1

A REVIEW OF ELEMENTARY MATRIX ALGEBRA

1.1 INTRODUCTION

In this chapter, we review some of the basic operations and fundamental properties involved in matrix algebra. In most cases, properties will be stated without proof, but in some cases, when instructive, proofs will be presented. We end the chapter with a brief discussion of random variables and random vectors, expected values of random variables, and some important distributions encountered elsewhere in the book.

1.2 DEFINITIONS AND NOTATION

Except when stated otherwise, a scalar such as α will represent a real number. A matrix A of size $m \times n$ is the $m \times n$ rectangular array of scalars given by

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix},$$

and sometimes it is simply identified as $A = (a_{ij})$. Sometimes it also will be convenient to refer to the (i, j)th element of A, as $(A)_{ij}$; that is, $a_{ij} = (A)_{ij}$. If $m = n$,

Matrix Analysis for Statistics, Third Edition. James R. Schott.

Companion Website: www.wiley.com/go/Schott/MatrixAnalysis3e

then A is called a square matrix of order m, whereas A is referred to as a rectangular matrix when $m \neq n$. An $m \times 1$ matrix

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}$$

is called a column vector or simply a vector. The element a_i is referred to as the ith component of $\boldsymbol{a}$. A $1 \times n$ matrix is called a row vector. The ith row and jth column of the matrix A will be denoted by $(A)_{i\cdot}$ and $(A)_{\cdot j}$, respectively. We will usually use capital letters to represent matrices and lowercase bold letters for vectors.

The diagonal elements of the $m \times m$ matrix A are $a_{11}, a_{22}, \dots, a_{mm}$. If all other elements of A are equal to 0, A is called a diagonal matrix and can be identified as $A = \text{diag}(a_{11}, \dots, a_{mm})$. If, in addition, $a_{ii} = 1$ for $i = 1, \dots, m$ so that $A = \text{diag}(1, \dots, 1)$, then the matrix A is called the identity matrix of order m and will be written as $A = I_m$ or simply $A = I$ if the order is obvious. If $A = \text{diag}(a_{11}, \dots, a_{mm})$ and b is a scalar, then we will use A^b to denote the diagonal matrix $\text{diag}(a_{11}^b, \dots, a_{mm}^b)$. For any $m \times m$ matrix A, D_A will denote the diagonal matrix with diagonal elements equal to those of A, and for any $m \times 1$ vector $\boldsymbol{a}$, $D_{\boldsymbol{a}}$ denotes the diagonal matrix with diagonal elements equal to the components of $\boldsymbol{a}$; that is, $D_A = \text{diag}(a_{11}, \dots, a_{mm})$ and $D_{\boldsymbol{a}} = \text{diag}(a_1, \dots, a_m)$.

A triangular matrix is a square matrix that is either an upper triangular matrix or a lower triangular matrix. An upper triangular matrix is one that has all of its elements below the diagonal equal to 0, whereas a lower triangular matrix has all of its elements above the diagonal equal to 0. A strictly upper triangular matrix is an upper triangular matrix that has each of its diagonal elements equal to 0. A strictly lower triangular matrix is defined similarly.

The ith column of the $m \times m$ identity matrix will be denoted by $\boldsymbol{e}_i$; that is, $\boldsymbol{e}_i$ is the $m \times 1$ vector that has its ith component equal to 1 and all of its other components equal to 0. When the value of m is not obvious, we will make it more explicit by writing $\boldsymbol{e}_i$ as $\boldsymbol{e}_{i,m}$. The $m \times m$ matrix whose only nonzero element is a 1 in the (i, j)th position will be identified as E_{ij}.

The scalar zero is written 0, whereas a vector of zeros, called a null vector, will be denoted by $\mathbf{0}$, and a matrix of zeros, called a null matrix, will be denoted by (0). The $m \times 1$ vector having each component equal to 1 will be denoted by $\mathbf{1}_m$ or simply $\mathbf{1}$ when the size of the vector is obvious.

1.3 MATRIX ADDITION AND MULTIPLICATION

The sum of two matrices A and B is defined if they have the same number of rows and the same number of columns; in this case,

$$A + B = (a_{ij} + b_{ij}).$$

The product of a scalar α and a matrix A is

$$\alpha A = A\alpha = (\alpha a_{ij}).$$

The premultiplication of the matrix B by the matrix A is defined only if the number of columns of A equals the number of rows of B. Thus, if A is $m \times p$ and B is $p \times n$, then $C = AB$ will be the $m \times n$ matrix which has its (i, j)th element, c_{ij}, given by

$$c_{ij} = (A)_{i\cdot}(B)_{\cdot j} = \sum_{k=1}^{p} a_{ik}b_{kj}.$$

A similar definition exists for BA, the postmultiplication of B by A, if the number of columns of B equals the number of rows of A. When both products are defined, we will not have, in general, $AB = BA$. If the matrix A is square, then the product AA, or simply A^2, is defined. In this case, if we have $A^2 = A$, then A is said to be an idempotent matrix.

The following basic properties of matrix addition and multiplication in Theorem 1.1 are easy to verify.

Theorem 1.1 Let α and β be scalars and A, B, and C be matrices. Then, when the operations involved are defined, the following properties hold:

(a) $A + B = B + A$.
(b) $(A + B) + C = A + (B + C)$.
(c) $\alpha(A + B) = \alpha A + \alpha B$.
(d) $(\alpha + \beta)A = \alpha A + \beta A$.
(e) $A - A = A + (-A) = (0)$.
(f) $A(B + C) = AB + AC$.
(g) $(A + B)C = AC + BC$.
(h) $(AB)C = A(BC)$.

1.4 THE TRANSPOSE

The transpose of an $m \times n$ matrix A is the $n \times m$ matrix A' obtained by interchanging the rows and columns of A. Thus, the (i, j)th element of A' is a_{ji}. If A is $m \times p$ and B is $p \times n$, then the (i, j)th element of $(AB)'$ can be expressed as

$$\begin{aligned}((AB)')_{ij} &= (AB)_{ji} = (A)_{j\cdot}(B)_{\cdot i} = \sum_{k=1}^{p} a_{jk}b_{ki} \\ &= (B')_{i\cdot}(A')_{\cdot j} = (B'A')_{ij}.\end{aligned}$$

Thus, evidently $(AB)' = B'A'$. This property along with some other results involving the transpose are summarized in Theorem 1.2.

Theorem 1.2 Let α and β be scalars and A and B be matrices. Then, when defined, the following properties hold:

(a) $(\alpha A)' = \alpha A'$.
(b) $(A')' = A$.
(c) $(\alpha A + \beta B)' = \alpha A' + \beta B'$.
(d) $(AB)' = B'A'$.

If A is $m \times m$, that is, A is a square matrix, then A' is also $m \times m$. In this case, if $A = A'$, then A is called a symmetric matrix, whereas A is called a skew-symmetric if $A = -A'$.

The transpose of a column vector is a row vector, and in some situations, we may write a matrix as a column vector times a row vector. For instance, the matrix E_{ij} defined in Section 1.2 can be expressed as $E_{ij} = \boldsymbol{e}_i \boldsymbol{e}_j'$. More generally, $\boldsymbol{e}_{i,m} \boldsymbol{e}_{j,n}'$ yields an $m \times n$ matrix having 1, as its only nonzero element, in the (i, j)th position, and if A is an $m \times n$ matrix, then

$$A = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} \boldsymbol{e}_{i,m} \boldsymbol{e}_{j,n}'.$$

1.5 THE TRACE

The trace is a function that is defined only on square matrices. If A is an $m \times m$ matrix, then the trace of A, denoted by $\mathrm{tr}(A)$, is defined to be the sum of the diagonal elements of A; that is,

$$\mathrm{tr}(A) = \sum_{i=1}^{m} a_{ii}.$$

Now if A is $m \times n$ and B is $n \times m$, then AB is $m \times m$ and

$$\begin{aligned}
\mathrm{tr}(AB) &= \sum_{i=1}^{m} (AB)_{ii} = \sum_{i=1}^{m} (A)_{i\cdot}(B)_{\cdot i} = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ji} \\
&= \sum_{j=1}^{n} \sum_{i=1}^{m} b_{ji} a_{ij} = \sum_{j=1}^{n} (B)_{j\cdot}(A)_{\cdot j} \\
&= \sum_{j=1}^{n} (BA)_{jj} = \mathrm{tr}(BA).
\end{aligned}$$

This property of the trace, along with some others, is summarized in Theorem 1.3.

Theorem 1.3 Let α be a scalar and A and B be matrices. Then, when the appropriate operations are defined, we have the following properties:

(a) $\operatorname{tr}(A') = \operatorname{tr}(A)$.
(b) $\operatorname{tr}(\alpha A) = \alpha \operatorname{tr}(A)$.
(c) $\operatorname{tr}(A + B) = \operatorname{tr}(A) + \operatorname{tr}(B)$.
(d) $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.
(e) $\operatorname{tr}(A'A) = 0$ if and only if $A = (0)$.

1.6 THE DETERMINANT

The determinant is another function defined on square matrices. If A is an $m \times m$ matrix, then its determinant, denoted by $|A|$, is given by

$$\begin{aligned} |A| &= \sum (-1)^{f(i_1, \dots, i_m)} a_{1i_1} a_{2i_2} \cdots a_{mi_m} \\ &= \sum (-1)^{f(i_1, \dots, i_m)} a_{i_1 1} a_{i_2 2} \cdots a_{i_m m}, \end{aligned}$$

where the summation is taken over all permutations $(i_1, \dots, i_m)$ of the set of integers $(1, \dots, m)$, and the function $f(i_1, \dots, i_m)$ equals the number of transpositions necessary to change $(i_1, \dots, i_m)$ to an increasing sequence of components, that is, to $(1, \dots, m)$. A transposition is the interchange of two of the integers. Although f is not unique, it is uniquely even or odd, so that $|A|$ is uniquely defined. Note that the determinant produces all products of m terms of the elements of the matrix A such that exactly one element is selected from each row and each column of A.

Using the formula for the determinant, we find that $|A| = a_{11}$ when $m = 1$. If A is 2×2, we have

$$|A| = a_{11}a_{22} - a_{12}a_{21},$$

and when A is 3×3, we get

$$\begin{aligned} |A| = {} & a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ & - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}. \end{aligned}$$

The following properties of the determinant in Theorem 1.4 are fairly straightforward to verify using the definition of a determinant.

Theorem 1.4 If α is a scalar and A is an $m \times m$ matrix, then the following properties hold:

(a) $|A'| = |A|$.
(b) $|\alpha A| = \alpha^m |A|$.

(c) If A is a diagonal matrix, then $|A| = a_{11} \cdots a_{mm} = \prod_{i=1}^{m} a_{ii}$.
(d) If all elements of a row (or column) of A are zero, $|A| = 0$.
(e) The interchange of two rows (or columns) of A changes the sign of $|A|$.
(f) If all elements of a row (or column) of A are multiplied by α, then the determinant is multiplied by α.
(g) The determinant of A is unchanged when a multiple of one row (or column) is added to another row (or column).
(h) If two rows (or columns) of A are proportional to one another, $|A| = 0$.

