MATH3841

Session 2, 2012

Review Notes on Multivariate Distributions

William T.M. Dunsmuir

July 9, 2013

Contents

These notes will be relied on for basic facts about joint distributions of random variables as well as properties of the multivariate normal distribution essential for MATH5845.

1 Joint Distribution and Density Functions

An n-dimensional random vector

$$X = (X_1, \ldots, X_n)'$$

has joint distribution function

$$F(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$$

for all real numbers $x_1, \ldots, x_n$.

The joint distribution of any sub-vector can be obtained by setting $x_i = \infty$ for the other random variables not in the sub-vector. For example, the distribution of X_1 is given by

$$F_{X_1}(x_1) = F(x_1, \infty, \ldots, \infty)$$

and the joint distribution of (X_i, X_j) by

$$F_{X_i,X_j}(x_i, x_j) = F(\infty, \ldots, \infty, x_i, \infty, \ldots, \infty, x_j, \infty, \ldots, \infty).$$

A random vector is said to be continuous if its distribution function can be written in terms of a non-negative density function $f(\cdot, \ldots, \cdot)$ as

$$F(x_1, \ldots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f(y_1, \ldots, y_n) dy_1 \ldots dy_n$$

where

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(y_1, \ldots, y_n) dy_1 \ldots dy_n = 1.$$

Note that we can derive the density by differentiating the distribution function

$$f(x_1, \ldots, x_n) = \frac{\partial F(x_1, \ldots, x_n)}{\partial x_1 \cdots \partial x_n}.$$

2 Independence

The random variables $X_1, \ldots X_n$ are said to be independent if

$$P(X_1 \leq x_1, \ldots, X_n \leq x_n) = P(X_1 \leq x_1) \cdots P(X_n \leq x_n)$$

for all real numbers $x_1, \ldots, x_n$. This is equivalent to

$$F(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n)$$

or

$$f(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$$

3 Conditional Distributions.

Let $X = (X_1, \ldots, X_n)'$ and $Y = (Y_1 \ldots, Y_m)'$ be two random vectors with joint density $f_{X,Y}$. The conditional density of Y given $X = x$ is

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)}.$$

Note that if X and Y are independent

$$f_{X,Y}(x, y) = f_X(x) f_Y(y)$$

so that $f_{Y|X}(y|x) = f_Y(y)$ in which case knowledge of $X = x$ does not alter the probabilities assigned to outcomes for Y. Conversely, if $f_{Y|X}(y|x) = f_Y(y)$ then X and Y are independent. Similary properties hold in terms of the distribution functions.

4 Expected Values.

Let $g(X)$ be a function of the random vector X. The expected value is

$$\begin{aligned} E(g(X)) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \ldots, x_n) f(x_1, \ldots, x_n) dx_1 \ldots dx_n \\ &= \int_{-\infty}^{\infty} g(x) f(x) dx. \end{aligned}$$

The mean $\mu = E(X)$ of a random variable corresponds to setting $g(X) = X$ while the variance $\sigma^2 =$var$(X) = E(X - \mu)^2$ corresponds to setting $g(X) = (X - \mu)^2$. The linearity property of expectation is

$$E(aX + b) = aE(X) + b.$$

Note also that

$$\mathrm{var}(aX + b) = a^2 \mathrm{var}(X).$$

5 Means and Covariances for Random Vectors

The mean vector is

$$\mu_X = E(X) = (E(X_1), \ldots, E(X_n))'$$

and the covariance between X_i and X_j is

$$\mathrm{cov}(X_i, X_j) = E(X_i - \mu_i)(X_j - \mu_j).$$

The correlation is

$$\mathrm{corr}(X_i, X_j) = \frac{\mathrm{cov}(X_i, X_j)}{\sqrt{\mathrm{var}(X_i)\ \mathrm{var}(X_j)}}.$$

For two random vectors X and Y the covariance matrix between them is

$$\Sigma_{XY} = \mathrm{cov}(X, Y) = E(X - EX)(Y - EY)' = E(XY') - (EX)(EY)'$$

with (i, j) element

$$(\Sigma_{XY})_{ij} = \mathrm{cov}(X_i, Y_j).$$

When $Y = X$, $\mathrm{cov}(X, Y)$ reduces to the covariance matrix of the random vector X. Note that if X and Y are independent then the covariance between them is the null matrix. The converse is not true in general but is true for the multivariate normal distribution - see below.

