

## 5.7: The Multivariate Normal Distribution

The multivariate normal distribution is among the most important of multivariate distributions, particularly in statistical inference and the study of Gaussian processes such as Brownian motion. The distribution arises naturally from linear transformations of independent normal variables. In this section, we consider the bivariate normal distribution first, because explicit results can be given and because graphical interpretations are possible. Then, with the aid of matrix notation, we discuss the general multivariate distribution.

### The Bivariate Normal Distribution

#### The Standard Distribution

Recall that the probability density function  $\phi$  of the standard normal distribution is given by

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad z \in \mathbb{R} \quad (5.7.1)$$

The corresponding distribution function is denoted  $\Phi$  and is considered a special function in mathematics:

$$\Phi(z) = \int_{-\infty}^z \phi(x) dx = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx, \quad z \in \mathbb{R} \quad (5.7.2)$$

Finally, the moment generating function  $m$  is given by

$$m(t) = \mathbb{E}(e^{tZ}) = \exp\left[\frac{1}{2}\text{var}(tZ)\right] = e^{t^2/2}, \quad t \in \mathbb{R} \quad (5.7.3)$$

Suppose that  $Z$  and  $W$  are independent random variables, each with the standard normal distribution. The distribution of  $(Z, W)$  is known as the *standard bivariate normal distribution*.

The basic properties of the standard bivariate normal distribution follow easily from independence and properties of the (univariate) normal distribution. Recall first that the graph of a function  $f: \mathbb{R}^2 \rightarrow \mathbb{R}$  is a surface. For  $c \in \mathbb{R}$ , the set of points  $\{(x, y) \in \mathbb{R}^2 : f(x, y) = c\}$  is the *level curve* of  $f$  at level  $c$ . The graph of  $f$  can be understood by means of the level curves.

The probability density function  $\phi_2$  of the standard bivariate normal distribution is given by

$$\phi_2(z, w) = \frac{1}{2\pi} e^{-\frac{1}{2}(z^2 + w^2)}, \quad (z, w) \in \mathbb{R}^2 \quad (5.7.4)$$

1. The level curves of  $\phi_2$  are circles centered at the origin.
2. The mode of the distribution is  $(0, 0)$ .
3.  $\phi_2$  is concave downward on  $\{(z, w) \in \mathbb{R}^2 : z^2 + w^2 < 1\}$

Proof

By independence,  $\phi_2(z, w) = \phi(z)\phi(w)$  for  $(z, w) \in \mathbb{R}^2$ . Parts (a) and (b) are clear. For part (c), the second derivative matrix of  $\phi_2$  is

$$\begin{bmatrix} \phi_2(z, w)(z^2 - 1) & \phi_2(z, w)zw \\ \phi_2(z, w)zw & \phi_2(z, w)(w^2 - 1) \end{bmatrix} \quad (5.7.5)$$

with determinant  $\phi_2^2(z, w)(1 - z^2 - w^2)$ . The determinant is positive and the diagonal entries negative on the circular region  $\{(z, w) \in \mathbb{R}^2 : z^2 + w^2 < 1\}$ , so the matrix is negative definite on this region.

Clearly  $\phi$  has a number of symmetry properties as well:  $\phi_2(z, w)$  is symmetric in  $z$  about 0 so that  $\phi_2(-z, w) = \phi_2(z, w)$ ;  $\phi_2(z, w)$  is symmetric in  $w$  about 0 so that  $\phi_2(z, -w) = \phi_2(z, w)$ ;  $\phi_2(z, w)$  is symmetric in  $(z, w)$  so that  $\phi_2(z, w) = \phi_2(w, z)$ . In short,  $\phi_2$  has the classical “bell shape” that we associate with normal distributions.

Open the bivariate normal experiment, keep the default settings to get the standard bivariate normal distribution. Run the experiment 1000 times. Observe the cloud of points in the scatterplot, and compare the empirical density functions to the probability density functions.

Suppose that  $(Z, W)$  has the standard bivariate normal distribution. The moment generating function  $m_2$  of  $(Z, W)$  is given by

$$m_2(s, t) = \mathbb{E}[\exp(sZ + tW)] = \exp\left[\frac{1}{2}\text{var}(sZ + tW)\right] = \exp\left[\frac{1}{2}(s^2 + t^2)\right], \quad (s, t) \in \mathbb{R}^2 \quad (5.7.6)$$

Proof

By independence,  $m_2(s, t) = m(s)m(t)$  for  $(s, t) \in \mathbb{R}^2$  where  $m$  is the standard normal MGF.

## The General Distribution

The general bivariate normal distribution can be constructed by means of an affine transformation on a standard bivariate normal vector. The distribution has 5 parameters. As we will see, two are location parameters, two are scale parameters, and one is a correlation parameter.

Suppose that  $(Z, W)$  has the standard bivariate normal distribution. Let  $\mu, \nu \in \mathbb{R}$ ;  $\sigma, \tau \in (0, \infty)$ ; and  $\rho \in (-1, 1)$ , and let  $X$  and  $Y$  be new random variables defined by

$$X = \mu + \sigma Z \quad (5.7.7)$$

$$Y = \nu + \tau \rho Z + \tau \sqrt{1 - \rho^2} W \quad (5.7.8)$$

The joint distribution of  $(X, Y)$  is called the *bivariate normal distribution* with parameters  $(\mu, \nu, \sigma, \tau, \rho)$

We can use the change of variables formula to find the joint probability density function.

Suppose that  $(X, Y)$  has the bivariate normal distribution with the parameters  $(\mu, \nu, \sigma, \tau, \rho)$  as specified above. The joint probability density function  $f$  of  $(X, Y)$  is given by

$$f(x, y) = \frac{1}{2\pi\sigma\tau\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu)^2}{\sigma^2} - 2\rho\frac{(x-\mu)(y-\nu)}{\sigma\tau} + \frac{(y-\nu)^2}{\tau^2}\right]\right\}, \quad (x, y) \in \mathbb{R}^2 \quad (5.7.9)$$

1. The level curves of  $f$  are ellipses centered at  $(\mu, \nu)$ .
2. The mode of the distribution is  $(\mu, \nu)$ .

Proof

Consider the transformation that defines  $(x, y)$  from  $(z, w)$  in the [definition](#). The inverse transformation is given by

$$z = \frac{x - \mu}{\sigma} \quad (5.7.10)$$

$$w = \frac{x - \mu}{\sigma\sqrt{1-\rho^2}} - \rho\frac{y - \nu}{\tau\sqrt{1-\rho^2}} \quad (5.7.11)$$

The Jacobian of the inverse transformation is

$$\frac{\partial(z, w)}{\partial(x, y)} = \frac{1}{\sigma\tau\sqrt{1-\rho^2}} \quad (5.7.12)$$

Note that the Jacobian is a constant, because the transformation is affine. The result now follows from the independence of  $Z$  and  $W$ , and the change of variables formula

1. Note that  $f$  has the form  $f(x, y) = a \exp[-bg(x, y)]$  where  $a$  and  $b$  are positive constants and

$$g(x, y) = \frac{(x - \mu)^2}{\sigma^2} - 2\rho\frac{(x - \mu)(y - \nu)}{\sigma\tau} + \frac{(y - \nu)^2}{\tau^2}, \quad (x, y) \in \mathbb{R}^2 \quad (5.7.13)$$

The graph of  $g$  is a paraboloid opening upward. The level curves of  $f$  are the same as the level curves of  $g$  (but at different levels of course).

