

第5章 統計的推測論の枠組み

この章では統計的推測の枠組みについて述べる。節 5.1 では、統計的推測理論で重要な役割を果たす統計的モデルを説明する。節 5.2 では、統計的推測理論を統一的な枠組みで議論する統計的決定理論を簡単に説明する。

5.1 統計的実験と母数モデル

$\mathbb{X}$ を距離空間とし、 $\mathcal{B}(\mathbb{X})$ を $\mathbb{X}$ 上の Borel 集合族¹ とする。すなわち、 $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ は可測空間である。

いま、 $X_1, X_2, \dots, X_n$ は確率空間 $(\Omega, \mathcal{A}, \Pr)$ 上の $\mathbb{X}$ 値確率要素² とする。さらに、 $\mathbf{X} = (X_1, X_2, \dots, X_n)^\top$ と書くことにする。典型的な例は $\mathbb{X} = \mathbb{R}^d$ ($d \in \mathbb{N}$) 等である。

この講義では $X_1, X_2, \dots, X_n$ は独立同一の分布に従うものとする。このような場合、 $X_1, X_2, \dots, X_n$ を ランダム標本 と呼ぶ。また、可測空間 $(\mathbb{X}^n, (\mathcal{B}(\mathbb{X}))^{\otimes n})$ を 標本空間 と呼ぶ。

可測空間 $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ 上の確率測度 P を

$$P(B) := \Pr(X_1 \in B) \quad (B \in \mathcal{B}(\mathbb{X}))$$

で定義する。確率変数 X_1 によって誘導された $\mathbb{X}$ 上の確率 P を X_1 の確率分布 (簡単に分布ともいうことがある) といい、 $X_1 \sim P$ と表記することにする。さらに、 $\mathbf{X}$ の分布を $P^{\otimes n}$ または $P^{\mathbf{X}}$ と書くことにする。したがって、 $B_j \in \mathcal{B}(\mathbb{X})$ ($j = 1, 2, \dots, n$) に対して

$$P^{\otimes n}(B_1 \times B_2 \times \cdots \times B_n) = \Pr(\mathbf{X} \in B_1 \times B_2 \times \cdots \times B_n)$$

となる³。

注意 5.1. $B \in \mathcal{B}(\mathbb{X})$ に対して、 $\Pr(X_1 \in B)$ と書いたとき

$$\Pr(X_1 \in B) = \Pr(\{\omega \in \Omega; X_1(\omega) \in B\})$$

¹開集合族を含む最小の σ 加法族

²確率変数, 確率ベクトル, 確率行列の総称を確率要素という。

³この記法 $P^{\otimes n}$ にも測度論的な背景がある。すなわち、 $P^{\otimes n}$ は P の n 個の測度の直積測度になっている。

の意味である. □

確率論的アプローチでは、観測データを生成するメカニズムを表現する確率分布 P は既知であり、確率要素 $\mathbf{X}$ の分布論的な特徴を調べる。一方、統計的推測のアプローチでは観測データを生成するメカニズムを表現する確率測度 P は観測者には未知であり、観測データ $\mathbf{X}$ に基づいて未知の確率測度 P を回復することを目指す。統計的推測は確率論的アプローチの逆問題と言える。

統計的推測の考え方は統計的実験という概念を基礎に組み立てられる。 $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} P^*$ とする。ここで、 P^* はデータを生成する真の確率分布とした。すなわち、 $\mathbf{X} := (X_1, \dots, X_n) \sim (P^*)^{\otimes n}$ である。統計的推測では観測データを生成するメカニズムを表現する未知の確率分布 P^* を観測データから回復することを目的とする。 $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ 上の確率分布全体の集合から真の分布を探すのは多くの困難を伴うので、候補となる確率分布の適当な集まりを設定する。この確率分布族の特定の集まりを $\mathcal{P}$ と書いたとき、この $\mathcal{P}$ を統計的モデルという。そして、真の確率分布 P^* は設定した統計的モデル $\mathcal{P}$ に含まれるとこの講義では仮定⁴する。

