STA261: Week 8

Confidence Intervals & Hypothesis Testing II

Alex Stringer

March 5th - 9th, 2018

Disclaimer

The materials in these slides are intended to be a companion to the course textbook, *Mathematical Statistics and Data Analysis, Third Edition*, by John A Rice. Material in the slides may or may not be taken directly from this source. These slides were organized and typeset by Alex Stringer.

A big thanks to Jerry Brunner as well for providing inspiration for assignment questions.

License

Recap

This week we will continue last week's discussion of

- Confidence Intervals
- Hypothesis Tests

We will use the CLT for the MLE that we derived in week 4, and we will extend the discussion to the case of comparing two samples.

CLT for the MLE

In lecture 4, we derived the following result. Let $X_i \sim F_\theta, i = 1 \ldots n$ be an IID random sample with distribution function F indexed by parameter θ with true value θ_0. Let $\hat{\theta} = \text{argmax}_\theta \ell(\theta)$ where $\ell(\theta) = \log f(\mathbf{x}|\theta) = \sum_{i=1}^n \log f(x_i|\theta)$ be the maximum likelihood estimator. Define

$$I(\theta) = -E\left(\frac{\partial^2 \ell(\theta)}{\partial \theta^2}\right)$$

to be the Fisher Information. Then under regularity conditions (described in lecture 4), as $n \to \infty$,

$$\sqrt{I(\theta_0)}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, 1)$$

CLT for the MLE

Corollary: an approximate, large-sample $1-\alpha$ Confidence Interval for θ is

$$\left(\hat{\theta} - \frac{1}{\sqrt{I(\hat{\theta})}} z_{1-\alpha/2}, \hat{\theta} + \frac{1}{\sqrt{I(\hat{\theta})}} z_{1-\alpha/2}\right)$$

where we have replaced θ_0 with a consistent estimator, $\hat{\theta}$, in the Fisher Information (on assignment 8 you will investigate the effect of this in small samples).

Example

Let $X_i \sim Bern(\theta_0)$. Derive a $1-\alpha$ confidence interval for θ.

We did this before using specific distributional information; let's do it now using MLE theory.

The MLE is $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

The Fisher Information is $I(\theta) = \frac{n}{\theta(1-\theta)}$

Example

A $1-\alpha$ confidence interval for θ is then

$$\left(\hat{\theta}-\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}z_{1-\alpha/2}, \hat{\theta}+\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}z_{1-\alpha/2}\right)$$

Last lecture, we threw a coin 10 times and got 7 heads, and we decided that this did not give sufficient evidence to reject $H_0 : \theta = 0.5$. We can compute an approximate 95% confidence interval for θ, given $n = 10$ and $\hat{\theta} = 0.7$:

$$(0.42, 0.98)$$

The fact that this interval contains $\theta = 0.5$ agrees with our previous inference.

Monotone Transformations

Because confidence intervals are just intervals, we can stretch/shrink them using monotone transformations.

This gives us an easy way to get confidence intervals for transformations of our parameters of interest.

Suppose we have a $1-\alpha$ CI for some parameter θ:

$$(L, U)$$

If $g(\cdot)$ is *monotonic increasing*, then a $1-\alpha$ CI for $g(\theta)$ is given by

$$(g(L), g(U))$$

If $g(\cdot)$ is *monotonic decreasing*, then a $1-\alpha$ CI for $g(\theta)$ is given by

$$(g(U), g(L))$$

Example

In the previous example, get a 95% confidence interval for $n\theta$, the expected number of heads in n throws (n fixed).

Our interval is

$$\left(n\hat{\theta} - n\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} z_{1-\alpha/2}, n\hat{\theta} + n\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} z_{1-\alpha/2} \right)$$
$$= (4.16, 9.84)$$

We just multiply the original interval by 10.

Example

If $X_i \sim N(\mu, \sigma^2)$ then on assignment 7 you showed that a $1 - \alpha$ CI for σ^2 is

$$\left(\frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}} \right)$$

where

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

and $\chi^2_{n-1,1-\alpha/2}$ refers to the $1 - \alpha/2$ quantile of a χ^2_{n-1} distribution.

That's for the variance σ^2- can we find a confidence interval for the standard deviation, σ?

Example

Yep- a $1-\alpha$ CI for $\sigma = \sqrt{\sigma^2}$ is

$$\left(\sqrt{\frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}}, \sqrt{\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}}\right)$$

A $1-\alpha$ CI for $\log\sigma^2$ would be

$$\left(\log\frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}, \log\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}\right)$$

Example

A $1-\alpha$ CI for the precision, $\psi = 1/\sigma^2$, would be

$$\left(\left(\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}} \right)^{-1}, \left(\frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}} \right)^{-1} \right)$$

Note that we switched the quantiles in the interval (flipped the interval around) because $g(x) = 1/x$ is a decreasing, not increasing, function.

