18.650
Statistics for Applications

Chapter 5: Parametric hypothesis testing

* hypothesis testing have little more of a Statistical flavor, when it comes to designing method's.

* most of the statistical analysis or use of the Central Limit theorem, all these thing's actually did not come in the building of the estimator's, in the designing of the Estimator, but really in the analysis of the Estimator.

* hypothesis testing is on the same Level as estimation. And on the same level as estimator will be the word "test".

* when we are going to devise a test, we are gonna actually need to understand random fluctuation's that arise from CLT better.

* It's not Just gonna to be in the analysis, it also gonna be in the design, and everything we have been doing before in understanding the behavior of an estimator is actually gonna come in and be extremely useful in actual design of tests.

Cherry Blossom run (1)

- The credit union Cherry Blossom Run is a 10 mile race that takes place every year in D.C.
- In 2009 there were 14974 participants
- Average running time was 103.5 minutes. (Actual value)
(True value)

Were runners faster in 2012?

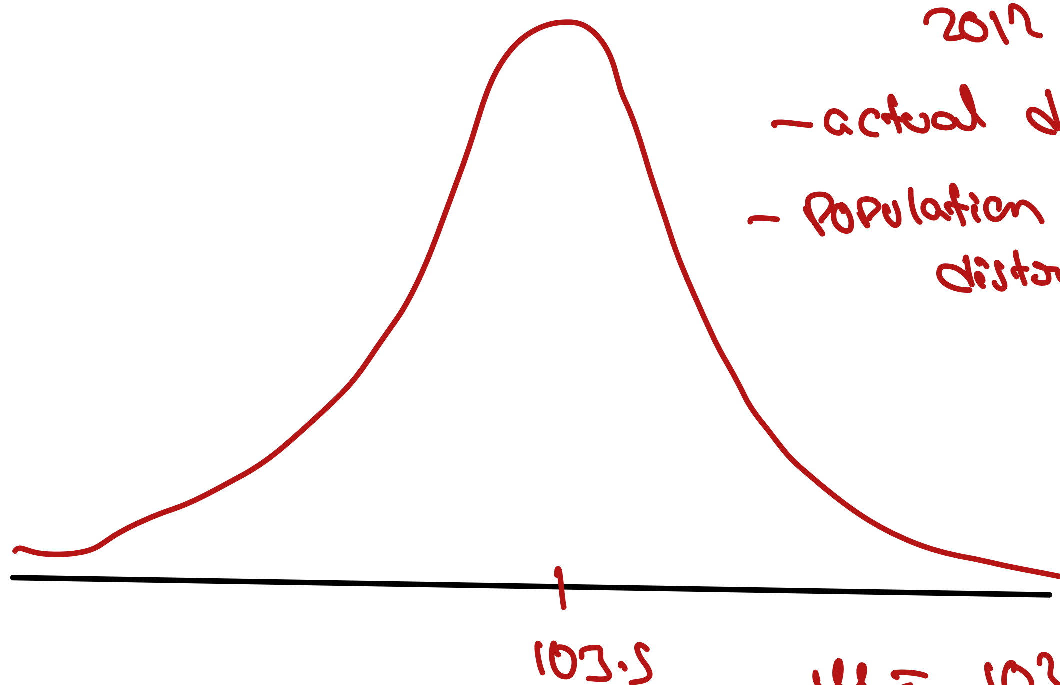
To answer this question, select n runners from the 2012 race at random and denote by $X_1, \ldots, X_n$ their running time.

Cherry Blossom run (2)

We can see from past data that the running time has Gaussian distribution.

2012

- actual distributi
- population distribution



The variance was 373.

$\mu = 103.5$

$\sigma^2 = 373$

Cherry Blossom run (3)

- We are given i.i.d r.v $X_1, \ldots, X_n$ and we want to know if $X_1 \sim \mathcal{N}(103.5, 373)$
- This is a **hypothesis testing** problem.
- There are many ways this could be false:
  1. $\mathbb{E}[X_1] \neq 103.5$
  2. var$[X_1] \neq 373$
  3. X_1 may not even be Gaussian.
- We are interested in a very specific question: is $\mathbb{E}[X_1] < 103.5$?

$x_1, x_2, \ldots, x_n \sim N(\mu, \sigma^2)$

2012

$\Rightarrow$ we want to know, is the distribution the same from past year's. we want to know the random variable that we are looking for X_1, does it have the same distribution in 2012? that it did in 2009?

$\Rightarrow$ The question is, is X_1 has Gaussian distri- with $\mu = 103.5$, $\sigma^2 = 373$? in 2012.

* So This question that call's for a yes or No answer is a hypothesis testing Problem.

* This is the Basis of basically all of Data driven Scientific inquiry. we Just ask question's. we formulate a scientific hypothesis, knocking down this gene is going to cure melanoma, is this true?

$\Rightarrow$ we are gonna observe some patient's on which I knock down this gene. we are gonna collect some measurements. and we are gonna try to answer this yes / No question.

It's different from the question, what is the mean running time for this year (2012)?

* This hypothesis testing is testing if this hypothesis is true? The hypothesis in common english we just said, were runner's running faster?

* What we have to do is to factor in as much knowledge as you can & make as many modeling assumption's, so that you can let the data speak about your particular question. Here the particular question is, Are the runner's running faster? ⟹ we are really asking a question about the expectation. we really wanna know if the $\mathbb{E}$ changed?

⟹ So as far as we are concerned, we are happy to assumption that the rest has been unchanged.

We are interested in a very specific Question: is $\mathbb{E}[X_i] < 103.5$? ⟹ But first we need to fix the rest. And fixing the rest is actually part of modeling Assumption.

Cherry Blossom run (4)

- We make the following **assumptions**:
  1. $\text{var}[X_1] = 373$ (variance is the same between 2009 and 2012)
  2. X_1 is Gaussian.
- The only thing that we did not fix is $\mathbb{E}[X_1] = \mu$.
- Now we want to test (only): "Is $\mu = 103.5$ or is $\mu < 103.5$"?
- By making **modeling assumptions**, we have reduced the number of ways the hypothesis $X_1 \sim \mathcal{N}(103.5, 373)$ may be rejected.
- The only way it can be rejected is if $X_1 \sim \mathcal{N}(\mu, 373)$ for some $\mu < 103.5$.
- We compare an expected value to a fixed reference number (103.5).

Where does this $X_i \sim$ Gaussian Assumption come from? It's Just natural, when we Just measure a bunch of things. The CLT appear's in the small thing's of Nature. $\Rightarrow$ that's why the Bell Shaped Curve is there everywhere.

Question:

"is $\mu = 103.5$ or is $\mu < 103.5$"?

* By making all these modeling Assumption's, the fact that variance doesn't changed, the fact that its still Gaussian, we have actually reduced the number of way's the hypothesis can be violated. The # Possible alternative realities for this hypothesis

* μ could be > 103.5, but we actually Just going to say that if its larger. all we going to be able to tell is that its not Smaller. we are not gonna able to tell that its actually larger.

The only way it can be rejected now, if x belong's to very specific family of distributi..
if it hae a distribution Gaussian mean μ and variance $\sigma^2 = 373$, for $\mu < 103.5$

— only Gaussian distribution's

— All possible distribution's of X

Cherry Blossom run (5)

Simple heuristic:

"If $\bar{X}_n < 103.5$, then $\mu < 103.5$"

This could go wrong if I randomly pick only fast runners in my sample $X_1, \ldots, X_n$.

Better heuristic:

"If $\bar{X}_n < 103.5 -$(something that $\xrightarrow[n\to\infty]{} 0$), then $\mu < 103.5$"

To make this intuition more precise, we need to take the size of the random fluctuations of $\bar{X}_n$ into account!

$\bar{X}_n$ will be close to true mean by, LLN , so we are gonna decide whether $\bar{X}_n < 103.5$

$$\Downarrow$$

$$\mu < 103.5$$

because $\bar{X}_n$, μ are close. we could do that , the Problem is that this could go Pretty wrong. Because if n is small, then we know $\bar{X}_n$ is not equal to μ. we know $\bar{X}_n$ is close to μ , But also know that there Pretty high chance that its not equal to μ, in Particular, we know its going to be somewhere at $\frac{1}{\sqrt{n}}$ away from mu.

Hidden truth (2012) $\mu = 103.4$ (we are not supposed to know this.

$n = 10$ we know

It is Possible that $\bar{X}_n = 103.4 + 0.3 = 103.7$

While the truth $\mu = 103.4 < 103.5$, But we are concluding $\bar{X}_n = 103.7$, $\mu > 103.5$, that's because we have not been very cautious.

* So, what we want to do is to have a little buffer to account for the fact that $\bar{X}_n$ is not a precise value for the true μ.

