Quadratic forms
Cochran's theorem, degrees of freedom, and all that…

Dr. Frank Wood

Why We Care

- Cochran's theorem tells us about the distributions of partitioned sums of squares of normally distributed random variables.
- Traditional linear regression analysis relies upon making statistical claims about the distribution of sums of squares of normally distributed random variables (and ratios between them)
  - i.e. in the simple normal regression model

$$SSE/\sigma^2 = \sum(Y_i - \hat{Y}_i)^2 \sim \chi^2(n-2)$$

- Where does this come from?

Outline

- Review some properties of multivariate Gaussian distributions and sums of squares
- Establish the fact that the multivariate Gaussian sum of squares is $\chi^2(n)$ distributed
- Provide intuition for Cochran's theorem
- Prove a lemma in support of Cochran's theorem
- Prove Cochran's theorem
- Connect Cochran's theorem back to matrix linear regression

Preliminaries

- Let Y_1, Y_2, …, Y_n be N(μ_i,σ_i^2) random variables.
- As usual define

$$Z_i = \frac{Y_i - \mu_i}{\sigma_i}$$

- Then we know that each Z_i ~ N(0,1)

From Wackerly et al, 306

Theorem 0 : Statement

- The sum of squares of n N(0,1) random variables is χ^2 distributed with n degrees of freedom

$$\left(\sum_{i=1}^{n} Z_i^2\right) \sim \chi^2(n)$$

Theorem 0: Givens

- Proof requires knowing both

1.

$$Z_i^2 \sim \chi^2(\nu),\ \nu = 1 \text{ or equivalently}$$
$$Z_i^2 \sim \Gamma(\nu/2, 2),\ \nu = 1$$

Homework, midterm ?

2. If Y_1, Y_2, …., Y_n are independent random variables with moment generating functions $m_{Y_1}(t)$, $m_{Y_2}(t)$, … $m_{Y_n}(t)$, then when U = $Y_1 + Y_2 + \ldots Y_n$

$$m_U(t) = m_{Y_1}(t) \times m_{Y_2}(t) \times \ldots \times m_{Y_n}(t)$$

and from the uniqueness of moment generating functions that $m_U(t)$ fully characterizes the distribution of U

Theorem 0: Proof

- The moment generating function for a $\chi^2(\nu)$ distribution is (Wackerley et al, back cover)

$$m_{Z_i^2}(t) = (1-2t)^{\nu/2}, \text{ where here } \nu = 1$$

- The moment generating function for

$$V = (\sum_{i=1}^n Z_i^2)$$

is (by given prerequisite)

$$m_V(t) = m_{Z_1^2}(t) \times m_{Z_2^2}(t) \times \cdots \times m_{Z_n^2}(t)$$

Theorem 0: Proof

- But $m_V(t) = m_{Z_1^2}(t) \times m_{Z_2^2}(t) \times \cdots \times m_{Z_n^2}(t)$
  is just

$$m_V(t) = (1-2t)^{1/2} \times (1-2t)^{1/2} \times \cdots \times (1-2t)^{1/2}$$

- Which is itself, by inspection, just the moment generating function for a χ^2(n) random variable

$$m_V(t) = (1-2t)^{n/2}$$

  which implies (by uniqueness) that

$$V = (\sum_{i=1}^{n} Z_i^2) \sim \chi^2(n)$$

Quadratic Forms and Cochran's Theorem

- Quadratic forms of normal random variables are of great importance in many branches of statistics
  - Least squares
  - ANOVA
  - Regression analysis
  - etc.
- General idea
  - Split the sum of the squares of observations into a number of quadratic forms where each corresponds to some cause of variation

Quadratic Forms and Cochran's Theorem

- The conclusion of Cochran's theorem is that, under the assumption of normality, the various quadratic forms are independent and χ^2 distributed.
- This fact is the foundation upon which many statistical tests rest.

