Lecture 3

Xiaoping Shi
Department of Statistics, University of Toronto
xpshi@utstat.toronto.edu
http://www.utstat.utoronto.ca/∼xpshi/sta261-2013s.html

July 7, 2013

- Bayes's Rule
- Prior and Posterior Distributions
- Bayesian Point Estimation
- Bayesian Interval Estimation
- Bayesian Testing Procedures

To understand the Bayesian inference, let us review **Bayes's Rule**:

Let A and $B_1, \cdots, B_m$ be events where the B_i are disjoint, $\cup_{i=1}^{m} = \Omega$, and $P(B_i) > 0$ for all i. Then

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{m} P(A|B_i)P(B_i)}$$

Suppose we have a Poisson distribution $P(\theta)$, and we know that the parameter is either equal to $\theta = 2$ or $\theta = 3$. In Bayesian inference, the parameter is treated as a random variable Θ.

Suppose for this example, we assign **prior** probabilities of $P(\Theta = 2) = 1/3$ and $P(\Theta = 3) = 2/3$ to the two possible values based upon past experiences.
Now suppose a random sample of size $n = 2$ results in the observations $X_1 = 2, X_2 = 4$. Given these data, what is the **posterior** probabilities of $\Theta = 2$ and $\Theta = 3$?

By Bayes's Rule, we have

$$P(\Theta = 2|X_1 = 2, X_2 = 4) =$$
$$\frac{P(X_1 = 2, X_2 = 4|\Theta = 2)P(\Theta = 2)}{P(X_1 = 2, X_2 = 4|\Theta = 2)P(\Theta = 2) + P(X_1 = 2, X_2 = 4|\Theta = 3)P(\Theta = 3)}$$
$$= \frac{\frac{1}{3}\frac{e^{-2}2^2}{2!}\frac{e^{-2}2^4}{4!}}{\frac{1}{3}\frac{e^{-2}2^2}{2!}\frac{e^{-2}2^4}{4!} + \frac{2}{3}\frac{e^{-3}3^2}{2!}\frac{e^{-3}3^4}{4!}} = 0.245$$

Similarly, $P(\Theta = 3|X_1 = 2, X_2 = 4) = 1 - 0.245 = 0.755$.

That is, with the observations $X_1 = 1, X_2 = 4$, the posterior probability of $\Theta = 2$ was smaller than the prior probability of $\Theta = 2$. Similarly, the posterior probability of $\Theta = 3$ was greater than the corresponding prior.

That is, the observations seemed to favor $\Theta = 3$ more than $\Theta = 2$.

We shall now describe the Bayesian approach to the problem of estimation. This approach is one application of a principle of statistical inference that may be called **Bayesian statistics**.

Consider a r.v. X that has a distribution of probability that depends upon the symbol θ, where θ is an element of a well-defined set Ω.

We have previously looked upon θ as being a unknown parameter in frequentist's viewpoint (250-year debate between Bayesians and frequentists).

Let us now introduce a r.v. Θ that has a distribution of probability over the set Ω; and we now look upon θ as a possible value of the r.v. Θ.

We shall denote the pdf of Θ by $h(\theta)$. Suppose that the sample $X_1, \cdots, X_n$ is a random sample (denoted by $\mathbf{X}$) from the conditional distribution of X given $\Theta = \theta$ with pdf $f(x|\theta)$.

$$\begin{aligned} X|\theta &\sim f(x|\theta) \\ \Theta &\sim h(\theta) \end{aligned}$$

Thus we can write the joint conditional pdf of $X_1, \cdots, X_n(\mathbf{X})$, given $\Theta = \theta$, as

$$L(\theta) = \prod_{i=1}^{n} f(X_i|\theta)$$

where $L(\theta)$ is the likelihood function of θ.

