STAT3602 Statistical Inference

Example Class 4

LIU Chen

Department of Statistics and Actuarial Science, HKU

07/10/2020

Outline

Bayesian Approach: Prior to posterior

Bayesian approach

- X: observable random variate with probability function $f(x \mid \theta)$
- θ: unobservable random variate with a specified prior probability function $\pi(\theta)$:

$$\int_{\Theta} \pi(\theta) d\theta = 1 \quad (\text{ continuous } \theta), \qquad \text{or}$$

$$\sum_{\theta \in \Theta} \pi(\theta) = 1 \quad (\text{ discrete } \theta)$$

Bayesian Approach: Prior to posterior

Posterior Probability Function

The posterior probability function of θ given the observed data $\boldsymbol{x}$ is defined to be the conditional probability function of θ given $\boldsymbol{X} = \boldsymbol{x}$, that is

$$\begin{aligned} \pi(\theta \mid x) &= \frac{f(\boldsymbol{x} \mid \theta)\pi(\theta)}{\int_{\Theta} f\left(x \mid \theta'\right) \pi\left(\theta'\right) d\theta'} \\ &\propto f(\boldsymbol{x} \mid \theta)\pi(\theta) \end{aligned}$$

Let the prior $\pi(\theta)$ be given for $\theta \in \Theta$. Consider a decision problem with loss function $L(\theta, a)$ for $\theta \in \Theta$ and action $a \in \mathcal{A}$ (action space). Definition. The expected posterior loss given data $\boldsymbol{x}$, incurred by taking action a, is

$$\mathbb{E}[L(\theta, a) \mid \boldsymbol{x}] = \int_{\Theta} L(\theta, a)\pi(\theta \mid \boldsymbol{x}) d\theta$$

Bayesian Approach: **Bayesian decision**

Definition.

A Bayesian decision is to take an action $a \in \mathcal{A}$ which minimises the expected posterior loss $\mathbb{E}[L(\theta, a) \mid \boldsymbol{x}]$

Writing $f(\boldsymbol{x}) = \int_\Theta \pi(\theta') f(x \mid \theta') d\theta'$, we have

$$\begin{aligned}
\mathbb{E}[L(\theta, a) \mid \boldsymbol{x}] &= \int_\Theta L(\theta, a) \frac{f(\boldsymbol{x} \mid \theta)\pi(\theta)}{f(\boldsymbol{x})} d\theta \\
&= \frac{1}{f(\boldsymbol{x})} \int_\Theta L(\theta, a) f(\boldsymbol{x} \mid \theta)\pi(\theta) d\theta
\end{aligned}$$

Thus, minimising $\mathbb{E}[L(\theta, a) \mid \boldsymbol{x}]$ w.r.t. $a \in \mathcal{A}$ is equivalent to minimising $\int_\Theta L(\theta, a) f(\boldsymbol{x} \mid \theta)\pi(\theta) d\theta$ w.r.t. $a \in \mathcal{A}$

Outline

Point estimation of $\boldsymbol{\theta} \in \mathbb{R}^k$

Consider two examples of loss function L:

1. $L(\boldsymbol{\theta}, \boldsymbol{a}) = \|\boldsymbol{\theta} - \boldsymbol{a}\|_2^2 \left(\boldsymbol{a} \in \mathbb{R}^k\right)$ is minimized by
$$\boldsymbol{a} = \mathbb{E}[\boldsymbol{\theta} \mid \boldsymbol{x}] = \begin{bmatrix} \mathbb{E}\left[\theta_1 \mid \boldsymbol{x}\right] \\ \vdots \\ \mathbb{E}\left[\theta_k \mid \boldsymbol{x}\right] \end{bmatrix} = \text{posterior mean of } \boldsymbol{\theta}$$
2. $L(\boldsymbol{\theta}, \boldsymbol{a}) = \|\boldsymbol{\theta} - \boldsymbol{a}\|_1 \left(\boldsymbol{a} \in \mathbb{R}^k\right)$ is minimized by $\boldsymbol{a}$ satisfying $\mathbb{P}\left(\theta_i \leq a_i \mid \boldsymbol{x}\right) = 1/2, \quad i = 1, \ldots, k$. Thus, the Bayesian decision is to set a_i to be the posterior median of θ_i

