APPENDIX 5

Multivariate Statistical Distributions

Multivariate Density Functions

An n-dimensional random vector $x \in \mathcal{R}$ is an ordered set of real numbers $[x_1, x_2, \ldots, x_n]'$ each of which represents some aspect of a statistical event. A scalar-valued function $F(x)$, whose value at $\phi = [\phi_1, \phi_2, \ldots, \phi_n]'$ is the probability of the event $(x_1 \leq \phi_1, x_2 \leq \phi_2, \ldots, x_n \leq \phi_n)$, is called a cumulative distribution function.

(1) If $F(x)$ has the representation

$$F(x) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f(x_1, \ldots, x_n) dx_1 \cdots dx_n,$$

which can also be written as

$$F(x) = \int_{-\infty}^{x} f(x) dx,$$

then it is said to be absolutely continuous; in which case $f(x) = f(x_1, \ldots, x_n)$ is called a continuous probability density function.

When x has the probability density function $f(x)$, it is said to be distributed as $f(x)$, and this is denoted by writing $x \sim f(x)$.

The function $f(x)$ has the following properties:

(2) (i) $f(x) \geq 0$ for all $x \in \mathcal{R}^n$.

(ii) If $\mathcal{A} \subset \mathcal{R}^n$ is a set of values for x, then the probability that x is in $\mathcal{A}$ is

$P(\mathcal{A}) = \int_{\mathcal{A}} f(x) dx.$

(iii) $P(x \in \mathcal{R}^n) = \int_x f(x) dx = 1.$

One may wish to characterise the statistical event in terms only of a subset of the elements in x. In that case, one is interested in the marginal distribution of the subset.

(3) Let the $n \times 1$ random vector $x \sim f(x)$ be partitioned such that $x' = [x_1, x_2]'$ where $x_1' = [x_1, \ldots, x_m]$ and $x_2' = [x_{m+1}, \ldots, x_n]$ Then, with $f(x) = f(x_1, x_2)$, the marginal probability density function of x_1 can be defined as

$$f(x_1) = \int_{x_2} f(x_1, x_2) dx_2,$$

which can also be written as

$$f(x_1, \ldots, x_m) \\ = \int_{x_n} \cdots \int_{x_m+1} f(x_1, \ldots, x_m, x_{m+1}, \ldots, x_n) dx_{m+1} \cdots dx_n.$$

Using the marginal probability density function, the probability that x_1 will assume a value in the set $\mathcal{B}$ can be expressed, without reference to the value of the vector x_2, as

$$P(\mathcal{B}) = \int_{\mathcal{B}} f(x_1) dx_1.$$

Next, we consider conditional probabilities.

(4) The probability of the event $x_1 \in \mathcal{A}$ given the event $x_2 \in \mathcal{B}$ is

$$P(\mathcal{A}|\mathcal{B}) = \frac{P(\mathcal{A} \cap \mathcal{B})}{P(\mathcal{B})} = \frac{\int_{\mathcal{B}} \int_{\mathcal{A}} f(x_1, x_2) dx_1 dx_2}{\int_{\mathcal{B}} f(x_2) dx_2}.$$

It is also necessary to define the probability $P(\mathcal{A}|x_2 = \phi)$ of the event $x_1 \in \mathcal{A}$ given that x_2 has the specific value ϕ. This problem can be approached by finding the limiting value of $P(\mathcal{A}|\phi < x_2 \leq \phi + \Delta x_2)$ as Δx_2 tends to zero. Defining the event $\mathcal{B} = \{x_2; \phi < x_2 \leq \phi + \Delta x_2\}$, it follows from the mean value theorem that

$$P(\mathcal{B}) = \int_{\phi}^{\phi + \Delta x_2} f(x_2) dx_2 = f(\phi^0) \Delta x_2,$$

