$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}.$$

Find the PDF of $\mathbf{Y}$ in terms of PDF of $\mathbf{X}$.

Solution

Since A is invertible, we can write

$$X = A^{-1}(Y - b).$$

We can also check that

$$J = \det(A^{-1}) = \frac{1}{\det(A)}.$$

Thus, we conclude that

$$f_Y(y) = \frac{1}{|\det(A)|} f_X\big(A^{-1}(y - b)\big).$$

Normal (Gaussian) Random Vectors:

We discussed two jointly normal random variables previously in Section 5.3.2. In particular, two random variables X and Y are said to be **bivariate normal** or **jointly normal**, if $aX + bY$ has normal distribution for all $a, b \in \mathbb{R}$. We can extend this definition to n jointly normal random variables.

Random variables X_1, X_2,..., X_n are said to be **jointly normal** if, for all a_1,a_2,..., $a_n \in \mathbb{R}$, the random variable

$$a_1X_1 + a_2X_2 + \ldots + a_nX_n$$

is a normal random variable.

As before, we agree that the constant zero is a normal random variable with zero mean and variance, i.e., $N(0, 0)$. When we have several jointly normal random variables, we often put them in a vector. The resulting random vector is a called a normal (Gaussian) random vector.

A random vector

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ . \\ . \\ . \\ X_n \end{bmatrix}$$

is said to be **normal** or **Gaussian** if the random variables X_1, X_2,..., X_n are jointly normal.

To find the general form for the PDF of a Gaussian random vector it is convenient to start from the simplest case where the X_i's are independent and identically distributed (i.i.d.), $X_i \sim N(0,1)$. In this case, we know how to find the joint PDF. It is simply the product of the individual (marginal) PDFs. Let's call such a random vector the **standard normal random vector**. So, let

$$\mathbf{Z} = \begin{bmatrix} Z_1 \\ Z_2 \\ . \\ . \\ . \\ Z_n \end{bmatrix},$$

where Z_i's are i.i.d. and $Z_i \sim N(0,1)$. Then, we have

$$\begin{aligned} f_{\mathbf{Z}}(\mathbf{z}) &= f_{Z_1,Z_2,\ldots,Z_n}(z_1, z_2, \ldots, z_n) \\ &= \prod_{i=1}^{n} f_{Z_i}(z_i) \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} z_i^2\right\} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2}\mathbf{z}^T\mathbf{z}\right\}. \end{aligned}$$

For a standard normal random vector **Z**, where Z_i's are i.i.d. and $Z_i \sim N(0,1)$, the PDF is given by

$$f_{\mathbf{Z}}(\mathbf{z}) = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2}\mathbf{z}^T\mathbf{z}\right\}.$$

Now, we need to extend this formula to a general normal random vector **X** with mean $\mathbf{m}$ and covariance matrix **C**. This is very similar to when we defined general normal random variables from the standard normal random variable. We remember that if $Z \sim N(0,1)$, then the random variable $X = \sigma Z + \mu$ has $N(\mu, \sigma^2)$ distribution. We would like to do the same thing for normal random vectors.

Assume that I have a normal random vector **X** with mean $\mathbf{m}$ and covariance matrix **C**. We write $\mathbf{X} \sim N(\mathbf{m}, \mathbf{C})$. Further, assume that $\mathbf{C}$ is a positive definite matrix.(The positive definiteness assumption here does not create any limitations. We already know that $\mathbf{C}$ is positive semi-definite (Theorem 6.2), so $\det(\mathbf{C}) \geq 0$. We also know that $\mathbf{C}$ is positive definite if and only if $\det(\mathbf{C}) > 0$ (Theorem 6.3). So here, we are only excluding the case $\det(\mathbf{C}) = 0$. If $\det(\mathbf{C}) = 0$, then you can show that you can write some X_i's as a linear combination of others, so indeed we can remove them from the vector without losing any information.) Then from linear algebra we know that there exists an n by n matrix **Q** such that

$$\mathbf{Q}\mathbf{Q}^T = \mathbf{I} \quad (\mathbf{I} \text{ is the identity matrix}),$$
$$\mathbf{C} = \mathbf{Q}\mathbf{D}\mathbf{Q}^T,$$

where **D** is a diagonal matrix

$$D = \begin{bmatrix} d_{11} & 0 & \dots & 0 \\ 0 & d_{22} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & d_{nn} \end{bmatrix}.$$

The positive definiteness assumption guarantees that all d_{ii}'s are positive. Let's define

$$D^{\frac{1}{2}} = \begin{bmatrix} \sqrt{d_{11}} & 0 & \dots & 0 \\ 0 & \sqrt{d_{22}} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sqrt{d_{nn}} \end{bmatrix}.$$

We have $D^{\frac{1}{2}} D^{\frac{1}{2}} = \mathbf{D}$ and $D^{\frac{1}{2}} = D^{\frac{1}{2}^T}$. Also define

$$\mathbf{A} = Q D^{\frac{1}{2}} Q^T.$$

Then,

$$\mathbf{A}\mathbf{A}^T = \mathbf{A}^T\mathbf{A} = \mathbf{C}.$$

Now we are ready to define the transformation that converts a standard Gaussian vector to $\mathbf{X} \sim N(\mathbf{m}, \mathbf{C})$. Let Z be a standard Gaussian vector, i.e., $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I})$. Define

$$\mathbf{X} = \mathbf{A}\mathbf{Z} + \mathbf{m}.$$

We claim that $\mathbf{X} \sim N(\mathbf{m}, \mathbf{C})$. To see this, first note that $\mathbf{X}$ is a normal random vector. The reason is that any linear combination of components of $\mathbf{X}$ is indeed a linear combination of components of $\mathbf{Z}$ plus a constant. Thus, every linear combination of components of $\mathbf{X}$ is a normal random variable. It remains to show that $E\mathbf{X} = \mathbf{m}$ and $\mathbf{C}_{\mathbf{X}} = \mathbf{C}$. First note that by linearity of expectation we have

$$\begin{aligned} E\mathbf{X} &= E\left[\mathbf{A}\mathbf{Z} + \mathbf{m}\right] \\ &= \mathbf{A}E[\mathbf{Z}] + \mathbf{m} \\ &= \mathbf{m}. \end{aligned}$$