Hypothesis testing about θ

Consider testing:

$$H_0 : \theta \in \Theta_0 \quad \text{vs} \quad H_1 : \theta \in \Theta_1 \equiv \Theta \backslash \Theta_0$$

Action space: $\mathcal{A} = \{a_0 \, (\leftrightarrow \text{ accept } H_0), a_1 \, (\leftrightarrow \text{ reject } H_0)\}$
Define, for some $L_0, L_1 > 0$, the **loss function**

$$L(\theta, a_0) = L_1 \mathbf{1}\{\theta \in \Theta_1\}, \quad L(\theta, a_1) = L_0 \mathbf{1}\{\theta \in \Theta_0\}$$

Given x, the **expected posterior loss**

$$\mathbb{E}[L(\theta, a_j) \mid \boldsymbol{x}] = L_{1-j} \mathbb{E}[\mathbf{1}\{\theta \in \Theta_{1-j}\} \mid \boldsymbol{x}] = L_{1-j} \mathbb{P}(\theta \in \Theta_{1-j} \mid \boldsymbol{x})$$

The Bayesian decision is to choose a_j such that $L_{1-j} \mathbb{P}(\theta \in \Theta_{1-j} \mid \boldsymbol{x})$ is minimised, or equivalently, that

$$\frac{L_{1-j}}{L_j} < \frac{\mathbb{P}(\theta \in \Theta_j \mid \boldsymbol{x})}{1 - \mathbb{P}(\theta \in \Theta_j \mid \boldsymbol{x})}$$

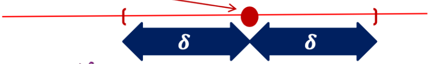
i.e.

$$\text{reject } H_0 \text{ if } \mathbb{P}(H_1 \mid \boldsymbol{x}) = \mathbb{P}(\theta \in \Theta_1 \mid \boldsymbol{x}) = \int_{\Theta_1} \pi(\theta \mid \boldsymbol{x}) d\theta > \frac{L_0}{L_0 + L_1}$$

Interval estimation of θ I

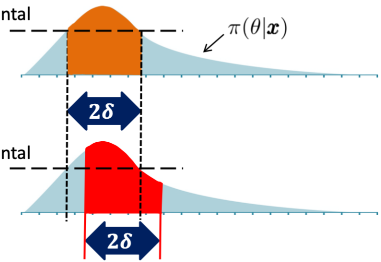
(a) Fix length $= 2\delta$ $\rightarrow$ maximise *posterior coverage probability*

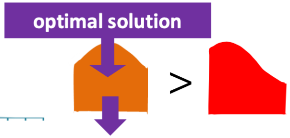
Any candidate can be represented by a



Maximise: $$\underbrace{\int_{a-\delta}^{a+\delta} \pi(\theta|\boldsymbol{x})\, d\theta}_{\text{posterior coverage probability of } [a-\delta,\, a+\delta]} \propto \int_{a-\delta}^{a+\delta} \pi(\theta) f(\boldsymbol{x}|\theta)\, d\theta \quad \text{w.r.t. } a$$

Special case: unimodal $\pi(\theta|x)$





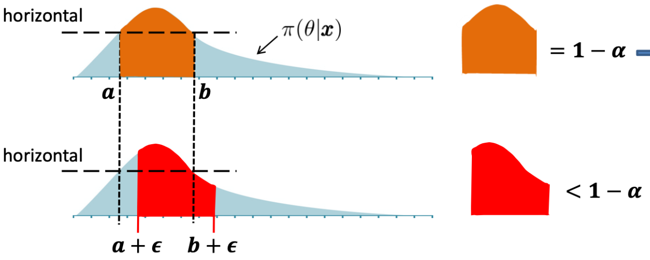
$[a-\delta, a+\delta]$ **satisfying:**

$$\boxed{\pi(a-\delta|\boldsymbol{x}) = \pi(a+\delta|\boldsymbol{x})}$$

(assuming $\pi(\theta|\boldsymbol{x})$ continuous at $\theta = a-\delta,\ a+\delta$)

