Bayesian Statistics

Bayesian Inference in Conjugate Cases

Brani Vidakovic
Professor
School of Industrial and Systems Engineering

Before We Begin…

- Estimation
- Credible Sets
- Testing Hypotheses
- Bayesian Prediction
- Examples

Bayesian Estimation

Experiment: $X_1, \dots, X_n \sim f(x|\theta)$

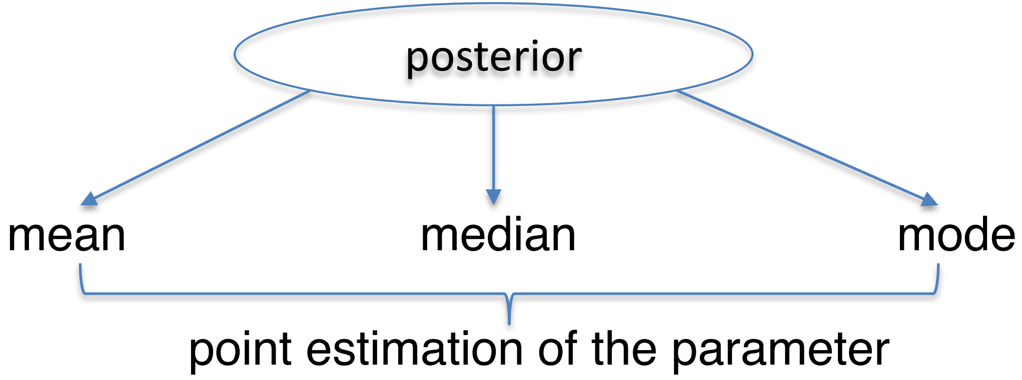
$$\begin{array}{ccc} \pi(\theta) & \longrightarrow & \pi(\theta | \underset{\sim}{X}) \\ \text{(prior)} & & \text{(posterior)} \end{array}$$

For a Bayesian, posterior is ultimate summary of experiment, prior information, and observed data.

The only coherent approach to incorporate prior information into the inference is via Bayes theorem.

<u>Inference</u>: Estimation (Point, interval), and Testing.

Estimators



Bayes' estimator is a function of $X_1, \ldots, X_n$. When $X_1, \ldots, X_n$ are observed, Bayes estimator is called an action, a.

Posterior mean is an action that minimizes

$$\mathbb{E}^{\theta|X}(\theta - a)^2,$$

with respect to a.

Posterior median is an action that minimizes

$$\mathbb{E}^{\theta|\mathrm{X}}|\theta - a|,$$

with respect to a.

Let $L_c(\theta, a) = \begin{cases} 0, & |\theta - a| \le c \\ 1, & else \end{cases}$, then **posterior mode** is an action that minimizes

$$\lim_{c \to 0} \mathbb{E}^{\theta|X} L_c(\theta, a),$$

with respect to a.

Most common Bayes estimator of the parameter is its **posterior mean**.

Frequentist rule that minimizes **Bayes risk**,

$$\mathbb{E}^{\theta}\mathbb{E}^{X|\theta}\big(\theta - \delta(X)\big)^2,$$

is the posterior mean $\delta_B(X)$. If X is observed, that is, if $\delta_B(X)$ is conditioned on X, the result is Bayes action.

Bayes risk is

$$r(\theta, \delta) = \mathbb{E}^{\theta}\mathbb{E}^{X|\theta}\big(\theta - \delta(X)\big)^2 \equiv \mathbb{E}^{X}\mathbb{E}^{\theta|X}\big(\theta - \delta(X)\big)^2.$$

Posterior mean minimizes the Bayes risk.