An alternative expression for $|A|$ can be given in terms of the cofactors of A. The minor of the element a_{ij}, denoted by m_{ij}, is the determinant of the $(m-1) \times (m-1)$ matrix obtained after removing the ith row and jth column from A. The corresponding cofactor of a_{ij}, denoted by A_{ij}, is then given as $A_{ij} = (-1)^{i+j} m_{ij}$.

Theorem 1.5 For any $i = 1, \ldots, m$, the determinant of the $m \times m$ matrix A can be obtained by expanding along the ith row,

$$|A| = \sum_{j=1}^{m} a_{ij} A_{ij}, \tag{1.1}$$

or expanding along the ith column,

$$|A| = \sum_{j=1}^{m} a_{ji} A_{ji}. \tag{1.2}$$

Proof. We will just prove (1.1), as (1.2) can easily be obtained by applying (1.1) to A'. We first consider the result when $i = 1$. Clearly

$$\begin{aligned} |A| &= \sum (-1)^{f(i_1, \ldots, i_m)} a_{1i_1} a_{2i_2} \cdots a_{mi_m} \\ &= a_{11} b_{11} + \cdots + a_{1m} b_{1m}, \end{aligned}$$

where

$$a_{1j} b_{1j} = \sum (-1)^{f(i_1, \ldots, i_m)} a_{1i_1} a_{2i_2} \cdots a_{mi_m},$$

and the summation is over all permutations for which $i_1 = j$. Since $(-1)^{f(j, i_2, \ldots, i_m)} = (-1)^{j-1} (-1)^{f(i_2, \ldots, i_m)}$, this implies that

$$b_{1j} = \sum (-1)^{j-1} (-1)^{f(i_2, \ldots, i_m)} a_{2i_2} \cdots a_{mi_m},$$

where the summation is over all permutations $(i_2, \ldots, i_m)$ of $(1, \ldots, j-1, j+1, \ldots, m)$. If C is the $(m-1) \times (m-1)$ matrix obtained from A by deleting its 1st row and jth column, then b_{1j} can be written

$$b_{1j} = (-1)^{j-1} \sum (-1)^{f(i_1, \dots, i_{m-1})} c_{1i_1} \cdots c_{m-1 i_{m-1}} = (-1)^{j-1}|C|$$
$$= (-1)^{j-1} m_{1j} = (-1)^{1+j} m_{1j} = A_{1j},$$

where the summation is over all permutations $(i_1, \dots, i_{m-1})$ of $(1, \dots, m-1)$ and m_{1j} is the minor of a_{1j}. Thus,

$$|A| = \sum_{j=1}^{m} a_{1j} b_{1j} = \sum_{j=1}^{m} a_{1j} A_{1j},$$

as is required. To prove (1.1) when $i > 1$, let D be the $m \times m$ matrix for which $(D)_{1\cdot} = (A)_{i\cdot}$, $(D)_{j\cdot} = (A)_{j-1\cdot}$, for $j = 2, \dots, i$, and $(D)_{j\cdot} = (A)_{j\cdot}$ for $j = i + 1, \dots, m$. Then $A_{ij} = (-1)^{i-1} D_{1j}$, $a_{ij} = d_{1j}$ and $|A| = (-1)^{i-1}|D|$. Thus, since we have already established (1.1) when $i = 1$, we have

$$|A| = (-1)^{i-1}|D| = (-1)^{i-1} \sum_{j=1}^{m} d_{1j} D_{1j} = \sum_{j=1}^{m} a_{ij} A_{ij},$$

and so the proof is complete. □

Our next result indicates that if the cofactors of a row or column are matched with the elements from a different row or column, the expansion reduces to 0.

Theorem 1.6 If A is an $m \times m$ matrix and $k \neq i$, then

$$\sum_{j=1}^{m} a_{ij} A_{kj} = \sum_{j=1}^{m} a_{ji} A_{jk} = 0. \tag{1.3}$$

Example 1.1 We will find the determinant of the 5×5 matrix given by

$$A = \begin{bmatrix} 2 & 1 & 2 & 1 & 1 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 2 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 2 & 2 & 1 \end{bmatrix}.$$

Using the cofactor expansion formula on the first column of A, we obtain

$$|A| = 2 \begin{vmatrix} 0 & 3 & 0 & 0 \\ 0 & 2 & 2 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 2 & 2 & 1 \end{vmatrix},$$

and then using the same expansion formula on the first column of this 4×4 matrix, we get

$$|A| = 2(-1)\begin{vmatrix} 3 & 0 & 0 \\ 2 & 2 & 0 \\ 1 & 1 & 1 \end{vmatrix}.$$

Because the determinant of the 3×3 matrix above is 6, we have

$$|A| = 2(-1)(6) = -12.$$

Consider the $m \times m$ matrix C whose columns are given by the vectors $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_m$; that is, we can write $C = (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_m)$. Suppose that, for some $m \times 1$ vector $\boldsymbol{b} = (b_1, \ldots, b_m)'$ and $m \times m$ matrix $A = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m)$, we have

$$\boldsymbol{c}_1 = A\boldsymbol{b} = \sum_{i=1}^{m} b_i \boldsymbol{a}_i.$$

Then, if we find the determinant of C by expanding along the first column of C, we get

$$\begin{aligned} |C| &= \sum_{j=1}^{m} c_{j1} C_{j1} = \sum_{j=1}^{m} \left(\sum_{i=1}^{m} b_i a_{ji} \right) C_{j1} \\ &= \sum_{i=1}^{m} b_i \left(\sum_{j=1}^{m} a_{ji} C_{j1} \right) = \sum_{i=1}^{m} b_i |(\boldsymbol{a}_i, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_m)|, \end{aligned}$$

so that the determinant of C is a linear combination of m determinants. If B is an $m \times m$ matrix and we now define $C = AB$, then by applying the previous derivation on each column of C, we find that

$$\begin{aligned} |C| &= \left| \left(\sum_{i_1=1}^{m} b_{i_1 1} \boldsymbol{a}_{i_1}, \ldots, \sum_{i_m=1}^{m} b_{i_m m} \boldsymbol{a}_{i_m} \right) \right| \\ &= \sum_{i_1=1}^{m} \cdots \sum_{i_m=1}^{m} b_{i_1 1} \cdots b_{i_m m} |(\boldsymbol{a}_{i_1}, \ldots, \boldsymbol{a}_{i_m})| \\ &= \sum b_{i_1 1} \cdots b_{i_m m} |(\boldsymbol{a}_{i_1}, \ldots, \boldsymbol{a}_{i_m})|, \end{aligned}$$

where this final sum is only over all permutations of $(1, \ldots, m)$, because Theorem 1.4(h) implies that

$$|(\boldsymbol{a}_{i_1}, \ldots, \boldsymbol{a}_{i_m})| = 0$$

if $i_j = i_k$ for any $j \neq k$. Finally, reordering the columns in $|(\boldsymbol{a}_{i_1}, \ldots, \boldsymbol{a}_{i_m})|$ and using Theorem 1.4(e), we have

$$|C| = \sum b_{i_1 1} \cdots b_{i_m m} (-1)^{f(i_1, \ldots, i_m)} |(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m)| = |B||A|.$$

This very useful result is summarized in Theorem 1.7.

Theorem 1.7 If both A and B are square matrices of the same order, then

$$|AB| = |A||B|.$$

1.7 THE INVERSE

An $m \times m$ matrix A is said to be a nonsingular matrix if $|A| \neq 0$ and a singular matrix if $|A| = 0$. If A is nonsingular, a nonsingular matrix denoted by A^{-1} and called the inverse of A exists, such that

$$AA^{-1} = A^{-1}A = I_m. \tag{1.4}$$

This inverse is unique because, if B is another $m \times m$ matrix satisfying the inverse formula (1.4) for A, then $BA = I_m$, and so

$$B = BI_m = BAA^{-1} = I_m A^{-1} = A^{-1}.$$

The following basic properties of the matrix inverse in Theorem 1.8 can be easily verified by using (1.4).

Theorem 1.8 If α is a nonzero scalar, and A and B are nonsingular $m \times m$ matrices, then the following properties hold:

(a) $(\alpha A)^{-1} = \alpha^{-1} A^{-1}$.
(b) $(A')^{-1} = (A^{-1})'$.
(c) $(A^{-1})^{-1} = A$.
(d) $|A^{-1}| = |A|^{-1}$.
(e) If $A = \mathrm{diag}(a_{11}, \ldots, a_{mm})$, then $A^{-1} = \mathrm{diag}(a_{11}^{-1}, \ldots, a_{mm}^{-1})$.
(f) If $A = A'$, then $A^{-1} = (A^{-1})'$.
(g) $(AB)^{-1} = B^{-1}A^{-1}$.

As with the determinant of A, the inverse of A can be expressed in terms of the cofactors of A. Let $A_\#$, called the adjoint of A, be the transpose of the matrix of cofactors of A; that is, the (i, j)th element of $A_\#$ is A_{ji}, the cofactor of a_{ji}. Then

$$AA_\# = A_\# A = \mathrm{diag}(|A|, \ldots, |A|) = |A| I_m,$$

because $(A)_{i\cdot}(A_\#)_{\cdot i} = (A_\#)_{i\cdot}(A)_{\cdot i} = |A|$ follows directly from (1.1) and (1.2), and $(A)_{i\cdot}(A_\#)_{\cdot j} = (A_\#)_{i\cdot}(A)_{\cdot j} = 0$, for $i \neq j$ follows from (1.3). The equation above then yields the relationship

$$A^{-1} = |A|^{-1} A_\#$$

when $|A| \neq 0$. Thus, for instance, if A is a 2×2 nonsingular matrix, then

$$A^{-1} = |A|^{-1} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}.$$

Similarly when $m = 3$, we get $A^{-1} = |A|^{-1} A_{\#}$, where

$$A_{\#} = \begin{bmatrix} a_{22}a_{33} - a_{23}a_{32} & -(a_{12}a_{33} - a_{13}a_{32}) & a_{12}a_{23} - a_{13}a_{22} \\ -(a_{21}a_{33} - a_{23}a_{31}) & a_{11}a_{33} - a_{13}a_{31} & -(a_{11}a_{23} - a_{13}a_{21}) \\ a_{21}a_{32} - a_{22}a_{31} & -(a_{11}a_{32} - a_{12}a_{31}) & a_{11}a_{22} - a_{12}a_{21} \end{bmatrix}.$$

The relationship between the inverse of a matrix product and the product of the inverses, given in Theorem 1.8(g), is a very useful property. Unfortunately, such a nice relationship does not exist between the inverse of a sum and the sum of the inverses. We do, however, have Theorem 1.9 which is sometimes useful.

