

第9章 Bayes 的推測

9.1 Bayes 的推測の考え方

前の節までの統計的手法は頻度論的手法 (frequentist methods) よ呼ばれるものである。頻度論的アプローチの考え型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。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。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 $\{\mathbb{P}_\theta; \theta \in \Theta\}$ 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 $\pi(\theta)$ 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) $p(x|\theta)$ を想定する。
3. 観測されたデータ $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ 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) $\pi_{\theta|X}(\theta|\mathbf{x})$ を求める。ただし $\mathbf{x} = (x_1, x_2, \dots, x_n)$ 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。いまは離散型確率変数の設定なので

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

となる。最後の等号は Bayes の公式を用いた。これを連続型分布の場合に書き直すと

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

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

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

$$\text{lik}_n(\theta) = p(\mathbf{x}|\theta)$$

と書くことにする。よって

$$\pi_{\Theta|X}(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta)\pi(\theta)}{\int p(\mathbf{x}|\theta)\pi(\theta) d\theta} =: \frac{\text{lik}_n(\theta)\pi(\theta)}{c_n} \propto \text{lik}_n(\theta)\pi(\theta)$$

と表すことができる. ただし

$$c_n := \int \text{lik}_n(\theta) \pi(\theta) d\theta$$

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. は事前分布の尤度関数倍である.

$$\pi_{\Theta|X}(\theta|\mathbf{x}) \propto \text{lik}_n(\theta) \pi(\theta).$$

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,

$$\int \{y - \theta\}^2 \pi_{\Theta|X}(\theta|\mathbf{x}) d\theta$$

を最小にする y の値を $\bar{\theta}$ と書くと

$$\bar{\theta} = \frac{\int \theta \pi_{\Theta|X}(\theta|\mathbf{x}) d\theta}{\int \pi_{\Theta|X}(\theta|\mathbf{x}) d\theta} = \frac{\int \theta \text{lik}_n(\theta) \pi(\theta) d\theta}{\int \text{lik}_n(\theta) \pi(\theta) d\theta}$$

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

$$\int_{-\infty}^a \pi_{\Theta|X}(\theta|\mathbf{x}) d\theta = \int_b^{\infty} \pi_{\Theta|X}(\theta|\mathbf{x}) d\theta$$

をみたく a, b を求めると区間 $C = (a, b)$ は

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

をみたく事後信頼区間となる.

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

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

すると

$$\pi_{\Theta|X}(\theta|\mathbf{x}) \propto \pi(\theta) \text{lik}_n(\theta) = \prod_{j=1}^n \theta^{x_j} (1 - \theta)^{1-x_j} = \theta^s (1 - \theta)^{n-s}$$

となる. ただし $s = \sum_{j=1}^n x_j$ である.

一方 $\alpha > 0, \beta > 0$ とし

$$\pi(\theta|\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1}$$

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

$$\pi_{\Theta|X}(\theta|\mathbf{x}) = \frac{\Gamma(n+2)}{\Gamma(s+1)\Gamma(n-s+1)} \theta^{(s+1)-1} (1-\theta)^{(n-s+1)-1}$$

となり

$$\Theta|X = \mathbf{x} \sim B(\alpha, \beta)$$

となる. このとき

$$\bar{\theta} = \int \theta \pi_{\Theta|X}(\theta|\mathbf{x}) d\theta$$

とおくと

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

を得る. $\hat{\theta} = \frac{s}{n}$, $\tilde{\theta} = \frac{1}{2}$ とおくと

$$\bar{\theta} = \lambda_n \hat{\theta} + (1-\lambda_n) \tilde{\theta}, \quad \lambda_n = \frac{n}{n+2}$$

と書ける.

信頼係数 0.95 の θ の事後信頼区間 $C = (a, b)$ は

$$\int_a^b \pi_{\Theta|X}(\theta|\mathbf{x}) d\theta = 0.95$$

をみたく a, b を数値計算で求めればよい. □

例 9.2. $X_1, X_2, \dots, X_n \sim \text{i.i.d. } N(\mu, \sigma^2)$ 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

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

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

$$\begin{aligned} \Theta|X &\sim N(\bar{\theta}, \tau^2), \\ \bar{\theta} &= w\bar{X} + (1-w)a, \\ w &= \frac{\frac{1}{se^2}}{\frac{1}{se^2} + \frac{1}{b^2}}, \quad \frac{1}{\tau^2} = \frac{1}{se^2} + \frac{1}{b^2}, \quad se = \frac{\sigma}{\sqrt{n}} \end{aligned}$$

となる. 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.

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

$$\Theta|X \approx N(\hat{\theta}, se^2)$$

となる. また n を固定する. $b \rightarrow \infty$ としたとき

$$\Theta | \mathbf{X} \approx N(\hat{\theta}, \text{se}^2)$$

となる. これは一様な事前分布に対応するものである.

区間 $C = (c, d)$ は

$$\Pr(\theta \in C | \mathbf{X}) = 0.95$$

をみたすものとする. したがって

$$\Pr(\theta < c | \mathbf{X}) = 0.025, \quad \Pr(\theta > d | \mathbf{X}) = 0.025$$

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.

$$\Pr(\theta < c | \mathbf{X}) = \Pr\left(\frac{\theta - \bar{\theta}}{\tau} < \frac{c - \bar{\theta}}{\tau} \middle| \mathbf{X}\right) = \Pr\left(Z < \frac{c - \bar{\theta}}{\tau}\right).$$

ただし $Z \sim N(0, 1)$ である. さらに

$$\Pr(Z < -1.96) = 0.025$$

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 $\hat{\theta} \approx \bar{\theta}$ かつ $\tau \approx \text{se}$ なので信頼係数 0.95 の Bayes 的信頼区間は近似的に $(\hat{\theta} - 1.96\text{se}, \hat{\theta} + 1.96\text{se})$ となり, 頻度論的信頼区間と同じになる.

9.3 事前分布の選択について

9.4 章末注釈と参考文献

9.5 演習問題