5

The Multivariate Normal Distribution

1 Preliminaries from Linear Algebra

In Chapter 1 we studied how to handle (linear transformations of) random vectors, that is, vectors whose components are random variables. Since the normal distribution is (one of) the most important distribution(s) and since there are special properties, methods, and devices pertaining to this distribution, we devote this chapter to the study of the *multivariate normal distribution*, or, equivalently, to the study of *normal random vectors*. We show, for example, that the sample mean and the sample variance in a (one-dimensional) sample are independent, a property that, in fact, characterizes this distribution and is essential, for example, in the so called t-test, which is used to test hypotheses about the mean in the (univariate) normal distribution when the variance is unknown. In fact, along the way we will encounter three different ways to show this independence. Another interesting fact that will be established is that if the components of a normal random vector are uncorrelated, then they are in fact independent. One section is devoted to quadratic forms of normal random vectors, which are of great importance in many branches of statistics. The main result, Cohran's theorem, states that, under certain conditions, one can split the sum of the squares of the observations into a number of quadratic forms, each of them pertaining to some cause of variation in an experiment in such a way that these quadratic forms are independent, and (essentially) χ^2-distributed random variables. This can be used to test whether or not a certain cause of variation influences the outcome of the experiment. For more on the statistical aspects, we refer to the literature cited in Appendix A.

We begin, however, by recalling some basic facts from linear algebra. Vectors are always column vectors (recall Remark 1.1.2). For convenience, however, we sometimes write $\mathbf{x} = (x_1, x_2, \ldots, x_n)'$. A square matrix $\mathbf{A} = \{a_{ij},\ i, j = 1, 2, \ldots, n\}$ is *symmetric* if $a_{ij} = a_{ji}$ and all elements are real. All *eigenvalues* of a real, symmetric matrix are real. In this chapter all matrices are real.

A. Gut, *An Intermediate course in Probabilty,* Springer Texts in Statistics,
DOI: 10.1007/978-1-4419-0162-0_5,

A square matrix $\mathbf{C}$ is orthogonal if $\mathbf{C}'\mathbf{C} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. Note that since, trivially, $\mathbf{C}^{-1}\mathbf{C} = \mathbf{C}\mathbf{C}^{-1} = \mathbf{I}$, it follows that

$$\mathbf{C}^{-1} = \mathbf{C}'. \tag{1.1}$$

Moreover, $\det \mathbf{C} = \pm 1$.

Remark 1.1. Orthogonality means that the rows (and columns) of an orthogonal matrix, considered as vectors, are orthonormal, that is, they have length 1 and are orthogonal; the scalar products between them are zero. □

Let $\mathbf{x}$ be an n-vector, let $\mathbf{C}$ be an orthogonal $n \times n$ matrix, and set $\mathbf{y} = \mathbf{C}\mathbf{x}$; $\mathbf{y}$ is also an n-vector. A consequence of the orthogonality is that $\mathbf{x}$ and $\mathbf{y}$ have the same length. Indeed,

$$\mathbf{y}'\mathbf{y} = (\mathbf{C}\mathbf{x})'\mathbf{C}\mathbf{x} = \mathbf{x}'\mathbf{C}'\mathbf{C}\mathbf{x} = \mathbf{x}'\mathbf{x}. \tag{1.2}$$

Now, let $\mathbf{A}$ be a symmetric matrix. A fundamental result is that there exists an orthogonal matrix $\mathbf{C}$ such that

$$\mathbf{C}'\mathbf{A}\mathbf{C} = \mathbf{D}, \tag{1.3}$$

where $\mathbf{D}$ is a diagonal matrix, the elements of the diagonal being the eigenvalues, $\lambda_1, \lambda_2, \ldots, \lambda_n$, of $\mathbf{A}$. It also follows that

$$\det \mathbf{A} = \det \mathbf{D} = \prod_{k=1}^{n} \lambda_k. \tag{1.4}$$

A *quadratic form* $Q = Q(\mathbf{x})$ based on the symmetric matrix $\mathbf{A}$ is defined by

$$Q(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x} \quad \Big(= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j\Big), \quad \mathbf{x} \in \mathbf{R}^n. \tag{1.5}$$

Q is *positive-definite* if $Q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$ and *nonnegative-definite* (positive-semidefinite) if $Q(\mathbf{x}) \geq 0$ for all $\mathbf{x}$.

One can show that Q is positive- (nonnegative-)definite iff all eigenvalues are positive (nonnegative). Another useful criterion is to check all subdeterminants of $\mathbf{A}$, that is, $\det \mathbf{A}_k$, where $\mathbf{A}_k = \{a_{ij},\ i, j = 1, 2, \ldots, k\}$ and $k = 1, 2, \ldots, n$. Then Q is positive- (nonnegative-)definite iff $\det \mathbf{A}_k > 0$ (≥ 0) for all $k = 1, 2, \ldots, n$.

A matrix is positive- (nonnegative-)definite iff the corresponding quadratic form is positive- (nonnegative-)definite.

Now, let $\mathbf{A}$ be a square matrix whose inverse exists. The *algebraic complement* $\mathbf{A}_{ij}$ of the element a_{ij} is defined as the matrix that remains after deleting the ith row and the jth column of $\mathbf{A}$. For the element a_{ij}^{-1} of the inverse $\mathbf{A}^{-1}$ of $\mathbf{A}$, we have

$$a_{ij}^{-1} = (-1)^{i+j} \frac{\det \mathbf{A}_{ji}}{\det \mathbf{A}}. \tag{1.6}$$

In particular, if $\mathbf{A}$ is symmetric, it follows that $\mathbf{A}_{ij} = \mathbf{A}'_{ji}$, from which we conclude that $\det \mathbf{A}_{ij} = \det \mathbf{A}_{ji}$ and hence that $a_{ij}^{-1} = a_{ji}^{-1}$ and that $\mathbf{A}^{-1}$ is symmetric.

Finally, we need to define the square root of a nonnegative-definite symmetric matrix. For a diagonal matrix $\mathbf{D}$ it is easy to see that the diagonal matrix whose diagonal elements are the square roots of those of $\mathbf{D}$ has the property that the square equals $\mathbf{D}$. For the general case we know, from (1.3), that there exists an orthogonal matrix $\mathbf{C}$ such that $\mathbf{C}'\mathbf{A}\mathbf{C} = \mathbf{D}$, that is, such that

$$\mathbf{A} = \mathbf{C}\mathbf{D}\mathbf{C}', \tag{1.7}$$

where $\mathbf{D}$ is the diagonal matrix whose diagonal elements are the eigenvalues of $\mathbf{A}$; $d_{ii} = \lambda_i$, $i = 1, 2, \ldots, n$.

Let us denote the square root of $\mathbf{D}$, as described above, by $\widetilde{\mathbf{D}}$. We thus have $\widetilde{d}_{ii} = \sqrt{\lambda_i}$, $i = 1, 2, \ldots, n$ and $\widetilde{\mathbf{D}}^2 = \mathbf{D}$. Set $\mathbf{B} = \mathbf{C}\widetilde{\mathbf{D}}\mathbf{C}'$. Then

$$\mathbf{B}^2 = \mathbf{B}\mathbf{B} = \mathbf{C}\widetilde{\mathbf{D}}\mathbf{C}'\mathbf{C}\widetilde{\mathbf{D}}\mathbf{C}' = \mathbf{C}\widetilde{\mathbf{D}}\widetilde{\mathbf{D}}\mathbf{C}' = \mathbf{C}\mathbf{D}\mathbf{C}' = \mathbf{A}\,, \tag{1.8}$$

that is, $\mathbf{B}$ is a square root of $\mathbf{A}$. A common notation is $\mathbf{A}^{1/2}$.

Now, this holds true for any of the 2^n choices of square roots. However, in order to ensure that the square root is nonnegative-definite we tacitly assume in the following that the nonnegative square root of the eigenvalues has been chosen, viz., that throughout $\widetilde{d}_{ii} = +\sqrt{\lambda_i}$.

If, in addition, $\mathbf{A}$ has an inverse, one can show that

$$(\mathbf{A}^{-1})^{1/2} = (\mathbf{A}^{1/2})^{-1}, \tag{1.9}$$

which is denoted by $\mathbf{A}^{-1/2}$.

Exercise 1.1. Verify formula (1.9).

Exercise 1.2. Show that $\det \mathbf{A}^{-1/2} = (\det \mathbf{A})^{-1/2}$. □

Remark 1.2. The reader who is less used to working with vectors and matrices might like to spell out certain formulas explicitly as sums or double sums, and so forth. □

2 The Covariance Matrix

Let $\mathbf{X}$ be a random n-vector whose components have finite variance.

Definition 2.1. *The* mean vector *of* $\mathbf{X}$ *is* $\boldsymbol{\mu} = E\,\mathbf{X}$, *the components of which are* $\mu_i = E\,X_i$, $i = 1, 2, \ldots, n$.

The covariance matrix *of* $\mathbf{X}$ *is* $\boldsymbol{\Lambda} = E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'$, *whose elements are* $\lambda_{ij} = E(X_i - \mu_i)(X_j - \mu_j)$, $i, j = 1, 2, \ldots, n$. □

Thus, $\lambda_{ii} = \operatorname{Var} X_i$, $i = 1, 2, \ldots, n$, and $\lambda_{ij} = \operatorname{Cov}(X_i, X_j) = \lambda_{ji}$, $i, j = 1, 2, \ldots, n$ (and $i \neq j$, or else $\operatorname{Cov}(X_i, X_i) = \operatorname{Var} X_i$). In particular, every covariance matrix is symmetric.

Theorem 2.1. *Every covariance matrix is nonnegative-definite.*

Proof. The proof is immediate from the fact that, for any $\mathbf{y} \in \mathbf{R}^n$,

$$Q(\mathbf{y}) = \mathbf{y}'\mathbf{\Lambda}\mathbf{y} = \mathbf{y}'E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'\mathbf{y} = \operatorname{Var}\left(\mathbf{y}'(\mathbf{X} - \boldsymbol{\mu})\right) \geq 0. \qquad \square$$

Remark 2.1. If $\det \mathbf{\Lambda} > 0$, the probability distribution of $\mathbf{X}$ is truly n-dimensional in the sense that it cannot be concentrated on a subspace of lower dimension. If $\det \mathbf{\Lambda} = 0$ it can be concentrated on such a subspace; we call it the *singular* case (as opposed to the nonsingular case). □

Next we consider linear transformations.

Theorem 2.2. *Let* $\mathbf{X}$ *be a random* n*-vector with mean vector* $\boldsymbol{\mu}$ *and covariance matrix* $\mathbf{\Lambda}$*. Further, let* $\mathbf{B}$ *be an* $m \times n$ *matrix, let* $\mathbf{b}$ *be a constant* m*-vector, and set* $\mathbf{Y} = \mathbf{B}\mathbf{X} + \mathbf{b}$*. Then*

$$E\,\mathbf{Y} = \mathbf{B}\boldsymbol{\mu} + \mathbf{b} \quad \textit{and} \quad \operatorname{Cov}\mathbf{Y} = \mathbf{B}\mathbf{\Lambda}\mathbf{B}'.$$

Proof. We have

$$E\,\mathbf{Y} = \mathbf{B}E\,\mathbf{X} + \mathbf{b} = \mathbf{B}\boldsymbol{\mu} + \mathbf{b}$$

and

$$\begin{aligned} \operatorname{Cov}\mathbf{Y} &= E(\mathbf{Y} - E\,\mathbf{Y})(\mathbf{Y} - E\,\mathbf{Y})' = E\,\mathbf{B}(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'\mathbf{B}' \\ &= \mathbf{B}E\big\{(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'\big\}\mathbf{B}' = \mathbf{B}\mathbf{\Lambda}\mathbf{B}'. \end{aligned} \qquad \square$$

Remark 2.2. Note that for $n = 1$ the theorem reduces to the well-known facts $E\,Y = aE\,X + b$ and $\operatorname{Var} Y = a^2 \operatorname{Var} X$ (where $Y = aX + b$).

