Lecture 16: Confidence Intervals

MATH 667-01
Statistical Inference
University of Louisville

November 14, 2017

Introduction

- We start with the definition of an interval estimator and a confidence interval in Section 9.1 of Casella and Berger (2002)[1].
- Then in Section 9.2, two methods (inverting a test, using a pivot) of finding confidence intervals are discussed. Bayesian credible intervals are also discussed.

[1] Casella, G. and Berger, R. (2002). *Statistical Inference, Second Edition*. Duxbury Press, Belmont, CA.

- In this chapter, we consider interval estimation of the unknown parameter θ (or function of the unknown parameter) in a parametric model $\boldsymbol{X} \sim f_{\boldsymbol{X}}(\boldsymbol{x}|\theta)$.
- For instance, we might wish to make an inference about p for the model where $X_1, \dots, X_n$ is a random sample from a Bernoulli(p) population. The results of this inferential procedure might be reported as:
  "A 95% confidence interval for p is $(0.445, 0.498)$."
- It is important to interpret confidence intervals correctly. A common incorrect interpretation of the above interval is:
  "The probability that p is in the interval $(0.445, 0.498)$ is 0.95."
  Frequentists assume p is fixed, so the probability of this event is either 0 (if p is not in the interval) or 1 (if p is in the interval).

Introduction

- The key idea in interpreting the probabilistic statement of a confidence interval is that, after we observe the sample, the endpoints of the confidence interval are not random so the statement is either correct or incorrect. Before we observe the sample, the endpoints are random and there is some probability that the interval defined by the two endpoints will include the true value of the parameter.
- *Definition L16.1* (Def 9.1.1 on p.417): An *interval estimate* of a real-valued parameter θ is any pair of functions $L(x_1, \ldots, x_n)$ and $U(x_1, \ldots, x_n)$, of a sample that satisfy $L(\boldsymbol{x}) \leq U(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathcal{X}$. If $\boldsymbol{X} = \boldsymbol{x}$ is observed, the inference $L(\boldsymbol{x}) \leq \theta \leq U(\boldsymbol{x})$ is made. The random interval $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ is called an *interval estimator*.
- If one of the endpoints is infinite, then the interval estimate is said to be *one-sided*. Otherwise, the interval estimate is *two-sided*.

- *Definition L16.2* (Def 9.1.4 on p.418): For an *interval estimator* $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ of a parameter θ, the *coverage probability* of $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ is the probability that the random interval $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ covers the true parameter, θ. In symbols, it is denoted by

$$P_\theta(L(\boldsymbol{X}) \leq \theta \leq U(\boldsymbol{X})).$$

- *Definition L16.3* (Def 9.1.5 on p.418): For an interval estimator $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ of a parameter θ, the *confidence coefficient* of $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ is the infimum of the coverage probabilities,

$$\inf_\theta P_\theta(L(\boldsymbol{X}) \leq \theta \leq U(\boldsymbol{X})).$$

  An interval estimator of θ with confidence coefficient $1 - \alpha$ is called a $100(1 - \alpha)$*% confidence interval for* θ.

- *Example L16.1*: Suppose $X_1, \ldots, X_n$ is a random sample from a Normal(μ, σ^2) population with both μ and σ^2 unknown. Let $L(X_1, \ldots, X_n) = \bar{X} - \frac{S}{\sqrt{n}} t_{n-1,.025}$ and $U(X_1, \ldots, X_n) = \bar{X} + \frac{S}{\sqrt{n}} t_{n-1,.025}$ where $\bar{X}$ is the sample mean, S is the sample standard deviation, and $t_{N,\alpha}$ is the value that satisfies $P(T > t_{N,\alpha}) = \alpha$ for $T \sim t_N$. Show that $[L(X_1, \ldots, X_n), U(X_1, \ldots, X_n)]$ is a 95% confidence interval for μ.
- *Answer to Example L16.1*: Here we will use the fact that $\dfrac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$ (see slide 13.10).
  Now we show that
  $P(L(X_1, \ldots, X_n) < \mu < U(X_1, \ldots, X_n)) = .95.$

