

## INTERVAL ESTIMATION

CB ch. 9, 10.4



In *point estimation*, we need a measure of variance of point estimator: implicitly, if variance is smaller, we have more confidence in our point estimator.

In *interval estimation*, we are very concerned about confidence, and so give up trying to get point estimate of parameter, but rather just try to know that it lies within some region with some prescribed probability.

Consider a model  $f(\vec{X}|\theta)$ , with data  $\vec{X}$ . We want to know what  $\theta$  is.

In interval estimation, we want to make statements like:

$$\theta \in C(\vec{X}) \subset \Theta,$$

where  $C(\vec{X})$  is an interval estimate, which depends on the observed data  $\vec{X}$ .  $C(\vec{X})$  is an interval on the real line (or a connected region in multidimensional Euclidean space). We are not interested in disconnected segments.

**Definition 9.1.1:** an interval estimate for a real-valued parameter  $\theta$  based on a sample  $\vec{X} \equiv (X_1, \dots, X_n)$  is a *pair of functions*  $L(\vec{X})$  and  $U(\vec{X})$  so that  $L(\vec{X}) \leq U(\vec{X})$  for all  $\vec{X}$ . For the observed data  $\vec{X}^o$ , the inference  $L(\vec{X}^o) \leq \theta \leq U(\vec{X}^o)$  is made.

Note:

- Both  $L(\vec{X})$  and  $U(\vec{X})$  are random variables, so that  $C(\vec{X}) \equiv [L(\vec{X}), U(\vec{X})]$  is a random interval.
- $[L(\vec{X}), U(\vec{X})]$  is a two-sided interval. Sometimes, you seek  $(-\infty, U(\vec{X})]$  or  $[L(\vec{X}), \infty)$ , which are one-sided intervals.

**Definition 9.1.4:** the **coverage probability** of an interval estimator is  $P_\theta \left( \theta \in [L(\vec{X}), U(\vec{X})] \right)$ .

This is the probability that the random interval  $[L(\vec{X}), U(\vec{X})]$  “covers” the true  $\theta$ .

Important: in the expression for the coverage probability,  $\theta$  is not random, but  $L(\vec{X})$  and  $U(\vec{X})$  are. So  $P_\theta \left( \theta \in [L(\vec{X}), U(\vec{X})] \right)$  means  $P_\theta \left( L(\vec{X}) \leq \theta, U(\vec{X}) \geq \theta \right)$ .

One problem about the coverage probability is that it can vary depend on what  $\theta$  is.

**Definition 9.1.5:** For an interval estimator  $[L(\vec{X}), U(\vec{X})]$  of a parameter  $\theta$ , the **confidence coefficient**  $\equiv \inf_\theta P_\theta \left( \theta \in [L(\vec{X}), U(\vec{X})] \right)$ .

The confidence coefficient does not depend on  $\theta$ .

Usually, we use the term **confidence interval** to refer to a combination of an interval estimate, along with a measure of confidence (such as the confidence coefficient). Hence, a confidence interval is a statement like “ $\theta$  is between 1.5 and 2.8 with probability 80%.”



**Example:**  $X_1, \dots, X_n \sim i.i.d. U[0, \theta]$ , and  $Y_n \equiv \max(X_1, \dots, X_n)$  (the sample maximum). Consider two interval estimators

1.  $[aY_n, bY_n]$ , where  $1 \leq a < b$
2.  $[Y_n + c, Y_n + d]$ , where  $0 \leq c < d$ .

What is the confidence coefficient of each?

(1) The coverage probability

$$\begin{aligned} P_\theta(\theta \in [aY_n, bY_n]) &= P_\theta(aY_n \leq \theta \leq bY_n) \\ &= P_\theta\left(\frac{\theta}{b} \leq y \leq \frac{\theta}{a}\right). \end{aligned}$$

From before, we know that density of  $Y_n$  is  $f(y) = \frac{1}{\theta^n} ny^{n-1}$ , for  $y \in [0, \theta]$ , so that

$$\begin{aligned} P_\theta\left(\frac{\theta}{b} \leq y \leq \frac{\theta}{a}\right) &= \frac{1}{\theta^n} \int_{\frac{\theta}{b}}^{\frac{\theta}{a}} ny^{n-1} dy \\ &= \frac{1}{\theta^n} \left[ \left(\frac{\theta}{a}\right)^n - \left(\frac{\theta}{b}\right)^n \right]. \end{aligned}$$

Since coverage probability is not a function of  $\theta$ , then this is also confidence coefficient.

