宗教文学講義スライド（2024年10月27）

本日の講義内容

- 5.1 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 十分統計量 → 次回

5.1 統計的実験と母数モデル

資料より詳しく説明する

$X_1, X_2, \dots, X_n$を確率空間$(\Omega, \mathcal{A}, \Pr)$上の確率変数とする.

$X_i : \Omega \to \mathbb{R}$で, $\forall r \in \mathbb{R}$に対して

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

をみたす.

$(\cdot)^T$は転置の記号

$\underline{X} = (X_1, X_2, \dots, X_n)^T$と記す.

仮定 この講義では、$X_1, X_2, \ldots, X_n$ 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

確率ベクトル $\underline{X}$ の値域 $\mathbb{R}^n$ を**標本空間**という。

$B(\mathbb{R})$ を $\mathbb{R}$ 上の Borel 集合族（$\mathbb{R}$ の開集合族を含む最小の σ 加法族）とする。

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

$$P(B) := \Pr(X_1 \in B) \qquad (\forall B \in \mathcal{B}(\mathbb{R}))$$

で定める.

確率測度 P を X_1 の分布 といい

$$X_1 \sim P$$

と記す.

$\underline{X}$ の分布を $P^{\otimes n}$ または $P^{\underline{X}}$ と記す.

すなわち、${}^{\forall} B_j \in \mathcal{B}(\mathbb{R})$ に対して

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

で定める。すると

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

となる。

確率論

$P^{\underline{X}}$ を既知として，$\underline{X}$ の分布論的特徴を調べる．

推測統計学

$P^{\underline{X}}$ を未知として，$\underline{X}$ の情報から $P^{\underline{X}}$ を回復することが目標．

いま、P^*を$(\mathbb{R}, \mathcal{B}(\mathbb{R}))$上の分布とし

$$X_1, X_2, \dots, X_n \overset{\text{i.i.d.}}{\sim} P^* \iff \underline{X} \sim (P^*)^{\otimes n}$$

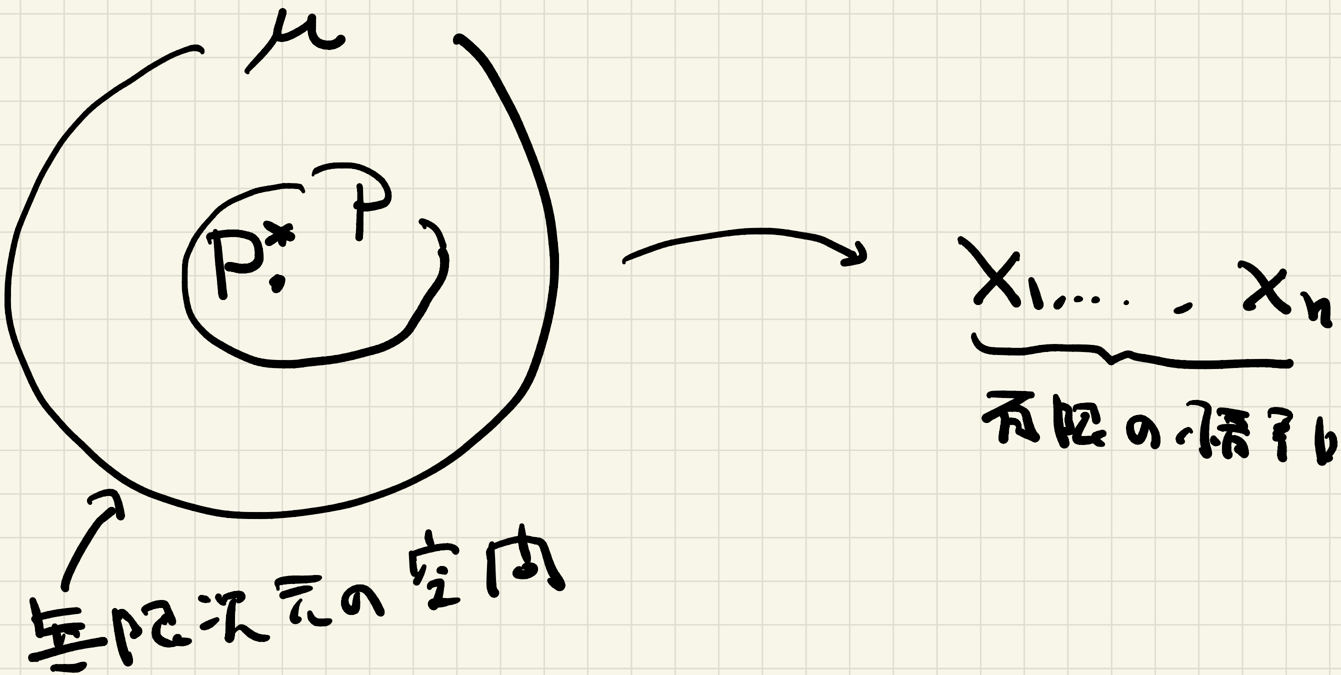
とする。　　　「真の分布」とよぶ

$\underline{X}$の情報に基づき、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。

$\mathcal{P}$：確率分布全体の集合とする



$\mathcal{P}$：統計的モデル（$\mathcal{M}$の都合のよい部分集合）

確率分布の「適当」な集まり。

仮定　想定(設定)した統計的モデルは真の分布を含むとする.