Also, by Example 6.12 we have

$$\begin{aligned} C_X &= AC_Z A^T \\ &= \mathbf{A}\mathbf{A}^T & \text{(since } C_Z = \mathbf{I}) \\ &= \mathbf{C}. \end{aligned}$$

Thus, we have shown that **X** is a random vector with mean $\mathbf{m}$ and covariance matrix **C**. Now we can use Example 6.15 to find the PDF of **X**. We have

$$
\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}) &= \frac{1}{|\det(\mathbf{A})|} f_{\mathbf{Z}}\left(\mathbf{A}^{-1}(\mathbf{x}-\mathbf{m})\right) \\
&= \frac{1}{(2\pi)^{\frac{n}{2}} |\det(\mathbf{A})|} \exp\left\{-\frac{1}{2}(\mathbf{A}^{-1}(\mathbf{x}-\mathbf{m}))^T(\mathbf{A}^{-1}(\mathbf{x}-\mathbf{m}))\right\} \\
&= \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det(\mathbf{C})}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T\mathbf{A}^{-T}\mathbf{A}^{-1}(\mathbf{x}-\mathbf{m})\right\} \\
&= \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det(\mathbf{C})}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{m})\right\}.
\end{aligned}
$$

For a normal random vector $\mathbf{X}$ with mean $\mathbf{m}$ and covariance matrix $\mathbf{C}$, the PDF is given by

$$
f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det \mathbf{C}}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{m})\right\} \qquad (6.1)
$$

Example 6.16

Let X and Y be two jointly normal random variables with $X \sim N(\mu_X, \sigma_X)$, $Y \sim N(\mu_Y, \sigma_Y)$, and $\rho(X, Y) = \rho$. Show that the above PDF formula for PDF of $\begin{bmatrix} X \\ Y \end{bmatrix}$ is the same as $f_{X,Y}(x, y)$ given in Definition 5.4 in Section 5.3.2. That is,

$$
\begin{aligned}
f_{XY}(x,y) = &\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \cdot \\
&\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]\right\}.
\end{aligned}
$$

Solution

Both formulas are in the form $ae^{-\frac{1}{2}b}$. Thus, it suffices to show that they have the same a and b. Here we have

$$
m = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}.
$$

We also have

$$C = \begin{bmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(Y,X) & \mathrm{Var}(Y) \end{bmatrix} = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}.$$

From this, we obtain

$$\det \mathbf{C} = \sigma_X^2\sigma_Y^2(1-\rho^2).$$

Thus, in both formulas for PDF a is given by

$$a = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}.$$

Next, we check b. We have

$$C^{-1} = \frac{1}{\sigma_X^2\sigma_Y^2(1-\rho^2)}\begin{bmatrix} \sigma_Y^2 & -\rho\sigma_X\sigma_Y \\ -\rho\sigma_X\sigma_Y & \sigma_X^2 \end{bmatrix}.$$

Now by matrix multiplication we obtain

$$\begin{aligned} &(x-m)^T\mathbf{C}^{-1}(x-m) = \\ &\qquad = \frac{1}{\sigma_X^2\sigma_Y^2(1-\rho^2)}\begin{bmatrix} x-\mu_X \\ y-\mu_Y \end{bmatrix}^T \begin{bmatrix} \sigma_Y^2 & -\rho\sigma_X\sigma_Y \\ -\rho\sigma_X\sigma_Y & \sigma_X^2 \end{bmatrix}\begin{bmatrix} x-\mu_X \\ y-\mu_Y \end{bmatrix} \\ &\qquad = -\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right], \end{aligned}$$

which agrees with the formula in Definition 5.4.

Remember that two jointly normal random variables X and Y are independent if and only if they are uncorrelated. We can extend this to multiple jointly normal random variables. Thus, if you have a normal random vector whose components are uncorrelated, you can conclude that the components are independent. To show this, note that if the X_i's are uncorrelated, then the covariance matrix $\mathbf{C_X}$ is diagonal, so its inverse $\mathbf{C_X^{-1}}$ is also diagonal. You can see that in this case the PDF (Equation 6.1) becomes the products of marginal PDFs.

If $\mathbf{X} = [X_1, X_2, \ldots, X_n]^T$ is a normal random vector, and we know $\mathrm{Cov}(X_i, X_j) = 0$ for all $i \neq j$, then X_1,X_2, ..., X_n are independent.

Another important result is that if $\mathbf{X} = [X_1, X_2, \ldots, X_n]^T$ is a normal random vector then $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$ is also a random vector because any linear combination of components of $\mathbf{Y}$ is also a linear combination of components of $\mathbf{X}$ plus a constant value.

If $\mathbf{X} = [X_1, X_2, \ldots, X_n]^T$ is a normal random vector, $\mathbf{X} \sim N(\mathbf{m}, \mathbf{C})$, $\mathbf{A}$ is an m by n fixed matrix, and $\mathbf{b}$ is an m-dimensional fixed vector, then the random vector $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$ is a normal random vector with mean $\mathbf{A}E\mathbf{X} + \mathbf{b}$ and covariance matrix $\mathbf{ACA}^T$.

$$\mathbf{Y} \sim N(\mathbf{A}E\mathbf{X} + \mathbf{b}, \mathbf{ACA}^T)$$