Two Samples (textbook, chapter 11, sections 11.1, 11.2.1, 11.3)

Up until now we have been dealing with a single sample from a single distribution. This lets us answer some questions in science, like “is Kellogs lying about the weight of boxes of Raisin Bran?” (they aren't; see assignment 7) and “does this poll provide evidence that candidate A has a popularity above 50%?”.

What is also common in practice are questions of the form “does the mean measurement of X differ between group A and group B?”. For example, do patients in the treatment group differ in some clinical measurement vs patients in the control group?

These are **two-sample** problems.

Two Samples

Suppose $X_i \sim N(\mu_x, \sigma^2)$ and $Y_i \sim N(\mu_y, \sigma^2), i = 1 \dots n$ are mutually independent random samples (each (X_i, Y_i) pair is independent as well as all the X_i and Y_i being independent from each other) from Normally distributed populations with potentially different means, but equal variances. We are interested in testing

$$H_0 : \mu_x = \mu_y$$

and finding a confidence interval for the mean difference (or difference in means, same thing)

$$d = \mu_x - \mu_y$$

Two Samples

We know that

$$\bar{X} \sim N(\mu_x, \sigma^2/n)$$
$$\bar{Y} \sim N(\mu_y, \sigma^2/n)$$

Because the X_i and Y_i are all mutually independent, so are $\bar{X}$ and $\bar{Y}$, and so they are jointly normally distributed. This means that

$$\bar{X} - \bar{Y} \sim N\left(\mu_x - \mu_y, \frac{2\sigma^2}{n}\right)$$

Two Samples

It follows that a suitable pivot for the mean difference $d = \mu_x - \mu_y$ is

$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sigma\sqrt{2/n}} \sim N(0, 1)$$

and a $1 - \alpha$ confidence interval for d is

$$\left(\bar{X} - \bar{Y} - \sigma\sqrt{2/n}z_{1-\alpha/2}, \bar{X} - \bar{Y} + \sigma\sqrt{2/n}z_{1-\alpha/2}\right)$$

Two Samples: unknown variance

Like with the one-sample case, in practice we don't know σ^2. We estimate this using the sample variance. In lecture 7 we defined the t-distribution, and showed that this is the distribution of the modified test statistic in which this replacement is made; the same result holds here, and we have

$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{s\sqrt{2/n}} \sim t_{2(n-1)}$$

where we estimate s^2 by pooling the variance from the two groups,

$$s^2 = \frac{s_x^2 + s_y^2}{2}$$

Two Samples: unknown variance

A $1-\alpha$ confidence interval for d is

$$\left(\bar{X}-\bar{Y}-s\sqrt{2/n}\times t_{2(n-1),1-\alpha/2}, \bar{X}-\bar{Y}+s\sqrt{2/n}\times t_{2(n-1),1-\alpha/2}\right)$$

In the hypothesis testing scenario, we would reject $H_0 : d = 0$ if

$$\frac{\left|\bar{X}-\bar{Y}\right|}{s\sqrt{2/n}} > t_{2(n-1),1-\alpha/2}$$

Example

Suppose a small clinical trial is run for a new cholesterol medication. Patients are assigned into treatment (X) and control (Y) groups, with $n = 10$ patients in each.

The question is whether the mean of group X's cholesterol measurement is different from group Y's.

Example

We perform the trial and get the following summary statistics:

$$\begin{aligned}
\bar{X} &= 8.2 \\
\bar{Y} &= 9.6 \\
\hat{d} &= 8.2 - 9.6 = -1.4 \\
s_x^2 &= 2.1 \\
s_y^2 &= 2.9 \\
s^2 &= \frac{2.1 + 2.9}{2} = 2.5
\end{aligned}$$

Is an observed difference of -1.4 enough to conclude that the medication actually lowers cholesterol, at this level of observed variability?

Example

The test statistic is

$$t(\mathbf{x}) = \frac{\left|\hat{d} - 0\right|}{s\sqrt{2/n}} = \frac{|-1.4|}{\sqrt{2.5 \times 2/10}} = 1.98$$

The corresponding p-value is

$$p_0 = 2 \times (1 - P_T(\left|t(\mathbf{x})\right|)) = 0.06$$

so we do not reject H_0 at the $\alpha = 0.05$ significance level.

