Applied Mathematical Sciences | Volume 6

W. Freiberger
U. Grenander

A Course in Computational Probability and Statistics

With 35 Illustrations

Springer-Verlag New York · Heidelberg · Berlin
1971

Walter Freiberger
Ulf Grenander
Division of Applied Mathematics
Brown University
Providence, Rhode Island

Library of Congress Catalog Card Number 76-176272.

ISBN-13: 978-0-387-90029-2 e-ISBN-13: 978-1-4612-9837-3
DOI: 10.1007/978-1-4612-9837-3

PREFACE

This book arose out of a number of different contexts, and numerous persons have contributed to its conception and development.

It had its origin in a project initiated jointly with the IBM Cambridge Scientific Center, particularly with Dr. Rhett Tsao, then of that Center. We are grateful to Mr. Norman Rasmussen, Manager of the IBM Scientific Center Complex, for his initial support.

The work is being carried on at Brown University with generous support from the Office of Computing Activities of the National Science Foundation (grants GJ-174 and GJ-710); we are grateful to Dr. John Lehmann of this Office for his interest and encouragement. Professors Donald McClure and Richard Vitale of the Division of Applied Mathematics at Brown University contributed greatly to the project and taught courses in its spirit. We are indebted to them and to Dr. Tore Dalenius of the University of Stockholm for helpful criticisms of the manuscript.

The final stimulus to the book's completion came from an invitation to teach a course at the IBM European Systems Research Institute at Geneva. We are grateful to Dr. J.F. Blackburn, Director of the Institute, for his invitation, and to him and his wife Beverley for their hospitality.

We are greatly indebted to Mrs. Katrina Avery for her splendid secretarial and editorial work on the manuscript.

TABLE OF CONTENTS

INTRODUCTION

The purpose of this book is to present an attitude. It has been designed with the aim of making students and perhaps also faculty aware of some of the consequences of modern computing technology for probability theory and mathematical statistics. Not only the increased speed and memory of modern computers are relevant here; of at least equal importance to our subject are the versatile input-output devices and the existence of interactive time-sharing systems and of powerful programming languages. Of the last-mentioned, we have found APL most useful for our purposes.

The work described in these notes was initiated because we felt the time was ripe for a systematic exploitation of modern computing techniques in mathematical statistics and applied probability. Model building, for instance in applied probability, is very different today from what it was in pre-computer days, although this change has not yet fully penetrated to the textbook level. This course is being presented to remedy this situation to some degree; through it, we hope, students will become aware of how computers have increased their freedom of choice of mathematical models and liberated them from the restraints imposed by traditional mathematical techniques.

The project which led to this set of lecture notes is a continuation, although in different form, of an activity organized several years ago at the University of Stockholm. The activity was intended as a complement to the already existing program for training graduate students in mathematical statistics and operations research. It was felt that the students received a well-balanced education in mathematical theory but that something else was lacking: they were not trained for solving real-life problems in raw and unpolished form (in which they usually appear), far removed from that of pure and idealized textbook problems. In particular, little attention was given to the first, difficult stages of problem solving, namely the building of models, the collection of data, and the crucial relation between the mathematical, formal apparatus and the real-life phenomena under study.

To give students an opportunity for acquiring a more realistic orientation and critical attitude towards data, they were exposed to real problems chosen from many fields in industry, government and research. With the help of advice from a teacher or from older and more experienced students, they were asked to study a problem, formulate their own model, scrutinize data and present an analysis to the person or agency from whom the problem had originated. The results were later discussed in laboratory sessions, often before the analysis was completed, in order to get suggestions and ideas.

It became obvious in the course of this experiment that a major defect in the conventional training of graduate students was the lack of attention paid to the role of computers in the applications of mathematical statistics. Students have often to pass through a more or less painful stage in which they reorient themselves in order to learn what is computationally feasible as distinct from analytically possible. It is desirable that this transition be made easier, quicker and more complete. Although most students will have had some exposure to the computer, they may be inexperienced in its use in, say, applied probability. This will affect their ability to formulate realistic models, forcing them to choose analytically tractable models rather than those which best fit the problem.

The purpose of the present project is to equip students in probability and statistics better for work on real problems. The emphasis in the latter effort is on model building as influenced by computer science.

A growing number of practicing statisticians are today aware of the need to exploit more fully the capability of computers in statistics. This is particularly true in applied statistics, for instance in data analysis, where some research workers have emphasized this for several years. Less attention has been paid to the computational side of probability theory, although the need for a computational reorientation also exists in this subject. We therefore chose to concentrate our efforts on probability theory and applied probability as well as on statistics.

We divided our work into several chapters. Each chapter represents some concept or technique or relation for which a sufficiently rich mathematical structure has been developed and in which, at the same time, the impact of computer science can be expected to be substantial. The chapters together will cover only a part of mathematical statistics, although, we hope, an important one. We are particularly interested in areas in which the interaction between the analytical and computational approach is strong. This will usually only be the case where the analytical technique has been extended so far that further extension seems possible or worthwhile only through computer use, and makes it necessary that students possess a certain degree of mathematical sophistication. A course designed in such a way should be distinguished from one aiming only at teaching the computer implementation of standard statistical techniques and the writing of statistical programs. A course of the latter type is certainly useful and necessary, but the present project is more ambitious in scope and perhaps also more difficult to accomplish in the classroom. Little, in fact, seems to have been done in the direction described here. We had originally hoped to be able to use some already developed material, but there is disappointingly little available.

The prerequisites for the course are familiarity with the elements of probability and statistics, calculus and linear algebra. It will be assumed that students have some programming experience; most computing in the course will be based on APL which is, from the point of view of our computing tasks, by far the most suitable currently available interactive programming language.

The APL programs in the book should not be interpreted as forming a comprehensive, debugged program library (see section 7.1 in this context). They are presented only to illustrate our approach to computational probability and statistics and undoubtedly contain several mistakes.

Since the degree of mathematical sophistication is expected to vary a good deal among students, more advanced sections are starred, and it is suggested that they be read only by those who feel they have sufficient mathematical background. These sections should be discussed in class, without detailed proofs being given; instead, their interpretation and practical consequences should be discussed by the lecturer.

We strongly recommend that students be encouraged to complete the assignments to help them in the development of a real understanding of the material. The extent and size of the assignments will depend in part upon the computational facilities available during the course.

For further reading, and to find more advanced developments of some of the subjects covered here, we recommend the series of reports published under the NSFsponsored "Computational Probability and Statistics" project at Brown University, the titles of which are listed in the References.

The curves in the book were produced by a TSP plotting system on-line with a DATEL terminal, operating under APL/360.