第9章　Bayes的推測

9.1　Bayes 的推測の考え方

前節までの統計的手法は頻度論的手法 (frequentist methods) と呼ばれるものである．頻度論的アプローチの考え方は以下のようにまとめられる．

【F1】確率は大量に観測されたデータの度数の極限と考える．

【F2】真の母数は未知だが, 固定した値であると考える．したがって真の母数は変動しないし, 真の母数に対する意味のある確率的な主張はない．

【F3】推測手法は大量にデータが観測されれば, うまく機能する保証があるように設計されている．たとえば信頼係数 0.95 の信頼区間は, 大量に観測されたデータの極限的な頻度において, 0.95 の信頼度が保証されている．

この考え方とは異なるアプローチの 1 つが Bayes 的推測である．Bayes 的アプローチは以下のような考え方に基づいている．

【B1】確率は極限的な頻度ではなく, 信頼の程度を表現するものであると考える．データはある確率変数の実現値と考えるだけでなく, それ以外の色々なものも確率変数の実現値と考える．

【B2】真の母数は固定された定数にも関わらず, 真の母数の確率的な主張をする．

【B3】真の母数に対して確率分布を想定し, 未知の母数の推測を行う．点推定値や区間推定値もこの分布からの実現値と考える．

確率に主観的な見方を導入する Bayes 的アプローチには, 頻度論的立場の主流派 (古典的な) 統計学者からの大きな批判がある．しかし, 統計学の隣接分野である機械学習やデータマイニングの分野では Bayes 的なアプローチが広く採用されている．

哲学的な議論は脇において, Bayes 的推測がどのような形式で行われるかをこの節ではみていく．

9.2 Bayes 的推測手法

Bayes 的推測は以下のステップに従い行われる．母数モデルを $\{\mathsf{P}_\theta;\, \theta \in \Theta\}$ とする．

1. 事前分布 (prior distribution) と呼ばれる母数 θ についての分布 $\mathsf{p}^{\Theta}(\theta)$ を想定する．

2. 母数 θ に与えられたときにデータの分布を表現する条件付きの分布 (頻度論的な立場では真の分布) $\mathsf{p}^{\boldsymbol{X}|\Theta}(\boldsymbol{x}|\,\theta)$ を想定する．

3. 観測されたデータ $X_1 = x_1,\, X_2 = x_2,\, \ldots,\, X_n = x_n$ に基づき, θ の信頼度をアップデートする．すなわち, **事後分布** (posterior distribution) $\mathsf{p}^{\Theta|\boldsymbol{X}}(\theta|\,\boldsymbol{x})$ を求める．ただし $\boldsymbol{x} = (x_1,\, x_2,\, \ldots,\, x_n)$ とした．

そして, 事後分布に基づき, 母数の推測を行う．3 番目のステップがどのように行われるかを θ の事前分布とデータ $\boldsymbol{X}$ の分布が共に離散型の場合で説明する．

真の母数 θ は確率変数 Θ の実現値と考える．いまは離散型確率変数の設定なので

$$\begin{aligned}\mathsf{Pr}\big(\Theta = \theta|\, \boldsymbol{X} = \boldsymbol{x}\big) &= \frac{\mathsf{Pr}\big(\Theta = \theta,\, \boldsymbol{X} = \boldsymbol{x}\big)}{\mathsf{Pr}(\boldsymbol{X} = \boldsymbol{x})} \\ &= \frac{\mathsf{Pr}\big(\Theta = \theta,\, \boldsymbol{X} = \boldsymbol{x}\big)}{\sum_\theta \mathsf{Pr}\big(\boldsymbol{X} = \boldsymbol{x}\big|\, \Theta = \theta\big)\mathsf{Pr}\big(\Theta = \theta\big)}\end{aligned}$$

となる．最後の等号は全確率の法則 (補題 1.11) を用いた．これを連続型分布の場合に形式的に書き直すと

$$\mathsf{p}^{\boldsymbol{\Theta}|\boldsymbol{X}}(\theta|\,\boldsymbol{x}) = \frac{\mathsf{p}^{\boldsymbol{X}|\Theta}(\boldsymbol{x}|\,\theta)\mathsf{p}^{\Theta}(\theta)}{\int \mathsf{p}^{\boldsymbol{X}|\Theta}(\boldsymbol{x}|\,\theta)\mathsf{p}^{\Theta}(\theta)\,\mathrm{d}\theta}$$

と書ける．ただし $\mathsf{p}^{\boldsymbol{\Theta}|\boldsymbol{X}}(\,\cdot\,|\,\boldsymbol{x})$ は $\boldsymbol{X} = \boldsymbol{x}$ が与えられたときの $\boldsymbol{\Theta}$ の条件付き p.d.f. であり, θ の事後分布である．$\mathsf{p}^{\boldsymbol{X}|\Theta}(\boldsymbol{x}|\,\theta)$ は真の母数が θ のときのデータの分布の p.d.f. で, $\mathsf{p}^{\Theta}(\theta)$ は Θ の分布の p.d.f. である．

n 個のランダム標本 $\boldsymbol{X} = (X_1,\, X_2,\, \ldots,\, X_n)$ に対して, 真の母数が θ のときの $\boldsymbol{X}$ の同時 p.d.f. $\mathsf{p}^{\boldsymbol{X}|\Theta}(\,\cdot\,|\,\theta)$ とし, $X_1 = x_1,\, X_2 = x_2,\, \ldots,\, X_n = x_n$ を観測したときの尤度関数 $\mathsf{lik}_n(\theta)$ を

$$\mathsf{lik}_n(\theta) = \mathsf{p}^{\boldsymbol{X}|\Theta}(\boldsymbol{x}|\,\theta)$$

と書くことにする．よって

$$\mathsf{p}^{\Theta|\boldsymbol{X}}(\theta|\,\boldsymbol{x}) = \frac{\mathsf{p}^{\boldsymbol{X}|\Theta}(\boldsymbol{x}|\,\theta)\mathsf{p}^{\Theta}(\theta)}{\int \mathsf{p}^{\boldsymbol{X}|\Theta}(\boldsymbol{x}|\,\theta)\mathsf{p}^{\Theta}(\theta)\,\mathrm{d}\theta} =: \frac{\mathsf{lik}_n(\theta)\mathsf{p}^{\Theta}(\theta)}{c_n} \propto \mathsf{lik}_n(\theta)\mathsf{p}^{\Theta}(\theta)$$

と表すことができる. ただし

$$c_n := \int \mathsf{lik}_n(\theta)\mathsf{p}^{\Theta}(\theta)\,\mathrm{d}\theta$$

は正規化定数と呼ばれる値である. c_n は θ に依存せず, データの実現値に依存した値である. よって事後分布の条件付き p.d.f. は事前分布と尤度関数の積の定数倍である.

$$\mathsf{p}^{\mathbf{\Theta}|\boldsymbol{X}}(\theta|\,\boldsymbol{x}) \propto \mathsf{lik}_n(\theta)\mathsf{p}^{\Theta}(\theta).$$

c_n を無視してもよいのだろうか? 必要なときは求めることができるので問題ない. 事後分布の平均やモード (最頻値) を推測に用いることが多い. たとえば,

$$\int \{y-\theta\}^2 \mathsf{p}^{\mathbf{\Theta}|\boldsymbol{X}}(\theta|\,\boldsymbol{x})\,\mathrm{d}\theta$$

を最小にする y の値を $\overline{\theta}$ と書くと

$$\overline{\theta} = \frac{\int \theta \mathsf{p}^{\mathbf{\Theta}|\boldsymbol{X}}(\theta|\,\boldsymbol{x})\,\mathrm{d}\theta}{\int \theta^2\, \mathsf{p}^{\mathbf{\Theta}|\boldsymbol{X}}(\theta|\,\boldsymbol{x})\,\mathrm{d}\theta} = \frac{\int \theta\, \mathsf{lik}_n(\theta)\mathsf{p}^{\Theta}(\theta)\,\mathrm{d}\theta}{\int \theta^2\, \mathsf{lik}_n(\theta)\mathsf{p}^{\Theta}(\theta)\,\mathrm{d}\theta}$$

と表現できる. また信頼区間であれば, $0<\alpha<1/2$ に対して

$$\int_{-\infty}^{a} \mathsf{p}^{\mathbf{\Theta}|\boldsymbol{X}}(\theta|\,\boldsymbol{x})\,\mathrm{d}\theta = \int_{b}^{\infty} \mathsf{p}^{\mathbf{\Theta}|\boldsymbol{X}}(\theta|\,\boldsymbol{x})\,\mathrm{d}\theta = \frac{\alpha}{2}$$

をみたす $a,\,b$ を求めると開区間 $C=(a,\,b)$ は

$$\mathsf{Pr}\big(\theta \in C\big) = 1-\alpha$$

をみたす事後信頼区間となる.

