A Bit of the Big Picture

In several lectures we have encountered quadratic forms like $q_i = \mathbf{y}'A_i\mathbf{y}$ where the $\mathbf{y}$ have normal distributions and the matrices A_i are idempotent, meaning that $A_i^2 = A_i$. We have derived results like how to calculate $E(\mathbf{y}'A_i\mathbf{y})$ and how to establish independence of $\mathbf{y}'A_i\mathbf{y}$ and $\mathbf{y}'A_j\mathbf{y}$ by checking to see if $A_iA_j = 0$. Why have we been so interested in these quadratic forms? Below I will outline what we have been trying to accomplish with these quadratic forms, and how they can give us a deeper understanding of the theory of linear models.

In our practice of fitting linear models to data we always encounter decompositions of sums of squares. In regression models we see:

$$\text{TSS} = \text{RegSS} + \text{RSS},$$

while in (balanced) analysis of variance models we see things like:

$$\text{TSS} = \text{SSA} + \text{SSB} + \text{SSAB} + \text{RSS}.$$

Our analyses are then conducted by forming suitable ratios of normalized sums of squares to create F statistics. These sums of squares decompositions can be written in matrix form almost like:

$$\mathbf{y}'\mathbf{y} = \mathbf{y}'A_1\mathbf{y} + \mathbf{y}'A_2\mathbf{y},$$

except that the total sum of squares is corrected for the mean, so that TSS $= \sum(y_i - \overline{y})^2$. We have already used a vector decomposition like the one above for regression:

$$\mathbf{y}'\mathbf{y} = \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} + \mathbf{y}'(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y}.$$

If we write the mean in the form $\overline{\mathbf{y}}=\frac{1}{n}\mathbf{J}_n\mathbf{y}$, (recall that $\mathbf{J}_n$ is a n x n matrix of 1's) then we can obtain a matrix version of the equation (TSS = RegSS + RSS) above as:

$$\mathbf{y}'(\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n)\mathbf{y} = \mathbf{y}'(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \frac{1}{n}\mathbf{J}_n)\mathbf{y} + \mathbf{y}'(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y}.$$

As we have already seen when testing a subset of regression coefficients, these matrices can become even more complicated-looking. A question of interest at this point is, given a sum of squares decomposition like $\mathbf{y}'\mathbf{y} = \mathbf{y}'A_1\mathbf{y} + \mathbf{y}'A_2\mathbf{y}$, when do the quadratic forms follow independent chi-square distributions, allowing us to form valid F ratios? If $\mathbf{y}$ has a multivariate normal distribution, then it turns out that the key is in the A_i matrices. If A_i is symmetric and idempotent of rank r, then the quadratic form $\mathbf{y}'A_i\mathbf{y}$ follows a (either central or non-central) chi-square distribution with r degrees of freedom. As we already have seen, the key to independence of two quadratic forms $\mathbf{y}'A_i\mathbf{y}$ and $\mathbf{y}'A_j\mathbf{y}$ is that $A_iA_j = 0$. There is a fascinating geometric interpretation of these matrices, which are called projection matrices, which we have discussed in previous lectures. Here are some more results about the use of these quadratic forms for linear model analyses:

Some Results about Distributions

Definition: If $\mathbf{y} \sim N_n(\mathbf{0}, \mathbf{I}_n)$, then $\mathbf{y}'\mathbf{y}$ has a (central) chi-squared distribution with n degrees of freedom, denoted by χ^2_n. We will denote its pdf by $f_{\chi^2}(x; n)$. The probability density function of a χ^2_n random variable is given by:

$$f_{\chi^2}(x; n) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}.$$

Definition: If $\mathbf{y} \sim N_n(\mu, \mathbf{I}_n)$, then $\mathbf{y}'\mathbf{y}$ has a (noncentral) chi-squared distribution with n degrees of freedom and noncentrality parameter $\lambda = \mu'\mu$, denoted by $\chi^2_n(\lambda)$. We will denote its pdf by $f_{\chi^2}(x; n, \lambda)$. The probability density function of a $\chi^2_n(\lambda)$ random variable is given by:

$$f_{\chi^2}(x; n, \lambda) = \sum_{j=0}^{\infty} (e^{-\lambda} \frac{\lambda^j}{j!})\, f_{\chi^2}(x; n + 2j),$$

