CHAPTER 5

Some Special Distributions

5.1 INTRODUCTION

In preceding chapters we studied probability distributions in general. In this chapter we study some commonly occurring probability distributions and investigate their basic properties. The results of this chapter will be of considerable use in theoretical as well as practical applications. We begin with some discrete distributions in Section 5.2 and follow with some continuous models in Section 5.3. Section 5.4 deals with bivariate and multivariate normal distributions, and in Section 5.5 we discuss the exponential family of distributions.

5.2 SOME DISCRETE DISTRIBUTIONS

In this section we study some well-known univariate and multivariate discrete distributions and describe their important properties.

5.2.1 Degenerate Distribution

The simplest distribution is that of an RV X *degenerate* at point k, that is, $P\{X = k\} = 1$ and $= 0$ elsewhere. If we define

(1) $$\varepsilon(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 & \text{if } x \geq 0, \end{cases}$$

the DF of the RV X is $\varepsilon(x - k)$. Clearly, $EX^l = k^l$, $l = 1, 2, \ldots$, and $M(t) = e^{tk}$. In particular, var$(X) = 0$. This property characterizes a degenerate RV. As we shall see, the degenerate RV plays an important role in the study of limit theorems.

5.2.2 Two-Point Distribution

We say that an RV X has a *two-point distribution* if it takes two values, x_1 and x_2, with probabilities

$$P\{X = x_1\} = p \quad \text{and} \quad P\{X = x_2\} = 1 - p, \qquad 0 < p < 1.$$

We may write

$$(2)\qquad X = x_1 I_{[X=x_1]} + x_2 I_{[X=x_2]},$$

where I_A is the indicator function of A. The DF of X is given by

$$(3)\qquad F(x) = p\varepsilon(x - x_1) + (1 - p)\varepsilon(x - x_2).$$

Also,

$$(4)\qquad EX^k = px_1^k + (1 - p)x_2^k, \qquad k = 1, 2, \dots,$$

and

$$(5)\qquad M(t) = pe^{tx_i} + (1 - p)e^{tx_2} \qquad \text{for all } t.$$

In particular,

$$(6)\qquad EX = px_1 + (1 - p)x_2$$

and

$$(7)\qquad \operatorname{var}(X) = p(1 - p)(x_1 - x_2)^2.$$

If $x_1 = 1, x_2 = 0$, we get the important *Bernoulli* RV:

$$(8)\qquad P\{X = 1\} = p \quad \text{and} \quad P\{X = 0\} = 1 - p, \qquad 0 < p < 1.$$

For a Bernoulli RV X with parameter p, we write $X \sim b(1, p)$ and have

$$(9)\qquad EX = p, \qquad \operatorname{var}(X) = p(1-p), \quad \text{and} \quad M(t) = 1 + p(e^t - 1), \qquad \text{all } t.$$

Bernoulli RVs occur in practice, for example, in coin-tossing experiments. Suppose that $P\{\text{H}\} = p, 0 < p < 1$, and $P\{\text{T}\} = 1 - p$. Define RV X so that $X(\text{H}) = 1$ and $X(\text{T}) = 0$. Then $P\{X = 1\} = p$ and $P\{X = 0\} = 1 - p$. Each repetition of the experiment will be called a *trial*. More generally, any nontrivial experiment can be dichotomized to yield a Bernoulli model. Let $(\Omega, \mathcal{S}, P)$ be the sample space of an experiment, and let $A \in \mathcal{S}$ with $P(A) = p > 0$. Then $P(A^c) = 1 - p$. Each performance of the experiment is a Bernoulli trial. It will be convenient to call the occurrence of event A a *success* and the occurrence of A^c a *failure*.

Example 1 (Sabharwal [95]). In a sequence of n Bernoulli trials with constant probability p of success (S), and $1 - p$ of failure (F), let Y_n denote the number of times the combination SF occurs. To find EY_n and $\operatorname{var}(Y_n)$, let X_i represent the event that occurs on the ith trial, and define RVs

$$f(X_i, X_{i+1}) = \begin{cases} 1 & \text{if } X_i = S,\ X_{i+1} = F \\ 0 & \text{otherwise} \end{cases} \qquad (i = 1, 2, \dots, n - 1).$$

Then

$$Y_n = \sum_{i=1}^{n-1} f(X_i, X_{i+1})$$

and

$$EY_n = (n-1)p(1-p).$$

Also,

$$EY_n^2 = E\left[\sum_{i=1}^{n-1} f^2(X_i, X_{i+1})\right] + E\left[\sum\sum_{i \neq j} f(X_i, X_{i+1}) f(X_j, X_{j+1})\right]$$
$$= (n-1)p(1-p) + (n-2)(n-3)p^2(1-p)^2,$$

so that

$$\text{var}(Y_n) = p(1-p)[n - 1 + p(1-p)(5-3n)].$$

If $p = \frac{1}{2}$, then

$$EY_n = \frac{n-1}{4} \quad \text{and} \quad \text{var}(Y_n) = \frac{n+1}{16}.$$

5.2.3 Uniform Distribution on *n* Points

X is said to have a *uniform distribution* on n points $\{x_1, x_2, \ldots, x_n\}$ if its PMF is of the form

(10) $$P\{X = x_i\} = \frac{1}{n}, \qquad i = 1, 2, \ldots, n.$$

Thus we may write

$$X = \sum_{i=1}^{n} x_i I_{[X=x_i]} \quad \text{and} \quad F(x) = \frac{1}{n}\sum_{i=1}^{n} \varepsilon(x - x_i),$$

(11) $$EX = \frac{1}{n}\sum_{i=1}^{n} x_i,$$

(12) $$EX^l = \frac{1}{n}\sum_{i=1}^{n} x_i^l, \qquad l = 1, 2, \ldots$$

and

(13) $$\text{var}(X) = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum x_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

if we write $\bar{x} = \sum_{i=1}^{n} x_i/n$. Also,

$$M(t) = \frac{1}{n}\sum_{i=1}^{n} e^{tx_i} \qquad \text{for all } t. \tag{14}$$

If, in particular, $x_i = i$, $i = 1, 2, \ldots, n$,

$$EX = \frac{n+1}{2}, \qquad EX^2 = \frac{(n+1)(2n+1)}{6}, \tag{15}$$

and

$$\operatorname{var}(X) = \frac{n^2-1}{12}. \tag{16}$$

Example 2. A box contains tickets numbered 1 to N. Let X be the largest number drawn in n random drawings with replacement.

Then $P\{X \le k\} = (k/N)^n$, so that

$$\begin{aligned} P\{X = k\} &= P\{X \le k\} - P\{X \le k-1\} \\ &= \left(\frac{k}{n}\right)^n - \left(\frac{k-1}{N}\right)^n. \end{aligned}$$

Also,

$$\begin{aligned} EX &= N^{-n}\sum_{1}^{N}[k^{n+1} - (k-1)^{n+1} - (k-1)^n] \\ &= N^{-n}\left[N^{n+1} - \sum_{1}^{N}(k-1)^n\right]. \end{aligned}$$

5.2.4 Binomial Distribution

We say that X has a *binomial distribution* with parameter p if its PMF is given by

$$p_k = P\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \ldots, n; \quad 0 \le p \le 1. \tag{17}$$

Since $\sum_{k=0}^{n} p_k = [p+(1-p)]^n = 1$, the p_k's indeed define a PMF. If X has PMF (17), we will write $X \sim b(n, p)$. This is consistent with the notation for a Bernoulli RV. We have

$$F(x) = \sum_{k=0}^{n}\binom{n}{k} p^k (1-p)^{n-k} \varepsilon(x-k).$$

In Example 3.2.5 we showed that

$$EX = np, \tag{18}$$

$$EX^2 = n(n-1)p^2 + np, \tag{19}$$

and

$$\operatorname{var}(X) = np(1-p) = npq, \tag{20}$$

where $q = 1 - p$. Also,

$$\begin{aligned} M(t) &= \sum_{k=0}^{n} e^{tk}\binom{n}{k}p^k(1-p)^{n-k} \\ &= (q + pe^t)^n \qquad \text{for all } t. \end{aligned} \tag{21}$$

The PGF of $X \sim b(n, p)$ is given by $P(s) = \{1 - p(1 - s)\}^n$, $|s| \le 1$.

Binomial distribution can also be considered as the distribution of the sum of n independent, identically distributed $b(1, p)$ random variables. If we toss a coin, with constant probability p of heads and $1 - p$ of tails, n times, the distribution of the number of heads is given by (17). Alternatively, if we write

$$X_k = \begin{cases} 1 & \text{if } k\text{th toss results in a head,} \\ 0 & \text{otherwise,} \end{cases}$$

the number of heads in n trials is the sum $S_n = X_1 + X_2 + \cdots + X_n$. Also

$$P\{X_k = 1\} = p \quad \text{and} \quad P\{X_k = 0\} = 1 - p, \qquad k = 1, 2, \ldots, n.$$

Thus

$$ES_n = \sum_{1}^{n} EX_i = np,$$

$$\operatorname{var}(S_n) = \sum_{1}^{n} \operatorname{var}(X_i) = np(1-p),$$

and

$$\begin{aligned} M(t) &= \prod_{i=1}^{n} Ee^{tX_i} \\ &= (q + pe^t)^n. \end{aligned}$$

Theorem 1. Let $X_i (i = 1, 2, \ldots, k)$ be independent RVs with $X_i \sim b(n_i, p)$. Then $S_k = \sum_{i=1}^{k} X_i$ has a $b(n_1 + n_2 + \cdots + n_k,\ p)$ distribution.

Corollary. If $X_i (i = 1, 2, \ldots, k)$ are iid RVs with common PMF $b(n, p)$, then S_k has a $b(nk, p)$ distribution.

Actually, the additive property described in Theorem 1 characterizes the binomial distribution in the following sense. Let X and Y be two independent, nonnegative, finite integer-valued RVs and let $Z = X+Y$. Then Z is a binomial RV with parameter p if and only if X and Y are binomial RVs with the same parameter p. The "only if" part is due to Shanbhag and Basawa [101] and will not be proved here.

Example 3. A fair die is rolled n times. The probability of obtaining exactly one 6 is $n(\frac{1}{6})(\frac{5}{6})^{n-1}$, the probability of obtaining no 6 is $(\frac{5}{6})^n$, and the probability of obtaining at least one 6 is $1 - (\frac{5}{6})^n$.

The number of trials needed for the probability of at least one 6 to be $\geq \frac{1}{2}$ is given by the smallest integer n such that

$$1 - \left(\frac{5}{6}\right)^n \geq \frac{1}{2}$$

so that

$$n \geq \frac{\log 2}{\log 1.2} \approx 3.8.$$

Example 4. Here r balls are distributed in n cells so that each of n^r possible arrangements has probability n^{-r}. We are interested in the probability p_k that a specified cell has exactly k balls $(k = 0, 1, 2, \ldots, r)$. Then the distribution of each ball may be considered as a trial. A success results if the ball goes to the specified cell (with probability $1/n$); otherwise, the trial results in a failure (with probability $1 - 1/n$). Let X denote the number of successes *in* r trials. Then

$$p_k = P\{X = k\} = \binom{r}{k}\left(\frac{1}{n}\right)^k\left(1 - \frac{1}{n}\right)^{r-k}, \qquad k = 0, 1, 2, \ldots, n.$$

5.2.5 Negative Binomial Distribution (Pascal or Waiting-Time Distribution)

Let $(\Omega, \mathcal{S}, P)$ be a probability space of a given statistical experiment, and let $A \in \mathcal{S}$ with $P(A) = p$. On any performance of the experiment, if A happens we call it a success, otherwise a failure. Consider a succession of trials of this experiment, and let us compute the probability of observing exactly r successes, where $r \geq 1$ is a fixed integer. If X denotes the number of failures that precede the rth success, $X + r$ is the total number of replications needed to produce r successes. This will happen if and only if the last trial results in a success and among the previous $(r + X - 1)$ trials there are exactly X failures. It follows by independence that

$$\text{(22)} \qquad P\{X = x\} = \binom{x + r - 1}{x} p^r (1 - p)^x, \qquad x = 0, 1, 2, \ldots.$$

Rewriting (22) in the form

$$(23) \qquad P\{X = x\} = \binom{-r}{x} p^r (-q)^x, \qquad x = 0, 1, 2, \ldots; \quad q = 1 - p,$$

we see that

$$(24) \qquad \sum_{x=0}^{\infty} \binom{-r}{x} (-q)^x = (1 - q)^{-r} = p^{-r}.$$

It follows that

$$\sum_{x=0}^{\infty} P\{X = x\} = 1.$$

Definition 1. For a fixed positive integer $r \geq 1$ and $0 < p < 1$, an RV with PMF given by (22) is said to have a *negative binomial distribution*. We use the notation $X \sim NB(r; p)$ to denote that X has a negative binomial distribution.

We may write

$$X = \sum_{x=0}^{\infty} x I_{[X=x]} \quad \text{and} \quad F(x) = \sum_{k=0}^{\infty} \binom{k + r - 1}{k} p^r (1 - p)^k \varepsilon(x - k).$$

For the MGF of X we have

$$(25) \qquad \begin{aligned} M(t) &= \sum_{x=0}^{\infty} \binom{x + r - 1}{x} p^r (1 - p)^x e^{tx} \\ &= p^r \sum_{x=0}^{\infty} (qe^t)^x \binom{x + r - 1}{x} \qquad (q = 1 - p) \\ &= p^r (1 - qe^t)^{-r} \qquad \text{for } qe^t < 1. \end{aligned}$$

The PGF is given by $P(s) = p^r (1 - sq)^{-r}$, $|s| \leq 1$. Also,

$$(26) \qquad \begin{aligned} EX &= \sum_{x=0}^{\infty} x \binom{x + r - 1}{x} p^r q^x \\ &= rp^r \sum_{x=0}^{\infty} \binom{x + r}{x} q^{x+1} \\ &= rp^r q (1 - q)^{-r-1} = \frac{rq}{p}. \end{aligned}$$

Similarly, we can show that

$$(27) \qquad \operatorname{var}(X) = \frac{rq}{p^2}.$$

If, however, we are interested in the distribution of the number of trials required to get r successes, we have, writing $Y = X + r$,

$$P\{Y = y\} = \binom{y-1}{r-1} p^r (1-p)^{y-r}, \qquad y = r,\ r+1, \ldots, \tag{28}$$

$$\begin{aligned} EY &= EX + r = \frac{r}{p}, \\ \operatorname{var}(Y) &= \operatorname{var}(X) = \frac{rq}{p^2}, \end{aligned} \tag{29}$$

and

$$M_Y(t) = (pe^t)^r (1 - qe^t)^{-r} \qquad \text{for } qe^t < 1. \tag{30}$$

Let X be a $b(n, p)$ RV, and let Y be the RV defined in (28). If there are r or more successes in the first n trials, at most n trials were required to obtain the first r of these successes. We have

$$P\{X \geq r\} = P\{Y \leq n\} \tag{31}$$

and also

$$P\{X < r\} = P\{Y > n\}. \tag{32}$$

In the special case when $r = 1$, the distribution of X in (22) is given by

$$P\{X = x\} = pq^x, \qquad x = 0, 1, 2, \ldots. \tag{33}$$

An RV X with PMF (33) is said to have a *geometric distribution*. Clearly, for the geometric distribution, we have

$$M(t) = p(1 - qe^t)^{-1}, \quad EX = \frac{q}{p}, \quad \text{and} \quad \operatorname{var}(X) = \frac{q}{p^2}. \tag{34}$$

***Example 5* (Banach's Matchbox Problem).** A mathematician carries one matchbox each in his right and left pockets. When he wants a match, he selects the left pocket with probability p and the right pocket with probability $1 - p$. Suppose that initially each box contains N matches. Consider the moment when the mathematician discovers that a box is empty. At that time the other box may contain $0, 1, 2 \ldots, N$ matches. Let us identify success with the choice of the left pocket. The left-pocket box will be empty at the moment when the right-pocket box contains exactly r matches if and only if exactly $N - r$ failures precede the $(N + 1)$st success. A similar argument applies to the right pocket, and we have

$$\begin{aligned} p_r &= \text{probability that the mathematician discovers a box empty while} \\ &\quad\ \text{the other contains } r \text{ matches} \\ &= \binom{2N-r}{N-r} p^{N+1} q^{N-r} + \binom{2N-r}{N-r} q^{N+1} p^{N-r}. \end{aligned}$$

Example 6. A fair die is rolled repeatedly. Let us compute the probability of event A that a 2 will show up before a 5. Let A_j be the event that a 2 shows up on the jth trial ($j = 1, 2, \ldots$) for the first time, and a 5 does not show up on the previous $j-1$ trials. Then $PA = \sum_{j=1}^{\infty} PA_j$, where $PA_j = \frac{1}{6}(\frac{4}{6})^{j-1}$. It follows that

$$P(A) = \sum_{j=1}^{\infty} \frac{1}{6}\left(\frac{4}{6}\right)^{j-1} = \frac{1}{2}.$$

Similarly, the probability that a 2 will show up before a 5 or a 6 is $\frac{1}{3}$, and so on.

Theorem 2. Let $X_1, X_2, \ldots, X_k$ be independent $NB(r_i; p)$ RVs, $i = 1, 2, \ldots, k$, respectively. Then $S_k = \sum_{i=1}^{k} X_i$ is distributed as $NB(r_1 + r_2 + \cdots + r_k; p)$.

Corollary. If $X_1, X_2, \ldots, X_k$ are iid geometric RVs, then S_k is an $NB(k; p)$ RV.

Theorem 3. Let X and Y be independent RVs with PMFs $NB(r_1; p)$ and $NB(r_2; p)$, respectively. Then the conditional PMF of X, given $X + Y = t$, is expressed by

$$P\{X = x | X + Y = t\} = \frac{\binom{x + r_1 - 1}{x}\binom{t + r_2 - x - 1}{t - x}}{\binom{t + r_1 + r_2 - 1}{t}}.$$

If, in particular, $r_1 = r_2 = 1$, the conditional distribution is uniform on $t + 1$ points.

Proof. By Theorem 2, $X + Y$ is an $NB(r_1 + r_2; p)$ RV. Thus

$$\begin{aligned} P\{X = x | X + Y = t\} &= \frac{P\{X = x,\ Y = t - x\}}{P\{X + Y = t\}} \\ &= \frac{\binom{x + r_1 - 1}{x} p^{r_1}(1-p)^x \binom{t - x + r_2 - 1}{t - x} p^{r_2}(1-p)^{t-x}}{\binom{t + r_1 + r_2 - 1}{t} p^{r_1 + r_2}(1-p)^t} \\ &= \frac{\binom{x + r_1 - 1}{x}\binom{t + r_2 - x - 1}{t - x}}{\binom{t + r_1 + r_2 - 1}{t}}, \qquad t = 0, 1, 2, \ldots. \end{aligned}$$

If $r_1 = r_2 = 1$, that is, if X and Y are independent geometric RVs, then

$$(35)\, P\{X = x | X + Y = t\} = \frac{1}{t + 1}, \qquad x = 0, 1, 2, \ldots, t; \quad t = 0, 1, 2, \ldots.$$

Theorem 4 (Chatterji [12]). Let X and Y be iid RVs, and let

$$P\{X = k\} = p_k > 0, \qquad k = 0, 1, 2, \ldots.$$

If

$$(36) \quad P\{X = t | X + Y = t\} = P\{X = t - 1 | X + Y = t\} = \frac{1}{t+1}, \qquad t \geq 0,$$

then X and Y are geometric RVs.

Proof. We have

$$(37) \qquad P\{X = t | X + Y = t\} = \frac{p_t p_0}{\sum_{k=0}^{t} p_k p_{t-k}} = \frac{1}{t+1}$$

and

$$(38) \qquad P\{X = t - 1 | X + Y = t\} = \frac{p_{t-1} p_1}{\sum_{k=0}^{t} p_k p_{t-k}} = \frac{1}{t+1}.$$

It follows that

$$\frac{p_t}{p_{t-1}} = \frac{p_1}{p_0}$$

and by iteration $p_t = (p_1/p_0)^t p_0$. Since $\sum_{t=0}^{\infty} p_t = 1$, we must have $p_1/p_0 < 1$. Moreover,

$$1 = p_0 \frac{1}{1 - (p_1/p_0)},$$

so that $p_1/p_0 = 1 - p_0$, and the proof is complete.

Theorem 5. If X has a geometric distribution, then for any two nonnegative integers m and n,

$$(39) \qquad P\{X > m + n | X > m\} = P\{X \geq n\}.$$

The proof is left as an exercise.

Remark 1. Theorem 5 says that the geometric distribution has *no memory*; that is, the information of no successes in m trials is forgotten in subsequent calculations.

The converse of Theorem 5 is also true.

Theorem 6. Let X be a nonnegative integer-valued RV satisfying

$$P\{X > m + 1 | X > m\} = P\{X \geq 1\}$$

for any nonnegative integer m. Then X must have a geometric distribution.

Proof. Let the PMF of X be written as

$$P\{X = k\} = p_k, \qquad k = 0, 1, 2, \ldots.$$

Then

$$P\{X \geq n\} = \sum_{k=n}^{\infty} p_k$$

and

$$P\{X > m\} = \sum_{m+1}^{\infty} p_k = q_m, \quad \text{say},$$

$$P\{X > m+1 | X > m\} = \frac{P\{X > m+1\}}{P\{X > m\}} = \frac{q_{m+1}}{q_m}.$$

Thus

$$q_{m+1} = q_m q_0,$$

where $q_0 = P\{X > 0\} = p_1 + p_2 + \cdots = 1 - p_0$. It follows that $q_k = (1 - p_0)^{k+1}$, and hence $p_k = q_{k-1} - q_k = (1 - p_0)^k p_0$, as asserted.

Theorem 7. Let $X_1, X_2, \ldots, X_n$ be independent geometric RVs with parameters $p_1, p_2, \ldots, p_n$, respectively. Then $X_{(1)} = \min(X_1, X_2, \ldots, X_n)$ is also a geometric RV with parameter

$$p = 1 - \prod_{i=1}^{n}(1 - p_i).$$

The proof is left as an exercise.

Corollary. Iid RVs $X_1, X_2, \ldots, X_n$ are $NB(1; p)$ if and only if $X_{(1)}$ is a geometric RV with parameter $1 - (1 - p)^n$.

