36-705: Intermediate Statistics **Fall 2019**

Lecture 25: October 30

Lecturer: Siva Balakrishnan

Today we will discuss confidence sets and ways to construct them. We have discussed point estimation so far where the goal is to construct an estimate $\widehat{\theta}(X_1, \ldots, X_n)$ of some parameter θ after observing $\{X_1, \ldots, X_n\}$.

The setting here is that we have a statistical model (i.e. a collection of distributions) $\mathcal{P}$. Let $C_n(X_1, \ldots, X_n)$ be a set constructed using the observed data $X_1, \ldots, X_n$. This is a random set. C_n is a $1 - \alpha$ confidence set for a parameter θ if:

$$P(\theta \in C_n(X_1, \ldots, X_n)) \geq 1 - \alpha, \quad \text{for all} \ \ P \in \mathcal{P}.$$

This means that no matter which distribution in $\mathcal{P}$ generated the data, the interval guarantees the coverage property described above. Some people would refer to such intervals as *honest* confidence intervals to make explicit the fact that the coverage is uniform over the model.

At a high-level, the confidence interval gives us some idea of how precise our estimate of the unknown parameter is, i.e. a wide interval indicates that our (point) estimate is imprecise.

Example: Suppose that we considered, $X_1, \ldots, X_n \sim U[0, \theta]$. Then we could construct the usual point estimate $\widehat{\theta} = X_{(n)}$. We could perhaps consider two types of confidence intervals:

$$\begin{aligned} C_1 &= [a_1\widehat{\theta}, b_1\widehat{\theta}], \quad 1 \leq a_1 \leq b_1 \\ C_2 &= [\widehat{\theta} + a_2, \widehat{\theta} + b_2], \quad a_2, b_2 \geq 0. \end{aligned}$$

Let us try to calculate the coverage probabilities of these two types of intervals. As a preliminary observe that:

$$\mathbb{P}(\widehat{\theta} \leq t) = \left(\frac{t}{\theta}\right)^n, \quad \text{for} \ \ 0 \leq t \leq \theta.$$

1. C_1: We can compute that,

$$\begin{aligned} \mathbb{P}(\theta \in C_1) &= \mathbb{P}(\widehat{\theta} \leq \theta/a_1, \widehat{\theta} \geq \theta/b_1) \\ &= \mathbb{P}(\widehat{\theta} \leq \theta/a_1) - \mathbb{P}(\widehat{\theta} \leq \theta/b_1) \\ &= \left(\frac{1}{a_1}\right)^n - \left(\frac{1}{b_1}\right)^n. \end{aligned}$$

   So we have that for instance choosing $a_1 = 1$, $b_1 = \left(\frac{1}{\alpha}\right)^{1/n}$ guarantees us that this confidence interval has coverage probability exactly $1 - \alpha$.

2. C_2: Similarly we have that,

$$\begin{aligned}\mathbb{P}(\theta \in C_2) &= \mathbb{P}(\widehat{\theta} \leq \theta - a_2, \widehat{\theta} \geq \theta - b_2) \\ &= \mathbb{P}(\widehat{\theta} \leq \theta - a_2) - \mathbb{P}(\widehat{\theta} \leq \theta - b_2) \\ &= \left(\frac{\theta - a_2}{\theta}\right)^n - \left(\frac{\theta - b_2}{\theta}\right)^n.\end{aligned}$$

   Notice that now the coverage probability depends on the unknown parameter θ (which is undesirable). Furthermore, if we choose any constants (a_2, b_2) (say depending only on the desired coverage probability α), then as $\theta \to \infty$ we have that the interval has coverage probability that tends to 0, i.e. the interval is not honest for any constants (a_2, b_2).

We will now discuss a few different ways of constructing confidence intervals. Although superficially different most of these methods are roughly the same.

25.1 Inverting a test

We discussed this method in the last lecture. We suppose that we have a (family of) test(s) for the hypotheses:

$$\begin{aligned}H_0 &: \theta = \theta_0 \\ H_1 &: \theta \neq \theta_0.\end{aligned}$$

These tests have a rejection region and a corresponding acceptance region (where we fail to reject the null). Denote the acceptance region for the test of $\theta = \theta_0$ as $A(\theta_0)$. This is a subset of the sample space.

