STA261: Week 7

Confidence Intervals & Hypothesis Testing I

Alex Stringer

Feb 26th - March 2nd, 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

Summary so far

So far, we have talked about

- Probability vs Statistics, the Inference problem
- Inference for parameters from known families of distributions
- Finding a function of the data (estimator) that gives us “good” estimates of parameters
- What “good” means

We basically introduced a new concept every week.

Going Forward

We will now switch direction a bit. The rest of the course is going to be focussed on

- Confidence Intervals
- Hypothesis Tests

We will introduce both concepts today, then study them in increasing generality for the next 5 weeks.

Going Forward

Specifically, here is the tentative lecture schedule for the rest of the course:

- **Lecture 7**: Confidence Intervals & Hypothesis Testing I
- **Lecture 8**: Confidence Intervals & Hypothesis Testing II
- **Lecture 9**: Likelihood Ratio Tests
- **Lecture 10**: Power and Sample Size Calculations
- **Lecture 11**: Computational Methods & The Bootstrap
- **Lecture 12**: Exam Review

Summary so far

So far, we have introduced the concepts of

- **Consistency**: more and more data, estimates should get "closer and closer" to true value
- **Sufficiency**: formalize the notion of information about parameter in the sample, then choose an estimator that uses all this information
- **Unbiasedness**: on average, across all possible datasets, our estimation procedure should give the right answer
- **Efficiency**: given that we want an unbiased estimator, we might as well pick the one with the lowest variance

Summary so far

Seems reasonable. But have we solved our problem?

Let's say we are faced with an actual sample, and we believe it came from a $N(\mu_0, 1)$ distribution.

- Find an estimate of μ based off of an estimator that is consistent, sufficient, unbiased, and efficent (no problem)
- Evaluate its standard deviation using the data (okay)

Example

Let's say $n = 5$ and we get $\mathbf{x} = (-0.89, -0.23, -0.65, -0.42, 0.21)$

Then we use $\bar{X}$ to estimate μ. We know that

$$\bar{X} \sim N\left(\mu_0, \frac{1}{\sqrt{5}}\right)$$

Example

Calculate $\bar{x} = -0.4$, $SD(\bar{X}) = 0.45$.

Our estimate, -0.4, is one realization of a $N(\mu_0, 0.45)$ random variable.

How do we use this idea to make an inference about μ? Is it good enough to just say $\mu_0 = -0.4$ and be done with it?

Plausible Values

We interpreted the MLE as the value that most plausibly generated the data we observed.

In this example, the most plausible value of μ is thus -0.4.

But there's variability in the sample. If we observed a different sample, we'd get a different $\bar{x}$.

Plausible Values

What about $\mu = -0.41$? Is this plausible, given the observed data?

What about $\mu = -0.5$? Could this value of μ have generated the data we saw?

What about $\mu = -10$?

There must, somehow, be a **range** of plausible values for μ, based on the observed data.

Confidence

Can we use the observed data to systematically generate a **range of plausible values** for the parameter?

Plausible means “could plausibly have generated the data we observed”.

We will look for an interval

$$C(\mathbf{X}) = (L(\mathbf{X}), U(\mathbf{X}))$$

that contains μ_0 with high probability.

If we find $L(\mathbf{X})$ and $U(\mathbf{X})$ in this way, we can say that this interval represents a range of values of μ that could plausibly have generated the data we observed.

Pivots

How can we measure probability if we don't know the true parameter value?

We introduce the concept of a **pivot**.

Definition: a **pivot** for parameter θ is a random variable that depends on the unknown parameter value θ_0, but has a known distribution that does not depend on θ_0.

Pivots

Example: for the previous example, find a pivot for μ.

We have $\bar{X} \sim N\left(\mu_0, \frac{1}{\sqrt{5}}\right)$. Using properties of the normal distribution, we can write

$$Z = \frac{\bar{X} - \mu_0}{1/\sqrt{5}} \sim N(0,1)$$

Z is a pivot for μ, because it depends on the true value μ_0, but has a known distribution $(N(0,1))$ which does not depend on μ_0.

Pivots

Example: for $X \sim N(0, \sigma_0^2)$, find a pivot for σ^2.

We will prove later in this lecture that

$$\frac{ns^2}{\sigma_0^2} \sim \chi_n^2$$

so $\frac{ns^2}{\sigma_0^2}$ is a pivot for σ^2.

Confidence Intervals

We can use Z to find our desired interval. We wish to find an interval

$$C(\mathbf{X}) = \left(L(\mathbf{X}), U(\mathbf{X})\right)$$

such that

$$P\left(L(\mathbf{X}) \leq \mu_0 \leq U(\mathbf{X})\right) = 1 - \alpha$$

for some $0 < \alpha < 0.5$. We call this a $1 - \alpha$ *confidence interval for* μ.

Most often in scientific applications $\alpha = 0.05$, in which case you hear the term 95% *confidence interval*.

