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

2026 年 01 月 17 日

講義内容

- 前回までの記号と重要事項の復習
- 9.1 Bayes 的な考え方
- 9.2 Bayes 的推測手法
- 9.3 事前分布の選択について

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

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

- 確率変数 X の分布 P^X

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

  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

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

- 累積分布関数 F^X の性質: ① 右連続, ② 非減少, ③ $\lim_{x\to\infty} \mathrm{F}^X(x) = 1$, $\lim_{x\to-\infty} \mathrm{F}^X(x) = 0$.
  - 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
$$\mathsf{p}^X(x) = \mathsf{F}^X(x) - \mathsf{F}^X(x-) \qquad (x \in \mathbb{R})$$
を確率関数 (p.m.f.) という.
- 確率変数 X は連続型のとき
$$\mathsf{F}^X(x) = \int_{-\infty}^{x} \mathsf{p}^X(t)\,\mathrm{d}t \qquad (x \in \mathbb{R})$$
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

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

Bayes 的な考え方 (1)

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.

Bayes 的な考え方 (2)

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.

Bayes 的推測手法 (1)

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 $\{\mathbf{P}_\theta;\ \theta \in \Theta\}$ とする.

- 事前分布 (prior distribution) と呼ばれる母数 θ についての分布 $\mathbf{p}^{\Theta}(\theta)$ を想定する.
- 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) $\mathbf{p}^{X|\Theta}(x|\theta)$ を想定する.
- 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) $\mathbf{p}^{\Theta|X}(\theta|x)$ を求める. ただし $x = (x_1, x_2, \ldots, x_n)$ 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 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.
- $$\Pr(\Theta = \theta \mid X = x) = \frac{\Pr(\Theta = \theta,\ X = x)}{\Pr(X = x)} = \frac{\Pr(\Theta = \theta,\ X = x)}{\sum_\theta \Pr(X = x \mid \Theta = \theta)\Pr(\Theta = \theta)}$$

  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

  $$p^{\Theta \mid X}(\theta \mid x) = \frac{p^{X \mid \Theta}(x \mid \theta)p^{\Theta}(\theta)}{\int p^{X \mid \Theta}(x \mid \theta)p^{\Theta}(\theta)\, d\theta}$$

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

- n 個のランダム標本を $X = (X_1, X_2, \ldots, X_n)$.
- 真の母数が θ のときの X の同時 p.d.f. $\mathsf{p}^{X|\Theta}(\cdot\,|\,\theta)$.
- $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ を観測したときの尤度関数 $\mathsf{lik}_n(\theta)$ を
$$\mathsf{lik}_n(\theta) = \mathsf{p}^{X|\Theta}(x|\,\theta)$$
と書くことにする.
- よって, 事後分布は
$$\mathsf{p}^{\Theta|X}(\theta|\,x) = \frac{\mathsf{p}^{X|\Theta}(x|\,\theta)\mathsf{p}^{\Theta}(\theta)}{\int \mathsf{p}^{X|\Theta}(x|\,\theta)\mathsf{p}^{\Theta}(\theta)\,\mathrm{d}\theta} =: \frac{\mathsf{lik}_n(\theta)\mathsf{p}^{\Theta}(\theta)}{c_n} \propto \mathsf{lik}_n(\theta)\mathsf{p}^{\Theta}(\theta)$$
と表すことができる.
  - $$c_n := \int \mathsf{lik}_n(\theta)\mathsf{p}^{\Theta}(\theta)\,\mathrm{d}\theta$$
    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.

- 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.
$$\mathsf{p}^{\Theta|X}(\theta|x) \propto \mathsf{lik}_n(\theta)\mathsf{p}^{\Theta}(\theta).$$
- 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,
$$\int \{y-\theta\}^2 \mathsf{p}^{\Theta|X}(\theta|x)\,d\theta$$
を最小にする y の値を $\bar{\theta}$ と書くと
$$\bar{\theta} = \frac{\int \theta \mathsf{p}^{\Theta|X}(\theta|x)\,d\theta}{\int \theta^2\, \mathsf{p}^{\Theta|X}(\theta|x)\,d\theta} = \frac{\int \theta\, \mathsf{lik}_n(\theta)\mathsf{p}^{\Theta}(\theta)\,d\theta}{\int \theta^2\, \mathsf{lik}_n(\theta)\mathsf{p}^{\Theta}(\theta)\,d\theta}$$
と表現できる.

- また信頼区間であれば, $0 < \alpha < 1/2$ に対して

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

をみたす a, b を求めると開区間 $C = (a, b)$ は

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

をみたす事後信頼区間となる.

- ただし, 事後分布は

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

であった.

