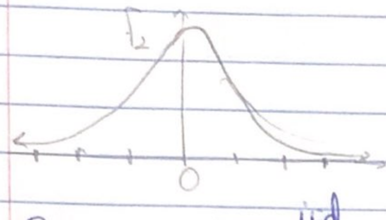
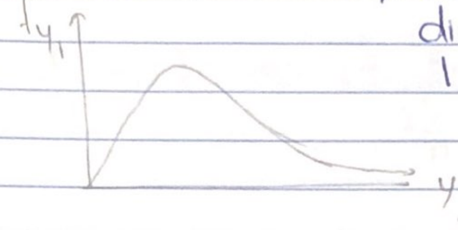
Lecture -15 10/28/2020

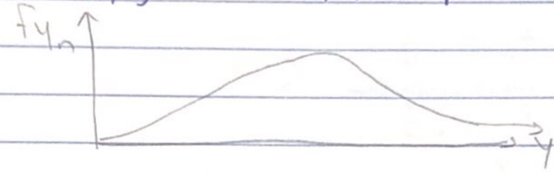
We need a couple of facts from Math 368 from distribution theory:

$Z \sim N(0,1) \implies Y_1 := Z^2 \sim \chi^2_1$ chi-squared distribution with 1 degree of freedom



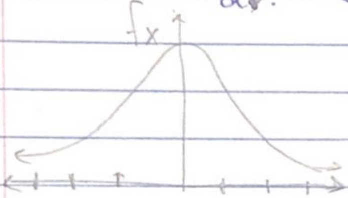


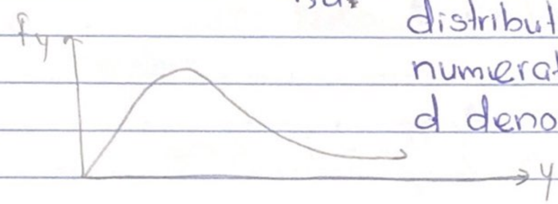
$Z_1, \ldots, Z_n \overset{iid}{\sim} N(0,1) \implies Y_n := Z_1^2 + \ldots + Z_n^2 \sim \chi^2_n$



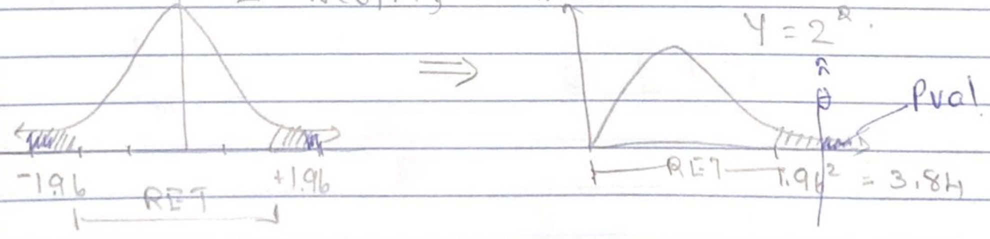
$X \sim T_{df} \implies Y = X^2 \sim F_{1,df}$ Fisher-Snecedor distribution with 1 numerator df and d denominator df.

tails are thicker than normal





This means that every 2-sided Z-test (exact or approx) is equivalent to a chi-squared test (exact or approx) and every 2-sided t-test (exact or approx) is equivalent to an F test (exact or approx).



DGP is iid Normal mean θ, variance σ^2, σ known and the estimator is the sample average and you're testing Ha: θ is not equal to θ_0,

$$\frac{\hat{\theta}-\theta_0}{\sigma/\sqrt{n}} \sim N(0,1) \implies \frac{(\hat{\theta}-\theta_0)^2}{\sigma^2/n} \sim \chi^2,$$

DGP is iid Bern(θ), and same as above,

$$\frac{\hat{\theta}-\theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}} \dot{\sim} N(0,1) \implies \frac{(\hat{\theta}-\theta_0)^2}{\frac{\theta_0(1-\theta_0)}{n}} \dot{\sim} \chi^2_1$$

DGP is iid normal with both θ and σ^2 unknown, the same as above,

$$\frac{\hat{\theta}-\theta_0}{s/\sqrt{n}} \sim T_{n-1} \implies \frac{(\hat{\theta}-\theta_0)^2}{s^2/n} \sim F_{1,n-1}$$

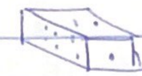
Let's say you want to prove a coin is weighted unfairly. So you assume its flips have the DGP iid Bern(θ), and you test Ha: θ is not ½.

$$\hat{\theta}_{std} = \frac{\hat{\theta}-\frac{1}{2}}{\frac{1/2}{\sqrt{n}}} \overset{?}{\in} [1.96, 1.96]$$

$$\vec{\theta} = \begin{bmatrix} \theta_1 = P(\text{die} = 1) \\ \theta_2 = P(\text{die} = 2) \\ \vdots \\ \theta_6 = P(\text{die} = 6) \end{bmatrix}$$

$H_a: \theta \neq \frac{1}{2}$
$H_0 = \theta = \frac{1}{2}$.

Let's say you want to prove a 6-sided die is fair.



At least one θ is not $\frac{1}{6}$.

H_a: die is unfair $\exists_j \, \theta_j \neq \frac{1}{6}$ $\quad \vec{\theta} \neq \vec{\theta}_0 = \frac{1}{6}T$
H_0 : die is fair. $\theta_1 = \theta_2 = \dots = \theta_6 = \frac{1}{6}$ or $\vec{\theta} = \vec{\theta}_0 = \frac{1}{6}T$.