Confidence Intervals

We have (dropping the notational dependence on $\mathbf{X}$):

$$\begin{aligned}1-\alpha &= P\left(L \le \mu_0 \le U\right)\\ &= P(\bar{X}-U \le \bar{X}-\mu_0 \le \bar{X}-L)\\ &= P\left(\frac{\bar{X}-U}{1/\sqrt{5}} \le \frac{\bar{X}-\mu_0}{1/\sqrt{5}} \le \frac{\bar{X}-L}{1/\sqrt{5}}\right)\\ &= P\left(\frac{\bar{X}-U}{1/\sqrt{5}} \le Z \le \frac{\bar{X}-L}{1/\sqrt{5}}\right)\end{aligned}$$

Confidence Intervals

But we know the distribution of Z, it's standard normal, so

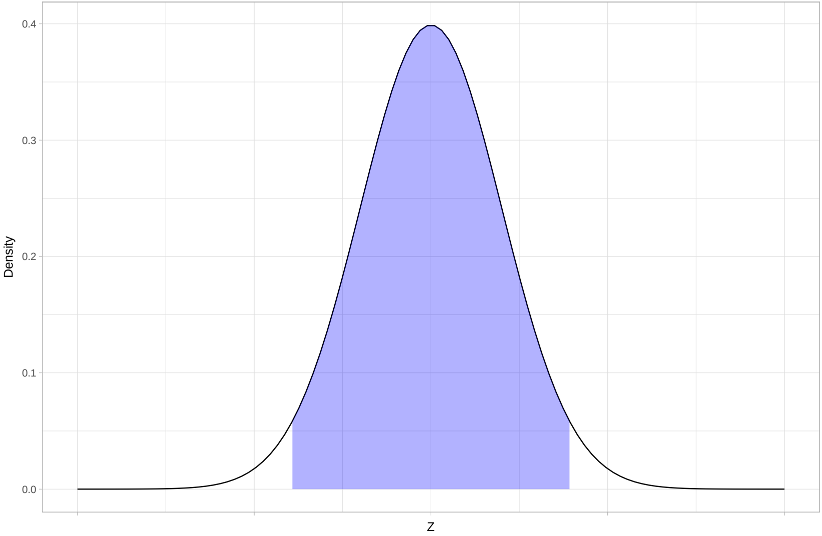
$$P\left(\frac{\bar{X}-U}{1/\sqrt{5}} \leq Z \leq \frac{\bar{X}-L}{1/\sqrt{5}}\right)$$

allows us to choose

$$\frac{\bar{X}-U}{1/\sqrt{5}} = z_{\alpha/2}$$
$$\frac{\bar{X}-L}{1/\sqrt{5}} = z_{1-\alpha/2}$$

where $z_{\alpha/2}$ is the $\alpha/2$ *quantile* of the standard normal distribution.

Confidence Intervals



Standard Normal Density

Shaded region has probability approximately 95%. Unshaded regions each have probability approximately 2.5%

Confidence Intervals

Due to the symmetry of the normal distribution, $z_{1-\alpha/2} = -z_{\alpha/2}$. Also because we assumed $\alpha < 0.5$, we know $z_{1-\alpha/2} > 0$.

Rearranging, we find that

$$\begin{aligned} L &= \bar{X} - \frac{1}{\sqrt{n}} z_{1-\alpha/2} \\ U &= \bar{X} + \frac{1}{\sqrt{n}} z_{1-\alpha/2} \end{aligned}$$

Confidence Interval for μ, known σ

Corollary: If $X_i \sim N(\mu_0, \sigma_0^2)$ is an IID sample from a normal distribution with unknown mean and *known* variance, then a $1-\alpha$ **confidence interval** for μ is given by

$$\left(\bar{X} - \frac{\sigma_0}{\sqrt{n}} z_{1-\alpha/2}, \bar{X} + \frac{\sigma_0}{\sqrt{n}} z_{1-\alpha/2}\right)$$

where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution, i.e.

$$\Phi(z_{1-\alpha/2}) = P(Z < z_{1-\alpha/2}) = 1 - \alpha/2$$

Standard Normal Quantiles

If $\alpha = 0.05$, then $\alpha/2 = 0.025$, $1 - \alpha/2 = 0.975$, and

$$z_{1-\alpha/2} = 1.96$$

You can find this using the `qnorm` command in `R`:

```
round(qnorm(0.975),2)
```

```
## [1] 1.96
```

```
round(qnorm(0.025),2)
```

```
## [1] -1.96
```

Aside

I recommend you use `R` to compute quantiles when doing the assignments, but you can also use tables or calculators that you find online.

On the exam, you'll get relevant quantiles provided in an easy to read format, because I don't want to spend time teaching you to read normal tables, which you won't use once your courses start *requiring* the use of `R`.

Please don't ask me whether you need to memorize quantiles! You **do not**.

