Measure Theory

Volume I

V.I. Bogachev

Measure Theory

Volume I

Springer

Vladimir I. Bogachev
Department of Mechanics and Mathematics
Moscow State University
119992 Moscow
Russia

Mathematics Subject Classification (2000): 28-00, 46GXX, 60-XX

Library of Congress Control Number: 2006933997

ISBN-10 3-540-34513-2 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-34513-8 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media
springer.com

Typesetting by the author using a Springer LaTeX macro package
Cover design: deblik, Berlin

Printed on acid-free paper SPIN: 11693215 41/SPi 5 4 3 2 1 0

Preface

This book gives an exposition of the foundations of modern measure theory and offers three levels of presentation: a standard university graduate course, an advanced study containing some complements to the basic course (the material of this level corresponds to a variety of special courses), and, finally, more specialized topics partly covered by more than 850 exercises. The target readership includes graduate students interested in deeper knowledge of measure theory, instructors of courses in measure and integration theory, and researchers in all fields of mathematics. The book may serve as a source for many advanced courses or as a reference.

Volume 1 (Chapters 1–5) is devoted to the classical theory of measure and integral, created chiefly by H. Lebesgue and developed by many other mathematicians, in particular, by E. Borel, G. Vitali, W. Young, F. Riesz, D. Egoroff, N. Lusin, J. Radon, M. Fréchet, H. Hahn, C. Carathéodory, and O. Nikodym, whose results are presented in these chapters. Almost all the results in Chapters 1–5 were already known in the first third of the 20th century, but the methods of presentation, certainly, take into account later developments. The basic material designed for graduate students and oriented towards beginners covers approximately 100 pages in the first five chapters (i.e., less than 1/4 of those chapters) and includes the following sections: §1.1–1.7, §2.1–2.11, §3.2–3.4, §3.9, §4.1, §4.3, and some fragments of §5.1–5.4. It corresponds to a one-semester university course of real analysis (measure and integration theory) taught by the author at the Department of Mechanics and Mathematics at the Lomonosov Moscow University. The curriculum of this course is found at the end of the Bibliographical and Historical Comments. The required background includes only the basics of calculus (convergence of sequences and series, continuity of functions, open and closed sets in the real line, the Riemann integral) and linear algebra. Although knowledge of the Riemann integral is not formally assumed, I am convinced that the Riemann approach should be a starting point of the study of integration; acquaintance with the basics of the Riemann theory enables one to appreciate the depth and beauty of Lebesgue's creation. Some additional notions needed in particular sections are explained in the appropriate places. Naturally, the classical basic material of the first five chapters (without supplements) does not differ much from what is contained in many well-known textbooks on measure and integration or probability theory, e.g., Bauer [**70**], Halmos [**404**], Kolmogorov, Fomin [**536**], Loève [**617**], Natanson [**707**], Neveu [**713**], Parthasarathy [**739**], Royden [**829**], Shiryaev [**868**], and other books. An important feature of our exposition is that the listed sections contain only minimal material covered in real lectures. In particular, less attention than usual is given to measures on semirings etc. In general, the technical set-theoretic ingredients are considerably shortened. However, the corresponding material is not completely excluded: it is just transferred to supplements and exercises. In this way, one can substantially ease the first acquaintance with the subject when the abundance of definitions and set-theoretical constructions often make obstacles for understanding the principal ideas. Other sections of the main body of the book, supplements and exercises contain many things that are very useful in applications but seldom included in textbooks. There are two reasons why the standard course is included in full detail (rather than just mentioned in prerequisites): it makes the book completely self-contained and available to a much broader audience, in addition, many topics in the advanced material continue our discussion started in the basic course; it would be unnatural to give a continuation of a discussion without its beginning and origins. It should be noted that brevity of exposition has not been my priority; moreover, due to the described structure of the book, certain results are first presented in more special cases and only later are given in more general form. For example, our discussion of measures and integrals starts from finite measures, since the consideration of infinite values does not require new ideas, but for the beginner may overshadow the essence by rather artificial troubles with infinities. The organization of the book does not suggest reading from cover to cover; in particular, almost all sections in the supplements are independent of each other and are directly linked only to specific sections of the main part. A detailed table of contents is given. Here are brief comments on the structure of chapters.

In Chapter 1, the principal objects are countably additive measures on algebras and σ-algebras, and the main theorems are concerned with constructions and extensions of measures.

Chapter 2 is devoted to the construction of the Lebesgue integral, for which measurable functions are introduced first. The main theorems in this chapter are concerned with passage to the limit under the integral sign. The Lebesgue integral — one of the basic objects in this book — is not the most general type of integral. Apparently, its role in modern mathematics is explained by two factors: it possesses a sufficient and reasonable generality combined with aesthetic attractiveness.

In Chapter 3, we consider the most important operations on measures and functions: the Hahn–Jordan decomposition of signed measures, product measures, multiplication of measures by functions, convolutions of functions and measures, transformations of measures and change of variables. We discuss in detail finite and infinite products of measures. Fundamental theorems due to Radon&Nikodym and Fubini are presented.

Chapter 4 is devoted to spaces of integrable functions and spaces of measures. We discuss the geometric properties of the space L^p, study the uniform integrability, and prove several important theorems on convergence and boundedness of sequences of measures. Considerable attention is given to weak convergence and the weak topology in L^1. Finally, the structure properties of spaces of functions and measures are discussed.

In Chapter 5, we investigate connections between integration and differentiation and prove the classical theorems on the differentiability of functions of bounded variation and absolutely continuous functions and integration by parts. Covering theorems and the maximal function are discussed. The Henstock–Kurzweil integral is introduced and briefly studied.

Whereas the first volume presents the ideas that go back mainly to Lebesgue, the second volume (Chapters 6–10) is to a large extent the result of the development of ideas generated in 1930–1960 by a number of mathematicians, among which primarily one should mention A.N. Kolmogorov, J. von Neumann, and A.D. Alexandroff; other chief contributors are mentioned in the comments. The central subjects in Volume 2 are: transformations of measures, conditional measures, and weak convergence of measures. These three themes are closely interwoven and form the heart of modern measure theory. Typical measure spaces here are infinite dimensional: e.g., it is often convenient to consider a measure on the interval as a measure on the space $\{0,1\}^\infty$ of all sequences of zeros and ones. The point is that in spite of the fact that any reasonable measure space is isomorphic to an interval, a significant role is played by diverse additional structures on measure spaces: algebraic, topological, and differential. This is partly explained by the fact that many problems of modern measure theory grew under the influence of probability theory, the theory of dynamical systems, information theory, the theory of representations of groups, nonlinear analysis, and mathematical physics. All these fields brought into measure theory not only problems, methods, and terminology, but also inherent ways of thinking. Note also that the most fruitful directions in measure theory now border with other branches of mathematics.

Unlike the first volume, a considerable portion of material in Chapters 6–10 has not been presented in such detail in textbooks. Chapters 6–10 require also a deeper background. In addition to knowledge of the basic course, it is necessary to be familiar with the standard university course of functional analysis including elements of general topology (e.g., the textbook by Kolmogorov and Fomin covers the prerequisites). In some sections it is desirable to be familiar with fundamentals of probability theory (for this purpose, a concise book, Lamperti [**566**], can be recommended). In the second volume many themes touched on in the first volume find their natural development (for example, transformations of measures, convergence of measures, Souslin sets, connections between measure and topology).

Chapter 6 plays an important technical role: here we study various properties of Borel and Souslin sets in topological spaces and Borel mappings of

Souslin sets, in particular, several measurable selection and implicit function theorems are proved here. The birth of this direction is due to a great extent to the works of N. Lusin and M. Souslin. The exposition in this chapter has a clear set-theoretic and topological character with almost no measures. The principal results are very elegant, but are difficult in parts in the technical sense, and I decided not to hide these difficulties in exercises. However, this chapter can be viewed as a compendium of results to which one should resort in case of need in the subsequent chapters.

In Chapter 7, we discuss measures on topological spaces, their regularity properties, and extensions of measures, and examine the connections between measures and the associated functionals on function spaces. The branch of measure theory discussed here grew from the classical works of J. Radon and A.D. Alexandroff, and was strongly influenced (and still is) by general topology and descriptive set theory. The central object of the chapter is Radon measures. We also study in detail perfect and τ-additive measures. A separate section is devoted to the Daniell–Stone method. This method could have been explained already in Chapter 2, but it is more natural to place it close to the Riesz representation theorem in the topological framework. There is also a brief discussion of measures on locally convex spaces and their characteristic functionals (Fourier transforms).

