Lecture 11

Confidence Sets

1 Introduction

So far, we have been considering point estimation. In this lecture, we will study interval estimation. Let X denote our data. Let $\theta \in \mathbb{R}$ be our parameter of interest. Our task is to construct a data-dependent interval $[l(X), r(X)]$ so that it contains θ with large probability. One possibility is to set $l(X) = -\infty$ and $r(X) = +\infty$. Such an interval will contain θ with probability 1. Of course, the problem with this interval is that it is too long. So, we want to construct an interval that will be shorter. More generally, instead of intervals, we can consider confidence set $C(X) \subset \mathbb{R}$ such that it contains θ with large probability. The concept of confidence sets can be also applied to any set of possible parameter values Θ, not just for $\mathbb{R}$.

Let us introduce the basic concepts related to confidence sets.

Definition 1. Coverage probability of the set $C(X) \subset \Theta$ is the probability (under the assumption that the true value is θ) that confidence set $C(X)$ contains θ, i.e. Coverage Probability$(\theta) = P_\theta\{\theta \in C(X)\}$.

Of course, in practice, we are interested in confidence sets that contain the true parameter value with large probability uniformly over the set of possible parameters values.

Definition 2. Confidence level is the minimum of coverage probabilities over the set of possible parameter values, i.e. Confidence Level $= \inf_{\theta \in \Theta} P_\theta\{\theta \in C(X)\}$. We say that confidence set $C(X)$ has confidence level α if $\inf_{\theta \in \Theta} P_\theta\{\theta \in C(X)\} \geq \alpha$.

Let us consider how we can construct confidence sets.

2 Test Inversion

For each possible parameter value $\theta_0 \in \Theta$, consider the problem of testing the null hypothesis, $H_0 : \theta = \theta_0$ against the alternative, $H_a : \theta \neq \theta_0$. Suppose that for each such hypothesis we have a test of size α. Then the confidence set $C(X) = \{\theta_0 \in \Theta : \text{the null hypothesis that } \theta = \theta_0 \text{ is not rejected}\}$ is of confidence level $1 - \alpha$. Indeed, suppose that the true value of the parameter is θ_0. Since the test of $\theta = \theta_0$ against $\theta \neq \theta_0$ has level α by construction, $P_{\theta_0}\{\text{the test rejects } \theta = \theta_0\} \leq \alpha$. So, with probability of at least $1 - \alpha$, $\theta_0 \in C(X)$. In other words, $P_{\theta_0}\{\theta_0 \in C(X)\} \geq 1 - \alpha$. The same holds for all $\theta_0 \in \Theta$. So, $\inf_{\theta \in \Theta} P_\theta\{\theta \in C(X)\} \geq 1 - \alpha$. This procedure is known as test inversion.

We may find that the confidence set construction is a dual problem for testing. Note that if we have a way to construct a confidence set, we can construct a test for any hypothesis $H_0 : \theta = \theta_0$. Indeed, once we have a confidence set $C(X)$ of level $1 - \alpha$, we can form a test of the null hypothesis, H_0, that $\theta = \theta_0$ against the alternative, H_a, that $\theta \neq \theta_0$ by accepting the null hypothesis if and only if $\theta_0 \in C(X)$. This test will be of size α.

Example 1 Let $X_1, ..., X_n$ be a random sample from a distribution with two finite moments. Let us use test inversion to construct a confidence set for $\mu = EX_i$ of (asymptotic) level $1 - \alpha$. Let us consider the problem of testing the null hypothesis, $H_0 : \mu = \mu_0$ against the alternative, $H_a : \mu \neq \mu_0$. Under the null hypothesis we have the following asymptotic statement: $t(\mu_0) = \sqrt{n}\frac{\overline{X} - \mu_0}{s} \Rightarrow N(0, 1)$. One possible test of size α is to accept the null hypothesis if and only if $z_{\alpha/2} \leq \frac{\sqrt{n}(\overline{X} - \mu_0)}{s} \leq z_{1-\alpha/2}$, where z_α is the quantile of the standard normal distribution. This test will accept the null hypothesis $\mu = \mu_0$ if and only if $\overline{X} - z_{1-\alpha/2}\frac{s}{\sqrt{n}} \leq \mu_0 \leq \overline{X} - z_{\alpha/2}\frac{s}{\sqrt{n}}$. So, the confidence set is $\left[\overline{X} - z_{1-\alpha/2}\frac{s}{\sqrt{n}}, \overline{X} - z_{\alpha/2}\frac{s}{\sqrt{n}}\right]$. Note that we actually end up with an interval in this example.

