To Do

Read Sections 4.3 - 4.5.

Do End-of-Chapter Problems 1-17 in preparation for Tutorial Test 2.

Important Ideas from Last Day

(1) Interval Estimation Using the Relative Likelihood Function

(2) Interval Estimation Using the Log Relative Likelihood Function

(3) Definition of Coverage Probability

(4) Interval Estimation Using Confidence Intervals

(5) Interpretation of a Confidence Interval

Definition of a Confidence Interval

Definition

A 100*p*% confidence interval for a parameter is an interval estimate $[L(y), U(y)]$ for which

$$P(\theta \in [L(Y), U(Y)]) = P[L(Y) \le \theta \le U(Y)] = p.$$

The value *p* is called the confidence coefficient for the confidence interval.

Interpretation of a Confidence Interval

Suppose

$$0.95 = P(\theta \in [L(Y), U(Y)]) = P[L(Y) \le \theta \le U(Y)]. \quad (1)$$

Suppose also that we draw repeated independent random samples from the same population and each time we construct the interval [L(y), U(y)] based on the observed data y.

Then (1) tells us that we should expect 95% of these constructed intervals to contain the true but unknown value of θ.

We also expect that 5% of these constructed intervals will not contain the true but unknown value of θ.

Interpretation of a Confidence Interval

Of course we usually only construct one such interval based on one set of data.

We hope that we are one of the lucky 95%!

For a particular sample we say that we are 95% confident that the true value of $\boldsymbol{\theta}$ is contained in the interval we have constructed.

Today's Lecture

(1) Definition of a Pivotal Quantity

(2) How to Use a Pivotal Quantity to Construct a Confidence Interval

(3) Approximate Pivotal Quantities

(4) Approximate Confidence Intervals for Binomial

Example from Last Day: 95% Confidence Interval for the Mean μ of a Gaussian Population, Known Standard Deviation σ

Suppose $Y_1, Y_2, \ldots, Y_n$ is a random sample from a G(μ,1) distribution where E(Y_i) = μ is unknown but sd(Y_i) = 1 is known. Since

$$P\left(\mu \in \left[\bar{Y} - \frac{1.96}{\sqrt{n}}, \bar{Y} + \frac{1.96}{\sqrt{n}} \right] \right) = 0.95$$

therefore

$$\left[\bar{y} - \frac{1.96}{\sqrt{n}}, \bar{y} + \frac{1.96}{\sqrt{n}} \right] \text{ or } \bar{y} \pm \frac{1.96}{\sqrt{n}}$$

is a 95% confidence interval for μ.

Confidence Intervals and Pivotal Quantities

Note that the statement

$$P\left(\mu \in \left[\overline{Y} - \frac{1.96}{\sqrt{n}}, \overline{Y} + \frac{1.96}{\sqrt{n}} \right] \right) = 0.95$$

holds no matter what the true value of *μ* is.

We now consider a general method for constructing confidence intervals which have this very useful property.

Definition of a Pivotal Quantity

Definition

A pivotal quantity $Q = Q(Y;\theta)$ is a function of the data Y and the unknown parameter θ such that the distribution of the random variable Q is completely known.

Note: This definition implies that probability statements such as $P(Q \le a)$ and $P(Q \ge b)$ depend on a and b but not on θ or any other unknown information.

Example

Suppose $Y_1, Y_2, \ldots, Y_n$ is a random sample from a G(μ,σ) distribution where E(Y_i) = μ is unknown but sd(Y_i) = σ is known.

A point estimator for μ is $\tilde{\mu} = \bar{Y}$ and the sampling distribution of this estimator is

$$\tilde{\mu} = \bar{Y} \sim G\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

Example Cont'd

Since

$$Q = Q(Y;\mu) = \frac{\bar{Y} - \mu}{\sigma / \sqrt{n}} \sim G(0,1)$$

has a completely known distribution, Q is a pivotal quantity.

Pivotal quantities can be used for constructing confidence intervals.

How to Use a Pivotal Quantity to Construct a Confidence Interval

(1) Determine numbers a and b such that
$P[a \le Q(Y;\theta) \le b] = p$.
(Why are you able to do this?)

(2) Re-express the inequality $a \le Q(Y;\theta) \le b$ in the form $L(Y) \le \theta \le U(Y)$, then

$$p = P[a \le Q(Y;\theta) \le b\ ] = P[L(Y) \le \theta \le U(Y)\ ]$$

so the coverage probability equals p.

(3) For observed data y, the interval $[L(y),U(y)]$ is a $100p\%$ confidence interval for θ.