Remark 2.3. We will permit ourselves, at times, to be somewhat careless about specifying dimensions of matrices and vectors. It will always be tacitly understood that the dimensions are compatible with the arithmetic of the situation at hand. □

3 A First Definition

We will provide three definitions of the multivariate normal distribution. In this section we present the first one, which states that a random vector is normal iff every linear combination of its components is normal. In Section 4 we provide a definition based on the characteristic function, and in Section 5 we give a definition based on the density function. We also prove that the first two definitions are always equivalent (i.e., when the covariance matrix is nonnegative-definite) and that the three of them are equivalent in the nonsingular case (i.e., when the covariance matrix is positive-definite). A fourth definition is given in Problem 10.1.

Definition I. *The random n-vector* $\mathbf{X}$ *is normal iff, for every n-vector* $\mathbf{a}$*, the (one-dimensional) random variable* $\mathbf{a}'\mathbf{X}$ *is normal. The notation* $\mathbf{X} \in N(\boldsymbol{\mu}, \mathbf{\Lambda})$ *is used to denote that* $\mathbf{X}$ *has a (multivariate) normal distribution with mean vector* $\boldsymbol{\mu}$ *and covariance matrix* $\mathbf{\Lambda}$. □

Remark 3.1. The actual distribution of $\mathbf{a}'\mathbf{X}$ depends, of course, on $\mathbf{a}$. The degenerate normal distribution (meaning variance equal to zero) is also included as a possible distribution of $\mathbf{a}'\mathbf{X}$.

Remark 3.2. Note that no assumption whatsoever is made about independence between the components of $\mathbf{X}$. □

Surprisingly enough, this somewhat abstract definition is extremely applicable and useful. Moreover, several proofs, which otherwise become complicated, become very "simple" (and beautiful). For example, the following three properties are immediate consequences of this definition:

(a) Every component of $\mathbf{X}$ is normal.
(b) $X_1 + X_2 + \cdots + X_n$ is normal.
(c) Every marginal distribution is normal.

Indeed, to see that X_k is normal for $k = 1, 2, \ldots, n$, we choose $\mathbf{a}$ such that $a_k = 1$ and $a_j = 0$ otherwise.

To see that the sum of all components is normal, we simply choose $a_k = 1$ for all k.

As for (c) we argue as follows: To show that $(X_{i_1}, X_{i_2}, \ldots, X_{i_k})'$ is normal for some $k = (1,)\,2, \ldots, n-1$, amounts to checking that all linear combinations of these components are normal. However, since we know that $\mathbf{X}$ is normal, we know that $\mathbf{a}'\mathbf{X}$ is normal for *every* $\mathbf{a}$, in particular for all $\mathbf{a}$, such that $a_j = 0$ for $j \neq i_1, i_2, \ldots, i_k$, which establishes the desired conclusion.

We also observe that, from a first course in probability theory, we know that any linear combination of *independent* normal random variables is normal (via the convolution formula and/or the moment generating function—recall Theorem 3.3.2), that is, the condition in Definition I is satisfied. It follows, in particular, that

(d) if $\mathbf{X}$ has independent normal components, then $\mathbf{X}$ is normal.

Another important result is as follows:

Theorem 3.1. *Suppose that* $\mathbf{X} \in N(\boldsymbol{\mu}, \mathbf{\Lambda})$ *and set* $\mathbf{Y} = \mathbf{B}\mathbf{X} + \mathbf{b}$. *Then* $\mathbf{Y} \in N(\mathbf{B}\boldsymbol{\mu} + \mathbf{b}, \mathbf{B}\mathbf{\Lambda}\mathbf{B}')$.

Proof. The first part of the proof merely amounts to establishing the fact that a linear combination of the components of $\mathbf{Y}$ is a (some other) linear combination of the components of $\mathbf{X}$. Namely, we wish to show that $\mathbf{a}'\mathbf{Y}$ is normal for every $\mathbf{a}$. However,

$$\mathbf{a}'\mathbf{Y} = \mathbf{a}'\mathbf{B}\mathbf{X} + \mathbf{a}'\mathbf{b} = (\mathbf{B}'\mathbf{a})'\mathbf{X} + \mathbf{a}'\mathbf{b} = \mathbf{c}'\mathbf{X} + d, \tag{3.1}$$

where $\mathbf{c} = \mathbf{B}'\mathbf{a}$ and $d = \mathbf{a}'\mathbf{b}$. Since $\mathbf{c}'\mathbf{X}$ is normal according to Definition I (and d is a constant), it follows that $\mathbf{a}'\mathbf{Y}$ is normal. The correctness of the parameters follows from Theorem 2.2. $\square$

Exercise 3.1. Let X_1, X_2, X_3, and X_4 be independent, $N(0,1)$-distributed random variables. Set $Y_1 = X_1 + 2X_2 + 3X_3 + 4X_4$ and $Y_2 = 4X_1 + 3X_2 + 2X_3 + X_4$. Determine the distribution of $\mathbf{Y}$.

Exercise 3.2. Let $\mathbf{X} \in N + \left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & -2 \\ -2 & 7 \end{pmatrix} \right)$. Set

$$Y_1 = X_1 + X_2 \quad \text{and} \quad Y_2 = 2X_1 - 3X_2.$$

Determine the distribution of $\mathbf{Y}$. $\square$

A word of caution is appropriate at this point. We noted above that all marginal distributions of a normal random vector $\mathbf{X}$ are normal. The *joint normality* of all components of $\mathbf{X}$ was essential here. In the following example we define two random variables that are normal but not jointly normal. This shows that a general converse does not hold; there exist normal random variables that are not jointly normal.

Example 3.1. Let $X \in N(0,1)$ and let Z be independent of X and such that $P(Z = 1) = P(Z = -1) = 1/2$. Set $Y = Z \cdot X$. Then

$$P(Y \le x) = \frac{1}{2}P(X \le x) + \frac{1}{2}P(-X \le x) = \frac{1}{2}\Phi(x) + \frac{1}{2}\big(1 - \Phi(-x)\big) = \Phi(x)\,,$$

that is, $Y \in N(0,1)$. Thus, X and Y are both (standard) normal. However, since

$$P(X + Y = 0) = P(Z = -1) = \frac{1}{2},$$

it follows from Definition I that $X + Y$ cannot be normal and, hence, that $(X, Y)'$ is not normal. $\square$

For a further example, see Problem 10.7.

Another kind of converse one might consider is the following. An obvious consequence of Theorem 3.1 is that if $\mathbf{X} \in N(\boldsymbol{\mu}, \boldsymbol{\Lambda})$, and if the matrices $\mathbf{A}$ and $\mathbf{B}$ are such that $\mathbf{A} = \mathbf{B}$, then $\mathbf{AX} \overset{d}{=} \mathbf{BX}$. A natural question is whether or not the converse holds, viz., if $\mathbf{AX} \overset{d}{=} \mathbf{BX}$, does it then follow that $\mathbf{A} = \mathbf{B}$?

Exercise 3.3. Let X_1 and X_2 be independent standard normal random variables and put

$$Y_1 = X_1 + X_2, \quad Y_2 = 2X_1 + X_2 \quad \text{and} \quad Z_1 = X_1\sqrt{2}, \quad Z_2 = \frac{3}{\sqrt{2}}X_1 + \frac{1}{\sqrt{2}}X_2.$$

(a) Determine the corresponding matrices $\mathbf{A}$ and $\mathbf{B}$?
(b) Check that $\mathbf{A} \neq \mathbf{B}$.
(c) Show that (nevertheless) $\mathbf{Y}$ and $\mathbf{Z}$ are have the same normal distribution (which one?). $\square$

4 The Characteristic Function: Another Definition

The characteristic function of a random vector $\mathbf{X}$ is (recall Definition 3.4.2)

$$\varphi_{\mathbf{X}}(\mathbf{t}) = E\, e^{i\mathbf{t}'\mathbf{X}}. \tag{4.1}$$

Now, suppose that $\mathbf{X} \in N(\boldsymbol{\mu}, \boldsymbol{\Lambda})$. We observe that $Z = \mathbf{t}'\mathbf{X}$ in (4.1) has a one-dimensional normal distribution by Definition I. The parameters are $m = E\, Z = \mathbf{t}'\boldsymbol{\mu}$ and $\sigma^2 = \operatorname{Var} Z = \mathbf{t}'\boldsymbol{\Lambda}\mathbf{t}$. Since

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \varphi_Z(1) = \exp\{im - \tfrac{1}{2}\sigma^2\}\,, \tag{4.2}$$

we have established the following result:

Theorem 4.1. *For* $\mathbf{X} \in N(\boldsymbol{\mu}, \boldsymbol{\Lambda})$, *we have*

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \exp\{i\mathbf{t}'\boldsymbol{\mu} - \tfrac{1}{2}\mathbf{t}'\boldsymbol{\Lambda}\mathbf{t}\}.$$

□

It turns out that we can, in fact, establish a converse to this result and thereby obtain another, equivalent, definition of the multivariate normal distribution. We therefore temporarily "forget" the above and begin by proving the following fact:

Lemma 4.1. *For any nonnegative-definite symmetric matrix* $\boldsymbol{\Lambda}$, *the function*

$$\varphi^*(\mathbf{t}) = \exp\{i\mathbf{t}'\boldsymbol{\mu} - \tfrac{1}{2}\mathbf{t}'\boldsymbol{\Lambda}\mathbf{t}\}$$

is the characteristic function of a random vector $\mathbf{X}$ *with* $E\,\mathbf{X} = \boldsymbol{\mu}$ *and* $\operatorname{Cov}\mathbf{X} = \boldsymbol{\Lambda}$.

Proof. Let $\mathbf{Y}$ be a random vector whose components $Y_1, Y_2, \ldots, Y_n$ are independent, $N(0,1)$-distributed random variables, and set

$$\mathbf{X} = \boldsymbol{\Lambda}^{1/2}\mathbf{Y} + \boldsymbol{\mu}. \tag{4.3}$$

Since $\operatorname{Cov}\mathbf{Y} = \mathbf{I}$, it follows from Theorem 2.2 that

$$E\,\mathbf{X} = \boldsymbol{\mu} \quad \text{and} \quad \operatorname{Cov}\mathbf{X} = \boldsymbol{\Lambda}. \tag{4.4}$$

Furthermore, an easy computation shows that

$$\varphi_{\mathbf{Y}}(\mathbf{t}) = E\exp\{i\mathbf{t}'\mathbf{Y}\} = \exp\{-\tfrac{1}{2}\mathbf{t}'\mathbf{t}\}. \tag{4.5}$$

It finally follows that

$$\begin{aligned}
\varphi_{\mathbf{X}}(\mathbf{t}) &= E\exp\{i\mathbf{t}'\mathbf{X}\} = E\exp\{i\mathbf{t}'(\boldsymbol{\Lambda}^{1/2}\mathbf{Y} + \boldsymbol{\mu})\} \\
&= \exp\{i\mathbf{t}'\boldsymbol{\mu}\} \cdot E\exp\{i\mathbf{t}'\boldsymbol{\Lambda}^{1/2}\mathbf{Y}\} \\
&= \exp\{i\mathbf{t}'\boldsymbol{\mu}\} \cdot E\exp\{i(\boldsymbol{\Lambda}^{1/2}\mathbf{t})'\mathbf{Y}\} \\
&= \exp\{i\mathbf{t}'\boldsymbol{\mu}\} \cdot \varphi_{\mathbf{Y}}(\boldsymbol{\Lambda}^{1/2}\mathbf{t}) \\
&= \exp\{i\mathbf{t}'\boldsymbol{\mu}\} \cdot \exp\big\{-\tfrac{1}{2}(\boldsymbol{\Lambda}^{1/2}\mathbf{t})'(\boldsymbol{\Lambda}^{1/2}\mathbf{t})\big\} \\
&= \exp\big\{i\mathbf{t}'\boldsymbol{\mu} - \tfrac{1}{2}\mathbf{t}'\boldsymbol{\Lambda}\mathbf{t}\big\},
\end{aligned}$$

as desired. □

Note that at this point we do not (yet) know that $\mathbf{X}$ is normal.