(2) The coverage probability

$$\begin{aligned} P_\theta(\theta \in [Y_n + c, Y_n + d]) &= P_\theta(Y_n + c \leq \theta \leq Y_n + d) \\ &= P_\theta\left(1 - \frac{d}{\theta} \leq \frac{y}{\theta} \leq 1 - \frac{c}{\theta}\right) \\ &= P_\theta(\theta - d \leq y \leq \theta - c) \\ &= \frac{1}{\theta^n} \int_{\theta-d}^{\theta-c} ny^{n-1} dy = \frac{1}{\theta^n} ((\theta - c)^n - (\theta - d)^n) \end{aligned}$$

so that coverage probability depends on  $\theta$ .

But note that  $\lim_{\theta \rightarrow \infty} \frac{1}{\theta^n} ((\theta - c)^n - (\theta - d)^n) = 0$ , so that confidence coefficient is 0.



## Methods of Finding Interval Estimators

General principle: “invert” a test statistic

Testing tells you: under a given  $H_0 : \theta = \theta_0$ ,

$$P_{\theta_0}(T_n \in R) = \alpha \Leftrightarrow P_{\theta_0}(T_n \notin R) = 1 - \alpha,$$

where  $T_n$  is a test statistic.

We can use this to construct a  $(1 - \alpha)$  confidence interval:

- If you *fix*  $\alpha$ , but vary the null hypothesis  $\theta_0$ , then you obtain  $R(\theta_0)$ , a rejection region for each  $\theta_0$  such that, by construction:

$$\forall \theta_0 \in \Theta : P_{\theta_0}(T_n \notin R(\theta_0)) = 1 - \alpha.$$

- Now, for a given sample  $\vec{X} \equiv X_1, \dots, X_n$ , consider the set

$$C(\vec{X}) \equiv \left\{ \theta : T_n(\vec{X}) \notin R(\theta) \right\}.$$

By construction:

$$P_{\theta}(\theta \in C(\vec{X})) = P_{\theta}(T_n(\vec{X}) \notin R(\theta)) = 1 - \alpha, \forall \theta \in \Theta.$$

Therefore,  $C(\vec{X})$  is a  $(1 - \alpha)$  confidence interval for  $\theta$ .

- The confidence interval  $C(\vec{X})$  is the set of  $\theta$ 's such that, for the given data  $\vec{X}$  and for each  $\theta_0 \in C(\vec{X})$ , you would not be able to reject the null hypothesis  $H_0 : \theta = \theta_0$ .
- In hypothesis testing, acceptance region is set of  $\vec{X}$  which are very likely for a fixed  $\theta_0$ .

In interval estimation, confidence interval is set of  $\theta$ 's which make  $\vec{X}$  very likely, for a fixed  $\vec{X}$ .

This “inversion” exercise is best illustrated graphically.



**Example:**  $X_1, \dots, X_n \sim i.i.d. N(\mu, 1)$ .

We want to construct a 95% CI for  $\mu$  by inverting the T-test.

- We know that, under each null  $H_0 : \mu = \mu_0$ ,  $\sqrt{n}(\bar{X}_n - \mu_0) \sim N(0, 1)$ .
- Hence, for each  $\mu_0$ , a 95% acceptance region is

$$\begin{aligned} & \{-1.96 \leq \sqrt{n}(\bar{X}_n - \mu_0) \leq 1.96\} \\ \Leftrightarrow & \left\{ \bar{X}_n - \frac{1}{\sqrt{n}}1.96 \leq \mu_0 \leq \bar{X}_n + \frac{1}{\sqrt{n}}1.96 \right\} = C(\bar{X}_n) \end{aligned}$$

- Graph:

- Now consider what happens when we invert one-sided test. Consider the hypotheses  $H_0 : \mu \leq \mu_0$  vs.  $H_1 : \mu > \mu_0$ . Then a 95% acceptance region is

$$\begin{aligned} & \{\sqrt{n}(\bar{X}_n - \mu_0) \leq 1.645\} \\ \Leftrightarrow & \left\{ \mu_0 \leq \bar{X}_n - \frac{1}{\sqrt{n}}1.65 \right\} = C(\bar{X}_n) \end{aligned}$$

