SPRINGER TEXTS IN STATISTICS

PROBABILITY

概率论

JIM PITMAN

Springer 世界图书出版公司
www.wpcbj.com.cn

JIM PITMAN

PROBABILITY

Springer

图书在版编目（CIP）数据

概率论 = Probability：英文/（美）皮特曼（Pitman，J.）著.
—北京：世界图书出版公司北京公司，2009.5
ISBN 978-7-5100-0465-0

Ⅰ.概… Ⅱ.皮… Ⅲ.概率论—英文 Ⅳ.0211

中国版本图书馆 CIP 数据核字（2009）第 055597 号

书　　名：	Probability
作　　者：	Jim Pitman
中 译 名：	概率论
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：	16 开
印　　张：	36
版　　次：	2009 年 05 月
版权登记：	图字：01-2009-1100
书　　号：	978-7-5100-0465-0/O·680　　定　　价：89.00 元

Jim Pitman
Department of Statistics
University of California
Berkeley, Ca 94720

Mathematical Subject Classification (1992): 60-01

Library of Congress Cataloging-in-Publication Data
Pitman, Jim.
Probability / Jim Pitman.
p. cm. — (Springer texts in statistics)
Includes bibliographical references and index.
ISBN 0-387-97974-3 (U.S.) — ISBN 3-540-97974-3
1. Probabilities. I. Title. II. Series.
QA273.P493 1993
519.2—dc20 92-39051

ISBN 0-387-97974-3

springer.com

Springer Texts in Statistics

Alfred	Elements of Statistics for the Life and Social Sciences
Berger	Introduction to Probability and Stochastic Processes, Second Edition
Bilodeau and Brenner	Theory of Multivariate Statistics
Blom	Probability and Statistics: Theory and Applications
Brockwell and Davis	An Introduction to Time Series and Forecasting
Chow and Teicher	Probability Theory: Independence, Interchangeability, Martingales, Third Edition
Christensen	Advanced Linear Modeling: Multivariate, Times Series, and Spatial Data; Nonparametric Regression and Response Surface Maximization, Second Edition
Christensen	Log-Linear Models and Logistic Regression, Second Edition
Christensen	Plane Answers to Complex Questions: The Theory of Linear Models, Second Edition
Creighton	A First Course in Probability Models and Statistical Inference
Davis	Statistical Methods for the Analysis of Repeated Measurements
Dean and Voss	Design and Analysis of Experiments
du Toit, Steyn, and Stumpf	Graphical Exploratory Data Analysis

Preface

Preface to the Instructor

This is a text for a one-quarter or one-semester course in probability, aimed at students who have done a year of calculus. The book is organized so a student can learn the fundamental ideas of probability from the first three chapters without reliance on calculus. Later chapters develop these ideas further using calculus tools.

The book contains more than the usual number of examples worked out in detail. It is not possible to go through all these examples in class. Rather, I suggest that you deal quickly with the main points of theory, then spend class time on problems from the exercises, or your own favorite problems. The most valuable thing for students to learn from a course like this is how to pick up a probability problem in a new setting and relate it to the standard body of theory. The more they see this happen in class, and the more they do it themselves in exercises, the better.

The style of the text is deliberately informal. My experience is that students learn more from intuitive explanations, diagrams, and examples than they do from theorems and proofs. So the emphasis is on problem solving rather than theory.