Example

Let's calculate a 95% confidence interval for μ in our example where $X \sim N(\mu_0, 1)$, $n = 5$ and

$$\mathbf{x} = (-0.89, -0.23, -0.65, -0.42, 0.21)$$

We find that

$$\bar{x} = -0.4$$
$$\sigma_0 = 1$$
$$z_{1-\alpha/2} = z_{0.975} = 1.96$$

which gives the 95% confidence interval for μ:

$$(-1.27, 0.48)$$

Interpretation

We then say that based on the observed data, any value of μ between -1.27 and 0.48 seems reasonable.

We say that the probability of this interval containing μ_0 is about 95%.

The language here is very important, especially when working with memebers of the scientific community who do not have formal stats backgrounds. We do *not* say "the probability that μ_0 lies in this interval is 95%", even though that is technically a mathematically correct statement.

We say "the probability that the interval contains μ_0 is 95%".

The *interval* is random. μ_0 is *not random*.

In the News

We hear about this concept somewhat often in the news. Whenever you read about a poll, and you hear "accurate to within plus/minus 2 percentage points, 19 times out of 20", you now know that

- "accurate to within plus/minus 2 percentage points": the width of a confidence interval for the sample proportion p that they are trying to estimate
- "19 times out of 20": the signficance level, in this case $19/20 = 95\%$

Simulation

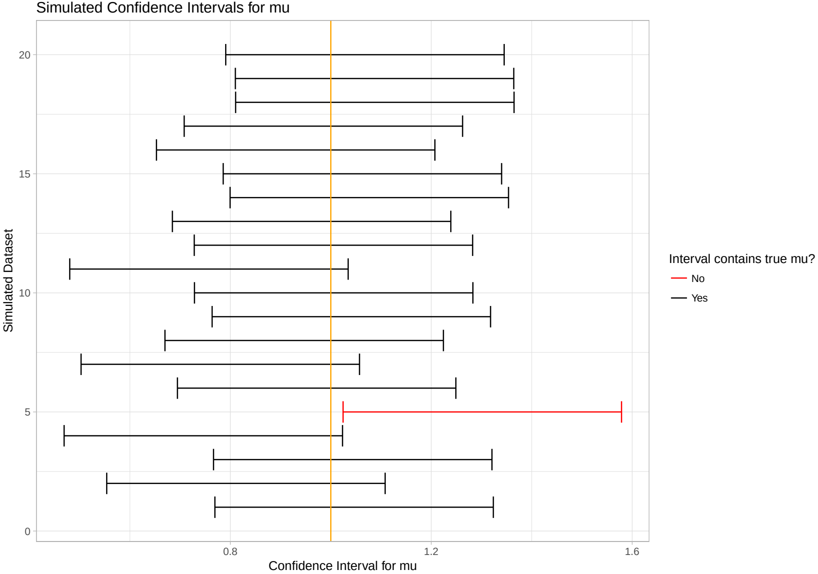
To illustrate the interpretation of a confidence interval, we run a simulation.

Generate a bunch of datasets from a $N(1, 1)$ distribution (so $\mu_0 = 1$), calculate the 95% CI for μ obtained from each dataset, and see how many contain 1.

It should be about $19/20$.

What's important is that μ_0 never changes. Each new dataset gives a different *interval*.

Simulation


Simulated Confidence Intervals for mu
Simulated Dataset
20
15
10
5
0
0.8
1.2
1.6
Confidence Interval for mu
Interval contains true mu?
No
Yes

Testing Whether μ Equals μ_0

Another question we might ask: is any particular value μ_0 supported by the data?

What if we think, say, $\mu_0 = 0$, and we'd like to formally test whether the data says this is plausible.

We could just look, and see whether our hypothesized value μ_0 lies within the confidence interval for μ.

There is another very popular approach to this.

Hypothesis Tests

Suppose we are interested in a particular value of μ, μ_0, and we would like to see whether the **hypothesis** that $\mu = \mu_0$ is supported by our data.

We develop a **Hypothesis Test** to answer this question.

We call the hypothesized value of μ the **null hypothesis**, and write

$$H_0 : \mu = \mu_0$$

We compare this to one or more *alternative* value(s) of μ, which we call the **alternative hypothesis**, and label H_1.

Simple vs Composite Hypotheses

More generally, we define two disjoint subsets of the parameter space Ω: $\Omega_0 \subset \Omega$ and $\Omega_1 \subset \Omega$, and write

$$\begin{aligned} H_0 :& \theta \in \Omega_0 \text{ vs} \\ H_1 :& \theta \in \Omega_1 \end{aligned}$$

If $\Omega_0 = \{\theta_0\}$, a single value, then the null hypothesis is called **simple**. Else, it is called **composite**.

Similarly for the alternative hypothesis.

Most often we test **simple null hypotheses** against **composite alternative hypotheses**.

Simple vs Composite Hypotheses

That is, most often we test

$$H_0 : \theta = \theta_0$$

against

$$H_1 : \theta \neq \theta_0$$

In this very common example, it is the case that $\Omega_0 \cup \Omega_1 = \Omega$, i.e. the two hypotheses partition the parameter space, but this isn't a formal requirement.

Hypothesis Tests

For a given scenario, we decide whether to **reject** the null hypothesis in favour of the alternative, or whether to not.