(b) Desire posterior coverage probability $\geq 1-\alpha \quad \rightarrow \quad$ minimise *length*

Special case: unimodal $\pi(\theta|x)$



Optimal solution:

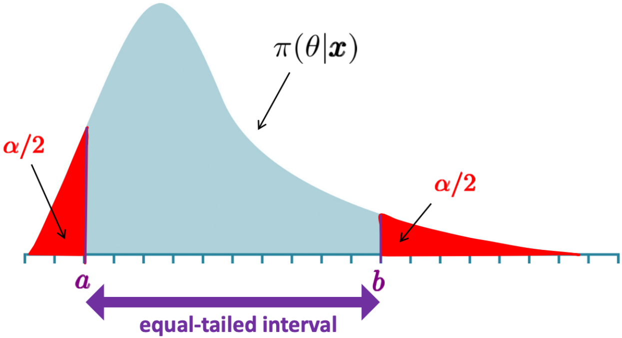
$[a, b]$ satisfying: $\boxed{\pi(a|\boldsymbol{x}) = \pi(b|\boldsymbol{x}) \quad \& \quad \int_a^b \pi(\theta|\boldsymbol{x})\, d\theta = 1-\alpha}$

(assuming $\pi(\theta|\boldsymbol{x})$ continuous at $\theta = a,\, b$)

Interval estimation of θ III

(c) Fix posterior coverage probability $= 1 - \alpha$ & require "*equal-tailed*"

i.e. $\begin{cases} \boxed{\begin{aligned} \mathbb{P}(\theta > b \mid \boldsymbol{x}) &= \int_b^{\infty} \pi(\theta|\boldsymbol{x})\, d\theta = \alpha/2 \\ \mathbb{P}(\theta < a \mid \boldsymbol{x}) &= \int_{-\infty}^{a} \pi(\theta|\boldsymbol{x})\, d\theta = \alpha/2 \end{aligned}} \end{cases}$



Predictive distribution

Observation

$\boldsymbol{\theta} \sim \pi(\theta)$ (prior)

$\boldsymbol{X} \sim f(\boldsymbol{x}|\theta)$

(posterior) $\pi(\theta|\boldsymbol{X}) \propto \pi(\theta) f(\boldsymbol{X}|\theta)$

unobserved/future $Y|(\boldsymbol{\theta}, \boldsymbol{X}) \sim g(y|\boldsymbol{\theta}, \boldsymbol{X})$

(conditional probability function of Y given $(\theta, \boldsymbol{X})$)

$$(\boldsymbol{\theta}, Y)|\boldsymbol{X} \sim g(y|\theta, \boldsymbol{X})\pi(\theta|\boldsymbol{X})$$

(interested in predicting Y)

$$g^*(y|\boldsymbol{X}) \triangleq \int_{\Theta} g(y|\theta, \boldsymbol{X})\, \pi(\theta|\boldsymbol{X})\, d\theta$$

posterior predictive distribution of Y

Outline

Exercise 1

Let $X_1, \ldots, X_n$ be i.i.d. random variables with density function $f(x \mid \theta) = \theta e^{-\theta x}, x > 0$ where $\theta \in (0, \infty)$ is an unknown parameter. Consider testing the null hypothesis $H_0 : \theta \leq 1$ against $H_1 : \theta > 1$. Denote the corresponding range of θ as Θ_0 and Θ_1. Suppose we give prior guess:

$$\pi(\theta) \sim \text{Gamma}(\alpha, \beta)$$

with density

$$\frac{1}{\Gamma(\alpha)} \beta^{\alpha} \theta^{\alpha-1} e^{-\theta\beta}$$

And define the loss function as $L(\theta, a_j) = \mathbf{1}(\theta \in \Theta_{1-j})$.