Let Y and X be linearly related as $Y = a + BX$ where a is a vector and B is a matrix (all with conforming dimensions). Then

$$\mu_Y = E(Y) = a + BE(X) = a + B\mu_X$$

and

$$\Sigma_{YY} = B\Sigma_{XX}B'.$$

Note also that any covariance matrix Σ is non-negative definite, that is $b'\Sigma b \geq 0$ for any vector b. The proof of this follows from the last identity. Let $Y = b'X$ where X has covariance matrix Σ. Then

$$0 \leq \mathrm{var}(Y) = b'\Sigma b.$$

6 The Multivariate Normal Distribution.

6.1 The general multivariate normal density

The random vector X has the multivariate normal distribution with mean μ and non-singular covariance matrix Σ if

$$f_X(x) = \frac{1}{(2\pi)^{n/2}(\det \Sigma)^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right\}.$$

Notation: $X \sim N(\mu, \Sigma)$.

6.2 The Bivariate Normal Density

As special case is the bivariate normal density from which most of the required insight about the multivariate normal is obtained. Let $X = (X_1, X_2)'$ and

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

with inverse

$$\Sigma^{-1} = \frac{1}{1-\rho^2}\begin{bmatrix} \sigma_1^{-2} & -\rho\sigma_1^{-1}\sigma_2^{-1} \\ -\rho\sigma_1^{-1}\sigma_2^{-1} & \sigma_2^{-2} \end{bmatrix}$$

and $\det(\Sigma) = \sigma_1^2\sigma_2^2(1-\rho^2)$. Substitution in the general multivariate normal density gives

$$f_X(x) = \frac{1}{2\pi[\sigma_1^2\sigma_2^2(1-\rho^2)]^{1/2}} \exp\left\{-\frac{1}{2}Q(x_1, x_2; \sigma_1, \sigma_2, \rho)\right\}$$

with quadratic form

$$Q(x_1, x_2; \sigma_1, \sigma_2, \rho) = \frac{1}{(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right].$$

Some important facts about the bivariate normal are:

1. The contours of equal density are ellipses

   $$\{(x_1, x_2) : Q(x_1, x_2; \sigma_1, \sigma_2, \rho) = k\}$$

   for any constant $k \geq 0$.

2. When the correlation $\rho = 0$ the two random variables X_1 and X_2 are independent. This can easily be concluded from the form of $Q(x_1, x_2; \sigma_1, \sigma_2, \rho = 0)$. Hence for the bivariate normal distribution, independence is equivalent to uncorrelatedness.

6.3 Standardised Bivariate Normal Density

Consider a special case of the bivariate normal density for the two standardised random variables

$$U = \frac{X_1 - \mu_1}{\sigma_1}, \quad V = \frac{X_2 - \mu_2}{\sigma_2}$$

have joint normal density with

$$\mu_U = \mu_V = 0$$
$$\sigma_U^2 = \sigma_V^2 = 1$$

so that

$$f_{U,V}(u, v) = \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left(-\frac{1}{2(1-\rho^2)}[u^2 + v^2 - 2\rho uv]\right). \tag{1}$$

Recall, for any pair of continuous random variables, the conditional density of $V|U = u$ is

$$f_{V|U}(v|u) = \frac{f_{U,V}(u, v)}{f_U(u)}$$

so that the joint density can be expressed as

$$f_{U,V}(u, v) = f_U(u) f_{V|U}(v|u). \tag{2}$$

Hence if we can find a factorization of the bivariate normal density (1) in the form (2) then we have derived the marginal density of U and the conditional density of $V|U = u$.

The key to the factorization is the completion of the square in the exponent as follows

$$\begin{aligned}\frac{u^2+v^2-2\rho uv}{1-\rho^2} &= \frac{(u^2-\rho^2u^2)+(v^2-2\rho uv+\rho^2u^2)}{1-\rho^2}\\ &= \frac{u^2(1-\rho^2)+(v-\rho u)^2}{1-\rho^2}\\ &= u^2+\frac{(v-\rho u)^2}{1-\rho^2}.\end{aligned}$$

Substituting this into the exponent in equation (1) we get

$$\begin{aligned}f_{U,V}(u,v) &= \frac{1}{2\pi(1-\rho^2)^{1/2}}\exp\left(-\frac{1}{2}u^2-\frac{1}{2}\frac{(v-\rho u)^2}{1-\rho^2}\right)\\ &= \left[\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}u^2\right)\right]\left[\frac{1}{\sqrt{2\pi}(1-\rho^2)^{1/2}}\exp\left(-\frac{1}{2}\frac{(v-\rho u)^2}{1-\rho^2}\right)\right].\end{aligned}$$

