Springer Texts in Statistics

Allan Gut

Probability: A Graduate Course

Springer

Allan Gut
Department of Mathematics
University of Uppsala
SE-751 06 Uppsala
Sweden

Library of Congress Cataloging-in-Publication Data
Gut, Allan, 1944–
Probability: a graduate course / Allan Gut.
p. cm. — (Springer texts in statistics)
Includes bibliographical references and index.
ISBN 0-387-22833-0 (alk. paper)
1. Probabilities. 2. Distribution (Probability theory) 3. Probabilistic number theory. I. Title. II. Series.
QA273.G869 2005
519.2—dc22 2004056469

ISBN 0-387-22833-0 Printed on acid-free paper.

Printed in the United States of America.

9 8 7 6 5 4 3 2 1 SPIN 10978764

Typesetting: Pages created by the author using Springer's TEX macro package.

springeronline.com

Preface

Toss a symmetric coin twice. What is the probability that both tosses will yield a head?

This is a well-known problem that anyone can solve. Namely, the probability of a head in each toss is $1/2$, so the probability of two consecutive heads is $1/2 \cdot 1/2 = 1/4$.

BUT! What did we do? What is involved in the solution? What are the arguments behind our computations? Why did we multiply the two halves connected with each toss?

This is reminiscent of the centipede[1] who was asked by another animal how he walks; he who has so many legs, in which order does he move them as he is walking? The centipede contemplated the question for a while, but found no answer. However, from that moment on he could no longer walk.

This book is written with the hope that we are not centipedes.

There exist two kinds of probabilists. One of them is the mathematician who views probability theory as a purely mathematical discipline, like algebra, topology, differential equations, and so on. The other kind views probability theory as the *mathematical modeling of random phenomena*, that is with a view toward applications, and as a companion to statistics, which aims at finding methods, principles and criteria in order to analyze data emanating from experiments involving random phenomena and other observations from the real world, with the ultimate goal of making wise decisions. I would like to think of myself as both.

What kind of a random process describes the arrival of claims at an insurance company? Is it one process or should one rather think of different processes, such as one for claims concerning stolen bikes and one for houses that have burnt down? How well should the DNA sequences of an accused offender and a piece of evidence match each other in order for a conviction? A

[1] Cent is 100, so it means an animal with 100 legs. In Swedish the name of the animal is *tusenfoting*, where "tusen" means 1000 and "fot" is foot; thus an animal with 1000 legs or feet.

milder version is how to order different species in a phylogenetic tree. What are the arrival rates of customers to a grocery store? How long are the service times? How do the clapidemia cells split? Will they create a new epidemic or can we expect them to die out? A classical application has been the arrivals of telephone calls to a switchboard and the duration of calls. Recent research and model testing concerning the Internet traffic has shown that the classical models break down completely and new thinking has become necessary. And, last but (not?) least, there are many games and lotteries.

The aim of this book is to provide the reader with a fairly thorough treatment of the main body of basic and classical probability theory, preceded by an introduction to the mathematics which is necessary for a solid treatment of the material. This means that we begin with basics from measure theory, such as σ-algebras, set theory, measurability (random variables) and Lebesgue integration (expectation), after which we turn to the Borel-Cantelli lemmas, inequalities, transforms and the three classical limit theorems: the law of large numbers, the central limit theorem and the law of the iterated logarithm. A final chapter on martingales – one of the most efficient, important, and useful tools in probability theory – is preceded by a chapter on topics that could have been included with the hope that the reader will be tempted to look further into the literature. The reason that these topics did not get a chapter of their own is that beyond a certain number of pages a book becomes deterring rather than tempting (or, as somebody said with respect to an earlier book of mine: "It is a nice format for bedside reading").

One thing that is *not* included in this book is a philosophical discussion of whether or not chance exist, whether or not randomness exists. On the other hand, probabilistic modeling is a wonderful, realistic, and efficient way to model phenomena containing uncertainties and ambiguities, regardless of whether or not the answer to the philosophical question is yes or no.

I remember having read somewhere a sentence like "There exist already so many textbooks [of the current kind], so, why do I write another one?" This sentence could equally well serve as an opening for the present book.

Luckily, I can provide an answer to that question. The answer is the short version of the story of the mathematician who was asked how one realizes that the fact he presented in his lecture (because this was really a he) was trivial. After 2 minutes of complete silence he mumbled

I know it's trivial, but I have forgotten why.

