Sidney I. Resnick

A Probability Path

概率论入门

Birkhäuser

世界图书出版公司

www.wpcbj.com.cn

Sidney I. Resnick

A Probability Path

Birkhäuser
Boston • Basel • Berlin

图书在版编目（CIP）数据

概率论入门 = A probability path：英文/（美）雷斯尼克（Resnick，S. I.）著.
—影印本.—北京：世界图书出版公司北京公司，2013. 3
ISBN 978-7-5100-5827-1

Ⅰ. ①概… Ⅱ. ①雷… Ⅲ. ①概率论—教材—英文 Ⅳ. ①O211

中国版本图书馆 CIP 数据核字（2013）第 035280 号

书　　名：	A Probability Path
作　　者：	Sidney I. Resnick
中 译 名：	概率论入门
责任编辑：	高蓉　刘慧
出 版 者：	世界图书出版公司北京公司
印 刷 者：	三河市国英印务有限公司
发　　行：	世界图书出版公司北京公司（北京朝内大街 137 号 100010）
联系电话：	010-64021602，010-64015659
电子信箱：	kjb@ wpcbj. com. cn
开　　本：	24 开
印　　张：	19. 5
版　　次：	2013 年 3 月
版权登记：	图字：01-2012-8539
书　　号：	978-7-5100-5827-1
定　　价：	69. 00 元

Preface

There are several good current probability books — Billingsley (1995), Durrett (1991), Port (1994), Fristedt and Gray (1997), and I still have great affection for the books I was weaned on — Breiman (1992), Chung (1974), Feller (1968, 1971) and even Loève (1977). The books by Neveu (1965, 1975) are educational and models of good organization. So why publish another? Many of the existing books are encyclopedic in scope and seem intended as reference works, with navigation problems for the beginner. Some neglect to teach any measure theory, assuming students have already learned all the foundations elsewhere. Most are written by mathematicians and have the built in bias that the reader is assumed to be a mathematician who is coming to the material for its beauty. Most books do not clearly indicate a one-semester syllabus which will offer the essentials.

I and my students have consequently found difficulties using currently available probability texts. There is a large market for measure theoretic probability by students whose primary focus is not mathematics for its own sake. Rather, such students are motivated by examples and problems in statistics, engineering, biology and finance to study probability with the expectation that it will be useful to them in their research work. Sometimes it is not clear where their work will take them, but it is obvious they need a deep understanding of advanced probability in order to read the literature, understand current methodology, and prove that the new technique or method they are dreaming up is superior to standard practice.

So the clientele for an advanced or measure theoretic probability course that is primarily motivated by applications outnumbers the clientele deeply embedded in pure mathematics. Thus, I have tried to show links to statistics and operations research. The pace is quick and disciplined. The course is designed for one semester with an overstuffed curriculum that leaves little time for interesting excursions or personal favorites. A successful book needs to cover the basics clearly. Equally important, the exposition must be efficient, allowing for time to cover the next important topic.

Chapters 1, 2 and 3 cover enough measure theory to give a student access to advanced material. Independence is covered carefully in Chapter 4 and expectation and Lebesgue integration in Chapter 5. There is some attention to comparing the Lebesgue vs the Riemann integral, which is usually an area that concerns students. Chapter 6 surveys and compares different modes of convergence and must be carefully studied since limit theorems are a central topic in classical probability and form the core results. This chapter naturally leads into laws of large numbers (Chapter 7), convergence in distribution, and the central limit theorem (Chapters 8 and 9). Chapter 10 offers a careful discussion of conditional expectation and martingales, including a short survey of the relevance of martingales to mathematical finance.

Suggested syllabi: If you have one semester, you have the following options: You could cover Chapters 1–8 plus 9, or Chapters 1–8 plus 10. You would have to move along at unacceptable speed to cover both Chapters 9 and 10. If you have two quarters, do Chapters 1–10. If you have two semesters, you could do Chapters 1–10, and then do the random walk Chapter 7 and the Brownian Motion Chapter 6 from Resnick (1992), or continue with stochastic calculus from one of many fine sources.

Exercises are included and students should be encouraged or even forced to do many of them.

Harry is on vacation.

Acknowledgements. Cornell University continues to provide a fine, stimulating environment. NSF and NSA have provided research support which, among other things, provides good computing equipment. I am pleased that AMS-TEXand LATEX merged into AMS-LATEX, which is a marvelous tool for writers. Rachel, who has grown into a terrific adult, no longer needs to share her mechanical pencils with me. Nathan has stopped attacking my manuscripts with a hole puncher and gives ample evidence of the fine adult he will soon be. Minna is the ideal companion on the random path of life. Ann Kostant of Birkhäuser continues to be a pleasure to deal with.

