smfl4.tex
Lecture 4. 24.1.2011

Proof. Write $\mathbf{Y} := \mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{X}$ ($\mathbf{\Sigma}^{-\frac{1}{2}}$ exists as $\mathbf{\Sigma} > \mathbf{0}$, by above). Then $\mathbf{Y}$ has covariance matrix $\mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{\Sigma}(\mathbf{\Sigma}^{-\frac{1}{2}})^T$. Since $\mathbf{\Sigma} = \mathbf{\Sigma}^T$ and $\mathbf{\Sigma} = \mathbf{\Sigma}^{\frac{1}{2}}\mathbf{\Sigma}^{\frac{1}{2}}$, $\mathbf{Y}$ has covariance matrix $\mathbf{I}$ (the components Y_i of $\mathbf{Y}$ are uncorrelated).

Change variables as above, with $\mathbf{y} = \mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{x}$, $\mathbf{x} = \mathbf{\Sigma}^{\frac{1}{2}}\mathbf{y}$. The Jacobian is (taking $\mathbf{A} = \mathbf{\Sigma}^{-\frac{1}{2}}$) $J = \partial\mathbf{x}/\partial\mathbf{y} = det(\mathbf{\Sigma}^{\frac{1}{2}}), = (det\mathbf{\Sigma})^{\frac{1}{2}}$ by the product theorem for determinants. Substituting, the integrand is

$$\exp\{-\frac{1}{2}(\mathbf{x}-\mu)^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)\} = \exp\{-\frac{1}{2}(\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{y}-\mathbf{\Sigma}^{\frac{1}{2}}(\mathbf{\Sigma}^{-\frac{1}{2}}\mu))^T\sigma^{-1}(\sigma^{\frac{1}{2}}\mathbf{y}-\sigma^{\frac{1}{2}}(\sigma^{-\frac{1}{2}}\mu))\}.$$

Writing $\nu := \sigma^{-\frac{1}{2}}\mu$, this is

$$\exp\{-\frac{1}{2}(\mathbf{y}-\nu)^T\sigma^{\frac{1}{2}}\sigma^{-1}\sigma^{\frac{1}{2}}(\mathbf{y}-\nu)\} = \exp\{-\frac{1}{2}(\mathbf{y}-\nu)^T(\mathbf{y}-\nu)\}.$$

So by the change of density formula, $\mathbf{Y}$ has density

$$g(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{1}{2}n}|\sigma|^{\frac{1}{2}}}.|\sigma|^{\frac{1}{2}}.\exp\{-\frac{1}{2}(\mathbf{y}-\nu)^T(\mathbf{y}-\nu)\}.$$

This factorises as

$$\Pi_{i=1}^n \frac{1}{(2\pi)^{\frac{1}{2}}}\exp\{-\frac{1}{2}(y_i-\nu_i)^2\}.$$

So the components Y_i of $\mathbf{Y}$ are independent $N(\nu_i, 1)$. So $\mathbf{Y}$ is multinormal, $N(\nu, I)$.
(i) Taking $A = B = \mathbf{R}^n$, $\int_{\mathbf{R}^n} f(\mathbf{x})d\mathbf{x} = \int_{\mathbf{R}^n} g(\mathbf{y})d\mathbf{y}, = 1$ as g is a probability density, as above. So f is also a probability density (non-negative and integrates to 1).
(ii) $\mathbf{X} = \sigma^{\frac{1}{2}}\mathbf{Y}$ is a linear transformation of $\mathbf{Y}$, and $\mathbf{Y}$ is multivariate normal, $N(\nu, I)$. So $\mathbf{X}$ is multivariate normal.
(iii) $E\mathbf{X} = \sigma^{\frac{1}{2}}E\mathbf{Y} = \sigma^{\frac{1}{2}}\nu = \sigma^{\frac{1}{2}}.\sigma^{-\frac{1}{2}}\mu = \mu$, $cov\mathbf{X} = \sigma^{\frac{1}{2}}cov\mathbf{Y}(\sigma^{\frac{1}{2}})^T = \sigma^{\frac{1}{2}}\mathbf{I}\sigma^{\frac{1}{2}} = \sigma$. So $\mathbf{X}$ is multinormal $N(\mu, \sigma)$. So its MGF is

$$M(\mathbf{t}) = \exp\{\mathbf{t}^T\mu + \frac{1}{2}\mathbf{t}^T\sigma\mathbf{t}\}. \qquad //$$

Independence of Linear Forms

Given a normally distributed random vector $\mathbf{x} \sim N(\mu, \Sigma)$ and a matrix

A, one may form the *linear form* $A\mathbf{x}$. One often encounters several of these together, and needs their joint distribution – in particular, to know when these are independent.

THEOREM 3. Linear forms $A\mathbf{x}$ and $B\mathbf{x}$ with $\mathbf{x} \sim N(\mu, \Sigma)$ are independent iff

$$A\Sigma B^T = 0.$$

In particular, if A, B are symmetric and $\Sigma = \sigma^2 I$, they are independent iff

$$AB = 0.$$

Proof. The joint MGF is

$$M(\mathbf{u}, \mathbf{v}) := E \exp\{\mathbf{u}^T A\mathbf{x} + i\mathbf{v}^T B\mathbf{x}\} = E \exp\{(A^T\mathbf{u} + B^T\mathbf{v})^T\mathbf{x}\}.$$

This is the MGF of $\mathbf{x}$ at argument $\mathbf{t} = A^T\mathbf{u} + B^T\mathbf{v}$, so

$$M(\mathbf{u}, \mathbf{v}) = \exp\{(\mathbf{u}^T A + \mathbf{v}^T B)\mu + \frac{1}{2}[\mathbf{u}^T A\Sigma A^T\mathbf{u} + \mathbf{u}^T A\Sigma B^T\mathbf{v} + \mathbf{v}^T B\Sigma A^T\mathbf{u} + \mathbf{v}^T B\Sigma B^T\mathbf{uv}]\}.$$

This factorises into a product of a function of $\mathbf{u}$ and a function of $\mathbf{v}$ iff the two cross-terms in $\mathbf{u}$ and $\mathbf{v}$ vanish, that is, iff $A\Sigma B^T = 0$ and $B\Sigma A^T = 0$; by symmetry of Σ, the two are equivalent.