Theorem 1.9 Suppose A and B are nonsingular matrices, with A being $m \times m$ and B being $n \times n$. For any $m \times n$ matrix C and any $n \times m$ matrix D, it follows that if $A + CBD$ is nonsingular, then

$$(A + CBD)^{-1} = A^{-1} - A^{-1}C(B^{-1} + DA^{-1}C)^{-1}DA^{-1}.$$

Proof. The proof simply involves verifying that $(A + CBD)(A + CBD)^{-1} = I_m$ for $(A + CBD)^{-1}$ given above. We have

$$\begin{aligned} &(A + CBD)\{A^{-1} - A^{-1}C(B^{-1} + DA^{-1}C)^{-1}DA^{-1}\} \\ &\quad = I_m - C(B^{-1} + DA^{-1}C)^{-1}DA^{-1} + CBDA^{-1} \\ &\qquad - CBDA^{-1}C(B^{-1} + DA^{-1}C)^{-1}DA^{-1} \\ &\quad = I_m - C\{(B^{-1} + DA^{-1}C)^{-1} - B \\ &\qquad + BDA^{-1}C(B^{-1} + DA^{-1}C)^{-1}\}DA^{-1} \\ &\quad = I_m - C\{B(B^{-1} + DA^{-1}C)(B^{-1} + DA^{-1}C)^{-1} - B\}DA^{-1} \\ &\quad = I_m - C\{B - B\}DA^{-1} = I_m, \end{aligned}$$

and so the result follows. □

The expression given for $(A + CBD)^{-1}$ in Theorem 1.9 involves the inverse of the matrix $B^{-1} + DA^{-1}C$. It can be shown (see Problem 7.12) that the conditions of the theorem guarantee that this inverse exists. If $m = n$ and C and D are identity matrices, then we obtain Corollary 1.9.1 of Theorem 1.9.

Corollary 1.9.1 Suppose that A, B and $A + B$ are all $m \times m$ nonsingular matrices. Then

$$(A + B)^{-1} = A^{-1} - A^{-1}(B^{-1} + A^{-1})^{-1}A^{-1}.$$

We obtain Corollary 1.9.2 of Theorem 1.9 when $n = 1$.

Corollary 1.9.2 Let A be an $m \times m$ nonsingular matrix. If $\boldsymbol{c}$ and $\boldsymbol{d}$ are both $m \times 1$ vectors and $A + \boldsymbol{cd}'$ is nonsingular, then

$$(A + \boldsymbol{cd}')^{-1} = A^{-1} - A^{-1}\boldsymbol{cd}'A^{-1}/(1 + \boldsymbol{d}'A^{-1}\boldsymbol{c}).$$

Example 1.2 Theorem 1.9 can be particularly useful when m is larger than n and the inverse of A is fairly easy to compute. For instance, suppose we have $A = I_5$,

$$B = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ -1 & 1 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \quad D' = \begin{bmatrix} 1 & -1 \\ -1 & 2 \\ 0 & 1 \\ 1 & 0 \\ -1 & 1 \end{bmatrix},$$

from which we obtain

$$G = A + CBD = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ -1 & 6 & 4 & 3 & 1 \\ -1 & 2 & 2 & 0 & 1 \\ -2 & 6 & 4 & 3 & 2 \\ -1 & 4 & 3 & 2 & 2 \end{bmatrix}.$$

It is somewhat tedious to compute the inverse of this 5×5 matrix directly. However, the calculations in Theorem 1.9 are fairly straightforward. Clearly, $A^{-1} = I_5$ and

$$B^{-1} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix},$$

so that

$$(B^{-1} + DA^{-1}C) = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} + \begin{bmatrix} -2 & 0 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 2 & 5 \end{bmatrix}$$

and

$$(B^{-1} + DA^{-1}C)^{-1} = \begin{bmatrix} 2.5 & 0.5 \\ -1 & 0 \end{bmatrix}.$$

Thus, we find that

$$G^{-1} = I_5 - C(B^{-1} + DA^{-1}C)^{-1}D$$

$$= \begin{bmatrix} -1 & 1.5 & -0.5 & -2.5 & 2 \\ -3 & 3 & -1 & -4 & 3 \\ 3 & -2.5 & 1.5 & 3.5 & -3 \\ 2 & -2 & 0 & 3 & -2 \\ -1 & 0.5 & -0.5 & -1.5 & 2 \end{bmatrix}.$$

1.8 PARTITIONED MATRICES

Occasionally we will find it useful to partition a given matrix into submatrices. For instance, suppose A is $m \times n$ and the positive integers m_1, m_2, n_1, n_2 are such that $m = m_1 + m_2$ and $n = n_1 + n_2$. Then one way of writing A as a partitioned matrix is

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

where A_{11} is $m_1 \times n_1$, A_{12} is $m_1 \times n_2$, A_{21} is $m_2 \times n_1$, and A_{22} is $m_2 \times n_2$. That is, A_{11} is the matrix consisting of the first m_1 rows and n_1 columns of A, A_{12} is the matrix consisting of the first m_1 rows and last n_2 columns of A, and so on. Matrix operations can be expressed in terms of the submatrices of the partitioned matrix. For example, suppose B is an $n \times p$ matrix partitioned as

$$B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix},$$

where B_{11} is $n_1 \times p_1$, B_{12} is $n_1 \times p_2$, B_{21} is $n_2 \times p_1$, B_{22} is $n_2 \times p_2$, and $p = p_1 + p_2$. Then the premultiplication of B by A can be expressed in partitioned form as

$$AB = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix}.$$

Matrices can be partitioned into submatrices in other ways besides this 2×2 partitioned form. For instance, we could partition only the columns of A, yielding the expression

$$A = \begin{bmatrix} A_1 & A_2 \end{bmatrix},$$

where A_1 is $m \times n_1$ and A_2 is $m \times n_2$. A more general situation is one in which the rows of A are partitioned into r groups and the columns of A are partitioned into c groups so that A can be written as

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1c} \\ A_{21} & A_{22} & \cdots & A_{2c} \\ \vdots & \vdots & & \vdots \\ A_{r1} & A_{r2} & \cdots & A_{rc} \end{bmatrix},$$

where the submatrix A_{ij} is $m_i \times n_j$ and the integers $m_1, \ldots, m_r$ and $n_1, \ldots, n_c$ are such that

$$\sum_{i=1}^{r} m_i = m \quad \text{and} \quad \sum_{j=1}^{c} n_j = n.$$

This matrix A is said to be in block diagonal form if $r = c$, A_{ii} is a square matrix for each i, and A_{ij} is a null matrix for all i and j for which $i \neq j$. In this case, we will write $A = \text{diag}(A_{11}, \ldots, A_{rr})$; that is,

$$\text{diag}(A_{11}, \ldots, A_{rr}) = \begin{bmatrix} A_{11} & (0) & \cdots & (0) \\ (0) & A_{22} & \cdots & (0) \\ \vdots & \vdots & & \vdots \\ (0) & (0) & \cdots & A_{rr} \end{bmatrix}.$$

Example 1.3 Suppose we wish to compute the transpose product AA', where the 5×5 matrix A is given by

$$A = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ -1 & -1 & -1 & 2 & 0 \\ -1 & -1 & -1 & 0 & 2 \end{bmatrix}.$$

The computation can be simplified by observing that A may be written as

$$A = \begin{bmatrix} I_3 & \mathbf{1}_3\mathbf{1}_2' \\ -\mathbf{1}_2\mathbf{1}_3' & 2I_2 \end{bmatrix}.$$

As a result, we have

$$\begin{aligned} AA' &= \begin{bmatrix} I_3 & \mathbf{1}_3\mathbf{1}_2' \\ -\mathbf{1}_2\mathbf{1}_3' & 2I_2 \end{bmatrix} \begin{bmatrix} I_3 & -\mathbf{1}_3\mathbf{1}_2' \\ \mathbf{1}_2\mathbf{1}_3' & 2I_2 \end{bmatrix} \\ &= \begin{bmatrix} I_3 + \mathbf{1}_3\mathbf{1}_2'\mathbf{1}_2\mathbf{1}_3' & -\mathbf{1}_3\mathbf{1}_2' + 2\mathbf{1}_3\mathbf{1}_2' \\ -\mathbf{1}_2\mathbf{1}_3' + 2\mathbf{1}_2\mathbf{1}_3' & \mathbf{1}_2\mathbf{1}_3'\mathbf{1}_3\mathbf{1}_2' + 4I_2 \end{bmatrix} \\ &= \begin{bmatrix} I_3 + 2\mathbf{1}_3\mathbf{1}_3' & \mathbf{1}_3\mathbf{1}_2' \\ \mathbf{1}_2\mathbf{1}_3' & 3\mathbf{1}_2\mathbf{1}_2' + 4I_2 \end{bmatrix} \\ &= \begin{bmatrix} 3 & 2 & 2 & 1 & 1 \\ 2 & 3 & 2 & 1 & 1 \\ 2 & 2 & 3 & 1 & 1 \\ 1 & 1 & 1 & 7 & 3 \\ 1 & 1 & 1 & 3 & 7 \end{bmatrix}. \end{aligned}$$

1.9 THE RANK OF A MATRIX

Our initial definition of the rank of an $m \times n$ matrix A is given in terms of submatrices. We will see an alternative equivalent definition in terms of the concept of linearly independent vectors in Chapter 2. Most of the material we include in this section can be found in more detail in texts on elementary linear algebra such as Andrilli and Hecker (2010) and Poole (2015).

In general, any matrix formed by deleting rows or columns of A is called a submatrix of A. The determinant of an $r \times r$ submatrix of A is called a minor of order r. For instance, for an $m \times m$ matrix A, we have previously defined what we called the minor of a_{ij}; this is an example of a minor of order $m - 1$. Now the rank of a nonnull $m \times n$ matrix A is r, written $\text{rank}(A) = r$, if at least one of its minors of order r is nonzero while all minors of order $r + 1$ (if there are any) are zero. If A is a null matrix, then $\text{rank}(A) = 0$. If $\text{rank}(A) = \min(m, n)$, then A is said to have full rank. In particular, if $\text{rank}(A) = m$, A has full row rank, and if $\text{rank}(A) = n$, A has full column rank.

The rank of a matrix A is unchanged by any of the following operations, called elementary transformations:

(a) The interchange of two rows (or columns) of A.
(b) The multiplication of a row (or column) of A by a nonzero scalar.
(c) The addition of a scalar multiple of a row (or column) of A to another row (or column) of A.

Thus, the definition of the rank of A is sometimes given as the number of nonzero rows in the reduced row echelon form of A.

Any elementary transformation of A can be expressed as the multiplication of A by a matrix referred to as an elementary transformation matrix. An elementary transformation of the rows of A will be given by the premultiplication of A by an elementary transformation matrix, whereas an elementary transformation of the columns corresponds to a postmultiplication. Elementary transformation matrices are nonsingular, and any nonsingular matrix can be expressed as the product of elementary transformation matrices. Consequently, we have Theorem 1.10.

Theorem 1.10 Let A be an $m \times n$ matrix, B be an $m \times m$ matrix, and C be an $n \times n$ matrix. Then if B and C are nonsingular matrices, it follows that

$$\text{rank}(BAC) = \text{rank}(BA) = \text{rank}(AC) = \text{rank}(A).$$

By using elementary transformation matrices, any matrix A can be transformed into another matrix of simpler form having the same rank as A.

