立教大学　数理統計学序論 1 ・数理統計学 1

今野 良彦

大阪公立大学

2025 年 11 月 08 日

講義内容

- 記号の復習
- 5.2 統計的実験と母数モデル
- 5.2 統計的決定問題
- 6.1 正則母数モデルと Fisher 情報量
- 6.2 指数型分布族
- 6.3 指数型分布族の正準表示

- $(\Omega, \mathcal{A}, \mathsf{Pr})$: 確率空間.
  - $\mathcal{A}$ は σ 加法族.
  - $\mathsf{Pr} : \mathcal{A} \longrightarrow [0, 1]$ は確率測度.
- X は確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率変数 $\Leftrightarrow \forall r \in \mathbb{R}$ に対して

$$X^{-1}((r, \infty)) = \{\omega \in \Omega;\ X(\omega) > r\}.$$

- 確率変数 X の分布 P^X

$$\mathsf{P}^X(B) = \mathsf{Pr}(\omega \in \Omega;\ X(\omega) \in B\}) \qquad (\forall B \in \mathcal{B}(\mathbb{R}))$$

ここで, $\mathcal{B}(\mathbb{R})$ は Borel 集合族. すなわち, $\mathbb{R}$ の開集合族を含む最小の σ 加法族.

- 確率変数 X の累積分布関数 F^X: $x \in \mathbb{R}$ に対して

$$\mathsf{F}^X(x) = \mathsf{P}^X((-\infty, x])] = \mathsf{Pr}(\omega \in \Omega;\ X(\omega) \le x).$$

- 累積分布関数 F^X の性質: ① 右連続, ② 非減少, ③ $\lim_{x\to\infty} \mathsf{F}^X(x) = 1,\ \lim_{x\to-\infty} \mathsf{F}^X(x) = 0.$
  - F^X が階段関数 $\Rightarrow$ X は離散型確率変数.
  - F^X が連続関数 $\Rightarrow$ X は連続型確率変数.

記号の復習 (2)

- 確率変数 X は離散型のとき

$$\mathsf{p}^X(x) = \mathsf{F}^X(x) - \mathsf{F}^X(x-)$$

を確率関数 (p.m.f.) という.

- 確率変数 X は連続型のとき

$$\mathsf{F}^X = \int_{-\infty}^{x} \mathsf{p}^X(t)\,\mathrm{d}t$$

をみたす非負関数 p^X(存在するならば) を確率密度関数 (p.d.f.) という.

- 確率変数 X の期待値を

$$\mathsf{E}[X] = \begin{cases} \sum_x x\mathsf{p}^X(x) & \text{(離散型の場合)} \\ \int_{\mathbb{R}} x\mathsf{p}^X(x)\,\mathrm{d}x & \text{(連続型の場合)} \end{cases}$$

ただし, 和および積分の中を絶対値をとったものが有限ののとき.

記号の復習 (3)

- 確率変数 X の分散: $\mathsf{E}[X^2] < \infty$ のとき

$$\mathsf{Var}[X] = \mathsf{E}[\{X - \mathsf{E}[X]\}^2]$$

- $\mathsf{F}^X(\cdot)$ が階段関数のとき, X を離散型確率変数という.
- 確率変数 X の積率母関数 (m.g.f.): 原点を含むある近傍 $(-z_0, z_0)$ があって, $\forall t \in (-z_0, z_0)$ に対して

$$\mathsf{m}^X(t) = \mathsf{E}[e^{tX}]$$

5.2 統計的実験と母数モデル (1)

- $X_1, X_2, \ldots, X_n$ を確率空間 $(\Omega, \mathcal{A}, \mathrm{Pr})$ 上の確率変数列とする. すなわち, $X_i : \Omega \longrightarrow \mathbb{R}\,(i = 1, 2, \ldots, n)$ で, $\forall r \in \mathbb{R}$ に対して

$$X_i^{-1}((r, \infty)) = \{\omega \in \Omega;\ X_i(\omega) > r\}$$

をみたす.
- $X = (X_1, X_2, \ldots, X_n)^\top$ と記す. ただし, $(\cdot)^\top$ は $(\cdot)$ の転置を表す.

仮定

- 以後では, $X_1, X_2, \ldots, X_n$ は独立同一分布に従うと仮定する.
- このような場合, $X_1, X_2, \ldots, X_n$ を「標本の大きさが n のランダム標本」ということにする.
- 確率ベクトル X の値域 $\mathbb{R}^n$ を標本空間という.
- $\mathcal{B}(\mathbb{R})$ を $\mathbb{R}$ 上の Borel 集合族 ($\mathbb{R}$ 上の開集合族を含む最小の σ 加法族) とする.
- 確率変数 X_1 で誘導される可測空間 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上の確率測度 P を

$$\mathsf{P}(B) := \mathrm{Pr}(X \in B) \qquad (\forall B \in \mathcal{B}(\mathbb{R}))$$

  で定める.
- 確率測度 P を $\underline{X_1 \text{ の分布}}$ といい

$$X_1 \sim \mathsf{P}$$

  と記す.
- 確率ベクトル X の分布を $\mathsf{P}^{\otimes n}$ または P^X と記す.

仮定の続き

- すなわち, $\forall B_j \in \mathcal{B}(\mathbb{R})\ (j = 1, 2, \ldots, n)$ に対して

$$\mathrm{P}^{\otimes n}(B_1 \times B_2 \times \cdots \times B_n) = \mathrm{Pr}(X \in B_1 \times B_2 \times \cdots \times B_n) = \mathrm{Pr}\left(\bigcap_{j=1}^{n}\{X_j \in B_j\}\right)$$

となる. すると

$$\mathrm{P}^{\otimes n}(B_1 \times B_2 \times \cdots \times B_n) = \prod_{j=1}^{n} \mathrm{P}(B_j)$$

となる.