Given observed data $\{X_1, \ldots, X_n\}$ we consider the random set:

$$C(X_1, \ldots, X_n) = \{\theta_0 : \{X_1, \ldots, X_n\} \in A(\theta_0)\}.$$

Our confidence set is simply the set of parameters θ_0 that we would fail to reject using our family of tests. If our family of tests has level α then the set $C(X_1, \ldots, X_n)$ is a $1 - \alpha$ confidence set.

To see this observe that since our test controls the Type I error we have that for any parameter θ_0,

$$\mathbb{P}_{\theta_0}(\{X_1, \ldots, X_n\} \notin A(\theta_0)) \leq \alpha,$$

so with probability at least $1 - \alpha$ we have that, $\{X_1, \ldots, X_n\} \in A(\theta_0)$ and hence that $\theta_0 \in C(X_1, \ldots, X_n)$.

We can also construct tests using confidence intervals, i.e. consider the test that rejects the null hypothesis $\theta = \theta_0$ if $\theta_0 \notin C(X_1, \ldots, X_n)$, then if $C(X_1, \ldots, X_n)$ is a $1 - \alpha$ confidence interval this test has level α, i.e.

$$\mathbb{P}_{\theta_0}(\text{reject null } \theta = \theta_0) = \mathbb{P}_{\theta_0}(\theta_0 \notin C(X_1, \ldots, X_n)) \leq \alpha.$$

Let us quickly re-visit the uniform example. Suppose we observe $X_1, \ldots, X_n \sim U[0, \theta]$ and would like to construct a confidence interval, then one method would be to invert the LRT, i.e. we compute the likelihood-ratio for testing $H_0 : \theta = \theta_0$ as if $\theta \geq \max_i X_i$ then:

$$\text{LR} = \frac{\frac{1}{\theta_0^n}}{\frac{1}{(\max_i X_i)^n}} = \frac{(\max_i X_i)^n}{\theta_0^n},$$

and we note that we reject the null for small values of this quantity, i.e. we reject the null if

$$\frac{(\max_i X_i)^n}{\theta_0^n} \leq k_\alpha,$$

for an appropriate choice of k_α. So if we consider the confidence interval obtained by inverting this test, we see that it has the form:

$$C(X_1, \ldots, X_n) = \left\{ \theta : \max_i X_i \leq \theta \leq \frac{\max_i X_i}{k_\alpha^{1/n}} \right\},$$

which is precisely the type of multiplicative interval that we studied in the last lecture (we also calculated a value for k_α that ensures that $C(X_1, \ldots, X_n)$ has coverage $1 - \alpha$ in that lecture). This just highlights that in this case, we could have obtained the right kind of interval in a less ad hoc manner (by inverting the LRT).

Example: Suppose we observe $X_1, \ldots, X_n \sim \text{Exp}(\lambda)$, and want to construct a confidence interval for λ.

As our test, suppose we use the LRT, i.e. we define the likelihood ratio:

$$\Lambda = \frac{\lambda_0^n \exp(-\lambda_0 \sum_i X_i)}{(\frac{1}{\overline{X}})^n \exp(-n)}.$$

The acceptance region has the form:

$$A(\lambda_0) = \left\{ \{X_1, \ldots, X_n\} : \left(\lambda_0 \sum X_i\right)^n \exp(-\lambda_0 \sum X_i) \geq k_\alpha(\lambda_0) \right\},$$

where $k_\alpha(\lambda_0)$ needs to be chosen appropriately to control the Type I error. Observe that since $X_i \sim \text{Exp}(\lambda_0)$, $\lambda_0 X_i \sim \text{Exp}(1)$, so $k_\alpha(\lambda_0)$ does not depend on λ_0. Once we determine the cut-off we would obtain the confidence interval by collecting:

$$C(X_1, \ldots, X_n) = \{\lambda : \left(\lambda \sum X_i\right)^n \exp(-\lambda \sum X_i) \geq k_\alpha\},$$

which is an expression that can be solved numerically. Determining k_α and then finding the confidence set can be quite tedious to do exactly (see the Casella and Berger book) and an alternative would be to use large-sample (asymptotic approximations).