母数化：統計的モデル$\mathcal{P}$の要素のラベリング集合$\Theta \subset \mathbb{R}^k$とし, 写像

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

を想定する. すなわち, $\mathcal{P}$と$\mathbb{R}^k$の部分集合との間に対応関係ができると想定する.

仮定 写像

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

は Θ から $\mathcal{P}$ への全単射とする.

すると, $^\exists \theta^* \in \Theta$ s.t.

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

となる.

θ^* を真の母数といい, Θ の元 θ を母数と,

Θ を母数空間という

この写像を統計的モデルの母数化という

注意 ： 母数化は一意ではない.

定義5.2 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$上の統計的モデル $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ は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) $\Theta \subset \mathbb{R}^d$ $(d \in \mathbb{N})$ は有限次元Euclid空間 $\mathbb{R}^d$ の「よい」部分集合である.

(2) 写像

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

は「滑らか」である.

(3) ${}^{\forall}\theta_1, \theta_2 \in \Theta$ に対して

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

をみたす.

注意　(3)をみたす母数化を識別可能という.

注意　母数空間 Θ の次元が有限次元であるとき,
$\mathcal{P} = \{P_\theta ; \theta \in \Theta\}$ をパラメトリックモデルとよぶ.
(後例になろう)

定義 5.5 (1) 可測空間 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))^n$ とそれ上の

母数モデル $\mathcal{P} = \{P_\theta ; \theta \in \Theta\}$ の組

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

を統計的実験という.

(2) 観測データの生成分布 P^* に対応する Θ の元を

θ^* とかく、真の母数という. すなわち

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

(3) $\underset{\sim}{X} = (X_1, X_2, \ldots, X_n)^T$ は

$$X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} P^*$$

と仮定し、$\underset{\sim}{X}$ を標本の大きさが n のランダム標本という.

目標　$\underset{\sim}{X}$ の情報から P^* を回復したい!

例15.6 (1) $X_1, X_2, \dots, X_n \overset{\text{i.i.d.}}{\sim} N(\mu^*, (\sigma^*)^2)$

とする. ただし, μ^* と σ^* は未知.

このときの統計的実験は

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

である. よって

$$P_\theta(B) = \int_B p(x \mid \theta)\, dx \qquad (B \in \mathcal{B}(\mathbb{R}))$$

である.

(2) $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathrm{Ber}(\theta)\ (0<\theta<1)$ とする.

このとき, 統計的実験は

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

となる.

5.2 統計的決定問題

$\mathbb{R}^n$: 標本空間

$\underline{X} = (X_1, X_2, \ldots, X_n)^T$: 標本

(1) $\mathbb{A}$: 行動空間　例: $\mathbb{A} = \mathbb{R}$; $\mathbb{A} = \{0, 1\}$.

決定関数

$$d : \mathbb{R}^n \ni \underline{x} \longmapsto d(\underline{x}) \in \mathbb{A}$$

ただし, $\underline{x} = (x_1, x_2, \ldots, x_n)^T$

決定関数全体を $\mathbb{D}$ とおき, $\mathbb{D}$ を決定空間という

(2) $\mathbb{A}\times\Theta$ 上の非負値関数

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

を用意する. 関数 L を損失関数という.

データ $\underset{\sim}{X}=\underset{\sim}{x}$ を観測したとき, $L(\theta^*, d(\underset{\sim}{x}))$ の値を小さくする決定関数 d が望ましい.

よって, ランダムな量 $L(\theta^*, d(\underset{\sim}{X}))$ を評価の指標と用いればよい. しかし. θ^* は未知なので, $L(\theta^*, d(\underset{\sim}{x}))$ も $L(\theta^*, d(\underset{\sim}{X}))$ の分布はわからない.

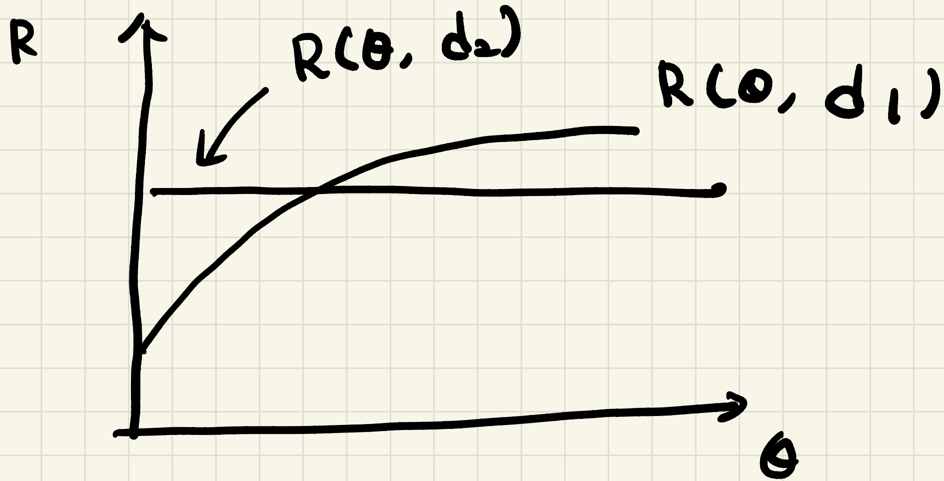
そこで，$L(\theta, d(\underline{x}))$ $(\theta \in \Theta)$ を $P_\theta^{\otimes n}$ に関して期待値を取って θ の