Proof. The necessity follows from Theorem 7. For the sufficiency part of the proof, let

$$P\{X_{(1)} \leq k\} = 1 - P\{X_{(1)} > k\} = 1 - (1 - p)^{n(k+1)}.$$

But

$$\begin{aligned} P\{X_{(1)} \leq k\} &= 1 - P\{X_1 > k, X_2 > k, \ldots, X_n > k\} \\ &= 1 - [1 - F(k)]^n, \end{aligned}$$

where F is the common DF of $X_1, X_2, \ldots, X_n$. It follows that

$$1 - F(k) = (1-p)^{k+1},$$

so that $P\{X_1 > k\} = (1-p)^{k+1}$, which completes the proof.

5.2.6 Hypergeometric Distribution

A box contains N marbles. Of these, M are drawn at random, marked, and returned to the box. The contents of the box are then thoroughly mixed. Next, n marbles are drawn at random from the box, and the marked marbles are counted. If X denotes the number of marked marbles, then

$$P\{X = x\} = \binom{N}{n}^{-1} \binom{M}{x} \binom{N-M}{n-x}. \tag{40}$$

Since x cannot exceed M or n, we must have

$$x \le \min(M, n). \tag{41}$$

Also, $x \ge 0$ and $N - M \ge n - x$, so that

$$x \ge \max(0, M + n - N). \tag{42}$$

Note that

$$\sum_{k=0}^{n} \binom{a}{k} \binom{b}{n-k} = \binom{a+b}{n}$$

for arbitrary numbers a, b and positive integer n. It follows that

$$\sum_x P\{X = x\} = \binom{N}{n}^{-1} \sum_x \binom{M}{x} \binom{N-M}{n-x} = 1.$$

Definition 2. An RV X with PMF given by (40) is called a *hypergeometric RV*.

It is easy to check that

$$EX = \frac{n}{N} M, \tag{43}$$

$$EX^2 = \frac{M(M-1)}{N(N-1)} n(n-1) + \frac{nM}{N}, \tag{44}$$

and

$$\operatorname{var}(X) = \frac{nM}{N^2(N-1)} (N-M)(N-n). \tag{45}$$

Example 7. A lot consisting of 50 bulbs is inspected by taking at random 10 bulbs and testing them. If the number of defective bulbs is at most 1, the lot is accepted; otherwise, it is rejected. If there are, in fact, 10 defective bulbs in the lot, the probability of accepting the lot is

$$\frac{\binom{10}{1}\binom{40}{9}}{\binom{50}{10}} + \frac{\binom{40}{10}}{\binom{50}{10}} = .3487$$

Example 8. Suppose that an urn contains b white and c black balls, $b + c = N$. A ball is drawn at random, and before drawing the next ball, $s + 1$ balls of the same color are added to the urn. The procedure is repeated n times. Let X be the number of white balls drawn in n draws, $X = 0, 1, 2, \dots, n$. We shall find the PMF of X.

First note that the probability of drawing k white balls in successive draws is

$$\frac{b}{N}\frac{b+s}{N+s}\frac{b+2s}{N+2s}\cdots\frac{b+(k-1)s}{N+(k-1)s},$$

and the probability of drawing k white balls in the first k draws and then $n - k$ black balls in the next $n - k$ draws is

$$\begin{aligned} p_k = \frac{b}{N}\frac{b+s}{N+s}\cdots\frac{b+(k-1)s}{N+(k-1)s}\frac{c}{N+ks}\frac{c+s}{N+(k+1)s} \\ \cdots\frac{c+(n-k-1)s}{N+(n-1)s}. \end{aligned} \tag{46}$$

Here p_k also gives the probability of drawing k white and $n - k$ black balls in any given order. It follows that

$$P\{X = k\} = \binom{n}{k} p_k. \tag{47}$$

An RV X with PMF given by (47) is said to have a *Polya distribution.* Let us write

$$Np = b, \quad N(1-p) = c, \quad \text{and} \quad N\alpha = s.$$

Then with $q = 1 - p$, we have

$$P\{X = k\} = \binom{n}{k}\frac{p(p+\alpha)\cdots[p+(k-1)\alpha]q(q+\alpha)\cdots[q+(n-k-1)\alpha]}{1(1+\alpha)\cdots[1+(n-1)\alpha]}.$$

Let us take $s = -1$. This means that the ball drawn at each draw is not replaced in the urn before drawing the next ball. In this case $\alpha = -1/N$, and we have

$$P\{X = k\} = \binom{n}{k} \frac{Np(Np-1)\cdots[Np-(k-1)]c(c-1)\cdots[c-(n-k-1)]}{N(N-1)\cdots[N-(n-1)]}$$

$$= \frac{\binom{Np}{k}\binom{Nq}{n-k}}{\binom{N}{n}}, \tag{48}$$

which is a hypergeometric distribution. Here

$$\max(0, n - Nq) \le k \le \min(n, Np). \tag{49}$$

Theorem 8. Let X and Y be independent RVs with PMFs $b(m, p)$ and $b(n, p)$, respectively. Then the conditional distribution of X, given $X + Y$, is hypergeometric.

5.2.7 Negative Hypergeometric Distribution

Consider the model of Section 5.2.6. A box contains N marbles; M of these are marked (or say defective) and $N - M$ are unmarked. A sample of size n is taken, and let X denote the number of defective marbles in the sample. If the sample is drawn without replacement, we saw that X has a hypergeometric distribution with PMF (40). If, on the other hand, the sample is drawn with replacement, then $X \sim b(n, p)$ where $p = M/N$.

Let Y denote the number of draws needed to draw the rth defective marble. If the draws are made with replacement, then Y has the negative binomial distribution given in (22) with $p = M/N$. What if the draws are made without replacement? In that case in order that the kth draw ($k \ge r$) be the rth defective marble drawn, the kth draw must produce a defective marble, whereas the previous $k - 1$ draws must produce $r - 1$ defectives. It follows that

$$P(Y = k) = \frac{\binom{M}{r-1}\binom{N-M}{k-r}}{\binom{N}{k-1}} \cdot \frac{M-r+1}{N-k+1}$$

for $k = r, r + 1, \ldots, N$. Rewriting, we see that

$$P(Y = k) = \binom{k-1}{r-1} \frac{\binom{N-k}{m-r}}{\binom{N}{M}}. \tag{50}$$

An RV Y with PMF (50) is said to have a *negative hypergeometric distribution.*

It is easy to see that

$$EY = r\frac{N+1}{M+1}, \quad EY(Y+1) = \frac{r(r+1)(N+1)(N+2)}{(M+1)(M+2)},$$

and

$$\text{var}(Y) = \frac{r(N-M)(N+1)(M+1-r)}{(M+1)^2(M+2)}.$$

Also, if $r/N \to 0$ and $k/N \to 0$ as $N \to \infty$, then

$$\binom{k-1}{r-1}\binom{N-k}{M-r}\Big/\binom{N}{M} \longrightarrow \binom{k-1}{r-1}\left(\frac{M}{N}\right)^r\left(1-\frac{M}{N}\right)^{k-r}$$

which is (22).

5.2.8 Poisson Distribution

Definition 3. An RV X is said to be a *Poisson RV* with parameter $\lambda > 0$ if its PMF is given by

$$(51) \qquad P\{X = k\} = \frac{e^{-\lambda}\lambda^k}{k!}, \qquad k = 0, 1, 2, \ldots.$$

We first check to see that (51) indeed defines a PMF. We have

$$\sum_{k=0}^{\infty} P\{X = k\} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = 1.$$

If X has the PMF given by (51), we will write $X \sim P(\lambda)$. Clearly,

$$X = \sum_{k=0}^{\infty} k I_{[X=k]}$$

and

$$F(x) = \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!}\varepsilon(x-k).$$

The mean and the variance are given by (see Problem 3.2.9)

$$(52) \qquad EX = \lambda, \qquad EX^2 = \lambda + \lambda^2,$$

and

$$(53) \qquad \text{var}(X) = \lambda.$$

The MGF of X is given by (see Example 3.3.7)

$$Ee^{tX} = \exp[\lambda(e^t - 1)] \tag{54}$$

and the PGF by $P(s) = e^{-\lambda(1-s)}$, $|s| \le 1$.

Theorem 9. Let $X_1, X_2, \ldots, X_n$ be independent Poisson RVs with $X_k \sim P(\lambda_k)$, $k = 1, 2, \ldots, n$. Then $S_n = X_1 + X_2 + \cdots + X_n$ is a $P(\lambda_1 + \lambda_2 + \cdots + \lambda_n)$ RV.

The converse of Theorem 9 is also true. Indeed, Raikov [82] showed that if $X_1, X_2, \ldots, X_n$ are independent and $S_n = \sum_{i=1}^n X_i$ has a Poisson distribution, each of the RVs $X_1, X_2, \ldots, X_n$ has a Poisson distribution.

Example 9. The number of female insects in a given region follows a Poisson distribution with mean λ. The number of eggs laid by each insect is a $P(\mu)$ RV. We are interested in the probability distribution of the number of eggs in the region.

Let F be the number of female insects in the given region. Then

$$P\{F = f\} = \frac{e^{-\lambda}\lambda^f}{f!}, \qquad f = 0, 1, 2, \ldots.$$

Let Y be the number of eggs laid by each insect. Then

$$\begin{aligned} P\{Y = y,\ F = f\} &= P\{F = f\}P\{Y = y|F = f\} \\ &= \frac{e^{-\lambda}\lambda^f}{f!}\frac{(f\mu)^y e^{-\mu f}}{y!}. \end{aligned}$$

Thus

$$P\{Y = y\} = \frac{e^{-\lambda}\mu^y}{y!}\sum_{f=0}^{\infty}\frac{(\lambda e^{-\mu})^f f^y}{f!}.$$

The MGF of Y is given by

$$\begin{aligned} M(t) &= \sum_{f=0}^{\infty}\frac{\lambda^f e^{-\lambda}}{f!}\sum_{y=0}^{\infty}\frac{e^{yt}(f\mu)^y}{y!}e^{-\mu f} \\ &= \sum_{f=0}^{\infty}\frac{\lambda^f e^{-\lambda}}{f!}\exp[f\mu(e^t - 1)] \\ &= e^{-\lambda}\sum_{f=0}^{\infty}\frac{[\lambda e^{\mu(e^t-1)}]^f}{f!} \\ &= e^{-\lambda}\exp[\lambda e^{\mu(e^t-1)}]. \end{aligned}$$

Theorem 10. Let X and Y be independent RVs with PMFs $P(\lambda_1)$ and $P(\lambda_2)$, respectively. Then the conditional distribution of X, given $X+Y$, is binomial.

Proof. For nonnegative integers m and n, $m < n$, we have

$$\begin{aligned} P\{X = m | X + Y = n\} &= \frac{P\{X = m,\ Y = n - m\}}{P\{X + Y = n\}} \\ &= \frac{e^{-\lambda_1}(\lambda_1^m/m!)e^{-\lambda_2}(\lambda_2^{n-m}/(n-m)!)}{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n/n!} \\ &= \binom{n}{m}\frac{\lambda_1^m\lambda_2^{n-m}}{(\lambda_1+\lambda_2)^n} \\ &= \binom{n}{m}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^m\left(1-\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{n-m}, \\ & \quad m = 0, 1, 2, \ldots, n, \end{aligned}$$

and the proof is complete.

Remark 2. The converse of this result is also true in the following sense. If X and Y are independent nonnegative integer-valued RVs such that $P\{X = k\} > 0$, $P\{Y = k\} > 0$, for $k = 0, 1, 2, \ldots$, and the conditional distribution of X, given $X + Y$, is binomial, both X and Y are Poisson. This result is due to Chatterji [12]. For the proof, see Problem 13.

Theorem 11. If $X \sim P(\lambda)$ and the conditional distribution of Y, given $X = x$, is $b(x, p)$, then Y is a $P(\lambda p)$ RV.

Example 10 (Lamperti and Kruskal [58]). Let N be a nonnegative integer-valued RV. Independent of each other, N balls are placed either in urn A with probability p $(0 < p < 1)$ or in urn B with probability $1 - p$, resulting in N_A balls in urn A and $N_B = N - N_A$ balls in urn B. We will show that the RVs N_A and N_B are independent if and only if N has a Poisson distribution. We have

$$P\{N_A = a \text{ and } N_B = b | N = a + b\} = \binom{a+b}{a}p^a(1-p)^b,$$

where a, b are integers ≥ 0. Thus

$$P\{N_A = a,\ N_B = b\} = \binom{a+b}{a}p^aq^bP\{N = n\}, \qquad q = 1 - p, \quad n = a + b.$$

If N has a Poisson (λ) distribution, then

$$P\{N_A = a,\ N_B = b\} = \frac{(a+b)!}{a!\,b!}p^aq^b\frac{e^{-\lambda}\lambda^{a+b}}{(a+b)!}$$

$$= \left(\frac{p^a \lambda^a e^{-\lambda/2}}{a!}\right)\left(\frac{q^b \lambda^b}{b!} e^{-\lambda/2}\right),$$

so that N_A and N_B are independent.

Conversely, if N_A and N_B are independent, then

$$P\{N = n\} n! = f(a) g(b)$$

for some functions f and g. Clearly, $f(0) \neq 0, g(0) \neq 0$ because $P\{N_A = 0,\ N_B = 0\} > 0$. Thus there is a function h such that $h(a+b) = f(a)g(b)$ for all nonnegative integers a, b. It follows that

$$h(1) = f(1)g(0) = f(0)g(1),$$
$$h(2) = f(2)g(0) = f(1)g(1) = f(0)g(2),$$

and so on. By induction,

$$f(a) = f(1)\left[\frac{g(1)}{g(0)}\right]^{a-1}, \qquad g(b) = g(1)\left[\frac{f(1)}{f(0)}\right]^{b-1}.$$

We may write, for some $\alpha_1, \alpha_2, \lambda$,

$$f(a) = \alpha_1 e^{-a\lambda}, \qquad g(b) = \alpha_2 e^{-b\lambda},$$

and

$$P\{N = n\} = \alpha_1 \alpha_2 \frac{e^{-\lambda(a+b)}}{(a+b)!},$$

so that N is a Poisson RV.

5.2.9 Multinomial Distribution

The binomial distribution is generalized in the following natural fashion. Suppose that an experiment is repeated n times. Each replication of the experiment terminates in one of k mutually exclusive and exhaustive events $A_1, A_2, \ldots, A_k$. Let p_j be the probability that the experiment terminates in A_j, $j = 1, 2, \ldots, k$, and suppose that p_j ($j = 1, 2, \ldots, k$) remains constant for all n replications. We assume that the n replications are independent.

Let $x_1, x_2, \ldots, x_{k-1}$ be nonnegative integers such that $x_1 + x_2 + \cdots + x_{k-1} \leq n$. Then the probability that exactly x_i trials terminate in A_i, $i = 1, 2, \ldots, k-1$, and hence that $x_k = n - (x_1 + x_2 + \cdots + x_{k-1})$ trials terminate in A_k is clearly

$$\frac{n!}{x_1! x_2! \cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}.$$

If $(X_1, X_2, \dots, X_k)$ is a random vector such that $X_j = x_j$ means that event A_j has occurred x_j times, $x_j = 0, 1, 2, \dots, n$, the joint PMF of $(X_1, X_2, \dots, X_k)$ is given by

$$(55)\qquad P\{X_1 = x_1,\ X_2 = x_2, \dots,\ X_k = x_k\} = \begin{cases} \dfrac{n!}{x_1!\,x_2!\cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k} & \text{if } n = \sum_1^k x_i, \\ 0 & \text{otherwise.} \end{cases}$$

Definition 4. An RV $(X_1, X_2, \dots, X_{k-1})$ with joint PMF given by

$$(56)\qquad P\{X_1 = x_1,\ X_2 = x_2, \dots,\ X_{k-1} = x_{k-1}\} = \begin{cases} \dfrac{n!}{x_1!\,x_2!\dots(n - x_1 - \cdots - x_{k-1})!} p_1^{x_1} p_2^{x_2} \cdots p_k^{n-x_1-\dots-x_{k-1}} \\ \qquad \text{if } x_1 + x_2 + \cdots + x_{k-1} \le n, \\ 0 \qquad \text{otherwise,} \end{cases}$$

is said to have a *multinomial distribution*.

For the MGF of $(X_1, X_2, \dots, X_{k-1})$ we have

$$(57)\qquad \begin{aligned} M(t_1, t_2, \dots, t_{k-1}) &= Ee^{t_1X_1+t_2X_2+\cdots+t_{k-1}X_{k-1}} \\ &= \sum_{\substack{x_1, x_2, \dots, x_{k-1}=0 \\ x_1+x_2+\dots x_{k-1}\le n}}^{n} e^{t_1x_1+\cdots+t_{k-1}x_{k-1}} \frac{n!\, p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}}{x_1!\,x_2!\cdots x_k!} \\ &= \sum_{\substack{x_1, x_2, \dots, x_{k-1}=0 \\ x_1+x_2+\dots x_{k-1}\le n}}^{n} \frac{n!}{x_1!\,x_2!\dots x_k!} (p_1e^{t_1})^{x_1} (p_2e^{t_2})^{x_2} \dots \\ &\quad \cdot (p_{k-1}e^{t_{k-1}})^{x_{k-1}} p_k^{x_k} \\ &= (p_1e^{t_1} + p_2e^{t_2} + \cdots + p_{k-1}e^{t_{k-1}} + p_k)^n \\ &\qquad \text{for all } t_1, t_2, \dots, t_{k-1} \in \mathcal{R}. \end{aligned}$$

Clearly,

$$M(t_1, 0, 0, \dots, 0) = (p_1e^{t_1} + p_2 + \cdots + p_k)^n = (1 - p_1 + p_1e^{t_1})^n,$$

which is binomial. Indeed, the marginal PMF of each X_i, $i = 1, 2, \dots, k-1$, is binomial. Similarly, the joint MGF of X_i, X_j, $i, j = 1, 2, \dots, k-1$ $(i \ne j)$, is

$$M(0, 0, \dots, 0, t_i, 0, \dots, 0, t_j, 0, \dots, 0) = [p_ie^{t_i} + p_je^{t_j} + (1 - p_i - p_j)]^n,$$

which is the MGF of a *trinomial distribution* with PMF

(58) $$f(x_i, x_j) = \frac{n!}{x_i!\,x_j!(n - x_i - x_j)!} p_i^{x_i} p_j^{x_j} p_k^{n-x_i-x_j}, \qquad p_k = 1 - p_i - p_j.$$

Note that the RVs $X_1, X_2, \ldots, X_{k-1}$ are dependent.

From the MGF of $(X_1, X_2, \ldots, X_{k-1})$ or directly from the marginal PMFs we can compute the moments. Thus

(59) $$EX_j = np_j \quad \text{and} \quad \text{var}(X_j) = np_j(1 - p_j), \qquad j = 1, 2, \ldots, k - 1,$$

and for $j = 1, 2, \ldots, k - 1$, and $i \neq j$,

(60) $$\text{cov}(X_i, X_j) = E[(X_i - np_i)(X_j - np_j)] = -np_i p_j.$$

It follows that the correlation coefficient between X_i and X_j is given by

(61) $$\rho_{ij} = -\left[\frac{p_i p_j}{(1 - p_i)(1 - p_j)}\right]^{1/2}, \qquad i, j = 1, 2, \ldots, k - 1 \quad (i \neq j).$$

Example 11. Consider the trinomial distribution with PMF

$$P\{X = x, Y = y\} = \frac{n!}{x!\,y!\,(n - x - y)!} p_1^x p_2^y p_3^{n-x-y},$$

where x, y are nonnegative integers such that $x + y \leq n$, and $p_1, p_2, p_3 > 0$ with $p_1 + p_2 + p_3 = 1$. The marginal PMF of X is given by

$$P\{X = x\} = \binom{n}{x} p_1^x (1 - p_1)^{n-x}, \qquad x = 0, 1, 2, \ldots, n.$$

It follows that

$$P\{Y = y | X = x\}$$
$$= \begin{cases} \dfrac{(n - x)!}{y!\,(n - x - y)!} \dfrac{p_2}{1 - p_1} \left(\dfrac{p_3}{1 - p_1}\right)^{n-x-y} & \text{if } y = 0, 1, 2, \ldots, n - x, \\ 0 & \text{otherwise,} \end{cases}$$

(62)

which is $b(n - x, p_2/(1 - p_1))$. Thus

(63) $$E\{Y|x\} = (n - x)\frac{p_2}{1 - p_1}.$$

Similarly,

(64) $$E\{X|y\} = (n - y)\frac{p_1}{1 - p_2}.$$

Finally, we note that if $\mathbf{X} = (X_1, X_2, \ldots, X_k)$ and $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_k)$ are two independent multinomial RVs with common parameter $(p_1, p_2, \ldots, p_k)$, then $\mathbf{Z} = \mathbf{X} + \mathbf{Y}$ is also a multinomial RV with probabilities $(p_1, p_2, \ldots, p_k)$. This follows easily if one employs the MGF technique, using (58). Actually, this property characterizes the multinomial distribution. If $\mathbf{X}$ and $\mathbf{Y}$ are k-dimensional, nonnegative, independent random vectors, and if $\mathbf{Z} = \mathbf{X} + \mathbf{Y}$ is a multinomial random vector with parameter $(p_1, p_2, \ldots, p_k)$, then $\mathbf{X}$ and $\mathbf{Y}$ also have multinomial distribution with the same parameter. This result is due to Shanbhag and Basawa [101] and will not be proved here.

5.2.10 Multivariate Hypergeometric Distribution

Consider an urn containing N items divided into k categories containing $n_1, n_2, \ldots, n_k$ items, respectively, where $\sum_{j=1}^{k} n_j = N$. A random sample, without replacement, of size n is taken from the urn. Let X_i = number of items in sample of type i. Then

$$P\{X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k\} = \prod_{j=1}^{k} \binom{n_j}{x_j} \Big/ \binom{N}{n}, \tag{65}$$

where $x_j = 0, 1, \ldots, \min(n, n_j)$, and $\sum_{j=1}^{k} x_j = n$.

