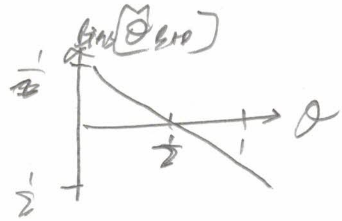


$X_1, \dots, X_n \overset{iid}{\sim} \text{Bern}(\theta)$. You want point estimator for θ.

$\hat{\theta} = \frac{1}{n} \sum X_i$ vs. $\hat{\theta}_{BAD} = \frac{1}{2}$

Common sense should tell you that always guessing the same single value is a <u>bad</u> idea. But let's see if our theoretical framework can show why it's bad.

First of all, $\hat{\theta}$ is unbiased.



$\text{Bias}[\hat{\theta}_{BAD}] = E[\hat{\theta}_{BAD}] - \theta = \frac{1}{2} - \theta$

$\Rightarrow \hat{\theta}_{BAD}$ is biased for almost all values of $\theta \in \Theta$. This is bad.

How about risk? Mean loss. If loss = sqd-err loss, risk = MSE.

$MSE[\hat{\theta}] = \frac{\theta(1-\theta)}{n}$ from lec 2

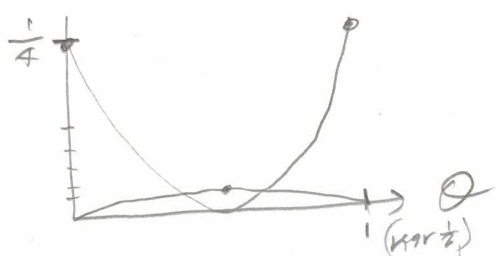
$MSE[\hat{\theta}_{BAD}] = Var[\hat{\theta}_{BAD}] + \text{Bias}[\hat{\theta}_{BAD}]^2 = 0 + \left(\frac{1}{2} - \theta\right)^2 = \left(\theta - \frac{1}{2}\right)^2$

Note: as $n \uparrow$, $MSE[\hat{\theta}] \downarrow$ but $MSE[\hat{\theta}_{BAD}]$ doesn't $\downarrow$! This is bad!

You want your estimator to get more accurate as your sample grows.

Let's graph for $n = 10$

Estimator Theory is really cool but unfortunately there is no more time for it



There is a small region of θ where $\hat{\theta}_{BAD}$ is more accurate than $\hat{\theta}$ (and this region shrinks as n increases). This is usually the case. Estimators cannot be expected to be the best everywhere. So how to compare ~~two~~ MSE curves? One option is the "maximal risk" metric i.e. $\sup_{\theta \in \Theta} \{R(\hat{\theta}, \theta)\}$. In our case

$\sup_{\theta \in (0,1)} \{MSE[\hat{\theta}_{BAD}]\} = \frac{1}{4}$ vs. $\sup_{\theta \in (0,1)} \{MSE[\hat{\theta}]\} = \frac{1}{4n}$ so $\hat{\theta}$ beats $\hat{\theta}_{BAD}$.

Are UFO's real? Big mystery. Nobody truly knows!
⇒ No way to convince you either way absolutely and unequivocally.

Let's say I want to convince ~~y~~ you that UFO's are real.

Two strategies

(I) I assert I'm right and you all have to disprove me. And by sheer force of charisma and you not being able to convince me that they are not real, you become convinced.

(II) I temporarily assume I'm wrong and then show you evidence. If I show you enough evidence, eventually you'll be convinced. Akin to proof by contradiction

Which strategy is better to convince?

II. It's more intellectually honest (I can't trick you). I will win more believers. It's more honest especially if I have a vested, monetary interest in convincing people I'm right. If I'm changing the status quo, this is the only way it works.

Goal #2 of statistical inference: theory testing.
This is a logical structure to prove a fundamental theory about a DGP.

For example, I would like to convince you, that the mean prop. of iPhones in CUNY 300-level STEM classes (our assumed pop. of interest from last week's survey), θ is **not** the American average of 52.4%.

A hypothesis is a mathematical statement about a pop. param of interest e.g. $\theta = .524$ or $\theta > .524$ or $\theta < .524$, $\theta \neq .524$ etc.

The "alternative hypothesis" H_a is what you'd like to prove and the null hypothesis H_0 is the opposite of H_a. By the logical argument from before, we first assume H_0, then introduce evidence, $\hat{\theta}$ which hopefully is "significant" enough to convince you that H_0 is wrong, so you accept H_a.

