsto Z_n(u; \theta; \mathbf{x}_n) = \prod_{j=1}^n \frac{dP_{\theta_0+n^{-1/2}u}}{dP_{\theta_0}}(x_j) = \prod_{j=1}^n p_{(\theta_0+n^{-1/2}u)/\theta_0}(x_j) \in C(\Theta)$$

は、母数空間  $\mathbb{R}^p$  上の確率場 ( $C(\Theta)$ -値確率変数) をなす。  $\theta_0$  からの変位の空間を  $U \subset \mathbb{R}^p$  とすると、  $\sup_{u \in U} C(\mathbb{R}^p; \mathbb{R}) \rightarrow \mathbb{R}$  は連続な汎関数であり、連続写像定理を通じて Ibragimov-Has'minskii 理論が流入する：

$$\sup_{u \in U} Z_n(u; \theta_0) \Rightarrow \sup_{u \in U} Z(u; \theta_0) \quad (n \rightarrow \infty).$$

**定理 3.1.14** . 確率変数  $Z_n : \mathcal{G} \rightarrow C(\mathbb{R}^p); \mathbf{x}_n \mapsto Z_n(-; \theta_0; \mathbf{x}_n)$  は、次の  $Z : \mathbb{R}^p \rightarrow \mathbb{R}$  に弱収束する：

$$Z(u; \theta_0) := \exp \left( u^\top \Delta(\theta_0) - \frac{1}{2} u^\top I(\theta_0) u \right).$$

**要諦 3.1.15** (局所漸近正規性再考). 各点  $u \in \mathbb{R}^p$  での  $Z_n(u; \theta_0; \mathbf{x}_n)$  の法則収束を、局所漸近正規性というのであった。

**要諦 3.1.16** (最尤推定量再考). 収束

$$\sup_{u \in U} Z_n(u; \theta_0) \Rightarrow \sup_{u \in U} Z(u; \theta_0) \quad (n \rightarrow \infty).$$

は、

$$\arg \max_{u \in \mathbb{R}^p} Z_n(u; \theta_0) \Rightarrow \arg \max_{u \in \mathbb{R}^p} Z(u; \theta_0) \quad (n \rightarrow \infty).$$

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \Rightarrow I(\theta_0)^{-1} \Delta(\theta_0) \quad (n \rightarrow \infty)$$

を含意する。同様の方法により、Bayes 推定量の漸近正規性も得られる。

## 3.2 接触構造

$P_n$  が帰無仮説、 $Q_n$  が対立仮説の法則という意味論がある。接触しているとは漸近的絶対連続であることをいい、統計量の極限において、測度変換を与える基礎になる。

### 3.2.1 尤度比と測度変換

測度変換の公式が成り立つためには、絶対連続性が必要である。

2つの確率変数があったときに、一方のもう一方に対する絶対連続部分の Radon-Nikodym 微分を尤度比確率変数という。これは零集合上の差を除いて一意に定まる。「尤度比」という名前は実用上の重要性が先行して考察されている対象かと思いがちだが、実際は本質的に数学的に自然に重要な対象である。

絶対連続性  $Q \ll P \Leftrightarrow P(A) = 0 \Rightarrow Q(A) = 0$  は、 $P$ -a.e. 性質はそのまま  $Q$ -a.e. に成り立つことを意味し、 $\text{supp } q \subset \text{supp } p$  だから  $P$  のグラフが上から覆いかぶさっている形になる。  $q/p \, dP$  なる形式は  $Q^a$  を表す。実はこの  $\Omega$  全域での積分が 1 であるとき、 $Q^\perp = 0$  である。実際、

$$\int f dQ \geq \int f \frac{dQ}{dP} dP \quad (f \geq 0)$$

が一般に成り立つが、等号が成り立って測度変換公式として使えるための必要十分条件が  $Q \ll P$  である。  
Lebesgue 測度は参照測度として優秀で、人類はみなこの測度に引き戻して計算を行う。

**動機 3.2.1** (尤度比とは、数学的には絶対連続な確率測度組の Radon-Nikodym 微分のことに他ならない)。

- (1) (参照測度  $\nu$  の下の) 任意の確率測度  $P, Q$  について、絶対連続部分と直交／特異部分とへの Lebesgue 分解  $Q = Q^a + Q^\perp$  が存在して、 $Q^a \ll P$  かつ  $Q^\perp \perp P$  かつ

$$\forall A \in \mathcal{F} \quad Q^a(A) = \int_A \frac{q}{p} dP, \quad \frac{dQ}{dP} = \frac{q}{p} \text{ } P\text{-a.e.}$$

が成り立つ。

- (2) こうして、Radon-Nikodym 微分という確率変数  $\frac{dQ}{dP} : \Omega \rightarrow \mathbb{R}_+$  を確率  $P$  の下で考えるというテーマが定まる。これを**尤度比**という。

- (3) 尤度比の平均値が 1 であることと、 $Q^\perp = 0$  であることと、絶対連続  $Q \ll P$  であることは同値。

**補題 3.2.2.**  $P, Q \ll \mu$  を確率測度とし、対応する密度を  $p, q$  とする。

- (1) 絶対連続部分  $Q^a$  の参照測度  $\mu$  が介在しない表示  $Q^a(A) = \int_A \frac{q}{p} dP$  が成り立つ。  
(2) 尤度比の  $P$ -平均値が 1 であること  $\int \frac{q}{p} dP = 1$  と、 $Q^\perp = 0 \Leftrightarrow Q[p=0] = 0$  であることと、絶対連続  $Q \ll P$  であることは同値。  
(3) 一般に  $\int f dQ \geq \int f \frac{dQ}{dP} dP$  が成り立ち、等号が成立するのは  $Q \ll P$  のとき。

### 3.2.2 接触性

$X$  の  $Q$  についての法則を、 $P$  の言葉で考えたいとき、 $q/p$  を積分核とすれば良いことがわかった。この測度変換の問題の漸近版を考える。いつ、極限実験  $P$  を用いて、対象の実験列  $(Q_n)$  の極限分布が計算可能になるか？

