Cochran's theorem

Kristian Wichmann

June 16, 2017

1 Basic ingredients

Let $U_1, U_2, \ldots, U_n$ be i.i.d. standard normally distributed random variables. Now, consider the sum:

$$\sum_{i=1}^{n} U_i^2 \tag{1}$$

By definition this quantity is χ_n^2 distributed. But assume that it can also be written as a sum of quadratic forms in the U_i variables:

$$\sum_{i=1}^{n} U_i^2 = Q_1 + Q_2 + \cdots + Q_k \tag{2}$$

That Q_i is a quadratic form means that:

$$Q_i = \sum_{j=1}^{n} \sum_{k=1}^{n} U_j B_{jk}^{(i)} U_k \tag{3}$$

Here, $B^{(i)}$ is a symmetric, positive semi-definite matrix.

From this we immediately see that the following must be true:

$$\sum_{i=1}^{k} B^{(i)} = I_n \tag{4}$$

Here I_n is the unit matrix of size n as usual. We will also use the notation r_i for the rank of $B^{(i)}$.

1.1 Example: Standard normal sample

As a simple example, consider a population that is standard normally distributed. A random sample of size n corresponds to the U_i's above. Of interest is the usual estimators of means and sum of squares (for estimating variance):

$$\bar{U} = \frac{1}{n}\sum_{i=1}^{n} U_i, \quad SS_U = \sum_{i=1}^{n}(U_i - \bar{U})^2 \tag{5}$$

Now, let's try to express the sum of the U_i^2 in terms of these. Consider the usual trick:

$$\sum_{i=1}^{n} U_i^2 = \sum_{i=1}^{n}(U_i - \bar{U} + \bar{U})^2 = \sum_{i=1}^{n}(U_i - \bar{U})^2 + \sum_{i=1}^{n}\bar{U}^2 + 2\bar{U}\sum_{i=1}^{n}(U_i - \bar{U}) \tag{6}$$

The last sum is equal to zero by definition of $\bar{U}$, so we're left with:

$$\sum_{i=1}^{n} U_i^2 = \sum_{i=1}^{n}(U_i - \bar{U})^2 + \sum_{i=1}^{n}\bar{U}^2 = SS_U + n\bar{U}^2 \tag{7}$$

We've now split the sum into two quadratic forms:

$$Q_1 = SS_U = \sum_{i=1}^{n}(U_i - \bar{U})^2, \quad Q_2 = n\bar{U}^2 \tag{8}$$

Let's rewrite these to make the corresponding B-matrices clear. Q_2 is the simplest:

$$Q_2 = n\bar{U}^2 = n\left(\frac{1}{n}\sum_{i=1}^{n} U_i\right)^2 = \frac{1}{n}\sum_{i,j=1}^{n} U_i U_j \tag{9}$$

In other words, Q_2 can be expressed as $\sum_{j=1}^{n}\sum_{k=1}^{n} U_j B_{jk}^{(2)} U_k$, where $B^{(2)} = \frac{1}{n}J_n$. Here, J_n is the $n \times n$ matrix containing all 1's. Now, let's tackle Q_1:

$$Q_1 = \sum_{i=1}^{n}(U_i - \bar{U})^2 = \sum_{i=1}^{n} U_i^2 + \sum_{i=1}^{n}\bar{U}^2 - 2\bar{U}\sum_{i=1}^{n} U_i = \sum_{i=1}^{n} U_i^2 - nU^2 \tag{10}$$

Once again, the definition of $\bar{U}$ has been used. This means (using the result for Q_1, that $B^{(1)} = I_n - \frac{1}{n}J_n$.

It may superficially seem that this is just a roundabout way of stating the trivial identity $I_n = \left(I_n - \frac{1}{n}J_n\right) + \left(\frac{1}{n}J_n\right)$, which is indeed what equation (4) already told us. But the important part is the decomposition into estimators of mean and variance.

1.2 Example: Normal sample

Things don't change too much if we consider a sample from a general, normal distributed population. If the stochastic variables $X_1, X_2, \ldots, X_n$ are i.i.d. and $X_i \sim N(\mu, \sigma^2)$, simply transform as usual:

$$U_i = \frac{X_i - \mu}{\sigma} \tag{11}$$

Since these will all be standard normally distributed, everything proceeds as in the previous section.

