Probability an introduction

Geoffrey Grimmett

Dominic Welsh

Probability
an introduction

Geoffrey Grimmett
School of Mathematics, University of Bristol
Dominic Welsh
Merton College, Oxford

CLARENDON PRESS · OXFORD
1986

Oxford University Press, Walton Street, Oxford OX2 6DP

Oxford New York Toronto
Delhi Bombay Calcutta Madras Karachi
Petaling Jaya Singapore Hong Kong Tokyo
Nairobi Dar es Salaam Cape Town
Melbourne Auckland

and associated companies in
Beirut Berlin Ibadan Nicosia

Oxford is a trade mark of Oxford University Press

Published in the United States
by Oxford University Press, New York

British Library Cataloguing in Publication Data

Grimmett, Geoffrey
Probability: an introduction.
1. Probabilities
I. Title II. Welsh, Dominic
519.2 QA273

ISBN 0-19-853272-5
ISBN 0-19-853264-4 (pbk.)

Library of Congress Cataloging in Publication Data

Grimmett, Geoffrey.
Probability: an introduction.
Bibliography: p.
Includes index.
1. Probabilities. I. Welsh, D. J. A. II. Title.
QA273.G735 1986 519.2 85-31035

ISBN 0-19-853272-5
ISBN 0-19-853264-4 (pbk.)

Filmset and printed in Northern Ireland by The Universities Press (Belfast) Ltd.

Preface

Probability and statistics are now taught widely in schools and are integral parts of many O-level and A-level syllabuses. Consequently the attitudes of universities towards these subjects have changed over the last few years and, at many universities, first-year mathematics students learn material which was previously taught in the third year only. This text is based upon first and second year courses in probability theory which are given at the Universities of Bristol and Oxford.

Broadly speaking we cover the usual material, but we hope that our account will have certain special attractions for the reader and we shall say what these may be in a moment. The first eight chapters form a course in basic probability, being an account of events, random variables, and distributions—we treat discrete and continuous random variables separately—together with simple versions of the law of large numbers and the central limit theorem. There is an account of moment generating functions and their applications. The last three chapters are about branching processes, random walks, and continuous-time random processes such as the Poisson process; we hope that these chapters are adequate at this level and are suitable appetizers for courses in applied probability and random processes. We have deliberately omitted various topics which are crucial in more advanced treatments, such as the theory of Markov chains, and we hope that most critics will agree with such decisions. In the case of Markov chains, we could not justify to ourselves the space required to teach more than mere fragments of the theory. On the other hand we have included a brief treatment of characteristic functions in two optional sections for the more advanced reader.

We have divided the text into three sections: (A) Probability, (B) Further Probability, and (C) Random Processes. In doing so we hope to indicate two things. First, the probability in Part A seems to us to be core material for first-year students, whereas the material in Part B is somewhat more difficult. Secondly, although random processes are collected together in the final three chapters, they may well be introduced much earlier in the course. The chapters on branching processes and random walks might well come after Chapter 5, and the chapter on continuous-time processes after Chapter 6.

We have two major aims: to be concise and to be honest about mathematical rigour. Some will say that this book reads like a set of lecture notes. We would not regard this as entirely unfair; indeed a principal reason for writing it was that we believe that most students benefit more from possessing a compact account of the subject in 200 printed pages or so (at a suitable price) than a diffuse account of 400 pages. Most undergraduates learn probability theory by attending lectures, at which they normally take copious and occasionally incorrect notes; they may also attend tutorials and classes. Few are the undergraduates who learn probability in private by relying on a textbook as the sole, or even principal, source of inspiration and learning. Although some will say that this book is too difficult, it is the case that first-year students at many universities learn some quite difficult things, such as axiomatic systems in algebra and ε/δ analysis, and we doubt if much of the material covered here is inherently more challenging than these. Also, lecturers and tutors have certain advantages over authors—they have the power to hear and speak to their audiences—and these advantages should help them to explain the harder things to their students.