**記法 3.2.3.** 台となる測度空間  $(\Omega^n, \mathcal{A}^n)$  を共通とする統計的実験の列  $(P_n), (Q_n)$  を考える。そしてそこでの統計量の列  $X_n : \Omega^n \rightarrow \mathbb{R}^k$  との振る舞いを考察する。

**定義 3.2.4 (contiguous).** モデル  $(Q_n)$  は  $(P_n)$  下に**接触している**  $Q_n \triangleleft P_n$  とは、任意の可測集合の列  $(A_n), A_n \in \mathcal{A}^n$  に対して、 $P_n(A_n) \rightarrow 0 \Rightarrow Q_n(A_n) \rightarrow 0$  が成り立つことをいう。

**注 3.2.5.**  $\frac{dQ_n}{dP_n}, \frac{dP_n}{dQ_n}$  はそれぞれ  $P_n, Q_n$  の下で一様に緊密であるから、Prohorov の定理から、弱コンパクトである。すなわち、任意の部分列は、弱収束する部分列をもつ。以降、この意味で  $U$  に収束するとき、 $\frac{dP_n}{dQ_n} \xrightarrow{Q_n}^* U$  と表す。

**補題 3.2.6 (接触性の特徴付け (LeCam 1)).** 次の 4 条件は同値。

- (1)  $Q_n \triangleleft P_n$ .  
(2)  $\frac{dP_n}{dQ_n} \xrightarrow{Q_n}^* U$  ならば、 $P[U > 0] = 1$  である。  
(3)  $\frac{dQ_n}{dP_n} \xrightarrow{P_n}^* V$  ならば、 $E[V] = 1$  である。  
(4) 任意の統計量  $T_n : \Omega_n \rightarrow \mathbb{R}^k$  について、 $T_n \xrightarrow{P_n} 0$  ならば  $T_n \xrightarrow{Q_n} 0$  である。

**要諦 3.2.7.** 漸近的でない場合の消息でいうと、

$$Q_n \left( \frac{dP_n}{dQ_n} = 0 \right) = 0 \quad \Leftrightarrow \quad E_P \left[ \frac{dQ}{dP} \right] = 1$$

は  $Q_n \ll P_n$  に同値だが、 $Q_n \triangleleft P_n$  と同値になるためには、上のような表現となる。

**例 3.2.8 (漸近的对数正規性).**

$$\frac{dP_n}{dQ_n} \xrightarrow{Q_n} e^{N(\mu, \sigma^2)}$$

が成り立つとき、 $Q_n \triangleleft P_n$  で、さらに互いに接触していることは  $\mu = -\frac{1}{2}\sigma^2$  に同値。 □

**定理 3.2.9 (漸近的測度変換 (LeCam 3)).**  $X_n : \Omega^n \rightarrow \mathbb{R}^k$  を確率変数列とする。

- (1)  $Q_n \triangleleft P_n$ ,
- (2)  $\left(X_n, \frac{dQ_n}{dP_n}\right) \xrightarrow{P_n} (X, V)$

ならば、 $L(B) := E[V, B] = E[1_B V]$  は確率測度を定め、これについて  $X_n \xrightarrow{Q_n} L$ 。

**補題 3.2.10 (接触性の十分条件).**  $P_n, Q_n$  はある  $\sigma$ -有限測度  $\nu_n$  に対して絶対連続であるとし、Radon-Nikodym 微分を  $p_n, q_n$  とおく。

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

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

### 3.3 局所漸近正規性

パラメータのスケーリングの違いを除いて、局所漸近正規なモデルはあるガウスモデルに収束する。

**議論 3.3.1.** ある点  $\theta_0 \in \mathbb{R}^k$  に注目して、その周りの局所パラメータを  $h := \sqrt{n}(\theta - \theta_0)$  とすると、元の統計的実験  $(\mathcal{X}^n, \mathcal{A}^n, (P_\theta^n)_{\theta \in \mathbb{R}^k})$  は、 $(\mathcal{X}^n, \mathcal{A}^n, (P_{\theta_0 + h/\sqrt{n}}^n)_{h \in \mathbb{R}^k})$  と変換される。するとこの統計的実験は、元の統計的実験  $\theta \mapsto P_\theta$  が滑らかであるとき、 $(\mathcal{X}^n, \mathcal{A}^n, (N(h, I_{\theta_0}^{-1}))_{h \in \mathbb{R}^k})$  に漸近する！

#### 3.3.1 尤度の拡張

**議論 3.3.2.**  $P_\theta \ll \mu$  とし、 $p_\theta := \frac{dP_\theta}{d\mu}$  と表す。ここでは  $\theta, h \in \mathbb{R}$  のような記法をするが、 $k$  次元としても  $\dot{l}(x)$  が  $k$  ベクトルになり、掛け算が内積または二次形式になるのみである。 $l_\theta(x) := \log p_\theta(x) \in C^2(\mathcal{X})$  と仮定すると（この仮定は強すぎるが）、Taylor の定理より、

$$\log P_{\theta+h}(x) = l_\theta(x) + \dot{l}_\theta(x)h + \frac{1}{2}\ddot{l}_\theta(x)h^2 + o_x(h^2). \quad h \in \mathbb{R}^k, h \rightarrow 0$$

$$\log \frac{p_{\theta+h}}{p_\theta}(x) = h\dot{l}_\theta(x) + \frac{1}{2}h^2\ddot{l}_\theta(x) + o_x(h^2). \quad h \in \mathbb{R}^k, h \rightarrow 0$$

ここで尤度比が現れる。特に  $x$  に観測値  $X_1, \dots, X_n$  を代入し、 $h$  を  $\frac{h}{\sqrt{n}}$  として  $n \rightarrow \infty$  の極限を考えると、

$$\log \prod_{i=1}^n \frac{p_{\theta+h/\sqrt{n}}}{p_\theta}(X_i) = \frac{h}{\sqrt{n}} \sum_{i=1}^n \dot{l}_\theta(X_i) + \frac{1}{2} \frac{h^2}{n} \sum_{i=1}^n \ddot{l}_\theta(X_i) + \sum_{i=1}^n o_x\left(\frac{h^2}{n}\right) \quad (n \rightarrow \infty)$$

