

Example 2.6.2. Let X_1, X_2 , and X_3 be three mutually independent random variables and let each have the pdf

$$f(x) = \begin{cases} 2x & 0 < x < 1 \\ 0 & \text{elsewhere.} \end{cases} \quad (2.6.8)$$

The joint pdf of X_1, X_2, X_3 is $f(x_1)f(x_2)f(x_3) = 8x_1x_2x_3$, $0 < x_i < 1$, $i = 1, 2, 3$, zero elsewhere. Then, for illustration, the expected value of $5X_1X_2^3 + 3X_2X_3^4$ is

$$\int_0^1 \int_0^1 \int_0^1 (5x_1x_2^3 + 3x_2x_3^4) 8x_1x_2x_3 dx_1 dx_2 dx_3 = 2.$$

Let Y be the maximum of X_1, X_2 , and X_3 . Then, for instance, we have

$$\begin{aligned} P(Y \leq \tfrac{1}{2}) &= P(X_1 \leq \tfrac{1}{2}, X_2 \leq \tfrac{1}{2}, X_3 \leq \tfrac{1}{2}) \\ &= \int_0^{1/2} \int_0^{1/2} \int_0^{1/2} 8x_1x_2x_3 dx_1 dx_2 dx_3 \\ &= (\tfrac{1}{2})^6 = \tfrac{1}{64}. \end{aligned}$$

In a similar manner, we find that the cdf of Y is

$$G(y) = P(Y \leq y) = \begin{cases} 0 & y < 0 \\ y^6 & 0 \leq y < 1 \\ 1 & 1 \leq y. \end{cases}$$

Accordingly, the pdf of Y is

$$g(y) = \begin{cases} 6y^5 & 0 < y < 1 \\ 0 & \text{elsewhere.} \end{cases} \quad \blacksquare$$

Remark 2.6.1. If X_1, X_2 , and X_3 are mutually independent, they are **pairwise independent** (that is, X_i and X_j , $i \neq j$, where $i, j = 1, 2, 3$, are independent). However, the following example, attributed to S. Bernstein, shows that pairwise independence does not necessarily imply mutual independence. Let X_1, X_2 , and X_3 have the joint pmf

$$p(x_1, x_2, x_3) = \begin{cases} \frac{1}{4} & (x_1, x_2, x_3) \in \{(1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 1)\} \\ 0 & \text{elsewhere.} \end{cases}$$

The joint pmf of X_i and X_j , $i \neq j$, is

$$p_{ij}(x_i, x_j) = \begin{cases} \frac{1}{4} & (x_i, x_j) \in \{(0, 0), (1, 0), (0, 1), (1, 1)\} \\ 0 & \text{elsewhere,} \end{cases}$$

whereas the marginal pmf of X_i is

$$p_i(x_i) = \begin{cases} \frac{1}{2} & x_i = 0, 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Obviously, if $i \neq j$, we have

$$p_{ij}(x_i, x_j) \equiv p_i(x_i)p_j(x_j),$$

and thus X_i and X_j are independent. However,

$$p(x_1, x_2, x_3) \neq p_1(x_1)p_2(x_2)p_3(x_3).$$

Thus X_1, X_2 , and X_3 are not mutually independent.

Unless there is a possible misunderstanding between *mutual* and *pairwise* independence, we usually drop the modifier *mutual*. Accordingly, using this practice in Example 2.6.2, we say that X_1, X_2, X_3 are independent random variables, meaning that they are mutually independent. Occasionally, for emphasis, we use *mutually independent* so that the reader is reminded that this is different from *pairwise independence*.

In addition, if several random variables are mutually independent and have the same distribution, we say that they are **independent and identically distributed**, which we abbreviate as **iid**. So the random variables in Example 2.6.2 are iid with the common pdf given in expression (2.6.8). ■

The following is a useful corollary to Theorem 2.6.1 for iid random variables. Its proof is asked for in Exercise 2.6.7.

Corollary 2.6.1. *Suppose $X_1, X_2, \dots, X_n$ are iid random variables with the common mgf $M(t)$, for $-h < t < h$, where $h > 0$. Let $T = \sum_{i=1}^n X_i$. Then T has the mgf given by*

$$M_T(t) = [M(t)]^n, \quad -h < t < h. \quad (2.6.9)$$

2.6.1 *Multivariate Variance-Covariance Matrix

This section makes explicit use of matrix algebra and it is considered as an optional section.

In Section 2.5 we discussed the covariance between two random variables. In this section we want to extend this discussion to the n -variate case. Let $\mathbf{X} = (X_1, \dots, X_n)'$ be an n -dimensional random vector. Recall that we defined $E(\mathbf{X}) = (E(X_1), \dots, E(X_n))'$, that is, the expectation of a random vector is just the vector of the expectations of its components. Now suppose $\mathbf{W}$ is an $m \times n$ matrix of random variables, say, $\mathbf{W} = [W_{ij}]$ for the random variables W_{ij} , $1 \leq i \leq m$ and $1 \leq j \leq n$. Note that we can always string out the matrix into an $mn \times 1$ random vector. Hence, we define the expectation of a random matrix

$$E[\mathbf{W}] = [E(W_{ij})]. \quad (2.6.10)$$

As the following theorem shows, the linearity of the expectation operator easily follows from this definition:

Theorem 2.6.2. *Let $\mathbf{W}_1$ and $\mathbf{W}_2$ be $m \times n$ matrices of random variables, let $\mathbf{A}_1$ and $\mathbf{A}_2$ be $k \times m$ matrices of constants, and let $\mathbf{B}$ be an $n \times l$ matrix of constants.*

Then

$$E[\mathbf{A}_1 \mathbf{W}_1 + \mathbf{A}_2 \mathbf{W}_2] = \mathbf{A}_1 E[\mathbf{W}_1] + \mathbf{A}_2 E[\mathbf{W}_2] \quad (2.6.11)$$

$$E[\mathbf{A}_1 \mathbf{W}_1 \mathbf{B}] = \mathbf{A}_1 E[\mathbf{W}_1] \mathbf{B}. \quad (2.6.12)$$

Proof: Because of the linearity of the operator E on random variables, we have for the (i, j) th components of expression (2.6.11) that

$$E \left[\sum_{s=1}^m a_{1is} W_{1sj} + \sum_{s=1}^m a_{2is} W_{2sj} \right] = \sum_{s=1}^m a_{1is} E[W_{1sj}] + \sum_{s=1}^m a_{2is} E[W_{2sj}].$$

Hence by (2.6.10), expression (2.6.11) is true. The derivation of expression (2.6.12) follows in the same manner. ■

Let $\mathbf{X} = (X_1, \dots, X_n)'$ be an n -dimensional random vector, such that $\sigma_i^2 = \text{Var}(X_i) < \infty$. The **mean** of $\mathbf{X}$ is $\boldsymbol{\mu} = E[\mathbf{X}]$ and we define its **variance-covariance matrix** as

$$\text{Cov}(\mathbf{X}) = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'] = [\sigma_{ij}], \quad (2.6.13)$$

where σ_{ii} denotes σ_i^2 . As Exercise 2.6.8 shows, the i th diagonal entry of $\text{Cov}(\mathbf{X})$ is $\sigma_i^2 = \text{Var}(X_i)$ and the (i, j) th off diagonal entry is $\text{Cov}(X_i, X_j)$.

Example 2.6.3 (Example 2.5.6, Continued). In Example 2.5.6, we considered the joint pdf

$$f(x, y) = \begin{cases} e^{-y} & 0 < x < y < \infty \\ 0 & \text{elsewhere,} \end{cases}$$

and showed that the first two moments are

$$\begin{aligned} \mu_1 &= 1, & \mu_2 &= 2 \\ \sigma_1^2 &= 1, & \sigma_2^2 &= 2 \\ E[(X - \mu_1)(Y - \mu_2)] &= 1. \end{aligned} \quad (2.6.14)$$

Let $\mathbf{Z} = (X, Y)'$. Then using the present notation, we have

$$E[\mathbf{Z}] = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \text{and} \quad \text{Cov}(\mathbf{Z}) = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}. \quad \blacksquare$$

Two properties of $\text{Cov}(X_i, X_j)$ needed later are summarized in the following theorem:

Theorem 2.6.3. Let $\mathbf{X} = (X_1, \dots, X_n)'$ be an n -dimensional random vector, such that $\sigma_i^2 = \sigma_{ii} = \text{Var}(X_i) < \infty$. Let $\mathbf{A}$ be an $m \times n$ matrix of constants. Then

$$\text{Cov}(\mathbf{X}) = E[\mathbf{X}\mathbf{X}'] - \boldsymbol{\mu}\boldsymbol{\mu}' \quad (2.6.15)$$

$$\text{Cov}(\mathbf{A}\mathbf{X}) = \mathbf{A}\text{Cov}(\mathbf{X})\mathbf{A}'. \quad (2.6.16)$$

Proof: Use Theorem 2.6.2 to derive (2.6.15); i.e.,

$$\begin{aligned}\text{Cov}(\mathbf{X}) &= E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'] \\ &= E[\mathbf{X}\mathbf{X}' - \boldsymbol{\mu}\mathbf{X}' - \mathbf{X}\boldsymbol{\mu}' + \boldsymbol{\mu}\boldsymbol{\mu}'] \\ &= E[\mathbf{X}\mathbf{X}'] - \boldsymbol{\mu}E[\mathbf{X}'] - E[\mathbf{X}]\boldsymbol{\mu}' + \boldsymbol{\mu}\boldsymbol{\mu}',\end{aligned}$$

which is the desired result. The proof of (2.6.16) is left as an exercise. ■

All variance-covariance matrices are **positive semi-definite** matrices; that is, $\mathbf{a}'\text{Cov}(\mathbf{X})\mathbf{a} \geq 0$, for all vectors $\mathbf{a} \in R^n$. To see this let $\mathbf{X}$ be a random vector and let $\mathbf{a}$ be any $n \times 1$ vector of constants. Then $Y = \mathbf{a}'\mathbf{X}$ is a random variable and, hence, has nonnegative variance; i.e.,

$$0 \leq \text{Var}(Y) = \text{Var}(\mathbf{a}'\mathbf{X}) = \mathbf{a}'\text{Cov}(\mathbf{X})\mathbf{a}; \quad (2.6.17)$$

hence, $\text{Cov}(\mathbf{X})$ is positive semi-definite.

EXERCISES

2.6.1. Let X, Y, Z have joint pdf $f(x, y, z) = 2(x + y + z)/3$, $0 < x < 1$, $0 < y < 1$, $0 < z < 1$, zero elsewhere.

- (a) Find the marginal probability density functions of X, Y , and Z .
- (b) Compute $P(0 < X < \frac{1}{2}, 0 < Y < \frac{1}{2}, 0 < Z < \frac{1}{2})$ and $P(0 < X < \frac{1}{2}) = P(0 < Y < \frac{1}{2}) = P(0 < Z < \frac{1}{2})$.
- (c) Are X, Y , and Z independent?
- (d) Calculate $E(X^2YZ + 3XY^4Z^2)$.
- (e) Determine the cdf of X, Y , and Z .
- (f) Find the conditional distribution of X and Y , given $Z = z$, and evaluate $E(X + Y|z)$.
- (g) Determine the conditional distribution of X , given $Y = y$ and $Z = z$, and compute $E(X|y, z)$.

2.6.2. Let $f(x_1, x_2, x_3) = \exp[-(x_1 + x_2 + x_3)]$, $0 < x_1 < \infty$, $0 < x_2 < \infty$, $0 < x_3 < \infty$, zero elsewhere, be the joint pdf of X_1, X_2, X_3 .

- (a) Compute $P(X_1 < X_2 < X_3)$ and $P(X_1 = X_2 < X_3)$.
- (b) Determine the joint mgf of X_1, X_2 , and X_3 . Are these random variables independent?

2.6.3. Let X_1, X_2, X_3 , and X_4 be four independent random variables, each with pdf $f(x) = 3(1 - x)^2$, $0 < x < 1$, zero elsewhere. If Y is the minimum of these four variables, find the cdf and the pdf of Y .

Hint: $P(Y > y) = P(X_i > y, i = 1, \dots, 4)$.

2.6.4. A fair die is cast at random three independent times. Let the random variable X_i be equal to the number of spots that appear on the i th trial, $i = 1, 2, 3$. Let the random variable Y be equal to $\max(X_i)$. Find the cdf and the pmf of Y .

Hint: $P(Y \leq y) = P(X_i \leq y, i = 1, 2, 3)$.

2.6.5. Let $M(t_1, t_2, t_3)$ be the mgf of the random variables X_1, X_2 , and X_3 of Bernstein's example, described in the remark following Example 2.6.2. Show that

$$M(t_1, t_2, 0) = M(t_1, 0, 0)M(0, t_2, 0), \quad M(t_1, 0, t_3) = M(t_1, 0, 0)M(0, 0, t_3),$$

and

$$M(0, t_2, t_3) = M(0, t_2, 0)M(0, 0, t_3)$$

are true, but that

$$M(t_1, t_2, t_3) \neq M(t_1, 0, 0)M(0, t_2, 0)M(0, 0, t_3).$$

Thus X_1, X_2, X_3 are pairwise independent but not mutually independent.

2.6.6. Let X_1, X_2 , and X_3 be three random variables with means, variances, and correlation coefficients, denoted by $\mu_1, \mu_2, \mu_3; \sigma_1^2, \sigma_2^2, \sigma_3^2$; and $\rho_{12}, \rho_{13}, \rho_{23}$, respectively. For constants b_2 and b_3 , suppose $E(X_1 - \mu_1 | x_2, x_3) = b_2(x_2 - \mu_2) + b_3(x_3 - \mu_3)$. Determine b_2 and b_3 in terms of the variances and the correlation coefficients.

2.6.7. Prove Corollary 2.6.1.

2.6.8. Let $\mathbf{X} = (X_1, \dots, X_n)'$ be an n -dimensional random vector, with the variance-covariance matrix given in display (2.6.13). Show that the i th diagonal entry of $\text{Cov}(\mathbf{X})$ is $\sigma_i^2 = \text{Var}(X_i)$ and that the (i, j) th off diagonal entry is $\text{Cov}(X_i, X_j)$.

2.6.9. Let X_1, X_2, X_3 be iid with common pdf $f(x) = \exp(-x)$, $0 < x < \infty$, zero elsewhere. Evaluate:

(a) $P(X_1 < X_2 | X_1 < 2X_2)$.

(b) $P(X_1 < X_2 < X_3 | X_3 < 1)$.

2.7 Transformations for Several Random Variables

In Section 2.2 it was seen that the determination of the joint pdf of two functions of two random variables of the continuous type was essentially a corollary to a theorem in analysis having to do with the change of variables in a twofold integral. This theorem has a natural extension to n -fold integrals. This extension is as follows. Consider an integral of the form

$$\int \cdots \int_A f(x_1, x_2, \dots, x_n) dx_1 dx_2 \cdots dx_n$$

taken over a subset A of an n -dimensional space $\mathcal{S}$. Let

$$y_1 = u_1(x_1, x_2, \dots, x_n), \quad y_2 = u_2(x_1, x_2, \dots, x_n), \dots, y_n = u_n(x_1, x_2, \dots, x_n),$$

together with the inverse functions

$$x_1 = w_1(y_1, y_2, \dots, y_n), \quad x_2 = w_2(y_1, y_2, \dots, y_n), \dots, x_n = w_n(y_1, y_2, \dots, y_n)$$

define a one-to-one transformation that maps $\mathcal{S}$ onto $\mathcal{T}$ in the $y_1, y_2, \dots, y_n$ space and, hence, maps the subset A of $\mathcal{S}$ onto a subset B of $\mathcal{T}$. Let the first partial derivatives of the inverse functions be continuous and let the n by n determinant (called the Jacobian)

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$$

not be identically zero in $\mathcal{T}$. Then

$$\begin{aligned} & \int \cdots \int_A f(x_1, x_2, \dots, x_n) dx_1 dx_2 \cdots dx_n \\ &= \int \cdots \int_B f[w_1(y_1, \dots, y_n), w_2(y_1, \dots, y_n), \dots, w_n(y_1, \dots, y_n)] |J| dy_1 dy_2 \cdots dy_n. \end{aligned}$$

Whenever the conditions of this theorem are satisfied, we can determine the joint pdf of n functions of n random variables. Appropriate changes of notation in Section 2.2 (to indicate n -space as opposed to 2-space) are all that are needed to show that the joint pdf of the random variables $Y_1 = u_1(X_1, X_2, \dots, X_n)$, $\dots$, $Y_n = u_n(X_1, X_2, \dots, X_n)$, where the joint pdf of $X_1, \dots, X_n$ is $f(x_1, \dots, x_n)$, is given by

$$g(y_1, y_2, \dots, y_n) = f[w_1(y_1, \dots, y_n), \dots, w_n(y_1, \dots, y_n)] |J|,$$

where $(y_1, y_2, \dots, y_n) \in \mathcal{T}$, and is zero elsewhere.

Example 2.7.1. Let X_1, X_2, X_3 have the joint pdf

$$f(x_1, x_2, x_3) = \begin{cases} 48x_1x_2x_3 & 0 < x_1 < x_2 < x_3 < 1 \\ 0 & \text{elsewhere.} \end{cases} \quad (2.7.1)$$

If $Y_1 = X_1/X_2$, $Y_2 = X_2/X_3$, and $Y_3 = X_3$, then the inverse transformation is given by

$$x_1 = y_1y_2y_3, \quad x_2 = y_2y_3, \quad \text{and} \quad x_3 = y_3.$$

The Jacobian is given by

$$J = \begin{vmatrix} y_2y_3 & y_1y_3 & y_1y_2 \\ 0 & y_3 & y_2 \\ 0 & 0 & 1 \end{vmatrix} = y_2y_3^2.$$

Moreover, inequalities defining the support are equivalent to

$$0 < y_1y_2y_3, \quad y_1y_2y_3 < y_2y_3, \quad y_2y_3 < y_3, \quad \text{and} \quad y_3 < 1,$$

which reduces to the support $\mathcal{T}$ of Y_1, Y_2, Y_3 of

$$\mathcal{T} = \{(y_1, y_2, y_3) : 0 < y_i < 1, i = 1, 2, 3\}.$$

Hence the joint pdf of Y_1, Y_2, Y_3 is

$$\begin{aligned} g(y_1, y_2, y_3) &= 48(y_1 y_2 y_3)(y_2 y_3) y_3 |y_2 y_3^2| \\ &= \begin{cases} 48 y_1 y_2^3 y_3^5 & 0 < y_i < 1, i = 1, 2, 3 \\ 0 & \text{elsewhere.} \end{cases} \end{aligned} \quad (2.7.2)$$

The marginal pdfs are

$$\begin{aligned} g_1(y_1) &= 2y_1, 0 < y_1 < 1, \text{ zero elsewhere} \\ g_2(y_2) &= 4y_2^3, 0 < y_2 < 1, \text{ zero elsewhere} \\ g_3(y_3) &= 6y_3^5, 0 < y_3 < 1, \text{ zero elsewhere.} \end{aligned}$$

Because $g(y_1, y_2, y_3) = g_1(y_1)g_2(y_2)g_3(y_3)$, the random variables Y_1, Y_2, Y_3 are mutually independent. ■

Example 2.7.2. Let X_1, X_2, X_3 be iid with common pdf

$$f(x) = \begin{cases} e^{-x} & 0 < x < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

Consequently, the joint pdf of X_1, X_2, X_3 is

$$f_{X_1, X_2, X_3}(x_1, x_2, x_3) = \begin{cases} e^{-\sum_{i=1}^3 x_i} & 0 < x_i < \infty, i = 1, 2, 3 \\ 0 & \text{elsewhere.} \end{cases}$$

Consider the random variables Y_1, Y_2, Y_3 defined by

$$Y_1 = \frac{X_1}{X_1 + X_2 + X_3}, \quad Y_2 = \frac{X_2}{X_1 + X_2 + X_3}, \quad \text{and} \quad Y_3 = X_1 + X_2 + X_3.$$

Hence, the inverse transformation is given by

$$x_1 = y_1 y_3, \quad x_2 = y_2 y_3, \quad \text{and} \quad x_3 = y_3 - y_1 y_3 - y_2 y_3,$$

with the Jacobian

$$J = \begin{vmatrix} y_3 & 0 & y_1 \\ 0 & y_3 & y_2 \\ -y_3 & -y_3 & 1 - y_1 - y_2 \end{vmatrix} = y_3^2.$$

The support of X_1, X_2, X_3 maps onto

$$0 < y_1 y_3 < \infty, \quad 0 < y_2 y_3 < \infty, \quad \text{and} \quad 0 < y_3(1 - y_1 - y_2) < \infty,$$

which is equivalent to the support $\mathcal{T}$ given by

$$\mathcal{T} = \{(y_1, y_2, y_3) : 0 < y_1, 0 < y_2, 0 < 1 - y_1 - y_2, 0 < y_3 < \infty\}.$$

Hence the joint pdf of Y_1, Y_2, Y_3 is

$$g(y_1, y_2, y_3) = y_3^2 e^{-y_3}, \quad (y_1, y_2, y_3) \in \mathcal{T}.$$

The marginal pdf of Y_1 is

$$g_1(y_1) = \int_0^{1-y_1} \int_0^\infty y_3^2 e^{-y_3} dy_3 dy_2 = 2(1-y_1), \quad 0 < y_1 < 1,$$

zero elsewhere. Likewise the marginal pdf of Y_2 is

$$g_2(y_2) = 2(1-y_2), \quad 0 < y_2 < 1,$$

zero elsewhere, while the pdf of Y_3 is

$$g_3(y_3) = \int_0^1 \int_0^{1-y_1} y_3^2 e^{-y_3} dy_2 dy_1 = \frac{1}{2} y_3^2 e^{-y_3}, \quad 0 < y_3 < \infty,$$

zero elsewhere. Because $g(y_1, y_2, y_3) \neq g_1(y_1)g_2(y_2)g_3(y_3)$, Y_1, Y_2, Y_3 are dependent random variables.

Note, however, that the joint pdf of Y_1 and Y_3 is

$$g_{13}(y_1, y_3) = \int_0^{1-y_1} y_3^2 e^{-y_3} dy_2 = (1-y_1)y_3^2 e^{-y_3}, \quad 0 < y_1 < 1, 0 < y_3 < \infty,$$

zero elsewhere. Hence Y_1 and Y_3 are independent. In a similar manner, Y_2 and Y_3 are also independent. Because the joint pdf of Y_1 and Y_2 is

$$g_{12}(y_1, y_2) = \int_0^\infty y_3^2 e^{-y_3} dy_3 = 2, \quad 0 < y_1, 0 < y_2, y_1 + y_2 < 1,$$

zero elsewhere, Y_1 and Y_2 are seen to be dependent. ■

We now consider some other problems that are encountered when transforming variables. Let X have the Cauchy pdf

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < \infty,$$

and let $Y = X^2$. We seek the pdf $g(y)$ of Y . Consider the transformation $y = x^2$. This transformation maps the space of X , namely $\mathcal{S} = \{x : -\infty < x < \infty\}$, onto $\mathcal{T} = \{y : 0 \leq y < \infty\}$. However, the transformation is not one-to-one. To each $y \in \mathcal{T}$, with the exception of $y = 0$, there correspond two points $x \in \mathcal{S}$. For example, if $y = 4$, we may have either $x = 2$ or $x = -2$. In such an instance, we represent $\mathcal{S}$ as the union of two disjoint sets A_1 and A_2 such that $y = x^2$ defines a one-to-one transformation that maps each of A_1 and A_2 onto $\mathcal{T}$. If we take A_1 to be $\{x : -\infty < x < 0\}$ and A_2 to be $\{x : 0 \leq x < \infty\}$, we see that A_1 is mapped onto $\{y : 0 < y < \infty\}$, whereas A_2 is mapped onto $\{y : 0 \leq y < \infty\}$, and these sets are not the same. Our difficulty is caused by the fact that $x = 0$ is an element of $\mathcal{S}$. Why, then, do we not return to the Cauchy pdf and take

$f(0) = 0$? Then our new $\mathcal{S}$ is $\mathcal{S} = \{-\infty < x < \infty \text{ but } x \neq 0\}$. We then take $A_1 = \{x : -\infty < x < 0\}$ and $A_2 = \{x : 0 < x < \infty\}$. Thus $y = x^2$, with the inverse $x = -\sqrt{y}$, maps A_1 onto $\mathcal{T} = \{y : 0 < y < \infty\}$ and the transformation is one-to-one. Moreover, the transformation $y = x^2$, with inverse $x = \sqrt{y}$, maps A_2 onto $\mathcal{T} = \{y : 0 < y < \infty\}$ and the transformation is one-to-one. Consider the probability $P(Y \in B)$, where $B \subset \mathcal{T}$. Let $A_3 = \{x : x = -\sqrt{y}, y \in B\} \subset A_1$ and let $A_4 = \{x : x = \sqrt{y}, y \in B\} \subset A_2$. Then $Y \in B$ when and only when $X \in A_3$ or $X \in A_4$. Thus we have

$$\begin{aligned} P(Y \in B) &= P(X \in A_3) + P(X \in A_4) \\ &= \int_{A_3} f(x) dx + \int_{A_4} f(x) dx. \end{aligned}$$

In the first of these integrals, let $x = -\sqrt{y}$. Thus the Jacobian, say J_1 , is $-1/2\sqrt{y}$; furthermore, the set A_3 is mapped onto B . In the second integral let $x = \sqrt{y}$. Thus the Jacobian, say J_2 , is $1/2\sqrt{y}$; furthermore, the set A_4 is also mapped onto B . Finally,

$$\begin{aligned} P(Y \in B) &= \int_B f(-\sqrt{y}) \left| -\frac{1}{2\sqrt{y}} \right| dy + \int_B f(\sqrt{y}) \frac{1}{2\sqrt{y}} dy \\ &= \int_B [f(-\sqrt{y}) + f(\sqrt{y})] \frac{1}{2\sqrt{y}} dy. \end{aligned}$$

Hence the pdf of Y is given by

$$g(y) = \frac{1}{2\sqrt{y}} [f(-\sqrt{y}) + f(\sqrt{y})], \quad y \in \mathcal{T}.$$

With $f(x)$ the Cauchy pdf we have

$$g(y) = \begin{cases} \frac{1}{\pi(1+y)\sqrt{y}} & 0 < y < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

In the preceding discussion of a random variable of the continuous type, we had two inverse functions, $x = -\sqrt{y}$ and $x = \sqrt{y}$. That is why we sought to partition $\mathcal{S}$ (or a modification of $\mathcal{S}$) into two disjoint subsets such that the transformation $y = x^2$ maps each onto the same $\mathcal{T}$. Had there been three inverse functions, we would have sought to partition $\mathcal{S}$ (or a modified form of $\mathcal{S}$) into three disjoint subsets, and so on. It is hoped that this detailed discussion makes the following paragraph easier to read.

