Lecture 12: Bayesian Inference — I

Bayes' Theorem:

Consider two events A and B, defined on the sample space S. Then

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B|A)}{P(B)}$$

$$= \frac{P(A)P(B|A)}{P(A \cap B) + P(A' \cap B)}$$

$$= \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A')P(B|A')}$$

Here $P(A)$ = prior probability of A

$P(A|B)$ = posterior probability of A

$P(B|A)$ = likelihood

Now consider a r.v. X that has a probability distribution that depends on θ, where θ is an element of a well-defined set Ω. For example, if the symbol θ is the mean of a normal distribution, then Ω may be the real line ($\mathbb{R}$).

Let Θ be a r.v. that is distributed over the set Ω. Let $h(\theta)$ be the pdf of Θ. Then $h(\theta)$ is the <u>prior</u> pdf of Θ.

Thus, $f(x|\theta)$, is the conditional pdf of X.

We have $X|\theta \sim f(x|\theta)$, $\Theta \sim h(\theta)$.

Suppose $X_1, X_2, \ldots, X_n$ is a random sample from $f(x|\theta)$. Then the likelihood function

$$L(\underline{x}|\theta) = f(x_1|\theta) f(x_2|\theta) \cdots f(x_n|\theta)$$

defines the joint conditional pdf of $\underline{X}$, given $\Theta = \theta$.

The joint pdf of $\underline{X}$ and Θ is

$$g(\underline{x}, \theta) = L(\underline{x}|\theta)h(\theta)$$

So $g_1(\underline{x}) = \int_{-\infty}^{\infty} g(\underline{x}, \theta)\,d\theta$ if Θ is continuous

The conditional pdf Θ given $\underline{X}$, is

$$k(\theta|\underline{x}) = \frac{g(\underline{x}, \theta)}{g_1(\underline{x})} = \frac{L(\underline{x}|\theta)h(\theta)}{g_1(\underline{x})}$$

Here, $k(\theta|\underline{x})$ is called the posterior pdf of Θ.

Example 1:

Consider the model

$$X_i|\theta \sim \text{iid Poisson}(\theta)$$

$$\Theta \sim \Gamma(\alpha, \beta), \text{ where } \alpha, \beta \text{ are known}$$

Let $\underline{X}' = (X_1, X_2, \ldots, X_n)$.

$$L(\underline{x}|\theta) = \prod_{i=1}^{n} \theta^{x_i} e^{-\theta}/x_i!, \quad x_i = 0, 1, 2, \ldots, \quad i = 1, 2, \ldots, n.$$

and the prior pdf is

$$h(\theta) = \frac{\theta^{\alpha-1} e^{-\theta/\beta}}{\Gamma(\alpha)\beta^{\alpha}}, \quad \theta < \infty$$

The joint pdf

$$g(\underline{x}, \theta) = L(\underline{x}|\theta) h(\theta)$$

$$= \left[\frac{\theta^{x_1} e^{-\theta}}{x_1!} \cdots \frac{\theta^{x_n} e^{-\theta}}{x_n!}\right] \left[\frac{\theta^{\alpha-1} e^{-\theta/\beta}}{\Gamma(\alpha)\beta^{\alpha}}\right],$$

provided that $x_i = 0, 1, 2, 3, \ldots$, $i = 1, 2, \ldots, n$ and

$0 < \theta < \infty$.

The marginal distribution of the sample

$$g_1(\underline{x}) = \int_0^{\infty} \frac{\theta^{\sum x_i + \alpha - 1} e^{-(n+1/\beta)\theta}}{x_1! x_2! \ldots x_n! \Gamma(\alpha)\beta^{\alpha}} d\theta$$

$$= \frac{\Gamma\left(\sum_1^n x_i + \alpha\right)}{x_1! \cdots x_n! \, \Gamma(\alpha) \beta^\alpha (n + 1/\beta)^{\Sigma x_i + \alpha}}.$$

