Topic: Hypothesis Testing Part I

10-5-22

(Ch. 9.1, 9.2)

Estimation Recap

point estimates - a single value, based on data, that is meant to represent the "best" guess as to the value of θ.

interval estimates - explicit recognition that conclusion is uncertain by providing a range of possible values for θ.

There are various statistical principles that can guide estimation in the sense of determining ways in which the data should effect statistical conclusions about a model parameter, θ:

1. Sufficiency Principle - provided the model is adequate, identical conclusions should be drawn from any different observations of the data if they have the same value of sufficient stat.

2. **Weak Likelihood Principle** – two observed data sets that have proportional likelihoods for θ should yield identical conclusions about θ, provided the model is adequate.

3. **Strong Likelihood Principle** –
two observed data sets from two different (but both adequate) prob models involving the same θ should yield identical conclusions about θ if their likelihoods are proportional

Other notable principles include:

4. Invariance Principle

5. Conditionality Principle

but these are beyond the scope of this class.

There are also specific principles regarding the form and interpretation of statistical conclusions about θ, from the data and an assumed model:

1. Strong Repeated Sampling Principle

statical procedures should be assessed by their hypothetical performance under identical sampling conditions (physical interpretation)

2. Weak Repeated Sampling Principle

we should avoid stat. procedures that (for some/any values of θ) will produce misleading conclusions most of the time

3. Bayesian Coherency Principle

all uncertaintires are described w/ prob. distbn's to ensures self-consistent "betting" behavior

4. Principle of Coherent Decision Making

ensures self-consistent decisions are made from stat. analyses

This context for statistical estimation and interpretation now brings us to **statistical inference**.

Q) What is statistical inference?

"the process of drawing conclusions about an unknown parameter that one wants to measure or estimate" — Encyclopedia Brittanica

There is a quantification of uncertainty or unknowables.

Principles for Statistical Inference

1. **Sampling Theory** — prioritizes strong or weak repeated sampling principle
2. **Likelihood Theory** — prioritizes strong or weak likelihood principle
3. **Bayesian Theory** — prioritizes Bayesian coherency principle
4. Decision Theory — the origins of significance testing (N-P paradigm); prioritizes coherent princp. of decision making

Tests of Significance

Setting: $\underset{\sim}{x}_{obs} = (x_1, x_2, \ldots, x_n)$ — The data

"null" hypothesis → $H_0 = \begin{cases} \text{a statement/hypothesis concerning the} \\ \text{distribution of } (X_1, \ldots, X_n) \end{cases}$

"alternative" hypothesis → $H_A = \begin{cases} \text{a different statement/hypothesis concerning} \\ \text{the distb'n of } (X_1, \ldots, X_n) \end{cases}$

- A simple hypothesis completely specifies the distribution of the (random) data.
- A composite hypothesis, on the other hand, does not completely (and unambiguously) specify the distribution hypothesized.

Null hypotheses can arise in many different settings:

- H_0 may correspond to the prediction of some scientific theory thought to be true

eg) astronomical model describes mass to light ratio of some galaxies

- H_0 may divide the possible distb'n's into two qualitatively different types

eg) coin is fair

- H_0 could represent a simple set of circumstances which, in absence of evidence to the contrary, we may wish to assume holds

eg) errors in obs. data are Normal

- H_0 could assert a complete absense of structure in some sense

eg) overall ANOVA F-test for all regression coefficients

Note: The null and alternative are not necessarily given equal footing in the context of significance tests. Notably, H_0 is of intrinsic interest whereas H_A serves only to indicate the direction of interesting departures.

The motivating question here is: Is there evidence (from the data) of inconsistency w/ H_0?

Given any H_0 and H_A, we can visualize our possible conclusions from a significance test as

Reality

Conclusion/Decision	H_0	Not H_0
Reject H_0	Type I error	✓
Failure to reject H_0	✓	Type II error

The **significance level** of a hypothesis test is the (often controlled, pre-determined) conditional probability of a Type I error.

Significance level $= \alpha \in (0,1)$

The **power** of a hypothesis test is the conditional probability of <u>not</u> making a Type II error.

power $= 1-\beta$, where $\beta =$ prob. of type II error $\in (0,1)$

Q) When we say "probability of type X error", what is random? ie. what are we describing w/ a probability law/distb'n?

10-7-22

Ex) Say $n=16$, $X_1, \dots, X_n \overset{IID}{\sim} N(\mu, 0.4^2)$

You want to test: $\begin{cases} H_0: \mu = 37 \\ \quad \text{vs} \\ H_A: \mu = 36.8 \end{cases}$

---

① If $A_\alpha = \{\bar{x} : \bar{x} < 36.9\}$ is our rejection region (meaning we reject H_0 only if we observe $\bar{X}_{obs} \in A_\alpha$) What is the significance level of our test, α?

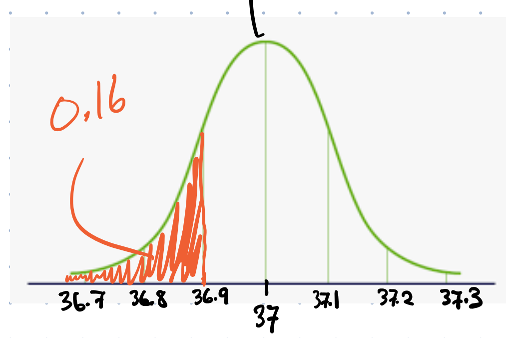
$\alpha = \Pr(\text{Type I error})$

$= \Pr(\text{Reject } H_0 \mid H_0 \text{ is true})$

$= \Pr(\bar{X} < 36.9 \mid \mu = 37)$

$= 0.16$

$\bar{X} \sim N\left(37, \frac{0.4^2}{16}\right)$, $\frac{0.4^2}{16} = 0.1^2$



What is the (approximate) power?

Power $= 1 - \beta$

$\beta = \Pr(\text{Type II error})$

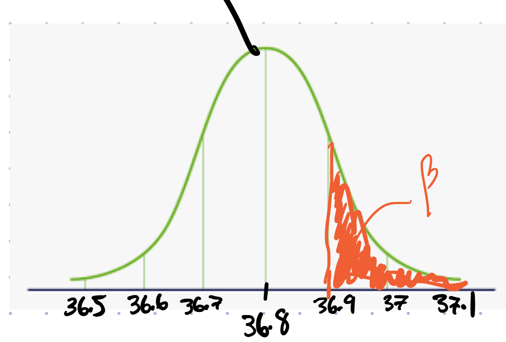
$= \Pr(\text{Fail to Reject } H_0 \mid H_0 \text{ is NOT true})$

$= \Pr(\bar{X} \geq 36.9 \mid \mu = 36.8)$

$= 0.16$

Power $= 1 - 0.16 = 0.84$

$\bar{X} \sim N\left(36.8, \frac{0.4^2}{16}\right)$



Takeaway: The key is using the sampling distb'n of $\bar{X}$!

called the "test statistic"

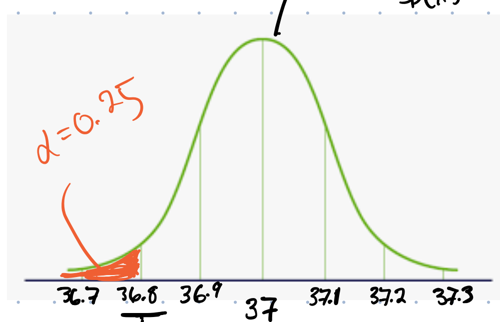
Group Exercize:

② Define a new rejection region, A_α, (w/ n=16) so that $\alpha \approx 0.025$. What is the power for this test?