Let $f(x_1, x_2, \dots, x_n)$ be the joint pdf of $X_1, X_2, \dots, X_n$, which are random variables of the continuous type. Let $\mathcal{S}$ denote the n -dimensional space where this joint pdf $f(x_1, x_2, \dots, x_n) > 0$, and consider the transformation $y_1 = u_1(x_1, x_2, \dots, x_n)$, $\dots$, $y_n = u_n(x_1, x_2, \dots, x_n)$, which maps $\mathcal{S}$ onto $\mathcal{T}$ in the $y_1, y_2, \dots, y_n$ space. To each point of $\mathcal{S}$ there corresponds, of course, only one point in $\mathcal{T}$; but to a point in $\mathcal{T}$ there may correspond more than one point in $\mathcal{S}$. That is, the transformation

may not be one-to-one. Suppose, however, that we can represent $\mathcal{S}$ as the union of a finite number, say k , of mutually disjoint sets $A_1, A_2, \dots, A_k$ so that

$$y_1 = u_1(x_1, x_2, \dots, x_n), \dots, y_n = u_n(x_1, x_2, \dots, x_n)$$

define a one-to-one transformation of each A_i onto $\mathcal{T}$. Thus to each point in $\mathcal{T}$ there corresponds exactly one point in each of $A_1, A_2, \dots, A_k$. For $i = 1, \dots, k$, let

$$x_1 = w_{1i}(y_1, y_2, \dots, y_n), x_2 = w_{2i}(y_1, y_2, \dots, y_n), \dots, x_n = w_{ni}(y_1, y_2, \dots, y_n),$$

denote the k groups of n inverse functions, one group for each of these k transformations. Let the first partial derivatives be continuous and let each

$$J_i = \begin{vmatrix} \frac{\partial w_{1i}}{\partial y_1} & \frac{\partial w_{1i}}{\partial y_2} & \dots & \frac{\partial w_{1i}}{\partial y_n} \\ \frac{\partial w_{2i}}{\partial y_1} & \frac{\partial w_{2i}}{\partial y_2} & \dots & \frac{\partial w_{2i}}{\partial y_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial w_{ni}}{\partial y_1} & \frac{\partial w_{ni}}{\partial y_2} & \dots & \frac{\partial w_{ni}}{\partial y_n} \end{vmatrix}, \quad i = 1, 2, \dots, k,$$

be not identically equal to zero in $\mathcal{T}$. Considering the probability of the union of k mutually exclusive events and by applying the change-of-variable technique to the probability of each of these events, it can be seen that the joint pdf of $Y_1 = u_1(X_1, X_2, \dots, X_n)$, $Y_2 = u_2(X_1, X_2, \dots, X_n), \dots, Y_n = u_n(X_1, X_2, \dots, X_n)$, is given by

$$g(y_1, y_2, \dots, y_n) = \sum_{i=1}^k f[w_{1i}(y_1, \dots, y_n), \dots, w_{ni}(y_1, \dots, y_n)] |J_i|,$$

provided that $(y_1, y_2, \dots, y_n) \in \mathcal{T}$, and equals zero elsewhere. The pdf of any Y_i , say Y_1 , is then

$$g_1(y_1) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(y_1, y_2, \dots, y_n) dy_2 \dots dy_n.$$

Example 2.7.3. Let X_1 and X_2 have the joint pdf defined over the unit circle given by

$$f(x_1, x_2) = \begin{cases} \frac{1}{\pi} & 0 < x_1^2 + x_2^2 < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Let $Y_1 = X_1^2 + X_2^2$ and $Y_2 = X_1^2 / (X_1^2 + X_2^2)$. Thus $y_1 y_2 = x_1^2$ and $x_2^2 = y_1(1 - y_2)$. The support $\mathcal{S}$ maps onto $\mathcal{T} = \{(y_1, y_2) : 0 < y_i < 1, i = 1, 2\}$. For each ordered pair $(y_1, y_2) \in \mathcal{T}$, there are four points in $\mathcal{S}$, given by

$$\begin{aligned} (x_1, x_2) & \text{ such that } x_1 = \sqrt{y_1 y_2} \text{ and } x_2 = \sqrt{y_1(1 - y_2)} \\ (x_1, x_2) & \text{ such that } x_1 = \sqrt{y_1 y_2} \text{ and } x_2 = -\sqrt{y_1(1 - y_2)} \\ (x_1, x_2) & \text{ such that } x_1 = -\sqrt{y_1 y_2} \text{ and } x_2 = \sqrt{y_1(1 - y_2)} \\ \text{and } (x_1, x_2) & \text{ such that } x_1 = -\sqrt{y_1 y_2} \text{ and } x_2 = -\sqrt{y_1(1 - y_2)}. \end{aligned}$$

The value of the first Jacobian is

$$\begin{aligned} J_1 &= \begin{vmatrix} \frac{1}{2}\sqrt{y_2/y_1} & \frac{1}{2}\sqrt{y_1/y_2} \\ \frac{1}{2}\sqrt{(1-y_2)/y_1} & -\frac{1}{2}\sqrt{y_1/(1-y_2)} \end{vmatrix} \\ &= \frac{1}{4} \left\{ -\sqrt{\frac{1-y_2}{y_2}} - \sqrt{\frac{y_2}{1-y_2}} \right\} = -\frac{1}{4} \frac{1}{\sqrt{y_2(1-y_2)}}. \end{aligned}$$

It is easy to see that the absolute value of each of the four Jacobians equals $1/4\sqrt{y_2(1-y_2)}$. Hence, the joint pdf of Y_1 and Y_2 is the sum of four terms and can be written as

$$g(y_1, y_2) = 4 \frac{1}{\pi 4\sqrt{y_2(1-y_2)}} = \frac{1}{\pi\sqrt{y_2(1-y_2)}}, \quad (y_1, y_2) \in \mathcal{T}.$$

Thus Y_1 and Y_2 are independent random variables by Theorem 2.4.1. ■

Of course, as in the bivariate case, we can use the mgf technique by noting that if $Y = g(X_1, X_2, \dots, X_n)$ is a function of the random variables, then the mgf of Y is given by

$$E(e^{tY}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{tg(x_1, x_2, \dots, x_n)} f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n,$$

in the continuous case, where $f(x_1, x_2, \dots, x_n)$ is the joint pdf. In the discrete case, summations replace the integrals. This procedure is particularly useful in cases in which we are dealing with linear functions of independent random variables.

Example 2.7.4 (Extension of Example 2.2.6). Let X_1, X_2, X_3 be independent random variables with joint pmf

$$p(x_1, x_2, x_3) = \begin{cases} \frac{\mu_1^{x_1} \mu_2^{x_2} \mu_3^{x_3} e^{-\mu_1 - \mu_2 - \mu_3}}{x_1! x_2! x_3!} & x_i = 0, 1, 2, \dots, i = 1, 2, 3 \\ 0 & \text{elsewhere.} \end{cases}$$

If $Y = X_1 + X_2 + X_3$, the mgf of Y is

$$\begin{aligned} E(e^{tY}) &= E(e^{t(X_1 + X_2 + X_3)}) \\ &= E(e^{tX_1} e^{tX_2} e^{tX_3}) \\ &= E(e^{tX_1}) E(e^{tX_2}) E(e^{tX_3}), \end{aligned}$$

because of the independence of X_1, X_2, X_3 . In Example 2.2.6, we found that

$$E(e^{tX_i}) = \exp\{\mu_i(e^t - 1)\}, \quad i = 1, 2, 3.$$

Hence,

$$E(e^{tY}) = \exp\{(\mu_1 + \mu_2 + \mu_3)(e^t - 1)\}.$$

This, however, is the mgf of the pmf

$$p_Y(y) = \begin{cases} \frac{(\mu_1 + \mu_2 + \mu_3)^y e^{-(\mu_1 + \mu_2 + \mu_3)}}{y!} & y = 0, 1, 2, \dots \\ 0 & \text{elsewhere,} \end{cases}$$

so $Y = X_1 + X_2 + X_3$ has this distribution. ■

Example 2.7.5. Let X_1, X_2, X_3, X_4 be independent random variables with common pdf

$$f(x) = \begin{cases} e^{-x} & x > 0 \\ 0 & \text{elsewhere.} \end{cases}$$

If $Y = X_1 + X_2 + X_3 + X_4$, then similar to the argument in the last example, the independence of X_1, X_2, X_3, X_4 implies that

$$E(e^{tY}) = E(e^{tX_1}) E(e^{tX_2}) E(e^{tX_3}) E(e^{tX_4}).$$

In Section 1.9, we saw that

$$E(e^{tX_i}) = (1 - t)^{-1}, \quad t < 1, \quad i = 1, 2, 3, 4.$$

Hence,

$$E(e^{tY}) = (1 - t)^{-4}.$$

In Section 3.3, we find that this is the mgf of a distribution with pdf

$$f_Y(y) = \begin{cases} \frac{1}{3!} y^3 e^{-y} & 0 < y < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

Accordingly, Y has this distribution. ■

EXERCISES

2.7.1. Let X_1, X_2, X_3 be iid, each with the distribution having pdf $f(x) = e^{-x}$, $0 < x < \infty$, zero elsewhere. Show that

$$Y_1 = \frac{X_1}{X_1 + X_2}, \quad Y_2 = \frac{X_1 + X_2}{X_1 + X_2 + X_3}, \quad Y_3 = X_1 + X_2 + X_3$$

are mutually independent.

2.7.2. If $f(x) = \frac{1}{2}$, $-1 < x < 1$, zero elsewhere, is the pdf of the random variable X , find the pdf of $Y = X^2$.

2.7.3. If X has the pdf of $f(x) = \frac{1}{4}$, $-1 < x < 3$, zero elsewhere, find the pdf of $Y = X^2$.

Hint: Here $\mathcal{T} = \{y : 0 \leq y < 9\}$ and the event $Y \in B$ is the union of two mutually exclusive events if $B = \{y : 0 < y < 1\}$.

2.7.4. Let X_1, X_2, X_3 be iid with common pdf $f(x) = e^{-x}$, $x > 0$, 0 elsewhere. Find the joint pdf of $Y_1 = X_1$, $Y_2 = X_1 + X_2$, and $Y_3 = X_1 + X_2 + X_3$.

2.7.5. Let X_1, X_2, X_3 be iid with common pdf $f(x) = e^{-x}$, $x > 0$, 0 elsewhere. Find the joint pdf of $Y_1 = X_1/X_2$, $Y_2 = X_3/(X_1 + X_2)$, and $Y_3 = X_1 + X_2$. Are Y_1, Y_2, Y_3 mutually independent?

2.7.6. Let X_1, X_2 have the joint pdf $f(x_1, x_2) = 1/\pi$, $0 < x_1^2 + x_2^2 < 1$. Let $Y_1 = X_1^2 + X_2^2$ and $Y_2 = X_2$. Find the joint pdf of Y_1 and Y_2 .

2.7.7. Let X_1, X_2, X_3, X_4 have the joint pdf $f(x_1, x_2, x_3, x_4) = 24$, $0 < x_1 < x_2 < x_3 < x_4 < 1$, 0 elsewhere. Find the joint pdf of $Y_1 = X_1/X_2$, $Y_2 = X_2/X_3$, $Y_3 = X_3/X_4$, $Y_4 = X_4$ and show that they are mutually independent.

2.7.8. Let X_1, X_2, X_3 be iid with common mgf $M(t) = ((3/4) + (1/4)e^t)^2$, for all $t \in R$.

- (a) Determine the probabilities, $P(X_1 = k)$, $k = 0, 1, 2$.
- (b) Find the mgf of $Y = X_1 + X_2 + X_3$ and then determine the probabilities, $P(Y = k)$, $k = 0, 1, 2, \dots, 6$.

2.8 Linear Combinations of Random Variables

In this section, we summarize some results on linear combinations of random variables that follow from Section 2.6. These results will prove to be quite useful in Chapter 3 as well as in succeeding chapters.

Let $(X_1, \dots, X_n)'$ denote a random vector. In this section, we consider linear combinations of these variables, writing them, generally, as

$$T = \sum_{i=1}^n a_i X_i, \quad (2.8.1)$$

for specified constants $a_1, \dots, a_n$. We obtain expressions for the mean and variance of T .

The mean of T follows immediately from linearity of expectation. For reference, we state it formally as a theorem.

Theorem 2.8.1. Suppose T is given by expression (2.8.1). Suppose $E(X_i) = \mu_i$, for $i = 1, \dots, n$. Then

$$E(T) = \sum_{i=1}^n a_i \mu_i. \quad (2.8.2)$$

In order to obtain the variance of T , we first state a general result on covariances.

Theorem 2.8.2. Suppose T is the linear combination (2.8.1) and that W is another linear combination given by $W = \sum_{i=1}^m b_i Y_i$, for random variables $Y_1, \dots, Y_m$ and specified constants $b_1, \dots, b_m$. Let $T = \sum_{i=1}^n a_i X_i$ and let $W = \sum_{i=1}^m b_i Y_i$. If $E[X_i^2] < \infty$, and $E[Y_j^2] < \infty$ for $i = 1, \dots, n$ and $j = 1, \dots, m$, then

$$\text{Cov}(T, W) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \text{Cov}(X_i, Y_j). \quad (2.8.3)$$

Proof: Using the definition of the covariance and Theorem 2.8.1, we have the first equality below, while the second equality follows from the linearity of E :

$$\begin{aligned}\text{Cov}(T, W) &= E \left[\sum_{i=1}^n \sum_{j=1}^m (a_i X_i - a_i E(X_i))(b_j Y_j - b_j E(Y_j)) \right] \\ &= \sum_{i=1}^n \sum_{j=1}^m a_i b_j E[(X_i - E(X_i))(Y_j - E(Y_j))],\end{aligned}$$

which is the desired result. ■

To obtain the variance of T , simply replace W by T in expression (2.8.3). We state the result as a corollary:

Corollary 2.8.1. *Let $T = \sum_{i=1}^n a_i X_i$. Provided $E[X_i^2] < \infty$, for $i = 1, \dots, n$,*

$$\text{Var}(T) = \text{Cov}(T, T) = \sum_{i=1}^n a_i^2 \text{Var}(X_i) + 2 \sum_{i < j} a_i a_j \text{Cov}(X_i, X_j). \quad (2.8.4)$$

Note that if $X_1, \dots, X_n$ are independent random variables, then by Theorem 2.5.2 all the pairwise covariances are 0; i.e., $\text{Cov}(X_i, X_j) = 0$ for all $i \neq j$. This leads to a simplification of (2.8.4), which we record in the following corollary.

Corollary 2.8.2. *If $X_1, \dots, X_n$ are independent random variables and $\text{Var}(X_i) = \sigma_i^2$, for $i = 1, \dots, n$, then*

$$\text{Var}(T) = \sum_{i=1}^n a_i^2 \sigma_i^2. \quad (2.8.5)$$

Note that we need only X_i and X_j to be uncorrelated for all $i \neq j$ to obtain this result.

Next, in addition to independence, we assume that the random variables have the same distribution. We call such a collection of random variables a *random sample* which we now state in a formal definition.

Definition 2.8.1. *If the random variables $X_1, X_2, \dots, X_n$ are independent and identically distributed, i.e. each X_i has the same distribution, then we say that these random variables constitute a **random sample** of size n from that common distribution. We abbreviate independent and identically distributed by **iid**. ■*

In the next two examples, we find some properties of two functions of a random sample, namely the sample mean and variance.

Example 2.8.1 (Sample Mean). Let $X_1, \dots, X_n$ be independent and identically distributed random variables with common mean μ and variance σ^2 . The **sample mean** is defined by $\bar{X} = n^{-1} \sum_{i=1}^n X_i$. This is a linear combination of the sample observations with $a_i \equiv n^{-1}$; hence, by Theorem 2.8.1 and Corollary 2.8.2, we have

$$E(\bar{X}) = \mu \text{ and } \text{Var}(\bar{X}) = \frac{\sigma^2}{n}. \quad (2.8.6)$$

Because $E(\bar{X}) = \mu$, we often say that $\bar{X}$ is **unbiased** for μ . ■

Example 2.8.2 (Sample Variance). Define the **sample variance** by

$$S^2 = (n-1)^{-1} \sum_{i=1}^n (X_i - \bar{X})^2 = (n-1)^{-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right), \quad (2.8.7)$$

where the second equality follows after some algebra; see Exercise 2.8.1.

In the average that defines the sample variance S^2 , the division is by $n-1$ instead of n . One reason for this is that it makes S^2 unbiased for σ^2 , as next shown. Using the above theorems, the results of the last example, and the facts that $E(X^2) = \sigma^2 + \mu^2$ and $E(\bar{X}^2) = (\sigma^2/n) + \mu^2$, we have the following:

$$\begin{aligned} E(S^2) &= (n-1)^{-1} \left(\sum_{i=1}^n E(X_i^2) - nE(\bar{X}^2) \right) \\ &= (n-1)^{-1} \{ n\sigma^2 + n\mu^2 - n[(\sigma^2/n) + \mu^2] \} \\ &= \sigma^2. \end{aligned} \quad (2.8.8)$$

Hence, S^2 is unbiased for σ^2 . ■

EXERCISES

2.8.1. Derive the second equality in expression (2.8.7).

2.8.2. Let X_1, X_2, X_3, X_4 be four iid random variables having the same pdf $f(x) = 2x$, $0 < x < 1$, zero elsewhere. Find the mean and variance of the sum Y of these four random variables.

2.8.3. Let X_1 and X_2 be two independent random variables so that the variances of X_1 and X_2 are $\sigma_1^2 = k$ and $\sigma_2^2 = 2$, respectively. Given that the variance of $Y = 3X_2 - X_1$ is 25, find k .

2.8.4. If the independent variables X_1 and X_2 have means μ_1, μ_2 and variances σ_1^2, σ_2^2 , respectively, show that the mean and variance of the product $Y = X_1X_2$ are $\mu_1\mu_2$ and $\sigma_1^2\sigma_2^2 + \mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2$, respectively.

2.8.5. Find the mean and variance of the sum $Y = \sum_{i=1}^5 X_i$, where $X_1, \dots, X_5$ are iid, having pdf $f(x) = 6x(1-x)$, $0 < x < 1$, zero elsewhere.

2.8.6. Determine the mean and variance of the sample mean $\bar{X} = 5^{-1} \sum_{i=1}^5 X_i$, where $X_1, \dots, X_5$ is a random sample from a distribution having pdf $f(x) = 4x^3$, $0 < x < 1$, zero elsewhere.

2.8.7. Let X and Y be random variables with $\mu_1 = 1$, $\mu_2 = 4$, $\sigma_1^2 = 4$, $\sigma_2^2 = 6$, $\rho = \frac{1}{2}$. Find the mean and variance of the random variable $Z = 3X - 2Y$.

2.8.8. Let X and Y be independent random variables with means μ_1, μ_2 and variances σ_1^2, σ_2^2 . Determine the correlation coefficient of X and $Z = X - Y$ in terms of $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$.

2.8.9. Let μ and σ^2 denote the mean and variance of the random variable X . Let $Y = c + bX$, where b and c are real constants. Show that the mean and variance of Y are, respectively, $c + b\mu$ and $b^2\sigma^2$.

2.8.10. Determine the correlation coefficient of the random variables X and Y if $\text{var}(X) = 4$, $\text{var}(Y) = 2$, and $\text{var}(X + 2Y) = 15$.

2.8.11. Let X and Y be random variables with means μ_1 , μ_2 ; variances σ_1^2 , σ_2^2 ; and correlation coefficient ρ . Show that the correlation coefficient of $W = aX + b$, $a > 0$, and $Z = cY + d$, $c > 0$, is ρ .

2.8.12. A person rolls a die, tosses a coin, and draws a card from an ordinary deck. He receives \$3 for each point up on the die, \$10 for a head and \$0 for a tail, and \$1 for each spot on the card (jack = 11, queen = 12, king = 13). If we assume that the three random variables involved are independent and uniformly distributed, compute the mean and variance of the amount to be received.

2.8.13. Let X_1 and X_2 be independent random variables with nonzero variances. Find the correlation coefficient of $Y = X_1X_2$ and X_1 in terms of the means and variances of X_1 and X_2 .

2.8.14. Let X_1 and X_2 have a joint distribution with parameters μ_1 , μ_2 , σ_1^2 , σ_2^2 , and ρ . Find the correlation coefficient of the linear functions of $Y = a_1X_1 + a_2X_2$ and $Z = b_1X_1 + b_2X_2$ in terms of the real constants a_1 , a_2 , b_1 , b_2 , and the parameters of the distribution.

2.8.15. Let X_1 , X_2 , and X_3 be random variables with equal variances but with correlation coefficients $\rho_{12} = 0.3$, $\rho_{13} = 0.5$, and $\rho_{23} = 0.2$. Find the correlation coefficient of the linear functions $Y = X_1 + X_2$ and $Z = X_2 + X_3$.

2.8.16. Find the variance of the sum of 10 random variables if each has variance 5 and if each pair has correlation coefficient 0.5.

2.8.17. Let X and Y have the parameters μ_1 , μ_2 , σ_1^2 , σ_2^2 , and ρ . Show that the correlation coefficient of X and $[Y - \rho(\sigma_2/\sigma_1)X]$ is zero.

2.8.18. Let S^2 be the sample variance of a random sample from a distribution with variance $\sigma^2 > 0$. Since $E(S^2) = \sigma^2$, why isn't $E(S) = \sigma$?

Hint: Use Jensen's inequality to show that $E(S) < \sigma$.

Chapter 3

Some Special Distributions

3.1 The Binomial and Related Distributions

In Chapter 1 we introduced the *uniform distribution* and the *hypergeometric distribution*. In this chapter we discuss some other important distributions of random variables frequently used in statistics. We begin with the binomial and related distributions.

A **Bernoulli experiment** is a random experiment, the outcome of which can be classified in but one of two mutually exclusive and exhaustive ways, for instance, success or failure (e.g., female or male, life or death, nondefective or defective). A sequence of **Bernoulli trials** occurs when a Bernoulli experiment is performed several independent times so that the probability of success, say p , remains the same from trial to trial. That is, in such a sequence, we let p denote the probability of success on each trial.

Let X be a random variable associated with a Bernoulli trial by defining it as follows:

$$X(\text{success}) = 1 \text{ and } X(\text{failure}) = 0.$$

That is, the two outcomes, success and failure, are denoted by one and zero, respectively. The pmf of X can be written as

$$p(x) = p^x(1-p)^{1-x}, \quad x = 0, 1, \quad (3.1.1)$$

and we say that X has a **Bernoulli distribution**. The expected value of X is

$$\mu = E(X) = (0)(1-p) + (1)(p) = p,$$

and the variance of X is

$$\sigma^2 = \text{var}(X) = p^2(1-p) + (1-p)^2p = p(1-p).$$

It follows that the standard deviation of X is $\sigma = \sqrt{p(1-p)}$.

In a sequence of n independent Bernoulli trials, where the probability of success remains constant, let X_i denote the Bernoulli random variable associated with the

i th trial. An observed sequence of n Bernoulli trials is then an n -tuple of zeros and ones. In such a sequence of Bernoulli trials, we are often interested in the total number of successes and not in the order of their occurrence. If we let the random variable X equal the number of observed successes in n Bernoulli trials, the possible values of X are $0, 1, 2, \dots, n$. If x successes occur, where $x = 0, 1, 2, \dots, n$, then $n - x$ failures occur. The number of ways of selecting the x positions for the x successes in the n trials is

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}.$$

Since the trials are independent and the probabilities of success and failure on each trial are, respectively, p and $1 - p$, the probability of each of these ways is $p^x(1-p)^{n-x}$. Thus the pmf of X , say $p(x)$, is the sum of the probabilities of these $\binom{n}{x}$ mutually exclusive events; that is,

$$p(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, 2, \dots, n \\ 0 & \text{elsewhere.} \end{cases} \quad (3.1.2)$$

It is clear that $p(x) \geq 0$. To verify that $p(x)$ sums to 1 over its range, recall the binomial series, expression (1.3.7) of Chapter 1, which is:

$$(a+b)^n = \sum_{x=0}^n \binom{n}{x} b^x a^{n-x},$$

for n a positive integer. Thus,

$$\begin{aligned} \sum_x p(x) &= \sum_{x=0}^n \binom{n}{x} p^x (1-p)^{n-x} \\ &= [(1-p) + p]^n = 1. \end{aligned}$$

Therefore, $p(x)$ satisfies the conditions of being a pmf of a random variable X of the discrete type. A random variable X that has a pmf of the form of $p(x)$ is said to have a **binomial distribution**, and any such $p(x)$ is called a **binomial pmf**. A binomial distribution is denoted by the symbol $b(n, p)$. The constants n and p are called the **parameters** of the binomial distribution.