例 9.1. $X_1,\,X_2,\,\ldots,\,X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Ber}(\theta)\,(0<\theta<1)$ とし, 事前分布は $(0,\,1)$ 上の一様分布とする. すなわち

$$\mathsf{p}^{\Theta}(\theta) = \begin{cases} 1 & (0<\theta<1) \\ 0 & (\text{その他の場合}) \end{cases}$$

である. すると

$$\begin{aligned} \mathsf{p}^{\mathbf{\Theta}|\boldsymbol{X}}(\theta|\,\boldsymbol{x}) \propto \mathsf{p}^{\Theta}(\theta)\mathsf{lik}_n(\theta) &= \prod_{j=1}^{n} \theta^{x_j}(1-\theta)^{1-x_j}\mathbb{1}_{\{0,\,1\}}(x_j) \\ &= \theta^{s_n}(1-\theta)^{n-s_n}\mathbb{1}_{\{0,\,1,\,\ldots,\,n\}}(s_n) \end{aligned}$$

となる. ただし $s_n = \sum_{j=1}^{n} x_j$ である. したがって

$$\Theta|\,\boldsymbol{X}=\boldsymbol{x} \,\sim\, \mathsf{Beta}(s_n+1,\,n-s_n+1)$$

となることがわかる.

一方 $\alpha > 0,\, \beta > 0$ とし

$$\mathsf{p}^{\Theta}(\theta|\,\alpha,\,\beta) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} & (x>0) \\ 0 & (\text{その他の場合}) \end{cases}$$

とする. ただし $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\,\mathrm{d}x$ である. すなわち θ の事前分布は $\mathsf{B}(\alpha,\,\beta)$ である. このとき

$$\begin{aligned} &\mathsf{p}^{\Theta|\boldsymbol{X}}(\theta|\,\boldsymbol{x}) \\ &\quad = \frac{\Gamma(n+\alpha+\beta)}{\Gamma(s_n+\alpha)\Gamma(n-s_n+\beta)}\theta^{(s_n+\alpha)-1}(1-\theta)^{(n-s+\beta)-1}\mathbb{1}_{\{0,\,1,\,\ldots,\,n\}}(s_n) \end{aligned} \tag{9.1}$$

となる. したがって

$$\Theta|\,\boldsymbol{X} = \boldsymbol{x} \,\sim\, \mathsf{B}(s_n+\alpha,\, n-s_n+\beta)$$

がわかる. このとき

$$\overline{\theta} = \int \theta\mathsf{p}^{\Theta|\boldsymbol{X}}(\theta|\,\boldsymbol{x})\,\mathrm{d}\theta$$

とおくと

$$\overline{\theta} = \frac{s_n+\alpha}{n+\alpha+\beta}$$

を得る. $\widehat{\theta} = \dfrac{s_n}{n},\, \widetilde{\theta} = \dfrac{1}{2}$ とおくと

$$\overline{\theta} = \lambda_n\widehat{\theta} + (1-\lambda_n)\widetilde{\theta}, \quad \lambda_n = \frac{n}{n+\alpha+\beta}$$

と書ける.

信頼係数 0.95 の θ の事後信頼区間 $C = (a,\, b)$ は

$$\int_a^b \mathsf{p}^{\Theta|\boldsymbol{X}}(\theta|\,\boldsymbol{x})\,\mathrm{d}\theta = 0.95$$

をみたす $a,\, b$ を数値計算で求めればよい. □

問 9.1. $X_1,\, X_2,\, \ldots,\, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Ber}(\theta)\ (0 < \theta < 1)$ とし,

$$\mathsf{p}^{\Theta}(\theta|\,\alpha,\,\beta) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} & (x>0) \\ 0 & (\text{その他の場合}) \end{cases}$$

とする. このとき, $X_1 = x_1,\, X_2 = x_2,\, \ldots,\, X_n = x_n$ が与えられたときの Θ の条件付き p.d.f. が (9.1) で与えられることを示せ.

問 9.2. $X \sim \mathsf{B}(\alpha, \beta)$ のとき, $\mathsf{E}[X] = \dfrac{\alpha}{\alpha+\beta}$ となることを示せ.

例 9.2. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}(\mu, \sigma^2)$ とする. ただし $\mu \in \mathbb{R}$ で σ^2 は既知とする. 事前分布として

$$\Theta \sim \mathsf{N}(a, b^2)$$

を仮定する. ただし $a \in \mathbb{R}, 0 < b < \infty$ である. すると

$$\begin{aligned} &\Theta| \boldsymbol{X} = \boldsymbol{x} \sim \mathsf{N}(\overline{\theta}, \tau^2), \qquad (9.2) \\ &\overline{\theta} = w\overline{x}_n + (1-w)a, \quad \overline{x}_n = \frac{1}{n}\sum_{j=1}^{n} x_j, \\ &w = \frac{\frac{1}{\mathsf{se}^2}}{\frac{1}{\mathsf{se}^2}+\frac{1}{b^2}}, \quad \frac{1}{\tau^2} = \frac{1}{\mathsf{se}^2} + \frac{1}{b^2}, \quad \mathsf{se} = \frac{\sigma}{\sqrt{n}} \end{aligned}$$

となる. 問 9.3 を参照のこと. ただし, $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$, $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ とした.

このように事前分布と事後分布が同じ母数モデルに属するとき, 事前分布はこのモデルに**随伴する** (conjugate) という. または, このような事前分布を随伴事前分布という.

$n \to \infty$ のとき $w \to 1$ かつ $\dfrac{\tau}{\mathsf{se}} \to 1$ となる. 標本数が大きいとき

$$\Theta| \boldsymbol{X} = \boldsymbol{x} \approx \mathsf{N}\big(\overline{x}_n, \mathsf{se}^2\big)$$

となる. また n を固定する. $b \to \infty$ としたとき

$$\Theta| \boldsymbol{X} = \boldsymbol{x} \approx \mathsf{N}\big(\overline{x}_n, \mathsf{se}^2\big)$$

となる. これは一様な事前分布に対応するものである.