The heuristic:

if we want to conclude that $\mu < 103.5$, maybe we need to have $\bar{X}_n < 103.5 - \underbrace{\text{buffer}}_{\downarrow\, n \to \infty \atop 0}$

So to make this intuition more precise, we need to understand those fluctuations, we need to actually put in something that is more precise than these little wiggles here. we need to have the CLT come in.

Clinical trials (1)

- Pharmaceutical companies use hypothesis testing to test if a new drug is efficient. (that's what they do, they want to know the new drug work?)
- To do so, they administer a drug to a group of patients (test group) and a placebo to another group (control group).
- Assume that the drug is a cough syrup.
- Let μ_{control} denote the expected number of expectorations per hour after a patient has used the placebo.
- Let μ_{drug} denote the expected number of expectorations per hour after a patient has used the syrup.
- We want to know if $\mu_{\text{drug}} < \mu_{\text{control}}$
- We compare two expected values. No reference number.

* Example of comparing 2 group's

=) There is placebo effect. its not like actually giving a drug that does not work, is going to have no effect on patients. It will have a small effect, but its very hard to quantify. we know its there but hard to quantify.

=) Rather than saying, the Ground truth is no improvement, the Ground truth is the placebo effect, we need to measure what the placebo effect is.

we split our patient's into two group's

① drug Group ② control group

cough syrup :- we measure the efficacy of the cough syrup is to measure How many time we cough in a hour

$\mu_{control}$ = # expectoration per hour (Expected Number)

=) Control Group :- those are the people who have been actually given just like sugar (maple syrup) ($\mu_{control}$)

Drug Group :- People who are given actual cough syrup. (μ_{Drug})

* we can imagine may be maple syrup have an effect on Expectoration per hour.
(we don't know what this gonna be, we Just want to measure if the drug is actually having better impact on Expectoration per hour than mapal syrup)

What we want to know "if $\mu_{drug} < \mu_{control}$" that would be enough.

* if we had access to all the population's, that will ever take the syrup for all ages, we will get μ_{drug}
* Problem is we don't have access to it
μ_{drug} is unknown.
* we are gonna have to make this desicion based on Sample's that give me imprecise knowledge about μ_{Drug} and $\mu_{control}$.

* in 1'st example we compared an unknown expected value to a fixed number 103.5,

* here we are comparing two unknown number's with each other. there is two Source of randomness. μ_{Drug}, $\mu_{control}$

Clinical trials (2)

- Let $\overbrace{X_1, \dots, X_{n_{\text{drug}}}}^{\text{Collect data from drug group}}$ denote n_{drug} i.i.d r.v. with distribution $\text{Poiss}(\mu_{\text{drug}})$
- Let $Y_1, \dots, Y_{n_{\text{control}}}$ denote n_{control} i.i.d r.v. with distribution $\text{Poiss}(\mu_{\text{control}})$ (Collect data from control group)
- We want to test if $\mu_{\text{drug}} < \mu_{\text{control}}$.

Heuristic:

fluctuation of $\bar{X}_{\text{drug}}$ & $\bar{X}_{\text{control}}$

"If $\bar{X}_{\text{drug}} < \bar{X}_{\text{control}} -$(something that $\xrightarrow[\substack{n_{\text{drug}} \to \infty \\ n_{\text{control}} \to \infty}]{}$ 0), then conclude that $\mu_{\text{drug}} < \mu_{\text{control}}$"

* we don't have to collect same # of people for drug & control group.

we can write $\bar{X}_{drug} < \bar{X}_{control} - (\text{goes to } 0)$

$\iff \bar{X}_{drug} - \bar{X}_{control} < -(\text{goes to } 0)$

$\bar{X}_{drug}$ ↓ Gaussian from CLT; $\bar{X}_{control}$ ↓ Gaussian from CLT

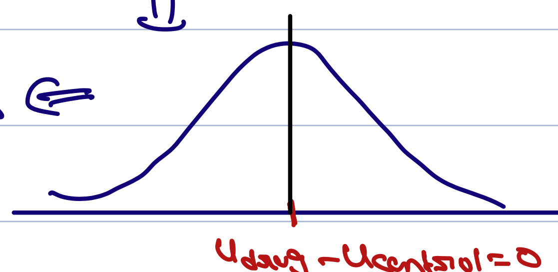
$\underbrace{\text{Gaussian from CLT} \quad \text{Gaussian from CLT}}_{\text{independent}}$

$\Rightarrow \bar{X}_{drug} - \bar{X}_{control} \sim N(\mu_{drug} - \mu_{control}, \text{some variance})$

let's assume $\mu_{drug} = \mu_{control}$ (huge scam :))

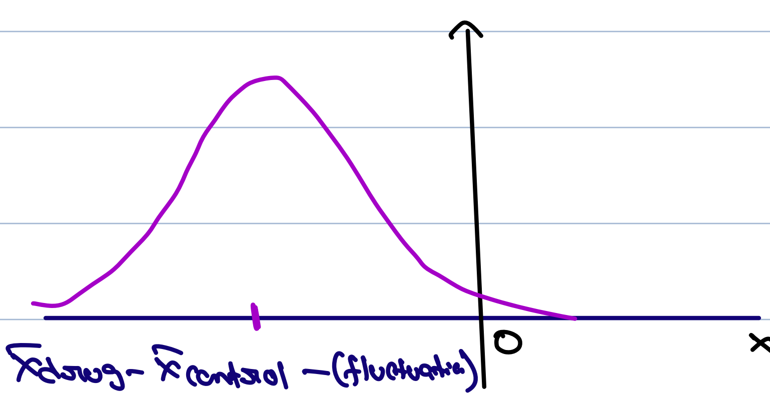
$\Rightarrow \bar{X}_{drug} - \bar{X}_{control} \sim N(0, Var)$

⇓



half of the time we would ⇐ actuall come up with the distribution $\bar{X}_{drug} - \bar{X}_{control}$ above this value (0), half of the time below this value (0) (frequentist confidence interval)

$\Rightarrow$ Half of the scam's go through FDA so what we are trying to do is



Heuristics (1)

Example 1: A coin is tossed 80 times, and Heads are obtained 54 times. Can we conclude that the coin is significantly unfair ?

- $n = 80$, $X_1, \ldots, X_n \stackrel{iid}{\sim} \mathsf{Ber}(p)$;
- $\bar{X}_n = 54/80 = .68$
- If it was true that $p = .5$: By CLT+Slutsky's theorem,

$$\sqrt{n}\frac{\bar{X}_n - .5}{\sqrt{.5(1-.5)}} \approx \mathcal{N}(0,1).$$

- $\sqrt{n}\frac{\bar{X}_n - .5}{\sqrt{.5(1-.5)}} \approx 3.22$
- Conclusion: It **seems quite** reasonable to reject the hypothesis $p = .5$.

we want know whether we have fair coin.

$\Rightarrow$ $P = 1/2$ (want to know) or $P \neq 1/2$

$$X_1, X_2, \ldots, X_{80} \overset{iid}{\sim} \text{Bernoulli}(P)$$

$\Rightarrow$ if we run this experiment 80 times, we need exactly 40 head's, 40 tail's, this is very unlikely to happen.

$\overline{X_n} = \frac{54}{80} = 0.68$ (question is can this be due to chance?) Chance alone?

$\Rightarrow$ what is the Likelihood that a fair coin would actually end up 54 head's out of 80, rather than 40?

$$\sqrt{n}\,\frac{\overline{X_n} - P}{\sqrt{P(1-P)}} \sim N(0,1) \quad \text{(CLT)}$$

$$\Rightarrow \quad \sqrt{n}\,\frac{\overline{X_n} - P}{\sqrt{\overline{X_n}(1-\overline{X_n})}} \sim N(0,1) \quad \text{(CLT + Slutsky)}$$

$$\Rightarrow \quad \frac{\sqrt{80}\,(0.68 - p)}{\sqrt{0.68\,(1-0.68)}} \sim N(0,1)$$

if fair $p = 0.5$

$\Rightarrow$ we want to know what is the Likelihood if a fair coin would give me 0.68? $\Rightarrow$ Plug in $p = 0.5$

$$\frac{\sqrt{n}\,(\bar{X}_n - 0.5)}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \sim N(0,1)$$

if coin is fair

* otherwise we are gonna have mean which is not zero here.

if coin is not fair (Assume $p = 0.6$)

$$\frac{\sqrt{n}\,(\bar{X}_n - 0.5)}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} = \frac{\sqrt{n}(\bar{X}_n - 0.6)}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} + \frac{\sqrt{n}(0.6-0.5)}{\sqrt{\bar{X}_n(1-\bar{X}_n)}}$$

$\sim N(0,1)$ + some value (Number)

* so if we are actually, want in the fact that

$$\frac{\sqrt{n}(\bar{x}_n - p)}{\sqrt{\bar{x}_n(1-\bar{x}_n)}} \sim N(0,1)$$

has normal distributio, we need to Plug in true value of P in above formula.