ここで、 $E_\theta[\dot{l}_\theta] = 0, -E_\theta[\ddot{l}_\theta] = E_\theta[\dot{l}_\theta^2] = I_\theta$  だから、

$$(1) \Delta_{n,\theta} := \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{l}_\theta(X_i) \in \mathbb{R} \text{ とおくと } \Delta_{n,\theta} \xrightarrow{d} N(0, I_\theta) \text{ (中心極限定理).}$$

$$(2) \text{ 大数の法則より } \frac{1}{n} \sum_{i=1}^n \ddot{l}_\theta(X_i) \xrightarrow{d} -I_\theta.$$

$$(3) \text{ 3項目はおそらく確率収束 (仮定をうまく設定すれば良い).}$$

よって、

$$\log \prod_{i=1}^n \frac{p_{\theta+h/\sqrt{n}}}{p_\theta}(X_i) = h\Delta_{n,\theta} - \frac{1}{2}I_\theta h^2 + o_{p_\theta}(1)$$

が成り立つ。実は、必要な過程は、 $L^2$ -微分の言葉で簡潔に書ける。

**定理 3.3.3 (局所漸近正規性の十分条件).** モデル  $(p_\theta)_{\theta \in \Theta \subset \mathbb{R}^k}$  は、密度関数の平方根  $\theta \mapsto \sqrt{p_\theta}$  が  $L^2$ -微分可能であるとする：

$$\exists l_\theta \in \mathcal{L}^2(\mathcal{X}) \quad \int \left( \sqrt{p_{\theta+h}} - \sqrt{p_\theta} - \frac{1}{2} h^\top l_\theta \sqrt{p_\theta} \right)^2 d\mu = o(\|h\|^2) \quad (h \rightarrow 0).$$

なお、通常の意味で微分可能であった場合は、

$$\frac{1}{2} l_\theta \sqrt{p_\theta} = \frac{\partial \sqrt{p_\theta}(x)}{\partial \theta} \Leftrightarrow l_\theta = \frac{\partial}{\partial \theta} \log p_\theta(x)$$

が必要であることに注意。このとき、

- (1)  $E_\theta[l] = 0 \in \mathbb{R}^k$  かつ  $I_\theta := E_\theta[l_\theta l_\theta^\top] \in M_k(\mathbb{R})$  が定まる。
- (2) 任意の収束列  $\{h_n\} \subset \Theta, h_n \rightarrow h$  について、 $\Delta_{n,\theta} := \frac{1}{\sqrt{n}} \sum_{i=1}^n l_\theta(X_i) \in \mathbb{R}^k$  とすると、

$$\log \prod_{i=1}^n \frac{p_{\theta+h_n/\sqrt{n}}(X_i)}{p_\theta} = h^\top \Delta_{n,\theta} - h^\top \frac{1}{2} I_\theta h + o_{p_\theta}(1)$$

**定義 3.3.4 (LAN: local asymptotic normal).** モデル  $(P_{n,\theta} | \theta \in \Theta)_{n \in \mathbb{N}}$  が  $\theta_0 \in \Theta$  において局所漸近正規であるとは、

- (1) 行列  $r_n, I_{\theta_0} \in M_k(\mathbb{R})$ ,
- (2)  $P_{\theta_0}$  の下で  $N(0, I_{\theta_0})$  に分布収束する確率変数  $\Delta_{n,\theta} : \Omega \rightarrow \mathbb{R}^k$ ,

が存在して、任意の収束列  $\{h_n\} \subset \mathbb{R}^k, h_n \rightarrow h$  について、

$$\log \frac{dP_{n,\theta+r_n^{-1}h_n}}{dP_{n,\theta}} = h^\top \Delta_{n,\theta} - \frac{1}{2} h^\top I_\theta h + o_{P_{n,\theta}}(1)$$

と展開できることをいう。

**例 3.3.5.**  $L^2$ -微分可能なモデル  $(P_\theta)_{\theta \in \Theta}$  が定める実験列  $(P_\theta^n)$  は、 $r_n = \sqrt{n}I$  を行列として局所漸近正規である。 □

### 3.3.2 正規実験への収束

上の「局所漸近正規性」の意味を、統計的実験列の収束概念以前の状態で考えたい。

**議論 3.3.6.** 統計的実験が収束するとき、元の実験で弱収束する統計量は、極限実験での弱収束先に法則同等である。

**定義 3.3.7.** ランダム化された統計量  $T = T(X, U)$  とは、 $U \sim U([0, 1])$  にも依存する統計量をいう。

**定理 3.3.8 (正規実験へ収束するための十分条件).** モデル  $(p_\theta)_{\theta \in \Theta \subset \mathbb{R}^k}$  は、密度関数の平方根  $\theta \mapsto \sqrt{p_\theta}$  が  $L^2$ -微分可能であり、Fisher 情報行列  $I_\theta := E_\theta[l_\theta l_\theta^\top] \in \text{GL}_k(\mathbb{R})$  が可逆であるとする。実験  $(P_{\theta+h/\sqrt{n}})_{h \in \mathbb{R}^k}$  における統計量  $T_n : \mathcal{X}^n \rightarrow \mathbb{R}^k$  が、任意の  $P_{\theta+h/\sqrt{n}} (h \in \mathbb{R}^k)$  について弱収束するとき、実験  $(N(h, I_\theta^{-1}))_{h \in \mathbb{R}^k}$  上のランダム化された統計量  $T := T(X, U) : \mathcal{X} \times [0, 1] \rightarrow \mathbb{R}^k$  が存在して、任意の  $h$  について同じ収束先へ分布収束する。

## 3.4 漸近決定理論

$h_0 \in H$  上の尤度比過程は、周りの  $h$  と  $h_0$  との「分離度」を表す。実際対数尤度比の  $P_h$  についての平均が  $h, h_0$  の間の KL-分離度である。そこで統計的実験が収束するとは、任意の尤度比過程がバージョンの違いを除いて収束することとすると、統計モデルがある種の "Riemann 多様体" として、各点において局所構造が等しいことと思える。極限的実験を定めると、ここでの効率限界が、元の実験での下界を与える。