Now the first factor is the standard normal $N(0,1)$ density. The second factor is a density of a $N(\rho u,(1-\rho^2))$ random variable. We can therefore identify the first factor as the marginal density for U,

$$f_U(u)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}u^2\right)$$

and the second factor as the conditional density for $V|U=u$,

$$f_{V|U}(v|u)=\frac{1}{\sqrt{2\pi}(1-\rho^2)^{1/2}}\exp\left(-\frac{1}{2}\frac{(v-\rho u)^2}{1-\rho^2}\right).$$

This proves that the marginal density for U,the standard normal density, and that the conditional density for $V|U=u$ is normal with (conditional) mean

$$E(V|U=u)=\rho u$$

and (conditional) variance

$$\mathrm{Var}(V|U=u)=1-\rho^2.$$

In summary

$$U\sim N(0,1)$$

and

$$V|u\sim N(\rho u,1-\rho^2).$$

Notes:

1. The parameter ρ is the correlation between U and V as is easily derived as follows

$$
\begin{aligned}
\operatorname{corr}(U,V) &= \frac{\operatorname{cov}(U,V)}{\sqrt{\operatorname{var}(U)\operatorname{var}(V)}} \\
&= \operatorname{cov}(U,V), \quad \text{(since } U \text{ and } V \text{ have unit variance)} \\
&= \int\int uv f_{U,V}(u,v)dvdu \\
&= \int\int uv f_U(u) f_{V|U}(v|u)dvdu \\
&= \int u f_U(u) \left[\int v f_{V|U}(v|u)dv\right] du
\end{aligned}
$$

But $\int v f_{V|U}(v|u)dv$ is the mean value for a random variable with the $N(\rho u, 1-\rho^2)$ density. Hence $\int v f_{V|U}(v|u)dv = \rho u$. Substituting this in the double integral we get

$$
\begin{aligned}
\operatorname{corr}(U,V) &= \int u f_U(u)[\rho u]du \\
&= \rho \int u^2 f_U(u)du.
\end{aligned}
$$

But $\int u^2 f_U(u)du$ is $E\left(U^2\right)$ where $U \sim N(0,1)$ and hence $\int u^2 f_U(u)du = 1$ leading to

$$
\operatorname{corr}(U,V) = \rho.
$$

This proves that for the bivariate normal density given by equation (1) the parameter ρ is the correlation between U and V.

2. When $\rho = 0$ the conditional density simplifies to

$$
f_{V|U}(v|u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}v^2\right) = f_V(v)
$$

so that U, V are independent. You can also verify independence directly by considering what happens in expression (1) when $\rho = 0$. That is

$$
\begin{aligned}
f_{U,V}(u,v) &= \frac{1}{2\pi} \exp\left(-\frac{1}{2}(u^2+v^2)\right) \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}v^2\right) \\
&= f_U(u) f_V(v).
\end{aligned}
$$

3. For both positive and negative $\rho \neq 0$, $\text{Var}(V|U=u) = 1-\rho^2 < 1$ so that use of $U = u$ information to predict V using the conditional mean $E(V|U=u) = \rho u$ will lead to lower conditional variance for this prediction of V. That is when U and V are not independent, conditioning on one improves precision of prediction of the other.

4. As $|\rho| \to 1$ so that correlation get large in absolute value, note that the conditional variance $\text{Var}(V|U=u) \to 0$ so that once $U = u$ is known V is known with certainty to be equal to u.

6.4 Properties of the General Multivariate Normal Distribution.

Some important facts about the general multivariate normal distribution are:

1. Any subvector of a multivariate normal vector has a multivariate normal distribution.

2. If $X \sim N(\mu, \Sigma)$, B is an $m \times n$ matrix of real numbers and a is a real $m \times 1$ vector then
$$Y = a + BX \sim N(a + B\mu_X, B\Sigma B').$$

3. In particular, any linear combination $b'X$ has a univariate normal distribution.

In general, consider a multivariate normal random vector $X \sim N(\mu, \Sigma)$. Partition as

$$X = \begin{bmatrix} X^{(1)} \\ X^{(2)} \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu^{(1)} \\ \mu^{(2)} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

where $\mu^{(j)} = E(X^{(j)})$ and $\Sigma_{ij} = E(X^{(i)} - \mu^{(j)})(X^{(j)} - \mu^{(j)})'$. Then

1. $X^{(1)}$ and $X^{(2)}$ are independent if and only if $\Sigma_{12} = 0$.

2. The conditional distribution of $X^{(1)}$ given $X^{(2)} = x^{(2)}$ is multivariate normal with conditional mean vector
$$E(X^{(1)}|X^{(2)} = x^{(2)}) = \mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(x^{(2)} - \mu^{(2)})$$

and covariance matrix

$$\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

This is a very important result.