Lecture Notes 11
Confidence Sets

1 Introduction

Let $\mathcal{P}$ be a statistical model. Let $C_n \equiv C_n(X_1,\ldots,X_n)$ be a set that is constructed from $X_1,\ldots,X_n$. Note that C_n is a random set. We say that C_n is a $1-\alpha$ *confidence set* for the parameter θ if

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

In other words

$$\inf_{P\in\mathcal{P}} P(\theta \in C_n) \geq 1-\alpha.$$

When

$$C_n = \Big[L(X_1,\ldots,X_n),\ U(X_1,\ldots,X_n)\Big]$$

we call C_n a *confidence interval.*

Important! C_n is random; θ is fixed.

Example 1 *Let $X_1,\ldots,X_n \sim N(\theta,\sigma)$. Suppose that σ is known. Let*

$$L = L(X_1,\ldots,X_n) = \overline{X} - c, \qquad U = U(X_1,\ldots,X_n) = \overline{X} + c.$$

Then

$$\begin{aligned}
P_\theta(L \leq \theta \leq U) &= P_\theta(\overline{X} - c \leq \theta \leq \overline{X} + c) \\
&= P_\theta(-c < \overline{X} - \theta < c) = P_\theta\left(-\frac{c\sqrt{n}}{\sigma} < \frac{\sqrt{n}(\overline{X}-\theta)}{\sigma} < \frac{c\sqrt{n}}{\sigma}\right) \\
&= P\left(-\frac{c\sqrt{n}}{\sigma} < Z < \frac{c\sqrt{n}}{\sigma}\right) = \Phi(c\sqrt{n}/\sigma) - \Phi(-c\sqrt{n}/\sigma) \\
&= 1 - 2\Phi(-c\sqrt{n}/\sigma) = 1-\alpha
\end{aligned}$$

if we choose $c = \sigma z_{\alpha/2}/\sqrt{n}$. So, if we define $C_n = \overline{X}_n \pm \sigma z_{\alpha/2}\sqrt{n}$ then

$$P_\theta(\theta \in C_n) = 1-\alpha$$

for all θ.

Example 2 *$X_i \sim N(\theta_i, 1)$ for $i = 1,\ldots,n$. Let*

$$C_n = \{\theta \in \mathbb{R}^n :\ ||X-\theta||^2 \leq \chi^2_{n,\alpha}\}.$$

Then

$$P_\theta(\theta \notin C_n) = P_\theta(||X-\theta||^2 > \chi^2_{n,\alpha}) = P(\chi^2_n > \chi^2_{n,\alpha}) = \alpha.$$

Four methods:

1. Probability Inequalities
2. Inverting a test
3. Pivots
4. Large Sample Approximations

NOTE: Optimal confidence intervals are confidence intervals that are as short as possible but we will not discuss optimality.

2 Using Probability Inequalities

Intervals that are valid for finite samples can be obtained by probability inequalities.

Example 3 *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 4 *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

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

3 Inverting a Test

For each θ_0, construct a level α test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. Define $\phi_{\theta_0}(x_1, \ldots, x_n) = 1$ if we reject and $\phi_{\theta_0}(x_1, \ldots, x_n) = 0$ if we don't reject. Let $A(\theta_0)$ be the acceptance region, that is,

$$A(\theta_0) = \Big\{ x_1, \ldots, x_n : \ \phi_{\theta_0}(x_1, \ldots, x_n) = 0 \Big\}.$$

Let

$$C_n \equiv C_n(x_1, \ldots, x_n) = \{\theta : \ (x_1, \ldots, x_n) \in A(\theta)\} = \{\theta : \ \phi_\theta(x_1, \ldots, x_n) = 0\}.$$

Theorem 5 *For each θ,*

$$P_\theta(\theta \in C(x_1, \ldots, x_n)) = 1 - \alpha.$$

Proof. Note that $1 - P_\theta(\theta \in C(x_1, \ldots, x_n))$ is the probability of rejecting θ when θ is true which is α. ■

The converse is also true:

Lemma 6 *If $C(x_1, \ldots, x_n)$ is a $1 - \alpha$ confidence interval then the test:*

$$\text{reject } H_0 \ \text{ if } \theta_0 \notin C(x_1, \ldots, x_n)$$

is a level α test.

