Topic 18. Sample correlation coefficient
Lehmann §5.4; Ferguson §8

Suppose that \((X_1, Y_1), (X_2, Y_2), \ldots\) are iid vectors with \(E X_i^t < \infty\) and \(E Y_i^t < \infty\). For the sake of simplicity, we will assume without loss of generality that \(E X_i = E Y_i = 0\) (alternatively, we could base all of the following derivations on the centered versions of the random variables).

We wish to find the asymptotic distribution of the sample correlation \(r = s_{xy}/(s_x s_y)\), where if we let

\[
\begin{pmatrix}
m_x \\
m_y \\
m_{xx} \\
m_{yy} \\
m_{xy}
\end{pmatrix} = \frac{1}{n} \begin{pmatrix}
\sum_{i=1}^{n} X_i \\
\sum_{i=1}^{n} Y_i \\
\sum_{i=1}^{n} X_i^2 \\
\sum_{i=1}^{n} Y_i^2 \\
\sum_{i=1}^{n} X_i Y_i
\end{pmatrix},
\]

then

\[
s_x^2 = m_{xx} - m_x^2, s_y^2 = m_{yy} - m_y^2, \text{ and } s_{xy} = m_{xy} - m_x m_y.
\]

Notice that we have suppressed the \(n\) in the notation above in order to keep things slightly simpler. According to the central limit theorem,

\[
\sqrt{n} \begin{pmatrix}
m_x \\
m_y \\
m_{xx} \\
m_{yy} \\
m_{xy}
\end{pmatrix} - \begin{pmatrix}
0 \\
0 \\
\sigma_x^2 \\
\sigma_y^2 \\
\sigma_{xy}
\end{pmatrix} \xrightarrow{\mathcal{D}} N_5 \left( \begin{pmatrix}
\text{Cov} (X_1, X_1) & \cdots & \text{Cov} (X_1, X_1 Y_1) \\
\text{Cov} (Y_1, X_1) & \cdots & \text{Cov} (Y_1, X_1 Y_1) \\
\vdots & \ddots & \vdots \\
\text{Cov} (X_1 Y_1, X_1) & \cdots & \text{Cov} (X_1 Y_1, X_1 Y_1)
\end{pmatrix} \right).
\]

Let \(\Sigma\) denote the covariance matrix in expression (37). Define a function \(g : R^5 \rightarrow R^3\) such that \(g\) applied to the vector of moments in equation (35) yields the vector \((s_x^2, s_y^2, s_{xy})\) as defined in expression (36). Then

\[
\hat{g} \begin{pmatrix}
a \\
b \\
c \\
d \\
e
\end{pmatrix} = \begin{pmatrix}
-2a & 0 & 1 & 0 & 0 \\
0 & -2b & 0 & 1 & 0 \\
-a & -b & 0 & 0 & 1
\end{pmatrix}.
\]

Therefore, if we let

\[
\Sigma^* = \hat{g} \begin{pmatrix}
0 & \sigma_x^2 \\
\sigma_y^2 & \sigma_{xy}
\end{pmatrix} \Sigma \begin{pmatrix}
0 & \sigma_x^2 \\
\sigma_y^2 & \sigma_{xy}
\end{pmatrix}^t = \begin{pmatrix}
\text{Cov} (X_1^2, X_1^2) & \text{Cov} (X_1^2, Y_1^2) & \text{Cov} (X_1^2, X_1 Y_1) \\
\text{Cov} (Y_1^2, X_1^2) & \text{Cov} (Y_1^2, Y_1^2) & \text{Cov} (Y_1^2, X_1 Y_1) \\
\text{Cov} (X_1 Y_1, X_1^2) & \text{Cov} (X_1 Y_1, Y_1^2) & \text{Cov} (X_1 Y_1, X_1 Y_1)
\end{pmatrix},
\]

then by the delta method,

\[
\sqrt{n} \begin{pmatrix}
s_x^2 \\
s_y^2 \\
s_{xy}
\end{pmatrix} - \begin{pmatrix}
\sigma_x^2 \\
\sigma_y^2 \\
\sigma_{xy}
\end{pmatrix} \xrightarrow{\mathcal{D}} N_3 (0, \Sigma^*).
\]
Finally, define the function \( h(a, b, c) = c/\sqrt{ab} \), so that we have \( h(s_x^2, s_y^2, s_{xy}) = r \). Then \( \dot{h}(a, b, c) = \frac{1}{2}(-c/\sqrt{a^3b}, -c/\sqrt{ab^3}, 2/\sqrt{ab}) \), so that

\[
\dot{h}(a, b, c) = \left( \frac{-\sigma_{xy}}{2\sigma_x^3 \sigma_y}, \frac{-\sigma_{xy}}{2\sigma_x \sigma_y^3}, \frac{2}{\sigma_x \sigma_y} \right).
\]