2. The maximum of  $f$  occurs at the minimum of  $g$ , at the point  $(\mu, \nu)$ .

The following theorem gives fundamental properties of the bivariate normal distribution.

Suppose that  $(X, Y)$  has the bivariate normal distribution with parameters  $(\mu, \nu, \sigma, \tau, \rho)$  as specified above. Then

1.  $X$  is normally distributed with mean  $\mu$  and standard deviation  $\sigma$ .
2.  $Y$  is normally distributed with mean  $\nu$  and standard deviation  $\tau$ .
3.  $\text{cor}(X, Y) = \rho$ .
4.  $X$  and  $Y$  are independent if and only if  $\rho = 0$ .

Proof

These result can be proved from the probability density function, but it's easier and more helpful to use the transformation definition. So, assume that  $(X, Y)$  is defined in terms of the standard bivariate normal pair  $(Z, W)$  as in the [definition](#).

1.  $X = \mu + \sigma Z$  so  $X$  has the normal distribution with mean  $\mu$  and standard deviation  $\sigma$ . This is a basic property of the normal distribution, and indeed is the way that the general normal variable is constructed from a standard normal variable.

2. Since  $Z$  and  $W$  are independent and each has the standard normal distribution,  $Y = \nu + \tau\rho Z + \tau\sqrt{1-\rho^2}W$  is normally distributed by another basic property. Because  $Z$  and  $W$  have mean 0, it follows from the linear property of expected value that  $\mathbb{E}(Y) = \nu$ . Similarly, since  $Z$  and  $W$  have variance 1, it follows from basic properties of variance that  $\text{var}(Y) = \tau^2\rho^2 + \tau^2(1-\rho^2) = \tau^2$ .
3. Using the bi-linear property of covariance and independence we have  $\text{cov}(X, Y) = \rho\tau\sigma \text{cov}(Z, Z) = \rho\tau\sigma$ , and hence from (a) and (b),  $\text{cor}(X, Y) = \rho$ .
4. As a general property, recall that if  $X$  and  $Y$  are independent then  $\text{cor}(X, Y) = 0$ . Conversely, if  $\rho = 0$  then  $X = \mu + \sigma Z$  and  $Y = \nu + \tau W$ . Since  $Z$  and  $W$  are independent, so are  $X$  and  $Y$ .

Thus, two random variables with a joint normal distribution are independent if and only if they are uncorrelated.

In the bivariate normal experiment, change the standard deviations of  $X$  and  $Y$  with the scroll bars. Watch the change in the shape of the probability density functions. Now change the correlation with the scroll bar and note that the probability density functions do *not* change. For various values of the parameters, run the experiment 1000 times. Observe the cloud of points in the scatterplot, and compare the empirical density functions to the probability density functions.

In the case of perfect correlation ( $\rho = 1$  or  $\rho = -1$ ), the distribution of  $(X, Y)$  is also said to be bivariate normal, but *degenerate*. In this case, we know from our study of covariance and correlation that  $(X, Y)$  takes values on the regression line  $\{(x, y) \in \mathbb{R}^2 : y = \nu + \rho\frac{\tau}{\sigma}(x - \mu)\}$ , and hence does not have a probability density function (with respect to Lebesgue measure on  $\mathbb{R}^2$ ). Degenerate normal distributions will be discussed in more detail [below](#).

In the bivariate normal experiment, run the experiment 1000 times with the values of  $\rho$  given below and selected values of  $\sigma$  and  $\tau$ . Observe the cloud of points in the scatterplot and compare the empirical density functions to the probability density functions.

1.  $\rho \in \{0, 0.3, 0.5, 0.7, 1\}$
2.  $\rho \in \{-0.3, -0.5, -0.7, -1\}$

The conditional distributions are also normal.

Suppose that  $(X, Y)$  has the bivariate normal distribution with parameters  $(\mu, \nu, \sigma, \tau, \rho)$  as specified above.

1. For  $x \in \mathbb{R}$ , the conditional distribution of  $Y$  given  $X = x$  is normal with mean  $\mathbb{E}(Y | X = x) = \nu + \rho\frac{\tau}{\sigma}(x - \mu)$  and variance  $\text{var}(Y | X = x) = \tau^2(1 - \rho^2)$ .
2. For  $y \in \mathbb{R}$ , the conditional distribution of  $X$  given  $Y = y$  is normal with mean  $\mathbb{E}(X | Y = y) = \mu + \rho\frac{\sigma}{\tau}(y - \nu)$  and variance  $\text{var}(X | Y = y) = \sigma^2(1 - \rho^2)$ .

Proof from density functions

By symmetry, we need only prove (a). The conditional PDF of  $Y$  given  $X = x$  is  $y \mapsto f(x, y)/g(x)$  where  $f$  is the [joint PDF](#), and where  $g$  is the PDF of  $X$ , namely the normal PDF with mean  $\mu$  and standard deviation  $\sigma$ . The result then follows after some algebra.

Proof from random variables

Again, we only need to prove (a). We can assume that  $(X, Y)$  is defined in terms of a standard normal pair  $(Z, W)$  as in the [definition](#). Hence

$$Y = \nu + \rho\tau\frac{X - \mu}{\sigma} + \tau\sqrt{1 - \rho^2}W \quad (5.7.14)$$

Since that  $X$  and  $W$  are independent, the conditional distribution of  $Y$  given  $X = x$  is the distribution of  $\nu + \rho\tau\frac{x - \mu}{\sigma} + \tau\sqrt{1 - \rho^2}W$ . The latter distribution is normal, with mean and variance specified in the theorem.

Note that the conditional variances do not depend on the value of the given variable.

In the bivariate normal experiment, set the standard deviation of  $X$  to 1.5, the standard deviation of  $Y$  to 0.5, and the correlation to 0.7.

1. Run the experiment 100 times.
2. For each run, compute  $\mathbb{E}(Y | X = x)$  the predicted value of  $Y$  for the given the value of  $X$ .
3. Over all 100 runs, compute the square root of the average of the squared errors between the predicted value of  $Y$  and the true value of  $Y$ .

You may be perplexed by the lack of symmetry in how  $(X, Y)$  is defined in terms of  $(Z, W)$  in the original [definition](#). Note however that the distribution is completely determined by the 5 parameters. If we define  $X' = \mu + \sigma\rho Z + \sigma\sqrt{1 - \rho^2}W$  and  $Y' = \nu + \tau Z$  then  $(X', Y')$  has the same distribution as  $(X, Y)$ , namely the bivariate normal distribution with parameters  $(\mu, \nu, \sigma, \tau, \rho)$  (although, of course  $(X', Y')$  and  $(X, Y)$  are different random vectors). There are other ways to define the same distribution as an affine transformation of  $(Z, W)$ —the situation will be clarified in the next subsection.

