Economics 520, Fall 2007

Lecture Note 19: Confidence Intervals

When we looked at estimation so far, we have looked at point estimation. Here we will look at interval estimation. We are trying to find intervals that contain the true parameter with high probability. At the same time we would like the intervals to be narrow.

Definition 1 *Let $X \sim f_X(x;\theta^*)$. An interval estimate of a scalar parameter θ^* is a pair of functions $L(x)$ and $U(x)$, with $L(x) \leq U(x)$. The random interval $(L(X), U(X))$ is an interval estimator.*

Definition 2 *For an interval estimator $(L(X), U(X))$ the confidence level is*

$$\inf_{\theta \in \Theta} Pr(L(X) \leq \theta \leq U(X)|\theta),$$

where

$$Pr(L(X) \leq \theta \leq U(X)|\theta) = \int_{L(x)}^{U(x)} f_X(x;\theta)dx.$$

Let us consider two examples.

Example 1
Let $X \sim \mathcal{N}(\mu, 1)$, for $-\infty < \mu < \infty$. Consider the three interval estimators, $CI_1 = (0, 1)$, $CI_2 = (x, \infty)$, and $CI_3 = (x - 0.674, x + 0.674)$. In the first case the confidence level of the interval is zero: for values of μ outside $(0, 1)$ the probability of the interval including μ is zero. For the second interval the confidence level is .5, because for any value of μ,

$$Pr(X \leq \mu \leq \infty|\mu) = 0.5.$$

For the third interval, the confidence interval is also .5. To see this, note that

$$\begin{aligned} Pr(X - 0.674 \leq \mu \leq X + 0.674|\mu) &= Pr(-0.674 \leq \mu - X \leq 0.674|\mu) \\ &= Pr(-0.674 \leq Z \leq 0.674), \end{aligned}$$

where Z is standard normal, since $\mu - X$ is standard normal. This probability is equal to 0.5, so the confidence level does not depend on μ and is equal to 0.5. The third interval is much narrower than the second, so it is a better confidence interval. □

Example 2
Let $X_1, X_2, \ldots, X_N$ iid with common density $f_X(x;\lambda) = \lambda \exp(-x\lambda)$. The maximum likelihood estimator is $\hat{\lambda} = 1/\bar{x}$. In large samples, we have that

$$\hat{\lambda} \xrightarrow{p} \lambda,$$

and

$$\sqrt{N}(\hat{\lambda} - \lambda) \xrightarrow{d} \mathcal{N}(0, \lambda^2).$$

A confidence interval with asymptotic level 0.95 exploits this normal distribution:

$$CI = \left(\hat{\lambda} - 1.96\sqrt{\frac{\hat{\lambda}^2}{N}}, \hat{\lambda} + 1.96\sqrt{\frac{\hat{\lambda}^2}{N}} \right).$$

To see how this works, note that by Slutsky's theorem,

$$\frac{1}{\hat{\lambda}}\sqrt{N}(\hat{\lambda} - \lambda) \xrightarrow{d} N(0,1).$$

Therefore,

$$Pr(-1.96 \leq \frac{\sqrt{N}(\hat{\lambda} - \lambda)}{\hat{\lambda}} \leq 1.96) \rightarrow 0.95$$
$$\Rightarrow Pr\left(-1.96 \leq \frac{\sqrt{N}(\lambda - \hat{\lambda})}{\hat{\lambda}} \leq 1.96\right) \rightarrow 0.95$$
$$\Rightarrow Pr\left(-1.96\sqrt{\frac{\hat{\lambda}^2}{N}} \leq (\lambda - \hat{\lambda}) \leq 1.96\sqrt{\frac{\hat{\lambda}^2}{N}}\right) \rightarrow 0.95$$
$$\Rightarrow Pr\left(\hat{\lambda} - 1.96\sqrt{\frac{\hat{\lambda}^2}{N}} \leq \lambda \leq \hat{\lambda} + 1.96\sqrt{\frac{\hat{\lambda}^2}{N}}\right) \rightarrow 0.95.$$

□

In general a way of finding confidence intervals with good properties is to invert test statistics.