In Chapter 8, directly linked only to Chapter 7, the theory of weak convergence of measures is presented. We prove several fundamental results due to A.D. Alexandroff, Yu.V. Prohorov and A.V. Skorohod, study the weak topology on spaces of measures and consider weak compactness. The topological properties of spaces of measures on topological spaces equipped with the weak topology are discussed. The concept of weak convergence of measures plays an important role in many applications, including stochastic analysis, mathematical statistics, and mathematical physics. Among many complementary results in this chapter one can mention a thorough discussion of convergence of measures on open sets and a proof of the Fichtenholz–Dieudonné–Grothendieck theorem.

Chapter 9 is devoted to transformations of measures. We discuss the properties of images of measures under mappings, the existence of preimages, various types of isomorphisms of measure spaces (for example, point, metric, topological), the absolute continuity of transformed measures, in particular, Lusin's (N)-property, transformations of measures by flows generated by vector fields, Haar measures on locally compact groups, the existence of invariant measures of transformations, and many other questions important for applications. The "nonlinear measure theory" discussed here originated in the 1930s in the works of G.D. Birkhoff, J. von Neumann, N.N. Bogolubov, N.M. Krylov, E. Hopf and other researchers in the theory of dynamical systems, and was also considerably influenced by other fields such as the integration on topological groups developed by A. Haar, A. Weil, and others. A separate section is devoted to the theory of Lebesgue spaces elaborated by V. Rohlin (such spaces are called here Lebesgue–Rohlin spaces).

Chapter 10 is close to Chapter 9 in its spirit. The principal ideas of this chapter go back to the works of A.N. Kolmogorov, J. von Neumann, J. Doob, and P. Lévy. It is concerned with conditional measures — the object that plays an exceptional role in measure theory as well as in numerous applications. We describe in detail connections between conditional measures and conditional expectations, prove the main theorems on convergence of conditional expectations, establish the existence of conditional measures under broad assumptions and clarify their relation to liftings. In addition, a concise introduction to the theory of martingales is given with views towards applications in measure theory. A separate section is devoted to ergodic theory — a fruitful field at the border of measure theory, probability theory, and mathematical physics. Finally, in this chapter we continue our study of Lebesgue–Rohlin spaces, and in particular, discuss measurable partitions.

Extensive complementary material is presented in the final sections of all chapters, where there are also a lot of exercises supplied with complete solutions or hints and references. Some exercises are merely theorems from the cited sources printed in a smaller font and are placed there to save space (so that the absence of hints means that I have no solutions different from the ones found in the cited works). The symbol ° marks exercises recommendable for graduate courses or self-study. Note also that many solutions have been borrowed from the cited works, but sometimes solutions simpler than the original ones are presented (this fact, however, is not indicated). It should be emphasized that many exercises given without references are either taken from the textbooks listed in the bibliographical comments or belong to the mathematical folklore. In such exercises, I omitted the sources (which appear in hints, though), since they are mostly secondary. It is possible that some exercises are new, but this is never claimed for the obvious reason that a seemingly new assertion could have been read in one of hundreds papers from the list of references or even heard from colleagues and later recalled.

The book contains an extensive bibliography and the bibliographical and historical comments. The comments are made separately on each volume, the bibliography in Volume 1 contains the works cited only in that volume, and Volume 2 contains the cumulative bibliography, where the works cited only in Volume 1 are marked with an asterisk. For each item in the list of references we indicate all pages where it is cited. The comments, in addition to remarks of a historical or bibliographical character, give references to works on many special aspects of measure theory, which could not be covered in a book of this size, but the information about which may be useful for the reader. A detailed subject index completes the book (Volume 1 contains only the index for that volume, and Volume 2 contains the cumulative index).

For all assertions and formulas we use the triple enumeration: the chapter number, section number, and assertion number (all assertions are numbered independently of their type within each section); numbers of formulas are given in brackets.

This book is intended as a complement to the existing large literature of advanced graduate-text type and provides the reader with a lot of material from many parts of measure theory which does not belong to the standard course but is necessary in order to read research literature in many areas. Modern measure theory is so vast that it cannot be adequately presented in one book. Moreover, even if one attempts to cover all the directions in a universal treatise, possibly in many volumes, due depth of presentation will not be achieved because of the excessive amount of required information from other fields. It appears that for an in-depth study not so voluminous expositions of specialized directions are more suitable. Such expositions already exist in a several directions (for example, the geometric measure theory, Hausdorff measures, probability distributions on Banach spaces, measures on groups, ergodic theory, Gaussian measures). Here a discussion of such directions is reduced to a minimum, in many cases just to mentioning their existence.

This book grew from my lectures at the Lomonosov Moscow University, and many related problems have been discussed in lectures, seminar talks and conversations with colleagues at many other universities and mathematical institutes in Moscow, St.-Petersburg, Kiev, Berlin, Bielefeld, Bonn, Oberwolfach, Paris, Strasburg, Cambridge, Warwick, Rome, Pisa, Vienna, Stockholm, Copenhagen, Zürich, Barcelona, Lisbon, Athens, Edmonton, Berkeley, Boston, Minneapolis, Santiago, Haifa, Kyoto, Beijing, Sydney, and many other places. Opportunities to work in the libraries of these institutions have been especially valuable. Through the years of work on this book I received from many individuals the considerable help in the form of remarks, corrections, additional references, historical comments etc. Not being able to mention here all those to whom I owe gratitude, I particularly thank H. Airault, E.A. Alekhno, E. Behrends, P.A. Borodin, G. Da Prato, D. Elworthy, V.V. Fedorchuk, M.I. Gordin, M.M. Gordina, V.P. Havin, N.V. Krylov, P. Lescot, G. Letta, A.A. Lodkin, E. Mayer-Wolf, P. Malliavin, P.-A. Meyer, L. Mejlbro, E. Priola, V.I. Ponomarev, Yu.V. Prohorov, M. Röckner, V.V. Sazonov, B. Schmuland, A.N. Shiryaev, A.V. Skorohod, O.G. Smolyanov, A.M. Stepin, V.N. Sudakov, V.I. Tarieladze, S.A. Telyakovskii, A.N. Tikhomirov, F. Topsøe, V.V. Ulyanov, H. von Weizsäcker, and M. Zakai. The character of presentation was considerably influenced by discussions with my colleagues at the chair of theory of functions and functional analysis at the Department of Mechanics and Mathematics of the Lomonosov Moscow University headed by the member of the Russian Academy of Science P.L. Ulyanov. For checking several preliminary versions of the book, numerous corrections, improvements and other related help I am very grateful to A.V. Kolesnikov, E.P. Krugova, K.V. Medvedev, O.V. Pugachev, T.S. Rybnikova, N.A. Tolmachev, R.A. Troupianskii, Yu.A. Zhereb'ev, and V.S. Zhuravlev. The book took its final form after Z. Lipecki read the manuscript and sent his corrections, comments, and certain materials that were not available to me. I thank J. Boys for careful copyediting and the editorial staff at Springer-Verlag for cooperation.

Moscow, August 2006 Vladimir Bogachev

Contents

Contents of Volume 2

CHAPTER 1

Constructions and extensions of measures

> I compiled these lectures not assuming from the reader any knowledge other than is found in the under-graduate programme of all departments; I can even say that not assuming anything except for acquaintance with the definition and the most elementary properties of integrals of continuous functions. But even if there is no necessity to know much before reading these lectures, it is yet necessary to have some practice of thinking in such matters.
>
> *H. Lebesgue.* Intégration et la récherche des fonctions primitives.