Example 2 Let $X_1, ..., X_n$ be a random sample from the distribution $N(\mu, \sigma^2)$. Let us use test inversion to construct a confidence set for σ^2 of level $1 - \alpha$. Consider the problem of testing the null hypothesis, $H_0 : \sigma^2 = \sigma_0^2$ against the alternative, $H_a : \sigma^2 \neq \sigma_0^2$. Under the null hypothesis, $(n-1)s^2/\sigma_0^2 \sim \chi^2(n-1)$. One possible test of size α is to accept the null hypothesis if and only if

$$\chi^2_{\alpha/2}(n-1) \leq (n-1)s^2/\sigma_0^2 \leq \chi^2_{1-\alpha/2}(n-1).$$

This test will accept the null hypothesis $\sigma^2 = \sigma_0^2$ if and only if

$$(n-1)s^2/\chi^2_{1-\alpha/2}(n-1) \leq \sigma_0^2 \leq (n-1)s^2/\chi^2_{\alpha/2}(n-1).$$

So, the confidence set is

$$\left[\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}(n-1)}, \frac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)}\right].$$

This is not the shortest interval, apparently we may cut off unequal tails, but tails whose probability would sum up to α: say

$$\left[\frac{(n-1)s^2}{\chi^2_{1-\alpha+\delta}(n-1)}, \frac{(n-1)s^2}{\chi^2_{\delta}(n-1)}\right]$$

for $0 < \delta < \alpha$ is also a valid confidence set, we may try to optimize over δ to find the shortest interval.

In general, if we can find a (asymptotically) pivotal quantity $Q = q(X_1, ..., X_n, \theta_0)$ such that distribution of Q under the null hypothesis $\theta = \theta_0$ does not depend on the choice of θ_0 (in finite samples or asymptotically), then we can use Q for testing and confidence set construction. Indeed, since distribution of Q is independent of the true parameter value, we can find numbers a and b such that $P_{\theta_0}\{a \leq Q \leq b\} = 1 - \alpha$ for all $\theta_0 \in \Theta$. Then one possible test is to accept the null hypothesis that $\theta = \theta_0$ if and only if $a \leq q(X_1, ..., X_n, \theta_0) \leq b$.

The confidence set will consists of all parameter values θ_0 which are accepted:

$$C(X) = \{\theta_0 : a \leq q(X_1, ..., X_n, \theta_0) \leq b\}.$$

If $q(X, \theta_0)$ considered as a function of θ_0 is continuous monotonic and there exists a unique inverse, then we would have $C(X)$ as an interval with end points $q^{-1}(X_1, ..., X_n, a)$ and $q^{-1}(X_1, ..., X_n, b)$. Though in many interesting cases inversion of a test will not lead to an interval.

3   Pratt's Theorem

Informally, the theorem states that if we use a uniformly-most-powerful test (UMP) for the confidence set construction, the expected length of the confidence set will be the shortest among all confidence sets of a given level.

Theorem 3. *Let $X \sim f(x|\theta)$ be our data. Let $C(X)$ be our confidence set for θ. Then, under some regularity conditions, for any θ_0,*

$$E_{\theta_0}[\text{length of } C(X)] = \int P_{\theta_0}\{\theta \in C(X)\}d\theta.$$

Moreover, if $C(X)$ is constructed by inverting a UMP test of size α, then $C(X)$ has the shortest expected length among all confidence sets of level $1 - \alpha$ for any θ_0.

Proof. The first result follows from

$$
\begin{aligned}
E_{\theta_0}[\text{length of } C(X)] &= E_{\theta_0}\left[\int_\theta I\{\theta \in C(X)\}d\theta\right] \\
&= \int_x \int_\theta I\{\theta \in C(X)\}d\theta f(x|\theta_0)dx \\
&= \int_\theta \int_x I\{\theta \in C(X)\}f(x|\theta_0)dx d\theta \\
&= \int P_{\theta_0}\{\theta \in C(X)\}d\theta.
\end{aligned}
$$