$$R(\theta, d) = E_\theta[L(\theta, d(\underline{x}))]$$

を考える．ただし，$E_\theta[\cdot]$ は $P_\theta^{\otimes n}$ $(\theta \in \Theta)$ のもとでの期待値である．$R(\cdot, d)$ を決定関数 d の危険関数 (risk function) という．すなわち

$$\Theta \ni \theta \longmapsto R(\theta, d) \in [0, \infty) \cup \{+\infty\}.$$

(3) θ^*における危険関数 $R(\theta^*, d)$ を小さくする d を求めればよい。しかし、$R(\theta^*, d)$ の値は未知。よって、危険関数 $R(\theta, d)$ を θ に関して一様に評価する



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

を**統計的決定問題**という.

(4) $\mathbb{D}$ は $\mathbb{R}^n$ から $\mathbb{A}$ への可測関数全体と
なるが,「よい」決定関数の探索は $\mathbb{D}$ の
「よい」部分集合の中で行うのが通常.

点推定　真の母数 θ^* を $\underline{X}$ に基づいて1点で図定するのが点推定

!したがって、$\mathbb{A} = \Theta$，決定関数 $d(\underline{X})$ を**推定量**といい、$\underline{X} = \underline{x}$ を観測したときの推定量の値 $d(\underline{x})$ を**推定値**という。

注意　推定量…確率変数…分布を持つ

推定値…数

$\Theta = \mathbb{R}$ とする。たとえば

$$L_1(\theta, a) = (a-\theta)^2 \; ; \quad a \in A, \ \theta \in \Theta$$

$$L_2(\theta, a) = |a-\theta| \; ; \quad \text{''}$$

$$R_i(\theta, d) = E_\theta[L_i(\theta, d(\underline{X}))] \quad (i=1,2)$$

$R_1(\theta, d)$ を d の平均2乗誤差

(Mean Squared Error = MSE)

$R_2(\theta, d)$ を d の平均絶対誤差

(Mean Absolute Error = MAE)

したがって，危険関数を Θ に関して何らかの意味で一様に評価して，「よい」/「最適」な推定量をみつけたい．

検定問題　母数空間 Θ を排反な部分集合に分割する。

$$\Theta = \Theta_0 \cup \Theta_1 ; \quad \Theta_0 \cap \Theta_1 = \phi.$$

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ことを考える。

よって、$A = \{0, 1\}$ と言える。決定関数は

標本空間 $\mathbb{R}^n$ の部分集合 W に対して

$$d(x) = \begin{cases} 1 & (x \in W) \iff \theta \in \Theta_1 \text{ と判断} \\ 0 & (x \notin W) \Longleftarrow \theta \in \Theta_0 \text{ と判断} \end{cases}$$

で定めることができる. 検定では, 決定関数 d を **検定関数** という

損失関数として

$$L(\theta, 0) = \begin{cases} 0 & (\theta \in \Theta_0) \longleftarrow \text{正しい判断} \\ 1 & (\theta \in \Theta_1) \longleftarrow \text{誤判断} \end{cases}$$

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

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

通常，$H_0 : \theta \in \Theta_0$ を帰無仮説 とよび，

$H_1 : \theta \in \Theta_1$ を対立仮説 という.

$R(d, \theta) = E_\theta[L(\theta, L(x))]$ とおく.

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

トレードオフ

実際、

(第1種の誤りの確率)+(第2種の誤りの確率)

$\geq$ 下限

となることが知られている。

いま,

$$\theta \in \Theta_0 \text{ のとき}, \quad \beta(\theta) := R(\theta, d)$$

$$\theta \in \Theta_1 \text{ のとき}, \quad \beta(\theta) = 1 - R(\theta, d)$$

としたとき, β を検出力関数という.

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 d の中で，$\theta \in \Theta_1$ において

一様に $\beta(\theta)$ を大きく d をさがしたい．

この α を有意水準 といい，$\sup_{\theta \in \Theta_0} R(\theta, d)$

を検定関数 d のサイズという

$$\sup_{\theta \in \Theta_0} R(\theta, d) \le \alpha \Longleftrightarrow (\text{サイズ}) \le \text{有意水準}$$

区間推定　観測 $\underline{X}=\underline{x}$ に基づく母数 θ の信頼区間を $[\ell(\underline{x}), u(\underline{x})]$ に対して，損失関数として

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

を考える．この場合は L は負の値を取ることもある．また

$$\mathbb{I}\{\cdots\} = \begin{cases} 1 & (\cdots \text{が真}) \\ 0 & (\cdots \text{が偽}) \end{cases}$$

決定関数 $d(\underline{x}) = [\ell(\underline{x}), u(\underline{x})]$ に対して

$$R(\theta, d) = E_\theta[u(\underline{x}) - \ell(\underline{x})]$$

$$- Pr_\theta[\theta \in [\ell(\underline{x}), u(\underline{x})]]$$

となる.