Therefore, if \( A \) denotes the \( 1 \times 3 \) matrix in equation (38), using the delta method once again yields

\[
\sqrt{n}(r - \rho) \xrightarrow{L} \mathcal{N}(0, A\Sigma^* A^t).
\]

Consider the special case of bivariate normal \((X_i, Y_i)\). In this case, we may derive

\[
\Sigma^* = \begin{pmatrix}
2\sigma_x^4 & 2\rho \sigma_x^2 \sigma_y^2 & 2\rho \sigma_x^2 \sigma_y^2 \\
2\rho \sigma_x^2 \sigma_y^2 & 2\sigma_y^4 & 2\rho \sigma_x \sigma_y^3 \\
2\rho \sigma_x \sigma_y^3 & 2\rho \sigma_x \sigma_y^3 & (1 + \rho^2) \sigma_y^2
\end{pmatrix}.
\]

In this case, \( A\Sigma^* A^t = (1 - \rho^2)^2 \), which implies that

\[
\sqrt{n}(r - \rho) \xrightarrow{L} \mathcal{N}\{0, (1 - \rho^2)^2\}.
\]

In the normal case, we may derive a variance-stabilizing transformation. According to equation (40), we should find a function \( f(x) \) satisfying \( f'(x) = (1 - x^2)^{-1} \). Since

\[
\frac{1}{1 - x^2} = \frac{1}{2(1 - x)} + \frac{1}{2(1 + x)},
\]

which is easy to integrate, we obtain

\[
f(x) = \frac{1}{2} \log \frac{1 + x}{1 - x}.
\]

This is called Fisher’s transformation; we conclude that

\[
\sqrt{n} \left( \frac{1}{2} \log \frac{1 + r}{1 - r} - \frac{1}{2} \log \frac{1 + \rho}{1 - \rho} \right) \xrightarrow{L} \mathcal{N}(0, 1).
\]

### Problems for Topic 18

**Problem 18.1** Verify expressions (39) and (40).

**Problem 18.2** Assume \((X_1, Y_1), \ldots, (X_n, Y_n)\) are iid from some bivariate normal distribution. Let \( \rho \) denote the population correlation coefficient and \( r \) the sample correlation coefficient.

(a) Describe a test of \( H_0 : \rho = 0 \) against \( H_1 : \rho \neq 0 \) based on the fact that

\[
\sqrt{n}[f(r) - f(\rho)] \xrightarrow{L} \mathcal{N}(0, 1),
\]

where \( f(x) \) is Fisher’s transformation \( f(x) = (1/2) \log[(1 + x)/(1 - x)] \). Use \( \alpha = .05 \).

(b) Based on 5000 repetitions each, estimate the actual level for this test in the case when \( E(X_i) = E(Y_i) = 0, \operatorname{Var}(X_i) = \operatorname{Var}(Y_i) = 1, \) and \( n \in \{3, 5, 10, 20\} \).
The first step in this topic is to review some facts regarding change of variables. Suppose \( X \) has density \( f_X(x) \) and \( Y = g(X) \), where \( g : R^k \rightarrow R^k \) is differentiable and has a well-defined inverse. We are interested in determining the density of \( Y \).

Note that \( \hat{g}(x) \) is a \( k \times k \) matrix, namely, the matrix of partial derivatives \( \frac{\partial g_i(x)}{\partial x_j} \). This square matrix is sometimes called the Jacobian matrix, or simply the Jacobian. In other contexts, however, “Jacobian” refers to the determinant of the Jacobian matrix. We will use the term “Jacobian” to refer to this Jacobian determinant in what follows; the matrix \( \hat{g} \) will be called the “Jacobian matrix”. Note that the Jacobian of \( g(x) \) is a real-valued function of \( x \). We denote this function \( J_g : R^k \rightarrow R \). Thus, \( J_g(x) = \text{Det} \{ \hat{g}(x) \} \).