注意

- 確率論: P^X を 既知 として, X ないしはそれの関数の分布論的特徴 (Stochastic properties) を調べる.
- 推測（数理）統計学: P^X を 未知 として, X の情報（実現値）から未知 P^X を回復（推測）する.
- 推測統計学の枠組み: P^* を測度空間 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上の 未知 の分布として

$$X_1, X_2, \ldots, X_n \sim \mathsf{P}^* \Leftrightarrow X \sim (\mathsf{P}^*)^{\otimes n}$$

とする.
- P^* を「真の分布」ということもある. すなわち, データ生成メカニズム.
- X の情報（観測値）に基づき, 確率分布全体の集合 $\mathcal{M}$ から未知の分布（真の分布）P^* を回復する問題は「不適切逆問題 (ill-posed inverse problem)」となることが知られている.

5.2 統計的実験と母数モデル (5)

Insert a picture!

不適切逆問題のひとつの対応: 統計的モデル

- 探索の範囲を確率分布全体の集合 $\mathcal{M}$ から狭い集合に限定する!
- $\mathcal{M}$ の都合のよい部分集合を設定して, そこに制限した探索を行う. この $\mathcal{M}$ の部分集合で都合のよいように設定したものを 統計的モデル ということにする.
- 設定した統計的モデルを $\mathcal{P} \subset \mathcal{M}$ と記すことにする.

仮定 この講義では, 想定（設定）した統計的モデル $\mathcal{P}$ は真の分布 $\mathbf{P}^*$ を含んでいるとする. すなわち

$$\mathbf{P}^* \in \mathcal{P}.$$

母数化: 統計的モデル $\mathcal{P}$ の要素（分布）のレベリングをする集合 $\boldsymbol{\Theta} \subset \mathbb{R}^k$ $(\exists k \in \mathbb{N})$ を想定する.

$$\boldsymbol{\Theta} \ni \theta \longleftrightarrow \mathbf{P}_\theta \in \mathcal{P}.$$

すなわち, 統計的モデル $\mathcal{P}$ と $\mathbb{R}^k$ の部分集合間に対応関係を想定する. $\boldsymbol{\Theta}$ を統計的モデルの 母数空間 といい, $\boldsymbol{\Theta}$ の元を 母数 ということにする.

仮定

- 以下では, 対応関係

$$\boldsymbol{\Theta} \ni \theta \mapsto \mathsf{P}_\theta \in \mathcal{P} \tag{1}$$

は母数空間 $\boldsymbol{\Theta}$ から統計的モデル $\mathcal{P}$ への全単射と仮定する.

- すると $\mathsf{P}^* \in \mathcal{P}$ と仮定していたので, ある $\exists\theta^* \in \boldsymbol{\Theta}$ があって

$$\mathsf{P}_{\theta^*} = \mathsf{P}^* \Leftrightarrow \mathsf{P}_{\theta^*}(B) = \mathsf{P}(B) \qquad (\forall B \in \mathcal{B}(\mathbb{R}))$$

となる.

- 真の分布 P^* に対応する母数 θ^* を「真の母数」という. この講義の仮定から, $\theta^* \in \boldsymbol{\Theta}$ となっていることに注意する.
- 写像 (1) を設定することを統計的モデルの「母数化」という.
- 母数化（写像と母数空間の設定）は一意ではないこと (便宜的である) に注意する.

5.2 統計的実験と母数モデル (8)

記号のまとめ

$\mathcal{P}$	統計的モデル
$\boldsymbol{\Theta}$	母数空間
$\mathbf{P}^* \in \mathcal{P}$	真の分布
$\boldsymbol{\theta}^* \in \boldsymbol{\Theta}$	真の母数

注意 「モデル」という言葉には注意をすること．この講義では「モデル」という言葉を単独では用いないようにする．

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

(2) 写像

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

は「滑らか」である [1].

[1] $\mathcal{P}$ 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.

$\boxed{\text{定義 5.2 の続き}}$ 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) $\forall \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \boldsymbol{\Theta}$ に対して

$$\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2 \Rightarrow \mathsf{P}_{\boldsymbol{\theta}_1} \neq \mathsf{P}_{\boldsymbol{\theta}_2}$$

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であるという.
$\boxed{\text{注意 5.3}}$ 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. □

注意 5.4 母数空間 $\boldsymbol{\Theta}$ 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. $\boldsymbol{\Theta}$ 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. 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[2]」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・モデルに対する統計的推測理論については Bickel *et al.* (1993), van der Vaart (1998), Kosorok (2007), 久保木・鈴木 (2015) を参照のこと. □

[2]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.

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

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

の組を統計的実験といい

$$\left(\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n), \{\mathsf{P}_\theta;\ \theta \in \Theta\}\right)$$

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 $\theta^* \in \Theta$ を書くことにする. θ^* を真の母数である. すなわち

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

である.

この講義では $X = (X_1, X_2, \ldots, X_n)$ 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 X のことを標本の大きさが n のランダム標本という.

この講義で仮定する統計的実験

(1) 観測データを $X = (X_1, X_2, \ldots, X_n)^\top$ 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

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

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 $\mathcal{P} = \{\mathsf{P}_\theta; \theta \in \Theta\}$ 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 $\mathsf{P}_\theta \in \mathcal{P}$ 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 $\theta^* \in \theta$ が唯一あって $\mathsf{P}^* = \mathsf{P}_{\theta^*}$ と書ける.

例 5.6 (1) $X_1, X_2, \ldots, X_n$ は正規分布 $\mathrm{N}(\mu, \sigma^2)$ 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

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

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 $\mathrm{Ber}(\theta)$ $(0 < \theta < 1)$ 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

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

を想定していることになる. □

注意 5.7 $X_1, X_2, \ldots, X_n$ 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,

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

のような統計的実験を仮定し, ある $\theta^* \in \Theta$ があって, $\mathbf{P}_{\theta^*} = \mathbf{P}^*$ 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 (1)

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 **9** 章で説明する. 以下では, 頻度論的推測論の枠組みを統計的決定理論[3]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 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

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

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.

