Bayesian Inference for the Normal Distribution

1. Posterior distribution with a sample size of 1

Eg. $X|\mu \sim N(\mu, \sigma^2)$. σ^2 is known. Suppose that we have an unknown parameter μ for which the prior beliefs can be express in terms of a normal distribution, so that

$$\mu \sim N(\mu_0, \sigma_0^2)$$

where μ_0 and σ_0^2 are known.

Please derive the posterior distribution of μ given that we have on observation x

$$f(\mu) = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{\frac{-(\mu-\mu_0)^2}{2\sigma_0^2}}$$

$$f(x|\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

and hence

$$f(\mu|x) = \frac{f(\mu)f(x|\mu)}{\int_{-\infty}^{\infty} f(\mu)f(x|\mu)d\mu} = \frac{f(\mu)f(x|\mu)}{f(x)}$$

$\propto f(\mu)f(x|\mu)$

$$= \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left\{\frac{-(\mu-\mu_0)^2}{2\sigma_0^2}\right\} \times \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}$$

$$= \frac{1}{2\pi\sqrt{\sigma_0^2\sigma^2}} \exp\left\{\frac{-\mu^2 + 2\mu\mu_0 - \mu_0^2}{2\sigma_0^2} - \frac{x^2 - 2\mu x + \mu^2}{2\sigma^2}\right\}$$

$= \text{const}$

$$\times \exp\left\{\frac{-\mu^2\sigma^2 + 2\mu\mu_0\sigma^2 - \mu_0^2\sigma^2 - \sigma_0^2x^2 + 2\mu\sigma_0^2x - \mu^2\sigma_0^2}{2\sigma_0^2\sigma^2}\right\}$$

$$\propto \exp\left\{\frac{-\mu^2(\sigma^2+\sigma_0^2) + 2\mu(\mu_0\sigma^2 + \sigma_0^2x) - (\mu_0^2\sigma^2 + \sigma_0^2x^2)}{2\sigma_0^2\sigma^2}\right\}$$

$$\propto \exp\left\{\frac{-\mu^2 + 2\mu\frac{\mu_0\sigma^2+\sigma_0^2x}{\sigma^2+\sigma_0^2} - \left(\frac{\mu_0\sigma^2+\sigma_0^2x}{\sigma^2+\sigma_0^2}\right)^2}{\frac{2\sigma_0^2\sigma^2}{\sigma^2+\sigma_0^2}}\right\}$$

$$\times \exp\left\{-\frac{\mu_0^2\sigma^2 + \sigma_0^2x^2}{2\sigma_0^2\sigma^2}\right\} \propto \exp\left\{\frac{-\left(\mu - \frac{\mu_0\sigma^2 + x\sigma_0^2}{\sigma^2+\sigma_0^2}\right)^2}{2\cdot\frac{\sigma^2\sigma_0^2}{\sigma^2+\sigma_0^2}}\right\}$$

Letting

$$\sigma_1^2 = \frac{\sigma^2\sigma_0^2}{\sigma^2+\sigma_0^2} = \frac{1}{\sigma^{-2}+\sigma_0^{-2}}$$

$$\mu_1 = \frac{\mu_0\sigma^2 + x\sigma_0^2}{\sigma^2+\sigma_0^2} = \frac{\mu_0\sigma_0^{-2} + x\sigma^{-2}}{\sigma^{-2}+\sigma_0^{-2}} = \sigma_1^2(\mu_0\sigma_0^{-2} + x\sigma^{-2})$$

so that

$$\sigma_1^{-2} = \sigma^{-2} + \sigma_0^{-2}$$

$$\mu_1\sigma_1^{-2} = \mu_0\sigma_0^{-2} + x\sigma^{-2}$$

and hence

$$f(\mu|x) \propto \exp\left\{\frac{-(\mu-\mu_1)^2}{2\sigma_1^2}\right\}$$

from which it follows that as a density must integrate to unity

$$f(\mu|x) = \frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left\{\frac{-(\mu-\mu_1)^2}{2\sigma_1^2}\right\}$$

that is the posterior density is

$$\Rightarrow \mu|X \sim N(\mu_1, \sigma_1^2)$$

2. Posterior distribution with a sample of size n, using the entire likelihood.

We can generalize the situation in the previous example by supposing that a priori

$$\mu \sim N(\mu_0, \sigma_0^2)$$

but that instead of having just one observation we have, (given μ, or conditioning on μ), n independent observations $\underline{X} = (X_1, X_2, \ldots, X_n)$ such that

