Probability and Statistics by Example: I

Probability and statistics are as much about intuition and problem solving, as they are about theorem proving. Because of this, students can find it very difficult to make a successful transition from lectures to examinations to practice, since the problems involved can vary so much in nature. Since the subject is critical in many modern applications such as mathematical finance, quantitative management, telecommunications, signal processing, bioinformatics, as well as traditional ones such as insurance, social science and engineering, the authors have rectified deficiencies in traditional lecture-based methods by collecting together a wealth of exercises for which they've supplied complete solutions. These solutions are adapted to the needs and skills of students. To make it of broad value, the authors supply basic mathematical facts as and when they are needed, and have sprinkled some historical information throughout the text.

Probability and Statistics by Example

Volume I. Basic Probability and Statistics

Y. SUHOV
University of Cambridge

M. KELBERT
University of Wales–Swansea

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521847667

First published 2005

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this publication is available from the British Library

ISBN-13 978-0-521-84766-7 hardback
ISBN-10 0-521-84766-4 hardback
ISBN-13 978-0521-61233-3 paperback
ISBN-10 0-521-61233-0 paperback

Contents

Preface

The original motivation for writing this book was rather personal. The first author, in the course of his teaching career in the Department of Pure Mathematics and Mathematical Statistics (DPMMS), University of Cambridge, and St John's College, Cambridge, had many painful experiences when good (or even brilliant) students, who were interested in the subject of mathematics and its applications and who performed well during their first academic year, stumbled or nearly failed in the exams. This led to great frustration, which was very hard to overcome in subsequent undergraduate years. A conscientious tutor is always sympathetic to such misfortunes, but even pointing out a student's obvious weaknesses (if any) does not always help. For the second author, such experiences were as a parent of a Cambridge University student rather than as a teacher.

We therefore felt that a monograph focusing on Cambridge University mathematics examination questions would be beneficial for a number of students. Given our own research and teaching backgrounds, it was natural for us to select probability and statistics as the overall topic. The obvious starting point was the first-year course in probability and the second-year course in statistics. In order to cover other courses, several further volumes will be needed; for better or worse, we have decided to embark on such a project.

Thus our essential aim is to present the Cambridge University probability and statistics courses by means of examination (and examination-related) questions that have been set over a number of past years. Following the decision of the Board of the Faculty of Mathematics, University of Cambridge, we restricted our exposition to the Mathematical Tripos questions from the years 1992–1999. (The questions from 2000–2004 are available online at http://www.maths.cam.ac.uk/ppa/.) Next, we included some IA Probability regular example sheet questions from the years 1992–2003 (particularly those considered as difficult by students). Further, we included the problems from Specimen Papers issued in 1992 and used for mock examinations (mainly in the beginning of the 1990s) and selected examples from the 1992 list of so-called sample questions. A number of problems came from example sheets and examination papers from the University of Wales-Swansea.

Of course, Cambridge University examinations have never been easy. On the basis of examination results, candidates are divided into classes: first, second (divided into two categories: 2.1 and 2.2) and third; a small number of candidates fail. (In fact, a more detailed list ranking all the candidates in order is produced, but not publicly disclosed.) The examinations are officially called the 'Mathematical Tripos', after the three-legged stools on which candidates and examiners used to sit (sometimes for hours) during oral examinations in ancient times. Nowadays all examinations are written. The first-year of the three-year undergraduate course is called Part IA, the second Part IB and the third Part II.

For example, in May–June of 2003 the first-year mathematics students sat four examination papers; each lasted three hours and included 12 questions from two subjects. The following courses were examined: algebra and geometry, numbers and sets, analysis, probability, differential equations, vector calculus, and dynamics. All questions on a given course were put in a single paper, except for algebra and geometry, which appears in two papers. In each paper, four questions were classified as short (two from each of the two courses selected for the paper) and eight as long (four from each selected course). A candidate might attempt all four short questions and at most five long questions, no more than three on each course; a long question carries twice the credit of a short one. A calculation shows that if a student attempts all nine allowed questions (which is often the case), and the time is distributed evenly, a short question must be completed in 12–13 minutes and a long one in 24–25 minutes. This is not easy and usually requires special practice; one of the goals of this book is to assist with such a training programme.

The pattern of the second-year examinations has similarities but also differences. In June 2003, there were four IB Maths Tripos papers, each three hours long and containing nine or ten short and nine or ten long questions in as many subjects selected for a given paper. In particular, IB statistics was set in Papers 1, 2 and 4, giving a total of six questions. Of course, preparing for Part IB examinations is different from preparing for Part IA; we comment on some particular points in the corresponding chapters.

