数学图书 影印版系列

Jean Jacod
Philip Protter 著

概率论基础

Probability Essentials

清华大学出版社

This introduction to Probability Theory can be used, at the beginning graduate level, for a one-semester course on Probability Theory or for self-direction without benefit of a formal course; the measure theory needed is developed in the text. It will also be useful for students and teachers in related areas such as Finance Theory (Economics), Electrical Engineering, and Operations Research. The text covers the essentials in a directed and lean way with 28 short chapters. Assuming of readers only an undergraduate background in mathematics, it brings them from a starting knowledge of the subject to a knowledge of the basics of Martingale Theory. After learning Probability Theory form this text, the interested student will be ready to continue with the study of more advanced topics, such as Brownian Motion and Ito Calculus, or Statistical Inference. The second edition contains some additions to the text and to the references and some parts are completely rewritten.

ISBN 978-7-302-24155-3

9 787302 241553 >

定价：35.00元

数学图书 影印版系列

Jean Jacod
Philip Protter 著

概率论基础

Probability Essentials

清华大学出版社
北京

Jean Jacod　Philip Protter
Probability Essentials
ISBN: 3-540-43871-8

北京市版权局著作权合同登记号　图字：01-2007-0343

图书在版编目(CIP)数据

概率论基础=Probability Essentials：英文 /（法）雅科蒂（Jacod,J.），（美）普罗特（Protter,P.）著. -- 北京：清华大学出版社，2010.12
ISBN 978-7-302-24155-3

Ⅰ.①概… Ⅱ.①雅… ②普… Ⅲ.①概率论－高等学校－教材－英文　Ⅳ.①O21

中国版本图书馆CIP数据核字(2010)第238758号

责任编辑：刘　颖
责任印制：杨　艳
出版发行：清华大学出版社　　**地　址**：北京清华大学学研大厦 A 座
http://www.tup.com.cn　　**邮　编**：100084
社　总　机：010-62770175　　**邮　购**：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质　量　反　馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
印 装 者：北京市清华园胶印厂
经　销：全国新华书店
开　本：185×230　**印　张**：16.5
版　次：2010 年 12 月第 1 版　　**印　次**：2010 年 12 月第1 次印刷
印　数：1～3000
定　价：35.00 元

产品编号：015870-01

To Diane and Sylvie
and
to Rachel, Margot, Olivier, Serge,
Thomas, Vincent and Martin

Preface to the Second Edition

We have made small changes throughout the book, including the exercises, and we have tried to correct if not all, then at least most of the typos. We wish to thank the many colleagues and students who have commented constructively on the book since its publication two years ago, and in particular Professors Valentin Petrov, Esko Valkeila, Volker Priebe, and Frank Knight.

Jean Jacod, Paris
Philip Protter, Ithaca
March, 2002

Preface to the Second Printing of the Second Edition

We have benefited greatly from the long list of typos and small suggestions sent to us by Professor Luis Tenorio. These corrections have improved the book in subtle yet important ways, and the authors are most grateful to him.

Jean Jacod, Paris
Philip Protter, Ithaca
January, 2004

Preface to the First Edition

We present here a one semester course on Probability Theory. We also treat measure theory and Lebesgue integration, concentrating on those aspects which are especially germane to the study of Probability Theory. The book is intended to fill a current need: there are mathematically sophisticated students and researchers (especially in Engineering, Economics, and Statistics) who need a proper grounding in Probability in order to pursue their primary interests. Many Probability texts available today are celebrations of Probability Theory, containing treatments of fascinating topics to be sure, but nevertheless they make it difficult to construct a lean one semester course that covers (what we believe) are the essential topics.

Chapters 1–23 provide such a course. We have indulged ourselves a bit by including Chapters 24–28 which are highly optional, but which may prove useful to Economists and Electrical Engineers.

This book had its origins in a course the second author gave in Perugia, Italy in 1997; he used the samizdat "notes" of the first author, long used for courses at the University of Paris VI, augmenting them as needed. The result has been further tested at courses given at Purdue University. We thank the indulgence and patience of the students both in Perugia and in West Lafayette. We also thank our editor Catriona Byrne, as well as Nick Bingham for many superb suggestions, an anonymous referee for the same, and Judy Mitchell for her extraordinary typing skills.