Here are a few words about our approach to rigour. It is clearly impossible to prove everything with complete rigour at this level; on the other hand it is important that students should understand why rigour is necessary. We try to be rigorous where possible, and elsewhere we go to some lengths to point out how and where we skate over thin ice. This can occasionally be tedious.

Most sections finish with a few exercises; these are usually completely routine, and students should do them as a matter of course. Each chapter finishes with a collection of problems; these are often much harder than the exercises, and include many parts of questions taken from examination papers set in Bristol and Oxford; we acknowledge permission from Bristol University and from Oxford University Press in this regard. There is a final chapter containing some hints for solving the problems. Problems marked with an asterisk may be rather difficult.

We hope that the remaining mistakes and misprints are not held against us too much, and that they do not pose overmuch of a hazard to the reader. Only with the kind help of our students have we reduced them to the present level. Finally we thank Rhoda Rees for typing the manuscript with such skill, speed and good cheer.

Bristol and Oxford G. G.
July 1985 D. W.

Contents

原书缺页

原书缺页

1
Events and probabilities

1.1 Experiments with chance

Many actions have outcomes which are largely unpredictable in advance—tossing a coin and throwing a dart are simple examples. Probability theory is about such actions and their consequences. The mathematical theory starts with the idea of an *experiment* (or *trial*), being a course of action whose consequence is not predetermined; this experiment is reformulated as a mathematical object called a *probability space*. In broad terms the probability space corresponding to a given experiment comprises three items:

(i) the set of all possible outcomes of the experiment,
(ii) a list of all the events which may possibly occur as consequences of the experiment,
(iii) an assessment of the likelihoods of these events.

For example, if the experiment is the throwing of a fair six-sided die, then the probability space contains

(i) the set $\{1, 2, 3, 4, 5, 6\}$ of possible outcomes,
(ii) a list of events such as 'the result is 3',
'the result is at least 4',
'the result is a prime number',
(iii) the assessment that each number $1, 2, 3, 4, 5, 6$ is equally likely to be the result of the throw.

Given any experiment involving chance, there is a corresponding probability space, and the study of such spaces is called *probability theory*. Next, we shall see how to construct such spaces more explicitly.

1.2 Outcomes and events

We use the letter $\mathscr{E}$ to denote a particular experiment whose outcome is not completely predetermined. The first thing which we do is to make a list of all the possible outcomes of $\mathscr{E}$; the set of all such possible outcomes is called the *sample space* of $\mathscr{E}$ and we usually denote it by Ω. The Greek letter ω denotes a typical member of Ω, and we call each member ω of Ω an *elementary event*.

If, for example, $\mathscr{E}$ is the experiment of throwing a fair die once, then $\Omega = \{1, 2, 3, 4, 5, 6\}$. There are many questions which we may wish to ask about the actual outcome of this experiment (questions such as 'is the outcome a prime number?'), and all such questions may be rewritten in terms of subsets of Ω (the previous question becomes 'does the outcome lie in the subset $\{2, 3, 5\}$ of Ω?'). The second thing which we do is to make a list of all the events which are interesting to us; this list takes the form of a collection of subsets of Ω, each such subset A representing the event 'the outcome of $\mathscr{E}$ lies in A'. Thus we ask 'which possible events are interesting to us' and then we make a list of the corresponding subsets of Ω. This relationship between *events* and *subsets* is very natural, especially because two or more events combine with each other in just the same way as the corresponding subsets combine; for example, if A and B are subsets of Ω then

the set $A \cup B$ corresponds to the event 'either A or B occurs',
the set $A \cap B$ corresponds to the event 'both A and B occur',
the set $\Omega \backslash A$ corresponds† to the event 'A does not occur',

where we say that a subset C of Ω 'occurs' whenever the outcome of $\mathscr{E}$ lies in C. Thus all set-theoretic statements and combinations may be interpreted in terms of events; for example, the formula

$$\Omega \backslash (A \cap B) = (\Omega \backslash A) \cup (\Omega \backslash B)$$

may be read as 'if A and B do not both occur, then either A does not occur or B does not occur'. In a similar way, if $A_1, A_2, \ldots$ are events then the sets $\bigcup_{i=1}^{\infty} A_i$ and $\bigcap_{i=1}^{\infty} A_i$ represent the events 'A_i occurs, for some i' and 'A_i occurs, for every i', respectively.