さらに統計的モデルの各要素はある集合 Θ の要素 θ で添え字付けられると仮定する。すなわち、添え字集合全体と統計的モデルとの対応である写像

$$\Theta \ni \theta \mapsto P_\theta \in \mathcal{P}$$

を想定⁵する。これを統計的モデルの母数化といい、 Θ を母数空間、その要素 θ を母数という。そして真の確率分布 P^* に対応するある母数 $\theta^* \in \Theta$ が存在⁶すると仮定する。この θ^* を真の母数ということにする。すなわち

$$P^* = P_{\theta^*}$$

である。もちろん $\theta^* \in \Theta$ ではあるが、真の母数 θ^* は観測者には未知である。したがって、統計的推測では未知の真の母数 θ^* を観測データ $\mathbf{X}$ から回復することが目的⁷となる。

次に統計的モデルのなかでも最も基本的な正則母数モデルを定義する。

⁴もちろん、真の確率分布 P^* が想定した統計的モデルに含まれないことを想定した議論もある。しかし、統計的モデルに真の分布が含まれない場合には、より高度な議論が必要となるので、この講義録ではより基本的な設定を考えている。

⁵この写像がどのような性質を持つかを数学的に定義することはできる。しかし、確率分布の集まりの空間に位相をいれることになり、数学的に議論が高度になるので、漠然とした写像と理解しておくことにしよう。

⁶母数化が識別可能（この概念は後に定義）であれば、この母数は一意である..

⁷統計的機械学習では、真の母数 θ^* の回復を目標とするより、真の確率測度と同等のメカニズムから生成される未来の観測 $\tilde{X}_{n+1}$ を観測データ $\mathbf{X}$ に基づいて予測することを目標にしている。統計的機械学習では、統計的モデルの役割が相対的に低くなる。

定義 5.2. 統計的モデル $\mathcal{P} = \{P_\theta; \theta \in \Theta\}$ が 正則母数モデル(regular parametric model) であるとは、次の条件をみたすときをいう。

- (1) 母数空間 Θ は 有限次元 Euclid 空間 $\mathbb{R}^d$ の「よい」部分集合である。ただし $d \in \mathbb{N}$ である。

- (2) 写像

$$\Theta \ni \theta \mapsto P_\theta \in \mathcal{P}$$

は「滑らか」である⁸。

条件 (2) を母数化の正則性という。正則母数モデルを簡単に母数モデルということもある。

さらにこの講義では次の条件も仮定する。

- (3) $\forall \theta_1, \theta_2 \in \Theta$ に対して

$$\theta_1 \neq \theta_2 \Rightarrow P_{\theta_1} \neq P_{\theta_2}$$

をみたす。

すなわち母数化を与える写像は単射である。このような母数化を 識別可能(identifiable) であるという。

注意 5.3. 定義 5.2 は数学的には厳密ではない表現である。数学的により厳密な母数モデルの定義については Bickel *et al.* (1993, pp.11-13) を参照のこと。

注意 5.4. 母数空間 Θ が有限次元ではないような統計的モデルを考えることも重要である。 Θ が無限次元の統計的モデルのことを「ノンパラメトリック・モデル」と統計学では慣例的に呼んでいることが多い。母数空間が無限次元とはいえ、統計的モデルは母数化されているので、「非母数モデル」と呼ぶのは奇異である。しかし歴史的にこの用語が使用されてきたので、この講義録では統計学の歴史的慣例に従うことにする。この言葉使いを嫌う数理統計学者は「ノンパラメトリック・モデル」のことを「無限次元統計的モデル」と呼んでいる。さすがに「無限次元母数モデル」とは言わないようである。さらに母数空間が有限次元の母数空間と無限次元の母数空間の直積で表現され、有限次元の母数を回復の対象とするような統計的モデルを「セミパラメトリック・モデル⁹」という。これも言葉の意味としては奇異であるが、統計学の習慣に従うことにする。

⁸ $\mathcal{P}$ は測度の集合なので位相をどのようにいれるかはすこし難しい議論になる。 $\mathcal{P}$ が Radon 測度の集まりならば、weak-star 位相を入れることができる。この議論は Tojo and Yoshino (2021) を参照のこと。

