Statistical Inference

Second Edition

George Casella Roger L. Berger

DUXBURY ADVANCED SERIES

**OIKONOMIKO ΠΑΝΕΠΙΣΤΗΜΙΟ**

**ΑΘΗΝΩΝ ΒΙΒΛΙΟΘΗΚΗ**

**E1*0.****722*6*5* Ap. 309.5

**Tag. *CAS***

**Statistical Inference**

Second Edition

**George Casella** *University of Florida*

**Roger L. Berger** *North Carolina State University*

**ONOM*I*A**

**DIKOM**

**BIBAIOONCH**

**UHI**

**AOHY**

***C*ANON**

**DUXBURY**

**THOMSON LEARNING**

Australia **• Canada • Mexico • Singapore • Spain • Un**ited Kingdom

United States

**илинди** H9146716

**DUXBURY**

**THOMSON LEARNING**

Sponsoring Edit A *Carolyn Crockett* **Marketing Representative: *Tom Ziolkowski* Editorial Assi*s*tant: *Jennifer Jenkins* Produotion Ed**itor: *Tom* N*ovack* **Assistan**t Editor: A*nn Day* Manuscript Editor: *Carol Reitz* **Permission**s Editor: *S*ue *Ewing*

**Cover Design: *Jennifer Mackres* I**nterior Illustratio**n: *Lori He*cke*lman* Print Buyer: Ven*a Dyer* Typesetting: *Integre Technical Publish****ing Co.* Cover Printing: *R. R. Donnelley & Sons Co.,*

***Crawfordsville*** Printing and Binding: *R. R. Donnelley*

*& Sons Co., Crawfo****rdsville***

All pro**ducts used herein are us**ed for identification p**urposes only and may be trademarks or registered trademarks of their respective owners.**

**COPYRI**GHT © 2002 the **Wadsworth Group. Duxbury is an imprint of the Wadsworth Group, a division of Thomson Learning Inc.**

**Thomson LearningTM is a trademark used herein under license.**

***For more inf*o*rmation about this or any other Duxbury products, contact:* DUXBURY 511 Forest Lodge Road** Pacific Grove, CA 93950 USA **www.duxbury.com** 1-800-423-0563 (Thom**son Learning Academi**c Resource Center)

All rights **reserved. N**o part of this work may be reproduced, transcribed **or used in any form or by any means-graphic, electronic, or mechanica**l, including photocopying, **recording, taping, Web distribution, or information storage and/or retrieval systems without the prior written permission** of the publisher.

***Fo*r *p*e*rmission to use ma*t**e*rial from t****h*is *an*d *c*ont*act us by* www.thomsonrights.com fax**: 1-800-730-2215 phone: 1-800-730-2214

**oints**

**P**

Printed in United States of A**meric**

10 9 8 7 6 5 4 3 2 1

**Library of Congress Cataloging-in-Publication Data**

**Casella, George.**

**Statistical inference */* George Casella, Ro*g*er L. Berger**.—2nd ed.

**p. cm. Includes bibliographical references and indexes. ISBN 0-534-24312-6**

**1. Mathematical statistics. 2. Probabilities. I. Berger, Roger L. II. Title. QA276*.*C37 2001 519.54dc21**

**2001025794**

**OK 18 P*A***

**WTED ON**

**PAPER**

*T*o A*nne and Vicki*

***Duxb*u*ry titles of related i*n*terest***

**Daniel, A*pplied Nonparametric Statisti****cs* 2nd Derr, S*tatistical Consulting:* A *Guide to Effective Co****mmunication*** Durrett, *Probability: Theory and Examples* and Graybill, *Theory a*n*d* A*pplication of the Linear Model*

Johnson, A*pplied Multivariate Methods for Data* A*nalysts* **Kuehl, *Design of Experiments: Statist****ical Principles of Research Design and Analysis* and Larsen, Marx, & Cooil, S*tatistics for Applied Problem Solving and Dec****is*i*on Making*** Lohr, S*ampling: Design and Analysis* Lunneborg*, Data* A*nalysis by Resampling: Co*n*cepts and Appl****ications*** Minh, Ap*plied Probability Models* Minitab Inc.*, MINITABTM Student Version 12 fo*r *Windows* **Myer*s, Classica****l and Modern Regr****ession with Applicati****on*s and Newton *&* Harvill, S*tatConcepts:* A V*isual Tour of Statistical Id****eas* Ramse**y & Schafer*, The Statistical Sleuth* 2nd SAS Institute Inc*., JMP-IN: Statistical Discovery Software* **Savag**e*, INSIGHT: B****usin*ess A*na****lysis Software for Microsoft® Excel* **Scheaff**er, Mendenhall, & Ott*, Elementary Survey Sampling* 5th Shapiro, *Modeling the Supply Chain* **Winston, S*imulation Model****ing Using @RISK.*.. To order copies contact your local bookstore or call 1-800-354-9706. For more **information contact Duxbury Press-8t 5**11 Forest Lodge Road, Pacific Grove, CA 93950, **or go to: www.duxbury.com**

**Preface to the Second Edition**

Although Sir Arthur Conan Doyle is responsible for most of the quotes in this book, perhaps the best description of the life of this book can be attributed to the Grateful Dead sentiment, “What a long, strange trip it's been."

Plans for the second edition started about six **years ago, a**nd for a long time we struggled with questions about what to add and what to delete. Thankfully, as time **passed, the answers becam**e clearer as the flow of the discipline of statistics became **clearer**. We see the trend moving away from elegant proofs of special cases to algo rithmic solutions of more complex and practical cases. This does not undermine the importance of mathematics and rigor; indeed, we have found that these have become more important. But the manner in which they are applied is changing.

For those familiar with the first edition, **we can su*mm*ar**i*z*e the changes succinctly as follows. Discussion of asymptotic methods has been greatly expanded into its own chapter. There is more emphasis on computing and simulation (see Section 5.5 and the computer algebra Appendix); coverage of the more applicable techniques has been expanded or added (for example, bootstrapping, the EM algorithm, p-values, logistic and robust regression); and there are many new Miscellan**ea and Exercises.** We have de-emphasized the more specialized theoretical topics, such **as equivariance** and decision theory, and have restructured so*m*e material in Chapters 3-11 for clarity.

There a*r*e two things that we want to note. First, with respect to computer algebra **programs, a**lthough we believe that they are becoming increasingly valuable tools, we did not want to force them on the instructor who does not share that belief. Thus, the treatment is “unobtrusive" in that it appears only in an appendix, with some hints throughout the book where it may be useful. Second, we have changed the numbering system to one that facilitates finding things. No**w theorems, lemmas, examp**les, and definitions are numbered together; for example, Definition 7.2.4 is followed by Example 7.2.5 and Theorem 10.1.3 precedes Example 10.1.4.

The first four chapters have received only minor changes. We reordered some ma terial (in particular, the inequalities and identities have been split), added some new **exam**ples and exercises, and did some general updating. Chapter 5 has also been re ordered, with the convergence section being moved further back, an**d a new section on** generating random variables added. The previous c**overa*g*e of invariance, which was** in Chapters 7*–9* of the first edition, has been greatly reduced and incorporated into Chapter 6, which otherwise has received only minor editing (mostly the addition of **new exercis**es). Chapter 7 has been expanded and updated, and includes a new section on the EM algorithm. Chapter 8 has also received minor editing and updating, and

**Yi X\*HOOLART**

**PREFACE TO THE SECOND EDITION**

**has a new seeriop;on p-values. In Chapter 9 we now put more emphasis o**n pivoting **(havingroalized th**at "guaranteeing an interval” was merely “pivoting the cdf”). Also, the material that was in Chapter 10 of the first edition (decision theory) has been re duced, and small sections on loss function optimality of point estimation, hypothesis testing, and interval estimation have been added to the appropriate chapters.

Chapter 10 is entirely new and attempts to lay out the fundamentals of large sample **infer**ence, including the delta method, consistency and asymptotic normality, boot strapping, robust estimators, score tests, etc. Chapter 11 is classic oneway ANO*V*A **and linear regression (which was covered** in two different chapters in the first edi tion). Unfortunately, co*v*erage of randomized block designs has been eliminated for **space reason**s. Chapter 12 covers regression with errors-in-variables and co**ntains new** material on robust and logistic **regression.**

After teaching from the first edition for a **number of years, we know (approximat**ely) what can be covered in a one-year course. From the second edition, it should be possible to cover the following in one year:

Chapter 1: Sections 1-*7* Chapter 6: Sections 1-3 Chapter 2: Sections 1-3 Chapter 7: Sections 1-3 Chapter 3: Sections 1-6 Chapter 8: Sections 1-3 Chapter 4: Sections 1-7 Chapter 9: Sections 1-3 Chapter 5: Sections 1-6 Chapter 10: Sections 1, 3, 4

Classes that begin the course with some probability background c**an cover more ma** terial from the later chapters.

Finally, it is almost impossible to thank all of the people who have contributed in **some way t**o making the second edition a reality (and help us correct the mistakes in the first edition). To all of our students, friends, and colleagues who took the time to send us a note or an e-mail, we thank you. A number of people made key suggestions that led to substantial changes i**n presentat**ion. Sometimes these suggestions were just short notes or comments, and some were longer reviews. Some were so long ago that their authors may have forgotten, but we haven't. So thanks to Arthur Cohen, Sir David Cox, Steve Samuels, Rob Strawderman and Tom Wehrly. We also owe much to **Jay Beder, who has sent us numerous comments and suggestions over the years and** possibly knows the first edition better than we do, and to Michael Perlman and his **class, who are sending comments and corrections even as we wr**ite this.

This book has seen a number of editors. We thank Alex Kugashev, who in the mid-1990s first suggested doing a second edition, and our editor, Carolyn Crockett, who constantly encouraged us. Perhaps the one person (other than us) who is most **respo**nsible for this book is our first editor, John Kimmel, who encouraged, published, **and marke**ted the first edition. Thanks, John.

*George Casella*

*Roger L. Berger*

**Preface to the First Edition**

When someone discovers that you are writing a textbook, one (or both) of two ques tions will be asked. The first is "Why are you writing a book?" and the second is "How is your book different from what's out there?” The first question is fairly e**asy to answer**. You are writing a book because you are not entirely satisfied with the available texts. The *s*econd question is harder to answer. The answer can't be put in a few sentences so, in order not to bore your audience (who may be asking the question only out of politeness), you try to say something quick and witty. It usually doesn't work.

The purpose of this book is to build theoretical statistics (as different from mathe matical statistics) from the first principles of probability theory. Logical development, **proofs, ideas, themes, et**c., evolve through statistical arguments. Thus, starting from the basics of probability, we develop the theory of statistical inference using tech niques, definitions, and concepts that are statistical **and are natural exte*nsio*ns and consequenc**es of previous concepts. When this end**eavor was started, we we*r*e not sure** how well it would work. The final judgment of our success is, of course, left to the **reader.**

The book is intended for first-year graduate students majoring in statistics or in a field wh**ere a statistics concentrat**ion is desirable. The prerequisite is one year of calculus. (Some familiarity with matrix manipulations would be useful, but is not **essent**ial.) The book can be used for a two-semester, or three-quarter, introductory course in statistics.

The fi*r*st four chapters cover basics of probability theory and introduce many fun damentals that are later necessary. Chapters 5 and 6 are the first statistical chapters. Chapter 5 is transitional (between probability and statistics) and can be the starting point for a course in statistical theory for students with some probability background. Chapter 6 is somewhat unique, detailing three statistical principles (sufficiency, like lihood, and invariance) and showing how these principles are important in modeling data. Not all instructors will cover this chapter in detail, although we strongly recom mend spending some time here. In particular, the likelihood and invariance principles are treated in detail. Along with the sufficiency principle, these principles, and the thinking behind them, are fundamental to total statistical understanding.

Chapters 7–9 represent the central core of statistical inference, **estimation (point** and interval) and hypothesis testing. *A* major feature of these chapters is the division into methods of *finding* appropriate statistical techniques and methods of *evaluating* these techniques. Finding and evaluating are of interest to both the theorist and the

**viii**

**PREFACE TO THE FIRST EDITION**

**pract**itioner, but we feel that it is important to **separate these endeavors. Different concerns a**re important, and different rules are invoked. Of further interest may be the sections of these chapters titled Other Considerations. Here, we indicate how the rules of statistical inference may be relaxed (as is done every day) and still produce meaningful inferences. Many of the techniques covered in these sectio**ns are ones that** are used in consulting and are helpful in analyzing and inferring from actual problems.

The final three chapters can be thought of as special topics, although we feel that **some fam**iliarity with the material is important in anyone's statistical education. Chapter 10 is a thorough introduction to decision theory and contains the most mod ern material we could include. Chapter 11 deals with the analysis of variance (**oneway** and randomized block), building the theory of the complete analysis from the more simple theory of treatment contrasts. Our experience has been that **experimenters are** most interested in inferences from contrasts, and using principles developed earlier, most tests and intervals can be derived from contrasts. Finally, Chapter 12 treats the theory of regression, dealing first with simple line**ar regression and then covering regression with "errors in variabl**es.” This latter topic is quite important, not only to show its own usefulness and inherent difficulties, but also to illustrate the limitations of inferences from ordinary regression.

As more concrete guidelines for basing a one-year course on this book, we offer the following suggestions. There can be two distinct types of courses taught from this book. One kind we might label “more mathematical," being a course appropriate for students majoring in statistics and having a solid mathematics background (at least 1 years of calculus, some matrix algebra, and perhaps a real analysis course). For such students we recommend covering Chapters 1-9 in their entirety (which should take approximately 22 weeks) and spend the remaining time customizing the course with selected topics from Chapters 10–12. Once the first nine chapters are covered, the material in each of the last three chapters is self-contained, and can be covered in any order.