例 9.1

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

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

- すると, 事後分布は

$$\mathrm{p}^{\Theta|X}(\theta|x) \propto \mathrm{p}^{\Theta}(\theta)\mathrm{lik}_n(\theta) = \prod_{j=1}^{n} \theta^{x_j}(1-\theta)^{1-x_j}\mathbb{1}_{\{0,1\}}(x_j)$$
$$= \theta^{s_n}(1-\theta)^{n-s_n}\mathbb{1}_{\{0,1,\ldots,n\}}(s_n).$$

ただし, $s_n = \sum_{j=1}^{n} x_j$.

- したがって

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

例 9.1 の続き

- 一方 $\alpha > 0, \beta > 0$ とし

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

ただし $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} \,\mathrm{d}x$.

- すなわち θ の事前分布は $\mathsf{B}(\alpha, \beta)$ である. このとき

$$\mathsf{p}^{\Theta | X}(\theta | x) = \frac{\Gamma(n+\alpha+\beta)}{\Gamma(s_n+\alpha)\Gamma(n-s_n+\beta)} \theta^{(s_n+\alpha)-1}(1-\theta)^{(n-s+\beta)-1} \mathbb{1}_{\{0, 1, \ldots, n\}}(s_n). \tag{1}$$

- したがって

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

例 9.1 の続き

- このとき

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

とおくと

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

- $\widehat{\theta} = \frac{s_n}{n}$, $\widetilde{\theta} = \frac{1}{2}$ とおくと

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

Bayes 的推測手法 (8)

例 9.1 の続き

- 信頼係数 $\mathbf{0.95}$ の $\boldsymbol{\theta}$ の事後信頼区間 $\boldsymbol{C = (a,\ b)}$ は

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

をみたす $\boldsymbol{a},\ \boldsymbol{b}$ を数値計算で求めればよい. □

例 9.2

- $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}(\mu, \sigma^2)$.
- 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

$$\Theta | X = x \sim \mathsf{N}(\bar{\theta}, \tau^2), \tag{2}$$

$$\bar{\theta} = w\bar{x}_n + (1 - w)a, \quad \bar{x}_n = \frac{1}{n}\sum_{j=1}^{n} x_j,$$

$$w = \frac{\frac{1}{\mathsf{se}^2}}{\frac{1}{\mathsf{se}^2} + \frac{1}{b^2}}, \quad \frac{1}{\tau^2} = \frac{1}{\mathsf{se}^2} + \frac{1}{b^2}, \quad \mathsf{se} = \frac{\sigma}{\sqrt{n}}.$$

  .
- 個々の計算は, 問 9.3 を参照のこと.
- ただし, $X = (X_1, X_2, \ldots, X_n)$, $x = (x_1, x_2, \ldots, x_n)$ とした.

例 9.2 の続き

- 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

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

となる.

- n を固定する. $b \to \infty$ としたとき

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

となる. これは一様な事前分布に対応するものである.

例 9.2 の続き

- また
  区間 $C = (c, d)$ は
  $$\Pr(\theta \in C \mid X = x) = 0.95$$
  をみたすものとする.
- したがって
  $$\Pr(\theta < c \mid X = x) = 0.025, \quad \Pr(\theta > d \mid X = x) = 0.025$$
  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.
  $$\Pr(\theta < c \mid X = x) = \Pr\left(\frac{\theta - \bar{\theta}}{\tau} < \frac{c - \bar{\theta}}{\tau} \middle| X = x\right) = \Pr\left(Z < \frac{c - \bar{\theta}}{\tau}\right).$$
  ただし $Z \sim \mathrm{N}(0, 1)$ である.

例 9.2 の続き

- $$\Pr(Z < -1.96) = 0.025$$
  なので
  $$\frac{c - \bar{\theta}}{\tau} = -1.96$$
  とすればよい.
- よって
  $$c = \bar{\theta} - 1.96\tau$$
  とする.
- 同様に
  $$d = \bar{\theta} + 1.96\tau$$
  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 $(\bar{\theta} - 1.96\tau, \bar{\theta} + 1.96\tau)$ となる.
- さらに $\widehat{\theta} \approx \widetilde{\theta}$ かつ $\tau \approx \mathrm{se}$ なので信頼係数 0.95 の Bayes 的信頼区間は近似的に $(\widehat{\theta} - 1.96\mathrm{se}, \widehat{\theta} + 1.96\mathrm{se})$ となり, 頻度論的信頼区間と同じになる.

事前分布の選択について (1)

- 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.

事前分布の選択について (2)

- 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.

