Lecture 23: Confidence Interval

18 April, 2018

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

Definition 23.1 *We say that an interval estimator* $[L(\mathbf{X}), U(\mathbf{X})]$ *of a parameter* θ *has confidence level* $1-\alpha$ *(*α *does not depend on* θ*) if*

$$P\{L(\mathbf{X}) \leq \theta, \theta \leq U(\mathbf{X})\} = 1-\alpha,$$

Equivalently, we say that $[L(\mathbf{X}), U(\mathbf{X})]$ *is a* $(1-\alpha)100\%$ *confidence interval for* θ*.*

We would like α to be small. Common values for α are $0.1, .05$, and $.01$ which correspond to confidence levels 90%, 95%, and 99% respectively. Thus, when we are asked to find a 95% confidence interval for a parameter θ, we need to find $L(\mathbf{X})$ and $U(\mathbf{X})$ such that

$$P\{L(\mathbf{X}) \leq \theta \leq U(\mathbf{X})\} \geq 0.95$$

Note that $P\{L(\mathbf{X}) \leq \theta, \theta \leq U(\mathbf{X})\}$ can be equivalently written as

$$P\{L(\mathbf{X}) \leq \theta \leq U(\mathbf{X})\}, \quad \text{or} \quad P\{\theta \in [L(\mathbf{X}), U(\mathbf{X})]\}. \tag{23.1}$$

But it is important to keep in mind that the interval is the random quantity, not the parameter. Therefore when we write probability statements such as in (23.1), these probability statements refer to $\mathbf{X}$, not θ.

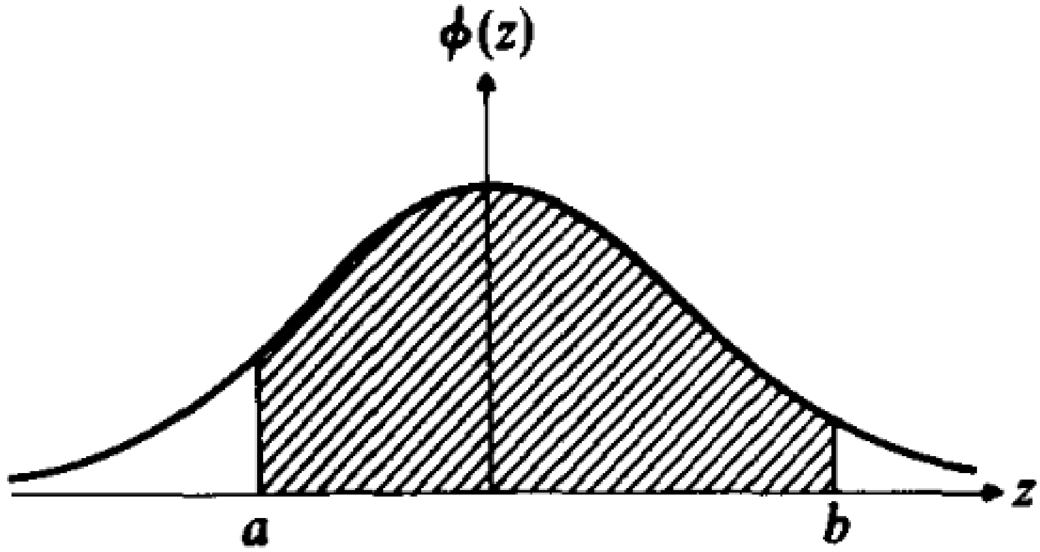
For a random sample X_1, X_2, X_3, X_4 from a $N(\mu, 1)$ distribution, an interval estimator of μ is $[\bar{X}-1, \bar{X}+1]$ has confidence level 0.9544. Similarly, intervals with any desired degree of confidence between 0 and 1 can be obtained. Thus, since

$$P[-2.58 < Z < 2.58] = 0.99, \text{ where } Z \sim N(0,1)$$

a 99% confidence interval for the true mean is obtained by converting the inequalities as before to get

$$P[\bar{X} - 3.58 \leq \mu \leq \bar{X} + 3.58] = 0.99.$$

It is to be observed that there are, in fact, many possible intervals with the same probability (with the same confidence coefficient). Any two numbers a and b such that 95 percent of the area under graph of normal density function $\phi(\cdot)$ lies between a and b will determine a 95 per-



cent confidence interval.

Finding Confidence Interval

Now, let's talk about how we can find interval estimators. The method of finding a confidence interval that has been illustrated in Lecture-22, followed by the Example 22.5, is a general method. The method entails finding, if possible, a function $\left(\frac{\bar{X} - \mu}{\sqrt{\frac{1}{4}}} \text{ in Example 22.5} \right)$ of the sample and the parameter to be estimated which has a distribution independent of the parameter and any other parameters. Then any probability statement of the form $P[a \leq Z \leq b] = 1 - \alpha$ for known a and b, where Z is the function, will give rise to a probability statement about the parameter that we hope can be rewritten to give a confidence interval. This technique is applicable in many important problems.