### 3.4.1 実験列の収束の定義

**定義 3.4.1 (likelihood ratio process, limit experiment).** 統計的実験  $(\mathcal{X}, \mathcal{A}, (P_h)_{h \in H})$  について、

- (1)  $h_0 \in H$  上の**尤度比過程**とは,  $\left(\frac{dP_h}{dP_{h_0}}(X)\right)_{h \in H}$ , または参照測度が存在するとき,  $\left(\frac{P_h}{P_{h_0}}(X)\right)_{h \in H}$  をいう. これは, 観測という確率変数  $X$  に, モデルの密度関数族が定める積写像  $(p_h/p_{h_0})_{h \in H}$  を合成して得たものである.
- (2) 実験列  $((\mathcal{G}_n, \mathcal{A}_n, P_{n,h}; h \in H))_{n \in \mathbb{N}}$  が**極限実験**  $(\mathcal{G}, \mathcal{A}, P_h; h \in H)$  に収束するとは, それらの任意の  $h_0 \in H$  上の尤度比過程の任意の有限次元周辺分布が,  $h_0$  が真のパラメータであるという過程の下で法則収束することをいう:

$$\forall_{h_0 \in H} \forall_{I \subset H} |I| < \infty \Rightarrow \left(\frac{dP_{n,h}}{dP_{n,h_0}}(X_n)\right)_{h \in I} \xrightarrow{h_0} \left(\frac{dP_h}{dP_{h_0}}(X)\right)_{h \in I}.$$

### 3.4.2 漸近表現定理

法則収束する統計量の漸近的な振る舞いは, 極限実験における統計量の考察に帰着する. 実験内の最適統計量の考察と, 実験の収束とは, 完全に独立に議論できるのである!

実は, さらに仮定をおくと, 最尤推定量は極限実験での最尤推定量に収束し, 尤度比統計量の列は尤度比統計量に収束することなども言える.

**記法 3.4.2.** 実験  $\mathcal{G}_n = (P_{n,h}|h \in H)_{n \in \mathbb{N}}$  上の統計量  $T_n: \mathcal{G}^n \rightarrow H$  を考える. 任意の分布  $P_h$  の下で,  $T_n \xrightarrow{P_h} L_h \in P(H)$  に法則収束するとする. このとき,  $T_n$  の漸近的な振る舞いは極限法則  $\{L_h\}_{h \in H}$  に支配される. ここで, 極限法則  $\{L_h\}_{h \in H}$  が, 極限実験における (少し変換した) 統計量の法則に対応する.

**定義 3.4.3 (dominated, randomized statistic).**

- (1) 実験  $\mathcal{G} = (P_h|h \in H)$  が**支配されている**とは, ある  $\sigma$ -有限測度  $\mu$  が存在して  $\forall_{h \in H} P_h \ll \mu$  が成り立つことをいう.
- (2)  $(\mathcal{G}, \mathcal{A}, P_h|h \in H)$  上の**ランダム化された統計量**  $T$  とは, 可測写像  $T: \mathcal{G} \times [0, 1] \rightarrow \mathbb{R}^k$  をいう. ただし,  $[0, 1]$  上には一様分布を与える.

**定理 3.4.4 (法則収束統計量の漸近表現定理).** 実験  $\mathcal{G}_n = (P_{n,h}|h \in H)_{n \in \mathbb{N}}$  は,  $\mu$  に支配された実験  $\mathcal{G} = (P_h|h \in H)$  に収束するとする.  $\mathcal{G}_n$  の統計量  $T_n$  が, 任意の  $h \in H$  について法則収束するならば,  $\mathcal{G}$  のランダム化された統計量  $T$  が存在して,  $\forall_{h \in H} T_n \xrightarrow{h} T$ .

### 3.4.3 漸近正規性

$h \in H$  にて局所漸近正規なモデルは, 正規分布モデル  $N(Jh, J)$  へ収束する実験列を定める.

**定理 3.4.5 (局所漸近正規ならば, 正規実験に収束する).**  $\mathcal{G}_n = (P_{n,h}|h \in H)$  を実験列で, パラメータ空間  $H \subset \mathbb{R}^d$  は  $0 \in H$  を満たすとする.  $h = 0$  下で  $N(0, J)$  ( $J \in M_d(\mathbb{R})$ ) に収束する確率変数列  $\Delta_n: \mathcal{G}^n \rightarrow \mathbb{R}^d$  を用いて

$$\log \frac{dP_{n,h}}{dP_{n,0}} = h^\top \Delta_n - \frac{1}{2} h^\top J h + o_{P_{n,0}}(1)$$

と展開できるとき, 実験列  $\mathcal{G}_n$  は  $(N(Jh, J))_{h \in H}$  に収束する.

### 3.4.4 一様分布

一様分布のモデル  $(U([0, \theta]))_{\theta \in \mathbb{R}_+}$  は  $L^2$ -微分可能ではなく, 正規な実験に収束しない. 実は, 指数実験に収束する.

### 3.4.5 Pareto 分布

Pareto 分布の極限実験は, 正規実験と指数実験の組み合わせとなる.

## 3.4.6 漸近混合正規性

## 3.5 尤度比確率場の局所漸近構造

統計推測の最適性の問題は、尤度解析に帰着される。

## 3.5.1 漸近決定理論

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

**歴史 3.5.1.** 漸近論への数学的アプローチで、初めて理論と言えるものは Wald 1943 と LeCam 1960, 1972, 1979 である。そこで、統計の実験とは、多様体  $\{P_\theta\}_{\theta \in \Theta} \subset P(\mathbb{R}^n)$  とされた。いまでは無限次元の場合（セミ／ノンパラメトリック）も考えられている。実験を繰り返すにつれて、実験は局所的に簡単なモデルで近似可能になっていく、この現象を Gaussian shift という。よって、問題は元の実験の研究から、Gaussian shift の近似へと移り変わる（極限定理）。これは本質的には実験の滑らかさに起因する。