⁹英語読みをすれば、「セマイパラメトリックモデル」という。「semi-parametric model」の最初の「i」は長母音であることに注意が必要である。

生存データ解析で広く使用される Cox の比例ハザード・モデルはセミパラメトリック・モデルの最高傑作であろう。20 世紀の数理統計学の到達点のひとつであるセミパラメトリック・モデルに対する統計的推測理論については Bickel *et al.* (1993), van der Vaart and Wellner (1996), van der Vaart (1998), Kosorok (2007), 久保木・鈴木 (2015) を参照のこと。□

定義 5.5. (1) 可測空間 $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ と母数モデル

$$\mathcal{P} = \{P_\theta; \theta \in \Theta, P_\theta \text{ は } (\mathbb{X}, \mathcal{B}(\mathbb{X})) \text{ 上の確率分布}\}$$

の組を 統計的実験(statistical experiment) といい

$$\left(\mathbb{X}^n, \mathcal{B}(\mathbb{X})^{\otimes n}, \{P_\theta^{\otimes n}; \theta \in \Theta\} \right)$$

と書く。

(2) 観測データを生成するメカニズムを表現する真の確率測度 P^* に対応する母数を $\theta^* \in \Theta$ と書くことにする。すなわち, θ^* は真の母数で

$$P^* = P_{\theta^*}$$

である。

この講義では $\mathbf{X} = (X_1, X_2, \dots, X_n)$ と書いたとき, $X_1, X_2, \dots, X_n$ は独立に同一分布 P^* に従うことを仮定する。上記の仮定をおいた観測データ $\mathbf{X}$ のことを 標本の大きさが n のランダム標本 という。

以上の議論から, この講義で扱う統計的実験をまとめると下記のようなになる。

この講義で仮定する統計的実験

(1) 観測データを

$$\mathbf{X} = (X_1, X_2, \dots, X_n)^\top$$

と書き, 可測空間 $(\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n))$ に値をとる。

(2) 観測データは真の分布 P^* からの標本である。すなわち

$$\mathbf{X} \sim (P^*)^{\otimes n}$$

である。ただし P^* は $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ 上の確率測度である。

(3) 統計的モデル $\mathcal{P} = \{P_\theta; \theta \in \Theta\}$ を設定する。ただし P_θ も P^* と同じ可測空間 $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ 上の確率測度である。

(4) 母数空間 Θ は Euclid 空間 $\mathbb{R}^d$ の「よい」部分集合である。ただし $d \in \mathbb{N}$ である。

- (5) $\Theta \ni \theta$ から $P_\theta \in \mathcal{P}$ への写像は「滑らか」かつ単射 (母数化の識別可能性を仮定).
- (6) ある $\theta^* \in \Theta$ が唯一あって $P^* = P_{\theta^*}$ と書ける.

これらをまとめて統計的実験といい

$$(\mathbb{X}^n, \mathcal{B}(\mathbb{X})^{\otimes n}, \{P_\theta^{\otimes n}; \theta \in \Theta\})$$

と書き, $\mathcal{P} = \{P_\theta^{\otimes n}; \theta \in \Theta\}$ を正則母数モデルという. ランダム標本の場合には, 簡単に $\mathcal{P} = \{P_\theta; \theta \in \Theta\}$ とも書くことにする. 以後は単に母数モデルということにする. そして統計的推測の目標は真の確率分布 P^* からのランダム標本 $(X_1, X_2, \dots, X_n)$ に基づき θ^* を回復することである.

例 5.6. (1) $X_1, X_2, \dots, X_n$ を正規分布 $N(\mu, \sigma^2)$ から標本の大きさが n のランダム標本とする. ただし μ, σ ($0 < \sigma < \infty$) が共に未知とする. このとき統計的実験

$$\left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \left\{ p_\theta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) (-\infty < x < \infty); \right. \right. \\ \left. \left. \theta := (\mu, \sigma) \in \Theta := \mathbb{R} \times (0, \infty) \right\} \right)$$

を想定していることになる. 統計的モデルは確率分布が特定できる表現でよいので, この場合には p.d.f. で表現している.