実数 α $(0 < \alpha < 1)$ に対して

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

のもとで 区間の長さの期待値

$$E_\theta[u(\underline{x}) - \ell(\underline{x})]$$

を短くする区間 $[\ell(\underline{x}), u(\underline{x})]$ がのぞましい.

$1-\alpha$ を信頼係数とよぶ.

統計的推測

- 推定論 $\begin{cases} \text{点推定論} \\ \text{区間推定論} \end{cases}$

- 検定論

決定空間の元の間に半順序 $<$ を導入

$^\forall d_1, d_2 \in \mathbb{D}$ に対して

$$d_1 < d_2$$

$$\Leftrightarrow R(\theta, d_1) \leq R(\theta, d_2) \quad (^\forall \theta \in \Theta)$$

かつ

$$R(\theta_0, d_1) < R(\theta_0, d_2) \quad (^\exists \theta_0 \in \Theta)$$

注意　半順序 < に関して一番よいものは存在しない!

∵) 一番よいものは真の母数を知っていること になる.

決定関数の最適性へのアプローチ

(1) ミニマックス基準, Bayes 基準

(2) 考察する決定関数を制限

…不偏性, 不変性 / Neyman-Pearson の補題

6.1 正則母数モデルと Fisher 情報量

定義 6.1 $d, n \in \mathbb{N}$ とし，$\mathcal{X}^n (\subset \mathbb{R}^n)$ を標本空間とする．$\mathcal{X}$ 上の母数モデル $\mathcal{P} = \{P_\theta ; \theta \in \Theta \subset \mathbb{R}^d\}$ が与えられているとする．分布 P_θ の p.m.f. / p.d.f を $p(\cdot \mid \underline{\theta})$ と記すことにする．

このとき，$\mathcal{P}$ は正則であるとは次の条件 (1)〜(5) をみたすときをいう

(1) Θ は $\mathbb{R}^d$ の開集合.

(2) $\{x \in \mathbb{R} ; p(x|\boldsymbol{\theta}) > 0\}$ は $\boldsymbol{\theta}$ によらず同じ.
上の集合を「分布の台」と言葉を乱用していうことにする.

(3) ${}^\forall \boldsymbol{\theta} \in \Theta$ とする. $\boldsymbol{\theta}$ に関する1次と2次の $p(x|\boldsymbol{\theta})$ の導関数は(ほとんど至る所の $x \in \mathcal{X}$ に対して連続である.

(4) $p(x|\underline{\theta})$のθに関する1次と2次の偏導関数は$x(\in \mathcal{X})$の関数として可積分

(5) $p(x|\underline{\theta})$の1次および2次の導関数は$\underline{\theta}$の微分記号とxの積分記号と交換ができる

$\underline{\theta}=(\theta_1,\dots,\theta_d)^T$と書いたとき

$$\int_{\mathcal{X}} \frac{\partial}{\partial\theta_j} p(x|\underline{\theta})\,dx = \frac{\partial}{\partial\theta_j}\int_{\mathcal{X}} p(x|\underline{\theta})\,dx$$

$(j=1,\dots,d)$

等が成立

$X_1=x_1,\ X_2=x_2,\ \dots,\ X_n=x_n$ を観測したとき、Θ 上の関数 l_n を

$$l_n(\underline{\theta}\mid\underline{x}) = \log\prod_{j=1}^{n} p(x_j\mid\underline{\theta}) \quad (\underline{x}=(x_1,\dots,x_n))$$

で定める。

$l_n(\cdot\mid\underline{x})$ を対数尤度関数といい

$\exp(l_n(\underline{\theta}\mid\underline{x}))$ を尤度関数という。

記法　$\underline{X} \sim P_{\underline{\theta}}^{\otimes n}$ とする．この分布に関する期待値，分散および共分散を

$$E_{\underline{\theta}}[\cdot],\quad \mathrm{Var}_{\underline{\theta}}[\cdot],\quad \mathrm{Cov}_{\underline{\theta}}[\cdot,\cdot]$$

と記す．すなわち，$f: \mathcal{X}^n \to \mathbb{R}$ に対して

$$E_{\underline{\theta}}[f(\underline{X})] = \begin{cases} \displaystyle\sum_{\underline{x} \in \mathcal{X}^n} f(\underline{x})\, p^{\underline{X}}(\underline{x} \mid \underline{\theta}) & (\text{離散型}) \\ \displaystyle\int_{\mathcal{X}^n} f(\underline{x})\, p^{\underline{X}}(\underline{x} \mid \underline{\theta})\, d\underline{x} & (\text{連続型}) \end{cases}$$

である．ただし，$d\underline{x} = dx_1\, dx_2 \cdots dx_n$ で

$p^{X}(\cdot\mid\underline{\theta})$ は $P_{\underline{\theta}}^{(n)}$ に対応する p.m.f. / p.d.f.