We say that $(X_1, X_2, \ldots, X_{k-1})$ has *multivariate hypergeometric distribution* if its joint PMF is given by (65). It is clear that each X_j has a marginal hypergeometric distribution. Moreover, the conditional distributions are also hypergeometric. Thus

$$P\{X_i = x_i | X_j = x_j\} = \frac{\binom{n_i}{x_i}\binom{N - n_i - x_j}{n - x_i - x_j}}{\binom{N - n_j}{n - x_j}},$$

and

$$P\{X_i = x_i | X_j = x_j, X_\ell = x_\ell\} = \frac{\binom{n_i}{x_i}\binom{N - n_i - n_j - n_\ell}{n - x_i - x_j - x_\ell}}{\binom{N - n_j - n_\ell}{n - x_j - x_\ell}},$$

and so on. It is therefore easy to write down the marginal and conditional means and variances. We leave the reader to show that

$$EX_j = n\frac{n_j}{N}$$

$$\operatorname{var}(X_j) = n\frac{n_j}{n}\frac{N - n_j}{N}\frac{N - n}{N - 1},$$

and

$$\operatorname{cov}(X_i, X_j) = -\frac{N-n}{N-1} n \left(\frac{n_j}{N}\right)^2.$$

5.2.11 Multivariate Negative Binomial Distribution

Consider the setup of Section 5.2.9, where each replication of an experiment terminates in one of k mutually exclusive and exhaustive events $A_1, A_2, \dots, A_k$. Let $p_j = P(A_j)$, $j = 1, 2, \dots, k$. Suppose that the experiment is repeated until event A_k is observed for the rth time, $r \geq 1$. Then

$$(66) \qquad P(X_1 = x_1,\ X_2 = x_2, \dots, X_k = r) = \frac{(x_1 + x_2 + \dots + x_{k-1} + r - 1)!}{\left(\prod_{j=1}^{k-1} x_j!\right)(r-1)!} p_k^r \prod_{j=1}^{k-1} p_j^{x_j}$$

for $x_i = 0, 1, 2, \dots$ $(i = 1, 2, \dots, k-1)$, $1 \leq r < \infty$, $0 < p_i < 1$, $\sum_{i=1}^{k-1} p_i < 1$, and $p_k = 1 - \sum_{j=1}^{k-1} p_j$.

We say that $(X_1, X_2, \dots, X_{k-1})$ has a *multivariate negative binomial* (or *negative multinomial) distribution* if its joint PMF is given by (66).

It is easy to see that the marginal PMF of any subset of $\{X_1, X_2, \dots, X_{k-1}\}$ is negative multinomial. In particular, each X_j has a negative binomial distribution.

We will leave the reader to show that

$$(67) \qquad M(s_1, s_2, \dots, s_{k-1}) = Ee^{\sum_{j=1}^{k-1} s_j X_j} = p_k^r \left(1 - \sum_{j=1}^{k-1} s_j p_j\right)^{-r}$$

and

$$(68) \qquad \operatorname{cov}(X_i, X_j) = \frac{r p_i p_j}{p_k^2}.$$

PROBLEMS 5.2

1. (a) Let us write

$$b(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, 2, \dots, n.$$

Show that as k goes from 0 to n, $b(k; n, p)$ first increases monotonically and then decreases monotonically. The greatest value is assumed when $k = m$, where m is an integer such that

$$(n+1)p - 1 < m \le (n+1)p$$

except that $b(m-1; n, p) = b(m; n, p)$ when $m = (n+1)p$.

(b) If $k \ge np$, then

$$P\{X \ge k\} \le b(k; n, p)\frac{(k+1)(1-p)}{k+1-(n+1)p};$$

and if $k \le np$, then

$$P\{X \le k\} \le b(k; n, p)\frac{(n-k+1)p}{(n+1)p-k}.$$

2. Generalize the result in Theorem 10 to n independent Poisson RVs; that is, if $X_1, X_2, \dots, X_n$ are independent RVs with $X_i \sim P(\lambda_i)$, $i = 1, 2, \dots, n$, the conditional distribution of $X_1, X_2, \dots, X_n$, given $\sum_{i=1}^n X_i = t$, is multinomial with parameters $t, \lambda_1/\sum_i^n \lambda_i, \dots, \lambda_n/\sum_1^n \lambda_i$.

3. Let X_1, X_2 be independent RVs with $X_i \sim b(n_i, \frac{1}{2})$, $i = 1, 2$. What is the PMF of $X_1 - X_2 + n_2$?

4. A box contains N identical balls numbered 1 through N. Of these balls, n are drawn at a time. Let $X_1, X_2, \dots, X_n$ denote the numbers on the n balls drawn. Let $S_n = \sum_{i=1}^n X_i$. Find $\text{var}(S_n)$.

5. From a box containing N identical balls marked 1 through N, M balls are drawn one after another without replacement. Let X_i denote the number on the ith ball drawn, $i = 1, 2, \dots, M$, $1 \le M \le N$. Let $Y = \max(X_1, X_2, \dots, X_M)$. Find the DF and the PMF of Y. Also find the conditional distribution of $X_1, X_2, \dots, X_M$, given $Y = y$. Find EY and $\text{var}(Y)$.

6. Let $f(x; r, p)$, $x = 0, 1, 2, \dots$, denote the PMF of an $NB(r; p)$ RV. Show that the terms $f(x; r, p)$ first increase monotonically and then decrease monotonically. When is the greatest value assumed?

7. Show that the terms

$$P_\lambda\{X = k\} = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k = 0, 1, 2, \dots,$$

of the Poisson PMF reach their maxima when k is the largest integer $\le \lambda$ and at $(\lambda - 1)$ and λ if λ is an integer.

8. Show that

$$\binom{n}{k}p^k(1-p)^{n-k} \to e^{-\lambda}\frac{\lambda^k}{k!}$$

as $n \to \infty$ and $p \to 0$, so that $np = \lambda$ remains constant. (*Hint:* Use Stirling's approximation, namely, $n! \approx \sqrt{2\pi}\, n^{n+1/2}e^{-n}$ as $n \to \infty$.)

9. A biased coin is tossed indefinitely. Let p $(0 < p < 1)$ be the probability of success (heads). Let Y_1 denote the length of the first run and Y_2 be the length of the second run. Find the PMFs of Y_1 and Y_2, and show that $EY_1 = q/p + p/q$, $EY_2 = 2$. If Y_n denotes the length of the nth run, $n \geq 1$, what is the PMF of Y_n? Find EY_n.

10. Show that

$$\binom{N}{n}^{-1}\binom{Np}{k}\binom{N(1-p)}{n-k} \to \binom{n}{k}p^k(1-p)^{n-k}$$

as $N \to \infty$.

11. Show that

$$\binom{r+k-1}{k}p^r(1-p)^k \to e^{-\lambda}\frac{\lambda^k}{k!}$$

as $p \to 1$ and $r \to \infty$ in such a way that $r(1-p) = \lambda$ remains fixed.

12. Let X and Y be independent geometric RVs. Show that $\min(X, Y)$ and $X - Y$ are independent.

13. Let X and Y be independent RVs with PMFs $P\{X = k\} = p_k$, $P\{Y = k\} = q_k$, $k = 0, 1, 2, \ldots$, where $p_k, q_k > 0$ and $\sum_{k=0}^{\infty} p_k = \sum_{k=0}^{\infty} q_k = 1$. Let

$$P\{X = k | X + Y = t\} = \binom{t}{k}\alpha_t^k(1-\alpha_t)^{t-k}, \qquad 0 \leq k \leq t.$$

Then $\alpha_t = \alpha$ for all t, and

$$p_k = \frac{e^{-\theta\beta}(\theta\beta)^k}{k!} \quad \text{and} \quad q_k = \frac{e^{-\theta}\theta^k}{k!},$$

where $\beta = \alpha/(1-\alpha)$, and $\theta > 0$ is arbitrary. (Chatterji [12])

14. Generalize the result of Example 10 to the case of k urns, $k \geq 3$.

15. Let $(X_1, X_2, \ldots, X_{k-1})$ have a multinomial distribution with parameters n, $p_1, p_2, \ldots, p_{k-1}$. Write

$$Y = \sum_{i=1}^{k} \frac{(X_i - np_i)^2}{np_i},$$

where $p_k = 1 - p_1 - \cdots - p_{k-1}$, and $X_k = n - X_1 - \cdots - X_{k-1}$. Find EY and $\operatorname{var}(Y)$.

16. Let X_1, X_2 be iid RVs with common DF F, having positive mass at $0, 1, 2, \ldots$. Also, let $U = \max(X_1, X_2)$ and $V = X_1 - X_2$. Then

$$P\{U = j,\ V = 0\} = P\{U = j\}P\{V = 0\}$$

for all j if and only if F is a geometric distribution. (Srivastava [107])

17. Let X and Y be mutually independent RVs, taking nonnegative integer values. Then

$$P\{X \le n\} - P\{X + Y \le n\} = \alpha P\{X + Y = n\}$$

holds for $n = 0, 1, 2, \ldots$ and some $\alpha > 0$ if and only if

$$P\{Y = n\} = \frac{1}{1+\alpha}\left(\frac{\alpha}{1+\alpha}\right)^n, \qquad n = 0, 1, 2, \ldots.$$

(*Hint:* Use Problem 3.3.8.) (Puri [81])

18. Let $X_1, X_2, \ldots$ be a sequence of independent $b(1, p)$ RVs with $0 < p < 1$. Also, let $Z_N = \sum_{i=1}^{N} X_i$, where N is a $P(\lambda)$ RV that is independent of the X_i's. Show that Z_N and $N - Z_N$ are independent.

19. Prove Theorems 5, 7, 8, and 11.

5.3 SOME CONTINUOUS DISTRIBUTIONS

In this section we study some most frequently used absolutely continuous distributions and describe their important properties. Before we introduce specific distributions it should be remarked that associated with each PDF f there is an *index* or a *parameter* θ (may be multidimensional) which takes values in an index set Θ. For any particular choice of $\theta \in \Theta$ we obtain a specific PDF f_θ from the family of PDFs $\{f_\theta,\ \theta \in \Theta\}$.

Let X be an RV with PDF $f_\theta(x)$, where θ is a real-valued parameter. We say that θ is a *location parameter* and $\{f_\theta\}$ is a *location family* if $X - \theta$ has PDF $f(x)$ which does not depend on θ. The parameter θ is said to be a *scale parameter* and $\{f_\theta\}$ is a *scale family* of PDFs if X/θ has PDF $f(x)$ which is free of θ. If $\theta = (\mu, \sigma)$ is two-dimensional, we say that θ is a *location-scale parameter* if the PDF of $(X - \mu)/\sigma$ is free of μ and σ. In that case, $\{f_\theta\}$ is known as a *location-scale family*.

It is easily seen that θ is a location parameter if and only if $f_\theta(x) = f(x - \theta)$, a scale parameter if and only $f_\theta(x) = (1/\theta)f(x)$, and a location-scale parameter if $f_\theta(x) = (1/\sigma)f((x - \mu)/\sigma)$, $\sigma > 0$ for some PDF f. The density f is called the *standard* PDF for the family $\{f_\theta, \theta \in \Theta\}$.

A location parameter simply relocates or shifts the graph of PDF f without changing its shape. A scale parameter stretches (if $\theta > 1$) or contracts (if $\theta < 1$) the graph of f. A location-scale parameter, on the other hand, stretches or contracts the graph of f with the scale parameter and then shifts the graph to locate at μ (see Fig. 1).

Some PDFs also have a shape parameter. Changing its value alters the shape of the graph. For the Poisson distribution λ is a shape parameter.

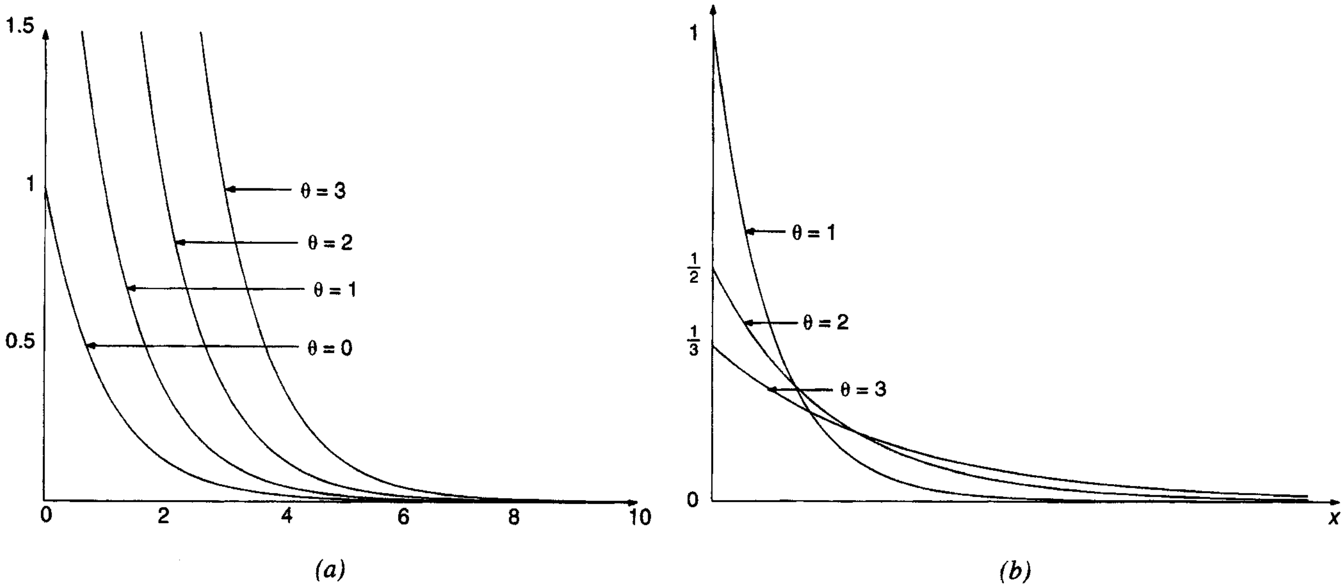


Fig. 1. (*a*) Exponential location family; (*b*) exponential scale family.

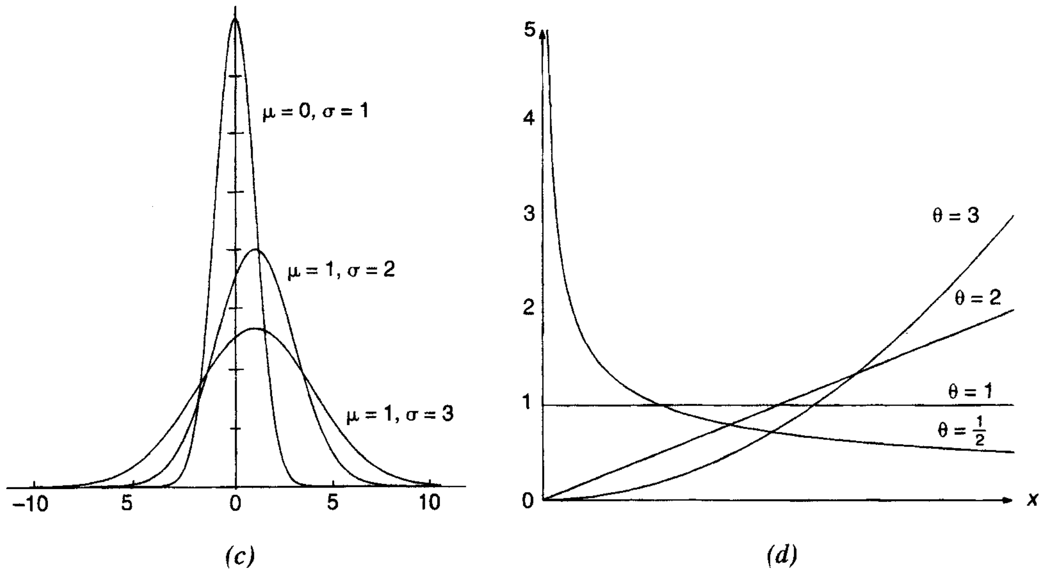


Fig. 1. (*continued*). (*c*) normal location-scale family; (*d*) shape parameter family $f_\theta(x) = \theta x^{\theta-1}$.

For the following PDF,

$$f(x;\mu,\beta,\alpha)=\frac{1}{\beta\Gamma(\alpha)}\left(\frac{x-\mu}{\beta}\right)^{\alpha-1}\exp\left\{-\frac{x-\mu}{\beta}\right\},\qquad x>\mu$$

and $=0$ otherwise, μ is a location, β a scale, and α a shape parameter. The standard density for this location-scale family is

$$f(x)=\frac{1}{\Gamma(\alpha)}x^{\alpha-1}e^{-x},\qquad x>0$$

and $=0$ otherwise. For the standard PDF f, α is a shape parameter.

5.3.1 Uniform Distribution (Rectangular Distribution)

Definition 1. An RV X is said to have a *uniform distribution* on the interval $[a,b]$, $-\infty<a<b<\infty$, if its PDF is given by

(1) $$f(x)=\begin{cases}\dfrac{1}{b-a}, & a\le x\le b,\\ 0, & \text{otherwise.}\end{cases}$$

We will write $X\sim U[a,b]$ if X has a uniform distribution on $[a,b]$.

The endpoint a or b or both may be excluded. Clearly,

$$\int_{-\infty}^{\infty}f(x)\,dx=1,$$

so that (1) indeed defines a PDF. The DF of X is given by

(2) $$F(x)=\begin{cases}0, & x<a,\\ \dfrac{x-a}{b-a}, & a\le x<b,\\ 1, & b\le x;\end{cases}$$

(3) $$EX=\frac{a+b}{2},\qquad EX^k=\frac{b^{k+1}-a^{k+1}}{(k+1)(b-a)},\qquad k>0 \text{ is an integer;}$$

(4) $$\operatorname{var}(X)=\frac{(b-a)^2}{12};$$

(5) $$M(t)=\frac{1}{t(b-a)}(e^{tb}-e^{ta}),\qquad t\ne 0.$$

Example 1. Let X have a PDF given by

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, & 0<x<\infty,\quad \lambda>0,\\ 0, & \text{otherwise.}\end{cases}$$

Then

$$F(x) = \begin{cases} 0 & x \le 0, \\ 1 - e^{-\lambda x}, & x > 0. \end{cases}$$

Let $Y = F(X) = 1 - e^{-\lambda X}$. The PDF of Y is given by

$$f_Y(y) = \frac{1}{\lambda} \cdot \frac{1}{1-y} \lambda e^{-\lambda(-1/\lambda)\log(1-y)} = 1, \quad 0 \le y < 1.$$

Let us define $f_Y(y) = 1$ at $y = 1$. Then we see that Y has density function

$$f_Y(y) = \begin{cases} 1, & 0 \le y \le 1, \\ 0, & \text{otherwise}, \end{cases}$$

which is the $U[0, 1]$ distribution. That this is not a mere coincidence is shown in the following theorem.

Theorem 1 (Probability Integral Transformation). Let X be an RV with a continuous DF F. Then $F(X)$ has the uniform distribution on $[0, 1]$.

The proof is left as an exercise.

The reader is asked to consider what happens in the case where F is the DF of a discrete RV. In the converse direction the following result holds.

Theorem 2. Let F be any DF, and let X be a $U[0, 1]$ RV. Then there exists a function h such that $h(X)$ has DF F, that is,

$$P\{h(X) \le x\} = F(x) \qquad \text{for all } x \in (-\infty, \infty). \tag{6}$$

Proof. If F is the DF of a discrete RV Y, let

$$P\{Y = y_k\} = p_k, \qquad k = 1, 2, \ldots.$$

Define h as follows:

$$h(x) = \begin{cases} y_1 & \text{if } 0 \le x < p_1, \\ y_2 & \text{if } p_1 \le x < p_1 + p_2, \\ \vdots & \vdots \end{cases}$$

Then

$$P\{h(X) = y_1\} = P\{0 \le X < p_1\} = p_1,$$
$$P\{h(X) = y_2\} = P\{p_1 \le X < p_1 + p_2\} = p_2,$$

and, in general,

$$P\{h(X) = y_k\} = p_k, \qquad k = 1, 2, \ldots .$$

Thus $h(X)$ is a discrete RV with DF F.

If F is continuous and strictly increasing, F^{-1} is well defined, and we take $h(X) = F^{-1}(X)$. We have

$$\begin{aligned} P\{h(X) \le x\} &= P\{F^{-1}(X) \le x\} \\ &= P\{X \le F(x)\} \\ &= F(x), \end{aligned}$$

as asserted.

In general, define

$$F^{-1}(y) = \inf\{x\colon F(x) \ge y\}, \tag{7}$$

and let $h(X) = F^{-1}(X)$. Then we have

$$\{F^{-1}(y) \le x\} = \{y \le F(x)\}. \tag{8}$$

Indeed, $F^{-1}(y) \le x$ implies that for every $\varepsilon > 0$, $y \le F(x + \varepsilon)$. Since $\varepsilon > 0$ is arbitrary and F is continuous on the right, we let $\varepsilon \to 0$ and conclude that $y \le F(x)$. Since $y \le F(x)$ implies that $F^{-1}(y) \le x$ by definition (7), it follows that (8) holds generally. Thus

$$P\{F^{-1}(X) \le x\} = P\{X \le F(x)\} = F(x).$$

Theorem 2 is quite useful in generating samples with the help of the uniform distribution.

Example 2. Let F be the DF defined by

$$F(x) = \begin{cases} 0, & x \le 0 \\ 1 - e^{-x}, & x > 0. \end{cases}$$

Then the inverse to $y = 1 - e^{-x}$, $x > 0$, is $x = -\log(1 - y)$, $0 < y < 1$. Thus

$$h(y) = -\log(1 - y),$$

and $-\log(1 - X)$ has the required distribution, where X is a $U[0, 1]$ RV.

Theorem 3. Let X be an RV defined on $[0, 1]$. If $P\{x < X \le y\}$ depends only on $y - x$ for all $0 \le x \le y \le 1$, then X is $U[0, 1]$.

Proof. Let $P\{x < X \le y\} = f(y-x)$; then $f(x+y) = P\{0 < X \le x+y\} = P\{0 < X \le x\} + P\{x < X \le x+y\} = f(x) + f(y)$. Note that f is continuous from the right. We have

$$f(x) = f(x) + f(0),$$

so that

$$f(0) = 0.$$

We will show that $f(x) = cx$ for some constant c. It suffices to prove the result for positive x. Let m be an integer; then

$$f(mx) = f(x) + \cdots + f(x) = mf(x).$$

Letting $x = n/m$, we get

$$f\left(m \cdot \left(\frac{n}{m}\right)\right) = mf\left(\frac{n}{m}\right),$$

so that

$$f\left(\frac{n}{m}\right) = \frac{1}{m} f(n) = \frac{n}{m} f(1),$$

for positive integers n and m. Letting $f(1) = c$, we have proved that

$$f(x) = cx$$

for rational numbers x.