2 Statement of Cochran's theorem

We're now ready to state the actual theorem:

Theorem 1. *Let $U_1, U_2, \ldots, U_n$ be i.i.d. random variables, $U_i \sim N(0,1)$. Further assume:*

$$\sum_{i=1}^{n} U_i^2 = Q_1 + Q_2 + \cdots + Q_k, \tag{12}$$

where $Q_i = \sum_{jk} U_j B_{jk}^{(i)} U_k$ with $B^{(i)}$ being a symmetric, positive definite matrix of rank r_i. The following statements are now equivalent:

1. *$\sum_{i=1}^{k} r_i = n$.*
2. *$Q_i \sim \chi^2_{r_i}$ for all i.*
3. *The Q_i's are all independent of each other.*

2.1 Example: Standard normal sample (cont.)

Let's determine the ranks of $B^{(1)}$ and $B^{(2)}$ to see if the prerequisites of the theorem holds. $B^{(2)}$ is the simplest: all the rows are exactly the same (and non-zero), so the rank must be 1. So we need to prove that the rank of Q_1 is $n-1$:

Theorem 2. *The rank of the matrix $I_n - \frac{1}{n} J_n$ is $n-1$.*

Proof. The proof is by induction.
1. Base case: For $n = 1$, the matrix is simply $[0]$, which clearly has rank 0. To avoid division by zero in the inductive step, we also need to show this for $n = 2$. In this case, the matrix is:

$$\begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \tag{13}$$

Each row is equal to the other multiplied by -1, so the rank is clearly 1.
2. Inductive step: Assume $I_{n-1} - \frac{1}{n-1}J_{n-1}$ to have rank $n-2$. The matrix $I_n - \frac{1}{n}J_n$ has the form:

$$\begin{pmatrix} \frac{n-1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & \frac{n-1}{n} & \cdots & -\frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \cdots & \frac{n-1}{n} \end{pmatrix} \tag{14}$$

To bring the matrix to echelon form, the upper row is normalized, i.e. multiplied by $\frac{n}{n-1}$ (this is why we needed $n = 2$ in the base case). The non-leading entries thus turn into:

$$-\frac{1}{n}\frac{n}{n-1} = -\frac{1}{n-1} \tag{15}$$

After this row operation we're therefore left with:

$$\begin{pmatrix} 1 & -\frac{1}{n-1} & \cdots & -\frac{1}{n-1} \\ -\frac{1}{n} & \frac{n-1}{n} & \cdots & -\frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \cdots & \frac{n-1}{n} \end{pmatrix} \tag{16}$$

$\frac{1}{n}$ times the first row is now added to all the other rows. There's two possibilities for the results. Diagonal elements:

$$\frac{n-1}{n} - \frac{1}{n(n-1)} = \frac{(n-1)^2 - 1}{n(n-1)} = \frac{n^2 + 1 - 2n - 1}{n(n-1)} = \frac{n(n-2)}{n(n-1)} = \frac{n-2}{n-1} \tag{17}$$

Off-diagonal elements:

$$-\frac{1}{n} - \frac{1}{n(n-1)} = \frac{-(n-1) - 1}{n(n-1)} = -\frac{n}{n(n-1)} = -\frac{1}{n-1} \tag{18}$$

So, we're left with the matrix:

$$\begin{pmatrix} 1 & -\frac{1}{n-1} & \cdots & -\frac{1}{n-1} \\ 0 & \frac{n-2}{n-1} & \cdots & -\frac{1}{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & -\frac{1}{n-1} & \cdots & \frac{n-2}{n-1} \end{pmatrix} \tag{19}$$

However, the lower right $(n-1)\times(n-1)$ submatrix is exactly $I_{n-1} - \frac{1}{n-1}J_{n-1}$, which by the induction hypothesis has rank $n-2$. Therefore, the total number of leading ones in the echelon form of $I_n - \frac{1}{n}J_n$ is $n-1$. $\square$

According to Cochran's theorem, the estimators Q_1 and Q_2 are therefore independent and distributed as follows:

$$nSS_U \sim \chi^2_{n-1} \quad \bar{U}^2 \sim \chi^2_1 \tag{20}$$

This is the source of the (in)famous $n-1$ term in the variance of a sample.

