Republic of Iraq Ministry of Higher Education & Research

University of Anbar

College of Education for Pure Sciences

محاضرات الاحصاء ٢

مدرس المادة : الاستاذ المساعد الدكتور فراس شاكر محمود

BAYESIAN ESTIMATION

We now describe another approach to estimation that is used by a group of Statisticians who call themselves Bayesians .To understand their approach Fully would require more text than we can allocate to this topic, but let us Begin this brief introduction by considering a simple application of the theorem of the Reverend Thomas Bayes.

Example:

Suppose we know that we are going to select an observation from a Poisson distribution with mean $\boldsymbol{\lambda}$ equal to 2 or 4. Moreover, prior to performing the experiment, we believe that $\lambda = 2$ has about four times as much chance Of being the parameter as does $\lambda = 4$; that is the prior probabilities are $P(\lambda = 2) = 0.8$ and $P(\lambda = 4) = 0.2$.

Solution:

The experiment is now performed and we observe that$x = 6$. At this point, our intuition tells us that $\lambda = 2$ seems less likely than before, as the observation $x = 6$ is much more probable with $\lambda = 4$ than with $\lambda = 2$, because, in an obvious notation,

$$P(X = 6/\lambda = 2) = 0.995 - 0.983 = 0.012$$

and

$$P(X = 6/\lambda = 4) = 0.889 - 0.785 = 0.104,$$

from Table .Our intuition can be supported by computing the conditional probability of $\lambda = 2$, given that $X = 6$:

$$P(\lambda = 2/X = 6) = \frac{P(\lambda = 2, X = 6)}{P(X = 6)}$$

$$= \frac{P(\lambda = 2)P(X = 6/\lambda = 2)}{P(\lambda = 2)P(X = 6/\lambda = 2) + P(\lambda = 4)P(X = 6/\lambda = 4)}$$

$$= \frac{(0.8)(0.012)}{(0.8)(0.012) + (0.2)(0.104)} = 0.316.$$

This conditional probability is called the posterior probability of $\lambda = 2$, given the single data point (here, $x = 6$).In a similar fashion, the posterior probability of $\lambda = 4$ is found to be 0.684 thus, we see that the probability of $\lambda = 2$ has decreased from 0.8 (the prior probability) to 0.316 (the posterior probability) with the observation of $x = 6$.

<u>Example:</u>

Suppose that Y has a binomial distribution with parameters n and $p = \theta$.

Then the pmf of Y, given $\boldsymbol{\theta}$, is

$$g(y/\theta) = \binom{n}{y} \theta^{y}(1-\theta)^{n-y}, \qquad y = 0,1,2, \dots, n$$

<u>Solution:</u>

Let us take the prior pdf of the parameter to be the beta pdf:-

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

Such a prior pdf provides a Bayesian a great deal of flexibility through the selection of the parameters α and β. Thus, the joint probabilities can be described by a product of a binomial pmf with parameters n and $\boldsymbol{\theta}$ and this beta pdf, namely,

$$k(y, \theta) = \binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1},$$

On the support given by $y = 0,1,2, \dots, n$ and $0 < \theta < 1$. We find

$$k_1(y) = \int_0^1 k(y, \theta) d\theta$$

$$= \binom{n}{y} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\,\Gamma(\beta)} \frac{\Gamma(\alpha + y)\,\Gamma(n + \beta - y)}{\Gamma(n + \alpha + \beta)}$$

On the support $y = 0,1,2, \ldots, n$ by comparing the integral with one involving a bete pdf with parameters $y + \alpha$ and $n - y + \beta$. Therefore,

$$k(\theta/y) = \frac{k(y, \theta)}{k_1(y)}$$

$$= \frac{\Gamma(n + \alpha + \beta)}{\Gamma(\alpha + y)\,\Gamma(n + \beta - y)} \theta^{y+\alpha-1}(1 - \theta)^{n-y+b-1}, \qquad 0 < \theta < 1$$

Which is a beta pdf with parameters $y + \alpha$ and $n - y + \beta$. With the squared error loss function we must minimize, with respect to w(y), the integral

$$\int_0^1 [\theta - w(y)]^2\, k(\theta/y)\, d\theta,$$

to obtain the Bayes estimator. But, as noted earlier, if Z is a random variable with

A second moment, then $E[(Z - b)^2]$ is minimized by $b = E(Z)$. In the preceding integration, $\boldsymbol{\theta}$ is like the Z with pdf $k(\theta/y)$, and w(y) is like the b, so the minimization is accomplished by taking

$$w(y) = E(\theta/y) = \frac{\alpha + \beta}{\alpha + \beta + n}$$

Which is the mean of the beta distribution with parameters $y + \alpha$ and $n - y + \beta$.