25.2 Inverting Probability Inequalities

In some simple cases, we can use tail bounds to derive confidence intervals. These typically have the advantage of being exact, finite-sample intervals. However, they are rarely used in practice for many reasons including: (1) we do not always have tail bounds for estimators of interest (2) there are usually imprecisely known constants in tails bounds (3) related to (2) they are often very conservative (i.e. the intervals are often too wide to be useful).

Here are a couple of examples:

Example 25.1 *Let* $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$. *By Hoeffding's inequality:*

$$\mathbb{P}(|\widehat{p} - p| > \epsilon) \leq 2e^{-2n\epsilon^2}.$$

Let

$$\epsilon_n = \sqrt{\frac{1}{2n} \log\left(\frac{2}{\alpha}\right)}.$$

Then

$$\mathbb{P}\left(|\widehat{p} - p| > \sqrt{\frac{1}{2n} \log\left(\frac{2}{\alpha}\right)}\right) \leq \alpha.$$

Hence, $\mathbb{P}(p \in C) \geq 1 - \alpha$ *where* $C = (\widehat{p} - \epsilon_n, \widehat{p} + \epsilon_n)$.

Example 25.2 *Let* $X_1, \ldots, X_n \sim F$. *Suppose we want a* **confidence band** *for* F. *We can use VC theory. Remember that*

$$\mathbb{P}\left(\sup_x |F_n(x) - F(x)| > \epsilon\right) \leq 2e^{-2n\epsilon^2}.$$

Let

$$\epsilon_n = \sqrt{\frac{1}{2n} \log\left(\frac{2}{\alpha}\right)}.$$

Then

$$\mathbb{P}\left(\sup_x |F_n(x) - F(x)| > \sqrt{\frac{1}{2n} \log\left(\frac{2}{\alpha}\right)}\right) \leq \alpha.$$

Hence,

$$P_F(L(t) \leq F(t) \leq U(t) \ \text{ for all } t) \geq 1 - \alpha$$

for all F, *where*

$$L(t) = \widehat{F}_n(t) - \epsilon_n, \quad U(t) = \widehat{F}_n(t) + \epsilon_n.$$

We can improve this by taking

$$L(t) = \max\left\{\widehat{F}_n(t) - \epsilon_n,\ 0\right\}, \quad U(t) = \min\left\{\widehat{F}_n(t) + \epsilon_n,\ 1\right\}.$$

25.2.1 Pivots

Another useful way of attempting to construct confidence intervals is to base the intervals on *pivots*. A pivot is a function of the data and the unknown parameter θ – $Q(X_1, \ldots, X_n, \theta)$ – whose distribution does not depend on θ.

Let us consider two examples:

1. Suppose that $X_1, \ldots, X_n \sim N(\theta, 1)$ then we can see that $Q(X_1, \ldots, X_n) = \overline{X_n} - \theta \sim N(0, 1/n)$ and so the distribution of Q does not depend on θ.

2. Suppose we consider $X_1, \ldots, X_n \sim U[0, \theta]$ and we consider the function:

$$Q(X_1, \ldots, X_n, \theta) = \frac{\max_i X_i}{\theta},$$

has distribution:

$$P(Q(X_1, \ldots, X_n, \theta) \leq t) = \begin{cases} t^n & 0 \leq t \leq 1 \\ 1 & t \geq 1. \end{cases}$$

Once again the distribution does not depend on θ.

