Chapter 9

Pearson's chi-square test

9.1 Null hypothesis asymptotics

Let $\mathbf{X}_1, \mathbf{X}_2, \cdots$ be independent from a multinomial$(1, \mathbf{p})$ distribution, where $\mathbf{p}$ is a k-vector with nonnegative entries that sum to one. That is,

$$P(X_{ij} = 1) = 1 - P(X_{ij} = 0) = p_j \quad \text{for all } 1 \le j \le k \tag{9.1}$$

and each $\mathbf{X}_i$ consists of exactly $k-1$ zeros and a single one, where the one is in the component of the "success" category at trial i. Note that the multinomial distribution is a generalization of the binomial distribution to the case in which there are k categories of outcome instead of only 2.

The purpose of this section is to derive the asymptotic distribution of the Pearson chi-square statistic

$$\chi^2 = \sum_{j=1}^{k} \frac{(n_j - np_j)^2}{np_j}, \tag{9.2}$$

where n_j is the random variable $n\overline{X}_j$, the number of successes in the jth category for trials $1, \ldots, n$. In a real application, the true value of $\mathbf{p}$ is not known, but instead we assume that $\mathbf{p} = \mathbf{p}^0$ for some null value $\mathbf{p}^0$. We will show that χ^2 converges in distribution to the chi-square distribution on $k-1$ degrees of freedom, which yields to the familiar chi-square test of goodness of fit for a multinomial distribution.

Equation (9.1) implies that Var $X_{ij} = p_j(1-p_j)$. Furthermore, Cov $(X_{ij}, X_{i\ell})$ = E $X_{ij}X_{i\ell} -$

$p_j p_\ell = -p_j p_\ell$ for $j \neq \ell$. Therefore, the random vector $\mathbf{X}_i$ has covariance matrix

$$\Sigma = \begin{pmatrix} p_1(1-p_1) & -p_1p_2 & \cdots & -p_1p_k \\ -p_1p_2 & p_2(1-p_2) & \cdots & -p_2p_k \\ \vdots & & \ddots & \vdots \\ -p_1p_k & -p_2p_k & \cdots & p_k(1-p_k) \end{pmatrix}. \tag{9.3}$$

Since E $\mathbf{X}_i = \mathbf{p}$, the central limit theorem implies

$$\sqrt{n}\left(\overline{\mathbf{X}}_n - \mathbf{p}\right) \xrightarrow{d} N_k(\mathbf{0}, \Sigma). \tag{9.4}$$

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

We now present two distinct derivations of this asymptotic distribution of the χ^2 statistic in equation (9.2), because each derivation is instructive. One derivation avoids dealing with the singular matrix Σ, whereas the other does not.

In the first approach, define for each i $\mathbf{Y}_i = (X_{i1}, \ldots, X_{i,k-1})$. That is, let $\mathbf{Y}_i$ be the $k-1$-vector consisting of the first $k-1$ components of $\mathbf{X}_i$. Then the covariance matrix of $\mathbf{Y}_i$ is the upper-left $(k-1) \times (k-1)$ submatrix of Σ, which we denote by Σ^*. Similarly, let $\mathbf{p}^*$ denote the vector $(p_1, \ldots, p_{k-1})$.

One may verify that Σ^* is invertible and that

$$(\Sigma^*)^{-1} = \begin{pmatrix} \frac{1}{p_1} + \frac{1}{p_k} & \frac{1}{p_k} & \cdots & \frac{1}{p_k} \\ \frac{1}{p_k} & \frac{1}{p_2} + \frac{1}{p_k} & \cdots & \frac{1}{p_k} \\ \vdots & & \ddots & \vdots \\ \frac{1}{p_k} & \frac{1}{p_k} & \cdots & \frac{1}{p_{k-1}} + \frac{1}{p_k} \end{pmatrix}. \tag{9.5}$$