区間 $C = (c, d)$ は

$$\mathsf{Pr}\big(\theta \in C \big| \boldsymbol{X} = \boldsymbol{x}\big) = 0.95$$

をみたすものとする. したがって

$$\mathsf{Pr}\big(\theta < c \big| \boldsymbol{X} = \boldsymbol{x}\big) = 0.025, \quad \mathsf{Pr}\big(\theta > d \big| \boldsymbol{X} = \boldsymbol{x}\big) = 0.025$$

となるように c, d を選べばよい. このことから c を以下をみたすように選べばよい.

$$\mathsf{Pr}\big(\theta < c \big| \boldsymbol{X} = \boldsymbol{x}\big) = \mathsf{Pr}\left(\frac{\theta - \overline{\theta}}{\tau} < \frac{c-\overline{\theta}}{\tau}\middle| \boldsymbol{X} = \boldsymbol{x}\right) = \mathsf{Pr}\left(Z < \frac{c-\overline{\theta}}{\tau}\right).$$

ただし $Z \sim \mathsf{N}(0,1)$ である．さらに

$$\mathsf{Pr}(Z < -1.96) = 0.025$$

なので

$$\frac{c - \overline{\theta}}{\tau} = -1.96$$

とすればよい．よって

$$c = \overline{\theta} - 1.96\tau$$

とする．同様に

$$d = \overline{\theta} + 1.96\tau$$

を得る．これらのことから信頼係数 0.95 の Bayes 的信頼区間は $(\overline{\theta} - 1.96\tau,\, \overline{\theta} + 1.96\tau)$ となる．さらに $\widehat{\theta} \approx \widetilde{\theta}$ かつ $\tau \approx \mathsf{se}$ なので信頼係数 0.95 の Bayes 的信頼区間は近似的に $(\widehat{\theta} - 1.96\mathsf{se},\, \widehat{\theta} + 1.96\mathsf{se})$ となり，頻度論的信頼区間と同じになる．

問 9.3. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}(\mu, \sigma^2)$ とする．ただし $\mu \in \mathbb{R}$ で σ^2 は既知とする．さらに，事前分布として

$$\Theta \sim \mathbf{N}(a, b^2)$$

を仮定する．ただし $a \in \mathbb{R}$, $0 < b < \infty$ である．このとき，$\boldsymbol{X} = \boldsymbol{x}$ が与えられたときの Θ の条件付き分布が (9.2) で与えられることを確かめよ．

9.3 事前分布の選択について

Bayes 的推測における重要な疑問は，事前分布 π の選択である．ある学派は，母数 θ に関する主観的な意見を反映させて事前分布を選択すればよいと主張している．ある場合には，このやり方は実行可能かもしれない．しかし，多変数の母数を含む複雑な問題に対しては，実行可能性に欠けることもあるだろう．さらに，客観的かつ科学的な決定を目標とするデータ解析において，主観的な意見を入り込ませることは相容れないという見方ができるであろう．

別の方策として，ある種の無情報事前分布を用いるやり方がある．たとえば，Bernoulli 試行において，母数 $\theta\,(0 < \theta < 1)$ に対する事前分布 π を

$$\mathsf{p}^{\Theta}(\theta) \equiv 1$$

と定まる．このとき，形式的な計算から

$$\Theta |\, \mathcal{D}_n \sim \mathsf{Beta}(s_n + 1,\, n - s_n + 1)$$

を得る．ただし，$\mathcal{D}_n = \{x_1, x_2, \ldots, x_n\}$ で $s_n = \sum_{j=1}^{n} x_j$ である．

9.3.1 一様事前分布

$X \sim \mathsf{N}(\theta, \sigma^2)$ とする. ただし, σ は既知とする. いま, $\mathcal{D}_n$ を $\mathsf{N}(\theta, \sigma^2)$ からの標本の大きさが n 個のランダム標本の実現値とする. θ の事前分布 p^{Θ} を

$$\mathsf{p}^{\Theta}(\theta) \propto c$$

を採用する. ただし, $c > 0$ は定数である. このとき

$$\int \mathsf{p}^{\Theta}(\theta)\,\mathrm{d}\theta = \infty$$

となるので, p^{Θ} は p.d.f. ではない. しかし, 形式的な計算により

$$\mathsf{p}^{\Theta|\boldsymbol{X}}(\theta|\,\mathcal{D}_n) \ \propto\ \mathsf{lik}_n(\theta)\mathsf{p}^{\Theta}(\theta) \ \propto\ \mathsf{lik}_n(\theta)$$

となることがわかる. $\overline{x}_n = (1/n)\sum_{j=1}^n x_j$ とおいたとき

$$\Theta|\,\mathcal{D}_n \ \sim\ \mathsf{N}\left(\overline{x}_n,\ \frac{\sigma^2}{n}\right)$$

となることがわかる. なぜならば

$$\begin{aligned}
\mathsf{lik}_n(\theta)\mathsf{p}^{\Theta}(\theta) &= c\prod_{j=1}^n \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x_j-\theta)^2}{2\sigma^2}\right\} \\
&= c\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\frac{\sum_{j=1}^n (x_j-\overline{x}_n)^2}{2\sigma^2} - \frac{(\overline{x}_n-\theta)^2}{2\sigma^2/n}\right\} \\
&\propto \exp\left\{-\frac{(\overline{x}_n-\theta)^2}{2\sigma^2/n}\right\}
\end{aligned}$$

であるからである. 以上の議論から Bayes 的推定量は頻度論的な推定量である標本平均と同じになる.

しかし, 一様事前分布は母数の変換に関して不変ではない. たとえば, $X \sim \mathsf{Ber}(\theta)\,(0 < \theta < 1)$ とし, 事前分布 p^{Θ} として

$$\mathsf{p}^{\Theta}(\theta) = \begin{cases} 1 & (0 < \theta < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

を採用する. いま, 母数の変換 Ψ を

$$\Psi := \Psi(\theta) = \log\left(\frac{\theta}{1-\theta}\right) \ \Leftrightarrow\ \Psi := \log\left(\frac{\Theta}{1-\Theta}\right)$$

と定める．すると

$$\begin{aligned}
\mathsf{p}^{\Psi}(\psi) &= \frac{\mathrm{d}}{\mathrm{d}\psi}\mathsf{Pr}^{\Psi}(\Psi \le \psi) = \frac{\mathrm{d}}{\mathrm{d}\psi}\mathsf{Pr}^{\Psi}\left(\log\left(\frac{\Theta}{1-\Theta}\right) \le \psi\right) \\
&= \frac{\mathrm{d}}{\mathrm{d}\psi}\mathsf{Pr}\left(\Theta \le \frac{e^{\psi}}{1+e^{\psi}}\right) = \frac{\mathrm{d}}{\mathrm{d}\psi}\left(\frac{e^{\psi}}{1+e^{\psi}}\right) = \frac{e^{\psi}}{(1+e^{\psi})^2} \qquad (\psi \in \mathbb{R})
\end{aligned}$$

となるので, Θ の分布は一様分布であるので, Θ についての事前情報はない．Θ に関して無知であれば, Θ の変換である Ψ に関しても事前情報はないはずである．しかし, Ψ の分布は一様分布でないので, Ψ についての事前情報があることになる．したがって, Ψ についても事前情報がないことがないことと p^{Ψ} の分布の間には齟齬が生じることになる．

9.3.2 Jeffreys 事前分布

母数の変換に関して不変な事前分布を定めるために, Jeffreys を

$$\mathsf{p}^{\Theta}(\theta) \ \propto \ \sqrt{|\mathcal{F}(\theta)|}$$

で定める．ここで, $\mathcal{F}(\theta)$ は Fisher 情報行列で

$$\mathcal{F}(\theta) = \mathsf{E}\left[\left(\frac{\partial}{\partial\theta}\log \mathsf{lik}_n(\theta|\,\boldsymbol{X})\right)\left(\frac{\partial}{\partial\theta}\log \mathsf{lik}_n(\theta|\,\boldsymbol{X})\right)^{\top}\middle|\,\Theta = \theta\right]$$

で定義され, $|\cdot|$ は行列式を表す．また, 期待値は $\boldsymbol{X}$ の分布に関して取っていることに注意せよ．

定理 9.3. Jeffreys 事前分布は母数の変換に対して不変である．

Proof. □

例 9.4. Benoulli 分布を考える．すると

$$\begin{aligned}
\mathcal{F}(\theta) &= \mathsf{E}\left[\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\log\bigl(\theta^{X}(1-\theta)^{1-X}\bigr)\right)^2\middle|\,\Theta = \theta\right] \\
&= \mathsf{E}\left[\left(\frac{X}{\theta} - \frac{1-X}{1-\theta}\right)^2\middle|\,\Theta = \theta\right] \\
&= \mathsf{E}\left[\left(\frac{X-\theta}{\theta(1-\theta)}\right)^2\middle|\,\Theta = \theta\right] = \frac{1}{\theta(1-\theta)}
\end{aligned}$$

