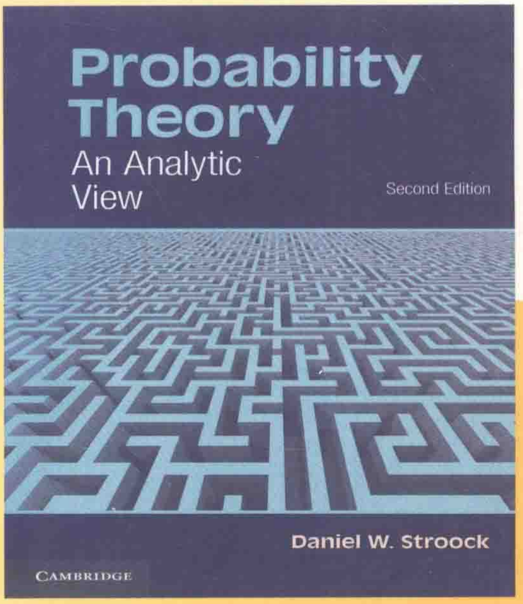
Daniel W. Stroock

Probability Theory

An Analytic View

Second Edition

概率论 第2版



CAMBRIDGE

世界图书出版公司
www.wpcbj.com.cn

Probability Theory

An Analytic View

Second Edition

Daniel W. Stroock
Massachusetts Institute of Technology

CAMBRIDGE
UNIVERSITY PRESS

图书在版编目（CIP）数据

概率论：第2版 = Probability Theory：An Analytic View Second Edition：英文/（美）斯特鲁克（Stroock，D. W.）著.—影印本.—北京：世界图书出版公司北京公司，2015. 11
ISBN 978-7-5192-0534-8

Ⅰ.①概… Ⅱ.①斯… Ⅲ.①概率论—英文 Ⅳ.①O211

中国版本图书馆 CIP 数据核字（2015）第 288006 号

Probability Theory
An Analytic View Second Edition
概率论 第2版

著　　者： Daniel W. Stroock
责任编辑： 刘　慧　岳利青
装帧设计： 任志远

出版发行： 世界图书出版公司北京公司
地　　址： 北京市东城区朝内大街 137 号
邮　　编： 100010
电　　话： 010 - 64038355（发行）　64015580（客服）　64033507（总编室）
网　　址： http://www. wpcbj. com. cn
邮　　箱： wpcbjst@ vip. 163. com
销　　售： 新华书店
印　　刷： 三河市国英印务有限公司
开　　本： 787mm × 1092mm　1/16
印　　张： 35
字　　数： 663 千
版　　次： 2016 年 7 月第 1 版　　2016 年 7 月第 1 次印刷
版权登记： 01-2016-0375
ISBN 978-7-5192-0534-8　　　　**定价：** 119.00 元

Probability Theory

An Analytic View, Second Edition

This second edition of Daniel W. Stroock's text is suitable for first-year graduate students with a good grasp of introductory undergraduate probability. It provides a reasonably thorough introduction to modern probability theory with an emphasis on the mutually beneficial relationship between probability theory and analysis. It includes more than 750 exercises and offers new material on Levy processes, large deviations theory, Gaussian measures on a Banach space, and the relationship between a Wiener measure and partial differential equations.

The first part of the book deals with independent random variables, Central Limit phenomena, the general theory of weak convergence and several of its applications, as well as elements of both the Gaussian and Markovian theories of measures on function space. The introduction of conditional expectation values is postponed until the second part of the book, where it is applied to the study of martingales. This part also explores the connection between martingales and various aspects of classical analysis and the connections between a Wiener measure and classical potential theory.

Dr. Daniel W. Stroock is the Simons Professor of Mathematics Emeritus at the Massachusetts Institute of Technology. He has published many articles and is the author of six books, most recently *Partial Differential Equations for Probabilists* (2008).

Probability Theory: An Analytic View Second Edition (978-0-521-13250-3) by Daniel W. Stroock, published by Cambridge University Press 2011

This reprint edition for the People's Republic of China is published by arrangement with the Press Syndicate of the University of Cambridge, Cambridge, United Kingdom.