Furthermore, the χ^2 statistic of equation (9.2) by be rewritten as

$$\chi^2 = n(\overline{\mathbf{Y}} - \mathbf{p}^*)^\top (\Sigma^*)^{-1} (\overline{\mathbf{Y}} - \mathbf{p}^*). \tag{9.6}$$

The facts in Equations (9.5) and (9.6) are checked in Problem 9.2. If we now define

$$\mathbf{Z}_n = \sqrt{n}(\Sigma^*)^{-1/2}(\overline{\mathbf{Y}} - \mathbf{p}^*),$$

then the central limit theorem implies $\mathbf{Z}_n \xrightarrow{d} N_{k-1}(\mathbf{0}, I)$. By definition, the χ^2_{k-1} distribution is the distribution of the sum of the squares of $k-1$ independent standard normal random variables. Therefore,

$$\chi^2 = (\mathbf{Z}_n)^\top \mathbf{Z}_n \xrightarrow{d} \chi^2_{k-1}, \tag{9.7}$$

which is the result that leads to the familiar chi-square test.

In a second approach to deriving the limiting distribution (9.7), we use some properties of projection matrices.

Definition 9.1 A symmetric matrix P is called a projection matrix if it is idempotent; that is, if $P^2 = P$.

The following lemmas, to be proven in Problem 9.3, give some basic facts about projection matrices.

Lemma 9.2 Suppose P is a projection matrix. Then every eigenvalue of P equals 0 or 1. Suppose that r denotes the number of eigenvalues of P equal to 1. Then if $\mathbf{Z} \sim N_k(\mathbf{0}, P)$, $\mathbf{Z}^\top \mathbf{Z} \sim \chi^2_r$.

Lemma 9.3 The trace of a square matrix M, $\mathrm{Tr}\,(M)$, is equal to the sum of its diagonal entries. For matrices A and B whose sizes allow them to be multiplied in either order, $\mathrm{Tr}\,(AB) = \mathrm{Tr}\,(BA)$.

Recall (Lemma 4.8) that if a square matrix M is symmetric, then there exists an orthogonal matrix Q such that $QMQ^\top$ is a diagonal matrix whose entries consist of the eigenvalues of M. By Lemma 9.3, $\mathrm{Tr}\,(QMQ^\top) = \mathrm{Tr}\,(Q^\top QM) = \mathrm{Tr}\,(M)$, which proves yet another lemma:

Lemma 9.4 If M is symmetric, then $\mathrm{Tr}\,(M)$ equals the sum of the eigenvalues of M.

Define $\Gamma = \mathrm{diag}\,(\mathbf{p})$, and let Σ be defined as in Equation (9.3). Equation (9.4) implies

$$\sqrt{n}\Gamma^{-1/2}(\overline{\mathbf{X}} - \mathbf{p}) \xrightarrow{d} N_k(\mathbf{0}, \Gamma^{-1/2}\Sigma\Gamma^{-1/2}).$$

Since Σ may be written in the form $\Gamma - \mathbf{p}\mathbf{p}^\top$,

$$\Gamma^{-1/2}\Sigma\Gamma^{-1/2} \;=\; I - \Gamma^{-1/2}\mathbf{p}\mathbf{p}^\top\Gamma^{-1/2} \;=\; I - \sqrt{\mathbf{p}}\sqrt{\mathbf{p}}^\top \tag{9.8}$$

has trace $k-1$; furthermore,

$$(I - \sqrt{\mathbf{p}}\sqrt{\mathbf{p}}^\top)(I - \sqrt{\mathbf{p}}\sqrt{\mathbf{p}}^\top) \;=\; I - 2\sqrt{\mathbf{p}}\sqrt{\mathbf{p}}^\top + \sqrt{\mathbf{p}}\sqrt{\mathbf{p}}^\top\sqrt{\mathbf{p}}\sqrt{\mathbf{p}}^\top \;=\; I - \sqrt{\mathbf{p}}\sqrt{\mathbf{p}}^\top$$

because $\sqrt{\mathbf{p}}^\top\sqrt{\mathbf{p}} = 1$, so the covariance matrix (9.8) is a projection matrix.