(2) $X_1, X_2, \dots, X_n$ を Bernoulli 分布 $\text{Ber}(\theta)$ ($0 < \theta < 1$) から標本の大きさが n のランダム標本とする. ただし θ が未知のときには, 統計的実験

$$\left(\{0, 1\}^n, 2^{\{0, 1\}^n}, \left\{ p_\theta(\mathbf{x}) = \theta^x (1-\theta)^{1-x} (x = 0, 1); \theta \in \Theta = (0, 1) \right\} \right)$$

を想定していることになる. □

注意 5.7. $X_1, X_2, \dots, X_n$ を分布 P^* から標本の大きさが n のランダム標本といったときには,

$$\left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{P_\theta; \theta \in \Theta\} \right)$$

のような統計的実験を仮定し, ある $\theta^* \in \Theta$ があって, $P_{\theta^*} = P^*$ であることを想定している. ことさらに統計的実験という用語を今後は用いないことにする. □

5.2 統計的決定問題

統計的推測理論には多くのアプローチがある。その中で代表的なアプローチが二つある。一つは頻度論的なもので、もう一方はベイズ論的なものである。以下では頻度論的推測理論の枠組みを説明することにする。Bayes 論的推測理論は第 9 章で説明する。以下では、頻度論的推測理論の枠組みを統計的決定理論¹⁰の言葉を使って説明する。

標本空間を $\mathbb{X}^n$ とし、観測データを $\mathbf{X}$ とする。

- (1) まず観測データに基づき行う行動のすべてを集めた集合を 行動空間 といい $\mathbb{A}$ で記す。この講義では $\mathbb{A} = \mathbb{R}$ や $\mathbb{A} = \{0, 1\}$ などである。観測者が観測データに基づき行動 $\mathbb{A}$ の要素を選択するルールを決定関数といい

$$d: \mathbb{X}^n \ni \mathbf{x} \mapsto d(\mathbf{x})$$

で記す。決定関数の集まりを 決定空間 といい、 $\mathbb{D}$ と記す。したがって観測者は合理的な行動 d が存在すればありがたいわけである。

- (2) 次に行動を評価するための道具として直積空間 $\Theta \times \mathbb{A}$ 上の非負値実数値関数

$$L: \Theta \times \mathbb{A} \mapsto [0, \infty) \cup \{\infty\}$$

を用意する¹¹。この関数を 損失関数 といい、 $L(\theta, a)$ の値が小さいほど望ましい行動である。決定関数の「よさ」を評価するには観測データの実現値 $\mathbf{X} = \mathbf{x}$ と真の母数 θ^* における損失関数の値 $L(\theta^*, d(\mathbf{x}))$ がわかればよい。したがって決定関数 d の「よさ」の評価に $L(\theta^*, d(\mathbf{X}))$ を使えばよいのだが、これは用いることができない。これはランダムな量であり、未知の母数 $\theta^* \in \Theta$ がわからないので、 $L(\theta^*, d(\mathbf{x}))$ は知ることができない量である。そこで母数 θ に対して観測データが $P_\theta^{\otimes n}$ によって生成されたと仮定し、損失関数 $L(\theta, d(\mathbf{X}))$ を $P_\theta^{\otimes n}$ に関して期待値を取ったもの

$$R(\theta, d) := E_\theta[L(\theta, d(\mathbf{X}))]$$

を考える。ただし、 $E_\theta[\cdot]$ は $\mathbf{X} \sim P_\theta^{\otimes n}$ ($\theta \in \Theta$) のもとでの期待値である。これを決定関数 d の 危険関数 という。なお、危険関数は $\theta \in \Theta$ の関数であることに注意せよ。

- (3) 決定関数 d の「よさ」の評価は真の母数 θ^* のもとで行いたいところである。しかしこれは未知である。危険関数の $\theta \in \Theta$ に関する