[3] 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.

(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

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

を用意する[4]. この関数を損失関数といい, $\mathsf{L}(\theta, a)$ 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 $X = x$ と真の母数 θ^* における損失関数の値 $\mathsf{L}(\theta^*, d(x))$ がわかればよい. したがって決定関数 d の「よさ」の評価に $\mathsf{L}(\theta^*, d(X))$ 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 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 $\mathsf{P}_\theta^{\otimes n}$ によって生成されたと仮定し, 損失関数 $\mathsf{L}(\theta, d(X))$ を $\mathsf{P}_\theta^{\otimes n}$ に関して期待値を取ったもの

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

を考える.

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

ただし, $\mathbf{E}_\theta[\,\cdot\,]$ は $X \sim \mathsf{P}_\theta^{\otimes n}$ $(\theta \in \Theta)$ 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

$$(\mathbb{X}^n, \{\mathsf{P}_\theta;\ \theta \in \Theta\}, \mathbb{A}, \mathbb{D}, \mathsf{L})$$

を統計的決定問題[5]という.

[5]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,

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

と書くべきである.

5.2 統計的決定問題 (4)

(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 $\mathbf{\Theta}$ 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[6]を設けるのが一般的である.

[6]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.

点推定問題

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

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

を取るのが代表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

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

を平均 $\mathbf{2}$ 乗誤差 と 平均絶対誤差という. したがって平均 $\mathbf{2}$ 乗誤差を $\boldsymbol{\Theta}$ 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

$$\boldsymbol{\Theta} = \boldsymbol{\Theta}_0 \cup \boldsymbol{\Theta}_1, \quad \boldsymbol{\Theta}_0 \cap \boldsymbol{\Theta}_1 = \varnothing.$$

行動は真の母数 $\boldsymbol{\theta}^*$ が $\boldsymbol{\Theta}_0$ に属するか, $\boldsymbol{\Theta}_1$ 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 $\mathbb{X}^n$ の部分集合 W に対して

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

で定めること[7] ができる. 検定問題では d のことを検定関数という.

[7] 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.

検定問題の続き

損失関数としては

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

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

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, $d(x)$ に形式的に X を代入した $d(X)$ 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 $\mathrm{R}(\theta,\,d) = \mathrm{E}_\theta[\mathrm{L}(\theta,\,d(X))]$ は以下のようになる.

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

検定問題の続き

ここで第 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. すなわち, 同時には二つの確率を小さくできない. 実は

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

ということになっているのである[8]. そこで $\theta \in \Theta_0$ のとき

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

とし, $\theta \in \Theta_1$ のとき

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

と定義したものを検出力関数という.

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

検定問題の続き

仮説検定では与えられた数 $\alpha\ (0 < \alpha < 1)$ に対して

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

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 $\mathsf{R}(d, \theta)$ 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. $\sup_{\theta\in\Theta_0} \mathsf{R}(\theta, d)$ 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 [9] を探したいわけである.

[9] 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.

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 $X = x$ に基づく母数 θ の推定区間 $[\ell(x), u(x)]$ に対して損失関数として

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

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

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

に対して危険関数は

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

となる.

5.2 統計的決定問題 (11)

区間推定の続き

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

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

のもとで区間の長さの期待値 $\mathrm{E}_\theta[u(X) - \ell(X)]$ 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.

5.2 統計的決定問題 (12)

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

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

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.

5.2 統計的決定問題 (13)

たとえば $X \sim \mathrm{N}(\mu, 1)$ によって μ を推定する問題を損失関数 $\mathrm{L}(\widehat{\mu}, \mu) = (\widehat{\mu} - \mu)^2$ 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. 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 $\mathbf{\Theta}$ 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.

定義 6.1 $d, n \in \mathbb{N}$ とし, $\mathbb{X}^n$ ($\subset \mathbb{R}^n$) を標本空間[10] とする. $\mathbb{X}$ 上の母数モデル $\mathcal{P} = \{\mathsf{P}_\theta;\ \theta \in \Theta \subset \mathbb{R}^d\}$ が与えられているとする. 分布 P_θ の p.d.f.(または p.m.f.) $\mathsf{p}_\theta(x)$ と表記する. このとき母数モデルは正則であるとは次の条件をみたすときをいう.

(1) $\Theta \subset \mathbb{R}^d$ は開集合.

(2) $\mathcal{P}$ に属する分布の p.d.f.(または p.m.f.) は同じ台をもつ. すなわち, 集合 $\{x \in \mathbb{R};\ \mathsf{p}_\theta(x) > 0\}$ は $\theta \in \Theta$ に依存しない.

(3) $\forall\theta \in \Theta$ とする. $\mathsf{p}_\theta(x)$ の θ の 1 次と 2 次の偏導関数は $x (\in \mathbb{X})$ に関して連続である.

(4) $\mathsf{p}_\theta(x)$ の θ に関する 1 次と 2 次の偏導関数は $x (\in \mathbb{X})$ の関数として可積分である.

(5) $\mathsf{p}_\theta(x)$ の 1 次と 2 次の偏導関数は θ の微分記号と x の積分記号と交換が可能である.

[10] すこし言葉を乱用して, $\mathbb{X}$ のことも標本空間と呼ぶことにする.

記号の導入

$X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ を観測したとき Θ 上の実数値関数 ℓ_n を

$$\ell_n(\theta | x) = \log \prod_{j=1}^{n} \mathrm{p}_\theta(x_j) \qquad (x = (x_1, x_2, \ldots, x_n))$$

で定義[11]する.

$X \sim \mathrm{P}_\theta^{\otimes n}$ のときこの分布に関する期待値, 分散および共分散を

$$\mathrm{E}_\theta[\,\cdot\,], \qquad \mathrm{Var}_\theta[\,\cdot\,], \qquad \mathrm{Cov}_\theta[\,\cdot, \cdot\,]$$

と表記する.