* if we know $P = 0.5$ then $\frac{\sqrt{n}(\bar{x}_n - 0.5)}{\sqrt{0.5(1-0.5)}} \sim N(0,1)$

$$\frac{\sqrt{80}(0.68 - 0.5)}{\sqrt{0.5(1-0.5)}} = 3.22$$

$\downarrow$

3; S.D away form 0

$\Rightarrow$ reasonable to reject the hypothesis that $P = 0.5$

* The number we got (3.22), we should see, should look like a number we would get if we ask computer to draw one random Gaussian number.

* ⟹ 99.9% this number would be $(-3, 3)$ (± 3 S.D)

⟹ 78% $(-2, 2)$ (± 2 S.D)

⟹ 68% $(-1, 1)$ (± 1 S.D)

* getting 3.22 is extremely unlikely to happen.

* we are extremely unlucky to happen. But it can happen right?, it could be the case out of 80, 80 head's.

⟹ So we cannot remove completely the uncertanity. right? it still possible 3.22 is due to noise,

⟹ 3.22 is very unlikely of a Gaussian.

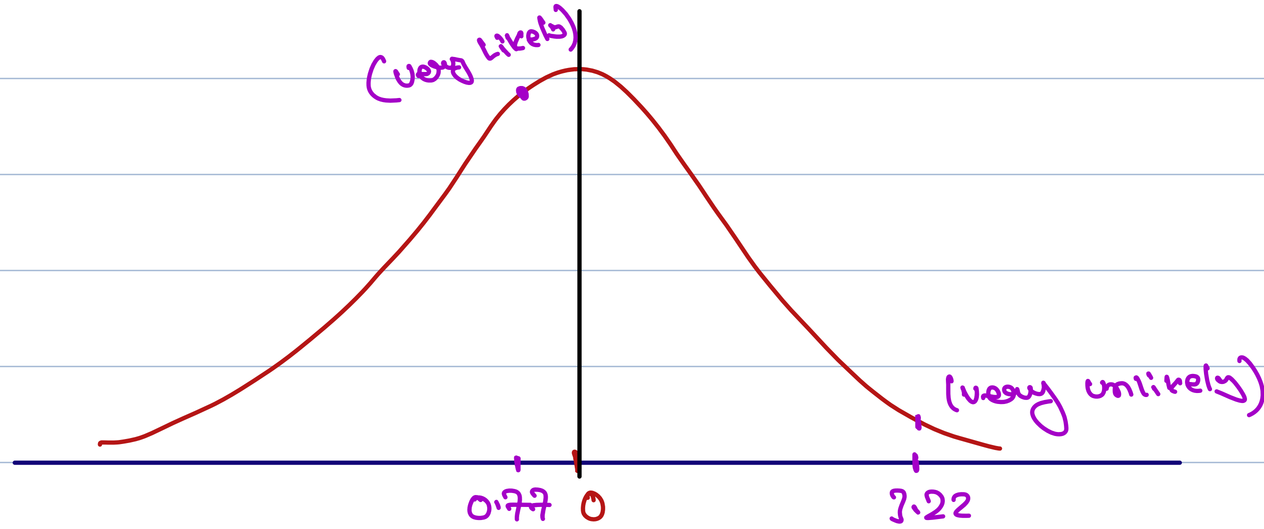
Heuristics (2)

Example 2: A coin is tossed 30 times, and Heads are obtained 13 times. Can we conclude that the coin is significantly unfair ?

- $n = 30, X_1, \ldots, X_n \overset{iid}{\sim} \mathsf{Ber}(p)$;
- $\bar{X}_n = 13/30 \approx .43$
- If it was true that $p = .5$: By CLT+Slutsky's theorem,

$$\sqrt{n}\frac{\bar{X}_n - .5}{\sqrt{.5(1-.5)}} \approx \mathcal{N}(0,1).$$

- Our data gives $\sqrt{n}\dfrac{\bar{X}_n - .5}{\sqrt{.5(1-.5)}} \approx -.77$
- The number $.77$ is a plausible realization of a random variable $Z \sim \mathcal{N}(0,1)$.
- Conclusion: our data does not suggest that the coin is unfair.



we are expecting is out of 30, we got 13/30 we are happy to put that on the account of randomness. Now of course, the question is gonna be, where do we draw the line? is 12 the right number? 11? 10? what is it?

* So Basically the answer is its whatever you want to be. The Problem is its hard to think on the scale. This scale is moving, everytime we have a new Problem, we have a new scale in mind. It's very difficult.

The Purpose of statistical analysis and in Particular this Process that, takes $\bar{x}$ and turns it into something that should be standard Gaussian

* Allow's us to map the value of $\bar{x}$ into a scale that is the standard scale of the Gaussian.

* What is large number? or an unsually large number for a Gaussian? that's all we need to know.

=> we need to draw the line on Gaussian.
on How large should a +ve Value of a Gaussian to become unreasonable to us?

* Probability is something that quantifies our error. and the question is What Percentage of error are we willing to tolerate?

Statistical formulation (1)

- Consider a sample $X_1, \ldots, X_n$ of i.i.d. random variables and a statistical model $(E, (\mathbb{P}_\theta)_{\theta \in \Theta})$.

- Let Θ_0 and Θ_1 be disjoint subsets of Θ.

- Consider the two hypotheses: $\begin{cases} H_0 : & \theta \in \Theta_0 \\ H_1 : & \theta \in \Theta_1 \end{cases}$

- H_0 is the *null hypothesis*, H_1 is the *alternative hypothesis*.

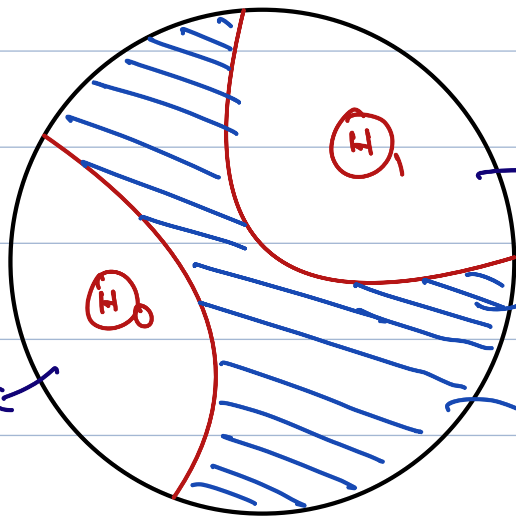
- If we believe that the true θ is either in Θ_0 or in Θ_1, we may want to *test* H_0 *against* H_1.

- We want to decide whether to *reject* H_0 (look for evidence against H_0 in the data).

$x_1, x_2, \ldots x_n \qquad (E, (\mathbb{P}_\theta)_{\theta \in \Theta})$

Space in which our parameter live

Θ_0, Θ_1 (two disjoint subsets)



$\mu < 103.5$

if it happen's it happen's, we are not gonna make any statement about it

removed already

Ex: $\mu > 103.5$

$\mu = 103.5$

in test we are focusing on Θ_0, or Θ_1, It mean's that a Priori we have already removed all the Possibilities in this region. already removed space ($\mu > 103.5$). if it happen's, it happen's, we are not making any statement about that case.

$\Theta_0 \Rightarrow$ Null hypothesis

$\Theta_1 \Rightarrow$ Alternative hypothesis

$\Theta_0 \Rightarrow$ Status Quote $\Theta_1 \Rightarrow$ Scientific discovery.

* if we are a regulator, we push toward's Θ_0, if we are a scientist, we push toward's Θ_1.

* if we can find evidence from data, we are tending toward's Θ_1, But if we say we need a lot of data before we can actually move away from status quote that's Θ_0.

Θ_0 is some status quote, Θ_1 being some discovery that goes against the status quote.

* if we Believe that the true θ is either in Θ_0 or in Θ_1, we may want to test H_0 against Θ_1. This is how our question's are formulated.

* we are gonna say, we have Null, we have an alternative, we want to test H_0 against H_1

* we want to test the null hypothesis against alternative hypothesis.

$H_0 : \theta \in \Theta_0 \quad (\mu = 103.5)$

$H_1 : \theta \in \Theta_1 \quad (\mu < 103.5)$

Statistical formulation (2)

- H_0 and H_1 do not play a symmetric role: the data is is only used to try to disprove H_0
- In particular lack of evidence, does not mean that H_0 is true ("innocent until proven guilty")
- A *test* is a statistic $\psi \in \{0, 1\}$ such that:
  - If $\psi = 0$, H_0 is not rejected;
  - If $\psi = 1$, H_0 is rejected.
- Coin example: H_0: $p = 1/2$ vs. H_1: $p \neq 1/2$.
- $\psi = \mathbb{1}\left\{ \left| \sqrt{n} \frac{\bar{X}_n - .5}{\sqrt{.5(1 - .5)}} \right| > C \right\}$, for some $C > 0$.
- How to choose the *threshold* C ?