Example 3.1.1 (Computation of Binomial Probabilities). Suppose we roll a fair six-sided die 3 times. What is the probability of getting exactly 2 sixes? For our notation, let X be the number of sixes obtained in the 3 rolls. Then X has a binomial distribution with $n = 3$ and $p = 1/6$. Hence,

$$P(X = 2) = p(2) = \binom{3}{2} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^1 = 0.06944.$$

We can do this calculation with a hand calculator. Suppose, though, we want to determine the probability of at least 16 sixes in 60 rolls. Let Y be the number of sixes in 60 rolls. Then our desired probability is given by the series

$$P(Y \geq 16) = \sum_{j=16}^{60} \binom{60}{j} \left(\frac{1}{6}\right)^j \left(\frac{5}{6}\right)^{60-j},$$

which is not a simple calculation. Most statistical packages provide procedures to calculate binomial probabilities. In R, if Y is $b(n, p)$ then the cdf of Y is computed as $P(Y \leq y) = \text{pbinom}(y, n, p)$. Hence, for our example, using R we compute the $P(Y \geq 16)$ as

$$P(Y \geq 16) = 1 - P(Y \leq 15) = 1 - \text{pbinom}(15, 60, 1/6) = 0.0338.$$

The R function `dbinom` computes the pmf of a binomial distribution. For instance, to compute the probability that $Y = 11$, we use the R code: `dbinom(11, 60, 1/6)`, which computes to 0.1246. ■

The mgf of a binomial distribution is easily obtained as follows:

$$\begin{aligned} M(t) &= \sum_x e^{tx} p(x) = \sum_{x=0}^n e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^n \binom{n}{x} (pe^t)^x (1-p)^{n-x} \\ &= [(1-p) + pe^t]^n \end{aligned}$$

for all real values of t . The mean μ and the variance σ^2 of X may be computed from $M(t)$. Since

$$M'(t) = n[(1-p) + pe^t]^{n-1} (pe^t)$$

and

$$M''(t) = n[(1-p) + pe^t]^{n-1} (pe^t) + n(n-1)[(1-p) + pe^t]^{n-2} (pe^t)^2,$$

it follows that

$$\mu = M'(0) = np$$

and

$$\sigma^2 = M''(0) - \mu^2 = np + n(n-1)p^2 - (np)^2 = np(1-p).$$

Suppose Y has the $b(60, 1/6)$ distribution as discussed in Example 3.1.1. Then $E(Y) = 60(1/6) = 10$ and $\text{Var}(Y) = 60(1/6)(5/6) = 8.33$

Example 3.1.2. If the mgf of a random variable X is

$$M(t) = \left(\frac{2}{3} + \frac{1}{3}e^t\right)^5,$$

then X has a binomial distribution with $n = 5$ and $p = \frac{1}{3}$; that is, the pmf of X is

$$p(x) = \begin{cases} \binom{5}{x} \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{5-x} & x = 0, 1, 2, \dots, 5 \\ 0 & \text{elsewhere.} \end{cases}$$

Here $\mu = np = \frac{5}{3}$ and $\sigma^2 = np(1-p) = \frac{10}{9}$. ■

Example 3.1.3. If Y is $b(n, \frac{1}{3})$, then $P(Y \geq 1) = 1 - P(Y = 0) = 1 - (\frac{2}{3})^n$. Suppose that we wish to find the smallest value of n that yields $P(Y \geq 1) > 0.80$. We have $1 - (\frac{2}{3})^n > 0.80$ and $0.20 > (\frac{2}{3})^n$. Either by inspection or by use of logarithms, we see that $n = 4$ is the solution. That is, the probability of at least one success throughout $n = 4$ independent repetitions of a random experiment with probability of success $p = \frac{1}{3}$ is greater than 0.80. ■

Example 3.1.4. Let the random variable Y be equal to the number of successes throughout n independent repetitions of a random experiment with probability p of success. That is, Y is $b(n, p)$. The ratio Y/n is called the relative frequency of success. Recall expression (1.10.3), the second version of Chebyshev's inequality (Theorem 1.10.3). Applying this result, we have for all $\epsilon > 0$ that

$$P\left(\left|\frac{Y}{n} - p\right| \geq \epsilon\right) \leq \frac{\text{Var}(Y/n)}{\epsilon^2} = \frac{p(1-p)}{n\epsilon^2}$$

[Exercise 3.1.3 asks for the determination of $\text{Var}(Y/n)$]. Now, for every fixed $\epsilon > 0$, the right-hand member of the preceding inequality is close to zero for sufficiently large n . That is,

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{Y}{n} - p\right| \geq \epsilon\right) = 0$$

and

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{Y}{n} - p\right| < \epsilon\right) = 1.$$

Since this is true for every fixed $\epsilon > 0$, we see, in a certain sense, that the relative frequency of success is for large values of n , close to the probability of p of success. This result is one form of the **Weak Law of Large Numbers**. It was alluded to in the initial discussion of probability in Chapter 1 and is considered again, along with related concepts, in Chapter 5. ■

Example 3.1.5. Let the independent random variables X_1, X_2, X_3 have the same cdf $F(x)$. Let Y be the middle value of X_1, X_2, X_3 . To determine the cdf of Y , say $F_Y(y) = P(Y \leq y)$, we note that $Y \leq y$ if and only if at least two of the random variables X_1, X_2, X_3 are less than or equal to y . Let us say that the i th "trial" is a success if $X_i \leq y$, $i = 1, 2, 3$; here each "trial" has the probability of success $F(y)$. In this terminology, $F_Y(y) = P(Y \leq y)$ is then the probability of at least two successes in three independent trials. Thus

$$F_Y(y) = \binom{3}{2}[F(y)]^2[1 - F(y)] + [F(y)]^3.$$

If $F(x)$ is a continuous cdf so that the pdf of X is $F'(x) = f(x)$, then the pdf of Y is

$$f_Y(y) = F'_Y(y) = 6[F(y)][1 - F(y)]f(y). \quad \blacksquare$$

Suppose we have several independent binomial distributions with the same probability of success. Then it makes sense that the sum of these random variables is binomial, as shown in the following theorem.

Theorem 3.1.1. Let $X_1, X_2, \dots, X_m$ be independent random variables such that X_i has binomial $b(n_i, p)$ distribution, for $i = 1, 2, \dots, m$. Let $Y = \sum_{i=1}^m X_i$. Then Y has a binomial $b(\sum_{i=1}^m n_i, p)$ distribution.

Proof: The mgf of X_i is $M_{X_i}(t) = (1 - p + pe^t)^{n_i}$. By independence it follows from Theorem 2.6.1 that

$$M_Y(t) = \prod_{i=1}^m (1 - p + pe^t)^{n_i} = (1 - p + pe^t)^{\sum_{i=1}^m n_i}.$$

Hence, Y has a binomial $b(\sum_{i=1}^m n_i, p)$ distribution. ■

For the remainder of this section, we discuss some important distributions that are related to the binomial distribution.

3.1.1 Negative Binomial and Geometric Distributions

Consider a sequence of independent Bernoulli trials with constant probability p of success. Let the random variable Y denote the total number of failures in this sequence before the r th success, that is, $Y + r$ is equal to the number of trials necessary to produce exactly r successes with the last trial as a success. Here r is a fixed positive integer. To determine the pmf of Y , let y be an element of $\{y : y = 0, 1, 2, \dots\}$. Then, since the trials are independent, $P(Y = y)$ is equal to the product of the probability of obtaining exactly $r - 1$ successes in the first $y + r - 1$ trials times the probability p of a success on the $(y + r)$ th trial. Thus the pmf of Y is

$$p_Y(y) = \begin{cases} \binom{y+r-1}{r-1} p^r (1-p)^y & y = 0, 1, 2, \dots \\ 0 & \text{elsewhere.} \end{cases} \quad (3.1.3)$$

A distribution with a pmf of the form $p_Y(y)$ is called a **negative binomial distribution** and any such $p_Y(y)$ is called a negative binomial pmf. The distribution derives its name from the fact that $p_Y(y)$ is a general term in the expansion of $p^r [1 - (1-p)]^{-r}$. It is left as an exercise to show that the mgf of this distribution is $M(t) = p^r [1 - (1-p)e^t]^{-r}$, for $t < -\log(1-p)$. The R call to compute $P(y \leq y)$ is `pnbinom(y, r, p)`.

Example 3.1.6. Suppose the probability that a person has blood type B is 0.12. In order to conduct a study concerning people with blood type B, patients are sampled independently of one another until 10 are obtained who have blood type B. Determine the probability that at most 30 patients have to have their blood type determined. Let Y have a negative binomial distribution with $p = 0.12$ and $r = 10$. Then, the desired probability is

$$P(Y \leq 20) = \sum_{j=0}^{20} \binom{j+9}{9} 0.12^{10} 0.88^j.$$

Its computation in R is `pnbinom(20, 10, 0.12) = 0.0019`. ■

If $r = 1$, then Y has the pmf

$$p_Y(y) = p(1-p)^y, \quad y = 0, 1, 2, \dots, \quad (3.1.4)$$

zero elsewhere, and the mgf $M(t) = p[1 - (1-p)e^t]^{-1}$. In this special case, $r = 1$, we say that Y has a **geometric distribution**. In terms of Bernoulli trials, Y is the number of failures until the first success. The geometric distribution was first discussed in Example 1.6.3 of Chapter 1. For the last example, the probability that exactly 11 patients have to have their blood type determined before the first patient with type B blood is found is given by $.88^{11}0.12$. This is computed in R by `dgeom(11,0.12) = 0.0294`.

3.1.2 Multinomial Distribution

The binomial distribution is generalized to the multinomial distribution as follows. Let a random experiment be repeated n independent times. On each repetition, there is one and only one outcome from one of k categories. Call the categories $C_1, C_2, \dots, C_k$. For example, the upface of a roll of a six-sided die. Then the categories are $C_i = \{i\}$, $i = 1, 2, \dots, 6$. For $i = 1, \dots, k$, let p_i be the probability that the outcome is an element of C_i and assume that p_i remains constant throughout the n independent repetitions. Define the random variable X_i to be equal to the number of outcomes that are elements of C_i , $i = 1, 2, \dots, k-1$. Because $X_k = n - X_1 - \dots - X_{k-1}$, X_k is determined by the other X_i 's. Hence, for the joint distribution of interest we need only consider $X_1, X_2, \dots, X_{k-1}$.

The joint pmf of $(X_1, X_2, \dots, X_{k-1})$ is

$$P(X_1 = x_1, X_2 = x_2, \dots, X_{k-1} = x_{k-1}) = \frac{n!}{x_1! \dots x_{k-1}! x_k!} p_1^{x_1} \dots p_{k-1}^{x_{k-1}} p_k^{x_k}, \quad (3.1.5)$$

for all $x_1, x_2, \dots, x_{k-1}$ that are nonnegative integers and such that $x_1 + x_2 + \dots + x_{k-1} \leq n$, where $x_k = n - x_1 - \dots - x_{k-1}$ and $p_k = 1 - \sum_{j=1}^{k-1} p_j$. We next show that expression (3.1.5) is correct. The number of distinguishable arrangements of x_1 C_1 's, x_2 C_2 's, $\dots$, x_k C_k 's is

$$\binom{n}{x_1} \binom{n-x_1}{x_2} \dots \binom{n-x_1-\dots-x_{k-2}}{x_{k-1}} = \frac{n!}{x_1! x_2! \dots x_k!}$$

and the probability of each of these distinguishable arrangements is

$$p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}.$$

Hence the product of these two latter expressions gives the correct probability, which is in agreement with expression (3.1.5).

We say that $(X_1, X_2, \dots, X_{k-1})$ has a **multinomial distribution** with parameters n and $p_1, \dots, p_{k-1}$. The joint mgf of $(X_1, X_2, \dots, X_{k-1})$ is $M(t_1, \dots, t_{k-1}) = E(\exp\{\sum_{i=1}^{k-1} t_i X_i\})$, i.e.,

$$M(t_1, \dots, t_{k-1}) = \sum \dots \sum \frac{n!}{x_1! \dots x_{k-1}! x_k!} (p_1 e^{t_1})^{x_1} \dots (p_{k-1} e^{t_{k-1}})^{x_{k-1}} p_k^{x_k},$$

where the multiple sum is taken over all nonnegative integers and such that $x_1 + x_2 + \cdots + x_{k-1} \leq n$. Let $m = \sum_{i=1}^{k-1} p_i e^{t_i} + p_{k-1}$. Recall that $x_k = n - \sum_{i=1}^{k-1} x_i$. Then since $m > 0$, we have

$$\begin{aligned} M(t_1, \dots, t_{k-1}) &= m^n \sum \cdots \sum \frac{n!}{x_1! \cdots x_{k-1}! x_k!} \\ &\quad \times \left(\frac{p_1 e^{t_1}}{m} \right)^{x_1} \cdots \left(\frac{p_{k-1} e^{t_{k-1}}}{m} \right)^{x_{k-1}} \left(\frac{p_k}{m} \right)^{x_k} \\ &= m^n \times 1 = \left(\sum_{i=1}^{k-1} p_i e^{t_i} + p_{k-1} \right)^n, \end{aligned} \quad (3.1.6)$$

where we have used the property that sum of a pmf over its support is 1.

We can use the joint mgf to determine marginal distributions. The mgf of X_i is

$$M(0, \dots, 0, t_i, 0, \dots, 0) = (p_i e^{t_i} + (1 - p_i))^n;$$

hence, X_i is binomial with parameters n and p_i . The mgf of (X_i, X_j) , $i < j$, is

$$M(0, \dots, 0, t_i, 0, \dots, 0, t_j, 0, \dots, 0) = (p_i e^{t_i} + p_j e^{t_j} + (1 - p_i - p_j))^n;$$

so that (X_i, X_j) has a multinomial distribution with parameters n , p_i , and p_j . At times, we say that (X_1, X_2) has a **trinomial distribution**.

Another distribution of interest is the conditional distribution of X_i given X_j . For convenience, we select $i = 2$ and $j = 1$. We know that (X_1, X_2) is multinomial with parameters n and p_1 and p_2 and that X_1 is binomial with parameters n and p_1 . Thus, the conditional pmf is,

$$\begin{aligned} p_{X_2|X_1}(x_2 | x_1) &= \frac{p_{X_1, X_2}(x_1, x_2)}{p_{X_1}(x_1)} \\ &= \frac{x_1!(n - x_1)!}{n! p_1^{x_1} [1 - p_1]^{n - x_1}} \frac{n! p_1^{x_1} p_2^{x_2} [1 - (p_1 + p_2)]^{n - (x_1 + x_2)}}{x_1! x_2! [n - (x_1 + x_2)]!} \\ &= \binom{n - x_1}{x_2} \frac{p_2^{x_2}}{(1 - p_1)^{x_2}} \frac{[(1 - p_1) - p_2]^{n - x_1 - x_2}}{(1 - p_1)^{n - x_1 - x_2}} \\ &= \binom{n - x_1}{x_2} \left(\frac{p_2}{1 - p_1} \right)^{x_2} \left(1 - \frac{p_2}{1 - p_1} \right)^{n - x_1 - x_2}, \end{aligned}$$

for $0 \leq x_2 \leq n - x_1$. Note that $p_2 < 1 - p_1$. Thus, the conditional distribution of X_2 given $X_1 = x_1$ is binomial with parameters $n - x_1$ and $p_2/(1 - p_1)$.

Based on the conditional distribution of X_2 given X_1 , we have $E(X_2 | X_1) = (n - X_1)p_2/(1 - p_1)$. Let ρ_{12} be the correlation coefficient between X_1 and X_2 . Since the conditional mean is linear with slope $-p_2/(1 - p_1)$, $\sigma_2 = \sqrt{np_2(1 - p_2)}$, and $\sigma_1 = \sqrt{np_1(1 - p_1)}$, it follows from expression (2.5.4) that

$$\rho_{12} = -\frac{p_2}{1 - p_1} \frac{\sigma_1}{\sigma_2} = -\sqrt{\frac{p_1 p_2}{(1 - p_1)(1 - p_2)}}.$$

Because the support of X_1 and X_2 has the constraint $x_1 + x_2 \leq n$, the negative correlation is not surprising.

3.1.3 Hypergeometric Distribution

In Chapter 1, for a particular problem, we introduced the hypergeometric distribution; see expression (1.6.4). We now formally define it. Suppose we have a lot of N items of which D are defective. Let X denote the number of defective items in a sample of size n . If the sampling is done with replacement and the items are chosen at random, then X has a binomial distribution with parameters n and D/N . In this case the mean and variance of X are $n(D/N)$ and $n(D/N)[(N-D)/N]$, respectively. Suppose, however, that the sampling is without replacement, which is often the case in practice. The pmf of X follows by noting in this case that each of the $\binom{N}{n}$ samples are equally likely and that there are $\binom{N-D}{n-x}\binom{D}{x}$ samples that have x defective items. Hence, the pmf of X is

$$p(x) = \frac{\binom{N-D}{n-x}\binom{D}{x}}{\binom{N}{n}}, \quad x = 0, 1, \dots, n, \quad (3.1.7)$$

where, as usual, a binomial coefficient is taken to be 0 when the top value is less than the bottom value. We say that X has a **hypergeometric distribution** with parameters (N, D, n) .

The mean of X is

$$\begin{aligned} E(X) &= \sum_{x=0}^n xp(x) = \sum_{x=1}^n x \frac{\binom{N-D}{n-x}\binom{D}{x}[D(D-1)!]/[x(x-1)!(D-x)!]}{[N(N-1)!]/[(N-n)!n(n-1)!]} \\ &= n \frac{D}{N} \sum_{x=1}^n \frac{\binom{(N-1)-(D-1)}{(n-1)-(x-1)} \binom{D-1}{x-1} \binom{N-1}{n-1}^{-1}}{\binom{D-1}{x-1}} = n \frac{D}{N}. \end{aligned}$$

In the next-to-last step, we used the fact that the probabilities of a hypergeometric $(N-1, D-1, n-1)$ distribution summed over its entire range is 1. So the means for both types of sampling (with and without replacement) are the same. The variances, though, differ. As Exercise 3.1.31 shows, the variance of a hypergeometric (N, D, n) is

$$\text{Var}(X) = n \frac{D}{N} \frac{N-D}{N} \frac{N-n}{N-1}. \quad (3.1.8)$$

The last term is often thought of as the correction term when sampling without replacement. Note that it is close to 1 if N is much larger than n .

The pmf (3.1.7) can be computed in R with the code `dhyper(x, D, N-D, n)`. Suppose we draw 2 cards from a well shuffled standard deck of 52 cards and record the number of aces. The next R segment shows the probabilities over the range $\{0, 1, 2\}$ for sampling with and without replacement, respectively:

```
rng <- 0:2; dbinom(rng, 2, 1/13); dhyper(rng, 4, 48, 2)
[1] 0.85207101 0.14201183 0.00591716
[1] 0.850678733 0.144796380 0.004524887
```

Notice how close the probabilities are.

EXERCISES

3.1.1. If the mgf of a random variable X is $(\frac{1}{3} + \frac{2}{3}e^t)^5$, find $P(X = 2 \text{ or } 3)$. Verify using the R function `dbinom`.

3.1.2. The mgf of a random variable X is $(\frac{2}{3} + \frac{1}{3}e^t)^9$.

(a) Show that

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = \sum_{x=1}^5 \binom{9}{x} \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{9-x}.$$

(b) Use R to compute the probability in Part (a).

3.1.3. If X is $b(n, p)$, show that

$$E\left(\frac{X}{n}\right) = p \quad \text{and} \quad E\left[\left(\frac{X}{n} - p\right)^2\right] = \frac{p(1-p)}{n}.$$

3.1.4. Let the independent random variables $X_1, X_2, \dots, X_{40}$ be iid with the common pdf $f(x) = 3x^2$, $0 < x < 1$, zero elsewhere. Find the probability that at least 35 of the X_i 's exceed $\frac{1}{2}$.

3.1.5. Over the years, the percentage of candidates passing an entrance exam to a prestigious law school is 20%. At one of the testing centers, a group of 50 candidates take the exam and 20 pass. Is this odd? Answer on the basis that $X \geq 20$ where X is the number that pass in a group of 50 when the probability of a pass is 0.2.

3.1.6. Let Y be the number of successes throughout n independent repetitions of a random experiment with probability of success $p = \frac{1}{4}$. Determine the smallest value of n so that $P(1 \leq Y) \geq 0.70$.

3.1.7. Let the independent random variables X_1 and X_2 have binomial distribution with parameters $n_1 = 3$, $p = \frac{2}{3}$ and $n_2 = 4$, $p = \frac{1}{2}$, respectively. Compute $P(X_1 = X_2)$.

Hint: List the four mutually exclusive ways that $X_1 = X_2$ and compute the probability of each.

3.1.8. For this exercise, the reader must have access to a statistical package that obtains the binomial distribution. Hints are given for R code, but other packages can be used too.

(a) Obtain the plot of the pmf for the $b(15, 0.2)$ distribution. Using R, the following commands return the plot:

```
x<-0:15; plot(dbinom(x,15,.2)~x)
```

(b) Repeat part (a) for the binomial distributions with $n = 15$ and with $p = 0.10, 0.20, \dots, 0.90$. Comment on the shapes of the pmf's as p increases. Use the following R segment:

```
x<-0:15; par(mfrow=c(3,3)); p <- 1:9/10
for(j in p){plot(dbinom(x,15,j)~x); title(paste("p=",j))}
```

- (c) Let $Y = \frac{X}{n}$, where X has a $b(n, 0.05)$ distribution. Obtain the plots of the pmfs of Y for $n = 10, 20, 50, 200$. Comment on the plots (what do the plots seem to be converging to as n gets large?).

3.1.9. If $x = r$ is the unique mode of a distribution that is $b(n, p)$, show that

$$(n+1)p - 1 < r < (n+1)p.$$

This substantiates the comments made in Part (b) of Exercise 3.1.8.

Hint: Determine the values of x for which $p(x+1)/p(x) > 1$.

3.1.10. Suppose X is $b(n, p)$. Then by definition the pmf is symmetric if and only if $p(x) = p(n-x)$, for $x = 0, \dots, n$. Show that the pmf is symmetric if and only if $p = 1/2$.

3.1.11. Toss two nickels and three dimes at random. Make appropriate assumptions and compute the probability that there are more heads showing on the nickels than on the dimes.

3.1.12. Let $X_1, X_2, \dots, X_{k-1}$ have a multinomial distribution.

- Find the mgf of $X_2, X_3, \dots, X_{k-1}$.
- What is the pmf of $X_2, X_3, \dots, X_{k-1}$?
- Determine the conditional pmf of X_1 given that $X_2 = x_2, \dots, X_{k-1} = x_{k-1}$.
- What is the conditional expectation $E(X_1 | x_2, \dots, x_{k-1})$?

3.1.13. Let X be $b(2, p)$ and let Y be $b(4, p)$. If $P(X \geq 1) = \frac{5}{9}$, find $P(Y \geq 1)$.

3.1.14. Let X have a binomial distribution with parameters n and $p = \frac{1}{3}$. Determine the smallest integer n can be such that $P(X \geq 1) \geq 0.85$.

3.1.15. Let X have the pmf $p(x) = (\frac{1}{3})(\frac{2}{3})^x$, $x = 0, 1, 2, 3, \dots$, zero elsewhere. Find the conditional pmf of X given that $X \geq 3$.

3.1.16. One of the numbers $1, 2, \dots, 6$ is to be chosen by casting an unbiased die. Let this random experiment be repeated five independent times. Let the random variable X_1 be the number of terminations in the set $\{x : x = 1, 2, 3\}$ and let the random variable X_2 be the number of terminations in the set $\{x : x = 4, 5\}$. Compute $P(X_1 = 2, X_2 = 1)$.

3.1.17. Show that the moment generating function of the negative binomial distribution is $M(t) = p^r[1 - (1-p)e^t]^{-r}$. Find the mean and the variance of this distribution.

Hint: In the summation representing $M(t)$, make use of the negative binomial series.¹

¹See, for example, *Mathematical Comments* referenced in the Preface.

3.1.18. One way of estimating the number of fish in a lake is the following **capture-recapture** sampling scheme. Suppose there are N fish in the lake where N is unknown. A specified number of fish T are captured, tagged, and released back to the lake. Then at a specified time and for a specified positive integer r , fish are captured until the r th tagged fish is caught. The random variable of interest is Y the number of nontagged fish caught.

- (a) What is the distribution of Y ? Identify all parameters.
- (b) What is $E(Y)$ and the $\text{Var}(Y)$?
- (c) The method of moment estimate of N is to set Y equal to the expression for $E(Y)$ and solve this equation for N . Call the solution $\hat{N}$. Determine $\hat{N}$.
- (d) Determine the mean and variance of $\hat{N}$.

3.1.19. Consider a multinomial trial with outcomes $1, 2, \dots, k$ and respective probabilities $p_1, p_2, \dots, p_k$. Let **ps** denote the R vector for $(p_1, p_2, \dots, p_k)$. Then a single random trial of this multinomial is computed with the command `multitrial(ps)`, where the required R functions are:²

```
psum <- function(v){
  p<-0; psum <- c()
  for(j in 1:length(v)){p<-p+v[j]; psum <- c(psum,p)}
  return(psum)}
multitrial <- function(p){
  pr <- c(0,psum(p))
  r <- runif(1); ic <- 0; j <- 1
  while(ic==0){if((r > pr[j]) && (r <= pr[j+1]))
    {multitrial <-j; ic<-1}; j<- j+1}
  return(multitrial)}
```

- (a) Compute 10 random trials if `ps=c(.3,.2,.2,.2,.1)`.
- (b) Compute 10,000 random trials for **ps** as in (a). Check to see how close the estimates of p_i are with p_i .

3.1.20. Using the experiment in part (a) of Exercise 3.1.19, consider a game when a person pays \$5 to play. If the trial results in a 1 or 2, she receives nothing; if a 3, she receives \$1; if a 4, she receives \$2; and if a 5, she receives \$20. Let G be her gain.

- (a) Determine $E(G)$.
- (b) Write R code that simulates the gain. Then simulate it 10,000 times, collecting the gains. Compute the average of these 10,000 gains and compare it with $E(G)$.

²Downloadable at the site listed in the Preface

3.1.21. Let X_1 and X_2 have a trinomial distribution. Differentiate the moment-generating function to show that their covariance is $-np_1p_2$.

3.1.22. If a fair coin is tossed at random five independent times, find the conditional probability of five heads given that there are at least four heads.

3.1.23. Let an unbiased die be cast at random seven independent times. Compute the conditional probability that each side appears at least once given that side 1 appears exactly twice.

3.1.24. Compute the measures of skewness and kurtosis of the binomial distribution $b(n, p)$.

3.1.25. Let

$$p(x_1, x_2) = \binom{x_1}{x_2} \left(\frac{1}{2}\right)^{x_1} \left(\frac{x_1}{15}\right), \quad \begin{matrix} x_2 = 0, 1, \dots, x_1 \\ x_1 = 1, 2, 3, 4, 5, \end{matrix}$$

zero elsewhere, be the joint pmf of X_1 and X_2 . Determine

(a) $E(X_2)$.

(b) $u(x_1) = E(X_2|x_1)$.

(c) $E[u(X_1)]$.

Compare the answers of parts (a) and (c).

Hint: Note that $E(X_2) = \sum_{x_1=1}^5 \sum_{x_2=0}^{x_1} x_2 p(x_1, x_2)$.

3.1.26. Three fair dice are cast. In 10 independent casts, let X be the number of times all three faces are alike and let Y be the number of times only two faces are alike. Find the joint pmf of X and Y and compute $E(6XY)$.

3.1.27. Let X have a geometric distribution. Show that

$$P(X \geq k + j | X \geq k) = P(X \geq j), \quad (3.1.9)$$

where k and j are nonnegative integers. Note that we sometimes say in this situation that X is **memoryless**.

3.1.28. Let X equal the number of independent tosses of a fair coin that are required to observe heads on consecutive tosses. Let u_n equal the n th Fibonacci number, where $u_1 = u_2 = 1$ and $u_n = u_{n-1} + u_{n-2}$, $n = 3, 4, 5, \dots$

(a) Show that the pmf of X is

$$p(x) = \frac{u_{x-1}}{2^x}, \quad x = 2, 3, 4, \dots$$

(b) Use the fact that

$$u_n = \frac{1}{\sqrt{5}} \left[\left(\frac{1 + \sqrt{5}}{2} \right)^n - \left(\frac{1 - \sqrt{5}}{2} \right)^n \right]$$

to show that $\sum_{x=2}^{\infty} p(x) = 1$.

3.1.29. Let the independent random variables X_1 and X_2 have binomial distributions with parameters n_1 , $p_1 = \frac{1}{2}$ and n_2 , $p_2 = \frac{1}{2}$, respectively. Show that $Y = X_1 - X_2 + n_2$ has a binomial distribution with parameters $n = n_1 + n_2$, $p = \frac{1}{2}$.

3.1.30. Consider a shipment of 1000 items into a factory. Suppose the factory can tolerate about 5% defective items. Let X be the number of defective items in a sample without replacement of size $n = 10$. Suppose the factory returns the shipment if $X \geq 2$.

- (a) Obtain the probability that the factory returns a shipment of items that has 5% defective items.
- (b) Suppose the shipment has 10% defective items. Obtain the probability that the factory returns such a shipment.
- (c) Obtain approximations to the probabilities in parts (a) and (b) using appropriate binomial distributions.

Note: If you do not have access to a computer package with a hypergeometric command, obtain the answer to (c) only. This is what would have been done in practice 20 years ago. If you have access to R, then the command `dhyper(x,D,N-D,n)` returns the probability in expression (3.1.7).

