

Wiley Series in Probability and Statistics

# Examples and Problems in Mathematical Statistics

*Shelemiyahu Zacks*

WILEY



## Examples and Problems in Mathematical Statistics

WILEY SERIES IN PROBABILITY AND STATISTICS

Established by WALTER A. SHEWHART and SAMUEL S. WILKS

Editors: *David J. Balding, Noel A. C. Cressie, Garrett M. Fitzmaurice,  
Harvey Goldstein, Iain M. Johnstone, Geert Molenberghs, David W. Scott,  
Adrian F. M. Smith, Ruey S. Tsay, Sanford Weisberg*  
Editors Emeriti: *Vic Barnett, J. Stuart Hunter, Joseph B. Kadane, Jozef L. Teugels*

A complete list of the titles in this series appears at the end of this volume.

# Examples and Problems in Mathematical Statistics

SHELEMYAHU ZACKS

Department of Mathematical Sciences  
Binghamton University  
Binghamton, NY

**WILEY**

Copyright © 2014 by John Wiley & Sons, Inc. All rights reserved.

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

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at [www.copyright.com](http://www.copyright.com). Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permission>.

**Limit of Liability/Disclaimer of Warranty:** While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at [www.wiley.com](http://www.wiley.com).

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

Zacks, Shelemyahu, 1932- author.

Examples and problems in mathematical statistics / Shelemyahu Zacks.  
pages cm

Summary: "This book presents examples that illustrate the theory of mathematical statistics and details how to apply the methods for solving problems" – Provided by publisher.

Includes bibliographical references and index.

ISBN 978-1-118-60550-9 (hardback)

1. Mathematical statistics—Problems, exercises, etc. I. Title.

QC32.Z265 2013

519.5—dc23

2013034492

Printed in the United States of America

ISBN: 9781118605509

10 9 8 7 6 5 4 3 2 1

To my wife Hanna,  
our sons Yuval and David,  
and their families, with love.





# Contents

|   |             |
|---|-------------|
| <b>Preface</b>  | <b>xv</b>   |
| <b>List of Random Variables</b>   | <b>xvii</b> |
| <b>List of Abbreviations</b>  | <b>xix</b>  |
| <b>1 Basic Probability Theory</b>                                       | <b>1</b>    |
| PART I: THEORY, 1   |             |
| 1.1 Operations on Sets, 1   |             |
| 1.2 Algebra and $\sigma$ -Fields, 2                                     |             |
| 1.3 Probability Spaces, 4   |             |
| 1.4 Conditional Probabilities and Independence, 6                       |             |
| 1.5 Random Variables and Their Distributions, 8                         |             |
| 1.6 The Lebesgue and Stieltjes Integrals, 12                            |             |
| 1.6.1 General Definition of Expected Value: The Lebesgue Integral, 12   |             |
| 1.6.2 The Stieltjes–Riemann Integral, 17                                |             |
| 1.6.3 Mixtures of Discrete and Absolutely Continuous Distributions, 19  |             |
| 1.6.4 Quantiles of Distributions, 19                                    |             |
| 1.6.5 Transformations, 20   |             |
| 1.7 Joint Distributions, Conditional Distributions and Independence, 21 |             |
| 1.7.1 Joint Distributions, 21   |             |
| 1.7.2 Conditional Expectations: General Definition, 23                  |             |
| 1.7.3 Independence, 26  |             |
| 1.8 Moments and Related Functionals, 26                                 |             |
| 1.9 Modes of Convergence, 35  |             |
| 1.10 Weak Convergence, 39   |             |
| 1.11 Laws of Large Numbers, 41  |             |

|  |  |    |
|--|--|----|
| 1.11.1                                   | The Weak Law of Large Numbers (WLLN),            | 41 |
| 1.11.2                                   | The Strong Law of Large Numbers (SLLN),          | 42 |
| 1.12                                     | Central Limit Theorem,                           | 44 |
| 1.13                                     | Miscellaneous Results,                           | 47 |
| 1.13.1                                   | Law of the Iterated Logarithm,                   | 48 |
| 1.13.2                                   | Uniform Integrability,                           | 48 |
| 1.13.3                                   | Inequalities,                                    | 52 |
| 1.13.4                                   | The Delta Method,                                | 53 |
| 1.13.5                                   | The Symbols $o_p$ and $O_p$ ,                    | 55 |
| 1.13.6                                   | The Empirical Distribution and Sample Quantiles, | 55 |
| PART II: EXAMPLES,                       |  | 56 |
| PART III: PROBLEMS,                      |  | 73 |
| PART IV: SOLUTIONS TO SELECTED PROBLEMS, |  | 93 |