Result 1 *Let $X \sim f_X(x;\theta)$. Suppose $C(\theta_0)$ is a critical region for a test of the null hypothesis $H_0: \theta = \theta_0$ against the alternative $H_a: \theta \neq \theta_0$ with level α. Then the region*

$$CI(x) = \{\theta \in \Theta | x \notin C(\theta)\},$$

is a $1 - \alpha$ level confidence interval.

To see what this means in practice, let us return to the previous exponential example. Consider a Wald test for the hypothesis $\lambda = \lambda_0$ against the alternative $\lambda \neq \lambda_0$. A possible test statistic is

$$WALD = N \cdot (\hat{\lambda}_{ml} - \lambda_0)^2 \cdot \mathcal{I}(\hat{\lambda}_{ml}),$$

which under the null hypothesis has a chi-squared distribution with degrees of freedom equal to one. The critical region for a 0.05 level test is

$$C(\lambda) = \{x_1, \ldots, x_N | N \cdot (\hat{\lambda}_{ml} - \lambda_0)^2 \cdot \mathcal{I}(\hat{\lambda}_{ml}) > 3.84\}$$
$$= \{x_1, \ldots, x_N | N \cdot (\hat{\lambda}_{ml} - \lambda_0)^2 \cdot \mathcal{I}(\hat{\lambda}_{ml}^2) > 3.84\}.$$

Inverting this leads to the 95% confidence interval we derived before

$$CI = (\hat{\lambda}_{ml} - 1.96\sqrt{1/\hat{\lambda}_{ml}^2}, \hat{\lambda}_{ml} + 1.96\sqrt{1/\hat{\lambda}_{ml}^2}).$$

Example 3
Suppose $X_1, \ldots, X_N$ are IID $\mathcal{N}(\mu, \sigma^2)$. Note that here we have two unknown parameters, μ and σ^2. Suppose we have 100 observations, with $\bar{x} = 1$ and $\bar{x^2} = 6$. We will find the maximum likelihood estimator for σ^2, obtain a 95% confidence interval, and test the hypothesis that $\sigma^2 = 4$ against the alternative that it differs from 4 at the 0.10 level.

(a) ML estimates
The log likelihood function is

$$L(\mu, \sigma) = \sum_{i=1}^{N} -\frac{1}{2}\ln\sigma^2 - \frac{x_i^2 - 2\mu x_u + \mu^2}{2\sigma^2}.$$

The first derivative with respect to μ is

$$\frac{\partial L}{\partial \mu}(\mu, \sigma^2) = \sum_{i=1}^{N} \frac{x_i - \mu}{\sigma^2},$$

and the first derivative with respect to σ^2 is

$$\frac{\partial L}{\partial \sigma^2}(\mu, \sigma^2) = \sum_{i=1}^{N} -\frac{1}{2\sigma^2} + \frac{x_i^2 - 2\mu x_i + \mu^2}{2\sigma^4}.$$

Solving for the maximum likelihood estimators by setting both first derivatives equal to zero leads to

$$\hat{\mu}_{ml} = \bar{x} = 1,$$

and

$$\hat{\sigma}_{ml}^2 = \bar{x^2} - (\bar{x})^2 = 5.$$

(b) information matrix estimates
The matrix of second derivatives of the log of the density function is

$$\mathcal{I} = -E\left[\begin{array}{cc} \frac{\partial^2 \ln f_X}{\partial\mu\partial\mu}(X;\mu,\sigma^2) & \frac{\partial^2 \ln f_X}{\partial\mu\partial\sigma^2}(X;\mu,\sigma^2) \\ \frac{\partial^2 \ln f_X}{\partial\sigma^2\partial\mu}(X;\mu,\sigma^2) & \frac{\partial^2 \ln f_X}{\partial\sigma^2\partial\sigma^2}(X;\mu,\sigma^2) \end{array}\right]$$

$$= -E\left[\begin{array}{cc} -\frac{1}{\sigma^2} & -\frac{x-\mu}{\sigma^4} \\ -\frac{x-\mu}{\sigma^4} & \frac{1}{2\sigma^4} - \frac{(x-\mu)^2}{\sigma^6} \end{array}\right].$$

Taking the expectation leads to

$$\mathcal{I} = \left[\begin{array}{cc} \mathcal{I}_{\mu\mu} & \mathcal{I}_{\mu\sigma^2} \\ \mathcal{I}_{\sigma^2\mu} & \mathcal{I}_{\sigma^2\sigma^2} \end{array}\right] = \left[\begin{array}{cc} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{array}\right].$$