How to Construct a 95% Confidence Interval for the Mean μ of a Gaussian Population, Known σ

(1) Find the value *a* from Normal tables such that

$$0.95 = P\left(-a \le \frac{\bar{Y} - \mu}{\sigma / \sqrt{n}} \le a \right)$$

(2) Solve the inequality for μ:

$$-a \le \frac{\bar{Y} - \mu}{\sigma / \sqrt{n}} \le a \;\text{ iff }\; \bar{Y} - a\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{Y} + a\frac{\sigma}{\sqrt{n}}$$

(3) A 95% confidence interval for μ based on the observed data $y_1, y_2, \ldots, y_n$ is $\left[\bar{y} - a\frac{\sigma}{\sqrt{n}}, \bar{y} + a\frac{\sigma}{\sqrt{n}} \right]$

95% Confidence Interval for the Mean μ of a Gaussian Population, Known σ

There are an infinite number of values *a* and *b* are such that

$$0.95 = P\left(a \leq \frac{\bar{Y} - \mu}{\sigma / \sqrt{n}} \leq b \right)$$

The interval

$$\left[\bar{y} - b\frac{\sigma}{\sqrt{n}}, \bar{y} - a\frac{\sigma}{\sqrt{n}} \right]$$

is also a 95% confidence interval.

Since the Gaussian distribution is symmetric about its mean we use *a* = -1.96 and *b* = 1.96 which gives the narrowest confidence interval.

95% Confidence Interval for the Mean μ of a Gaussian Population, Known σ

Note: **The confidence interval**

$$\left[\bar{y} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{y} - 1.96\frac{\sigma}{\sqrt{n}}\right] \text{ or } \bar{y} \pm 1.96\frac{\sigma}{\sqrt{n}}$$

is symmetric about the point estimate of *μ*, $\hat{\mu} = \bar{y}.$

100*p*% Confidence Interval for the Mean *μ* of a Gaussian Population, Known *σ*

How to construct a 100*p*% confidence interval for *μ* for Gaussian data when *σ* is known (short version):

Use the Normal table to find the value *a* such that P(-*a* ≤ *Z* ≤ *a*) = *p* where *Z* ~ G(0,1) or equivalently $P(Z \le a) = \frac{1+p}{2}$.

A 100*p*% confidence interval for *μ* is:

$$\bar{y} \pm a\frac{\sigma}{\sqrt{n}}$$

Useful Numbers

For a 100p% confidence interval for μ we find the value a in the Normal table such that $P(-a \le Z \le a) = p$ or $P(Z \le a) = (1+p)/2$ where $Z \sim G(0,1)$.

For a 90% confidence interval, $p = 0.90$, $(1+p)/2 = 0.95$ and $a = 1.645$.

For a 95% confidence interval, $p = 0.95$, $(1+p)/2 = 0.975$ and $a = 1.960$.

For a 99% confidence interval, $p = 0.99$, $(1+p)/2 = 0.995$ and $a = 2.576$.

“Two-sided, Equal-tailed” Confidence Intervals

A 100p% confidence interval for μ is of the form:

point estimate ± (table value) × sd(estimator)

Such an interval is often called a “two-sided, equal-tailed” confidence interval.

We will encounter other examples of two-sided, equal-tailed confidence intervals in this course.

Approximate Pivotal Quantities

For most statistical models it is not possible to find "exact" pivotal quantities or confidence intervals for θ.

Fortunately, we can often find random variables $Q_n = Q_n(Y_1, Y_2, \ldots, Y_n; \theta)$ such that as $n \to \infty$, the distribution of Q_n ceases to depend on θ or other unknown information.

We call Q_n an asymptotic or approximate pivotal quantity.

Approximate Confidence Intervals for Binomial

For a Binomial experiment, *Y* ~ Binomial(n,*θ*) and the point estimator of *θ* is

$$\tilde{\theta} = \frac{Y}{n}$$

For large n, the approximate sampling distribution of $\tilde{\theta} = \frac{Y}{n}$ **is**

$$\frac{\tilde{\theta} - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \sim G(0,1) \ \text{ approximately}$$

by the Central Limit Theorem.

Approximate Confidence Intervals for Binomial

It can be also shown for large *n* that

$$Q_n = \frac{\tilde{\theta} - \theta}{\sqrt{\dfrac{\tilde{\theta}\left(1 - \tilde{\theta}\right)}{n}}} \sim G(0,1) \quad \text{approximately.}$$

***Q_n* is an approximate pivotal quantity which can be used to construct approximate confidence intervals for *θ*.**

Approximate Confidence Intervals for Binomial

Since

$$0.95 \approx P\left(-1.96 \le \frac{\tilde{\theta}-\theta}{\sqrt{\frac{\tilde{\theta}\left(1-\tilde{\theta}\right)}{n}}} \le 1.96\right)$$

$$= P\left(\tilde{\theta}-1.96\sqrt{\frac{\tilde{\theta}\left(1-\tilde{\theta}\right)}{n}} \le \theta \le \tilde{\theta}+1.96\sqrt{\frac{\tilde{\theta}\left(1-\tilde{\theta}\right)}{n}}\right)$$

therefore an approximate 95% confidence interval for θ is given by

$$\hat{\theta} \pm 1.96\sqrt{\frac{\hat{\theta}\left(1-\hat{\theta}\right)}{n}}.$$