Preliminaries: A Common Quadratic Form

- Let

$$\mathbf{x} \sim \mathbf{N}(\mu, \mathbf{\Lambda})$$

- Consider the (important) quadratic form that appears in the exponent of the normal density

$$(\mathbf{x} - \mu)' \mathbf{\Lambda}^{-1} (\mathbf{x} - \mu)$$

- In the special case of μ = 0 and Λ = I this reduces to x'x which by what we just proved we know is χ^2 (n) distributed
- Let's prove that this holds in the general case

Lemma 1

- Suppose that x~N(μ, Λ) with $|\Lambda| > 0$ then (where n is the dimension of x)

$$(\mathbf{x}-\mu)'\mathbf{\Lambda^{-1}}(\mathbf{x}-\mu) \sim \chi^{\mathbf{2}}(\mathbf{n})$$

- Proof: Set y = $\Lambda^{-1/2}$(x-μ) then
  - E(y) = 0
  - Cov(y) = $\Lambda^{-1/2}\ \Lambda\ \Lambda^{-1/2}$ = I
  - That is y ~N(0,I) and thus

$$(\mathbf{x}-\mu)'\mathbf{\Lambda^{-1}}(\mathbf{x}-\mu) = \mathbf{y}'\mathbf{y} \sim \chi^{\mathbf{2}}(\mathbf{n})$$

Note: this is sometimes called “sphering” data

The Path

- What do we have?

$$(\mathbf{x}-\mu)'\mathbf{\Lambda}^{-\mathbf{1}}(\mathbf{x}-\mu) = \mathbf{y}'\mathbf{y} \sim \chi^{\mathbf{2}}(\mathbf{n})$$

- Where are we going?
  - (Cochran's Theorem) Let X_1, X_2, …, X_n be independent N(0,σ^2)-distributed random variables, and suppose that

    $$\sum_{i=1}^{n} X_i^2 = Q_1 + Q_2 + \ldots + Q_k$$

    Where Q_1, Q_2, …, Q_k are positive semi-definite quadratic forms in the random variables X_1, X_2, …, X_m, that is,

    $$Q_i = \mathbf{X}'\mathbf{A_i}\mathbf{X},\ i = 1, 2, \ldots, k$$

Cochran's Theorem Statement

Set Rank $A_i = r_i$, i=1,2,…, k. If

$$r_1 + r_2 + \ldots + r_k = n$$

- Then
  1. $Q_1, Q_2, \ldots, Q_k$ are *independent*
  2. $Q_i \sim \sigma^2\chi^2(r_i)$

Reminder: the rank of a matrix is the number of linearly independent rows / columns in the matrix, or, equivalently, the number of its non-zero eigenvalues

Closing the Gap

- We start with a lemma that will help us prove Cochran's theorem
- This lemma is a linear algebra result
- We also need to know a couple results regarding linear transformations of normal vectors
  - We attend to those first.

Linear transformations

- Theorem 1: Let X be a normal random vector. The components of X are independent iff they are uncorrelated.
  - Demonstrated in class by setting $Cov(X_i, X_j) = 0$ and then deriving product form of joint density

Linear transformations

- Theorem 2: Let X ~ N(μ, Λ) and set Y = C'X where the orthogonal matrix C is such that C'Λ C = D. Then Y ~ N(C'μ, D); the components of Y are independent; and Var Y_k = λ_k, k =1…n, where λ_1, λ_2,…, λ_n are the eigenvalues of Λ

Look up singular value decomposition.

Orthogonal transforms of iid N(0,σ^2) variables

- Let X ~ N(μ, σ^2 I) where $\sigma^2 > 0$ and set Y = CX where C is an orthogonal matrix. Then Cov{Y} = Cσ^2IC' = σ^2 I
- This leads to
- Theorem 2: Let X ~ N(μ, σ^2 I) where $\sigma^2 > 0$, let C be an arbitrary orthogonal matrix, and set Y=CX. The Y $\sim$ N(Cμ, σ^2I); in particular, Y_1, Y_2, …, Y_n are independent normal random variables with the same variance σ^2.