¹⁰統計的決定理論はゲーム論の概念を借用して、統計的推測論の枠組みと最適理論を定式化(言語化)したものである。

¹¹ただし、区間推定のばあいには、 $L: \mathbb{A} \times \Theta \mapsto [-1, \infty) \cup \{\infty\}$ とすることもある。

なんらかの様な評価が必要になってくる. このことから危険関数の母数空間に関する様な評価が統計的推測論の深みと困難の淵源である. またこれが統計的推測理論のわかりにくさの原因でもあろう. 前節で説明した統計的実験に行動空間, 決定空間, そして損失関数を加えた組

$$(\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n), \{P_\theta; \theta \in \Theta\}, \mathbb{A}, \mathbb{D}, L)$$

を統計的決定問題¹²という.

- (4) 決定空間 $\mathbb{D}$ は $\mathbb{X}^n$ から $\mathbb{A}$ への可測関数全体とすることもできる. しかし目標は危険関数の Θ に関する様な評価であるので, $\mathbb{D}$ には可測性以外の合理的な制限¹³を設けるのが一般的である.

点推定問題

真の母数 θ^* を観測データ $\mathbf{X}$ に基づいて 1 点で回復するのが点推定である. したがって $\mathbb{A} = \Theta$ となる. 点推定の場合には決定関数 $d(\mathbf{X})$ を推定量といい, 観測データの実現値 $\mathbf{X} = \mathbf{x}$ における推定量の値 $d(\mathbf{x})$ を推定値という. $\Theta = \mathbb{R}$ ならば損失関数として

$$L(a, \theta) = (a - \theta)^2, \quad L(a, \theta) = |a - \theta|$$

を取るのが代表的なアプローチである. 上記の損失関数それぞれに対応する危険関数

$$R(\theta, d) = E_\theta[L(\theta, d(\mathbf{X}))]$$

を平均 2 乗誤差と平均絶対誤差という. したがって平均 2 乗誤差を Θ に関してなんらかの意味で一様に評価することで考えている推定量の族 $\mathbb{D}$ の中から「最適」な推定量ないしは合理的な観点から正当化される推定量をみつきたいわけである.

検定問題

母数空間を 2 つの排反な部分集合に分ける. すなわち

$$\Theta = \Theta_0 \cup \Theta_1, \quad \Theta_0 \cap \Theta_1 = \emptyset$$

である. 行動は真の母数 θ^* が Θ_0 に属するか, Θ_1 に属するかを判断する. したがって行動空間は $\mathbb{A} = \{0, 1\}$ と書ける. すなわち, $\mathbb{A}$ の元 0 は

¹²本来であれば, どこで可測であるかを考える必要があるので,

$$\left((\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n)), \{P_\theta; \theta \in \Theta\}, (\mathbb{A}, \mathcal{B}(\mathbb{A})), (\mathbb{D}, \mathcal{B}(\mathbb{D})), L \right)$$

と書くべきである.

¹³合理的な制限の概念として不変性や不偏性などがある. また尤度に基づく方法に限定するといった考え方もある.

$\theta^* \in \Theta_0$ であると判断したことを表し, 1 は $\theta^* \in \Theta_1$ であると判断したことを表す. 決定関数 $d: \mathbb{X}^n \rightarrow \mathbb{A}$ は標本空間 $\mathbb{X}^n$ の部分集合 W に対して

$$d(\mathbf{x}) = \begin{cases} 1 & (\mathbf{x} \in W) \\ 0 & (\mathbf{x} \notin W) \end{cases}$$

で定めること¹⁴ができる. 検定問題では d のことを 検定関数 という. 損失関数としては

$$L(\theta, 0) = \begin{cases} 0 & (\theta \in \Theta_0) \\ 1 & (\theta \in \Theta_1) \end{cases}; \quad L(\theta, 1) = \begin{cases} 1 & (\theta \in \Theta_0) \\ 0 & (\theta \in \Theta_1) \end{cases}$$

と取る. すなわち

	$\theta \in \Theta_0$	$\theta \in \Theta_1$
$d = 0$	0	1
$d = 1$	1	0

である.