The next step is to show that if $\mathbf{X}$ has a characteristic function given as in the lemma, then $\mathbf{X}$ is normal in the sense of Definition I. Thus, let $\mathbf{X}$ be given as described and let $\mathbf{a}$ be an arbitrary n-vector. Then

$$\begin{aligned}\varphi_{\mathbf{a}'\mathbf{X}}(u) &= E\exp\{iu\,\mathbf{a}'\mathbf{X}\} = \varphi_{\mathbf{X}}(u\mathbf{a})\\ &= \exp\bigl\{i(u\mathbf{a})'\boldsymbol{\mu} - \tfrac{1}{2}(u\mathbf{a})'\boldsymbol{\Lambda}(u\mathbf{a})\bigr\}\\ &= \exp\{ium - \tfrac{1}{2}u^2\sigma^2\}\,,\end{aligned}$$

where $m = \mathbf{a}'\boldsymbol{\mu}$ and $\sigma^2 = \mathbf{a}'\boldsymbol{\Lambda}\mathbf{a} \geq 0$, which proves that $\mathbf{a}'\mathbf{X} \in N(m, \sigma^2)$ and hence that $\mathbf{X}$ is normal in the sense of Definition I.

Alternatively, we may argue as in the proof of Theorem 3.1:

$$\mathbf{a}'\mathbf{X} = \mathbf{a}'\bigl(\boldsymbol{\Lambda}^{1/2}\mathbf{Y} + \boldsymbol{\mu}\bigr) = \mathbf{a}'\boldsymbol{\Lambda}^{1/2}\mathbf{Y} + \mathbf{a}'\boldsymbol{\mu} = \bigl(\boldsymbol{\Lambda}^{1/2}\mathbf{a}\bigr)'\mathbf{Y} + \mathbf{a}'\boldsymbol{\mu}\,,$$

which shows that a linear combination of the components of $\mathbf{X}$ is equal to (another) linear combination of the components of $\mathbf{Y}$, which, in turn, we know is normal, since $\mathbf{Y}$ has *independent components*.

We have thus shown that the function defined in Lemma 4.1 is, indeed, a characteristic function and that the linear combinations of the components of the corresponding random vector are normal. This motivates the following alternative definition of the multivariate normal distribution.

Definition II. *A random vector* $\mathbf{X}$ *is normal iff its characteristic function is of the form*

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \exp\{i\mathbf{t}'\boldsymbol{\mu} - \tfrac{1}{2}\mathbf{t}'\boldsymbol{\Lambda}\mathbf{t}\}\,,$$

for some vector $\boldsymbol{\mu}$ *and nonnegative-definite matrix* $\boldsymbol{\Lambda}$. □

We have also established the following fact:

Theorem 4.2. *Definitions I and II are equivalent.* □

Remark 4.1. The definition and expression for the moment generating function are the obvious ones:

$$\psi_{\mathbf{X}}(\mathbf{t}) = E\,e^{\mathbf{t}'\mathbf{X}} = \exp\bigl\{\mathbf{t}'\boldsymbol{\mu} + \tfrac{1}{2}\mathbf{t}'\boldsymbol{\Lambda}\mathbf{t}\bigr\}.$$ □

Exercise 4.1. Suppose that $\mathbf{X} = (X_1, X_2)'$ has characteristic function

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \exp\{it_1 + 2it_2 - \tfrac{1}{2}t_1^2 + 2t_1t_2 - 6t_2^2\}.$$

Determine the distribution of $\mathbf{X}$.

Exercise 4.2. Suppose that $\mathbf{X} = (X_1, X_2)'$ has characteristic function

$$\varphi(t, u) = \exp\{it - 2t^2 - u^2 - tu\}.$$

Find the distribution of $X_1 + X_2$.

Exercise 4.3. Suppose that X and Y have a (joint) moment generating function given by

$$\psi(t, u) = \exp\{t^2 + 2tu + 4u^2\}.$$

Compute $P(2X < Y + 2)$. □

5 The Density: A Third Definition

Let $\mathbf{X} \in N(\boldsymbol{\mu}, \boldsymbol{\Lambda})$. If $\det \boldsymbol{\Lambda} = 0$, the distribution is singular, as mentioned before, and no density exists. If, however, $\det \boldsymbol{\Lambda} > 0$, then there exists a density function that, moreover, is uniquely determined by the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Lambda}$.

In order to determine the density, it is therefore sufficient to find it for a normal distribution constructed in some convenient way. To this end, let $\mathbf{Y}$ and $\mathbf{X}$ be defined as in the proof of Lemma 4.1, that is, $\mathbf{Y}$ has independent, standard normal components and $\mathbf{X} = \boldsymbol{\Lambda}^{1/2}\mathbf{Y} + \boldsymbol{\mu}$. Then $\mathbf{X} \in N(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ by Theorem 3.1, as desired.

Now, since the density of $\mathbf{Y}$ is known, it is easy to compute the density of $\mathbf{X}$ with the aid of the transformation theorem. Namely,

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}) &= \prod_{k=1}^{n} f_{Y_k}(y_k) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-y_k^2/2} \\ &= \Big(\frac{1}{2\pi}\Big)^{n/2} e^{-\frac{1}{2}\sum_{k=1}^{n} y_k^2} = \Big(\frac{1}{2\pi}\Big)^{n/2} e^{-\frac{1}{2}\mathbf{y}'\mathbf{y}}, \quad \mathbf{y} \in \mathbf{R}^n. \end{aligned}$$

Further, since $\det \boldsymbol{\Lambda} > 0$, we know that the inverse $\boldsymbol{\Lambda}^{-1}$ exists, that

$$\mathbf{Y} = \boldsymbol{\Lambda}^{-1/2}(\mathbf{X} - \boldsymbol{\mu}), \tag{5.1}$$

and hence that the Jacobian is $\det \boldsymbol{\Lambda}^{-1/2} = (\det \boldsymbol{\Lambda})^{-1/2}$ (Exercise 1.2). The following result emerges.

Theorem 5.1. *For* $\mathbf{X} \in N(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ *with* $\det \boldsymbol{\Lambda} > 0$, *we have*

$$f_{\mathbf{X}}(\mathbf{x}) = \Big(\frac{1}{2\pi}\Big)^{n/2} \frac{1}{\sqrt{\det \boldsymbol{\Lambda}}} \exp\big\{-\tfrac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Lambda}^{-1}(\mathbf{x} - \boldsymbol{\mu})\big\}. \qquad \square$$

Exercise 5.1. We have tacitly used the fact that if $\mathbf{X}$ is a random vector and $\mathbf{Y} = \mathbf{BX}$ then

$$\Big|\frac{d(\mathbf{y})}{d(\mathbf{x})}\Big| = \det \mathbf{B}.$$

Prove that this is correct. $\square$

We are now ready for our third definition.

Definition III. *A random vector* $\mathbf{X}$ *with* $E\,\mathbf{X} = \boldsymbol{\mu}$ *and* $\mathrm{Cov}\mathbf{X} = \boldsymbol{\Lambda}$, *such that* $\det \boldsymbol{\Lambda} > 0$, *is* $N(\boldsymbol{\mu}, \boldsymbol{\Lambda})$*-distributed iff the density equals*

$$f_{\mathbf{X}}(\mathbf{x}) = \Big(\frac{1}{2\pi}\Big)^{n/2} \frac{1}{\sqrt{\det \boldsymbol{\Lambda}}} \exp\big\{-\tfrac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Lambda}^{-1}(\mathbf{x} - \boldsymbol{\mu})\big\}, \quad \mathbf{x} \in \mathbf{R}^n. \qquad \square$$

Theorem 5.2. *Definitions I, II, and III are equivalent (in the nonsingular case).*

Proof. The equivalence of Definitions I and II was established in Section 4. The equivalence of Definitions II and III (in the nonsingular case) is a consequence of the uniqueness theorem for characteristic functions. □

Now let us see how the density function can be computed explicitly. Let $\mathbf{\Lambda}_{ij}$ be the algebraic complement of $\lambda_{ij} = \operatorname{Cov}(X_i, X_j)$ and set $\triangle_{ij} = (-1)^{i+j}\det \mathbf{\Lambda}_{ij}$ ($= \triangle_{ji}$, since $\mathbf{\Lambda}$ is symmetric). Since the elements of $\mathbf{\Lambda}^{-1}$ are $\triangle_{ij}/\triangle$, $i, j = 1, 2, \ldots, n$, where $\triangle = \det \mathbf{\Lambda}$, it follows that

$$f_{\mathbf{X}}(\mathbf{x}) = \Big(\frac{1}{2\pi}\Big)^{n/2} \frac{1}{\sqrt{\triangle}} \exp\Big\{-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\triangle_{ij}}{\triangle}(x_i - \mu_i)(x_j - \mu_j)\Big\}. \tag{5.2}$$

In particular, the following holds for the case $n = 2$: Set $\mu_k = E\,X_k$ and $\sigma_k^2 = \operatorname{Var} X_k$, $k = 1, 2$, and $\sigma_{12} = \operatorname{Cov}(X_1, X_2)$, and let $\rho = \sigma_{12}/\sigma_1\sigma_2$ be the correlation coefficient, where $|\rho| < 1$ (since $\det \mathbf{\Lambda} > 0$). Then $\triangle = \sigma_1^2\sigma_2^2(1-\rho^2)$, $\triangle_{11} = \sigma_2^2$, $\triangle_{22} = \sigma_1^2$, $\triangle_{12} = \triangle_{21} = -\rho\sigma_1\sigma_2$, and hence

$$\mathbf{\Lambda} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \quad \text{and} \quad \mathbf{\Lambda}^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} \dfrac{1}{\sigma_1^2} & -\dfrac{\rho}{\sigma_1\sigma_2} \\ -\dfrac{\rho}{\sigma_1\sigma_2} & \dfrac{1}{\sigma_2^2} \end{pmatrix}.$$

It follows that

$$\begin{aligned} f_{X_1,X_2}(x_1, x_2) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \\ &\times \exp\Big\{-\frac{1}{2(1-\rho^2)}\Big(\big(\frac{x_1-\mu_1}{\sigma_1}\big)^2 - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \big(\frac{x_2-\mu_2}{\sigma_2}\big)^2\Big)\Big\}. \end{aligned}$$

Exercise 5.2. Let the (joint) moment generating function of $\mathbf{X}$ be

$$\psi(t, u) = \exp\{t^2 + 3tu + 4u^2\}.$$

Determine the density function of $\mathbf{X}$.