where $\phi \leq \phi^0 \leq \phi + \Delta x_2$. Likewise, there is

$$P(\mathcal{A} \cap \mathcal{B}) = \int_{\mathcal{A}} f(x_1, \phi^*) \Delta x_2 dx_1,$$

where $\phi \leq \phi^* \leq \phi + \Delta x_2$. Thus, provided that $f(\phi^0) > 0$, it follows that

$$P(\mathcal{A}|\mathcal{B}) = \frac{\int_{\mathcal{A}} f(x_1, \phi^*)dx}{f(\phi_0)};$$

and the probability $P(\mathcal{A}|x_2 = \phi)$ can be defined as the limit this integral as Δx_2 tends to zero and both ϕ^0 and ϕ^* tend to ϕ. Thus, in general,

(5) If $x' = [x_1', x_2']$, then the conditional probability density function of x_1 given x_2 is defined as

$$f(x_1|x_2) = \frac{f(x)}{f(x_2)} = \frac{f(x_1, x_2)}{f(x_2)}.$$

Notice that the probability density function of x can now be written as $f(x) = f(x_1|x_2)f(x_2) = f(x_2|x_1)f(x_1)$.

A definition of statistical independence follows immediately:

(6) The vectors x_1, x_2 are statistically independent if their joint distribution is $f(x_1, x_2) = f(x_1)f(x_2)$ or, equivalently, if $f(x_1|x_2) = f(x_1)$ and $f(x_2|x_1) = f(x_2)$.

Functions of Random Vectors

Consider a random vector $y \sim g(y)$, which is a continuous function $y = y(x)$ of another random vector $x \sim f(x)$, and imagine that the inverse function $x = x(y)$ is uniquely defined. If $\mathcal{A}$ is a statistical event defined as a set of values of x, and if $\mathcal{B} = \{y = y(x), x \in \mathcal{A}\}$ is the same event defined in terms of y, then it follows that

$$\begin{aligned} \int_{\mathcal{A}} f(x)dx &= P(\mathcal{A}) \\ &= P(\mathcal{B}) = \int_{\mathcal{B}} g(y)dy. \end{aligned} \tag{7}$$

When the probability density function $f(x)$ is know, it should be straightforward to find $g(y)$.

For the existence of a uniquely defined inverse transformation $x = x(y)$, it is necessary and sufficient that the determinant $|\partial x/\partial y|$, known as the Jacobian, should be nonzero for all values of y; which means that it must be either strictly positive or strictly negative. The Jacobian can be used in changing the variable under the integral in (7) from x to y to give the identity

$$\int_{\mathcal{B}} f\{x(y)\} \left|\frac{dx}{dy}\right| dy = \int_{\mathcal{B}} g(y)dy.$$

Within this expression, there are $f\{x(y)\} \geq 0$ and $g(y) \geq 0$. Thus, if $|\partial x/\partial y| > 0$, the probability density function of y can be identified as $g(y) = f\{x(y)\}|\partial x/\partial y|$. However, if $|\partial x/\partial y| < 0$, then $g(y)$, defined in this way, is no longer positive. The recourse is to change the signs of the axes of y. Thus, in general, the probability density function of y is defined as $g(y) = f\{x(y)\}\|\partial x/\partial y\|$, where $\|\partial x/\partial y\|$ is the absolute value of the determinant. The result may be summarised as follows:

(8) If $x \sim f(x)$ and $y = y(x)$ is a monotonic transformation with a uniquely defined inverse $x = x(y)$, then $y \sim g(y) = f\{x(y)\}\|\partial x/\partial y\|$, where $\|\partial x/\partial y\|$ is the absolute value of the determinant of the matrix $\partial x/\partial y$ of the partial derivatives of the inverse transformation.

Even when $y = y(x)$ has no uniquely defined inverse, it is still possible to find a probability density function $g(y)$ by the above method provided that the transformation is surjective, which is to say that the range of the transformation is coextensive with the vector space within which the random vector y resides.