Thus the joint pdf of the sample and Θ is

$$g(\mathbf{X}, \Theta) = L(\theta)h(\theta)$$

and the joint marginal distribution of the sample is

$$g_1(\mathbf{X}) = \int g(\mathbf{X}, \Theta) d\theta$$

The conditional pdf of Θ, given the sample $\mathbf{X}$, is

$$k(\theta|\mathbf{X}) = \frac{g(\mathbf{X}, \Theta)}{g_1(\mathbf{X})}$$

called the **posterior pdf**.

$$k(\theta|\mathbf{X}) \propto g(\mathbf{X}, \Theta)$$

Example 1: Consider the model

$$\begin{aligned} X_i|\theta &\sim \text{iid Poisson}(\theta) \\ \Theta &\sim \Gamma(\alpha, \lambda) \end{aligned}$$

Hence, the random sample is drawn from a Poisson distribution with mean θ and the prior distribution is $\Gamma(\alpha, \lambda)$ distribution.

$$L(\theta) = \prod_{i=1}^{n} \frac{\theta^{X_i} e^{-\theta}}{X_i!}$$

and the prior pdf is

$$h(\theta) = \frac{\lambda^\alpha \theta^{\alpha-1} e^{-\lambda\theta}}{\Gamma(\alpha)}, \quad 0 < \theta < \infty$$

Hence, the joint mixed continuous discrete pdf is given by

$$g(\mathbf{X}, \Theta) = \prod_{i=1}^{n} \frac{\theta^{X_i} e^{-\theta}}{X_i!} \frac{\lambda^{\alpha} \theta^{\alpha-1} e^{-\lambda\theta}}{\Gamma(\alpha)}$$

Posterior pdf.

$$\begin{aligned} k(\theta|\mathbf{X}) \qquad & \propto g(\mathbf{X}, \Theta) \\ & \propto \theta^{(\sum_{i=1}^{n} X_i + \alpha) - 1} e^{-(n+\lambda)\theta} \end{aligned}$$

Clearly, $k(\theta|\mathbf{X})$ must be gamma pdf with parameters $\alpha^* = \sum_{i=1}^{n} X_i + \alpha$ and $\lambda^* = (n + \lambda)$, i.e., $G(\sum_{i=1}^{n} X_i + \alpha, (n + \lambda))$.

Suppose we want a joint estimator of θ. From the Bayesian viewpoint, this really amounts to selecting a decision function δ, so that $\delta(\mathbf{X})$ is a predicted value of θ .

It seems desirable that the choice of the decision function δ should depend upon a loss function $\mathcal{L}(\theta, \delta(\mathbf{X}))$. One way in which this dependence upon the loss function can be reflected is to select the decision function δ in such a way that the conditional expectation of the loss is a minimum.

A **Bayes' estimate** is a decision function δ that minimizes

$$E\{\mathcal{L}(\theta, \delta(\mathbf{X}))\} = \int \mathcal{L}(\theta, \delta(\mathbf{X}))k(\theta|\mathbf{X})d\theta$$

$$\delta(\mathbf{X}) = \arg\min E\{\mathcal{L}(\theta, \delta(\mathbf{X}))\}$$

The $\delta(\mathbf{X})$ is called a **Bayes' estimate** of θ.

If the loss function is give by $\mathcal{L}(\theta, \delta(\mathbf{X})) = [\theta - \delta(\mathbf{X})]^2$ (called **square loss**), then the Bayes' estimate is

$$\delta(\mathbf{X}) = E(\Theta|\mathbf{X})$$

the mean of **Posterior pdf**. Why?

This follows from the fact that $E[(W - b)^2]$ is a minimum when $b = E(W)$.

If the loss function is give by $\mathcal{L}(\theta, \delta(\mathbf{X})) = |\theta - \delta(\mathbf{X})|$, then a median of the **Posterior pdf** is the Bayes' solution. See Example 10 in Lecture 2.

Example 1 continued: The Posterior pdf is $G(\sum_{i=1}^{n} X_i + \alpha, (n+\lambda))$. What's Bayes' estimate of θ given square loss.

Posterior mean: $\frac{\sum_{i=1}^{n} X_i + \alpha}{n+\lambda}$.

