本日の講義内容

- 9.1 Bayes的の考え方
- 9.2 Bayes的推測手法
- 9.3 事前分布の選択について
- 9.4 乱数の発生法
- 9.5 MCMC法
- 9.8 1次元の van Tree 不等式

9.1 Bayes的推測の考え方

・頻度論的アプローチ

(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) 統計的モデルの母数は未知だが固定した値と考える.

(F3) 推測手法は大量のデータが観測されればうまく行くように設計されている.

- Bayes論的アプローチ

(B1) 確率は極限的頻度ではなく、信頼の程度を表現するものと考える。

(B2) 真の母数は固定された定数にもかかわらず、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。

9.2 Bayes的推測手法

Bayes的推測のステップ

0. 統計的母数モデル $\mathcal{P}=\{P_\theta : \theta \in \Theta\}$ を想定

1. 事前分布 (prior dis.) と呼ばれる母数 θ についての分布 $\pi(\theta)$ を想定する

2. $P^{X|\Theta}(\cdot|\theta)$: θ が与えられたときのデータ生成分布

3. 観測データ $\underline{X}=\underline{x}$ に基づき, θ の信頼度をアップデートする. 事後分布 $\pi^{\Theta|\underline{X}}(\cdot|\underline{x})$ を求める

3のステップの説明

Θ：母数に対応する確率変数

X：データに対応する確率変数

ΘとXは離散型確率変数とする

数学的にデリケートなところを気にしなくてすむので

事後分布は

$$\Pr(\Theta=\theta \mid X=x) = \frac{\Pr(\Theta=\theta,\ X=x)}{\Pr(X=x)}$$

$$= \frac{\Pr(\Theta=\theta,\ X=\theta)}{\sum_{\theta} \Pr(X=x,\ \Theta=\theta)}$$

$$= \frac{Pr(\Theta=\theta,\ X=x)}{\sum_{\theta} Pr(X=x \mid \Theta=\theta)\, Pr(\Theta=\theta)}$$ ← Bayesの公式での計算と同じ

これを形式的に連続分布のスタイルに直す?

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

と書ける。ただし π : 事前分布

$\pi^{\Theta|X}(\cdot|x)$: $X=x$ のときの Θ の条件付き分布(事後分布)

$p^{X|\Theta}(\cdot|\theta)$: $\Theta=\theta$ のときのデータの分布

$X=x$ を観測したときの Θ の確率密度を求める

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

とおくと

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

θ に関係ない

$$=: \frac{\mathrm{lik}_n(\theta)\,\pi(\theta)}{C_n}$$

$$\propto \mathrm{lik}_n(\theta)\,\pi(\theta)$$

と書ける。ただし

$$C_n = \int \mathrm{lik}_n(\theta)\pi(\theta)\, d\theta$$

したがって

$$\pi^{\Theta|X}(\theta|x) \propto \mathrm{lik}_n(\theta)\pi(\theta)$$

θの推測するために事後分布の平均等を用いる。

たとえば、yをxの関数としてとる

$$\int_{\Theta}\{y-\theta\}^2 \pi^{\Theta|X}(\theta|x)\, d\theta$$

を最小にする y の値を $\bar{\theta}$ とおくと

$$\bar{\theta} = \frac{\int_{\Theta} \theta \, \pi^{\Theta|X}(\theta|x) \, d\theta}{\int_{\Theta} \theta^2 \, \pi^{\Theta|X}(\theta|x) \, dx}$$

$$= \frac{\int_{\Theta} \theta \, \mathrm{lik}_n(\theta) \pi(\theta) \, d\theta}{\int_{\Theta} \theta^2 \, \mathrm{lik}_n(\theta) \pi(\theta) \, d\theta}$$

と表現できる。

また, $0 < \alpha < \frac{1}{2}$ に対して

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

をみたす a, b を求める. すると区間 $C = (a, b)$

は

$$\Pr(\Theta \in C) = 1 - \alpha$$

をみたす Θ の信頼区間となる.