**定義 3.5.2 (experiment, deficiency).**  $\Theta \neq \emptyset$  に関する実験とは、3つ組  $E = (\Omega, \mathcal{A}, \mathcal{P})$  をいう。パラメータ空間  $\Theta$  に関する実験全体の集合を  $\mathcal{E}(\Theta)$  で表す。  $\delta(E, F) := \inf \left\{ \epsilon > 0 \mid E \stackrel{\epsilon}{\subset} F \right\}$  ( $E, F \in \mathcal{E}(\Theta)$ ) に対して、  $\Delta(E, F) := \max \{ \delta(E, F), \delta(F, E) \}$  を欠損という。これは  $\mathcal{E}(\Theta)$  上に擬距離を定める。

$$\Delta(E, F) = \sup \{ \}$$

## 3.5.2 統計的実験

漸近正規性を定義する枠組みを整備する。Frechet 微分のような考え方をするが、 $L^2$  ノルムにより条件が弱められている [12]。

**定義 3.5.3.** 可測空間とその上の確率分布族  $(P_\theta)_{\theta \in \Theta}$  の組  $(\mathcal{X}, \mathcal{A}, \{P_\theta\}_{\theta \in \Theta})$  を統計的実験という。大標本理論では、直積による列  $(\mathcal{X}^n, \mathcal{A}^n, \{P_\theta^n\}_{\theta \in \Theta})$  を考える。

**記法 3.5.4 (正則性の仮定).**  $\Theta \stackrel{\text{open}}{\subset} \mathbb{R}^p$  とし、 $P_\theta$  の真値  $P_{\theta_0}$  に対する Lebesgue 分解における、絶対連続部分の Radon-Nikodym 微分を  $dP_\theta/dP_{\theta_0}$ 、特異部分を  $\sigma_{\theta/\theta_0}$  で表すと、

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

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

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

で定める。

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

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

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

**補題 3.5.5 ( $L^2$ -微分の表示).** G1, G2 を仮定し、 $P_\theta$  は確率密度関数族  $p(x, \theta)$  を持ち、対応  $\theta \mapsto p(x, \theta)$  は  $\theta_0$  において微分可能とする。このとき、

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



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

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

とおく.

**補題 3.5.6.**

### 3.5.3 局所漸近正規性

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

対応  $\mathbb{R}^p \ni u \mapsto Z_n(u; \theta_0) \in C(\mathbb{R}^p)$  を尤度比確率場という. Ibragimov HasMinskii は, この関数の  $C(\mathbb{R}^p)$  上での弱収束を示した:  $Z_n(-; \theta_0) \xrightarrow{d} Z(-; \theta_0)$ .

**記法 3.5.7** (normalized likelihood ratio). 正規化された尤度比を

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

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

**定理 3.5.8 (LeCam).** 正規な統計的実験の列  $E_i$  について,  $I(\theta_0)$  ( $\theta_0 \in \Theta$ ) は非退化とする. [G1],[G2] を仮定し  $\varphi$  を  $D_u$  の  $L^2$ -微分とする. このとき, ある  $p$  次元と 1 次元確率変数  $\Delta_n(\theta_0), \rho_n(u, \theta_0)$  が存在して, 正規化された尤度比は次の表現を持つ:

$$Z_n(u; \theta_0) := \exp \left( u^\top \Delta_n(\theta_0) - \frac{1}{2} u^\top I(\theta_0) u + \rho_n(u, \theta_0) \right).$$

また,

$$L[\Delta_n(\theta_0) | P_{\theta_0}^n] \rightarrow N_p(0, I(\theta_0)), \quad \rho_n(u, \theta_0) \xrightarrow{P_{\theta_0}^n} 0.$$

**[証明].**

$$\Delta_n(\theta_0) := \frac{2}{\sqrt{n}} \sum_{j=1}^n \varphi(x_j), \quad \rho_n(u, \theta) := \log Z_n(u; \theta_0) - u^\top \Delta_n(\theta_0) + \frac{1}{2} u^\top I(\theta_0) u.$$

とすれば良い. ■

**要諦 3.5.9.** この定理が保証する尤度比の性質は, 推定量の性質を調べるに当たって強力な道具となる. 特に, 尤度比の対数  $\log Z_n$  は  $N\left(-\frac{1}{2} u^\top I(\theta_0) u, u^\top I(\theta_0) u\right)$  に分布収束する.

**定義 3.5.10 (LAN: Locally Asymptotically Normal (LeCam 1960)).** 定理の  $Z_n(u; \theta_0)$  の展開が成り立つとき, すなわち, G1, G2 が成り立つとき, 統計的実験  $(\mathcal{X}^n, \mathcal{A}^n, \{P_\theta^n\}_{\theta \in \Theta})$  は  $\theta_0$  において局所漸近正規であるという.

**要諦 3.5.11.** LAN のとき,  $P_{\theta_0+n^{-1/2}u}^n$  は  $P_{\theta_0}^n$  に接触している.

### 3.5.4 局所漸近正規な例

統計モデルの列が正規モデルで近似可能であるという性質が, 局所漸近正規性である. 正則なパラメトリックモデルの独立同分布モデルは局所漸近正規で, これについての LeCam の定理は中心極限定理に他ならない.

**定理 3.5.12 (Hajek の十分条件).**  $\Theta \subset \mathbb{R}^1$  で添字付けられた実験  $E_i$  の分布族  $f(x, \theta) = \frac{dP_\theta}{d\nu}$  が次の 3 条件を満たすならば, 族  $P_\theta^n$  は LAN 条件を  $\theta = t$  について満たす.

- (1) 任意の  $x \in \mathcal{X}$  について  $f(x, -) : \Theta \rightarrow [0, 1]$  は  $\theta = t$  の近傍  $U \subset^{\text{open}} \Theta$  で絶対連続である.
- (2)  $\nu$ -a.e.  $x \in \mathcal{X}$  について, 微分係数  $\frac{\partial f(x, \theta)}{\partial \theta}$  が  $\theta \in U$  上存在する.
- (3)  $I(\theta)$  ( $\theta = t$ ) は連続で半正定値である.