- *Answer to Example L16.1 continued*:

$$
\begin{aligned}
& P\left(\bar{X} - \frac{S}{\sqrt{n}} t_{n-1,.025} \le \mu \le \bar{X} + \frac{S}{\sqrt{n}} t_{n-1,.025}\right) \\
= \; & P\left(\bar{X} - \frac{S}{\sqrt{n}} t_{n-1,.025} \le \mu \text{ and } \mu \le \bar{X} + \frac{S}{\sqrt{n}} t_{n-1,.025}\right) \\
= \; & P\left(\bar{X} \le \mu + \frac{S}{\sqrt{n}} t_{n-1,.025} \text{ and } \mu - \frac{S}{\sqrt{n}} t_{n-1,.025} \le \bar{X}\right) \\
= \; & P\left(\frac{\bar{X} - \mu}{S/\sqrt{n}} \le t_{n-1,.025} \text{ and } -t_{n-1,.025} \le \frac{\bar{X} - \mu}{S/\sqrt{n}}\right) \\
= \; & P\left(-t_{n-1,.025} \le \frac{\bar{X} - \mu}{S/\sqrt{n}} \text{ and } \frac{\bar{X} - \mu}{S/\sqrt{n}} \le t_{n-1,.025}\right) \\
= \; & 1 - P\left(-t_{n-1,.025} > \frac{\bar{X} - \mu}{S/\sqrt{n}} \text{ or } \frac{\bar{X} - \mu}{S/\sqrt{n}} > t_{n-1,.025}\right) \\
= \; & 1 - \left\{P\left(\frac{\bar{X} - \mu}{S/\sqrt{n}} < -t_{n-1,.025}\right) + P\left(\frac{\bar{X} - \mu}{S/\sqrt{n}} > t_{n-1,.025}\right)\right\} \\
= \; & 1 - (.025 + .025) = 1 - .05 = .95
\end{aligned}
$$

Normal simulation:

- Consider the previous example with population parameters $\mu = 8$ and $\sigma^2 = 4$, and sample size $n = 5$.
- Suppose that the 5 values in the sample are 8.9, 8.8, 5.5, 6.1, and 11.5. Based on this sample, the 95% confidence interval for μ is $(5.15, 11.17)$. Note that $\mu = 8 \in (5.15, 11.17)$.
- However, if we used a different sample of size 5, we would obtain a different confidence interval.
- If we repeat this process for many samples of size 5, the true value $\mu = 8$ will be in confidence interval about 95% of the time.

Confidence Interval Simulation

Sample #	Sample	95% CI	Result
1	$8.9, 8.8, 5.5, 6.1, 11.5$	$(5.15, 11.17)$	$8 \in (5.15, 11.17)$
2	$9.2, 7.2, 9.2, 5.6, 5.9$	$(5.27, 9.57)$	$8 \in (5.27, 9.57)$
3	$9.3, 8.8, 8.9, 8.7, 10.2$	$(8.42, 9.94)$	$8 \notin (8.42, 9.94)$
4	$8.6, 9.0, 5.4, 6.1, 9.8$	$(5.40, 10.16)$	$8 \in (5.40, 10.16)$
5	$9.8, 10.0, 4.3, 5.9, 5.4$	$(3.80, 10.36)$	$8 \in (3.80, 10.36)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
10000	$7.4, 5.8, 8.0, 4.0, 4.2$	$(3.63, 8.13)$	$8 \in (3.63, 8.13)$

The proportion of times that μ is in the 95% CI for this simulation is 0.9485.

Confidence Interval Simulation

Code for simulation with 10000 samples:

```
set.seed(100000)
n=5;mu=8;sigma=2
count=0
est.mean=rep(0,10000)
lower.limit=rep(0,10000)
upper.limit=rep(0,10000)
for (i in 1:10000){
 x=rnorm(n,mean=mu,sd=sigma)
 est.mean[i]=mean(x)
 lower.limit[i]=est.mean[i]-qt(.975,df=n-1)*sd(x)/sqrt(n)
 upper.limit[i]=est.mean[i]+qt(.975,df=n-1)*sd(x)/sqrt(n)
}
accept.H0=(lower.limit<mu)&(mu<upper.limit)
proportion=mean(accept.H0)
proportion
```

Methods of Finding Interval Estimators

- Now, we consider methods for finding interval estimators of parameters of interest.
- Specifically, in class we will cover three widely used methods:
  1. inverting a hypothesis test
  2. using a pivotal quantity
  3. finding a Bayesian credible interval/set.

