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.

5.1　統計的実験と母数モデル

$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上で定義された $\mathbb{X}$ 値確率要素[1]とする. ただし $\mathbb{X}$ は距離空間とする. 典型的な例は $\mathbb{X} = \mathbb{R}^n$ 等である. 可測空間を $\big(\mathbb{X}, \mathcal{B}(\mathbb{X})\big)$ と書くことにする. だたし $\mathcal{B}(\mathbb{X})$ は $\mathbb{X}$ 上の Borel 集合族[2]である.

この講義では $X_1, X_2, \ldots, X_n$ は独立同一の分布に従うものとする. $X_1, X_2, \ldots, X_n$ を**ランダム標本**, $\mathbb{X}$ を**標本空間**と呼ぶことにする. $\boldsymbol{X}$ の測度空間 $\big(\mathbb{X}, \mathcal{B}(\mathbb{X})\big)$ 上の確率測度 $\mathsf{P}^{\otimes n}$ を

$$\mathsf{P}^{\otimes n}(B) := \mathsf{Pr}(\boldsymbol{X} \in B) \qquad (B \in \mathcal{B}(\mathbb{X}))$$

で定義する. 確率測度 $\mathsf{P}^{\otimes n}$ を $\boldsymbol{X}$ の**分布**といい, $\boldsymbol{X} \sim \mathsf{P}^{\otimes n}$ と表記することにする. また $\mathsf{P}^{\otimes n} = \mathsf{Pr} \circ \boldsymbol{X}^{-1}$ とも表現することがある. 各 $X_j\,(j = 1, 2, \ldots, n)$ の周辺確率測度を P と書くことにする. すると

$$\mathsf{P}^{\otimes n} = \underbrace{\mathsf{P} \times \mathsf{P} \times \cdots \times \mathsf{P}}_{n\ \text{個}}, \qquad \mathsf{P} = \mathsf{Pr} \circ X_1^{-1}$$

と書ける.

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

$$\begin{aligned}\mathsf{Pr}(\boldsymbol{X} \in B) &= \mathsf{Pr}\big(\{\omega \in \Omega;\, (X_1(\omega), X_2(\omega), \ldots, X_n(\omega)) \in B\}\big) \\ &= \mathsf{Pr}\big(\{\omega \in \Omega;\, \boldsymbol{X}(\omega) \in B\}\big)\end{aligned}$$

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測度 $\mathbb{P}$ は既知であり, 確率要素 $\boldsymbol{X}$ の分布論的な特徴を調べる. 一方

[1] 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.
[2] 開集合族を含む最小の σ 加法族

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 $\boldsymbol{X}$ 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. $\boldsymbol{X} \sim \mathsf{P}^{\otimes n}$ とする. 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測度 $\mathsf{P}^{\otimes n}$ (これを**真の確率測度**[3]ということにする) を観測データから回復するために, 候補となる $\big(\mathbb{X}, \mathcal{B}(\mathbb{X})\big)$ 上の確率測度の集まり $\mathcal{P}$ を設定する. この $\mathcal{P}$ を**統計的モデル**という. そして真の確率測度 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 で添え字付けれると仮定する. すなわち写像

$$\Theta \ni \theta \mapsto \mathsf{P}_\theta \in \mathcal{P}$$

を想定する. 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. そして真の確率測度 P に対応する母数を**真の母数**といい, θ^* と書くことにする. すなわち

$$\mathsf{P} = \mathsf{P}_{\theta^*}$$

である. もちろん $\theta^* \in \Theta$ ではあるが, 真の母数 θ^* は観測者には未知のものである. 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 $\boldsymbol{X}$ から回復することが目的[4]となる.

次に統計的モデルのなかでも最も基本的な母数モデルを定義する.

定義 5.2. 統計的モデル
$\mathcal{P} = \{\mathsf{P}_\theta; \theta \in \Theta\}$ が**母数モデル** 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 $\mathbb{R}^k$ の「よい」部分集合である. ただし $k \in \mathbb{N}$ である.

(2) 写像
$$\Theta \ni \theta \mapsto \mathsf{P}_\theta \in \mathcal{P}$$
は「滑らか」である[5].

[3]誤解のおそれがないときには, 真のモデルを P と書くことにする.