Example 7 *Suppose we use the LRT. We reject H_0 when*

$$\frac{L(\theta_0)}{L(\widehat{\theta})} \leq c.$$

So

$$C = \left\{ \theta : \ \frac{L(\theta)}{L(\widehat{\theta})} \geq c \right\}.$$

Example 8 *Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ with σ^2 known. The LRT of $H_0 : \mu = \mu_0$ rejects when*

$$|\overline{X} - \mu_0| \geq \frac{\sigma}{\sqrt{n}} z_{\alpha/2}.$$

So

$$A(\mu) = \left\{ x^n : \ |\overline{X} - \mu_0| < \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \right\}$$

and so $\mu \in C(X^n)$ if and only if

$$|\overline{X} - \mu| \leq \frac{\sigma}{\sqrt{n}} z_{\alpha/2}.$$

In other words,

$$C_n = \overline{X}_n \pm \frac{\sigma}{\sqrt{n}} z_{\alpha/2}.$$

If σ is unknown, then this becomes

$$C_n = \overline{X}_n \pm \frac{S}{\sqrt{n}} t_{n-1,\alpha/2}.$$

(Good practice question.)

4 Pivots

A function $Q(X_1, \ldots, X_n, \theta)$ is a *pivot* if the distribution of Q does not depend on θ. For example, if $X_1, \ldots, X_n \sim N(\theta, 1)$ then

$$\overline{X}_n - \theta \sim N(0, 1/n)$$

so $Q = \overline{X}_n - \theta$ is a pivot.

Let a and b be such that

$$P_\theta(a \leq Q(X, \theta) \leq b) \geq 1 - \alpha$$

for all θ. We can find such an a and b because Q is a pivot. It follows immediately that

$$C(x) = \{\theta : \ a \leq Q(x, \theta) \leq b\}$$

has coverage $1 - \alpha$.

Example 9 *Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$. ($\sigma$ known.) Then*

$$Z = \frac{\sqrt{n}(\overline{X} - \mu)}{\sigma} \sim N(0, 1).$$

We know that

$$P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$$

and so

$$P\left(-z_{\alpha/2} \leq \frac{\sqrt{n}(\overline{X} - \mu)}{\sigma} \leq z_{\alpha/2}\right) = 1 - \alpha.$$

Thus

$$C = \overline{X} \pm \frac{\sigma}{\sqrt{n}} z_{\alpha/2}.$$

If σ is unknown, then this becomes

$$C = \overline{X} \pm \frac{S}{\sqrt{n}} t_{n-1,\alpha/2}$$

because

$$T = \frac{\sqrt{n}(\overline{X} - \mu)}{S} \sim t_{n-1}.$$

Example 10 *Let* $X_1, \ldots, X_n \sim \text{Uniform}(0, \theta)$. *Let* $Q = X_{(n)}/\theta$. *Then*

$$\mathbb{P}(Q \leq t) = \prod_i \mathbb{P}(X_i \leq t\theta) = t^n$$

so Q *is a pivot. Let* $c_n = \alpha^{1/n}$. *Then*

$$\mathbb{P}(Q \leq c_n) = \alpha.$$

Also, $\mathbb{P}(Q \leq 1) = 1$. *Therefore,*

$$\begin{aligned} 1 - \alpha &= \mathbb{P}(c \leq Q \leq 1) = \mathbb{P}\left(c \leq \frac{X_{(n)}}{\theta} \leq 1\right) \\ &= \mathbb{P}\left(\frac{1}{c} \geq \frac{\theta}{X_{(n)}} \geq 1\right) \\ &= \mathbb{P}\left(X_{(n)} \leq \theta \leq \frac{X_{(n)}}{c}\right) \end{aligned}$$

so a $1 - \alpha$ *confidence interval is*

$$\left(X_{(n)},\ \frac{X_{(n)}}{\alpha^{1/n}}\right).$$

5 Large Sample Confidence Intervals

The Wald Interval. We know that, under regularity conditions,

$$\frac{\widehat{\theta}_n - \theta}{\text{se}} \rightsquigarrow N(0, 1)$$

where $\widehat{\theta}_n$ is the mle and $\text{se} = 1/\sqrt{I_n(\widehat{\theta})}$. So this is an asymptotic pivot and an approximate confidence interval is

$$\widehat{\theta}_n \pm z_{\alpha/2}\text{se}.$$

By the delta method, a confidence interval for $\tau(\theta)$ is

$$\tau(\widehat{\theta}_n) \pm z_{\alpha/2}\text{se}(\widehat{\theta})|\tau'(\widehat{\theta}_n)|.$$