Given a pivot we can construct confidence intervals in a simple way. Since the distribution of Q does not depend on θ, we can find a, b which do not depend on θ such that:

$$\mathbb{P}_\theta(a \leq Q(X_1, \ldots, X_n, \theta) \leq b) = 1 - \alpha, \quad \text{for all } \theta \in \Theta.$$

Now, we construct our confidence interval as:

$$C(X_1, \ldots, X_n) = \{\theta : a \leq Q(X_1, \ldots, X_n, \theta) \leq b\}.$$

By our construction:

$$\mathbb{P}_\theta(\theta \in C(X_1, \ldots, X_n)) = \mathbb{P}_\theta(a \leq Q(X_1, \ldots, X_n, \theta) \leq b) = 1 - \alpha.$$

Going back to our two examples we find that we will once again obtain the now standard intervals for the two problems (the additive interval for the Gaussian mean, and the multiplicative scale interval for the uniform parameter).

25.3 Tests Versus Confidence Intervals

Confidence intervals are more informative than tests. Intuitively, p-values are more informative than an accept/reject decision because it summarizes all the significance levels for which

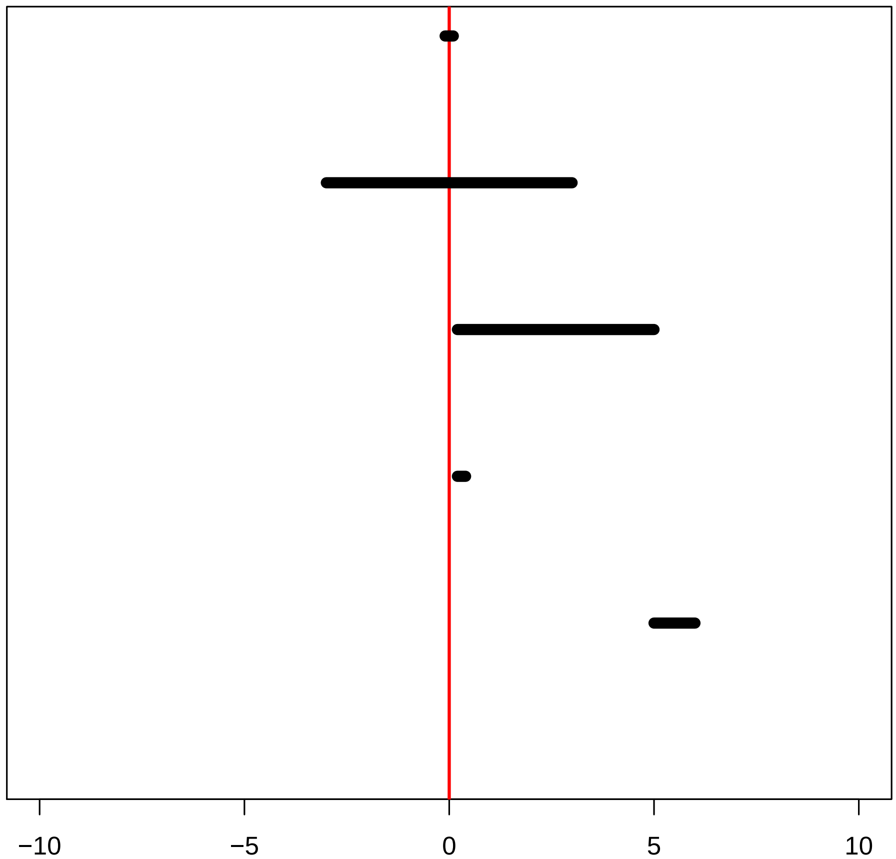

Figure 25.1: Five examples: 1. Not significant, precise. 2. Not significant, imprecise. 3. Barely significant, imprecise. 4. Barely significant, precise. 5. Significant and precise.

we would reject the null hypothesis. Similarly, a confidence interval is more informative that a test because it summarizes all the parameters for which we would (fail to) reject the null hypothesis. More practically, a confidence interval tells us something about the "effect size" as well as something about the uncertainty in our estimate of the "effect size".

Look at Figure 25.1. Suppose we are testing $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$. We see 5 different confidence intervals. The first two cases (top two) correspond to not rejecting H_0. The other three correspond to rejecting H_0. Reporting the confidence intervals is much more informative than simply reporting "reject" or "don't reject."