I strongly dislike the arrogance and snobbism that encompasses mathematics and many mathematicians. Books and papers are filled with expressions such as "it is easily seen", "it is trivial", "routine computations yield", and so on. The last example is sometimes modified into "routine, but tedious, computations yield". And we all know that behind things that are easily seen there may be years of thinking and/or huge piles of scrap notes that lead nowhere, and one sheet where everything finally worked out nicely.

Clearly, things become routine after many years. Clearly, facts become, at least *intuitively*, obvious after some decades. But in writing papers and books we try to help those who do not know yet, those who want to learn. We wish to attract people to this fascinating part of the world. Unfortunately though, phrases like the above ones are repellent, rather than being attractive. If a reader understands immediately that's fine. However, it is more likely that he or she starts off with something that either results in a pile of scrap notes or in frustration. Or both. And nobody is made happier, certainly not the reader. I have therefore avoided, or, at least, tried to avoid, expressions like the above unless they are adequate.

The main aim of a book is to be helpful to the reader, to help her or him to understand, to inform, to educate, and to attract (and not for the author to prove himself to the world). It is therefore essential to keep the flow, not only in the writing, but also in the reading. In the writing it is therefore of great importance to be rather extensive and not to leave too much to the (interested) reader.

A related aspect concerns the style of writing. Most textbooks introduce the reader to a number of topics in such a way that further insights are gained through exercises and problems, some of which are not at all easy to solve, let alone trivial. We take a somewhat different approach in that several such "would have been" exercises are given, together with their solutions as part of the ordinary text – which, as a side effect, reduces the number of exercises and problems at the end of each chapter. We also provide, at times, results for which the proofs consist of variations of earlier ones, and therefore are left as an exercise, with the motivation that doing almost the same thing as somebody else has done provides a much better understanding than reading, nodding and agreeing. I also hope that this approach creates the atmosphere of a dialogue rather than of the more traditional monologue (or sermon).

The ultimate dream is, of course, that this book contains no errors, no slips, no misprints. Henrik Wanntorp has gone over a substantial part of the manuscript with a magnifying glass, thereby contributing immensely to making that dream come true. My heartfelt thanks, Henrik. I also wish to thank Raimundas Gaigalas for several perspicacious remarks and suggestions concerning his favorite sections, and a number of reviewers for their helpful comments and valuable advice. As always, I owe a lot to Svante Janson for being available for any question at all times, and, more particularly, for always providing me with an answer. John Kimmel of Springer-Verlag has seen me through the process with a unique combination of professionalism, efficiency, enthusiasm and care, for which I am most grateful.

Finally, my hope is that the reader who has digested this book is ready and capable to attack any other text, for which a solid probabilistic foundation is necessary or, at least, desirable.

Uppsala *Allan Gut*
November 2004

Contents

Outline of Contents

In this extended list of contents, we provide a short expansion of the headings into a quick overview of the contents of the book.

Chapter 1. Introductory Measure Theory

The mathematical foundation of probability theory is measure theory and the theory of Lebesgue integration. The bulk of the introductory chapter is devoted to measure theory: sets, measurability, σ-algebras, and so on. We do not aim at a full course in measure theory, rather to provide enough background for a solid treatment of what follows.

Chapter 2. Random Variables

Having set the scene, the first thing to do is to forget probability spaces (!). More precisely, for modeling random experiments one is interested in certain specific quantities, called *random variables*, rather than in the underlying probability space itself. In Chapter 2 we introduce random variables and present the basic concepts, as well as concrete applications and examples of probability models. In particular, Lebesgue integration is developed in terms of expectation of random variables.

Chapter 3. Inequalities

Some of the most useful tools in probability theory and mathematics for proving finiteness or convergence of sums and integrals are *inequalities*. There exist many useful ones spread out in books and papers. In Chapter 3 we make an attempt to present a sizable amount of the most important inequalities.

Chapter 4. Characteristic Functions

Just as there are i.a. Fourier transforms that transform convolution of functions into multiplication of their corresponding transforms, there exist probabilistic transforms that "map" addition of independent random variables into multiplication of their transforms, the most prominent one being the characteristic function.

Chapter 5. Convergence

Once we know how to add random variables the natural problem is to investigate asymptotics. We begin by introducing some convergence concepts, prove uniqueness, after which we investigate how and when they imply each other. Other important problems are when, and to what extent, limits and expectations (limits and integrals) can be interchanged, and when, and to what extent, functions of convergent sequences converge to the function of the limit.