If H_0 is true: $\bar{X} \sim N\left(37, \frac{0.4^2}{16}\right)$

$SD(\bar{X}) = 0.1$

$$\alpha = Pr(\bar{X} < ? \mid \mu = 37)$$



$$A_\alpha = \{\bar{x} : \bar{x} < 36.8\}$$

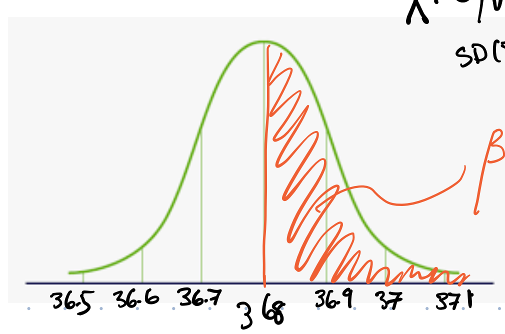
$$\beta = Pr(\bar{X} \geq 36.8 / \mu = 36.8)$$

$$= 0.50$$

So Power $= 1 - 0.5 = 0.5$

If H_0 is false, our only other option is H_A and: $\bar{X} \sim N\left(36.8, \frac{0.4^2}{16}\right)$

$SD(\bar{X}) = 0.1$



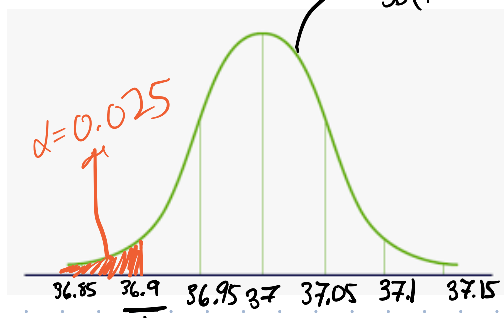
③ Repeat #2 but now assume $n=64$.

$\alpha = Pr(\bar{X} < ? \mid \mu = 37)$

$A_\alpha = \{\bar{x} : \bar{x} < 36.9\}$

If H_0 is true: $\bar{X} \sim N\left(37, \frac{0.4^2}{64}\right)$

$SD(\bar{X}) = 0.05$

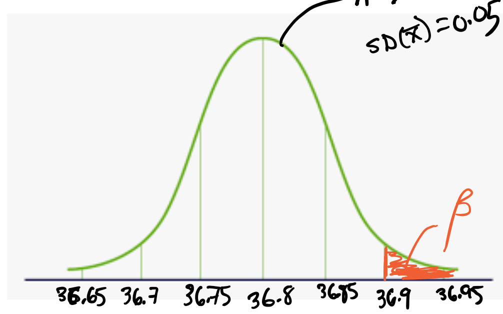


$\beta = Pr(\bar{X} \geq 36.9 \mid \mu = 36.8)$
$= 0.025$

power $= 1 - 0.025$
$= 0.975$

If H_0 is false, our only other option is H_A and: $\bar{X} \sim N\left(36.8, \frac{0.4^2}{64}\right)$

$SD(\bar{X}) = 0.05$



④ How would the power in #2 + #3 change if we used a smaller (or larger) α?

If we fix n then:

- As α decreases, so does the power (b/c β increases).

- As α increases, so does the power.

Notes on Quiz 2

#2) There is a small, but consequential typo in $\hat{\theta}_1$. This typo makes it v. difficult to determine whether $\hat{\theta}_2$ or $\hat{\theta}_3$ has a smaller MSE for $n=4$. Here is the correct version:

$$\hat{\theta}_1 = \frac{0.3}{n}\sum_{i=1}^{n}\frac{1}{X_i} \; ; \; \hat{\theta}_2 = E\left[\hat{\theta}_1 \,\middle|\, \sum_{i=1}^{n} X_i\right]$$

$$\hat{\theta}_3 = \frac{4n}{\sum_{i=1}^{n} X_i}$$

Q) What is $\hat{\theta}_2$? How is it random?

Recall some former HW problems:

HW 2 #3

Given joint density: $f(x,y) = \frac{6}{7}(x+y)^2 \mathbb{I}\{0 \le x \le 1\}\mathbb{I}\{0 \le y \le 1\}$

Find $E[Y|X=x]$

HW3 #1 for $T = \sum_{i=1}^{N} X_i$ w/ X_i all IID

$E[T|N=n] = NE[X]$

and

$E[E(T|N=n)] = E(N)E(X)$

Midsemester Adjustments

10-17-22

43% response rate

- OH Th switch to M 4-5pm
- Must visit OH by end of Unit 2 to be elligible for full participation grade.
- More in-class examples
- No more than 2 HW/wk somewhat shortened assignments
- opportunity to redo a quiz for partial credit

Topic: Hypothesis Testing Part II (Ch. 9.3 & 9.4)

<u>Setting:</u> $\underset{\sim}{x}_{obs} = (x_1, x_2, \ldots, x_n)$

$$H_0 = \begin{cases} \text{a statement/hypothesis concerning the} \\ \text{distb'n of } (X_1, \ldots, X_n) \end{cases}$$

$$(H_1 \text{ or}) \; H_A = \begin{cases} \text{a different statement/hypothesis concerning} \\ \text{the distb'n of } (X_1, \ldots, X_n) \end{cases}$$

Typically, we also assume

$$(X_1, \ldots, X_n) \overset{IID}{\sim} f(x; \theta)$$

So the likelihood for θ given $\underset{\sim}{x}_{obs}$ is:

$$L(\theta) = \prod_{i=1}^{n} f(x_i; \theta) \quad \text{or} \quad \ell(\theta) = \sum_{i=1}^{n} \ln(f(x_i; \theta))$$

Therefore, we are often interested in hypotheses about the value of θ.

$$\begin{cases} H_0: \theta = \theta_0 \\ \\ H_1: \theta = \theta_1 \end{cases}$$

To conduct a "level α" hypothesis test we need

1) A test statistic
   ie. a function of the data, say, $T(\underset{\sim}{x})$
2) A rejection rule

ie. some $A_\alpha = \{ \underset{\sim}{x} : T(\underset{\sim}{x}) \text{ behaves some atypical way} \}$

A popular choice of test statistic is the ratio of the likelihoods specified by H_0 & H_1:

The **Likelihood ratio (LHR)** is the ratio

$$\Lambda(\theta_i; x_{obs}) = \Lambda = \frac{L(\theta_0)}{L(\theta_1)}$$

FYI this is the Greek capital "lambda"

Q) When is Λ large? What does $\Lambda = 1$ mean about H_0 & H_1? What is the smallest/largest possible value for Λ?

Now, the idea is to define a rejection region (or rule) so that

$$\left\{ \begin{array}{l} \text{we reject } H_0 \text{ in favor of } H_1 \\ \text{only if } T(x_{obs}) \text{ is } \underline{\text{improbable}} \\ \text{if we assume } H_0 \text{ is correct.} \end{array} \right.$$

ie. $$A_\alpha = \left\{ \underset{\sim}{x} : \begin{array}{l} T(\underset{\sim}{x}_{obs}) \text{ occurs w/ low probability} \\ \text{under the assumption that} \\ H_0 \text{ is correct.} \end{array} \right.$$

Correspondingly, a p-value is

$$\text{p-value} = \Pr\left(\begin{array}{l} T(\underset{\sim}{X}) \text{ is } T(\underset{\sim}{x}_{obs}) \text{ or anything less} \\ \text{likely than } T(\underset{\sim}{x}_{obs}) \text{ under the} \\ \text{assumption that } H_0 \text{ is correct} \end{array}\right)$$

Recall our example from before:

Ex) Say $X_1, \dots, X_n \overset{IID}{\sim} N(\mu, 0.4^2)$

Test $H_0: \mu = 37$ vs $H_1: \mu = 36.9$

Case 1: $n = 16$

Case 2: $n = 64$

- If $\alpha = 0.025$, how does increasing n change the power of the test?
- How does changing α impact the power of the test?

Q) How do we choose the direction in A_α?

A) $A_\alpha = \left\{ \underset{\sim}{x} : \frac{L(\theta_0)}{L(\theta_1)} \geq c_\alpha \right\}$

which can often be simplified into statements about a sufficient stat, $T(X) \gtrless k_\alpha$.

Ex cont'd) Say $X_1, \ldots, X_n$ IID $N(\mu, 0.4^2)$

Test $H_0: \mu = 37$ vs $H_1: \mu = 36.9$

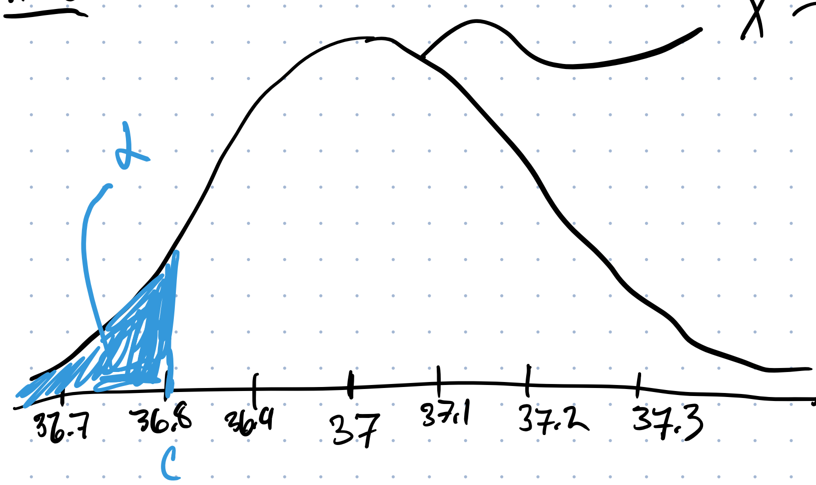
Case 1: $n=16$

Case 2: $n=64$

Suppose we change H_1 to $H_1: \mu = 36$ and fix $\alpha = 0.025$.

