An Introduction to Measure-Theoretic Probability

Second edition

An Introduction to Measure-Theoretic Probability

Second edition

by

GEORGE G. ROUSSAS

Department of Statistics
University of California, Davis

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Academic Press is an imprint of Elsevier

Academic Press is an imprint of Elsevier
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK
Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands
225 Wyman Street, Waltham, MA 02451, USA
525 B Street, Suite 1800, San Diego, CA 92101-4495, USA

Second edition **2014**

Library of Congress Cataloging-in-Publication Data
Application submitted

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

For information on all **Academic Press** publications
visit our web site at store.elsevier.com

Printed and bound in USA
14 15 16 17 18 10 9 8 7 6 5 4 3 2 1

ISBN: 978-0-12-800042-7

This book is dedicated 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.

Pictured on the Cover

Carathéodory, Constantine (1873–1950)

He was born in Berlin to Greek parents and grew up in Brussels, Belgium. In high school, he twice won a prize as the best Mathematics student in Belgium. He studied Military Engineering in Belgium, and pursued graduate studies in Göttingen under the supervision of Hermann Minkowski.

He is known for his contributions to the theory of functions, the calculus of variations, and measure theory. His name is identified with the theory of outer measure, an application of which in measure theory is the so-called Carathéodory Extension Theorem. Also, he did work on the foundations of thermodynamics, and in 1909, he published the "first axiomatic rigid foundation of thermodynamics," which was acclaimed by Max Planck and Max Born.

From correspondence between Albert Einstein and Constantine Carathéodory, it may be deduced that Carathéodory's work helped Einstein in shaping some of his theories.

In 1924, he was appointed professor of Mathematics at the University of Munich, where he stayed until his death in 1950.

Preface to First Edition

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 extension 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 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 the 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 (measurable) 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 independence of the random variables involved. 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.

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

Preface to Second Edition

This is a revised version of the first edition of the book with copyright year 2005.

The basic character of the book remains the same, although its style is slightly different. Whatever changes were effected were made to correct misprints and oversights; add some clarifying points, as well as insert more references to previous parts of the book in support of arguments made; make minor modifications in the formulation, and in particular, the proof of some results; and supply additional exercises.

Specifically, the formulation of Theorem 8 in Chapter 3 has been rearranged. The proof of Theorem 3, case 3, in Chapter 4, has been simplified. The proof of Theorem 12 in Chapter 5 has been modified. Proposition 1, replaces Remark 6(ii) in Chapter 7. The proof of Theorem 3(iii) in Chapter 8 has been modified, and so has the concluding part of the proof of Theorem 5 in the same chapter. Likewise for the proofs of Theorems 7 and 8 in the same chapter. Remark 2 was inserted in Chapter 9 in order to further illustrate the abstract definition and the significance of the conditional expectation.

Section 3 of Chapter 11 has been restructured. Theorem 3 has been split into two parts, Theorem 3 and Theorem 3^*. Part (i) is the same in both of these theorems, as well as in the original theorem. There is a difference, however, in the formulation of part (ii) in the new versions of the theorems. Theorem 3 here is formulated along the familiar lines involving distribution functions and characteristic functions of random variables. Its formulation is followed by two lemmas, which facilitate its proof. The formulation of the second part of Theorem 3^* is more general, and along the same lines as the converse of the original Theorem 3. Theorem 3^* is also followed by two lemmas and one proposition, which lead to its justification. This section is concluded with two propositions, Propositions 2 and 3, where some restrictions imposed in the formulation of Lemmas 1–4 and Proposition 1 are lifted. In the same chapter, Chapter 11, the proof of Theorem 8 is essentially split into two parts, with the insertion of a "Statement" (just a few lines below relation (11.28)), in order to emphasize the asserted uniformity in the convergence.

In Chapter 12, Example 3 was added, first to illustrate the process of checking the Lindeberg-Feller condition, and second to provide some insight into this condition.

In Chapter 14, the second part of Lemma 7 has been modified, and so has its proof .

Finally, a new chapter, Chapter 16, has been added. This Chapter discusses some material on statistical inference, and it was added for the benefit of statistically oriented users of the book and upon the suggestion of a reviewer of the revised edition of the book. Its main purpose, however, is to demonstrate as to how some of the theorems, corollaries, etc. discussed in the book apply in establishing statistically inference results.