Exercise 5.3. Suppose that $\mathbf{X} \in N(\mathbf{0}, \mathbf{\Lambda})$, where

$$\mathbf{\Lambda} = \begin{pmatrix} \frac{7}{2} & \frac{1}{2} & -1 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ -1 & 0 & \frac{1}{2} \end{pmatrix}.$$

Put $Y_1 = X_2 + X_3$, $Y_2 = X_1 + X_3$, and $Y_3 = X_1 + X_2$. Determine the density function of $\mathbf{Y}$. □

6 Conditional Distributions

Let $\mathbf{X} \in N(\boldsymbol{\mu}, \boldsymbol{\Lambda})$, and suppose that $\det \boldsymbol{\Lambda} > 0$. The density thus exists as given in Section 5. Conditional densities are defined (Chapter 2) as the ratio of the relevant joint and marginal densities. One can show that all marginal distributions of a nonsingular normal distribution are nonsingular and hence possess densities.

Let us consider the case $n = 2$ in some detail. Suppose that $(X, Y)' \in N(\boldsymbol{\mu}, \boldsymbol{\Lambda})$, where $E\,X = \mu_x$, $E\,Y = \mu_y$, $\operatorname{Var} X = \sigma_x^2$, $\operatorname{Var} Y = \sigma_y^2$, and $\rho_{X,Y} = \rho$, where $|\rho| < 1$. Then

$$
\begin{aligned}
f_{Y|X=x}(y) &= \frac{f_{X,Y}(x,y)}{f_X(x)} \\
&= \frac{\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\{-\frac{1}{2(1-\rho^2)}\big((\frac{x-\mu_x}{\sigma_x})^2 - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + (\frac{y-\mu_y}{\sigma_y})^2\big)\}}{\frac{1}{\sqrt{2\pi}\sigma_x}\exp\{-\frac{1}{2}(\frac{x-\mu_x}{\sigma_x})^2\}} \\
&= \frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}} \exp\Big\{-\tfrac{1}{2(1-\rho^2)}\Big((\tfrac{x-\mu_x}{\sigma_x})^2\rho^2 - 2\rho\tfrac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + (\tfrac{y-\mu_y}{\sigma_y})^2\Big)\Big\} \\
&= \frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}} \exp\Big\{-\frac{1}{2\sigma_y^2(1-\rho^2)}\Big(y - \mu_y - \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)\Big)^2\Big\}. \qquad (6.1)
\end{aligned}
$$

This density is easily recognized as the density of a normal distribution with mean $\mu_y + \rho\frac{\sigma_y}{\sigma_x}(x - \mu_x)$ and variance $\sigma_y^2(1-\rho^2)$. It follows, in particular, that

$$
\begin{aligned}
E(Y \mid X = x) &= \mu_y + \rho\frac{\sigma_y}{\sigma_x}(x - \mu_x), \\
\operatorname{Var}(Y \mid X = x) &= \sigma_y^2(1-\rho^2).
\end{aligned} \qquad (6.2)
$$

As a special feature we observe that the regression function is linear (and coinciding with the regression line) and that the conditional variance equals the residual variance. For the former statement we refer back to Remark 2.5.4 and for the latter to Theorem 2.5.3. Further, recall that the residual variance is independent of x.

Example 6.1. Suppose the density of $(X, Y)'$ is given by

$$f(x,y) = \frac{1}{2\pi}\exp\{-\tfrac{1}{2}(x^2 - 2xy + 2y^2)\}.$$

Determine the conditional distributions, particularly the conditional expectations and the conditional variances.

Solution. The function $x^2 - 2xy + 2y^2 = (x-y)^2 + y^2$ is positive-definite. We thus identify the joint distribution as normal. An inspection of the density shows that

$$E\,X = E\,Y = 0 \quad \text{and} \quad \mathbf{\Lambda}^{-1} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}, \tag{6.3}$$

which implies that

$$\begin{pmatrix} X \\ Y \end{pmatrix} \in N(\mathbf{0}, \mathbf{\Lambda}), \quad \text{where} \quad \mathbf{\Lambda} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}. \tag{6.4}$$

It follows that $\operatorname{Var} X = 2$, $\operatorname{Var} Y = 1$, and $\operatorname{Cov}(X, Y) = 1$, and hence that $\rho_{X,Y} = 1/\sqrt{2}$.

A comparison with (6.2) shows that

$$\begin{aligned} E(Y \mid X = x) &= \frac{x}{2} \quad \text{and} \quad \operatorname{Var}(Y \mid X = x) = \frac{1}{2}, \\ E(X \mid Y = y) &= y \quad \text{and} \quad \operatorname{Var}(X \mid Y = y) = 1. \end{aligned}$$

The conditional distributions are the normal distributions with corresponding parameters. □

Remark 6.1. Instead of having to remember formula (6.2), it is often as simple to perform the computations leading to (6.1) directly in each case. Indeed, in higher dimensions this is necessary. As an illustration, let us compute $f_{Y|X=x}(y)$.

Following (6.4) or by using the fact that $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy$, we have

$$\begin{aligned} f_{Y|X=x}(y) &= \frac{\frac{1}{2\pi} \exp\{-\frac{1}{2}(x^2 - 2xy + 2y^2)\}}{\frac{1}{\sqrt{2\pi}\sqrt{2}} \exp\{-\frac{1}{2} \cdot \frac{x^2}{2}\}} \\ &= \frac{1}{\sqrt{2\pi}\sqrt{1/2}} \exp\Big\{-\frac{1}{2}\Big(\frac{x^2}{2} - 2xy + 2y^2\Big)\Big\} \\ &= \frac{1}{\sqrt{2\pi}\sqrt{1/2}} \exp\Big\{-\frac{1}{2}\frac{(y - x/2)^2}{1/2}\Big\}, \end{aligned}$$

which is the density of the $N(x/2, 1/2)$-distribution. □

Exercise 6.1. Compute $f_{X|Y=y}(x)$ similarly. □

Example 6.2. Suppose that $\mathbf{X} \in N(\boldsymbol{\mu}, \mathbf{\Lambda})$, where $\boldsymbol{\mu} = \mathbf{1}$ and

$$\mathbf{\Lambda} = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}.$$

Find the conditional distribution of $X_1 + X_2$ given that $X_1 - X_2 = 0$.

Solution. We introduce the random variables $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$ to reduce the problem to the standard case; we are then faced with the problem of finding the conditional distribution of Y_1 given that $Y_2 = 0$.

Since we can write $\mathbf{Y} = \mathbf{B}\mathbf{X}$, where

$$\mathbf{B} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

it follows that $\mathbf{Y} \in N(\mathbf{B}\boldsymbol{\mu}, \mathbf{B}\boldsymbol{\Lambda}\mathbf{B}')$, that is, that

$$\mathbf{Y} \in N\Big(\begin{pmatrix} 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 7 & 1 \\ 1 & 3 \end{pmatrix} \Big),$$

and hence that

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{2\pi\sqrt{20}} \exp\Big\{-\frac{1}{2}\Big(\frac{3(y_1-2)^2}{20} - \frac{(y_1-2)y_2}{10} + \frac{7y_2^2}{20}\Big)\Big\}.$$

Further, since $Y_2 \in N(0, 3)$, we have

$$f_{Y_2}(y_2) = \frac{1}{\sqrt{2\pi}\sqrt{3}} \exp\Big\{-\frac{1}{2} \cdot \frac{y_2^2}{3}\Big\}.$$

Finally,

$$\begin{aligned} f_{Y_1 \mid Y_2 = 0}(y_1) &= \frac{f_{Y_1,Y_2}(y_1, 0)}{f_{Y_2}(0)} = \frac{\frac{1}{2\pi\sqrt{20}} \exp\{-\frac{1}{2} \cdot \frac{3(y_1-2)^2}{20}\}}{\frac{1}{\sqrt{2\pi}\sqrt{3}} \exp\{-\frac{1}{2} \cdot 0\}} \\ &= \frac{1}{\sqrt{2\pi}\sqrt{20/3}} \exp\Big\{-\frac{1}{2}\frac{(y_1-2)^2}{20/3}\Big\}, \end{aligned}$$

which we identify as the density of the $N(2, 20/3)$-distribution. □

Remark 6.2. It follows from the general formula (6.1) that the final exponent must be a square. This provides an extra check of one's computations. Also, the variance appears twice (in the last example it is 20/3) and must be the same in both places. □

Let us conclude by briefly considering the general case $n \geq 2$. Thus, $\mathbf{X} \in N(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ with $\det \boldsymbol{\Lambda} > 0$. Let $\widetilde{\mathbf{X}}_1 = (X_{i_1}, X_{i_2}, \ldots, X_{i_k})'$ and $\widetilde{\mathbf{X}}_2 = (X_{j_1}, X_{j_2}, \ldots, X_{j_m})'$ be subvectors of $\mathbf{X}$, that is, vectors whose components consist of k and m of the components of $\mathbf{X}$, respectively, where $1 \leq k < n$ and $1 \leq m < n$. The components of $\widetilde{\mathbf{X}}_1$ and $\widetilde{\mathbf{X}}_2$ are assumed to be different. By definition we then have

$$f_{\widetilde{\mathbf{X}}_2 \mid \widetilde{\mathbf{X}}_1 = \widetilde{\mathbf{x}}_1}(\widetilde{\mathbf{x}}_2) = \frac{f_{\widetilde{\mathbf{X}}_1, \widetilde{\mathbf{X}}_2}(\widetilde{\mathbf{x}}_1, \widetilde{\mathbf{x}}_2)}{f_{\widetilde{\mathbf{X}}_1}(\widetilde{\mathbf{x}}_1)}. \tag{6.5}$$

Given the formula for normal densities (Theorem 5.1) and the fact that the coordinates of $\widetilde{\mathbf{x}}_1$ are constants, the ratio in (6.5) must be the density of some normal distribution. The conclusion is that *conditional distributions of multivariate normal distributions are normal.*

Exercise 6.2. Let $\mathbf{X} \in N(\mathbf{0}, \mathbf{\Lambda})$, where

$$\mathbf{\Lambda} = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 6 & 0 \\ -1 & 0 & 4 \end{pmatrix}.$$

Set $Y_1 = X_1 + X_3$, $Y_2 = 2X_1 - X_2$, and $Y_3 = 2X_3 - X_2$. Find the conditional distribution of Y_3 given that $Y_1 = 0$ and $Y_2 = 1$.

7 Independence

A very special property of the multivariate normal distribution is the following:

Theorem 7.1. *Let* $\mathbf{X}$ *be a normal random vector. The components of* $\mathbf{X}$ *are independent iff they are uncorrelated.*

Proof. We only need to show that uncorrelated components are independent, the converse always being true.

Thus, by assumption, $\mathrm{Cov}(X_i, X_j) = 0$, $i \neq j$. This implies that the covariance matrix is diagonal, the diagonal elements being $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$. If some $\sigma_k^2 = 0$, then that component is degenerate and hence independent of the others. We therefore may assume that all variances are positive in the following. It then follows that the inverse $\mathbf{\Lambda}^{-1}$ of the covariance matrix exists; it is a diagonal matrix with diagonal elements $1/\sigma_1^2, 1/\sigma_2^2, \ldots, 1/\sigma_n^2$. The corresponding density function therefore equals

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= \Big(\frac{1}{2\pi}\Big)^{n/2} \frac{1}{\prod_{k=1}^n \sigma_k} \cdot \exp\Big\{-\frac{1}{2}\sum_{k=1}^n \frac{(x_k - \mu_k)^2}{\sigma_k^2}\Big\} \\ &= \prod_{k=1}^n \frac{1}{\sqrt{2\pi}\sigma_k} \cdot \exp\Big\{-\frac{(x_k-\mu_k)^2}{2\sigma_k^2}\Big\}, \end{aligned}$$

which proves the desired independence. □

Example 7.1. Let X_1 and X_2 be independent, $N(0, 1)$-distributed random variables. Show that $X_1 + X_2$ and $X_1 - X_2$ are independent.