which is a Poisson-weighted-average of central chi-squared densities. Note that the noncentral chi-squared distribution with n degrees of freedom and $\lambda = 0$ is just the central chi-squared distribution with n degrees of freedom. Clearly if $\mathbf{y} \sim N_n(\mu, \sigma^2_\varepsilon \mathbf{I}_n)$ then $(\mathbf{y} - \mu)'(\mathbf{y} - \mu)/\sigma^2_\varepsilon \sim \chi^2_n$, and also if $\mathbf{y} \sim N_n(\mu, \sigma^2_\varepsilon \mathbf{I}_n)$ then $\mathbf{y}'\mathbf{y}/\sigma^2_\varepsilon \sim \chi^2_n(\lambda)$, with $\lambda = \mu'\mu/\sigma^2_\varepsilon$. Note that some authors use different definitions of the noncentrality parameter λ, with some using $\lambda = \mu'\mu/2$ and others using $\lambda = \sqrt{\mu'\mu}$.

Definition: If $u \sim \chi^2_m$ and $v \sim \chi^2_n$ are two independent (central) chi-squared random variables, then $F_{m,n} = (u/m)/(v/n)$ has a (central) F distribution with m and n degrees of freedom.

Definition: If $u \sim \chi^2_m(\lambda)$ and $v \sim \chi^2_n$ are two independent (one noncentral and one central) chi-squared random variables, then $F_{m,n} = (u/m)/(v/n)$

has a noncentral F distribution with m and n degrees of freedom and noncentrality parameter λ.

There are related random variables such as the noncentral t distribution and the doubly noncentral F distribution, but we will not concern ourselves with them. The noncentral chi-squared and F distributions will be important for investigating the power of hypothesis tests of interest.

Some Results about Idempotent Matrices

Definition: A matrix A is called idempotent if $A^2 = AA = A$. We will generally assume that our idempotent matrices are symmetric. Matrices that are idempotent and symmetric are called projection matrices.

Result 7: For a projection matrix A, all of its eigenvalues are either 0 or 1. Also, rank(A) =trace(A) = the number of eigenvalues it has equal to 1. If A is of dimension p x p then rank(A)+rank($I_p - A$) = p.

Some Final Results about the Big Picture

Result 8: If $\mathbf{y} \sim N_n(\mu, \mathbf{I}_n)$ and A is a symmetric matrix, then $\mathbf{y}'A\mathbf{y}$ has a noncentral chi-squared distribution with r =trace(A) degrees of freedom if and only if A is an idempotent matrix. The noncentrality parameter is $\lambda = \mu' A \mu$.

We have already seen that two quadratic forms $q_1 = \mathbf{y}'A_1\mathbf{y}$ and $q_2 = \mathbf{y}'A_2\mathbf{y}$ are independent if and only if $A_1 \mathbf{\Sigma} A_2 = \mathbf{0}$ (where $V(\mathbf{y}) = \mathbf{\Sigma}$). Thus the key themes for pairwise independence of these quadratic forms are idempotence and orthogonality. When the TSS is decomposed into more than two quadratic forms, Cochran's Theorem addresses the mutual independence of the set of quadratic forms. Here it is stated like it appears in Searle (1971):

Result 9 (Cochran's Theorem): Let $\mathbf{y} \sim N_n(\mathbf{0}, \mathbf{I}_n)$ and A_i, $i = 1, 2, \cdots, k$, be symmetric matrices with rank(A_i) = r_i with $\mathbf{I}_n = \sum_{i=1}^{k} A_i$. Then the quadratic forms $q_i = \mathbf{y}'A_i\mathbf{y}$, $i = 1, 2, \cdots, k$, are mutually independent, $\chi^2_{r_i}$ chi-squared random variables if and only if $\sum_{i=1}^{k} r_i = n$.

References

Hocking, R.R. 1996. Methods and Applications of Linear Models: Regression and the Analysis of Variance, New York: John Wiley & Sons, Inc.

Searle, S.R. 1971. Linear Models, New York: John Wiley & Sons, Inc.

Seber, G.A.F., and Lee, A.J. 2003. Linear Regression Analysis, Second Edition, Hoboken, NJ: Wiley-Interscience.