2.2 Normal sample (cont.)

When the samples X_i are normal distributed as opposed to standard normally distributed, i.e. $X_i \sim N(\mu, \sigma^2)$, equation (20) still holds, but now the U's are given by:

$$U_i = \frac{X_i - \mu}{\sigma} \tag{21}$$

This means that the mean estimator is:

$$\bar{U} = \frac{1}{n}\sum_{i=1}^{n} \frac{X_i - \mu}{\sigma} = \frac{1}{n\sigma}\left(\sum_{i=1}^{n} X_i - n\mu\right) = \frac{1}{\sigma}\left(\bar{X} - \mu\right) \tag{22}$$

So the sum of squares is:

$$SS_U = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma} - \frac{1}{\sigma}\left(\bar{X} - \mu\right)\right)^2 \tag{23}$$

The inside of the parenthesis is simply $\frac{1}{\sigma}(X_i - \bar{X})$, which means that the sum of squares is:

$$SS_U = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{\sigma^2} SS_X \tag{24}$$

Therefore the distribution of the sum of squares of the X'es can be expressed:

$$nSS_X \sim \sigma^2 \chi^2_{n-1} \tag{25}$$

3 A detour: Idempotent matrices

It turns out that the concept of an *idempotent* matrix will be of use to us here. A square matrix A is called idempotent if $A^2 = A$. Such matrices have some interesting properties. Let's start with this:

Theorem 3. *The eigenvalues of an idempotent matrix A are either* 0 *or* 1.

Proof. Assume λ is an eigenvalue of A. I.e. there exists a vector v such that $Av = \lambda v$. Apply A to both sides of the equation to get:

$$A(Av) = A(\lambda v) \Leftrightarrow Av = \lambda^2 v \tag{26}$$

This means that $\lambda v = \lambda^2 v$. This is only possible if $\lambda = \lambda^2$, which has only two solutions: $\lambda = 0$ and $\lambda = 1$ $\square$

The converse of theorem 3 is also true:

Theorem 4. *If a diagonalizable matrix A has only eigenvalues 0 and 1, A is idempotent.*

Proof. Let r be the multiplicity of the eigenvalue 1. It follows that the multiplicity of the eigenvalue 0 is $n - r$. Since A is diagonalizable, it means that an orthogonal matrix O exists, such that $A = O^T DO$ where $D = \text{diag}\{\underbrace{1, 1, \cdots, 1}_{r \text{ times}}, \underbrace{0, 0, \cdots, 0}_{n-r \text{ times}}\}$. Now calculate:

$$A^2 = A = (O^T DO)^2 = O^T DOO^T DO = O^T D^2 O \tag{27}$$

Since the diagonal of D only has zeroes and ones in it, in fact $D^2 = D$. Therefore $A^2 = O^T DO = A$. $\square$

In fact, it turns out that all idempotent matrices are diagonalizable:

Theorem 5. *An idempotent matrix A is diagonalizable.*

Proof. Let r be the rank of r. Then the image space of A has a dimension of r. Let $w_1, w_2, \cdots, w_r$ be a basis. Since each of these vectors are in the image, it means that there exists vectors $v_1, v_2, \cdots, v_r$ such that $Av_i = w_i$. So:

$$Aw_i = A(Av_i) = Av_i = w_i \tag{28}$$

Hence the multiplicity of the eigenvalue 1 is (at least) r. By the dimensionality theorem, the dimension of the null space is $n - r$. This means that 0 is an eigenvalue with multiplicity $n - r$. Thus A has n eigenvalues and is therefore diagonalizable. $\square$

From the proof we also see that the multiplicity of the eigenvalue 1 is equal to the rank of A. This leads to the following neat result:

Theorem 6. *The rank r of an idempotent matrix A is equal to its trace.*

Proof. A is diagonalizable, so there exists an orthogonal matrix O such that $A = O^T DO$, where D is the diagonal matrix $\text{diag}\{\underbrace{1, 1, \cdots, 1}_{r \text{ times}}, \underbrace{0, 0, \cdots, 0}_{n-r \text{ times}}\}$. Since trace is invariant under similarity transformations, this means that $\text{tr}A = \text{tr}D = r$. □

Theorem 7. *Let A and B be matrices such that $A = BA$ and $B = AB$. Then A and B are both idempotent.*

Proof. The two results are shown in almost exactly the same way:

- For A:
$$A = BA = (AB)A = A(BA) = AA = A^2 \tag{29}$$

- For B:
$$B = AB = (BA)B = B(AB) = BB = B^2 \tag{30}$$

□

Theorem 8. *Let A and B be idempotent matrices which satisfies $AB = BA = 0$. Then $A + B$ is idempotent.*

Proof. Compute the square of $A + B$:

$$(A + B)^2 = (A + B)(A + B) = A^2 + AB + BA + B^2 \tag{31}$$

The two mixed terms disappear by assumption, and since A and B are both idempotent, this is simply equal to $A + B$. □

Theorem 9. *Let A be idempotent. Then $B = I - A$ is also idempotent, and $AB = BA = 0$.*

Proof. Calculate the square of B:

$$B^2 = (I-A)^2 = (I-A)(I-A) = I^2-IA-AI+A^2 = I-2A+A = I-A = B \tag{32}$$

In the penultimate step A's idempotency was used. Now calculate AB:

$$AB = A(I - A) = A - A^2 = A - A = 0 \tag{33}$$

Similarly for BA:

$$BA = (I - A)A = A - A^2 = A - A = 0 \tag{34}$$

□

3.1 Example: Normal sample yet again

It turns out, that the matrices $B^{(1)}$ and $B^{(2)}$ are indeed idempotent. Recall that $B^{(1)} = I_n - \frac{1}{n}J_n$. So:

$$\begin{aligned}(B^{(1)})^2 = \left(I_n - \frac{1}{n}J_n\right)^2 = I_n^2 + \frac{1}{n^2}J_n^2 - \frac{1}{n}(I_nJ_n + J_nI_n) = \\ I_n + \frac{1}{n}J_n - 2\frac{1}{n}J_n = I_n - \frac{1}{n}J_n = B^{(1)}\end{aligned} \tag{35}$$

Here, we've used that $J_n^2 = nJ_n$. Similarly, since $B^{(2)} = \frac{1}{n}J_n$:

$$(B^{(2)})^2 = \left(\frac{1}{n}J_n\right)^2 = \frac{1}{n}J_n = B^{(2)} \tag{36}$$

So both matrices are indeed idempotent, and we can immediately conclude that the traces, and therefore the ranks are $n-1$ and 1 respectively.

4 Idempotency of the $B^{(i)}$ matrices

In our example. The $B^{(i)}$ matrices happened to be idempotent. The question is if this is a coincidence? It turns out it is not:

Theorem 10. *With the notation from section* 1*, if* $\sum_{i=1}^{k} r_i = n$*, then the* $B^{(i)}$ *matrices are all idempotent.*

Proof. Start by recalling equation (4): $\sum_{i=1}^{k} B^{(i)} = I_n$. For a given i this can also be stated as:

$$I_n = B^{(i)} + \underbrace{\sum_{j \neq i} B^{(j)}}_{C^{(i)}} \tag{37}$$

Here, we've defined $C^{(i)}$ as shown by the brace. By subadditivity of matrix rank, this implies:

$$\text{rank}(I_n) \leq \text{rank}(B^{(i)}) + \text{rank}(C^{(i)}) \Leftrightarrow \text{rank}(C^{(i)}) \geq n - r_i \tag{38}$$

However it also means:

$$\text{rank}(C^{(i)}) \leq \sum_{j \neq i}^{k} \text{rank}(B^{(j)}) = n - r_i \tag{39}$$

The last equality follows from the assumption $\sum_{i=1}^{k} r_i = n$. So the only possibility is $\text{rank}(C^{(i)}) = n - r_i$.

