pfsl16.tex

Lecture 16. 14.11.2012

Corollary. The components of X are independent iff Σ is diagonal – that is, iff the components are uncorrelated. So in the Gaussian case, 'independent' is the same as 'uncorrelated'.

Proof. The components are independent iff the joint CF factors into the product of the marginal CFs. This factorization takes place, into $\Pi_i \exp\{\mu_i t_i - \frac{1}{2}\sigma_{ii} t_i^2\}$, in the diagonal case only. //

Note. Recall that we need random variables to be in L_2 (square-integrable) for their variances and covariances to be defined. Then, 'independent' implies 'uncorrelated': if X, Y are independent,

$$cov(X,Y) := E[(X - E[X])(Y - E[Y])] = E[X - E[X]].E[Y - E[Y]] = 0,$$

by the Multiplication Theorem. But the converse is far from true in general. For example, if

$$U \sim U(0,1), \qquad X := \cos 2\pi U, \qquad Y := \sin 2\pi U,$$

then $E[X] = \int_0^{2\pi} \cos u du = 0$, $E[Y] = 0$ similarly, and $E[XY] = \int_0^{2\pi} \cos 2\pi u \sin 2\pi u du = \frac{1}{2}\int_0^{2\pi} \sin 4\pi u du = 0$. So X, Y are uncorrelated. But they are very heavily dependent: knowing an angle's sine, the angle is determined to within two values, and thus its cosine is also.

This identification of 'independent' with 'uncorrelated' is a very special, and very useful, property of normality/Gaussianity.

Recall that a covariance matrix Σ is always (i) symmetric: ($\sigma_{ij} = \sigma_{ji}$, as $\sigma_{ij} = cov(X_i, X_j)$);
(ii) non-negative definite: $a^T \Sigma a \geq 0$ for all n-vectors a.
Suppose that Σ is, further, *positive definite*:

$$a^T \Sigma a > 0 \qquad \text{unless} \qquad a = 0.$$

[We write $\Sigma > 0$ for 'Σ is positive definite', $\Sigma \geq 0$ for 'Σ is non-negative definite'.]

Recall from Linear Algebra that λ is an *eigenvalue* of a matrix A with *eigenvector* x ($\neq 0$) if

$$Ax = \lambda x$$

(x is *normalized* if $x^T x = \Sigma_i x_i^2 = 1$, as is always possible), and
(i) a symmetric matrix has all its eigenvalues real;
(ii) a symmetric non-negative definite matrix has all its eigenvalues non-negative;
(iii) a symmetric positive definite matrix is non-singular (has an inverse), and has all its eigenvalues positive.

We quote

Theorem (Spectral Decomposition). If A is a symmetric matrix, A can be written

$$A = \Gamma\Lambda\Gamma^T,$$

where Λ is a diagonal matrix of eigenvalues of A, Γ is an orthogonal matrix whose columns are normalized eigenvectors.

Corollary. (i) For Σ a covariance matrix, we can define its *square root* matrix $\Sigma^{\frac{1}{2}}$ by $\Sigma^{\frac{1}{2}} := \Gamma\Lambda^{\frac{1}{2}}\Gamma^T$, $\Lambda^{\frac{1}{2}} := diag(\lambda_i^{\frac{1}{2}})$, with $\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}} = \Sigma$.

For Σ a non-singular (i.e. positive definite) covariance matrix, we can define its *inverse square root* matrix $\Sigma^{-\frac{1}{2}}$ by

$$\Sigma^{-\frac{1}{2}} := \Gamma\Lambda^{-\frac{1}{2}}\Gamma^T, \qquad \Lambda^{-\frac{1}{2}} := diag(\lambda^{-\frac{1}{2}}), \qquad \text{with} \qquad \Lambda^{-\frac{1}{2}}\Lambda^{-\frac{1}{2}} = \Lambda^{-1}.$$

Theorem. If X_i are independent (univariate) normal, any linear combination of the X_i is normal. That is, $X = (X_1, \cdots, X_n)^T$, with X_i independent normal, is multinormal.