Imagine that x is a vector in $\mathcal{R}^n$ and that y is a vector in $\mathcal{R}^m$ where $m < n$. Then the technique is to devise an invertible transformation $q = q(x)$ where $q' = [y', z']$ comprises, in addition to the vector y, a vector z of $n - m$ dummy variables. Once the probability density function of q has been found, the marginal probability density function $g(y)$ can be obtained by a process of integration.

Expectations

(9) If $x \sim f(x)$ is a random variable, its expected value is defined by

$$E(x) = \int_x f(x)dx.$$

In determining the expected value of a variable which is a function of x, one can rely upon the probability density function of x. Thus

(10) If $y = y(x)$ is a function of $x \sim f(x)$, and if $y \sim g(y)$, then

$$E(y) = \int_y g(y)dy = \int_x y(x)f(x)dx.$$

It is helpful to define an expectations operator E, which has the following properties, amongst others:

(11) (i) If $x \geq 0$, then $E(x) \geq 0$.

(ii) If c is a constant, then $E(c) = c$.

(iii) If c is a constant and x is a random variable, then $E(cx) = cE(x)$.

(iv) $E(x_1 + x_2) = E(x_1) + E(x_2)$

(v) If x_1, x_2 are independent random variables, then $E(x_1 x_2) = E(x_1)E(x_2)$.

These are readily established from the definitions (9) and (10). Taken together, the properties (iii) and (iv) imply that

$$E(c_1 x_1 + c_2 x_2) = c_1 E(x_1) + c_2 E x_2)$$

when c_1, c_2 are constants. Thus, the expectations operator is seen to be a linear operator.

Moments of a Multivariate Distribution

Some of the more important moments of a multivariate distribution can now be defined and some of their properties can be recorded.

(12) The expected value of the element x_i of the random vector $x \sim f(x)$ is defined by

$$E(x_i) = \int_x x_i f(x) dx = \int_{x_i} x_i f(x_i) dx_i,$$

where $f(x_i)$ is the marginal distribution of x_i.

The variance of x_i is defined by

$$\begin{aligned} V(x_i) &= E\left\{[x_i - E(x_i)]^2\right\} \\ &= \int_x [x_i - E(x_i)]^2 f(x) dx = \int_{x_i} [x_i - E(x_i)]^2 f(x_i) dx_i. \end{aligned}$$

The covariance of x_i and x_j is defined as

$$\begin{aligned} C(x_i, x_j) &= E[x_i - E(x_i)][x_j - E(x_j)] \\ &= \int_x [x_i - E(x_i)][x_j - E(x_j)] f(x) dx \\ &= \int_{x_j} \int_{x_i} [x_i - E(x_i)][x_j - E(x_j)] f(x_i, x_j) dx_i dx_j, \end{aligned}$$

where $f(x_i, x_j)$ is the marginal distribution of x_i and x_j.

The expression for the covariance can be expanded to give $C(x_i, x_j) = E[x_i x_j - E(x_i)x_j - E(x_j)x_i + E(x_i)E(x_j)] = E(x_i x_j) - E(x_i)E(x_j)$. By setting $x_j = x_i$, a similar expression is obtained for the variance $V(x_i) = C(x_i, x_i)$. Thus

(13)
$$\begin{aligned} C(x_i, x_j) &= E(x_i x_j) - E(x_i)E(x_j), \\ V(x_i) &= E(x_i^2) - [E(x_i)]^2. \end{aligned}$$

The property of the expectations operator given under (11)(i) implies that $V(x_i) \geq 0$. Also, by applying the property under (11)(v) to the expression for $C(x_i, x_j)$, it can be deduced that

(14) If x_i, x_j are independently distributed, then $C(x_i, x_j) = 0$.