Thus we write down a collection $\mathscr{F} = \{A_i : i \in I\}$ of subsets of Ω which are interesting to us; each $A \in \mathscr{F}$ is called an *event*. In simple cases, such as the die-throwing example above, we usually take $\mathscr{F}$ to be the set of *all* subsets of Ω (called the *power set* of Ω), but for reasons which may be appreciated later there are many circumstances in which we take $\mathscr{F}$ to be a very much smaller collection than the entire power set. In all cases we demand a certain consistency of $\mathscr{F}$, in the following sense. If $A, B, C, \ldots \in \mathscr{F}$ then we may reasonably be interested also in the events 'A does *not* occur' and 'at least one of $A, B, C, \ldots$ occurs'. With this in mind we require that $\mathscr{F}$ satisfy the following definition.

The collection $\mathscr{F}$ of subsets of the sample space Ω is called an *event space* if

(1) $\mathscr{F}$ is non-empty,

† For any subset A of Ω, the *complement* of A is the set of all members of Ω which are not members of A. We denote the complement of A by either $\Omega \backslash A$ or A^c, depending on the context.

(2) $$\text{if } A \in \mathscr{F} \text{ then } \Omega \backslash A \in \mathscr{F},$$

(3) $$\text{if } A_1, A_2, \ldots \in \mathscr{F} \text{ then } \bigcup_{i=1}^{\infty} A_i \in \mathscr{F}.$$

We speak of an event space $\mathscr{F}$ as being 'closed under the operations of taking complements and countable unions'. An elementary consequence of axioms (1)–(3) is that an event space $\mathscr{F}$ must contain the empty set $\varnothing$ and the whole set Ω. This holds since $\mathscr{F}$ contains some set A (from (1)), and hence $\mathscr{F}$ contains $\Omega \backslash A$ (from (2)), giving also that $\mathscr{F}$ contains the union $\Omega = A \cup (\Omega \backslash A)$ together with the complement $\Omega \backslash \Omega = \varnothing$ of this last set.

Here are some examples of pairs $(\Omega, \mathscr{F})$ of sample spaces and event spaces.

Example 4 Ω is any set and $\mathscr{F}$ is the power set of Ω. □

Example 5 Ω is any set and $\mathscr{F} = \{\varnothing, A, \Omega \backslash A, \Omega\}$ where A is a given subset of Ω. □

Example 6 $\Omega = \{1, 2, 3, 4, 5, 6\}$ and $\mathscr{F}$ is the collection

$$\varnothing, \{1,2\}, \{3,4\}, \{5,6\}, \{1,2,3,4\}, \{3,4,5,6\}, \{1,2,5,6\}, \Omega$$

of subsets of Ω. This event space is unlikely to arise naturally in practice. □

Exercises In these exercises, Ω is a set and $\mathscr{F}$ is an event space of subsets of Ω.

1. If $A, B \in \mathscr{F}$, show that $A \cap B \in \mathscr{F}$.
2. The *difference* $A \backslash B$ of two subsets A and B of Ω is the set $A \cap (\Omega \backslash B)$ of all points of Ω which are in A but not in B. Show that if $A, B \in \mathscr{F}$, then $A \backslash B \in \mathscr{F}$.
3. The *symmetric difference* $A \triangle B$ of two subsets A and B of Ω is defined to be the set of points of Ω which are in either A or B but not in both. If $A, B \in \mathscr{F}$, show that $A \triangle B \in \mathscr{F}$.
4. If $A_1, A_2, \ldots, A_m \in \mathscr{F}$ and k is a positive integer, show that the set of points in Ω which belong to exactly k of the A's belongs to $\mathscr{F}$ (the previous exercise is the case when $m = 2$ and $k = 1$).
5. Show that if Ω is a finite set then $\mathscr{F}$ contains an even number of subsets of Ω.

