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 $\theta$ 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

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

we call $C_n$ a confidence interval.

Important! $C_n$ is random; $\theta$ is fixed.

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

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

Then

$$P_\theta(L \leq \theta \leq U) = P_\theta(\overline{X} - c \leq \theta \leq \overline{X} + c)$$

$$= P_\theta(-c < \overline{X} - \theta < c) = P_\theta \left( -\frac{c\sqrt{n}}{\sigma} < \frac{\sqrt{n}(\overline{X} - \theta)}{\sigma} < \frac{c\sqrt{n}}{\sigma} \right)$$

$$= P \left( -\frac{c\sqrt{n}}{\sigma} < Z < \frac{c\sqrt{n}}{\sigma} \right) = \Phi\left(c\sqrt{n}/\sigma\right) - \Phi\left(-c\sqrt{n}/\sigma\right)$$

$$= 1 - 2\Phi\left(-c\sqrt{n}/\sigma\right) = 1 - \alpha$$

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 $\theta$.

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

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

1
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:
\[
P(|\hat{p} - p| > \varepsilon) \leq 2e^{-2n\varepsilon^2}.
\]
Let
\[
\varepsilon_n = \sqrt{\frac{1}{2n} \log \left(\frac{2}{\alpha}\right)}.
\]
Then
\[
P\left(|\hat{p} - p| > \sqrt{\frac{1}{2n} \log \left(\frac{2}{\alpha}\right)}\right) \leq \alpha.
\]

Hence, \(p \in C\) \(\geq 1 - \alpha\) where \(C = (\hat{p} - \varepsilon_n, \hat{p} + \varepsilon_n)\).

Example 4 Let \(X_1, \ldots, X_n \sim F\). Suppose we want a confidence band for \(F\). Remember that
\[
P\left(\sup_x |F_n(x) - F(x)| > \varepsilon\right) \leq 2e^{-2n\varepsilon^2}.
\]
Let
\[
\varepsilon_n = \sqrt{\frac{1}{2n} \log \left(\frac{2}{\alpha}\right)}.
\]
Then
\[
P\left(\sup_x |F_n(x) - F(x)| > \sqrt{\frac{1}{2n} \log \left(\frac{2}{\alpha}\right)}\right) \leq \alpha.
\]

Hence,
\[
P_F(L(t) \leq F(t) \leq U(t) \text{ for all } t \geq 1 - \alpha
\]
for all \(F\), where
\[
L(t) = \hat{F}_n(t) - \varepsilon_n, \quad U(t) = \hat{F}_n(t) + \varepsilon_n.
\]

We can improve this by taking
\[
L(t) = \max \left\{\hat{F}_n(t) - \varepsilon_n, 0\right\}, \quad U(t) = \min \left\{\hat{F}_n(t) + \varepsilon_n, 1\right\}.
\]
3 Inverting a Test

For each \( \theta_0 \), construct a level \( \alpha \) 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,

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

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 \( \theta \),

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

**Proof.** Note that \( 1 - P_\theta(\theta \in C(x_1, \ldots, x_n)) \) is the probability of rejecting \( \theta \) when \( \theta \) is true which is \( \alpha \). \[\blacksquare\]

The converse is also true:

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

reject \( H_0 \) if \( \theta_0 \notin C(x_1, \ldots, x_n) \)

is a level \( \alpha \) test.

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

\[
\frac{L(\theta_0)}{L(\hat{\theta})} \leq c.
\]

So

\[
C = \left\{ \theta : \frac{L(\theta)}{L(\hat{\theta})} \geq c \right\}.
\]

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

\[
|\bar{X} - \mu_0| \geq \frac{\sigma}{\sqrt{n}} z_{\alpha/2}.
\]

So

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

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

\[
|\bar{X} - \mu| \leq \frac{\sigma}{\sqrt{n}} z_{\alpha/2}.
\]
In other words,
\[ C_n = \bar{X}_n \pm \frac{\sigma}{\sqrt{n}} z_{\alpha/2}. \]

*If \( \sigma \) is unknown, then this becomes*
\[ C_n = \bar{X}_n \pm \frac{S}{\sqrt{n}} t_{n-1, \alpha/2}. \]

(Good practice question.)

4 Large Sample Confidence Intervals

**The Wald Interval.** We know that, under regularity conditions,
\[ \frac{\hat{\theta}_n - \theta}{\text{se}} \rightarrow N(0, 1) \]
where \( \hat{\theta}_n \) is the mle and \( \text{se} = 1/\sqrt{I_n(\theta)} \). So this is an asymptotic pivot and an approximate confidence interval is
\[ \hat{\theta}_n \pm z_{\alpha/2} \text{se}. \]

By the delta method, a confidence interval for \( \tau(\theta) \) is
\[ \tau(\hat{\theta}_n) \pm z_{\alpha/2} \text{se}(\hat{\theta}) | \tau'(\hat{\theta}_n)|. \]

**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 \( \theta \). We don’t reject if
\[ -2 \log \left( \frac{L(\theta_0)}{L(\theta)} \right) \leq \chi^2_{k, \alpha} \]
that is, if
\[ \frac{L(\theta_0)}{L(\theta)} > e^{-\chi^2_{k, \alpha}/2}. \]

So, the set of non-rejected nulls is
\[ C_n = \left\{ \theta : \frac{L(\theta)}{L(\theta)} > e^{-\chi^2_{k, \alpha}/2} \right\}. \]

This is an upper level set of the likelihood function. Then
\[ P_\theta(\theta \in C) \rightarrow 1 - \alpha \]
for each \( \theta \).
Example 9 Let $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$. Using the Wald statistic

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

so an approximate confidence interval is

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

Using the LRT we get

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

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

5 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 informative 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.