Another important result is that

(15)
$$V(x_i + x_j) = V(x_i) + V(x_j) + 2C(x_i, x_j).$$

This comes from expanding the final expression in

$$\begin{aligned} V(x_i + x_j) &= E\left\{[(x_i + x_j) - E(x_i + x_j)]^2\right\} \\ &= E\left([x_i - E(x_i)] + [x_j - E(x_j)]^2\right). \end{aligned}$$

It is convenient to assemble the expectations, variances and covariances of a multivariate distribution into matrices.

(16) If $x \sim f(x)$ is an $n \times 1$ random vector, then its expected value

$$E(x) = [E(x_1), \ldots, E(x_n)]'$$

is a vector comprising the expected values of the n elements. Its dispersion matrix

$$\begin{aligned} D(x) &= E\{[x - E(x)][x - E(x)]'\} \\ &= E(xx') - E(x)E(x') \end{aligned}$$

is a symmetric $n \times n$ matrix comprising the variances and covariances of its elements. If x is partitioned such that $x' = [x_1', x_2']$, then the covariance matrix

$$\begin{aligned} C(x_1, x_2) &= E\{[x_1 - E(x_1)][x_2 - E(x_2)]'\} \\ &= E(x_1 x_2') - E(x_1)E(x_2') \end{aligned}$$

is a matrix comprising the covariances of the two sets of elements.

The dispersion matrix is nonnegative definite. This is confirmed via the identity $a'D(x)a = a'\{E[x - E(x)][x - E(x)]'\}a = E\{[a'x - E(a'x)]^2\} = V(a'x) \geq 0$, which reflects the fact that variance of any scalar is nonnegative. The following are some of the properties of the operators:

(17) If x, y, z are random vectors of appropriate orders, then

(i) $E(x + y) = E(x) + E(y)$,

(ii) $D(x + y) = D(x) + D(y) + C(x, y) + C(y, x)$,

(iii) $C(x + y, z) = C(x, z) + C(y, z)$.

Also

(18) If x, y are random vectors and A, B are matrices of appropriate orders, then

(i) $E(Ax) = AE(x)$,

(ii) $D(Ax) = AD(x)A'$,

(iii) $C(Ax, By) = AC(x, y)B'$.

The Multivariate Normal Distribution

The $n \times 1$ random vector x is normally distributed with a mean $E(x) = \mu$ and a dispersion matrix $D(x) = \Sigma$ if its probability density function is

$$N(x; \mu, \Sigma) = (2\pi)^{-n/2}|\Sigma|^{-1/2} \exp\left\{-\tfrac{1}{2}(x - \mu)'\Sigma^{-1}(x - \mu)\right\}. \tag{19}$$

It is understood that x is nondegenerate with $\operatorname{rank}(\Sigma) = n$ and $|\Sigma| \neq 0$. To denote that x has this distribution, we can write $x \sim N(\mu, \Sigma)$.

Two notable features of the normal distribution will be demostrated. The first feature is that the conditional and marginal distributions associated with a normally distributed vector are also normal. The second is that any linear function of a normally distributed vector is itself normally distributed. The arguments will be based on two fundamental facts. The first fact is that

(20) If $x \sim N(\mu, \Sigma)$ and if $y = A(x - b)$, where A is nonsingular, then $y \sim N\{A(\mu - b), A\Sigma A'\}$.

This may be illustrated by considering the case where $b = 0$. Then, according to the result in (8), y has the distribution

$$
\begin{aligned}
&N(A^{-1}y;\mu,\Sigma)\|\partial x/\partial y\| \\
(21)\qquad &\quad = (2\pi)^{-n/2}|\Sigma|^{-1/2}\exp\left\{-\tfrac{1}{2}(A^{-1}y-\mu)'\Sigma^{-1}(A^{-1}y-\mu)\right\}\|A^{-1}\| \\
&\quad = (2\pi)^{-n/2}|A\Sigma A'|^{-1/2}\exp\left\{-\tfrac{1}{2}(y-A\mu)'(A\Sigma A')^{-1}(y-A\mu)\right\};
\end{aligned}
$$

so, clearly, $y \sim N(A\mu, A\Sigma A')$.