3.1.31. Show that the variance of a hypergeometric (N, D, n) distribution is given by expression (3.1.8).

Hint: First obtain $E[X(X-1)]$ by proceeding in the same way as the derivation of the mean given in Section 3.1.3.

3.2 The Poisson Distribution

Recall that the following series expansion³ holds for all real numbers z :

$$1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots = \sum_{x=0}^{\infty} \frac{z^x}{x!} = e^z.$$

Consider the function $p(x)$ defined by

$$p(x) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!} & x = 0, 1, 2, \dots \\ 0 & \text{elsewhere,} \end{cases} \quad (3.2.1)$$

where $\lambda > 0$. Since $\lambda > 0$, then $p(x) \geq 0$ and

$$\sum_{x=0}^{\infty} p(x) = \sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1;$$

³See, for example, the discussion on Taylor series in *Mathematical Comments* referenced in the Preface.

that is, $p(x)$ satisfies the conditions of being a pmf of a discrete type of random variable. A random variable that has a pmf of the form $p(x)$ is said to have a **Poisson distribution** with parameter λ , and any such $p(x)$ is called a **Poisson pmf** with parameter λ .

As the following remark shows, Poisson distributions occur in many areas of applications.

Remark 3.2.1. Consider a process that counts the number of certain events occurring over an interval of time; for example, the number of tornados that touch down in Michigan per year, the number of cars entering a parking lot between 8:00 and 12:00 on a weekday, the number of car accidents at a busy intersection per week, the number of typographical errors per page of a manuscript, and the number of blemishes on a manufactured car door. As in the third and fourth examples, the occurrences need not be over time. It is convenient, though, to use the time representation in the following derivation. Let X_t denote the number of occurrences of such a process over the interval $(0, t]$. The range of X_t is the set of nonnegative integers $\{0, 1, 2, \dots\}$. For a nonnegative integer k and a real number $t > 0$, denote the pmf of X_t by $P(X_t = k) = g(k, t)$. Under the following three axioms, we next show that X_t has a Poisson distribution.

1. $g(1, h) = \lambda h + o(h)$, for a constant $\lambda > 0$.
2. $\sum_{t=2}^{\infty} g(t, h) = o(h)$.
3. The number of occurrences in nonoverlapping intervals are independent of one another.

Here the $o(h)$ notation means that $o(h)/h \rightarrow 0$ as $h \rightarrow 0$. For instance, $h^2 = o(h)$ and $o(h) + o(h) = o(h)$. Note that the first two axioms imply that in a small interval of time h , either one or no events occur and that the probability of one event occurring is proportional to h .

By the method of induction, we now show that the distribution of X_t is Poisson with parameter λt . First, we obtain $g(k, t)$ for $k = 0$. Note that the boundary condition $g(0, 0) = 1$ is reasonable. No events occur in time $(0, t + h]$ if and only if no events occur in $(0, t]$ and no events occur in $(t, t + h]$. By Axioms (1) and (2), the probability that no events occur in the interval $(0, h]$ is $1 - \lambda h + o(h)$. Further, the intervals $(0, t]$ and $(t, t + h]$ do not overlap. Hence, by Axiom (3) we have

$$g(0, t + h) = g(0, t)[1 - \lambda h + o(h)]. \quad (3.2.2)$$

That is,

$$\frac{g(0, t + h) - g(0, t)}{h} = -\lambda g(0, t) + \frac{g(0, t)o(h)}{h} \rightarrow -\lambda g(0, t), \quad \text{as } h \rightarrow 0.$$

Thus, $g(0, t)$ satisfies the differential equation

$$\frac{d_t g(0, t)}{g(0, t)} = -\lambda$$

Integrating both side with respect to t , we have for some constant c that

$$\log g(0, t) = -\lambda t + c \quad \text{or} \quad g(0, t) = e^{-\lambda t} e^c.$$

Finally, using the boundary condition $g(0, 0) = 1$, we have $e^c = 1$. Hence,

$$g(0, t) = e^{-\lambda t}. \quad (3.2.3)$$

So the result holds for $k = 0$.

For the remainder of the proof, assume that, for k a nonnegative integer, $g(k, t) = e^{-\lambda t} (\lambda t)^k / k!$. By induction, the proof follows if we can show that the result holds for $g(k+1, t)$. Another reasonable boundary condition is $g(k+1, 0) = 0$. Consider $g(k+1, t+h)$. In order to have $k+1$ occurrences in $(0, t+h]$ either there are $k+1$ occurrences in $(0, t]$ and no occurrences in $(t, t+h]$ or there are k occurrences in $(0, t]$ and one occurrence in $(t, t+h]$. Because these events are disjoint we have by the independence of Axiom 3 that

$$g(k+1, t+h) = g(k+1, t)[1 - \lambda h + o(h)] + g(k, t)[\lambda h + o(h)],$$

that is,

$$\frac{g(k+1, t+h) - g(k+1, t)}{h} = -\lambda g(k+1, t) + g(k, t)\lambda + [g(k+1, t) + g(k, t)] \frac{o(h)}{h}.$$

Letting $h \rightarrow 0$ and using the value of $g(k, t)$, we obtain the differential equation

$$\frac{d}{dt} g(k+1, t) = -\lambda g(k+1, t) + \lambda e^{-\lambda t} [(\lambda t)^k / k!].$$

This is a linear differential equation of first order. Appealing to a theorem in differential equations, its solution is

$$e^{\int \lambda dt} g(k+1, t) = \int e^{\int \lambda dt} \lambda e^{-\lambda t} [(\lambda t)^k / k!] dt + c.$$

Using the boundary condition $g(k+1, 0) = 0$ and carrying out the integration, we obtain

$$g(k+1, t) = e^{-\lambda t} [(\lambda t)^{k+1} / (k+1)!]$$

Therefore, X_t has a Poisson distribution with parameter λt . ■

Let X have a Poisson distribution with parameter λ . The mgf of X is given by

$$\begin{aligned} M(t) &= \sum_{x=0}^{\infty} e^{tx} p(x) = \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x e^{-\lambda}}{x!} \\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} \\ &= e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)} \end{aligned}$$

for all real values of t . Since

$$M'(t) = e^{\lambda(e^t-1)}(\lambda e^t)$$

and

$$M''(t) = e^{\lambda(e^t-1)}\lambda e^t \lambda e^t + e^{\lambda(e^t-1)}\lambda e^t$$

then

$$\mu = M'(0) = \lambda$$

and

$$\sigma^2 = M''(0) - \mu^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

That is, a Poisson distribution has $\mu = \sigma^2 = \lambda > 0$.

If X has a Poisson distribution with parameter λ , then $P(X = k)$ is computed by the R command `dpois(k,lambda)` and the cumulative probability $P(X \leq k)$ is calculated by `ppois(k,lambda)`.

Example 3.2.1. Let X be the number of automobile accidents at a busy intersection per week. Suppose that X has a Poisson distribution with $\lambda = 2$. Then the expected number of accidents per week is 2 and the standard deviation of the number of accidents is $\sqrt{2}$. The probability of at least one accident in a week is

$$P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-2} = 1 - \text{dpois}(0,2) = 0.8647$$

and the probability that there are between 3 and 8 (inclusive) accidents is

$$P(3 \leq X \leq 8) = P(X \leq 8) - P(X \leq 2) = \text{ppois}(8,2) - \text{ppois}(2,2) = 0.3231.$$

Suppose we want to determine the probability that there are exactly 16 accidents in a 4 week period. By Remark 3.2.1, the number of accidents over a 4 week period has a Poisson distribution with parameter $2 \times 4 = 8$. So the desired probability is `dpois(16,8) = 0.0045`. The following R code computes a spiked plot of the pmf of X over $\{0, 1, \dots, 7\}$, a subset of the range of X .

```
rng=0:7; y=dpois(rng,2); plot(y~rng,type="h",ylab="pmf",xlab="Rng");
points(y~rng,pch=16,cex=2)
```

■

Example 3.2.2. Let the probability of exactly one blemish in 1 foot of wire be about $\frac{1}{1000}$ and let the probability of two or more blemishes in that length be, for all practical purposes, zero. Let the random variable X be the number of blemishes in 3000 feet of wire. If we assume the independence of the number of blemishes in nonoverlapping intervals, then by Remark 3.2.1 the postulates of the Poisson process are approximated, with $\lambda = \frac{1}{1000}$ and $t = 3000$. Thus X has an approximate Poisson distribution with mean $3000(\frac{1}{1000}) = 3$. For example, the probability that there are five or more blemishes in 3000 feet of wire is

$$P(X \geq 5) = \sum_{k=5}^{\infty} \frac{3^k e^{-3}}{k!} = 1 - \text{ppois}(4,3) = 0.1847. \quad \blacksquare$$

The Poisson distribution satisfies the following important additive property.

Theorem 3.2.1. *Suppose $X_1, \dots, X_n$ are independent random variables and suppose X_i has a Poisson distribution with parameter λ_i . Then $Y = \sum_{i=1}^n X_i$ has a Poisson distribution with parameter $\sum_{i=1}^n \lambda_i$.*

Proof: We obtain the result by determining the mgf of Y , which by Theorem 2.6.1 is given by

$$M_Y(t) = E(e^{tY}) = \prod_{i=1}^n e^{\lambda_i(e^t - 1)} = e^{\sum_{i=1}^n \lambda_i(e^t - 1)}.$$

By the uniqueness of mgfs, we conclude that Y has a Poisson distribution with parameter $\sum_{i=1}^n \lambda_i$. ■

Example 3.2.3 (Example 3.2.2, Continued). Suppose, as in Example 3.2.2, that a bail of wire consists of 3000 feet. Based on the information in the example, we expect three blemishes in a bail of wire, and the probability of five or more blemishes is 0.1847. Suppose in a sampling plan, three bails of wire are selected at random and we compute the mean number of blemishes in the wire. Now suppose we want to determine the probability that the mean of the three observations has five or more blemishes. Let X_i be the number of blemishes in the i th bail of wire for $i = 1, 2, 3$. Then X_i has a Poisson distribution with parameter 3. The mean of X_1, X_2 , and X_3 is $\bar{X} = 3^{-1} \sum_{i=1}^3 X_i$, which can also be expressed as $Y/3$, where $Y = \sum_{i=1}^3 X_i$. By the last theorem, because the bails are independent of one another, Y has a Poisson distribution with parameter $\sum_{i=1}^3 3 = 9$. Hence, the desired probability is

$$P(\bar{X} \geq 5) = P(Y \geq 15) = 1 - \text{ppois}(14, 9) = 0.0415.$$

Hence, while it is not too odd that a bail has five or more blemishes (probability is 0.1847), it is unusual (probability is 0.0415) that three independent bails of wire average five or more blemishes. ■

EXERCISES

3.2.1. If the random variable X has a Poisson distribution such that $P(X = 1) = P(X = 2)$, find $P(X = 4)$.

3.2.2. The mgf of a random variable X is $e^{4(e^t - 1)}$. Show that $P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.931$.

3.2.3. In a lengthy manuscript, it is discovered that only 13.5 percent of the pages contain no typing errors. If we assume that the number of errors per page is a random variable with a Poisson distribution, find the percentage of pages that have exactly one error.

3.2.4. Let the pmf $p(x)$ be positive on and only on the nonnegative integers. Given that $p(x) = (4/x)p(x-1)$, $x = 1, 2, 3, \dots$, find the formula for $p(x)$.

Hint: Note that $p(1) = 4p(0)$, $p(2) = (4^2/2!)p(0)$, and so on. That is, find each $p(x)$ in terms of $p(0)$ and then determine $p(0)$ from

$$1 = p(0) + p(1) + p(2) + \dots$$

3.2.5. Let X have a Poisson distribution with $\mu = 100$. Use Chebyshev's inequality to determine a lower bound for $P(75 < X < 125)$. Next, calculate the probability using R. Is the approximation by Chebyshev's inequality accurate?

3.2.6. The following R code segment computes a page of plots for Poisson pmfs with means 2, 4, 6, ..., 18. Run this code and comment on the the shapes and modes of the distributions.

```
par(mfrow=c(3,3)); x= 0:35; lam=seq(2,18,2);
for(y in lam){plot(dpois(x,y)~x); title(paste("Mean is ",y))}
```

3.2.7. By Exercise 3.2.6 it seems that the Poisson pmf peaks at its mean λ . Show that this is the case by solving the inequalities $[p(x+1)/p(x)] > 1$ and $[p(x+1)/p(x)] < 1$, where $p(x)$ is the pmf of a Poisson distribution with parameter λ .

3.2.8. Using the computer, obtain an overlay plot of the pmfs of the following two distributions:

(a) Poisson distribution with $\lambda = 2$.

(b) Binomial distribution with $n = 100$ and $p = 0.02$.

Why would these distributions be approximately the same? Discuss.

3.2.9. Continuing with Exercise 3.2.8, make a page of four overlay plots for the following 4 Poisson and binomial combinations: $\lambda = 2, p = 0.02$; $\lambda = 10, p = 0.10$; $\lambda = 30, p = 0.30$; $\lambda = 50, p = 0.50$. Use $n = 100$ in each situation. Plot the subset of the binomial range that is between $np \pm \sqrt{np(1-p)}$. For each situation, comment on the goodness of the Poisson approximation to the binomial.

3.2.10. The approximation discussed in Exercise 3.2.8 can be made precise in the following way. Suppose X_n is binomial with the parameters n and $p = \lambda/n$, for a given $\lambda > 0$. Let Y be Poisson with mean λ . Show that $P(X_n = k) \rightarrow P(Y = k)$, as $n \rightarrow \infty$, for an arbitrary but fixed value of k .

Hint: First show that:

$$P(X_n = k) = \frac{\lambda^k}{k!} \left[\frac{n(n-1) \cdots (n-k+1)}{n^k} \left(1 - \frac{\lambda}{n}\right)^{-k} \right] \left(1 - \frac{\lambda}{n}\right)^n.$$

3.2.11. Let the number of chocolate chips in a certain type of cookie have a Poisson distribution. We want the probability that a cookie of this type contains at least two chocolate chips to be greater than 0.99. Find the smallest value of the mean that the distribution can take.

3.2.12. Compute the measures of skewness and kurtosis of the Poisson distribution with mean μ .

3.2.13. On the average, a grocer sells three of a certain article per week. How many of these should he have in stock so that the chance of his running out within a week is less than 0.01? Assume a Poisson distribution.

3.2.14. Let X have a Poisson distribution. If $P(X = 1) = P(X = 3)$, find the mode of the distribution.

3.2.15. Let X have a Poisson distribution with mean 1. Compute, if it exists, the expected value $E(X!)$.

3.2.16. Let X and Y have the joint pmf $p(x, y) = e^{-2}/[x!(y-x)!]$, $y = 0, 1, 2, \dots$, $x = 0, 1, \dots, y$, zero elsewhere.

- (a) Find the mgf $M(t_1, t_2)$ of this joint distribution.
- (b) Compute the means, the variances, and the correlation coefficient of X and Y .
- (c) Determine the conditional mean $E(X|y)$.

Hint: Note that

$$\sum_{x=0}^y [\exp(t_1 x)] y! / [x!(y-x)!] = [1 + \exp(t_1)]^y.$$

Why?

3.2.17. Let X_1 and X_2 be two independent random variables. Suppose that X_1 and $Y = X_1 + X_2$ have Poisson distributions with means μ_1 and $\mu > \mu_1$, respectively. Find the distribution of X_2 .

3.3 The Γ , χ^2 , and β Distributions

In this section we introduce the continuous gamma Γ -distribution and several associated distributions. The support for the Γ -distribution is the set of positive real numbers. This distribution and its associated distributions are rich in applications in all areas of science and business. These applications include their use in modeling lifetimes, failure times, service times, and waiting times.

The definition of the Γ -distribution requires the Γ function from calculus. It is proved in calculus that the integral

$$\int_0^{\infty} y^{\alpha-1} e^{-y} dy$$

exists for $\alpha > 0$ and that the value of the integral is a positive number. The integral is called the **gamma function** of α , and we write

$$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y} dy.$$

If $\alpha = 1$, clearly

$$\Gamma(1) = \int_0^{\infty} e^{-y} dy = 1.$$

If $\alpha > 1$, an integration by parts shows that

$$\Gamma(\alpha) = (\alpha - 1) \int_0^\infty y^{\alpha-2} e^{-y} dy = (\alpha - 1)\Gamma(\alpha - 1). \quad (3.3.1)$$

Accordingly, if α is a positive integer greater than 1,

$$\Gamma(\alpha) = (\alpha - 1)(\alpha - 2) \cdots (3)(2)(1)\Gamma(1) = (\alpha - 1)!.$$

Since $\Gamma(1) = 1$, this suggests we take $0! = 1$, as we have done. The Γ function is sometimes called the factorial function.

We say that the continuous random variable X has a Γ -distribution with parameters $\alpha > 0$ and $\beta > 0$, if its pdf is

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} & 0 < x < \infty \\ 0 & \text{elsewhere.} \end{cases} \quad (3.3.2)$$

In which case, we often write that X has $\Gamma(\alpha, \beta)$ distribution.

To verify that $f(x)$ is a pdf, note first that $f(x) > 0$, for all $x > 0$. To show that it integrates to 1 over its support, we use the change-of-variable $z = x/\beta$, $dz = (1/\beta)dx$ in the following derivation:

$$\begin{aligned} \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty (\beta z)^{\alpha-1} e^{-z} \beta dz \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \beta^\alpha \Gamma(\alpha) = 1; \end{aligned}$$

hence, $f(x)$ is a pdf. This change-of-variable used is worth remembering. We use a similar change-of-variable in the following derivation of X 's mgf:

$$\begin{aligned} M(t) &= \int_0^\infty e^{tx} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx \\ &= \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x(1-\beta t)/\beta} dx. \end{aligned}$$

Next, use the change-of-variable $y = x(1 - \beta t)/\beta$, $t < 1/\beta$, or $x = \beta y/(1 - \beta t)$, to obtain

$$M(t) = \int_0^\infty \frac{\beta/(1 - \beta t)}{\Gamma(\alpha)\beta^\alpha} \left(\frac{\beta y}{1 - \beta t} \right)^{\alpha-1} e^{-y} dy.$$

That is,

$$\begin{aligned} M(t) &= \left(\frac{1}{1 - \beta t} \right)^\alpha \int_0^\infty \frac{1}{\Gamma(\alpha)} y^{\alpha-1} e^{-y} dy \\ &= \frac{1}{(1 - \beta t)^\alpha}, \quad t < \frac{1}{\beta}. \end{aligned}$$

Now

$$M'(t) = (-\alpha)(1 - \beta t)^{-\alpha-1}(-\beta)$$

and

$$M''(t) = (-\alpha)(-\alpha - 1)(1 - \beta t)^{-\alpha-2}(-\beta)^2.$$

Hence, for a gamma distribution, we have

$$\mu = M'(0) = \alpha\beta$$

and

$$\sigma^2 = M''(0) - \mu^2 = \alpha(\alpha + 1)\beta^2 - \alpha^2\beta^2 = \alpha\beta^2.$$

Suppose X has a $\Gamma(\alpha, \beta)$ distribution. To calculate probabilities for this distribution in R, let $a = \alpha$ and $b = \beta$. Then the command `pgamma(x, shape=a, scale=b)` returns $P(X \leq x)$, while the value of the pdf of X at x is returned by the command `dgamma(x, shape=a, scale=b)`.

Example 3.3.1. Let X be the lifetime in hours of a certain battery used under extremely cold conditions. Suppose X has a $\Gamma(5, 4)$ distribution. Then the mean lifetime of the battery is 20 hours with standard deviation $\sqrt{5 \times 16} = 8.94$ hours. The probability that battery lasts at least 50 hours is `1-pgamma(50, shape=5, scale=4) = 0.0053`. The median lifetime of the battery is `qgamma(.5, shape=5, scale=4) = 18.68` hours. The probability that the lifetime is within one standard deviation of its mean lifetime is

`pgamma(20+8.94, shape=5, scale=4)-pgamma(20-8.94, shape=5, scale=4)=.700`.

Finally, this line of R code presents a plot of the pdf:

`x=seq(.1, 50, .1); plot(dgamma(x, shape=5, scale=4)~x)`.

On this plot, the reader should locate the above probabilities and the mean and median lifetimes of the battery. ■

The main reason for the appeal of the Γ -distribution in applications is the variety of shapes of the distribution for different values of α and β . This is apparent in Figure 3.3.1 which depicts six Γ -pdfs.⁴

Suppose X denotes the failure time of a device with pdf $f(x)$ and cdf $F(x)$. In practice, the pdf of X is often unknown. If a large sample of failure times of these devices is at hand then estimates of the pdf can be obtained as discussed in Chapter 4. Another function that helps in identifying the pdf of X is the *hazard function* of X . Let x be in the support of X . Suppose the device has not failed at time x , i.e., $X > x$. What is the probability that the device fails in the next instance? We answer this question in terms of the rate of failure at x , which is:

$$\begin{aligned} r(x) &= \lim_{\Delta \rightarrow 0} \frac{P(x \leq X < x + \Delta | X \geq x)}{\Delta} = \frac{1}{1 - F(x)} \lim_{\Delta \rightarrow 0} \frac{P(x \leq X < x + \Delta)}{\Delta} \\ &= \frac{f(x)}{1 - F(x)}. \end{aligned} \quad (3.3.3)$$

The rate of failure at time x , $r(x)$, is defined as the **hazard function** of X at x .

⁴The R function for these plots is `newfigc3s3.1.R`, at the site listed in the Preface.

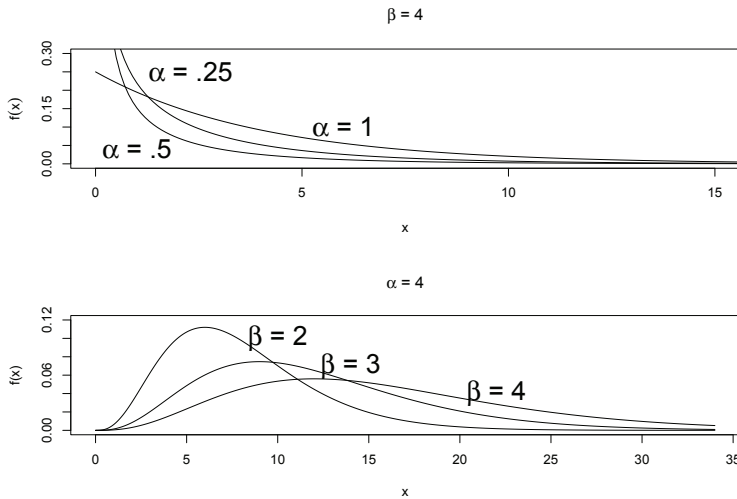


Figure 3.3.1: Several gamma densities

Note that the hazard function $r(x)$ satisfies $-(d/dx) \log[1 - F(x)]$; that is,

$$1 - F(x) = e^{-\int r(x) dx + c}, \quad (3.3.4)$$

for a constant c . When the support of X is $(0, \infty)$, $F(0) = 0$ serves as a boundary condition to solve for c . In practice, often the scientist can describe the hazard rate and, hence, $F(x)$ can be determined from expression (3.3.4). For example, suppose the hazard rate of X is constant; i.e., $r(x) = 1/\beta$ for some $\beta > 0$. Then

$$1 - F(x) = e^{-\int (1/\beta) dx + c} = e^{-x/\beta} e^c.$$

Since $F(0) = 0$, $e^c = 1$. So the pdf of X is

$$f(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta} & x > 0 \\ 0 & \text{elsewhere.} \end{cases} \quad (3.3.5)$$

Of course, this is a $\Gamma(1, \beta)$ distribution, but it is also called the **exponential distribution** with parameter $1/\beta$. An important property of this distribution is given in Exercise 3.3.25.

Using R, hazard functions can be quickly plotted. Here is the code for an overlay plot of the hazard functions of the exponential distribution with $\beta = 8$ and the $\Gamma(4, 2)$ -distribution.

```
x=seq(.1,15,.1); t=dgamma(x,shape=4,scale=2)
b=(1-pgamma(x,shape=4,scale=2));y1=t/b;plot(y1~x);abline(h=1/8)
```

Note that the hazard function of this Γ -distribution is an increasing function of x ; i.e., the rate of failure increases as time progresses. Other examples of hazard functions are given in Exercise 3.3.26.

One of the most important properties of the gamma distribution is its additive property.

Theorem 3.3.1. *Let $X_1, \dots, X_n$ be independent random variables. Suppose, for $i = 1, \dots, n$, that X_i has a $\Gamma(\alpha_i, \beta)$ distribution. Let $Y = \sum_{i=1}^n X_i$. Then Y has a $\Gamma(\sum_{i=1}^n \alpha_i, \beta)$ distribution.*

Proof: Using the assumed independence and the mgf of a gamma distribution, we have by Theorem 2.6.1 that for $t < 1/\beta$,

$$M_Y(t) = \prod_{i=1}^n (1 - \beta t)^{-\alpha_i} = (1 - \beta t)^{-\sum_{i=1}^n \alpha_i},$$

which is the mgf of a $\Gamma(\sum_{i=1}^n \alpha_i, \beta)$ distribution. ■

Γ -distributions naturally occur in the Poisson process, also.

Remark 3.3.1 (Poisson Processes). For $t > 0$, let X_t denote the number of events of interest that occur in the interval $(0, t]$. Assume X_t satisfies the three assumptions of a Poisson process. Let k be a fixed positive integer and define the continuous random variable W_k to be the waiting time until the k th event occurs. Then the range of W_k is $(0, \infty)$. Note that for $w > 0$, $W_k > w$ if and only if $X_w \leq k - 1$. Hence,

$$P(W_k > w) = P(X_w \leq k - 1) = \sum_{x=0}^{k-1} P(X_w = x) = \sum_{x=0}^{k-1} \frac{(\lambda w)^x e^{-\lambda w}}{x!}.$$

In Exercise 3.3.5, the reader is asked to prove that

$$\int_{\lambda w}^{\infty} \frac{z^{k-1} e^{-z}}{(k-1)!} dz = \sum_{x=0}^{k-1} \frac{(\lambda w)^x e^{-\lambda w}}{x!}.$$

Accepting this result, we have, for $w > 0$, that the cdf of W_k satisfies

$$F_{W_k}(w) = 1 - \int_{\lambda w}^{\infty} \frac{z^{k-1} e^{-z}}{\Gamma(k)} dz = \int_0^{\lambda w} \frac{z^{k-1} e^{-z}}{\Gamma(k)} dz,$$

and for $w \leq 0$, $F_{W_k}(w) = 0$. If we change the variable of integration in the integral that defines $F_{W_k}(w)$ by writing $z = \lambda y$, then

$$F_{W_k}(w) = \int_0^w \frac{\lambda^k y^{k-1} e^{-\lambda y}}{\Gamma(k)} dy, \quad w > 0,$$

and $F_{W_k}(w) = 0$ for $w \leq 0$. Accordingly, the pdf of W_k is

$$f_{W_k}(w) = F'_{W_k}(w) = \begin{cases} \frac{\lambda^k w^{k-1} e^{-\lambda w}}{\Gamma(k)} & 0 < w < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

That is, the waiting time until the k th event, W_k , has the gamma distribution with $\alpha = k$ and $\beta = 1/\lambda$. Let T_1 be the waiting time until the first event occurs, i.e., $k = 1$. Then the pdf of T_1 is

$$f_{T_1}(w) = \begin{cases} \lambda e^{-\lambda w} & 0 < w < \infty \\ 0 & \text{elsewhere.} \end{cases} \quad (3.3.6)$$

Hence, T_1 has the $\Gamma(1, 1/\lambda)$ -distribution. The mean of $T_1 = 1/\lambda$, while the mean of X_1 is λ . Thus, we expect λ events to occur in a unit of time and we expect the first event to occur at time $1/\lambda$.