### 3.5.5 漸近有効性

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

**定理 3.5.13 (Hajek の不等式).** 統計的実験は  $\theta_0$  にて局所漸近正規であるとする。  $\theta$  の任意の推定量系列  $T_n$  について、

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

## 3.6 漸近有効性

滑らかなパラメトリックモデルの漸近下界は  $N(0, I_\theta^{-1})$  が与える。

**記法 3.6.1.**  $\sqrt{n}(T_n - \psi(\theta))$  は全ての  $\theta \in \Theta$  について分布収束すると仮定して（この仮定を正則性という）、その効率性を比較する。

### 3.6.1 実験の下界

漸近有効性の議論は、正規な位置母数モデル (Gaussian shift モデル) での有効推定量の問題に帰着する。

**定義 3.6.2 (regular).** 統計量の列  $(T_n : \mathcal{X}^n \rightarrow \Theta)$  が  $\theta$  において正則または漸近的法則同等であるとは、ある極限分布  $L_\theta$  が存在して、

$$\forall h \in \Theta \quad \sqrt{n} \left( T_n - \psi \left( \theta + \frac{h}{\sqrt{n}} \right) \right) \xrightarrow[\theta + h/\sqrt{n}]{d} L_\theta$$

が成り立つことをいう。極限分布  $L_\theta$  が  $h$  に依ることを許すとき、**前正則**であるとする。前正則な統計量が正則であるとは、局所一様に取れる消息を表す。

**定理 3.6.3.**

(E1) 実験  $(P_\theta)_{\theta \in \Theta}$  は点  $\theta \in \Theta \stackrel{\text{open}}{\subset} \mathbb{R}^k$  で  $\theta \mapsto dP^{1/2}$  が  $L^2$ -微分可能で、可逆な Fisher 行列  $I_\theta \in \text{GL}_k(\mathbb{R})$  を持つとする。

(E2)  $\psi : \Theta \rightarrow \mathbb{R}$  は  $\theta$  で微分可能であるとする。

(E3) 実験  $(P_{\theta+h/\sqrt{n}})_{h \in \mathbb{R}^k}$  の統計量  $T_n$  は前正則であるとする。

このとき、ランダム化された統計量  $T$  が実験  $(N(h, I_\theta^{-1}))_{h \in \mathbb{R}^k}$  上に存在して、 $T - \dot{\psi}_\theta h \sim L_{\theta, h}$ 。

### 3.6.2 Gauss モデルの平均の推定

### 3.6.3 畳み込み定理

正則な統計量の中で、最適なものは解析的に定まる。

**定理 3.6.4.**

(E1) 実験  $(P_\theta)_{\theta \in \Theta}$  は点  $\theta \in \Theta \stackrel{\text{open}}{\subset} \mathbb{R}^k$  で  $\theta \mapsto dP^{1/2}$  が  $L^2$ -微分可能で、可逆な Fisher 行列  $I_\theta \in \text{GL}_k(\mathbb{R})$  を持つとする。

(E2)  $\psi : \Theta \rightarrow \mathbb{R}$  は  $\theta$  で微分可能であるとする。

(E3) 実験  $(P_\theta^n)_{\theta \in \Theta}$  の統計量  $T_n$  は正則で極限分布  $L_\theta$  を持つとする。

このとき、ある確率測度  $M_\theta \in P(\mathcal{X})$  が存在して、 $L_\theta = N(0, \dot{\psi}_\theta I_\theta^{-1} \dot{\psi}_\theta^\top) * M_\theta$  と表せる。



## 3.6.4 局所漸近 minimax 定理

全く違うアプローチで、最適な統計量が正規分布  $N(0, \dot{\psi}_\theta I_\theta^{-1} \dot{\psi}_\theta^\top)$  に漸近的に従うことを示す方法を考える。

## 3.6.5 下界を達成する推定量

漸近分布  $N(0, \dot{\psi}_\theta I_\theta^{-1} \dot{\psi}_\theta^\top)$  を持つ推定量が最適とわかったから、このようなものの必要十分条件を探す。これは実は漸近線型性である。線形性が正規性に対応するのである！

**補題 3.6.5 (漸近線型性: 最適性の特徴づけ).**

(E1) 実験  $(P_\theta)_{\theta \in \Theta}$  は点  $\theta \in \Theta \stackrel{\text{open}}{\subset} \mathbb{R}^k$  で  $\theta \mapsto dP^{1/2}$  が  $L^2$ -微分可能で、可逆な Fisher 行列  $I_\theta \in \text{GL}_k(\mathbb{R})$  を持つとする。

(E2)  $\psi: \Theta \rightarrow \mathbb{R}$  は  $\theta$  で微分可能であるとする。

このとき、実験  $(P_\theta^n)_{\theta \in \mathbb{R}^k}$  の統計量列  $(T_n)$  について、次の2条件は同値。

$$(1) \text{ 漸近線型である: } \sqrt{n}(T_n - \psi(\theta)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{\psi}_\theta I_\theta^{-1} \dot{l}_\theta^\top(X_i) + o_P(1).$$

$$(2) T_n \text{ は } \psi(\theta) \text{ に対する正則な統計量の中で最適である: } \forall h \in \Theta \quad \sqrt{n} \left( T_n - \psi \left( \theta + \frac{h}{\sqrt{n}} \right) \right) \xrightarrow[\theta+h/\sqrt{n}]{d} N(0, \dot{\psi}_\theta I_\theta^{-1} \dot{\psi}_\theta^\top).$$

**要諦 3.6.6.**  $\Delta_{n,\theta} := \frac{1}{\sqrt{n}} \sum \dot{l}_\theta(X_i)$  の部分が  $N(0, I_\theta)$  に収束する。

**例 3.6.7.** 適切な正則条件の下で、最尤推定量は漸近線型である。したがってデルタ法によって、任意の汎関数  $\psi(\theta)$  に対して、plug-in 推定量は漸近線型である。  $\square$

## 第 4 章

# 漸近展開

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

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

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

### 4.1 漸近展開

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

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

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

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

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

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

### 4.2 平滑化補題

分布の差と特性関数の差の関係を与える補題を準備する。

### 4.3 特性関数の展開

### 4.4 漸近展開の正当性の証明

### 4.5 漸近展開の変換

### 4.6 最尤推定量の漸近展開

### 4.7 歴史

物理学者も解析学者も参入した対象であり、中心極限定理とはやはり夢のテーマであった。ロシアが強いことに疑問もない。正規分布には、算術平均が最尤推定量になるということと、微小分布へ無限分解可能な分布としての2つの普遍性がある。