For a typical Cambridge University student, specific preparation for the examinations begins in earnest during the Easter (or Summer) Term (beginning in mid-April). Ideally, the work might start during the preceding five-week vacation. (Some of the examination work for Parts IB and II, the computational projects, is done mainly during the summer vacation period.) As the examinations approach, the atmosphere in Cambridge can become rather tense and nervous, although many efforts are made to diffuse the tension. Many candidates expend a great deal of effort in trying to calculate exactly how much work to put into each given subject, depending on how much examination credit it carries and how strong or weak they feel in it, in order to optimise their overall performance. One can agree or disagree with this attitude, but one thing seemed clear to us: if the students receive (and are able to digest) enough information about and insight into the level and style of the Tripos questions, they will have a much better chance of performing to the best of their abilities. At present, owing to great pressures on time and energy, most of them are not in a position to do so, and much is left to chance. We will be glad if this book helps to change this situation by alleviating pre-examination nerves and by stripping Tripos examinations of some of their mystery, at least in respect of the subjects treated here.

Thus, the first reason for this book was a desire to make life easier for the students. However, in the course of working on the text, a second motivation emerged, which we feel is of considerable professional interest to anyone teaching courses in probability and statistics. In 1991–2 there was a major change in Cambridge University to the whole approach to probabilistic and statistical courses. The most notable aspect of the new approach was that the IA Probability course and the IB Statistics course were redesigned to appeal to a wide audience (200 first-year students in the case of IA Probability and nearly the same number of the second-year students in the case of IB Statistics). For a large number of students, these are the only courses from the whole of probability and statistics which they attend during their undergraduate years. Since more and more graduates in the modern world have to deal with theoretical and (especially) applied problems of a probabilistic or statistical nature, it is important that these courses generate and maintain a strong and wide appeal. The main goal shifted, moving from an academic introduction to the subject towards a more methodological approach which equips students with the tools needed to solve reasonable practical and theoretical questions in a 'real life' situation.

Consequently, the emphasis in IA Probability moved further away from sigma-algebras, Lebesgue and Stiltjies integration and characteristic functions to a direct analysis of various models, both discrete and continuous, with the aim of preparing students both for future problems and for future courses (in particular, Part IB Statistics and Part IB/II Markov chains). In turn, in IB Statistics the focus shifted towards the most popular practical applications of estimators, hypothesis testing and regression. The principal determination of examination performance in both IA Probability and IB Statistics became students' ability to choose and analyse the right model and accurately perform a reasonable amount of calculation rather than their ability to solve theoretical problems.

Certainly such changes (and parallel developments in other courses) were not always unanimously popular among the Cambridge University Faculty of Mathematics, and provoked considerable debate at times. However, the student community was in general very much in favour of the new approach, and the 'redesigned' courses gained increased popularity both in terms of attendance and in terms of attempts at examination questions (which has become increasingly important in the life of the Faculty of Mathematics). In addition, with the ever-growing prevalence of computers, students have shown a strong preference for an 'algorithmic' style of lectures and examination questions (at least in the authors' experience).

In this respect, the following experience by the first author may be of some interest. For some time I have questioned former St John's mathematics graduates, who now have careers in a wide variety of different areas, about what parts of the Cambridge University course they now consider as most important for their present work. It turned out that the strongest impact on the majority of respondents is not related to particular facts, theorems, or proofs (although jokes by lecturers are well remembered long afterwards). Rather they appreciate the ability to construct a mathematical model which represents a real-life situation, and to solve it analytically or (more often) numerically. It must therefore be acknowledged that the new approach was rather timely. As a consequence of all this, the level and style of Maths Tripos questions underwent changes. It is strongly suggested (although perhaps it was not always achieved) that the questions should have a clear structure where candidates are led from one part to another.

The second reason described above gives us hope that the book will be interesting for an audience outside Cambridge. In this regard, there is a natural question: what is the book's place in the (long) list of textbooks on probability and statistics. Many of the references in the bibliography are books published in English after 1991, containing the terms 'probability' or 'statistics' in their titles and available at the Cambridge University Main and Departmental Libraries (we are sure that our list is not complete and apologise for any omission).

As far as basic probability is concerned, we would like to compare this book with three popular series of texts and problem books, one by S. Ross [Ros1–Ros6], another by D. Stirzaker [St1–St4], and the third by G. Grimmett and D. Stirzaker [GriS1–GriS3]. The books by Ross and Stirzaker are commonly considered as a good introduction to the basics of the subject. In fact, the style and level of exposition followed by Ross has been adopted in many American universities. On the other hand, Grimmett and Stirzaker's approach is at a much higher level and might be described as 'professional'. The level of our book is intended to be somewhere in-between. In our view, it is closer to that of Ross or Stirzaker, but quite far away from them in several important aspects. It is our feeling that the level adopted by Ross or Stirzaker is not sufficient to get through Cambridge University Mathematical Tripos examinations with Class 2.1 or above. Grimmett and Stirzaker's books are of course more than enough – but in using them to prepare for an examination the main problem would be to select the right examples from among a thousand on offer.

On the other hand, the above monographs, as well as many of the books from the bibliography, may be considered as good complementary reading for those who want to take further steps in a particular direction. We mention here just a few of them: [Chu], [Dur1], [G], [Go], [JP], [Sc] and [ChaY]. In any case, the (nostalgic) time when everyone learning probability had to read assiduously through the (excellent) two-volume Feller monograph [Fe] had long passed (though in our view, Feller has not so far been surpassed).