Theorem 1.11 If A is an $m \times n$ matrix of rank $r > 0$, then nonsingular $m \times m$ and $n \times n$ matrices B and C exist, such that $H = BAC$ and $A = B^{-1}HC^{-1}$, where H is given by

(a) I_r if $r = m = n$, (b) $\begin{bmatrix} I_r & (0) \end{bmatrix}$ if $r = m < n$,

(c) $\begin{bmatrix} I_r \\ (0) \end{bmatrix}$ if $r = n < m$, (d) $\begin{bmatrix} I_r & (0) \\ (0) & (0) \end{bmatrix}$ if $r < m, r < n$.

Corollary 1.11.1 is an immediate consequence of Theorem 1.11.

Corollary 1.11.1 Let A be an $m \times n$ matrix with $\text{rank}(A) = r > 0$. Then an $m \times r$ matrix F and an $r \times n$ matrix G exist, such that $\text{rank}(F) = \text{rank}(G) = r$ and $A = FG$.

1.10 ORTHOGONAL MATRICES

An $m \times 1$ vector $\boldsymbol{p}$ is said to be a normalized vector or a unit vector if $\boldsymbol{p}'\boldsymbol{p} = 1$. The $m \times 1$ vectors, $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n$, where $n \leq m$, are said to be orthogonal if $\boldsymbol{p}_i'\boldsymbol{p}_j = 0$ for all $i \neq j$. If in addition, each $\boldsymbol{p}_i$ is a normalized vector, then the vectors are said to be orthonormal. An $m \times m$ matrix P whose columns form an orthonormal set of vectors is called an orthogonal matrix. It immediately follows that

$$P'P = I_m.$$

Taking the determinant of both sides, we see that

$$|P'P| = |P'||P| = |P|^2 = |I_m| = 1.$$

Thus, $|P| = +1$ or -1, so that P is nonsingular, $P^{-1} = P'$, and $PP' = I_m$ in addition to $P'P = I_m$; that is, the rows of P also form an orthonormal set of $m \times 1$ vectors. Some basic properties of orthogonal matrices are summarized in Theorem 1.12.

Theorem 1.12 Let P and Q be $m \times m$ orthogonal matrices and A be any $m \times m$ matrix. Then

(a) $|P| = \pm 1$,
(b) $|P'AP| = |A|$,
(c) PQ is an orthogonal matrix.

One example of an $m \times m$ orthogonal matrix, known as the Helmert matrix, has the form

$$H = \begin{bmatrix} \frac{1}{\sqrt{m}} & \frac{1}{\sqrt{m}} & \frac{1}{\sqrt{m}} & \cdots & \frac{1}{\sqrt{m}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & \cdots & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{m(m-1)}} & \frac{1}{\sqrt{m(m-1)}} & \frac{1}{\sqrt{m(m-1)}} & \cdots & -\frac{(m-1)}{\sqrt{m(m-1)}} \end{bmatrix}.$$

For instance, if $m = 4$, the Helmert matrix is

$$H = \begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} & 0 \\ 1/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} & -3/\sqrt{12} \end{bmatrix}.$$

Note that if $m \neq n$, it is possible for an $m \times n$ matrix P to satisfy one of the identities, $P'P = I_n$ or $PP' = I_m$, but not both. Such a matrix is sometimes referred to as a semiorthogonal matrix.

An $m \times m$ matrix P is called a permutation matrix if each row and each column of P has a single element 1, while all remaining elements are zeros. As a result, the columns of P will be $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m$, the columns of I_m, in some order. Note then that the (h, h)th element of $P'P$ will be $\boldsymbol{e}_i'\boldsymbol{e}_i = 1$ for some i, and the (h, l)th element of $P'P$ will be $\boldsymbol{e}_i'\boldsymbol{e}_j = 0$ for some $i \neq j$ if $h \neq l$; that is, a permutation matrix is a special orthogonal matrix. Since there are $m!$ ways of permuting the columns of I_m, there are $m!$ different permutation matrices of order m. If A is also $m \times m$, then PA creates an $m \times m$ matrix by permuting the rows of A, and AP produces a matrix by permuting the columns of A.

1.11 QUADRATIC FORMS

Let $\boldsymbol{x}$ be an $m \times 1$ vector, $\boldsymbol{y}$ an $n \times 1$ vector, and A an $m \times n$ matrix. Then the function of $\boldsymbol{x}$ and $\boldsymbol{y}$ given by

$$\boldsymbol{x}'A\boldsymbol{y} = \sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j a_{ij}$$

is sometimes called a bilinear form in $\boldsymbol{x}$ and $\boldsymbol{y}$. We will be most interested in the special case in which $m = n$, so that A is $m \times m$, and $\boldsymbol{x} = \boldsymbol{y}$. In this case, the function above reduces to the function of $\boldsymbol{x}$,

$$f(\boldsymbol{x}) = \boldsymbol{x}'A\boldsymbol{x} = \sum_{i=1}^{m}\sum_{j=1}^{m} x_i x_j a_{ij},$$

which is called a quadratic form in $\boldsymbol{x}$; A is referred to as the matrix of the quadratic form. We will always assume that A is a symmetric matrix because, if it is not, A may be replaced by $B = \frac{1}{2}(A + A')$, which is symmetric, without altering $f(\boldsymbol{x})$; that is,

$$\begin{aligned}\boldsymbol{x}'B\boldsymbol{x} &= \frac{1}{2}\boldsymbol{x}'(A + A')\boldsymbol{x} = \frac{1}{2}(\boldsymbol{x}'A\boldsymbol{x} + \boldsymbol{x}'A'\boldsymbol{x}) \\ &= \frac{1}{2}(\boldsymbol{x}'A\boldsymbol{x} + \boldsymbol{x}'A\boldsymbol{x}) = \boldsymbol{x}'A\boldsymbol{x}\end{aligned}$$

because $\boldsymbol{x}'A'\boldsymbol{x} = (\boldsymbol{x}'A'\boldsymbol{x})' = \boldsymbol{x}'A\boldsymbol{x}$. As an example, consider the function

$$f(\boldsymbol{x}) = x_1^2 + 3x_2^2 + 2x_3^2 + 2x_1x_2 - 2x_2x_3,$$

where $\boldsymbol{x}$ is 3×1. The symmetric matrix A satisfying $f(\boldsymbol{x}) = \boldsymbol{x}'A\boldsymbol{x}$ is given by

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 3 & -1 \\ 0 & -1 & 2 \end{bmatrix}.$$

Every symmetric matrix A and its associated quadratic form is classified into one of the following five categories:

(a) If $\boldsymbol{x}'A\boldsymbol{x} > 0$ for all $\boldsymbol{x} \neq \mathbf{0}$, then A is positive definite.

(b) If $\boldsymbol{x}'A\boldsymbol{x} \geq 0$ for all $\boldsymbol{x}$ and $\boldsymbol{x}'A\boldsymbol{x} = 0$ for some $\boldsymbol{x} \neq \mathbf{0}$, then A is positive semidefinite.

(c) If $\boldsymbol{x}'A\boldsymbol{x} < 0$ for all $\boldsymbol{x} \neq \mathbf{0}$, then A is negative definite.

(d) If $\boldsymbol{x}'A\boldsymbol{x} \leq 0$ for all $\boldsymbol{x}$ and $\boldsymbol{x}'A\boldsymbol{x} = 0$ for some $\boldsymbol{x} \neq \mathbf{0}$, then A is negative semidefinite.

(e) If $\boldsymbol{x}'A\boldsymbol{x} > 0$ for some $\boldsymbol{x}$ and $\boldsymbol{x}'A\boldsymbol{x} < 0$ for some $\boldsymbol{x}$, then A is indefinite.

Note that the null matrix is actually both positive semidefinite and negative semidefinite.

Positive definite and negative definite matrices are nonsingular, whereas positive semidefinite and negative semidefinite matrices are singular. Sometimes the term *nonnegative definite* will be used to refer to a symmetric matrix that is either positive definite or positive semidefinite. An $m \times m$ matrix B is called a square root of the nonnegative definite $m \times m$ matrix A if $A = BB'$. Sometimes we will denote such a matrix B as $A^{1/2}$. If B is also symmetric, so that $A = B^2$, then B is called the symmetric square root of A.

Quadratic forms play a prominent role in inferential statistics. In Chapter 11, we will develop some of the most important results involving quadratic forms that are of particular interest in statistics.

1.12 COMPLEX MATRICES

Throughout most of this text, we will be dealing with the analysis of vectors and matrices composed of real numbers or variables. However, there are occasions in which an analysis of a real matrix, such as the decomposition of a matrix in the form of a product of other matrices, leads to matrices that contain complex numbers. For this reason, we will briefly summarize in this section some of the basic notation and terminology regarding complex numbers.

Any complex number c can be written in the form

$$c = a + ib,$$

where a and b are real numbers and i represents the imaginary number $\sqrt{-1}$. The real number a is called the real part of c, whereas b is referred to as the imaginary part of c. Thus, the number c is a real number only if b is 0. If we have two complex numbers, $c_1 = a_1 + ib_1$ and $c_2 = a_2 + ib_2$, then their sum is given by

$$c_1 + c_2 = (a_1 + a_2) + i(b_1 + b_2),$$

whereas their product is given by

$$c_1 c_2 = a_1 a_2 - b_1 b_2 + i(a_1 b_2 + a_2 b_1).$$

Corresponding to each complex number $c = a + ib$ is another complex number denoted by $\overline{c}$ and called the complex conjugate of c. The complex conjugate of c is given by $\overline{c} = a - ib$ and satisfies $c\overline{c} = a^2 + b^2$, so that the product of a complex number and its conjugate results in a real number.

A complex number can be represented geometrically by a point in the complex plane, where one of the axes is the real axis and the other axis is the complex or imaginary axis. Thus, the complex number $c = a + ib$ would be represented by the point (a, b) in this complex plane. Alternatively, we can use the polar coordinates (r, θ), where r is the length of the line from the origin to the point (a, b) and θ is the angle between this line and the positive half of the real axis. The relationship between a and b, and r and θ is then given by

$$a = r\cos(\theta), \quad b = r\sin(\theta).$$

Writing c in terms of the polar coordinates, we have

$$c = r\cos(\theta) + ir\sin(\theta),$$

or, after using Euler's formula, simply $c = re^{i\theta}$. The absolute value, also sometimes called the modulus, of the complex number c is defined to be r. This is, of course, always a nonnegative real number, and because $a^2 + b^2 = r^2$, we have

$$|c| = |a + ib| = \sqrt{a^2 + b^2}.$$

We also find that

$$\begin{aligned}|c_1c_2| &= \sqrt{(a_1a_2 - b_1b_2)^2 + (a_1b_2 + a_2b_1)^2} \\ &= \sqrt{(a_1^2 + b_1^2)(a_2^2 + b_2^2)} = |c_1||c_2|.\end{aligned}$$

Using this identity repeatedly, we also see that for any complex number c and any positive integer n, $|c^n| = |c|^n$.

