Handbook of THE NORMAL DISTRIBUTION

Jagdish K. Patel • Campbell B. Read

Department of Mathematics and Statistics
University of Missouri-Rolla
Rolla, Missouri

Department of Statistics
Southern Methodist University
Dallas, Texas

MARCEL DEKKER, INC. New York and Basel

Library of Congress Cataloging in Publication Data

Patel, Jagdish K.
Handbook of the normal distribution.

(Statistics, textbooks and monographs; v. 40)
Includes bibliographical references and index.
1. Gaussian distribution. I. Read, Campbell B.
II. Title. III. Series.
QA273.6.P373 519.2 81-17422
ISBN 0-8247-1541-1 AACR2

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

Current printing (last digit):
10 9 8 7 6 5 4 3 2 1

PRINTED IN THE UNITED STATES OF AMERICA

To my wife, Prabha, and children, Sanjay and Sonal--J.K.P.

To my mother and father--C.B.R.

PREFACE

This book contains a collection of results relating to the normal distribution. It is a compendium of properties, and problems of analysis and proof are not covered. The aim of the authors has been to list results which will be useful to theoretical and applied researchers in statistics as well as to students.

Distributional properties are emphasized, both for the normal law itself and for statistics based on samples from normal populations. The book covers the early historical development of the normal law (Chapter 1); basic distributional properties, including references to tables and to algorithms suitable for computers (Chapters 2 and 3); properties of sampling distributions, including order statistics (Chapters 5 and 8), Wiener and Gaussian processes (Chapter 9); and the bivariate normal distribution (Chapter 10). Chapters 4 and 6 cover characterizations of the normal law and central limit theorems, respectively; these chapters may be more useful to theoretical statisticians. A collection of results showing how other distributions may be approximated by the normal law completes the coverage of the book (Chapter 7).

Several important subjects are not covered. There are no tables of distributions in this book, because excellent tables are available elsewhere; these are listed, however, with the accuracy and coverage in the sources. The multivariate normal distribution other than the bivariate case is not discussed; the general linear model and regression models based on normality have been amply documented elsewhere; and the applications of normality in the methodology of statistical inference and decision theory would provide material for another volume on their own.

In citing references, the authors have tried to balance the aim of giving historical credit where it is due with the desirability of citing easily obtainable sources which may be consulted for further detail. In the latter case, we do not aim to cite every such work, but only enough to give the researcher or student a readily available source to which to turn.

We would like to thank the following persons for reviewing parts of the manuscript and giving helpful suggestions: Lee J. Bain, Herbert A. David, Maxwell E. Engelhardt, C. H. Kapadia, C. G. Khatri, Samuel Kotz, Lloyd S. Nelson, Donald B. Owen--who also gave editorial guidance, Stephen M. Stigler, Farroll T. Wright, and a referee. For assistance in typing the manuscript and for their infinite patience, we thank Connie Brewster, Sheila Crain, Millie Manley, and Dee Patterson; we would like to thank Dr. Maurits Dekker and the staff at MDI for their work in taking the manuscript through production. We would also like to thank Southern Methodist University for giving one of us leave for a semester in order to do research for the manuscript.

J.K.P. C.B.R.

CONTENTS

chapter 1

GENESIS: A HISTORICAL BACKGROUND

> I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the "Law of Frequency of Error." The law would have been personified by the Greeks and deified, if they had known of it. It reigns with serenity and in complete self-effacement amidst the wildest confusion.

So wrote Sir Francis Galton (1889, p. 66) about the normal distribution, in an age when the pursuit of science was tinged with the romanticism of the nineteenth century. In this age of computers, it is hard to find enthusiasm expressed with the sense of wonder of these men of letters, so much do we take for granted from modern technology.

In the seventeenth century Galileo (trans. 1953; 1962, pp. 303-309) expressed his conclusions regarding the measurement of distances to the stars by astronomers (Maistrov, 1974, pp. 31-34). He reasoned that random errors are inevitable in instrumental observations, that small errors are more likely to occur than large ones, that measurements are equally prone to err in one direction (above) or the other (below), and that the majority of observations tend to cluster around the true value. Galileo revealed here many of the characteristics of the normal probability distribution law, and also asserted that (random) errors made in observation are distinct from (systematic) final errors arising out of computation.

Although the study of probability began much earlier, modern statistics made its first great stride with the publication in 1713 of Jacob Bernoulli's *Ars Conjectandi*, in which Bernoulli proved the weak law of large numbers. The normal distribution first appeared in 1733 as an approximation to the probability for sums of binomially distributed quantities to lie between two values, when Abraham de Moivre communicated it to some of his contemporaries. A search by Daw and Pearson (1972) confirmed that several copies of this note had been bound up with library copies of de Moivre's *Miscellanea Analytica* which were printed in 1733 or later.

The theorem appeared again in de Moivre's book *The Doctrine of Chances* (1738, 1756; 1967; see also David, 1962, where it appears as an appendix). Although the main result is commonly termed "the de Moivre-Laplace limit theorem" (see [3.4.8]), the same approximation to binomial probabilities was obtained by Daniel Bernoulli in 1770-1771, but because he published his work through the Imperial Academy of Sciences in St. Petersburg, it remained there largely unnoticed until recently (Sheynin, 1970). Bernoulli also compiled the earliest known table of the curve $y = \exp(-\mu^2/100)$; see Table 1.1.

The natural development of probability theory into mathematical statistics took place with Pierre Simon de Laplace, who "was more responsible for the early development of mathematical statistics than any other man" (Stigler, 1975, p. 503). Laplace (1810, 1811; 1878-1912) developed the characteristic function as a tool for large sample theory and proved the first general central limit theorem; broadly speaking, central limit theorems show how sums of random variables tend to behave, when standardized to have mean zero and unit variance, like standard normal variables as the sample size becomes large; this happens, for instance, when they are drawn as random samples from "well-behaved" distributions; see [6.1].

Laplace showed that a class of linear unbiased estimators of linear regression coefficients is approximately normally distributed if the sample size is large; in 1812 (Laplace, 1812, chap. VIII) he proved that the probability distribution of the expectation of life at any specified age tends to the normal (Seal, 1967, p. 207).

TABLE 1.1 Some Early Tables of Normal Functions, with Coverage and Accuracy

Source	Function	Coverage	Accuracy
D. Bernoulli (1770-1771)	$\exp(-x^2/100)$	$x = 1(1)5(5)30$	4 sig.
Kramp (1799)	$\log_{10}\{\sqrt{\pi}[1 - \Phi(\sqrt{2}x)]\}$	$x = 0(0.01)3$	7 dec.
Legendre (1826)	$2\sqrt{\pi}[1 - \Phi(\sqrt{2}x)]$	$\begin{cases} x = 0(0.01)5 \\ \exp(-x^2) = 0(0.01)0.8 \end{cases}$	10 dec.
de Morgan (1837)	$2\Phi(\sqrt{2}x) - 1$	$x = 0(0.01)2$	7 dec.
de Morgan (1838)		$x = 0(0.01)3$	7 dec.
Glaisher (1871)	$\sqrt{\pi}[1 - \Phi(\sqrt{2}x)]$	$x = 3(0.01)4.5$	11 dec.
Markov (1888)		$x = 0(0.001)3(0.01)4.8$	11 dec.
Burgess (1898)	$2\Phi(\sqrt{2}x - 1)$	$x = 0(0.0001)1(0.002)3$ $(0.1)5(0.5)6$	15 dec.
Sheppard (1898)	$\sqrt{2\pi}\phi(x)$	$x = 0(0.05)4(0.1)5$	5 dec
Sheppard (1898)	Z_α	$2\alpha - 1 = \begin{cases} 0.1(0.1)0.9 \\ 0(0.01)0.99 \end{cases}$	10 dec. 5 dec.
Sheppard (1903)[a]	$\Phi(x)$, $\phi(x)$	$x = 0(0.01)6$	7 dec.
Sheppard (1907)[a]	Z_α	$\alpha = 0.5(0.001)0.999$	4 dec.

[a]Incorporated into Pearson and Hartley (1958).

Source: Greenwood and Hartley, 1962.

He derived the asymptotic distribution of a single-order statistic in a linear regression problem as normal, when the parent distribution is symmetric about zero and well behaved. In 1818 he showed that when the parent distribution is normal, the least squares estimator (LSE) has smaller variance than any linear combination of observations (Stigler, 1973). In the course of deriving this result, Laplace showed that the asymptotic joint distribution of the LSE and his order statistic estimator is bivariate normal (see Chapter 9), and obtained the minimum variance property of the LSE under normality while trying to combine the two estimators to reduce the variance.

These results pertain to Laplace's work as it relates to the normal distribution. The scope of his work is much fuller; for further discussion, see Stigler (1973, 1975a).

Problems arising from the collection of observations in astronomy led Legendre in 1805 to state the least squares principle, that of minimizing the sum of squares of "errors" of observations about what we would call in modern terms a regression plane; Legendre also obtained the normal equations. In 1809, Carl Friedrich Gauss published his *Theoria Motus Corporum Coelestium*, stating that he had used the least squares principle since 1795. This led to some controversy as to priority, involving Gauss, Laplace, Legendre, and several colleagues of Gauss (Plackett, 1972), but it all hinged upon whether publication should be the criterion for settling the issue or not. In the nineteenth century, research was often done independently, without knowledge of the achievements of others, as we shall see later. It comes as no surprise, then, that Gauss knew nothing of Legendre's earlier work when he published his *Theoria Motus*.

In this work, Gauss showed that the distribution of errors, assumed continuous, must be normal if the location parameter has (again in modern terminology) a uniform prior, so that the arithmetic mean is the mode of the posterior distribution (Seal, 1967). Gauss's linear least squares model was thus appropriate when the "errors" come from a normal distribution. An American mathematician, Robert Adrain (1808), who knew nothing of Gauss' work but who may have seen Legendre's book, derived the univariate and bivariate normal distributions as distributions of errors, and hence the method of least squares (Stigler, 1977), but his work did not influence the development of the subject.

The study of least squares, or the theory of errors, was to proceed for several decades without much further interaction with developing statistical theory. The normal distribution had not yet found its place in either theoretical or applied branches of the subject, and Gauss gave little further consideration to it (Seal, 1967). However, he points out (see Maistrov, 1974, pp. 155-156) that under the normal law, errors of any magnitude are possible. Once the universality of the normal law was accepted and then assumed, as it was later for some time, scientists also assumed that all observations should therefore be retained, resulting in a delay in developing methods for identifying and discarding outliers. For a good summary of Gauss' contributions to statistics and the theory of least squares, see Sprott (1978) or Whittaker and Robinson (1924, 1926).

The astronomer Friedrich Wilhelm Bessel (1818) published a comparison of the observed residuals and those expected from Gauss' normal law of errors and found a remarkably close agreement, from sets of 300 or more measurements of angular coordinates of stars. The publication of a book by Hagen (1837), which contained a derivation of the normal law as an approximation to the distribution of the total error, when that error is assumed to be the resultant of an infinitely large number of equal but equally likely positive or negative elementary errors, may have led Bessel in 1838 to develop the hypothesis of elementary errors. Bessel thus derived the normal law as an approximation for the total error, assumed now to be the sum of a large number of mutually independent, but not identically distributed elementary errors with well-behaved properties, including symmetrical distribution about zero.

The hypothesis of elementary errors became firmly established, particularly among astronomers like G. B. Airy (1861), who interpreted Laplace's central limit theorem from just such a point of view; in fact, the development of Laplace's theorem does not depend on the "elementary errors" assumption, although it could be made to do so. For further discussion, see Adams (1974, pp. 59-67).

In 1860 the Scottish mathematical physicist James Clerk Maxwell published the first of his two great papers on the kinetic theory of gases. Using geometrical considerations, he derived the normal distribution as the distribution of orthogonal velocity components of particles moving freely in a vacuum (Maxwell, 1860, pp. 22-23; 1952, pp. 380-381). His results lead through the work of Boltzmann to the modern theory of statistical mechanics and are notable as the first attempt to describe the motion of gases by a statistical function rather than a deterministic one. "The velocities are distributed among the particles" he wrote, "according to the same law as the errors are distributed among the observations in the theory of the method of least squares."

One of the first to fit a normal curve to data outside the field of astronomy was the Belgian scientist Adolphe Quetelet (1846), who did so, for example, by fitting a symmetric binomial distribution having 1000 outcomes, about the median height of 100,000 French conscripts (Stigler, 1975b). Quetelet was familiar with Laplace's central limit theorem, but this indirect approach avoided the use of calculus or of normal probability tables; some early tables associated with the normal distribution are cited in Table 1.1.

The first scientist in England to make use of earlier work on the continent was Francis Galton, who had a remarkable career in exploration, geography, the study of meteorology and, above all, anthropometry; the last-named is the study of anthropology through analysis of physical measurements. In his *Natural Inheritance* (1889), Galton drew on the work of Quetelet, noting his application of normal curve-fitting to human measurements and developing it to fit a model to describe the dependence of such measurements on those of an offspring's parents. Galton plotted his data in two dimensions and noted that data points of equal intensity appeared to lie on elliptical curves. From all of this he developed the linear regression model, the concept of correlation (1888), and the equation of the bivariate normal distribution (1886), with the help of a mathematician, Hamilton Dickson (see Chapter 9). The bivariate and trivariate normal distributions had already been developed independently by Bravais (1846), with application to data by Schols (1875), but although aware of the notion of correlation, these workers did not find it to have the degree of importance which Galton gave it.

A good account of how Galton arrived at the elliptical curves and bivariate normal distribution is given by Pearson (1920; 1970, p. 196), including the diagram from which he discovered observationally the form of the bivariate normal surface.

The publication of *Natural Inheritance* was the catalyst for the English school of biometry to begin to take major strides. The zoologist W. F. R. Weldon sought the help of an applied mathematician, Karl Pearson, realizing that statistical methods might establish evidence to support Darwin's theory of natural selection. The "law of error," as Galton termed it, prompted Weldon to assume initially that all physical characters in homogeneous animal populations would be normally distributed (Pearson, 1965; 1970, p. 328).

Galton realized, however, that sets of data might well follow some other frequency law. The geometric mean of a set of observations, he wrote, might better represent the most probable value of a distribution, and if so, the logarithms of the observations might be assumed to follow the normal law. This led to the lognormal distribution, but it stimulated Pearson to develop a system of frequency curves, depending on a set of parameters, which would cover all distributions occurring in nature, at least those which are continuous (Pearson, 1967; 1970, pp. 344-345).

Karl Pearson generalized and gave precision to Galton's discussion of correlation by developing a theory of multiple correlation and multiple regression. Such a theory, he realized, would be necessary to answer the kind of questions which were being posed by Weldon. Working from two, three, and four variables, Francis Edgeworth (1892) provided the first statement of the multivariate normal distribution, and Pearson (1896) gave an explicit derivation of it (Seal, 1967).

It is noteworthy that the work of Legendre, Gauss, and the least squares school was neglected and unnoticed during the early years of biometry in England (1885-1908), possibly because of Pearson's preoccupation with the multivariate normal distribution (Seal, 1967). Pearson's predictive regression equation (1896) was not seen to be identical in form and solution to Gauss's 1809 model. Duplication of research in least squares theory continued in the work of R. A. Fisher and others at Rothamsted until well after 1930 (Seal, 1967).

In trying to fit data to his frequency curves, Pearson was faced with the need to test the goodness of fit. Ernst Abbe (1863; 1906) had derived the distribution of ΣX_i^2, where $X_1, \ldots, X_n$ is a random sample from a normal distribution with mean zero (Sheynin, 1966; Kendall, 1971), and Helmert (1876) derived the distribution of $\Sigma(X_i - \bar{X})^2$, where $\bar{X}$ is the arithmetic mean of $X_1, \ldots, X_n$. These yield the chi-square distribution with n and n - 1 degrees of freedom, respectively. Abbe's work went unnoticed until 1966, credit prior to that time having been given to Helmert. The matrix transformation of Helmert, however, is still used as an instructional tool for deriving the distribution of $\Sigma(X_i - \bar{X})^2$ and for establishing the independence of $\bar{X}$ and the sample variance $\Sigma(X_i - \bar{X})^2/(n - 1)$ in a normal sample.

The problem is quite different if the parent population is not normal. Lancaster (1966) shows how Laplace provided the necessary techniques for Bienaymé in 1838 and 1852 to obtain chi-square as an asymptotic large-sample distribution without any assumption of normality. Bienaymé obtained something close to "Σ(observed - expected)^2/(expected)," which is Pearson's chi-square statistic, summation being over classes.

Weldon had found a noticeable exception to the fit of the normal curve to a set of data for the relative frontal breadth of Naples crabs (Pearson, 1965; 1970, p. 328). Thinking it might be compounded from two subspecies, he fitted a compound normal distribution (see

[1.6]) or "double-humped curve," as he called it, but the fit was done purely by trial and error in 1892. Karl Pearson's first statistical paper in 1894 tackled this problem, introducing the method of moments as a technique in fitting a frequency distribution. But the question of whether the fitted distribution was reasonable led Pearson to the work which resulted in his paper of 1900 and established the chi-square goodness-of-fit test firmly as a cornerstone of modern statistics. Ironically, the first sacred cow to fall with this new tool was the law of errors; Pearson soundly berated the astronomer G. B. Airy (1861), who had tried to illustrate the universality of the normal law with an unlucky set of data. Using the same data and his new chi-square test, Pearson showed that the normal law gave an unacceptable fit after all. "How healthy is the spirit of scepticism," he wrote, "in all inquiries concerning the accordance of theory and nature." (Pearson, 1900, p. 172).

Pearson and the English school of biometry preferred to work with large data sets for the statistical analysis of the various problems which they faced. But W. S. Gosset ("Student") was compelled by circumstances arising in the Guinness brewery company in Dublin to solve problems for which small samples only could be taken. A year of work in the Biometric Laboratory under Pearson led to his famous paper "The Probable Error of a Mean" (Student, 1908) in which he deduced the t distribution, the ratio of the sample mean to standard deviation in a normal sample. Pearson seemed to have little interest in Gosset's results, perhaps because he felt wary of letting any biologist or medical research worker believe that there was a simple method of drawing conclusions from scanty data.

Gossett was unaware of Helmert's derivation of the sampling distribution of $\Sigma(X_i - \bar{X})^2$, but he inferred its distribution and showed it to be uncorrelated with $\bar{X}$. For other work by Gosset, see Pearson (1970, pp. 348-351, 360-403), or Pearson and Wishart (1958).

The English school of biometry did not undertake a serious study of probability theory, unlike the Russian school of Pafnuti Lvovich Tchebyshev and his pupils Markov and Lyapunov. From the middle of the nineteenth century this school applied mathematical rigor to laws of large numbers, dependent events, and central limit properties; with the introduction of the concept of a random variable, it was able to establish sufficient conditions for standardized sums of dependent, as well as of independent random variables to converge to the normal law. The first clear statement of the problem, together with a proof which later required revisions and additional proof (Maistrov, 1974, pp. 202-208) was given by Tchebyshev in 1887 (reprinted 1890) using the method of moments. The importance of Tchebyshev's approach to central limit theorems lay in the clearly defined mathematical character he ascribed to random variables. He did this by establishing restrictions on the applicability of results in probability theory, so that in every set of circumstances one might determine whether or not the limit theorems hold. It was left to Andrei Andreevich Markov (1898) to correct Tchebyshev's theorem, and to Alexander Mikhailovich Lyapunov (1901; 1954-1965, Vol. I, pp. 157-176) to produce a central limit theorem of great generality, rigorously proved with the tools of classical analysis, including that of characteristic functions. Since this does not require the existence of moments of any order, it might appear that Tchebyshev's method of moments had been outlived; apparently challenged by this, Markov (1913; 1951, pp. 319-338) proved Lyapunov's theorem by the method of moments with the now well-known device of introducing truncated random variables. For a statement of these and other limit theorems as they developed historically, see [6.1] and Uspensky (1937, app. II). This work was to be continued and put on a firm axiomatic basis by Bernstein, Khinchine, and Kolmogorov. For further accounts of the early history of the Russian school, see Maistrov (1974) and Adams (1974).

In the same paper, Tchebyshev (1890) developed a series expansion and an upper bound for the difference between the cumulative distribution function $F_n(x)$ of a standardized sum $(\Sigma X_i)/(\sqrt{n}\sigma)$ and that of a standard normal variable, where σ is the standard deviation in the population of interest. In 1905, Charlier introduced a series to improve the central limit theorem approximation to the density function of $(\Sigma X_i)/(\sqrt{n}\sigma)$ in terms of the standard normal density ϕ: Edgeworth (1905) developed his series expansion for $F_n(x)$ with the same purpose in mind, although he had produced versions for symmetrical densities as early as 1883. See [3.5], Gnedenko and Kolmogorov (1968, chap. 8), and Stigler (1978).

The early story of the normal distribution is largely the story of the beginnings of statistics as a science. We leave the story at this point; with R. A. Fisher modern statistics begins to branch out and accelerate. From 1915 onward Fisher found the distribution of the correlation coefficient, of the absolute deviation $\Sigma|X_i - \bar{X}|/n$ in normal samples, of regression coefficients, correlation ratios, multiple regression and partial correlation coefficients, and of the ratio F of sample variances from two normal populations. At the same time he developed his ideas of estimation, sufficiency, likelihood, inference, the analysis of variance, and experimental design (Kendall, 1970; Savage, 1976). Normality assumptions have played a key role in statistical analysis through the years, but since the 1960s considerably more attention has been given to questioning these assumptions, requiring estimators that are robust when such assumptions are violated, and to devising further enlightening tests of their validity.

A final word about nomenclature. The normal distribution has been named after various scientists, including Laplace and Gauss. Kac (1975, pp. 6-7) recalls that it has also been named after Quetelet and Maxwell. Stigler (1980) points out that although a few modern writers refer to it as the Laplace or Laplace-Gauss distribution, and engineers name it after Gauss, no modern writers name it after its originator Abraham de Moivre, as far as is known. Francis Galton called it by several names, indicating that before 1900 no one term had received common acceptance. Among these names were *law of frequency of error* and *the exponential law* (1875), *law of deviation from an average*, and *law of errors of observation* (1869).

In 1877, Galton first used the name *normal law* in the sense that it is commonly encountered in statistics, but the earliest known use of the terms, and that in the same sense, was in 1873, by the American Charles Sanders Peirce (Kruskal, 1978). The label *normal law* or *normal distribution* gained acceptance with the English school of biometry; Karl Pearson (1921-1933; 1978, p. 156) claimed to have coined it, but was apparently unaware of earlier uses.

REFERENCES

Abbe, E. (1906; 1863). Uber die Gesetzmässigkeit in der Vertheilung der Fehler bei Beobachtungsreihen, Gesammelte Abhandlungen, Vol. II, Jena: Gustav Fisher. English translation (1968), ref. P. B. 191, 292, T., from National Technical Information Service, Springfield, Va.

Adams, W. J. (1974). *The Life and Times of the Central Limit Theorem*, New York: Caedmon.

Adrain, R. (1808). Research concerning the probabilities of errors which happen in making observations, etc., *The Analyst; or Mathematical Museum* 1(4), 93-109.

Airy, G. B. (1861). *On the Algebraical and Numerical Theory of Errors of Observations and the Combination of Observations*, London: Macmillan.

Bernoulli, D. (1770-1771). Mensura sortis ad fortuitam successionem rerum naturaliter contingentium applicata, *Novi Commentarii Academiae Scientiarum Imperialis Petropolitanae* 14, 26-45; 15, 3-28.

Bernoulli, J. (1713). Ars Conjectandi; German version (1899): *Wahrscheinlichkeitsrechnung*, Leipzig: Engelmann.

Bessel, F. W. (1818). Astronomiae pro anno MDCCLV deducta ex observationibus viri incomporabilis James Bradley specula Grenovicensi per annos 1750-1762 institutis, Regiomonti.

Bessel, F. W. (1838). Untersuchungen über der Wahrscheinlichkeit der Beobachtungsfehler, *Astronomische Nachrichten* 15, 368-404.

Bienaymé, J. (1838). Sur la probabilité des resultats moyens des observations; demonstration directe de la règle de Laplace, *Mémoires des Savans Étrangers, Académie (Royale) des Sciences de l'Institut de France*, Paris, 5, 513-558.

Bienaymé, J. (1852). Sur la probabilité des erreurs d'après la méthode des moindres carrés; reprinted (1858): *Mémoires des Savans Étrangers, Académie (Royale) des Sciences de l'Institut de France*, Paris, 15, 615-663.

Bravais, A. (1846). Analyse mathematique sur les probabilités des erreurs de situation d'un point, *Mémoires des Savans Étrangers, Académie (Royale) des Sciences de l'Institut de France*, Paris, 9, 255-332.

Burgess, J. (1898). On the definite integral... [i.e., ...], with extended tables of values, *Transactions of the Royal Society of Edinburgh* 39, 257-321.

Charlier, C. V. L. (1905). Über die Darstellung willkürlicher Funktionen, *Arkiv för Matematik, Astronomi och Fysik* 2(20), 1-35.

David, F. N. (1962). *Games, Gods, and Gambling*, New York: Hafner.

Daw, R. H., and Pearson, E. S. (1972). Studies in the history of probability and statistics, XXX: Abraham de Moivre's 1733 derivation of the normal curve: A bibliographical note, *Biometrika* 59, 677-680.

Edgeworth, F. Y. (1883). The law of error, *Philosophical Magazine* 16, 300-309.

Edgeworth, F. Y. (1892). Correlated averages, *Philosophical Magazine* Ser. 5, 34, 190-204.

Edgeworth, F. Y. (1905). The law of error, *Proceedings of the Cambridge Philosophical Society* 20, 36-65.

Galileo, G. (trans. 1953; 2nd ed. 1962). *Dialogue Concerning the Two Chief World Systems--Ptolemaic and Copernican* (trans. S. Drake), Berkeley: University of California Press.

Galton, F. (1869). *Hereditary Genius*, London: Macmillan.

Galton, F. (1875). Statistics by intercomparison, with remarks on the law of frequency of error, *Philosophical Magazine* 49, 33-46.

Galton, F. (1886). Family likeness in stature, *Proceedings of the Royal Society* 40, 42-73.

Galton, F. (1888). Co-relations and their measurement, chiefly from anthropometric data, *Proceedings of the Royal Society* 45, 135-145.

Galton, F. (1889). *Natural Inheritance*, London: Macmillan.

Gauss, C. F. (1809). *Theoria Motus Corporum Coelestium*, Lib. 2, Sec. III, 205-224, Hamburg: Perthes u. Besser.

Gauss, C. F. (1816). Bestimmung der Genauigkeit der Beobachtungen, *Zeitschrift Astronomi* 1, 185-197.

Glaisher, J. W. L. (1871). On a class of definite integrals--Part II, *Philosophical Magazine* 42, 421-436.

Gnedenko, B. V., and Kolmogorov, A. N. (1968). *Limit Distributions for Sums of Independent Random Variables* (trans. from Russian), Reading, Mass.: Addison-Wesley.

Greenwood, J. A., and Hartley, H. O. (1962). *Guide to Tables in Mathematical Statistics*, Princeton, N.J.: Princeton University Press.

Helmert, F. R. (1876). Die Genauigkeit der Formel von Peters zur Berechnung des wahrscheinlichen Beobachtungsfehlers director

Beobachtungen tungen gleicher Genauigkeit, *Astronomische Nachrichten* 88, 113-120.

Kac, M. (1975). Some reflections of a mathematician on the nature and the role of statistics, *Proceedings of the Conference on Directions for Mathematical Statistics*, 5-11, Applied Probability Trust.

Kendall, M. G. (1963). Ronald Aylmer Fisher, 1890-1962, *Biometrika* 50, 1-15; reprinted (1970) in *Studies in the History of Probability and Statistics* (E. S. Pearson and M. G. Kendall, eds.), 439-454, New York: Hafner.

Kendall, M. G. (1971). Studies in the history of probability and statistics, XXVI: The work of Ernst Abbe, *Biometrika* 58, 369-373; Corrigendum (1972): *Biometrika* 59, 498.

Kramp, C. (1799). Analyse des réfractions astronomiques et terrestres, Leipsic: Schwikkert; Paris: Koenig.

Kruskal, W. (1978). Formulas, numbers, words: Statistics in prose, *The American Scholar* 47, 223-229.

Lancaster, H. O. (1966). Forerunners of the Pearson χ^2, *Australian Journal of Statistics* 8, 117-126.

Laplace, P. S. (1809-1810). Mémoire sur les approximations des formules qui sont fonctions de très grands nombres et sur leur application aux probabilités, *Mémoires de la Classe des Sciences mathématiques et physiques de l'Institut*, Paris, 353-415, 559-565, et. passim (see *Oeuvres Complètes* 12, 301-353).

Laplace, P. S. (1810). Mémoire sur les intégrales définies et leur application aux probabilités, *Mémoires de la Classe des Sciences mathématiques et physiques de l'Institut*, Paris, 279-347, et. passim (1810-1812: see *Oeuvres Complètes* 12, 357-412).

Laplace, P. S. (1812). *Théorie Analytique des Probabilités*, Paris (*Oeuvres Complètes*, Vol. 7).

Laplace, P. S. (1878, 1912). *Oeuvres Complètes de Laplace*, 14 volumes, Paris: Gauthier-Villars.

Legendre, A. M. (1805). *Nouvelles Methodes pour la Determination des Orbites des Comètes*, Paris.

Legendre, A. M. (1826). *Traite des Fonctions elliptiques et des Integrales Euleriennes, avec des Tables pour en faciliter le Calcul numerique*, Vol. 2, Paris: Huzard-Courcier.

Lyapunov, A. M. (1901). Nouvelle forme du théorème sur la limite de probabilité, *Mémoires de l'Académie Impériale des Sciences de St. Pétersbourg* 12, 1-24.

Lyapunov, A. M. (1954-1965). *Izbrannye Trudi* (Selected Works), Academy of Sciences, USSR.

Maistrov, L. E. (1967, 1974). *Probability Theory: A Historical Sketch* (trans. by S. Kotz), New York: Academic Press.

Markov, A. A. (1888). Table des valeurs de l'integrale..., St. Pêtersbourg: Acadêmie Impériale des Sciences.

Markov, A. A. (1899-1900). The law of large numbers and the method of least squares, *Izvestia Physiko-mathematicheskago Obschvestva pri Imperatorskom Kazanskom Universitet* 8, 110-128; 9, 41-43.

Markov, A. A. (1913). A probabilistic limit theorem for the cases of Academician A. M. Lyapounov, *Izvlyechenye iz knigi ischislyenye veroyatnostyei* (supplement to "Theory of Probability"), 4-e.

Markov, A. A. (1951). *Izbrannye Trudi* (Selected Works), Academy of Sciences, USSR.

Maxwell, J. C. (1860). Illustrations of the dynamical theory of gases, *Philosophical Magazine* 19, 19-32; 20, 21-37; (1952): *Scientific Papers of James Clerk Maxwell*, 377-409.

de Moivre, A. (1733). Approximatio ad Summam Ferminorum Binomii $(a + b)^n$ in Seriem expansi.

de Moivre, A. (1738, 1756). *The Doctrine of Chances*; reprint (1967): New York: Chelsea.

de Morgan, A. (1837). Theory of probabilities, in *Encyclopaedia Metropolitana* 2, 359-468.

de Morgan, A. (1838). An essay on probabilities and on their application to life contingencies and insurance offices, London: Longmans.

Pearson, E. S. (1965). Some incidents in the early history of Biometry and Statistics, *Biometrika* 52, 3-18; reprinted (1970) in *Studies in the History of Statistics and Probability* (E. S. Pearson and M. G. Kendall, eds.), 323-338, New York: Hafner.

Pearson, E. S. (1967). Some reflexions on continuity in the development of mathematical statistics, 1885-1920, *Biometrika* 54, 341-355; reprinted (1970), as in Pearson (1965), 339-353.

Pearson, E. S. (1970). William Sealy Gosset, 1876-1937: "Student" as a statistician, *Studies in the History of Statistics and Probability* (E. S. Pearson and M. G. Kendall, eds.), 360-403, New York: Hafner.

Pearson, E. S., and Hartley, H. O. (1958). *Biometrika Tables for Statisticians* 1 (2nd ed.), London: Cambridge University Press.

Pearson, E. S., and Wishart, J., eds. (1958). *"Student's" Collected Papers*, Cambridge: Cambridge University Press.

Pearson, K. (1896). Mathematical contributions to the theory of evolution, III: Regression, heredity, and panmixia, *Philosophical Transactions of the Royal Society of London* A187, 253-318.

Pearson, K. (1900). On a criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, *Philosophical Magazine* (5)50, 157-175.

Pearson, K. (1920). Notes on the history of correlation, *Biometrika* 13, 25-45; reprinted (1970) in *Studies in the History of Statistics and Probability* (E. S. Pearson and M. G. Kendall, eds.), 185-205, New York: Hafner.

Pearson, K. (1921-1933). *The History of Statistics in the 17th and 18th Centuries, against the Changing Background of Intellectual, Scientific, and Religious Thought* (lectures, E. S. Pearson, ed., 1978), New York: Macmillan.

Plackett, R. L. (1972). Studies in the history of probability and statistics, XXIX: The discovery of the method of least squares, *Biometrika* 59, 239-251.

Quetelet, L. A. J. (1846). *Lettres à S. A. R. Le Duc Regnant de Saxe-Cobourg et Gotha, sur la Theorie des Probabilités appliquée aux Sciences Morales et Politiques* (English trans., 1849), Brussels: Hayez.

Savage, L. J. (1976). On rereading R. A. Fisher (with discussion), *Annals of Statistics* 4, 441-500.

Schols, C. M. (1875). Over de theorie des fouten in de ruimte en in het platte vlak, *Verhandelingen der koninklijke Akademie van Wetenschappen*, Amsterdam, 15, 1-75.

Seal, H. L. (1967). Studies in the history of probability and statistics, XV: The historical development of the Gauss linear model, *Biometrika* 54, 1-24; reprinted (1970) in *Studies in the History of Probability and Statistics* (E. S. Pearson and M. G. Kendall, eds.), 207-230, New York: Hafner.

Sheppard, W. F. (1898). On the application of the theory of error to cases of normal distribution and normal correlation, *Philosophical Transactions of the Royal Society of London* A192, 101-167.

Sheppard, W. F. (1903). New tables of the probability integral, *Biometrika* 2, 174-190.

Sheppard, W. F. (1907). Table of deviates of the normal curve, *Biometrika* 5, 404-406.

Sheynin, O. B. (1966). Origin of the theory of errors, *Nature* 211, 1003-1004.

Sheynin, O. B. (1968). Studies in the history of probability and statistics, XXI: On the early history of the law of large numbers, *Biometrika* 55, 459-467; reprinted (1970) in *Studies in the History of Statistics and Probability* (E. S. Pearson and M. G. Kendall, eds.), 231-239, New York: Hafner.

Sheynin, O. B. (1970). Studies in the history of probability and statistics, XXIII: Daniel Bernoulli on the normal law, *Biometrika* 57, 199-202.

Sprott, D. A. (1978). Gauss's contributions to statistics, *Historia Mathematica* 5, 183-203.

Stigler, S. M. (1973). Laplace, Fisher, and the discovery of the concept of sufficiency, *Biometrika* 60, 439-445.

Stigler, S. M. (1975a). Studies in the history of probability and statistics, XXXIV: Napoleonic statistics: The work of Laplace, *Biometrika* 62, 503-517.

Stigler, S. M. (1975b). The transition from point to distribution estimation, *Bulletin of the International Statistical Institute* 46 (Proceedings of the 40th Session, Book 2), 332-340.

Stigler, S. M. (1977). An attack on Gauss, published by Legendre in 1820, *Historia Mathematica* 4, 31-35.

Stigler, S. M. (1978). Francis Ysidro Edgeworth, statistician, *Journal of the Royal Statistical Society* A141, 287-313, followed by discussion.

Stigler, S. M. (1980). Stigler's law of eponymy, *Transactions of the New York Academy of Sciences II* 39 (Science and Social Structure: A Festschrift for Robert K. Merton), 147-157.

Student (1908). The probable error of a mean, *Biometrika* 6, 1-25.

Tchebyshev, P. L. (1890). Sur deux théorèmes relatifs aux probabilités, *Acta Mathematica* 14, 305-315; reprinted (1962) in *Oeuvres*, Vol. 2, New York: Chelsea.

Uspensky, J. V. (1937). *Introduction to Mathematical Probability*, New York: McGraw-Hill.

Whittaker, E. T., and Robinson, G. (1926). *The Calculus of Observations*, London: Blackie.

chapter 2

SOME BASIC AND MISCELLANEOUS RESULTS

2.1 DEFINITIONS AND PROPERTIES

Some of the following results and others appear in Abramowitz and Stegun (1964, pp. 931-935). Further properties may be found in Chapters 3 and 4; for distributions of sampling statistics from normal populations, see Chapters 5 and 7.

[*2.1.1*] The probability density function (pdf) of a normal random variable X is given by

$$f(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \qquad -\infty < x < \infty,\ -\infty < \mu < \infty,\ \sigma > 0$$

The constants μ, σ, and σ^2 are, respectively, the mean, standard deviation, and variance of the normal distribution. Let $F(x;\mu,\sigma^2)$ denote the cumulative distribution function (cdf) of X.

[*2.1.2*] The pdf of a standard normal random variable Z is given by

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right), \qquad -\infty < z < \infty$$

Let $\Phi(z)$ denote the cdf of Z.

Throughout the book the notation $N(\mu,\sigma^2)$ will denote the normal distribution $F(\cdot,\mu,\sigma^2)$ with mean μ and variance σ^2. Also, the notations f, F, ϕ, and Φ are used only in reference to normal distributions.

We shall denote the quantiles of Z by z_β, where

$$\Pr(Z \le z_\beta) = \Phi(z_\beta) = 1 - \beta$$

[*2.1.3*] The random variables X and Z are linearly related and the relationship is given by $Z = (X - \mu)/\sigma$. Hence, for all x and z, $F(x) = \Phi[(x - \mu)/\sigma]$ and $\Phi(z) = F(\mu + \sigma z)$.

[*2.1.4*] The curves f(x) and $\phi(z)$ are symmetric about $x = \mu$ and $z = 0$, respectively. Hence, for all x and z, $f[(\mu - x)/\sigma] = f[(x - \mu)/\sigma]$,

$$F(-x) = 1 - F(x + 2\mu) \qquad \phi(-z) = \phi(z) \qquad \Phi(-z) = 1 - \Phi(z)$$
$$z_\beta = -z_{1-\beta} \qquad \Phi(z_\beta) - \Phi(z_{1-\beta}) = 1 - 2\beta, \qquad 0 < \beta < 0.5$$

[*2.1.5*] The pdf f(x) is unimodal with mode, median, and expected value at $x = \mu$. The curve f(x) has two points of inflection, at $\mu - \sigma$ and $\mu + \sigma$. Similar properties hold for $\phi(x)$.

[*2.1.6*] The function f(x) is logconcave in x; also $\Phi(z)$ is logconcave in z (Tong, 1978, p. 661). This result is useful in deriving inequalities for normal probabilities.

[*2.1.7*] The error function erf(x) is defined by $\mathrm{erf}(x) = 2\Phi(x\sqrt{2}) - 1$, $x \ge 0$. The incomplete gamma function ratio P(a,x) and $\phi(x)$ are related by $P(1/2,x) = 2\Phi(\sqrt{2x}) - 1$, $x \ge 0$, where $P(a,x) = \left(\int_0^x t^{a-1}e^{-t}\,dt\right)/\Gamma(a)$.

[*2.1.8*] Let X be a $N(\mu,\sigma^2)$ random variable, and $Y = aX + b$, where a and b are any real numbers. Then the random variable Y has a $N(a\mu + b, a^2\sigma^2)$ distribution.

[*2.1.9*] *Repeated Derivatives of* $\phi(x)$. (a) Let

$$\left(-\frac{d}{dx}\right)^r \phi(x) = H_r(x)\phi(x)$$

where $H_r(x)$ is a Tchebyshev-Hermite polynomial of degree r in x. The polynomial $H_r(x)$ is the coefficient of $t^r/r!$ in $\exp(tx - t^2/2)$, so that

$$H_r(x) = x^r - \frac{r^{(2)}}{2\cdot 1!} x^{r-2} + \frac{r^{(4)}}{2^2\cdot 2!} x^{r-4} - \frac{r^{(6)}}{2^3\cdot 3!} x^{r-6} + \cdots$$

where $r^{(a)} = r(r - 1)\ldots(r - a + 1)$. The first nine polynomials are given by

$$H_0(x) = 1$$
$$H_1(x) = x$$
$$H_2(x) = x^2 - 1$$
$$H_3(x) = x^3 - 3x$$
$$H_4(x) = x^4 - 6x^2 + 3$$
$$H_5(x) = x^5 - 10x^3 + 15x$$
$$H_6(x) = x^6 - 15x^4 + 45x^2 - 15$$
$$H_7(x) = x^7 - 21x^5 + 105x^3 - 105x$$
$$H_8(x) = x^8 - 28x^6 + 210x^4 - 420x^2 + 105$$

Expressions for $H_r(x)$ of higher degree can be obtained from the recurrence relation

$$H_r(x) = xH_{r-1}(x) - (r - 1)H_{r-2}(x)$$

See Kendall and Stuart (1977, pp. 167-168) and Draper and Tierney (1973, pp. 507-508), where expressions for $H_r(x)$ are given explicitly when $r = 0, 1, 2, \ldots, 27$.

(b) Let

$$\phi^{(m)}(x) = \frac{d^m}{dx^m} \phi(x)$$

Then the differential equation

$$\phi^{(m+2)}(x) + x\phi^{(m+1)}(x) + (m + 1)\phi^{(m)}(x) = 0$$

is satisfied. The value of $\phi^{(m)}(x)$ at $x = 0$ is (Abramowitz and Stegun, 1964, p. 933)

$$\phi^{(m)}(0) = \begin{cases} \dfrac{(-1)^{m/2} m!}{(\sqrt{2\pi}) 2^{m/2} (m/2)!}, & m = 2r, \quad r = 0, 1, 2, \ldots \\ 0, & m \text{ odd} \quad m > 0 \end{cases}$$

See [3.1] and Table 3.1 for listings of tables of these derivatives.

[*2.1.10*] Repeated Integrals of $\phi(x)$. (a) Let

$$I_n(x) = \int_x^\infty I_{n-1}(y)\, dy, \qquad n \geq 0$$

where $I_{-1}(x) = \phi(x)$.

(b) $I_{-n}(x) = \left(-\dfrac{d}{dx}\right)^{n-1} \phi(x) = (-1)^{n-1}\phi^{(n-1)}(x), \qquad n \geq -1$

(c) $\left(\dfrac{d^2}{dx^2} + x\dfrac{d}{dx} - n\right) I_n(x) = 0$

(d) $(n + 1)I_{n+1}(x) + xI_n(x) - I_{n-1}(x) = 0, \qquad n > -1$

(e) $I_n(x) = \displaystyle\int_x^\infty \frac{(y - x)^n}{n!}\phi(y)\, dy, \qquad n > -1$

(f) $I_n(0) = I_{-n}(0) = \left[\left(\dfrac{n}{2}\right)! 2^{1+n/2}\right]^{-1}, \qquad n$ even

(Abramowitz and Stegun, 1964, pp. 934-935)

[*2.1.11*] Sheppard (1898, pp. 104-106) gives an interesting geometrical property of the normal curve. Suppose that the mean is zero and the standard deviation σ. Let P be a point and let PM be the ordinate at P, so that OM = x. Let P move so that if the tangent PT to the locus of P intersects the x axis at T, then

$$OM \cdot MT = \sigma^2$$

The locus of points of P is then a normal curve, but the area under the curve will not necessarily be unity. The additional condition that the ordinate at 0 is of height $1/\sqrt{2\pi}$ yields the curve y = $f(x;0,\sigma^2)$, the normal pdf.

2.2 MOMENTS, CUMULANTS, AND GENERATING FUNCTIONS

[*2.2.1*] Let X be a $N(\mu,\sigma^2)$ random variable (rv). For r = 1, 2, 3, ..., let $E(X^r) = \mu'_r$; $E(X - \mu'_1)^r = \mu_r$; $E|X - \mu'_1|^r = \nu_r$; and κ_r be the cumulant of order r, defined by the equation (Kendall and Stuart, 1977, p. 69)

$$\exp\left[\kappa_1 t + \frac{\kappa_2 t^2}{2!} + \cdots + \frac{\kappa_r t^r}{r!} + \cdots\right] = 1 + \mu'_1 t + \frac{\mu'_2 t^2}{2!} + \cdots + \frac{\mu'_r t^r}{r!} + \cdots$$

for values of the dummy variable t in some interval. Other than κ_1, all cumulants remain unchanged if a rv has its origin shifted by some fixed quantity, say, a.

For a normal distribution, all absolute moments ν_r and hence moments μ'_r, μ_r, and cumulants κ_r of all orders exist (Kendall and Stuart, 1977, p. 75). Carleman's condition, that a set of moments uniquely determines a distribution if $\Sigma_{i=0}^{\infty}(\mu_{2i})^{-1/(2i)}$ diverges, is satisfied by the normal distribution (Kendall and Stuart, 1977, p. 114).

[*2.2.2*] The following results hold for $N(\mu,\sigma^2)$ variables:

Mean = $\mu'_1 = \mu$

Variance = $\mu_2 = \sigma^2$

Mode = median = μ

$\mu_3 = 0$, $\mu_4 = 3\sigma^4$, $\mu_5 = 0$, $\mu_6 = 15\sigma^6$, $\mu_7 = 0$, $\mu_8 = 105\sigma^8$

Mean deviation = $\nu_1 = \sqrt{(2\sigma^2/\pi)}$, $\nu_2 = \sigma^2$, $\nu_3 = 2\sigma^3\sqrt{2/\pi}$,

$$\nu_4 = 3\sigma^4, \quad \nu_5 = 8\sigma^5\sqrt{2/\pi}$$

Cumulants: $\kappa_1 = \mu$, $\kappa_2 = \sigma^2$, $\kappa_r = 0$ if $r \ge 3$

Coefficient of variation = σ/μ

Coefficient of skewness = $\beta_1 = \dfrac{\mu_3^2}{\mu_2^3} = 0$

$$\text{Kurtosis} = \beta_2 = \frac{\mu_4}{\mu_2^2} = 3$$

$$\text{Coefficient of excess} = \beta_2 - 3 = 0$$

[*2.2.3*]

(a) $$\mu'_{2r-1} = \sigma^{2r-1} \sum_{i=1}^{r} \frac{(2r-1)!(\mu)^{2i-1}}{(2i-1)!(r-i)!2^{r-i}\sigma^{2i-1}}, \quad r = 1, 2, 3, \ldots$$

$$\mu'_{2r} = \sigma^{2r} \sum_{i=0}^{r} \frac{(2r)!\mu^{2i}}{(2i)!(r-i)!2^{r-i}\sigma^{2i}}, \quad r = 1, 2, 3, \ldots$$

(Bain, 1969, p. 34)

Recurrence relation:

$$\mu'_{r+1} = r\sigma^2\mu'_{r-1} + \mu\mu'_r, \quad r = 1, 2, 3, \ldots \quad \text{(Bain, 1969, p. 34)}$$

(b) $$\mu_{2r+1} = 0, \quad r = 0, 1, 2, 3, \ldots$$

$$\mu_{2r} = \nu_{2r} = (\sigma^2/2)^r(2r)!/r!, \quad r = 0, 1, 2, 3, \ldots$$

(Bain, 1969, p. 34; Kendall and Stuart, 1977, p. 143)

$$\nu_{2r-1} = (2\pi)^{-1/2}(r-1)!2^r\sigma^{2r-1}, \quad r = 1, 2, 3, \ldots$$

(Lukacs, 1970, p. 13)

$$\nu_r = 2^{r/2}\Gamma\{(r+1)/2\}\sigma^r/\sqrt{\pi}, \quad r = 1, 2, \ldots$$

(Kamat, 1953, p. 21)

[*2.2.4*] *Incomplete Moments*. Let $I_r(a) = \int_a^\infty y^r\phi(y)\,dy$; $r = 0, 1, 2, \ldots$ and $I_r(a) = 0$ if $r < 0$. Then $I_0(a) = 1 - \Phi(a)$ and (Elandt, 1961, p. 551)

$$I_r(a) = a^{r-1}\phi(a) + (r-1)I_{r-2}(a)$$

$$I_r(a) = [a^{r-1} + (r-1)a^{r-3} + (r-1)(r-3)\cdots5\cdot3\cdot a]\phi(a) + (r-1)(r-3)\cdots5\cdot3\cdot I_0(a), \quad r \text{ even}$$

$$I_r(a) = [a^{r-1} + (r-1)a^{r-3} + \cdots + (r-1)(r-3)\cdots4.2]\phi(a), \quad r \text{ odd}.$$

[*2.2.5*] (Kendall and Stuart, 1977, p. 143)

Moment-generating function = $E(e^{tX}) = \exp(t\mu + t^2\sigma^2/2)$

Characteristic function = $E(e^{itX}) = \exp(it\mu - t^2\sigma^2/2)$

Cumulant-generating function = $\log E(e^{itX}) = it\mu - t^2\sigma^2/2$

2.3 DISTRIBUTIONS RELATED TO $\Phi(x)$

[*2.3.1*] Let X be a N(0,1) random variable. Then X^2 has a chi-square distribution with one degree of freedom (Mood et al., 1974, p. 243), and pdf $(2\pi)^{-1/2}y^{-1/2}e^{-(-1/2)y}$ $(y > 0)$.

[*2.3.2*] *Lognormal Distribution*

(a) Let X be a $N(\mu,\sigma^2)$ random variable. Then the random variable $Y = \exp(X)$ has a lognormal distribution with pdf (Mood et al., 1974, p. 117)

$$g(y;\mu,\sigma^2) = \frac{1}{\sigma y\sqrt{2\pi}} \exp\left\{-\frac{(\log y - \mu)^2}{2\sigma^2}\right\}, \qquad y > 0,\ -\infty < \mu < \infty,\ \sigma > 0$$

(b) (i) Let $G(y;\mu,\sigma^2)$ be the cdf corresponding to the pdf $g(y;\mu,\sigma^2)$ of (a) and $F(x;\mu,\sigma^2)$ be the cdf of the rv X of (a). Then

$$G(y;\mu,\sigma^2) = F(\log y;\mu,\sigma^2)$$

$$F(x;\mu,\sigma^2) = G(e^x;\mu,\sigma^2)$$

(ii) Let z_p and y_p be the quantiles of order p of a N(0,1) rv and $g(y;\mu,\sigma^2)$, respectively, so that $\Phi(z_p) = \Pr(Y \le y_p) = 1 - p$; then (Aitchison and Brown, 1957, p. 9)

$$y_p = \exp(\mu + z_p\sigma)$$

(iii) The mean and variance of $g(y;\mu,\sigma^2)$ are $\exp(\mu + \sigma^2/2)$ and $\exp(2\mu + \sigma^2)\{\exp(\sigma^2) - 1\}$, respectively.

For further information about the lognormal distribution, see Aitchison and Brown (1957) or Johnson and Kotz (1970, chap. 14).

[*2.3.3*] Let Y be a rv with pdf $g(y;\lambda,\mu)$ given by

$$g(y;\lambda,\eta) = \{\lambda/(2\pi y^3)\}^{1/2}\exp[-\lambda(y - \eta)^2/(2\eta^2 y)],$$
$$y > 0,\ \lambda > 0,\ \eta > 0$$

Then Y has a standard inverse Gaussian or Wald distribution. It is the distribution of the first passage time of a particle to reach a single absorbing barrier at a distance $\sqrt{\lambda}$ from the origin, when the particle is moving under a Wiener process with drift $\sqrt{\lambda/\eta}$ and variance parameter 1; see [11.1.9] and Johnson and Kotz (1970, chap. 15).

(a) The mean and variance of Y are η and η^3/λ, respectively.

(b) The rv $\lambda(Y - \eta)^2/(\eta^2 Y)$ has a chi-square distribution with one degree of freedom (Shuster, 1968, p. 1514).

(c) The cdf $G(y;\lambda,\eta)$ corresponding to $g(y;\lambda,\eta)$ above is given by (Chhikara and Folks, 1974, p. 251)

$$G(y;\lambda,\eta) = \Phi[(\lambda/y)^{1/2}(y/\eta - 1)]$$
$$+ \exp(2\lambda/\eta)\Phi[-(\lambda/y)^{1/2}(1 + y/\eta)]$$

Thus the cdf of the inverse Gaussian distribution can be evaluated from tables of the standard normal cdf Φ; see [3.1]. Shuster (1968) has expressed G in terms of the cdf of a chi-square rv with one degree of freedom.

[*2.3.4*] ***Uniform Distribution***. Let X be a normal random variable with cdf $F(x) = F(x;\mu,\sigma^2)$. Then the random variable $Y = F(X)$ has a uniform distribution with pdf

$$g(y) = 1, \qquad 0 < y < 1$$

2.4 MEMBER OF SOME WELL-KNOWN FAMILIES OF DISTRIBUTION

The normal distribution is a member of some well-known families of distributions such as the exponential, Pearson, stable, and infinitely divisible families. Hence the general properties of these families also hold for the normal distribution.

[*2.4.1*] Exponential Family (Lehmann, 1959, pp. 50-54; Ferguson, 1967, pp. 125-131; Mood et al., 1974, pp. 312-314; Roussas, 1973, pp. 217-223; Bickel and Doksum, 1977, pp. 67-73).

[*2.4.1.1*] The pdf $g(x;\theta)$ of a *one-parameter exponential family* member, defined for θ in an interval I, is of the form

$$g(x;\theta) = \begin{cases} \exp[Q(\theta)T(x) + S(x) + h(\theta)], & a < x < b \\ 0, & \text{elsewhere} \end{cases}$$

If, further, neither a nor b depends upon θ, Q is a nontrivial continuous function of θ, T is differentiable with nonzero derivative almost surely if $a < x < b$, and S is continuous in x if $a < x < b$, then g represents a regular case of the exponential family (Hogg and Craig, 1978, p. 357).

(a) The normal pdf with $\theta = \mu$ and σ^2 known is a regular member of this family with

$$Q(\theta) = \theta/\sigma^2 \qquad T(x) = x \qquad S(x) = -x^2/(2\sigma^2) - \frac{1}{2}\log(2\pi\sigma^2)$$
$$h(\theta) = -\theta^2/2\sigma^2, \qquad a = -\infty,\ b = +\infty$$

(b) The normal pdf with $\theta = \sigma^2$ and μ known is a regular member of this family with

$$Q(\theta) = -1/(2\theta) \qquad T(x) = (x - \mu)^2 \qquad S(x) = 0$$
$$h(\theta) = -\frac{1}{2}\log(2\pi\theta), \qquad a = -\infty,\ b = +\infty$$

[*2.4.1.2*] The pdf $g(x;\theta_1,\theta_2)$ of a two-parameter exponential family member, defined for θ_j in an interval I_j ($j = 1, 2$), is of the form

$$g(x;\theta_1,\theta_2) = \exp\left[\sum_{i=1}^{2} Q_i(\theta_1,\theta_2)T_i(x) + S(x) + h(\theta_1,\theta_2)\right], \qquad a < x < b$$

Further, g represents a *regular case* if neither a nor b depends on θ_1 or on θ_2, the functions $Q_i(\theta_1,\theta_2)$ are nontrivial, functionally independent, and continuous in θ_1 and in θ_2 ($\theta_j \in I_j$; $j = 1, 2$), the functions $T_i(x)$ are differentiable in x, with continuous derivatives $T_i(x)$ for $a < x < b$, no one of the $T_i(x)$ is a linear combination of the others, and S is continuous in x ($a < x < b$) (Hogg and Craig, 1978, p. 366).

The normal pdf with $\theta_1 = \mu$ and $\theta_2 = \sigma^2$ is a regular member of this family, where

$$Q_1(\theta_1,\theta_2) = -1/(2\theta_2) \qquad Q_2(\theta_1,\theta_2) = \theta_1/\theta_2$$

$$T_1(x) = x^2 \qquad T_2(x) = x \qquad S(x) = 0$$

$$h(\theta_1,\theta_2) = -\theta_1^2/(2\theta_2) - \frac{1}{2}\log(2\pi\theta_2), \qquad a = -\infty,\ b = +\infty$$

[*2.4.2*] *Linear Exponential-type Distributions* (Blackwell and Girshick, 1954, pp. 179-180; Patil, 1963, p. 205; Patil and Shorrock, 1965, pp. 94-99). The pdf $g(x;\theta)$ of a *linear exponential-type distribution* is of the form

$$g(x;\theta) = a(x)\ \exp(\theta x)/g(\theta)$$

where $a(x)$ is a nonnegative function which depends only on x, $\theta \in I$, and $g(\theta)$ is finite and differentiable (Patil, 1963).

The normal pdf with $\theta = \mu/\sigma^2$, σ^2 known, is of linear exponential type:

$$a(x) = \exp\{-x^2/(2\sigma^2)\} \qquad g(\theta) = (\sigma\sqrt{2\pi})^{-1}\exp(-\theta^2\sigma^2/2) \quad .$$

The normal pdf with $\theta = \sigma^2$ and μ known is not of linear exponential type.

[*2.4.3*] *The Pearson System* (Pearson, 1895; Johnson and Kotz, 1970, pp. 9-15; Kendall and Stuart, 1977, pp. 159-166; Ord, 1972, chap. 1). The pdf $g(x)$ of a member of the *Pearson system* satisfies the differential equation

$$\frac{d}{dx}\{g(x)\} = \frac{(x - a)g(x)}{b_0 + b_1x + b_2x^2}$$

The normal pdf is a member of this system with

$$a = \mu,\ b_0 = -\sigma^2,\ b_1 = 0,\ b_2 = 0$$

[2.4.4] *Monotone Likelihood Ratio (MLR) Distributions* (Lehmann, 1959, p. 68; Ferguson, 1967, pp. 208-210; Mood et al., 1974, p. 423). In what follows, $L(x_1, x_2, \ldots, x_n; \theta)$ denotes the likelihood function or joint pdf

$$\prod_{i=1}^{n} g(x_i; \theta)$$

of a random sample of size n from a distribution with pdf $g(x;\theta)$.

The pdf $g(x;\theta)$ for θ in an interval I is said to have a MLR in a statistic $T(x_1, x_2, \ldots, x_n)$ if the likelihood ratio

$$\frac{L(x_1, x_2, \ldots, x_n; \theta_1)}{L(x_1, x_2, \ldots, x_n; \theta_2)}$$

is either a nonincreasing function of T for every $\theta_1 < \theta_2$ or a nondecreasing function of T for every $\theta_1 < \theta_2$; $n = 1, 2, \ldots$.

If $n = 1$, the normal pdf with $\theta = \mu$ and σ^2 known has MLR in $T(x) = x$. The normal pdf with $\theta = \sigma^2$ and μ known has MLR in $T(x) = (x - \mu)^2$, but not in $T(x) = x$.

[2.4.5] *Stable Distributions* (Feller, 1966, pp. 165-173; Holt and Crow, 1973, pp. 143-198; Lukacs, 1970, pp. 128-161). Let X, X_1, ..., X_n be independently and identically distributed (iid) random variables and let $S_n = \Sigma_{i=1}^n X_i$. The distribution of X is *stable* if for each n there exist constants $c_n > 0$, and γ_n such that S_n has the same distribution as $c_n X + \gamma_n$ and the distribution of X is not concentrated at the origin. The distribution of X is *strictly stable* if $\gamma_n = 0$. Stable distributions are absolutely continuous and unimodal (Lukacs, 1970, pp. 138, 158).

The normal distribution is a stable distribution with $c_n = \sqrt{n}$ and $\gamma_n = \mu\sqrt{n}(\sqrt{n} - 1)$. It is strictly stable when $\mu = 0$.

[2.4.6] *Infinitely Divisible Distributions* (Feller, 1966, pp. 173-176; Fisz, 1962; Lukacs, 1970, chap. 5). A distribution G is an *infinitely divisible distribution* (IDD) if for each n it can be represented as the distribution of the sum $S_n = \Sigma_{i=1}^n X_i$ of n iid random variables with common distribution G_n; all stable distributions are IDD (Feller, 1966, p. 173).

The normal distribution $F(\cdot;\mu,\sigma^2)$ is IDD with G_n as the common $N(\mu/n,\sigma^2/n)$ distribution.

[*2.4.7*] *Unimodal Distributions* (Lukacs, 1970, pp. 91-99). A distribution G(x) is said to be *unimodal* if there exists at least one value a such that G(x) is convex for x < a and concave for x > a. The point a is called a *mode* or *vertex* of G(x) (Lukacs, 1970, p. 91). All stable distributions are unimodal (Lukacs, 1970, p. 91). An equivalent definition of unimodality is that "the density function g (if it exists) is such that $\{x \mid g(x) \ge c\}$ is a convex set for all $c \ge 0$" (Anderson, 1955, pp. 170-171).

The normal distribution $F(x;\mu,\sigma^2)$ is unimodal with mode at μ.

[*2.4.8*] *Polya-type Distributions* (Karlin, 1957, p. 282; Johnson and Kotz, 1970, p. 32). A family of distributions with pdfs $f(x_i,\theta)$ which are continuous in the real variables θ and x is said to be *of Polya type n* if for all $x_1 < x_2 < \cdots < x_m$ and $\theta_1 < \theta_2 < \cdots < \theta_m$ and for every $1 \le m \le n$

$$\begin{vmatrix} f(x_1;\theta_1) & f(x_1;\theta_2) & \cdots & f(x_1;\theta_m) \\ f(x_2;\theta_1) & f(x_2;\theta_2) & \cdots & f(x_2;\theta_m) \\ \cdots & & & \\ f(x_m;\theta_1) & f(x_m;\theta_2) & \cdots & f(x_m;\theta_m) \end{vmatrix} \ge 0$$

and *strictly of Polya type n* if strict inequality holds in the above determinant. If the family of distributions is of Polya type n for every n, then the family is said to be of *Polya type* ∞.

The normal distribution with θ = μ or with $\theta = \sigma^2$ is strictly of Polya type ∞ (Karlin, 1957, p. 282).

[*2.4.9*] *Symmetric Power Distributions* (Box, 1953, p. 467; Box and Tiao, 1964, p. 170). A distribution is a *symmetric power distribution* if it has a pdf given by

$$g(x;\mu,\sigma,\beta) = k \exp\{-\tfrac{1}{2}|(x - \mu)/\sigma|^{2/(1+\beta)}\}, \qquad -\infty < x < \infty,$$
$$|\mu| < \infty,\ 0 < \sigma < \infty,\ |\beta| < 1$$
$$k = 1/[\Gamma\{1 + (1 + \beta)/2\}2^{\{1+(1+\beta)/2\}}\sigma]$$

The $N(\mu,\sigma^2)$ distribution results when $\beta = 0$. If $\beta < 0$, the distribution is platykurtic ($\kappa_2 < 3$) with the uniform distribution in the limit as $\beta \to -1$. If $\beta > 0$, the distribution is leptokurtic ($\kappa_2 > 3$), with the double exponential distribution when $\beta = 1$. This family has been studied in connection with some tests of normality, and of robustness.

2.5 COMPOUND AND MIXED NORMAL DISTRIBUTION

The term *mixture* of normal distributions has been used in two different senses by different writers. In the first sense, outlined in [2.5.1], a mixture is formed when one or both parameters of the normal distribution are continuous random variables. In the more commonly occurring sense, compound or mixed distributions result from the "mixing" of two or more component distributions ([2.5.2]). Studies of compound normal distributions in this sense go back to Karl Pearson (1894); the pair (μ,σ) may be thought of as being sampled from a multinomial distribution.

[*2.5.1*] If the conditional pdf of a rv X, given one or both parameters μ and σ^2, is normal, then the marginal pdf of X, if this exists, is a *mixture of normal distributions*.

[*2.5.1.1*] *Mean μ as a rv*. We suppose throughout that the conditional pdf of X, given μ, is $N(\mu,\sigma^2)$.

(a) If the marginal pdf of μ is $N(\theta,\sigma_1^2)$, then the marginal pdf of X is $N(\theta, \sigma^2 + \sigma_1^2)$.

(b) Suppose that the marginal pdf of μ is uniform, given by

$$g(\mu) = 1/r, \qquad a_1 < \mu < a_2,\ a_2 - a_1 = r$$

Then the marginal pdf and cdf of X are, respectively,

$$h(x) = \{\Phi(t_1) - \Phi(t_2)\}/r$$
$$H(x) = \sigma\{t_1\Phi(t_1) - t_2\Phi(t_2) + \phi(t_1) - \phi(t_2)\}/r$$

where $t_i = (x - a_i)/\sigma$, i = 1, 2. See Bhattacharjee et al. (1963, pp. 404-405), who also derive the cdf of the sum of two such independent rvs.

[*2.5.1.2*] <u>*Variance* σ^2 *as a rv*</u>. Suppose that, given σ^2, X has a $N(0,\sigma^2)$ distribution, and that the pdf of σ^2 is given by

$$g(\lambda) = e^{-\alpha\lambda}\alpha^m\lambda^{m-1}/\Gamma(m), \qquad \lambda > 0,\ \alpha > 0,\ m > 0$$

which is a gamma distribution. Then X has marginal pdf h(x), where

$$h(x) = \sqrt{\alpha}(x\sqrt{2\alpha})^{m-(1/2)}k_{m-(1/2)}(x\sqrt{2\alpha})/\{\sqrt{\pi}2^{m-1}\Gamma(m)\}$$

and $k(\cdot)$ is a modified Hankel function (Teichroew, 1957, p. 511; Abramowitz and Stegun, 1964, pp. 358, 510). When m is an integer,

$$h(x) = \frac{\sqrt{2\alpha}}{(m-1)!}\,\frac{e^{-\alpha x\sqrt{2}}}{2^{2m-1}}\left\{\sum_{v=0}^{m-1}\frac{(2m-2v)!(2\alpha x\sqrt{2})^v}{v!(m-v-1)!}\right\}$$

Teichroew (1957, p. 511) gives an expression for the cdf H(x), for any m > 0. The characteristic function of X is $\{1 + t^2/(2\alpha)\}^{-m}$, from which the following moments are found:

$$E(X^{2k-1}) = 0, \qquad k = 1, 2, 3, \ldots$$

$$\text{Var}(X) = m/\alpha, \qquad \mu_4(X) = 3m(m+1)/\alpha^2$$

The kurtosis $\mu_4(X)/\{\text{Var}(X)\}^2 = 3(1 + m^{-1}), > \mu_4(Z)/\{\text{Var}(Z)\}^2 = 3$, where Z is a N(0,1) rv (Teichroew, 1957, p. 512).

[*2.5.1.3*] If, given μ, X has a $N(\mu,\sigma^2)$ distribution, and the marginal distribution of σ^{-2} is gamma, then the mixture with X has marginally a Pearson Type VII distribution (Johnson and Kotz, 1970, p. 88).

[*2.5.2*] If a rv X comes with probability p_i (i = 1,2,...,k) from a $N(\mu_i,\sigma_i^2)$ distribution, where $p_1 + p_2 + \cdots + p_k = 1$, then X has a *compound* or *mixed normal distribution* with pdf g(x) given by

$$g(x) = \sum_{i=1}^{k} p_i\phi[(x - \mu_i)/\sigma_i], \qquad -\infty < x < \infty$$

In what follows, we restrict attention to the case $k = 2$. We may write $p_1 = p$, $p_2 = 1 - p$, and $g(x;p) = g(x)$. Intuitively, $g(x;p)$ may be bimodal, or it may be unimodal. Some of the research into mixtures has addressed the issue of modality, as several of the following results indicate.

[*2.5.2.1*] *Moments of the Mixture of Two Normal Distributions.* Let $\theta_i' = E(X^i)$, and $\theta_i = E[X - \theta_1']^i$, $m_i = \mu_i - \theta_1'$, $i = 1, 2, \ldots$. The first five moments are as follows (Cohen, 1967, p. 16; Johnson and Kotz, 1970, p. 89):

$$\theta_1' = \sum_{i=1}^{2} p_i\mu_i, \qquad \theta_1 = 0$$

$$\theta_2' = \sum_{i=1}^{2} p_i(\mu_i^2 + \sigma_i^2), \qquad \theta_2 = \sum_{i=1}^{2} p_i(m_i^2 + \sigma_i^2)$$

$$\theta_3' = \sum_{i=1}^{2} p_i\mu_i(\mu_i^2 + 3\sigma_i^2), \qquad \theta_3 = \sum_{i=1}^{2} p_i m_i(m_i^2 + 3\sigma_i^2)$$

$$\theta_4' = \sum_{i=1}^{2} p_i(\mu_i^4 + 6\mu_i^2\sigma_i^2 + 3\sigma_i^4), \qquad \theta_4 = \sum_{i=1}^{2} p_i(m_i^4 + 6m_i^2\sigma_i^2 + 3\sigma_i^4)$$

$$\theta_5' = \sum_{i=1}^{2} p_i(\mu_i^5 + 10\mu_i^3\sigma_i^2 + 15\mu_i\sigma_i^4), \qquad \theta_5 = \sum_{i=1}^{2} p_i(m_i^5 + 10m_i^3\sigma_i^2 + 15m_i\sigma_i^4)$$

[*2.5.2.2*] The mixed normal distribution $g(x;p)$ is *symmetrical* in two cases:

(a) $p = \frac{1}{2}$ and $\sigma_1 = \sigma_2 = \sigma$

or

(b) $\mu_1 = \mu_2$

In case (b) the distribution is always unimodal; in case (a) $g(x;p)$ is unimodal if $|\mu_1 - \mu_2| \le 2\sigma$, and bimodal if $|\mu_1 - \mu_2| > 2\sigma$ (Cohen, 1967, pp. 23-24).

[*2.5.2.3*] A sufficient condition that $g(x;p)$ be unimodal *for all p*, $0 < p < 1$, is that (Eisenberger, 1964, p. 359)

$$(\mu_1 - \mu_2)^2 \le 27\sigma_1^2\sigma_2^2/\{4(\sigma_1^2 + \sigma_2^2)\}$$

A sharper inequality is that (Behboodian, 1970, p. 138)

$$|\mu_1 - \mu_2| \le 2 \min(\sigma_1, \sigma_2)$$

[*2.5.2.4*] A sufficient condition for g(x;p) to be unimodal when $\sigma_1 = \sigma_2 = \sigma$ and *for a given value of p* is that

$$|\mu_1 - \mu_2| \le 2\sigma\sqrt{(1 + |\log p - \log(1 - p)|/2)}$$

This condition is also necessary when p = 1/2 (Behboodian, 1970, pp. 138-139).

[*2.5.2.5*] A sufficient condition that *there exist values of p*, $0 < p < 1$, for which g(x;p) is bimodal, is that (Eisenberger, 1964, pp. 359-360)

$$(\mu_1 - \mu_2)^2 > 8\sigma_1^2\sigma_2^2/(\sigma_1^2 + \sigma_2^2)$$

[*2.5.2.6*] Behboodian (1970, pp. 135-138) gives an iterative procedure for finding the mode or modes of g(x;p), and a brief table.

2.6 FOLDED AND TRUNCATED NORMAL DISTRIBUTIONS: MILLS' RATIO

[*2.6.1*] Most of what follows is discussed by Elandt (1961, pp. 551-554). Let X be a $N(\mu,\sigma^2)$ random variable, and let $Y = |X|$. Then Y has a *folded normal distribution* with pdf g(y), where

$$g(y) = \begin{cases} (\sqrt{2\pi}\sigma)^{-1}[\exp\{-(y - \mu)^2/(2\sigma^2)\} \\ \qquad + \exp\{-(y + \mu)^2/(2\sigma^2)\}], & y \ge 0 \\ 0, & y < 0 \end{cases}$$

where $|\mu| < \infty$, $\sigma > 0$. If $y \ge 0$,

$$g(y) = \sqrt{2}(\sigma\sqrt{\pi})^{-1}\cosh(\mu y/\sigma^2)\exp\{-(y^2 + \mu^2)/(2\sigma^2)\}$$

[*2.6.1.1*] *Relation with Noncentral Chi-square Distribution* (see [5.3.6]). If a rv U has a noncentral chi-square distribution with one degree of freedom and noncentrality parameter μ^2/σ^2, then the distribution of the rv $\sigma\sqrt{U}$ is given by the pdf g.

[*2.6.1.2*] *Moments of the Folded Normal Distribution*

$$E(Y^r) = \sigma^r \sum_{j=0}^{r} \binom{r}{j} \theta^{r-j}[I_j(-\theta) + (-1)^{r-j} I_j(\theta)], \qquad r = 1, 2, \ldots$$

where $\theta = \mu/\sigma$ and $I_r(a)$ is as defined in [2.3.4]. The mean μ_f and variance σ_f^2 of Y are given by

$$\mu_f = \sigma\sqrt{2/\pi}\ \exp(-\theta^2/2) - \mu[1 - 2\Phi(\theta)]$$

$$\sigma_f^2 = \mu^2 + \sigma^2 - \mu_f^2$$

[*2.6.1.3*] *Half-Normal Distribution*. If $\mu = 0$, the folded normal distribution becomes a *half-normal distribution* with a simplified pdf given by

$$g(y) = \sqrt{2}(\sigma\sqrt{\pi})^{-1} \exp\{-y^2/(2\sigma^2)\}, \qquad y \geq 0,\ \sigma > 0$$

and $E(Y^r) = 2\sigma^r I_r(0)$; $r = 1, 2, \ldots$. The mean and variance simplify to

$$\mu_f = \sigma\sqrt{2/\pi}, \qquad \sigma_f^2 = \sigma^2(1 - 2/\pi)$$

[*2.6.1.4*] The family of folded normal distributions is included between two extremes, the half-normal, for which $\mu_f/\sigma_f = 1.3237$, and the normal, for which μ_f/σ_f is infinite. Approximate normality is attained if $\mu_f/\sigma_f > 3$ (Elandt, 1961, p. 554).

[*2.6.1.5*] Tables of the folded normal cdf G(y) for $\mu_f/\sigma_f = 1.3236$, 1.4(0.1)3, and y = 0.1(0.1)7 are given by Leone et al. (1961, pp. 546-549). They have also given values of μ and σ_f for μ_f/σ_f = 1.3236, 1.325, 1.33(0.01)1.50(0.02)1.70(0.05)2.40(0.10)3.00, on p. 545.

[*2.6.2*] Let X be a $N(\mu,\sigma^2)$ random variable. The conditional distribution of X, given that $a < X < b$, is a *doubly truncated normal distribution* with pdf g(y), where

$$g(y) = \begin{cases} \dfrac{1}{\sigma}\,\phi\left(\dfrac{y-\mu}{\sigma}\right) \Big/ \left[\Phi\left(\dfrac{b-\mu}{\sigma}\right) - \Phi\left(\dfrac{a-\mu}{\sigma}\right)\right], & a < y < b \\ 0, & \text{otherwise} \end{cases}$$

If $a = -\infty$, the distribution is *singly truncated from above*, and is *singly truncated from below* if $b = \infty$.

(a) When $a = 0$, $b = \infty$, and $\mu = 0$, the pdf is that of the half-normal distribution discussed in [2.6.1.3].

(b) Let $h = (a - \mu)/\sigma$, $k = (b - \mu)/\sigma$. Then the mean and variance μ_t and σ_t^2 of the doubly truncated normal distribution are given by (Johnson and Kotz, 1970, pp. 81, 83; see [2.6.3.2])

$$\mu_t = \mu + \sigma\{\phi(h) - \phi(k)\}/[\Phi(k) - \Phi(h)]$$

$$\sigma_t^2 = \sigma^2[1 + \{h\phi(h) - k\phi(k)\}/\{\Phi(k) - \Phi(h)\}] - (\mu_t - \mu)^2$$

(c) Shah and Jaiswal (1966, pp. 107-111) give the pdf and moments of g(y) shifted so that the origin is at a. Johnson and Kotz (1970, p. 84) give a table of μ_t, σ_t, and the ratio (mean deviation)/σ_t for selected values of $\Phi(h)$ and of $1 - \Phi(k)$.

[*2.6.3*] Consider a continuous rv Y with cdf G(y) and pdf $g(\cdot)$ such that $G(0) = 0$. The conditional pdf of Y, given that $Y > y$, is h(y), where

$$h(y) = g(y)/[1 - G(y)], \qquad y > 0$$

which is the *hazard rate* or *failure rate*. The name comes from the use of this conditional pdf to describe survival time distributions, given that an individual survives to time y. Distributions have been classified into increasing failure rate (IFR) and decreasing failure rate (DFR) families, for which the failure rate $h(\cdot)$ is increasing and decreasing, respectively, for all $y > 0$.

[*2.6.3.1*] The half-normal distribution (see [2.6.1.3]) is IFR (Patel, 1973, pp. 281-284) with failure rate

$$(1/\sigma)\phi(x/\sigma)/[1 - \Phi(x/\sigma)], \qquad x > 0$$

If $\sigma = 1$, the reciprocal of the failure rate, $[1 - \Phi(x)]/\phi(x)$, is known as *Mills' ratio*, $R(x)$.

[*2.6.3.2*] Let Y be a $N(\mu,\sigma^2)$ rv, singly truncated from below (see [2.4.2]) at a. If $u = (a - \mu)/\sigma$, then the mean and variance of Y are given by

$$\mu_t = \mu + \sigma/R(u)$$

$$\sigma_t^2 = \sigma^2[1 + h/R(h) - \{R(h)\}^{-2}]$$

See also [3.1.1(c)].

[*2.6.3.3*] Since $\Phi(x) = 1 - \phi(x)R(x)$, the cdf of a N(0,1) rv may be obtained from tables of Mills' ratio, and vice versa. See Table 2.1; the most extensive tables of R(x) are those of Sheppard (1939), to 12 decimal places for x = 0.00(0.01)9.50 and 24 places for x = 0(0.1)10. Approximations and inequalities for R(x) are presented in Sections 3.3 to 3.7.

2.7 "NORMAL" DISTRIBUTIONS ON THE CIRCLE

In an attempt to find distributions which have the properties of normal rvs for a random angle θ, the study of directional data has developed. One can think of a point on the circumference of a circle of fixed radius as being the realization of a random direction. Two distributions (the von Mises and wrapped normal) come closer to direct analogy with the normal distribution than any others, and we introduce them briefly here. The best source for study of this subject is Mardia (1972, chap. 3; see pp. 68-69).

[*2.7.1*] *The von Mises Distribution*. Suppose that a direction is denoted by a point P on a unit circle, center O, or more precisely, by the vector OP, and let Θ be the angular polar coordinate of P. The von Mises distribution is often used to describe direction; it plays a prominent role in statistical inference on the circle, and its importance there is almost the same as that of the normal distribution on the line (Mardia, 1972, p. 58).

A rv Θ is said to have a *von Mises distribution* if its pdf is given by

$$g(\theta) = \{2\pi I_0(\kappa)\}^{-1} \exp[\kappa \cos(\theta - \mu_0)], \quad 0 < \theta \le 2\pi,\ \kappa > 0,\quad 0 \le \mu_0 < 2\pi$$

where $I_0(\kappa)$ is the modified Bessel function of the first kind and order zero; i.e.,

$$I_0(\kappa) = \sum_{r=0}^{\infty} \frac{\left(\frac{1}{2}\kappa\right)^{2r}}{(r!)^2}$$

The distribution $g(\theta)$ can also be defined over the interval $-\pi < \theta < \pi$ (Stephens, 1969, p. 149).

Some properties of this distribution are discussed by Gumbel et al. (1953) and its applications by Gumbel (1954). We present a few properties here.

[*2.7.1.1*] The distribution is unimodal and symmetric about $\theta = \mu_0$. The mode and the mean direction are at μ_0, and the anti-mode at $\theta - \mu_0 + \pi$ (Mardia, 1972, pp. 57-58; Gumbel et al., 1953, p. 138).

[*2.7.1.2*] The curve $g(\theta)$ has two points of inflection, at $\pm\arccos\{-1/(2\kappa) + \sqrt{1 + 1/(4\kappa^2)}\}$ (Mardia, 1972, p. 58; Gumbel et al., 1953, p. 138). The shape of the curve depends on κ, which is a concentration parameter. When $\kappa = 0$, the distribution $g(\theta)$ reduces to the uniform distribution on $(0,2\pi)$. For small values of κ, it approximates to a cardioid distribution, with pdf given by

$$g_1(\theta) = (2\pi)^{-1}\left\{1 + \frac{1}{2}\kappa \cos(\theta - \mu_0)\right\}, \quad 0 < \theta \le 2\pi,\ 0 \le \mu_0 < 2\pi$$

For large values of κ, the rv Θ is approximately distributed as a $N(\mu_0, 1/\kappa)$ rv (Mardia, 1972, p. 60; Gumbel et al., 1953, pp. 138-139).

[*2.7.1.3*] The von Mises distribution is related to the bivariate normal distribution (Chapter 9) as follows: let X be $N(\cos\mu_0, 1/\kappa)$ and Y be $N(\sin\mu_0, 1/\kappa)$, and suppose X and Y are independent. Let $X = R \cos \Theta$, and $Y = R \sin \Theta$. Then the conditional distribution of Θ for $r = 1$ is the von Mises distribution $g(\theta)$ (Mardia, 1972, pp. 60-61).

[*2.7.1.4*] Mardia (1972, pp. 63-64) gives Fourier and integral expansions for $g(\theta)$ and for the cdf of the distribution. Tables of values of the cdf are found in Gumbel et al. (1953, pp. 147-150), Batschelet (1965), and Mardia (1972, pp. 290-296).

[*2.7.1.5*] The von Mises family is not additive. The distribution of the sum of two iid rvs each having the von Mises distribution is given by Mardia (1972, p. 67).

[*2.7.2*] *The Wrapped Normal Distribution*. Any given distribution on the line can be wrapped around the circumference of a circle of unit radius. If X is a rv on the line with pdf $g(x)$, the rv X_W of the wrapped distribution is given by $X_W = X(\text{mod } 2\pi)$ and its pdf by (Mardia, 1972, p. 53)

$$h(\theta) = \sum_{p=-\infty}^{\infty} g(\theta + 2\pi p)$$

Let X be $N(0,\sigma^2)$. Then the pdf of the derived wrapped normal distribution is $h(\theta)$, where

$$\begin{aligned} h(\theta) &= (\sigma\sqrt{2\pi})^{-1} \sum_{p=-\infty}^{\infty} \exp\left\{\frac{-(\theta + 2\pi p)^2}{(2\sigma^2)}\right\}, \qquad 0 < \theta \le 2\pi \\ &= (2\pi)^{-1}\left\{1 + 2 \sum_{p=1}^{\infty} \rho^{p^2} \cos(p\theta)\right\}, \qquad 0 < \theta \le 2\pi,\ 0 \le \rho \le 1 \end{aligned}$$

where $\rho = \exp(-\sigma^2/2)$ (Mardia, 1972, p. 55). Some properties and brief remarks follow.

[*2.7.2.1*] The distribution is unimodal and is symmetric about the mean direction at zero. In general, it has two points of inflection (Mardia, 1972, p. 55).

[*2.7.2.2*] As $\rho \to 0$, h tends to the uniform distribution, while as $\rho \to 1$, it is concentrated at one single point (Mardia, 1972, p. 55).

[*2.7.2.3*] The wrapped normal distribution possesses the additive property. Let $\Theta_1, \Theta_2, \ldots, \Theta_n$ be independent rvs and let Θ_i have the pdf $g(\theta)$ with $\rho = \rho_i$. Then the sum of $\Theta_1, \Theta_2, \ldots, \Theta_n$ has the pdf $g(\theta)$ with $\rho = \Pi_{i=1}^n \rho_i$ (Mardia, 1972, p. 56).

[*2.7.2.4*] *Central Limit Theorem.* Let Θ_j, $j = 1, 2, \ldots, n$ be iid rvs with $-\pi < \theta_j \le \pi$ for each j. Suppose also that $E(\Theta) = 0$. Then the distribution of $(\theta_1 + \theta_2 + \cdots + \theta_n)/\sqrt{n}$ (mod 2π) tends as $n \to \infty$ to the wrapped normal distribution with parameter σ^2, where $\sigma^2 = E(\Theta^2)$ (Mardia, 1972, p. 90).

[*2.7.3*] The von Mises and wrapped normal distributions can be made to approximate one another fairly closely. In [2.7.1], let $\mu_0 = 0$, and

$$\rho = \exp(-\sigma^2/2) = I_1(\kappa)/I_0(\kappa)$$

where these parameters appear in [1.8.1] and [1.8.2], and

$$I_1(\kappa) = \sum_{r=0}^{\infty} \frac{(\kappa/2)^{2r+1}}{(r+1)(r!)^2}$$

which is the modified Bessel function of the first kind and the first order. Then there is exact agreement between the two distributions as $\kappa \to 0$ (and $\sigma \to \infty$) or as $\kappa \to \infty$ (and $\sigma \to 0$). Numerical calculations indicate satisfactory agreement for intermediate parameter values (Stephens, 1963, pp. 388-389; Mardia, 1972, pp. 66-67).

REFERENCES

The numbers in square brackets give the sections in which the corresponding reference is cited.

Abramowitz, M., and Stegun, I. A. (eds.) (1964). *Handbook of Mathematical Functions*, Washington, D.C.: National Bureau of Standards. [2.1; 2.1.9, 10; 2.5.1.2]

Aitchison, J., and Brown, J. A. C. (1957). *The Lognormal Distribution*, London: Cambridge University Press. [2.3.2]

Anderson, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities, *Proceedings of the American Mathematical Society* 6, 170-176. [2.4.7]

Bain, L. J. (1969). Moments of a non-central t and non-central F distribution, *The American Statistician* 23(4), 33-34. [2.2.3]

Batschelet, E. (1965). Statistical methods for the analysis of problems in animal orientation and certain biological rhythms, American Institute of Biological Sciences, Washington, D.C. [2.7.1.4]

Behboodian, J. (1970). On the modes of a mixture of two normal distributions, *Technometrics* 12, 131-139. [2.5.2.3, 6]

Bhattacharjee, G. P., Pandit, S. N. N., and Mohan, R. (1963). Dimensional chains involving rectangular and normal error-distributions, *Technometrics* 5, 404-406. [2.5.1.1]

Bickel, P. J., and Doksum, K. A. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics*, San Francisco: Holden-Day. [2.4.1]

Blackwell, D., and Girshick, M. A. (1954). *Theory of Games and Statistical Decisions*, New York: Wiley. [2.4.2]

Box, G. E. P. (1953). A note on regions for tests of kurtosis, *Biometrika* 40, 465-468. [2.4.9]

Box, G. E. P., and Tiao, G. C. (1964). A note on criterion robustness and inference robustness, *Biometrika* 51, 169-173. [2.4.9]

Chhikara, R. J., and Folks, J. L. (1974). Estimation of the inverse Gaussian distribution, *Journal of the American Statistical Association* 69, 250-254. [2.3.3]

Cohen, A. C. (1967). Estimation in mixtures of two normal distributions, *Technometrics* 9, 15-28. [2.5.2.1, 2]

Draper, N. R., and Tierney, D. E. (1973). Exact formulas for additional terms in some important series expansions, *Communications in Statistics* 1, 495-524. [2.1.9]

Eisenberger, I. (1964). Genesis of bimodal distributions, *Technometrics* 6, 357-363. [2.5.2.3, 5]

Elandt, R. C. (1961). The folded normal distribution: Two methods of estimating parameters from moments, *Technometrics* 3, 551-562. [2.2.4; 2.6.1]

Feller, W. (1966). *An Introduction to Probability Theory and its Applications*, Vol. 2, New York: Wiley. [2.4.5, 6]

Ferguson, T. S. (1967). *Mathematical Statistics. A Decision Theoretic Approach*, New York: Academic. [2.4.1, 4]

Fisz, M. (1962). Infinitely divisible distributions: Recent results and applications, *Annals of Mathematical Statistics* 33, 68-84. [2.4.6]

Gumbel, E. J. (1954). Applications of the circular normal distribution, *Journal of the American Statistical Association* 49, 267-297. [2.7.1]

Gumbel, E. J., Greenwood, J. A., and Durand, D. (1953). The circular normal distribution: Theory and tables, *Journal of the American Statistical Association* 48, 131-152. [2.7.1]

Hogg, R. V., and Craig, A. T. (1978). *Introduction to Mathematical Statistics*, New York: Macmillan. [2.4.1.1, 2]

Holt, D. R., and Crow, E. L. (1973). Tables and graphs of stable probability density functions, *Journal of Research* (National Bureau of Standards) 77B, 143-198. [2.4.5]

Johnson, N. L., and Kotz, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions*, Vol. 1, New York: Wiley. [2.3.2, 3; 2.4.3, 8; 2.5.1.3; 2.5.2.1; 2.6.2]

Kamat, A. R. (1953). Incomplete and absolute moments of the multivariate normal distribution with some applications, *Biometrika* 40, 20-34. [2.2.3]

Karlin, S. (1957). Polya type distributions, II, *Annals of Mathematical Statistics* 28, 281-308. [2.4.8]

Kendall, M. G., and Stuart, A. (1977). *The Advanced Theory of Statistics*, Vol. 1 (4th ed.), New York: Macmillan. [2.1.9; 2.2.1, 3, 5; 2.4.3]

Lehmann, E. L. (1959). *Testing Statistical Hypotheses*, New York: Wiley. [2.4.1, 4]

Leone, F. C., Nelson, L. S., and Nottingham, R. B. (1961). The folded normal distribution, *Technometrics* 3, 543-550. [2.6.1.5]

Lukacs, E. (1970). *Characteristic Functions*, London: Griffin. [2.2.3; 2.4.5, 6, 7]

Mardia, K. V. (1972). *Statistics of Directional Data*, New York: Academic. [2.7.1, 2, 3]

Mood, A. M., Graybill, F. A., and Boes, D. C. (1974). *Introduction to the Theory of Statistics* (3rd ed.), New York: McGraw-Hill. [2.3.1, 2; 2.4.1, 4]

Ord, J. K. (1972). *Families of Frequency Distributions*, New York: Hafner. [2.4.3]

Patel, J. K. (1973). A catalog of failure distributions, *Communications in Statistics* 1, 281-284. [2.6.3.1]

Patil, G. P. (1963). A characterization of the exponential-type distribution, *Biometrika* 50, 205-207. [2.4.2]

Patil, G. P., and Shorrock, R. (1965). On certain properties of the exponential type families, *Journal of the Royal Statistical Society* B27, 94-99. [2.4.2]

Pearson, K. (1894). Contributions to the mathematical theory of evolution, *Philosophical Transactions of the Royal Society* 185, 71-110. [2.5]

Pearson, K. (1895). Contributions to the mathematical theory of evolution, II: Skew variations in homogeneous material, *Philosophical Transactions of the Royal Society of London* A186, 343-414. [2.4.3]

Roussas, G. G. (1973). *A First Course in Mathematical Statistics*, Reading, Mass.: Addison-Wesley. [2.4.1]

Shah, S. M., and Jaiswal, M. C. (1966). Estimation of parameters of doubly truncated normal distribution from first four sample moments, *Annals of the Institute of Statistical Mathematics* 18, 107-111. [2.6.2]

Sheppard, W. F. (1939). *The Probability Integral*, British Association Mathematical Tables, Vol. 7, London: Cambridge University Press. [2.6.3.3]

Sheppard, W. F. (1898). On the application of the theory of error to cases of normal distribution and normal correlation, *Philosophical Transactions of the Royal Society* 192, 101-167. [2.1.11]

Shuster, J. (1968). On the inverse Gaussian distribution function, *Journal of the American Statistical Association* 63, 1514-1516. [2.3.3]

Stephens, M. A. (1963). Random walk on a circle, *Biometrika* 50, 385-390. [2.7.3]

Stephens, M. A. (1969). Tests for the von Mises distribution, *Biometrika* 56, 149-160. [2.7.1]

Teichroew, D. (1957). The mixture of normal distributions with different variances, *Annals of Mathematical Statistics* 28, 510-512. [2.5.1.2]

Tong, Y. L. (1978). An adaptive solution to ranking and selection problems, *Annals of Statistics* 6, 658-672. [2.1.6]

chapter 3

THE NORMAL DISTRIBUTION: TABLES, EXPANSIONS, AND ALGORITHMS

The standard normal cdf $\Phi(x)$ and related functions, such as the quantiles and Mills' ratio, can be expanded in series and in other forms such as continued functions. Such expressions frequently give rise to approximations to $\Phi(x)$ and other functions, by truncating an expansion at a point which provides a suitable degree of accuracy. The main use of such expansions is in providing suitable algorithms for computing numerical values of functions of interest. We have therefore included a discussion of available tables in the same chapter as our listing of expansions and approximations; some of the latter were used in one form or another to compute the values appearing in published tables of the normal distribution.

Where approximations are scrutinized for accuracy, or compared with one another, writers have generally examined the absolute error or absolute relative error; in approximating to the cdf $\Phi(x)$, for example, these would be $|G(x) - \Phi(x)|$ or $|G(x) - \Phi(x)|/\Phi(x)$, respectively, where G(x) is the approximating function. In this chapter, the absolute error will be denoted briefly by |error|, and the absolute relative error by |relative error|.

A large number of sources are included in this chapter, and some of these contain a few errors which appear in later corrigenda. We have not claimed to have detected all such errors, but those made between 1943 and 1969 may be cited in the *Cumulative Index to "Mathematics of Computation,"* vols. 1 to 23. Subsequent errors, regardless of the source, are frequently reported in *Mathematics of Computation*.

3.1 TABLES, NOMOGRAMS, AND ALGORITHMS

[*3.1.1*] (a) Rather than attempt to describe the large choice of tables available of $\Phi(x)$, $\phi(x)$, z_α, derivatives of $\phi(x)$, Mills' ratio $R(x)$, and so on, we shall list a few of the more accessible sources, concentrating on recent sources, and including a cross section of American, British, Russian, Indian, and Japanese tables. Greenwood and Hartley (1962) give a comprehensive list of tables published up to 1958, as do Fletcher et al. (1962a, pp. 318-356). The latter (Fletcher et al., 1962b) include a special discussion (pp. 781-932) of errors appearing in published tables cited, together with corrections. There is an excellent comprehensive discussion in Johnson and Kotz (1970a, pp. 41-45).

(b) Table 3.1 summarizes the information found in seven main sources; some of them can be seen to be very detailed, particularly for the cdf $\Phi(x)$. The most accurate table that we could find of z_α [when $1 - \alpha$, or $\Phi(z_\alpha)$, is given] was that of White (1970, pp. 636-637), giving z_α to 20 decimal places for $1 - \alpha = 0.50(0.005)0.995$ and for $\alpha = 5 \times 10^{-k}$, 2.5×10^{-k}, and 10^{-k}; $k = 1(1)20$; see also Kelley (1948), with z_α to 8 places.

In order to use these tables, which relate to the upper 50 percent of the normal area only, results in [2.1.4] should be noted.

(c) Clark (1957, pp. 527-536) gives a table to four decimal places of values of the mean μ_{ab} and standard deviation σ_{ab} of a standard normal distribution truncated at a and b ($a < b$), for $a = -3.0(0.25)0.50$ and $b = 0.0(0.25)3.0$. Note that $\mu_{-b,-a} = \mu_{ab}$ and $\sigma_{-b,-a} = \sigma_{ab}$; see [2.6.2] and [2.6.3.2].

(d) It is of historical interest to note the list of some of the earliest known tables relating to the normal distribution in Table 1.1. Although these were prone to errors, we must count it remarkable that such degrees of accuracy were obtained without the help of computers and, for the most part, without the use of any but the most primitive hand calculators.

TABLE 3.1 Coverage in Some Standard Sources of Tables of the Unit Normal cdf Φ, pdf ϕ, Quantiles z_α, $[\Phi(z_\alpha) = 1 - \alpha]$, Derivatives of ϕ, and Mills' Ratio $R(x)$, $[R(x) = \{1 - \Phi(x)\}/\phi(x)]$

Source	Function	Coverage	Decimal places	Significant figures
Abramowitz and Stegun (1964, pp. 966-977)[a]	$\Phi(x)$, $\phi(x)$, $\phi'(x)$	$x = 0.0(0.02)3.00$	15	
	$\Phi(x)$	$x = 3.0(0.05)5.0$	10	
	$\phi(x)$	$x = 3.0(0.05)5.0$	9	
	$\phi'(x)$	$x = 3.0(0.05)5.0$	7	
	$\phi^{(r)}(x)$, $r = 2, \ldots, 6$	$x = 0.0(0.02)3.00(0.05)5.0$	7-10	
	$-\log_{10}[1 - \Phi(x)]$	$x = 5(1)50(10)100(50)500$	5	
	$\phi^{(r)}(x)$; $r = 7, \ldots, 12$	$x = 0.0(0.1)5.0$	7	
	$\phi(z_\alpha)$, z_α	$1 - \alpha = 0.500(0.001)0.999$	5	
	z_α	$1 - \alpha = 0.975(0.0001)0.9999$	5	
	z_α	$\alpha = 10^{-r}$; $r = 4, \ldots, 23$	5	
Owen (1962, pp. 3-13)	$\Phi(x)$, $\phi(x)$	$x = 0.0(0.01)3.99$	6	
	$1 - \Phi(x)$	$x = 3.0(0.1)6.0(0.2)10(1)20, 25, 30(10)100(25)200(50)500$		5
	z_α, $\phi(z_\alpha)$	$1 - \alpha = 0.50(0.01)0.90(0.005)0.99(0.001)0.999(0.0001)0.9999$, etc. to $1-10^9$	5	
	$\Phi(x)/\phi(x)$	$x = 0(0.01)3.99$	4	
	$R(x)$, $\phi'(x)$, $\phi''(x)$, $\phi'''(x)$	$x = 0.0(0.01)3.99$	5	

TABLE 3.1 (continued)

Source	Function	Coverage	Decimal places	Significant figures
Owen (cont'd)	$R(x)$	$x = 3.0(0.1)6.0(0.2)10(1)20, 25, 30(10)100(25)200(50)500$		5
Pearson and Hartley (1966, tables 1-5)	$\Phi(x)$, $\phi(x)$	$x = 0.0(0.01)4.50$	7	
		$x = 4.50(0.01)6.00$	10	
	$-\log_{10}[1 - \Phi(x)]$	$x = 5(1)50(10)100(50)500$	5	
	z_α	$1 - \alpha = 0.50(0.001)0.999$	4	
		$0.98(0.0001)0.9999$	4	
		$1-10^{-r}$ $(r = 4, \ldots, 9)$	4	
	$\phi(z_\alpha)$	$1 - \alpha = 0.50(0.001)0.999$	5	
Pearson and Hartley (1972, tables 1, 2)	z_α, $\phi(z_\alpha)$	$1 - \alpha = 0.50(0.001)0.999$	10	
	z_α	$1 - \alpha = 0.999(0.0001)0.9999$	8	
	$\phi(z_\alpha)$	$1 - \alpha = 0.999(0.0001)0.9999$	9	
	$\Phi(x) - 0.5$, $\phi(x)$, $\phi'(x)$, $\phi''(x)$	$x = 0.0(0.02)6.20$	6	
	$\phi^{(r)}(x)$; $r = 3, \ldots, 9$	$x = 0.0(0.02)6.20$	5 to 1	
Rao et al. (1966, tables 3.1, 3.2)	$\Phi(x) - 0.5$	$x = 0.00(0.001)3.0(0.01)4(0.1)4.9$	6	
	$\phi(x)$	$x = 0.0(0.01)3.0(0.1)4$	6	

	z_α	$2(1 - \alpha) = 0.01(0.01)0.99$;	6	
		10^{-r} (r = 3, ..., 9)	5	
	$2[1 - \Phi(x)]$	x = 0.25,0.5(0.5)5.0	6	
Smirnov (1965, tables I, II, III)	$\Phi(x) - \frac{1}{2}$, $\phi(x)$	x = 0.00(0.001)2.50(0.002) 3.40(0.005)4.00(0.01) 4.50	7	
		x = 4.50(0.01)6.00	10	
	$-\log_{10}\{1 - \Phi(x)\}$	x = 5(1)50(10)100(50)500	5	
Yamauti (1972, tables A1, A2, A3)	$\phi(x)$	x = 0.0(0.01)4.99		5
	$1 - \Phi(x)$	x = 0.0(0.01)4.99		5
	$1 - \Phi(x)$	x = 0.1(0.1)10.0		35
	z_α	α = 0.0(0.001)0.499	5	

[a]Tables compiled by M. Zelen and N. C. Severo.

[*3.1.2*] There is a nomogram for $\Phi(x)$ by Varnum (1950, pp. 32-34) and for the mean of a normal distribution singly truncated from below by Altman (1950, pp. 30-31) (see [1.7.2]).

[*3.1.3*] A large body of computing algorithms for the normal cdf and quantiles is available, and for the generation of random normal deviates, that is, of simulated "samples" from normal populations. In this section we shall identify some of these programs.

[*3.1.3.1*] Some algorithms for computing $\Phi(x)$, *given* x, use polynomial and rational approximations; these appear in [3.2.1] and are based on sources such as Hastings (1955). Algorithms of Ibbetson (1963, p. 616) and Adams (1969, pp. 197-198) for computing the standard normal cdf $\Phi(x)$ are unrivaled for speed if accuracy to seven and nine decimal places (respectively) is sufficient; see Hill (1969, p. 299). These algorithms are designed for everyday statistical use; if more accuracy is required, the computation time will increase, in general. Hill and Joyce (1967, pp. 374-375) give an algorithm providing accuracy within at least two decimal places of the machine capability. IMSL (1977, pp. 11-12) has subroutines for computing normal percentiles, cdf values $\Phi(x)$, pdf values $\phi(x)$, and functions related to the distributions of chi-square, t, F, bivariate normal distribution function values, as well as cdfs for other statistical distributions. These subroutines can be incorporated into larger programs, and are available on a wide variety of machines.

[*3.1.3.2*] An algorithm for computing N(0,1) *quantiles* z_α is that of Beasley and Springer (1977, pp. 118-121); see Odeh and Evans (1974, pp. 96-97). On an ICL System 470 with a 24-bit mantissa, the largest error in a test of 10,000 starting values of α was 1.14×10^{-6}. This algorithm may be useful when working in Fortran or another high-level language. In machine code, the "fast" method of Marsaglia et al. (1964, pp. 4-10) should be considered.

[*3.1.3.3*] (a) Tables of "random numbers" may be used to generate *random normal deviates* in conjunction with algorithms discussed in [3.1.3.2], but good computer programs are available. Marsaglia and Bray (1964, pp. 260-264) developed a program which is almost as fast as that of Marsaglia et al. (1964, pp. 4-10), but easier to implement, and with smaller storage capacity. See also Abramowitz and Stegun (1964, pp. 952-953).

With user accessibility as their criterion for comparison, Kinderman and Ramage (1976, pp. 893-896) recommend an algorithm of their own and one by Ahrens and Dieter (1972, pp. 873-882). By user accessibility is meant a suitable combination of machine independence, brevity, and implementation in a programmer-oriented language.

All of the preceding programs use probability mixing; among assembler implementations which are machine-dependent, a second algorithm of Ahrens and Dieter (1973, pp. 927-937) seems a good choice.

(b) Box and Muller (1958, pp. 610-611) suggested the following transformations of pairs of independent random numbers U_1 and U_2 to generate a pair (Z_1, Z_2) of independent standard normal variables:

$$Z_1 = (-2 \ln U_1)^{1/2} \cos(2\pi U_2)$$

$$Z_2 = (-2 \ln U_1)^{1/2} \sin(2\pi U_2)$$

Here U_1 and U_2 are independent random variables uniformly distributed on the unit interval (0,1). The Box-Muller transformation can be used on small as well as large computers; it has become a standard method for generating N(0,1) deviates.

(c) Tables of random normal deviates by Wold (1948) or by Sengupta and Bhattacharya (1958, pp. 250-286) are available.

3.2 EXPRESSIONS FOR THE DISTRIBUTION FUNCTION

Many of the results for evaluating the normal cdf $\Phi(x)$ are in terms of Mills' ratio $[1 - \Phi(x)]/\phi(x)$, but there are some expressions which do not directly involve the pdf $\phi(x)$, and we present these first.

[*3.2.1*] *Power Series Expansions of* $\Phi(x)$. (a) If $x \ge 0$ (Laplace, 1785; Adams, 1974, p. 29),

$$\Phi(x) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}}\left(x - \frac{x^3}{6} + \frac{x^5}{40} - \frac{x^7}{336} + \cdots + \frac{(-1)^r x^{2r+1}}{(2r + 1)2^r(r!)} + \cdots\right)$$

This series converges slowly except when x is small (Shenton, 1954, p. 182; Kendall and Stuart, 1977, p. 143). If we stop the summation when $r = n - 1$, and the remainder is $E_n(x)$, then

$$|E_n(x)| < x^{2n+1} \exp(x^2/2)/[2^n(2n + 1)\sqrt{2\pi}]$$

Badhe (1976, pp. 173-176) derived an approximation from the above expansion for $0 < x \le 2$:

$$\Phi(x) = 0.5 + x(a + y[b + y\{c + y(d + y[e + y\{f + y(g + hy)\}])\}])$$

$$y = x^2/32, \quad a = 0.3989,4227,84 \quad b = -2.1276,9007,9$$
$$c = 10.2125,6621,21 \quad d = -38.8830,3149,09$$
$$e = 120.2836,3707,87 \quad f = -303.2973,1534,19$$
$$g = 575.0731,3191,7 \quad h = -603.9068,0920,58$$

This approximation is suitable for use with a desk calculator, and has maximum |error| equal to 0.20×10^{-8} if $0 \le x \le 2$.

(b) The following expansion (Kerridge and Cook, 1976, p. 402) is convenient over the whole range of values of x, unlike many other expressions appearing in this chapter. The most computationally attractive form is given by

$$\Phi(x) = \frac{1}{2} + (\sqrt{2\pi})^{-1} x \exp\left(\frac{-x^2}{8}\right) \sum_{n=0}^{\infty} \frac{\theta_{2n}\left(\frac{1}{2} x\right)}{2n + 1}, \qquad -\infty < x < \infty$$

where $\theta_0(x) = 1$, $\theta_1(x) = x^2/2$, and $\theta_{2n}(x)$ is obtained from the recurrence relation

$$\theta_{r+1}(x) = x^2\{\theta_r(x) - \theta_{r-1}(x)\}/(r + 1), \qquad r = 1, 2, \ldots$$

Alternatively, $\theta_r(x) = x^r H_r(x)/r!$, where $H_r(x)$ is the rth Hermite polynomial ([2.1.9]; Kendall and Stuart, 1977, pp. 166-167).

The number of terms in the above series needed to gain accuracy to 5, 10, or 15 decimal places are tabled (for 10 decimal places, 6, 13, and 24 terms when x = 1, 3, and 6, respectively). The expansion requires uniformly fewer such terms for $1 \le x \le 6$ than the Laplace series in (a) above and also than that in [3.5.2], or than Shenton's (1954) continued fraction ([3.4.1(b)]). It requires fewer terms than Laplace's (1805) continued fraction if $x \le 2$, but if $x \ge 3$, the latter begins to perform better [3.4.1(a)], as does Laplace's (1785) asymptotic series ([3.5.1]).

[3.2.2] *Rational Approximations*. The first three of the following appear in Abramowitz and Stegun (1964, p. 932) and in Hastings (1955, pp. 185-187):

(a)
$$2[1 - \Phi(x)] = [1 + (0.1968{,}54)x + (0.1151{,}94)x^2 + (0.0003{,}44)x^3 + (0.0195{,}27)x^4]^{-4} + 2\varepsilon(x), \qquad x \ge 0$$
$$|\varepsilon(x)| < 2.5 \times 10^{-4}$$

(b)
$$2[1 - (x)] = [1 + (0.0997{,}9271)x + (0.0443{,}2014)x^2 + (0.0096{,}9920)x^3 - (0.0000{,}9862)x^4 + (0.0058{,}1551)x^5]^{-8} + 2\varepsilon(x), \qquad x \ge 0$$
$$|\varepsilon(x)| < 2 \times 10^{-5}$$

(c)
$$2[1 - \Phi(x)] = [1 + (0.0498{,}6734{,}70)x + (0.0211{,}4100{,}61)x^2 + (0.0032{,}7762{,}63)x^3 + (0.0000{,}3800{,}36)x^4 + (0.0000{,}4889{,}06)x^5 + (0.0000{,}0538{,}30)x^6]^{-16} + 2\varepsilon(x), \qquad x \ge 0$$
$$|\varepsilon(x)| < 1.5 \times 10^{-7}$$

(d) Carta (1975, pp. 856-862) obtains fractions of the form

$$1 - \Phi(x) \simeq \frac{1}{2}[c_1 + c_2 x + \cdots + c_n x^{n-1}]^{-2^q}, \qquad x > 0$$

A range of such approximations, together with charts of minimum absolute and relative error criteria, is obtained, including tables of c_1, c_2, ... for n = 4, 5, 6 and for q = 0(1)5. Carta includes approximations with one coefficient constrained to be zero, when these compare favorably with the above.

[*3.2.3*] Five approximations involving exponential functions:

(a) $$\Phi(x) \simeq [\exp(-2x\sqrt{2/\pi}) + 1]^{-1}, \qquad x \geq 0$$
$$= \frac{1}{2}\{1 + \tanh(x\sqrt{2/\pi})\} \qquad \text{(Tocher, 1963)}$$

If $0 \leq x \leq 4.0$, the largest |error| is 0.017670 when x = 1.5.

(b) The following limit provides an approximation when $x \to \infty$ (Feller, 1968, p. 194):

$$\{1 - \Phi(x + a/x)\}/\{1 - \Phi(x)\} \to e^{-a} \qquad \text{as } x \to \infty,\ a > 0$$

(c) $$\Phi(x) \simeq [1 + \exp(-2y)]^{-1} = \{1 + \tanh(y)\}/2, \qquad x \geq 0$$
$$y = (\sqrt{2/\pi})x\{1 + (0.0447,15)x^2\} \qquad \text{(Page, 1977, p. 75)}$$

The maximum |error| is 0.000179. This approximation was developed to improve on that by Tocher above, and is suitable for hand calculators.

(d) $$2\Phi(x) - 1 < [1 - \exp(-2x^2/\pi)]^{1/2}, \qquad x > 0$$
(Pólya, 1949, p. 64)

If the bound on the right is used as an approximation, then the maximum |relative error| when x > 0 is less than 0.71×10^{-2}. Hamaker (1978, p. 77) obtained a modification, as the approximation

$$\Phi(x) \simeq \frac{1}{2}[1 + \text{sgn}(x)\{1 - \exp(-t^2)\}^{1/2}], \qquad -\infty < x < \infty$$
$$t = 0.806|x|(1 - 0.018|x|), \qquad \text{sgn}(x) = |x|/x \qquad (x \neq 0)$$

This is an improvement on the Pólya approximation.

[*3.2.4*] Some inequalities have an upper bound, say, related to Mills' ratio R(x), while a lower bound may belong in this section. For the sake of completeness, these results are collected together in [3.7]; see [3.7.2] and [3.7.3].

[3.2.5] Moran (1980, pp. 675-676) gives two approximations to $\Phi(x)$ which are accurate to nine decimal places when $|x| \le 7$:

(a) $$\Phi(x) \simeq \frac{1}{2} + \frac{1}{\pi}\left\{\frac{x}{3\sqrt{2}} + \sum_{n=1}^{12} \frac{\exp(-n^2/9)}{n} \sin\left(\frac{nx\sqrt{2}}{3}\right)\right\}$$

(b) $$\Phi(x) \simeq \frac{1}{2} + \frac{1}{\pi} \sum_{n=0}^{12} \frac{\exp\left\{-\left(n + \frac{1}{2}\right)^2/9\right\}}{n + \frac{1}{2}} \sin\left\{\frac{\left(n + \frac{1}{2}\right)x\sqrt{2}}{3}\right\}$$

If $|x| > 7$, the accuracy decreases rapidly.

3.3 EXPRESSIONS FOR THE DENSITY FUNCTION

[3.3.1] Using the exponential series,

$$\phi(x) = (\sqrt{2\pi})^{-1}\left[1 + \sum_{i=1}^{\infty} \frac{(-x^2/2)^i}{(i!)}\right]$$

[3.3.2] *Rational Approximations* (Abramowitz and Stegun, 1964, pp. 932-933; Hastings, 1955, pp. 151-153). For all values of x,

(a) $$\phi(x) = [2.4908{,}95 + (1.4660{,}03)x^2 - (0.0243{,}93)x^4 + (0.1782{,}57)x^6]^{-1} + \varepsilon(x)$$

$$|\varepsilon(x)| < 2.7 \times 10^{-3}$$

(b) $$\phi(x) = [2.5112{,}61 + (1.1728{,}01)x^2 + (0.4946{,}18)x^4 - (0.0634{,}17)x^6 + (0.0294{,}61)x^8]^{-1} + \varepsilon(x)$$

$$|\varepsilon(x)| < 0.8 \times 10^{-3}$$

(c) $$\phi(x) = [2.5052{,}367 + (1.2831{,}204)x^2 + (0.2264{,}718)x^4 + (0.1306{,}469)x^6 - (0.0202{,}490)x^8 + (0.0039{,}132)x^{10}]^{-1} + \varepsilon(x)$$

$$|\varepsilon(x)| < 2.3 \times 10^{-4}$$

(d) Derenzo (1977, p. 217) gives a simple approximation to $\Phi(x)$ with integer coefficients; it may be useful for hand calculators:

If

$$0 < x \le 5.5 \quad \text{and} \quad \frac{1}{2} > 1 - \Phi(x) \ge 1.9 \times 10^{-8}$$

then

$$1 - \Phi(x) = \frac{1}{2} \exp\left[- \frac{(83x + 351)x + 562}{703x^{-1} + 165}\right] + \varepsilon(x),$$

$$|\varepsilon(x)|/\{1 - \Phi(x)\} < 0.42 \times 10^{-4}$$

If

$$x \ge 5.5 \quad \text{and} \quad 1 - \Phi(x) \le 1.9 \times 10^{-8}$$

then

$$1 - \Phi(x) = (2\pi)^{1/2}x^{-1} \exp(-x^2/2 - 0.94/x^2) + \varepsilon(x),$$

$$|\varepsilon(x)|/\{1 - \Phi(x)\} < 0.40 \times 10^{-4}$$

3.4 CONTINUED FRACTION EXPANSIONS: MILLS' RATIO

Let

$$P = \cfrac{x}{a + \cfrac{y}{c + \cfrac{z}{d + \cdots}}} = \frac{x}{a +}\,\frac{y}{c +}\,\frac{z}{d +} \cdots$$

that is, a continued fraction.

[*3.4.1*] We compare two expansions for Mills' ratio:

(a) $R(x) = \frac{1}{x +}\,\frac{1}{x +}\,\frac{2}{x +}\,\frac{3}{x +}\,\frac{4}{x +} \cdots, \quad x > 0$ (Laplace, 1805)

(b) $\{2\phi(x)\}^{-1} - R(x) = [\Phi(x) - 1/2]/\phi(x)$

$$= \frac{x}{1 -}\,\frac{x^2}{3 +}\,\frac{2x^2}{5 -}\,\frac{3x^2}{7 +}\,\frac{4x^2}{9 -} \cdots, \quad x > 0$$

(Shenton, 1954, pp. 182-184)

Shenton (1954) compared (a) and (b) for rapidity of convergence. Expansion (a) converges more rapidly for moderate or large x; the larger the value of x, the more rapid the convergence. Expansion (b) converges more rapidly when x is small; the smaller the value of x, the more rapid the convergence. To achieve an accuracy of 2.5×10^{-7}, expansion (a) is preferred when $x \ge 2.5$, and expansion (b) when $x < 2.5$. See also Kendall and Stuart (1977, pp. 145-146) and [3.2.1(b)].

[3.4.2] Patry and Keller (1964, pp. 89-97) give the following continued fraction expansion:

$$R(x) = \left\{\sqrt{\frac{\pi}{2}}\right\}\left[\frac{1}{a_0w +}\,\frac{1}{a_1w +}\,\frac{1}{a_2w +}\,\cdots\right], \qquad x \ge 0 \qquad w = \frac{x}{\sqrt{2}}$$

$$a_0 = \sqrt{\pi}, \qquad a_1 = \frac{2}{\sqrt{\pi}}$$

$$\left.\begin{aligned} a_{2n} &= \sqrt{\pi}(2n - 1)!!/(2n)!!, \\ a_{2n+1} &= 2(2n)!!/[(2n + 1)!!\sqrt{\pi}] \end{aligned}\right\} \qquad n = 1, 2, \ldots$$

where $n!! = n(n - 2)!!$, and $n!! = 1$ if $n = 0$ or $n = 1$. Convergence is poor near zeroes of the expansion, and so the following approximation was derived from the above (Patry and Keller, 1964, p. 93):

$$R(x) \simeq t/(tx + \sqrt{2}), \qquad x \ge 0$$

$$t = \sqrt{\pi} + (2 - q)x, \qquad \sqrt{\pi} = 1.7724{,}5385$$

$$\begin{aligned} q = a/b;\ a &= 0.8584{,}0765{,}7 + x[0.3078{,}1819{,}3 \\ &\quad + x\{0.0638{,}3238{,}91 - (0.0001{,}8240{,}5075)x\}] \\ b &= 1 + x[0.6509{,}7426{,}5 + x\{0.2294{,}8581{,}9 \\ &\quad + (0.0340{,}3018{,}23)x\}] \end{aligned}$$

As an approximation to $1 - \Phi(x)$, this algorithm has $|\text{error}|$ less than 12.5×10^{-9} for the range $0 \le x \le 6.38$, and relative error less than 30×10^{-9}, 100×10^{-9}, and 500×10^{-9} if $0 \le x \le 2$, $2 < x \le 4.6$, and $4.6 < x \le 6.38$, respectively (Patry and Keller, 1964, p. 93). Badhe (1976, p. 174) states that this approximation is better than those of Hart (1966, pp. 600-602), Hastings (1955, pp. 167-169), and Schucany and Gray (1968, pp. 201-202); see [2.6.2], [2.6.1], and [2.6.3], respectively.

3.5 EXPRESSIONS FOR MILLS' RATIO, BASED ON EXPANSIONS

Mills' ratio R(x) is equal to $[1 - \Phi(x)]/\phi(x)$, and hence the following results can be rephrased in terms of the cdf $\Phi(x)$. Thus algorithms and inequalities for R(x) can be used as algorithms and inequalities for Φ.

[3.5.1] Laplace (1785) gives the series expansion

$$R(x) = \frac{1}{x} - \frac{1}{x^3} + \frac{1\cdot 3}{x^5} - \cdots + (-1)^n \frac{1\cdot 2\cdot 5\cdots(2n-1)}{x^{2n+1}} + \cdots, \qquad x \ge 0$$

which diverges as $n \to \infty$, but is reasonably effective for large values of x. If we write (Rahman, 1968, pp. 144-145; Kendall and Stuart, 1977, pp. 144-145)

$$R(x) = \sum_{r=1}^{n} (-1)^n (2r!)/[2^r x^{2r+1}(r!)] + E_n(x)$$

then

$$|E_n(x)| < (2n!)/[2^n x^{2n+1}(n!)]$$

The sum to n terms is an upper bound to R(x) when n is even, and a lower bound when n is odd (Feller, 1957, p. 145; 1968, pp. 175, 193-194); see also [3.2.1(b)].

[3.5.2] $\Phi(x) = \frac{1}{2} + \phi(x) \Sigma_{r=0}^{\infty} x^{2r+1}/\{1\cdot 3\cdot 5\cdot\cdots\cdot(2r+1)\}$, $x \ge 0$ (Laplace, 1812, p. 103; Abramowitz and Stegun, 1964, p. 932; Pólya, 1949, p. 66). This expression is not a serious rival to the expansions given in [3.2.1]; see Shenton (1954, p. 180). However, the following bounds have been derived from the above by Gupta and Waknis (1965, pp. 144-145):

$$\frac{\Phi(x) - \frac{1}{2}}{\phi(x)} > \sum_{r=0}^{n-1} \frac{x^{2r+1}}{1\cdot 3\cdot 5\cdot \cdots (2r+1)}, \qquad x > 0,\ n = 1, 2, \ldots$$

$$\frac{\Phi(x) - \frac{1}{2}}{\phi(x)} < \sum_{r=0}^{n-1} \frac{x^{2r+1}}{1\cdot 3\cdot 5\cdot \cdots (2r+1)} + \frac{(2n+3)x^{2n+1}2^{-n-1}}{(2n+3-x^2)\left(n+\frac{1}{2}\right)_{n+1}},$$

$$0 < x < \sqrt{2n+3}$$

where $(n)_k = n(n-1)\cdots(n-k+1)$, $k \geq 1$. Gupta and Waknis (1965) tabulate bounds for $[\Phi(x) - 1/2]/\phi(x)$ for selected values of n and x. When $0.1 \leq x \leq 3.0$, the lower bound above appears to be closer to the exact value of $\Phi(x) - 1/2$ than is the upper bound.

[*3.5.3*] Shenton (1954, p. 188) quotes the expansion

$$xR(x) = 1 - \frac{1}{x^2+2} + \frac{1}{(x^2+2)(x^2+4)} - \frac{5}{(x^2+2)(x^2+4)(x^2+6)}$$
$$+ \frac{9}{(x^2+2)\cdots(x^2+8)} - \frac{129}{(x^2+2)\cdots(x^2+10)} + \cdots,$$
$$x > 0$$

If the general term is $(-1)^n a_n/[(x^2+2)(x^2+4)\cdots(x^2+2n)]$, then (Badhe, 1976, pp. 174-175) $a_6 = 57$, $a_7 = 9141$, $a_8 = 36{,}879$, $a_9 = 1{,}430{,}049$, $a_{10} = 19{,}020{,}019$, and $a_{11} = 1{,}689{,}513{,}233$; see also Abramowitz and Stegun (1964, p. 932). Badhe (1976) gives an approximation based on the sum to the term in which n = 9, as follows:

$$xR(x) = 1 - y\{1 + y(7 + y[55 + y\{445 + 3745\, Q_1(x)y\}])\}, \quad x > 0$$
$$y = (x^2+10)^{-1}$$
$$Q_1(x) = 1 + 8.50\{x^2 - (0.4284{,}6397{,}53)x^{-2} + 1.2409{,}6410{,}9\}^{-1}$$

This algorithm can be computed on a desk calculator. Based on relative error as a criterion, it is more accurate when $x > 4$ than Patry and Keller (1964); see [3.4.2]. However, it is not suitable when $x < 2$.

[*3.5.4*] Ruben (1962, pp. 178-179) derived the expansion

$$R(x) = \sum_{k=1}^{n} (k-1)!\,g_{k-1}\left(\frac{k}{x}\right)x^{-(2k-1)} + O(x^{-2n-1}), \qquad x > 0$$

$$g_0(y) = \exp\left(\frac{-y^2}{2}\right)$$

and

$$g_k(y) = \frac{d}{dy}\left\{\frac{g_{k-1}(y) - g_{k-1}(k/x)}{y - k/x}\right\}, \qquad k = 1, 2, \ldots$$

Ruben gives the expressions

$$g_1(x^{-1})^* = x^2 \exp\left(-\frac{1}{2}x^{-2}\right) - (x^2 + 2)\exp(-2x^{-2})$$

$$g_2(3x^{-1}) = \frac{1}{2}x^4 \exp\left(-\frac{1}{2}x^{-2}\right) - (x^4 + 2x^2)\exp(-2x^{-2})$$
$$+ \frac{1}{2}(x^4 + 5x^2 + 9)\exp(-9x^{-2}/2)$$

If the expansion is taken as far as the term in $g_2(\cdot)$, then the $|\text{error}|$ is less than 0.0004 in the range $3 \le x \le 10$.

[3.5.5] $R(x) = \Sigma_{k=0}^{n=1} \beta_k(x) + E_n(x)$, $x > 0$ (Ruben, 1963, p. 360); $n = 1, 2, \ldots$, where β_k is a rational function defined by

$$\beta_k(x) = \frac{1\cdot 3\cdots(2k - 1)2^{-k}}{(k!)x}\sum_{i=0}^{k}(-1)^i\binom{k}{i}\alpha_i$$

$$\alpha_i = \gamma_i x^2\left\{1 + \sum_{j=1}^{i} i(i - 1)\cdots(i - j + 1)(2\gamma_i)^j\right\}$$

$$\gamma_i = (x^2 + 2i + 1)^{-1}$$

Then, if $x > 1$,

$$0 < E_n(x) \le (2\pi)^{-1/2}\, 3\cdot 5\cdots(2n - 1)\, x^2(x^2 - 1)^{-n-1}$$
$$\times \min[e^{-n}, (2n)!(x^2 - 1)^{-n}/n!]$$

If $S_n(x) = \Sigma_{k=0}^{n-1}\beta_k(x)$, then (Ruben, 1963, pp. 360, 362-363)

$$S_1(x) = \frac{x}{x^2 + 1} \qquad S_2(x) = \frac{x(x^4 + 6x^2 + 11)}{(x^2 + 1)(x^2 + 3)^2}$$

---

*This should perhaps read $g_1(2x^{-1})$.

$$S_3(x) = \frac{x(x^{10} + 21x^8 + 176x^6 + 740x^4 + 1611x^2 + 1507)}{(x^2 + 1)(x^2 + 3)^2(x^2 + 5)^3}$$

Each partial sum is a rational approximation; $S_3(x)$ compares favorably when $2 \le x \le 10$ with the approximation of Gray and Schucany (1968, pp. 718-719); see [3.6.3]: $S_3(x)$ is also better in that range than $C_2(x)$ (Ruben, 1964, pp. 340-343), discussed next.

[3.5.6]

$$R(x) = \sum_{k=0}^{n-1} \gamma_k(x) + E_n(x), \qquad x > 0;\ n = 1, 2, \ldots$$

$$\gamma_0(x) = (x^2 + 2)^{-1/2}$$

$$\gamma_k(x) = \sum_{j=0}^{k} (-1)^j \binom{k}{j} (x^2 + 2j + 2)^{-1/2} \delta_j(x), \qquad k = 1, 2, \ldots$$

$$\delta_0(x) = 1$$

and

$$\delta_j(x) = 1 + \sum_{s=1}^{j} 1 \cdot 3 \cdots (2s - 1) \binom{j}{s} (x^2 + 2j + 2)^{-s}, \qquad j = 1, 2, \ldots$$

Further (Ruben, 1964, pp. 340-343)

$$0 < E_n(x) \le \min[(e^{-n}) 1 \cdot 3 \cdots (2n - 1) x^{-2n-1},$$
$$(2^{-n}) 1 \cdot 3 \cdots (4n - 1) x^{-4n-1}]$$

Let $\Sigma_{k=0}^{n-1} \gamma_k(x) = C_n(x)$. Then the first three partial sums [approximations to R(x)] are given by

$$C_1(x) = (x^2 + 2)^{-1/2}$$

$$C_2(x) = 2(x^2 + 2)^{-1/2} - [1 + (x^2 + 4)^{-1}]/(x^2 + 4)^{1/2}$$

$$C_3(x) = 3(x^2 + 2)^{-1/2} - 3[1 + (x^2 + 4)^{-1}]/(x^2 + 4)^{1/2}$$
$$+ [1 + 2(x^2 + 6)^{-1} + 3(x^2 + 6)^{-2}]/(x^2 + 6)^{1/2}$$

See Ruben (1964, p. 341) and [3.5.5] above.

[3.5.7] Using Tchebyshev polynomial representations, Ray and Pitman (1963) derived two expansions for R(x), one valid in the interval [0.1] and the second in the range $[1,\infty)$.

(a) $$R(x) = \frac{1}{2} a_0 T_0(x) + \sum_{i=1}^{\infty} a_i T_i(x), \qquad 0 \le x \le 1$$

where

$$T_0(x) = 1,\ T_1(x) = x,\ T_{r+1}(x) = 2xT_r(x) - T_{r-1}(x);\ r = 1, 2, \ldots$$

Choosing an integer n large enough to assume $a_r = 0$ when $r > n$, the precision of the expansion can be controlled; see Ray and Pitman (1963, pp. 893-895). If n = 11, the values of the coefficients are:

a_0 = 3.2690,615	a_1 = -1.2974,425	a_2 = 0.4054,731
a_3 = -0.1076,724	a_4 = 0.0252,763	a_5 = -0.0053,740
a_6 = 0.0010,518	a_7 = -0.0001,917	a_8 = 0.0000,328
a_9 = -0.0000,053	a_{10} = 0.0000,008	a_{11} = -0.0000,001

When $0.2 \le x \le 1.0$, the number of terms required to achieve accuracy to 3, 4, 5, or 6 decimal places is 7, 8, 10, or 11, respectively; see [2.5.9] also.

(b) $$R(x) = x^{-1}\left[\frac{1}{2} b_0 T_0(x^{-1}) + \sum_{i=1}^{\infty} b_{2i} T_{2i}(x^{-1})\right], \qquad x \ge 1$$

where

$$T_0(y) = 1,\ T_2(y) = 2y^2 - 1,\ (2y^2 - 2)T_r(y) = T_{r+2}(y) + T_{r-2}(y),$$
$$r = 2, 4, 6, \ldots$$

The precision can be controlled as in (a); see Ray and Pitman (1963, pp. 895-896). If $b_{2i} = 0$ when $i > 13$, the coefficients are:

b_0 = 1.5792,832	b_2 = -0.1611,570	b_4 = 0.0345,335
b_6 = -0.0096,581	b_8 = 0.0031,367	b_{10} = -0.0011,253
b_{12} = 0.0004,341	b_{14} = -0.0001,771	b_{16} = 0.0000,754
$b_{18} = -(0.331) \times 10^{-4}$	$b_{20} = (0.149) \times 10^{-4}$	$b_{22} = -(0.070) \times 10^{-}$
$b_{24} = (0.038) \times 10^{-4}$	$b_{26} = -(0.027) \times 10^{-4}$	

When $2 \le x \le 10$, the number of terms required to achieve accuracy to 3, 4, 5, or 6 decimal places is 5, 8, 11, or 14, respectively; see also [2.5.9].

[*3.5.8*] Ray and Pitman (1963, pp. 898-901). If $y = \exp(-2/x^2)$,

$$
\begin{aligned}
xR(x) = 1 &+ \frac{1}{2}(y - 1) + (y^4 - 4y + 3)/12 \\
&+ 23(y^9 - 6y^4 + 15y - 10)/1440 \\
&+ 37(y^{16} - 8y^9 + 28y^4 - 56y + 35)/13440 \cdots
\end{aligned}
$$

The terms in this expansion contain exponentials only. The approximation shown here has maximum |error| less than 0.0029 when $2 \le x \le 10$, and less than 0.00002 when $4 \le x \le 10$. See also [3.5.9] below.

[*3.5.9*] Ray and Pitman (1963, pp. 899-901) give some approximations to $R(x)$ based on a Laguerre-Gauss expansion. Three of these are as follows:

$$R_1(x) = x^{-1} \exp(-0.5/t^2), \qquad |\text{error}| < 1/(2x^3)$$

$$
\begin{aligned}
R_2(x) = x^{-1}[&(0.853553) \exp(-0.171573/x^2) \\
&+ (0.146447) \exp(-5.828429/x^2)], \qquad |\text{error}| < 1/(2x^5)
\end{aligned}
$$

$$
\begin{aligned}
R_5(x) = x^{-1}[&(0.521756) \exp(-0.034732/x^2) \\
&+ (0.398667) \exp(-0.998854/x^2) \\
&+ (0.075942) \exp(-6.46714/x^2) \\
&+ (0.003612) \exp(-25.1044/x^2) \\
&+ (0.000023) \exp(-79.8949/x^2)], \qquad |\text{error}| < 15/(4x^{11})
\end{aligned}
$$

The approximation $R_5(x)$ is more accurate than that of [3.5.8] when $2 \le x \le 10$, but the latter is simpler in form; $R_2(x)$ is similarly more accurate than the sum of the first two terms of [3.5.8].

[*3.5.10*] Expansions of the same type as in [3.5.7], but which appear to require fewer terms to achieve the same accuracy are given by Rabinowitz (1969, pp. 647-651). These are

$$R(x) \simeq \frac{1}{2} b_0(a) + \sum_{r=1}^{k} b_r(a) T_r(2xa^{-1} - 1), \qquad 0 \le x \le a$$

$$R(x) \simeq x^{-1}\left[\frac{1}{2} d_0(a) + \sum_{r=1}^{k} d_r(a) T_r(2ax^{-1} - 1)\right], \qquad x \ge a > 0$$

The Tchebyshev polynomials $T_r(\cdot)$ are defined in [3.5.7]. Values of the constants $b_r(1)$, $b_r(2)$, $d_r(1)$, and $d_r(2)$ for r = 0(1)9, r = 0(1)11, r = 0(1)18, and r = 0(1)13, respectively, are tabulated. When a = 1, the number of terms required to achieve an accuracy of 7 decimal places in the range $0 \le x \le 1$ is 8, compared with 12 in the expansion of [3.5.7(a)]; in the range $x \ge 1$, 10 terms are needed for 6-place accuracy, compared with 14 in the expansion of [3.5.7(b)].

3.6 OTHER APPROXIMATIONS TO MILLS' RATIO

[3.6.1] The following approximations are based on Hastings (1955, pp. 167-169); see also Abramowitz and Stegun (1964, p. 932).

(a) $$\Phi(x) = 1 - \phi(x)[(0.4361,836)t - (0.1201,676)t^2 + (0.9372,980)t^3] + \varepsilon(x), \qquad x \ge 0$$

$$t = \{1 + (0.3326,7)x\}^{-1}, \qquad |\varepsilon(x)| < 1.2 \times 10^{-5}$$

(b) $$\Phi(x) = 1 - \phi(x)[(0.1806,1682)t + (0.7652,0181)t^2 - (0.7616,8891)t^3 + (1.0691,8442)t^4] + \varepsilon(x), \qquad x \ge 0$$

$$t = \{1 + (0.2700,90)x\}^{-1}, \qquad |\varepsilon(x)| < 10^{-6}$$

(c) $$\Phi(x) = 1 - \phi(x)[(0.3193,8153,0)t - (0.3565,6378,2)t^2 + (1.7814,7793,7)t^3 - (1.8212,5597,8)t^4 + (1.3302,7442,9)t^5] + \varepsilon(x), \qquad x \ge 0$$

$$t = \{1 + (0.2316,419)x\}^{-1}, \qquad |\varepsilon(x)| < 7.5 \times 10^{-8}$$

[3.6.2] Hart (1966) gives the approximation

$$xR(x) \simeq 1 - y[x\sqrt{\pi/2} + \{(\pi x^2/2) + y\exp(-x^2/2)\}^{1/2}]^{-1}, \quad |x| < \infty$$

$$y = (1 + 2\pi a^2 x^2)^{1/2}/(1 + ax^2)$$

$$a = [1 + (1 + 6\pi - 2\pi^2)^{1/2}]/(2\pi)$$

The maximum |error| is 0.00013, near x = ±1; the maximum |relative error| is 0.00055, near x = ±1.7; see [3.6.3] below.

[*3.6.3*] Gray and Schucany (1968, p. 718) give the rational approximation

$$R(x) \simeq \frac{x}{x^2 + 2}\left[\frac{x^6 + 6x^4 - 14x^2 - 28}{x^6 + 5x^4 - 20x^2 - 4}\right], \qquad x > 2$$

which is uniformly more accurate than that in [3.6.2] when $3 \le x \le 10$, in that the |relative error| is smaller. The maximum |relative error| is 0.0020 when x = 2; if $x \ge 4$, the same quantity is 0.00001 when x = 4. This approximation has a |relative error| uniformly less than that of $S_2(x)$ in [3.5.5], and compares favorably with $S_3(x)$ in [3.5.5], being of simpler structure. See also Schucany and Gray (1968, pp. 201-202).

[*3.6.4*] Andrews (1973) derived an approximation formula for tail probabilities in a class of distributions. In the case of the N(0,1) distribution,

$$1 - \Phi(x) \simeq \phi(x)\{x^{-1} - (2x^3)^{-1}\}, \qquad x > 0$$

As an approximation to $1 - \Phi(x)$, the |relative error| when $1 - \Phi(x)$ has values 0.05, 0.01, and 5×10^{-6} is 0.02, 0.04, and 0.02, respectively.

[*3.6.5*] Pearson and Hartley (1958, p. 117) point out that, when x > 50, the approximation $1 - \Phi(x) \simeq \phi(x)/x$ is adequate. Note, however, that $R(x) \le 1/x$, x > 50; see [3.7.1].

3.7 INEQUALITIES FOR MILLS' RATIO AND OTHER QUANTITIES

There is a good discussion of many of the following results in Johnson and Kotz (1970b, pp. 278-282) and Mitrinović (1970, pp. 177-180). There are two main points of concern in comparing inequalities.

(i) How useful is a bound as an approximation, i.e., how sharp is the bound?

(ii) Do bounds converge in some sense to the function of interest?

We have already stated some inequalities which are partial sums of an expansion, as in [3.5.1], and for $\Phi(x)$, as in [3.2.3(c)]; see also [3.5.2].

[*3.7.1*] $x/(x^2 + 1) \le [(x^2 + 4)^{1/2} - x]/2 \le R(x) \le 1/x; \ x > 0$ (Gordon, 1941, p. 365; Birnbaum, 1942, p. 245). Equivalently,

$$1 - \phi(x)/x \le \Phi(x) \le 1 - \phi(x)[(x^2 + 4)^{1/2} - x]/2$$

$$\phi(x)[(x^2 + 4)^{1/2} - x]/2 \le 1 - \Phi(x) \le \phi(x)/x$$

[*3.7.2*] Tate (1953, p. 133) derived inequalities for $1 - \Phi(x)$ such that, when $x > 0$, the upper bound improves upon that in [3.7.1]:

$$[1 - \{1 - \exp(-x^2)\}^{1/2}]/2 \le 1 - \Phi(x) \le \frac{1}{2} + \phi(x)x^{-1} - h(x), \quad x \ge 0$$

$$\frac{1}{2} + \phi(x)x^{-1} + h(x) \le 1 - \Phi(x) \le [1 + \{1 - \exp(-x^2)\}^{1/2}]/2, \quad x \le 0$$

$$h(x) = [(1/4) + \exp(-x^2)/(2\pi x^2)]^{1/2}$$

When $x \ge 0$, neither lower bound for $1 - \Phi(x)$ in the above or in [3.7.1] is uniformly sharper than the other.

[*3.7.3*] Chu (1955, p. 263) obtained the inequality

(a) $\Phi(x) \ge [1 + \{1 - \exp(-x^2/2)\}^{1/2}]/2, \quad x \ge 0$

He also derived the Pólya (1949) upper bound of [3.2.3(c)], and found it to be sharper than that derived from [3.7.2] when $x > 0$. In comparing the Pólya bound with the upper bounds derived from [3.7.1], Chu (1955) found the Polya bound to be better for values of x close to zero, and the bounds of [3.7.2] to be preferable for large positive values of x. The above lower bound for $\Phi(x)$ is sharper than that obtained from [3.7.2] $(x \ge 0)$ if and only if $x \le 1.01$, approximately. The lower bound for $\Phi(x)$ in [3.7.1] is sharper than that above if and only if $x \ge 1.45$ approximately.

(b) $\Phi(x) \geq \frac{1}{2} + x/[2(\pi + 2x^2)]^{1/2}$, $x \geq 0$ (Chu, 1955, p. 264)

[*3.7.4*] D'Ortenzio (1965, pp. 4-7--with corrections):

$$\Phi(x) \leq \frac{1}{2} + [\{1 - \exp(-x^2)\}/4 - \{(1/4) - (2\pi)^{-1}\}x^2\exp(-x^2)]^{1/2}$$

$$\Phi(x) \geq \frac{1}{2} + [\{1 - \exp(-x^2/2)\}/4 + \{(2\pi)^{-1} - (1/8)\}x^2\exp(-x^2)]^{1/2}, \quad x \geq 0$$

The average of these bounds is accurate to two decimal places only ($0 \leq x \leq 2.5$).

[*3.7.5*] Bounds for R(x) of the form $\alpha/[\gamma x + (x^2 + \beta)^{1/2}]$:

(a) $R(x) < 4/[3x + (x^2 + 8)^{1/2}]$, $x > -1$ (Sampford, 1953, p. 132)

For large values of x (Boyd, 1959, pp. 44-46), a sharper upper bound is given by

$$3/[5x + (x^2 + 12)^{1/2}] + 1/[x + (x^2 + 4)^{1/2}]$$

(b) The sharpest bounds for R(x) of the form $2/[x + (x^2 + \beta)^{1/2}]$ are given by the inequalities (Mitrinović, 1970, p. 178)

$$2/[x + (x^2 + 4)^{1/2}] < R(x) < 2/[x + (x^2 + 8/\pi)^{1/2}], \quad x > 0$$

The lower bound is Birnbaum's lower bound in [3.7.1]. Only the upper bound converges to R(0), i.e., $\sqrt{\pi/2}$, as $x \to 0$.

(c) The sharpest bounds for R(x) of the form $\alpha/[\gamma x + (x^2 + \beta)^{1/2}]$ are given by the inequalities (Boyd, 1959)

$$\pi/[(\pi - 1)x + (x^2 + 2\pi)^{1/2}] < R(x) < \pi/[2x + \{(\pi - 2)^2x^2 + 2\pi\}^{1/2}], \quad x > 0$$

These bounds are better than those in (b) above; both converge to R(0) as $x \to 0$ (Mitrinović, 1970, pp. 178-179). The ratio (upper bound)/(lower bound) is never more than 1.02.

[*3.7.6*] Rational functions as bounds for R(x) (see [3.7.1]) (Gross and Hosmer, 1978, pp. 1354-1355):

(a) $x/(1 + x^2) < R(x) < (x^2 + 2)/[x(x^2 + 3)], \quad x > 0$

The percent relative error of the upper bound derived as an approximation to $1 - \Phi(x)$ decreases from 1.84 to 0.62 as $1 - \Phi(x)$ decreases from 0.025 to 0.005.

(b) Shenton (1954, p. 188) gives the following inequalities for positive x; some improve the lower bound in (a):

$$R(x) > [x + 2x/(x^4 + 6x^2 + 15)](x^2 + 1)^{-1}$$

$$(x^2 + 1)R(x) > x + 2(x^5 + 14x^3 + 75x)/(x^8 + 20x^6 + 150x^4 + 420x^2 + 525)$$

$$(x^3 + 3x)R(x) < x^2 + 2 - 6/(x^4 + 10x^2 + 35)$$

$$(x^3 + 3x)R(x) < x^2 + 2 - 6(x^4 + 18x^2 + 119)/(x^8 + 28x^6 + 294x^4 + 1260x^2 + 2205)$$

$$(x^4 + 6x^2 + 3)R(x) > x^3 + 5x + 4!x/(x^6 + 15x^4 + 105x^2 + 315)$$

$$(x^5 + 10x^3 + 15x)R(x) < x^4 + 9x^2 + 8 - 5!/(x^6 + 21x^4 + 189x^2 + 693)$$

3.8 QUANTILES

Suppose that Z is a rv with a N(0,1) distribution, and that $\Phi(z_p) = \Pr(Z \le z_p) = 1 - p$. In what follows, z_p is expressed in terms of a given value of p.

[3.8.1] Two approximations are given in Hastings (1955, pp. 191-192) (Abramowitz and Stegun, 1964, pp. 932-933):

(a) $$z_p = t - [(2.30753) + (0.27061)t]/[1 + (0.99229)t + (0.04481)t^2] + \varepsilon(p),\ 0 < p \le \tfrac{1}{2},\ t = (-2 \log p)^{1/2}$$

Then

$$|\varepsilon(p)| < 3 \times 10^{-3}$$

(b) $$z_p = t - [(2.515517) + (0.802853)t + (0.010328)t^2]/[1 + (1.432788)t + (0.189269)t^2 + (0.001308)t^3] + \varepsilon(p),\ 0 < p \le \tfrac{1}{2},\ t = (-2 \log p)^{1/2}$$

Then

$$|\varepsilon(p)| < 4.5 \times 10^{-4}$$

[3.8.2] Wetherill (1965, pp. 202-203) compared three approximations to z_p, noting that the quantiles in the central and the tail regions of $\Phi(\cdot)$ require different methods of approximation in order to achieve accuracy.

(a) $$z_p = (1.253314)w + (0.328117)w^3 + (0.180392)w^5 + (0.122403)w^7 + (0.113945)w^9 + \varepsilon(p), \quad 0 \le p \le \frac{1}{2}$$
$$w = 1 - 2p$$

Then

$$|\varepsilon(p)| < 1.6 \times 10^{-4} \quad \text{if } 0.2 < p < 0.3$$

(b) A satisfactory approximation when $|w| \le 0.91$ (corresponding to $p = 0.045$) is given by

$$z_p \simeq (0.328117)w^3 + (0.180392)w^5 + (0.3630)w^7 - (0.8559)w^9 + (1.0480)w^{11}$$

$$w = 1 - 2p, \quad 0 < p \le 0.5$$

Then

$$|\text{relative error}| < 0.002 \quad \text{if } 0.045 < p < 0.50$$

(c) If $0.00001 \le p \le 0.05$, then a satisfactory approximation is given by

$$z_p \simeq 0.401703 - (0.625600)v - (0.039811)v^2 - (0.001416)v^3, \quad 0 < p < 0.5, \quad v = \log[1 - |1 - 2p|]$$

[3.8.3] A simple approximation to z_p, with a maximum error of 0.02 when $0 \le z_p \le 4.0$, is given by (Hamaker, 1978, pp. 76-77)

$$z_p \simeq t - (0.50 + 0.30t)^{-1}, \quad 0 < p \le \frac{1}{2}, \ t = (-2 \log p)^{1/2}$$

[3.8.4] Hamaker (1978, p. 77) gives an approximation which holds when $0 < p < 1$, unlike those above, for which adjustments need to be made when $1/2 \le p < 1$:

$$z_p = \text{sgn}(p - 0.50)[1.238u\{1 + (0.0262)u\}], \qquad 0 < p < 1$$
$$u = \{-\log(4p - 4p^2)\}^{1/2}$$

This gives results almost as accurate as that of [3.8.1(a)].

[*3.8.5*] Derenzo (1977, pp. 217-218) gives a simple approximation to quantiles, with integer coefficients and useful for hand calculators. Let $\Phi(z_p) = 1 - p$ and $y = -\log(2p)$. Then if $1/2 > p > 10^{-7}$, or $0 < z_p < 5.2$,

$$z_p = \left[\frac{\{(4y + 100)y + 205\}y^2}{\{(2y + 56)y + 192\}y + 131}\right]^{1/2} + \varepsilon(p), \quad |\varepsilon(p)| < 1.3 \times 10^{-4}$$

If $10^{-7} > p > 1/2 \times 10^{-112}$, or $5.2 < z_p < 22.6$,

$$z_p = \left[\frac{\{(2y + 280)y + 572\}y}{(y + 144)y + 603}\right]^{1/2} + \varepsilon(p), \qquad |\varepsilon(p)| < 4 \times 10^{-4}$$

[*3.8.6*] Schmeiser (1979, pp. 175-176) compares two approximations z^* and $\hat{z}$ to z_p with a third approximation z' of Page (1977, pp. 75-76), where $\Phi(z_p) = 1 - p$. All may be used on pocket calculators, and are as follows:

(a) $z^* = \{(1 - p)^{0.135} - p^{0.135}\}/0.1975, \qquad 0 < p < 1$

(b) $\hat{z} = 0.20 + \{(1 - p)^{0.14} - p^{0.14}\}/0.1596, \quad 0.5 < 1 - p < 1$

(c) $z' = u - (0.134145u)^{-1}$

$$u = [\{y + (y^2 + 3.31316)^{1/2}\}/0.08943]^{1/3}$$
$$y = \log\{(1 - p)/p\}/1.59577, \qquad 0.5 < 1 - p < 1$$

If $1 - p < 0.5$, z_{1-p} can be approximated by (b) or (c) and the sign of $\hat{z}$ or of z' changed. If $|z_p| \le 2$, i.e., $p < 0.0228$, then z' is most accurate ($|z' - zp| \le 0.0011$) and $|z^* - z_p| \le 0.0093$; if $|z_p| > 2$, then $\hat{z}$ is most accurate, although the error increases in the tail of the distribution (if $2 < |z_p| \le 4$, $|\hat{z} - z_p| < 0.0145$) (Schmeiser, 1979).

3.9 APPROXIMATING THE NORMAL BY OTHER DISTRIBUTIONS

A number of writers have sought to fit other distributions to the normal, and then to approximate the normal cdf and quantiles from them. The more successful of these efforts are discussed in [3.9.3] and [3.9.4].

[3.9.1] Chew (1968, pp. 22-24) has fitted the uniform, triangular, cosine, logistic, and Laplace distributions, all symmetric about zero, to the standard normal, choosing parameter values to match the second moment, where possible. Near zero, none of these fits very well. Chew claims that beyond about one standard deviation the pdfs of these distributions are "fairly close" together with $\phi(x)$, but the relative errors may be quite large.

[3.9.2] Hoyt (1968, p. 25) approximates the N(0,1) pdf by that of a sum S of three iid rvs, each having a uniform distribution on (-1,1). The rv S has pdf g(s), given by

$$g(s) = \begin{cases} (3 - s^2)/8, & |s| \le 1 \\ (3 - |s|)^2/16, & 1 \le |s| \le 3 \\ 0, & |s| \ge 3 \end{cases}$$

The maximum error in the pdf is 0.024 when s = 0, and in the cdf is 0.010 when s = 0.60.

3.9.3 *<u>Approximation by Burr Distributions</u>*. Consider the family of distributions with cdf G given by

$$G(x) = 1 - (1 + x^c)^{-k}, \qquad x \ge 0,\ c > 0,\ k > 0$$

Burr (1967, p. 648) approximated $\Phi(x)$ by $G_1(x)$ and $G_2(x)$, where

$$G_1(x) = \begin{cases} 1 - [1 + \{(0.6446,93) + (0.1619,84)x\}^{4.874}]^{-6.158}, \\ \qquad\qquad\qquad\qquad x > -3.9799,80 \\ 0, \qquad\qquad\qquad\quad\ \ x < -3.9799,80 \end{cases}$$

$$G_2(x) = \{G_1(x) + 1 - G_1(-x)\}/2$$

Then $G_2(x)$ provides the better approximation. The largest difference between $G_2(x)$ and $\Phi(x)$ is 4.6×10^{-4}, when $x = \pm 0.6$.

If we solve the equation

$$G_2(y_p) = 1 - p$$

for y_p when p is given, we obtain an approximation to the quantile z_p, where $\Phi(z_p) = 1 - p$. Burr (1967) gives the approximation

$$z_p \simeq y_p \simeq [(p^{-1/6.158} - 1)^{1/4.874} - \{(1 - p)^{-1/6.158} - 1\}^{1/4.874}]/0.323968$$

When $0.008 < p < 1/2$, $|\text{error}| < 3.5 \times 10^{-3}$, and when $0.045 < p < 1/2$, the approximation compares favorably with that of Hastings (1955, p. 191) given in [3.8.1(a)].

[3.9.4] *Approximation by Weibull Distributions*. The Weibull family has cdf W(x) given by

$$W(x) = \Pr(X \le x) = 1 - \exp[-\{(x - a)/\theta\}^m], \quad x > a,\ \theta > 0,\ m > 0$$

where a, θ, and m are location, scale, and shape parameters, respectively. If $c \simeq 3.60$, the third central moment is zero, and if $c \simeq 2.20$, the kurtosis is 3.0, as for a normal pdf (Johnson and Kotz, 1970a, p. 253). Plait (1962, pp. 23, 26) found that the Weibull distribution for which $m = 3.25$, $a = 0$, and $\theta = 1$ fits an $N(0.8963, 0.303^2)$ distribution closely.

By equating several measures of skewness to zero, Dubey (1967, pp. 69-79) arrived at four approximations to $\Phi(x)$, of which we give two that appear to fit the greatest range of values of x in the interval [-3,3]. The approximations $W_1(x)$ and $W_2(x)$ following were derived by putting the measures (median-mode) and the third central moment μ_3, respectively, equal to zero.

	m	a	θ	x: preferred range
$W_1(x)$	3.25889	-2.96357	3.30589	[-1.5,-1.1],[0.2,0.9],[2.2,3.0]
$W_2(x)$	3.60232	-3.24311	3.59893	[-3.0,-1.6],[-1.0,0.1],[1.0,2.1]

The preferred range minimizes the |error|, although Dubey (1967, p. 78) finds a few values of x for which his other approximations do slightly better. The combination above, however, has a maximum |error| of 0.0078 when x = -0.9.

REFERENCES

The numbers in square brackets give the sections in which the corresponding reference is cited.

Abramowitz, M., and Stegun, I. A. (1964). *Handbook of Mathematical Functions*, Washington, D.C.: National Bureau of Standards. [3.1.3.3; 3.2.2; 3.3.2; 3.5.2, 3; 3.6.1; 3.8.1, 4]

Adams, A. G. (1969). Algorithm 39: Areas under the normal curve, *Computer Journal* 12, 197-198. [3.1.3.1]

Adams, W. J. (1974). *The Life and Times of the Central Limit Theorem*, New York: Caedmon. [3.2.1]

Ahrens, J. H., and Dieter, U. (1972). Computer methods for sampling from the exponential and normal distributions, *Communications of the Association for Computing Machinery* 15(10), 873-882. [3.1.3.3]

Ahrens, J. H., and Dieter, U. (1973). Extensions of Forsythe's method for random sampling from the normal distribution, *Mathematics of Computation* 27, 927-937. [3.1.3.3]

Altman, I. B. (1950). The effect of one-sided truncation on the average of a normal distribution, *Industrial Quality Control* 7(3), 30-31.

Andrews, D. F. (1973). A general method for the approximation of tail areas, *Annals of Statistics* 1, 367-372. [3.6.4]

Badhe, S. K. (1977). New approximation of the normal distribution function, *Communications in Statistics* B5, 173-176. [3.2.1; 3.4.2; 3.5.3]

Beasley, J. D., and Springer, S. G. (1977). Algorithm AS111: The percentage points of the normal distribution, *Applied Statistics* 26, 118-121. [3.1.3.2]

Birnbaum, Z. W. (1942). An equality for Mill's ratio, *Annals of Mathematical Statistics* 13, 245-246. [3.7.1, 2, 3]

Box, G. E. P., and Muller, M. E. (1958). A note on the generation of random normal deviates, *Annals of Mathematical Statistics* 29, 610-611. [3.1.3.3]

Boyd, A. V. (1959). Inequalities for Mills' ratio, *Reports of Statistical Application Research* (Union of Japanese Scientists and Engineers) 6, 44-46. [3.7.5]

Burr, I. W. (1967). A useful approximation to the normal distribution function, with application to simulation, *Technometrics* 9, 647-651. [3.9.3]

Carta, D. G. (1975). Low-order approximations for the normal probability integral and the error function, *Mathematics of Computation* 29, 856-862. [3.2.2]

Chew, V. (1968). Some useful alternatives to the normal distribution, *The American Statistician* 22(3), 22-24. [3.9.1]

Chu, J. T. (1955). On bounds for the normal integral, *Biometrika* 42, 263-265. [3.7.3]

Clark, F. E. (1957). Truncation to meet requirements on means, *Journal of the American Statistical Association* 52, 527-536. [3.1.1]

Derenzo, S. E. (1977). Approximations for hand calculators using small integer coefficients, *Mathematics of Computation* 31, 214-222. [3.2.3; 3.8.5]

D'Ortenzio, R. J. (1965). Approximating the normal distribution function, *Systems Design* 9, 4-7. [3.7.4]

Dubey, S. D. (1967). Normal and Weibull distributions, *Naval Research Logistic Quarterly* 14, 69-79. [3.9.4]

Feller, W. (1957, 1968). *An Introduction to Probability Theory and Its Applications*, Vol. 1 (2nd, 3rd eds.), New York: Wiley. [3.2.3; 3.5.1]

Fletcher, A., Miller, J. C. P., Rosenhead, L., and Comrie, L. J. (1962a). *An Index of Mathematical Tables*, Vol. I (2nd ed.), Reading, Mass.: Addison-Wesley/Scientific Computing Service. [3.1.1]

Fletcher, A., Miller, J. C. P., Rosenhead, L., and Comrie, L. J. (1962b). *An Index of Mathematical Tables*, Vol. II (2nd ed.), Reading, Mass.: Addison-Wesley/Scientific Computing Service. [3.1.1]

Gordon, R. D. (1941). Values of Mills' ratio of area to bounding ordinate and of the normal probability integral for large values of the argument, *Annals of Mathematical Statistics* 12, 364-366. [3.6.5; 3.7.1, 2, 3]

Gray, H. L., and Schucany, W. R. (1968). On the evaluation of distribution functions, *Journal of the American Statistical Association* 63, 715-720. [3.5.5; 3.6.3]

Greenwood, J. A., and Hartley, H. O. (1962). *Guide to Tables in Mathematical Statistics*, Princeton, N.J.: Princeton University Press. [3.1.1]

Gross, A. J., and Hosmer, D. W. (1978). Approximating tail areas of probability distributions, *Annals of Statistics* 6, 1352-1359. [3.7.6]

Gupta, S. S., and Waknis, M. N. (1965). A system of inequalities for the incomplete gamma function and the normal integral, *Annals of Mathematical Statistics* 36, 139-149. [3.5.2]

Hamaker, H. C. (1978). Approximating the cumulative normal distribution and its inverse, *Applied Statistics* 27, 76-77. [3.2.3; 3.8.3, 4]

Hart, R. G. (1966). A close approximation related to the error function, *Mathematics of Computation* 20, 600-602. [3.4.2; 3.6.2, 3]

Hastings, C. (1955). *Approximations for Digital Computers*, Princeton, N.J.: Princeton University Press. [3.1.3.1; 3.2.2; 3.3.2; 3.4.2; 3.6.1; 3.8.1, 4; 3.9.3]

Hill, I. D. (1969). Remark ASR2: A remark on Algorithm AS2 "The normal integral," *Applied Statistics* 18, 299-300. [3.1.3.1]

Hill, I. D., and Joyce, S. A. (1967). Algorithm 304: Normal curve integral, *Communications of Association for Computing Machinery* 10(6), 374-375. [3.1.3.1]

Hoyt, J. P. (1968). A simple approximation to the standard normal probability density function, *The American Statistician* 22(2), 25-26. [3.9.2]

Ibbetson, D. (1963). Algorithm 209: Gauss, *Communications of the Association for Computing Machinery* 6, 616. [3.1.3.1]

IMSL Library 3, Edition 6 (1977). IMSL Inc., 7500 Bellaire Blvd., 6th floor, GNB Bldg., Houston, Tex. 77036. [3.1.3.1]

Johnson, N. L., and Kotz, S. (1970a). *Distributions in Statistics--Continuous Univariate Distributions*, Vol. 1, New York: Wiley. [3.1.1; 3.9.4]

Johnson, N. L., and Kotz, S. (1970b). *Distributions in Statistics--Continuous Univariate Distributions*, Vol. 2, New York: Wiley. [3.7]

Kelley, T. L. (1948). *The Kelley Statistical Tables*, Cambridge, Mass.: Harvard University Press. [3.1.1]

Kendall, M. G., and Stuart, A. (1977). *The Advanced Theory of Statistics*, Vol. 1 (4th ed.), New York: Macmillan. [2.2.1; 2.4.1; 2.5.1]

Kerridge, D. F., and Cook, G. W. (1976). Yet another series for the normal integral, *Biometrika* 63, 401-403. [3.2.1]

Kinderman, A. J., and Ramage, J. G. (1976). Computer generation of normal random variables, *Journal of the American Statistical Association* 71, 893-896. [3.1.3.3]

Laplace, P. S. (1785). Mémoire sur les approximations des formules qui sont fonctions de très-grands nombres, *Histoire de l'Académie Royale des Sciences de Paris*, 1-88; reprinted in *Oeuvres Complètes*, Vol. 10, 209-291. [3.2.1; 3.5.1, 2]

Laplace, P. S. (1805). *Traité de Mécanique Celeste*, Vol. 4, Paris. [3.2.1; 3.4.1]

Laplace, P. S. (1812). *Théorie Analytique de Probabilitês*, Vol. 2, Paris: Courcier. [3.2.1; 3.5.2]

Marsaglia, G., and Bray, T. A. (1964). A convenient method for generating normal variables, *SIAM Review* 6(3), 260-264. [2.1.3.3]

Marsaglia, G., MacLaren, M. D., and Bray, T. A. (1964). A fast procedure for generating normal random variables, *Communications of the Association for Computing Machinery* 7(1), 4-10. [3.1.3.2, 3]

Mitrinović, D. S. (1970). *Analytic Inequalities*, New York: Springer-Verlag. [3.7]

Moran, P. A. P. (1980). Calculation of the normal distribution function, *Biometrika* 67, 675-676.

Odeh, R. E., and Evans, J. O. (1974). Algorithm AS70: The percentage points of the normal distribution, *Applied Statistics* 23, 96-97. [3.1.3.2]

Owen, D. B. (1962). *Handbook of Statistical Tables*, Reading, Mass.: Addison-Wesley. [3.1]

Page, E. (1977). Approximations to the cumulative normal function and its inverse for use on a pocket calculator, *Applied Statistics* 26, 75-76. [3.2.3; 3.8.6]

Patry, J., and Keller, J. (1964). Zur berechnung des fehlerintegrals, *Numerische Mathematik* 6, 89-97. [3.4.2; 3.5.3]

Pearson, E. S., and Hartley, H. O. (1966). *Biometrika Tables for Statisticians*, Vol. 1 (3rd ed.), London: Cambridge University Press. [3.1; 3.6.5]

Pearson, E. S., and Hartley, H. O. (1972). *Biometrika Tables for Statisticians*, Vol. 2, London: Cambridge University Press. [3.1]

Plait, A. (1962). The Weibull distribution--with tables, *Industrial Quality Control* 19(5), 17-26. [3.9.4]

Pólya, G. (1949). Remarks on computing the probability integral in one and two dimensions, *Proceedings of the First Berkeley Symposium on Mathematical Statistics and Probability*, 63-78, Berkeley: University of California Press. [3.2.3; 3.5.2; 3.7.3]

Rabinowitz, P. (1969). New Chebyshev polynomial approximations to Mills' ratio, *Journal of the American Statistical Association* 64, 647-654. [3.5.10]

Rahman, N. A. (1968). *A Course in Theoretical Statistics*, New York: Hafner. [3.5.1]

Rao, C. R., Mitra, S. K., and Matthai, A., eds. (1966). *Formulae and Tables for Statistical Work*, Calcutta: Statistical Publishing Society. [3.1]

Ray, W. D., and Pitman, A. E. N. T. (1963). Chebyshev polynomial and other new approximations to Mills' ratio, *Annals of Mathematical Statistics* 34, 892-902. [3.5.7, 8, 9, 10]

Ruben, H. (1962). A new asymptotic expansion for the normal probability integral and Mills' ratio, *Journal of the Royal Statistical Society* B24, 177-179. [3.5.4]

Ruben, H. (1963). A convergent asymptotic expansion for Mills' ratio and the normal probability integral in terms of rational functions, *Mathematische Annalen* 151, 355-364. [3.5.5, 6; 3.6.3]

Ruben, H. (1964). Irrational fraction approximations to Mills' ratio, *Biometrika* 51, 339-345. [3.5.5, 6]

Sampford, M. R. (1953). Some inequalities on Mills' ratio and related functions, *Annals of Mathematical Statistics* 24, 130-132. [3.7.5]

Schmeiser, B. (1979). Approximations to the inverse cumulative normal function for use on hand calculators, *Applied Statistics* 28, 175-176. [3.8.6]

Schucany, W. R., and Gray, H. L. (1968). A new approximation related to the error function, *Mathematics of Computation* 22, 201-202. [3.4.2; 3.6.3]

Sengupta, J. M., and Bhattacharya, N. (1958). Tables of random normal deviates, *Sankhȳa* 20, 250-286. [3.1.3.3]

Shenton, L. R. (1954). Inequalities for the normal integral including a new continued fraction, *Biometrika* 41, 177-189. [3.2.1; 3.4.1; 3.5.2, 3; 3.7.6]

Smirnov, N. V. (1965). *Tables of the Normal Probability Integral, the Normal Density, and its Normalized Derivatives*, New York: Macmillan; translated from the Russian (1960) ed., Moscow: Academy of Sciences of the USSR. [3.1]

Tate, R. F. (1953). On a double inequality of the normal distribution, *Annals of Mathematical Statistics* 24, 132-134. [3.7.2, 3]

Tocher, K. D. (1963). *The Art of Simulation*, London: English Universities Press. [3.2.3]

Varnum, E. C. (1950). Normal area nomogram, *Industrial Quality Control* 6(4), 32-34. [3.1.2]

Wetherill, G. B. (1965). An approximation to the inverse normal function suitable for the generation of random normal deviates on electronic computers, *Applied Statistics* 14, 201-205. [3.8.2]

White, J. S. (1970). Tables of Normal Percentile Points, *Journal of the American Statistical Association* 65, 635-638. [3.1.1]

Wold, H. (1948). *Random normal deviates*, Tracts for Computers, Vol. 25, London: Cambridge University Press. [3.1.3.3]

Yamauti, Z. (ed.) (1972). *Statistical Tables and Formulas with Computer Applications*, Tokyo: Japanese Standards Association. [3.1]

chapter 4

CHARACTERIZATIONS

If a normally distributed rv, or if a random sample from a normal distribution, has some property P, then it may be of interest to know whether such a property characterizes the normal law. Serious research in this area has been done since 1935, with most results dating from after 1950; we have tended to concentrate on those results which might be of practical interest, and users may therefore find some of the special cases of more concern than the most general theorems. Some of the more mathematical and theoretical characterizations have been omitted, but can be found in the books and expository papers on the subject. These include Johnson and Kotz (1970, pp. 50-53); Kendall and Stuart (1977, pp. 388-392); Kotz (1974, pp. 39-65); Lukacs (1956, pp. 195-214); Lukacs and Laha (1964, chaps. 5 and 10); Mathai and Pederzoli (1977); and Kagan et al. (1973). The review of the last-named by Diaconis et al. (1977, pp. 583-592) is well worth reading also, giving as it does an excellent summary and exposition of the field.

While characterizations of the exponential distribution largely involve order statistics (see [4.8], however), those of the normal law involve distributional or independence properties of linear and quadratic forms to a great degree. Linear forms alone are considered in [4.1], and along with quadratic forms in [4.2]; these are based on sets of independent and sometimes iid random variables. Characterizations based on conditional distributions and regression properties are also mainly based on linear and quadratic forms; these are given in [4.3], and users may note that the conditioning statistic in all of these characterizations is invariably linear. Special topics are covered in [4.5] (characteristic functions), [4.6] (polar coordinate transformations), [4.7] (characterizations based on sufficiency, information, and concepts from decision theory and statistical inference), and [4.8] (order statistics). The question of how close to a characterizing property a distribution can be and be nearly normal is touched upon in [4.9] (but see also Diaconis et al., 1977, pp. 590-591).

Several characterizations of the bivariate normal distribution appear in [10.1.5]; others belong as special cases of characterizations of the multivariate normal distribution, which lies outside the scope of this book. References for these multivariate characterizations include Kagan et al. (1973), Mathai and Pederzoli (1977), Johnson and Kotz (1972, pp. 59-62), Khatri and Rao (1972, pp. 162-173), Khatri and Rao (1976, pp. 81-94), and Khatri (1979, pp. 589-598).

4.1 CHARACTERIZATION BY LINEAR STATISTICS

[*4.1.1*] (a) Let X_1 and X_2 be independent nondegenerate rvs, not necessarily identically distributed; let $Y = X_1 + X_2$. Then Y is normally distributed if and only if X_1 and X_2 are normally distributed (Mathai and Pederzoli, 1977, p. 6; Cramér, 1936, pp. 405-414).

(b) The sum $X_1 + \cdots + X_n$ of n independent rvs (n = 2, 3, ...) is normally distributed if and only if each component variable X_1, X_2, ..., X_n is itself normally distributed (Cramér, 1946, pp. 212-213).

(c) Let X_1, X_2, ..., X_n be a random sample from a certain population, and let $L = a_1X_1 + a_2X_2 + \cdots + a_nX_n$, a linear statistic, where a_1, ..., a_n are nonzero constants. Then the population is normal if and only if L is normally distributed (Lukacs, 1956, p. 198).

(d) Let $X_1, X_2, \ldots, X_n$ be a random sample from a certain population and let constants $a_1, \ldots, a_n$ be given, such that

$$A_s = \sum_{j=1}^{n} a_j^s \neq 0, \qquad s = 1, 2, 3, \ldots$$

for all s. Then the statistic $a_1X_1 + \cdots + a_nX_n$ is normally distributed with mean μ and variance σ^2 if and only if each of X_1, ..., X_n is normally distributed with mean μ/A_1 and variance σ^2/A_2 (Lukacs, 1956, pp. 197-198).

[*4.1.2*] Let $X_1, X_2, \ldots, X_n$ be a random sample from a certain population, and suppose that all absolute moments $E(|X_i|^k)$ exist ($i = 1, \ldots, n$; $k = 1, 2, \ldots$). Consider the linear statistics

$$L_1 = a_1X_1 + \cdots + a_nX_n, \qquad L_2 = b_1X_1 + \cdots + b_nX_n$$

Suppose that $\{|a_1|,\ldots,|a_n|\}$ is not a permutation of $\{|b_1|,\ldots,|b_n|\}$, and that $\Sigma_{i=1}^n a_i = \Sigma_{i=1}^n b_i$, $\Sigma_{i=1}^n a_i^2 = \Sigma_{i=1}^n b_i^2$. Then the parent population is normally distributed if and only if L_1 and L_2 are identically distributed (Lukacs, 1956, p. 210).

Marcinkiewicz (1939, pp. 612-618) gives a result which generalizes this to infinite sums.

[*4.1.3*] Let $X_1, \ldots, X_n$ be a random sample from a certain population. Then every linear statistic $a_1X_1 + \cdots + a_nX_n$ is distributed like the rv $(a_1^2 + \cdots + a_n^2)^{1/2}X_1$ if and only if the parent population is normal. This is a special case of a theorem of Lukacs (1956, p. 200). Shimizu (1962, pp. 173-178) has a similar result, requiring the assumption of finite variances.

[*4.1.4*] (a) Let X_1 and X_2 be independent rvs with finite variances. Then the sum $X_1 + X_2$ and the difference $X_1 - X_2$ are independent if and only if X_1 and X_2 are normally distributed (Bernstein, 1941, pp. 21-22).

(b) Let $X_1, \ldots, X_n$ be a sample from a certain population and consider the linear statistics

$$L_1 = a_1X_1 + \cdots + a_nX_n, \qquad L_2 = b_1X_1 + \cdots + b_nX_n$$

such that $\Sigma_{j=1}^{n} a_j b_j = 0$, while $\Sigma_{j=1}^{n} (a_j b_j)^2 \neq 0$. Then the parent population is normal if and only if L_1 and L_2 are independently distributed (Lukacs, 1956, p. 205). This result does not require any assumption of the existence of moments.

(c) Let $X_1, \ldots, X_n$ be mutually independent rvs (not necessarily identically distributed). Let L_1 and L_2 be linear statistics $\Sigma_{i=1}^{n} a_i X_i$ and $\Sigma_{i=1}^{n} b_i X_i$, where $a_1, \ldots, a_n, b_1, \ldots, b_n$ are constant coefficients.

If L_1 and L_2 are independent, then the rvs X_i for which the product $a_i b_i \neq 0$ are all normally distributed (Kagan et al., 1973, p. 89; Mathai and Pederzoli, 1977, p. 25; Darmois, 1951, pp. 1999-2000; Skitovitch, 1953, pp. 217-219). The form of the converse is given by: If $\Sigma_{j=1}^{n} \sigma_j^2 a_j b_j = 0$ [where $\mathrm{Var}(X_j) = \sigma_j^2$] and those X_i are normally distributed for which $a_j b_j \neq 0$, then L_1 and L_2 are independent.

[4.1.5] An orthogonal transformation of n iid normally distributed rvs preserves mutual independence and normality. Lancaster (1954, p. 251) showed that this property characterizes the normal distribution.

Let $X_1, \ldots, X_n$ be mutually independent rvs each with zero mean and variance one, and let n nontrivial linear transformations

$$Y_j = \sum_{i=1}^{n} a_{ij} X_j$$

be made, which exclude those of the type $X_i + a = cY_j$, and such that $Y_1, \ldots, Y_n$ are mutually independent. Then $X_1, \ldots, X_n$ are normally distributed and the transformation is orthogonal (Kendall and Stuart, 1977, p. 391).

[4.1.6] In the following result we consider an infinite number of variables (Kagan et al., 1973, p. 94; Mathai and Pederzoli, 1977, p. 30).

Let $X_1, X_2, \ldots$ be a sequence of independent rvs and $\{a_j\}$, $\{b_j\}$ be two sequences of real constants such that

(i) The sequences $\{a_j/b_j : a_j b_j \neq 0\}$ and $\{b_j/a_j : a_j b_j \neq 0\}$ are both bounded.

(ii) $\Sigma_{j=1}^{\infty} a_j X_j$ and $\Sigma_{j=1}^{\infty} b_j X_j$ converge with probability one to rvs U and V, respectively.

(iii) U and V are independent.

Then for every j such that $a_j b_j \neq 0$, X_j is normally distributed.

4.2 LINEAR AND QUADRATIC CHARACTERIZATIONS

[*4.2.1*] (a) Let $X_1, X_2, \ldots, X_n$ be a random sample from a certain population; let $\bar{X} = \Sigma X_i/n$, the sample mean, and let $S^2 = \Sigma(X_i - \bar{X})^2/(n - 1)$, the sample variance. For $n \geq 2$, a necessary and sufficient condition for the independence of $\bar{X}$ and S^2 is that the parent population be normal (Kagan et al., 1973, p. 103; Kendall and Stuart, 1977, pp. 308-309; Lukacs, 1956, p. 200). This result was first obtained by Geary (1936, pp. 178-184) and by Lukacs (1942, pp. 91-93) under more restrictive conditions involving the existence of moments.

(b) Let $X_1, X_2, \ldots, X_n$ be iid random variables with common finite mean and variance. Let the first-order mean square successive difference D_k^2 be defined by

$$D_k^2 = \frac{1}{2}(n - k)^{-1} \sum_{i=1}^{n-k} (X_{i+k} - X_i)^2, \qquad k = 1, 2, \ldots, n - 1$$

and let $\bar{X} = \Sigma X_i/n$. Then a necessary and sufficient condition for $\bar{X}$ and D_k^2 to be independent is that the parent population is normal (Geisser, 1956, p. 858).

(c) The following theorem by Rao (1958, p. 915) generalizes results in (a) and (b) above. Let $X_1, \ldots, X_n$ be iid random variables with common finite mean and variance, and let

$$\delta^2 = \left(\sum_{t=1}^{m} \sum_{j=1}^{n} h_{tj}^2\right)^{-1} \sum_{t=1}^{m} (h_{t1}X_1 + \cdots + h_{tn}X_n)^2, \qquad m \geq 1$$

where $\Sigma_{j=1}^{n} h_{tj} = 0$ when $t = 1, 2, \ldots, m$. Then a necessary and sufficient condition for the parent population to be normal is that $\bar{X}$ and δ^2 are independent. See also Laha (1953, pp. 228-229).

In (a), $m = n$ and $h_{jj} = 1 - n^{-1}$; $h_{tj} = -n^{-1}$ if $t \neq j$.

In (b), $m = n - k$ and $h_{jj} = -1$, $h_{tj} = 1$ if $j = t + k$ and $h_{tj} = 0$ otherwise.

(d) Let $X_1, \ldots, X_n$ be iid random variables ($n \geq 4$), and let $Y = (X_1 - X_2)/S$. If Y is stochastically independent of the pair $(\bar{X}, S)$, where $\bar{X} = \Sigma X_i/n$ and $S^2 = \Sigma(X_i - \bar{X})^2/(n - 1)$, then the parent population is normal (Kelkar and Matthes, 1970, p. 1088).

(e) Kaplansky (1943, p. 197) weakened the conditions for the characterization in (a). Let $X_1, \ldots, X_n$ be a random sample from a continuous population with sample mean $\bar{X}$ and sample variance S^2. Then the joint pdf of $X_1, \ldots, X_n$ is a function $h(\bar{x}, s)$ of the values of $\bar{X}$ and S and also the joint pdf of $\bar{X}$ and S is given by $h(\bar{x}, s)s^{n-2}$ if and only if the parent population is normal ($n \geq 3$).

[*4.2.2*] (a) Let $X_1, \ldots, X_n$ be a random sample from a certain population. Let $L = \Sigma_{i=1}^n a_i X_i$ and $Q = \Sigma_{i=1}^n X_i^2 - L^2$, where $\Sigma_{i=1}^n a_i^2 = 1$. Then a necessary and sufficient condition for the independence of L and Q is that the parent population is normal (Kagan et al., 1973, pp. 105-106; Mathai and Pederzoli, 1977, p. 35).

(b) Let $X_1, \ldots, X_n$ be a random sample from a population where the second moment exists, and let $\bar{X} = \Sigma X_i/n$, $Q = \Sigma_{i=1}^n \Sigma_{j=1}^n a_{ij} X_i X_j$, where $\Sigma_{i=1}^n a_{ii} \neq 0$ and $\Sigma_{i=1}^n \Sigma_{j=1}^n a_{ij} = 0$. Then $\bar{X}$ and Q are independent if and only if the parent population is normal (Mathai and Pederzoli, 1977, p. 38; Kagan et al., 1973, pp. 106-107). See also Lukacs and Laha (1964, p. 81).

(c) Let $X_1, \ldots, X_n$ be iid random variables, and let

$$Q = \sum_{j=1}^{n} \sum_{k=1}^{n} a_{jk} X_j X_k + \sum_{j=1}^{n} b_j X_j$$

such that $\Sigma_{j=1}^n a_{jj} \neq 0$ and $\Sigma_{j=1}^n \Sigma_{j=1}^n a_{jk} = 0$. Then $X_1 + \cdots + X_n$ and Q are independent if and only if

(i) The parent population is normal,

(ii) $\Sigma_{j=1}^n a_{jk} = 0$ when $k = 1, \ldots, n$, and

(iii) $\Sigma_{j=1}^{n} b_j = 0$ (Lukacs and Laha, 1964, p. 100).

[*4.2.3*] Based on a standard normal sample, the distribution of $\Sigma_1^n (X_i + a_i)^2$ depends on $a_1, \ldots, a_n$ through the noncentral chi-square noncentrality parameter $a_1^2 + \cdots + a_n^2$; see [5.3.6]. It turns out that this property is a characterization of the normal law.

(a) If $X_1, \ldots, X_n$ are iid random variables ($n \geq 2$), and the distribution of the statistic $\Sigma_{i=1}^n (X_i + a_i)^2$ depends on $a_1, \ldots, a_n$ only through $\Sigma_{i=1}^n a_i^2$, where $a_1, \ldots, a_n$ are real, then the parent population is normal (Kagan et al., 1973, p. 453).

(b) If the random vector $(X_1,\ldots,X_m)$ is independent of the random vector $(X_{m+1},\ldots,X_n)$, $1 \leq m < n$, and if for arbitrary constants $a_1, \ldots, a_n$, the distribution of $\Sigma_{j=1}^n (X_j + a_j)^2$ depends on $a_1, \ldots, a_n$ only through $\Sigma_{j=1}^n a_j^2$, then $X_1, \ldots, X_n$ are mutually independent and normally distributed with zero means and common variance (Kotz, 1974, p. 48).

[*4.2.4*] The following related results are due to Geisser (1973, pp. 492-494).

(a) Let X and Y be independent rvs, and let X have a N(0,1) distribution. Then the rv $(aX + bY)^2/(a^2 + b^2)$ has a chi-square distribution with one degree of freedom (χ_1^2) for some nonzero a and b if and only if Y^2 is χ_1^2.

(b) Let X and Y be independent with X^2 and Y^2 each having a χ_1^2 distribution. Then the rv $(aX + bY)^2/(a^2 + b^2)$ has a χ_1^2 distribution for some nonzero a and b if and only if at least one of X and Y has a N(0,1) distribution.

(c) Let X and Y be iid random variables. Then X and Y are N(0,1) rvs if and only if, for some nonzero a and b, the rvs $(aX + bY)^2/(a^2 + b^2)$ and $(aX - bY)^2/(a^2 + b^2)$ are χ_1^2 rvs.

[*4.2.5*] The distribution of the sample variance provides a characterization of normality.

(a) Let $X_1, \ldots, X_n$ ($n \geq 2$) be a random sample from a nondegenerate symmetric distribution with finite variance σ^2 and sample mean $\bar{X}$. Then $\Sigma_{i=1}^{n}(X_i - \bar{X})^2/\sigma^2$ is distributed as chi-square with (n - 1) degrees of freedom (χ_{n-1}^2) if and only if the parent distribution is normal (Ruben, 1974, p. 379).

(b) Let $X_1, X_2, \ldots$ be an iid sequence of nondegenerate rvs with finite variance σ^2, and let m and n be distinct integers not less than 2. Denote $\Sigma_{i=1}^{m} X_i/m$ by $\bar{X}_m$, and $\Sigma_{i=1}^{n} X_i/n$ by $\bar{X}_n$. Then $\Sigma_{i=1}^{m}(X_i - \bar{X}_m)^2/\sigma^2$ and $\Sigma_{i=1}^{n}(X_i - \bar{X}_n)^2/\sigma^2$ are distributed as χ_{m-1}^2 and χ_{n-1}^2, respectively, if and only if the parent distribution is normal (Ruben, 1975, p. 72; Bondesson, 1977, pp. 303-304).

(c) Consider a vector of observations following the usual assumptions of the general linear model (Rao, 1973, chap. 4), having iid and symmetrically distributed error components. Let σ^2 be the error variance and ν the degrees of freedom of the residual sum of squares (RSS). Then the distribution of (RSS/σ^2) is χ_ν^2 if and only if the error components are normally distributed. See Ruben (1976, p. 186) for a formal statement and proof.

4.3 CHARACTERIZATIONS BY CONDITIONAL DISTRIBUTIONS AND REGRESSION PROPERTIES

[*4.3.1*] Let X and Y be rvs, each with a nonzero pdf at zero. If the conditional distribution of X, given X + Y = x + y, is the normal distribution with mean $\frac{1}{2}(x + y)$ for all x and y, then X and Y are iid normal rvs (Patil and Seshadri, 1964, p. 289).

[*4.3.2*] Let X_1, X_2 be iid random variables with mean zero. If $E(X_1 - \alpha X_2 | X_1 + \beta X_2) = 0$ and $E(X_1 + \beta X_2 | X_1 - \alpha X_2) = 0$, where $\alpha \neq 0$, $\beta \neq 0$, then X_1 and X_2 are normal (possibly degenerate) when $\beta\alpha = 1$, and degenerate otherwise (Kagan et al., 1973, p. 161; see also p. 158).

[*4.3.3*] Let $X_1, \ldots, X_k$ be independent rvs and a_j, b_j (j = 1, ..., k) be nonzero constants such that, when $i \neq j$, $a_i b_i^{-1} + a_j b_j^{-1} \neq 0$. If the conditional distribution of $\Sigma_{j=1}^{k} a_j X_j$, given $\Sigma_{j=1}^{k} b_j X_j$, is symmetric, then $X_1, \ldots, X_k$ are normal rvs (possibly degenerate). If the characteristic function of X_j is $\exp(itA_j - B_j t^2)$, with A_j real and B_j nonnegative, then $\Sigma_{j=1}^{k} A_j a_j = 0$ and $\Sigma_{j=1}^{k} B_j a_j b_j = 0$ (Kagan et al., 1973, p. 419).

[*4.3.4*] (a) *The Kagan-Linnik-Rao Theorem* (Kagan et al., 1973, p. 155). Let $X_1, \ldots, X_n$ be iid rvs ($n \ge 3$) with mean zero, and let $\bar{X} = \Sigma X_i/n$. Then if

$$E(\bar{X}|X_1 - \bar{X}, \ldots, X_n - \bar{X}) = 0$$

the parent population is normal.

An alternative condition, given by Kagan et al. (1965, p. 405), is that

$$E(\bar{X}|X_2 - X_1, X_3 - X_1, \ldots, X_n - X_1) = 0$$

If $n = 2$, this property holds for any symmetric parent distribution.

(b) Let $X_1, \ldots, X_n$ ($n \ge 3$) be mutually independent rvs, each with mean zero, and let $L_1, \ldots, L_n$ be linearly independent linear functions of $X_1, \ldots, X_n$, with all the coefficients in L_1 nonzero. Then, if $E(L_1|L_2, \ldots, L_n) = 0$, the rvs $X_1, \ldots, X_n$ are all normally distributed (Kagan et al., 1973, p. 156). This is a generalization of the result in (a).

(c) Let $X_1, \ldots, X_n$ be iid nondegenerate rvs, and $L_i = \Sigma_{j=1}^n a_{ij}X_j$ ($i = 1, \ldots, n - 1$) be linearly independent linear functions of $X_1, \ldots, X_n$, while $L_n = \Sigma_j a_{nj}X_j$ is a linear form such that the vector $(a_{ni},\ldots,a_{nn})$ is not a multiple of any vector with components 0, 1, -1. Then, from the conditions $E(L_i|L_n) = 0$ ($i = 1, \ldots, n - 1$), it follows that the parent population is normal if and only if $\Sigma_{j=1}^n a_{ij}a_{nj} = 0$ ($i = 1, \ldots, n - 1$) (Kagan et al., 1973, p. 161).

(d) Cacoullos (1967b, p. 1897) gives a result analogous to that in (c); $X_1, \ldots, X_n$ are iid random variables with mean zero and positive finite variance. The forms $L_1, \ldots, L_{n-1}$ are linearly independent statistics defined as in (c). Define $L_n = b_1X_1 + \cdots + b_nX_n$ (with no restrictions on the coefficients $b_1, \ldots, b_n$); then if $E(L_i|L_n) = 0$ ($i = 1, \ldots, n - 1$), the parent distribution is normal.

[*4.3.5*] (a) The following result is related to the Darmois-Skitovitch theorem of [4.1.4(c)]; it weakens the condition of independence of the linear statistics; on the other hand, it requires the assumption of finite variances.

Let $X_1, \ldots, X_n$ be a random sample from a population with mean zero and variance σ^2. Let linear statistics

$$Y_1 = a_1X_1 + \cdots + a_nX_n \qquad Y_2 = b_1X_1 + \cdots + b_nX_n$$

be defined, such that $a_n \neq 0$, $|b_n| > \max(|b_1|, \ldots, |b_{n-1}|)$, and $E(Y_1|Y_2) = 0$.

If $\sigma^2 < \infty$, $\Sigma_{i=1}^n a_ib_i = 0$ (or $\sigma^2 > 0$), and $a_ib_i/(a_nb_n) < 0$ ($i = 1, \ldots, n - 1$), then the parent population is normal (Rao, 1967, p. 5).

(b) A special case of (a) is as follows (Rao, 1967, pp. 1-2, 6). Let $X_1, \ldots, X_n$ be a random sample from a population with mean zero and finite variance ($n \geq 3$), and let $\bar{X}$ be the sample mean. Then if, for *any fixed* value of i ($1 \leq i \leq n$),

$$E(\bar{X}|X_i - \bar{X}) = 0$$

the parent population is normal.

(c) Pathak and Pillai (1968, p. 142) give a characterization which removes the assumption of a finite variance in (a). Let X_0, $X_1, \ldots, X_n$ be $n + 1$ iid random variables with common mean zero. Let

$$Y_1 = X_0 - a_1X_1 - \cdots - a_nX_n \qquad Y_2 = X_0 + b_1X_1 + \cdots + b_nX_n$$

where $a_ib_i > 0$ and $|b_i| < 1$ ($i = 1, \ldots, n$), and such that $E(Y_1|Y_2) = 0$. Then the parent distribution is normal if and only if $\Sigma_{i=1}^n a_ib_i = 1$.

(d) Rao (1967, p. 8) extended the characterization in (a) as follows. Let $X_1, \ldots, X_n$ ($n \geq 3$) be mutually independent (but not necessarily identically distributed) rvs, and suppose that there exist n linear statistics

$$Y_i = a_{i1}X_1 + \cdots + a_{in}X_n, \qquad i = 1, \ldots, n$$

such that the determinant $|(a_{ij})| \neq 0$ and $a_{11}, \ldots, a_{1n}$ are all nonzero. If $E(X_i) = 0$ ($i = 1, \ldots, n$) and $E(Y_1|Y_2, \ldots, Y_n) = 0$, then $X_1, \ldots, X_n$ are all normally distributed.

[*4.3.6*] The following result may be compared with that in [4.1.6]. Let $X_1, X_2, \ldots$ be a sequence of iid nondegenerate rvs with mean zero and having moments of all orders. Suppose that $\{a_i\}$ and $\{b_i\}$ are sequences of real constants such that $\Sigma_{i=1}^{\infty} |a_i|$ converges, $\Sigma_{i=1}^{\infty} b_i X_i$ converges with probability one, and

$$E\left(\sum_{i=1}^{\infty} a_i X_i \,\middle|\, \sum_{i=1}^{\infty} b_i X_i\right) = 0$$

Then the parent population is normal (Kagan et al., 1973, p. 190).

[*4.3.7*] *Linear Regression and Homoscedasticity*. Let $X_1, \ldots, X_n$ be mutually independent rvs with finite variances σ_i^2 (1 = 1, ..., n). Let

$$L_1 = a_1 X_1 + \cdots + a_n X_n \qquad L_2 = b_1 X_1 + \cdots + b_n X_n$$

where $a_i b_i \neq 0$ (i = 1, ..., n). Then

$$E(L_1 | L_2) = \alpha + \beta L_2 \quad \text{and} \quad \text{Var}(L_1 | L_2) = \sigma_0^2 \qquad \text{(constant)}$$

for some constants α and β, if and only if

(i) X_i is normal whenever $b_i \neq \beta a_i$, and
(ii) $\beta = (\Sigma' a_i b_i \sigma_i^2)/(\Sigma' a_i^2 \sigma_i^2)$, $\sigma_0^2 = \Sigma'\{(b_i - \beta a_i)^2 \sigma_i^2\}$

where Σ' denotes summation over all i such that $b_i \neq \beta a_i$ (Kagan et al., 1973, p. 191; Lukacs and Laha, 1964, p. 123).

[*4.3.8*] Let $X_1, \ldots, X_n$ be iid random variables with finite variance σ^2. If the conditional expectation of any unbiased quadratic estimator of $c\sigma^2$ $(c \neq 0)$, given the fixed sum $X_1 + \cdots + X_n$, does not involve the latter, the distribution of the parent population is normal (Laha, 1953, p. 228). See also [4.2.1(a)].

[*4.3.9*] We shall require the notion of constancy of regression in the remainder of this section. A rv Y with finite mean is defined to have *constant regression* on a rv X if $E(Y|X) = E(Y)$ with probability one with respect to the probability distribution of X (Lukacs, 1956, p. 204). Another way of stating this is that the conditional expectation of Y, given that X = x, is free of x almost surely.

[*4.3.10*] (a) Let $X_1, \ldots, X_n$ be a random sample from a certain population with finite variance. Let L and Q be linear and quadratic statistics, defined by

$$L = X_1 + \cdots + X_n, \qquad Q = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} X_i X_j + \sum_{i=1}^{n} b_i X_i$$

where $\Sigma_{i=1}^n a_{ii} \neq 0$, $\Sigma_{i=1}^n \Sigma_{j=1}^n a_{ij} = 0$, and $\Sigma_{i=1}^n b_i = 0$. Then the parent population is normal if and only if Q has constant regression on L (Kagan et al., 1973, p. 215; Mathai and Pederzoli, 1977, p. 46; Lukacs and Laha, 1964, p. 106).

(b) Lukacs (1956, p. 205) gives the special case of (a) in which $b_1 = \cdots = b_n = 0$.

[*4.3.11*] (a) Let $X_1, \ldots, X_n$ be iid rvs with moments up to order m. Define the linear statistics

$$L_1 = \sum_{i=1}^{n} a_i X_i, \qquad L_2 = \sum_{i=1}^{n} b_i X_i$$

where $\Sigma_{i=1}^n a_i b_i = 0$ and $a_i a_j > 0$ (i, j = 1, ..., n). Then L_2^2 has constant regression on L_1 if and only if the parent distribution is normal (Cacoullos, 1967a, pp. 399-400; see also Cacoullos, 1967b, p. 1895).

(b) Let $X_1, \ldots, X_n$ be as in (a), with a symmetric distribution. Define L_1 and L_2 as in (a), where $a_i \neq 0$ (i = 1, ..., n) and $\Sigma_{i=1}^n a_i b_i = 0$. Then L_2^2 has constant regression on L_1 if and only if the parent distribution is normal (Cacoullos, 1967a, p. 401).

[*4.3.12*] Let $X_1, \ldots, X_n$ be iid random variables such that $E(|X_1|^p) < \infty$ $(p \geq 2)$. Let $P = P(X_1, \ldots, X_n)$ be a homogeneous polynomial statistic of degree p which is an unbiased estimator of the pth cumulant κ_p. Then P has constant regression on the sum $X_1 + \cdots + X_n$ if and only if the underlying population is normal (Kagan et al., 1973, p. 216).

4.4 INDEPENDENCE OF OTHER STATISTICS

[*4.4.1*] A natural question to ask is whether the independence of the sample mean $\bar{X}$ and some function of $X_1 - \bar{X}, \ldots, X_n - \bar{X}$ in a random sample from a population (see [5.2.3]) characterizes the normal law. For a discussion, see Johnson and Kotz (1970, pp. 51-52) and [5.2.3]; particular cases are presented in [4.2.1], [4.4.2], and [4.4.6].

[*4.4.2*] Let $X_1, \ldots, X_n$ be iid random variables and let $M_p = \Sigma_{i=1}^n (X_i - \bar{X})^p/n$, the pth sample moment (p = 2, 3, ...). Suppose that (p - 1)! is not divisible by (n - 1); then M_p and $\bar{X}$ are independent if and only if the parent population is normal (Lukacs and Laha, 1964, p. 101).

[*4.4.3*] A polynomial $P(x_1,\ldots,x_n)$ of degree p is said to be *nonsingular* if it contains the pth power of at least one variable and if for all positive integers k, $\Pi(k) \neq 0$, where $\Pi(k)$ is the polynomial formed if we replace each positive power x_s^j by $k(k-1)\cdots(k-j+1)$ in $P(x_1,\ldots,x_n)$ (Lukacs and Laha, 1964, p. 96).

Let $X_1, \ldots, X_n$ be iid random variables, and $P = P(X_1,\ldots,X_n)$ be a nonsingular polynomial statistic. If P and $X_1 + \cdots + X_n$ are independent, then the parent distribution is normal (Lukacs and Laha, 1964, p. 98).

[*4.4.4*] Let X and Y be iid random variables. Let the quotient X/Y follow the Cauchy law distributed symmetrically about zero and be independent of $U = X^2 + Y^2$. Then the rvs X and Y follow the normal law (Seshadri, 1969, p. 258). See also [5.1.1].

[*4.4.5*] Let $X_0, X_1, \ldots, X_n$ $(n \geq 2)$ be n + 1 independent rvs such that $\Pr(X_i = 0) = 0$ (i = 0, 1, ..., n) and with distributions symmetrical about zero. Let $Y_1 = X_1/|X_0|$, $Y_2 = X_2\sqrt{2}/(X_0^2 + X_1^2)^{1/2}$, ..., $Y_n = X_n\sqrt{n}/(X_0^2 + X_1^2 + \cdots + X_{n-1}^2)^{1/2}$.

A necessary and sufficient condition for $X_0, \ldots, X_n$ to be iid $N(0,\sigma^2)$ rvs is that $Y_1, \ldots, Y_n$ are mutually independent rvs having Student's t distribution with 1, 2, ..., n degrees of freedom, respectively (Kotlarski, 1966, p. 603; Mathai and Pederzoli, 1977, pp. 69, 71).

[*4.4.6*] For a definition of k-statistics in the following characterization, see Kendall and Stuart (1977, pp. 296-346). Let $X_1, \ldots, X_n$ be a random sample from a population with cdf G(x), and let p be an integer greater than one, such that the pth moment of G exists. The distribution G is normal if and only if the k-statistic of order p is independent of the sample mean $\bar{X}$ (Lukacs, 1956, p. 201).

4.5 CHARACTERISTIC FUNCTIONS AND MOMENTS

[*4.5.1.1*] A distribution is characterized by its moments, if these exist for all orders, but also if $E[\{\min(X_1,X_2,\ldots,X_n)\}^k]$ is specified for all $k \geq 1$, where $X_1, X_2, \ldots, X_n$ is a random sample from the distribution of interest (Chan, 1967, pp. 950-951). See also Arnold and Meeden (1975, pp. 754-758).

[*4.5.1.2*] An rv X has a normal distribution if and only if all the cumulants κ_r of order r vanish, for all r such that $r > r_0 > 2$ (Kendall and Stuart, 1977, p. 125).

[*4.5.2*] Let m_1, m_2, and n be integers such that $0 \leq m_i \leq n$, $n > 1$ $(i = 1, 2)$. Suppose $\psi_1(t)$ and $\psi_2(t)$ are two characteristic functions such that

$$\psi_1(t)\psi_2(t) = \exp(ict)\{\psi_1(t/\sqrt{n})\}^{m_1}\{\psi_1(-t/\sqrt{n})\}^{n-m_1} \cdot \{\psi_2(t/\sqrt{n})\}^{m_2}\{\psi_2(-t/\sqrt{n})\}^{n-m_2}$$

for all values of t. Then $\psi_1(t)$ and $\psi_2(t)$ correspond to normal distributions (Blum, 1956, p. 59). See also Blum (1956, p. 61).

[*4.5.3*] The following result is useful in the study of particle size distributions. Let Y be a rv with pdf p(x) and moment-generating function M(t), and let X be the rv with pdf g(x) given by

$$g(x) = \exp(tx)p(x)/M(t), \qquad -\infty < x < \infty$$

Then Var(X) = Var(Y) for all values of t if and only if Y has a normal pdf (Ziegler, 1965, p. 1203). Under these conditions, if Y is a $N(\mu,\sigma^2)$ rv, then X has a $N(\mu + \sigma^2 t, \sigma^2)$ distribution.

[*4.5.4*] Let ψ be a nontrivial characteristic function such that, in some neighborhood of the origin where it does not vanish,

$$\psi(t) = \prod_{j=1}^{\infty} [\{\psi(\pm\beta_j t)\}^{v_j}], \qquad 0 < \beta_j < 1,\ v_j > 0\ (j = 1, 2, \ldots)$$

Then (a) $\Sigma_{j=1}^{\infty} v_j \beta_j^2 \le 1$ and (b) ψ is the characteristic function of a normal law, if and only if $\Sigma_{j=1}^{\infty} v_j \beta_j^2 = 1$ (Kagan et al., 1973, p. 178).

[*4.5.5*] The rv W^2 has a χ_1^2 distribution (see [5.3]) if and only if the characteristic function of W, $\psi(t)$, satisfies the equation (Geisser, 1973, p. 492)

$$\psi(t) + \psi(-t) = 2\exp(-t^2/2)$$

The distribution of W is then that of a N(0,1) rv (see [2.3.1]).

[*4.5.6*] Let Z_1, Z_2, and Z_3 be three mutually independent rvs, symmetrically distributed about zero with cdf continuous at zero. Then the joint characteristic function of the ratios Z_1/Z_3 and Z_2/Z_3 is $\exp[-\{t_1^2 + t_2^2\}^{1/2}]$ if and only if Z_1, Z_2, and Z_3 are iid normal rvs with mean zero (Kotz, 1974, p. 47; Pakshirajan and Mohan, 1969, p. 532).

4.6 CHARACTERIZATIONS FROM PROPERTIES OF TRANSFORMATIONS

[*4.6.1*] Let $x_1, x_2, \ldots, x_n$ be the Cartesian coordinates, not all zero, of a point in Euclidean space of n dimensions. We can transform to *polar coordinates* $(r, \Theta_1, \Theta_2, \ldots, \Theta_{n-1})$, if $n \ge 2$, as follows: define the transformation T by

$$x_1 = r\cos\theta_1 \qquad x_2 = r\sin\theta_1\cos\theta_2$$

$$x_3 = r\sin\theta_1\sin\theta_2\cos\theta_3, \ \cdots$$

$$x_{n-1} = r \sin \theta_1 \sin \theta_2 \cdots \sin \theta_{n-2} \cos \theta_{n-1}$$

$$x_n = r \sin \theta_1 \cdots \sin \theta_{n-2} \sin \theta_{n-1}, \quad -\infty < x_i < \infty \ (i = 1, \ldots, n),$$

$$r \geq 0, \quad 0 \leq \theta_i < 2\pi \quad (i = 1, \ldots, n-1)$$

If $X_1, \ldots, X_n$ are rvs with joint pdf $p_1(x_1,\ldots,x_n)$, transformation T leads to "*polar rvs*" R, $\Theta_1, \ldots, \Theta_{n-1}$ with joint pdf $p_2(r, \theta_1,\ldots,\theta_{n-1})$ given by

$$p_2(r,\theta_1,\ldots,\theta_{n-1}) = p_1(x_1,\ldots,x_n) r^{n-1} \cdot \left|\sin^{n-2} \theta_1 \sin^{n-3} \theta_2 \cdots \sin \theta_{n-2}\right|$$

We can speak of R as the *radial variable*, and of $\Theta_1, \ldots, \Theta_{n-1}$ as the *angular variables*; note that $R^2 = X_1^2 + \cdots + X_n^2$.

(a) Let $X_1, \ldots, X_n$ be a random sample from a certain population, which is absolutely continuous. Then the joint pdf of the sample, $p_1(x_1,\ldots,x_n)$, is constant on the "sphere" $x_1^2 + x_2^2 + \cdots + x_n^2 = r^2$ if and only if the parent population is standard normal (Bartlett, 1934, pp. 327-340; Kendall and Stuart, 1977, p. 390).

(b) Let $X_1, \ldots, X_n$ $(n \geq 2)$ be rvs and suppose that, under transformation T above, the ratio $p_2(r,\theta_1,\ldots,\theta_{n-1})/(r^{n-1} \sin^{n-2} \theta_1 \cdots \sin \theta_{n-2})$* is well defined and nonzero everywhere, continuous in r and equal to $p_1(x_1,\ldots,x_n)$, which is in turn continuous in each x_i $(i = 1, \ldots, n)$. Then $X_1, \ldots, X_n$ are mutually independent and R is independent of $\Theta_1, \Theta_2, \ldots, \Theta_{n-1}$ if and only if $X_1, \ldots, X_n$ are iid $N(0,\sigma^2)$ random variables (Tamhanker, 1967, p. 1924; Kotz, 1974, p. 50; Johnson and Kotz, 1970, p. 52). This result essentially gives conditions under which the independence of the radial and (joint) angular variables characterizes normality. Tamhanker (1967) gives the transformed pdf $p_2(r,\theta_1,\ldots,\theta_n)$ without absolute value notation.

---

*The denominator should perhaps be $r^{n-1}\left|\sin^{n-2} \theta_1 \cdots \sin \theta_{n-2}\right|$.

(c) Let $X_1, \ldots, X_n$ $(n \geq 2)$ be rvs, and suppose that, under transformation T above, $p_1(x_1,\ldots,x_n)$ is continuous everywhere, $p_1(0,\ldots,0) > 0$, and $p_2(r,\theta_1,\ldots,\theta_{n-1})/|r^{n-1} \sin^{n-2} \theta_1 \cdots \sin \theta_{n-2}|$ is well defined and equal to $p_1(x_1,\ldots,x_n)$. Then $X_1, \ldots, X_n$ are iid $N(0,\sigma^2)$ rvs if and only if (i) two of $X_1, \ldots, X_n$ are conditionally independent, given the values of the other $n - 2$ rvs; and (ii) R and θ_1 are conditionally independent, given $\theta_2, \theta_3, \ldots, \theta_{n-1}$ (Alam, 1971, p. 523).

[*4.6.2*] Let X_1 and X_2 be iid random variables with (continuous) pdf g(x). Let Y_1 and Y_2 be rvs defined by the transformation

$$Y_1 + iY_2 = (X_1 + iX_2)^k(X_1^2 + X_2^2)^{-(k-1)/2}, \qquad i = \sqrt{-1}$$

where k is an integer not less than 2. Then X_1 and X_2 are iid $N(0,\sigma^2)$ rvs if and only if Y_1 and Y_2 are independent with the same distribution as X_1 and X_2 (Beer and Lukacs, 1973, pp. 100, 103; Kotz, 1974, pp. 49-50).

[*4.6.3*] Let X_1 and X_2 be iid random variables with (continuous) pdf g(x). Let Y_1 and Y_2 be rvs defined by the transformation

$$Y_1 = X_1 \cos\{a(X_1^2 + X_2^2)\} + X_2 \sin\{a(X_1^2 + X_2^2)\}$$
$$Y_2 = -X_1 \sin\{a(X_1^2 + X_2^2)\} + X_2 \cos\{a(X_1^2 + X_2^2)\}$$

where a is a nonzero constant. Then X_1 and X_2 are iid $N(0,\sigma^2)$ rvs if and only if Y_1 and Y_2 are independent with the same distribution as X_1 and X_2 (Beer and Lukacs, 1973, pp. 102, 106-107).

[*4.6.4*] (a) The following result was designed for testing normality. Let $n = 2k + 3$, where $k \geq 2$; let $X_1, \ldots, X_n$ be iid random variables with (unknown) mean μ and (unknown) variance σ^2, $|\mu| < \infty$, $\sigma^2 > 0$. Let

$$Z_1 = (X_1 - X_2)/\sqrt{2}, \qquad Z_2 = (X_1 + X_2 - 2X_3)/\sqrt{3.2}, \quad \cdots$$
$$Z_{n-1} = \{X_1 + \cdots + X_{n-1} - (n - 1)X_n\}/\sqrt{n(n - 1)}$$

$$Z_n = X_1 + \cdots + X_n$$

Then define

$$Y_1 = Z_1^2 + Z_2^2, \qquad Y_2 = Z_3^2 + Z_4^2, \qquad \ldots$$

$$Y_{k+1} = Z_{n-2}^2 + Z_{n-1}^2, \qquad S_{k+1} = Y_1 + Y_2 + \cdots + Y_k$$

where $k + 1 = \frac{1}{2}(n - 1)$. Finally, we define

$$U_{(r)} = (Y_1 + \cdots + Y_r)/S_{k+1}, \qquad r = 1, \ldots, k$$

Then $U_{(1)}, U_{(2)}, \ldots, U_{(k)}$ behave like the k-order statistics of k iid random variables having a uniform distribution over (0,1), if and only if the parent distribution of $X_1, \ldots, X_n$ is $N(\mu,\sigma^2)$ (Csörgö and Seshadri, 1971, pp. 333-339; Kotz, 1974, p. 49).

(b) The sources in (a) also give an analogous characterization which is suitable when μ is known; here $n = 2k$, $Z_i = X_i - \mu$ $(i = 1, \ldots, n)$, the Y_j are similarly defined, with k + 1 replaced by k; $S_k = Y_1 + \cdots + Y_k$ and $U_{(r)} = (Y_1 + \cdots + Y_r)/S_k$, where now $r = 1, \ldots, k - 1$. The conclusion is directly analogous, replacing k by k - 1.

4.7 SUFFICIENCY, ESTIMATION, AND TESTING

The following material concerns the role of the normal distribution in statistical inference rather than distributional properties, but is given here for convenience. Terms such as information, sufficiency, completeness, and admissibility are defined and their properties are developed in sources such as Bickel and Doksum (1977), Ferguson (1967), and Lehmann (1959). See also Mood et al. (1974) for a discussion of concepts used in the theory of estimation and testing of hypotheses.

[*4.7.1*] It turns out that, among all continuous distributions depending on a location parameter, and having certain regularity conditions, the location parameter is estimated worst of all in the case of normal samples. This important result is due to Kagan et al. (1973, pp. 405-406). Formally, let $p(x - \theta)$ denote the pdf of a family of continuous distributions with location parameter θ; assume $p(\cdot)$ to be continuously differentiable, having fixed finite variance σ^2, and such that $|x|p(x) \to 0$ as $|x| \to \infty$. The *Fisher information measure* $I_p(\theta)$ is given here by

$$I_p(\theta) = E[\{\partial \log p(x - \theta)/\partial\theta\}^2 | \theta] = I_p(0), = I_p, \text{ say}$$

Then in the class of all such densities p, $\min_p I_p$ is attained when X has a $N(\theta,\sigma^2)$ distribution.

[*4.7.2*] If p(x) is a pdf, then a measure of its closeness to a uniform distribution (Mathai and Pederzoli, 1977, p. 15) is given by its *entropy*, defined as $J = -\int_a^b \log p(x)\cdot p(x)\, dx$, where $\Pr(a < X < b) = 1$. Let X be a rv with pdf p(x), and with given mean and variance. Among all such continuous distributions, the $N(\mu,\sigma^2)$ pdf maximizes the entropy (Mathai and Pederzoli, 1977, p. 16; Kagan et al., 1973, pp. 408-410).

[*4.7.3*] Let $p(x;\theta)$ be a pdf indexed by a parameter θ. If $\theta_1 \neq \theta_2$, but $p(\cdot;\theta_1) = p(\cdot;\theta_2)$, where θ, θ_1, θ_2 lie in the parameter space, we say that θ is *unidentifiable*; otherwise, θ is *identifiable* (Bickel and Doksum, 1977, p. 60).

Let (X_1,X_2) be (jointly) normally distributed, and let Y be a rv independent of (X_1,X_2). Consider a linear structure given by

$$U = Y + X_1, \qquad V = \beta Y + X_2$$

Then the parameter β is identifiable if and only if Y is not normally distributed (Lukacs and Laha, 1964, p. 126).

[*4.7.4*] (a) The following result modernizes a result of Gauss (1809; 1906). Let $\{G(x - \theta);\ |\theta| < \infty\}$ be the cdf of a location parameter family of absolutely continuous distributions on the real line and let the pdf g(x) be lower semicontinuous at zero. If for all random samples of sizes 2 and 3, a maximum likelihood estimator (MLE) of θ is $\bar{X}$, then $G(\cdot)$ is a normal distribution with mean zero (Teicher, 1961, p. 1215).

(b) Let $\{G(x/\sigma);\ \sigma > 0\}$ be a scale parameter family of absolutely continuous distributions with the pdf $g(x)$ satisfying

(i) $g(\cdot)$ is continuous on $(-\infty,\infty)$

(ii) $\lim_{y\to 0}[g(\lambda y)/g(y)] = 1$, whenever $\lambda > 0$

If, for all sample sizes a MLE of σ is $(\Sigma_{i=1}^n X_i^2/n)^{1/2}$, then $G(\cdot)$ has a $N(0,1)$ distribution (Teicher, 1961, p. 1221).

[*4.7.5*] (a) Let $X_1, X_2, \ldots, X_n$ $(n \ge 2)$ be mutually independent nondegenerate rvs; θ is a location parameter, so that X_i has cdf $G_i(x - \theta)$, $(i = 1, 2, \ldots, n)$, $-\infty < \theta < \infty$. A necessary and sufficient condition for $b_1X_1 + \cdots + b_nX_n$ $(b_1, \ldots, b_n$ all nonzero) to be a sufficient statistic for θ is that each X_i is a normal variable with variance a/b_i for some constant a (Kelkar and Matthes, 1970, p. 1086).

(b) Let $X_1, \ldots, X_n$ $(n \ge 2)$ be defined as in (a). Then the sample mean $\bar{X}$ is sufficient for θ if and only if $X_1, \ldots, X_n$ are iid normal rvs (Dynkin, 1961, pp. 17-40; Mathai and Pederzoli, 1977, p. 75). This is the case in (a) for which $b_1 = \cdots = b_n = 1/n$.

[*4.7.6*] (a) Let $X_1, \ldots, X_n$ $(n \ge 2)$ be mutually independent rvs with a common scale parameter σ, so that X_i has cdf $G_i(x/\sigma)$ $(i = 1, \ldots, n)$, $\sigma > 0$. Let G_i be absolutely continuous (wrt Lebesgue measure) in a neighborhood of zero, and let $\partial g_i/\partial x$ be nonzero and continuous when $x = 0$. Then if $\Sigma_{i=1}^n X_i^2$ is sufficient for σ, $X_1, \ldots, X_n$ have normal distributions with mean zero (Kelkar and Matthes, 1970, p. 1087).

(b) Let $X_1, X_2, \ldots, X_n$ $(n \ge 4)$ be iid random variables, each X_i having a location parameter θ $(-\infty < \theta < \infty)$ and a scale parameter σ $(\sigma > 0)$. Let $\bar{X} = \Sigma_{i=1}^n X_i/n$ and $S^2 = \Sigma_{i=1}^n (X_i - \bar{X})^2/(n - 1)$. If $(\bar{X},S^2)$ is a sufficient statistic for (θ,σ), then each X_i has a normal distribution (Kelkar and Matthes, 1970, p. 1087).

[*4.7.7*] If $X_1, X_2, \ldots, X_n$ are independent random variables such that their joint density function involves an unknown location parameter θ, then a necessary and sufficient condition in order that $\Sigma_{i=1}^{n} b_i X_i$ is a boundedly complete sufficient statistic for θ is that $b_i > 0$ and that X_i is a normal variable with mean θ and variance proportional to $1/b_i$ $(i = 1, \ldots, n)$ (from Basu, 1955, p. 380). In particular, $\bar{X}$ is boundedly complete sufficient for θ if and only if $X_1, \ldots, X_n$ are iid normal rvs.

[*4.7.8*] Let $X_1, \ldots, X_n$ be continuous iid random variables with pdf $g(x)$ and variance σ^2. Let $\bar{X}$ be the sample mean, and $h(X_1, \ldots, X_n)$ be any distribution-free unbiased estimator of σ^2.

(i) If $\bar{X}$ and $h(X_1,\ldots,X_n)$ are independent, $g(\cdot)$ is the normal pdf.

(ii) If g is the normal pdf, $\bar{X}$ and $h(X_1,\ldots,X_n)$ are uncorrelated (Shimizu, 1961, p. 53).

[*4.7.9*] The following characterizations hold under a *quadratic loss* function $(\hat{\theta} - \theta)^2$, where $\hat{\theta}$ estimates a parameter θ.

(a) Let θ be a location parameter and $X_1, \ldots, X_n$ $(n \geq 3)$ be mutually independent rvs such that X_j has cdf $G_j(x - \theta)$, $E(X_j) = 0$ and $\mathrm{Var}(X_j) = \sigma_j^2 < \infty$; $j = 1, \ldots, n$. The optimal linear estimator of θ under quadratic loss, given by

$$\hat{L} = \left(\sum_{i=1}^{n} X_i/\sigma_i^2\right)\left(\sum_{i=1}^{n} 1/\sigma_i^2\right)^{-1}$$

is admissible in the class of all unbiased estimators of θ if and only if all the cdfs $G_j(\cdot)$ are normal $(j = 1, \ldots, n)$ (Kagan et al., 1973, p. 227).

(b) In particular in (a), when $X_1, \ldots, X_n$ are iid random variables $(n \geq 3)$, $\hat{L} = \bar{X}$, the sample mean. Hence the admissibility under quadratic loss of $\bar{X}$ in the class of unbiased estimators of θ characterizes the parent distribution as normal (Kagan et al., 1973, p. 228).

(c) Let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ be mutually independent rvs $(n \geq 3)$ with zero means and finite variances, where X_j has cdf $G_j(x - \theta - \Delta)$, Y_j has cdf $H_j(x - \theta)$, $(j = 1, \ldots, n)$ and $|\theta| < \infty$, $|\Delta| < \infty$. In order that an estimator of Δ of the form $\Sigma a_j(X_j - Y_j)$

be admissible under quadratic loss in the class of unbiased estimators of Δ, it is necessary and sufficient that $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ all be normally distributed and that $\mathrm{Var}(X_j)/\mathrm{Var}(Y_j)$ be constant; $j = 1, \ldots, n$ (Kagan et. al., 1973, p. 228).

(d) Kagan et al. (1973, chap. 7) give further characterizations of normality through admissibility of certain estimators under quadratic loss; these include polynomials, least squares estimators, and linear estimators of shift parameters in autoregressive processes.

[*4.7.10*] *Absolute error loss* for an estimator $\hat{\theta}$ of θ is $|\hat{\theta} - \theta|$. Let $X_1, \ldots, X_n$ $(n \geq 6)$ be a random sample from a unimodal distribution with cdf $F(x - \theta)$ and a continuously differentiable pdf. Then the sample mean $\bar{X}$ is admissible as an estimator of θ under absolute error loss if and only if the parent distribution is normal (Kagan et al., 1973, p. 247; see also p. 251).

[*4.7.11*] Let $X_1, \ldots, X_n$ be a random sample $(n \geq 3)$ from a population with continuous cdf $F(x - \theta)$ and finite variance. We wish to test the null hypothesis H_0: $\theta = 0$ against the composite alternative H_1: $\theta > 0$. For every α $(0 < \alpha < 1)$, the test having critical region determined by the relation

$$\bar{X} > c_\alpha, \qquad c_\alpha = \max\{c: \Pr(\bar{X} > c \mid \theta = 0) = \alpha\}$$

is uniformly most powerful of size α for testing H_0 against H_1 if and only if the parent population is normal (Kagan et al., 1973, p. 451).

4.8 MISCELLANEOUS CHARACTERIZATIONS

We begin this section with some characterizations derived through order statistics.

[*4.8.1*] Let $X_1, \ldots, X_n$ be a random sample from a continuous population, and let $Y_i = \Sigma_{j=1}^n a_{ij} X_j$ $(i = 1, \ldots, n)$ be an orthogonal transformation. Further, let $W_1 = \max(X_1,\ldots,X_n) - \min(X_1,\ldots,X_n)$ and $W_2 = \max(Y_1,\ldots,Y_n) - \min(Y_1,\ldots,Y_n)$ be the corresponding sample ranges. Then W_1 and W_2 are identically distributed if and only if the parent population of $X_1, \ldots, X_n$ is normal (Kotz, 1974, p. 50).

[*4.8.2*] In what follows below, $X_1, \ldots, X_n$ are iid nontrivial rvs, having cdf G and finite variance. The order statistics are given by

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}, \quad n \ge 2, \text{ and } \Phi \text{ is the N(0,1) cdf}$$

The characterizations are due to Govindarajulu (1966, pp. 1011-1015).

(a) $E[X_{(n)}^2 - X_{(n-1)}X_{(n)}] = 1$ if and only if there is an extended real number A ($-\infty \le A < \infty$) such that

$$G(x) = \begin{cases} 0, & x \le A \\ \{\Phi(x) - \Phi(A)\}/\{1 - \Phi(A)\}, & A < x < \infty \end{cases}$$

a normal distribution truncated from below.

(b) $E[X_{(1)}^2 - X_{(1)}X_{(2)}] = 1$ if and only if there is an extended real number B ($-\infty < B \le \infty$) such that

$$G(x) = \begin{cases} \Phi(x)/\Phi(B), & -\infty < x < B \\ 1, & x \ge B \end{cases}$$

a normal distribution truncated from above.

(c) Assume that $G(0) = 1$. Then $\Sigma_{i=1}^n E[X_{(i)}X_{(n)}] = 1$ if and only if

$$G(x) = \begin{cases} 2\Phi(x), & -\infty < x \le 0 \\ 1, & x > 0 \end{cases}$$

a normal distribution, folded on the left.

(d) Suppose that the mean of the parent distribution is at zero. Then for $i = 1, 2, \ldots, n$, $\Sigma_{j=1}^n E[X_{(i)}X_{(j)}] = 1$ if and only if $G(x) = \Phi(x)$, $|x| < \infty$.

(e) Let the parent distribution be symmetric about zero; then for $i = 1, \ldots, n$, $\Sigma_{j=1}^n E[X_{(i)}X_{(j)}] = 1$ if and only if $G(x) = \Phi(x)$, $|x| < \infty$.

[*4.8.3*] (a) Let X_1, X_2, X_3 be mutually independent rvs, symmetrical about the origin, and such that $\Pr(X_k = 0) = 0$, $k = 1, 2, 3$. Let $Y_1 = X_1/X_3$, $Y_2 = X_2/X_3$. The necessary and sufficient condition for X_1, X_2, and X_3 to be normally distributed with a common standard deviation σ is that the joint distribution of (Y_1, Y_2) is the bivariate Cauchy distribution with joint pdf $g(y_1,y_2)$, where (Kotlarski, 1967, p. 75)

$$g(y_1,y_2) = [2\pi(1 + y_1^2 + y_2^2)^{3/2}]^{-1}, \qquad -\infty < y_1,y_2 < \infty$$

(b) Define X_1, X_2, and X_3 as in (a); and let $V_1 = X_1(X_1^2 + X_2^2)^{-1/2}$, $V_2 = (X_1^2 + X_2^2)^{1/2}(X_1^2 + X_2^2 + X_3^2)^{-1/2}$. Then X_1, X_2, and X_3 are iid normally distributed rvs if and only if V_1 and V_2 are independent with respective pdfs $h_1(u)$ and $h_2(u)$ given by (Kotlarski, 1967, pp. 75-76)

$$h_1(u) = \pi^{-1}(1 - u^2)^{-1/2}, \qquad |u| < 1$$

$$h_2(u) = u(1 - u^2)^{-1/2}, \qquad 0 < u < 1$$

4.9 NEAR-CHARACTERIZATIONS

Suppose that a random sample $X_1, \ldots, X_n$ has some property P which characterizes a distribution G (or family of distributions). The question then arises: if the sample *nearly* has the property P, is the parent distribution nearly G, and vice versa? If the answer is yes, then (with formal definitions) the characterization is termed *stable*. Work in this field is limited, but we can give a few results.

[*4.9.1*] Two rvs X and Y are ε-*independent* if, for all x and y,

$$|\Pr(X < x, Y < y) - \Pr(X < x)\Pr(Y < y)| < \varepsilon$$

The rv X is ε-*normal* with parameters μ and σ if its cdf G satisfies the inequality (Nhu, 1968, p. 299)

$$|G(x) - \Phi\{(x - \mu)/\sigma\}| < \varepsilon, \qquad |x| < \infty$$

See also Meshalkin (1968, p. 1747).

[*4.9.2*] (a) Let X_1 and X_2 be iid random variables with mean zero, variance one, and $E|X_i|^3 < M < \infty$ $(i = 1, 2)$. If $X_1 + X_2$ and $X_1 - X_2$ are ε-independent, then the cdf G of X_1 and X_2 satisfies the inequality

$$\sup_x |G(x) - \Phi(x)| < C_1 \varepsilon^{1/3}$$

where C_1 depends on M only (Meshalkin, 1968, p. 1748; Kagan et al., 1973, p. 298; see also Nhu, 1968, p. 300). This result relates to that of Bernstein (1941), given in [4.1.4(a)].

(b) Let X_1 and X_2 be defined as in (a), with cdf G, and let $(X_1 + X_2)/\sqrt{2}$ have cdf H. If $\sup_x |G(x) - H(x)| \le \varepsilon$, then

$$\sup_x |G(x) - \Phi(x)| < C_2 \varepsilon^{1/3}$$

where C_2 depends on M only (Meshalkin, 1968, p. 1748). This result relates to that of Cramér (1936), given in [4.1.1(a)].

[4.9.3] Let $X_1, \ldots, X_n$ be iid random variables with mean μ, variance σ^2, and $E[|X_i|^{2(1+\delta)}] = \beta_\delta < \infty$; $0 < \delta \le 1$. If $\bar{X}$ and S^2 are ε-independent, then the parent population is $A(\varepsilon)$-normal with parameters μ and σ ([4.9.1]), where $A(\varepsilon) = C_3/\sqrt{\log(1/\varepsilon)}$ and C_3 depends only on μ, σ, δ, and β_δ (Nhu, 1968, p. 300). This result relates to the fundamental characterization in [4.2.1(a)].

REFERENCES

The numbers in square brackets give the sections in which the corresponding references are cited.

Alam, K. (1971). A characterization of normality. *Annals of the Institute of Statistical Mathematics* 23, 523-525. [4.6.1]

Arnold, B. C., and Meeden, G. (1975). Characterization of distributions by sets of moments of order statistics, *Annals of Statistics* 3, 754-758. [4.5.1.1]

Bartlett, M. S. (1934). The vector representation of a sample. *Proceedings of the Cambridge Philosophical Society* 30, 327-340. [4.6.1]

Basu, D. (1955). On statistics independent of a complete sufficient statistic, *Sankhyā* 15, 377-380. [4.7.7]

Beer, S., and Lukacs, E. (1973). Characterizations of the normal distribution by suitable transformations, *Journal of Applied Probability* 10, 100-108. [4.6.2, 3]

Bernstein, S. (1941). Sur une propriété caractéristique de la loi de Gauss, *Transactions of the Leningrad Polytechnic Institute* 3, 21-22; reprinted (1964) in *Collected Works*, Vol. 4, 314-315. [4.1.4; 4.9.2]

Bickel, P. J., and Doksum, K. A. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics*, San Francisco: Holden-Day. [4.7; 4.7.3]

Blum, J. R. (1956). On a characterization of the normal distribution, *Skandinavisk Aktuarietidskrift* 39, 59-62. [4.5.2]

Bondesson, L. (1977). The sample variance, properly normalized, is χ^2-distributed for the normal law only, *Sankhyā* A39, 303-304. [4.2.5]

Chan, L. K. (1967). On a characterization of distribution by expected values of extreme order statistics, *American Mathematical Monthly* 74, 950-951. [4.5.1.1]

Cacoullos, T. (1967a). Some characterizations of normality, *Sankhyā* A29, 399-404. [4.3.11]

Cacoullos, T. (1967b). Characterizations of normality by constant regression of linear statistics on another linear statistic, *Annals of Mathematical Statistics* 38, 1894-1898. [4.3.4, 11]

Cramér, H. (1936). Über eine Eigenschaft der normalen Verteilungsfunktion, *Mathematische Zeitschrift* 41, 405-414. [4.1.1; 4.9.2]

Cramér, H. (1946). *Mathematical Methods of Statistics*, Princeton, N.J.: Princeton University Press. [4.1.1]

Csörgö, M., and Seshadri, V. (1971). Characterizing the Gaussian and exponential laws via mappings onto the unit interval, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 18, 333-339. [4.6.4]

Darmois, G. (1951). Sur une propriété caractéristique de la loi de probabilité de Laplace, *Comptes Rendus de l'Académie des Sciences* (Paris) 232, 1999-2000. [4.1.4; 4.3.5]

Diaconis, P., Olkin, I., and Ghurye, S. G. (1977). Review of *Characterization Problems in Mathematical Statistics* (by A. M. Kagan, Yu. V. Linnik, and C. R. Rao), *Annals of Statistics* 5, 583-592. [Introduction]

Dynkin, E. B. (1961). Necessary and sufficient statistics for a family of probability distributions, *Selected Translations in Mathematics, Statistics, and Probability* 1, 17-40. [4.7.5]

Ferguson, T. S. (1967). *Mathematical Statistics--A Decision Theoretic Approach*, New York: Academic. [4.7]

Gauss, C. F. (1809). *Theoria Motus Corporum Coelestium*; reprinted (1906) in *Werke*, Liber 2, Sectio III, pp. 240-244. [4.7.4]

Geary, R. C. (1936). Distribution of "Student's" ratio for non-normal samples, *Journal of the Royal Statistical Society* B3, 178-184. [4.2.1]

Geisser, S. (1956). A note on the normal distribution, *Annals of Mathematical Statistics* 27, 858-859. [4.2.1]

Geisser, S. (1973). Normal characterizations via the square of random variables, *Sankhyā* A35, 492-494. [4.2.4; 4.5.5]

Govindarajulu, Z. (1966). Characterization of normal and generalized truncated normal distributions using order statistics, *Annals of Mathematical Statistics* 37, 1011-1015. [4.8.2]

Johnson, N. L., and Kotz, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions*, Vol. 1, New York: Wiley. [4.4.1; 4.6.1]

Johnson, N. L., and Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions*, New York: Wiley. [Introduction]

Kagan, A. M., Linnik, Yu. V., and Rao, C. R. (1965). On a characterization of the normal law based on a property of the sample average. *Sankhyā* A27, 405-406. [4.3.4]

Kagan, A. M., Linnik, Yu. V., and Rao, C. R. (1973). *Characterization Problems in Mathematical Statistics* (trans. B. Ramachandran), New York: Wiley. [4.1.4, 6; 4.2.1, 2, 3; 4.3.2, 3, 4, 6, 7, 10, 12; 4.5.4; 4.7.1, 2, 9, 10, 11; 4.9.2]

Kaplansky, I. (1943). A characterization of the normal distribution, *Annals of Mathematical Statistics* 14, 197-198. [4.2.1]

Kelkar, D., and Matthes, T. K. (1970). A sufficient statistics characterization of the normal distribution, *Annals of Mathematical Statistics* 41, 1086-1090. [4.2.1; 4.7.5, 6]

Kendall, M. G., and Stuart, A. (1977). *The Advanced Theory of Statistics*, Vol. 1 (4th ed.), New York: Macmillan. [4.1.5; 4.2.1; 4.5.1; 4.6.1]

Khatri, C. G. (1979). Characterization of multivariate normality, II: Through linear regressions, *Journal of Multivariate Analysis* 9, 589-598. [Introduction]

Khatri, C. G., and Rao, C. R. (1972). Functional equations and characterization of probability laws through linear functions of random variables, *Journal of Multivariate Analysis* 2, 162-173. [Introduction]

Khatri, C. G., and Rao, C. R. (1976). Characterizations of multivariate normality, I: Through independence of some statistics, *Journal of Multivariate Analysis* 6, 81-94. [Introduction]

Kotlarski, I. (1966). On characterizing the normal distribution by Student's law, *Biometrika* 53, 603-606. [4.4.5]

Kotlarski, I. (1967). On characterizing the gamma and the normal distribution, *Pacific Journal of Mathematics* 20, 69-76. [4.8.3]

Kotz, S. (1974). Characterization of statistical distributions: A supplement to recent surveys, *International Statistical Review* 42, 39-65. [4.2.3; 4.5.6; 4.6.1, 2, 4; 4.8.1]

Laha, R. G. (1953). On an extension of Geary's theorem, *Biometrika* 40, 228-229. [4.2.1; 4.3.8]

Lancaster, H. O. (1954). Traces and cumulants of quadratic forms in normal variables, *Journal of the Royal Statistical Society* B16, 247-254. [4.1.5]

Lehmann, E. L. (1959). *Testing Statistical Hypotheses*, New York: Wiley. [4.7]

Lukacs, E. (1942). A characterization of the normal distribution, *Annals of Mathematical Statistics* 13, 91-93. [4.2.1]

Lukacs, E. (1956). Characterization of populations by properties of suitable statistics, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* 2, 195-214. [4.1.1, 2, 3, 4; 4.2.1; 4.3.9, 10; 4.4.6]

Lukacs, E., and Laha, R. G. (1964). *Applications of Characteristic Functions*, New York: Hafner. [4.2.2; 4.3.7, 10; 4.4.2, 3; 4.7.3]

Marcinkiewicz, J. (1939). Sur une proprieté de la loi de Gauss, *Mathematische Zeitschrift* 44, 612-618. [4.1.2]

Mathai, A. M., and Pederzoli, G. (1977). *Characterizations of the Normal Probability Law*, New York: Halsted/Wiley. [4.1.1, 4, 6; 4.2.2; 4.3.10; 4.4.5; 4.7.2, 5]

Meshalkin, L. D. (1968). On the robustness of some characterizations of the normal distribution, *Annals of Mathematical Statistics* 39, 1747-1750. [4.9.1, 2]

Mood, A. M., Graybill, F. A., and Boes, D. C. (1974). *Introduction to the Theory of Statistics*, New York: McGraw-Hill. [4.7]

Nhu, H. H. (1968). On the stability of certain characterizations of a normal population, *Theory of Probability and Its Applications* 13, 299-304. [4.9.1, 2, 3]

Pakshirajan, R. P., and Mohan, N. R. (1969). A characterization of the normal law, *Annals of the Institute of Statistical Mathematics* 21, 529-532. [4.5.6]

Pathak, P. K., and Pillai, R. N. (1968). On a characterization of the normal law, *Sankhyā* A30, 141-144. [4.3.5]

Patil, G. P., and Seshadri, V. (1964). Characterization theorems for some univariate probability distributions, *Journal of the Royal Statistical Society* B26, 286-292. [4.3.1]

Rao, C. R. (1967). On some characterizations of the normal law, *Sankhyā* A29, 1-14. [4.3.5]

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed., New York: Wiley. [4.2.5]

Rao, J. N. K. (1958). A characterization of the normal distribution, *Annals of Mathematical Statistics* 29, 914-919; addendum (1959), op. cit. 30, 610. [4.2.1]

Ruben, H. (1974). A new characterization of the normal distribution through the sample variance, *Sankhyā* A36, 379-388. [4.2.5]

Ruben, H. (1975). A further characterization of normality through the sample variance, *Sankhyā* A37, 72-81. [4.2.5]

Ruben, H. (1976). A characterization of normality through the general linear model, *Sankhyā* A38, 186-189. [4.2.5]

Seshadri, V. (1969). A characterization of the normal and Weibull distributions, *Canadian Mathematical Bulletin* 12, 257-260. [4.4.4]

Shimizu, R. (1961). A characterization of the normal distribution, *Annals of the Institute of Statistical Mathematics* 13, 53-56. [4.7.8]

Shimizu, R. (1962). Characterization of the normal distribution, II, *Annals of the Institute of Statistical Mathematics* 14, 173-178. [4.1.3]

Skitovitch, V. P. (1953). On a property of the normal distribution, *Doklady Akademii Nauk SSSR* 89, 217-219 (in Russian). [4.1.4; 4.3.5]

Tamhanker, M. V. (1967). A characterization of normality, *Annals of Mathematical Statistics* 38, 1924-1927. [4.6.1]

Teicher, H. (1961). Maximum likelihood characterization of distributions, *Annals of Mathematical Statistics* 32, 1214-1222. [4.7.4]

Ziegler, R. K. (1965). A uniqueness theorem concerning moment distributions, *Journal of the American Statistical Association* 60, 1203-1206. [4.5.3]

chapter 5

SAMPLING DISTRIBUTIONS

We have already listed in Section 2.3 distributions arising from some transformations of normal random variables. In this chapter we summarize distributions and their properties when they arise out of random samples from normal populations, that is, from mutually independent normal random variables $X_1, \ldots, X_n$, which will usually but not in every case be identically distributed. One class of sampling distributions which will not be considered in this chapter is that involving the order statistics; Chapter 7 will cover this class separately. Also covered separately in Chapter 6 are approximations based on the normal law to the distributions introduced in Sections 5.3 to 5.5. The abbreviation *iid* in what follows means "independently and identically distributed."

5.1 SAMPLES NOT LARGER THAN FOUR

[5.1.1] (a) Let X_1 and X_2 be iid $N(0,\sigma^2)$ rvs. Then the rv Y given by $Y = X_1/X_2$ has a *Cauchy distribution* with pdf (Kendall and Stuart, 1977, p. 285)

$$g(y) = [\pi(1 + y^2)]^{-1}, \qquad -\infty < y < \infty$$

This is also a Student's t distribution with one degree of freedom; see [5.4.1]. The converse is false (Steck, 1958, p. 604; Fox, 1965, p. 631).

(b) If Z_1 and Z_2 are independent N(0,1) rvs, and if

$$Y = (a + Z_1)/(b + Z_2)$$

then the pdf of Y is g(y;a,b), where

$$g(y;a,b) = \frac{\exp\{-(a^2 + b^2)/2\}}{\pi(1 + t^2)}\left[1 + \frac{q}{\phi(q)}\left\{\Phi(q) - \frac{1}{2}\right\}\right],$$

$$q = (b + ay)/(1 + y^2)^{1/2}$$

Marsaglia (1965, pp. 194-199) gives representations of the cdf of this distribution in terms of bivariate normal probabilities. If b > 3, then

$$\Pr(Y \le t) \simeq \Phi\{(bt - a)/(1 + t^2)^{1/2}\}$$

[*5.1.2*] Let X_1 and X_2 be independent normal rvs with common variance. Then the rvs $aX_1 + bX_2$ and $cX_1 + dX_2$ are independent if and only if $ac + bd = 0$. In particular, $X_1 + X_2$ and $X_1 - X_2$ are independent (Mood et al., 1974, p. 244).

[*5.1.3*] Let X_1 and X_2 be iid N(0,1) rvs. Then the random variable Y, where $Y = X_1X_2$, has pdf g(y), given by

$$g(y) = k_0(y)/\pi, \qquad -\infty < y < \infty$$

Here $k_0(y)$ is Bessel's function of the second kind with a purely imaginary argument of zero order (Epstein, 1948, pp. 375-377; Kendall and Stuart, 1977, pp. 285-286; Craig, 1936, pp. 1-15; Aroian et al., 1978, pp. 165-172).

[*5.1.4*] Let X_1 and X_2 be independent $N(0,\sigma_1^2)$ and $N(0,\sigma_2^2)$ rvs, respectively, and let the rv Y be defined by

$$Y = X_1X_2/(X_1^2 + X_2^2)^{1/2}$$

Then Y has a normal distribution with variance $[(\sigma_1^2)^{-1} + (\sigma_2^2)^{-1}]^{-1}$. If, in addition, $\sigma_1^2 = \sigma_2^2$, and $W = (X_1^2 - X_2^2)/(X_1^2 + X_2^2)$, then W has a normal distribution; further, Y and W are independent (Feller, 1966, p. 63; Shepp, 1964, p. 459).

[*5.1.5*] *Laplace Distribution*. Let Z_1, Z_2, Z_3, Z_4 be iid standard normal random variables. Then the rv $Y = Z_1Z_2 + Z_3Z_4$ has a Laplace distribution with pdf (Mantel, 1973, p. 31)

$$g(y) = \frac{1}{2}\exp(-|y|), \qquad -\infty < y < \infty$$

[*5.1.6*] The Birnbaum-Saunders distribution arises as a fatigue-life model. Let the rv X have a $N(0,\alpha^2/4)$ distribution, and let

$$T = \beta\{1 + 2X^2 + 2X(1 + X^2)^{1/2}\}$$

Then T has pdf $g(t;\alpha,\beta)$ and cdf $G(t;\alpha,\beta)$, where

$$g(t;\alpha,\beta) = \frac{1}{2\sqrt{2\pi}\,\alpha^2\beta t^2}\,\frac{t^2 - \beta^2}{\sqrt{t/\beta} - \sqrt{\beta/t}}\exp\left[-\frac{1}{2\alpha^2}\left(\frac{t}{\beta} + \frac{\beta}{t} - 2\right)\right],$$

$$t > 0;\ \alpha,\beta > 0$$

$$G(t;\alpha,\beta) = \Phi\{\alpha^{-1}(\sqrt{t/\beta} - \sqrt{\beta/t})\}, \qquad \alpha,\beta > 0$$

The mean and variance of T are $\beta(1 + \frac{1}{2}\alpha^2)$ and $\alpha^2\beta^2(1 + 5\alpha^2/4)$, respectively (Birnbaum and Saunders, 1969, pp. 323, 324; Mann et al., 1974, pp. 150-155). Further, the rv 1/T has cdf $G(\cdot;\alpha,\beta^{-1})$ in our notation, and the rv aT has cdf $G(\cdot;\alpha,a\beta)$ if $a > 0$.

5.2 THE SAMPLE MEAN: INDEPENDENCE

[*5.2.1*] Let X_i (i = 1, 2, ..., n) be n independent $N(\mu_i,\sigma_i^2)$ rvs. Then the random variable $Y = \Sigma_{i=1}^n a_iX_i$, where a_i, ..., a_n are real numbers, has a $N(\Sigma_{i=1}^n a_i\mu_i, \Sigma_{i=1}^n a_i^2\sigma_i^2)$ distribution. In particular, let $\bar{X} = \Sigma X_i/n$, where $a_1 = \cdots = a_n = 1$ and $\sigma_1^2 = \cdots = \sigma_n^2 = \sigma^2$. Then $\bar{X}$ has a $N(\mu,\sigma^2/n)$ distribution.

[*5.2.2*] (a) Let X_i (i = 1, 2, ..., n) be iid $N(\mu,\sigma^2)$ rvs. Then the rvs $a_1X_1 + a_2X_2 + \cdots + a_nX_n$ and $b_1X_1 + b_2X_2 + \cdots + b_nX_n$ are independent if and only if $ab_1 + a_2b_2 + \cdots + a_nb_n = 0$.

(b) Let $\underline{a}_1$, $\underline{a}_2$, ..., $\underline{a}_r$ be any n × 1 mutually orthogonal vectors ($2 \le r \le n$). Then the rvs $\underline{a}_1'\underline{X}$, $\underline{a}_2'\underline{X}$, ..., $\underline{a}_r'\underline{X}$ are mutually independent, where $\underline{X}' = (X_1,X_2,\ldots,X_n)$. If, in addition, $\underline{a}_1$, ...,

$\underline{a}_r$ are orthonormal, then $(\underline{a}_1'\underline{X}, \ldots, \underline{a}_r'\underline{X})$ are iid $N(\mu,\sigma^2)$ rvs also (Rao, 1973, pp. 183-184).

[*5.2.3*] Let X_i ($i = 1, 2, \ldots, n$) be iid $N(\mu,\sigma^2)$ rvs. Let $\bar{X}$ be the sample mean $\Sigma X_i/n$, $S^2 = \Sigma_{i=1}^n (X_i - \bar{X})^2/(n - 1)$ be the sample variance, $Y = \Sigma_{i=1}^n |X_i - \bar{X}|/n$ be the sample absolute deviation, and $W = \max(X_i) - \min(X_i)$ be the sample range. Then

(i) $\bar{X}$ and S^2 are independent.

(ii) $\bar{X}$ and Y are independent.

(iii) $\bar{X}$ and W are independent.

(iv) $\bar{X}$ and any function of the set of statistics $(X_1 - \bar{X}, \ldots, X_n - \bar{X})$ are independent (Johnson and Kotz, 1970a, p. 50).

In fact, any function $g(x_1, x_2, \ldots, x_n)$ of a sample of n independent observations from a normally distributed population and the sample mean $\bar{X}$ are independent provided only that $g(x_1, x_2, \ldots, x_n) = g(x_1 + a, x_2 + a, \ldots, x_n + a)$ (Daly, 1946, p. 71; Herrey, 1965, pp. 261-262). Another result for obtaining certain uncorrelated sample statistics from any symmetric distribution is given by Hogg (1960, pp. 265-266).

[*5.2.4*] Let $X_1, X_2, \ldots$ be an iid sequence of $N(\mu,\sigma^2)$ random variables and define $Y_n = [nX_{n+1} - (X_1 + \cdots + X_n)]/\sqrt{n(n + 1)}$; $n = 1, 2, \ldots$. Then $Y_1, Y_2, \ldots$ is an iid sequence of $N(0,\sigma^2)$ random variables (Lehmann, 1959, p. 201).

5.3 SAMPLING DISTRIBUTIONS RELATED TO CHI-SQUARE

Of particular interest in what follows is the statistic S^2, or $\Sigma(X_i - \bar{X})^2/(n - 1)$, the sample variance in a random sample of size n from a $N(\mu,\sigma^2)$ distribution.

[*5.3.1*] Let $Z_1, Z_2, \ldots, Z_n$ be iid $N(0,1)$ random variables. Then the rv Y, where $Y = \Sigma_{i=1}^n Z_i^2$, has a *chi-square distribution* with n degrees of freedom, denoted by χ_n^2. The quantiles of Y are $\chi_{n;\beta}^2$, where $\Pr(Y \le \chi_{n;\beta}^2) = 1 - \beta$, and Y has pdf $g(y;n)$, given by

$g(y;n) = [2^{n/2}\Gamma(n/2)]^{-1}y^{(n-2)/2}e^{-y/2}$; $y > 0$, $n = 1, 2, \ldots$. The mean, variance, skewness, and kurtosis of a χ^2_ν rv are, respectively (Johnson and Kotz, 1970a, p. 168)

$$\nu, \quad 2\nu, \quad \sqrt{8/\nu}, \quad \text{and } 3 + 12/\nu$$

See Johnson and Kotz (1970a, chap. 17) and Lancaster (1969) for a fuller discussion.

[5.3.2] *Some Basic Properties of Chi-Square from Normal Samples.* Let $X_1, \ldots, X_n$ be iid $N(\mu,\sigma^2)$ random variables.

(a) The rv $\Sigma_{i=1}^n (X_i - \mu)^2/\sigma^2$ has a χ^2_n distribution.

(b) The rv $\Sigma_{i=1}^n (X_i - \bar{X})^2/\sigma^2 = (n - 1)\, S^2/\sigma^2$ has a χ^2_{n-1} distribution (Mood et al., 1974, pp. 243-245).

(c) See [5.3.7] for some properties of quadratic forms, applicable here when the noncentrality parameter is zero, i.e., when $\underline{\mu} = \underline{0}$.

[5.3.3] Let $G(y;\nu)$ be the cdf of a χ^2_ν rv. Then

$$\text{(i)}\quad 1 - G(y;\, 2m + 1) = 2[1 - \Phi(\sqrt{y})] + 2\sum_{r=1}^{m} g(y;\, 2r + 1)$$

$$= 2[1 - \Phi(\sqrt{y})] + 2\phi(\sqrt{y}) \times \sum_{r=1}^{m} \frac{y^{(2r-1)/2}}{1\cdot 3\cdot 5\cdots(2r - 1)}$$

$$\text{(ii)}\quad 1 - G(y;\, 2m) = 2\sum_{r=1}^{m} g(y;\, 2r)$$

$$= \sqrt{2\pi}\,\phi(\sqrt{y})\left[1 + \sum_{r=1}^{m-1} \frac{y^r}{2\cdot 4\cdots(2r)}\right]$$

$$\text{(iii)}\quad \Phi(\sqrt{y}) = \sum_{r=1}^{\infty} g(y;\, 2r + 1) + \frac{1}{2} = \sum_{r=2}^{\infty} g(y;\, r)$$

(Puri, 1973, p. 63; Lancaster, 1969, p. 21; Abramowitz and Stegun, 1964, p. 941)

[5.3.4] The *sample variance* S^2 has a $\{\sigma^2/(n - 1)\}\chi^2_{n-1}$ distribution, so that $\Pr(S^2 \le y) = \Pr(\chi^2_{n-1} \le (n - 1)y/\sigma^2)$. Note that the second sample moment m_2 is sometimes termed the *sample variance*, where $m_2 = \Sigma(X_i - \bar{X})^2/n = (n - 1)S^2/n$ (e.g., Kendall and Stuart, 1977, p. 273). If $\nu = n - 1$, the pdf of S^2 is $g(y;\nu)$, where (Mood et al., 1974, p. 245)

$$g(y;\nu) = \left(\tfrac{1}{2}\nu/\sigma^2\right)^{\nu/2} y^{\nu/2-1} \exp\{-\nu y/(2\sigma^2)\}/\Gamma(\nu/2), \qquad y > 0$$

The first four moments and the shape factors of the sampling distribution of S^2 are:

$$E(S^2) = \sigma^2$$

$$\mathrm{Var}(S^2) = 2\sigma^4/(n - 1)$$

$$\mu_3(S^2) = 8\sigma^6/(n - 1)^2$$

$$\mu_4(S^2) = 12(n + 3)\sigma^8/(n - 1)^3$$

Skewness: $\mu_3(S^2)/\{\mathrm{Var}(S^2)\}^{3/2} = 2\sqrt{2/(n - 1)}$

Kurtosis: $\mu_4(S^2)/\{\mathrm{Var}(S^2)\}^2 = 3 + 12/(n - 1)$

Generally,

$$\mu_r'(S^2) = \{\sigma^2/(n - 1)\}^r E\{(\chi^2_{n-1})^r\}$$

and

$$\mu_r(S^2) = \{\sigma^2/(n - 1)\}^r E\{(\chi^2_{n-1} - n + 1)^r\}$$

where χ^2_{n-1} denotes the chi-square distribution with n - 1 degrees of freedom.

[*5.3.5*] The sampling distribution of the *sample standard deviation* S, where $S = \{\Sigma_{i=1}^n (X_i - \bar{X})^2/(n - 1)\}^{1/2}$. The pdf of S is $g(s;n)$, where

$$g(s;n) = \frac{s^{\nu-1}\nu^{\nu/2} \exp\{-\nu s^2/(2\sigma^2)\}}{2^{(\nu-2)/2}\sigma^\nu \Gamma(\nu/2)}, \qquad s > 0,\ \nu = n - 1 \ge 1$$

This is the pdf of $(\sigma/\sqrt{n - 1}) \times$ (chi with n - 1 degrees of freedom). Probability statements about S are most easily made from the distribution of S^2. Thus $\Pr(a \le S \le b) = \Pr(a^2 \le S^2 \le b^2)$; $b > a > 0$. From Johnson and Welch (1939, pp. 216-218) we find the moments of S about the origin to be given by

$$E(S) = \sigma\sqrt{2/(n-1)}\,\Gamma(n/2)/\Gamma\{(n-1)/2\}$$

$$E(S^2) = \sigma^2$$

$$E(S^3) = n\sigma^2 E(S)/(n-1)$$

$$E(S^4) = \sigma^4(n+1)/(n-1)$$

$$E(S^r) = \{2\sigma^2/(n-1)\}^{r/2}\Gamma\{(n+r-1)/2\}/\Gamma\{(n-1)/2\}, \quad r = 1, 2, \ldots$$

Further, Jensen's inequality gives

$$E(S) = E(\{S^2\}^{1/2}) < \{E(S^2)\}^{1/2} = \sigma$$

There are no concise expressions for the central moments, but David (1949, p. 390) gives expansions accurate to $O(n^{-2})$: if $\nu = n - 1$,

$$E(S) \simeq \sigma(1 - \nu^{-1}/4 + \nu^{-2}/32)$$

$$\mathrm{Var}(S) \simeq \sigma^2\left(\frac{1}{2}\nu^{-1} - \nu^{-2}/8\right)$$

$$\mu_3(S) \simeq \sigma^3\nu^{-2}/4$$

$$\mu_4(S) \simeq 3\sigma^4\nu^{-2}/4$$

Pearson and Hartley (1966, table 35) tabulate values of $E(S/\sigma)$, $\sqrt{\mathrm{Var}(S/\sigma)}$, $\mu_3(S)/\{\mathrm{Var}(S)\}^{3/2}$, and $\mu_4(S)/\{\mathrm{Var}(S)\}^2$ to 6, 5, 4, and 4 decimal places, respectively, for $\nu = 1(1)20(5)50(10)100$.

[*5.3.6*] Let $X_1, X_2, \ldots, X_n$ be mutually independent rvs such that X_i has a $N(\mu_i, 1)$ distribution, and let $\lambda = \Sigma_{i=1}^n \mu_i^2$. Then the rv Y, where $Y = \Sigma_{i=1}^n X_i^2$ has a *noncentral chi-square distribution* with n degrees of freedom and noncentrality parameter λ, denoted by $\chi_n^2(\lambda)$. The pdf $g(y;n,\lambda)$ of Y is given by (Johnson and Kotz, 1970b, p. 132)

$$g(y;n,\lambda) = \frac{e^{-(1/2)(\lambda+y)}}{2^{n/2}} \sum_{j=0}^{\infty} \frac{y^{(1/2)n+j-1}\lambda^j}{\Gamma\left(\frac{1}{2}n + j\right)2^{2j}j!}, \qquad y > 0,\ \lambda \ge 0$$

A word of caution needs to be given about differences in the literature over the definition of the noncentrality parameter. The majority of sources which we consulted, going back to Patnaik (1949, pp. 202-232), use the definition above. But Searle (1971, p. 49) and Graybill (1961, p. 74) define $\frac{1}{2}\Sigma\mu_i^2$, while Guenther (1964, pp. 957-960) defines $\sqrt{\Sigma\mu_i^2}$ as the noncentrality parameter. These discrepancies in definition may lead to some confusion if one is unaware of them, for example, in determining moments.

The mean, variance, skewness, and kurtosis of a $\chi_\nu^2(\lambda)$ rv (as defined above) are $\nu + \lambda$, $2(\nu + 2\lambda)$, $\sqrt{8}(\nu + 3\lambda)/(\nu + 2\lambda)^{3/2}$, and $3 + 12(\nu + 4\lambda)/(\nu + 2\lambda)^2$, respectively; see Johnson and Kotz (1970b, p. 134) and their chap. 28 for further discussion.

[*5.3.7*] If X has a noncentral χ^2 distribution with two degrees of freedom and noncentrality parameter λ, then

$$\Pr(X \le r) = (2\pi)^{-1} \iint_\Gamma \exp\{-(x^2 + y^2)/2\}\, dy\, dx$$

where Γ is the circle $(x - \sqrt{\lambda})^2 + y^2 \le r^2$. This is the probability that (X,Y) lies in an offset circle of radius r at a distance $\sqrt{\lambda}$ from the origin, when X and Y are independent N(0,1) rvs. See Owen (1962, pp. 172-180), where these probabilities are tabled.

[*5.3.8*] *Some Basic Properties of $\chi_\nu^2(\cdot)$ Relating to Normal Samples*. Let $X_1, \ldots, X_n$ be defined as in [5.3.6] above, writing $\underset{\sim}{X}' = (X_1,\ldots,X_n)$ and $\underset{\sim}{\mu}' = (\mu_1,\mu_2,\ldots,\mu_n)$. Let $\underset{\sim}{A}$, $\underset{\sim}{A}_1$, $\underset{\sim}{A}_2$, ..., $\underset{\sim}{A}_k$ be $n \times n$ symmetric matrices. The following six theorems give properties of linear combinations of the sample variables and of quadratic forms.

(a) The quadratic form $\underset{\sim}{X}'\underset{\sim}{A}\underset{\sim}{X}$ has a noncentral chi-square distribution if and only if $\underset{\sim}{A}$ is idempotent (that is, $\underset{\sim}{A}^2 = \underset{\sim}{A}$), in which case $\underset{\sim}{X}'\underset{\sim}{A}\underset{\sim}{X}$ has degrees of freedom equal to the rank of $\underset{\sim}{A}$, and noncentrality parameter $\underset{\sim}{\mu}'\underset{\sim}{A}\underset{\sim}{\mu}$ (Rao, 1973, pp. 185-186; Searle, 1971, pp. 57-58).

(b) Let $\underset{\sim}{X}'\underset{\sim}{A}_1\underset{\sim}{X}$ and $\underset{\sim}{X}'\underset{\sim}{A}_2\underset{\sim}{X}$ have noncentral chi-square distributions. Then they are independent if and only if $\underset{\sim}{A}_1\underset{\sim}{A}_2 = \underset{\sim}{0}$ (Rao, 1973, p. 187; Searle, 1971, pp. 59-60).

(c) Let $\underset{\sim}{B}$ be a q × n matrix. Then $\underset{\sim}{X}'\underset{\sim}{A}\underset{\sim}{X}$ and $\underset{\sim}{B}\underset{\sim}{X}$ are independent if and only if $\underset{\sim}{B}\underset{\sim}{A} = \underset{\sim}{0}$ (Searle, 1971, p. 59). This result holds whether or not $\underset{\sim}{X}'\underset{\sim}{A}\underset{\sim}{X}$ has a noncentral chi-square distribution.

(d) *Cochran's Theorem.* Let the rank of $\underset{\sim}{A}_j$ be r_j (j = 1, 2, ..., k) and $Q_j = \underset{\sim}{X}'\underset{\sim}{A}_j\underset{\sim}{X}$, such that

$$\sum_{i=1}^{n} X_i^2 = \underset{\sim}{X}'\underset{\sim}{X} = Q_1 + \cdots + Q_k$$

Then Q_j has a $\chi^2_{r_j}(\lambda_j)$ distribution (j = 1, ..., k) and $Q_1, \ldots, Q_k$ are mutually independent if and only if either

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

or

$$\underset{\sim}{A}_j \text{ is idempotent, } j = 1, \ldots, k \quad \text{and} \quad \underset{\sim}{A}_j\underset{\sim}{A}_s = 0 \text{ for all } s \neq j$$

Then $\lambda_j = \underset{\sim}{\mu}'\underset{\sim}{A}_j\underset{\sim}{\mu}$ and $\Sigma_{i=1}^n \mu_i^2 = \Sigma_{i=1}^k \lambda_j$ (Rao, 1973, pp. 185-188; Searle, 1971, pp. 60-64).

(e) If $\underset{\sim}{X}'\underset{\sim}{X} = Q_1 + Q_2$, where Q_1 has a χ^2_a distribution, then Q_2 has a χ^2_{n-a} distribution (Rao, 1973, p. 187), n > a > 0.

(f) If $Q = Q_1 + Q_2$, where Q has a χ^2_a distribution, Q_1 has a χ^2_b distribution (b < a), and Q_2 is nonnegative, then Q_2 has a χ^2_{a-b} distribution (Rao, 1973, p. 187).

(g) $E(\underset{\sim}{X}'\underset{\sim}{A}\underset{\sim}{X}) = \underset{\sim}{\mu}'\underset{\sim}{A}\underset{\sim}{\mu} + \text{trace}(\underset{\sim}{A})$ (Searle, 1971, p. 55). This holds whether or not $\underset{\sim}{X}$ is a normal sample.

[5.3.9] Let $G(y;\nu,\lambda)$ be the cdf corresponding to the pdf $g(y;\nu,\lambda)$ of [5.3.6]. Then, if $a = \sqrt{\lambda} + \sqrt{y}$ and $b = \sqrt{\lambda} - \sqrt{y}$ (Han, 1975, pp. 213-214),

$$G(y;1,\lambda) = \Phi(a) - \Phi(b)$$

$$G(y;3,\lambda) = \Phi(a) - \Phi(b) + \{\phi(a) - \phi(b)\}/\sqrt{\lambda}$$

$$G(y;5,\lambda) = \Phi(a) - \Phi(b) + (2 - \lambda^{-1})\{\phi(a) - \phi(b)\}/\sqrt{\lambda} - \lambda^{-1}\{a\phi(a) - b\phi(b)\}$$

$$G(y;7,\lambda) = \Phi(a) - \Phi(b) + (3 - 4\lambda^{-1} + 3\lambda^{-2})\{\phi(a) - \phi(b)\}/\sqrt{\lambda} - 3(\lambda^{-1} - \lambda^{-2})\{a\phi(a) + b\phi(b)\} + \lambda^{-3/2}\{a^2\phi(a) + b^2\phi(b)\}$$

5.4 SAMPLING DISTRIBUTIONS RELATED TO t

[*5.4.1*] Let Z be a standard normal rv and U a χ_k^2 random variable, independent of Z. Then the rv Y, defined by

$$Y = Z\sqrt{k/U}$$

has a *Student t distribution* with k degrees of freedom, denoted by t_k. The quantiles are $t_{k;\beta}$, where $\Pr(Y \le t_{k;\beta}) = 1 - \beta$, and the pdf g(y;k) of Y is given by (Mood et al., 1974, p. 250)

$$g(y;k) = (k\pi)^{-1/2} \frac{\Gamma\{(k+1)/2\}}{\Gamma(k/2)} \left(1 + \frac{y^2}{k}\right)^{-(k+1)/2}$$

$$-\infty < y < \infty \qquad k = 1, 2, \ldots$$

This is a Pearson Type VII distribution. The distribution is symmetrical about zero; the odd moments vanish, when they exist. The rth moment exists if and only if $k > r$. The variance and kurtosis of Y are $k/(k - 2)$ and $3 + 6/(k - 4)$, respectively, and they exist only if $k > 2$ and $k > 4$, respectively (Johnson and Kotz, 1970b, p. 96); see also [5.1.1].

[*5.4.2*] With the notation of [5.4.1],

$$\lim_{k\to\infty} g(y;k) = \phi(y)$$

Further (Johnson and Kotz, 1970b, p. 95),

$$\lim_{k\to\infty} \Pr(Y \le y) = \Phi(y)$$

[*5.4.3*] (a) Let $\bar{X}$ be the mean $\Sigma X_i/n$ and S^2 be the sample variance $\Sigma(X_i - \bar{X})^2/(n - 1)$ of a sample of size n under normality $(n \ge 2)$. If $T = \sqrt{n}(\bar{X} - \mu)/S$, then T has a t_{n-1} distribution (Mood et al., 1974, p. 250).

(b) If $U = \sqrt{n}(\bar{X} - \mu)/\{\Sigma_{i=1}^n (X_i - \mu)^2/n\}$, then U does not have a t distribution, because $\bar{X}$ and $\Sigma(X_i - \mu)^2$ are not independent. In fact, $\Pr(|U| \le \sqrt{n}) = 1$, since $U = \sqrt{n}T/(T^2 + n - 1)^{1/2}$, where T is defined in (a) above (D. B. Owen, personal communication).

[*5.4.4*] Let $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ be independent iid samples of $N(\mu,\sigma^2)$ and $N(\xi,\sigma^2)$ random variables, respectively.

If

$$T = \frac{\sqrt{mn/(m+n)}\{\bar{X} - \bar{Y} - (\mu - \xi)\}}{[\{\Sigma(X_i - \bar{X})^2 + \Sigma(Y_j - \bar{Y})^2\}/(m+n-2)]^{1/2}}$$

where $\bar{X} = \Sigma X_i/m$ and $\bar{Y} = \Sigma Y_j/n$, then T has a t_{m+n-2} distribution (Mood et al., 1974, p. 435).

[*5.4.5*] If X_0 and X_1 are independent χ_k^2 rvs, then the random variable Y has a t_k distribution, where (Cacoullos, 1965, p. 528)

$$Y = \frac{1}{2}\sqrt{k}(X_1 - X_0)/\sqrt{X_0 X_1}$$

[*5.4.6*] Chu (1956, pp. 783-784) derived the following inequalities for the cdf $G(y;\nu)$ of a t_ν distribution; if $a \geq 0$, $b \geq 0$, and $\nu \geq 3$,

$$\begin{aligned} \frac{\nu}{\nu+1}\left\{\Phi\left(b\sqrt{\frac{\nu+1}{\nu}}\right) - \Phi\left(-a\sqrt{\frac{\nu+1}{\nu}}\right)\right\} &\leq G(b;\nu) - G(-a;\nu) \\ &\leq \sqrt{\frac{7\nu-3}{7\nu-14}}\left\{\Phi\left(b\sqrt{\frac{\nu-2}{\nu}}\right) \right. \\ &\qquad \left. - \Phi\left(-a\sqrt{\frac{\nu-2}{\nu}}\right)\right\} \end{aligned}$$

For these inequalities, putting one of the arguments equal to zero gives

$$\begin{aligned} \frac{\nu}{\nu+1}\Phi\left(a\sqrt{\frac{\nu+1}{\nu}}\right) + \frac{1}{2(\nu+1)} \leq G(a;\nu) &\leq \sqrt{\frac{7\nu-3}{7\nu-14}}\Phi\left(a\sqrt{\frac{\nu-2}{\nu}}\right) \\ &\qquad + \frac{1}{2}\left(1 - \sqrt{\frac{7\nu-3}{7\nu-14}}\right) \end{aligned}$$

where $a \geq 0$ and $\nu \geq 3$. If $\bar{\Phi}(x) = 1 - \Phi(x)$ and $\bar{G}(a,\nu) = 1 - G(a,\nu)$, these inequalities are reversed when Φ and G are replaced by $\bar{\Phi}$ and $\bar{G}$, respectively.

Chu (1956) gives sharper inequalities than the above, i.e.,

$$\begin{aligned} c_\nu\sqrt{\frac{2}{\nu+1}}\left\{\Phi\left(b\sqrt{\frac{\nu+1}{\nu}}\right) - \Phi\left(-a\sqrt{\frac{\nu+1}{\nu}}\right)\right\} &\leq G(b;\nu) - G(-a;\nu) \\ \leq c_\nu\sqrt{\frac{2}{\nu-2}}\left\{\Phi\left(b\sqrt{\frac{\nu-2}{\nu}}\right) - \Phi\left(-a\sqrt{\frac{\nu-2}{\nu}}\right)\right\} & \end{aligned}$$

$$c_\nu = \frac{\Gamma\{(\nu + 1)/2\}}{\Gamma(\nu/2)}, \qquad a \ge 0,\ b \ge 0,\ \nu \ge 3$$

[*5.4.7*] Let X be a $N(\mu,\sigma^2)$ rv and U/σ^2 a χ_k^2 rv, independent of X. Then the rv Y defined by

$$Y = X\sqrt{k/U}$$

has a *noncentral t distribution* with k degrees of freedom and noncentrality parameter δ, where $\delta = \mu/\sigma$; we denote this by $t_k(\delta)$. The pdf is $g(y;k,\delta)$, where (Johnson and Kotz, 1970b, p. 205)

$$g(y;k,\delta) = \frac{k^{k/2}\exp(-\delta^2/2)}{\sqrt{\pi}(k + y^2)^{(k+1)/2}} \sum_{i=0}^{\infty} \left[\frac{\Gamma\{(k + i + 1)/2\}}{i!\Gamma(k/2)}\left(\frac{2^{i/2}}{k + y^2}\right)(\delta y)^i\right]$$

$$-\infty < y < \infty,\ -\infty < \delta < \infty,\ k = 1, 2, \ldots$$

Merrington and Pearson (1958, p. 485) give the first four moments of Y about the origin:

$$E(Y) = \frac{\sqrt{\nu/2}\,\delta\Gamma\{(\nu - 1)/2\}}{\Gamma(\nu/2)}$$

$$\mu_2'(Y) = \frac{(1 + \delta^2)\nu}{\nu - 2}$$

$$\mu_3'(Y) = \frac{(\nu/2)^{3/2}\delta(3 + \delta^2)\Gamma\{(\nu - 3)/2\}}{\Gamma(\nu/2)}$$

$$\mu_4'(Y) = \frac{(3 + 6\delta^2 + \delta^4)\nu^2}{(\nu - 2)(\nu - 4)}$$

[*5.4.8*] Let $X_1, \ldots, X_n$ be iid $N(\mu,\sigma^2)$ rvs, and let S^2 be the sample variance. If $Y = \sqrt{n}\,\bar{X}/S$, then Y has a $t_{n-1}(\sqrt{n}\mu/\sigma)$ distribution.

[*5.4.9*] Let $G(y;k,\delta)$ be the cdf corresponding to the pdf $g(y;k,\delta)$ of [5.4.7]. Then (Owen, 1968, pp. 465-466)

$$G(0;k,\delta) = \Phi(-\delta)$$

$$G(1;1,\delta) = 1 - [\Phi(\delta/\sqrt{2})]^2$$

$$G(y;k,\delta) = \frac{\sqrt{2\pi}}{\Gamma(k/2)2^{(k-2)/2}} \int_0^\infty \Phi\left(\frac{yx}{\sqrt{k}} - \delta\right) x^{k-1}\phi(x)\,dx$$

See also Hawkins (1975, p. 43).

[*5.4.10*] The *sample coefficient of variation* (CV)v is defined by

$$v = S/\bar{X}$$

In a random sample of size n from a normal population, let $Y = \sqrt{n}/v$. Then Y has a noncentral $t_{n-1}(\sqrt{n}/V)$ distribution, where $V = \sigma/\mu$, the coefficient of variation in the underlying population (Johnson and Welch, 1940, pp. 362-363).

[*5.4.11*] Iglewicz et al. (1968, p. 581) give the following approximation when $V \le 0.5$ to the percentiles of the sample CV v, where $\Pr(v \le v_p) = 1 - p$ and $\chi^2_{n-1;p}$ is defined similarly in [5.3.1]:

$$v_p \simeq V\left(\frac{\chi^2_{n-1;p}}{n-1}\right)^{1/2}\left[1 + \frac{V^2}{2n}\{\chi^2_{n-1;p} - (n-2)\} + \frac{V^4}{8n^2}\{3(\chi^2_{n-1;p})^2 - 8(n-2)\chi^2_{n-1;p} + (n-2)(5n-12)\}\right]$$

This approximation is accurate to four decimal places if $n \ge 5$ and if *either* $V \le 0.3$ *or* the first two terms only are used and $V \le 0.2$. If $0.3 \le v \le 0.5$, the approximation is still "quite good."

[*5.4.12*] Iglewicz and Myers (1970, pp. 166-169) compared six approximations to percentiles of v. Three of the best of these follow.

(a) Pearson distributions were fitted to the (approximate) first four moments of v, and approximate percentage points were obtained from tabled values in Johnson et al. (1963, pp. 459-498).

(b) One can approximate $Bv^2/(1 + v^2)$ as a χ^2_{n-1} rv, where $B = n(1 + V^2)/V^2$. If the probability of v being negative is negligible, one finds that (McKay, 1932, pp. 695-698)

$$v_p \simeq \{n/(n-1)\}^{1/2}\{\chi^2_{n-1;p}/(B - \chi^2_{n-1;p})\}^{1/2}$$

(c) Approximate the distribution of v as a $N(V, (V^4 + \frac{1}{2}V^2)/n)$ rv. For moderate or large values of n, the Pearson approximation in (a) gives the best results, but it is cumbersome to use.

If simplicity is a factor, then approximation (b) is uniformly more accurate than that in (c).

Iglewicz and Myers (1970) table exact and approximate values of v_p for p = 0.99, 0.95, 0.05, and 0.01, for V = 0.1(0.1)0.4, and for n = 10, 20, 30, 50.

[*5.4.13*] David (1949, p. 387) gives approximations to the first four moments of the distribution of v under normality: to terms of order n^{-2},

$$E\left(\frac{v}{V}\right) \simeq 1 + \frac{V^2}{n} - \frac{1}{4(n-1)} + \frac{3V^4}{n^2} - \frac{V^2}{4n(n-1)} + \frac{1}{32(n-1)^2}$$

$$\mathrm{Var}\left(\frac{v}{V}\right) \simeq \frac{V^2}{n} + \frac{1}{2(n-1)} + \frac{8V^4}{n^2} + \frac{V^2}{n(n-1)} - \frac{1}{8(n-1)^2}$$

$$\mu_3\left(\frac{v}{V}\right) \simeq \frac{6V^4}{n^2} + \frac{3V^2}{n(n-1)} + \frac{1}{4(n-1)^2}$$

$$\mu_4\left(\frac{v}{V}\right) \simeq \frac{3V^4}{n^2} + \frac{3V^2}{n(n-1)} + \frac{3}{4(n-1)^2}$$

These approximations assume that V is not large. If μ is close to zero, however, V could be large, and a correspondingly large value of n would be required.

[*5.4.14*] Let v be the sample CV and r the number of negative sample values out of n. Then $v \ge \{n/(n-1)\}\{r/(n-r)\}^{1/2}$. This inequality does not depend upon normality (Summers, 1965, p. 67).

5.5 DISTRIBUTIONS RELATED TO F

[*5.5.1*] Let U_1 and U_2 be independent χ^2_m and χ^2_n rvs, respectively. Then the rv Y defined by

$$Y = U_1 m^{-1}/(U_2 n^{-1})$$

has an *F distribution* with (m,n) degrees of freedom, denoted by $F_{m,n}$. The quantiles are $F_{m,n;\beta}$, where $\Pr(Y \le F_{m,n;\beta}) = 1 - \beta$, and the pdf g(y;m,n) is given by (Mood et al., 1974, p. 246)

$$g(y;m,n) = m^{m/2}n^{n/2}\{B(\tfrac{1}{2}m, \tfrac{1}{2}n)\}^{-1}y^{(1/2)m-1}(my + n)^{-(1/2)(m+n)}$$

$$y > 0, \quad m,n = 1, 2, \ldots$$

where $B(\cdot,\cdot)$ is the beta function; further,

$$\Pr(Y \le y) = I_\gamma(m/2,n/2)$$

where $\gamma = my/(n + my)$ and $I_\gamma(\cdot,\cdot)$ is the incomplete beta function ratio, tabled by Pearson (1934). The mean and variance of Y are $n/(n - 2)$ and $2n^2(m + n - 2)/\{m(n - 2)^2(n - 4)\}$, respectively ($n > 2$ and $n > 4$).

[*5.5.2*] Let $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ be independent iid random samples of $N(\mu,\sigma^2)$ and $N(\xi,\sigma^2)$ rvs, respectively, with sample means $\bar{X}$ and $\bar{Y}$, so that $\bar{X} = \Sigma X_i/m$ and $\bar{Y} = \Sigma Y_j/n$. Then the rv W, defined by

$$W = \frac{\Sigma_{i=1}^m (X_i - \bar{X})^2(m - 1)^{-1}}{\Sigma_{j=1}^n (Y_j - \bar{Y})^2(n - 1)^{-1}}$$

has a $F_{m-1,n-1}$ distribution (Mood et al., 1974, pp. 246-247). Further, the rv $\Sigma(X_i - \bar{X})^2/\{\Sigma(X_i - \bar{X})^2 + \Sigma(Y_j - \bar{Y})^2\}$, which is equal to $(m - 1)W/\{n - 1 + (m - 1)W\}$, has a beta distribution with pdf $x^{(m-3)/2}(1 - x)^{(n-3)/2}/B(\frac{1}{2}(m - 1),\frac{1}{2}(n - 1))$, $0 \le x \le 1$; $m,n \ge 2$ (Johnson and Kotz, 1970b, p. 78).

[*5.5.3*] If T is a rv with a t_ν distribution, then the rv T^2 has a $F_{1,\nu}$ distribution. Cacoullos (1965, p. 529) showed that

$$F_{n,n;\alpha} = 1 + 2n^{-1}t_{n;\alpha}^2 + 2t_{n;\alpha}(n^{-1} + n^{-2}t_{n;\alpha}^2)^{1/2}$$

$$t_{n;\alpha} = \frac{1}{2}\sqrt{n}(F_{n,n;\alpha}^{1/2} - F_{n,n;\alpha}^{-1/2})$$

[*5.5.4*] Let U have a noncentral $\chi_m^2(\lambda)$ distribution, and V a χ_n^2 distribution, where U and V are independent. Then the rv Y defined by

$$Y = Um^{-1}/Vn^{-1}$$

has a *noncentral F distribution* with (m,n) degrees of freedom and noncentrality parameter λ. See [5.3.6] above for a caution regarding various differences in defining λ in the literature; here it is based on that in [5.3.6]. Denote noncentral F by $F_{m,n}(\lambda)$, and its quantiles by $F_{m,n;\beta}(\lambda)$, where $\Pr(Y \le F_{m,n;\beta}(\lambda)) = 1 - \beta$. The pdf $g(y;m,n,\lambda)$ is given by

$$g(y;m,n,\lambda) = \sum_{j=0}^{\infty} \left[\frac{e^{-(1/2)\lambda}[(1/2)\lambda]^j}{j!B[(1/2)m + j,(1/2)n]}\left(\frac{m}{n}\right)^{(1/2)m+j} y^{(1/2)m+j-1} \times \left(1 + \frac{m}{n}y\right)^{-(1/2)(m+n)-j}\right], y > 0,\ \lambda > 0,\ m,n = 1, 2, \ldots$$

where $B(\cdot,\cdot)$ is the beta function. The mean and variance of Y are $n(m + \lambda)/\{m(n - 2)\}$; $(n > 2)$, and $2(n/m)^2\{(m + \lambda)^2 + (m + 2\lambda)/(n - 2)\}/\{(n - 2)^2(n - 4)\}$; $(n > 4)$, respectively (Johnson and Kotz, 1970b, pp. 190-191).

[*5.5.5*] Let U, V, and Y be defined as in [5.5.4], except that V has a noncentral $\chi_n^2(\eta)$ distribution. Then Y has a *doubly noncentral F distribution* with (m,n) degrees of freedom and noncentrality parameters (λ,η). See Johnson and Kotz (1970b, chap. 30) for further discussion.

[*5.5.6*] If the rv X has a noncentral $F_{m,n}(\lambda)$ distribution, and if $U = mX/(n + mX)$, then the rv U has a *noncentral beta distribution* denoted by $B[(1/2)m,(1/2)n;\lambda]$, with pdf $g(u;m,n,\lambda)$ given by

$$g(u;m,n,\lambda) = \frac{e^{-\lambda/2} \sum_{j=0}^{\infty} \left(\frac{1}{2}\lambda\right)^j (j!)^{-1} u^{(1/2)m+j-1}(1 - u)^{(1/2)n-1}}{B[(1/2)m + j,\ (1/2)n]}$$

$$0 \le u \le 1,\quad \lambda > 0;\quad m,n = 1, 2, \ldots$$

See Graybill (1961, p. 79), where his noncentrality parameter λ is half the corresponding λ here; the caution given in [5.3.6] applies.

The cdf of the above distribution is $G(u;m,n,\lambda)$, where

$$G(u;m,n,\lambda) = e^{-\lambda/2} \sum_{j=0}^{\infty} \left(\frac{1}{2}\lambda\right)^j (j!)^{-1} I_u\left(\frac{1}{2}m + j, \frac{1}{2}n\right), \quad 0 \le u \le 1$$

where $I_u(\cdot,\cdot)$ is the incomplete beta function ratio, tabled by Pearson (1934). An alternative form for the pdf is given by Seber (1963, p. 542).

[5.5.7] Let X_i be a $N(\mu_i,1)$ rv ($i = 1, \ldots, m$) and Y_j a $N(0,1)$ rv ($j = 1, \ldots, n$), where $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ are mutually independent. Let $S = \Sigma X_i^2$, $T = \Sigma Y_j^2$, and $\lambda = \Sigma \mu_i^2$. Then the rv $S/(S + T)$ has a noncentral $B(\frac{1}{2}m,\frac{1}{2}n;\lambda)$ distribution, given above in [5.5.6]. See also Hodges (1955, pp. 648-653), who defines noncentral beta in terms of $T/(S + T)$, that is, $1 - S(S + T)^{-1}$, rather than by $S/(S + T)$.

5.6 THE SAMPLE MEAN DEVIATION

[5.6.1] The *mean deviation* d of a normal sample of size n is defined by

$$d = \Sigma|X_i - \bar{X}|/n$$

Godwin (1945, pp. 254-256) obtained the distribution of d, but it is not simple or easily tractable. Pearson and Hartley (1966, table 21) give upper and lower 0.1, 0.5, 1.0, 2.5, 5.0, and 10.0 percentiles of the distribution of d in a normal sample when $\sigma = 1$, for $n = 2(1)10$, to three decimal places. Pearson and Hartley (1972, table 8) give values of the cdf of d when $\sigma = 1$ for $d = 0.0(0.01)3.0$ and $n = 2(1)10$, to five decimal places. A method for approximating to the distribution when $n > 10$ is given in Pearson and Hartley (1966, p. 89). Cadwell (1954, pp. 12-17) shows that $c(d/\sigma)^{1.8}$ has an approximate χ_q^2 distribution for values of c and q tabled when $n = 4(1)10(5)50$. See also Cadwell (1953, pp. 336-339, 342-343). Note that the distribution of d has an approximate $N(\sigma\sqrt{2/\pi}, (1 - 2\pi^{-1})\sigma^2/n)$ distribution, but only when n is very large.

[5.6.2] Moments of d (Herrey, 1965, p. 259; Kamat, 1954, pp. 541-542). Exact values of the first three moments of d/σ under normality are as follows:

$$E(d) = \sigma\sqrt{\frac{2(n - 1)}{n\pi}}$$

$$\mathrm{Var}(d) = \frac{2\sigma^2(n - 1)[(1/2)\pi + (n^2 - 2n)^{1/2} - n + \arcsin\{(n - 1)^{-1}\}]}{n^2\pi}$$

$$\mu_3(d) = \sigma^3\left(\frac{n - 1}{n}\right)^{3/2}\left[\sqrt{\frac{2}{\pi}}\,\frac{(4 - n)}{n^2(n - 1)} + \left(\frac{2}{\pi}\right)^{3/2}\left\{\frac{(n - 2)}{n}\left(\frac{n - 3}{n - 1}\right)^{1/2}\right.\right.$$
$$+ 2 + \frac{3(n - 2)^2}{n^2(n - 1)}\arcsin\left(\frac{1}{n - 2}\right) - \frac{3(n - 1)}{n}$$
$$\left.\left.\times\left[1 - \frac{1}{(n - 1)^2}\right]^{1/2} - \frac{3}{n}\arcsin\left(\frac{1}{n - 1}\right)\right\}\right]$$

Kamat (1954) also gives an expression for the fourth moment.

Geary (1936, p. 301) gave expansions for several moments; with a refinement for μ_3 and μ_4 by Pearson (1945, p. 252), these are ($\nu = n - 1$):

$$\sigma^{-2}\mathrm{Var}(d) = \{(n - 1)/n\}\{(0.0450,70)\nu^{-1} - (0.1246,48)\nu^{-2}$$
$$+ (0.0848,59)\nu^{-3} + (0.0063,23)\nu^{-4}\} + O(n^{-5})$$
$$\sigma^{-3}\mu_3(d) = \{(n - 1)/n\}^{3/2}\{(0.2180,14)\nu^{-2} - (0.0741,70)\nu^{-3}$$
$$+ (0.0573,13)\nu^{-4} - (0.0404,57)\nu^{-5}\} + O(n^{-6})$$
$$\sigma^{-4}\mu_4(d) = 3\{\mathrm{Var}(d)/\sigma^2\}^2 + \{(n - 1)/n\}^2\{(0.1147,71)\nu^{-3}$$
$$- (0.0685,09)\nu^{-4} + (0.0333,71)\nu^{-5}\} + O(n^{-6})$$

Johnson (1958, p. 481) gives the approximation

$$\sigma^{-2}\mathrm{Var}(d) \simeq n^{-1}(1 - 2/\pi)\{1 - (0.12)n^{-1}\}$$

See Pearson and Hartley (1966, pp. 41-42) for references to fuller information for small samples; in their table 20, values of E(d), Var(d), s.d.(d), and the shape factors are given for n = 2(1)20,30, 60 to decimal places varying from three to six.

[*5.6.3*] The ratio of the sample mean deviation to sample standard deviation, d/S, is used for detecting changes in kurtosis, i.e., as a test of normality. The statistic which has been studied is known as *Geary's* a, where

$$a = \Sigma|X_i - \bar{X}|/[n\Sigma\{(X_i - \bar{X})^2\}]^{1/2} = d/\sqrt{m_2} = \{\sqrt{n/(n - 1)}\}d/S$$

The corresponding value in the normal population is $\nu_1/\sqrt{\nu_2} = \nu_1/\sqrt{\mu_2} = \sqrt{2/\pi}$, or 0.7978,8456. Geary (1936, p. 303) gives the following approximation to the pdf g(a) of a: if $E(a) = \eta$ and $\mu_3(a)/\{\text{Var}(a)\}^{3/2} = \lambda$, the shape factor for skewness of the distribution of a, then

$$g(a) \simeq \frac{1}{\sigma\sqrt{2\pi}}\left[1 - \frac{\lambda}{6}\left\{\frac{3(a - \eta)}{\sigma} - \frac{(a - \eta)^3}{\sigma^3}\right\}\right]\exp\left\{-\frac{(a - \eta)^2}{2\sigma^2}\right\}, \quad a > 0$$

Expressions for the moments appearing in this approximation are given in [5.6.6] below; the approximation is the basis of Pearson and Hartley's (1966) table 34A, which gives upper and lower 1, 5, and 10 percentiles of g(a) to four decimal places for n - 1 = 10(5)50(10)100(100)1000, and corresponding values of E(a) and s.d.(a) to five decimal places. Geary (1936, pp. 304-305) gives charts from which the values of the above percentiles can be read off for sample sizes between 11 and 1001.

[*5.6.4*] Geary's a and the sample variance S^2 are independent (Geary, 1936, p. 296).

[*5.6.5*] The moments about the origin of a satisfy the relation

$$E(a^r) = \{n/(n - 1)\}^{r/2}E(d^r)/E(S^r)$$

where d is the sample mean deviation and S^2 the sample variance (Geary, 1936, p. 296).

[*5.6.6*] Exact expressions for the first four moments of a about the origin are given by

$$E(a) = \sqrt{(n - 1)/\pi}\,\Gamma\{(n - 1)/2\}/\Gamma(n/2)$$

$$E(a^2) = n^{-1} + \{2/(n\pi)\}[(n^2 - 2n)^{1/2} + \arcsin\{(n - 1)^{-1}\}]$$

$$E(a^3) = \{n/(n - 1)\}^{1/2}E(d^3/\sigma^3)/E(S/\sigma)$$

$$E(a^4) = \{n^2/(n^2 - 1)\}E(d^4/\sigma^4)$$

Geary (1936, p. 301) gives the following expansions in powers of ν^{-1}, where $\nu = n - 1$ and $\sqrt{2/\pi} = 0.7978{,}846$:

$$E(a) = (\sqrt{2/\pi}) + (0.1994{,}71)\nu^{-1} + (0.0249{,}34)\nu^{-2} - (0.0311{,}68)\nu^{-3} - (0.0081{,}82)\nu^{-4} + 0(n^{-5})$$

$$\text{Var}(a) = (0.0450{,}70)\nu^{-1} - (0.1246{,}48)\nu^{-2} + (0.1098{,}49)\nu^{-3} + (0.0063{,}23)\nu^{-4} + 0(n^{-5})$$

$$\mu_3(a) = -(0.0168{,}57)\nu^{-2} + (0.0848{,}59)\nu^{-3} - (0.2418{,}25)\nu^{-4} + 0(n^{-5})$$

$$\mu_4(a) = 3\{\text{Var}(a)\}^2 + (0.0110{,}51)\nu^{-3} - (0.1454{,}43)\nu^{-4} + 0(n^{-5})$$

and for the shape factors

$$\mu_3(a)/\{\text{Var}(a)\}^{3/2} = -1.7618\{1 - (2.3681)\nu^{-1} - (8.8646)\nu^{-2}\}/\sqrt{\nu} + 0(n^{-7/2})$$

$$\mu_4(a)/\{\text{Var}(a)\}^2 = 3 + 5.441\{1 - (7.628)\nu^{-1}\}/\nu + 0(\nu^{-3})$$

5.7 THE MOMENT RATIOS $\sqrt{b_1}$ AND b_2

The sample rth moment m_r of a random sample $X_1, \ldots, X_n$ from any population is $\Sigma_{i=1}^n\{(X_i - \bar{X})^r\}/n$, calculated about the sample mean $\bar{X}$. The results which follow were obtained only after R. A. Fisher had developed the structure which he called k-statistics. These are symmetric functions of the observations; the rth k-statistic k_r has as its mean the rth cumulant κ_r of the underlying population; see Kendall and Stuart (1977, pp. 296-300).

[*5.7.1*] The third sample moment m_3, where $m_3 = \Sigma(X_i - \bar{X})^3/n$, has been studied in standardized form rather than on its own. We consider $\sqrt{b_1}$, a measure of *skewness* of the sample, where

$$\sqrt{b_1} = m_3/m_2^{3/2} = \sqrt{n}\ \Sigma\{(X_i - \bar{X})^3\}/\sqrt{\Sigma}\{(X_i - \bar{X})^2\}]^{3/2}$$

a statistic which has been used in tests of normality. The sampling distribution under normality has been approximated by a Pearson Type VII (Student t form) distribution (Pearson, 1963, pp. 95-112; 1965, pp. 282-285), as well as by a symmetric Johnson (1949, pp. 149-176) S_u distribution, the former having been used to tabulate percentage points. Pearson and Hartley (1966, table 34B) table 5 and 1 percentiles and values of the standard deviation of $\sqrt{b_1}$ under normality for n = 25(5)50(10)100(25)200(50)1000(200)2000(500)5000. Note, however, that while the third edition of *Biometrika Tables for Statisticians*, Vol. 1 (Pearson and Hartley, 1966) contains corrections by Pearson (1965), the earlier editions do not have these.

[5.7.2] *Moments of* $\sqrt{b_1}$ *under Normality* (Fisher, 1930, pp. 16-28; Geary, 1947, p. 68). The distribution being symmetric about the origin, the odd moments all vanish.

$$\mathrm{Var}(\sqrt{b_1}) = 6(n - 2)/\{(n + 1)(n + 3)\}$$

$$\mu_4(\sqrt{b_1}) = \frac{108(n - 2)(n^2 + 27n - 70)}{(n + 1)(n + 3)(n + 5)(n + 7)(n + 9)}$$

Pearson (1930, p. 242) has given expansions in powers of 1/n of the standard deviation and kurtosis:

$$\mathrm{s.d.}(\sqrt{b_1}) = \sqrt{(6/n)}\ (1 - 3n^{-1} + 6n^{-2} - 15n^{-3}) + 0(n^{-9/2})$$

$$\mu_4(\sqrt{b_1})/\{\mathrm{Var}(\sqrt{b_1})\}^2 = 3 + 36n^{-1} - 864n^{-2} + 1{,}2096n^{-3} + 0(n^{-4})$$

The statistic $[(n + 1)(n + 3)/\{6(n - 2)\}]^{1/2}\sqrt{b_1}$ has unit variance and fourth moment $3 + 36n^{-1} + 0(n^{-2})$, indicating a reasonably rapid approach to normality (Kendall and Stuart, 1977, p. 317).

[5.7.3] A measure of the *kurtosis* of a sample is b_2, where

$$b_2 = n\Sigma\{X_i - \bar{X})^4\}/[\Sigma\{(X_i - \bar{X})^2\}]^2 = m_4/m_2^2$$

and $m_4 = \Sigma\{(X_i - \bar{X})^4\}/n$, the sample fourth moment. The statistic b_2 has also been used in tests of normality; its sampling distribution under normality can be approximated by a skew Pearson Type IV or skew Johnson (1949, pp. 149-176) S_u distribution, by equating the first four moments (Pearson, 1963, pp. 95-112). Using the closeness of these to the exact fifth moment as a criterion of accuracy of approximation, Pearson (1965, pp. 282-285) found the S_u fit to be better if $n < 80$, and the Pearson Type IV to be better if $n > 80$. Pearson and Hartley (1966, table 34C) tabulate upper and lower 5 and 1 percentiles of the distribution of b_2 under normality for n = 50(25)150(50)700(100)1000(200)2000(500)5000. Note, however, that while the third edition of *Biometrika Tables for Statisticians*, Vol. 1 (Pearson and Hartley, 1966) contains the percentage points for $25 < n < 200$ given by Pearson (1965, p. 284), the earlier editions do not have these. Pearson (1963, pp. 95-112) noted that even when n = 200, the distribution of b_2 is far from being normal.

[*5.7.4*] *Moment of* b_2 *under Normality* (Fisher, 1930, pp. 16-28; Hsu and Lawley, 1940, p. 246). Recall that for a $N(\mu,\sigma^2)$ rv, the kurtosis is equal to 3.

$$\text{Mean} = \mu_1'(b_2) = 3(n-1)/(n+1)$$

$$\text{Var}(b_2) = 24n(n-2)(n-3)/\{(n+1)^2(n+3)(n+5)\}$$

$$\mu_3(b_2) = \frac{1728n(n-2)(n-3)(n^2-5n+2)}{(n+1)^3(n+3)(n+5)(n+7)(n+9)}$$

$$\mu_4(b_2) = \frac{1728n(n-2)(n-3)(n^5+207n^4-1707n^3+4105n^2-1902n+720)}{(n+1)^4(n+3)(n+5)(n+7)(n+9)(n+11)(n+13)}$$

Pearson (1930, p. 243) gave the following expansions:

$$\text{s.d.}(b_2) = \sqrt{24/n}\{1 - 15(2n)^{-1} + 271(8n^2)^{-1} - 2319(16n^3)^{-1}\} + O(n^{-9/2})$$

$$\text{Skewness} = \frac{\mu_3(b_2)}{\{\mu_2(b_2)\}^{3/2}} = \frac{216}{n}\left(1 - \frac{29}{n} + \frac{519}{n^2} - \frac{7637}{n^3}\right) + O(n^{-5})$$

$$\text{Kurtosis} = \frac{\mu_4(b_2)}{\{\mu_2(b_2)\}^2} = 3 + \frac{540}{n} - \frac{20,196}{n^2} + \frac{470,412}{n^3} + O(n^{-4})$$

5.8 MISCELLANEOUS RESULTS

[5.8.1] *Sampling Distributions under Linear Regression.* Suppose that $Y_1, Y_2, \ldots, Y_n$ are iid and Y_i has an $N(\alpha + \beta(x_i - \bar{x}), \sigma^2)$ distribution; $i = 1, 2, \ldots, n$, where $\bar{x} = \Sigma x_i/n$. Here $x_1, \ldots, x_n$ may be predetermined constants such as temperature, pressure, thickness, etc., or may be realizations of one variable in a bivariate normal sample (see Chapter 10). Let Σ denote summation from $i = 1$ to $i = n$, and

$$a = \bar{Y} = \Sigma Y_i/n, \qquad b = \Sigma\{(x_i - \bar{x})Y_i\}/\Sigma\{(x_i - \bar{x})^2\}$$

$$S^2 = \Sigma_{i=1}^n\{Y_i - a - b(x_i - \bar{x})\}^2/(n - 2)$$

Then $\Sigma\{Y_i - a - b(x_i - \bar{x})\}^2/\sigma^2$ has a χ^2_{n-2} and $(b - \beta)^2\Sigma(x_i - \bar{x})^2$ has independently a χ^2_1 distribution; thus

$$(b - \beta)\sqrt{\{\Sigma(x_i - \bar{x})^2\}}/(n - 2)$$

has a t_{n-2} distribution. If Y is a future observation, with mean $\alpha + \beta(x - \bar{x})$, then

$$\{Y - \alpha - \beta(x - \bar{x})\}/(S[n^{-1} + (x - \bar{x})^2/\Sigma\{(x_i - \bar{x})^2\}]^{1/2})$$

has a t_{n-2} distribution, and so also does (Brownlee, 1965, pp. 335-342)

$$\{Y - a - b(x - \bar{x})\}/(S[1 + n^{-1} + (x - \bar{x})^2/\Sigma\{(x_i - \bar{x})^2\}]^{1/2})$$

[5.8.2] If X has a $N(\theta,1)$ distribution and Y has independently a χ^2_n distribution, let $Q = X/(X^2 + Y^2)^{1/2}$. Then $\sqrt{n}\,Q/(1 - Q^2)^{1/2}$ has a noncentral $t_n(\theta)$ distribution, and Q^2 a noncentral $B(\frac{1}{2},\frac{1}{2}n;\theta^2)$ distribution (see [5.5.6]). Hogben et al. (1964, pp. 298-314) derive moments and (Hogben et al., 1964, p. 316) an approximation to the distribution of Q. The first four central moments are tabled to six decimal places for $\theta = 0.1(0.1)1.0(0.2)2(0.5)6(1)10$ and $n = 1(1)24(5)40,50,60,80,100$.

[*5.8.3*] If X has a $N(\theta,1)$ distribution and Y has independently a noncentral χ_n^2 distribution, then the variables X/Y and $X/(X^2 + Y^2)^{1/2}$ are *generalized noncentral* t and *generalized noncentral beta* variables, respectively. Park (1964, pp. 1584, 1588) has obtained approximations to the first few moments, and asymptotic expressions for the pdfs in the tails of these distributions.

[*5.8.4*] Let $X_1, \ldots, X_n$ be iid N(0,1) random variables, $S_r = X_1 + \cdots + X_r$ ($r = 1, 2, \ldots, n$), and U_n the maximum of the partial sums $S_1, \ldots, S_n$. The first two moments about the origin of U_n are given by

$$\mu_1'(U_n) = (2\pi)^{-1/2} \sum_{s=1}^{n-1} s^{-1/2}$$

$$\mu_2'(U_n) = (n + 1)/2 + (2\pi)^{-1} \sum_{r=1}^{n-2} \sum_{s=1}^{r} \{s(r - s + 1)\}^{-1/2}$$

See, for example, Anis (1956, pp. 80-83), where the first four moments about the mean are tabulated for n = 2(1)15, or Solari and Anis (1957, pp. 706-716), where the moments of the maximum of adjusted partial sums $\{S_r - r\bar{X}\}$ are discussed.

These statistics may be of interest in storage problems.

[*5.8.5*] Let $X_1, \ldots, X_n$ be iid $N(0,\sigma^2)$ rvs, and

$$D^2 = \nu^{-1} \sum_{j=1}^{\nu} (X_{j+1} - X_j)^2, \qquad \nu = n - 1$$

the mean square successive difference. Shah (1970, pp. 193-198) represents the cdf of $D^2/(2\sigma^2)$ in terms of Laguerre polynomials. Harper (1967, p. 421) derived moments of q^2, where $q^2 = (1/2)D^2$:

$$E(q^2) = 1, \quad \mathrm{Var}(q^2) = \nu^{-2}(3\nu - 1), \qquad \nu \ge 1$$

$$\mu_3(q^2) = 4\nu^{-3}(5\nu - 3), \qquad \nu \ge 1$$

$$\mu_4(q^2) = 3\nu^{-4}(9\nu^2 + 64\nu - 57), \qquad \nu \ge 2$$

Harper (1967) gives the first eight moments and cumulants, obtains a table of percentage points of the q^2 distribution, and considers some normal and chi-square approximations.

REFERENCES

The numbers in square brackets give the sections in which the corresponding reference is cited.

Abramowitz, M., and Stegun, I. A. (1964). *Handbook of Mathematical Functions*, Washington, D.C.: National Bureau of Standards. [5.3.3]

Anis, A. A. (1956). On the moments of the maximum of partial sums of a finite number of independent normal variates, *Biometrika* 43, 79-84. [5.8.4]

Aroian, L. A., Taneja, V. S., and Cornwell, L. W. (1978). Mathematical forms of the distribution of the product of two normal variables, *Communications in Statistics* A7(2), 165-172. [5.1.3]

Birnbaum, Z. W., and Saunders, S. C. (1969). A new family of life distributions, *Journal of Applied Probability* 6, 319-327. [5.1.6]

Brownlee, K. A. (1965). *Statistical Theory and Methodology in Science and Engineering*, 2nd ed., New York: Wiley. [5.8.1]

Cacoullos, T. (1965). A relation between t and F distributions, *Journal of the American Statistical Association* 60, 528-531. [5.4.5; 5.5.3]

Cadwell, J. H. (1953). Approximating to the distributions of measures of dispersion by a power of χ^2, *Biometrika* 40, 336-346. [5.6.1]

Cadwell, J. H. (1954). The statistical treatment of mean deviation, *Biometrika* 41, 12-18. [5.6.1]

Chu, J. T. (1956). Errors in normal approximations to t, τ, and similar types of distribution, *Annals of Mathematical Statistics* 27, 780-789. [5.4.6]

Craig, C. C. (1936). On the frequency function of xy, *Annals of Mathematical Statistics* 7, 1-15. [5.1.3]

Daly, J. F. (1946). On the use of the sample range in an analogue of Student's t-test, *Annals of Mathematical Statistics* 17, 71-74. [5.2.3]

David, F. N. (1949). Note on the application of Fisher's k-statistics, *Biometrika* 36, 383-393. [5.3.5; 5.4.13]

Epstein, B. (1948). Some applications of the Mellin transform in statistics, *Annals of Mathematical Statistics* 19, 370-379. [5.1.3]

Feller, W. (1966). *An Introduction to Probability Theory and Its Applications*, Vol. 2, New York: Wiley. [5.1.4]

Fisher, R. A. (1930). The moments of the distribution for normal samples of measures of departure from normality, *Proceedings of the Royal Society of London* A130, 16-28. [5.7.2, 4]

Fox, C. (1965). A family of distributions with the same ratio property as normal distribution, *Canadian Mathematical Bulletin* 8, 631-636. [5.1.1]

Geary, R. C. (1936). Moments of the ratio of the mean deviation to the standard deviation for normal samples, *Biometrika* 28, 295-305. [5.6.2, 3, 4, 5, 6]

Geary, R. C. (1947). The frequency distribution of $\sqrt{b_1}$ for samples of all sizes drawn at random from a normal population, *Biometrika* 34, 68-97. [5.7.2]

Godwin, H. J. (1945). On the distribution of the estimate of mean deviation obtained from samples from a normal population, *Biometrika* 33, 254-256. [5.6.1]

Graybill, F. A. (1961). *An Introduction to Linear Statistical Models*, Vol. 1, New York: McGraw-Hill. [5.3.6; 5.5.6]

Guenther, W. C. (1964). Another derivation of the non-central chi-square distribution, *Journal of the American Statistical Association* 59, 957-960. [5.3.6]

Han, C. P. (1975). Some relationships between noncentral chi-squared and normal distributions, *Biometrika* 62, 213-214. [5.3.9]

Harper, W. M. (1967). The distribution of the mean half-square successive difference, *Biometrika* 54, 419-433. [5.8.5]

Hawkins, D. M. (1975). From the noncentral t to the normal integral, *American Statistician* 29, 42-43. [5.4.9]

Herrey, E. M. J. (1965). Confidence intervals based on the mean absolute deviation of a normal sample, *Journal of the American Statistical Association* 60, 257-269. [5.2.3; 5.6.2]

Hodges, J. L. (1955). On the non-central beta distribution, *Annals of Mathematical Statistics* 26, 648-653. [5.5.7]

Hogben, D., Pinkham, R. S., and Wilk, M. B. (1964). (1) The moments of a variate related to the non-central "t"; (2) An approximation to the distribution of Q, *Annals of Mathematical Statistics* 35, 298-318. [5.8.2]

Hogg, R. V. (1960). Certain uncorrelated statistics, *Journal of the American Statistical Association* 62, 265-267. [5.2.3]

Hsu, C. T., and Lawley, D. N. (1939). The derivation of the fifth and sixth moments of the distribution of b_2 in samples from a normal population, *Biometrika* 31, 238-248. [5.7.4]

Iglewicz, B., and Myers, R. H. (1970). Comparisons of approximations to the percentage points of the sample coefficient of variation, *Technometrics* 12, 166-169. [5.4.12]

Iglewicz, B., Myers, R. H., and Howe, R. B. (1968). On the percentage points of the sample coefficient of variation, *Biometrika* 55, 580-581. [5.4.11]

Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation, *Biometrika* 36, 149-176. [5.7.1, 3]

Johnson, N. L. (1958). The mean deviation, with special reference to samples from a Pearson Type III population, *Biometrika* 45, 478-483. [5.6.2]

Johnson, N. L., and Kotz, S. (1970a). *Distributions in Statistics: Continuous Univariate Distributions*, Vol. 1, New York: Wiley. [5.2.3; 5.3.1]

Johnson, N. L., and Kotz, S. (1970b). *Distributions in Statistics: Continuous Univariate Distributions*, Vol. 2, New York: Wiley. [5.3.6; 5.4.1, 2, 7; 5.5.2, 4, 5]

Johnson, N. L., Nixon, E., Amos, D. E., and Pearson, E. S. (1963). Tables of percentage points of Pearson curves, for given $\sqrt{\beta_1}$ and β_2, expressed in standard measure, *Biometrika* 50, 459-498. [5.4.12]

Johnson, N. L., and Welch, B. L. (1939). On the calculation of the cumulants of the χ-distribution, *Biometrika* 31, 216-218. [5.3.5]

Johnson, N. L., and Welch, B. L. (1940). Applications of the noncentral t distribution, *Biometrika* 31, 362-389. [5.4.10]

Kamat, A. R. (1954). Moments of the mean deviation, *Biometrika* 41, 541-542. [5.6.2]

Kendall, M. G., and Stuart, A. (1977). *The Advanced Theory of Statistics*, Vol. 1 (4th ed.), New York: Macmillan. [5.1.1, 3; 5.3.4; 5.7; 5.7.2]

Lancaster, H. O. (1969). *The Chi-Squared Distribution*, New York: Wiley. [5.3.1, 3]

Lehmann, E. L. (1959). *Testing Statistical Hypotheses*, New York: Wiley. [5.2.4]

McKay, A. (1932). Distribution of the coefficient of variation and the extended "t" distribution, *Journal of the Royal Statistical Society* 95, 695-698. [5.4.12]

Mann, N. R., Schafer, R. E., and Singpurwalla, N. D. (1974). *Methods for Statistical Analysis of Reliability and Life Data*, New York: Wiley. [5.1.6]

Mantel, N. (1973). A characteristic function exercise, *The American Statistician* 27(1), 31. [5.1.5]

Marsaglia, G. (1965). Ratios of normal variables and ratios of sums of uniform variables, *Journal of the American Statistical Association* 60, 193-204. [5.1.1]

Merrington, M., and Pearson, E. S. (1958). An approximation to the distribution of noncentral "t," *Biometrika* 45, 484-491. [5.4.7]

Mood, A. M., Graybill, F. A., and Boes, D. C. (1974). *Introduction to the Theory of Statistics*, 3rd ed., New York: McGraw-Hill. [5.1.2; 5.3.2, 4; 5.4.1, 3, 4; 5.5.1, 2]

Owen, D. B. (1962). *Handbook of Statistical Tables*, Reading, Mass.: Addison-Wesley. [5.3.7]

Owen, D. B. (1968). A survey of properties and applications of the noncentral t-distribution, *Technometrics* 10, 445-478. [5.4.9]

Park, J. H. (1964). Variations of the non-central t and beta distributions, *Annals of Mathematical Statistics* 35, 1583-1593. [5.8.3]

Patnaik, P. B. (1949). The non-central χ^2- and F-distributions and their applications, *Biometrika* 36, 202-232. [5.3.6]

Pearson, E. S. (1930). A further development of tests for normality, *Biometrika* 22, 239-249. [5.7.2, 4]

Pearson, E. S. (1945). The probability integral of the mean deviation, *Biometrika* 33, 252-253. [5.6.2]

Pearson, E. S. (1963). Some problems arising in approximating to probability distributions, using moments, *Biometrika* 50, 95-112. [5.7.1, 3]

Pearson, E. S. (1965). Tables of percentage points of $\sqrt{b_1}$ and b_2 in normal samples: A rounding off, *Biometrika* 52, 282-285. [5.7.1, 3]

Pearson, E. S., and Hartley, H. O. (1966). *Biometrika Tables for Statisticians*, Vol. 1, 3rd ed., London: Cambridge University Press. [5.3.5; 5.6.1, 2, 3; 5.7.1, 3]

Pearson, E. S., and Hartley, H. O. (1972). *Biometrika Tables for Statisticians*, Vol. 2, London: Cambridge University Press. [5.6.1]

Pearson, K., ed. (1934). *Tables of the Incomplete Beta Function*, London: Cambridge University Press. [5.5.1, 6]

Puri, P. S. (1973). On a property of exponential and geometric distributions and its relevance to multivariate failure rate, *Sankhyā* A35, 61-78. [5.3.3]

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed., New York: Wiley. [5.2.2; 5.3.8]

Searle, S. R. (1971). *Linear Models*, New York: Wiley. [5.3.6, 8]

Seber, G. A. F. (1963). The noncentral chi-squared and beta distributions, *Biometrika* 50, 542-545. [5.5.6]

Shah, B. K. (1970). On the distribution of half the mean square successive difference, *Biometrika* 57, 193-198. [5.8.5]

Shepp, L. (1964). Problem 62-9: Normal functions of normal random variables, *SIAM Review* 6, 459. [5.1.4]

Solari, M. E., and Anis, A. A. (1957). The mean and variance of the maximum of the adjusted partial sums of a finite number of independent normal variates, *Annals of Mathematical Statistics* 28, 706-716. [5.8.4]

Steck, G. P. (1958). A uniqueness property not enjoyed by the normal distribution, *Annals of Mathematical Statistics* 29, 604-606. [5.1.1]

Summers, R. D. (1965). An inequality for the sample coefficient of variation and an application to variables sampling, *Technometrics* 7, 67. [5.4.14]

chapter 6

LIMIT THEOREMS AND EXPANSIONS

The normal distribution holds a chief place among statistical distributions on account of the Central Limit Theorem, discussed in Section 6.1. The most widely used version of this theorem states that the standardized sample mean of a random sample of size n from an infinite population with common mean μ and finite variance tends to the standard normal distribution as $n \to \infty$. That is, if $\bar{G}_n$ is the cdf of the standardized sample mean, then $\bar{G}_n(x) \to \Phi(x)$ for every value of x. In fact, if $\{\bar{G}_n(x): n = 1, 2, \ldots\}$ is a sequence of cdfs such that $\bar{G}_n(x) \to \Phi(x)$ pointwise for all x as $n \to \infty$, then the convergence is uniform in x (Ash, 1972, p. 358). For some distributions, convergence may be rapid, while for others it can be very slow. This is an important consideration in trying to use a normal approximation based on the theorem, which in itself gives no information about the rapidity of convergence. The Berry-Esseen theorem, given in Section 6.3, gives some information in the form of upper bounds to $|\bar{G}_n(x) - \Phi(x)|$ which are proportional to $1/\sqrt{n}$. A number of expansions of $\bar{G}_n(x)$ about $\Phi(x)$ provide, in addition (see Section 6.4) approximations to $\bar{G}_n(x)$ for large or even moderate values of n.

In the next chapter, we consider normal approximations to the commonly used distributions in statistics; many of these approximations involve modifications of the central limit property.

6.1 CENTRAL LIMIT THEOREMS FOR INDEPENDENT VARIABLES

In this section, $X_1, \ldots, X_n, \ldots$ is a sequence of mutually independent random variables. For more extensive discussions, see Gnedenko (1968, pp. 94-121, 302-317), Gnedenko and Kolmogorov (1968, pp. 125-132), and Cramér (1970, pp. 53-69). We shall present in sequence some early historical forms of the Central Limit Theorem (see Chapter 1) as well as more commonly occurring special cases of it.

[*6.1.1*] *The de Moivre-Laplace Limit Theorem*. Let X_n be a binomially distributed rv such that $\Pr(X_n = x) = \binom{n}{x}p^x(1 - p)^{n-x}$; $x = 0, 1, 2, \ldots, n$; $0 < p < 1$. Let $Y_n = (X_n - np)/\sqrt{np(1 - p)}$, so that Y_n is standardized, with mean zero and variance one, and let Y_n have cdf $G_n(y)$. Then, if $a < b$,

$$\lim_{n\to\infty} \Pr(a \le Y_n \le b) = \Phi(b) - \Phi(a)$$

$$\lim_{n\to\infty} \Pr(a \le Y_n < b) = \Phi(b) - \Phi(a)$$

and

$$\lim_{n\to\infty}\left[\frac{\Pr(a \le Y_n \le b)}{\Phi\{b + \frac{1}{2}(np(1 - p))^{-1/2}\} - \Phi\{a - \frac{1}{2}(np(1 - p))^{-1/2}\}}\right] = 1$$

This last form is more accurate if the denominator is used as an approximation to $\Pr(a \le Y_n \le b)$ (Feller, 1968, pp. 182-186; Gnedenko, 1968, p. 104; Woodroofe, 1975, pp. 97-106).

[*6.1.2*] *The de Moivre-Laplace Local Limit Theorem*. With the conditions and notation of [6.1.1] (Gnedenko, 1968, p. 98; Woodroofe, 1975, pp. 98, 105),

$$\lim_{n\to\infty}\left[\frac{\sqrt{np(1 - p)}\,\Pr(X_n = m)}{\phi\{(m - np)/\sqrt{np(1 - p)}\}}\right] = 1$$

[*6.1.3*] *Tchebyshev's Central Limit Theorem* (Maistrov, 1974, pp. 202-203). Let $X_1, X_2, \ldots$ be mutually independent rvs with zero means and finite moments of all orders, and let $\sigma_i^2 = \mathrm{Var}(X_i)$. Further,

(i) $\lim_{n\to\infty}[(\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2)/n]$ exists and is finite.

(ii) $|E(X_i^r)| < A_r < \infty, \quad r = 2, 3, \ldots$.

Then

$$\lim_{n\to\infty} \Pr\{a \le (X_1 + \cdots + X_n)/(\sigma_1^2 + \cdots + \sigma_n^2)^{1/2} \le b\} = \Phi(b) - \Phi(a)$$

Condition (i) is necessary and was added by Kolmogorov.

[*6.1.4*] *The Markov-Tchebyshev Central Limit Theorem* (Modern Version). Let $X_1, X_2, \ldots$ be a sequence of mutually independent rvs such that $E(x_i) = \xi_i$ and $\mathrm{Var}(X_i) = \sigma_i^2$; $i = 1, 2, \ldots$; if also $S_n = X_1 + \cdots + X_n$ and (Maistrov, 1974, pp. 210-211)

$$\lim_{n\to\infty}\left[\sum_{k=1}^{n} E(|X_k - \xi_k|^r)/(\sigma_1^2 + \cdots + \sigma_n^2)\right] = 0$$

then

$$\lim_{n\to\infty} \Pr\left[\{S_n - (\xi_1 + \cdots + \xi_n)\}/(\sigma_1^2 + \cdots + \sigma_n^2)^{1/2} < x\right] = \Phi(x)$$

[*6.1.5*] *The Central Limit Theorem*--iid case (Lêvy, 1925, p. 233; Lindeberg, 1922, pp. 211-225; Cramér, 1970, pp. 53-55; Woodroofe, 1975, p. 251; Feller, 1971, p. 259). Let $X_1, X_2, \ldots$ be iid random variables with common mean μ and common finite variance σ^2, and let $\bar{G}_n(x)$ be the cdf of the rv $(X_1 + \cdots + X_n - n\mu)/(\sigma\sqrt{n})$. Then

$$\lim_{n\to\infty} \{\bar{G}_n(x)\} = \Phi(x)$$

uniformly in x.

[*6.1.6*] *Lyapunov's Theorem* (Gnedenko, 1968, p. 310; Maistrov, 1974, pp. 222-223). Let $X_1, X_2, \ldots$ be mutually independent rvs, such that $E(X_i) = \mu_i$ and $\mathrm{Var}(X_i) = \sigma_i^2 < \infty$; $i = 1, 2, \ldots$. Let $S_n = X_1 + \cdots + X_n$, $m_n = \mu_1 + \cdots + \mu_n$, and $B_n^2 = \sigma_1^2 + \cdots + \sigma_n^2$. If there exists a positive number δ such that

$$\lim_{n\to\infty}\left[B_n^{-(2+\delta)} \sum_{k=1}^{n} E(|X_k - \mu_k|^{2+\delta})\right] = 0$$

then

$$\lim_{n\to\infty} \bar{G}_n(x) = \Phi(x)$$

uniformly in x, where $\bar{G}_n(\cdot)$ is the cdf of $(S_n - m_n)B_n$ (Lyapunov, 1901, pp. 1-24).

[*6.1.7*] <u>*Cramér's Central Limit Theorem*</u> (Cramér, 1970, pp. 57-58). Let $X_1, X_2, \ldots$ be a sequence of mutually independent rvs with means $\mu_1, \mu_2, \ldots$ and finite variances $\sigma_1^2, \sigma_2^2, \ldots$. With the notation of [6.1.6], suppose further that $B_n \to \infty$ and $\sigma_n/B_n \to 0$ as $n \to \infty$ (or, equivalently, that $\sigma_\nu/B_n \to 0$ as $n \to \infty$, uniformly for $\nu = 1, 2, \ldots, n$). Then

$$\lim_{n\to\infty} \bar{G}_n(x) = \Phi(x), \qquad \text{for all } x$$

The condition(s) given for this result mean that the total s.d. of S_n tends to ∞ while each component σ_n contributes only a small fraction of the total s.d.

[*6.1.8*] Let $X_1, X_2, \ldots$ be a sequence of mutually independent rvs with S_n, B_n, m_n, and $\bar{G}_n(\cdot)$ defined as in [6.1.6], and let G_k be the cdf of X_k; $k = 1, 2, \ldots$.

(i) Given $\varepsilon > 0$, if

$$\lim_{n\to\infty}\left[B_n^{-2} \sum_{k=1}^{n} \int_{|x-\mu_k|\geq\varepsilon B_n} (x - \mu_k)^2 dG_k(x)\right] = 0 \qquad (*)$$

then (Lindeberg, 1922, pp. 211-225)

$$\lim_{n\to\infty} \bar{G}_n(x) = \Phi(x) \qquad \text{for all } x$$

The condition (*) is known as the *Lindeberg-Feller condition*, and it is not only sufficient, but in a certain sense necessary for the central limit property to hold (Woodroofe, 1975, pp. 255-257; Feller, 1971, pp. 518-521).

(ii) (Feller, 1971, p. 520). Suppose that the conditions

$$B_n \to \infty \qquad \sigma_n/B_n \to 0 \qquad (**)$$

hold, as $n \to \infty$. Then the condition (*) is necessary for $\bar{G}_n(x)$ to converge to $\Phi(x)$.

See also Feller (1935, pp. 521-559), Ash (1972, pp. 336-342), and Cramér (1970, pp. 57-60), who points out that the conditions (**) are jointly equivalent to the condition that $\sigma_\nu/B_n \to 0$ as $n \to \infty$ uniformly for $\nu = 1, 2, \ldots, n$. Cramér (1970, p. 58) also states that the Lindeberg-Feller condition (*) implies the condition (**).

Remarks. (a) The condition (*) implies that $\bar{G}_n(x) \to \Phi(x)$ as above in several cases (Ash, 1972, pp. 336-338):

(i) $\Pr(|X_k| \le A < \infty) = 1$ and $B_n \to \infty$; $k = 1, 2, \ldots$. This is the "uniformly bounded independent rvs" case.

(ii) The "iid" case of [6.1.5].

(iii) The "independent Bernoulli rvs" case, leading to the de Moivre-Laplace theorem of [6.1.1].

(iv) The Lyapunov condition of [6.1.6], leading to Lyapunov's theorem.

(b) Ash (1972, pp. 341-342) states that if, for all $\varepsilon > 0$,

$$\Pr\{|X_k - \mu_k|/B_n \ge \varepsilon\} \to 0$$

as $n \to \infty$, uniformly for $k = 1, 2, \ldots, n$, then the Lindeberg-Feller condition (*) is necessary and sufficient for the central limit property to hold.

(c) While central limit theorems are stated in terms of standardized partial sums such as $(S_n - m_n)/B_n$, applications in practice often refer to S_n being approximately normal for large n, with mean m_n and variance B_n^2. This is true in the following sense. Let Y_n be a rv with a $N(m_n, B_n^2)$ distribution. Then (Ash, 1972, pp. 356-357)

$$|\Pr(S_n \le y) - \Pr(Y_n \le y)| = \left|\bar{G}_n\left(\frac{y - m_n}{B_n}\right) - \Phi\left(\frac{y - m_n}{B_n}\right)\right| \to 0$$

$$\text{as } n \to \infty$$

for all y, because $\bar{G}_n \to \Phi$ uniformly over the real line.

[*6.1.9*] *A Conditioned Central Limit Theorem* (Rényi, 1958, pp. 215-228). Let $X_1, X_2, \ldots$ be an iid sequence of rvs with mean zero and variance one, and let B be an event (in the σ-algebra of the defining probability space) such that Pr(B) > 0. Let $T_n = \Sigma_{i=1}^n X_i/\sqrt{n}$ (n = 1, 2, ...); then

$$\lim_{n\to\infty} \Pr(T_n < x|B) = \Phi(x) \qquad \text{for all } x$$

See also [6.3.9], and Landers and Rogge (1977, p. 595).

6.2 FURTHER LIMIT THEOREMS

In this section we give limit theorems for densities and results for sequences of rvs which may not be independent.

[*6.2.1*] We shall require here, and in Section 6.3 later, the notion of a *lattice distribution*, that is, for which every realization of some rv X has the form a + kh, with k an integer and a and h fixed numbers. The number h is a *span*, and is a *maximal span* of X if it is the largest possible choice of span.

If the maximal span is unity, we say that X is a *unit lattice rv*. Most of the well-known discrete distributions are unit lattice distributions with a = 0 and h = 1, such as the Bernoulli, binomial, Poisson, hypergeometric, geometric, and negative binomial distributions. We shall state all relevant results for this class separately.

[*6.2.2*] Suppose that $X_1, X_2, \ldots$ is an iid sequence of rvs with a lattice distribution having a span h, common mean μ, and common finite variance σ^2. Let $S_n = X_1 + \cdots + X_n$ and define

$$P_n(k) = \Pr(S_n = an + kh), \qquad k = 0, \pm 1, \pm 2, \ldots$$

Then

$$(\sqrt{n}\sigma/h)P_n(k) - \phi\{(an + kh - n\mu)/\sqrt{n}\sigma\} \to 0$$

as $n \to \infty$, uniformly for all k if and only if the span h is maximal (Gnedenko, 1968, pp. 313-316; Feller, 1971, pp. 517-518).

If the sequence has a unit lattice distribution (see [6.2.1]), the result takes the form

$$\sqrt{n}\sigma \Pr(S_n = k) - \phi\{(k - n\mu)/\sqrt{n}\sigma\} \to 0$$

[*6.2.3*] *Limit Theorems for Densities*. (a) (Gnedenko, 1968, p. 317). Let X_1, X_2, ... be an iid sequence of absolutely continuous rvs with common mean μ and finite variance σ^2. For all $n \ge n_0$, let the standardized sum $\{(X_1 + \cdots + X_n) - n\mu\}/(\sqrt{n}\sigma)$ have pdf $u_n(x)$. Then

$$u_n(x) \to \phi(x)$$

as $n \to \infty$ uniformly in x ($|x| < \infty$) if and only if there exists a number m such that $u_m(x)$ is bounded.

This result also holds if the boundedness condition for $u_m(x)$ is replaced by the condition that the common pdf g(x) of X_1, X_2, ... is bounded (Rényi, 1970, p. 449).

(b) (Gnedenko and Kolmogorov, 1968, p. 224). Let X_1, X_2, ... be an iid sequence of absolutely continuous rvs with common mean zero, finite variance σ^2, and pdf g(x). If for some $m \ge 1$ the pdf $g_m(x)$ of the sum $X_1 + \cdots + X_m$ satisfies the relation

$$\int_{-\infty}^{\infty} \{g_m(x)\}^r \, dx < \infty$$

whenever $1 < r \le 2$, then

$$\sigma\sqrt{n}\; g_n(\sigma\sqrt{n}x) \to \phi(x)$$

uniformly in x; $|x| < \infty$.

See also Feller (1971, p. 516), who derives the same result when a sufficient condition is that the characteristic function of X_k (k = 1, 2, ...) be integrable.

[6.2.4] *Limit Theorem for Dependent Variables*. A sequence X_1, X_2, ... of rvs is called *m-dependent* if and only if $\{X_b, X_{b+1}, \ldots, X_{b+s}\}$ and $\{X_{a-r}, X_{a-r+1}, \ldots, X_a\}$ are independent sets of variables if $b - a > m$ (Serfling, 1968, p. 1162).

Theorem 1. X_1, X_2, ... is a sequence of m-dependent, uniformly bounded rvs and $S_n = X_1 + \cdots + X_n$, with standard deviation B_n. Then, if $B_n/n^{1/3} \to \infty$ as $n \to \infty$, $\bar{G}_n(x) \to \Phi(x)$ for all x as $n \to \infty$, where $\bar{G}_n$ is the cdf of $\{S_n - E(S_n)\}/B_n$ (Chung, 1974, p. 214).

Theorem 2. Let $T_a = \Sigma_{i=a+1}^{a+n} X_i/\sqrt{n}$, where X_1, X_2, ... is an m-dependent sequence of rvs such that $E(X_i) = 0$, $E(|X_i|^{2+\delta}) \le M < \infty$ for some $\delta > 0$, and $E(T_a^2) \to A^2 > 0$ uniformly in a as $n \to \infty$. Then the limiting distribution of $(nA^2)^{-1/2} \Sigma_{i=1}^n X_i$ is N(0,1) (Serfling, 1968, pp. 1158-1159; Hoeffding and Robbins, 1948, pp. 773-780).

Serfling (1968) and Brown (1971, pp. 59-66) give several central limit theorems under various dependency conditions, including some for stationary sequences, bounded sequences, and martingales, in addition to references to work by a number of other writers.

The sequence X_1, X_2, ... of rvs is called *stationary* if, for all r, the joint distribution of X_i, X_{i+1}, ..., X_{i+r} does not depend on i.

Theorem 3. (Fraser, 1957, p. 219). If X_1, X_2, ... is a stationary m-dependent sequence of rvs such that $E(X_1) = \mu$ and $E\{|X_1|^3\}$ exists, then $n^{-1/2} \Sigma_{j=1}^n X_j$ is asymptotically normal, with mean $\sqrt{n}\mu$ and variance

$$\text{Var}(X_1) + 2[\text{Cov}(X_1,X_2) + \cdots + \text{Cov}(X_1,X_{m+1})]$$

[6.2.5] *Random Number of Terms*. Let X_1, X_2, ... be a sequence of iid random variables with mean zero and variance one, and let V(1), V(2), ... be a sequence of positive integer-valued rvs independent of X_1, X_2, ... and such that $V(n)/n \to c$ in probability, for some positive constant c. Then, for all x,

$$\Pr(S_{V(n)}/\sqrt{V(n)} \le x) \to \Phi(x) \qquad \text{where } S_k = X_1 + \cdots + X_k$$

as $n \to \infty$ (Chung, 1974, p. 216; Feller, 1971, p. 265; Anscombe, 1952, p. 601).

[*6.2.6*] Let $X_1, X_2, \ldots$ be an iid sequence of rvs with mean zero and variance one, and let $Y_n = \max(S_1, S_2, \ldots, S_n)$, where $S_n = X_1 + \cdots + X_n$. Then $Y_n/\sqrt{n}$ converges in distribution to the half-normal distribution of [2.6.1.3]; that is, for all $y \geq 0$, as $n \to \infty$ (Chung, 1974, p. 222),

$$\Pr(Y_n/\sqrt{n} \leq y) \to 2\Phi(y) - 1$$

[*6.2.7*] The rvs $X_{n,k(n)}$ for $k = 1, 2, \ldots, h(n)$ are *infinitesimal* if

$$\sup_{1 \leq k \leq h(n)} \Pr\{|X_{n,k}| \geq \varepsilon\} \to 0$$

as $n \to \infty$ for every positive ε, where $h(n)$ is a positive-integer-valued function of the positive integer n.

Suppose further that $X_{n,1}, X_{n,2}, \ldots, X_{n,h(n)}$ are mutually independent for each value of n, and that

$$Y_n = X_{n,1} + \cdots + X_{n,h(n)}$$

Then, if $X_{n,k}$ has cdf $G_{n,k}$, and if the sequence of rvs $\{Y_n\}$ $(n = 1, 2, \ldots)$ converges to a limit as $n \to \infty$,

$$\sum_{k=1}^{h(n)} \int_{|x| \geq \varepsilon} dG_{n,k}(x) \to 0$$

as $n \to \infty$ for every $\varepsilon > 0$ if and only if the limiting distribution is normal (Gnedenko and Kolmogorov, 1968, pp. 95, 126).

[*6.2.8*] A sequence $\{G_n\}$ of cumulative distribution functions converges to $\Phi(\cdot)$ if and only if the sequence $\{\psi_n(t)\}$ of their characteristic functions converges to $\exp(-t^2/2)$, where if X_n is a rv with cdf G_n, $\psi_n(t) = E(\exp(itX_n))$. The convergence of $\psi_n(t)$ to $\exp(-t^2/2)$ is uniform for every finite interval (Feller, 1971, p. 508).

[*6.2.9*] Let T_1, T_2, ... be a sequence of statistics such that for θ lying in an interval, $\sqrt{n}(T_n - \theta)$ has asymptotically (as $n \to \infty$) a $N(0,\sigma^2(\theta))$ distribution. Let $g(\cdot)$ be a single-valued function having first derivative g'.

(a) If $g'(\theta) \neq 0$, then $\sqrt{n}[g(T_n) - g(\theta)]$ has asymptotically a normal distribution with mean zero and variance $\{g'(\theta)\sigma(\theta)\}^2$.

(b) If, in addition, $g'(\cdot)$ is continuous at θ, then $\sqrt{n}[g(T_n) - g(\theta)]/g'(T_n)$ has asymptotically a $N(0,\sigma^2(\theta))$ distribution.

(c) If, further, $\sigma(\cdot)$ is continuous at θ, then $\sqrt{n}[g(T_n) - g(\theta)]/[g'(T_n)\sigma(\theta)]$ has asymptotically a $N(0,1)$ distribution (Rao, 1973, pp. 385-386).

6.2.10 *Sample Fractiles*. Let X be a rv, and let x_p be any value such that $\Pr(X \le x_p) \ge p$ and $\Pr(X \ge x_p) \ge 1 - p$. (This allows for discrete rvs, where the fractiles may not be unique.) The pth *fractile statistic* in a sample of n observations from the distribution of X is a value $\hat{x}_p$ such that the number of observations less than or equal to $\hat{x}_p$ is at least [np] and the number greater than or equal to $\hat{x}_p$ is at least [n(1 - p)], where [x] denotes the largest integer less than or equal to x.

Let X be continuous with pdf f(x) which is continuous in x, and suppose that x_p is unique with $f(x_p)$ strictly positive. Then $\sqrt{n}(\hat{x}_p - x_p)$ has asymptotically a normal distribution with mean zero and variance $p(1 - p)/\{f(x_p)\}^2$ (Rao, 1973, pp. 422-423). Further, if $0 < p(1) < p(2) < \cdots < p(k) < 1$, with corresponding unique quantiles $x_{p(1)} < \cdots < x_{p(k)}$ such that $f(x_{p(1)}) > 0, \ldots, f(x_{p(k)}) > 0$, then the asymptotic joint distribution of $\sqrt{n}(\hat{x}_{p(1)} - x_{p(1)}), \ldots, \sqrt{n}(\hat{x}_{p(k)} - x_{p(k)})\}$ is multivariate normal with mean vector $(0,\ldots,0)$, the asymptotic covariance between $\sqrt{n}\,\hat{x}_{p(i)}$ and $\sqrt{n}\,\hat{x}_{p(j)}$ is $p(i)\{1 - p(j)\}/[f(x_{p(i)})f(x_{p(j)})]$, $1 \le i < j \le k$ (Mosteller, 1946, pp. 383-384). See also Wilks (1962, pp. 271-274).

[*6.2.11*] If X_1, X_2, ... are equicorrelated standard rvs, i.e.,

$$EX_i = 0 \qquad EX_i^2 = 1 \qquad EX_iX_{i+N} = \rho, \qquad i = 1, 2, \ldots;\ N = 1, 2, \ldots$$

then the limiting distribution of $X_{(n;n)} - \sqrt{2(1 - \rho) \log n}$ is normal when $\rho \geq 0$, with mean zero and variance ρ (David, 1970, p. 215), where $X_{(n;n)} = \max(X_1, X_2, \ldots, X_n)$.

[*6.2.12*] Let $X_{(1;n)} \leq X_{(2;n)} \leq \cdots \leq X_{(n;n)}$ be the order statistics in a random sample $X_1, \ldots, X_n$ from some distribution (see Chapter 8). The *trimmed mean* T_n is defined in [8.8.5] as

$$T_n = \{[\beta n] - [\alpha n]\}^{-1} \sum_{i=[\alpha n]+1}^{[\beta n]} X_{(i;n)}, \qquad 0 < \alpha < \beta < 1$$

where proportions α and $1 - \beta$ of the ordered sample are trimmed off the lower and upper ends, respectively, and $[x]$ is the largest integer less than or equal to x.

Let G(x) be the cdf of the underlying population. If G is continuous and strictly increasing in $\{x: 0 < G(x) < 1\}$, then the asymptotic distribution of T_n is normal, with mean and variance as given in [8.8.5], where μ_t and σ_t^2 are the mean and variance of the distribution $G(\cdot)$ truncated below and above at a and b, respectively.

For other cases, $G(\cdot)$ is truncated to include $a \leq x < b$ only where $a = \sup\{x: G(x) \geq \alpha\}$, $b = \inf\{x: G(x) \geq \beta\}$; μ_t and σ_t^2 are the mean and variance of this truncated distribution. Let

$$A = a - \inf\{x: G(x) \geq \alpha\} \qquad B = \sup\{x: G(x) \leq \beta\} - b$$

Then the limiting distribution of $\sqrt{n}(T_n - \mu_t)$ is that of a rv Z, where

$$Z = (\beta - \alpha)^{-1}\{Y_1 + (b - \mu_t)Y_2 + (a - \mu_t)Y_3 + B \max(0, Y_2) - A \max(0, Y_3)\}$$

$$E(Z) = [B\sqrt{\beta(1 - \beta)} - A\sqrt{\alpha(1 - \alpha)}]/\{\sqrt{2\pi}(\beta - \alpha)\}$$

where $Y_1 \sim N(0, (\beta - \alpha)\sigma_t^2)$, (Y_2, Y_3) and Y_1 are independent, and (Y_2, Y_3) is bivariate normal with mean vector zero and variance-covariance matrix (Stigler, 1973, p. 473)

$$\begin{pmatrix} \beta(1-\beta) & -\alpha(1-\beta) \\ -\alpha(1-\beta) & \alpha(1-\alpha) \end{pmatrix}$$

Notice that, for a nonnormal limiting distribution to occur, trimming must take place when α or β corresponds to a nonunique percentile of the underlying distribution $G(\cdot)$. Stigler (1973, p. 477) suggests a more smoothly trimmed mean which is asymptotically normal for any distribution $G(\cdot)$; see [6.2.13.3].

[*6.2.13*] Using the notation of [6.2.12], let $X_1, \ldots, X_n$ be an iid sample from a distribution with cdf G, and let

$$S_n = \sum_{i=1}^{n} c_{in} X_{(i;n)}$$

a linear function of the order statistics, with "weights" c_{in}. Considerable research, particularly in the mid-1960s to mid-1970s, addressed the question of what conditions lead to the asymptotic normality of S_n. For a good readable discussion and some references, see Stigler (1969, pp. 770-772; 1974, pp. 676-677, 687-688, 690-692). There is generally a trade-off between the amount of weight to be permitted to the extreme order statistics and smoothness restrictions on $G(\cdot)$, particularly in the tails and on the density of G in the support of G.

[*6.2.13.1*] Let

$$S_n = \sum_{i=1}^{n} n^{-1} J\left(\frac{i}{n+1}\right) X_{(i;n)}$$

where J is a suitable weight function, and let

$$\sigma^2(J,G) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} J(G(x))J(G(y))\{G(\min(x,y)) - G(x)G(y)\}\, dy\, dx$$

The following results are due to Stigler (1974, pp. 681-687).

(i) If $E(X_i^2) < \infty$, if J is bounded and continuous (almost everywhere w.r.t. G^{-1}), and if $\sigma^2(J,G) > 0$, then $\{S_n - E(S_n)\}/\sigma(S_n)$ is asymptotically $N(0,1)$.

(ii) (A version in which the extremes are trimmed.) Suppose that, for some $\varepsilon > 0$, $x^{\varepsilon}[1 - G(x) + G(-x)] \to 0$ as $x \to \infty$, J is bounded and continuous as in (i), that $J(u) = 0$ if $0 < u < \alpha$ and $1 - \alpha < u < 1$, and that $\sigma^2(J,G) > 0$. Then the conclusion in (i) holds.

We state two applications of these results.

[*6.2.13.2*] Let

$$D_n = [n(n - 1)]^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} |X_{(i;n)} - X_{(j;n)}|$$

which is called *Gini's mean difference*. Then (Stigler, 1974, pp. 690-691) if $J(u) = u - 1/2$, $D_n = 4(n + 1)S_n/(n - 1)$, and result (i) of [6.2.13] above applies when $E(X_i^2) < \infty$.

[*6.2.13.3*] Let

$$S_n = \sum_{i=1}^{n} n^{-1} J\{i/(n + 1)\} X_{(i;n)}$$

where

$$J(u) = \begin{cases} (u - \alpha)(\tfrac{1}{2} - \alpha)^{-1}, & \alpha \le u \le \tfrac{1}{2} \\ (1 - \alpha - u)(\tfrac{1}{2} - \alpha)^{-1}, & \tfrac{1}{2} \le u \le 1 - \alpha \\ 0, & \text{otherwise} \end{cases}$$

This is a form of smoothly trimmed mean (see [6.2.12] and Stigler, 1973, p. 477) which is asymptotically normal, whatever the form of the underlying distribution G may be.

6.2.14 *A Central Limit Theorem for U-Statistics* (Fraser, 1957, p. 137). We give two simplified versions which have practical uses.

Theorem 1 (Lehmann, 1975, pp. 366-368). Let $X_1, \ldots, X_n$ be a random sample from a common distribution, and let $h(x,y)$ be a symmetric function, i.e., $h(x,y) = h(y,x)$.

Let

$$E\,h(X_i,X_j) = \theta, \qquad i \neq j$$
$$h_1(x) = E\,h(X_i,x) = E\,h(x,X_i)$$
$$\sigma_1^2 = \mathrm{Var}\{h_1(X_i)\}$$
$$U = \Sigma_{i<j}\{h(X_i,X_j)\}, \qquad \text{so that } E(U) = \theta$$

Then $\sqrt{n}(U - \theta)$ is asymptotically $N(0,4\sigma_1^2)$ as $n \to \infty$. A more general symmetric function than $h(x,y)$, involving subsets of r variables ($3 \leq r < n$), leads to a main theorem of Hoeffding (1948, p. 305); see Fraser (1957, p. 225).

Theorem 2 A Two-sample Version (Lehmann, 1975, pp. 362-365). Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be iid samples, each sample from possibly different distributions, and independent of one another. Let $f(x,y)$ be a function of two variables, let $m \leq n$, and suppose that $m/n \to \lambda$ as m and $n \to \infty$, where λ may be zero. Further, let

$$E[f(X_i,Y_j)] = \theta \qquad f^*(X_i,Y_j) = f(X_i,Y_j) - \theta$$
$$f_{10}(x) = E[f^*(x,Y_j)] \qquad f_{01}(y) = E[f^*(X_i,y)]$$
$$\sigma_{10}^2 = \mathrm{Var}\ f_{10}(X_i) \qquad \sigma_{01}^2 = \mathrm{Var}\ f_{01}(Y_j)$$
$$U = (mn)^{-1}\ \Sigma_{i=1}^{m}\Sigma_{j=1}^{n}\ f(X_i,Y_j), \qquad \text{so that } E(U) = \theta$$

Then $\sqrt{m}(U - \theta)$ is asymptotically $N(0, \sigma_{10}^2 + \sigma_{01}^2)$ as m and $n \to \infty$.

For an extension to more general functions, involving subsets of r and s variables ($3 \leq r < m$, $3 \leq s < n$), and leading to more general U-statistics, see Fraser (1957, pp. 229-230).

The theorems stated here have applications to rank statistics which are useful in nonparametric one- and two-sample hypothesis-testing problems. For example, the asymptotic normality of the Wilcoxon one- and two-sample rank-sum statistics and the Spearman rank correlation coefficient can be shown to be asymptotically normal; see Lehmann (1975, pp. 365-366, 368-371); Fraser (1957, pp. 231, 234-235). A number of test statistics of this kind have asymptotic normality, as described in Hollander and Wolfe (1973).

[*6.2.15*] The following result is a central limit theorem for a finite population. Let N = population size, M = number of "successes" in the population, and n = number drawn at random from the population without replacement. Let X = number of "successes" in the sample, so that X has a hypergeometric distribution with

$$\Pr(X = k) = g(k;N,M,n) = \frac{\binom{M}{k}\binom{N - M}{n - k}}{\binom{M}{n}}$$

Suppose further that $1 \le M \le (1/2)N$, and $1 \le n \le (1/2)N$. Then, if $p = M/N$ and $\lambda = n/N$,

$$\lim_{Np\lambda \to +\infty} \Sigma' g(k;N,M,n) = \Phi(x)$$

where Σ' denotes summation over the set $k \le np + x\sqrt{\{np(1 - p)(1 - \lambda)\}}$ (Rényi, 1970, p. 466).

[*6.2.16*] The following is a central limit theorem for order statistics. Let $X_1, X_2, \ldots$ be an iid sequence of rvs with an absolutely continuous cdf F(x) and a density function f(x) continuous and positive if $a \le x < b$. Let $X_{(1;n)} \le X_{(2;n)} \le \cdots \le X_{(n;n)}$ be the order statistics based on $X_1, X_2, \ldots, X_n$, and q a number such that $0 < F(a) < q < F(b) < 1$. If $\{k_n\}$ is a sequence of integers such that

$$\lim_{n \to \infty} \{\sqrt{n}\,|(k_n/n) - q|\} = 0$$

and if $q = F(Q)$, $D = \sqrt{q(1 - q)}/f(Q)$, then (Rényi, 1970, p. 490)

$$\lim_{n \to \infty} \Pr\left\{\frac{X_{(k_n;n)} - Q}{D/\sqrt{n}} < x\right\} = \Phi(x)$$

6.3 RAPIDITY OF CONVERGENCE TO NORMALITY

None of the results in Section 6.1 or Section 6.2 gives us any information on how large the value of n needs to be for the cdf $\tilde{G}_n(\cdot)$ of the standardized sum of n rvs to approach within a specified amount δ, say, of the standard normal cdf Φ. The first to derive an upper bound to $|\bar{G}_n(x) - \Phi(x)|$ was Tchebyshev; see Adams (1974, p. 75). We give below the bound obtained by Lyapunov for historical interest, followed by the important Berry-Esseen theorem and its ramifications; a crucial assumption is the existence of finite absolute third moments, although William Feller proved in one of his last published papers (see [6.3.6]) that a form of the main result holds without the third moment assumption. See also [6.4.2, 4, 7].

[6.3.1] Let $X_1, X_2, \ldots$ be an iid sequence of rvs with common mean μ, finite variance σ^2, and finite absolute third moment $\nu_3 = E(|X_i - \mu|^3)$. Let $\bar{G}_n(\cdot)$ be the cdf of $(X_1 + \cdots + X_n - n\mu)/(\sqrt{n}\sigma)$. Then there is a positive constant γ such that, for all real x (Lyapunov, 1901, pp. 1-24; Gnedenko and Kolmogorov, 1968, p. 201),

$$|\bar{G}_n(x) - \Phi(x)| < \gamma(\nu_3/\sigma^3) \log n/\sqrt{n}, \qquad n = 1, 2, \ldots$$

[6.3.2] *The Berry-Esseen Theorem*. The following improves upon Lyapunov's result of [6.3.1] by removing the factor log n from the upper bound for $|\bar{G}_n(x) - \Phi(x)|$.

(a) Let $X_1, X_2, \ldots$ be an iid sequence of rvs under the conditions of [6.3.1]. Then, for some positive constant C, and for all x (Feller, 1971, p. 542; Gnedenko and Kolmogorov, 1968, p. 201),

$$\sqrt{n}|\bar{G}_n(x) - \Phi(x)| \le C\nu_3/\sigma^3, \qquad n = 1, 2, \ldots$$

(b) Let $X_1, X_2, \ldots$ be a sequence of mutually independent rvs, such that $E(X_i) = \mu_i$, $\mathrm{Var}(X_i) = \sigma_i^2$, and $E(|X_i - \mu_i|^3) = \beta_i < \infty$; $i = 1, 2, \ldots$. Further, let $\rho_n = \Sigma_{i=1}^n \beta_i/(\Sigma_{i=1}^n \sigma_i^2)^{3/2}$, and let $\bar{G}_n(\cdot)$ be the cdf of the standardized sum $\{(X_1 - \mu_1) + \cdots + (X_n - \mu_n)\}/(\sigma_1^2 + \cdots + \sigma_n^2)^{1/2}$. Then, for all real x and for $n = 1, 2, \ldots$, there is a positive constant C such that (Berry, 1941, pp. 122-136; Esseen, 1942, pp. 1-19; Cramér, 1970, p. 78)

$$|\bar{G}_n(x) - \Phi(x)| \le C\rho_n$$

(c) Under the conditions of (a) or (b) (Esseen, 1956, pp. 160-170; Beek, 1972, pp. 188, 196),

$$(\sqrt{10} + 3)/(6\sqrt{2\pi}) = 0.4097, 32 \le C < 0.7975$$

The infimum value of C is attained for a particular Bernoulli distribution, and hence it cannot be improved upon (Bhattacharya and Rao, 1976, p. 240). Values of C may be derived for particular families or kinds of distributions, but the upper bound above is the most recent result in a series of progressively sharper bounds obtained by a number of workers using the assumptions in (a) or (b) only, the earliest of which was Esseen's (1956, pp. 160-170) bound of 7.5 for C.

Beek's upper bound for C of 0.7975 has not been improved upon for iid sequences of rvs; prior to his paper of 1972, however, it appeared (Zolotarev, 1966, pp. 95-105) that sharper upper bounds for C might be found in this case.

[*6.3.3*] Zahl (1966, pp. 1225-1245) obtained a modified Berry-Esseen result. Using the notation of [6.3.1] and [6.3.2], let

$$\beta_i' = \begin{cases} \beta_i, & \beta_i \ge 3\sigma_i^3/\sqrt{2} \\ \sigma_i^3(0.7804 - 0.1457\beta_i/\sigma_i^3)^{-1}, & \text{otherwise, } i = 1, 2, \ldots \end{cases}$$

$$\rho_n' = \sum_{i=1}^{n} \beta_i' \Big/ \left(\sum_{i=1}^{n} \sigma_i^2\right)^{3/2}$$

Then

$$\sup_x \left|\bar{G}_n(x) - \Phi(x)\right| \le (0.650)\rho_n'$$

This result is sharper than that of Beek (1972, pp. 185, 196) if and only if $\Sigma_{i=1}^n \beta_i'/\Sigma_{i=1}^n \beta_i < 0.7975/0.650$ (Beek, 1972).

[*6.3.4*] Let X_1, X_2, ... be an iid sequence of rvs with the conditions and notation of [6.3.1]. If the rvs of the sequence are symmetrically distributed about μ and their common cdf is continuous at μ, then (Gnedenko and Kolmogorov, 1968), p. 218)

$$\lim_{n\to\infty} \sup_{|x|<\infty} [\sqrt{n}|\bar{G}_n(x) - \Phi(x)|] \le 1/\sqrt{2\pi} = 0.3989,423$$

Equality holds when $\Pr(X_i = -a) = 1/2 = \Pr(X_i = a)$ for some real number a.

[*6.3.5*] Let $X_1, X_2, \ldots$ be an iid sequence of rvs. Under the conditions of [6.3.1] and using the same notation,

$$\lim_{n\to\infty} \inf_{a,b} \sup_{|x|<\infty} \left[\sqrt{n}\left|\bar{G}_n(x) - \Phi\left(\frac{x - a}{b}\right)\right|\right] \le (2\pi)^{-1/2}\nu_3/\sigma^3$$

with equality if $\Pr(X_i = -h) = 1/2 = \Pr(X_i = h)$ for some real h (Rogozin, 1960, pp. 114-117).

[*6.3.6*] The following theorem gives forms of the Berry-Esseen theorem from truncation of the variables, and is due to Feller (1968, pp. 261-263). Notice that it is independent of the method of truncation and, more important, does not require the assumptions in [6.3.1] to [6.3.5] of third moments.

(a) Let $X_1, X_2, \ldots$ be a sequence of mutually independent rvs, and let

$$X_k' = \begin{cases} X_k, & \text{if } -\tau_k < X_k < \tau_k', \quad -\infty \le -\tau_k < 0 < \tau_k' \le \infty \\ 0, & \text{otherwise} \end{cases}$$

Suppose that $E(X_k) = 0$ and $E(X_k^2) = \sigma_k^2 < \infty$; $k = 1, 2, \ldots$. Let

$$\beta_k' = E(X_k - X_k')^2 \qquad \gamma_k = E(|X_k'|^3)$$

$$b_n' = \beta_1' + \cdots + \beta_n'$$

$$c_n = \gamma_1 + \cdots + \gamma_n$$

$$B_n^2 = \sigma_1^2 + \cdots + \sigma_n^2$$

Then if $\bar{G}_n(\cdot)$ is the cdf of the normalized sum of $X_1, X_2, \ldots, X_n$ as in [6.3.2(b)],

$$\sup_x |\bar{G}_n(x) - \Phi(x)| \le 6\{(c_n/B_n^3) + (b_n'/B_n^2)\}, \qquad n = 1, 2, \ldots$$

(b) In addition to conditions (a), suppose that, if $G_k(\cdot)$ is the cdf of X_k,

$$\int_{-t_k}^{\tau_k} x\, dF_k(x) \le 0 \qquad \text{and} \qquad \int_{-\tau_k}^{t_k'} x\, dF_k(x) \ge 0$$

$$-\infty \le -t_k \le -\tau_k, \quad \tau_k' \le t_k' \le \infty$$

Let

$$\beta_k = E(X_k'^2), \quad b_n = \beta_1 + \cdots + \beta_n$$

$$\pi_k = \Pr(X_k \ne X_k')$$

Then

$$\sup_x |\bar{G}_n(x) - \Phi(x)| \le 6\{(c_n/B_n^3) + 1 - (b_n/B_n^2)\} + \pi_1 + \cdots + \pi_n$$

[*6.3.7*] The following two theorems are given by Gnedenko and Kolmogorov (1968, pp. 210-213).

(a) $X_1, X_2, \ldots$ is an iid sequence of nonlattice rvs, having common mean μ, common variance σ^2, and common finite (central) third moment μ_3. Let $\bar{G}_n(\cdot)$ be the cdf of $\{(X_1 + \cdots + X_n) - n\mu\}/(\sigma\sqrt{n})$. Then, if $Q(x) = \mu_3(1 - x^2)/(6\sigma^3)$,

$$\bar{G}_n(x) - \Phi(x) = \phi(x)Q(x)/\sqrt{n} + o(1/\sqrt{n})$$

(b) $X_1, X_2, \ldots$ is an iid sequence of unit lattice rvs (see [6.2.1]), having common mean, variance, and third moment as in (a). With the notation of (a), let $R(x) = [x] - x + 1/2$, where $[x]$ is the largest integer less than or equal to x. Then

$$\bar{G}_n(x) - \Phi(x) = \phi(x)[Q(x) + \sigma^{-1}R(\sigma\sqrt{n}x)]/\sqrt{n} + o(1/\sqrt{n})$$

See Gnedenko and Kolmogorov (1968) for a version of this result for other lattice distributions.

[*6.3.8*] The following results illustrate the possible deviation from normality of distributions having a limited number of their moments equal to the corresponding moments of a N(0,1) rv

(McGillivray and Kaller, 1966, pp. 509-514). Let the equations

$$f_{2n}(x) = \phi(x)(a_0 + a_2x^2 + \cdots + a_{2n}x^{2n})$$
$$f_{2n}(-x) = f_{2n}(x), \qquad n \geq 2$$

define a pdf with moments up to order 2(n - 1) equal to those of $\phi(\cdot)$,

$$\int_{-\infty}^{\infty} x^{2r} f_{2r}(x)\, dx = (2r)!/(2^r r!), \qquad r = 0, 1, 2, \ldots, 2(n - 1)$$

and let $H_k(\cdot)$ be the Hermite polynomial of order k (see [2.1.9]).

(a) Then $f_{2n}(x) = \phi(x)\{1 + a_{2n}H_{2n}(x)\}$, where

$$0 \leq a_{2n} \leq (|\inf_x H_{2n}(x)|)^{-1} = A_{2n}, \qquad \text{say, } n = 2, 3, \ldots$$

(b) Further,

$$\sup_x |f_{2n}(x) - \phi(x)| \leq A_{2n}(2n)!/(\sqrt{2\pi}\ 2^n n!), \qquad n = 2, 3, \ldots$$

When n = 2, 3, or 4, the upper bound is 0.20000, 0.05812, or 0.01385, respectively.

(c) If F_{2n} is the cdf corresponding to f_{2n}, then

$$\sup_x |F_{2n}(x) - \Phi(x)| \leq A_{2n} \sup_x\{\phi(x)|H_{2n-1}(x)|\}, \quad n = 2, 3, \ldots$$

When n = 2, 3, or 4, this bound is 0.10, 0.03, or 0.005, respectively. The above pdf f_{2n} has one parameter free to vary, a_{2n}. "Permitting more than one free parameter in the polynomial would only serve to increase the maximum possible derivation" (McGillivray and Keller, 1966, p. 514).

[6.3.9] *A Berry-Esseen Type Bound for the Conditioned Central Limit Theorem of [6.1.10]*. Let $X_1, X_2, \ldots$ be an iid sequence of rvs with mean zero and variance one, where $E(|X_i|^q) < \infty$ for some $q \geq 3$; and let B_k be an event depending on $X_1, \ldots, X_k$ only, $1 \leq k < n$; so B_k is a member of the σ-algebra generated by $X_1, \ldots, X_k$. Then for each r such that $2 \leq r \leq q$, there exists a constant c_r such that whenever $\Pr(B_k) > 0$ (Landers and Rogge, 1977, p. 598),

$$\sup_x |\Pr(\bar{X}_n < x | B_k) - \Phi(x)| \le c_r \sqrt{k/n}/\{\Pr(B_k)\}^{1/r}$$

[*6.3.10.1*] De Wet (1976, pp. 77-96) gives a Berry-Esseen result for the *trimmed mean* (see [6.2.12]) of an iid sample, with a bound $Cn^{-1/2}$ for $\sqrt{n}|\bar{G}_n(x) - \Phi(x)|$, when $\bar{G}_n(\cdot)$ is the cdf of the standardized trimmed mean of a sample of size n, and the parent distribution has a finite variance and finite absolute third moment.

[*6.3.10.2*] Bjerve (1977, p. 365) gives a Berry-Essen bound $Kn^{-1/2}$ for $|\bar{H}_n(x) - \Phi(x)|$, where $\bar{H}_n(\cdot)$ is the cdf of trimmed linear functions $\Sigma_{i=1}^n c_{in} X_{(i;n)}$ of order statistics. This allows fairly general weights on the observations which are not trimmed, but requires a smoothness condition for the cdf.

Helmers (1977, p. 941) obtains a similar bound, where $\bar{H}_n(\cdot)$ is the cdf of a standardized untrimmed linear function $n^{-1}\Sigma_{i=1}^n J(i/(n + 1))X_{(i;n)}$; the weight function $J(\cdot)$ must be bounded and continuous on (0,1) and must satisfy a smoothness condition, and $\bar{H}_n(\cdot)$ must not have too much weight in the tails.

[*6.3.11*] Callaert and Janssen (1978, pp. 417-418) derive a Berry-Esseen bound for *U-statistics* of order two (see [6.2.14]), as follows: let $h(x,y)$ be a symmetric function, so that $h(x,y) = h(y,x)$, and let $X_1, \ldots, X_n$ be a random sample from a common distribution. Suppose further that $E\, h(X_i,X_j) = 0$, $i \ne j$, and that $\sigma_1^2 > 0$, where σ_1^2 is defined as in [6.2.14]. Let

$$U = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} h(X_i, X_j)$$

If $\nu_3 = E|h(X_1,X_2)|^3 < \infty$, then for some positive constant C,

$$\sqrt{n}|\Pr(\sigma_n^{-1} U \le x) - \Phi(x)| \le C\nu_3\sigma_1^{-3}, \qquad n = 1, 2, \ldots$$

where $\mathrm{Var}(U) = \sigma_n^2$. The authors indicate that the result remains valid for symmetric functions $h(\cdot)$ of higher order than two, and also for U-statistics based upon two or more samples (Callaert and Janssen, 1978, p. 420).

6.4 EXPANSIONS

[*6.4.1*] Consider a cumulative distribution function G(x) with mean zero and variance one. We give expansions for G(x) in terms of $\Phi(x)$ and its derivatives, and for the important case where $\bar{G}_n(x)$ is the cdf of a normalized sum of iid rvs, in powers of $1/\sqrt{n}$ and in terms of $\Phi(x)$ and of $\phi(x)$. The genesis for such expansions is to be found in a paper by Tchebyshev (1890; 1962), whose ideas lead to the Edgeworth series (see [6.4.2] below). We also give expansions for quantiles of G in terms of those of Φ, and vice versa, the Cornish-Fisher series discussed in [6.4.12] to [6.4.14].

Notice that in [6.4.2] to [6.4.5], [6.4.8], and [6.4.9] below, the *formal* expansions are given, but the question of their validity (i.e., do they converge, and if so, how rapidly?) is a different matter. Some conditions for convergence appear in [6.4.6], and the discussion in [6.4.7], [6.4.10], and [6.4.11] is also relevant. A more meaningful approach to these expansions is to view them as *approximations* when a limited number of terms are used only, and then to see whether an approximation in any given instance is uniformly good over some interval for x, and (in the case of $\bar{G}_n$) for what values of n.

For more detailed discussion, see Cramér (1946, pp. 221-230), Cramér (1970, pp. 81, 86-88), Gnedenko and Kolmogorov (1968, pp. 190-196, 220-222), Draper and Tierney (1973, pp. 495-524), Johnson and Kotz (1970, pp. 16-19, 33-35), Kendall and Stuart (1977, pp. 168-178), and Wilks (1962, pp. 262-266).

[*6.4.2*] Let $X_1, X_2, \ldots$ be an iid sequence of rvs with absolutely continuous cdf G(x) and let $\bar{G}_n(x)$ be the cdf of the normalized sum $(X_1 + \cdots + X_n - n\mu)/(\sqrt{n}\sigma)$, where μ and σ^2 are the common mean and variance of G; n = 1, 2, ... Suppose further that G has cumulants $\kappa_1 = \mu, \kappa_2, \kappa_3, \ldots$, and let $\lambda_r = \kappa_r/\sigma^r$; r = 1, 2, ...

The formal *Edgeworth expansion* for $\bar{G}_n(x)$ is given by (Edgeworth, 1905, pp. 36-65, 113-141)

$$\bar{G}_n(x) = \Phi(x) - \phi(x)\left[\frac{\lambda_3 H_2(x)}{3!\sqrt{n}} + \left\{\frac{\lambda_4 H_3(x)}{4!} + \frac{10\lambda_3^2 H_5(x)}{6!}\right\}\frac{1}{n} + \left\{\frac{\lambda_5 H_4(x)}{5!} + \frac{35\lambda_3\lambda_4 H_6(x)}{7!} + \frac{280\lambda_3^3 H_8(x)}{9!}\right\}\frac{1}{n^{3/2}} + \cdots + O(n^{-5/2})\right]$$

where $H_r(x)$ is the rth Hermite polynomial, as defined in [2.1.9]; see Cramér (1970, pp. 86-87) and Draper and Tierney (1973, pp. 499, 502-507), where coefficients of terms in $1/n^{r/2}$ are given up to $r = 10$. We can write the above in terms of the central moments of G and with explicit expressions for the polynomials:

$$\bar{G}_n(x) = \Phi(x) - \phi(x)\left[\frac{\mu_3(x^2 - 1)}{\sigma^3 3!\sqrt{n}} + \left\{\left(\frac{\mu_4}{\sigma^4} - 3\right)\frac{x^3 - 3x}{4!} + \frac{10\mu_3^2(x^5 - 10x^3 + 15x)}{\sigma^6 6!}\right\}\frac{1}{n} + \left\{\left(\frac{\mu_5}{\sigma^5} - 10\frac{\mu_3}{\sigma^3}\right)\frac{x^4 - 6x^2 + 3}{5!} + \frac{\mu_3}{\sigma^3}\left(\frac{\mu_4}{\sigma^4} - 3\right)\frac{35(x^6 - 15x^4 + 45x^2 - 15)}{7!} + \frac{280\mu_3^3(x^8 - 28x^6 + 210x^4 - 420x^2 + 105)}{\sigma^9 9!}\right\}\frac{1}{n^{3/2}}\right] + O(n^{-5/2})$$

[*6.4.3*] If, in the expressions in [6.4.2], we put $n = 1$, we obtain the Edgeworth expansion for the cdf of $(X_1 - \mu)/\sigma$.

[*6.4.4*] Let $g_n(x)$ be the pdf of $\bar{G}_n(x)$ as defined in [6.4.2]. The formal Edgeworth expansion of $g_n(x)$ is obtained by differentiating both sides of the expansion for $\bar{G}_n(x)$:

$$g_n(x) = \phi(x)\left[1 + \frac{\lambda_3 H_3(x)}{3!\sqrt{n}} + \left\{\frac{\lambda_4 H_4(x)}{4!} + \frac{10\lambda_3^2 H_6(x)}{6!}\right\}\frac{1}{n} + \left\{\frac{\lambda_5 H_5(x)}{5!} + \frac{35\lambda_3\lambda_4 H_7(x)}{7!} + \frac{280\lambda_3^3 H_9(x)}{9!}\right\}\frac{1}{n^{3/2}}\right] + O(n^{-5/2})$$

or explicitly,

$$g_n(x) = \phi(x)\left[1 + \frac{\mu_3(x^3 - 3x)}{\sigma^3 3!\sqrt{n}} + \left\{\left(\frac{\mu_4}{\sigma^4} - 3\right)\frac{x^4 - 6x^2 + 3}{4!} + \frac{10\mu_3^2(x^6 - 15x^4 + 45x^2 - 15)}{\sigma^6 6!}\right\}\frac{1}{n} + \left\{\left(\frac{\mu_5}{\sigma^5} - 10\frac{\mu_3}{\sigma^3}\right)\frac{x^5 - 10x^3 + 15x}{5!} + \frac{\mu_3}{\sigma^3}\left(\frac{\mu_4}{\sigma^4} - 3\right)\frac{35(x^7 - 21x^5 + 105x^3 - 105x)}{7!} + \frac{280\mu_3^3(x^9 - 36x^7 + 378x^5 - 1260x^3 + 945x)}{\sigma^9 9!}\right\}\frac{1}{n^{3/2}}\right] + O(n^{-5/2})$$

See Cramér (1946, p. 229), Draper and Tierney (1973, pp. 499, 502-503) with coefficients of $1/n^{r/2}$ up to r = 10, and Feller (1971, p. 535).

[*6.4.5*] If we put n = 1 in the expressions in [6.4.4], we obtain the Edgeworth expansion for the pdf of $(X_1 - \mu)/\sigma$. In applications, only terms up to that in $1/n$ (with n = 1 here) are generally used (Johnson and Kotz, 1970, p. 19), although we have given more here. See [6.4.11] for further discussion.

[*6.4.6*] In the Edgeworth expansions of $\bar{G}_n(x)$ in [6.4.2] and of $g_n(x)$ in [6.4.4], the highest moment of $X_1, X_2, \ldots$ appearing in the coefficient of $n^{-r/2}$ is μ_{r+2}. If higher moments than μ_{r+2} do not exist, then in a sense to be stated presently, we can write the expansion up to terms in $n^{-r/2}$, with a remainder equal to $O(n^{-(r+2)/2})$. The following is due to Cramér (1970, pp. 81-82); see also Gnedenko and Kolmogorov (1968, p. 220).

Let $X_1, X_2, \ldots$ be an iid sequence of rvs with finite absolute moments ν_s of the sth order, where $s > 3$, and let the sequence have common characteristic function $\psi(t) = E[\exp(itX_1)]$. If

$$\limsup_{|t|\to\infty} |\psi(t)| < 1$$

then there exist polynomials $Q_1(x)$, $Q_2(x)$, ... such that

$$\bar{G}_n(x) - \Phi(x) = \phi(x)\left\{\frac{Q_1(x)}{n^{1/2}} + \frac{Q_2(x)}{n} + \cdots + \frac{Q_{s-3}(x)}{n^{(s-3)/2}}\right\} + R_{s,n}$$

and

$$|R_{s,n}| < M/n^{(s-2)/2}, \qquad \text{uniformly in } x$$

where M depends on s and the cdf G, but is functionally independent of n and of x.

If the expansion is taken to the term in $Q_{s-2}(x)/n^{(s-2)/2}$, the remainder is $o(1/n^{(s-2)/2})$. The polynomials are derived as in [6.4.2]; see Draper and Tierney (1973, pp. 499-508). The condition on the characteristic function $\psi(t)$ holds for all absolutely continuous distributions (Cramér, 1970, p. 81), but not for lattice distributions. The expression in [6.3.7(b)] is indicative of this.

[*6.4.7*] If for the iid sequence $X_1, X_2, \ldots$, the cdf G is absolutely continuous and $\mu_3 = 0$, $\bar{G}_n(x)$ is approximated by $\Phi(x)$ except for terms of order $1/n$; if, in addition, the kurtosis $\mu_4/\sigma^4 = 3$, $\bar{G}_n(x)$ is approximated by $\Phi(x)$ except for terms of order $1/n^{3/2}$ (Wilks, 1962, pp. 265-266).

[*6.4.8*] Let G(x) be a cumulative distribution function having pdf g(x). The formal *Gram-Charlier expansion of Type A* for g(x) is given by

$$g(x) = \sum_{j=0}^{\infty} c_j H_j(x)\phi(x) = \sum_{j=0}^{\infty} b_j D(j)\{\phi(x)\}$$

where $H_j(x)$ is the Hermite polynomial of order j (see [2.1.9]), $D(j)\{\phi(x)\}$ is the jth derivative of $\phi(x)$ with respect to x, and $\{c_j\}$, $\{b_j\}$ are sequences of constants (Kendall and Stuart, 1977, pp. 168-169; Johnson and Kotz, 1970, pp. 16-17). If the mean is zero, then

$$g(x) = \phi(x)\{1 + \tfrac{1}{2}(\mu_2 - 1)H_2(x) + \mu_3 H_3(x)/6$$
$$+ (\mu_4 - 6\mu_2 + 3)H_4(x)/24 + \cdots\}$$

If additionally the variance of g(x) is one, then

$$g(x) = \phi(x)\{1 + \mu_3 H_3(x)/6 + (\mu_4 - 3)H_4(x)/24 + \cdots\}$$

$$= \phi(x)\{1 + \mu_3(x^3 - 3x)/6 + (\mu_4 - 3)(x^4 - 6x^2 + 3)/24 + \cdots\}$$

Integration gives the Gram-Charlier expansion of G(x); again, $\mu_2 = 1$:

$$G(x) = \Phi(x) - \phi(x)\{\mu_3 H_2(x)/6 + (\mu_4 - 3)H_3(x)/24 + \cdots\}$$

$$= \Phi(x) - \phi(x)\{\mu_3(x^2 - 1)/6 + (\mu_4 - 3)(x^3 - 3x)/24 + \cdots\}$$

(See Cramér, 1946, pp. 222-223; Charlier, 1905, 1-35.)

[*6.4.9*] Cramér (1970, pp. 87-88) gives the Gram-Charlier Type A expansion of the cdf $\bar{G}_n(x)$ of the normalized sum of n iid random variables having a common pdf (see [6.4.2]): using notation from [6.4.2],

$$\bar{G}_n(x) = \Phi(x) - \phi(x)\left[\frac{\lambda_3}{3!\sqrt{n}} H_2(x) + \frac{\lambda_4}{4!n} H_3(x) + \frac{\lambda_5}{5!n^{3/2}} H_4(x) + \frac{1}{6!}\left(\frac{\lambda_6}{n^2} + \frac{10\lambda_3^2}{n}\right)H_5(x) + \frac{1}{7!}\left(\frac{\lambda_7}{n^{5/2}} + \frac{35\lambda_3\lambda_4}{n^{3/2}}\right)H_6(x) + \cdots\right]$$

$$g_n(x) = \phi(x)\left[1 + \frac{\lambda_3}{3!\sqrt{n}} H_3(x) + \frac{\lambda_4}{4!n} H_4(x) + \frac{\lambda_5}{5!n^{3/2}} H_5(x) + \frac{1}{6!}\left(\frac{\lambda_6}{n^2} + \frac{10\lambda_3^2}{n}\right)H_6(x) + \frac{1}{7!}\left(\frac{\lambda_7}{n^{5/2}} + \frac{35\lambda_3\lambda_4}{n^{3/2}}\right)H_7(x) + \cdots\right]$$

In terms of moments of the iid sequence (see Kendall and Stuart, 1977, p. 73), explicit expressions are given by

$$\bar{G}_n(x) = \Phi(x) - \phi(x)\left[\frac{\mu_3}{\sigma^3 3!\sqrt{n}}(x^2 - 1) + \left(\frac{\mu_4}{\sigma^4} - 3\right)\frac{1}{4!n}(x^3 - 3x) + \left(\frac{\mu_5}{\sigma^5} - 10\frac{\mu_3}{\sigma^3}\right)\frac{1}{5!n^{3/2}}(x^4 - 6x^2 + 3) + \left\{\left(\frac{\mu_6}{\sigma^6} - 15\frac{\mu_4}{\sigma^4} - 10\frac{\mu_3^2}{\sigma^6} + 30\right)\frac{1}{n^2} + \frac{10\mu_3^2}{\sigma^6 n}\right\}\frac{1}{6!}(x^5 - 10x^3 + 15x) + \cdots\right]$$

$$g_n(x) = \phi(x)\left[1 + \frac{\mu_3}{\sigma^3 3!\sqrt{n}}(x^3 - 3x) + \left(\frac{\mu_4}{\sigma^4} - 3\right)\frac{1}{4!n}(x^4 - 6x^2 + 3)\right.$$
$$+ \left(\frac{\mu_5}{\sigma^5} - 10\frac{\mu_3}{\sigma^3}\right)\frac{1}{5!n^{3/2}}(x^5 - 10x^3 + 15x)$$
$$+ \left\{\left(\frac{\mu_6}{\sigma^6} - 15\frac{\mu_4}{\sigma^4} - 10\frac{\mu_3^2}{\sigma^6} + 30\right)\frac{1}{n^2} + \frac{10\mu_3^2}{\sigma^6 n}\right\}\frac{1}{6!}$$
$$\left. \times (x^6 - 15x^4 + 45x^2 - 15) + \cdots\right]$$

[*6.4.10*] We have given the expansions in [6.4.9] to enough terms to illustrate how they differ from those in [6.4.2] and [6.4.4]. When they converge, the Edgeworth and Gram-Charlier series are essentially the same, differing only in order. The Edgeworth expansions order the terms in ascending powers of $(1/\sqrt{n})$; the Gram-Charlier expansions order terms in derivatives of $\Phi(x)$ in ascending order, or, equivalently, in Tchebyshev-Hermite polynomials in increasing order. It therefore seems natural to employ Edgeworth series for $\bar{G}_n(x)$ or $g_n(x)$ as a device for approximations to the distribution of $(X_1 + \cdots + X_n - n\mu)/(\sigma\sqrt{n})$ if n is large or even moderately large, and the highest moment appearing in the coefficient of $n^{k/2}$ is μ_{k+2} (Johnson and Kotz, 1970, p. 17).

[*6.4.11*] If n = 1, the above expressions yield expansions for G(x) and its pdf g(x) in terms of $\Phi(x)$ and $\phi(x)$. For some intervals in x, either the Edgeworth or Gram-Charlier approximations or both may give negative density functions, particularly in the tails of the distribution. In the same way, they may not give unimodal curves, even when g(x) is unimodal. Barton and Dennis (1952, pp. 425-427) discuss these shortcomings when only terms containing moments up to μ_4 are used; they give a diagram in the (β_1, β_2) plane $(\beta_1 = \mu_3^2/\sigma^6,\ \beta_2 = \mu_4/\sigma^4)$ showing regions in which each type of expansion of g(x) is unimodal, and also regions in which the approximation to g(x) is nonnegative for all values of x. The Gram-Charlier approximation does better than the Edgeworth by both criteria in their study; the approximations tend to perform best by those criteria when the skewness $\sqrt{\beta_1}$ and kurtosis β_2 are near to zero and ≤ 5, respectively. See Kendall and Stuart (1977, p. 174) and Johnson and Kotz (1970, pp. 18-20), where the diagram of Barton and Dennis (1952) is reproduced.

Note also that the sum to k terms, say, of the Gram-Charlier series for g(x) in [6.4.8] may fluctuate irregularly from one value of k to the next. It has been customary in both series for g(x) to include terms involving μ_2, μ_3, and μ_4 only, and sometimes with the Gram-Charlier series for g(x), to include terms as far as that in $H_6(x)$ (Kendall and Stuart, 1977, pp. 147, 174).

[6.4.12] Cornish and Fisher (1937, pp. 307-320) developed expansions of quantiles of continuous distributions in terms of corresponding quantiles of a standard normal distribution, and vice versa; these were further extended by Fisher and Cornish (1960, pp. 209-226). We shall give these here in the context of normalized sums $(X_1 + \cdots + X_n - n\mu)/(\sqrt{n}\sigma)$ with cdf $\bar{G}_n(x)$, where $X_1, X_2, \ldots$ is an iid sequence of rvs with common pdf g(x), mean μ, variance σ^2, and cumulants $\kappa_3, \kappa_4, \ldots$. As in [6.4.2], $\bar{G}_n$ has cumulants λ_{rn}, where $\lambda_{rn} = \kappa_r/(\sigma^r n^{1-r/2}) = \lambda_r/n^{1-r/2}$; $r = 3, 4, \ldots$.

Let $\Phi(z_p) = \bar{G}_n(x_p)$. Then

$$\begin{aligned}
x_p = z_p &+ \frac{\kappa_3}{6\sigma^3 n^{1/2}}(z_p^2 - 1) + \left[\frac{\kappa_4}{24\sigma^4}(z_p^2 - 3z_p) + \frac{\kappa_3^2}{36\sigma^6}(-2z_p^3 + 5z_p)\right]\frac{1}{n} \\
&+ \left[\frac{\kappa_5}{120\sigma^5}(z_p^4 - 6z_p^2 + 3) + \frac{\kappa_3\kappa_4}{24\sigma^7}(-z_p^4 + 5z_p^2 - 2)\right. \\
&\left. + \frac{\kappa_3^3}{324\sigma^9}(12z_p^4 - 53z_p^2 + 17)\right]\frac{1}{n^{3/2}} + \left[\frac{\kappa_6}{720\sigma^6}(z_p^5 - 10z_p^3 + 15z_p)\right. \\
&+ \frac{\kappa_3\kappa_5}{180\sigma^8}(-2z_p^5 + 17z_p^3 - 21z_p) + \frac{\kappa_4^2}{384\sigma^8}(-3z_p^5 + 24z_p^3 - 29z_p) \\
&+ \frac{\kappa_3^2\kappa_4}{288\sigma^{10}}(14z_p^5 - 103z_p^3 + 107z_p) + \frac{\kappa_4^3}{7776\sigma^{12}}(-252z_p^5 + 1688z_p^3 \\
&\left. - 1511z_p)\right]\frac{1}{n^2} + \cdots
\end{aligned}$$

The inverse expansion of z_p in terms of x_p is given by (Draper and Tierney, 1973, pp. 503, 511-516; Johnson and Kotz, 1970, p. 34)

$$
\begin{aligned}
z_p = x_p &+ \frac{\kappa_3}{6\sigma^3 n^{1/2}}(-x_p^2 + 1) + \left[\frac{\kappa_4}{24\sigma^4}(-x_p^3 + 3x_p) + \frac{\kappa_3^2}{36\sigma^6}(4x_p^3 - 7x_p)\right]\frac{1}{n} \\
&+ \left[\frac{\kappa_5}{120\sigma^5}(-x_p^4 + 6x_p^2 - 3) + \frac{\kappa_3\kappa_4}{144\sigma^7}(11x_p^4 - 42x_p^2 + 15)\right. \\
&+ \left.\frac{\kappa_3^3}{648\sigma^9}(-69x_p^4 + 187x_p^2 - 52)\right]\frac{1}{n^{3/2}} + \left[\frac{\kappa_6}{720\sigma^6}(-x_p^5 + 10x_p^3 - 15x_p)\right. \\
&+ \frac{\kappa_3\kappa_5}{360\sigma^8}(7x_p^5 - 48x_p^3 + 51x_p) + \frac{\kappa_4^2}{384\sigma^8}(5x_p^5 - 32x_p^3 + 35x_p) \\
&+ \frac{\kappa_3^2\kappa_4}{864\sigma^{10}}(-111x_p^5 + 547x_p^3 - 456x_p) \\
&+ \left.\frac{\kappa_4^3}{7776\sigma^{12}}(948x_p^5 - 3628x_p^3 + 2473x_p)\right]\frac{1}{n^2} + \cdots
\end{aligned}
$$

[*6.4.13*] Cornish-Fisher expansions for the quantile x_p of a *standardized* random variable $(X - \mu)/\sigma$, where X is a rv with pdf $g(x)$, mean μ, variance σ^2, and cumulants κ_3, κ_4, ..., are obtained from [6.4.12] by putting n equal to one, and similarly for the inverse expansion of z_p. The former then becomes a transformation to normality.

[*6.4.14*] The remarks in [6.4.7] apply with $\bar{G}_n(x)$ and $\Phi(x)$ replaced by x_p and z_p, respectively, and vice versa. It has been more common, however, for Cornish-Fisher expansions to be used as transformations of a unit normal rv (with n = 1 in [6.4.12]), but the form in which they appear in [6.4.12] should make them easier to use for the normalized sums with which this chapter has been mainly concerned.

Note that the negative frequencies and multimodalities of some Edgeworth series do not apply, and terms up to those in $1/n^2$ (involving κ_6) are more frequently included in approximations than they are in [6.4.2], say (Johnson and Kotz, 1970, p. 35). Draper and Tierney (1973, pp. 503-518) give coefficients of terms involving cumulants up to κ_{10}, accounting in both forms of Cornish-Fisher expansion for terms up to those in $1/n^4$.

Fisher and Cornish (1960, pp. 211-213) give values to 12 decimal places of Hermite polynomials $H_r(x_p)$; $r = 1(1)7$; $p = 0.0005$ $(\times 10, 10^2, 10^3)$, $0.001(\times 10, 10^2)$, $0.0025(\times 10, 10^2)$; and for the same values of p, numerical values of the coefficients of the adjusted polynomials appearing in the first expansion in [6.4.12], to 5 decimal places.

REFERENCES

The numbers in square brackets give the sections in which the corresponding reference is cited.

Adams, W. J. (1974). *The Life and Times of the Central Limit Theorem*, New York: Caedmon. [6.3]

Anscombe, F. J. (1952). Large-sample theory of sequential estimation, *Proceedings of the Cambridge Philosophical Society* 48, 600-607. [6.2.5]

Ash, R. B. (1972). *Real Analysis and Probability*, New York: Academic. [6.1.8]

Barton, D. E., and Dennis, K. E. (1952). The conditions under which Gram-Charlier and Edgeworth curves are positive definite and unimodal, *Biometrika* 39, 425-427. [6.4.11]

Beek, P. van (1972). An application of the Fourier method to the problem of sharpening the Berry-Esseen inequality. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 23, 187-197. [6.3.2, 3]

Berry, A. C. (1941). The accuracy of the Gaussian approximation to the sum of independent variates, *Transactions of the American Mathematical Society* 49, 122-136. [6.3.2]

Bhattacharya, R. N., and Rao, R. R. (1976). *Normal Approximation and Asymptotic Expansions*, New York: Wiley. [6.3.2]

Bjerve, S. (1977). Error bounds for linear combinations of order statistics, *Annals of Statistics* 5, 357-369. [6.3.10.2]

Brown, B. M. (1971). Martingale central limit theorems, *Annals of Mathematical Statistics* 42, 59-66. [6.2.4]

Callaert, H., and Janssen, P. (1978). The Berry-Esseen theorem for U-statistics, *Annals of Statistics* 6, 417-421. [6.3.11]

Charlier, C. V. L. (1905). Über die Darstellung willkürlicher Funktionen, *Arkiv för Matematik, Astronomi, och Fysik* 2(20), 1-35. [6.4.8]

Chung, K. L. (1974). *A Course in Probability Theory*, New York: Academic. [6.2.4, 5, 6]

Cornish, E. A., and Fisher, R. A. (1937). Moments and cumulants in the specification of distributions, *Review of the International Statistical Institute* 5, 307-320. [6.4.12]

Cramér, H. (1946). *Mathematical Methods of Statistics*, Princeton, N.J.: Princeton University Press. [6.4.1, 4, 8]

Cramér, H. (1970). *Random Variables and Probability Distributions*, 3rd ed. (1st ed., 1937), London: Cambridge University Press. [6.1; 6.1.5, 7, 8; 6.3.2; 6.4.1, 2, 6, 9]

David, H. A. (1970). *Order Statistics*, New York: Wiley. [6.2.11]

de Wet, T. (1976). Berry-Essen results for the trimmed mean, *South African Statistical Journal* 10, 77-96. [6.3.10.1]

Draper, N. R., and Tierney, D. E. (1973). Exact formulas for additional terms in some important series expansions, *Communications in Statistics* 1, 495-524. [6.4.1, 2, 4, 6, 12, 14]

Edgeworth, F. Y. (1905). The law of error, *Transactions of the Cambridge Philosophical Society* 20, 36-65, 113-141. [6.4.2]

Esseen, C. G. (1942). On the Liapounoff limit of error in the theory of probability, *Arkiv för Matematik, Astronomi, och Fysik* 28A, 1-19. [6.3.2]

Esseen, C. G. (1956). A moment inequality with an application to the central limit theorem, *Skandinavisk Aktuarietidskrift* 39, 160-170. [6.3.2]

Feller, W. (1935). Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung, *Mathematische Zeitschrift* 40, 521-559. [6.1.8]

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, Vol. 1 (3rd ed.), New York: Wiley. [6.1.1]

Feller, W. (1968). On the Berry-Esseen theorem. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 10, 261-268. [6.3.6]

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. 2 (2nd ed.), New York: Wiley. [6.1.5.8; 6.2.2, 3, 5, 8; 6.3.2; 6.4.4]

Fisher, R. A., and Cornish, E. A. (1960). The percentile points of distributions having known cumulants, *Technometrics* 2, 209-225. [6.4.12, 14]

Fraser, D. A. S. (1957). *Nonparametric Methods in Statistics*, New York: Wiley. [6.2.4, 14]

Gnedenko, B. V. (1962, 1968). *The Theory of Probability* (4th ed., trans. B. D. Seckler), New York: Chelsea. [6.1; 6.1,1, 2; 6.2.2, 3]

Gnedenko, B. V., and Kolmogorov, A. N. (1968). *Limit Distributions for Sums of Independent Random Variables* (rev. ed., trans. K. L. Chung), Reading, Mass.: Addison-Wesley. [6.1; 6.2.3, 7; 6.3.1, 2, 4, 7; 6.4.1, 6]

Helmers, R. (1977). The order of the normal approximation for linear combinations of order statistics with smooth weight functions, *Annals of Probability* 5, 940-953. [6.3.10.2]

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution, *Annals of Mathematical Statistics* 19, 293-325. [6.2.14]

Hoeffding, W., and Robbins, H. (1948). The central limit theorem for dependent random variables, *Duke Mathematical Journal* 15, 773-780. [6.2.4]

Hollander, M., and Wolfe, D. A. (1973). *Nonparametric Statistical Methods*, New York: Wiley. [6.2.14]

Johnson, N. L., and Kotz, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions*, Vol. 1, New York: Wiley. [6.4.1; 6.4.5, 8, 10, 11, 12, 14]

Kendall, M. G., and Stuart, A. (1977). *The Advanced Theory of Statistics*, Vol. 1 (4th ed.), New York: Macmillan. [6.4.1, 8, 9, 11]

Landers, D., and Rogge, L. (1977). Inequalities for conditioned normal approximations, *Annals of Probability* 5, 595-600. [6.1.9; 6.3.9]

Lehmann, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks*, San Francisco: Holden-Day. [6.2.14]

Lévy, P. (1925). *Calcul des Probabilités*, Paris. [6.1.5]

Lindeberg, J. W. (1922). Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung *Mathematische Zeitschrift* 15, 211-225. [6.1.5, 8]

Lyapunov, A. M. (1901). Nouvelle forme du théorème sur la limite de probabilité, *Mémoires de l'Académie Impériale des Sciences de St. Pétersbourg* 12, 1-24. [6.1.6; 6.3.1]

McGillivray, W. R., and Kaller, C. L. (1966). A characterization of deviation from normality under certain moment assumptions, *Canadian Mathematical Bulletin* 9, 509-514. [6.3.8]

Maistrov, L. E. (1974). *Probability Theory: A Historical Sketch* (trans. S. Kotz), New York: Academic. [6.1.3, 4, 6]

Mosteller, F. (1946). On some useful "inefficient" statistics, *Annals of Mathematical Statistics* 17, 377-408. [6.2.10]

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications* (2nd ed.), New York: Wiley. [6.2.9, 10]

Rényi, A. (1958). On mixing sequences of sets, *Acta Mathematica* (Hungarian Academy of Sciences) 9, 215-228. [6.1.9]

Rényi, A. (1970). *Probability Theory*, Budapest: Akadémiai Kiadó. [6.2.3, 15, 16]

Rogozin, B. A. (1960). A remark on Esseen's paper: "A moment inequality with an application to the central limit theorem," *Theory of Probability and Its Applications* 5, 114-117. [6.3.5]

Serfling, R. J. (1968). Contributions to central limit theory for dependent variables, *Annals of Mathematical Statistics* 39, 1158-1175. [6.2.4]

Stigler, S. M. (1969). Linear functions of order statistics, *Annals of Mathematical Statistics* 40, 770-788. [6.2.13]

Stigler, S. M. (1973). The asymptotic distribution of the trimmed mean, *Annals of Statistics* 1, 472-477. [6.2.12; 6.2.13.3]

Stigler, S. M. (1974). Linear functions of order statistics with smooth weight functions, *Annals of Statistics* 2, 676-693. [6.2.13]

Tchebyshev, P. L. (1890). Sur deux théorèmes relatifs aux probabilités, *Acta Mathematica* 14, 305-315; reprinted (1962) in *Oeuvres*, Vol. 2, New York: Chelsea. [6.4.1]

Wilks, S. S. (1962). *Mathematical Statistics*, New York: Wiley. [6.4.1, 7; 6.2.10]

Woodroofe, M. (1975). *Probability with Applications*, New York: McGraw-Hill. [6.1.5; 6.1.1, 2, 8]

Zahl, S. (1966). Bounds for the central limit theorem error, *SIAM Journal on Applied Mathematics* 14, 1225-1245. [6.3.3]

Zolotarev, V. M. (1966). An absolute estimate of the remainder term in the central limit theorem, *Theory of Probability and Its Applications* 11, 95-105. [6.3.2]

chapter 7

NORMAL APPROXIMATIONS TO DISTRIBUTIONS

Many of the distributions which cannot be evaluated in closed form can be approximated, and it is not surprising that many of these approximations are based on the normal law. The best are those which are asymptotic, converging in some sense to normality; one class of distributions of this kind are those which can be represented (like the binomial, Poisson, and gamma distributions) as that of the sum of iid random variables, so that the central limit theorem can be used to provide suitable approximations.

Another source of distributions which lend themselves to normality approximations are those whose distribution functions can be expressed directly in terms of the incomplete gamma and beta function ratios. The latter is the right-tail probability of a binomial rv (which has the central limit property), and can be directly related to the cdfs of the negative binomial, beta, Student t, and F distributions; see Abramowitz and Stegun (1964, pp. 946-947) for details. It is often (but not always) the case, then, that approximations which are good for the binomial are good in some sense for these other distributions.

We have not attempted to list all approximations based on normality. A poor approximation may be given if it is of historical interest, but in general, the aim has been to present those combining accuracy with simplicity, having the user with a desk calculator or looking for a straightforward algorithm in mind. The criteria for accuracy or absolute and relative error are mentioned whenever a clear statement of the accuracy achieved can be given.

In each section, the rough order of the material is as follows: simple approximations to the cdf, accurate approximations to the cdf, exact normal deviates, and approximations to percentiles. Bounds are included in this chapter, when these are based on normality properties.

The source material contains much more information than is given here. Extensive discussions appear in the books by Johnson and Kotz (1969, 1970a, 1970b) and Molenaar (1970), and in Peizer and Pratt (1968, pp. 1416-1456).

7.1 THE BINOMIAL DISTRIBUTION

[*7.1.1*] *Notation*. Let the probability function (pf) of a binomial rv Y be g(y;n,p), where

$$\Pr(Y = y) = g(y;n,p) = \binom{n}{y} p^y (1 - p)^{n-y}, \quad y = 0, 1, \ldots, n; \quad 0 < p < 1$$

and let $q = 1 - p$. Denote the cdf of Y by $G(y;n,p) = \Pr(Y \le y)$, and note that

$$G(k;n,p) = 1 - G(n - k - 1; n, q)$$

a result which may give a choice for each approximation, in trying to reduce error.

7.1.2 *Classical Normal Approximation with Continuity Correction*. This approximation is derived directly from the de Moivre-Laplace limit theorem (see [6.1.1]) (Parzen, 1960, p. 239; Roussas, 1973, pp. 141-142):

$$G(y;n,p) \simeq \Phi\{(y + 1/2 - np)/\sqrt{npq}\}, \qquad y = 0, 1, 2, \ldots, n$$

$$\Pr(a \le Y \le b) \simeq \Phi\{(b + 1/2 - np)/\sqrt{npq}\} - \Phi\{(a - 1/2 - np)/\sqrt{npq}\},$$
$$0 \le a \le b \le n, \text{ a and b integers}$$

If a, b, or y are not integers (but $0 \le a, b, y \le n$ still holds), they should be replaced in the normal approximation by [a], [b],

[y], where [a], for example, denotes "the largest integer not greater than a."

This approximation is alternatively given by

$$\Pr\{\alpha \le (Y - np)/\sqrt{npq} \le \beta\} \simeq \Phi\{([\beta\sqrt{npq} + np] + 1/2 - np)/\sqrt{npq}\} - \Phi\{([\alpha\sqrt{npq} + np] + 1/2 - np)/\sqrt{npq}$$

where [x] is defined as above, and $0 \le \alpha\sqrt{npq} + np \le \beta\sqrt{npq} + np \le n$. Raff (1956, pp. 294-295) investigated the maximum error which could arise in estimating *any* sum of consecutive binomial terms. If this is M(n,p), then $M(n,p) \le 0.140/\sqrt{npq}$ always, and if $np^{3/2} > 1.07$, then $M(n,p) \le 0.05$. If n is fixed, M(n,p) decreases as p increases to 0.5, except that the trend is reversed when p is close to zero. If p is fixed, then M(n,p) decreases as n increases.

[*7.1.3*] From the local de Moivre-Laplace limit theorem of [6.1.2],

$$g(y;n,p) = \binom{n}{y} p^y (1 - p)^{n-y} \simeq \phi[(y - np)/\sqrt{npq}]/\sqrt{npq} \simeq f(y;np,npq)$$

Then $\Sigma_{k=0}^{n} |g(y;n,p) - f(y;np,npq)| = \{1 + 4\exp(-3/2)\}/(3\sqrt{2\pi npq}) + O(n^{-1}p^{-1}q^{-1})$ (Johnson and Kotz, 1969, p. 67; Govindarajulu, 1965, p. 152). Alternatively, g(y;n,p) can be approximated as in [7.1.2], with b = y, a = y - 1.

[*7.1.4*] ***Recommended***. For a combination of accuracy and simplicity, Molenaar (1970, pp. 7-8, 86-88, 109-110) recommends versions of

$$\Phi[\sqrt{4k + a}\,\sqrt{q} - \sqrt{4n - 4k + b}\,\sqrt{p}]$$

to approximate G(k;n,p); k = 0, 1, ..., n (Freeman and Tukey, 1950, pp. 607-611, where a = 4, b = 0).

Suppose that p is close to 1/2; that is, say,

If n = 3, $0.25 \le p \le 0.75$

If n = 30, $0.40 \le p \le 0.60$

If n = 300, $0.46 \le p \le 0.54$

Then

$$G(k;n,p) \simeq \Phi[\sqrt{4k + 3}\,\sqrt{q} - \sqrt{4n - 4k - 1}\,\sqrt{p}]$$
$$\text{if } 0.45 \le |G(k;n,p) - \tfrac{1}{2}| \le 0.495$$

and

$$G(k;n,p) \simeq \Phi[\sqrt{4k + 2.5}\,\sqrt{q} - \sqrt{4n - 4k - 1.5}\,\sqrt{p}]$$
$$\text{if } 0.05 \le G(k;n,p) \le 0.93$$

If p is not close to 1/2 in the sense above, then

$$G(k;n,p) \simeq \Phi[\sqrt{4k + 4}\,\sqrt{q} - \sqrt{4n - 4k}\,\sqrt{p}]$$
$$\text{if } 0.45 \le |G(k;n,p) - \tfrac{1}{2}| \le 0.495$$

$$G(k;n,1) \simeq \Phi[\sqrt{4k + 3}\,\sqrt{q} - \sqrt{4n - 4k - 1}\,\sqrt{p}]$$
$$\text{if } 0.05 \le G(k;n,p) \le 0.93$$

Molenaar (1970, pp. 111-114) tables values of the relative error in these approximations for certain values of n between 5 and 100, when p = 0.05, 0.20, 0.40, and 0.50. The absolute error is $O(1/\sqrt{n})$ if $p \ne 1/2$, and $O(1/n)$ if $p = 1/2$ (Molenaar, 1970, p. 92).

[*7.1.5*] Improved classical normal approximations are given by

$$G(k;n,p) \simeq \Phi[(k + c - np)\{(n + d)pq + \delta\}^{-1/2}]$$

(in [7.1.2], $c = 1/2$, $d = \delta = 0$).

If $c = (2 - p)/3$, $d = 1/3$, and $\delta = 0$, the maximum error in the approximation is reduced to about half of that in [7.1.2] unless p is close to 1/2, when this property is reversed (Gebhardt, 1969, pp. 1641-1642). The problem of choosing c, d, and δ optimally in some sense, however, is complicated; see Molenaar (1970, pp. 76-79).

[*7.1.6*] *The Angular Approximation*. (a) Raff (1956, pp. 293-303) gives the approximation

$$G(k;n,p) \simeq \Phi[2\sqrt{n}\{\arcsin\sqrt{(k + 1/2)/n} - \arcsin\sqrt{p}\}]$$

If M(n,p) is the maximum error as defined in [7.1.1], then $M(n,p) \le 0.140/\sqrt{npq}$. This approximation is preferable to that in [7.1.1] unless p is close to 1/2.

(b) An improved angular approximation is given by

$$G(k;n,p) \simeq \Phi[2\sqrt{n+\delta}\{\arcsin\sqrt{(k+1/2+\beta)/(n+\gamma)} - \arcsin\sqrt{p}\}]$$

where $\beta = q\gamma$, $\beta = \gamma/2$, or $\beta = p\gamma$ (Molenaar, 1970, p. 82). In (a) above, $\beta = \gamma = \delta = 0$. Suitable choices of β, γ, and δ lead to an approximation more accurate than that in [7.1.5]; see Molenaar (1970, pp. 82-84).

Gebhardt (1969, p. 1642) takes $\beta = 1/6$, $\gamma = 1/3$, and $\delta = 1/3$; the maximum error is one-half to one-tenth (when np is small) of that in the approximation in (a). Johnson and Kotz (1969, p. 65) take $\beta = 3/8$, $\gamma = 3/4$, $\delta = 0$. See also [7.15.4].

[*7.1.7*] The *Camp-Paulson* approximation is more accurate than those listed above, but also more complex (Camp, 1951, pp. 130-131):

$$G(k;n,p) \simeq \Phi\{-x/(3\sqrt{z})\}$$

$$x = \left[\frac{(n-k)p}{(k+1)q}\right]^{1/3}\left(9 - \frac{1}{n-k}\right) + \frac{1}{k+1} - 9$$

$$z = \left[\frac{(n-k)p}{(k+1)q}\right]^{2/3}\left(\frac{1}{n-k}\right) + \frac{1}{k+1}$$

See Molenaar (1970, pp. 92-94). The error is $0(npq)^{-1})$ for all p, and is $0((npq)^{-3/2})$ if $p = 1/2$ or $p = 0.042$ (Molenaar, 1970, p. 94). The maximum absolute error $M(n,p)$, defined in the same way as in [7.1.2], is less than 0.0122 for any values of n and p, decreasing as np increases above a value of 0.02. Further, $M(n,p) \le 0.007/\sqrt{npq}$ (Raff, 1956, pp. 293-303). See also remarks below in [7.1.8].

[*7.1.8*] *The Borges Approximation* (Borges, 1970, pp. 189-199). Let k be an integer, $0 \le k < n$. Then

$$G(k;n,p) \simeq \Phi(y_{k+(1/2)})$$

$$y_k = (pq)^{-1/6}\sqrt{n+1/3}\int_p^{h(k,n)}\{s(1-s)\}^{-1/3}\,ds$$

$$h(k,n) = (k+1/6)/(n+1/3)$$

The maximum absolute error $M(n,p)$, defined as in [7.1.2], is $0(1/(npq))$ for all values of p, and is $0(1/(npq)^{3/2})$ if $p = 1/2$ (Molenaar, 1970, p. 95). Gebhardt (1971, pp. 189-190) has tabled the incomplete beta integral

$$B(x;\ 2/3,\ 2/3) = \int_{1/2}^{x}\{s(1 - s)\}^{-1/3}\,ds = -B(1 - x;\ 2/3,\ 2/3)$$

for $x = 0.50(0.001)1.00$ and $x = 0.990(0.0001)1.000$ to 5 decimal places; also he tables $(pq)^{-1/6}$ for $p = 0.00(0.001)0.500$ to 5 decimal places.

Molenaar (1970, p. 95) suggests the approximation

$$\int_0^x \{s(1 - s)\}^{-1/3}\,ds \simeq 1.5x^{2/3}(60 - 17x)/(60 - 25x),\ 0 < x \le \frac{1}{2}$$
$$= J(x),\ \text{say}$$

and

$$\simeq 2.0533902 - J(1 - x),\ \frac{1}{2} < x < 1$$

The agreement is good to 2 decimal places if $x = 1/2$, improving to 5 decimal places if $x = 0.10$.

Both the Borges and Camp-Paulson approximations are superior to those of [7.1.5] with $c = (2 - p)/3$, $d = 1/3$, and $\delta = 0$, and of [7.1.6(b)] (arcsine) with $\beta = 1/6$, $\gamma = 1/3$, and $\delta = 1/3$ (Gebhardt, 1969, p. 1645). The Camp-Paulson is better than the Borges approximation (by not very much) for all values of p when $n \le 20$, say, and if $0.2 < p < 0.8$ when $n \ge 50$, say (Molenaar, 1970, p. 98; Gebhardt, 1969, pp. 1643, 1645), although these conclusions consider the amount of computation as well as relative error as criteria; based on maximum absolute error alone, the Borges approximation is in fact preferable if $n \le 20$ and np is near to 2 or 3.

[*7.1.9*] A more accurate approximation than any described so far, but also more cumbersome, is due to Peizer and Pratt (1968, pp. 1417-1420, 1448); this is accurate to $O(1/(npq)^{3/2})$ if $p \ne 1/2$, and to $O(1/(npq)^2)$ if $p = 1/2$. It is given by $G(k;n,p) \simeq \Phi(z)$, where

$$z = \frac{d[1 + qT\{(k + 1/2)/(np)\} + pT\{(n - k - 1/2)/(nq)\}]^{1/2}}{\sqrt{pq(n + 1/6)}}$$

$$d = k + 2/3 - (n + 1/3)p + 0.02\left(\frac{q}{k + 1} - \frac{p}{n - k} + \frac{q - 1/2}{n + 1}\right)$$

$$T(x) = (1 - x^2 + 2x \log x)/(1 - x)^2, \qquad x \neq 1$$
$$T(1) = 0$$

See Molenaar (1970, p. 102) for a refinement of d. For the above (Peizer and Pratt, 1968, p. 1418),

$$|G(k;n,p) - \Phi(z)| \le \begin{cases} 0.001 & \text{if } k \ge 1,\ n - k \ge 2 \\ 0.01 & \text{if } k \ge 0,\ n - k \ge 1 \end{cases}$$

[*7.1.10*] A *recommended* approximation, more accurate than that given in [7.1.4], has been proposed by Molenaar (1970, pp. 100, 105, 110):

$$G(k;n,p) \simeq \Phi\{2\sqrt{(k + 1)q + A} - 2\sqrt{(n - k)p + B}\}, \qquad p \neq 1/2$$
$$A = \{(4 - 10p + 7p^2)(k + 1/2 - np)^2/(36npq)\} - (8 - 11p + 5p^2)/18$$
$$B = \{(1 - 4p + 7p^2)(k + 1/2 - np)^2/(36npq)\} - (2 + p + 5p^2)/18$$
$$G(k;n,1/2) \simeq \Phi(\sqrt{2k + 2 + \beta} - \sqrt{2n - 2k + \beta})$$
$$\beta = \{(2k + 1 - n)^2 - 10n\}/(12n)$$

These are accurate to $O(1/(npq)^{3/2})$ if $p \neq 1/2$, and to $O(1/(npq)^2)$ if $p = 1/2$. See Molenaar (1970, pp. 100-101, 104) for slightly improved but more cumbersome alternatives.

[*7.1.11*] Let $u = (k + 1/2 - np)/\sigma$ and $\sigma = \sqrt{npq}$. Then (Molenaar, 1970, pp. 72, 76) the *exact normal deviate* z defined by $G(k;n,p) = \Phi(z)$ satisfies

$$\begin{aligned} z = {} & u + \sigma^{-1}(q - p)(-u^2 + 1)/6 \\ & + \sigma^{-2}\{(5 - 14pq)u^3 + (-2 + 2pq)u\}/72 \\ & + \sigma^{-3}(q - p)\{(-249 + 438pq)u^4 + (79 - 28pq)u^2 \\ & + 128 - 26pq\}/6480 + O(\sigma^{-4}) \end{aligned}$$

$$u = z + \sigma^{-1}(1 - 2p)(z^2 - 1)/6$$
$$+ \sigma^{-2}\{z^3(2p^2 - 2p - 1) + z(-14p^2 + 14p - 2\}/72$$
$$+ \sigma^{-3}\{z^4(6p^3 - 9p^2 - 9p + 6) + z^2(14p^3 - 21p^2 - 21p + 14)$$
$$- 32p^3 + 48p^2 + 48p - 32\}/1620 + 0(\sigma^{-4})$$

See also Pratt (1968, p. 1467).

[*7.1.12*] If an approximation is desired to be *accurate near an assigned probability* α or $1 - \alpha$, and if $\Phi(z_\alpha) = 1 - \alpha(0 < \alpha \leq 1/2)$, then

$$G(k;n,p) = \Phi(z) \simeq \Phi(u)$$
$$u = 2\sqrt{(k + 1 + b)q} - 2\sqrt{(n - k + b)p}$$
$$b = (z_\alpha^2 - 4)/12$$

The error is $(q - p)(z_\alpha^2 - z^2)/(12\sqrt{npq}) + 0(1/(npq))$. See Molenaar (1970, pp. 88-91) for discussion and possible refinements.

[*7.1.13*] Bounds. Johnson and Kotz (1969, pp. 68-70) give several bounds on cumulative and individual binomial probabilities; these bounds involve exponential functions. Slud (1977, pp. 406-410), however, gives several inequalities involving the normal cdf ($k = 0, 1, 2, \ldots, n$):

(a) If $0 \leq p \leq 1/4$, $np \leq k \leq n$, or $np \leq k \leq nq$, then

$$\Pr(Y \geq k) = 1 - G(k - 1; n, p) \geq 1 - \Phi\{(k - np)/\sqrt{npq}\}$$

(b) (Slud, 1977, p. 404) If $k \leq np$, the conclusion in (a) holds.

(c) If $np \leq k \leq nq$ and $k \geq 2$, then

$$g(k;n,p) > \Phi\{(k - np + 1)/\sqrt{npq}\} - \Phi\{(k - np)/\sqrt{npq}\}$$

(d) If $p \leq 1/4$ and either (i) $k \geq nq$ or (ii) $k \geq n/3$, $n \geq 27$, then

$$g(k;n,p) - [\Phi\{(k - np + 1)/\sqrt{npq}\} - \Phi\{(k - np)/\sqrt{npq}\}]$$
$$> \gamma\phi\{(k - np)/\sqrt{npq}\}/\sqrt{npq}$$

where, if (ii) holds, $\gamma = \delta = \{(k - np)^2/(2npq)\}(k^{-1} - (3npq)^{-1}$, and if (i) holds, $\gamma = \max(\delta, 0.16)$.

7.2 THE POISSON DISTRIBUTION

[7.2.1] *Notation*. Let the pf of a Poisson rv Y be $g(y;\lambda)$, where

$$\Pr(Y = y) = g(y;\lambda) = e^{-\lambda}\lambda^y/y!, \qquad y = 0, 1, 2, \ldots$$

Denote the cdf of Y by $G(y;\lambda)$, $= \Pr(Y \le y)$.

[7.2.2] *Classical Normal Approximation with Continuity Correction*

$$G(k;\lambda) \simeq \Phi[(k + 1/2 - \lambda)/\sqrt{\lambda}], \qquad k = 0, 1, 2, \ldots$$

Asymptotic behavior indicates that, for large λ, the correction of 1/2 should only be used for probabilities between 0.057 and 0.943; if $\lambda = 10$, for probabilities between 0.067 and 0.953; and if $\lambda = 4$, between 0.073 and 0.958. The approximation may overestimate probabilities less than 0.16 and underestimate $1 - \Pr(Y \le y)$ when the latter is less than 0.16. See Molenaar (1970, pp. 34-36) for discussion of these and other points.

With these cautions in mind, for nonnegative integers a and b (Parzen, 1960, p. 248),

$$\Pr(a \le Y \le b) \simeq \Phi[(b + 1/2 - \lambda)/\sqrt{\lambda}] - \Phi[(a + 1/2 - \lambda)/\sqrt{\lambda}]$$
$$g(k;\lambda) \simeq \Phi[(k + 1/2 - \lambda)/\sqrt{\lambda}] - \Phi[(k - 1/2 - \lambda)/\sqrt{\lambda}]$$

[7.2.3] *Recommended Simple Approximation, Accurate to* $O(\lambda^{-1/2})$ (Molenaar, 1970, pp. 8-9, 38-40, 64).

$$G(k;\lambda) \simeq \Phi(2\sqrt{k + 1} - 2\sqrt{\lambda}) \qquad \text{for tails}$$
$$G(k;\lambda) \simeq \Phi(2\sqrt{k + 3/4} - 2\sqrt{\lambda}), \quad 0.09 < G(k;\lambda) < 0.94; \quad \lambda \ge 15$$
$$0.05 < G(k;\lambda) < 0.93; \quad \lambda < 15$$

This approximation is more accurate than that of [7.2.2], and is based on that of Freeman and Tukey (1950, p. 607). Molenaar (1970, pp. 38-41) discusses $\Phi(2\sqrt{k + \alpha} - 2\sqrt{\lambda + \beta})$ for various choices of α and β as an approximation to $G(k;\lambda)$.

[*7.2.4*] There are also variance-stabilizing transformations $\sqrt{Y + \alpha}$ and $\sqrt{Y} + \sqrt{Y + 1}$, described in [7.15.3], which can be used as normal approximations.

[*7.2.5*] (a) An approximation of Makabe and Morimura (1955, pp. 37-38) has bounds which also give accuracy to $O(1/\lambda)$ as $\lambda \to \infty$; see Govindarajulu (1965, p. 156). Let k and m be nonnegative integers; then if $k < m$ and $\lambda \geq 1$,

$$\begin{aligned} G(m;\lambda) - G(k;\lambda) &= \Phi(b) - \Phi(a) + \{\phi(b)(1 - b^2) \\ &\quad - \phi(a)(1 - a^2)\}/(6\sqrt{\lambda}) + R_1 \\ |R_1| &< (0.0544)/\lambda + (0.0108)/\lambda^{3/2} + (0.2743)/\lambda^2 \\ &\quad + (0.0065)/\lambda^{5/2} + (1 + (1/2)\lambda^{-1/2}) \exp(-2\sqrt{\lambda}), \\ a &= (k - 1/2 - \lambda)/\sqrt{\lambda}, \quad b = (m + 1/2 - \lambda)/\sqrt{\lambda} \end{aligned}$$

See Molenaar (1970, p. 53) and (c) below.

(b) Cheng (1949, p. 396) gives, similarly,

$$\begin{aligned} G(k;\lambda) &= \Phi(c) + (1 - c^2)\phi(c)/(6\sqrt{\lambda}) + R_2 \\ c &= (k + 1/2 - \lambda)/\sqrt{\lambda} \\ |R_2| &< 0.076/\lambda + 0.043/\lambda^{3/2} + 0.13/\lambda^2 \end{aligned}$$

(c) If $\lambda \geq 1$,

$$\begin{aligned} g(k;\lambda) &= \phi(y)\{1/\sqrt{\lambda} + (3y - y^3)/(6\lambda)\} + R_3 \\ y &= (k - \lambda)/\sqrt{\lambda} \\ |R_3| &= (0.0748)/\lambda^{3/2} + (0.00554)/\lambda^2 + (0.3724)/\lambda^{5/2} \\ &\quad - (0.5595)/\lambda^3 + \{1 + (1/2)\lambda^{-1/2}\}\lambda^{-1/4} \exp(-2\sqrt{\lambda}) \end{aligned}$$

See sources listed in (a).

[*7.2.6*] The Wilson-Hilferty (1931, pp. 684-688) approximation to chi-square leads here to

$$G(k;\lambda) \simeq \Phi[3\sqrt{k + 1} - \{3\sqrt{k + 1}\}^{-1} - 3\lambda^{1/3}(k + 1)^{1/6}]$$

If $G(k;\lambda) = \Phi(z)$, the error in the approximation is (Molenaar, 1970, pp. 48-49)

$$\phi(z)(3z - z^3)/(108\lambda) + O(\lambda^{-3/2})$$

[7.2.7] <u>*A Recommended Accurate Approximation, with Error* $O(\lambda^{-3/2})$</u> (Molenaar, 1970, pp. 62, 64)

$$G(k;\lambda) \approx \Phi\{2\sqrt{k + (t + 4)/9} - 2\sqrt{\lambda + (t - 8)/36}\}$$

$$t = (k - \lambda + 1/6)^2/\lambda$$

The error, if $G(k;\lambda) = \Phi(z)$, is

$$\phi(z)(-6z^4 + 26z^2 + 7)/(6480\lambda^{3/2}) + O(\lambda^{-2})$$

[7.2.8] Another accurate approximation (except near $k = 0$ when $\lambda \le 5$) is that of *Peizer and Pratt* (1968, pp. 1417-1420); see also Molenaar (1970, pp. 59-61). Thus

$$G(k;\lambda) \approx \Phi(u)$$

where

$$u = \{k - \lambda + (2/3) + \varepsilon/(k + 1)\}[1 + T\{(k + (1/2))/\lambda\}]^{1/2}\sqrt{\lambda}$$

$$T(x) = (1 - x^2 + 2x \log x)/(1 - x)^2, \quad x \neq 1$$

$$T(1) = 0$$

If tail probabilities are being approximated, ε increases optimally from 0.02077 to 0.02385 as α (or $1 - \alpha$) decreases from 0.10 to 0.005. If

$$G(k;\lambda) = \Phi(z)$$

exactly, the error is (Molenaar, 1970, p. 59)

$$\phi(z)(-z^2 + 1620\varepsilon - 32)/(1620\lambda^{3/2}) + O(\lambda^{-2})$$

When $\varepsilon = 0.02$, the maximum absolute error is 0.001 if $k \ge 1$ and is 0.01 if $k \ge 0$ (Peizer and Pratt, 1968, pp. 1417-1424).

[7.2.9] (a) If $\Pr(Y \le k - 1) = G[k - 1; \lambda) = \Phi(x)$, then the *exact normal deviate* x is given by (Riordan, 1949, pp. 417, 420; Molenaar, 1970, p. 32)

$$x = u + \frac{u^2 - 1}{3k^{1/2}} + \frac{7u^3 - u}{36k} + \frac{219u^4 - 14u^2 - 13}{1620k^{3/2}}$$

$$+ \frac{3993u^5 - 152u^3 + 119u}{40320k^2} + O(k^{-5/2})$$

$$u = (k - \lambda)/\sqrt{k}$$

(b) Molenaar (1970, p. 33) gives two expressions in powers of $\lambda^{-1/2}$, and we reproduce one of them here. If $G(k;\lambda) = \Phi(z)$, then

$$z = v + \frac{-v^2 + 1}{6\lambda^{1/2}} + \frac{5v^3 - 2v}{72\lambda} + \frac{-249v^4 + 79v^2 + 128}{6480\lambda^{3/2}} + O(\lambda^{-2})$$

$$v = (k + (1/2) - \lambda)/\sqrt{\lambda}$$

[*7.2.10*] For an approximation designed to be most *accurate near an assigned probability* α or $1 - \alpha$, where $0 < \alpha < 1/2$ and $\Phi(z_\alpha) = 1 - \alpha$, we have

$$G(k;\lambda) = \Phi(z) \simeq \Phi(u)$$

$$u = 2\{k + (z_\alpha^2 + 11)/18\}^{1/2} - 2\{\lambda - (z_\alpha^2 + 2)/36\}^{1/2}$$

The error is $(z_\alpha^2 - z^2)\phi(z)(12\sqrt{\lambda}) + O(\lambda^{-3/2})$. See Molenaar (1970, pp. 41-44) for a discussion of this and other possible approximations.

[*7.2.11*] $G(k;\lambda) \le \Phi\{(k + 1 - \lambda)/\sqrt{\lambda}\}$ (Bohman, 1963, pp. 47-52).

7.3 THE NEGATIVE BINOMIAL DISTRIBUTION

[*7.3.1*] The pf of a negative binomial rv Y is $h(y;s,p)$, where

$$\Pr(Y = y) = h(y;s,p) = \binom{s + y - 1}{y} p^s q^y, \qquad s > 0,\ q = 1 - p$$

$$0 < p < 1;\ y = 0, 1, 2, \ldots$$

Let the cdf of Y be $H(y;s,p)$. The parameter s is usually, but need not be, a positive integer; alternative forms of the distribution are given in Johnson and Kotz (1969, pp. 122, 124); when $s = 1$, we have a *geometric distribution*, in which case probabilities can be computed exactly in simple form.

[*7.3.2*] Approximations to H(y;s,p) when s is a positive integer can be deduced from those for binomial probabilities discussed in [7.1], by virtue of the fact that (using the notation G for the binomial cdf in [7.1]) (Patil, 1960, p. 501; Johnson and Kotz, 1969, p. 127)

$$H(y;s,p) = 1 - G(s - 1;\ s + y,\ p) = G(y;\ s + y,\ q)$$

Hence from the approximations to G(k;n,p) in [7.1] we can construct approximations to H(y;s,p) as follows:

(i) Replace k by y.
(ii) Replace n by y + s and n - k by s.
(iii) Interchange p and q.
(iv) If $G(k;n,p) \simeq \Phi(z)$, then $H(y;s,p) \simeq \Phi(z)$, where z is expressed in terms of y, s, and p from (i), (ii), and (iii).

[*7.3.3*] To illustrate the procedure of [7.3.2], we adapt Molenaar's recommended simple yet accurate binomial approximation, given in [7.1.4]. Suppose that p is close to 1/2; that is,

If $s + y = 3$, $\quad 0.25 \le p \le 0.75$
If $s + y = 30$, $\quad 0.40 \le p \le 0.60$
If $s + y = 300$, $\quad 0.46 \le p \le 0.54$

Then

$$H(y;s,p) \simeq \Phi[\sqrt{4y + 3}\ \sqrt{p} - \sqrt{4s - 1}\ \sqrt{q}] \quad \text{if } 0.45 \le |H(y;s,p) - (1/2)| \le 0.495$$

and

$$H(y;s,p) \simeq \Phi[\sqrt{4y + 2.5}\ \sqrt{p} - \sqrt{4s - 1.5}\ \sqrt{q}] \quad \text{if } 0.07 \le H(y;s,p) \le 0.95$$

If p is not close to 1/2 in the sense above, then

$$H(y;s,p) \simeq \Phi[\sqrt{4y + 4}\ \sqrt{p} - \sqrt{4sq}] \quad \text{if } 0.45 \le |H(y;s,p) - (1/2)| \le 0.495$$

$$H(y;s,p) \simeq \Phi[\sqrt{4y + 3}\ \sqrt{p} - \sqrt{4s - 1}\ \sqrt{q}] \quad \text{if } 0.07 \le H(y;s,p) \le 0.95$$

The absolute error is $O(1/\sqrt{s + y})$ if $p \neq 1/2$, and $O(1/(s + y))$ if $p = 1/2$.

[*7.3.4*] As a second illustration of the rule in [7.3.2], we adapt the *recommended* accurate approximation of Molenaar, given in [7.1.10]:

$$H(y;s,p) \simeq \Phi\{2\sqrt{(y + 1)p + A} - 2\sqrt{sq + B}\}, \qquad p \neq 1/2$$

$$A = \{(4 - 10q + 7q^2)(py - sq + (1/2))^2/(36(s + y)pq)\}$$
$$- (8 - 11q + 5q^2)/18$$

$$B = \{(1 - 4q + 7q^2)(py - sq + (1/2))^2/(36(s + y)pq)\}$$
$$- (2 + q + 5q^2)/18$$

$$H(y;s,1/2) \simeq \Phi(\sqrt{2y + 2 + \beta} - \sqrt{2s + \beta})$$

$$\beta = \{(s - y - 1)^2 - 10(s + y)\}/\{12(s + y)\}$$

These are accurate to $O(\{(s + y)pq\}^{-3/2})$ if $p \neq 1/2$, and to $O(\{(s + y)pq\}^{-2})$ if $p = 1/2$.

[*7.3.5*] The Camp-Paulson approximation (see [7.1.7]) was adapted by Bartko (1966, p. 349; 1967, p. 498):

$$H(y;s,p) \simeq \Phi\{-x/(3\sqrt{z})\}$$

$$x = (9 - s^{-1})[sq/\{p(y + 1)\}]^{1/3} - (9y + 8)/(y + 1)$$

$$z = s^{-1}[sq/\{p(y + 1)\}]^{2/3} + 1/(y + 1)$$

This result is also obtained from [7.3.2] and [7.1.7].

Bartko's table of maximum absolute errors in selected cases is reproduced in Johnson and Kotz (1969, p. 130). If $0.05 \leq p \leq 0.95$ and $s = 5$, 10, 25, or 50, this error is no greater than 0.004, 0.005, 0.003, or 0.002, respectively; for some values of p, it may be considerably smaller.

[*7.3.6*] From *Peizer and Pratt* (1968, pp. 1417-1422) comes the following accurate approximation, also obtained from [7.3.2] and [7.1.9]:

$$H(y;s,p) \simeq \Phi(u)$$

$$u = d[1 + pT\{(y + 1/2)/(yq + sq)\} + qT\{(s - 1/2)/(yp + sp)\}]^{1/2}/\sqrt{pq(y + s + 1/6)}$$

$$d = y + 2/3 - (y + s + 1/3)q + 0.02\left(\frac{p}{y + 1} - \frac{q}{s} + \frac{p - 1/2}{y + s + 1}\right)$$

$$T(x) = (1 - x^2 + 2x \log x)/(1 - x)^2, \qquad x \neq 1$$

$$T(1) = 0$$

For the above,

$$|H(y;s,p) - \Phi(u)| \le \begin{cases} 0.001 & \text{if } y \ge 1,\ s \ge 2 \\ 0.01 & \text{if } y \ge 0,\ s \ge 1 \end{cases}$$

The approximation is accurate to $O(1/\{(y + s)pq\}^{3/2}$ if $p \neq 1/2$, and to $O(1/\{(y + s)pq\}^2)$ if $p = 1/2$ (see [7.1.9]).

[*7.3.7*] From [7.3.2] and [7.1.10], *the exact normal deviate* defined by $H(y;s,p) = \Phi(z)$, where $u = py - qs + 1/2$ and $\sigma = \sqrt{(y + s)pq}$, is given by

$$z = u + \sigma^{-1}(p - q)(-u^2 + 1)/6 + \sigma^{-2}\{(5 - 14pq)u^3 + (-2 + 2pq)u\}/72 + \sigma^{-3}(p - q)\{(-249 + 438pq)u^4 + (79 - 28pq)u^2 + 128 - 26pq\}/6480 + O(\sigma^{-4})$$

[*7.3.8*] From [7.3.2] and [7.1.11], we obtain an approximation desired to be accurate near probability α or $1 - \alpha$. If $\Phi(z_\alpha) = 1 - \alpha$ $(0 < \alpha \le 1/2)$, then

$$H(y;s,p) = \Phi(z) \simeq \Phi(u)$$

$$u = 2\sqrt{(y + 1 + b)p} - 2\sqrt{(s + b)q}$$

$$b = (z_\alpha^2 - 4)/12$$

The error is $(q - p)(z_\alpha^2 - z^2)/(12\sqrt{(s + y)pq}) + O(\{(s + y)pq\}^{-1})$.

7.4 THE HYPERGEOMETRIC DISTRIBUTION

[*7.4.1*] Let the pf of a hypergeometric rv Y be g(y;n,M,N), where $g(y;n,M,N) = \binom{M}{y}\binom{N - M}{n - y}/\binom{N}{n}$; $\max(0, n - N + M) \le y \le \min(n,M)$, y a nonnegative integer. We assume that $n \le M \le (1/2)N$; the notation of the distribution can always be arranged so that these inequalities hold (Molenaar, 1970, p. 116). Let G(y;n,M,N) be the cdf of Y.

[*7.4.2*] *Classical Normal Approximation with Continuity Correction*. If n is large, but M/N is not small, then (Johnson and Kotz, 1969, p. 149; see also Hemelrijk, 1967, pp. 226-227, and Molenaar, 1970, p. 119)

$$G(k;n,M,N) \simeq \Phi(u)$$

$$u = \frac{k + 1/2 - nM/N}{\{(N - n)(N - 1)^{-1}n(M/N)(1 - M/N)\}^{1/2}}$$

Note that $u = (k + 1/2 - E(Y))/\sqrt{\mathrm{Var}(Y)}$.

[*7.4.3*] *A Recommended Simple Approximation Having Accuracy over a Wide Range* (Molenaar, 1970, pp. 6-7, 132-133, 148)

$$G(k;n,M,N) \simeq \Phi(u),$$

$$u = \begin{cases} 2\{\sqrt{(k + 1)(N - M - n + k + 1)} \\ \quad - \sqrt{(n - k)(M - k)}\}/\sqrt{N - 1} \\ \qquad\qquad \text{for tail probabilities} \\ 2\{\sqrt{(k + 3/4)(N - M - n + k + 3/4)} \\ \quad - \sqrt{(n - k - 1/4)(M - k - 1/4)}\}/\sqrt{N} \\ \qquad\qquad 0.05 < G(k;n,M,N) < 0.93 \end{cases}$$

These approximations are generally superior to that in [7.4.2].

[*7.4.4*] Let $n/N \to t$, $M/N \to p$, $(N - M)/N \to q$, $h(k - np) \to x$ as $N \to \infty$, where $1/h = \{Npqt(1 - t)\}^{1/2}$. Then, when N is large (Feller, 1968, p. 194),

$$g(k;n;M,N) \simeq h\phi(x)$$

[7.4.5] (a) The following *bounds* were derived by Nicholson (1956, pp. 474-475). Define

$$p = 1 - q = (M + 1)/(N + 2)$$
$$s = 1 - t = (n + 1)/(N + 2)$$
$$\sigma^2 = (n + 1)pqt$$
$$a = (p - q)(t - s)/6$$
$$d_k = k - (n + 1)p$$
$$R = 5(1 - pq)(1 - st)/(36\sigma^2) + 2/(3N + 6)$$

Suppose that $\sigma > 3$, $b \geq (n + 1)p$, $c + 1/2 \leq (n + 1)p + 2\sigma^2/3$, $n - c \geq 4$ and $M - c \geq 4$, where b and c are integers such that $b \leq c \leq n$. Then

$$\Pr(b \leq Y \leq c) = G(c; n, M, N) - G(b - 1; n, M, N)$$
$$\leq \{1 - (N + 2)^{-1}\}e^R\{\Phi(y_{c+1}) - \Phi(y_b)\}$$

where

$$y_k = \frac{d_k}{\sigma} + \frac{a}{\sigma}\left(\frac{d_k}{\sigma}\right)^2 + \frac{2a}{\sigma} - \frac{1}{2}\left(\frac{1}{\sigma}\right)^2$$

The inequality is reversed if we define

$$y_k = \frac{d_k}{\sigma} + \frac{a}{\sigma}\left(\frac{d_k}{\sigma}\right)^2 + \frac{2a}{\sigma} + \frac{(d_c + 1/2)^3}{6}\left(\frac{1}{\sigma}\right)^5 + \frac{1}{7\sigma}$$

As long as $\{(c + 1/2 - (n + 1)p)/\sigma\}^3 = O(\sigma)$, the gap between the bounds is $O(\sigma^{-1})$, and an approximation lying between the bounds is good for all combinations of n and p for which $(n + 1)pqt > 9$ (Nicholson, 1956, pp. 473, 475).

(b) Molenaar (1970, p. 135) states that an improvement to approximations to $G(c;n,M,N)$ taken from Nicholson's bounds in (a) above is given by

$$G(c;n,M,N) \simeq \Phi[\{1 - (N + 2)^{-1}\}e^R y^*_{c+1}]$$

where R is defined in (a) above, and

$$y_k^* = \frac{d_k}{\sigma} + \frac{a}{\sigma}\left(\frac{d_k}{\sigma}\right)^2 + \frac{2a}{\sigma}$$

This approximation is superior to that derived from (a) in situations where the conditions stated in (a) hold (Molenaar, 1970, p. 135).

[*7.4.6*] A *recommended more accurate* approximation than that in [7.4.3], yet simple in form, is given by Molenaar (1970, p. 136):

Let $p^* = M/N$, $s^* = n/N$, $\tau^2 = (N - n)nM(N - M)/N^3 = N\ \mathrm{Var}(Y)/(N - 1)$, and $z_k = (k + 1/2 - np^*)/\tau$. Then

$$G(c;n,M,N) \simeq \Phi(u)$$

$$u = z_c + (z_c^2 - 1)\left[-\frac{(2s^* - 1)(2p^* - 1)}{6\tau} + z_c\frac{1 - 3s^* + 3s^{*2}}{48\tau^2}\right]$$

This approximation is recommended as being rather accurate, unless $s^* \le p^* \le 0.25$; for the latter case, see Molenaar (1970, pp. 126, 136).

[*7.4.7*] The *exact normal deviate* z defined by $G(c;n,M,N) = \Phi(z)$ satisfies

$$\begin{aligned} z = z_c &+ (2s^* - 1)(2p^* - 1)(1 - z_c^2)/(6\tau) \\ &+ [z_c^3\{5 - 14s^*(1 - s^*) - 14p^*(1 - p^*) + 38s^*(1 - s^*)p^*(1 - p^*)\} \\ &+ z_c\{-2 + 2s^*(1 - s^*) + 2p^*(1 - p^*) \\ &+ 10s^*(1 - s^*)p^*(1 - p^*)\}]/(72\tau^2) + O(\tau^{-2}) \end{aligned}$$

where $z_c = (k + 1/2 - np^*)/\tau$; the quantities p^*, s^*, and τ are as defined above in [7.4.6] (Molenaar, 1970, p. 120).

Inverting the expansion (Molenaar, 1970, p. 124),

$$\begin{aligned} z_c = z &+ (2s^* - 1)(2p^* - 1)(z^2 - 1)/(6\tau) \\ &+ [z^3\{-1 - 2s^*(1 - s^*) - 2p^*(1 - p^*) + 26s^*(1 - s^*)p^*(1 - p^*)\} \\ &+ z\{-2 + 14s^*(1 - s^*) + 14p^*(1 - p^*) \\ &- 74s^*(1 - s^*)p^*(1 - p^*)\}]/(72\tau^2) + O(\tau^{-2}) \end{aligned}$$

7.5 MISCELLANEOUS DISCRETE DISTRIBUTIONS

[*7.5.1*] *Neyman's Type A* distribution, obtained by compounding two Poisson distributions, has pf $g(k;\lambda,\eta)$ given by

$$g(k;\lambda,\eta) = \begin{cases} \sum_{j=0}^{\infty} e^{-\lambda}(\lambda^j/j!)e^{-j\eta}\{(j\eta)^k/k!\}, & k = 1, 2, \ldots \\ \exp\{-\lambda(1 - e^{-\eta})\}, & k = 0 \end{cases}$$

If λ is large and η is not too small, the standardized rv

$$(Y - \lambda\eta)/\sqrt{\lambda\eta(1 + \eta)}$$

has an approximate N(0,1) distribution, where Y has the pf above (Neyman, 1939, p. 46; Martin and Katti, 1962, pp. 355, 358-359; see also Johnson and Kotz, 1969, pp. 217, 221).

[*7.5.2*] *Absorption Distributions*. Suppose that a large number n of particles cross or fail to cross an infinite slab (in two dimensions) of width b, containing a large number M of absorption points, and with an initial expected number θ of absorption points to be encountered by each particle. Let Y be the number of absorptions. Then Y has pf $g(k;\theta,n,M)$, where

$$g(k;\theta,n,M) = \frac{q^{(M-k)(n-k)}\Pi_{j=n-k+1}^{n}(1 - q^j)\Pi_{j=M-k+1}^{M}(1 - q^j)}{\Pi_{j=1}^{k}(1 - q^j)}$$

$$q = 1 - \theta/M, \quad k = 0, 1, \ldots, \min(M,n)$$

Then the standardized rv $\{Y - (1 - q^M)(1 - q^n)\}/\sigma$ tends rapidly to normality, where (Borenius, 1953, pp. 151-157; Johnson and Kotz, 1969, pp. 267-269)

$$\sigma^2 = (1 - q^M)q^M(1 - q^n)q^n/[(1 - q)\{1 - (1 - q^M)(1 - q^n)\}^2]$$

[*7.5.3*] Let $X_1, \ldots, X_n$ be a sequence of iid random variables, usually assumed continuous, and let the rv Y be the total number of *runs* of increasing and of decreasing values in the sequence $X_1, \ldots,$

X_n. Then if n > 12, the distribution of the standardized rv $\{Y - (2n - 1)/3\}\sqrt{90/(16n - 29)}$ is approximately N(0,1). Wallis and Moore (1941, p. 405) suggest replacing Y - (2n - 1)/3 by $\mathrm{sgn}\{|Y - (2n - 1)/3| - 1/2\}$ for a continuity correction. See also Johnson and Kotz (1969, pp. 257-258).

[*7.5.4*] Let $X_1, \ldots, X_N$ be a sequence of rvs with realized values $x_1, \ldots, x_N$. The value x_j is a *record* if, for some $j \le k \le N$, x_j is the largest (or smallest) of $x_1, \ldots, x_k$. Let Y_1 be the total number of records (largest and smallest) in $X_1, \ldots, X_N$; and let

$$\begin{aligned} Y_2 = {} & \text{(total number of upper records)} \\ & - \text{(total number of lower records)} \end{aligned}$$

Then as $N \to \infty$, Y_1 and Y_2 are asymptotically independent; the asymptotic distributions of Y_1 and Y_2 are $N(\Sigma_{j=2}^N j^{-1},\ 2\Sigma_{j=2}^N j^{-1} - 4\Sigma_{j=2}^N j^{-2})$ and $N(0,\ 2\Sigma_{j=2}^N j^{-1})$, respectively. Foster and Stuart (1954, p. 6) table exact and approximate cdfs of Y_1 and Y_2 when N = 15 and N = 6, respectively; Johnson and Kotz (1969, pp. 259-260) reproduce the first of these tables and discuss the asymptotic normality of local records distributions, in which x_j is a local record over some sequence of r successive values, say.

7.6 THE BETA DISTRIBUTION

[*7.6.1*] Let the probability density function (pdf) of a beta rv Y be h(y;a,b), where

$$h(y;a,b) = y^{a-1}(1 - y)^{b-1}/B(a,b), \qquad 0 < y < 1,\ a > 0,\ b > 0$$

where B(a,b) is the beta function. Let H(y;a,b) be the cdf of Y.

[*7.6.2*] When a and b are positive integers, approximations to H(y;a,b) can be deduced from those for binomial probabilities discussed in [7.1]. Using the notation G for the binomial cdf in [7.1],

$$H(y;a,b) = G(b - 1;\ a + b - 1,\ 1 - y) = 1 - G(a - 1;\ a + b - 1,\ y)$$

(See Johnson and Kotz, 1969, p. 63.) Either of these forms give rise to approximations, the choice depending on which reduces the error by most. Using the first form, we can construct approximations to $H(y;a,b)$ from those for $G(k;n,p)$ as follows:

(i) Replace k by b - 1.

(ii) Replace n by a + b - 1, and n - k by a.

(iii) Replace p by 1 - y, and q by y.

(iv) If $G(k;n,p) \simeq \Phi(z)$, then $H(y;a,b) \simeq \Phi(z)$, where z is expressed in terms of y, a, and b from (i), (ii), and (iii).

[*7.6.3*] From [7.6.2] and Molenaar's *recommended* simple yet accurate approximation in [7.1.4], we derive the following: let y be close to 1/2 in the sense that, if a + b = 4, 31, or 301, then $|y - 1/2| \le 0.25$, 0.1, or 0.04, respectively.

Then

$$H(y;a,b) \simeq \Phi[\sqrt{(4b - 1)y} - \sqrt{(4a - 1)(1 - y)}] \quad \text{if } 0.45 \le |H(y;a,b) - 1/2| \le 0.495$$

and

$$H(y;a,b) \simeq \Phi[\sqrt{(4b - 3/2)y} - \sqrt{(4a - 3/2)(1 - y)}] \quad \text{if } 0.05 \le H(y;a,b) \le 0.93$$

If y is not close to 1/2 in the above sense, then

$$H(y;a,b) \simeq \Phi[\sqrt{4by} - \sqrt{4a(1 - y)}] \quad \text{if } 0.45 \le |H(y;a,b) - 1/2| \le 0.495$$

and

$$H(y;a,b) \simeq \Phi[\sqrt{(4b - 1)y} - \sqrt{(4a - 1)(1 - y)}] \quad \text{if } 0.05 \le H(y;a,b) \le 0.93$$

The absolute error is $O(1/\sqrt{a + b})$ if $y \ne 1/2$ and $O(1/(a + b))$ if $y = 1/2$.

[*7.6.4*] From [7.6.2] and the *Camp-Paulson* approximation of [7.1.7], the following approximation obtains:

$$H(y;a,b) \simeq \Phi\{-x/(3\sqrt{z})\}$$

$$x = \left[\frac{a(1 - y)}{by}\right]^{1/3}\left(9 - \frac{1}{a}\right) + \frac{1}{b} - 9$$

$$z = \left[\frac{a(1 - y)}{by}\right]^{2/3}\left(\frac{1}{a}\right) + \frac{1}{b}$$

the error is $O(\{(a + b)y(1 - y)\}^{-1})$ for all y, and is $O(\{(a + b)y(1 - y)\}^{-3/2})$ if $y = 1/2$ or $y = 0.042$.

[7.6.5] The Peizer and Pratt approximation can be obtained from [7.6.2] and [7.1.9]; or see Peizer and Pratt (1968, pp. 1417-1422):

$$H(y;a,b) \simeq \Phi(u), \qquad 0 < y < 1$$

$$u = d\left[1 + yT\left\{\frac{b - 1/2}{(a + b - 1)(1 - y)}\right\} + (1 - y)T\left\{\frac{a - 1/2}{(a + b - 1)y}\right\}\right]^{1/2} \Bigg/ \sqrt{y(1 - y)(a + b - 5/6)}$$

$$d = (a + b - 2/3)y - (a - 1/3) + 0.02\{yb^{-1} - (1 - y)a^{-1} + (y - 1/2)(a + b)^{-1}\}$$

$$T(x) = (1 - x^2 + 2x \log x)/(1 - x)^2, \qquad x \neq 1$$

$$T(1) = 0$$

Then

$$|H(y;a,b) - \Phi(u)| \le \begin{cases} 0.001 & \text{if } a \ge 2,\ b \ge 2 \\ 0.01 & \text{if } a \ge 1,\ b \ge 1 \end{cases}$$

Generally, the approximation is accurate to $O(1/\{(a + b)y(1 - y)\}^{3/2})$ if $y \neq 1/2$, and to $O(1/\{(a + b)y(1 - y)\}^2)$ if $y = 1/2$ (see [7.1.9]).

[7.6.6] Molenaar's *recommended more accurate* approximation of [7.1.10] adapts, with [7.6.2], to give the following:

$$H(y;a,b) \simeq \Phi\{2\sqrt{by + A} - 2\sqrt{a(1 - y) + B}\}, \qquad 0 < y < 1,\ y \neq 1/2$$

$$A = [(1 - 4y + 7y^2)\{a - 1/2 - (a + b - 1)y\}^2/\sigma] - (2 + y + 5y^2)/18$$

$$B = [(4 - 10y + 7y^2)\{a - 1/2 - (a + b - 1)y\}^2/\sigma] - (8 - 11y + 5y^2)/18$$

$$\sigma^2 = 36(a + b - 1)y(1 - y)$$

$$H(1/2;a,b) \simeq \Phi(\sqrt{2b + \beta} - \sqrt{2a + \beta})$$

$$\beta = \{(a - b)^2 - 10(a + b - 1)\}/\{12(a + b - 1)\}$$

This is accurate to $O(1/\{(a + b)y(1 - y)\}^{3/2})$ if $y \neq 1/2$, and to $O(1/\{(a + b - 1)y(1 - y)\}^2)$ if $y = 1/2$ (see [7.1.10]).

[*7.6.7*] Suppose that $(a + b - 1)(1 - y) \geq 0.8$ and $a + b > 6$. Then (Abramowitz and Stegun, 1964, p. 945)

$$H(y;a,b) = \Phi(3x/\sqrt{z}) + \varepsilon$$

$$x = (by)^{1/3}\{1 - (9b)^{-1}\} - \{a(1 - y)\}^{1/3}\{1 - (9a)^{-1}\}$$

$$z = (by)^{2/3}b^{-1} + \{a(1 - y)\}^{2/3}a^{-1}$$

$$|\varepsilon| < 5 \times 10^{-3}$$

7.7 THE VON MISES DISTRIBUTION

[*7.7.1*] A von Mises or circular normal rv Y has pdf $g(y;a)$, where

$$g(y;a) = \{2\pi I_0(a)\}^{-1} \exp(a \cos y), \quad -\pi < y < \pi,\ a > 0$$

(See [2.7.1].) Let $G(y;a)$ be the cdf of Y. Normal approximations to $G(y;a)$ improve as $a \to \infty$, so we assume that a is large.

[*7.7.2*] A number of approximations have been compared by Upton (1974, pp. 369-371), who also introduced the last [(iv) below]:

(i) $G(y;a) \simeq \Phi(y\sqrt{a})$

(ii) $G(y;a) \simeq \Phi(y\sqrt{a - 1/2})$, $\quad a > 10$

(iii) $G(y;a) \simeq \Phi[y\sqrt{a}\{1 - (8a)^{-1}\}]$

(iv) $G(y;a) \simeq \Phi[y\sqrt{a}\{1 - (8a)^{-1}\} - y^3\sqrt{a}\{1 + (4a)^{-1}\}/24]$

Based on comparisons for which $0.6 \leq G(y;a) \leq 0.99$ and $1 \leq a \leq 15$, (iii) is better than (i), (ii) has errors in the tails, to $O(a^{-1})$, but (iv) is recommended as most accurate, with 2-, 3-, or 4-decimal-place accuracy if $a \geq 3$, $a \geq 6$, or $a \geq 12$, respectively, and errors to $O(a^{-2})$.

[7.7.3] G. W. Hill (1976, p. 674) gives an approximation which improves on those in [7.7.2]. Let $w = 2\sqrt{a}\sin(y/2)$ or $y = 2 \arcsin((1/2)w/\sqrt{a})$. Then

$$G(y;a) \simeq \Phi(u),$$

$$u = w - \frac{w}{8a} - \frac{2w^3 + 7w}{128a^2} - \frac{8w^5 + 46w^3 + 177w}{3072a^3} - \cdots$$

$$w = u + \frac{u}{8a} + \frac{2u^3 + 9u}{128a^3} + \frac{8u^5 + 70u^3 + 225u}{3072a^3} + \cdots$$

[7.7.4] <u>*An Approximation Accurate to* $O(a^{-3})$</u> (G. W. Hill, 1976, pp. 674-675)

$$G(y;a) \simeq \Phi(u)$$

$$u = (y/|y|)\left[2a\left(1 - \frac{1}{4a} - \frac{3}{32a^2}\right)(1 - \cos y) \times \left\{1 - \frac{1}{16a}(1 - \cos y)\right\}\right]^{1/2}, \qquad y \neq 0$$

[7.7.5] G. W. Hill (1976, p. 675) gives an approximation which is accurate to $O(a^{-5})$, but which (unlike the others above) requires access to tables of the modified Bessel function $I_0(a)$; see G. W. Hill (1976), however, for details of a computer algorithm. If a is large,

$$I_0(a) \simeq (2\pi a)^{-1/2}e^a\left[1 + \frac{1}{8a} + \frac{1\cdot 9}{2!(8a)^2} + \frac{1\cdot 9\cdot 25}{3!(8a)^3} + \cdots\right]$$

The following form is recommended. Let $z = \{(\sqrt{2/\pi})e^a/I_0(a)\} \times \sin(y/2)$; then

$$G(y;a) \simeq \Phi(u)$$

$$u = z - z^3\left\{8a - \frac{2z^2 + 16}{3} - \frac{4z^4 + 7z^2 + 334}{96a}\right\}^{-2}$$

$$z = u + u^3\left\{8a - \frac{2u^2 + 16}{3} - \frac{4u^4 + 25u^2 + 334}{96a}\right\}^{-2}$$

7.8 THE CHI-SQUARE AND GAMMA DISTRIBUTIONS

[*7.8.1*] (a) Let the pdf of a chi-square rv Y be $g(y;\nu)$, so that

$$g(y;\nu) = \{2^{\nu/2}\Gamma(\nu/2)\}^{-1/2} y^{(1/2)\nu-1} e^{-(1/2)y}, \quad y > 0, \ \nu = 1, 2, \ldots$$

Let the cdf of Y be $G(y;\nu)$, such that $G(y_p;\nu) = 1 - p = \Phi(z_p)$ (see [5.3.1] to [5.3.3]). The distribution is that of the sum of squares of ν iid N(0,1) rvs.

(b) The most general form of the pdf of a gamma random variable X is given by $h(x;a,b,c) = \{b^a\Gamma(a)\}^{-1}(x - c)^{a-1} \exp\{-(x - c)/b\}$; $x > c$, $a > 0$, $b > 0$, where in most applications $c = 0$. Probability and other statements about the distribution can be taken from the transformation

$$Y = 2(X - c)/b$$

since Y has a χ^2_{2a} distribution. The approximations in this section are presented in the context of chi-square rvs, but this transformation can easily be used to apply the results to a general gamma distribution. Thus $\Pr(X \le x) = G(2(y - c)/b;\ 2a)$ and $y_p = 2(x_p - c)/b$, with obvious notation for quantiles x_p of X.

[*7.8.2*] A simple normal approximation:

$$G(y;\nu) \simeq \Phi\{(y - \nu)/\sqrt{2\nu}\}$$

This is not very accurate, unless the degrees of freedom ν are very large (Johnson and Kotz, 1970a, p. 176).

[*7.8.3*] *The Fisher Approximation*

$$G(y;\nu) \simeq \Phi(\sqrt{2y} - \sqrt{2\nu - 1})$$

This is an improvement over that of [7.8.2], but is not as accurate as that of [7.8.4]; it is satisfactory if $\nu > 100$ (Severo and Zelen, 1960, p. 411); see Johnson and Kotz (1970a, pp. 176-177) and Kendall and Stuart (1977, p. 402) for tables comparing the accuracies of the Fisher and Wilson-Hilferty approximations.

[7.8.4] (a) *The Wilson Hilferty Approximation* (1931, pp. 684-688)

$$G(y;\nu) \simeq \Phi[\{(y/\nu)^{1/3} - 1 + 2/(9\nu)\}\sqrt{9\nu/2}] = \Phi(x), \text{ say}$$

This accurate approximation (when $\nu > 30$) uses the fact that $(Y/\nu)^{1/3}$ is approximately a $N(1 - 2/(9\nu), 2/(9\nu))$ rv, and its superior performance arises because $(Y/\nu)^{1/3}$ tends to symmetry and to normality more rapidly than does $\sqrt{2Y}$ in [7.8.3], and $\sqrt{2Y}$ tends to normality more rapidly than does Y in [7.8.2] (Kendall and Stuart, 1977, pp. 400-401; Johnson and Kotz, 1970a, pp. 176-177). The minimum absolute error when $\nu \geq 3$ is less than 0.007, and is less than 0.001 when $\nu \geq 15$ (Mathur, 1961, pp. 103-105).

(b) In (a), replace x by $x + h_\nu$, where

$$h_\nu = -\frac{2}{27\nu}\left\{\frac{2\sqrt{2}(x^2 - 1)}{3\sqrt{\nu}} - \frac{x^3 - 3x}{4}\right\}$$

Severo and Zelen (1960, pp. 413-414) suggest approximating h_ν by $(60/\nu)h_{60}$ when $\nu \geq 30$, say; and by $(20/\nu)h_{20}$ when $5 \leq \nu \leq 20$. Values of h_{60} and h_{20} are tabulated to five decimal places for $x = -3.5(0.05)3.5$. See also Abramowitz and Stegun (1964, p. 941) for values of h_{60}. This approximation gives an improvement over that in (a).

[7.8.5] Bounds on $\Pr(Y > y)$ (Wallace, 1959, p. 1127). Let

$$W(y) = \sqrt{y - \nu - \nu \log(y/\nu)} \qquad W_1(y) = W(y) + (\sqrt{2/\nu})/3$$

Then

$$d_\nu \exp\{1/(9\nu)\}[1 - \Phi\{W_1(y)\}] < 1 - G(y;\nu)$$
$$< d_\nu[1 - \Phi\{W(y)\}]$$

$$d_\nu = \sqrt{2\pi}\exp[-(1/2)\nu][(1/2)\nu]^{(\nu-1)/2}/\Gamma[(1/2)\nu]$$

[7.8.6] *The Peizer and Pratt Approximation* (1968, pp. 1418, 1421, 1424)

$$G(y;\nu) \simeq \Phi(u)$$

$$u = d[1 + T\{(\nu - 1)/y\}]^{1/2}/\sqrt{2y}$$
$$d = y - \nu + (2/3) - 0.08/\nu$$
$$T(x) = (1 - x^2 + 2x \log x)/(1 - x)^2, \quad x \neq 1$$
$$T(1) = 0$$

Then

$$|G(y;\nu) - \Phi(u)| \le 0.001 \text{ (resp. } 0.01) \quad \text{if } \nu \ge 4 \text{ (resp. } \nu \ge 2)$$

[*7.8.7*] If Y has a χ_ν^2 distribution, then the distribution of log Y is more nearly normal than that of Y (Johnson and Kotz, 1970a, p. 181). The resulting approximation is not used very much in practice, but it is likely to be useful in the approximation to the distribution of the sample generalized variance (see [7.14.4]). Let $\psi(\cdot)$ and $\psi'(\cdot)$ denote the diagamma and trigamma functions, respectively (Abramowitz and Stegun, 1964, pp. 255, 260). Then (Johnson and Kotz, 1972, p. 198; 1969, p. 7)

$$G(y;\nu) \simeq \Phi(u)$$
$$u = \{\log y - E(\log Y)\}/\sqrt{\mathrm{Var}(\log Y)}$$
$$E(\log Y) = \log 2 + \psi[(1/2)\nu] \simeq \log(\nu - 1), \qquad \nu \ge 4$$
$$\mathrm{Var}(\log Y) = \psi'[(1/2)\nu] \simeq 2/(\nu - 1), \qquad \nu \ge 4$$

Further (Abramowitz and Stegun, 1964, pp. 255, 258, 260, 807), if γ is Euler's constant (γ = 0.5772, 1566, 49),

$$E(\log Y) = \begin{cases} \log 2 - \gamma + \{1 + 1/2 + 1/3 + \cdots \\ \qquad + [(1/2)\nu - 1]^{-1}\}, & \nu \text{ even} \\ -\log 2 - \gamma + 2\{1 + (1/3) + (1/5) \\ \qquad + \cdots + 1/(\nu - 2)\}, & \nu \text{ odd} \end{cases}$$

$$\mathrm{Var}(\log Y) = \begin{cases} (\pi^2/6) - \{1 + 2^{-2} + 3^{-2} + \cdots \\ \qquad + [(1/2)\nu - 1]^{-2}\}, & \nu \text{ even} \\ (\pi^2/2) - 4\{1 + 3^{-2} + 5^{-2} + \cdots \\ \qquad + (\nu - 2)^{-2}\}, & \nu \text{ odd} \end{cases}$$

Gnanadesikan and Gupta (1970, pp. 108-113) have investigated the log Y approximation; it improves as ν increases. The difference between the true and approximate quantiles is always less than 0.40, and for probability levels between 0.05 and 0.95 is less than 0.05. The maximum absolute error in probability when the standardized percentiles of log Y are approximated by those of a N(0,1) rv is less than 0.02 when $\nu \le 25$, decreasing as ν increases.

[*7.8.8*] Expressions for the *percentile* y_p in terms of z_p, where $G(y_p;\nu) = \Phi(z_p)$

(a) $y_p \simeq (1/2)\{z_p + \sqrt{2\nu - 1}\}^2$, from [7.8.3]

(b) $y_p \simeq \nu[z_p\sqrt{2/(9\nu)} + 1 - 2/(9\nu)]^3$, from [7.8.4]

See [7.8.9] for a discussion of accuracy.

[*7.8.9*] The Cornish-Fisher (CF) expansion (see [6.4.12]) gives for *percentiles* (Fisher and Cornish, 1960, p. 215)

$$\begin{aligned}
y_p \simeq{} & \nu + \sqrt{\nu}(z_p\sqrt{2}) + 2(z_p^2 - 1)/3 + (z_p^3 - 7z_p)/(9\sqrt{2\nu}) \\
& - (6z_p^4 + 14z_p^2 - 32)/(405\nu) \\
& + (9z_p^5 + 256z_p^3 - 433z_p)/(4860\sqrt{2}\nu^{3/2}) \\
& + (12z_p^6 - 243z_p^4 - 923z_p^2 + 1472)/(25515\nu^2) \\
& - (3753z_p^7 + 4353z_p^5 - 289517z_p^3 - 289717z_p)/(9185400\sqrt{2}\nu^{5/2}) \\
& + O(\nu^{-3})
\end{aligned}$$

Peiser (1943, pp. 56-62) and Goldberg and Levine (1946, pp. 216-225) derived approximations to the quantiles y_p which are essentially the first four and six terms, respectively, of the above CF expansion.

Sahai and Thompson (1974, pp. 86-88) and Johnson and Kotz (1970a, pp. 176-177) have compared the above with the approximations in [7.8.8]. All closely approximate y_p if $0.1 < p < 0.9$, and are better if p is close to zero than if p is close to one. If $\nu \ge 40$, the CF expansion (Peiser approximation) to four terms has an error less than 0.1 percent; the Wilson-Hilferty does better for smaller values of ν (see [7.8.8(b)]) and is always better than that of Fisher [7.8.8(a)].

If $\nu \geq 30$, the CF expansion to six terms (Goldberg-Levine) has an error in y_p no greater than 0.0001 when $0.1 \leq p \leq 0.9$. Only if ν or $1 - p$ is very small is there much need for more than six terms of the CF expansion.

[*7.8.10*] An improved Wilson-Hilferty approximation to y_p (see [7.8.4]) is given by (Severo and Zelen, 1960, pp. 411, 413)

$$y_p \simeq \nu\{(z_p - h_\nu)\sqrt{2/(9\nu)} + 1 - 2/(9\nu)\}^3$$

$$h_\nu = -\frac{2}{27\nu}\left\{\frac{2\sqrt{2}(z_p^2 - 1)}{3\sqrt{\nu}} - \frac{z_p^3 - 3z_p}{4}\right\}$$

Approximately, $h_\nu \simeq (60/\nu)h_{60}$ or $h_\nu \simeq (20/\nu)h_{20}$ under the conditions discussed in [7.8.4(b)] above.

7.9 NONCENTRAL CHI-SQUARE

[*7.9.1*] The pdf $g(y;\nu,\lambda)$ of a noncentral chi-square rv with ν degrees of freedom and noncentrality parameter λ is given in [5.3.6]; let $G(y;\nu,\lambda)$ be the cdf of Y. The distribution of Y is that of $\Sigma_{i=1}^{\nu}(X_i - \mu_i)^2$, where $X_1, \ldots, X_\nu$ are iid N(0,1) rvs, and $\lambda = \mu_1^2 + \cdots + \mu_\nu^2$.

[*7.9.2*] Two simple normal approximations, not very accurate, but having error $O(1/\sqrt{\lambda})$ as $\lambda \to \infty$, uniformly in y, are given by (Johnson, 1959, p. 353; Johnson and Kotz, 1970b, p. 141)

(a) $G(y;\nu,\lambda) \simeq \Phi[(y - \nu - \lambda)/\sqrt{2(\nu + 2\lambda)}]$

(b) $G(y;\nu,\lambda) \simeq \Phi[(y - \nu - \lambda + 1)/\sqrt{2(\nu + 2\lambda)}]$

[*7.9.3*] The following approximations to noncentral χ^2 involve first a central χ^2 approximation, and then a normal approximation to central χ^2. As one would expect, the first stage (direct central χ^2 approximation) is generally more accurate than the normality approximation superimposed on it; the reference sources contain further discussion (see also Johnson and Kotz, 1970b, pp. 139-142).

(a) Abdel-Aty (1954, p. 538) applied the Wilson-Hilferty approximation of [7.8.4(a)] to $\{Y/(\nu + \lambda)\}^{1/3}$. Thus (Abramowitz and Stegun, 1964, p. 942)

$$G(y;\nu,\lambda) \simeq \Phi(u)$$

$$u = \frac{\{y/(\nu + \lambda)\}^{1/3} - 1 + 2(\nu + 2\lambda)/\{9(\nu + \lambda)^2\}}{[2(\nu + 2\lambda)/\{9(\nu + \lambda)^2\}]^{1/2}}$$

As one would expect, this approximation does better when the noncentrality parameter is small (and the distribution of Y is more like that of central chi-square), and deteriorates as λ increases; see [7.9.5(a)].

(b) Patnaik (1949, p. 209) applied Fisher's approximation of [7.8.3] to a multiple of Y having adjusted degrees of freedom; then

$$G(y;\nu,\lambda) \simeq \Phi[\sqrt{2y(\nu + \lambda)/(\nu + 2\lambda)} - \{2(\nu + \lambda)^2/(\nu + 2\lambda) - 1\}^{1/2}]$$

From a few comparisons (Johnson, 1959, pp. 352-363), this approximation is not any better than that of [7.9.2(b)].

[*7.9.4*] (a) One of three approximations of Sankaran (1959, pp. 235-236) is accurate even when ν is small, over a wide range. Let

$$h = 1 - 2(\nu + \lambda)(\nu + 3\lambda)/\{3(\nu + 2\lambda)^2\}$$

$$= (1/3) + (2/3)\lambda^2(\nu + 2\lambda)^{-2}$$

$$G(y;\nu,\lambda) \simeq \Phi[(\{y/(\nu + \lambda)\}^h - a)/b]$$

$$a = 1 + h(h - 1)\frac{\nu + 2\lambda}{(\nu + \lambda)^2} - h(h - 1)(2 - h)(1 - 3h)\frac{(\nu + 2\lambda)^2}{2(\nu + \lambda)^4}$$

$$b = h\frac{\sqrt{2(\nu + 2\lambda)}}{\nu + \lambda}\left\{1 - (1 - h)(1 - 3h)\frac{\nu + 2\lambda}{2(\nu + \lambda)^2}\right\}$$

to terms of $O(\lambda^{-2})$. Although complicated, this approximation performs better than any so far listed. For upper and lower 5 percent points, Johnson and Kotz (1970b, p. 142) compare the absolute error for $\nu = 2$, 4, and 7, and $\lambda = 1$, 4, 16, and 25 with those of [7.9.2(b)] and [7.9.3(a)]; this error is never greater than 0.06 ($\lambda \geq 1$) or 0.03 ($\lambda \geq 4$).

(b) In the context of a wider problem (see [7.14.1]), and apparently unaware of the Sankaran approximation in (a) above, Jensen and Solomon (1972, pp. 899-901) gave the formula in (a), with terms truncated to $O(\lambda^{-1})$; that is, defining h as above

$$G(y;\nu,\lambda) \simeq \Phi[(\{y/(\nu + \lambda)\}^h - a')/b']$$

$$a' = 1 + h(h - 1)(\nu + 2\lambda)/(\nu + \lambda)^2$$

$$b' = h\sqrt{2(\nu + 2\lambda)}/(\nu + \lambda)$$

For selection of values, in which $4 \leq \nu \leq 24$ and $4 \leq \lambda \leq 24$, the absolute error in approximating $G(y;\nu,\lambda)$ is less than 0.005.

[7.9.5] If $G(y_p;\nu,\lambda) = 1 - p = \Phi(z_p)$, approximations to percentiles are as follows:

(a) $$y_p \simeq (\nu + \lambda)[z_p\sqrt{C} + 1 - C]^3$$

$$C = 2(\nu + 2\lambda)/\{9(\nu + \lambda)^2\}$$ (Abramowitz and Stegun, 1964, p. 942)

This is derived from [7.9.3(a)].

(b) From [7.9.4], we get

$$y_p \simeq (\nu + \lambda)(a + bz_p)^{1/h}$$

where a, b, and h are as defined in [7.9.4], and we may replace a and b by a' and b', respectively; see Johnson and Kotz (1970b, p. 142).

7.10 STUDENT'S t DISTRIBUTION

[7.10.1] The pdf $g(y;\nu)$ of a Student t rv Y with ν degrees of freedom is given by

$$g(y;\nu) = [\sqrt{\nu}B(1/2,\nu/2)]^{-1}(1 + t^2/\nu)^{-(\nu+1)/2}, \quad \nu = 1, 2, \ldots$$

where $B(\cdot,\cdot)$ is the beta function. Let $G(y;\nu)$ be the cdf of Y. Normal approximations to t tend to improve as ν increases.

[*7.10.2*] *Simple Normal Approximations*

(a) $G(y;\nu) \simeq \Phi(y\sqrt{1 - 2/\nu})$

This is only moderately good if $\nu < 20$ (Johnson and Kotz, 1970b, p. 101).

(b) $G(y;\nu) \simeq \Phi[y\{1 - 1/(4\nu)\}/\{1 + (1/2)y^2/\nu\}^{1/2}$

for ν large (Abramowitz and Stegun, 1964, p. 949).

[*7.10.3*] *Bounds by Wallace* (1959, pp. 1124-1125). Let $G(y;\nu) = \Phi\{z(y)\}$ and $u(y) = \{\nu \log(1 + y^2/\nu)\}^{1/2}$. Suppose that $y > 0$. Then

(a) $z(y) \le u(y), \quad \nu > 0$

(b) $z(y) \ge u(y)\{1 - 1/(2\nu)\}^{1/2}, \quad \nu > 1/2$

(c) $z(y) \ge u(y) - 0.368/\sqrt{\nu}, \quad \nu \ge 1/2$

As an approximation to $z(y)$, these bounds show that $u(y)$ has an absolute error not exceeding $0.368/\sqrt{\nu}$. Except for very large values of y, the bound in (c) is much poorer than that in (b); the maximum of the bounds in (b) and (c) is "a good approximation." See Johnson and Kotz (1970b, pp. 108-109).

As a variant of the approximation in (a), Mickey (1975, pp. 216-217) found

$$z(y) \simeq \pm\{[\nu - (1/2)] \log(1 + y^2/\nu)\}^{1/2}$$

to give good results if $\nu > 10$, and gives a further modification which may be used when ν is as small as 2.

[*7.10.4*] (a) *A Recommended Approximation* (Wallace, 1959, p. 1125). If $G(y;\nu) = \Phi\{u(y)\}$ and $u(y) = \{\nu \log(1 + y^2/\nu)\}^{1/2}$, then

$$z(y) \simeq u(y)[1 - 2\{1 - \exp(-S^2)\}^{1/2}/(8\nu + 3)], \quad y > 0$$
$$S = 0.184(8\nu + 3)/\{u(y)\sqrt{\nu}\}$$

This is amenable to calculation with a slide rule, and seems to be within 0.02 of $z(y)$ for a wide range; it is better than the approximation of Peizer and Pratt (1968, p. 1428).

(b) A simpler approximation than that of (a) is given by

$$z_1(y) \simeq u(y)(8\nu + 1)/(8\nu + 3), \qquad y > 0$$

which seems to be within 0.02 of $z(y)$ when $y^2/\nu < 5$ (Wallace, 1959, p. 1125). Johnson and Kotz (1970b, pp. 108-109) give some comparisons of (a), (b), and [7.10.3].

(c) Peizer and Pratt's approximation (1968, pp. 1427-1428) is as follows:

$$z_2(y) \simeq \pm\left(\nu - \frac{2}{3} + \frac{1}{10\nu}\right)\left\{\frac{1}{\nu - 5/6}\log(1 + y^2/\nu)\right\}^{1/2}$$

where the sign chosen is to agree with the sign of y. This is as good as the approximation in (b) when $\nu \geq 2$, and not as good when $\nu < 1.5$; it is never as accurate as that in (a).

[*7.10.5*] Bounds by Chu (1956, p. 784). If $x \geq 0$, $y \geq 0$, and $\nu \geq 3$,

$$G(y;\nu) - G(-x;\nu) \leq \sqrt{(\nu - 3/7)/(\nu - 2)}\,[\Phi(y\sqrt{1 - 2/\nu}) - \Phi(-x\sqrt{1 - 2/\nu})]$$

$$G(y;\nu) - G(-x;\nu) \geq (1 + 1/\nu)^{-1}[\Phi(y\sqrt{1 + 1/\nu}) - \Phi(-x\sqrt{1 + 1/\nu})]$$

The proportional error is less than $1/\nu$ for all $x \geq 0$ and $y \geq 0$ when $\nu \geq 8$.

[*7.10.6*] *Fisher's Expansions for the pdf and cdf of* t

$$\begin{aligned} g(y;\nu) = \phi(y)\Big[1 &+ \frac{1}{4}(y^4 - 2y^2 - 1)\nu^{-1} + \frac{1}{96}(3y^8 - 28y^6 \\ &+ 30y^4 + 12y^2 + 3)\nu^{-2} \\ &+ \frac{1}{384}(y^{12} - 22y^{10} + 113y^8 - 92y^6 - 33y^4 - 6y^2 + 15)\nu^{-3} \\ &+ \frac{1}{92{,}160}(15y^{16} - 600y^{14} + 7100y^{12} - 26616y^{10} \\ &+ 18330y^8 + 6360y^6 + 1980y^4 - 1800y^2 - 945)\nu^{-4} + \cdots\Big] \end{aligned}$$

$$G(y;\nu) = \Phi(y) - \phi(y)\left[\frac{y}{4}(y^2 + 1)\nu^{-1} + \frac{y}{96}(3y^6 - 7y^4 - 5y^2 - 3)\nu^{-2}\right.$$
$$+ \frac{y}{384}(y^{10} - 11y^8 + 14y^6 + 6y^4 - 3y^2 - 15)\nu^{-3}$$
$$+ \frac{y}{92,160}(15y^{14} - 375y^{12} + 2225y^{10} - 2141y^8 - 939y^6$$
$$\left. - 213y^4 - 915y^2 + 945)\nu^{-4} + \cdots\right]$$

If the term in ν^{-4} is omitted, the maximum absolute error in the approximation to $G(y;\nu)$ is 5×10^{-6} (Johnson and Kotz, 1970b, pp. 101-102; Fisher, 1925, pp. 109-112).

[*7.10.7*] An inverse hyperbolic sine approximation, designed to be suitable for smaller as well as larger values of ν (Anscombe, 1950, pp. 228-229) is

$$G(y;\nu) \simeq \Phi(z), \qquad z = \pm\sqrt{(2\nu - 1)/3}\,\sinh^{-1}(\sqrt{3y^2/(2\nu)})$$

[*7.10.8*] A *recommended accurate* approximation has been given by Hill (1970a, pp. 617-618). Let $G(y;\nu) = \Phi\{z(y)\}$ and $w = w(y) = \{(\nu - 1/2)\log(1 + y^2/\nu)\}^{1/2}$. Then if $y > 0$,

$$z(y) \simeq w + (w^3 + 3w)/b - (4w^7 + 33w^5 + 240w^3$$
$$+ 855w)/\{10b(b + 0.8w^4 + 100\}, \qquad b = 48(\nu - 1/2)^2$$

The maximum $|\text{error}|$ in $G(y;\nu)$ for all values of y is less than 10^{-1}, 10^{-3}, 10^{-5}, or 10^{-7} if $\nu \geq 1$, 2, 4, or 6, respectively.

[*7.10.9*] Moran (1966, pp. 225-230) has given normalizing transformations which are suitable at specific percentage points; see Johnson and Kotz (1970b, pp. 110-111), also Scott and Smith (1970, pp. 681-682).

[*7.10.10*] The Cornish-Fisher (CF) expansion (see [6.4.12]) gives for the *percentile* $t_{\nu,p}$, where $G(t_{\nu,p};\nu) = \Phi(z_p) = 1 - p$ (Fisher and Cornish, 1960, p. 216),

$$t_{\nu,p} = z_p + (z_p^3 + z_p)/(4\nu) + (5z_p^5 + 16z_p^3 + 3z_p)/(96\nu^2)$$
$$+ (3z_p^7 + 19z_p^5 + 17z_p^3 - 15z_p)/(384\nu^3)$$
$$+ (79z_p^9 + 776z_p^7 + 1482z_p^5 - 1920z_p^3 - 945z_p)/(92160\nu^4)$$
$$+ \cdots$$

Peiser (1943, pp. 56-62) and Goldberg and Levine (1946, pp. 216-225) derived approximations to $t_{\nu,p}$ which are essentially the first two and three terms, respectively, of the above CF expansion. If $\nu \geq 6$ and $|p - 1/2| < 0.49$, so that percentiles are not being approximated far out in the tails, then the first three terms of the expansion should be accurate enough for most purposes; but it is worthwhile to compute all five terms if ν is likely not to be large or if extreme tail probabilities may be involved (Sahai and Thompson, 1974, p. 83). See also Hill (1970b, pp. 619-620) for a CF type expansion.

[*7.10.11*] Prescott (1974, pp. 178-180) compared inverse transformations of several discussed so far. The closest to the CF expansion above, and the one which gives the best approximation of those which he compared, is the inversion of [7.10.4(b)], Wallace's transformation. This gives for the percentile $t_{\nu,p}$, where $t_{\nu,p} > 0$,

$$t_{\nu,p} \simeq \sqrt{\nu}\{\exp(z_p^2 b^2/\nu) - 1\}^{1/2}, \qquad b = (8\nu + 3)/(8\nu + 1)$$

The scaling factor b makes this a good approximation to $t_{\nu,p}$ over most of the distribution, including small values of ν and large values of z.

7.11 NONCENTRAL t

[*7.11.1*] The distribution of noncentral t is given in [5.4.7]. Let the pdf be $g(y;\nu,\lambda)$ and the cdf $G(y;\nu,\delta)$, where ν is the degrees of freedom and δ the noncentrality parameter. The distribution can be represented as that of $(Z + \delta)/(\chi_\nu/\sqrt{\nu})$, where Z is a N(0,1) rv and χ_ν a chi rv with ν df, distributed independently of Z.

[*7.11.2*] Let Y have a noncentral t distribution, represented as in [7.11.1]. Then $G(y;\nu,\delta) = \Pr(Z - y\sqrt{\nu}\chi_\nu \le -\delta)$, and many approximations are based on the approximation to normality of the rv $Z - y\sqrt{\nu}\chi_\nu$ (Johnson and Kotz, 1970b, p. 207). Thus

$$G(y;\nu,\delta) \simeq \Phi(u)$$
$$u = \{-\delta + yE(\chi_\nu)/\sqrt{\nu}\}/\{1 + y^2 \operatorname{Var}(\chi_\nu)/\nu\}^{1/2}$$

where $E(\chi_\nu)$ and $\operatorname{Var}(\chi_\nu)$ are the mean and variance, respectively, of χ_ν, and $\operatorname{Var}(\chi_\nu) = \nu - \{E(\chi_\nu)\}^2$. The approximations following have been or may be used in the above.

(a) $E(\chi_\nu) \simeq \sqrt{\nu}$ $\quad \operatorname{Var}(\chi_\nu) \simeq 1/2$ (Johnson and Kotz, 1970b, p. 207).

(b) $E(\chi_\nu) \simeq \sqrt{\nu}\{1 - 1/(4\nu)\}$ $\quad \operatorname{Var}(\chi_\nu) \simeq 1/2$ (Abramowitz and Stegun, 1964, p. 949).

(c) $E(\chi_\nu) \simeq \sqrt{\nu}\{1 - 1/(4\nu) + 1/(32\nu^2)\}$ $\quad \operatorname{Var}(\chi_\nu) \simeq 1/2 + 1/(8\nu)$ This approximation results from taking further terms in series expansions for $E(\chi_\nu)$; see [5.3.5], or Read (1973, pp. 183-185).

(d) We have exactly that

$$E(\chi_\nu) = \sqrt{2}\Gamma\{(1/2)(\nu + 1)\}/\Gamma[(1/2)\nu]$$

Tables of $E(\chi_\nu/\sqrt{\nu})$ can be used to substitute appropriate values; sources are given in [5.3.5], where $E(S/\sigma)$ and $\operatorname{Var}(S/\sigma)$ can be replaced by $E(\chi_\nu/\sqrt{\nu})$ and $\operatorname{Var}(\chi_\nu/\sqrt{\nu})$, respectively.

While these approximations have not all been compared, it seems plausible that those which give $E(\chi_\nu)$ and $\operatorname{Var}(\chi_\nu)$ most accurately will perform best. All improve as $\nu \to \infty$. For some cautionary remarks concerning percentiles, however, see [7.11.3] and Johnson and Kotz (1970b, pp. 207-208).

[*7.11.3*] If $G(y_p;\nu,\delta) = 1 - p = \Phi(z_p)$, we obtain the following approximation to percentiles y_p of the noncentral t distribution from the approach used in [7.11.2]: putting $b_\nu = E(\chi_\nu/\sqrt{\nu})$ (Johnson and Kotz, 1970b, p. 207),

$$y_p \approx \frac{\delta b_\nu + z_p\{b_\nu^2 + (\delta^2 - z_p^2)(1 - b_\nu^2)\}^{1/2}}{b_\nu^2 - z_p^2(1 - b_\nu^2)}$$

Various approximations to y_p arise from the substitutions for b_ν given in (a), (b), and (c), as well as from (d) of [7.11.2], and the remarks given there regarding accuracy apply here also. However, the range of values of p for which real approximations to y_p obtain is restricted; we must have

$$z_p^2 < \delta^2 + b_\nu^2/(1 - b_\nu^2)$$

with analogous inequalities if (a), (b), or (c) of [7.11.2] is applied. The greatest range of values of p is not necessarily given by the most accurate approximation; (a) of [7.11.2] gives a wider range than (d), for example; see Johnson and Kotz (1970b, pp. 207-208)

[*7.11.4*] Johnson and Kotz (1970b, pp. 208-209) give expressions based on Cornish-Fisher expansions for percentiles y_p.

If $p > 0.5$ and $z = z_p$,

$$\begin{aligned} y_p = {} & z + \delta + \{z^3 + z + (2z^2 + 1)\delta + z\delta^2\}/(4\nu) \\ & + \{5z^5 + 16z^3 + 3z + 3(4z^4 + 12z^2 + 1)\delta + 6(z^3 + 4z)\delta^2 \\ & - 4(z^2 - 1)\delta^3 - 3z\delta^4\}/(96\nu^2) + O(\nu^{-3}) \end{aligned}$$

and if $p < 0.5$,

$$\begin{aligned} y_p = {} & t_{\nu,p} + \delta + \delta(1 + 2z^2 + z\delta)/(4\nu) + \delta\{3(4z^4 + 12z^2 + 1) \\ & + 6(z^3 + 4z)\delta - 4(z^2 - 1)\delta^2 - 3z\delta^3\}/(96\nu^2) + O(\nu^{-3}) \end{aligned}$$

where $t_{\nu,p}$ is the corresponding percentile of a Student t rv having ν degrees of freedom. If values of $t_{\nu,p}$ are unavailable, the first expression may be used when $p < 0.5$.

7.12 THE F DISTRIBUTION

[*7.12.1*] The F distribution is given in [5.5.1]. A random variable Y with an $F_{m,n}$ distribution can be represented as $(\chi_m^2/m)/(\chi_n^2/n)$, where the χ_m^2 and χ_n^2 variables are independent. Let G(y;m,n) be the cdf of Y.

[*7.12.2*] The distribution of z, where $z = (1/2)\log Y$, is more nearly normal than that of Y (Johnson and Kotz, 1970b, p. 81), and early investigations by Fisher (1924, pp. 805-813) were made of the properties of z rather than those of Y.

If m and n are both large, then (Wishart, 1947, pp. 172, 174; Kendall and Stuart, 1977, p. 407)

$$E(z) \simeq \frac{1}{2}(n^{-1} - m^{-1}) + (n^{-2} - m^{-2})/6$$

$$\mathrm{Var}(z) \simeq \frac{1}{2}(m^{-1} + n^{-1}) + \frac{1}{2}(m^{-2} + n^{-2}) + (m^{-3} + n^{-3})/3$$

For special expressions for the higher cumulants of z, see Wishart (1947) or Johnson and Kotz (1970b, pp. 78-80).

[*7.12.3*] If $z = \frac{1}{2}\log Y$,

$$G(y;m,n) \simeq \Phi(u)$$

$$u \simeq \left\{\frac{1}{2}\log y - E(z)\right\}/\sqrt{\mathrm{Var}(z)}$$

This approximation is good when m and n are both large; the moments of z may be approximated as in [7.12.2], with the first terms only for the simplest approximation, and the fuller expressions giving more accuracy.

[*7.12.4*] The Wilson-Hilferty approximation of [7.8.4] leads to the property that the rv

$$U_1 = \left\{Y^{1/3}\left(1 - \frac{2}{9n}\right) - \left(1 - \frac{2}{9m}\right)\right\}\Big/\{(2/9m) + 2Y^{2/3}/(9n)\}^{1/2}$$

has an approximate N(0,1) distribution (Paulson, 1942, pp. 233-235). This should only be used if $n \ge 3$ and for lower tail probabilities if $m \ge 3$ also; it is quite accurate if $n \ge 10$.

Let $h_1 = 2/(9m)$ and $h_2 = 2/(9n)$. Then the function

$$u = \frac{(1 - h_2)x - (1 - h_1)}{(h_2x^2 + h_1)^{1/2}}$$

is monotonic increasing when $x > 0$, so that, putting $x = y^{1/3}$, $Y \le y$ $(y > 0)$ corresponds to $U_1 \le u$, and $G(y;m,n) \approx \Phi(u)$. See also [7.12.8], Kendall and Stuart (1977, pp. 408-409), and Johnson and Kotz (1970b, pp. 82-83).

[*7.12.5*] The Fisher transformation of [7.8.3] leads to the property that the rv (Laubscher, 1960, p. 1111; Abramowitz and Stegun, 1964, p. 947)

$$U_2 = (\sqrt{2n - 1}\sqrt{mY/n} - \sqrt{2m - 1})/\sqrt{(mY/n) + 1}$$

has approximately a $N(0,1)$ distribution. Then

$$G(y;m,n) \simeq \Phi(u)$$

$$u = (\sqrt{2n - 1}\sqrt{my/n} - \sqrt{2m - 1})/\sqrt{(my/n) + 1}$$

[*7.12.6*] A *recommended* accurate approximation is that of *Peizer and Pratt* (1968, pp. 1416-1423, 1427). When $m \neq n$,

$$G(y;m,n) \simeq \Phi(x)$$

$$x = d\left[\left\{1 + qT\left(\frac{n - 1}{p(m + n - 2)}\right) + pT\left(\frac{m - 1}{q(m + n - 2)}\right)\right\}\Big/ \left\{pq\left(\frac{1}{2}m + \frac{1}{2}n - 5/6\right)\right\}\right]^{1/2}$$

$$p = n/(my + n), \qquad q = 1 - p$$

$$d = \frac{n}{2} - \frac{1}{3} - \left(\frac{m + n}{2} - \frac{2}{3}\right)p + \varepsilon\left(\frac{q}{n} - \frac{p}{m} + \frac{q - \frac{1}{2}}{m + n}\right)$$

where $\varepsilon = 0$ or $\varepsilon = 0.04$;

$$T(x) = (1 - x^2 + 2x \log x)/(1 - x)^2, \qquad x \neq 1$$

$$T(1) = 0$$

The simpler computation is given by $\varepsilon = 0$; a slight improvement in accuracy occurs when $\varepsilon = 0.04$.

When m = n, the following simplification holds:

$$x = \pm\left(n - \frac{2}{3} + \frac{0.1}{n}\right)\left[\frac{-\log\{y/(y + 1)^2\}}{n - 5/6}\right]^{1/2}$$

where the sign should agree with the sign of $1/2 - 1/(y + 1)$.

The absolute error when $\varepsilon = 0.04$ satisfies

$$|G(y;m,n) - \Phi(x)| < \begin{cases} 0.001 & \text{if } m,n \geq 4 \\ 0.01 & \text{if } m,n \geq 2 \end{cases}$$

[*7.12.7*] Expressions for the percentiles of F are of some interest. Let $G(y_p;m,n) = 1 - p = \Phi(z_p)$. Then from properties discussed in [7.12.2, 3], where percentiles of Fisher's z transform are denoted by w_p, $w_p = 1/2 \log y_p \simeq E(z) + z_p\sqrt{\text{Var}(z)}$ (Johnson and Kotz, 1970b, p. 81). The approximate expressions in [7.12.2] for E(z) and Var(z) may be used when m and n are large, taking the first terms only, or the full expressions for better accuracy; then $y_p = \exp(2w_p)$.

[*7.12.8*] Working from Cornish-Fisher-type expansions (see [7.12.10]), Fisher (1924, pp. 805-813) derived the simple approximation in (a) following. Successive improvements in accuracy resulted from modifications by Cochran (1940, pp. 93-95) and by Carter (1947, pp. 356-357) given in (b) and (c) following. The best of these approximations is that of Carter in (c).

(a) $$w_p = \frac{1}{2}\log y_p \simeq z_p/\sqrt{(h - 1)} - (m^{-1} - n^{-1})(\lambda - 1/6)$$

$$\lambda = (z_p^2 + 3)/6 \qquad 2h^{-1} = m^{-1} + n^{-1} \qquad \text{(Fisher, 1924)}$$

(b) $$w_p = \frac{1}{2}\log y_p \simeq z_p/\sqrt{(h - \lambda)} - (m^{-1} - n^{-1})(\lambda - 1/6)$$

with the notation of (a) (Cochran, 1940; Kendall and Stuart, 1977, pp. 409-410; Johnson and Kotz, 1970b, p. 82).

(c) $$w_p = \frac{1}{2}\log y_p \simeq z_p h'^{-1}\sqrt{(h' + \lambda')} - d'[\lambda' + (5/6) - 2/(3h')]$$

$$\lambda' = (z_p^2 - 3)/6 \qquad 2h'^{-1} = (m - 1)^{-1} + (n - 1)^{-1}$$

$$d' = (m - 1)^{-1} - (n - 1)^{-1}$$ (Carter, 1947; Johnson and Kotz, 1970b)

This approximation compares favorably with that discussed next.

[*7.12.9*] From the approximation of Paulson (1942, pp. 233-235) and the monotonicity property discussed in [7.12.4] above,

$$y_p^{1/3} \simeq \frac{(1 - h_1)(1 - h_2) \pm [z_p^2\{(h_1 + h_2)(1 + h_1 h_2) - h_1 h_2 (z_p^2 + 4)\}]^{1/2}}{(1 - h_2)^2 - z_p^2 h_2}$$

$$h_1 = 2/(9m) \qquad h_2 = 2/(9n)$$

The sign should be chosen so that $y_p^{1/3} < (1 - h_1)/(1 - h_2)$ if $z_p < 0$ and $y_p^{1/3} > (1 - h_1)/(1 - h_2)$ if $z_p > 0$; see Ashby (1968, p. 209). This approximation should be used only if $n \geq 3$, and for lower tail probabilities if $m \geq 3$ also; it is quite accurate if $n \geq 10$; see also the sources listed in [7.12.4].

Ashby (1968, p. 209) has given improved approximations for $n \leq 10$. If y_p is obtained as above, then the linear relation

$$y_p' = ky_p + c$$

gives more accuracy; Ashby has tabled empirical values of k and c for n = 1(1)10 at the upper 5, 1, and 0.1 percent tail probabilities.

[*7.12.10*] Using approximations to the cumulants of $z = \frac{1}{2} \log Y$ when m and n are large, Fisher and Cornish (1960, pp. 209-225) gave an expansion for w_p (see [7.12.7]) which appears to be better than that of [7.12.9] if $n \leq 5$, and frequently better if $n \geq 10$, for upper tail probabilities (see Sahai and Thompson, 1974, pp. 89-91): If $b = m^{-1} + n^{-1}$, $d = m^{-1} - n^{-1}$, and $z = z_p$,

$$\begin{aligned} w_p = \frac{1}{2} \log y_p \simeq{}& z\sqrt{b/2} - d(z^2 + 2)/6 + \sqrt{\frac{b}{2}}\left\{\frac{b}{24}(z^3 + 3z) + \frac{d^2}{72b}(z^3 + 11z)\right\} - \frac{bd}{120}(z^4 + 9z^2 + 8) + \frac{d^3}{(3240)b}(3z^4 + 7z^2 - 16) \\ &+ \sqrt{\frac{b}{2}}\left\{\frac{b^2}{1920}(z^5 + 20z^3 + 15z) + \frac{d^2}{2880}(z^5 + 44z^3 + 183z) + \frac{d^4}{(155{,}520)b^2}(9z^5 - 284z^3 - 1513z)\right\} \end{aligned}$$

7.13 NONCENTRAL F

[*7.13.1*] The distribution of noncentral F is given in [5.5.4]. Let the pdf and cdf be $g(y;m,n,\lambda)$ and $G(y;m,n,\lambda)$, respectively. A rv Y with an $F_{m,n}(\lambda)$ distribution can be represented as $\{\chi_m^2(\lambda)/m\}/(\chi_n^2/n)$, where the $\chi_m^2(\lambda)$ and χ_n^2 variables are independent.

[*7.13.2*] (a) Analogous to the discussion in [7.12.4], the rv

$$U_1 = \{(1 - h_2)(aY)^{1/3} - (1 - h_1)\}/\{h_2(aY)^{2/3} + h_1\}^{1/2}$$

has an approximate N(0,1) distribution, where

$$a = m/(m + \lambda) \qquad h_1 = 2(m + 2\lambda)/\{9(m + \lambda)^2\} \qquad h_2 = 2/(9n)$$

Then (Severo and Zelen, 1960, p. 416; Laubscher, 1960, p. 1111; Abramowitz and Stegun, 1964, p. 948; Johnson and Kotz, 1970b, p. 83; see (b) below)

$$G(y;m,n,\lambda) \simeq \Phi(u)$$
$$u = \{(1 - h_2)(ay)^{1/3} - (1 - h_1)\}/\{h_2(ay)^{2/3} + h_1\}^{1/2}$$

(b) Analogous to the discussion in [7.12.5], the rv

$$U_2 = \{\sqrt{2n - 1}\sqrt{mY/n} - \sqrt{2(m + \lambda) - b}\}/\sqrt{(mY/n) + b}$$

has an approximate N(0,1) distribution, where

$$b = (m + 2\lambda)/(m + \lambda)$$

Then (Laubscher, 1960, p. 1111)

$$G(y;m,n,\lambda) \simeq \Phi(u)$$
$$u = \{\sqrt{2n - 1}\sqrt{my/n} - \sqrt{2(m + \lambda) - b}\}/\sqrt{(my/n) + b}$$

Laubscher compared the approximations in (a) and (b) for a few values of λ and y, and for limited choices of m and n ($3 \le m \le 8$, $10 \le n \le 30$). Both were accurate to two decimal places; U_2 performed better than U_1 for the values tabled, although not uniformly so.

[*7.13.3*] Using the first few terms of an Edgeworth expansion ([6.4]), Mudholkar et al. (1976), pp. 353, 357) give an approximation which numerical studies indicate to be more accurate than those in [7.13.2]. Thus

$$G(y;m,n,\lambda) \simeq \Phi(x) - \phi(x)[\beta_1(x^2 - 1)/6 + \beta_2(x^3 - 3x)/24 + \beta_1^2(x^5 - 10x^3 + 15x)/72]$$

$$x = -\kappa_1/\sqrt{\kappa_2} \qquad \beta_1 = \kappa_3/\kappa_2^{3/2} \qquad \beta_2 = \kappa_4/\kappa_2^2$$

and κ_1, κ_2, κ_3, and κ_4 are the first four cumulants of the rv $\{\chi_m^2(\lambda)/m\}^{1/3} - y^{1/3}(\chi_n^2/n)^{1/3}$; here $\chi_m^2(\lambda)$ and χ_n^2 are independent noncentral and central chi-square rvs with m and n degrees of freedom, respectively. Mudholkar et al. (1976) give approximate expressions for κ_1, κ_2, κ_3, and κ_4; these are rather cumbersome, but the authors provide a Fortran routine.

7.14 MISCELLANEOUS CONTINUOUS DISTRIBUTIONS

[*7.14.1*] *Quadratic Forms*. Let $X_1, \ldots, X_k$ be iid N(0,1) rvs and $c_1, \ldots, c_k, a_1, \ldots, a_k$ be bounded constants such that $c_j > 0$; $j = 1, \ldots, k$. Define

$$Q_k = \sum_{j=1}^{k} c_j(X_j + a_j)^2$$

Further, if

$$\theta_s = \sum_{j=1}^{k} c_j^s(1 + sa_j^2), \qquad s = 1, 2, \ldots$$

the sth cumulant of the distribution of Q_k is $2^{s-1}(s - 1)!\theta_s$. Using a Wilson-Hilferty-type transformation, Jensen and Solomon (1972, pp. 898-900) take the rv Z to be an approximate N(0,1) variable, where

$$Z = \theta_1[(Q_k/\theta_1)^h - 1 - \theta_2 h(h - 1)/\theta_1^2]/(2\theta_2 h^2)^{1/2}$$

$$h = 1 - 2\theta_1\theta_3/(3\theta_2)^2$$

This approximation compares well with its (nonnormal) competitors.

For cases in which $a_1 = \cdots = a_k = 0$ and $k = 2, 3, 4, 5$, the approximation to $\Pr(Q_k \le t)$ tends to improve as t increases, and as variation among the parameters $c_1, \ldots, c_k$ decreases (Jensen and Solomon, 1972, pp. 901-902). The case in which $c_1 = \cdots = c_k = 1$ reduces to the approximation for noncentral chi-square in [7.9.4(b)], and, if in addition $a_1 = \cdots = a_k = 0$, it yields the Wilson-Hilferty approximation for chi-square in [7.8.4(a)].

For further discussion of the distribution of quadratic forms, see Johnson and Kotz (1970b, chap. 29).

[*7.14.2*] *Distance Distributions*. Let $(X_1, X_2, \ldots, X_p)$ and $(Y_1, Y_2, \ldots, Y_p)$ be two random points in a p-dimensional sphere of unit radius, and let

$$R = \left\{ \sum_{i=1}^{p} (X_i - Y_i)^2 \right\}^{1/2}$$

denote the distance between them. Then when p is large, R has an approximate $N(\sqrt{2}, (2p)^{-1})$ distribution (Johnson and Kotz, 1970b, p. 267).

[*7.14.3*] The Birnbaum-Saunders distribution (Birnbaum and Saunders, 1969, pp. 319-327) has been described in [5.1.6] as that of a rv $\theta[U\sigma + \{U^2\sigma^2 + 1\}^{1/2}]^2$, where U is a N(0,1) rv.

As $\sigma \to 0$, the distribution tends to normality (Johnson and Kotz, 1970b, p. 269).

[*7.14.4*] Let $(n - 1)^{-1} \underset{\sim}{S}$ be the sample variance-covariance matrix in a sample of size n from a multivariate normal population of dimension m, with mean vector $\underset{\sim}{\mu}$ and population variance-covariance matrix $\underset{\sim}{V}$, assumed nonsingular; suppose that $n > m$. The determinant $|(n - 1)^{-1}\underset{\sim}{S}|$ is the sample *generalized variance*, and $|\underset{\sim}{V}|$ the population generalized variance.

(a) Then $X = |\underset{\sim}{S}|/|\underset{\sim}{V}|$ has the distribution of $\Pi_{j=1}^{m} Y_{n-j}$, where $Y_{n-j} \sim \chi^2_{n-j}$ and $Y_{n-1}, Y_{n-2}, \ldots, Y_{n-m}$ are mutually independent (Johnson and Kotz, 1972, pp. 197, 198). The distribution of log X can thus be approximated by treating $\Sigma_{j=1}^{m} \log Y_{n-j}$ as a normal rv (see [7.8.7]), with mean $\Sigma_{j=1}^m E(\log Y_{n-j})$ and variance $\Sigma_{j=1}^m \text{Var}(\log Y_{n-j})$. In the formulas given in [7.8.7], exact expressions for the mean and variance of $\log Y_{n-j}$ are obtained if ν is replaced by $n - j$. If $n - m \geq 4$,

$$E[\log X] = m \log 2 + \sum_{j=1}^{m} \psi\left(\tfrac{1}{2}(n - j)\right) \simeq \sum_{j=1}^{m} \log(n - j - 1)$$

$$\text{Var}(\log X) = \sum_{j=1}^{m} \psi'\left(\tfrac{1}{2}(n - j)\right) \simeq 2 \sum_{j=1}^{m} (n - j - 1)^{-1}$$

where $\psi(\cdot)$ and $\psi'(\cdot)$ are defined in [7.8.7]. Gnanadesikan and Gupta (1970, p. 113) found that the normal approximation to log X improves with both m and n.

(b) Anderson (1958, p. 173) shows that as $n \to \infty$, $\{(n - 1)^{-[m-(1/2)]}|\underset{\sim}{S}|/|\underset{\sim}{V}|\} - \sqrt{n - 1}$ has asymptotically a N(0,2m) distribution, leading to a second normal approximation. Note that Anderson's $\underset{\sim}{S}$ is the equivalent of $(n - 1)^{-1}\underset{\sim}{S}$ as defined here; also that his n has the meaning of n - 1 as defined here.

7.15 NORMALIZING TRANSFORMATIONS

[*7.15.1*] Transformations of rvs in statistics are directed toward three purposes, frequently in the context of the general linear model and regression problems;

(i) Additivity of main effects and removal of interactions

(ii) Variance stabilizing, to make $\sigma^2(X)$ functionally free of both the mean E(X) and the sample size n

(iii) Normality of the observations

A transformation which achieves or nearly achieves one of these aims frequently helps toward the achievement of one or both of the others, but only by not seeking optimal achievement in more than one direction (Kendall and Stuart, 1966, p. 93; Hoyle, 1973, pp. 203-204). Thus, variance stabilizing may normalize approximately, but not optimally.

That a transformation leads to normality is usually inferred when it alters the skewness and kurtosis in that direction. More is required for normality than knowledge of suitable values of the first four moments, but there appear to be no important occurring cases in which small values of $|\mu_3/\mu_2^{3/2}|$ and of $|(\mu_4/\mu_2^2) - 3|$ give a misleading picture of normal approximation (Kendall and Stuart, 1966, p. 94).

Certain transformations are listed elsewhere in this book. Variance-stabilizing transformations which also tend to normalize include the angular transformation for binomial variables ([7.1.6] and [7.15.4] below), the logarithmic transformation for chi-square and gamma rvs of [7.8.7], and the inverse tanh transformation for a bivariate normal sample correlation coefficient in [10.5.8]. Transformations directed toward normality do not necessarily stabilize the variance; those considered elsewhere include the cube root transformation of chi-square in [7.8.4(a)], the power transformation for general quadratic forms in [7.14.1], and the Cornish-Fisher expansions (see [6.4.12] to [6.4.14], where z_p and x_p may be replaced by z and x, respectively). Transformations of ordered observations to normal scores in a random sample from an unknown distribution are discussed in [8.2.11].

Variance-stabilizing transformations are given further below in [7.15.2] to [7.15.6], and normalizing transformations in [7.15.6] to [7.15.8]. For further discussion, see Hoyle (1973, pp. 203-223).

[7.15.2] If T_ν is a statistic such that $\sqrt{\nu}(T_\nu - \theta)$ has asymptotically a $N(0,\sigma^2(\theta))$ distribution as $\nu \to \infty$, then for suitable single-valued differentiable functions $g(\cdot)$, $\sqrt{\nu}[g(T_\nu) - g(\theta)]/[g'(t_\nu)\sigma(\theta)]$ has asymptotically a N(0,1) distribution ([6.1.9]). A form of this result by Laubscher (1960, p. 1105) shows how stabilizing the variance may sometimes lead to asymptotic normality: the rv Y_ν has a variance asymptotically stabilized at c^2, where

$$Y_\nu = c \int_k^{\sqrt{\nu}T_\nu} \{\sigma(\theta)^{-1}\}\, d\theta$$

See also Kendall and Stuart (1977, p. 425). Further improvements can often be made (Anscombe, 1948, pp. 246-254; Hotelling, 1953, pp. 193-232) if the variance of g(X) is not quite constant.

[*7.15.3*] <u>*Square Root Transformations*</u>. (a) Let X have a Poisson distribution with mean λ. If, for arbitrary α,

$$g(x) = \begin{cases} \sqrt{x + \alpha}, & x \geq -\alpha \\ 0, & x < -\alpha \end{cases}$$

then $g(X) - \sqrt{\lambda + \alpha}$ converges to a $N(0,1/4)$ distribution as $\lambda \to \infty$ (Curtiss, 1943, pp. 113-114). Anscombe (1948, pp. 246-254) found that $\alpha = 3/8$ is optimal for variance stabilizing, except for small values of λ, with variance $\{1 + 1/(16\lambda^2)\}/4$. Freeman and Tukey (1950, pp. 607-611) found that the *chordal transformation*

$$h(X) = \sqrt{X} + \sqrt{X + 1}$$

stabilizes the variance over a larger range of values of λ; see Hoyle (1973, pp. 207-208).

(b) Let X have a gamma distribution with pdf proportional to $x^{(1/2)\nu-1}e^{-hx}$, $x > 0$. If, for arbitrary α,

$$g(x) = \begin{cases} \sqrt{x + \alpha}, & x \geq -\alpha \\ 0, & x < -\alpha \end{cases}$$

then $g(X) - \sqrt{\alpha + [(1/2)\nu/h]}$ converges to a $N(0,(4h)^{-1})$ distribution as $\nu \to \infty$ (Curtiss, 1943, p. 115). The case in which $\alpha = 0$ is related to the Fisher approximation of [7.8.3].

[*7.15.4*] <u>*Inverse Transformations*</u>. The arctanh transformation is discussed in [10.5.8]; it can arise as a variance-stabilizing device.

(a) *Angular* transformations of binomial rvs were introduced as approximations in [7.1.6]. From a variance-stabilizing point of view,

$$g(x) = \begin{cases} \arcsin\{(4n)^{-1/2}\}, & x = 0 \\ \sqrt{n + 1/2}\arcsin[\{(x + 3/8)/(n + 3/4)\}^{1/2}], & x = 1, \ldots, n - 1 \\ (1/2)\pi - \arcsin\{(4n)^{-1/2}\}, & x = n \end{cases}$$

gives a variance of $1/4 + O(n^{-2})$, the remainder tending faster to zero as $n \to \infty$ than in other such transformations of this kind (Anscombe, 1948, pp. 246-254; Hoyle, 1973, p. 210).

The angular transformation should not be confused with the less suitable *arcsin* transformation (Hoyle, 1973, p. 209)

$$h(x) = \arcsin x$$

Curtiss (1943, pp. 116, 117) gives several limit theorems for transformations of binomial rvs. Let X be a binomial rv based on n trials with mean np. Let

$$g(x) = \begin{cases} \sqrt{n} \arcsin[\{(x + \alpha)/n\}^{1/2}], & -\alpha \le x \le 1 - \alpha \\ 0 & x < -\alpha \text{ or } x > 1 - \alpha \end{cases}$$

Then $g(X) - \sqrt{n} \arcsin \sqrt{p + \alpha/n}$ has an asymptotic $N(0,1/4)$ distribution as $n \to \infty$. Other results he states give analogous properties for $\sqrt{n}\ \text{arcsinh} \sqrt{X/n}$, $\sqrt{n} \log(X/n)$, and $(1/2)\sqrt{n} \log[X/(n - X)]$.

(b) Let X be a negative binomial rv with pf as given in [7.3.1]. The transformation defined by

$$g(x) = \text{arcsinh}[\{(x + 3/8)/(s - 3/4)\}^{1/2}]$$

stabilizes the variance at $(1/4) + O(s^{-2})$ (Anscombe, 1948; Hoyle, 1973, p. 210).

(c) The inverse sinh transformation $g(X) = \alpha\{\text{arcsinh}(\beta X) - \text{arcsinh}(\beta\mu)\}$ is also used when X has a noncentral $t_\nu(\delta)$ distribution [5.4.7; 7.11.1], where (Laubscher, 1960, pp. 1106, 1107) $\mu = \delta\sqrt{(1/2)\nu}\,\Gamma\{(\nu - 1)/2\}/\Gamma[(1/2)\nu]$,

$$\alpha = b^{-1} \qquad \beta = b\sqrt{(\nu - 2)/\nu}$$

$$b = \left[\frac{2\{\Gamma[(1/2)\nu]\}^2}{(\nu - 2)[\{\Gamma(\nu - 1)/2)\}^2 - 1]}\right]^{1/2}, \qquad \nu \ge 4$$

The variable $g(X)$ approximates to a $N(0,1)$ rv, but the accuracy deteriorates if ν is small and δ is large simultaneously; larger values of ν improve the approximation.

See also Azorín (1953, pp. 173-198, 307-337) for simpler transformations $\sqrt{c\nu}$ arcsinh$(X/\sqrt{c\nu}) - \delta$, where c = 1 or c = 2/3.

(d) Let X have a noncentral $F_{m,n}(\lambda)$ distribution [5.5.4; 7.13.1]; if $\lambda \geq 4$, the transformation

$$g(X) = \sqrt{(1/2)n - 2}\ \text{arccosh}[a(X + n/m)]$$

$$a = (m/n)\{(n - 2)/(m + n - 2)\}^{1/2}$$

stabilizes the variance, but does not give such a satisfactory normalizing approximation as those in [7.13.2] (Laubscher, 1960, pp. 1109-1111).

[7.15.5] *Other Variance-Stabilizing Transformations*

(a) If E(X) = θ and Var(X) $\propto \theta^4$, then the appropriate transformation is given by (Hoyle, 1973, p. 207)

$$g(X) \propto X^{-1}$$

This may reduce extreme skewness.

(b) If E(X) = θ and Var(X) $\propto \theta^2$, the logarithmic transformation

$$g(X) = \log(X + a)$$

may be considered; it also reduces skewness, but to a lesser extent than in (b) (Hoyle, 1973, p. 207). If X has a normal distribution, however, g(X) is a lognormal rv ([2.3.2]).

[7.15.6] *Power Transformations*. Let X be a rv; consider the class of transformations defined by

$$g(x) = (x + a)^b, \qquad b \neq 0$$

(The transformation corresponding to b = 0 is logarithmic, as in [7.15.5(b)].) Box and Cox (1964, pp. 211-252) have developed a computer routine to estimate b so that Y = g(X) satisfies all three criteria, (i), (ii), and (iii), of [7.15.1]; see also the graphic approach of Draper and Hunter (1969, pp. 23-40).

Moore (1957, pp. 237-246) treated g(·) as a normalizing transformation only, where 0 < b < 1.

[7.15.7] *The Johnson System of Curves*. Let X be a rv and

$$g(x) = \gamma + \delta h\{(x - \alpha)/\beta\}$$

where γ, δ, α, and β are parameters, and $h(\cdot)$ is monotonic. Johnson (1949, pp. 149-176) developed a system of frequency curves such that $Y = g(X)$ is normally distributed or approximately so; these have the property that two rvs X_1 and X_2 having the same skewness and kurtosis lead to a unique curve in the system.

Three families comprise the system: if $x' = (x - \alpha)/\beta$,

(a) Lognormal: $h(x') = \log x', \quad x' \geq 0$

(b) The S_B system: $h(x') = \log\{x'/(1 - x')\}, \quad 0 \leq x' \leq 1$

(c) The S_u system: $h(x') = \text{arcsinh } x', \quad -\infty < x' < \beta$

The labels S_B and S_u refer to the bounded and unbounded range of values of X, respectively. There are comprehensive discussions of the Johnson system in Johnson and Kotz (1970a, pp. 22-27) and in Kendall and Stuart (1977, pp. 182-185).

Johnson (1965, pp. 547-558) gives tables of values of z and δ corresponding to pairs of values of skewness and kurtosis of X. Hill et al. (1976, pp. 180-189) fit Johnson curves by a computer algorithm from the moments of X; I. D. Hill (1976, pp. 190-192) gives algorithms for normal-Johnson and Johnson-normal transformations.

[7.15.8] *Other Normalizing Transformations*. (a) Let X be a continuous rv with cdf $\psi(x)$, and let

$$y = g(x) = \Phi^{-1}\{\psi(x)\}$$

the *coordinate transformation* (Hoyle, 1973, p. 211), where Φ^{-1} is the inverse of the standard normal cdf Φ. If $Y = g(X)$, then Y has a N(0,1) distribution, since $\Phi\{(g(x)\} = \psi(x)$.

The function $\Phi^{-1}(\cdot)$ cannot be expressed in closed form. Any of the approximations to normal quantiles given in [3.8] may be used, however, noting that, if $\Phi(z) = 1 - p$, then $z = \Phi^{-1}(1 - p)$; in [3.8], replace z_p by z.

(b) *The Probit Transformation* (Bliss, 1935, pp. 134-167; Finney, 1971, p. 19). This is defined by

$$p = \Phi(y - 5)$$

where p is an observable proportion, estimating the probability with which the equivalent normal deviate y - 5 is exceeded in a standard normal distribution. The equivalent normal deviate is reduced by 5 to make the probability of negative values of y negligibly small. The quantity y is the *probit* of the proportion p; the corresponding rv is Y, where

$$Y = (X - \mu)/\sigma$$

and in bioassay problems, X is the logarithm of "dose" (Finney, 1971, p. 11). For an interesting history of the subject, see Finney (1971, pp. 38-42). Fisher and Yates (1964, table IX) give probits corresponding to values of p.

REFERENCES

The numbers in square brackets give the sections in which the corresponding reference is cited.

Abdel-Aty, S. H. (1954). Approximate formulae for the percentage points and the probability integral of the non-central χ^2 distribution, *Biometrika* 41, 538-540. [7.9.3]

Abramowitz, M., and Stegun, I. A. (eds.) (1964). *Handbook of Mathematical Functions*, Washington, D.C.: National Bureau of Standards. [7.6.7; 7.8.4, 7; 7.9.3, 5; 7.10.2; 7.11.2; 7.12.5; 7.13.2]

Anderson, T. W. (1958). *Introduction to Multivariate Statistical Analysis*, New York: Wiley. [7.14.4]

Anscombe, F. J. (1948). The transformation of Poisson, binomial, and negative binomial data, *Biometrika* 35, 246-254. [7.15.2, 3, 4]

Anscombe, F. J. (1950). Table of the hyperbolic transformation $\sinh^{-1}\sqrt{x}$, *Journal of the Royal Statistical Society* A113, 228-229. [7.10.7]

Ashby, T. (1968). A modification to Paulson's approximation to the variance ratio distribution, *The Computer Journal* 11, 209-210. [7.12.9]

Azorín, P. F. (1953). Sobre la distribución t no central, I, II, *Trabajos de Estadística* 4, 173-198, 307-337. [7.15.4]

Bartko, J. J. (1966). Approximating the negative binomial, *Technometrics* 8, 345-350; (1967): erratum, *Technometrics* 9, 498. [7.3.5]

Birnbaum, Z. W., and Saunders, S. C. (1969). A new family of life distributions, *Journal of Applied Probability* 6, 319-327. [7.14.3]

Bliss, C. I. (1935). The calculation of the dosage mortality curve, *Annals of Applied Biology* 22, 134-167. [7.15.8]

Bohman, H. (1963). Two inequalities for Poisson distributions, *Skandinavisk Aktuarietidskrift* 46, 47-52. [7.2.11]

Borenius, G. (1953). On the statistical distribution of mine explosions, *Skandinavisk Aktuarietidskrift* 36, 151-157. [7.5.2]

Borges, R. (1970). Eine Approximation der Binomialverteilung durch die Normalverteilung der Ordnung 1/n, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 14, 189-199. [7.1.8]

Box, G. E. P., and Cox, D. R. (1964). An analysis of transformations, *Journal of the Royal Statistical Society* B26, 211-252. [7.15.6]

Camp, B. H. (1951). Approximation to the point binomial, *Annals of Mathematical Statistics* 22, 130-131. [7.1.7]

Carter, A. H. (1947). Approximation to percentage points of the z distribution, *Biometrika* 34, 352-358. [7.12.8]

Cheng, T. T. (1949). The normal approximation to the Poisson distribution and a proof of a conjecture of Ramanujan, *Bulletin of the American Mathematical Society* 55, 396-401. [7.2.5]

Chu, J. T. (1956). Errors in normal approximations to t, τ, and similar types of distribution, *Annals of Mathematical Statistics* 27, 780-789. [7.10.5]

Cochran, W. G. (1940). Note on an approximate formula for the significance levels of z, *Annals of Mathematical Statistics* 11, 93-95. [7.12.8]

Curtiss, J. H. (1943). On transformations used in the analysis of variance, *Annals of Mathematical Statistics* 14, 107-122. [7.15.3, 4]

Draper, N. R., and Hunter, W. G. (1969). Transformations: Some examples revisited, *Technometrics* 11, 23-40. [7.15.6]

Feller, W. (1968). *Introduction to Probability Theory and Its Applications*, Vol. 1 (3rd ed.), New York: Wiley. [7.4.4]

Finney, D. J. (1971). *Probit Analysis* (3rd ed.), Cambridge: Cambridge University Press. [7.15.8]

Fisher, R. A. (1924). On a distribution yielding the error functions of several well-known statistics, *Proceedings of the International Mathematical Congress*, Toronto, 805-813. [7.12.2, 8]

Fisher, R. A. (1925). Expansion of Student's integral in powers of n^{-1}, *Metron* 5(3), 109-112. [7.10.6]

Fisher, R. A., and Cornish, E. A. (1960). The percentile points of distributions having known cumulants, *Technometrics* 2, 205-226. [7.8.9; 7.10.10; 7.12.10]

Fisher, R. A., and Yates, F. (1964). *Statistical Tables for Biological, Agricultural, and Medical Research*, London and Edinburgh: Oliver & Boyd. [7.15.8]

Foster, F. G., and Stuart, A. (1954). Distribution-free tests in time series based on the breaking of records, *Journal of the Royal Statistical Society* B16, 1-16. [7.5.4]

Freeman, M. F., and Tukey, J. W. (1950). Transformations related to the angular and the square root, *Annals of Mathematical Statistics* 21, 607-611. [7.1.4; 7.2.3; 7.15.3]

Gebhardt, F. (1969). Some numerical comparisons of several approximations to the binomial distribution, *Journal of the American Statistical Association* 64, 1638-1646. [7.1.5, 6, 8]

Gebhardt, F. (1971). Incomplete Beta-integral $B(x;2/3,2/3)$ and $[p(1 - p)]^{-1/6}$ for use with Borges' approximation of the binomial distribution, *Journal of the American Statistical Association* 66, 189-191. [7.1.8]

Gnanadesikan, M., and Gupta, S. S. (1970). A selection procedure for multivariate normal distributions in terms of the generalized variances, *Technometrics* 12, 103-117. [7.8.7; 7.14.4]

Goldberg, G., and Levine, H. (1946). Approximate formulas for the percentage points and normalization of t and χ^2, *Annals of Mathematical Statistics* 17, 216-225. [7.8.9; 7.10.10]

Govindarajulu, Z. (1965). Normal approximations to the classical discrete distributions, *Sankhȳa* A27, 143-172. [7.1.3; 7.2.5]

Hemelrijk, J. (1967). The hypergeometric, the normal and chi-squared, *Statistica Neerlandica* 21, 225-229. [7.4.2]

Hill, G. W. (1970a). Algorithm 395: Student's t-distribution, *Communications of the Association for Computing Machinery* 13, 617-619. [7.10.8]

Hill, G. W. (1970b). Algorithm 396: Student's t-quantiles, *Communications of the Association for Computing Machinery* 13, 619-620. [7.10.10]

Hill, G. W. (1976). New approximations to the von Mises distribution, *Biometrika* 63, 673-676. [7.7.3, 4, 5]

Hill, I. D. (1976). Algorithm AS 100: Normal-Johnson and Johnson-normal transformations, *Applied Statistics* 25, 190-192. [7.15.7]

Hill, I. D., Hill, R., and Holder, R. L. (1976). Algorithm AS 99: Fitting Johnson curves by moments, *Applied Statistics* 25, 180-189. [7.15.7]

Hotelling, H. (1953). New light on the correlation coefficient and its transforms, *Journal of the Royal Statistical Society* B15, 193-232. [7.15.2]

Hoyle, M. H. (1973). Transformations--An introduction and a bibliography, *International Statistical Review* 41, 203-223; erratum (1976): *International Statistical Review* 44, 368. [7.15.1, 3, 4, 5, 8]

Jensen, D. R., and Solomon, H. (1972). A Gaussian approximation to the distribution of a definite quadratic form, *Journal of the American Statistical Association* 67, 898-902. [7.9.4; 7.14.1]

Johnson, N. L. (1949). Systems of Frequency curves generated by methods of translation, *Biometrika* 36, 149-176. [7.15.7]

Johnson, N. L. (1959). On an extension of the connexion between Poisson and χ^2-distributions, *Biometrika* 46, 352-363. [7.9.2, 3]

Johnson, N. L. (1965). Tables to facilitate fitting S_U frequency curves, *Biometrika* 52, 547-558. [7.15.7]

Johnson, N. L., and Kotz, S. (1969). *Distributions in Statistics: Discrete Distributions*, New York: Wiley. [7.1.3, 6, 13; 7.3.1, 2, 5; 7.4.2; 7.5.1, 2, 3, 4; 7.6.2; 7.8.7]

Johnson, N. L., and Kotz, S. (1970a). *Distributions in Statistics: Continuous Univariate Distributions*, Vol. 1, New York: Wiley. [7.8.2, 3, 4, 7, 9; 7.15.7]

Johnson, N. L., and Kotz, S. (1970b). *Distributions in Statistics: Continuous Univariate Distributions*, Vol. 2, New York: Wiley. [7.9.2, 3, 4, 5; 7.10.2, 3, 4, 6; 7.11.2, 3, 4; 7.12.2, 4, 7, 8; 7.13.2; 7.14.1, 2, 3]

Johnson, N. L., and Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions*, New York: Wiley. [7.8.7; 7.14.4]

Kendall, M. G., and Stuart, A. (1966). *The Advanced Theory of Statistics*, Vol. 3, New York: Hafner. [7.15.1]

Kendall, M. G., and Stuart, A. (1977). *The Advanced Theory of Statistics*, Vol. 1 (4th ed.), New York: Macmillan. [7.8.3, 4; 7.12.2, 4, 8; 7.15.2, 7]

Laubscher, N. F. (1960). Normalizing the noncentral t and F distributions, *Annals of Mathematical Statistics* 31, 1105-1112. [7.12.5; 7.13.2; 7.15.2, 4]

Makabe, H., and Morimura, H. (1955). A normal approximation to the Poisson distribution, *Reports on Statistical Applications Research* (Union of Japanese Scientists and Engineers) 4, 37-46. [7.2.5]

Martin, D. C., and Katti, S. K. (1962). Approximations to the Neyman Type A distribution for practical problems, *Biometrics* 18, 354-364. [7.5.1]

Mathur, R. K. (1961). A note on Wilson-Hilferty transformation, *Calcutta Statistical Association Bulletin* 10, 103-105. [7.8.4]

Mickey, M. R. (1975). Approximate test probabilities for Student's t distribution, *Biometrika* 62, 216-217. [7.10.3]

Molenaar, W. (1970). *Approximations to the Poisson, Binomial, and Hypergeometric Distribution Functions*, Mathematical Centre Tracts, Vol. 31, Amsterdam: Mathematisch Centrum. [7.1.4, 5, 6, 7, 8, 9, 10, 11, 12; 7.2.2, 3, 5, 6, 7, 8, 9, 10; 7.4.1, 2, 3, 5, 6, 7]

Moore, P. G. (1957). Transformations to normality using fractional powers of the variables, *Journal of the American Statistical Association* 52, 237-246. [7.15.6]

Moran, P. A. P. (1966). Accurate approximations for t-tests, in *Research Papers in Statistics: Festschrift for J. Neyman* (F. N. David, ed.), 225-230. [7.10.9]

Mudholkar, G. S., Chaubey, Y. P., and Lin, C.-C. (1976). Some approximations for the noncentral F-distribution, *Technometrics* 18, 351-358. [7.13.3]

Neyman, J. (1939). On a new class of "contagious" distributions, applicable in entomology and bacteriology, *Annals of Mathematical Statistics* 10, 35-57. [7.5.1]

Nicholson, W. L. (1956). On the normal approximation to the hypergeometric distribution, *Annals of Mathematical Statistics* 27, 471-483. [7.4.5]

Parzen, E. (1960). *Modern Probability Theory and Its Applications*, New York: Wiley. [7.1.2; 7.2.2]

Patil, G. P. (1960). On the evaluation of the negative binomial distribution with examples, *Technometrics* 2, 501-505. [7.3.2]

Patnaik, P. B. (1949). The non-central χ^2- and F-distributions and their applications, *Biometrika* 36, 202-232. [7.9.3]

Paulson, E. (1942). An approximate normalization of the analysis of variance distribution, *Annals of Mathematical Statistics* 13, 233-235. [7.12.4, 9]

Peiser, A. M. (1943). Asymptotic formulas for significance levels of certain distributions, *Annals of Mathematical Statistics* 14, 56-62. [7.8.9; 7.10.10]

Peizer, D. B., and Pratt, J. W. (1968). A normal approximation for binomial, F, beta, and other common, related tail probabilities, I, *Journal of the American Statistical Association* 63, 1416-1456. [7.1.9; 7.2.8; 7.3.6; 7.6.5; 7.8.6; 7.10.4; 7.12.6]

Pratt, J. (1968). A normal approximation for binomial, F, beta, and other common, related tail probabilities, II, *Journal of the American Statistical Association* 63, 1457-1483. [7.1.11]

Prescott, P. (1974). Normalizing transformations of Student's t distribution, *Biometrika* 61, 177-180. [7.10.11]

Raff, M. S. (1956). On approximating the point binomial, *Journal of the American Statistical Association* 51, 293-303. [7.1.2, 6, 7]

Read, C. B. (1973). An application of a result of Watson to estimation of the normal standard deviation, *Communications in Statistics* 1, 183-185. [7.11.2]

Riordan, J. (1949). Inversion formulas in normal variable mapping, *Annals of Mathematical Statistics* 20, 417-425. [7.2.9]

Roussas, G. G. (1973). *A First Course in Mathematical Statistics*, Reading, Mass.: Addison-Wesley. [7.1.2]

Sahai, H., and Thompson, W. O. (1974). Comparisons of approximations to the percentiles of the t, χ^2, and F distributions, *Journal of Statistical Computation and Simulation* 3, 81-93. [7.8.9; 7.10.10; 7.12.10]

Sankaran, M. (1959). On the noncentral chi-square distribution, *Biometrika* 46, 235-237. [7.9.4]

Scott, A., and Smith, T. M. F. (1970). A note on Moran's approximation to Student's t, *Biometrika* 57, 681-682. [7.10.9]

Severo, N. C., and Zelen, M. (1960). Normal approximation to the chi-square and noncentral F probability functions, *Biometrika* 47, 411-416. [7.8.3, 4, 10; 7.13,2]

Slud, E. V. (1977). Distribution inequalities for the binomial law, *Annals of Probability* 5, 404-412. [7.1.13]

Upton, G. J. G. (1974). New approximations to the distribution of certain angular statistics, *Biometrika* 61, 369-373. [7.7.2]

Wallace, D. L. (1959). Bounds on normal approximations to Student's and the chi-square distributions, *Annals of Mathematical Statistics* 30, 1121-1130; correction (1960): *Annals of Mathematical Statistics* 31, 810. [7.8.5; 7.10.3, 4]

Wallis, W. A., and Moore, G. H. (1941). A significance test for time series analysis, *Journal of the American Statistical Association* 36, 401-412. [7.5.3]

Wilson, E. B., and Hilferty. M. M. (1931). The distribution of chi-square, *Proceedings of the National Academy of Science* 17, 684-688. [7.2.6; 7.8.4]

Wishart, J. (1947). The cumulants of the z and of the logarithmic χ^2 and t distributions, *Biometrika* 34, 170-178, 374. [7.12.2]

chapter 8

ORDER STATISTICS FROM NORMAL SAMPLES

In this chapter a number of results are listed which relate to order statistics from normally distributed parent populations. Asymptotic distributions are included along with discussion of each statistic of interest, but limiting normal distributions of linear combinations and other functions of order statistics with other parent populations are listed in [6.2.9] and the remainder of Section 6.2.

The notation is defined in [8.1.1], and the warning given there about conflicting notation in the literature should be noted. Basic properties of independence and distributions of order statistics appear in Section 8.1, while moments are given in Section 8.2 and in Tables 8.2(a) to 8.2(d) which collect together for the first time results for the first four moments of the order statistics in samples not larger than 5. Properties of deviates of order statistics from the sample mean are given in Section 8.3, as well as properties of the ratio of such deviates to an estimate of the standard deviation. In Section 8.4 appear properties of the sample range and of the ratio of the range to an estimate of standard deviation. Quasi-ranges are discussed in Section 8.5, and several results relating to the sample median and to the sample midrange are given in Section 8.6. Sample quantiles other than the median are briefly covered in Section 8.7; properties of Gini's mean difference, of the trimmed mean, and of some statistics based on samples from several populations are given in Section 8.8.

A good source book with proofs and extensive discussion is David (1970); see also chap. 14 of Kendall and Stuart (1977); Galambos (1978) has a detailed exposition of the asymptotic theory of extreme order statistics, which updates Gumbel's classic work of 1958.

8.1 ORDER STATISTICS: BASIC RESULTS

[8.1.1] Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a normal distribution. If these rvs are arranged in ascending order of magnitude and written

$$X_{(1;n)} \le X_{(2;n)} \le \cdots \le X_{(n;n)}$$

then $X_{(r;n)}$ is the *rth order statistic* ($r = 1, 2, \ldots, n$). When no confusion arises, we write more simply

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$$

In some sources, the labeling is made in reverse order, with $X_{(1)}$ as the largest- rather than as the smallest-order statistic, and it is wise to check in each case, particularly when tables are consulted.

[8.1.2.1] *Independence*. The order statistics are not mutually independent. However, for fixed values of r and s, and as $n \to \infty$ (David, 1970, pp. 211, 213),

(a) $X_{(r;n)}$ and $X_{(n-s+1;n)}$ are asymptotically independent,

(b) both are asymptotically independent of the central order statistics,

(c) both are asymptotically independent of the sample mean $\bar{X}$.

Further, the order statistics of a random sample from any population have a *Markov property*; for a $N(\mu,\sigma^2)$ parent with cdf F(x), this is given for $1 \le r \le n - 1$ by (Pyke, 1965, p. 399)

$$\Pr\{X_{(r+1;n)} \le x \mid X_{(r;n)} = y\} = 1 - \{1 - F(x)\}^{n-r} \cdot \{1 - F(y)\}^{-(n-r)}, \qquad x > y$$

See [8.8.1] for certain functions of order statistics which are uncorrelated.

[*8.1.2.2*] Let $X_1, X_2, \ldots, X_n$ be a random sample from a $N(\mu,\sigma^2)$ distribution, and let $X_{(1)}, \ldots, X_{(n)}$ be the order statistics. Let

$$U = \sum_{i=1}^{n} c_i X_{(i)} \qquad \sum_{i=1}^{n} c_i = 0$$

Then if $S^2 = (n - 1)^{-1}\Sigma_{i=1}^{n}(X_i - \bar{X})^2$ is the sample variance, U/S, $\bar{X}$, and S are mutually independent (David, 1970, pp. 71, 86). This result applies to several linear combinations of order statistics discussed in this chapter.

Deviate from the mean: $(X_{(i)} - \bar{X})/S$, $\bar{X}$ and S are mutually independent ([8.3.1]).

Range: $(X_{(n)} - X_{(1)})/S$, $\bar{X}$ and S are mutually independent (Section 8.4).

Quasi-range: $(X_{(n-r+1)} - X_{(r)})/S$, $\bar{X}$ and S are mutually independent (Section 8.5).

[*8.1.3*] *Distributions*. Let the parent distribution be $N(0,1)$, and let $F_r(x)$ and $f_r(x)$ denote the cdf and pdf of $X_{(r;n)}$, respectively. Then, when $1 \le r \le n$,

(a) $$F_r(x) = \sum_{i=r}^{n} \binom{n}{i} p^i (1 - p)^{n-i} = I_p(r, n - r + 1)$$

(b) $$f_r(x) = r\binom{n}{r} p^{r-1}(1 - p)^{n-r}\phi(x)$$

(c) $$F_r(-x) = 1 - F_{n-r+1}(x)$$

(d) $$f_r(-x) = f_{n-r+1}(x)$$

where $p = \Phi(x)$, $I_p(a,b)$ is the incomplete beta function, and $B(a,b)$ is the beta function (David, 1970, p. 8). Special cases when $r = 1$ or $r = n$ are given by

$$\Pr(X_{(1)} \le x) = 1 - (1 - p)^n$$

$$\Pr(X_{(n)} \le x) = p^n$$

$$f_n(x) = np^{n-1}\phi(x)$$

$$f_1(x) = n(1 - p)^{n-1}\phi(x)$$

For a discussion of the extremes $X_{(1)}$ and $X_{(n)}$, see Gumbel (1958, pp. 129-140).

[8.1.4] *Percentage Points of Order Statistics with a* $N(\mu,\sigma^2)$ *Parent*. Let $\Pr[X_{(r)} < x_{(r),p}] = 1 - p$, and let Y have a F distribution with $2(n - r + 1)$ and $2r$ degrees of freedom, such that

$$\Pr(Y < y_{2(n-r+1),2r;\beta}) = 1 - \beta \qquad \text{(see [7.12.7])}$$

Then (Guenther, 1977, pp. 319-320)

$$\Phi\left(\frac{x_{(r),1-\alpha} - \mu}{\sigma}\right) = \frac{r}{r + (n - r + 1)y_{2(n-r+1),2r;\alpha}} = q, \text{ say}$$

To obtain $x_{(r),1-\alpha}$, first derive the percentile $y_{2(n-r+1),2r;\alpha}$ of F from tables of F or by means of a desk calculator (Guenther, 1977, p. 319; see [7.12.7] to [7.12.10] and Johnson and Kotz, 1970, chap. 26); then solve $\Phi(y) = q$ for y using standard normal tables (see also Section 3.8). For the extreme-order statistics, solve (Gupta, 1961, p. 889)

$$\Phi(\{x_{(1),\alpha} - \mu\}/\sigma) = 1 - \alpha^{1/n}$$

$$\Phi(\{x_{(n),\alpha} - \mu\}/\sigma) = (1 - \alpha)^{1/n}$$

[8.1.5] Tables of the cdf and percentage points of $X_{(1;n)}$ and $X_{(n;n)}$ are given in several standard sources, listed in Table 8.1. In addition, Govindarajulu and Hubacker (1964, pp. 77-78) table percentage points of $X_{(r;n)}$ corresponding to α, $1 - \alpha = 0.50, 0.25, 0.10, 0.025, 0.01$; $n = 1(1)30$; $r = 1(1)(1 + [(1/2)n])$, when the parent distribution is N(0,1); Gupta (1961, pp. 890-891) tables similar percentage points corresponding to $1 - \alpha = 0.50, 0.75, 0.90, 0.95, 0.99$; $n = 1(1)10$, $r = 1(1)n$, and values of r corresponding to the extreme and central order statistics when $n = 11(1)20$, to four decimal places. See also David (1970, pp. 225-233) for references to other tables.

[8.1.6] *Joint Distributions*. Let the parent distribution be N(0,1), and let $f(x_1,x_2,\ldots,x_k;r_1,r_2,\ldots,r_k)$ be the joint pdf of

TABLE 8.1 Normal Order Statistics: Tables and Coverages in Some Standard Sources

Function	Source[a]	Coverage	Decimal places	Significant figures
$EX_{(i;n)}$	PH1, p. 190	$n = 2(1)26(2)50$, $n - i + 1 = 1(1)[\frac{1}{2}n]$	3, 2	
	OW, pp. 151-154	$n = 2(1)50$, $n - i + 1 = 1(1)[\frac{1}{2}n]$	4	
	RMM, pp. 89-90	$n = 2(1)40$, $n - i + 1 = 1(1)[\frac{1}{2}n]$	2	
	PH2, pp. 27, 205-210	$n = 2(1)100(25)200$, $n - i + 1 = 1(1)[\frac{1}{2}n]$	5	
	Hart 2, pp. 425-455	$n = 2(1)100$ and values to 400, $n + i + 1 = 1(1)[\frac{1}{2}n]$	5	
	SG, p. 193	$n = 2(1)20$, $n - i + 1 = 1(1)[\frac{1}{2}n]$	10	
	Y, pp. 33-37	$n = 2(1)50$, $n - i + 1 = 1(1)[\frac{1}{2}n]$	20	
$E\{X_{(i;n)}X_{(j;n)}\}$	SG, pp. 191, 194-199	$n = 1(1)20$, $j = 1(1)n$, $n - i + 1 = 1(1)[\frac{1}{2}n]$	10	
	Y, pp. 38-49	$n = 2(1)30$, $j = 1(1)n$, $n - i + 1 = 1(1)[\frac{1}{2}n]$	8	
$\mathrm{Cov}\{X_{(i;n)}, X_{(j;n)}\}$	OW, pp. 163-169	$n = 2(1)20$, $j = 1(1)n$, $n - i + 1 = 1(1)[\frac{1}{2}n]$	4	
	PH2, pp. 211-213	$n = 2(1)20$, $j = 1(1)n$, $n - i + 1 = 1(1)[\frac{1}{2}n]$	6	
	SG, pp. 191, 200-205	$n = 2(1)20$, $j = 1(1)n$, $n - i + 1 = 1(1)[\frac{1}{2}n]$	10	
$\sqrt{\mathrm{Var}\ X_{(i;n)}}$	Y, pp. 33-37	$n = 2(1)50$, $n - i + 1 = 1(1)[\frac{1}{2}n]$	20	
$(X_{n;n}) - \mu)/\sigma$, $(\mu - X_{(1;n)})/\sigma$:				
(a) Percent points	(a) RMM, p. 94	$n = 1(1)30$, $\alpha = 0.05, 0.01, 0.001$	3	
	PH1, p. 184	$n = 1(1)30$, α, $1 - \alpha = 0.10, 0.05, 0.025$		
	SG, p. 322	$0.01, 0.005, 0.001$	3	
(b) cdf	(b) PH2, pp. 184-187	$n = 1(1)25(5)60, 100$, $-2.6 \le x \le 6.1$	7	
$X_{(1;n)}, X_{(n;n)}$: μ_2, μ_3, μ_4, ratios	PH2, p. 216 SG, p. 188	$n = 1(1)\ 50$		
$E\{X^r_{(n;n)}\}$	SG, p. 186	$n = 1(1)50$, $r = 1(1)10$	--	10

$(X_{(n;n)} - \bar{X})/\sigma$, $(\bar{X} - X_{(1;n)})/\sigma$:			
Percent points	PH1, p. 184 SG, p. 322	n = 3(1)9; α, 1 - α = 0.10,0.05,0.025, 0.01, 0.005, 0.001	3
cdf	PH2, pp. 189-199	n = 3(1)25; 0.00 ≤ x ≤ 4.90	4, 5, 6
$(X_{(n;n)} - \bar{X})/S_\nu, S_\nu$ from indpt. sample; percent points	PH1, pp. 185-186	n = 3(1)10,12; α = 0.10,0.05,0.025,0.01, 0.005,0.001, ν = 10(1)20,24,30,40,60,120,∞ Also for α = 0.05,0.01; ν = 5(1)9	2
	SG, p. 326	n = 3(1)10,12; α = 0.05,0.01; ν = 10(1)20, 24,30,40,60,120,∞	2
$(X_{(n;n)} - \bar{X})/S, S$ from same sample; percent points	SG, p. 324 Y, p. 90	n = 3(1)25; α = 0.10,0.05,0.025,0.01 n = 3(1)74; α = 0,10,0.05,0.01	3 3
<u>Range W</u>			
W: percent points	OW, p. 139	n = 2(1)20(2)40(10)100; α, 1 - α = 0.10,0.05,0.025,0.01,0.001	3
	PH1, p. 177	n = 2(1)20; α, 1 - α = 0.10,0.05,0.025 0.01,0.005,0.001	2
	Hart 1, pp. 372-374	. = 2(1)20(2)40(10)100; α = 0.10(0.1)0.90; α, 1 - α = 0.05,0.025,0.01,0.005,0.001, 0.0005,0.0001	6
	SG, p. 327	n = 2(1)20; α, 1 - α = 0.10,0.05,0.025,0.01, 0.005,0.001	2
	Y, p. 61	n = 2(1)40(5)50(10)100; α, 1 - α = 0.5,0.25, 0.10,0.05,0.025,0.01,0.005	3
W: cdf	PH1, pp. 178-183 Hart 1, pp. 240-370	n = 2(1)20; x = 0.00(0.05)7.30 n = 2(1)20(2)40(10)100; x = 0.00(0.01)10.47	4 8
W: pdf	Hart 1, pp. 38-97 Y, pp. 54-59	n = 2(1)16; x = 0.00(0.01)9.99 n = 3(1)20; x = 0.00(0.05)7.65	8 4

TABLE 8.1 (continued)

Function	Source[a]	Coverage	Decimal places	Significant figures
Range W (continued)				
W: moments	OW, p. 140	$\mu,\sigma^2,\sqrt{\beta_1},\beta_2$; n = 2(1)20(2)40(10)100	3	
	PH1, p. 176	$\mu,\sigma,\sigma^2,\beta_1,\beta_2,\mu/\sigma^2,\mu^2/\sigma^2$; n = 2(1)20	5,4,3,2	
	Hart 1, pp. 376, 377	$\mu,\sigma^2,\sqrt{\beta_1},\beta_2$; n = 2(1)100	10,8,7	
	Y, p. 60	$\mu,\sigma^2,\sqrt{\beta_1},\beta_2$; n = 2(1)100	7,6,5	
Studentized range W/S_ν (S_ν from independent sample)				
W/S_ν: percent points	OW, pp. 144-148	n = 2(1)20,24,30,40,60,100; α, $1-\alpha$ = 0.10,0.05,0.025,0.01,0.005; ν = 1,3,5(5)20,60,∞	3	
	PH1, pp. 191-193	n = 2(1)20; α = 0.10,0.05,0.01; ν = 1(1)20,24,30,40,60,120,∞	2	
	Hart 1, pp. 624-661	n = 2(1)20(2)40(10)100; α, $1-\alpha$ = 0.001, 0.005,0.01,0.025,0.05,0.10(0.1)0.50; ν = 1(1)20,24,30,40,60,120,∞	3	
	SG, pp. 114-115	n = 2(1)20; α = 0.05,0.01; ν = 1(1)20,24,30,40,60,120,∞	2	
	Y, p. 63	n = 2(1)6,8,10,15,20,30; α = 0.05,0.01; ν = 1(1)10(2)20,24,30,40,60,120,∞	4	
W/S_ν: cdf	Hart 1, pp. 382-622	n = 2(1)20(2)40(10)100; ν = 1(1)20,24, 30,40,60,120; $0.0 \le x \le 2000.0$	6	

W/S (S from same sample)				
W/S: percent points	PH1, p. 200	$n = 3(1)20(5)100,150,200,500,1000$; α, $1-\alpha$ = 0.10,0.05,0.025,0.01,0.005,0.001	3,2	
	SG, pp. 328-329	As for PH1, p. 200; but lower % points missing for $n = 3(1)9$	3,2	
$(\bar{X}-\mu)/W$, $\bar{X}$ and W from same sample				
Percent points	OW, p. 142	$n = 2(1)20$; α = 0.05,0.025,0.01,0.005, 0.001,0.0005	3,2	
$\lvert\bar{X}-\mu\rvert/W$: percent points	SG, p. 120	$n = 2(1)12$; α = 0.05,0.01	3,2	
Indpt. ranges W_1, W_2				
W_1/W_2: percent points	PH1, pp. 196-199	$n_1 = 2(1)15$, $n_2 = 2(1)15$; α = 0.5,0.25, 0.10,0.05,0.025,0.01,0.005,0.001	--	4
	Hart 1, pp. 224-227	As for PH1		4
W_1/W_2: cdf	Hart 1, pp. 172-221	$n_1 = 2(1)15$, $n_2 = 2(1)15$; $1 \le x \le 600$	5	
W_1/W_2: pdf	Hart 1, pp. 100-169	$n_1 = 2(1)15$, $n_2 = 2(1)15$; $0 \le x \le 600$	6	
$\lvert\bar{X}_1-\bar{X}_2\rvert/(W_1+W_2)$: percent points	PH1, pp. 194-195	$n_1 = 2(1)20$, $n_2 = 2(1)20$; α = 0.10,0.05,0.02,0.01	3	
$\lvert\bar{X}_1-\bar{X}_2\rvert/\{\frac{1}{2}(W_1+W_2)\}$	SG, p. 120	$n_1 = n_2 = 2(1)12$; α = 0.05,0.01	3	
Quasi-range: $W_r = X_{(n-r;n)} - X_{(r+1;n)}$				
E W_r	Hart 2, pp. 136-139	$n = 2(1)100$; $r = 0(1)8$	6	
Var W_r	Hart 2, pp. 142-144	$n = 2(1)100$; $r = 0(1)8$	5	
s.d. W_r	Hart 2, pp. 146-148	$n = 2(1)100$; $r = 0(1)8$	5	

TABLE 8.1 (continued)

Function	Source[a]	Coverage	Decimal places	Significant figures
Quasi-range (continued)				
W_r: percent points	Hart 2, pp. 296-319	n = 2(1)20(2)40(10)100; r = 0(1)8; α = 0.1(0.1)0.9; α, 1 - α = 0.05,0.025, 0.01,0.005,0.001,0.0005,0.0001	6	
W_r: cdf	Hart 2, pp. 160-294	n = 2(1)20(2)40(10)100; r = 0(1)8; x = 0.05(0.05)10.0	8	
k indpt. samples of size n				
S^2_{max}/S^2_{min}: percent points	OW, p. 101	n = 3(1)11,13,16,21,31,61,∞; k = 2(1)12; α = 0.05,0.01	2,1,0	
	PH1, p. 202	n = 3(1)11,13,16,21,31,61,∞; k = 2(1)12; α = 0.05,0.01	2,1,0	
	Y, pp. 72-75	n = 3(1)31,41,61,121,∞; k = 2(1)20; α = 0.05,0.01	2,1,0	
$S^2_{max}/\Sigma S^2_j$: percent points	PH1, p. 203	n = 2(1)11,17,37,145,∞; k = 2(1)10,12, 15,20; α = 0.05,0.01	4	
	Y, pp. 76-79	n = 2(1)31,41,61,121,∞; k = 2(1)20; α = 0.05,0.01	4	
S^2_{max}/S^2_0, S^2_0 indpt: percent points	PH1, p. 176	n = 2; k = 1(1)10; ν = 10,12,15,20,30, 60,∞; α = 0.05,0.01	2	
	Y, pp. 68-71	n = 2; k = 1(1)10(2)30; ν = 9(1)30(2)50, 60,80,120,240,∞; α = 0.05,0.01	2	

Range: $W_{max}/\Sigma W_j$: percent points	PH1, p. 205	n = 2(1)10; k = 2(1)10,12,15,20; α = 0.05	3
Y half-normal: $E\ Y_{(i;n)}$	PH2, p. 226	n = 1(1)30; i = 1(1)n	4

[a] OW = Owen (1962).
RMM = Rao, Mitra, and Matthai (1966).
PH1 = Pearson and Hartley (1966), vol. 1.
PH2 = Pearson and Hartley (1972), vol. 2.
Hart 1 = Harter (1969a).
Hart 2 = Harter (1969b).
SG = Sarhan and Greenberg (1962).
Y = Yamauti (1972).

$X_{(r_1)}, X_{(r_2)}, \ldots, X_{(r_k)}$; $1 \le k \le n$; $1 \le r_1 \le r_2 \le \cdots \le r_k \le n$. Then (David, 1970, p. 9)

$$f(x_1,\ldots,x_k;r_1,\ldots,r_k) = \begin{cases} c\{\Phi(x_1)\}^{r_1-1} \prod_{i=1}^{k-1} [\{\Phi(x_{i+1}) \\ \quad - \Phi(x_i)\}^{r_{i+1}-r_i-1}]\{1 - \Phi(x_k)\}^{n-r_k} \\ \quad \times \phi(x_1)\cdots\phi(x_k), & x_1 \le x_2 \le \cdots \le x_k \\ 0, & \text{otherwise} \end{cases}$$

where $c = n!/[(r_1 - 1)!(\Pi_{i=1}^{k-1}\{(r_{i+1} - r_i - 1)!\})(n - r_k)!]$. The joint pdf of all n order statistics is

$$\begin{cases} n!\phi(x_1)\cdots\phi(x_k), & x_1 \le x_2 \le \cdots \le x_k \\ 0, & \text{otherwise} \end{cases}$$

If the parent distribution is $N(\mu,\sigma^2)$, the joint pdf is

$$\sigma^{-k} f\left(\frac{x_1 - \mu}{\sigma}, \ldots, \frac{x_k - \mu}{\sigma}; r_1, \ldots, r_k\right)$$

[*8.1.7.1*] Haldane and Jayakar (1963, pp. 89-94) show that a linear combination of $X^2_{(n;n)}$ or of $X^2_{(1;n)}$ converges in distribution to the extreme-value form. Let

$$v^2 \exp(v^2) = n^2/(2\pi)$$

$$Y_n = (X^2_{(n;n)} - v^2 + 2v^{-2} - 7v^{-4})/\{2(1 - v^{-2} + 3v^{-4})\}$$

Then

$$\lim_{n\to\infty} \Pr(Y_n \le y) = \exp(-e^{-y}), \qquad -\infty < y < \infty$$

The pdf of this distribution is $\exp(-y - e^{-y})$. More precisely, the pdf of Y_n is

$$\exp(-y - e^{-y})\{1 + v^{-4}y(y - 2 - ye^{-y}) + O(v^{-6}) + O(n^{-1})\}$$

Note that when $n = 10^3$, 10^5, and 10^7, $v = 3.12$, 4.28, and 5.21, respectively (Haldane and Jayakar, 1963, p. 90). Further,

$$E\{X^2_{(n;n)}\} = v^2 - 2v^{-2} + 3v^{-4} + \gamma + O(v^{-6})$$

$$\mathrm{Var}\{X^2_{(n;n)}\} = (2\pi/3)(1 - 2v^{-2} + 7v^{-4}) + O(v^{-6})$$

where γ is Euler's constant, 0.577215665.

[*8.1.7.2*] Fisher and Tippett (1928, p. 183) give a "penultimate form" of limiting distribution for a suitably normed linear function of $X_{(n;n)}$; this has pdf of the form $k(-x)^{k-1} \exp\{-(-x)^k\}$, for suitable choice of k. This is Gumbel's "third asymptote" (Gumbel, 1958, chaps. 5, 7).

[*8.1.7.3*] Galambos (1978, pp. 52, 65-67) shows that, for a N(0,1) parent,

$$\lim_{n\to\infty} \Pr[\sqrt{2 \log n}\{X_{(n;n)} - a_n\} < x] = \exp(-e^{-x}), \qquad -\infty < x < \infty$$

where

$$a_n = \sqrt{2 \log n} - \tfrac{1}{2}(\log \log n + \log(4\pi))/\sqrt{2 \log n}$$

$$\lim_{n\to\infty} \Pr[\sqrt{2 \log n}\{X_{(1;n)} + a_n\} < x] = 1 - \exp(-e^{x}), \qquad -\infty < x < \infty$$

This choice of a_n goes back to Cramér (1946, pp. 374-378); see also David (1970, pp. 205, 209).

[*8.1.8.1*] The asymptotic distribution of $X_{(n-r+1;n)}$, suitably normed, in a sample from a N(0,1) parent, has pdf $g(y;r,n)$, where

$$g(y;r,n) = r^r\{(r - 1)!\}^{-1} \exp(-ry - re^{-y}), \qquad -\infty < y < \infty$$

If the parent distribution is N(0,1) and $X_{(n-r+1;n)}$ has a mode at m, then $g(y;r,n)$ is the asymptotic distribution of (Kendall and Stuart, 1977, pp. 356-359, 369; Woodroofe, 1975), p. 259)

$$(r/n)\phi(m)\{X_{(n-r+1;n)} - m\}$$

As for the case in which $m = 1$, convergence to the limiting form is slow.

[*8.1.8.2*] Haldane and Jayakar (1963, p. 92) give the following result for the limiting distribution of $X_{(m;n)}$ and of $X_{(n-m+1;n)}$. Define v^2 and Y_n as in [8.1.7.1], replacing $X^2_{(n;n)}$ by $X^2_{(n-m+1;n)}$. Then

$$\lim_{n\to\infty} \Pr(Y_n \le y) = \frac{1}{(m-1)!}\int_{-\infty}^{y} \exp(-mt - e^{-t})\, dt$$

More precisely, Y_n has pdf

$$\{(m-1)!\}^{-1} \exp(-my - e^{-y})\{1 + v^{-4}y(my - 2 - ye^{-y}) + O(m^{-6}) + O(n^{-1})\}, \quad m = 2, 3, \ldots$$

Haldane and Jayakar warn, however, that the approximation should only be used if m^2/n is negligible in comparison with unity; if $n = 1000$, m should not be greater than 3. They give cumulants of the limiting distribution when m is small or large. See also David (1970, p. 210).

[*8.1.8.3*] For the same values of the constants a_n defined in [8.1.7.3], and for fixed $r \ge 1$, with a N(0,1) parent (Galambos, 1978, p. 105),

$$\lim_{n\to\infty} \Pr[\sqrt{2 \log n}\{X_{(n-r+1;n)} - a_n\} < x] = \exp(-e^{-x}) \sum_{t=0}^{r-1} \frac{e^{-tx}}{t!}$$

$$\lim_{n\to\infty} \Pr[\sqrt{2 \log n}\{X_{(r;n)} + a_n\} < x] = 1 - \exp(-e^{x}) \sum_{t=0}^{r-1} \frac{e^{tx}}{t!},$$

$$-\infty < x < \infty$$

8.2 MOMENTS

[*8.2.1*] *Notation*. Let

$$\mu_{(r;n)} = E(X_{(r;n)})$$

$$\mu_{(r,s;n)} = E(X_{(r;n)}X_{(s;n)})$$

$$\sigma_{(r,s;n)} = \text{Cov}(X_{(r;n)},X_{(s;n)}) = E[\{X_{(r;n)} - \mu_{(r;n)}\} \times \{X_{(s;n)} - \mu_{(s;n)}\}]$$

$$\mu^{(k)}_{(r;n)} = E(X_{(r;n)}{}^{k}), \qquad 1 \le r \le n,\ 1 \le s \le n$$

[*8.2.2*] The following *moment relations* hold when the order statistics have a N(0,1) parent (Sarhan and Greenberg, 1962, p. 191; Govindarajulu, 1963, p. 636; Jones, 1948, pp. 271-273):

$$\mu_{(r;n)} = -\mu_{(n-r+1;n)}$$

$$\sigma_{(r,s;n)} = \sigma_{(s,r;n)} = \sigma_{(n-r+1,n-s+1;n)} = \sigma_{(n-s+1,n-r+1;n)}$$

$$= \mu_{(r,s;n)} - \mu_{(r;n)}\mu_{(s;n)}$$

$$\sum_{r=1}^{n} \sum_{s=1}^{n} \sigma_{(r,s;n)} = n, \qquad 1 \le r \le s \le n$$

Harter (1969b, p. 26) gives the following:

$$\mu_{(r+1;n)} = \{n\mu_{(r;n-1)} - (n-r)\mu_{(r;n)}\}/n$$

$$\mu_{(r;n-1)} = \{r\mu_{(r+1;n)} + (n-r)\mu_{(r;n)}\}/n, \qquad 1 \le r \le n - 1$$

Using these recursive relations, the values of $\mu_{(1;n)}$ $(1 \le n \le N)$ are sufficient to compute the set of values of $\mu_{(r;n)}$ $(2 \le r \le n,\ 2 \le n \le N)$. See [8.2.3] below.

If n is even, then (David, 1970, p. 37)

$$\mu^{(k)}_{((1/2)n;n-1)} = \begin{cases} \mu^{(k)}_{((1/2)n;n)}, & k \text{ even} \\ 0, & k \text{ odd} \end{cases}$$

This result holds for any parent distribution symmetric about the origin, and includes a correction by David.

[*8.2.3*] <u>*Moment Values*</u>. (a) Define $F_r(x)$ as in [8.1.3]. Then for a $N(0,\sigma^2)$ parent (David, 1970, p. 29),

$$\mu_{(r;n)} = \int_0^\infty \{F_{n-r+1}(x) - F_r(x)\}\, dx$$

(b) Bose and Gupta (1959, p. 437) give expressions for $\mu^{(k)}_{(r;n)}$, k = 1, 2, 3, 4; these are in terms of certain integral functions. If r = n and $I_n(k) = \int_{-\infty}^{\infty} \{\Phi(x)\}^{n-k} \exp(-(1/2)kx^2)\, dx$, then for an N(0,1) parent,

$$\mu_{(n;n)} = 2\binom{n}{2}(2\pi)^{-1}I_n(2)$$

$$\mu^{(2)}_{(n;n)} = 1 + 3\binom{n}{3}(2\pi)^{-3/2}I_n(3)$$

$$\mu^{(3)}_{(n;n)} = (5\mu_{(n;n)}/2) + 4\binom{n}{4}(2\pi)^{-2}I_n(4)$$

$$\mu^{(4)}_{(n;n)} = (-4/3) + (13\mu^{(2)}_{(r;n)}/3) + 5\binom{n}{5}(2\pi)^{-5/2}I_n(5)$$

(c) From Jones (1948, p. 270) and Godwin (1949, p. 284) we can derive *exact values* for $\mu_{(r;n)}$ and $\mu^{(2)}_{(r;n)}$ for n = 2, 3, 4, and 5. These appear in Table 8.2A and B, respectively. From Bose and Gupta (1959, pp. 438-439) we can derive third and fourth moments; these appear in Table 8.2C and D. These references also include numerical approximations which are accurate to several decimal places; the tables clearly show how these moments differ from the corresponding first four moments (0, 1, 0, and 3, respectively) of unordered N(0,1) random variables.

[*8.2.4*] By inverting the probability integral transformation, David and Johnson (1954, pp. 228-240) developed a *series* in powers of $(n + 2)^{-1}$, giving moments and joint cumulants of N(0,1) order statistics. Convergence may be slow or may not hold at all when r/n is close to zero or one; however, a series by Plackett (1958, pp. 131-142) may converge a little more rapidly, although it has less computational advantage (David, 1970, p. 66). See also Saw (1960, pp. 79-86), who developed bounds for the remainder of the David-Johnson series after an even number of terms, and compared these with bounds derived by Plackett (1958) for his series.

[*8.2.5*] David (1970, pp. 64-65) gives the following *bounds and approximations* for $\mu_{(r;n)}$, of which the lower bound is rather poor: if $r \geq (n + 1)/2$,

$$\Phi^{-1}\{(r - 1)/n\} \leq \mu_{(r;n)} \leq \min[\Phi^{-1}\{r/(n + 1/2)\}, \Phi^{-1}\{(r - 1/2)/n\}]$$

$$\mu_{(r;n)} \le \min\left[\Phi^{-1}\left\{1 - \exp\left(-\sum_{i=n-r+1}^{n} i^{-1}\right)\right\},\right.$$

$$\left.\times\ \Phi^{-1}\left\{1 + \exp\left(-\sum_{i=n-r+1}^{n} i^{-1}\right)^{-1}\right\}\right]$$

The author has communicated that an improved lower bound is $\Phi^{-1}\{r/(n + 1)\}$.

[8.2.6] *Approximations* of the form

$$\mu_{(r;n)} \simeq \Phi^{-1}\{(r - \alpha_{r,n})/(n - 2\alpha_{r,n} + 1)\}$$

due to Blom (1958) are discussed in Harter (1961, pp. 153-156) and in Harter (1969b, p. 456), where values of $\alpha_{r,n}$ are tabled which make the approximation accurate; these are for n = 25, 50, 100, 200, 400. Approximate algorithms for $\alpha_{r,n}$ for intermediate values of n are also given. If $n \le 20$, then $0.33 \le \alpha_{r,n} \le 0.39$ for all r; so that taking $\alpha_{r,n} = 3/8$ (David, 1970, pp. 65, 67),

$$\mu_{r,b} \simeq \Phi^{-1}\{(r - 3/8)/(n + 1 - 3/4)\}, \qquad n \le 20$$

[8.2.7] Bounds and approximations to means, variances and covariances of order statistics can be obtained using results in David (1970, pp. 54-57). General bounds for $E\{X_{(s;n)} - X_{(r;n)}\}$ and for $|EX_{(r;n)}|$ are given in David (1970, pp. 51-54, 68, 69); some of these apply to general classes of parent distributions, such as those which are symmetrical about zero. For example, for independent samples from any parent distribution, $\sigma_{(r,s;n)} \ge 0$ for all r, s, and n (Bickel, 1967, p. 575).

[8.2.8] Tables of moments of N(0,1) order statistics appear in general standard sources, listed in Table 8.1. Others are referenced in David (1970, pp. 225-233); Harter (1961, pp. 158-165) tables values of $\mu_{(r;n)}$ to 5 decimal places for n = 2(1)100(25)250(50)400, and Teichroew (1956, pp. 416-422) tables values of $\mu_{(r;n)}$ and of $E\{X_{(i;n)}X_{(j;n)}\}$ to 10 decimal places for n = 2(1)20. Ruben (1954, pp. 224-226) tables values of $\mu_{(n;n)}^{(k)}$ to 10 significant figures for

TABLE 8.2

A. Exact Values of $EX_{(r;n)}$: n = 2, 3, 4, or 5; N(0,1) Parent

r \ n:	2	3	4	5
1	$-\frac{1}{\sqrt{\pi}}$	$-\frac{3}{2\sqrt{\pi}}$	$-\frac{3}{2\sqrt{\pi}}\left(1+\frac{2a}{\pi}\right)$	$-\frac{5}{4\sqrt{\pi}}\left(1+\frac{6a}{\pi}\right)$
2	$\frac{1}{\sqrt{\pi}}$	0	$-\frac{3}{2\sqrt{\pi}}\left(1-\frac{6a}{\pi}\right)$	$-\frac{5}{2\sqrt{\pi}}\left(1-\frac{6a}{\pi}\right)$
3	--	$\frac{3}{2\sqrt{\pi}}$	$\frac{3}{2\sqrt{\pi}}\left(1-\frac{6a}{\pi}\right)$	0
4	--	--	$\frac{3}{2\sqrt{\pi}}\left(1+\frac{2a}{\pi}\right)$	$\frac{5}{2\sqrt{\pi}}\left(1-\frac{6a}{\pi}\right)$
5	--	--	--	$\frac{5}{4\sqrt{\pi}}\left(1+\frac{6a}{\pi}\right)$

B. Exact Values of $EX^2_{(r;n)}$: n = 2, 3, 4, 5; N(0,1) Parent

r \ n:	2	3	4	5
1	1	$1+\frac{\sqrt{3}}{2\pi}$	$1+\frac{\sqrt{3}}{\pi}$	$1+\frac{5\sqrt{3}}{4\pi}+\frac{5\sqrt{3}b}{2\pi^2}$
2	1	$1-\frac{\sqrt{3}}{\pi}$	$1-\frac{\sqrt{3}}{\pi}$	$1-\frac{10\sqrt{3}}{\pi^2}b$
3	--	$1+\frac{\sqrt{3}}{2\pi}$	$1-\frac{\sqrt{3}}{\pi}$	$1-\frac{5\sqrt{3}}{2\pi}+\frac{15\sqrt{3}b}{\pi^2}$
4	--	--	$1+\frac{\sqrt{3}}{\pi}$	$1-\frac{10\sqrt{3}}{\pi^2}b$
5	--	--	--	$1+\frac{5\sqrt{3}}{4\pi}+\frac{5\sqrt{3}b}{2\pi^2}$

Note: $a = \arcsin(1/3) \simeq 0.33983\ 69094$
$b = \arcsin(1.4) \simeq 0.25268\ 02552$
$\pi \simeq 3.14159\ 2654$; $\sqrt{\pi} \simeq 1.77245\ 3851$

Table 8.2 (continued)

C. Exact Values of $EX^3_{(r;n)}$: n = 2, 3, 4, 5; N(0,1) Parent

r \ n:	2	3	4	5
1	$-\frac{5}{2\sqrt{\pi}}$	$-\frac{15}{4\sqrt{\pi}}$	$-\frac{1}{\pi^{3/2}}\left(\frac{1}{2\sqrt{2}} + 15c\right)$	$-\frac{25}{4\pi^{3/2}}\left(-\pi + \frac{1}{5\sqrt{2}} + 6c\right)$
2	$\frac{5}{2\sqrt{\pi}}$	0	$-\frac{15}{\pi^{3/2}}\left(\pi - \frac{1}{10\sqrt{2}} - 3c\right)$	$-\frac{25}{\pi^{3/2}}\left(\pi - \frac{1}{10\sqrt{2}} - 3c\right)$
3	--	$\frac{15}{4\sqrt{\pi}}$	$\frac{15}{\pi^{3/2}}\left(\pi - \frac{1}{10\sqrt{2}} - 3c\right)$	0
4	--	--	$\frac{1}{\pi^{3/2}}\left(\frac{1}{2\sqrt{2}} + 15c\right)$	$\frac{25}{\pi^{3/2}}\left(\pi - \frac{1}{10\sqrt{2}} - 3c\right)$
5	--	--	--	$\frac{25}{4\pi^{3/2}}\left(-\pi + \frac{1}{5\sqrt{2}} + 6c\right)$

D. Exact Values of $EX^4_{(r;n)}$: n = 2, 3, 4, 5; N(0,1) Parent

r \ n:	2	3	4	5
1	3	$3 + \frac{13}{2\pi\sqrt{3}}$	$3 + \frac{13}{\pi\sqrt{3}}$	$3 + \frac{\sqrt{5}}{4\pi^2} + d$
2	3	$3 - \frac{13}{2\pi\sqrt{3}}$	$3 - \frac{13}{\pi\sqrt{3}}$	$3 + \frac{65}{\pi\sqrt{3}} - \frac{\sqrt{5}}{\pi^2} - 4d$
3	--	$3 + \frac{13}{2\pi\sqrt{3}}$	$3 - \frac{13}{\pi\sqrt{3}}$	$3 - \frac{130}{\pi\sqrt{3}} + \frac{3\sqrt{5}}{2\pi^2} + 6d$
4	--	--	$3 + \frac{13}{\pi\sqrt{3}}$	$3 + \frac{65}{\pi\sqrt{3}} - \frac{\sqrt{5}}{\pi^2} - 4d$
5	--	--	--	$3 + \frac{\sqrt{5}}{4\pi^2} + d$

Note: $c = \arctan(\sqrt{2}) \simeq 0.95531\ 6618$
$d = 65(\pi^2\sqrt{3})^{-1} \arctan(\sqrt{5/3}) \simeq 3.46675\ 52225\ 38$
$\pi \neq 3.14159\ 2654$; $\sqrt{\pi} \simeq 1.77245\ 3851$

Sources: See [8.2.3].

$n = 1(1)50$ and $k = 1(1)10$; he also tables values of the variance, standard deviation, third and fourth central and standardized moments of $X_{(n;n)}$ to 8 and 7 decimal places for $n = 1(1)50$. All of these are based on $N(0,1)$ parents. Tables to 10 decimal places of $\sigma_{r,s;n}$ appear in Owen et al. (1977, table I) for $s \le r$; $r = 1(1)([(1/2)n] + 1)$; $n = 2(1)50$.

[*8.2.9*] Some *exact values of joint moments* $\mu_{(r,s;n)}$ were obtained by Jones (1948, p. 270) and Godwin (1949, pp. 284-285); these are as follows:

$n = 2$: $\mu_{(1,2;2)} = 0 \qquad \sigma_{(1,2;2)} = 1/\pi$

$n = 3$: $\mu_{(1,2;3)} = \mu_{(2,3;3)} = \sqrt{3}/(2\pi)$

$\mu_{(1,3;3)} = -\sqrt{3}/\pi \qquad \sigma_{(1,2;3)} = \sqrt{3}/(2\pi)$

$\sigma_{(2,3;3)} = \sqrt{3}/(2\pi) \qquad \sigma_{(1,3;3)} = (9 - 4\sqrt{3})/(4\pi)$

$n = 4$: $\mu_{(1,2;4)} = \mu_{(3,4;4)} = \sqrt{3}/\pi$

$\mu_{(1,3;4)} = \mu_{(2,4;4)} = -(2\sqrt{3} - 3)/\pi$

$\mu_{(1,4;4)} = -3/\pi \qquad \mu_{(2,3;4)} = (2\sqrt{3} - 3)/\pi$

$n = 5$: $\mu_{(1,2;5)} = \mu_{(4,5;5)} = 5\sqrt{3}/(4\pi)$

$$\mu_{(1,3;5)} = \mu_{(3,5;5)} = \frac{5(3 - \sqrt{3})}{2\pi} - (5\sqrt{3}/\pi^2)\alpha - (15/\pi^2)\beta$$

$\mu_{(1,4;5)} = \mu_{(2,5;5)} = -15/(2\pi) + (45/\pi^2)\beta$

$\mu_{(1,5;5)} = (30/\pi^2)\beta$

$\pi = 3.14159\ 2654$

$\alpha = \arcsin(1/4) \simeq 0.25268\ 02552$

$\beta = \arcsin(1/\sqrt{6}) \simeq 0.42053\ 4335$

[*8.2.10*] Davis and Stephens (1978, pp. 206-212) give a Fortran computer program for approximating the covariance matrix of order statistics from a $N(0,1)$ parent. If $n \le 20$, the maximum error is less than 0.00005.

[*8.2.11*] Let $X_{(1;n)}$, $X_{(2;n)}$, ..., $X_{(n;n)}$ be the order statistics for a combined sample from one or more unknown parent distributions, not necessarily normal. Nonparametric approaches often replace these order statistics by their ranks, or by their expected values $E\{X_{(1;n)}\}$, ..., $E\{X_{(n;n)}\}$ for a N(0,1) parent, that is, by the *normal scores* of the ordered sample. This happens, for example, when X_1, ..., X_n is a combined sample from two populations, the *normal scores statistic* is then the sum of the normal scores in the combined sample of the measurements belonging to the second population. See Lehmann (1975, pp. 96-97) for further discussion and references; see also [7.15.1] and [8.7.6].

Tables of normal scores for different (combined) sample sizes n are referenced in [8.2.8] and Table 8.1. David et al. (1968, table III) table normal scores $E\{Z_{(r;n)}\}$ to 5 decimal places for $n = 2(1)100(25)250(50)400$ and $r = \{n + 1 - [n/2]\}(1)n$.

8.3 ORDERED DEVIATES FROM THE SAMPLE MEAN

[*8.3.1.1*] The statistics $X_{(i;n)} - \bar{X}$, based on a sample of size n from a $N(\mu,\sigma^2)$ population, are *ordered deviates from the mean*. Let S or S' be an estimate of σ with ν degrees of freedom; then $\{X_{(i;n)} - \bar{X}\}/S'$ is a *Studentized deviate* from $\bar{X}$ if S' is based on a second independent sample X_1^*, X_2^*, ..., $X_{\nu+1}^*$ and

$$S'^2 = \nu^{-1} \sum_{i=1}^{\nu+1} (X_j^* - \bar{X}^*)^2 \qquad \bar{X}^* = (\nu + 1)^{-1} \sum_{i=1}^{\nu+1} X_i^*$$

Frequently, however, S is based on the same sample X_1, ..., X_n, and then $\{X_{(i;n)} - \bar{X}\}/S$ is not Studentized, since $X_{(i;n)} - \bar{X}$ and S are not then independent. However, $\{X_{(i;n)} - \bar{X}\}/S$, $\bar{X}$, and S are mutually independent in this situation (see [8.1.2.2]). Some writers (e.g., Berman, 1962, p. 154) have used the term *Studentized* in the latter sense.

[*8.3.1.2*] In a normal sample of at least two observations, $X_{(r;n)} - \bar{X}$ and $\bar{X}$ are independent (Kendall and Stuart, 1977, p. 370); $r = 1, 2, \ldots, n$.

[8.3.2.1] In this subsection, $\bar{X}$ is the sample mean, whatever the sample size may be, and the order statistics come from a $N(\mu,1)$ parent.

If $n = 2$, then $X_{(2;2)} - \bar{X} = \{X_{(2;2)} - X_{(1;2)}\}/2$, with pdf $g_2(x)$ given by

$$g_2(x) = 2\pi^{-1/2} \exp(-x^2) \qquad x > 0$$

a half-normal distribution (see McKay, 1935, p. 469; [2.6.1.3]).

If $n = 3$, then $X_{(3;3)} - \bar{X}$ has pdf $g_3(x)$ given by (McKay, 1935, p. 469)

$$g_3(x) = 3(3/\pi)^{1/2} \exp(-3x^2/4)\{\Phi(3x/\sqrt{2}) - 1/2\}, \qquad x > 0$$

Recursively, if $X_{(n;n)} - \bar{X}$ has pdf $g_n(x)$ and cdf $G_n(x)$, then

$$g_n(x) = \frac{n}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{n}{n-1}x^2\right)\sqrt{\frac{n}{n-1}}\,G_{n-1}\left(\frac{nx}{n-1}\right), \qquad n = 2, 3, \ldots$$

See also David (1970, pp. 72-73) and Grubbs (1950, pp. 41-44).

[8.3.2.2] For samples of size n from a $N(\mu,1)$ parent, let $X_{(n;n)} - \bar{X}$ have cdf $G_n(x)$. Then, when x is very small (McKay, 1935, p. 469),

$$1 - G_n(x) \simeq n[1 - \Phi\{x(n/(n-1))^{1/2}\}]$$

This approximation can be used to obtain a good approximation to the upper percent points based on the usual range of probabilities. Thus, if $G_n(x_\alpha) = 1 - \alpha$, tables of $N(0,1)$ probabilities can be used to solve for x_α from the relation

$$\alpha/n \simeq 1 - \Phi[x_\alpha\{n/(n-1)\}^{1/2}]$$

If this approximation is x'_α, then (David, 1956a, p. 86)

$$\alpha - \tfrac{1}{2}(n-1)\alpha^2/n < 1 - G_n(x'_\alpha) < \alpha$$

In particular, for all n,

$$0.04875 < 1 - G_n(x'_{0.05}) < 0.05$$

$$0.00995 < 1 - G_n(x'_{0.01}) < 0.01$$

A "generally very accurate second approximation" $\tilde{x}_\alpha$ follows from the relation

$$\alpha + \frac{1}{2}(n - 1)\alpha^2/n \simeq n - n\Phi[\tilde{x}_\alpha\{n/(n - 1)\}^{1/2}]$$

[*8.3.3*] Based on a $N(\mu,\sigma^2)$ parent, let the cumulants of $X_{(i;n)} - \bar{X}$ be $\kappa'^{(k)}_{(i)}$, and let those of $X_{(i;n)}$ be $\kappa^{(k)}_{(i)}$; k = 1, 2, ... Then (David, 1970, p. 86)

$$\kappa'^{(1)}_{(i)} = E\{X_{(i;n)} - \bar{X}\} = \kappa^{(1)}_{(i)} - \mu$$

$$\kappa'^{(2)}_{(i)} = \text{Var}\{X_{(i;n)} - \bar{X}\} = \kappa^{(2)}_{(i)} - \sigma^2/n$$

$$\kappa'^{(k)}_{(i)} = \kappa^{(k)}_{(i)}, \qquad k = 3, 4, \ldots$$

[*8.3.4*] Grubbs (1950, pp. 31-37) gives a table of the cdf $G_n(x)$ of $(X_{(n;n)} - \bar{X})$ for an $N(\mu,\sigma^2)$ parent, to 5 decimal places, for x = 0.00(0.05)4.90 and n = 2(1)25. He also tables (p. 45) upper percent points of the distribution for n = 2(1)25 and α = 0.10, 0.05, 0.01, and 0.005, to 3 decimal places; and (p. 46) the mean, standard deviation, skewness, and kurtosis for n = 2(1)15, 20, 60, 100, 200, 500, 1000, also to 3 decimal places. See Table 8.1 for coverage in some standard sources.

[*8.3.5*] Let $Y_{(i)} = \{X_{(i;n)} - \bar{X}\}/S$, where $S^2 = \Sigma(X_i - \bar{X})^2/(n-1)$, so that $Y_{(i)}$ is not externally Studentized; the parent population is $N(\mu,\sigma^2)$.

Then $Y_{(i)}$, $\bar{X}$, and S are mutually independent ([8.1.2.2]), from which it follows for the moments about zero of $Y_{(i)}$ that (see, for example, Quesenberry and David, 1961, p. 381)

$$E\{Y^r_{(i)}\} = E\{(X_{(i;n)} - \bar{X})^r\}/E(S^r)$$

[*8.3.6*] With the notation and conditions in [8.3.5], let $Y = (X_i - \bar{X})/S$, one of the unordered Y terms (see [5.4.3]). Then (David, 1970, pp. 86-87)

$$Y_{(n-1)} \le \sqrt{\{(n-1)(n-2)/(2n)\}}$$

$$\Pr\{Y_{(n)} > y \ge \sqrt{(n-1)(n-2)/(2n)}\} = n \Pr(Y > y)$$

$$\Pr\{Y_{(1)} < y \le -\sqrt{(n-1)(n-2)/(2n)}\} = n \Pr(Y < y)$$

[*8.3.7*] Upper 10, 5, 2.5, and 1 percent points for $Y_{(n)}$ as defined in [8.3.5] are tabled for n = 3(1)25 to 3 decimal places by Grubbs (1950, p. 29). Pearson and Chandra Sekar (1936, p. 318) table some corresponding percent points for n = 3(1)19, but of $\sqrt{n/(n-1)}Y_{(n)}$. See also Table 8.1.

[*8.3.8*] Let a_n be defined as in [8.1.7.3] and S^2 as in [8.3.5]. Then the normed extreme deviate

$$\sqrt{2 \log n}\left\{\frac{X_{(n;n)} - \bar{X}}{S} - a_n\right\}$$

has the limiting distribution as $n \to \infty$ with cdf $\exp(-e^{-x})$, $-\infty < x < \infty$ (David, 1970, p. 213). This holds for a $N(\mu,\sigma^2)$ parent, and is the same extreme-value distribution given in [8.1.7] for $X_{(n;n)}$, suitably normed.

[*8.3.9*] Consider the absolute difference ratios $|X_1 - \bar{X}|/S$, ..., $|X_n - \bar{X}|/S$, where S is defined in [8.3.5], and let the ordered set of these statistics be

$$0 \le U_{(1)} \le U_{(2)} \le \cdots \le U_{(n)}$$

Then (Pearson and Chandra Sekar, 1936, pp. 313-315) the $U_{(i)}$ terms are bounded almost surely as follows:

If n - i is even and i > 1,

$$U_{(i)} \le \left\{\frac{n-1}{n-i+1+(i-1)^{-1}}\right\}^{1/2}$$

If n is odd,

$$U_{(1)} \le \left\{\frac{(n - 1)^2}{n^2 + n}\right\}^{1/2}$$

If n - i is odd,

$$U_{(i)} \le \sqrt{\frac{n - 1}{n - i + 1}}$$

Borenius (1958, pp. 152-156) gives formulas for the pdf of $\sqrt{n/(n - 1)}U_{(n)}$.

[*8.3.10*] The *Studentized extreme deviate* is $\{X_{(n;n)} - \bar{X}\}/S'$ or $\{\bar{X} - X_{(1;n)}\}/S'$, where S' is defined in [8.3.1] from an independent sample; the parent population is $N(\mu,\sigma^2)$ and S' has ν degrees of freedom. An approximation to the upper 100α percent point is given by (David, 1970, p. 86)

$$(\sqrt{(n - 1)/n})t_{\nu;\alpha/n}$$

where $t_{\nu;\alpha/n}$ is the upper $100\alpha/n$ percent point of a Student t rv with ν degrees of freedom.

Tables of upper percent points of the Studentized extreme deviate are listed in Table 8.1, from standard sources. They also appear in David (1956b, p. 450) to 2 decimal places for α = 0.10, 0.05, 0.025, 0.01, and 0.005, n = 3(1)10, 12, and ν = 10(1)20, 24, 30, 40, 60, 120, ∞. Corresponding values appear to 1 decimal place for α = 0.001. The use of the approximation above leads to errors no greater than 0.07 for the values covered in David's tables. The third edition of *Biometrika Tables for Statisticians*, vol. 1 (Pearson and Hartley, 1966, table 26) contains corrections by David (1956b) to some earlier approximate values when $\nu > 20$.

Pillai (1959, p. 473) gives upper 5 and 1 percent points to 2 places for n = 2(1)10, 12, and ν = 3(1)10; also to 1 decimal place for ν = 2 and the nearest integer for ν = 1; his values are reproduced in Pearson and Hartley (1966), except for ν = 1(1)4 and for n = 2.

[*8.3.11*] Halperin et al. (1955, pp. 187-188) give tables to 2 decimal places of upper and lower bounds for the upper 5 and 1 percent points of the maximum absolute Studentized deviate, i.e., of

$$\max\left\{\frac{X_{(n;n)} - \bar{X}}{S'}, \frac{\bar{X} - X_{(1;n)}}{S'}\right\}$$

where S' is defined in [8.3.1]. These tables are for n = 3(1)10, 15, 20, 30, 40, 60, and ν = 3(1)10, 15, 20, 30, 40. 60, 120, ∞.

[*8.3.12*] Quesenberry and David (1961, pp. 379-390) considered statistics $\{X_{(i;n)} - \bar{X}\}/S^*$, where for a $N(\mu,\sigma^2)$ parent,

$$S^{*2} = (n - 1)S^2 + \nu S_\nu^2$$

S^2 is the sample variance and S_ν^2 is an independent mean-square estimate of σ^2 based on ν degrees of freedom. Let

$$V = \max[\{X_{(i;n)} - \bar{X}\}]/S^*$$

$$V^* = \max[|X_{(i;n)} - \bar{X}|/S^*], \qquad i = 1, 2, \ldots, n$$

V is a pooling of information in $Y_{(n)}$ and $U_{(n)}$, defined above; but note that S^{*2} is not divided by its degrees of freedom, $n + \nu - 1$.

Then (Quesenberry and David, 1961, p. 381) $E(V^r) = E(\{X_{(n;n)} - \bar{X}\}^r)/E(S^{*r})$. Tables of percent points of V (Quesenberry and David, 1961, pp. 388-390; Pearson and Hartley, 1966, pp. 187-188) for α = 0.05 and n = 3(1)10, 12, 15, 20, ν = 0(1)10, 12, 15, 20, 24, 30, 40, 50 are given to 3 decimal places, and for α = 0.01 to 4 decimal places. Analogous tables of percent points of V* are given, but to 3 decimal places; in Quesenberry and David (1961) these appear as bounds, with a condensed version in Pearson and Hartley (1966).

8.4 THE SAMPLE RANGE

[*8.4.1.1*] The range W of a sample of size n is $X_{(n;n)} - X_{(1;n)}$. For a N(0,1) parent, the pdf g(w;n) and cdf G(w;n) are given by (David, 1970, pp. 10-11; Harter, 1969a, pp. 4, 13)

$$g(w;n) = n(n-1)\int_{-\infty}^{\infty} [\Phi(x+w) - \Phi(x)]^{n-2}\phi(x)\phi(x+w)\,dx, \quad w > 0$$

$$G(w;n) = n\int_{-\infty}^{\infty} [\Phi(x+w) - \Phi(x)]^{n-1}\phi(x)\,dx$$

$$= n\int_{0}^{\infty} [\{\Phi(x+w) - \Phi(x)\}^{n-1} + \{\Phi(w-x)$$

$$+ \Phi(x) - 1\}^{n-1}]\phi(x)\,dx, \quad w > 0$$

Hartley (1942, p. 342):

$$G(2w;n) = \{2\Phi(w) - 1\}^n + 2n\int_0^{\infty} \{\Phi(x+w)$$

$$- \Phi(x-w)\}^{n-1}\phi(x+w)\,dx$$

[*8.4.1.2*] Special cases of the pdf and cdf in [8.4.1.1] are given by (Harter, 1969a, p. 4)

$$g(0;2) = 1/\sqrt{\pi} \qquad g(0;n) = 0 \text{ if } n > 2$$

If $w > 0$ (Lord, 1947, p. 44),

$$g(w;2) = \sqrt{2}\phi(w/\sqrt{2}) \qquad G(w;2) = 2\Phi(w/\sqrt{2}) - 1$$

McKay and Pearson (1933, p. 417):

$$g(w;3) = 3\sqrt{2}[2\Phi(w/\sqrt{6}) - 1]\phi(w/\sqrt{2})$$

Bland et al. (1966, p. 246):

$$G(w;3) = 1 - 12T(w/\sqrt{2}, 1/\sqrt{3})$$

where $T(h,a) = \int_0^a \{\phi(h)\phi(hx)/(1+x^2)\}\,dx$, and $T(\cdot,\cdot)$ has been tabled by Owen (1956, pp. 1080-1087) (see [10.2.3]).

When $n = 3$, W has the same distribution as $3\sqrt{3}\,M/2$ or $3M_1$, where $M = \Sigma_{i=1}^3 |X_i - \bar{X}|/3$ and $M_1 = \Sigma_{i=1}^3 |X_i - X_{(2;3)}|/3$.

When $n = 4$ (Bland et al., 1966),

$$G(w;4) = 12\Phi(w/\sqrt{2}) - 1 - 48S(w/\sqrt{2}, 1/\sqrt{3}, \sqrt{2}) - 48S(w/\sqrt{2}, 1/\sqrt{3}, 1/\sqrt{2})$$

where $S(y,a,b) = \int_{-\infty}^{y} T(ax,b)\phi(x)\,dx$, and $S(\cdot,\cdot,\cdot)$ is tabulated by Steck (1958, pp. 790-799).

Bland et al. (1966) give an expression for $G(w;5)$ and $g(w;5)$.

[*8.4.2.1*] The *moments* of the range can be derived from those of $X_{(n)}$ and $X_{(1)}$, discussed above in Section 8.2. However, for some small sample sizes, we give the following results for a N(0,1) parent (Ruben, 1956, p. 460; Sarhan and Greenberg, 1962, p. 185):

$$n = 2: \quad E(W) = 2/\sqrt{\pi} \qquad E(W_2^2) = 2 \qquad E(W_2^3) = 8/\sqrt{\pi}$$

$$n = 4: \quad E(W) = 6 \arccos(-1/3)/(\pi^{3/2})$$

$$= 6\{(1/2)\pi + \arcsin(1/3)\}/\pi^{3/2}$$

$$E(W^2) = 2 + 6\{1 + (1/\sqrt{3})\}/\pi$$

From [8.2.9], Tables 8.2A and 8.2B, we have

$$n = 3: \quad E(W) = 3/\sqrt{\pi} \qquad E(W^2) = 2 + (3\sqrt{3}/\pi)$$

$$n = 5: \quad E(W) = \frac{5}{2\sqrt{\pi}}\left\{1 + \frac{6 \arcsin(1/3)}{\pi}\right\}$$

$$E(W^2) = 2 + \frac{5\sqrt{3}}{2\pi} + \frac{5\sqrt{3}}{\pi^2} \arcsin(1.4) + \frac{60}{\pi^2} \arcsin \frac{1}{\sqrt{6}}$$

Generally, when n = 3 (McKay and Pearson, 1933, p. 417),

$$E(W^k) = \frac{3}{\pi}\, 2^{k+1}\Gamma((1/2)k + 1) \int_0^{\pi/6} \cos^k\theta \, d\theta$$

[*8.4.2.2*] With the notation of [8.4.2.1] and of [8.2.1], for a N(0,1) parent (David, 1970, p. 28),

$$E(W) = 2\mu_{(n;n)} \qquad \mathrm{Var}(W) = 2\{\sigma_{(n,n;n)} - \sigma_{(n,1;n)}\}$$

[*8.4.3*] Tables giving the cdf, percentage points, and moments of the range are listed from several standard sources in Table 8.1. Barnard (1978, pp. 197-198) gives a Fortran computer program for the cdf of W, from a N(0,1) parent. Harter (1960, pp. 1125-1131) gives tables of percent points and of moments of W; the coverage is that given in Table 8.1 for Harter's later tables (Harter, 1969a, pp. 372-377).

[*8.4.4.1*] An *approximation to the pdf of* W for a $N(\mu,1)$ parent is the *asymptotic distribution* given by Cadwell (1953a, p. 607); this is $g(y;n)$, where

$$g(2y;n) \simeq \frac{n(n-1)\sqrt{\pi}\{\phi(y)\}^2\{2\Phi(y)-1\}^{n-3/2}}{\{2\Phi(y)-1-(n-2)\phi'(y)\}^{1/2}}$$

Accuracy, as measured by the lower moments, shows this to improve upon earlier approximations; an improvement upon it is also indicated. See David (1970, pp. 211-212).

[*8.4.4.2*] David (1970, pp. 141-142) discusses three approximations to the distribution of W/σ, based on chi or chi-square. The most accurate is given by (Cadwell, 1953b, pp. 337-338, 344)

$$W/\sigma \simeq (\chi_\nu^2/c)^\alpha$$

where ν, c, and α are appropriate constants. Cadwell's table 1 gives values of ν, $1/\alpha$, and log c for n = 2(1)20.

[*8.4.4.3*] Cadwell (1954, pp. 803-805) gives an *approximation to the cdf* of the range W from N(0,1) samples. This is $G(w;n)$, where

$$G(2y;n) \simeq \{2\Phi(y)-1\}^n + 2na\{2\Phi(y)-1\}^{n-1}$$
$$\times \exp\{-(1/2)y^2(1-a^2)\{\Phi 1 - \Phi(ay)\}$$
$$a^{-2} = 1 + (n-1)y\phi(y)/\{\Phi(y)-1/2\}$$

When n = 20, 60, or 100, the maximum error of this approximation is 0.0031, 0.0040, or 0.0043, respectively.

An improved approximation to $G(w;n)$ results if the last bracket $1 - \Phi(ay)$ is replaced by

$$\{1 - \Phi(ay) - (n-1)a^4P(y)Q(ay)\}$$

$$P(x) = \frac{x^2}{8}\left\{\frac{\phi(x)}{1/2-\Phi(x)}\right\}^2 + \frac{x^3-3x}{24}\left\{\frac{\phi(x)}{1/2-\Phi(x)}\right\}$$

$$Q(x) = (x^4+6x^2+3)\{1-\Phi(x)\} - (x^3+5x)\phi(x)$$

If $n = 20$, 60, or 100, the maximum error is only -0.00052, -0.00070, or -0.00075, respectively. Cadwell (1954, p. 804) notes that errors in each of these approximations will initially increase with n (from $n = 2$) and then fall asymptotically to zero.

[*8.4.4.4*] Johnson (1952, p. 418) gives approximations to the cdf $G(y;n)$ which are useful when $n \le 15$ and $y \le 2$. He also gives approximations to the quantiles y_α, where $G(y_\alpha;n) = 1 - \alpha$, but these hold only when $n \le 5$ and for the lower quantiles $(1 - \alpha \le 0.025)$. The simpler of these is given by

$$y_\alpha \simeq \sqrt{2\pi}(\alpha/\sqrt{n})^{1/(n-1)}$$

[*8.4.5*] For the same values of the constants a_n defined in [8.1.7.3], i.e.,

$$a_n = \sqrt{2 \log n} - (1/2)(\log \log n + \log (4\pi))/\sqrt{2 \log n}$$

and with a $N(\mu,1)$ parent, the limiting distribution of the range W is given by (Galambos, 1978, p. 109)

$$\lim_{n\to\infty} \Pr[\sqrt{2 \log n}\{W - 2a_n\} < x] = \int_{-\infty}^{\infty} e^{-y} \exp(-e^{-y} - e^{y-x})\, dy$$

The range W and the extreme value $X_{(n;n)}$ are asymptotically independent (Galambos, 1978, p. 109).

[*8.4.6*] As in [8.3.1], let S or S' be an estimate of σ with ν degrees of freedom. Then W/S' is the *Studentized range* if S' is based on a second independent sample of size $\nu + 1$; frequently S is based on the same sample, and W/S is not "Studentized," in the sense that W and S are not independent. Sometimes W/S is referred to as *internally Studentized*, and W/S' as *externally studentized*. In both cases, however, the statistic has a distribution free both of μ and of σ^2, for a general $N(\mu,\sigma^2)$ parent.

[*8.4.7*] The distribution of W/S is studied by David et al. (1954, 482-493). The rvs W/S, $\bar{X}$, and S are mutually independent ([8.1.2.2]). The authors constructed a table of percent points of W/S, extended along with moments by Pearson and Stephens (1964, pp.

484-487). The latter quote exact results when n = 3: if H is the cdf of W/S, then

$$x = 2\cos\{(1 - H(x))\pi/6\}, \qquad \sqrt{3} \le x \le 2$$

[*8.4.8*] The following bounds hold for samples from any distribution:

$$W/S \le \sqrt{2(n-1)}$$

$$W/S \ge \begin{cases} \sqrt{2(n-1)/n}, & n \text{ even} \\ \sqrt{2n/(n+1)}, & n \text{ odd} \end{cases} \quad \text{(Thomson, 1955, p. 268)}$$

$$\{X_{(n-1;n)} - X_{(1;n)}\}/S \le \sqrt{3(n-1)/2} \quad \text{(David et al., 1954, p. 492)}$$

[*8.4.9*] The moments of W/S satisfy (David et al., 1954, p. 483)

$$E\{(W/S)^r\} = E(W^r)/E(S^r)$$

These moments may thus be determined from those of W (see [8.4.2.3]) and of S (see [5.3.5]). For a $N(\mu,\sigma^2)$ parent (David, 1970, p. 72),

$$E\{(W/S)^r\} = \{(n-1)/2\}^{(1/2)r}[\Gamma(n-1)/2)/\Gamma((1/2)(n-1 + r))]E\{(W/\sigma)^r\}$$

[*8.4.10*] The cdf of the Studentized range Q, where Q = W/S', as defined in [8.4.6], is $G(q;\nu,n)$, where (Harter, 1969a, p. 18)

$$G(q;\nu,n) = 2\{[(1/2)\nu]^{(1/2)\nu}/\Gamma[(1/2)\nu]\}\int_0^\infty x^{\nu-1}e^{-(1/2)\nu x^2} \times G_0(qx;n)\,dx$$

and $G_0(\cdot;n)$ is the cdf of the range W (see [8.4.1, 5]). Pillai (1952, p. 195) gives the pdf of Q in the form of an infinite series.

[*8.4.11*] The moments of the Studentized range Q satisfy (David et al., 1954, p. 483; David, 1970, p. 71)

$$E(Q^r) = E(W^r)E(S'^{-r}), \qquad 0 \le r < \nu$$

$$= [(1/2)\nu]^{(1/2)r}\Gamma((\nu - r)/2)E\{(W/\sigma)^r\}/\Gamma[(1/2)\nu], \qquad 0 \le r < \nu$$

[*8.4.12*] Tables of the cdf and percentage points of Q and of W/S are listed in Table 8.1 for standard sources. See also Pearson and Stephens (1964, pp. 485-486) for tables of percent points and moments of Q, and Harter (1960, pp. 1132-1143) for percent points of Q [$n = 2(1)20(2)40(10)100$; $\alpha = 0.10, 0.05, 0.025, 0.01, 0.005, 0.001$; $\nu = 1(1)20, 24, 30, 40, 60, 120, \infty$; to 3 decimal places or 4 significant figures].

[*8.4.13*] The distribution of the statistic $(\tilde{X} - \mu)/W$ is discussed by Daly (1946, pp. 71-74). Tables of percent points are listed in Table 8.1.

8.5 QUASI-RANGES

[*8.5.1*] The *rth quasi-range* of a random sample of size n is W_r', where $W_r' = X_{(n-r;n)} - X_{(r+1;n)}$, and is used to estimate the standard deviation σ. The sample range is W_0' and the standardized quasi-range is W_r'/σ, denoted by W_r.

[*8.5.2*] The pdf and cdf of W_r for a $N(\mu,\sigma^2)$ parent are given by $g_r(w;n)$ and $G_r(w;n)$, where

$$g_r(w;n) = \frac{n!}{(n - 2r - 2)!(r!)^2} \int_{-\infty}^{\infty} [\Phi(x)]^r [1 - \Phi(x + w)]^r$$

$$\cdot [\Phi(x + w) - \Phi(x)]^{n-2r-2} \phi(x)\phi(x + w)\, dx$$

(Cadwell, 1953a, p. 604; Harter, 1969b, p. 12)

$$G_r(w;n) = \int_{-\infty}^{\infty} \sum_{k=0}^{r} \frac{n(n - 1)\cdots(n - 2r + k)}{r!(r - k)!} [1 - \Phi(x + w)]^{r-k}$$

$$\cdot [\Phi(x + w) - \Phi(x)]^{n-2r+k-1} [\Phi(x)]^r \phi(x)\, dx$$

(Harter, 1969b, p. 12)

Cadwell (1953a, p. 605) gives an asymptotic series for the pdf, viz.,

$$g_r(w;n) \simeq C\{\phi(x)\}^2 \{2\Phi(x) - 1\}^{n-2r-2} \{1 - \Phi(x)\}^{2r} k[1 - (n - 2r - 2)(3\theta'^2 - \theta''')k^4/8 - \cdots]$$

where

$$k^{-2} = 2 + 2r(\psi^2 + \psi') - (n - 2r - 2)\theta'$$

$$\theta^{(\ell)}(x) = \frac{(d/dx)^{\ell}\phi(x)}{\Phi(x) - \frac{1}{2}}$$

$$\psi^{(\ell)}(x) = \frac{(d/dx)^{\ell}\phi(x)}{1 - \Phi(x)}$$

and C is chosen to make the total probability of the approximate pdf equal to one. Cadwell gives an expression for the next term, of $0(n^{-2})$.

[*8.5.3*]

$$E(W_r) = 2(r + 1)\binom{n}{r+1}\int_{-\infty}^{\infty} x[1 - \Phi(x)]^r[\Phi(x)]^{n-r-1}\phi(x)\, dx$$

$$= 2E\{X_{(n-r)}\}$$

$$E(W_r^2) = \frac{n!}{(n - 2r - 2)!(r!)^2}\int_{-\infty}^{\infty}\left\{\int_0^{\infty} H(x,y)\, dy\right\}[\Phi(x)]^r\phi(x)\, dx$$

$$H(x,y) = y^2[1 - \Phi(x + y)]^r[\Phi(x + y) - \Phi(x)]^{n-2r-2}\phi(x + y)$$

(Harter, 1969b, pp. 3-4)

[*8.5.4*] Extensive tables of the cdf, percent points and first two moments appearing in Harter (1969b, pp. 136-319) are listed in Table 8.1. Harter (1959, pp. 982-987) tables $E(W_r)$ to 6 decimal places and $\mathrm{Var}(W_r)$ to 5 places for n = 2(1)100, r = 0(1)8. Cadwell (1953a, p. 610) tables the first four moments of the first quasi-range W_1, as well as upper and lower 5, 2.5, 1, and 0.1 percent points, for n = 10(1)30, to 2 (percent points) and 3 or 4 (moments) decimal places.

8.6 MEDIAN AND MIDRANGE

[*8.6.1*] If n = 2m + 1, where m is an integer, the *sample median* is $X_{(m+1;n)}$, i.e., $X_{((n+1)/2;n)}$; if n = 2m, it is $(1/2)\{X_{(m;n)} + X_{(m+1;n)}\}$, i.e., $(1/2)\{X_{(n/2;n)} + X_{((n/2)+1;n)}\}$. The *sample midrange* is $(1/2)\{X_{(n;n)} + X_{(1;n)}\}$.

[*8.6.2.1*] If n is odd, the distribution of the median $\tilde{X}$ is easier to express. Let the pdf of $\tilde{X}$ for a sample of size n from a N(0,1) parent be g(y;n).

(a) Then (Kendall and Stuart, 1977, p. 348)

$$g(y;2m+1) = \frac{(2m+1)!}{m!m!}[\Phi(y)\{1-\Phi(y)\}]^m \phi(y)$$

and g(y;2m + 1) is approximately proportional to (Cadwell, 1952, p. 208)

$$\{1 + 2(\pi - 3)my^4/(3\pi^2)\} \exp\{-(1/2)y^2 - (2my^2/\pi)\}$$

(b) If n = 2m, g(y;2m) is proportional to (Cadwell, 1952, p. 209)

$$\exp(-y^2) \int_0^\infty [\Phi(y-x)\{1-\Phi(y+x)\}]^{m-1} \exp(-x^2)\, dx$$

[*8.6.2.2*] Let the parent population be N(μ,1), and let n = 2m + 1. The cdf H(y;n) of the rv $\{X_{(m+1;n)} - \mu\}/\sigma_m$ satisfies the following equalities, where $\sigma_m^2 = (1/2)\pi/(2m+1) = (1/2)\pi/n$ (Chu, 1955, pp. 114-115):

$$\begin{aligned} 0.9929(1 + m/8)\sqrt{1 - 1/(2m+2)}\{\phi(y) - \phi(-x)\} \\ \le H(y;n) - H(-x;n) \\ \le (1 + m/8)\sqrt{1 + 1/(2m)}\{\phi(y) - \phi(-x)\}, \quad y > 0,\ x > 0 \end{aligned}$$

Note that σ_m^2 is the asymptotic variance of the median, so that H is the cdf of an "asymptotically standardized" rv. See [8.6.4] below.

[*8.6.3.1*] The mean of the sample median is zero, as is the third moment, for a N(0,1) parent.

If n = 2m + 1,

$$\text{Var}\{X_{(m+1;n)}\} = 1 + \frac{(2m+1)!}{4\pi m!m!} \sum_{j=0}^{m} (-1)^j \binom{m}{j} (m+j)(m+j-1)$$

$$\cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \{\Phi(x)\}^{m+j-2} \exp \frac{-3x^2}{2}\, dx$$

(Kendall and Stuart, 1977, p. 349)

$$\text{Var}\{X_{(m+1;n)}\} = \frac{\pi}{2(n + 2)} + \frac{\pi^2}{4(n + 2)(n + 4)} + O(n^{-3})$$

(Kendall and Stuart, 1977, p. 351)

when n is large. If σ_m^2 is defined as in [8.6.2.2], then (Chu and Hotelling, 1955, p. 601)

$$\text{Var}\{X_{(m+1;n)}\} = \sigma_m^2[1 - (2 - (1/2)\pi)n^{-1} - (3\pi - 4 - 13\pi^2/24)n^{-2}] + O(n^{-4}), \qquad n \geq 7$$

Further (Chu and Hotelling, 1955, p. 602),

$$B_m\{1 - 1/(n + 1)\}^{3/2} \leq \text{Var}\{X_{(m+1;n)}\}/\sigma_m^2 \leq B_m\{1 + 1/(n - 1)\}^{3/2}$$

$$B_m = \frac{(2m + 1)!}{m!m!} 2^{-n}\sqrt{2\pi/n}$$

In the latter inequalities, if $n \geq 4$,

$$1 + \frac{1}{8m} - \frac{7m + 3}{24m^2(2m + 1)} < B_m < 1 + \frac{1}{8m} + \frac{1}{16m(8m - 1)}$$

See also Chu and Hotelling (1955, pp. 601-604), in particular their expressions (49) and (56) and remarks 1 and 2.

[*8.6.3.2*] Hodges and Lehmann (1967, pp. 927-928) have given a large-sample approximation to the variance of the sample median. Whether $n = 2m$ or $n = 2m + 1$, this is given for a $N(\mu,\sigma^2)$ parent by

$$\text{Var}(\tilde{X}) \simeq \frac{\pi\sigma^2}{4m}\left[1 - \frac{(6 - \pi)\pi\sigma^2}{2m}\right]$$

Insofar as this formula is accurate, it pays to base the median on an even number $(n = 2m)$ of observations, rather than the next highest odd number $(2m + 1)$ (Hodges and Lehmann, 1967, p. 928). Dixon (1957, p. 807) tables the efficiency of $\tilde{X}$ with respect to the sample mean $\bar{X}$, and the exact variance of $\tilde{X}$, for $n = 2(1)20, \infty$.

[*8.6.4*] For a sample of size $n = 2m + 1$ from a $N(\mu,\sigma^2)$ parent, the distribution of $\tilde{X}$ is asymptotically normal with mean μ and variance $\pi\sigma^2/\{2(2m + 1)\} = (1/2)\pi(\sigma^2/n)$ (Chu, 1955, p. 114).

[*8.6.5.1*] The ratio c_n is of interest, where

$$c_n = \frac{\text{standard deviation of sample median } \tilde{X} \text{ of n values}}{\text{standard deviation of sample mean } \bar{X} \text{ of n values}}$$

Asymptotically, $c_n \to \sqrt{\pi/2}$ as $n \to \infty$. If $n = 2m + 1$ (Cadwell, 1952, p. 208; Pearson, 1931, p. 363),

$$c_n \simeq \left\{\frac{\pi(2m + 1)}{\pi + 4m}\right\}^{1/2}\left\{1 + \frac{4(\pi - 3)m}{(\pi + 4m)^2}\right\}$$

$$= \sqrt{\pi/2} - \frac{0.2690}{n} - \frac{0.0782}{n^2} + \cdots$$

If $n = 2m$ (Cadwell, 1952, pp. 209-210),

$$c_n \simeq \left\{\frac{\pi m}{\pi + 2m - 2}\right\}^{1/2}\left\{1 - \frac{4 - \pi}{4(\pi + 2m - 2)} + \frac{(\pi - 3)(m - 1)}{(\pi + 2m - 2)^2}\right\}$$

$$= \sqrt{\pi/2} - \frac{0.8941}{n} + O(n^{-2})$$

Kendall and Stuart (1977, p. 350) give a table of values of c_n for $n = 2(2)12, 20, \infty$ to 3 decimal places, showing that c_n increases from 1.000 to 1.214 as n increases from 2 to 20, and that $c_n \to 1.253$ as $n \to \infty$.

[*8.6.5.2*] Hodges and Lehmann (1967, pp. 926-927) have tabled the exact and approximate efficiency e(n) of $\tilde{X}$ for $n = 1(1)20, \infty$, where

$$e(n) = \text{Var}(\bar{X})/\text{Var}(\tilde{X})$$

for samples of size n from a $N(\mu,\sigma^2)$ parent, $\bar{X}$ being the sample mean. The approximate efficiency is given by

$$e(n) \simeq 2/\pi + a/n$$

where $a = 4/\pi - 1$ if n is odd, and $a = 6/\pi - 1$ if n is even. The efficiency of $\tilde{X}$ decreases from 1.0 to 0.637 as n increases from one to ∞, but does so through persistently higher values when n is odd. See also [8.6.13].

[*8.6.6*] The fourth moment ratio $\beta_2 = \mu_4/\mu_2^2$ of $\tilde{X}$ when n is large satisfies (Cadwell, 1952, pp. 208, 209)

$$\beta_2 \simeq 3 + \frac{16(\pi - 3)m}{(\pi + 4m)^3}, \qquad n = 2m + 1$$

$$\beta_2 \simeq 3 + \frac{4(\pi - 3)(m - 1)}{(\pi + 2m - 2)^2}, \qquad n = 2m$$

[*8.6.7*] Let $\bar{X}$ and $\tilde{X}$ be the sample mean and median, respectively, from a sample of size n from an $N(\mu,\sigma^2)$ population. Then $(\bar{X},\tilde{X})$ have a joint bivariate normal distribution asymptotically, as $n \to \infty$, where (Wilks, 1962, p. 275)

$$E(\bar{X}) = \mu \qquad E(\tilde{X}) = \mu$$

$$\mathrm{Var}(\bar{X}) = \sigma^2/n \qquad \mathrm{Var}(\tilde{X}) \simeq (1/2)\pi\sigma^2/n$$

$$\mathrm{Cov}(\bar{X},\tilde{X}) = \sigma^2/n$$

[*8.6.8*] The cdf H(y;n) of the *midrange* $(1/2)\{X_{(1)} + X_{(n)}\}$ is given by (Gumbel, 1958, p. 109)

$$H(y;n) = n\int_{-\infty}^{y} [\Phi(2y - x) - \Phi(x)]^{n-1}\phi(x)\,dx$$

(Notice that Gumbel refers to $X_{(1)} + X_{(n)}$ as the "midrange.")

[*8.6.9*] The pdf h(y;n) of the midrange is given by

$$h(y;n) = n(n - 1)\pi^{-1}\exp(-(1/2)ny^2)\sum_{i=0}^{\infty} B_i y^{2i}$$

where Pillai (1950, pp. 101-102) gives expressions for B_0, B_1, and B_2, and tables the first five B coefficients when $3 \le n \le 10$.

When n = 2,

$$h(y;2) = \pi^{-1/2}\exp(-y^2) = \phi(\sqrt{2}y)$$

[*8.6.10*] The midrange of a sample of size n from a $N(\mu,\sigma^2)$ parent converges in probability to μ as $n \to \infty$. The mean of the midrange is μ, and the asymptotic variance is $\pi^2\sigma^2/(24 \log n)$ (Kendall and Stuart, 1977, p. 365); the latter tends to zero more slowly than σ^2/n, the variance of $\bar{X}$.

[*8.6.11*] The asymptotic distribution of Gumbel's "reduced" midrange for a N(0,1) parent has for its pdf the logistic form (Galambos, 1978, p. 109; Gumbel, 1958, p. 311)

$$\exp(-y)\{1 + \exp(-y)\}^{-2}$$

$$\lim_{n\to\infty} \Pr[(2\sqrt{2 \log n})1/2\{X_{(1;n)} + X_{(n;n)}\} < x] = (1 + e^{-x})^{-1}$$

This result is an exception to the more common phenomenon that statistical measures of location, at least for symmetric populations, have an asymptotic normal distribution.

[*8.6.12*] The midrange becomes increasingly sensitive to outlying values in the tails of the parent distribution, as the sample size becomes larger, at least when the parent distribution is not restricted to a finite interval. Thus, for a $N(\mu,\sigma^2)$ parent, let

$$d_n = \frac{\text{standard deviation of sample midrange of n values}}{\text{standard deviation of sample mean of n values}}$$

Then, as n increases from two to 20, d_n increases from 1.000 to 1.691, and $d_n \to \infty$ as $n \to \infty$ (Kendall and Stuart, 1977, p. 350). Dixon (1957, p. 807) tables the exact variance of the midrange and its efficiency with respect to the sample mean $\bar{X}$, for n = 2(1)20, ∞.

[*8.6.13*] The *rth midrange* is $(1/2)\{X_{(n-r+1;n)} + X_{(r;n)}\}$; r = 1, 2, ..., [(1/2)n]. Leslie and Culpin (1970, pp. 317-322) have tabled the standard deviations for N(0,1) samples; n = 2(1)21; also the cdfs, n = 3(2)15, 18, and r = 1, 2. For n = 5, 11, 13, and 15, the cdf is given for r = 3, and then n = 18 for r = 4. All tables are to 4 decimal places. See also [8.8.1].

When r = 1 and r = [(1/2)n], this statistic becomes the midrange and median, respectively. As r increases from 1 to [(1/2)n], the standard deviation of the rth midrange decreases, and then increases again. Leslie and Culpin's table 1 (1970, p. 317) indicates that the approximate values of r giving the least variance are given for $n = 3(1)21$ by $r \approx [n/3]$. Mosteller (1946, pp. 387-389) found that for large n, the value of r which minimizes the variance of the rth midrange is given by $r/n \approx 0.2702$. Dixon (1957, p. 807) tables the efficiencies of these optimum rth midranges with respect to the sample mean $\bar{X}$ for $n = 2(1)20, \infty$, when the data come from a $N(\mu,\sigma^2)$ parent. As n increases from 3 to 20, the efficiency decreases from 0.92 to 0.824, with a limiting efficiency of 0.810 as $n \to \infty$.

The sequence of rth midrange statistics is equivalent to the *quasi-medians* of Hodges and Lehmann (1967, p. 928). These are given by $(1/2)\{X_{(m+1-r;2m+1)} + X_{(m+1+r;2m+1)}\}$ or by $(1/2)\{X_{(m-r;2m)} + X_{(m+1+r;2m)}\}$; $r = 0, 1, 2, \ldots, [(1/2)n]$. Hodges and Lehmann (1967, pp. 929-931) give an approximation to the pdf of these statistics; for a $N(\mu,\sigma^2)$ parent, the variance (Hodges and Lehmann, 1967, p. 928) is approximately

$$\frac{\pi\sigma^2}{4m}\left[1 - \frac{4r + 6 - \pi}{4m}\right]$$

whether $n = 2m$ or $n = 2m + 1$, when r is fixed and n is large; see [8.6.3.2].

8.7 QUANTILES

[*8.7.1*] If Z is an N(0,1) random variable, then with the notation used throughout this book, the p-quantile of Z is z_{1-p}, where

$$\Pr(Z \le z_{1-p}) = \Phi(z_{1-p}) = p$$

In a sample of size n, and if np is not an integer, then the order statistic $X_{([np]+1;n)}$ is the unique sample p-quantile, where [np] is the greatest integer less than or equal to p.

If np is an integer, then any quantity in the interval between (and including) $X_{([np];n)}$ and $X_{([np]+1;n)}$ can be used to define the sample p-quantile. In what follows we assume that np is not an integer.

[*8.7.2*] Let $Y = X_{([np+1];n)}$, where the sample has a $N(\mu,\sigma^2)$ parent having cdf $F(\cdot,\mu,\sigma^2)$ and pdf $f(\cdot;\mu,\sigma^2)$. The pdf of Y is $g(y;\mu,\sigma^2)$ where (Cramér, 1946, p. 368)

$$g(y;\mu,\sigma^2) = \binom{n}{[np]}(n - [np])\{F(y;\mu,\sigma^2)\}^{[np]} \cdot \{1 - F(y;\mu,\sigma^2)\}^{n-[np]-1} f(y;\mu,\sigma^2), \qquad -\infty < y < \infty$$

[*8.7.3*] The asymptotic distribution of Y as $n \to \infty$ is normal, with mean $(z_{1-p} - \mu)/\sigma$ and variance $\dfrac{\sigma^2}{\{\phi(z_{1-p})\}^2}\,\dfrac{p(1 - p)}{n}$ (Cramér, 1946, p. 369; Kendall and Stuart, 1977, p. 252; Wilks, 1962, p. 273; see also Caldwell, 1952, pp. 210-211).

[*8.7.4*] Mosteller (1946, pp. 383-384) gives the asymptotic joint distribution of the sample quantiles corresponding to p_1, p_2, ..., p_k, $0 < p_1 < p_2 < \cdots < p_k < 1$. For a $N(\mu,\sigma^2)$ parent, this asymptotic joint distribution (as $n \to \infty$) is multivariate normal, with mean vector $(z_{1-p_1}, z_{i-p_2}, \ldots, z_{1-p_k})$ and

$$\mathrm{Cov}(X_{([np_i]+1;n)}, X_{([np_j]+1;n)}) \approx \frac{p_i(1 - p_j)\sigma^2}{n\phi(z_{1-p_i})\phi(z_{1-p_j})}, \qquad 1 \le i \le j \le k$$

See also Cramér (1946, pp. 369-370), Kendall and Stuart (1977, pp. 253-254).

[*8.7.5*] The median is the 0.5 quartile, and is discussed in Section 8.6. The *semiinterquartile range* in a N(0,1) distribution is $(1/2)(z_{1/4} - z_{3/4})$; the sample semiinterquartile range is $(1/2)\{X_{([3n/4]+1;n)} - X_{([n/4]+1;n)}\}$, and for a $N(\mu,\sigma^2)$ parent has an asymptotic normal distribution as $n \to \infty$ with mean $(1/2)\sigma(z_{1/4} - z_{3/4})$, or $(0.6744898)\sigma$, and asymptotic variance

$$\frac{\sigma^2}{16n} \frac{1}{\{\phi(z_{1/4})\}^2} = \left\{(0.786716)\ \frac{\sigma}{\sqrt{n}}\right\}^2$$

(Cramer, 1946, p. 370; Kendall and Stuart, 1977, p. 254).

[*8.7.6*] Let $X_{(1;n)}$, $X_{(2;n)}$, ..., $X_{(n;n)}$ be order statistics in a combined sample from one or more unknown parent distributions, replaced by rank statistics as in [8.2.11]. As an alternative to normal scores, replace $X_{(r;n)}$ by $\Phi^{-1}[r/(n + 1)]$; in the notation used throughout this book, if $p(r) = r/(n + 1)$, then

$$\Phi(z_{1-p(r)}) = p(r) \qquad \Phi^{-1}(p(r)) = z_{1-p(r)}$$

The *Van der Waerden statistic* is the sum of the scores $\Phi^{-1}(p(r))$ over those values r for measurements belonging to the second of two populations involved in a combined sample. See Lehmann (1975, p. 97) for further discussion and references.

8.8 MISCELLANEOUS RESULTS

[*8.8.1*] Certain statistics discussed in this chapter and in Chapter 5 are uncorrelated. The sample mean, sample median, sample midrange, and sample rth midranges are all *odd location statistics* $T(X_1,\dots,X_n)$ such that, for all h

$$T(x_1 + h,\ \dots,\ x_n + h) = T(x_1,\ \dots,\ x_n) + h$$
$$T(-x_1,\ \dots,\ -x_n) = -T(x_1,\ \dots,\ x_n)$$

The sample variance, sample range, quasi-ranges, sample mean deviation from the mean or median, and the sample interquartile range are all *even location-free statistics* $T(X_1,\dots,X_n)$ such that, for all h,

$$T(x_1 + h,\ \dots,\ x_n + h) = T(x_1,\ \dots,\ x_n)$$
$$T(-x_1,\ \dots,\ -x_n) = T(x_1,\ \dots,\ x_n)$$

Hogg (1960, pp. 265-267) shows that for $N(\mu,\sigma^2)$ samples, or for any symmetric distribution, the correlation coefficient, if any, of an odd location statistic and an even location-free statistic is zero.

One other even location-free statistic is Gini's mean difference, discussed next.

[*8.8.2.1*] *Gini's mean difference* G is defined by (Kendall and Stuart, 1977, p. 257)

$$G = \{n(n - 1)\}^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} |X_i - X_j|$$

Computationally, however, one of two other forms, functions of the order statistics, may be more suitable (David, 1970, pp. 146, 147, 167):

$$G = \{n(n - 1)\}^{-1} \sum_{i=1}^{[(1/2)n]} (n - 2i + 1)W'_r$$

$$= 4\{n(n - 1)\}^{-1} \sum_{i=1}^{n} \{i - (1/2)(n + 1)\}X_{(i;n)}$$

where W'_r, the rth quasi-range, is defined in [8.5.1].

For a sample of size n from a $N(\mu,\sigma^2)$ parent (David, 1970, p. 167; Kendall and Stuart, 1969, p. 241),

$$E(G) = 2\sigma/\sqrt{\pi} = 4 \int_{-\infty}^{\infty} x\{F(x;\mu,\sigma^2) - 1/2\}f(x;\mu,\sigma^2)\, dx$$

$$\text{Var}(G) = 4\sigma^2\{\pi n(n - 1)\}^{-1}\{2(n - 2)\sqrt{3} - 2(2n - 3) + (n + 1)\pi/3\}$$

$$\simeq (0.8068)^2\sigma^2/n \quad \text{when n is large}$$

Further,

$$E(G) = E|X_i - X_j|, \qquad i \neq j$$

in the unordered sample, and $E(\sqrt{\pi}G/2) = \sigma$.

[*8.8.2.2*] The limiting distribution of $\{G - E(G)\}/\sqrt{\text{Var}(G)}$ as $n \to \infty$ is standard normal for independent and identically distributed samples, whenever the parent distribution has a finite second moment (Stigler, 1974, pp. 683, 690-691).

[*8.8.3.1*] For a random sample of size n from a normal population with mean μ and variance σ^2, and for fixed k, $\{nF(X_{(k;n)};\mu,\sigma^2)\}$

is a sequence of rvs converging in distribution as $n \to \infty$ $(n \ge k)$ to the gamma distribution with cdf (Wilks, 1962, p. 269)

$$G(y;k) = \int_0^y (1/\Gamma(k))x^{k-1}e^{-x}\,dx, \qquad y > 0$$

[*8.8.3.2*] With the conditions of [8.8.3.1], let np^* be an integer such that $p^* = p + O(1/n)$, $0 < p < 1$. Then, for large n, $F(X_{(np^*;n)};\mu,\sigma^2)$ is asymptotically distributed as a normal rv with mean p and variance $p(1 - p)/n$ (Wilks, 1962, p. 271).

[*8.8.4*] Let π_n be the proportion of observations which are greater than the sample mean $\bar{X}$ in a normally distributed sample of size n. Then $\sqrt{n}(\pi_n - 1/2)$ has an asymptotic $N(0, 1/4 - 1/(2\pi))$ distribution (David, 1962, p. 1161). Let

$$P(n,k) = \Pr\{X_{(k;n)} \le \bar{X} \le X_{(k+1;n)}\}$$

Kendall (1954, pp. 560-564) gives an Edgeworth series as an approximation to P(n,k); David (1963, pp. 49-54) gives some exact expressions and bounds for n = 4, 5, and arbitrary n. Specifically, if

$$C(n) = n\sqrt{n-1}\,\Gamma\{(n-1)/2\}/\{2(n-2)!\,\pi^{(1/2)(n-1)}\}$$

$$C(n)(1/2)^{(1/2)(n-2)} \le P(n,1) \le C(n)(1 - n^{-1})^{(1/2)(n-2)}$$

The exact values of P(4,1), P(5,1), and P(6,1) are 0.175, 0.049, and 0.011, respectively.

The asymptotic value of P(n,k) as $n \to \infty$ and for fixed values of k is (David, 1963, p. 53)

$$(k^{n-k-1}e^{n-2k})^{1/2}/\{2^{n-k+1}(k!)^2 n^{n-3k-1}\pi^{n-k}\}^{1/2}$$

[*8.8.5*] Let $X_1, X_2, \ldots, X_n$ be a random sample from a $N(\mu,\sigma^2)$ distribution. If a proportion α of the ordered sample is "trimmed" off the lower end, and a proportion $1 - \beta$ is "trimmed" off the upper end, then the average of the remaining observations is the *trimmed mean*, T_n, say, where, if $0 < \alpha < \beta < 1$, and [x] is the largest integer less than or equal to x,

$$T_n = \{[\beta n] - [\alpha n]\}^{-1} \sum_{i=[\alpha n]+1}^{[\beta n]} X_{(i;n)}$$

For given α, β, and n, the mean and variance of T_n can be computed from results given in [8.2.3], [8.2.4], and [8.2.6]; the trimmed mean is less sensitive to outliers in the sample than is the sample mean $\bar{X}$. Stigler (1973, p. 473) gives the asymptotic distribution of T_n (see [6.2.12]); for a $N(\mu,\sigma^2)$ parent, this is as follows:

Let μ_t and σ_t^2 be the mean and variance, respectively, of the doubly truncated $N(\mu,\sigma^2)$ distribution, truncated below at a and above at b, where

$$\Pr(X_i \le a) = \alpha \qquad \Pr(X_i \ge b) = 1 - \beta$$

Expressions for μ_t and σ_t^2 appear in [2.6.2]; note that, if for a N(0,1) random variable Z,

$$\Pr(Z \le z_{1-\alpha}) = \Phi(z_{1-\alpha}) = \alpha$$

as in [2.1.2], then

$$a = \mu + z_{1-\alpha}\sigma \qquad b = \mu + z_{1-\beta}\sigma$$

The asymptotic distribution of $\sqrt{n}(T_n - \mu_t)$ as $n \to \infty$ is normal with mean zero and variance σ^{*2}, where

$$\sigma^{*2} = (\beta - \alpha)^{-2}\{(\beta - \alpha)\sigma_t^2 + \beta(1 - \beta)(b - \mu_t)^2$$
$$+ \alpha(1 - \alpha)(\mu_t - a)^2 + 2\alpha(1 - \beta)(b - \mu_t)(\mu_t - a)\}$$

If $\alpha = 1 - \beta$, so that the trimming is symmetrical, then $\mu_t = \mu$ and

$$\sigma^{*2} = \{1 - 2\alpha + 2z_{1-\alpha}\phi(z_{1-\alpha})$$
$$+ 2\alpha z_{1-\alpha}^2\}\sigma^2/(1 - 2\alpha)^2, \qquad 0 < \alpha < 1/2$$

[*8.8.6.1*] *The Ratio of Two Ranges*. Let R_1 and R_2 be the sample ranges from two independent random samples of sizes n_1 and n_2, respectively, from normal populations having the same variance, σ^2.

Consider the ratio $R = R_1/R_2$; the statistic R is sometimes used in place of the usual F statistic. The pdf and cdf of R are, respectively,

$$h(r;n_1,n_2) = \int_0^\infty x g_2(x;n_2) g_1(xr;n_1)\, dx$$

$$H(r;n_1,n_2) = \int_0^r h(x;n_1,n_2)\, dx = 1 - H(1/r;n_2,n_1)$$

where g_i is the pdf of R_i/σ (i = 1, 2) (Harter, 1969a, pp. 5-6).

[*8.8.6.2*] Some particular cases of $H(r;n_1,n_2)$ are given by $H(r;2,2) = (2/\pi)\arctan(r)$, $H(r;2,3) = (6/\pi)\arctan\{r/(4 + 3r^2)^{1/2}\}$, $H(r;3,2) = (6/\pi)\arctan\{(3 + 4r^2)^{1/2}\} - 2$ (Link, 1950, p. 113).

[*8.8.6.3*] Sources for tables of the pdf, cdf, and percent points of R are listed in Table 8.1. Also listed are sources for tables of percent points of $|\bar{X}_1 - \bar{X}_2|/(W_1 + W_2)$, where $\bar{X}_1$ and $\bar{X}_2$ are the sample means of the two samples.

[*8.8.7.1*] *Mean Range*. The mean range is used as an estimator of standard deviation in a one-way and the associated analysis of variance (David, 1970, pp. 158-162). Suppose m samples, all of size n, are taken from the same normal distribution with $\sigma = 1$. Then the mean range $\bar{W}_{n,m}$ is defined as the mean of the m sample ranges. Let g(y;n,m) be the pdf of $\bar{W}_{n,m}$, and G(y;n,m) the cdf. If $T(h,\lambda)$ is the function related to bivariate normal probabilities defined in [10.2.3] and tabled by Owen (1956, pp. 1080-1087), then (Bland et al., 1966, pp. 246-247)

$$G(y;2,2) = \Pr(\bar{W}_{2,2} \le y) = [2\Phi(y) - 1]^2, \qquad y > 0$$

$$g(y;2,3) = 4\sqrt{6}\phi(y\sqrt{3/2})\{1 - 6T((1/2)y\sqrt{3},\sqrt{3})\}, \qquad y > 0$$

$\bar{R}_{3,2}$ has mean $3/\sqrt{\pi}$ and variance $1 - 3(3 - \sqrt{3})/(2\pi)$; Bland et al. (1966, pp. 246-247) give expressions for g(y;3,2) and g(y;2,4). For a discussion of some approximations to the distribution of $\bar{W}_{n,m}$ by chi-square, see David (1970, pp. 141-142) and Cadwell (1953, pp. 336-346).

[*8.8.7.2*] Bliss et al. (1956, p. 420) have tabled upper 5 percent points of the distribution of $T = W_{max}/\Sigma_{j=1}^{k} W_j$, where W_{max} is the largest among k mutually independent sample ranges $W_1, \ldots, W_k$, each based upon a random sample of size n from a common $N(\mu,\sigma^2)$ parent. The statistic T has been suggested to test for the presence of outliers. The percent points are based upon the use of two approximations (Bliss et al., 1956, p. 421). See also Table 8.1 for other sources of tables.

[*8.8.8.1*] Let $S_1^2, \ldots, S_k^2$ be sample variances of k mutually independent random samples, each of size n, based on a common $N(\mu,\sigma^2)$ distribution, and let

$$U = S_{max}^2 / \sum_{j=1}^{k} S_j^2$$

where S_{max}^2 is the largest of $S_1^2, \ldots, S_k^2$. Cochran (1941, p. 50) has tabled the upper 5 percent points of U for k = 3(1)10 and n = 2(1)11. When n = 3, and each S_j^2 has two degrees of freedom,

$$\Pr(U > y) = k(1 - y)^{k-1} - \binom{k}{2}(1 - 2y)^{k-1} + \cdots + (-1)^{h-1}\binom{k}{h}(1 - hy)^{k-1}$$

where h is the greatest integer less than 1/y (Cochran, 1941, p. 47).

If the number k of samples is large, then approximately,

$$\Pr(U \le y) \simeq (1 - e^{-(1/2)k(n-1)y})^k$$

and U has an approximate pdf

$$g(u;k,n) \simeq (1/2)k^2(n - 1)e^{-(1/2)k(n-1)y} \cdot (1 - e^{-(1/2)k(n-1)y})^{k-1}, \qquad y > 0$$

with (approximate) mean value $2(\gamma + \log k)/\{(n - 1)k\}$ (Cochran, 1941, p. 51); γ is Euler's constant. See also Table 8.1 for other sources with tables.

[*8.8.8.2*] Hartley (1950, p. 308) introduced the ratio $F_{max} = S_{max}^2/S_{min}^2$ to test for heterogeneity of variance, where S_{max}^2 is defined in [8.8.8.1] above, and S_{min}^2 is the smallest of $S_1^2, \ldots, S_k^2$. David (1952, p. 424) gives a table of upper 5 and 1 percent points of the distribution of F_{max}, for the degrees of freedom n - 1 = 2(1)10, 12, 15, 20, 30, 60, ∞, and k = 2(1)12. See Table 1.1 for other sources of tables.

Let $g(y;\nu)$ and $G(y;\nu)$ be the pdf and cdf, respectively, of a chi-square rv with ν degrees of freedom. Then (David, 1952, p. 423)

$$\Pr(F_{max} \le y) = k \int_0^\infty g(x;\, n-1)[G(xy;\, n-1) - G(x;n-1)]^{k-1}\, dx$$

REFERENCES

The numbers in square brackets give the sections in which the corresponding references are cited.

Barnard, J. (1978). Algorithm AS126: Probability integral of the normal range, *Applied Statistics* 27, 197-198. [8.4.3]

Berman, S. (1962). Limiting distribution of the studentized largest observation, *Skandinavisk Aktuarietidskrift* 45, 154-161. [8.3.1.1]

Bickel, P. J. (1967). Some contributions to the theory of order statistics, *Proceedings of the Fifth Berkeley Symposium* on *Mathematical Statistics and Probability* 1, 575-591. [8.2.7]

Bland, R. P., Gilbert, R. D., Kapadia, C. H., and Owen, D. B. (1966). On the distributions of the range and mean range for samples from a normal distribution, *Biometrika* 53, 245-248. [8.4.1.2; [8.8.7.1]

Bliss, C. I., Cochran, W. G., and Tukey, J. W. (1956). A rejection criterion based upon the range, *Biometrika* 43, 418-422. [8.8.7.2]

Blom, G. (1958). *Statistical Estimates and Transformed Beta-Variables*, Uppsala, Sweden: Almqvist & Wiksell; New York: Wiley. [8.2.6]

Borenius, G. (1958). On the distribution of the extreme values in a sample from a normal distribution, *Skandinavisk Aktuarietidskrift* 41, 131-166. [8.3.9]

Bose, R. C., and Gupta, S. S. (1959). Moments of order statistics from a normal population, *Biometrika* 46, 433-440. [8.2.3]

Cadwell, J. H. (1952). The distribution of quantiles of small samples, *Biometrika* 39, 207-211. [8.6.1; 8.6.5.1; 8.6.6; 8.7.3]

Cadwell, J. H. (1953a). The distribution of quasi-ranges in samples from a normal population, *Annals of Mathematical Statistics* 24, 603-613. [8.4.4.1; 8.5.2, 4]

Cadwell, J. H. (1953b). Approximating to the distributions of measures of dispersion by a power of χ^2, *Biometrika* 40, 336-346. [8.4.4.2; 8.8.7.1]

Cadwell, J. H. (1954). The probability integral of range for samples from a symmetrical unimodal population, *Annals of Mathematical Statistics* 25, 803-806. [8.4.4.3]

Chu, J. T. (1955). On the distribution of the sample median, *Annals of Mathematical Statistics* 26, 112-116. [8.6.2, 4]

Chu, J. T., and Hotelling, H. (1955). The moments of the sample median, *Annals of Mathematical Statistics* 26, 593-606. [8.6.3.1]

Cochran, W. G. (1941). The distribution of the largest of a set of estimated variances as a fraction of their total, *Annals of Human Genetics* 11, 47-52. [8.8.8.1]

Cramér, H. (1946). *Mathematical Methods of Statistics*, Princeton, N.J.: Princeton University Press. [8.1.7.3; 8.7.2, 3, 4, 5]

Daly, J. D. (1946). On the use of the sample range in an analogue of Student's t-test, *Annals of Mathematical Statistics* 17, 71-74. [8.4.13]

David, F. N., Barton, D. E., et al. (1968). *Normal Centroids, Medians, and Scores for Ordinal Data*, Tracts for Computers, Vol. XXIX, Cambridge: Cambridge University Press. [8.2.11]

David, F. N., and Johnson, N. L. (1954). Statistical treatment of censored data, *Biometrika* 41, 228-240. [8.2.4]

David, H. A. (1952). Upper 5% and 1% points of the maximum F-ratio, *Biometrika* 39, 422-424. [8.8.8.2]

David, H. A. (1956a). On the application to statistics of an elementary theorem in probability, *Biometrika* 43, 85-91. [8.3.2.2]

David, H. A. (1956b). Revised upper percentage points of the extreme studentized deviate from the sample mean, *Biometrika* 43, 449-451. [8.3.10]

David, H. A. (1970). *Order Statistics*, New York: Wiley. [8.1.2.1, 2; 8.1.3, 5, 6; 8.1.7.3; 8.1.8.2; 8.2.2, 3, 4, 5, 6, 7, 8; 8.3.2.1; 8.3.3, 6, 8, 10; 8.4.1.1; 8.4.2.2; 8.4.4.1, 2; 8.4.9, 11; 8.8.2.1; 8.8.7.1]

David, H. A., Hartley, H. O., and Pearson, E. S. (1954). The distribution of the ratio, in a single normal sample, of range to standard deviation, *Biometrika* 41, 482-493. [8.4.7, 8, 9, 11]

David, H. T. (1962). The sample mean among the moderate order statistics, *Annals of Mathematical Statistics* 33, 1160-1166. [8.8.4]

David, H. T. (1963). The sample mean among the moderate order statistics, *Annals of Mathematical Statistics* 34, 33-55. [8.8.4]

Davis, C. S., and Stephens, M. A. (1978). Approximating the covariance matrix of normal order statistics, *Applied Statistics* 27, 206-212. [8.2.10]

Dixon, W. J. (1957). Estimates of the mean and standard deviation of a normal population, *Annals of Mathematical Statistics* 28, 806-809. [8.6.3.2; 8.6.12, 13]

Fisher, R. A., and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample, *Proceedings of the Cambridge Philosophical Society* 24, 180-190. [8.1.7.2]

Galambos, J. (1978). *The Asymptotic Theory of Extreme Order Statistics*, New York: Wiley. [8.1.7.3; 8.1.8.3; 8.4.5; 8.6.11]

Godwin, H. J. (1949). Some low moments of order statistics, *Annals of Mathematical Statistics* 20, 279-285. [8.2.3, 9]

Govindarajulu, Z. (1963). On moments of order statistics and quasi-ranges from normal populations, *Annals of Mathematical Statistics* 34, 633-651. [8.2.2]

Govindarajulu, Z., and Hubacker, N. W. (1964). Percentiles of order statistics in samples from uniform, normal, chi (ld.f.) and Weibull populations, *Reports on Statistical Applications Research, JUSE*, 11, 64-90. [8.1.5]

Grubbs, F. E. (1950). Sample criteria for testing outlying observations, *Annals of Mathematical Statistics* 21, 27-58. [8.3.2.1; 8.3.4, 7]

Guenther, W. C. (1977). An easy method for obtaining percentage points of order statistics, *Technometrics* 19, 319-322. [8.1.4]

Gumbel, E. J. (1958). *Statistics of Extremes,* New York: Columbia University Press. [8.1.3; 8.1.7.2; 8.6.8, 11]

Gupta, S. S. (1961). Percentage points and modes of order statistics from the normal distribution, *Annals of Mathematical Statistics* 32, 888-893. [8.1.4, 5]

Haldane, J. B. S., and Jayakar, S. D. (1963). The distribution of extremal and nearly extremal values in samples from a normal distribution, *Biometrika* 50, 89-94. [8.1.7.1; 8.1.8.2]

Halperin, M., Greenhouse, S. W., Cornfield, J., and Zalokar, J. (1955). Tables of percentage points for the Studentized maximum absolute deviate in normal samples, *Journal of the American Statistical Association* 50, 185-195. [8.3.11]

Harter, H. L. (1959). The use of sample quasi-ranges in estimating population standard deviation, *Annals of Mathematical Statistics* 30, 980-999. [8.5.4]

Harter, H. L. (1960). Tables of range and Studentized range, *Annals of Mathematical Statistics* 31, 1122-1147. [8.4.3, 12]

Harter, H. L. (1961). Expected values of normal order statistics, *Biometrika* 48, 151-165; correction: 476. [8.2.6, 8]

Harter, H. L. (1969a). *Order Statistics and Their Use in Testing and Estimation*, Vol. 1, Aerospace Research Laboratories, USAF. [8.4.1.1, 2; 8.4.3, 10; 8.8.6.1; Table 8.1]

Harter, H. L. (1969b). *Order Statistics and Their Use in Testing and Estimation*, Vol. 2, Aerospace Research Laboratories, USAF. [8.2.2, 6; 8.5.2, 3, 4; Table 8.1]

Hartley, H. O. (1942). The range in normal samples, *Biometrika* 32, 334-348. [8.4.1.1]

Hartley, H. O. (1950). The maximum F ratio as a short-cut test for heterogeneity of variance, *Biometrika* 37, 308-312. [8.8.8.2]

Hodges, J. L., and Lehmann, E. L. (1967). On medians and quasi medians, *Journal of the American Statistical Association* 62, 926-931. [8.6.3.2; 8.6.5.2; 8.6.13]

Hogg, R. V. (1960). Certain uncorrelated statistics, *Journal of the American Statistical Association* 55, 265-267. [8.8.1]

Johnson, N. L. (1952). Approximations to the probability integral of the distribution of range, *Biometrika* 39, 417-418. [8.4.4.4]

Johnson, N. L., and Kotz, S. (1970). *Distributions in Statistics, Continuous Univariate Distributions*, Vol. 2, New York: Wiley. [8.1.4]

Jones, H. L. (1948). Exact lower moments of order statistics in small samples from a normal distribution, *Annals of Mathematical Statistics* 19, 270-273. [8.2.2, 3, 9]

Kendall, M. G. (1954). Two problems in sets of measurements, *Biometrika* 41, 560-564. [8.8.4]

Kendall, M. G., and Stuart, A. (1977). *The Advanced Theory of Statistics*, Vol. 1 (4th ed.), New York: Macmillan. [8.1.8.1; 8.3.1.2; 8.6.2; 8.6.3.1; 8.6.5.1; 8.6.10, 12; 8.7.3, 4, 5; 8.8.2.1]

Lehmann, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks*, San Francisco: Holden-Day. [8.2.11; 8.7.6]

Leslie, R. T., and Culpin, D. (1970). Distribution of quasimid-ranges and associated mixtures, *Technometrics* 12, 311-325. [8.6.13]

Link, R. F. (1950). The sampling distribution of the ratio of two ranges from independent samples, *Annals of Mathematical Statistics* 21, 112-116. [8.8.6.2]

Lord, E. (1947). The use of range in place of standard deviation in the t-test, *Biometrika* 34, 41-67. [8.4.1.2]

McKay, A. T. (1935). The distribution of the difference between the extreme observation and the sample mean in samples of n from a normal universe, *Biometrika* 27, 466-471. [8.3.2.1, 2]

McKay, A. T., and Pearson, E. S. (1933). A note on the distribution of range in samples of n, *Biometrika* 25, 415-420. [8.4.1.2; 8.4.2.1]

Mosteller, F. (1946). On some useful "inefficient" statistics, *Annals of Mathematical Statistics* 17, 377-408. [8.6.13; 8.7.4]

Owen, D. B. (1956). Tables for computing bivariate normal probabilities, *Annals of Mathematical Statistics* 27, 1075-1090. [8.4.1.2; 8.8.7.1]

Owen, D. B. (1962). *Handbook of Statistical Tables*, Reading, Mass.: Addison-Wesley. [Table 8.1]

Owen, D. B., Odeh, R. E., and Davenport, J. M. (1977). *Selected Tables in Mathematical Statistics*, Vol. V, Providence, R.I.: American Mathematical Society. [8.2.8]

Pearson, E. S., and Chandra Sekar, C. (1936). The efficiency of statistical tools and a criterion for the rejection of outlying observations, *Biometrika* 28, 308-320. [8.3.7, 9]

Pearson, E. S., and Hartley, H. O. (1966). *Biometrika Tables for Statisticians*, Vol. 1 (3rd ed.), London: Cambridge University Press. [8.3.10, 12; Table 8.1]

Pearson, E. S., and Hartley, H. O. (1972). *Biometrika Tables for Statisticians*, Vol. 2, London: Cambridge University Press. [Table 8.1]

Pearson, E. S., and Stephens, M. A. (1964). The ratio of range to standard deviation in the same normal sample, *Biometrika* 51, 484-487. [8.4.7, 12]

Pearson, K. (1931). On the standard error of the median to a third approximation..., *Biometrika* 23, 361-363. [8.6.5.1]

Pillai, K. C. S. (1950). On the distributions of midrange and semirange in samples from a normal population, *Annals of Mathematical Statistics* 21, 100-105. [8.6.9]

Pillai, K. C. S. (1952). On the distribution of "Studentized" range, *Biometrika* 39, 194-195. [8.4.10]

Pillai, K. C. S. (1959). Upper percentage points of the extreme Studentized deviate from the sample mean, *Biometrika* 46, 473-474. [8.3.10]

Plackett, R. L. (1958). Linear estimation from censored data, *Annals of Mathematical Statistics* 29, 131-142. [8.2.4]

Pyke, R. (1965). Spacings, *Journal of the Royal Statistical Society* B27, 395-449 (with discussion). [8.1.2.1]

Quesenberry, C. P., and David, H. A. (1961). Some tests for outliers, *Biometrika* 48, 379-390. [8.3.5, 12]

Rao, C. R., Mitra, S. K., and Matthai, A. (eds.) (1966). *Formulae and Tables for Statistical Work*, Calcutta: Statistical Publishing Society. [Table 8.1]

Ruben, H. (1954). On the moments of order statistics in samples from normal populations, *Biometrika* 41, 200-227. [8.2.8]

Ruben, H. (1956). On the moments of the range and product moments of extreme order statistics in normal samples, *Biometrika* 43, 458-460. [8.4.2.1]

Sarhan, A. E., and Greenberg, B. G. (eds.) (1962). *Contributions to Order Statistics*, New York: Wiley. [8.2.2; 8.4.2.1; Table 8.1]

Saw, J. G. (1960). A note on the error after a number of terms of the David-Johnson series for the expected values of normal order statistics, *Biometrika* 47, 79-86. [8.2.4]

Steck, G. P. (1958). A table for computing trivariate normal probabilities, *Annals of Mathematical Statistics* 29, 780-800. [8.4.1.2]

Stigler, S. (1973). The asymptotic distribution of the trimmed mean, *Annals of Statistics* 3, 472-477. [8.8.5]

Stigler, S. M. (1974). Linear functions of order statistics with smooth weight functions, *Annals of Statistics* 2, 676-693. [8.8.2.2]

Teichroew, D. (1956). Tables of expected values of order statistics and products of order statistics for samples of size twenty and less from the normal distribution, *Annals of Mathematical Statistics* 27, 410-426. [8.2.8]

Thomson, G. W. (1955). Bounds for the ratio of range to standard deviation, *Biometrika* 42, 268-269. [8.4.8]

Wilks, S. S. (1962). *Mathematical Statistics*, New York: Wiley. [8.6.7; 8.7.3; 8.8.3.1, 2]

Woodroofe, M. (1975). *Probability with Applications*, New York: McGraw-Hill. [8.1.8.1]

Yamauti, Z. (ed.) (1972). *Statistical Tables and Formulas with Computer Applications*, Tokyo: Japanese Standards Association. [Table 8.1]

chapter 9

THE WIENER AND GAUSSIAN PROCESSES

9.1 BROWNIAN MOTION

[*9.1.1*] *Introduction*. The Wiener or Wiener-Lévy process plays an important role in the theory of stochastic processes and its applications to quantum mechanics, to the study of thermal noise in circuits, as a model for price fluctuations in stock and commodity markets, and to the asymptotic distribution of goodness-of-fit tests for distributions (Parzen, 1962, chaps. 1, 3; Bhat, 1972, p. 184; Cox and Miller, 1965, chap. 5). But most particularly, the Weiner process (see Section 9.2) and Ornstein-Uhlenbeck process (Section 9.5) are used as models to describe *Brownian motion*, the ceaseless, irregular movements of a small particle immersed in liquid or gas--for example, a smoke particle suspended in air.

[*9.1.2*] The normal distribution first appeared in this context as the pdf of the position X(t) of a particle at time t in one-dimensional space, given that $x = x_0$ when $t = 0$, and such that the diffusion equation

$$\frac{\partial g}{\partial t} = \frac{1}{2}\frac{\partial^2 g}{\partial x^2}$$

is satisfied, where $g(x;t,x_0)$ is the required pdf. Bachelier (1900, pp. 21-86) showed heuristically that a solution to this equation is given by

$$g(x;t,x_0) = (\sqrt{2\pi t})^{-1} \exp\{-(x - x_0)^2/(2t)\}$$

Einstein (1905, 1956) obtained this solution to the same diffusion equation from physical considerations; he explained Brownian motion by assuming that each particle in the surrounding liquid or gas is continually bombarded by molecules. A rigorous mathematical theory to describe Brownian motion was developed by Norbert Wiener (1923, pp. 131-174). This involves measure-theoretic considerations, which we shall avoid here; see also Wax (1954).

[*9.1.3*] The process described in [9.1.2] is the Wiener process, defined formally in [9.2.2] below. When $x_0 = 0$, it is also called the Brownian motion process (Karlin and Taylor, 1975, pp. 340-355). But Cox and Miller (1965, pp. 206-207, 225) point out that, while this stochastic process describes the physical phenomenon of Brownian motion very well if t is reasonably large, it is not so satisfactory when t is small (see [9.2.9.2]). For this reason, and because the Ornstein-Uhlenbeck process (Section 9.5) is a satisfactory alternative to describe Brownian motion, the term *Brownian motion process* may be confusing, and we shall not use it further.

9.2 THE WIENER PROCESS

[*9.2.1*] A *stochastic process* is a collection of random variables $X(t)$, $t \in T$, defined on a common probability space, where T is a subset of $(-\infty,\infty)$; T is commonly a time parameter set. If T is an interval of finite length, $X(t)$ is called a continuous time or continuous parameter process. If the rvs $X(t)$ all take values from a fixed set A, then A is the state space; if A is a continuum such as an interval, then $X(t)$ is a continuous (state space) process, and if A is a discrete set of points (finite or countable), then $X(t)$ is a discrete (state space) process (Hoel et al., 1972, p. vii; Cox and Miller, 1965, p. 14).

[*9.2.2*] *The Wiener process with drift* μ *and variance parameter* σ^2 is a stochastic process $X(t)$ on $[0,\infty)$ such that $X(0) = 0$ and the joint distribution of $X(t_0)$, $X(t_1)$, ..., $X(t_n)$ when

$$t_n > t_{n-1} > \cdots > t_1 > t_0 > 0$$

satisfies the conditions:

(i) The differences $X(t_k) - X(t_{k-1})$ ($k = 1, 2, \ldots, n$) are mutually independent and normally distributed rvs.

(ii) The mean of the difference $X(t_k) - X(t_{k-1}) = (t_k - t_{k-1})\mu$.

(iii) $\mathrm{Var}\{X(t_k) - X(t_{k-1})\} = (t_k - t_{k-1})\sigma^2$.

The distribution of the displacement $X(t) - X(s)$, where $t > s > 0$, thus depends on $t - s$, and not upon s or any other function of s and t; and it is identical to the distribution of $X(t + h) - X(s + h)$. In particular, the displacement $X(t) - X(0)$ has a $N(\mu t, \sigma^2 t)$ distribution; if $\mu = 0$ and $\sigma^2 = 1$, $X(t)$ is the *standardized Wiener process*, as is $\{X(t) - \mu t\}/\sigma$ in the general case.

The Wiener process is denoted below by $W(\mu,\sigma^2;t)$.

[*9.2.3*] The $W(\mu,\sigma^2;t)$ process has the *Markov property* that the conditional distribution of $X(t) - a$, given that $X(u) = a$ and where $t > u$, is that of

$$\mu(t - u) + \sigma Z\sqrt{t - u}$$

where $Z \sim N(0,1)$; this distribution is independent of the history of the process up to time u (Cox and Miller, 1965, p. 206; Breiman, 1968, p. 268). Thus if $t > t_1 > t_0$ (Karlin and Taylor, 1975, p. 343),

$$\Pr\{X(t) \le x \mid X(t_0) = x_0, X(t_1) = t_1\} = \Pr\{X(t) \le x \mid X(t_1) = t_1\}$$

[*9.2.4*] $Z(t)$ is a $W(0,1;t)$ process. Then if $t_0 < t < t_1$, the conditional distribution of $Z(t)$, given that $Z(t_0) = a$ and $Z(t_1) = b$, is normal, with mean $a + (b - a)(t - t_0)/(t_1 - t_0)$ and variance $(t_1 - t)(t - t_0)/(t_1 - t_0)$ (Karlin and Taylor, 1975, p. 345).

[*9.2.5*] Suppose that $Z(t)$ is a $W(0,1;t)$ process and that $a > 0$. Then

$$\Pr\{\max_{0<u\le T} Z(u) \ge a\} = \sqrt{2/(\pi T)} \int_a^\infty \exp\left\{-\frac{1}{2}x^2/T\right\} dx$$
$$= 2\{1 - \Phi(a/\sqrt{T})\}$$
$$\to 1 \text{ as } T \to \infty$$

Hence Z(t) will exceed any bound with probability as close to one as desired if the time period (0,T) is large enough. It can also be shown that a particle subject to the W(0,1;t) law will return to the origin with probability one, no matter how far away it drifts (Karlin and Taylor, 1975, pp. 345-346; Lévy, 1948).

[*9.2.6*] Let X(t) be a $W(\mu,\sigma^2;t)$ process, so that X(0) = 0, and let $Y = \max_{t>0} X(t)$. Then if $\mu \ge 0$, $\Pr(Y \le y) = 0$ $(y < \infty)$, and if $\mu < 0$,

$$\Pr(Y \le y) = 1 - \exp\{-2y|\mu|/\sigma^2\}, \qquad y > 0$$

the negative exponential distribution (Cox and Miller, 1965, p. 212).

[*9.2.7*] *The Law of the Iterated Logarithm.* If Z(t) is a W(0,1;t) process (Breiman, 1968, pp. 263, 266),

$$\overline{\lim_{t\downarrow 0}} \frac{X(t)}{\sqrt{2t \log(\log 1/t)}} = 1 \text{ a.s.}$$

$$\overline{\lim_{t\to\infty}} \frac{X(t)}{\sqrt{2t \log(\log t)}} = 1 \text{ a.s.}$$

[*9.2.8*] If X(t) is a $W(\mu,\sigma^2;t)$ process and $t_n > t_{n-1} > \cdots > t_0 \ge 0$, the rvs $X(t_n)$, ..., $X(t_0)$ have a multivariate normal distribution with $E\{X(t_k)\} = \mu t_k$, and a variance-covariance matrix for which $\text{Var } X(t_k) = t_k\sigma^2$ and (Breiman, 1968, p. 250)

$$\text{Cov}\{X(t_j), X(t_k)\} = \sigma^2 \min(t_j, t_k)$$

[*9.2.9.1*] The sample paths of a Wiener process X(t), plotted as realizations against t, are continuous, but the derivative exists nowhere (Karlin and Taylor, 1975, p. 345; Breiman, 1968, pp. 257-261).

[*9.2.9.2*] In a small time interval Δt, the order of magnitude of the change $X(t + \Delta t) - X(t)$ for a $W(0,\sigma^2;t)$ process is $\sqrt{\Delta t}$, and hence the "velocity" would have order of magnitude $(\Delta t)^{-1/2}$, which becomes infinite as $\Delta t \to 0$. This explains why the Wiener process is unsatisfactory for describing Brownian motion when t is small (see [9.1.3] and [9.5.3]), since in Brownian motion, particles move under a quick succession of impacts of neighboring particles (Cox and Miller, 1965, p. 207).

[*9.2.10*] Let $Z(t)$ be a $W(0,1;t)$ process; then the total variation of the path of $Z(t)$ is infinite with probability one. This can be expressed more exactly by the result (Karlin and Taylor, 1975, p. 37)

$$\lim_{n\to\infty} \sum_{k=1}^{2^n} |Z(kt/2^n) - Z\{(k - 1)t/2^n\}| = \infty \text{ a.s.}$$

[*9.2.11*] Let $X(t) - x_0$ be a $W(\mu,\sigma^2;t)$ process, so that $X(0) = x_0$, and let $p(x_0,x;t)$ be the pdf of $X(t)$ at time t, so that

$$p(x_0,x;t) = \frac{1}{\sigma\sqrt{2\pi t}} \exp\left\{-\frac{(x - x_0 - \mu t)^2}{2\sigma^2 t}\right\}$$

Then $X(t)$ satisfies the *forward diffusion equation*

$$\frac{1}{2}\sigma^2 \frac{\partial^2}{\partial x^2} p(x_0,x;t) - \mu \frac{\partial}{\partial x} p(x_0,x;t) = \frac{\partial}{\partial t} p(x_0,x;t)$$

with the initial boundary condition that $p(x_0,x;0)$ has all its probability mass at $x = x_0$. With the same boundary condition, the process satisfies the *backward diffusion equation* (Cox and Miller, 1965, pp. 208-210)

$$\frac{1}{2}\sigma^2 \frac{\partial^2}{\partial x_0^2} p(x_0,x;t) + \mu \frac{\partial}{\partial x_0} p(x_0,x;t) = \frac{\partial}{\partial t} p(x_0,x;t)$$

[9.2.12] Let Z(t) be a W(0,1;t) process. Then the probability that X(t) = 0 at least once when $0 < t_0 < t < t_1$ is $2\pi^{-1} \arccos\sqrt{t_0/t_1}$ (Karlin and Taylor, 1975, p. 348).

9.3 WIENER PROCESS WITH ABSORBING BARRIERS

[9.3.1] *First Passage Time to the Point* $x = a > 0$. The following may be considered with or without the restriction of a single absorbing barrier at x = a. Let T be the first passage time to a, so that

$$X(0) \equiv 0 \qquad X(t) < a \ (0 < t < T) \qquad X(T) \equiv a$$

Let $g(t|a)$ be the pdf of T. Then (see [9.2.6])

$$\Pr(T < \infty) = \begin{cases} 1, & \mu \geq 0 \\ \exp(-2a|\mu|/\sigma^2), & \mu < 0 \end{cases}$$

Hence the probability of ultimate absorption at x = a > 0 is unity if and only if the drift μ is nonnegative (Cox and Miller, 1965, pp. 210-212).

(i) If $\mu < 0$, $\Pr(T < \infty) < 1$ and T has an improper distribution.

(ii) If $\mu > 0$,

$$g(t|a) = \frac{a}{\sigma\sqrt{2\pi t^3}} \exp\left\{-\frac{(a - \mu t)^2}{2\sigma^2 t}\right\}$$

Thus, if $\mu > 0$, the first passage time T has an inverse Gaussian distribution; see [2.3.3] and Johnson and Kotz (1970, chap. 1).

Mean passage time: $E(T|a) = a/\mu$

and

$$\operatorname{Var}(T|a) = a\sigma^2/(2\mu^3)$$

(iii) If $\mu = 0$,

$$g(t|a) = \frac{a}{\sigma\sqrt{2\pi t^3}} \exp\{-a^2/(2\sigma^2 t)\}$$

but T has no finite moments (Cox and Miller, 1965, pp. 220-222).

[*9.3.2*] The presence of absorbing barriers alters the boundary conditions for the diffusion equations in [9.2.11], so that these no longer hold. For a $W(\mu,\sigma^2;t)$ process with *one absorbing barrier* at $x = a > 0$, the pdf $p(x;t|a)$ is now given by (Cox and Miller, 1965, pp. 220-221)

$$p(x;t|a) = \begin{cases} \dfrac{1}{\sigma\sqrt{2\pi t}}\left[\exp\left\{-\dfrac{(x-\mu t)^2}{2\sigma^2 t}\right\} - \exp\left\{\dfrac{2\mu a}{\sigma^2} - \dfrac{(x-2a-\mu t)^2}{2\sigma^2 t}\right\}\right], & x < a \\ 0, & x = a \end{cases}$$

[*9.3.3*] Let $p(x;t|a,b)$ be the pdf of a $W(\mu,\sigma^2;t)$ process with *two absorbing barriers*, at $x = a > 0$ and at $x = b < 0$, with $X(0) = 0$. Then (Cox and Miller, 1965, p. 222)

(i) $$p(x;t|a,b) = \frac{1}{\sigma\sqrt{2\pi t}}\sum_{n=-\infty}^{\infty}\left[\exp\left\{\frac{\mu x_n'}{\sigma^2} - \frac{(x - x_n' - \mu t)^2}{2\sigma^2 t}\right\} - \exp\left\{\frac{\mu x_n''}{\sigma^2} - \frac{(x - x_n'' - \mu t)^2}{2\sigma^2 t}\right\}\right], \quad b < x < a$$

where $x_n' = 2n(a - b)$, $x_n'' = 2a - x_n'$; $n = 0, \pm 1, \pm 2, \ldots$. If $X(0) = u$, and $b < u < a$, replace a, b, and x in the above result by $a - u$, $b - u$, and $x - u$, respectively (Cox and Miller, 1965, p. 223); see also Feller (1971, pp. 341-342).

(ii) The probability of absorption at the upper barrier a when $X(0) = u$ is $P(a)$, where

$$P(a) = \begin{cases} \dfrac{\exp(hu) - \exp(hb)}{\exp(ha) - \exp(hb)}, & h = -\dfrac{2\mu}{\sigma^2} \quad \mu \neq 0 \\ \dfrac{u - b}{a - b}, & \mu = 0 \end{cases}$$

The probability of absorption at the lower barrier b is then $1 - P(a)$ (Karlin and Taylor, 1975, pp. 360-361).

These results have analogs in sequential analysis, for processes in discrete time.

[*9.3.4*] Let $X(t) - x_0$ be a $W(\mu,\sigma^2;t)$ process, so that $X(0) = x_0 > 0$, and suppose that there is a *reflecting barrier* at $x = 0$, so that $X(t)$ is on the positive half-line only. If $X(t)$ has pdf $q(x_0,x;t)$, then (Cox and Miller, 1965, pp. 223-225)

$$q(x,x_0;t) = \frac{1}{\sigma\sqrt{2\pi t}}\left[\exp\left\{-\frac{(x - x_0 - \mu t)^2}{2\sigma^2 t}\right\} + \exp\left\{-\frac{(x + x_0 - \mu t)^2 + 4x_0\mu t}{2\sigma^2 t}\right\} + \frac{2\mu}{\sigma^2}\exp\left(\frac{2\mu x}{\sigma^2}\right)\left\{1 - \Phi\left(\frac{x + x_0 + \mu t}{\sigma\sqrt{t}}\right)\right\}\right], \quad x > 0$$

If $\mu \geq 0$, $q(x_0,x;t) \to 0$ as $t \to \infty$; if $\mu < 0$, $q(x_0,x;t) \to 2|\mu|\sigma^{-2} \times \exp(-2|\mu|x/\sigma^2)$, $x > 0$, as $t \to \infty$, an exponential distribution free of the starting point x_0. This is an equilibrium distribution to which the process stabilizes after an infinite time.

9.4 GAUSSIAN PROCESSES

[*9.4.1*] A discrete stochastic process $\{X_n\}$ ($n = 0, 1, 2, \ldots$) is a *Gaussian or normal process in discrete time* if every finite linear combination of the rvs $X_0, X_1, X_2, \ldots$ is normally distributed (see Hoel et al., 1972, pp. 119-120; Parzen, 1962, pp. 89-90), or equivalently, if the joint distribution of any set $(X_n,X_m,\ldots,X_k)$ is multivariate normal.

[*9.4.2*] A discrete Gaussian process $\{X_n\}$ is *strictly stationary* if the sets $\{X_{n+s},X_{m+s},\ldots,X_{k+s}\}$ and $\{X_n,X_m,\ldots,X_k\}$ have the same joint distribution, where $n, m, \ldots, k$ are nonnegative integers and s is a positive integer. A Gaussian process $\{X_n\}$ is strictly stationary if and only if (i) $E\{X_n\} = \mu$ (constant); and (ii) the covariance $\text{Cov}\{X_n,X_{n+s}\} = \gamma(s)$, a function of the increment s only

(Cox and Miller, 1965, p. 276). Then $\gamma(s)$ is the *autocovariance function*; the correlation coefficient $\rho(s)$ between X_n and X_{n+s} is the *autocorrelation function*. For any strictly stationary Gaussian process in discrete time,

$$\rho(s) = \gamma(s)/\gamma(0): \quad \gamma(0) = \mathrm{Var}(X_n)$$

[*9.4.3*] A special case of the process of [9.4.2] is an unrestricted random walk X_1', $X_1' + X_2'$, ..., $\Sigma_{i=1}^n X_i'$, ... in discrete time, in which $\{X_n'\}$ is an iid sequence of $N(\mu,\sigma^2)$ rvs. The $W(\mu,\sigma^2;t)$ process of [9.2.2] is an approximation to the above random walk if $X(0) = 0$, in the sense that the imbedded process $X(1)$, $X(2)$, ... is a representation of the random walk (Cox and Miller, 1965, p. 213).

[*9.4.4*] Let $\{X_n\}$ be a strictly stationary discrete Gaussian process. If, for $n = 1, 2, \ldots$, the distribution of X_n given that $X_{n-1} = x$ does not depend on $(X_1, X_2, \ldots, X_{n-2})$, then $\{X_n\}$ is a *strictly stationary Markov Gaussian process*.

A Gaussian stationary process $\{X_n\}$ is a Markov process if and only if

$$\rho(s) = \{\rho(1)\}^s, \qquad s = 2, 3, 4, \ldots$$

where $\rho(s)$ is the autocorrelation function (Cox and Miller, 1965, p. 289).

[*9.4.5*] If $X_n = \lambda X_{n-1} + Z_n$, where $\{Z_n\}$ is a sequence of uncorrelated rvs with zero mean and common variance, then $\{X_n\}$ is a *first-order autoregressive process*. If the components $\{Z_n\}$ are iid $N(0,\sigma^2)$ rvs, then $\{X_n\}$ is a Gaussian Markov process (Cox and Miller, 1965, p. 281). If $\lambda = 1$, we have a random walk, imbedded as in [9.4.3] in a Wiener process.

If $|\lambda| < 1$, the process becomes stationary as $n \to \infty$, with autocorrelation function $\rho(s)$, given by $\rho(s) = \lambda^s$ (Cox and Miller, 1965, pp. 279-280).

[*9.4.6*] A continuous stochastic process $X(t)$ is a *Gaussian or normal process in continuous time* if the joint distribution of $X(t_1)$,

$X(t_2)$, ..., $X(t_k)$ is multivariate normal ($k = 1, 2, 3, \ldots$), for any choice of t_1, t_2, ..., t_k (Parzen, 1962, p. 89).

[*9.4.7*] A continuous Gaussian process is *strictly stationary* if $\{X(t_1), \ldots, X(t_k)\}$ and $\{X(t_1 + s), \ldots, X(t_k + s)\}$ have the same joint distribution, where s is any real number; then a continuous Gaussian process is strictly stationary if and only if $\text{Var}\{X(t)\} < \infty$, $E\{X(t)\} = \mu$, constant, and $\text{Cov}\{X(t), X(t + s)\}$ is a function of s only (Karlin and Taylor, 1975, p. 445), say $\gamma(s)$, which is free of t. If for such a process, $\text{Var}\{X(t)\} = \sigma^2$, the autocorrelation function $\rho(s)$ is given by (Cox and Miller, 1965, p. 293)

$$\rho(s) = \gamma(s)/\sigma^2$$

[*9.4.8*] The Wiener process (see Section 9.2) is a Gaussian process with stationary independent increments (Parzen, 1962, pp. 28, 91).

[*9.4.9*] Let X(t) be a Gaussian process. Then the following are also Gaussian processes (Parzen, 1962, p. 90):

(i) $\displaystyle\int_0^t X(s)\, dx$

(ii) $\dfrac{\partial}{\partial t}\{X(t)\}$

(iii) $\displaystyle\sum_{i=1}^{k} a_i X(t + \alpha_i)$, $\quad a_1, \ldots, a_k, \alpha_1, \ldots, \alpha_k$ constant

In order to understand what is meant by the integral and derivative of a stochastic variable, see Parzen (1962, pp. 78-84).

[*9.4.10*] A Markov process in continuous time X(t), or *diffusion process*, has the property that $\Pr[X(u) < y \mid X(t) = x;\ t < u]$ depends only on u and t, and not on events that occurred at times earlier than t (Cox and Miller, 1965, pp. 204-205). A Wiener process is a Gaussian diffusion process.

A continuous Gaussian process is Markov if and only if the autocorrelation function $\rho(s)$ is of the form $\rho(s) = e^{-\lambda|s|}$ for some constant λ (Cox and Miller, 1965, p. 289).

[*9.4.11*] Let Z(t) be a normalized Wiener process ($\mu = 0$, $\sigma^2 = 1$; [1.6.1]). Let f(t) be continuous, nonnegative, and strictly increasing in t, and let a(t) be a real continuous function. If

$$X(t) = a(t)Z\{f(t)\}$$

then X(t) is a Gaussian diffusion process (Cox and Miller, 1965, pp. 228-229).

[*9.4.12*] Let X(t) be a stochastic process with stationary independent increments. Then the paths are continuous with probability one if and only if X(t) is Gaussian (Feller, 1971, pp. 304-305); and then X(t) is a Wiener process, from [9.4.8].

9.5 THE ORNSTEIN-UHLENBECK PROCESS

We present a model due to Uhlenbeck and Ornstein (1930, pp. 823-841) that overcomes the defect for describing Brownian motion which was noted in [9.1.3]. There appear to be two definitions of the OU process, as it will be called, given in [9.5.1] and [9.5.2].

[*9.5.1*] The OU process U(t) with $U(0) = u_0$ is a continuous Gaussian Markov process such that for $t > 0$ (Cox and Miller, 1965, pp. 225-227),

$$E\{U(t)\} = u_0 e^{-\beta t} \qquad \mathrm{Var}\{U(t)\} = \tfrac{1}{2}(\sigma^2/\beta)(1 - e^{-2\beta t}), \qquad \beta > 0$$

One can show that

$$\mathrm{Cov}\{U(t),\ U(t + s)\} = \tfrac{1}{2}(\sigma^2/\beta)e^{-\beta|s|}(1 - e^{-2\beta t}), \quad t > 0,\ \beta > 0$$

The autocorrelation function of U(t) is then $\rho(s)$, where $\rho(s) = e^{-\beta|s|}$. However, the process U(t) does not have independent increments.

[*9.5.2*] As $t \to \infty$, the OU process U(t) is an "equilibrium distribution," a Gaussian diffusion process U*(t) such that (Cox and Miller, 1965, p. 228)

$$E\{U^*(t)\} = 0 \qquad \mathrm{Var}\{U^*(t)\} = (1/2)\sigma^2/\beta, \qquad \beta > 0$$

$$\mathrm{Cov}\{U^*(t),\ U^*(t + s)\} = (1/2)(\sigma\ /\beta)e^{-\beta|s|}, \qquad \beta > 0$$

Parzen (1962, pp. 96-97) defines the OU process in terms of this equilibrium distribution, which is stationary.

[*9.5.3*] The OU process U(t) in [9.5.1] was derived as a model for the *velocity* of a Brownian particle, rather than its displacement X(t). If U(t) represents the velocity of a particle, then X(t) - X(0) is a Gaussian process, but without the Markov property (Cox and Miller, 1965, pp. 227-228):

$$E\{X(t) - X(0)\} = 0$$

$$\begin{aligned} \text{Var}\{X(t) - X(0)\} &= \sigma^2\{\beta t - 1 + \exp(-\beta t)\}/\beta^3, \qquad \beta > 0,\ t > 0 \\ &\simeq \sigma^2 t/\beta^2 \qquad \text{if } t \text{ is large} \\ &\simeq \sigma^2 t^2/(2\beta) \qquad \text{if } t \text{ is small} \end{aligned}$$

A comparison with [9.2.9.2] shows the OU process to be more satisfactory for describing Brownian motion than the $W(0,\sigma^2;t)$ process. See also Feller (1971, pp. 335-336).

[*9.5.4*] Dirkse (1975, p. 596) gives an asymptotic expression for an absorption probability of the equilibrium OU process U*(t) defined in [9.5.2]: if $\beta = 1$ and $\sigma^2 = 2$, so that $\text{Cov}\{U^*(t), U^*(t + s)\} = e^{-|s|}$, then

$$\lim_{c\to\infty} \Pr[\sup_{0\le t\le a} |U^*(t)| > c] = 2\phi(c)\{ac - (a - 1)/c\}$$

where $\phi(\cdot)$ is the N(0,1) pdf.

REFERENCES

The numbers in square brackets give the sections in which the corresponding reference is cited.

Bachelier, L. (1900). Théorie de la speculation, *Annales de l'École Normale Supérieure* 17, 21-86. [9.1.2]

Bhat, U. N. (1972). *Elements of Applied Stochastic Processes*, New York: Wiley. [9.1.1]

Breiman, L. (1968). *Probability*, Reading, Mass.: Addison-Wesley. [9.2.7, 8; 9.2.9.1]

Cox, D. R., and Miller, H. D. (1965). *The Theory of Stochastic Processes*, London: Methuen. [9.1.1, 3; 9.2.1, 3, 6; 9.2.9.2; 9.2.11; 9.3.1, 2, 3, 4; 9.4.2, 3, 4, 5, 7, 10, 11; 9.5.1, 2, 3]

Dirkse, J. P. (1975). An absorption probability for the Ornstein-Uhlenbeck process, *Journal of Applied Probability* 12, 595-599. [9.5.4]

Einstein, A. (1905, 1956). *Investigations on the Theory of the Brownian Movement*, New York: Dover. [9.1.2]

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. 2 (2nd ed.), New York: Wiley. [9.3.3; 9.4.8; 9.5.3]

Hoel, P. G., Port, S. C., and Stone, C. J. (1972). *Introduction to Stochastic Processes*, Boston: Houghton Mifflin. [9.2.1; 9.4.1]

Johnson, N. L., and Kotz, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions*, Vol. 1, New York: Wiley. [9.3.1]

Karlin, S., and Taylor, H. M. (1975). *A First Course in Stochastic Processes*, New York: Academic Press. [9.1.3; 9.2.3, 4, 5; 9.2.9.1; 9.2.10, 12; 9.3.3; 9.4.7]

Lévy, P. (1948). *Processus stochastiques et mouvement Brownien*, Paris: Gauthier-Villars. [9.2.5]

Parzen, E. (1962). *Stochastic Processes*, San Francisco: Holden-Day. [9.1.1; 9.4.1, 6, 8, 9]

Uhlenbeck, G. E., and Ornstein, L. S. (1930). On the theory of Brownian motion, *Physical Review* 36, 823-841. [9.5]

Wax, N. (ed.) (1954). *Selected Papers on Noise and Stochastic Processes*, New York: Dover. [9.1.2]

Wiener, N. (1923). Differential space, *Journal of Mathematics and Physics* 2, 131-174. [9.1.2]

chapter 10

THE BIVARIATE NORMAL DISTRIBUTION

Although we shall not give such detailed attention to correlated normal variables as to the univariate normal distribution, we shall present the basic properties of the bivariate normal distribution. This includes basic properties of the distribution (Section 10.1), algorithms and sources for tables (in Section 10.2), offset coverage probabilities for circles and ellipses (Section 10.3), expected values (Section 10.4), and sampling distributions (Section 10.5), particularly that of the sample correlation coefficient. Some miscellaneous results appear in Section 10.6.

10.1 DEFINITIONS AND BASIC PROPERTIES

[*10.1.1*] The random vector (X,Y) has a ***bivariate normal distribution*** if the joint pdf of (X,Y) is given (Bickel and Doksum, 1977, pp. 23-24) by

$$f(x,y;\mu_1,\mu_2;\sigma_1^2,\sigma_2^2,\rho) = f(\underline{x};\underline{\theta})$$

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

We denote the joint cdf of (X,Y) by $F(\underset{\sim}{x};\underset{\sim}{\theta})$, where $\underset{\sim}{\theta} = (\mu_1,\mu_2;\sigma_1^2,\sigma_2^2,\rho)$. Then X and Y have marginal $N(\mu_1,\sigma_1^2)$ and $N(\mu_2,\sigma_2^2)$ distributions, respectively. (X,Y) has *variance-covariance matrix* $\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, so that the *correlation coefficient* of X and Y is ρ. If we standardize X and Y, so that

$$Z_1 = (X - \mu_1)/\sigma_1 \qquad Z_2 = (Y - \mu_2)/\sigma_2$$

then (Z_1,Z_2) has a standard bivariate normal (SBVN) distribution with joint pdf given by

$$\psi(z_1,z_2;\rho) = \frac{1}{2\pi\sqrt{1 - \rho^2}} \exp\left[\frac{-1}{2(1 - \rho^2)}\{z_1^2 - 2\rho z_1 z_2 + z_2^2\}\right]$$

We shall say that (X,Y) has a $BVN(\underset{\sim}{\theta})$ distribution, and that (Z_1,Z_2) has a $SBVN(\rho)$ distribution, with cdf $\Psi(z_1,z_2;\rho)$. If $\rho = 0$, then (X,Y) is said to have an *elliptical normal* distribution, with joint pdf

$$\frac{1}{2\pi\sigma_1\sigma_2} \exp\left[-\frac{1}{2}\left\{\frac{(x - \mu_1)^2}{\sigma_1^2} + \frac{(y - \mu_2)^2}{\sigma_2^2}\right\}\right]$$

If, in addition, $\sigma_1 = \sigma_2$, (X,Y) is said to have a *circular normal* distribution. These names arise because in the (x,y) plane, the contours of constant density are ellipses and circles, respectively, centered at (μ_1,μ_2).

[*10.1.2*] If $\rho = \pm 1$ above, then $\Pr(Y = aX + b) = 1$, for some constants a and b. If $\rho = 0$, then X and Y are independent; but if, for some bivariate distribution, zero correlation implies independence, it does not follow that (X,Y) has a BVN distribution. An example is the generalized Farlie-Gumbel-Morgenstern distribution, with joint pdf (Johnson and Kotz, 1975, pp. 415-416)

$$h(x,y) = g_1(x)g_2(y)[1 + \alpha\{1 - 2G_1(x)\}\{1 - 2G_2(y)\}], \qquad |\alpha| < 1$$

and joint cdf

$$H(x,y) = G_1(x)G_2(y)[1 + \alpha\{1 - G_1(x)\}\{1 - G_2(y)\}]$$

The marginal cdfs of X and Y in this example are $G_1(\cdot)$ and $G_2(\cdot)$, respectively, and if we choose $G_1 \equiv \Phi$ and $G_2 \equiv \Phi$, the same example demonstrates that (X,Y) may have a non-BVN joint distribution with normal marginals for X and Y.

[*10.1.3*] Derived Distributions. If (X,Y) has a BVN($\underline{\theta}$) distribution, then (aX + bY, cX + dY) has a BVN distribution with mean vector $\begin{pmatrix} a\mu_1 + b\mu_2 \\ c\mu_1 + d\mu_2 \end{pmatrix}$ and variance-covariance matrix

$$\begin{pmatrix} a^2\sigma_1^2 + 2ab\sigma_1\sigma_2 + b^2\sigma_2^2 & ac\sigma_1^2 + \rho(bc + ad)\sigma_1\sigma_2 + bd\sigma_2^2 \\ ac\sigma_1^2 + \rho(bc + ad)\sigma_1\sigma_2 + bd\sigma_2^2 & c^2\sigma_1^2 + 2\rho cd\sigma_1\sigma_2 + d^2\sigma_2^2 \end{pmatrix}$$

[*10.1.4*] The conditional distribution of Y, given X = x, has the conditional pdf $f(y|x)$, where

$$f(y|x) = \frac{1}{\sqrt{2\pi\sigma_2^2(1 - \rho^2)}} \exp\left[-\frac{1}{2\sigma_2^2(1 - \rho^2)} \cdot \left\{y - \mu_2 - \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1)\right\}^2\right]$$

that is, given X = x, Y has a $N[\mu_2 + \rho(\sigma_2/\sigma_1)(x - \mu_1), \sigma_2^2(1 - \rho^2)]$ distribution, in which the homoscedastic property obtains, that the conditional variance $\sigma_2^2(1 - \rho^2)$ is free of the observed value x (Bickel and Doksum, 1977, p. 27).

The *regression equation* for this distribution is the conditional expectation of Y, given X = x, i.e.,

$$E(Y|x) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1)$$

and the *regression coefficient* β is the slope $\rho\sigma_2/\sigma_1$. Two properties are of interest: (i) The variables X and Y - E(Y|X) are independent; (ii) the *regression function* $E(Y|x)$ is not only linear, but for the BVN distribution it is identical to that obtained by the method of least squares (Wilks, 1962, 87-88, 163). The linear combination $\alpha' + \beta' x$ which minimizes $E(Y - \alpha' - \beta' x)^2$ is given by $E(Y|x)$, so that $\alpha' = \mu_2 - \beta\mu_1$, and $\beta' = \beta$. Further, the resulting minimum value of $E(Y - \alpha' - \beta' x)^2$ is $\sigma_2^2(1 - \rho^2)$, the variance of the conditional distribution of Y, given X = x, also known as the (least squares) *residual variance* of Y on X. It is zero if and only if $\rho = \pm 1$.

[*10.1.5*] *Characterizations*. Many of the characterizations of the univariate normal distribution listed in Chapter 4 generalize to the multivariate case; see Johnson and Kotz (1972, pp. 59-62). We give a few of the most straightforward characterizations of the BVN distribution here.

(a) For any constants a and b, not both zero, aX + bY has a normal distribution (Johnson and Kotz, 1972, p. 59).

(b) The independence of the sample mean vector $(\bar{X},\bar{Y})$ of a random sample of n observations from a bivariate population and the elements (S_1^2, S_2^2, R) which determine the sample covariance matrix; the notation is defined in [10.5.1]; see Kendall and Stuart (1977, p. 388).

(c) The regression of Y on X and also that of X on Y (see [10.1.4]) are linear; that is,

$$Y = cX + d + \varepsilon$$
$$X = c'Y + d' + \varepsilon'$$

and the rvs ε and ε' have identical distributions. This is a valid characterization unless X and Y are independent or satisfy some functional relationship (Kendall and Stuart, 1973, p. 367). Fisk (1970, p. 486) shows that no assumptions about moments of higher order than the first are needed. If second moments do exist, however, then the cases in which this characterization is not valid correspond to $\rho = 0$, $\rho = \pm 1$. See also Bhattacharyya (1943, pp. 399-406) for other characterizations involving regression.

(d) Let X and Y be rvs such that (i) X - aY and Y are independent; and (ii) Y - bX and X are independent. Then (X,Y) has a BVN distribution if $ab \neq 0$ and $ab \neq 1$ (Rao, 1975, pp. 1-13).

(e) Let X, Y, U_1, and U_2 be rvs and a and b be constants such that (i) $Z_1 = X + aY + U_1$ and (Y,U_1,U_2) are independent; (ii) $Z_2 = bX + Y + U_2$ and (X,U_1,U_2) are independent. Then (Z_1,Z_2) has a BVN distribution if $a \neq 0$ and $b \neq 0$; further (Z_1,Z_2) and (U_1,U_2) are independent (Khatri and Rao, 1976, pp. 83-84).

[*10.1.6*] If (X,Y) has a SBVN(ρ) distribution, the ratio X/Y has a Cauchy distribution with pdf given by (Fieller, 1932, pp. 428-440)

$$a(u) = (1 - \rho^2)^{1/2}/\{\pi(1 - 2\rho u + u^2)\}$$

Hinkley (1969, pp. 635-639) derived the distribution of X/Y when (X,Y) has a BVN $(\theta_1,\theta_2;\sigma_1^2,\sigma_2^2,\rho)$ distribution. Let G(u) be the cdf and let

$$a(u) = \left(\frac{u^2}{\sigma_1^2} - \frac{2\rho u}{\sigma_1\sigma_2} + \frac{1}{\sigma_2^2}\right)^{1/2}$$

Then $G(u) \to G^*(u) = (\theta_2 u - \theta_1)/(\sigma_1\sigma_2 a(u))$ as $\theta_2/\sigma_2 \to \infty$, i.e., as $\Pr(Y > 0) \to 1$. Hinkley (1969) shows the approximation $G(u) \simeq G^*(u)$ to be useful when θ_2 is much larger than σ_2; also

$$|G(u) - G^*(u)| \le \Phi(-\theta_2/\sigma_2)$$

[*10.1.7*] If (X,Y) has a circular normal distribution with zero means, and $\sigma_x^2 = \sigma_y^2 = \sigma^2$, the radial error R, where $R^2 = X^2 + Y^2$, has a Rayleigh distribution with pdf

$$g(r) = r\sigma^{-2} \exp(-(1/2)r^2/\sigma^2), \qquad r > 0$$

The circular probable error is the 50-percentile radial error 0.774σ. For the distribution of R when (X,Y) have a general BVN distribution but are uncorrelated, see Weil (1954, pp. 168-170) and Laurent (1957, pp. 75-89).

[*10.1.8*] Lancaster (1957, p. 290) proves a theorem which has useful applications in contingency table and correspondence analysis: Let (X,Y) have a BVN($\underline{\theta}$) distribution; if a bivariate distribution with correlation coefficient ρ' is derived by separate transformations on X and Y, then

$$|\rho'| < |\rho|$$

(See also Kendall and Stuart, 1967, p. 569.) Thus, we have bivariate grouped data (or an $r \times s$ contingency table with qualitative categories), and we then seek separate scoring systems for each classification which will maximize the product-moment correlation coefficient between them, then we are effectively trying to fit the closest to a BVN distribution that the data will allow.

[*10.1.9*] Craig (1936, pp. 1-15) gives a series form for the distribution of XY, and Aroian et al. (1978, pp. 165-172) give an integral form; (X,Y) has a BVN distribution.

10.2 TABLES AND ALGORITHMS

[*10.2.1*] *Probabilities*. In the historical development of the study of the bivariate normal distribution, interest centered around algorithms for computing $\Pr(Z_1 \geq h,\ Z_2 \geq k)$, denoted by $L(h,k,\rho)$, where (Z_1,Z_2) has an SBVN (ρ) distribution, that is,

$$L(h,k,\rho) = \int_h^\infty \int_k^\infty \psi(x,y;\rho)\ dy\ dx$$

Then (Sheppard, 1898, pp. 101-167)

$$L(0,0,\rho) = 1/4 + \arcsin\{\rho/(2\pi)\}$$

and

$$\Psi(h,k;\rho) = \Psi(k,h;\rho) = L(-h,-k,\rho) = L(-k,-h,\rho)$$

Available tables of $L(h,k,\rho)$ are given for nonnegative values of h and k only, and so the following relations need to be used (Abramowitz and Stegun, 1964, pp. 936-937; National Bureau of Standards, 1959, pp. vi-vii):

$$\Psi(h,k;\rho) = L(-h,-k,\rho) = L(h,k,\rho) + \Phi(h) + \Phi(k) - 1$$

$$\Psi(-h,k;\rho) = L(h,-k,\rho) = 1 - \Phi(h) - L(h,k,-\rho)$$

$$\Psi(h,-k;\rho) = L(-h,k,\rho) = 1 - \Phi(k) - L(h,k,-\rho)$$

$$\Pr(|Z_1| \le h,\ |Z_2| \le k) = 2[L(h,k,\rho) + L(h,k,-\rho) + \Phi(h) + \Phi(k) - 1]$$

$$\Psi(h,k,0) = \Phi(h)\Phi(k) \quad \text{from independence}$$

$$\Psi(h,k,-1) = \begin{cases} 0, & h + k \le 0 \\ \Phi(h) + \Phi(k) - 1, & h + k \ge 0 \end{cases}$$

Note that $L(h,k,-1) = 1 - \Phi(h) - \Phi(k)$ if $h + k \le 0$ and that equation (26.3.16) of Abramowitz and Stegun (1964) is incorrect:

$$\Psi(h,k,1) = \begin{cases} \Phi(k), & k \le h \\ \Phi(h), & k \ge h \end{cases}$$

Further,

$$L(h,k,\rho) = \int_h^\infty \phi(x)[1 - \Phi\{(k - \rho x)/\sqrt{1 - \rho^2}\}]\, dx$$

$$= \int_k^\infty \phi(y)[1 - \Phi\{(h - \rho y)/\sqrt{1 - \rho^2}\}]\, dy$$

Pólya (1949, pp. 70-73) proved that, if $\rho h > k$, and $0 < \rho < 1$,

$$1 - \Phi(h) - (1 - \rho^2)(\rho h - k)^{-1}\phi(k)[1 - \Phi\{(h - \rho k)/\sqrt{1 - \rho^2}\}] < L(h,k,\rho) < 1 - \Phi(h)$$

[*10.2.2*] Another quantity which has attracted some attention is $V(h,k)$, where

$$V(h,k) = \frac{1}{2\pi}\int_0^h\int_0^{kx/h} \exp[-(x^2 + y^2)/2]\, dy\, dx$$

which is the integral of the standard circular normal density over the triangle with vertices (0,0), (h,0), and (h,k). Then $V(0,k) = 0 = V(h,0)$ and

$$L(h,k,\rho) = L(0,0,\rho) - 1/2[\Phi(h) + \Phi(k) - 1] + V(h,(k - \rho h)/\sqrt{1 - \rho^2}) + V(k,(h - \rho k)/\sqrt{1 - \rho^2})$$

and

$$V(h,k) = L(h,0,\tau) - L(0,0,\tau) + \Phi(h) - 1/4$$

where $\tau = -k\,\mathrm{sgn}(h)/\sqrt{h^2 + k^2}$ and $L(0,0,\rho) = 1/4 + \arcsin\{\rho/(2\pi)\}$ (see [10.2.1]). Further,

$$V(h,k) = V(-h,-k) = -V(-h,k) = V(h,-k)$$
$$V(h,k) + V(k,h) = \{\Phi(h) - 1/2\}\{\Phi(k) - 1/2\}$$

[*10.2.3*] For computational purposes, it is easier to work with the quantity $T(h,\lambda)$ than with $V(h,\lambda h)$, where (Owen, 1956, p. 1078)

$$T(h,\lambda) = (2\pi)^{-1}\arctan(\lambda) - V(h,\lambda h)$$

Here $(2\pi)^{-1}\arctan(\lambda)$ is the integral of the circular normal density $\psi(x,y;0)$ over the sector in the positive quadrant of the (x,y)-plane bounded by the lines $y = 0$, $y = \lambda x$. We have that

$$T(\lambda h,\lambda^{-1}) = 1/4 - \{\Phi(h) - 1/2\}\{\Phi(\lambda h) - 1/2\} - T(h,\lambda)$$
$$T(h,\lambda) = T(-h,\lambda) = -T(h,-\lambda)$$
$$T(h,0) = 0 \qquad V(h,0) = 0$$
$$T(0,\lambda) = (2\pi)^{-1}\arctan(\lambda)$$
$$T(h,1) = (1/2)\Phi(h)\{1 - \Phi(h)\}$$

These results render computation of $T(h,\lambda)$ for positive values of h and λ (such that $\lambda \le 1$) to be sufficient to obtain $T(h,\lambda)$ for any other values of the arguments. The joint standard bivariate normal cdf may be evaluated from the relation

$$\Psi(x,y;\rho) = (1/2)\Phi(h) + (1/2)\Phi(k) - T(h,(k - \rho h)/\{h\sqrt{1 - \rho^2}\}) - T(k,(h - \rho k)/\{k\sqrt{1 - \rho^2}\}) - (1/2)J(h,k)$$

where $J(h,k) = 0$ if $hk > 0$, and if $hk = 0$ but $h + k \ge 0$; and where $J(h,k) = 1$ otherwise (Owen, 1962, p. 184).

[*10.2.4*] *Tables*. The results given above in [10.2.1] to [10.2.3] are frequently used in conjunction with available tables to evaluate $\Psi(h,k;\rho)$, $L(h,k,\rho)$, $V(h,k)$, and $T(h,\lambda)$. We list some of the most accessible tables in Table 10.1, along with the coverage and number of decimal places given. Linear interpolation in

TABLE 10.1 Bivariate Normal Probabilities: Tabled Functions with Coverages

Source	Function	Coverage	Decimal places
Pearson (1931)	$L(h,k,\rho)$	$h, k = 0.0(0.1)2.6$ $\rho = 0(0.05)1$	6
		$h, k = 0.0(0.1)2.6$ $-\rho = 0.0(0.05)1$	7
Nicholson (1943)	$V(h,k)$	$h, k = 0.1(0.1)3.0;\ k = \infty$	6
National Bureau of Standards (1959)	$L(h,k,\rho)$	$h, k = 0.0(0.1)4.0$ $\rho = 0.0(0.05)0.95(0.01)1$	6
		$h, k = 0.0(0.1)4.0$ $-\rho = 0.0(0.05)0.95(0.01)1$	7
	$V(h,\lambda h)$	$\lambda = 0.1(0.1)1.0$ $h = 0.0(0.01)4.0(0.02)$ $4.6(0.1)5.6,\infty$	7
	$V(\lambda h,h)$	$\lambda = 0.1(0.1)1.0$ $h = 0.0(0.01)4.0(0.02)5.6,\infty$	7
Yamauti (1972)	$L(h,k,\rho)$	$\rho = 1.0(-0.01)0.95(-0.05)$ $0.60,-0.60(-0.05)-0.95$ $(-0.01)-0.99$ $h, k = 0.0(0.1)4.0$ $\rho = 0.55,0.50(-0.10)-0.50,$ -0.55 $h, k = 0.0(0.1)2.6$	7
	$V(h,\lambda h)$	$\lambda = 0.1(0.1)1.0$ $h = 0.0(0.01)4.0(0.02)4.6$ $(0.1)5.0,6.0,\infty$	7
	$(2\pi)^{-1}\tan^{-1}\lambda$, $(2\pi)^{-1}\sin^{-1}\lambda$	$\lambda = 0.01,(0.01)1.00$	10
Owen (1956)	$T(h,\lambda)$	$h = 0.0(0.01)2.0(0.02)3.0$ $\lambda = 0.25(0.25)1.0$ $h = 0.0(0.25)3.0$ $\lambda = 0.0(0.01)1.0,\infty$ $h = 3.0(0.05)3.5(0.1)4.7,4.76$ $\lambda = 0.1,0.2(0.05)0.5(0.1)0.8,$ $1.0,\infty$	6

TABLE 10.1 (continued)

Source	Function	Coverage	Decimal places
Owen (1962) Table 8.5	$T(h,\lambda)$	$h = 0.0(0.01)3.14$; $\lambda = 0.25(0.25)1.0$ $h = 0.0(0.25)3.25$; $\lambda = 0.0(0.01)1.0, 1.25, 1.50, 2.00, \infty$ $h = 3.0(0.05)3.50(0.1)4.0(0.2)4.6, 4.76$; $\lambda = 0.1, 0.2(0.05)0.5(0.1)0.8, 1.0, \infty$	6
Smirnov and Bol'shev (1926)	$T(h,\lambda)$	$h = 0(0.01)3.0$; $\lambda = 0(0.01)1.0$ $h = 3.0(0.05)4.0$; $\lambda = 0.05(0.05)1.0$ $h = 4.0(0.1)5.2$; $\lambda = 0.1(0.1)1.0$	7
	$T(h,1)$	$h = 0(0.001)3.0(0.005)4.0(0.01)5.0(0.1)6.0$	
	$(2\pi)^{-1} \tan^{-1}\lambda$	$\lambda = 0(0.001)1.0$	

these and other tables gives varying degrees of accuracy, and the recommended interpolation procedures given in the appropriate sources should be followed if a high degree of accuracy is important.

The first extensive set of tables of bivariate normal probabilities was for $L(h,k,\rho)$, edited by Karl Pearson (1931), and containing a number of earlier sets of tables. Nicholson (1943, pp. 59-72) and Owen (1956, pp. 1075-1090) were the first to table values of $V(h,k)$ and $T(h,\lambda)$, respectively. Zelen and Severo (1960, pp. 621-623) (see Abramowitz and Stegun, 1964, pp. 937-939) provide charts to read off values of $L(h,0,\rho)$, and show how to use the charts to obtain values of $L(h,k,\rho)$.

[*10.2.5*] To 2 decimal places, Fieller et al. (1955) list 3000 random pairs from each of nine SBVN distributions with correlation coefficients 0.1(0.1)0.9. Each page of 50 entries includes sums, sums of squares and products, and sample correlation coefficients. Wold (1948, p. xii) suggested using $(X, \rho X + \sqrt{1 - \rho^2}Y)$ for such pairs, where X and Y are independent N(0,1) rvs, and chosen from tables of random normal deviates.

[*10.2.6*] Owen et al. (1975, pp. 127-138) have tabled values of β such that, when (X,Y) has an SBVN (ρ) distribution,

$$\Pr(Y \geq z_\gamma \mid X \geq z_\beta) = \delta$$

for given values of γ, δ, and ρ, listed below. These appear in the context of a screening variable X, which is used to indicate the acceptability of a performance variable Y. Coverage in the table is to 4 decimal places, for $\rho = 0.3(0.05)1.0$; $\gamma = 0.75(0.01)0.94$ with $\delta = 0.95, 0.99, 0.999$; also for $\gamma = 0.95(0.01)0.98$ with $\delta = 0.99, 0.999$; and for $\gamma = 0.99$ with $\delta = 0.999$.

Algorithms and Approximations

[*10.2.7*] Owen (1956, p. 1079) showed that

$$T(h,\lambda) = \frac{1}{2\pi}\arctan\lambda - \frac{1}{2\pi}\sum_{j=0}^{\infty} \cdot \left[\frac{(-1)^j\lambda^{2j+1}}{2j+1}\{1 - \exp(-(1/2)h^2)\sum_{i=0}^{j}\frac{((1/2)h^2)^i}{i!}\right]$$

which converges rapidly when h and λ are small. Borth (1973, pp. 82-85) suggests using this as an approximation to a desired accuracy of 10^{-7} if $h \leq 1.6$ or $\lambda \leq 0.3$, but for faster convergence with higher values of h or λ, he gives the following modification. Let

$$P_{2m}(x) = \frac{8q}{1-q^2}\left\{\frac{1}{2} + \sum_{k=1}^{m-1}(-1)^k q^k S_{2k}(x) + (-1)^m\frac{q^m}{1-q^2}S_{2m}(x)\right\}$$

where

$$|x| \leq 1$$
$$q = 3 - 2\sqrt{2}$$
$$S_{2k}(x) = \cos(2k \arccos x)$$

which is the Tchebyshev polynomial of degree 2k. Now write

$$P_{2m}(x) = \sum_{k=0}^{m} C_{2k}x^{2k}$$

This approximates $(1 + x^2)^{-1}$ on [-1,1]. Then

$$T(h,\lambda) \simeq (2\pi)^{-1} \exp(-(1/2)h^2) \sum_{k=0}^{m} C_{2k} I_{2k}(h\lambda/\sqrt{2})(h/\sqrt{2})^{-(2k+1)}$$

where

$$I_{2k}(w) = \frac{1}{2}\{(2k - 1)I_{2k-2}(w) - w^{2k-1} \exp(-w^2)\}$$

$$I_0(w) = \sqrt{\pi}\{\Phi(w\sqrt{2}) - 1/2\}$$

Borth (1973) recommends this modification if $h > 1.6$ and $\lambda > 0.3$, noting that if $h > 5.2$, $T(h,\lambda) < 10^{-7}$, and that the required accuracy is attained if $m = 6$. This use of Owen's algorithm with Borth's modification retains speed of computation with accuracy. When $m = 6$, the values of C_{2k} corresponding to k = 0, 1, 2, 3, 4, 5, 6 are, respectively, 0.9^536, -0.9^32989, 0.9872976, -0.9109973, 0.6829098, -0.3360210, and 0.07612251.

[*10.2.8*] Sowden and Ashford (1969, pp. 169-180) suggest a composite method of computing $L(h,k,\rho)$, incorporating Owen's algorithm of [10.2.7] for $T(h,\lambda)$ when h and λ are small, and introducing Hermite-Gauss quadrature or Simpson's rule otherwise. Daley (1974, pp. 435-438) asserts that Simpson's rule integration alone, based on the normal distribution of X, given Y = y, is better, when it is iterated once to give $T(h,\lambda)$, in that no slow convergence problem arises as $\lambda \to \pm 1$; the level of accuracy attained is 3×10^{-7}. Young and Minder (1974, pp. 455-457) give an algorithm for $T(h,\lambda)$ only, over all values of λ.

Daley (1974, pp. 435-438) also gives an approximation to $T(h,\lambda)$, based on that of Cadwell (1951, pp. 475-479):

$$T(h,\lambda) \simeq (2\pi)^{-1}(\arctan \lambda) \exp(-(1/2)\lambda h^2/\arctan \lambda) \cdot \{1 + (0.00868)h^4\lambda^4\}$$

with a maximum error for all h when $|\lambda| \le 1$.

[*10.2.9*] If $h \ge 5.6$ and $0.1 \le \lambda \le 1$, $V(h, h)$ is equal (to 7 decimal places) to $V(\infty,\lambda\infty)$, that is, to $(2\pi)^{-1} \arctan \lambda$; but for small values of λ, $V(\lambda h,h)$ differs considerably at $h = 5.6$ from $V(\lambda\infty,\infty)$, that is, $(2\pi)^{-1}$ arccot λ; if $h \geq 5.6$ in this region, a better approximation is given by (National Bureau of Standards, 1959, pp. vii-viii)

$$V(\lambda h,h) \simeq (2\pi)^{-1} \operatorname{arccot} \lambda - 1/2\{1 - \Phi(\lambda h)\}$$

The error in this approximation, for small values of λ, is less than $1/2 \times 10^{-7}$. If $h \leq 0.8$ and $\lambda \leq 1$ (Johnson and Kotz, 1972, p. 97),

$$V(h,\lambda h) \simeq (4\pi)^{-1}\lambda h^2\{1 - (1/4)\lambda^2(1 + (1/3)\lambda^2)\}$$

In using the preceding tables, linear interpolation gives variable degrees of accuracy, and the recommended procedures given in the sources listed in Table 10.1 should be followed.

[*10.2.10*] Drezner (1978, pp. 277-279) gives a computational algorithm for $\Pr(X \leq h, Y \leq k)$ when (X,Y) has a SBVN (ρ) distribution. The algorithm is based on Gaussian quadrature. See also Gideon and Gurland (1978, pp. 681-684), who give a polynomial approximation which is suitable for minicomputers. The function approximated is $D(r,\theta)/D(r,(1/2)\pi)$, where $D(r,\theta)$ is the probability that (X,Y) lies in a half-infinite triangle. The authors' equation (3) relates $D(r,\theta)$ to the joint cdf of (X,Y), and a table of coefficients for the polynomial approximation is included. The error in approximating $D(r,\theta)$ is not greater than 5×10^{-6}.

10.3 OFFSET CIRCLES AND ELLIPSES

[*10.3.1*] The circular coverage function $p(R,d)$ is defined by

$$p(R,d) = (2\pi)^{-1} \iint_\Gamma \exp\{-(x^2 + y^2)/2\}\, dy\, dx$$
$$\Gamma = \{(x,y): \ (x - d)^2 + y^2 \leq R^2\}$$

Thus $p(R,d)$ is the probability mass of that portion of a standard circular normal or SBVN(0) distribution (defined in [10.1.1]) which lies in a circle Γ of radius R, with the center of Γ offset by a distance d from the mean (0,0) of the distribution. In polar coordinates,

$$p(R,d) = (2\pi)^{-1} \iint_{\Gamma} r \exp(-r^2/2)\, d\theta\, dr$$

$$p(R,0) = 1 - \exp(-(1/2)R^2)$$

where $x = r \cos\theta$, $y = r \sin\theta$, and $\Gamma = \{(r,\theta): r^2 - 2rd \cos\theta + d^2 \le R^2\}$;

$$p(R,d) = \exp(-d^2/2) \int_0^R r \exp(-r^2/2) I_0(rd)\, dr$$

where $I_0(\cdot)$ is a modified Bessel function of the first kind of order zero, viz., $I_0(z) = \pi^{-1} \int_0^{\pi} \exp(-z \cos\theta)\, d\theta$ (Di Donato and Jarnagin, 1962, p. 348).

[*10.3.2*] The circular coverage function has the following reproducing property. Let (X,Y) have a BVN$(a,b;\sigma^2,\sigma^2,0)$ distribution, let $d^2 = a^2 + b^2$, $\sigma_2^2 = \sigma^2 + \sigma_1^2$, and $r^2 = x^2 + y^2$. Then (Read, 1971, p. 1733)

$$p(R/\sigma_2, d/\sigma_2) = (2\pi\sigma^2)^{-1} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(R/\sigma_1, r/\sigma_1) \exp[-\{(x - a)^2 + (y - b)^2\}/(2\pi\sigma^2)]\, dx\, dy$$

[*10.3.3*] Let (Z_1,Z_2) have a standard circular normal [i.e., SBVN(0)] distribution. Then (Owen, 1962, p. 172)

$$\Pr(a_1Z_1^2 + a_2Z_2^2 \le t) = p(A,B) - p(B,A)$$

where $A = (1/2)[\sqrt{t/a_1} + \sqrt{t/a_2}]$ and $B = (1/2)|\sqrt{t/a_1} - \sqrt{t/a_2}|$. This gives the cdf of a quadratic form in two dimensions; further (Owen, 1962, p. 172),

$$\Pr(W \le t) = p(t,\sqrt{\lambda})$$

where W has a noncentral chi-square distribution with two degrees of freedom and noncentrality parameter λ.

[*10.3.4*] The circular coverage function arises as the probability of destruction of a point target at the origin, when damage is complete for a distance $r \le R$ from the point of impact (the "cookie cutter") and zero otherwise, and the point of impact has a circular normal distribution about a mean impact point; p(R,d) is also relevant in certain problems involving areal targets. Two surveys covering material in this and the next subsection in detail are those by Guenther and Terragno (1964, pp. 232-260) and by Eckler (1969, pp. 561-589); these list tables, some of which appear only in technical reports with limited circulation. The latter tables generally have the most detailed coverage, but are unfortunately the least easily obtained. We cite only sources which are readily accessible or are not referenced in the above surveys. Table 10.2 lists some of the available tables and charts, with coverages, of p(R,d). The algorithms used to generate the tables are described in the relevant sources.

TABLE 10.2 Tables and Charts of p(R,d) with Coverages

Source	Function	Coverage	Decimal places
Solomon (1953, Fig. 1)	p(R,r)	$p = 0.05(0.05)0.95$ $0 \le d \le 10$ (horizontally) $0 \le R \le 8$ (vertically)	Graphed
Burington and May (1970, tables 11.10.1, 11.11.1)	R	$p(R,0) = 0.0(0.05)1.0$	4
	p(R,d)	$d = 0(0.1)3.0(0.2)6$ $R = 0.1(0.1)1.0(0.2)3$	3 or 4
Groves and Smith (1957)	R	$R = 0.1, 0.5, 1(1)12$ $0 \le d \le 9$ (horizontally) $0.0^31 \le 1 - p(R,d) \le 0.9^4$	Graphed
Owen (1962, tables 8.1, 8.2)	R	$d = 0$ $p = 0.01(0.01)0.99(0.001)$ $0.9^3(0.0^31)0.9^4$ to 0.9^9	4
	1 - p(R,d)	$R - d = -3.9(0.1)4.0$ $d = 0.1(0.1)6(0.5)10(1)20, \infty$	3
DiDonato and Jarnagin (1962)	R	$p(r,d) = 0.01(0.05)0.95, 0.97, 0.99, 0.995, 0.9^3$ to 0.9^6	6
		$d = 0.1, 0.5(0.5)2(1)6, 8, 10$ $d = 20, 30, 50, 80, 120$	4-5
Bark et al. (1964, table I)	1 - p(R,d)	$R = 0.0(0.02)7.84$ $d = 0.0(0.02)3.00$ See sec. 2 of source for wider coverage.	6

[*10.3.5*] The following recursive procedure is well suited for rapid machine computation of $1 - p(R,d)$, and gets around the difficult problem of interpolating from values in the tables (Brennan and Reed, 1965, pp. 312-313). It may also be less complicated than other methods described in the literature.

Replace the zero-order Bessel function by its expansion to get

$$I_0(rd) = \sum_{n=0}^{\infty} [(1/2)rd]^{2n}/(n!)^2$$

and substitute in the integral of [10.3.1] to obtain

$$1 - p(R,d) = 1 - \sum_{n=0}^{\infty} g_n k_n$$

where

$$g_0 = 1 - \exp(-(1/2)R^2)$$

$$g_n = g_{n-1} - ((1/2)R^2)^n \exp(-(1/2)R^2)/n!$$

$$k_n = ((1/2)d^2)^n \exp(-(1/2)d^2)/n! = k_{n-1}((1/2)d^2)/n$$

After N iterations, the remainder is R_N, where $R_N = \Sigma_N^\infty g_n k_n$. Then

$$R_N < k_N g_N \{1 - ((1/2)Rd/N)^2\}^{-1} \qquad N > (1/2)Rd$$

$$R_{N+1} \le k_N g_N \qquad N > Rd/\sqrt{2}$$

For large enough values of N, the summation can be stopped at any desired accuracy level; the second bound on R_N avoids computation of R_N at each iteration. The procedure is stable in that convergence is rapid once the series approaches its limit, and in that an initial error in g_0 or in k_0 will not grow, but will retain the same percentage error in $1 - p(R,d)$. The series $\Sigma g_n k_n$ for $p(R,d)$ is essentially that derived by Gilliland (1962, pp. 758-767).

[*10.3.6.1*] Approximations to $p(R,d)$ are also available; the first terms of the series in [10.3.5] provide one. Grubbs (1964, pp. 52-55) suggested the following:

$$p(R,d) \simeq \Pr(\chi^2_\nu \le mR^2/v)$$

where χ^2_ν is a central chi-square rv, $\nu = 2m^2/v$ df, $m = 1 + d^2/2$, and $v = 1 + d^2$. Hence $p(R,d) \simeq I((1/2)R^2/\sqrt{v},\ m^2/v - 1)$, where $I(\cdot,\cdot)$ is Pearson's incomplete gamma function (Pearson, 1922). The Wilson-Hilferty approximation (see [10.2.6]) then gives

$$p(R,d) \simeq \Phi[\{(R^2/m)^{1/3} - (1 - vm^{-2}/9)\}/\sqrt{vm^{-2}/9}]$$

for which the greatest observed discrepancy was 0.023.

[*10.3.6.2*] Wegner's (1963) approximation (Read, 1971, p. 1731),

$$p(R,d) \simeq \begin{cases} 1 - \Phi\{d - (R^2 - 1)^{1/2}\}, & \\ \qquad R \ge 3;\ d > 1.5,\ R > 1.8 & \\ R^2(2 + (1/2)R^2)^{-1} \exp[-d^2(2 + (1/2)R^2)^{-1}], & \\ \qquad d \le 1.5,\ R \le 1;\ d \ge 1.5,\ R \le 1.8 & \\ 1 - \exp[-R^2\{1.416 + (0.397 - 0.0159R^2)d^2\}^{-2}], & \\ \qquad \text{otherwise} & \end{cases}$$

has a maximum error 0.02 in the neighborhood of $(R,d) = (2,1.5)$.

[*10.3.6.3*] If d is large, then (DiDonato and Jarnagin, 1962, p. 354)

$$p(R,d) \simeq \Phi(R - d - (1/2)d^{-1}) + D$$

where $D = O(d^{-2})$. The error D is given in table II of Bark et al. (1964). DiDonato and Jarnagin (1962, p. 353) table values of R inversely as a function of $p(\cdot,\cdot)$ and d; for coverage, see Table 10.2.

[*10.3.7*] *Offset Circles and Ellipses: Elliptical Normal Variables*. Let (X,Y) have a $\text{BVN}(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2,0)$ distribution; we wish to evaluate

$$\iint_{x^2+y^2\le R^2} f(x,y;\mu_1,\mu_2;\sigma_1^2,\sigma_2^2,0)\ dy\ dx$$

where the offset circle is centered at the origin and $\sigma_1^2 \ne \sigma_2^2$.

[*10.3.8.1*] *Case I:* $\mu_1 = \mu_2 = 0$. Without loss of generality, $\sigma_1^2 \geq \sigma_2^2$. Let $R' = R/\sigma_1$ and $c = \sigma_2/\sigma_1$, so that $0 \leq c \leq 1$. The probability in [10.3.7] then depends only on R' and c, and is denoted by P(R',c); other sources use the form $P(a_1,a_2,t)$ or $P(a_2,a_1,t)$, where $a_1 \leq a_2$, $c^2 = a_1/a_2$, and $R'^2 = t/a_2$. Then (Guenther and Terragno, 1964, p. 240)

$$P(R',c) = P(a_1,a_2,t) = \Pr(Z_1^2 + c^2Z_2^2 \leq R'^2)$$
$$= \Pr(a_1Z_1^2 + a_2Z_2^2 \leq t)$$

the cdf of a quadratic form in iid N(0,1) variables Z_1 and Z_2.

[*10.3.8.2*] Suppose we require $\Pr(b_1X^2 + b_2Y^2 \leq r)$, the probability of falling inside an ellipse centered at the origin, where $b_1 > 0$ and $b_2 > 0$. If $c' = b_1/\sigma_1^2 + b_2/\sigma_2^2$; $a_i' = b_i/(c\sigma_i^2)$, where $i = 1, 2$; and $t' = r/c$, then (Owen, 1962, p. 181)

$$\Pr(b_1X^2 + b_2Y^2 \leq r) = P(a_1',a_2',t)$$

and we have reduced the problem to that of [10.3.7].

[*10.3.9*] Table 10.3 describes some available tabulations relating to P(R',c); for others, see Eckler (1969, p. 564). Harter (1960, p. 724) used the following form for numerical integration by the trapezoidal rule, with some advice on chopping and truncation errors:

$$P(R',c) = \frac{2c}{\pi}\int_0^{\pi} \frac{[1 - \exp\{-R'^2/(4c^2)\cdot[(1 + c^2) - (1 - c^2)\cos\phi]\}]}{(1 + c^2) - (1 - c^2)\cos\phi}\, d\phi$$

[*10.3.10*] Guenther and Terragno (1964, p. 241) show that

$$P(R',c) = p[(1/2)R'(c^{-1} + 1),\ (1/2)R'(c^{-1} - 1)]$$
$$- p[(1/2)R'(c^{-1} - 1),\ (1/2)R'(c^{-1} + 1)]$$

so that P(R',c) may be obtained from the circular coverage function.

[*10.3.11*] The approximation to the circular coverage function p(R,d) in [10.3.6.1] by Grubbs (1964, pp. 52-55) applies also to

TABLE 10.3 Tables of $P(R',c)$ and $P(a_1,a_2,t)$ with Coverages

Source	Function	Coverage	Decimal places
Grad and Solomon (1955)	$P(a_1,a_2,t)$	$(a_2,a_1) = (0.5,0.5),(0.6,0.4),(0.7,0.3),(0.8,0.2),(0.9,0.1),(0.95,0.05),(0.99,0.01),(1,0)$ $t = 0.1(0.1)1(0.5)2(1)5$	4
Owen (1962, tables 8.3, 8.4)	$P(a_1,a_2t)$	$(a_2,a_1) = (0.5,0.5),(0.6,0.4),(2/3,1/3),(0.7,0.3),(0.75,0.25),(0.8,0.2),(0.875,0.125),(0.9,0.1),(0.95,0.05),(0.99,0.01)$; $t = 0.1(0.1)1,1.5,2.0(1)5$	5
	t	$P(a_1,a_2,t) = 0.05(0.05)0.30(0.1)0.7(0.05)0.95$	
Harter (1960)	$P(R',c)$	$c = 0(0.1)1.0$ $R' = 0.1(0.1)6.0$	7
	R'	$P = 0.5,0.75,0.95,0.975,0.99,0.995,0.9975,0.999$	5
Weingarten and DiDonato (1961)	R'	$c = 0.05(0.05)1$ $P(R',c) = 0.05(0.05)0.95(0.01)0.99$	5
Beyer (1966)	$P(R',c)$	$c = 0(0.1)1.0$ $R' = 0.1(0.1)5.8$	7

$P(R',c)$, being based on fitting the first two moments of a weighted sum of noncentral chi-square variables to those of a central chi-square rv. Thus (see [10.3.8.1]),

$$P(R',c) \simeq \Pr(\chi^2_\nu \le R'^2/\{v(1 + c^2)\})$$

where $\nu = 2/v$ and $v = 2(1 + c^4)/(1 + c^2)^2$; and

$$P(R',c) \simeq \Phi[\{(R'^2/(1 + c^2))^{1/3} - (1 - v/9)\}/\sqrt{v/9}]$$

[*10.3.12*] Let $P_n(\lambda) = \Pr(Y > n)$, where Y has a Poisson distribution with mean λ. Then (Gilliland, 1962, pp. 758-767; Eckler, 1969, p. 565)

$$P(R',c) \simeq \frac{2c}{1 + c^2} \sum_{n=0}^{\infty} \left[\frac{(2n)!}{2^{2n}(n!)^2} \left(\frac{1 - c^2}{1 + c^2} \right)^{2n} P_{2n} \left\{ \frac{(1 + c^2)R'^2}{4c^2} \right\} \right]$$

[*10.3.13*] Lilliefors (1957, pp. 416-421) gives an approximation requiring no tabled functions. Let

$$A = 1 + \frac{1}{4} R'^2 \qquad \text{and} \qquad B = \frac{1}{2} (c^2 - 1)$$

Then

$$P(R',c) \simeq \sum_{n=1}^{\infty} \frac{(-1)^{n+1} R'^{2n}}{2^n n! c^{2n-1}} a_n$$

where

$$a_1 = 1 \qquad a_2 = A \qquad a_3 = A^2 + \frac{1}{2} B^2 \qquad a_4 = A^3 + 3B^2 A/2$$
$$a_5 = A^4 + 3B^2 A^2 + 3B^4/8$$

For algorithms to derive higher factors a_n, see the source paper or Eckler (1969, p. 565).

[*10.3.14.1*] *Case II:* $\mu_1 \neq 0$, $\mu_2 \neq 0$. The probability P in [10.3.7] has been tabulated to three decimal places by Lowe (1960, pp. 177-187). Four parameter values are involved: Lowe's tables cover σ_1/σ_2 = 1, 2, 4, 8; R/σ_2 = 1, 2, 4, 8, 16, 32, 64; and a range of values of μ_1/σ_1 and μ_2/σ_2 chosen to cover that region with substantial variation in the values of P.

Some extensive tables by Groenewoud et al. (1967), which give values of P to 5 decimal places, were derived using a recursive scheme operating on a series expansion; the authors describe the technique. The coverage is as follows:

$$\mu_1/\sigma_1 = \{0.0(0.5)5.0\}(\sigma_2/\sigma_1) \qquad \mu_2/\sigma_2 = 0.0(0.5)5.0$$
$$\sigma_2/\sigma_1 = 0.1(0.1)1.0 \qquad R/\sigma_1 = 2A,\ 3A,\ 4A$$

where values of A are tabulated to 2 or 3 decimal places, and are chosen so that the probability range is reasonably covered.

[*10.3.14.2*] Gilliland (1962, pp. 759-761) gives a series expansion for P; see Eckler (1969, p. 565). The approximation of Grubbs (1964, pp. 51-62), given in [10.3.6.1], holds for P if

$$m = 1 + (\mu_1^2 + \mu_2^2)/(\sigma_1^2 + \sigma_2^2)$$

$$V = 2(\sigma_1^4 + \sigma_2^4 + 2\mu_1^2\sigma_1^2 + 2\mu_2^2\sigma_2^2)/(\sigma_1^2 + \sigma_2^2)^2$$

Then (see also Eckler, 1969, pp. 566-567), if χ_ν^2 is a central chi-square rv, where $\nu = 2m^2/v$ df, and $\sigma^2 = \sigma_1^2 + \sigma_2^2$,

$$P \simeq \Pr(\chi_\nu^2 \le mR^2/v)$$

$$\simeq \Phi[\{(R^2/(m\sigma^2))^{1/3} - (1 - vm^{-2}/9)\}/\sqrt{vm^{-2}/9}]$$

[*10.3.15*] For further discussion, see Johnson and Kotz (1970, chaps. 28 and 29).

10.4 MOMENTS

[*10.4.1*] Let (X,Y) have a BVN($\underline{\theta}$) distribution as defined in [10.1.1]. Define $\mu'_{rs} = E(X^r Y^s)$. We have

$$E(X) = \mu_1 \qquad \mathrm{Var}(X) = \sigma_1^2$$

$$E(Y) = \mu_2 \qquad \mathrm{Var}(Y) = \sigma_2^2$$

Then μ'_{rs} is the *joint moment* of order (r,s) and μ_{rs} the *joint central moment* of order (r,s), where

$$\mu_{rs} = E\{(X - \mu_1)^r (Y - \mu_2)^s\}$$

so that $\mathrm{Cov}(X,Y) = \mu_{11}$, the *covariance* of the BVN($\underline{\theta}$) distribution. From [10.1.1], $\mu_{11} = \rho\sigma_1\sigma_2$.

[*10.4.2*] The BVN($\underline{\theta}$) distribution is uniquely determined by the joint central moments μ_{10}, μ_{01}, μ_{20}, μ_{02}, and μ_{11}, that is, by μ_1, μ_2, σ_1^2, σ_2^2, and $\rho\sigma_1\sigma_2$, respectively.

[*10.4.3*] The *joint moment-generating function* of a BVN($\underline{\theta}$) distribution (defined in [10.1.1]) is $M(t_1,t_2)$, where

$$M(t_1,t_2) = \exp[t_1\mu_1 + t_2\mu_2 + (1/2)(t_1^2\sigma_1^2 + 2\rho t_1 t_2 \sigma_1\sigma_2 + t_2^2\sigma_2^2)]$$

and the *joint cumulant-generating function* is $K(t_1,t_2)$, where

$$K(t_1,t_2) = t_1\mu_1 + t_2\mu_2 + (1/2)(t_1^2\sigma_1^2 + 2\rho t_1 t_2 \sigma_1\sigma_2 + t_2^2\sigma_2^2)$$

Hence all cumulants κ_{rs} vanish if $r > 2$ or $s > 2$.

[*10.4.4*] Let $\lambda_{rs} = \mu_{rs}/(\sigma_1^r\sigma_2^s)$, the *standardized joint central moment* of order (r,s) of (X,Y). Then

$$\lambda_{rs} = 0 \qquad \text{whenever } r + s \text{ is odd}$$

Further,

$$\lambda_{11} = \rho \qquad \lambda_{31} = \lambda_{13} = 3\rho \qquad \lambda_{51} = \lambda_{15} = 15\rho$$

$$\lambda_{71} = \lambda_{17} = 105\rho \qquad \lambda_{22} = 1 + 2\rho^2$$

$$\lambda_{24} = \lambda_{42} = 3(1 + 4\rho^2) \qquad \lambda_{26} = \lambda_{62} = 15(1 + 6\rho^2)$$

$$\lambda_{33} = 3\rho(3 + 2\rho^2) \qquad \lambda_{35} = \lambda_{53} = 15\rho(3 + 4\rho^2)$$

$$\lambda_{44} = 3(3 + 24\rho^2 + 8\rho^4)$$

See Kendall and Stuart (1977, p. 94) who also give the following recursion formulas and expansions for $\{\lambda_{rs}\}$:

$$\lambda_{rs} = (r + s - 1)\rho\lambda_{r-1,s-1} + (r - 1)(s - 1)(1 - \rho^2)\lambda_{r-2,s-2}$$

If $t = \min(r,s)$,

$$\lambda_{2r,2s} = \frac{(2r)!(2s)!}{2^{r+s}} \sum_{j=0}^{t} \frac{(2\rho)^{2j}}{(r-j)!(s-j)!(2j)!}$$

$$\lambda_{2r+1,2s+1} = \frac{(2r+1)!(2s+1)!}{2^{r+s}} \rho \sum_{j=0}^{t} \frac{(2\rho)^{2j}}{(r-j)!(s-j)!(2j+1)!}$$

[*10.4.5*] If (X,Y) has a BVN($\underline{\theta}$) distribution as defined in [10.1.1], then $E|X^rY^s|$ is the *joint absolute moment* of order (r,s), and $E|(X - \mu_1)^r(Y - \mu_2)^s|$ is the *joint absolute central moment* of order (r,s). Let

$$\nu_{rs} = E\left|\frac{(X - \mu_1)^r(Y - \mu_2)^s}{\sigma_1^r\sigma_2^s}\right|$$

the *standardized joint absolute moment* of order (r,s). If (Z_1,Z_2) has an SBVN(ρ) distribution, then $\nu_{rs} = E|Z_1^r Z_2^s|$. If $\tau = \sqrt{1 - \rho^2}$, then (Kamat, 1953, pp. 26-27)

$$\nu_{11} = 2\{\tau + \rho \arcsin \rho\}/\pi$$

$$\nu_{12} = \nu_{21} = (1 + \rho^2)\sqrt{2/\pi}$$

$$\nu_{13} = \nu_{31} = 2\{\tau(2 + \rho^2) + 3\rho \arcsin \rho\}/\pi \qquad \nu_{22} = 1 + 2\rho^2$$

$$\nu_{14} = \nu_{41} = (3 + 6\rho^2 - \rho^4)\sqrt{2/\pi} \qquad \nu_{23} = \nu_{32} = 2(1 + 3\rho^2)\sqrt{2/\pi}$$

$$\nu_{24} = \nu_{42} = 3(1 + 4\rho^2)$$

$$\nu_{33} = 2\{(4 + 11\rho^2)\tau + 3\rho(3 + 2\rho^2)\arcsin \rho\}/\pi$$

In series form,

$$\nu_{m,n} = \pi^{-1}2^{(m+n)/2}(1 - \rho^2)^{(m+n+1)/2} \sum_{k=0}^{\infty} \Gamma\{(m + 1)/2 + k\} \cdot \Gamma\{(n + 1)/2 + k\}(2\rho)^{2k}/(2k)!$$

[*10.4.6*] If (X,Y) has a SBVN(ρ) distribution, and if Z = max(X,Y), then $E(Z) = \sqrt{(1 - \rho)/\pi}$ (David, 1970, p. 41).

10.5 SAMPLING DISTRIBUTIONS

[*10.5.1*] The *joint pdf of a sample* (X_1,Y_1); (X_2,Y_2); ...; (X_n,Y_n) of size n from a BVN distribution is given by

$$f(\underline{x}_n,\underline{y}_n;\mu_1,\mu_2;\sigma_1^2,\sigma_2^2,\rho)$$

$$= [(2\pi\sigma_1\sigma_2)^n(1 - \rho^2)^{n/2}]^{-1} \exp[-\tfrac{1}{2}(1 - \rho^2)^{-1}\{\Sigma(x_i - \mu_1)^2/\sigma_1^2 - 2\rho\Sigma[(x_i - \mu_1)(y_i - \mu_2)/(\sigma_1\sigma_2)] + \Sigma(y_i - \mu_2)^2/\sigma_2^2\}]$$

$$= f(\bar{x},\bar{y};\mu_1,\mu_2,\sigma_1^2/n,\sigma_2^2/n,\rho)h(s_1^2,s_2^2,r;\sigma_1^2,\sigma_2^2,\rho)$$

the product of the joint pdf of $(\bar{X},\bar{Y})$ and the joint pdf of (S_1^2,S_2^2,R), given in [10.5.2] below, where

$$S_1^2 = \Sigma(X_i - \bar{X})^2/(n - 1) \qquad S_2^2 = \Sigma(Y_i - \bar{Y})^2/(n - 1)$$

and

$$R = \frac{\Sigma(x_i - \bar{X})(Y_i - \bar{Y})}{[\Sigma(X_i - \bar{X})^2\Sigma(Y_i - \bar{Y})^2]^{1/2}}$$

which is the sample or product-moment correlation coefficient (Kendall and Stuart, 1977, pp. 412-414). Thus $(\bar{X},\bar{Y},S_1^2,S_2^2,R)$ is jointly sufficient for $\underline{\theta}$, and $(\bar{X},\bar{Y})$ is jointly independent of (S_1^2,S_2^2,R).

[*10.5.2*] The distribution of $(\bar{X},\bar{Y})$ is $\mathrm{BVN}(\mu_1,\mu_2;\sigma_1^2/n,\sigma_2^2/n,\rho)$ (Kendall and Stuart, 1977, p. 413).

The distribution of (S_1^2,S_2^2,R) has joint pdf given by (Kendall and Stuart, 1977, p. 413; Fisher, 1915, pp. 507-521)

$$\begin{aligned} h(s_1^2,s_2^2,r;\sigma_1^2,\sigma_2^2,\rho) &= [n\{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}\}^{n-1}]^{-1}(s_1s_2)^{n-2} \\ &\quad \cdot (1 - r^2)^{(n-4)/2} \\ &\quad \cdot \exp\left[-\frac{n}{2(1 - \rho^2)}\left\{\frac{s_1^2}{\sigma_1^2} - 2\frac{\rho r s_1 s_2}{\sigma_1\sigma_2} + \frac{s_2^2}{\sigma_2^2}\right\}\right], \\ &\quad s_1 > 0,\ s_2 > 0, \qquad |r| \le 1 \end{aligned}$$

If in addition, $\rho = 0$, then R, S_1^2, and S_2^2 are mutually independent (Kshirasagar, 1972, pp. 32-34).

[*10.5.3*] The *distribution of the sample correlation coefficient* R has attracted considerable attention in the literature. Fisher (1915) derived the pdf (Kendall and Stuart, 1977, pp. 413-415)

$$\begin{aligned} g(r;\rho) &= \frac{(1 - \rho^2)^{(n-1)/2}}{\pi\Gamma(n - 2)}(1 - r^2)^{(n-4)/2} \\ &\quad \cdot \frac{d^{n-2}}{d(r\rho)^{n-2}}\left\{\frac{\arccos(-r\rho)}{\sqrt{1 - r^2\rho^2}}\right\}, \qquad |r| \le 1 \end{aligned}$$

Cramér (1946, p. 398) gives an infinite series representation for $g(r;\rho)$:

$$g(r;\rho) = \frac{2^{n-3}}{\pi(n-3)!}(1-\rho^2)^{(n-1)/2}(1-r^2)^{(n-4)/2} \cdot \sum_{i=0}^{\infty}\left\{\Gamma\left(\frac{n+i-1}{2}\right)\right\}^2 \frac{(2\rho r)^i}{i!}$$

if $|r| \le 1$ and $|\rho| < 1$; $g(r,\rho) = 0$ otherwise. Integration term by term leads to rapid convergence of $\Pr(0 < R < r)$ with a good bound on error; see Guenther (1977, pp. 45-48), who also gives procedures for computing such probabilities which can be used with desk calculators programmed to compute probabilities based on the F distribution. See also Johnson and Kotz (1970, pp. 221-233) for some other expressions for $g(r;\rho)$.

[*10.5.4*] David (1938, 1954) has tabled percentage points and ordinates of the distribution of R as follows:

$n = 3(1)25$

$\rho = 0.0(0.1)0.4$	$r = -1.00(0.05)1.00$	Ordinate: 2 decimal places
$\rho = 0.5(0.1)0.9$	$r = -1.00(0.05)0.60(0.025)1.00$	
$\rho = 0.90$	$r = 0.80(0.01)0.95(0.005)1.00$	

$n = 50$

$\rho = 0.0(0.1)0.4$	$r = -0.75(0.05)1.00$	Ordinate: 2 decimal places
$\rho = 0.5(0.1)0.7$	$r = -0.25(0.05)1.00$	
$\rho = 0.80$	$r = 0.22(0.02)1.00$	
$\rho = 0.90$	$r = 0.61(0.01)1.00$	Ordinate: 1 dec. place

Probabilities are given to 5 decimal places; the 1954 edition contains a few corrections to errors in the 1938 edition on pp. viii and xxxii.

[*10.5.5*] The *moments of the distribution of* R are as follows (Ghosh, 1966, pp. 258-262):

$$\mu = E(R) = \frac{2}{n-1}\left[\frac{\Gamma(n/2)}{\Gamma\{(n-1)/2\}}\right]^2 \rho F(1/2, 1/2, (n+1)/2; \rho^2)$$

where

$$F(a,b,c;\rho^2) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{j=0}^{\infty} \left[\frac{\Gamma(a + j)\Gamma(b + j)}{\Gamma(c + j)} \frac{\rho^{2j}}{j!}\right]$$

which is the hypergeometric function, and

$$E(R^2) = 1 - (n - 2)(1 - \rho^2)(n - 1)^{-1}F(1, 1, (n + 1)/2; \rho^2)$$

If $m = n + 6$, then (Ghosh, 1966, pp. 258-262):

$$E(R) = \rho - \frac{\rho(1 - \rho^2)}{2m}\left\{1 + \frac{9}{4m}(3 + \rho^2) + \frac{3(121 + 70\rho^2 + 25\rho^4)}{8m^2}\right\} + 0(m^{-4})$$

$$\mathrm{Var}(R) = \frac{(1 - \rho^2)^2}{m}\left\{1 + \frac{14 + 11\rho^2}{2m} + \frac{98 + 130\rho^2 + 75\rho^4}{2m^2}\right\} + 0(m^{-4})$$

Hotelling (1953, p. 212) shows that

$$E(R) = \rho + (1 - \rho^2)\left\{-\frac{\rho}{2(n - 1)} + \frac{\rho - 9\rho^3}{8(n - 1)^2} + \frac{\rho + 42\rho^3 - 75\rho^5}{16(n - 1)^3}\right\} + 0(n^{-4})$$

$$\mathrm{Var}(R) = (1 - \rho^2)^2\left\{\frac{1}{n - 1} + \frac{11\rho^2}{2(n - 1)^2} + \frac{-24\rho^2 + 75\rho^4}{2(n - 1)^3}\right\} + 0(n^{-4})$$

$$\mu_3(R) = (1 - \rho^2)^3\left\{-\frac{6\rho}{(n - 1)^2} + \frac{15\rho - 88\rho^3}{(n - 1)^3}\right\} + 0(n^{-4})$$

and

$$\mu_4(R) = (1 - \rho^2)^4\left\{\frac{3}{(n - 1)^2} + \frac{-6 + 105\rho^2}{(n - 1)^3}\right\} + 0(n^{-4})$$

Thus if n is large, the bias in E(R) is of order (n^{-1}), the variance is of order (n^{-1}), while the skewness and kurtosis differ from normality by order $(n^{-1/2})$ and (n^{-1}), respectively, i.e.,

$$\mu_3(R)/\sigma^3(R) = -6\rho n^{-1/2} + 0(n^{-3/2})$$

$$\mu_4(R)/\sigma^4(R) = 3 - 6(1 - 12\rho^2)n^{-1} + 0(n^{-2})$$

[*10.5.6*] *Distribution of* R *When* $\rho = 0$. We have (Kendall and Stuart, 1977, p. 416; 'Student', 1908)

$$g(r;0) = (1 - r^2)^{(n-4)/2}/B(1/2, (n - 2)/2), \qquad 0 < |r| < 1$$

If $T = \{(n - 2)R^2/(1 - R^2)\}^{1/2}$, then when $\rho = 0$, T has a Student t distribution with n - 2 degrees of freedom. This property is very useful for tests of independence of X and Y in a BVN population.

Thus the distribution g(r;0) is symmetric about zero with mean zero and variance 1/(n - 1). The kurtosis is given by

$$\mu_4(R \mid 0)/\sigma^4(R \mid 0) = 3 - 6/(n + 1)$$

[*10.5.7*] *The Arcsine Transformation of* R. The statistic arcsin R is unbiased, i.e., E(arcsin R) = arcsin ρ (Harley, 1956, pp. 219-224; Harley, 1957, pp. 273-275; Kendall and Stuart, 1967, p. 294).

Let $X = (R - \rho)/(1 - \rho R)$ and Y = arcsin X. Then Y has an approximate normal distribution when n is large, where

$$E(Y) = \frac{\rho}{2(n - 1)} + \frac{3\rho + \rho^3}{8(n - 1)^2} + \frac{3\rho + 3\rho^5}{16(n - 1)^3} + O(n^{-4})$$

$$\mathrm{Var}(Y) = \frac{1}{n - 1} + \frac{2 - \rho^2}{2(n - 1)^2} + \frac{8 - 3\rho^2 - 6\rho^4}{12(n - 1)^3} + O(n^{-4})$$

This approximation (Sankaran, 1958, pp. 567-571) is accurate to 3 decimal places in giving probabilities for R; in a simplified version, $Y \sim N(0, (n - 2)^{-1})$, approximately, but this is less accurate.

[*10.5.8*] Various transformations of R leading to approximate normality have been suggested, the most well known of which is Fisher's z, where (Fisher, 1921), pp. 3-32; Kendall and Stuart, 1977, p. 419)

$$z = \frac{1}{2}\ln[(1 + R)/(1 - R)]$$

Fisher expanded the exponent of the pdf of z in inverse powers of n - 1, and in powers of ζ, where $\zeta = (1/2)\ln[(1 + \rho)/(1 - \rho)]$.

Although the distribution of R tends to normality with increasing sample size, it does so rather slowly; this, together with the mathematical difficulties inherent in $g(r;\rho)$ when $\rho \neq 0$ led to the search for a suitable approximation. Fisher's z tends to normality much faster, but also has a variance which is (almost) independent of ρ.

Fisher (1921) expanded the moments of z as follows, where corrections of Gayen (see Hotelling, 1953, p. 216) are included (Kendall and Stuart, 1977, p. 419)

$$E(z) = \zeta + \frac{\rho}{2(n-1)}\left\{1 + \frac{5+\rho^2}{4(n-1)} + \frac{11 + 2\rho^2 + 3\rho^4}{8(n-1)^2}\right\} + 0(n^{-4})$$

$$\text{Var}(z) = \frac{1}{n-1}\left\{1 + \frac{4-\rho^2}{2(n-1)} + \frac{22 - 6\rho^2 - 3\rho^4}{6(n-1)^2}\right\} + 0(n^{-4})$$

$$\frac{\mu_3}{\sigma^3} = \frac{\rho^3}{(n-1)^{3/2}} + 0(n^{-2}) \qquad \text{(skewness)}$$

$$\frac{\mu_4}{\sigma^4} = 3 + \frac{2}{n-1} + \frac{4 + 2\rho^2 - 3\rho^4}{(n-1)^2} + 0(n^{-3}) \qquad \text{(kurtosis)}$$

The most common approximation treats z as a $N(\zeta, 1/(n-3))$ variable, and others treat z as normal, taking Var(z) as $(n-3)^{-1}$ or as above to terms of order $(n-1)^{-3}$, and taking E(z) to terms of order $(n-1)^{-1}$ or $(n-1)^{-2}$. David (1954) and Kraemer (1973, pp. 1004-1008) state that great accuracy is achieved with higher-order approximations, even when $n \leq 10$, provided that $|\rho|$ is not too close to unity; when n = 11 and $\rho = 0.90$, the agreement with the exact sampling distribution of R is reasonably good.

Hotelling (1953) suggested two modifications of z to stabilize the variance further and give a closer approximation to normality. These are given by (Hotelling, 1953, pp. 223-224)

$$z^* = z - (3z + R)/\{4(n-1)\}$$

and

$$z^{**} = z^* - (23z + 33R - 5R^3)/\{96(n-1)^2\}$$

Then

$$E(z^*) = \zeta - \frac{3\zeta + \rho}{4(n-1)} + \frac{\rho}{2(n-1)} + \frac{3\rho}{8(n-1)^2} + O(n^{-3})$$

and $E(z^{**})$ differs from $E(z^*)$ only by terms of order n^{-3}. However,

$$\mathrm{Var}(z^*) = (n-1)^{-1} + O(n^{-3})$$

while

$$\mathrm{Var}(z^{**}) = (n-1)^{-1} + O(n^{-4})$$

[*10.5.9.1*] Ruben (1966, pp. 518-519) approximates the distribution of R by the relation

$$\Pr(R \le r) \simeq \Phi\left[\frac{\sqrt{(n - 5/2)}\tilde{r} - \sqrt{(n - 3/2)}\tilde{\rho}}{\{1 + (1/2)(\tilde{r}^2 + \tilde{\rho}^2)\}^{1/2}}\right]$$

where $\tilde{r} = r/\sqrt{1 - r^2}$, $\tilde{\rho} = \rho/\sqrt{1 - \rho^2}$, and Φ is the standard normal cdf. This approximation improves upon Fisher's z transform, but lacks the variance-stabilizing property of z. See also Kshirsagar (1972, pp. 88-92).

[*10.5.9.2*] Kraemer (1973, pp. 1004-1008) has developed another approximation. If $\rho' = \rho'(\rho,n)$, such that $|\rho'| \ge \rho$, $\rho' = \rho$ whenever ρ is 0, 1, or -1, $\rho'(-\rho,n) = -\rho'(\rho,n)$ and as $n \to \infty$ we have that $\rho' \to \rho$, then

$$(R - \rho')\sqrt{\frac{(n-2)}{(1 - R^2)(1 - \rho'^2)}}$$

is approximately distributed as Student's t with n - 2 degrees of freedom. If $n > 10$, the optimum choice for ρ' is the median of the distribution of R, given ρ and n (tabled in Kraemer, 1973), but if $n > 25$, one may take ρ' equal to ρ. Kraemer compared her approximation with those for z mentioned in [10.5.8] and found it to give more accurate probabilities for R if $|\rho| < 0.8$.

[*10.5.9.3*] Mudholkar and Chaubey (1976, pp. 163-172) approximate the distribution of Fisher's z by adjusting the kurtosis with a mixture of normal and logistic distributions. Using the notation of [10.5.8], let $y = [z - E(z)]/\sqrt{Var(z)}$, where $E(z)$ and $Var(z)$ are approximated as in [10.5.8]. Then

$$\Pr(R \le r) \simeq \lambda\Phi(y) + (1 - \lambda)\{1 + \exp(-\pi y/\sqrt{3})\}^{-1}$$

where Φ is the standard normal cdf and

$$1 - \lambda = (\mu_4(z)/\sigma^4(z) - 3)/1.2$$
$$\simeq \left[\frac{2}{n - 1} + \frac{4 + 2\rho^2 - 3\rho^4}{(n - 1)^2}\right]/1.2$$

The 100(α)th percentile of z is approximated by $\lambda z_{1-\alpha} - (1 - \lambda) \times \sqrt{3}\{\ln(\alpha^{-1} - 1)\}/\pi$, where $\Phi(z_{1-\alpha}) = \alpha$. The authors compare exact and approximate probabilities based on their mixture distribution with those of Fisher's z transformation (see [10.5.8]), of Ruben (1966), and of Kraemer (1973) for a range of values of R when $\rho = 0.0(0.1)0.90$, and $n = 11, 21$; and they compare percentiles of R computed by these methods for $\rho = 0.5, 0.9$; $n = 11, 25, 50$; $\alpha > 0.95$, with exact values. The approximations of Kraemer and of Mudholkar and Chaubey are uniformly accurate to 2 or 3 decimal places; those of Ruben are almost as good, while Fisher's z approximation sacrifices accuracy for simplicity, unless n is large.

[*10.5.10*] Let $m = n - (5/2) + \rho^2/4$. Then (Konishi, 1978, pp. 654-655)

$$\Pr\{\sqrt{m}(z - \zeta) < x\} = \Phi(x) - (1/2)(\rho m^{-1/2} + x^3 m^{-1}/6)\phi(x) + O(m^{-3/2})$$

This gives an improvement upon the $N(\zeta, (n - 3)^{-1})$ approximation for Fisher's z and can be used to give an accurate approximation to the distribution of R over its whole domain. Based on calculations for n = 11, 25, and 50, and $\rho = 0.1(0.2)0.9$, Konishi (1978) found that the above appears to be more accurate when $\rho \ge 0.3$ than approximations due to Ruben and Kraemer, discussed in [10.5.9.1] and [10.5.9.2], respectively; even when n is as small as 11, the above gives high accuracy over all values of R. For the range of values of n and of ρ tabled, Konishi found that, for $\Pr(R \le x)$,

$$\text{Max}|\text{exact value - approximate value}| \le 0.00102$$

[*10.5.11*] If n pairs of observations from a BVN distribution are available, the *sample regression coefficient* of Y on X is b, where $b = RS_2/S_1$ (see [10.5.1]). Let $\beta = \rho\sigma_2/\sigma_1$, the population regression coefficient of Y on X; then the variable $\sqrt{n - 1}(b - \beta) \times \sigma_1/\{\sigma_2(1 - \rho^2)\}^{1/2}$ has a Student t distribution with n - 1 degrees of freedom (Kendall and Stuart, 1977, p. 421).

The variable $(b - \beta)S_1\sqrt{n - 2}/\{S_2(1 - R^2)^{1/2}\}$ has a Student t distribution with n - 2 degrees of freedom (Kendall and Stuart, 1977, p. 422). This result is more useful, since ρ, σ_1, and σ_2 are generally unknown. When $\rho = 0$, this is the variable described in [10.5.6].

The sampling distribution of b has (Kendall and Stuart, 1973, pp. 325-326) mean β,

$$\text{Var}(b) = (\sigma_2^2/\sigma_1^2)(1 - \rho^2)/(n - 3)$$

$$\mu_3(b)/\sigma^3(b) = 0 \qquad \text{(skewness)}$$

$$\mu_4(b)/\sigma^4(b) = 6/(n - 5) + 3 \qquad \text{(kurtosis)}$$

[*10.5.12*] If $\sigma_1^2 = \sigma_2^2$, the sample *intraclass correlation coefficient* U may be of interest--for example, when it is not clear which variable should be labeled X and which Y. Thus, X and Y might denote the heights of twin brothers, and a measure of association between them is desired (Kendall and Stuart, 1973, pp. 314-316). The statistic U is defined by

$$U = 2S_{12}/(S_1^2 + S_2^2)$$

where S_1^2 and S_2^2 are defined in [10.5.1] and

$$S_{12} = \sum_{i=1}^{n} \{(X_i - \bar{X})(Y_i - \bar{Y})\}/(n - 1)$$

which is the sample covariance. The pdf of U (De Lury, 1938, pp. 149-151) is given by

$$g(u) = \begin{cases} \dfrac{\Gamma\{(n+1)/2\}}{\sqrt{\pi}\,\Gamma(n/2)} (1-\rho^2)^{(1/2)n}(1-\rho u)^{-n} \\ \quad \cdot\ (1-u^2)^{(n-1)/2}, & |u| \le 1 \\ 0, & |u| > 1 \end{cases}$$

If $\rho = 0$, the distribution of U is identical to that of R in [10.5.6], except that n is replaced by n + 1.

[*10.5.13*] $(X_1,Y_1), \ldots, (X_n,Y_n)$ is a random sample from a BVN(θ) distribution, as defined in [10.1.1]. Suppose

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$$

are order statistics for $(X_1,\ldots,X_n)$; let $Y_{[1]}, Y_{[2]}, \ldots, Y_{[n]}$ be the corresponding Y terms not necessarily ordered. Further, let $\mu_{r;n}$, $\sigma^2_{r;n}$, and $\sigma_{rs;n}$ be the expected values of $Z_{(r)}$, $\mathrm{Var}(Z_{(r)})$, and $\mathrm{Cov}(Z_{(r)},Z_{(s)})$, respectively, where $Z_{(1)} \le \cdots \le Z_{(n)}$ are order statistics from a sample of size n from a N(0,1) population. Then (David, 1970, p. 41)

$$EY_{[r]} = \mu_2 + \rho\sigma_2\mu_{r;n}$$

$$\mathrm{Var}\ Y_{[r]} = (1-\rho^2)\sigma_2^2 + \rho^2\sigma_2^2\sigma_{r;n}^2$$

$$\mathrm{Cov}(X_{(r)},Y_{[s]}) = \rho\sigma_1\sigma_2\sigma_{rs;n}$$

$$\mathrm{Cov}(Y_{[r]},Y_{[s]}) = \rho^2\sigma_2^2\sigma_{rs;n}, \qquad r \ne s \qquad r,s = 1, 2, \ldots, n$$

See also Watterson (1959, pp. 814-824).

10.6 MISCELLANEOUS RESULTS

[*10.6.1*] Most of this section will concern *truncated BVN distributions*. A good discussion appears in Johnson and Kotz (1972, pp. 112-117), with results for cases in which both variables are singly or doubly truncated. We present some results and properties for a singly truncated SBVN(ρ) distribution, where the variable X is truncated to include values greater than x = a only. The joint pdf g(x,y) is given by

$$g(x,y) = \frac{1}{1 - \Phi(a)} \frac{1}{2\pi\sqrt{1 - \rho^2}} \exp\left\{- \frac{1}{2(1 - \rho^2)}(x^2 - 2\rho xy + y^2)\right\}, \quad x > a$$

and $g(x,y) = 0$ if $x < a$.

[*10.6.2.1*] Let $E_T(X)$, $Var_T(X)$, and ρ_T denote the mean and variance of X and the correlation coefficient between X and Y in a SBVN(ρ) distribution singly truncated so that $X > a$. Also, let

$$q(a) = \phi(a)/\{1 - \Phi(a)\}$$
$$= \{R(x)\}^{-1}$$

where R(x) is Mills' ratio (see [3.4] to [3.7]). Then (Rao et al., 1961, pp. 434-435)

$$E_T(X) = q(a) \qquad E_T(Y) = \rho q(a)$$
$$E_T(X^2) = 1 + aq(a) \qquad E_T(Y^2) = 1 + \rho^2 aq(a)$$
$$Var_T(X) = 1 - q(a)\{q(a) - a\} \qquad Var_T(Y) = 1 - \rho^2 q(a)\{q(a) - a\}$$
$$\rho_T = \rho\sqrt{Var_T(X)/Var_T(Y)}$$

Since $Var_T(X) \le Var_T(Y)$, it follows that

$$|\rho_T| \le |\rho|$$

See also Weiler (1959, pp. 73-81).

[*10.6.2.2*] Tallis (1961, p. 225) shows that the moment generating function of the truncated distribution above is given by

$$E(e^{t_1X+t_2Y}) = (\{1 - \Phi(a - t_1 - \rho t_2)\}/\{1 - \Phi(a)\})$$
$$\cdot \exp\{(1/2)(t_1^2 + 2\rho t_1t_2 + t_2^2)\}$$

[*10.6.2.3*] Gajjar and Subrahmaniam (1978, p. 456) give the following recursive relation between the moments of the above distribution:

$$\mu'_{r+1,s} - a\mu'_{r,s} = r(1 - \rho^2)\mu'_{r-1,s} + \rho\mu'_{r,s+1} - a(r - 1)(1 - \rho^2)\mu'_{r-2,s} - a\rho\mu'_{r-1,s-1}, \quad r \geq 2 \quad s \geq 0$$

See also Johnson and Kotz (1972, p. 114).

[*10.6.3*] Truncation of a SBVN(ρ) distribution so that $X > a$ does not affect the regression of Y on X (Rao et al., 1968, p. 435); see [10.1.4]. With the notation of [10.6.2.1] above, however, the regression of X on Y is the expected value of X, given $Y = y$, i.e. (Johnson and Kotz, 1972, p. 113),

$$E(X \mid Y = y) = \rho y + \sqrt{1 - \rho^2}\, q((a - \rho y)/\sqrt{1 - \rho^2}$$

[*10.6.4*] Rao et al. (1968, pp. 433-436) show that, if ρ_T^* is the correlation coefficient between the sample variances S_1^2 and S_2^2 of X and Y, respectively, in samples of size n from a SBVN(ρ) distribution truncated as above, and if μ_{rs} is the (r,s)th central moment $E\{(X - E_T(X))^r(Y - E_T(Y))^s\}$, then to $O(n^{-1})$,

$$\rho_T^* = (\mu_{22} - \mu_{20}\mu_{02})/\{(\mu_{40} - \mu_{20}^2)(\mu_{04} - \mu_{02}^2)\}^{1/2}$$

Further (see also Gajjar and Subrahmaniam, 1978, p. 458),

$$\rho_T^* \geq \rho_T^2 \qquad \text{and} \qquad |\rho_T^*| \leq |\rho_T|$$

for all choices of a and of ρ. Gajjar and Subrahmaniam (1978) give expressions for the moments of the sample correlation coefficient r_T; see also Cook (1951, pp. 368-376).

[*10.6.5*] Let $X_1, X_2, \ldots, X_n$ be a random sample from a $N(\mu,\sigma^2)$ population, and $a_1, \ldots, a_n, b_1, \ldots, b_n$ be constants. Then $(\Sigma_{i=1}^n a_iX_i, \Sigma_{i=1}^n b_iX_i)$ has a BVN($\underline{\theta}$) distribution, where, in the notation of [10.1.1],

$$\underline{\theta}' = [(\Sigma a_i)\mu, (\Sigma b_i)\mu; (\Sigma a_i^2)\sigma^2, (\Sigma b_i^2)\sigma^2, (\Sigma a_i b_i)/\sqrt{(\Sigma a_i^2)(\Sigma b_i^2)}]$$

$$= [(\Sigma a_i)\mu, (\Sigma b_i)\mu; \sigma^2, \sigma^2, \Sigma a_i b_i] \quad \text{if} \quad \Sigma a_i^2 = \Sigma b_i^2 = 1$$

and

$$\underline{\theta}' = [0, 0; \sigma^2, \sigma^2, \Sigma a_i b_i] \quad \text{if, in addition,} \quad \Sigma a_i = \Sigma b_i = 0$$

Hence $\Sigma a_i X_i$ and $\Sigma b_i X_i$ are independent if and only if $\Sigma a_i b_i = 0$, that is, these linear transformations are orthogonal. In the design of experiments, $X_1, \ldots, X_n$ may represent treatment means under equal group sizes; $\Sigma a_i X_i$ is termed a contrast whenever $\Sigma a_i = 0$, the normalizing condition $\Sigma a_i^2 = 1$ being added for convenience.

REFERENCES

The numbers in square brackets give the sections in which the corresponding references are cited.

Abramowitz, M., and Stegun, I. A. (eds.) (1964). *Handbook of Mathematical functions*, Washington, D.C.: National Bureau of Standards. [10.2.1, 4]

Aroian, L. A., Taneja, V. S., and Cornwell, L. W. (1978). Mathematical forms of the distribution of the product of two normal variables, *Communications in Statistics* A7, 165-172. [10.1.9]

Bark, L. S., Bol'shev, L. N., Kuznetsov, P. E., and Cherenkov, A. P. (1964). *Tables of the Rayleigh-Rice Distributions*, Moscow: Voychislityelnoy Tsentr. [10.3.6.3; Table 10.2]

Beyer, W. H. (ed.) (1966). *Handbook of Tables for Probability and Statistics*, Cleveland: Chemical Rubber Co. [Table 10.3]

Bhattacharyya, A. (1943). On some sets of sufficient conditions leading to the normal bivariate distribution, *Sankhyā* 6, 399-406. [10.1.5]

Bickel, P. J., and Doksum, K. A. (1977). *Mathematical Statistics: Basic Ideas and Topics*, San Francisco: Holden-Day. [10.1.1, 4]

Borth, D. M. (1973). A modification of Owen's method for computing the bivariate normal integral, *Applied Statistics* 22, 82-85. [10.2.7]

Brennan, L. E., and Reed, I. S. (1965). A recursive method of computing the Q Function, *IEEE Transactions on Information Theory*, 312-313. [10.3.5]

Burington, R. S., and May, D. C. (1970). *Handbook of Probability and Statistics with Tables*, Sandusky, Ohio: Handbook Publishers. [Table 10.2]

Cadwell, J. H. (1951). The bivariate normal integral, *Biometrika* 38, 475-479. [10.2.8]

Cook, M. B. (1951). Two applications of bivariate k-statistics, *Biometrika* 38, 368-376. [10.6.4]

Craig, C. C. (1936). On the frequency function of xy, *Annals of Mathematical Statistics* 7, 1-15. [10.1.9]

Cramér, H. (1946). *Mathematical Methods of Statistics*, Princeton, N.J.: Princeton University Press. [10.5.3]

Daley, D. J. (1974). Computation of bi- and tri-variate normal integrals, *Applied Statistics* 23, 435-438. [10.2.8]

David, F. N. (1938, 1954). *Tables of the Correlation Coefficient*, Cambridge, England: Cambridge University Press. [10.5.4, 8]

David, H. A. (1970). *Order Statistics*, New York: Wiley. [10.4.6; 10.5.13]

De Lury, D. B. (1938). Note on correlations, *Annals of Mathematical Statistics* 9, 149-151. [10.5.12]

Di Donato, A. R., and Jarnagin, M. P. (1962). A method for computing the circular coverage function, *Mathematics of Computation* 16, 347-355. [10.3.1; 10.3.6.3; Table 10.2]

Drezner, Z. (1978). Computation of the bivariate normal integral, *Mathematics of Computation* 32, 277-279. [10.2.10]

Eckler, A. R. (1969). A survey of coverage problems associated with point and area targets, *Technometrics* 11, 561-589. [10.3.4, 9, 13; 10.3.14.2]

Fieller, E. C. (1932). The distribution of the index in a normal bivariate population, *Biometrika* 24, 428-440. [10.1.6]

Fieller, E. C., Lewis, T., and Pearson, E. S. (1955). *Correlated Random Normal Deviates*, Tracts for Computers, Vol. 26, London: Cambridge University Press; corrigenda (1956): *Biometrika* 43, 496. [10.2.5]

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population, *Biometrika* 10, 507-521. [10.5.2]

Fisher, R. A. (1921). On the probable error of a coefficient of correlation deduced from a small sample, *Metron* 1(4), 3-32. [10.5.8]

Fisk, P. R. (1970). A note on a characterization of the multivariate normal distribution, *Annals of Mathematical Statistics* 41, 486-494. [10.1.5]

Gajjar, A. V., and Subrahmaniam, K. (1978). On the sample correlation coefficient in the truncated bivariate normal population, *Communications in Statistics* B7, 455-478. [10.6.2.3, 4]

Ghosh, B. K. (1966). Asymptotic expansions for the moments of the distribution of sample correlation coefficient, *Biometrika* 53, 258-262. [10.5.5]

Gideon, R. A., and Gurland, J. (1978). A polynomial type approximation for bivariate normal variates, *SIAM Journal on Applied Mathematics* 34, 681-684.

Gilliland, D. C. (1962). Integral of the bivariate normal distribution over an offset circle, *Journal of the American Statistical Association* 57, 758-768. [10.3.5, 12; 10.3.14.2]

Grad, A., and Solomon, H. (1955). Distribution of quadratic forms and some applications, *Annals of Mathematical Statistics* 26, 464-477. [Table 10.3]

Groenewoud, C., Hoaglin, D. C., and Vitalis, J. A. (1967). *Bivariate Normal Offset Circle Probability Tables*, 2 volumes, Buffalo: Cornell Aeronautical Laboratory. [10.3.14.1]

Groves, A. D., and Smith, E. S. (1957). Salvo hit probabilities for offset circular targets, *Operations Research* 5, 222-228. [Table 10.2]

Grubbs, F. E. (1964). Approximate circular and noncircular offset probabilities of hitting, *Operations Research* 12, 51-62. [10.3.6.1; 10.3.11; 10.3.14.2]

Guenther, W. C. (1977). Desk calculation of probabilities for the distribution of the sample correlation coefficient, *The American Statistician* 31(1), 45-48. [10.5.3]

Guenther, W. C., and Terragno, P. J. (1964). A review of the literature on a class of coverage problems, *Annals of Mathematical Statistics* 35, 232-260. [10.3.4; 10.3.8.1; 10.3.10]

Harley, B. I. (1956). Some properties of an angular transformation for the correlation coefficient, *Biometrika* 43, 219-224. [10.5.7]

Harley, B. I. (1957). Relation between the distribution of noncentral t and of a transformed correlation coefficient, *Biometrika* 44, 273-275. [10.5.7]

Harter, H. L. (1960). Circular error probabilities, *Journal of the American Statistical Association* 55, 723-731. [10.3.9; Table 10.3]

Hinkley, D. V. (1969). On the ratio of two correlated normal random variables, *Biometrika* 56, 635-639. [10.1.6]

Hotelling, H. (1953). New light on the correlation coefficient and its transforms, *Journal of the Royal Statistical Society* B15, 193-232 (with discussion). [10.5.5, 8]

Johnson, N. L., and Kotz, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions*, Vol. 2, New York: Wiley. [10.3.15; 10.5.4]

Johnson, N. L., and Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions*, New York: Wiley. [10.1.5; 10.2.9; 10.6.1; 10.6.2.3; 10.6.3]

Johnson, N. L., and Kotz, S. (1975). On some generalized Farlie-Gumbel-Morgenstern distributions, *Communications in Statistics* 4, 415-427. [10.1.2]

Kamat, A. R. (1953). Incomplete and absolute moments of the multivariate normal distribution with some applications, *Biometrika* 40, 20-34. [10.4.5]

Kendall, M. G., and Stuart, A. (1973). *The Advanced Theory of Statistics*, Vol. 2 (3rd ed.), New York: Hafner. [10.1.5, 8; 10.5.7, 11, 12]

Kendall, M. G., and Stuart, A. (1977). *The Advanced Theory of Statistics*, Vol. 1 (4th ed.), New York: Macmillan. [10.1.5; 10.4.4; 10.5.1, 2, 3, 6, 8, 11]

Khatri, C. G., and Rao, C. R. (1976). Characterizations of multivariate normality, I: Through independence of some statistics, *Journal of Multivariate Analysis* 6, 81-94. [10.1.5]

Konishi, S. (1978). An approximation to the distribution of the sample correlation coefficient, *Biometrika* 65, 654-656. [10.5.10]

Kraemer, H. C. (1973). Improved approximation to the non-null distribution of the correlation coefficient, *Journal of the American Statistical Association* 68, 1004-1008. [10.5.8; 10.5.9.2, 3; 10.5.10]

Kshirsagar, A. (1972). *Multivariate Analysis*, New York: Dekker. [10.5.2; 10.5.9.1]

Lancaster, H. O. (1957). Some properties of the bivariate normal distribution considered in the form of a contingency table, *Biometrika* 44, 289-292. [10.1.8]

Laurent, A. G. (1957). Bombing problems--a statistical approach, *Operations Research* 5, 75-89. [10.1.7]

Lilliefors, H. W. (1957). A hand-computation determination of kill probability for weapons having spherical lethal volume, *Operations Research* 5, 416-421. [10.3.13]

Lowe, J. R. (1960). A table of the integral of the bivariate normal distribution over an offset circle, *Journal of the Royal Statistical Society* B22, 177-187. [10.3.14.1]

Mudholkar, G. S., and Chaubey, Y. P. (1976). On the distribution of Fisher's transformation of the correlation coefficient, *Communications in Statistics* B5, 163-172. [10.5.9.3]

National Bureau of Standards (1959). *Tables of the Bivariate Normal Distribution Function and Related Functions*, Applied Mathematics Series 50, Washington, D.C.: U.S. Government Printing Office. [10.2.1, 9]

Nicholson, C. (1943). The probability integral for two variables, *Biometrika* 33, 59-72. [10.2.4]

Owen, D. B. (1956). Tables for computing bivariate normal probabilities, *Annals of Mathematical Statistics* 27, 1075-1090. [10.2.3, 4, 7]

Owen, D. B. (1962). *Handbook of Statistical Tables*, Reading, Mass.: Addison-Wesley. [10.2.3; 10.3.3; 10.3.8.2]

Owen, D. B., McIntire, D., and Seymour, E. (1975). Tables using one or two screening variables to increase acceptable product under one-sided specifications, *Journal of Quality Technology* 7, 127-138. [10.2.6]

Pearson, K. (ed.) (1922). *Tables of the Incomplete Γ-Function*, London: H.M. Stationery Office. [10.3.6.1]

Pearson, K. (1931). *Tables for Statisticians and Biometricians*, Vol. 2, London: Cambridge University Press. [10.2.4]

Pólya, G. (1949). Remarks on computing the probability integral in one and two dimensions, *Proceedings of the First Berkeley Symposium on Mathematical Statistics and Probability,* 63-78, Berkeley: University of California Press. [10.2.1]

Rao, B. R., Garg, M. L., and Li, C. C. (1968). Correlation between the sample variances in a singly truncated bivariate normal distribution, *Biometrika* 55, 433-436. [10.6.2.1; 10.6.3, 4]

Rao, C. R. (1975). Some problems in characterization of the multivariate normal distribution, *A Modern Course on Statistical Distributions in Scientific Work,* Vol. 3, *Characterizations and Applications* (G. P. Patil et al., eds.), New York: D. Reidel. [10.1.5]

Read, C. B. (1971). Two casualty-estimation problems associated with area targets and the circular coverage function, *Operations Research* 19, 1730-1741. [10.3.2; 10.3.6.2]

Ruben, H. (1966). Some new results on the distribution of the sample correlation coefficient, *Journal of the Royal Statistical Society* B28, 513-525. [10.5.9.1, 3; 10.5.10]

Sankaran, M. (1958). On Nair's transformation of the correlation coefficient, *Biometrika* 45, 567-571. [10.5.7]

Sheppard, W. F. (1898). On the application of the theory of error to cases of normal distributions and normal correlation, *Philosophical Transactions of the Royal Society of London* 192, 101-167. [10.2.1]

Smirnov, N. V., and Bol'shev, L. N. (1962). *Tables for Evaluating a Function of a Bivariate Normal Distribution*, Moscow: Is'datel'stov Akademii Nauk SSR. [Table 10.1]

Solomon, H. (1953). Distribution of the measure of a random two-dimensional set, *Annals of Mathematical Statistics* 24, 650-656. [Table 10.2]

Sowden, R. R., and Ashford, J. R. (1969). Computation of the bi-variate normal integral, *Applied Statistics* 18, 169-180. [10.2.8]

"Student" (1908). On the probable error of a correlation coefficient, *Biometrika* 6, 302-310. [10.5.6]

Tallis, G. M. (1961). The moment generating function of the truncated multinormal distribution, *Journal of the Royal Statistical Society* B23, 223-229. [10.6.2.2]

Watterson, G. A. (1959). Linear estimation in censored samples from multivariate normal populations, *Annals of Mathematical Statistics* 30, 814-824. [10.5.13]

Wegner, L. H. (1963). Quick-count, a general war casualty estimation model, RM 3811-PR, The Rand Corporation. [10.3.6.2]

Weil, H. (1954). The distribution of radial error, *Annals of Mathematical Statistics* 25, 168-170. [10.1.7]

Weiler, H. (1959). Mean and standard deviations of a truncated normal bivariate distribution, *Australian Journal of Statistics* 1, 73-81. [10.6.2.1]

Weingarten, H., and Di Donato, A. R. (1961). A table of generalized circular error, *Mathematics of Computation* 15, 169-173. [Table 10.3]

Wilks, S. S. (1962). *Mathematical Statistics*, New York: Wiley. [10.1.4]

Wold, H. (1948). *Random Normal Deviates*, Tracts for Computers 25, Cambridge, England: Cambridge University Press. [10.2.5]

Yamauti, Z. (ed.) (1972). *Statistical Tables and Formulas with Computer Applications*, Tokyo: Japanese Standards Association. [Table 10.1]

Young, J. C., and Minder, C. E. (1974). Algorithm AS 76: An integral useful in calculating non-central t and bi-variate normal probabilities, *Applied Statistics* 23, 455-457. [10.2.8]

Zelen, M., and Severo, N. C. (1960). Graphs for bivariate normal probabilities, *Annals of Mathematical Statistics* 31, 619-624. [10.2.4]

INDEX