1.3 Probabilities

From our experiment $\mathscr{E}$, we have so far constructed a sample space Ω and an event space $\mathscr{F}$ associated with $\mathscr{E}$, but there has been no mention yet of probabilities. The third thing which we do is to allocate probabilities to each event in $\mathscr{F}$, writing $\mathsf{P}(A)$ for the probability of the event A. We shall assume that this can be done in such a way that the probability function P satisfies certain intuitively attractive conditions:

(i) each event A in the event space should have a probability $\mathsf{P}(A)$ which lies between 0 and 1;

(ii) the event Ω, that 'something happens', should have probability 1, and the event $\varnothing$, that 'nothing happens', should have probability 0;

(iii) if A and B are disjoint events (so that $A \cap B = \varnothing$) then $\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B)$.

We collect these conditions into a formal definition as follows.

A mapping $\mathsf{P} : \mathscr{F} \to \mathbb{R}$ is called a *probability measure* on $(\Omega, \mathscr{F})$ if

(7) $$\mathsf{P}(A) \geq 0 \quad \text{for all} \quad A \in \mathscr{F},$$

(8) $$\mathsf{P}(\Omega) = 1 \quad \text{and} \quad \mathsf{P}(\varnothing) = 0,$$

(9) if $A_1, A_2, \ldots$ are disjoint events in $\mathscr{F}$ (so that $A_i \cap A_j = \varnothing$ whenever $i \neq j$) then

$$\mathsf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathsf{P}(A_i).$$

We emphasize that a probability measure P on $(\Omega, \mathscr{F})$ is defined only on those subsets of Ω which lie in $\mathscr{F}$. The second part of condition (8) is superfluous in the above definition; to see this, note that $\varnothing$ and Ω are disjoint events with union $\Omega \cup \varnothing = \Omega$ and so

$$\mathsf{P}(\Omega) = \mathsf{P}(\Omega \cup \varnothing) = \mathsf{P}(\Omega) + \mathsf{P}(\varnothing) \qquad \text{by (9).}$$

Condition (9) requires that the probability of the union of a countable† collection of non-overlapping sets is the sum of the individual probabilities.

Example 10 Let Ω be a set and A be a proper subset of Ω (so that $A \neq \varnothing, \Omega$). If $\mathscr{F}$ is the event space $\{\varnothing, A, \Omega \backslash A, \Omega\}$ then all probability measures on $(\Omega, \mathscr{F})$ have the form

$$\mathsf{P}(\varnothing) = 0, \qquad \mathsf{P}(A) = p,$$
$$\mathsf{P}(\Omega \backslash A) = 1 - p, \qquad \mathsf{P}(\Omega) = 1,$$

for some p satisfying $0 \leq p \leq 1$. □

†A set S is called *countable* if it may be put in one–one correspondence with a subset of the natural numbers $\{1, 2, 3, \ldots\}$.

Example 11 Let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ be a finite set of exactly N points, and let $\mathcal{F}$ be the power set of Ω. It is easy to check that the function P defined by†

$$\mathrm{P}(A) = \frac{1}{N}|A| \qquad \text{for } A \in \mathcal{F}$$

is a probability measure on $(\Omega, \mathcal{F})$. □

Exercises 6. Let $p_1, p_2, \ldots, p_N$ be non-negative numbers such that $p_1 + p_2 + \cdots + p_N = 1$, and let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$, with $\mathcal{F}$ the power set of Ω, as in Example 11 above. Show that the function Q given by

$$Q(A) = \sum_{i:\omega_i \in A} p_i \qquad \text{for } A \in \mathcal{F},$$

is a probability measure on $(\Omega, \mathcal{F})$. Is Q a probability measure if $\mathcal{F}$ is not the power set of Ω but merely some event space of subsets of Ω?