For a chapter-by-chapter brief description of the material discussed in the book, and also advice us to how the book can be used, the reader should go over the preface of its first edition.

The *Answers Manual* has been revised along the same lines as the text of the book. Thus, misprints and oversights have been corrected, and a handful of solutions have been modified. Of course, solutions to all new exercises are supplied. Again, the *Answers Manual*, in its revised version, will be made available to all those instructors who adopt the book as the textbook in their course.

Misprints and oversights were located in the usual way; that is, by teaching from the book. Many of the misprints and oversights were pointed out by attentive students. In this respect, special mention should be made of my students Qiuyan Xu and Gabriel Becker. Clarifications, modifications, and rearrangement of material, as described earlier, were also stimulated, to a large extent, by observations made and questions posed by students. Warm thanks are extended to all those who took my two-quarter course the last two offerings. Also, I am grateful to Stacy Hill and Paul Ressel for constructive comments. In particular, I am indebted to Michael McAssey, a formerly graduate student in the Department of Statistics, for the significant role he played toward the revision of the book and the *Answers Manual*. The accuracy and efficiency by which he handled the material was absolutely exemplary. Thanks are also due to Chu Shing (Randy) Lai for most efficiently implementing some corrections and inserting additional material into the book and the Answer Manual.

In closing, I consider it imperative to mention the following facts. Each chapter is introduced by a brief summary, describing the content of the chapter. In addition, there is an appendix in the book, Appendix A, where a much more extensive description is provided, chapter-by-chapter. It is my opinion that the reader would benefit greatly by reading this appendix before embarking on the study of the chapters. In this revision, a new appendix, Appendix B, has been added, providing a brief review of the Riemann–Stieltjes integral, and its relationship to the Riemann integral and the Lebesgue integral on the real line. The Riemann–Stieltjes integral is used explicitly in parts of Chapter 8, and implicitly in part of Chapters 11 through 13. Finally, it is mentioned here that some notation and abbreviations have been added to refresh readers' memory and ensure uniformity in notation.

George G. Roussas
Davis, California
September 2013

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.

1.1 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 nonempty.
(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 $A_1 \cap A_2$.

Examples. (Recall that a set is *countable* if it is either finite or it has the same cardinality as the set of integers. In the latter case it is *countably infinite*. A set is *uncountable* if it has the same cardinality as the real numbers.)

(1) $\mathcal{C} = \{\oslash, \Omega\}$ is a field called the *trivial* field. (It is the smallest possible field.)
(2) $\mathcal{C} = \{\text{all subsets of } \Omega\} = \mathcal{P}(\Omega)$ is a field called the *discrete* field. (It is the largest possible field.)
(3) $\mathcal{C} = \{\oslash, A, A^c, \Omega\}$ for some A with $\oslash \subset A \subset \Omega$.
(4) Let Ω be infinite (countably or not) and let $\mathcal{C} = \{A \subseteq \Omega;\ A \text{ is finite or } A^c \text{ is finite}\}$. Then $\mathcal{C}$ is a field.
(5) Let $\mathcal{C}$ be the class of all (finite) sums (unions of pairwise disjoint sets) of the partitioning sets of a finite partition of an arbitrary set Ω (see Definition 12 below). Then $\mathcal{C}$ is a field (*induced* or *generated* by the underlying partition). ■

An Introduction to Measure-Theoretic Probability, Second Edition. http://dx.doi.org/10.1016/B978-0-12-800042-7.00001-3

Remark 2. In Example 4, it is to be observed that if Ω is finite rather than infinite, then $\mathcal{C} = \mathcal{P}(\Omega)$.

Consequences of Definition 1.

(1) $\Omega, \oslash \in \mathcal{F}$ for every $\mathcal{F}$.
(2) If $A_j \in \mathcal{F}$, $j = 1, \ldots, n$, then $\bigcup_{j=1}^{n} A_j \in \mathcal{F}$.
(3) If $A_j \in \mathcal{F}$, $j = 1, \ldots, n$, then $\bigcap_{j=1}^{n} A_j \in \mathcal{F}$.