Example 2: Consider the model

$$\begin{aligned} X_i|\theta &\sim \text{ iid Binomial, } B(1,\theta) \\ \Theta &\sim \text{ Beta distribution, } Beta(\alpha,\beta), \alpha \text{ and } \beta \text{ are known} \end{aligned}$$

that is, the prior pdf is

$$h(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}, 0<\theta<1$$

$$L(\theta) = \frac{n!}{(\sum_{i=1}^n X_i)!(n-\sum_{i=1}^n X_i)!}\theta^{\sum_{i=1}^n X_i}(1-\theta)^{n-\sum_{i=1}^n X_i}$$

see subsection of MLE of Multinomial cell probabilities for pdf of Binomial distribution in Lecture 2.

Posterior pdf.

$$\begin{aligned} k(\theta|\mathbf{X}) &\propto g(\mathbf{X},\Theta) \\ &\propto \theta^{(\sum_{i=1}^n X_i+\alpha)-1}(1-\theta)^{n-\sum_{i=1}^n X_i+\beta-1} \end{aligned}$$

The Posterior pdf is $Beta(\sum_{i=1}^{n} X_i + \alpha, n - \sum_{i=1}^{n} X_i + \beta)$. What's Bayes' estimate of θ given square loss.

Posterior mean: $\frac{\sum_{i=1}^{n} X_i + \alpha}{n+\alpha+\beta}$, why?

Example 3: Consider the model

$$\begin{aligned} X_i|\theta &\sim \text{iid Normal}, N(\theta, \sigma^2), \text{where } \sigma^2 \text{ is known} \\ \Theta &\sim N(\theta_0, \sigma_0^2), \text{ where } \theta_0 \text{ and } \sigma_0^2 \text{ are known} \end{aligned}$$

Posterior pdf.

$$\begin{aligned} k(\theta|\mathbf{X}) &\propto \frac{1}{[\sqrt{2\pi}\sigma]^n} \exp\left[-\frac{\sum_{i=1}^n (X_i-\theta)^2}{2\sigma^2}\right] \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left[-\frac{(\theta-\theta_0)^2}{2\sigma_0^2}\right] \\ &\propto \exp\left[-\frac{\sum (X_i-\theta)^2}{2\sigma^2} - \frac{(\theta-\theta_0)^2}{2\sigma_0^2}\right] \\ &\propto \exp\left[-\frac{\theta^2}{2}\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right) + \frac{2\theta}{2}\left(\frac{\sum X_i}{\sigma^2}+\frac{\theta_0}{\sigma_0^2}\right)\right] \\ &\propto \exp\left[-\frac{1}{2\frac{\sigma^2\sigma_0^2}{n\sigma_0^2+\sigma^2}}\left(\theta - \frac{\sum X_i\sigma_0^2+\theta_0\sigma^2}{n\sigma_0^2+\sigma^2}\right)^2\right] \end{aligned}$$

The Posterior pdf is $N(\frac{\sum X_i\sigma_0^2+\theta_0\sigma^2}{n\sigma_0^2+\sigma^2}, \frac{\sigma^2\sigma_0^2}{n\sigma_0^2+\sigma^2})$. What's Bayes' estimate of θ given square loss.

Posterior mean: $\frac{\sum X_i\sigma_0^2+\theta_0\sigma^2}{n\sigma_0^2+\sigma^2}$, i.e. $\frac{\sigma_0^2}{\sigma_0^2+\sigma^2/n}\bar{X} + \frac{\sigma^2/n}{\sigma_0^2+\sigma^2/n}\theta_0$

Note that it is a weighted average of the MLE $\bar{X}$ and the prior mean θ_0. As in the last example, for large n the Bayes's estimator is close to the MLE and $\delta(\mathbf{X})$ is a consistent estimator of θ.

Thus the Bayesian procedures permit the decision make to enter his or her prior opinions into the solution in a very formal ways such that the influences of these prior notions will be less and less as n increases.