Suppose that  $(X, Y)$  has the bivariate normal distribution with parameters  $(\mu, \nu, \sigma, \tau, \rho)$ . Then  $(X, Y)$  has moment generating function  $M$  given by

$$M(s, t) = \mathbb{E}[\exp(sX + tY)] = \exp\left[\mathbb{E}(sX + tY) + \frac{1}{2}\text{var}(sX + tY)\right] = \exp\left[\mu s + \nu t + \frac{1}{2}(\sigma^2 s^2 + 2\rho\sigma\tau st + \tau^2 t^2)\right], \quad (s, t) \in \mathbb{R}^2 \quad (5.7.15)$$

Proof

Using the representation of  $(X, Y)$  in terms of the standard bivariate normal vector  $(Z, W)$  in the [definition](#) and collecting terms gives

$$M(s, t) = \mathbb{E}\left(\exp\left[(\mu s + \nu t) + (\sigma s + \rho\tau t)Z + \tau\sqrt{1 - \rho^2}tW\right]\right) \quad (5.7.16)$$

Hence from independence we have

$$M(s, t) = \exp(\mu s + \nu t)m(\sigma s + \rho\tau t)m\left(\tau\sqrt{1 - \rho^2}t\right) \quad (5.7.17)$$

where  $m$  is the standard normal MGF. Substituting and simplifying gives the result.

We showed [above](#) that if  $(X, Y)$  has a bivariate normal distribution then the marginal distributions of  $X$  and  $Y$  are also normal. The converse is not true.

Suppose that  $(X, Y)$  has probability density function  $f$  given by

$$f(x, y) = \frac{1}{2\pi}e^{-(x^2+y^2)/2}\left[1 + xy e^{-(x^2+y^2-2)/2}\right], \quad (x, y) \in \mathbb{R}^2 \quad (5.7.18)$$

1.  $X$  and  $Y$  each have standard normal distributions.
2.  $(X, Y)$  does not have a bivariate normal distribution.

Proof

Note that  $f(x, y) = \phi_2(x, y)[1 + u(x)u(y)]$  for  $(x, y) \in \mathbb{R}^2$ , where  $\phi_2$  is the [bivariate standard normal PDF](#) and where  $u$  is given by  $u(t) = te^{-(t^2-1)/2}$  for  $t \in \mathbb{R}$ . From simple calculus,  $u$  is symmetric about 0, has a local maximum at  $t = 1$ , and  $u(t) \rightarrow 0$  as  $t \rightarrow \infty$ . In particular,  $|u(t)| \leq 1$  for  $t \in \mathbb{R}$  and hence  $f(x, y) \geq 0$  for  $(x, y) \in \mathbb{R}^2$ . Next, a helpful trick is that we can write integrals of  $f$  as expected values of functions of a standard normal pair  $(Z, W)$ . In particular,

$$\int_{\mathbb{R}^2} f(x, y)dx dy = \mathbb{E}[1 + u(Z)u(W)] = 1 + \mathbb{E}[u(Z)]\mathbb{E}[u(W)] = 1 \quad (5.7.19)$$

since  $\mathbb{E}[u(Z)] = \mathbb{E}[u(W)] = 0$  by the symmetry of the standard normal distribution and the symmetry of  $u$  about 0. Hence  $f$  is a valid PDF on  $\mathbb{R}^2$ . Suppose now that  $(X, Y)$  has PDF  $f$ .

1. The PDF of  $X$  at  $x \in \mathbb{R}$  is

$$\int_{\mathbb{R}} f(x, y)dy = \int_{\mathbb{R}} \phi_2(x, y)dy + u(x)\mathbb{E}[u(W)] = \phi(x) \quad (5.7.20)$$

where as usual,  $\phi$  is the standard normal PDF on  $\mathbb{R}$ . By symmetry,  $Y$  also has the standard normal distribution.

2.  $f$  does not have the form of a [bivariate normal PDF](#) and hence  $(X, Y)$  does not have a bivariate normal distribution.

## Transformations

Like its univariate counterpart, the family of bivariate normal distributions is preserved under two types of transformations on the underlying random vector: affine transformations and sums of independent vectors. We start with a preliminary result on affine transformations that should help clarify the original [definition](#). Throughout this discussion, we assume that the parameter vector  $(\mu, \nu, \sigma, \tau, \rho)$  satisfies the usual conditions:  $\mu, \nu \in \mathbb{R}$ , and  $\sigma, \tau \in (0, \infty)$ , and  $\rho \in (-1, 1)$ .

Suppose that  $(Z, W)$  has the standard bivariate normal distribution. Let  $X = a_1 + b_1Z + c_1W$  and  $Y = a_2 + b_2Z + c_2W$  where the coefficients are in  $\mathbb{R}$  and  $b_1c_2 - c_1b_2 \neq 0$ . Then  $(X, Y)$  has a bivariate normal distribution with parameters given by

1.  $\mathbb{E}(X) = a_1$
2.  $\mathbb{E}(Y) = a_2$
3.  $\text{var}(X) = b_1^2 + c_1^2$
4.  $\text{var}(Y) = b_2^2 + c_2^2$
5.  $\text{cov}(X, Y) = b_1b_2 + c_1c_2$

### Proof

A direct proof using the change of variables formula is possible, but our goal is to show that  $(X, Y)$  can be written in the form given above in the [definition](#). First, parts (a)–(e) follow from basic properties of expected value, variance, and covariance. So, in the notation used in the definition, we have  $\mu = a_1$ ,  $\nu = a_2$ ,  $\sigma = \sqrt{b_1^2 + c_1^2}$ ,  $\tau = \sqrt{b_2^2 + c_2^2}$ , and

$$\rho = \frac{b_1 b_2 + c_1 c_2}{\sqrt{b_1^2 + c_1^2} \sqrt{b_2^2 + c_2^2}} \quad (5.7.21)$$

(Note from the assumption on the coefficients that  $b_1^2 + c_1^2 > 0$  and  $b_2^2 + c_2^2 > 0$ ). Simple algebra shows that

$$\sqrt{1 - \rho^2} = \frac{b_1 c_2 - c_1 b_2}{\sqrt{b_1^2 + c_1^2} \sqrt{b_2^2 + c_2^2}} \quad (5.7.22)$$

Next we define

$$U = \frac{b_1 Z + c_1 W}{\sqrt{b_1^2 + c_1^2}} \quad (5.7.23)$$

$$V = \frac{c_1 Z - b_1 W}{\sqrt{b_1^2 + c_1^2}} \quad (5.7.24)$$

The transformation that defines  $(u, v)$  from  $(z, w)$  is its own inverse, and has Jacobian 1. Hence it follows that  $(U, V)$  has the same joint distribution as  $(Z, W)$ , namely the standard bivariate normal distribution. Simple algebra shows that

$$X = a_1 + \sqrt{b_1^2 + c_1^2} U = \mu + \sigma U \quad (5.7.25)$$

$$Y = a_2 + \frac{b_1 b_2 + c_1 c_2}{\sqrt{b_1^2 + c_1^2}} U + \frac{c_1 b_2 - b_1 c_2}{\sqrt{b_1^2 + c_1^2}} V = \nu + \tau \rho U + \tau \sqrt{1 - \rho^2} V \quad (5.7.26)$$

This is the form given in the [definition](#), so it follows that  $(X, Y)$  has a bivariate normal distribution.