Define $\mathbf{A}_n = \sqrt{n}\Gamma^{-1/2}(\overline{\mathbf{X}} - \mathbf{p})$. Then we may check (in problem 9.3) that

$$\chi^2 = (\mathbf{A}_n)^\top \mathbf{A}_n. \tag{9.9}$$

Therefore, since the covariance matrix (9.8) is a projection with trace $k-1$, Lemma 9.4 and Lemma 9.2 prove that $\chi^2 \xrightarrow{d} \chi^2_{k-1}$ as desired.

Exercises for Section 9.1

Exercise 9.1 *Hotelling's T^2.* Suppose $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots$ are independent and identically distributed from some k-dimensional distribution with mean $\boldsymbol{\mu}$ and finite nonsingular covariance matrix Σ. Let S_n denote the sample covariance matrix

$$S_n \;=\; \frac{1}{n-1}\sum_{j=1}^{n}(\mathbf{X}^{(j)} - \overline{\mathbf{X}})(\mathbf{X}^{(j)} - \overline{\mathbf{X}})^\top.$$

To test $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}^0$ against $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}^0$, define the statistic

$$T^2 \;=\; (\mathbf{V}^{(n)})^\top S_n^{-1}(\mathbf{V}^{(n)}),$$

where $\mathbf{V}^{(n)} = \sqrt{n}(\overline{\mathbf{X}} - \boldsymbol{\mu}^0)$. This is called Hotelling's T^2 statistic.

[Notes: This is a generalization of the square of a unidimensional t statistic. If the sample is multivariate normal, then $[(n-k)/(nk-k)]T^2$ is distributed as $F_{k,n-k}$. A Pearson chi square statistic may be shown to be a special case of Hotelling's T^2.]

(a) You may assume that $S_n^{-1} \xrightarrow{P} \Sigma^{-1}$ (this follows from the weak law of large numbers since $P(S_n \text{ is nonsingular}) \to 1$). Prove that under the null hypothesis, $T^2 \xrightarrow{d} \chi^2_k$.

(b) Let $\{\boldsymbol{\mu}^{(n)}\}$ be alternatives such that $\sqrt{n}(\boldsymbol{\mu}^{(n)} - \boldsymbol{\mu}^0) \to \boldsymbol{\delta}$. You may assume that under $\{\boldsymbol{\mu}^{(n)}\}$,

$$\sqrt{n}(\overline{\mathbf{X}} - \boldsymbol{\mu}^{(n)}) \;\xrightarrow{d}\; N_k(\mathbf{0}, \Sigma).$$

Find (with proof) the limit of the power against the alternatives $\{\boldsymbol{\mu}^{(n)}\}$ of the test that rejects H_0 when $T^2 \geq c_\alpha$, where $P(\chi^2_k > c_\alpha) = \alpha$.

(c) An approximate $1-\alpha$ confidence set for $\boldsymbol{\mu}$ based on the result in part (a) may be formed by plotting the elliptical set

$$\{\boldsymbol{\mu} : n(\overline{\mathbf{X}} - \boldsymbol{\mu})^\top S_n^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu}) = c_\alpha\}.$$

For a random sample of size 100 from $N_2(\mathbf{0}, \Sigma)$, where $\Sigma = \begin{pmatrix} 1 & 3/5 \\ 3/5 & 1 \end{pmatrix}$, produce a scatterplot of the sample and plot 90% and 99% confidence sets on this scatterplot.