となる．よって, Jeffreys 事前分布 p^{Θ} は

$$\mathsf{p}^{\Theta}(\theta) \ \propto \ \sqrt{|\mathcal{F}(\theta)|} = \theta^{-1/2}(1-\theta)^{-1/2}$$

となる．これは $\mathsf{Beta}(1/2, 1/2)$ の p.d.f. であり, 一様な分布に非常に近い． □

9.3.3 共役事前分布

定義 9.5. 事前分布は**共役**であるとは, 事後分布が事前分布の分布族に属すときをいう. すなわち, 事前分布の分布族を $\mathcal{P}$ としたとき

$$\mathsf{p}^{\Theta|\boldsymbol{X}}(\,\cdot\,|\,\boldsymbol{x}) \in \mathcal{P}$$

が成立することである. ただし, $\mathsf{p}_\theta^{\boldsymbol{X}}$ を標本分布の p.d.f. で

$$\mathsf{p}^{\Theta|\boldsymbol{X}}(\theta|\,\boldsymbol{x}) = \frac{\mathsf{p}_\theta^{\boldsymbol{X}|\Theta}(\boldsymbol{x})\mathsf{p}^{\Theta}(\theta)}{\int \mathsf{p}_\theta^{\boldsymbol{X}|\Theta}(\boldsymbol{x})\mathsf{p}^{\Theta}(\theta)\,\mathrm{d}\theta}$$

である.

指数型分布族に対する共役事前分布を導出してみよう. σ 有限な測度 μ に関する p.d.f. $\mathsf{p}^{\boldsymbol{X}|\Theta}(\,\cdot\,|\,\theta)$ は

$$\mathsf{p}^{\boldsymbol{X}|\Theta}(\boldsymbol{x}|\,\boldsymbol{\theta}) = \exp\bigl\{\boldsymbol{\theta}^\top \boldsymbol{x} - \kappa(\boldsymbol{\theta})\bigr\} \tag{9.3}$$

で与えられたとする. ただし, $\boldsymbol{\theta} \in \mathbb{R}^d$ で, 母数空間 $\Theta \subset \mathbb{R}^d$ は部分集合

$$\left\{\boldsymbol{\theta} \in \mathbb{R}^d;\, \kappa(\boldsymbol{\theta}) = \log\left(\int \exp\bigl\{\boldsymbol{\theta}^\top \boldsymbol{x} - \kappa(\boldsymbol{x})\bigr\}\,\mathrm{d}\mu(\boldsymbol{x})\right) < \infty\right\}$$

の内部で定義する.

定理 9.6. (9.3) で与えられる指数型分布族に対する共役事前分布の p.d.f. は

$$\mathsf{p}_{\boldsymbol{x}_0,\,n_0}^{\Theta}(\boldsymbol{\theta}) = \frac{\exp\Bigl(n_0 \boldsymbol{x}_0^\top \boldsymbol{\theta} - n_0 \kappa(\boldsymbol{\theta})\Bigr)}{\displaystyle\int \exp\Bigl(n_0 \boldsymbol{x}_0^\top \boldsymbol{\theta} - n_0 \kappa(\boldsymbol{\theta})\Bigr)\,\mathrm{d}\mu(\boldsymbol{\theta})}$$

で与えられる. ただし, $\boldsymbol{x}_0 \in \mathbb{R}^d$, $n_0 \in \mathbb{R}$ である.

Proof. 以下の計算からわかる.

$$\begin{aligned}
\mathsf{p}^{\Theta|\boldsymbol{X}}(\boldsymbol{x}|\,\boldsymbol{\theta})\mathsf{p}_{\boldsymbol{x}_0,\,n_0}^{\Theta}(\boldsymbol{\theta}) &\propto \exp\Bigl(\boldsymbol{\theta}^\top \boldsymbol{x} - \kappa(\boldsymbol{\theta})\Bigr)\exp\Bigl(n_0 \boldsymbol{x}_0^\top \boldsymbol{\theta} - n_0 \kappa(\boldsymbol{\theta})\Bigr) \\
&= \exp\Bigl((\boldsymbol{x} + \boldsymbol{x}_0)^\top \boldsymbol{\theta} - (1 + n_0)\kappa(\boldsymbol{\theta})\Bigr) \\
&= \exp\left\{(1 + n_0)\left(\frac{1}{1 + n_0}\boldsymbol{x} + \frac{n_0}{1 + n_0}\boldsymbol{x}_0\right)^\top - (1 + n_0)\kappa(\boldsymbol{\theta})\right\}\Bigr) \\
&\propto \mathsf{p}^{\Theta}_{\frac{1}{1+n_0}\boldsymbol{x} + \frac{n_0}{1+n_0}\boldsymbol{x}_0,\,1+n_0}(\boldsymbol{\theta}).
\end{aligned}$$

□

$\boldsymbol{X}_1,\,\boldsymbol{X}_2,\,\ldots,\,\boldsymbol{X}_n$ を (9.3) からの標本の大きさが n のランダム標本とする. すると $(\boldsymbol{X}_1,\,\boldsymbol{X}_2,\,\ldots,\,\boldsymbol{X}_n)$ の同時 p.d.f. は

$$\prod_{j=1}^{n} \mathsf{p}^{\boldsymbol{X}|\Theta}(\boldsymbol{x}_j|\,\boldsymbol{\theta}) = \exp\Bigl(n\boldsymbol{\theta}^\top\overline{\boldsymbol{x}}_n - n\kappa(\boldsymbol{\theta})\Bigr), \qquad \overline{\boldsymbol{x}}_n = \frac{1}{n}\sum_{j=1}^{n}\boldsymbol{x}_j$$

となるので

$$\begin{aligned}
&\mathsf{p}^{\Theta|(\boldsymbol{X}_1,\boldsymbol{X}_2,\ldots,\boldsymbol{X}_n)}\bigl(\boldsymbol{\theta}|\,\boldsymbol{x}_1,\,\boldsymbol{x}_2,\,\ldots,\,\boldsymbol{x}_n\bigr)\\
&\qquad\propto \exp\Bigl(n\boldsymbol{\theta}^\top\overline{\boldsymbol{x}}_n - n\kappa(\boldsymbol{\theta})\Bigr)\exp\Bigl(n_0\boldsymbol{x}_0^\top\boldsymbol{\theta} - n_0\kappa(\boldsymbol{\theta})\Bigr)\\
&\qquad= \exp\biggl((n+n_0)\Bigl(\frac{n}{1+n_0}\overline{\boldsymbol{x}}_n + \frac{n_0}{1+n_0}\boldsymbol{x}\Bigr) - (n+n_0)\kappa(\boldsymbol{\theta})\biggr)
\end{aligned}$$

となる. よって, 事後分布の p.d.f. は $\mathsf{p}^{\Theta}_{\frac{n}{n+n_0}\overline{\boldsymbol{x}}_n+\frac{n_0}{n+n_0}\boldsymbol{x}_0,\,n+n_0}(\boldsymbol{\theta})$ で与えらることがわかる.

