Lecture 19. Confidence Interval

December 5, 2011

The phrase "**confidence interval**" can refer to a random interval, called an interval estimator, that covers the true value θ_0 of a parameter of interest with a prespecified probability. The prespecified probability (usually something 90%, 95% or 99%) is called the "coverage probability" or the "confidence level". The phrase could also refer to a realization of that random interval. Theoretical econometricians usually use the first meaning whereas applied econometricians usually use the second meaning. In this lecture, we do not explicitly distinguish the two. It is worth noting that when the second meaning is used, it does not make sense to say the true value lies in the confidence interval with such and such probability.[1] This is because in classical statistics, θ_0 is not considered random, the realization of a random interval is also nonrandom. The probability that a non-random interval covers a nonrandom true value is either zero or one and cannot be anything in between. Thus, it is more conventional to say that the confidence interval of confidence level $100(1-\alpha)\%$ is $[-0.01, 0.01]$.

Construct a confidence interval

To construct a confidence interval, one typically start with an estimator and build an interval around it. The idea is that, if the estimator is good then it is likely to be close to the true value. Thus an interval around the estimator is likely to cover the true value.

Example 1. Suppose we would like to construct a confidence interval for the expection, μ, of X using a random n-sample $\{X_1, ..., X_n\}$. We start with an estimator, say the sample mean $\bar{X}_n$ and construct an interval of the shape:

$$[\bar{X}_n - c_n, \bar{X}_n + c_n],$$

for some positive (possibly random) c_n. The c_n is chosen to make the interval have a pre-

[1] For example, it does not make sense to say that θ_0 lies in $[-0.01, 0.01]$ with $100(1-\alpha)\%$ probability.

specified coverage probability either exactly:

$$\Pr{}_\mu(\bar{X}_n - c_n \leq \mu \leq \bar{X}_n + c_n) = 1 - \alpha; \tag{1}$$

or asymptotically:

$$\lim_{n\to\infty} \Pr{}_\mu(\bar{X}_n - c_n \leq \mu \leq \bar{X}_n + c_n) = 1 - \alpha. \tag{2}$$

Case 1. $X \sim N(\mu, \sigma^2)$ with known σ^2. Then, $\bar{X}_n \sim N(\mu, \sigma^2/n)$, and hence $\sqrt{n}(\bar{X}_n - \mu)/\sigma \sim N(0, 1)$. Using this fact, we can determine c_n as follows:

$$\begin{aligned}
\Pr{}_\mu(\bar{X}_n - c_n \leq \mu \leq \bar{X}_n + c_n) &= \Pr{}_\mu(|\bar{X}_n - \mu| \leq c_n) \\
&= \Pr{}_\mu(|\sqrt{n}(\bar{X}_n - \mu)/\sigma| \leq \sqrt{n}c_n/\sigma) \\
&= 1 - 2(1 - \Phi(\sqrt{n}c_n/\sigma)) \\
&= 2\Phi(\sqrt{n}c_n/\sigma) - 1. \qquad (3)
\end{aligned}$$

The second to the last step uses the fact that the density of standard normal is symmetric around zero, thus $\Phi(x) = 1 - \Phi(-x)$. Set the coverage probability to $1 - \alpha$, and we have

$$\Phi(\sqrt{n}c_n/\sigma) = 1 - \alpha/2. \tag{4}$$

Use z_α to denote the $(1-\alpha)$ quantile of $N(0, 1)$. Then the appropriate c_n should be $\sigma z_{\alpha/2}/\sqrt{n}$. And the confidence interval for μ is

$$[\bar{X}_n - \sigma z_{\alpha/2}/\sqrt{n}, \bar{X}_n + \sigma z_{\alpha/2}/\sqrt{n}]. \tag{5}$$

Remark. This is a "symmetric confidence interval" in that it is symmetric around the estimator $\bar{X}_n$. It is also an "equal-tailed confidence interval" in that the probability that the whole interval falls on one side of μ equals the probability that it falls on the other side of μ:

$$\Pr{}_\mu(\bar{X}_n - \sigma z_{\alpha/2}/\sqrt{n} \geq \mu) = \Pr{}_\mu(\bar{X}_n + \sigma z_{\alpha/2}/\sqrt{n} \leq \mu) = \alpha/2. \tag{6}$$

The symmetric confidence interval and the equal-tailed confidence interval coincide in this case because $\bar{X}_n$ (the estimator)'s distribution is symmetric around μ (the true value)

Why does one want to restrict the confidence interval to be symmetric or "equal-tailed"? Why not consider confidence intervals of the shape, for example, $[\bar{X}_n - c_n, \bar{X}_n + 10c_n]$? It is as easy to find c_n that makes this confidence interval have the prespecified coverage probability as above. But the problem is that this interval will be longer than the symmetric one. (c_n will be smaller than $\sigma z_{\alpha/2}/\sqrt{n}$, but $10c_n$ will be much bigger than $\sigma z_{\alpha/2}/\sqrt{n}$.)