Hints: In part (c), to produce a random vector with the $N_2(\mathbf{0}, \Sigma)$ distribution, take a $N_2(\mathbf{0}, I)$ random vector and left-multiply by a matrix A such that $AA^\top =$

Σ. It is not hard to find such an A (it may be taken to be lower triangular). One way to graph the ellipse is to find a matrix B such that $B^\top S_n^{-1} B = I$. Then note that

$$\{\boldsymbol{\mu} : n(\overline{\mathbf{X}} - \boldsymbol{\mu})^\top S_n^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu}) = c_\alpha\} \;=\; \{\overline{\mathbf{X}} - B\boldsymbol{\nu} : \boldsymbol{\nu}^\top \boldsymbol{\nu} = c_\alpha/n\},$$

so it remains only to find points $\boldsymbol{\nu}$, closely spaced, such that $\boldsymbol{\nu}^\top\boldsymbol{\nu}$ equals a constant. To find a matrix B such as the one specified, note that the matrix of eigenvectors of S_n, properly normalized, gives an orthogonal matrix that diagonalizes.

Exercise 9.2 Verify Equations (9.5) and (9.6).

Exercise 9.3 Prove Lemma 9.2 and Lemma 9.3, then verify Equation (9.9).

Exercise 9.4 Pearson's chi-square for a 2-way table: Product multinomial model. If A and B are categorical variables with 2 and k levels, respectively, and we collect random samples of size m and n from levels 1 and 2 of A, then classify each individual according to its level of the variable B, the results of this study may be summarized in a $2 \times k$ table. The standard test of the independence of variables A and B is the Pearson chi-square test, which may be written as

$$\sum_{\text{all cells in table}} \frac{(O_j - E_j)^2}{E_j},$$

where O_j is the observed count in cell j and E_j is the estimate of the expected count under the null hypothesis. Equivalently, we may set up the problem as follows: If $\mathbf{X}$ and $\mathbf{Y}$ are independent Multinomial$(m, \mathbf{p})$ and Multinomial $(n, \mathbf{p})$ random vectors, respectively, then the Pearson chi-square statistic is

$$\mathrm{W}^2 \;=\; \sum_{j=1}^{k} \left\{ \frac{(X_j - mZ_j/N)^2}{mZ_j/N} + \frac{(Y_j - nZ_j/N)^2}{nZ_j/N} \right\},$$

where $\mathbf{Z} = \mathbf{X} + \mathbf{Y}$ and $N = n + m$. (Note: I used W^2 to denote the chi-square statistic to avoid using yet another variable that looks like an X.)

Prove that if $N \to \infty$ in such a way that $n/N \to \alpha \in (0, 1)$, then

$$\mathrm{W}^2 \;\xrightarrow{d}\; \chi^2_{k-1}.$$

Exercise 9.5 Pearson's chi-square for a 2-way table: Multinomial model. Now consider the case in which $(\mathbf{X}, \mathbf{Y})$ is a single multinomial $(N, \mathbf{q})$ random $2k$-vector.

X_i will still denote the $(1, i)$ entry in a $2 \times k$ table, and Y_i will still denote the $(2, i)$ entry.

(a) In this case, $\mathbf{q}$ is a $2k$-vector. Let $\alpha = q_1/(q_1 + q_{k+1})$ and define $\mathbf{p}$ to be the k-vector such that $(q_1, \ldots, q_k) = \alpha\mathbf{p}$. Prove that under the usual null hypothesis that variable A is independent of variable B (i.e., the row variable and the column variable are independent), $\mathbf{q} = (\alpha\mathbf{p}, (1-\alpha)\mathbf{p})$ and $p_1 + \cdots + p_k = 1$.

(b) As in Problem 9.4, let $\mathbf{Z} = \mathbf{X} + \mathbf{Y}$. Assume the null hypothesis is true and suppose that for some reason α is known. The Pearson chi-square statistic may be written as

$$\mathrm{W}^2 \;=\; \sum_{j=1}^{k} \left\{ \frac{(X_j - \alpha Z_j)^2}{\alpha Z_j} + \frac{(Y_j - (1-\alpha)Z_j)^2}{(1-\alpha)Z_j} \right\}. \tag{9.10}$$

Find the joint asymptotic distribution of

$$\sqrt{N\alpha(1-\alpha)} \left(\frac{X_1}{N\alpha} - \frac{Y_1}{N(1-\alpha)}, \ldots, \frac{X_k}{N\alpha} - \frac{Y_k}{n(1-\alpha)} \right)$$

and use this result to prove that $\mathrm{W}^2 \xrightarrow{d} \chi^2_k$.