Another type of course is "more practical." Such a course may also be a first course for mathematically sophisticated students, but is aimed at students with one year of calculus who may not be majoring in statistics. It stresses the more practical uses of statistical theory, being more concerned with understanding basic statistical concepts and deriving reasonable statistical procedures for a variety of situations, and less concerned with formal optimality investigations. Such a course will necessarily omit a certain amount of material, but the following list of sections can be covered in a one-year course:

Chapter Sections

All 2.1, 2.2, 2.3 .3.1, 3.2

4.1, 4.2, 4.3, 4.5 5.1, 5.2, 5.3.1, 5.4 6.1.1, 6.2.1 7.1, 7.2.1, *7.2.2,* 7.2.3, 7.3.1, 7.3.3, 7.4 8.1, 8.2.1, 8.2.3, 8.2.4, 8.3.1, 8.3.2, 8.4

oor A WNA

**PREFACE TO THE FIRST EDITION** 9.1, 9.2.1, 9.2.2, 9.2.4, 9.3.1, 9.4 11.1, 11.2

12.1, 12.2 If time permits, there can be some discussion (with little emphasis on details) of the material in Sections 4.4, 5.5, and 6.1.2, 6.1.3, 6.1.4. The material in Sections 11.3 and

12.3 may also be considered.

The exercises have been gathered from many sources and are quite plentiful. We feel that, perhaps, the only way to master this material is through practice, and thus we have included much opportunity to do so. The exercises are as varied as we could make them, and many of them illustrate points that are either new or compl**ementary** to the material in the text. Some exercises are even taken from research papers. (It makes you feel old when you can include exercises based on papers that were new research during your own student days!) Although the exercises are not subdivided like the chapters, their ordering roughly follows that of the chapter. (Subdivisions often give too many hints.) Furthermore, the exercises become (again, roughly) more challenging as their numbers become higher.

As this is an introductory book with a relatively broad scope, the topics are not covered in great depth. Ho**wever, we fe**lt some obligation to guide **the reader one** step further in the topics that may be of interest. Thus, we have included many **references**, pointing to the path to deeper understanding of any particular topic. (The *Encyclopedia of Statistical Sciences*, edited by Kotz, Johnson, and Read, provides a fine introduction to many topics.)

To write this book, we have drawn on both our past teachings and current work. We have also drawn on many people, to whom we are extremely grateful. We thank our colleagues at Cornell, North Carolina State, and Purdue—in particular, Jim Berger, Larry Brown, Sir David Cox, Ziding Feng, Janet Johnson, Leon Gleser, Costas Goutis, Dave Lansky, George McCabe, Chuck McCulloch, Myra Samuels, Steve Schwager, and Shayle Searle, who have given their time and expertise in reading parts of this **manuscr**ipt, offered assistance, and taken part in many conversations leading to con structive suggestions. We also thank Shanti Gupta for his hospitality, and the li brary at Purdue, which was essential. We are grateful for the detailed reading and helpful suggestions of Shayle Searle and of our reviewers, both anonymous and non **anon**ymous (Jim Albert, Dan Coster, and Tom Wehrly). We also thank David Moore and George McCabe for allowing us to use their tables, and Steve Hirdt for supplying us with data. Since this book was written by two people who, for most of the time, were at least 600 miles apart, we lastly thank Bitnet for making this entire thing possible.

*George Casella Roger L. Berger*

“*We have got to the deductions and the inferences," said Lestrade, winking at me. "I find it hard enough to tackle facts, Holmes, without flying away*

*after theories and fancies."* **Inspector Lestrade to Sherlock Holmes**

*The Boscombe Valley Mystery*

**Contents**

coer er

**1 Probability Theory**

1.1 Set Theory 1.2 Basics of Probability Theory

1.2.1 Axiomatic Foundations 1.2.2 The Calculus of Probabilities 1.2.3 Counting

1.2.4 Enu*m*e*r*ating *O*utco*m*es 1.3 Conditional Probability and Independence 1.4 Random Variables 1.5 Distribution Functions 1.6 Density and Mass Functions 1.7 Exercises 1.8 Miscellanea

&

8

sammen og

S

er

**2 Transformations and Expectations**

2.1 Distributions of Functions of a Random Variable 2.2 Expected Values 2.3 Moments and Moment Generating Functions 2.4 Differentiating Under an Integral Sign 2.5 Exercises 2.6 Miscellanea

Så

**3 Common Families of Distributions**

3.1 Introduction 3.2 Discrete Distributions 3.3 Continuous Distributions 3.4 Exponential Families 3.5 Location and Scale Families

111 **116**

xiv

**CONTENTS**

**3.6 Inequalities and I**dentities

3.6.1 Probability Inequalities

**3.6.2 Identities** 3.7 Exercises 3.8 Miscellanea

**121 122** 123

***1*27**

135

***139***

139 1*4*7

156

**162**

**169**

**4 Multiple Random Variables**

4.1 Joint and Marginal Distributions 4.2 Conditional Distributions and Independence 4.3 Bivariate Transformations **4.4 Hierar**chical Models and Mixture Distributions 4.5 Covariance and Correlation 4.6 Multivariate Distributions 4.7 Inequalities

*4.*7.1 Numerical Inequalities

4.7.2 Functional Inequalities 4.8 Exercises 4.9 Miscellanea

***1*77 186**

**186**

189 **192** 203

**207**

207

211

**218 218**

**222 226**

232 232

**5 Properties of a Random Sample**

5.1 Basic Concepts of Random Samples 5.2 Sums of Random Variables from a Random Sample 5.3 Sampling from the Normal Distribution

5.3.1 Properties of the Sample Mea**n and Variance**

5.3.2 The Derived Distributions: Student's *t* and Snedecor's *F* 5.4 Order Statistics 5.5 Convergence Concepts

5.5.1 Convergence in Probability 5.5.2 Almost Sure **Convergence** 5.5.3 Convergence in Distribution

5.5.4 The Delta Method 5.6 Generating a Random Sample

5.6.1 Direct Methods 5.6.2 Indirect Methods

5.6.3 The Accept/Reject Algorithm 5.7 Exercises 5.8 Miscellanea

234

235

240

245 247

251

253 255

2*67*

**271** 271

*272*

**6 Principles of Data Reduction**

6.1 Introduction 6.2 The Sufficiency Principle

6.2.1 Sufficient Statistics 6.2.2 Minimal Sufficient Statistics 6.2.3 Ancillary Statistics 6.2.4 Şufficient, Ancillary, and Complete Statistics

272 279 282 284

**CONTENTS**

**XV**

6.3 The Likelihood Principle

6.3.1 The Likelihood Function

6.3.2 The Formal Likelihood Principle 6.4 The Equivariance Principle **6.5 Exercises** 6.6 Miscellanea

290 290 292 296

300

307

**311**

311

**312** 312

315

***7* Point Estimation**

7.1 Introduction 7.2 Methods of Finding Estimators

7.2.1 Method of Moments 7.2.2 Maximum Likelihood Estimators 7.2.3 Bayes Estimators

7.2.4 The EM Algorithm 7.3 Methods of Evaluating Estimators

7.3.1 Mean Squared Error 7.3.2 Best Unbiased Estimators 7.3.3 Sufficiency and Unbiasedness

7.3.4 Loss Function Optimality 7.4 Exercises 7.5 Miscellanea

324 326 330

330

**334 342 348**

355

367

**373**

*3*73 *3*74 *3*74 *3*79

**380**

**8 Hypothesis Testing**

8.1 Introduction 8.2 Methods of Finding Tests

8.2.1 Likelihood Ratio Tests 8.2.2 **Bayesian Tests**

8.2.3 Union-Intersection and Intersection-Union Tests 8.3 Methods of Evaluating Tests

8.3.1 Error Probabilities and the Power Function 8.3.2 Most Powerful Tests 8.3.3 Sizes of Union-Intersection and Intersection-Union Tests 8.3.4 p-Values

8.3.5 Loss Function Optimality 8.4 Exercises 8.5 Miscellanea

**382 382**

**387 394**

**397** 400 402 413

**9 Interval Estimation**

9.1 Introduction 9.2 Methods of Finding Interval Estimators

9.2.1 Inverting a Test Statistic 9.2.2 Pivotal Quantities 9.2.3 Pivoting the CDF 9.2.4 Bayesian Intervals

**417 417** 420 *4*20 **427** 430

435

**хvі**

**CONTENTS**

.

**440**

440

9.3 Methods of Evaluating Interval Estimators

9.3.1 Size and Coverage Probability 9.3.2 Test-Related Optimality 9.3.3 Bayesian Optimality

9.3.4 Loss Function Optimality 9.4 Exercises 9.5 Miscellanea

**444** 4*47* **449 451 463**

**467 467** 467 *4*70 **473**

478

**10 Asymptotic Evaluations**

10.1 Point Estimation

10.1.1 Consistency 10.1.2 Efficiency 10.1.3 Calculations and Comparisons

10.1.4 Bootstrap Standard Errors 10.2 Robustness

10.2.1 The Mean and the Median

10.2.2 M-Estimators 10.3 Hypothesis Testing

10.3.1 Asymptotic Distribution of LRTs

10.3.2 Other Large-Sample Tests 10.4 Interval Estimation

10.4.1 Approximate Maximum Likelihood Intervals

10.4.2 Other Large-Sample Intervals 10.5 Exercises 10.6 Miscellanea

**488**

492

**496 496** 499 504 **515**

**521**

**521**

522 **524** 525

527

530 534

536

**11 Analysis of Variance and Regression**

11.1 Introduction **11.2 Onewa*y A*nalysis of Variance**

11.2.1 Model and Distribution Assumptions 11.2.2 The Classic ANO*V*A Hypothesis 11.2.3 Inferences Regarding Linear Combinations of Means 11.2.4 The ANOVA *F* Test 11.2.5 Simultaneous Estimation of Contrasts

11.2.6 Partitioning Sums of Squares 11.3 Simple Linear Regression

11.3.1 Least Squares: A Mathematical Solution 11.3.2 Best Linear Unbiased Estimators: A Statistical Solution 11.3.3 Models and Distribution Assumptions 11.3.4 Estimation and Testing with Normal Errors 11.3.5 Estimation and Prediction at a Specified x = xo

11.3.6 Simultaneous Estimation and Confidence Bands 11.4 Exercises 11.5 Miscellanea

539 542 **544 548** 550 557 559 563 572

**CONTENTS**

**xvii**

***577*** 577 577 579

**581**

**12 Regression Models**

12.1 Introduction **12.2 Regression with Errors in Variables**

12.2.1 Functional and Structural Relationships 12.2.2 A Least Squares Solution 12.2.3 Maximum Likelihood Estimation

12.2.4 Confidence Sets 12.3 Logistic Regression

12.3.1 The Model

12.3.2 Estimation 12.4 Robust Reg**ression 12.5 Exercises** 12.6 Miscell*a*nea

583 **588 59*1***

**591**

**593** 597 602 **608**

**Appendix: Computer Algebra**

**613**

**Table of Common Distributions**

**621**

**References**

**629**

**Author Index**

**645**

**Subject Index**

**649**

**List of Tables**

1.2.1

Number o**f arrangements**

**16**

4.1.1

Values of the joint pmf *f(x, y*)

141

7.3.1

354

Three estimators for a binomial p Counts of leukemia cases

360

8.3.1

Two types of errors in hypothesis testing

383

9.2.1 9.2.2

Location-scale pivots Sterne's acceptance region and confidence set Three 90% normal confidence intervals

*4*27 **431 441**

10.1.1 Bootstrap and Delta Method variances

Median*/*mean asymptotic relative efficiencies 10.2.1 Huber estimators

Huber estimator asymptotic relative efficiencies, k = 1.5

Poisson LRT statistic 10.3.1 Power of robust tests 10.4.1 Confidence coefficient for a pivotal interval 10.4.2 Confidence coefficients for intervals based on Huber's M-estimator

**480 484 485 487** 490 497

500

504

11.2.1 ANOVA table for oneway classification 11.3.1 Data pictured in Figure 11.3.1 11.3.2 ANOVA table for simple linear regression

538 542 556

12.3.1 Challenger data 12.4.1 Potoroo data

Regression M-estimator asymptotic relative efficiencies

**594** 598 601

**List of Figures**

1.2.1 1.2.2 1.5.1 1.5.2 1.6.1

Dart board for Example 1.2.7 Histogram of **averages** Cdf of Example 1.5.2 Geometric cdf, *p*=.3 Area under logistic curve

2.1.1 2.1.2 2.3.1 2.3.2 2.3.3

Transformation of Example 2.1.2 Increasing and nondecreasing cdfs Exponential densities Two pdfs with the same moments Poisson approximation to the binomial

106

**107** 108 109

3.3.1 3.3.2 3.3.3 3.3.4 3.3.5 3.3.6 3.5.1 3.5.2 3.5.3 3.5.4

Standard normal density Normal approximation to the binomial Beta densities Symmetric beta densities Standard normal density and Cauchy density Lognormal and gamma pdfs Location densities Exponential location densities Members of the same scale family Location-scale families

110

117 **118 119** 120

4.1.1 4.5.1 4.5.2 4.7.1 *4.*7.2

Regions for Example 4.1.12 Regions for Examples 4.5.4 and 4.5.8 Regions for Example 4.5.9 Convex function Jensen's Inequality

147 170 175

**189**

190

232 248

5.4.1 5.6.1 5.6.2 5.6.3

Region on which fr*,*v(r, v) >0 for Example 5.4.7 Histogram of exponential pdf Histogram of Poisson sample variances Beta distribution in accept*/*reject sampling

251

252

**XX**

**LIST OF FIGURES**

7.3.1 **7.3.2**

**Binomial MSE comparison Risk functions for *v*ariance estimators**

333 351

8.2.1 8.3.1 8.3.2 8.3.3 8.3.4

LRT statistic Power functions for Example 8.3.2 Power functions for Example 8.3.3 Power functions for three tests in Example 8.3.19 Risk function for test in Example 8.3.31

377 **384** 384 394 401

9.2.1 9.2.2 9.2.3 9.2.4 9.2.5 9.3.1

Confidence interval-acceptance region relationship Acceptance region and confidence interval for Example 9.2.3 Credible and confidence intervals from Example 9.2.16 Credible probabilities of the intervals from Example 9.2.16 Coverage probabilities of the intervals from Example 9.2.16 Three interval estimators from Example 9.2.16

