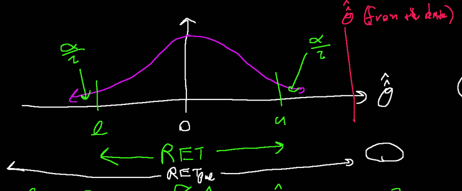
Let's go with (a) the difference in sample averages. So, under the null, $\bar{X}_{b,1} - \bar{X}_{b,2} \approx 0$, It won't be exactly zero due to chance variations, so what is the threshold of nonzero to reject H_0? We need the sampling distribution of this difference in sample averages under H_0.

There are many such fake samples 200-choose-100 = 10^58. So with a big computer, let's take B = 1,000,000. The higher the B is, the more accurate, but not much more accurate. Let's calculate the test statistic for all B "resamplings" and plot a histogram:



$$\ell := \text{Quantile}\left[\{\hat{\hat{\theta}}_1, \ldots, \hat{\hat{\theta}}_B\}, \frac{\alpha}{2}\right]$$

$$u := \text{Quantile}\left[\{\hat{\hat{\theta}}_1, \ldots, \hat{\hat{\theta}}_B\}, 1 - \frac{\alpha}{2}\right]$$

How to get the 1-sided pval. You count the number of thetahathat_b's that are more extreme than the thetahathat from the original data.

You can also used Fisher's permutation idea to get a CI for theta. This is difficult computationally so we won't study it.

We will now see a resampling method that is extremely famous and etremely useful: Efron's non-parametric bootstrap (1979). ch8 AoS and the idea is remarkably simple. Imagine you have an iid DGP f(x; theta_1, ..., theta_K) and you have some parameter / function of parameters you want inference for e.g.

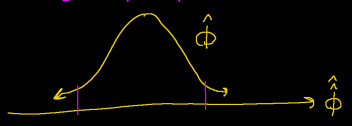
$$\phi = g(\theta_1, \ldots, \theta_K) \text{ estimated by } \hat{\hat{\phi}} = w(x_1, \ldots, x_n)$$

e.g. $\phi = \text{Med}[X] = g(\theta_1, \ldots, \theta_K)$ but $\hat{\hat{\phi}} = \text{Median}[\{x_1, \ldots, x_n\}]$

We don't know how to get the estimator's distribution. Which means we can't test hypotheses or construct CI's. But... the bootstrap gives us a way to find the asymptotic distribution of the estimator:

For b = 1, ..., B, a large number of resamplings,
* take a fake sample of size n by sampling with replacement of the original n data points (this will give you about 2/3 of the unique data points and 1/3 left out)
* for each fake sample, I calculate phihathat_b. The collection of these B statistics is called the "boostrap distribution"

Thm: the boostrap distribution approximates the real distribution of the estimator as B --> infinity and n --> infinity. So B should be as big as possible given your practical constraints.



"Bootstrap" comes from the English expression "pull oneself up by the boostraps" which means to miraculously succeed kind of like defying gravity.

One way to get a CI is to just use the quantiles of the boostrap distribution:

$$CI_{\phi, 1-\alpha} = \left[\text{Quantile}\left[\{\hat{\hat{\phi}}_1, \ldots, \hat{\hat{\phi}}_B\}, \frac{\alpha}{2}\right], \text{Quantile}\left[\{\hat{\hat{\phi}}_1, \ldots, \hat{\hat{\phi}}_B\}, 1 - \frac{\alpha}{2}\right]\right]$$

there are other ways too that we won't study. One way to do a two-sided hypothesis test (i.e. against H_0: phi = phi_0) is to just check

$$\phi_0 \in CI_{\phi, 1-\alpha} \overset{\text{yes}}{\Longrightarrow} \text{Retain } H_0.$$

It's very simple but very approximate. It's always better to use a distributional result if you have it.

In finance, people care about the Sharpe Ratio defined as:

$$\phi = \frac{E[X] - r_{free}}{SD[X]}$$

a constant representing the risk free rate

est.

$$\hat{\hat{\phi}} = \frac{\bar{X} - r_{free}}{\sqrt{\frac{1}{n} \sum (X_i - \bar{X})^2}} \qquad \hat{\phi} \sim ?$$

Randomization and Causal Estimation

Consider two populations.

Scenario 1:
Pop 1: students who hand in their test at the end of the exam and
Pop 2: students who hand in their test early.

Let y_1,1, ..., y_1,n_1 denote test score data from Pop 1
Let y_2,1, ..., y_2,n_2 denote test score data from Pop 2

We want to test if the means of these two populations are different i.e. against H_0: theta_1 = theta_2. We know nothing about the DGP's but we can use the Wald test (2-sample z-test):

$$\text{If } \frac{\hat{\hat{\theta}}_1 - \hat{\hat{\theta}}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \in [-1.96, 1.96] \Rightarrow \text{Retain } H_0 \text{ at } \alpha = 5\%.$$

Let's say you reject and then you can conclude: "there is a difference in the mean test scores betwen students who hand in their exam early and hand in their exam on time". (statement 1a)