- What is $T(\underset{\sim}{X})$?
- What does "atypical" mean in this context?
- What is A_α for $n=16$? For $n=64$?

n=16



$\bar{X} \overset{H_0}{\sim} N\left(37, \frac{0.4^2}{16}\right)$

$\Downarrow$ sd $= 0.1$

$A_\alpha = \{\underset{\sim}{x} : \bar{x} < c\}$

$Pr(\text{Reject } H_0 \mid H_0 \text{ is correct}) = 0.025$

In short: changing the value of μ in H_1 doesn't change the rejection region(s) A_α at all

Note:

In testing some hypothesis, there is opportunity for creative choices of the test statistic, $T(\underset{\sim}{X})$ and also for the rejection region, A_α.

When comparing different tests of the same hypotheses, if both tests have the same significance level, then the test w/ the highest power is preferable.

10-19-22

The Neyman-Pearson (NP) Lemma for Most Powerful Tests

For a test of two simple hypotheses the LHR test is at least as powerful as any other test w/ the same (or more restrictive - smaller) α.

Ie. The most powerful test of $H_0: \theta = \theta_0$ vs $H_1: \theta = \theta_1$ is the test w/

$$T(\underset{\sim}{X}) = \frac{L(\theta_0)}{L(\theta_1)} = \frac{f(X_1, X_2, \ldots, X_n; \theta_0)}{f(X_1, X_2, \ldots, X_n; \theta_1)}$$

and rejection region

$$A_\alpha = \left\{ \underset{\sim}{X} : \frac{L(\theta_0)}{L(\theta_1)} < C_\alpha \right\}$$

where C_α is chosen so that

$$P_r\left(\frac{L(\theta_0)}{L(\theta_1)} < C_\alpha \,\Big|\, H_0 \text{ is correct} \right) = \alpha.$$

Recall:

The likelihood function for $(X_1, \ldots, X_n)$ is

$$L(\theta) = f(X_1, \ldots, X_n; \theta).$$

So, for any value of θ, $L(\theta)$ is the probability of observing a particular set of data, $(X_1, \ldots, X_n)$.

But for an observed data set, $(x_1, \ldots, x_n)$, $L(\theta)$ is a deterministic function of the possible values for θ.

Proof of NP Lemma:

Define $Y = \mathbb{I}\{\underset{\sim}{X} \in A_\alpha\}$ and $Y^* = \mathbb{I}\{\underset{\sim}{X} \in A_\alpha^*\}$

where $\alpha^* \leq \alpha$, where $\underset{\sim}{X} = (X_1, \dots, X_n)$.

Note that $E[Y \mid H_0] = \Pr(Y=1 \mid H_0) = \alpha$

and $E[Y \mid H_1] = \Pr(Y=1 \mid H_1) = 1-\beta$ (ie. the power).

And since $\alpha^* \leq \alpha$, $E[Y^* \mid H_0] = \alpha^* \leq \alpha$.

It remains to show that $E[Y \mid H_1] \geq E[Y^* \mid H_1]$.

For observed data $\underset{\sim}{x} = (x_1, \dots, x_n)$, write the corresponding observed values of Y and Y^* by y and y^*, respectively.

Note $y^*[c_\alpha f(\underset{\sim}{x};\theta_1) - f(\underset{\sim}{x};\theta_0)] \leq y\,[c_\alpha f(\underset{\sim}{x};\theta_1) - f(\underset{\sim}{x};\theta_0)]$.

> To see this consider the case where $y=1, y^*=0$:
>
> Since $y=1$, $\dfrac{L(\theta_0)}{L(\theta_1)} < c_\alpha$, ie. $0 < c_\alpha f(\underset{\sim}{x};\theta_1) - f(\underset{\sim}{x};\theta_0)$
>
> Next, consider the case where $y=0$, $y^*=1$:
>
> Here, $y=0$ means $c_\alpha f(\underset{\sim}{x};\theta_1) - f(\underset{\sim}{x};\theta_0) \leq 0$.

Thus, multiplying both sides of the inequality by $f(\underset{\sim}{x};\theta)$ and integrating w.r.t. $\underset{\sim}{x}$ yields

$$c_\alpha E[Y^* \mid \theta_1] - E[Y^* \mid \theta_0] \leq c_\alpha E[Y \mid \theta_1] - E[Y \mid \theta_0]$$

which can be rearranged so that

$$E[Y \mid \theta_0] - E[Y^* \mid \theta_0] \leq c_\alpha \left(E[Y \mid \theta_1] - E[Y^* \mid \theta_1]\right)$$

where the LHS is $\alpha - \alpha^* \geq 0$.

Hence, $E[Y \mid \theta_1] - E[Y^* \mid \theta_1] \geq 0$. ∎

Uniformly Most Powerful (UMP) 1-sided tests

For simple hypotheses $H_0: \theta = \theta_0$ vs $H_1: \theta = \theta_1$, the rejection region, A_α, depends on the sign (+/−) of the difference $\theta_0 - \theta_1$.

For a given α-level, the rejection region for this test stays the same for any $\theta_1 < \theta_0$. Every value for θ_1 that remains on the same side of θ_0 has the same most powerful test by the NP lemma. Therefore, the LRT is uniformly most powerful for tests of composite alternatives: $H_0: \theta = \theta_0$ vs. $H_1: \theta < \theta_0$.

(or $H_1: \theta > \theta_0$)

Q) Is there a UMP test for a two sided H_1?

Unfortunately, no!

A test of $H_0: \theta = \theta_0$ vs. $H_1: \theta \neq \theta_0$ allows for differences from θ_0 in either direction.

We can define the rejection region for this test as the <u>union</u> of rejection regions for each one-sided test at an $\alpha/2$-level of significance.

But, either of these 1-sided tests will have greater power for certain values in the parameter space, Θ.

nice!

Ex) Suppose $X_1, \dots, X_n \overset{IID}{\sim} N(\theta, 1)$ and test

$H_0: \theta = \theta_0$ vs. $H_A: \theta = \theta_1$, at $\alpha = 0.05$ level.

1-parameter exponential family! $\Rightarrow$ $T(\underset{\sim}{X}) = \overline{X}$ is a sufficient stat for θ

Furthermore, $T(\underset{\sim}{X}) \sim N(\theta, \frac{1}{n})$

$$L(\theta) = f(\underset{\sim}{x}; \theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x_i - \theta)^2\right\}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2} \sum_{i=1}^{n} (x_i - \theta)^2\right\}$$

$$\Lambda = \frac{L(\theta_0)}{L(\theta_1)} = \frac{\cancel{\left(\frac{1}{\sqrt{2\pi}}\right)^n} \exp\left\{-\frac{1}{2} \sum_{i=1}^{n} (x_i - \theta_0)^2\right\}}{\cancel{\left(\frac{1}{\sqrt{2\pi}}\right)^n} \exp\left\{-\frac{1}{2} \sum_{i=1}^{n} (x_i - \theta_1)^2\right\}}$$

$$= \exp\left\{-\frac{1}{2} \sum_{i=1}^{n} (x_i - \theta_0)^2 - \left(-\frac{1}{2} \sum_{i=1}^{n} (x_i - \theta_1)^2\right)\right\}$$

$$= \exp\left\{-\frac{1}{2} \sum_{i=1}^{n} \left((x_i - \theta_0)^2 - (x_i - \theta_1)^2\right)\right\}$$

$\pm \overline{x}$ $\qquad$ $\pm \overline{x}$

$$= \exp\left\{-\frac{1}{2} \sum_{i=1}^{n} \left[(\underbrace{x_i - \overline{x}}_{A} + \underbrace{\overline{x} - \theta_0}_{B})^2 - (\underbrace{x_i - \overline{x}}_{A'} + \underbrace{\overline{x} - \theta_1}_{B'})^2 \right]\right\}$$

$$= \exp\left\{-\frac{1}{2} \sum_{i=1}^{n} \left[(x_i - \overline{x})^2 + 2(x_i - \overline{x})(\overline{x} - \theta_0) + (\overline{x} - \theta_0)^2 \right.\right.$$

$$\left.\left. - \left[(x_i - \overline{x})^2 + 2(x_i - \overline{x})(\overline{x} - \theta_1) + (\overline{x} - \theta_1)^2 \right] \right]\right\}$$

$$= \underset{(*)}{\cdots} = \exp\left\{ n\overline{x}(\theta_0 - \theta_1) + \frac{n}{2}(\theta_1^2 - \theta_0^2) \right\}$$

Supplemental notes for 10-19-22

(Note: I encourage you to try this out on your own before reading the soln.)