* we collect data, what we are going to try to find on this data is evidence against Ho. and the alternative is going to guide you into which direction we should be lookins for evidence against this guy.

* We want to decide whether we reject Ho. so we look for evidence agaist Ho in the data.

* Ho and H_1 do not play a symmetric role:
The data is only used to try to disaprove Ho
⟹ its very important to know which one youre going to plac in Ho, and which one youre going to place in H_1. Becaus we are gonna always favour in side of Ho. so we have to be careful about this. If it's a close call b/w Ho and H_1, if data does not carry lot of evidence we are gonna side with Ho.

* we actually never saying that Ho is true, we are Just saying we did not

find evidence against H_0.

* we don't say I accepted H_0, we say I failed to reject H_0.

"Innocent until Proven Guilty" => in Particular lack of evidence does not mean you are not Guilty, OJ Simpson was found not Guilty, it was not found innocent.

* The Prosecutor bring's their evidence, and then the Jury has to decide whether they were convinced that this person was Guilty of anything (crime). and the Question is do you have enough, but if you don't have evidence, it's not Burden of defender to Prove that they are innocent. No body is Proving they are innocent, we Just have to make sure there is enough evidence against you.

Youre H_0 until Proven H_1

* The role of Estimator's in hypothesis testing is played by something called Test.

* And a test is statistic. we are asking only yes and No Answer from this test

⇒ A test is a statistic $\varphi \in \{0, 1\}$

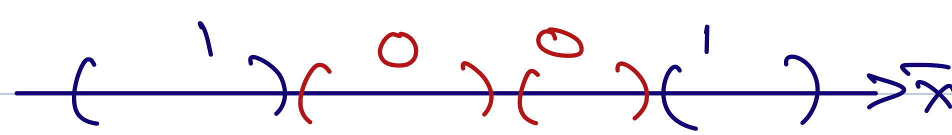
① if $\varphi = 0$, H_0 is not rejected

② if $\varphi = 1$, H_0 is rejected

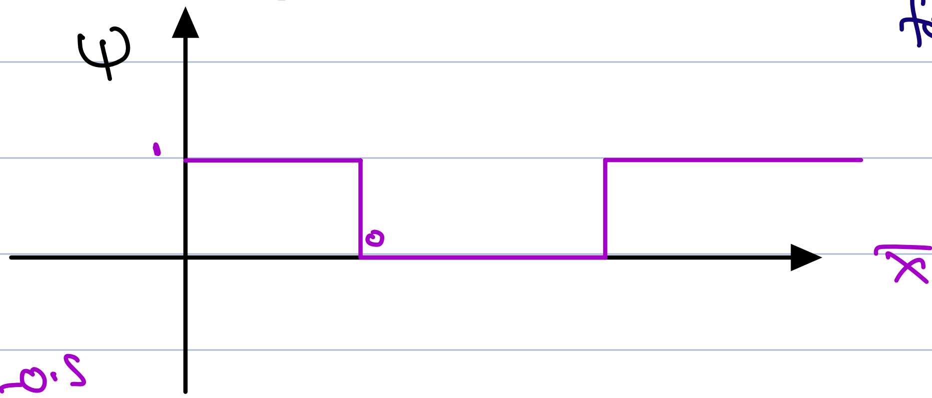
⇒ So we have two possibilites (yes, No) , we look at our data & turned it into yes, No answer

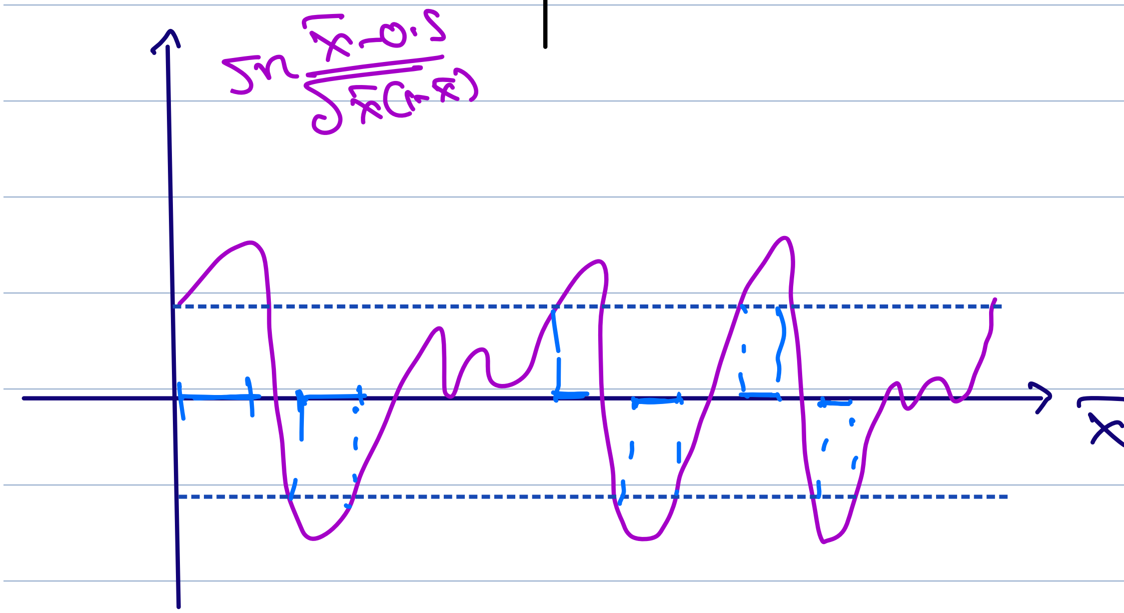
⇒ Any thing that Gives only 0 or 1 is an indicator function.

Ex: coin : $H_0 : P = 1/2$ vs $H_1 : P \neq 1/2$



deviding the sample space of statistic into 0 or 1 if $\psi = I\left\{ \left| \sqrt{n}\, \frac{\bar{x}-0.5}{\sqrt{p(1-p)}} \right| > c \right\}$ for some $c>0$





* The test is not just specification of $\bar{x}$, But also specification of constant C.

* And the constant c is was exactly where our belief about large value for Gaussian is.

Statistical formulation (3)

- *Rejection region* of a test ψ:

$$R_\psi = \{x \in E^n : \psi(x) = 1\}.$$

- *Type 1 error* of a test ψ (rejecting H_0 when it is actually true):

$$\begin{array}{rcccl} \alpha_\psi & : & \Theta_0 & \to & \mathbb{R} \\ & & \theta & \mapsto & \mathbb{P}_\theta[\psi = 1]. \end{array}$$

- *Type 2 error* of a test ψ (not rejecting H_0 although H_1 is actually true):

$$\begin{array}{rcccl} \beta_\psi & : & \Theta_1 & \to & \mathbb{R} \\ & & \theta & \mapsto & \mathbb{P}_\theta[\psi = 0]. \end{array}$$

- *Power* of a test ψ:

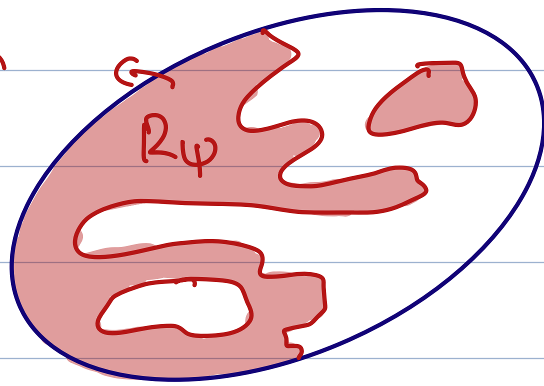
$$\pi_\psi = \inf_{\theta \in \Theta_1} (1 - \beta_\psi(\theta)).$$

Rejection region of a test φ

$$R_\varphi = \{ x \in E^n : \varphi(x) = 1 \}$$

$\Rightarrow$ The value of the sample $(x_1, x_2, \ldots x_n)$ that we will collect will lead to a test that says 1. (rejection region)

rejection region.



$E^n = \mathcal{X} = (x_1, x_2, \ldots x_n)$

The Test is the indicator of rejection region.

There are two way's we can make error when there's test.

① Either the truth is in H_0 and you are saying actually its H_1.

