pfsl14.tex
Lecture 14. 7.11.2013 (half-hour: Problems)

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. β. 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

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

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

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

So

$$SS = SSR + SSE :$$

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

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* $trace(A)$ of a square matrix $A = (a_{ij})$, defined as the sum of its diagonal elements:

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

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

$$O^T A O = D.$$

(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 $x^T P x$ 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):

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

So by definition of the chi-square distribution,

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

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,

$$I = I^2 = \sum\nolimits_i P_i^2 + \sum\nolimits_{i<j} P_i P_j = \sum\nolimits_i P_i + \sum\nolimits_{i<j} P_i P_j.$$