A useful identity relating a complex number c and its conjugate to the absolute value of c is

$$c\overline{c} = |c|^2.$$

Applying this identity to the sum of two complex numbers $c_1 + c_2$ and noting that $c_1\overline{c}_2 + \overline{c}_1c_2 \leq 2|c_1||c_2|$, we get

$$\begin{aligned}|c_1 + c_2|^2 &= (c_1 + c_2)\overline{(c_1 + c_2)} = (c_1 + c_2)(\overline{c}_1 + \overline{c}_2) \\ &= c_1\overline{c}_1 + c_1\overline{c}_2 + c_2\overline{c}_1 + c_2\overline{c}_2 \\ &\leq |c_1|^2 + 2|c_1||c_2| + |c_2|^2 \\ &= (|c_1| + |c_2|)^2.\end{aligned}$$

From this result, we get the important inequality, $|c_1 + c_2| \leq |c_1| + |c_2|$, known as the triangle inequality.

A complex matrix is simply a matrix whose elements are complex numbers. As a result, a complex matrix can be written as the sum of a real matrix and an imaginary matrix; that is, if C is an $m \times n$ complex matrix then it can be expressed as

$$C = A + iB,$$

where both A and B are $m \times n$ real matrices. The complex conjugate of C, denoted $\overline{C}$, is simply the matrix containing the complex conjugates of the elements of C; that is,

$$\overline{C} = A - iB.$$

The conjugate transpose of C is $C^* = \overline{C}'$. If the complex matrix C is square and $C^* = C$, so that $c_{ij} = \overline{c}_{ji}$, then C is said to be Hermitian. Note that if C is Hermitian and C is a real matrix, then C is symmetric. The $m \times m$ matrix C is said to be unitary if $C^*C = I_m$, which is the generalization of the concept of orthogonal matrices to complex matrices because if C is real, then $C^* = C'$.

1.13 RANDOM VECTORS AND SOME RELATED STATISTICAL CONCEPTS

In this section, we review some of the basic definitions and results in distribution theory that will be needed later in this text. A more comprehensive treatment of this subject can be found in books on statistical theory such as Casella and Berger (2002) or Lindgren (1993). To be consistent with our notation in which we use a capital letter to denote a matrix, a bold lowercase letter for a vector, and a lowercase letter for a scalar, we will use a lowercase letter instead of the more conventional capital letter to denote a scalar random variable.

A random variable x is said to be discrete if its collection of possible values, R_x, is a countable set. In this case, x has a probability function $p_x(t)$ satisfying $p_x(t) = P(x = t)$, for $t \in R_x$, and $p_x(t) = 0$, for $t \notin R_x$. A continuous random variable x, on the other hand, has for its range, R_x, an uncountably infinite set. Associated with each continuous random variable x is a density function $f_x(t)$ satisfying $f_x(t) > 0$, for $t \in R_x$, and $f_x(t) = 0$, for $t \notin R_x$. Probabilities for x are obtained by integration; if B is a subset of the real line, then

$$P(x \in B) = \int_B f_x(t)\, dt.$$

For both discrete and continuous x, we have $P(x \in R_x) = 1$.

The expected value of a real-valued function of x, $g(x)$, gives the average observed value of $g(x)$. This expectation, denoted $E[g(x)]$, is given by

$$E[g(x)] = \sum_{t \in R_x} g(t) p_x(t),$$

if x is discrete, and

$$E[g(x)] = \int_{-\infty}^{\infty} g(t) f_x(t)\, dt,$$

if x is continuous. Properties of the expectation operator follow directly from properties of sums and integrals. For instance, if x is a random variable and α and β are constants, then the expectation operator satisfies the properties

$$E(\alpha) = \alpha$$

and

$$E[\alpha g_1(x) + \beta g_2(x)] = \alpha E[g_1(x)] + \beta E[g_2(x)],$$

where g_1 and g_2 are any real-valued functions. The expected values of a random variable x given by $E(x^k)$, $k = 1, 2, \ldots$ are known as the moments of x. These moments are important for both descriptive and theoretical purposes. The first few moments can be used to describe certain features of the distribution of x. For instance, the first moment or mean of x, $\mu_x = E(x)$, locates a central value of the distribution. The variance of x, denoted σ_x^2 or $\operatorname{var}(x)$, is defined as

$$\sigma_x^2 = \operatorname{var}(x) = E[(x - \mu_x)^2] = E(x^2) - \mu_x^2,$$

so that it is a function of the first and second moments of x. The variance gives a measure of the dispersion of the observed values of x about the central value μ_x. Using properties of expectation, it is easily verified that

$$\text{var}(\alpha + \beta x) = \beta^2 \, \text{var}(x).$$

All of the moments of a random variable x are embedded in a function called the moment generating function of x. This function is defined as a particular expectation; specifically, the moment generating function of x, $m_x(t)$, is given by

$$m_x(t) = E(e^{tx}),$$

provided this expectation exists for values of t in a neighborhood of 0. Otherwise, the moment generating function does not exist. If the moment generating function of x does exist, then we can obtain any moment from it because

$$\left. \frac{d^k}{dt^k} \, m_x(t) \right|_{t=0} = E(x^k).$$

More importantly, the moment generating function characterizes the distribution of x in that, under certain conditions, no two different distributions have the same moment generating function.

We now focus on some particular families of distributions that we will encounter later in this text. A random variable x is said to have a univariate normal distribution with mean μ and variance σ^2, indicated by $x \sim N(\mu, \sigma^2)$, if the density of x is given by

$$f_x(t) = \frac{1}{\sqrt{2\pi}\sigma} \, e^{-(t-\mu)^2/2\sigma^2}, \quad -\infty < t < \infty.$$

The corresponding moment generating function is

$$m_x(t) = e^{\mu t + \sigma^2 t^2/2}.$$

A special member of this family of normal distributions is the standard normal distribution $N(0, 1)$. The importance of this distribution follows from the fact that if $x \sim N(\mu, \sigma^2)$, then the standardizing transformation $z = (x - \mu)/\sigma$ yields a random variable z that has the standard normal distribution. By differentiating the moment generating function of $z \sim N(0, 1)$, it is easy to verify that the first six moments of z, which we will need in Chapter 11, are 0, 1, 0, 3, 0, and 15, respectively.

If r is a positive integer, then a random variable v has a chi-squared distribution with r degrees of freedom, written $v \sim \chi^2_r$, if its density function is

$$f_v(t) = \frac{t^{(r/2)-1} e^{-t/2}}{2^{r/2}\Gamma(r/2)}, \quad t > 0,$$

where $\Gamma(r/2)$ is the gamma function evaluated at $r/2$. The moment generating function of v is given by $m_v(t) = (1 - 2t)^{-r/2}$, for $t < \frac{1}{2}$. The importance of the chi-squared distribution arises from its connection to the normal distribution. If $z \sim N(0,1)$, then $z^2 \sim \chi_1^2$. Further, if $z_1, \dots, z_r$ are independent random variables with $z_i \sim N(0,1)$ for $i = 1, \dots, r$, then

$$\sum_{i=1}^{r} z_i^2 \sim \chi_r^2. \tag{1.5}$$

The chi-squared distribution mentioned above is sometimes referred to as a central chi-squared distribution because it is actually a special case of a more general family of distributions known as the noncentral chi-squared distributions. These noncentral chi-squared distributions are also related to the normal distribution. If $x_1, \dots, x_r$ are independent random variables with $x_i \sim N(\mu_i, 1)$, then

$$\sum_{i=1}^{r} x_i^2 \sim \chi_r^2(\lambda), \tag{1.6}$$

where $\chi_r^2(\lambda)$ denotes the noncentral chi-squared distribution with r degrees of freedom and noncentrality parameter

$$\lambda = \frac{1}{2}\sum_{i=1}^{r} \mu_i^2;$$

that is, the noncentral chi-squared density, which we will not give here, depends not only on the parameter r but also on the parameter λ. Since (1.6) reduces to (1.5) when $\mu_i = 0$ for all i, we see that the distribution $\chi_r^2(\lambda)$ corresponds to the central chi-squared distribution χ_r^2 when $\lambda = 0$.

A distribution related to the chi-squared distribution is the F distribution with r_1 and r_2 degrees of freedom, denoted by F_{r_1,r_2}. If $y \sim F_{r_1,r_2}$, then the density function of y is

$$f_y(t) = \frac{\Gamma\{(r_1+r_2)/2\}}{\Gamma(r_1/2)\Gamma(r_2/2)}\left(\frac{r_1}{r_2}\right)^{r_1/2} t^{(r_1-2)/2}\left(1+\frac{r_1}{r_2}t\right)^{-(r_1+r_2)/2}, \quad t > 0.$$

The importance of this distribution arises from the fact that if v_1 and v_2 are independent random variables with $v_1 \sim \chi_{r_1}^2$ and $v_2 \sim \chi_{r_2}^2$, then the ratio

$$t = \frac{v_1/r_1}{v_2/r_2}$$

has the F distribution with r_1 and r_2 degrees of freedom.

The concept of a random variable can be extended to that of a random vector. A sequence of related random variables $x_1, \dots, x_m$ is modeled by a joint or multivariate probability function $p_{\boldsymbol{x}}(\boldsymbol{t})$ if all of the random variables are discrete, and a multivariate density function $f_{\boldsymbol{x}}(\boldsymbol{t})$ if all of the random variables are continuous, where $\boldsymbol{x} = (x_1, \dots, x_m)'$ and $\boldsymbol{t} = (t_1, \dots, t_m)'$. For instance, if they are continuous and B is a region in R^m, then the probability that $\boldsymbol{x}$ falls in B is

$$P(\boldsymbol{x} \in B) = \int \cdots \int_B f_{\boldsymbol{x}}(\boldsymbol{t})\ dt_1 \cdots dt_m,$$

whereas the expected value of the real-valued function $g(\boldsymbol{x})$ of $\boldsymbol{x}$ is given by

$$E[g(\boldsymbol{x})] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\boldsymbol{t}) f_{\boldsymbol{x}}(\boldsymbol{t})\ dt_1 \cdots dt_m.$$

The random variables $x_1, \ldots, x_m$ are said to be independent, a concept we have already referred to, if and only if the joint probability function or density function factors into the product of the marginal probability or density functions; that is, in the continuous case, $x_1, \ldots, x_m$ are independent if and only if

$$f_{\boldsymbol{x}}(\boldsymbol{t}) = f_{x_1}(t_1) \cdots f_{x_m}(t_m),$$

for all $\boldsymbol{t}$.