This book is dedicated to my teachers:

M. Kac, H.P. McKean, Jr., and S.R.S. Varadhan

Preface

From the Preface to the First Edition

When writing a graduate level mathematics book during the last decade of the twentieth century, one probably ought not inquire too closely into one's motivation. In fact, if ones own pleasure from the exercise is not sufficient to justify the effort, then one should seriously consider dropping the project. Thus, to those who (either before or shortly after opening it) ask *for whom was this book written*, my pale answer is *me*; and, for this reason, I thought that I should preface this preface with an explanation of who I am and what were the peculiar educational circumstances that eventually gave rise to this somewhat peculiar book.

My own introduction to probability theory began with a private lecture from H.P. McKean, Jr. At the time, I was a (more accurately, *the*) graduate student of mathematics at what was then called The Rockefeller Institute for Biological Sciences. My official mentor there was M. Kac, whom I had cajoled into becoming my adviser after a year during which I had failed to insert even one micro-electrode into the optic nerves of innumerable limuli. However, as I soon came to realize, Kac had accepted his role on the condition that it would not become a burden. In particular, he had no intention of wasting much of his own time on a reject from the neurophysiology department. On the other hand, he was most generous with the time of his younger associates, and that is how I wound up in McKean's office. Never one to bore his listeners with a lot of dull preliminaries, McKean launched right into a wonderfully lucid explanation of P. Lévy's interpretation of the infinitely divisible laws. I have to admit that my appreciation of the lucidity of his lecture arrived nearly a decade after its delivery, and I can only hope that my reader will reserve judgment of my own presentation for an equal length of time.

In spite of my perplexed state at the end of McKean's lecture, I was sufficiently intrigued to delve into the readings that he suggested at its conclusion. Knowing that the only formal mathematics courses that I would be taking during my graduate studies would be given at N.Y.U. and guessing that those courses would be oriented toward partial differential equations, McKean directed me to material which would help me understand the connections between partial differential equations and probability theory. In particular, he suggested that I start with the, then recently translated, two articles by E.B. Dynkin which had appeared originally in the famous 1956 volume of *Teoriya Veroyatnostei i ee Primeneniya*. Dynkin's articles turned out to be a godsend. They were beautifully crafted to tell the reader enough so that he could understand the ideas and not so much that he would become bored by them. In addition, they gave me an introduction to a host of ideas and techniques (e.g., stopping times and the strong Markov property), all of which Kac himself consigned to the category of overelaborated measure theory. In fact, it would be reasonable to say that my thesis was simply the application of techniques which I picked up from Dynkin to a problem that I picked up by reading some notes by Kac. Of course, along the way I profited immeasurably from continued contact with McKean, a large number of courses at N.Y.U. (particularly ones taught by M. Donsker, F. John, and L. Nirenberg), and my increasingly animated conversations with S.R.S. Varadhan.

As I trust the preceding description makes clear, my graduate education was anything but deprived; I had ready access to some of the very best analysts of the day. On the other hand, I never had a *proper* introduction to my field, probability theory. The first time that I ever summed independent random variables was when I was summing them in front of a class at N.Y.U. Thus, although I now admire the magnificent body of mathematics created by A.N. Kolmogorov, P. Lévy, and the other twentieth-century heroes of the field, I am not a *dyed-in-the-wool* probabilist (i.e., what Donsker would have called a true *coin-tosser*). In particular, I have never been able to develop sufficient sensitivity to the distinction between a *proof* and a *probabilistic proof*. To me, a proof is clearly *probabilistic* only if its punch-line comes down to an argument like $P(A) \leq P(B)$ because $A \subseteq B$; and there are breathtaking examples of such arguments. However, to base an entire book on these examples would require a level of genius that I do not possess. In fact, I myself enjoy probability theory best when it is inextricably interwoven with other branches of mathematics and not when it is presented as an entity unto itself. For this reason, the reader should not be surprised to discover that he finds some of the material presented in this book *does not belong here*; but I hope that he will make an effort to figure out why I disagree with him.