- Graph:



**Example:**  $X_1, \dots, X_n \sim U[0, \theta]$ . Construct a confidence set based on inverting the LRT.

- Consider the one-sided hypotheses  $H_0 : \theta \leq \theta_0$  vs.  $H_1 : \theta > \theta_0$ .
- As derived before:  $\lambda(\vec{X}) = \begin{cases} 0 & \text{if } \max(X_1, \dots, X_n) > \theta_0 \\ 1 & \text{otherwise.} \end{cases}$
- So acceptance region is  $\{Y_n \equiv \max(X_1, \dots, X_n) \leq \theta_0\}$  and confidence set is  $\{\theta_0 : \theta_0 \geq Y_n\} = C(Y_n)$ .
- Graph

- What is confidence coefficient?  $P_\theta(\theta \in C(Y_n)) = P_\theta(\theta \geq Y_n) = 1$ , for all  $\theta$ !



### Pivotal quantities

- Some confidence intervals are easier (more convenient) to analyze than others
- Consider the problem  $X_1, \dots, X_n \sim i.i.d. N(\mu, \sigma^2)$ .
- Let  $\bar{X}_n \equiv \frac{1}{n} \sum_i X_i$ , and consider two confidence intervals for  $\mu$ :
  1.  $\mu \in \left[ \bar{X}_n - \frac{a}{\sqrt{n}}, \bar{X}_n + \frac{a}{\sqrt{n}} \right], a > 0$
  2.  $\mu \in \left[ \bar{X}_n - \frac{a}{\sqrt{n}}\sigma, \bar{X}_n + \frac{a}{\sqrt{n}}\sigma \right], a > 0$

What is the coverage probability and confidence coefficient for each interval?

1.

$$\begin{aligned} & P_{(\mu, \sigma^2)} \left( \mu \in \left[ \bar{X}_n - \frac{a}{\sqrt{n}}, \bar{X}_n + \frac{a}{\sqrt{n}} \right] \right) \\ &= P_{(\mu, \sigma^2)} \left( -\frac{a}{\sqrt{n}} \leq \mu - \bar{X}_n \leq \frac{a}{\sqrt{n}} \right) \\ &= P_{(\mu, \sigma^2)} \left( -a \leq \sqrt{n}(\bar{X}_n - \mu) \leq a \right) \\ &= \Phi(a; 0, \sigma^2) - \Phi(-a; 0, \sigma^2) \\ &= 2 \cdot \Phi(a; 0, \sigma^2) - 1. \end{aligned}$$

Hence, confidence coefficient is  $\inf_{\sigma^2} 2 \cdot \Phi(a; 0, \sigma^2) - 1$ .

2.

$$\begin{aligned} & P_{(\mu, \sigma^2)} \left( \mu \in \left[ \bar{X}_n - \frac{a}{\sqrt{n}}\sigma, \bar{X}_n + \frac{a}{\sqrt{n}}\sigma \right] \right) \\ &= P_{(\mu, \sigma^2)} \left( -a \leq \sqrt{n}(\bar{X}_n - \mu)/\sigma \leq a \right) \\ &= 2 \cdot \Phi(a; 0, 1) - 1 \end{aligned}$$

so that confidence coefficient is also  $2 \cdot \Phi(a; 0, 1) - 1$ .

This case is easier to analyze, because the coverage probability does not depend on any parameters. For this reason, we prefer confidence intervals based on “pivotal quantities”.



**Definition 9.2.6:** A random variable  $Q(\vec{X}; \theta)$  is a **pivotal quantity** if its distribution is independent of all parameters. That is,  $Q(\vec{X}; \theta)$  has the same distribution for all values of  $\theta$ .

When we create CI's by inverting tests, the relevant pivotal quantity is the test statistic:

- In the previous example, the two test statistics were (i)  $\sqrt{n}(\bar{X}_n - \mu)$ ; (ii)  $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ .

Pivotalness implies that the distribution of the test statistic does not depend on parameters: (i) is distributed  $N(0, \sigma^2)$  (so not pivotal), whereas (ii) is pivotal (distributed  $N(0, 1)$ ).

- For the uniform distribution case, the statistic  $Y_n \equiv \max(X_1, \dots, X_n)$  is not pivotal, but  $Y_n/\theta$  is. (Recall earlier example.)