Since $B^{(i)}$ is symmetric, it is diagonalizable. So there exists an orthogonal matrix such that

$$B^{(i)} = O^T D_i O, \tag{40}$$

where D_i is diagonal. r_i of the diagonal elements are non-zero:

$$D_i = \text{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_{r_i}, \underbrace{0, \cdots, 0}_{n-r_i \text{ zeroes}}\} \tag{41}$$

Equation (40) can be rewritten as $O^T B^{(i)} O = D_i$. This comes in handy as equation (37) is sandwiched by O and O^T in the same way:

$$O I_n O^T = O B^{(i)} O^T + O C^{(i)} O^T \Leftrightarrow I_n = D_i + O C^{(i)} O^T \tag{42}$$

Therefore $OC^{(i)}O^T$ must be a diagonal matrix. Arranging the orthogonal matrices back this means

$$C^{(i)} = O^T \left[\text{diag}\{1-\lambda_1, 1-\lambda_2, \cdots, 1-\lambda_{r_i}, \underbrace{1, \cdots, 1}_{n-r_i \text{ ones}}\} \right] O \tag{43}$$

So $C^{(i)}$ is diagonalized by the same orthogonal transformation as $B^{(i)}$. Now, we know the rank of $C^{(i)}$ is $n - r_i$, but there's already $n - r_i$ eigenvalues equal to one, so the rest of the eigenvalues must be zero. In other words $1 - \lambda_1 = 0$, $1 - \lambda_2 = 0, \ldots, 1 - \lambda_{r_i} = 0$ which means that the first r_i λ's must be zero! So $B^{(i)}$ has only eigenvalues of 0 and 1 and therefore is idempotent. $\square$

5 Quadratic forms of idempotent matrices

As the last theorem shows, we'll be dealing with such quadratic forms. The following will be useful:

Theorem 11. *Let $U_1, U_2, \cdots, U_n$ be i.i.d. standard normal stochastic variables and A be a symmetric $n \times n$ matrix. Consider the quadratic form:*

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{n} U_i A_{ij} U_j \tag{44}$$

Then A is idempotent with rank r if and only if $Q \sim \chi^2_r$.

Proof. Since A is symmetric, it is therefore diagonalizable. So $A = O^T D O$, where O is an orthogonal and D a diagonal matrix, respectively. So, in matrix form:

$$Q = U^T (O^T D O) U = (U^T O^T) D (OU) = (OU)^T D (OU) \tag{45}$$

We need to show both implications. First assume A is idempotent with rank r. Then:

$$D = \text{diag}\{\underbrace{1, 1, \cdots, 1}_{r \text{ times}}, \underbrace{0, 0, \cdots, 0}_{n-r \text{ times}}\} \tag{46}$$

Since an orthogonal transformation of a multidimensional standard normal is once again a multidimensional standard normal, it follows that Q is indeed a sum of r squares of independent standard normals. So $Q \sim \chi^2_r$.

Conversely, assume $Q \sim \chi^2_r$. This means that Q can be written as a sum of r squares of i.i.d. standard normals V_i:

$$Q = \sum_{i=1}^{r} V_i^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} V_i D'_{ij} V_j \tag{47}$$

Here $D' = \text{diag}\{\underbrace{1, 1, \cdots, 1}_{r \text{ times}}, \underbrace{0, 0, \cdots, 0}_{n-r \text{ times}}\}$. So in matrix form:

$$(OU)^T D (OU) = V^T D' V \tag{48}$$

But eigenvalues must remain the same no matter what basis is used, so D and D' must have the same diagonal elements, which means that A is idempotent with rank r. □

Another result we will need is the following (which is a special case of what is sometimes known as Craig's theorem):

Theorem 12. *Let $U_1, U_2, \cdots, U_n$ be i.i.d. standard normally distributed variables and A and B be symmetric $n \times n$ matrices. The quadratic forms*

$$Q_A = U^T A U, \quad Q_B = U^T B U \tag{49}$$

are now independent if and only if $AB = 0$.

