SPRINGER TEXTS IN STATISTICS

Measure Theory and Probability Theory

Krishna B. Athreya
Soumendra N. Lahiri

Springer

Springer Texts in Statistics

Advisors:
George Casella   Stephen Fienberg   Ingram Olkin

Springer Texts in Statistics

Alfred: Elements of Statistics for the Life and Social Sciences
Athreya and Lahiri: Measure Theory and Probability Theory
Berger: An Introduction to Probability and Stochastic Processes
Bilodeau and Brenner: Theory of Multivariate Statistics
Blom: Probability and Statistics: Theory and Applications
Brockwell and Davis: Introduction to Times Series and Forecasting, Second Edition
Carmona: Statistical Analysis of Financial Data in S-Plus
Chow and Teicher: Probability Theory: Independence, Interchangeability, Martingales, Third Edition
Christensen: Advanced Linear Modeling: Multivariate, Time Series, and Spatial Data—Nonparametric Regression and Response Surface Maximization, Second Edition
Christensen: Log-Linear Models and Logistic Regression, Second Edition
Christensen: Plane Answers to Complex Questions: The Theory of Linear Models, Third Edition
Creighton: A First Course in Probability Models and Statistical Inference
Davis: Statistical Methods for the Analysis of Repeated Measurements
Dean and Voss: Design and Analysis of Experiments
du Toit, Steyn, and Stumpf: Graphical Exploratory Data Analysis
Durrett: Essentials of Stochastic Processes
Edwards: Introduction to Graphical Modelling, Second Edition
Finkelstein and Levin: Statistics for Lawyers
Flury: A First Course in Multivariate Statistics
Ghosh, Delampady and Samanta: An Introduction to Bayesian Analysis: Theory and Methods
Gut: Probability: A Graduate Course
Heiberger and Holland: Statistical Analysis and Data Display: An Intermediate Course with Examples in S-PLUS, R, and SAS
Jobson: Applied Multivariate Data Analysis, Volume I: Regression and Experimental Design
Jobson: Applied Multivariate Data Analysis, Volume II: Categorical and Multivariate Methods
Kalbfleisch: Probability and Statistical Inference, Volume I: Probability, Second Edition
Kalbfleisch: Probability and Statistical Inference, Volume II: Statistical Inference, Second Edition
Karr: Probability
Keyfitz: Applied Mathematical Demography, Second Edition
Kiefer: Introduction to Statistical Inference
Kokoska and Nevison: Statistical Tables and Formulae
Kulkarni: Modeling, Analysis, Design, and Control of Stochastic Systems
Lange: Applied Probability
Lange: Optimization
Lehmann: Elements of Large-Sample Theory

(continued after index)

Krishna B. Athreya
Soumendra N. Lahiri

Measure Theory and Probability Theory

Springer

Krishna B. Athreya
Department of Mathematics and
Department of Statistics
Iowa State University
Ames, IA 50011
kba@iastate.edu

Soumendra N. Lahiri
Department of Statistics
Iowa State University
Ames, IA 50011
snlahiri@iastate.edu

Editorial Board

George Casella
Department of Statistics
University of Florida
Gainesville, FL 32611-8545
USA

Stephen Fienberg
Department of Statistics
Carnegie Mellon University
Pittsburgh, PA 15213-3890
USA

Ingram Olkin
Department of Statistics
Stanford University
Stanford, CA 94305
USA

Library of Congress Control Number: 2006922767

ISBN-10: 0-387-32903-X e-ISBN: 0-387-35434-4
ISBN-13: 978-0387-32903-1

Printed on acid-free paper.

©2006 Springer Science+Business Media, LLC
All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excepts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.
The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed in the United States of America. (MVY)

9 8 7 6 5 4 3 2 1

springer.com

Dedicated to our wives
Krishna S. Athreya and Pubali Banerjee
and
to the memory of
Uma Mani Athreya and Narayani Ammal

Preface

This book arose out of two graduate courses that the authors have taught during the past several years; the first one being on measure theory followed by the second one on advanced probability theory.

The traditional approach to a first course in measure theory, such as in Royden (1988), is to teach the Lebesgue measure on the real line, then the differentation theorems of Lebesgue, L^p-spaces on $\mathbb{R}$, and do general measure at the end of the course with one main application to the construction of product measures. This approach does have the pedagogic advantage of seeing one concrete case first before going to the general one. But this also has the disadvantage in making many students' perspective on measure theory somewhat narrow. It leads them to think only in terms of the Lebesgue measure on the real line and to believe that measure theory is intimately tied to the topology of the real line. As students of statistics, probability, physics, engineering, economics, and biology know very well, there are mass distributions that are typically nonuniform, and hence it is useful to gain a general perspective.