Typical tests in a Stat 101 class

$H_0 : \theta \leq \theta_0 \iff H_a : \theta > \theta_0$ right-tailed or right-sided test

$H_0 : \theta \geq \theta_0 \iff H_a : \theta < \theta_0$ left-tailed or left-sided test

$H_0 : \theta = \theta_0 \iff H_a : \theta \neq \theta_0$ two-tailed or two-sided test

$H_0 : \theta \neq \theta_0 \iff H_a : \theta = \theta_0$ equivalence test (NOT COVERED)

Upon doing a hypothesis test, there are two results:

(I) There is "statistically significant" evidence against H_0 $\Rightarrow$ "H_a is accepted" or "H_0 is rejected" you are convinced

(II) There is not stat sig evidence against H_0 $\Rightarrow$ "H_0 is retained" or "we fail to reject H_0". You are <u>not</u> convinced.

How ~~do~~ we decide the outcome of the test?

First note there's many different types of tests, each with different subtleties and assumptions. They don't always agree!!!

All tests examine a "test statistic". The test statistic gauges departure from H_0's sampling distribution.

First example: I want to prove the iPhone mean prop is different than the Android mean of .524. What are the hypotheses?

$H_a: \theta \neq .524 \iff H_0: \theta = .524$

Test statistic? $\hat{\theta} = \frac{1}{n} \sum x_i = \frac{10}{21} = .476$

~~It is~~ clear $.476 \neq .524$. Should we reject H_0?

No, because $\hat{\theta}$ may not be significantly weird if $\theta = .524$ because the sampling distr of $\hat{\theta} \mid H_0$ may expect $\hat{\theta}$.

1:1 function $t \iff \hat{\theta}$

What is $\hat{\theta} \mid H_0$? Well $\hat{\theta} = \frac{1}{21}\underbrace{(X_1 + \ldots + X_{21})}_{\text{we assumed } \overset{iid}{\sim} \text{Bern}(.524)} \iff 21\hat{\theta} = X_1 + \ldots + X_{21} \sim \text{Binom}(21, .52$

$$= \binom{21}{21\hat{\theta}} .524^{21\hat{\theta}} (1-.524)^{1-}$$

Let's plot it

What is reasonable given H_0 is true?

What is "statistically significantly" weird?

$RET = \{ \cdots \}$

This is called 2-sided "Binomial ~~Exact~~ Test"

we used the binomial distr

we knew the exact PMF

of one binomial proportion

θ

Now we examine the test statistic. If ..

$\hat{\hat{\theta}} \notin RET \Rightarrow$ Reject H_0

$\hat{\hat{\theta}} \in RET \Rightarrow$ Retain H_0

Since $\hat{\hat{\theta}} \in RET$, we retain H_0 and conclude "there is not significant evidence to conclude the true iPhone proportion is not .527"

Does this mean we accept $H_0: \theta = .527$! No. It just means it's a good enough theory for now. We wouldn't reject $H_0: \theta = .527001$ either nor infinite other H_0's. This is actually how science works. We had $F = ma$, Newton's Laws for 300+ yr before Einstein showed they're wrong

Let's say it came out differently, we rejected H_0.

Is it possible we can make a mistake when rejecting H_0?

Yes. The REJ set still is possible but unlikely.

$$P(\hat{\theta} \in REJ) = P(\hat{\theta} = 0) + \ldots + P(\hat{\theta} = \quad) + P(\hat{\theta} = \quad) + \ldots + P(\hat{\theta} = 1)$$

$$= \qquad\qquad = P(\text{Type I error}) = \text{"size" of a test}$$

If you're going to run a test that rejects H_0 "size" of the time, you have to be okay with being wrong "size" of the time

Def of level: if a test has size $\geq$ level $= \alpha$, the test is said to have "level α". $\alpha = 5\%$? ~~yes~~

$\alpha = 1\%$? yes

Can size be 5%? No.

Can size be 1%? No.

Not all sizes possible if sampling distr. discrete

CDF of Binomial

$$\text{size} \in \left\{ P(\hat{\theta} \leq L) + P(\hat{\theta} \geq U) = F_B(21L, 21, .524) + (1 - F_B(21(U-1), 21, .524)) : L < U,\ L, U \in \{0, 1, \ldots, 21\} \right\}$$

We retained H_0. Is it possible we made a mistake?

Yes. We should've rejected H_0 since it is truly false.

This is a Type II error. $P(\text{Type II err})$ = complicated

		Decision	
		Retain H_0	Reject H_0
Unknown Truth	H_0 true	✓	Type I err
	H_0 false	Type II err	✓

You control max prob type I error via picking size.

You indirectly control prob type II err through size as well.

$$P(\text{Type I err}) \downarrow \iff P(\text{Type II err}) \uparrow$$

$$P(\text{Type II err}) \downarrow \iff P(\text{Type I err}) \uparrow$$

Cosmic whack-a-mole

Scientific convention: $\alpha = 5\%$ or $\alpha = 1\%$.

This's considered "statistically significant". We are conservative to retain H_0 since we don't want to accept crazy theories