1. What is the Bayesian decision rule?

Exercise 1: Solution I

The bayesian decision rule is to choose a_0 if $P(\theta \in \Theta_0 \mid X) > 0.5$, otherwise to a_1

Easy to see that

$$\pi(\theta \mid x_1 \ldots x_n) \propto \theta^{\alpha+n-1} e^{-(n\bar{x}+\beta)\theta}$$

which is of the form as gamma distribution with $\alpha' = \alpha + n, \beta' = n\bar{x} + \beta$.
Thus the bayesian rule is to accept H_0 when $P(\theta \leq 1 \mid x) > 0.5$ for $\theta \mid x \sim \text{Gamma}(\alpha', \beta')$

Outline

Exercise 2

Given the value of λ, the number X_i of transactions made by customer i at an online store in a year has a Poisson(λ) distribution, with X_i independent of X_j for $i \neq j$. The value of λ is unknown. Our prior distribution for λ is a gamma (5,1) distribution. We observe the numbers of transactions in a year for 45 customers and $\sum_{i=1}^{45} = 182$.

1. Find the 95% equal-tailed posterior interval of λ.
2. Find an expression for the posterior predictive probability that a customer makes m transactions in a year.

A random variable X that is gamma-distributed, $X \sim \Gamma(\alpha, \beta)$. *The corresponding probability density function* $f(x \mid \alpha, \beta) = \frac{1}{\Gamma(\alpha)} \beta^{\alpha} x^{\alpha - 1} e^{-\beta x}$

Exercise 2: Solution I

1. $\lambda \sim \text{Gamma}(5,1)$. The prior pdf of λ is

$$\pi(\lambda) = \frac{1}{\Gamma(5)} e^{-\lambda}\lambda^4$$

$$f(x_1,\ldots,x_n \mid \lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}}{x_i!}\lambda^{x_i} = \frac{e^{-n\lambda}\lambda^{\sum x_i}}{\prod x_i!}$$

$$\lambda \mid x_1,\ldots,x_n \sim \text{Gamma}\left(\sum x_i + 5, n+1\right) = \text{Gamma}(187,46)$$

Thus, the 95% equal-tailed posterior interval of λ is

$$[\,\text{Gamma}_{0.025}(187,46), \text{Gamma}_{\,0.975}(187,46)] = [3.503418, 4.668181]$$

$R : qgamma(0.025, 187, 46)$ $Excel := GAMMA.INV(0.025, 187, 1/46)$

Exercise 2: Solution II

$$
\begin{aligned}
g^*(y \mid x) &= \int_0^\infty g(y \mid \lambda)\pi\left(\lambda \mid x_1, \ldots, x_n\right) d\lambda \\
&= \int_0^\infty \frac{e^{-\lambda}\lambda^y}{y!} \frac{1}{\Gamma(187)} 46^{187} \lambda^{186} e^{-46\lambda} d\lambda \\
&= \frac{46^{187}}{y!\Gamma(187)} \int_0^\infty \lambda^{186+y} e^{-47\lambda} d\lambda \\
&= \frac{46^{187}}{y!\Gamma(187)} \times \frac{\Gamma(187+y)}{47^{187+y}} \int_0^\infty \frac{47^{187+y}}{\Gamma(187+y)} \lambda^{186+y} e^{-47\lambda} d\lambda \\
&= \frac{46^{187}}{y!\Gamma(187)} \times \frac{\Gamma(187+y)}{47^{187+y}} \\
&= \frac{(186+y)!}{186!y!} \left(\frac{46}{47}\right)^{187} \left(\frac{1}{47}\right)^y
\end{aligned}
$$

Exercise 2: Solution III

The posterior predictive probability that a customer makes m transactions in a year is $g^*(y = m \mid x)$

Outline

Exercise 3

Suppose $X_1, ..., X_n$ are independent, identically distributed from the normal distribution $N(\mu, 1/\tau)$. The normal variance as well as the normal mean is unknown. Consider the prior in which τ has a Gamma distribution with parameters $\alpha > 0, \beta > 0$ and, conditionally on τ, μ has distribution $N(v, 1/(k\tau))$ for some constants $k > 0, v \in \mathbb{R}$. What is the posterior distribution, i.e. $\pi(\tau, \mu \mid x)$?