Chapter 6. The Law of Large Numbers

The law of large numbers states that (the distribution of) the arithmetic mean of a sequence of independent trials stabilizes around the center of gravity of the underlying distribution (under suitable conditions). There exist *weak* and *strong* laws and several variations and extensions of them. We shall meet some of them as well as some applications.

Chapter 7. The Central Limit Theorem

The central limit theorem, which (in its simplest form) states that if the variance is finite, then the arithmetic mean, properly rescaled, of a sequence of independent trials approaches a normal distribution as the number of observations increases. There exist many variations and generalizations, of the theorem, the central one being the Lindeberg-Lévy-Feller theorem. We also prove results on moment convergence, and rate results, the foremost one being the celebrated Berry-Esseen theorem.

Chapter 8. The Law of the Iterated Logarithm

This is a special, rather delicate and technical, and very beautiful, result, which provides precise bounds on the oscillations of sums of the above kind. The name obviously stems from the iterated logarithm that appears in the expression of the parabolic bound.

Chapter 9. Limit Theorems; Extensions and Generalizations

There are a number of additional topics that would fit well into a text like the present one, but for which there is no room. In this chapter we shall meet a number of them – stable distributions, domain of attraction, infinite divisibility, sums of dependent random variables, extreme value theory, the Stein-Chen method – in a somewhat more sketchy or introductory style. The reader who gets hooked on such a topic will be advised to some relevant literature (more can be found via the Internet).

Chapter 10. Martingales

This final chapter is devoted to one of the most central topics, not only in probability theory, but also in more traditional mathematics. Following some introductory material on conditional expectations and the definition of a martingale, we present several examples, convergence results, results for stopped martingales, regular martingales, uniformly integrable martingales, stopped random walks, and reversed martingales.

In Addition

A list of notation and symbols precedes the main body of text, and an appendix with some mathematical tools and facts, a bibliography, and an index conclude the book. References are provided for more recent results, for more nontraditional material, and to some of the historic sources, but in general not to the more traditional material. In addition to cited material, the list of references contains references to papers and books that are relevant without having been specifically cited.

Suggestions for a Course Curriculum

One aim with the book is that it should serve as a graduate probability course – as the title suggests. In the same way as the sections in Chapter 9 contain materials that no doubt would have deserved chapters of their own, Chapters 6,7, and 8 contain sections entitled "Some Additional Results and Remarks", in which a number of additional results and remarks are presented, results that are not as central and basic as earlier ones in those chapters.

An adequate course would, in my opinion, consist of Chapters 1-8, and 10, except for the sections "Some Additional Results and Remarks", plus a skimming through Chapter 9 at the level of the instructor's preferences.

Notation and Symbols

Ω	the sample space
ω	an elementary event
$\mathcal{F}$	the σ-algebra of events
x^+	$\max\{x, 0\}$
x^-	$-\min\{x, 0\}$
$[x]$	the integer part of x
$\log^+ x$	$\max\{1, \log x\}$
$\sim$	the ratio of the quantities on either side tends to 1
$\mathbb{N}$	the (positive) natural numbers
$\mathbb{Z}$	the integers
$\mathbb{R}$	the real numbers
$\mathcal{R}$	the Borel σ-algebra on $\mathbb{R}$
$\lambda(\,\cdot\,)$	Lebesgue measure
$\mathbb{Q}$	the rational numbers
$\mathbb{C}$	the complex numbers
C	the continuous functions
C_0	the functions in C tending to 0 at $\pm\infty$
C_B	the bounded continuous functions
$C[a, b]$	the functions in C with support on the interval $[a, b]$
D	the right-continuous, functions with left-hand limits
$D[a, b]$	the functions in D with support on the interval $[a, b]$
D^+	the non-decreasing functions in D
$\mathbb{J}_G$	the discontinuities of $G \in D$
$I\{A\}$	indicator function of (the set) A
$\#\{A\}$	number of elements in (cardinality of) A
$\|A\|$	number of elements in (cardinality of) A
A^c	complement of A
∂A	the boundary of A
$P(A)$	probability of A