Exercise 9.6 In Problem 9.5(b), it was assumed that α was known. However, in most problems this assumption is unrealistic. Therefore, we replace all occurrences of α in Equation (9.10) by $\hat{\alpha} = \sum_{i=1}^{k} X_i/N$. This results in a different asymptotic distribution for the W^2 statistic. Suppose we are given the following multinomial probabilities for a 2×2 table with independent row and column variables:

$P(X_1 = 1) = .1$	$P(X_2 = 1) = .15$	.25
$P(Y_1 = 1) = .3$	$P(Y_2 = 1) = .45$	.75
.4	.6	1

Note that $\alpha = .25$ in the above table. Let $N = 50$ and simulate 1000 multinomial random vectors with the above probabilities. For each, calculate the value of W^2 using both the known value $\alpha = .25$ and the value $\hat{\alpha}$ estimated from the data. Plot the empirical distribution function of each of these two sets of 1000 values. Compare with the theoretical distribution functions for the χ^2_1 and χ^2_2 distributions.

Hint: To generate a multinomial random variable with expectation vector matching the table above, because of the independence inherent in the table you can generate two independent Bernoulli random variables with respective success probabilities equal to the margins: That is, let $P(A = 2) = 1 - P(A = 1) = .6$

and $P(B = 2) = 1 - P(B = 1) = .75$, then classify the multinomial observation into the correct cell based on the random values of A and B.

Exercise 9.7 The following example comes from genetics. There is a particular characteristic of human blood (the so-called MN blood group) that has three types: M, MN, and N. Under idealized circumstances known as Hardy-Weinberg equilibrium, these three types occur in the population with probabilities $p_1 = \pi_M^2$, $p_2 = 2\pi_M\pi_N$, and $p_3 = \pi_N^2$, respectively, where π_M is the frequency of the M allele in the population and $\pi_N = 1 - \pi_M$ is the frequency of the N allele.

We observe data $\mathbf{X}_1, \ldots, \mathbf{X}_n$, where $\mathbf{X}_i$ has one of three possible values: $(1, 0, 0)^T$, $(0, 1, 0)^T$, or $(0, 0, 1)^T$, depending on whether the ith individual has the M, MN, or N blood type. Denote the total number of individuals of each of the three types by n_1, n_2, and n_3; in other words, $n_j = n\overline{X}_j$ for each j.

If the value of π_M were known, then the results of this section would show that the Pearson χ^2 statistic converges in distribution to a chi-square distribution on 2 degrees of freedom. However, of course we usually don't know π_M. Instead, we estimate it using the maximum likelihood estimator $\hat{\pi}_M = (2n_1 + n_2)/2n$. By the invariance principle of maximum likelihood estimation, this gives $\hat{\mathbf{p}} = (\hat{\pi}_M^2, 2\hat{\pi}_M\hat{\pi}_N, \hat{\pi}_N^2)^T$ as the maximum likelihood estimator of $\mathbf{p}$.

(a) Define $\mathbf{B}_n = \sqrt{n}(\overline{\mathbf{X}} - \hat{\mathbf{p}})$. Use the delta method to derive the asymptotic distribution of $\Gamma^{-1/2}\mathbf{B}_n$, where $\Gamma = \text{diag}(p_1, p_2, p_3)$.

(b) Define $\hat{\Gamma}$ to be the diagonal matrix with entries $\hat{p}_1, \hat{p}_2, \hat{p}_3$ along its diagonal. Derive the asymptotic distribution of $\hat{\Gamma}^{-1/2}\mathbf{B}_n$.

(c) Derive the asymptotic distribution of the Pearson chi-square statistic

$$\chi^2 = \sum_{j=1}^{3} \frac{(n_j - n\hat{p}_j)^2}{n\hat{p}_j}. \tag{9.11}$$

Exercise 9.8 Take $\pi_M = .75$ and $n = 100$ in the situation described in Problem 9.7. Simulate 500 realizations of the data.

