A hypothesis is a mathematical statement about the DGP usually about one of its parameters θ e.g θ = 0.9, θ ≠ 0.4, θ ∈ [0.43, 0.78], θ ∉ [0.11, 0.56], etc.

The "alternative hypothesis" (H_a) is the theory you wish to prove. In (II), we assume the opposite temporarily which we call the "null hypothesis" (H_0).

The 3 usual cases for the iid Bernoulli DGP are:

(1) $H_0: \theta \le \theta_0 \iff H_a: \theta > \theta_0$ right-tailed / sided test

(2) $H_0: \theta \ge \theta_0 \iff H_a: \theta < \theta_0$ left-tailed / sided test

(3) $H_0: \theta = \theta_0 \iff H_a: \theta \ne \theta_0$ two-tailed / sided test

(4) $H_0: \theta \ne \theta_0 \iff H_a: \theta = \theta_0$ equivalence test (not covered in 369)

There are two outcomes of the hypothesis test (decisions). Either (I) there is "sufficient" evidence against H_0 and H_0 is "rejected". (II) there is no "sufficient" evidence against H_0 and H_0 fails to be rejected / H_0 is "retained".

How do we decide whether to reject or retain H_0? There are many such tests for every DGP. We'll only study a few tests. How do we run a test? First, we select a "test statistic". This statistic captures a "departure" from H_0's sampling distribution.

As an example. Let DGP = iid Bernoulli(θ), n = 20 and

$H_a: \theta \ne 0.524 \iff H_0: \theta = 0.524$ ← 2020 avg iPhones in USA

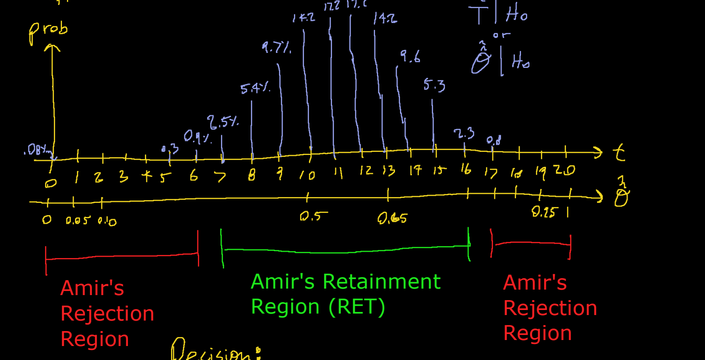
What is the "test statistic"? $\hat{\hat{\theta}} = \frac{\sum x_i}{n} = 0.65$

Is this test statistic "far enough" from H_0 to yield a rejection? We don't know because we don't know yet how variable our point estimate is (nor its distribution) under the assumed H_0. Let's get the sampling distribution under H_0.

$$\hat{T}_n \mid H_0 = \hat{T}_n \mid \theta = 0.524 \sim \text{Binom}(20, 0.524)$$

$$\hat{\theta}_n = \frac{\hat{T}_n}{n} \qquad \hat{\hat{\theta}} = 0.65 \iff t = 13$$

We can use t (the total number of iPhones) also as a test statistic since t is 1:1 with $\hat{\hat{\theta}}$. We know the exact sampling distribution of $\hat{T}_n \mid \theta$ so let's graph the PMF:



Decision:

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

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

What is the probability that the test statistic lands in the Rejection region if H_0 is actually true?

$$\underbrace{P(\hat{\theta} = 0) + P(\hat{\theta} = 0.05) + \ldots + P(\hat{\theta} = 0.30)}_{\text{left tail}} + \underbrace{P(\hat{\theta} = 0.85) + P(\hat{\theta} = 0.90) + P(\hat{\theta} = 0.95) + P(\hat{\theta} = 1)}_{\text{right tail}}$$

= 2.3% = P(Type I error) which is called the "size of the test".

If you're going to reject some of the time, you have to be okay with making a Type I error that some of the time.

If a test has size ≥ level = α the test is said to be a level α test.