Order of Topics. The basic rules of probability all appear in Chapter 1. Intuition for probabilities is developed using Venn and tree diagrams. Only finite additivity of probability is treated in this chapter. Discussion of countable additivity is postponed to Section 3.4. Emphasis in Chapter 1 is on the concept of a probability distribution and elementary applications of the addition and multiplication rules. Combinatorics appear via study of the binomial and hypergeometric distributions in Chapter 2. The concepts of mean and standard deviation appear in a preliminary form in this chapter, motivated by the normal approximation, without the notation of random variables. These concepts are then developed for discrete random variables in Chapter 3. The main object of the first three chapters is to get to the circle of ideas around the normal approximation for sums of independent random variables. This is achieved by Section 3.3. Sections 3.4 and 3.5 deal with the standard distributions on the non-negative integers. Conditional distributions and expectations, covariance and correlation for discrete distributions are postponed to Chapter 6, nearby treatment of the same concepts for continuous distributions. The discrete theory could be done right after Chapter 3, but it seems best to get as quickly as possible to continuous things. Chapters 4 and 5 treat continuous distributions assuming a calculus background. The main emphasis here is on how to do probability calculations rather than rigorous development of the theory. In particular, differential calculations are used freely from Section 4.1 on, with only occasional discussion of the limits involved.

Optional Sections. These are more demanding mathematically than the main stream of ideas.

Terminology. Notation and terms are standard, except that *outcome space* is used throughout instead of sample space. Elements of an outcome space are called *possible outcomes.*

Pace. The earlier chapters are easier than later ones. It is important to get quickly through Chapters 1 and 2 (no more than three weeks). Chapter 3 is more substantial and deserves more time. The end of Chapter 3 is the natural time for a midterm examination. This can be as early as the sixth week. Chapters 4, 5, and 6 take time, much of it spent teaching calculus.

Preface to the Student

Prerequisites. This book assumes some background of mathematics, in particular, calculus. A summary of what is taken for granted can be found in Appendices I to III. Look at these to see if you need to review this material, or perhaps take another mathematics course before this one.

How to read this book. To get most benefit from the text, work one section at a time. Start reading each section by skimming lightly over it. Pick out the main ideas, usually boxed, and see how some of the examples involve these ideas. Then you may already be able to do some of the first exercises at the end of the section, which you should try as soon as possible. Expect to go back and forth between the exercises and the section several times before mastering the material.

Exercises. Except perhaps for the first few exercises in a section, do not expect to be able to plug into a formula or follow exactly the same steps as an example in the text. Rather, expect some variation on the main theme, perhaps a combination with ideas of a previous section, a rearrangement of the formula, or a new setting of the same principles. Through working problems you gain an active understanding of the concepts. If you find a problem difficult, or can't see how to start, keep in mind that it will always be related to material of the section. Try re-reading the section with the problem in mind. Look for some similarity or connection to get started. Can you express the problem in a different way? Can you identify relevant variables? Could you draw a diagram? Could you solve a simpler problem? Could you break up the problem into simpler parts? Most of the problems will yield to this sort of approach once you have understood the basic ideas of the section. For more on problem-solving techniques, see the book *How to Solve It* by G. Polya (Princeton University Press).

Solutions. Brief solutions to most odd numbered exercises appear at the end of the book.

Chapter Summaries. These are at the end of every chapter.

Review Exercises. These come after the summaries at the end of every chapter. Try these exercises when reviewing for an examination. Many of these exercises combine material from previous chapters.

Distribution Summaries. These set out the properties of the most important distributions. Familiarity with these properties reduces the amount of calculation required in many exercises.

Examinations. Some midterm and final examinations from courses taught from this text are provided, with solutions a few pages later.

Acknowledgments

Thanks to many students and instructors who have read preliminary versions of this book and provided valuable feedback. In particular, David Aldous, Peter Bickel, Ed Chow, Steve Evans, Roman Fresnedo, David Freedman, Alberto Gandolfi, Hank Ibser, Barney Krebs, Bret Larget, Russ Lyons, Lucien Le Cam, Maryse Loranger, Deborah Nolan, David Pollard, Roger Purves, Joe Romano, Tom Salisbury, David Siegmund, Anne Sheehy, Philip Stark, and Ruth Williams made numerous corrections and suggestions. Thanks to Ani Adhikari, David Aldous, David Blackwell, David Brillinger, Lester Dubins, Persi Diaconis, Mihael Perman and Robin Pemantle for providing novel problems. Thanks to Ed Chow, Richard Cutler, Bret Larget, Kee Won Lee, and Arunas Rudvalis who helped with solutions to the problems. Thanks to Carol Block and Chris Colbert who typed an early draft. Special thanks to Ani Adhikari, who provided enormous assistance with all aspects of this book. The graphics are based on her library of mathematical graphics routines written in PostScript. The graphics were further developed by Jianqing Fan, Ed Chow, and Ofer Licht. Thanks to Ed Chow for organizing drafts of the book on the computer, and to Bret Larget and Ofer Licht for their assistance in final preparation of the manuscript.

