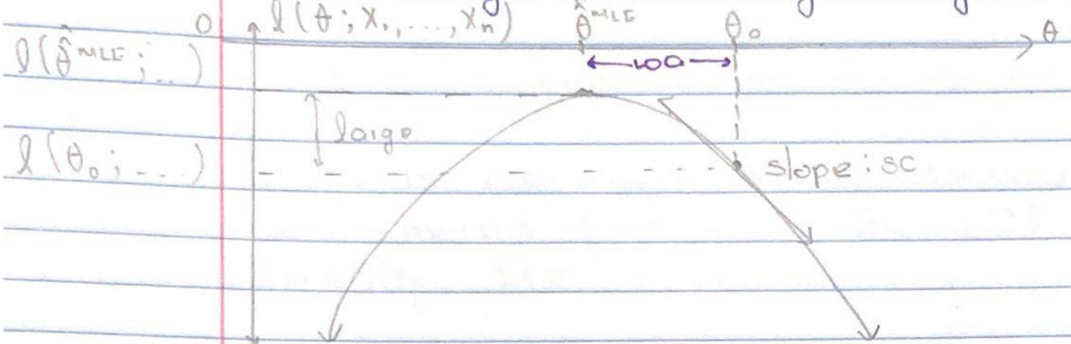
Lecture 21 11/25/2020

Thm: The wald test, the score test and the LR test are all "asymptotically equivalent" which means as n gets large, the decision is the same for all three tests. Here's an illustration to show how they similar they really are. For $H_0: \theta = \theta_0$,



How about a general test for goodness of fit? Pearson χ^2 required k categories. What if I have the following data:

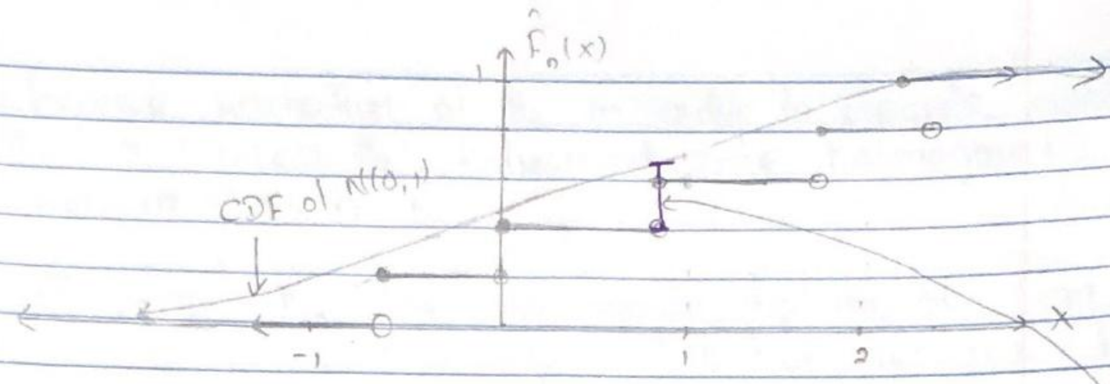
$X_1 = 1.73$, $X_2 = -0.49$, $X_3 = 0.93$, $X_4 = 2.16$, $X_5 = 0.03$

and I want to test against H_0: DGP is iid $N(0,1)$ vs H_a: any other DGP

You can't use the LRT here because the LRT would force you to have an $H_a: N(\theta_1, \theta_2)$. For continuous data (DGP), we can use the Kolmogorov - Smirnov (KS) test. Here's how it works.

We first compute the "empirical distribution function", $\hat{F}_n$:

$$F_n(x) := \frac{\#[X_i \leq x]}{n}$$ which is monotonically increasing (but not strictly because it goes flat between X's).



$\hat{F}_n$ is a "function estimator" for the "true" CDF, $F(x)$.