Longer interval estimators are considered less informative typically. There is one special case, though, which is the "one-sided confidence interval". The one sided confidence interval is of the shape $(-\infty, \bar{X}_n + c_n]$ or $[\bar{X}_n - c_n, \infty)$. These intervals are of infinite length, but they are frequently used when one side of the parameter space (space of μ) is more interesting than the other. For example, when it is of interest to conclude "we have 95% confidence that μ is above (some number)", one can use $[\bar{X}_n - c_n, \infty)$ instead of the two-sided intervals. At the same confidence level, this interval gives you the largest lower bound number that you can insert in your conclusion.

(Exercise: find c_n for the confidence interval $[\bar{X}_n - c_n, \bar{X}_n + 10c_n]$ and for the confidence interval $[\bar{X}_n - c_n, \infty)$ to have 95% coverage probability.)

Case 1. In the above case, we assume that σ is known. If σ is not known, then the confidence interval above is not feasible. In order to find c_n, in the second step of (3), we should not divide by σ. Instead, we can divide by the sample variance S_X. Then

$$\Pr{}_\mu(\bar{X}_n - c_n \le \mu \le \bar{X}_n + c_n) = \Pr{}_\mu(\sqrt{n}|\bar{X}_n - \mu|/S_X \le \sqrt{n}c_n/S_X). \tag{7}$$

The random variable $\sqrt{n}(\bar{X}_n - \mu)/S_X$ is not $N(0,1)$, but it has the student t-distribution, which is shown next. As above we know $\sqrt{n}(\bar{X}_n - \mu)/\sigma \sim N(0,1)$. The random variable $(n-1)S_X^2/\sigma^2 \sim \chi^2(1)$ because:

$$\begin{aligned}
(n-1)S_X^2/\sigma^2 &\equiv \sum_{i=1}^n (X_i - \bar{X}_n)^2/\sigma^2 \\
&= \begin{pmatrix} \sigma^{-1}(X_1-\mu) \\ \dots \\ \sigma^{-1}(X_n-\mu) \end{pmatrix}^{\prime} (I_n - n^{-1}1_n 1_n^{\prime}) \begin{pmatrix} \sigma^{-1}(X_1-\mu) \\ \dots \\ \sigma^{-1}(X_n-\mu) \end{pmatrix},
\end{aligned}$$

where I_n is a n-dimensional identity matrix, 1_n is a n-vector of ones. The matrix $I_n - n^{-1}1_n 1_n^{\prime}$ is idempotent, and the vector $\begin{pmatrix} \sigma^{-1}(X_1-\mu) \\ \dots \\ \sigma^{-1}(X_n-\mu) \end{pmatrix} \sim N(0, I_n)$. The quadractic form of a standard normal vector with an idempotent weight matrix follows a chi-squared distribution of degree equal to the trace of the weight matrix. Thus,

$$\begin{aligned}(n-1)S_X^2/\sigma^2 &\sim \chi^2(trace(I_n - n^{-1}1_n1_n^{'})) \\ &= \chi^2(trace(I_n) - n^{-1}trace(1_n1_n^{'})) \\ &= \chi^2(trace(I_n) - n^{-1}trace(1_n^{'}1_n)) \\ &= \chi^2(n-1). \end{aligned} \tag{8}$$

It can also be shown that $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ is independent of $(n-1)S_X^2/\sigma^2$. Therefore,

$$\begin{aligned}\sqrt{n}(\bar{X}_n - \mu)/S_X &= \frac{\sqrt{n}(\bar{X}_n - \mu)/\sigma}{\sqrt{((n-1)S_X^2/\sigma^2)/n-1}} \\ &= \frac{Z}{\sqrt{W/n-1}},\end{aligned}$$

where $Z \sim N(0,1)$ and $W \sim \chi^2(n-1)$ and Z is independent of W. By direct calculation of the density of $\frac{Z}{\sqrt{W/n-1}}$, one can show that it is the density for the student t distribution with degree of freedom $n-1$: $t(n-1)$. Thus, setting (7) to $1-\alpha$ gives:

$$\begin{aligned}1-\alpha &= \Pr(|t(n-1)| \leq \sqrt{n}c_n/S_X) \\ &= 2F_{t(n-1)}(\sqrt{n}c_n/S_X) - 1.\end{aligned}$$

Thus, $c_n = t_{\alpha/2,n-1}S_X/\sqrt{n}$, where $t_{\alpha/2,n-1}$ is the $1-\alpha/2$ quantile of $t(n-1)$. The confidence interval for μ is

$$[\bar{X}_n - t_{\alpha/2,n-1}S_X/\sqrt{n}, \bar{X}_n + t_{\alpha/2,n-1}S_X/\sqrt{n}].$$