The second of the fundamental facts is that

(22) If $z \sim N(\mu, \Sigma)$ can be written in partitioned form as

$$
\begin{bmatrix} y \\ x \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_y \\ \mu_x \end{bmatrix}, \begin{bmatrix} \Sigma_{yy} & 0 \\ 0 & \Sigma_{xx} \end{bmatrix}\right),
$$

then $y \sim N(\mu_y, \Sigma_{yy})$ and $x \sim N(\mu_x, \Sigma_{xx})$ are independently distributed normal variates.

This can be seen by considering the quadratic form

$$
(23)\qquad (z-\mu)'\Sigma^{-1}(z-\mu) = (y-\mu_y)'\Sigma_{yy}^{-1}(y-\mu_y) + (x-\mu_x)'\Sigma_{xx}^{-1}(x-\mu_x)
$$

which arises in this particular case. Substituting the RHS into the expression for $N(z;\mu,\Sigma)$ in (19) and using $|\Sigma| = |\Sigma_{yy}||\Sigma_{xx}|$, gives

$$
\begin{aligned}
&N(z;\mu,\Sigma) = (2\pi)^{-m/2}|\Sigma_{yy}|^{-1/2}\exp\left\{-\tfrac{1}{2}(y-\mu_y)'\Sigma_{yy}^{-1}(y-\mu_y)\right\} \\
(24)\qquad &\quad \times (2\pi)^{(m-n)/2}|\Sigma_{xx}|^{-1/2}\exp\left\{-\tfrac{1}{2}(x-\mu_x)'\Sigma_{xx}^{-1}(x-\mu_x)\right\} \\
&= N(y;\mu_y,\Sigma_{yy}) \times N(x;\mu_x,\Sigma_{xx}).
\end{aligned}
$$

The latter can only be the product of the marginal distributions of y and x, which proves that these vectors are independently distributed.

The essential feature of the result is that

(25) If y and x are normally distributed with $C(y, x) = 0$, then they are mutually independent.

A zero covariance does not generally imply statistical independence.

Even when y, x are not independently distributed, their marginal distributions are still formed in the same way from the appropriate components of μ and Σ. This is entailed in the first of the two main results which is that

(26) If $z \sim N(\mu, \Sigma)$ is partitioned as

$$
\begin{bmatrix} y \\ x \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_y \\ \mu_x \end{bmatrix}, \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix}\right),
$$

then the marginal distribution of x is $N(\mu_x \Sigma_{xx})$ and the conditional distribution of y given x is

$$N(y|x; \mu_y + \Sigma_{yx}\Sigma_{xx}^{-1}(x - \mu_x), \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}).$$

Proof. The joint distribution of x and y can be factored as the product of the marginal distribution of x and the conditional distribution of y given x:

$$N(y, x) = N(y|x)N(x). \tag{27}$$

The following notation may be adopted:

$$z = \begin{bmatrix} y - E(y) \\ x - E(x) \end{bmatrix}, \qquad w = \begin{bmatrix} y - E(y|x) \\ x - E(x) \end{bmatrix} = \begin{bmatrix} \varepsilon \\ x - E(x) \end{bmatrix}. \tag{28}$$

Then, the mapping from z to $w = Qz$ may be represented by

$$\begin{bmatrix} \varepsilon \\ x - E(x) \end{bmatrix} = \begin{bmatrix} I & -B' \\ 0 & I \end{bmatrix} \begin{bmatrix} y - E(y) \\ x - E(x) \end{bmatrix}, \tag{29}$$

wherein

$$\varepsilon = y - E(y|x) = y - E(y) - B'\{x - E(x)\}. \tag{30}$$

The following dispersion matrices are defined:

$$D(z) = \Sigma_{zz} = \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix}, \qquad D(w) = \Sigma_{zz} = \begin{bmatrix} \Sigma_{\varepsilon\varepsilon} & 0 \\ 0 & \Sigma_{xx} \end{bmatrix}. \tag{31}$$

The off-diaogonal blocks of $D(w)$, which are $C\{y - E(y|x), x\} = 0$ and its transpose, bear witness to the fact that the prediction error $\varepsilon = y - E(y|x)$ is uncorrelated with x, which is the instrument of the prediction.