[4]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 $\widetilde{\boldsymbol{X}}$ を観測データ $\boldsymbol{X}$ 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.

[5]
$\mathcal{P}$ 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. このあたりの議論は Tojo and Yoshino (2021) を参照のこと.

この条件を母数化の正則性という. したがって条件 (1)(2) をみたすものを正則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) $\theta_1, \theta_2 \in \Theta$ に対して

$$\theta_1 \neq \theta_2 \Longrightarrow \mathsf{P}_{\theta_1} \neq \mathsf{P}_{\theta_2}.$$

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**であるという.

注意 5.3. 定義 5.2 は数学的にはすぼらな表現である. 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.

$X_1, X_2, \ldots, X_n$ は独立同一分布に従っているので, X_1 の確率測度の族 $\{\mathsf{P}_\theta; \theta \in \Theta\}$ を母数モデルと同一視する. □

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」と統計学では呼んでいる. 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, 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人は「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[6]」という. これも言葉の意味のしては奇異であるが, 統計学の習慣に従うことにする. 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・モデルの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

$$\mathcal{P} = \{\mathsf{P}_\theta; \theta \in \Theta, \mathsf{P}_\theta \text{ は } (\mathbb{X}, \mathcal{B}(\mathbb{X})) \text{ 上の確率測度 }\}$$

[6]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.

の組を**統計的実験**といい,

$$\big(\mathbb{X},\, \mathcal{B}(\mathbb{X}),\, \{\mathsf{P}_\theta;\, \theta \in \Theta\}\big)$$

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 $\theta^* \in \Theta$ を書くことにする. θ^* を**真の母数**という. すなわち,

$$\mathsf{P} = \mathsf{P}_{\theta^*}$$

である.

通常 $\boldsymbol{X}$ は $\mathbb{R}^n$ に値をとる確率ベクトルとすることが多い. また

$$\mathsf{P} := \mathsf{Pr} \circ X_1^{-1}, \quad \mathsf{P}_{\theta^*} := \mathsf{Pr} \circ X_1^{-1}$$

と書くことにする. P のことを**真の分布**といい, $\{\mathsf{P}_\theta;\, \theta \in \Theta\}$ もだらしなく母数モデルということにする.

この講義では $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ と書いたとき, $X_1, X_2, \ldots, X_n$ 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 $\boldsymbol{X}$ 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にある.

この講義で仮定する統計的実験

(1) 観測データを $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ と書き, 可測空間 $\big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)\big)$ に値をとる.

(2) 観測データは真の分布 P からのランダム標本である. すなわち

$$X_1, X_2, \ldots, X_n \sim \text{i.i.d.}\, P \quad \text{もしくは} \quad \boldsymbol{X} \sim \mathsf{P}^{\otimes n}$$

である. ただし P は $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ または $(\mathcal{N}, 2^{\mathcal{N}})$ 上の確率測度である. ここで $\mathcal{N}$ は高々可算の集合とした.

(3) 統計的モデル
$\mathcal{P} = \big\{P_\theta;\, \theta \in \Theta\big\}$ $\big($もしくは$\mathcal{P} = \big\{\mathsf{P}_\theta;\, \theta \in \Theta\big\}\big)$ を設定する. ただし P_θ も P と同じ可測空間上の確率測度である.

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

(5) $\theta \ni \theta$ から $\mathsf{P}_\theta \in \mathcal{P}$ 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 $\theta^* \in \Theta$ があって $\mathsf{P} = \mathsf{P}_{\theta^*}$.

これらをまとめて統計的実験といい

$$\big(\mathbb{R},\, \mathcal{B}(\mathbb{R}),\, \{\mathsf{P}_\theta;\, \theta \in \Theta\}\big) \text{ もしくは } \big(\mathcal{N},\, 2^{\mathcal{N}},\, \{\mathsf{P}_\theta;\, \theta \in \Theta\}\big)$$

と書き,
$\mathcal{P}$ を正則母数モデルという. 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分布 P からのランダム標本 $(X_1,\, X_2,\, \ldots,\, X_n)$ に基づき θ^* を回復することである.