Jean Jacod, Paris
Philip Protter, West Lafayette

Table of Contents

1 Introduction

Almost everyone these days is familiar with the concept of Probability. Each day we are told the probability that it will rain the next day; frequently we discuss the probabilities of winning a lottery or surviving the crash of an airplane. The insurance industry calculates (for example) the probability that a man or woman will live past his or her eightieth birthday, given he or she is 22 years old and applying for life insurance. Probability is used in business too: for example, when deciding to build a waiting area in a restaurant, one wants to calculate the probability of needing space for more than n people each day; a bank wants to calculate the probability a loan will be repaid; a manufacturer wants to calculate the probable demand for his product in the future. In medicine a doctor needs to calculate the probability of success of various alternative remedies; drug companies calculate the probability of harmful side effects of drugs. An example that has recently achieved spectacular success is the use of Probability in Economics, and in particular in Stochastic Finance Theory. Here interest rates and security prices (such as stocks, bonds, currency exchanges) are modelled as varying randomly over time but subject to specific probability laws; one is then able to provide insurance products (for example) to investors by using these models. One could go on with such a list. Probability theory is ubiquitous in modern society and in science.

Probability theory is a reasonably old subject. Published references on games of chance (i.e., gambling) date to J. Cardan (1501–1576) with his book *De Ludo Alae* [4]. Probability also appears in the work of Kepler (1571–1630) and of Galileo (1564–1642). However historians seem to agree that the subject really began with the work of Pascal (1623–1662) and of Fermat (1601–1665). The two exchanged letters solving gambling "paradoxes" posed to them by the aristocrat de Méré. Later the Dutch mathematician Christian Huygens (1629–1695) wrote an influential book [13] elaborating on the ideas of Pascal and Fermat. Finally in 1685 it was Jacques Bernoulli (1654–1705) who proposed such interesting probability problems (in the "Journal des Scavans") (see also [3]) that it was necessary to develop a serious theory to answer them. After the work of J. Bernoulli and his contemporary A. De Moivre (1667–1754) [6], many renowned mathematicians of the day worked on probability problems, including Daniel Bernoulli (1700–1782), Euler (1707–1803), Gauss (1777–1855), and Laplace (1749–1827). For a nice history of Probability before 1827 (the year of the death of Laplace) one can consult [21]. In the twentieth century it was Kolmogorov (1903–1987) who saw the connection between the ideas of Borel and Lebesgue and probability theory and he gave probability theory its rigorous measure theory basis. After the fundamental work of Kolmogorov, the French mathematician Paul Lévy (1886–1971) set the tone for modern Probability with his seminal work on Stochastic Processes as well as characteristic functions and limit theorems.

We think of Probability Theory as a mathematical model of chance, or random events. The idea is to start with a few basic principles about how the laws of chance behave. These should be sufficiently simple that one can believe them readily to correspond to nature. Once these few principles are accepted, we then deduce a mathematical theory to guide us in more complicated situations. This is the goal of this book.

We now describe the approach of this book. First we cover the bare essentials of discrete probability in order to establish the basic ideas concerning probability measures and conditional probability. We next consider probabilities on countable spaces, where it is easy and intuitive to fix the ideas. We then extend the ideas to general measures and of course probability measures on the real numbers. This represents Chapters 2–7. Random variables are handled analogously: first on countable spaces and then in general. Integration is established as the expectation of random variables, and later the connection to Lebesgue integration is clarified. This brings us through Chapter 12.

Chapters 13 through 21 are devoted to the study of limit theorems, the central feature of classical probability and statistics. We give a detailed treatment of Gaussian random variables and transformations of random variables, as well as weak convergence.