1.4 Probability spaces

We now combine the previous ideas and define a *probability space* to be a triple $(\Omega, \mathcal{F}, \mathrm{P})$ of objects such that

(i) Ω is a set,
(ii) $\mathcal{F}$ is an event space of subsets of Ω,
(iii) P is a probability measure on $(\Omega, \mathcal{F})$.

There are many elementary consequences of the axioms which underlie this definition, and we describe some of these. Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space.

(12) If $A, B \in \mathcal{F}$ then‡ $A \setminus B \in \mathcal{F}$.

Proof The complement of $A \setminus B$ equals $(\Omega \setminus A) \cup B$, which is the union of events and is therefore an event. Hence $A \setminus B$ is an event, by (2). □

(13) If $A_1, A_2, \ldots \in \mathcal{F}$ then $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$.

Proof The complement of $\bigcap_{i=1}^{\infty} A_i$ equals $\bigcup_{i=1}^{\infty} (\Omega \setminus A_i)$ which is the union of the complements of events and is therefore an event. Hence the intersection of the A's is an event also, as before. □

(14) If $A \in \mathcal{F}$ then $\mathrm{P}(A) + \mathrm{P}(\Omega \setminus A) = 1$.

† The *cardinality* $|A|$ of a set A is the number of points in A.

‡ $A \setminus B = A \cap (\Omega \setminus B)$ is the set of points in A which are not in B.

Proof A and $\Omega\backslash A$ are disjoint events with union Ω, and so

$$1 = \mathsf{P}(\Omega) = \mathsf{P}(A) + \mathsf{P}(\Omega\backslash A). \qquad \square$$

(15) If $A, B \in \mathcal{F}$ then $\mathsf{P}(A \cup B) + \mathsf{P}(A \cap B) = \mathsf{P}(A) + \mathsf{P}(B)$.

Proof The set A is the union of the disjoint sets $A\backslash B$ and $A \cap B$, and hence

$$\mathsf{P}(A) = \mathsf{P}(A\backslash B) + \mathsf{P}(A \cap B) \qquad \text{by (9).}$$

A similar remark holds for the set B, giving that

$$\begin{aligned}\mathsf{P}(A) + \mathsf{P}(B) &= \mathsf{P}(A\backslash B) + 2\mathsf{P}(A \cap B) + \mathsf{P}(B\backslash A)\\ &= \mathsf{P}((A\backslash B) \cup (A \cap B) \cup (B\backslash A)) + \mathsf{P}(A \cap B) \qquad \text{by (9)}\\ &= \mathsf{P}(A \cup B) + \mathsf{P}(A \cap B). \qquad \square\end{aligned}$$

(16) If $A, B \in \mathcal{F}$ and $A \subseteq B$ then $\mathsf{P}(A) \leq \mathsf{P}(B)$.

Proof $\mathsf{P}(B) = \mathsf{P}(A) + \mathsf{P}(B\backslash A) \geq \mathsf{P}(A)$. $\square$

It is often useful to draw a Venn diagram when working with probabilities. For example, to show the formula in (15) above we might draw the diagram in Fig. 1.1, and note that the probability of $A \cup B$ is the sum of $\mathsf{P}(A)$ and $\mathsf{P}(B)$ minus $\mathsf{P}(A \cap B)$, since this latter probability is counted twice in the simple sum $\mathsf{P}(A) + \mathsf{P}(B)$.

Exercises In these exercises $(\Omega, \mathcal{F}, \mathsf{P})$ is a probability space.