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

とし、

$$\pi(\theta) = \begin{cases} 1 & (0<\theta<1) \\ 0 & (\text{otherwise}) \end{cases}$$

とする。すると

$$\underline{X} = (X_1, \ldots, X_n)$$

$$\underline{x} = (x_1, \ldots, x_n)$$

$$\pi^{\Theta|\underline{X}}(\theta|\underline{x}) \propto \mathrm{lik}_n(\theta)\,\pi(\theta)$$

$$= \prod_{j=1}^{n} \theta^{x_j}(1-\theta)^{1-x_j}\,\mathbb{I}_{\{0,1\}}(x_j)$$

$$= \theta^{s}(1-\theta)^{n-s}\,\mathbb{1}_{\{0,1,\dots,n\}}(s)$$

となる。ただし $s = \sum_{j=1}^{n} x_j$.

$$\int_0^1 \{y-\theta\}^2 \pi^{\Theta|X}(\theta|x)\, d\theta$$

を最小にするのは $\bar{\theta} = \frac{s}{n}$ となる。

一方、$\alpha > 0$, $\beta > 0$ とし

$$\pi(\theta|\alpha,\beta) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} & (0<\theta<1) \\ 0 & (\text{otherwise}) \end{cases}$$

とする。ただし、$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$ $(\alpha > 0)$ である。

このとき

$$\pi^{\Theta|X}(\theta|\underline{x}) \propto \mathrm{lik}_n(\theta)\pi(\theta)$$

$$= \theta^{(s+\alpha)-1}(1-\theta)^{(n-s+\beta)-1}$$

がわかる。よって

$$\Theta|\underline{X}=\underline{x} \sim B(s+\alpha,\ n-s+\beta)$$

例19.2 $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ とする

ただし、$\mu \in \mathbb{R}$ は未知で、σ^2 は既知、とする.

事前分布として

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

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

すると

$$\Theta \mid \underline{X} = \underline{x} \sim N(\bar{\theta}, \tau^2)$$

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

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

となる

注意　事前分布と事後分布が同じ母数モデルに属するとき，事前分布はこの統計的母数モデルに共役（conjugate）するという．このような事前分布を共役事前分布という

$\Theta \mid \underline{X} = \underline{x} \sim N(\bar{\theta}, \tau^2)$ に対して、

$n \to \infty$ のとき、$w \to 1$ かつ $\dfrac{\tau}{se} \to 1$ となる.

したがって、n が大きいとき

$$\Theta \mid \underline{X} = \underline{x} \approx N(\bar{x}, se^2)$$

となる.

また、nを固定する. $b \to \infty$ としたとき

$$\Theta \mid \underline{X} = \underline{x} \approx N(\bar{x}, se^2)$$

となる.

信頼区間のBayes的構成

区間 $C = (c, d)$ を

$$\Pr(\theta \in C \mid \underline{X} = \underline{x}) = \alpha \quad \left(0 < \alpha < \tfrac{1}{2}\right)$$

をみたすものとする。したがって

$$\Pr(\theta < c \mid \underline{X} = \underline{x}) = \frac{\alpha}{2}$$

$$\Pr(\theta > d \mid \underline{X} = \underline{x}) = \frac{\alpha}{2}$$

となるように c と d を選べばよい。

すなわち，$\Theta \mid \underline{X} = \underline{x} \sim N(\bar{\theta}, \tau^2)$ に注意すると

$$\Pr(\theta < c \mid \underline{X} = \underline{x}) = \Pr\left(\frac{\theta - \bar{\theta}}{\tau} < \frac{c - \bar{\theta}}{\tau} \,\middle|\, \underline{X} = \underline{x}\right)$$

$$= \Pr\left(Z < \frac{c - \bar{\theta}}{\tau}\right) \qquad Z \sim N(0, 1)$$

$Z_{\alpha/2}$ を $N(0,1)$ の上側 $\frac{\alpha}{2}$ 点とすると

$$\frac{c - \bar{\theta}}{\tau} = -Z_{\alpha/2} \iff c = \bar{\theta} - Z_{\alpha/2}\tau$$

同様に

$$d = \bar{\theta} + Z_{\alpha/2}\tau$$

よって、信頼係数 $1-\alpha$ の Bayes 的信頼区間は

$$\left(\bar{\theta} - z_{\frac{\alpha}{2}} \tau ,\ \bar{\theta} + z_{\alpha/2} \tau \right)$$

となる。ただし

$$\bar{\theta} = w \bar{x} + (1-w) a$$

$$w = \frac{\frac{1}{se^2}}{\frac{1}{se^2} + \frac{1}{b^2}}, \quad \frac{1}{\tau^2} = \frac{1}{se^2} + \frac{1}{b^2}$$

□

9.3 事前分布の選択について

ここで事前分布を p^{Θ} を表記することにする.