We **never** “accept” the null hypothesis. Many sources use this language (including our textbook), but in reality, all we can say that “our experiment failed to provide sufficient evidence against the null hypothesis”.

The absence of evidence is not the evidence of absence

Hypothesis Tests

For a given scenario, we decide whether to **reject** the null hypothesis in favour of the alternative, or whether to not.

There are two types of mistakes we could make.

- A **Type I Error**: *reject* the null hypothesis when it is *true*
- A **Type II Error**: *fail to reject* the null hypothesis when it is *false*

Which is worse?

Type I & II Error

The classic analogy is one of a courtroom trial: the defendent is either guilty or not, and the jury must decide whether to convict.

In our legal system, there is no "proven innocent"- we *assume* innocence and prove guilt.

Failing to prove guilt $\implies$ fail to reject H_0.

We *want* to do this if H_0 is true. We don't want to do this if H_0 is false.

But we **definitely** don't want to prove guilt when the defendent is actually innocent; we really don't want to reject a true H_0.

Type I & II Error

Reject a true $H_0 \implies$ send an innocent person to jail.

Fail to reject a false $H_0 \implies$ let a guilty person go free.

We develop our test to have a fixed chosen probability of Type I error, which we call α.

We then find the testing procedure, subject to this, which has the lowest probability of Type II error, which we call β.

$$P(\text{type I error}) = P(\text{reject } H_0 | H_0 \text{ true}) = \alpha$$
$$P(\text{type II error}) = P(\text{fail to reject } H_0 | H_0 \text{ false}) = \beta$$

Decision Rule

We wish to come up with a decision rule, based on the observed data, for deciding whether to reject H_0.

We define a **test statistic** $T(\mathbf{X})$.

We define a **critical region** $R_\alpha(T)$ for this test statistic, such that we reject H_0 if $T(\mathbf{X}) \in R_\alpha(T)$. Usually this will be a subset of $\mathbb{R}$, so it's less abstract than it sounds.

Decision Rule

How to pick the pair $T(\mathbf{X}), R_\alpha(T)$? We said earlier that our strategy will be to choose a test with a fixed, known probability of type I error, α. That is, we choose $R_\alpha(T)$ such that

$$P(T(\mathbf{X}) \in R_\alpha(T) | H_0 \text{ true}) = \alpha$$

We can choose $T(\mathbf{X})$ so that we can measure this probability.

Test Statistic

We choose $T(\mathbf{X})$ such that

- It has a known distribution *if H_0 is true*
- It depends on the data through an estimator of some kind
- The critical region is tractable, i.e. we know how to evaluate $P(T(\mathbf{X}) \in R_\alpha(T) | H_0 \text{ true})$

Example: Normal Mean

For example, if $X \sim N(\mu_0, 1)$, we can construct a test statistic.

Earlier we phrased our notion of plausible values for μ in terms of how far away they were from $\bar{X}$, in units of standard deviation $\frac{1}{\sqrt{n}}$.

This implies that a good test statistic might be our pivot from before,

$$T(\mathbf{X}) = \frac{\bar{X} - \mu_0}{1/\sqrt{5}}$$

Under $H_0 : \mu = \mu_0$, this has a known distribution,

$$T(\mathbf{X}) | H_0 \sim N(0, 1)$$

Example: Normal Mean

We want

$$P(T(\mathbf{X}) \in R_\alpha(T) | \mu = \mu_0) = \alpha$$

where $T(\mathbf{X})|\mu = \mu_0 \sim N(0, 1)$.

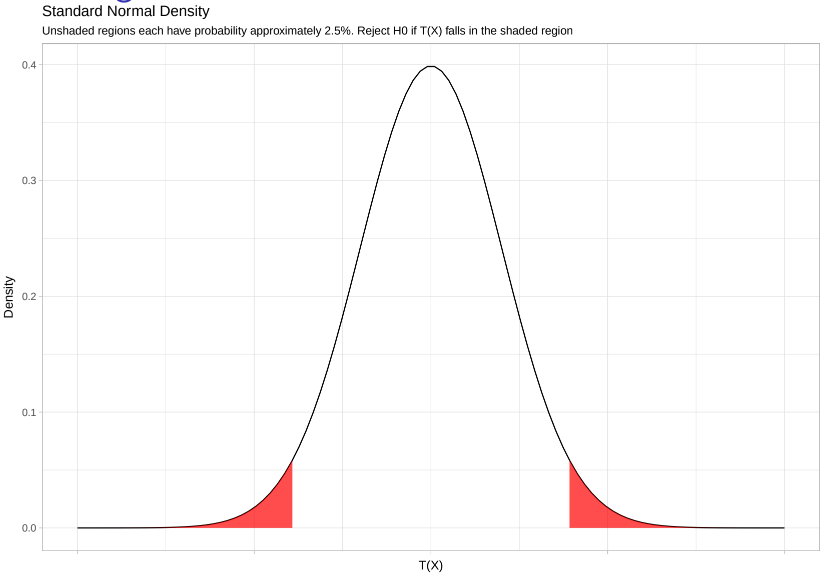
We could, in principle, choose $R_\alpha(T)$ to be any interval which has probability α under the standard normal distribution.