Conditional expectation is not presented via the Radon-Nikodym theorem and the Hahn–Jordan decomposition, but rather we use Hilbert Space projections. This allows a rapid approach to the theory. To this end we cover the necessities of Hilbert space theory in Chapter 22; we nevertheless extend the concept of conditional expectation beyond the Hilbert space setting to include integrable random variables. This is done in Chapter 23. Last, in Chapters 24–28 we give a beginning taste of martingales, with an application to the Radon–Nikodym Theorem. These last five chapters are not really needed for a course on the "essentials of probability". We include them however because many sophisticated applications of probability use martingales; also martingales serve as a nice introduction to the subject of stochastic processes.

We have written the book independent of the exercises. That is, the important material is in the text itself and not in the exercises. The exercises provide an opportunity to absorb the material by working with the subject. Starred exercises are suspected to be harder than the others.

We wish to acknowledge that Allan Gut's book [11] was useful in providing exercises, and part of our treatment of martingales was influenced by the delightful introduction to the book of Richard Bass [1].

No probability background is assumed. The reader should have a good knowledge of (advanced) calculus, some linear algebra, and also "mathematical sophistication".

Random Experiments

Random experiments are experiments whose output cannot be surely predicted in advance. But when one repeats the same experiment a large number of times one can observe some "regularity" in the average output. A typical example is the toss of a coin: one cannot predict the result of a single toss, but if we toss the coin many times we get an average of about 50% of "heads" if the coin is fair.

The theory of probability aims towards a mathematical theory which describes such phenomena. This theory contains three main ingredients:

a) The state space: this is the set of all possible outcomes of the experiment, and it is usually denoted by Ω.

Examples:

1) A toss of a coin: $\Omega = \{\text{h,t}\}$.
2) Two successive tosses of a coin: $\Omega = \{\text{hh,tt,ht,th}\}$.
3) A toss of two dice: $\Omega = \{(i,j) : 1 \le i \le 6, 1 \le j \le 6\}$.
4) The measurement of a length L, with a measurement error: $\Omega = \mathbf{R}_+$, where $\mathbf{R}_+$ denotes the positive real numbers $[0,\infty)$; $\omega \in \Omega$ denotes the result of the measurement, and $\omega - L$ is the measurement error.
5) The lifetime of a light-bulb: $\Omega = \mathbf{R}_+$.

b) The events: An "event" is a property which can be observed either to hold or not to hold *after* the experiment is done. In mathematical terms, an event is a subset of Ω. If A and B are two events, then

- the *contrary* event is interpreted as the complement set A^c;
- the event "A **or** B" is interpreted as the union $A \cup B$;
- the event "A **and** B" is interpreted as the intersection $A \cap B$;
- the *sure* event is Ω;
- the *impossible* event is the empty set $\emptyset$;
- an **elementary event** is a "singleton", i.e. a subset $\{\omega\}$ containing a single outcome ω of Ω.

We denote by $\mathcal{A}$ the family of all events. Often (but not always: we will see why later) we have $\mathcal{A} = 2^\Omega$, the set of all subsets of Ω. The family $\mathcal{A}$ should be "stable" by the logical operations described above: if $A, B \in \mathcal{A}$, then we must have $A^c \in \mathcal{A}$, $A \cap B \in \mathcal{A}$, $A \cup B \in \mathcal{A}$, and also $\Omega \in \mathcal{A}$ and $\emptyset \in \mathcal{A}$.

c) The probability: With each event A one associates a number denoted by $P(A)$ and called the "probability of A". This number measures the likelihood of the event A to be realized *a priori*, before performing the experiment. It is chosen between 0 and 1, and the more likely the event is, the closer to 1 this number is.

To get an idea of the properties of these numbers, one can imagine that they are the limits of the "frequency" with which the events are realized: let us repeat the same experiment n times; the n outcomes might of course be different (think of n successive tosses of the same die, for instance). Denote by $f_n(A)$ the frequency with which the event A is realized (i.e. the number of times the event occurs, divided by n). Intuitively we have:

$$P(A) \;=\; \text{limit of } \; f_n(A) \;\text{ as }\; n \uparrow +\infty. \tag{1.1}$$

(we will give a precise meaning to this "limit" later). From the obvious properties of frequencies, we immediately deduce that:

1. $0 \leq P(A) \leq 1$,
2. $P(\Omega) = 1$,
3. $P(A \cup B) = P(A) + P(B) \quad \text{if} \quad A \cap B = \emptyset$.