Remark 3. It is shown by examples that $A_j \in \mathcal{F}$, $j \geq 1$, need not imply $\bigcup_{j=1}^{\infty} A_j \in \mathcal{F}$, and similarly for $\bigcap_{j=1}^{\infty} A_j$ (see Remark 5 below).

Definition 2. $\mathcal{C}$ is said to be a σ-*field*, usually denoted by $\mathcal{A}$, if it is a field and (iii) in Definition 1 is strengthened to

(iii′) If $A_j \in \mathcal{C}$, $j = 1, 2, \ldots$, then $\bigcup_{j=1}^{\infty} A_j \in \mathcal{C}$. ■

Remark 4. In view of (ii) and (iii), the union in (iii′) may be replaced by the intersection $\bigcap_{j=1}^{\infty} A_j$.

Examples.

(6) $\mathcal{C} = \{\oslash, \Omega\}$ is a σ-field called the *trivial* σ-*field*.
(7) $\mathcal{C} = \mathcal{P}(\Omega)$ is a σ-field called the *discrete* σ-*field*.
(8) Let Ω be uncountable and let $\mathcal{C} = \{A \subseteq \Omega;\ A$ is countable or A^c is countable$\}$. Then $\mathcal{C}$ is a σ-field. (Of course, if Ω is countably infinite, then $\mathcal{C} = \mathcal{P}(\Omega)$).
(9) Let $\mathcal{C}$ be the class of all countable sums of the partitioning sets of a countable partition of an arbitrary set Ω. Then $\mathcal{C}$ is a σ-field (*induced* or *generated* by the underlying partition). ■

Remark 5. A σ-field is always a field, but a field need not be a σ-field. In fact, in Example 4 take $\Omega = \Re$ (real line), and let $A_j = \{k$ integer; $-j \leq k \leq j\}$, $j = 0, 1, \ldots$ Then $A_j \in \mathcal{C}$, $\bigcup_{j=0}^{n} A_j \in \mathcal{C}$ for any $n = 0, 1, \ldots$ but $\bigcup_{j=0}^{\infty} A_j$ (= set of all integers) $\notin \mathcal{C}$.

Let I be any index set. Then

Theorem 1.

(i) *If* $\mathcal{F}_j$, $j \in I$ *are fields, so is* $\bigcap_{j \in I} \mathcal{F}_j = \{A \subseteq \Omega;\ A \in \mathcal{F}_j,\ j \in I\}$.
(ii) *If* $\mathcal{A}_j$, $j \in I$ *are* σ-*fields, so is* $\bigcap_{j \in I} \mathcal{A}_j = \{A \subseteq \Omega;\ A \in \mathcal{A}_j,\ j \in I\}$. ■

Proof. Immediate. ■

Let $\mathcal{C}$ be any class of subsets of Ω. Then

Theorem 2.

(i) There is a unique minimal field containing $\mathcal{C}$. This is denoted by $\mathcal{F}(\mathcal{C})$ and is called the *field generated by* $\mathcal{C}$.

(ii) There is a unique minimal σ-field containing $\mathcal{C}$. This is denoted by $\sigma(\mathcal{C})$ and is called the *σ-field generated by* $\mathcal{C}$. ■

Proof.

(i) $\mathcal{F}(\mathcal{C}) = \bigcap_{j \in I} \mathcal{F}_j$, where $\{\mathcal{F}_j, j \in I\}$ is the nonempty class of all fields containing $\mathcal{C}$.

(ii) $\sigma(\mathcal{C}) = \bigcap_{j \in I} \mathcal{A}_j$, where $\{\mathcal{A}_j, j \in I\}$ is the nonempty class of all σ-fields containing $\mathcal{C}$. ■

Remark 6. Clearly, $\sigma(\mathcal{F}(\mathcal{C})) = \sigma(\mathcal{C})$. Indeed, $\mathcal{C} \subseteq \mathcal{F}(\mathcal{C})$, which implies $\sigma(\mathcal{C}) \subseteq \sigma(\mathcal{F}(\mathcal{C}))$. Also, for every σ-field $\mathcal{A}_i \supseteq \mathcal{C}$ it holds $\mathcal{A}_i \supseteq \mathcal{F}(\mathcal{C})$, since $\mathcal{A}_i$ is a field (being a σ-field), and $\mathcal{F}(\mathcal{C})$ is the minimal field (over $\mathcal{C}$). Hence $\sigma(\mathcal{C}) = \bigcap_i \mathcal{A}_i \supseteq \mathcal{F}(\mathcal{C})$. Since $\sigma(\mathcal{C})$ is a σ-field, it contains the minimal σ-field over $\mathcal{F}(\mathcal{C})$, $\sigma(\mathcal{F}(\mathcal{C}))$; i.e., $\sigma(\mathcal{C}) \supseteq \sigma(\mathcal{F}(\mathcal{C}))$. Hence $\sigma(\mathcal{C}) = \sigma(\mathcal{F}(\mathcal{C}))$.

Application 1. Let $\Omega = \Re$ and $\mathcal{C}_0 = \{$all intervals in $\Re\} = \{(x, y), (x, y], [x, y), [x, y], (-\infty, a), (-\infty, a], (b, \infty), [b, \infty);\ x, y \in \Re, x < y, a, b \in \Re\}$. Then $\sigma(\mathcal{C}_0)$ is denoted by $\mathcal{B}$ and is called the *Borel σ-field* over the real line. The sets in $\mathcal{B}$ are called *Borel sets*. Let $\bar{\Re} = \Re \cup \{-\infty, \infty\}$. $\bar{\Re}$ is called the *extended* real line and the σ-field $\bar{\mathcal{B}}$ generated by $\mathcal{B} \cup \{-\infty\} \cup \{\infty\}$ the *extended* Borel σ-field.

Remark 7. $\{x\} \in \mathcal{B}$ for every $x \in \Re$. Indeed, $\{x\} = \bigcap_{n=1}^{\infty} \left[x, x + \frac{1}{n}\right]$ with $\left[x, x + \frac{1}{n}\right] \in \mathcal{B}$. Hence $\bigcap_{n=1}^{\infty} \left[x, x + \frac{1}{n}\right] \in \mathcal{B}$, or $\{x\} \in \mathcal{B}$. Alternatively, with $a < x < b$, we have $\{x\} = (a, x] \cap [x, b) \in \mathcal{B}$.

Definition 3. The pair $(\Omega, \mathcal{A})$ is called a *measurable space* and the sets in $\mathcal{A}$ *measurable sets*. In particular, $(\Re, \mathcal{B})$ is called the *Borel real line*, and $(\bar{\Re}, \bar{\mathcal{B}})$ the *extended Borel real line*. ■

Let $\mathcal{C}$ again be a class of subsets of Ω. Then

Definition 4. $\mathcal{C}$ is called a *monotone* class if $A_j \in \mathcal{C}$, $j = 1, 2, \ldots$ and $A_j \uparrow$ (i.e., $A_1 \subseteq A_2 \subseteq \cdots$) or $A_j \downarrow$ (i.e., $A_1 \supseteq A_2 \supseteq \cdots$), then $\lim_{j\to\infty} A_j \stackrel{def}{=} \bigcup_{j=1}^{\infty} A_j \in \mathcal{C}$ and $\lim_{j\to\infty} A_j \stackrel{def}{=} \bigcap_{j=1}^{\infty} A_j \in \mathcal{C}$, respectively. ■

Theorem 3. A σ-field $\mathcal{A}$ is a monotone field (i.e., a field that is also a monotone class) and conversely. ■

Proof. One direction is immediate. As for the other, let $\mathcal{F}$ be a monotone field and let any $A_j \in \mathcal{F}$, $j = 1, 2, \ldots$. To show that $\bigcup_{j=1}^{\infty} A_j \in \mathcal{F}$. We have: $\bigcup_{j=1}^{\infty} A_j = A_1 \cup (A_1 \cup A_2) \cup \cdots \cup (A_1 \cup \cdots \cup A_n) \cup \cdots = \bigcup_{n=1}^{\infty} B_n$, where $B_n = \bigcup_{j=1}^{n} A_j$, and hence $B_n \in \mathcal{F}$, $n = 1, 2, \ldots$ and $B_n \uparrow$. Thus $\bigcup_{n=1}^{\infty} B_n \in \mathcal{F}$. ■

Theorem 4. If $\mathcal{M}_j$, $j \in I$, are monotone classes, so is $\bigcap_{j \in I} \mathcal{M}_j = \{A \subseteq \Omega;\ A \in \mathcal{M}_j, j \in I\}$. ■

Proof. Immediate. ■

Theorem 5. There is a unique minimal monotone class $\mathcal{M}$ containing $\mathcal{C}$. ■

Proof. $\mathcal{M} = \bigcap_{j \in I} \mathcal{M}_j$, where $\{\mathcal{M}_j, j \in I\}$ is the nonempty class of all monotone classes containing $\mathcal{C}$. ■

Remark 8. $\{\mathcal{M}_j, j \in I\}$ is nonempty since $\sigma(\mathcal{C})$ or $\mathcal{P}(\Omega)$ belong in it.

Remark 9. It may be seen by means of examples (see Exercise 12) that a monotone class need not be a field.

However, see the next lemma, as well as Theorem 6.

Lemma 1. Let $\mathcal{C}$ be a field and $\mathcal{M}$ be the minimal monotone class containing $\mathcal{C}$. Then $\mathcal{M}$ is a field.

Proof. In order to prove that $\mathcal{M}$ is a field, it suffices to prove that relations (*) hold, where

$$(*) \left\{ \text{for every } A, B \in \mathcal{M}, \text{ we have: } \begin{array}{l} \text{(i) } A \cap B \in \mathcal{M} \\ \text{(ii) } A^c \cap B \in \mathcal{M} \\ \text{(iii) } A \cap B^c \in \mathcal{M} \end{array} \right\}.$$

(That is, for every $A, B \in \mathcal{M}$, their intersection is in $\mathcal{M}$, and so is the intersection of any one of them by the complement of the other.)

In fact, $\mathcal{M} \supseteq \mathcal{C}$, implies $\Omega \in \mathcal{M}$. Taking $B = \Omega$, we get that for every $A \in \mathcal{M}$, $A^c \cap \Omega = A^c \in \mathcal{M}$ (by (ii)). Since also $A \cap B \in \mathcal{M}$ (by (i)) for all $A, B \in \mathcal{M}$, the proof would be completed.

In order to establish (*), we follow the following three steps:

Step 1. For any $A \in \mathcal{M}$, define $\mathcal{M}_A = \{B \in \mathcal{M}; (*) \text{ holds}\}$, so that $\mathcal{M}_A \subseteq \mathcal{M}$. Obviously $A \in \mathcal{M}_A$, since $\oslash \in \mathcal{M}$. It is asserted that $\mathcal{M}_A$ is a *monotone class*. Let $B_j \in \mathcal{M}_A$, $j = 1, 2, \ldots$, $B_j \uparrow$. To show that $\bigcup_{j=1}^{\infty} B_j \stackrel{def}{=} B \in \mathcal{M}_A$; i.e., to show that (*) holds. We have: $A \cap B = A \cap (\bigcup_j B_j) = \bigcup_j (A \cap B_j) \in \mathcal{M}$, since $\mathcal{M}$ is monotone and $A \cap B_j \uparrow$. Next, $A^c \cap B = A^c \cap (\bigcup_j B_j) = \bigcup_j (A^c \cap B_j) \in \mathcal{M}$ since $A^c \cap B_j \in \mathcal{M}$, by (*)(ii), and $A^c \cap B_j \uparrow$. Finally, $A \cap B^c = A \cap (\bigcup_j B_j)^c = A \cap (\bigcap_j B_j^c) = \bigcap_j (A \cap B_j^c)$ with $A \cap B_j^c \in \mathcal{M}$ by (*)(iii) and $A \cap B_j^c \downarrow$, so that $\bigcap_j (A \cap B_j^c) \in \mathcal{M}$ since $\mathcal{M}$ is monotone. The case that $B_j \downarrow$ is treated similarly, and the proof that $\mathcal{M}_A$ is a monotone class is complete.

Step 2. If $A \in \mathcal{C}$, *then* $\mathcal{M}_A = \mathcal{M}$. As already mentioned, $\mathcal{M}_A \subseteq \mathcal{M}$. So it suffices to prove that $\mathcal{M} \subseteq \mathcal{M}_A$. Let $B \in \mathcal{C}$. Then (*) holds and hence $B \in \mathcal{M}_A$. Therefore $\mathcal{C} \subseteq \mathcal{M}_A$. By step 1, $\mathcal{M}_A$ is a monotone class and $\mathcal{M}$ is the minimal monotone class containing $\mathcal{C}$. Thus $\mathcal{M} \subseteq \mathcal{M}_A$ and hence $\mathcal{M}_A = \mathcal{M}$.

Step 3. If A is any set in $\mathcal{M}$, *then* $\mathcal{M}_A = \mathcal{M}$. We show that $\mathcal{C} \subseteq \mathcal{M}_A$, which implies $\mathcal{M} \subseteq \mathcal{M}_A$ since $\mathcal{M}_A$ is a monotone class containing $\mathcal{C}$ and $\mathcal{M}$ is the minimal monotone class over $\mathcal{C}$. Since also $\mathcal{M}_A \subseteq \mathcal{M}$, the result $\mathcal{M}_A = \mathcal{M}$ would follow. To show $\mathcal{C} \subseteq \mathcal{M}_A$, take $B \in \mathcal{C}$ and consider $\mathcal{M}_B$. Then $\mathcal{M}_B = \mathcal{M}$ by step 2. Since $A \in \mathcal{M}$, we have $A \in \mathcal{M}_B$, which implies that $B \cap A$, $B^c \cap A$, and $B \cap A^c$ all belong in $\mathcal{M}$; or $A \cap B$, $A^c \cap B$, and $A \cap B^c$ belong in $\mathcal{M}$, which means that $B \in \mathcal{M}_A$. ■

Theorem 6. Let $\mathcal{C}$ be a field and $\mathcal{M}$ be the minimal monotone class containing $\mathcal{C}$. Then $\mathcal{M} = \sigma(\mathcal{C})$. ■

Proof. Evidently, $\mathcal{M} \subseteq \sigma(\mathcal{C})$ since every σ-field is a monotone class. By Lemma 1, $\mathcal{M}$ is a field, and hence a σ-field, by Theorem 3. Thus, $\mathcal{M} \supseteq \sigma(\mathcal{C})$ and hence $\mathcal{M} = \sigma(\mathcal{C})$. ■

Remark 10. Lemma 1 and Theorem 6 just discussed provide an illustration of the intricate relation of fields, monotone classes, and σ-fields in a certain setting. As will also be seen in several places in this book, monotone classes are often used as tools in arguments meant to establish results about σ-fields. In this kind of arguments, the roles of a field and of a monotone class may be substituted by the so-called *π-systems* and *λ-systems*, respectively. The definition of these concepts may be found, for example, in page 41 in Billingsley (1995). A result analogous to Theorem 6 is then Theorem 1.3 in page 5 of the reference just cited, which states that: If $\mathcal{P}$ is a π-system and $\mathcal{G}$ is a λ-system, then $\mathcal{P} \subset \mathcal{G}$ implies $\sigma(\mathcal{P}) \subset \mathcal{G}$.

1.2 Product Measurable Spaces

Consider the measurable spaces $(\Omega_1, \mathcal{A}_1)$, $(\Omega_2, \mathcal{A}_2)$. Then

Definition 5. The *product space* of Ω_1, Ω_2, denoted by $\Omega_1 \times \Omega_2$, is defined as follows: $\Omega_1 \times \Omega_2 = \{\omega = (\omega_1, \omega_2); \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$. In particular, for $A \in \mathcal{A}_1$, $B \in \mathcal{A}_2$ the product of A, B, denoted by $A \times B$, is defined by: $A \times B = \{\omega = (\omega_1, \omega_2); \omega_1 \in A, \omega_2 \in B\}$, and the subsets $A \times B$ of $\Omega_1 \times \Omega_2$ for $A \in \mathcal{A}_1$, $B \in \mathcal{A}_2$ are called (measurable) *rectangles*. A, B are called the *sides* of the rectangle. ■

From Definition 5, one easily verifies the following lemma.

Lemma 2. Consider the rectangle $E = A \times B$. Then, with "+" denoting union of disjoint events,

(i) $E^c = (A \times B^c) + (A^c \times \Omega_2) = (A^c \times B) + (\Omega_1 \times B^c)$.
Consider the rectangles $E_1 = A_1 \times B_1$, $E_2 = A_2 \times B_2$. Then

(ii) $E_1 \cap E_2 = (A_1 \cap A_2) \times (B_1 \cap B_2)$. Hence $E_1 \cap E_2 = \oslash$ if and only if at least one of the sets $A_1 \cap A_2$, $B_1 \cap B_2$ is $\oslash$.
Consider the rectangles E_1, E_2 as above, and the rectangles $F_1 = A'_1 \times B'_1$, $F_2 = A'_2 \times B'_2$. Then

(iii)
$$\begin{aligned}(E_1 \cap F_1) \cap (E_2 \cap F_2) &= [(A_1 \cap A'_1) \times (B_1 \cap B'_1)] \cap [(A_2 \cap A'_2) \\ &\quad \times (B_2 \cap B'_2)] \qquad \text{(by (ii))} \\ &= [(A_1 \cap A'_1) \cap (A_2 \cap A'_2)] \\ &\quad \times [(B_1 \cap B'_1) \cap (B_2 \cap B'_2)] \qquad \text{(by (ii))} \\ &= [(A_1 \cap A_2) \cap (A'_1 \cap A'_2)] \\ &\quad \times [(B_1 \cap B_2) \cap (B'_1 \cap B'_2)].\end{aligned}$$

Hence, the left-hand side is $\oslash$ if and only if at least one of $(A_1 \cap A_2) \cap (A'_1 \cap A'_2)$, $(B_1 \cap B_2) \cap (B'_1 \cap B'_2)$ is $\oslash$.

Theorem 7. Let $\mathcal{C}$ be the class of all finite *sums* (i.e., unions of pairwise disjoint) of rectangles $A \times B$ with $A \in \mathcal{A}_1$, $B \in \mathcal{A}_2$. Then $\mathcal{C}$ is a field (of subsets of $\Omega_1 \times \Omega_2$). ■

Proof. Clearly, $\mathcal{C} \neq \oslash$. Next, let $E, F \in \mathcal{C}$. Then we show that $E \cap F \in \mathcal{C}$. In fact, $E, F \in \mathcal{C}$ implies that $E = \sum_{i=1}^{m} E_i$, $F = \sum_{j=1}^{n} F_j$ with $E_i = A_i \times B_i, i = 1, \ldots, m$, $F_j = A'_j \times B'_j, j = 1, \ldots, n$. Thus $E \cap F = \bigcup_{i=1}^{m} \bigcup_{j=1}^{n} (E_i \cap F_j)$ and $E_i \cap F_j, E_{i'} \cap F_{j'}$ are disjoint for $(i, j) \neq (i', j')$ by Lemma 2 (ii), (iii). Indeed, in Lemma 2 (iii), make the identification: $A_1 = A_i, B_1 = B_i, A_2 = A_{i'}, B_2 = B_{i'}$, $A'_1 = A'_j, B'_1 = B'_j, A'_2 = A'_{j'}, B'_2 = B'_{j'}$ to get $(E_i \cap F_j) \cap (E_{i'} \cap F_{j'}) = [(A_i \cap A'_i) \cap (A'_j \cap A'_{j'})] \times [(B_i \cap B_{i'}) \cap (B'_j \cap B'_{j'})]$ by the third line on the right-hand side in Lemma 2(iii), and at least one of $(A_i \cap A_{i'}) \cap (A'_j \cap A'_{j'})$, $(B_i \cap B_{i'}) \cap (B'_j \cap B'_{j'})$ is equal to $\oslash$. Then, by Lemma 2(iii) again, $(E_i \cap F_j) \cap (E_{i'} \cap F_{j'}) = \oslash$, and therefore $E \cap F = \sum_{i=1}^{m} \sum_{j=1}^{n} (E_i \cap F_j)$. However, $E_i \cap F_j = (A_i \cap A'_j) \times (B_i \cap B'_j)$ (by Lemma 2(ii)), and $A_i \cap A'_j \in \mathcal{A}_1, B_i \cap B'_j \in \mathcal{A}_2, i = 1, \ldots, m, j = 1, \ldots, n$. Thus $E \cap F$ is the sum of finitely many rectangles and hence $E \cap F \in \mathcal{C}$. (By induction it is also true that if $E_k \in \mathcal{C}, k = 1, \ldots, \ell$, then $\bigcap_{k=1}^{\ell} E_k \in \mathcal{C}$.) Finally, $E^c = (\sum_{i=1}^{m} E_i)^c = \bigcap_{i=1}^{m} E_i^c = \bigcap_{i=1}^{m} [(A_i \times B_i^c) + (A_i^c \times \Omega_2)]$ (by Lemma 2(i)), and $A_i \times B_i^c, A_i^c \times \Omega_2$ are disjoint rectangles so that their sum is in $\mathcal{C}$. But then so is their intersection over $i = 1, \ldots, m$ by the induction just mentioned. The proof is completed. ∎