Continuing in this way, for $i \geq 1$, let T_i denote the interarrival time of the i th event; i.e., T_i is the time between the occurrence of event $(i - 1)$ and event i . As shown T_1 has the $\Gamma(1, 1/\lambda)$. Note that Axioms (1) and (2) of the Poisson process only depend on λ and the length of the interval; in particular, they do not depend on the endpoints of the interval. Further, occurrences in nonoverlapping intervals are independent of one another. Hence, using the same reasoning as above, T_j , $j \geq 2$, also has the $\Gamma(1, 1/\lambda)$ -distribution. Furthermore, $T_1, T_2, T_3, \dots$ are independent. Note the waiting time until the k th event satisfies $W_k = T_1 + \dots + T_k$. Thus by Theorem 3.3.1, W_k has a $\Gamma(k, 1/\lambda)$ distribution, confirming the derivation above. Although this discussion has been intuitive, it can be made rigorous; see, for example, Parzen (1962). ■

3.3.1 The χ^2 -Distribution

Let us now consider a special case of the gamma distribution in which $\alpha = r/2$, where r is a positive integer, and $\beta = 2$. A random variable X of the continuous type that has the pdf

$$f(x) = \begin{cases} \frac{1}{\Gamma(r/2)2^{r/2}} x^{r/2-1} e^{-x/2} & 0 < x < \infty \\ 0 & \text{elsewhere,} \end{cases} \quad (3.3.7)$$

and the mgf

$$M(t) = (1 - 2t)^{-r/2}, \quad t < \frac{1}{2},$$

is said to have a **chi-square distribution** (χ^2 -distribution), and any $f(x)$ of this form is called a **chi-square pdf**. The mean and the variance of a chi-square distribution are $\mu = \alpha\beta = (r/2)2 = r$ and $\sigma^2 = \alpha\beta^2 = (r/2)2^2 = 2r$, respectively. We call the parameter r the number of degrees of freedom of the chi-square distribution (or of the chi-square pdf). Because the chi-square distribution has an important role in statistics and occurs so frequently, we write, for brevity, that X is $\chi^2(r)$ to mean that the random variable X has a chi-square distribution with r degrees of freedom. The R function `pchisq(x,r)` returns $P(X \leq x)$ and the command `dchisq(x,r)` returns the value of the pdf of X at x when X has a chi-squared distribution with r degrees of freedom.

Example 3.3.2. Suppose X has a χ^2 -distribution with 10 degrees of freedom. Then the mean of X is 10 and its standard deviation is $\sqrt{20} = 4.47$. Using R, its median and quartiles are `qchisq(c(.25,.5,.75),10) = (6.74, 9.34, 12.55)`. The following

command plots the density function over the interval (0, 24):

```
x=seq(0,24,.1);plot(dchisq(x,10)~x).
```

Compute this line of code and locate the mean, quartiles, and median of X on the plot. ■

Example 3.3.3. The quantiles of the χ^2 -distribution are frequently used in statistics. Before the advent of modern computation, tables of these quantiles were compiled. Table I in Appendix D offers a typical χ^2 -table of the quantiles for the probabilities 0.01, 0.025, 0.05, 0.1, 0.9, 0.95, 0.975, 0.99 and degrees of freedom 1, 2, ..., 30. As discussed, the R function `qchisq` easily computes these quantiles. Actually, the following two lines of R code performs the computation of Table I.

```
rs=1:30; ps=c(.01,.025,.05,.1,.9,.95,.975,.99);
for(r in rs){print(c(r,round(qchisq(ps,r),digits=3)))}
```

Note that the code rounds the critical values to 3 places. ■

The following result is used several times in the sequel; hence, we record it as a theorem.

Theorem 3.3.2. *Let X have a $\chi^2(r)$ distribution. If $k > -r/2$, then $E(X^k)$ exists and it is given by*

$$E(X^k) = \frac{2^k \Gamma\left(\frac{r}{2} + k\right)}{\Gamma\left(\frac{r}{2}\right)}, \quad \text{if } k > -r/2. \quad (3.3.8)$$

Proof: Note that

$$E(X^k) = \int_0^\infty \frac{1}{\Gamma\left(\frac{r}{2}\right) 2^{r/2}} x^{(r/2)+k-1} e^{-x/2} dx.$$

Make the change of variable $u = x/2$ in the above integral. This results in

$$E(X^k) = \int_0^\infty \frac{1}{\Gamma\left(\frac{r}{2}\right) 2^{(r/2)-1}} 2^{(r/2)+k-1} u^{(r/2)+k-1} e^{-u} du.$$

This simplifies to the desired result provided that $k > -(r/2)$. ■

Notice that if k is a nonnegative integer, then $k > -(r/2)$ is always true. Hence, all moments of a χ^2 distribution exist and the k th moment is given by (3.3.8).

Example 3.3.4. Let X have a gamma distribution with $\alpha = r/2$, where r is a positive integer, and $\beta > 0$. Define the random variable $Y = 2X/\beta$. We seek the pdf of Y . Now the mgf of Y is

$$\begin{aligned} M_Y(t) &= E(e^{tY}) = E\left[e^{(2t/\beta)X}\right] \\ &= \left[1 - \frac{2t}{\beta}\right]^{-r/2} = [1 - 2t]^{-r/2}, \end{aligned}$$

which is the mgf of a χ^2 -distribution with r degrees of freedom. That is, Y is $\chi^2(r)$. ■

Because the χ^2 -distributions are a subfamily of the Γ -distributions, the additivity property for Γ -distributions given by Theorem 3.3.1 holds for χ^2 -distributions, also. Since we often make use of this property, we state it as a corollary for easy reference.

Corollary 3.3.1. *Let $X_1, \dots, X_n$ be independent random variables. Suppose, for $i = 1, \dots, n$, that X_i has a $\chi^2(r_i)$ distribution. Let $Y = \sum_{i=1}^n X_i$. Then Y has a $\chi^2(\sum_{i=1}^n r_i)$ distribution.*

3.3.2 The β -Distribution

As we have discussed, in terms of modeling, the Γ -distributions offer a wide variety of shapes for skewed distributions with support $(0, \infty)$. In the exercises and later chapters, we offer other such families of distributions. How about continuous distributions whose support is a bounded interval in R ? For example suppose the support of X is (a, b) where $-\infty < a < b < \infty$ and a and b are known. Without loss of generality, for discussion, we can assume that $a = 0$ and $b = 1$, since, if not, we could consider the random variable $Y = (X - a)/(b - a)$. In this section, we discuss the **β -distribution** whose family offers a wide variety of shapes for distributions with support on bounded intervals.

One way of defining the β -family of distributions is to derive it from a pair of independent Γ random variables. Let X_1 and X_2 be two independent random variables that have Γ distributions and the joint pdf

$$h(x_1, x_2) = \frac{1}{\Gamma(\alpha)\Gamma(\beta)} x_1^{\alpha-1} x_2^{\beta-1} e^{-x_1-x_2}, \quad 0 < x_1 < \infty, \quad 0 < x_2 < \infty,$$

zero elsewhere, where $\alpha > 0$, $\beta > 0$. Let $Y_1 = X_1 + X_2$ and $Y_2 = X_1/(X_1 + X_2)$. We next show that Y_1 and Y_2 are independent.

The space $\mathcal{S}$ is, exclusive of the points on the coordinate axes, the first quadrant of the x_1, x_2 -plane. Now

$$\begin{aligned} y_1 &= u_1(x_1, x_2) = x_1 + x_2 \\ y_2 &= u_2(x_1, x_2) = \frac{x_1}{x_1 + x_2} \end{aligned}$$

may be written $x_1 = y_1 y_2$, $x_2 = y_1(1 - y_2)$, so

$$J = \begin{vmatrix} y_2 & y_1 \\ 1 - y_2 & -y_1 \end{vmatrix} = -y_1 \neq 0.$$

The transformation is one-to-one, and it maps $\mathcal{S}$ onto $\mathcal{T} = \{(y_1, y_2) : 0 < y_1 < \infty, 0 < y_2 < 1\}$ in the $y_1 y_2$ -plane. The joint pdf of Y_1 and Y_2 on its support is

$$\begin{aligned} g(y_1, y_2) &= (y_1) \frac{1}{\Gamma(\alpha)\Gamma(\beta)} (y_1 y_2)^{\alpha-1} [y_1(1 - y_2)]^{\beta-1} e^{-y_1} \\ &= \begin{cases} \frac{y_2^{\alpha-1}(1-y_2)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)} y_1^{\alpha+\beta-1} e^{-y_1} & 0 < y_1 < \infty, \quad 0 < y_2 < 1 \\ 0 & \text{elsewhere.} \end{cases} \end{aligned}$$

In accordance with Theorem 2.4.1 the random variables are independent. The marginal pdf of Y_2 is

$$\begin{aligned} g_2(y_2) &= \frac{y_2^{\alpha-1}(1-y_2)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)} \int_0^\infty y_1^{\alpha+\beta-1} e^{-y_1} dy_1 \\ &= \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y_2^{\alpha-1}(1-y_2)^{\beta-1} & 0 < y_2 < 1 \\ 0 & \text{elsewhere.} \end{cases} \end{aligned} \quad (3.3.9)$$

This pdf is that of the **beta distribution** with parameters α and β . Since $g(y_1, y_2) \equiv g_1(y_1)g_2(y_2)$, it must be that the pdf of Y_1 is

$$g_1(y_1) = \begin{cases} \frac{1}{\Gamma(\alpha+\beta)} y_1^{\alpha+\beta-1} e^{-y_1} & 0 < y_1 < \infty \\ 0 & \text{elsewhere,} \end{cases}$$

which is that of a gamma distribution with parameter values of $\alpha + \beta$ and 1.

It is an easy exercise to show that the mean and the variance of Y_2 , which has a beta distribution with parameters α and β , are, respectively,

$$\mu = \frac{\alpha}{\alpha + \beta}, \quad \sigma^2 = \frac{\alpha\beta}{(\alpha + \beta + 1)(\alpha + \beta)^2}.$$

The package R calculates probabilities for the beta distribution. If X has a beta distribution with parameters $\alpha = a$ and $\beta = b$, then the command `pbeta(x,a,b)` returns $P(X \leq x)$ and the command `dbeta(x,a,b)` returns the value of the pdf of X at x .

Example 3.3.5 (Shapes of β -Distributions). The following 3 lines of R code⁵, will obtain a 4×4 page of plots of β pdfs for all combinations of integer values of α and β between 2 and 5. Those distributions on the main diagonal of the page of plots are symmetric, those below the main diagonal are left-skewed, and those above the main diagonal are right-skewed.

```
par(mfrow=c(4,4));r1=2:5; r2=2:5;x=seq(.01,.99,.01)
for(a in r1){for(b in r2){plot(dbeta(x,a,b)~x);
  title(paste("alpha = ",a,"beta = ",b))}}
```

■

Note that if $\alpha = \beta = 1$, then the β -distribution is the uniform distribution with support $(0, 1)$.

We close this section with another example of a random variable whose distribution is derived from a transformation of gamma random variables.

Example 3.3.6 (Dirichlet Distribution). Let $X_1, X_2, \dots, X_{k+1}$ be independent random variables, each having a gamma distribution with $\beta = 1$. The joint pdf of these variables may be written as

$$h(x_1, x_2, \dots, x_{k+1}) = \begin{cases} \prod_{i=1}^{k+1} \frac{1}{\Gamma(\alpha_i)} x_i^{\alpha_i-1} e^{-x_i} & 0 < x_i < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

⁵Download the R function `betaplts` at the site listed in the Preface.

Let

$$Y_i = \frac{X_i}{X_1 + X_2 + \cdots + X_{k+1}}, \quad i = 1, 2, \dots, k,$$

and $Y_{k+1} = X_1 + X_2 + \cdots + X_{k+1}$ denote $k+1$ new random variables. The associated transformation maps $\mathcal{A} = \{(x_1, \dots, x_{k+1}) : 0 < x_i < \infty, i = 1, \dots, k+1\}$ onto the space:

$$\mathcal{B} = \{(y_1, \dots, y_k, y_{k+1}) : 0 < y_i, i = 1, \dots, k, y_1 + \cdots + y_k < 1, 0 < y_{k+1} < \infty\}.$$

The single-valued inverse functions are $x_1 = y_1 y_{k+1}, \dots, x_k = y_k y_{k+1}, x_{k+1} = y_{k+1}(1 - y_1 - \cdots - y_k)$, so that the Jacobian is

$$J = \begin{vmatrix} y_{k+1} & 0 & \cdots & 0 & y_1 \\ 0 & y_{k+1} & \cdots & 0 & y_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & y_{k+1} & y_k \\ -y_{k+1} & -y_{k+1} & \cdots & -y_{k+1} & (1 - y_1 - \cdots - y_k) \end{vmatrix} = y_{k+1}^k.$$

Hence the joint pdf of $Y_1, \dots, Y_k, Y_{k+1}$ is given by

$$\frac{y_{k+1}^{\alpha_1 + \cdots + \alpha_{k+1} - 1} y_1^{\alpha_1 - 1} \cdots y_k^{\alpha_k - 1} (1 - y_1 - \cdots - y_k)^{\alpha_{k+1} - 1} e^{-y_{k+1}}}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_k) \Gamma(\alpha_{k+1})},$$

provided that $(y_1, \dots, y_k, y_{k+1}) \in \mathcal{B}$ and is equal to zero elsewhere. By integrating out y_{k+1} , the joint pdf of $Y_1, \dots, Y_k$ is seen to be

$$g(y_1, \dots, y_k) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_{k+1})}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_{k+1})} y_1^{\alpha_1 - 1} \cdots y_k^{\alpha_k - 1} (1 - y_1 - \cdots - y_k)^{\alpha_{k+1} - 1}, \quad (3.3.10)$$

when $0 < y_i, i = 1, \dots, k, y_1 + \cdots + y_k < 1$, while the function g is equal to zero elsewhere. Random variables $Y_1, \dots, Y_k$ that have a joint pdf of this form are said to have a **Dirichlet pdf**. It is seen, in the special case of $k = 1$, that the Dirichlet pdf becomes a beta pdf. Moreover, it is also clear from the joint pdf of $Y_1, \dots, Y_k, Y_{k+1}$ that Y_{k+1} has a gamma distribution with parameters $\alpha_1 + \cdots + \alpha_k + \alpha_{k+1}$ and $\beta = 1$ and that Y_{k+1} is independent of $Y_1, Y_2, \dots, Y_k$. ■

EXERCISES

3.3.1. Suppose $(1 - 2t)^{-6}$, $t < \frac{1}{2}$ is the mgf of the random variable X .

(a) Use R to compute $P(X < 5.23)$.

(b) Find the mean μ and variance σ^2 of X . Use R to compute $P(|X - \mu| < 2\sigma)$.

3.3.2. If X is $\chi^2(5)$, determine the constants c and d so that $P(c < X < d) = 0.95$ and $P(X < c) = 0.025$.

3.3.3. Suppose the lifetime in months of an engine, working under hazardous conditions, has a Γ distribution with a mean of 10 months and a variance of 20 months squared.

- (a) Determine the median lifetime of an engine.
- (b) Suppose such an engine is termed successful if its lifetime exceeds 15 months. In a sample of 10 engines, determine the probability of at least 3 successful engines.

3.3.4. Let X be a random variable such that $E(X^m) = (m+1)!2^m$, $m = 1, 2, 3, \dots$. Determine the mgf and the distribution of X .

Hint: Write out the Taylor series⁶ of the mgf.

3.3.5. Show that

$$\int_{\mu}^{\infty} \frac{1}{\Gamma(k)} z^{k-1} e^{-z} dz = \sum_{x=0}^{k-1} \frac{\mu^x e^{-\mu}}{x!}, \quad k = 1, 2, 3, \dots$$

This demonstrates the relationship between the cdfs of the gamma and Poisson distributions.

Hint: Either integrate by parts $k-1$ times or obtain the “antiderivative” by showing that

$$\frac{d}{dz} \left[-e^{-z} \sum_{j=0}^{k-1} \frac{\Gamma(k)}{(k-j-1)!} z^{k-j-1} \right] = z^{k-1} e^{-z}.$$

3.3.6. Let X_1 , X_2 , and X_3 be iid random variables, each with pdf $f(x) = e^{-x}$, $0 < x < \infty$, zero elsewhere.

- (a) Find the distribution of $Y = \text{minimum}(X_1, X_2, X_3)$.
Hint: $P(Y \leq y) = 1 - P(Y > y) = 1 - P(X_i > y, i = 1, 2, 3)$.
- (b) Find the distribution of $Y = \text{maximum}(X_1, X_2, X_3)$.

3.3.7. Let X have a gamma distribution with pdf

$$f(x) = \frac{1}{\beta^2} x e^{-x/\beta}, \quad 0 < x < \infty,$$

zero elsewhere. If $x = 2$ is the unique mode of the distribution, find the parameter β and $P(X < 9.49)$.

3.3.8. Compute the measures of skewness and kurtosis of a gamma distribution that has parameters α and β .

3.3.9. Let X have a gamma distribution with parameters α and β . Show that $P(X \geq 2\alpha\beta) \leq (2/e)^\alpha$.

Hint: Use the result of Exercise 1.10.5.

⁶See, for example, the discussion on Taylor series in *Mathematical Comments* referenced in the Preface.

3.3.10. Give a reasonable definition of a chi-square distribution with zero degrees of freedom.

Hint: Work with the mgf of a distribution that is $\chi^2(r)$ and let $r = 0$.

3.3.11. Using the computer, obtain plots of the pdfs of chi-squared distributions with degrees of freedom $r = 1, 2, 5, 10, 20$. Comment on the plots.

3.3.12. Using the computer, plot the cdf of a $\Gamma(5, 4)$ distribution and use it to guess the median. Confirm it with a computer command that returns the median [In R, use the command `qgamma(.5, shape=5, scale=4)`].

3.3.13. Using the computer, obtain plots of beta pdfs for $\alpha = 1, 5, 10$ and $\beta = 1, 2, 5, 10, 20$.

3.3.14. In a warehouse of parts for a large mill, the average time between requests for parts is about 10 minutes.

- (a) Find the probability that in an hour there will be at least 10 requests for parts.
- (b) Find the probability that the 10th request in the morning requires at least 2 hours of waiting time.

3.3.15. Let X have a Poisson distribution with parameter m . If m is an experimental value of a random variable having a gamma distribution with $\alpha = 2$ and $\beta = 1$, compute $P(X = 0, 1, 2)$.

Hint: Find an expression that represents the joint distribution of X and m . Then integrate out m to find the marginal distribution of X .

3.3.16. Let X have the uniform distribution with pdf $f(x) = 1$, $0 < x < 1$, zero elsewhere. Find the cdf of $Y = -2 \log X$. What is the pdf of Y ?

3.3.17. Find the uniform distribution of the continuous type on the interval (b, c) that has the same mean and the same variance as those of a chi-square distribution with 8 degrees of freedom. That is, find b and c .

3.3.18. Find the mean and variance of the β distribution.

Hint: From the pdf, we know that

$$\int_0^1 y^{\alpha-1} (1-y)^{\beta-1} dy = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

for all $\alpha > 0$, $\beta > 0$.

3.3.19. Determine the constant c in each of the following so that each $f(x)$ is a β pdf:

- (a) $f(x) = cx(1-x)^3$, $0 < x < 1$, zero elsewhere.
- (b) $f(x) = cx^4(1-x)^5$, $0 < x < 1$, zero elsewhere.
- (c) $f(x) = cx^2(1-x)^8$, $0 < x < 1$, zero elsewhere.

3.3.20. Determine the constant c so that $f(x) = cx(3-x)^4$, $0 < x < 3$, zero elsewhere, is a pdf.

3.3.21. Show that the graph of the β pdf is symmetric about the vertical line through $x = \frac{1}{2}$ if $\alpha = \beta$.

3.3.22. Show, for $k = 1, 2, \dots, n$, that

$$\int_p^1 \frac{n!}{(k-1)!(n-k)!} z^{k-1} (1-z)^{n-k} dz = \sum_{x=0}^{k-1} \binom{n}{x} p^x (1-p)^{n-x}.$$

This demonstrates the relationship between the cdfs of the β and binomial distributions.

3.3.23. Let X_1 and X_2 be independent random variables. Let X_1 and $Y = X_1 + X_2$ have chi-square distributions with r_1 and r degrees of freedom, respectively. Here $r_1 < r$. Show that X_2 has a chi-square distribution with $r - r_1$ degrees of freedom. *Hint:* Write $M(t) = E(e^{t(X_1+X_2)})$ and make use of the independence of X_1 and X_2 .

3.3.24. Let X_1, X_2 be two independent random variables having gamma distributions with parameters $\alpha_1 = 3$, $\beta_1 = 3$ and $\alpha_2 = 5$, $\beta_2 = 1$, respectively.

(a) Find the mgf of $Y = 2X_1 + 6X_2$.

(b) What is the distribution of Y ?

3.3.25. Let X have an exponential distribution.

(a) For $x > 0$ and $y > 0$, show that

$$P(X > x + y \mid X > x) = P(X > y). \quad (3.3.11)$$

Hence, the exponential distribution has the **memoryless** property. Recall from Exercise 3.1.9 that the discrete geometric distribution has a similar property.

(b) Let $F(x)$ be the cdf of a continuous random variable Y . Assume that $F(0) = 0$ and $0 < F(y) < 1$ for $y > 0$. Suppose property (3.3.11) holds for Y . Show that $F_Y(y) = 1 - e^{-\lambda y}$ for $y > 0$.

Hint: Show that $g(y) = 1 - F_Y(y)$ satisfies the equation

$$g(y+z) = g(y)g(z),$$

3.3.26. Let X denote time until failure of a device and let $r(x)$ denote the hazard function of X .

(a) If $r(x) = cx^b$; where c and b are positive constants, show that X has a **Weibull** distribution; i.e.,

$$f(x) = \begin{cases} cx^b \exp \left\{ -\frac{cx^{b+1}}{b+1} \right\} & 0 < x < \infty \\ 0 & \text{elsewhere.} \end{cases} \quad (3.3.12)$$

- (b) If $r(x) = ce^{bx}$, where c and b are positive constants, show that X has a **Gompertz** cdf given by

$$F(x) = \begin{cases} 1 - \exp\left\{\frac{c}{b}(1 - e^{bx})\right\} & 0 < x < \infty \\ 0 & \text{elsewhere.} \end{cases} \quad (3.3.13)$$

This is frequently used by actuaries as a distribution of the length of human life.

- (c) If $r(x) = bx$, linear hazard rate, show that the pdf of X is

$$f(x) = \begin{cases} bxe^{-bx^2/2} & 0 < x < \infty \\ 0 & \text{elsewhere.} \end{cases} \quad (3.3.14)$$

This pdf is called the **Rayleigh** pdf.

3.3.27. Write an R function that returns the value $f(x)$ for a specified x when $f(x)$ is the Weibull pdf given in expression (3.3.12). Next write an R function that returns the associated hazard function $r(x)$. Obtain side-by-side plots of the pdf and hazard function for the three cases: $c = 5$ and $b = 0.5$; $c = 5$ and $b = 1.0$; and $c = 5$ and $b = 1.5$.

3.3.28. In Example 3.3.5, a page of plots of β pdfs was discussed. All of these pdfs are mound shaped. Obtain a page of plots for all combinations of α and β drawn from the set $\{.25, .75, 1, 1.25\}$. Comment on these shapes.

3.3.29. Let $Y_1, \dots, Y_k$ have a Dirichlet distribution with parameters $\alpha_1, \dots, \alpha_k, \alpha_{k+1}$.

- (a) Show that Y_1 has a beta distribution with parameters $\alpha = \alpha_1$ and $\beta = \alpha_2 + \dots + \alpha_{k+1}$.
- (b) Show that $Y_1 + \dots + Y_r$, $r \leq k$, has a beta distribution with parameters $\alpha = \alpha_1 + \dots + \alpha_r$ and $\beta = \alpha_{r+1} + \dots + \alpha_{k+1}$.
- (c) Show that $Y_1 + Y_2$, $Y_3 + Y_4$, $Y_5, \dots, Y_k$, $k \geq 5$, have a Dirichlet distribution with parameters $\alpha_1 + \alpha_2$, $\alpha_3 + \alpha_4$, $\alpha_5, \dots, \alpha_k, \alpha_{k+1}$.

Hint: Recall the definition of Y_i in Example 3.3.6 and use the fact that the sum of several independent gamma variables with $\beta = 1$ is a gamma variable.

3.4 The Normal Distribution

Motivation for the normal distribution is found in the Central Limit Theorem, which is presented in Section 5.3. This theorem shows that normal distributions provide an important family of distributions for applications and for statistical inference, in general. We proceed by first introducing the standard normal distribution and through it the general normal distribution.