$$X_i|\mu \sim N(\mu, \sigma^2)$$

then

$$f(\mu|\underline{x}) \propto f(\mu)f(\underline{x}|\mu) = f(\mu)f(x_1|\mu)f(x_2|\mu)\ldots f(x_n|\mu)$$

$$= \frac{1}{\sqrt{2\pi\sigma_0^2}}\exp\left\{\frac{-(\mu-\mu_0)^2}{2\sigma_0^2}\right\} \times \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{\frac{-(x_i-\mu)^2}{2\sigma^2}\right\}$$

$$= \frac{1}{(2\pi)^{\frac{(n+1)}{2}}\sqrt{\sigma_0^2\sigma^{2n}}}\exp\left\{\frac{-\mu^2+2\mu\mu_0-\mu_0^2}{2\sigma_0^2}\right.$$

$$\left. - \sum_{i=1}^{n}\frac{x_i^2-2\mu x_i+\mu^2}{2\sigma^2}\right\}$$

$$\propto \exp\left\{\frac{-\mu^2(\sigma^2+n\sigma_0^2)+2\mu(\mu_0\sigma^2+\sigma_0^2x_1+\cdots+\sigma_0^2x_n)-(\mu_0^2\sigma^2+\sigma_0^2x_1^2+\cdots+\sigma_0^2x_n^2)}{2\sigma_0^2\sigma^2}\right\}$$

$$\propto \exp\left\{\frac{-\mu^2+2\mu\frac{\mu_0\sigma^2+\sum_{i=1}^n\sigma_0^2x_i}{\sigma^2+n\sigma_0^2}-\left(\frac{\mu_0\sigma^2+\sum_{i=1}^n\sigma_0^2x_i}{\sigma^2+n\sigma_0^2}\right)^2}{2\frac{\sigma_0^2\sigma^2}{\sigma^2+n\sigma_0^2}}\right\}$$

$$\times \exp\left\{-\frac{\mu_0^2\sigma^2+\sum_{i=1}^n\sigma_0^2x_i^2}{2\sigma_0^2\sigma^2}\right\}$$

$$\propto \exp\left\{-\frac{\left(\mu-\frac{\mu_0\sigma^2+\sum_{i=1}^n\sigma_0^2x_i}{\sigma^2+n\sigma_0^2}\right)^2}{2\frac{\sigma_0^2\sigma^2}{\sigma^2+n\sigma_0^2}}\right\}$$

Proceeding just as we did in the previous example when we had only one observation, we see that the posterior distribution is

$$\mu\big|\underline{x}\sim N(\mu_1,\sigma_1^2)$$

where

$$\sigma_1^2=\frac{\sigma^2\sigma_0^2}{\sigma^2+n\sigma_0^2}=\frac{1}{\sigma_0^{-2}+n\sigma^{-2}}$$

$$\mu_1=\frac{\mu_0\sigma^2+\sum_{i=1}^n\sigma_0^2x_i}{\sigma^2+n\sigma_0^2}=\frac{\mu_0\sigma_0^{-2}+\sum_{i=1}^n x_i\sigma^{-2}}{\sigma_0^{-2}+n\sigma^{-2}}$$

$$=\sigma_1^2\left(\mu_0\sigma_0^{-2}+\sum_{i=1}^n x_i\sigma^{-2}\right)$$

We could alternatively write these formulae as

$$\sigma_1^2=\left(\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2/n}\right)^{-1}$$

$$\mu_1=\sigma_1^2\left(\frac{\mu_0}{\sigma_0^2}+\frac{\bar{x}}{\sigma^2/n}\right)$$

which shows that, assuming a normal prior and likelihood, the result is just the same as the posterior distribution obtained from the single observation of the mean $\bar{X}$, since we know that

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

and the above formulae are the ones we had before with σ^2 replaced σ^2/n and X by $\bar{X}$.

3. Posterior distribution with a sample of size n, using the sufficient statistic $\bar{X}$.

Let $X_1, X_2, \ldots, X_n$ (given μ, or conditioning on μ) be i.i.d. $N(\mu, \sigma^2)$ where the variance is known. The sample mean

$$T(\underline{X}) = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

is the sufficient statistic for μ, such that

$$f(T(\underline{x})|\mu) = f(\bar{x}|\mu) = \frac{1}{\sqrt{2\pi\sigma^2/n}} e^{\frac{-n(\bar{x}-\mu)^2}{2\sigma^2}}$$