通常 $H_0: \theta \in \Theta_0$ のことを 帰無仮説 とよぶ. さらに, $d(\mathbf{x})$ に形式的に $\mathbf{X}$ を代入した $d(\mathbf{X})$ を観測データに基づく検定の決定を表す. 簡単に, これを 検定統計量 という. $H_1: \theta \in \Theta_1$ のことを 対立仮説 という. 危険関数 $R(\theta, d) = E_\theta[L(\theta, d(\mathbf{X}))]$ は以下ようになる. ここで第 1 種の誤

	$\theta \in \Theta_0$	$\theta \in \Theta_1$
$d = 0$	正しい判断	第 2 種の誤り
$d = 1$	第 1 種の誤り	正しい判断

りの確率と第 2 種の誤りの確率はトレード・オフの関係になっていることが鍵である. すなわち, 同時には二つの確率を小さくできない. 実は

$$(\text{第 1 種の誤りの確率}) + (\text{第 2 種の誤りの確率}) \geq \text{下限}$$

ということになっているのである¹⁵. そこで $\theta \in \Theta_0$ のとき

$$\beta(\theta) := R(\theta, d)$$

とし, $\theta \in \Theta_1$ のとき

$$\beta(\theta) := 1 - R(\theta, d)$$

¹⁴正確には確率化決定関数を考える必要があるが, 議論を簡単にするためにこれは考えないことにする.

¹⁵このことは第 7 章で説明する Neyman-Pearson の補題からわかる.

と定義したものを 検出力関数 という. 仮説検定では与えられた数 α ($0 < \alpha < 1$) に対して

$$\sup_{\theta \in \theta_0} R(\theta, d) \leq \alpha$$

をみたす検定関数の中から $\theta \in \Theta_1$ において $\beta(\theta)$ を大きくするもの, すなわち $R(\theta, d)$ を小さくするものと選ぶことを目指す. ちなみに α のことを 有意水準 という. $\sup_{\theta \in \Theta_0} R(\theta, d)$ を検定関数 d の サイズ という. したがってサイズが有意水準より小さい検定関数の中から $\theta \in \Theta_1$ において検出力関数の値が一様に大きなもの¹⁶を探したいわけである.

区間推定

議論を簡単にするために $\Theta = \mathbb{R}$ とする. 区間推定において行動は $\mathbb{R}$ の区間となる. したがって決定空間は観測データから区間への対応となる. 観測データの実現値 $\mathbf{X} = \mathbf{x}$ に基づく母数 θ の推定区間 $[\ell(\mathbf{x}), u(\mathbf{x})]$ に対して損失関数として

$$L(\theta, [\ell, u]) = (u - \ell) - \mathbb{1}\{\theta \in [\ell, u]\}$$

などが考えられる. ただし, $\mathbb{1}\{\theta \in [\ell, u]\} = \begin{cases} 1 & (\theta \in [\ell, u]) \\ 0 & (\theta \notin [\ell, u]) \end{cases}$ である. この場合には, L は負の値を取ることもある. 決定関数

$$d(\mathbf{X}) = [\ell(\mathbf{X}), u(\mathbf{X})]$$

に対して危険関数は

$$R(\theta, d) = E_{\theta}[u(\mathbf{X}) - \ell(\mathbf{X})] - \Pr_{\theta}(\theta \in [\ell(\mathbf{X}), u(\mathbf{X})])$$

となる. ただし, $\forall \mathcal{B}(\mathbb{X}^n)$ と $\mathbf{X} \sim P^{\otimes n}$ に対して, $\Pr_{\theta}(\mathbf{X} \in B) := E_{\theta}[\mathbb{1}_B(\mathbf{X})]$ で定義した.

実数 α ($0 < \alpha < 1$) が与えられたとき

$$\Pr_{\theta}(\theta \in [\ell(\mathbf{X}), u(\mathbf{X})]) \geq 1 - \alpha$$

のもとで区間の長さの期待値 $E_{\theta}[u(\mathbf{X}) - \ell(\mathbf{X})]$ を短くする区間が望ましい区間といえよう. $1 - \alpha$ を 信頼係数 とよぶ.