Inverting a Hypothesis Test

- *Theorem L16.1* (Thm 9.2.2 on p.421): For each $\theta_0 \in \Theta$, let $A(\theta_0)$ be the acceptance region of a level α test of $H_0 : \theta = \theta_0$. For each $\boldsymbol{x} \in \mathcal{X}$, define a set $C(\boldsymbol{x})$ in the parameter space by

$$C(\boldsymbol{x}) = \{\theta_0 : \boldsymbol{x} \in A(\theta_0)\}.$$

  Then the random set $C(\boldsymbol{X})$ is a $1 - \alpha$ confidence set. Conversely, let $C(\boldsymbol{X})$ be a $1 - \alpha$ confidence set. For any $\theta_0 \in \Theta$, define $A(\theta_0) = \{\boldsymbol{x} : \theta_0 \in C(\boldsymbol{x})\}$. Then $A(\theta_0)$ is the acceptance region of the level α test of $H_0 : \theta = \theta_0$.

Inverting a Hypothesis Test

- *Proof of Theorem L16.1*: Suppose $A(\theta_0)$ is the acceptance region of a level α test. Then $P_{\theta_0}(\boldsymbol{X} \notin A(\theta_0)) \leq \alpha$ which implies that $P_{\theta_0}(\boldsymbol{X} \in A(\theta_0)) = 1 - P_{\theta_0}(\boldsymbol{X} \notin A(\theta_0)) \geq 1 - \alpha$. Since θ_0 is arbitraty, we write θ instead of θ_0 so that $P_\theta(\theta \in C(\boldsymbol{X})) = P_\theta(\boldsymbol{X} \in A(\theta)) \geq 1 - \alpha$ for all $\theta \in \Theta$. By *Definition L16.3*, $C(\boldsymbol{X})$ is a $100(1-\alpha)\%$ confidence set for θ.
- Conversely, suppose $C(\boldsymbol{X})$ is a $100(1-\alpha)\%$ confidence set for θ. Then the probability of rejecting H_0 in testing $H_0 : \theta = \theta_0$ using acceptance region $A(\theta_0)$ is

$$P_{\theta_0}(\boldsymbol{X} \notin A(\theta)) = P_\theta(\theta \notin C(\boldsymbol{X})) = 1 - P_\theta(\theta \in C(\boldsymbol{X})) \leq 1 - (1-\alpha) = \alpha.$$

So, by *Definition L12.6*, this is a level α test.

Inverting a Hypothesis Test

- *Example L16.2*: Suppose $X_1, \ldots, X_n$ is a random sample from a distribution with pdf

$$f(x|\mu) = e^{-(x-\mu)} I_{[\mu,\infty)}(x).$$

  The likelihood ratio test of $H_0 : \mu = \mu_0$ versus $H_a : \mu > \mu_0$ has a rejection region of the form $\left\{ \boldsymbol{x} : x(1) \geq \mu_0 - \frac{1}{n} \ln \alpha \right\}$ (see *Example L14.3*). Find a $100(1-\alpha)\%$ confidence interval for μ by inverting this hypothesis test.

Inverting a Hypothesis Test

- *Answer to Example L16.2*: By *Theorem L16.1*, a $100(1-\alpha)\%$ confidence interval for μ is

$$\begin{aligned} C(\boldsymbol{x}) &= \{\mu : \boldsymbol{x} \in A(\mu)\} \\ &= \left\{\mu : \mu \leq x_{(1)} < \mu - \frac{1}{n}\ln\alpha\right\} \\ &= \left\{\mu : x_{(1)} + \frac{1}{n}\ln\alpha < \mu \leq x_{(1)}\right\}. \end{aligned}$$

Equivalently, we can write that $\left(X_{(1)} + \frac{1}{n}\ln\alpha, X_{(1)}\right]$ is a $100(1-\alpha)\%$ confidence interval for μ.

Using a Pivotal Quantity

- *Definition L16.4* (Def 9.2.6 on p.427): A random variable $Q(\boldsymbol{X}, \theta) = Q(X_1, \ldots, X_n, \theta)$, is a *pivotal quantity* (or *pivot*) if the distribution of $Q(\boldsymbol{X}, \theta)$ is independent of all parameters. That is, if $\boldsymbol{X} \sim F(\boldsymbol{x}|\theta)$, then $Q(\boldsymbol{X}, \theta)$ has the same distribution for all values of θ.
- *Example L16.3*: Suppose $X_1, \ldots, X_n$ are iid Normal random variables with both μ and σ^2 unknown.