The Bayes Estimator:

$$\delta_B(x) = \frac{\int_\Theta \theta f(x|\theta)\pi(\theta)d\theta}{\int_\Theta f(x|\theta)\pi(\theta)d\theta} = \int_\Theta \theta \cdot \pi(\theta|x)d\theta$$

$$\Rightarrow \quad \pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{m(x)} = \frac{f(x|\theta)\pi(\theta)}{\int_\Theta f(x|\theta)\pi(\theta)d\theta}$$

$\delta_B(X) \longrightarrow$ minimizes Bayes Risk

$$r(\pi,\delta) = \mathbb{E}^{\theta}\mathbb{E}^{X|\theta}(\theta-\delta)^2$$
$$= \mathbb{E}^{X}\mathbb{E}^{\theta|X}(\theta-\delta)^2$$

$a^* \longrightarrow$ minimizes posterior expected loss

$$\mathbb{E}^{\theta|X}(\theta-a)^2$$

GTx

Credible Sets

Credible Sets

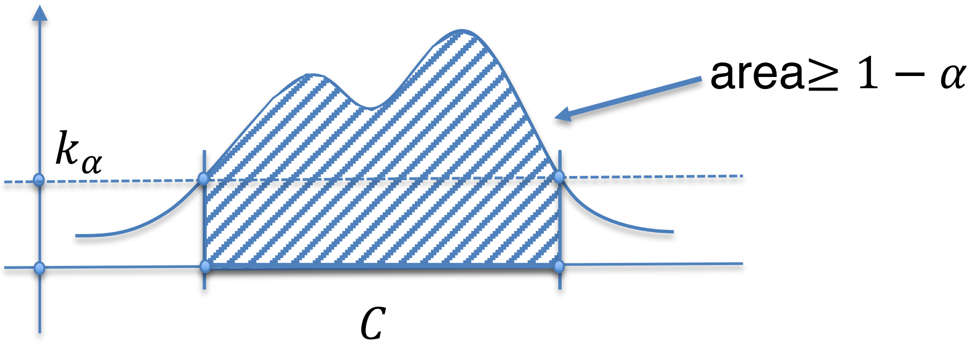
- Posterior $\pi(\theta|X)$ found.
- Assume $C \subset \Theta \equiv$ parameter space
- C is credible set with credibility $1-\alpha$, if

$$\int_C \pi(\theta|x)d\theta \geq 1-\alpha$$

- HPD-credible set (HPD $\equiv$ highest posterior density)
- Equitailed credible set

HPD Credible Set

$$C = \{\theta \in \Theta \mid \pi(\theta|x) \geq k(\alpha)\}$$
$$\mathbb{P}^{\theta|X}(\theta \in C) \geq 1-\alpha$$



Equitailed Credible Set

$$\int_{-\infty}^{L} \pi(\theta|x)d\theta \leq \frac{\alpha}{2}\,;\, \int_{U}^{+\infty} \pi(\theta|x)d\theta \leq \frac{\alpha}{2}$$
$$\mathbb{P}^{\theta|X}(\theta \in [\mathrm{L}, \mathrm{U}]) \geq 1-\alpha$$

<u>Example</u>: Jeremy's IQ

Recall, $X|\theta \sim N(\theta, \sigma^2)$

$\theta \sim N(\mu, \tau^2)$

$X = 98$

$\sigma^2 = 80$

$\tau^2 = 120$

$\theta|X \sim N(102.8, 48)$

95% CS: $102.8 \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{48}$, $z_{0.975} = 1.96$

$\theta \in [89.2207, 116.3793]$; $L = 27.1586$

95% CI: $98 \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{80}$, $z_{0.975} = 1.96$

$\theta \in [80.4692, 115.5308]$; $L = 35.0615$

Interpretations of CS and CI are different.

<u>Example</u>: $X_1, X_2, \dots, X_n \sim \mathrm{Exp}(\lambda),\ \lambda$ – rate

$$\sum_{i=1}^{n} X_i \sim \mathrm{Ga}(n, \lambda) \qquad f(x|\lambda) \propto \lambda^n \cdot e^{-\lambda \sum x_i}$$

$$\lambda \sim \mathrm{Ga}(\alpha, \beta) \qquad \pi(\lambda) \propto \lambda^{\alpha-1} e^{-\beta \cdot \lambda}$$

Posterior: $\pi(\lambda | \sum X_i) \sim \mathrm{Ga}(n + \alpha, \sum x_i + \beta)$

Let $X_1 = 2$, $X_2 = 7$, and $X_3 = 12$ be the lifetimes of a particular device. Assume that $X_i' s$ are exponential $\mathrm{Exp}(\lambda)$ with unknown parameter λ. Let prior on λ be gamma $\mathrm{Ga}(1,8)$.

