M

→ Ksenia : part I.

→ Boris : part II → mark. scheme, → scans, → grades

→ Max : ha.

Contingency tables.

1-d cont. table:

Grades	A	B	C	D
# Stud	20	30	20	10

observed frequencies

n - number of obs.

Assumpt. $X_1, X_2 \dots X_n$ iid

$X_i \in \{A, B, C, D\}$

prob-s in H_0

H_0: $p_A = 0.25$ $p_B = 0.25$ $p_C = 0.25$ $p_D = 0.25$

H_A: not all prob are correctly specified in H_0.

$\alpha = 0.01$

How to test H_0?

→ stand. LR / LM / Wald *[2]
MaxLik.

→ historical test (Pearson) *[1]

r - number of cat.

$$F.S. = \sum_{i=1}^{r} \frac{(N_i - E(N_i))^2}{E(N_i)}$$

wiki: $$\sum_{i=1}^{r} \frac{(O_i - E_i)^2}{E_i}$$

I'am scared → $$\sum \frac{(O - E)^2}{E}$$

zero? !!

Under H_0: $$\sum_{i=1}^{r} \frac{(N_i - E(N_i))^2}{E(N_i)} \xrightarrow[n \to \infty]{dist} \chi^2_{r-1}$$

Grades	A	B	C	D
# Stud	20	30	20	10

observed frequencies

n - number of obs.

Assumpt.

$X_1, X_2, \dots X_n$ iid

$X_i \in \{A, B, C, D\}$

prob-s in H_0

H_0: $p_A = 0.25 \quad p_B = 0.25 \quad p_C = 0.25 \quad p_D = 0.25$

H_A: not all prob are correctly specified in H_0.

$N_1 = 20 \qquad N_4 = 10$

$N_2 = 30$

$N_3 = 20$

$n = 20 + 30 + 20 + 10 = 80$

$$N_1 \overset{H_0}{\sim} Bin(n = 80, \; p_A = 0.25)$$

$$E(N_1) = 80 \cdot 0.25 = 20$$

$$N_2 \overset{H_0}{\sim} Bin(n = 80, \; p_B = 0.25)$$

$$E(N_2) = \dots = 20$$

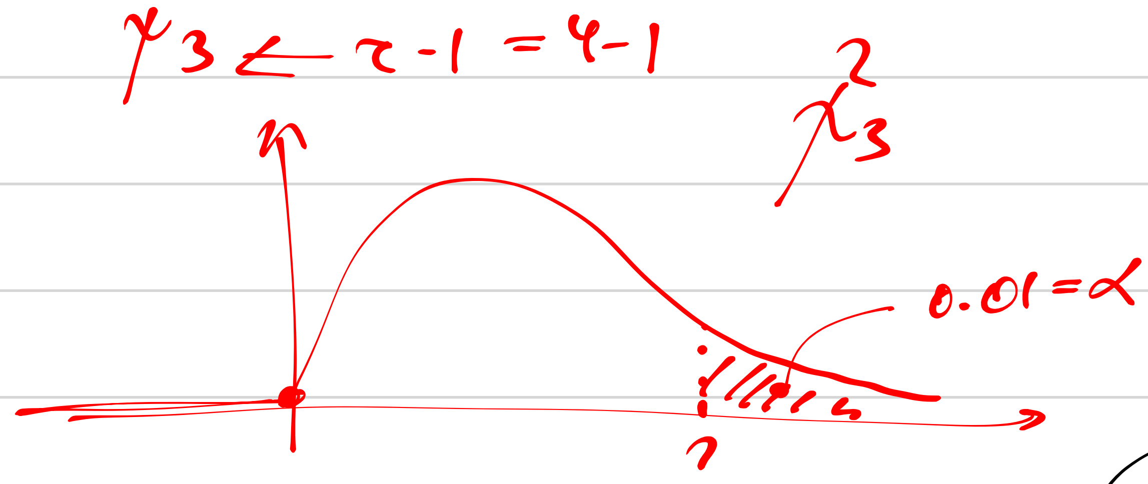
$$\dots E(N_3) = 20 \qquad E(N_4) = 20$$

Pearson's statistic

$$P.S. = \frac{(20-20)^2}{20} + \frac{(30-20)^2}{20} + \frac{(20-20)^2}{20} + \frac{(10-20)^2}{20} =$$

$$= \frac{100}{10} = 10. = PS_{obs}$$

$$P.S. \overset{approx}{\sim} \chi^2_3 \leftarrow r - 1 = 4 - 1$$



stat. tables / software: $\chi^2_{crit} = 11.34$

Stat. Conclusion: Do not reject H_0: $p = (0.25, 0.25, 0.25, 0.25)$.