いま, $\boldsymbol{\theta} = (\theta_1,\,\theta_2,\,\ldots,\,\theta_d)^\top$ と $\nabla = \left(\dfrac{\partial}{\partial\theta_1},\,\dfrac{\partial}{\partial\theta_2},\,\ldots,\,\dfrac{\partial}{\partial\theta_d}\right)^\top$ としたとき

$$\nabla\mathsf{p}^{\Theta}_{\boldsymbol{x}_0,\,n_0}(\boldsymbol{\theta}) = n_0\bigl(\boldsymbol{x}_0 - \nabla\kappa(\boldsymbol{\theta})\bigr)\mathsf{p}^{\Theta}_{\boldsymbol{x}_0,\,n_0}(\boldsymbol{\theta})$$

となる. 一方

$$\int \nabla\mathsf{p}^{\Theta}_{\boldsymbol{x}_0,\,n_0}(\boldsymbol{\theta})\,\mathrm{d}\mu(\boldsymbol{\theta}) = \nabla\underbrace{\int \mathsf{p}^{\Theta}_{\boldsymbol{x}_0,\,n_0}(\boldsymbol{\theta})\,\mathrm{d}\mu(\boldsymbol{\theta})}_{=1} = \mathbf{0}_d$$

となる. これらの 2 つの等式から

$$\begin{aligned}
\mathsf{E}\bigl[\nabla\kappa(\boldsymbol{\theta})\bigr] &= \mathsf{E}\biggl[\boldsymbol{x}_0 - \frac{1}{n_0}\nabla\mathsf{p}^{\Theta}_{\boldsymbol{x}_0,\,n_0}(\boldsymbol{\theta})\biggr]\\
&= \boldsymbol{x}_0 - \frac{1}{n_0}\int\nabla\mathsf{p}^{\Theta}_{\boldsymbol{x}_0,\,n_0}(\boldsymbol{\theta})\,\mathrm{d}\mu(\boldsymbol{\theta})\\
&= \boldsymbol{x}_0
\end{aligned}$$

を得る. ただし, 期待値は $\mathsf{p}^{\Theta}_{\boldsymbol{x}_0,\,n_0}(\boldsymbol{\theta})$ に関してである. もっと一般的には

$$\mathsf{E}\bigl[\nabla\kappa(\boldsymbol{\theta})\bigm|\boldsymbol{X}_1,\,\boldsymbol{X}_2,\,\ldots,\,\boldsymbol{X}_n\bigr] = \frac{n}{n+n_0}\overline{\boldsymbol{X}}_n + \frac{n_0}{n+n_0}\boldsymbol{x}_0$$

となる.

9.4 1 次元の van Trees 不等式

母数空間 Θ を $\mathbb{R}$ の開部分集合とする状態空間 $(\mathbb{X}, \mathcal{B})$ 上の統計的モデル

$$\mathcal{P} := \left\{\mathsf{P}_\theta;\, \theta \in \Theta, \mathsf{P}_\theta \text{ は } (\mathbb{X}, \mathcal{B}) \text{ 上の確率分布で}, \mathsf{P}_\theta \ll \mu\right\}$$

を考える．ただし, μ は $(\mathbb{X}, \mathcal{B})$ 上の σ 有限な測度である．さらに

$$\mathsf{p}_\theta := \frac{\mathrm{d}\mathsf{P}_\theta}{\mathrm{d}\mu}$$

とし, 写像

$$\Theta \times \mathbb{X} \ni (\theta, x) \mapsto \mathsf{p}_\theta(x)$$

は可測[1]と仮定する．さらに

$$\sqrt{\mathsf{p}_\theta} \in L^2(\mu)$$

と仮定する．以降, $\|\cdot\|_\mu$ は $L^2(\mu)$ における L^2 ノルムとする．すなわち, $L^2(\mu)$ の元 $g : \mathbb{X} \longrightarrow \mathbb{R}$ に対して

$$\|g\|_\mu = \sqrt{\int_{\mathbb{X}} g^2(x)\,\mathrm{d}\mu(x)}$$

と定める．また, $\mathbb{X}$ のノルムを $\|\cdot\|$ と記すことにする．

定義 9.7 (L^2 における可微分性)**.** σ 有限な測度 μ によって支配される統計的モデル $\mathcal{P}$ は $\theta_0 \in \Theta$ で $L^2(\mu)$ **に関して微分可能**であるとは, 任意の $\sqrt{\mathsf{p}_\theta} \in L^2(\mu)$ に対して, ある $\dot{\sqrt{\mathsf{p}}}_{\theta_0} \in L^2(\mu)$ が存在して,

$$\left\|\sqrt{\mathsf{p}_\theta} - \sqrt{\mathsf{p}_{\theta_0}} - (\theta - \theta_0)\, \dot{\sqrt{\mathsf{p}}}_{\theta_0}\right\|_\mu = o(|\theta - \theta_0|) \qquad (\theta \to \theta_0)$$

をみたすときをいう．ここで, $\dot{\sqrt{\mathsf{p}}}_{\theta_0}$ を θ_0 **における** $L^2(\mu)$ **微分**とよぶ.

注意 9.8. 統計的モデル $\mathcal{P}$ が $L^2(\mu)$ に関して微分可能のとき, $\theta_0\,(\in \Theta)$ における Fisher 情報量 $\mathcal{F}_{\mathcal{P}}(\theta_0)$ は

$$\begin{aligned}
\mathcal{F}_{\mathcal{P}}(\theta_0) &= \int_{\mathbb{X}} \left\{\frac{\mathrm{d}}{\mathrm{d}\theta}\log \mathsf{p}_\theta(x)\bigg|_{\theta=\theta_0}\right\}^2 \mathsf{p}_{\theta_0}(x)\,\mathrm{d}\mu(x) \\
&= 4\int_{\mathbb{X}} \left\{\dot{\sqrt{\mathsf{p}}}_{\theta_0}(x)\right\}^2 \mathrm{d}\mu(x) = \int_{\mathbb{X}} \left\{\frac{\dot{\mathsf{p}}_\theta(x)}{\mathsf{p}_\theta(x)}\right\}^2 \mathrm{d}\mu(x)
\end{aligned}$$

と表現できる.

[1] $\|\cdot\|_\mu$ によって誘導される位相における開集合族の Borel 集合族に関して可測と考えるのが自然であろう.

定義 9.9. 開集合 $\Theta\,(\subset \mathbb{R})$ 上に集中する確率分布 Q は**よい振る舞いの事前分布**であるとは, Q は $\mathbb{R}$ 上の Lebesgue 測度に関して絶対連続な密度関数 q を持ち, さらにほとんど至る所存在する導関数 $\overset{\bullet}{\mathsf{q}}$ を持ち

$$\mathcal{F}_{\mathsf{Q}} := \int_{\Theta} \big\{\overset{\bullet}{\mathsf{q}}\,(\theta)\big\}^2 \frac{\mathbb{1}\{\mathsf{q}(\theta) > 0\}}{\mathsf{q}(\theta)}\,\mathrm{d}\mathsf{m}(\theta) < \infty$$

をみたすときをいう. ただし, m は $\mathbb{R}$ 上の Lebesgue 測度である.

統計量 $T : \mathbb{X} \longrightarrow \mathbb{R}$ は関数 $\psi(\theta)$ の推定量とする. ここで, ψ は絶対連続で, ほとんど至る所導関数 $\overset{\bullet}{\psi}$ をもつとする. E_θ は P_θ に関する期待値としたとき, van Trees 不等式は

$$\int_{\Theta} \mathsf{E}_\theta\big[\{T - \psi(\theta)\}^2\big]\,\mathrm{d}\mathsf{Q}(\theta) \geq \frac{\left(\int_{\Theta} \overset{\bullet}{\psi}\,(\theta)\,\mathrm{d}\mathsf{Q}(\theta)\right)^2}{\mathcal{F}_{\mathsf{Q}} + \int_{\Theta} \mathcal{F}_{\mathcal{P}}(\theta)\,\mathrm{d}\mathsf{Q}(\theta)} \tag{9.4}$$

と表現できる. この不等式のために以下の仮定をおく.