This book attempts to provide that general perspective right from the beginning. The opening chapter gives an informal introduction to measure and integration theory. It shows that the notions of σ-algebra of sets and countable additivity of a set function are dictated by certain very natural approximation procedures from practical applications and that they are not just some abstract ideas. Next, the general extension theorem of Carathedory is presented in Chapter 1. As immediate examples, the construction of the large class of Lebesgue-Stieltjes measures on the real line and Euclidean spaces is discussed, as are measures on finite and countable

spaces. Concrete examples such as the classical Lebesgue measure and various probability distributions on the real line are provided. This is further developed in Chapter 6 leading to the construction of measures on sequence spaces (i.e., sequences of random variables) via Kolmogorov's consistency theorem.

After providing a fairly comprehensive treatment of measure and integration theory in the first part (Introduction and Chapters 1–5), the focus moves onto probability theory in the second part (Chapters 6–13). The feature that distinguishes probability theory from measure theory, namely, the notion of independence and dependence of random variables (i.e., measureable functions) is carefully developed first. Then the laws of large numbers are taken up. This is followed by convergence in distribution and the central limit theorems. Next the notion of conditional expectation and probability is developed, followed by discrete parameter martingales. Although the development of these topics is based on a rigorous measure theoretic foundation, the heuristic and intuitive backgrounds of the results are emphasized throughout. Along the way, some applications of the results from probability theory to proving classical results in analysis are given. These include, for example, the density of normal numbers on (0,1) and the Wierstrass approximation theorem. These are intended to emphasize the benefits of studying both areas in a rigorous and combined fashion. The approach to conditional expectation is via the mean square approximation of the "unknown" given the "known" and then a careful approximation for the L^1-case. This is a natural and intuitive approach and is preferred over the "black box" approach based on the Radon-Nikodym theorem.

The final part of the book provides a basic outline of a number of special topics. These include Markov chains including Markov chain Monte Carlo (MCMC), Poisson processes, Brownian motion, bootstrap theory, mixing processes, and branching processes. The first two parts can be used for a two-semester sequence, and the last part could serve as a starting point for a seminar course on special topics.

This book presents the basic material on measure and integration theory and probability theory in a self-contained and step-by-step manner. It is hoped that students will find it accessible, informative, and useful and also that they will be motivated to master the details by carefully working out the text material as well as the large number of exercises. The authors hope that the presentation here is found to be clear and comprehensive without being intimidating.

Here is a quick summary of the various chapters of the book. After giving an informal introduction to the ideas of measure and integration theory, the construction of measures starting with set functions on a small class of sets is taken up in Chapter 1 where the Caratheodory extension theorem is proved and then applied to construct Lebesgue-Stieltjes measures. Integration theory is taken up in Chapter 2 where all the basic convergence theorems including the MCT, Fatou, DCT, BCT, Egorov's, and Scheffe's are

proved. Included here are also the notion of uniform integrability and the classical approximation theorem of Lusin and its use in L^p-approximation by smooth functions. The third chapter presents basic inequalities for L^p-spaces, the Riesz-Fischer theorem, and elementary theory of Banach and Hilbert spaces. Chapter 4 deals with Radon-Nikodym theory via the Riesz representation on L^2-spaces and its application to differentiation theorems on the real line as well as to signed measures. Chapter 5 deals with product measures and the Fubini-Tonelli theorems. Two constructions of the product measure are presented: one using the extension theorem and another via iterated integrals. This is followed by a discussion on convolutions, Laplace transforms, Fourier series, and Fourier transforms. Kolmogorov's consistency theorem for the construction of stochastic processes is taken up in Chapter 6 followed by the notion of independence in Chapter 7. The laws of large numbers are presented in a unified manner in Chapter 8 where the classical Kolmogorov's strong law as well as Etemadi's strong law are presented followed by Marcinkiewicz-Zygmund laws. There are also sections on renewal theory and ergodic theorems. The notion of weak convergence of probability measures on $\mathbb{R}$ is taken up in Chapter 9, and Chapter 10 introduces characteristic functions (Fourier transform of probability measures), the inversion formula, and the Levy-Cramer continuity theorem. Chapter 11 is devoted to the central limit theorem and its extensions to stable and infinitely divisible laws. Chapter 12 discusses conditional expectation and probability where an L^2-approach followed by an approximation to L^1 is presented. Discrete time martingales are introduced in Chapter 13 where the basic inequalities as well as convergence results are developed. Some applications to random walks are indicated as well. Chapter 14 discusses discrete time Markov chains with a discrete state space first. This is followed by discrete time Markov chains with general state spaces where the regeneration approach for Harris chains is carefully explained and is used to derive the basic limit theorems via the iid cycles approach. There are also discussions of Feller Markov chains on Polish spaces and Markov chain Monte Carlo methods. An elementary treatment of Brownian motion is presented in Chapter 15 along with a treatment of continuous time jump Markov chains. Chapters 16–18 provide brief outlines respectively of the bootstrap theory, mixing processes, and branching processes. There is an Appendix that reviews basic material on elementary set theory, real and complex numbers, and metric spaces.