**歴史 4.7.1** (Stirling の公式と de Moivre-Laplace の定理). Fermat と Pascal の往復書簡 (1654) ののちに、Huygens が De Ratiociniis in Ludo Aleae (1657) を書き、これを James Bernoulli が、まとめて Ars Conjectandi (1713) が死後に出版された。この時点で第四部では推論に応用することを考えたのである。復元抽出の標本数  $n$  をどこまで大きくすれば、確率  $p$  を一定の誤差以下で評価できるかを考えた。これは多項分布の問題であるから、Newton の二項公式や Bernoulli 数の技法が第二部でまとめられている。この問題を乗り越えたのが De Moivre (1733) で、多項分布の確率が正規曲線  $e^{-x^2}$  の積分で近似できることを発見した。この自分の論文を英訳し、The Doctrine of Chances, 2nd Ed. (1738) に収録して、友人に配った。さらに  $n!$  の近似の問題を解いて精緻化したのが Stirling (1730, Methodus differentialis) である。De Moivre の友人であった。この点で本質的に解析学化したのが de Moivre であったが、本格的に微積分学の言葉で書き直したのが Laplace であった。こうして、de Moivre-Laplace の定理が示された。解析学とは極限の技法で、極限とは近似の基礎である。二項分布とは、二点分布に従う独立同分布列の和の分布でもあることに注意すれば、これは中心極限定理の系である。

**歴史 4.7.2** (誤差分布と天文学). Simpson (1775) の London 王立学会に当てた手紙に、天文学者は算術平均を推定量として使うが、注意深く測定した1回の方が信頼できるという意見の知識人も多いという現状を書いている。そこで Simpson, Lagrange らは誤差分布として考え得る分布たち（裾が軽く、対称な分布）の算術平均の分布が調べられた。

回帰分析だが、方程式の形をパラメトリックに仮定し、観測値を集めると、解析的には解無しになるが、実際は仮定も正しくなく、観測誤差もあるので、損失関数を最小にする係数を推定量とすることになる（決定問題）。Gauss が最終的にこれを解決し、標本平均が位置母数の最尤推定量になるのは誤差法則が正規分布の場合に限ることを解析的に示し、このとき最小二乗法が必然的に最良になることを示した。算術平均を信じるなら、誤差分布は正規性を仮定することになるのだ。

**歴史 4.7.3** (誤差とは何か). Young は誤差という概念の安定性に疑問を持ち、「誤差とは、根元誤差の独立和に分解できるのはいか」「そして独立和は、元々の誤差分布に依らずに特定の分布に従うのでは？」と、裏の数理論に目を向けた。これは数学的には新規性はないが、無限可分分布という概念に目を向けさせる。こうして Gauss とは独立に、もう一度誤差分布に正規分布を用いる妥当性が導かれる。つまり、3 次以上の積率が無視できるほど小さいような根元誤差の話の分布は、正規分布に従うことになる。「微小分布の和」としての普遍性があるわけだ。

**歴史 4.7.4.** Chebyshev, Markov の順に十分条件が厳密に見つかっていき、Lyapunov が特性関数の方法で数学的に正しいものをひとまず1つ確定させた。Lindeberg, Levy, Feller, Kolmogorov が引き継ぎ、現代論になる。

## 第 5 章

# 情報幾何学

### 5.1 漸近展開と情報幾何

最尤推定量の漸近展開の係数の一部は、接続係数として解釈できる。統計量の漸近展開式に幾何学的解釈を与え、その高次の統計的性質を説明しようとする研究の流れは、Efron から始まり、甘利俊一による  $\alpha$ -接続の重要性の認識によって情報幾何学として確立された。特に、1 次漸近有効性の枠組みを脱して、高次漸近有効性の議論も可能になる。

### 5.2 変分

#### 5.2.1 定義

**歴史 5.2.1.** 1910 年頃から統計的な距離の意味で使われており、論文では Bhattacharya (1943). On a Measure of Divergence between Two Statistical Populations Defined by Their Probability Distributions. で Bhattacharya 距離 (距離ではない) を定義し、これに Kullback and Leibler (1951) が続いたものが始まりである。ただし、Kullback (1959) では現在の KL-変分を "directed divergence" と呼んで、対称化を考えている。現代の意味での用法は Amari (1985) からであるが、それまでは "quasi-distance" (Amari 1982), "contrast function" (Eguchi 1985) などを読んだ。

**定義 5.2.2 (divergence).**  $(M, (\xi_x))$  を多様体とする。  $C^2$ -級関数  $D: M \times M \rightarrow \mathbb{R}_+$  が分離度であるとは、

- (1) 非退化性:  $D[P|Q] = 0 \Leftrightarrow P = Q$ .
- (2) 第二変数が計量の二次形式である: 第二変数に関する微分の Hesse 行列  $G(\xi) = (g_{ij}(\xi))$  は正定値対称である。
- (3) 任意の点  $P \in M$  において、開近傍系  $(U_r(P) := \{Q \in M \mid D[P, Q] < r\})_{r>0}$  は単調増大。

#### 5.2.2 例

**定義 5.2.3 (f-divergence).**

**定理 5.2.4 (Fisher 計量).**  $C^3$ -級のパラメトリック模型  $\mathcal{P} = (P_\theta)_{\theta \in \Theta}$  ( $\Theta \subset \mathbb{R}^p$ ) について、

- (1) 1 次の微小項が消える:  $\forall_{j \in [p]} \left. \frac{\partial}{\partial \theta_j} \right|_{\theta=\theta_0} D(P_\theta|P_{\theta_0}) = 0$ .
- (2) KL-変分の合成  $\theta \mapsto D(P_\theta|P_{\theta_0})$  の Hesse 行列を  $G = \left( g_{jk}(\theta_0) := \left. \frac{\partial^2}{\partial \theta_j \partial \theta_k} \right|_{\theta=\theta_0} D(P_\theta|P_{\theta_0}) \right)$  とする:

$$D(P_\theta|P_{\theta_0}) = \frac{1}{2} \Delta \theta_j \Delta \theta_k g_{jk}(\theta_0) + o(|\theta - \theta_0|^2) \quad \Delta \theta_j := (\theta - \theta_0)_j.$$

