Pearson’s Chi-square Test

\[ (Y_1, \ldots, Y_K) \sim \text{Multinomial}(n, p_1, \ldots, p_K) \Rightarrow \sum_{k=0}^{K} \frac{(Y_k - np_k)^2}{np_k} \xrightarrow{L} \chi^2_{K-1} \]

Examples:

1. **Independent Representation of Multinomial Distribution** Let \( X_1, X_2, \ldots, X_n \) independently drawn from discrete distribution \( \mathbb{P}(X_i = k) = p_k, \ k = 1, \ldots, m \) and \( p_1 + p_2 + \cdots + p_m = 1 \). Define \( Y_k = \sum_{i=1}^{n} I(X_i = k), \ k = 1, \ldots, m \). Show \( (Y_1, Y_2, \ldots, Y_m) \sim \text{Multinom}(n, p_1, \ldots, p_m) \).

2. **Discrete Goodness of Fit** Test the independence of \( X \) and \( Y \) in the two-way contingency table

<table>
<thead>
<tr>
<th>Variable</th>
<th>Y=1</th>
<th>Y=0</th>
<th>Marginal of X</th>
</tr>
</thead>
<tbody>
<tr>
<td>X=1</td>
<td>16</td>
<td>3</td>
<td>19</td>
</tr>
<tr>
<td>X=0</td>
<td>398</td>
<td>225</td>
<td>623</td>
</tr>
<tr>
<td>Marginal of Y</td>
<td>414</td>
<td>228</td>
<td>642</td>
</tr>
</tbody>
</table>

Model the problem as \( (16, 398, 3, 225) \sim \text{Multinomial}(642, pq, (1-p)q, p(1-q), (1-p)(1-q)) \) under null.

(a) Compute the MLE for \( p \) and \( q \).

(b) Do Pearson’s goodness of test using \( \hat{p} \) and \( \hat{q} \).

(c) Derive another test based on the likelihood ratio test.

3. **Continuous Goodness of Fit** Given the independent samples \( X_1, \ldots, X_{100} \), we are interested in testing of the following hypothesis

\[ H_0 : X_1, \ldots, X_{100} \sim \text{Uniform}[0, \theta] \quad \text{versus} \quad H_1 : \text{Otherwise} \]

We observe 15 samples in \([0, 1] \), 20 samples in \((1, 2] \), 30 samples in \((2, 3] \), 25 samples in \((3, 10] \), 10 samples in \((10, +\infty) \) and \( X_{\text{max}} = 40 \). Test the goodness-of-fit.