If H_0 is assumed, then I assume a DGP explicity which means I assume the CDF of the data explicitly, $F(x) = F_{H_0}(x)$

So now to test, we need a test statistic that gauges the data's departure from H_0. What should that look like?

$$D_n = \text{difference}(\hat{F}_n(x), F_{H_0}(x))$$

The KS test uses the "supremum norm difference" which means it measures the "largest absolute difference between the two over all x":

$$D_n := \sup[|\hat{F}_n(x) - F_{H_0}(x)|] \leq 1 \quad \text{e.g.}$$

Advanced note: the Glivenko-Cantelli thm (1933) proves that D_n converges to zero under H_0. This means that the empirical distribution function converges to the real CDF at all x. They also prove that if H_0 is false then it converges to something > 0 meaning that the power of the KS test converges to 100%.

We need the sampling distribution to see if our sample d_n value ($\hat{\theta}$) is within tolerance limits chance variation of H_0 in order to decide "retain H_0" or "reject H_0". Advanced note: Kolmogorov proved in 1933 that:

$\sqrt{n}\, D_n \xrightarrow{d} K$, the "Kolmogorov distribution", an amazing "distribution-free" result kind of like the CLT.

$$\Downarrow$$

$\sqrt{n}\, D_n \dot{\sim} K$

Tables of critical values are precomputed. For example at $\alpha = 5\%$, the critical cutoff is a K value of 1.359. However, these critical value are very approximate for $n < 50$, so there are better tables for finite n. We won't bother with that in 369.

There is an extension to the KS test for non-continuous DGP's which we won't cover in this class.

A major limitation to the KS test is you need a null hypothesis which is an explicit DGP i.e. parameter values specified e.g. H_0: iid $N(0,1)$. You can't say "H_0: normal! You need a different test for that situation. One example is the Shapiro-Wilk test which we don't study.

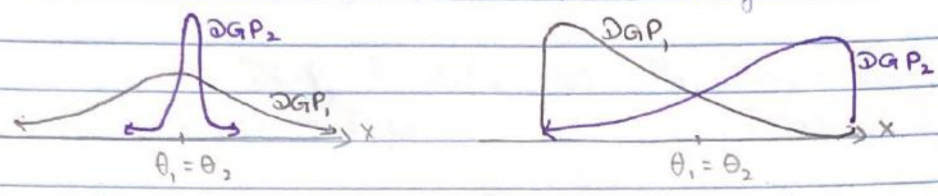
What if you have the following setup: you are sampled from two different populations independently:
$X_{11}, X_{12}, \ldots, X_{1n} \overset{iid}{\sim} DGP_1$, independent of $X_{21}, X_{22}, \ldots, X_{2n_2} \overset{iid}{\sim} DGP_2$
and you want to test against.

H_0: $DGP_1 = DGP_2$ i.e. $F_1(x) = F_2(x)$ vs
H_a: DGP_1 is not the same as DGP_2 i.e. $F_1(x) \neq F_2(x)$

Let $\theta_1 := E[X_1]$ i.e. DGP_1, $\theta_2 := E[X_2]$ i.e. DGP_2.

If $\theta_1 \neq \theta_2 \Rightarrow F_1(x) \neq F_2(x)$ so why not just test via Wald $H_a: \theta_1 \neq \theta_2$?

Because $\theta_1 = \theta_2 \not\Rightarrow F_1(x) = F_2(x)$ e.g.



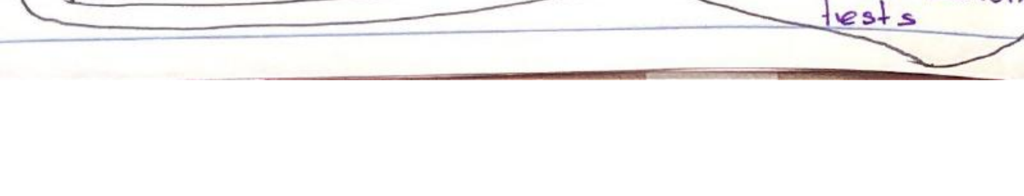
For continuous DGP's, Kolmogorov also proved the following result:

$$\sqrt{\frac{n_1 n_2}{n_1 + n_2}}\; D_{n,m} \xrightarrow{d} K \text{ where } D_{n,m} := \sup_x \left[\left| \hat{F}_1(x) - \hat{F}_2(x) \right| \right]$$