The mean vector of $\boldsymbol{x}$, denoted $\boldsymbol{\mu}$, is the vector of expected values of the x_i's; that is,

$$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)' = E(\boldsymbol{x}) = [E(x_1), \ldots, E(x_m)]'.$$

A measure of the linear relationship between x_i and x_j is given by the covariance of x_i and x_j, which is denoted $\operatorname{cov}(x_i, x_j)$ or σ_{ij} and is defined by

$$\sigma_{ij} = \operatorname{cov}(x_i, x_j) = E[(x_i - \mu_i)(x_j - \mu_j)] = E(x_i x_j) - \mu_i \mu_j. \tag{1.7}$$

When $i = j$, this covariance reduces to the variance of x_i; that is, $\sigma_{ii} = \sigma_i^2 = \operatorname{var}(x_i)$. When $i \neq j$ and x_i and x_j are independent, then $\operatorname{cov}(x_i, x_j) = 0$ because in this case $E(x_i x_j) = \mu_i \mu_j$. If α_1, α_2, β_1 and β_2 are constants, then

$$\operatorname{cov}(\alpha_1 + \beta_1 x_i, \alpha_2 + \beta_2 x_j) = \beta_1 \beta_2 \operatorname{cov}(x_i, x_j).$$

The matrix Ω, which has σ_{ij} as its (i, j)th element, is called the variance–covariance matrix, or simply the covariance matrix, of $\boldsymbol{x}$. This matrix will be also denoted sometimes by var($\boldsymbol{x}$) or cov($\boldsymbol{x}, \boldsymbol{x}$). Clearly, $\sigma_{ij} = \sigma_{ji}$ so that Ω is a symmetric matrix. Using (1.7), we obtain the matrix formulation for Ω,

$$\Omega = \operatorname{var}(\boldsymbol{x}) = E[(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})'] = E(\boldsymbol{x}\boldsymbol{x}') - \boldsymbol{\mu}\boldsymbol{\mu}'.$$

If $\boldsymbol{\alpha}$ is an $m \times 1$ vector of constants and we define the random variable $y = \boldsymbol{\alpha}'\boldsymbol{x}$, then

$$\begin{aligned} E(y) = E(\boldsymbol{\alpha}'\boldsymbol{x}) = E\left(\sum_{i=1}^m \alpha_i x_i\right) &= \sum_{i=1}^m \alpha_i E(x_i) \\ &= \sum_{i=1}^m \alpha_i \mu_i = \boldsymbol{\alpha}'\boldsymbol{\mu}. \end{aligned}$$

If, in addition, $\boldsymbol{\beta}$ is another $m \times 1$ vector of constants and $w = \boldsymbol{\beta}'\boldsymbol{x}$, then

$$\begin{aligned} \operatorname{cov}(y, w) = \operatorname{cov}(\boldsymbol{\alpha}'\boldsymbol{x}, \boldsymbol{\beta}'\boldsymbol{x}) &= \operatorname{cov}\left(\sum_{i=1}^{m} \alpha_i x_i, \sum_{j=1}^{m} \beta_j x_j\right) \\ &= \sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \beta_j \operatorname{cov}(x_i, x_j) = \sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \beta_j \sigma_{ij} = \boldsymbol{\alpha}'\Omega\boldsymbol{\beta}. \end{aligned}$$

In particular, $\operatorname{var}(y) = \operatorname{cov}(y, y) = \boldsymbol{\alpha}'\Omega\boldsymbol{\alpha}$. Because this holds for any choice of $\boldsymbol{\alpha}$ and because the variance is always nonnegative, Ω must be a nonnegative definite matrix. More generally, if A is a $p \times m$ matrix of constants and $\boldsymbol{y} = A\boldsymbol{x}$, then

$$E(\boldsymbol{y}) = E(A\boldsymbol{x}) = AE(\boldsymbol{x}) = A\boldsymbol{\mu}, \tag{1.8}$$

$$\begin{aligned} \operatorname{var}(\boldsymbol{y}) &= E[\{\boldsymbol{y} - E(\boldsymbol{y})\}\{\boldsymbol{y} - E(\boldsymbol{y})\}'] = E[(A\boldsymbol{x} - A\boldsymbol{\mu})(A\boldsymbol{x} - A\boldsymbol{\mu})'] \\ &= E[A(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})'A'] = A\{E[(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})']\}A' \\ &= A\Omega A'. \end{aligned} \tag{1.9}$$

Thus, the mean vector and covariance matrix of the transformed vector, $A\boldsymbol{x}$, is $A\boldsymbol{\mu}$ and $A\Omega A'$. If $\boldsymbol{v}$ and $\boldsymbol{w}$ are random vectors, then the matrix of covariances between components of $\boldsymbol{v}$ and components of $\boldsymbol{w}$ is given by

$$\operatorname{cov}(\boldsymbol{v}, \boldsymbol{w}) = E(\boldsymbol{v}\boldsymbol{w}') - E(\boldsymbol{v})E(\boldsymbol{w})'.$$

In particular, if $\boldsymbol{v} = A\boldsymbol{x}$ and $\boldsymbol{w} = B\boldsymbol{x}$, then

$$\operatorname{cov}(\boldsymbol{v}, \boldsymbol{w}) = A \operatorname{cov}(\boldsymbol{x}, \boldsymbol{x})B' = A \operatorname{var}(\boldsymbol{x})B' = A\Omega B'.$$

A measure of the linear relationship between x_i and x_j that is unaffected by the measurement scales of x_i and x_j is called the correlation. We denote this by ρ_{ij} and it is defined as

$$\rho_{ij} = \frac{\operatorname{cov}(x_i, x_j)}{\sqrt{\operatorname{var}(x_i)\operatorname{var}(x_j)}} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}.$$

When $i = j$, $\rho_{ij} = 1$. The correlation matrix P, which has ρ_{ij} as its (i, j)th element, can be expressed in terms of the corresponding covariance matrix Ω and the diagonal matrix $D_\Omega^{-1/2} = \operatorname{diag}(\sigma_{11}^{-1/2}, \ldots, \sigma_{mm}^{-1/2})$; specifically,

$$P = D_\Omega^{-1/2}\Omega D_\Omega^{-1/2}. \tag{1.10}$$

For any $m \times 1$ vector $\boldsymbol{\alpha}$, we have

$$\boldsymbol{\alpha}'P\boldsymbol{\alpha} = \boldsymbol{\alpha}'D_\Omega^{-1/2}\Omega D_\Omega^{-1/2}\boldsymbol{\alpha} = \boldsymbol{\beta}'\Omega\boldsymbol{\beta},$$

where $\boldsymbol{\beta} = D_\Omega^{-1/2}\boldsymbol{\alpha}$, and so P must be nonnegative definite because Ω is. In particular, if $\boldsymbol{e}_i$ is the ith column of the $m \times m$ identity matrix, then

$$\begin{aligned}(\boldsymbol{e}_i + \boldsymbol{e}_j)'P(\boldsymbol{e}_i + \boldsymbol{e}_j) &= (P)_{ii} + (P)_{ij} + (P)_{ji} + (P)_{jj}\\ &= 2(1 + \rho_{ij}) \geq 0\end{aligned}$$

and

$$\begin{aligned}(\boldsymbol{e}_i - \boldsymbol{e}_j)'P(\boldsymbol{e}_i - \boldsymbol{e}_j) &= (P)_{ii} - (P)_{ij} - (P)_{ji} + (P)_{jj}\\ &= 2(1 - \rho_{ij}) \geq 0,\end{aligned}$$

from which we obtain the inequality, $-1 \leq \rho_{ij} \leq 1$.

Typically, means, variances, and covariances are unknown and so they must be estimated from a sample. Suppose $x_1, \ldots, x_n$ represents a random sample of a random variable x that has some distribution with mean μ and variance σ^2. These quantities can be estimated by the sample mean and the sample variance given by

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i,$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right).$$

In the multivariate setting, we have analogous estimators for $\boldsymbol{\mu}$ and Ω; if $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ is a random sample of an $m \times 1$ random vector $\boldsymbol{x}$ having mean vector $\boldsymbol{\mu}$ and covariance matrix Ω, then the sample mean vector and sample covariance matrix are given by

$$\overline{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i,$$

$$S = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})' = \frac{1}{n-1}\left(\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i' - n\overline{\boldsymbol{x}}\,\overline{\boldsymbol{x}}'\right).$$

The sample covariance matrix can be then used in (1.10) to obtain an estimator of the correlation matrix, P; that is, if we define the diagonal matrix $D_S^{-1/2} = \mathrm{diag}(s_{11}^{-1/2}, \ldots, s_{mm}^{-1/2})$, then the correlation matrix can be estimated by the sample correlation matrix defined as

$$R = D_S^{-1/2} S D_S^{-1/2}.$$

One particular joint distribution that we will consider is the multivariate normal distribution. This distribution can be defined in terms of independent standard normal random variables. Let $z_1, \ldots, z_m$ be independently distributed as $N(0, 1)$, and put $\boldsymbol{z} = (z_1, \ldots, z_m)'$. The density function of $\boldsymbol{z}$ is then given by

$$f(\boldsymbol{z}) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} z_i^2\right) = \frac{1}{(2\pi)^{m/2}} \exp\left(-\frac{1}{2}\boldsymbol{z}'\boldsymbol{z}\right).$$

Because $E(\boldsymbol{z}) = \mathbf{0}$ and $\mathrm{var}(\boldsymbol{z}) = I_m$, this particular m-dimensional multivariate normal distribution, known as the standard multivariate normal distribution, is denoted as $N_m(\mathbf{0}, I_m)$. If $\boldsymbol{\mu}$ is an $m \times 1$ vector of constants and T is an $m \times m$ nonsingular matrix, then $\boldsymbol{x} = \boldsymbol{\mu} + T\boldsymbol{z}$ is said to have the m-dimensional multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\Omega = TT'$. This is indicated by $\boldsymbol{x} \sim N_m(\boldsymbol{\mu}, \Omega)$. For instance, if $m = 2$, the vector $\boldsymbol{x} = (x_1, x_2)'$ has a bivariate normal distribution and its density, induced by the transformation $\boldsymbol{x} = \boldsymbol{\mu} + T\boldsymbol{z}$, can be shown to be

$$\begin{aligned} f(\boldsymbol{x}) = \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}(1-\rho^2)}} \exp\Bigg(-\frac{1}{2(1-\rho^2)} \Bigg\{ \frac{(x_1-\mu_1)^2}{\sigma_{11}} \\ -2\rho\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right) + \frac{(x_2-\mu_2)^2}{\sigma_{22}} \Bigg\}\Bigg), \end{aligned} \qquad (1.11)$$

for all $\boldsymbol{x} \in R^2$, where $\rho = \rho_{12}$ is the correlation coefficient. When $\rho = 0$, this density factors into the product of the marginal densities, so x_1 and x_2 are independent if and only if $\rho = 0$. The cumbersome-looking density function given in (1.11) can be more conveniently expressed by using matrix notation. It is straightforward to verify that this density is identical to

$$f(\boldsymbol{x}) = \frac{1}{2\pi|\Omega|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Omega^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}. \qquad (1.12)$$

The density function of an m-variate normal random vector is very similar to the function given in (1.12). If $\boldsymbol{x} \sim N_m(\boldsymbol{\mu}, \Omega)$, then its density is

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{m/2}|\Omega|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Omega^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}, \qquad (1.13)$$

for all $\boldsymbol{x} \in R^m$.

If Ω is positive semidefinite, then $\boldsymbol{x} \sim N_m(\boldsymbol{\mu}, \Omega)$ is said to have a singular normal distribution. In this case, Ω^{-1} does not exist and so the multivariate normal density cannot be written in the form given in (1.13). However, the random vector $\boldsymbol{x}$ can still be expressed in terms of independent standard normal random variables. Suppose that $\mathrm{rank}(\Omega) = r$ and U is an $m \times r$ matrix satisfying $UU' = \Omega$. Then $\boldsymbol{x} \sim N_m(\boldsymbol{\mu}, \Omega)$ if $\boldsymbol{x}$ is distributed the same as $\boldsymbol{\mu} + U\boldsymbol{z}$, where now $\boldsymbol{z} \sim N_r(\mathbf{0}, I_r)$.