A mathematical model for our experiment is thus a triple $(\Omega, \mathcal{A}, P)$, consisting of the space Ω, the family $\mathcal{A}$ of all events, and the family of all $P(A)$ for $A \in \mathcal{A}$; hence we can consider that P is a map from $\mathcal{A}$ into $[0, 1]$, which satisfies at least the properties (2) and (3) above (plus in fact an additional property, more difficult to understand, and which is given in Definition 2.3 of the next Chapter).

A fourth notion, also important although less basic, is the following one:

d) Random variable: A random variable is a quantity which depends on the outcome of the experiment. In mathematical terms, this is a map from Ω into a space E, where often $E = \mathbf{R}$ or $E = \mathbf{R}^d$. **Warning:** this terminology, which is rooted in the history of Probability Theory going back 400 years, is quite unfortunate; a random "variable" is not a variable in the analytical sense, but a function !

Let X be such a random variable, mapping Ω into E. One can then "transport" the probabilistic structure onto the target space E, by setting

$$P^X(B) = P(X^{-1}(B)) \quad \text{for} \quad B \subset E,$$

where $X^{-1}(B)$ denotes the pre-image of B by X, i.e. the set of all $\omega \in \Omega$ such that $X(\omega) \in B$. This formula defines a new probability, denoted by P^X, but on the space E instead of Ω. This probability P^X is called the **law of the variable** X.

Example (toss of two dice): We have seen that $\Omega = \{(i,j) : 1 \leq i \leq 6, 1 \leq j \leq 6\}$, and it is natural to take here $\mathcal{A} = 2^{\Omega}$ and

$$P(A) = \frac{\#(A)}{36} \qquad \text{if } A \subset \Omega,$$

where $\#(A)$ denotes the number of points in A. One easily verifies the properties (1), (2), (3) above, and $P(\{\omega\}) = \frac{1}{36}$ for each singleton. The map $X : \Omega \to \mathbf{N}$ defined by $X(i,j) = i + j$ is the random variable "sum of the two dice", and its law is

$$P_X(B) = \frac{\text{number of pairs } (i,j) \text{ such that } i + j \in B}{36}$$

(for example $P^X(\{2\}) = P(\{1,2\}) = \frac{1}{36}$, $P^X(\{3\}) = \frac{2}{36}$, etc...).

We will formalize the concept of a probability space in Chapter 2, and random variables are introduced with the usual mathematical rigor in Chapters 5 and 8.

2 Axioms of Probability

We begin by presenting the minimal properties we will need to define a Probability measure. Hopefully the reader will convince himself (or herself) that the two axioms presented in Definition 2.3 are reasonable, especially in view of the frequency approach (1.1). From these two simple axioms flows the entire theory. In order to present these axioms, however, we need to introduce the concept of a σ-algebras.

Let Ω be an abstract space, that is with no special structure. Let 2^{Ω} denote all subsets of Ω, including the empty set denoted by $\emptyset$. With $\mathcal{A}$ being a subset of 2^{Ω}, we consider the following properties:

1. $\emptyset \in \mathcal{A}$ and $\Omega \in \mathcal{A}$;
2. If $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$, where A^c denotes the complement of A;
3. $\mathcal{A}$ is closed under finite unions and finite intersections: that is, if $A_1, \dots,$ A_n are all in $\mathcal{A}$, then $\cup_{i=1}^{n} A_i$ and $\cap_{i=1}^{n} A_i$ are in $\mathcal{A}$ as well (for this it is enough that $\mathcal{A}$ be stable by the union and the intersection of any *two* sets);
4. $\mathcal{A}$ is closed under countable unions and intersections: that is, if $A_1, A_2,$ $A_3, \dots$ is a countable sequence of events in $\mathcal{A}$, then $\cup_{i=1}^{\infty} A_i$ and $\cap_{i=1}^{\infty} A_i$ are both also in $\mathcal{A}$.

