pfsl15(14).tex
Lecture 15. 11.11.2014 (half-hour – Problems)

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 [$= 1 \times 1$ 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) gives the CF of X as required:

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