Then

$$f(\underline{x}|\mu) = f(T(\underline{x})|\mu)H(\underline{x})$$

so the posterior distribution is

$$f(\mu|\underline{x}) = \frac{f(\mu)f(\underline{x}|\mu)}{\int_{-\infty}^{\infty} f(\mu)f(\underline{x}|\mu)d\mu} = \frac{f(\mu)f(T(\underline{x})|\mu)H(\underline{x})}{\int_{-\infty}^{\infty} f(\mu)f(T(\underline{x})|\mu)H(\underline{x})d\mu}$$

$$\propto \frac{f(\mu)f(T(\underline{x})|\mu)}{\int_{-\infty}^{\infty} f(\mu)f(T(\underline{x})|\mu)d\mu} \propto f(\mu)f(T(\underline{x})|\mu)$$

$$= f(\mu)f(\bar{x}|\mu)$$

$$= \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left\{\frac{-(\mu-\mu_0)^2}{2\sigma_0^2}\right\}$$

$$\times \frac{1}{\sqrt{2\pi\sigma^2/n}} \exp\left\{\frac{-n(\bar{x}-\mu)^2}{2\sigma^2}\right\}$$

$$= \frac{1}{2\pi\sqrt{\sigma_0^2\sigma^2/n}} \exp\left\{\frac{-\mu^2 + 2\mu\mu_0 - \mu_0^2}{2\sigma_0^2} - \frac{\bar{x}^2 - 2\mu\bar{x} + \mu^2}{2\sigma^2/n}\right\}$$

$$= \text{const}$$

$$\times \exp\left\{\frac{-\mu^2\sigma^2/n + 2\mu\mu_0\sigma^2/n - \mu_0^2\sigma^2/n - \sigma_0^2\bar{x}^2 + 2\mu\sigma_0^2\bar{x} - \mu^2\sigma_0^2}{2\sigma_0^2\sigma^2/n}\right\}$$

$$\propto \exp\left\{\frac{-\mu^2(\sigma^2/n + \sigma_0^2) + 2\mu(\mu_0\sigma^2/n + \sigma_0^2\bar{x}) - (\mu_0^2\sigma^2/n + \sigma_0^2\bar{x}^2)}{2\sigma_0^2\sigma^2/n}\right\}$$

$$\propto \exp\left\{\frac{-\mu^2 + 2\mu\frac{\mu_0\sigma^2/n + \sigma_0^2\bar{x}}{\sigma^2/n + \sigma_0^2} - \left(\frac{\mu_0\sigma^2/n + \sigma_0^2\bar{x}}{\sigma^2/n + \sigma_0^2}\right)^2}{\frac{2\sigma_0^2\sigma^2/n}{\sigma^2/n + \sigma_0^2}}\right\}$$

$$\times \exp\left\{-\frac{\mu_0^2\sigma^2/n + \sigma_0^2\bar{x}^2}{2\sigma_0^2\sigma^2/n}\right\}$$

$$\propto \exp\left\{\frac{-\left(\mu - \frac{\mu_0\sigma^2/n + \bar{x}\sigma_0^2}{\sigma^2/n + \sigma_0^2}\right)^2}{2\cdot\frac{\sigma_0^2\sigma^2/n}{\sigma^2/n + \sigma_0^2}}\right\}$$

The same as previous examples, we are setting the posterior distribution as

$$\mu \big| \underline{X} \sim N(\mu_1, \sigma_1^2)$$

where

$$\sigma_1^2 = \frac{\sigma_0^2\sigma^2/n}{\sigma^2/n + \sigma_0^2} = \left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2/n}\right)^{-1}$$

$$\mu_1 = \frac{\mu_0\sigma^2/n + \bar{x}\sigma_0^2}{\sigma^2/n + \sigma_0^2} = \frac{\mu_0/\sigma_0^2 + \bar{x}/(\sigma^2/n)}{1/\sigma_0^2 + 1/(\sigma^2/n)} = \sigma_1^2\left(\frac{\mu_0}{\sigma_0^2} + \frac{\bar{x}}{\sigma^2/n}\right)$$

Therefore as know from the general theorem, the posterior distribution using the sufficient statistic $\bar{X}$ yields the same result as the one using the entire likelihood in example 2.

4. Frequentist Properties of Bayesian Estimators.

Given a random sample $\{X_1, X_2, \cdots, X_n\}$ from a Normal population with mean μ and variance 4. Please

(a) Derive a sufficient statistic for μ.

(b) Derive the maximum likelihood estimator (MLE) of μ.

(c) Assuming the prior of μ is $\mathrm{N}(1,9)$. Derive the the Bayes estimator of μ.