である. さらに, $\underline{X}$ を確率空間 $(\Omega, \mathcal{A}, \Pr)$

上の確率ベクトルとしたとき, 可測空間 $(\Omega, \mathcal{A})$

上の測度 $\Pr_{\underline{\theta}}$ を

$$\Pr_{\underline{\theta}}(\underline{X}^{-1}(B)) := \Pr_{\underline{\theta}}(\underline{X} \in B)$$

$$= E_{\underline{\theta}}[\mathbb{I}_B(\underline{X})]$$

$$({}^{\forall}B \in \mathcal{B}(\mathcal{X}^n) \quad)$$

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

である.

さらに, $i, j \in \{1, 2, \dots, d\}$ に対して

$$\dot{\ell}_{n,i} := \dot{\ell}_{n,i}(\underline{\theta} \mid \underline{x})$$

$$= \frac{\partial \ell_n(\underline{\theta} \mid \underline{x})}{\partial \theta_i}$$

$$\ddot{\ell}_{n,ij} = \ddot{\ell}_{n,ij}(\underline{\theta} \mid \underline{x}) = \frac{\partial^2 \ell_n(\underline{\theta} \mid \underline{x})}{\partial \theta_i \partial \theta_j}$$

$$\dot{\underline{\ell}}_n = (\dot{\ell}_{n,1}, \dot{\ell}_{n,2}, \ldots, \dot{\ell}_{n,d})^T$$

$$\ddot{\underline{\ell}}_n = \left(\ddot{\ell}_{n,ij} \right)_{i,j=1,2,\ldots,d}$$

$d \times d$ 行列

$\underline{X} \sim P_{\underline{\theta}}^{\otimes n}$ のとき, Fisher情報量 $I_{\underline{X}}(\underline{\theta})$ を

$$I_{\underline{X}}(\underline{\theta}) = \left(E_{\theta}\left[\dot{\ell}_{n,i}(\underline{X}|\underline{\theta})\, \dot{\ell}_{n,j}(\underline{X}|\underline{\theta}) \right] \right)_{i,j=1,\ldots,d}$$

で定める.

注) 定義6.1(4)より, $I_{\underline{X}}(\underline{\theta})$ は 存在する.

定理 6.2 標本空間 $\mathcal{X}$ 上の母数モデル $\mathcal{P}=\{P_\theta : \theta\in\Theta\}$ は正則とする。$\underline{X}\sim P_\theta^{\otimes n}$ としたとき、${}^\forall\underline{\theta}\in\Theta$ に対して以下が成立する。

(1) $E_{\underline{\theta}}[\underline{\dot{\ell}}_n(\underline{\theta}\mid\underline{X})]=\underline{0}_d$.

(2) $F_{\underline{X}}(\underline{\theta})=\mathrm{Cov}[\underline{\dot{\ell}}_n(\underline{\theta}\mid\underline{X})]$

$$=E_{\underline{\theta}}[\underline{\dot{\ell}}_n(\underline{\theta}\mid\underline{X})\underline{\dot{\ell}}_n^T(\underline{\theta}\mid\underline{X})].$$

(3) $F_{\underline{X}}(\underline{\theta})=-E_{\underline{\theta}}[\underline{\ddot{\ell}}_n(\underline{\theta}\mid\underline{X})]$.

証明 (1)と(3)について，$d=1$で連続型の場合に示す.

(1) $\tilde{\mathcal{X}} = \{ x \in \mathbb{R} ; p(x|\theta) > 0 \}$と定める.

$$E_\theta[\dot{\ell}_n(\theta | \underline{X})]$$

$$= E_\theta\left[\frac{\partial}{\partial\theta} \log \left(\prod_{k=1}^{n} p(x_k|\theta) \right) \right]$$

$$= E_\theta\left[\sum_{k=1}^{n} \frac{\partial}{\partial\theta} \log p(x_k|\theta) \right]$$

$$= \sum_{k=1}^{n} \int_{\mathcal{X}^n} \left(\frac{\partial}{\partial \theta} \log p(x_k | \theta) \right) \prod_{j=1}^{n} p(x_j | \theta)\, dx_j$$

$$= \sum_{k=1}^{n} \left\{ \int_{\mathcal{X}} \left(\frac{\partial}{\partial \theta} \log p(x_k | \theta) \right) p(x_k | \theta)\, dx_k \right\}$$

$$\times \prod_{j \neq k} \underbrace{\int_{\mathcal{X}} p(x_j | \theta)\, dx_j}_{=1}$$

$$= \sum_{k=1}^{n} \left\{ \int_{\mathcal{X}^*} \left(\frac{\partial}{\partial \theta} \log p(x_k | \theta) \right) p(x_k | \theta)\, dx_k \right.$$

$$\left. + \int_{\mathcal{X} \setminus \mathcal{X}^*} \left(\frac{\partial}{\partial \theta} \log p(x_k | \theta) \right) \underbrace{p(x_k | \theta)}_{=0}\, dx_k \right\}$$

$$= \sum_{k=1}^{n} \int_{\tilde{X}} \frac{1}{p(x_k|\theta)} \left(\frac{\partial}{\partial \theta} p(x_k|\theta) \right) p(x_k|\theta)\, dx_k$$

$$= \sum_{k=1}^{n} \int_{\tilde{X}} \left(\frac{\partial}{\partial \theta} p(x_k|\theta) \right) dx_k$$

$$= \sum_{k=1}^{n} \frac{\partial}{\partial \theta} \int_{\tilde{X}} p(x_k|\theta)\, dx_k$$

$$= \sum_{k=1}^{n} \frac{\partial}{\partial \theta} \left\{ \int_{\tilde{X}} p(x_k|\theta)\, dx_k + \int_{X \setminus \tilde{X}} \underbrace{p(x_k|\theta)}_{=0}\, dx_k \right\}$$

$$= \sum_{k=1}^{n} \frac{\partial}{\partial \theta} \underbrace{\int_{\mathcal{X}} p(x_k \mid \theta)\, dx_k}_{=1} = 0 .$$