The most common and logical choice is to set

$$R_\alpha(T) = (-\infty, z_{\alpha/2}) \cup (z_{1-\alpha/2}, \infty) = (-\infty, -z_{1-\alpha/2}) \cup (z_{1-\alpha/2}, \infty)$$

which means we reject the null hypothesis for values of $T(\mathbf{X}) > z_{1-\alpha/2}$, or values of $T(\mathbf{X}) < -z_{1-\alpha/2}$.

Critical Region



Example

Let's test the hypothesis that $\mu = 0$ at the $\alpha = 0.05$ significance level in our example where $X \sim N(\mu_0, 1)$, $n = 5$ and

$$\mathbf{x} = (-0.89, -0.23, -0.65, -0.42, 0.21)$$

We find that

$$\begin{aligned} \bar{x} &= -0.4 \\ \sigma_0 &= 1 \\ R_\alpha(T) &= (-\infty, -1.96) \cup (1.96, \infty) \end{aligned}$$

which gives the test statistic

$$t(\mathbf{x}) = -0.89$$

P-Value

Because $t(\mathbf{x}) \notin R_\alpha(T)$, or equivalently $t(\mathbf{x}) \in R_\alpha(T)^c$, we fail to reject H_0 at the $\alpha = 0.05$ significance level.

How close were we? We call the probability of observing a test statistic with equal or greater evidence against H_0 the **p-value** of the test, p_0.

In this example, this means values farther away from 0.

P-Value

In this example, this means values farther away from 0, so the p-value for this test is

$$
\begin{aligned}
p_0 = P\big(|T(\mathbf{X})| > |t(\mathbf{x})|\big) &= P\big(|T(\mathbf{X})| > 0.89\big) \\
&= P(T(\mathbf{X}) > 0.89 \text{ or } T(\mathbf{X}) < -0.89) \\
&= 1 - \Phi(0.89) + \Phi(-0.89) \\
&= 2 \times (1 - \Phi(0.89)) \\
&= 0.38
\end{aligned}
$$

Null Distribution of the P-Value

Note the form of the p-value when the null hypothesis is true

$$p_0 = 2 \times (1 - \Phi(\left|T(\mathbf{X})\right|))$$

The p-value is a random variable, because it depends on the data, through the test statistic. When H_0 is true, we can find its probability distribution.

Proposition: when H_0 is true,

$$p_0 \sim Unif(0, 1)$$

Null Distribution of the P-Value

Proof: when H_0 is true, $T(\mathbf{X}) \sim N(0,1)$ and

$$\begin{aligned} P(\Phi(|T(\mathbf{X})|) < x) &= P(|T(\mathbf{X})| < \Phi^{-1}(x)) \\ &= P(T(\mathbf{X}) < \Phi^{-1}(x)) - P(T(\mathbf{X}) < -\Phi^{-1}(x)) \\ &= \Phi(\Phi^{-1}(x)) - (1 - \Phi(\Phi^{-1}(x))) \\ &= 2x - 1 \end{aligned}$$

so $\Phi(|T(\mathbf{X}|) \sim Unif(1/2, 1)$. It follows that

$$p_0 = 2 \times (1 - \Phi(|T(\mathbf{X})|) \sim Unif(0,1)$$

Null Distribution of the P-Value

If H_0 is true,

$$p_0 \sim Unif(0, 1)$$

In particular if H_0 is true then

$$P(p_0 < \alpha) = \alpha$$

Since p_0 is a monotone function of $\left|T(\mathbf{X})\right|$, our decision rule is:

Reject H_0 if $p_0 < \alpha$

P-Values

This distributional result holds more generally, and gives us a way to define tests in other situations.

The definition of the p-value was: *the probability of observing a test statistic with equal or greater evidence against H_0, if H_0 is true*.

Our testing procedure is then:

- Find a test statistic with a known distribution if H_0 is true
- Compute the test statistic for our dataset
- Compute the p-value, which is the probability of observing a test statistic that gives equal or greater evidence against H_0 than what we observed, if H_0 is true
- Reject H_0 if $p_0 < \alpha$. Remember, we choose α

Example

Consider the coin flip example: we flip a coin n times, and want to estimate θ, the probability of heads. We know a good estimator is

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i = \text{ sample proportion of heads}$$

Let's test the hypothesis that $\theta = \theta_0 = 0.5$, i.e. test that the coin is fair.

Example

Use $T(\mathbf{X}) = \hat{\theta}$ as a test statistic. Any value of $\hat{\theta}$ that is equal or farther from 0.5 than our observed value $t(\mathbf{x})$ gives greater evidence against the null hypothesis. That is, our p-value is

$$p_0 = P\left(\left|T\left(\mathbf{X}\right) - \theta_0\right| > \left|t(\mathbf{x}) - \theta_0\right|\right)$$

and we reject H_0 if $p_0 < \alpha$.