The Likelihood-Based Confidence Set. Let's consider inverting the asymptotic LRT. We test

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0.$$

Let k be the dimension of θ. We don't reject if

$$-2\log\left(\frac{L(\theta_0)}{L(\widehat{\theta})}\right) \leq \chi^2_{k,\alpha}$$

that is, if

$$\frac{L(\theta_0)}{L(\widehat{\theta})} > e^{-\chi^2_{p,\alpha}/2}.$$

So, the set of non-rejected nulls is

$$C_n = \left\{ \theta : \frac{L(\theta)}{L(\widehat{\theta})} > e^{-\frac{\chi^2_{p,\alpha}}{2}} \right\}.$$

This is an upper level set of the likelihood function. Then

$$P_\theta(\theta \in C) \to 1 - \alpha$$

for each θ.

Example 11 *Let* $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$. *Using the Wald statistic*

$$\frac{\widehat{p} - p}{\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}} \rightsquigarrow N(0,1)$$

so an approximate confidence interval is

$$\widehat{p} \pm z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}.$$

Using the LRT we get

$$C = \left\{ p : \ -2\log\left(\frac{p^Y(1-p)^{n-Y}}{\widehat{p}^Y(1-\widehat{p})^{n-Y}}\right) \le \chi^2_{1,\alpha} \right\}.$$

These intervals are different but, for large n, *they are nearly the same. A finite sample interval can be constructed by inverting a test.*

6 Tests Versus Confidence Intervals

Confidence intervals are more informative than tests. Look at Figure 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 informaive than simply reporting "reject" or "don't reject."

