SECOND EDITION

PROBABILITY, STATISTICS, AND STOCHASTIC PROCESSES

PETER OLOFSSON
MIKAEL ANDERSSON

WILEY

PROBABILITY, STATISTICS, AND STOCHASTIC PROCESSES

Second Edition

PETER OLOFSSON
MIKAEL ANDERSSON

WILEY
A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

Library of Congress Cataloging-in-Publication Data:

Olofsson, Peter, 1963–
Probability, statistics, and stochastic processes / Peter Olofsson, Mikael Andersson. – 2nd ed.
p. cm.
ISBN 978-0-470-88974-9 (hardback)
1. Stochastic processes–Textbooks. 2. Probabilities–Textbooks. 3. Mathematical statistics–Textbooks. I. Andersson, Mikael. II. Title.
QA274.O46 2012
519.2′3–dc23

2011040205

Printed in the United States of America

ISBN: 9780470889749

10 9 8 7 6 5 4 3 2 1

PREFACE

The second edition was motivated by comments from several users and readers that the chapters on statistical inference and stochastic processes would benefit from substantial extensions. To accomplish such extensions, I decided to bring in Mikael Andersson, an old friend and colleague from graduate school. Being five days my junior, he brought a vigorous and youthful perspective to the task and I am very pleased with the outcome. Below, Mikael will outline the major changes and additions introduced in the second edition.

PETER OLOFSSON

San Antonio, Texas, 2011

The chapter on statistical inference has been extended, reorganized, and split into two new chapters. Chapter 6 introduces the principles and concepts behind standard methods of statistical inference in general, while the important case of normally distributed samples is treated separately in Chapter 7. This is a somewhat different structure compared to most other textbooks in statistics since common methods such as t tests and linear regression come rather late in the text. According to my experience, if methods based on normal samples are presented too early in a course, they tend to overshadow other approaches such as nonparametric and Bayesian methods and students become less aware that these alternatives exist.

New additions in Chapter 6 include consistency of point estimators, large sample theory, bootstrap simulation, multiple hypothesis testing, Fisher's exact test, Kolmogorov–Smirnov test and nonparametric confidence intervals, as well as a discussion of informative versus noninformative priors and credibility intervals in Section 6.8.

Chapter 7 starts with a detailed treatment of sampling distributions, such as the t, chi-square, and F distributions, derived from the normal distribution. There are also new sections introducing one-way analysis of variance and the general linear model.

Chapter 8 has been expanded to include three new sections on martingales, renewal processes, and Brownian motion. These areas are of great importance in probability theory and statistics, but since they are based on quite extensive and advanced mathematical theory, we offer only a brief introduction here.

It has been a great privilege, responsibility, and pleasure to have had the opportunity to work with such an esteemed colleague and good friend. Finally, the joint project that we dreamed about during graduate school has come to fruition!

I also have a victim of preoccupation and absentmindedness, my beloved Eva whom I want to thank for her support and all the love and friendship we have shared and will continue to share for many days to come.

MIKAEL ANDERSSON

Stockholm, Sweden, 2011

PREFACE TO THE FIRST EDITION

THE BOOK

In November 2003, I was completing a review of an undergraduate textbook in probability and statistics. In the enclosed evaluation sheet was the question "Have you ever considered writing a textbook?" and I suddenly realized that the answer was "Yes," and had been for quite some time. For several years I had been teaching a course on calculus-based probability and statistics mainly for mathematics, science, and engineering students. Other than the basic probability theory, my goal was to include topics from two areas: statistical inference and stochastic processes. For many students this was the only probability/statistics course they would ever take, and I found it desirable that they were familiar with confidence intervals and the maximum likelihood method, as well as Markov chains and queueing theory. While there were plenty of books covering one area or the other, it was surprisingly difficult to find one that covered both in a satisfying way and on the appropriate level of difficulty. My solution was to choose one textbook and supplement it with lecture notes in the area that was missing. As I changed texts often, plenty of lecture notes accumulated and it seemed like a good idea to organize them into a textbook. I was pleased to learn that the good people at Wiley agreed.