To complete the proof we consider the case where x is a positive irrational number. Then we can find a decreasing sequence of positive rationals $x_1, x_2, \ldots$ such that $x_n \to x$. Since f is right continuous,

$$f(x) = \lim_{x_n \downarrow x} f(x_n) = \lim_{x_n \downarrow x} cx_n = cx.$$

Now, for $0 \le x \le 1$,

$$\begin{aligned} F(x) &= P\{X \le 0\} + P\{0 < X \le x\} \\ &= F(0) + P\{0 < X \le x\} \\ &= f(x) \\ &= cx, \qquad 0 \le x \le 1. \end{aligned}$$

Since $F(1) = 1$, we must have $c = 1$, so that

$$F(x) = x, \qquad 0 \le x \le 1.$$

This completes the proof.

5.3.2 Gamma Distribution

The integral

$$\Gamma(\alpha) = \int_{0+}^{\infty} x^{\alpha-1}e^{-x}\,dx \tag{9}$$

converges or diverges according as $\alpha > 0$ or ≤ 0. For $\alpha > 0$ the integral in (9) is called the *gamma function*. In particular, if $\alpha = 1$, $\Gamma(1) = 1$. If $\alpha > 1$, integration by parts yields

$$\Gamma(\alpha) = (\alpha - 1)\int_{0}^{\infty} x^{\alpha-2}e^{-x}\,dx = (\alpha - 1)\Gamma(\alpha - 1). \tag{10}$$

If $\alpha = n$ is a positive integer, then

$$\Gamma(n) = (n - 1)!. \tag{11}$$

Also writing $x = y^2/2$ in $\Gamma(\frac{1}{2})$, we see that

$$\Gamma\left(\frac{1}{2}\right) = \frac{1}{\sqrt{2}}\int_{-\infty}^{\infty} e^{-y^2/2}\,dy.$$

Now consider the integral $I = \int_{-\infty}^{\infty} e^{-y^2/2}\,dy$. We have

$$I^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left[-\frac{(x^2 + y^2)}{2}\right]dx\,dy,$$

and changing to polar coordinates, we get

$$I^2 = \int_{0}^{2\pi}\int_{0}^{\infty} r\exp\left(-\frac{r^2}{2}\right)dr\,d\theta = 2\pi.$$

It follows that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

Let us write $x = y/\beta$, $\beta > 0$, in the integral in (9). Then

$$\Gamma(\alpha) = \int_{0}^{\infty} \frac{y^{\alpha-1}}{\beta^{\alpha}}e^{-y/\beta}\,dy, \tag{12}$$

so that

$$\int_{0}^{\infty} \frac{1}{\Gamma(\alpha)\beta^{\alpha}}y^{\alpha-1}e^{-y/\beta}\,dy = 1. \tag{13}$$

Since the integrand in (13) is positive for $y > 0$, it follows that the function

$$f(y) = \begin{cases} \dfrac{1}{\Gamma(\alpha)\beta^\alpha} y^{\alpha-1} e^{-y/\beta}, & 0 < y < \infty, \\ 0, & y \le 0. \end{cases} \tag{14}$$

defines a PDF for $\alpha > 0$, $\beta > 0$.

Definition 2. An RV X with PDF defined by (14) is said to have a *gamma distribution* with parameters α and β. We will write $X \sim G(\alpha, \beta)$.

Figure 2 gives graphs of some gamma PDFs.

The DF of a $G(\alpha, \beta)$ RV is given by

$$F(x) = \begin{cases} 0, & x \le 0, \\ \dfrac{1}{\Gamma(\alpha)\beta^\alpha} \displaystyle\int_0^x y^{\alpha-1} e^{-y/\beta}\, dy, & x > 0. \end{cases} \tag{15}$$

The MGF of X is easily computed. We have

$$\begin{aligned} M(t) &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty e^{x(t-1/\beta)} x^{\alpha-1}\, dx \\ &= \left(\frac{1}{1-\beta t}\right)^\alpha \int_0^\infty \frac{y^{\alpha-1} e^{-y}}{\Gamma(\alpha)}\, dy, \qquad t < \frac{1}{\beta} \\ &= (1-\beta t)^{-\alpha}, \qquad t < \frac{1}{\beta}. \end{aligned} \tag{16}$$

It follows that

$$EX = M'(t)|_{t=0} = \alpha\beta \tag{17}$$

and

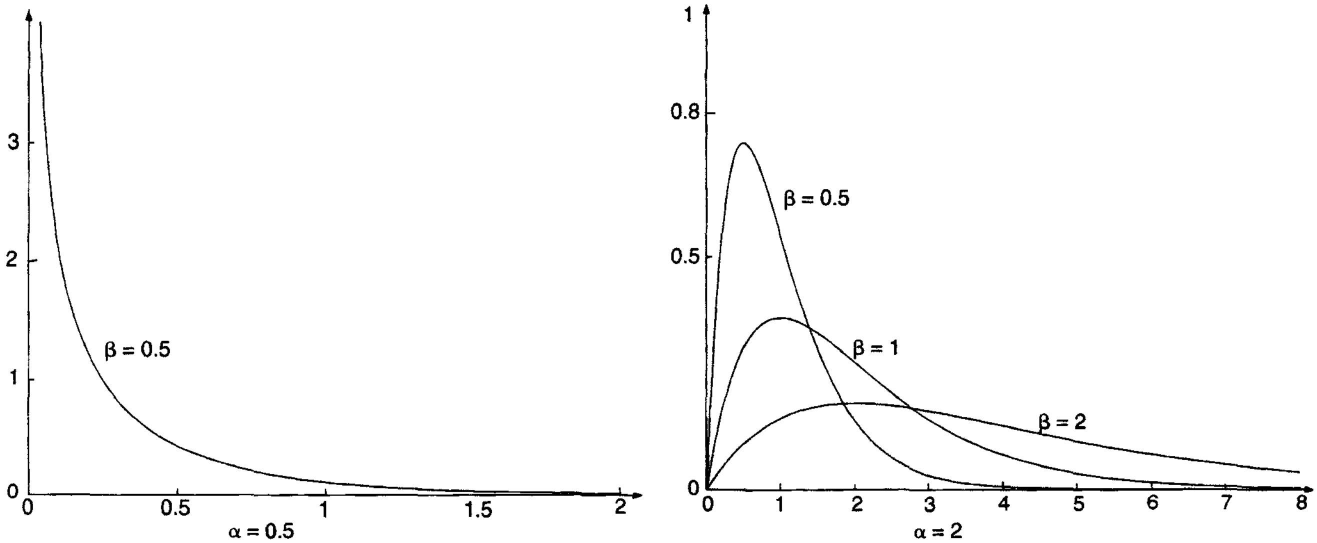
$$EX^2 = M''(t)|_{t=0} = \alpha(\alpha+1)\beta^2, \tag{18}$$

so that

$$\operatorname{var}(X) = \alpha\beta^2. \tag{19}$$

Indeed, we can compute the moment of order n such that $\alpha + n > 0$ directly from the density. We have

$$\begin{aligned} EX^n &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty e^{-x/\beta} x^{\alpha+n-1}\, dx \\ &= \beta^n \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)} \\ &= \beta^n (\alpha+n-1)(\alpha+n-2)\cdots\alpha \end{aligned} \tag{20}$$



(a) (b)

Fig. 2. Gamma density functions.

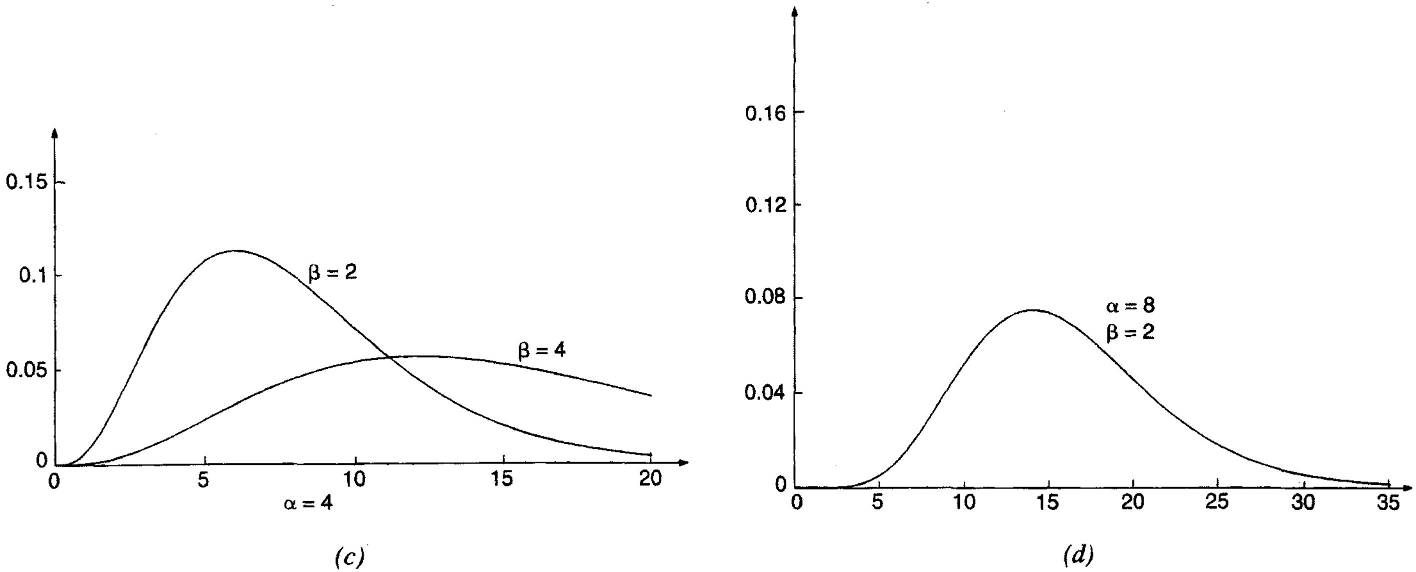


Fig. 2. (*continued*). Gamma density functions.

The special case when $\alpha = 1$ leads to the *exponential distribution with parameter* β. The PDF of an exponentially distributed RV is therefore

$$(21) \qquad f(x) = \begin{cases} \beta^{-1}e^{-x/\beta}, & x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Note that we can speak of the exponential distribution on $(-\infty, 0)$. The PDF of such an RV is

$$(22) \qquad f(x) = \begin{cases} \beta^{-1}e^{x/\beta}, & x < 0, \\ 0, & x \geq 0. \end{cases}$$

Clearly, if $X \sim G(1, \beta)$, we have

$$(23) \qquad EX^n = n!\,\beta^n$$

$$(24) \qquad EX = \beta \quad \text{and} \quad \text{var}(X) = \beta^2,$$

and

$$(25) \qquad M(t) = (1 - \beta t)^{-1} \qquad \text{for } t < \beta^{-1}.$$

Another special case of importance is when $\alpha = n/2$, $n > 0$ (an integer) and $\beta = 2$.

Definition 3. An RV X is said to have a *chi-square distribution* (χ^2*-distribution*) with n degrees of freedom where n is a positive integer if its PDF is given by

$$(26) \qquad f(x) = \begin{cases} \dfrac{1}{\Gamma(n/2)2^{n/2}}e^{-x/2}x^{n/2-1}, & 0 < x < \infty, \\ 0, & x \leq 0. \end{cases}$$

We will write $X \sim \chi^2(n)$ for a χ^2 RV with n degrees of freedom (d.f.).

If $X \sim \chi^2(n)$, then

$$(27) \qquad EX = n, \qquad \text{var}(X) = 2n,$$

$$(28) \qquad EX^k = \frac{2^k\Gamma[(n/2) + k]}{\Gamma(n/2)},$$

and

$$(29) \qquad M(t) = (1 - 2t)^{-n/2} \qquad \text{for } t < \tfrac{1}{2}.$$

Theorem 4. Let $X_1, X_2, \ldots, X_n$ be independent RVs such that $X_j \sim G(\alpha_j, \beta)$, $j = 1, 2, \ldots, n$. Then $S_n = \sum_{k=1}^n X_k$ is a $G(\sum_{j=1}^n \alpha_j, \beta)$ RV.

Corollary 1. Let $X_1, X_2, \ldots, X_n$ be iid RVs, each with an exponential distribution with parameter β. Then S_n is a $G(n, \beta)$ RV.

Corollary 2. If $X_1, X_2, \ldots, X_n$ are independent RVs such that $X_j \sim \chi^2(r_j)$, $j = 1, 2, \ldots, n$, then S_n is a $\chi^2(\sum_{i=1}^n r_j)$ RV.

Theorem 5. Let $X \sim U(0, 1)$. Then $Y = -2 \log X$ is $\chi^2(2)$.

Corollary. Let $X_1, X_2, \ldots, X_n$ be iid RVs with common distribution $U(0, 1)$. Then $-2 \sum_{i=1}^n \log X_i = 2 \log(1/\prod_{i=1}^n X_i)$ is $\chi^2(2n)$.

Theorem 6. Let $X \sim G(\alpha_1, \beta)$ and $Y \sim G(\alpha_2, \beta)$ be independent RVs. Then $X + Y$ and X/Y are independent.

Corollary. Let $X \sim G(\alpha_1, \beta)$ and $Y \sim G(\alpha_2, \beta)$ be independent RVs. Then $X + Y$ and $X/(X + Y)$ are independent.

The converse of Theorem 6 is also true. The result is due to Lukacs [66], and we state it without proof.

Theorem 7. Let X and Y be two nondegenerate RVs that take only positive values. Suppose that $U = X + Y$ and $V = X/Y$ are independent. Then X and Y have gamma distribution with the same parameter β.

Theorem 8. Let $X \sim G(1, \beta)$. Then the RV X has "no memory," that is,

$$P\{X > r + s | X > s\} = P\{X > r\} \tag{30}$$

for any two positive real numbers r and s.

The proof is left as an exercise.

The converse of Theorem 8 is also true in the following sense.

Theorem 9. Let F be a DF such that $F(x) = 0$ if $x < 0$, $F(x) < 1$ if $x > 0$, and

$$\frac{1 - F(x + y)}{1 - F(y)} = 1 - F(x) \qquad \text{for all } x, y > 0. \tag{31}$$

Then there exists a constant $\beta > 0$ such that

$$1 - F(x) = e^{-x\beta}, \qquad x > 0. \tag{32}$$

Proof. Equation (31) is equivalent to

$$g(x + y) = g(x) + g(y)$$

if we write $g(x) = \log\{1 - F(x)\}$. From the proof of Theorem 3 it is clear that the only right continuous solution is $g(x) = cx$. Hence $F(x) = 1 - e^{cx}$, $x \geq 0$. Since $F(x) \to 1$ as $x \to \infty$, it follows that $c < 0$ and the proof is complete.

Theorem 10. Let $X_1, X_2, \dots, X_n$ be iid RVs. Then $X_i \sim G(1, n\beta)$, $i = 1, 2, \dots, n$, if and only if $X_{(1)}$ is $G(1, \beta)$.

Note that, if $X_1, X_2, \dots, X_n$ are independent with $X_i \sim G(1, \beta_i), i = 1, 2, \dots, n$, then $X_{(1)}$ is a $G(1, 1/\sum_{i=1}^{b} \beta_i^{-1})$ RV.

The following result describes the relationship between exponential and Poisson RVs.

Theorem 11. Let $X_1, X_2, \dots$ be a sequence of iid RVs having common exponential density with parameter $\beta > 0$. Let $S_n = \sum_{k=1}^{n} X_k$ be the nth partial sum, $n = 1, 2, \dots$, and suppose that $t > 0$. If $Y =$ number of $S_n \in [0, t]$, then Y is a $P(t/\beta)$ RV.

Proof. We have

$$P\{Y = 0\} = P\{S_1 > t\} = \frac{1}{\beta} \int_t^\infty e^{-x/\beta} \, dx = e^{-t/\beta},$$

so that the assertion holds for $Y = 0$. Let n be a positive integer. Since the X_i's are nonnegative, S_n is nondecreasing, and

$$P\{Y = n\} = P\{S_n \leq t, \; S_{n+1} > t\}. \tag{33}$$

Now

$$P\{S_n \leq t\} = P\{S_n \leq t, \; S_{n+1} > t\} + P\{S_{n+1} \leq t\}. \tag{34}$$

It follows that

$$P\{Y = n\} = P\{S_n \leq t\} - P\{S_{n+1} \leq t\}, \tag{35}$$

and since $S_n \sim G(n, \beta)$, we have

$$\begin{aligned} P\{Y = n\} &= \int_0^t \frac{1}{\Gamma(n)\beta^n} x^{n-1} e^{-x/\beta} \, dx - \int_0^t \frac{1}{\Gamma(n+1)\beta^{n+1}} x^n e^{-x/\beta} \, dx \\ &= \frac{t^n e^{-t/\beta}}{\beta^n n!}, \end{aligned}$$

as asserted.

Theorem 12. If X and Y are independent exponential RVs with parameter β, then $Z = X/(X + Y)$ has a $U(0, 1)$ distribution.

Note that in view of Theorem 7, Theorem 12 characterizes the exponential distribution in the following sense. Let X and Y be independent RVs that are nondegenerate and take only positive values. Suppose that $X + Y$ and X/Y are independent. If $X/(X + Y)$ is $U(0, 1)$, X and Y both have the exponential distribution with parameter β. This follows since by Theorem 7, X and Y must have the gamma distribution with parameter β. Thus $X/(X + Y)$ must have (see Theorem 14) the PDF

$$f(x) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} x^{\alpha_1 - 1}(1 - x)^{\alpha_2 - 1}, \qquad 0 < x < 1,$$

and this is the uniform density on $(0, 1)$ if and only if $\alpha_1 = \alpha_2 = 1$. Thus X and Y both have the $G(1, \beta)$ distribution.

Theorem 13. Let X be a $P(\lambda)$ RV. Then

$$P\{X \le K\} = \frac{1}{K!}\int_\lambda^\infty e^{-x}x^K\,dx \tag{36}$$

expresses the DF of X in terms of an incomplete gamma function.

Proof.

$$\begin{aligned}\frac{d}{d\lambda}P\{X \le K\} &= \sum_{j=0}^{K} \frac{1}{j!}(je^{-\lambda}\lambda^{j-1} - \lambda^j e^{-\lambda})\\ &= \frac{-\lambda^K e^{-\lambda}}{K!},\end{aligned}$$

and it follows that

$$P\{X \le K\} = \frac{1}{K!}\int_\lambda^\infty e^{-x}x^K\,dx,$$

as asserted.

An alternative way of writing (36) is the following:

$$P\{X \le K\} = P\{Y \ge 2\lambda\},$$

where $X \sim P(\lambda)$, and $Y \sim \chi^2(2K + 2)$.

5.3.3 Beta Distribution

The integral

$$B(\alpha, \beta) = \int_{0+}^{1-} x^{\alpha-1}(1 - x)^{\beta-1}\,dx \tag{37}$$

converges for $\alpha > 0$, $\beta > 0$ and is called a *beta function*. For $\alpha \le 0$ or $\beta \le 0$ the integral in (37) diverges. It is easy to see that for $\alpha > 0$, $\beta > 0$,

$$B(\alpha, \beta) = B(\beta, \alpha), \tag{38}$$

$$B(\alpha, \beta) = \int_{0+}^{\infty} x^{\alpha-1}(1+x)^{-\alpha-\beta}\,dx, \tag{39}$$

and

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}. \tag{40}$$

It follows that

$$f(x) = \begin{cases} \dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}, & 0 < x < 1, \\ 0, & \text{otherwise}, \end{cases} \tag{41}$$

defines a PDF.

Definition 4. An RV X with PDF given by (41) is said to have a *beta distribution* with parameters α and β, $\alpha > 0$, $\beta > 0$. We will write $X \sim B(\alpha, \beta)$ for a beta variable with density (41).

Figure 3 gives graphs of some beta PDFs.

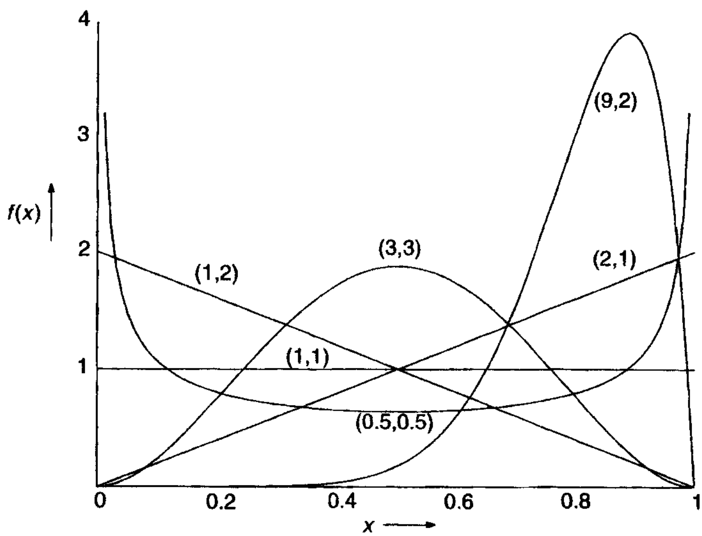


Fig. 3. Beta density functions

The DF of a $B(\alpha, \beta)$ RV is given by

$$(42)\qquad F(x) = \begin{cases} 0, & x \le 0, \\ [B(\alpha, \beta)]^{-1} \int_{0+}^{x} y^{\alpha-1}(1-y)^{\beta-1}\,dy, & 0 < x < 1, \\ 1, & x \ge 1. \end{cases}$$

If n is a positive number, then

$$(43)\qquad \begin{aligned} EX^n &= \frac{1}{B(\alpha, \beta)} \int_0^1 x^{n+\alpha-1}(1-x)^{\beta-1}\,dx \\ &= \frac{B(n+\alpha, \beta)}{B(\alpha, \beta)} = \frac{\Gamma(n+\alpha)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(n+\alpha+\beta)}, \end{aligned}$$

using (40). In particular,

$$(44)\qquad EX = \frac{\alpha}{\alpha+\beta}$$

and

$$(45)\qquad \operatorname{var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

For the MGF of $X \sim B(\alpha, \beta)$, we have

$$(46)\qquad M(t) = \frac{1}{B(\alpha, \beta)} \int_0^1 e^{tx} x^{\alpha-1}(1-x)^{\beta-1}\,dx.$$

Since moments of all order exist, and $E|X|^j < 1$ for all j, we have

$$(47)\qquad \begin{aligned} M(t) &= \sum_{j=0}^{\infty} \frac{t^j}{j!} EX^j \\ &= \sum_{j=0}^{\infty} \frac{t^j}{\Gamma(j+1)} \frac{\Gamma(\alpha+j)\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+j)\Gamma(\alpha)}. \end{aligned}$$

Remark 1. Note that in the special case where $\alpha = \beta = 1$ we get the uniform distribution on $(0, 1)$.

