Lecture 14. 16.11.2015

2. Quadratic forms in normal variates

In deriving the normal equations, we minimised the total sum of squares

$$SS := (y - A\beta)^T(y - A\beta)$$

w.r.t. $\beta$. The minimum value is called the sum of squares for error,

$$SSE := (y - A\hat{\beta})^T(y - A\hat{\beta}).$$

From the normal equations (NE) and the definition of the projection matrix $P$,

$$A\hat{\beta} = Py.$$

So

$$SSE = (y - Py)^T(y - Py) = y^T y - y^T Py - y^T Py + y^T P^T Py = y^T (I - P) y,$$

using $P^T = P$ and $P^2 = P$, and a little matrix algebra (see e.g. [BF], 3.4) gives also

$$SSE = (y - A\beta)^T (I - P)(y - A\beta).$$

The sum of squares for regression is

$$SSR := (\hat{\beta} - \beta)^T C (\hat{\beta} - \beta).$$

Again, a little matrix algebra (see e.g. [BF], 3.4) gives

$$SSR = (y - A\beta)^T P (y - A\beta).$$

So

$$SS = SSR + SSE :$$

$$(y - A\beta)^T (y - A\beta) = (y - A\beta)^T P (y - A\beta) + (y - A\beta)^T (I - P) (y - A\beta); \ (SSD)$$

either of both of these are called the sum-of-squares decomposition. Now from the model equations (ME), $y - A\beta = \epsilon$ is a random $n$-vector whose components are iid $N(0, \sigma^2)$. So (SSD) decomposes a quadratic form in normal variates $\epsilon = (\epsilon_1, \ldots, \epsilon_n)^T$ with matrix $I$ into the sum of two quadratic forms with matrices $P$ and $I - P$. Now by Craig’s theorem ([KS1], (15.55))
such quadratic forms with matrices $A$, $B$ are independent iff $AB = 0$. But since

$$P(I - P) = P - P^2 = P - P = 0,$$

this shows that $SSR$ and $SSE$ are independent. Thus $(SSD)$ decomposes the total sum of squares into a sum of independent sums of squares – the main tool used in regression.

We recall some results from Linear Algebra (see e.g. [BF] Ch. 3 and the references cited there). We need the trace $\text{trace}(A)$ of a square matrix $A = (a_{ij})$, defined as the sum of its diagonal elements:

$$\text{trace}(A) = \sum a_{ii}.$$

(i) A real symmetric matrix $A$ can be diagonalised by an orthogonal transformation $O$ to a diagonal matrix $D$:

$$O^T AO = D.$$

(ii) For $A$ idempotent (a projection), its eigenvalues are 0 or 1.

(iii) For $A$ idempotent, its trace is its rank. So if we have a quadratic form $x^TPx$ with $P$ a projection of rank $r$ and $x$ an $n$-vector $(x_1, \ldots, x_n)^T$ with $x_i$ iid $N(0, \sigma^2)$, we can diagonalise by an orthogonal transformation $y = Ox$ to a sum of squares of $r$ normals (wlog the first $r$):

$$x^TPx = y_1^2 + \ldots + y_r^2, \quad y_i \text{ iid } N(0, \sigma^2).$$

So by definition of the chi-square distribution,

$$x^TPx \sim \sigma^2 \chi^2(r).$$

Sums of Projections

Suppose that $P_1, \ldots, P_k$ are symmetric projection matrices with sum the identity:

$$I = P_1 + \ldots + P_k.$$

Take the trace of both sides: the $n \times n$ identity matrix $I$ has trace $n$. Each $P_i$ has trace its rank $n_i$, so as trace is additive

$$n = n_1 + \ldots + n_k.$$

Then squaring,

$$I = I^2 = \sum_i P_i^2 + \sum_{i<j} P_i P_j = \sum_i P_i + \sum_{i<j} P_i P_j.$$
Taking the trace,
\[ n = \sum n_i + \sum_{i<j} \text{trace}(P_iP_j) = n + \sum_{i<j} \text{trace}(P_iP_j) : \]
\[ \sum_{i<j} \text{trace}(P_iP_j) = 0. \]