$X, Y, Z, \ldots$	random variables
$F(x),\ F_X(x)$	distribution function (of X)
$X \in F$	X has distribution (function) F
$C(F_X)$	the continuity set of F_X
$p(x),\ p_X(x)$	probability function (of X)
$f(x),\ f_X(x)$	density (function) (of X)
$\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \ldots$	random (column) vectors
$\mathbf{X}', \mathbf{Y}', \mathbf{Z}', \ldots$	the transpose of the vectors
$F_{X,Y}(x,y)$	joint distribution function (of X and Y)
$p_{X,Y}(x,y)$	joint probability function (of X and Y)
$f_{X,Y}(x,y)$	joint density (function) (of X and Y)
$E,\ E\,X$	expectation (mean), expected value of X
$\mathrm{Var}\,,\ \mathrm{Var}\,X$	variance, variance of X
$\mathrm{Cov}\,(X,Y)$	covariance of X and Y
$\rho,\ \rho_{X,Y}$	correlation coefficient (between X and Y)
$\mathrm{med}\,(X)$	median of X
$g(t),\ g_X(t)$	(probability) generating function (of X)
$\psi(t),\ \psi_X(t)$	moment generating function (of X)
$\varphi(t),\ \varphi_X(t)$	characteristic function (of X)
$X \sim Y$	X and Y are equivalent random variables
$X \overset{a.s.}{=} Y$	X and Y are equal (point-wise) almost surely
$X \overset{d}{=} Y$	X and Y are equidistributed
$X_n \overset{a.s.}{\to} X$	X_n converges almost surely (a.s.) to X
$X_n \overset{p}{\to} \mathrm{X}$	X_n converges in probability to X
$X_n \overset{r}{\to} \mathrm{X}$	X_n converges in r-mean (L^r) to X
$X_n \overset{d}{\to} X$	X_n converges in distribution to X
$X_n \overset{a.s.}{\not\to}$	X_n does not converge almost surely
$X_n \overset{p}{\not\to}$	X_n does not converge in probability
$X_n \overset{r}{\not\to}$	X_n does not converge in r-mean (L^r)
$X_n \overset{d}{\not\to}$	X_n does not converge in distribution
$\Phi(x)$	standard normal distribution function
$\phi(x)$	standard normal density (function)
$F \in \mathcal{D}(G)$	F belongs to the domain of attraction of G
$g \in \mathcal{RV}\,(\rho)$	g varies regularly at infinity with exponent ρ
$g \in \mathcal{SV}$	g varies slowly at infinity

$\mathrm{Be}(p)$	Bernoulli distribution
$\beta(r,s)$	beta distribution
$\mathrm{Bin}(n,p)$	binomial distribution
$C(m,a)$	Cauchy distribution
$\chi^2(n)$	chi-square distribution
$\delta(a)$	one-point distribution
$\mathrm{Exp}(a)$	exponential distribution
$F(m,n)$	(Fisher's) F-distribution
$\mathrm{Fs}(p)$	first success distribution
$\Gamma(p,a)$	gamma distribution
$\mathrm{Ge}(p)$	geometric distribution
$H(N,n,p)$	hypergeometric distribution
$L(a)$	Laplace distribution
$\mathrm{LN}(\mu,\sigma^2)$	log-normal distribution
$N(\mu,\sigma^2)$	normal distribution
$N(0,1)$	standard normal distribution
$\mathrm{NBin}(n,p)$	negative binomial distribution
$\mathrm{Pa}(k,\alpha)$	Pareto distribution
$\mathrm{Po}(m)$	Poisson distribution
$\mathrm{Ra}(\alpha)$	Rayleigh distribution
$t(n)$	(Student's) t-distribution
$\mathrm{Tri}(a,b)$	triangular distribution on (a,b)
$U(a,b)$	uniform or rectangular distribution on (a,b)
$W(a,b)$	Weibull distribution
$X \in P(\theta)$	X has a P-distribution with parameter θ
$X \in P(\alpha,\beta)$	X has a P-distribution with parameters α and β

a.e.	almost everywhere
a.s.	almost surely
cf.	*confer*, compare, take counsel
i.a.	*inter alia*, among other things, such as
i.e.	*id est*, that is
i.o.	infinitely often
iff	if and only if
i.i.d.	independent, identically distributed
viz.	*videlicet*, in which
w.l.o.g.	without loss of generality
♠	hint for solving a problem
♣	bonus remark in connection with a problem
□	end of proof, definitions, exercises, remarks, etc.