$\hat{\theta}_{MLE} = \underbrace{s(x_1, \ldots, x_n)}_{\text{some function of the data x\_1, ..., x\_n}} = \bar{x} \in \mathbb{R}$ e.g. 0.1989

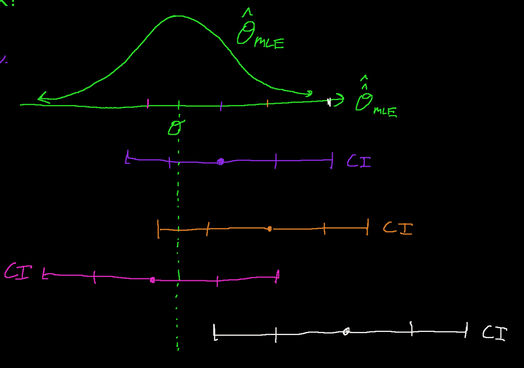
$\mathcal{F}$: iid Bern(θ)

$\hat{\theta}_{MLE} = \underbrace{s(X_1, X_2, \ldots, X_n)}_{\text{the same function except of the rv's}} = \bar{X} \sim N(\theta, SE[\hat{\theta}_{MLE}]^2)$ by property 2

$\bar{X}$: a rv, not a value, rv's have distributions

We use this normality to create the confidence interval (CI).
Why do CI's work?

$\alpha = 5\% \Rightarrow 1 - \alpha = 95\%$
$\Downarrow$
$z_{\alpha/2} = 1.96$



CI's have 95% probability (before they're realized) of capturing (i.e. including) the real value of theta.

Inference goal #3: "hypothesis testing" also called "theory testing".

Consider a situation where *I* am trying to convince *you* of something.

Scenario I:

I declare
H_a: Aliens and UFO's exist.
If they don't exist *you* need to provide *me* *sufficient* evidence that:
H_0: Aliens and UFO's don't exist.

Scenario II:

I'll assume for the moment that
H_0: Aliens and UFO's don't exist
and I will provide *you* *sufficient* evidence to the point that *you're* convinced that
H_a: Aliens and UFO's exist.

Scenario II is more convincing and it is how science generally works. The theory I'm trying to demonstrate is called the "alternative hypothesis" (H_a) since it's alternative to maybe business-as-usual. In scenario II, you assume the oppositive of the theory which is called the "null hypothesis" (H_0). This is the "Hypothesis Testing" procedure.

In our context, theories are phrases as mathematical statements about theta, the unknown parameter. We will study three types of H_a's:

constant value

$H_a: \theta \neq \theta_0$ vs. $H_0: \theta = \theta_0$ → two-sided test / two tailed test

$H_a: \theta > \theta_0$ vs $H_0: \theta \leq \theta_0$ → right-sided test / right-tailed test

$H_a: \theta < \theta_0$ vs. $H_0: \theta \geq \theta_0$ → left-sided test / left-tailed test

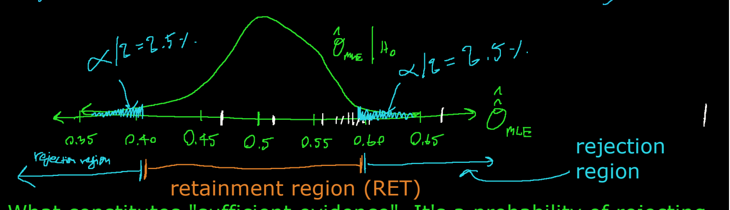
There are two outcomes of a hypothesis test:
(A) You were not shown sufficient evidence of H_a. Thus, you "fail to reject H_0" or "retain H_0".
(B) You were indeed shown sufficient evidence of H_a. Thus, you "reject H_0" or "accept H_a".

Imagine you're flipping a coin n=100 times and you're counting the number of HEADS, then F: iid Bern(theta). You want to prove the coin is unfair.