例9.1の再考 Beroulli試行において事前分布を

$$p^{\Theta}(\theta)=\begin{cases}1 & (0<\theta<1)\\ 0 & (\text{otherwise})\end{cases} \longleftarrow \text{一様分布}$$

と定める. $D_n=\{x_1,\dots,x_n\}$ を $\underline{X}=(X_1,\dots,X_n)$ の実現値

とし, $s_n=\sum_{j=1}^{n} x_j$ と書いたとき

$$\Theta \mid D_n \sim \mathrm{Beta}(s_n+1,\ n-s_n+1)$$

となる.

9.3.1 一様事前分布

$X \sim N(\theta, \sigma^2)$ とする. ただし, $\theta \in \mathbb{R}$ は未知とし,
σ^2 は既知とする. $\underbrace{D_n}_{=\{x_1, \dots, x_n\}}$ を $N(\theta, \sigma^2)$ の n 個の i.i.d.
標本の実現値とする

$\theta \in \mathbb{R}$ の事前分布 p^{Θ} を　　ℝ上の「一様分布」

$$p^{\Theta}(\theta) \propto c \qquad (c > 0)$$

とする. このとき

$$\int_{-\infty}^{\infty} p^{\Theta}(\theta)\,d\theta = +\infty$$

となるので、$p^{\Theta}(\cdot)$ は p.d.f. でない。

しかし、形式的な計算により

$$p^{\Theta|\underline{X}}(\theta \mid D_n) \propto \mathrm{lik}_n(\theta)\, p^{\Theta}(\theta)$$

$$\propto \mathrm{lik}_n(\theta)$$

となることがわかる。ここで

$$\mathrm{lik}_n(\theta) \propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}(x_j-\theta)^2\right\}$$

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

$$\Theta \mid D_n \sim N\left(\bar{x}_n, \frac{\sigma^2}{n}\right)$$

となる.

∵) $\sum_{j=1}^{n}(x_j-\theta)^2 = n(\bar{x}_n-\theta) + \underbrace{\sum_{j=1}^{n}(x_j-\bar{x}_n)^2}_{\text{定数項}}$

より

$$\mathrm{lik}_n(\theta) \propto \exp\left\{-\frac{n}{2\sigma^2}(\theta-\bar{x}_n)^2\right\}$$

$$= \exp\left\{-\frac{1}{2\cdot\frac{\sigma^2}{n}}(\theta-\bar{x}_n)^2\right\}$$

□

以上から、Bayes的推測は頻度論的推測と一致する。

9.3.2 Jeffreys 事前分布

$\underline{X} = (X_1, \ldots, X_n)$ の同時 p.d.f. を $p^{\underline{X}|\Theta}(\cdot|\theta)$ と書いたとき、

$$\mathcal{I}(\theta) = E\left[\left(\frac{\partial}{\partial\theta}\log p^{\underline{X}|\Theta}(\underline{X}|\theta)\right)\left(\frac{\partial}{\partial\theta}\log p^{\underline{X}|\Theta}(\underline{X}|\theta)\right)^T\right]$$

と定める。ただし、$\theta = (\theta_1, \ldots, \theta_d)$ に対して

$$\frac{\partial}{\partial\theta} = \left(\frac{\partial}{\partial\theta_1}, \ldots\ldots, \frac{\partial}{\partial\theta_d}\right)^T$$

