14.30 Introduction to Statistical Methods in Economics
Spring 2009

14.30 Introduction to Statistical Methods in Economics
Lecture Notes 20

Konrad Menzel

April 30, 2009

1 Confidence Intervals (continued)

The following example illustrates one way of constructing a confidence interval when the distribution of the estimator is not normal.

Example 1 *Suppose $X_1, \ldots, X_n$ are i.i.d. with $X_i \sim U[0, \theta]$, and we want to construct a 90% confidence interval for θ_0. Let*

$$\hat{\theta} = \max\{X_1, \ldots, X_n\} = X_{(n)}$$

the nth order statistic (as we showed last time, this is also the maximum-likelihood estimator). Even though, as we saw, $\hat{\theta}$ is not unbiased for θ, we can use it to construct a confidence interval for θ. From results for order statistics, we saw that the c.d.f. of $\hat{\theta}$ is given by the c.d.f. of $\hat{\theta}$ is given by

$$F_{\hat{\theta}}(\theta) = \begin{cases} 0 & \theta \leq 0 \\ \left(\frac{\theta}{\theta_0}\right)^n & \text{if } 0 < \theta \leq \theta_0 \\ 1 & \text{if } \theta > \theta_0 \end{cases}$$

where we plugged in the c.d.f. of a $U[0, \theta_0]$ random variable, $F(x) = \frac{x}{\theta_0}$.
In order to obtain the functions for A and B, let us first find constants a and b such that

$$P_{\theta_0}(a \leq \hat{\theta} \leq b) = F_{\hat{\theta}}(b) - F_{\hat{\theta}}(b) = 0.95 - 0.05 = 0.9$$

We can find a and b by solving

$$F_{\hat{\theta}}(a) = 0.05 \text{ and } F_{\hat{\theta}}(b) = 0.95$$

so that we obtain $a = \sqrt[n]{0.05}\theta_0$ and $b = \sqrt[n]{0.95}\theta_0$. This doesn't give us a confidence interval yet, since looking at the definition of a CI, we want the true parameter θ_0 in the middle of the inequalities, and the functions on either side depend only on the data and other known quantities.
However, we can rewrite

$$0.9 = P_{\theta_0}(a \leq \hat{\theta} \leq b) = P_{\theta_0}\left(\sqrt[n]{0.05}\theta_0 \leq \hat{\theta} \leq \sqrt[n]{0.95}\theta_0\right) = P_{\theta_0}\left(\frac{\hat{\theta}}{\sqrt[n]{0.95}} \leq \theta_0 \leq \frac{\hat{\theta}}{\sqrt[n]{0.05}}\right)$$

Therefore

$$[A, B] = [A(X_1, \ldots, X_n), B(X_1, \ldots, X_n)] = \left[\frac{\max\{X_1, \ldots, X_n\}}{\sqrt[n]{0.95}}, \frac{\max\{X_1, \ldots, X_n\}}{\sqrt[n]{0.05}}\right]$$

is a 90% confidence interval for θ_0. Notice that in this case, the bounds of the confidence intervals depend on the data only through the estimator $\hat{\theta}(X_1, \ldots, X_n)$. This need not be true in general.

Let's recap how we arrived at the confidence interval:

1. first get estimator/statistic $\hat{\theta}(X_1, \ldots, X_n)$ and the distribution of $\hat{\theta}$.

2. find $a(\theta), b(\theta)$ such that
$$P(a(\theta) \leq \hat{\theta} \leq b(\theta)) = 1 - \alpha$$

3. rewrite the event by solving for θ
$$P(A(X) \leq \theta \leq B(X)) = P(A(\hat{\theta}) \leq \theta \leq B(\hat{\theta})) = 1 - \alpha$$

4. evaluate $A(X), B(X)$ for the observed sample $X_1, \ldots, X_n$

5. the $1 - \alpha$ confidence interval is then given by
$$\widehat{CI} = [A(X_1, \ldots, X_n), B(X_1, \ldots, X_n)]$$