It is now more than a year later, and the book has been written. The first three chapters develop probability theory and introduce the axioms of probability, random variables, and joint distributions. The following two chapters are shorter and of an "introduction to" nature: Chapter 4 on limit theorems and Chapter 5 on simulation. Statistical inference is treated in Chapter 6, which includes a section on Bayesian statistics, too often a neglected topic in undergraduate texts. Finally, in Chapter 7, Markov chains in discrete and continuous time are introduced. The reference list at the end of the book is by no means intended to be comprehensive; rather, it is a subjective selection of the useful and the entertaining.

Throughout the text I have tried to convey an intuitive understanding of concepts and results, which is why a definition or a proposition is often preceded by a short discussion or a motivating example. I have also attempted to make the exposition entertaining by choosing examples from the rich source of fun and thought-provoking probability problems. The data sets used in the statistics chapter are of three different kinds: real, fake but realistic, and unrealistic but illustrative.

THE PEOPLE

Most textbook authors start by thanking their spouses. I know now that this is far more than a formality, and I would like to thank $A\lambda\kappa\mu\acute{\eta}\nu\eta$ not only for patiently putting up with irregular work hours and an absentmindedness greater than usual but also for valuable comments on the aesthetics of the manuscript.

A number of people have commented on various parts and aspects of the book. First, I would like to thank Olle Häggström at Chalmers University of Technology, Göteborg, Sweden for valuable comments on all chapters. His remarks are always accurate and insightful, and never obscured by unnecessary politeness. Second, I would like to thank Kjell Doksum at the University of Wisconsin for a very helpful review of the statistics chapter. I have also enjoyed the Bayesian enthusiasm of Peter Müller at the University of Texas MD Anderson Cancer Center.

Other people who have commented on parts of the book or been otherwise helpful are my colleagues Dennis Cox, Kathy Ensor, Rudy Guerra, Marek Kimmel, Rolf Riedi, Javier Rojo, David W. Scott, and Jim Thompson at Rice University; Prof. Dr. R.W.J. Meester at Vrije Universiteit, Amsterdam, The Netherlands; Timo Seppäläinen at the University of Wisconsin; Tom English at Behrend College; Robert Lund at Clemson University; and Jared Martin at Shell Exploration and Production. For help with solutions to problems, I am grateful to several bright Rice graduate students: Blair Christian, Julie Cong, Talithia Daniel, Ginger Davis, Li Deng, Gretchen Fix, Hector Flores, Garrett Fox, Darrin Gershman, Jason Gershman, Shu Han, Shannon Neeley, Rick Ott, Galen Papkov, Bo Peng, Zhaoxia Yu, and Jenny Zhang. Thanks to Mikael Andersson at Stockholm University, Sweden for contributions to the problem sections, and to Patrick King at ODS–Petrodata, Inc. for providing data with a distinct Texas flavor: oil rig charter rates. At Wiley, I would like to thank Steve Quigley, Susanne Steitz, and Kellsee Chu for always promptly answering my questions. Finally, thanks to John Haigh, John Allen Paulos, Jeffrey E. Steif, and an anonymous Dutchman for agreeing to appear and be mildly mocked in footnotes.

Peter Olofsson

Houston, Texas, 2005

CONTENTS

1

BASIC PROBABILITY THEORY

1.1 INTRODUCTION

Probability theory is the mathematics of randomness. This statement immediately invites the question "What is randomness?" This is a deep question that we cannot attempt to answer without invoking the disciplines of philosophy, psychology, mathematical complexity theory, and quantum physics, and still there would most likely be no completely satisfactory answer. For our purposes, an informal definition of randomness as "what happens in a situation where we cannot predict the outcome with certainty" is sufficient. In many cases, this might simply mean lack of information. For example, if we flip a coin, we might think of the outcome as random. It will be either heads or tails, but we cannot say which, and if the coin is fair, we believe that both outcomes are equally likely. However, if we knew the force from the fingers at the flip, weight and shape of the coin, material and shape of the table surface, and several other parameters, we would be able to predict the outcome with certainty, according to the laws of physics. In this case we use randomness as a way to describe uncertainty due to lack of information.[1]

Next question: "What is probability?" There are two main interpretations of probability, one that could be termed "objective" and the other "subjective." The first is

[1]To quote the French mathematician Pierre-Simon Laplace, one of the first to develop a mathematical theory of probability: "Probability is composed partly of our ignorance, partly of our knowledge."