Here are some suggestions on how to use the book.

1. For a one-semester course on real analysis (i.e., measure end integration theory), material up to Chapter 5 and the Appendix should provide adequate coverage with Chapter 6 being optional.

2. A one-semester course on advanced probability theory for those with the necessary measure theory background could be based on Chapters 6–13 with a selection of topics from Chapters 14–18.

3. A one-semester course on combined treatment of measure theory and probability theory could be built around Chapters 1, 2, Sections 3.1–3.2 of Chapter 3, all of Chapter 4 (Section 4.2 optional), Sections 5.1 and 5.2 of Chapter 5, Chapters 6, 7, and Sections 8.1, 8.2, 8.3 (Sections 8.5 and 8.6 optional) of Chapter 8. Such a course could be followed by another that includes some coverage of Chapters 9–12 before moving on to other areas such as mathematical statistics or martingales and financial mathematics. This will be particularly useful for graduate programs in statistics.

4. A one-semester course on an introduction to stochastic processes or a seminar on special topics could be based on Chapters 14–18.

A word on the numbering system used in the book. Statements of results (i.e., Theorems, Corollaries, Lemmas, and Propositions) are numbered consecutively within each section, in the format $a.b.c$, where a is the chapter number, b is the section number, and c is the counter. Definitions, Examples, and Remarks are numbered individually within each section, also of the form $a.b.c$, as above. Sections are referred to as $a.b$ where a is the chapter number and b is the section number. Equation numbers appear on the right, in the form $(b.c)$, where b is the section number and c is the equation number. Equations in a given chapter a are referred to as $(b.c)$ within the chapter but as $(a.b.c)$ outside chapter a. Problems are listed at the end of each chapter in the form $a.c$, where a is the chapter number and c is the problem number.

In the writing of this book, material from existing books such as Apostol (1974), Billingsley (1995), Chow and Teicher (2001), Chung (1974), Durrett (2004), Royden (1988), and Rudin (1976, 1987) has been freely used. The authors owe a great debt to these books. The authors have used this material for courses taught over several years and have benefited greatly from suggestions for improvement from students and colleagues at Iowa State University, Cornell University, the Indian Institute of Science, and the Indian Statistical Institute. We are grateful to them.

Our special thanks go to Dean Issacson, Ken Koehler, and Justin Peters at Iowa State University for their administrative support of this long project. Krishna Athreya would also like to thank Cornell University for its support.

We are most indebted to Sharon Shepard who typed and retyped several times this book, patiently putting up with our never-ending “final” versions. Without her patient and generous help, this book could not have been written. We are also grateful to Denise Riker who typed portions of an earlier version of this book.

John Kimmel of Springer got the book reviewed at various stages. The referee reports were very helpful and encouraging. Our grateful thanks to both John Kimmel and the referees.

We have tried hard to make this book free of mathematical and typographical errors and misleading or ambiguous statements, but we are aware that there will still be many such remaining that we have not caught. We will be most grateful to receive such corrections and suggestions for improvement. They can be e-mailed to us at *kba@iastate.edu* or *snlahiri@iastate.edu*.

On a personal note, we would like to thank our families for their patience and support. Krishna Athreya would like to record his profound gratitude to his maternal granduncle, the late Shri K. Venkatarama Iyer, who opened the door to mathematical learning for him at a crucial stage in high school, to the late Professor D. Basu of the Indian Statistical Institute who taught him to think probabilistically, and to Professor Samuel Karlin of Stanford University for initiating him into research in mathematics.