以上のように統計的決定問題の枠組みで統計的推測の問題である点推定, 区間推定, および検定を統一的に扱うことができる.

次に決定空間の元の間順序 $\prec$ を導入しよう. 決定関数 $\forall d_1, d_2 \in \mathbb{D}$ に対して

$$d_1 \prec d_2$$

¹⁶これは究極の目標であり, 一様に検出力関数の値を最大にする検定統計量は存在しないかもしれない.

$$\Leftrightarrow R(\theta, d_1) \leq R(\theta, d_2) (\forall \theta \in \Theta) \text{ かつ } R(\theta_0, d_1) < R(\theta_0, d_2) (\exists \theta_0 \in \Theta)$$

定める. すると決定空間 $\mathbb{D}$ を標本空間 $\mathbb{X}$ から行動空間 $\mathbb{A}$ への可測関数すべてから成る集合とすれば順序 $\prec$ は半順序になる. すなわち順序 $\prec$ の意味で一番よいものは存在しない.

たとえば $X \sim N(\mu, 1)$ によって μ を推定する問題を損失関数 $L(\mu, \hat{\mu}) = (\hat{\mu} - \mu)^2$ のもとで考える. ただし $\hat{\mu}$ は μ の推定量である. このとき

$$\hat{\mu}_0(X) = 0$$

なる推定量より, 母数 ν に関して一様によい推定量は存在しない. なぜならば $\mu = 0$ において $\hat{\mu}_0$ の危険関数の値は 0 となるので, $\hat{\mu}_0$ より一様によい推定量は存在しないこと¹⁷がわかる.

最良の決定関数が存在しない場合には決定関数を比較するための別の観点の導入が必要となる. 主なもので次の二つがある.

- (1) 決定関数の最適性について別の概念を導入する. 代表的なものとしてミニマックス基準と Bayes 基準がある.
- (2) 考察する決定関数を制限し, その中で危険関数を母数 Θ に関して一様に小さくする決定関数をみつける. たとえば不偏性, 不変性などを導入して, 考察する決定関数を制限する方法がある. また Neyman-Pearson の補題による議論がある.

さらに決定空間のなかからよい決定関数をみつけるのではなく, 一定の原理によって導かれる決定関数を考えて, それについてなんらかの合理性を証明する方針がある. 統計的決定問題の枠組みからははずれるが, ある原理に基づきなんらかのかたちで合理的な正当化ができる決定関数を導出することが考えられてきた. 導出の原理として推定ではモーメント法, 最尤法 (第 7 章) などが知られている. 検定法では尤度比検定, スコア検定, Wald 検定, Rao 検定 (第 8 章) がある. 区間推定では検定統計量の反転, ピボット法 (第 8 章) などがある.

5.3 章末注釈と参考文献

この章の話題を数学的に厳密に議論をするためには, 数学の進んだ道具立てが必要となるので, 直観的な説明に留めた.

¹⁷ある推定量に対して, それより一様によい推定量が存在しないとき, その推定量を許容的という.

5.4 演習問題

演習問題 5.1. $0 < \theta < 1$ とし, $X \sim \text{Bino}(100, \theta)$ とする. 母数 θ に対する損失関数 $L(\theta, d) = (\theta - d)^2$ を考える. さらに, θ の推定量

$$d_0(X) = \frac{X}{100}, \quad d_1(X) = \frac{X+3}{100}, \quad d_2(X) = \frac{X+3}{106}$$

を考える.

- (1) 推定量 d_0 の危険関数 $R(\theta, d_0)$ を求めよ.
- (2) 推定量 d_1 の危険関数 $R(\theta, d_1)$ を求めよ.
- (3) 推定量 d_2 の危険関数 $R(\theta, d_2)$ を求めよ.
- (4) 危険関数 $R(\theta, d_0)$ と $R(\theta, d_1)$ の大小を比較せよ.
- (5) 危険関数 $R(1/2, d_0)$ と $R(1/2, d_2)$ および $R(1/8, d_0)$ と $R(1/8, d_2)$ の大小を比較せよ.