Solution. It is easily checked that $\mathrm{Cov}(X_1 + X_2, X_1 - X_2) = 0$, which implies that $X_1 + X_2$ and $X_1 - X_2$ are uncorrelated. By Theorem 7.1 they are also independent. □

Remark 7.1. We have already encountered Example 7.1 in Chapter 1; see Example 1.2.4. There independence was proved with the aid of transformation (Theorem 1.2.1) and factorization. The solution here illustrates the power of Theorem 7.1. □

Exercise 7.1. Let X and Y be jointly normal with correlation coefficient ρ and suppose that $\operatorname{Var} X = \operatorname{Var} Y$. Show that X and $Y - \rho X$ are independent.

Exercise 7.2. Let X and Y be jointly normal with $E\,X = E\,Y = 0$, $\operatorname{Var} X = \operatorname{Var} Y = 1$, and correlation coefficient ρ. Find θ such that $X\cos\theta + Y\sin\theta$ and $X\cos\theta - Y\sin\theta$ are independent.

Exercise 7.3. Generalize the results of Example 7.1 and Exercise 7.1 to the case of nonequal variances. □

Remark 7.2. In Example 3.1 we stressed the importance of the assumption that the distribution was *jointly* normal. The example is also suited to illustrate the importance of that assumption with respect to Theorem 7.1. Namely, since $E\,X = E\,Y = 0$ and $E\,XY = E\,X^2 Z = E\,X^2 \cdot E\,Z = 0$, it follows that X and Y are *uncorrelated*. However, since $|X| = |Y|$, it is clear that X and Y are *not independent*. □

We conclude by stating the following generalization of Theorem 7.1, the proof of which we leave as an exercise:

Theorem 7.2. *Suppose that $\mathbf{X} \in N(\boldsymbol{\mu}, \mathbf{\Lambda})$, where $\mathbf{\Lambda}$ can be partitioned as follows:*

$$\mathbf{\Lambda} = \begin{pmatrix} \mathbf{\Lambda}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Lambda}_k \end{pmatrix}$$

(possibly after reordering the components), where $\mathbf{\Lambda}_1, \mathbf{\Lambda}_2, \ldots, \mathbf{\Lambda}_k$ are matrices along the diagonal of $\mathbf{\Lambda}$. Then $\mathbf{X}$ can be partitioned into vectors $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(k)}$ with $\operatorname{Cov}(\mathbf{X}^{(i)}) = \mathbf{\Lambda}_i$, $i = 1, 2, \ldots, k$, in such a way that these random vectors are independent. □

Example 7.2. Suppose that $\mathbf{X} \in N(\mathbf{0}, \mathbf{\Lambda})$, where

$$\mathbf{\Lambda} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 4 \\ 0 & 4 & 9 \end{pmatrix}.$$

Then X_1 and $(X_2, X_3)'$ are independent. □

8 Linear Transformations

A major consequence of Theorem 7.1 is that it is possible to make linear transformations of normal vectors in such a way that the new vector has independent components. In particular, any orthogonal transformation of a normal vector whose components are independent and have common variance produces a new normal random vector with independent components. As a major application, we show in Example 8.3 how these relatively simple facts can be used to prove the rather delicate result that states that the sample mean and the sample variance in a normal sample are independent. For further details concerning applications in statistics we refer to Appendix A, where some references are given.

We first recall from Section 3 that a linear transformation of a normal random vector is normal. Now suppose that $\mathbf{X} \in N(\boldsymbol{\mu}, \mathbf{\Lambda})$. Since $\mathbf{\Lambda}$ is nonnegative-definite, there exists (formula (1.3)) an orthogonal matrix $\mathbf{C}$, such that $\mathbf{C}'\mathbf{\Lambda}\mathbf{C} = \mathbf{D}$, where $\mathbf{D}$ is a diagonal matrix whose diagonal elements are the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of $\mathbf{\Lambda}$.

Set $\mathbf{Y} = \mathbf{C}'\mathbf{X}$. It follows from Theorem 3.1 that $\mathbf{Y} \in N(\mathbf{C}'\boldsymbol{\mu}, \mathbf{D})$. The components of $\mathbf{Y}$ are thus uncorrelated and, in view of Theorem 7.1, *independent*, which establishes the following result:

Theorem 8.1. *Let $\mathbf{X} \in N(\boldsymbol{\mu}, \mathbf{\Lambda})$, and set $\mathbf{Y} = \mathbf{C}'\mathbf{X}$, where the orthogonal matrix $\mathbf{C}$ is such that $\mathbf{C}'\mathbf{\Lambda}\mathbf{C} = \mathbf{D}$. Then $\mathbf{Y} \in N(\mathbf{C}'\boldsymbol{\mu}, \mathbf{D})$. Moreover, the components of $\mathbf{Y}$ are independent and $\operatorname{Var} Y_k = \lambda_k$, $k = 1, 2, \ldots, n$, where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of $\mathbf{\Lambda}$.* □

Remark 8.1. In particular, it may occur that some eigenvalues are equal to zero, in which case the corresponding component is degenerate.

Remark 8.2. As a special corollary it follows that the statement "$\mathbf{X} \in N(\mathbf{0}, \mathbf{I})$" is equivalent to the statement "$X_1, X_2, \ldots, X_n$ are independent, standard normal random variables."

Remark 8.3. The primary use of Theorem 8.1 is in proofs and for theoretical arguments. In practice it may be cumbersome to apply the theorem when n is large, since the computation of the eigenvalues of $\mathbf{\Lambda}$ amounts to solving an algebraic equation of degree n. □

Another situation of considerable importance in statistics is orthogonal transformations of independent, normal random variables with the same variance, the point being that the transformed random variables also are independent. That this is indeed the case may easily be proved with the aid of Theorem 8.1. Namely, let $\mathbf{X} \in N(\boldsymbol{\mu}, \sigma^2\mathbf{I})$, where $\sigma^2 > 0$, and set $\mathbf{Y} = \mathbf{C}\mathbf{X}$, where $\mathbf{C}$ is an orthogonal matrix. Then $\operatorname{Cov} \mathbf{Y} = \mathbf{C}\sigma^2\mathbf{I}\mathbf{C}' = \sigma^2\mathbf{I}$, which, in view of Theorem 7.1, yields the following result:

Theorem 8.2. *Let $\mathbf{X} \in N(\boldsymbol{\mu}, \sigma^2\mathbf{I})$, where $\sigma^2 > 0$, let $\mathbf{C}$ be an arbitrary orthogonal matrix, and set $\mathbf{Y} = \mathbf{C}\mathbf{X}$. Then $\mathbf{Y} \in N(\mathbf{C}\boldsymbol{\mu}, \sigma^2\mathbf{I})$; in particular, $Y_1, Y_2, \ldots, Y_n$ are independent normal random variables with the same variance, σ^2.* □

As a first application we reexamine Example 7.1.

Example 8.1. Thus, X and Y are independent, $N(0,1)$-distributed random variables, and we wish to show that $X+Y$ and $X-Y$ are independent.

It is clearly equivalent to prove that $U = (X+Y)/\sqrt{2}$ and $V = (X-Y)/\sqrt{2}$ are independent. Now, $(X,Y)' \in N(\mathbf{0},\mathbf{I})$ and

$$\begin{pmatrix} U \\ V \end{pmatrix} = \mathbf{B} \begin{pmatrix} X \\ Y \end{pmatrix}, \quad \text{where} \quad \mathbf{B} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix},$$

that is, $\mathbf{B}$ is orthogonal. The conclusion follows immediately from Theorem 8.2.

Example 8.2. Let $X_1, X_2, \ldots, X_n$ be independent, $N(0,1)$-distributed random variables, and let $a_1, a_2, \ldots, a_n$ be reals, such that $\sum_{k=1}^n a_k^2 \neq 0$. Find the conditional distribution of $\sum_{k=1}^n X_k^2$ given that $\sum_{k=1}^n a_k X_k = 0$.

Solution. We first observe that $\sum_{k=1}^n X_k^2 \in \chi^2(n)$ (recall Exercise 3.3.6 for the case $n = 2$). In order to determine the desired conditional distribution, we define an orthogonal matrix $\mathbf{C}$, whose first row consists of the elements a_1/a, a_2/a, $\ldots$, a_n/a, where $a = \sqrt{\sum_{k=1}^n a_k^2}$; note that $\sum_{k=1}^n (a_k/a)^2 = 1$. From linear algebra we know that the matrix $\mathbf{C}$ can be completed in such a way that it becomes an orthogonal matrix. Next we set $\mathbf{Y} = \mathbf{CX}$, note that $\mathbf{Y} \in N(\mathbf{0},\mathbf{I})$ by Theorem 8.2, and observe that, in particular, $aY_1 = \sum_{k=1}^n a_k X_k$. Moreover, since $\mathbf{C}$ is orthogonal, we have $\sum_{k=1}^n Y_k^2 = \sum_{k=1}^n X_k^2$ (formula (1.2)). It follows that the desired conditional distribution is the same as the conditional distribution of $\sum_{k=1}^n Y_k^2$ given that $Y_1 = 0$, that is, as the distribution of $\sum_{k=2}^n Y_k^2$, which is $\chi^2(n-1)$. □

Exercise 8.1. Study the case $n = 2$ and $a_1 = a_2 = 1$ in detail. Try also to reach the conclusion via the random variables U and V in Example 8.1. □

Example 8.3. There exists a famous characterization of the normal distribution to the effect that it is the only distribution such that the arithmetic mean and the sample variance are independent. This independence is, for example, exploited in order to verify that the t-statistic, which is used for testing the mean in a normal population when the variance is unknown, actually follows a t-distribution.

Here we prove the "if" part; the other one is much harder. Thus, let $X_1, X_2, \ldots, X_n$ be independent, $N(0,1)$-distributed random variables, set $\bar{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$ and $s_n^2 = \frac{1}{n-1}\sum_{k=1}^n (X_k - \bar{X}_n)^2$.