Note that this example shows that we don't necessarily *have* to base our test statistic off of a pivot.

Example

Our p-value is

$$p_0 = P\left(\left|T\left(\mathbf{X}\right) - \theta_0\right| \geq \left|t(\mathbf{x}) - \theta_0\right|\right)$$
$$= 1 - P\left(n\theta_0 - \left|nt(\mathbf{x}) - n\theta_0\right| < \sum_{i=1}^{n} X_i < n\theta_0 + \left|nt(\mathbf{x}) - n\theta_0\right|\right)$$

where

- $n\theta_0$: expected number of heads in n flips
- $nt(\mathbf{x})$: observed number of heads
- $\left|nt(\mathbf{x}) - n\theta_0\right|$: absolute difference in observed vs expected number of heads

Any value of $\sum_{i=1}^{n} X_i$ that is between $n\theta_0 - \left|nt(\mathbf{x}) - n\theta_0\right|$ and $n\theta_0 + \left|nt(\mathbf{x}) - n\theta_0\right|$ gives *less* evidence against $H_0 : \theta = \theta_0$

Example

Suppose we throw the coin 10 times and get 7 heads. Does this give sufficient evidence to reject $H_0 : \theta = 1/2$? That is, does this give us sufficient evidence to suggest that the coin is not fair?

We have

- $n\theta_0 = 5$ (we would expect to see 5 heads if the coin were fair)
- $nt(\mathbf{x}) = 7$ (we saw 7 heads)
- $p_0 = 1 - P\left(5 - |7-5| < \sum_{i=1}^{n} X_i < 5 + |7-5|\right) = P\left(\sum_{i=1}^{n} X_i \in \{0,1,2,3,7,8,9,10\}\right)$

Where $\sum_{i=1}^{n} \sim Binom(n, \theta_0)$

Example

For this example, compute

$$\begin{aligned} p_0 &= 1 - P\left(5 - |7-5| < \sum_{i=1}^{n} X_i < 5 + |7-5|\right) \\ &= P\left(\sum_{i=1}^{n} X_i \in \{0, 1, 2, 3, 7, 8, 9, 10\}\right) \\ &= \sum_{k=0}^{3} P\left(\sum_{i=1}^{n} X_i = k\right) + \sum_{k=7}^{10} P\left(\sum_{i=1}^{n} X_i = k\right) \\ &= 0.34 \end{aligned}$$

Example

Since the probability of observing a result with greater or equal evidence against H_0 when H_0 is true is 0.34, we don't reject H_0 based on these data (at the usual arbitrary 0.05 significance level).

That is, observing 7 heads in 10 flips is not terribly unlikely if $\theta = 0.5$.

What if $n = 20$ and we observe 14 heads? What would the p-value be?

Example

```r
# P-Value from first example
round(sum(dbinom(0:3,10,.5)) + sum(dbinom(7:10,10,.5)),2)
```

```
## [1] 0.34
```

```r
# P-Value from second example
round(sum(dbinom(0:6,20,.5)) + sum(dbinom(14:20,20,.5)),2)
```

```
## [1] 0.12
```

Observing the same $\hat{\theta} = 0.7$ gives greater evidence against $H_0 : \theta = 0.5$ when we flip the coin more times.

Example

If we flipped the coin 30 times and got 21 heads:

```
round(sum(dbinom(0:9,30,.5)) + sum(dbinom(21:30,30,.5)),2)
```

```
## [1] 0.04
```

If we flipped the coin 100 times and got 70 heads:

```
round(sum(dbinom(0:30,100,.5)) + sum(dbinom(70:100,100,.5))
```

```
## [1] 8e-05
```

Back to the Normal Distribution

Recap: to test $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$ at the α signficance level when $X \sim N(\mu_0, 1)$:

- Compute the test statistic $t(\mathbf{x}) = \frac{\bar{x} - \mu_0}{1/\sqrt{n}}$
- Compute the p-value $p_0 = 2 \times (1 - \Phi(\left|t(\mathbf{x})\right|))$
- Reject H_0 if $p_0 < \alpha$, otherwise fail to reject

Connection to the CLT

This is where the CLT becomes extremely useful. In reality, we don't know the distribution of our data. But we do know that in finite samples,

$$\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$$

approximately. So we may apply this hypothesis testing procedure anyways if the data are "large enough".

Example

For example, if the number of coin flips is “large enough”, then we can apply the CLT to our binomial example.

We have

$$\sum_{i=1}^{n} X_i \sim Binom(n, \theta_0)$$

$$E\left(\sum_{i=1}^{n} X_i\right) = n\theta_0$$

$$Var\left(\sum_{i=1}^{n} X_i\right) = n\theta_0(1-\theta_0)$$

$$\implies \frac{\sum_{i=1}^{n} X_i - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}} \xrightarrow{d} N(0,1)$$

Example

Use the test statistic

$$t(\mathbf{x}) = \frac{\sum_{i=1}^{n} x_i - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}}$$

and reject H_0 if $t(\mathbf{x}) < -z_{1-\alpha/2}$ or $t(\mathbf{x}) > z_{1-\alpha/2}$.