The posterior pdf of Θ, given $\underline{X} = \underline{x}$ is

$$k(\theta | \underline{x}) = \frac{L(\underline{x}|\theta) h(\theta)}{g_1(x)}$$

$$= \frac{\theta^{\Sigma x_i + \alpha - 1} e^{-\theta/(\beta/n\beta+1)}}{\Gamma(\Sigma x_i + \alpha) \left[\beta/(n\beta+1)\right]^{\Sigma x_i + \alpha}}$$

provided that $0 < \theta < \infty$. This conditional pdf is of the gamma type, with parameters

$$\alpha^* = \sum_1^n x_i + \alpha$$

$$\beta^* = \beta/(n\beta+1)$$

Note that it is not really necessary to determine the marginal pdf $g_1(\underline{x})$ to find the posterior pdf $k(\theta|\underline{x})$. Dividing $L(\underline{x}|\theta)h(\theta)$ by $g_1(\underline{x})$ yields $c(\underline{x})\,\theta^{\sum x_i+\alpha-1}e^{-\theta/[\beta/(n\beta+1)]}$, where $c(\underline{x})$ does not depend on θ. Therefore

$$k(\theta|\underline{x}) = c(\underline{x})\,\theta^{\sum x_i+\alpha-1}e^{-\theta/[\beta/(n\beta+1)]},$$

provided that $0<\theta<\infty$, and $x_i = 0,1,2,\ldots$, $i=1,2,\ldots,n$. However, $c(\underline{x})$ must be a constant needed to make $k(\theta|\underline{x})$ a pdf i.e. $c(\underline{x}) = \dfrac{1}{\Gamma(\sum x_i+\alpha)\,[\beta/(n\beta+1)]^{\sum x_i+\alpha}}$

So, $k(\theta|\underline{x}) \propto L(\underline{x}|\theta)\,h(\theta)$

Here, we write

$$k(\theta|\underline{x}) \propto \theta^{\sum_1^n x_i + \alpha - 1} e^{-\theta/[\beta/(n\beta+1)]},$$

$0 < \theta < \infty$. Clearly, $k(\theta|\underline{x})$ is a gamma pdf with parameters

$$\alpha^* = \sum_1^n x_i + \alpha, \quad \beta^* = \beta/(n\beta+1)$$

Now suppose that there exists a sufficient statistic $Y = u(\underline{X})$ for θ so that $L(\underline{x}|\theta) = g[u(\underline{x})|\theta]\, H(x)$, where $g(y|\theta)$ is the pdf of Y, given $\Theta = \theta$. Then $k(\theta|\underline{x}) \propto g[u(\underline{x})|\theta] h(\theta)$, since $H(\underline{x})$ does not depend on θ and hence can be dropped. We can then write $k(\theta|y) \propto g(y|\theta) h(\theta)$. In the continous case, $g_1(y) = \int_{-\infty}^{\infty} g(y|\theta) h(\theta)\, d\theta$.

Bayesian Point Estimation

Suppose we wish to find a point estimator of θ. Then we must select a decision function $\delta(\underline{x})$, so that $\delta(\underline{x})$ is a predicted value of θ when both the computed value $\underline{x}$ and conditional pdf $k(\theta|\underline{x})$ are known. The choice of this decision function depends on a loss function $\mathcal{L}(\theta, \delta(x))$ in such a way that the conditional expectation of the loss is a minimum. So a <u>Bayes estimate</u> is a decision function $\delta(\underline{x})$ that minimizes

$$E\left\{ \mathcal{L}(\Theta, \delta(\underline{x})) \mid \underline{X} = \underline{x} \right\} = \int_{-\infty}^{\infty} \mathcal{L}(\theta, \delta(\underline{x}))\, k(\theta \mid \underline{x})\, d\theta$$

i.e $\delta(\underline{x}) = \operatorname{argmin} \int_{-\infty}^{\infty} \mathcal{L}(\theta, \delta(x))\, k(\theta \mid \underline{x})\, d\theta.$

Note:

a) If $\mathcal{L}(\theta, \delta(\underline{x})) = (\theta - \delta(\underline{x}))^2$, then

$\delta(\underline{x}) = E(\Theta \mid \underline{x})$, the mean of the conditional distribution of Θ, given $\underline{X} = \underline{x}$.

