BAYESIAN STATISTICS

Chapter 8

Instructor: Seokho Lee

Hankuk University of Foreign Studies

8. Population Mean Estimation

8.1. Introduction

In the previous chapters we dealt with parameters associated with the dichotomous and count random variables: proportion for binomial and mean for Poisson.

We consider here the mean parameter of the continuous random variables. For example, individual income/expenditure or height/weight are the case.

The parameters of interest are mean and variance of such random variables. In this chapter we consider the mean parameter.

Normal, uniform, beta, and gamma distributions can be used for the continuous random variables. Each distribution has the associated model parameters: $\mathrm{N}(\mu, \sigma^2)$, Uniform$(a, b)$, Beta$(\alpha, \beta)$, and Gamma$(\alpha, \beta)$. The mean and variance are obtained by the model parameters. Thus, inference on the model parameters plays an important role.

Here we consider the mean parameter, μ, for normal random variables.

8.2. Bayes estimation for μ

Definition (Likelihood)

Notation: $X \sim \mathrm{N}(\mu, \sigma^2)$

$$\ell(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

Definition (Conjugate prior)

Notation: $\mu \sim \mathrm{N}(\nu, \tau^2)$

$$g(\mu) = \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{(\mu-\nu)^2}{2\tau^2}\right\}$$

with the known $-\infty < \nu < \infty$ and $\tau^2 > 0$.

8.2.1. Single data example

Suppose σ^2 is known to be σ_0^2. Then $X \sim \mathrm{N}(\mu, \sigma_0^2)$. The prior for μ is assumed to be $\mu \sim \mathrm{N}(\nu, \tau^2)$ with the known ν and τ^2.

- prior: $g(\mu) = \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{(\mu-\nu)^2}{2\tau^2}\right\}$
- likelihood: $\ell(x|\mu) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma_0^2}\right\}$
- posterior:

$$
\begin{aligned}
h(\mu|x) &\propto g(\mu)\ell(x|\mu) \\
&= \frac{1}{\sqrt{2\pi\tau^2}} \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left\{-\frac{(\mu-\nu)^2}{2\tau^2}\right\} \exp\left\{-\frac{(x-\mu)^2}{2\sigma_0^2}\right\} \\
&\propto \exp\left\{-\frac{(\mu-\nu)^2}{2\tau^2} - \frac{(x-\mu)^2}{2\sigma_0^2}\right\}
\end{aligned}
$$

Note that the exponent is

$$\frac{(\mu-\nu)^2}{\tau^2} + \frac{(x-\mu)^2}{\sigma_0^2} \;=\; \frac{\sigma_0^2+\tau^2}{\sigma_0^2\tau^2}\left(\mu - \frac{\sigma_0^2\nu+\tau^2 x}{\sigma_0^2+\tau^2}\right)^2 + \frac{(x-\nu)^2}{\sigma_0^2+\tau^2}.$$

Inserting this to the expression in the previous page, we get

$$\begin{aligned} h(\mu|x) \quad &\propto \quad \exp\left\{-\frac{\sigma_0^2+\tau^2}{2\sigma_0^2\tau^2}\left(\mu - \frac{\sigma_0^2\nu+\tau^2 x}{\sigma_0^2+\tau^2}\right)^2\right\} \times \exp\left\{-\frac{(x-\nu)^2}{2(\sigma_0^2+\tau^2)}\right\} \\ &\propto \quad \exp\left\{-\left(\mu - \frac{\sigma_0^2\nu+\tau^2 x}{\sigma_0^2+\tau^2}\right)^2 \Big/ \left(\frac{2\sigma_0^2\tau^2}{\sigma_0^2+\tau^2}\right)\right\} \\ &\sim \quad \mathrm{N}(\mu^*, \sigma^{2*}) \end{aligned}$$

with

$$\mu^* = \frac{\sigma_0^2\nu+\tau^2 x}{\sigma_0^2+\tau^2}, \qquad \sigma_0^{2*} = \frac{\sigma_0^2\tau^2}{\sigma_0^2+\tau^2}.$$

μ^* is the Bayes estimator and we denote it by $\hat{\mu}_B$. The associated 95% EPD and HPD credible interval for μ is $\mu^* \pm z_{\alpha/2}\sigma^*$.

Example (8-1)

Suppose $X \sim N(\mu, \sigma_0^2)$ where $\sigma_0^2 = 9$ and μ is unknown. The mean μ suppose follow $\mu \sim N(\nu = 100, \tau^2 = 16)$. If X is observed as $x = 112.5$, then what is the posterior of μ? Give the Bayes estimator for μ.