Consider the integral

$$I = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right) dz. \quad (3.4.1)$$

This integral exists because the integrand is a positive continuous function that is bounded by an integrable function; that is,

$$0 < \exp\left(\frac{-z^2}{2}\right) < \exp(-|z| + 1), \quad -\infty < z < \infty,$$

and

$$\int_{-\infty}^{\infty} \exp(-|z| + 1) dz = 2e.$$

To evaluate the integral I , we note that $I > 0$ and that I^2 may be written

$$I^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-\frac{z^2 + w^2}{2}\right) dz dw.$$

This iterated integral can be evaluated by changing to polar coordinates. If we set $z = r \cos \theta$ and $w = r \sin \theta$, we have

$$\begin{aligned} I^2 &= \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} e^{-r^2/2} r dr d\theta \\ &= \frac{1}{2\pi} \int_0^{2\pi} d\theta = 1. \end{aligned}$$

Because the integrand of display (3.4.1) is positive on R and integrates to 1 over R , it is a pdf of a continuous random variable with support R . We denote this random variable by Z . In summary, Z has the pdf

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right), \quad -\infty < z < \infty. \quad (3.4.2)$$

For $t \in R$, the mgf of Z can be derived by a completion of a square as follows:

$$\begin{aligned} E[\exp\{tZ\}] &= \int_{-\infty}^{\infty} \exp\{tz\} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}z^2\right\} dz \\ &= \exp\left\{\frac{1}{2}t^2\right\} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(z-t)^2\right\} dz \\ &= \exp\left\{\frac{1}{2}t^2\right\} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}w^2\right\} dw, \end{aligned} \quad (3.4.3)$$

where for the last integral we made the one-to-one change of variable $w = z - t$. By the identity (3.4.2), the integral in expression (3.4.3) has value 1. Thus the mgf of Z is

$$M_Z(t) = \exp\left\{\frac{1}{2}t^2\right\}, \quad \text{for } -\infty < t < \infty. \quad (3.4.4)$$

The first two derivatives of $M_Z(t)$ are easily shown to be

$$\begin{aligned} M'_Z(t) &= t \exp\left\{\frac{1}{2}t^2\right\} \\ M''_Z(t) &= \exp\left\{\frac{1}{2}t^2\right\} + t^2 \exp\left\{\frac{1}{2}t^2\right\}. \end{aligned}$$

Upon evaluating these derivatives at $t = 0$, the mean and variance of Z are

$$E(Z) = 0 \text{ and } \text{Var}(Z) = 1. \quad (3.4.5)$$

Next, define the continuous random variable X by

$$X = bZ + a,$$

for $b > 0$. This is a one-to-one transformation. To derive the pdf of X , note that the inverse of the transformation and the Jacobian are $z = b^{-1}(x - a)$ and $J = b^{-1}$, respectively. Because $b > 0$, it follows from (3.4.2) that the pdf of X is

$$f_X(x) = \frac{1}{\sqrt{2\pi}b} \exp \left\{ -\frac{1}{2} \left(\frac{x-a}{b} \right)^2 \right\}, \quad -\infty < x < \infty.$$

By (3.4.5), we immediately have $E(X) = a$ and $\text{Var}(X) = b^2$. Hence, in the expression for the pdf of X , we can replace a by $\mu = E(X)$ and b^2 by $\sigma^2 = \text{Var}(X)$. We make this formal in the following:

Definition 3.4.1 (Normal Distribution). *We say a random variable X has a **normal distribution** if its pdf is*

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2 \right\}, \quad \text{for } -\infty < x < \infty. \quad (3.4.6)$$

The parameters μ and σ^2 are the mean and variance of X , respectively. We often write that X has a $N(\mu, \sigma^2)$ distribution.

In this notation, the random variable Z with pdf (3.4.2) has a $N(0, 1)$ distribution. We call Z a **standard normal** random variable.

For the mgf of X , use the relationship $X = \sigma Z + \mu$ and the mgf for Z , (3.4.4), to obtain

$$\begin{aligned} E[\exp\{tX\}] &= E[\exp\{t(\sigma Z + \mu)\}] = \exp\{\mu t\} E[\exp\{t\sigma Z\}] \\ &= \exp\{\mu t\} \exp \left\{ \frac{1}{2} \sigma^2 t^2 \right\} = \exp \left\{ \mu t + \frac{1}{2} \sigma^2 t^2 \right\}, \end{aligned} \quad (3.4.7)$$

for $-\infty < t < \infty$.

We summarize the above discussion, by noting the relationship between Z and X :

X has a $N(\mu, \sigma^2)$ distribution if and only if $Z = \frac{X-\mu}{\sigma}$ has a $N(0, 1)$ distribution. (3.4.8)

Let X have a $N(\mu, \sigma^2)$ distribution. The graph of the pdf of X is seen in Figure 3.4.1 to have the following characteristics: (1) symmetry about a vertical axis through $x = \mu$; (2) having its maximum of $1/(\sigma\sqrt{2\pi})$ at $x = \mu$; and (3) having the x -axis as a horizontal asymptote. It should also be verified that (4) there are

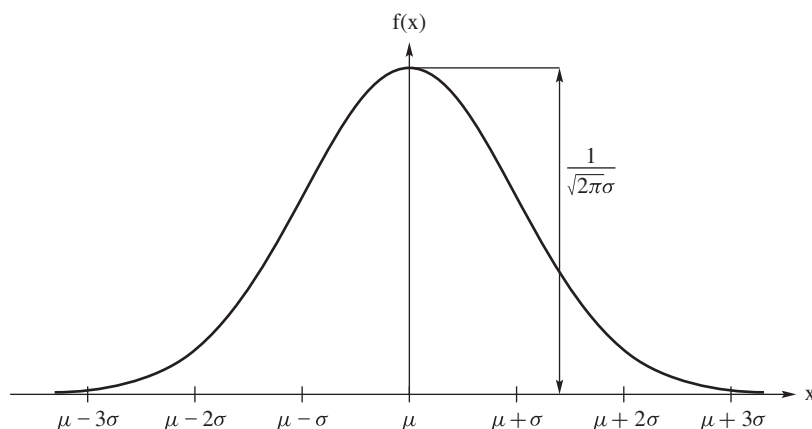


Figure 3.4.1: The normal density $f(x)$, (3.4.6).

points of inflection at $x = \mu \pm \sigma$; see Exercise 3.4.7. By the symmetry about μ , it follows that the median of a normal distribution is equal to its mean.

If we want to determine $P(X \leq x)$, then the following integration is required:

$$P(X \leq x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-(t-\mu)^2/(2\sigma^2)} dt.$$

From calculus we know that the integrand does not have an antiderivative; hence, the integration must be carried out by numerical integration procedures. The R software uses such a procedure for its function `pnorm`. If X has a $N(\mu, \sigma^2)$ distribution, then the R call `pnorm(x, mu, sigma)` computes $P(X \leq x)$, while `q = qnorm(p, mu, sigma)` gives the p th quantile of X ; i.e., q solves the equation $P(X \leq q) = p$. We illustrate this computation in the next example.

Example 3.4.1. Suppose the height in inches of an adult male is normally distributed with mean $\mu = 70$ inches and standard deviation $\sigma = 4$ inches. As a graph of the pdf of X use Figure 3.4.1 replacing μ by 70 and σ by 4. Suppose we want to compute the probability that a man exceeds six feet (72 inches) in height. Locate 72 on the figure. The desired probability is the area under the curve over the interval $(72, \infty)$ which is computed in R by `1 - pnorm(72, 70, 4) = 0.3085`; hence, 31% of males exceed six feet in height. The 95th percentile in height is `qnorm(0.95, 70, 4) = 76.6` inches. What percentage of males have heights within one standard deviation of the mean? Answer: `pnorm(74, 70, 4) - pnorm(66, 70, 4) = 0.6827`. ■

Before the age of modern computing tables of probabilities for normal distributions were formulated. Due to the fact (3.4.8), only tables for the standard normal distribution are required. Let Z have the standard normal distribution. A graph of

its pdf is displayed in Figure 3.4.2. Common notation for the cdf of Z is

$$P(Z \leq z) = \Phi(z) = \text{dfn} \int_0^z \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt, \quad -\infty < z < \infty. \quad (3.4.9)$$

Table II of Appendix D displays a table for $\Phi(z)$ for specified values of $z > 0$. To compute $\Phi(-z)$, where $z > 0$, use the identity

$$\Phi(-z) = 1 - \Phi(z). \quad (3.4.10)$$

This identity follows because the pdf of Z is symmetric about 0. It is apparent in Figure 3.4.2 and the reader is asked to show it in Exercise 3.4.1.

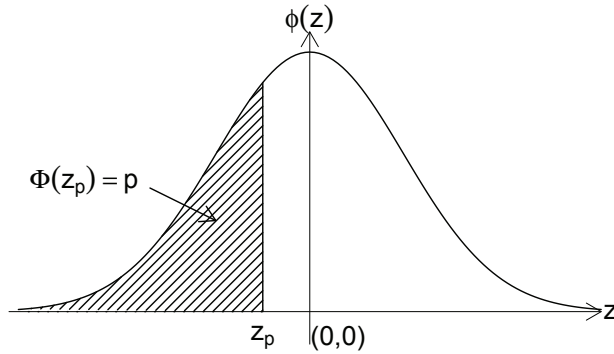


Figure 3.4.2: The standard normal density: $p = \Phi(z_p)$ is the area under the curve to the left of z_p .

As an illustration of the use of Table II, suppose in Example 3.4.1 that we want to determine the probability that the height of an adult male is between 67 and 71 inches. This is calculated as

$$\begin{aligned} P(67 < X < 71) &= P(X < 71) - P(X < 67) \\ &= P\left(\frac{X - 70}{4} < \frac{71 - 70}{4}\right) - P\left(\frac{X - 70}{4} < \frac{67 - 70}{4}\right) \\ &= P(Z < 0.25) - P(Z < -0.75) = \Phi(0.25) - 1 + \Phi(0.75) \\ &= 0.5987 - 1 + 0.7734 = 0.3721 \end{aligned} \quad (3.4.11)$$

$$= \text{pnorm}(71, 70, 4) - \text{pnorm}(67, 70, 4) = 0.372079. \quad (3.4.12)$$

Expression (3.4.11) is the calculation by using Table II, while the last line is the calculation by using the R function `pnorm`. More examples are offered in the exercises. As a final note on Table II, it is generated by the R function:

```
normtab <- function(){ za <- seq(0.00,3.59,.01);
  pz <- t(matrix(round(pnorm(za),digits=4),nrow=10))
  colnames(pz) <- seq(0,.09,.01)
  rownames(pz) <- seq(0.0,3.5,.1); return(pz)}
```

The function `normtab` can be downloaded at the site mentioned in the Preface.

Example 3.4.2 (Empirical Rule). Let X be $N(\mu, \sigma^2)$. Then, by Table II or R,

$$\begin{aligned} P(\mu - 2\sigma < X < \mu + 2\sigma) &= \Phi\left(\frac{\mu + 2\sigma - \mu}{\sigma}\right) - \Phi\left(\frac{\mu - 2\sigma - \mu}{\sigma}\right) \\ &= \Phi(2) - \Phi(-2) \\ &= 0.977 - (1 - 0.977) = 0.954. \end{aligned}$$

Similarly, $P(\mu - \sigma < X < \mu + \sigma) = 0.6827$ and $P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9973$. Sometimes these three intervals and their corresponding probabilities are referred to as the **empirical rule**. Note that we can use Chebyshev's Theorem (Theorem 1.10.3), to obtain lower bounds for these probabilities. While the empirical rule is much more precise, it also requires the assumption of a normal distribution. On the other hand, Chebyshev's theorem requires only the assumption of a finite variance.

■

Example 3.4.3. Suppose that 10% of the probability for a certain distribution that is $N(\mu, \sigma^2)$ is below 60 and that 5% is above 90. What are the values of μ and σ ? We are given that the random variable X is $N(\mu, \sigma^2)$ and that $P(X \leq 60) = 0.10$ and $P(X \leq 90) = 0.95$. Thus $\Phi[(60 - \mu)/\sigma] = 0.10$ and $\Phi[(90 - \mu)/\sigma] = 0.95$. From Table II we have

$$\frac{60 - \mu}{\sigma} = -1.28, \quad \frac{90 - \mu}{\sigma} = 1.64.$$

These conditions require that $\mu = 73.1$ and $\sigma = 10.2$ approximately. ■

Remark 3.4.1. In this chapter we have illustrated three types of **parameters** associated with distributions. The mean μ of $N(\mu, \sigma^2)$ is called a **location parameter** because changing its value simply changes the location of the middle of the normal pdf; that is, the graph of the pdf looks exactly the same except for a shift in location. The standard deviation σ of $N(\mu, \sigma^2)$ is called a **scale parameter** because changing its value changes the spread of the distribution. That is, a small value of σ requires the graph of the normal pdf to be tall and narrow, while a large value of σ requires it to spread out and not be so tall. No matter what the values of μ and σ , however, the graph of the normal pdf is that familiar “bell shape.” Incidentally, the β of the gamma distribution is also a scale parameter. On the other hand, the α of the gamma distribution is called a **shape parameter**, as changing its value modifies the shape of the graph of the pdf, as can be seen by referring to Figure 3.3.1. The parameters p and μ of the binomial and Poisson distributions, respectively, are also shape parameters. ■

Continuing with the first part of Remark 3.4.1, if X is $N(\mu, \sigma^2)$ then we say that X follows the **location model** which we write as

$$X = \mu + e, \tag{3.4.13}$$

where e is a random variable (often called random error) with a $N(0, \sigma^2)$ distribution. Conversely, it follows immediately that if X satisfies expression (3.4.13) with e distributed $N(0, \sigma^2)$ then X has a $N(\mu, \sigma^2)$ distribution.

We close this part of the section with three important results.

Example 3.4.4 (All the Moments of a Normal Distribution). Recall that in Example 1.9.7, we derived all the moments of a standard normal random variable by using its moment generating function. We can use this to obtain all the moments of X , where X has a $N(\mu, \sigma^2)$ distribution. From expression (3.4.13), we can write $X = \sigma Z + \mu$, where Z has a $N(0, 1)$ distribution. Hence, for all nonnegative integers k a simple application of the binomial theorem yields

$$E(X^k) = E[(\sigma Z + \mu)^k] = \sum_{j=0}^k \binom{k}{j} \sigma^j E(Z^j) \mu^{k-j}. \quad (3.4.14)$$

Recall from Example 1.9.7 that all the odd moments of Z are 0, while all the even moments are given by expression (1.9.3). These can be substituted into expression (3.4.14) to derive the moments of X . ■

Theorem 3.4.1. *If the random variable X is $N(\mu, \sigma^2)$, $\sigma^2 > 0$, then the random variable $V = (X - \mu)^2 / \sigma^2$ is $\chi^2(1)$.*

Proof. Because $V = W^2$, where $W = (X - \mu) / \sigma$ is $N(0, 1)$, the cdf $G(v)$ for V is, for $v \geq 0$,

$$G(v) = P(W^2 \leq v) = P(-\sqrt{v} \leq W \leq \sqrt{v}).$$

That is,

$$G(v) = 2 \int_0^{\sqrt{v}} \frac{1}{\sqrt{2\pi}} e^{-w^2/2} dw, \quad 0 \leq v,$$

and

$$G(v) = 0, \quad v < 0.$$

If we change the variable of integration by writing $w = \sqrt{y}$, then

$$G(v) = \int_0^v \frac{1}{\sqrt{2\pi}\sqrt{y}} e^{-y/2} dy, \quad 0 \leq v.$$

Hence the pdf $g(v) = G'(v)$ of the continuous-type random variable V is

$$g(v) = \begin{cases} \frac{1}{\sqrt{\pi}\sqrt{2}} v^{1/2-1} e^{-v/2} & 0 < v < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

Since $g(v)$ is a pdf

$$\int_0^\infty g(v) dv = 1;$$

hence, it must be that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ and thus V is $\chi^2(1)$. ■

One of the most important properties of the normal distribution is its additivity under independence.

Theorem 3.4.2. *Let $X_1, \dots, X_n$ be independent random variables such that, for $i = 1, \dots, n$, X_i has a $N(\mu_i, \sigma_i^2)$ distribution. Let $Y = \sum_{i=1}^n a_i X_i$, where $a_1, \dots, a_n$ are constants. Then the distribution of Y is $N(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2)$.*

Proof: By Theorem 2.6.1, for $t \in R$, the mgf of Y is

$$\begin{aligned} M_Y(t) &= \prod_{i=1}^n \exp \{ t a_i \mu_i + (1/2) t^2 a_i^2 \sigma_i^2 \} \\ &= \exp \left\{ t \sum_{i=1}^n a_i \mu_i + (1/2) t^2 \sum_{i=1}^n a_i^2 \sigma_i^2 \right\}, \end{aligned}$$

which is the mgf of a $N(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2)$ distribution. ■

A simple corollary to this result gives the distribution of the sample mean $\bar{X} = n^{-1} \sum_{i=1}^n X_i$ when $X_1, X_2, \dots, X_n$ represents a random sample from a $N(\mu, \sigma^2)$.

Corollary 3.4.1. *Let $X_1, \dots, X_n$ be iid random variables with a common $N(\mu, \sigma^2)$ distribution. Let $\bar{X} = n^{-1} \sum_{i=1}^n X_i$. Then $\bar{X}$ has a $N(\mu, \sigma^2/n)$ distribution.*

To prove this corollary, simply take $a_i = (1/n)$, $\mu_i = \mu$, and $\sigma_i^2 = \sigma^2$, for $i = 1, 2, \dots, n$, in Theorem 3.4.2.

3.4.1 *Contaminated Normals

We next discuss a random variable whose distribution is a mixture of normals. As with the normal, we begin with a standardized random variable.

Suppose we are observing a random variable that most of the time follows a standard normal distribution but occasionally follows a normal distribution with a larger variance. In applications, we might say that most of the data are “good” but that there are occasional outliers. To make this precise let Z have a $N(0, 1)$ distribution; let $I_{1-\epsilon}$ be a discrete random variable defined by

$$I_{1-\epsilon} = \begin{cases} 1 & \text{with probability } 1 - \epsilon \\ 0 & \text{with probability } \epsilon, \end{cases}$$

and assume that Z and $I_{1-\epsilon}$ are independent. Let $W = Z I_{1-\epsilon} + \sigma_c Z (1 - I_{1-\epsilon})$. Then W is the random variable of interest.

The independence of Z and $I_{1-\epsilon}$ imply that the cdf of W is

$$\begin{aligned} F_W(w) = P[W \leq w] &= P[W \leq w, I_{1-\epsilon} = 1] + P[W \leq w, I_{1-\epsilon} = 0] \\ &= P[W \leq w | I_{1-\epsilon} = 1] P[I_{1-\epsilon} = 1] \\ &\quad + P[W \leq w | I_{1-\epsilon} = 0] P[I_{1-\epsilon} = 0] \\ &= P[Z \leq w](1 - \epsilon) + P[Z \leq w/\sigma_c] \epsilon. \\ &= \Phi(w)(1 - \epsilon) + \Phi(w/\sigma_c) \epsilon \end{aligned} \tag{3.4.15}$$

Therefore, we have shown that the distribution of W is a mixture of normals. Further, because $W = Z I_{1-\epsilon} + \sigma_c Z (1 - I_{1-\epsilon})$, we have

$$E(W) = 0 \text{ and } \text{Var}(W) = 1 + \epsilon(\sigma_c^2 - 1); \tag{3.4.16}$$

see Exercise 3.4.24. Upon differentiating (3.4.15), the pdf of W is

$$f_W(w) = \phi(w)(1 - \epsilon) + \phi(w/\sigma_c) \frac{\epsilon}{\sigma_c}, \quad (3.4.17)$$

where ϕ is the pdf of a standard normal.

Suppose, in general, that the random variable of interest is $X = a + bW$, where $b > 0$. Based on (3.4.16), the mean and variance of X are

$$E(X) = a \text{ and } \text{Var}(X) = b^2(1 + \epsilon(\sigma_c^2 - 1)). \quad (3.4.18)$$

From expression (3.4.15), the cdf of X is

$$F_X(x) = \Phi\left(\frac{x-a}{b}\right)(1 - \epsilon) + \Phi\left(\frac{x-a}{b\sigma_c}\right)\epsilon, \quad (3.4.19)$$

which is a mixture of normal cdfs.

Based on expression (3.4.19) it is easy to obtain probabilities for contaminated normal distributions using R. For example, suppose, as above, W has cdf (3.4.15). Then $P(W \leq w)$ is obtained by the R command `(1-eps)*pnorm(w) + eps*pnorm(w/sigc)`, where `eps` and `sigc` denote ϵ and σ_c , respectively. Similarly, the pdf of W at w is returned by `(1-eps)*dnorm(w) + eps*dnorm(w/sigc)/sigc`. The functions `pcn` and `dcn`⁷ compute the cdf and pdf of the contaminated normal, respectively. In Section 3.7, we explore mixture distributions in general.

EXERCISES

3.4.1. If

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-w^2/2} dw,$$

show that $\Phi(-z) = 1 - \Phi(z)$.

3.4.2. If X is $N(75, 100)$, find $P(X < 60)$ and $P(70 < X < 100)$ by using either Table II or the R command `pnorm`.

3.4.3. If X is $N(\mu, \sigma^2)$, find b so that $P[-b < (X - \mu)/\sigma < b] = 0.90$, by using either Table II of Appendix D or the R command `qnorm`.

3.4.4. Let X be $N(\mu, \sigma^2)$ so that $P(X < 89) = 0.90$ and $P(X < 94) = 0.95$. Find μ and σ^2 .

3.4.5. Show that the constant c can be selected so that $f(x) = c2^{-x^2}$, $-\infty < x < \infty$, satisfies the conditions of a normal pdf.

Hint: Write $2 = e^{\log 2}$.

3.4.6. If X is $N(\mu, \sigma^2)$, show that $E(|X - \mu|) = \sigma\sqrt{2/\pi}$.

3.4.7. Show that the graph of a pdf $N(\mu, \sigma^2)$ has points of inflection at $x = \mu - \sigma$ and $x = \mu + \sigma$.

⁷Downloadable at the site listed in the Preface.

3.4.8. Evaluate $\int_2^3 \exp[-2(x-3)^2] dx$.

3.4.9. Determine the 90th percentile of the distribution, which is $N(65, 25)$.

3.4.10. If e^{3t+8t^2} is the mgf of the random variable X , find $P(-1 < X < 9)$.

3.4.11. Let the random variable X have the pdf

$$f(x) = \frac{2}{\sqrt{2\pi}} e^{-x^2/2}, \quad 0 < x < \infty, \quad \text{zero elsewhere.}$$

(a) Find the mean and the variance of X .

(b) Find the cdf and hazard function of X .

Hint for (a): Compute $E(X)$ directly and $E(X^2)$ by comparing the integral with the integral representing the variance of a random variable that is $N(0, 1)$.

3.4.12. Let X be $N(5, 10)$. Find $P[0.04 < (X - 5)^2 < 38.4]$.

3.4.13. If X is $N(1, 4)$, compute the probability $P(1 < X^2 < 9)$.

3.4.14. If X is $N(75, 25)$, find the conditional probability that X is greater than 80 given that X is greater than 77. See Exercise 2.3.12.

3.4.15. Let X be a random variable such that $E(X^{2m}) = (2m)!/(2^m m!)$, $m = 1, 2, 3, \dots$ and $E(X^{2m-1}) = 0$, $m = 1, 2, 3, \dots$. Find the mgf and the pdf of X .

3.4.16. Let the mutually independent random variables X_1 , X_2 , and X_3 be $N(0, 1)$, $N(2, 4)$, and $N(-1, 1)$, respectively. Compute the probability that exactly two of these three variables are less than zero.

3.4.17. Compute the measures of skewness and kurtosis of a distribution which is $N(\mu, \sigma^2)$. See Exercises 1.9.14 and 1.9.15 for the definitions of skewness and kurtosis, respectively.

3.4.18. Let the random variable X have a distribution that is $N(\mu, \sigma^2)$.

(a) Does the random variable $Y = X^2$ also have a normal distribution?

(b) Would the random variable $Y = aX + b$, a and b nonzero constants have a normal distribution?

Hint: In each case, first determine $P(Y \leq y)$.

3.4.19. Let the random variable X be $N(\mu, \sigma^2)$. What would this distribution be if $\sigma^2 = 0$?

Hint: Look at the mgf of X for $\sigma^2 > 0$ and investigate its limit as $\sigma^2 \rightarrow 0$.

3.4.20. Let Y have a **truncated** distribution with pdf $g(y) = \phi(y)/[\Phi(b) - \Phi(a)]$, for $a < y < b$, zero elsewhere, where $\phi(x)$ and $\Phi(x)$ are, respectively, the pdf and distribution function of a standard normal distribution. Show then that $E(Y)$ is equal to $[\phi(a) - \phi(b)]/[\Phi(b) - \Phi(a)]$.

3.4.21. Let $f(x)$ and $F(x)$ be the pdf and the cdf, respectively, of a distribution of the continuous type such that $f'(x)$ exists for all x . Let the mean of the truncated distribution that has pdf $g(y) = f(y)/F(b)$, $-\infty < y < b$, zero elsewhere, be equal to $-f(b)/F(b)$ for all real b . Prove that $f(x)$ is a pdf of a standard normal distribution.

3.4.22. Let X and Y be independent random variables, each with a distribution that is $N(0, 1)$. Let $Z = X + Y$. Find the integral that represents the cdf $G(z) = P(X + Y \leq z)$ of Z . Determine the pdf of Z .

Hint: We have that $G(z) = \int_{-\infty}^{\infty} H(x, z) dx$, where

$$H(x, z) = \int_{-\infty}^{z-x} \frac{1}{2\pi} \exp[-(x^2 + y^2)/2] dy.$$

Find $G'(z)$ by evaluating $\int_{-\infty}^{\infty} [\partial H(x, z)/\partial z] dx$.

3.4.23. Suppose X is a random variable with the pdf $f(x)$ which is symmetric about 0; i.e., $f(-x) = f(x)$. Show that $F(-x) = 1 - F(x)$, for all x in the support of X .

3.4.24. Derive the mean and variance of a contaminated normal random variable. They are given in expression (3.4.16).

3.4.25. Investigate the probabilities of an “outlier” for a contaminated normal random variable and a normal random variable. Specifically, determine the probability of observing the event $\{|X| \geq 2\}$ for the following random variables (use the R function `pcn` for the contaminated normals):

- (a) X has a standard normal distribution.
- (b) X has a contaminated normal distribution with cdf (3.4.15), where $\epsilon = 0.15$ and $\sigma_c = 10$.
- (c) X has a contaminated normal distribution with cdf (3.4.15), where $\epsilon = 0.15$ and $\sigma_c = 20$.
- (d) X has a contaminated normal distribution with cdf (3.4.15), where $\epsilon = 0.25$ and $\sigma_c = 20$.

3.4.26. Plot the pdfs of the random variables defined in parts (a)–(d) of the last exercise. Obtain an overlay plot of all four pdfs also. In R the domain values of the pdfs can easily be obtained by using the `seq` command. For instance, the command `x<-seq(-6,6,.1)` returns a vector of values between -6 and 6 in jumps of 0.1 . Then use the R function `dcn` for the contaminated normal pdfs.

3.4.27. Consider the family of pdfs indexed by the parameter α , $-\infty < \alpha < \infty$, given by

$$f(x; \alpha) = 2\phi(x)\Phi(\alpha x), \quad -\infty < x < \infty, \quad (3.4.20)$$

where $\phi(x)$ and $\Phi(x)$ are respectively the pdf and cdf of a standard normal distribution.

- (a) Clearly $f(x; \alpha) > 0$ for all x . Show that the pdf integrates to 1 over $(-\infty, \infty)$.

Hint: Start with

$$\int_{-\infty}^{\infty} f(x; \alpha) dx = 2 \int_{-\infty}^{\infty} \phi(x) \int_{-\infty}^{\alpha x} \phi(t) dt.$$

Next sketch the region of integration and then combine the integrands and use the polar coordinate transformation we used after expression (3.4.1).

- (b) Note that $f(x; \alpha)$ is the $N(0, 1)$ pdf for $\alpha = 0$. The pdfs are left skewed for $\alpha < 0$ and right skewed for $\alpha > 0$. Using R, verify this by plotting the pdfs for $\alpha = -3, -2, -1, 1, 2, 3$. Here's the code for $\alpha = -3$:

```
x=seq(-5,5,.01); alp=-3; y=2*dnorm(x)*pnorm(alp*x);plot(y~x)
```

This family is called the **skewed normal family**; see Azzalini (1985).