421 423 **437 438** 439 **449**

10.1.1 Asymptotic relative efficiency for ga**mma mean estimators** 10.3.1 Poisson LRT histogram 10.4.1 LRT intervals for a binomial proportion 10.4.2 C*ov*erage probabilities for nominal .9 binomial confidence procedures

**478** 490 502 503

11.3.1 Vertical distances that are measured by RSS 11.3.2 Geometric description of the BLUE 11.3.3 Scheffé bands, t interval, and Bonferroni intervals

**542** 547 562

12.2.1 Distance minimized by orthogonal least squares 12.2.2 Three regression lines 12.2.3 Creasy-Williams *F* statistic 12.3.1 Challenger data logistic curve 12.4.1 Least squares, LAD, and M-estimate fits

581 **583** 590 595

599

**List of Examples**

1.1.3 1.2.2 1.2.3 1.2.5 1.2.7 1.2.10 1.2.12 1.2.13 1.2.15 1.2.18 1.2.19 1.2.20 1.3.1 1.3.3 1.3.4 1.3.6 1.3.8 1.3.10 1.3.11 1.3.13 1.4.2 1.4.3 1.4.4 1.5.2 1.5.4 1.5.5 1.5.6 1.5.9 1.6.2 1.6.4

Event operations Sigma algebra-I Sigma algebra-II Defining probabilities-I Defining probabilities-II Bonferroni's Inequality Lottery-I Tour**na*m*ent** Lottery-II Poker Sampling with replacement Calculating an average Four aces Continuation of Example 1.3.1 **Three prisoners** Coding Chevalier de Meré Tossing two dice Letters Three coin tosses-I **Random** variables Three coin tosses-II Distribution of a random variable Tossing three coins Tossing for a head Continuous cdf Cdf with jumps Identically distributed random variables Geometric probabilities Logistic probabilities

2.1.1 2.1.2 2.1.4 2.1.6

Binomial transformation Uniform transformation Uniform-exponential relationship-I Inverted gamma pdf

**xodii**

**LIST OF EXAMPLES**

2.1.7 S**quare transformation** 2.1.9 **Normal-chi squared relationship** *2*.2.2

Exponential mean 2.2.3 Binomial mean 2.2.4 **Cauchy mean** 2.2.6 Minimizing distance 2.2.7 Uniform-exponential relationship-II 2.3.3 Exponential v**ariance** 2.3.5 Binomi**al variance** 2.3.8 Gamma mgf 2.3.9 Binomial mgf 2.3.10 Nonunique moments 2.3.13 Poisson approximation 2.4.5 Interchanging integration and differentiation-I 2.4.6 Interchanging integration and differentiation-II 2.4.7 Interchanging summation and differentiation 2.4.9 Continuation of Example 2.4.7

3.2.1 3.2.3 3.2.4 3.2.5 3.2.6 3.2.7 3.3.1 3.3.2 3.4.1 3.4.3 3.4.4 3.4.6 3.4.8 3.4.9 3.5.3 3.6.2 3.6.3 **3.6.6** 3.6.9

Acceptance sampling Dice probabilities Waiting time Poisson approximation **Invers**e binomial sampling Failure times Gamma-Poisson relationship **Normal approximation** Binomial exponential family Binomial mean and variance Normal exponential family Continuation of Example 3.4.4 A curved exponential family Normal approximations Exponential location family Illustrating Chebychev A normal probability inequality Higher-order normal moments Higher-order Poisson moments

100 105 **111** 112 **113 114 115** 115 **118 122** 123 **125** 127

**140**

**141**

4.1.2 4.1.4 4.1.5 4.1.7 4.1.8 4.1.9 4.1.11

Sample space for dice Continuation of Example 4.1.2 Joint pmf for dice Marginal pmf for dice Dice probabilities Same marginals, different joint pmf Calculating joint probabilities-I

**142** 143 **144 144 145**

**LIST OF EX*A*MPLES**

**xxiii**

**146 148** 150 152 153 154

155

4.1.12 Calculating joint probabilities-II 4.2.2 Calculating conditional probabilities 4.2.4 Calculating conditional pdfs 4.2.6 Checking indepen**dence-I** 4.2.8 Checking independence-II

Joint probability model 4.2.11 Expectations of independent variables 4.2.13 Mgf of a sum of normal variables *4*.3.1 Distribution of **the sum *o*f Poisso*n v*ariables** 4.3.3 Distribution of the product of beta variables 4.3.4 Sum and difference of normal variables 4.3.6 Distribution of the ratio of normal variables 4.4.1 Binomial-Poisson hierarchy *4.*4.2 Continuation of Example 4.4.1 4.4.5

**Gene**ralization of Example 4.4.1 **4.4.6** Beta-binomial hierarchy 4.4.8 Continuation of Example 4.4.6 4.5.4 Correlation-I 4.5.8 Correlation-II 4.5.9 Correlation-III 4.6.1 Multivariate pdfs 4.6.3 Multivariate pmf **4.6.8** Mgf of a sum of gamma variables 4.6.13 Multivariate change of variables

Covariance inequality 4*.*7.8 An inequality for **means**

156 157 **158 159 162** 163 163 **165 167 168** 170 173

174

**178**

***1*8*1***

**183 185 188**

191

208

2*10*

215 **216** 217 22*4* 225 230

23*1*

5.1.2 5.1.3 5.2.8 5.2.10 5.2.12 5.3.5 5.3.7 5.4.5 5.4.7 5.5.3 5.5.5 5.5.7 5.5.8 5.5.11 5.5.16 5.5.18 5.5.19 5.5.22

Sample pdf-exponential Finite population model Distribution of the **mean** Sum of Cauch**y random variables** Sum of Bernoulli random variables Variance ratio distribution Continuation of Example 5.3.5 Uniform order statistic pdf Distribution of the midrange and range Consistency of 52 Consistency of S Almost sure convergence Convergence in probability, not almost surely Maximum of uniforms Normal approximation to the negative binomial Normal approximation with estimated variance Estimating the odds Continuation of Example 5.5.19

233 233 **234**

234

235 239 240 **240 242**

**LIST OF EXAMPLES**

242 **243**

*2*44

5.5.23 Approx**imate mean and variance** 5.5.25 Continuation o**f Exampl**e 5.5.23 5.5.27 Moments of a ratio estimator 5.6.1 Exponential lifetime 5.6.2 Continuation of Example 5.6.1 5.6.3 Probability Integral **Transform** .5.6.4 Box-Muller algorithm

5.6.5 Binomial random variable generation 5.6.6 Distribution of the Poisson variance 5.6.7 **Beta random variable generation—I** 5.6.9 Beta random variable generation-II

246 246 **247 249** 249 250 251

254

6.2.3 Binomial sufficient statistic 6.2.4 Normal sufficient statistic 6.2.5 Sufficient order statistics 6.2.7 Continuation of Example 6.2.4 6.2.8 Uniform sufficient statistic 6.2.9 Normal sufficient statistic, both parameters unknown 6.2.12 Two normal sufficient statistics 6.2.14 Normal minimal sufficient statistic 6.2.15 Uniform minimal sufficient statistic 6.2.17 Uniform ancillary statistic 6.2.18 Location family ancillary statistic 6.2.19 Scale family ancillary statistic 6.2.20 Ancillary precision 6.2.22 Binomial complete sufficient statistic 6.2.23 Uniform complete sufficient statistic 6.2.26 Using Basu's Theorem-I 6.2.27 Using Basu's Theorem-II 6.3.2 Negative binomial likelihood 6.3.3 Normal fiducial distribution 6.3.4 Evidence function 6.3.5 Binomial/negative binomial experiment 6.3.7 Continuation of Example 6.3.5 6.4.1 Binomial equivariance 6.4.3 Continuation of Example 6.4.1 6.4.5 Conclusion of Example 6.4.1 6.4.6 Normal location invariance

27*4* 274 *2*75 *27*7 277 **279 280** 281 282 **282** 283 284 285 285 **286 288 289** 290 291 292 293 295 297 298

299

299

313 **313**

7.2.1 7.2.2 7.2.3 7.2.5 7.2.6 7*.2.*7

Normal method of moments Binomial method of moments Satterthwaite approximation Normal likelihood Continuation of Example 7.2.5 Bernoulli MLE

***314***

**316**

**317**

**317**

**LIST OF EXAMPLES**

**Xorv**

**318** 318 321 **322**

**323**

324 326 326 327 **328**

331

331

7.2.**8 Restricted range MLE** 7.2.9 Binomial MLE, unknown number of trials 7.2.11 Normal MLEs, *u* and o unknown 7.2.12 Continuation of Example 7.2.11 7.2.13 Continuation of Example 7.2.2 7.2.14 Binomial Bayes estimation 7.2.16 Normal Bayes estimators *7.2.1*7 Multiple Poisson rates *7.*2.18 Continuation of Example 7.2*.1*7 7.2.19 Conclusion of Example 7.2.17 7.3.3 Normal MSE 7.3.4 Continuation of Example 7.3.3 7.3.5 MSE of binomial Bayes estimator 7.3.6 MSE of equivariant estimators 7.3.8 Poisson unbiased estimation 7.3.12 Conclusion of Example 7.3.8 7.3.13 Unbiased estimator for the scale uniform 7.3.14 Normal variance bound 7.3.16 Continuation of Example 7.3.14 7.3.18 Conditioning on an insufficient statistic 7.3.21 Unbiased estimators of zero 7.3.22 C*o*ntinuat*i*on of Example 7.3.13 7.3.24 Binomial best unbiased estimation 7.3.25 Binomial risk functions 7.3.26 Risk of normal variance 7.3*.*27 Variance estimation using Stein's loss 7.3.28 Two Bayes rules 7.3.29 Normal Bayes estimates 7.3.30 Binomial Bayes estimates

**332 333** 335 338 339 340 **341** 343 **345** 346 347 350 350

**35*1***

353 353 354

*3*75 *3*76 378 *3*78

*3*7*9*

8.2.2 8.2.3 8.2.5 8.2.6 8.2.7 8.2.8 8.2.9 8.3.2 8.3.3 8.3.4 8.3.7 8.3.8 8.3.10 8.3.14 8.3.15

Normal LRT Exponential LRT LRT and sufficiency Normal LRT with unk**nown variance Normal Bayesian test** Normal union-intersection test Acceptance sampling Binomial power function Normal power function Continuation of Example 8.3.3 Size of LRT Size of union-intersection test Conclusion of Example 8.3.3 UMP binomial test UMP normal test

380 **382** 383 **384** 385 **386 387 387**

390

390

**Xorvi**

**LIST OF EXAMPLES**

**8.3.18 Continuation of Examp**le 8.3.15 8.3.19 N**onexisten**ce of UMP test 8.3.20 Unbiased test 8.3.22 An equivalence 8.3.25 Intersection-union test 8.3.28 Two-sided normal p-value 8.3.29 One-sided normal p-value 8.3.30 Fisher's Exact Test 8.3.31 Risk of UMP test

**392** 392 **393** 395 396 398 **398** 399 401

**418 418** 419 420

**423**

425

425

427 **428**

429

429 **431**

9.1.2 9.1.3 9.1.6 9.2.1 9.2.3 9.2*.*4 9.2.5 9.2.7 9.2.8 9.2.9 9.2.10 9.2.11 9.2.13 9.2.15 9.2.16 9.2.17 9.2.18 9.3.1 9.3.3 9.3.4 9.3.6 9.3.8 9.3.11 9.3.12 9.3.13

Interval estimator Continuation of Example 9.1.2 Scale uniform interval estimator Inverting a normal test Inverting an LRT **Normal one-sid**ed confidence bound Binomial one-sided confidence bound Location-scale pivots **Gamma pivot** Continuation of Example 9.2.8 Normal pivotal interval Shortest length binomial set Location exponential interval Poisson interval estimator Poisson credible set Poisson credible and coverage probabilities Coverage of a normal credible set Optimizing length Optimizing expected length Shortest pivotal interval UMA confidence bound Continuation of Example 9.3.6 Poisson HPD region Normal HPD region Normal interval estimator

**433**

**434** 436

*4*37

**438** 441 **443 443**

4*4*5

**446 448 449** 450

10.1.2 Consistency of X 10.1.4 Continuation of Example 10.1.2 10.1.8 Limiting variances 10.1.10 Large-sample mixture v**ariances** 10.1.13 Asymptotic normality and consistency 10.1.14 Approximate binomial variance 10.1.15 Continuation of Example 10.1.14 10.1.17 AREs of Poisson estimators 10.1.18 Estimating **a gamma mean**

**468 469 470** *4*71 472 474 475 *4*76

*4*77

**LIST OF EXAMPLES**

**Xorvii**

**478 479** *4*79 480

**482**

**483 484 485** 486

*487*

10.1.19 Boo**tstrapping a variance 10.1.20 Bootstrapping a binomial variance** 10.1.21 Conclusion of Example 10.1.20 10.1.22 Parametric boot**strap 10.2.1 Robustness of the sample mean 10.2.3 Asymptotic normality of the median** 10.2.4 AREs of the **median to the mean 10.2.5 Huber estimator** 10.2.6 Limit distribution of the Huber es**timator** *10.*2.7 **ARE of the Huber esti*m*ator** 10.3.2 Poisson LRT **10.3.4 Multinomial LRT** 10.3.5 Large-sample binomial tests 10.3.6 Binomial score test 10.3.7 Tests based on the Huber estimator 10.4.1 Continuation of Example 10.1.14 **10.4.2 Binomial score interval** 10.4.3 Binomial LRT in**terval 10.4.4 Approximate interval 10.4.5 Approximate Poisson interval** 10.4.6 More on the binomial scor**e interval 10.4.7 Compariso**n of binomial intervals **10.4.8 Intervals based on the Huber estimator 10.4.9 Negative binomial interval 10**.6.2 Influence functions of the m**ean and median**

**489 491 493** 495 **496** *4*97 **498 499 499**

*500*

501 502 503 504

**518**

11.2.**1 Oneway ANOVA** 11.2.3 The ANOVA hypothesis **11.2.6 ANOVA contrasts 11.2.9 Pairwise differences** 11.2.12 Continuation of Example 11.2.1 11.3.1 Predicting **grape crops** 11.3.4 C**ontinuation of Exam**ple 11.3.1

**522** 525 **529** 534 538 540 555

12.2.1 12.3.1 12.3.2 12.4.1 12.4.2 12.4.3 **12.4.4** 12.4.5

Estim**ating atmospheric pressure** Challenger **data Challenger data continued Robustness of least squares estimates Catastrophic observations** Asymptotic normality of the LAD estimator **Regression M-estimator** S**imulation of regression ARES**

579 594 **596** 597 598 599 601

601

A.0.1 A.0.2 A.0.3

Unordered sampling Uni**variate transformation Bivariate transformations**

**613 614 614**

**Xorviji**

**LIST OF EXAMPLES**

A.0.4 A.0.5 A.0.6 A.0.7 A.0.8

Normal probability **Density of a sum** Fourth moment of sum of uniforms **ARE for a gamma mean** Limit of chi squared mgfs

**616 616 617**

***618***

**619**

**ww*V*O*wKM*EL**

**WOHN**

Chapter 1

**QANON**

**Probability Theory**

**II I**

"You *can, for erample, never foretell what any one man will do, but you can say with precision what an average number will be up to. Individuals vary, but percentages remain constant. So says the statistician."*

**Sherlock Holmes**

*The Sign of Four*

The subject of probability theory is the foundation upon which all of statistics is built, providing a means for modeling populations, experiments, or almost anything else that could be considered a random phenomenon. Through these models, statisti cians are able to draw inferences about populations, inferences based on **examination** of only a part of the whole.