Remark 2. If X is a beta RV with parameters α and β, then $1 - X$ is a beta variate with parameters β and α. In particular, X is $B(\alpha, \alpha)$ if and only if $1 - X$ is $B(\alpha, \alpha)$. A special case is the uniform distribution on $(0, 1)$. If X and $1 - X$ have the same distribution, it does not follow that X has to be $B(\alpha, \alpha)$. All this entails is that the PDF satisfies

$$f(x) = f(1-x), \qquad 0 < x < 1.$$

Take

$$f(x) = \frac{1}{B(\alpha, \beta) + B(\beta, \alpha)}[x^{\alpha-1}(1-x)^{\beta-1} + (1-x)^{\alpha-1}x^{\beta-1}], \qquad 0 < x < 1.$$

Example 3. Let X be distributed with PDF

$$f(x) = \begin{cases} \frac{1}{12}x^2(1-x), & 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Then $X \sim B(3, 2)$ and

$$EX^n = \frac{\Gamma(n+3)\Gamma(5)}{\Gamma(3)\Gamma(n+5)} = \frac{4!}{2!} \cdot \frac{(n+2)!}{(n+4)!} = \frac{12}{(n+4)(n+3)},$$

$$EX = \frac{12}{20}, \qquad \text{var}(X) = \frac{6}{5^2 \cdot 6} = \frac{1}{25},$$

$$M(t) = \sum_{j=0}^{\infty} \frac{t^j}{j!} \cdot \frac{(j+2)!\,4!}{(j+4)!\,2!}$$

$$= \sum_{j=0}^{\infty} \frac{12}{(j+4)(j+3)} \cdot \frac{t^j}{j!},$$

and

$$P\{0.2 < X < 0.5\} = \frac{1}{12}\int_{0.2}^{0.5} (x^2 - x^3)\, dx = 0.023.$$

Theorem 14. Let X and Y be independent $G(\alpha_1, \beta)$ and $G(\alpha_2, \beta)$, respectively, RVs. Then $X/(X+Y)$ is a $B(\alpha_1, \alpha_2)$ RV.

Let $X_1, X_2, \ldots, X_n$ be iid RVs with the uniform distribution on $[0, 1]$. Let $X_{(k)}$ be the kth-order statistic.

Theorem 15. The RV $X_{(k)}$ has a beta distribution with parameters $\alpha = k$ and $\beta = n - k + 1$.

Proof. Let X be the number of X_i's that lie in $[0, t]$. Then X is $b(n, t)$. We have

$$P\{X_{(k)} \le t\} = P\{X \ge k\} = \sum_{j=k}^{n} \binom{n}{j} t^j (1-t)^{n-j}.$$

Also,

$$\begin{aligned}\frac{d}{dt}P\{X \geq k\} &= \sum_{j=k}^{n}\binom{n}{j}[jt^{j-1}(1-t)^{n-j} - (n-j)t^{j}(1-t)^{n-j-1}]\\ &= \sum_{j=k}^{n}\left[n\binom{n-1}{j-1}t^{j-1}(1-t)^{n-j} - n\binom{n-1}{j}t^{j}(1-t)^{n-j-1}\right]\\ &= n\binom{n-1}{k-1}t^{k-1}(1-t)^{n-k}.\end{aligned}$$

On integration, we get

$$P\{X_{(k)} \leq t\} = n\binom{n-1}{k-1}\int_0^t x^{k-1}(1-x)^{n-k}\,dx,$$

as asserted.

Remark 3. Note that we have shown that, if X is $b(n, p)$, then

$$1 - P\{X < k\} = n\binom{n-1}{k-1}\int_0^p x^{k-1}(1-x)^{n-k}\,dx, \tag{48}$$

which expresses the DF of X in terms of the DF of a $B(k, n-k+1)$ RV.

Theorem 16. Let $X_1, X_2, \ldots, X_n$ be independent RVs. Then $X_1, X_2, \ldots, X_n$ are iid $B(\alpha, 1)$ RVs if and only if $X_{(n)} \sim B(\alpha n, 1)$.

5.3.4 Cauchy Distribution

Definition 5. An RV X is said to have a *Cauchy distribution* with parameters μ and θ if its PDF is given by

$$f(x) = \frac{\mu}{\pi}\frac{1}{\mu^2 + (x-\theta)^2}, \qquad -\infty < x < \infty, \quad \mu > 0. \tag{49}$$

We will write $X \sim C(\mu, \theta)$ for a Cauchy RV with density (49).

Figure 4 gives graph of a Cauchy PDF.

We first check that (49) in fact defines a PDF. Substituting $y = (x-\theta)/\mu$, we get

$$\int_{-\infty}^{\infty} f(x)\,dx = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{dy}{1+y^2} = \frac{2}{\pi}(\tan^{-1} y)_0^{\infty} = 1.$$

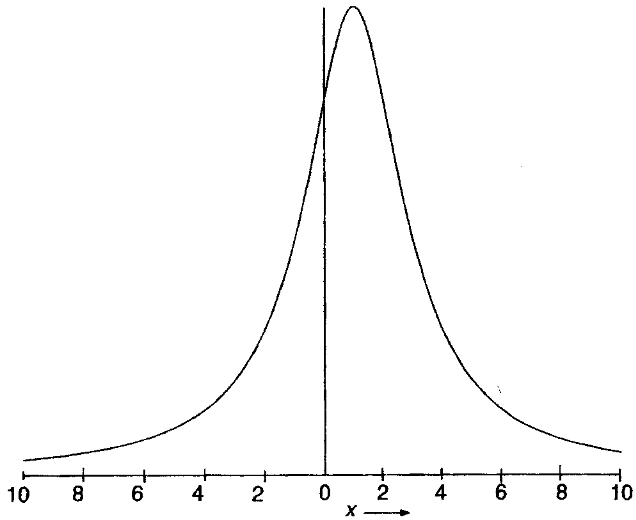


Fig. 4. Cauchy density function.

The DF of a $C(1, 0)$ RV is given by

(50) $$F(x) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1} x, \qquad -\infty < x < \infty.$$

Theorem 17. Let X be a Cauchy RV with parameters μ and θ. The moments of order < 1 exist, but the moments of order ≥ 1 do not exist for the RV X.

Proof. It suffices to consider the PDF

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

$$E|X|^\alpha = \frac{2}{\pi}\int_0^\infty x^\alpha \frac{1}{1+x^2}\,dx,$$

and, letting $z = 1/(1+x^2)$ in the integral, we get

$$E|X|^\alpha = \frac{1}{\pi}\int_0^1 z^{(1-\alpha)/2-1}(1-z)^{[(\alpha+1)/2]-1}\,dz,$$

which converges for $\alpha < 1$ and diverges for $\alpha \geq 1$. This completes the proof of the theorem.

It follows from Theorem 17 that the MGF of a Cauchy RV does not exist. This creates some manipulative problems. We note, however, that the cf of $X \sim C(\mu, 0)$ is given by

(51) $$\phi(t) = e^{-\mu|t|}.$$

Theorem 18. Let $X \sim \mathcal{C}(\mu_1, \theta_1)$ and $Y \sim \mathcal{C}(\mu_2, \theta_2)$ be independent RVs. Then $X + Y$ is a $\mathcal{C}(\mu_1 + \mu_2, \theta_1 + \theta_2)$ RV.

Proof. For notational convenience we will prove the result in the special case where $\mu_1 = \mu_2 = 1$ and $\theta_1 = \theta_2 = 0$, that is, where X and Y have the common PDF

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

The proof in the general case follows along the same lines. If $Z = X + Y$, the PDF of Z is given by

$$fz(z) = \frac{1}{\pi^2} \int_{-\infty}^{\infty} \frac{1}{1 + x^2} \cdot \frac{1}{1 + (z - x)^2}\, dx.$$

Now

$$\frac{1}{(1 + x^2)[1 + (z - x)^2]}$$
$$= \frac{1}{z^2(z^2 + 4)} \left[\frac{2zx}{1 + x^2} + \frac{z^2}{1 + x^2} + \frac{2z^2 - 2zx}{1 + (z - x)^2} + \frac{z^2}{1 + (z - x)^2} \right],$$

so that

$$fz(z) = \frac{1}{\pi^2} \frac{1}{z^2(z^2 + 4)} \left[z \log \frac{1 + x^2}{1 + (z - x)^2} + z^2 \tan^{-1} x + z^2 \tan^{-1}(x - z) \right]_{-\infty}^{\infty}$$
$$= \frac{1}{\pi} \frac{2}{z^2 + 2^2}, \qquad -\infty < z < \infty.$$

It follows that if X and Y are iid $\mathcal{C}(1, 0)$ RVs, then $X + Y$ is a $\mathcal{C}(2, 0)$ RV. We note that the result follows effortlessly from (51).

Corollary. Let $X_1, X_2, \ldots, X_n$ be independent Cauchy RVs, $X_k \sim \mathcal{C}(\mu_k, \theta_k)$, $k = 1, 2, \ldots, n$. Then $S_n = \sum_1^n X_k$ is a $\mathcal{C}(\sum_1^n \mu_k, \sum_1^n \theta_k)$ RV.

In particular, if $X_1, X_2, \ldots, X_n$ are iid $\mathcal{C}(1, 0)$ RVs, $n^{-1} S_n$ is also a $\mathcal{C}(1, 0)$ RV. This is a remarkable result, the importance of which will become clear in Chapter 6. Actually, this property uniquely characterizes the Cauchy distribution. If F is a nondegenerate DF with the property that $n^{-1} S_n$ also has DF F, then F must be a Cauchy distribution (see Thompson [112, p. 112]).

The proof of the following result is simple.

Theorem 19. Let X be $\mathcal{C}(\mu, 0)$. Then λ / X, where λ is a constant, is a $\mathcal{C}(|\lambda| / \mu, 0)$ RV.

Corollary. X is $\mathcal{C}(1, 0)$ if and only if $1/X$ is $\mathcal{C}(1, 0)$.

We emphasize that if X and $1/X$ have the same PDF on $(-\infty, \infty)$, it does not follow* that X is $C(1, 0)$, for let X be an RV with PDF

$$f(x) = \begin{cases} \dfrac{1}{4} & \text{if } |x| \le 1, \\ \dfrac{1}{4x^2} & \text{if } |x| > 1. \end{cases}$$

Then X and $1/X$ have the same PDF, as can easily be checked.

Theorem 20. Let X be a $U(-\pi/2, \pi/2)$ RV. Then $Y = \tan X$ is a Cauchy RV.

Many important properties of the Cauchy distribution can be derived from this result (see Pitman and Williams [78]).

5.3.5 Normal Distribution (Gaussian Law)

One of the most important distributions in the study of probability and mathematical statistics is the *normal distribution*, which we examine presently.

Definition 6. An RV X is said to have a *standard normal distribution* if its PDF is given by

$$(52) \qquad \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-(x^2/2)}, \qquad -\infty < x < \infty.$$

We first check that f defines a PDF. Let

$$I = \int_{-\infty}^{\infty} e^{-x^2/2}\, dx.$$

Then

$$0 < e^{-x^2/2} < e^{-|x|+1}, \qquad -\infty < x < \infty,$$

$$\int_{-\infty}^{\infty} e^{-|x|+1}\, dx = 2e,$$

and it follows that I exists. We have

$$I = \int_0^{\infty} y^{-1/2} e^{-y/2}\, dy$$

*Menon [71] has shown that we need the condition that both X and $1/X$ be *stable* to conclude that X is Cauchy.

A nondegenerate distribution function F is said to be stable if for two iid RVs X_1, X_2 with common DF F, and given constants $a_1, a_2 > 0$, we can find $\alpha > 0$ and $\beta(a_1, a_2)$ such that the RV

$$X_3 = \alpha^{-1}(a_1 x_1 + a_2 X_2 - \beta)$$

again has the same distribution F. Examples are the Cauchy (see the corollary to Theorem 18) and normal (discussed in Section 5.3.5) distributions.

$$= \Gamma\left(\frac{1}{2}\right) \cdot 2^{1/2}$$
$$= \sqrt{2\pi}.$$

Thus $\int_{-\infty}^{\infty} \varphi(x)\,dx = 1$, as required.

Let us write $Y = \sigma X + \mu$, where $\sigma > 0$. Then the PDF of Y is given by

$$\psi(y) = \frac{1}{\sigma}\varphi\left(\frac{y-\mu}{\sigma}\right)$$

$$(53) \qquad = \frac{1}{\sigma\sqrt{2\pi}} e^{-[(y-\mu)^2/2\sigma^2]}, \qquad -\infty < y < \infty; \quad \sigma > 0, \quad -\infty < \mu < \infty.$$

Definition 7. An RV X is said to have a *normal distribution* with parameters μ $(-\infty < \mu < \infty)$ and $\sigma (> 0)$ if its PDF is given by (53).

If X is a normally distributed RV with parameters μ and σ, we will write $X \sim \mathcal{N}(\mu, \sigma^2)$. In this notation, φ defined by (52) is the PDF of an $\mathcal{N}(0, 1)$ RV. The DF of an $\mathcal{N}(0, 1)$ RV will be denoted by $\Phi(x)$, where

$$(54) \qquad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\,du.$$

Clearly, if $X \sim \mathcal{N}(\mu, \sigma^2)$, then $Z = (X - \mu)/\sigma \sim N(0, 1)$. Z is called a *standard normal* RV. For the MGF of an $\mathcal{N}(\mu, \sigma^2)$ RV, we have

$$(55) \qquad M(t) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left(\frac{-x^2}{2\sigma^2} + x\frac{t\sigma^2 + \mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2}\right) dx$$
$$= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left[\frac{-(x - \mu - \sigma^2 t)^2}{2\sigma^2} + \mu t + \frac{\sigma^2 t^2}{2}\right] dx$$
$$= \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right),$$

for all real values of t. Moments of all order exist and may be computed from the MGF. Thus

$$(56) \qquad EX = M'(t)|_{t=0} = (\mu + \sigma^2 t)M(t)|_{t=0} = \mu$$

and

$$(57) \qquad EX^2 = M''(t)|_{t=0} = [M(t)\sigma^2 + (\mu + \sigma^2 t)^2 M(t)]_{t=0}$$
$$= \sigma^2 + \mu^2.$$

Thus

$$\text{(58)} \qquad \operatorname{var}(X) = \sigma^2.$$

Clearly, the central moments of odd order are all zero. The central moments of even order are as follows:

$$\begin{aligned}
\text{(59)}\quad E(X-\mu)^{2n} &= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} x^{2n} e^{-x^2/2\sigma^2}\,dx \qquad (n \text{ is a positive integer})\\
&= \frac{\sigma^{2n}}{\sqrt{2\pi}} \cdot 2^{n+1/2}\Gamma\left(n+\frac{1}{2}\right)\\
&= [(2n-1)(2n-3)\cdots 3\cdot 1]\sigma^{2n}.
\end{aligned}$$

As for the absolute moment of order α, for a standard normal RV Z we have

$$\begin{aligned}
\text{(60)}\qquad E|Z|^{\alpha} &= \frac{1}{\sqrt{2\pi}}\cdot 2\int_0^{\infty} z^{\alpha} e^{-z^2/2}\,dz\\
&= \frac{1}{\sqrt{2\pi}}\int_0^{\infty} y^{[(\alpha+1)/2)]-1} e^{-y/2}\,dy\\
&= \frac{\Gamma[(\alpha+1)/2]2^{\alpha/2}}{\sqrt{\pi}}.
\end{aligned}$$

As remarked earlier, the normal distribution is one of the most important distributions in probability and statistics, and for this reason the standard normal distribution is available in tabular form. Table ST2 at the end of the book gives the probability $P\{Z > z\}$ for various values of $z(> 0)$ in the tail of an $\mathcal{N}(0, 1)$ RV. In this book we write z_α for the value of Z that satisfies $\alpha = P\{Z > z_\alpha\}$, $0 \le \alpha \le 1$.

Example 4. By Chebychev's inequality, if $E|X|^2 < \infty$, $EX = \mu$, and $\operatorname{var}(X) = \sigma^2$, then

$$P\{|X-\mu| \ge K\sigma\} \le \frac{1}{K^2}.$$

For $K = 2$, we get $P\{|X-\mu| \ge K\sigma\} \le 0.25$, and for $K = 3$, we have $P\{|X-\mu| \ge K\sigma\} \le \frac{1}{9}$. If X is, in particular, $\mathcal{N}(\mu\ \sigma^2)$, then

$$P\{|X-\mu| \ge K\sigma\} = P\{|Z| \ge K\},$$

where Z is $\mathcal{N}(0, 1)$. From Table ST2,

$$P\{|Z| \ge 1\} = 0.318, \quad P\{|Z| \ge 2\} = 0.046, \quad \text{and} \quad P\{|Z| \ge 3\} = 0.002.$$

Thus practically all the distribution is concentrated within three standard deviations of the mean.

Example 5. Let $X \sim \mathcal{N}(3, 4)$. Then

$$\begin{aligned} P\{2 < X \le 5\} &= P\left\{\frac{2-3}{2} < \frac{X-3}{2} \le \frac{5-3}{2}\right\} = P\{-0.5 < Z \le 1\} \\ &= P\{Z \le 1\} - P\{Z \le -0.5\} \\ &= 0.841 - P\{Z \ge 0.5\} \\ &= 0.0841 - 0.309 = 0.532. \end{aligned}$$

Theorem 21 (Feller [22, p. 175]). Let Z be a standard normal RV. Then

$$P\{Z > x\} \approx \frac{1}{\sqrt{2\pi}\,x} e^{-x^2/2} \qquad \text{as } x \to \infty. \tag{61}$$

More precisely, for every $x > 0$,

$$\frac{1}{\sqrt{2\pi}} e^{-x^2/2}\left(\frac{1}{x} - \frac{1}{x^3}\right) < P\{Z > x\} < \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}. \tag{62}$$

Proof. We have

$$\frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-(1/2)y^2}\left(1 - \frac{3}{y^4}\right) dy = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\left(\frac{1}{x} - \frac{1}{x^3}\right) \tag{63}$$

and

$$\frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-y^2/2}\left(1 + \frac{1}{y^2}\right) dy = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\frac{1}{x}, \tag{64}$$

as can be checked on differentiation. Approximation (61) follows immediately.

Theorem 22. Let $X_1, X_2, \ldots, X_n$ be independent RVs with $X_k \sim \mathcal{N}(\mu_k, \sigma_k^2)$, $k = 1, 2, \ldots, n$. Then $S_n = \sum_{k=1}^n X_k$ is an $\mathcal{N}(\sum_{k=1}^n \mu_k, \sum_1^n \sigma_k^2)$ RV.

Corollary 1. If $X_1, X_2, \ldots, X_n$ are iid $\mathcal{N}(\mu, \sigma^2)$ RVs, then S_n is an $\mathcal{N}(n\mu, n\sigma^2)$ RV and $n^{-1}S_n$ is an $\mathcal{N}(\mu, \sigma^2/n)$ RV.

Corollary 2. If $X_1, X_2, \ldots, X_n$ are iid $\mathcal{N}(0, 1)$ RVs, then $n^{-1/2}S_n$ is also an $\mathcal{N}(0, 1)$ RV.

We remark that if $X_1, X_2, \ldots, X_n$ are iid RVs with $EX = 0$, $EX^2 = 1$ such that $n^{-1/2}S_n$ also has the same distribution for each $n = 1, 2, \ldots$, that distribution can only be $\mathcal{N}(0, 1)$. This characterization of the normal distribution will become clear when we study the central limit theorem in Chapter 6.

Theorem 23. Let X and Y be independent RVs. Then $X + Y$ is normally distributed if and only if X and Y are both normal.

If X and Y are independent normal RVs, $X + Y$ is normal by Theorem 22. The converse is due to Cramér [15] and will not be proved here.

Theorem 24. Let X and Y be independent RVs with common $\mathcal{N}(0, 1)$ distribution. Then $X + Y$ and $X - Y$ are independent.

The converse is due to Bernstein [3] and is stated here without proof.

Theorem 25. If X and Y are independent RVs with the same distribution, and if $Z_1 = X + Y$ and $Z_2 = X - Y$ are independent, all RVs X, Y, Z_1, and Z_2 are normally distributed.

The following result generalizes Theorem 24.

Theorem 26. If $X_1, X_2, \ldots, X_n$ are independent normal RVs and $\sum_{i=1}^n a_i b_i \operatorname{var}(X_i) = 0$, then $L_1 = \sum_{i=1}^n a_i X_i$ and $L_2 = \sum_{i=1}^n b_i X_i$ are independent. Here $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$ are fixed (nonzero) real numbers.

Proof. Let $\operatorname{var}(X_i) = \sigma_i^2$, and assume without loss of generality that $EX_i = 0$, $i = 1, 2, \ldots, n$. For any real numbers α, β, and t,

$$\begin{aligned}
Ee^{(\alpha L_1 + \beta L_2)t} &= E \exp\left[t \sum_1^n (\alpha a_i + \beta b_i) X_i\right] \\
&= \prod_{i=1}^n \exp\left[\frac{t^2}{2} (\alpha a_i + \beta b_i)^2 \sigma_i^2\right] \\
&= \exp\left(\frac{\alpha^2 t^2}{2} \sum_1^n a_i^2 \sigma_i^2 + \frac{\beta^2 t^2}{2} \sum_1^n b_i^2 \sigma_i^2\right) \quad \left(\text{since } \sum_i^n a_i b_i \sigma_i^2 = 0\right) \\
&= \prod_{i=1}^n \exp\left(\frac{t^2 \alpha^2}{2} a_i^2 \sigma_i^2\right) \prod_{i=1}^n \exp\left(\frac{t^2 \beta^2}{2} b_i^2 \sigma_i^2\right) \\
&= \prod_1^n Ee^{t\alpha a_i X_i} \prod_1^n Ee^{t\beta b_i X_i} \\
&= E \exp\left(t\alpha \sum_1^n a_i X_i\right) E \exp\left(t\beta \sum_1^n b_i X_i\right) = Ee^{\alpha t L_1} Ee^{\beta t L_2}.
\end{aligned}$$

Thus we have shown that

$$M(\alpha t, \beta t) = M(\alpha t, 0)M(0, \beta t) \qquad \text{for all } \alpha, \beta, t.$$

It follows that L_1 and L_2 are independent.

