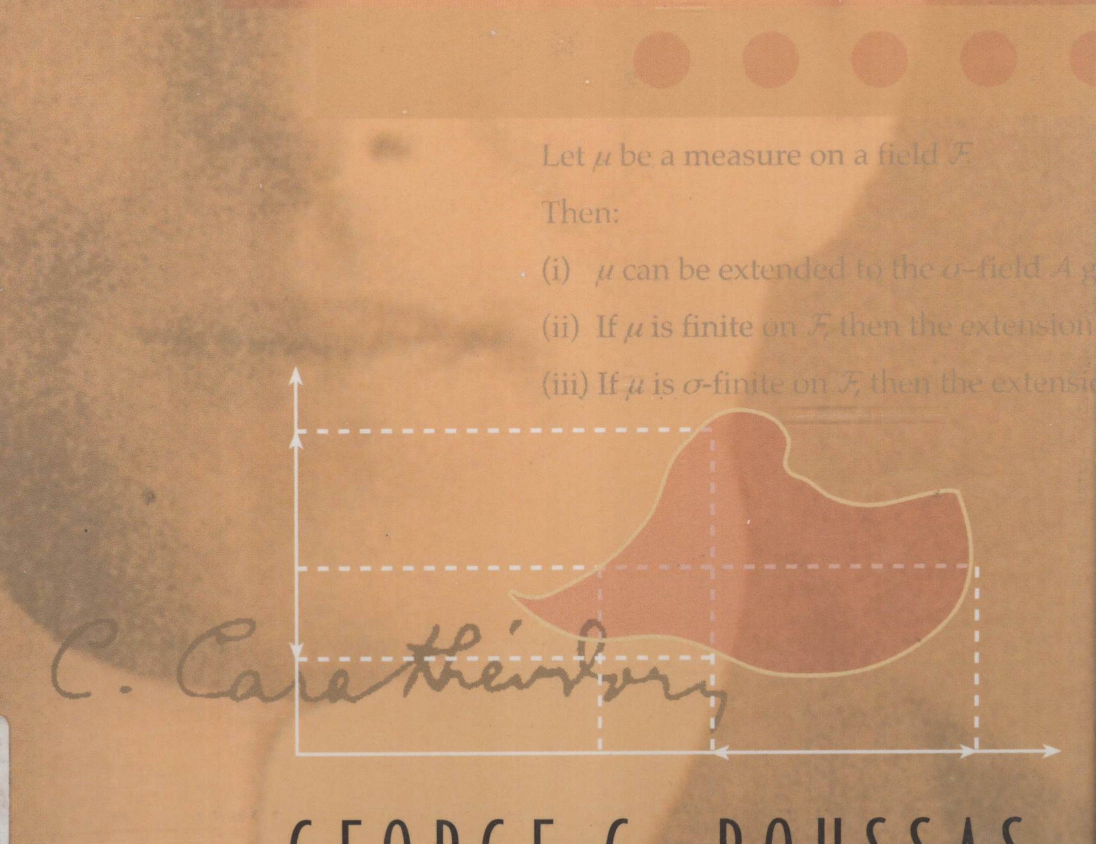
ELSEVIER
ACADEMIC PRESS

An Introduction to Measure-Theoretic Probability



GEORGE G. ROUSSAS

AN INTRODUCTION TO MEASURE–THEORETIC PROBABILITY

George G. Roussas
University of California, Davis

ELSEVIER
ACADEMIC
PRESS

Amsterdam • Boston • Heidelberg • London • New York • Oxford
Paris • San Diego • San Francisco • Singapore • Sydney • Tokyo

Senior Editor	Barbara Holland
Project Manager	Anne B. McGee
Associate Editor	Tom Singer
Marketing Manager	Linda Beattie
Cover Design	Dick Hannus, Hannus Design Assoc.
Composition	Kolam Information Services
Interior Printer	The Maple-Vail Book Manufacturing Group

Elsevier Academic Press
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA
525 B Street, Suite 1900, San Diego, California 92101-4495, USA
84 Theobald's Road, London WC1X 8RR, UK

This book is printed on acid-free paper. ∞

Library of Congress Cataloging-in-Publication Data
Roussas, George G.
An introduction to measure-theoretic probability/George G. Roussas.–1st ed.
p.cm.
Includes bibliographical references and index.
ISBN 0-12-599022-7 (alk. paper)
1. Probabilities. 2. Measure theory. I. Title.
QA273.R864 2004
519.2–dc22
2004017135

British Library Cataloging in Publication Data
A catalogue record for this book is available from the British Library

For all information on all Elsevier Academic Press publications
visit our Web site at www.books.elsevier.com

Printed in the United States of America
04 05 06 07 08 09 9 8 7 6 5 4 3 2 1

This book is devoted to the memory of Edward W. Barankin, the probabilist, mathematical statistician, classical scholar, and philosopher, for his role in stimulating my interest in probability with emphasis on detail and rigor.