Now
\[ \text{trace}(P_iP_j) = \text{trace}(P_i^2P_j^2) \quad (P_i, P_j \text{ projections}) \]
\[ = \text{trace}((P_jP_i)(P_iP_j)) \quad (\text{trace}(AB) = \text{trace}(BA)) \]
\[ = \text{trace}((P_iP_j)^T(P_iP_j)) \quad ((AB)^T = B^TA^T; P_i, P_j \text{ symmetric}) \]
\[ \geq 0, \]
since for a matrix \( M \)
\[ \text{trace}(M^TM) = \sum_i (M^TM)_{ii} \]
\[ = \sum_i \sum_j (M^T)_{ij}(M)_{ji} \]
\[ = \sum_i \sum_j m_{ij}^2 \]
\[ \geq 0. \]

So we have a sum of non-negative terms being zero. So each term must be zero. That is, the square of each element of \( P_iP_j \) must be zero. So each element of \( P_iP_j \) is zero, so matrix \( P_iP_j \) is zero:
\[ P_iP_j = 0 \quad (i \neq j). \]

This is the condition that the linear forms \( P_1x, \ldots, P_kx \) be independent (below). Since the \( P_i x \) are independent, so are the \( (P_i x)^T(P_i x) = x^TP_i^TP_i x \), i.e. \( x^TP_i x \) as \( P_i \) is symmetric and idempotent. That is, the quadratic forms \( x^TP_1x, \ldots, x^TP_kx \) are also independent.

We now have
\[ x^T x = x^TP_1x + \ldots + x^TP_kx. \]
The left is \( \sigma^2 \chi^2(n) \); the \( i \)th term on the right is \( \sigma^2 \chi^2(n_i) \).

We summarise our conclusions.

**Theorem (Chi-Square Decomposition Theorem).** If
\[ I = P_1 + \ldots + P_k, \]

3
with each $P_i$ a symmetric projection matrix with rank $n_i$, then
(i) the ranks sum:
\[ n = n_1 + \ldots + n_k; \]
(ii) each quadratic form $Q_i := x^TP_ix$ is chi-squared:
\[ Q_i \sim \sigma^2 \chi^2(n_i); \]
(iii) the $Q_i$ are mutually independent.

This fundamental result gives all the distribution theory commonly needed for the Linear Model (for which see e.g. [BF]). In particular, since $F$-distributions are defined in terms of distributions of independent chi-squares, it explains why we constantly encounter $F$-statistics, and why all the tests of hypotheses that we encounter will be $F$-tests. This is so throughout the Linear Model – Multiple Regression, as here, Analysis of Variance, Analysis of Covariance and more advanced topics.

Note. The result above generalises beyond our context of projections. With the projections $P_i$ replaced by symmetric matrices $A_i$ of rank $n_i$ with sum $I$, the corresponding result (Cochran’s Theorem, 1934, also known as the Fisher-Cochran theorem) is that (i), (ii) and (iii) are equivalent. The proof is harder (one needs to work with quadratic forms, where we were able to work with linear forms). For monograph treatments, see e.g. Rao [R], sections 1c.1 and 3b.4 and Kendall & Stuart [KS1], sections 15.16 - 15.21.

3. The multivariate normal (Gaussian) distribution

In $n$ dimensions, for a random $n$-vector $X = (X_1, \ldots, X_n)^T$, one needs
(i) a mean vector $\mu = (\mu_1, \ldots, \mu_n)^T$ with $\mu_i = EX_i, \mu = E[X]$;
(ii) a covariance matrix $\Sigma = (\sigma_{ij})$, with $\sigma_{ij} = \text{cov}(X_i, X_j)$: $\Sigma = \text{cov}(X)$.

First, note how mean vectors and covariance matrices transform under linear changes of variable:

**Proposition.** If $Y = AX + b$, with $Y, b$ $m$-vectors, $A$ an $m \times n$ matrix and $X$ an $n$-vector, (i) the mean vectors are related by $E[Y] = AE[X] + b = A\mu + b$; (ii) the covariance matrices are related by $\Sigma_Y = A\Sigma_XA^T$. 
