A PRIMER ON STATISTICAL DISTRIBUTIONS

N. BALAKRISHNAN
McMaster University
Hamilton, Canada

V. B. NEVZOROV
St. Petersburg State University
Russia

WILEY-INTERSCIENCE

A JOHN WILEY & SONS, INC., PUBLICATION

A PRIMER ON STATISTICAL DISTRIBUTIONS

This Page Intentionally Left Blank

A PRIMER ON STATISTICAL DISTRIBUTIONS

N. BALAKRISHNAN
McMaster University
Hamilton, Canada

V. B. NEVZOROV
St. Petersburg State University
Russia

WILEY-INTERSCIENCE

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data:

Balakrishnan, N., 1956–
A primer on statistical distributions / N. Balakrishnan and V.B. Nevzorov.
p. cm.
Includes bibliographical references and index.
ISBN 0-471-42798-5 (acid-free paper)
1. Distribution (Probability theory) I. Nevzorov, Valery B., 1946– II. Title.

QA273.B25473 2003
519.2'4—dc21 2003041157

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

To my lovely daughters, Sarah and Julia
(N.B.)

To my wife, Ludmila
(V.B.N.)

This Page Intentionally Left Blank

CONTENTS

This Page Intentionally Left Blank

PREFACE

Distributions and their properties and interrelationships assume a very important role in most upper-level undergraduate as well as graduate courses in the statistics program. For this reason, many introductory statistics textbooks discuss in a chapter or two a few basic statistical distributions, such as binomial, Poisson, exponential, and normal. Yet a good knowledge of some other distributions, such as geometric, negative binomial, Pareto, beta, gamma, chi-square, logistic, Laplace, extreme value, multinomial, multivariate normal, and Dirichlet will be immensely useful to those students who go on to upper-level undergraduate or graduate courses in statistics. Students in applied programs such as psychology, sociology, biology, geography, geology, economics, business, and engineering will also benefit significantly from an exposure to different distributions and their properties as statistical modelling of observed data is an integral part of these disciplines.

It is for this reason we have prepared this textbook, which is tailor-made for a one-term course (of about 35 lectures) on *statistical distributions.* All the preliminary concepts and definitions are presented in Chapter 1. The rest of the material is divided into three parts, with Part I covering discrete distributions, Part II covering continuous distributions, and Part III covering multivariate distributions. In each chapter we have included a few pertinent exercises (at an appropriate level for students taking the course) which may be handed out as homework at the end of each chapter. A biographical sketch of some of the leading contributors to the area of *statistical distribution theory* is presented in the Appendix to present students with a historical sense of developments in this important and fundamental area in the field of statistics.

From our experience, we would suggest the following lecture allocation for teaching a course on *statistical distributions* based on this book:

5	**lectures**	on	*preliminaries*	(Chapter 1)
9	**lectures**	on	*discrete distributions*	(Part I)
17	**lectures**	on	*continuous distributions*	(Part II)
4	**lectures**	on	*multivariate distributions*	(Part III)

We welcome comments and criticisms from all those who teach a course based on this book. Any suggestions for improvement or "necessary" addition (omission of which in this version should be regarded as a consequence of our ignorance, not of personal nonscientific antipathy) sent to us will be much appreciated and will be acted upon when the opportunity arises.

It is important to mention here that many authoritative and encyclopedic volumes on statistical distribution theory exist in the literature. For example:

- Johnson, Kotz, and Kemp (1992), describing discrete univariate distributions
- Stuart and Ord (1993), discussing general distribution theory
- Johnson, Kotz, and Balakrishnan (1994, 1995), describing continuous univariate distributions
- Johnson, Kotz, and Balakrishnan (1997), describing discrete multivariate distributions
- Wimmer and Altmann (1999), providing a thesaurus on discrete univariate distributions
- Evans, Peacock, and Hastings (2000), describing discrete and continuous distributions
- Kotz, Balakrishnan, and Johnson (2000), discussing continuous multivariate distributions

are some of the prominent ones. In addition, there are separate books dedicated to some specific distributions, such as Poisson, generalized Poisson, chi-square, Pareto, exponential, lognormal, logistic, normal, and Laplace (which have all been referred to in this book at appropriate places). These books may be consulted for any additional information.

