NULL - HYPOTESIS SIGNIFICANCE TESTING

Mauro Díaz

Exercise 16.4 (Chi-squared test)

a. Show that the $\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$ test statistic is approximately χ^2 distributed when we have two categories.

A chi-squared distribution with k degrees of freedom, denoted χ^2_k is defined as

$$\sum_{j=1}^{k} z_j^2 \sim \chi^2_k \quad , \text{ where } \quad z_1, \dots, z_k \sim N(0,1) \text{ independent}$$

Now, the statement defines: $\quad \chi^2 = \sum_{i=1}^{2} \frac{(O_i - E_i)^2}{E_i}$, is $\chi^2 \xrightarrow{d} \chi^2_1$?

First of all:

- $O_i \equiv$ n° of observations actually observed in category i.
- $E_i \equiv$ n° of observations we would expect in category i if H_0 were true.

With 2 categories, we can reduce a binomial model:

$$O_1 \sim Bin(n, p) \qquad O_2 \sim n - O_1$$

$$E_1 = np \qquad E_2 = n(1-p)$$

So: $$\chi^2 = \frac{(O_1 - np)^2}{np} + \frac{(n - O_1 - n(1-p))^2}{n(1-p)} = (O_1 - np)^2 \left(\frac{1}{np} + \frac{1}{n(1-p)} \right) =$$

$$= \frac{(O_1 - np)^2}{np(1-p)} = \left(\frac{O_1 - np}{np(1-p)} \right)^2$$

As $O_1 \sim Bin(n,p) \Rightarrow E[O_1] = np$, $V(O_1) = np(1-p)$, using CLT:

$$\frac{O_1 - np}{np(1-p)} \xrightarrow{d} N(0,1)$$

and by the continuous mapping th: $\quad \left(\frac{O_1 - np}{np(1-p)} \right)^2 \xrightarrow{d} N(0,1)^2 \equiv \chi^2_1$