Note that $\int P_{\theta_0}\{\theta \in C(X)\}d\theta = \int_{\theta \neq \theta_0} P_{\theta_0}\{\theta \in C(X)\}d\theta$ and, for any $\theta \neq \theta_0$, $P_{\theta_0}\{\theta \in C(X)\}$ equals 1 minus the power of the test based on confidence set $C(X)$. So, if $\tilde{C}(X)$ denotes the confidence set constructed by inverting a UMP test,

$$P_{\theta_0}\{\theta \in C(X)\} \geq P_{\theta_0}\{\theta \in \tilde{C}(X)\}$$

and

$$\int P_{\theta_0}\{\theta \in C(X)\}d\theta \geq \int P_{\theta_0}\{\theta \in \tilde{C}(X)\}d\theta.$$

Combining this inequality with the first result yields the second result of Pratt's theorem. □

Example 3   Let $X_1, ..., X_n$ be a random sample from distribution $N(\mu, \sigma^2)$. We have already seen that the UMP test of the null hypothesis, H_0, that $\mu = \mu_0$ against the alternative, H_a, that $\mu \neq \mu_0$ accepts the

null hypothesis if and only if $|(\overline{X}_n - \mu)/\sqrt{s^2/n}| \leq t_{1-\alpha/2}(n-1)$. So, the confidence interval with shortest expected length is

$$\left[\overline{X}_n - \frac{s}{\sqrt{n}}t_{1-\alpha/2}, \overline{X}_n + \frac{s}{\sqrt{n}}t_{1-\alpha/2}\right].$$

4 Asymptotic Theory for Interval Construction

Let $X_1, ..., X_n$ be a random sample from distribution $f(x|\theta)$ with $\theta \in \Theta$. Under some regularity conditions,

$$\sqrt{n}(\hat{\theta}_{ML} - \theta) \Rightarrow N(0, I^{-1}(\theta)).$$

For any function $h : \Theta \to \mathbb{R}$, under some regularity conditions, by the delta-method,

$$\sqrt{n}(h(\hat{\theta}_{ML}) - h(\theta)) \Rightarrow N(0, (h'(\theta))^2 I^{-1}(\theta)).$$

We can consistently estimate $(h'(\theta))^2 I^{-1}(\theta)$ by $n(h'(\hat{\theta}_{ML}))^2(-\partial^2 l_n(\hat{\theta}_{ML})/\partial\theta^2)^{-1}$. Denote

$$\hat{V}(h(\hat{\theta}_{ML})) = (h'(\hat{\theta}_{ML})^2(-\partial^2 l_n(\hat{\theta}_{ML})/\partial\theta^2)^{-1}$$

By the Slutsky theorem,

$$\frac{h(\hat{\theta}_{ML}) - h(\theta)}{\sqrt{\hat{V}(h(\hat{\theta}_{ML}))}} \Rightarrow N(0, 1).$$

So, we can construct a confidence interval for $h(\theta)$ as

$$\left[h(\hat{\theta}_{ML}) + z_{\alpha/2}\sqrt{\hat{V}(h(\hat{\theta}_{ML}))}, h(\hat{\theta}_{ML}) + z_{1-\alpha/2}\sqrt{\hat{V}(h(\hat{\theta}_{ML}))}\right].$$

Note that this confidence set is essentially constructed based on the Wald statistic.

Example 4 Let $X_1, ..., X_n$ be a random sample from distribution Bernoulli(p). Suppose we want to construct a confidence set for $h(p) = p/(1-p)$. Denote $\hat{p} = \overline{X}_n$. Then

$$\sqrt{n}(\hat{p} - p) \Rightarrow N(0, p(1-p)).$$

In addition,

$$h'(p) = \frac{(1-p)+p}{(1-p)^2} = \frac{1}{(1-p)^2}.$$

By delta-method,

$$\sqrt{n}(h(\hat{p}) - h(p)) \Rightarrow N(0, p/(1-p)^3).$$

So, $\hat{V}(h(\hat{p})) = \hat{p}/((1-\hat{p})^3 n)$. Thus, a confidence interval for $p/(1-p)$ is

$$\left[\frac{\hat{p}}{1-\hat{p}} + z_{\alpha/2}\sqrt{\frac{\hat{p}}{(1-\hat{p})^3 n}}, \frac{\hat{p}}{1-\hat{p}} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}}{(1-\hat{p})^3 n}}\right].$$

4.1 Confidence Sets Based on LM and LR Tests

In addition to the Wald statistic, we can invert tests based on the LM and LR statistics as well. However, these confidence sets are usually more involved, as the inversion procedure is less straightforward.