The quadratic exponent of the joint distribution of x and y may be expressed in terms either of z or w. Thus, $z'\Sigma_{zz}^{-1}z = w'\Sigma_{ww}^{-1}w = z'Q'\Sigma_{ww}^{-1}Qz$, which indicates that $\Sigma_{zz} = Q^{-1}\Sigma_{ww}Q'^{-1}$. This is written more explicitly as

$$\begin{aligned} \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix} &= \begin{bmatrix} I & B' \\ 0 & I \end{bmatrix} \begin{bmatrix} \Sigma_{\varepsilon\varepsilon} & 0 \\ 0 & \Sigma_{xx} \end{bmatrix} \begin{bmatrix} I & 0 \\ B & I \end{bmatrix} \\ &= \begin{bmatrix} \Sigma_{\varepsilon\varepsilon} + B'\Sigma_{xx}B & B'\Sigma_{xx} \\ \Sigma_{xx}B & \Sigma_{xx} \end{bmatrix}. \end{aligned} \tag{32}$$

The equation is solved for

$$B = \Sigma_{xx}^{-1}\Sigma_{xy} \qquad \text{and} \qquad \Sigma_{\varepsilon\varepsilon} = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}. \tag{33}$$

Therefore, the joint density function of x and $y|x$ can be written as

$$N(x;\mu_x,\Sigma_{xx})N(y|x;\mu_y-\Sigma_{yx}\Sigma_{xx}^{-1}\mu_x,\Sigma_{yy}-\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}). \tag{34}$$

Integrating the conditional distribution $N(y|x)$ with respect to x gives the marginal distribution $N(y;\mu_y,\Sigma_{yy})$.

The linear function

$$\begin{aligned} E(y|x)&=E(y)+C(y,x)D^{-1}(x)\{x-E(x)\}\\ &=\mu_y+\Sigma_{yx}\Sigma_{xx}^{-1}(x-\mu_x), \end{aligned} \tag{35}$$

which defines the expected value of x for given values of y, is described as the regression of y on x. The matrix $B=\Sigma_{xx}^{-1}\Sigma_{xy}$ is the matrix of the regression coefficients.

Now that the general the form of the marginal distribution has been established, it can be shown that any nondegenerate random vector which represents a linear function of a normal vector is itself normally distributed. To this end, it can be proved that

(36) If $x\sim N(\mu,\Sigma)$ and $y=B(x-b)$ where $\text{null}(B')=0$ or, equivalently, B has full row rank, then $y\sim N(B(\mu-b),B\Sigma B')$.

Proof. If B has full row rank, then there exists a nonsingular matrix $A'=[B',C']$ such that

$$q=\begin{bmatrix} y\\ z\end{bmatrix}=\begin{bmatrix} B\\ C\end{bmatrix}(x-b). \tag{37}$$

Then q has the distribution $N(q;A(\mu-b),A\Sigma A')$ where

$$A(\mu-b)=\begin{bmatrix} B(\mu-b)\\ C(\mu-b)\end{bmatrix},\qquad A\Sigma A'=\begin{bmatrix} B\Sigma B' & B\Sigma C'\\ C\Sigma B' & C\Sigma C'\end{bmatrix}. \tag{38}$$

It follows from (27) that y has the marginal distribution

$$N\{B(\mu-b),B\Sigma B'\}. \tag{39}$$

Distributions Associated with the Normal Distribution

(40) Let $\eta\sim N(0,I)$ be an $n\times 1$ vector of independently and identically distributed normal variates $\eta_i\sim N(0,1);i=1,\ldots,n$. Then $\eta'\eta$

has a chi-square distribution of n degrees of freedom denoted by $\chi^2(n)$.