- Later, talk about test statistics which are *asymptotically* pivotal (including T-statistics, LRT)



## Methods for Evaluating Interval Estimators

Small sample properties (Section 9.3)

Generally, we want to have confidence intervals with high confidence coefficients, as well as small size/length.

Problem is: for a given confidence coefficient  $(1 - \alpha)$ , find the CI with the *shortest length*.

**Example:**  $X_1, \dots, X_n \sim i.i.d. N(\mu, \sigma^2)$  with  $\sigma^2$  known.

- As derived above,  $Z_n \equiv \sqrt{n}(\bar{X}_n - \mu)/\sigma$  is pivotal, therefore any  $(a, b)$  satisfying

$$P_\mu(a \leq Z_n \leq b) = \Phi(b) - \Phi(a) = 1 - \alpha$$

yields a corresponding  $(1 - \alpha)$ -CI for  $\mu$ :

$$\left\{ \mu : \bar{X}_n - b \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n - a \frac{\sigma}{\sqrt{n}} \right\}.$$

- Now we want to choose  $(a, b)$  so that  $b - a$  is the shortest length possible, for a given confidence coefficient  $(1 - \alpha)$ .
- It turns out that the *symmetric* solution  $a = -b$  is optimal here. The symmetric solution is

$$1 - \alpha = \Phi(-a) - \Phi(a) = 1 - 2\Phi(a) \Rightarrow a = \Phi^{-1}\left(\frac{\alpha}{2}\right).$$

This result generalizes to any sampling distribution that is unimodal.

**Theorem 9.3.2:** Let  $f(x)$  be a unimodal pdf. If the interval satisfies

(i)  $\int_a^b f(x)dx = 1 - \alpha$

(ii)  $f(a) = f(b) > 0$

(iii)  $a \leq x^* \leq b$ , where  $x^*$  is a mode of  $f(x)$ , then  $[a, b]$  is the shortest lengthèd interval satisfying (i).

Note:

- $x$  in the above theorem denotes the pivotal statistic upon which the CI is based
- $f(x)$  need not be symmetric: (graph)

- However, when  $f(x)$  is symmetric, and  $x^* = 0$ , then  $a = -b$ . This is the case for the  $N(0, 1)$  density.



### Large-sample confidence intervals

Many common tests (T-test, LRT) are based on asymptotically-valid reject regions. CI derived by inverting these test are then asymptotically-valid CI's.

Consider the T-statistic:  $Z_n = \frac{(\hat{\theta}_n - \theta_0)}{\sqrt{\frac{1}{n}\hat{V}}} \xrightarrow{d} N(0, 1)$ .

A two-sided T-test with the rejection region  $\{|Z_n| > 1.96\}$  would have asymptotic size  $\alpha = 0.05$ .

The corresponding asymptotic  $(1 - \alpha)$ -confidence interval would be

$$\left\{ \hat{\theta}_n - 1.96\sqrt{\frac{1}{n}\hat{V}} \leq \theta_0 \leq \hat{\theta}_n + 1.96\sqrt{\frac{1}{n}\hat{V}} \right\}.$$

Moreover, note that the  $N(0, 1)$  distribution satisfies Thm 9.3.2, and is also symmetric, with mode  $x^* = 0$ , so that this asymptotic CI is the shortest lengthèd CI for the confidence coefficient  $(1 - \alpha)$ .

Finally, note that the T-statistic is asymptotically pivotal, so that the coverage probability of the corresponding CI is identical for all  $\theta_0$ .



Similarly, you can derive CI's by inverting the likelihood ratio test statistic: a size-0.05

test would be

$$\mathbf{1} \left[ -2 \cdot \sum_i \left( \log f(X_i|\theta_0) - \log f(X_i|\hat{\theta}_n) \right) \geq 3.84 \right].$$

The corresponding asymptotic 95% CI would be

$$\left\{ \theta_0 : -2 \cdot \sum_i \left( \log f(X_i|\theta_0) - \log f(X_i|\hat{\theta}_n) \right) \leq 3.84 \right\}.$$

This is a very complicated calculation. However, consider the example where

$$X_1, \dots, X_n \sim i.i.d. N(\mu, 1).$$

Can you derive the 95% CI corresponding to inverting the 0.05-LRT?