The first step is to determine the distribution of

$$(\bar{X}_n, X_1 - \bar{X}_n, X_2 - \bar{X}_n, \ldots, X_n - \bar{X}_n)'.$$

Since the vector can be written as $\mathbf{BX}$, where

$$\mathbf{B} = \begin{pmatrix} \frac{1}{n} & \frac{1}{n} & \frac{1}{n} \dots & \frac{1}{n} \\ 1-\frac{1}{n} & -\frac{1}{n} & -\frac{1}{n} \dots & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & -\frac{1}{n} \dots & -\frac{1}{n} \\ \vdots & \vdots & \ddots \ \vdots & \\ -\frac{1}{n} & -\frac{1}{n} & -\frac{1}{n} \dots & 1-\frac{1}{n} \end{pmatrix},$$

we know that the vector is normal with mean $\mathbf{0}$ and covariance matrix

$$\mathbf{BB}' = \begin{pmatrix} \frac{1}{n} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \end{pmatrix},$$

where $\mathbf{A}$ is some matrix the exact expression of which is of no importance here. Namely, the point is that we may apply Theorem 7.2 in order to conclude that $\bar{X}_n$ and $(X_1 - \bar{X}_n, X_2 - \bar{X}_n, \dots, X_n - \bar{X}_n)$ are independent, and since s_n^2 is simply a function of $(X_1 - \bar{X}_n, X_2 - \bar{X}_n, \dots, X_n - \bar{X}_n)$ it follows that $\bar{X}_n$ and s_n^2 are independent random variables. □

Exercise 8.2. Suppose that $\mathbf{X} \in N(\boldsymbol{\mu}, \sigma^2\mathbf{I})$, where $\sigma^2 > 0$. Show that if $\mathbf{B}$ is any matrix such that $\mathbf{BB}' = \mathbf{D}$, a diagonal matrix, then the components of $\mathbf{Y} = \mathbf{BX}$ are independent, normal random variables; this generalizes Theorem 8.2. As an application, reconsider Example 8.1. □

Theorem 8.3. *(Daly's theorem) Let* $\mathbf{X} \in N(\boldsymbol{\mu}, \sigma^2\mathbf{I})$ *and set* $\bar{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$. *Suppose that* $g(\mathbf{x})$ *is translation invariant, that is, for all* $\mathbf{x} \in \mathbf{R}^n$, *we have* $g(\mathbf{x} + a \cdot \mathbf{1}) = g(\mathbf{x})$ *for all* a. *Then* $\bar{X}_n$ *and* $g(\mathbf{X})$ *are independent.*

Proof. Throughout the proof we assume, without restriction, that $\boldsymbol{\mu} = \mathbf{0}$ and that $\sigma^2 = 1$. The translation invariance of g implies that g is, in fact, living in the $(n-1)$-dimensional hyperplane $x_1 + x_2 + \dots + x_n =$ constant, on which $\bar{X}_n$ is constant. We therefore make a change of variable similar to that of Example 8.2. Namely, define an orthogonal matrix $\mathbf{C}$ such that the first row has all elements equal to $1/\sqrt{n}$, and set $\mathbf{Y} = \mathbf{CX}$. Then, by construction, we have $Y_1 = \sqrt{n} \cdot \bar{X}_n$ and, by Theorem 8.2, that $\mathbf{Y} \in N(\mathbf{0}, \mathbf{I})$. The translation invariance implies, in view of the above, that g depends only on Y_2, Y_3, ..., Y_n and hence, by Theorem 7.2, is independent of Y_1. □

Example 8.4. Since the sample variance s_n^2 as defined in Example 8.3 is translation invariant, the conclusion of that example follows, alternatively, from Daly's theorem. Note, however, that Daly's theorem can be viewed as an extension of that very example.

Example 8.5. The range $R_n = X_{(n)} - X_{(1)}$ (which was defined in Section 4.2) is obviously translation invariant. It follows that $\bar{X}_n$ and R_n are independent (in normal samples). □

There also exist useful linear transformations that are not orthogonal. One important example, in the two-dimensional case, is the following, a special case of which was considered in Exercise 7.1.

Suppose that $\mathbf{X} \in N(\boldsymbol{\mu}, \mathbf{\Lambda})$, where

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{and} \quad \mathbf{\Lambda} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

with $|\rho| < 1$. Define $\mathbf{Y}$ through the relations

$$\begin{aligned} X_1 &= \mu_1 + \sigma_1 Y_1, \\ X_2 &= \mu_2 + \rho\sigma_2 Y_1 + \sigma_2\sqrt{1-\rho^2}\,Y_2. \end{aligned} \tag{8.1}$$

This means that $\mathbf{X}$ and $\mathbf{Y}$ are connected via $\mathbf{X} = \boldsymbol{\mu} + \mathbf{B}\mathbf{Y}$, where

$$\mathbf{B} = \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sigma_2\sqrt{1-\rho^2} \end{pmatrix},$$

which is not orthogonal. However, a simple computation shows that $\mathbf{Y} \in N(\mathbf{0}, \mathbf{I})$, that is, Y_1 and Y_2 are independent, standard normal random variables.

Example 8.6. If X_1 and X_2 are independent and $N(0,1)$-distributed, then X_1^2 and X_2^2 are independent, $\chi^2(1)$-distributed random variables, from which it follows that $X_1^2 + X_2^2 \in \chi^2(2)$ (Exercise 3.3.6(b)). Now, assume that $\mathbf{X}$ is normal with $E\,X_1 = E\,X_2 = 0$, $\operatorname{Var} X_1 = \operatorname{Var} X_2 = 1$, and $\rho_{X_1,X_2} = \rho$ with $|\rho| < 1$. Find the distribution of $X_1^2 - 2\rho X_1 X_2 + X_2^2$.

To solve this problem, we first observe that for $\rho = 0$ it reduces to Exercise 3.3.6(b) (why?). In the general case,

$$X_1^2 - 2\rho X_1 X_2 + X_2^2 = (X_1 - \rho X_2)^2 + (1-\rho^2)X_2^2. \tag{8.2}$$

From above (or Exercise 7.1) we know that $X_1 - \rho X_2$ and X_2 are independent, in fact,

$$\begin{pmatrix} X_1 - \rho X_2 \\ X_2 \end{pmatrix} = \begin{pmatrix} 1 & -\rho \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \in N\Big(\mathbf{0}, \begin{pmatrix} 1-\rho^2 & 0 \\ 0 & 1 \end{pmatrix}\Big).$$

It follows that

$$X_1^2 - 2\rho X_1 X_2 + X_2^2 = (1-\rho^2)\Big\{\Big(\frac{X_1 - \rho X_2}{\sqrt{1-\rho^2}}\Big)^2 + X_2^2\Big\} \in (1-\rho^2)\cdot\chi^2(2),$$

and since $\chi^2(2) = \operatorname{Exp}(2)$ we conclude, from the scaling property of the exponential distribution, that $X_1^2 - 2\rho X_1 X_2 + X_2^2 \in \operatorname{Exp}(2(1-\rho^2))$.

We shall return to this example in a more general setting in Section 9; see also Problem 10.37. □

9 Quadratic Forms and Cochran's Theorem

Quadratic forms of normal random vectors are of great importance in many branches of statistics, such as least-squares methods, the analysis of variance, regression analysis, and experimental design. The general idea is to split the sum of the squares of the observations into a number of quadratic forms, each corresponding to some cause of variation. In an agricultural experiment, for example, the yield of crop varies. The reason for this may be differences in fertilization, watering, climate, and other factors in the various areas where the experiment is performed. For future purposes one would like to investigate, if possible, how much (or if at all) the various treatments influence the variability of the result. The splitting of the sum of squares mentioned above separates the causes of variability in such a way that each quadratic form corresponds to one cause, with a final form—the residual form—that measures the random errors involved in the experiment. The conclusion of Cochran's theorem (Theorem 9.2) is that, under the assumption of normality, the various quadratic forms are independent and χ^2-distributed (except for a constant factor). This can then be used for testing hypotheses concerning the influence of the different treatments. Once again, we remind the reader that some books on statistics for further study are mentioned in Appendix A.

We begin by investigating a particular quadratic form, after which we prove the important Cochran's theorem.

Let $\mathbf{X} \in N(\boldsymbol{\mu}, \boldsymbol{\Lambda})$, where $\boldsymbol{\Lambda}$ is nonsingular, and consider the quadratic form $(\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Lambda}^{-1}(\mathbf{X} - \boldsymbol{\mu})$, which appears in the exponent of the normal density. In the special case $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Lambda} = \mathbf{I}$ it reduces to $\mathbf{X}'\mathbf{X}$, which is $\chi^2(n)$-distributed (n is the dimension of $\mathbf{X}$). The following result shows that this is also true in the general case.

Theorem 9.1. *Suppose that* $\mathbf{X} \in N(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ *with* $\det \boldsymbol{\Lambda} > 0$. *Then*

$$(\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Lambda}^{-1}(\mathbf{X} - \boldsymbol{\mu}) \in \chi^2(n),$$

where n is the dimension of $\mathbf{X}$.

Proof. Set $\mathbf{Y} = \boldsymbol{\Lambda}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$. Then

$$E\,\mathbf{Y} = \mathbf{0} \quad \text{and} \quad \operatorname{Cov} \mathbf{Y} = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{-1/2} = \mathbf{I},$$

that is, $\mathbf{Y} \in N(\mathbf{0}, \mathbf{I})$, and it follows that

$$(\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Lambda}^{-1}(\mathbf{X} - \boldsymbol{\mu}) = (\boldsymbol{\Lambda}^{-1/2}(\mathbf{X} - \boldsymbol{\mu}))'(\boldsymbol{\Lambda}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})) = \mathbf{Y}'\mathbf{Y} \in \chi^2(n),$$

as was shown above. □

Remark 9.1. Let $n = 2$. With the usual notation the theorem amounts to the fact that

$$\frac{1}{1-\rho^2}\Big\{\frac{(X_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(X_1-\mu_1)(X_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(X_2-\mu_2)^2}{\sigma_2^2}\Big\} \in \chi^2(2). \qquad \square$$

As an introduction to Cochran's theorem, we study the following situation. Suppose that $X_1, X_2, \ldots, X_n$ is a sample of $X \in N(0, \sigma^2)$. Set $\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$, and consider the following identity:

$$\sum_{k=1}^{n} X_k^2 = \sum_{k=1}^{n} (X_k - \bar{X}_n)^2 + n \cdot \bar{X}_n^2. \tag{9.1}$$

The first term on the right-hand side equals $(n-1)s_n^2$, where s_n^2 is the sample variance. It is a $\sigma^2 \cdot \chi^2(n-1)$-distributed quadratic form. The second term is $\sigma^2 \cdot \chi^2(1)$-distributed. The terms are independent. The left-hand side is $\sigma^2 \cdot \chi^2(n)$-distributed. We have thus split the sum of the squares of the observations into a sum of two independent quadratic forms that both follow some χ^2-distribution (except for the factor σ^2).

The statistical significance of this is that the splitting of the sum of the squares $\sum_{k=1}^{n} X_k^2$ is the following. Namely, the first term on the right-hand side of (9.1) is large if the sample is very much spread out, and the second term is large if the mean is not "close" to zero. Thus, if the sum of squares is large we may, via the decomposition (9.1) find out the cause; is the variance large or is it not true that the mean is zero (or both)?

In Example 8.3 we found that the terms on the right-hand side of (9.1) were independent. This leads to the t-test, which is used for testing whether or not the mean equals zero. More generally, representations of the sum of squares as a sum of nonnegative-definite quadratic forms play a fundamental role in statistics, as pointed out before. The problem is to assert that the various terms on the right-hand side of such representations are independent and χ^2-distributed. Cochran's theorem provides a solution to this problem.