[11] 対数尤度(ゆうど)関数と呼ぶ.

記号の導入の続き

すなわち, 可積分な関数 $h : \mathbb{X}^n \to \mathbb{R}$ に対して

$$\mathrm{E}_\theta[h(X)] := \begin{cases} \sum_{x \in \mathbb{X}^n} h(x) \prod_{j=1}^n \mathrm{p}_\theta(x_j) & (\text{離散型の場合}) \\ \int_{\mathbb{X}^n} h(x) \prod_{j=1}^n \mathrm{p}_\theta(x_j)\, \mathrm{d}x & (\text{連続型の場合}) \end{cases}$$

と定める. ただし $\mathrm{d}x = \mathrm{d}x_1 \mathrm{d}x_2 \times \cdots \times \mathrm{d}x_n$ である. さらに, X を確率空間 $(\Omega, \mathcal{A}, \mathrm{Pr})$ 上の確率ベクトルとしたとき, 任意の $B \in \mathcal{B}(\mathbb{X}^n)$ に対して, 可測空間 $(\Omega, \mathcal{A})$ 上の確率測度 Pr_θ を

$$\mathrm{Pr}_\theta(X^{-1}(B)) := \mathrm{Pr}_\theta(X \in B) := \mathrm{E}_\theta[\mathbb{1}_B(X)]$$

で定める. ただし, $X^{-1}(B) := \{\omega \in \Omega;\ X(\omega) \in B\} \in \mathcal{A}$ である.

記号の導入の続き

つぎに, $i, j = 1, 2, \ldots, d$ に対して

$$\dot{\ell}_{n,i} := \dot{\ell}_{n,i}(\theta | x) = \frac{\partial \ell_n(\theta | x)}{\partial \theta_i}, \quad \ddot{\ell}_{n,ij} := \ddot{\ell}_{n,ij}(x | \theta) = \frac{\partial^2 \ell_n(\theta | x)}{\partial \theta_j \partial \theta_k},$$

$$\dot{\ell}_n = (\dot{\ell}_{n,1}, \dot{\ell}_{n,2}, \ldots, \dot{\ell}_{n,d})^\top, \quad \ddot{\ell}_n = (\ddot{\ell}_{n,jk}), \quad \theta = (\theta_1, \theta_2, \ldots, \theta_d)^\top$$

と定義する. $X \sim \mathsf{P}_\theta^{\otimes n}$ のとき Fisher 情報量 $\mathcal{F}_X(\theta)$ は $d \times d$ 行列で $\mathcal{F}_X(\theta)$ の (i, j) 成分 $\{\mathcal{F}_X(\theta)\}_{jk}$ は

$$\{\mathcal{F}_X(\theta)\}_{ij} = \mathsf{E}_\theta[\dot{\ell}_{n,i}(X | \theta)\ \dot{\ell}_{n,j}(X | \theta)] \qquad (1 \leq i, j \leq d)$$

で定義される.

$\boxed{\text{定理 6.2}}$ 標本空間 $\mathbb{X}$ 上の母数モデル $\mathcal{P} = \{\mathbf{P}_\theta;\ \theta \in \Theta \subset \mathbb{R}^d\}$ は正則であるとする. $X \sim \mathbf{P}_\theta^{\otimes n}$ としたとき以下が成立する.

(1) $\forall \theta \in \Theta$ に対して $\mathbf{E}_\theta[\dot{\ell}_n(\theta | X)] = \mathbf{0}_d$ となる.

(2) $\mathcal{F}_X(\theta) = \mathbf{Cov}_\theta[\dot{\ell}_n(X | \theta)] = \mathbf{E}_\theta[\dot{\ell}_n(\theta | X)\,\dot{\ell}_n^\top(\theta | X)]$ が成り立つ.

(3) $\mathcal{F}_X(\theta) = -\mathbf{E}_\theta[\ddot{\ell}_n(\theta | X)]$ が成り立つ.

6.1 正則母数モデルと Fisher 情報量 (6)

定理 **6.2** の証明　証明は連続型分布の場合を示す. $\mathbf{P}_\theta$ の p.d.f. を $\mathbf{p}_\theta(x)$ と表記する. さらに, $\widetilde{\mathbb{X}} = \{x \in \mathbb{X};\ \mathbf{p}_\theta(x) > 0\}$ とおく. 母数モデルの正則性から微分記号と積分記号の交換が保証されているので, $i = 1, 2, \ldots, d$ に対して

定理 6.2 の証明の続き

$$
\begin{aligned}
\mathrm{E}_\theta[\dot{\ell}_{n,i}(\theta | X)] &= \sum_{k=1}^{n} \int_{\mathbb{X}} \left(\frac{\partial}{\partial\theta_i} \log \mathrm{p}_\theta(x_k)\right) \mathrm{p}_\theta(x_k)\, dx_k \\
&= \sum_{k=1}^{n} \left\{ \int_{\tilde{\mathbb{X}}} \left(\frac{\partial}{\partial\theta_i} \log \mathrm{p}_\theta(x_k)\right) \mathrm{p}_\theta(x_k)\, dx_k \right. \\
&\qquad \left. + \int_{\mathbb{X}\setminus\tilde{\mathbb{X}}} \left(\frac{\partial}{\partial\theta_i} \log \mathrm{p}_\theta(x_k)\right) \underbrace{\mathrm{p}_\theta(x_k)}_{=0}\, dx_k \right\} \\
&= \sum_{k=1}^{n} \int_{\tilde{\mathbb{X}}} \frac{1}{\mathrm{p}_\theta(x_k)} \left(\frac{\partial}{\partial\theta_i} \mathrm{p}_\theta(x_k)\right) \mathrm{p}_\theta(x_k)\, dx_k \\
&= \sum_{k=1}^{n} \int_{\tilde{\mathbb{X}}} \frac{\partial}{\partial\theta_i} \mathrm{p}_\theta(x_k)\, dx_k =: (*)
\end{aligned}
$$