Where we are

- Now we can transform N(μ, $\sum$) random variables into N(0,D) random variables.
- We know that orthogonal transformations of a random vector X ~ N(μ,σ^2I) results in a transformed vector whose elements are still independent
- The preliminaries are over, now we proceed to proving a lemma that forms the backbone of Cochran's theorem.

Lemma 1

- Let $x_1, x_2, \ldots, x_n$ be real numbers. Suppose that $\sum x_i^2$ can be split into a sum of positive semidefinite quadratic forms, that is,

$$\sum_{i=1}^{n} x_i^2 = Q_1 + Q_2 + \ldots + Q_k$$

where $Q_i = x'A_ix$ and (rank Q_i =) rank $A_i = r_i$, $i=1,2,\ldots,k$. If $\sum r_i = n$ then there exists an orthogonal matrix C such that, with $x = Cy$ we have…

Lemma 2 cont.

$$\begin{aligned} Q_1 &= y_1^2 + y_2^2 + \ldots + y_{r_1}^2 \\ Q_2 &= y_{r_1+1}^2 + y_{r_1+2}^2 + \ldots + y_{r_1+r_2}^2 \\ Q_3 &= y_{r_1+r_2+1}^2 + y_{r_1+r_2+2}^2 + \ldots + y_{r_1+r_2+r_3}^2 \\ &\vdots \\ Q_k &= y_{n-r_k+1}^2 + y_{n-r_k+2}^2 + \ldots + y_n^2 \end{aligned}$$

- Remark: Note that different quadratic forms contain different y-variables and that the number of terms in each Q_i equals the rank, r_i, of Q_i

What’s the point?

- We won’t construct this matrix C, it’s just useful for proving Cochran’s theorem.
- We care that
  - The y_i^2’s end up in different sums – we’ll use this to prove independence of the different quadratic forms.

Proof

- We prove the n=2 case. The general case is obtained by induction. [Gut 95]
- For n=2 we have

$$Q = \sum_{i=1}^{n} x_i^2 = \mathbf{x}'\mathbf{A_1}\mathbf{x} + \mathbf{x}'\mathbf{A_2}\mathbf{x} \quad (= \mathbf{Q_1} + \mathbf{Q_2})$$

where A_1 and A_2 are positive semi-definite matrices with ranks r_1 and r_2 respectively and $r_1 + r_2 = n$

Proof: Cont.

- By assumption there exists an orthogonal matrix C such that

$$\mathbf{C'A_1C = D}$$

where D is a diagonal matrix, the diagonal elements of which are the eigenvalues of A_1; λ_1, λ_2, …, λ_n.

- Since Rank(A_1) =r_1 then r_1 eigenvalues are positive and n-r_1 eigenvalues equal zero.
- Suppose without restriction that the first r_1 eigenvalues are positive and the rest are zero.

Proof : Cont

- Set $$\mathbf{x} = \mathbf{C}\mathbf{y}$$

  and remember that when C is an orthogonal matrix that

$$\mathbf{x}'\mathbf{x} = (\mathbf{C}\mathbf{y})'\mathbf{C}\mathbf{y} = \mathbf{y}'\mathbf{C}'\mathbf{C}\mathbf{y} = \mathbf{y}'\mathbf{y}$$

then

$$Q = \sum y_i^2 = \sum_{i=1}^{r_1} \lambda_i y_i^2 + \mathbf{y}'\mathbf{C}'\mathbf{A_2}\mathbf{C}\mathbf{y}$$

Proof : Cont

- Or, rearranging terms slightly and expanding the second matrix product

$$\sum_{i=1}^{r_i}(1-\lambda_i)y_i^2 + \sum_{i=r_1+1}^{n} y_i^2 = \mathbf{y'C'A_2Cy}$$

- Since the rank of the matrix A_2 equals r_2 (= $n-r_1$) we can conclude that

$$\lambda_1 = \lambda_2 = \ldots = \lambda_{r_1} = 1$$

$$Q_1 = \sum_{i=1}^{r_1} y_i^2 \text{ and } Q_2 = \sum_{i=r_1+1}^{n} y_i^2$$

which proves the lemma for the case n=2.