1.1. Measurement of length: introductory remarks

Many problems discussed in this book grew from the following question: which sets have length? This question clear at the first glance leads to two other questions: what is a "set" and what is a "number" (since one speaks of a qualitative measure of length)? We suppose throughout that some answers to these questions have been given and do not raise them further, although even the first constructions of measure theory lead to situations requiring greater certainty. We assume that the reader is familiar with the standard facts about real numbers, which are given in textbooks of calculus, and for "set theory" we take the basic assumptions of the "naive set theory" also presented in textbooks of calculus; sometimes the axiom of choice is employed. In the last section the reader will find a brief discussion of major set-theoretic problems related to measure theory. We use throughout the following set-theoretic relations and operations (in their usual sense): $A \subset B$ (the inclusion of a set A to a set B), $a \in A$ (the inclusion of an element a in a set A), $A \cup B$ (the union of sets A and B), $A \cap B$ (the intersection of sets A and B), $A \backslash B$ (the complement of B in A, i.e., the set of all points from A not belonging to B). Finally, let $A \triangle B$ denote the symmetric difference of two sets A and B, i.e., $A \triangle B = (A \cup B) \backslash (A \cap B)$. We write $A_n \uparrow A$ if $A_n \subset A_{n+1}$ and $A = \bigcup_{n=1}^{\infty} A_n$; we write $A_n \downarrow A$ if $A_{n+1} \subset A_n$ and $A = \bigcap_{n=1}^{\infty} A_n$.

The restriction of a function f to a set A is denoted by $f|_A$.

The standard symbols $\mathbb{N} = \{1, 2, \ldots\}$, $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}^n$ denote, respectively, the sets of all natural, integer, rational numbers, and the n-dimensional Euclidean space. The term "positive" means "strictly positive" with the exception of some special situations with the established terminology (e.g., the positive part of a function may be zero); similarly with "negative".

The following facts about the set $\mathbb{R}^1$ of real numbers are assumed to be known.

1) The sets $U \subset \mathbb{R}^1$ such that every point x from U belongs to U with some interval of the form $(x-\varepsilon, x+\varepsilon)$, where $\varepsilon > 0$, are called open; every open set is the union of a finite or countable collection of pairwise disjoint intervals or rays. The empty set is open by definition.

2) The closed sets are the complements to open sets; a set A is closed precisely when it contains all its limit points. We recall that a is called a limit point for A if every interval centered at a contains a point $b \neq a$ from A. It is clear that any unions and finite intersections of open sets are open. Thus, the real line is a topological space (more detailed information about topological spaces is given in Chapter 6).

It is clear that any intersections and finite unions of closed sets are closed. An important property of $\mathbb{R}^1$ is that the intersection of any decreasing sequence of nonempty bounded closed sets is nonempty. Depending on the way in which the real numbers have been introduced, this claim is either an axiom or is derived from other axioms. The principal concepts related to convergence of sequences and series are assumed to be known.

Let us now consider the problem of measurement of length. Let us aim at defining the length λ of subsets of the interval $I = [0,1]$. For an interval J of the form (a,b), $[a,b)$, $[a,b]$ or $(a,b]$, we set $\lambda(J) = |b-a|$. For a finite union of disjoint intervals $J_1,\ldots,J_n$, we set $\lambda\bigl(\bigcup_{i=1}^n J_i\bigr) = \sum_{i=1}^n \lambda(J_i)$. The sets of the indicated form are called *elementary*. We now have to make a non-trivial step and extend measure to non-elementary sets. A natural way of doing this, which goes back to antiquity, consists of approximating non-elementary sets by elementary ones. But how to approximate? The construction that leads to the so-called *Jordan measure* (which should be more precisely called the *Peano–Jordan measure* following the works Peano [**741**], Jordan [**472**]), is this: a set $A \subset I$ is Jordan measurable if for any $\varepsilon > 0$, there exist elementary sets A_ε and B_ε such that $A_\varepsilon \subset A \subset B_\varepsilon$ and $\lambda(B_\varepsilon \backslash A_\varepsilon) < \varepsilon$. It is clear that when $\varepsilon \to 0$, the lengths of A_ε and B_ε have a common limit, which one takes for $\lambda(A)$. Are all the sets assigned lengths after this procedure? No, not at all. For example, the set $\mathbb{Q}\cap I$ of rational numbers in the interval is not Jordan measurable. Indeed, it contains no elementary set of positive measure. On the other hand, any elementary set containing $\mathbb{Q}\cap I$ has measure 1. The question arises naturally about extensions of λ to larger domains. It is desirable to preserve the nice properties of length, which it possesses on the class of Jordan measurable sets. The most important of these properties are the additivity (i.e., $\lambda(A\cup B) = \lambda(A) + \lambda(B)$ for any disjoint sets A and B in the domain) and the invariance with respect to translations. The first property is even fulfilled in the following stronger form of countable additivity: if disjoint sets A_n together with their union $A = \bigcup_{n=1}^\infty A_n$ are Jordan measurable, then $\lambda(A) = \sum_{n=1}^\infty \lambda(A_n)$. As we shall see later, this problem admits solutions. The most important of them suggested by Lebesgue a century ago and leading to Lebesgue measurability consists of changing the way of approximating by elementary sets. Namely, by analogy with the ancient construction one introduces the outer measure λ^* for *every* set $A \subset I$ as the infimum of sums of measures of elementary sets forming countable covers of A. Then a set A is called Lebesgue measurable if the equality $\lambda^*(A) + \lambda^*(I\backslash A) = \lambda(I)$ holds, which can also be expressed in the form of the equality $\lambda^*(A) = \lambda_*(A)$, where the inner measure λ_* is defined *not* by means of inscribed sets as in the case of the Jordan measure, but by the equality $\lambda_*(A) = \lambda(I) - \lambda^*(I\backslash A)$. An equivalent description of the Lebesgue measurability in terms of approximations by elementary sets is this: for any $\varepsilon > 0$ there exists an elementary set A_ε such that $\lambda^*(A \triangle A_\varepsilon) < \varepsilon$. Now, unlike the Jordan measure, no inclusion of sets is required, i.e., "skew approximations" are admissible. This minor nuance leads to a substantial enlargement of the class of measurable sets. The enlargement is so great that the question of the existence of sets to which no measure is assigned becomes dependent on accepting or not accepting certain special set-theoretic axioms. We shall soon verify that the collection of Lebesgue measurable sets is closed with respect to countable unions, countable intersections, and complements. In addition, if we define the measure of a set A as the limit of measures of elementary sets approximating it in the above sense, then the extended measure turns out to be countably additive. All these claims will be derived from more general results. The role of the countable additivity is obvious from the very beginning: if one approximates a disc by unions of rectangles or triangles, then countable unions arise with necessity.

It follows from what has been said above that in the discussion of measures the key role is played by issues related to domains of definition and extensions. So the next section is devoted to principal classes of sets connected with domains of measures. It turns out in this discussion that the specifics of length on subsets of the real line play no role and it is reasonable from the very beginning to speak of measures of an arbitrary nature. Moreover, this point of view becomes necessary for considering measures on general spaces, e.g., manifolds or functional spaces, which is very important for many branches of mathematics and theoretical physics.

1.2. Algebras and σ-algebras

One of the principal concepts of measure theory is an algebra of sets.

1.2.1. Definition. *An algebra of sets $\mathcal{A}$ is a class of subsets of some fixed set X (called the space) such that*

(i) *X and the empty set belong to $\mathcal{A}$;*

(ii) *if $A,\ B \in \mathcal{A}$, then $A \cap B \in \mathcal{A}$, $A \cup B \in \mathcal{A}$, $A\backslash B \in \mathcal{A}$.*

In place of the condition $A\backslash B \in \mathcal{A}$ one could only require that $X\backslash B \in \mathcal{A}$ whenever $B \in \mathcal{A}$, since $A\backslash B = A \cap (X\backslash B)$ and $A \cup B = X\backslash\bigl((X\backslash A) \cap (X\backslash B)\bigr)$. It is sufficient as well to require in (ii) only that $A\backslash B \in \mathcal{A}$ for all $A, B \in \mathcal{A}$, since $A \cap B = A\backslash(A\backslash B)$.

Sometimes in the definition of an algebra the inclusion $X \in \mathcal{A}$ is replaced by the following wider assumption: there exists a set $E \in \mathcal{A}$ called the unit of the algebra such that $A \cap E = A$ for all $A \in \mathcal{A}$. It is clear that replacing X by E we arrive at our definition on a smaller space. It should be noted that not all of the results below extend to this wider concept.

1.2.2. Definition. *An algebra of sets $\mathcal{A}$ is called a σ-algebra if for any sequence of sets A_n in $\mathcal{A}$ one has $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.*

1.2.3. Definition. *A pair $(X, \mathcal{A})$ consisting of a set X and a σ-algebra $\mathcal{A}$ of its subsets is called a measurable space.*

The basic set (space) on which a σ-algebra or measure are given is most often denoted in this book by X; other frequent symbols are E, M, S (from "ensemble", "Menge", "set"), and Ω, a generally accepted symbol in probability theory. For denoting a σ-algebra it is traditional to use script Latin capitals (e.g., $\mathcal{A}$, $\mathcal{B}$, $\mathcal{E}$, $\mathcal{F}$, $\mathcal{L}$, $\mathcal{M}$, $\mathcal{S}$), Gothic capitals $\mathfrak{A}$, $\mathfrak{B}$, $\mathfrak{F}$, $\mathfrak{L}$, $\mathfrak{M}$, $\mathfrak{S}$ (i.e., A, B, F, L, M and S) and Greek letters (e.g., $\varSigma$, $\varLambda$, $\varGamma$, $\varXi$), although when necessary other symbols are used as well.

In the subsequent remarks and exercises some other classes of sets are mentioned such as semialgebras, rings, semirings, σ-rings, etc. These classes slightly differ in the operations they admit. It is clear that in the definition of a σ-algebra in place of stability with respect to countable unions one could require stability with respect to countable intersections. Indeed, by the formula $\bigcup_{n=1}^{\infty} A_n = X \backslash \bigcap_{n=1}^{\infty} (X \backslash A_n)$ and the stability of any algebra with respect to complementation it is seen that both properties are equivalent.

1.2.4. Example. The collection of finite unions of all intervals of the form $[a, b]$, $[a, b)$, $(a, b]$, (a, b) in the interval $[0, 1]$ is an algebra, but not a σ-algebra.

Clearly, the collection 2^X of all subsets of a fixed set X is a σ-algebra. The smallest σ-algebra is $(X, \varnothing)$. Any other σ-algebra of subsets of X is contained between these two trivial examples.

1.2.5. Definition. *Let $\mathcal{F}$ be a family of subsets of a space X. The smallest σ-algebra of subsets of X containing $\mathcal{F}$ is called the σ-algebra generated by $\mathcal{F}$ and is denoted by the symbol $\sigma(\mathcal{F})$. The algebra generated by $\mathcal{F}$ is defined as the smallest algebra containing $\mathcal{F}$.*

The smallest σ-algebra and algebra mentioned in the definition exist indeed.

1.2.6. Proposition. *Let X be a set. For any family $\mathcal{F}$ of subsets of X there exists a unique σ-algebra generated by $\mathcal{F}$. In addition, there exists a unique algebra generated by $\mathcal{F}$.*

PROOF. Set $\sigma(\mathcal{F}) = \bigcap_{\mathcal{F} \subset \mathcal{A}} \mathcal{A}$, where the intersection is taken over all σ-algebras of subsets of the space X containing all sets from $\mathcal{F}$. Such σ-algebras exist: for example, 2^X; their intersection by definition is the collection of all sets that belong to each of such σ-algebras. By construction, $\mathcal{F} \subset \sigma(\mathcal{F})$. If we are given a sequence of sets $A_n \in \sigma(\mathcal{F})$, then their intersection, union and complements belong to any σ-algebra $\mathcal{A}$ containing $\mathcal{F}$, hence belong to $\sigma(\mathcal{F})$, i.e., $\sigma(\mathcal{F})$ is a σ-algebra. The uniqueness is obvious from the fact that the existence of a σ-algebra $\mathcal{B}$ containing $\mathcal{F}$ but not containing $\sigma(\mathcal{F})$ contradicts the definition of $\sigma(\mathcal{F})$, since $\mathcal{B}\cap\sigma(\mathcal{F})$ contains $\mathcal{F}$ and is a σ-algebra. The case of an algebra is similar. □

Note that it follows from the definition that the class of sets formed by the complements of sets in $\mathcal{F}$ generates the same σ-algebra as $\mathcal{F}$. It is also clear that a countable class may generate an uncountable σ-algebra. For example, the intervals with rational endpoints generate the σ-algebra containing all single-point sets.

The algebra generated by a family of sets $\mathcal{F}$ can be easily described explicitly. To this end, let us add to $\mathcal{F}$ the empty set and denote by $\mathcal{F}_1$ the collection of all sets of this enlarged collection together with their complements. Then we denote by $\mathcal{F}_2$ the class of all finite intersections of sets in $\mathcal{F}_1$. The class $\mathcal{F}_3$ of all finite unions of sets in $\mathcal{F}_2$ is the algebra generated by $\mathcal{F}$. Indeed, it is clear that $\mathcal{F} \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3$ and that $\varnothing \in \mathcal{F}_3$. The class $\mathcal{F}_3$ admits any finite intersections, since if $A = \bigcup_{i=1}^n A_i$, $B = \bigcup_{j=1}^k B_j$, where $A_i, B_j \in \mathcal{F}_2$, then we have $A \cap B = \bigcup_{i\le n, j\le k} A_i \cap B_j$ and $A_i \cap B_j \in \mathcal{F}_2$. In addition, $\mathcal{F}_3$ is stable under complements. Indeed, if $E = E_1 \cup \cdots \cup E_n$, where $E_i \in \mathcal{F}_2$, then $X\backslash E = \bigcap_{i=1}^n (X\backslash E_i)$. Since $E_i = E_{i,1} \cap \cdots \cap E_{i,k_i}$, where $E_{i,j} \in \mathcal{F}_1$, one has $X\backslash E_i = \bigcup_{j=1}^{k_i}(X\backslash E_{i,j})$, where $D_{i,j} := X\backslash E_{i,j} \in \mathcal{F}_1$. Hence $X\backslash E = \bigcap_{i=1}^n \bigcup_{j=1}^{k_i} D_{i,j}$, which belongs to $\mathcal{F}_3$ by the stability of $\mathcal{F}_3$ with respect to finite unions and intersections. On the other hand, it is clear that $\mathcal{F}_3$ belongs to the algebra generated by $\mathcal{F}$.

One should not attempt to imagine the elements of the σ-algebra generated by the class $\mathcal{F}$ in a constructive form by means of countable unions, intersections or complements of the elements in $\mathcal{F}$. The point is that the above-mentioned operations can be repeated in an unlimited number of steps in any order. For example, one can form the class $\mathcal{F}_\sigma$ of countable unions of closed sets in the interval, then the class $\mathcal{F}_{\sigma\delta}$ of countable intersections of sets in $\mathcal{F}_\sigma$, and continue this process inductively. One will be obtaining new classes all the time, but even their union does not exhaust the σ-algebra generated by the closed sets (the proof of this fact is not trivial; see Exercises 6.10.30, 6.10.31, 6.10.32 in Chapter 6). In §1.10 we study the so-called A-operation, which gives all sets in the σ-algebra generated by intervals, but produces also other sets. Let us give an example where one can explicitly describe the σ-algebra generated by a class of sets.

1.2.7. Example. Let $\mathcal{A}_0$ be a σ-algebra of subsets in a space X. Suppose that a set $S \subset X$ does not belong to $\mathcal{A}_0$. Then the σ-algebra $\sigma\bigl(\mathcal{A}_0 \cup \{S\}\bigr)$, generated by $\mathcal{A}_0$ and the set S coincides with the collection of all sets of the form

$$E = (A \cap S)\cup\bigl(B \cap (X\backslash S)\bigr), \quad \text{where } A, B \in \mathcal{A}_0. \tag{1.2.1}$$

PROOF. All sets of the form (1.2.1) belong to the σ-algebra $\sigma\bigl(\mathcal{A}_0\cup\{S\}\bigr)$. On the other hand, the sets of the indicated type form a σ-algebra. Indeed,

$$X\backslash E = \bigl((X\backslash A)\cap S\bigr)\cup\bigl((X\backslash B)\cap(X\backslash S)\bigr),$$

since x does not belong to E precisely when either x belongs to S but not to A, or x belongs neither to S, nor to B. In addition, if the sets E_n are represented in the form (1.2.1) with some $A_n, B_n \in \mathcal{A}_0$, then $\bigcap_{n=1}^\infty E_n$ and $\bigcup_{n=1}^\infty E_n$ also have the form (1.2.1). For example, $\bigcap_{n=1}^\infty E_n$ has the form (1.2.1) with $A = \bigcap_{n=1}^\infty A_n$ and $B = \bigcap_{n=1}^\infty B_n$. Finally, all sets in $\mathcal{A}_0$ are obtained in the form (1.2.1) with $A = B$, and for obtaining S we take $A = X$ and $B = \varnothing$. □

In considerations involving σ-algebras the following simple properties of the set-theoretic operations are often useful.

1.2.8. Lemma. *Let $(A_\alpha)_{\alpha\in\Lambda}$ be a family of subsets of a set X and let $f\colon\ E\to X$ be an arbitrary mapping of a set E to X. Then*

$$X\backslash\bigcup_{\alpha\in\Lambda}A_\alpha = \bigcap_{\alpha\in\Lambda}(X\backslash A_\alpha),\quad X\backslash\bigcap_{\alpha\in\Lambda}A_\alpha = \bigcup_{\alpha\in\Lambda}(X\backslash A_\alpha),\tag{1.2.2}$$

$$f^{-1}\Bigl(\bigcup_{\alpha\in\Lambda}A_\alpha\Bigr) = \bigcup_{\alpha\in\Lambda}f^{-1}(A_\alpha),\quad f^{-1}\Bigl(\bigcap_{\alpha\in\Lambda}A_\alpha\Bigr) = \bigcap_{\alpha\in\Lambda}f^{-1}(A_\alpha).\tag{1.2.3}$$

PROOF. Let $x\in X\backslash\bigcup_{\alpha\in\Lambda}A_\alpha$, i.e., $x\notin A_\alpha$ for all $\alpha\in\Lambda$. The latter is equivalent to the inclusion $x\in\bigcap_{\alpha\in\Lambda}(X\backslash A_\alpha)$. Other relationships are proved in a similar manner. □

1.2.9. Corollary. *Let $\mathcal{A}$ be a σ-algebra of subsets of a set X and f an arbitrary mapping from a set E to X. Then the class $f^{-1}(\mathcal{A})$ of all sets of the form $f^{-1}(A)$, where $A\in\mathcal{A}$, is a σ-algebra in E.*

In addition, for an arbitrary σ-algebra $\mathcal{B}$ of subsets of E, the class of sets $\{A\subset X\colon\ f^{-1}(A)\in\mathcal{B}\}$ is a σ-algebra. Furthermore, for any class of sets $\mathcal{F}$ in X, one has $\sigma\bigl(f^{-1}(\mathcal{F})\bigr) = f^{-1}\bigl(\sigma(\mathcal{F})\bigr)$.

PROOF. The first two assertions are clear from the lemma. Since the class $f^{-1}\bigl(\sigma(\mathcal{F})\bigr)$ is a σ-algebra by the first assertion, we obtain the inclusion $\sigma\bigl(f^{-1}(\mathcal{F})\bigr)\subset f^{-1}\bigl(\sigma(\mathcal{F})\bigr)$. Finally, by the second assertion, we have $f^{-1}\bigl(\sigma(\mathcal{F})\bigr)\subset\sigma\bigl(f^{-1}(\mathcal{F})\bigr)$ because $f^{-1}(\mathcal{F})\subset\sigma\bigl(f^{-1}(\mathcal{F})\bigr)$. □

Simple examples show that the class $f(\mathcal{B})$ of all sets of the form $f(B)$, where $B\in\mathcal{B}$, is not always an algebra.

1.2.10. Definition. *The Borel σ-algebra of $\mathbb{R}^n$ is the σ-algebra $\mathcal{B}(\mathbb{R}^n)$ generated by all open sets. The sets in $\mathcal{B}(\mathbb{R}^n)$ are called Borel sets. For any set $E\subset\mathbb{R}^n$, let $\mathcal{B}(E)$ denote the class of all sets of the form $E\cap B$, where $B\in\mathcal{B}(\mathbb{R}^n)$.*

The class $\mathcal{B}(E)$ can also be defined as the σ-algebra generated by the intersections of E with open sets in $\mathbb{R}^n$. This is clear from the following: if the latter σ-algebra is denoted by $\mathcal{E}$, then the family of all sets $B \in \mathcal{B}(\mathbb{R}^n)$ such that $B \cap E \in \mathcal{E}$ is a σ-algebra containing all open sets, i.e., it coincides with $\mathcal{B}(\mathbb{R}^n)$. The sets in $\mathcal{B}(E)$ are called Borel sets of the space E and $\mathcal{B}(E)$ is called the Borel σ-algebra of the space E. One should keep in mind that such sets may not be Borel in $\mathbb{R}^n$ unless, of course, E itself is Borel in $\mathbb{R}^n$. For example, one always has $E \in \mathcal{B}(E)$, since $E \cap \mathbb{R}^n = E$.

It is clear that $\mathcal{B}(\mathbb{R}^n)$ is also generated by the class of all closed sets.

1.2.11. Lemma. *The Borel σ-algebra of the real line is generated by any of the following classes of sets*:

(i) *the collection of all intervals*;

(ii) *the collection of all intervals with rational endpoints*;

(iii) *the collection of all rays of the form $(-\infty, c)$, where c is rational*;

(iv) *the collection of all rays of the form $(-\infty, c]$, where c is rational*;

(v) *the collection of rays of the form $(c, +\infty)$, where c rational*;

(vi) *the collection of all rays of the form $[c, +\infty)$, where c is rational. Finally, the same is true if in place of rational numbers one takes points of any everywhere dense set.*

Proof. It is clear that all the sets indicated above are Borel, since they are either open or closed. Therefore, the σ-algebras generated by the corresponding families are contained in $\mathcal{B}(\mathbb{R}^1)$. Since every open set on the real line is the union of an at most countable collection of intervals, it suffices to show that any interval (a, b) is contained in the σ-algebras corresponding to the classes (i)–(vi). This follows from the fact that (a, b) is the union of intervals of the form (a_n, b_n), where a_n and b_n are rational, and also is the union of intervals of the form $[a_n, b_n)$ with rational endpoints, whereas such intervals belong to the σ-algebra generated by the rays $(-\infty, c)$, since they can be written as differences of rays. In a similar manner, the differences of the rays of the form (c, ∞) give the intervals $(a_n, b_n]$, from which by means of unions one constructs the intervals (a, b). □

It is clear from the proof that the Borel σ-algebra is generated by the closed intervals with rational endpoints. It is seen from this, by the way, that disjoint classes of sets may generate one and the same σ-algebra.

1.2.12. Example. The collection of all single-point sets in a space X generates the σ-algebra consisting of all sets that are either at most countable or have at most countable complements. In addition, this σ-algebra is strictly smaller than the Borel one if $X = \mathbb{R}^1$.

Proof. Denote by $\mathcal{A}$ the family of all sets $A \subset X$ such that either A is at most countable or $X \backslash A$ is at most countable. Let us verify that $\mathcal{A}$ is a σ-algebra. Since X is contained in $\mathcal{A}$ and $\mathcal{A}$ is closed under complementation, it suffices to show that $A := \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ whenever $A_n \in \mathcal{A}$. If all A_n are at most countable, this is obvious. Suppose that among the sets A_n there is at least one set A_{n_1} whose complement is at most countable. The complement of A is contained in the complement of A_{n_1}, hence is at most countable as well, i.e., $A \in \mathcal{A}$. All one-point sets belong to $\mathcal{A}$, hence the σ-algebra $\mathcal{A}_0$ generated by them is contained in $\mathcal{A}$. On the other hand, it is clear that any set in $\mathcal{A}$ is an element of $\mathcal{A}_0$, whence it follows that $\mathcal{A}_0 = \mathcal{A}$. □

Let us give definitions of several other classes of sets employed in measure theory.

1.2.13. Definition. (i) *A family $\mathcal{R}$ of subsets of a set X is called a ring if it contains the empty set and the sets $A \cap B$, $A \cup B$ and $A \backslash B$ belong to $\mathcal{R}$ for all $A, B \in \mathcal{R}$;*

(ii) *A family $\mathcal{S}$ of subsets of a set X is called a semiring if it contains the empty set, $A \cap B \in \mathcal{S}$ for all $A, B \in \mathcal{S}$ and, for every pair of sets $A, B \in \mathcal{S}$ with $A \subset B$, the set $B \backslash A$ is the union of finitely many disjoint sets in $\mathcal{S}$. If $X \in \mathcal{S}$, then $\mathcal{S}$ is called a semialgebra*;

(iii) *A ring is called a σ-ring if it is closed with respect to countable unions. A ring is called a δ-ring if it is closed with respect to countable intersections.*

As an example of a ring that is not an algebra, let us mention the collection of all bounded sets on the real line. The family of all intervals in the interval $[a, b]$ gives an example of a semiring that is not a ring. According to the following lemma, the collection of all finite unions of elements of a semiring is a ring (called the ring generated by the given semiring). It is clear that this is the minimal ring containing the given semiring.

1.2.14. Lemma. *For any semiring $\mathcal{S}$, the collection of all finite unions of sets in $\mathcal{S}$ forms a ring $\mathcal{R}$. Every set in $\mathcal{R}$ is a finite union of pairwise disjoint sets in $\mathcal{S}$. If $\mathcal{S}$ is a semialgebra, then $\mathcal{R}$ is an algebra.*

Proof. It is clear that the class $\mathcal{R}$ admits finite unions. Suppose that $A = A_1 \cup \cdots \cup A_n$, $B = B_1 \cup \cdots \cup B_k$, where $A_i, B_j \in \mathcal{S}$. Then we have $A \cap B = \bigcup_{i \le n, j \le k} A_i \cap B_j \in \mathcal{R}$. Hence $\mathcal{R}$ admits finite intersections. In addition,

$$A \backslash B = \bigcup_{i=1}^{n} \Bigl(A_i \backslash \bigcup_{j=1}^{k} B_j \Bigr) = \bigcup_{i=1}^{n} \bigcap_{j=1}^{k} (A_i \backslash B_j).$$

Since the set $A_i \backslash B_j = A_i \backslash (A_i \cap B_j)$ is a finite union of sets in $\mathcal{S}$, one has $A \backslash B \in \mathcal{R}$. Clearly, A can be written as a union of a finitely many disjoint sets in $\mathcal{S}$ because $\mathcal{S}$ is closed with respect to intersections. The last claim of the lemma is obvious. □

Note that for any σ-algebra $\mathcal{B}$ in a space X and any set $A \subset X$, the class $\mathcal{B}_A := \{B \cap A\colon\ B \in \mathcal{B}\}$ is a σ-algebra in the space A. This σ-algebra is called the trace σ-algebra.

1.3. Additivity and countable additivity of measures

Functions with values in $(-\infty, +\infty)$ will be called real or real-valued. In the cases where we discuss functions with values in the extended real line $[-\infty, +\infty]$, this will always be specified.

1.3.1. Definition. *A real-valued set function μ defined on a class of sets $\mathcal{A}$ is called additive (or finitely additive) if*

$$\mu\Bigl(\bigcup_{i=1}^{n} A_i\Bigr) = \sum_{i=1}^{n} \mu(A_i) \tag{1.3.1}$$

for all n and all disjoint sets $A_1, \ldots, A_n \in \mathcal{A}$ such that $\bigcup_{i=1}^{n} A_i \in \mathcal{A}$.

In the case where $\mathcal{A}$ is closed with respect to finite unions, the finite additivity is equivalent to the equality

$$\mu(A \cup B) = \mu(A) + \mu(B) \tag{1.3.2}$$

for all disjoint sets $A, B \in \mathcal{A}$.

If the domain of definition of an additive real-valued set function μ contains the empty set $\varnothing$, then $\mu(\varnothing) = 0$. In particular, this is true for any additive set function on a ring or an algebra.

It is also useful to consider the property of *subadditivity* (also called the *semiadditivity*):

$$\mu\Bigl(\bigcup_{i=1}^{n} A_i\Bigr) \le \sum_{i=1}^{n} \mu(A_i) \tag{1.3.3}$$

for all $A_i \in \mathcal{A}$ with $\bigcup_{i=1}^{n} A_i \in \mathcal{A}$. Any additive nonnegative set function on an algebra is subadditive (see below).

1.3.2. Definition. *A real-valued set function μ on a class of sets $\mathcal{A}$ is called countably additive if*

$$\mu\Bigl(\bigcup_{n=1}^{\infty} A_n\Bigr) = \sum_{n=1}^{\infty} \mu(A_n) \tag{1.3.4}$$

for all pairwise disjoint sets A_n in $\mathcal{A}$ such that $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$. A countably additive set function defined on an algebra is called a measure.

It is readily seen from the definition that the series in (1.3.4) converges absolutely because its sum is independent of rearrangements of its terms.

1.3.3. Proposition. *Let μ be an additive real set function on an algebra (or a ring) of sets $\mathcal{A}$. Then the following conditions are equivalent:*

(i) *the function μ is countably additive,*

(ii) *the function μ is continuous at zero in the following sense: if $A_n \in \mathcal{A}$, $A_{n+1} \subset A_n$ for all $n \in \mathbb{N}$ and $\bigcap_{n=1}^{\infty} A_n = \varnothing$, then*

$$\lim_{n\to\infty} \mu(A_n) = 0, \tag{1.3.5}$$

(iii) *the function* μ *is continuous from below, i.e., if* $A_n \in \mathcal{A}$ *are such that* $A_n \subset A_{n+1}$ *for all* $n \in \mathbb{N}$ *and* $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$*, then*

$$\mu\Bigl(\bigcup_{n=1}^{\infty} A_n\Bigr) = \lim_{n\to\infty} \mu(A_n). \tag{1.3.6}$$

PROOF. (i) Let μ be countably additive and let the sets $A_n \in \mathcal{A}$ decrease monotonically to the empty set. Set $B_n = A_n \backslash A_{n+1}$. The sets B_n belong to $\mathcal{A}$ and are disjoint and their union is A_1. Hence the series $\sum_{n=1}^{\infty} \mu(B_n)$ converges. Then $\sum_{n=N}^{\infty} \mu(B_n)$ tends to zero as $N \to \infty$, but the sum of this series is $\mu(A_N)$, since $\bigcup_{n=N}^{\infty} B_n = A_N$. Hence we arrive at condition (ii).

Suppose now that condition (ii) is fulfilled. Let $\{B_n\}$ be a sequence of pairwise disjoint sets in $\mathcal{A}$ whose union B is an element of $\mathcal{A}$ as well. Set $A_n = B \backslash \bigcup_{k=1}^{n} B_k$. It is clear that $\{A_n\}$ is a sequence of monotonically decreasing sets in $\mathcal{A}$ with the empty intersection. By hypothesis, $\mu(A_n) \to 0$. By the finite additivity this means that $\sum_{k=1}^{n} \mu(B_k) \to \mu(B)$ as $n \to \infty$. Hence μ is countably additive. Clearly, (iii) follows from (ii), for if the sets $A_n \in \mathcal{A}$ increase monotonically and their union is the set $A \in \mathcal{A}$, then the sets $A \backslash A_n \in \mathcal{A}$ decrease monotonically to the empty set. Finally, by the finite additivity (iii) yields the countable additivity of μ. □

The reader is warned that there is no such equivalence for semialgebras (see Exercise 1.12.75).

1.3.4. Definition. *A countably additive measure* μ *on a* σ*-algebra of subsets of a space* X *is called a probability measure if* $\mu \geq 0$ *and* $\mu(X) = 1$.

1.3.5. Definition. *A triple* $(X, \mathcal{A}, \mu)$ *is called a measure space if* μ *is a nonnegative measure on a* σ*-algebra* $\mathcal{A}$ *of subset of a set* X*. If* μ *is a probability measure, then* $(X, \mathcal{A}, \mu)$ *is called a probability space.*

Nonnegative not identically zero measures are called *positive measures*.

Additive set functions are also called additive measures, but to simplify the terminology we use the term measure only for *countably additive measures on algebras or rings*. Countably additive measures are also called σ-additive measures.

1.3.6. Definition. *A measure defined on the Borel* σ*-algebra of the whole space* $\mathbb{R}^n$ *or its subset is called a Borel measure.*

It is clear that if $\mathcal{A}$ is an algebra, then the additivity is just equality (1.3.2) for arbitrary disjoint sets in $\mathcal{A}$. Similarly, if $\mathcal{A}$ is a σ-algebra, then the countable additivity is equality (1.3.4) for arbitrary sequences of disjoint sets in $\mathcal{A}$. The above given formulations are convenient for two reasons. First, the validity of the corresponding equalities is required only for those collections of sets for which both parts make sense. Second, as we shall see later, under natural hypotheses, additive (or countably additive) set functions admit additive (respectively, countably additive) extensions to larger classes of sets that admit unions of the corresponding type.

1.3.7. Example. Let $\mathcal{A}$ be the algebra of sets $A \subset \mathbb{N}$ such that either A or $\mathbb{N}\backslash A$ is finite. For finite A, let $\mu(A) = 0$, and for A with a finite complement let $\mu(A) = 1$. Then μ is an additive, but not countably additive set function.

PROOF. It is clear that $\mathcal{A}$ is indeed an algebra. Relation (1.3.2) is obvious for disjoint sets A and B if A is finite. Finally, A and B in $\mathcal{A}$ cannot be infinite simultaneously being disjoint. If μ were countably additive, we would have had $\mu(\mathbb{N}) = \sum_{n=1}^{\infty} \mu(\{n\}) = 0$. □

There exist additive, but not countably additive set functions on σ-algebras (see Example 1.12.28). The simplest countably additive set function is identically zero. Another example: let X be a nonempty set and let $a \in X$; Dirac's measure δ_a at the point a is defined as follows: for every $A \subset X$, $\delta_a(A) = 1$ if $a \in A$ and $\delta_a(A) = 0$ otherwise. Let us give a slightly less trivial example.

1.3.8. Example. Let $\mathcal{A}$ be the σ-algebra of all subsets of $\mathbb{N}$. For every set $A = \{n_k\}$, let $\mu(A) = \sum_k 2^{-n_k}$. Then μ is a measure on $\mathcal{A}$.

In order to construct less trivial examples (say, Lebesgue measure), we need auxiliary technical tools discussed in the next section.

Note several simple properties of additive and countably additive set functions.

1.3.9. Proposition. *Let μ be a nonnegative additive set function on an algebra or a ring $\mathcal{A}$.*

(i) *If $A, B \in \mathcal{A}$ and $A \subset B$, then $\mu(A) \le \mu(B)$.*

(ii) *For any collection $A_1, \ldots, A_n \in \mathcal{A}$ one has*

$$\mu\Bigl(\bigcup_{i=1}^{n} A_i\Bigr) \le \sum_{i=1}^{n} \mu(A_i).$$

(iii) *The function μ is countably additive precisely when in addition to the additivity it is countably subadditive in the following sense: for any sequence $\{A_n\} \subset \mathcal{A}$ with $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ one has*

$$\mu\Bigl(\bigcup_{n=1}^{\infty} A_n\Bigr) \le \sum_{n=1}^{\infty} \mu(A_n).$$

PROOF. Assertion (i) follows, since $\mu(B\backslash A) \ge 0$. Assertion (ii) is easily verified by induction taking into account the nonnegativity of μ and the relation $\mu(A \cup B) = \mu(A\backslash B) + \mu(B\backslash A) + \mu(A \cap B)$.

If μ is countably additive and the union of sets $A_n \in \mathcal{A}$ belongs to $\mathcal{A}$, then according to Proposition 1.3.3 one has

$$\mu\Bigl(\bigcup_{i=1}^{n} A_i\Bigr) \to \mu\Bigl(\bigcup_{i=1}^{\infty} A_i\Bigr),$$

which by (ii) gives the estimate indicated in (iii). Finally, such an estimate combined with the additivity yields the countable additivity. Indeed, let B_n be pairwise disjoint sets in $\mathcal{A}$ whose union B belongs to $\mathcal{A}$ as well. Then for any $n \in \mathbb{N}$ we have

$$\sum_{k=1}^{n} \mu(B_k) = \mu\Bigl(\bigcup_{k=1}^{n} B_k\Bigr) \le \mu(B) \le \sum_{k=1}^{\infty} \mu(B_k),$$

whence it follows that $\sum_{k=1}^{\infty} \mu(B_k) = \mu(B)$. □

1.3.10. Proposition. *Let $\mathcal{A}_0$ be a semialgebra* (*see Definition* 1.2.13). *Then every additive set function μ on $\mathcal{A}_0$ uniquely extends to an additive set function on the algebra $\mathcal{A}$ generated by $\mathcal{A}_0$* (*i.e., the family of all finite unions of sets in $\mathcal{A}_0$*). *This extension is countably additive provided that μ is countably additive on $\mathcal{A}_0$. The same is true in the case of a semiring $\mathcal{A}$ and the ring generated by it.*

Proof. By Lemma 1.2.14 the collection of all finite unions of elements of $\mathcal{A}_0$ is an algebra (or a ring when $\mathcal{A}_0$ is a semiring). It is clear that any set in $\mathcal{A}$ can be represented as a union of disjoint elements of $\mathcal{A}_0$. Set

$$\mu(A) = \sum_{i=1}^{n} \mu(A_i)$$

if $A_i \in \mathcal{A}_0$ are pairwise disjoint and their union is A. The indicated extension is obviously additive, but we have to verify that it is well-defined, i.e., is independent of partitioning A into parts in $\mathcal{A}_0$. Indeed, if $B_1, \ldots, B_m$ are pairwise disjoint sets in $\mathcal{A}_0$ whose union is A, then by the additivity of μ on the algebra $\mathcal{A}_0$ one has the equality $\mu(A_i) = \sum_{j=1}^{m} \mu(A_i \cap B_j)$, $\mu(B_j) = \sum_{i=1}^{n} \mu(A_i \cap B_j)$, whence the desired conclusion follows. Let us verify the countable additivity of the indicated extension in the case of the countable additivity on $\mathcal{A}_0$. Let $A, A_n \in \mathcal{A}$, $A = \bigcup_{n=1}^{\infty} A_n$ be such that $A_n \cap A_k = \varnothing$ if $n \neq k$. Then

$$A = \bigcup_{j=1}^{N} B_j, \quad A_n = \bigcup_{i=1}^{N_n} B_{n,i},$$

where $B_j, B_{n,i} \in \mathcal{A}_0$. Set $C_{n,i,j} := B_{n,i} \cap B_j$. The sets $C_{n,i,j}$ are pairwise disjoint and

$$B_j = \bigcup_{n=1}^{\infty} \bigcup_{i=1}^{N_n} C_{n,i,j}, \quad B_{n,i} = \bigcup_{j=1}^{N} C_{n,i,j}.$$

By the countable additivity of μ on $\mathcal{A}_0$ we have

$$\mu(B_j) = \sum_{n=1}^{\infty} \sum_{i=1}^{N_n} \mu(C_{n,i,j}), \quad \mu(B_{n,i}) = \sum_{j=1}^{N} \mu(C_{n,i,j}),$$

and by the definition of μ on $\mathcal{A}$ one has the following equality:

$$\mu(A) = \sum_{j=1}^{N} \mu(B_j), \quad \mu(A_n) = \sum_{i=1}^{N_n} \mu(B_{n,i}).$$

We obtain from these equalities that $\mu(A) = \sum_{n=1}^{\infty} \mu(A_n)$, since both quantities equal the sum of all $\mu(C_{n,i,j})$. That it is possible to interchange the summations in n and j is obvious from the fact that the series in n converge and the sums in j and i are finite. □

1.4. Compact classes and countable additivity

In this section, we give a sufficient condition for the countable additivity, which is satisfied for most of the measures encountered in real applications.

1.4.1. Definition. *A family $\mathcal{K}$ of subsets of a set X is called a compact class if, for any sequence K_n of its elements with $\bigcap_{n=1}^{\infty} K_n = \varnothing$, there exists N such that $\bigcap_{n=1}^{N} K_n = \varnothing$.*

The terminology is explained by the following basic example.

1.4.2. Example. An arbitrary family of compact sets in $\mathbb{R}^n$ (more generally, in a topological space) is a compact class.

PROOF. Indeed, let K_n be compact sets whose intersection is empty. Suppose that for every n the set $E_n = \bigcap_{i=1}^{n} K_i$ contains some element x_n. We may assume that no element of the sequence $\{x_n\}$ is repeated infinitely often, since otherwise it is a common element of all E_n. By the compactness of K_1 there exists a point x each neighborhood of which contains infinitely many elements of the sequence $\{x_n\}$. All sets E_n are compact and $x_i \in E_n$ whenever $i \geq n$, hence the point x belongs to all E_n, which is a contradiction. □

Note that some authors call the above-defined compact classes countably compact or semicompact and in the definition of compact classes require the following stronger property: if the intersection of a (possibly uncountable) collection of sets in $\mathcal{K}$ is empty, then the intersection of some its finite subcollection is empty as well. See Exercise 1.12.105 for an example distinguishing the two properties. Although such a terminology is more consistent from the point of view of topology (see Exercise 6.10.66 in Chapter 6), we shall not follow it.