定理 6.2 の証明の続き

$$
\begin{aligned}
(*) &= \sum_{k=1}^{n} \frac{\partial}{\partial \theta_i} \int_{\widetilde{\mathbb{X}}} \mathrm{p}_\theta(x_k)\, \mathrm{d}x_k \\
&= \sum_{k=1}^{n} \frac{\partial}{\partial \theta_i} \left\{ \int_{\widetilde{\mathbb{X}}} \mathrm{p}_\theta(x_k)\, \mathrm{d}x_k + \int_{\mathbb{X}\setminus\widetilde{\mathbb{X}}} \underbrace{\mathrm{p}_\theta(x_k)}_{=0}\, \mathrm{d}x_k \right\} \\
&= \sum_{k=1}^{n} \frac{\partial}{\partial \theta_i} \underbrace{\int_{\mathbb{X}} \mathrm{p}_\theta(x_k)\, \mathrm{d}x_k}_{=1} \\
&= 0
\end{aligned}
$$

となる. よって (1) は示された.

定理 6.2 の証明の続き

(2) は明らか.

(3) を示すために次に注意する. $i, j = 1, 2, \ldots, d$ と $x \in \widetilde{\mathbb{X}}$ に対して

$$\frac{\partial^2 \log \mathrm{p}_\theta(x)}{\partial\theta_i\partial\theta_j} = \frac{1}{\mathrm{p}_\theta(x)} \frac{\partial^2 \mathrm{p}(x|\,\theta)}{\partial\theta_i\partial\theta_j} - \left(\frac{\partial}{\partial\theta_i} \log \mathrm{p}_\theta(x)\right)\left(\frac{\partial}{\partial\theta_j} \log \mathrm{p}_\theta(x)\right)$$

となる.

定理 6.2 の証明の続き　このとき

$$\int_{\widetilde{\mathbb{X}}} \frac{1}{\mathsf{p}_\theta(x)}\left(\frac{\partial^2 \mathsf{p}_\theta(x)}{\partial\theta_i\partial\theta_j}\right)\mathsf{p}_\theta(x)\,\mathrm{d}x = \frac{\partial^2}{\partial\theta_i\partial\theta_j}\left\{\int_{\widetilde{\mathbb{X}}} \mathsf{p}_\theta(x)\,\mathrm{d}x + \int_{\mathbb{X}\setminus\widetilde{\mathbb{X}}} \mathsf{p}_\theta(x)\,\mathrm{d}x\right\}$$
$$= \frac{\partial^2}{\partial\theta_i\partial\theta_j}\underbrace{\int_{\mathbb{X}} \mathsf{p}_\theta(x)\,\mathrm{d}x}_{=1} = 0$$

と

$$\int_{\widetilde{\mathbb{X}}}\left(\frac{\partial}{\partial\theta_i}\log \mathsf{p}_\theta(x)\right)\left(\frac{\partial}{\partial\theta_j}\log \mathsf{p}_\theta(x)\right)\mathsf{p}_\theta(x)\,\mathrm{d}x$$
$$= \mathsf{E}_\theta\left[\left(\frac{\partial}{\partial\theta_i}\log \mathsf{p}_\theta(X_1)\right)\left(\frac{\partial}{\partial\theta_j}\log \mathsf{p}_\theta(X_1)\right)\right]$$

となる．

定理 6.2 の証明の続き　このことより

$$
\begin{aligned}
\mathsf{E}_\theta\left[\frac{\partial^2 \log \mathsf{p}_\theta(X_1)}{\partial\theta_i\partial\theta_j}\right] &= \int_{\widetilde{\mathbb{X}}} \frac{\partial^2 \log \mathsf{p}_\theta(x)}{\partial\theta_i\partial\theta_j}\mathsf{p}_\theta(x)\,\mathrm{d}x \\
&\quad + \int_{\mathbb{X}\setminus\widetilde{\mathbb{X}}} \frac{\partial^2 \log \mathsf{p}_\theta(x)}{\partial\theta_i\partial\theta_j}\mathsf{p}_\theta(x)\,\mathrm{d}x \\
&= \int_{\widetilde{\mathbb{X}}} \frac{\partial^2 \log \mathsf{p}_\theta(x)}{\partial\theta_i\partial\theta_j}\mathsf{p}_\theta(x)\,\mathrm{d}x \\
&= \underbrace{\int_{\widetilde{\mathbb{X}}} \frac{1}{\mathsf{p}_\theta(x)}\left(\frac{\partial^2 \mathsf{p}_\theta(x)}{\partial\theta_i\partial\theta_j}\right)\mathsf{p}_\theta(x)\,\mathrm{d}x}_{=0} \\
&\quad - \mathsf{E}_\theta\left[\left(\frac{\partial}{\partial\theta_i}\log \mathsf{p}_\theta(X_1)\right)\left(\frac{\partial}{\partial\theta_j}\log \mathsf{p}_\theta(X_1)\right)\right] \\
&= -\mathsf{E}_\theta\left[\left(\frac{\partial}{\partial\theta_i}\log \mathsf{p}_\theta(X_1)\right)\left(\frac{\partial}{\partial\theta_j}\log \mathsf{p}_\theta(X_1)\right)\right]. \qquad (2)
\end{aligned}
$$