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

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

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) $X_1,\, X_2,\, \ldots,\, X_n$ は Bernoulli 分布 $\mathsf{Ber}(\theta)\,(0 < \theta < 1)$ 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

$$\left(\{0,\, 1\}^n,\, 2^{\{0,1\}^n},\, \big\{\mathsf{p}(x|\,\theta) = \theta^x(1-\theta)^{1-x}\,(x = 0,\, 1)\big\};\, \theta \in \Theta = (0,\, 1)\big\}\right)$$

を想定していることになる. □

注意 5.7. $X_1,\, X_2,\, \ldots,\, X_n$ は分布 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, 上記のような統計的実験を仮定したことになる. 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. 統計的実験を考えるなどとは言わないことにする. □

5.2　統計的決定問題

統計的推測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のものである. 以下では頻度論的推測論の枠組みを説明することにする. Bayes 論的推測論は第 **??** 章で説明する.

頻度論的推測論の枠組みを統計的決定理論[7]の言葉を使って説明する.

標本空間を $\mathbb{R}^n$ とし, 観測データを $\boldsymbol{X}$ とする.

[7]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.

(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,

$$d : \mathbb{R}^n \ni \boldsymbol{x} \mapsto d(\boldsymbol{x})$$

で記す. 行動関数の集まりを行動空間といい, $\mathbb{D}$ 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 $\mathbb{A} \times \Theta$ 上の非負値実数値関数

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

を用意する[8]. この関数を**損失関数**といい, $\mathsf{L}(a, \theta)$ の値が小さいほど望ましい行動であるとする. 決定関数の「よさ」を評価するには観測データの実現値 $\boldsymbol{X} = \boldsymbol{x}$ と真の母数 θ^* における損失関数の値 $\mathsf{L}\big(d(\boldsymbol{x}), \theta^*\big)$ がわかればよい. したがって決定関数 d の「よさ」の評価に $\mathsf{L}\big(d(\boldsymbol{X}), \theta^*\big)$ 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 θ^* がわからないと知ることができない量であるからである. そこで母数 θ に対して観測データが P_θ によって生成されたと仮定し, 損失関数 $\mathsf{L}\big(d(\boldsymbol{X}), \theta\big)$ を P_θ に関して期待値を取ったもの

$$\mathsf{R}(d, \theta) := \mathsf{E}_\theta\big[\mathsf{L}\big(d(\boldsymbol{X}), \theta\big)\big]$$

を考える. これを決定関数 d の母数 θ に対する**危険関数**という.

(3) 決定関数 d の「よさ」の評価は真の母数 θ^* のもとで行たいところである. しかしこれは未知である. 危険関数の $\theta \in \Theta$ に関するなんらかの一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一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

$$\big(\mathbb{R}^n, \{\mathsf{P}_\theta; \theta \in \Theta\}, \mathbb{A}, \mathbb{D}, \mathsf{L}\big) \quad \text{もしくは} \quad \big(\mathbb{R}^n, \{\mathsf{P}_\theta; \theta \in \Theta\}, \mathbb{A}, \mathbb{D}, \mathsf{L}\big)$$

を統計的決定問題[9]という.

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

[9] 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,

$$\Big(\big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)\big), \{\mathsf{P}_\theta; \theta \in \Theta\}, \big(\mathbb{A}, \mathcal{B}(\mathbb{A})\big), \big(\mathbb{D}, \mathcal{B}(\mathbb{D})\big), \mathsf{L}\Big)$$

と書くべきである.

(4) 決定空間 $\mathbb{D}$ は $\mathbb{R}^n$ 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 に関する一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[10]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 $\boldsymbol{X}$ 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 $d(\boldsymbol{X})$ を**推定量**といい, 観測データの実現値 $\boldsymbol{X} = \boldsymbol{x}$ における推定量の値 $d(\boldsymbol{x})$ を**推定値**という. $\Theta = \mathbb{R}$ ならば損失関数として

$$\mathsf{L}\big(a, \theta\big) = \big(a - \theta\big)^2, \quad \mathsf{L}\big(a, \theta\big) = \big|a - \theta\big|$$

を取るのが代表なアプローチである. 前者の損失関数に対応する危険関数

$$\mathsf{R}(d, \theta) = \mathsf{E}_\theta\big[\mathsf{L}\big(d(\boldsymbol{X}), \theta\big)\big]$$

を**平均 2 乗誤差** 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けたいわけである.