K. B. Athreya
S. N. Lahiri
May 12, 2006

Contents

Measures and Integration: An Informal Introduction

For many students who are learning measure and integration theory for the first time, the notions of a σ-algebra of subsets of a set Ω, countable additivity of a set function λ, measurability of a function, the definition of an integral, and the interchange of limits and integration are not easy to understand and often seem not so intuitive. The goals of this informal introduction to this subject are (1) to show that the notions of σ-algebra and countable additivity are logical consequences of certain natural approximation procedures; (2) the dividends for the assumption of these two properties are great, and they lead to a nice and natural theory that is also very powerful for the handling of limits. Of course, as the saying goes, the devil is in the details. After this informal introduction, the necessary details are given in the next few sections. It is hoped that after this heuristic explanation of the subject, the motivation for and the process of mastering the details on the part of the students will be forthcoming.

What is Measure Theory?

A simple answer is that it is a theory about the distribution of mass over a set $\mathbb{S}$. If the mass is uniformly distributed and $\mathbb{S}$ is an Euclidean space $\mathbb{R}^k$, it is the theory of Lebesgue measure on $\mathbb{R}^k$ (i.e., length in $\mathbb{R}$, area in $\mathbb{R}^2$, volume in $\mathbb{R}^3$, etc.). Probability theory is concerned with the case when $\mathbb{S}$ is the sample space of a random experiment and the total mass is one. Consider the following example.

Imagine an open field $\mathbb{S}$ and a snowy night. At daybreak one goes to the field to measure the amount of snow in as many of the subsets of $\mathbb{S}$ as

possible. Suppose now that one has the tools to measure the snow exactly on a class of subsets, such as triangles, rectangles, circular shapes, elliptic shapes, etc., no matter how small. It is natural to try to approximate oddly-shaped regions by combinations of these "standard shapes," and then use a limiting process to obtain a measure for the oddly-shaped regions and reach some limit for such sets. Let $\mathcal{B}$ denote the class of subsets of $\mathbb{S}$ whose measure is obtained this way and let $\lambda(B)$ denote the amount of snow in each $B \in \mathcal{B}$. Call $\mathcal{B}$ the class of all (snow) measurable sets and $\lambda(B)$ the measure (of snow) on B for each $B \in \mathcal{B}$. It is reasonable to expect that the following properties of $\mathcal{B}$ and $\lambda(\cdot)$ hold:

Properties of $\mathcal{B}$

(i) $A \in \mathcal{B} \Rightarrow A^c \in \mathcal{B}$ (i.e., if one can measure the amount of snow on A and knows the total amount on $\mathbb{S}$, then one knows the amount of snow on A^c).

(ii) $A_1, A_2 \in \mathcal{B} \Rightarrow A_1 \cup A_2 \in \mathcal{B}$ (i.e., if one can measure the amount of snow on A_1 and A_2, then one can do the same for $A_1 \cup A_2$).

(iii) If $\{A_n : n \geq 1\} \subset \mathcal{B}$, and $A_n \subset A_{n+1}$ for all $n \geq 1$, then $\lim_{n\to\infty} A_n \equiv \bigcup_{n=1}^{\infty} A_n \in \mathcal{B}$ (i.e., if one can measure the amount of snow on A_n for each $n \geq 1$ on an *increasing sequence* of sets, then one can do so on the limit of A_n).

(iv) $\mathcal{C} \subset \mathcal{B}$ where $\mathcal{C}$ is the class of nice sets such as triangles, squares, etc., that one started with.

Properties of $\lambda(\cdot)$

(i) $\lambda(A) \geq 0$ for $A \in \mathcal{B}$ (i.e., the amount of snow on any set is nonnegative!)

(ii) If $A_1, A_2 \in B, A_1 \cap A_2 = \emptyset, \lambda(A_1 \cup A_2) = \lambda(A_1) + \lambda(A_2)$ (i.e., the amounts of snow on two disjoint sets simply add up! This property of λ is referred to as *finite additivity*).

(iii) If $\{A_n : n \geq 1\} \subset \mathcal{B}$, are such that $A_n \subset A_{n+1}$ for all n, then $\lambda(\lim_{n\to\infty} A_n) = \lambda(\bigcup_{n=1}^{\infty} A_n) = \lim_{n\to\infty} \lambda(A_n)$ (i.e., if we can approximate a set A by an increase sequence of sets $\{A_n\}_{n\geq 1}$ from $\mathcal{B}$, then $\lambda(A) = \lim_{n\to\infty} \lambda(A_n)$. This property of λ is referred to as *monotone continuity from below*, or *m.c.f.b.* in short).