Let $X_1, ..., X_n$ be a random sample from the distribution Bernoulli(p). Then the joint log-likelihood is

$$l_n = \log\left(p^{\sum X_i}(1-p)^{n-\sum X_i}\right) = \sum X_i \log p + (n - \sum X_i)\log(1-p).$$

So,

$$\frac{\partial l_n}{\partial p} = \frac{\sum X_i}{p} - \frac{n - \sum X_i}{1-p},$$

and

$$I(p) = \frac{1}{p(1-p)}.$$

Thus,

$$
\begin{aligned}
LM &= \left(\frac{\sum X_i/p - (n - \sum X_i)/(1-p)}{\sqrt{n/(p(1-p))}}\right)^2 \\
&= \left(\frac{(1-p)\sum X_i - (n - \sum X_i)p}{\sqrt{np(1-p)}}\right)^2 \\
&= \left(\frac{\sum X_i - np}{\sqrt{np(1-p)}}\right)^2.
\end{aligned}
$$

We know that $LM \Rightarrow \chi_1^2$. So, the confidence set based on inverting the LM test is

$$\left\{p \in (0,1): \left|\frac{\sum X_i - np}{\sqrt{np(1-p)}}\right| \leq z_{1-\alpha/2}\right\},$$

which is the solution to a quadratic inequality.

As for the LR test,

$$l_n^{ur} - l_n^r = \sum X_i \log(\hat{p}/p_0) + (n - \sum X_i)\log((1-\hat{p})/((1-p_0)).$$

So, the confidence set based on inverting the LR test is

$$\left\{p \in (0,1): 2\left(\sum X_i \log(\hat{p}/p) + (n - \sum X_i)\log((1-\hat{p})/((1-p)))\right) \leq \chi_{1-\alpha}^2(1)\right\}.$$

It is the solution to a nonlinear inequality.

5 Bootstrap confidence sets

Assume one wants to create a confidence set for a parameter θ for which s/he has a consistent estimator $\hat{\theta} = \delta(X)$ where X is a random draw from unknown distribution F. One way to construct a bootstrap confidence set is by bootstrapping a statistic $T(\theta_0) = \hat{\theta} - \theta_0$ whenever testing $H_0 : \theta = \theta_0$. In particular, we would draw a bootstrapped sample X^* from an approximating distribution $\hat{F}$ and calculate $T^* = \delta(X^*) - \hat{\theta}$. Then quantiles of T^* will serve as critical values for the corresponding test. Notice that inverting this test is exceptionally easy!

- For $b = 1, ..., B$ repeat the following:

  - Draw a random sample X_b^* from distribution $\hat{F}$;
  - Calculate $T_b^* = \delta(X_b^*) - \hat{\theta}$;

- Order the bootstrapped statistics from smallest to largest: $T_{(1)}^* \leq ... \leq T_{(B)}^*$.

- Test of $H_0 : \theta = \theta_0$ accepts if $T_{([\frac{\alpha}{2}B])}^* \leq \hat{\theta} - \theta_0 \leq T_{([(1-\frac{\alpha}{2})B])}^*$.

- Confidence set is $\hat{\theta} - T_{([(1-\frac{\alpha}{2})B])}^* \leq \theta_0 \leq \hat{\theta} - T_{([\frac{\alpha}{2}B])}^*$.

When does this work? When the difference between the distributions of T and T^* converges almost surely to zero as the sample size increases. Notice that this interval implicitly bias-corrects. Most often the application of this method happens when $\hat{\theta}$ is asymptotically Gaussian (and you choose not to calculate standard errors).

Though if one decided to calculate standard errors then s/he may bootstrap a t-statistic. One would then bootstrap statistics $Z^* = \frac{\delta(X_b^*) - \hat{\theta}}{s.e.^*}$ and find the proper quantiles of it. The resulting confidence set will be $\hat{\theta} - Z_{([(1-\frac{\alpha}{2})B])}^* s.e. \leq \theta_0 \leq \hat{\theta} - Z_{([\frac{\alpha}{2}B])}^* s.e.$

Grid bootstrap One may construct a confidence set by inverting other statistics as well, though the inversion is less obvious. This is called a grid bootstrap. One would impose a fine grid on the space of Θ and will test each value θ_0 on this grid. In particular, one would calculate the test statistic testing $H_0 : \theta = \theta_0$, say, $G(\theta_0, X)$, and find its bootstrapped critical values. That is the hypothesis is accepted for example iff $G(\theta_0, X) < G_{([(1-\alpha)B])}^*$ (if the test uses only one-side of the final distribution). Then on the grid we would decide which θ_0 are accepted, and the result does not have to be an interval.