<u>Find Bayes estimators for λ and credible sets (HPD and Equitailed).</u>

Here $\sum X_i = 12 + 7 + 2 = 21, \quad n = 3$

$\lambda | \sum X_i \sim \text{Ga}\,(3 + 1, 21 + 8) \equiv \text{Ga}(4{,}29)$

Posterior Mean : $\mathbb{E}(\lambda | \sum X_i) = \frac{4}{29} = \boxed{0.1379}$

Posterior Mode: Mode$= \frac{4-1}{29} = \boxed{0.1034}$

Posterior Median: (Numerical: gaminv$\left(0.5,\ 4,\ \frac{1}{29}\right) = \boxed{0.1266)}$

Recall $\mathbb{E}(X_i | \lambda) = \frac{1}{\lambda} \quad \Rightarrow$ lifetimes $= \{7.2516, 9.6712, 7.8975\}$.

Credible sets (illustration): gammagamma.m

Bayesian Testing

Bayesian Testing

Assume that Θ_0 and Θ_1 are two non-overlapping sets of parameter θ. We want to test

$$H_0 \colon \theta \in \Theta_0 \qquad \text{v.s.} \qquad H_1 \colon \theta \in \Theta_1$$

Conceptually simple:

$$p_0 = \int_{\Theta_0} \pi(\theta|x)d\theta = \mathbb{P}^{\theta|X}(H_0)$$

$$p_1 = \int_{\Theta_1} \pi(\theta|x)d\theta = \mathbb{P}^{\theta|X}(H_1)$$

Choose hypothesis with larger posterior probability.

Prior probabilities of hypotheses

$$\pi_0 = \int_{\Theta_0} \pi(\theta)d\theta, \qquad \pi_1 = \int_{\Theta_1} \pi(\theta)d\theta$$

B_{01} – Bayes Factor in favor of H_0

$$B_{01} = \frac{p_0/p_1}{\pi_0/\pi_1}, \quad \text{(posterior odds / prior odds)}$$

$$B_{10} = \frac{1}{B_{01}}$$

Precise null $H_0\!: \theta = \theta_0$ requires prior with point mass at θ_0.

Precise null

$$H_0 : \theta = \theta_0 \qquad \text{v.s.} \qquad H_1 : \theta \neq \theta_0$$

$$\pi(\theta) = \pi_0 \cdot \delta_{\theta_0} + \overbrace{(1 - \pi_0)}^{\pi_1} \cdot \xi(\theta)$$

$$m(x) = \pi_0 \cdot f(x|\theta_0) + \pi_1 \cdot m_1(x)$$

$$m_1(x) = \int_{\{\theta \neq \theta_0\}} f(x|\theta)\xi(\theta)d\theta$$

$$\pi(\theta|x) = \frac{f(x|\theta_0)\pi_0}{m(x)} = \frac{f(x|\theta_0)\pi_0}{\pi_0 f(x|\theta_0) + \pi_1 m_1(x)} = \left(1 + \frac{\pi_1}{\pi_0} \cdot \frac{m_1(x)}{f(x|\theta_0)}\right)^{-1}.$$

Bayes Factor:

$$B_{01} = \frac{f(x|\theta_0)}{m_1(x)}.$$

Bayes Factor (BF) Calibration

B_{10} – BF in favor of H_1

Value	Evidence against $\boldsymbol{H_0}$
$0 \leq \log_{10} B_{10} \leq 0.5$	Poor
$0.5 < \log_{10} B_{10} \leq 1$	Substantial
$1 < \log_{10} B_{10} \leq 1.5$	Strong
$1.5 < \log_{10} B_{10} \leq 2$	Very Strong
$\log_{10} B_{10} > 2$	Decisive

Example: Jeremy's IQ

In the context of Jeremy's IQ example. Test the hypotheses

$$H_0: \theta \leq 100 \quad \text{v.s.} \quad H_1: \theta > 100$$

$$\theta | x \sim N(102.8, 48)$$

- $p_0 = \mathbb{P}^{\theta|X}(H_0) = \int_{-\infty}^{100} \frac{1}{\sqrt{2\pi \cdot 48}} \cdot e^{-\frac{(\theta - 102.8)^2}{2 \cdot 48}} d\theta$
  $= \text{normcdf}(100, 102.8, \ \text{sqrt}(48))$
  $= \boxed{0.3431}$