一様事前分布

- $X \sim \mathrm{N}(\theta, \sigma^2)$ とする. ただし, σ は既知とする.
- いま, $\mathcal{D}_n$ を $\mathrm{N}(\theta, \sigma^2)$ 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
$$\mathrm{p}^{\Theta}(\theta) \propto c$$
を採用する. ただし, $c > 0$ は定数である.
- このとき
$$\int \mathrm{p}^{\Theta}(\theta)\,\mathrm{d}\theta = \infty$$
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
$$\mathrm{p}^{\Theta|X}(\theta|\mathcal{D}_n) \propto \mathrm{lik}_n(\theta)\mathrm{p}^{\Theta}(\theta) \propto \mathrm{lik}_n(\theta)$$
となることがわかる.

事前分布の選択について(4)

$\bar{x}_n = (1/n)\sum_{j=1}^{n} x_j$ とおいたとき

- $$\Theta \mid \mathcal{D}_n \sim \mathsf{N}\left(\bar{x}_n, \frac{\sigma^2}{n}\right)$$
  となることがわかる.
- なぜならば
  $$\mathsf{lik}_n(\theta)\mathsf{p}^{\Theta}(\theta) = c\prod_{j=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x_j-\theta)^2}{2\sigma^2}\right\}$$
  $$= c\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\frac{\sum_{j=1}^{n}(x_j-\bar{x}_n)^2}{2\sigma^2} - \frac{(\bar{x}_n-\theta)^2}{2\sigma^2/n}\right\}$$
  $$\propto \exp\left\{-\frac{(\bar{x}_n-\theta)^2}{2\sigma^2/n}\right\}$$
  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, $X \sim \mathrm{Ber}(\theta)\ (0 < \theta < 1)$.
- 事前分布 p^{Θ} として

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

  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

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

事前分布の選択について (5)

- $\mathbf{\Theta}$ の分布は一様分布である. $\mathbf{\Theta}$ に関して無知 (一様分布なので) であれば, $\mathbf{\Psi}$ に関しても無知であるはずである. このことと $\mathbf{p}^{\mathbf{\Psi}}$ の分布は齟齬 (一様分布でないので) が生じることになる.
- 一様事前分布は母数の変換に関して不変ではないことがわかる.

$\mathbf{\Theta}$ の事前分布は一様分布（$\mathbf{\Theta}$ について事前情報なし）

$\Rightarrow$ 変換された $\mathbf{\Psi}$ を考えると

$\Rightarrow$ $\mathbf{\Psi}$ の分布は一様分布でない

$\Rightarrow$ $\mathbf{\Psi}$ についての事前情報があることになる.

事前分布の選択について (5)

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

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

で定める.

  - ここで, $\mathcal{F}(\theta)$ は Fisher 情報行列で

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

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

事前分布の選択について (6)

$\boxed{\textbf{定理 9.3}}$ Jeffreys 事前分布は母数の変換に対して不変である.

$\boxed{\textbf{定理 9.3 の証明}}$ 省略.

例 9.4

- Benoulli 分布を考える. $0 < \theta < 1$ としたとき,
  $p_\theta(x) = \theta^x(1-\theta)^{1-x}$ $(x = 0,\ 1)$.
- すると, Benoulli 分布の母数 θ の Fisher 情報量を求めると

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

- よって, Jeffreys 事前分布 p^Θ は

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

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

共役(きょうやく) 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

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

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

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

である.

事前分布の選択について (9)

- 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. $\mathsf{p}^{X|\Theta}(\cdot\,|\,\theta)$ は

$$\mathsf{p}^{X|\Theta}(x|\theta) = \exp\{\theta^\top x - \kappa(\theta)\} \tag{3}$$

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

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

の内部で定義する.

事前分布の選択について (10)

$\boxed{\text{定理 } \mathbf{9.6}}$ **(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

$$\mathbf{p}^{\mathbf{\Theta}}_{\boldsymbol{x}_0,\, \boldsymbol{n}_0}(\boldsymbol{\theta}) = \frac{\exp\left(\boldsymbol{n}_0 \boldsymbol{x}_0^\top \boldsymbol{\theta} - \boldsymbol{n}_0 \boldsymbol{\kappa}(\boldsymbol{\theta})\right)}{\int \exp\left(\boldsymbol{n}_0 \boldsymbol{x}_0^\top \boldsymbol{\theta} - \boldsymbol{n}_0 \boldsymbol{\kappa}(\boldsymbol{\theta})\right) \mathrm{d}\boldsymbol{\mu}(\boldsymbol{\theta})}$$

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

事前分布の選択について (11)

$\boxed{\text{定理 } \mathbf{9.6} \text{ の証明}}$ 資料を参照のこと.

本日の講義内容のまとめ

- 9.1 Bayes 的な考え方
- 9.2 Bayes 的推測手法
  - 事前分布
  - 事後分布
- 9.3 事前分布の選択について
  - 一様事前分布
  - Jefferys 事前分布