(d) Which of the two estimators (the Bayes estimator and the MLE) are better? Why?

Solution:

(a)

$$f(X|\mu) = \prod_{i=1}^{n} f(x_i|\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi \times 4}} exp\left(-\frac{(x_i-\mu)^2}{2\times 4}\right)$$

$$= \left(\frac{1}{\sqrt{2\pi \times 4}}\right)^n exp\left(-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\times 4}\right)$$

$$= (8\pi)^{-\frac{n}{2}} exp\left(-\frac{\sum_{i=1}^{n} x_i^2 + n\mu^2 - 2n\bar{x}\mu}{8}\right)$$

$$= (8\pi)^{-\frac{n}{2}} exp\left(-\frac{\sum_{i=1}^{n} x_i^2 + n\mu^2}{8}\right) exp(\frac{n\bar{x}\mu}{4})$$

By the factorization theorem, $\bar{X}$ is a SS for μ.

(b) Likelihood function:

$$L(\boldsymbol{X}|\mu) = \prod_{i=1}^{n} f(x_i|\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi \times 4}} exp\left(-\frac{(x_i-\mu)^2}{2\times 4}\right)$$

$$= (8\pi)^{-\frac{n}{2}} exp\left(-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{8}\right)$$

$$l(\mu) = -\frac{n}{2}ln(8\pi) - \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{8}$$

$$\frac{\partial l(\mu)}{\partial \mu} = -\frac{1}{4}\sum_{i=1}^{n}(x_i - \mu) = 0 \quad \Longrightarrow \hat{\mu} = \bar{x}$$

$$\frac{\partial^2 l(\mu)}{\partial^2 \mu} = -\frac{1}{4}n < 0$$

$\Longrightarrow \hat{\mu} = \bar{X}$ is the MLE

(c)

<u>The prior distribution is:</u>

$$\mu \sim N(1,9)$$

$$f(\mu) = \frac{1}{\sqrt{2\pi \times 9}} exp(-\frac{(\mu - 1)^2}{2 \times 9})$$

<u>The posterior distribution is:</u>

$$h(\mu|\boldsymbol{X}) = L(\boldsymbol{X}|\mu)f(\mu)$$

$$h(\mu|\boldsymbol{X}) \propto exp\left[-\frac{(4+9n)\mu^2 - 2(4+9n\bar{x})\mu}{72}\right]$$

$$\propto exp\left[-\frac{(\mu - \frac{4+9n\bar{x}}{4+9n})^2}{2 \times \frac{36}{4+9n}}\right]$$

$$\mu|\boldsymbol{X} \sim N(\frac{4+9n\bar{x}}{4+9n}, \frac{36}{4+9n})$$

The Bayes estimator under the squared error loss:

$$\tilde{\mu} = E(\mu|\boldsymbol{X}) = \frac{4+9n\bar{x}}{4+9n} = \frac{9n}{4+9n}\bar{x} + \frac{4}{4+9n} \times 1$$

$$\text{as } n \to \infty, \tilde{\mu} \to \bar{X}$$

(d) For the MLE $\hat{\mu} = \bar{X}$, we have:

$$E(\hat{\mu}) = E(\bar{X}) = \mu$$

$$Var(\hat{\mu}) = Var(\bar{X}) = \frac{4}{n}$$

$$MSE(\hat{\mu}) = Var(\hat{\mu}) + [E(\hat{\mu}) - \mu]^2 = \frac{4}{n}$$

Here the MLE is indeed also the best unbiased estimator for μ.

Therefore, the MLE is better when we do not have reliable prior information on the underlying distribution.

For the Bayes estimator $\tilde{\mu} = \frac{4+9n\bar{x}}{4+9n}$, we have:

$$E(\tilde{\mu}) = \mu + \frac{4(1-\mu)}{4+9n}$$

$$Var(\tilde{\mu}) = \frac{4}{n}\left(\frac{9n}{4+9n}\right)^2$$

$$MSE(\tilde{\mu}) = Var(\tilde{\mu}) + [E(\tilde{\mu}) - \mu]^2 = \frac{4}{n}\left(\frac{9n}{4+9n}\right)^2 + \left[\frac{4(1-\mu)}{4+9n}\right]^2$$

One can see that in general, the Bayes estimator is biased and could have larger MSE than the MLE. However, when the prior information is accurate, for example, taking the extreme case of $\mu = 1$. In this case, the Bayes estimator is not only unbiased, but also has smaller MSE than the MLE. Therefore, when we do have reliable prior information, the Bayesian estimator is preferred.