  (a) Show that the random variable $\dfrac{(n-1)S^2}{\sigma^2}$ is a pivot.
  (b) Use the pivot to construct a $100(1-\alpha)\%$ confidence interval for σ^2.
  (c) Find a level α test of $H_0 : \sigma^2 = \sigma_0^2$ corresponding to the confidence interval in (b).

- *Answer to Example L16.3*: (a) Since $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ has the same distribution for all μ and σ^2, it is a pivot.
- (b) If $Q \sim \chi^2_N$, let $\chi^2_{N,\alpha}$ be the value such that $P(Q > \chi^2_{N,\alpha}) = \alpha$. For any nonnegative α_1, α_2 such that $\alpha_1 + \alpha_2 = \alpha$, we have

$$\begin{aligned}
1-\alpha &= 1-\alpha_1-\alpha_2 \\
&= P\left(\frac{(n-1)S^2}{\sigma^2} \geq \chi^2_{n-1,1-\alpha_1}\right) - P\left(\frac{(n-1)S^2}{\sigma^2} > \chi^2_{n-1,\alpha_2}\right) \\
&= 1 - P\left(\frac{(n-1)S^2}{\sigma^2} < \chi^2_{n-1,1-\alpha_1}\right) - P\left(\frac{(n-1)S^2}{\sigma^2} > \chi^2_{n-1,\alpha_2}\right) \\
&= 1 - P\left(\frac{(n-1)S^2}{\sigma^2} < \chi^2_{n-1,1-\alpha_1} \text{ or } \frac{(n-1)S^2}{\sigma^2} > \chi^2_{n-1,\alpha_2}\right) \\
&= P_{(\mu,\sigma^2)}\left(\chi^2_{n-1,1-\alpha_1} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{n-1,\alpha_2}\right) \\
&= P_{(\mu,\sigma^2)}\left(\frac{(n-1)S^2}{\chi^2_{n-1,\alpha_2}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha_1}}\right).
\end{aligned}$$

- *Answer to Example L16.3 continued*: This shows that

$$C(\boldsymbol{X}) = \left\{ \sigma^2 : \sigma^2 \in \left[\frac{(n-1)S^2}{\chi^2_{n-1,\alpha_2}}, \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha_1}} \right] \right\}$$

  is a $100(1-\alpha)\%$ confidence interval for σ^2.
- (c) By *Theorem L16.1*,

$$\begin{aligned} A(\sigma_0^2) &= \left\{ \boldsymbol{x} : \sigma_0^2 \in C(\boldsymbol{x}) \right\} \\ &= \left\{ \boldsymbol{x} : \chi^2_{n-1,1-\alpha_1} \leq \frac{(n-1)s^2}{\sigma_0^2} \leq \chi^2_{n-1,\alpha_2} \right\} \end{aligned}$$

  is the acceptance region of a level α test of $H_0 : \sigma^2 = \sigma_0^2$.

- *Example L16.4*: Suppose $X_1, \ldots, X_n$ are iid random variables from a Pareto distribution with pdf

$$f(x|\theta) = \frac{\theta}{x^{\theta+1}} I_{(1,\infty)}(x)$$

where $\theta > 0$.
(a) Let $Y = \max\{X_1, \ldots, X_n\}$. Find the cdf $F_Y(y) = P(Y \leq y)$ for $y > 1$.
(b) Find the inverse of the cdf $F_Y^{-1}(p)$ for $p \in (0, 1)$.
(c) Show that $F_Y(Y)$ is a pivot.
(d) Use the pivot in part (c) to find a $100(1-\alpha)\%$ confidence interval for θ with the form $[\theta_L, \infty)$.