Corollary. If X_1, X_2 are independent $\mathcal{N}(\mu_1, \sigma^2)$ and $\mathcal{N}(\mu_2, \sigma^2)$ RVs, then $X_1 - X_2$ and $X_1 + X_2$ are independent. (This gives Theorem 24.)

Darmois [19] and Skitovitch [104] provided the converse of Theorem 26, which we state without proof.

Theorem 27. If $X_1, X_2, \ldots, X_n$ are independent RVs, $a_1, a_2, \ldots, a_n$, $b_1, b_2, \ldots, b_n$ are real numbers none of which equals zero, and if the linear forms

$$L_1 = \sum_{i=1}^{n} a_i X_i \quad \text{and} \quad L_2 = \sum_{i=1}^{n} b_i X_i$$

are independent, all the RVs are normally distributed.

Corollary. If X and Y are independent RVs such that $X + Y$ and $X - Y$ are independent, $X, Y, X + Y$, and $X - Y$ are all normal.

Yet another result of this type is the following theorem.

Theorem 28. Let $X_1, X_2, \ldots, X_n$ be iid RVs. Then the common distribution is normal if and only if

$$S_n = \sum_{k=1}^{n} X_k \quad \text{and} \quad Y_n = \sum_{i=1}^{n} (X_i - n^{-1} S_n)^2$$

are independent.

In Chapter 7 we prove the necessity part of this result, which is basic to the theory of t-tests in statistics (Chapter 10; see also Example 4.4.6). The sufficiency part was proved by Lukacs [65], and we will not prove it here.

Theorem 29. $X \sim \mathcal{N}(0, 1) \Rightarrow X^2 \sim \chi^2(1)$.

See Example 2.5.7 for the proof.

Corollary 1. If $X \sim \mathcal{N}(\mu, \sigma^2)$, the RV $Z^2 = (X - \mu)^2/\sigma^2$ is $\chi^2(1)$.

Corollary 2. If $X_1, X_2, \ldots, X_n$ are independent RVs and $X_k \sim \mathcal{N}(\mu_k, \sigma_k^2)$, $k = 1, 2, \ldots, n$, then $\sum_{k=1}^{n} (X_k - \mu_k)^2/\sigma_k^2$ is $\chi^2(n)$.

Theorem 30. Let X and Y be iid $\mathcal{N}(0, \sigma^2)$ RVs. Then X/Y is $\mathcal{C}(1, 0)$.

For the proof, see Example 2.5.7.

We remark that the converse of this result does not hold; that is, if $Z = X/Y$ is the quotient of two iid RVs and Z has a $\mathcal{C}(1, 0)$ distribution, it does not follow that X and Y are normal, for take X and Y to be iid with PDF

$$f(x) = \frac{\sqrt{2}}{\pi} \frac{1}{1 + x^4}, \qquad -\infty < x < \infty.$$

We leave the reader to verify that $Z = X/Y$ is $\mathcal{C}(1, 0)$.

5.3.6 Some Other Continuous Distributions

Several other distributions that are related to distributions studied earlier also arise in practice. We record briefly some of these and their important characteristics. We will use these distributions infrequently. We say that X has a *lognormal distribution* if $Y = \ln X$ has a normal distribution. The PDF of X is then

$$\text{(65)} \qquad f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\log x - \mu)^2}{2\sigma^2}\right], \qquad x \geq 0,$$

and $f(x) = 0$ for $x < 0$, where $-\infty < \mu < \infty$, $\sigma > 0$. In fact for $x > 0$

$$\begin{aligned} P(X \leq x) &= P(\ln X \leq \ln x) \\ &= P(Y \leq \ln x) = P\left(\frac{Y - \mu}{\sigma} \leq \frac{\ln x - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{\ln x - \mu}{\sigma}\right) \end{aligned}$$

where Φ is the DF of a $\mathcal{N}(0, 1)$ RV which easily leads to (65). It is easily seen that for $n \geq 0$,

$$\text{(66)} \begin{cases} EX^n = \exp\left(n\mu + \dfrac{n^2\sigma^2}{2}\right) \\ EX = \exp\left(\mu + \dfrac{\sigma^2}{2}\right), \quad \operatorname{var}(X) = \exp(2\mu + 2\sigma^2) - \exp(2\mu + \sigma^2). \end{cases}$$

The MGF of X does not exist.

We say that the RV X has a *Pareto* distribution with parameters $\theta > 0$ and $\alpha > 0$ if its PDF is given by

$$\text{(67)} \qquad f(x) = \frac{\alpha\theta^\alpha}{(x + \theta)^{\alpha+1}}, \quad x > 0$$

and zero otherwise. Here θ is scale parameter and α is a shape parameter. It is easy to check that

$$
(68)\qquad \begin{cases} F(x) = P(X \le x) = 1 - \dfrac{\theta^\alpha}{(\theta + x)^\alpha}, & x > 0 \\ EX = \dfrac{\theta}{\alpha - 1}, \quad \alpha > 1, \quad \text{and} \quad \operatorname{var}(X) = \dfrac{\alpha\theta^2}{(\alpha-2)(\alpha-1)^2} \end{cases}
$$

for $\alpha > 2$. The MGF of X does not exist since all moments of X do not.

Suppose that X has a Pareto distribution with parameters θ and α. Writing $Y = \ln(X/\theta)$, we see that Y has PDF

$$
(69)\qquad f_Y(y) = \frac{\alpha e^y}{(1+e^y)^{\alpha+1}}, \qquad -\infty < y < \infty,
$$

and DF

$$
F_Y(y) = 1 - (1 + e^y)^{-\alpha} \qquad \text{for all } y.
$$

The PDF in (69) is known as a *logistic distribution*. We introduce location and scale parameters μ and σ by writing $Z = \mu + \sigma Y$, taking $\alpha = 1$, and then the PDF of Z is easily seen to be

$$
(70)\qquad f_Z(z) = \frac{1}{\sigma} \frac{\exp[(z-\mu)/\sigma]}{\{1 + \exp[(z-\mu)/\sigma]\}^2}
$$

for all real z. This is the PDF of a logistic RV with location and scale parameters μ and σ. We leave the reader to check that

$$
(71)\qquad \begin{cases} F_Z(z) = \exp\left(\dfrac{z-\mu}{\sigma}\right)\left[1 + \exp\left(\dfrac{z-\mu}{\sigma}\right)\right]^{-1} \\ EZ = \mu, \qquad \operatorname{var}(Z) = \dfrac{\pi^2\sigma^2}{3} \\ M_Z(t) = \exp(\mu t)\Gamma(1 - \sigma t)\Gamma(1 + \sigma t), \qquad t < \dfrac{1}{\sigma}. \end{cases}
$$

Pareto distribution is also related to an exponential distribution. Let X have Pareto PDF of the form

$$
(72)\qquad f_X(s) = \frac{\alpha\sigma^\alpha}{x^{\alpha+1}}, \qquad x > \sigma
$$

and zero otherwise. A simple transformation leads to PDF (72) from (67). Then it is easily seen that $Y = \ln(X/\sigma)$ has an exponential distribution with mean $1/\alpha$. Thus some properties of exponential distribution that are preserved under monotone transformations can be derived for Pareto PDF (72) by using the logarithmic transformation.

Some other distributions are related to the gamma distribution. Suppose that $X \sim G(1, \beta)$. Let $Y = X^{1/\alpha}$, $\alpha > 0$. Then Y has PDF

$$f_Y(y) = \frac{\alpha}{\beta} y^{\alpha-1} \exp\left(\frac{-y^\alpha}{\beta}\right), \qquad y > 0 \tag{73}$$

and zero otherwise. The RV Y is said to have a *Weibull distribution*. We leave the reader to show that

$$\begin{cases} F_Y(y) = 1 - \exp\left(\frac{-y^\alpha}{\beta}\right), & y > 0 \\ EY^n = \beta^{n/\alpha}\Gamma\left(1 + \frac{n}{\alpha}\right), & EY = \beta^{1/\beta}\Gamma\left(1 + \frac{1}{\alpha}\right), \\ \operatorname{var}(Y) = \beta^{2/\alpha}\left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right)\right]. & \end{cases} \tag{74}$$

The MGF of Y exists only for $\alpha \geq 1$ but for $\alpha > 1$ it does not have a form useful in applications. The special case $\alpha = 2$, and $\beta = \theta^2$ is known as a *Rayleigh distribution*.

Suppose that X has a Weibull distribution with PDF (73). Let $Y = \ln X$. Then Y has DF

$$F_Y(y) = 1 - \exp\left(-\frac{1}{\beta}e^{\alpha y}\right), \qquad -\infty < y < \infty.$$

Setting $\theta = (1/\alpha)\ln\beta$ and $\sigma = 1/\alpha$, we get

$$F_Y(y) = 1 - \exp\left[-\exp\left(\frac{y - \theta}{\sigma}\right)\right] \tag{75}$$

with PDF

$$f_Y(y) = \frac{1}{\sigma}\exp\left[\frac{y - \theta}{\sigma} - \exp\left(\frac{y - \theta}{\sigma}\right)\right] \tag{76}$$

for $-\infty < y < \infty$ and $\sigma > 0$. An RV with PDF (76) is called an *extreme value distribution* with location and scale parameters θ and σ. It can be shown that

$$\begin{aligned} EY &= \theta - \gamma\sigma, \qquad \operatorname{var}(Y) = \frac{\pi^2\sigma^2}{6}, \\ M_Y(t) &= e^{\theta t}\Gamma(1 + \sigma t) \end{aligned} \tag{77}$$

where $\gamma \approx 0.577216$ is the Euler constant.

The final distribution we consider is also related to a $G(1, \beta)$ RV. Let f_1 be the PDF of $G(1, \beta)$ and f_2 the PDF

$$f_2(x) = \frac{1}{\beta}\exp\left(\frac{x}{\beta}\right), \qquad x < 0, \quad = 0 \text{ otherwise}.$$

Clearly, f_2 is also an exponential PDF defined on $(-\infty, 0)$. Consider the *mixture PDF*

(78) $$f(x) = \tfrac{1}{2}[f_1(x) + f_2(x)], \qquad -\infty < x < \infty.$$

Clearly,

(79) $$f(x) = \frac{1}{2}\exp\left(-\frac{|x|}{\beta}\right), \qquad -\infty < x < \infty$$

and the PDF f defined in (79) is called a *Laplace* or *double exponential PDF*. It is convenient to introduce a location parameter μ and consider instead the PDF

(80) $$f(x) = \frac{1}{2}\exp\left(-\frac{|x-\mu|}{\beta}\right), \qquad -\infty < x < \infty,$$

where $-\infty < \mu < \infty$, $\beta > 0$. It is easy to see that for RV X with PDF (80), we have

(81) $$EX = \mu, \quad \text{var}(X) = 2\beta^2, \quad \text{and} \quad M(t) = e^{\mu t}[1 - (\beta t)^2]^{-1},$$

for $|t| < 1/\beta$.

For completeness let us define a *mixture PDF* (PMF). Let $g(x|\theta)$ be a PDF and let $h(\theta)$ be a *mixing PDF*. Then the PDF

(82) $$f(x) = \int g(x|\theta)h(\theta)\,d\theta$$

is called a *mixture density function*. If h is a PMF with support set $\{\theta_1, \theta_2, \ldots, \theta_k\}$, then (82) reduces to a *finite mixture density function*

(83) $$f(x) = \sum_{i=1}^{k} g(x|\theta_i)h(\theta_i).$$

The quantities $h(\theta_i)$ are called *mixing proportions*. The PDF (78) is an example with $k = 2$, $h(\theta_1) = h(\theta_2) = \frac{1}{2}$, $g(x|\theta_1) = f_1(x)$, and $g(x|\theta_2) = f_2(x)$.

PROBLEMS 5.3

1. Prove Theorem 1.
2. Let X be an RV with PMF $p_k = P\{X = k\}$ given below. If F is the corresponding DF, find the distribution of $F(X)$, in the following cases:

   (a) $p_k = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 0, 1, 2, \ldots, n$; $0 < p < 1$.

   (b) $p_k = e^{-\lambda}(\lambda^k/k!)$, $k = 0, 1, 2, \ldots$; $\lambda > 0$.

3. Let $Y_1 \sim U[0,1]$, $Y_2 \sim U[0, Y_1], \dots, Y_n \sim U[0, Y_{n-1}]$. Show that

$$Y_1 \sim X_1, \quad Y_2 \sim X_1X_2, \quad \dots, \quad Y_n \sim X_1X_2\cdots X_n,$$

where $X_1, X_2, \dots, X_n$ are iid $U[0,1]$ RVs. If U is the number of $Y_1, Y_2, \dots, Y_n$ in $[t, 1]$, where $0 < t < 1$, show that U has a Poisson distribution with parameter $-\log t$.

4. Let $X_1, X_2, \dots, X_n$ be iid $U[0,1]$ RVs. Prove by induction or otherwise that $S_n = \sum_{k=1}^n X_k$ has the PDF

$$f_n(x) = [(n-1)!]^{-1} \sum_{k=0}^{n} (-1)^k \binom{n}{k} [\varepsilon(x-k)]^{n-1} (x-k)^{n-1},$$

where $\varepsilon(x) = 1$ if $x \geq 0$, $= 0$ if $x < 0$.

5. (a) Let X be an RV with PMF $p_j = P(X = x_j)$, $j = 0, 1, 2, \dots$, and let F be the DF of X. Show that

$$EF(X) = \frac{1}{2}\left(1 + \sum_{j=0}^{\infty} p_j^2\right)$$

and

$$\operatorname{var} F(X) = \sum_{j=0}^{\infty} p_j q_{j+1}^2 - \frac{1}{2}\left(1 - \sum_{j=0}^{\infty} p_j^2\right)^2$$

where $q_{j+1} = \sum_{i=j+1}^{\infty} p_i$.

(b) Let $p_j > 0$ for $j = 0, 1, \dots, N$ and $\sum_{j=0}^{N} p_j = 1$. Show that

$$EF(X) \geq \frac{N+2}{2(N+1)}$$

with equality if and only if $p_j = 1/(N+1)$ for all j. (Rohatgi [89])

6. Prove (a) Theorem 6 and its corollary, and (b) Theorem 10.

7. Let X be a nonnegative RV of the continuous type, and let $Y \sim U(0, X)$. Also, let $Z = X - Y$. Then the RVs Y and Z are independent if and only if X is $G(2, 1/\lambda)$ for some $\lambda > 0$. (Lamperti [57])

8. Let X and Y be independent RVs with common PDF $f(x) = \beta^{-\alpha}\alpha x^{\alpha-1}$ if $0 < x < \beta$, and $= 0$ otherwise; $\alpha \geq 1$. Let $U = \min(X, Y)$ and $V = \max(X, Y)$. Find the joint PDF of U and V and the PDF of $U + V$. Show that U/V and V are independent.

9. Prove Theorem 14.

10. Prove Theorem 8.

11. Prove Theorems 19 and 20.

12. Let $X_1, X_2, \ldots, X_n$ be independent RVs with $X_i \sim \mathcal{C}(\mu_i, \lambda_i)$, $i = 1, 2, \ldots, n$. Show that the RV $X = 1/\sum_{i=1}^{n} X_i^{-1}$ is also a Cauchy RV with parameters $\mu/(\lambda^2 + \mu^2)$ and $\lambda/(\lambda^2 + \mu^2)$, where

$$\lambda = \sum_{i=1}^{n} \frac{\lambda_i}{\lambda_i^2 + \mu_i^2} \quad \text{and} \quad \mu = \sum_{i=1}^{n} \frac{\mu_i}{\lambda_i^2 + \mu_i^2}.$$

13. Let $X_1, X_2, \ldots, X_n$ be iid $\mathcal{C}(1, 0)$ RVs and $a_i \neq 0$, b_i, $i = 1, 2, \ldots, n$, be any real numbers. Find the distribution of $\sum_{i=1}^{n} 1/(a_i X_i + b_i)$.

14. Suppose that the load of an airplane wing is a random variable X with $\mathcal{N}(1000, 14400)$ distribution. The maximum load that the wing can withstand is an RV Y, which is $\mathcal{N}(1260, 2500)$. If X and Y are independent, find the probability that the load encountered by the wing is less than its critical load.

15. Let $X \sim \mathcal{N}(0, 1)$. Find the PDF of $Z = 1/X^2$. If X and Y are iid $\mathcal{N}(0, 1)$, deduce that $U = XY/\sqrt{X^2 + Y^2}$ is $\mathcal{N}(0, \frac{1}{4})$.

16. In Problem 15 let X and Y be independent normal RVs with zero means. Show that $U = XY/\sqrt{X^2 + Y^2}$ is normal. If, in addition, $\text{var}(X) = \text{var}(Y)$, show that $V = (X^2 - Y^2)/\sqrt{X^2 + Y^2}$ is also normal. Moreover, U and V are independent. (Shepp [102])

17. Let X_1, X_2, X_3, X_4 be independent $\mathcal{N}(0, 1)$. Show that $Y = X_1X_2 + X_3X_4$ has the PDF $f(y) = \frac{1}{2}e^{-|y|}$, $-\infty < y < \infty$.

18. Let $X \sim \mathcal{N}(15, 16)$. Find (a) $P\{X \leq 12\}$, (b) $P\{10 \leq X \leq 17\}$, (c) $P\{10 \leq X \leq 19 \mid X \leq 17\}$, and (d) $P\{|X - 15| \geq 0.5\}$.

19. Let $X \sim \mathcal{N}(-1, 9)$. Find x such that $P\{X > x\} = 0.38$. Also find x such that $P\{|X + 1| < x\} = 0.4$.

20. Let X be an RV such that $\log(X - a)$ is $\mathcal{N}(\mu, \sigma^2)$. Show that X has PDF

$$f(x) = \begin{cases} \dfrac{1}{\sigma(x-a)\sqrt{2\pi}} \exp\left\{-\dfrac{[\log(x-a) - \mu]^2}{2\sigma^2}\right\} & \text{if } x > a, \\ 0 & \text{if } x \leq a. \end{cases}$$

If m_1, m_2 are the first two moments of this distribution and $\alpha_3 = \mu_3/\mu_2^{3/2}$ is the coefficient of skewness, show that a, μ, σ are given by

$$a = m_1 - \frac{\sqrt{m_2 - m_1^2}}{\eta}, \qquad \sigma^2 = \log(1 + \eta^2),$$

and

$$\mu = \log(m_1 - a) - \tfrac{1}{2}\sigma^2,$$

where η is the real root of the equation $\eta^3 + 3\eta - \alpha_3 = 0$.

21. Let $X \sim G(\alpha, \beta)$ and let $Y \sim U(0, X)$.

(a) Find the PDF of Y.

(b) Find the conditional PDF of X given $Y = y$.

(c) Find $P(X + Y \le 2)$.

22. Let X and Y be iid $\mathcal{N}(0, 1)$ RVs. Find the PDF of $X/|Y|$. Also, find the PDF of $|X|/|Y|$.

23. It is known that $X \sim B(\alpha, \beta)$, and $P(X < 0.2) = 0.22$. If $\alpha + \beta = 26$, find α and β. (*Hint:* Use Table ST1.)

24. Let $X_1, X_2, \ldots, X_n$ be iid $\mathcal{N}(\mu, \sigma^2)$ RVs. Find the distribution of

$$Y_n = \frac{\sum_{k=1}^n kX_k - \mu \sum_{k=1}^n k}{\left(\sum_{k=1}^n k^2\right)^{1/2}}.$$

25. Let $F_1, F_2, \ldots, F_n$ be n DFs. Show that $\min[F_1(x_1), F_2(x_2), \ldots, F_n(x_n)]$ is an n-dimensional DF with marginal DFs $F_1, F_2, \ldots, F_n$. (Kemp [48])

26. Let $X \sim NB(1; p)$ and $Y \sim G(1, 1/\lambda)$. Show that X and Y are related by the equation

$$P\{X \le x\} = P\{Y \le [x]\} \quad \text{for } x > 0, \qquad \lambda = \log\left(\frac{1}{1-p}\right).$$

where $[x]$ is the largest integer $\le x$. Equivalently, show that

$$P\{Y \in (n, n+1]\} = P_\theta\{X = n\},$$

where $\theta = 1 - e^{-\lambda}$. (Prochaska [80])

27. Let T be an RV with DF F and write $S(t) = 1 - F(t) = P(T > t)$. The function F is called the *survival* (or *reliability*) *function* of X (or DF F). The function $\lambda(t) = f(t)/S(t)$ is called the *hazard* (or *failure-rate*) *function*. For the following PDF, find the hazard function:

(a) Rayleigh: $f(t) = (t/\alpha^2)\exp(-t^2/2\alpha^2),\ t > 0$.

(b) Lognormal: $f(t) = 1/(t\sigma\sqrt{2\pi})\exp[-(\ln\, t - \mu)^2/2\sigma^2]$.

(c) Pareto: $f(t) = \alpha\theta^\alpha/t^{\alpha+1},\ t > \theta$, and $= 0$ otherwise.

(d) Weibull: $f(t) = (\alpha/\beta)t^{\alpha-1}\exp(-t^\alpha/\beta),\ t > 0$.

(e) Logistic: $f(t) = (1/\beta)\exp[-(t-\mu)/\beta]\{1 + \exp[-(t-\mu)/\beta]\}^{-2}$, $-\infty < t < \infty$.

28. Consider the PDF

$$f(x) = \left(\frac{\lambda}{2\pi x^3}\right)^{1/2} \exp\left[-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right], \qquad x > 0$$

and $= 0$ otherwise. An RV X with PDF f is said to have an *inverse Gaussian distribution* with parameters μ and λ, both positive. Show that

$$EX = \mu, \quad \operatorname{var}(X) = \frac{\mu^3}{\lambda}, \quad \text{and}$$

$$M(t) = E\exp(tX) = \exp\left\{\frac{\lambda}{\mu}\left[1 - \left(1 - \frac{2t\mu^2}{\lambda}\right)^{1/2}\right]\right\}.$$

29. Let f be the PDF of a $\mathcal{N}(\mu, \sigma^2)$ RV.

(a) For what value of c is the function cf^n, $n > 0$, a PDF?