Given n rolls of the die $x_1, \dots, x_n$, how do we our test? We need some way to measure/gauge departure from H_0 is (a statistically or a set of statistics). Let's look at a frequency table e.g.

	Roll #							
	1	2	3	4	5	6	Total	rvs
Observed Quantity	4	1	3	2	1	4	n = 15	$Q_1 Q_2 \dots Q_6$:
Expected quantity	2.5	2.5	2.5	2.5	2.5	2.5	n = 15	

$E_1, E_2, \dots, E_6$ = constant.

$\hat{\phi} = (O_1 - E_1) + (O_2 - E_2) + \dots + (O_6 - E_6)$ if $\hat{\phi}$ is large $\Rightarrow$ Reject H_0

Maybe . . .

$\hat{\phi} = |O_1 - E_1| + \dots + |O_6 - E_6|$

This is a good estimator for departure from the null hypothesis . . . but we don't know its sampling distribution making is unusable in practice.

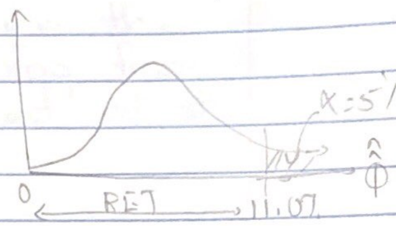
$$\hat{\phi} = \frac{[Q_1 - E_1]^2}{E_1} + \dots + \frac{(Q_6 - E_6)^2}{E_6} \xrightarrow{d} \chi^2_5$$

this fact is proved in Math 368 if we had more time.

Karl Pearson (1900) and it's named the "chi-squard goodness of fit test". In general, if there are k categories (e.g. here $k=6$), then the following:

$$\hat{\phi} = \sum_{k=1}^{K} \frac{(O_k - E_k)^2}{E_k} \xrightarrow{d} \chi^2_{K-1}$$

Let's run our "die unfair test" for the data above at $\alpha = 5\%$; $F_{\chi^2_5}(11.07) = 95\%$.



$$\hat{\phi} = \frac{(4-2.5)^2}{2.5} + \dots + \frac{(4-2.5)^2}{2.5}$$

$= 3.8 \in RET \Longrightarrow$ Retail H_0

New situation. Let's look at data for $n = 279$ men and record their hair color and eye color. Here's the raw data as a "contingency table" or "cross tabulation":

\# rows, $r = 4$

\# colums, $c = 4$

		EYE COLOR				
		Brown (EB)	Blue (EL)	Hazel (EZ)	Green (EG)	Total
Hair color	Black (HB)	32	11	10	3	$n_{HB} = 56 = n_{1.}$
	Brown (HO)	53	50	25	15	$n_{HO} = 143 = n_{2.}$
	Red (HR)	10	10	7	7	$n_{HR} = 34 = n_{3.}$
	Blonde (HL)	3	30	5	8	$n_{HL} = 46 = n_{4.}$
		$98 = n_{EB} = n_{.1}$	$101 = n_{EL} = n_{.2}$	$47 = n_{EZ} = n_{.3}$	$33 = n_{EG} = n_{.4}$	$n = 279$

I want to test if hair

H_a: hair color and eye color are dependent events.

H_0: hair color and eye color are independent events.

Let θ denote a true population probability
e.g. $\theta_{HB,EB} = \theta_{1,1}$ = P(black hair and brown eyes),
$\theta_{HB} = \theta_{1,\cdot}$ = P(black hair)

$H_a: \exists_{j,k}$ s.t $\theta_{jk} \neq \theta_{j\cdot}\theta_{\cdot k}$ i.e. at last one is unequal

$$H_0: \theta_{1,1} = \theta_{1\cdot}\theta_{\cdot 1}, \theta_{1,2} = \theta_{1\cdot}\theta_{\cdot 2}, \ldots, \theta_{4,4} = \theta_{4\cdot}\theta_{\cdot 4}$$

H_0 is $r \times c = 4 \times 4 = 16$ equalities.

We need a statistic to gauge the departure from H_0. Let's follow the reasoning from the previous example. We first looked at the data we expect if H_0 was true.

Eye color

Hair color	1	2	3	4	Tot
1	$E_{11} = n\theta_{1\cdot}\theta_{\cdot 1}$	$E_{12} = n\theta_{1\cdot}\theta_{\cdot 2}$			
2					
3			$E_{33} = n\theta_{3\cdot}\theta_{\cdot 3}$		
4				etc	
Tot					

$$\phi = \frac{(O_{11} - E_{11})^2}{E_{11}} + \ldots + \frac{(O_{44} - E_{44})^2}{E_{44}}$$

$$= \frac{(O_{11} - n\theta_{1\cdot}\theta_{\cdot 1})^2}{n\theta_{1\cdot}\theta_{\cdot 1}} + \ldots + \frac{(O_{44} - n\theta_{4\cdot}\theta_{\cdot 4})^2}{n\theta_{4\cdot}\theta_{\cdot 4}}$$

Can we compute phihathat? No. You do not know any of the θ_i's or any of the θ_j's.

How about we [Richardify and] replace the θ_i's & θ_j's with $\hat{\theta}_i$'s and $\hat{\theta}_j$'s. Yes....