Proof. If X_i are independent $N(\mu_i, \sigma_i^2) \qquad (i = 1, \cdots, n)$, $Y := \sum_i a_i X_i + c$ is a linear combination, Y has CF

$$\begin{aligned}
\phi_Y(t) &:= E[\exp\{it(c + \textstyle\sum_i a_i X_i)\}] \\
&= e^{tc} E[\Pi \exp\{ita_i X_i\}] \qquad \text{(property of exponentials)} \\
&= e^{itc} \Pi E[\exp\{ita_i X_i\}] \qquad \text{(independence)} \\
&= e^{itc} \Pi \exp\{\mu_i i(a_i t) - \frac{1}{2}\sigma_i^2 (a_i t)^2\} \qquad \text{(normal CF)} \\
&= \exp\{i[c + \textstyle\sum_i a_i \mu_i]t - \frac{1}{2}[\textstyle\sum_i a_i^2 \sigma_i^2] t^2,
\end{aligned}$$

so Y is $N(c + \sum_i a_i \mu_i, \sum_i a_i^2 \sigma_i^2)$, from its CF. //

The Multinormal Density

If X is n-variate normal, $N(\mu, \mathbf{\Sigma})$, its density (in n dimensions) need not exist (e.g. the singular case $\rho = \pm 1$ with $n = 2$ of the bivariate normal – see e.g. [BF], 1.5). But if $\Sigma > 0$ (so Σ^{-1} exists), X has a density. The link between the multinormal density below and the multinormal CF above is due to the English statistician F. Y. Edgeworth (1845-1926).

Theorem (Edgeworth, 1893). If μ is an n-vector, $\Sigma > 0$ a symmetric positive definite $n \times n$ matrix, then (i)

$$f(x) := \frac{1}{(2\pi)^{\frac{1}{2}n}|\Sigma|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\}$$

is an n-dimensional probability density function (of a random n-vector X, say);
(ii) X has CF $\phi(t) = \exp\{it^T\mu - \frac{1}{2}t^T\Sigma t\}$; (iii) X is multinormal $N(\mu, \Sigma)$.

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

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

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

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

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

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

$$g(y) = \frac{1}{(2\pi)^{\frac{1}{2}n}|\Sigma|^{\frac{1}{2}}}.|\Sigma|^{\frac{1}{2}}.\exp\{-\frac{1}{2}(y-\nu)^T(y-\nu)\}, \quad = \prod_{i=1}^{n}\frac{1}{(2\pi)^{\frac{1}{2}}}\exp\{-\frac{1}{2}(y_i-\nu_i)^2\}.$$

So the components Y_i are independent $N(\nu_i, 1)$. So Y is $N(\nu, I)$. //

Note. (i) Taking $A = B = \mathbf{R}^n$, $\int_{\mathbf{R}^n} f(x)dx = \int_{\mathbf{R}^n} g(y)dy, = 1$ as g is a probability density, as above. So f is also a probability density.

(ii) $X = \Sigma^{\frac{1}{2}} Y$ is a linear transf. of Y, so is multivariate normal as Y is.
(ii) $E[X] = \Sigma^{\frac{1}{2}} E[Y] = \Sigma^{\frac{1}{2}} \nu = \Sigma^{\frac{1}{2}}.\Sigma^{-\frac{1}{2}} \mu = \mu$, $cov(X) = \Sigma^{\frac{1}{2}} cov(Y) (\Sigma^{\frac{1}{2}})^T = \Sigma^{\frac{1}{2}} I \Sigma^{\frac{1}{2}} = \Sigma$. So X is multinormal $N(\mu, \Sigma)$. So its CF is

$$\phi(t) = \exp\{t^T \mu - \frac{1}{2} t^T \Sigma t\}. \qquad //$$

Independence of Linear Forms Given a normally distributed random vector $x \sim N(\mu, \Sigma)$ and a matrix A, one may form the *linear form* Ax. One often needs to know when such linear forms are independent.

Theorem. Linear forms Ax and Bx with $x \sim N(\mu, \Sigma)$ 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 CF is

$$\phi(u, v) := E[\exp\{iu^T A + iv^T Bx\}] = E[\exp\{i(A^T u + B^T v)^T x\}].$$

This is the CF of x at argument $t = A^T u + B^T v$, so

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

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

Independence of quadratic forms. If the matrix of a quadratic form is a symmetric projection P, then the quadratic form is

$$x^T P x = x^T P P x = (Px)^T (Px) = \|Px\|^2.$$

So the question of independence of such quadratic forms – the only ones that we shall encounter – reduces to that of linear forms Px. This is dealt with by the above. This explains why $P_i P_j = 0$ $(i \neq j)$ – the *orthogonality condition* between projections P_i, P_j – is needed, in the Chi-Square Decomposition Theorem of IV.2 (Cochran's theorem).