## 2 Statistical Distributions

106

|                 |   |     |
|-----------------|---|-----|
| PART I: THEORY, |   | 106 |
| 2.1             | Introductory Remarks,                                   | 106 |
| 2.2             | Families of Discrete Distributions,                     | 106 |
| 2.2.1           | Binomial Distributions,                                 | 106 |
| 2.2.2           | Hypergeometric Distributions,                           | 107 |
| 2.2.3           | Poisson Distributions,                                  | 108 |
| 2.2.4           | Geometric, Pascal, and Negative Binomial Distributions, | 108 |
| 2.3             | Some Families of Continuous Distributions,              | 109 |
| 2.3.1           | Rectangular Distributions,                              | 109 |
| 2.3.2           | Beta Distributions,                                     | 111 |
| 2.3.3           | Gamma Distributions,                                    | 111 |
| 2.3.4           | Weibull and Extreme Value Distributions,                | 112 |
| 2.3.5           | Normal Distributions,                                   | 113 |
| 2.3.6           | Normal Approximations,                                  | 114 |
| 2.4             | Transformations,  | 118 |
| 2.4.1           | One-to-One Transformations of Several Variables,        | 118 |
| 2.4.2           | Distribution of Sums,                                   | 118 |
| 2.4.3           | Distribution of Ratios,                                 | 118 |
| 2.5             | Variances and Covariances of Sample Moments,            | 120 |
| 2.6             | Discrete Multivariate Distributions,                    | 122 |
| 2.6.1           | The Multinomial Distribution,                           | 122 |
| 2.6.2           | Multivariate Negative Binomial,                         | 123 |
| 2.6.3           | Multivariate Hypergeometric Distributions,              | 124 |

|  |  |
|--|--|
| 2.7  | Multinormal Distributions, 125   |
| 2.7.1  | Basic Theory, 125  |
| 2.7.2  | Distribution of Subvectors and Distributions of Linear Forms, 127                                |
| 2.7.3  | Independence of Linear Forms, 129  |
| 2.8  | Distributions of Symmetric Quadratic Forms of Normal Variables, 130                              |
| 2.9  | Independence of Linear and Quadratic Forms of Normal Variables, 132                              |
| 2.10   | The Order Statistics, 133  |
| 2.11   | $t$ -Distributions, 135  |
| 2.12   | $F$ -Distributions, 138  |
| 2.13   | The Distribution of the Sample Correlation, 142  |
| 2.14   | Exponential Type Families, 144   |
| 2.15   | Approximating the Distribution of the Sample Mean: Edgeworth and Saddlepoint Approximations, 146 |
| 2.15.1                                       | Edgeworth Expansion, 147   |
| 2.15.2                                       | Saddlepoint Approximation, 149   |
| PART II: EXAMPLES, 150                       |  |
| PART III: PROBLEMS, 167                      |  |
| PART IV: SOLUTIONS TO SELECTED PROBLEMS, 181 |  |

### **3 Sufficient Statistics and the Information in Samples** **191**

|  |   |
|--|---|
| PART I: THEORY, 191                          |   |
| 3.1  | Introduction, 191   |
| 3.2  | Definition and Characterization of Sufficient Statistics, 192 |
| 3.2.1  | Introductory Discussion, 192                                  |
| 3.2.2  | Theoretical Formulation, 194                                  |
| 3.3  | Likelihood Functions and Minimal Sufficient Statistics, 200   |
| 3.4  | Sufficient Statistics and Exponential Type Families, 202      |
| 3.5  | Sufficiency and Completeness, 203                             |
| 3.6  | Sufficiency and Ancillarity, 205                              |
| 3.7  | Information Functions and Sufficiency, 206                    |
| 3.7.1  | The Fisher Information, 206                                   |
| 3.7.2  | The Kullback–Leibler Information, 210                         |
| 3.8  | The Fisher Information Matrix, 212                            |
| 3.9  | Sensitivity to Changes in Parameters, 214                     |
| 3.9.1  | The Hellinger Distance, 214                                   |
| PART II: EXAMPLES, 216                       |   |
| PART III: PROBLEMS, 230                      |   |
| PART IV: SOLUTIONS TO SELECTED PROBLEMS, 236 |   |

## **4 Testing Statistical Hypotheses 246**

### **PART I: THEORY, 246**

- 4.1 The General Framework, 246
- 4.2 The Neyman–Pearson Fundamental Lemma, 248
- 4.3 Testing One-Sided Composite Hypotheses in MLR Models, 251
- 4.4 Testing Two-Sided Hypotheses in One-Parameter Exponential Families, 254
- 4.5 Testing Composite Hypotheses with Nuisance Parameters—Unbiased Tests, 256
- 4.6 Likelihood Ratio Tests, 260
  - 4.6.1 Testing in Normal Regression Theory, 261
  - 4.6.2 Comparison of Normal Means: The Analysis of Variance, 265
- 4.7 The Analysis of Contingency Tables, 271
  - 4.7.1 The Structure of Multi-Way Contingency Tables and the Statistical Model, 271
  - 4.7.2 Testing the Significance of Association, 271
  - 4.7.3 The Analysis of  $2 \times 2$  Tables, 273
  - 4.7.4 Likelihood Ratio Tests for Categorical Data, 274
- 4.8 Sequential Testing of Hypotheses, 275
  - 4.8.1 The Wald Sequential Probability Ratio Test, 276