An important property of the multivariate normal distribution is that a linear transformation of a multivariate normal vector yields a multivariate normal vector; that is, if $\boldsymbol{x} \sim N_m(\boldsymbol{\mu}, \Omega)$ and A is a $p \times m$ matrix of constants, then $\boldsymbol{y} = A\boldsymbol{x}$ has a p-variate normal distribution. In particular, from (1.8) and (1.9), we know that $\boldsymbol{y} \sim N_p(A\boldsymbol{\mu}, A\Omega A')$.

We next consider spherical and elliptical distributions that are extensions of multivariate normal distributions. In particular, a spherical distribution is an extension of the standard multivariate normal distribution $N_m(\mathbf{0}, I_m)$, whereas an elliptical distribution is an extension of the multivariate normal distribution $N_m(\boldsymbol{\mu}, \Omega)$. An $m \times 1$ random vector $\boldsymbol{x}$ has a spherical distribution if $\boldsymbol{x}$ and $P\boldsymbol{x}$ have the same distribution for all $m \times m$ orthogonal matrices P. If $\boldsymbol{x}$ has a spherical distribution with a density function, then this density function depends on $\boldsymbol{x}$ only through the value of $\boldsymbol{x}'\boldsymbol{x}$; that is, the density function of $\boldsymbol{x}$ can be written as $g(\boldsymbol{x}'\boldsymbol{x})$ for some function g. The term *spherical* distribution then arises from the fact that the density function is the same for all points $\boldsymbol{x}$ that lie on the sphere $\boldsymbol{x}'\boldsymbol{x} = c$, where c is a nonnegative constant. Clearly $\boldsymbol{z} \sim N_m(\mathbf{0}, I_m)$ has a spherical distribution because for any $m \times m$ orthogonal matrix P, $P\boldsymbol{z} \sim N_m(\mathbf{0}, I_m)$. An example of a nonnormal spherical distribution is the uniform distribution; that is, if $\boldsymbol{u}$ is a randomly selected point on the surface of the unit sphere in R^m, then $\boldsymbol{u}$ has a spherical distribution. In fact, if the $m \times 1$ random vector $\boldsymbol{x}$ has a spherical distribution, then it can be expressed as

$$\boldsymbol{x} = w\boldsymbol{u}, \tag{1.14}$$

where $\boldsymbol{u}$ is uniformly distributed on the m-dimensional unit sphere, w is a nonnegative random variable, and $\boldsymbol{u}$ and w are independently distributed. It is easy to verify that when $\boldsymbol{z}$ has the distribution $N_m(\mathbf{0}, I_m)$, then (1.14) takes the form

$$\boldsymbol{z} = v\boldsymbol{u},$$

where $v^2 \sim \chi^2_m$. Thus, if the $m \times 1$ random vector $\boldsymbol{x}$ has a spherical distribution, then it can also be expressed as

$$\boldsymbol{x} = w\boldsymbol{u} = wv^{-1}\boldsymbol{z} = s\boldsymbol{z},$$

where again $\boldsymbol{z}$ has the distribution $N_m(\mathbf{0}, I_m)$, $s = wv^{-1}$ is a nonnegative random variable, and $\boldsymbol{z}$ and s are independently distributed. The contaminated normal distributions and the multivariate t distributions are other examples of spherical distributions. A random vector $\boldsymbol{x}$ having a contaminated normal distribution can be expressed as $\boldsymbol{x} = s\boldsymbol{z}$, where $\boldsymbol{z} \sim N_m(\mathbf{0}, I_m)$ independently of s, which takes on the values σ and 1 with probabilities p and $1 - p$, respectively, and $\sigma \neq 1$ is a positive constant. If $\boldsymbol{z} \sim N_m(\mathbf{0}, I_m)$ independently of $v^2 \sim \chi^2_n$, then the random vector $\boldsymbol{x} = n^{1/2}\boldsymbol{z}/v$ has a multivariate t distribution with n degrees of freedom.

We generalize from spherical distributions to elliptical distributions in the same way that $N_m(\mathbf{0}, I_m)$ was generalized to $N_m(\boldsymbol{\mu}, \Omega)$. An $m \times 1$ random vector $\boldsymbol{y}$ has an elliptical distribution with parameters $\boldsymbol{\mu}$ and Ω if it can be expressed as

$$\boldsymbol{y} = \boldsymbol{\mu} + T\boldsymbol{x},$$

where T is $m \times r$, $TT' = \Omega$, $\text{rank}(\Omega) = r$, and the $r \times 1$ random vector $\boldsymbol{x}$ has a spherical distribution. Using (1.14), we then have

$$\boldsymbol{y} = \boldsymbol{\mu} + wT\boldsymbol{u},$$

where the random variable $w \geq 0$ is independent of $\boldsymbol{u}$, which is uniformly distributed on the r-dimensional unit sphere. If Ω is nonsingular and $\boldsymbol{y}$ has a density, then it depends on $\boldsymbol{y}$ only through the value of $(\boldsymbol{y}-\boldsymbol{\mu})'\Omega^{-1}(\boldsymbol{y}-\boldsymbol{\mu})$; that is, the density is the same for all points $\boldsymbol{y}$ that lie on the ellipsoid $(\boldsymbol{y}-\boldsymbol{\mu})'\Omega^{-1}(\boldsymbol{y}-\boldsymbol{\mu})=c$, where c is a nonnegative constant. A more detailed discussion about spherical and elliptical distributions can be found in Fang et al. (1990).

One of the most widely used procedures in statistics is regression analysis. We will briefly describe this analysis here and later use regression analysis to illustrate some of the matrix methods developed in this text. Some good references on regression are Kutner et al. (2005), Rencher and Schaalje (2008), and Sen and Srivastava (1990). In the typical regression problem, one wishes to study the relationship between some response variable, say y, and k explanatory variables $x_1, \dots, x_k$. For instance, y might be the yield of some product of a manufacturing process, whereas the explanatory variables are conditions affecting the production process, such as temperature, humidity, pressure, and so on. A model relating the x_j's to y is given by

$$y=\beta_0+\beta_1x_1+\cdots+\beta_kx_k+\epsilon, \tag{1.15}$$

where $\beta_0, \dots, \beta_k$ are unknown parameters and ϵ is a random error, that is, a random variable, with $E(\epsilon)=0$. In what is known as ordinary least squares regression, we also have the errors as independent random variables with common variance σ^2; that is, if ϵ_i and ϵ_j are random errors associated with the responses y_i and y_j, then $\mathrm{var}(\epsilon_i)=\mathrm{var}(\epsilon_j)=\sigma^2$ and $\mathrm{cov}(\epsilon_i,\epsilon_j)=0$. The model given in (1.15) is an example of a linear model because it is a linear function of the parameters. It need not be linear in the x_j's so that, for instance, we might have $x_2=x_1^2$. Because the parameters are unknown, they must be estimated and this will be possible if we have some observed values of y and the corresponding x_j's. Thus, for the ith observation, suppose that the explanatory variables are set to the values $x_{i1}, \dots, x_{ik}$ yielding the response y_i, and this is done for $i=1, \dots, N$, where $N>k+1$. If model (1.15) holds, then we should have, approximately,

$$y_i=\beta_0+\beta_1x_{i1}+\cdots+\beta_kx_{ik}$$

for each i. This can be written as the matrix equation

$$\boldsymbol{y}=X\boldsymbol{\beta}$$

if we define

$$\boldsymbol{y}=\begin{bmatrix} y_1\\ y_2\\ \vdots\\ y_N\end{bmatrix},\quad \boldsymbol{\beta}=\begin{bmatrix}\beta_0\\ \beta_1\\ \vdots\\ \beta_k\end{bmatrix},\quad X=\begin{bmatrix}1 & x_{11} & \cdots & x_{1k}\\ 1 & x_{21} & \cdots & x_{2k}\\ \vdots & \vdots & & \vdots\\ 1 & x_{N1} & \cdots & x_{Nk}\end{bmatrix}.$$

One method of estimating the β_j's, which we will discuss from time to time in this text, is called the method of least squares. If $\hat{\boldsymbol{\beta}}=(\hat{\beta}_0, \dots, \hat{\beta}_k)'$ is an estimate of the parameter vector $\boldsymbol{\beta}$, then $\hat{\boldsymbol{y}} = X\hat{\boldsymbol{\beta}}$ is the vector of fitted values, whereas $\boldsymbol{y} - \hat{\boldsymbol{y}}$ gives the vector of errors or deviations of the actual responses from the corresponding fitted values, and

$$f(\hat{\boldsymbol{\beta}}) = (\boldsymbol{y} - X\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - X\hat{\boldsymbol{\beta}})$$

gives the sum of squares of these errors. The method of least squares selects as $\hat{\boldsymbol{\beta}}$ any vector that minimizes the function $f(\hat{\boldsymbol{\beta}})$. We will see later that any such vector satisfies the system of linear equations, sometimes referred to as the normal equations,

$$X'X\hat{\boldsymbol{\beta}} = X'\boldsymbol{y}.$$

If X has full column rank, that is, $\text{rank}(X) = k + 1$, then $(X'X)^{-1}$ exists and so the least squares estimator of $\boldsymbol{\beta}$ is unique and is given by

$$\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\boldsymbol{y}.$$

PROBLEMS

1.1 Show that the scalar properties $ab = 0$ implies $a = 0$ or $b = 0$, and $ab = ac$ for $a \neq 0$ implies that $b = c$ do not extend to matrices by finding

  (a) 2×2 nonnull matrices A and B for which $AB = (0)$,

  (b) 2×2 matrices A, B, and C, with A being nonnull, such that $AB = AC$, yet $B \neq C$.

1.2 Let A be an $m \times m$ idempotent matrix. Show that

  (a) $I_m - A$ is idempotent,

  (b) BAB^{-1} is idempotent, where B is any $m \times m$ nonsingular matrix.

1.3 Let A and B be $m \times m$ symmetric matrices. Show that AB is symmetric if and only if $AB = BA$.

1.4 Prove Theorem 1.3(e); that is, if A is an $m \times n$ matrix, show that $\text{tr}(A'A) = 0$ if and only if $A = (0)$.

1.5 Show that

  (a) if $\boldsymbol{x}$ and $\boldsymbol{y}$ are $m \times 1$ vectors, $\text{tr}(\boldsymbol{x}\boldsymbol{y}') = \boldsymbol{x}'\boldsymbol{y}$,

  (b) if A and B are $m \times m$ matrices and B is nonsingular, $\text{tr}(BAB^{-1}) = \text{tr}(A)$.

1.6 Suppose A is $m \times n$ and B is $n \times m$. Show that $\text{tr}(AB) = \text{tr}(A'B')$.

1.7 Suppose that A, B, and C are $m \times m$ matrices. Show that if they are symmetric matrices, then $\text{tr}(ABC) = \text{tr}(ACB)$.