(solution) The posterior is

$$\mu | X = x \sim \mathrm{N}\left(\frac{\sigma_0^2\nu + \tau^2 x}{\sigma_0^2 + \tau^2}, \frac{\sigma_0^2\tau^2}{\sigma_0^2 + \tau^2}\right).$$

The Bayes estimator, or the mean of the posterior, is

$$\mu^* = \hat{\mu}_B = \frac{\sigma_0^2\nu + \tau^2 x}{\sigma_0^2 + \tau^2} = \frac{(9)(100) + (16)(112.5)}{9 + 18} = 108.$$

The variance of the posterior is

$$\sigma^{*2} = \frac{\sigma_0^2\tau^2}{\sigma_0^2 + \tau^2} = \frac{(9)(16)}{9 + 16} = (2.4)^2.$$

Thus,

$$\mu | X = 112.5 \sim \mathrm{N}(108, (2.4)^2).$$

8.2.2. *n*-sample example

First, we introduce the re-parametrization:

$$\pi_{rec} = \frac{1}{\tau^2}, \quad p_{rec} = \frac{1}{\sigma_0^2}.$$

The reciprocal of the variance is usually called "precision". With this expression, the posterior is rewritten as

$$\mu | X = x \sim \mathrm{N}\left(\frac{\sigma_0^2 \nu + \tau^2 x}{\sigma_0^2 + \tau^2}, \frac{\sigma_0^2 \tau^2}{\sigma_0^2 + \tau^2} \right) = \mathrm{N}\left(\frac{\pi_{rec} \nu + p_{rec} x}{\pi_{rec} + p_{rec}}, \frac{1}{\pi_{rec} + p_{rec}} \right)$$

Suppose we observe a single evaluation of X as x_1. Then,

- prior : $\mathrm{N}\left(\nu, \frac{1}{\pi_{rec}}\right)$
- likelihood : data x_1
- posterior : $\mathrm{N}\left(\frac{\pi_{rec} \nu + p_{rec} x_1}{\pi_{rec} + p_{rec}}, \frac{1}{\pi_{rec} + p_{rec}} \right)$

Again, we observe one more single observation of X as x_2. The posterior from the observation x_1 can be regarded as a prior in this case. Then.

- prior : $\mathrm{N}\left(\frac{\pi_{rec}\nu+p_{rec}x_1}{\pi_{rec}+p_{rec}}, \frac{1}{\pi_{rec}+p_{rec}}\right) = \mathrm{N}\left(\nu_1, \frac{1}{\pi_{1,rec}}\right)$
- likelihood : data x_2
- posterior : $\mathrm{N}\left(\frac{\pi_{1,rec}\nu_1+p_{rec}x_2}{\pi_{1,rec}+p_{rec}}, \frac{1}{\pi_{1,rec}+p_{rec}}\right) = \mathrm{N}\left(\nu_2, \frac{1}{\pi_{2,rec}}\right)$

Note that

$$\nu_2 = \frac{\pi_{1,rec}\nu_1+p_{rec}x_2}{\pi_{1,rec}+p_{rec}} = \frac{\pi_{rec}\nu+p_{rec}(x_1+x_2)}{\pi_{rec}+2p_{rec}}$$

and

$$\frac{1}{\pi_{2,rec}} = \frac{1}{\pi_{1,rec}+p_{rec}} = \frac{1}{\pi_{rec}+2p_{rec}}.$$

Now, suppose we have n samples from X. Then we can generalize this result:

- prior : $\mathrm{N}\left(\nu, \frac{1}{\pi_{rec}}\right)$
- likelihood : data $x_1, \ldots, x_n$
- posterior : $\mathrm{N}\left(\nu_n, \frac{1}{\pi_{n,rec}}\right) = \mathrm{N}\left(\frac{\pi_{rec}\nu+np_{rec}\bar{x}}{\pi_{rec}+np_{rec}}, \frac{1}{\pi_{rec}+np_{rec}}\right)$

Some notable properties:

- The posterior mean is the weighted average of the prior mean and the sample mean with weight is the reciprocals of variances; π_{rec} and np_{rec}.
- When the sample size is large, the posterior mean gets close to the sample mean and the prior mean does not affect the posterior mean.
- The posterior mean depends only on the sample mean, which is a sufficient statistic of n data. The configuration of $(x_1, \ldots, x_n)$ does not affect the posterior mean unless their mean is changed.
- The posterior variance becomes smaller as the sample size gets larger.

Suppose there is no prior information on μ. This implies that there is no preference on any value on the real line, so that the prior distribution becomes wider and its variance gets larger ($\tau^2 \to \infty$ or $\pi_{rec} \to 0$). In such case, the posterior becomes

$$\mu | \mathbf{X} = \mathbf{x} \sim \mathrm{N}\left(\bar{x}, \frac{1}{np_{rec}}\right) = \mathrm{N}\left(\bar{x}, \frac{\sigma_0^2}{n}\right).$$

Note that the posterior distribution can also be obtained as followings:

- prior $\propto \exp\left\{-\frac{(\mu-\nu)^2}{2\tau^2}\right\}$
- likelihood $\propto \exp\left\{-\frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma_0^2}\right\}$
- posterior $=$ prior $\times$ likelihood $=$? (exercise)

Example (8-2)

We observe 25 samples of $\bar{x} = 135$ from $X \sim N(\mu, 14^2)$. Suppose there is no information available on μ. Find the posterior distribution and 95% EPD credible interval for μ.

(solution) The posterior is

$$N\left(\bar{x}, \frac{\sigma_0^2}{n}\right) = N(135, 2.8^2).$$

The 95% EPD credible interval is

$$135 \pm (1.96)(2.8) = (129.51, 140.49).$$