At the restricted estimates $\hat{\mu}_r = 1$ and $\sigma^2 = 4$ this is equal to

$$\hat{\mathcal{I}}_r = \begin{bmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{32} \end{bmatrix}.$$

At the unrestricted estimates $\hat{\mu}_r = 1$ and $\hat{\sigma}^2 = 5$ this is equal to

$$\hat{\mathcal{I}}_r = \begin{bmatrix} \frac{1}{5} & 0 \\ 0 & \frac{1}{50} \end{bmatrix}.$$

The normalized asymptotic variance is

$$V = \mathcal{I}^{-1} = \begin{bmatrix} V_{\mu\mu} & V_{\mu\sigma^2} \\ V_{\sigma^2\mu} & V_{\sigma^2\sigma^2} \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{bmatrix}.$$

(c) 95% confidence interval

A 95% confidence interval for σ^2 is based on the normal approximation to the sampling distribution. The variance for the 2×1 vector $\sqrt{N}(\hat{\mu} - \mu, \hat{\sigma}^2 - \sigma^2)$ is V, so the variance for $\sqrt{N}(\hat{\sigma}^2 - \sigma^2)$ is $V_{\sigma^2\sigma^2} = 2\sigma^4$, estimated at $\sigma^2 = 5$ to be equal to 50.

$$CI = (\hat{\sigma}^2_{ml} - 1.96 \cdot \sqrt{50/N}, \hat{\sigma}^2_{ml} + 1.96 \cdot \sqrt{50/N}) = (3.61, 6.39).$$

(d) Wald test

The Wald test statistic is

$$WALD = N \cdot (\hat{\sigma}^2 - \sigma^2)/V_{\sigma^2\sigma^2} = 2,$$

so at the 10% level we do not reject the null hypothesis that $\sigma^2 = 4$. (the cutoff point for the chi-squared one distribution is 2.71.)

(e) Lagrange multiplier test

The Lagrange multiplier test is

$$LM = \frac{1}{N} \frac{\partial L}{\partial(\mu, \sigma^2)'}(\hat{\mu}_r, \sigma_r^2)' \cdot \mathcal{I}^{-1} \cdot \frac{\partial L}{\partial(\mu, \sigma^2)'}(\hat{\mu}_r, \sigma_r^2)$$

$$= \frac{1}{N} \begin{pmatrix} \sum_{i=1}^N \frac{x_i - \hat{\mu}_r}{\sigma^2} \\ \sum_{i=1}^N -\frac{1}{2\sigma_r^2} + \frac{(x_i - \hat{\mu}_r)^2}{2\sigma_r^4} \end{pmatrix}' \cdot \begin{bmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{bmatrix} \cdot \begin{pmatrix} \sum_{i=1}^N \frac{x_i - \hat{\mu}_r}{\sigma^2} \\ \sum_{i=1}^N -\frac{1}{2\sigma_r^2} + \frac{(x_i - \hat{\mu}_r)^2}{2\sigma_r^4} \end{pmatrix}$$

$$= \frac{1}{N} \begin{pmatrix} 0 \\ -\frac{N}{8} + \frac{N \cdot 5}{32} \end{pmatrix}' \cdot \begin{bmatrix} 4 & 0 \\ 0 & 32 \end{bmatrix} \cdot \begin{pmatrix} 0 \\ -\frac{N}{8} + \frac{N \cdot 5}{32} \end{pmatrix} = 3.125.$$

So, here we reject the null hypothesis, though only barely.

(f) likelihood ratio test

The log likelihood function is

$$L(\mu, \sigma) = \sum_{i=1}^N -\frac{1}{2} \ln \sigma^2 - \frac{x_i^2 - 2\mu x_u + \mu^2}{2\sigma^2}.$$

At $\hat{\mu}_r = 1$ and $\sigma_r^2 = 4$, this is equal to -131.81. At $\hat{\mu}_u = 1$ and $\hat{\sigma}_u = 5$ this is equal to -130.47. Hence

$$LR = 2 \cdot (L(\hat{\mu}_u, \hat{\sigma}_u^2) - L(\hat{\mu}_r, \hat{\sigma}_r)) = 2.69,$$

and so we (barely) accept the null hypothesis.