It is instructive to note that this Bayes estimator can be written as

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

Which is a weighted average of the maximum likelihood estimate y/n of $\boldsymbol{\theta}$ and the mean $\alpha/(\alpha+\beta)$ of the prior pdf of the parameter. Moreover, the respective weights are $n/(\alpha+\beta+n)$ and $(\alpha+\beta)/(\alpha+\beta+n)$.Thus, we see that α and β should be selected so that not only is $\alpha/(\alpha+\beta)$ the desired prior mean, but also the sum $(\alpha+\beta)$ plays a role corresponding to a sample size. That is, if we want our prior opinion to have as much weight as a sample size of 20, we would take $\alpha+\beta=20$. So if our prior mean is 3/4, we select α=15 and β=5. That is, the prior pdf of $\boldsymbol{\theta}$ is beta (15, 5).If we observe n=40 and y=28, then the posterior pdf is beta (28+15=43, 12+5=17).

Example:

Let us consider again Example2, but now say that $X_1, X_2, \dots X_n$ is a random sample from the Bernoulli distribution with pmf

$$f(x/\theta) = \theta^x(1-\theta)^{1-x}, \quad x = 0,1.$$

With the same prior pdf of $\boldsymbol{\theta}$, the joint distribution of $X_1, X_2, \dots X_n$ and $\boldsymbol{\theta}$

given by

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\theta^{\sum_{i=1}^{n}x_i}(1-\theta)^{n-\sum_{i=1}^{n}x_i}, 0<\theta<1, x_i=0,1.$$

Of course, the posterior pdf of $\boldsymbol{\theta}$, given that $X_1=x_1, X_2=x_2, \dots, X_n=x_n$,

Is such that

$$k(\theta/x_1, x_2, \dots, x_n) \propto \theta^{\sum_{i=1}^{n}x_i+\alpha-1}(1-\theta)^{n-\sum_{i=1}^{n}x_i+\beta-1}, 0<\theta<1,$$

Which is beta with $\alpha^* = \sum_{i=1}^{n} x_i + \alpha$, $\beta^* = n - \sum_{i=1}^{n} x_i + \beta$,the conditional mean of $\boldsymbol{\theta}$ is

$$\frac{\sum_{i=1}^{n}x_i+\alpha}{n+\alpha+\beta} = \left(\frac{n}{n+\alpha+\beta}\right)\left(\frac{\sum_{i=1}^{n}x_i}{n}\right) + \left(\frac{\alpha+\beta}{n+\alpha+\beta}\right)\left(\frac{\alpha}{\alpha+\beta}\right),$$

Which, with $=\sum x_i$, is exactly the same result as that of Example 2.

MORE BAYESIAN CONCEPTS

Let $X_1, X_2, \dots X_n$ be a random sample from a distribution with pdf (pmf) $f(x/\theta)$, and let h ($\boldsymbol{\theta}$) be the prior pdf. Then the distribution associated with the marginal pdf of $X_1, X_2, \dots X_n$ namely,

$$k_1(x_1, x_2, \dots, x_n) = \int_{-\infty}^{\infty} f(x_1/\theta) f(x_2/\theta) \dots f(x_n/\theta) h(\theta) d\theta,$$

Is called the predictive distribution because it provides the best description of the Probabilities on $X_1, X_2, \dots X_n$. Often this creates some interesting distributions. For example, suppose there is only one X with the normal pdf

$$f(x/\theta) = \frac{\sqrt{\theta}}{\sqrt{2\pi}} e^{-(\theta x^2)/2}, \quad -\infty < x < \infty.$$

Here, $\theta = 1/\sigma^2$, the inverse of the variance, is called the precision of X. Say this precision has the gamma pdf

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

Then the predictive pdf is

$$k_1(x) = \int_0^{\infty} \frac{\theta^{\alpha+\frac{1}{2}-1} e^{-(\frac{x^2}{2}+\frac{1}{\beta})\theta}}{\Gamma(\alpha)\, \beta^{\alpha}\sqrt{2\pi}} d\theta$$

$$= \frac{\Gamma(\alpha + 1/2)}{\Gamma(\alpha)\, \beta^{\alpha}\sqrt{2\pi}} \frac{1}{(1/\beta + x^2/2)^{\alpha+1/2}}, \quad -\infty < x < \infty$$