Now it is easy to show more generally that the bivariate normal distribution is closed with respect to affine transformations.

Suppose that  $(X, Y)$  has the bivariate normal distribution with parameters  $(\mu, \nu, \sigma, \tau, \rho)$ . Define  $U = a_1 + b_1 X + c_1 Y$  and  $V = a_2 + b_2 X + c_2 Y$ , where the coefficients are in  $\mathbb{R}$  and  $b_1 c_2 - c_1 b_2 \neq 0$ . Then  $(U, V)$  has a bivariate normal distribution with parameters as follows:

1.  $\mathbb{E}(U) = a_1 + b_1 \mu + c_1 \nu$
2.  $\mathbb{E}(V) = a_2 + b_2 \mu + c_2 \nu$
3.  $\text{var}(U) = b_1^2 \sigma^2 + c_1^2 \tau^2 + 2b_1 c_1 \rho \sigma \tau$
4.  $\text{var}(V) = b_2^2 \sigma^2 + c_2^2 \tau^2 + 2b_2 c_2 \rho \sigma \tau$
5.  $\text{cov}(U, V) = b_1 b_2 \sigma^2 + c_1 c_2 \tau^2 + (b_1 c_2 + b_2 c_1) \rho \sigma \tau$

### Proof

From our original [construction](#), we can write  $X = \mu + \sigma Z$  and  $Y = \nu + \tau \rho Z + \tau \sqrt{1 - \rho^2} W$  where  $(Z, W)$  has the standard bivariate normal distribution. Then by simple substitution,  $U = A_1 + B_1 Z + C_1 W$  and  $V = A_2 + B_2 Z + C_2 W$  where  $A_i = a_i + b_i \mu + c_i \nu$ ,  $B_i = b_i \sigma + c_i \tau \rho$ ,  $C_i = c_i \tau \sqrt{1 - \rho^2}$  for  $i \in \{1, 2\}$ . By simple algebra,

$$B_1 C_2 - C_1 B_2 = \sigma \tau \sqrt{1 - \rho^2} (b_1 c_2 - c_1 b_2) \neq 0 \quad (5.7.27)$$

Hence  $(U, V)$  has a bivariate normal distribution from the previous theorem. Parts (a)–(e) follow from basic properties of expected value, variance, and covariance.

The bivariate normal distribution is preserved with respect to sums of independent variables.

Suppose that  $(X_i, Y_i)$  has the bivariate normal distribution with parameters  $(\mu_i, \nu_i, \sigma_i, \tau_i, \rho_i)$  for  $i \in \{1, 2\}$ , and that  $(X_1, Y_1)$  and  $(X_2, Y_2)$  are independent. Then  $(X_1 + X_2, Y_1 + Y_2)$  has the bivariate normal distribution with parameters given by

1.  $\mathbb{E}(X_1 + X_2) = \mu_1 + \mu_2$
2.  $\mathbb{E}(Y_1 + Y_2) = \nu_1 + \nu_2$
3.  $\text{var}(X_1 + X_2) = \sigma_1^2 + \sigma_2^2$
4.  $\text{var}(Y_1 + Y_2) = \tau_1^2 + \tau_2^2$

$$5. \operatorname{cov}(X_1 + X_2, Y_1 + Y_2) = \rho_1 \sigma_1 \tau_1 + \rho_2 \sigma_2 \tau_2$$

Proof

Let  $M_i$  denote the MGF of  $(X_i, Y_i)$  for  $i \in \{1, 2\}$  and let  $M$  denote the MGF of  $(X_1 + X_2, Y_1 + Y_2)$ . By independence,  $M(s, t) = M_1(s, t)M_2(s, t)$  for  $(s, t) \in \mathbb{R}^2$ . Using the [bivariate normal MGF](#), and basic properties of the exponential function,

$$M(s, t) = \exp\left(\mathbb{E}[s(X_1 + X_2) + t(Y_1 + Y_2)] + \frac{1}{2}\operatorname{var}[s(X_1 + X_2) + t(Y_1 + Y_2)]\right), \quad (s, t) \in \mathbb{R}^2 \quad (5.7.28)$$

Of course from basic properties of expected value, variance, and covariance,

$$\mathbb{E}[s(X_1 + X_2) + t(Y_1 + Y_2)] = s(\mu_1 + \mu_2) + t(\nu_1 + \nu_2) \quad (5.7.29)$$

$$\operatorname{var}[s(X_1 + X_2) + t(Y_1 + Y_2)] = s(\sigma_1^2 + \sigma_2^2) + t(\tau_1^2 + \tau_2^2) + 2st(\rho_1 \sigma_1 \tau_1 + \rho_2 \sigma_2 \tau_2) \quad (5.7.30)$$

Substituting gives the result.

The following result is important in the simulation of normal variables.

Suppose that  $(Z, W)$  has the standard bivariate normal distribution. Define the polar coordinates  $(R, \Theta)$  of  $(Z, W)$  by the equations  $Z = R \cos \Theta$ ,  $W = R \sin \Theta$  where  $R \geq 0$  and  $0 \leq \Theta < 2\pi$ . Then

1.  $R$  has probability density function  $g$  given by  $g(r) = r e^{-\frac{1}{2}r^2}$  for  $r \in [0, \infty)$ .
2.  $\Theta$  is uniformly distributed on  $[0, 2\pi)$ .
3.  $R$  and  $\Theta$  are independent.

Proof

The Jacobian of the polar coordinate transformation that gives  $(z, w)$  from  $(r, \theta)$  is  $r$ , as we all remember from calculus. Hence by the change of variables theorem, the PDF  $g$  of  $(R, \Theta)$  in terms of the from [standard normal PDF](#)  $\phi_2$  is given by

$$g(r, \theta) = \phi_2(r \cos \theta, r \sin \theta) r = \frac{1}{2\pi} r e^{-r^2/2}, \quad (r, \theta) \in [0, \infty) \times [0, 2\pi) \quad (5.7.31)$$

The result then follows from the factorization theorem for independent random variables.

The distribution of  $R$  is known as the *standard Rayleigh distribution*, named for William Strutt, Lord Rayleigh. The Rayleigh distribution studied in more detail in a separate section.