The theory of probability has a long and rich history, dating back at least to the **sevent**eenth century when, at the request of their friend, the Chevalier de Meré, Pascal and Fermat developed a mathematical formulation of gambling odds.

The aim of this chapter is not to give a thorough introduction to probability theory; such an attempt would be foolhardy in so short a space. Rather, we attempt to outline some of the basic ideas of probability theory that are fundamental to the study of statistics.

Just as statistics builds upon the foundation of probability theory, probability the ory in turn builds upon set theory, which is where we begin.

**1.1 Set Theory** One of the main objectives of a statistician is to draw conclusions about a population of objects by conducting an experiment. The first step in this endeavor is to identify the possible outcomes or, in statistical terminology, the sample space.

**Definition 1.1**.1 The set, S, of all possible outcomes of a particular experiment is called the *sample spac*e for the experiment.

If the experiment consists of tossing a coin, the sample space contains two outcomes, heads and tails; thus,

*S*={H, T}. If, on the other hand, the experiment consists of observing the reported SAT scores of randomly selected students at a certain university, the sample space would be

***t*ibaL HYMOQINATE**

*IT*

**PROBABILITY THEORY**

**Section 1.1**

**thler.det** of positive id*tede*rs between 200 and 800 that are multiples of ten-that is, $ ${200, 210, 220,,.,780,790, 800). Finally, consider an experiment where the **obséivation je reartigni ti**me to a certain stimulus. Here, the sample space would **consist of all positive numbe**rs, that is, S = (0,0).

We can classify sample spaces into two types according to the number of elements they contain. Sample spaces can be either countable or uncountable; if the elements of **a sample space ca**n be put into 1-1 correspondence with a subset of the integers, the sample space is countable. Of course, if the sample space contains only a finite number of elements, it is countable. Thus, the coin-toss and SAT score sample spaces are both countable (in fact, finite), whereas the reaction time sample space is uncountable, since the positive real numbers cannot be put into 1-1 correspondence with the integers. If, how**ever, we measured re**action time to the **nearest s**econd, then the sample space would be in seconds) S = {0,1,2,3,...), which is then countable.

This distinction between countable and uncountable sample spaces is important only in that it dictates the way in which probabilities can be assigned. For the most part, this causes no problems, although the mathematical treatment of the situations is different. On a philosophical level, it might be argued that there can only be count able sample spaces, s**ince measurements ca**nnot be made with infinite accuracy. (A **sam**ple space consisting of, say, all ten-digit numbers is a countable sample space.) While in practice this is true, probabilistic and statistical methods associated with uncountable sample spaces are, in general, less cumbersome than those for countable sample spaces, and provide a close approximation to the true (countable) situation.

Once the sample space has been defined, we are in a position to consider collections of possible outcomes of an experiment.

**Definition 1.1.2** An *event* is any collection of possible outcomes of an experiment, that is, any subset of S (including S itself).

Let A be an event, a subset of S. We say the event A occurs if the outcome of the **experim**ent is in the set A. When speaking of probabilities, we generally speak of the probability of an event, rather than a set. But we may use the terms interchangeably.

We first need to define formally the following two relationships, which allow us to order and equate sets:

А*с Ве*те Aх*є В;*

(containment) A*= B* AC*B* and *B*CA.

(equality) Given any two events (or sets) A and *B,* we have the following elementary set operations: **Union**: The union of A and *B*, written AU*B*, is the set of elements that belong to

either A or *B* or both:

*A*U*B =* {x:1 € *A* or IE *B}*.

**Intersection**: The intersection of *A* and *B*, written An*B*, is the set of elements that

belong to both A and *B:*

An*B =* {1:1 € A and X*E B}.*

**Section 1.1**

**SET THEORY**

***C*omplementation: The complement of** A, written *A*C, is the set of all elements

that are not in A:

Ao = {x: *x¢* A}.

**Example 1.1.3 (Event operations)** Consider the experiment of selecting a card at random from a standard deck and noting its suit: clubs (C), diamonds (D), hearts (H), or spades (S). The sample space is

*S*={C, D, H, S),

and some possible events are

A = {C, D} and *B* = {D, H, SH.

From these events we **can form**

AU*B=* {C, D, H,S}, A*nB* = {D), and A = {H,S).

Furthermore, notice that AU*B=*S (the event S) and (AUB) = 0, where 0 denotes the *empty set* (the set consisting of no elements).

The elementary set operations can be combined, somewhat akin to the way addition and multiplication can be combined. As long **as we are c**areful, we can treat sets as if they were numbers. We can now state the following useful properties of set operations.

**Theorem 1.1**.4 *For any three ev*en*ts, A, B, and C, defined on a sample spac*e *S,*

a. Commutativity AU*B= B*U*A*,

AN*B=B*nA; **b. Ass**ociativity AU*(B*U*C*) = (AU*B*)U*C,*

An *(BO*C)=(An *B*) n*C*; c. Distributive Laws An*(B*U*C*) =(AN*B*)U(ANC),

AU*(BNC*) = (AU*B*) *n*(AU*C*); d. DeMorgan's Laws (AU*B*) = Ao*n Bo,*

(A*NB*) = A' U *BC.* **Proof**: The proof of much of this theorem is left as Exercise 1.3. Also, Exercises 1.9 and 1.10 generalize the theorem. To illustrate the technique, however, we will prove the Distributive Law:

AN(*B*UC) = (An B)U(AN*C*). (You might be familiar with the use of Venn diagrams to “prove” theorems in set theory. We caution that although Venn di**agrams are sometim**es helpful in visualizing a situation, they do not constitute a formal proof.) To prove that two sets are equal, it must be demonstrated that each set contains the other. Formally, then

An*(B*U*C*) = {X E*S:* XE A and 2 E *(B*U*C*)}; (AN*B*)U(ANC) = {Z ES:x E(AN*B*) or IE(ANC)}.

**PROBABILITY THEORY**

**Section 1.1**

**We first show tha**t A*n(B*U*C*) C*(*ANB)U(*A*N*C*). Let 2 € *(*An*(B*U*C*)). By the **def**inition of intersection, it must be that I*E (BUC*), that is, either \* E *B o*r IE*C*. Since x also must be in *A*, we have that either a E (*A*NB) or x E (AN*C*); therefore,

I € *(*(An *B*)U(AN*C*)),

**and the containment is established.**

**Now assume** x *= (*(AN*B*)U(ANC)). This implies that 2 (AN*B*) or IE *(*ANC). If x E (AN*B)*, then x is in both A and *B*. Since x E *B, 2 E (B*UC) and thus IE(AN*(B*U*C*)). If, on the other hand, 2 € (*A*n*c*), the argument is similar, and we again conclude that x € (An(*B*U*C*)). Thus, we have established (An *B*u(AN*C*) *A*n(*B*U*C*), showing containment in the other direction and, hence, proving the Distributive Law.

The operations of union and intersection c**an be extende**d to infinite collections of sets as well. If A1, A2, A3, ... is a collection of sets, all defined on a sample space *S,* **then**

JA;= {Z ES:3€ Aį for some i},

**i=1**

**8 CT**

Ai= {Z ES:2€ Ai for all i}.

For example, let S = (0,1] and define Ai = [(1*/i*), 1]. Then

(1/1), 1] =

{x € (0,1] : < € [(1*/1*), 1] for some i}

**i=1**

**=1**

= {x € (0,1}} = (0,1); 0 4: = N(1/),1] = {2 € (0,11 : 2 € [(1*/*), 1] for all i} izi. i=1

= {z € (0,1] : 1 € [1,1]} = {1}.

(the point 1)

It is also possible to define unions and intersections over uncountable collections of sets. If I is an index set (a set of elements to be used as indices), then

UAg = {2 ES:2 € Ag for some a},

**aer**

Ag = {Z ES:I E Ag for all a}.

aer If, for example, we take r = {all positive real numbers) and Ag = (0,a], then Vaer Ag = (0,00) is an uncountable union. While uncountable unions and intersec tions do not play a major role in statistics, they sometimes provide a useful mec**hanism** for obtaining an answer (see Section 8.2.3).

Finally, we discuss the idea of a partition of the sample space.

**BASICS OF PROBABILITY THEORY**

**Section 1.2**

**on**

**Definition 1.1.5 Two events A and *B* ar*e dis****joint (o*r *mutually exclusiv*e) if An*B =* 0. The events A1, A2, ... are *pairwise disjoint* (or *mutually exclusive*) if Ain Aj = 0 for all i + *j.*

Disjoint sets are sets with no points in common. If we draw a Venn diagram for two disjoint sets, the sets do not overlap. The collection

Ai = si, i + 1),

i = 0,1,2, ...,

consists of pairwise disjoint sets. Note further that von Ai = [0,00).

**Definition 1.1.6** If A1, A2,... are pairwise disjoint and Up Ai = S, then the collection A1, A2, ... forms a *partition* of S.

The sets Aį = [i, i + 1) form a partition of (0,0). In general, parti**tions are very** useful, allowing us to divide the sample space into small, nonoverlapping pieces.

**1.2 Basics of Probability Theory**

When an experiment is performed, the realization of the experiment is an outcome in the sample space. If the experiment is performed a number of times, different outcomes **may** occur each time or some outcomes may repeat. This “frequency of occurrence of **an o**utcome can be thought of as a probability. More probable outcomes occur **more fre**quently. If the outcomes of an experiment can be described probabilistically, we **are** on our way to analyzing the experiment statistically.

In this section we describe some of the basics of probability theory. We do not define probabilities in terms of frequencies but instead take the mathematically simpler **ax*io*matic app*r*oach**. As will b*e s*een, the axiomatic approach is not concerned with the interpretations of probabilities, but is concerned only that the probabilities are defined by a function satisfying the axioms. Interpretations of the probabilities are quite another matter. The "frequency of occurrence of an event is one example of a **pa**rticular interpretation of probability. Another possible interpretation is a subjective one, where rather than thinking of probability as frequency, we can think of it as a belief in the chance of an event occurring.

*1.2.1* A*riomatic Fou****ndatio*n*s*** For each event A in the sample space S we want to associate with *A* a number between zero and one that will be called the probability of A, denoted by *P*(A). It would seem natural to define the domain of *P* (the set where the arguments of the function *P(*) are defined) as all subsets of S; that is, for each *A*C*S* we define *P(A*) as the probability that A occurs. Unfortunately, matters are not that simple. There are some technical difficulties to overcome. We will not dwell on these technicalities; although they are of importance, they are usually of more interest to probabilists *t*han to statisticians. However, a firm understanding of statistics re**quires at least a** passing familiarity with the following.

**PROBABILITY THEORY**

**Section 1.2**

**Definition 1.2.1 *A* collection of subsets of S is called a *sigma algebra (*or *Borel*** *field)*, denoted by B, if it satisfies the following three properties: a. D E *B* (the empty set is an element of B). b. If AEB, then A E B (B is closed under complementation). C. If A1, A2,..*. EB*, then US*A*į E*B (B* is closed under countable unions).

The empty set 0 is a subset of any set. Thus, OC S. Property (a) states that this subset is always in a sigma algebra. Since S = 0°, properties (a) and (b) imply that *S* is always in *B* also. In addition, from DeMorgan's Laws it follows that B is closed under countable intersections. If A1, A2, ..*. EB,* then As, AS,... E *B* by property (b), and therefore V A E *B.* However, using De Morgan's Law (as in Exercise 1.9), we **have**

(1.2.1)

(Ů4) =

Thus, again by property (b), n, A; *E B.*

Associated with sample space S we can have many different sigma algebras. For **exam**ple, the collection of the two sets {0, S} is a sigma algebra, usually called the trivial sigma algebra. The only sigma algebra we will be concerned with is the smallest one that contains all of the open sets in a given sample space S.

**Example 1.2.2 (Sigma algebra-I)** If S is finite or countable, then these techni calities really do not arise, for we define for a given sample space *S,*

*B* = {all subsets of S, including S itself).

If S has n elements, there are 2" sets in *B (*see Exercise 1.14). For example, if S = {1,2,3), then B is the following collection of 23 = 8 sets:

{1} {1,2} {1,2,3} {2} {1,3} {3} {2,3}

In general, if S is uncountable, it is not an easy task to describe B. However, *B* is **chosen t**o contain any set of interest.

**Example 1.2.3 (Sigma algebra-II**) chosen to contain all sets of the form

Let S = *(*-0,00), the real line. Then B is

[a,b],

(a,b],

*(a,b),*

and

(*a, b*)

for all real numbers a and b. Also, from the properties of *B*, it follows that B con tains all sets that can be formed by taking (possibly countably infinite) unions and intersections of sets of the above varieties.