3.4.28. For Z distributed $N(0, 1)$, it can be shown that

$$E[\Phi(hZ + k)] = \Phi[k/\sqrt{1 + h^2}];$$

see Azzalini (1985). Use this fact to obtain the mgf of the pdf (3.4.20). Next obtain the mean of this pdf.

3.4.29. Let X_1 and X_2 be independent with normal distributions $N(6, 1)$ and $N(7, 1)$, respectively. Find $P(X_1 > X_2)$.

Hint: Write $P(X_1 > X_2) = P(X_1 - X_2 > 0)$ and determine the distribution of $X_1 - X_2$.

3.4.30. Compute $P(X_1 + 2X_2 - 2X_3 > 7)$ if X_1, X_2, X_3 are iid with common distribution $N(1, 4)$.

3.4.31. A certain job is completed in three steps in series. The means and standard deviations for the steps are (in minutes)

Step	Mean	Standard Deviation
1	17	2
2	13	1
3	13	2

Assuming independent steps and normal distributions, compute the probability that the job takes less than 40 minutes to complete.

3.4.32. Let X be $N(0, 1)$. Use the moment generating function technique to show that $Y = X^2$ is $\chi^2(1)$.

Hint: Evaluate the integral that represents $E(e^{tX^2})$ by writing $w = x\sqrt{1 - 2t}$, $t < \frac{1}{2}$.

3.4.33. Suppose X_1, X_2 are iid with a common standard normal distribution. Find the joint pdf of $Y_1 = X_1^2 + X_2^2$ and $Y_2 = X_2$ and the marginal pdf of Y_1 .

Hint: Note that the space of Y_1 and Y_2 is given by $-\sqrt{y_1} < y_2 < \sqrt{y_1}, 0 < y_1 < \infty$.

3.5 The Multivariate Normal Distribution

In this section we present the multivariate normal distribution. In the first part of the section, we introduce the bivariate normal distribution, leaving most of the proofs to the later section, Section 3.5.2.

3.5.1 Bivariate Normal Distribution

We say that (X, Y) follows a **bivariate normal distribution** if its pdf is given by

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-q/2}, \quad -\infty < x < \infty, \quad -\infty < y < \infty, \quad (3.5.1)$$

where

$$q = \frac{1}{1-\rho^2} \left[\left(\frac{x-\mu_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{x-\mu_1}{\sigma_1} \right) \left(\frac{y-\mu_2}{\sigma_2} \right) + \left(\frac{y-\mu_2}{\sigma_2} \right)^2 \right], \quad (3.5.2)$$

and $-\infty < \mu_i < \infty$, $\sigma_i > 0$, for $i = 1, 2$, and ρ satisfies $\rho^2 < 1$. Clearly, this function is positive everywhere in R^2 . As we show in Section 3.5.2, it is a pdf with the mgf given by:

$$M_{(X,Y)}(t_1, t_2) = \exp \left\{ t_1\mu_1 + t_2\mu_2 + \frac{1}{2}(t_1^2\sigma_1^2 + 2t_1t_2\rho\sigma_1\sigma_2 + t_2^2\sigma_2^2) \right\}. \quad (3.5.3)$$

Thus, the mgf of X is

$$M_X(t_1) = M_{(X,Y)}(t_1, 0) = \exp \left\{ t_1\mu_1 + \frac{1}{2}t_1^2\sigma_1^2 \right\};$$

hence, X has a $N(\mu_1, \sigma_1^2)$ distribution. In the same way, Y has a $N(\mu_2, \sigma_2^2)$ distribution. Thus μ_1 and μ_2 are the respective means of X and Y and σ_1^2 and σ_2^2 are the respective variances of X and Y . For the parameter ρ , Exercise 3.5.3 shows that

$$E(XY) = \frac{\partial^2 M_{(X,Y)}}{\partial t_1 \partial t_2}(0, 0) = \rho\sigma_1\sigma_2 + \mu_1\mu_2. \quad (3.5.4)$$

Hence, $\text{cov}(X, Y) = \rho\sigma_1\sigma_2$ and thus, as the notation suggests, ρ is the correlation coefficient between X and Y . We know by Theorem 2.5.2 that if X and Y are independent then $\rho = 0$. Further, from expression (3.5.3), if $\rho = 0$ then the joint mgf of (X, Y) factors into the product of the marginal mgfs and, hence, X and Y are independent random variables. Thus if (X, Y) has a bivariate normal distribution, then X and Y are independent if and only if they are uncorrelated.

The bivariate normal pdf, (3.5.1), is mound shaped over R^2 and peaks at its mean (μ_1, μ_2) ; see Exercise 3.5.4. For a given $c > 0$, the points of equal probability (or density) are given by $\{(x, y) : f(x, y) = c\}$. It follows with some algebra that these sets are ellipses. In general for multivariate distributions, we call these sets **contours** of the pdfs. Hence, the contours of bivariate normal distributions are

elliptical. If X and Y are independent then these contours are circular. The interested reader can consult a book on multivariate statistics for discussions on the geometry of the ellipses. For example, if $\sigma_1 = \sigma_2$ and $\rho > 0$, the main axis of the ellipse goes through the mean at a 45° angle; see Johnson and Wichern (2008) for discussion.

Figure 3.5.1 displays a three-dimensional plot of the bivariate normal pdf with $(\mu_1, \mu_2) = (0, 0)$, $\sigma_1 = \sigma_2 = 1$, and $\rho = 0.5$. For location, the peak is at $(\mu_1, \mu_2) = (0, 0)$. The elliptical contours are apparent. Locate the main axis. For a region A in the plane, $P[(X, Y) \in A]$ is the volume under the surface over A . In general such probabilities are calculated by numerical integration methods.

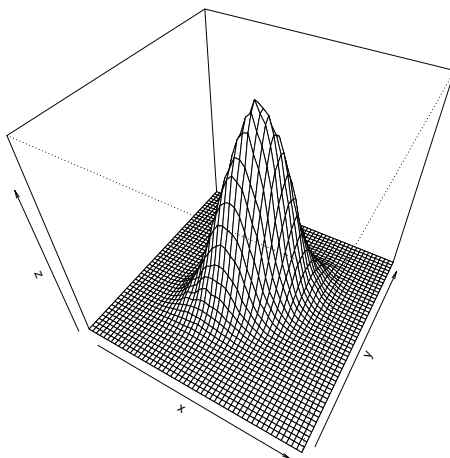


Figure 3.5.1: A sketch of the surface of a bivariate normal distribution with mean $(0, 0)$, $\sigma_1 = \sigma_2 = 1$, and $\rho = 0.5$.

In the next section, we extend the discussion to the general multivariate case; however, Remark 3.5.1, below, returns to the bivariate case and can be read with minor knowledge of vector and matrices.

3.5.2 *Multivariate Normal Distribution, General Case

In this section we generalize the bivariate normal distribution to the n -dimensional multivariate normal distribution. As with Section 3.4 on the normal distribution, the derivation of the distribution is simplified by first discussing the standardized variable case and then proceeding to the general case. Also, in this section, vector and matrix notation are used.

Consider the random vector $\mathbf{Z} = (Z_1, \dots, Z_n)'$, where $Z_1, \dots, Z_n$ are iid $N(0, 1)$ random variables. Then the density of $\mathbf{Z}$ is

$$\begin{aligned} f_{\mathbf{Z}}(\mathbf{z}) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}z_i^2\right\} = \left(\frac{1}{2\pi}\right)^{n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^n z_i^2\right\} \\ &= \left(\frac{1}{2\pi}\right)^{n/2} \exp\left\{-\frac{1}{2}\mathbf{z}'\mathbf{z}\right\}, \end{aligned} \quad (3.5.5)$$

for $\mathbf{z} \in R^n$. Because the Z_i s have mean 0, have variance 1, and are uncorrelated, the mean and covariance matrix of $\mathbf{Z}$ are

$$E[\mathbf{Z}] = \mathbf{0} \text{ and } \text{Cov}[\mathbf{Z}] = \mathbf{I}_n, \quad (3.5.6)$$

where $\mathbf{I}_n$ denotes the identity matrix of order n . Recall that the mgf of Z_i evaluated at t_i is $\exp\{t_i^2/2\}$. Hence, because the Z_i s are independent, the mgf of $\mathbf{Z}$ is

$$\begin{aligned} M_{\mathbf{Z}}(\mathbf{t}) = E[\exp\{\mathbf{t}'\mathbf{Z}\}] &= E\left[\prod_{i=1}^n \exp\{t_i Z_i\}\right] = \prod_{i=1}^n E[\exp\{t_i Z_i\}] \\ &= \exp\left\{\frac{1}{2}\sum_{i=1}^n t_i^2\right\} = \exp\left\{\frac{1}{2}\mathbf{t}'\mathbf{t}\right\}, \end{aligned} \quad (3.5.7)$$

for all $\mathbf{t} \in R^n$. We say that $\mathbf{Z}$ has a **multivariate normal distribution** with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{I}_n$. We abbreviate this by saying that $\mathbf{Z}$ has an $N_n(\mathbf{0}, \mathbf{I}_n)$ distribution.

For the general case, suppose Σ is an $n \times n$, symmetric, and positive semi-definite matrix. Then from linear algebra, we can always decompose Σ as

$$\Sigma = \mathbf{\Gamma}'\mathbf{\Lambda}\mathbf{\Gamma}, \quad (3.5.8)$$

where $\mathbf{\Lambda}$ is the diagonal matrix $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$ are the eigenvalues of Σ , and the columns of $\mathbf{\Gamma}'$, $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$, are the corresponding eigenvectors. This decomposition is called the **spectral decomposition** of Σ . The matrix $\mathbf{\Gamma}$ is orthogonal, i.e., $\mathbf{\Gamma}^{-1} = \mathbf{\Gamma}'$, and, hence, $\mathbf{\Gamma}\mathbf{\Gamma}' = \mathbf{I}$. As Exercise 3.5.19 shows, we can write the spectral decomposition in another way, as

$$\Sigma = \mathbf{\Gamma}'\mathbf{\Lambda}\mathbf{\Gamma} = \sum_{i=1}^n \lambda_i \mathbf{v}_i \mathbf{v}_i'. \quad (3.5.9)$$

Because the λ_i s are nonnegative, we can define the diagonal matrix $\mathbf{\Lambda}^{1/2} = \text{diag}\{\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}\}$. Then the orthogonality of $\mathbf{\Gamma}$ implies

$$\Sigma = [\mathbf{\Gamma}'\mathbf{\Lambda}^{1/2}\mathbf{\Gamma}][\mathbf{\Gamma}'\mathbf{\Lambda}^{1/2}\mathbf{\Gamma}].$$

We define the matrix product in brackets as the **square root** of the positive semi-definite matrix Σ and write it as

$$\Sigma^{1/2} = \mathbf{\Gamma}'\mathbf{\Lambda}^{1/2}\mathbf{\Gamma}. \quad (3.5.10)$$

Note that $\Sigma^{1/2}$ is symmetric and positive semi-definite. Suppose Σ is positive definite; that is, all of its eigenvalues are strictly positive. Based on this, it is then easy to show that

$$\left(\Sigma^{1/2}\right)^{-1} = \Gamma' \Lambda^{-1/2} \Gamma; \quad (3.5.11)$$

see Exercise 3.5.13. We write the left side of this equation as $\Sigma^{-1/2}$. These matrices enjoy many additional properties of the law of exponents for numbers; see, for example, Arnold (1981). Here, though, all we need are the properties given above.

Suppose $\mathbf{Z}$ has a $N_n(\mathbf{0}, \mathbf{I}_n)$ distribution. Let Σ be a positive semi-definite, symmetric matrix and let $\boldsymbol{\mu}$ be an $n \times 1$ vector of constants. Define the random vector $\mathbf{X}$ by

$$\mathbf{X} = \Sigma^{1/2} \mathbf{Z} + \boldsymbol{\mu}. \quad (3.5.12)$$

By (3.5.6) and Theorem 2.6.3, we immediately have

$$E[\mathbf{X}] = \boldsymbol{\mu} \text{ and } \text{Cov}[\mathbf{X}] = \Sigma^{1/2} \Sigma^{1/2} = \Sigma. \quad (3.5.13)$$

Further, the mgf of $\mathbf{X}$ is given by

$$\begin{aligned} M_{\mathbf{X}}(\mathbf{t}) = E[\exp\{\mathbf{t}'\mathbf{X}\}] &= E\left[\exp\{\mathbf{t}'\Sigma^{1/2}\mathbf{Z} + \mathbf{t}'\boldsymbol{\mu}\}\right] \\ &= \exp\{\mathbf{t}'\boldsymbol{\mu}\} E\left[\exp\left\{\left(\Sigma^{1/2}\mathbf{t}\right)'\mathbf{Z}\right\}\right] \\ &= \exp\{\mathbf{t}'\boldsymbol{\mu}\} \exp\left\{(1/2)\left(\Sigma^{1/2}\mathbf{t}\right)'\Sigma^{1/2}\mathbf{t}\right\} \\ &= \exp\{\mathbf{t}'\boldsymbol{\mu}\} \exp\{(1/2)\mathbf{t}'\Sigma\mathbf{t}\}. \end{aligned} \quad (3.5.14)$$

This leads to the following definition:

Definition 3.5.1 (Multivariate Normal). *We say an n -dimensional random vector $\mathbf{X}$ has a multivariate normal distribution if its mgf is*

$$M_{\mathbf{X}}(\mathbf{t}) = \exp\{\mathbf{t}'\boldsymbol{\mu} + (1/2)\mathbf{t}'\Sigma\mathbf{t}\}, \text{ for all } \mathbf{t} \in R^n. \quad (3.5.15)$$

where Σ is a symmetric, positive semi-definite matrix and $\boldsymbol{\mu} \in R^n$. We abbreviate this by saying that $\mathbf{X}$ has a $N_n(\boldsymbol{\mu}, \Sigma)$ distribution.

Note that our definition is for positive semi-definite matrices Σ . Usually Σ is positive definite, in which case we can further obtain the density of $\mathbf{X}$. If Σ is positive definite, then so is $\Sigma^{1/2}$ and, as discussed above, its inverse is given by expression (3.5.11). Thus the transformation between $\mathbf{X}$ and $\mathbf{Z}$, (3.5.12), is one-to-one with the inverse transformation

$$\mathbf{Z} = \Sigma^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$$

and the Jacobian $|\Sigma^{-1/2}| = |\Sigma|^{-1/2}$. Hence, upon simplification, the pdf of $\mathbf{X}$ is given by

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\}, \text{ for } \mathbf{x} \in R^n. \quad (3.5.16)$$

In Section 3.5.1, we discussed the contours of the bivariate normal distribution. We now extend that discussion to the general case, adding probabilities to the contours. Let $\mathbf{X}$ have a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution. In the n -dimensional case, the contours of constant probability for the pdf of $\mathbf{X}$, (3.5.16), are the ellipsoids

$$(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = c^2,$$

for $c > 0$. Define the random variable $Y = (\mathbf{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu})$. Then using expression (3.5.12), we have

$$Y = \mathbf{Z}' \boldsymbol{\Sigma}^{1/2} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{1/2} \mathbf{Z} = \mathbf{Z}' \mathbf{Z} = \sum_{i=1}^n Z_i^2.$$

Since $Z_1, \dots, Z_n$ are iid $N(0, 1)$, Y has χ^2 -distribution with n degrees of freedom. Denote the cdf of Y by $F_{\chi_n^2}$. Then we have

$$P[(\mathbf{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \leq c^2] = P(Y \leq c^2) = F_{\chi_n^2}(c^2). \quad (3.5.17)$$

These probabilities are often used to label the contour plots; see Exercise 3.5.5. For reference, we summarize the above proof in the following theorem. Note that this theorem is a generalization of the univariate result given in Theorem 3.4.1.

Theorem 3.5.1. *Suppose $\mathbf{X}$ has a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, where $\boldsymbol{\Sigma}$ is positive definite. Then the random variable $Y = (\mathbf{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu})$ has a $\chi^2(n)$ distribution.*

The following two theorems are very useful. The first says that a linear transformation of a multivariate normal random vector has a multivariate normal distribution.

Theorem 3.5.2. *Suppose $\mathbf{X}$ has a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution. Let $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$, where $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{b} \in R^m$. Then $\mathbf{Y}$ has a $N_m(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$ distribution.*

Proof: From (3.5.15), for $\mathbf{t} \in R^m$, the mgf of $\mathbf{Y}$ is

$$\begin{aligned} M_{\mathbf{Y}}(\mathbf{t}) &= E[\exp\{\mathbf{t}'\mathbf{Y}\}] \\ &= E[\exp\{\mathbf{t}'(\mathbf{A}\mathbf{X} + \mathbf{b})\}] \\ &= \exp\{\mathbf{t}'\mathbf{b}\} E[\exp\{(\mathbf{A}'\mathbf{t})'\mathbf{X}\}] \\ &= \exp\{\mathbf{t}'\mathbf{b}\} \exp\{(\mathbf{A}'\mathbf{t})'\boldsymbol{\mu} + (1/2)(\mathbf{A}'\mathbf{t})'\boldsymbol{\Sigma}(\mathbf{A}'\mathbf{t})\} \\ &= \exp\{\mathbf{t}'(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}) + (1/2)\mathbf{t}'\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\mathbf{t}\}, \end{aligned}$$

which is the mgf of an $N_m(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$ distribution. ■

A simple corollary to this theorem gives marginal distributions of a multivariate normal random variable. Let $\mathbf{X}_1$ be any subvector of $\mathbf{X}$, say of dimension $m < n$. Because we can always rearrange means and correlations, there is no loss in generality in writing $\mathbf{X}$ as

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}, \quad (3.5.18)$$

where $\mathbf{X}_2$ is of dimension $p = n - m$. In the same way, partition the mean and covariance matrix of $\mathbf{X}$; that is,

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \quad (3.5.19)$$

with the same dimensions as in expression (3.5.18). Note, for instance, that $\boldsymbol{\Sigma}_{11}$ is the covariance matrix of $\mathbf{X}_1$ and $\boldsymbol{\Sigma}_{12}$ contains all the covariances between the components of $\mathbf{X}_1$ and $\mathbf{X}_2$. Now define $\mathbf{A}$ to be the matrix

$$\mathbf{A} = [\mathbf{I}_m : \mathbf{O}_{mp}],$$

where $\mathbf{O}_{mp}$ is an $m \times p$ matrix of zeroes. Then $\mathbf{X}_1 = \mathbf{A}\mathbf{X}$. Hence, applying Theorem 3.5.2 to this transformation, along with some matrix algebra, we have the following corollary:

Corollary 3.5.1. *Suppose $\mathbf{X}$ has a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, partitioned as in expressions (3.5.18) and (3.5.19). Then $\mathbf{X}_1$ has a $N_m(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ distribution.*

This is a useful result because it says that any marginal distribution of $\mathbf{X}$ is also normal and, further, its mean and covariance matrix are those associated with that partial vector.

Recall in Section 2.5, Theorem 2.5.2, that if two random variables are independent then their covariance is 0. In general, the converse is not true. However, as the following theorem shows, it is true for the multivariate normal distribution.

Theorem 3.5.3. *Suppose $\mathbf{X}$ has a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, partitioned as in the expressions (3.5.18) and (3.5.19). Then $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent if and only if $\boldsymbol{\Sigma}_{12} = \mathbf{O}$.*

Proof: First note that $\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}'_{12}$. The joint mgf of $\mathbf{X}_1$ and $\mathbf{X}_2$ is given by

$$M_{\mathbf{X}_1, \mathbf{X}_2}(\mathbf{t}_1, \mathbf{t}_2) = \exp \left\{ \mathbf{t}'_1 \boldsymbol{\mu}_1 + \mathbf{t}'_2 \boldsymbol{\mu}_2 + \frac{1}{2} (\mathbf{t}'_1 \boldsymbol{\Sigma}_{11} \mathbf{t}_1 + \mathbf{t}'_2 \boldsymbol{\Sigma}_{22} \mathbf{t}_2 + \mathbf{t}'_2 \boldsymbol{\Sigma}_{21} \mathbf{t}_1 + \mathbf{t}'_1 \boldsymbol{\Sigma}_{12} \mathbf{t}_2) \right\} \quad (3.5.20)$$

where $\mathbf{t}' = (\mathbf{t}'_1, \mathbf{t}'_2)$ is partitioned the same as $\boldsymbol{\mu}$. By Corollary 3.5.1, $\mathbf{X}_1$ has a $N_m(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ distribution and $\mathbf{X}_2$ has a $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$ distribution. Hence, the product of their marginal mgfs is

$$M_{\mathbf{X}_1}(\mathbf{t}_1) M_{\mathbf{X}_2}(\mathbf{t}_2) = \exp \left\{ \mathbf{t}'_1 \boldsymbol{\mu}_1 + \mathbf{t}'_2 \boldsymbol{\mu}_2 + \frac{1}{2} (\mathbf{t}'_1 \boldsymbol{\Sigma}_{11} \mathbf{t}_1 + \mathbf{t}'_2 \boldsymbol{\Sigma}_{22} \mathbf{t}_2) \right\}. \quad (3.5.21)$$

By (2.6.6) of Section 2.6, $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent if and only if the expressions (3.5.20) and (3.5.21) are the same. If $\boldsymbol{\Sigma}_{12} = \mathbf{O}'$ and, hence, $\boldsymbol{\Sigma}_{21} = \mathbf{O}$, then the expressions are the same and $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent. If $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent, then the covariances between their components are all 0; i.e., $\boldsymbol{\Sigma}_{12} = \mathbf{O}'$ and $\boldsymbol{\Sigma}_{21} = \mathbf{O}$. ■

Corollary 3.5.1 showed that the marginal distributions of a multivariate normal are themselves normal. This is true for conditional distributions, too. As the

following proof shows, we can combine the results of Theorems 3.5.2 and 3.5.3 to obtain the following theorem.

Theorem 3.5.4. *Suppose $\mathbf{X}$ has a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, which is partitioned as in expressions (3.5.18) and (3.5.19). Assume that $\boldsymbol{\Sigma}$ is positive definite. Then the conditional distribution of $\mathbf{X}_1 | \mathbf{X}_2$ is*

$$N_m(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}). \quad (3.5.22)$$

Proof: Consider first the joint distribution of the random vector $\mathbf{W} = \mathbf{X}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{X}_2$ and $\mathbf{X}_2$. This distribution is obtained from the transformation

$$\begin{bmatrix} \mathbf{W} \\ \mathbf{X}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{I}_m & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \mathbf{O} & \mathbf{I}_p \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}.$$

Because this is a linear transformation, it follows from Theorem 3.5.2 that the joint distribution is multivariate normal, with $E[\mathbf{W}] = \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2$, $E[\mathbf{X}_2] = \boldsymbol{\mu}_2$, and covariance matrix

$$\begin{bmatrix} \mathbf{I}_m & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \mathbf{O} & \mathbf{I}_p \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I}_m & \mathbf{O}' \\ -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \mathbf{I}_p \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \mathbf{O}' \\ \mathbf{O} & \boldsymbol{\Sigma}_{22} \end{bmatrix}.$$

Hence, by Theorem 3.5.3 the random vectors $\mathbf{W}$ and $\mathbf{X}_2$ are independent. Thus the conditional distribution of $\mathbf{W} | \mathbf{X}_2$ is the same as the marginal distribution of $\mathbf{W}$; that is,

$$\mathbf{W} | \mathbf{X}_2 \text{ is } N_m(\boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}).$$

Further, because of this independence, $\mathbf{W} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{X}_2$ given $\mathbf{X}_2$ is distributed as

$$N_m(\boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{X}_2, \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}), \quad (3.5.23)$$

which is the desired result. ■

In the following remark, we return to the bivariate normal using the above general notation.

Remark 3.5.1 (Continuation of the Bivariate Normal). Suppose (X, Y) has a $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, where

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}. \quad (3.5.24)$$

Substituting $\rho\sigma_1\sigma_2$ for σ_{12} in $\boldsymbol{\Sigma}$, it is easy to see that the determinant of $\boldsymbol{\Sigma}$ is $\sigma_1^2\sigma_2^2(1 - \rho^2)$. Recall that $\rho^2 \leq 1$. For the remainder of this remark, assume that $\rho^2 < 1$. In this case, $\boldsymbol{\Sigma}$ is invertible (it is also positive definite). Further, since $\boldsymbol{\Sigma}$ is a 2×2 matrix, its inverse can easily be determined to be

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1 - \rho^2)} \begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}. \quad (3.5.25)$$

This shows the equivalence of the bivariate normal pdf notation, (3.5.1), and the general multivariate normal distribution with $n = 2$ pdf notation, (3.5.16).

To simplify the conditional normal distribution (3.5.22) for the bivariate case, consider once more the bivariate normal distribution that was given in Section 3.5.1. For this case, reversing the roles so that $Y = X_1$ and $X = X_2$, expression (3.5.22) shows that the conditional distribution of Y given $X = x$ is

$$N \left[\mu_2 + \rho \frac{\sigma_2}{\sigma_1} (x - \mu_1), \sigma_2^2 (1 - \rho^2) \right]. \quad (3.5.26)$$

Thus, with a bivariate normal distribution, the conditional mean of Y , given that $X = x$, is linear in x and is given by

$$E(Y|x) = \mu_2 + \rho \frac{\sigma_2}{\sigma_1} (x - \mu_1).$$

Although the mean of the conditional distribution of Y , given $X = x$, depends upon x (unless $\rho = 0$), the variance $\sigma_2^2(1 - \rho^2)$ is the same for all real values of x . Thus, by way of example, given that $X = x$, the conditional probability that Y is within $(2.576)\sigma_2\sqrt{1 - \rho^2}$ units of the conditional mean is 0.99, whatever the value of x may be. In this sense, most of the probability for the distribution of X and Y lies in the band

$$\mu_2 + \rho \frac{\sigma_2}{\sigma_1} (x - \mu_1) \pm 2.576\sigma_2\sqrt{1 - \rho^2}$$

about the graph of the linear conditional mean. For every fixed positive σ_2 , the width of this band depends upon ρ . Because the band is narrow when ρ^2 is nearly 1, we see that ρ does measure the intensity of the concentration of the probability for X and Y about the linear conditional mean. We alluded to this fact in the remark of Section 2.5.

In a similar manner we can show that the conditional distribution of X , given $Y = y$, is the normal distribution

$$N \left[\mu_1 + \rho \frac{\sigma_1}{\sigma_2} (y - \mu_2), \sigma_1^2 (1 - \rho^2) \right]. \quad \blacksquare$$

Example 3.5.1. Let us assume that in a certain population of married couples the height X_1 of the husband and the height X_2 of the wife have a bivariate normal distribution with parameters $\mu_1 = 5.8$ feet, $\mu_2 = 5.3$ feet, $\sigma_1 = \sigma_2 = 0.2$ foot, and $\rho = 0.6$. The conditional pdf of X_2 , given $X_1 = 6.3$, is normal, with mean $5.3 + (0.6)(6.3 - 5.8) = 5.6$ and standard deviation $(0.2)\sqrt{(1 - 0.36)} = 0.16$. Accordingly, given that the height of the husband is 6.3 feet, the probability that his wife has a height between 5.28 and 5.92 feet is

$$P(5.28 < X_2 < 5.92 | X_1 = 6.3) = \Phi(2) - \Phi(-2) = 0.954.$$

The interval $(5.28, 5.92)$ could be thought of as a 95.4% *prediction interval* for the wife's height, given $X_1 = 6.3$. $\blacksquare$

3.5.3 *Applications

In this section, we consider several applications of the multivariate normal distribution. These the reader may have already encountered in an applied course in statistics. The first is *principal components*, which results in a linear function of a multivariate normal random vector that has independent components and preserves the “total” variation in the problem.

