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】真の母数は未知だか, 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 $\boldsymbol{\pi}(\theta)$ 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) $\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)$ 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) $\boldsymbol{\pi}_{\theta|\,\boldsymbol{X}}(\theta|\,\boldsymbol{x})$ 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 $\boldsymbol{\Theta}$ の実現値と考える. いまは離散型確率変数の設定なので

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

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

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

と書ける. ただし $\boldsymbol{\pi}_{\boldsymbol{\Theta}|\,\boldsymbol{X}}(\,\cdot\,|\,\boldsymbol{x})$ 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. $\mathsf{p}(\boldsymbol{x}|\,\theta)$ は真の母数が θ のときのデータの分布の p.d.f. で, $\boldsymbol{\pi}(\theta)$ は $\boldsymbol{\Theta}$ 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. $\mathsf{p}^{\boldsymbol{X}}(\,\cdot\,|\,\theta)$ 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

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

と書くことにする. よって

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

と表すことができる．ただし

$$c_n := \int \mathsf{lik}_n(\theta)\boldsymbol{\pi}(\theta)\,\mathrm{d}\theta$$

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．

$$\boldsymbol{\pi}_{\boldsymbol{\Theta}|\,\boldsymbol{X}}(\theta|\,\boldsymbol{x}) \propto \mathsf{lik}_n(\theta)\boldsymbol{\pi}(\theta).$$

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,

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

を最小にする y の値を $\overline{\theta}$ と書くと

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

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

$$\int_{-\infty}^{a}\boldsymbol{\pi}_{\boldsymbol{\Theta}|\,\boldsymbol{X}}(\theta|\,\boldsymbol{x})\,\mathrm{d}\theta = \int_{b}^{\infty}\boldsymbol{\pi}_{\boldsymbol{\Theta}|\,\boldsymbol{X}}(\theta|\,\boldsymbol{x})\,\mathrm{d}\theta$$

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

$$\boldsymbol{\pi}(\theta) = \begin{cases} 1 & (0<\theta<1) \\ 0 & (\text{その他の場合}) \end{cases}$$

である．すると

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

となる．ただし $s=\sum_{j=1}^{n}x_j$ である．

一方 $\alpha > 0,\, \beta > 0$ とし

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

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

$$\boldsymbol{\pi}_{\boldsymbol{\Theta}|\,\boldsymbol{X}}(\theta|\,\boldsymbol{x}) = \frac{\Gamma(n+2)}{\Gamma(s+1)\Gamma(n-s+1)}\theta^{(s+1)-1}(1-\theta)^{(n-s+1)-1} \tag{9.1}$$

となる．したがって

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

がわかる．このとき

$$\overline{\theta} = \int \theta\boldsymbol{\pi}_{\boldsymbol{\Theta}|\,\boldsymbol{X}}(\theta|\,\boldsymbol{x})\,\mathrm{d}\theta$$

とおくと

$$\overline{\theta} = \frac{s+1}{n+2}$$

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

$$\overline{\theta} = \lambda_n\widehat{\theta} + (1-\lambda)\widetilde{\theta}, \quad \lambda_n = \frac{n}{n+2}$$

と書ける．

信頼係数 0.95 の θ の事後信頼区間 $C = (a,\,b)$ は

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

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，

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

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 $\boldsymbol{\Theta}$ 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

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

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

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

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

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

となる．また n を固定する．$b \to \infty$ としたとき

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

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

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

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 $\mathbf{\Theta}$ 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

9.4 章末注釈と参考文献

この章は [26, pp.175 − 192] を借用した.

9.5 演習問題

演習問題 9.1.

演習問題 9.2.