(b) Let Φ be the DF of $Z \sim \mathcal{N}(0, 1)$. Find $E[Z\Phi(Z)]$ and $E[Z^2\Phi(Z)]$.

5.4 BIVARIATE AND MULTIVARIATE NORMAL DISTRIBUTIONS

In this section we introduce the bivariate and multivariate normal distributions and investigate some of their important properties. We note that bivariate analogs of other PDFs are known, but they are not always uniquely identified. For example, there are several versions of bivariate exponential PDFs so-called because each has exponential marginals. We will not encounter any of these bivariate PDFs in this book.

Definition 1. A two-dimensional RV (X, Y) is said to have a *bivariate normal distribution* if the joint PDF is of the form

$$(1) \quad f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-Q(x,y)/2}, \qquad -\infty < x < \infty, \qquad -\infty < y < \infty,$$

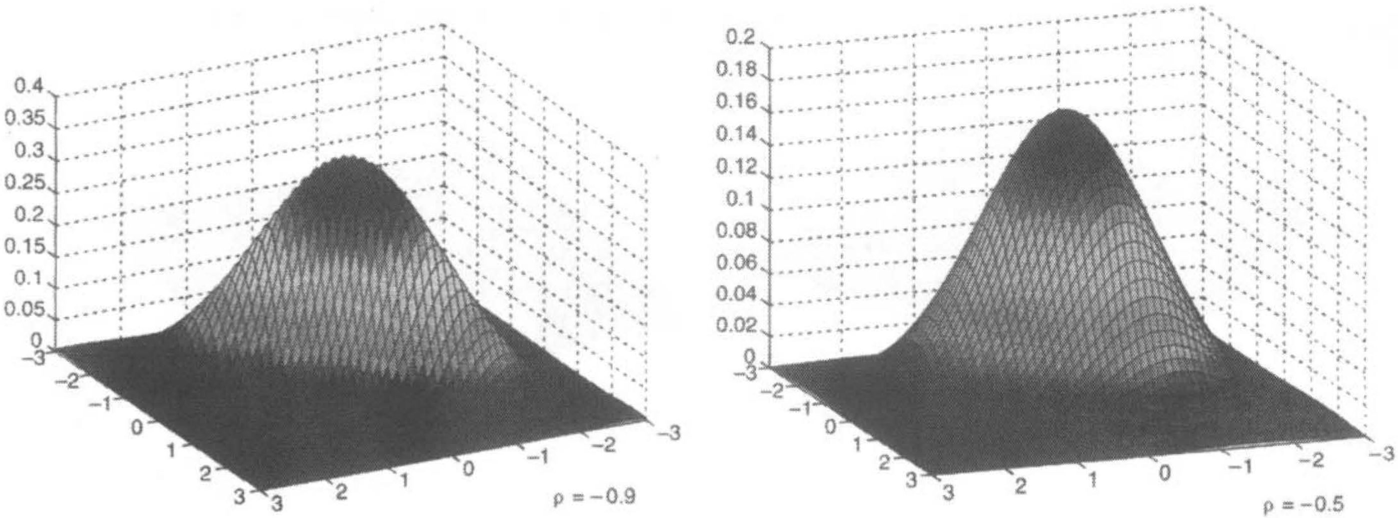
where $\sigma_1 > 0, \sigma_2 > 0, |\rho| < 1$, and Q is the positive definite quadratic form

$$(2) \quad Q(x, y) = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{x-\mu_1}{\sigma_1}\frac{y-\mu_2}{\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right].$$

Figure 1 gives graphs of bivariate normal PDF for selected values of ρ.

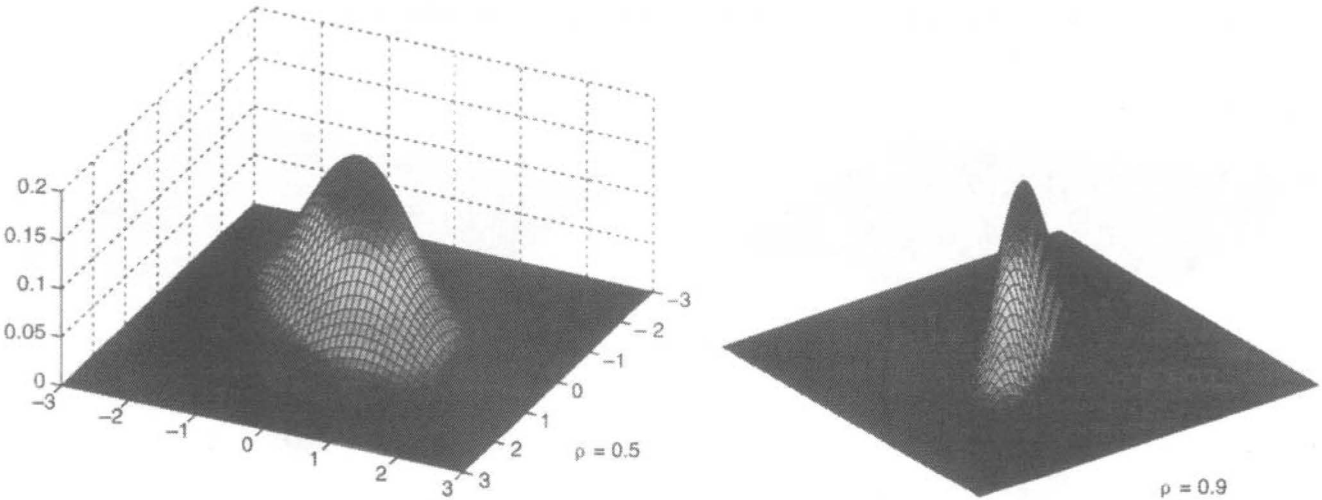
We first show that (1) indeed defines a joint PDF. In fact, we prove the following result.

Theorem 1. The function defined by (1) and (2) with $\sigma_1 > 0, \sigma_2 > 0, |\rho| < 1$ is a joint PDF. The marginal PDFs of X and Y are, respectively, $\mathcal{N}(\mu_1, \sigma_1^2)$ and $\mathcal{N}(\mu_2, \sigma_2^2)$, and ρ is the correlation coefficient between X and Y.



(a) *(b)*

Fig. 1. Bivariate normal with $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$, and $\rho = -0.9, -0.5, 0.5, 0.9$.



(c) (d)

Fig. 1. (*continued*). Bivariate normal with $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$, and $\rho = -0.9, -0.5, 0.5, 0.9$.

Proof. Let $f_1(x) = \int_{-\infty}^{\infty} f(x, y)\,dy$. Note that

$$(1-\rho^2)Q(x, y) = \left(\frac{y-\mu_2}{\sigma_2} - \rho\frac{x-\mu_1}{\sigma_1}\right)^2 + (1-\rho^2)\left(\frac{x-\mu_1}{\sigma_1}\right)^2$$

$$= \left\{\frac{y-[\mu_2 + \rho(\sigma_2/\sigma_1)(x-\mu_1)]}{\sigma_2}\right\}^2 + (1-\rho^2)\left(\frac{x-\mu_1}{\sigma_1}\right)^2.$$

It follows that

$$f_1(x) = \frac{1}{\sigma_1\sqrt{2\pi}}\exp\left[\frac{-(x-\mu_1)^2}{2\sigma_1^2}\right]\int_{-\infty}^{\infty}\frac{\exp\{-(y-\beta_x)^2/[2\sigma_2^2(1-\rho^2)]\}}{\sigma_2\sqrt{1-\rho^2}\sqrt{2\pi}}\,dy, \tag{3}$$

where we have written

$$\beta_x = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1). \tag{4}$$

The integrand is the PDF of an $\mathcal{N}(\beta_x, \sigma_2^2(1-\rho^2))$ RV, so that

$$f_1(x) = \frac{1}{\sigma_1\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right], \qquad -\infty < x < \infty.$$

Thus

$$\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} f(x, y)\,dy\right]dx = \int_{-\infty}^{\infty} f_1(x)\,dx = 1,$$

and $f(x, y)$ is a joint PDF of two RVs of the continuous type. It also follows that f_1 is the marginal PDF of X, so that X is $\mathcal{N}(\mu_1, \sigma_1^2)$. In a similar manner we can show that Y is $\mathcal{N}(\mu_2, \sigma_2^2)$.

Furthermore, we have

$$\frac{f(x, y)}{f_1(x)} = \frac{1}{\sigma_2\sqrt{1-\rho^2}\sqrt{2\pi}}\exp\left[\frac{-(y-\beta_x)^2}{2\sigma_2^2(1-\rho^2)}\right], \tag{5}$$

where β_x is given by (4). It is clear, then, that the conditional PDF $f_{Y|X}(y \mid x)$ given by (5) is also normal, with parameters β_x and $\sigma_2^2(1-\rho^2)$. We have

$$E\{Y \mid x\} = \beta_x = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1) \tag{6}$$

and

$$\operatorname{var}\{Y|x\} = \sigma_2^2(1-\rho^2). \tag{7}$$

In order to show that ρ is the correlation coefficient between X and Y, it suffices to show that $\operatorname{cov}(X, Y) = \rho\sigma_1\sigma_2$. We have from (6)

$$\begin{aligned} E(XY) &= E\{E\{XY|X\}\} \\ &= E\left\{X\left[\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(X - \mu_1)\right]\right\} \\ &= \mu_1\mu_2 + \frac{\rho\sigma_2}{\sigma_1}\sigma_1^2. \end{aligned}$$

It follows that

$$\operatorname{cov}(X, Y) = E(XY) - \mu_1\mu_2 = \rho\sigma_1\sigma_2.$$

Remark 1. If $\rho^2 = 1$, then (1) becomes meaningless. But in that case we know (Theorem 4.5.1) that there exist constants a and b such that $P\{Y = aX + b\} = 1$. We thus have a univariate distribution, which is called the ***bivariate degenerate*** (or ***singular***) ***normal distribution***. The bivariate degenerate normal distribution does not have a PDF but corresponds to an RV (X, Y) whose marginal distributions are normal or degenerate and are such that (X, Y) falls on a fixed line with probability 1. It is for this reason that degenerate distributions are considered as normal distributions with variance 0.

Next we compute the MGF $M(t_1, t_2)$ of a bivariate normal RV (X, Y). If $f(x, y)$ is the PDF given in (1) and f_1 is the marginal PDF of X, we have

$$\begin{aligned} M(t_1, t_2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t_1x+t_2y} f(x, y)\,dx\,dy, \\ &= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} f_{Y|X}(y \mid x)e^{t_2y}\,dy\right] e^{t_1x} f_1(x)\,dx \\ &= \int_{-\infty}^{\infty} e^{t_1x} f_1(x)\left\{\exp\left[\frac{1}{2}\sigma_2^2t_2^2(1-\rho^2) + t_2\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1)\right)\right]\right\} dx \\ &= \exp\left[\frac{1}{2}\sigma_2^2t_2^2(1-\rho^2) + t_2\mu_2 - \rho t_2\frac{\sigma_2}{\sigma_1}\mu_1\right]\int_{-\infty}^{\infty} e^{t_1x}e^{(\rho\sigma_2/\sigma_1)xt_2} f_1(x)\,dx. \end{aligned}$$

Now

$$\int_{-\infty}^{\infty} e^{(t_1+\rho t_2\sigma_2/\sigma_1)x} f_1(x)\,dx = \exp\left[\mu_1\left(t_1 + \rho\frac{\sigma_2}{\sigma_1}t_2\right) + \frac{1}{2}\sigma_1^2\left(t_1 + \rho t_2\frac{\sigma_2}{\sigma_1}\right)^2\right].$$

Therefore,

$$M(t_1, t_2) = \exp\left(\mu_1t_1 + \mu_2t_2 + \frac{\sigma_1^2t_1^2 + \sigma_2^2t_2^2 + 2\rho\sigma_1\sigma_2t_1t_2}{2}\right). \tag{8}$$

The following result is an immediate consequence of (8).

Theorem 2. If (X, Y) has a bivariate normal distribution, X and Y are independent if and only if $\rho = 0$.

Remark 2. It is quite possible for an RV (X, Y) to have a bivariate density such that the marginal densities of X and Y are normal and the correlation coefficient is 0, yet X and Y are not independent. Indeed, if the marginal densities of X and Y are normal, it does not follow that the joint density of (X, Y) is a bivariate normal. Let

$$(9) \qquad f(x, y) = \frac{1}{2}\left\{\frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left[\frac{-1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right] + \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left[\frac{-1}{2(1-\rho^2)}(x^2 + 2\rho xy + y^2)\right]\right\}.$$

Here $f(x, y)$ is a joint PDF such that both marginal densities are normal, $f(x, y)$ is not bivariate normal, and X and Y have zero correlation. But X and Y are not independent. We have

$$f_1(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad -\infty < x < \infty,$$

$$f_2(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}, \qquad -\infty < y < \infty,$$

and

$$EXY = 0.$$

Example 1 (Rosenberg [91]). Let f and g be PDFs with corresponding DFs F and G. Also, let

$$(10) \qquad h(x, y) = f(x)g(y)[1 + \alpha(2F(x) - 1)(2G(y) - 1)],$$

where $|\alpha| \le 1$ is a constant. It was shown in Example 4.3.1 that h is a bivariate density function with given marginal densities f and g.

In particular, take f and g to be the PDF of $\mathcal{N}(0, 1)$, that is,

$$(11) \qquad f(x) = g(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad -\infty < x < \infty,$$

and let (X, Y) have the joint PDF $h(x, y)$. We will show that $X + Y$ is not normal except in the trivial case $\alpha = 0$, when X and Y are independent.

Let $Z = X + Y$. Then

$$EZ = 0, \qquad \operatorname{var}(Z) = \operatorname{var}(X) + \operatorname{var}(Y) + 2\operatorname{cov}(X, Y).$$

It is easy to show (Problem 2) that $\text{cov}(X, Y) = \alpha/\pi$, so that $\text{var}(Z) = 2[1+(\alpha/\pi)]$. If Z is normal, its MGF must be

$$M_z(t) = e^{t^2[1+(\alpha/\pi)]}. \tag{12}$$

Next we compute the MGF of Z directly from the joint PDF (10). We have

$$\begin{aligned}
M_1(t) &= E\{e^{tX+tY}\} \\
&= e^{t^2} + \alpha \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{tx+ty}[2F(x)-1][2F(y)-1]f(x)f(y)\,dx\,dy \\
&= e^{t^2} + \alpha \left\{\int_{-\infty}^{\infty} e^{tx}[2F(x)-1]f(x)\,dx\right\}^2.
\end{aligned}$$

Now

$$\begin{aligned}
\int_{-\infty}^{\infty} e^{tx}[2F(x)-1]f(x)\,dx &= -2\int_{-\infty}^{\infty} e^{tx}[1-F(x)]f(x)\,dx + e^{t^2/2} \\
&= e^{t^2/2} - 2\int_{-\infty}^{\infty}\int_{x}^{\infty} \frac{1}{2\pi}\exp\left[-\tfrac{1}{2}(x^2+u^2-2tx)\right] du\,dx \\
&= e^{t^2/2} - \int_{-\infty}^{\infty}\int_{0}^{\infty} \frac{\exp\left\{-\frac{1}{2}[x^2+(v+x)^2-2tx]\right\}}{\pi}\,dv\,dx \\
&= e^{t^2/2} - \int_{0}^{\infty} \frac{\exp[-v^2/2+(v-t)^2/4]}{\sqrt{\pi}} \int_{-\infty}^{\infty} \frac{\exp\{-[x+(v-t)/2]^2\}}{\sqrt{\pi}}\,dx\,dv \\
&= e^{t^2/2} - 2e^{t^2/2}\int_{0}^{\infty} \frac{\exp\left\{-\frac{1}{2}[(v+t)^2/2]\right\}}{2\sqrt{\pi}}\,dv \\
&= e^{t^2/2} - 2e^{t^2/2}P\left\{Z_1 > \frac{t}{\sqrt{2}}\right\},
\end{aligned} \tag{13}$$

where Z_1 is an $\mathcal{N}(0, 1)$ RV.

It follows that

$$\begin{aligned}
M_1(t) &= e^{t^2} + \alpha\left(e^{t^2/2} - 2e^{t^2/2}P\left\{Z_1 > \frac{1}{\sqrt{2}}\right\}\right)^2 \\
&= e^{t^2}\left[1+\alpha\left(1-2P\left\{Z_1 > \frac{t}{\sqrt{2}}\right\}\right)^2\right].
\end{aligned} \tag{14}$$

If Z were normally distributed, we must have $M_Z(t) = M_1(t)$ for all t and all $|\alpha| \leq 1$, that is,

$$e^{t^2}e^{(\alpha/\pi)t^2} = e^{t^2}\left[1+\alpha\left(1-2P\left\{Z_1 > \frac{t}{\sqrt{2}}\right\}\right)^2\right]. \tag{15}$$

For $\alpha = 0$, the equality clearly holds. The expression within the brackets on the right side of (15) is bounded by $1+\alpha$, whereas the expression $e^{(\alpha/\pi)t^2}$ is unbounded, so the equality cannot hold for all t and α.

Next we investigate the multivariate normal distribution of dimension n, $n \geq 2$. Let $\mathbf{M}$ be an $n \times n$ real, symmetric, and positive definite matrix. Let $\mathbf{x}$ denote the $n \times 1$ column vector of real numbers $(x_1, x_2, \ldots, x_n)'$, and let $\boldsymbol{\mu}$ denote the column vector $(\mu_1, \mu_2, \ldots, \mu_n)'$, where $\mu_i\ (i = 1, 2, \ldots, n)$ are real constants.

Theorem 3. The nonnegative function

$$f(\mathbf{x}) = c\exp\left[-\frac{(\mathbf{x}-\boldsymbol{\mu})'\mathbf{M}(\mathbf{x}-\boldsymbol{\mu})}{2}\right], \qquad -\infty < x_i < \infty, \quad i = 1, 2, \ldots, n, \tag{16}$$

defines the joint PDF of some random vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)'$, provided that the constant c is chosen appropriately. The MGF of $\mathbf{X}$ exists and is given by

$$M(t_1, t_2, \ldots, t_n) = \exp\left(\mathbf{t}'\boldsymbol{\mu} + \frac{\mathbf{t}'\mathbf{M}^{-1}\mathbf{t}}{2}\right), \tag{17}$$

where $\mathbf{t} = (t_1, t_2, \ldots, t_n)'$ and $t_1, t_2, \ldots, t_n$ are arbitrary real numbers.

Proof. Let

$$I = c\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left[\mathbf{t}'\mathbf{x} - \frac{(\mathbf{x}-\boldsymbol{\mu})'\mathbf{M}(\mathbf{x}-\boldsymbol{\mu})}{2}\right]\prod_{i=1}^{n} dx_i. \tag{18}$$

Changing the variables of integration to $y_1, y_2, \ldots, y_n$ by writing $x_i - \mu_i = y_i$, $i = 1, 2, \ldots, n$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$, we have $\mathbf{x} - \boldsymbol{\mu} = \mathbf{y}$ and

$$I = c\exp(\mathbf{t}'\boldsymbol{\mu})\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left(\mathbf{t}'\mathbf{y} - \frac{\mathbf{y}'\mathbf{M}\mathbf{y}}{2}\right)\prod_{i=1}^{n} dy_i. \tag{19}$$

Since $\mathbf{M}$ is positive definite, it follows that all the n characteristic roots of $\mathbf{M}$, say $m_1, m_2, \ldots, m_n$, are positive. Moreover, since $\mathbf{M}$ is symmetric, there exists an $n \times n$ orthogonal matrix $\mathbf{L}$ such that $\mathbf{L}'\mathbf{M}\mathbf{L}$ is a diagonal matrix with diagonal elements $m_1, m_2, \ldots, m_n$. Let us change the variables to $z_1, z_2, \ldots, z_n$ by writing $\mathbf{y} = \mathbf{L}\mathbf{z}$, where $\mathbf{z}' = (z_1, z_2, \ldots, z_n)$, and note that the Jacobian of this orthogonal transformation is $|\mathbf{L}|$. Since $\mathbf{L}'\mathbf{L} = \mathbf{I}_n$, where $\mathbf{I}_n$ is an $n \times n$ unit matrix, $|\mathbf{L}| = 1$ and we have

$$(20) \qquad I = c\exp(\mathbf{t}'\boldsymbol{\mu}) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left(\mathbf{t}'\mathbf{L}\mathbf{z} - \frac{\mathbf{z}'\mathbf{L}'\mathbf{M}\mathbf{L}\mathbf{z}}{2}\right) \prod_{i=1}^{n} dz_i.$$

If we write $\mathbf{t}'\mathbf{L} = \mathbf{u}' = (u_1, u_2, \ldots, u_n)$, then $\mathbf{t}'\mathbf{L}\mathbf{z} = \sum_{i=1}^{n} u_i z_i$. Also, $\mathbf{L}'\mathbf{M}\mathbf{L} = \operatorname{diag}(m_1, m_2, \ldots, m_n)$, so that $\mathbf{z}'\mathbf{L}'\mathbf{M}\mathbf{L}\mathbf{z} = \sum_{i=1}^{n} m_i z_i^2$. The integral in (20) can therefore be written as

$$\prod_{i=1}^{n}\left[\int_{-\infty}^{\infty} \exp\left(u_i z_i - \frac{m_1 z_i^2}{2}\right) dz_i\right] = \prod_{i=1}^{n}\left[\sqrt{\frac{2\pi}{m_i}} \exp\left(\frac{u_i^2}{2m_i}\right)\right].$$

If follows that

$$(21) \qquad I = c\ \exp(\mathbf{t}'\mathbf{u}) \frac{(2\pi)^{n/2}}{(m_1 m_2 \cdots m_n)^{1/2}} \exp\left(\sum_{i=1}^{n} \frac{u_i^2}{2m_i}\right).$$

Setting $t_1 = t_2 = \cdots = t_n = 0$, we see from (18) and (21) that

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \cdots dx_n = \frac{c(2\pi)^{n/2}}{(m_1 m_2 \cdots m_n)^{1/2}}.$$

By choosing

$$(22) \qquad c = \frac{(m_1 m_2 \cdots m_n)^{1/2}}{(2\pi)^{n/2}}$$

we see that f is a joint PDF of some random vector $\mathbf{X}$, as asserted.

