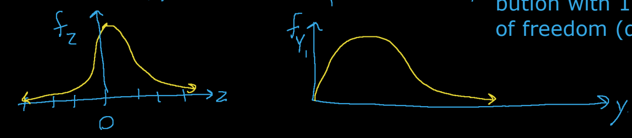
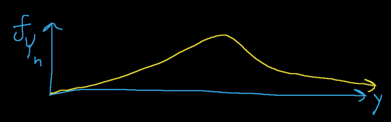
We need a couple of facts from Math 368 from distribution theory:

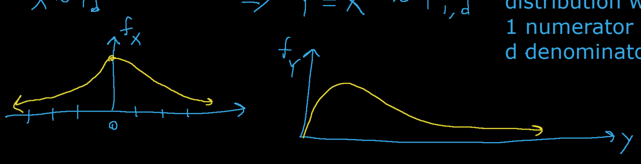
$Z \sim N(0,1) \Rightarrow Y_1 := Z^2 \sim \chi^2_1$ chi-squared distribution with 1 degree of freedom (df)



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

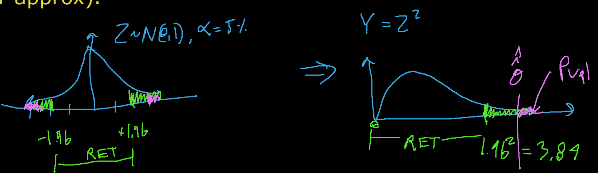


$X \sim T_d \Rightarrow Y = X^2 \sim F_{1,d}$ Fisher-Snecedor distribution with 1 numerator df and d denominator df.



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).

$Z \sim N(0,1), \alpha = 5\% \Rightarrow Y = Z^2$



DGP is iid Normal mean theta, variance sigsq, sigsq known and the estimator is the sample average and you're testing H_a: theta is not equal to theta_0,

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

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

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

DGP is iid normal with both theta and sigsq unknown, and same as above,

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

Let's say you want to prove a coin is weighted unfairly. So you assume its flips have the DGP iid Bern(theta), and you test H_a: theta is not 1/2.

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

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

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

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

At least one θ is not $\frac{1}{6}$.

H_a: die is unfair $\exists j \; \theta_j \neq \frac{1}{6}$ $\quad \vec{\theta} \neq \vec{\theta}_0 = \frac{1}{6}\vec{1}$

H_0: die is fair $\theta_1 = \theta_2 = \ldots = \theta_6 = \frac{1}{6}$ or $\vec{\theta} = \vec{\theta}_0 = \frac{1}{6}\vec{1}$

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

Roll #	1	2	3	4	5	6	Total
Observed Quantity	4	1	3	2	1	4	n=15
Expected Quantity	2.5	2.5	2.5	2.5	2.5	2.5	n=15

$O_1, O_2, \ldots, O_6$ rv's

$E_1, E_2, \ldots, E_6$ constants

$$\hat{\Phi} = (O_1 - E_1) + (O_2 - E_2) + \ldots + (O_6 - E_6) \quad \text{if } \hat{\Phi} \text{ large} \Rightarrow \text{Reject } H_0$$

Maybe...

$$\hat{\Phi} = |O_1 - E_1| + \ldots + |O_6 - E_6|$$

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

$$\hat{\Phi} = \frac{(O_1 - E_1)^2}{E_1} + \ldots + \frac{(O_6 - E_6)^2}{E_6} \xrightarrow{d} \chi^2_5$$

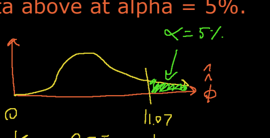
this fact is proved in Math 368 if we had more time

Karl Pearson (1900) and it's named the "chi-squared 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%.

$F_{\chi^2_5}(11.07) = 95\%$



$$\hat{\Phi} = \frac{(4-2.5)^2}{2.5} + \ldots + \frac{(4-2.5)^2}{2.5} = 3.8 \in RET \Rightarrow \text{Retain } H_0$$

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, $r = 4$
cols, $c = 4$

Eye Color

Hair Color	Brown (EB)	Blue (EL)	Hazel (EZ)	Green (EG)	Total
Black (HB)	$32 = O_{11}$	11	10	3	$56 = n_{HB} = n_{1\cdot}$
Brown (HO)	$53 = O_{21}$	50	25	$15 = O_{24}$	$143 = n_{HO} = n_{2\cdot}$
Red (HR)	10	10	$7 = O_{33}$	7	$34 = n_{HR} = n_{3\cdot}$
Blonde (HL)	3	30	5	8	$46 = n_{HL} = n_{4\cdot}$
Total	$98 = n_{EB} = n_{\cdot 1}$	$101 = n_{EL} = n_{\cdot 2}$	$47 = n_{EZ} = n_{\cdot 3}$	$33 = n_{EG} = n_{\cdot 4}$	$n = 279$

I want to test
H_a: hair color and eye color are dependent events
H_0: hair color and eye color are independent events
Let theta denote a true population probability e.g.
theta_HB,EB = theta_1,1 = P(black hair and brown eyes),
theta_HB = theta_1,. = 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 x c = 4 x 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					

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

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

Can we compute phihathat? NO. You do not know any of the theta_i.'s or any of the theta_.j's.

How about we [Richardify and] replace the theta_i.'s and theta_.j's with thetahat_i.'s and thetahat_.j's. Yes...