Jim Pitman, January 1993

Contents

1 Introduction

This chapter introduces the basic concepts of probability theory. These are the notions of:

- ***an outcome space***, or set of all possible outcomes of some kind;
- ***events*** represented mathematically as ***subsets*** of an outcome space; and
- ***probability*** as a function of these events or subsets.

The word "event" is used here for the kind of thing that has a probability, like getting a six when you roll a die, or getting five heads in a row when you toss a coin five times. The probability of an event is a measure of the likelihood or chance that the event occurs, on a scale from 0 to 1. Section 1.1 introduces these ideas in the simplest setting of equally likely outcomes. Section 1.2 treats two important interpretations of probability: approximation of long-run frequencies and subjective judgment of uncertainty. However probabilities are understood or interpreted, it is generally agreed that they must satisfy certain rules, in particular the basic ***addition rule***. This rule is built in to the idea of a ***probability distribution***, introduced in Section 1.3. The concepts of ***conditional probability***, and ***independence*** appear in Section 1.4. These concepts are further developed in Section 1.5 on ***Bayes' rule*** and Section 1.6 on sequences of events.

1.1 Equally Likely Outcomes

Probability is an extension of the idea of a ***proportion***, or ratio of a part to a whole. If there are 300 men and 700 women in a group, the proportion of men in the group is

$$\frac{300}{300+700} = 0.3 = 30\%$$

Suppose now that someone is picked at random from this population of men and women. For example, the choice could be made by drawing at random from a box of 1000 tickets, with different tickets corresponding to different people. It would then be said that

- the ***probability*** of choosing a woman is 70%;
- the ***odds in favor*** of choosing a woman are 7 to 3 (or 7/3 to 1); and
- the ***odds against*** choosing a woman are 3 to 7 (or 3/7 to 1).

So in thinking about someone picked at random from a population, a proportion in the population becomes a probability, and something like a sex ratio becomes an odds ratio.

There is an implicit assumption here: ***"picked at random"*** means everyone has the same chance of being chosen. In practice, for a draw at random from a box, this means the tickets are similar, and well mixed up before the draw. Intuitively, we say different tickets are ***equally likely***, or that they have the ***same chance***. In other words, the draw is ***honest***, ***fair***, or ***unbiased***. In more mathematical language, the ***probability*** of each ticket is the same, namely, 1/1000 for an assumed total of 1000 tickets.

For the moment, take for granted this intuitive idea of equally likely outcomes. Represent the set of all possible outcomes of some situation or experiment by Ω (capital omega, the last letter in the Greek alphabet). For instance, Ω would be the set of 1000 people (or the 1000 corresponding tickets) in the previous example. Or $\Omega = \{head, tail\}$ for the result of tossing a coin, or $\Omega = \{1, 2, 3, 4, 5, 6\}$ for rolling an ordinary six-sided die. The set Ω is called the ***outcome space***. Something that might or might not happen, depending on the outcome, is called an ***event***. Examples of events are "person chosen at random is a woman", "coin lands heads", "die shows an even number". An event A is represented mathematically by a subset of the outcome space Ω. For the examples above, A would be the set of women in the population, the set comprising the single outcome $\{head\}$, and the set of even numbers $\{2, 4, 6\}$.