**Section 1.2**

**BASICS OF PROBABILITY THEORY**

**We are now in a position t**o define a probability function.

**Definition 1.2.4 Given a sample space S and an associate**d sigma algebra *B*, a *probability function* is a function *P* with domain *B* that satisfies 1*. P*(A) > 0 for all AE*B.* 2. *P*(S) = 1. 3. If A1, A2,... E*B* are pairwise disjoint, then *P*(

V A ) = Li*-P*(Ai).

The three properties given in Definition 1.2.4 are usually referred to as the A**xioms** of Probability (or the Kolmogorov Axioms, after A. Kolm*og*oro*v, o*ne of the fathers of probability theory). Any function P that satisfies the Axioms of Probability is called a probability function. The axiomatic definition makes no attempt to tell what partic ular function *P* to choose; it merely requires *P* to satisfy the axioms. For **any sample** space many different probability functions can be defined. Which one(s) reflec**ts what** is likely to be observed in a particular experiment is still to be discussed.

**Example 1.2.5 (Defining probabilities-I)** Consider the simple experiment of tossing a fair coin, so S = {H, T}. By a "fair" coin we mean a balanced coin that is equally as likely to land heads up as tails up, and hence the reasonable probability function is the one that assign*s* equal probabiliti**es to heads and tails, that** is,

(1.2.2)

*P*({H}) = *P*({T}).

Note that (1.2.2) does not follow from the Axioms of Probability but rather is out side of the axioms. We have used a symmetry interpretation of probability (or just intuition) to impose the requirement that heads and tails be equally probable. Since S = {H} U*{*T}, we have, from Axiom *1, P{*{H} U*{*T}) = 1. Also, {H} and {T} are disjoint, so *P*({H} U{T}) = *P*({H}) + *P*({T}) and

(1.2.3)

*P*({H}) + *P*({T}) = 1.

Simultaneously solving (1.2.2) and (1.2.3) shows that *P*({H}) = *P*({T})= 1.

Since (1.2.2) is based on our knowledge of the particular experiment, not the axioms, **any nonnegative val**ues for *P({*H}) an*d P({*T}) that satisfy (1.2.3) define a legitimate probability function. For example, we might choose *P*({H}) = and *P*{{T})= . ||

We need general methods of defining probability functions that we know will always satisfy Kolmogorov's Axioms. We do not want to have to check the Axioms for each new probability function, like we did in Example 1.2.5. The following giv**es a common** method of defining a legitimate probability function.

**Theorem 1**.2.6 *Let* S = {81, ...,*S*n} *be a finite set. Let B be any sigma algebra of subsets of S. Let P*1, ..*., Pn be nonnegative numbers that sum to 1. For a*n*y A e B, define P(A) by*

*P*(A) =

pi.

**{*i:*9;E*A*}**

**PROBABILITY THEORY**

**Section 1.2**

*(The* ***sum over an empty set is defi****ned to be 0.) Then P is a probability function o*n ***B. This remains t****r*u*e if S* = {81, 52,...} ***is a countable set.* Proof:** We will give the proof for finite S. For any A E*B, P(*A) = fiib*ita*? *Pi* 20, **because ever**y *pi* > 0. Thus, Axiom 1 is true. Now,

P(S) – I r=1

**{1:9;ES)**

**i=1**

Thus, Axiom 2 is true. Let A1,..., A*k* denote pairwise disjoint events. *(B c*ontains only a finite number of sets, so we need consider only finite disjoint unions.) Then,

(U4) -

*Pj =* ;}

PÉLE, > - ŠPA).

*P(*A).

*Pj =*) i=1 {*3:9*;€ Ai} i=1

**\i=*1***

{*j:9;* EUA

The first and third equalities are true by the definition of *P(*A). The disjointedness of the Ais ensures that the second equality is true, because the same *pj*s appear exactly once on each side of the equality. Thus, Axiom 3 is true and Kolmogorov's Axioms **are sa**tisfied.

The physical reality of the experiment might dictate the probability as**signment, as the next exa**mple illustrates.

**Example 1.2.7 (Defining probabilities-II)** The game of darts is played by throwing a dart at a board and receiving a score corresponding to the numbe**r assigned** to the region in which the dart lands. For a novice player, it **seems reasonable to assum**e that the probability of the dart hitting a particular region is proportional to the area of the region. Thus, a bigger region has a higher probability of being hit.

Referring to Figure 1.2.1, we see that the dart board has radius r and the d**istance** between rings is *r/*5. If we make the assumption that the board is always hit (see Exercise 1.7 for a variation on this), **then we have**

Area of region i *P(*scoring i points) = -

Area of dart board

For example

**a*p*2** *- 1 (41/*5)2 *P(*scoring 1 point) = -

7m2

=1-1-.

**Orl**

It is easy to derive the general formula, and we find that

*P (*scoring i points) = 10

)

, i = 1, ..., 5,

independent of 7 and r. The sum of the areas of the disjoint regions equals the area of the dart board. Thus, the probabilities that have been assigned to the five outcomes sum to 1, and, by Theorem 1.2.6, this is a probability function (see Exercise 1.8). 1

**Section 1.2**

**BASICS OF PROBABILITY THEORY**

Figure 1.2.1*. Dart board for Example 1.2.7*

**Before we leave the axioma**tic development of probability, there is one further point **to consider. Axio**m 3 of Definition 1.2.4, which is commonly known as the Axiom of C*o*untable Additivity, is not universally accepted among statistic**ians. Indeed, it can** be argued that axioms should be simple, self-evident statements. Comparing Axiom 3 to the other axioms, which are simple and self-evident, may lead us to doubt whether it is reasonable to assume the truth of Axiom 3.

The Axiom of Countable Additivity is rejected by a school of statisticians led by deFinetti (1972), who chooses to replace this axiom with the Axiom of Finite Additivity.

***Axio****m of Finite Additivity:* If A E *B a*n*d B EB* are disjoint, then

*P(A*U*B) = P(A*) + *P(B)*.

While this axiom may not be entirely self-evident, it is certainly simpler than the **Axi**om of Countable Additivity (and is implied by it - see Exercise 1.12).

**Assu**ming only finite additivity, while perhaps more plausible, can lead to unex pected complications in statistical theory - complications that, at this level, do not **necessar**ily enhance understanding of the subject. We therefore proceed under the assumption that the Axiom of Countable Additivity holds.

*1*.2*.2 The Calculus of Probabilities* . From the Axioms of Probability we can build up many properties of the probability

· function, properties that are quite helpful in the calculation of *mor*e complicated

probabilities. Some of these manipulations will be discussed in detail in this section; others will be left **as exercises.**

We start with some (fairly self-evident) properties of the probability function when applied to a single event.

.

..

*.:*

**PROBABILITY THEORY**

**Section 1.2**

**Theorem 1.2.*8 If P is a probability function and* A *is any set in B, then*** a*. P(*O) = *0, where 0 is the empty set;* b. *P*(A) <1; c. *P(*AC) = 1- *P(*A).

**Proof: It is easiest t**o prove (c) first. The sets A and Ao form a partition of the **sample spa**ce, that is, S = *A*UA. Therefore,

*(*1.2.4)

*P(*AU*A*) = *P(*S) = 1

by the second axiom. Also, A and Ao are disjoint, so by the third axiom,

(1.2.5)

P(*A*U AC) *= P*(A) + *P*(AC).

Combining (1.2.4) and (1.2.5) gives (c).

Since *P*(A) 20, (b) is immediately implied by (c). To prove (a), we use a simi**lar argument o**n S=SUV. (Recall that both S and 0 are always in B.) Since S and 0 are disjoint, we have

1 = *P*(S*) = P(*SUO) = *P*(S) + P*(*O),

and thus *P(O*) = 0.

Theorem 1.2.8 contains properties that are so basic that they also have the fla **vor of ax**ioms, although we have formally proved them using only the original three Kolmogorov Axioms. The next theorem, which is similar in spirit to Theorem 1.2.8, **contains statements that are n**ot so self-evident.

**Theorem 1.2.9 *I****f P is a probability function and* A *and B are any sets in B, then* a*. P(*

*B* AC) *= P(B) - P*(An *B);* b. *P(*AU*B) = P(A*) + *P(B) - P*(

*A B* ); *C. If A CB, then P*(A) < *P(B*).

**Proof: T**o establish (a) note that for any sets A and *B* we **have**

*B = {Bn* A} U{*B* n A},

and therefore

(1.2.6) *P(B) = P*({BN A}U{BN A°}) = *P(BO* A) + *P(BN*A),

where the last equality in (1.2.6) follows from the fact that *Bn* A and B n AC are disjoint. Rearranging (1.2.6) gives (a).

To establish (b), we use the identity

(1.2.*7)*

AU*B=*AU{*B*RAC}.

**Section 1.2**

**BASICS OF PROBABILITY THEORY**

**11**

**A Venn diagram w**ill show why (1.2.7) holds, although a formal proof is not difficult (see Exercise 1.2). Using (1.2*.*7) and the fact that A and *B* n AC are disjoint (since A **and AC are), we have**

(1.2.8)

P(AU*B) =* P(A) + *P(B*NA) = P(A) + *P(B) –* P(A*NB)*

from (a).

If AC*B*, then *AnB* = A. Therefore, using (a) we have

O*SP(BN*A) *= P(B) - P(*A),

establishing (c).

Formula (b) of Theorem 1.2.9 gives a useful inequality for the probability of an intersection. Since *P*(AU*B*) < 1, we have from (1.2.8), after **some rearranging,**

(1.2.9)

*P(*AN*B) > P*(A) + *P(B)* - 1.

This inequality is a special case of what is known as *Bonferroni's Inequality (*Miller 1981 is a good reference). Bonferroni's Inequality allows us to bound the probability of **a simultaneous e**vent (the intersection) in terms of the probabilities of the individual **events.**

**Example 1.2.10 (Bonferroni's Inequal**ity) Bonferroni's Inequality is partic ularly useful when it is difficult (or even impossible) to calculate the intersection probability, but some idea of the size of this probability is desired. Suppos**e A and *B* are two events and** each has probability .95. Then the probability that both will occur is bounded below by

*P(*AN*B*) > P(A) + *P(B)* – 1 = .95 +.95 - 1 = .90.

Note that unless the probabilities of the individual events are sufficiently large, the Bonferroni bound is a useless (but correct!) negative number.

We close this section with a theorem that gives some useful results for dealing with a collection of sets.

**Theorem 1.2.11** *If P is a probability function, then* a. P(A) = *P*(An *Ci*) *for any partition C1, C*2,...; b. *P*(

U 1A;) < Li*-P(*Aj) *for any sets* A1, A2,.....

*(Boole's Inequality)*

**Proof:** Since *C*1, *C2*, ... form a partition, we have that Cin*C;* = 0 for all i + *j,* and S=up *C*ủ. Hence,

A = Ans = n(0) -JAC),

A = AnS = An

in*ci)*,

**i=1**

**i=1**

**12**

**PROBABILITY THEORY**

**Section 1.2**

**where the last equality follows from** the Distributive Law (Theorem 1.1.4). W**e there fore have**

PLA) = P(ÜAnca).

Now, since the *Ci* are disjoint, the sets An*C*i are also disjoint, and from the prope**rties** of a probability function we have

*P*I

PŮLANC:)) - ŠPANC),

*C*i)] =

(*ANC)*,

**12-**

**1**

**i=**

**1**

establishing (a).

To establish (b) we first construct a disjoint collection Ai, A,..., with the property that V AT = VIA;. We define Aby

A1 = A1,

Az = 41, A; = 4: (UA), 1=2,3..

AT = Ail

*11*-1

A b=1

;

*į* = 2,3,..., .

where the notation A *B* denotes the part of A that does not intersect with B. In more familiar symbols, A\*B* = An *B*C. It should be easy to see that A = V Ai, and we therefore have

*r*ool *P*İTAL = *P* U A =

li=1 */* **Vi=1 l , s=1** where the last equality follows since the Ar are disjoint. To see this, we write

p(ÜA.) - P (Ů4:) - ŠP(47),

P(A),

*11*-1

*1\*-1* An At={ALU jyn ALLUA;

(definition of A:)

***j*=*1***

**-**

**1**

4:045 = {« («)}u{a (U4)}(detition of Asy

-{ar (Ux)}{4^(14) } cestion et non ={um74}n{um7as Destingai Leroy

i\_1

***k***

**-1**

= *{* din na naterina

(DeMorgan's Laws) ***j*=1** l*j*=1 Now if i > k, the first intersection above will be contained in the set Af, which will **have an em**pty intersection with Ak. If *k > i*, the argument is similar. Further, by construction Am C Ai, so *P*(A) < *P*(A) and we have

ŠP(47)\*P(A),

***i*=**

**1**

establishing (b).

**Section 1.2**

**BASICS OF PROBABILITY THEORY**

**There is a similarity between Boole's Inequality and Bonferroni's Inequality. In fact, they are essentially the same** thing. We could have used Boole's Inequality to **der**ive (1.2.9). If we apply Boole's Inequality to A', w**e have**

P(«) <ŠP(43).

**li=1**

**/**

**i=1**

and using the facts that UA = (NA) and *P*(A) = 1 - *P*(Ai), we obtain

1-P(@m) sn-Pra).

1-*P*IN Aila*n*

**Vi=1**

This becomes, o**n rearranging terms,**

(1.2.10)

P(4) > ŠP(43) = (n − 1),

li=1

*/*

**i=**

which is a more general version of the Bonferroni Inequality of (1.2.9).

*1.*2*.3 Counting* The elementary process of counting can become quite sophisticated when placed in **the hands of a statistic**ian. Most often, methods of counting are used in order to construct probability **assignments** on finite sample spaces, although they can be used **to answer other questions also.**

**Example 1.2.12 (Lotte**ry-I) For a number of years the New York state lottery **operated** according to the following scheme. From the numbers 1, 2, ..., 44, a person **may pick any six f**or her ticket. The winning number is then decided by randomly selecting six numbers from the forty-four. To be able to calculate the probability of winning we first must count how many different groups of six numbers can be chosen from the forty-four.