As a preliminary we need the following lemma:

Lemma 9.1. *Let $x_1, x_2, \ldots, x_n$ be real numbers. Suppose that $\sum_{i=1}^{n} x_i^2$ can be split into a sum of nonnegative-definite quadratic forms, that is, suppose that*

$$\sum_{i=1}^{n} x_i^2 = Q_1 + Q_2 + \cdots + Q_k ,$$

where $Q_i = \mathbf{x}'\mathbf{A}_i\mathbf{x}$ and (Rank Q_i =) Rank $\mathbf{A}_i = r_i$ *for $i = 1, 2, \ldots, k$. If $\sum_{i=1}^{k} r_i = n$, then there exists an orthogonal matrix $\mathbf{C}$ such that, with $\mathbf{x} = \mathbf{C}\mathbf{y}$, we have*

$$\begin{aligned}
Q_1 &= y_1^2 + y_2^2 + \cdots + y_{r_1}^2 , \\
Q_2 &= y_{r_1+1}^2 + y_{r_1+2}^2 + \cdots + y_{r_1+r_2}^2 , \\
Q_3 &= y_{r_1+r_2+1}^2 + y_{r_1+r_2+2}^2 + \cdots + y_{r_1+r_2+r_3}^2 , \\
&\vdots \\
Q_k &= y_{n-r_k+1}^2 + y_{n-r_k+2}^2 + \cdots + y_n^2 .
\end{aligned}$$

□

Remark 9.2. Note that different quadratic forms contain different y variables and that the number of terms in each Q_i equals the rank r_i of Q_i. □

We confine ourselves to proving the lemma for the case $k = 2$. The general case is obtained by induction.

Proof. Recall the assumption that $k = 2$. We thus have

$$Q = \sum_{i=1}^{n} x_i^2 = \mathbf{x}'\mathbf{A}_1\mathbf{x} + \mathbf{x}'\mathbf{A}_2\mathbf{x} \quad \bigl(= Q_1 + Q_2\bigr)\,, \tag{9.2}$$

where $\mathbf{A}_1$ and $\mathbf{A}_2$ are nonnegative-definite matrices with ranks r_1 and r_2, respectively, and $r_1 + r_2 = n$. Since A_1 is nonnegative-definite, there exists an orthogonal matrix $\mathbf{C}$ such that

$$\mathbf{C}'\mathbf{A}_1\mathbf{C} = \mathbf{D},$$

where $\mathbf{D}$ is a diagonal matrix, the diagonal elements $\lambda_1, \lambda_2, \ldots, \lambda_n$ of which are the eigenvalues of $\mathbf{A}_1$. Since $\operatorname{Rank}\mathbf{A}_1 = r_1$, r_1 λ-values are positive and $n - r_1$ λ-values equal zero. Suppose, without restriction, that $\lambda_i > 0$ for $i = 1, 2, \ldots, r_1$ and that $\lambda_{r_1+1} = \lambda_{r_1+2} = \cdots = \lambda_n = 0$, and set $\mathbf{x} = \mathbf{C}\mathbf{y}$. Then (recall (1.2) for the first equality)

$$Q = \sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{r_1} \lambda_i \cdot y_i^2 + \mathbf{y}'\mathbf{C}'\mathbf{A}_2\mathbf{C}\mathbf{y}\,,$$

or, equivalently,

$$\sum_{i=1}^{r_1}(1-\lambda_i)\cdot y_i^2 + \sum_{i=r_1+1}^{n} y_i^2 = \mathbf{y}'\mathbf{C}'\mathbf{A}_2\mathbf{C}\mathbf{y}\,. \tag{9.3}$$

Since the rank of the right-hand side of (9.3) equals $r_2\,(= n - r_1)$, it follows that $\lambda_1 = \lambda_2 = \cdots = \lambda_{r_1} = 1$, which shows that

$$Q_1 = \sum_{i=1}^{r_1} y_i^2 \quad \text{and} \quad Q_2 = \sum_{i=r_1+1}^{n} y_i^2\,. \tag{9.4}$$

This proves the lemma for the case $k = 2$. □

Theorem 9.2. *(Cochran's theorem) Suppose that $X_1, X_2, \ldots, X_n$ are independent, $N(0, \sigma^2)$-distributed random variables, and that*

$$\sum_{i=1}^{n} X_i^2 = Q_1 + Q_2 + \cdots + Q_k\,,$$

where $Q_1, Q_2, \ldots, Q_k$ are nonnegative-definite quadratic forms in the random variables $X_1, X_2, \ldots, X_n$, that is,

$$Q_i = \mathbf{X}'\mathbf{A}_i\mathbf{X}\,, \quad i = 1, 2, \ldots, k.$$

Set Rank $\mathbf{A}_i = r_i$, $i = 1, 2, \ldots, k$. *If*

$$r_1 + r_2 + \cdots + r_k = n,$$

then

(a) $Q_1, Q_2, \ldots, Q_k$ *are independent;*
(b) $Q_i \in \sigma^2\chi^2(r_i), \quad i = 1, 2, \ldots, k.$

Proof. It follows from Lemma 9.1 that there exists an orthogonal matrix $\mathbf{C}$ such that the transformation $\mathbf{X} = \mathbf{CY}$ yields

$$\begin{aligned} Q_1 &= Y_1^2 + Y_2^2 + \cdots + Y_{r_1}^2, \\ Q_2 &= Y_{r_1+1}^2 + Y_{r_1+2}^2 + \cdots + Y_{r_1+r_2}^2, \\ &\vdots \\ Q_k &= Y_{n-r_k+1}^2 + Y_{n-r_k+2}^2 + \cdots + Y_n^2. \end{aligned}$$

Since, by Theorem 8.2, $Y_1, Y_2, \ldots, Y_n$ are independent, $N(0, \sigma^2)$-distributed random variables, and since every Y^2 occurs in exactly one Q_j, the conclusion follows. □

Remark 9.3. It suffices to assume that Rank $\mathbf{A}_i \leq r_i$ for $i = 1, 2, \ldots, k$, with $r_1 + r_2 + \cdots + r_k = n$, in order for Theorem 9.2 to hold. This follows from a result in linear algebra, namely that if $\mathbf{A}, \mathbf{B}$, and $\mathbf{C}$ are matrices such that $\mathbf{A} + \mathbf{B} = \mathbf{C}$, then Rank $\mathbf{C} \leq$ Rank $\mathbf{A}$ + Rank $\mathbf{B}$. An application of this result yields

$$n \leq \sum_{i=1}^{k} \operatorname{Rank} \mathbf{A}_i \leq \sum_{i=1}^{k} r_i = n\,, \tag{9.5}$$

which, in view of the assumption, forces Rank $\mathbf{A}_i$ to be equal to r_i for all i. □

Example 9.1. We have already proved (twice) in Section 8 that the sample mean and the sample variance are independent in a normal sample. By using the partition in formula (9.1) and Cochran's theorem (and Remark 9.2) we may obtain a third proof of that fact. □

In applications the quadratic forms can frequently be written as

$$Q = L_1^2 + L_2^2 + \cdots + L_p^2\,, \tag{9.6}$$

where $L_1, L_2, \ldots, L_p$ are linear forms in $X_1, X_2, \ldots, X_n$. It may therefore be useful to know some method for determining the rank of a quadratic form of this kind.

Theorem 9.3. *Suppose that the nonnegative-definite form* $Q = Q(\mathbf{x})$ *is of the form (9.6), where*

$$L_i = \mathbf{a}_i'\mathbf{x}\,, \quad i = 1, 2, \ldots, p,$$

and set $\mathbf{L} = (L_1, L_2, ..., L_p)'$. *If there exist exactly* m *linear relations* $\mathbf{d}_j'\mathbf{L} = 0$, $j = 1, 2, \ldots, m$, *then* Rank $Q = p - m$.

Proof. Put $\mathbf{L} = \mathbf{Ax}$, where $\mathbf{A}$ is a $p \times n$ matrix. Then $\operatorname{Rank} \mathbf{A} = p - m$. However, since

$$Q = \mathbf{L}'\mathbf{L} = \mathbf{x}'\mathbf{A}'\mathbf{A}\mathbf{x}\,,$$

it follows (from linear algebra) that $\operatorname{Rank} \mathbf{A}'\mathbf{A} = \operatorname{Rank} \mathbf{A}$. □

Example 9.1 (continued). Thus, let $\mathbf{X} \in N(\mathbf{0}, \sigma^2\mathbf{I})$, and consider the partition (9.1). Then $Q_1 = \sum_{k=1}^{n}(X_k - \bar{X}_n)^2$ is of the kind described in Theorem 9.3, since $\sum_{k=1}^{n}(X_k - \bar{X}_n) = 0$. □

10 Problems

1. In this chapter we have (so far) met three equivalent definitions of a multivariate normal distribution. Here is a fourth one: $\mathbf{X}$ is normal if and only if there exists an orthogonal transformation $\mathbf{C}$ such that the random vector $\mathbf{CX}$ has independent, normal components. Show that this definition is indeed equivalent to the usual ones (e.g., by showing that it is equivalent to the first one).
2. Suppose that X and Y have a two-dimensional normal distribution with means 0, variances 1, and correlation coefficient ρ, $|\rho| < 1$. Let (R, Θ) be the polar coordinates. Determine the distribution of Θ.
3. The random variables X_1 and X_2 are independent and $N(0,1)$-distributed. Set
$$Y_1 = \frac{X_1^2 - X_2^2}{\sqrt{X_1^2 + X_2^2}} \quad \text{and} \quad Y_2 = \frac{2X_1 \cdot X_2}{\sqrt{X_1^2 + X_2^2}}\,.$$
Show that Y_1 and Y_2 are independent, $N(0,1)$-distributed random variables.
4. The random vector $(X, Y)'$ has a two-dimensional normal distribution with $\operatorname{Var} X = \operatorname{Var} Y$. Show that $X + Y$ and $X - Y$ are independent random variables.
5. Suppose that X and Y have a joint normal distribution with $E\,X = E\,Y = 0$, $\operatorname{Var} X = \sigma_x^2$, $\operatorname{Var} Y = \sigma_y^2$, and correlation coefficient ρ. Compute $E\,XY$ and $\operatorname{Var} XY$.
Remark. One may use the fact that X and a suitable linear combination of X and Y are independent.
6. The random variables X and Y are independent and $N(0,1)$-distributed. Determine
(a) $\mathrm{E}(X \mid X > Y)$,
(b) $\mathrm{E}(X + Y \mid X > Y)$.
7. We know from Section 7 that if X and Y are jointly normally distributed then they are independent iff they are uncorrelated. Now, let $X \in N(0,1)$ and $c \geq 0$. Define Y as follows:
$$Y = \begin{cases} X, & \text{for} \quad |X| \leq c, \\ -X, & \text{for} \quad |X| > c. \end{cases}$$

(a) Show that $Y \in N(0,1)$.
(b) Show that X and Y are not jointly normal.
Next, let $g(c) = \mathrm{Cov}\,(X,Y)$.
(c) Show that $g(0) = -1$ and that $g(c) \to 1$ as $c \to \infty$. Show that there exists c_0 such that $g(c_0) = 0$ (i.e., such that X and Y are uncorrelated).
(d) Show that X and Y are not independent (when $c = c_0$).