(*) To see how we get this final expression consider:

$$\sum (x_i-\theta_0)^2 = \sum (x_i - \bar{x} + \bar{x} - \theta_0)^2$$
$$= \sum \left[(x_i-\bar{x})^2 + (\bar{x}-\theta_0)^2 + 2(x_i-\bar{x})(\bar{x}-\theta_0)\right]$$
$$= \sum_{i=1}^{n} \left[(x_i^2 - 2x_i\bar{x} + \bar{x}^2) + (\bar{x}^2 - 2\theta_0\bar{x} + \theta_0^2) + (2x_i - 2\bar{x})(\bar{x}-\theta_0)\right]$$
$$= \sum_{i=1}^{n} \left(x_i^2 - 2x_i\bar{x} + \bar{x}^2 + \bar{x}^2 - 2\theta_0\bar{x} + \theta_0^2 + 2x_i\bar{x} - 2\bar{x}^2 - 2\theta_0 x_i + 2\theta_0\bar{x}\right)$$
$$= \sum x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 + n\bar{x}^2 - 2n\theta_0\bar{x} + n\theta_0^2 + 2n\bar{x}^2 - 2n\bar{x}^2 - 2n\theta_0\bar{x} + 2n\theta_0\bar{x}$$
$$= \sum x_i^2 - 2n\theta_0\bar{x} + n\theta_0^2.$$

Similarly, for $(x_i-\theta_1)^2$ we have:

$$\sum (x_i-\theta_1)^2 = \sum x_i^2 - 2n\theta_1\bar{x} + n\theta_1^2$$

And therefore,

$$\sum (x_i-\theta_0)^2 - \sum (x_i-\theta_1)^2 = 2n\bar{x}(\theta_1-\theta_0) + n(\theta_1^2-\theta_0^2)$$

10-21-22

Duality of Confidence Intervals and Hypothesis Tests

Recall $\boxed{A_\alpha = \{\underset{\sim}{x} : T(\underset{\sim}{x}) \text{ is unusual enough under } H_0\}}$

represents the subset of the joint sample space, $\mathcal{X}$, for all values of $\underset{\sim}{x} = (x_1, x_2, \ldots, x_n)$ such that we reject $H_0 : \theta = \theta_0$ at the level α.

All other possible values of $\underset{\sim}{x}$ lie in what we call the "acceptance region",

$$\boxed{A(\theta_0) = \{\underset{\sim}{x} : \underset{\sim}{x} \notin A_\alpha\}},$$

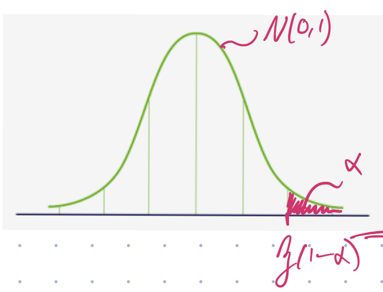
that is, the subset of $\mathcal{X}$ for which we would fail to reject $H_0 : \theta = \theta_0$ at the level α.

Ex) For $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\theta, 1)$ if we test $H_0 : \theta = \theta_0$ vs $H_1 : \theta > \theta_0$ at the α-level and reject for $\underset{\sim}{X} \in A_\alpha = \{\underset{\sim}{X} : \bar{X} - \theta_0 > c_\alpha\}$ where c_α is the quantile such that $\Pr(\bar{X} - \theta > c_\alpha \mid H_0) = \alpha$, ie. $c_\alpha = z_{(1-\alpha)}/\sqrt{n}$.

$\bar{X} \sim N(\theta, 1/n)$

$\bar{X} - \theta \sim N(0, \frac{1}{n})$

$\sqrt{n}(\bar{X} - \theta) \sim N(0, 1)$



Then we fail to reject H_0 when $\bar{X} - \theta_0 < z_{(\alpha)}/\sqrt{n}$

ie. $\bar{X} - z_{(\alpha)}/\sqrt{n} < \theta_0$.

Hence $A(\theta_0) = \{\underset{\sim}{x} : \bar{X} - z_{(\alpha)}/\sqrt{n} < \theta_0\}$.

Setting:

Formally, let θ be a parameter for a family of probability distb'ns and denote the set of all possible values of θ be the parameter space, Θ. Let $\underset{\sim}{X} = (X_1, \ldots, X_n)$ be the random vector of data.

Theorem: From Tests to Confidence Regions

If for every value of $\theta_0 \in \Theta$ there is a level-α hypothesis test of $H_0: \theta = \theta_0$ w/ corresponding acceptance region, $A(\theta_0)$,

Then, the set $\boxed{C(\underset{\sim}{X}) = \{\theta : \underset{\sim}{X} \in A(\theta)\}}$ is a $(1-\alpha) \times 100\%$ confidence region for θ.

Proof:

For A to be the acceptance region of a level-α test means $\Pr\left(\underset{\sim}{X} \in A(\theta_0) \,\middle|\, \theta = \theta_0\right) = 1 - \alpha$.

So, $\Pr\left(\theta_0 \in C(\underset{\sim}{X}) \,|\, \theta = \theta_0\right) = \Pr\left(\underset{\sim}{X} \in A(\theta_0) \,\middle|\, \theta = \theta_0\right) = 1 - \alpha$

by definition of $C(\underset{\sim}{X})$.

Main Idea

A $(1-\alpha) \times 100\%$ conf. region for θ consists of all those values of $\theta_0 \in \Theta$ for which the hypothesis $H_0: \theta = \theta_0$ will NOT be rejected at level α.

Theorem: From Confidence Regions to Tests

If $C(\underset{\sim}{X})$ is a $(1-\alpha)*100\%$ confidence region for θ;
ie. for every θ_0, $Pr(\theta_0 \in C(\underset{\sim}{X}) \mid \theta=\theta_0) = 1-\alpha$.

Then an acceptance region for a test at level α of the hypothesis $H_0: \theta=\theta_0$ is

$$A(\theta_0) = \{\underset{\sim}{X} : \theta_0 \in C(\underset{\sim}{X})\}.$$

Proof:

If $C(\underset{\sim}{X})$ is such that $Pr(\theta_0 \in C(\underset{\sim}{X}) \mid \theta=\theta_0) = 1-\alpha$, then the test of $H_0: \theta=\theta_0$ has level α because

$$Pr(\underset{\sim}{X} \in A(\theta_0) \mid \theta=\theta_0) = Pr(\theta_0 \in C(\underset{\sim}{X}) \mid \theta=\theta_0) = 1-\alpha.$$

Main Idea

The hypothesis $H_0: \theta=\theta_0$ is NOT rejected if θ_0 lies in the confidence region for θ.

Hint: Pause & ask yourself – what is random?

X – random
x – observed/fixed

θ – unknown, fixed
θ_0, θ_1 – fixed

Worksheet practice.