- $p_1 = \mathbb{P}^{\theta|X}(H_1) = 1 - 0.3431 = \boxed{0.6569}$

- $\pi_0 = \mathbb{P}^\theta(H_0) = \int_{-\infty}^{100} \frac{1}{\sqrt{2\pi \cdot 120}} e^{-\frac{(\theta-110)^2}{2\cdot 120}} d\theta$

  $= \text{normcdf}(100, 110, \text{sqrt}(120))$

  $= \boxed{0.1807}$

  $\pi_1 = 1 - 0.1807 = \boxed{0.8193}$

- $B_{10} = \frac{p_1/p_0}{\pi_1/\pi_0} = \frac{0.6569/0.3431}{0.8193/0.1807} = \frac{1.9146}{4.5340} = \boxed{0.4223}$

$$\boxed{\frac{p_1}{p_0} = B_{10} \times \frac{\pi_1}{\pi_0}}$$

- $\log_{10} B_{01} = -\log_{10} B_{10} = \boxed{0.3744}$ (poor evidence in favor of H_0)

GTx

Example: 10 flips of a coin revised

$$X|p \sim \mathrm{Bin}(n,p);\; p \sim \mathrm{Be}(500{,}500), X = 0$$

Posterior: $p|X \sim \mathrm{Be}(500{,}510)$

- We already found posterior mean in one of previous examples.

$$\mathbb{E}(p|X) = \frac{500}{1010} = 0.4950495\ldots$$

- The **mode** for $\mathrm{Be}(\alpha,\beta)$ is $\frac{\alpha-1}{\alpha+\beta-2}$; here the posterior mode is $\frac{499}{1008} = 0.4950397\ldots$
- The **median** (not explicit, uses special functions)
  $\mathrm{betainv}(0.5, 500, 510) = 0.4950462\ldots$
  (approximation $\frac{\alpha-1/3}{\alpha+\beta-2/3} = \frac{499.666\dot{6}}{1009.333\dot{3}} = 0.4950462\ldots$)

Test $H_0 : p \leq 0.5 \quad \text{v.s.} \quad H_1 : p > 0.5$

$$p_0 = \int_0^{0.5} \frac{1}{B(500,510)} p^{500-1}(1-p)^{510-1} dp$$

$$= \text{betacdf}(0.5, 500, 510)$$

$$= \boxed{0.6235}$$

$$p_1 = 1 - p_0 = \boxed{0.3765}$$

$$\pi_0 = \int_0^{0.5} \frac{1}{B(500,500)} p^{500-1}(1-p)^{500-1} dp$$

$$= \text{betacdf}(0.5, 500, 500)$$

$$= \boxed{0.5}$$

$$\pi_1 = 1 - \pi_0 = \boxed{0.5}$$

$$B_{01} = \frac{p_0/p_1}{\pi_0/\pi_1} = \frac{0.6235}{0.3765} = \boxed{1.656}$$

$\log_{10} B_{01} = \boxed{0.2191}$ **(Poor evidence against H_1)**

Test $H_0: p = 0.5$ v.s. $H_1: p \neq 0.5$

$\pi(p) = 0.8 \cdot \delta_{0.5} + 0.2 \cdot \mathrm{Be}(500, 500)$

$\pi_0 = 0.8,\ \pi_1 = 0.2$

$$m_1(x)\Big|_{X=0} = m_1(0)$$

$$= \int_0^1 \binom{10}{0} p^0 (1-p)^{10} \frac{1}{B(500,500)} p^{500-1}(1-p)^{500-1} dp$$

$$= \frac{B(500,510)}{B(500,500)} = \boxed{0.001021}$$

$$f(x|p)\Big|_{\substack{X=0\\ p=0.5}} = f(0|0.5)$$

$$= \binom{10}{0} 0.5^0 \cdot 0.5^{10} = \frac{1}{1024} = \boxed{0.0009765}$$

$$B_{10} = \frac{f(0|0.5)}{m_1(0)} = \frac{0.0009765}{0.001021} = \boxed{0.9564}$$

$$\log_{10} B_{01} = -\log_{10} B_{10} = -\log_{10} 0.9564 = \boxed{0.0194}$$

(Very poor evidence against H_1)

GTx

Summary