Definition 2.1. *$\mathcal{A}$ is an* algebra *if it satisfies* (1), (2) *and* (3) *above. It is a* σ-algebra, *(or a σ-field) if it satisfies* (1), (2), *and* (4) *above.*

Note that under (2), (1) can be replaced by either (1′): $\emptyset \in \mathcal{A}$ or by (1″): $\Omega \in \mathcal{A}$. Note also that (1)+(4) implies (3), hence any σ-algebra is an algebra (but there are algebras that are not σ-algebras: see Exercise 2.17).

Definition 2.2. *If $\mathcal{C} \subset 2^{\Omega}$, the* σ-algebra generated by $\mathcal{C}$, *and written* $\sigma(\mathcal{C})$, *is the smallest σ-algebra containing $\mathcal{C}$. (It always exists because 2^{Ω} is a σ-algebra, and the intersection of a family of σ-algebras is again a σ-algebra*: See Exercise 2.2.)

Example:

(i) $\mathcal{A} = \{\emptyset, \Omega\}$ (the trivial σ-algebra).
(ii) A is a subset; then $\sigma(A) = \{\phi, A, A^c, \Omega\}$.

(iii) If $\Omega = \mathbf{R}$ (the Real numbers) (or more generally if Ω is a space with a topology, a case we treat in Chapter 8), the *Borel σ-algebra* is the σ-algebra generated by the open sets (or by the closed sets, which is equivalent).

Theorem 2.1. *The Borel σ-algebra of $\mathbf{R}$ is generated by intervals of the form $(-\infty, a]$, where $a \in \mathbf{Q}$ ($\mathbf{Q}$ = rationals).*

Proof. Let $\mathcal{C}$ denote all open intervals. Since every open set in $\mathbf{R}$ is the countable union of open intervals, we have $\sigma(\mathcal{C})$ = the Borel σ-algebra of $\mathbf{R}$.

Let $\mathcal{D}$ denote all intervals of the form $(-\infty, a]$, where $a \in \mathbf{Q}$. Let $(a, b) \in \mathcal{C}$, and let $(a_n)_{n\geq 1}$ be a sequence of rationals decreasing to a and $(b_n)_{n\geq 1}$ be a sequence of rationals increasing strictly to b. Then

$$\begin{aligned}(a,b) &= \cup_{n=1}^{\infty}(a_n, b_n] \\ &= \cup_{n=1}^{\infty}\left((-\infty, b_n] \cap (-\infty, a_n]^c\right),\end{aligned}$$

Therefore $\mathcal{C} \subset \sigma(\mathcal{D})$, whence $\sigma(\mathcal{C}) \subset \sigma(\mathcal{D})$. However since each element of $\mathcal{D}$ is a closed set, it is also a Borel set, and therefore $\sigma(\mathcal{D})$ is contained in the Borel sets $\mathcal{B}$. Thus we have

$$\mathcal{B} = \sigma(\mathcal{C}) \subset \sigma(\mathcal{D}) \subset \mathcal{B},$$

and hence $\sigma(\mathcal{D}) = \mathcal{B}$. □

On the state space Ω the family of all events will always be a σ-algebra $\mathcal{A}$: the axioms (1), (2) and (3) correspond to the "logical" operations described in Chapter 1, while Axiom (4) is necessary for mathematical reasons. The probability itself is described below:

Definition 2.3. *A* probability measure *defined on a σ-algebra $\mathcal{A}$ of Ω is a function $P : \mathcal{A} \to [0, 1]$ that satisfies:*

1. $P(\Omega) = 1$
2. *For every countable sequence $(A_n)_{n\geq 1}$ of elements of $\mathcal{A}$, pairwise disjoint (that is, $A_n \cap A_m = \emptyset$ whenever $n \neq m$), one has*

$$P\left(\cup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

Axiom (2) above is called *countable additivity*; the number $P(A)$ is called the *probability* of the event A.

In Definition 2.3 one might imagine a more naïve condition than (2), namely:

$$A, B \in \mathcal{A},\ A \cap B = \emptyset \quad \Rightarrow \quad P(A \cup B) = P(A) + P(B). \tag{2.1}$$