Hypothesis Testing

We have some data $\{X_1, X_2, \ldots, X_n\} = \{X_i\}_{i \leq n}$

Hypothesis: $E[X] = \mu_0$ is H_0

CLT says: if $\{X_i\}_{i \leq n}$ are i.i.d. and $0 < \sigma_x^2 < \infty$,

$$\bar{X} \sim N\left(\mu_x, \frac{\sigma_x^2}{n}\right) \text{ when } n \text{ is "large"}$$

So under H_0, $\bar{X} \sim N\left(\mu_0, \frac{\sigma_x^2}{n}\right)$

What is the likelihood of observing $\bar{X}^{data}$, if H_0 were true?

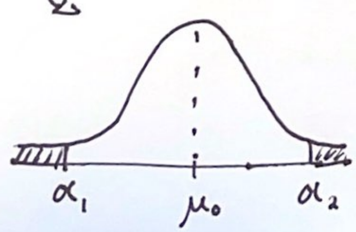
Well, suppose we are comfortable with a 5% chance of falsely rejecting H_0.

Then choose α_1 and α_2 st:

$$P(\alpha_1 \leq \bar{X} \leq \alpha_2) = 0.95$$

ie there is a 95% chance $\bar{X}^{data}$ is in this region

$N\left(\mu_0, \frac{\sigma_x^2}{n}\right)$



Since N is symmetric, $|\alpha_1| = |\alpha_2|$

This is equivalent to

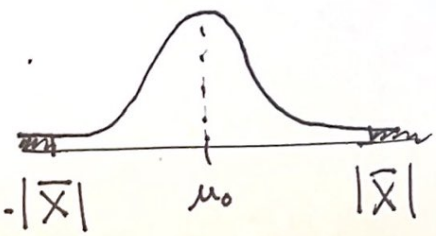
$$1 - P(\alpha_1 \leq \bar{X} \leq \alpha_2) = 0.05 \leftrightarrow P(\bar{X} < \alpha_1 \mid \bar{X} > \alpha_2) = 0.05$$

So mass in <u>each</u> tail is 0.025

INTUITION: If H_0 is true, very unlikely that we draw a value in shaded region!

So, conditional on having drawn a value in the tail, it is not likely H_0 is true.

Alternative way to think about this is using the p-value. Again, draw the distribution under H_0; then look at the likelihood of that observation.



P-value is the sum of probability (area) in the tails.

If p-value is small, H_0 is unlikely!

Exact same idea as before, but perhaps more informative.

Note: in reality (and in lecture) we use standardized variables. It does not change the logic or intuition.