$H_a: \theta \neq 0.5$

$H_0: \theta = 0.5$

If H_0 is true, $\theta = 0.5 \Rightarrow \underbrace{\hat{\theta}_{MLE}}_{=\bar{X}} \sim N(\underbrace{\theta}_{=0.5}, \underbrace{SE[\hat{\theta}_{MLE}]}_{=\sqrt{\frac{0.5(1-0.5)}{n}} = 0.05}{}^2)$



What constitutes "sufficient evidence". It's a probability of rejecting when H_0 is true (denoted alpha). Everyone is different.

If alpha = 5%, in a 2-tailed test, we put 1/2 the probability in each tail. 5% is the most common scientific standard. Thus, we retain H_0 for the most non-weird 95% of the thetahathats and reject H_0 for the 5% most weird thetahathats.

$$RET = [\theta_0 \pm z_{\alpha/2} SE[\hat{\theta}_{MLE}]] = \Big[0.5 \pm 1.96 \underbrace{\sqrt{\tfrac{0.5(1-0.5)}{100}}}_{=0.05}\Big] = [0.402, 0.598]$$

e.g. if $\bar{x} = \frac{61}{100} = 0.61$, $\alpha = 5\%$

$\hat{\theta}_{MLE} = 0.61 \notin RET \Rightarrow$ reject H_0 and conclude coin is unfair

e.g. if $\bar{x} = \frac{59}{100} = 0.59$

$\hat{\theta}_{MLE} = 0.59 \in RET \Rightarrow$ fail to reject H_0 and conclude there's not enough evidence of coin being unfair

We've covered the "frequentist" approach to statistical inference. But there are problems with it....

1. $\mathcal{F}$ = iid Bern(θ), $x = \langle 0, 0, 0 \rangle$

$\hat{\theta}_{MLE} = \bar{x} = 0$

Is that a good point estimate? NO. You shouldn't be able to say something is absolutely impossible after n=3 trials.

$$CI_{\theta, 1-\alpha} = \Big[0 \pm 1.96 \underbrace{\sqrt{\tfrac{0(1-0)}{3}}}_{0}\Big] = \{0\}$$

Is this a good confidence set? No. This is not a good set of "reasonable values".

2. What if you had prior knowledge that Θ was restricted to e.g. [0.1, 0.2] and not the full (0,1). You can't "enter that into" your inference.

3. Consider the frequentist interpretation of a CI:
   (1) Before you do the experiment, you have a 95% probability of capturing theta. But this doesn't tell you anything about after your experiment. After your experiment you have an interval e.g. [0.37, 0.43] and you can't say:

$$\underbrace{P(\theta \in [0.37, 0.43])}_{\text{no randomness!!}} = 0.95$$

   (2) 95% of CI's will cover theta. But again, I only make one!!! So this interpretation doesn't help me!

In conclusion, any specific CI means NOTHING.

4. Hypothesis tests result in a binary outcome: either you reject H_0 or you fail to reject H_0. What if you want to know

$$P(H_0 \mid x) \quad \text{or} \quad P(H_a \mid x)?$$

You cannot!!! One thing you can do is:

$p_{val} := P($seeing thetahathat or more extreme | H_0$) \neq P(H_0 \mid x)$

5. $\mathcal{F}$: iid Bern(θ), $x = \langle 0, 1, 0 \rangle \Rightarrow \hat{\theta}_{MLE} = 0.33$

$$CI_{\theta, 95\%} = \Big[0.33 \pm 1.96 \sqrt{\tfrac{0.33 \cdot 0.67}{3}}\Big] = [-0.20, 0.87]$$

Is this a reasonable confidence set? No. It's outside of the legal parameter space which is (0,1).

$\alpha = 5\%$,
$H_0: \theta = 0.5 \Rightarrow RET = \Big[0.5 \pm 1.96 \sqrt{\tfrac{0.5 \cdot 0.5}{3}}\Big] = [-0.066, 1.066]$

Is this a good hypothesis test? No becaues you NEVER reject!!!

The problem in #5 is because the asymptotic normality of the MLE doesn't "kick in" until n is large (MLE property 2 is not true yet).

We will solve all these problems with Bayesian Inference. Unfortunately, we will get other problems instead. It's a tradeoff and a personal decision.