(3) $E_\theta\left[\frac{\partial^2}{\partial \theta^2} \log p(X_1 \mid \theta) \right]$

$$= \int_{\tilde{\mathcal{X}}} \left(\frac{\partial^2}{\partial \theta^2} \log p(x \mid \theta) \right) p(x \mid \theta)\, dx$$

$$+ \int_{\mathcal{X} \setminus \tilde{\mathcal{X}}} \left(\frac{\partial^2}{\partial \theta^2} \log p(x \mid \theta) \right) p(x \mid \theta)\, dx$$

$$= \int_{\tilde{\mathcal{X}}} \frac{\partial}{\partial \theta} \left[\frac{1}{p(x \mid \theta)} \left(\frac{\partial}{\partial \theta} p(x \mid \theta) \right) \right] p(x \mid \theta)\, dx$$

$$= \int_{\mathcal{X}} \left\{ \frac{1}{p(x|\theta)} \left(\frac{\partial^2}{\partial \theta^2} p(x|\theta) \right) - \frac{1}{p^2(x|\theta)} \left(\frac{\partial}{\partial \theta} p(x|\theta) \right)^2 \right\} \times p(x|\theta)\, dx$$

$$= \int_{\mathcal{X}} \left(\frac{\partial^2}{\partial \theta^2} p(x|\theta) \right) dx - \int_{\mathcal{X}} \left(\frac{\partial}{\partial \theta} \log p(x|\theta) \right)^2 p(x|\theta)\, dx$$

$$= \frac{\partial^2}{\partial \theta^2} \underbrace{\int_{\mathcal{X}} p(x|\theta)\, dx}_{=1} - E_\theta\left[\left(\frac{\partial}{\partial \theta} \log p(X_1|\theta) \right)^2 \right]$$

$$= - E_\theta\left[\left(\frac{\partial}{\partial \theta} \log p(X_1|\theta) \right)^2 \right]$$

よって

$$\tilde{I}_{\underline{x}}(\underline{\theta}) = E_\theta\left[\left(\sum_{k_1=1}^{n}\frac{\partial}{\partial\theta}\log p(x_{k_1}|\theta)\right)\left(\sum_{k_2=1}^{n}\frac{\partial}{\partial\theta}p(x_{k_2}|\theta)\right)\right]$$

$$= \sum_{k=1}^{n}E_\theta\left[\left(\frac{\partial}{\partial\theta}\log p(x_k|\theta)\right)^2\right]$$

$$+ \sum_{k_1\neq k_2}E_\theta\left[\left(\frac{\partial}{\partial\theta}\log p(x_{k_1}|\theta)\right)\left(\frac{\partial}{\partial\theta}\log p(x_{k_2}|\theta)\right)\right]$$

$$= -\sum_{k=1}^{n}E_\theta\left[\frac{\partial^2}{\partial\theta^2}\log p(x_k|\theta)\right]$$

$$+\sum_{k_1 \neq k_2} \underbrace{E_\theta\left[\frac{\partial}{\partial\theta}\log p(x_{k_1}|\theta)\right]}_{=0} E_\theta\left[\frac{\partial}{\partial\theta}\log p(x_{k_2}|\theta)\right]$$

$$= -E_\theta\left[\frac{\partial^2}{\partial\theta^2}\log\prod_{k=1}^{n} p(x_k|\theta)\right]$$

$$= -E_\theta[\ddot{\ell}_n]. \qquad \square$$

(2) の証明　共分散公式から

$$\mathrm{Cov}\left[\frac{\partial}{\partial\theta}\log p(x_i|\theta), \frac{\partial}{\partial\theta}\log p(x_j|\theta)\right]$$

$$= E\left[\left(\frac{\partial}{\partial \theta} \log p(x_i \mid \theta)\right)\left(\frac{\partial}{\partial \theta} \log p(x_j \mid \theta)\right)\right]$$

$$- \underbrace{E\left[\frac{\partial}{\partial \theta} \log p(x_i \mid \theta)\right]}_{=0} E\left[\frac{\partial}{\partial \theta} \log p(x_j \mid \theta)\right]$$

6.2 指数型分布族

定義6.4 $d \in \mathbb{N}$ とし, $\mathcal{X} \subset \mathbb{R}^d$ は空でない部分集合とする. $\Theta \subset \mathbb{R}$ を母数空間とする. $\mathcal{X}$ 上の統計的モデル $\mathcal{P} = \{P_\theta ; \theta \in \Theta\}$ は 1母数指数型分布族であるとは, Θ 上の実数値関数 $A(\theta)$ と $u(\theta)$, $\mathcal{X}$ 上の実数値関数 T と h が存在して, P_θ の p.m.f. / p.d.f. $p(x \mid \theta)$ が次の形で書けるときをいう

$$p(x|\theta) = \begin{cases} h(x)\exp\{A(\theta)T(x) - u(\theta)\} & (x \in \mathcal{X}) \\ 0 & (x \notin \mathcal{X}) \end{cases}.$$

注意6.5. A, h, T の表現は一意的ではない.