Probability, Statistics, and Stochastic Processes, Second Edition. Peter Olofsson and Mikael Andersson.

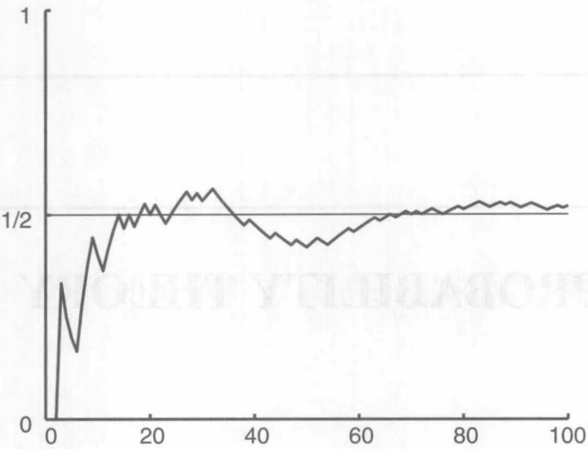

FIGURE 1.1 Consecutive relative frequencies of heads in 100 coin flips.

the interpretation of a probability as a *limit of relative frequencies*; the second, as a *degree of belief*. Let us briefly describe each of these.

For the first interpretation, suppose that we have an experiment where we are interested in a particular outcome. We can repeat the experiment over and over and each time record whether we got the outcome of interest. As we proceed, we count the number of times that we got our outcome and divide this number by the number of times that we performed the experiment. The resulting ratio is the *relative frequency* of our outcome. As it can be observed empirically that such relative frequencies tend to stabilize as the number of repetitions of the experiment grows, we might think of the limit of the relative frequencies as the probability of the outcome. In mathematical notation, if we consider n repetitions of the experiment and if S_n of these gave our outcome, then the relative frequency would be $f_n = S_n/n$, and we might say that the probability equals $\lim_{n\to\infty} f_n$. Figure 1.1 shows a plot of the relative frequency of heads in a computer simulation of 100 hundred coin flips. Notice how there is significant variation in the beginning but how the relative frequency settles in toward $\frac{1}{2}$ quickly.

The second interpretation, probability as a degree of belief, is not as easily quantified but has obvious intuitive appeal. In many cases, it overlaps with the previous interpretation, for example, the coin flip. If we are asked to quantify our degree of belief that a coin flip gives heads, where 0 means "impossible" and 1 means "with certainty," we would probably settle for $\frac{1}{2}$ unless we have some specific reason to believe that the coin is not fair. In some cases it is not possible to repeat the experiment in practice, but we can still imagine a sequence of repetitions. For example, in a weather forecast you will often hear statements like "there is a 30% chance of rain tomorrow." Of course, we cannot repeat the experiment; either it rains tomorrow or it does not. The 30% is the meteorologist's measure of the chance of rain. There is still a connection to the relative frequency approach; we can imagine a sequence of days with similar weather conditions, same time of year, and so on, and that in roughly 30% of the cases, it rains the following day.

The "degree of belief" approach becomes less clear for statements such as "the Riemann hypothesis is true" or "there is life on other planets." Obviously, these are statements that are either true or false, but we do not know which, and it is not unreasonable to use probabilities to express how strongly we believe in their truth. It is also obvious that different individuals may assign completely different probabilities.

How, then, do we actually *define* a probability? Instead of trying to use any of these interpretations, we will state a strict mathematical definition of probability. The interpretations are still valid to develop intuition for the situation at hand, but instead of, for example, *assuming* that relative frequencies stabilize, we will be able to *prove* that they do, within our theory.

1.2 SAMPLE SPACES AND EVENTS

As mentioned in the introduction, probability theory is a mathematical theory to describe and analyze situations where randomness or uncertainty are present. Any specific such situation will be referred to as a *random experiment*. We use the term "experiment" in a wide sense here; it could mean an actual physical experiment such as flipping a coin or rolling a die, but it could also be a situation where we simply observe something, such as the price of a stock at a given time, the amount of rain in Houston in September, or the number of spam emails we receive in a day. After the experiment is over, we call the result an *outcome*. For any given experiment, there is a set of possible outcomes, and we state the following definition.

Definition 1.1. The set of all possible outcomes in a random experiment is called the *sample space*, denoted S.