The Anderson - Darling (AD) test is very similar (same setup for both one-sample and two-sample goodness of fit tests) so we won't study it. For non-continuous you can use the Mann-Whitney U test, but we won't study that either.

The two-sample KS, AD, U tests are examples of "nonparametric tests" which means we make no explicit assumptions on the functional forms of the DGP's. They're also called "distribution-free" tests.

There is a completely different way of doing this type of testing which are called "resampling methods". And there are many different ones:



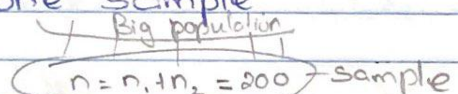
We will now study peimutation tests.

Assume the some setup as the 2-sample KS test. We have iid samples from DGP$_1$ indepdent of iid sample from DGP$_2$ (the two populations) and the same test objection:

$$H_a: DGP_1 \neq DGP_2, \quad vs \quad H_0: DGP_1 = DGP_2,$$
$$F_1(x) \neq F_2(x) \qquad\qquad F_1(x) = F_2(x)$$

Fisher in 1936 had the following thought experiment. Imagine $n_1 = 100$ Englishmen and $n_2 = 100$ Frechmen and you measure their heights denoted $x_{1,1}, \ldots, x_{1,100}$ and $x_{2,1}, \ldots x_{2,100}$.

Under H_0, the DGP's are the same so there's no distinction between population 1 (sample 1) and population 2 (sample 2). So we can imagine just one all-inclusive population and one sample:



You've seen this idea before. Remember the 2-prop z-test? We estimated θ with

$$\hat{\theta}_{pooled} := \frac{\Sigma x_{1i} + \Sigma x_{2i}}{n_1 + n_2}.$$

Thus, we can imagine arbitrarily dividing this "master sample" into two 100-sized pieces. To draw the first piece, take a random sample of the master sample of size 100 and then second piece is the data left over:

1^{st} fake sample 1: Some subset of size n_1 of $\{X_{11}, X_{12}, \ldots, X_{1n_1}, X_{21}, \ldots X_{2n_2}\}$

1^{st} fake sample 2: $\{X_{11}, \ldots, X_{1n_1}, X_{21}, \ldots, X_{2n_2}\} \setminus 1^{st}$ fake sample 1.

Draw B of these fake samples where B is a large number.

Let $R_{b,1} \subset \{1, 2, \ldots, n\}$ and $R_{b,2} \subset \{1, 2, \ldots, n\}$, $|R_{b,1}| = n_1$, $|R_{b,2}| = n_2$, $R_{b,1} \cup R_{b,2} = \{1, 2, \ldots, n\}$ where R represents a set of indicates in b^{th} fake sample 1 and the b^{th} fake sample 2.

Now calculate some statistic $\hat{\theta}_b$ which measures depature from H_0 and there are lots of choices:

(a) $\hat{\theta}_b = \bar{X}_{b,1} - \bar{X}_{b,2} = \frac{1}{n_1} \sum_{i \in R_{b,1}} X_i - \frac{1}{n_2} \sum_{i \in R_{b,2}} X_i$

(b) $\hat{\theta}_b = \text{Med}[\{X_i : i \in R_{b,1}\}] - \text{Med}[\{X_i : i \in R_{b,2}\}]$

(c) $\hat{\theta}_b = d_{n_1,n_2}$ from the 2-sample KS test.

(d) $\hat{\theta}_b = \dfrac{\bar{X}_{b,1}}{\bar{X}_{b,2}}$