Multivariate Gaussian Distribution

The random vector

$$X = (X_1, X_2, \ldots, X_p)$$

is said to have a *multivariate Gaussian distribution* if the joint distribution of $X_1, X_2, \ldots, X_p$ has density

$$f_X(x_1, x_2, \ldots, x_p) = \frac{1}{(2\pi)^{p/2}\det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu)\right) \quad (1)$$

where Σ is a $p \times p$ symmetric, positive definite matrix. The notation is as follows: x is the column vector

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix},$$

μ is the column vector

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix},$$

Σ^{-1} is the inverse of the matrix Σ and ^{t} denotes matrix transposition. Thus the quantity appearing in the exponential is a $1 \times p$ matrix times a $p \times p$ matrix times a $p \times 1$ matrix; and hence, a 1×1 matrix, i.e. a real number. Explicitly

$$(x-\mu)^t\Sigma^{-1}(x-\mu) = \sum_{k,\ell=1}^{p} (x_k - \mu_k)\Sigma^{-1}_{k\ell}(x_\ell - \mu_\ell)$$

where $\Sigma^{-1}_{k\ell}$ is the (k, ℓ)th matrix element of Σ^{-1}. The constants in front of the exponential are normalization constants; that is, if (1) is integrated over $\mathbb{R}^p$ then the result equals 1. The vector μ is the mean vector since

$$E(X) = \mu$$

as an exercise in integration shows. It is convenient for theoretical purposes to center X; that is, if $E(X) \neq 0$, the replace X by $X - \mu$. From now on we assume $E(X) = 0$ in which case the multivariate Gaussian (1) becomes

$$f_X(x_1, x_2, \ldots, x_p) = \frac{1}{(2\pi)^{p/2} \det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2}\, x^t \Sigma^{-1} x\right) \tag{2}$$

Now the matrix XX^t is a $p \times p$ matrix with elements $X_i X_j$. (Note $X^t X$ is 1×1 but XX^t is $p \times p$.). One can show (by evaluating integrals) that (recall we are setting $\mu = 0$)

$$E(XX^t) = \Sigma,$$

that is, $E(X_i X_j) = \Sigma_{ij}$. The matrix Σ is called the *covariance matrix.*

Important Remark: If the covariance matrix Σ is diagonal, then the density f_X factors and the random variables are independent.

The p = 2 case

We examine the case $p = 2$ in more detail. That is we have a random vector $X = (X_1, X_2)$ whose distribution is given by (2) for $p = 2$. In this case it is customary to parametrize Σ (for reasons that will become clear) as follows:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\,\sigma_1\sigma_2 \\ \rho\,\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Since

$$\det \Sigma = \sigma_1^2 \sigma_2^2 (1 - \rho^2)$$

and $\det \Sigma > 0$ (recall Σ is positive definite), we must have

$$-1 < \rho < 1.$$

The coefficient ρ is called the *correlation coefficient* since when it equals 0 the random variables X_1 and X_2 are independent. A calculation of the inverse of Σ gives

$$\Sigma^{-1} = \begin{pmatrix} \frac{1}{\sigma_1^2(1-\rho^2)} & -\frac{\rho}{\sigma_1\sigma_2(1-\rho^2)} \\ -\frac{\rho}{\sigma_1\sigma_2(1-\rho^2)} & \frac{1}{\sigma_2^2(1-\rho^2)} \end{pmatrix}.$$

Substituting this into (2) gives the *bivariate normal density*

$$f_X(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{x_1}{\sigma_1^2} - 2\rho\frac{x_1x_2}{\sigma_1\sigma_2} + \frac{x_2^2}{\sigma_2^2}\right)\right) \tag{3}$$

Recall the integral

$$\int_{-\infty}^{\infty} \mathrm{e}^{-ax^2+2bx}\,dx = \sqrt{\frac{\pi}{a}}\,\mathrm{e}^{b^2/a},\ a > 0. \tag{4}$$

Using this it is easy to show that the marginal densities are gaussians

$$\begin{aligned} f_{X_1}(x_1) &= \int_{-\infty}^{\infty} f_X(x_1,x_2)\,dx_2 = \frac{1}{\sqrt{2\pi}\,\sigma_1}\,\mathrm{e}^{-x_1^2/(2\sigma_1^2)}\,, \\ f_{X_2}(x_2) &= \int_{-\infty}^{\infty} f_X(x_1,x_2)\,dx_1 = \frac{1}{\sqrt{2\pi}\,\sigma_2}\,\mathrm{e}^{-x_2^2/(2\sigma_2^2)}\,. \end{aligned} \tag{5}$$

We now calculate $E(X_1|X_2)$, the expected value of X_1 given X_2. To do this we first find the conditional density $f(x_1|x_2)$. It is, by definition,

$$f(x_1|x_2) = \frac{f_X(x_1,x_2)}{f_{X_2}(x_2)}$$

where f_X is given by (3) and f_{X_2} by (5). Carrying out the algebra we see the conditional density is of the form

$$f(x_1|x_2) = d\,\mathrm{e}^{-ax_1^2+2bx_1x_2-cx_2^2}$$

where

$$\begin{aligned} a &= \frac{1}{2(1-\rho^2)\sigma_1^2} \\ b &= \frac{\rho}{2(1-\rho^2)\sigma_1\sigma_2} \\ c &= \frac{\rho^2}{2(1-\rho^2)\sigma_2^2} \\ d &= \frac{1}{\sqrt{2\pi(1-\rho^2)}\,\sigma_1} \end{aligned}$$

Then

$$E(X_1|X_2) = \int_{-\infty}^{\infty} x_1 f(x_1|x_2)\, dx_1.$$

This last integral can be computed using (4) with the result that

$$E(X_1|X_2) = (\frac{\sigma_1}{\sigma_2})\, \rho X_2. \tag{6}$$

Remarks:

1. If $\rho = 0$ then X_1 and X_2 are independent and $E(X_1) = E(X_1|X_2) = 0$.
2. If $\rho > 0$ then the expected value of X_1 given X_2 is positively correlated with X_2. (Similarly, it is negatively correlated when $\rho < 0$.)
3. The conditional expectation is *linear* in X_2.