This follows from the fact that

$E[(W-b)^2]$ is a minimum

when $b = E(W)$.

b) If $\mathcal{L}(\theta, \delta(\underline{x})) = |\theta - \delta(\underline{x})|$, then

$\delta(\underline{x}) = \operatorname{med}(\Theta \mid \underline{x})$, the median of the conditional distribution.

This follows from the fact that
$E(|W-b|)$ is a minimum when
$b=$ median of the distribution of W.
But the conditional expectation of the loss function, given $\underline{X}=\underline{x}$ is a r.v that is a function of $\underline{x}$. Its expected value is

$$\int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} \mathcal{L}(\theta,\delta(\underline{x}))\,k(\theta|\underline{x})\,d\theta\right\} g_1(\underline{x})\,d\underline{x}$$

$$= \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} \mathcal{L}(\theta,\delta(\underline{x}))\,L(\underline{x}|\theta)\,d\underline{x}\right\} h(\theta)\,d\theta \quad \text{— (1)}$$

Here $\int_{-\infty}^{\infty} \mathcal{L}(\theta,\delta(\underline{x}))\,L(\underline{x}|\theta)\,d\underline{x} = R(\theta,\delta)$, the <u>risk function</u>.

So (1) above defines the mean risk or expected risk

Thus, a Bayes estimate $\delta(\underline{x})$ minimizes

$$\int_{-\infty}^{\infty} \mathcal{L}(\theta, \delta(\underline{x}))\, k(\theta|\underline{x})\, d\theta \quad \text{for every } \underline{x}$$

for which $g(\underline{x}) > 0$, also minimizes the mean value of the risk.

<u>Example 2</u>:

Consider $X_i|\theta \sim \text{iid } b(1, \theta)$

$\Theta \sim \text{beta}(\alpha, \beta)$, α and β are known

i.e. $h(\theta) = \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}, \quad 0<\theta<1$

We are seeking for $\delta(\underline{x})$ that is a Bayes solution.

Now, $Y = \sum_{i=1}^{n} X_i$ is a sufficient statistic for θ, and $Y \sim b(n, \theta)$

$$g(y|\theta) = \binom{n}{y}\theta^{y}(1-\theta)^{n-y}, \quad y = 0,1,2,\ldots$$

Therefore, $k(\theta|y) \propto \theta^{y}(1-\theta)^{n-y}\theta^{\alpha-1}(1-\theta)^{\beta-1}$,
$0<\theta<1$.

i.e $$k(\theta|y) = \frac{\Gamma(n+\alpha+\beta)}{\Gamma(\alpha+y)\Gamma(n+\beta-y)}\theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}$$

$$\equiv \text{beta}(\alpha+y, \beta+n-y)$$

For $\mathcal{L}(\theta, \delta(y)) = [\theta-\delta(y)]^2$, (square error loss),

$$\delta(y) = \frac{\alpha+y}{\alpha+\beta+n}$$

$$= \left(\frac{n}{\alpha+\beta+n}\right)\frac{y}{n} + \left(\frac{\alpha+\beta}{\alpha+\beta+n}\right)\frac{\alpha}{\alpha+\beta},$$

a weighted average of the MLE of θ and the mean $\alpha/(\alpha+\beta)$ of the prior pdf of the parameter.

For large n, the Bayes estimate is close to the MLE of θ and $\delta(Y)$ is a consistent estimator of θ.

Exercise (fill in missing steps)

Consider $X_i|\theta \sim \text{iid } N(\theta, \sigma^2)$, σ^2 known

$\Theta \sim N(\theta_0, \sigma_0^2)$, where θ_0 and σ_0^2 are known.

Then $Y = \bar{X}$ is a sufficient statistic.