What does this mean again?

- This lemma only has to do with real numbers, not random variables.
- It says that if $\sum x_i^2$ can be split into a sum of positive semi-definite quadratic forms then there is a orthogonal (projection) matrix x=Cy (or C'x = y) that makes each of the quadratic forms have some very nice properties, foremost of which is that
  - Each y_i appears in only one resulting sum of squares.

Cochran's Theorem

Let X_1, X_2, …, X_n be independent N(0,σ^2)-distributed random variables, and suppose that

$$\sum_{i=1}^{n} X_i^2 = Q_1 + Q_2 + \ldots + Q_k$$

Where Q_1, Q_2, …, Q_k are positive semi-definite quadratic forms in the random variables X_1, X_2, …, X_m, that is,

$$Q_i = \mathbf{X}'\mathbf{A_i}\mathbf{X},\ i = 1, 2, \ldots, k$$

Set Rank A_i = r_i, i=1,2,…, k. If

$$r_1 + r_2 + \ldots + r_k = n$$

then

1. Q_1, Q_2, …, Q_k are *independent*
2. Q_i ~ $\sigma^2\chi^2(r_i)$

Proof [from Gut 95]

- From the previous lemma we know that there exists an orthogonal matrix C such that the transformation X=CY yields

$$\begin{aligned} Q_1 &= Y_1^2 + Y_2^2 + \ldots + Y_{r_1}^2 \\ Q_2 &= Y_{r_1+1}^2 + Y_{r_1+2}^2 + \ldots + Y_{r_1+r_2}^2 \\ Q_3 &= Y_{r_1+r_2+1}^2 + Y_{r_1+r_2+2}^2 + \ldots + Y_{r_1+r_2+r_3}^2 \\ &\vdots \\ Q_k &= Y_{n-r_k+1}^2 + Y_{n-r_k+2}^2 + \ldots + Y_n^2 \end{aligned}$$

- But since every Y^2 occurs in exactly one Q_j and the Y_i's are all independent N(0, σ^2) RV's (because C is an orthogonal matrix) Cochran's theorem follows.

Huh?

- Best to work an example to understand why this is important
- Let's consider the distribution of a sample variance (not regression model yet). Let Y_i, i=1…n be samples from Y ~ N(0, σ^2). We can use Cochran's theorem to establish the distribution of the sample variance (and it's independence from the sample mean).

Example

- Recall form of SSTO for regression model and note that the form of SSTO = (n-1) $s^2\{Y\}$

$$SSTO = \sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n}$$

- Recognize that this can be rearranged and the re-expressed in matrix form

$$\sum Y_i^2 = \sum (Y_i - \bar{Y})^2 + \frac{(\sum Y_i)^2}{n}$$

$$\mathbf{Y}'\mathbf{I}\mathbf{Y} = \mathbf{Y}'(\mathbf{I} - \frac{1}{n}\mathbf{J})\mathbf{Y} + \mathbf{Y}'(\frac{1}{n}\mathbf{J})\mathbf{Y}$$

Example cont.

- From earlier we know that

$$\mathbf{Y}'\mathbf{I}\mathbf{Y} \sim \sigma^2\chi^2(n)$$

  but we can read off the rank of the quadratic form as well (rank(I) = n)
- The ranks of the remaining quadratic forms can be read off too (with some linear algebra reminders)

$$\mathbf{Y}'\mathbf{Y} = \mathbf{Y}'(\mathbf{I} - \tfrac{1}{n}\mathbf{J})\mathbf{Y} + \mathbf{Y}'(\tfrac{1}{n}\mathbf{J})\mathbf{Y}$$