仮定 9.10. (1) Θ は $\mathbb{R}$ の開集合.
(2) 統計的モデル $\mathcal{P} = \{\mathsf{P}_\theta;\, \theta \in \Theta\}$ は $(\mathbb{X}, \mathcal{B})$ 上の σ 有限測度 μ によって支配される.
(3) μ に関する P_θ の密度を $\mathsf{p}_\theta = \dfrac{\mathrm{d}\mathsf{P}_\theta}{\mathrm{d}\mu}$ と書いたとき, 写像

$$\Theta \times \mathbb{X} \ni (\theta, x) \;\mapsto\; \mathsf{p}_\theta(x)$$

は可測である.
(4) $\mathcal{P}$ は $\Theta \cap \mathrm{supp}(\mathsf{q})$ 上のほとんど至る所で可微分である. ただし, $\mathrm{supp}(\mathsf{q}) := \left\{\theta \in \Theta;\, \dfrac{\mathrm{d}\mathsf{Q}}{\mathrm{dm}} = \mathsf{q}(\theta) > 0\right\}$ であり, m は $\mathbb{R}$ 上の Lebesgue 測度である.
(5) 関数 $\psi : \Theta \longrightarrow \mathbb{R}$ は絶対連続である.
(6) 関数 ψ^2 と $\overset{\bullet}{\psi}$ は Q 可積である.
(7)

$$\int_{\Theta} \mathsf{E}_\theta[T^2]\,\mathrm{d}\mathsf{Q}(\theta) < \infty; \qquad \int_{\Theta} \mathcal{F}_{\mathcal{P}}(\theta)\,\mathrm{d}\mathsf{Q}(\theta) < \infty$$

である.
(8) 任意の $B \in \mathcal{B}$ に対して, 関数

$$\Theta \cap \mathrm{supp}(\mathsf{q}) \ni \theta \mapsto \mathsf{P}_\theta(B)$$

は絶対連続である.
(9) θ が Θ の有限な境界に近づくとき, $\mathsf{q}(\theta) \to 0$ が成立する.

定理 9.11. 仮定 9.10 が成立し, $\mathcal{F}_{\mathsf{Q}} > 0$ とする. このとき, van Trees 不等式 (9.4) が成立する.

証明のために次の補題を準備する.

補題 9.12. 統計的モデル $\mathcal{P} = \{\mathsf{P}_\theta; \theta \in \Theta, \mathsf{P}_\theta \ll \mu\}$ は $\theta_0 \in \Theta$ において, $L^2(\mu)$ に関して可微分とする. 有界な統計量 $T : \mathbb{X} \longrightarrow \mathbb{R}$ を考える. すなわち, ある $M > 0$ が存在して, $\mu\big\{|T| > M\big\} = 0$ である. このとき

$$\gamma_T : \Theta \ni \theta \mapsto \mathsf{E}_\theta[T]$$

は θ_0 において微分可能で, 導関数

$$\dot{\gamma}_T\,(\theta_0) = 2\int_{\mathbb{X}} \dot{\sqrt{\mathsf{p}}}_{\theta_0}\,(x)\sqrt{\mathsf{p}}_{\theta_0}(x)T(x)\,\mathrm{d}\mu(x) \tag{9.5}$$

をもつ.

Proof. 導関数を求めるために

$$\mathsf{p}_\theta - \mathsf{p}_{\theta_0} - 2(\theta-\theta_0)\,\dot{\sqrt{\mathsf{p}}}_{\theta_0}\,\sqrt{\mathsf{p}}_{\theta_0}$$

を評価する. そのために, $r_\theta = \sqrt{\mathsf{p}}_\theta - \sqrt{\mathsf{p}}_{\theta_0} - (\theta-\theta_0)\,\dot{\sqrt{\mathsf{p}}}_{\theta_0}$ とおく. すると

$$\begin{aligned}
\mathsf{p}_\theta - \mathsf{p}_{\theta_0} - 2(\theta-\theta_0)\,\dot{\sqrt{\mathsf{p}}}_{\theta_0}\,\sqrt{\mathsf{p}}_{\theta_0} &= \big\{\sqrt{\mathsf{p}}_\theta\big\}^2 - \big\{\sqrt{\mathsf{p}}_{\theta_0}\big\}^2 - 2(\theta-\theta_0)\,\dot{\sqrt{\mathsf{p}}}_{\theta_0}\,\sqrt{\mathsf{p}}_{\theta_0}\\
&= \big\{\sqrt{\mathsf{p}}_{\theta_0} + (\theta-\theta_0)\,\dot{\sqrt{\mathsf{p}}}_{\theta_0} + r_\theta\big\}^2 - \big\{\sqrt{\mathsf{p}}_{\theta_0}\big\}^2\\
&\quad - 2(\theta-\theta_0)\,\dot{\sqrt{\mathsf{p}}}_{\theta_0}\,\sqrt{\mathsf{p}}_{\theta_0}\\
&= (\theta-\theta_0)^2\big\{\dot{\sqrt{\mathsf{p}}}_{\theta_0}\big\}^2 + r_\theta^2 + 2\sqrt{\mathsf{p}}_{\theta_0} r_\theta\\
&\quad + 2(\theta-\theta_0)\,\dot{\sqrt{\mathsf{p}}}_{\theta_0}\,r_\theta
\end{aligned} \tag{9.6}$$

となる. 最右辺を各項を積分したものを評価する.

$$\begin{aligned}
\int_{\mathbb{X}} (\theta-\theta_0)^2\big\{\dot{\sqrt{\mathsf{p}}}_{\theta_0}\big\}^2\,\mathrm{d}\mu &= O(|\theta-\theta_0|^2), \qquad (\because \mathcal{P}\text{ の可微分性})\\
\int_{\mathbb{X}} r_\theta^2\,\mathrm{d}\mu &= o(|\theta-\theta_0|^2), \qquad (\because \mathcal{P}\text{ の可微分性})\\
\left|\int_{\mathbb{X}} \sqrt{\mathsf{p}}_{\theta_0} r_\theta\,\mathrm{d}\mu\right| &\le \underbrace{\sqrt{\int_{\mathbb{X}} \mathsf{p}_{\theta_0}\,\mathrm{d}\mu}}_{=1}\sqrt{\int_{\mathbb{X}} r_\theta^2\,\mathrm{d}\mu} = o(|\theta-\theta_0|)\\
\left|\int_{\mathbb{X}} (\theta-\theta_0)\,\dot{\sqrt{\mathsf{p}}}_{\theta_0}\,r_\theta\,\mathrm{d}\mu\right| &\le |\theta-\theta_0|\sqrt{\int_{\mathbb{X}} \big\{\dot{\sqrt{\mathsf{p}}}_{\theta_0}\big\}^2\,\mathrm{d}\mu}\sqrt{\int_{\mathbb{X}} r_\theta^2\,\mathrm{d}\mu}\\
&= o(|\theta-\theta_0|^2)
\end{aligned}$$

となる. 以上から

$$\begin{aligned}
&\left|\gamma_S(\theta) - \gamma_S(\theta_0) - 2(\theta-\theta_0)\int_{\mathbb{X}} \dot{\sqrt{\mathsf{p}}}_{\theta_0}(x)\sqrt{\mathsf{p}}_\theta(x) S(x)\,\mathrm{d}\mu(x)\right| \\
&\quad = \Bigg| \int_{\mathbb{X}} S(x)\mathsf{p}_\theta(x)\,\mathrm{d}\mu(x) - \int_{\mathbb{X}} S(x)\mathsf{p}_{\theta_0}(x)\,\mathrm{d}\mu(x) \\
&\qquad\quad - 2(\theta-\theta_0)\int_{\mathbb{X}} \dot{\sqrt{\mathsf{p}}}_{\theta_0}(x) r_\theta(x) S(x)\,\mathrm{d}\mu(x)\Bigg| \\
&\quad \le M\Bigg| \int_{\mathbb{X}} \mathsf{p}_\theta(x)\,\mathrm{d}\mu(x) - \int_{\mathbb{X}} \mathsf{p}_{\theta_0}(x)\,\mathrm{d}\mu(x) \\
&\qquad\quad - 2(\theta-\theta_0)\int_{\mathbb{X}} \dot{\sqrt{\mathsf{p}}}_{\theta_0}(x) r_\theta(x)\,\mathrm{d}\mu(x)\Bigg| \\
&\quad = o(|\theta-\theta_0|)
\end{aligned}$$

がわかる. 以上から, $\gamma_S(\theta_0)$ は θ_0 で可微分で, (9.5) で与えられる導関数 $\dot{\gamma}_S(\theta_0)$ をもつことがわかる. □