LR (likelihood ratio) approach

$\hat{p}_1, \hat{p}_B, \hat{p}_C$

Mod_UR

Mod_R

Mod_UR : $p_A + p_B + p_C + p_D = 1$

Mod_R: $p_A = 0.25$, $p_B = 0.25$, $p_C = 0.25$, $p_D = 0.25$.

$$LR = 2\left(\ln L_{UR} - \ln L_R\right) \xrightarrow[n\to\infty \text{ dist}]{H_0} \chi^2_{df}$$

* (free)

$df = p_{UR} - p_R$

p_{UR} - number* of params in UR model

p_R -- // -- in R model.

$p_{UR} \overset{?}{=} 3$

$p_R \overset{?}{=} 0$

$df = 3 - 0 = 3$

binomial case

$$P(N_1 = 60, N_2 = 40) = \frac{100!}{60!\,40!} p_1^{60} \cdot (1-p_1)^{40}$$

Theorem

Under H_0: $\text{plim} \frac{PS_n}{LR_n} = 1$

multinomial case

$$L = P(N_1 = 20, N_2 = 30, N_3 = 20, N_4 = 10) =$$

$$= \frac{80!}{20!\,30!\,20!\,10!} \cdot \boxed{p_A^{20} \cdot p_B^{30} \cdot p_C^{20} \cdot (1 - p_A - p_B - p_C)^{10}}$$

M

$$\ln L = \ln M + 20 \ln p_A + 30 \ln p_B + 20 \ln p_C + 10 \log(1 - p_A - p_B - p_C)$$

$\underbrace{AA\ldots A}_{20\text{ letters}}\ \underbrace{B\ldots B}_{30\text{ letters}}\ \underbrace{C\ldots C}_{20}\ \underbrace{D.D}_{10}$

$$LR = (\ln L_{UR} - \ln L_R) \cdot 2$$

$+\ln M$ $\qquad$ $+\ln M$

$N_1 \sim Bin$

$N_2 \sim Bin$

$N_3 \sim Bin$

$N_4 \sim Bin$

$N_1 + N_2 + N_3 + N_4 = 80$

$$LR = 2\cdot\left(\underline{20}\cdot\left(\ln \hat{p}_1^{UR} - \ln \hat{p}_1^{R}\right) + \underline{30}\cdot\left(\ln \hat{p}_2^{UR} - \ln \hat{p}_2^{R}\right) + \underline{20}\left(\ldots\right) + \underline{10}\cdot\left(\ln \hat{p}_4^{UR} - \ln \hat{p}_4^{R}\right)\right)$$

$\boxed{\text{in R-model:}}$

$\hat{p}_1^R = 0,25$
$\hat{p}_2^R = 0,25$
$\hat{p}_3^R = 0.25$
$\hat{p}_4^R = 0,25$

$\boxed{\text{in UR-model}}$

$\max_{\hat{p}_1^{UR},\, \hat{p}_2^{UR},\, \hat{p}_1^{UR}}$

$$20 \ln \hat{p}_1^{UR} + 30 \ln \hat{p}_2^{UR} + 20 \ln \hat{p}_3^{UR} + 10 \cdot \ln\left(1 - \hat{p}_1^{UR} - \hat{p}_2^{UR} - \hat{p}_3^{UR}\right)$$

intuit.