We will describe one method of finding confidence intervals and call it the pivotal-quantity method.

As before, we assume a random sample $X_1, \cdots, X_n$ from some pmf or pdf $f(\cdot|\theta)$ parameterized by a real-constant θ. Our object is to find a confidence-interval estimate of θ.

Definition 23.2 *Let $X_1, \cdots, X_n$ be a random sample from the population $f(\cdot|\theta)$. Let $Q = f(X_1, \cdots, X_n, \theta)$; that is, let Q be a function of $X_1, \cdots, X_n$ and θ. If Q has a distribution that does not depend on θ and any other parameter, then Q is defined to be a pivotal quantity.*

Example 23.3 *Let $X_1, \cdots, X_n$ be a random sample from $N(\mu, 9)$. $\bar{X}-\mu$ is a pivotal quantity since $\bar{X}-\mu$ is normally distributed with mean 0 and variance $\frac{9}{n}$. Also $\dfrac{\bar{X}-\mu}{\sqrt{\frac{9}{n}}}$ has a standard normal distribution and, hence, is a pivotal quantity. On the other hand, $\dfrac{\bar{X}}{\mu}$ is not a pivotal quantity since $\dfrac{\bar{X}}{\mu}$ is normally distributed with mean unity and variance $\dfrac{9}{\mu^2 n}$, which depends on μ.*

EXAMPLE **8.4** Suppose that we are to obtain a single observation Y from an exponential distribution with mean θ. Use Y to form a confidence interval for θ with confidence coefficient .90.

Solution The probability density function for Y is given by

$$f(y) = \begin{cases} \left(\dfrac{1}{\theta}\right) e^{-y/\theta}, & y \geq 0, \\ 0, & \text{elsewhere.} \end{cases}$$

By the transformation method of Chapter 6 we can see that $U = Y/\theta$ has the exponential density function given by

$$f_U(u) = \begin{cases} e^{-u}, & u > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

The density function for U is graphed in Figure 8.5. $U = Y/\theta$ is a function of Y (the sample measurement) and θ, and the distribution of U does not depend on θ. Thus, we can use $U = Y/\theta$ as a pivotal quantity. Because we want an interval estimator with confidence coefficient equal to .90, we find two numbers a and b such that

$$P(a \leq U \leq b) = .90.$$

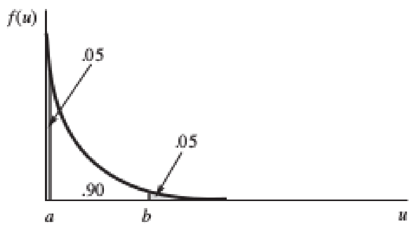


FIGURE 8.5 Density function for U, Example 8.4

One way to do this is to choose a and b to satisfy

$$P(U < a) = \int_0^a e^{-u}\,du = .05 \quad \text{and} \quad P(U > b) = \int_b^\infty e^{-u}du = .05.$$

These equations yield

$$1 - e^{-a} = .05 \quad \text{and} \quad e^{-b} = .05 \quad \text{or, equivalently,} \quad a = .051,\ b = 2.996.$$

It follows that

$$.90 = P(.051 \le U \le 2.996) = P\left(.051 \le \frac{Y}{\theta} \le 2.996\right).$$

Because we seek an interval estimator for θ, let us manipulate the inequalities describing the event to isolate θ in the middle. Y has an exponential distribution, so $P(Y > 0) = 1$, and we maintain the direction of the inequalities if we divide through by Y. That is,

$$.90 = P\left(.051 \le \frac{Y}{\theta} \le 2.996\right) = P\left(\frac{.051}{Y} \le \frac{1}{\theta} \le \frac{2.996}{Y}\right).$$

Taking reciprocals (and hence reversing the direction of the inequalities), we obtain

$$.90 = P\left(\frac{Y}{.051} \ge \theta \ge \frac{Y}{2.996}\right) = P\left(\frac{Y}{2.996} \le \theta \le \frac{Y}{.051}\right).$$

Thus, we see that $Y/2.996$ and $Y/.051$ form the desired lower and upper confidence limits, respectively. To obtain numerical values for these limits, we must observe an actual value for Y and substitute that value into the given formulas for the confidence limits. We know that limits of the form $(Y/2.996, Y/.051)$ will include the true (unknown) values of θ for 90% of the values of Y we would obtain by repeatedly sampling from this exponential distribution. ■