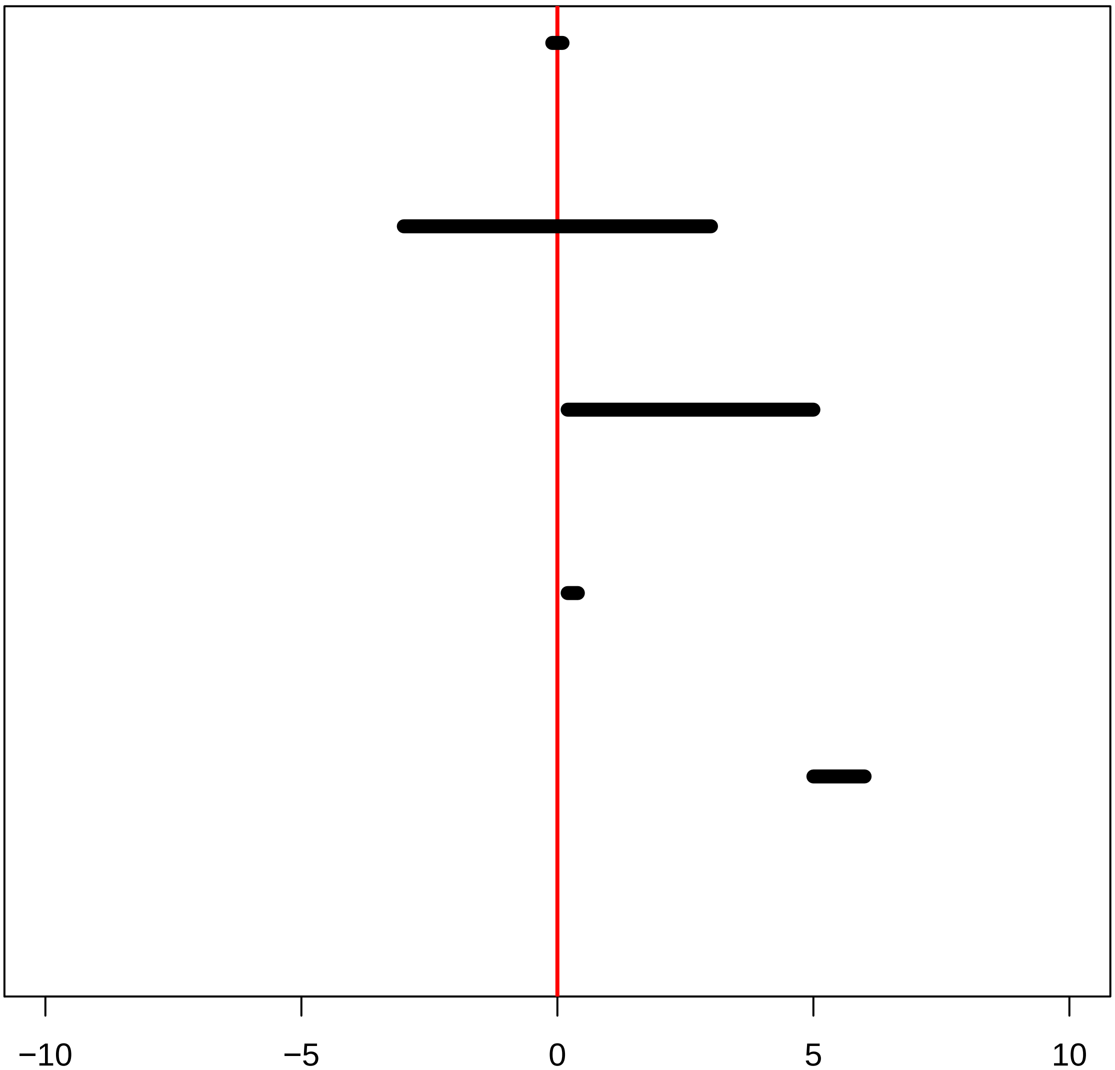

Figure 1: Five examples: 1. Not significant, precise. 2. Not significant, imprecise. 3. Barely significant, imprecise. 4. Barely significant, precise. 5. Significant and precise.

7 Confidence Sets for the cdf

Let $X_1, \ldots, X_n \sim F$. Recall that the Dvorestsky-Kiefer-Wolfowitz (DKW) inequality says that

$$\mathbb{P}(\sup_x |\widehat{F}_n(x) - F(x)| > \epsilon) \leq 2e^{-2n\epsilon^2}$$

where

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x) = \frac{\#\{X_i \leq x\}}{n}$$

is the empirical distribution function. (This is a bit sharper than the bound we get from the VC theorem). Let

$$L_n(x) = \max\{\widehat{F}_n(x) - \epsilon_n,\ 0\}, \qquad U_n(x) = \min\{\widehat{F}_n(x) + \epsilon_n,\ 1\}$$

where

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

It follows from the DKW inequality that, for any F,

$$P(L_n(x) \leq F(x) \leq U_n(x) \text{ for all } x) \geq 1 - \alpha.$$

We want to construct two functions $L(t) \equiv L(t, X)$ and $U(t) \equiv U(t, X)$ such that

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

for all F.

Let

$$K_n = \sup_x |F_n(x) - F(x)|$$

where

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x) = \frac{\#\{X_i \leq x\}}{n}$$

is the empirical distribution function. We claim that K_n is a pivot. To see this, let $U_i =$

$F(X_i)$. Then $U_1, \ldots, U_n \sim \text{Uniform}(0,1)$. So

$$
\begin{aligned}
K_n &= \sup_x |F_n(x) - F(x)| \\
&= \sup_x \left| \frac{1}{n} \sum_{i=1}^n I(X_i \leq x) - F(x) \right| \\
&= \sup_x \left| \frac{1}{n} \sum_{i=1}^n I(F(X_i) \leq F(x)) - F(x) \right| \\
&= \sup_x \left| \frac{1}{n} \sum_{i=1}^n I(U_i \leq F(x)) - F(x) \right| \\
&= \sup_{0 \leq t \leq 1} \left| \frac{1}{n} \sum_{i=1}^n I(U_i \leq t) - t \right|
\end{aligned}
$$

and the latter has a distribution depending only on $U_1, \ldots, U_n$. We could find, by simulation, a number c such that

$$
\mathbb{P}\left(\sup_{0 \leq t \leq 1} \left| \frac{1}{n} \sum_{i=1}^n I(U_i \leq t) - t \right| > c \right) = \alpha.
$$

A confidence set is then

$$
C = \{F : \sup_x |F_n(x) - F(x)| < c\}.
$$