Example

From our first example, we get

$$t(\mathbf{x}) = \frac{7-5}{\sqrt{10 \times (5/10) \times (5/10)}} = 1.26$$

which gives an approximate p-value of

$$p_0 = 0.21$$

The CLT approximation is a bit rough with $n = 10$, but at least our conclusion remains the same.

Example

```
# Approximate p-value based on CLT; n = 10 example
round(2*(1-pnorm(round((7 - 5)/(sqrt(10*5*5/(10*10))),2)))
```

```
## [1] 0.21
```

```
# Approximate p-value based on CLT; n = 20 example
round(2*(1-pnorm(round((14 - 10)/(sqrt(20*5*5/(10*10))),2))
```

```
## [1] 0.07
```

Unknown Variance

For general data where we don't know its distribution, we also don't know σ. How can we still use our normal-theory test statistic?

We saw in lecture 2 that the CLT still holds when we replace σ with a consistent estimator, like

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

But we can do better. We can get the *exact* distribution of the modified test statistic,

$$T(\mathbf{X}) = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

Which is sometimes called *Student's Statistic*.

Distribution of Sample Variance

To accurately assess the distribution of Student's Statistic, we need to account for the uncertainty in estimating σ^2 by s^2.

On assignment 1, you were asked to find the distribution of $Y = Z^2$ when $Z \sim N(0,1)$. You may have found that the density of Y is given by

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} y^{-1/2} e^{-y/2}$$

which is called the χ_1^2 distribution, the Chi-Squared distribution with 1 degree of freedom.

$\chi_1^2 = Gamma(1/2, 1/2)$

You may recognize this as a $Gamma(1/2, 1/2)$ distribution, because using the fact that $\Gamma(1/2) = \sqrt{\pi}$,

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} y^{-1/2} e^{-y/2} = \frac{(1/2)^{1/2}}{\Gamma(1/2)} y^{1-1/2} e^{-(1/2)y}$$

which is recognized as the density of a $Gamma(1/2, 1/2)$ random variable.

χ^2_1 Distribution (textbook section 6.2)

The MGF of $Y \sim \chi^2_1$ is

$$M_Y(t) = (1-2t)^{-1/2}$$

It follows that if $Z_i, i = 1 \ldots n$ is an IID sample from a $N(0,1)$ distribution, that

$$S = \sum_{i=1}^{n} Z_i^2 \sim \chi^2_n$$

which is another fact you were asked to show on Assignment 1.

χ^2_n Distribution (textbook section 6.2)

This is a $Gamma(n/2, 1/2)$ distribution, with density

$$f_S(s) = \frac{1}{\Gamma(n/2)2^{n/2}} s^{n/2-1} e^{-s/2}$$

For $X \sim Gamma(\alpha, \beta)$ distribution parametrized in this way, we know that $E(X) = \frac{\alpha}{\beta}$ and $Var(X) = \frac{\alpha}{\beta^2}$, so

$$E(S) = n$$
$$Var(S) = 2n$$

Distribution of the Sample Variance

It follows that if $X_i \sim N(\mu_0, \sigma_0^2)$ is a IID sample, that

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu_0}{\sigma_0} \right)^2 \sim \chi_n^2$$

since each summand is $N(0,1)$ and they are all independent.

Hence if μ is *known*, and we estimate σ^2 with $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu_0)^2$, then

$$\frac{n\hat{\sigma}^2}{\sigma_0^2} \sim \chi_n^2$$

Distribution of the Sample Variance

Of course, μ is not known. Theorem B of section 6.3 in the textbook (page 197) describes how to modify the above distributional result to obtain

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

where

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

We "lose a degree of freedom" in estimating μ with $\bar{X}$.

A Pivot for σ^2

We have just derived an exact pivot for σ^2, which you can use for Hypothesis Tests and Confidence Intervals (see assignment 7).

Note also that Student's Statistic can be decomposed into the product of two pivots (sort of):

$$T = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}\right) \times \left(\frac{\sigma}{s}\right)$$

$$= \left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}\right) \times \left(\frac{\left(\frac{(n-1)s^2}{\sigma^2}\right)}{(n-1)}\right)^{-1/2}$$

We need one more big result.

Joint Distribution of $\bar{X}$ and s^2.

For an IID random sample from a $N(\mu_0, \sigma_0^2)$ distribution, we have

$$\frac{\bar{X} - \mu_0}{\sigma_0} \sim N(0, 1)$$

and

$$\frac{(n-1)s^2}{\sigma_0^2} \sim \chi^2_{n-1}$$

To get the joint distribution of these two quantities, we prove the following important result.

Proposition: $\bar{X} \perp s^2$

Joint Distribution of $\bar{X}$ and s^2.

The textbook proves this using moment generating functions (section 6.3, page 195, theorem A), but it's all algebra. The intuitive, more elegant proof uses linear algebra and properties of the multivariate normal distribution.