このとき、 $G$  は半正定値で、 $\Theta$  上に Riemann 計量を定める。

**例 5.2.5** (正規分布族の KL 分離度).  $M_n(\mathbb{R})$  中の半正定値行列のなす部分多様体  $\text{GL}_n(\mathbb{R})_+$  は  $n(n+1)/2$  次元である. この多様体に

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

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

## 第 6 章

# 再標本化法

データから推定して得た分布  $\hat{P}_n$  から再度標本を取る手続きを含む手法を、**リサンプリング法**という。

### 6.1 ブートストラップ法

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

**模型 6.1.1** (ブートストラップ法が適用可能な模型).  $P$  を真の分布,  $\{X_j\}_{j \in [n]}$  を独立な観測とする. 確率変数  $R_n(\mathbf{x}_n; P) : \mathbb{R}^n \times P(\mathbb{R}) \rightarrow \mathbb{R}$  の分布関数

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

に対して,  $H_n(c_\alpha; P) = \alpha$  を満たす点  $c_\alpha \in \mathbb{R}$  を求める問題を考える. 例えば, 統計的汎関数  $\theta : P(\mathbb{R}) \rightarrow \mathbb{R}$  の推定量  $\hat{\theta}_n : \mathbb{R}^n \rightarrow \mathbb{R}$  に対して,  $R_n(\mathbf{x}_n; P) := \sqrt{n}(\hat{\theta}_n(\mathbf{x}_n) - \theta(P))$  とすれば, 検定が構成できる.

(1)  $P$  の推定量  $\hat{P}_n$  が得られたとする. このとき, プラグイン分布

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

を **Bootstrap 分布**という.

(2) プラグイン分布  $\hat{P}_n$  から大きさ  $n$  の無作為標本  $\mathbf{X}_n^* = (X_1^*, \dots, X_n^*)$  を用意する. これを **Bootstrap 標本**という.

特に, Bootstrap 分布が陽に表せないとき, (2) の手続きが困難になる. これは**計算機集約的** (computer-intensive) な統計手法である.

**記法 6.1.2.**  $(X_j)$  を確率分布  $P$  に従う独立な実確率変数列とし,  $R_n(\mathbf{X}_n, P) : \mathcal{X} \rightarrow \mathbb{R}$  を確率変数とする.  $R_n(\mathbf{X}_n, P)$  の分布関数を

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

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

**例 6.1.3.**  $P$  が未知で,  $\theta = \theta(P)$  の信頼区間を構成したいとき, 推定量  $\widetilde{\theta}_n(\mathbf{X}_n)$  から定まる確率変数  $R_n(\mathbf{X}_n, P) := \sqrt{n}(\widetilde{\theta}_n(\mathbf{X}_n) - \theta(P))$  の分位点  $c_\alpha$  を求めたい. □

**定義 6.1.4.**  $\hat{P}_n$  を  $P$  の推定量とする.

(1) 真の分布関数  $H_n(\mathbf{x}; P)$  に対する近似  $H_n(\mathbf{x}; \hat{P}_n) := \int_{\mathbb{R}^n} \mathbf{1}_{R_n(\mathbf{x}, \hat{P}_n) \leq x} (\hat{P}_n)^n(d\mathbf{x})$  を, **ブートストラップ分布**という.

(2) 推定分布  $\hat{P}_n$  に従う大きさ  $n$  の無作為標本  $\mathbf{X}^* = (X_1^*, \dots, X_n^*)$  を**ブートストラップ標本**という.

**注 6.1.5** (computer-intensive). 推定分布  $\hat{P}_n$  が複雑であるとき, 積分  $H_n(\mathbf{x}; \hat{P}_n)$  はモンテカルロ法で近似する. これが, ブートストラップ法が計算機集約的と言われる所以である.



**注 6.1.6** (studentization).  $R_n(\mathbf{X}_n, P)$  をステューデント化された確率変数で置き換えることで、ブートストラップ分布の近似精度の改善が可能である.

## 第 7 章

## 参考文献

## 参考文献

- [Hampel, Frank R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel W. A., 2005] Hampel, Frank R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel W. A. (2005). Robust Statistics: The Approach Based on Influence Functions. Wiley Series in Probability and Statistics. Wiley.
- [Huber, Peter J., and Ronchetti, E. M., 2009] Huber, Peter J., and Ronchetti, E. M. (2009). Robust Statistics. Wiley Series in Probability and Statistics. Wiley, 2 edition.

## 参考文献

- [1] 吉田朋広 (2006) 『数理統計学』(朝倉書店)
- [2] 鍋谷清治 (1978) 『数理統計学』(共立出版)
- [3] 竹内啓 『統計的推定の漸近理論』
- [4] 前園宜彦 (2019). 『ノンパラメトリック統計』(共立講座, 数学の輝き 12).
- [5] 前園宜彦 (2001). 『統計的推測の漸近理論』(九州大学出版会).
- [6] van der Vaart. (1998). *Asymptotic Statistics*. Cambridge Univ. Press.
- [7] Ferguson, T. S. (1996). *A Course in Large Sample Theory*. Taylor and Francis Group.  
漸近展開について参照すべきは以下である.
- [8] 柴田義貞 (1981) 『正規分布』(UP 応用数学選書, 東京大学出版会).
- [9] Bhattacharya, R. N., and Ghosh, J. K. (1978). On the Validity of the Formal Edgeworth Expansion. *Annals of Statistics*. 6(2):434-451.
- [10] Sweeting, T. J. (1977). Speeds of Convergence for the Multidimensional Central Limit Theorem. *Annals of Probability*. 5:29-41.
- [11] Rabi Bhattacharya - Course in Mathematical Statistics and Large Sample Theory
- [12] Ibragimov and Has'minskii - Statistical Estimation
- [13] Helmut Strasser "Mathematical Theory of Statistics"
- [14] 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.
- [15] 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.