This last assumption (iii) is what guarantees that different approximations lead to consistent limits. Thus, if there are two increasing sequences $\{A_n'\}_{n\geq 1}$ and $\{A_n''\}_{n\geq 1}$ having the same limit A but $\{\lambda(A_n')\}_{n\geq 1}$ and $\{\lambda(A_n'')\}_{n\geq 1}$ have different limits, then the approximating procedures are not consistent.

It turns out that the above set of reasonable and natural assumptions lead to a very rich and powerful theory that is widely applicable.

A triplet $(\mathbb{S}, \mathcal{B}, \lambda)$ that satisfies the above two sets of assumptions is called a *measure space*. The assumptions on $\mathcal{B}$ and λ are *equivalent* to the following:

On $\mathcal{B}$

$\mathcal{B}(\mathrm{i})'$: $\emptyset$, the empty set, lies in $\mathcal{B}$

$\mathcal{B}(\mathrm{ii})'$: $A \in \mathcal{B} \Rightarrow A^c \in \mathcal{B}$ (same as (i) before)

$\mathcal{B}(\mathrm{iii})'$: $A_1, A_2, \ldots \in \mathcal{B} \Rightarrow \cup_i A_i \in \mathcal{B}$ (combines (ii) and (iii) above) (*Closure under countable unions*).

On λ

$\lambda(\mathrm{i})'$: $\lambda(\cdot) \geq 0$ (same as (i) before) and $\lambda(\emptyset) = 0$.

$\lambda(\mathrm{ii})'$: $\lambda(\cup_{n\geq 1} A_n) = \sum_{n=1}^{\infty} \lambda(A_n)$ if $\{A_n\}_{n\geq 1} \subset \mathcal{B}$ are *pairwise disjoint*, i.e., $A_i \cap A_j = \emptyset$ for $i \neq j$ (*Countable additivity*).

Any collection $\mathcal{B}$ of subsets of $\mathbb{S}$ that satisfies $\mathcal{B}(\mathrm{i})'$, $\mathcal{B}(\mathrm{ii})'$, $\mathcal{B}(\mathrm{iii})'$ above is called a *σ-algebra*. Any set function λ on a σ-algebra $\mathcal{B}$ that satisfies $\lambda(\mathrm{i})'$ and $\lambda(\mathrm{ii})'$ above is called a *measure*. Thus, a *measure space* is a triplet $(\mathbb{S}, \mathcal{B}, \lambda)$ where $\mathbb{S}$ is a nonempty set, $\mathcal{B}$ is a *σ-algebra* of subsets of $\mathbb{S}$ and λ is a *measure* on $\mathcal{B}$. Notice that the σ-algebra structure on $\mathcal{B}$ and the countable additivity of λ are necessary consequences of the very natural assumptions (i), (ii), and (iii) on $\mathcal{B}$ and λ defined at the beginning.

It is not often the case that one is given $\mathcal{B}$ and λ explicitly. Typically, one starts with a small collection $\mathcal{C}$ of subsets of $\mathbb{S}$ that have properties resembling intervals or rectangles and a set function λ on $\mathcal{C}$. Then, $\mathcal{B}$ is the smallest σ-algebra containing $\mathcal{C}$ obtained from $\mathcal{C}$ by various operations such as countable unions, intersections, and their limits. The key properties on $\mathcal{C}$ that one needs are:

(i) $A, B \in \mathcal{C} \Rightarrow A \cap B \in \mathcal{C}$ (e.g., intersection of intervals is an interval).

(ii) $A \in \mathcal{C} \Rightarrow A^c$ is a finite union of sets from $\mathcal{C}$ (e.g., the complement of an interval is the union of two intervals or an interval itself).

A collection $\mathcal{C}$ satisfying (i) and (ii) is called a *semialgebra*. The function λ on $\mathcal{B}$ is an extension of λ on $\mathcal{C}$. For this extension to be a measure on $\mathcal{B}$, the conditions needed are

(i) $\lambda(A) \geq 0 \quad \text{for all} \quad A \in \mathcal{C}$

(ii) If $A_1, A_2, \ldots \in \mathcal{C}$ are pairwise disjoint and $A = \bigcup_{n\geq 1} A_n \in \mathcal{C}$, then $\lambda(A) = \sum_{n=1}^{\infty} \lambda(A_n)$.