Here are some examples of random experiments and their associated sample spaces.

Example 1.1. Roll a die and observe the number.

Here we can get the numbers 1 through 6, and hence the sample space is

$$S = \{1, 2, 3, 4, 5, 6\}$$

□

Example 1.2. Roll a die repeatedly and count the number of rolls it takes until the first 6 appears.

Since the first 6 may come in the first roll, 1 is a possible outcome. Also, we may fail to get 6 in the first roll and then get 6 in the second, so 2 is also a possible outcome. If we continue this argument we realize that any positive integer is a possible outcome and the sample space is

$$S = \{1, 2, \ldots\}$$

the set of positive integers. □

Example 1.3. Turn on a lightbulb and measure its lifetime, that is, the time until it fails.

Here it is not immediately clear what the sample space should be since it depends on how accurately we can measure time. The most convenient approach is to note that the lifetime, at least in theory, can assume any nonnegative real number and choose as the sample space

$$S = [0, \infty)$$

where the outcome 0 means that the lightbulb is broken to start with. □

In these three examples, we have sample spaces of three different kinds. The first is *finite*, meaning that it has a finite number of outcomes, whereas the second and third are infinite. Although they are both infinite, they are different in the sense that one has its points separated, $\{1, 2, \ldots\}$ and the other is an entire continuum of points. We call the first type *countable infinity* and the second *uncountable infinity*. We will return to these concepts later as they turn out to form an important distinction.

In the examples above, the outcomes are always numbers and hence the sample spaces are subsets of the real line. Here are some examples of other types of sample spaces.

Example 1.4. Flip a coin twice and observe the sequence of heads and tails.

With H denoting heads and T denoting tails, one possible outcome is HT, which means that we get heads in the first flip and tails in the second. Arguing like this, there are four possible outcomes and the sample space is

$$S = \{HH, HT, TH, TT\}$$

□

Example 1.5. Throw a dart at random on a dartboard of radius r.

If we think of the board as a disk in the plane with center at the origin, an outcome is an ordered pair of real numbers (x, y), and we can describe the sample space as

$$S = \{(x, y) : x^2 + y^2 \leq r^2\}$$

□

Once we have described an experiment and its sample space, we want to be able to compute probabilities of the various things that may happen. What is the probability that we get 6 when we roll a die? That the first 6 does not come before the fifth roll? That the lightbulb works for at least 1500 h? That our dart hits the bull's eye? Certainly, we need to make further assumptions to be able to answer these questions, but before that, we realize that all these questions have something in common. They all ask for probabilities of either single outcomes or groups of outcomes. Mathematically, we can describe these as subsets of the sample space.

Definition 1.2. A subset of S, $A \subseteq S$, is called an *event*.

Note the choice of words here. The terms "outcome" and "event" reflect the fact that we are describing things that may happen in real life. Mathematically, these are described as elements and subsets of the sample space. This duality is typical for probability theory; there is a verbal description and a mathematical description of the same situation. The verbal description is natural when real-world phenomena are described and the mathematical formulation is necessary to develop a consistent theory. See Table 1.1 for a list of set operations and their verbal description.

Example 1.6. If we roll a die and observe the number, two possible events are that we get an odd outcome and that we get at least 4. If we view these as subsets of the sample space, we get

$$A = \{1, 3, 5\} \quad \text{and} \quad B = \{4, 5, 6\}$$

If we want to use the verbal description, we might write this as

$$A = \{\text{odd outcome}\} \quad \text{and} \quad B = \{\text{at least } 4\}$$

□

We always use "or" in its nonexclusive meaning; thus, "A or B occurs" includes the possibility that both occur. Note that there are different ways to express combinations of events; for example, $A \setminus B = A \cap B^c$ and $(A \cup B)^c = A^c \cap B^c$. The latter is known as one of *De Morgan's laws*, and we state these without proof together with some other basic set theoretic rules.

TABLE 1.1 Basic Set Operations and Their Verbal Description

Notation	Mathematical Description	Verbal Description
$A \cup B$	The union of A and B	A or B (or both) occurs
$A \cap B$	The intersection of A and B	Both A and B occur
A^c	The complement of A	A does not occur
$A \setminus B$	The difference between A and B	A occurs but not B
Ø	The empty set	Impossible event