1.8 Prove Theorem 1.4.

1.9 Show that any square matrix can be written as the sum of a symmetric matrix and a skew-symmetric matrix.

1.10 Let A and B be $m \times m$ symmetric matrices. Show that $AB - BA$ is a skew-symmetric matrix.

1.11 Suppose that A is an $m \times m$ skew-symmetric matrix. Show that $-A^2$ is a nonnegative definite matrix.

1.12 Define the $m \times m$ matrices A, B, and C as

$$A = \begin{bmatrix} b_{11} + c_{11} & b_{12} + c_{12} & \cdots & b_{1m} + c_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{bmatrix},$$

$$B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{bmatrix},$$

$$C = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{bmatrix}.$$

Prove that $|A| = |B| + |C|$.

1.13 Verify the results of Theorem 1.8.

1.14 Suppose that A and B are $m \times m$ nonnull matrices satisfying $AB = (0)$. Show that both A and B must be singular matrices.

1.15 Consider the 4×4 matrix

$$A = \begin{bmatrix} 1 & 2 & 1 & 1 \\ 0 & 1 & 2 & 0 \\ 1 & 2 & 2 & 1 \\ 0 & -1 & 1 & 2 \end{bmatrix}.$$

Find the determinant of A by using the cofactor expansion formula on the first column of A.

1.16 Using the matrix A from the previous problem, verify (1.3) when $i = 1$ and $k = 2$.

1.17 Prove Theorem 1.6.

1.18 Let λ be a variable, and consider the determinant of $A - \lambda I_m$, where A is an $m \times m$ matrix, as a function of λ. What type of function of λ is this?

1.19 Find the adjoint matrix of the matrix A given in Problem 1.15. Use this to obtain the inverse of A.

1.20 Using elementary transformations, determine matrices B and C so that $BAC = I_4$ for the matrix A given in Problem 1.15. Use B and C to compute the inverse of A; that is, take the inverse of both sides of the equation $BAC = I_4$ and then solve for A^{-1}.

1.21 Compute the inverse of

(a) $I_m + \mathbf{1}_m\mathbf{1}'_m$,

(b) $I_m + \boldsymbol{e}_1\mathbf{1}'_m$.

1.22 Show that

(a) the determinant of a triangular matrix is the product of its diagonal elements,

(b) the inverse of a lower triangular matrix is a lower triangular matrix.

1.23 Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be $m \times 1$ vectors and D be an $m \times m$ diagonal matrix. Use Corollary 1.9.2 to find an expression for the inverse of $D + \alpha\boldsymbol{a}\boldsymbol{b}'$, where α is a scalar.

1.24 Let $A_\#$ be the adjoint matrix of an $m \times m$ matrix A. Show that

(a) $|A_\#| = |A|^{m-1}$,

(b) $(\alpha A)_\# = \alpha^{m-1}A_\#$, where α is a scalar.

1.25 Consider the $m \times m$ partitioned matrix

$$A = \begin{bmatrix} A_{11} & (0) \\ A_{21} & A_{22} \end{bmatrix},$$

where the $m_1 \times m_1$ matrix A_{11} and the $m_2 \times m_2$ matrix A_{22} are nonsingular. Obtain an expression for A^{-1} in terms of A_{11}, A_{22}, and A_{21}.

1.26 Let

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A'_{12} & A_{22} \end{bmatrix},$$

where A_{11} is $m_1 \times m_1$, A_{22} is $m_2 \times m_2$, and A_{12} is $m_1 \times m_2$. Show that if A is positive definite, then A_{11} and A_{22} are also positive definite.

1.27 Find the rank of the 4×4 matrix

$$A = \begin{bmatrix} 2 & 0 & 1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & 2 & 0 \\ 2 & 0 & 0 & -2 \end{bmatrix}.$$

1.28 Use elementary transformations to transform the matrix A given in Problem 1.27 to a matrix H having the form given in Theorem 1.11. Consequently, determine matrices B and C so that $BAC = H$.

1.29 Prove parts (b) and (c) of Theorem 1.12.

1.30 List all permutation matrices of order 3.

1.31 Consider the 3×3 matrix

$$P = \frac{1}{\sqrt{6}}\begin{bmatrix} \sqrt{2} & \sqrt{2} & \sqrt{2} \\ \sqrt{3} & -\sqrt{3} & 0 \\ p_{31} & p_{32} & p_{33} \end{bmatrix}.$$

Find values for p_{31}, p_{32}, and p_{33} so that P is an orthogonal matrix. Is your solution unique?

1.32 Give the conditions on the $m \times 1$ vector $\boldsymbol{x}$ so that the matrix $H = I_m - 2\boldsymbol{x}\boldsymbol{x}'$ is orthogonal.

1.33 Suppose the $m \times m$ orthogonal matrix P is partitioned as $P = [P_1 \;\; P_2]$, where P_1 is $m \times m_1$, P_2 is $m \times m_2$, and $m_1 + m_2 = m$. Show that $P_1'P_1 = I_{m_1}$, $P_2'P_2 = I_{m_2}$, and $P_1P_1' + P_2P_2' = I_m$.

1.34 Let A, B, and C be $m \times n$, $n \times p$, and $n \times n$ matrices, respectively, while $\boldsymbol{x}$ is an $n \times 1$ vector. Show that

(a) $A\boldsymbol{x} = \mathbf{0}$ for all choices of $\boldsymbol{x}$ if and only if $A = (0)$,

(b) $A\boldsymbol{x} = \mathbf{0}$ if and only if $A'A\boldsymbol{x} = \mathbf{0}$,

(c) $A = (0)$ if $A'A = (0)$,

(d) $AB = (0)$ if and only if $A'AB = (0)$,

(e) $\boldsymbol{x}'C\boldsymbol{x} = 0$ for all $\boldsymbol{x}$ if and only if $C' = -C$.

1.35 For each of the following, find the 3×3 symmetric matrix A so that the given identity holds:

(a) $\boldsymbol{x}'A\boldsymbol{x} = x_1^2 + 2x_2^2 - x_3^2 + 4x_1x_2 - 6x_1x_3 + 8x_2x_3$.

(b) $\boldsymbol{x}'A\boldsymbol{x} = 3x_1^2 + 5x_2^2 + 2x_3^2 + 2x_1x_2 + 2x_1x_3 + 4x_2x_3$.

(c) $\boldsymbol{x}'A\boldsymbol{x} = 2x_1x_2 + 2x_1x_3 + 2x_2x_3$.

1.36 Let $\boldsymbol{x}$ be a 4×1 vector. Find symmetric matrices A_1 and A_2 such that

$$\boldsymbol{x}'A_1\boldsymbol{x} = (x_1 + x_2 - 2x_3)^2 + (x_3 - x_4)^2,$$
$$\boldsymbol{x}'A_2\boldsymbol{x} = (x_1 - x_2 - x_3)^2 + (x_1 + x_2 - x_4)^2.$$

1.37 Let A be an $m \times m$ matrix, and suppose that a real $n \times m$ matrix T exists such that $T'T = A$. Show that A must be nonnegative definite.

1.38 Prove that a nonnegative definite matrix must have nonnegative diagonal elements; that is, show that if a symmetric matrix has any negative diagonal elements, then it is not nonnegative definite. Show that the converse is not true; that is, find a symmetric matrix that has nonnegative diagonal elements but is not nonnegative definite.

1.39 Let A be an $m \times m$ nonnegative definite matrix, while B is an $n \times m$ matrix. Show that BAB' is a nonnegative definite matrix.

1.40 Define A as

$$A = \begin{bmatrix} 5 & 1 \\ 1 & 4 \end{bmatrix}.$$

Find an upper triangular square root matrix of A; that is, find a 2×2 upper triangular matrix B satisfying $BB' = A$.

1.41 Use the standard normal moment generating function, $m_z(t) = e^{t^2/2}$, to show that the first six moments of the standard normal distribution are 0, 1, 0, 3, 0, and 15.

1.42 Use properties of expectation to show that for random variables x_1 and x_2, and scalars α_1, α_2, β_1, and β_2,

$$\text{cov}(\alpha_1 + \beta_1 x_1, \alpha_2 + \beta_2 x_2) = \beta_1 \beta_2 \,\text{cov}(x_1, x_2).$$

1.43 Let S be the sample covariance matrix computed from the sample $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, where each $\boldsymbol{x}_i$ is $m \times 1$. Define the $m \times n$ matrix X to be $X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$. Find a matrix expression for the symmetric matrix A satisfying $S = (n-1)^{-1} X A X'$.

1.44 Show that if $\boldsymbol{x} \sim N_m(\boldsymbol{\mu}, \Omega)$, where Ω is positive definite, then $(\boldsymbol{x} - \boldsymbol{\mu})'\Omega^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \sim \chi^2_m$.

1.45 Suppose $\boldsymbol{x} \sim N_3(\boldsymbol{\mu}, \Omega)$, where

$$\boldsymbol{\mu} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \Omega = \begin{bmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 3 \end{bmatrix},$$

and let the 3×3 matrix A and 2×3 matrix B be given by

$$A = \begin{bmatrix} 2 & 2 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 0 \end{bmatrix}.$$

(a) Find the correlation matrix of $\boldsymbol{x}$.

(b) Determine the distribution of $u = \mathbf{1}_3'\boldsymbol{x}$.

(c) Determine the distribution of $\boldsymbol{v} = A\boldsymbol{x}$.

(d) Determine the distribution of

$$\boldsymbol{w} = \begin{bmatrix} A\boldsymbol{x} \\ B\boldsymbol{x} \end{bmatrix}.$$

(e) Which, if any, of the distributions obtained in (b), (c), and (d) are singular distributions?

1.46 Suppose $\boldsymbol{x}$ is an $m \times 1$ random vector with mean vector $\boldsymbol{\mu}$ and covariance matrix Ω. If A is an $n \times m$ matrix of constants and $\boldsymbol{c}$ is an $m \times 1$ vector of constants, give expressions for

(a) $E[A(\boldsymbol{x} + \boldsymbol{c})]$,

(b) $\text{var}[A(\boldsymbol{x} + \boldsymbol{c})]$.

1.47 Let $x_1, \ldots, x_m$ be a random sample from a normal population with mean μ and variance σ^2, so that $\boldsymbol{x} = (x_1, \ldots, x_m)' \sim N_m(\mu \mathbf{1}_m, \sigma^2 I_m)$.

(a) What is the distribution of $\boldsymbol{u} = H\boldsymbol{x}$, where H is the Helmert matrix?

(b) Show that $\sum_{i=1}^m (x_i - \overline{x})^2 = \sum_{i=2}^m u_i^2$, and use this to establish that s^2 is distributed independently of $\overline{x}$.

1.48 Use the stochastic representation given in Section 1.13 for a random vector $\boldsymbol{x}$ having a contaminated normal distribution to show that $E(\boldsymbol{x}) = \mathbf{0}$ and $\text{var}(\boldsymbol{x}) = \{1 + p(\sigma^2 - 1)\}I_m$.

1.49 Show that if $\boldsymbol{x}$ has the multivariate t distribution with n degrees of freedom as given in Section 1.13, then $E(\boldsymbol{x}) = \mathbf{0}$ and $\text{var}(\boldsymbol{x}) = \frac{n}{n-2}I_m$ if $n > 2$.