Case 2. The confidence intervals in the above two cases are "exact" confidence intervals, i.e. their coverage probabilities equals the prespecified confidence level $1-\alpha$ (called "nominal confidence level") exactly for any sample size. This is possible only because the estimator (and its sample standard deviation)'s distribution is known and tractable, which is so because we assume the X has normal distribution. Without the normality assumption, the distributions of $\bar{X}_n$ and S_X are either unknown or in principal known but very untractable. Then, we will need to use confidence intervals that do not have exact coverage probability (1), but have asymptotic coverage probability (2). We have shown that (suppose $E[X^2] < \infty$ and $Var(X) > 0$)

$$\sqrt{n}(\bar{X}_n - \mu) \to_d N(0, \sigma^2) \text{ and } S_X \to_p \sigma.$$

By the continuous mappint theorem: $\sqrt{n}(\bar{X}_n - \mu)/S_X \to_d N(0,1)$. If we set $\sqrt{n}c_n/S_X$ to a fixed number c, then

$$\lim_{n\to\infty} \Pr{}_\mu(\bar{X}_n - c_n \le \mu \le \bar{X}_n + c_n) = \lim_{n\to\infty} \Pr{}_\mu(|\sqrt{n}(\bar{X}_n - \mu)/S_X| \le c)$$
$$=2\Phi(c) - 1.$$

Setting it to $1-\alpha$ gives $c = z_{\alpha/2}$. Therefore, we can set $c_n = z_{\alpha/2}S_X/\sqrt{n}$ and the random interval $[\bar{X}_n - c_n, \bar{X}_n + c_n]$ has coverage probability $1-\alpha$ approximately in large samples.

Example 2. The large sample confidence interval can be constructed for any θ that has a $\sqrt{n}$-consistent and asymptotically normal estimator $\hat{\theta}_n$ in the same way as Case 3 above. Suppose that $\sqrt{n}(\hat{\theta}_n - \theta) \to_d N(0, \sigma^2)$, $\sigma^2 > 0$ and there is a consistent estimator $\hat{\sigma}_n$ for σ. Then, a two-sided symmetric confidence interval for θ of nominal confidence level $1-\alpha$ can be

$$[\hat{\theta}_n - z_{\alpha/2}\hat{\sigma}_n/\sqrt{n}, \hat{\theta}_n + z_{\alpha/2}\hat{\sigma}_n/\sqrt{n}].$$

(Exercise, find the one-sided confidence interval of nominal confidence level $1-\alpha$ for θ and show that it asymptotically has coverage probability $1-\alpha$.)

Example 3 (Confidence interval for Variance)**.** Confidence intervals do not have to be of the shape as above. Suppose $X \sim N(\mu, \sigma^2)$ with unknown σ^2. Find a confidence interval for σ^2. Suppose that the interval is $[c_{1n}, c_{2n}]$. Then we want it to satisfy:

$$\Pr(c_{1n} \le \sigma^2 \le c_{2n}) = 1 - \alpha.$$

In Case 2 above, we have found that $(n-1)S_X^2/\sigma^2 \sim \chi^2(n-1)$. This suggests that we can do the following manipulation to the above equation:

$$\begin{aligned}
1 - \alpha &= \Pr(c_{1n} \le \sigma^2 \le c_{2n}) \\
&= \Pr(c_{1n}/((n-1)S_X^2) \le \sigma^2/((n-1)S_X^2) \le c_{2n}/((n-1)S_X^2)) \\
&= \Pr((n-1)S_X^2/c_{2n} \le (n-1)S_X^2/\sigma^2 \le (n-1)S_X^2/c_{1n}) \\
&= F_{\chi^2(n-1)}((n-1)S_X^2/c_{1n}) - F_{\chi^2(n-1)}((n-1)S_X^2/c_{2n}).
\end{aligned}$$

Because $F_{\chi^2(n-1)}()$, S_X^2 and n are known, we can choose c_{1n} and c_{2n} to make the last expression $1-\alpha$. Clearly, there are many pairs (c_{1n}, c_{2n}) that can satisfy the coverage probability requirement. Which ones to use depends on what one is interested in. If one has no preference, a good choice is the equal-tailed confidence interval, which makes

$$\Pr(c_{2n} \le \sigma^2) = \alpha/2 = \Pr(c_{1n} > \sigma^2)$$

or equivalently:

$$1 - F_{\chi^2(n-1)}((n-1)S_X^2/c_{1n}) = F_{\chi^2(n-1)}((n-1)S_X^2/c_{2n}) = \alpha/2,$$

or $(n-1)S_X^2/c_{1n} = \chi^2_{\alpha/2,n-1}$ and $(n-1)S_X^2/c_{2n} = \chi^2_{1-\alpha/2,n-1}$, where $\chi^2_{\alpha,n-1}$ is the $1-\alpha$ quantile of $\chi^2(n-1)$. Consequently, $c_{1n} = (n-1)S_X^2/\chi^2_{\alpha/2,n-1}$ and $c_{2n} = (n-1)S_X^2/\chi^2_{1-\alpha/2,n-1}$.

(Think: how to construct (large sample) confidence intervals for σ^2 without the normality assumption on X? Hint: we have derived the asymptotic distribution of S_X^2 in a previous lecture.)