定理 6.2 の証明の続き　よって

$$
\begin{aligned}
\{\mathcal{F}_X(\theta)\}_{ij} &= \mathsf{E}\left[\left(\sum_{k_1=1}^{n}\frac{\partial \log \mathsf{p}_\theta(X_{k_1})}{\partial\theta_i}\right)\left(\sum_{k_1=2}^{n}\frac{\partial \log \mathsf{p}_\theta(X_{k_2})}{\partial\theta_j}\right)\right] \\
&= \sum_{k=1}^{n}\mathsf{E}\left[\frac{\partial \log \mathsf{p}(\mathsf{p}_\theta(X_k)}{\partial\theta_i}\frac{\partial \log \mathsf{p}_\theta(X_k)}{\partial\theta_j}\right] \\
&\quad + \sum_{k_1\neq k_2}\mathsf{E}\left[\frac{\partial \log \mathsf{p}_\theta(X_{k_1})}{\partial\theta_i}\frac{\partial \log \mathsf{p}_\theta(X_{k_2})}{\partial\theta_j}\right] \\
&= \sum_{k=1}^{n}\mathsf{E}\left[\frac{\partial \log \mathsf{p}_\theta(X_k)}{\partial\theta_i}\frac{\partial \log \mathsf{p}_\theta(X_k)}{\partial\theta_j}\right] \\
&\quad + \sum_{k_1\neq k_2}\underbrace{\mathsf{E}\left[\frac{\partial \log \mathsf{p}_\theta(X_{k_1})}{\partial\theta_i}\right]}_{=0}\mathsf{E}\left[\frac{\partial \log \mathsf{p}_\theta(X_{k_2})}{\partial\theta_j}\right] =: (**)
\end{aligned}
$$

定理 6.2 の証明の続き

$$(**) = -\sum_{k=1}^{n} \mathrm{E}_\theta\left[\frac{\partial^2 \log \mathrm{p}_\theta(X_k)}{\partial\theta_i\partial\theta_j}\right] \qquad (\because (2))$$

$$= -\mathrm{E}\left[\frac{\partial^2 \ell_n(X|\,\theta)}{\partial\theta_i\partial\theta_j}\right] = -\mathrm{E}[\ddot{\ell}_{n,ij}\,(X|\,\theta)]$$

がわかる. □

例 6.3 $X = (X_1, X_2, \ldots, X_n)$ で
$X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathrm{N}(\mu, \sigma^2)$ とする. このとき, 統計的モデルは $\mathcal{P} = \{\mathrm{N}(\mu, \sigma^2);\, \mu \in \mathbb{R}, \sigma > 0\}$ で与えられるとする. 具体的な計算は資料を見ながら, 皆さんでやってみるとよい.

定義 6.4 $d \in \mathbb{N}$ とし, $\mathbb{X} \subset \mathbb{R}^d$ を空でない部分集合とする. さらに $\Theta \subset \mathbb{R}$ も空でない部分集合とする. $\mathbb{X}$ 上の統計的モデル $\{\mathsf{P}_\theta;\ \theta \in \Theta\}$ は 1 母数指数型分布族であるとは, Θ 上の実数値関数 A と $\kappa^\vee$, $\mathbb{X}$ 上の実数値関数 T と h が存在して [12], P_θ の p.d.f. または p.m.f. $\mathsf{p}_\theta^\vee(x)$ [13] が

$$\mathsf{p}_\theta^\vee(x) = \begin{cases} h(x)\exp\{A(\theta)T(x) - \kappa^\vee(\theta)\} & (x \in \mathbb{X} \subset \mathbb{R}^d) \\ 0 & (x \notin \mathbb{X}) \end{cases} \tag{3}$$

の形で書けるときをいう.

注意 6.5 関数 A, h, T の表現は一意ではない. □

[12] A と B は Greek letters 大文字の α と β である.

[13] 後に別の母数化である自然指数型分布族を導入するときに記号を簡単にするために, ここでは「$\mathsf{p}^\vee$ と $\kappa^\vee$」という少し不自然な記号を用いた.

例 6.6(Poisson 分布族) $X \sim \mathsf{Po}(\theta)\ (\theta > 0)$ とする. その p.m.f. は

$$\mathsf{p}^{\vee}_{\theta}(x) = \frac{\theta^x e^{-\theta}}{x!} = \frac{1}{x!} \exp\{x \log\theta - \theta\} \qquad (x = 0, 1, \ldots)$$

である. したがって $\{\mathsf{Po}(\theta);\ \theta \in (0, \infty)\}$ は

$$d = 1,\ A(\theta) = \log\theta,\ \kappa^{\vee}(\theta) = \theta,\ T(x) = x,\ h(x) = \frac{1}{x!}$$

によって生成される 1 母数指数型分布族である. □

例 6.7(2 項分布族) $X \sim \mathsf{Bino}(n, \theta)$ とする. ただし $n \in \mathbb{N}$, $0 < \theta < 1$ である. このとき X の p.m.f. は

$$\begin{aligned} \mathsf{p}_\theta^\vee(x) &= \binom{n}{x} \theta^x (1-\theta)^{n-x} \\ &= \binom{n}{x} \exp\left\{ x \log\left(\frac{\theta}{1-\theta}\right) + n \log(1-\theta) \right\} \qquad (x = 0, 1, \ldots, n) \end{aligned}$$

と書ける. ただし

$$\binom{n}{x} = \frac{n!}{x! \times (n-x)!}, \qquad 0! = 1$$

である. したがって $\{\mathsf{Bino}(n, \theta);\ \theta \in (0, 1)\}$ は

$$d = 1,\ A(\theta) = \log\left(\frac{\theta}{1-\theta}\right),\ \kappa^\vee(\theta) = -n \log(1-\theta),\ T(x) = x,$$

$$h(x) = \binom{n}{x}$$

によって生成される 1 母数指数型分布族である.

例 6.8 $X = (Y, Z)^\top$, $Z, W \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, 1)$ とし, $Y = Z + \theta W$ $(\theta > 0)$ とする. このとき X の同時 p.d.f. は次で与えられる. $x = (y, z)$ としたとき

$$\mathrm{p}_\theta^X(x) = \mathrm{p}_\theta^X(y, z) = \mathrm{p}^Z(z)\mathrm{p}^{Y|Z}(y|z) = \varphi(z)\theta^{-1}\varphi\left(\frac{y-z}{\theta}\right)$$
$$= \frac{1}{2\pi\theta}\exp\left\{-\frac{1}{2}\left(z^2 + \frac{(y-z)^2}{\theta^2}\right)\right\}$$
$$= \frac{1}{2\pi}\exp\left\{-\frac{z^2}{2}\right\}\exp\left\{-\frac{(y-z)^2}{2\theta^2} - \log\theta\right\}$$

と書ける. ただし p_Z は Z の p.d.f. とし, $\mathrm{p}_{Y|Z}$ は $Z = z$ を与えたときの Y の条件付き p.d.f., φ は $\mathrm{N}(0, 1)$ の p.d.f. で

$$\varphi(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2} \qquad (z \in \mathbb{R})$$

である.