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

$$p(x|\theta) = \begin{cases} \dfrac{\theta^x e^{-\theta}}{x!} = \dfrac{1}{x!}\exp\{x\log\theta - \theta\} & (x = 0, 1, 2, \ldots) \\ 0 & (\text{otherwise}) \end{cases}.$$

である. $\{Po(\theta);\ \theta \in (0,\infty)\}$ は

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

である.

例16.7 $X \sim Bino(n,\theta)\ (n\in\mathbb{N},\ 0<\theta<1)$

とする. X の p.m.f. は

$$p(x|\theta)=\binom{n}{x}\theta^{x}(1-\theta)^{n-x}$$

$$=\binom{n}{x}\exp\left\{x\log\left(\frac{\theta}{1-\theta}\right)+n\log(1-\theta)\right\}$$

$$(x = 0, 1, 2, \dots, n)$$

と書ける. ただし

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!}\ ; \quad 0! = 1.$$

したがって, $\{\mathrm{Bino}(n, \theta) : 0 < \theta < 1\}$ は

$$d = 1,\ A(\theta) = \log\left(\frac{\theta}{1-\theta}\right),\ u(\theta) = -n\log(1-\theta)$$

$$T(x) = x,\ h(x) = \binom{n}{x}.$$

例16.8　$\underline{X} = (Y, Z)^T$, $Z, W \overset{\text{i.i.d.}}{\sim} N(0,1)$ とし.

$Y := Z + \theta W$ ($\theta > 0$) とする.

$\underline{X}$ の p.d.f. $p^{\underline{X}}(y, z)$ は次で与えられる.

$\underline{x} = (y, z)^T$ としたとき.

$$p^{\underline{X}}(\underline{x} \mid \theta) = p^{Z}(z)\, p^{Y|Z}(y \mid z)$$

$$= \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \frac{1}{\sqrt{2\pi\theta^2}} \exp\left\{ - \frac{(y-z)^2}{2\theta^2} \right\}$$

$$= \frac{1}{2\pi\theta} \exp\left\{-\frac{1}{2}\left(z^2 + \frac{(y-z)^2}{\theta^2}\right)\right\}$$

$$= \frac{1}{2\pi} \exp\left\{-\frac{z^2}{2}\right\} \exp\left\{-\frac{(y-z)^2}{2\theta^2} - \log\theta\right\}$$

と書ける.

$$d=2,\quad A(\theta) = -\frac{1}{2\theta^2},\quad k(\theta) = \log\theta$$

$$T(\underline{x}) = T(y, z) = (y-z)^2$$

$$h(\underline{x}) = \frac{1}{2\pi}\exp\left\{-\frac{z^2}{2}\right\}.$$

□

$\{P_\theta ; \theta \in \Theta\}$ が

$$p(x|\theta) = h(x) \exp\{A(\theta)T(x) - u(\theta)\}$$

を満たすとする。

$X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} P_\theta$ とします

$\underline{X} = (X_1, X_2, \ldots, X_n)^T$ の同時 p.m.f / p.d.f.

を $\prod_{j=1}^{n} p(x_j|\theta) =: p^{\underline{X}}(\underline{x}|\theta)$ とします

$$\prod_{j=1}^{n} p(x_j|\theta) = \prod_{j=1}^{n} h(x_j) \exp\{A(\theta)T(x_j) - u(\theta)\}$$

$$= \left(\prod_{j=1}^{n} h(x_j)\right) \exp\left\{ A(\theta)\sum_{j=1}^{n} T(x_j) - n u(\theta) \right\}$$

によって、$\{P_\theta^{\otimes n} ; \theta \in \Theta\}$ は

$d=1$, $A(\theta)$, $\sum_{j=1}^{n} T(x_j)$, $nu(\theta)$, $\prod_{j=1}^{n} h(x_j)$

の1母数指数型分布族

指数型分布族の正準表示

$$p(x|\theta) = h(x)\exp\{A(\theta)T(x) - u(\theta)\} \quad (x \in \mathcal{X})$$

において

$$\eta = A(\theta)$$

とおくと

少し記号を乱用している.

$$p(x|\eta) = h(x)\exp\{\eta T(x) - u^{\vee}(\eta)\} \quad (x \in \mathcal{X})$$

と書ける. ただし, 形式的に

$$u^{\vee}(\eta) = \begin{cases} \log\left(\sum_{x \in \mathcal{X}} h(x) \exp\{\eta T(x)\}\right) & (\text{離散型}) \\ \log\left(\int_{\mathcal{X}} h(x) \exp\{\eta T(x)\} dx\right) & (\text{連続型}) \end{cases}$$