Linear Algebra Reminders

- For a symmetric and idempotent matrix A, rank(A) = trace(A), the number of non-zero eigenvalues of A.
  – Is (1/n)J symmetric and idempotent?
  – How about (I-(1/n)J)?
- trace(A + B) = trace(A) + trace(B)
- Assuming they are we can read off the ranks of each quadratic form

$$\mathbf{Y}'\mathbf{I}\mathbf{Y} = \mathbf{Y}'(\mathbf{I} - \frac{1}{n}\mathbf{J})\mathbf{Y} + \mathbf{Y}'(\frac{1}{n}\mathbf{J})\mathbf{Y}$$

rank: n   rank: n-1   rank: 1

Cochran's Theorem Usage

$$\mathbf{Y}'\mathbf{I}\mathbf{Y} = \mathbf{Y}'(\mathbf{I} - \frac{1}{n}\mathbf{J})\mathbf{Y} + \mathbf{Y}'(\frac{1}{n}\mathbf{J})\mathbf{Y}$$

rank: n rank: n-1 rank: 1

- Cochran's theorem tells us, immediately, that

$$\sum Y_i^2 \sim \sigma^2\chi^2(n), \sum(Y_i - \bar{Y})^2 \sim \sigma^2\chi^2(n-1), \frac{(\sum Y_i)^2}{n} \sim \sigma^2\chi^2(1)$$

because each of the quadratic forms is χ^2 distributed with degrees of freedom given by the rank of the corresponding quadratic form and each sum of squares is independent of the others.

What about regression?

- Quick comment: in the preceding, one can think about having modeled the population with a single parameter model – the parameter being the mean. The number of degrees of freedom in the sample variance sum of squares is reduced by the number of parameters fit in the linear model (one, the mean)
- Now – regression.

Rank of ANOVA Sums of Squares

	Rank
$SSTO = \mathbf{Y}'[\mathbf{I} - \frac{1}{n}\mathbf{J}]\mathbf{Y}$	n-1
$SSE = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y}$	n-p
$SSR = \mathbf{Y}'[\mathbf{H} - \frac{1}{n}\mathbf{J}]\mathbf{Y}$	p-1

good midterm question

- Slightly stronger version of Cochran's theorem needed (will assume it exists) to prove the following claim(s).

- From Cochran's theorem, knowing the ranks of

$$
\begin{aligned}
SSTO &= \mathbf{Y}'[\mathbf{I} - \frac{1}{n}\mathbf{J}]\mathbf{Y} \\
SSE &= \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y} \\
SSR &= \mathbf{Y}'[\mathbf{H} - \frac{1}{n}\mathbf{J}]\mathbf{Y}
\end{aligned}
$$

gives you this immediately

$$
\begin{aligned}
SSTO &\sim \sigma^2\chi^2(n-1) \\
SSE &\sim \sigma^2\chi^2(n-p) \\
SSR &\sim \sigma^2\chi^2(p-1)
\end{aligned}
$$

F Test for Regression Relation

- Now the test of whether there is a regression relation between the response variable Y and the set of X variables $X_1, \ldots, X_{p-1}$ makes more sense
- The F distribution is defined to be the ratio of χ^2 distributions that have themselves been normalized by their number of degrees of freedom.

F Test Hypotheses

- If we want to choose between the alternatives
  - $H_0 : \beta_1 = \beta_2 = \beta_3 \ldots = \beta_{p-1} = 0$
  - H_1 : not all β_k k=1…n equal zero

- We can use the defined test statistic

$$F^* = \frac{MSR}{MSE} \sim \frac{\frac{\sigma^2 \chi^2(p-1)}{p-1}}{\frac{\sigma^2 \chi^2(n-p)}{n-p}}$$

- The decision rule to control the Type I error at α is

$$\text{If } F^* \leq F(1-\alpha; p-1, n-p), \text{conclude } H_0$$
$$\text{If } F^* > F(1-\alpha; p-1, n-p), \text{conclude } H_a$$