Stat 61 In-Class Worksheet

Original group members:

Oct 21, 2020

$$L(\mu) = \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi\, 0.16}}\right) \exp\left\{-\frac{(x_i-\mu)^2}{2\cdot 0.16}\right\} = (\text{const})^n \cdot \exp\left\{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2(0.16)}\right\}$$

Suppose we observe $X_1, \ldots, X_n$ IID data points from a $N(\mu, 0.4^2)$ distribution where $n = 16$ and we wish to test $H_0 : \mu = 37$.

see supplemental notes from 10-19-22

1. For a simple alternative, $H_1 : \mu = 36.8$ with $\alpha = 0.025$, what is the rejection region A_α? What is the power of this test?

$$\Lambda = \frac{L(\mu_0)}{L(\mu_1)} = \frac{\exp\left\{-\frac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{2(0.16)}\right\}}{\exp\left\{-\frac{\sum_{i=1}^{n}(x_i-\mu_1)^2}{2(0.16)}\right\}} = \frac{\exp\left\{-\frac{\sum_{i=1}^{n}(x_i-37)^2}{2(0.16)}\right\}}{\exp\left\{-\frac{\sum_{i=1}^{n}(x_i-36.8)^2}{2(0.16)}\right\}} = \ldots = \text{const} \times \exp\left\{2n\bar{X}(37-36.8)\right\}$$

So the rejection region is:

$$A_\alpha = \{x : \exp\{2(0.2)n\bar{X}\} \times \text{const} < C_\alpha\}$$

$$\Leftrightarrow \{x : n \cdot 0.4\bar{X} < C''_\alpha\} \quad (\text{where } C''_\alpha = \ln(C'_\alpha/\text{const}))$$

$$\Leftrightarrow \{x : \bar{X} < C'''_\alpha\} \quad (\text{where } C'''_\alpha = \ln[C'_\alpha/\text{const}/n(0.2)])$$

and $\Pr(\bar{X} < C'''_\alpha \mid H_0) = 0.025.$

Hence $C'''_\alpha = 36.8$

and $\boxed{A_\alpha = \{x : \bar{x} < 36.8\}}$.

see class notes on 10-7-22

And finally,

$\boxed{\text{power} = \Pr(\bar{X} < 36.8 \mid \mu = 36.8) = 0.5}$

2. How does the rejection region change if we increase n to $n = 64$ but keep everything else the same?

A_α will be more broad w/ $n = 64$ that it is w/ $n = 16$.

3. How would the power change if we decreased α but kept everything else the same?

Decreasing α will decrease the power

4. How would A_α change if we instead tested against $H_1 : \mu = 36$?

A_α doesn't change at all

The Neyman-Pearson lemma implies that, for testing $H_0 : \mu = \mu_0$ vs $H_1 : \mu = \mu_1$ with $\mu_1 < \mu_0$, the test that rejects for $\bar{X} < c_\alpha$ is most powerful of all tests with comparable α.

5. Is the test in (1) above uniformly most powerful for any pair of hypotheses? If so which ones and why?

Yes, is UMP for $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$ since both the test stat. & Rej. Reg. remain unchanged & MP by N-P lemma for any other $\mu_1 < \mu_0$

6. For the test in (1) above, suppose you observe data where $\bar{x}_{obs} = 36.85$. What is the p-value for this one-sided test? (I.e. what is the smallest α level that would lead to rejecting H_0?)

$$\text{p-val} = \Pr(\Lambda \le \Lambda_{obs} \mid H_0 : \mu = 37)$$

$$= \Pr(\bar{X} \le \bar{x}_{obs} \mid H_0 : \mu = 37)$$

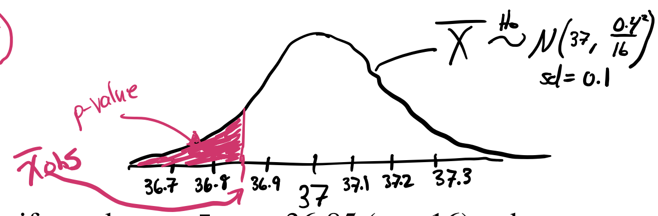
$$= \Pr(\bar{X} \le 36.85 \mid \mu = 37)$$

$$= \text{pnorm}(36.85, \text{mean}=37, \text{sd}=0.1, \text{lower.tail}=T)$$

$\Lambda = \text{const} \times \exp\{2n\bar{X}(37 - 36.8)\}$

$\Lambda_{obs} = \text{const} \times \exp\{2n \cdot 36.85(37 - 36.8)\}$

In general : p-val $= \Pr(\Lambda \le \Lambda_{obs} \mid H_0 : \theta = \theta_0)$



7. For testing $H_0 : \mu = \mu_0$ vs $H_1 : \mu < \mu_0$ at an $\alpha = 0.025$ level, if we observe $\bar{x}_{obs} = 36.85$ ($n = 16$), what are the range of μ_0 values that would NOT be rejected? (I.e. find a one-sided confidence interval for μ.)

$A_{0.025} = \{\underline{x} : \bar{x} < 36.8\}$ for any $\mu < \mu_0$ and $0.025 = \Pr(\bar{X} < 36.8 \mid H_0 : \mu = 37)$ (for $\mu_0 = 37$).

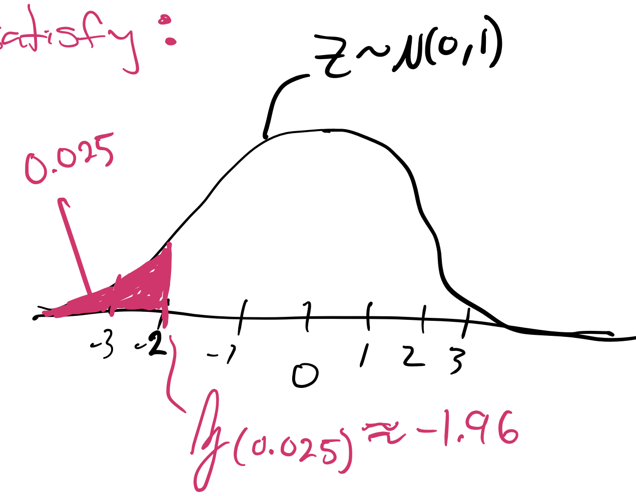
We must find which other μ_0 values satisfy:

$0.025 = \Pr(\bar{X} < 36.8 \mid H_0 : \mu = \mu_0)$. Since

$$0.025 = \Pr(\bar{X} - \mu_0 < 36.8 - \mu_0 \mid H_0 : \mu = \mu_0)$$

$$= \Pr\left(\frac{\bar{X} - \mu_0}{0.1} < \frac{36.8 - \mu_0}{0.1} \,\Big|\, H_0 : \mu = \mu_0\right)$$

$$= \Pr\left(Z < \frac{36.8 - \mu_0}{0.1}\right)$$



We have that $\Pr\left(\frac{\bar{X} - \mu_0}{0.1} \le \Phi(0.025)\right) = 0.025$ if H_0 is true.

ie. $\mu_0 \ge \bar{X} - 0.1\,\Phi(0.025)$ is a 1-sided 97.5% CI for μ.

Topic: Hypothesis Testing Part III

10-24-22

Re-cap

Simple H_0 vs simple H_1

The NP lemma concludes that for testing $\begin{cases} H_0: \theta=\theta_0 \\ H_1: \theta=\theta_1 \end{cases}$, $\theta_0 \neq \theta_1$

the most powerful level-α test is the likelihood ratio test.

Test statistic: $\Lambda = \dfrac{L(\theta_0)}{L(\theta_1)}$ where $L(\theta) = f((X_1, \dots, X_n); \theta)$ is the likelihood for θ.

Rejection Region: $A_\alpha = \left\{ \underset{\sim}{X} : \dfrac{L(\theta_0)}{L(\theta_1)} < C_\alpha \right\}$

where C_α is such that $Pr\left(\underset{\sim}{X} \in A_\alpha \mid H_0: \theta=\theta_0\right) = \alpha$

Error Probabilities:

$$\alpha = Pr(\text{Type I error}) = Pr\left(\frac{L(\theta_0)}{L(\theta_1)} < C_\alpha \,\Big|\, H_0: \theta=\theta_0\right) = \alpha$$

$$\beta = Pr(\text{Type II error}) = Pr\Big(\underbrace{\frac{L(\theta_0)}{L(\theta_1)} > C_\alpha}_{\underset{\sim}{X} \notin A_\alpha} \,\Big|\, H_1: \theta=\theta_1\Big)$$

Thus, power $= 1-\beta = Pr\Big(\underbrace{\underset{\sim}{X} \in A_\alpha}_{\frac{L(\theta_0)}{L(\theta_1)} < C_\alpha} \,\Big|\, H_1: \theta=\theta_1\Big)$

Finally, a p-value is a summary of the strength of evidence against H_0:

$$\text{p-value} = \cancel{\begin{cases} Pr(\Lambda > \Lambda_{obs} \mid H_0: \theta=\theta_0), & \text{if } \theta_1 > \theta_0 \\ Pr(\Lambda < \Lambda_{obs} \mid H_0: \theta=\theta_0), & \text{if } \theta_1 < \theta_0 \end{cases}}$$

Incorrect!