Since the quantile function  $\Phi^{-1}$  of the normal distribution cannot be given in a simple, closed form, we cannot use the usual random quantile method of simulating a normal random variable. However, the quantile method works quite well to simulate a Rayleigh variable, and of course simulating uniform variables is trivial. Hence we have a way of simulating a standard bivariate normal vector with a pair of random numbers (which, you will recall are independent random variables, each with the standard uniform distribution, that is, the uniform distribution on  $[0, 1)$ ).

Suppose that  $U$  and  $V$  are independent random variables, each with the standard uniform distribution. Let  $R = \sqrt{-2 \ln U}$  and  $\Theta = 2\pi V$ . Define  $Z = R \cos \Theta$  and  $W = R \sin \Theta$ . Then  $(Z, W)$  has the standard bivariate normal distribution.

Proof

The Rayleigh distribution function  $F$  is given by  $F(r) = 1 - e^{-r^2/2}$  for  $r \in [0, \infty)$  and hence the quantile function is given by  $F^{-1}(p) = \sqrt{-2 \ln(1-p)}$  for  $p \in [0, 1)$ . Hence if  $U$  has the standard uniform distribution, then  $\sqrt{-2 \ln(1-U)}$  has the Rayleigh distribution. But  $1 - U$  also has the standard uniform distribution so  $R = \sqrt{-2 \ln U}$  also has the Rayleigh distribution. If  $V$  has the standard uniform distribution then of course  $2\pi V$  is uniformly distributed on  $[0, 2\pi)$ . If  $U$  and  $V$  are independent, then so are  $R$  and  $\Theta$ . By the previous theorem, if  $Z = R \cos \Theta$  and  $W = R \sin \Theta$ , then  $(Z, W)$  has the standard bivariate normal distribution.

Of course, if we can simulate  $(Z, W)$  with a standard bivariate normal distribution, then we can simulate  $(X, Y)$  with the general bivariate normal distribution, with parameter  $(\mu, \nu, \sigma, \tau, \rho)$  by definition (5), namely  $X = \mu + \sigma Z$ ,  $Y = \nu + \tau \rho Z + \tau \sqrt{1 - \rho^2} W$ .

## The General Multivariate Normal Distribution

The general multivariate normal distribution is a natural generalization of the bivariate normal distribution studied above. The exposition is very compact and elegant using expected value and covariance matrices, and would be horribly complex without these tools. Thus, this section requires some prerequisite knowledge of linear algebra. In particular, recall that  $\mathbf{A}^T$  denotes the transpose of a matrix  $\mathbf{A}$  and that we identify a vector in  $\mathbb{R}^n$  with the corresponding  $n \times 1$  column vector.

### The Standard Distribution

Suppose that  $\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)$  is a vector of independent random variables, each with the standard normal distribution. Then  $\mathbf{Z}$  is said to have the  $n$ -dimensional *standard normal distribution*.

1.  $\mathbb{E}(\mathbf{Z}) = \mathbf{0}$  (the zero vector in  $\mathbb{R}^n$ ).
2.  $\text{vc}(\mathbf{Z}) = I$  (the  $n \times n$  identity matrix).

$\mathbf{Z}$  has probability density function  $\phi_n$  given by

$$\phi_n(\mathbf{z}) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\mathbf{z} \cdot \mathbf{z}\right) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2} \sum_{i=1}^n z_i^2\right), \quad \mathbf{z} = (z_1, z_2, \dots, z_n) \in \mathbb{R}^n \quad (5.7.32)$$

where as usual,  $\phi$  is the standard normal PDF.

Proof

By independence,  $\phi_n(\mathbf{z}) = \phi(z_1)\phi(z_2) \cdots \phi(z_n)$ .

$\mathbf{Z}$  has moment generating function  $m_n$  given by

$$m_n(\mathbf{t}) = \mathbb{E}[\exp(\mathbf{t} \cdot \mathbf{Z})] = \exp\left[\frac{1}{2}\text{var}(\mathbf{t} \cdot \mathbf{Z})\right] = \exp\left(\frac{1}{2}\mathbf{t} \cdot \mathbf{t}\right) = \exp\left(\frac{1}{2} \sum_{i=1}^n t_i^2\right), \quad \mathbf{t} = (t_1, t_2, \dots, t_n) \in \mathbb{R}^n \quad (5.7.33)$$

Proof

By independence,  $\mathbb{E}[\exp(\mathbf{t} \cdot \mathbf{Z})] = m(t_1)m(t_2) \cdots m(t_n)$  where  $m$  is the standard normal MGF.

### The General Distribution

Suppose that  $\mathbf{Z}$  has the  $n$ -dimensional standard normal distribution. Suppose also that  $\boldsymbol{\mu} \in \mathbb{R}^n$  and that  $\mathbf{A} \in \mathbb{R}^{n \times n}$  is invertible. The random vector  $\mathbf{X} = \boldsymbol{\mu} + \mathbf{A}\mathbf{Z}$  is said to have an  $n$ -dimensional normal distribution.

1.  $\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu}$ .
2.  $\text{vc}(\mathbf{X}) = \mathbf{A}\mathbf{A}^T$ .

Proof

1. From the linear property of expected value,  $\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu} + \mathbf{A}\mathbb{E}(\mathbf{Z}) = \boldsymbol{\mu}$ .
2. From basic properties of the variance-covariance matrix,  $\text{vc}(\mathbf{X}) = \mathbf{A}\mathbf{A}^T \text{vc}(\mathbf{Z}) = \mathbf{A}\mathbf{A}^T$ .

In the context of this result, recall that the variance-covariance matrix  $\text{vc}(\mathbf{X}) = \mathbf{A}\mathbf{A}^T$  is symmetric and positive definite (and hence also invertible). We will now see that the multivariate normal distribution is completely determined by the expected value vector  $\boldsymbol{\mu}$  and the variance-covariance matrix  $\mathbf{V}$ , and hence these give the basic parameters of the distribution.