② Are the truth is in H_1 and you are saying H_0

Type 1 error:

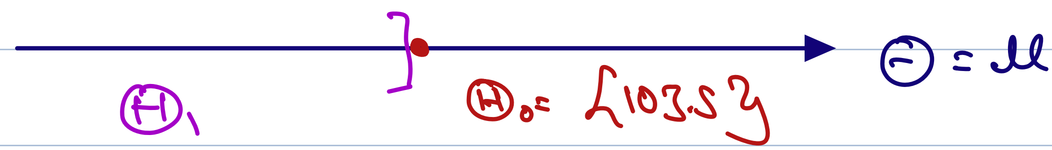
Probability of type 1 error:-

$$\alpha_\psi : \Theta_0 \longrightarrow \mathbb{R}$$

$$\theta \longrightarrow \mathbb{P}_\theta[\psi = 1]$$

* Probability of saying $\psi = 1$ when $\theta \in \Theta_0$

Ex: Cherry Blosome RUN



we need to compute the function α_ψ

$$\alpha_\psi : \theta \in \Theta_0 \longrightarrow \mathbb{P}_\theta(\psi = 1)$$

$$\Rightarrow \mathbb{P}_{103.5}(\psi = 1) \quad (\text{Becau } \Theta_0 = \{103.5\})$$

Probability that we reject H_0 when the True mean $\mu^* = 103.5$

Type 2 error:

$$\beta_\psi : \Theta_1 \longrightarrow \mathbb{R}$$

$$\theta \longrightarrow \mathbb{P}_\theta(\psi = 0)$$

β_φ function that map's $\theta \in \Theta_1$ to the Probability that we failed to reject H_0

$\mathbb{P}$(Saying the true mean < 103.5 when its actualy 103.5)

$$\beta_\varphi(\theta) = \mathbb{P}_{\theta \in \Theta_1}(\varphi = 0)$$ (Probability of type 2 error)

$$= 1 - \underbrace{\mathbb{P}_{\theta \in \Theta_1}(\varphi = 1)}_{\text{Probabiliti that we reject } H_0 \text{ and we should have reject.}}$$

$$\mathbb{P}_{\theta \in \Theta_1}(\varphi = 1) = 1 - \beta_\varphi(\theta)$$ (good thing)

this is good thing, this number being larg is good thing. it mean's, we should have rejected H_0, and we rejected. and we want this to hapen with large Probability.

\# and we are gonna look at most conservative choice of this number.

$$\mathbb{P}_{\theta \in \Theta_1}(\varphi = 1) = \underbrace{1 - \beta_\varphi(\theta)}$$

smallest value it can take on Θ_1 is called Power of φ

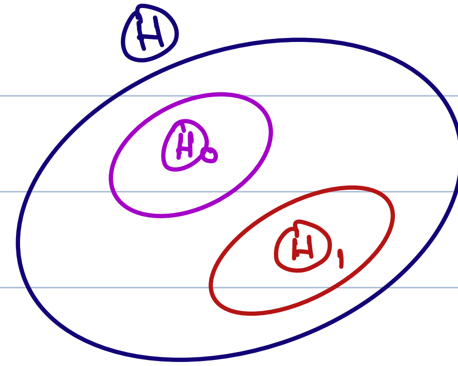
$\Rightarrow$ smallest possible value it can take.

Desision

	Reject Ho	Fail to reject Ho
Ho is true	Type 1 error	Correct decision
H1 is true	Correct decision	Type 2 error

Reality

① $\Theta \to \Theta_0 \subset \Theta$, $\Theta_1 \subset \Theta$ disjoint



② $H_0 : \theta \in \Theta_0$ Null hypothesis

$H_1 : \theta \in \Theta_1$ Alternative hypothesis

They have an asymmetric role. The Null represents the status quote, and data is here is to bring evidence against this guy (H_0). And we can really never conclude that H_0 is true, because all we could conclude is that H_1 is not true.

③ Test φ a statistic

$$\varphi(x_1, x_2, \dots x_n) \in \{0, 1\}$$

Rejection region $R_\varphi = \{x_1 \dots x_n : \varphi(x_1, x_2 \dots x_n) = 1\}$

④ Type-I error: reject H_0, whereas H_0 is actually correct

Type-II error: fail's to reject H_0, where H_1 is correct

$$P(\text{Type I error}) = \alpha_\varphi = \mathbb{P}_{\theta \in \Theta_0}(\varphi = 1)$$

$$P(\text{Type II error}) = \beta_\varphi = \mathbb{P}_{\theta \in \Theta_1}(\varphi = 0)$$

Power of a test φ :-

$$\pi_\varphi = \inf_{\theta \in \Theta_1} (1 - \beta_\varphi(\theta))$$

Statistical formulation (4)

- A test ψ has *level* α if

$$\alpha_\psi(\theta) \leq \alpha, \qquad \forall \theta \in \Theta_0.$$

- A test ψ has *asymptotic level* α if

$$\lim_{n \to \infty} \alpha_\psi(\theta) \leq \alpha, \qquad \forall \theta \in \Theta_0.$$

- In general, a test has the form

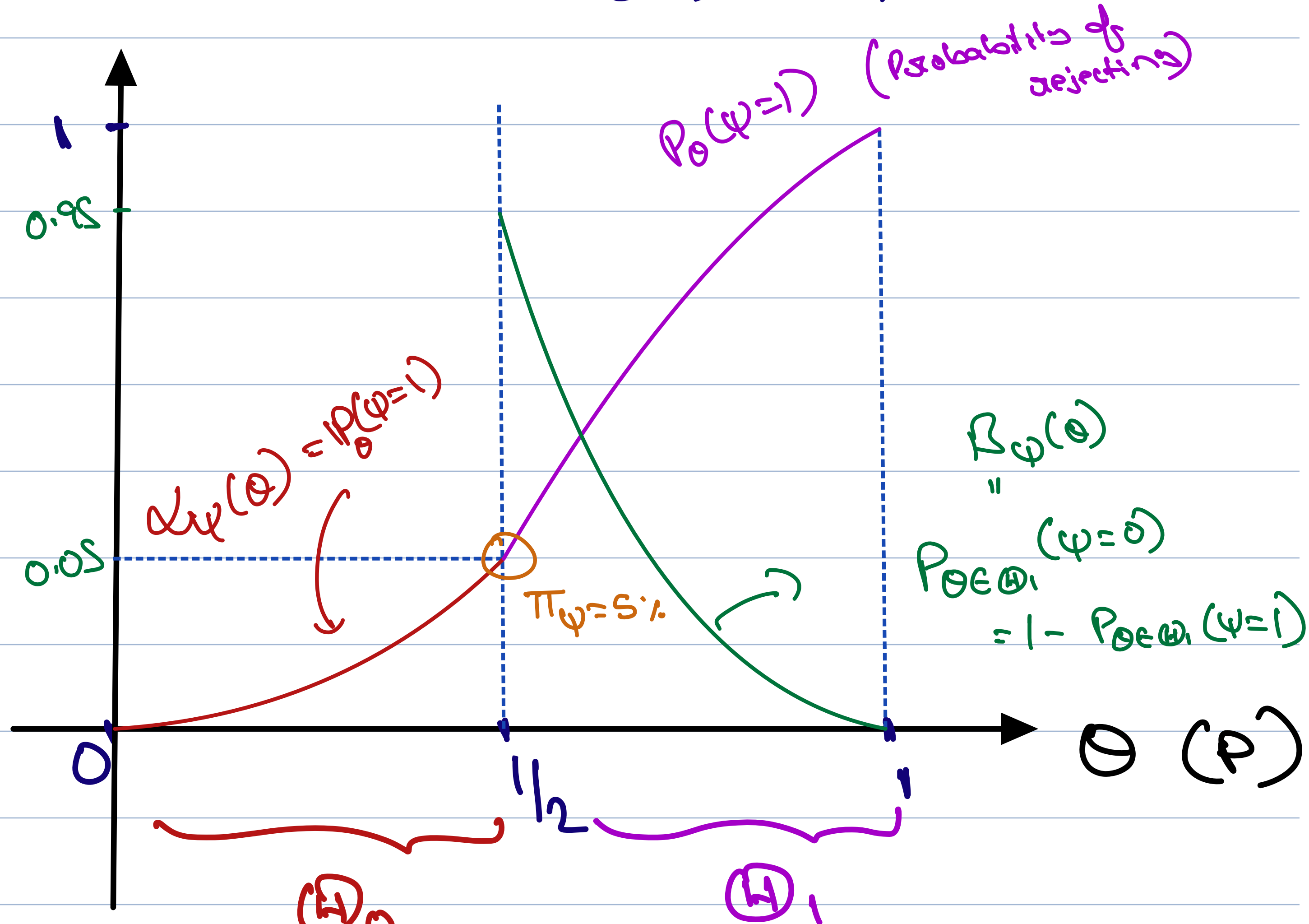
$$\psi = \mathbb{1}\{T_n > c\},$$

for some statistic T_n and threshold $c \in \mathbb{R}$.

- T_n is called the *test statistic*. The rejection region is $R_\psi = \{T_n > c\}$.