Suppose now that \( h(y) \) denotes the inverse of \( g(x) \), so that \( h \circ g(x) = x \) for all \( x \) and \( g \circ h(y) = y \) for all \( y \). Then we may write the density function for \( Y = g(X) \) in terms of the density function for \( X \) as follows:

\[
f_Y(y) = |J_h(y)| f_X \circ h(y).
\] (41)

Equation (41) is often summarized in words as something like “the new density is the old density times the Jacobian”; it is important to note, however, that the Jacobian referred to is the Jacobian of the inverse transformation, not the transformation itself.

**Example 19.1** Suppose \( X_1, \ldots, X_{n+1} \) are iid standard exponential random variables. For \( j = 1, \ldots, n \), define

\[
Y_j = \frac{\sum_{i=1}^{j} X_i}{\sum_{i=1}^{n+1} X_i}.
\]

We derive the joint density of \((Y_1, \ldots, Y_n)\) as follows. First, we observe that the joint density of \((X_1, \ldots, X_{n+1})\) is

\[
f_X(x) = \exp \left\{ -\sum_{i=1}^{n+1} x_i \right\} I\{x_1 > 0, \ldots, x_{n+1} > 0\}.
\]

As an intermediate step, define \( Z_j = \sum_{i=1}^{j} X_i \) for \( j = 1, \ldots, n+1 \). Then the \( X_i \) may be expressed in terms of the \( Z_i \) as

\[
X_i = \begin{cases} 
Z_i & \text{if } i = 1 \\
Z_i - Z_{i-1} & \text{if } i > 1.
\end{cases}
\]

The Jacobian is clearly 1 for this transformation, since the Jacobian matrix is lower triangular with ones on the diagonal. Therefore, we obtain

\[
F_Z(z) = \exp\{-z_{n+1}\} I\{0 < z_1 < z_2 < \cdots < z_{n+1}\}
\]

as the density of \( Z \).
If we define \( Y_{n+1} = Z_{n+1} \), then we may express the \( Z_i \) in terms of the \( Y_i \) as
\[
Z_i = \begin{cases} 
Y_{n+1}Y_i & \text{if } i < n+1 \\
Y_{n+1} & \text{if } i = n+1.
\end{cases}
\] (42)

The Jacobian matrix of the transformation in equation (42) is upper triangular, with \( y_{n+1} \) along the diagonal except for a 1 in the lower right corner. Thus, the Jacobian equals \( y_{n+1}^n \), so the density of \( Y \) is
\[
f_Y(y) = y_{n+1}^n \exp\{-y_{n+1}\}I\{y_{n+1} > 0\}I\{0 < y_1 < \cdots < y_n < 1\}.
\]

Thus, \( (Y_1, \ldots, Y_n) \) is independent of \( Y_{n+1} \) and the density of \( (Y_1, \ldots, Y_n) \) is proportional to \( I\{0 < y_1 < \cdots < y_n < 1\} \). Note that \( Y_{n+1} \) may be seen to have a Gamma\((n+1, 1)\) density, which is also evident when we consider that \( Y_{n+1} \) is the sum of \( n+1 \) iid standard exponential random variables.

Since the joint density of the order statistics from a sample of size \( n \) from Uniform\((0,1)\) is \( n!I\{0 < u_1 < \cdots < u_n < 1\} \), Example 19.1 proves the following lemma.

**Lemma 19.1** The joint distribution of the order statistics from a sample of size \( n \) from Uniform\((0,1)\) is the same as the joint distribution of
\[
\left( \frac{X_1}{\sum_{i=1}^{n+1} X_i}, \frac{X_1 + X_2}{\sum_{i=1}^{n+1} X_i}, \ldots, \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n+1} X_i} \right),
\] (43)
where \( X_1, \ldots, X_{n+1} \) are iid standard exponential random variables. Furthermore, the joint distribution of expression (43) is independent of \( \sum_{i=1}^{n+1} X_i \).

We now use Lemma 19.1 to achieve a powerful result, namely deriving the joint asymptotic distribution of a set of sample quantiles. Recall that we have already derived the asymptotic distribution of the sample median, back in Example 12.4; here, we generalize that result using an entirely different method than the method of Example 12.4.