Suppose that  $\mathbf{X}$  has an  $n$ -dimensional normal distribution with expected value vector  $\boldsymbol{\mu}$  and variance-covariance matrix  $\mathbf{V}$ . The probability density function  $f$  of  $\mathbf{X}$  is given by

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} \sqrt{\det(\mathbf{V})}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}) \cdot \mathbf{V}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right], \quad \mathbf{x} \in \mathbb{R}^n \quad (5.7.34)$$

Proof

From the definition can assume that  $\mathbf{X} = \boldsymbol{\mu} + \mathbf{A}\mathbf{Z}$  where  $\mathbf{A} \in \mathbb{R}^{n \times n}$  is invertible and  $\mathbf{Z}$  has the  $n$ -dimensional standard normal distribution, so that  $\mathbf{V} = \mathbf{A}\mathbf{A}^T$ . The inverse of the transformation  $\mathbf{x} = \boldsymbol{\mu} + \mathbf{A}\mathbf{z}$  is  $\mathbf{z} = \mathbf{A}^{-1}(\mathbf{x} - \boldsymbol{\mu})$  and hence the Jacobian of the inverse transformation is  $\det(\mathbf{A}^{-1}) = 1/\det(\mathbf{A})$ . Using the multivariate change of variables theorem,

$$f(\mathbf{x}) = \frac{1}{|\det(\mathbf{A})|} \phi_n[\mathbf{A}^{-1}(\mathbf{x} - \boldsymbol{\mu})] = \frac{1}{(2\pi)^{n/2} |\det(\mathbf{A})|} \exp\left[-\frac{1}{2}\mathbf{A}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \cdot \mathbf{A}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right], \quad \mathbf{x} \in \mathbb{R}^n \quad (5.7.35)$$

But  $\det(\mathbf{V}) = \det(\mathbf{A}\mathbf{A}^T) = \det(\mathbf{A}) \det(\mathbf{A}^T) = [\det(\mathbf{A})]^2$  and hence  $|\det(\mathbf{A})| = \sqrt{\det(\mathbf{V})}$ . Also,

$$\mathbf{A}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \cdot \mathbf{A}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = [\mathbf{A}^{-1}(\mathbf{x} - \boldsymbol{\mu})]^T \mathbf{A}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = (\mathbf{x} - \boldsymbol{\mu})^T (\mathbf{A}^{-1})^T \mathbf{A}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \quad (5.7.36)$$

$$= (\mathbf{x} - \boldsymbol{\mu})^T (\mathbf{A}^T)^{-1} \mathbf{A}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = (\mathbf{x} - \boldsymbol{\mu})^T (\mathbf{A}\mathbf{A}^T)^{-1}(\mathbf{x} - \boldsymbol{\mu}) \quad (5.7.37)$$

$$= (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{V}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = (\mathbf{x} - \boldsymbol{\mu}) \cdot \mathbf{V}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \quad (5.7.38)$$

Suppose again that  $\mathbf{X}$  has an  $n$ -dimensional normal distribution with expected value vector  $\boldsymbol{\mu}$  and variance-covariance matrix  $\mathbf{V}$ . The moment generating function  $M$  of  $\mathbf{X}$  is given by

$$M(\mathbf{t}) = \mathbb{E}[\exp(\mathbf{t} \cdot \mathbf{X})] = \exp\left[\mathbb{E}(\mathbf{t} \cdot \mathbf{X}) + \frac{1}{2}\text{var}(\mathbf{t} \cdot \mathbf{X})\right] = \exp\left(\mathbf{t} \cdot \boldsymbol{\mu} + \frac{1}{2}\mathbf{t} \cdot \mathbf{V}\mathbf{t}\right), \quad \mathbf{t} \in \mathbb{R}^n \quad (5.7.39)$$

Proof

Once again we start with the [definition](#) and assume that  $\mathbf{X} = \boldsymbol{\mu} + \mathbf{A}\mathbf{Z}$  where  $\mathbf{A} \in \mathbb{R}^{n \times n}$  is invertible. we have  $\mathbb{E}[\exp(\mathbf{t} \cdot \mathbf{X})] = \exp(\mathbf{t} \cdot \boldsymbol{\mu}) \mathbb{E}[\exp(\mathbf{t} \cdot \mathbf{A}\mathbf{Z})]$ . But  $\mathbf{t} \cdot \mathbf{A}\mathbf{Z} = (\mathbf{A}^T \mathbf{t}) \cdot \mathbf{Z}$  so using the [MGF](#) of  $\mathbf{Z}$  we have

$$\mathbb{E}[\exp(\mathbf{t} \cdot \mathbf{A}\mathbf{Z})] = \exp\left[\frac{1}{2}(\mathbf{A}^T \mathbf{t}) \cdot (\mathbf{A}^T \mathbf{t})\right] = \exp\left[\frac{1}{2}\mathbf{t}^T \mathbf{A} \mathbf{A}^T \mathbf{t}\right] = \exp\left[\frac{1}{2}\mathbf{t} \cdot \mathbf{V} \mathbf{t}\right] \quad (5.7.40)$$

Of course, the moment generating function completely determines the distribution. Thus, if a random vector  $\mathbf{X}$  in  $\mathbb{R}^n$  has a moment generating function of the form given above, for some  $\boldsymbol{\mu} \in \mathbb{R}^n$  and symmetric, positive definite  $\mathbf{V} \in \mathbb{R}^{n \times n}$ , then  $\mathbf{X}$  has the  $n$ -dimensional normal distribution with mean  $\boldsymbol{\mu}$  and variance-covariance matrix  $\mathbf{V}$ .

Note again that in the representation  $\mathbf{X} = \boldsymbol{\mu} + \mathbf{A}\mathbf{Z}$ , the distribution of  $\mathbf{X}$  is uniquely determined by the expected value vector  $\boldsymbol{\mu}$  and the variance-covariance matrix  $\mathbf{V} = \mathbf{A}\mathbf{A}^T$ , but not by  $\boldsymbol{\mu}$  and  $\mathbf{A}$ . In general, for a given positive definite matrix  $\mathbf{V}$ , there are many invertible matrices  $\mathbf{A}$  such that  $\mathbf{V} = \mathbf{A}\mathbf{A}^T$  (the matrix  $\mathbf{A}$  is a bit like a square root of  $\mathbf{V}$ ). A theorem in matrix theory states that there is a unique lower triangular matrix  $\mathbf{L}$  with this property. The representation  $\mathbf{X} = \boldsymbol{\mu} + \mathbf{L}\mathbf{Z}$  is known as the *canonical representation* of  $\mathbf{X}$ .

If  $\mathbf{X} = (X, Y)$  has bivariate normal distribution with parameters  $(\mu, \nu, \sigma, \tau, \rho)$ , then the lower triangular matrix  $\mathbf{L}$  such that  $\mathbf{L}\mathbf{L}^T = \text{vc}(\mathbf{X})$  is

$$\mathbf{L} = \begin{bmatrix} \sigma & 0 \\ \tau\rho & \tau\sqrt{1-\rho^2} \end{bmatrix} \quad (5.7.41)$$

Proof

Note that

$$\mathbf{L}\mathbf{L}^T = \begin{bmatrix} \sigma^2 & \sigma\tau\rho \\ \sigma\tau\rho & \tau^2 \end{bmatrix} = \text{vc}(X, Y) \quad (5.7.42)$$

Note that the matrix  $\mathbf{L}$  above gives the canonical representation of  $(X, Y)$  in terms of the standard normal vector  $(Z, W)$  in the original [definition](#), namely  $X = \mu + \sigma Z$ ,  $Y = \nu + \tau\rho Z + \tau\sqrt{1-\rho^2} W$ .