7. If $A, B \in \mathcal{F}$, show that

$$\mathsf{P}(A\backslash B) = \mathsf{P}(A) - \mathsf{P}(A \cap B).$$

8. If $A, B, C \in \mathcal{F}$, show that

$$\begin{aligned}\mathsf{P}(A \cup B \cup C) = \mathsf{P}(A) + \mathsf{P}(B) + \mathsf{P}(C) - \mathsf{P}(A \cap B)\\ - \mathsf{P}(A \cap C) - \mathsf{P}(B \cap C) + \mathsf{P}(A \cap B \cap C).\end{aligned}$$

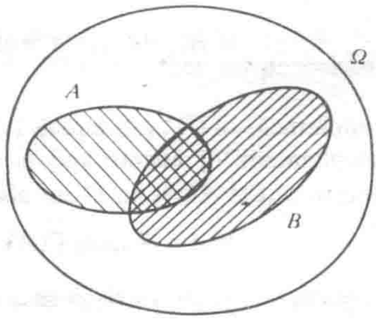

Fig. 1.1 A Venn diagram which illustrates the fact that $\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cap B)$

9. Let A, B, C be three events such that

$$\mathsf{P}(A)=\tfrac{5}{10},\quad \mathsf{P}(B)=\tfrac{7}{10},\quad \mathsf{P}(C)=\tfrac{6}{10},$$
$$\mathsf{P}(A\cap B)=\tfrac{3}{10},\quad \mathsf{P}(B\cap C)=\tfrac{4}{10},\quad \mathsf{P}(A\cap C)=\tfrac{2}{10},$$
$$\mathsf{P}(A\cap B\cap C)=\tfrac{1}{10}.$$

By drawing a Venn diagram or otherwise, find the probability that exactly two of the events A, B, C occur.

10. A fair coin is tossed 10 times (so that heads appears with probability $\frac{1}{2}$ at each toss). Describe the appropriate probability space in detail for the two cases when
(i) the outcome of every toss is of interest,
(ii) only the total number of tails is of interest.
In the first case your event space should have $2^{2^{10}}$ events, but in the second case it should have only 2^{11} events.

1.5 Discrete sample spaces

Let $\mathscr{E}$ be an experiment with probability space $(\Omega, \mathscr{F}, \mathsf{P})$. The structure of this space depends greatly upon whether Ω is a countable set (that is, a finite or countably infinite set) or an uncountable set. If Ω is a countable set then we normally take $\mathscr{F}$ to be the set of *all* subsets of Ω, for the following reason. Suppose that $\Omega = \{\omega_1, \omega_2, \ldots\}$ and, for each $\omega \in \Omega$, we are interested in whether or not this given ω is the actual outcome of $\mathscr{E}$; then we require that each singleton set $\{\omega\}$ belongs to $\mathscr{F}$. Let $A \subseteq \Omega$. Then A is countable (since Ω is countable) and so A may be expressed as the union of the countably many ω's which belong to A, giving that $A = \bigcup_{\omega\in A}\{\omega\} \in \mathscr{F}$ by (3). The probability $\mathsf{P}(A)$ of the event A is determined by the collection $\{\mathsf{P}(\{\omega\}) : \omega \in \Omega\}$ of probabilities since, by (9),

$$\mathsf{P}(A) = \sum_{\omega\in A} \mathsf{P}(\{\omega\}).$$

We usually write $\mathsf{P}(\omega)$ for the probability $\mathsf{P}(\{\omega\})$ of an event containing only one point in Ω.

Example 17 *Equiprobable outcomes*. If $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ and $\mathsf{P}(\omega_i) = \mathsf{P}(\omega_j)$ for all i and j, then $\mathsf{P}(\omega) = N^{-1}$ for all $\omega \in \Omega$, and $\mathsf{P}(A) = |A|/N$ for all $A \subseteq \Omega$. □

Example 18 *Random integers*. There are "intuitively-clear" statements which are without meaning in probability theory, and here is an example: *if we pick a positive integer at random, then it is an even integer with probability* $\frac{1}{2}$. Interpreting "at random" to mean that each positive integer is equally likely to be picked, then this experiment would