6.2 指数型分布族 (4)

$\boxed{\text{例 } \mathbf{6.8} \text{ の続き}}$ したがって X の分布は

$$d = 2,\ A(\theta) = -\frac{1}{2\theta^2},\ \kappa^\vee(\theta) = \log\theta,\ T(x) = (y - z)^2,$$
$$h(x) = \frac{1}{2\pi}\exp\left\{-\frac{z^2}{2}\right\}$$

で生成される **1** 母数指数型分布族に属する. □

1 母数指数型分布族に属する分布からのランダム標本の分布も 1 母数指数型分布族に属する. 特に, $n \in \mathbb{N}$ とし, $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{P}_\theta$ とする. ただし, $\{\mathsf{P}_\theta;\ \theta \in \Theta\}$ は (3) で与えられる指数型分布族とする. すると $\{\mathsf{P}_\theta^{\otimes n};\ \theta \in \Theta\}$ は $X^{(n)} = (X_1, X_2, \ldots, X_m)$ に対する指数型分布族となる. ただし, $\mathbb{R}^{dn}$ 上の確率測度を

$$\mathsf{P}_\theta^{\otimes n} = \underbrace{\mathsf{P}_\theta \times \mathsf{P}_\theta \times \cdots \times \mathsf{P}_\theta}_{n\text{個}}$$

と定めてた. このとき, $\mathsf{P}_\theta^{\otimes n}$ の同時 p.d.f. $\prod_{j=1}^n \mathsf{p}_\theta^\vee(\cdot)$ は次で与えられる. すると

$$\prod_{j=1}^n \mathsf{p}_\theta^\vee(x_j) = \prod_{j=1}^n h(x_j) \exp\left\{A(\theta)T(x_j) - \kappa^\vee(\theta)\right\}$$

$$= \left[\prod_{j=1}^n h(x_j)\right] \exp\left\{A(\theta) \sum_{j=1}^n T(x_j) - n\kappa^\vee(\theta)\right\}.$$

6.2 指数型分布族 (6)

したがって $\{\mathbf{P}_\theta^{\otimes n}; \theta \in \Theta\}$ は

$$A(\theta), \quad \sum_{j=1}^{n} T(x_j), \quad n\kappa^\vee(\theta), \quad \prod_{j=1}^{n} h(x_j)$$

で生成される $\mathbb{R}^{dn}$ 上の **1** 母数指数型分布族となる.
統計量 $\sum_{j=1}^n T(X_j)$ は θ の自然十分統計量[14]となる.

[14] 十分統計量の定義は定義 **??** を参照のこと.

(3) で表現された指数型分布族を θ でなく $\eta(= A(\theta))$ で添え字付けることを考える. 記号の乱用 [15] すると **1** 母数指数型分布族は

$$\mathsf{p}_\eta(x) = \begin{cases} h(x)\exp\{\eta T(x) - \kappa(\eta)\} & (x \in \mathbb{X} \subset \mathbb{R}^d) \\ 0 & (x \notin \mathbb{X}) \end{cases} \tag{4}$$

と書くことができる. ただし

$$\kappa(\eta) = \begin{cases} \log\left(\sum_{x\in\mathbb{X}} h(x)\exp\{\eta T(x)\}\right) & (\text{離散型の場合}) \\ \log\left(\int_{\mathbb{X}} h(x)\exp\{\eta T(x)\}\,\mathrm{d}x\right) & (\text{連続型の場合}) \end{cases}$$

である. κ の定義より, 直ちに $\kappa(\eta) = \kappa(A(\theta)) = \kappa^{\vee}(\theta)$ がわかる.

[15] ここでは, 記号の乱用をしている. 本来であれば, 領域を制限すると A の逆関数は存在するので, $\mathsf{p}(x|\,\eta)$ は $\mathsf{p}(x|\,A^{-1}(\eta))$ と書くべきであろう.

さらに

$$\mathcal{E} := \{\eta \in \mathbb{R};\ \kappa(\eta) < \infty\}$$

とおく. $\mathcal{E}$ は $\mathbb{R}$ の区間となる. すると確率分布族

$$\left\{\mathrm{p}(x|\,\eta) = h(x)\exp\{\eta T(x) - \kappa(\eta)\};\ \eta \in \mathcal{E}\right\}$$

は確率分布族

$$\left\{\mathrm{p}_\theta(x) = h(x)\exp\{A(\theta)T(x) - \kappa(\theta)\};\ \theta \in \Theta\right\}$$

を含む. 確率分布族 $\{\mathrm{p}(x|\,\eta);\ \eta \in \mathcal{E}\}$ は T, h によって生成された正準(せいじゅん)指数型分布族または自然指数型分布族といい, $\mathcal{E}$ を自然母数空間といい, T を自然十分統計量という.

例 6.9(例 6.6 の続き) Poisson 分布族は

$$\mathsf{p}_\eta(x) = \frac{1}{x!}\exp\{\eta x - e^\eta\} \qquad (x = 0,\ 1,\ 2, \ldots),$$

$$\eta = \log\theta,\ h(x) = \frac{1}{x!},\ T(x) = x,$$

$$\exp(\kappa(\eta)) = \sum_{x=0}^{\infty} h(x)\exp(\eta T(x)) = \sum_{x=0}^{\infty}\frac{\exp(\eta x)}{x!} = \sum_{x=0}^{\infty}\frac{(e^\eta)^x}{x!} = \exp(e^\eta),$$

$$\mathcal{E} = \mathbb{R}$$

となる. □

補題 6.10

$$\mathcal{E} := \left\{\eta \in \mathbb{R};\ \kappa(\eta) := \mathrm{E}[\exp\{\eta T(X)\}] < \infty\right\}$$

とする. このとき, 関数 $\kappa(\eta)$ は $\mathcal{E}^\circ$ 上の無限回微分可能である. ただし $\mathcal{E}^\circ$ は $\mathcal{E}$ の内部である. さらに積分記号と微分記号の交換は可能である.