Not all sizes are possible in all tests! E.g. we can't get a size = 5% test in this case. The possible sizes are:

$$\left\{ \underbrace{P(\hat{\theta} \le L) + P(\hat{\theta} \ge U)} : L < U,\ L \in \{0, \ldots, 20\},\ U \in \{0, \ldots, 20\} \right\}$$

L is the heighest point estimate to reject before retaining and U is the lowest test point estimate to reject before retaining.

$$\rightarrow F_B(20L, 20, 0.524) + (1 - F_B(20(U-1), 20, 0.524))$$

↳ binomial CDF function (need a calculator)

Thus you cannot "attain" a size for all possible level values here.

However, if the sampling distribution was continuous, then size and level are the same.

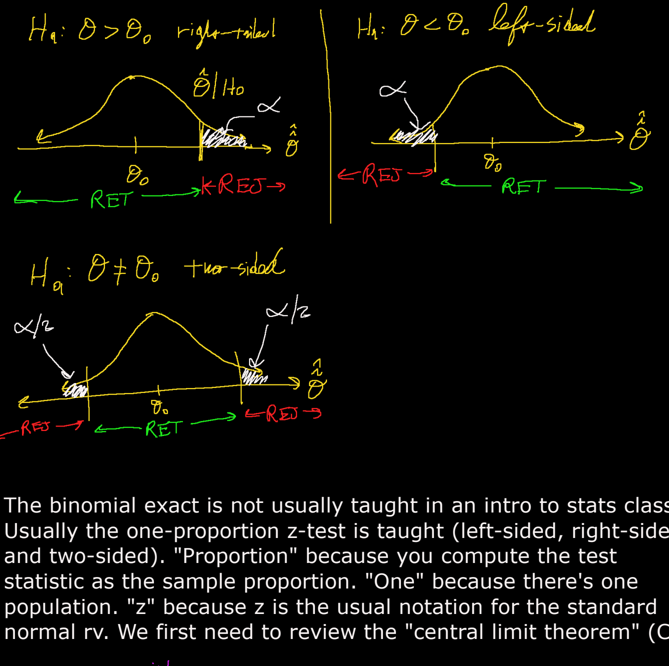
Everything we did above is called a "binomial exact test".

		Decision	
		Retain H_0	Reject H_0
Truth	H_0	✓	Type I error
	not H_0	Type II error	✓

You select / control / pick the Type I error through your choice of α.

Can we find the P(Type II error)? Not yet... we need a few more concepts. But we do know something. If ɑ increases => P(Type II error) decreases and if ɑ decreases => P(Type II error) increases.

Illustrations of what the three types of tests (left, right, two-sided tests) look like if the sampling distribution is continuous and somewhat normally shaped and centered at θ_0.



The binomial exact is not usually taught in an intro to stats class. Usually the one-proportion z-test is taught (left-sided, right-sided and two-sided). "Proportion" because you compute the test statistic as the sample proportion. "One" because there's one population. "z" because z is the usual notation for the standard normal rv. We first need to review the "central limit theorem" (CLT).

$X_1, X_2, \ldots, X_n \overset{iid}{\sim}$ some PMF or PDF (could be unknown) with mean μ and variance σ^2. Then:

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

converges in distribution. It means the CDF of the left hand side becomes more and more like the CDF of the right hand side as n → ∞.

Limits aren't real. n is our sample size and it never gets "infinitely" large. Thus, the CLT has the following useful implications:

$$\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \overset{\cdot}{\sim} N(0, 1),$$

approximately distributed as with the approximation getting better as n gets larger

$$\bar{X}_n \overset{\cdot}{\sim} N\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right), \qquad \bar{X}_n = \frac{T_n}{n}$$

$$\underline{T_n \overset{\cdot}{\sim} N(n\mu, (\sqrt{n}\sigma)^2)}$$

$\Longrightarrow$ In the case of our iid Bernoulli(θ) DGP setting, then

$$\hat{\theta} \overset{\cdot}{\sim} N\left(\theta, \left(\sqrt{\frac{\theta(1-\theta)}{n}}\right)^2\right)$$

This is what will can use as our sampling distribution to do tests. But... it is approximate! So this is an "approximate test". But it's fairly accurate if θ is far from 0 or far from 1.