Also, to my dearest sisters, who provided material means in my needy student years, and unrelenting moral support throughout my career.

PREFACE

This book in measure–theoretic probability has resulted from classroom lecture notes that this author has developed over a number of years, by teaching such a course at both the University of Wisconsin, Madison, and the University of California, Davis. The audience consisted of graduate students primarily in statistics and mathematics. There were always some students from engineering departments, and a handful of students from other disciplines such as economics.

The book is not a comprehensive treatment of probability, nor is it meant to be one. Rather, it is an excursion in measure–theoretic probability with the objective of introducing the student to the basic tools in measure theory and probability as they are commonly used in statistics, mathematics, and other areas employing this kind of moderately advanced mathematical machinery. Furthermore, it must be emphasized that the approach adopted here is entirely classical. Thus, characteristic functions are a tool employed extensively; no use of martingale or empirical process techniques is made anywhere.

The book does not commence with probabilistic concepts, and there is a good reason for it. As many of those engaged in teaching advanced probability and statistical theory know, very few students, if any, have been exposed to a measure theory course prior to attempting a course in advanced probability. This has been invariably the experience of this author throughout the years. This very fact necessitates the study of the basic measure-theoretic concepts and results—in particular, the study of those concepts and results that apply immediately to probability, and also in the form and shape they are used in probability.

On the basis of such considerations, the framework of the material to be dealt with is therefore determined. It consists of a brief introduction to measure theory, and then the discussion of those probability results that constitute the backbone of the subject matter. There is minimal flexibility allowed, and that is exploited in the form of the final chapter of the book. From many interesting and important candidate topics, this author has chosen to present a brief discussion of some basic concepts and results of ergodic theory.

From the very outset, there is one point that must be abundantly clarified, and that is the fact that everything is discussed in great detail with all proofs included;

no room is allowed for summary unproven statements. This approach has at least two side benefits, as this author sees them. One is that students have at their disposal a comprehensive and detailed proof of what are often deep theorems. Second, the instructor may skip the reproduction of such proofs by assigning their study to students.

In the experience of this author, there are no topics in this book which can be omitted, except perhaps for the final chapter. With this in mind, the material can be taught in two quarters, and perhaps even in one semester with appropriate calibration of the rate of presentation, and the omission of proofs of judiciously selected theorems. With all details presented, one can also cover an entire year of instruction, perhaps with some supplementation.

Most chapters are supplied with examples, and all chapters are concluded with a varying number of exercises. An unusual feature here is that an *Answers Manual* of all exercises will be made available to those instructors who adopt the book as the textbook in their course. Furthermore, an overview of each one of the 15 chapters is included in an appendix to the main body of the book. It is believed that the reader will benefit significantly by reviewing the overview of a chapter before the material in the chapter itself is discussed.

The remainder of this preface is devoted to a brief presentation of the material discussed in the 15 chapters of the book, chapter by chapter.

Chapter 1 commences with the introduction of the important classes of sets in an abstract space, which are those of a field, a σ-field, including the Borel σ-field, and a monotone class. They are illustrated by concrete examples, and their relationships are studied. Product spaces are also introduced, and some basic results are established. The discussion proceeds with the introduction of the concept of measurable functions, and in particular of random vectors and random variables. Some related results are also presented. This chapter is concluded with a fundamental theorem, Theorem 17, which provides for pointwise approximation of any random variable by a sequence of so-called simple random variables.

Chapter 2 is devoted to the introduction of the concept of a measure, and the study of the most basic results associated with it. Although a field is the class over which a measure can be defined in an intuitively satisfying manner, it is a σ-field—the one generated by an underlying field—on which a measure must be defined. One way of carrying out the construction of a measure on a σ-field is to use as a tool the so-called outer measure. The concept of an outer measure is then introduced, and some of its properties are studied in the second section of the chapter. Thus, starting with a measure on a field, utilizing the associated outer measure and the powerful Carathéodory theorem, one ensures the definition of a measure over the σ-field generated by the underlying field. The chapter is concluded with a study of the relationship between a measure over the Borel σ-field in the real line and certain point functions. A measure always determines a class of point functions, which are nondecreasing and right-continuous. The important thing, however, is that each such point function uniquely determines a measure on the Borel σ-field.