The idea is that

$$\begin{pmatrix} \bar{X} \\ X_1 - \bar{X} \\ \vdots \\ X_n - \bar{X} \end{pmatrix}$$

can be written as a linear combination of the datapoints $\mathbf{X} = (X_1, \ldots, X_n)$, which are jointly multivariate normal (because they are marginally normal and independent).

Joint Distribution of $\bar{X}$ and s^2.

Because of this, the vector

$$\begin{pmatrix} \bar{X} \\ X_1 - \bar{X} \\ \vdots \\ X_n - \bar{X} \end{pmatrix}$$

is multivariate normal. The proof then proceeds by showing that its elements are *uncorrelated*, which **because of their joint normality**, implies that they are independent.

It follows that $\bar{X} \perp s^2$, since these quantities can be written as functions of the above.

Important Note

We couldn't go through the major details of the proof, because they are slightly out of scope of this course. We can, though, touch on a very important point: the need to show **joint normality** in order to conclude independence from zero correlation.

You *cannot* conclude that because two arbitrary random variables are uncorrelated, they are independent; you also *cannot* conclude that because two random variables are normally distributed, that they are *jointly* normally distributed.

See Assignment 1, Question 2, or even better, check out Prof. Jeff Rosenthal's rant on this topic:
http://probability.ca/jeff/teaching/uncornor.html

t-distribution

We can now state the answer to our original problem.

Definition: let $Z \sim N(0,1)$, $U \sim \chi^2_\nu$, and $Z \perp U$. Then the distribution of the quantity

$$T = \frac{Z}{\sqrt{U/\nu}}$$

is called *Student's t-distribution with ν degrees of freedom*.

The density of T is given by

$$f_T(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi} \times \Gamma(\nu/2)} \times \left(1 + \frac{t^2}{\nu}\right)^{\frac{-(\nu+1)}{2}}$$

You will walk through this derivation step by step on assignment 7.

Student's Statistic

Corollary:

$$T = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim t_{n-1}$$

Proof: write

$$\frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \right) \times \left(\frac{s^2}{\sigma^2} \right)^{-1/2}$$

from which the result follows from the defition of the t-distribution and the joint distribution of $\bar{X}$ and s^2.

Properties

The t-distribution is symmetric, $f(t) = f(-t)$.

$E(T) = 0$, $Var(T) = \frac{\nu}{\nu-2}$ for $\nu > 2$.

As $\nu \to \infty$, $T \xrightarrow{d} Z$.

As the sample size gets bigger, the uncertainty in estimating s^2 becomes negligible, and the distribution of our test statistic using the sample variance tends towards the distribution of the test statistic we would use if σ^2 were known.

Unknown Variance

It follows that a $1-\alpha$ confidence interval for μ, when σ^2 is unknown, can be obtained as

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

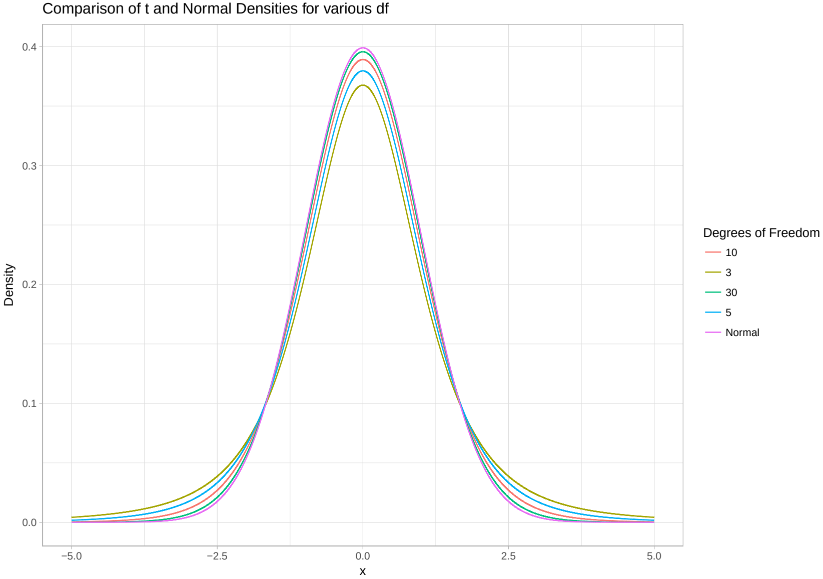
because we can use the same symmetry-related arguments as we did with the normal distribution.

So in practice, just replace the z quantiles with t quantiles.

For testing $H_0 : \mu = \mu_0$ when σ^2 is unknown, the rejection region is

$$R_\alpha(T) = (-\infty, -t_{1-\alpha/2}) \cup (t_{1-\alpha/2}, \infty)$$

t vs Normal


Comparison of t and Normal Densities for various df
Density
0.0
0.1
0.2
0.3
0.4
x
−5.0
−2.5
0.0
2.5
5.0
Degrees of Freedom
10
3
30
5
Normal