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

$$\Theta = \Theta_0 \cup \Theta_1, \quad \Theta_0 \cap \Theta_1 = \varnothing.$$

行動は真の母数 θ^* が Θ_0 に属するか, Θ_1 の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. 決定関数は標本空間 $\mathbb{X}$ の部分集合 R に対して,

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

で定めること[11]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

$$\mathsf{L}\big(0, \theta\big) = \begin{cases} 0 & (\theta \in \Theta_0) \\ 1 & (\theta \in \Theta_1) \end{cases} \qquad \mathsf{L}\big(1, \theta\big) = \begin{cases} 1 & (\theta \in \Theta_0) \\ 0 & (\theta \in \Theta_1) \end{cases}$$

と取る.

通常 $H_0 : \theta \in \Theta_0$ のことを**帰無仮説**とよぶ. さらに, $d(\boldsymbol{x})$ に形式的に $\boldsymbol{X}$ を代入した $d(\boldsymbol{X})$ 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 $\mathsf{R}\big(d, \theta\big) = \mathsf{E}_\theta\big[\mathsf{L}\big(d(\boldsymbol{X}), \theta\big)\big]$ は以下のようになる. こ

[10] 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.

[11] 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.

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

	$\theta \in \Theta_0$	$\theta \in \Theta_1$
$d = 0$	0	第 2 種の誤り
$d = 1$	第 1 種の誤り	0

こで第 1 種の誤りの確率と第 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. 同時には二つの確率を小さくできない. 実は

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

ということになっているのである[12]. そこで $\theta \in \Theta_0$ のとき

$$\beta(\theta) := \mathsf{R}\big(d, \theta\big)$$

$\theta \in \Theta_1$ のとき

$$\beta(\theta) := 1 - \mathsf{R}\big(d, \theta\big)$$

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 $\alpha\,(0 < \alpha < 1)$ に対して

$$\sup_{\theta \in \Theta_0} \mathsf{R}\big(d, \theta\big) \leq \alpha$$

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 $\mathsf{R}\big(d, \theta\big)$ 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. $\sup_{\theta \in \Theta_0} \mathsf{R}\big(d, \theta\big)$ 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 $\theta \in \Theta_1$ のおける検出力関数の値の大きなもの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. したがって行動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 $\boldsymbol{X} = \boldsymbol{x}$ に基づく母数 θ の推定区間 $[\ell(\boldsymbol{x}), u(\boldsymbol{x})]$ に対して損失関数として

$$\mathsf{L}\big([\ell, u], \theta\big) = (d_2 - \ell) - \mathbb{1}\big\{\theta \in [\ell, u]\big\}$$

などが考えられる. この場合には, L は負の値を取ることもある. 決定関数

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

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

に対して危険関数は

$$\mathsf{R}\big(d,\,\theta\big) = \mathsf{E}_\theta\big[u(\boldsymbol{X}) - \ell(\boldsymbol{X})\big] - \mathsf{Pr}_\theta\big(\theta \in [\ell(\boldsymbol{X}),\, u(\boldsymbol{X})]\big)$$

となる.

実数 $\alpha\,(0 < \alpha < 1)$ が与えられたとき

$$\mathsf{Pr}_\theta\big(\theta \in [\ell(\boldsymbol{X}),\, u(\boldsymbol{X})]\big) \geq 1 - \alpha$$

のもとで区間の長さの期待値 $\mathsf{E}_\theta\big[u(\boldsymbol{X}) - \ell(\boldsymbol{X})\big]$ を短くする区間が望ましい区間といえよう. α 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, および検定が統一的に扱うことができる.

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

$$\begin{aligned} &d_1 \prec d_2 \\ &\quad\Longleftrightarrow\ \mathsf{R}(d_1,\,\theta) \leq \mathsf{R}(d_2,\,\theta)\,(\forall\theta \in \Theta)\ \text{かつ}\ \mathsf{R}(d_1,\,\theta_0) < \mathsf{R}(d_2,\,\theta_0)\,(\exists\theta_0 \in \Theta) \end{aligned}$$

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

$$\widehat{\mu}_0 = 0$$

なる推定量は許容的になる. なぜならば $\mu = 0$ において $\widehat{\mu}_0$ の危険関数の値は 0 となるので, $\widehat{\mu}_0$ よりよい推定量は存在しないわけである.

最小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. おも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, 一定の原理の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

5.4　演習問題