In general:

$$\boxed{\text{p-val} = Pr(\Lambda \leq \Lambda_{obs} \mid H_0: \theta=\theta_0)}$$

correct!

Simple H_0 vs Composite H_1

1)

Since A_α does not change for any other value of Θ_i that lies on the same side of Θ_0 as Θ, in the tests of simple H_0 vs simple H_1, the uniformly most powerful test of $\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta > \theta_0 \end{cases}$ or $\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta < \theta_0 \end{cases}$ is the likelihood ratio test.

As before,

$$\alpha = Pr(\text{Type I error}) = Pr(\underset{\sim}{X} \in A_\alpha \mid H_0: \theta = \theta_0)$$

But β and the power may vary for different parameter values in H_1:

$$\text{eg. } \beta = Pr(\text{Type II error}) = Pr(\underset{\sim}{X} \notin A_\alpha \mid H_1: \theta > \theta_0)$$

$$\text{power} = Pr(\underset{\sim}{X} \in A_\alpha \mid H_1: \theta > \theta_0)$$

2)

For a two-sided alternative $\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta \neq \theta_0 \end{cases}$ there no uniformly most powerful test.

However, we can derive a level-α test by finding $A_{\alpha/2} \cup A'_{\alpha/2}$, where $A_{\alpha/2}$ is the level-$\alpha/2$ RR for $\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta > \theta_0 \end{cases}$ and $A'_{\alpha/2}$ is the level-$\frac{\alpha}{2}$ RR for $\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta < \theta_0 \end{cases}$.

Confidence Intervals

For significance level $\alpha \in (0,1)$, we call $(1-\alpha)$ the "confidence level".

Given data $(X_1, \dots, X_n) \sim f((x_1, \dots, x_n); \theta)$,
a (say) 95% confidence interval (CI) is a random interval that contains the value of θ that is associated w/ the observed data, $\underset{\sim}{X}_{obs}$, 95% of the time.

i.e. the "true" value of θ

(In general, the same is not true for Bayesian credible intervals.)

1-Sided CIs

A 1-sided $(1-\alpha)100\%$ CI for θ has one fixed bound (usually $\pm\infty$ or 0) and one random bound.
One way to create a 1-sided $(1-\alpha)100\%$ CI for θ is to invert a level-α 1-sided hypothesis test.
If A_α is the rejection region for this level-α hypothesis test, then the set of all θ_0 values that would NOT be rejected for the observed data, $\underset{\sim}{X}_{obs}$, forms a $(1-\alpha)100\%$ CI for θ.

2-Sided CIs

The intersection of two 1-sided $(1-\frac{\alpha}{2})100\%$ CIs forms a $(1-\alpha)100\%$ CI.

The two 1-sided intervals imply a random upper and lower bound, C_u and C_L such that

$$Pr(C_u > \theta) = Pr(C_L < \theta) = \frac{\alpha}{2}.$$

Assuming $C_u > C_L$, the intersection is non-empty and we have

$$Pr(C_L < \theta \leq C_u) = Pr(C_u \leq \theta) - Pr(C_L < \theta) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha.$$

Note

To construct a 2-sided $(1-\alpha)100\%$ CI, first construct two 1-sided $(1-\frac{\alpha}{2})100\%$ CIs and then find their intersection. Although this may not yield the narrowest possible interval w/ the desired coverage rate, this method does ensure that, in the case where the interval fails to contain the true value of θ, it is equally likely that it missed high or missed low, which is a desirable property for an interval estimate.

Generalized LHR Test

For composite H_1, we can use a general version of the LHR test where the test statistic is now

$$\Lambda = \frac{\max_{\theta \in \omega_0} L(\theta)}{\max_{\theta \in \Omega} L(\theta)}$$

read as "the max value of the likelihood for all $\theta \in \omega_0$."

where $\Omega = \Theta$ is the entire parameter space and
$\omega_0 \subseteq \Theta$ is the subspace specified by H_0,
and the rejection region, A_α, consists of <u>small</u> values of Λ.

[Recall, $\hat{\theta}_{MLE}$ is the maximizer of $L(\theta)$ for $\theta \in \Omega$.

Ex) $X_1, \dots, X_n \overset{IID}{\sim} N(\mu, 0.4^2)$ test $\begin{cases} H_0 : \mu = 37 \\ H_1 : \mu \neq 37 \end{cases}$ at α-level.

First, $L(\theta) \propto \exp\left\{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2(0.16)}\right\}$ and

$$\begin{cases} \Omega = (-\infty, \infty) \\ \omega_0 = \{37\} \end{cases} \quad \text{and} \quad \hat{\mu}_{MLE} = \bar{X}.$$

Then we have $\Lambda = \exp\left\{\frac{-1}{2(0.16)}\left[\sum_{i=1}^n (X_i - 37)^2 - \sum_{i=1}^n (X_i - \bar{X})^2\right]\right\}$

and $A_\alpha = \{\underline{X} : \Lambda < c_1\} \equiv \{\underline{X} : \sum_{i=1}^n (X_i - 37)^2 - \sum_{i=1}^n (X_i - \bar{X})^2 > c_2\}$

where c_2 is such that $\Pr\left(\sum_{i=1}^n (X_i - 37)^2 - \sum_{i=1}^n (X_i - \bar{X})^2 > c_2 \mid H_0 : \mu = 37\right) = \alpha.$

Theorem: Asymptotic Dist'b'n of the Gen. LHR

Provided the joint density (or mass) function $f((x_1,\dots,x_n);\theta)$ is "smooth enough",

$$-2\ln(\Lambda) \overset{H_0}{\sim} \chi^2_{(m)} \quad \text{as} \quad n\to\infty,$$

where $m = \mathrm{Dim}(\Omega) - \mathrm{Dim}(\omega_0)$.

Useful Identity: $\displaystyle\sum_{i=1}^{n}(X_i-\mu_0)^2 = \left(\sum_{i=1}^{n}(X_i-\bar{X})^2\right) + n(\bar{X}-\mu_0)^2$

Proof: See supplementary material for 10-19-22.

Ex cont'd)

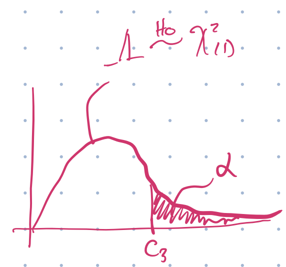
$$A_\alpha = \left\{\underset{\sim}{x} : \sum_{i=1}^{n}(X_i-37)^2 - \sum_{i=1}^{n}(X_i-\bar{X})^2 > c_2\right\}$$

$$= \left\{\underset{\sim}{x} : \sum_{i=1}^{n}(X_i-\bar{X})^2 + n(\bar{X}-37)^2 - \sum_{i=1}^{n}(X_i-\bar{X})^2 > c_2\right\}$$

$$= \left\{\underset{\sim}{x} : n(\bar{X}-37)^2 > c_2\right\} \equiv \left\{\underset{\sim}{x} : \frac{(\bar{X}-37)^2}{(0.16/n)} > c_3\right\}.$$

Recall, if $\bar{X}\sim N\left(\mu,\frac{\sigma^2}{n}\right)$ then $\frac{(\bar{X}-\mu)}{(\sigma/\sqrt{n})}\sim N(0,1)$ and so

$$\frac{(\bar{X}-\mu)^2}{\left(\frac{\sigma^2}{n}\right)} \sim \chi^2_{(1)}.$$



$\therefore\ c_3 = \text{qchisq}(1-\alpha,\ df=1,\ \text{lower.tail}=T)$

· Review Worksheet Solutions

Note: For two-sided H_1, we will use the generalized LHR test.

Ex: Two-Sided Test

Suppose $X \sim Bin(n,p)$ and test $\begin{cases} H_0 : p = \frac{1}{2} \\ H_1 : p \neq \frac{1}{2} \end{cases}$ at an $\alpha = 0.05$ significance level. Suppose $n = 10$.