Take \( 0 < p_1 < p_2 < 1 \) and define \( a_n = \lfloor .5 + np_1 \rfloor \) and \( b_n = \lfloor .5 + np_2 \rfloor \), where \( \lfloor x \rfloor \) denotes the greatest integer less than or equal to \( x \) (in particular, note that \( .5 + x \) is simply \( x \) rounded to the nearest integer). Then for an iid sample \( U_1, \ldots, U_n \), the \( a_n \)th and \( b_n \)th order statistics \( U(a_n) \) and \( U(b_n) \) may be taken to be the \( p_1 \) and \( p_2 \) sample quantiles, respectively.

Suppose that the \( U_i \) are Uniform\((0,1)\). Then by Lemma 19.1, \( (U(a_n), U(b_n)) \) has the same distribution as
\[
\left( \frac{\sum_{i=1}^{a_n} X_i}{\sum_{i=1}^{n+1} X_i}, \frac{\sum_{i=1}^{b_n} X_i}{\sum_{i=1}^{n+1} X_i} \right),
\]
where \( X_1, \ldots, X_{n+1} \) are iid standard exponential random variables. Let \( A = \sum_{i=1}^{a_n} X_i \), \( B = \sum_{i=a_n+1}^{b_n} X_i \), and \( C = \sum_{i=b_n+1}^{n+1} X_i \). Then the joint asymptotic distribution of \( (U(a_n), U(b_n)) \) is the same as that of
\[
g(A, B, C) \overset{\text{def}}{=} \left( \frac{A}{A + B + C}, \frac{B}{A + B + C} \right).
\] (44)

This joint asymptotic distribution may be easily determined using the delta method if we can determine the joint asymptotic distribution of \( (A, B, C) \). But this is easy, since \( A, B, \) and \( C \) are by definition sums of iid random variables and they are independent of one another. Consider, for example, the fact that a bit of algebra gives
\[
\sqrt{n} \left( \frac{A}{n} - p_1 \right) = \sqrt{\frac{a_n}{n}} \sqrt{\frac{a_n}{n}} \left( \frac{A}{a_n} - \frac{np_1}{a_n} \right) = \sqrt{\frac{a_n}{n}} \sqrt{\frac{a_n}{n}} \left( \frac{A}{a_n} - 1 \right) + \sqrt{n} \left( 1 - \frac{np_1}{a_n} \right).
\]
By the central limit theorem, \( \sqrt{n}(A/a_n) \xrightarrow{d} N(0,1) \) because a standard exponential variable has mean 1 and variance 1. Furthermore, \( a_n \) was defined so that
\[
\sqrt{n} \left( 1 - \frac{np_1}{a_n} \right) \to 0,
\]
which of course also implies that \( a_n/n \to p_1 \). Therefore, Slutsky’s theorem gives
\[
\sqrt{n} \left( \frac{A}{n} - p_1 \right) \xrightarrow{d} N(0,p_1).
\]
Similar arguments applied to \( B \) and to \( C \), along with the fact that \( A, B, \) and \( C \) are independent, gives
\[
\sqrt{n} \left\{ \begin{pmatrix} A/n \\ B/n \\ C/n \end{pmatrix} - \begin{pmatrix} p_1 \\ p_2 - p_1 \\ 1 - p_2 \end{pmatrix} \right\} \xrightarrow{d} N_3 \left\{ 0, \begin{pmatrix} p_1 & 0 & 0 \\ 0 & p_2 - p_1 & 0 \\ 0 & 0 & 1 - p_2 \end{pmatrix} \right\}
\]
by Theorem 15.3.

Now recall the definition of \( g(A,B,C) \) in equation (44). For this \( g : R^3 \to R^2 \), we obtain
\[
g(a, b, c) = \begin{pmatrix} \frac{b+c}{(a^2+b^2+c^2)} & -\frac{a}{(a+b+c)^2} & -\frac{a}{(a+b+c)^2} \\ \frac{c}{(a+b+c)^2} & \frac{b}{(a+b+c)^2} & \frac{a}{(a+b+c)^2} \end{pmatrix}.
\]

Therefore,
\[
g(p_1, p_2 - p_1, 1 - p_2) = \begin{pmatrix} 1-p_1 & -p_1 & -p_1 \\ 1-p_2 & 1-p_2 & 1-p_2 \end{pmatrix}
\]
so the delta method gives
\[
\sqrt{n} \left\{ \begin{pmatrix} U_{(a_n)} \\ U_{(b_n)} \end{pmatrix} - \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} \right\} \xrightarrow{d} N_2 \left\{ 0, \begin{pmatrix} p_1(1-p_1) & p_1(1-p_2) \\ p_1(1-p_2) & p_2(1-p_2) \end{pmatrix} \right\}. \tag{45}
\]