Let the random vector $\mathbf{X}$ have the multivariate normal distribution $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma}$ is positive definite. As in (3.5.8), write the spectral decomposition of $\boldsymbol{\Sigma}$ as $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}'\boldsymbol{\Lambda}\boldsymbol{\Gamma}$. Recall that the columns, $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$, of $\boldsymbol{\Gamma}'$ are the eigenvectors corresponding to the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ that form the main diagonal of the matrix $\boldsymbol{\Lambda}$. Assume without loss of generality that the eigenvalues are decreasing; i.e., $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n > 0$. Define the random vector $\mathbf{Y} = \boldsymbol{\Gamma}(\mathbf{X} - \boldsymbol{\mu})$. Since $\boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}' = \boldsymbol{\Lambda}$, by Theorem 3.5.2 $\mathbf{Y}$ has a $N_n(\mathbf{0}, \boldsymbol{\Lambda})$ distribution. Hence the components $Y_1, Y_2, \dots, Y_n$ are independent random variables and, for $i = 1, 2, \dots, n$, Y_i has a $N(0, \lambda_i)$ distribution. The random vector $\mathbf{Y}$ is called the vector of **principal components**.

We say the **total variation**, (TV), of a random vector is the sum of the variances of its components. For the random vector $\mathbf{X}$, because $\boldsymbol{\Gamma}$ is an orthogonal matrix

$$\text{TV}(\mathbf{X}) = \sum_{i=1}^n \sigma_i^2 = \text{tr } \boldsymbol{\Sigma} = \text{tr } \boldsymbol{\Gamma}'\boldsymbol{\Lambda}\boldsymbol{\Gamma} = \text{tr } \boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{\Gamma}' = \sum_{i=1}^n \lambda_i = \text{TV}(\mathbf{Y}).$$

Hence, $\mathbf{X}$ and $\mathbf{Y}$ have the same total variation.

Next, consider the first component of $\mathbf{Y}$, which is given by $Y_1 = \mathbf{v}_1'(\mathbf{X} - \boldsymbol{\mu})$. This is a linear combination of the components of $\mathbf{X} - \boldsymbol{\mu}$ with the property $\|\mathbf{v}_1\|^2 = \sum_{j=1}^n v_{1j}^2 = 1$, because $\boldsymbol{\Gamma}'$ is orthogonal. Consider any other linear combination of $(\mathbf{X} - \boldsymbol{\mu})$, say $\mathbf{a}'(\mathbf{X} - \boldsymbol{\mu})$ such that $\|\mathbf{a}\|^2 = 1$. Because $\mathbf{a} \in R^n$ and $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ forms a basis for R^n , we must have $\mathbf{a} = \sum_{j=1}^n a_j \mathbf{v}_j$ for some set of scalars $a_1, \dots, a_n$. Furthermore, because the basis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is orthonormal

$$\mathbf{a}'\mathbf{v}_i = \left(\sum_{j=1}^n a_j \mathbf{v}_j \right)' \mathbf{v}_i = \sum_{j=1}^n a_j \mathbf{v}_j' \mathbf{v}_i = a_i.$$

Using (3.5.9) and the fact that $\lambda_i > 0$, we have the inequality

$$\begin{aligned} \text{Var}(\mathbf{a}'\mathbf{X}) &= \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} \\ &= \sum_{i=1}^n \lambda_i (\mathbf{a}'\mathbf{v}_i)^2 \\ &= \sum_{i=1}^n \lambda_i a_i^2 \leq \lambda_1 \sum_{i=1}^n a_i^2 = \lambda_1 = \text{Var}(Y_1). \end{aligned} \quad (3.5.27)$$

Hence, Y_1 has the maximum variance of any linear combination $\mathbf{a}'(\mathbf{X} - \boldsymbol{\mu})$, such that $\|\mathbf{a}\| = 1$. For this reason, Y_1 is called the **first principal component** of $\mathbf{X}$.

What about the other components, $Y_2, \dots, Y_n$? As the following theorem shows, they share a similar property relative to the order of their associated eigenvalue. For this reason, they are called the **second**, **third**, through the **n th principal components**, respectively.

Theorem 3.5.5. *Consider the situation described above. For $j = 2, \dots, n$ and $i = 1, 2, \dots, j-1$, $\text{Var}[\mathbf{a}'\mathbf{X}] \leq \lambda_j = \text{Var}(Y_j)$, for all vectors $\mathbf{a}$ such that $\mathbf{a} \perp \mathbf{v}_i$ and $\|\mathbf{a}\| = 1$.*

The proof of this theorem is similar to that for the first principal component and is left as Exercise 3.5.20. A second application concerning linear regression is offered in Exercise 3.5.22.

EXERCISES

3.5.1. Let X and Y have a bivariate normal distribution with respective parameters $\mu_x = 2.8$, $\mu_y = 110$, $\sigma_x^2 = 0.16$, $\sigma_y^2 = 100$, and $\rho = 0.6$. Using R, compute:

- (a) $P(106 < Y < 124)$.
- (b) $P(106 < Y < 124 | X = 3.2)$.

3.5.2. Let X and Y have a bivariate normal distribution with parameters $\mu_1 = 3$, $\mu_2 = 1$, $\sigma_1^2 = 16$, $\sigma_2^2 = 25$, and $\rho = \frac{3}{5}$. Using R, determine the following probabilities:

- (a) $P(3 < Y < 8)$.
- (b) $P(3 < Y < 8 | X = 7)$.
- (c) $P(-3 < X < 3)$.
- (d) $P(-3 < X < 3 | Y = -4)$.

3.5.3. Show that expression (3.5.4) is true.

3.5.4. Let $f(x, y)$ be the bivariate normal pdf in expression (3.5.1).

- (a) Show that $f(x, y)$ has an unique maximum at (μ_1, μ_2) .
- (b) For a given $c > 0$, show that the points $\{(x, y) : f(x, y) = c\}$ of equal probability form an ellipse.

3.5.5. Let $\mathbf{X}$ be $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Recall expression (3.5.17) which gives the probability of an elliptical contour region for $\mathbf{X}$. The R function⁸ `ellipmake` plots the elliptical contour regions. To graph the elliptical 95% contour for a multivariate normal distribution with $\boldsymbol{\mu} = (5, 2)'$ and $\boldsymbol{\Sigma}$ with variances 1 and covariance 0.75, use the code

⁸Part of this code was obtained from an anonymous author at the site <http://stats.stackexchange.com/questions/9898/>

```
ellipmake(p=.95,b=matrix(c(1,.75,.75,1),nrow=2),mu=c(5,2)).
```

This R function can be found at the site listed in the Preface.

- (a) Run the above code.
- (b) Change the code so the probability is 0.50.
- (c) Change the code to obtain an overlay plot of the 0.50 and 0.95 regions.
- (d) Using a loop, obtain the overlay plot for a vector of probabilities.

3.5.6. Let U and V be independent random variables, each having a standard normal distribution. Show that the mgf $E(e^{t(UV)})$ of the random variable UV is $(1 - t^2)^{-1/2}$, $-1 < t < 1$.

Hint: Compare $E(e^{tUV})$ with the integral of a bivariate normal pdf that has means equal to zero.

3.5.7. Let X and Y have a bivariate normal distribution with parameters $\mu_1 = 5$, $\mu_2 = 10$, $\sigma_1^2 = 1$, $\sigma_2^2 = 25$, and $\rho > 0$. If $P(4 < Y < 16 | X = 5) = 0.954$, determine ρ .

3.5.8. Let X and Y have a bivariate normal distribution with parameters $\mu_1 = 20$, $\mu_2 = 40$, $\sigma_1^2 = 9$, $\sigma_2^2 = 4$, and $\rho = 0.6$. Find the shortest interval for which 0.90 is the conditional probability that Y is in the interval, given that $X = 22$.

3.5.9. Say the correlation coefficient between the heights of husbands and wives is 0.70 and the mean male height is 5 feet 10 inches with standard deviation 2 inches, and the mean female height is 5 feet 4 inches with standard deviation $1\frac{1}{2}$ inches. Assuming a bivariate normal distribution, what is the best guess of the height of a woman whose husband's height is 6 feet? Find a 95% prediction interval for her height.

3.5.10. Let

$$f(x, y) = (1/2\pi) \exp \left[-\frac{1}{2}(x^2 + y^2) \right] \left\{ 1 + xy \exp \left[-\frac{1}{2}(x^2 + y^2 - 2) \right] \right\},$$

where $-\infty < x < \infty$, $-\infty < y < \infty$. If $f(x, y)$ is a joint pdf, it is not a normal bivariate pdf. Show that $f(x, y)$ actually is a joint pdf and that each marginal pdf is normal. Thus the fact that each marginal pdf is normal does not imply that the joint pdf is bivariate normal.

3.5.11. Let X , Y , and Z have the joint pdf

$$\left(\frac{1}{2\pi} \right)^{3/2} \exp \left(-\frac{x^2 + y^2 + z^2}{2} \right) \left[1 + xyz \exp \left(-\frac{x^2 + y^2 + z^2}{2} \right) \right],$$

where $-\infty < x < \infty$, $-\infty < y < \infty$, and $-\infty < z < \infty$. While X , Y , and Z are obviously dependent, show that X , Y , and Z are pairwise independent and that each pair has a bivariate normal distribution.

3.5.12. Let X and Y have a bivariate normal distribution with parameters $\mu_1 = \mu_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 1$, and correlation coefficient ρ . Find the distribution of the random variable $Z = aX + bY$ in which a and b are nonzero constants.

3.5.13. Establish formula (3.5.11) by a direct multiplication.

3.5.14. Let $\mathbf{X} = (X_1, X_2, X_3)$ have a multivariate normal distribution with mean vector $\mathbf{0}$ and variance-covariance matrix

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

Find $P(X_1 > X_2 + X_3 + 2)$.

Hint: Find the vector $\mathbf{a}$ so that $\mathbf{aX} = X_1 - X_2 - X_3$ and make use of Theorem 3.5.2.

3.5.15. Suppose $\mathbf{X}$ is distributed $N_n(\boldsymbol{\mu}, \Sigma)$. Let $\bar{X} = n^{-1} \sum_{i=1}^n X_i$.

(a) Write $\bar{X}$ as $\mathbf{aX}$ for an appropriate vector $\mathbf{a}$ and apply Theorem 3.5.2 to find the distribution of $\bar{X}$.

(b) Determine the distribution of $\bar{X}$ if all of its component random variables X_i have the same mean μ .

3.5.16. Suppose $\mathbf{X}$ is distributed $N_2(\boldsymbol{\mu}, \Sigma)$. Determine the distribution of the random vector $(X_1 + X_2, X_1 - X_2)$. Show that $X_1 + X_2$ and $X_1 - X_2$ are independent if $\text{Var}(X_1) = \text{Var}(X_2)$.

3.5.17. Suppose $\mathbf{X}$ is distributed $N_3(\mathbf{0}, \Sigma)$, where

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

Find $P((X_1 - 2X_2 + X_3)^2 > 15.36)$.

3.5.18. Let X_1, X_2, X_3 be iid random variables each having a standard normal distribution. Let the random variables Y_1, Y_2, Y_3 be defined by

$$X_1 = Y_1 \cos Y_2 \sin Y_3, \quad X_2 = Y_1 \sin Y_2 \sin Y_3, \quad X_3 = Y_1 \cos Y_3,$$

where $0 \leq Y_1 < \infty$, $0 \leq Y_2 < 2\pi$, $0 \leq Y_3 \leq \pi$. Show that Y_1, Y_2, Y_3 are mutually independent.

3.5.19. Show that expression (3.5.9) is true.

3.5.20. Prove Theorem 3.5.5.

3.5.21. Suppose $\mathbf{X}$ has a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix

$$\Sigma = \begin{bmatrix} 283 & 215 & 277 & 208 \\ 215 & 213 & 217 & 153 \\ 277 & 217 & 336 & 236 \\ 208 & 153 & 236 & 194 \end{bmatrix}.$$

- (a) Find the total variation of $\mathbf{X}$.
- (b) Find the principal component vector $\mathbf{Y}$.
- (c) Show that the first principal component accounts for 90% of the total variation.
- (d) Show that the first principal component Y_1 is essentially a rescaled $\overline{X}$. Determine the variance of $(1/2)\overline{X}$ and compare it to that of Y_1 .

Note that the R command `eigen(amat)` obtains the spectral decomposition of the matrix `amat`.

3.5.22. Readers may have encountered the multiple regression model in a previous course in statistics. We can briefly write it as follows. Suppose we have a vector of n observations $\mathbf{Y}$ which has the distribution $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, where $\mathbf{X}$ is an $n \times p$ matrix of known values, which has full column rank p , and $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters. The least squares estimator of $\boldsymbol{\beta}$ is

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

- (a) Determine the distribution of $\hat{\boldsymbol{\beta}}$.
- (b) Let $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$. Determine the distribution of $\hat{\mathbf{Y}}$.
- (c) Let $\hat{\mathbf{e}} = \mathbf{Y} - \hat{\mathbf{Y}}$. Determine the distribution of $\hat{\mathbf{e}}$.
- (d) By writing the random vector $(\hat{\mathbf{Y}}', \hat{\mathbf{e}}')'$ as a linear function of $\mathbf{Y}$, show that the random vectors $\hat{\mathbf{Y}}$ and $\hat{\mathbf{e}}$ are independent.
- (e) Show that $\hat{\boldsymbol{\beta}}$ solves the least squares problem; that is,

$$\|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2 = \min_{\mathbf{b} \in R^p} \|\mathbf{Y} - \mathbf{X}\mathbf{b}\|^2.$$

3.6 t - and F -Distributions

It is the purpose of this section to define two additional distributions that are quite useful in certain problems of statistical inference. These are called, respectively, the (Student's) t -distribution and the F -distribution.

3.6.1 The t -distribution

Let W denote a random variable that is $N(0, 1)$; let V denote a random variable that is $\chi^2(r)$; and let W and V be independent. Then the joint pdf of W and V , say $h(w, v)$, is the product of the pdf of W and that of V or

$$h(w, v) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-w^2/2} \frac{1}{\Gamma(r/2)2^{r/2}} v^{r/2-1} e^{-v/2} & -\infty < w < \infty, \quad 0 < v < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

Define a new random variable T by writing

$$T = \frac{W}{\sqrt{V/r}}. \quad (3.6.1)$$

The transformation technique is used to obtain the pdf $g_1(t)$ of T . The equations

$$t = \frac{w}{\sqrt{v/r}} \quad \text{and} \quad u = v$$

define a transformation that maps $\mathcal{S} = \{(w, v) : -\infty < w < \infty, 0 < v < \infty\}$ one-to-one and onto $\mathcal{T} = \{(t, u) : -\infty < t < \infty, 0 < u < \infty\}$. Since $w = t\sqrt{u}/\sqrt{r}$, $v = u$, the absolute value of the Jacobian of the transformation is $|J| = \sqrt{u}/\sqrt{r}$. Accordingly, the joint pdf of T and $U = V$ is given by

$$\begin{aligned} g(t, u) &= h\left(\frac{t\sqrt{u}}{\sqrt{r}}, u\right)|J| \\ &= \begin{cases} \frac{1}{\sqrt{2\pi}\Gamma(r/2)2^{r/2}} u^{r/2-1} \exp\left[-\frac{u}{2}\left(1 + \frac{t^2}{r}\right)\right] \frac{\sqrt{u}}{\sqrt{r}} & |t| < \infty, 0 < u < \infty \\ 0 & \text{elsewhere.} \end{cases} \end{aligned}$$

The marginal pdf of T is then

$$\begin{aligned} g_1(t) &= \int_{-\infty}^{\infty} g(t, u) du \\ &= \int_0^{\infty} \frac{1}{\sqrt{2\pi r}\Gamma(r/2)2^{r/2}} u^{(r+1)/2-1} \exp\left[-\frac{u}{2}\left(1 + \frac{t^2}{r}\right)\right] du. \end{aligned}$$

In this integral let $z = u[1 + (t^2/r)]/2$, and it is seen that

$$\begin{aligned} g_1(t) &= \int_0^{\infty} \frac{1}{\sqrt{2\pi r}\Gamma(r/2)2^{r/2}} \left(\frac{2z}{1 + t^2/r}\right)^{(r+1)/2-1} e^{-z} \left(\frac{2}{1 + t^2/r}\right) dz \\ &= \frac{\Gamma[(r+1)/2]}{\sqrt{\pi r}\Gamma(r/2)} \frac{1}{(1 + t^2/r)^{(r+1)/2}}, \quad -\infty < t < \infty. \end{aligned} \quad (3.6.2)$$

Thus, if W is $N(0, 1)$, V is $\chi^2(r)$, and W and V are independent, then $T = W/\sqrt{V/r}$ has the pdf $g_1(t)$, (3.6.2). The distribution of the random variable T is usually called a ***t*-distribution**. It should be observed that a *t*-distribution is completely determined by the parameter r , the number of degrees of freedom of the random variable that has the chi-square distribution.

The pdf $g_1(t)$ satisfies $g_1(-t) = g_1(t)$; hence, the pdf of T is symmetric about 0. Thus, the median of T is 0. Upon differentiating $g_1(t)$, it follows that the unique maximum of the pdf occurs at 0 and that the derivative is continuous. So, the pdf is mound shaped. As the degrees of freedom approach ∞ , the *t*-distribution converges to the $N(0, 1)$ distribution; see Example 5.2.3 of Chapter 5.

The R command `pt(t, r)` computes the probability $P(T \leq t)$ when T has a *t*-distribution with r degrees of freedom. For instance, the probability that a *t*-distributed random variable with 15 degrees of freedom is less than 2.0 is computed

as `pt(2.0,15)`, while the command `qt(.975,15)` returns the 97.5th percentile of this distribution. The R code `t=seq(-4,4,.01)` followed by `plot(dt(t,3)~t)` yields a plot of the t -pdf with 3 degrees of freedom.

Before the age of modern computing, tables of the distribution of T were used. Because the pdf of T does depend on its degrees of freedom r , the usual t -table gives selected quantiles versus degrees of freedom. Table III in Appendix D is such a table. The following three lines of R code, however, produce this table.

```
ps = c(.9, .925, .950, .975, .99, .995, .999); df = 1:30; tab=c()
for(r in df){tab=rbind(tab,qt(ps,r))}; df=c(df,Inf);nq=qnorm(ps)
tab=rbind(tab,nq);tab=cbind(df,tab)
```

This code is the body of the R function `ttable` found at the site listed in the Preface. Due to the fact that t -distribution converges to the $N(0,1)$ distribution, only the degrees of freedom from 1 to 30 are used in such tables. This is, also, the reason that the last line in the table are the standard normal quantiles.

Remark 3.6.1. The t -distribution was first discovered by W. S. Gosset when he was working for an Irish brewery. Gosset published under the pseudonym Student. Thus this distribution is often known as **Student's t -distribution**. ■

Example 3.6.1 (Mean and Variance of the t -Distribution). Let the random variable T have a t -distribution with r degrees of freedom. Then, as in (3.6.1), we can write $T = W(V/r)^{-1/2}$, where W has a $N(0,1)$ distribution, V has a $\chi^2(r)$ distribution, and W and V are independent random variables. Independence of W and V and expression (3.3.8), provided $(r/2) - (k/2) > 0$ (i.e., $k < r$), implies the following:

$$E(T^k) = E \left[W^k \left(\frac{V}{r} \right)^{-k/2} \right] = E(W^k) E \left[\left(\frac{V}{r} \right)^{-k/2} \right] \quad (3.6.3)$$

$$= E(W^k) \frac{2^{-k/2} \Gamma(\frac{r}{2} - \frac{k}{2})}{\Gamma(\frac{r}{2}) r^{-k/2}} \quad \text{if } k < r. \quad (3.6.4)$$

Because $E(W) = 0$, the mean of T is 0, as long as the degrees of freedom of T exceed 1. For the variance, use $k = 2$ in expression (3.6.4). In this case the condition $r > k$ becomes $r > 2$. Since $E(W^2) = 1$, by expression (3.6.4), the variance of T is given by

$$\text{Var}(T) = E(T^2) = \frac{r}{r-2}. \quad (3.6.5)$$

Therefore, a t -distribution with $r > 2$ degrees of freedom has a mean of 0 and a variance of $r/(r-2)$. ■

3.6.2 The F -distribution

Next consider two independent chi-square random variables U and V having r_1 and r_2 degrees of freedom, respectively. The joint pdf $h(u,v)$ of U and V is then

$$h(u,v) = \begin{cases} \frac{1}{\Gamma(r_1/2)\Gamma(r_2/2)2^{(r_1+r_2)/2}} u^{r_1/2-1} v^{r_2/2-1} e^{-(u+v)/2} & 0 < u, v < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

We define the new random variable

$$W = \frac{U/r_1}{V/r_2}$$

and we propose finding the pdf $g_1(w)$ of W . The equations

$$w = \frac{u/r_1}{v/r_2}, \quad z = v,$$

define a one-to-one transformation that maps the set $\mathcal{S} = \{(u, v) : 0 < u < \infty, 0 < v < \infty\}$ onto the set $\mathcal{T} = \{(w, z) : 0 < w < \infty, 0 < z < \infty\}$. Since $u = (r_1/r_2)zw$, $v = z$, the absolute value of the Jacobian of the transformation is $|J| = (r_1/r_2)z$. The joint pdf $g(w, z)$ of the random variables W and $Z = V$ is then

$$g(w, z) = \frac{1}{\Gamma(r_1/2)\Gamma(r_2/2)2^{(r_1+r_2)/2}} \left(\frac{r_1zw}{r_2}\right)^{\frac{r_1-2}{2}} z^{\frac{r_2-2}{2}} \exp\left[-\frac{z}{2}\left(\frac{r_1w}{r_2} + 1\right)\right] \frac{r_1z}{r_2},$$

provided that $(w, z) \in \mathcal{T}$, and zero elsewhere. The marginal pdf $g_1(w)$ of W is then

$$\begin{aligned} g_1(w) &= \int_{-\infty}^{\infty} g(w, z) dz \\ &= \int_0^{\infty} \frac{(r_1/r_2)^{r_1/2}(w)^{r_1/2-1}}{\Gamma(r_1/2)\Gamma(r_2/2)2^{(r_1+r_2)/2}} z^{(r_1+r_2)/2-1} \exp\left[-\frac{z}{2}\left(\frac{r_1w}{r_2} + 1\right)\right] dz. \end{aligned}$$

If we change the variable of integration by writing

$$y = \frac{z}{2} \left(\frac{r_1w}{r_2} + 1\right),$$

it can be seen that

$$\begin{aligned} g_1(w) &= \int_0^{\infty} \frac{(r_1/r_2)^{r_1/2}(w)^{r_1/2-1}}{\Gamma(r_1/2)\Gamma(r_2/2)2^{(r_1+r_2)/2}} \left(\frac{2y}{r_1w/r_2 + 1}\right)^{(r_1+r_2)/2-1} e^{-y} \\ &\quad \times \left(\frac{2}{r_1w/r_2 + 1}\right) dy \\ &= \begin{cases} \frac{\Gamma[(r_1+r_2)/2](r_1/r_2)^{r_1/2}}{\Gamma(r_1/2)\Gamma(r_2/2)} \frac{w^{r_1/2-1}}{(1+r_1w/r_2)^{(r_1+r_2)/2}} & 0 < w < \infty \\ 0 & \text{elsewhere.} \end{cases} \quad (3.6.6) \end{aligned}$$

Accordingly, if U and V are independent chi-square variables with r_1 and r_2 degrees of freedom, respectively, then $W = (U/r_1)/(V/r_2)$ has the pdf $g_1(w)$, (3.6.6). The distribution of this random variable is usually called an ***F*-distribution**; and we often call the ratio, which we have denoted by W , F . That is,

$$F = \frac{U/r_1}{V/r_2}. \quad (3.6.7)$$

It should be observed that an *F*-distribution is completely determined by the two parameters r_1 and r_2 .

In terms of R computation, the command `pf(2.50, 3, 8)` computes to the value 0.8665 which is the probability $P(F \leq 2.50)$ when F has the F -distribution with 3 and 8 degrees of freedom. The 95th percentile of F is `qf(.95, 3, 8) = 4.066` and the code `x=seq(.01, 5, .01); plot(df(x, 3, 8)~x)` draws a plot of the pdf of this F random variable. Note that the pdf is right-skewed. Before the age of modern computation, tables of the quantiles of F -distributions for selected probabilities and degrees of freedom were used. Table IV in Appendix D displays the 95th and 99th quantiles for selected degrees of freedom. Besides its use in statistics, the F -distribution is used to model lifetime data; see Exercise 3.6.13.

Example 3.6.2 (Moments of F -Distributions). Let F have an F -distribution with r_1 and r_2 degrees of freedom. Then, as in expression (3.6.7), we can write $F = (r_2/r_1)(U/V)$, where U and V are independent χ^2 random variables with r_1 and r_2 degrees of freedom, respectively. Hence, for the k th moment of F , by independence we have

$$E(F^k) = \left(\frac{r_2}{r_1}\right)^k E(U^k) E(V^{-k}),$$

provided, of course, that both expectations on the right side exist. By Theorem 3.3.2, because $k > -(r_1/2)$ is always true, the first expectation always exists. The second expectation, however, exists if $r_2 > 2k$; i.e., the denominator degrees of freedom must exceed twice k . Assuming this is true, it follows from (3.3.8) that the mean of F is given by

$$E(F) = \frac{r_2}{r_1} r_1 \frac{2^{-1} \Gamma\left(\frac{r_2}{2} - 1\right)}{\Gamma\left(\frac{r_2}{2}\right)} = \frac{r_2}{r_2 - 2}. \quad (3.6.8)$$

If r_2 is large, then $E(F)$ is about 1. In Exercise 3.6.7, a general expression for $E(F^k)$ is derived. ■

3.6.3 Student's Theorem

Our final note in this section concerns an important result for the later chapters on inference for normal random variables. It is a corollary to the t -distribution derived above and is often referred to as Student's Theorem.

Theorem 3.6.1. *Let $X_1, \dots, X_n$ be iid random variables each having a normal distribution with mean μ and variance σ^2 . Define the random variables*

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{and} \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Then

- (a) $\bar{X}$ has a $N\left(\mu, \frac{\sigma^2}{n}\right)$ distribution.
- (b) $\bar{X}$ and S^2 are independent.
- (c) $(n-1)S^2/\sigma^2$ has a $\chi^2(n-1)$ distribution.