**Example 1.2.13 (Tournament) In a sin**gle-elimination tournament, such as the U.S. Op**en tennis tournament, players advance** only if they win (in contrast to double elimination or round-robin tou**rnaments). If we have 16 entrants, we m**ight be inter **est**ed in the number of paths a particular player can take to victory, where a path is **taken to mean a sequence of opponents.**

Counting problems, in general, sound complicated, and often we must do our count ing subject to many restrictions. The way to solve such problems is to break them down into a series of simple tasks that are easy to count, **and employ known rules** of combining tasks. The following theorem is a first step in such a pr**ocess and is sometimes known as th**e Fundamental Theorem of Counting.

**Theorem 1**.2.14 *If a job consists of k separate tasks, the ith of which can be done in ni ways, i* = 1,...,*k, then the entire job can be done in n*i x *7*2 x ...*x nk ways.*

**14**

**:**

**PROBABILITY THEORY**

**Section 1.2**

**Proof: It suffices to prove the theorem for k = *2* (see Exerci**se 1.15). The proof is **just a matter of care**ful counting. The first task can be done in ni **ways, and for each** of these ways we have n2 choices for the second task. Thus, we can do the job in

(1 xn2) + (1 x n2)+...+(1 x n2) = ni **x na**

**ny terms** ways, establishing the theorem for *k* = 2.

**Example 1.2.15 *(*Lottery-II)** Although the Fundamental Theorem of Counting is a reasonable place to start, in applications there are usually more aspects of a **problem to consider. For examp**le, in the New York state lottery the first **number** can be chosen in 44 ways, and the second number in 43 ways, making a total of 44 x 43 = 1,892 ways of choosing the first two numbers. However, if a person is allowed to choose **the same number t**wice, then the first **two numbers can be chosen** in 44 x 44 = 1,936 ways.

The distinction being made in Example 1.2.15 is between counting *with replacement* and counting *without replacement.* There is a second crucial element in any counting problem, whether or not the ordering of the tasks is important. To illustrate with the lottery example, suppose the winning numbers are selected in the order 12, 37, 35, 9, 13, 22. D**oes a person w**ho selected 9, 12, 13, 22, 35, 37 qualify as a winner? In other words, does the order in which the task is performed actually matter? Taking all of these considerations into account, we can construct a 2 x 2 table of possibilities:

*Possible methods of counting*

Without With

replacement replacement **Ordered**

Unordered

Before we begin to count, the following definition gives us s**ome extremel**y helpful notation.

**Definition 1.2.16 F**or a positive integer n, n! *(*read n factorial) is the product of all of the positive integers less than or equal to n. That is,

a! = n *= (*m - 1) x (m - 2) x x 3 x 2 x 1.

**Furthermore, we d**efine 0! = 1.

Let us now consider counting all of the possible lottery tickets under each of these **four cases.** 1. *Ordered, without* r*eplaceme*nt From the Fundamental Theorem of Counting, the

first number can be selected in 44 ways, the second in 43 ways, etc. So there are

**44**

X 39 = 291 = 5,082,517,440

possible tickets.

**Section 1.2**

**BASICS OF PROBABILITY THEORY**

2*. O****rdered, with replacem*en*t* Since each number can now be selected in 44 ways**

**(because the chosen number** is replaced), there **are**

**44 x 44 x 44 x 44 x 44 x 44** = 446 = 7,256,313,856 possible tickets. *3. Unordered, without replacement* We know the number of possible tickets when the

ordering must be accounted for, so what we must do is divide out the re**dundant** orderings. Again from the Fundamental Theorem, six numbers can be arranged in 6 x 5 x 4 x 3 x 2 x 1 ways, so the total number of unordered tickets is

44 x 43 x 42 x 41 x 40 x 39 44!

6 5 \* 4\*3\*2\*1 This form of counting plays a central role in much of statistics so much, in fact, that it has earned its ow*n* notation.

**6**1*201 = 7,*059,052

**2X*1***

**Definition 1.2.17** For nonnegative integers n and r, where n >r, we define the **sym**bol (), read *n choose r*, as

*in*

*n!*

r! *(n* - )! In our lottery **exam**ple, the number of possible tickets (unordered, without replace ment) is (\*\*). These numbers are also referred to as *binomial coefficien****ts,* for reasons** that will become clear in Chapter 3.

4*. Unordered, with replacement* This is the most difficult case to count. You might

first guess that the answer is 448/(6 x 5 x 4 x3 x 2 x 1), but this is not correct (it is too small). To count in this case, it is easiest to think of placing 6 markers on the 44 n**umbers.** In fact, we can think of the 44 numbers defining bins in which we can place the six markers, M, as shown, for example, in this figure.

| ML | MM | M | ... | M | M | L |

1 2 3 4 5 ... **41 42 43 *4*4**

The number of possible tickets is then equal to the number o**f ways that we can** put the 6 **markers i**nto the 44 bins. But this can be further reduced by noting that all we need to keep track of is the **arrangement of the markers a**nd the walls of the bins. Note further that **the two outermost wa**lls play no ***p*art. Thus, we *h*ave to** count all of the arrangements of 43 walls (44 bins yield 45 walls, but we d**isregard** the two end walls) and 6 markers. We therefore have 43 + 6 = 49 objects, which can be arranged in 49! ways. However, to eliminate the redundant orderings we must divide by both 6! and 43!, so the total n**umber of arrangements is**

49!

= 13,983,816.

6! 43! Although all of the preceding derivations were done in terms of an example, it should be easy to see that they hold in general. For co**mpleteness, we can summarize these s**ituations in Table 1.2.1.

**\***

**16**

**PROBABILITY THEORY**

**Soction 1.2**

**Table 1.2.1*. Number of possibl*e a*rr*an*g*ements o*f size r from n objec*ts**

With **replacement**

**Ordered**

Without **replacement**

***n!*** in – „)!

6

Unordered

Intr–1\

*1.2.4 Enumerating Outcomes*

The counting techniques of the previous section are useful **when the sample space** *S* is a finite set and all the outcomes in S are equally likely. Then probabilities of **events ca**n be calculated by simply counting the number of outcomes in the event. To see this, suppose that *S = {8*1,...,*s*n} is a finite sample space. Saying that all the outcomes are equally likely means that *P(*{si}) = *1/N f*or every outcome si. Then, using Axiom 3 from Definition 1.2.4, we have, for any event A,

1 *#* of elements in A *P*(A) = 2 P({s*i})* = 2 N = # of elements in 5

**SiEA** For large sample spaces, the counting techniques might be used to calculate both the numerator and denominator of this **expression.**

SE*A*

**Example 1.2.18 (Poker**) Consider choosing a five-card poker hand from a stan dard deck of 52 playing cards. Obviously, we are sampling without replacement from the deck. But to specify the possible outcomes (possible hands), we must decide whether we think of the hand as being dealt sequentially (ordered) or all at once (unordered). If we wish to calculate probabilities for events that depend on the or der, such as the probability of an ace in the first two car**ds, then we must use the ordered outcomes. But if our events do not d**epend on the ord**er, we can use the** unordered outcomes. For this example we use the unordered outco**mes, so the sample space c**onsists of all the five-card hands that can be chose**n from the 52-card deck.** There are *(9*3) = 2,598,960 possible hands. If the deck is well shuffled and the cards **are ran**domly dealt, it is **reasona**ble to assign probability 1*/*2,598,960 to each possible **hand.**

We now calculate some probabilities by counting outcomes in events. What is the probability of having four aces? How many different hands are there with four ace*s?* If we specify that four of the cards are aces, then there are 48 different ways of specifying the fifth card. Thus,

**48**

*P*(four aces) = 2.598.960'

**less t**han 1 chance in 50,000. Only slightly more complicated counting, using Theorem 1.2.14, allows us to calculate the probability of having four of a kind. There are 13

**Section 1.2**

**BASICS OF PROBABILITY THEORY**

***17***

**ways to s**pecify which denomination there will be four of. After we specify **these four** cards, there are 48 ways of specifying the fifth. Thus, the total number of hands with four of a kind is (13)(48) and

*P*(four of a kind*)* = 2.598.960 = 2 598.960

(13)(48) 624

To calculate the probability of exactly one pair (not two pairs, no three of a kind, etc.) we combine some of the counting techniques. The number of hands with exactly one pair is

(1.2.11)

143 = 1,098,240.

| *3*

Expression (1.2.11) comes from Theorem 1.2.14 because

13 = *#* of ways to specify the denomination for the pair,

1) *= # o*f ways to specify the two cards from that denomination,

*= # o*f ways of specifying the other **three de*nom*inati*o*ns,**

*43 = #* of ways of specifying the other three cards from those **denominations.**

Thus,

1,098,240

*P*(exactly one pair) = 2.598,960

When sampling without replacement, as in Example 1.2.18, if we want to calculate the probability of an event that does not depend on the order, we can use either **the ordere**d or unordered sample space. Each outcome in the **unordered sample space corres**ponds to r! outcomes in the ordered sample space. So, when counting ou**tcomes** in the ordered sample space, we use a factor of r! in both the numerator and denom **inat**or that will cancel to give the same probability as if we counted in the unordered **sample space.**

The situation is different if we sample with replacement. Each outcome in the unordered sample space corresponds to some outcomes in the ordered sample space, but the number of outcomes differs.

**Example 1.2.19 (Sampling with replacement)** Consider sampling r = 2 items from n = 3 items, with rep**lacement. The outcomes in the ordered and unordered sample spaces are these.**

Unordered (1,1} {2,2} {3,3} Ordered (1, 1) (2, 2) (3,3) Probability 1/

9*1 /*

*91 /*

*9*

{1,2} (1,2), (2, 1)

*2 /9*

{1,3} (1,3), (3, 1)

{2,3} (2,3), (3, 2)

2/9

**18**

**PROBABILITY THEORY**

**Section 1.2**

**The probabilities come from considering the nine outcomes in the ordered sample space to be equally likely. This corresponds to the common interpretat**ion of "sampling with replacement"; namely, one of the three items is chosen, each with probability 1*/*3; the item is noted and replaced; the **items are mixed and again on**e of the three items **is chosen, eac**h with probability 1/3. It is seen that the six outcomes in the unordered **sample space are** not equally likely under this kind of sampling. The formula for the **num**ber of outco**mes in the unordered sample space i**s useful for enumerating the outcomes, but ordered outcomes must be counted to correctly calculate probabilities.

Some authors argue that it is appropriate to assign equal probabilities to the un ordered outcomes when "randomly distributing r indistinguishable balls into n dis tinguishable urns." That is**, an urn is chosen at random an**d a ball placed in it, and this is repeated r times. The order in which the balls are placed is not recorded so, in the end, an outcome such as {1,3} means one ball is in urn 1 and one ball is in **urn 3.**

But here is the problem with this interpretation. Suppose two people observe this process, and Observer 1 records the order in which the balls are placed but Observer 2 does not. Observer 1 will assign probability *2/*9 to the event (1,3). Observer 2, **who is observing exactly the same process**, should also assign probability 2/9 to this event. But if the six **unordered outcomes are writt**en on identical pieces of paper and one is randomly chosen to determine the placement of the balls, then the unordered **outcomes each hav**e probability 1*/*6. So Observer 2 will assign probability 1*/*6 to the event (1,3).

The confusion arises because the phrase "with replacement” will typically be inter preted with the sequential kind of sampling we described above, leading **to assigning** a probability 2/9 to the event (1,3). This is the co**rrect way to proceed, as proba** bilities should be determined by the sampling mechanism, not whether the balls are distinguishable or indistinguishable.

**Example 1.2.20 (Calculating an average) A**s an illustration of the distinguish able/indistinguishable approach, suppose that we are going to calculate all possible **averages of four numbers se**lected from

2,4,9,12

**where we draw the numbers with replacement. For examp**le, possible d**raws are** {2, 4, 4,9} with average 4.75 and {4,4,9,9} with average 6.5. If we are only inter **ested in the average of the sampled** numbers, the ordering is unimportant, and thus the total number of distinct samples is obtained by counting according to **unordered, with-replacement sa**mpling.

The total number of distinct samples is (n+n-). But now, to calculate the proba bility distribution of the sampled **averages, we m**ust count the different ways that a particula**r average can occur.**

The value 4.75 can occur only if the sample con**tains on**e 2, two 4s, and one 9. The number of possible samples that have this configuration is given in the following table:

**i*n***

**Section 1.2**

**BASICS OF PROBABILITY THEORY**

**19**

**Probability**

**.12**

***.00***

**uur**

**2**

**4**

**10**

**12**

**6 8 Average**

Figure 1.2.2*. Histogram of averages of samples with replacement from the fo****ur numbers*** {2,4,4,9}

Unordered

{2,4,4,9}

Ordered (2, 4, 4,9), (2, 4, 9,4), (2, 9, 4, 4),(4, 2, 4,9), *(*4,2, 9,4),(4,4, 2,9),(4,4,9,2), (4,9, 2, 4), (4,9,4, 2), (9,2, 4, 4), (9, 4, 2, 4), (9, 4, 4, 2)

The total number of ordered samples is n" = 44 = 256, so the probability o**f drawing** the unordered sample {2,4,4,9} is 12/256. Compare this to the probability that we would have obtained if we regarded the unordered samples as equally likely - **we would have as**signed probability 1*/(*a+n *1*Q = 1/35 to {2,4,4,9} and to every other unordered sample.

To count the number of ordered samples that would result in {2,4,4,9}, we argue **as follows. We need to enumer**ate the possible orders of the four numbers {2, 4, 4,9}, **80 we are es**sentially using counting method 1 of Section 1.2.3. We can order the **sample in 4x3**x2x1 = 24 ways. But there is a bit of double counting here, **since we canno**t count distinct **arrangement**s of the two 4s. For example, the 24 ways would count {9,4, 2,4} twice (which would be OK if the 4s were different). To correct this, we divide by 2! (there are 2! ways to arrange the two 4s) and obtain 2*4/*2 = 12 ordered samples. In general, if there are k places **and we have *m* different numbers repeated** *ki, k2, ...,* km times, then the number of ordered samples is niz of counting is related to the *multinomial distribution,* which we will see in Section 4.6. Figure 1.2.2 is a histogram of the probability distribution of the sample averages, reflecting the multinomial counting of the samples.