Equivalent formulation:

$$Y|\theta \sim N(\theta, \sigma^2/n)$$

$$\Theta \sim N(\theta_0, \sigma_0^2)$$

Here $k(\theta|y) \propto \frac{1}{\sqrt{2\pi}\,\sigma/\sqrt{n}} \cdot \frac{1}{\sqrt{2\pi}\,\sigma_0} \exp\left[\frac{-(y-\theta)^2}{2(\sigma^2/n)} - \frac{(\theta-\theta_0)^2}{2\sigma_0^2}\right]$

After eliminating all constant factors, we have

$$k(\theta|y) \propto \exp\left[-\frac{[\sigma_0^2+(\sigma^2/n)]\theta^2-2[y\sigma_0^2+\theta_0(\sigma^2/n)]\times\theta}{2(\sigma^2/n)\sigma_0^2}\right]$$

This can be simplified by completing the square to become:

$$k(\theta|y) \propto \exp\left[-\frac{\left(\theta-\frac{y\sigma_0^2+\theta_0(\sigma^2/n)}{\sigma_0^2+(\sigma^2/n)}\right)^2}{\frac{2(\sigma^2/n)\sigma_0^2}{[\sigma_0^2+(\sigma^2/n)]}}\right]$$

So, the posterior pdf of the parameter is normal with mean

$$\frac{y\sigma_0^2+\theta_0(\sigma^2/n)}{\sigma_0^2+(\sigma^2/n)} = \left(\frac{\sigma_0^2}{\sigma_0^2+(\sigma^2/n)}\right)y + \left(\frac{\sigma^2/n}{\sigma_0^2+(\sigma^2/n)}\right)\theta_0$$

and variance $(\sigma^2/n)\sigma_0^2/[\sigma_0^2+(\sigma^2/n)]$

If the squared-error loss function is used, then this posterior mean is the Bayes estimator, which is a weighted average of MLE $y = \bar{x}$ and the prior mean θ_0.

For large n, the Bayes estimator is close to the MLE and $\delta(Y)$ is consistent θ.

If the absolute-error loss function was used, the Bayes solution, $\delta(Y)$, would be the median of the posterior distribution.

Bayesian Interval Estimation

To obtain an interval estimate of θ, we find two functions $u(\underline{x})$ and $v(\underline{x})$ so that $P(u(\underline{x}) < \Theta < v(\underline{x}) \mid \underline{X} = \underline{x})$

$$= \int_{u(\underline{x})}^{v(\underline{x})} k(\theta \mid \underline{x})\, d\theta$$

is large, say 0.95. The interval $(u(\underline{x}), v(\underline{x}))$ is an interval estimate of θ in that the conditional probability of Θ being there is 0.95 (say). Such an interval is called credible or probability interval.

In the Exercise above, the posterior pdf of Θ given $Y = y$ had mean $\dfrac{y\sigma_0^2 + \theta_0(\sigma^2/n)}{\sigma_0^2 + (\sigma^2/n)}$ and variance $(\sigma^2/n)\sigma_0^2 / (\sigma_0^2 + (\sigma^2/n))$.

Thus, a credible interval of probability 0.95 for θ is
$$\frac{y\sigma_0^2 + \theta_0(\sigma^2/n)}{\sigma_0^2+(\sigma^2/n)} \pm 1.96\sqrt{\frac{(\sigma^2/n)\sigma_0^2}{\sigma_0^2+(\sigma^2/n)}}$$

In Example 1, the posterior pdf was $\Gamma\left(y+\alpha,\ \beta/(n\beta+1)\right)$, where $y=\Sigma x_i$ (the sufficient statistic for θ).

$$\delta(y) = \frac{\beta(y+\alpha)}{n\beta+1} = \left(\frac{n\beta}{n\beta+1}\right)\frac{y}{n} + \left(\frac{\alpha\beta}{n\beta+1}\right)$$

Note that $\frac{2(n\beta+1)}{\beta}\Theta \sim \chi^2\left(df = 2(\Sigma x_i+\alpha)\right)$

Therefore, the $100(1-\alpha)\%$ credible interval for θ is:

$$\left(\frac{\beta}{2(n\beta+1)}\chi^2_{1-\alpha/2}\left[2\left(\sum_1^n x_i+\alpha\right)\right],\ \frac{\beta}{2(n\beta+1)}\chi^2_{\alpha/2}\left[2\left(\sum_1^n x_i+\alpha\right)\right]\right)$$