Using a Pivotal Quantity

- *Answer to Example L16.4*: (a) The cdf of Y is

$$\begin{aligned}F_Y(y) &= P(X_1 \leq y, X_2 \leq y, \ldots, X_n \leq y) \\ &= \prod_{i=1}^{n} P(X_i \leq y) \\ &= \left(\int_1^y \theta x^{-\theta-1} dx \right)^n \\ &= \left(\left[-x^{-\theta} \right]_1^y \right)^n = (1 - y^{-\theta})^n.\end{aligned}$$

- (b) The inverse of this function is $F_Y^{-1}(p) = (1 - \sqrt[n]{p})^{-1/\theta}$ since

$$(1 - y^{-\theta})^n = p \Rightarrow 1 - y^{-\theta} = \sqrt[n]{p} \Rightarrow y = (1 - \sqrt[n]{p})^{-1/\theta}.$$

Note that $F_Y^{-1}(p)$ is an increasing function of p.

Using a Pivotal Quantity

- *Answer to Example L16.4 continued*:
  (c) $F_Y(Y) \sim$ Uniform$(0,1)$ since

$$
\begin{aligned}
P(F_Y(Y) \leq y) &= P(Y \leq F_Y^{-1}(y)) \\
&= P(Y \leq (1 - \sqrt[n]{y})^{-1/\theta}) \\
&= (1 - ((1 - \sqrt[n]{y})^{-1/\theta})^{-\theta})^n \\
&= (1 - (1 - \sqrt[n]{y}))^n \\
&= (\sqrt[n]{y})^n = y
\end{aligned}
$$

  So, it is a pivot since the distribution of $F_Y(Y)$ has the same distribution for all θ.

- *Answer to Example L16.4 continued*:
  (d) The $100(1-\alpha)\%$ confidence interval $[0, -\ln(1-\sqrt[n]{1-\alpha})/\ln Y]$ for θ can be obtained as follows.

$$
\begin{aligned}
1-\alpha &= P(F_Y(Y) \leq 1-\alpha) \\
&= P((1-Y^{-\theta})^n \leq 1-\alpha) \\
&= P(1-Y^{-\theta} \leq \sqrt[n]{1-\alpha}) \\
&= P(Y^{-\theta} \geq 1-\sqrt[n]{1-\alpha}) \\
&= P(-\theta \ln Y \geq \ln(1-\sqrt[n]{1-\alpha})) \\
&= P\left(\theta \leq \frac{-\ln(1-\sqrt[n]{1-\alpha})}{\ln Y}\right)
\end{aligned}
$$

Finding a Bayesian Credible Interval

- Interval estimators are much different from a Bayesian perspective where the parameter is considered random.
- *Definition L16.5* (p.436): If $\pi(\theta|\boldsymbol{x})$ is the posterior distribution of θ given $\boldsymbol{X} = \boldsymbol{x}$, then for any set $A \subset \Theta$, the credible probability of A is

$$P(\theta \in A|\boldsymbol{x}) = \int_A \pi(\theta|\boldsymbol{x})\ d\theta \quad \text{(assuming } \theta|\boldsymbol{x} \text{ is continuous)}$$

and A is a *credible set* for θ.

Finding a Bayesian Credible Interval

- *Example L16.5*: Suppose $X_1, \ldots, X_n$ are iid Bernoulli(p) random variables and suppose we consider a Bayesian model which assumes that p follows a Uniform$(0, 1)$ prior.
  Find a 90% credible set for p for the data set with 4 successes and 14 failures.

- *Answer to Example L16.5*:
  From slide 7.23, $p|\sum_{i=1}^{n} X_i = y \sim$ beta$(y+\alpha, n-y+\beta)$, we have $p|\boldsymbol{X} = \boldsymbol{x} \sim$ beta$(4+1=5, 14+1=15)$.
  So we can find p_L such that $\int_0^{p_L} \pi(p|\boldsymbol{x})\ d\theta = .05$ and p_U such that $\int_{p_U}^{1} \pi(p|\boldsymbol{x})\ d\theta = .05$ where $\pi(p|\boldsymbol{x}) = 58140 p^4 (1-p)^{14}$.
- Using the R commands `qbeta(.05,5,15)` and `qbeta(.95,5,15)`, we obtain the 90% credible set $(.1099, .4191)$. The shortest 90% credible set $(.0953, .3991)$ can be obtained with the R commands `alpha=.02931685;qbeta(c(alpha,.9+alpha),5,15)` since

```
> dbeta(qbeta(c(alpha,.9+alpha),5,15),5,15)
[1] 1.180588 1.180588
```