4. ESTIMATION THEORY FOR THE MULTIVARIATE NORMALl.

Given a sample $x_1, \ldots, x_n$ from the multivariate normal $N_p(\mu, \Sigma)$, form the *sample mean* (vector)

$$\bar{x} := \frac{1}{n}\sum_{i=1}^{n} x_i,$$

as in the one-dimensional case, and the *sample covariance matrix*

$$S := \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^T(x_i - \bar{x}).$$

The likelihood for a sample of size 1 is

$$L(x|\mu, \Sigma) = (2\pi)^{-p/2}|\Sigma|^{-1/2} \exp\{-\frac{1}{2}(x - \mu)^T\Sigma^{-1}(x - \mu)\},$$

so the likelihood for a sample of size n is

$$L = (2\pi)^{-np/2}|\Sigma|^{-n/2}\exp\{-\frac{1}{2}\sum_1^n (x_i-\mu)^T\Sigma^{-1}(x_i-\mu)\}.$$

Writing

$$x_i - \mu = (x_i - \bar{x}) - (\mu - \bar{x}),$$

$$\sum_1^n (x_i-\mu)^T\Sigma^{-1}(x_i-\mu) = \sum_1^n (x_i-\bar{x})^T\Sigma^{-1}(x_i-\bar{x}) + n(\bar{x}-\mu)^T\Sigma^{-1}(\bar{x}-\mu)$$

(the cross-terms cancel as $\sum(x_i - \bar{x}) = 0$). The summand in the first term on the right is a scalar, so is its own trace. Since $trace(AB) = trace(BA)$ and $trace(A+B) = trace(B+A)$,

$$trace(\sum_1^n (x_i-\bar{x})^T\Sigma^{-1}(x_i-\bar{x})) = trace(\Sigma^{-1}\sum_1^n (x_i-\bar{x})^T(x_i-\bar{x}))$$

$$= trace(\Sigma^{-1}.nS) = n\ trace(\Sigma^{-1}S).$$

Combining,

$$L = (2\pi)^{-np/2}|\Sigma|^{-n/2}\exp\{-\frac{1}{2}n\ trace(\Sigma^{-1}S) - \frac{1}{n}n(\bar{x}-\mu)^T\Sigma^{-1}(\bar{x}-\mu)\}.$$

This involves the data only through $\bar{x}$ and S. We expect the sample mean $\bar{x}$ to be informative about the population mean μ and the sample covariance matrix S to be informative about the population covariance matrix S. In fact $\bar{x}$, S are fully informative about μ, Σ, in a sense that can be made precise using the theory of *sufficient statistics* (for which we must refer to a good book on statistical inference – see e.g. Casella and Berger [CB], Ch. 6, or III.5 below). These natural estimators are in fact the maximum likelihood estimators (Introductory Lectures in Statistics):

Theorem. For the multivariate normal $N_p(\mu, \Sigma)$, $\bar{x}$ and S are the maximum likelihood estimators for μ, Σ.

Proof. Write $V = (v_{ij}) := \Sigma^{-1}$. By above, the likelihood is

$$L = const.|V|^{n/2}\exp\{-\frac{1}{2}n\ trace(VS) - \frac{1}{2}n(\bar{x}-\mu)^T V(\bar{x}-\mu)\},$$

so the log-likelihood is

$$\ell = c + \frac{1}{2} n \log |V| - \frac{1}{2} n \; trace(VS) - \frac{1}{2} n (\bar{x} - \mu)^T V (\bar{x} - \mu).$$

The MLE $\hat{\mu}$ for μ is $\bar{x}$, as this reduces the last term (the only one involving μ) to its minimum value, 0. For a square matrix $A = (a_{ij})$, its determinant is

$$|A| = \sum_j a_{ij} A_{ij}$$

for each i, or

$$|A| = \sum_j a_{ij} A_{ij}$$

for each j, expanding by the ith row or jth column, where A_{ij} is the *cofactor* (signed minor) of a_{ij}. From either,

$$\partial |A| / \partial a_{ij} = A_{ij},$$

so

$$\partial \log |A| / \partial a_{ij} = A_{ij} / |A| = (A^{-1})_{ji},$$

the (j, i) element of A^{-1}, recalling the formula for the matrix inverse (or $(A^{-1})_{ij}$ if A is symmetric). Also, if B is symmetric,

$$trace(AB) = \sum_i \sum_j a_{ij} b_{ji} = \sum_{i,j} a_{ij} b_{ij},$$

so

$$\partial trace(AB) / \partial a_{ij} = b_{ij}.$$

Using these, and writing $S = (s_{ij})$,

$$\partial \log |V| / \partial v_{ij} = (V^{-1})_{ij} = (\Sigma)_{ij} = \sigma_{ij} \qquad (V := \Sigma^{-1}),$$

$$\partial trace(VS) / \partial v_{ij} = s_{ij}.$$

So

$$\partial \ell / \partial v_{ij} = \frac{1}{2} n (\sigma_{ij} - s_{ij}),$$

which is 0 for all i and j iff $\Sigma = S$. This says that S is the MLE for Σ, as required. //