Note that if $\alpha = r/2$ and $\beta = 2/r$, where r is a positive integer, then

$$k_1(x) \propto \frac{1}{(1 + x^2/r)^{(r+1)/2}}, \quad -\infty < x < \infty$$

Which is a t pdf with r degrees of freedom. So if the inverse of the variance- or precision $\boldsymbol{\theta}$-of a normal distribution varies as a gamma random variable , a generalization of a t distribution has been created that has heavier tails than the normal distribution. This mixture of normal (different from a mixed distribution) is attained by weighing with the gamma distribution in a process often called compounding.

Another illustration of compounding is given in the next example.

<u>Example:</u>

Suppose X has a gamma distribution with the two parameters k and θ^{-1}. (That is, the usual α is replaced by k and $\boldsymbol{\theta}$ by its reciprocal). Say h($\boldsymbol{\theta}$) is gamma with parameters α and β, so that

$$k_1(x) = \int_0^\infty \frac{\theta^k x^{k-1} e^{-\theta x}}{\Gamma(k)} \frac{1}{\Gamma(\alpha)\, \beta^\alpha} \theta^{\alpha-1} e^{-\theta/\beta} d\theta$$

$$= \int_0^\infty \frac{x^{k-1} \theta^{k+\alpha-1} e^{-\theta(x+1/\beta)}}{\Gamma(k)\, \Gamma(\alpha)\, \beta^\alpha} d\theta$$

$$= \frac{\Gamma(k+\alpha) x^{k-1}}{\Gamma(k)\, \Gamma(\alpha) \beta^\alpha} \frac{1}{(x+1/\beta)^{k+\alpha}}$$

$$= \frac{\Gamma(k+x) \beta^k x^{k-1}}{\Gamma(k)\, \Gamma(\alpha) (1+\beta x)^{k+\alpha}} \;, 0 < x < \infty.$$

Of course, this is a generalization of the F distribution, which we obtain by letting

$$\alpha = r_2/2\,, \qquad k = r_1/2\;, \qquad and\;\; \beta = r_1/r_2.$$

<u>Example:</u>

(Berry, 1996) This example deals with predictive probabilities, and it concerns the breakage of glass panels in high-rise buildings. One such case involved 39 panels, and of the 39 panels that broke, it was known that 3 broke due to nickel sulfide (NiS) stones found in them. Loss of evidence prevented the causes of breakage of the other 36 panels from being known. So the court wanted to know whether the manufacturer of the panels or the builder was at fault for the breakage of these 36 panels.

From expert testimony, it was thought that usually about 5% breakage is caused By NiS stones. That is, if this value of p is selected from a beta distribution, we have

$$\frac{\alpha}{\alpha+\beta} = 0.05$$

Moreover, the expert thought that if two panels from the same lot break and one breakage was caused by NiS stones, then, due to the pervasive nature of the manufacturing process, the probability of the second panel breaking due to NiS stones increases to about 95%. Thus, the posterior estimate of p (see Example 2) with one "success" after one trial is

$$\frac{\alpha+1}{\alpha+\beta+1} = 0.95$$

Solving Equations 3 and 4 for α and β, we obtain

$\alpha = \frac{1}{360}$ and $\beta = \frac{19}{360}$

Now updating the posterior probability with 3 "success" out of 3 trials, we obtain the posterior estimate of p:

$$\frac{\alpha+3}{\alpha+\beta+3} = \frac{1/360+3}{20/360+3}$$

$$= \frac{1081}{1100} = 0.983.$$

Of course, the court that heard the case wanted to know the expert's opinion about the probability that all of the remaining 36 panels broke because of NiS stones. Using updated probabilities after the third break, then the fourth, and so on, we obtain the product

$$\left(\frac{1/360+3}{20/360+3}\right)\left(\frac{1/360+4}{20/360+4}\right)\left(\frac{1/360+5}{20/360+5}\right)\ldots\left(\frac{1/360+38}{20/360+38}\right) = 0.8664.$$

That is, the expert held that the probability that all 36 breakages were caused by NiS stones was about 87%, which is the needed value in the court's decision.