1.1 Important Cases

1. $\hat{\theta}$ is normally distributed, $\text{Var}(\hat{\theta}) \equiv \sigma_{\hat{\theta}}^2$ is known: can form confidence interval
$$[A(X), B(X)] = \left[\hat{\theta} - \sqrt{\sigma_{\hat{\theta}}^2} \Phi^{-1} \left(1 - \frac{\alpha}{2} \right), \hat{\theta} + \sqrt{\sigma_{\hat{\theta}}^2} \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \right]$$

2. $\hat{\theta}$ is normally distributed, $\text{Var}(\hat{\theta})$ unknown, but have estimator $\hat{S}^2 = \widehat{\text{Var}(\hat{\theta})}$: confidence interval is given by
$$[A(X), B(X)] = \left[\hat{\theta} - \sqrt{\hat{S}^2} t_{n-1} \left(1 - \frac{\alpha}{2} \right), \hat{\theta} + \sqrt{\hat{S}^2} t_{n-1} \left(1 - \frac{\alpha}{2} \right) \right]$$
where $t_{n-1}(p)$ is the pth percentile of a t-distribution with $n - 1$ degrees of freedom.

3. $\hat{\theta}$ is not normal, but $n > 30$ or so: it turns out that all estimators we've seen (except for the maximum of the sample for the uniform distribution) will be asymptotically normal by the central limit theorem (it is not always straightforward how we apply the CLT in a given case). So we'll construct confidence intervals the same way as in case 2.

4. $\hat{\theta}$ not normal, n small: if the p.d.f. of $\hat{\theta}$ is known, can form confidence intervals from first principles (as in the last example). If the p.d.f. of $\hat{\theta}$ is not known, there is nothing we can do.

The reason for using the t-distribution in the second case is the following: since $\hat{\theta} \sim N\left(\mu, \frac{\sigma^2}{n} \right)$,
$$\frac{\hat{\theta} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1)$$

On the other hand, we can check that
$$\frac{(n-1)\hat{S}^2}{\sigma^2} \sim \chi_{n-1}^2$$

since in this setting, $\hat{S}$ can usually be written as a sum of squared normal residuals with mean zero and variance σ^2. Therefore,
$$\frac{\hat{\theta} - \mu}{\sqrt{\hat{S}^2 / n}} = \frac{\frac{\hat{\theta} - \mu}{\sigma / \sqrt{n}}}{\sqrt{\frac{(n-1)\hat{S}^2}{\sigma^2} / \sqrt{n-1}}} \sim \frac{N(0, 1)}{\sqrt{\chi_{n-1}^2}} \sim t_{n-1}$$

2

Also note that in the general case 4 (and in the last example involving a uniform), we did not require that the statistic $\hat{\theta}(X_1, \ldots, X_n)$ be an unbiased or consistent estimator of anything, but it just had to be strictly monotonic in the true parameter. However, the way we constructed confidence intervals for the normal cases (with or without knowledge of the variance of $\hat{\theta}$, the estimator has to be unbiased, and in case 3 (n large), it would have to be consistent.

2 Hypothesis Testing

2.1 Main Idea

Idea: given a random sample from a population, is there enough evidence to contradict some assertion about the population? Let's first define a number of important concepts:

- a *hypothesis* is an assumption about the distribution of a random variable in a population

- the *maintained hypothesis* is a hypothesis which cannot be tested, but which we will assume to be true no matter what.

- a *testable hypothesis* is a hypothesis which can and will be tested using evidence from a random sample

- the *null hypothesis* is the hypothesis to be tested

- the *alternative hypothesis* are other possible assumptions about the population other than the null

The testing problem can be stated as whether the parameter θ_0 corresponding to the density $f(x|\theta_0)$ which our sample $X_1, \ldots, X_n$ is drawn from belongs to a certain set of possible parameter values, Θ_0. We usually write the null hypothesis as

$$H_0 : \theta \in \Theta_0$$

which we test against the alternative

$$H_A : \theta \in \Theta_A$$

where $\Theta_0 \cap \Theta_A = \emptyset$.
If $\Theta_0 = \{\theta_0\}$ contains only a single parameter value, we say that the hypothesis is *simple*. A *composite* hypothesis is given by a set Θ containing multiple points, or an entire range of values.