Proof. Both A and B are symmetric and hence diagnalizable. So there exists orthogonal matrices O_A and O_B such that:

$$A = O_A^T D_A O_A, \quad B = O_B^T D_B O_B \tag{50}$$

Here D_A and D_B are diagonal matrices. So we have:

$$Q_A = U^T O_A^T D_A O_A U^T = (O_A U)^T D_A (O_A U), \quad Q_B = (O_B U)^T D_B (O_B U) \tag{51}$$

Using the standard formula for the covariance of a product, we get:

$$\begin{aligned}\mathrm{Cov}(Q_A, Q_B) = E[Q_A Q_b] - E[Q_A]E[Q_B] = \\ E[(O_A U)^T D_A (O_A U)(O_B U)^T D_B (O_B U)] - \\ E[(O_A U)^T D_A (O_A U)] E[(O_B U)^T D_B (O_B U)]\end{aligned}$$

If Q_A and Q_B are independent, then they're also uncorrelated. The covariance must therefore be zero. □

6 Properties of the χ^2-distribution

6.1 Moment generating function

The moment generating function (MGF) of the χ^2-distribution will be usedul in the following. Recall that the MGF of a random variable X is defined as:

$$M_X(t) = E\left[e^{tX}\right] \tag{52}$$

Now, assume that $X \sim \chi^2_r$, where r is the number of degrees of freedom. Recall that the probability density function (pdf) is:

$$f(x) = \frac{x^{r/2-1} e^{-x/2}}{2^{r/2}\Gamma(r/2)}, \quad x > 0 \tag{53}$$

This means that the MGF is:

$$M_X(t) = \int_0^\infty e^{tx} f(x)\ dx = \frac{1}{2^{r/2}\Gamma(r/2)} \int_0^\infty x^{r/2-1} \exp\left[x(t - 1/2)\right]\ dx \tag{54}$$

Recall the standard integral:

$$\int_0^\infty x^n \exp(-ax)\ dx = \frac{\Gamma(n+1)}{a^{n+1}} \tag{55}$$

Here, $n = r/2 - 1$ and $a = 1/2 - t$, so the integral part of equation 54 is:

$$\frac{\Gamma(r/2)}{(1/2 - t)^{r/2}} \tag{56}$$

So the factorials cancel and we're left with

$$M_X(t) = \frac{1}{2^{r/2}(1/2 - t)^{r/2}} = \frac{1}{(1 - 2t)^{r/2}} = (1 - 2t)^{-r/2} \tag{57}$$

7 Proof of Cochran's theorem

By now we've done most of the heavy lifting needed to actually prove theorem 1. We need to show that the three conditions are equivalent.

7.1 Proof that 1 $\Rightarrow$ 2

Proof. By theorem 10 the matrix $B^{(i)}$ is idempotent. By theorem 11 $Q_i \sim \chi^2_{r_i}$. $\square$

7.2 Proof that 2 $\Rightarrow$ 1

Proof. This follows directly from theorem 11 and the definition of χ^2 distributions. $\square$

7.3 Proof that 2 $\Rightarrow$ 3

Proof. From the proof of theorem 10 we know that the same orthogonal matrix O diagonalizes all the $B^{(i)}$ matrices simultaneously: $B^{(i)} = O^T D_i O$, where each D_i is diagonal. Furthermore idempotency also means all diagonal elements are either 0 or 1. We can now write:

$$U^T U = U^T \left[O^T \left(\sum_{i=1}^{k} D_i \right) O \right] U = (OU)^T \left(\sum_{i=1}^{k} D_i \right) (OU) \tag{58}$$

Since eigenvalues does not depend on basis, we must have $I_n = \sum_{i=1}^{k} D_i$. But since all diagonal D elements are 0 or 1 there can be only exactly one 1 in a single D matrix for each of the n basis vectors. Since orthogonal transformation maps standard multinormal to standard multinormal, the basis $V = OU$ consists of i.i.d. standard normally distributed stochastic variables. But each of the n basis vectors V_i will appear only in exactly one quadratic form. Hence the Q's must be pairwise independent. $\square$

7.4 Proof that 3 $\Rightarrow$ 2

Proof. According to equation (4) for any natural number m we must have:

$$\left(\sum_{i=1}^{k} B^{(i)} \right)^m = I_n \tag{59}$$

$\square$