The cumulative chi-square distribution is tabulated in most statistics textbooks; typically for degrees of freedom from $n = 1$ to $n = 30$. One need not bother, at present. with the formula for the density function; but it should be noted that, if $w \sim \chi^2(n)$, then $E(w) = n$ and $V(w) = 2n$.

(41) Let $x \sim N(0,1)$ be a standard normal variate, and let $w \sim \chi^2(n)$ be a chi-square variate of n degrees of freedom. Then the ratio $t = x/\sqrt{w/n}$ has a t distribution of n degrees of freedom denoted by $t(n)$.

The t distribution, which is perhaps the most important of the sampling distributions, is also extensively tabulated. Again, we shall not give the formula for the density function; but we may note that the distribution is symmetrical and that $E(t) = 0$ and $V(t) = n/(n-2)$. The distribution $t(n)$ approaches the standard normal $N(0,1)$ as n tends to infinity. This results from the fact that, as n tends to infinity, the distribution of the denominator in the ratio defining the t variate becomes increasingly concentrated around the value of unity, with the effect that the variate is dominated by its numerator. Finally,

(42) Let $w_1 \sim \chi^2(n)$ and $w_2 \sim \chi^2(m)$ be independently distributed chi-square variates of n and m degrees of freedom respectively. Then $F = \{(w_1/n)/(w_2/m)\}$ has an F distribution of n and m degrees of freedom denoted by $F(n,m)$.

We may record that $E(F) = m/(m-2)$ and $V(F) = 2m^2[1 + (m-2)/n]/(m-2)^2(m-4)$.

It should be recognised that

(43) If $t \sim t(n)$, then $t^2 \sim F(1,n)$.

This follows from (30) which indicates that $t^2 = \{(x^2/1)/(w/n)\}$, where $w \sim \chi^2(n)$ and $x^2 \sim \chi^2(1)$, since $x \sim N(0,1)$.

Quadratic Functions of Normal Vectors

Next, we shall establish a number of specialised results concerning quadratic functions of normally distributed vectors. The standard notation for the dispersion of the random vector ε now becomes $D(\varepsilon) = Q$. When it is important to know that the random vector $\varepsilon \sim N(0,Q)$ has the order $p \times 1$, we shall write $\varepsilon \sim N_p(0,Q)$.

We begin with some specialised results concerning the standard normal distribution $N(\eta; 0, I)$.

(44) If $\eta \sim N(0, I)$ and C is an orthonormal matrix such that $C'C = CC' = I$, then $C'\eta \sim N(0, I)$.

This is a straightforward specialisation of the basic result in (36). More generally,

(45) If $\eta \sim N_n(0, I)$ is an $n \times 1$ vector and C is an $n \times r$ matrix of orthonormal vectors, where $r \leq n$, such that $C'C = I_r$, then $C'\eta \sim N_r(0, I)$.

Occasionally, it is necessary to transform a nondegenerate vector $\varepsilon \sim N(0, Q)$ to a standard normal vector.

(46) Let $\varepsilon \sim N(0, Q)$, where $\text{null}(Q) = 0$. Then there exists a nonsingular matrix T such that $T'T = Q^{-1}$, $TQT' = I$, and it follows that $T\varepsilon \sim N(0, I)$.

This result can be used immediately to prove the first result concerning quadratic forms:

(47) If $\varepsilon \sim N_n(0, Q)$ and Q^{-1} exists, then $\varepsilon' Q^{-1} \varepsilon \sim \chi^2(n)$.

This follows since, if T is a matrix such that $T'T = Q$, $TQT' = I$, then $\eta = T\varepsilon \sim N_n(0, I)$; whence, from (40), it follows that $\eta'\eta = \varepsilon' T'T\varepsilon = \varepsilon' Q^{-1}\varepsilon \sim \chi^2(n)$.