$X_1, X_2, \dots, X_n \sim Ber(p)$

$$\psi = \mathbb{1}\left(\left| \sqrt{n} \frac{\bar{X}_n - p}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \right| > c \right) \quad \text{for some } c > 0$$



Power $\pi_\psi = \sup_{\theta\in\Theta_0} \mathbb{P}_{\theta\in\Theta_0}(\psi = 1)$

sup rejecting Ho, when $\theta \in \Theta_0$

$$= \inf_{\theta\in\Theta_1}\left(1 - \mathbb{P}_{\theta\in\Theta_1}(\psi=0)\right)$$

inf (rejecting Ho, when $\theta\in\Theta_1$)

Power = Smallest Prob of rejecting H_0, when $\theta \in \Theta_1$

= Largest Probability when we make a Type-I error

* This largest Prob of rejecting H_0 when $\theta = \Theta_1$ is called Level of test φ

$$\alpha := \sup_{\theta \in \Theta_0} \alpha_\varphi(\theta)$$

in general they will give α value (Ex: $\alpha = 5\%$)

then $\varphi = \mathbb{1}\left(\left| \sqrt{n} \frac{\bar{X} - p}{\sqrt{\bar{X}(1-\bar{X})}} \right| > C_\alpha \right)$

we calculat C_α from α value

* in many instance we do not know what's the Prob, we do not know How to compute the Prob of type 1 error. This is the maximum value for Type 1 error. we don't know how to compute it.

* It might be a very complicated r.v, may be its some weird Binomial, we could compute it, But it would be Painful. But we know how to compute is its asymptotic value. Because of the CLT, convergence in distribution, tells us that

$\mathbb{P}(\text{Type 1 error}) = \mathbb{P}(\text{Gaussian at some region})$

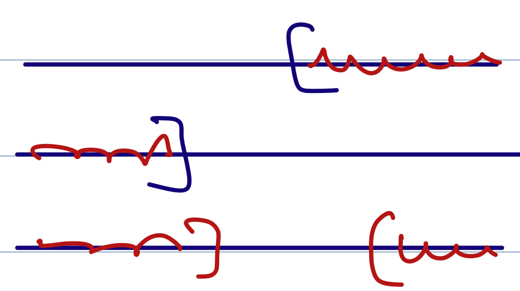
$\Rightarrow$ we will compute Asymptotic level α i.e.

$$\max \lim_{n\to\infty} \alpha_\psi(\theta) \le \alpha \quad \forall \theta \in \Theta_0$$

$$R_\psi : \{T_n(X_1, X_2, .. X_n) > c\}$$

This includes $\underbrace{|Q_n|}_{T_n} > c \iff Q_n > c$ or $Q_n < -c$

Example's of Rejection



T_n is called Test statistic :

What is test statistic?

→ what we are gonna build from our data $(x_1, x_2, .. x_n)$ then compare to some fixed value.

Ex: $$\psi = \mathbb{1}\left(\underbrace{\sqrt{n}\,\frac{\bar{X}_n - p}{\sqrt{\bar{X}_n(1-\bar{X}_n)}}}_{\text{test statistic}} > C_\alpha \right)$$

Is this a statistic?

It should not depend on parameter.

+ we will replace p by θ in Θ_0

Example (1)

- Let $X_1, \dots, X_n \stackrel{iid}{\sim} \mathsf{Ber}(p)$, for some unknown $p \in (0,1)$.
- We want to test:

$$H_0\colon\ p = 1/2 \ \text{vs.}\ H_1\colon\ p \neq 1/2$$

  with asymptotic level $\alpha \in (0,1)$.

- Let $T_n = \sqrt{n}\ \dfrac{\hat{p}_n - 0.5}{\sqrt{.5(1-.5)}}$, where $\hat{p}_n$ is the MLE.

- If H_0 is true, then by CLT and Slutsky's theorem,

$$\mathbb{P}[T_n > q_{\alpha/2}] \xrightarrow[n\to\infty]{} 0.05$$

- Let $\psi_\alpha = \mathbb{1}\{T_n > q_{\alpha/2}\}$.

$X_1, X_2, \ldots X_n \sim Ber(p)$

$H_0: p = 1/2$

$H_1: p \neq 1/2$

from CLT + slutsky

Wald's under H_0

$$\sqrt{n}\,\frac{\bar{X}_n - 0.5}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \xrightarrow{(d)} N(0,1)$$

our test statistic gonna be

$$T_n = \left|\sqrt{n}\,\frac{\bar{X}_n - 0.5}{\sqrt{\bar{X}_n(1-\bar{X}_n)}}\right|$$

$$\underbrace{\lim_{n\to\infty} \mathbb{P}_p(T_n > c)} \leq \alpha$$

level of test (given)

$$\mathbb{P}_p\left(\left|\frac{\sqrt{n}(\bar{X}_n - 0.5)}{\sqrt{\bar{X}_n(1-\bar{X}_n)}}\right| > c\right) \leq \alpha$$

is this correct?

even in the limit? NO

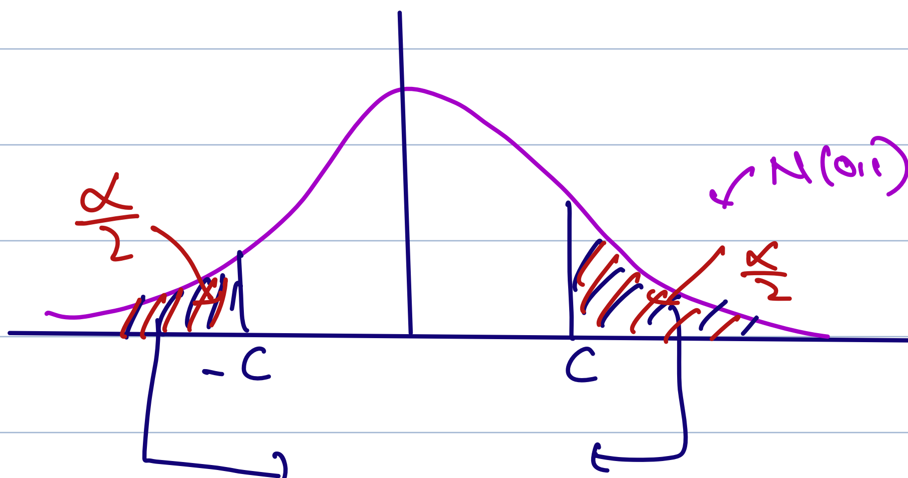
$\mathbb{P}(|N(0,1)| > c)$

$$\mathbb{P}_p\left(\left| \frac{\sqrt{n}\,(\overline{X}_n - 0.5)}{\sqrt{\overline{X}_n(1-\overline{X}_n)}} \right| > c \right)$$

$$\approx \mathbb{P}\left(|N(0,1)| > c \right)$$

$\forall p \in \Theta_0 \iff p = 0.5$ if $p \in \Theta_0$ $(\iff p = 0.5)$

then solve for c.



we want sum of the's area to be α

$$\mathbb{P}(N(0,1) > c) = \frac{\alpha}{2} \iff c = q_{\frac{\alpha}{2}}$$

for $\alpha = 5\%$ $q_{\frac{\alpha}{2}} = 1.96$

__in Ex1__ $|3.45| = 3.45 \Rightarrow 3.45 > 1.96$

reject H_0

__in Ex2__ $|-0.77| = 0.77 \Rightarrow 0.77 < 1.96$

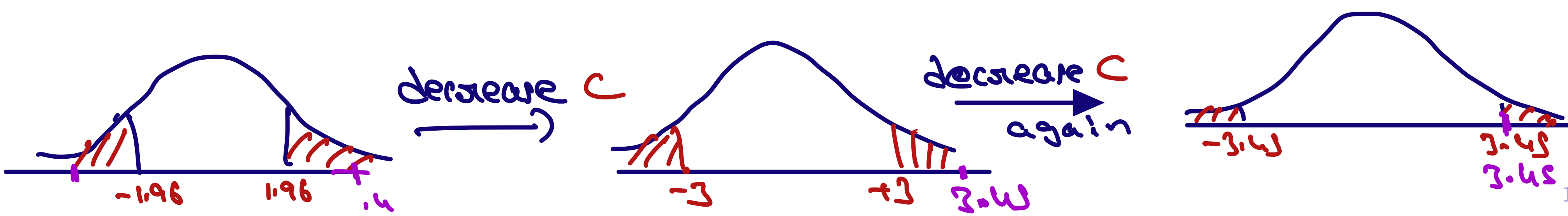
failed to reject H_0

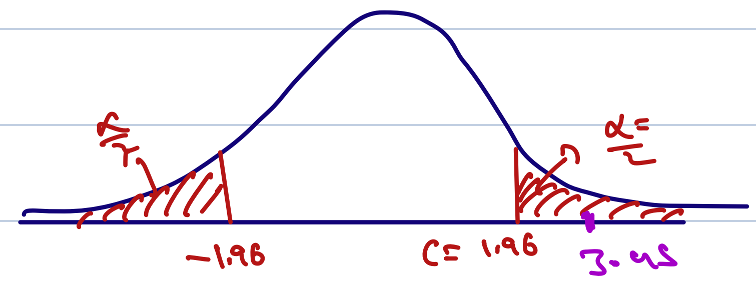
Example (2)

Coming back to the two previous coin examples: For $\alpha = 5\%$, $q_{\alpha/2} = 1.96$, so:

- In **Example 1**, H_0 is rejected at the asymptotic level 5% by the test $\psi_{5\%}$; $|3.45| > 1.96 \Rightarrow$ reject H_0
- In **Example 2**, H_0 is not rejected at the asymptotic level 5% by the test $\psi_{5\%}$. $|-0.77| < 1.96 \Rightarrow$ failed to reject H_0

Question: In **Example 1**, for what level α would ψ_α not reject H_0 ? And in **Example 2**, at which level α would ψ_α reject H_0 ?





in Example what should be the value of c so that for (3.45) we are rejecting Ho? for C= 3.45

* so for what α' value for $C = 3.45$ $\quad q_{\frac{\alpha'}{2}} = 3.45$

* that mean's if we actuall fix this α'-Level, that would be exactly the tipping point at which we would go from accepting to rejecting.