Finally, since

$$(\mathbf{L}'\mathbf{M}\mathbf{L})^{-1} = \operatorname{diag}(m_1^{-1}, m_2^{-1}, \ldots, m_n^{-1}),$$

we have

$$\sum_{i=1}^{n} \frac{u_i^2}{m_i} = \mathbf{u}'(\mathbf{L}'\mathbf{M}^{-1}\mathbf{L})\mathbf{u} = \mathbf{t}'\mathbf{M}^{-1}\mathbf{t}.$$

Also,

$$|\mathbf{M}^{-1}| = |\mathbf{L}'\mathbf{M}^{-1}\mathbf{L}| = (m_1 m_2 \cdots m_n)^{-1}.$$

It follows from (21) and (22) that the MGF of $\mathbf{X}$ is given by (17), and we may write

$$(23) \qquad c = \frac{1}{[(2\pi)^n |\mathbf{m}^{-1}|]^{1/2}}.$$

This completes the proof of Theorem 3.

Let us write $\mathbf{M}^{-1} = ((\sigma_{ij}))_{i,j=1,2,\dots,n}$. Then

$$M(0, 0, \dots, 0, t_i, 0, \dots, 0) = \exp\left(t_i\mu_i + \sigma_{ii}\frac{t_i^2}{2}\right)$$

is the MGF of X_i, $i = 1, 2, \dots, n$. Thus each X_i is $\mathcal{N}(\mu_i, \sigma_{ii})$, $i = 1, 2, \dots, n$. For $i \neq j$, we have for the MGF of X_i and X_j

$$\begin{aligned} &M(0, 0, \dots, 0, t_i, 0, \dots, 0, t_j, 0, \dots, 0) \\ &\quad = \exp\left(t_i\mu_i + t_j\mu_j + \frac{\sigma_{ii}t_i^2 + 2\sigma_{ij}t_it_j + t_j^2\sigma_{jj}}{2}\right). \end{aligned}$$

This is the MGF of a bivariate normal distribution with means μ_i, μ_j, variances σ_{ii}, σ_{jj}, and covariance σ_{ij}. Thus we see that

$$\boldsymbol{\mu}' = (\mu_1, \mu_2, \dots, \mu_n) \tag{24}$$

is the mean vector of $\mathbf{X}' = (X_1, \dots, X_n)$,

$$\sigma_{ii} = \sigma_i^2 = \operatorname{var}(X_i), \qquad i = 1, 2, \dots, n, \tag{25}$$

and

$$\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j, \qquad i \neq j; \quad i, j = 1, 2, \dots, n. \tag{26}$$

The matrix $\mathbf{M}^{-1}$ is called the *dispersion* (*variance–covariance*) *matrix* of the multivariate normal distribution.

If $\sigma_{ij} = 0$ for $i \neq j$, the matrix $\mathbf{M}^{-1}$ is a diagonal matrix, and it follows that the RVs $X_1, X_2, \dots, X_n$ are independent. Thus we have the following analog of Theorem 2.

Theorem 4. The components $X_1, X_2, \dots, X_n$ of a jointly normally distributed RV $\mathbf{X}$ are independent if and only if the covariances $\sigma_{ij} = 0$ for all $i \neq j$ $(i, j = 1, 2, \dots, n)$.

The following result is stated without proof. The proof is similar to the two-variate case except that now we consider the quadratic form in n variables: $E\{\sum_{i=1}^n t_i(X_i - \mu_i)\}^2 \geq 0$.

Theorem 5. The probability that the RVs $X_1, X_2, \dots, X_n$ with finite variances satisfy at least one linear relationship is 1 if and only if $|\mathbf{M}| = 0$.

Accordingly, if $|\mathbf{M}| = 0$, all the probability mass is concentrated on a hyperplane of dimension $< n$.

Theorem 6. Let $(X_1, X_2, \ldots, X_n)$ be an n-dimensional RV with a normal distribution. Let $Y_1, Y_2, \ldots, Y_k$, $k \le n$, be linear functions of X_j $(j = 1, 2, \ldots, n)$. Then $(Y_1, Y_2, \ldots, Y_k)$ also has a multivariate normal distribution.

Proof. Without loss of generality let us assume that $EX_i = 0$, $i = 1, 2, \ldots, n$. Let

$$(27)\qquad Y_p = \sum_{j=1}^{n} A_{pj} X_j, \qquad p = 1, 2, \ldots, k; \quad k \le n.$$

Then $EY_p = 0$, $p = 1, 2, \ldots, k$, and

$$(28)\qquad \operatorname{cov}(Y_p, Y_q) = \sum_{i,j=1}^{n} A_{pi} A_{qj} \sigma_{ij},$$

where $E(X_i X_j) = \sigma_{ij}$, $i, j = 1, 2, \ldots, n$.

The MGF of $(Y_1, Y_2, \ldots, Y_k)$ is given by

$$M^*(t_1, t_2, \ldots, t_k) = E\left[\exp\left(t_1 \sum_{j=1}^{n} A_{1j} X_j + \cdots + t_k \sum_{j=1}^{n} A_{kj} X_j\right)\right].$$

Writing $u_j = \sum_{p=1}^{k} t_p A_{pj}$, $j = 1, 2, \ldots, n$, we have

$$\begin{aligned}
(29)\qquad M^*(t_1, t_2, \ldots, t_k) &= E\left[\exp\left(\sum_{i=1}^{n} u_i X_i\right)\right] \\
&= \exp\left(\frac{1}{2} \sum_{i,j=1}^{n} \sigma_{ij} u_i u_j\right) \qquad \text{by (17)} \\
&= \exp\left(\frac{1}{2} \sum_{i,j=1}^{n} \sigma_{ij} \sum_{l,m=1}^{k} t_l t_m A_{li} A_{mj}\right) \\
&= \exp\left(\frac{1}{2} \sum_{l,m=1}^{k} t_l t_m \sum_{i,j=1}^{n} A_{li} A_{mj} \sigma_{ij}\right) \\
&= \exp\left[\frac{1}{2} \sum_{l,m=1}^{k} t_l t_m \operatorname{cov}(Y_l, Y_m)\right].
\end{aligned}$$

When (17) and (29) are compared, the result follows.

Corollary 1. Every marginal distribution of an n-dimensional normal distribution is univariate normal. Moreover, any linear function of $X_1, X_2, \ldots, X_n$ is univariate normal.

Corollary 2. If $X_1, X_2, \ldots, X_n$ are iid $\mathcal{N}(\mu, \sigma^2)$, and $\mathbf{A}$ is an $n \times n$ orthogonal transformation matrix, the components $Y_1, Y_2, \ldots, Y_n$ of $\mathbf{Y} = \mathbf{AX}'$, where $\mathbf{X} = (X_1, \ldots, X_n)'$, are independent RVs, each normally distributed with the same variance σ^2.

We have from (27) and (28)

$$\begin{aligned}\operatorname{cov}(Y_p, Y_q) &= \sum_{i=1}^{n} A_{pi} A_{qi} \sigma_{ii} + \sum_{i \neq j} A_{pi} A_{qj} \sigma_{ij} \\ &= \begin{cases} 0 & \text{if } p \neq q, \\ \sigma^2 & \text{if } p = q, \end{cases}\end{aligned}$$

since $\sum_{i=1}^{n} A_{pi} A_{qi} = 0$ and $\sum_{j=1}^{n} A_{pj}^2 = 1$. It follows that

$$M^*(t_1, t_2, \ldots, t_n) = \exp\left(\frac{1}{2} \sum_{l=1}^{n} t_l^2 \sigma^2\right).$$

and Corollary 2 follows.

Theorem 7. Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)'$. Then $\mathbf{X}$ has an n-dimensional normal distribution if and only if every linear function of $\mathbf{X}$,

$$\mathbf{X}'\mathbf{t} = t_1 X_1 + t_2 X_2 + \cdots + t_n X_n$$

has a univariate normal distribution.

Proof. Suppose that $\mathbf{X}'\mathbf{t}$ is normal for any $\mathbf{t}$. Then the MGF of $\mathbf{X}'\mathbf{t}$ is given by

$$M(s) = \exp\left(bs + \tfrac{1}{2}\sigma^2 s^2\right). \tag{30}$$

Here $b = E\{\mathbf{X}'\mathbf{t}\} = \sum_1^n t_i \mu_i = \mathbf{t}'\boldsymbol{\mu}$, where $\boldsymbol{\mu}' = (\mu_1, \ldots, \mu_n)$, and $\sigma^2 = \operatorname{var}(\mathbf{X}'\mathbf{t}) = \operatorname{var}(\sum t_i X_i) = \mathbf{t}'\mathbf{M}^{-1}\mathbf{t}$, where $\mathbf{M}^{-1}$ is the dispersion matrix of $\mathbf{X}$. Thus

$$M(s) = \exp\left(\mathbf{t}'\boldsymbol{\mu} s + \tfrac{1}{2}\mathbf{t}'\mathbf{M}^{-1}\mathbf{t} s^2\right). \tag{31}$$

Let $s = 1$; then

$$M(1) = \exp\left(\mathbf{t}'\boldsymbol{\mu} + \tfrac{1}{2}\mathbf{t}'\mathbf{M}^{-1}\mathbf{t}\right), \tag{32}$$

and since the MGF is unique, it follows that $\mathbf{X}$ has a multivariate normal distribution. The converse follows from Corollary 1 to Theorem 6.

Many characterization results for the multivariate normal distribution are now available. We refer the reader to Lukacs and Laha [67, p. 79].

PROBLEMS 5.4

1. Let (X, Y) have joint PDF

$$f(x, y) = \frac{1}{6\pi\sqrt{7}} \exp\left[-\frac{8}{7}\left(\frac{x^2}{16} - \frac{31}{32}x + \frac{xy}{8} + \frac{y^2}{9} - \frac{4}{3}y + \frac{71}{16}\right)\right],$$

for $-\infty < x < \infty$, $-\infty < y < \infty$.

(a) Find the means and variances of X and Y. Also find ρ.

(b) Find the conditional PDF of Y given $X = x$ and $E\{Y|x\}$, $\text{var}\{Y|x\}$.

(c) Find $P\{4 \le Y \le 6|X = 4\}$.

2. In Example 1, show that $\text{cov}(X, Y) = \alpha/\pi$.

3. Let (X, Y) be a bivariate normal RV with parameters μ_1, μ_2, σ_1^2, σ_2^2, and ρ. What is the distribution of $X + Y$? Compare your result with that of Example 1.

4. Let (X, Y) be a bivariate normal RV with parameters $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$, and ρ, and let $U = aX + b$, $a \neq 0$, and $V = cY + d$, $c \neq 0$. Find the joint distribution of (U, V).

5. Let (X, Y) be a bivariate normal RV with parameters $\mu_1 = 5$, $\mu_2 = 8$, $\sigma_1^2 = 16$, $\sigma_2^2 = 9$, and $\rho = 0.6$. Find $P\{5 < Y < 11 \mid X = 2\}$.

6. Let X and Y be jointly normal with means 0. Also, let

$$W = X\cos\theta + Y\sin\theta, \qquad Z = X\cos\theta - Y\sin\theta.$$

Find θ such that W and Z are independent.

7. Let (X, Y) be a normal RV with parameters μ_1, μ_2, σ_1^2, σ_2^2, and ρ. Find a necessary and sufficient condition for $X + Y$ and $X - Y$ to be independent.

8. For a bivariate normal RV with parameters $\mu_1, \mu_2, \sigma_1, \sigma_2$, and ρ show that

$$P(X > \mu_1,\ Y > \mu_2) = \frac{1}{4} + \frac{1}{2\pi}\tan^{-1}\frac{\rho}{\sqrt{1-\rho^2}}.$$

[*Hint*: The required probability is $P\big((X - \mu_1)/\sigma_1 > 0,\ (Y - \mu_2)/\sigma_2 > 0\big)$. Change to polar coordinates and integrate.]

9. Show that every variance–covariance matrix is symmetric positive semidefinite and conversely. If the variance–covariance matrix is not positive definite, then with probability 1 the random (column) vector $\mathbf{X}$ lies in some hyperplane $\mathbf{c}'\mathbf{X} = a$ with $\mathbf{c} \neq \mathbf{0}$.

10. Let (X, Y) be a bivariate normal RV with $EX = EY = 0$, $\text{var}(X) = \text{var}(Y) = 1$, and $\text{cov}(X, Y) = \rho$. Show that the RV $Z = Y/X$ has a Cauchy distribution.

11. (a) Show that

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{\sum x_i^2}{2}\right)\left[1 + \prod_1^n \left(x_i e^{-x_i^2/2}\right)\right]$$

is a joint PDF on $\mathcal{R}_n$.

(b) Let $(X_1, X_2, \ldots, X_n)$ have PDF f given in (a). Show that the RVs in any proper subset of $\{X_1, X_2, \ldots, X_n\}$ containing two or more elements are independent standard normal RVs.

5.5 EXPONENTIAL FAMILY OF DISTRIBUTIONS

Most of the distributions that we have so far encountered belong to a general family of distributions that we now study. Let Θ be an interval on the real line, and let $\{f_\theta : \theta \in \Theta\}$ be a family of PDFs (PMFs). Here and in what follows we write $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ unless otherwise specified.

Definition 1. If there exist real-valued functions $Q(\theta)$ and $D(\theta)$ on Θ and Borel-measurable functions $T(x_1, x_2, \ldots, x_n)$ and $S(x_1, x_2, \ldots, x_n)$ on $\mathcal{R}_n$ such that

$$(1) \qquad f_\theta(x_1, x_2, \ldots, x_n) = \exp[Q(\theta)T(\mathbf{x}) + D(\theta) + S(\mathbf{x})],$$

we say that the family $\{f_\theta, \theta \in \Theta\}$ is a *one-parameter exponential family*.

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m$ be iid with PMF (PDF) f_θ. Then the joint distribution of $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m)$ is given by

$$\begin{aligned} g_\theta(\mathbf{x}) &= \prod_{i=1}^m f_\theta(\mathbf{x}_i) = \prod_{i=1}^m \exp[Q(\theta)T(\mathbf{x}_i) + D(\theta) + S(\mathbf{x}_i)] \\ &= \exp\left[Q(\theta)\sum_{i=1}^m T(\mathbf{x}_i) + mD(\theta) + \sum_{i=1}^m S(\mathbf{x}_i)\right], \end{aligned}$$

where $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m)$, $\mathbf{x}_j = (x_{j1}, x_{j2}, \ldots, x_{jn})$, $j = 1, 2, \ldots, m$, and it follows that $\{g_\theta : \theta \in \Theta\}$ is again a one-parameter exponential family.

Example 1. Let $X \sim \mathcal{N}(\mu_0, \sigma^2)$, where μ_0 is known and σ^2 unknown. Then

$$\begin{aligned} f_{\sigma^2}(x) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu_0)^2}{2\sigma^2}\right] \\ &= \exp\left[-\log(\sigma\sqrt{2\pi}) - \frac{(x-\mu_0)^2}{2\sigma^2}\right] \end{aligned}$$

is a one-parameter exponential family with

$$Q(\sigma^2) = -\frac{1}{2\sigma^2}, \qquad T(x) = (x-\mu_0)^2, \qquad S(x) = 0, \quad \text{and}$$
$$D(\sigma^2) = -\log(\sigma\sqrt{2\pi}).$$

If $X \sim \mathcal{N}(\mu, \sigma_0^2)$, where σ_0 is known but μ is unknown, then

$$\begin{aligned} f_\mu(x) &= \frac{1}{\sigma_0\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma_0^2}\right] \\ &= \frac{1}{\sigma_0\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma_0^2} + \frac{\mu x}{\sigma_0^2} - \frac{\mu^2}{2\sigma_0^2}\right) \end{aligned}$$

is a one-parameter exponential family with

$$Q(\mu) = \frac{\mu}{\sigma_0^2}, \qquad D(\mu) = -\frac{\mu}{2\sigma_0^2}, \qquad T(x) = x,$$

and

$$S(x) = -\left[\frac{x^2}{2\sigma_0^2} + \frac{1}{2}\log(2\pi\sigma_0^2)\right].$$

Example 2. Let $X \sim P(\lambda)$, $\lambda > 0$ unknown. Then

$$P_\lambda\{X = x\} = e^{-\lambda}\frac{\lambda^x}{x!} = \exp[-\lambda + x\log\lambda - \log(x!)],$$

and we see that the family of Poisson PMFs with parameter λ is a one-parameter exponential family.

Some other important examples of one-parameter exponential families are binomial, $G(\alpha, \beta)$ (provided that one of α, β is fixed), $B(\alpha, \beta)$ (provided that one of α, β is fixed), negative binomial, and geometric. The Cauchy family of densities and the uniform distribution on $[0, \theta]$ do not belong to this class.

Theorem 1. Let $\{f_\theta : \theta \in \Theta\}$ be a one-parameter exponential family of PDFs (PMFs) given in (1). Then the family of distributions of $T(\mathbf{X})$ is also a one-parameter exponential family of PDFs (PMFs), given by

$$g_\theta(t) = \exp[tQ(\theta) + D(\theta) + S^*(t)]$$

for suitable $S^*(t)$.

The proof of Theorem 1 is a simple application of the transformation of variables technique studied in Section 4.4 and is left as an exercise, at least for the cases considered in Section 4.4. For the general case we refer to Lehmann [63, p. 58].

Let us now consider the k-parameter exponential family, $k \geq 2$. Let $\Theta \subseteq \mathcal{R}_k$ be a k-dimensional interval.

Definition 2. If there exist real-valued functions $Q_1, Q_2, \ldots, Q_k, D$ defined on Θ, and Borel-measurable functions $T_1, T_2, \ldots, T_k, S$ on $\mathcal{R}_n$ such that

$$f_{\boldsymbol{\theta}}(\mathbf{x}) = \exp\left[\sum_{i=1}^{k} Q_i(\boldsymbol{\theta}) T_i(\mathbf{x}) + D(\boldsymbol{\theta}) + S(\mathbf{x})\right], \tag{2}$$

we say that the family $\{f_{\boldsymbol{\theta}},\ \boldsymbol{\theta} \in \Theta\}$ is a *k-parameter exponential family*.

Once again, if $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m)$ and $\mathbf{X}_j$ are iid with common distribution (2), the joint distributions of $\mathbf{X}$ form a k-parameter exponential family. An analog of Theorem 1 also holds for the k-parameter exponential family.

Example 3. The most important example of a k-parameter exponential family is $\mathcal{N}(\mu, \sigma^2)$ when both μ and σ^2 are unknown. We have

$$\boldsymbol{\theta} = (\mu, \sigma^2), \qquad \Theta = \{(\mu, \sigma^2) : -\infty < \mu < \infty, \sigma^2 > 0\}$$

and

$$\begin{aligned} f_{\boldsymbol{\theta}}(x) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2 - 2\mu x + \mu^2}{2\sigma^2}\right) \\ &= \exp\left\{-\frac{x^2}{2\sigma^2} + \frac{\mu}{\sigma^2}x - \frac{1}{2}\left[\frac{\mu^2}{\sigma^2} + \log(2\pi\sigma^2)\right]\right\}. \end{aligned}$$

It follows that $f_{\boldsymbol{\theta}}$ is a two-parameter exponential family with

$$Q_1(\boldsymbol{\theta}) = -\frac{1}{2\sigma^2}, \qquad Q_2(\boldsymbol{\theta}) = \frac{\mu}{\sigma^2}, \qquad T_1(x) = x^2, \qquad T_2(x) = x,$$

$$D(\boldsymbol{\theta}) = -\frac{1}{2}\left[\frac{\mu^2}{\sigma^2} + \log(2\pi\sigma^2)\right], \quad \text{and} \quad S(x) = 0.$$

Other examples are the $G(\alpha, \beta)$ and $B(\alpha, \beta)$ distributions when both α, β are unknown, and the multinomial distribution. $U[\alpha, \beta]$ does not belong to this family, nor does $\mathcal{C}(\alpha, \beta)$.

Some general properties of exponential families will be studied in Chapter 8, and the importance of these families will then become evident.

Remark 1. The form in (2) is not unique, as easily seen by substituting αQ_i for Q_i and $(1/\alpha)T_i$ for T_i. This, however, is not going to be a problem in statistical considerations.

Remark 2. The integer k in Definition 2 is also not unique since the family $\{1, Q_1, \dots, Q_k\}$ or $\{1, T_1, \dots, T_k\}$ may be linearly dependent. In general, k need not be the dimension of Θ.

Remark 3. The support $\{\mathbf{x} : f_{\boldsymbol{\theta}}(\mathbf{x}) > 0\}$ does not depend on $\boldsymbol{\theta}$.

Remark 4. In (2), one can change parameters to $\eta_i = Q_i(\boldsymbol{\theta})$, $i = 1, 2, \dots, k$, so that

$$f_{\boldsymbol{\eta}}(\mathbf{x}) = \exp\left[\sum_{i=1}^{k} \eta_i T_i(\mathbf{x}) + D(\boldsymbol{\eta}) + S(\mathbf{x})\right] \tag{3}$$

where the parameters $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_k)$ are called *natural parameters*. Again, η_i may be linearly dependent so that one of η_i may be eliminated.

PROBLEMS 5.5

1. Show that the following families of distributions are one-parameter exponential families:
   (a) $X \sim b(n, p)$.
   (b) $X \sim G(\alpha, \beta)$, (i) if α is known, and (ii) if β is known.
   (c) $X \sim B(\alpha, \beta)$, (i) if α is known, and (ii) if β is known.
   (d) $X \sim NB(r; p)$, where r is known, p unknown.
2. Let $X \sim \mathcal{C}(1, \theta)$. Show that the family of distributions of X is not a one-parameter exponential family.
3. Let $X \sim U[0, \theta]$, $\theta \in [0, \infty)$. Show that the family of distributions of X is not an exponential family.
4. Is the family of PDFs

   $$f_\theta(x) = \tfrac{1}{2}e^{-|x-\theta|}, \qquad -\infty < x < \infty, \theta \in (-\infty, \infty),$$

   an exponential family?

5. Show that the following families of distributions are two-parameter exponential families:
   (a) $X \sim G(\alpha, \beta)$, both α and β unknown.
   (b) $X \sim B(\alpha, \beta)$, both α and β unknown.
6. Show that the families of distributions $U[\alpha, \beta]$ and $C(\alpha, \beta)$ do not belong to the exponential families.
7. Show that the multinomial distributions form an exponential family.