<u>定理 9.11 の証明</u>: いま

$$\begin{aligned}
\Delta : \mathbb{X}\times\Theta \ni (x,\,\theta) \mapsto \dot{\mathsf{q}}(\theta)\frac{\mathbb{1}\{\mathsf{q}(\theta)>0\}}{2\sqrt{\mathsf{q}(\theta)}}\sqrt{\mathsf{p}_\theta(x)} \\
+\sqrt{\mathsf{q}(\theta)}\,\dot{\sqrt{\mathsf{p}_\theta}}(x)
\end{aligned} \tag{9.7}$$

とおく. ただし, ほとんど至所の $\theta \in \Theta_q := \Theta \cap \mathrm{supp}(\mathsf{q})$ 上では, Δ は well-defined であり, $\theta \notin \mathrm{supp}(\mathsf{q})$ 上では, $\Delta = 0$ となる. m を $\mathbb{R}$ 上の Lebesgue 測度としたとき, 以下の主張を証明する.
主張

$$2\int_{\Theta\times\mathbb{X}} \Delta(x,\,\theta)\sqrt{\mathsf{q}(\theta)}\sqrt{\mathsf{p}_\theta(x)}\big\{T(x)-\psi(\theta)\big\}\,\mathrm{d}\mathsf{m}(\theta)\,\mathrm{d}\mu(x) = \int_\Theta \dot{\psi}(\theta)\,\mathrm{d}\mathsf{Q}(\theta) \tag{9.8}$$

が成り立つ.
<u>主張の証明</u>: (9.4) もしくは (9.8) を T が指示関数の有限和である場合について証明し, あとは標準機械の議論を経由して, 河積な T に対しても成り立つことを示せばよい. さらに, $n \in \mathbb{N}$ に対して, $\psi_n = \max\{-n,\,\min\{\psi,\,n\}\}$ とおくと, ほとんど至るところで, $\psi_n \overset{n\to\infty}{\longrightarrow} \psi$ かつ $\dot{\psi}_n \overset{n\to\infty}{\longrightarrow} \dot{\psi}$ である. 有界な ψ_n と $\dot{\psi}_n$ に対して (9.4) もしくは (9.8) が成立すれば, 優収束定理により, 一般の ψ と $\dot{\psi}$ に対しても成立することがわかる. 以上の議論から, 一般性を失うことなく, $T,\,\psi,\,\dot{\psi}$ は有界と仮定してよい.

(9.8) の左辺の項ごとに評価していく:

$$
\begin{aligned}
&2\int_{\Theta_q} \overset{\bullet}{\mathsf{q}}(\theta)\frac{\mathbb{1}\{\mathsf{q}(\theta)>0\}}{2\sqrt{\mathsf{q}(\theta)}}\sqrt{\mathsf{q}(\theta)}T(x)\mathsf{p}_\theta(x)\,\mathrm{d}\mathsf{m}(\theta)\,\mathrm{d}\mu(x)\\
&\qquad=\int_{\Theta_q\times\mathbb{X}} \overset{\bullet}{\mathsf{q}}(\theta)\left\{\int_{\mathbb{X}} T(x)\mathsf{p}_\theta(x)\,\mathrm{d}\mu(x)\right\}\mathrm{d}\mathsf{m}(\theta)\\
&\qquad=-\int_{\Theta_q} \overset{\bullet}{\mathsf{q}}(\theta)\gamma_T(\theta)\,\mathrm{d}\mathsf{m}(\theta), \qquad (9.9)\\
&-2\int_{\Theta_q\times\mathbb{X}} \overset{\bullet}{\mathsf{q}}(\theta)\frac{\mathbb{1}\{\mathsf{q}(\theta)>0\}}{2\sqrt{\mathsf{q}(\theta)}}\sqrt{\mathsf{q}(\theta)}\psi(\theta)\mathsf{p}_\theta(x)\,\mathrm{d}\mathsf{m}(\theta)\,\mathrm{d}\mu(x)\\
&\qquad=-\int_{\Theta_q} \overset{\bullet}{\mathsf{q}}(\theta)\psi(\theta)\left\{\int_{\mathbb{X}} \mathsf{p}_\theta(x)\,\mathrm{d}\mu(x)\right\}\mathrm{d}\mathsf{m}(\theta)\\
&\qquad=\int_{\Theta_q} \overset{\bullet}{\mathsf{q}}(\theta)\gamma_T(\theta)\,\mathrm{d}\mathsf{m}(\theta), \qquad (9.10)\\
&2\int_{\Theta_q\times\mathbb{X}} \sqrt{\mathsf{q}(\theta)}\,\overset{\bullet}{\sqrt{\mathsf{p}_\theta}}(x)\sqrt{\mathsf{q}(\theta)}\sqrt{\mathsf{p}_\theta(x)}T(x)\,\mathrm{d}\mathsf{m}(\theta)\,\mathrm{d}\mu(x)\\
&\qquad=2\int_{\Theta_q} \mathsf{q}(\theta)\left\{\int_{\mathbb{X}} \overset{\bullet}{\sqrt{\mathsf{p}_\theta}}(x)\sqrt{\mathsf{p}_\theta(x)}T(x)\,\mathrm{d}\mu(x)\right\}\mathrm{d}\mathsf{m}(\theta)\\
&\qquad=\int_{\Theta_q} \mathsf{q}(\theta)\,\overset{\bullet}{\gamma}_T(\theta)\,\mathrm{d}\mathsf{m}(\theta), \qquad (9.11)\\
&2\int_{\Theta_q\times\mathbb{X}} \sqrt{\mathsf{q}(\theta)}\,\overset{\bullet}{\sqrt{\mathsf{p}_\theta}}(x)\sqrt{\mathsf{q}(\theta)}\sqrt{\mathsf{p}_\theta(x)}\psi(\theta)\,\mathrm{d}\mathsf{m}(\theta)\,\mathrm{d}\mu(x)\\
&\qquad=-2\int_{\Theta_q} \psi(\theta)\mathsf{q}(\theta)\left\{\overset{\bullet}{\sqrt{\mathsf{p}_\theta}}(x)\sqrt{\mathsf{p}_\theta(x)}\,\mathrm{d}\mu(x)\right\}\mathrm{d}\mathsf{m}(\theta)\\
&\qquad=0
\end{aligned}
$$

となる. 4 番目の式の最後の等号は, 補題 9.12 において, $T=1$ とおくと $\gamma_T(\theta)=1$ となるので, $\overset{\bullet}{\gamma}_T(\theta)=2\int_{\mathbb{X}}\overset{\bullet}{\sqrt{\mathsf{p}_\theta}}\sqrt{\mathsf{p}_\theta}\,\mathrm{d}\mu=0$ となることからわかる.

次に, Θ_q は可算個の互いに排反な区間 $(a_r,\,b_r)\,(r\in R)$ の和集合として表現できることに注意する. Θ_q の有限な各境界は, Θ の有限な境界, Θ の内点, $\mathrm{supp}(\mathsf{q})$ の有限な境界のいずれかである. ψ の連続性から, $\mathrm{supp}(\mathsf{q})$ の境界に θ が近づくと ψ と $\overset{\bullet}{\psi}$ の値は 0 に収束する. γ_T と ψ の有界性から, θ が a_r または b_r に近づくとき, $\gamma_T(\theta)\mathsf{q}(\theta)$ と $\psi(\theta)\mathsf{q}(\theta)$ の値は 0 に近づく. さらに, $\lim_{\theta\to\pm\infty}\mathsf{q}(\theta)=0$ である.