First note that $L(p) = \binom{n}{x} p^x (1-p)^{n-x}$

and that $\hat{p}_{MLE} = \frac{X}{n}$. Also, $\Omega = [0,1]$ and $\omega_0 = \{\frac{1}{2}\}$.

Our test statistic is

$$\Lambda = \frac{\max_{p \in \{\frac{1}{2}\}} L(p)}{\max_{p \in \Omega} L(p)} = \frac{L(\frac{1}{2})}{L(\frac{X}{n})} = \frac{\binom{n}{X} \frac{1}{2}^X (1-\frac{1}{2})^{n-X}}{\binom{n}{X} (\frac{X}{n})^X (1-\frac{X}{n})^{n-X}}$$

$$= \ldots = \left(\frac{n}{2X}\right)^X \left(\frac{n}{2(n-X)}\right)^{n-X}$$

and our rejection region is

$$A_\alpha = \left\{ X : \left(\frac{n}{2X}\right)^X \left(\frac{n}{2(n-X)}\right)^{n-X} < c_\alpha \right\}$$

where c_α satisfies $Pr\left(\left(\frac{n}{2X}\right)^X \left(\frac{n}{2(n-X)}\right)^{n-X} < c_\alpha \,\middle|\, p = \frac{1}{2} \right) = 0.05$

$\boxed{\text{(highlight)} = \text{there's work you must show here}}$

To solve for an explicit rejection region note that

$$\Lambda = \left(\frac{n}{2x}\right)^x \left(\frac{n}{2(n-x)}\right)^{n-x} = \cdots = \underbrace{\left(\frac{n}{2}\right)^n}_{\text{const}} \frac{1}{x^x (n-x)^{n-x}}$$

So $A_\alpha = \left\{x : \left(\frac{n}{2x}\right)^x \left(\frac{n}{2(n-x)}\right)^{n-x} < c_\alpha\right\}$

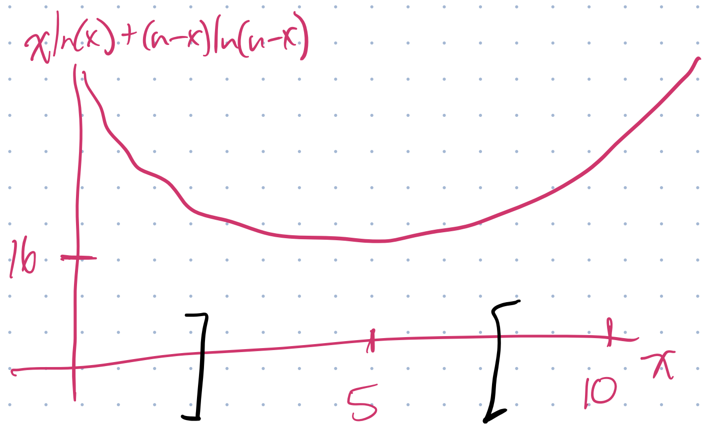
$$= \left\{x : \frac{1}{x^x (n-x)^{n-x}} < c'_\alpha\right\}$$

$$= \left\{x : \ln(\cancel{1}) - [x\ln(x) + (n-x)\ln(n-x)] < c'_\alpha\right\}$$

$$= \left\{x : \underbrace{x\ln(x) + (n-x)\ln(n-x)} > c''_\alpha\right\}$$

plot this! $n=10$

$$A_\alpha = \left\{x : \left|x - \frac{n}{2}\right| > k_\alpha\right\}$$



Takeaway:

$\left(\frac{n}{2x}\right)^x \left(\frac{n}{2(n-x)}\right)^{n-x}$ is small when x is far from $\frac{n}{2}$

Now we can explicitly solve for an $\alpha = 0.05$ rejection region:

$$0.05 = \Pr\left(\left|X - \tfrac{n}{2}\right| > k \,\middle|\, H_0: p = \tfrac{1}{2}\right)$$
$$= \Pr\left(X - \tfrac{n}{2} > k \text{ or } X - \tfrac{n}{2} < -k \,\middle|\, H_0: p = \tfrac{1}{2}\right)$$
$$= \Pr\left(X > k + \tfrac{n}{2} \,\middle|\, p = \tfrac{1}{2}\right) + \Pr\left(X < \tfrac{n}{2} - k \,\middle|\, p = \tfrac{1}{2}\right)$$

Which implies

$$\begin{cases} k + \frac{n}{2} = \text{qbinom}\left(1 - \frac{0.05}{2}, n=10, p=\frac{1}{2}, \text{lower.tail} = T\right) \\ \frac{n}{2} - k = \text{qbinom}\left(\frac{0.05}{2}, n=10, p=\frac{1}{2}, \text{lower.tail} = T\right) \end{cases}$$

Ie. $k = 3$, And finally,

$$A_\alpha = \left\{x : \left|x - \tfrac{n}{2}\right| > 3\right\}$$

$\blacksquare$

Ex cont'd: Two-Sided CI $(n=10)$

Find a 95% CI for p, supposing $x_{obs}=3$.

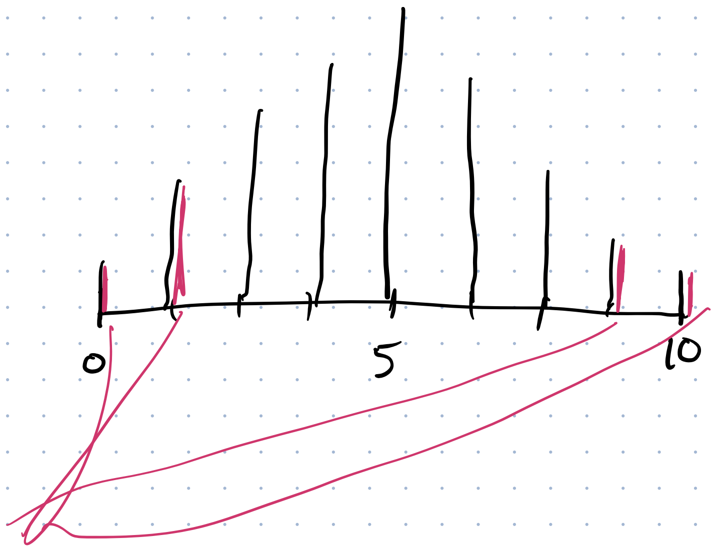
By definition of α:

$$0.05 = Pr\left(|X-\tfrac{n}{2}| > 3 \mid H_0: p=\tfrac{1}{2}\right)$$

$$= Pr\left(X > 3+\tfrac{n}{2} \mid p=\tfrac{1}{2}\right) + Pr\left(X < \tfrac{n}{2}-3 \mid p=\tfrac{1}{2}\right)$$

$$= Pr\left(X > 8 \mid p=\tfrac{1}{2}\right) + Pr\left(X < 2 \mid p=\tfrac{1}{2}\right)$$

$$X \overset{H_0}{\sim} Bin(10, \tfrac{1}{2})$$



All red mass functions sum to 0.05

thus all black mass functions sum to 0.95

and a 95% CI for p is $\left(\frac{2}{10}, \frac{8}{10}\right)$

↑ uninclusive

or, $\left[\frac{3}{10}, \frac{7}{10}\right]$.