Let $\#(A)$ be the number of outcomes in A. Informally, this is the number of chances for A to occur, or the number of different ways A can happen. Assuming equally likely outcomes, the probability of A, denoted $P(A)$, is defined to be the corresponding proportion of outcomes. This would be 700/1000, 1/2, and 3/6 in the three examples.

Equally Likely Outcomes

If all outcomes in a finite set Ω are equally likely, the probability of A is the number of outcomes in A divided by the total number of outcomes:

$$P(A) = \frac{\#(A)}{\#(\Omega)}$$

Probabilities defined by this formula for equally likely outcomes are fractions between 0 and 1. The number 1 represents certainty: $P(\Omega) = 1$. The number 0 represents impossibility: $P(A) = 0$ if there is no way that A could happen. Then A corresponds to the empty set, or set with no elements, denoted $\emptyset$. So $P(\emptyset) = 0$. Intermediate probabilities may be understood as various degrees of certainty.

Example 1. Picking a number between 1 and 100.

Suppose there is a box of 100 tickets marked $1, 2, 3, \ldots, 100$. A ticket is drawn at random from the box. Here are some events, with their descriptions as subsets and their probabilities obtained by counting. All possible numbers are assumed equally likely.

Event	Subset of $\{1, 2, \ldots, 100\}$	Probability
the number drawn has one digit	$\{1, 2, \ldots, 9\}$	9%
the number drawn has two digits	$\{10, 11, \ldots, 99\}$	90%
the number drawn is less than or equal to the number k	$\{1, 2, \ldots, k\}$	k%
the number drawn is strictly greater than k	$\{k+1, \ldots, 100\}$	$(100 - k)$%
the sum of the digits in the number drawn is equal to 3	$\{3, 12, 21, 30\}$	4%

Example 2. Rolling two dice.

A fair die is rolled and the number on the top face is noted. Then another fair die is rolled, and the number on its top face is noted.

Problem 1. What is the probability that the sum of the two numbers showing is 5?

Solution. Think of each possible outcome as a pair of numbers. The first element of the pair is the first number rolled, and the second element is the second number rolled. The first number can be any integer between 1 and 6, and so can the second number. Here are all the possible ways the dice could roll:

$$\begin{array}{cccccc} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{array}$$

The collection of these 36 pairs forms the outcome space Ω. Assume these 36 outcomes are equally likely. The event "the sum of the two numbers showing is 5" is represented by the subset $\{(1,4),(4,1),(2,3),(3,2)\}$. Since this subset has 4 elements,

$$P(\text{sum of two numbers showing is 5}) = \frac{4}{36} = \frac{1}{9}$$

Problem 2. What is the probability that one of the dice shows 2, and the other shows 4?

Solution. The subset corresponding to this event is $\{(2,4),(4,2)\}$. So the required probability is $2/36 = 1/18$.

Problem 3. What is the probability that the second number rolled is greater than the first number?

Solution. Look at the pairs in the outcome space Ω above, to see that this event corresponds to the subset

$$\begin{array}{ccccc} (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ & (2,3) & (2,4) & (2,5) & (2,6) \\ & & (3,4) & (3,5) & (3,6) \\ & & & (4,5) & (4,6) \\ & & & & (5,6) \end{array}$$

These are the pairs above the diagonal in Ω. There are 15 such pairs, so the probability that the second number rolled is greater than the first is $15/36$.

Problem 4. What is the probability that the second number rolled is less than the first number rolled?

Solution. The subset of Ω corresponding to this event is the set of pairs below the diagonal. There are just as many pairs below the diagonal as above. So the probability that the second number rolled is less than the first number is also $15/36$.

Example 3. Rolling two n-sided dice.

Repeat the above example for two rolls of a die with n faces numbered $1, 2, \ldots, n$, assuming $n \geq 4$.

Problem 1. Find the chance that the sum is 5.