したがって, 閉区間 $[c,\,d]\subset\Theta_q$ に対して, (9.9) + (9.11) は

$$
\begin{aligned}
&\int_{[c,\,d]}\left\{\overset{\bullet}{\mathsf{q}}(\theta)\gamma_T(\theta)+\mathsf{q}(\theta)\,\overset{\bullet}{\gamma}_T(\theta)\right\}\mathrm{d}\mathsf{m}(\theta)=\left[\mathsf{q}(\theta)\gamma_T(\theta)\right]_c^d,\\
&\int_{[c,\,d]}\psi(\theta)\,\overset{\bullet}{\mathsf{q}}(\theta)\,\mathrm{d}\mathsf{m}(\theta)=\left[\psi(\theta)\mathsf{q}(\theta)\right]_c^d-\int_{[c,\,d]}\overset{\bullet}{\psi}(\theta)\mathsf{q}(\theta)\,\mathrm{d}\mathsf{m}(\theta)
\end{aligned}
$$

となる．よって，$c \to a_r$ と $d \to b_r$ とすれば

$$\int_{a_r}^{b_r} \left\{ \dot{\mathsf{q}}(\theta)\gamma_T(\theta) + \mathsf{q}(\theta)\,\dot{\gamma}_T(\theta) \right\} \mathrm{dm}(\theta) = 0,$$

$$\int_{a_r}^{b_r} \psi(\theta)\,\dot{\mathsf{q}}(\theta)\,\mathrm{dm}(\theta) = -\int_{a_r}^{b_r} \dot{\psi}(\theta)\mathsf{q}(\theta)\,\mathrm{dm}(\theta)$$

がわかる．最後に，$r \in R$ について和を取れば，(9.8) は示せた．

(9.4) の証明: Cauchy-Schwarz の不等式を (9.8) に適用すると

$$\begin{aligned} &4\int_{\Theta\times\mathbb{X}} \Delta(x,\,\theta)^2\,\mathrm{dm}(\theta)\,\mathrm{d}\mu(x) \int_{\Theta\times\mathbb{X}} \left\{T(x) - \psi(\theta)\right\}^2 \mathrm{dP}_\theta(x)\,\mathrm{dQ}(\theta) \\ &\quad\geq \left(\int_\Theta \dot{\psi}(\theta)\,\mathrm{dQ}(\theta)\right)^2 \end{aligned}$$

となる．さらに

$$\begin{aligned} &4\int_{\Theta\times\mathbb{X}} \Delta(x,\,\theta)^2\,\mathrm{dm}(\theta)\,\mathrm{d}\mu(x) \\ &\quad= 4\int_{\Theta\times\mathbb{X}} \left\{\dot{\mathsf{q}}(\theta)\right\}^2 \frac{\mathbb{1}\{\mathsf{q}(\theta) > 0\}}{4\mathsf{q}(\theta)} \mathsf{p}_\theta(x)\,\mathrm{dm}(\theta)\,\mathrm{d}\mu(x) \\ &\qquad+ 4\int_{\Theta\times\mathbb{X}} \left\{\dot{\sqrt{\mathsf{p}_\theta}}(x)\right\}^2 \mathsf{q}(\theta)\,\mathrm{dm}(\theta)\,\mathrm{d}\mu(x) \\ &\qquad+ 4\int_{\Theta\times\mathbb{X}} \dot{\mathsf{q}}(\theta)\mathbb{1}\{\mathsf{q}(\theta) > 0\}\,\dot{\sqrt{\mathsf{p}_\theta}}(x)\sqrt{\mathsf{p}_\theta}(x)\,\mathrm{dm}(\theta)\,\mathrm{d}\mu(x) \\ &\quad= \mathcal{F}_\mathsf{Q} + \int_\Theta \mathcal{F}_\mathcal{P}(\theta)\,\mathrm{dQ}(\theta) \\ &\qquad+ \int_\Theta \dot{\mathsf{q}}(\theta)\mathbb{1}\{\mathsf{q}(\theta) > 0\} \underbrace{\left\{\int_\mathbb{X} \dot{\sqrt{\mathsf{p}_\theta}}(x)\sqrt{\mathsf{p}_\theta}(x)\,\mathrm{d}\mu(x)\right\}}_{=0\ (\because \text{補題 } 9.12 \text{ で } S=1\ )} \mathrm{d}\mu(x) \\ &\quad= \mathcal{F}_\mathsf{Q} + \int_\Theta \mathcal{F}_\mathcal{P}(\theta)\,\mathrm{dQ}(\theta) \end{aligned}$$

から

$$\int_{\Theta\times\mathbb{X}} \left\{T(x) - \psi(\theta)\right\}^2 \mathrm{dP}_\theta(x)\,\mathrm{dQ}(\theta) \geq \frac{\left(\int_\Theta \dot{\psi}(\theta)\,\mathrm{dQ}(\theta)\right)^2}{\mathcal{F}_\mathsf{Q} + \int_\Theta \mathcal{F}_\mathcal{P}(\theta)\,\mathrm{dQ}(\theta)}$$

がわかる．よって，定理の不等式は証明された． □

9.5 統計的決定理論からみた Bayes 法

竹村 (2020, pp.326-334) と草間 (1975, pp.115-119) から借用すること．

9.6 章末注釈と参考文献

この章は [32, pp.175 − 192] を借用した.

9.7 演習問題

演習問題 9.1. $0 < \theta < 1$ と $n \in \mathbb{N}$ に対して, $X \sim \mathsf{Bino}(n, \theta)$ とする. θ の事前分布を $\Theta \sim \mathsf{B}(\alpha, \beta)$ とする. ただし, $\alpha > 0$, $\beta > 0$ である. すなわち, Θ の p.d.f. と $\Theta = \theta$ が与えられたときの X の条件付き p.m.f. はそれぞれ

$$\mathsf{p}^{\Theta}(\theta) = \frac{1}{B(\alpha, \beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\mathbb{1}_{(0,1)}(\theta),$$
$$\mathsf{p}^{X|\Theta}(x|\theta) = \begin{pmatrix} n \\ x \end{pmatrix}\theta^{x}(1-\theta)^{n-x}\mathbb{1}_{\{0,1,\dots,n\}}(x)$$

である. ただし

$$B(\alpha,, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,\mathrm{d}x$$

である. このとき, 以下の問いに答えよ.
(1) X の周辺 p.m.f. $\mathsf{p}^{X}(\cdot)$ を求めよ.
(2) $x = 0, 1, \dots n$ に対して, $X = x$ を与えたときの Θ の条件付き分布 (事後分布) は $B(\alpha + x, \beta + n - x)$ となることを示せ.
(3) $d^{\mathrm{Bayes}}(X) = \mathsf{E}[\Theta|X]$ とおいたとき

$$d^{\mathrm{Bayes}}(X) = \frac{\alpha+\beta}{\alpha+\beta+x}\cdot\frac{\alpha}{\alpha+\beta} + \frac{n}{\alpha+\beta+n}\cdot\frac{X}{n}$$

であることを示せ.
(4)

$$\mathsf{R}_{d^{\mathrm{Bayes}}}(\theta) := \mathsf{E}\left[\{d^{\mathrm{Bayes}}(X) - \theta\}^2\right]$$

を計算し, $\mathsf{R}_{d^{\mathrm{Bayes}}}(\theta)$ が θ に依存しない定数関数になるための条件は

$$\alpha = \beta = \frac{\sqrt{n}}{2}$$

であることを示せ.
(5) $\alpha = \sqrt{n}/2$, $\beta = \sqrt{n}/2$ のときの d^{Bayes} を d^{M} とする. すなわち

$$d^{\mathrm{M}} = \frac{1}{1+\sqrt{n}}\cdot\frac{1}{2} + \frac{\sqrt{n}}{1+\sqrt{n}}\cdot\frac{X}{n}$$

である. このとき

$$\mathsf{R}_{d^{\mathrm{M}}}(\theta) = \frac{1}{4(1+\sqrt{n})^2} < \sup_{\theta \in (0,1)} \mathsf{R}_{X/n}(\theta)$$

を示せ. ただし

$$\mathsf{R}_{X/n}(\theta) := \mathsf{E}\left[\left(\frac{X}{n} - \theta\right)^2\right]$$

である.

$\boxed{\text{ヒント}}$

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{G(\alpha+\beta)}, \quad \Gamma\alpha = \int_0^\infty x^{\alpha-1} e^{-x} \, \mathrm{d}x, \quad \Gamma(\alpha+1) = \alpha\Gamma(\alpha)$$

である.

演習問題 9.2.