とした。すなわち、$\mathcal{I}(\theta)$ は Fisher 情報行列である

以下では，$\mathcal{I}(\theta)$ は正定値と仮定する

θ の事前分布を

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

と定める．ただし，$|\ \ |$ は行列式である．

例9.4（例9.1の続き） Beroulli試行の場合（1回のみ）を考える．すると

$$\mathcal{I}(\theta) = E\left[\left(\frac{d}{d\theta}\log\{\theta^{x}(1-\theta)^{1-x}\}\right)^{2} \middle| \Theta = \theta\right]$$

$$= E\left[\left(\frac{X}{\theta} - \frac{1-X}{1-\theta} \right)^2 \middle| \Theta = \theta \right]$$

$$= E\left[\left(\frac{X - \theta}{\theta(1-\theta)} \right)^2 \middle| \Theta = \theta \right]$$

$$= \frac{1}{\{\theta(1-\theta)\}^2} E[(X-\theta) \mid \Theta = \theta]$$

$$= \frac{1}{\{\theta(1-\theta)\}^2} \mathrm{Var}[X \mid \Theta = \theta]$$

$$= \frac{1}{\{\theta(1-\theta\}^2} \theta(1-\theta) = \frac{1}{\theta(1-\theta)}.$$

よって

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

なので

$$p^{\Theta}(\theta) \sim \mathrm{Beta}\left(\frac{1}{2}, \frac{1}{2}\right).$$

注意　一様事前分布とJeffreys事前分布のように「超母数」を含まない事前分布を無情報事前分布とよぶ。

9.3.3 共役事前分布

定義 9.5 事前分布 p^{Θ} は共役 (conjugate) であるとは, 事前分布の分布族を $\mathcal{P}$ と書いたとき

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

が成立することをいう。ただし $p^{X|\Theta}$ を標本分布の p.d.f. で

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

データの分布族が指数型分布族のとき.

共役事前分布を求めてみよう.

◇ 有限な測度 μ に関する p.d.f. $p(\cdot|\boldsymbol{\theta})$

は

$p^{\boldsymbol{x}|\Theta}(\boldsymbol{x}|\boldsymbol{\theta})$ とかくべきところを簡単に書く

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \exp\{\boldsymbol{\theta}^T \boldsymbol{x} - \psi(\boldsymbol{\theta})\} \qquad (9.3)$$

で与えられているとする. $\Theta \subset \mathbb{R}^d$ を母数空間とし

$$h : \Theta \longrightarrow [0, \infty)$$

で

$$\log\left\{\int_{\mathbb{R}^d} p(\underline{x}\mid\underline{\theta})\, d\mu(\underline{x})\right\} = u(\theta)$$

である。すなわち

$$\bar{\Theta} = \left\{\underline{\theta}\in\mathbb{R}^d \,;\, u(\theta) = \log\left(\int_{\mathbb{R}^d} \exp\{\underline{\theta}^T\underline{x} - u(\theta)\}\, d\mu(\underline{x})\right) < \infty\right\}$$

としたとき、Θ は $\bar{\Theta}$ の内部である。

定理 9.6 事前分布の p.d.f. を

$$p^{\Theta}_{\underline{x}_0, n_0}(\underline{\theta}) = \frac{\exp\{ n_0 \underline{x}_0^T \theta - n_0 u(\underline{\theta}) \}}{\int_{\Theta} \exp(n \underline{x}_0^T \underline{\theta} - n_0 u(\underline{\theta}) \} d\mu(\underline{\theta})}$$

は、指数型分布 (9.3) に対して共役である.
ただし. $\underline{x}_0 \in \mathbb{R}^d$, $n_0 \in \mathbb{R}$.

証明

$$p(\underline{x} | \underline{\theta})\, p^{\Theta}_{\underline{x}_0, n_0}(\underline{\theta})$$

$$\propto \exp\{\underline{\theta}^T \underline{x} - u(\underline{\theta})\}\exp\{n_0 \underline{x}_0^T \underline{\theta} - n_0 u(\underline{\theta})\}$$

$$= \exp\{(\underline{x} + \underline{x}_0)^T \underline{\theta} - (1+n_0) u(\underline{\theta})\}$$

$$= \exp\left\{(1+n_0)\left(\frac{1}{1+n_0}\underline{x} + \frac{1}{1+n_0}\underline{x}_0\right)^T \underline{\theta} - (1+n_0) u(\underline{\theta})\right\}$$

$$\propto p^{\Theta}_{\frac{1}{1+n_0}\underline{x} + \frac{1}{1+n_0}\underline{x}_0,\ 1+n_0}(\underline{\theta}).$$

$\boxed{\text{Q}}$