Exercise 3: Solution I

The full prior density is $\pi(\tau, \mu) = \pi(\tau)\pi(\mu \mid \tau)$, which may be written as

$$\pi(\tau, \mu) = \frac{\beta^\alpha}{\Gamma(\alpha)} \tau^{\alpha-1} e^{-\beta\tau} \cdot (2\pi)^{-1/2} (k\tau)^{1/2} \exp\left\{-\frac{k\tau}{2}(\mu - v)^2\right\}$$

or more simply

$$\pi(\tau, \mu) \propto \tau^{\alpha-1/2} \exp\left[-\tau\left\{\beta + \frac{k}{2}(\mu - v)^2\right\}\right]$$

We have $X_1, \ldots, X_n$ independent, identically distributed from $N(\mu, 1/\tau)$, so the likelihood is

$$f(x; \mu, \tau) = (2\pi)^{-n/2} \tau^{n/2} \exp\left\{-\frac{\tau}{2} \sum (x_i - \mu)^2\right\}$$

Thus

$$\pi(\tau, \mu \mid x) \propto \tau^{\alpha+n/2-1/2} \exp\left[-\tau\left\{\beta + \frac{k}{2}(\mu - v)^2 + \frac{1}{2}\sum (x_i - \mu)^2\right\}\right]$$

Exercise 3: Solution II

Complete the square to see that

$$\begin{gathered} k(\mu - v)^2 + \sum (x_i - \mu)^2 \\ = (k+n)\left(\mu - \tfrac{kv+n\bar{x}}{k+n}\right)^2 + \tfrac{nk}{n+k}(\bar{x} - v)^2 + \sum (x_i - \bar{x})^2 \end{gathered}$$

Hence the posterior satisfies

$$\pi(\tau, \mu \mid x) \propto \tau^{\alpha' - 1/2} \exp\left[-\tau\left\{\beta' + \frac{k'}{2}\left(\mu - v'\right)^2\right\}\right]$$

where

$$\begin{aligned} \alpha' &= \alpha + \tfrac{n}{2} \\ \beta' &= \beta + \tfrac{1}{2}\tfrac{nk}{n+k}(\bar{x} - v)^2 + \tfrac{1}{2}\sum (x_i - \bar{x})^2 \\ k' &= k + n \\ v' &= \tfrac{kv+n\bar{x}}{k+n} \end{aligned}$$

Exercise 3: Solution III

Thus the posterior distribution is of the same parametric form as the prior (the above form of prior is a conjugate family), but with (α, β, k, v) replaced by $(\alpha', \beta', k', v')$.

Sometimes we are particularly interested in the posterior distribution of μ alone. This may be simplified if we assume $\alpha = m/2$ for integer m. Then we may write the prior distribution, equivalently to the above, as

$$\tau = \frac{W}{2\beta}, \quad \mu = \nu + \frac{Z}{\sqrt{k\tau}}$$

where W and Z are independent random variables with the distributions χ^2_m (the chi-squared distribution on m degrees of freedom) and $N(0, 1)$ respectively. Recalling that $Z\sqrt{m/W}$ has a t_m distribution (the t distribution on m degrees of freedom), we see that under the prior distribution,

$$\sqrt{\frac{km}{2\beta}}(\mu - v) \sim t_m$$

Exercise 3: Solution IV

For the posterior distribution of μ, we replace m by $m' = m + n$, etc., to obtain

$$\sqrt{\frac{k'm'}{2\beta'}}\,(\mu - v') \sim t_{m'}$$

In general, the marginal posterior for a parameter μ of interest is obtained by integrating a joint posterior of μ and τ with respect to τ :

$$\pi(\mu \mid x) = \int \pi(\tau, \mu \mid x) d\tau$$