We take this opportunity to express our sincere thanks to Mr. Steve Quigley (of John Wiley & Sons, New York) for his support and encouragement during the preparation of this book. Our special thanks go to Mrs. Debbie Iscoe (Mississauga, Ontario, Canada) for assisting us with the camera-ready production of the manuscript, and to Mr. Weiquan Liu for preparing all the figures. We also acknowledge with gratitude the financial support provided by the Natural Sciences and Engineering Research Council of Canada and the Russian Foundation of Basic Research (Grants 01-01-00031 and 00-15-96019) during the course of this project.

N. BALAKRISHNAN
Hamilton, Canada

V. B. NEVZOROV
St. Petersburg, Russia

April 2003

CHAPTER 1

PRELIMINARIES

In this chapter we present some basic notations, notions, and definitions which a reader of this book must absolutely know in order to follow subsequent chapters.

1.1 Random Variables and Distributions

Let $(\Omega, \mathcal{T}, P)$ be a probability space, where $\Omega = \{\omega\}$ is a set of elementary events, $\mathcal{T}$ is a σ-algebra of events, and P is a probability measure defined on $(\Omega, \mathcal{T})$. Further, let B denote an element of the Borel σ-algebra of subsets of the real line $\mathcal{R}$.

Definition 1.1 A finite single-valued function $X = X(\omega)$ which maps Ω into $\mathcal{R}$ is called a *random variable* if for any Borel set B in $\mathcal{R}$, the inverse image of B, i.e.,

$$X^{-1}(B) = \{\omega : X(\omega) \in B\}$$

belongs to the σ-algebra $\mathcal{T}$.

It means that for all Borel sets B, one can define probabilities

$$P\{X \in B\} = P\{X^{-1}(B)\}.$$

In particular, if for any x $(-\infty < x < \infty)$ we take $B = (-\infty, x]$, then the function

$$F(x) = P\{X \leq x\} \tag{1.1}$$

is defined for the random variable X.

Definition 1.2 The function $F(x)$ is called the *distribution function* or *cumulative distribution function* (cdf) of the random variable X.

Remark 1.1 Quite often, the cumulative distribution function of a random variable X is defined as

$$G(x) = P\{X < x\}.$$

Most of the properties of both these versions of cdf (i.e., F and G) coincide. Only one important difference exists between functions $F(x)$ and $G(x)$: F is right continuous, while G is left continuous. In our treatment we use the cdf as given in Definition 1.2.

There are three types of distributions: absolutely continuous, discrete and singular, and any cdf $F(x)$ can be represented as a mixture

$$F(x) = p_1F_1(x) + p_2F_2(x) + p_3F_3(x) \tag{1.2}$$

of absolutely continuous F_1, discrete F_2, and singular F_3 cdf's, with nonnegative weights p_1, p_2, and p_3 such that $p_1 + p_2 + p_3 = 1$. In this book we restrict ourselves to distributions which are either purely absolutely continuous or purely discrete.

Definition 1.3 A random variable X is said to have a *discrete distribution* if there exists a countable set $B = \{x_1, x_2, \ldots\}$ such that

$$P\{X \in B\} = 1.$$

Remark 1.2 To determine a random variable having a discrete distribution, one must fix two sequences: a sequence of values $x_1, x_2, \ldots$ and a sequence of probabilities $p_k = P\{X = x_k\}$, $k = 1, 2, \ldots$, such that

$$\sum_k p_k = 1.$$

In this case, the cdf of X is given by

$$F(x) = P\{X \leq x\} = \sum_{k:\ x_k \leq x} p_k. \tag{1.3}$$

Definition 1.4 A random variable X with a cdf F is said to have an *absolutely continuous distribution* if there exists a nonnegative function $p(x)$ such that

$$F(x) = \int_{-\infty}^{x} p(t)\ dt \tag{1.4}$$

for any real x.