Sidney I. Resnick
School of Operations Research and Industrial Engineering
Cornell University

Contents

1

Sets and Events

1.1 Introduction

The core classical theorems in probability and statistics are the following:

- *The law of large numbers (LLN)*: Suppose $\{X_n, n \geq 1\}$ are independent, identically distributed (iid) random variables with common mean $E(X_n) = \mu$. The LLN says the sample average is approximately equal to the mean, so that

  $$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mu.$$

  An immediate concern is what does the convergence arrow "$\to$" mean? This result has far-reaching consequences since, if

  $$X_i = \begin{cases} 1, & \text{if event } A \text{ occurs,} \\ 0, & \text{otherwise} \end{cases}$$

  then the average $\sum_{i=1}^{n} X_i/n$ is the relative frequency of occurrence of A in n repetitions of the experiment and $\mu = P(A)$. The LLN justifies the frequency interpretation of probabilities and much statistical estimation theory where it underlies the notion of *consistency* of an estimator.

- *Central limit theorem (CLT)*: The central limit theorem assures us that sample averages when centered and scaled to have mean 0 and variance 1 have a distribution that is approximately normal. If $\{X_n, n \geq 1\}$ are iid with common mean $E(X_n) = \mu$ and variance $\text{Var}(X_n) = \sigma^2$, then

$$P\left[\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \le x\right] \to N(x) := \int_{-\infty}^{x} \frac{e^{-u^2/2}}{\sqrt{2\pi}} du.$$

This result is arguably the most important and most frequently applied result of probability and statistics. How is this result and its variants proved?

- *Martingale convergence theorems and optional stopping*: A martingale is a stochastic process $\{X_n, n \ge 0\}$ used to model a fair sequence of gambles (or, as we say today, investments). The conditional expectation of your wealth X_{n+1} after the next gamble or investment given the past equals the current wealth X_n. The martingale results on convergence and optimal stopping underlie the modern theory of stochastic processes and are essential tools in application areas such as mathematical finance. What are the basic results and why do they have such far reaching applicability?

Historical references to the CLT and LLN can be found in such texts as Breiman (1968), Chapter I; Feller, volume I (1968) (see the background on coin tossing and the de Moivre-Laplace CLT); Billingsley (1995), Chapter 1; Port (1994), Chapter 17.

1.2 Basic Set Theory

Here we review some basic set theory which is necessary before we can proceed to carve a path through classical probability theory. We start by listing some basic notation.

- Ω: An abstract set representing the sample space of some experiment. The points of Ω correspond to the outcomes of an experiment (possibly only a thought experiment) that we want to consider.

- $\mathcal{P}(\Omega)$: The power set of Ω, that is, the set of all subsets of Ω.

- Subsets $A, B, \ldots$ of Ω which will usually be written with roman letters at the beginning of the alphabet. Most (but maybe not all) subsets will be thought of as *events*, that is, collections of simple events (points of Ω).

  The necessity of restricting the class of subsets which will have probabilities assigned to them to something perhaps smaller than $\mathcal{P}(\Omega)$ is one of the sophistications of modern probability which separates it from a treatment of discrete sample spaces.

- Collections of subsets $\mathcal{A}, \mathcal{B}, \ldots$ which will usually be written by calligraphic letters from the beginning of the alphabet.

- An individual element of Ω: $\omega \in \Omega$.

- The empty set $\emptyset$, not to be confused with the Greek letter ϕ.

$\mathcal{P}(\Omega)$ has the structure of a Boolean algebra. This is an abstract way of saying that the usual set operations perform in the usual way. We will proceed using naive set theory rather than by axioms. The *set operations which you should know and will be commonly used* are listed next. These are often used to manipulate sets in a way that parallels the construction of complex events from simple ones.

1. *Complementation*: The complement of a subset $A \subset \Omega$ is

$$A^c := \{\omega : \omega \notin A\}.$$

2. *Intersection over arbitrary index sets*: Suppose T is some index set and for each $t \in T$ we are given $A_t \subset \Omega$. We define

$$\bigcap_{t \in T} A_t := \{\omega : \omega \in A_t, \quad \forall\, t \in T\}.$$

The collection of subsets $\{A_t, t \in T\}$ is *pairwise disjoint* if whenever $t, t' \in T$, but $t \neq t'$, we have

$$A_t \cap A_{t'} = \emptyset.$$

A synonym for pairwise disjoint is *mutually disjoint*. Notation: When we have a small number of subsets, perhaps two, we write for the intersection of subsets A and B

$$AB = A \cap B,$$

using a "multiplication" notation as shorthand.