Preface to the Second Edition

My favorite "preface to a second edition" is the one that G.N. Watson wrote for the second edition of his famous treatise on Bessel functions. The first edition appeared in 1922, the second came out in 1941, and Watson had originally intended to stay abreast of developments and report on them in the second edition. However, in his preface to the second edition Watson admits that his interest in the topic had "waned" during the intervening years and apologizes that, as a consequence, the new edition contains less new material than he had thought it would.

My excuse for not incorporating more new material into this second edition is related to but somewhat different from Watson's. In my case, what has waned is not my interest in probability theory but instead my ability to assimilate the transformations that the subject has undergone. When I was a student, probabilists were still working out the ramifications of Kolmogorov's profound insights into the connections between probability and analysis, and I have spent my career investigating and exploiting those connections. However, about the time when the first edition of this book was published, probability theory began a return to its origins in combinatorics, a topic in which my abilities are woefully deficient. Thus, although I suspect that, for at least a decade, the most exciting developments in the field will have a strong combinatorial component, I have not attempted to prepare my readers for those developments. I repeat that my decision not to incorporate more combinatorics into this new edition in no way reflects my assessment of the direction in which probability is likely to go but instead reflects my assessment of my own inability to do justice to the beautiful combinatorial ideas that have been introduced in the recent past.

In spite of the preceding admission, I believe that the material in this book remains valuable and that, no matter how probability theory evolves, the ideas and techniques presented here will play an important role. Furthermore, I have made some substantive changes. In particular, I have given more space to infinitely divisible laws and their associated Lévy processes, both of which are now developed in $\mathbb{R}^N$ rather than just in $\mathbb{R}$. In addition, I have added an entire chapter devoted to Gaussian measures in infinite dimensions from the perspective of the Segal–Gross school. Not only have recent developments in Malliavin calculus and conformal field theory sparked renewed interest in this topic, but it seems to me that most modern texts pay either no or too little attention to this beautiful material. Missing from the new edition is the treatment of singular integrals. I included it in the first edition in the hope that it would elucidate the similarity between cancellations that underlie martingale theory, especially Burkholder's Inequality, and Calderon–Zygmund theory. I still believe that these similarities are worth thinking about, but I have decided that my explanation of them led me too far astray and was more of a distraction than a pedagogically valuable addition.

Besides those mentioned above, minor changes have been made throughout. For one thing, I have spent a lot of time correcting old errors and, undoubtedly, inserting new ones. Secondly, I have made several organizational changes as well as others that are remedial. A summary of the contents follows.

Summary

1: Chapter 1 contains a sampling of the standard, pointwise convergence theorems dealing with partial sums of independent random variables. These include the Weak and Strong Laws of Large Numbers as well as Hartman–Wintner's Law of the Iterated Logarithm. In preparation for the Law of the Iterated Logarithm, Cramér's theory of large deviations from the Law of Large Numbers is developed in § 1.4. Everything here is very standard, although I feel that my passage from the bounded to the general case of the Law of the Iterated Logarithm has been considerably smoothed by the ideas that I learned during a conversation with M. Ledoux.

2: The whole of Chapter 2 is devoted to the classical Central Limit Theorem. After an initial (and slightly flawed) derivation of the basic result via moment considerations, Lindeberg's general version is derived in § 2.1. Although Lindeberg's result has become a *sine qua non* in the writing of probability texts, the Berry–Esseen estimate has not. Indeed, until recently, the Berry–Esseen estimate required a good many somewhat tedious calculations with characteristic functions (i.e., Fourier transforms), and most recent authors seem to have decided that the rewards did not justify the effort. I was inclined to agree with them until P. Diaconis brought to my attention E. Bolthausen's adaptation of C. Stein's techniques (the so-called *Stein's method*) to give a proof that is not only brief but also, to me, aesthetically pleasing. In any case, no use of Fourier methods is made in the derivation given in § 2.2. On the other hand, Fourier techniques are introduced in § 2.3, where it is shown that even elementary Fourier analytic tools lead to important extensions of the basic Central Limit Theorem to more than one dimension. Finally, in § 2.4, the Central Limit Theorem is applied to the study of Hermite multipliers and (following Wm. Beckner) is used to derive both E. Nelson's hypercontraction estimate for the Mehler kernel as well as Beckner's own estimate for the Fourier transform. I am afraid that, with this flagrant example of *the sort of thing that does not belong here*, I may be trying the patience of my purist colleagues. However, I hope that their indignation will be somewhat assuaged by the fact that the rest of the book is essentially independent of the material in § 2.4.

3: This chapter is devoted to the study of infinitely divisible laws. It begins in § 3.1 with a few refinements (especially The Lévy Continuity Theorem) of the Fourier techniques introduced in § 2.3. These play a role in § 3.2, where the Lévy–Khinchine formula is first derived and then applied to the analysis of stable laws.

4: In Chapter 4 I construct the Lévy processes (a.k.a. independent increment processes) corresponding to infinitely divisible laws. Secton 4.1 provides the requisite information about the pathspace $D(\mathbb{R}^N)$ of right-continuous paths with left limits, and § 4.2 gives the construction of Lévy processes with discontinuous paths, the ones corresponding to infinitely divisible laws having no Gaussian part. Finally, in § 4.3 I construct Brownian motion, the Lévy process with continuous paths, following the prescription given by Lévy.

5: Because they are not needed earlier, conditional expectations do not appear until Chapter 5. The advantage gained by this postponement is that, by the time I introduce them, I have an ample supply of examples to which conditioning can be applied; the disadvantage is that, with considerable justice, many probabilists feel that one is not doing *probability theory* until one is conditioning. Be that as it may, Kolmogorov's definition is given in § 5.1 and is shown to extend naturally both to σ-finite measure spaces as well as to random variables with values in a Banach space. Section 5.2 presents Doob's basic theory of real-valued, discrete parameter martingales: Doob's Inequality, his Stopping Time Theorem, and his Martingale Convergence Theorem. In the last part of § 5.2, I introduce reversed martingales and apply them to DeFinetti's theory of exchangeable random variables.

6: Chapter 6 opens with extensions of martingale theory in two directions: to σ-finite measures and to random variables with values in a Banach space. The results in § 6.1 are used in § 6.2 to derive Birkhoff's Individual Ergodic Theorem and a couple of its applications. Finally, in § 6.3 I prove Burkholder's Inequality for martingales with values in a Hilbert space. The derivation that I give is essentially the same as Burkholder's second proof, the one that gives optimal constants.

7: Section 7.1 provides a brief introduction to the theory of martingales with a continuous parameter. As anyone at all familiar with the topic knows, anything approaching a full account of this theory requires much more space than a book like this can give it. Thus, I deal with only its most rudimentary aspects, which, fortunately, are sufficient for the applications to Brownian motion that I have in mind. Namely, in § 7.2 I first discuss the intimate relationship between continuous martingales and Brownian motion (Lévy's martingale characterization of Brownian motion), then derive the simplest (and perhaps most widely applied) case of the Doob–Meyer Decomposition Theory, and finally show what Burkholder's Inequality looks like for continuous martingales. In the concluding section, § 7.3, the results in §§ 7.1–7.2 are applied to derive the Reflection Principle for Brownian motion.

8: In § 8.1 I formulate the description of Brownian motion in terms of its Gaussian, as opposed to its independent increment, properties. More precisely, following Segal and Gross, I attempt to convince the reader that Wiener measure (i.e., the distribution of Brownian motion) would like to be the standard Gauss measure on the Hilbert space $H^1(\mathbb{R}^N)$ of absolutely continuous paths with a square integrable derivative, but, for technical reasons, cannot live there and has to settle for a Banach space in which $H^1(\mathbb{R}^N)$ is densely embedded. Using Wiener measure as the model, in § 8.2 I show that, at an abstract level, any non-degenerate, centered Gaussian measure on an infinite dimensional, separable Banach space shares the same structure as Wiener measure in the sense that there is always a densely embedded Hilbert space, known as the Cameron–Martin space, for which it would like to be the standard Gaussian measure but on which it does not fit. In order to carry out this program, I need and prove Fernique's Theorem for Gaussian measures on a Banach space. In § 8.3 I begin by going in the opposite direction, showing how to pass from a Hilbert space H to a Gaussian measure on a Banach space E for which H is the Cameron–Martin space. The rest of § 8.3 gives two applications: one to "pinned Brownian" motion and the second to a very general statement of orthogonal invariance for Gaussian measures. The main goal of § 8.4 is to prove a large deviations result, known as Schilder's Theorem, for abstract Wiener spaces; and once I have Schilder's Theorem, I apply it to derive a version of Strassen's Law of the Iterated Logarithm. Starting with the Ornstein–Uhlenbeck process, I construct in § 8.5 a family of Gaussian measures known in the mathematical physics literature as Euclidean free fields. In the final section, § 8.6, I first show how to construct Banach space–valued Brownian motion and then derive the original form of Strassen's Law of the Iterated Logarithm in that context.

9: The central topic here is the abstract theory of weak convergence of probability measures on a Polish space. The basic theory is developed in § 9.1. In § 9.2 I apply the theory to prove the existence of regular conditional probability distributions, and in § 9.3 I use it to derive Donsker's Invariance Principle (i.e., the pathspace statement of the Central Limit Theorem).

10: Chapter 10 is an introduction to the connections between probability theory and partial differential equations. At the beginning of § 10.1 I show that martingale theory provides a link between probability theory and partial differential equations. More precisely, I show how to represent in terms of Wiener integrals solutions to parabolic and elliptic partial differential equations in which the Laplacian is the principal part. In the second part of § 10.1, I use this link to calculate various Wiener integrals. In § 10.2 I introduce the Markov property of Wiener measure and show how it not only allows one to evaluate other Wiener integrals in terms of solutions to elliptic partial differential equations but also enables one to prove interesting facts about solutions to such equations as a consequence of their representation in terms of Wiener integrals. Continuing in the same spirit, I show in § 10.2 how to represent solutions to the Dirichlet problem in terms of Wiener integrals, and in § 10.3 I use Wiener measure to construct and discuss heat kernels related to the Laplacian.

11: The final chapter is an extended example of the way in which probability theory meshes with other branches of analysis, and the example that I have chosen is the marriage between Brownian motion and classical potential theory. Like an ideal marriage, this one is simultaneously intimate and mutually beneficial to both partners. Indeed, the more one knows about it, the more convinced one becomes that the properties of Brownian paths are a perfect reflection of properties of harmonic functions, and vice versa. In any case, in § 11.1 I sharpen the results in § 10.2.3 and show that, in complete generality, the solution to the Dirichlet problem is given by the Wiener integral of the boundary data evaluated at the place where Brownian paths exit from the region. Next, in § 11.2, I discuss the Green function for a region and explain how its existence reflects the recurrence and transience properties of Brownian paths. In preparation for § 11.4, § 11.3 is devoted to the Riesz Decomposition Theorem for excessive functions. Finally, in § 11.4, I discuss the capacity of regions, derive Chung's representation of the capacitory measure in terms of the last place where a Brownian path visits a region, apply the probabilistic interpretation of capacity to give a derivation of Wiener's test for regularity, and conclude with two asymptotic calculations in which capacity plays a crucial role.

Suggestions about the Use of This Book

In spite of the realistic assessment contained in the first paragraph of its preface, when I wrote the first edition of this book I harbored the naïve hope that it might become *the standard* graduate text in probability theory. By the time that I started preparing the second edition, I was significantly older and far less naïve about its prospects. Although the first edition has its admirers, it has done little to dent the sales record of its competitors. In particular, the first edition has seldom been adopted as the text for courses in probability, and I doubt that the second will be either. Nonetheless, I close this preface with a few suggestions for anyone who does choose to base a course on it.

