Exercises, Section 9, Pearson’s Chi-Square.

1. A die was tossed 300 times and the uppermost face was recorded. The data are

<table>
<thead>
<tr>
<th>face</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>frequency</td>
<td>46</td>
<td>58</td>
<td>59</td>
<td>35</td>
<td>45</td>
<td>57</td>
</tr>
</tbody>
</table>

It is desired to test the hypothesis that the die is fair, \( H_0 : p_i = 1/6 \) for \( i = 1, \ldots, 6 \). Compute (a) Pearson’s \( \chi^2 \), (b) the Neyman \( \chi^2 \), (c) the Hellinger \( \chi^2 \), for testing \( H_0 \) with this data, and compare with the 5% cut-off point of the appropriate distribution.

2. Find the transformed \( \chi^2 \) where each cell is transformed by the reciprocal transformation. What is the modified transformed \( \chi^2 \) for this transformation?

3. (a) One measure of the homogeneity of a multinomial population with \( k \) cells and probabilities, \( p = (p_1, \ldots, p_k) \), is the sum of the squares of the probabilities, \( S(p) = \sum_1^k p_i^2 \). Note that \( 1/k \leq S(p) \leq 1 \), with higher values indicating greater heterogeneity. Given a sample of size \( n \) from this population (with replacement), we may estimate \( S(p) \) by \( S(\hat{p}) \), where \( \hat{p} = (\hat{p}_1, \ldots, \hat{p}_k) \) and \( \hat{p}_i \) is the proportion of the observations that fall in cell \( i \). What is the asymptotic distribution of \( S(\hat{p}) \)?

(b) Another measure of homogeneity often used is Shannon entropy, defined as \( H(p) = -\sum_1^k p_i \log p_i \), with \( 0 \leq H(p) \leq \log k \), and with higher values indicating greater homogeneity. What is the asymptotic distribution of \( H(\hat{p}) \)?

4. Consider a multinomial experiment with 4 cells, sample size \( n \), and vector of probabilities \( p = (p_1, p_2, p_3, p_4) \). Let \( n_i \) denote the number of observations falling in cell \( i \) for \( i = 1, \ldots, 4 \), where \( n_1 + n_2 + n_3 + n_4 = n \). Let \( X_n = n_1 + n_2 \) and \( Y_n = n_1 + n_3 \). Find the joint asymptotic distribution of \( X_n \) and \( Y_n \).

5. Modification of Pearson’s chi-square, \( \chi^2_M = (1/n) \sum_1^k (\hat{p}_i - p_i)^2 / P_i \), may be achieved by replacing the \( p_i \) in the denominator by any estimate, \( \hat{p}_i = f(p_i, \hat{p}_i) \), such \( \hat{p}_i \xrightarrow{P} p_i \) for all \( i \) as \( n \to \infty \). The resulting modified chisquare, \( \chi^2_M = (1/n) \sum_1^k (\hat{p}_i - p_i)^2 / f(p_i, \hat{p}_i) \), still has an asymptotic \( \chi^2_{n-1} \) distribution. Show that Hellinger’s \( \chi^2 \), in addition to being a transformed \( \chi^2 \), is also a modified \( \chi^2 \). In particular, find \( f(p_i, \hat{p}_i) \) such that \( f(p_i, \hat{p}_i) \xrightarrow{P} p_i \) and \( \chi^2_M = \chi^2_H \).

6. Let \( X \) and \( Y \) be 2-valued random variables taking on values 1 and 2, and let \( p_{ij} = P(X = i, Y = j) \) for \( i = 1, 2 \) and \( j = 1, 2 \), where \( \sum_i \sum_j p_{ij} = 1 \). The parameter \( \theta = \frac{p_{11}p_{22}}{p_{12}p_{21}} \) is called the odds-ratio and may be used as a measure of association between \( X \) and \( Y \). \( X \) and \( Y \) are independent if \( \theta = 1 \) (Show this), positively associated if \( \theta > 1 \), and negatively associated if \( \theta < 1 \).

Suppose a sample of size \( n \) is taken from the distribution of \( (X, Y) \), with \( n_{ij} \) observations falling in “cell” \( (i, j) \), where \( \sum_i \sum_j n_{ij} = n \).
(a) The sample estimate of $\theta$ is $\hat{\theta}_n = \frac{\hat{p}_{11}\hat{p}_{22}}{\hat{p}_{12}\hat{p}_{21}}$, where $\hat{p}_{ij} = n_{ij}/n$. What is the asymptotic distribution of $\hat{\theta}_n$ as $n \to \infty$?

(b) Let $\vartheta = \log(\theta)$ be the log odds-ratio. Find the asymptotic distribution of $\hat{\vartheta}_n = \log(\hat{\theta}_n)$ as an estimate of $\vartheta$. 