Exercise 12

Two Histograms

Given are two histograms with the same binning (r bins). The null hypothesis is that the two histograms represent random numbers from different distributions. However, it is suspected that both populations stem from the same distribution. This means there are r probabilities $p_1, \dots, p_r$ for an observation to lie in the i-th bin ($\sum_{i=1}^r p_i = 1$. The entries in the i-th bin of the first histogram are denoted n_i and in the second m_i. The number of observations in the first histogram is $N = \sum_{i=1}^r n_i$ and in the second $M = \sum_{i=1}^r m_i$.

> **(a) What distribution do the count rates in the individual bins follow? State the PDF for a single bin for both histograms (n_i and m_i) under the null hypothesis.**

The count rates in the individual bins follow a poisson-distribution:

first histogram: $\frac{\exp(-Np_i)(Np_i)^{n_i}}{n_i!}$

second histogram: $\frac{\exp(-Mp_i)(Mp_i)^{m_i}}{m_i!}$

> **(b) State the likelihood function for the null hypothesis. Find the estimator $\hat{p_i}$ that maximises the likelihood.**

$$L = \frac{\exp(-Np_i)(Np_i)^{n_i}}{n_i!} \cdot \frac{\exp(-Mp_i)(Mp_i)^{m_i}}{m_i!} = \frac{\exp(-p_i(N+M))(Np_i)^{n_i}(Mp_i)^{m_i}}{n_i! \cdot m_i!}$$

To find $\hat{p_i}$ that maximises the likelihood, we have to find the minimum of the log-likelihood:

$$l = \ln(L) = -p_i(N+M) + (n_i + m_i)\ln(p_i) + (n_i + m_i)\ln(N \cdot M) - \ln(n_i! \cdot m_i!)$$

$$\frac{\partial l}{\partial p_i} = -N - M + \frac{(n_i+m_i)}{\hat{p_i}} \stackrel{!}{=} 0$$

$$\Leftrightarrow \hat{p_i} = \frac{(n_i+m_i)}{N+M}$$

> **(c) State the χ^2 test statistic assuming the null hypothesis. (No simplification of the term is necessary)**

$$\chi^2 = \sum_{i=1}^r \left[\frac{(n_i - N\frac{n_i+m_i}{N+M})^2}{N \cdot \frac{n_i+m_i}{N+M}} + \frac{(m_i - M\frac{n_i+m_i}{N+M})^2}{M \cdot \frac{n_i+m_i}{N+M}} \right]$$

> **(d) How many degrees of freedom does the χ^2 distribution have? Does the test statistic for small bin contents ($n_i, m_i < 10$) still follow a χ^2 distribution? If not, why not?**

Through the histograms we have $2r$ degrees of freedom through the sum, $(r-1)$ p_i values given trough the estimator and M and N are given trough $M = \sum_{i=1}^r m_i$ and $N = \sum_{i=1}^r n_i$. So in total we have $r - 1$ degrees of freedom.

When the bin contents are too small, the uncertainties grow too big and the test statistic does not follow a χ^2 distribution anymore.

> **Given are the histograms: ...**

> **(e) It can be shown that the test statistic can be simplified to $\chi^2 = \frac{1}{NM}\sum_{i=1}^r \frac{(Nm_i - Mn_i)^2}{n_i+m_i}$. Check wether the null hypothesis for the given histograms $\alpha = 0.1, 0.05, 0.01$ is to be rejected. What does the Type II error describe in this case?**

$N == \sum_{i=1}^r n_i = 632$ and $M = \sum_{i=1}^r m_i = 81$

$$\hat{p_i} = \frac{(n_i+m_i)}{N+M}$$

$$\hat{p_1} = 0,177, \hat{p_2} = 0,314, \hat{p_3} = 0,509$$

$\chi^2 = \frac{1}{NM} \sum_{i=1}^{r} \frac{(Nm_i - Mn_i)^2}{n_i + m_i} \approx 8,43$

The χ^2 values for the different hisograms can be looked up in a table:

$\chi^2_{0,1} = 4,61$, $\chi^2_{0,05} = 5,99$, $\chi^2_{0,01} = 9,21$.

Therefore the null hyptothesis is accepted for $\alpha = 0,1$ and $= 0,05$ but is rejected for $\alpha = 0,01$.

A Type II error describes the error that occurs when one fails to reject a null hypothesis that is actually false.

In [ ]: