WILEY SERIES IN PROBABILITY
AND MATHEMATICAL STATISTICS

Normal Approximation and Asymptotic Expansions

R. N. Bhattacharya

Associate Professor of Mathematics
University of Arizona

R. Ranga Rao

Professor of Mathematics
University of Illinois

John Wiley & Sons, New York • London • Sydney • Toronto

TO GOURI AND SHANTHA

Library of Congress Cataloging in Publication Data

Bhattacharya, Rabindra Nath, 1937-
Normal approximation and asymptotic expansions.

(Wiley series in probability and mathematical statistics)
Bibliography: p.
Includes index.
1. Central limit theorem. 2. Convergence. 3. Asymptotic expansions. I. Rao, Ramaswamy Ranga, 1935- joint author. II. Title.

QA273.67.B48 519.2 75-35876
ISBN 0-471-07201-X

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Preface

This monograph presents in a unified way various refinements of the classical central limit theorem for independent random vectors and includes recent research on the subject. Most of the multidimensional results in this area are fairly recent, and significant advances over the last 15 years have led to a fresh outlook. The increasing demands of application (e.g., to the large sample theory of statistics) indicate that the present generality is useful. It is rather fortunate that in our context precision and generality go hand in hand.

Apart from some material that most students in probability and statistics encounter during the first year of their graduate studies, this book is essentially self-contained. It is unavoidable that lengthy computations frequently appear in the text. We hope that in addition to making it easier for someone to check the veracity of a particular result of interest, the detailed computations will also be helpful in estimations of constants that appear in various error bounds in the text. To facilitate comprehension each chapter begins with a brief indication of the nature of the problem treated and its solution. Notes at the end of each chapter provide some history and references and, occasionally, additional facts. There is also an Appendix devoted partly to some elementary notions in probability and partly to some auxiliary results used in the book.

We have not discussed many topics closely related to the subject matter (not to mention applications). Some of these topics are "large deviation," extension of the results of this monograph to the dependence case, and rates of convergence for the invariance principle. It would take another book of comparable size to cover these topics adequately.

We take this opportunity to thank Professors Raghu Raj Bahadur and Patrick Billingsley for encouraging us to write this book and giving us advice. We owe a special debt of gratitude to Professor Billingsley for his many critical remarks, suggestions, and other help. We thank Professor John L. Denny for graciously reviewing the manuscript and pointing out a number of errors. We gratefully acknowledge partial support from the National Science Foundation (Grant. No. MPS 75-07549). Miss Kanda Kunze and Mrs. Sarah Oordt, who did an excellent job of typing the manuscript, have our sincere appreciation.

R. N. BHATTACHARYA
R. RANGA RAO

Tucson, Arizona
Urbana, Illinois
August 1975

List of Symbols

$A \setminus B$	set of all elements of A not in B: (1.4)
$A+y$	$\{x+y : x \in A\}$: (5.5)
A^{ϵ}	set of all points at distances less than ϵ from A: (1.17)
$A^{-\epsilon}$	set of all x such that the open ball of radius ϵ centered at x is contained in A: (2.38)
$\mathcal{A}$	a generic class of Borel sets
$\mathcal{A}_{\alpha}^{*}(d:\mu)$, $\mathcal{A}_{\alpha}(d:\Phi_{0,V})$	special classes of Borel Sets: (17.3), (17.52)
a_n	(14.64)
α, β	usually nonnegative vectors with integral coordinates; sometimes positive numbers
$\lvert\alpha\rvert$	sum of coordinates of a nonnegative integral vector
B	a generic Borel set
B, B_n	positive square roots of the inverses of matrices V, V_n: (9.7), (19.28)
$B(x:\epsilon)$	open ball of radius ϵ centered at x: (1.10)
$\mathcal{B}^{k}$	Borel sigma-field of R^{k}
$\mathrm{Cl}(A)$	closure of A
$\mathcal{C}$	class of all convex Borel subsets of R^{k}
$c(B)$	convex hull of B: Section 3

$\operatorname{cov}(\mathbf{X},\mathbf{Y})$	covariance between random variables $\mathbf{X}$, $\mathbf{Y}$: (A.1.5)
$\operatorname{Cov}(\mathbf{X})$	matrix of covariances between coordinates of a random vector $\mathbf{X}$: Appendix A.1
D	average covariance matrix of centered truncated random vectors $\mathbf{Z}_1,\dots,\mathbf{Z}_n$: (14.5)
D^{α}	αth derivative: (4.3)
$d(0,\partial A)$	euclidean distance between the origin and ∂A: Section 17
$d_0(G_1,G_2)$	(17.50)
d_p	Prokhorov distance: (1.16)
d_{BL}	bounded Lipschitzian distance: (2.51)
$d(P,\Phi)$	(12.47)
$\operatorname{Det} V$, $\operatorname{Det} D$	determinant of a matrix V or D
$\det L$	absolute value of the determinant of the matrix of basis vectors of a lattice L: (21.20)
$\Delta(A,B)$	Hausdorff distance between sets A and B: (2.62)
$\Delta_{n,j,s}$, $\overline{\Delta}_{n,s}$	(14.4)
$\overline{\Delta}_{n,s}(\epsilon)$	(14.105), (14.106)
$\Delta^*_{n,s}$	(15.7)
$\tilde{\Delta}^*_{n,s}$	(17.55)
$\delta^*_{n,s}$	(18.4)
∂A	topological boundary of A: (1.15)
$E\mathbf{X}$, $E(\mathbf{X})$	expectation or mean of a random variable or random vector $\mathbf{X}$: (A.1.2), (A.1.3)
ϵ, ε	generic small numbers
$\in$	symbol for "belongs to"
$\hat{f}$	Fourier transform of a function f: (4.5)
$\tilde{f}$	(4.4)
$f_y(x)$	$f(x+y)$: (11.5)
$f*g$	convolution of functions f and g: (4.9)
f^{*n}	n-fold convolution of a function f: (4.11)

$\mathcal{F}$	a generic class of Borel-measurable functions
$\mathcal{F}^*$	fundamental domain for the dual lattice L^*: (21.22)
Φ	normal distribution on R^k with zero mean and identity covariance matrix
ϕ	density of Φ
$\Phi_{m,V}$	normal distribution with mean m and covariance V
$\phi_{m,V}$	density of $\Phi_{m,V}$: (6.31)
Φ_{r_0}	(15.5), (18.10)
$G_{a,m}$, $g_{a,m}$	a special probability measure and its density: (10.7)
g_T	(16.7)
$\gamma(f:\epsilon)$, $\gamma^*(f:\epsilon)$	(11.8), (11.18)
η_r, $\bar{\eta}_{r,n}$	(9.8), (19.32)
I	$k\times k$ identity matrix
I_A	indicator function of the set A
Int(A)	interior of A
K_ϵ	a smooth kernel probability measure assigning either all or more than half its mass to the sphere $B(x:\epsilon)$: (11.6), (11.16), (15.26)
χ_ν	νth cumulant, average of νth cumulants of $\mathbf{X}_1,\ldots,\mathbf{X}_n$: (6.9), (9.6), (14.1)
χ'_ν	average of νth cumulants of centered truncated random vectors $\mathbf{Z}_1,\ldots,\mathbf{Z}_n$: (14.3)
$\chi_{\nu,j}$, $\bar{\chi}_{\nu,n}$	νth cumulant of $\mathbf{X}_j$, $1\leqslant j\leqslant n$, and their average: (9.6), Sections 19, 20
$\chi_s(z)$	(6.16)
L	a lattice: Section 21
L^*	lattice of periods of $\|f\|$, f being the characteristic function of a lattice random vector: (21.9), (21.19)
$L(c,d)$	a Lipschitzian class of functions: (2.50)
$l_{s,n}$	Liapounov coefficient: (8.10)
λ, Λ	smallest and largest eigenvalues of an average covariance matrix V: Section 16

λ_k	Lebesgue measure on R^k
$\Lambda_{r,n}(F)$	(23.8)
$M_r(f)$, $M_0(f)$	(15.4)
$M_f(x:\epsilon)$, $m_f(x:\epsilon)$	supremum and infimum of f in $B(x:\epsilon)$: (11.2)
$\mathfrak{M}$	set of all finite signed measures on a metric space
μ^+, μ^-, $\|\mu\|$	positive, negative, and total variations of a finite signed measure μ: (1.1)
$\|\|\mu\|\|$	variation norm of a signed measure μ: (1.5)
$\hat{\mu}$	Fourier–Stieltjes transform of μ: (5.2)
$\mu * \nu$	convolution of two finite signed measures μ, ν: (5.4)
μ^{*n}	n-fold convolution of μ: (5.6)
$\mu \circ T^{-1}$	signed measure induced by the map T: (5.7)
μ_α	αth moment, average of αth moments of $\mathbf{X}_1,\dots,\mathbf{X}_n$: (6.1), (14.1)
μ'_α	average of αth moments of centered truncated random vectors $\mathbf{Z}_1,\dots,\mathbf{Z}_n$: (14.3)
$\mu_r(t)$, $\beta_s(t)$	(8.4)
$\nu!$	$\nu_1!\nu_2!\dots\nu_k!$ where $\nu=(\nu_1,\dots,\nu_k)$ is a non-negative integral vector
ν_r, ν_0	special signed measures: (15.5)
P	a probability measure, a polyhedron
$\mathfrak{P}$	set of all probability measures on a metric space
$\hat{P}$	characteristic function of a probability measure P: (5.2)
$\tilde{P}_s(z:\{\chi_\nu\})$	a special polynomial in z: (7.3)
$P_r(-\phi_{0,V}:\{\chi_\nu\})$	a polynomial multiple of $\phi_{0,V}$: (7.11)
$P_r(-\Phi_{0,V}:\{\chi_\nu\})$	signed measure whose density is $P_r(-\phi_{0,V}:\{\chi_\nu\})$
P_a	a special polyhedron: (3.19)
$p_n(y_{\alpha,n})$	point masses of normalized lattice random vectors $\mathbf{X}_1,\dots,\mathbf{X}_n$: (22.3)
$p'_n(y'_{\alpha,n})$	point masses of normalized truncated lattice random vectors: (22.3)

Q_n	distribution of $n^{-1/2}(\mathbf{X}_1+\cdots+\mathbf{X}_n)$, where $\mathbf{X}_1,\ldots,\mathbf{X}_n$ are independent random vectors having zero means and average covariance matrix V (or I)
Q_n''	distribution of $n^{-1/2}(\mathbf{Y}_1+\cdots+\mathbf{Y}_n)$, where $\mathbf{Y}_j$'s are truncations of $\mathbf{X}_j$'s: (14.2)
Q_n'	distribution of $n^{-1/2}(\mathbf{Z}_1+\cdots+\mathbf{Z}_n)$, where $\mathbf{Z}_j=\mathbf{Y}_j-E\mathbf{Y}_j$: (14.2)
$q_{n,m}$, $q'_{n,m}$	local expansions of point masses of Q_n, Q_n' in the lattice case: (22.3), (22.38), (23.2)
$\rho(x,A)$	distance between a point x and a set A: (1.18)
ρ_s	sth absolute moment, average of sth absolute moments of $\mathbf{X}_1,\ldots,\mathbf{X}_n$: (6.2), (9.6), (14.1)
ρ_s'	average of sth absolute moments of centered truncated tandom vectors $\mathbf{Z}_1,\ldots,\mathbf{Z}_n$: (14.3)
$\rho_{s,j}$, $\bar{\rho}_{s,n}$	sth absolute moment of $\mathbf{X}_j$, $1\leqslant j\leqslant n$, and their average: (14.1), (17.55)
S_j, S_α	special periodic functions: (A.4.2), (A.4.14)
$\mathcal{S}$	Schwartz space: (A.4.13)
σ_{k-1}	surface area measure on the unit sphere of R^k: Section 3
$\\|T\\|$	norm of a matrix T: (14.17)
τ_r	(16.6)
$\tau(f:2\epsilon)$, $\tau^*(f:2\epsilon)$	(11.8), (11.18)
V	average of covariance matrices of random vectors $\mathbf{X}_1,\ldots,\mathbf{X}_n$: (9.6), (14.5)
$\omega_f(A)$	oscillation of f on A: (2.7), (11.1)
$\omega_f(x:\epsilon)$	oscillation of f on $B(x:\epsilon)$: (2.7), (11.3)
$\bar{\omega}_f(\epsilon:\mu)$	average modulus of oscillation of f with respect to a measure μ: (11.23)
$\omega_f^*(\epsilon:\mu)$	$\sup_y \bar{\omega}_{f_y}(\epsilon:\mu)$: (11.24)
$\|x\|$	$\|x_1\|+\cdots+\|x_k\|$, where $x=(x_1,\ldots,x_k)$: (4.8)

$y_{\alpha,n}$, $y'_{\alpha,n}$	(22.3)
$\mathbf{Z}^+$	set of all nonnegative integers
$(\mathbf{Z}^+)^k$	set of all k-tuples of nonnegative integers
$\|\cdot\|$, $\langle,\rangle$	euclidean norm and inner product
$\|\cdot\|_p$	L^p-norm
$\mathbf{Z}$	set of all integers

Contents

CHAPTER 1

Weak Convergence of Probability Measures and Uniformity Classes

Let Q be a probability measure on a separable metric space S every open ball of which is connected (e.g., $S=R^k$). In the present chapter we characterize classes $\mathcal{F}$ of bounded Borel-measurable functions such that

$$\sup_{f\in\mathcal{F}}\left|\int_S f\,dQ_n-\int_S f\,dQ\right|\to 0 \qquad (n\to\infty), \tag{1}$$

for every sequence $\{Q_n : n \geqslant 1\}$ of probability measures converging weakly to Q. Such a class is called a Q-uniformity class. It turns out that $\mathcal{F}$ is a Q uniformity class if and only if

$$\sup_{f\in\mathcal{F}}\omega_f(S)<\infty, \qquad \lim_{\epsilon\downarrow 0}\left[\sup_{f\in\mathcal{F}}\int_S \omega_f(x:\epsilon)\,Q(dx)\right]=0, \tag{2}$$

where $\omega_f(S)$ is the (total) oscillation of f on S, and $\omega_f(x:\epsilon)$ its oscillation on the open ball of radius ϵ centered at x. This suggests that the appropriate characteristics of f on which the rate of convergence $\int f\,dQ_n \to \int f\,dQ$ depends are (i) $\omega_f(S)$ and (ii) the average oscillation function $\epsilon\to\int\omega_f(x:\epsilon)\,Q(dx)$. Specialized to indicator functions of Borel sets A, this says that the rate of convergence $Q_n(A)\to Q(A)$ depends on the function $\epsilon\to Q((\partial A)^\epsilon)$, where ∂A is the boundary of A and $(\partial A)^\epsilon$ is the set of all points whose distances from ∂A are less than ϵ. We have pursued this line of thinking in Chapters 3 and 4 to obtain appropriate rates of convergence for the central limit theorem.

Section 1 contains a brief review of those aspects of weak convergence theory that are relevant for proving results on characterization of uniformity classes in Section 2. These two sections are not used in the sequel (except as motivation). In Section 3 we obtain estimates such as

$$\sup_{C \in \mathcal{C}} \Phi((\partial C)^{\epsilon}) \leqslant d(k)\epsilon \qquad (\epsilon > 0), \tag{3}$$

where Φ is the standard normal distribution on R^k, $\mathcal{C}$ is the class of all (Borel-measurable) convex subsets of R^k, and $d(k)$ is a positive number depending only on k. We have several occasions in Chapters 3 and 4 to use these estimates for deriving rates of convergence $Q_n(C) \to \Phi(C)$, $C \in \mathcal{C}$, where Q_n is the distribution of the normalized sum of n independent random vectors.

1. WEAK CONVERGENCE

In this section we briefly review some standard results in the theory of weak convergence of probability measures.

Throughout this section S denotes a *metric space* with a metric ρ. The *Borel sigma-field* $\mathcal{B}$ *of* S is the smallest sigma-field containing the clas all open subsets of S. We say μ is a *(signed) measure on* S if it is a (sig measure defined on $\mathcal{B}$. The class of all finite signed measures on denoted by $\mathfrak{M}$, and the subcla of $\mathfrak{M}$ comprising all probability m sures is denoted by $\mathcal{P}$. Given a finite signed measure μ on S, one defines three associated set functions μ^+, μ^-, $|\mu|$, called the *positive, negative*, and *total variations* of μ, respectively, by

$$\begin{aligned} \mu^+(B) &= \sup\{\mu(A): A \subset B,\ A \in \mathcal{B}\}, \\ \mu^-(B) &= -\inf\{\mu(A): A \subset B,\ A \in \mathcal{B}\}, \qquad (B \in \mathcal{B}) \\ |\mu| &= \mu^+ + \mu^-. \end{aligned} \tag{1.1}$$

The so-called *Jordan–Hahn decomposition*† asserts that μ^+ and μ^- (and, therefore $|\mu|$) are finite measures on S satisfying

$$\mu = \mu^+ - \mu^-. \tag{1.2}$$

For every finite signed measure μ on a *separable* metric space S, we define the *support of* μ as the smallest closed subset of S whose complement has $|\mu|$-measure zero: that is,

$$\text{support of } \mu = \cap\{F: F \text{ closed}, |\mu|(S \backslash F) = 0\}, \tag{1.3}$$

†See Halmos [1], pp. 121–123.

where for any two sets A, B we write

$$A \setminus B = \{x : x \in A, x \notin B\}. \tag{1.4}$$

Note that the separability of the metric space S ensures that the complement of the right side of (1.3) has zero $|\mu|$-measure.

The class $\mathfrak{M}$ of (set) functions on $\mathfrak{B}$ into R^1 is a real linear space with respect to pointwise addition and multiplication by real scalars. It is a Banach space when endowed with the *variation norm*

$$\|\mu\| = |\mu|(S) \qquad (\mu \in \mathfrak{M}). \tag{1.5}$$

Let $C(S)$ denote the *class of all complex-valued, bounded, continuous functions on* S. The *weak totopogy* on $\mathfrak{M}$ is the weakest topology (on $\mathfrak{M}$) that makes the maps

$$\mu \to \int f d\mu \qquad [f \in C(S)] \tag{1.6}$$

on $\mathfrak{M}$ into the complex field **C** continuous. The right side of (1.6) always stands for the *Lebesgue integral of* (a μ-integrable, complex-valued, Borel-measurable function) f *on* S. The *Lebesgue integral of* f *on a Borel set* B is denoted by

$$\int_B f d\mu. \tag{1.7}$$

When it becomes necessary to indicate the variable of integration, we also write

$$\int f(x)\, \mu(dx) \tag{1.8}$$

instead of $\int f d\mu$.

In this monograph we are particularly concerned with the relativized weak topology on the class $\mathfrak{P}$ of all probability measures on S. In this topology *convergence of a sequence* $\{Q_n\}$ *of probability measures to a probability measure* Q means

$$\lim_n \int f dQ_n = \int f dQ \tag{1.9}$$

for every f in $C(S)$. The following theorem gives several characterizations of weak convergence of a sequence of probability measures.