(a) Compute

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

for each of your 500 datasets. Compare the empirical distribution function of these statistics with both the χ_1^2 and χ_2^2 distribution functions. Comment on what you observe.

(b) Compute the χ^2 statistic of Equation (9.11) for each of your 500 datasets. Compare the empirical distribution function of these statistics with both the χ_1^2 and χ_2^2 distribution functions. Comment on what you observe.

9.2 Power of Pearson's chi-square test

Suppose the k-vector $\mathbf{X}$ is distributed as multinomial $(n, \mathbf{p})$, and we wish to test the null hypothesis $H_0 : \mathbf{p} = \mathbf{p}^0$ against the alternative $H_1 : \mathbf{p} \neq \mathbf{p}^0$ using the Pearson chi-square test. We are given a sequence of specific alternatives $\mathbf{p}_n$ satisfying $\sqrt{n}(\mathbf{p}^{(n)} - \mathbf{p}) \to \boldsymbol{\delta}$ for some constant matrix $\boldsymbol{\delta}$. Note that this means $\sum_{i=1}^{k} \delta_i = 0$, a fact that will be used later. Our task is to derive the limit of the power of the sequence of tests under the sequence of alternatives $\mathbf{p}_n$.

The notion of a noncentral chi-square distribution will be important in this development, so we first give a definition.

Definition 9.5 If $A_1, \ldots, A_n$ are independent random variables with $A_i \sim N(\mu_i, 1)$, then the distribution of $A_1^2 + A_2^2 + \cdots + A_n^2$ is noncentral chi-square with n degrees of freedom and noncentrality parameter $\phi = \mu_1^2 + \cdots + \mu_n^2$. (In particular, the distribution depends on the μ_i only through ϕ.) We denote this distribution $\chi_n^2(\phi)$. Equivalently, we can say that if $\mathbf{A} \sim N_n(\boldsymbol{\mu}, I)$, then $\mathbf{A}^\top\mathbf{A} \sim \chi_n^2(\phi)$ where $\phi = \boldsymbol{\mu}^\top\boldsymbol{\mu}$.

Note that this is not a valid definition unless we may prove that the distribution of $\mathbf{A}^\top\mathbf{A}$ depends on $\boldsymbol{\mu}$ only through $\phi = \boldsymbol{\mu}^\top\boldsymbol{\mu}$. We prove this as follows. First, note that if $\phi = 0$ then there is nothing to prove. Otherwise, define $\boldsymbol{\mu}^* = \boldsymbol{\mu}/\sqrt{\phi}$. Next, find an orthogonal matrix Q whose first row is $(\boldsymbol{\mu}^*)^\top$. (It is always possible to do this, though we do not explain the details here. One method is the process of Gram-Schmidt orthogonalization). Then $Q\mathbf{A} \sim N_k(Q\boldsymbol{\mu}, I)$. Since $Q\boldsymbol{\mu}$ is a vector with first element $\sqrt{\phi}$ and remaining elements 0, $Q\mathbf{A}$ has a distribution that depends on $\boldsymbol{\mu}$ only through ϕ. But $\mathbf{A}^\top\mathbf{A} = (Q\mathbf{A})^\top(Q\mathbf{A})$, proving that the distribution of $\mathbf{A}^\top\mathbf{A}$ depends on the μ_i only through ϕ.

We will derive the power of the chi-square test by adapting the projection matrix technique of Section 9.1. First, we prove a lemma that generalizes Lemma 9.2.

Lemma 9.6 Suppose $\mathbf{Z} \sim N_k(\boldsymbol{\mu}, P)$, where P is a projection matrix of rank $r \leq k$ and $P\boldsymbol{\mu} = \boldsymbol{\mu}$. Then $\mathbf{Z}^\top\mathbf{Z} \sim \chi_r^2(\boldsymbol{\mu}^\top\boldsymbol{\mu})$.