Solution. Now there are n^2 possible pairs instead of $6^2 = 36$. But there are still just 4 possible pairs with sum 5. Hence

$$P(\text{sum is } 5) = 4/n^2$$

Problem 2. Find the chance that one roll is a 2, the other is a 4.

Solution. By the same argument $P(\text{a 2 and a 4}) = 2/n^2$.

Problem 3. Find the chance that the second number is greater than the first.

Solution. Now all pairs above the diagonal must be counted in an $n \times n$ matrix of pairs. There are no such pairs in the bottom row, 1 in the next, 2 in the next, and so on up to $(n-1)$ pairs in the top row, so the number of pairs above the diagonal is

$$\#(\text{above}) = 1 + 2 + 3 + \cdots + (n-1) = \frac{1}{2}n(n-1)$$

pairs altogether (see Appendix 2 on sums.) This gives

$$P(\text{second number is greater}) = \frac{\#(\text{above})}{\#(\text{total})} = \frac{\frac{1}{2}n(n-1)}{n^2} = \frac{1}{2}\left(1 - \frac{1}{n}\right)$$

Remark. Here is another way to find #(above), which gives the formula for the sum of the first $n-1$ integers (used above) as a consequence. Since

$$\#(\text{below}) + \#(\text{above}) + \#(\text{diagonal}) = \#(\text{total}) = n^2$$

and #(below) = #(above) by symmetry, and $\#(\text{diagonal}) = n$,

$$\#(\text{above}) = (n^2 - n)/2 = \frac{1}{2}n(n-1)$$

Problem 4. Find the chance that the first number is bigger.

Solution. Same as above, by the symmetry used already.

Note. As $n \to \infty$,

$$P(\text{two numbers are equal}) = \frac{\#(\text{diagonal})}{\#(\text{total})} = \frac{n}{n^2} = \frac{1}{n} \to 0$$

hence

$$P(\text{second bigger}) = P(\text{first bigger}) = \frac{1}{2}\left(1 - \frac{1}{n}\right) \to \frac{1}{2}$$

Odds

In a setting of equally likely outcomes, odds in favor of A give the ratio of the number of ways that A happens to the number of ways that A does not happen. The same ratio is obtained using probabilities instead of numbers of ways. Odds against A give the inverse ratio. More generally, just about any ratio of chances or probabilities can be called an odds ratio.

Gamblers are concerned with another sort of odds, which must be distinguished from odds defined as a ratio of chances. These are the odds offered by a casino or bookmaker in a betting contract, called here ***payoff odds*** to make the distinction clear. If you place a \$1 bet on an event A, and the payoff odds against A are 10 to 1, you stand to win \$10 if A occurs, and lose your \$1 if A does not occur. In a casino you first pay your \$1. If A occurs you get back a total of \$11. This is your winnings of \$10 plus your \$1 back. If A does not occur, the casino keeps your \$1. The price of \$1 is ***your stake***, the \$10 is the ***casino's stake***, and the \$11 is the ***total stake***.

The connection between payoff odds and chance odds is an ancient principle of gambling, understood long before mathematicians decided that probabilities were best represented as numbers between 0 and 1. Around 1584, a colorful gambler and scholar of the Italian Renaissance, named Cardano, wrote a book on games of chance. Considering the roll of a die, Cardano said,

> *I am as able to throw a 1, 3 or 5 as 2, 4 or 6. The wagers are therefore laid in accordance with this equality if the die is honest, and if not, they are made so much the larger or smaller in proportion to the departure from true equality.*

First there is the idea of equally likely outcomes, then a heuristic connecting payoff odds and chance odds:

The Fair Odds Rule

In a fair bet, the payoff odds equal the chance odds.

That is to say, in a fair bet on an event A, where you win if A occurs and the casino wins otherwise, the ratio of your stake to the casino's stake should be the ratio of probabilities $P(A)$ to $1 - P(A)$. Put another way, your stake should be proportion $P(A)$ of the total stake.