If an interval estimate of θ is desired, we can find two functions $u(\mathbf{X})$ and $v(\mathbf{X})$ so that the conditional probability

$$P(u(\mathbf{X}) \leq \Theta \leq v(\mathbf{X})) = \int_{u(\mathbf{X})}^{v(\mathbf{X})} k(\theta|\mathbf{X} d\theta$$

is $1-\alpha$. Then the interval $u(\mathbf{X})$ to $v(\mathbf{X})$ is an interval estimate of θ in the sense that the conditional probability of Θ belonging to that interval is equal to $1-\alpha$. These intervals are often called **Bayesian Interval**, so as not to confuse them with confidence intervals.

Example 3 continued: The Posterior pdf is

$$N\left(\frac{\bar{X}\sigma_0^2 + \theta_0\sigma^2/n}{\sigma_0^2 + \sigma^2/n}, \frac{\sigma^2/n\sigma_0^2}{\sigma_0^2 + \sigma^2/n}\right)$$

$1 - \alpha$ Bayesian Interval is

$$\frac{\bar{X}\sigma_0^2 + \theta_0\sigma^2/n}{\sigma_0^2 + \sigma^2/n} \pm z(\alpha/2)\sqrt{\frac{\sigma^2/n\sigma_0^2}{\sigma_0^2 + \sigma^2/n}}$$

For $z(\alpha/2)$, see Example 3 continued about Large Sample Theory for MLE in Lecture 2.

Example 1 continued: The Posterior pdf is

$$G(\sum_{i=1}^{n} X_i + \alpha, (n+\lambda)).$$

Recall if $Y \sim G(\alpha, \beta)$, then for any $c > 0$,

$$cY \sim G(\alpha, \beta/c)$$

$$G(\alpha, \beta = 1/2) = \chi_{2\alpha}$$

$$2(n+\lambda)G(\sum_{i=1}^{n} X_i + \alpha, (n+\lambda)) \sim \chi_{2(\sum_{i=1}^{n} X_i + \alpha)}$$

$1 - \alpha$ Bayesian Interval is

$$\left(\frac{1}{2(n+\lambda)}\chi_{2(\sum_{i=1}^{n} X_i + \alpha)}(1-\alpha/2), \frac{1}{2(n+\lambda)}\chi_{2(\sum_{i=1}^{n} X_i + \alpha)}(\alpha/2)\right)$$

Suppose we are interested in testing the hypotheses

$$H_0 : \theta \in \omega_0 \quad \text{versus} \quad H_a : \theta \in \omega_1$$

where $\omega_0 \cup \omega_1 = \Omega$ and $\omega_0 \cap \omega_1 = \emptyset$

We use the posterior distribution to compute the following probabilities

$$P(\Theta \in \omega_0 | \mathbf{X}) \quad \text{and} \quad P(\Theta \in \omega_1 | \mathbf{X})$$

A simple rule is to

$$\text{Accept} \quad H_0 \quad \text{if} \quad P(\Theta \in \omega_0 | \mathbf{X}) \geq P(\Theta \in \omega_1 | \mathbf{X})$$

Example 1 continued: Suppose we are interested in testing

$$H_0 : \theta \leq 10 \quad \text{versus} \quad H_a : \theta > 10$$

Suppose we think θ is about 12 but are not quite sure. Hence, we choose the $G(10, 1/1.2)$ pdf as our prior. The data for the problem are (11 7 11 6 5 9 14 10 9 5 8 10 8 10 12 9 3 12 14 4) with $\sum_{i=1}^{20} X_i = 177$.
The Posterior pdf is

$$G(177 + 10, (20 + 1/1.2)) = G(187, 20.833).$$

We compute the posterior probability of H_0 as

$$P(\Theta \leq 10|\mathbf{X}) = P(G(187, 20.833) \leq 10) = 0.9368$$

Thus $P(\Theta > 10|\mathbf{X}) = 1 - 0.9368 = 0.0632$; consequently, our rule would accept H_0.
The 95% Bayesian Interval, see previous Example, is (7.77, 10.31), which does contain 10.