The method used above to derive the joint distribution (45) of two sample quantiles may be easily extended to any number of quantiles. We may thus state the following theorem that relies on a generalization of the above argument:

**Theorem 19.1** Suppose that for given constants \( p_1, \ldots, p_k \) with \( 0 < p_1 < \cdots < p_k < 1 \), there exist sequences \( \{a_{1n}\}, \ldots, \{a_{kn}\} \) such that for all \( 1 \leq i \leq k \),
\[
\sqrt{n} \left( 1 - \frac{np_i}{a_{in}} \right) \to 0. \tag{46}
\]

Then if \( U_1, \ldots, U_n \) is a sample from Uniform(0,1),
\[
\sqrt{n} \left\{ \begin{pmatrix} U_{(a_{1n})} \\ \vdots \\ U_{(a_{kn})} \end{pmatrix} - \begin{pmatrix} p_1 \\ \vdots \\ p_k \end{pmatrix} \right\} \xrightarrow{d} N_k \left\{ 0, \begin{pmatrix} p_1(1-p_1) & \cdots & p_1(1-p_k) \\ \vdots & \ddots & \vdots \\ p_k(1-p_k) & \cdots & p_k(1-p_k) \end{pmatrix} \right\}.
\]

Note that in the covariance matrix in the above theorem, the \((i, j)\) entry is either \( p_i(1-p_j) \) or \( p_j(1-p_i) \), depending on whether \( i \leq j \) or \( j \leq i \).

As a corollary, we may restate Theorem 5.4.5 on page 314 of Lehmann:
Corollary 19.1 Suppose that there exists a cdf $F$ and points $\xi_1 < \cdots < \xi_k$ such that $F'(\xi_i)$ exists and is positive for all $i$. Let $p_i = F(\xi_i)$ for $1 \leq i \leq k$. Then if equation (46) is satisfied for sequences $\{a_{1n}\}, \ldots, \{a_{kn}\}$ and $X_1, \ldots, X_n$ is an iid sample from $F$, 

\[
\sqrt{n} \left\{ \begin{pmatrix} X_{(a_{1n})} \\ \vdots \\ X_{(a_{kn})} \end{pmatrix} - \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_k \end{pmatrix} \right\} \xrightarrow{d} \mathcal{N}_k \left\{ 0, \begin{pmatrix} \frac{p_1(1-p_1)}{F(\xi_1)^2} & \cdots & \frac{p_1(1-p_k)}{F(\xi_1)^2} \\ \vdots & \ddots & \vdots \\ \frac{p_k(1-p_k)}{F(\xi_k)^2} & \cdots & \frac{p_k(1-p_k)}{F(\xi_k)^2} \end{pmatrix} \right\}.
\]

The corollary is proved by a simple application of the delta method, since the hypotheses imply that $F(t)$ is continuous and strictly increasing in a neighborhood of each $\xi_i$. Thus, $F^{-1}(u)$ is well-defined in a neighborhood of each $p_i$. Defining $g(u_1, \ldots, u_k) = (F^{-1}(u_1), \ldots, F^{-1}(u_k))$, the matrix $g(u_1, \ldots, u_k)$ is a diagonal matrix with $i$th element 

\[
\frac{\partial F^{-1}(u_i)}{\partial u_i} = \frac{1}{F' \circ F^{-1}(u_i)}.
\]

(Incidentally, expression (47) may be derived by differentiating both sides of the equation $F \circ F^{-1}(u_i) = u_i$.)

Problems for Topic 19

Problem 19.1 Solve Problem 3.16 on page 357 of Lehmann.

Problem 19.2 Let $X_1$ be Uniform(−$\frac{\pi}{2}$, $\frac{\pi}{2}$) and let $X_2$ be standard exponential, independent of $X_1$. Find the joint distribution of $(Y_1, Y_2) = (\sqrt{2X_2} \cos X_1, \sqrt{2X_2} \sin X_1)$.

Problem 19.3 Suppose $X$ and $Y$ are iid standard normal random variables. What is the distribution of $X/Y$?