Proof: Since P is a covariance matrix, it is symmetric, which means that there exists an orthogonal matrix Q with $QPQ^{-1} = \text{diag}(\boldsymbol{\lambda})$, where $\boldsymbol{\lambda}$ is the vector of eigenvalues of P. Since P is a projection matrix, all of its eigenvalues are 0 or 1. Since P has rank r, exactly

r of the eigenvalues are 1. Without loss of generality, assume that the first r entries of $\boldsymbol{\lambda}$ are 1 and the last $k-r$ are 0. The random vector $Q\mathbf{Z}$ is $N_n(Q\boldsymbol{\mu}, \text{diag}\,(\boldsymbol{\lambda}))$, which implies that $\mathbf{Z}^\top\mathbf{Z} = (Q\mathbf{Z})^\top(Q\mathbf{Z})$ is by definition distributed as $\chi^2_r(\phi) + \varphi$, where

$$\phi \;=\; \sum_{i=1}^{r}(Q\boldsymbol{\mu})_i^2 \;\text{ and }\; \varphi \;=\; \sum_{i=r+1}^{k}(Q\boldsymbol{\mu})_i^2.$$

Note, however, that

$$Q\boldsymbol{\mu} \;=\; QP\boldsymbol{\mu} \;=\; QPQ^\top Q\boldsymbol{\mu} \;=\; \text{diag}\,(\boldsymbol{\lambda})Q\boldsymbol{\mu}. \tag{9.12}$$

Since entries $r+1$ through k of $\boldsymbol{\lambda}$ are zero, the corresponding entries of $Q\boldsymbol{\mu}$ must be zero because of Equation (9.12). This implies two things: First, $\varphi = 0$; and second,

$$\phi \;=\; \sum_{i=1}^{r}(Q\boldsymbol{\mu})_i^2 \;=\; \sum_{i=1}^{k}(Q\boldsymbol{\mu})_i^2 \;=\; (Q\boldsymbol{\mu})^\top(Q\boldsymbol{\mu}) \;=\; \boldsymbol{\mu}^\top\boldsymbol{\mu}.$$

Thus, $\mathbf{Z}^\top\mathbf{Z} \sim \chi^2_r(\boldsymbol{\mu}^\top\boldsymbol{\mu})$, which proves the result. ■

Define $\Gamma = \text{diag}\,(\mathbf{p}^0)$. Let $\Sigma = \Gamma - \mathbf{p}^0(\mathbf{p}^0)^\top$ be the usual multinomial covariance matrix under the null hypothesis; i.e., $\sqrt{n}(\mathbf{X}^{(n)}/n - \mathbf{p}^0) \xrightarrow{d} N_k(\mathbf{0}, \Sigma)$ if $X_n \sim \text{multinomial}(n, \mathbf{p}^0)$. Consider X_n to have instead a multinomial $(n, \mathbf{p}_n)$ distribution. Under the assumption made earlier that $\sqrt{n}(\mathbf{p}_n - \mathbf{p}^0) \to \boldsymbol{\delta}$, it may be shown that

$$\sqrt{n}(\mathbf{X}^{(n)}/n - \mathbf{p}^{(n)}) \;\xrightarrow{d}\; N_k(\mathbf{0}, \Sigma). \tag{9.13}$$

We claim that the limit (9.13) implies that the chi square statistic $n(\mathbf{X}^{(n)}/n - \mathbf{p}^0)^\top\Gamma^{-1}(\mathbf{X}^{(n)}/n - \mathbf{p}^0)$ converges in distribution to $\chi^2_{k-1}(\boldsymbol{\delta}^\top\Gamma^{-1}\boldsymbol{\delta})$, a fact that we now prove.