3. *Union over arbitrary index sets*: As above, let T be an index set and suppose $A_t \subset \Omega$. Define the union as

$$\bigcup_{t \in T} A_t := \{\omega : \omega \in A_t, \quad \text{for some} \quad t \in T\}.$$

When sets $A_1, A_2, \dots$ are mutually disjoint, we sometimes write

$$A_1 + A_2 + \dots$$

or even $\sum_{i=1}^{\infty} A_i$ to indicate $\cup_{i=1}^{\infty} A_i$, the union of mutually disjoint sets.

4. *Set difference* Given two sets A, B, the part that is in A but not in B is

$$A \setminus B := AB^c.$$

This is most often used when $B \subset A$; that is, when $AB = B$.

5. *Symmetric difference*: If A, B are two subsets, the points that are in one but not in both are called the symmetric difference

$$A \,\triangle\, B = (A \setminus B) \cup (B \setminus A).$$

You may wonder why we are interested in arbitrary index sets. Sometimes the natural indexing of sets can be rather exotic. Here is one example. Consider the space $USC_+([0, \infty))$, the space of non-negative *upper semi-continuous* functions with domain $[0, \infty)$. For $f \in USC_+([0, \infty))$, define the hypograph hypo(f) by

$$\text{hypo}(f) = \{(s, x) : 0 \le x \le f(s)\},$$

so that hypo(f) is the portion of the plane between the horizontal axis and the graph of f. Thus we have a family of sets indexed by the upper semi-continuous functions, which is a somewhat more exotic index set than the usual subsets of the integers or real line.

The previous list described common ways of constructing new sets from old. Now we list ways sets can be compared. Here are some simple *relations between sets*.

1. *Containment*: A is a subset of B, written $A \subset B$ or $B \supset A$, iff $AB = A$ or equivalently iff $\omega \in A$ implies $\omega \in B$.

2. *Equality*: Two subsets A, B are equal, written $A = B$, iff $A \subset B$ and $B \subset A$. This means $\omega \in A$ iff $\omega \in B$.

Example 1.2.1 Here are two simple examples of set equality on the real line for you to verify.

(i) $\bigcup_{n=1}^{\infty}[0, n/(n+1)) = [0, 1)$.

(ii) $\bigcap_{n=1}^{\infty}(0, 1/n) = \emptyset$. □

Here are some straightforward *properties of set containment* that are easy to verify:

$$\begin{aligned}
&A \subset A,\\
&A \subset B \text{ and } B \subset C \text{ implies } A \subset C,\\
&A \subset C \text{ and } B \subset C \text{ implies } A \cup B \subset C,\\
&A \supset C \text{ and } B \supset C \text{ implies } AB \supset C,\\
&A \subset B \text{ iff } B^c \subset A^c
\end{aligned}$$

Here is a list of *simple connections* between the set operations:

1. *Complementation*:

$$(A^c)^c = A, \quad \emptyset^c = \Omega, \quad \Omega^c = \emptyset.$$

2. *Commutativity* of set union and intersection:

$$A \cup B = B \cup A, \quad A \cap B = B \cap A.$$

Note as a consequence of the definitions, we have

$$\begin{array}{ll} A \cup A = A, & A \cap A = A, \\ A \cup \emptyset = A, & A \cap \emptyset = \emptyset \\ A \cup \Omega = \Omega, & A \cap \Omega = A, \\ A \cup A^c = \Omega, & A \cap A^c = \emptyset. \end{array}$$

3. *Associativity* of union and intersection:

$$(A \cup B) \cup C = A \cup (B \cup C), \ (A \cap B) \cap C = A \cap (B \cap C).$$

4. *De Morgan's laws*, a relation between union, intersection and complementation: Suppose as usual that T is an index set and $A_t \subset \Omega$. Then we have

$$(\bigcup_{t \in T} A_t)^c = \bigcap_{t \in T}(A_t^c), \quad (\bigcap_{t \in T} A_t)^c = \bigcup_{t \in T}(A_t^c).$$

The two De Morgan's laws given are equivalent.