Remark 1.3 The function $p(x)$ then satisfies the condition

$$\int_{-\infty}^{\infty} p(t)\ dt = 1, \tag{1.5}$$

and it is called the *probability density function* (pdf) of X. Note that any nonnegative function $p(x)$ satisfying (1.5) can be the pdf of some random variable X.

Remark 1.4 If a random variable X has an absolutely continuous distribution, then its cdf $F(x)$ is continuous.

Definition 1.5 We say that random variables X and Y have the same distribution, and write

$$X \stackrel{d}{=} Y \tag{1.6}$$

if the cdf's of X and Y (i.e., F_X and F_Y) coincide; that is,

$$F_X(x) = P\{X \leq x\} = P\{Y \leq x\} = F_Y(x) \ \forall \ x.$$

Exercise 1.1 Construct an example of a probability space $(\Omega, \mathcal{T}, P)$ and a finite single-valued function $X = X(\omega), \omega \in \Omega$, which maps Ω into $\mathcal{R}$, that is not a random variable.

Exercise 1.2 Let $p(x)$ and $q(x)$ be probability density functions of two random variables. Consider now the following functions:

$$(a)\ 2p(x) - q(x);\ \ (b)\ p(x) + 2q(x);\ \ (c)\ |p(x) - q(x)|;\ \ (d)\ \frac{1}{2}\left(p(x) + q(x)\right).$$

Which of these functions are probability density functions of some random variable for any choice of $p(x)$ and $q(x)$? Which of them can br valid probability density functions under suitably chosen $p(x)$ and $q(x)$? Is there a function that can never be a probability density function of a random variable?

Exercise 1.3 Suppose that $p(x)$ and $q(x)$ are probability density functions of X and Y, respectively, satisfying

$$p(x) = 2 - q(x) \quad \text{for} \quad 0 < x < 1.$$

Then, find $P\{X < -1\} + P\{Y < 2\}$.

The *quantile function* of a random variable X with cdf $F(x)$ is defined by

$$Q(u) = \inf\{x : F(x) \geq u\}, \quad 0 < u < 1.$$

In the case when X has an absolutely continuous distribution, then the quantile function $Q(u)$ may simply be written as

$$Q(u) = F^{-1}(u), \quad 0 < u < 1.$$

The corresponding *quantile density function* is given by

$$q(u) = \frac{dQ(u)}{du} = \frac{1}{p(Q(u))}, \quad 0 < u < 1,$$

where $p(x)$ is the pdf corresponding to the cdf $F(x)$.

It should be noted that just as forms of $F(x)$ may be used to propose families of distributions, general forms of the quantile function $Q(u)$ may also be used to propose families of statistical distributions. Interested readers may refer to the recent book by Gilchrist (2000) for a detailed discussion on statistical modelling with quantile functions.

1.2 Type of Distribution

Definition 1.6 Random variables X and Y are said to *belong to the same type of distribution* if there exist constants a and $h > 0$ such that

$$Y \stackrel{d}{=} a + hX. \tag{1.7}$$

Note then that the cdf's F_X and F_Y of the random variables X and Y satisfy the relation

$$F_Y(x) = F_X\left(\frac{x-a}{h}\right) \ \forall\ x. \tag{1.8}$$

One can, therefore, choose a certain cdf F as the standard distribution function of a certain distribution family. Then this family would consist of all cdf's of the form

$$F(x, a, h) = F\left(\frac{x-a}{h}\right), \qquad -\infty < x < \infty,\ h > 0, \tag{1.9}$$

and

$$F(x) = F(x, 0, 1).$$

Thus, we have a two-parameter family of cdf's $F(x, a, h)$, where a is called the *location parameter* and h is the *scale parameter*.

For absolutely continuous distributions, one can introduce the corresponding two-parameter families of probability density functions:

$$p(x, a, h) = \frac{1}{h}\, p\left(\frac{x-a}{h}\right), \tag{1.10}$$

where $p(x) = p(x, 0, 1)$ corresponds to the random variable X with cdf F, and $p(x, a, h)$ corresponds to the random variable $Y = a + hX$ with cdf $F(x, a, h)$.

1.3 Moment Characteristics

There are some classical numerical characteristics of random variables and their distributions. The most popular ones are expected values and variances. More general characteristics are the *moments*. Among them, we emphasize moments about zero (about origin) and central moments.

Definition 1.7 For a discrete random variable X taking on values $x_1, x_2, \ldots$ with probabilities

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

we define the nth *moment of X about zero* as

$$\alpha_n = EX^n = \sum_k x_k^n p_k. \tag{1.11}$$

We say that α_n exists if

$$\sum_k |x_k|^n p_k < \infty.$$

Note that the expected value EX is nothing but α_1. EX is also called the *mean of* X or the *mathematical expectation of* X.

Definition 1.8 The nth *central moment* of X is defined as

$$\beta_n = E(X - EX)^n = \sum_k (x_k - EX)^n p_k, \tag{1.12}$$

given that

$$\sum_k |x_k - EX|^n p_k < \infty.$$

If a random variable X has an absolutely continuous distribution with a pdf $p(x)$, then the moments about zero and the central moments have the following expressions:

$$\alpha_n = EX^n = \int_{-\infty}^{\infty} x^n p(x)\, dx \tag{1.13}$$

and

$$\beta_n = E(X - EX)^n = \int_{-\infty}^{\infty} (x - EX)^n p(x)\, dx. \tag{1.14}$$

We say that moments (1.13) exist if

$$\int_{-\infty}^{\infty} |x|^n p(x)\, dx < \infty. \tag{1.15}$$

The *variance of* X is simply the second central moment:

$$\text{Var } X = \beta_2 = E(X - EX)^2. \tag{1.16}$$

Central moments are easily expressed in terms of moments about zero as follows:

$$\begin{aligned} \beta_n &= E(X - EX)^n = \sum_{k=0}^{n} (-1)^k \binom{n}{k} (EX)^k EX^{n-k} \\ &= \sum_{k=0}^{n} (-1)^k \binom{n}{k} \alpha_1^k \alpha_{n-k}. \end{aligned} \tag{1.17}$$

In particular, we have

$$\text{Var } X = \beta_2 = \alpha_2 - \alpha_1^2 \tag{1.18}$$

and

$$\beta_3 = \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^3 \quad \text{and} \quad \beta_4 = \alpha_4 - 4\alpha_1\alpha_3 + 6\alpha_1^2\alpha_2 - 3\alpha_1^4. \tag{1.19}$$

Note that the first central moment $\beta_1 = 0$.

The inverse problem cannot be solved, however, because all central moments save no information about EX; hence, the expected value cannot be expressed in terms of β_n ($n = 1, 2, \ldots$). Nevertheless, the relation

$$\begin{aligned} \alpha_n &= EX^n = E[(X - EX) + EX]^n \\ &= \sum_{k=0}^{n} \binom{n}{k} (EX)^k E(X - EX)^{n-k} \\ &= \sum_{k=0}^{n} \binom{n}{k} \alpha_1^k \beta_{n-k} \end{aligned} \tag{1.20}$$

will enable us to express α_n ($n = 2, 3, \ldots$) in terms of $\alpha_1 = EX$ and the central moments $\beta_2, \ldots, \beta_n$. In particular, we have

$$\alpha_2 = \beta_2 + \alpha_1^2, \tag{1.21}$$

$$\alpha_3 = \beta_3 + 3\beta_2\alpha_1 + \alpha_1^3 \quad \text{and} \quad \alpha_4 = \beta_4 + 4\beta_3\alpha_1 + 6\beta_2\alpha_1^2 + \alpha_1^4. \tag{1.22}$$

Let X and Y belong to the same type of distribution [see (1.7)], meaning that

$$Y \stackrel{d}{=} a + hX$$

for some constants a and $h > 0$. Then, the following equalities allow us to express moments of Y in terms of the corresponding moments of X:

$$EY^n = E(a + hX)^n = \sum_{k=0}^{n} \binom{n}{k} a^k h^{n-k} EX^{n-k} \tag{1.23}$$

and

$$E(Y - EY)^n = E[h(X - EX)]^n = h^n E(X - EX)^n. \tag{1.24}$$

Note that the central moments of Y do not depend on the location parameter a. As particular cases of (1.23) and (1.24), we have

$$EY = a + hEX, \tag{1.25}$$

$$EY^2 = a^2 + 2ahEX + h^2EX^2, \quad \text{Var } Y = h^2 \text{ Var } X, \tag{1.26}$$

$$EY^3 = a^3 + 3a^2hEX + 3ah^2EX^2 + h^3EX^3, \tag{1.27}$$

$$EY^4 = a^4 + 4a^3hEX + 6a^2h^2EX^2 + 4ah^3EX^3 + h^4EX^4. \tag{1.28}$$

Definition 1.9 For random variables taking on values $0, 1, 2, \ldots$, the *factorial moments of positive order* are defined as

$$\mu_r = EX(X - 1) \cdots (X - r + 1), \qquad r = 1, 2, \ldots, \tag{1.29}$$

while the *factorial moments of negative order* are defined as

$$\mu_{-r} = E\left[\frac{1}{(X + 1)(X + 2) \cdots (X + r)}\right], \qquad r = 1, 2, \ldots. \tag{1.30}$$

While dealing with discrete distributions, it is quite often convenient to work with these factorial moments rather than regular moments. For this reason, it is useful to note the following relationships between the factorial moments and the moments:

$$\mu_1 = \alpha_1, \tag{1.31}$$

$$\mu_2 = \alpha_2 - \alpha_1, \tag{1.32}$$

$$\mu_3 = \alpha_3 - 3\alpha_2 + 2\alpha_1, \tag{1.33}$$

$$\mu_4 = \alpha_4 - 6\alpha_3 + 11\alpha_2 - 6\alpha_1, \tag{1.34}$$

$$\alpha_2 = \mu_2 + \mu_1, \tag{1.35}$$

$$\alpha_3 = \mu_3 + 3\mu_2 + \mu_1, \tag{1.36}$$

$$\alpha_4 = \mu_4 + 6\mu_3 + 7\mu_2 + \mu_1. \tag{1.37}$$

Exercise 1.4 Present two different random variables having the same expectations and the same variances.

Exercise 1.5 Let X be a random variable with expectation EX and variance $\text{Var } X$. What is the sign of $r(X) = E(X - |X|)(\text{Var } X - \text{Var } |X|)$? When does the quantity $r(X)$ equal 0?

Exercise 1.6 Suppose that X is a random variable such that $P\{X > 0\} = 1$ and that both EX and $E(1/X)$ exist. Then, show that $EX + E(1/X) \geq 2$.

Exercise 1.7 Suppose that $P\{0 \leq X \leq 1\} = 1$. Then, prove that $EX^2 \leq EX \leq EX^2 + \frac{1}{4}$. Also, find all distributions for which the left and right bounds are attained.

Exercise 1.8 Construct a variable X for which $EX^3 = -5$ and $EX^6 = 24$.

1.4 Shape Characteristics

For any distribution, we are often interested in some characteristics that are associated with the shape of the distribution. For example, we may be interested in finding out whether it is unimodal, or skewed, and so on. Two important measures in this respect are Pearson's measures of skewness and kurtosis.

Definition 1.10 Pearson's measures of skewness and kurtosis are given by

$$\gamma_1 = \frac{\beta_3}{\beta_2^{3/2}}$$

and

$$\gamma_2 = \frac{\beta_4}{\beta_2^2}.$$

Since these measures are functions of central moments, it is clear that they are free of the location. Similarly, due to the fractional form of the measures, it can readily be verified that they are free of scale as well. It can also be seen that the measure of skewness γ_1 may take on positive or negative values depending on whether β_3 is positive or negative, respectively. Obviously, when the distribution is symmetric about its mean, we may note that β_3 is 0, in which case the measure of skewness γ_1 is also 0. Hence, distributions with $\gamma_1 > 0$ are said to be *positively skewed distributions*, while those with $\gamma_1 < 0$ are said to be *negatively skewed distributions.*

Now, without loss of generality, let us consider an arbitrary distribution with mean 0 and variance 1. Then, by writing

$$\left[\int x^3\, p(x)\, dx\right]^2 = \left[\int \left\{x\sqrt{p(x)}\right\} \left\{(x^2-1)\sqrt{p(x)}\right\} dx\right]^2$$

and applying the Cauchy–Schwarz inequality, we readily obtain the inequality

$$\gamma_2 \geq \gamma_1^2 + 1.$$

Later, we will observe the coefficient of kurtosis of a normal distribution to be 3. Based on this value, distributions with $\gamma_2 > 3$ are called *leptokurtic distributions*, while those with $\gamma_2 < 3$ are called *platykurtic distributions.* Incidentally, distributions for which $\gamma_2 = 3$ (which clearly includes the normal) are called *mesokurtic distributions.*

Remark 1.5 Karl Pearson (1895) designed a system of continuous distributions wherein the pdf of every member satisfies a differential equation. By studying their moment properties and, in particular, their coefficients of skewness and kurtosis, he proposed seven families of distributions which all occupied different regions of the (γ_1, γ_2)-plane. Several prominent distributions (such as beta, gamma, normal, and t that we will see in subsequent chapters) belong to these families. This development was the first and historic attempt to propose a unified mechanism for developing different families of statistical distributions.

1.5 Entropy

One more useful characteristic of distributions (called *entropy*) was introduced by Shannon.

Definition 1.11 For a discrete random variable X taking on values $x_1, x_2, \ldots$ with probabilities $p_1, p_2, \ldots$, the *entropy* $H(X)$ is defined as

$$H(X) = -\sum_n p_n \log p_n. \tag{1.38}$$

If X has an absolutely continuous distribution with pdf $p(x)$, then the entropy is defined as

$$H(X) = -\int_D p(x) \log p(x)\; dx, \tag{1.39}$$

where

$$D = \{x : p(x) > 0\}.$$

In the case of discrete distributions, the transformation

$$Y = a + hX, \qquad -\infty < a < \infty,\; h > 0$$

does not change the probabilities p_n and, consequently, we have

$$H(Y) = H(X).$$

On the other hand, if X has a pdf $p(x)$, then $Y = a + hX$ has the pdf

$$g(x) = \frac{1}{h}\, p\left(\frac{x-a}{h}\right)$$

and

$$H(Y) = -\int_{D_1} g(x) \log g(x)\; dx,$$

where

$$D_1 = \{x : g(x) > 0\} = \left\{x : p\left(\frac{x-a}{h}\right) > 0\right\} = \left\{x : \frac{x-a}{h} \in D\right\}.$$

It is then easy to verify that

$$\begin{aligned} H(Y) &= -\int_{D_1} \frac{1}{h}\, p\left(\frac{x-a}{h}\right) \log\left\{\frac{1}{h}\, p\left(\frac{x-a}{h}\right)\right\} dx \\ &= -\int_D p(x) \log\left\{\frac{1}{h}\, p(x)\right\} dx \\ &= \log h \int_D p(x)\; dx - \int_D p(x) \log p(x)\; dx \\ &= \log h + H(X). \end{aligned} \tag{1.40}$$

1.6 Generating Function and Characteristic Function

In this section we present some functions that are useful in generating the probabilities or the moments of the distribution in a simple and unified manner. In addition, they may also help in identifying the distribution of an underlying random variable of interest.

Definition 1.12 Let X take on values $0, 1, 2, \ldots$ with probabilities $p_n = P\{X = n\}$, $n = 0, 1, \ldots$. All the information about this distribution is contained in the *generating function*, which is defined as

$$P(s) = Es^X = \sum_{n=0}^{\infty} p_n s^n, \tag{1.41}$$

with the right-hand side (RHS) of (1.41) converging at least for $|s| \leq 1$.

Some important properties of generating functions are as follows:

(a) $P(1) = 1$;

(b) for $|s| < 1$, there exist derivatives of $P(s)$ of any order;

(c) for $0 \leq s < 1$, $P(s)$ and all its derivatives $P^{(k)}(s), k = 1, 2, \ldots$, are nonnegative increasing convex functions;

(d) the generating function $P(s)$ uniquely determines probabilities p_n, $n = 1, 2, \ldots$, and the following relations are valid:

$$\begin{aligned} p_0 &= P(0), \\ p_n &= \frac{P^{(n)}(0)}{n!}, \qquad n = 1, 2, \ldots; \end{aligned}$$

(e) if random variables $X_1, \ldots, X_n$ are independent and have generating functions

$$P_k(s) = Es^{X_k}, \qquad k = 1, \ldots, n,$$

then the generating function of the sum $Y = X_1 + \cdots + X_n$ satisfies the relation

$$P_Y(s) = \prod_{k=1}^{n} P_k(s); \tag{1.42}$$

(f) the factorial moments can be determined from the generating function as

$$\mu_k = EX(X-1)\cdots(X-k+1) = P^{(k)}(1), \tag{1.43}$$

where

$$P^{(k)}(1) = \lim_{s \uparrow 1} P^{(k)}(s).$$

Definition 1.13 The *characteristic function* $f(t)$ of a random variable X is defined as

$$f(t) = E\exp\{itX\} = E\cos tX + i\ E\sin tX. \tag{1.44}$$

If X takes on values x_k $(k = 1, 2, \ldots)$ with probabilities $p_k = P\{X = x_k\}$, then

$$\begin{aligned} f(t) &= \sum_k \exp(itx_k)\ p_k \\ &= \sum_k \cos(tx_k)\ p_k + i \sum_k \sin(tx_k)\ p_k. \end{aligned} \tag{1.45}$$

For a random variable having a pdf $p(x)$, the characteristic function takes on an analogous form:

$$\begin{aligned} f(t) &= \int_{-\infty}^{\infty} e^{itx} p(x)\ dx \\ &= \int_{-\infty}^{\infty} \cos(tx)\ p(x)\ dx + i \int_{-\infty}^{\infty} \sin(tx)\ p(x)\ dx. \end{aligned} \tag{1.46}$$

For random variables taking on values $0, 1, 2, \ldots$, there exists the following relationship between the characteristic function and the generating function:

$$f(t) = P(e^{it}). \tag{1.47}$$

Some of the useful properties of characteristic functions are as follows:

(a) $f(0) = 1$;

(b) $|f(t)| \leq 1$;

(c) $f(t)$ is uniformly continuous;

(d) $f(t)$ uniquely determines the distribution of the corresponding random variable X;

(e) if X has the characteristic function f, then $Y = a + hX$ has the characteristic function

$$g(t) = e^{iat}\ f(ht);$$

(f) if random variables $X_1, \ldots, X_n$ are independent and their characteristic functions are $f_1(t), \ldots, f_n(t)$, respectively, then the characteristic function of the sum $Y = X_1 + \cdots + X_n$ is given by

$$f_Y(t) = \prod_{k=1}^{n} f_k(t); \tag{1.48}$$

(g) if the nth moment EX^n of the random variable X exists, then the characteristic function $f(t)$ of X has the first n derivatives, and

$$\alpha_k = EX^k = \frac{f^{(k)}(0)}{i^k}, \qquad k = 1, 2, \ldots, n; \tag{1.49}$$

moreover, in this situation, the following expansion is valid for the characteristic function:

$$\begin{aligned} f(t) &= 1 + \sum_{k=1}^{n} f^{(k)}(0)t^k + r_n(t) \\ &= 1 + \sum_{k=1}^{n} \alpha_k (it)^k + r_n(t), \end{aligned} \tag{1.50}$$

where

$$r_n(t) = o(t^n)$$

as $t \to 0$;