6 Some notes on joint confidence sets and the projection method

Imagine you are trying to estimate two parameters, α and β, from the same data set, and you have estimates $\hat{\alpha}$ and $\hat{\beta}$ that are consistent and jointly gaussian, that is,

$$\sqrt{n} \begin{pmatrix} \hat{\alpha} - \alpha \\ \hat{\beta} - \beta \end{pmatrix} \Rightarrow N(0, \Sigma), \quad \Sigma = \begin{pmatrix} \sigma_\alpha^2 & \sigma_{\alpha\beta} \\ \sigma_{\alpha\beta} & \sigma_\beta^2 \end{pmatrix},$$

and you have a consistent estimator for Σ, say $\hat{\Sigma}$. You can easily construct confidence sets for α: $[\hat{\alpha} - 1.96\frac{\hat{\sigma}_\alpha}{\sqrt{n}}, \hat{\alpha} + 1.96\frac{\hat{\sigma}_\alpha}{\sqrt{n}}]$ and for β: $[\hat{\beta} - 1.96\frac{\hat{\sigma}_\beta}{\sqrt{n}}, \hat{\beta} + 1.96\frac{\hat{\sigma}_\beta}{\sqrt{n}}]$. However, these two confidence intervals are not

jointly valid, that is the probability that both of them cover the true values simultaneously is less than 95%. To get a joint confidence set one should invert the test for the joint hypothesis $H_0 : \alpha = \alpha_0, \beta = \beta_0$. For example (if we still decide to stick with the Wald statistic), we may accept the null iff:

$$W(\alpha_0, \beta_0) = n \left(\hat{\alpha} - \alpha_0, \hat{\beta} - \beta_0 \right) \hat{\Sigma}^{-1} \begin{pmatrix} \hat{\alpha} - \alpha_0 \\ \hat{\beta} - \beta_0 \end{pmatrix} \leq \chi^2_{2,1-\alpha}.$$

The set of (α_0, β_0) described by the inequality above is ellipse, call it A. Then to construct confidence sets for α and β, which would be jointly valid, we may project it onto two axes. That is,

$$C_\beta = \{\beta_0 : \exists \alpha \ \ s.t. \ (\alpha, \beta_0) \in A\} = \{\beta_0 : \min_\alpha Wald(\alpha, \beta_0) \leq \chi^2_{2,1-\alpha}\} = \left[\hat{\beta} - \sqrt{\chi^2_{2,1-\alpha}} \frac{\hat{\sigma}_\beta}{\sqrt{n}}, \hat{\beta} + \sqrt{\chi^2_{2,1-\alpha}} \frac{\hat{\sigma}_\beta}{\sqrt{n}} \right].$$

This confidence set is constructed using the projection method. It is conservative, in the sense that

$$P_{\alpha_0, \beta_0}\{\alpha_0 \in C_\alpha \text{ and } \beta_0 \in C_\beta\} \geq 1 - \alpha,$$

as the event under the probability sign is the rectangle containing the confidence ellipse A. Another idea, may be to construct a rectangle to start with. For that we may consider other statistics, like:

$$S(\alpha_0, \beta_0) = \max \left\{ \frac{\hat{\alpha} - \alpha_0}{\hat{\sigma}_\alpha} \sqrt{n} \ , \ \frac{\hat{\beta} - \beta_0}{\hat{\sigma}_\beta} \sqrt{n} \right\}.$$

We may calculate the critical values from asymptotics (and they would depend on Σ) or by the bootstrap. Assuming that the bootstrap gives us critical value C, the confidence sets will be $C_\alpha = [\hat{\alpha} - C \frac{\hat{\sigma}_\alpha}{\sqrt{n}}, \hat{\alpha} + C \frac{\hat{\sigma}_\alpha}{\sqrt{n}}]$ and $C_\beta = [\hat{\beta} - C \frac{\hat{\sigma}_\beta}{\sqrt{n}}, \hat{\beta} + C \frac{\hat{\sigma}_\beta}{\sqrt{n}}]$.