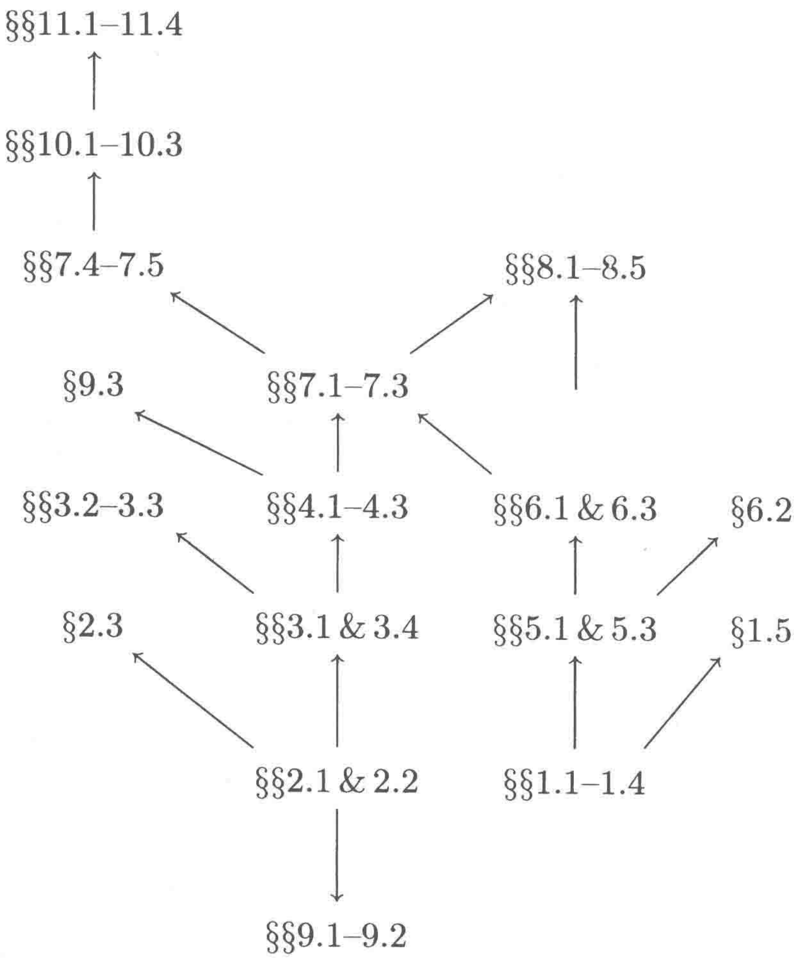
I am well aware that, except for those who find their way into the poorly stocked library of some prison camp, few copies of this book will be read from cover to cover. For this reason, I have attempted to organize it in such a way that, with the help of the table of dependence that follows, a reader can select a path which does not require his reading all the sections preceding the information he is seeking. For example, the contents of §§ 1.1–1.2, § 1.4, § 2.1, § 2.3, and § 5.1–5.2 constitute the backbone of a one semester, graduate level introduction to probability theory. What one attaches to this backbone depends on the speed with which these sections are covered and the content of the courses for which the course is the introduction. If the goal is to prepare the students for a career as a "quant" in what is left of the financial industry, an obvious choice is § 4.3 and as much of Chapter 7 as time permits, thereby giving one's students a reasonably solid introduction to Brownian motion. On the other hand, if one wants the students to appreciate that white noise is not the only noise that they may encounter in life, one might defer the discussion of Brownian motion and replace it with the material in Chapter 3 and §§ 4.1–4.2.

Alternatively, one might use this book in a more advanced course. An introduction to stochastic processes with an emphasis on their relationship to partial differential equations can be constructed out of Chapters 6, 7, 10, and 11, and § 4.3 combined with Chapter 8 could be used to provide background for a course on Gaussian processes.

Whatever route one takes through this book, it will be a great help to your students for you to suggest that they consult other texts. Indeed, it is a familiar fact that the third book one reads on a subject is always the most lucid, and so one should suggest at least two other books. Among the many excellent choices available, I mention Wm. Feller's *An Introduction to Probability Theory and Its Applications, Vol. II,* and M. Loéve's classic *Probability Theory.* In addition, for background, precision (including accuracy of attribution), and supplementary material, R. Dudley's *Real Analysis and Probability* is superb. Finally, an ever growing list of errata can be found at

www.mit-math.edu/~dws/prob2errata.pdf

Table of Dependence



Contents