Roadmap

- Clicker Questions for the quiz
- Goodness of fit tests
- Tests for independence of categorical variables

Likelihood Ratio Test Statistic for Multinomial

Pearson's Chi-Squared.

Clicker 1

$Y_1, \ldots Y_n$ 'SLRM.

$Y_i \sim G(\alpha + \beta x_i, \sigma)$.

$i = 1, \ldots n$

$n = 30$

independent.

What does $\frac{28 \hat{S}_e^2}{\sigma^2}$ follow?

(a) Z

(b) T_{28}

(c) T_{29}

$\boxed{\text{(d) } \chi^2_{28}} \longrightarrow 64\%$

$\frac{(n-2) S_e^2}{\sigma^2} \sim \chi^2_{n-2}$.

Q2 LRTs:

$$Y_1, \dots Y_n \sim f(\theta)$$

n is large:

(i) MLE $\hat{\theta} = 5$

(ii) $H_0: \theta = 4 \rightarrow \theta_0$

$R(4) = 0.3$ → Value of the Relative Likelihood function

The value of $\lambda(\theta_0)$

GOODNESS OF FIT TESTS

Facts that are relevant

(i).

$$H_0 : \theta = 4$$

$$\hat{\theta} = 5$$

$$\lambda = -2 \log \frac{L(\theta_0)}{L(\hat{\theta})}$$

$$= -2 \log R(\theta_0)$$

$$= -2 \log \underline{R(4)}$$

$$= -2 \log (0.3)$$

$$r(\theta) = \log R(\theta)$$

$$r(4) = \log R(4)$$

LRTS

$$H_0 : \theta = \theta_0$$

$$\Delta(\theta_0) = -2 \log \frac{L(\theta_0)}{L(\hat{\theta})} \sim \chi^2_1$$

This formula is true when θ is a scalar.

$$\underset{\sim}{\theta} = (\theta_1, \ldots \theta_m)$$

The test gets modified.

$$\Delta(\theta_0) = -2 \log \frac{L(\theta_0)}{L(\hat{\theta})} \sim \chi^2_n$$

where n = # of unrestricted parameters of θ

− # of parameters that we need to estimate under H_0.

<u>Example</u>: Test whether a die is fair.

$$\theta = (\overbrace{\theta_1, \theta_2, \ldots \theta_6}^{5})$$

$$\theta_i = \text{Prob}(i).$$

$$H_0 : \theta_1 = \theta_2 = \ldots = \theta_6 = 1/6$$

$df = 5 - 0 = 5$

$\therefore$

Example 2

$X_1, \ldots X_n$

We suspect that the data is Poisson.

$$H_0 : X_i \sim Poi(\theta)$$

~~F0~~ Collect a sample

0	
1	
2	
3	
4	

0·1

SAMPLE

Outcome	Frequencies
0	10
1	15
2	20
3	10
4	5
≥ 5	5

Suppose the data was Poisson.

$\hat{\theta} = \bar{x}$

$$\frac{e^{-\bar{x}}\,\bar{x}^{0}}{0!}$$

0·1

~~With~~

We want to find the expected frequencies of each category ~~given~~ assuming the data is Poisson.

Expected frequency = Estimated probability of that category × sample size

Example 3

$X_1, \ldots X_n$

$H_0: X_i \sim G(\mu, \sigma)$

Categories	Observed
[0,10)	15
[10,20)	
[20,30)	
[30,40)	
[40,50)	
≥50	

$n = 6$

$k =$

Result.

$$\Lambda(\theta_0) = -2 \log \frac{L(\theta_0)}{L(\tilde{\theta})}$$

$$= 2 \sum Y_i \log \frac{Y_i}{E_i}$$

where Y_i = observed freque[...]
of category i

E_i = Expected
frequency of
category i under H[...]

$$\boxed{\Lambda(\theta_0) \sim \chi^2_{n-1-k} \ df}$$

n = # of categories

k = # of parameters one has to eshmate under H_0

To run the test, we have to follow the following steps

Step 1 Calculate $\lambda(\theta_0)$

$$\lambda(\theta_0) = \underbrace{\boxed{2 \sum y_i \log \frac{y_i}{e_i}}}_{\text{known \#}}$$

Step 2 Calculate the p-value.

$$\text{p-value} = P(\Lambda \geq \lambda)$$

$$\chi^2_{n-k-1}$$

y_i's e_i

	Freq	Expected
1	40	50
2	60	50
3	30	50
4	70	50
5	50	50
6	50	50

Die problem

$$df = n - 1 - k$$

$$6 - 1 - 0 = 5$$

$X_1, \ldots, X_n \sim Poi(\theta)$

	y_i	e_i	
0			
1			
2			
3			
4			
≥ 5			

$n = 6 \qquad k = 1$

$$\Delta \sim \chi^2_4$$

This is called the test of goodness of fit using the LRTS.

- This works
  - if the sample size is large
  - if all the observed frequencies in each category ≥ 5

This can be used to check for model assumptions objectively.

$$\boxed{\Delta = 2\sum Y_i \ln \frac{Y_i}{E_i}}$$

Pearson's Chi-Square

$$\boxed{P = \sum \frac{(Y_i - E_i)^2}{E_i}} \sim \chi^2_{n-k-1}$$

Y_i = observed

$r(\theta_0) = \log R(\theta_0)$

(a) $-2 \log \frac{L(4)}{L(5)}$

0.3

$\Downarrow$

(b) $-2 \log 0.3 = -2 \log \underbrace{R(\theta_0)}$

(c) $-2\, r(\theta_0)$

$r(4)$

r = log relative likelihood function

(d) All of the above.

80%

$r(\theta) = \log R(\theta)$

$r(4)$ -

—