Remark 11. Clearly, the theorem also holds true if we start out with fields $\mathcal{F}_1$ and $\mathcal{F}_2$ rather than σ-fields $\mathcal{A}_1$ and $\mathcal{A}_2$.

Definition 6. The σ-field generated by the field $\mathcal{C}$ is called the *product σ-field* of $\mathcal{A}_1, \mathcal{A}_2$ and is denoted by $\mathcal{A}_1 \times \mathcal{A}_2$. The pair $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \times \mathcal{A}_2)$ is called the *product measurable space* of the (measurable) spaces $(\Omega_1, \mathcal{A}_1), (\Omega_2, \mathcal{A}_2)$. ∎

If we have $n \geq 2$ measurable spaces $(\Omega_i, \mathcal{A}_i), i = 1, \ldots, n$, the product measurable space $(\Omega_1 \times \cdots \times \Omega_n, \mathcal{A}_1 \times \cdots \times \mathcal{A}_n)$ is defined in an analogous way. In particular, if $\Omega_1 = \cdots = \Omega_n = \Re$ and $\mathcal{A}_1 = \cdots = \mathcal{A}_n = \mathcal{B}$, then the product space $(\Re^n, \mathcal{B}^n)$ is the *n-dimensional Borel space*, where $\Re^n = \Re \times \cdots \times \Re, \mathcal{B}^n = \mathcal{B} \times \cdots \times \mathcal{B}$ (n factors), and $\mathcal{B}^n$ is called the *n-dimensional Borel σ-field*. The members of $\mathcal{B}^n$ are called the *n-dimensional Borel sets.*

Now we consider the case of infinitely (countably or not) many measurable spaces $(\Omega_t, \mathcal{A}_t), t \in T$, where the $(\neq \oslash)$ index set T will usually be the real line or the positive half of it or the unit interval (0, 1) or [0,1].

Definition 7. The *product space* of $\Omega_t, t \in T$, denoted by $\prod_{t \in T} \Omega_t$ or Ω_T, is defined by $\Omega_T = \prod_{t \in T} \Omega_t = \{\omega = (\omega_t, t \in T); \omega_t \in \Omega_t, t \in T\}$. ∎

By forming the point $\omega = (\omega_t, t \in T)$ with $\omega_t \in \Omega_t, t \in T$, we tacitly assume, by invoking the *axiom of choice*, that there exists a function on T into $\bigcup_{t \in T} \Omega_t$ with $\Omega_t \neq \oslash, t \in T$, whose value at t, ω_t, belongs in Ω_t.

Now for $T = \{1, 2\}, \Omega_1 \times \Omega_2 = \{\omega = (\omega_1, \omega_2); \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$. Also, let $f : T \to \Omega_1 \cup \Omega_2$ such that $f(1) \in \Omega_1, f(2) \in \Omega_2$. Then $(f(1), f(2)) \in \Omega_1 \times \Omega_2$. Conversely, any $(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$ is the (ordered) pair of values of a function f on T into $\Omega_1 \cup \Omega_2$ with $f(1) \in \Omega_1, f(2) \in \Omega_2$; namely, the function for which $f(1) = \omega_1, f(2) = \omega_2$. Thus, $\Omega_1 \times \Omega_2$ may be looked upon as the collection of *all* functions f on T into $\Omega_1 \cup \Omega_2$ with $f(1) \in \Omega_1, f(2) \in \Omega_2$. Similar interpretation