In statistics, the picture is more complex. Even the definition of the subject of statistics is still somewhat controversial (see Section 3.1). The style of lecturing and examining the basic statistics course (and other statistics-related courses) at Cambridge University was always rather special. This style resisted a trend of making the exposition 'fully rigorous', despite the fact that the course is taught to mathematics students. A minority of students found it difficult to follow, but for most of them this was never an issue. On the other hand, the level of rigour in the course is quite high and requires substantial mathematical knowledge. Among modern books, the closest to the Cambridge University style is perhaps [CaB]. As an example of a very different approach, we can point to [Wil] (whose style we personally admire very much but would not consider as appropriate for first reading or for preparing for Cambridge examinations).

A particular feature of this book is that it contains repetitions: certain topics and questions appear more than once, often in slightly different form, which makes it difficult to refer to previous occurrences. This is of course a pattern of the examination process which becomes apparent when one considers it over a decade or so. Our personal attitudes here followed a proverb 'Repetition is the mother of learning', popular (in various forms) in several languages. However, we apologise to those readers who may find some (and possibly many) of these repetitions excessive.

This book is organised as follows. In the first two chapters we present the material of the IA Probability course (which consists of 24 one-hour lectures). In this part the Tripos questions are placed within or immediately following the corresponding parts of the expository text. In Chapters 3 and 4 we present the material from the 16-lecture IB Statistics course. Here, the Tripos questions tend to embrace a wider range of single topics, and we decided to keep them separate from the course material. However, the various pieces of theory are always presented with a view to the rôle they play in examination questions.

Displayed equations, problems and examples are numbered by chapter: for instance, in Chapter 2 equation numbers run from (2.1) to (2.102), and there are Problems 2.1–2.55.

Symbol □ marks the end of a solution of a given problem. Symbol ■ marks the end of an example.

A special word should be said about solutions in this book. In part, we use students' solutions or our own solutions (in a few cases solutions are reduced to short answers or hints). However, a number of the so-called examiners' model solutions have also been used; these were originally set by the corresponding examiners and often altered by relevant lecturers and co-examiners. (A curious observation by many examiners is that, regardless of how perfect their model solutions are, it is rare that any of the candidates follow them.) Here, we aimed to present all solutions in a unified style; we also tried to correct mistakes occurring in these solutions. We should pay the highest credit to all past and present members of the DPMMS who contributed to the painstaking process of supplying model solutions to Tripos problems in IA Probability and IB Statistics: in our view their efforts definitely deserve the deepest appreciation, and this book should be considered as a tribute to their individual and collective work.

On the other hand, our experience shows that, curiously, students very rarely follow the ideas of model solutions proposed by lecturers, supervisors and examiners, however impeccable and elegant these solutions may be. Furthermore, students understand each other much more quickly than they understand their mentors. For that reason we tried to preserve whenever possible the style of students' solutions throughout the whole book.

Informal digressions scattered across the text have been borrowed from [Do], [Go], [Ha], the St Andrew's University website www-history.mcs.st-andrews.ac.uk/history/ and the University of Massachusetts website www.umass.edu/wsp/statistics/tales/. Conversations with H. Daniels, D.G. Kendall and C.R. Rao also provided a few subjects. However, a number of stories are just part of folklore (most of them are accessible through the Internet); any mistakes are our own responsibility. Photographs and portraits of many of the characters mentioned in this book are available on the University of York website www.york.ac.uk/depts/maths/histstat/people/ and (with biographies) on http://members.aol.com/jayKplanr/images.htm.

The advent of the World Wide Web also had another visible impact: a proliferation of humour. We confess that much of the time we enjoyed browsing (quite numerous) websites advertising jokes and amusing quotations; consequently we decided to use some of them in this book. We apologise to the authors of these jokes for not quoting them (and sometimes changing the sense of sentences).

Throughout the process of working on this book we have felt both the support and the criticism (sometimes quite sharp) of numerous members of the Faculty of Mathematics and colleagues from outside Cambridge who read some or all of the text or learned about its existence. We would like to thank all these individuals and bodies, regardless of whether they supported or rejected this project. We thank personally Charles Goldie, Oliver Johnson, James Martin, Richard Samworth and Amanda Turner, for stimulating discussions and remarks. We are particularly grateful to Alan Hawkes for the limitless patience with which he went through the preliminary version of the manuscript. As stated above, we made wide use of lecture notes, example sheets and other related texts prepared by present and former members of the Statistical Laboratory, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, and Mathematics Department and Statistics Group, EBMS, University of Wales-Swansea. In particular, a large number of problems were collected by David Kendall and put to great use in Example Sheets by Frank Kelly. We benefitted from reading excellent lecture notes produced by Richard Weber and Susan Pitts. Damon Wischik kindly provided various tables of probability distributions. Statistical tables are courtesy of R. Weber.

Finally, special thanks go to Sarah Shea-Simonds and Maureen Storey for carefully reading through parts of the book and correcting a great number of stylistic errors.