If the matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$  in the [definition](#) is not invertible, then the variance-covariance matrix  $\mathbf{V} = \mathbf{A}\mathbf{A}^T$  is symmetric, but only positive semi-definite. The random vector  $\mathbf{X} = \boldsymbol{\mu} + \mathbf{A}\mathbf{Z}$  takes values in a lower dimensional affine subspace of  $\mathbb{R}^n$  that has measure 0 relative to  $n$ -dimensional Lebesgue measure  $\lambda_n$ . Thus,  $\mathbf{X}$  does not have a probability density function relative to  $\lambda_n$ , and so the distribution is degenerate. However, the formula for the [moment generating function](#) still holds. Degenerate normal distributions are discussed in more detail [below](#).

## Transformations

The multivariate normal distribution is invariant under two basic types of transformations on the underlying random vectors: affine transformations (with linearly independent rows), and concatenation of independent vectors. As simple corollaries of these two results, the normal distribution is also invariant with respect to subsequences of the random vector, re-orderings of the terms in the random vector, and sums of independent random vectors. The main tool that we will use is the moment generating function. We start with the first main result on affine transformations.

Suppose that  $\mathbf{X}$  has the  $n$ -dimensional normal distribution with mean vector  $\boldsymbol{\mu}$  and variance-covariance matrix  $\mathbf{V}$ . Suppose also that  $\mathbf{a} \in \mathbb{R}^m$  and that  $\mathbf{A} \in \mathbb{R}^{m \times n}$  has linearly independent rows (thus,  $m \leq n$ ). Then  $\mathbf{Y} = \mathbf{a} + \mathbf{A}\mathbf{X}$  has an  $m$ -dimensional normal distribution, with

1.  $\mathbb{E}(\mathbf{Y}) = \mathbf{a} + \mathbf{A}\boldsymbol{\mu}$
2.  $\text{vc}(\mathbf{Y}) = \mathbf{A}\mathbf{V}\mathbf{A}^T$

Proof

For  $\mathbf{t} \in \mathbb{R}^m$ ,  $\mathbb{E}[\exp(\mathbf{t} \cdot \mathbf{Y})] = \exp(\mathbf{t} \cdot \mathbf{a}) \mathbb{E}[\exp(\mathbf{t} \cdot \mathbf{A}\mathbf{X})]$ . but  $\mathbf{t} \cdot \mathbf{A}\mathbf{X} = (\mathbf{A}^T \mathbf{t}) \cdot \mathbf{X}$ , so using the [MGF](#) of  $\mathbf{X}$  we have

$$\mathbb{E}[\exp(\mathbf{t} \cdot \mathbf{A}\mathbf{X})] = \exp\left[(\mathbf{A}^T \mathbf{t}) \cdot \boldsymbol{\mu} + \frac{1}{2}(\mathbf{A}^T \mathbf{t}) \cdot \mathbf{V}(\mathbf{A}^T \mathbf{t})\right] \quad (5.7.43)$$

But  $(\mathbf{A}^T \mathbf{t}) \cdot \boldsymbol{\mu} = \mathbf{t} \cdot \mathbf{A}\boldsymbol{\mu}$  and  $(\mathbf{A}^T \mathbf{t}) \cdot \mathbf{V}(\mathbf{A}^T \mathbf{t}) = \mathbf{t} \cdot (\mathbf{A}\mathbf{V}\mathbf{A}^T) \mathbf{t}$ , so letting  $\mathbf{b} = \mathbf{a} + \mathbf{A}\boldsymbol{\mu}$  and  $\mathbf{U} = \mathbf{A}\mathbf{V}\mathbf{A}^T$  and putting the pieces together, we have  $\mathbb{E}[\exp(\mathbf{t} \cdot \mathbf{Y})] = \exp[\mathbf{b} \cdot \mathbf{t} + \frac{1}{2}\mathbf{t} \cdot \mathbf{U} \mathbf{t}]$ .

A clearly important special case is  $m = n$ , which generalizes the [definition](#). Thus, if  $\mathbf{a} \in \mathbb{R}^n$  and  $\mathbf{A} \in \mathbb{R}^{n \times n}$  is invertible, then  $\mathbf{Y} = \mathbf{a} + \mathbf{A}\mathbf{X}$  has an  $n$ -dimensional normal distribution. Here are some other corollaries:

Suppose that  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  has an  $n$ -dimensional normal distribution. If  $\{i_1, i_2, \dots, i_m\}$  is a set of distinct indices, then  $\mathbf{Y} = (X_{i_1}, X_{i_2}, \dots, X_{i_m})$  has an  $m$ -dimensional normal distribution.

### Proof

Let  $A \in \mathbb{R}^{m \times n}$  be the matrix defined by the condition that for  $j \in \{1, 2, \dots, m\}$ , row  $j$  has 1 in position  $i_j$  and has 0 in all other positions. Then  $A$  has linearly independent rows (since the  $i_j$  are distinct in  $j$ ) and  $Y = AX$ . Thus the result follows from the general theorem on [affine transformations](#).

In the context of the previous result, if  $X$  has mean vector  $\mu$  and variance-covariance matrix  $V$ , then  $Y$  has mean vector  $A\mu$  and variance-covariance matrix  $AVAT^T$ , where  $A$  is the 0-1 matrix defined in the proof. As simple corollaries, note that if  $X = (X_1, X_2, \dots, X_n)$  has an  $n$ -dimensional normal distribution, then any permutation of the coordinates of  $X$  also has an  $n$ -dimensional normal distribution, and  $(X_1, X_2, \dots, X_m)$  has an  $m$ -dimensional normal distribution for any  $m \leq n$ . Here is a slight extension of the last statement.

Suppose that  $X$  is a random vector in  $\mathbb{R}^m$ ,  $Y$  is a random vector in  $\mathbb{R}^n$ , and that  $(X, Y)$  has an  $(m+n)$ -dimensional normal distribution. Then

1.  $X$  has an  $m$ -dimensional normal distribution.
2.  $Y$  has an  $n$ -dimensional normal distribution.
3.  $X$  and  $Y$  are independent if and only if  $\text{cov}(X, Y) = \mathbf{0}$  (the  $m \times n$  zero matrix).

### Proof

As we already noted, parts (a) and (b) are a simple consequence of the [previous theorem](#). Thus, we just need to verify (c). In block form, note that

$$\text{vc}(X, Y) = \begin{bmatrix} \text{vc}(X) & \text{cov}(X, Y) \\ \text{cov}(Y, X) & \text{vc}(Y) \end{bmatrix} \quad (5.7.44)$$

Now let  $M$  denote the moment generating function of  $(X, Y)$ ,  $M_1$  the MGF of  $X$ , and  $M_2$  the MGF of  $Y$ . From the form of the [MGF](#), note that  $M(s, t) = M_1(s)M_2(t)$  for all  $s \in \mathbb{R}^m$ ,  $t \in \mathbb{R}^n$  if and only if  $\text{cov}(X, Y) = \mathbf{0}$ , the  $m \times n$  zero matrix.

Next is the converse to part (c) of the previous result: concatenating independent normally distributed vectors produces another normally distributed vector.