In Chapter 3, sequences of random variables are considered, and two basic kinds of convergences are introduced. One of them is the almost everywhere convergence, and the other is convergence in measure. The former convergence is essentially the familiar pointwise convergence, whereas convergence in measure is a mode of convergence not occurring in a calculus course. A precise expression of the set of pointwise convergence is established, which is used for formulating necessary and sufficient conditions for almost everywhere convergence. Convergence in measure is weaker than almost everywhere convergence, and the latter implies the former for finite measures. *Almost everywhere convergence* and *mutual almost everywhere convergence* are equivalent, as is easily seen. Although the same is true when convergence in measure is involved, its justification is fairly complicated and also requires the introduction of the concept of *almost uniform convergence*. Actually, a substantial part of the chapter is devoted in proving the equivalence just stated. In closing, it is to be mentioned that, in the presence of a probability measure, *almost everywhere convergence* and *convergence in measure* become, respectively, *almost sure convergence* and *convergence in probability*.

Chapter 4 is devoted to the introduction of the concept of the integral of a random variable with respect to a measure, and the proof of some fundamental properties of the integral. When the underlying measure is a probability measure, the integral of a random variable becomes its expectation. The procedure of defining the concept of the integral follows three steps. The integral is first defined for a simple random variable, then for a nonnegative random variable, and finally for any random variable, provided the last step produces a meaningful quantity. This chapter is concluded with a result, Theorem 13, which transforms integration of a function of a random variable on an abstract probability space into integration of a real-valued function defined on the real line with respect to a probability measure on the Borel σ-field, which is the probability distribution of the random variable involved.

Chapter 5 is the first chapter where much of what was derived in the previous chapters is put to work. This chapter provides results that in a real sense constitute the workhorse whenever convergence of integrals is concerned, or differentiability under an integral sign is called for, or interchange of the order of integration is required. Some of the relevant theorems here are known by names such as the Lebesgue Monotone Convergence Theorem, the Fatou–Lebesgue Theorem, the Dominated Convergence Theorem, and the Fubini Theorem. Suitable modifications of the basic theorems in the chapter cover many important cases of both theoretical and applied interest. This is also the appropriate point to mention that many properties involving integrals are established by following a standard methodology; namely, the property in question is first proved for indicator functions, then for nonnegative simple random variables, next for nonnegative random variables, and finally for any random variables. Each step in this process relies heavily on the previous step, and the Lebesgue Monotone Convergence Theorem plays a central role.

Chapter 6 is the next chapter in which results of great utilitarian value are established. These results include the standard inequalities (Hölder (Cauchy–Schwarz), Minkowski, c_r, Jensen), and a combination of a probability/moment inequality, which produces the Markov and Tchebichev inequalities. A third kind of convergence—convergence in the rth mean—is also introduced and studied to a considerable extent. It is shown that convergence in the rth mean is equivalent to mutual convergence in the rth mean. Also, necessary and sufficient conditions for convergence in the rth mean are given. These conditions typically involve the concepts of uniform continuity and uniform integrability, which are important in their own right. It is an easy consequence of the Markov inequality that convergence in the rth mean implies convergence in probability. No direct relation may be established between convergence in the rth mean and almost sure convergence.

In Chapter 7, the concept of absolute continuity of a measure relative to another measure is introduced, and the most important result from utilitarian viewpoint is derived; this is the Radon–Nikodym Theorem, Theorem 3. This theorem provides the representation of a dominated measure as an indefinite integral of a nonnegative random variable with respect to the dominating measure. Its corollary provides the justification for what is done routinely in statistics, namely, employing a probability density function in integration. The Radon–Nikodym Theorem follows easily from the Lebesgue Decomposition Theorem, which is a deep result, and this in turn is based on the Hahn–Jordan Decomposition Theorem. Although all these results are proved in great detail, this is an instance where an instructor may choose to give the outlines of the first two theorems, and assign to students the study of the details.

Chapter 8 revolves around the concept of distribution functions and their basic properties. These properties include the fact that a distribution function is uniquely determined by its values on a set that is dense in the real line, that the discontinuities, being jumps only, are countably many, and that every distribution function is uniquely decomposed into two distribution functions, one of which is a step function and the other a continuous function. Next, the concepts of weak and complete convergence of a sequence of distribution functions are introduced, and it is shown that a sequence of distribution functions is weakly compact. In the final section of the chapter, the so-called Helly–Bray type results are established. This means that sufficient conditions are given under which weak or complete convergence of a sequence of distribution functions implies convergence of the integrals of a function with respect to the underlying distribution functions.

The purpose of Chapter 9 is to introduce the concept of conditional expectation of a random variable in an abstract setting; the concept of conditional probability then follows as a special case. A first installment of basic properties of conditional expectations is presented, and then the discussion proceeds with the derivation of the conditional versions of the standard inequalities dealt with in Chapter 6. Conditional versions of some of the standard convergence theorems of Chapter 5 are also derived, and the chapter is concluded with the discussion of further properties of conditional expectations, and an application linking the abstract definition of conditional probability with its elementary definition.

In Chapter 10, the concept of independence is considered first for events and then for σ-fields and random variables. A number of interesting results are discussed, including the fact that real-valued functions of independent random variables are independent random variables, and that the expectation of the product of independent random variables is the product of the individual expectations. However, the most substantial result in this chapter is the fact that factorization of the joint distribution function of a finite number of random variables implies their independence. This result is essentially based on the fact that σ-fields generated by independent fields are themselves independent.

Chapter 11 is devoted to characteristic functions, their basic properties, and their usage for probabilistic purposes. Once the concept of a characteristic function is defined, the fundamental result, referred to in the literature as the inversion formula, is established in a detailed manner, and several special cases are considered; also, the applicability of the formula is illustrated by means of two concrete examples. One of the main objectives in this chapter is that of establishing the Paul Lévy Continuity Theorem, thereby reducing the proof of weak convergence of a sequence of distribution functions to that of a sequence of characteristic functions, a problem much easier to deal with. This is done in Section 3, after a number of auxiliary results are first derived. The multidimensional version of the continuity theorem is essentially reduced to the one-dimensional case through the so called Cramér–Wold device; this is done in Section 4. Convolution of two distribution functions and several related results are discussed in Section 5, whereas in the following section additional properties of characteristic functions are established. These properties include the expansion of a characteristic function in a Taylor-like formula around zero with a remainder given in three different forms. A direct application of this expansion produces the Weak Law of Large Numbers and the Central Limit Theorem. In Section 8, the significance of the moments of a random variable is dramatized by showing that, under certain conditions, these moments completely determine the distribution of the random variable through its characteristic function. The rigorous proof of the relevant theorem makes use of a number of results from complex analysis, which for convenient reference are cited in the final section of the chapter.

In the next two chapters—Chapters 12 and 13—what may be considered as the backbone of classical probability is taken up: namely, the study of the central limit problem is considered under two settings, one for centered random variables and one for noncentered random variables. In both cases, a triangular array of row-wise independent random variables is considered, and, under some general and weak conditions, the totality of limiting laws—in the sense of weak convergence—is obtained for the row sums. As a very special case, necessary and sufficient conditions are given for convergence to the normal law for both the centered and the noncentered case. In the former case, sets of simpler sufficient conditions are also given for convergence to the normal law, whereas in the latter case, necessary and sufficient conditions are given for convergence to the Poisson law. The Central Limit Theorem in its usual simple form and the convergence of binomial probabilities to Poisson probabilities are also derived as very special cases of general results.

The main objective of Chapter 14 is to present a complete discussion of the Kolmogorov Strong Law of Large Numbers. Before this can be attempted, a long series of other results must be established, the first of which is the Kolmogorov inequalities. The discussion proceeds with the presentation of sufficient conditions for a series of centered random variables to convergence almost surely, the Borel–Cantelli Lemma, the Borel Zero–One Criterion, and two analytical results known as the Toeplitz Lemma and the Kronecker Lemma. Still the discussion of another two results is needed—one being a weak partial version of the Kolmogorov Strong Law of Large Numbers, and the other providing estimates of the expectation of a random variable in terms of sums of probabilities—before the Kolmogorov Strong Law of Large Numbers, Theorem 7, is stated and proved. In Section 4, it is seen that, if the expectation of the underlying random variable is not finite, as is the case in Theorem 7, a version of Theorem 7 is still true. However, if said expectation does not exist, then the averages are unbounded with probability 1. The chapter is concluded with a brief discussion of the tail σ-field of a sequence of random variables and pertinent results, including the Kolmogorov Zero–One Law for independent random variables, and the so-called Three Series Criterion.