⟹ we know $C = 3.45$ is the tipping point, But if I try to map this back and understand what level would have been giving me this particular tiping point $(C = 3.45) = \alpha'$ (that's a number b/w 0 & 1.

⟹ the smaller α' is, the larger the c value. the more evidence we have in our data against ho.

p-value

Definition

The (asymptotic) *p-value* of a test ψ_α is the smallest (asymptotic) level α at which ψ_α rejects H_0. It is random, it depends on the sample.

Golden rule

p-value $\leq \alpha \quad \Leftrightarrow \quad H_0$ is rejected by ψ_α, at the (asymptotic) level α.

The smaller the p-value, the more confidently one can reject H_0.

- Example 1: p-value $= \mathbb{P}[|Z| > 3.21] \ll .01$.
- Example 2: p-value $= \mathbb{P}[|Z| > .77] \approx .44$.

Neyman-Pearson's paradigm

Idea: For given hypotheses, among all tests of level/asymptotic level α, is it possible to find one that has maximal power ?

Example: The trivial test $\psi = 0$ that never rejects H_0 has a perfect level ($\alpha = 0$) but poor power ($\pi_\psi = 0$).

Neyman-Pearson's theory provides (the most) powerful tests with given level. In 18.650, we only study several cases.

The χ^2 distributions

Definition

For a positive integer d, the χ^2 *(pronounced "Kai-squared") distribution with d degrees of freedom* is the law of the random variable $Z_1^2 + Z_2^2 + \ldots + Z_d^2$, where $Z_1, \ldots, Z_d \overset{iid}{\sim} \mathcal{N}(0,1)$.

Examples:

- If $Z \sim \mathcal{N}_d(\mathbf{0}, I_d)$, then $\|Z\|_2^2 \sim \chi_d^2$.
- Recall that the sample variance is given by
$$S_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\bar{X}_n)^2$$
- Cochran's theorem implies that for $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, if S_n is the sample variance, then
$$\frac{nS_n}{\sigma^2} \sim \chi_{n-1}^2.$$
- $\chi_2^2 = \mathsf{Exp}(1/2)$.

Student's T distributions

Definition

For a positive integer d, the *Student's T distribution with* d *degrees of freedom* (denoted by t_d) is the law of the random variable $\dfrac{Z}{\sqrt{V/d}}$, where $Z \sim \mathcal{N}(0,1)$, $V \sim \chi^2_d$ and $Z \perp\!\!\!\perp V$ (Z is independent of V).

Example:

- Cochran's theorem implies that for $X_1, \dots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, if S_n is the sample variance, then

$$\sqrt{n-1}\ \frac{\bar{X}_n - \mu}{\sqrt{S_n}} \sim t_{n-1}.$$

Wald's test (1)

- Consider an i.i.d. sample $X_1, \dots, X_n$ with statistical model $(E, (\mathbb{P}_\theta)_{\theta \in \Theta})$, where $\Theta \subseteq \mathbb{R}^d$ $(d \geq 1)$ and let $\theta_0 \in \Theta$ be fixed and given.
- Consider the following hypotheses:

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

- Let $\hat{\theta}^{MLE}$ be the MLE. Assume the MLE technical conditions are satisfied.

- If H_0 is true, then

$$\sqrt{n}\ I(\hat{\theta}^{MLE})^{1/2} \left(\hat{\theta}_n^{MLE} - \theta_0\right) \xrightarrow[n\to\infty]{(d)} \mathcal{N}_d\left(0, I_d\right) \quad \text{w.r.t. } \mathbb{P}_{\theta_0}.$$

Wald's test (2)

- Hence,

$$\underbrace{n\left(\hat{\theta}_n^{MLE}-\theta_0\right)^\top I(\hat{\theta}^{MLE})\left(\hat{\theta}_n^{MLE}-\theta_0\right)}_{T_n} \xrightarrow[n\to\infty]{(d)} \chi^2_d \quad \text{w.r.t. } \mathbb{P}_{\theta_0}.$$

- Wald's test with asymptotic level $\alpha \in (0,1)$:

$$\psi = \mathbb{1}\{T_n > q_\alpha\},$$

  where q_α is the $(1-\alpha)$-quantile of χ^2_d (see tables).

- Remark: Wald's test is also valid if H_1 has the form "$\theta > \theta_0$" or "$\theta < \theta_0$" or "$\theta = \theta_1$"...

Likelihood ratio test (1)

- Consider an i.i.d. sample $X_1, \ldots, X_n$ with statistical model $(E, (\mathbb{P}_\theta)_{\theta \in \Theta})$, where $\Theta \subseteq \mathbb{R}^d$ $(d \geq 1)$.

- Suppose the null hypothesis has the form

$$H_0 : (\theta_{r+1}, \ldots, \theta_d) = (\theta_{r+1}^{(0)}, \ldots, \theta_d^{(0)}),$$

  for some fixed and given numbers $\theta_{r+1}^{(0)}, \ldots, \theta_d^{(0)}$.

- Let

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmax}} \; \ell_n(\theta) \quad \text{(MLE)}$$

  and

$$\hat{\theta}_n^c = \underset{\theta \in \Theta_0}{\operatorname{argmax}} \; \ell_n(\theta) \quad \text{(“constrained MLE”)}$$

Likelihood ratio test (2)

- Test statistic:

$$T_n = 2\left(\ell_n(\hat{\theta}_n) - \ell_n(\hat{\theta}_n^c)\right).$$

- Theorem
  Assume H_0 is true and the MLE technical conditions are satisfied. Then,

$$T_n \xrightarrow[n\to\infty]{(d)} \chi^2_{d-r} \quad \text{w.r.t. } \mathbb{P}_\theta.$$

- Likelihood ratio test with asymptotic level $\alpha \in (0,1)$:

$$\psi = \mathbb{1}\{T_n > q_\alpha\},$$

  where q_α is the $(1-\alpha)$-quantile of χ^2_{d-r} (see tables).

Testing implicit hypotheses (1)

- Let $X_1, \ldots, X_n$ be i.i.d. random variables and let $\theta \in \mathbb{R}^d$ be a parameter associated with the distribution of X_1 (e.g. a moment, the parameter of a statistical model, etc...)

- Let $g : \mathbb{R}^d \to \mathbb{R}^k$ be continuously differentiable (with $k < d$).

- Consider the following hypotheses:

$$\begin{cases} H_0 : & g(\theta) = 0 \\ H_1 : & g(\theta) = 0. \end{cases}$$

- E.g. $g(\theta) = (\theta_1, \theta_2)$ $(k = 2)$, or $g(\theta) = \theta_1 - \theta_2$ $(k = 1)$, or...

Testing implicit hypotheses (2)

- Suppose an asymptotically normal estimator $\hat{\theta}_n$ is available:

$$\sqrt{n} \quad \hat{\theta}_n - \theta \quad \xrightarrow[n\to\infty]{(d)} \mathcal{N}_d(0, \Sigma(\theta)).$$

- Delta method:

$$\sqrt{n} \quad g(\hat{\theta}_n) - g(\theta) \quad \xrightarrow[n\to\infty]{(d)} \mathcal{N}_k\left(0, \Gamma(\theta)\right),$$

where $\Gamma(\theta) = \nabla g(\theta)^\top \Sigma(\theta) \nabla g(\theta) \in \mathbb{R}^{k\times k}$.

- Assume $\Sigma(\theta)$ is invertible and $\nabla g(\theta)$ has rank k. So, $\Gamma(\theta)$ is invertible and

$$\sqrt{n}\,\Gamma(\theta)^{-1/2} \quad g(\hat{\theta}_n) - g(\theta) \quad \xrightarrow[n\to\infty]{(d)} \mathcal{N}_k\left(0, I_k\right).$$

Testing implicit hypotheses (3)

- Then, by Slutsky's theorem, if $\Gamma(\theta)$ is continuous in θ,

$$\sqrt{n}\,\Gamma(\hat{\theta}_n)^{-1/2} \quad g(\hat{\theta}_n) - g(\theta) \quad \xrightarrow[n\to\infty]{(d)} \mathcal{N}_k\left(0, I_k\right).$$

- Hence, if H_0 is true, i.e., $g(\theta) = 0$,

$$\underbrace{n g(\hat{\theta}_n)^\top \Gamma^{-1}(\hat{\theta}_n) g(\hat{\theta}_n)}_{T_n} \xrightarrow[n\to\infty]{(d)} \chi^2_k.$$

- Test with asymptotic level α:

$$\psi = \mathbb{1}\{T_n > q_\alpha\},$$

where q_α is the $(1-\alpha)$-quantile of χ^2_k (see tables).