Example

The corresponding confidence interval for the mean difference is

$$\begin{aligned}
&\left(\hat{d} - s\sqrt{2/n} \times t_{2(n-1),1-\alpha/2}, \hat{d} + s\sqrt{2/n} \times t_{2(n-1),1-\alpha/2}\right) \\
&= \left(-1.4 - \sqrt{2.5 * 2/10} \times 2.1, -1.4 + \sqrt{2.5 * 2/10} \times 2.1\right) \\
&= (-2.89, 0.09)
\end{aligned}$$

which contains 0. Since we conclude that 0 is a plausible value of $\mu_x - \mu_y$ given the observed data, we conclude that the current study did not demonstrate that the treatment was effective, at the $\alpha = 0.05$ significance level.

Watch the language. We did *not* conclude that the treatment is ineffective.

Unequal Group Sizes

You don't need to make the two groups the same size for a two-sample t-test based on independent samples. Textbook section 11.2.1, and Assignment 8, give the details of the calculations when the group sizes are unequal.

We did, though, have to assume equal variances. This is a very common assumption, although that doesn't mean it's a good one.

We won't go further, but textbook page 428 talks briefly about the unequal variances case.

Paired Samples

Often in scientific practice, to save time and money, and increase the precision of the associated tests (more on that in lecture 10), studies that aim to compare two groups are conducted in a manner in which the observations are *paired*.

For example, cholesterol measurements could be taken on each patient before, and after, the treatment was applied. Patients act as their own control.

This is reasonable, and often a very clever thing to do, but it does break our independence assumption.

How to proceed?

Paired Samples

Suppose

$$(X_i, Y_i) \sim N\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \sigma_x\sigma_y\rho \\ \sigma_x\sigma_y\rho & \sigma_y^2 \end{pmatrix} \right)$$

independently, $i = 1 \ldots n$. This notation means that each X_i, Y_i pair follows a bivariate normal distribution with

$$\begin{aligned} E(X_i) &= \mu_x \\ E(Y_i) &= \mu_y \\ Var(X_i) &= \sigma_x^2 \\ Var(Y_i) &= \sigma_y^2 \\ Cor(X_i, Y_i) &= \rho \neq 0 \end{aligned}$$

and that each (X_i, Y_i) pair is independent of the other pairs.

Paired Samples

We wish to perform inference on only the difference between means, $d = \mu_x - \mu_y$. To that end, note that

$$D_i = X_i - Y_i \sim N(\mu_x - \mu_y, \sigma_d^2)$$

where σ_d^2 depends on σ_x^2, σ_y^2, and ρ.

It doesn't matter though- we don't need to estimate these quantites.

Paired Samples

To perform inference for paired samples,

- Calculate the sample differences, $d_i = x_i - y_i, i = 1 \dots n$
- Perform 1-sample inference on d_i

That is, to test $H_0 : \mu_x = \mu_y$ for a paired sample, simply test $H_0 : \mu_d = 0$ using

$$\frac{\hat{d}}{s_d/\sqrt{n}} \sim t_{n-1,1-\alpha/2}$$

where

$$\hat{d} = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)$$

$$s_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}((x_i - y_i) - (\bar{x} - \bar{y}))^2$$

are the sample mean and variance of the differences $d_i = x_i - y_i$.

Paired Samples

It follows that a $1 - \alpha$ CI for $\mu_x - \mu_y$ in this case is given by

$$\left(\bar{X} - \bar{Y} - \frac{s_d}{\sqrt{n}} \times t_{n-1,1-\alpha/2}, \bar{X} - \bar{Y} + \frac{s_d}{\sqrt{n}} \times t_{n-1,1-\alpha/2} \right)$$

Example

In our previous example, suppose we only sampled $n = 10$ patients total, and we took their cholesterol measurements before and after treatment. Suppose we observe equivalent summary statistics:

$$\hat{d} = -1.4$$
$$s_d^2 = 2.5$$

Note: it is **not** the case in general that the pooled variance from an independent samples t-test will be the same as the variance of the differences from a paired t-test. I made the example this way intentionally.

Example

This would give a test statistic of

$$t(\mathbf{x}) = \frac{\left|\hat{d}\right|}{s_d/\sqrt{n}} = 2.8$$

with corresponding p-value

$$p_0 = 0.02$$

so we would now reject $H_0 : \mu_x = \mu_y$ at the same significance level as before.

And we only had to sample half the number of patients.

Example

The corresponding confidence interval is

$$\left(\hat{d} - \frac{s_d}{\sqrt{n}} \times t_{n-1,1-\alpha/2}, \hat{d} + \frac{s_d}{\sqrt{n}} \times t_{n-1,1-\alpha/2}\right)$$
$$= (-2.53, -0.27)$$

which is much narrower than before.

Our inference in the paired case is more precise.