5. *Distributivity* laws providing connections between union and intersection:

$$\begin{aligned} B \cap \left(\bigcup_{t \in T} A_t\right) &= \bigcup_{t \in T}(BA_t), \\ B \cup \left(\bigcap_{t \in T} A_t\right) &= \bigcap_{t \in T}(B \cup A_t). \end{aligned}$$

1.2.1 Indicator functions

There is a very nice and useful duality between sets and functions which emphasizes the algebraic properties of sets. It has a powerful expression when we see later that taking the expectation of a random variable is theoretically equivalent to computing the probability of an event. If $A \subset \Omega$, we define the *indicator function of A* as

$$1_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{if } \omega \in A^c. \end{cases}$$

This definition quickly yields the simple properties:

$$1_A \leq 1_B \text{ iff } A \subset B,$$

and

$$1_{A^c} = 1 - 1_A.$$

Note here that we use the convention that for two functions f, g with domain Ω and range $\mathbb{R}$, we have

$$f \leq g \text{ iff } f(\omega) \leq g(\omega) \text{ for all } \omega \in \Omega$$

and

$$f = g \text{ if } f \leq g \text{ and } g \leq f.$$

1.3 Limits of Sets

The definition of convergence concepts for random variables rests on manipulations of sequences of events which require limits of sets. Let $A_n \subset \Omega$. We define

$$\inf_{k \geq n} A_k := \bigcap_{k=n}^{\infty} A_k, \quad \sup_{k \geq n} A_k := \bigcup_{k=n}^{\infty} A_k$$

$$\liminf_{n \to \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k,$$

$$\limsup_{n \to \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

The *limit* of a sequence of sets is defined as follows: If for some sequence $\{B_n\}$ of subsets

$$\limsup_{n \to \infty} B_n = \liminf_{n \to \infty} B_n = B,$$

then B is called the limit of B_n and we write $\lim_{n \to \infty} B_n = B$ or $B_n \to B$. It will be demonstrated soon that

$$\liminf_{n \to \infty} A_n = \lim_{n \to \infty} \left(\inf_{k \geq n} A_k \right)$$

and

$$\limsup_{n \to \infty} A_n = \lim_{n \to \infty} \left(\sup_{k \geq n} A_k \right).$$

To make sure you understand the definitions, you should check the following example as an exercise.

Example 1.3.1 Check

$$\liminf_{n \to \infty} [0, n/(n+1)) = \limsup_{n \to \infty} [0, n/(n+1)) = [0, 1). \qquad \square$$

We can now give an interpretation of $\liminf_{n \to \infty} A_n$ and $\limsup_{n \to \infty} A_n$.

Lemma 1.3.1 *Let $\{A_n\}$ be a sequence of subsets of Ω.*

(a) For lim sup *we have the interpretation*

$$\limsup_{n \to \infty} A_n = \left\{ \omega : \sum_{n=1}^{\infty} 1_{A_n}(\omega) = \infty \right\}$$
$$= \left\{ \omega : \omega \in A_{n_k}, k = 1, 2 \ldots \right\}$$

for some subsequence n_k depending on ω. Consequently, we write

$$\limsup_{n \to \infty} A_n = [A_n \text{ i.o. }]$$

where i.o. stands for infinitely often.

(b) For lim inf *we have the interpretation*

$$\liminf_{n\to\infty} A_n = \{\omega : \omega \in A_n \text{ for all } n \text{ except a finite number }\}$$
$$= \{\omega : \sum_n 1_{A_n^c}(\omega) < \infty\}$$
$$= \{\omega : \omega \in A_n, \forall n \geq n_0(\omega)\}.$$

Proof. (a) If

$$\omega \in \limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k,$$

then for every n, $\omega \in \cup_{k\geq n} A_k$ and so for all n, there exists some $k_n \geq n$ such that $\omega \in A_{k_n}$, and therefore

$$\sum_{j=1}^{\infty} 1_{A_j}(\omega) \geq \sum_n 1_{A_{k_n}}(\omega) = \infty,$$

which implies

$$\omega \in \left\{\omega : \sum_{n=1}^{\infty} 1_{A_n}(\omega) = \infty\right\};$$

thus

$$\limsup_{n\to\infty} A_n \subset \{\omega : \sum_{j=1}^{\infty} 1_{A_j}(\omega) = \infty\}.$$

Conversely, if

$$\omega \in \{\omega : \sum_{j=1}^{\infty} 1_{A_j}(\omega) = \infty\},$$

then there exists $k_n \to \infty$ such that $\omega \in A_{k_n}$, and therefore for all n, $\omega \in \cup_{j\geq n} A_j$ so that $\omega \in \limsup_{n\to\infty} A_n$. By defininition

$$\{\omega : \sum_{j=1}^{\infty} 1_{A_j}(\omega) = \infty\} \subset \limsup_{n\to\infty} A_n.$$

This proves the set inclusion in both directions and shows equality.

The proof of (b) is similar. □

The properties of lim sup and lim inf are analogous to what we expect with real numbers. The link is through the indicator functions and will be made explicit shortly. Here are two simple connections:

1. The relationship between lim sup and lim inf is

$$\liminf_{n\to\infty} A_n \subset \limsup_{n\to\infty} A_n$$