Example 2 *In the most common setup, H_0 is simple, and H_A composite, e.g. $X \sim N(\mu, \sigma_0^2)$, where σ_0^2 is known, and we want to test whether $\mu = 0$. In this setting, the maintained hypothesis is that the X_is are i.i.d. normal and $\mathrm{Var}(X_i) = \sigma_0^2$. The null hypothesis is $H_0 : \mu = 0$ (simple), and we test against the alternative hypothesis $H_A : \mu \neq 0$ (composite).*

In order to test the hypothesis we have to gather data, and then either accept or reject the null hypothesis based on the data. However, since our data will always be only a sample from the entire population, there is always a possibility that we will make a mistake in our decision: The probability of

		True State of Nature	
		H_0 true	H_0 false
	reject H_0	Type I error	correct
Decision			
	not reject H_0	correct	Type II error

a type I error of a given test,

$$\alpha = P(\text{Type I error}) = P(\text{reject}|H_0)$$

is called the *significance level* (or also the *size*) of the test. If we write

$$\beta = P(\text{Type II error}) = P(\text{don't reject}|H_A)$$

then $1 - \beta$ is the *power* of the test.

Usually we'll fix the significance level of the test, e.g. at 5%, and then try to construct a test that has maximal power given that significance level. So in a sense we prefer to err on not rejecting the null hypothesis.

The logic behind this is a little counter-intuitive at first, but it comes from the empirical problem of generalizing from a few observations to an entire population or a scientific law: even though we may only have observed instances which conform with our hypothesis about the population, it is sufficient to observe one which doesn't in order to disprove it. Therefore we can use empirical evidence only to reject a hypothesis, but never to confirm it. The following is a famous example by the philosopher Bertand Russell:

> "Domestic animals expect food when they see the person who usually feeds them. We know that all these rather crude expectations of uniformity are liable to be misleading. The man who has fed the chicken every day throughout its life at last wrings its neck instead, showing that more refined views as to the uniformity of nature would have been useful to the chicken. [..] The mere fact that something has happened a certain number of times causes animals and men to expect that it will happen again. Thus our instincts certainly cause us to believe that the sun will rise to-morrow, but we may be in no better a position than the chicken which unexpectedly has its neck wrung." (Russell, The Problems of Philosophy)

Therefore, if we want to present evidence that e.g. a certain drug significantly improves a patient's condition, we define the null hypothesis as H_0: "the drug has no effect on the patient's condition." Rejecting this hypothesis means that we have strong evidence *for* an effect of the drug. I.e. we always choose the null hypothesis as the statement we actually want to disprove.

As another illustration, we could think of the criminal justice system: in a process, both parties present data (evidence) in order to produce a decision "guilty" or "not guilty," and the jury can again make two errors: falsely convicting an innocent person (Type I error), or not convicting a criminal (Type II error). Most modern legal systems base criminal trials on the presumption of innocence, i.e. the accused is assumed to be "innocent until proven guilty", or in other words, the burden of proof is on the prosecution which has to produce evidence to convince the judge/jury that the accused is in fact guilty.

Note that decisions taken according to hypothesis tests need not be optimal in the sense that we ignore our ex ante probabilities for the null vs. the alternative hypothesis being true, and do not take into account the respective costs of making type I or type II errors. For criminal justice, proponents of preemption often argue that in many contexts - e.g. terrorism - a type II error may be prohibitively costly, so that the legal system should allow for exceptions of the presumption of innocence in some cases.

In sum, we'd like to formulate a rule which maps each possible outcome of the sample $X_1, \ldots, X_n$ to a decision "reject" or "do not reject."