First, recall that we have already shown that $\Gamma^{-1/2}\Sigma\Gamma^{-1/2}$ is a projection matrix of rank $k-1$. Define $\mathbf{V}^{(n)} = \sqrt{n}(\mathbf{X}^{(n)}/n - \mathbf{p}^0)$. Then

$$\mathbf{V}^{(n)} \;=\; \sqrt{n}(\mathbf{X}^{(n)}/n - \mathbf{p}^{(n)}) + \sqrt{n}(\mathbf{p}^{(n)} - \mathbf{p}^0).$$

The first term on the right hand side converges in distribution to $N_k(\mathbf{0}, \Sigma)$ and the second term converges to $\boldsymbol{\delta}$. Therefore, Slutsky's theorem implies that $\mathbf{V}^{(n)} \xrightarrow{d} N_k(\boldsymbol{\delta}, \Sigma)$, which gives

$$\Gamma^{-1/2}\mathbf{V}^{(n)} \xrightarrow{d} N_k(\Gamma^{-1/2}\boldsymbol{\delta}, \Gamma^{-1/2}\Sigma\Gamma^{-1/2}).$$

Thus, if we can show that $(\Gamma^{-1/2}\Sigma\Gamma^{-1/2})(\Gamma^{-1/2}\boldsymbol{\delta}) = (\Gamma^{-1/2}\boldsymbol{\delta})$, then the result we wish to prove follows from Lemma 9.6. But

$$\Gamma^{-1/2}\Sigma\Gamma^{-1}\boldsymbol{\delta} \;=\; \Gamma^{-1/2}[\Gamma - \mathbf{p}^0(\mathbf{p}^0)^\top]\Gamma^{-1}\boldsymbol{\delta} \;=\; \Gamma^{-1/2}[\boldsymbol{\delta} - \mathbf{p}^0(\mathbf{1})^\top\boldsymbol{\delta}] \;=\; \Gamma^{-1/2}\boldsymbol{\delta}$$

since $\mathbf{1}^\top\boldsymbol{\delta} = \sum_{i=1}^k \delta_i = 0$. Thus, we conclude that the chi-square statistic converges in distribution to $\chi^2_{k-1}(\boldsymbol{\delta}^\top\Gamma^{-1}\boldsymbol{\delta})$ under the sequence of alternatives $\mathbf{p}_1, \mathbf{p}_2, \ldots$.

Example 9.7 For a particular trinomial experiment with $n = 200$, suppose the null hypothesis is $H_0 : \mathbf{p} = \mathbf{p}^0 = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$. (This hypothesis might arise in the context of a genetics experiment.) We may calculate the approximate power of the Pearson chi-square test at level $\alpha = 0.01$ against the alternative $\mathbf{p} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.

First, set $\boldsymbol{\delta} = \sqrt{n}(\mathbf{p} - \mathbf{p}^0) = \sqrt{200}(\frac{1}{12}, -\frac{1}{6}, \frac{1}{12})$. Under the alternative $\mathbf{p}$, the chi square statistic is approximately noncentral χ^2_2 with noncentrality parameter

$$\boldsymbol{\delta}^\top \text{diag}\,(\mathbf{p}^0)^{-1}\boldsymbol{\delta} \;=\; 200\left(\frac{4}{144} + \frac{2}{36} + \frac{4}{144}\right) \;=\; \frac{200}{9}.$$

Since the test rejects H_0 whenever the statistic is larger than the .99 quantile of χ^2_2, namely 9.210, the power is approximated by $P\{\chi^2_2(\frac{200}{9}) > 9.210\} = 0.965$. These values were found using R as follows:

```
> qchisq(.99,2)
[1] 9.21034
> 1-pchisq(.Last.value, 2, ncp=200/9)
[1] 0.965006
```

Exercises for Section 9.2

Exercise 9.9 Suppose we have a tetranomial experiment and wish to test the hypothesis $H_0 : \mathbf{p} = (1/4, 1/4, 1/4, 1/4)$ against the alternative $H_1 : \mathbf{p} \neq (1/4, 1/4, 1/4, 1/4)$ at the .05 level.

(a) Approximate the power of the test against the alternative $(1/10, 2/10, 3/10, 4/10)$ for a sample of size $n = 200$.

(b) Give the approximate sample size necessary to give power of 80% against the alternative in part (a).