8. In Section 6 we found that conditional distributions of normal vectors are normal. The converse is, however, not true. Namely, consider the bivariate density
$$f_{X,Y}(x,y) = C \cdot \exp\{-(1+x^2)(1+y^2)\}, \quad -\infty < x, y < \infty,$$
where C is a normalizing constant. This is *not* a bivariate normal density. Show that in spite of this the conditional distributions are normal, that is, compute the conditional densities $f_{Y|X=x}(y)$ and $f_{X|Y=y}(x)$ and show that they are normal densities.
9. Suppose that the random variables X and Y are independent and $N(0,\sigma^2)$-distributed.
(a) Show that $X/Y \in C(0,1)$.
(b) Show that $X+Y$ and $X-Y$ are independent.
(c) Determine the distribution of $(X-Y)/(X+Y)$ (see also Problem 1.43(b)).
10. Suppose that the moment generating function of $(X,Y)'$ is
$$\psi_{X,Y}(t,u) = \exp\{2t + 3u + t^2 + atu + 2u^2\}.$$
Determine a so that $X+2Y$ and $2X-Y$ become independent.
11. Let $\mathbf{X}$ have a three-dimensional normal distribution. Show that if X_1 and X_2+X_3 are independent, X_2 and X_1+X_3 are independent, and X_3 and X_1+X_2 are independent, then X_1, X_2, and X_3 are independent.
12. Let X_1 and X_2 be independent, $N(0,1)$-distributed random variables. Set $Y_1 = X_1 - 3X_2 + 2$ and $Y_2 = 2X_1 - X_2 - 1$. Determine the distribution of
(a) $\mathbf{Y}$, and
(b) $Y_1 \mid Y_2 = y$.
13. Let X_1, X_2, and X_3 be independent, $N(1,1)$-distributed random variables. Set $U = 2X_1 - X_2 + X_3$ and $V = X_1 + 2X_2 + 3X_3$. Determine the conditional distribution of V given that $U = 3$.
14. Let X_1, X_2, X_3 be independent $N(2,1)$-distributed random variables. Determine the distribution of $X_1 + 3X_2 - 2X_3$ given that $2X_1 - X_2 = 1$.
15. Let Y_1, Y_2, and Y_3 be independent, $N(0,1)$-distributed random variables, and set
$$\begin{aligned} X_1 &= Y_1 - Y_3, \\ X_2 &= 2Y_1 + Y_2 - 2Y_3, \\ X_3 &= -2Y_1 + 3Y_3. \end{aligned}$$

Determine the conditional distribution of X_2 given that $X_1 + X_3 = x$.

16. The random variables X_1, X_2, and X_3 are independent and $N(0,1)$-distributed. Consider the random variables

$$\begin{aligned} Y_1 &= X_2 + X_3, \\ Y_2 &= X_1 + X_3, \\ Y_3 &= X_1 + X_2. \end{aligned}$$

Determine the conditional density of Y_1 given that $Y_2 = Y_3 = 0$.

17. The random vector $\mathbf{X}$ has a three-dimensional normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{\Lambda}$ given by

$$\mathbf{\Lambda} = \begin{pmatrix} 2 & 0 & -1 \\ 0 & 3 & 1 \\ -1 & 1 & 5 \end{pmatrix}.$$

Find the distribution of X_2 given that $X_1 - X_3 = 1$ and that $X_2 + X_3 = 0$.

18. The random vector $\mathbf{X}$ has a three-dimensional normal distribution with expectation $\mathbf{0}$ and covariance matrix $\mathbf{\Lambda}$ given by

$$\mathbf{\Lambda} = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 4 & 0 \\ -1 & 0 & 7 \end{pmatrix}.$$

Find the distribution of X_3 given that $X_1 = 1$.

19. The random vector $\mathbf{X}$ has a three-dimensional normal distribution with expectation $\mathbf{0}$ and covariance matrix $\mathbf{\Lambda}$ given by

$$\mathbf{\Lambda} = \begin{pmatrix} 2 & 1 & -1 \\ 1 & 3 & 0 \\ -1 & 0 & 5 \end{pmatrix}.$$

Find the distribution of X_2 given that $X_1 + X_3 = 1$.

20. The random vector $\mathbf{X}$ has a three-dimensional normal distribution with mean vector $\boldsymbol{\mu} = \mathbf{0}$ and covariance matrix

$$\mathbf{\Lambda} = \begin{pmatrix} 3 & -2 & 1 \\ -2 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

Find the distribution of $X_1 + X_3$ given that
(a) $X_2 = 0$,
(b) $X_2 = 2$.

21. Let $\mathbf{X} \in N(\boldsymbol{\mu}, \mathbf{\Lambda})$, where

$$\boldsymbol{\mu} = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} \quad \text{and} \quad \mathbf{\Lambda} = \begin{pmatrix} 3 & -2 & 1 \\ -2 & 2 & 0 \\ 1 & 0 & 2 \end{pmatrix}.$$

Determine the conditional distribution of $X_1 - X_3$ given that $X_2 = -1$.

22. Let $\mathbf{X} \in N(\boldsymbol{\mu}, \mathbf{\Lambda})$, where

$$\boldsymbol{\mu} = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} \quad \text{and} \quad \mathbf{\Lambda} = \begin{pmatrix} 3 & -2 & 1 \\ -2 & 2 & 0 \\ 1 & 0 & 3 \end{pmatrix}.$$

Determine the conditional distribution of $X_1 + X_2$ given that $X_3 = 1$.

23. The random vector $\mathbf{X}$ has a three-dimensional normal distribution with expectation $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Lambda}$ given by

$$\boldsymbol{\mu} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{\Lambda} = \begin{pmatrix} 4 & -2 & 1 \\ -2 & 3 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

Find the conditional distribution of $X_1 + 2X_2$ given that
(a) $X_2 - X_3 = 1$.
(b) $X_2 + X_3 = 1$.

24. The random vector $\mathbf{X}$ has a three-dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Lambda}$ given by

$$\boldsymbol{\mu} = \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix} \quad \text{and} \quad \mathbf{\Lambda} = \begin{pmatrix} 3 & -2 & 1 \\ -2 & 4 & -1 \\ 1 & -1 & 2 \end{pmatrix}.$$

Find the conditional distribution of X_1 given that $X_1 = -X_2$.

25. Let $\mathbf{X}$ have a three-dimensional normal distribution with mean vector and covariance matrix

$$\boldsymbol{\mu} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \mathbf{\Lambda} = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 2 \end{pmatrix},$$

respectively. Set $Y_1 = X_1 + X_2 + X_3$ and $Y_2 = X_1 + X_3$. Determine the conditional distribution of Y_1 given that $Y_2 = 0$.

26. Let $\mathbf{X} \in N(\mathbf{0}, \mathbf{\Lambda})$, where

$$\mathbf{\Lambda} = \begin{pmatrix} 2 & 1 & -1 \\ 1 & 3 & 0 \\ -1 & 0 & 5 \end{pmatrix}$$

Find the conditional distribution of X_1 given that $X_1 = X_2$ and $X_1 + X_2 + X_3 = 0$.

27. The random vector $\mathbf{X}$ has a three-dimensional normal distribution with expectation $\mathbf{0}$ and covariance matrix $\mathbf{\Lambda}$ given by

$$\mathbf{\Lambda} = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}.$$

Find the distribution of X_2 given that $X_1 = X_2 = X_3$.

28. Let $\mathbf{X} \in N(\mathbf{0}, \mathbf{\Lambda})$, where

$$\mathbf{\Lambda} = \begin{pmatrix} 1 & -\frac{1}{2} & \frac{3}{2} \\ -\frac{1}{2} & 2 & -1 \\ \frac{3}{2} & -1 & 4 \end{pmatrix}.$$

Determine the conditional distribution of $(X_1, X_1 + X_2)'$ given that $X_1 + X_2 + X_3 = 0$.

29. Suppose that the characteristic function of $(X, Y, Z)'$ is

$$\varphi(s,t,u) = \exp\{2is - s^2 - 2t^2 - 4u^2 - 2st + 2su\}.$$

Compute the conditional distribution of $X + Z$ given that $X + Y = 0$.

30. Let X_1, X_2, and X_3 have a joint moment generating function as follows:

$$\psi(t_1, t_2, t_3) = \exp\{2t_1 - t_3 + t_1^2 + 2t_2^2 + 3t_3^2 + 2t_1t_2 - 2t_1t_3\}.$$

Determine the conditional distribution of $X_1 + X_3$ given that $X_1 + X_2 = 1$.

31. The moment generating function of $(X, Y, Z)'$ is

$$\psi(s,t,u) = \exp\left\{\frac{s^2}{2} + t^2 + 2u^2 - \frac{st}{2} + \frac{3su}{2} - \frac{tu}{2}\right\}.$$

Determine the conditional distribution of X given that $X + Z = 0$ and $Y + Z = 1$.

32. Suppose $(X, Y, Z)'$ is normal with density

$$C \cdot \exp\Big\{-\frac{1}{2}(4x^2 + 3y^2 + 5z^2 + 2xy + 6xz + 4zy)\Big\},$$

where C is a normalizing constant. Determine the conditional distribution of X given that $X + Z = 1$ and $Y + Z = 0$.

33. Let X and Y be random variables, such that

$$Y \mid X = x \in N(x, \tau^2) \quad \text{with} \quad X \in N(\mu, \sigma^2).$$

(a) Compute $E\,Y$, $\operatorname{Var} Y$ and $\operatorname{Cov}(X, Y)$.
(b) Determine the distribution of the vector $(X, Y)'$.
(c) Determine the (posterior) distribution of $X \mid Y = y$.

34. Let X and Y be jointly normal with means 0, variances 1, and correlation coefficient ρ. Compute the moment generating function of $X \cdot Y$ for
(a) $\rho = 0$, and
(b) general ρ.

35. Suppose X_1, X_2, and X_3 are independent and $N(0,1)$-distributed. Compute the moment generating function of $Y = X_1X_2 + X_1X_3 + X_2X_3$.

36. If X and Y are independent, $N(0,1)$-distributed random variables, then $X^2 + Y^2 \in \chi^2(2)$ (recall Exercise 3.3.6). Now, let X and Y be jointly normal with means 0, variances 1, and correlation coefficient ρ. In this case $X^2 + Y^2$ has a *noncentral* $\chi^2(2)$-distribution. Determine the moment generating function of that distribution.

37. Let $(X, Y)'$ have a two-dimensional normal distribution with means 0, variances 1, and correlation coefficient ρ, $|\rho| < 1$. Determine the distribution of $(X^2 - 2\rho XY + Y^2)/(1 - \rho^2)$ by computing its moment generating function.
    Remark. Recall Example 8.6 and Remark 9.1.
38. Let X_1, X_2, $\dots$, X_n be independent, $N(0,1)$-distributed random variables, and set $\bar{X}_k = \frac{1}{k-1}\sum_{i=1}^{k-1} X_i$, $2 \le k \le n$. Show that

$$Q = \sum_{k=2}^{n} \frac{k-1}{k}\big(X_k - \bar{X}_k\big)^2$$

    is χ^2-distributed. What is the number of degrees of freedom?
39. Let X_1, X_2, and X_3 be independent, $N(1,1)$-distributed random variables. Set $U = X_1 + X_2 + X_3$ and $V = X_1 + 2X_2 + 3X_3$. Determine the constants a and b so that $E\big(U - a - bV\big)^2$ is minimized.
40. Let X and Y be independent, $N(0,1)$-distributed random variables. Then $X + Y$ and $X - Y$ are independent; see Example 7.1. The purpose of this problem is to point out a (partial) converse. Suppose that X and Y are independent random variables with common distribution function F. Suppose, further, that F is symmetric and that $\sigma^2 = E\,X^2 < \infty$. Let φ be the characteristic function of X (and Y). Show that if $X + Y$ and $X - Y$ are independent then we have

$$\varphi(t) = \big(\varphi(t/2)\big)^4.$$

    Use this relation to show that $\varphi(t) = e^{-\sigma^2 t^2/2}$. Finally, conclude that F is the distribution function of a normal distribution ($N(0, \sigma^2)$).
    Remark 1. The assumptions that the distribution is symmetric and the variance is finite are not necessary. However, without them the problem becomes much more difficult.
    Remark 2. Results of this kind are called *characterization theorems*. Another characterization of the normal distribution is provided by the following famous theorem due to the Swedish probabilist and statistician Harald Cramér (1893–1985): If X and Y are independent random variables such that $X + Y$ has a normal distribution, then X and Y are both normal.