補題 6.10 の証明 資料を参照のこと.

$\boxed{\text{定理 6.11}}$ X は (4) で与えられた自然指数型分布族に属する分布 $\mathrm{p}(x|\,\eta)$ に従うとする. η は $\mathcal{E}$ の内点としたとき, $T(X)$ の積率母関数は原点の近傍で存在し

$$M_T(s) = \exp[\kappa(s+\eta) - \kappa(\eta)]$$

で与えられる. ただし, s s は $\mathbf{0}$ のある近傍に含まれるとする. さらに

$$\mathrm{E}[T(X)] = \dot{\kappa}(\eta); \qquad \mathrm{Var}[T(X)] = \ddot{\kappa}(\eta) \tag{5}$$

である. ただし

$$\dot{\kappa}(\eta) = \frac{\mathrm{d}\kappa}{\mathrm{d}\eta}(\eta), \quad \ddot{\kappa}(\eta) = \frac{\mathrm{d}^2\kappa}{\mathrm{d}\eta^2}(\eta)$$

である.

定理 6.11 の証明 連続型分布の場合について証明する. κ の定義に注意すると

$$
\begin{aligned}
M(s) &= \mathrm{E}[\exp(sT(X))] \\
&= \int_{\mathbb{X}} h(x) \exp[(s+\eta)T(x) - \kappa(\eta)] \, \mathrm{d}x \\
&= \exp[\kappa(s+\eta) - \kappa(\eta)] \int_{\mathbb{X}} h(x) \exp[(s+\eta)T(x) - \kappa(s+\eta)] \, \mathrm{d}x \\
&= \exp[\kappa(s+\eta) - \kappa(\eta)]
\end{aligned}
$$

から 1 番目の主張はわかる. 最後の等号は $s + \eta \in \mathcal{E}^\circ$ となるように s をとると

$$
\int_{\mathbb{X}} h(x) \exp[(s+\eta)T(x) - \kappa(s+\eta)] \, \mathrm{d}x = 1
$$

となることよりわかる.

定理 6.11 の証明の続き　次に, $\kappa(\cdot)$ の定義と補題 6.10 から

$$\frac{\mathrm{d}\kappa}{\mathrm{d}\eta}(\eta) = \frac{\mathrm{d}}{\mathrm{d}\eta}\log\left(\int_{\mathbb{X}} \exp\{\eta T(x)\}h(x)\,\mathrm{d}x\right)$$

$$= \frac{\dfrac{\mathrm{d}}{\mathrm{d}\eta}\displaystyle\int_{\mathbb{X}} \exp\{\eta T(x)\}h(x)\,\mathrm{d}x}{\displaystyle\int_{\mathbb{X}} \exp\{\eta T(x)\}h(x)\,\mathrm{d}x}$$

$$= \frac{\displaystyle\int_{\mathbb{X}} \frac{\mathrm{d}}{\mathrm{d}\eta}\exp\{\eta T(x)\}h(x)\,\mathrm{d}x}{\exp\{\kappa(\eta)\}}$$

$$= \frac{\displaystyle\int_{\mathbb{X}} T(x)\exp\{\eta T(x)\}h(x)\,\mathrm{d}x}{\exp\{\kappa(\eta)\}}$$

$$= \int_{\mathbb{X}} T(x)\exp\{\eta T(x) - \kappa(\eta)\}h(x)\,\mathrm{d}x = \mathrm{E}[T(X)].$$

$\boxed{\text{定理 6.11 の証明の続き}}$ 前のスライドの式から (5) の 1 番目の等号が示せた. 同様に

$$
\begin{aligned}
\frac{\mathrm{d}^2\kappa}{\mathrm{d}\eta^2}(\eta) &= \frac{\mathrm{d}}{\mathrm{d}\eta}\int_{\mathbb{X}} T(x)\exp\{\eta T(x)-\kappa(\eta)\}h(x)\,\mathrm{d}x \\
&= \int_{\mathbb{X}} T(x)h(x)\frac{\mathrm{d}}{\mathrm{d}\eta}\exp\{\eta T(x)-\kappa(\eta)\}\,\mathrm{d}x \\
&= \int_{\mathbb{X}} T(x)h(x)\exp\{\eta T(x)-\kappa(\eta)\}\frac{\mathrm{d}}{\mathrm{d}\eta}\Big(\eta T(x)-\kappa(\eta)\Big)\,\mathrm{d}x \\
&= \int_{\mathbb{X}} T^2(x)h(x)\exp\{\eta T(x)-\kappa(\eta)\}\,\mathrm{d}x \\
&\qquad - \underbrace{\frac{\mathrm{d}}{\mathrm{d}\eta}\kappa(\eta)}_{=\mathsf{E}[T(X)]}\int_{\mathbb{X}} T(x)h(x)\exp\{\eta T(x)-\kappa(\eta)\}\,\mathrm{d}x \\
&= \mathsf{E}[T^2(X)] - \{\mathsf{E}[T(X)]\}^2 \\
&= \mathrm{Var}[T(X)].
\end{aligned}
$$

本日の講義のまとめ

- 5.2 統計的実験と母数モデル
  - 統計的モデル, 母数空間, 真の分布（母集団分布）, 真の母数
- 5.2 統計的決定問題
  - 点推定論, 検定論, 区間推定論
- 6.1 正則母数モデルと Fisher 情報量
- 6.2 指数型分布族
- 6.3 指数型分布族の正準表示