There is also one further refinement that is reflected in Figure 1.2.2. It is possible that two different unordered samples will result in the **same mean. For exa**mple, the unordered samples {4, 4, 12, 12) and {2,9,9,12} both result in **an average val**ue of 8. The first sample has probability 3/128 and the second has probability 3*/*64, giving the value 8 a probability of 9*/*128 = .07. See Example A.0.1 in Appendix A for details on constructing such a histogram. The calculation that we have done in this example is **an elementary versio**n of a very important statistical technique known as the *bootstrap* (Efron and Tibshirani 1993). We will return to the bootstrap in Section 10.1.4. ||

ki!k?!*...k*i. *T*his type

**PROBABILITY: THEORY**

**Section 1.3**

**1.3 Conditional Probability and Independence** All of the probabilities **that we have dea**lt with thus far have been unconditional probabilities. A **sample space was defin**ed and all probabilities **were calculated with respect to that sample space. In many instances, however, we are in a position to update the sample space based on new informatio**n. In such c**ases, we want to be able** to update probability calculations or to calculate co**n*ditional probabiliti*e*s.***

**Example 1.3.1 *(*Four aces)** Four cards are dealt from the top of a well-shuffled deck. What is the probability that they are the four aces*?* We can calculate this probability by the methods of the previous section. The number of distinct **groups of four cards is**

= 270,725.

**Only one of these groups consist**s of the four aces and every group is equally likely, so the probability of being dealt all four aces is *1/270,*725.

We can also calculate this probability by an “updating" argument, as follows. The probability that the first card is an ace is 4/52*. Given that the first card is a*n a*ce,* the probability that the second card is an ace is 3/51 (there are 3 aces and 51 cards left). Continuing this argument, we get the desired probability as

4 3 2 1 1 52 \* 51 \* 50 \* 49 270*,7*2*5*

In our second method of solving the problem, we up**dated the sample space after** each draw of a card; we calculated conditional probabilities.

**Definition 1.3.2 If A and B are event**s in S, and *P*(*B*) > 0, then the *conditional probability of A given B,* written *P*(AB), is

(1.3.1)

*P(ANB) P(*A\*B*) = ?

*P(B)*

Note that what happens in the conditional probability calculation is that B becomes the sample space: *P(B\B*) = 1. The intuition is that our original sample space, S*,* has been updated to *B*. All further occur**rences are then calibrated with respect to** their relation to *B*. In particular, note what happens to conditional probabilities of disjoint sets. Suppose *A* and *B* are disjoint, s*o P*(

*A B* ) = 0. It then follows that *P(AB*) *= P(BA*) = 0.

**Example 1.3.3 (Continuation of Example 1.3.1) A**lthough the probability of getting all four aces is quite small, let us see how the conditional probabilities c**hange given that some aces have already been drawn. Four cards will again be dealt from a** well-shuffled deck, and we now calculate

*Ps*4 aces in 4 cards | i aces in *i* cards),

*į* = 1,2,3.

**Section 1.3**

**CONDITIONAL PROBABILITY AND INDEPENDENCE**

**The event {4 aces in 4 cards) is a subset of the event {i aces in i cards). Thus, from** the definition of conditional probability, (1.3.1), we know **that**

*P(*4 aces in 4 cards i aces in i cards)

*P(*{4 aces in 4 cards} n {i aces in i cards))

*P(i* aces in i cards) *P*(4 aces in 4 cards)

*P(i* aces in i cards) The numerator has already been calculated, and the denominator can be calculated with a similar argument. The number of distinct groups of i cards is (?), and

***P(i* aces in i cards) =**

52

**Therefore, th**e conditional probability is given by

*P(*4 aces in 4 cards | i aces in i cards) =

(52) *(*52) (1)

(4 – 1)!48! (52 – 1)!

1 *75*2–

For i = 1, 2, and 3, the conditional probabilities are .00005, .00082, and .02041, **respectively.**

For an*y B* for which *P(B*) > 0, it is straightforward to verify that the probability function *P(*:*|B*) satisfies Kolmogorov's Axioms (see Exercise 1.35). You may suspect that requiring *P(B*) > 0 is redundant. Who would want to condition on an event of probability 0? Interestingly, sometimes this is a particularly useful way of thinking of things. However, we will defer these considerations until Chapter 4.

Conditional probabilities can be particularly slippery entities **and sometimes require** careful thought. Consider the following often-told tale.

**Example 1.3.4 (Three prisoners) Three prisoners**, A, B, and C, are on death Tow. The governor decides to pardon one of the three and **chooses at random the prisoner to pardon. H**e informs the warden of his choice but requests that the name be kept secret for a few days.

**The next day, A tries to get the warden t**o tell him who **had been pardoned. The warden refuses. A then asks** which of B or C will be executed. The warden thinks for a while, then tells A that B is to be **executed. Warden's reasoning: Each prisoner has** a chance of being pardoned. Clearly,

either B or C must be executed, so I have given A no information about whether

A will be pardoned. **A's reasoning:** Given that B will be executed, then either A or C will be pardoned.

My chance of being pardoned has risen to.

It should be clear that the w**arden's reasonin**g is correct, but let us see why. Let A*, B*, and *C* denote the events that A, B, or C is pardoned, respecti**vely. We know**

**PROBABILITY THEORY**

**Bo*c*tion 1.3**

that *P(*A) *= P(B) = P(C*)= . **Let W denote the event that the warden says B wi**ll die. Using (1.3.1), A can update his probability of bein**g pardoned to**

*P(*AW) = *P*(Anw)

*P*(W)

What is happening can be summarized in this table:

**Prisoner pardon**ed

Warden tells A

B dies C dies C dies B dies

each with equal

probability

Using this table, we can calculate

*P*(W) = *P(*warden says B dies)

*= P*(warden says B dies and A pardoned)

+ P(warden says B dies and C pardoned) + *P*(warden says B dies and B pardoned)

**=**

**.**

**= + = + 0** 6'3'

Thus, using the **warden's reasoning, we have**

*P(*A\W) =

P(Anw)

*P*(W) *P*(warden says B dies and A pardoned)

*P*(warden says B dies)

(1.3.2)

*1/6* 1 = 1*/*2 = 3

**However**, A falsely interprets the event was equal to the **event *B*o and calculates**

*P(*An *B)*

*P(*BC)

*P*(A*B*C) =

1*/*3 = 2/3

1

We see that conditional probabilities can be quite slippery and require careful interpretation. For some other variations of this problem, see Exercise 1.37.

**Re-expressin**g (1.3.1) gives a useful form for calculating intersection probabilities,

(1.3.3)

*P(A*N*B) = P*(A*B)P*(*B)*,

which is essentially the formula that was used in Example 1.3.1. We ca**n take advan tage o**f the symmetry of (1.3.3) and also write

(1.3.4)

*P*(AN*B) = P(BA*)*P*(A).

**Section 1.3**

**CONDITIONAL PROBABILITY AND INDEPENDENCE**

**29**

**When faced with seeming**ly difficult calcul**ations, we can break up our calculations** according to (1.3.3) or (1.3.4), w**hichever is easier. Furthermore, we can equate the** right-h**and sides of these equation**s to obtain (afte**r rearrangement)**

*P*(AB) = *P*(*B*A) P*IB*)

*P(*A)

(1.3.5)

**which gives us a formu**la for "turning around" conditional probabilities. Equation (1.3.5) is often called Bayes' Rule for its discoverer, Sir Thomas Bayes (although see Stigler 1983).

Bayes' Rule **has a more general f**orm than (1.3.5), one that applies to partitions of a sample space. We therefore take the following as the definition of Bayes' Rule.

**Theorem 1.3.5 (Bayes' Rule**) *Let* A1, A2, .*.. be a partition of the sample spa*ce, *and let B be any set. Then, for each i* = 1, 2, ...,

*P(B*1A*)P(A*)

*P*(AiB) = no *P(B*A;)*P(*A;)

**Example 1.3.6 (Coding)** When coded m**essages are sent, there are sometimes errors in transmission. In pa**rticular, Morse code uses "dots" and "dashes," which are **known to** occur in the proportion of 3:4. This means that for any given symbol,

*P(*dot sent) =

and P(dash sent) =

Suppose there is interference on the transmission line, and with probability a dot **is mistakenly received as a dash, and vice versa. If we receive a dot, can we be sure** that a dot was sent? Using Bayes' Rule, we can write

*P*(dot sent) *P*(dot sent | dot received) = P*(*dot received | dot sent)

P(dot received)

Now, from the information given, we know that P(dot sent) = ? and P(dot received, dot sent) = . Furthermore, we can also write

*P(*dot received) = P*(*dot received n dot sent) + *P*(dot received n dash sent)

*= P(*dot received | dot sent*)P(*dot sent)

+ *P*(dot received dash sent)*P*(dash sent) 7 3 1 4 25 = 3x3 + 3 \* = 56

Combining these results, we have that the probability of correctly receiving a dot is

21 *P*(dot sent dot received) – *(7/8)* X (3/7)

ved) = 2*5/5*6 =25

**PROBABILITY THEORY**

**Section 1.3**

**In some cases it may happen that the occurrence of a particular event, *B*, has no effect on the probability of another event**, A. Symbolically, **we are saying that**

(1.3.6)

*P(AB) = P(A*).

If this holds, then by Bayes' Rule (1.3.5) and using (1.3.6) we have

(1.3.7*)*

*P(B*[4) = P(A/B*)* PCE) = P(4) P(B) = P*(B),*

**P*A***

so the occurrence of A has no effect on *B. Mo*reover, since *P(B\AP*(A) = *P(*A*B*), it then follows that

*P(ANB) = P(A)P(B),*

which we take as the definition of statistical independence.

**Definition 1**.3.7

Two events, A and *B,* a*re statistically independent* if

(1.3.8)

*P(*AN*B) = P(A)P(B*).

Note that independence could **have been equivale**ntly defined by either (1.3.6) or (1.3.7) (as long as either *P(A*) >0 or *P(B*) > 0). The advantage of (1.3.8) is that it treats the events symmetrically and will be easier to generalize to more than two **events.**

Many gambling games provide models of independent events. The spins of a roulette wheel and the tosses of a pair of dice are both series of independent events.

**Example 1.3.8 (Chevalier de Meré)** The gambler introduced at the start of the chapter, the Chevalier de Meré, was particularly interested in the event that he could throw at least 1 six in 4 rolls of a die. We have

*P*(at least 1 six in 4 rolls) = *1 - P*(no six in 4 rolls)

IP(no six on roll i),

**i=1**

where the last equality follows by independence of the rolls. On any roll, the proba bility of *no*t rolling a six is, so

*P(*at least 1 six in 4 rolls) = 1 -

=.518.

Independence of A and *B* implies independence of the complements also. In fact, we have the following the**orem.**

we have the momenting the

implies indepen

**Section 1.3**

**CONDITIONAL PROBABILITY AND INDEPENDENCE**

**25**

***If A and B are independen****t events, then the following* ***pairs are***

**Theorem 1.3.9** *also independent:* a. A *and BC,* b. A *and B,* C. A *and BC,*

**Proof: W**e will prove only (a), leaving the rest as Exercise 1.40. To prove (a) we must show that *P(A BC*) *= P(A)P(BC)*. From Theorem 1.2.9a we have

(A and *B* are independent)

*P*(

A*B*C) = *P(*A) - *P*(AN*B*)

*P*(A) – *P*(A*)P(B*) *= P*(A)(1 – *P(B)*) *= P(*A*)P(BC*).

Independence of more **than two events can be defined in a manner** similar to (1.3.8), but we must be careful. For example, we might think that we could say A*, B*, and C are independent if *P*(

A *B C* ) = *P*(A)*P(B)P(C*). However, this is not the correct condition.

**Example 1.3.10 (Tossing two dice)** Let an experiment consist of tossing two dice. For this experi**ment the sample space is**

S = {(1, 1), (1, 2),...,(1,6), (2, 1), ..., (2,6),..., (6,1), ...,(6,6)};

that is, S consists of the 36 ordered pairs formed from the numbers 1 to 6. Define the following events:

A = {doubles appear} = {(1, 1), (2, 2), (3, 3), (4,4), (5,5), (6,6)}, *B*= {the sum is between *7* and 10}, C = {the sum is 2 or 7 or 8).

The probabilities can be calculated by counting among the 36 possible outcomes. We **have**

P(A) = a; *P(B*) = ? and PC) =

Furthermore,

*P*(AN*BOC) = P*(the sum is 8, composed of double 4s)

36

**1 1 1** - X - Xa **6 2 3**

*= P(A)P(B)P(C)*.

**20**

.

.

.

.

**PROBABILITY THEORY**

**Section 1.3**

**Howe*v*er,**

*P(BNC) = P(*sum equals 7 or 8) = 1*1 + P(B)P(C).*

**Sim**ilarly, it can be shown that *P(A*NB*) + P(*A*)P(B*); therefore, the **requirement** *P(*An *BnC*) *= P(A*)*P(B)P(C*) is **not a strong enough condition to guarantee pairw**ise independence.

A second attempt at a general definition of independence, in light of **the previ ous exam**ple, might be to define A, *B, a*nd *C* to be independent if all the p**airs are** independent. Alas, this condition also fails.

**Example 1.3.11 (Letters)** Let the sample space S consist of the 3! permutations of the letters a, b, and c along with the three triples of each letter. Thus,

S=

aaa bbb ccc abc bca cba. **( acb bac cab**

Furthermore, let each element of S have probability. **Define**

Ai = {ith place in the triple is occupied by a}.

It is then easy to count that

P(Aj) = 5; i = 1,2,3,

**and**

*P*(Ain A2) *= P(A*in A3) = *P*(Azn A3) =

**01**

so the A*i*s are pairwise independent. But

P(A1 N A2 n 43) = 5 + P(A1)P(A3)*P*(A3),

so the Ais do not satisfy the probability **requirement.**

**The preceding two examples show that** simultaneous *(*or mutual) independence of a collection o**f events requires an extremely stron**g definition. The following definition **works.**

**Definition 1.3.12** A collection of events A1,..., *A*n are *mutually independent* if for any subcollection Aing..., Air, w**e have**

(4.) - P(0.).