(h) let random variables $X, X_1, X_2, \ldots$ have cdf's $F, F_1, F_2, \ldots$ and characteristic functions $f, f_1, f_2, \ldots$, respectively. If for any fixed t, as $n \to \infty$,

$$f_n(t) \to f(t), \tag{1.51}$$

then

$$F_n(x) \to F(x) \tag{1.52}$$

for any x, where the limiting cdf is continuous. Note that (1.52) also implies (1.51).

There exist inversion formulas for characteristic functions which will enable us to determine the distribution that corresponds to a certain characteristic function. For example, if

$$\int_{-\infty}^{\infty} |f(t)|\, dt < \infty,$$

where $f(t)$ is the characteristic function of a random variable X, then X has the pdf $p(x)$ given by

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} f(t)\, dt. \tag{1.53}$$

Remark 1.6 Instead of working with characteristic functions, one could define the *moment generating function* of a random variable X as $E\exp\{tX\}$ (a real function this time) and work with it. However, there are instances where this moment generating function may not exist, while the characteristic function always exists. A classic example of this may be seen later when we discuss Cauchy distributions. Nonetheless, when the moment generating function does exist, it uniquely determines the distribution just as the characteristic function does.

Exercise 1.9 Consider a random variable X which takes on values $0, 1, 2, \ldots$ with probabilities $p_n = P\{X = n\}, n = 0, 1, 2, \ldots$. Let $P(s)$ be its generating function. If it is known that $P(0) = 0$ and $P(\frac{1}{3}) = \frac{1}{3}$, find the probabilities p_n.

Exercise 1.10 Let $P(s)$ and $Q(s)$ be the generating functions of the random variables X and Y. Suppose it is known that both EX and EY exist and that $P(s) \geq Q(s), 0 \leq s < 1$. What can be said about $E(X - Y)$? Can this expectation be positive, negative, or zero?

Exercise 1.11 If $f(t)$ is a characteristic function, then prove that the functions

$$f_1(t) = \frac{1}{2 - f(t)}, \quad f_2(t) = |f(t)|^2, \quad \text{and} \quad f_3(t) = \text{Re } f(t),$$

where Re $f(t)$ denotes the real part of $f(t)$, are also characteristic functions.

Exercise 1.12 If $f(t)$ is a characteristic function that is twice differentiable, prove that the function $g(t) = f''(t)/f''(0)$ is also a characteristic function.

Exercise 1.13 Consider the functions $f(t)$ and $g(t) = 2f(t) - 1$. Then, prove that if $g(t)$ is a characteristic function, $f(t)$ also ought to be a characteristic function. The reverse may not be true. To prove this, construct an example of a characteristic function $f(t)$ for which $g(t)$ is not a characteristic function.

Exercise 1.14 Find the only function among the following which is a characteristic function:

$$f(t), \quad f^2(2t), \quad f^3(3t), \quad \text{and} \quad f^6(6t).$$

Exercise 1.15 Find the only function among the following which is not a characteristic function:

$$f(t), \quad 2f(t) - 1, \quad 3f(t) - 2, \quad \text{and} \quad 4f(t) - 3.$$

Exercise 1.16 It is easy to verify that $f(t) = \cos t$ is a characteristic function of a random variable that takes on values 1 and -1 with equal probability of $\frac{1}{2}$. Consider now the following functions:

$$\cos^2 3t, \quad \cos^3 2t \cos^4 3t, \quad \cos t^2, \quad \cos(\cos t), \quad e^{\cos^3 t - 1} \quad \text{and} \quad \frac{1}{2 - \cos t}.$$

Which of these are characteristic functions?

Exercise 1.17 Prove that the functions $f_n(t) = \cos^n t - \sin^n t, n = 1, 2, \ldots$ are characteristic functions only if n is an even integer.