Problem 19.4 Suppose $X_1, \ldots, X_n$ is an iid sample with $P(X_1 \leq x) = F(x - \theta)$, where $F(x)$ is symmetric about zero. We wish to estimate $\theta$ by $(Q_p + Q_{1-p})/2$, where $Q_p$ and $Q_{1-p}$ are the $p$ and $1 - p$ sample quantiles, respectively. Find the smallest possible asymptotic variance for the estimator and the $p$ for which it is achieved for each of the following forms of $F(x)$:

(a) Standard Cauchy

(b) Standard normal

(c) Standard double exponential

Hint: If you cannot solve a problem analytically, try attacking it numerically. Part (c) is a bit of a trick question.

Problem 19.5 When we use a boxplot to assess the symmetry of a distribution, one of the main things we do is visually compare the lengths of $Q_3 - Q_2$ and $Q_2 - Q_1$, where $Q_i$ denotes the $i$th sample quartile.

(a) If we have a random sample of size $n$ from $N(0, 1)$, find the asymptotic distribution of $(Q_3^{(n)} - Q_2^{(n)}) - (Q_2^{(n)} - Q_1^{(n)})$.

(b) Repeat part (a) if the sample comes from a standard logistic distribution.

(c) Using 1000 simulations from each distribution, use graphs to assess the accuracy of each of the asymptotic approximations above for $n = 5$ and $n = 13$. (For a sample of size $4n + 1$, define $Q_i$ to be the $in + 1$ order statistic.) For each value of $n$ and each distribution, plot the empirical distribution function against the theoretical limiting cdf.
Hint: In parts (a) and (b), use Theorem 5.4.5 on p. 314. If you are using R for part (c), the function `cdf.compare` is a very useful function for comparing an empirical cdf with another cdf. For example, to compare the empirical cdf of a vector `x` with a standard normal distribution function, type `cdf.compare(x, dist="normal")`

Problem 19.6 Let $X_1, \ldots, X_n$ be a random sample from Uniform$(0, 2\theta)$. Find the asymptotic distributions of the median, the midquartile range, and $\frac{2}{3}X_{\lfloor 3n/4 \rfloor}$. (The midquartile range is the mean of the 1st and 3rd quartiles.) Compare these three estimates of $\theta$.

---

**Topic 20. Pearson’s chi-square statistic**

_Lehmann §5.5; Ferguson §9_

Let $X_1^{(1)}, X_2^{(2)}, \ldots$ be iid from a multinomial$(1, p)$ distribution, where $p$ is a $k$-vector with nonnegative entries that sum to one. That is,

$$P(X_j^{(i)} = 1) = 1 - P(X_j^{(i)} = 0) = p_j \quad \text{for all } 1 \leq j \leq k$$

and each $X_j^{(i)}$ consists of exactly $k-1$ zeros and a single one, where the one is in the component of the “success” category at trial $i$.

Since equation (48) implies that $\text{Var} X_j^{(i)} = p_j(1-p_j)$, and since $\text{Cov}(X_j^{(i)}, X_\ell^{(i)}) = -p_j p_\ell$ for $j \neq \ell$, the random vector $X^{(i)}$ has covariance matrix

$$\Sigma = \begin{pmatrix} p_1(1-p_1) & -p_1 p_2 & \cdots & -p_1 p_k \\ -p_1 p_2 & p_2(1-p_2) & \cdots & -p_2 p_k \\ \vdots & \ddots & \vdots & \vdots \\ -p_1 p_k & -p_2 p_k & \cdots & p_k(1-p_k) \end{pmatrix}. \quad (49)$$

Since $E X^{(i)} = p$, the central limit theorem implies

$$\sqrt{n} (\bar{X} - p) \xrightarrow{\mathcal{L}} N_k(0, \Sigma).$$

(50)

Note that the sum of the $j$th column of $\Sigma$ is $p_j - p_j(p_1 + \cdots + p_k) = 0$, which is to say that the sum of the rows of $\Sigma$ is the zero vector, so $\Sigma$ is not invertible.

We wish to derive the asymptotic distribution of Pearson’s chi-square statistic

$$\chi^2 = \sum_{j=1}^k \frac{(n_j - np_j)^2}{np_j},$$

where $n_j$ is the random variable $nX_j$, the number of successes in the $j$th category for trials $1, \ldots, n$. We will discuss two different ways to do this. One way avoids dealing with the singular matrix $\Sigma$, whereas the other does not.