**Section 1.4**

**RANDOM V*A*RIABLES**

**Example 1.3.13 (Three coin tosses-I) Consider the experiment of tossing a** coin three times. A sample point for this experiment must indicate the result of each toss. For example, HKT could indicate that two heads and then a tai**l were observed.** The sample space for this experiment has eight points, namely,

{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT). Let *Hi, i* = 1,2,3, denote the event that the ith toss is a head. For **example,**

(1.3.9)

*H1* = {HHH, HHT, HTH, HTT).

**If we assign pr**obability to each sample point, then using enumerations such as

(1.3.9), we see that *P(H*1) *= P(H2) = P(H*3) = . This says that the coin is fair and

**· has an equa**l probability of landing heads or tails on each toss.

Under this probability model, the events *H1, H2,* and *H3* are also mutually inde **pend**ent. To verify this we note **that**

1 1 1 1 *P(Hin Han H3*) = P({HHH

2.= *P*(*H*1*)P(H2)P(Hz).* To verify the condition in Definition 1.3.12, we also must check each pair. For example,

*P(H, N H2*) = P({HHH, HHT}) = = *= = P(H1)P(Hz).* The equality is also true for the other two pairs. Thus, *H1, H2,* and *H*3 are mutually independent. That is, the occurrence of a head on any toss has no effect on any of the other **tosses.**

It can be verified that the assignment of probability to each sample point is the only probability model that has *P(Hi*) = *P(H2) = P(H*3) = and *H1, H2, a*nd *H3* mutually independent.

**1.4 Random Variables In many ex**periments it is easier to deal with **a summary vari**able than with the original probability structure. For example, in an opinion poll, we might decide to **ask** 50 people whether they agree or disagree with a certain issue. If we record a "1" for agree and "O" for disagree, the sample space for this experiment has 250 elements, each an ordered string of 1s and Os of length 50. We should be able to reduce this to **a reas**onable size! It may be that the only quantity of interest is the number of people who agree (equivalently, disagree) out of 50 and, if we define a v**ariable X = number** of ls recorded out of 50, we have captured the essence of the problem. Note that the **sam**ple space for X is the set of integers {0, 1, 2, ...,50} and is much easier to deal with than the original sample space.

In defining the quantity X, we have defined a mapping (a function) from the original **sample space to a new sample sp**ace, usually a set of real numbers. In general, we have the following definition.

***A rando****m variable* is a function from a sample space S into the

**Definition 1.4.1 real numbers.**

**PROBABILITY THEORY**

**Section 1.4**

**Example 1.4.2 (Random variables) In some experiments random variables are** implicitly u**sed; some examples are these.**

*Examples of random variables* Experiment

**Random variable**

Toss two dice

X = sum of the numbers Toss a coin 25 times X = number of heads in 25 tosses Apply different amounts of

fertilizer to corn plants X = yield*/*acre

In defining a random variable, we have also defined a new sample space (the range of the r**andom varia**ble). We must now check formally that our probability function, which is defined on the original sample space, can be used for the random variable.

**Suppose we have a sam**ple space

S={81, ...,Sn}

with a probability function *P* and we define a random variable X with range x =

{21,..., Im}. We can define a probability function *P*x on X in the following way. We will observe X = I; if and only if the outcome of t**he random experim**ent is an si ES such that X($;) = Iį. Thus,

(1.4.1)

*P*X(X = x*i)* = *P(S*jES:X*(33*) = Ii}).

Note that the left-hand side of (1.4.1), the function *P*x, is an *induced* probability function on X, defined in terms of the original function *P*. Equation (1.4.1) formally defines a probability function, Px, for the random variable X. Of cou**rse, we have** to verify that *P*x satisfies the Kolmogorov Axioms, but that is not a very difficult job (see Exercise 1.45). Because of the equivalence in (1.4.1), we will simply write

*P*(X = Ii) rather than *P*x*(X* = Iį). A n*ote on notation*: Random variables will always be denoted with uppercase letters and the realized values of the variable (or its range) will be denoted by the corre sponding lo**wercase le**tters. Thus, the random variable X can take the value I.

**Example 1.4.3 (Three coin tosses-II) Consider again the experimen**t of tossing a fair coin three times from Example 1.3.13. Def**ine the random varia**ble X to be the **number of head**s obtained in the th**ree tosses. A complete enumeration of the value** of X for each point in the sample space is

*9* X(8)

HHH HHT HTH THH TTH THT HTT TTT

3 2 2 **2 1 1 1 0**

The range for the random variable X is X = {0,1,2,3). Assuming that all eight points in S have probability, by simply counting in the above display we see that

**Section 1.5**

**DISTRIBUTION FUNCTIONS**

*2*9

the induced probability function on X is given by

2

0

1

2

3

For example, *P*x(X = 1) = P({HTT, THT, TTH}) =

**ex**

**Example 1.4.4 (Distribution of a random variable)** It may be possible to determine *P*x even if a complete listing, as in Example 1.4.3, is not possible. Let S be the 250 strings of 50 Os and 1s, X = number of 1s, and X = {0, 1, 2, ..., 50), as mentioned at the beginning of this section. Suppose that each of the 250 strings is equally likely. The probability that X = 27 can be obtained by counting all of the strings with 27 1s in the original sample space. Since each string is equally likely, it follows that

*1*501 **127*)***

*Px(*X = 27) *– #* strings with 27 ls – (2*7)*

***#* strings**

In general, for any i ex,

*P*X(X = *i)* =

The previous illustrations had both a finite S and finite X, and the definition of ***Px* was s**traightforward. Such is also the case if X is countable. If X is uncountable, we define the induced probability function, *P*x, in a manner similar to (1.4.1). For **any** set ACX,

(1.4.2)

*PX(*X E A) = *P*({s ES: X(s) E A}).

This does define a legitimate probability function for which the Kolmogorov Axioms can be verified. *(T*o be precise, we use *(1*.4.2) to define probabilities only for a cer tain sigma algebra of subsets of X. But we will no**t concern our**selves with these technicalities.)

**1.5 Distribution Functions** With every random variable X, we associate a function called the cumulative distri bution function of X.

**Definition 1**.5.1 The cu*mulative distribution function* or *cdf* of a random variable X, denoted by *F*x(2), is defined by

*Fx(x*) *= Px*(X SX), for all z.

**PROBABILITY THEORY**

**Section 1.5**

*Fx* (\*) ***1.*000**

***.875***

**.375**

**.250**

Figure 1.5.1. *Cdf of Example 1.5.2*

**Example 1.5.2 *(*Tossing three coins)** Consider the experiment of tossing three fair coins, and let X = number of heads observed. The cdf of X is

(1.5.1)

if -0 <x<0

if I SI<1 *F*x(x) = { } if 15 x < 2

if 2 5 2 <3 1 if 3 <3 <.

The step function *F*x() is graphed in Figure 1.5.1. There are several points to note from Figure 1.5.1. *Fx* is defined for all values of I, not just those in X = {0,1,2,3). Thus, **for example,**

*F*x (2.5) *= P*(X < 2.5) = P(X = 0, 1, or 2) =

Note that *F*x has jumps at the values of Ii EX and the size of the jump at Xi is equal to *P(*X = *x*i). Also, *F*x(x) = 0 for x < 0 since X cannot be negative, and *F*x(x) = 1 for x>3 since x is certain to be less than or equal to such a value. |

As is apparent from Figure 1.5.1, *F*x can be discontinuous, with jumps at certain values of x. By the way in which *F*x is defined, however, at the jump points *F*x takes the value at the top of the jump. (Note the different inequalities in (1.5.1).) This is **known as *ri****ght-continuity—*the function is continuous when a point is approached from the right. The property of right-continuity is a consequence of the definition of the cdf. In contrast, if we had defined *Fx(*x) = *Px(*x < x) (note strict inequality), *F*x would then be *left-continuous.* The size of the jump at any point x is equal to *P(*X = 3).

Every cdf satisfies certain properties, some of which are obvious when we think of the definition of *F*x (x) in terms of probabilities.

**Section 1.5**

**DISTRIBUTION FUNCTIONS**

**31**

**Theorem 1.5.3 *Th****e functio*n *F(x) is a cdf if and only if the following three con* ***ditions*** *hold:* **a.** lim.--*-F(*x) = 0 *and* limg-00*F(*I) = 1. **b. *F*(x) *is a nondecreasin****g function of I.* c. *F*(x) *is right-continuous; that is, for every number* xo, limx170 *F(x) = F(2*0).

**Outline of proof: To prove necessity, the three properties can** be verified by writing *F* in terms of the probability function (see Exercise 1.48). To prove sufficiency, that if a function *F* **satisfies the three conditions of the theorem then it is a cdf *f*or some random variabl**e, is much harder. It must be established t**hat there exists a sample spa**ce S, a probability function *P* on S, **and a random vari**able X defined on S such that *F* is the cdf of X.

**Example 1.5.4 (Tossing for a head) Suppose we do an experiment that consists of tossing a coin until a head appears.** Let p = probability of a head on any given toss, **and define a random variable X = number of tosses req**uired to get a head. Then, for **any I** = 1,2,...,

*(*1.5.2)

*P(X* = *2*) = (1 - *p)\* -p*,

**since we must g**et 2 – 1 tails followed by a head for the event to occur and all trials **are independen**t. From (1.5.2) we calculate, for any positive integer 2,

(1.5.3)

P(X3 =) = ŚP(X = 6) = 30-.

**The partial sum of the geometric series is**

(1.5.4)

***\*\*-1***

***1- t*n** 1-*7*

*t +1*.

***k=*1**

**a fact that can be establi**shed by induction (see Exercise 1.50). Applying (1.5.4) to **our probabi**lity, we find that the cdf of **the random variable X is**

*F*x(x) *= P(*X<2)

1- (1 - p)\* -1-(1 – p) P =1-(1-p)\*,

=1,2,... .

The cdf *F*x(x) is fl**at between the nonnegative integers, as in Examp**le 1.5.2.

It is easy to show that if I <p<1, then *F*x (2) satisfies the conditions of Theorem 1.5.3. First,

lim 14-0

*Fx*(x) = 0

**PROBABILITY THEORY**

**Section 1.5**

**0**

**1**

**2**

**3**

**4**

**5**

**6**

***7***

**8**

**9**

**10**

**11**

**12**

**13**

**14**

**15**

Figure 1.5.2. *Geometric cdf, p=* .3

since *Fx(*x) = 0 for all x < 0, and

100

lim *F*x(x) = lim 1-(1-2)\* = 1,

1+00 where x goes through only integer values when this limit is taken. To verify property (b), we simply note that the sum in (1.5.3) contains more *positiv*e terms as I in**creases.** Finally, to verify (c), note that, for any z*, Fx(x* + 1) = *Fx*() if € > 0 is sufficiently small. Hence,

lim *F*x(*x* +€) = *Fx(x*),

€10

so *F*x(x) is right-continuous*. Fx(*x) is the cdf of a distribution called th*e geometric distribution* (after the series) and is pictured in Figure 1.5.2.

**Example 1.5.5 (Continuous cdf) An exam**ple of a continuous cdf is the function

(1.5.5)

*F*x (2) =,

Items which satisfies the conditions of Theorem 1.5.3. For example,

lim I- -

*F*x(x) = 0

since

lim

e\* = 0

-

-

**and**

lim *F*x(x)=1 100

since

lim

e \* = 0.

Differentiating *F*x(x) gives

PX *(2)* = (a + 2 y 30,

**Section 1.5**

**DISTRIBUTION FUNCTIONS**

**33**

showing that *F*x() is inc**reasing***. F*x is not only right-continuous, but also continuous. This is a special case of the *logistic distribution.*

**Example 1.*5.6 (*Cdf with jumps) I**f *F*x i*s n*ot a continuous function of x, it is possible for it to be a mixture of continuous pieces and jumps. For example, if we modify *F*x() of (1.5.5) to be, for some €, 1 > € > 0,

***1*-E**

if y < 0

1+*e*

(1.5.6)

*Fy(y*) =

Tat (1-€) if y 20,

*1+e-*

*y*

Yz0,

then *Fy(y*) is the cdf of a **random variab**le Y (see Exercise 1.47). The function *Fy* has a jump of height € at *y* = 0 and otherwise is continuous. This model might be appropriate if we were observing the reading from a gauge, a reading that could (theoretically) be **anywhere between - and o**o. This particular gauge, however, **sometime**s sticks at 0. We could then model our observations with *F*y, where e is the probability that the g**auge sticks.**

Whether a cdf is continuous or has jumps cor**responds to the associated random varia**ble being continuous or not. In fact, the association is such that it is co**nvenient** to define continuo**us random va**riables in this way.

**Definition 1.5.7 A random varia**ble X is *continuous* if *F*x(1) is a continuous function of I. A random variable X i*s discrete* if *F*x() is a step function of I.

We close this section with a theorem formally stating that *Fx c*ompletely deter **min**es the probability distribution of a random variable X. This is true if *P(*X E A) is defined only for events A in *B*?, the smallest sigma algebra containing all the intervals of real numbers of the form (*a, b*), [*a,b), (a*,b], and [a, b]. If probabilities are defined for a larger class of events, it is possible for two random variables to have the same distribution function but not the same probability for every event (see Chung 1974, page *2*7). In this book, as in most statistical applications, we are concerned only with events that are intervals, countable unions or intersections of intervals, etc. S*o* we do not consider such pathological cases. We first need the notion of two random variables being identically distributed.

**Definiti**on 1.5.8 The random variables X and Y are *identically distributed* if, for *e*very set AE*BP, P(*XE A) *= P(Y* E A).

Note that two random variables that are identically distributed are not necessarily equal. That is, Definition 1.5.8 does not say that X = Y.

**Example 1.5.9 (Identically distributed random variables**) Consider the ex periment of tossing a fair coin three times as in Example 1.4.3. Define the r**andom** variables X and Y by

X = number of heads observed and

Y = number of tails observed.