1.4.3. Theorem. *Let μ be a nonnegative additive set function on an algebra $\mathcal{A}$. Suppose that there exists a compact class $\mathcal{K}$ approximating μ in the following sense: for every $A \in \mathcal{A}$ and every $\varepsilon > 0$, there exist $K_\varepsilon \in \mathcal{K}$ and $A_\varepsilon \in \mathcal{A}$ such that $A_\varepsilon \subset K_\varepsilon \subset A$ and $\mu(A \backslash A_\varepsilon) < \varepsilon$. Then μ is countably additive. In particular, this is true if the compact class $\mathcal{K}$ is contained in $\mathcal{A}$ and for any $A \in \mathcal{A}$ one has the equality*

$$\mu(A) = \sup_{K \subset A,\ K \in \mathcal{K}} \mu(K).$$

Proof. Suppose that the sets $A_n \in \mathcal{A}$ are decreasing and their intersection is empty. Let us show that $\mu(A_n) \to 0$. Let us fix $\varepsilon > 0$. By hypothesis, there exist $K_n \in \mathcal{K}$ and $B_n \in \mathcal{A}$ such that $B_n \subset K_n \subset A_n$ and $\mu(A_n \backslash B_n) < \varepsilon 2^{-n}$. It is clear that $\bigcap_{n=1}^{\infty} K_n \subset \bigcap_{n=1}^{\infty} A_n = \varnothing$. By the definition of a compact class, there exists N such that $\bigcap_{n=1}^{N} K_n = \varnothing$. Then $\bigcap_{n=1}^{N} B_n = \varnothing$. Note that one has

$$A_N = \bigcap_{n=1}^{N} A_n \subset \bigcup_{n=1}^{N} (A_n \backslash B_n).$$

Indeed, let $x \in A_N$, i.e., $x \in A_n$ for all $n \leq N$. If x does not belong to $\bigcup_{n=1}^{N} (A_n \backslash B_n)$, then $x \notin A_n \backslash B_n$ for all $n \leq N$. Then $x \in B_n$ for every $n \leq N$, whence we obtain $x \in \bigcap_{n=1}^{N} B_n$, which is a contradiction. The above proved equality yields the estimate

$$\mu(A_N) \leq \sum_{n=1}^{N} \mu(A_n \backslash B_n) \leq \sum_{n=1}^{N} \varepsilon 2^{-n} \leq \varepsilon.$$

Hence $\mu(A_n) \to 0$, which implies the countable additivity of μ. □

1.4.4. Example. Let I be an interval in $\mathbb{R}^1$, $\mathcal{A}$ the algebra of finite unions of intervals in I (closed, open and half-open). Then the usual length λ_1, which assigns the value $b - a$ to the interval with the endpoints a and b and extends by additivity to their finite disjoint unions, is countably additive on the algebra $\mathcal{A}$.

Proof. Finite unions of closed intervals form a compact class and approximate from within finite unions of arbitrary intervals. □

1.4.5. Example. Let I be a cube in $\mathbb{R}^n$ of the form $[a, b]^n$ and let $\mathcal{A}$ be the algebra of finite unions of the parallelepipeds in I that are products of intervals in $[a, b]$. Then the usual volume λ_n is countably additive on $\mathcal{A}$. We call λ_n *Lebesgue measure.*

Proof. As in the previous example, finite unions of closed parallelepipeds form a compact approximating class. □

It is shown in Theorem 1.12.5 below that the compactness property can be slightly relaxed.

The previous results justify the introduction of the following concept.

1.4.6. Definition. *Let m be a nonnegative function on a class $\mathcal{E}$ of subsets of a set X and let $\mathcal{P}$ be a class of subsets of X, too. We say that $\mathcal{P}$ is an approximating class for m if, for every $E \in \mathcal{E}$ and every $\varepsilon > 0$, there exist $P_\varepsilon \in \mathcal{P}$ and $E_\varepsilon \in \mathcal{E}$ such that $E_\varepsilon \subset P_\varepsilon \subset E$ and $|m(E) - m(E_\varepsilon)| < \varepsilon$.*

1.4.7. Remark. (i) The reasoning in Theorem 1.4.3 actually proves the following assertion. Let μ be a nonnegative additive set function on an algebra $\mathcal{A}$ and let $\mathcal{A}_0$ be a subalgebra in $\mathcal{A}$. Suppose that there exists a compact class $\mathcal{K}$ approximating μ on $\mathcal{A}_0$ with respect to $\mathcal{A}$ in the following sense: for any $A \in \mathcal{A}_0$ and any $\varepsilon > 0$, there exist $K_\varepsilon \in \mathcal{K}$ and $A_\varepsilon \in \mathcal{A}$ such that $A_\varepsilon \subset K_\varepsilon \subset A$ and $\mu(A\backslash A_\varepsilon) < \varepsilon$. Then μ is countably additive on $\mathcal{A}_0$.

(ii) The compact class $\mathcal{K}$ in Theorem 1.4.3 need not be contained in $\mathcal{A}$. For example, if $\mathcal{A}$ is the algebra generated by all intervals in $[0, 1]$ with rational endpoints and μ is Lebesgue measure, then the class $\mathcal{K}$ of all finite unions of closed intervals with irrational endpoints is approximating for μ and has no intersection with $\mathcal{A}$. However, it will be shown in §1.12(ii) that one can always replace $\mathcal{K}$ by a compact class $\mathcal{K}'$ that is contained in $\sigma(\mathcal{A})$ and approximates the countably additive extension of μ on $\sigma(\mathcal{A})$. It is worth noting that there exists a countably additive extension of μ to the σ-algebra generated by $\mathcal{A}_0$ and $\mathcal{K}$ (see Theorem 1.12.34).

Note that so far in the considered examples we have been concerned with the countable additivity on algebras. However, as we shall see below, any countably additive measure on an algebra automatically extends (in a unique way) to a countably additive measure on the σ-algebra generated by this algebra.

We shall see in Chapter 7 that the class of measures possessing a compact approximating class is very large (so that it is not easy even to construct an example of a countably additive measure without compact approximating classes). Thus, the described sufficient condition of countable additivity has a very universal character. Here we only give the following result.

1.4.8. Theorem. *Let μ be a nonnegative countably additive measure on the Borel σ-algebra $\mathcal{B}(\mathbb{R}^n)$ in the space $\mathbb{R}^n$. Then, for any Borel set $B \subset \mathbb{R}^n$ and any $\varepsilon > 0$, there exist an open set U_ε and a compact set K_ε such that $K_\varepsilon \subset B \subset U_\varepsilon$ and $\mu(U_\varepsilon\backslash K_\varepsilon) < \varepsilon$.*

PROOF. Let us show that for any $\varepsilon > 0$ there exists a closed set $F_\varepsilon \subset B$ such that

$$\mu(B\backslash F_\varepsilon) < \varepsilon/2.$$

Then, by the countable additivity of μ, the set F_ε itself can be approximated from within up to $\varepsilon/2$ by $F_\varepsilon \cap U$, where U is a closed ball of a sufficiently large radius. Denote by $\mathcal{A}$ the class of all sets $A \in \mathcal{B}(\mathbb{R}^n)$ such that, for any $\varepsilon > 0$, there exist a closed set F_ε and an open set U_ε with $F_\varepsilon \subset A \subset U_\varepsilon$ and $\mu(U_\varepsilon\backslash F_\varepsilon) < \varepsilon$. Every closed set A belongs to $\mathcal{A}$, since one can take for F_ε the set A itself, and for U_ε one can take some open δ-neighborhood A^δ of the set A, i.e., the union of all open balls of radius δ with centers at the points in A. When δ is decreasing to zero, the open sets A^δ are decreasing to A, hence their measures approach the measure of A. Let us show that $\mathcal{A}$ is a σ-algebra. If this is done, then the theorem is proven, for the closed sets generate the Borel σ-algebra. By construction, the class $\mathcal{A}$ is closed with respect to the operation of complementation. Hence it remains to verify the stability of $\mathcal{A}$ with respect to countable unions. Let $A_j \in \mathcal{A}$ and let $\varepsilon > 0$. Then there exist a closed set F_j and an open set U_j such that $F_j \subset A_j \subset U_j$ and $\mu(U_j\backslash F_j) < \varepsilon 2^{-j}$, $j \in \mathbb{N}$.