The multinomial case: χ^2 test (1)

Let $E = \{a_1, \ldots, a_K\}$ be a finite space and $(\mathbb{P}_{\mathbf{p}})_{\mathbf{p} \in \Delta_K}$ be the family of all probability distributions on E:

- $$\Delta_K = \left\{ \mathbf{p} = (p_1, \ldots, p_K) \in (0,1)^K : \sum_{j=1}^{K} p_j = 1 \right\}.$$
- For $\mathbf{p} \in \Delta_K$ and $X \sim \mathbb{P}_{\mathbf{p}}$,

$$\mathbb{P}_{\mathbf{p}}[X = a_j] = p_j, \qquad j = 1, \ldots, K.$$

The multinomial case: χ^2 test (2)

- Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathbb{P}_{\mathbf{p}}$, for some unknown $\mathbf{p} \in \Delta_K$, and let $\mathbf{p}^0 \in \Delta_K$ be fixed.

- We want to test:

$$H_0\colon\ \mathbf{p} = \mathbf{p}^0 \text{ vs. } H_1\colon\ \mathbf{p} = \mathbf{p}^0$$

  with asymptotic level $\alpha \in (0, 1)$.

- Example: If $\mathbf{p}^0 = (1/K, 1/K, \ldots, 1/K)$, we are testing whether $\mathbb{P}_{\mathbf{p}}$ is the uniform distribution on E.

The multinomial case: χ^2 test (3)

- Likelihood of the model:

$$L_n(X_1, \ldots, X_n, \mathbf{p}) = p_1^{N_1} p_2^{N_2} \ldots p_K^{N_K},$$

where $N_j = \#\{i = 1, \ldots, n : X_i = a_j\}$.

- Let $\hat{\mathbf{p}}$ be the MLE:

$$\hat{\mathbf{p}}_j = \frac{N_j}{n}, \quad j = 1, \ldots, K.$$

⚠ $\hat{\mathbf{p}}$ maximizes $\log L_n(X_1, \ldots, X_n, \mathbf{p})$ **under the constraint**

$$\sum_{j=1}^{K} p_j = 1.$$

The multinomial case: χ^2 test (4)

- If H_0 is true, then $\sqrt{n}(\hat{\mathbf{p}} - \mathbf{p}^0)$ is asymptotically normal, and the following holds.

Theorem

$$\underbrace{n \sum_{j=1}^{K} \frac{\left(\hat{\mathbf{p}}_j - \mathbf{p}_j^0\right)^2}{\mathbf{p}_j^0}}_{T_n} \xrightarrow[n \to \infty]{(d)} \chi^2_{K-1}.$$

- χ^2 test with asymptotic level α: $\psi_\alpha = \mathbb{1}\{T_n > q_\alpha\}$, where q_α is the $(1-\alpha)$-quantile of χ^2_{K-1}.
- Asymptotic p-value of this test: $p - \text{value} = \mathbb{P}\left[Z > T_n | T_n\right]$, where $Z \sim \chi^2_{K-1}$ and $Z \perp\!\!\!\perp T_n$.

The Gaussian case: Student's test (1)

- Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, for some unknown $\mu \in \mathbb{R}, \sigma^2 > 0$ and let $\mu_0 \in \mathbb{R}$ be fixed, given.

- We want to test:

$$H_0\colon\ \mu = \mu_0 \text{ vs. } H_1\colon\ \mu = \mu_0$$

  with asymptotic level $\alpha \in (0, 1)$.

- **If σ^2 is known:** Let $T_n = \sqrt{n}\ \dfrac{\bar{X}_n - \mu_0}{\sigma}$. Then, $T_n \sim \mathcal{N}(0, 1)$ and

$$\psi_\alpha = \mathbb{1}\{|T_n| > q_{\alpha/2}\}$$

  is a test with (non asymptotic) level α.

The Gaussian case: Student's test (2)

If σ^2 is unknown:

- Let $\widetilde{T_n} = \sqrt{n-1}\ \dfrac{\bar{X}_n - \mu_0}{\sqrt{S_n}}$, where S_n is the sample variance.
- Cochran's theorem:
  - $\bar{X}_n \perp\!\!\!\perp S_n$;
  - $\dfrac{nS_n}{\sigma^2} \sim \chi^2_{n-1}$.
- Hence, $\widetilde{T_n} \sim t_{n-1}$: Student's distribution with $n-1$ degrees of freedom.

The Gaussian case: Student's test (3)

- Student's test with (non asymptotic) level $\alpha \in (0,1)$:

$$\psi_\alpha = \mathbb{1}\{|\widetilde{T_n}| > q_{\alpha/2}\},$$

where $q_{\alpha/2}$ is the $(1-\alpha/2)$-quantile of t_{n-1}.

- If H_1 is $\mu > \mu_0$, Student's test with level $\alpha \in (0,1)$ is:

$$\psi'_\alpha = \mathbb{1}\{\widetilde{T_n} > q_\alpha\},$$

where q_α is the $(1-\alpha)$-quantile of t_{n-1}.

- Advantage of Student's test:
  - Non asymptotic
  - Can be run on small samples

- Drawback of Student's test: It relies on the assumption that the sample is Gaussian.

Two-sample test: large sample case (1)

- Consider two samples: $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$, of independent random variables such that

$$\mathbb{E}[X_1] = \cdots = \mathbb{E}[X_n] = \mu_X$$

, and

$$\mathbb{E}[Y_1] = \cdots = \mathbb{E}[Y_m] = \mu_Y$$

- Assume that the variances of are known so assume (without loss of generality) that

$$\mathsf{var}(X_1) = \cdots = \mathsf{var}(X_n) = \mathsf{var}(Y_1) = \cdots = \mathsf{var}(Y_m) = 1$$

- We want to test:

$$H_0\colon \mu_X = \mu_Y \text{ vs. } H_1\colon \mu_X = \mu_Y$$

with asymptotic level $\alpha \in (0, 1)$.

Two-sample test: large sample case (2)

From CLT:

$$\sqrt{n}(\bar{X}_n - \mu_X) \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0,1)$$

and

$$\sqrt{m}(\bar{Y}_m - \mu_Y) \xrightarrow[m\to\infty]{(d)} \mathcal{N}(0,1) \quad \Rightarrow \quad \sqrt{n}(\bar{Y}_m - \mu_Y) \xrightarrow[\substack{n\to\infty\\ m\to\infty\\ \frac{m}{n}\to\gamma}]{(d)} \mathcal{N}(0,\gamma)$$

Moreover, the two samples are independent so

$$\sqrt{n}(\bar{X}_n - \bar{Y}_m) + \sqrt{n}(\mu_X - \mu_Y) \xrightarrow[\substack{n\to\infty\\ m\to\infty\\ \frac{m}{n}\to\gamma}]{(d)} \mathcal{N}(0,1+\gamma)$$

Under $H_0 : \mu_X = \mu_Y$:

$$\sqrt{n}\frac{\bar{X}_n - \bar{Y}_m}{\sqrt{1+m/n}} \xrightarrow[\substack{n\to\infty\\ m\to\infty\\ \frac{m}{n}\to\gamma}]{(d)} \mathcal{N}(0,1)$$

Test: $$\psi_\alpha = \mathbb{1}\left\{ \sqrt{n}\frac{\bar{X}_n - \bar{Y}_m}{\sqrt{1+m/n}} > q_{\alpha/2} \right\}$$

Two-sample T-test

- If the variances are unknown but we know that $X_i \sim \mathcal{N}(\mu_X, \sigma_X^2),\ Y_i \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$.
- Then
$$\bar{X}_n - \bar{Y}_m \sim \mathcal{N}\Big(\mu_X - \mu_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}\Big)$$
- Under H_0:
$$\frac{\bar{X}_n - \bar{Y}_m}{\sqrt{\sigma_X^2/n + \sigma_Y^2/m}} \sim \mathcal{N}(0,1)$$
- For unknown variance:
$$\frac{\bar{X}_n - \bar{Y}_m}{\sqrt{S_X^2/n + S_Y^2/m}} \sim t_N$$

where
$$N = \frac{\left(S_X^2/n + S_Y^2/m\right)^2}{\frac{S_X^4}{n^2(n-1)} + \frac{S_Y^4}{m^2(m-1)}}$$