である。すると

$$u^{\vee}(\eta) = u^{\vee}(A(\theta)) = u(\theta)$$

がわかる。さらに

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

以上から

$$\{p(x|\eta) = h(x)\exp\{\eta T(x) - \psi(\eta)\} ; \eta \in \mathcal{E}\}$$

は1母数指数型分布族で、これを

正準指数型分布族または自然指数型分布族

という。$\mathcal{E}$ を自然母数空間という。

$\dot{\psi}(\eta) = \frac{d}{d\eta}\psi(\eta)$ としたとき、

$$\dot{\psi}(\mathcal{E}) := \{\dot{\psi}(\eta) ; \eta \in \mathcal{E}\}$$

を期待値母数空間とよぶ

例6.9（例6.6の続き）

$$p(x|\theta) = \frac{1}{x!}\exp\{x \underbrace{\log\theta}_{=\eta} - \theta\}$$

$$\eta = \log\theta \iff \theta = e^{\eta}$$

よって

$$p(x|\eta) = \frac{1}{x!}\exp\{\eta x - e^{\eta}\}$$

$$\exp(\kappa^{\vee}(\eta)) = \sum_{x=0}^{\infty}\frac{1}{x!}\exp\{\eta x\} = \sum_{x=0}^{\infty}\frac{(e^{\eta})^{x}}{x!} = \exp(e^{\eta})$$

よって、$\kappa^{\vee}(\eta) = e^{\eta}$ $\curvearrowright$ $\mathcal{E} = \{\eta \in \mathbb{R} : e^{\eta} < \infty\} = \mathbb{R}$.

さらに、$\dot{\kappa}^{\vee}(\eta) = e^{\eta}$ より

$$\dot{\kappa}^{\vee}(\Xi) = \underbrace{(0,\infty)}_{\text{もともとのPoisson分布の母数空間}}$$

もともとのPoisson分布の母数空間 い

$$\left\{ \frac{1}{x!} \exp\{\eta x - e^{\eta}\} : \eta \in \mathbb{R} \right\}$$

$$= \left\{ \frac{1}{x!} \exp\{x \log\theta - \theta\} : \theta \in (0,\infty) \right\}$$

がいえそう。

Poisson分布族を $\mathbb{R}$ と $(0,\infty)$ 別々の空間で

添え字付けている.

補題 6.10 $\{h(x)\exp(\eta T(x)-\psi(\eta)) : \eta\in\Xi\}$ を考える. このとき, ψ は Ξ の内部上で無限回微分可能である. さらに, 積分記号と微分記号の入れかえも可能.

注意 1母数指数型分布族は正則母数モデル!

68

<u>定理 6.11</u> $\{h(x)\exp\{\eta T(x) - \psi(\eta)\} : \eta \in \mathcal{E}^\circ\}$

$\mathcal{E}$の内部

に対して、$T(x)$の積率母関数は原点のまわりで存在して ($X \sim h(x)\exp\{\eta T(x) - \psi(\eta)\}$ とする)

$$M_T(s) = \exp\{\psi(s+\eta) - \psi(\eta)\}$$

となり

$$E[T(X)] = \dot{\psi}(\eta) = \frac{d\psi}{d\eta}(\eta)$$

$$Var[T(X)] = \ddot{\psi}(\eta) = \frac{d^2\psi}{d\eta^2}(\eta)$$

証明　$u^\vee$の定義より，$\eta \in \mathcal{E}^\circ$に対して

$$M_T(s) = E[\exp(sT(x))]$$

指数型のとき

$$= \int_{\mathcal{X}} \exp(sT(x))\, h(x) \exp\{\eta T(x) - u^\vee(\eta)\}\, dx$$

$$= \int_{\mathcal{X}} h(x) \exp[(s+\eta)T(x) - u^\vee(\eta)]\, dx$$

ηは$\mathcal{E}$の内点なので，sを十分小さくとると

$\eta + s$も$\mathcal{E}$の内点

$$= \exp\left[u^{\nu}(\lambda+\eta) - u^{\nu}(\eta) \right]$$

$$\times \underbrace{\int_{\mathcal{X}} h(x) \exp\left\{ (\lambda+\eta) T(x) - u^{\nu}(\lambda+\eta) \right\} dx}_{=1}$$

$$= \exp\left[u^{\nu}(\lambda+\eta) - u^{\nu}(\eta) \right] \quad //$$

ゆえに:

$$\frac{d u^{\nu}}{d\eta}(\eta) = \frac{d}{d\eta} \log\left(\int_{\mathcal{X}} h(x) \exp\{ \eta T(x) \} \, dx \right)$$

$$= \frac{\frac{d}{d\eta} \int_{\mathcal{X}} h(x) \exp\{\eta T(x)\} dx}{\int_{\mathcal{X}} h(x) \exp\{\eta T(x)\} dx}$$

$$= \frac{\int_{\mathcal{X}} \frac{d}{d\eta} h(x) \exp\{\eta T(x)\} dx}{\exp\{u^{\vee}(\eta)\}}$$

$$= \frac{\int_{\mathcal{X}} T(x) h(x) \exp\{\eta T(x)\} dx}{\exp\{u^{\vee}(\eta)\}}$$

$$= \int_{\mathcal{X}} T(x)\, h(x) \exp\{\eta T(x) - \psi(\eta)\}\, dx$$

$$= E[T(X)].$$

$\square$