Suppose that  $X$  has the  $m$ -dimensional normal distribution with mean vector  $\mu$  and variance-covariance matrix  $U$ ,  $Y$  has the  $n$ -dimensional normal distribution with mean vector  $\nu$  and variance-covariance matrix  $V$ , and that  $X$  and  $Y$  are independent. Then  $Z = (X, Y)$  has the  $m+n$ -dimensional normal distribution with

1.  $\mathbb{E}(X, Y) = (\mu, \nu)$
2.  $\text{vc}(X, Y) = \begin{bmatrix} \text{vc}(X) & \mathbf{0} \\ \mathbf{0}^T & \text{vc}(Y) \end{bmatrix}$  where  $\mathbf{0}$  is the  $m \times n$  zero matrix.

### Proof

For  $t \in \mathbb{R}^{m+n}$ , write  $t$  in block form as  $t = (r, s)$  where  $r \in \mathbb{R}^m$  and  $s \in \mathbb{R}^n$ . By independence, the MGF of  $(X, Y)$  is

$$\mathbb{E}(\exp[t \cdot (X, Y)]) = \mathbb{E}[r \cdot X + s \cdot Y] = \mathbb{E}[\exp(r \cdot X)] \mathbb{E}[\exp(s \cdot Y)] \quad (5.7.45)$$

Using the formula for the [normal MGF](#) we have

$$\mathbb{E}(\exp[t \cdot (X, Y)]) = \exp\left(r \cdot \mu + \frac{1}{2} r \cdot U r\right) \exp\left(s \cdot \nu + \frac{1}{2} s \cdot V s\right) = \exp\left[\left(r \cdot \mu + s \cdot \nu\right) + \frac{1}{2} (r \cdot U r + s \cdot V s)\right] \quad (5.7.46)$$

But  $r \cdot \mu + s \cdot \nu = t \cdot (\mu, \nu)$  and  $r \cdot U r + s \cdot V s = t \cdot \begin{bmatrix} \text{vc}(X) & \mathbf{0} \\ \mathbf{0}^T & \text{vc}(Y) \end{bmatrix} t$  so the proof is complete

Just as in the univariate case, the normal family of distributions is closed with respect to sums of independent variables. The proof follows easily from the previous result.

Suppose that  $X$  has the  $n$ -dimensional normal distribution with mean vector  $\mu$  and variance-covariance matrix  $U$ ,  $Y$  has the  $n$ -dimensional normal distribution with mean vector  $\nu$  and variance-covariance matrix  $V$ , and that  $X$  and  $Y$  are independent. Then  $X + Y$  has the  $n$ -dimensional normal distribution with

1.  $\mathbb{E}(X + Y) = \mu + \nu$
2.  $\text{vc}(X + Y) = U + V$

### Proof

From the previous result  $(X, Y)$  has a  $2n$ -dimensional normal distribution. Moreover,  $X + Y = A(X, Y)$  where  $A$  is the  $n \times 2n$  matrix defined by the condition that for  $i \in \{1, 2, \dots, n\}$ , row  $i$  has 1 in positions  $i$  and  $n+i$  and 0 in all other positions. The matrix  $A$  has linearly

independent rows and thus the result follows from the general theorem on [affine transformations](#). Parts (a) and (b) are standard results for sums of independent random vectors.

We close with a trivial corollary to the general result on [affine transformation](#), but this corollary points the way to a further generalization of the multivariate normal distribution that includes the degenerate distributions.

Suppose that  $\mathbf{X}$  has an  $n$ -dimensional normal distribution with mean vector  $\boldsymbol{\mu}$  and variance-covariance matrix  $\mathbf{V}$ , and that  $\mathbf{a} \in \mathbb{R}^n$  with  $\mathbf{a} \neq \mathbf{0}$ . Then  $Y = \mathbf{a} \cdot \mathbf{X}$  has a (univariate) normal distribution with

1.  $\mathbb{E}(Y) = \mathbf{a} \cdot \boldsymbol{\mu}$
2.  $\text{var}(Y) = \mathbf{a} \cdot \mathbf{V} \mathbf{a}$

Proof

Note again that  $\mathbf{a} \cdot \mathbf{X} = \mathbf{a}^T \mathbf{X}$ . Since  $\mathbf{a} \neq \mathbf{0}$ , the single row of  $\mathbf{a}^T$  is linearly independent and hence the result follows from the general theorem on [affine transformations](#).

## A Further Generalization

The last result can be used to give a simple, elegant definition of the multivariate normal distribution that includes the degenerate distributions as well as the ones we have considered so far. First we will adopt our general definition of the univariate normal distribution that includes constant random variables.

A random variable  $\mathbf{X}$  that takes values in  $\mathbb{R}^n$  has an  $n$ -dimensional normal distribution if and only if  $\mathbf{a} \cdot \mathbf{X}$  has a univariate normal distribution for every  $\mathbf{a} \in \mathbb{R}^n$ .

Although an  $n$ -dimensional normal distribution may not have a probability density function with respect to  $n$ -dimensional Lebesgue measure  $\lambda_n$ , the form of the moment generating function is unchanged.

Suppose that  $\mathbf{X}$  has mean vector  $\boldsymbol{\mu}$  and variance-covariance matrix  $\mathbf{V}$ , and that  $\mathbf{X}$  has an  $n$ -dimensional normal distribution. The moment generating function of  $\mathbf{X}$  is given by

$$\mathbb{E}[\exp(\mathbf{t} \cdot \mathbf{X})] = \exp\left[\mathbb{E}(\mathbf{t} \cdot \mathbf{X}) + \frac{1}{2}\text{var}(\mathbf{t} \cdot \mathbf{X})\right] = \exp\left(\mathbf{t} \cdot \boldsymbol{\mu} + \frac{1}{2}\mathbf{t} \cdot \mathbf{V} \mathbf{t}\right), \quad \mathbf{t} \in \mathbb{R}^n \quad (5.7.47)$$

Proof

If  $\mathbf{t} \in \mathbb{R}^n$ , then by definition,  $\mathbf{t} \cdot \mathbf{X}$  has a univariate normal distribution. Thus  $\mathbb{E}[\exp(\mathbf{t} \cdot \mathbf{X})]$  is simply the moment generating function of  $\mathbf{t} \cdot \mathbf{X}$ , evaluated at the argument 1. The results then follow from the univariate MGF.

Our new general definition really is a generalization.

Suppose that  $\mathbf{X}$  has an  $n$ -dimensional normal distribution in the sense of the [general definition](#), and that the distribution of  $\mathbf{X}$  has a probability density function on  $\mathbb{R}^n$  with respect to Lebesgue measure  $\lambda_n$ . Then  $\mathbf{X}$  has an  $n$ -dimensional normal distribution in the sense of our [original definition](#).

Proof

This follows from our previous results, since both the MGF and the PDF completely determine the distribution of  $\mathbf{X}$ .

This page titled [5.7: The Multivariate Normal Distribution](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist](#) ([Random Services](#)) via [source content](#) that was edited to the style and standards of the LibreTexts platform.