The final chapter of the book, Chapter 15, is not an entirely integral part of the body of basic and fundamental results of measure-theoretic probability. Rather, it is one of the many possible choices of topics that this author could have covered. It serves as a very brief introduction to an important class of discrete parameter stochastic processes—stationary and ergodic or nonergodic processes—with a view toward proving the fundamental result, the Ergodic Theorem. In this framework, the concept of a stationary stochastic process is introduced, and some characterizations of stationarity are presented. The convenient and useful coordinate process is also introduced at this point. Next, the concepts of a transformation as well as a measure-preserving transformation are discussed, and it is shown that a measure-preserving transformation along with an arbitrary random variable define a stationary process. Associated with each transformation is a class of invariant sets and a class of almost sure invariant sets, both of which are σ-fields. A special class of transformations is the class of ergodic transformations, which are defined at this point. Invariance with respect to a transformation can also be defined for a stationary sequence of random variables, and it is so done. At this point, all the required machinery is available for the formulation of the Ergodic Theorem; also, its proof is presented, after some additional preliminary results are established. In the final section of the chapter, invariance of sets and of random variables is defined relative to a stationary process. Also, an alternative form of the Ergodic Theorem is given for nonergodic as well as ergodic processes. In closing, it is to be pointed out that one direction of the Kolmogorov Strong Law of Large Numbers is a special case of the Ergodic Theorem, as a sequence of independent identically distributed random variables forms a stationary and ergodic process.

Throughout the years, this author has drawn upon a number of sources in organizing his lectures. Some of those sources are among the books listed in the Selected References Section. However, the style and spirit of the discussions in this book lie closest to those of Loève's book. At this point, I would like to mention a recent worthy addition to the literature in measure theory—the book by Eric Vestrup, not least because Eric was one of our Ph.D. students at the University of California, Davis.

The lecture notes that eventually resulted in this book were revised, modified, and supplemented several times throughout the years; comments made by several of my students were very helpful in this respect. Unfortunately, they will have to remain anonymous, as I have not kept a complete record of them, and I do not want to provide an incomplete list. However, I do wish to thank my colleague and friend Costas Drossos for supplying a substantial number of exercises, mostly accompanied by answers. I would like to thank the following reviewers: Ibrahim Ahmad, University of Central Florida; Richard Johnson, University of Wisconsin; Madan Puri, Indiana University; Doraiswamy Ramachandran, California State University at Sacramento; and Zongwu Cai, University of North Carolina at Charlotte. Finally, thanks are due to my Project Assistant Newton Wai, who very skillfully turned my manuscript into an excellent typed text.

George G. Roussas
Davis, California
November 2003

TABLE OF CONTENTS

Chapter 1

CERTAIN CLASSES OF SETS, MEASURABILITY, AND POINTWISE APPROXIMATION

In this introductory chapter, the concepts of a field and of a σ-field are introduced, they are illustrated by means of examples, and some relevant basic results are derived. Also, the concept of a monotone class is defined and its relationship to certain fields and σ-fields is investigated. Given a collection of measurable spaces, their product space is defined, and some basic properties are established. The concept of a measurable mapping is introduced, and its relation to certain σ-fields is studied. Finally, it is shown that any random variable is the pointwise limit of a sequence of simple random variables.

1 MEASURABLE SPACES

Let Ω be an abstract set (or space) and let $\mathcal{C}$ be a class of subsets of Ω; i.e., $\mathcal{C} \subseteq \mathcal{P}(\Omega)$, the class of all subsets of Ω.

Definition 1

$\mathcal{C}$ is said to be a *field*, usually denoted by $\mathcal{F}$, if

(i) $\mathcal{C}$ is non-empty.

(ii) If $A \in \mathcal{C}$, then $A^c \in \mathcal{C}$.

(iii) If $A_1, A_2 \in \mathcal{C}$, then $A_1 \cup A_2 \in \mathcal{C}$. ■

Remark 1

In view of (ii) and (iii), the union $A_1 \cup A_2$ may be replaced by the intersection $A_1 \cap A_2$.