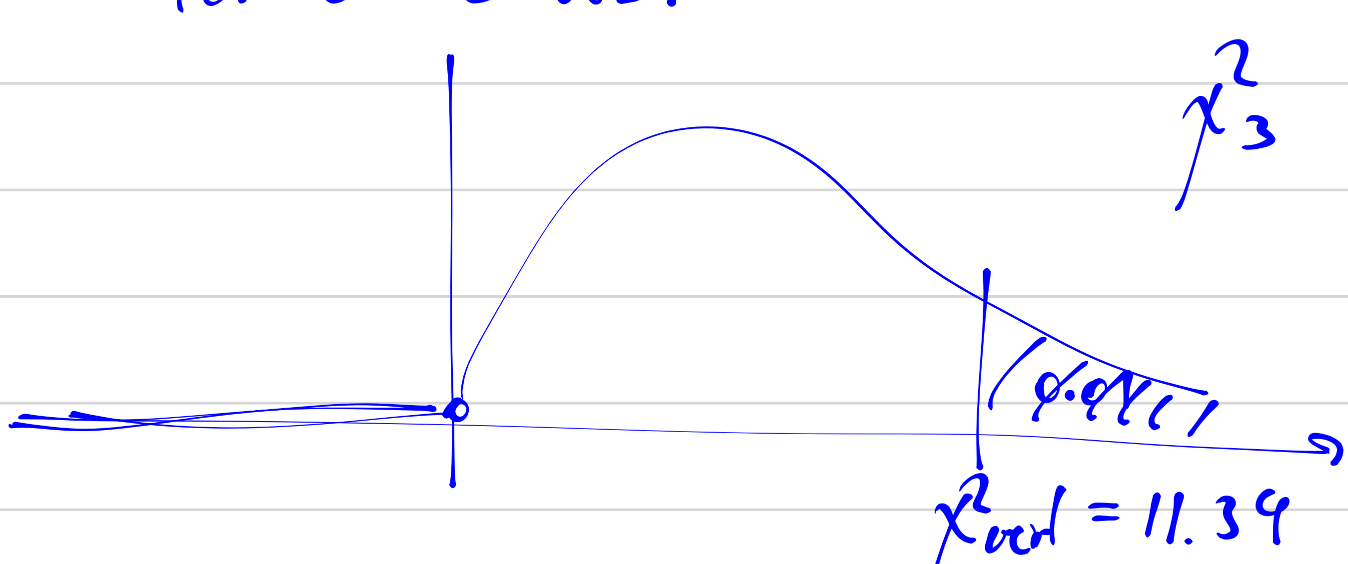
$\hat{p}_1^{UR} = 20/80$
$\hat{p}_2^{UR} = 30/80$
$\hat{p}_3^{UR} = 20/80$

$$\begin{cases} \dfrac{\partial \ln L}{\partial \hat{p}_1} = \dfrac{20}{\hat{p}_1} - \dfrac{10}{1-\hat{p}_1-\hat{p}_2-\hat{p}_3} = 0 \\ \dfrac{\partial \ln L}{\partial \hat{p}_2} = \ldots\ldots = 0 \\ \dfrac{\partial \ln L}{\partial \hat{p}_3} = \ldots\ldots = 0 \end{cases}$$

$$LR = 2\cdot\left[20\left(\ln\frac{20}{80} - \ln 0.25\right) + 30\left(\ln\frac{30}{80} - \ln 0.25\right) + \left(\ln\frac{20}{80} - \ln 0.25\right)20 + \left(\ln\frac{10}{80} - \ln 0.25\right)\cdot 10\right]$$

$$LR = 2\cdot\sum_{i=1}^{2}\left(\ln\frac{N_i}{n} - \ln p_i^{H_0}\right)\cdot N_i$$

$$= 10.46 = LR_{obs}.$$



Conclusion: do not reject H_0.

LR is more than this simple H_0 !!

R $\quad H_0: \; p_A = p_B = p_C$

UR $\quad H_A: \; p_A = p_B$

A	B	C	D
20	30	20	10

free params $p_{UR} \overset{?}{=} 2$

free params $p_R \overset{?}{=} 1$

$df = p_{UR} - p_R = 2 - 1 = 1$

$$LR = 2 \cdot \sum N_i \left(\ln \hat{p}_i^{UR} - \ln \hat{p}_i^{R} \right)$$

R model: $\max_{\hat{p}_A} \; 20 \ln \hat{p}_A + 30 \ln \hat{p}_A + 20 \ln \hat{p}_A + 10 \ln(1 - 3\hat{p}_A)$

allows no simple intuition

$$\cancel{\begin{array}{c|c} ABC & D \\ 3\hat{p}_A & 1-3\hat{p}_A \end{array}}$$

$$\ln L_R = 80 \ln \hat{p}_A + 10 \cdot \ln(1 - 3\hat{p}_A)$$

$$\frac{\partial \ln L_R}{\partial \hat{p}_A} = \frac{80}{\hat{p}_A} - \frac{10}{1 - 3\hat{p}_A} \cdot 3 = 0$$

$$80(1 - 3 \cdot \hat{p}_A) = 30 \hat{p}_A$$

$$80 = (80 \cdot 3 + 30) \hat{p}_A$$

$$\hat{p}_A^R = \frac{80}{80 \cdot 3 + 30} = \frac{8}{24 + 3} = \frac{8}{27}$$

UR-model: $\max_{\hat{p}_A, \hat{p}_C} \; 20 \ln \hat{p}_A + 30 \cdot \ln \hat{p}_A + 20 \ln \hat{p}_C + 10 \cdot \ln(1 - 2\hat{p}_A - \hat{p}_C)$

$\hat{p}_A^{UR} \quad \hat{p}_C^{UR}$

$$LR = 2 \cdot (\ln L_{UR} - \ln L_R) \quad \text{vs} \quad \boxed{\chi^2_{crit}, \; df = 1}$$