pfsl15.tex

Lecture 15. 13.11.2012

Taking the trace,

$$n = \sum n_i + \sum_{i<j} trace(P_i P_j) = n + \sum_{i<j} trace(P_i P_j) :$$

$$\sum_{i<j} trace(P_i P_j) = 0.$$

Now

$$\begin{aligned} trace(P_i P_j) &= trace(P_i^2 P_j^2) \qquad (P_i,\ P_j \text{ projections}) \\ &= trace((P_j P_i).(P_i P_j)) \qquad (trace(AB) = trace(BA)) \\ &= trace((P_i P_j)^T.(P_i P_j)) \quad ((AB)^T = B^T A^T;\ P_i,\ P_j \text{ symmetric}) \\ &\geq 0, \end{aligned}$$

since for a matrix M

$$\begin{aligned} trace(M^T M) &= \sum_i (M^T M)_{ii} \\ &= \sum_i \sum_j (M^T)_{ij} (M)_{ji} \\ &= \sum_i \sum_j m_{ij}^2 \\ &\geq 0. \end{aligned}$$

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 $P_i P_j$ must be zero. So each element of $P_i P_j$ is zero, so matrix $P_i P_j$ is zero:

$$P_i P_j = 0 \qquad (i \neq j).$$

This is the condition that the *linear forms* $P_1 x, \ldots, P_k x$ be independent (below). Since the $P_i x$ are independent, so are the $(P_i x)^T (P_i x) = x^T P_i^T P_i x$, i.e. $x^T P_i x$ as P_i is symmetric and idempotent. That is, the *quadratic forms* $x^T P_1 x, \ldots, x^T P_k \vec{x}$ are also independent.

We now have

$$x^T x = x^T P_1 x + \ldots + x^T P_k x.$$

The left is $\sigma^2 \chi^2(n)$; the ith term on the right is $\sigma^2 \chi^2(n_i)$.

We summarise our conclusions.

Theorem (Chi-Square Decomposition Theorem). If

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

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 $Q_i := x^T P_i x$ 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 $\mathbf{X} = (X_1, \cdots, X_n)^T$, one needs
(i) a *mean vector* $\mu = (\mu_1, \cdots, \mu_n)^T$ with $\mu_i = EX_i$, $\mu = E[X]$;
(ii) a *covariance matrix* $\Sigma = (\sigma_{ij})$, with $\sigma_{ij} = cov(X_i, X_j)$: $\Sigma = cov(X)$.

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 $\Sigma_Y = A\Sigma_X A^T$.

Proof. (i) This is just linearity of the expectation operator E: $Y_i = \sum_j a_{ij} X_j + b_i$, so

$$EY_i = \sum_j a_{ij} EX_j + b_i = \sum_j a_{ij} \mu_j + b_i,$$

for each i. In vector notation, this is $\mu_{\mathbf{Y}} = A\mu + \beta$.

(ii) $Y_i - EY_i = \sum_k a_{ik}(X_k - EX_k) = \sum_k a_{ik}(X_k - \mu_k)$, so

$$cov(Y_i, Y_j) = E[\sum_r a_{ir}(X_r - \mu_r) \sum_s a_{js}(X_s - \mu_s)] = \sum_{rs} a_{ir} a_{js} E[(X_r - \mu_r)(X_s - \mu_s)]$$

$$= \sum_{rs} a_{ir} a_{js} \sigma_{rs} = (A\Sigma A^T)_{ij},$$

identifying the elements of the matrix product $A\Sigma A^T$. //

Corollary. Covariance matrices Σ are non-negative definite.

Proof. Let a be any $n \times 1$ matrix (row-vector of length n); then $Y := aX$ is a scalar. So $Y = Y^T = Xa^T$. Taking $a = A^T, b = 0$ above, Y has variance [= 1×1 covariance matrix] $a^T \Sigma a$. But variances are non-negative. So $a^T \Sigma a \geq 0$ for all n-vectors a. This says that Σ is non-negative definite. //

We turn now to a technical result, which is important in reducing n-dimensional problems to one-dimensional ones.

Theorem (Cramér-Wold device). The distribution of a random n-vector X is completely determined by the set of all one-dimensional distributions of linear combinations $t^T X = \sum_i t_i X_i$, where t ranges over all fixed n-vectors.

Proof. $Y := t^T X$ has CF

$$\phi_Y(s) := E[\exp\{isY\}] = E[\exp\{ist^T X\}].$$

If we know the distribution of each Y, we know its CF $\phi_Y(s)$. In particular, taking $s = 1$, we know $E[\exp\{it^T X\}]$. But this is the CF of $X = (X_1, \cdots, X_n)^T$ evaluated at $t = (t_1, \cdots, t_n)^T$. But this determines the distribution of X. //

The Cramér-Wold device suggests a way to *define* the multivariate normal distribution. The definition below seems indirect, but it has the advantage of handling the full-rank and singular cases together ($\rho = \pm 1$ as well as $-1 < \rho < 1$ for the bivariate case).

Definition. An n-vector X has an *n-variate normal* (or *Gaussian*) distribution iff $a^T X$ is univariate normal for all constant n-vectors a.

Proposition. (i) Any linear transformation of a multinormal n-vector is multinormal;
(ii) Any vector of elements from a multinormal n-vector is multinormal.
In particular, the components are univariate normal.

Proof. (i) If $y = AX + c$ (A an $m \times n$ matrix, c an m-vector) is an m-vector, and b is any m-vector,

$$b^T Y = b^T(AX + c) = (b^T A)X + b^T c.$$

If $a = A^T b$ (an m-vector), $a^T X = b^T AX$ is univariate normal as X is multinormal. Adding the constant $b^T c$, $b^T Y$ is univariate normal. This holds for all b, so Y is m-variate normal.
(ii) Take a suitable matrix A of 1s and 0s to choose the required sub-vector. //

Theorem. If X is n-variate normal with mean μ and covariance matrix Σ, its CF is

$$\phi(t) := E[\exp\{it^T X\} = \exp\{it^T \mu - \frac{1}{2}t^T \Sigma t\}.$$

Proof. By the Proposition, $Y := t^T X$ has mean $t^T \mu$ and variance $t^T \Sigma t$. By definition of multinormality, $Y = t^T X$ is univariate normal. So Y is $N(t^T \mu, t^T \Sigma t)$. So Y has CF

$$\phi_Y(s) := E[\exp\{isY\}] = \exp\{ist^T \mu - \frac{1}{2}t^T \Sigma t\}.$$

But $E[(e^{isY})] = E[\exp\{ist^T X\}]$, so taking $s = 1$ (as in the proof of the Cramér-Wold device),

$$E[\exp\{it^T X\} = \exp\{it^T \mu - \frac{1}{2}t^T \Sigma t\},$$

giving the CF of X as required. //