↖ inclusive

∎

Ex)

Suppose $X \sim Bin(n,p)$ and test $\begin{cases} H_0: p = \frac{1}{2} \\ H_1: p \neq \frac{1}{2} \end{cases}$ at an $\alpha = 0.05$ significance level. Suppose $\underline{n = 170}$, $\underline{x_{obs} = 63}$.

$$Pr\left(\left|X - \frac{n}{2}\right| > 3 \,\middle|\, H_0: p = \tfrac{1}{2}\right) = \alpha = 0.05$$

So, $$0.05 = Pr\left(X - \frac{n}{2} > 3 \text{ or } x - \frac{n}{2} < -3 \,\middle|\, H_0: p = \tfrac{1}{2}\right)$$
$$= Pr\left(X > 3 + \frac{n}{2} \,\middle|\, H_0\right) + Pr\left(X < \frac{n}{2} - 3 \,\middle|\, H_0\right)$$

Since $X = \sum_{i=1}^{n} Y_i$, where $Y_i \overset{iid}{\sim} Bern(p)$, by the CLT, as $n \to \infty$ we have

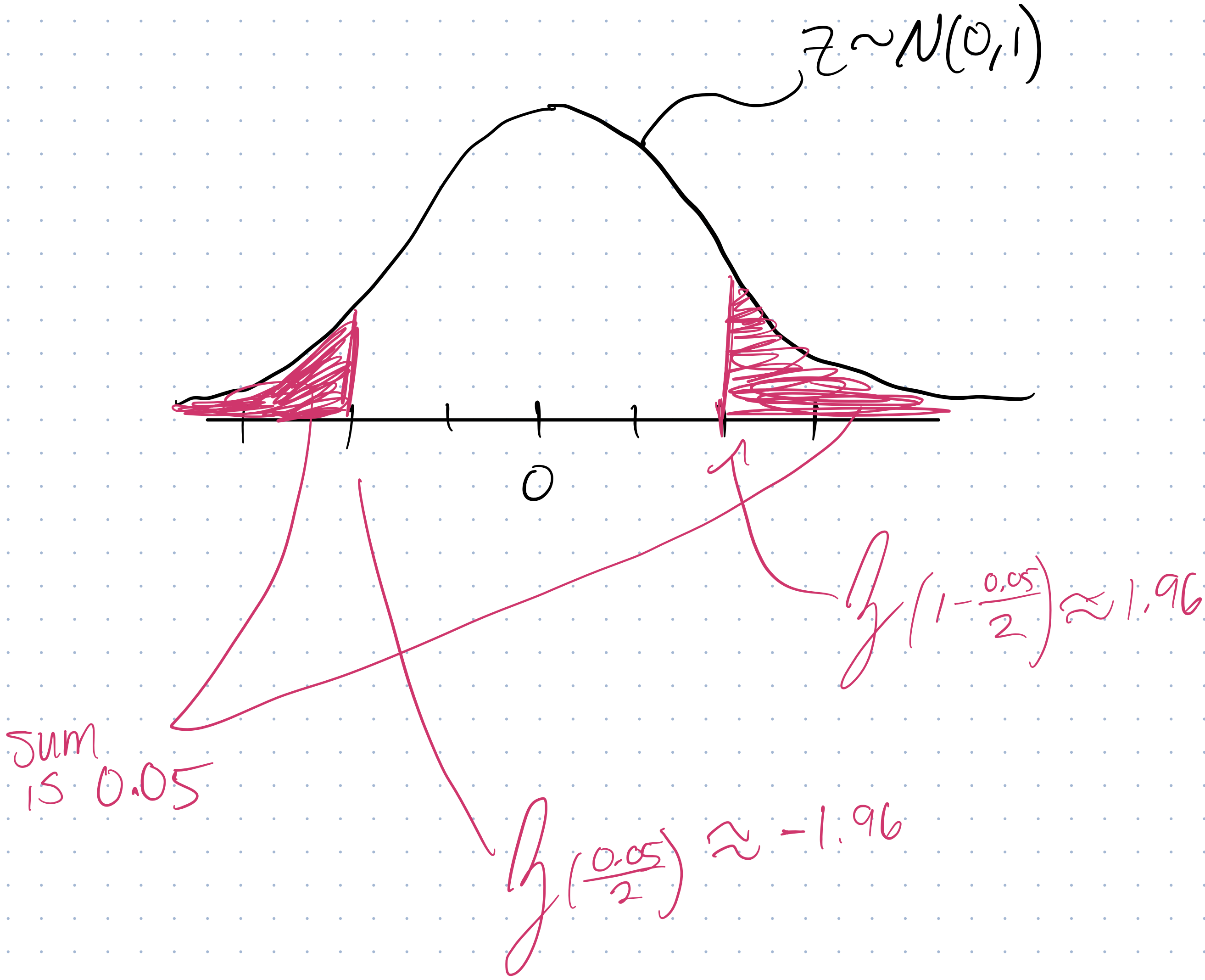
$$\frac{\overline{Y} - p}{\sqrt{\frac{p(1-p)}{n}}} \overset{n\to\infty}{\sim} N(0,1)$$

ie. $$\frac{\frac{1}{n}X - p}{\sqrt{\frac{p(1-p)}{n}}} \overset{n\to\infty}{\sim} N(0,1)$$

doesn't depend on any unknowns!

So $$0.05 = Pr\left(\frac{\frac{1}{n}X - p}{\sqrt{p(1-p)/n}} > \frac{\frac{1}{n}\left(3 + \frac{n}{2}\right) - p}{\sqrt{p(1-p)/n}} \,\middle|\, H_0: p = \tfrac{1}{2}\right)$$
$$+ Pr\left(\frac{\frac{1}{n}X - p}{\sqrt{p(1-p)/n}} < \frac{\frac{1}{n}\left(\frac{n}{2} - 3\right) - p}{\sqrt{p(1-p)/n}} \,\middle|\, H_0: p = \tfrac{1}{2}\right)$$

$$\overset{H_0}{=} Pr\left(Z > \frac{\frac{1}{n}\left(3 + \frac{n}{2}\right) - p}{\sqrt{p(1-p)/n}}\right) + Pr\left(Z < \frac{\frac{1}{n}\left(\frac{n}{2} - 3\right) - p}{\sqrt{p(1-p)/n}}\right)$$



By definition of quantiles we have

$$0.05 = \Pr\left(Z > z_{(1-\alpha/2)}\right) + \Pr\left(Z < z_{(\alpha/2)}\right)$$

$$= \Pr\left(\frac{\frac{1}{n}X - p_0}{\sqrt{p_0(1-p_0)/n}} > z_{(1-\alpha/2)} \,\middle|\, H_0: p = p_0\right)$$

$$+ \Pr\left(\frac{\frac{1}{n}X - p_0}{\sqrt{p_0(1-p_0)/n}} < z_{(\alpha/2)} \,\middle|\, H_0: p = p_0\right)$$

Hence we will fail to reject $H_0: p = p_0$

if $\dfrac{\frac{1}{n} X_{obs} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} > 1.96$

or if $\dfrac{\frac{1}{n} X_{obs} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} < -1.96$

By rearranging the above (takes some work and analysis of terms) we plug in $X_{obs} = 63$ and find a 95% (asymptotic) CI for p to be

$$[0.00698, 0.7399].$$

See the next two pages for additional details. This pg. & the previous are the most important to understand.

Since we FtR when $\dfrac{\frac{1}{n}X_{obs} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} > 1.96$

ie. $\dfrac{63}{170} - p_0 > 1.96\sqrt{\dfrac{p_0(1-p_0)}{170}}$

ie. $\left(\dfrac{\frac{63}{170} - p_0}{1.96}\right)^2 > \dfrac{p_0(1-p_0)}{170}$

ie. $\dfrac{\left(\frac{63}{170} - p_0\right)^2}{p_0(1-p_0)} > \dfrac{(1.96)^2}{170}$

ie. $\dfrac{\left(\frac{63}{170}\right)^2 - \frac{2(63)}{170}p_0 + p_0^2}{p_0 - p_0^2} > \dfrac{1.96^2}{170}$

ie. $\left(\dfrac{63}{170}\right)^2 - \dfrac{2(63)}{170}p_0 + p_0^2 > \dfrac{1.96^2}{170}p_0 - \dfrac{1.96^2}{170}p_0^2$

ie. $\left(\dfrac{63}{170}\right)^2 > p_0^2\left(-1 - \dfrac{1.96^2}{170}\right) + p_0\left(\dfrac{2(63)}{170} + \dfrac{1.96^2}{170}\right)$

ie. $0 > p_0^2\left(-1 - \dfrac{1.96^2}{170}\right) + p_0\left(\dfrac{2(63)}{170} + \dfrac{1.96^2}{170}\right) - \left(\dfrac{63}{170}\right)^2$

is a quadratic fnctn of p_0

$$\begin{cases} a = -1 - \dfrac{1.96^2}{170} \\ b = \dfrac{2(63)}{170} + \dfrac{1.96^2}{170} \\ c = -\left(\dfrac{63}{170}\right)^2 \end{cases}$$

Thus $p_0 < \dfrac{-\left(\frac{2\cdot 63}{170} + \frac{1.96^2}{170}\right) \pm \sqrt{4\left(-1 - \frac{1.96^2}{170}\right)\left(-\left(\frac{63}{170}\right)^2\right)}}{2\left(-1 - \frac{1.962}{170}\right)}$

Then, do the same analysis for

$\dfrac{\frac{1}{n}X_{obs} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} < -1.96$ and solve for CI bands. ∎