### **PART II: EXAMPLES, 283**

### **PART III: PROBLEMS, 298**

### **PART IV: SOLUTIONS TO SELECTED PROBLEMS, 307**

## **5 Statistical Estimation 321**

### **PART I: THEORY, 321**

- 5.1 General Discussion, 321
- 5.2 Unbiased Estimators, 322
  - 5.2.1 General Definition and Example, 322
  - 5.2.2 Minimum Variance Unbiased Estimators, 322
  - 5.2.3 The Cramér–Rao Lower Bound for the One-Parameter Case, 323
  - 5.2.4 Extension of the Cramér–Rao Inequality to Multiparameter Cases, 326
  - 5.2.5 General Inequalities of the Cramér–Rao Type, 327
- 5.3 The Efficiency of Unbiased Estimators in Regular Cases, 328
- 5.4 Best Linear Unbiased and Least-Squares Estimators, 331
  - 5.4.1 BLUEs of the Mean, 331
  - 5.4.2 Least-Squares and BLUEs in Linear Models, 332
  - 5.4.3 Best Linear Combinations of Order Statistics, 334

|          |  |            |
|----------|--|------------|
| 5.5      | Stabilizing the LSE: Ridge Regressions,  | 335        |
| 5.6      | Maximum Likelihood Estimators,   | 337        |
| 5.6.1    | Definition and Examples,   | 337        |
| 5.6.2    | MLEs in Exponential Type Families,   | 338        |
| 5.6.3    | The Invariance Principle,  | 338        |
| 5.6.4    | MLE of the Parameters of Tolerance Distributions,                                  | 339        |
| 5.7      | Equivariant Estimators,  | 341        |
| 5.7.1    | The Structure of Equivariant Estimators,   | 341        |
| 5.7.2    | Minimum MSE Equivariant Estimators,  | 343        |
| 5.7.3    | Minimum Risk Equivariant Estimators,   | 343        |
| 5.7.4    | The Pitman Estimators,   | 344        |
| 5.8      | Estimating Equations,  | 346        |
| 5.8.1    | Moment-Equations Estimators,   | 346        |
| 5.8.2    | General Theory of Estimating Functions,  | 347        |
| 5.9      | Pretest Estimators,  | 349        |
| 5.10     | Robust Estimation of the Location and Scale Parameters of Symmetric Distributions, | 349        |
|          | PART II: EXAMPLES,   | 353        |
|          | PART III: PROBLEMS,  | 381        |
|          | PART IV: SOLUTIONS OF SELECTED PROBLEMS,   | 393        |
| <b>6</b> | <b>Confidence and Tolerance Intervals</b>  | <b>406</b> |
|          | PART I: THEORY,  | 406        |
| 6.1      | General Introduction,  | 406        |
| 6.2      | The Construction of Confidence Intervals,  | 407        |
| 6.3      | Optimal Confidence Intervals,  | 408        |
| 6.4      | Tolerance Intervals,   | 410        |
| 6.5      | Distribution Free Confidence and Tolerance Intervals,                              | 412        |
| 6.6      | Simultaneous Confidence Intervals,   | 414        |
| 6.7      | Two-Stage and Sequential Sampling for Fixed Width Confidence Intervals,            | 417        |
|          | PART II: EXAMPLES,   | 421        |
|          | PART III: PROBLEMS,  | 429        |
|          | PART IV: SOLUTION TO SELECTED PROBLEMS,  | 433        |
| <b>7</b> | <b>Large Sample Theory for Estimation and Testing</b>                              | <b>439</b> |
|          | PART I: THEORY,  | 439        |
| 7.1      | Consistency of Estimators and Tests,   | 439        |
| 7.2      | Consistency of the MLE,  | 440        |

|     |  |     |
|-----|--|-----|
| 7.3 | Asymptotic Normality and Efficiency of Consistent Estimators,  | 442 |
| 7.4 | Second-Order Efficiency of BAN Estimators,   | 444 |
| 7.5 | Large Sample Confidence Intervals,   | 445 |
| 7.6 | Edgeworth and Saddlepoint Approximations to the Distribution of the MLE: One-Parameter Canonical Exponential Families, | 446 |
| 7.7 | Large Sample Tests,  | 448 |
| 7.8 | Pitman's Asymptotic Efficiency of Tests,   | 449 |
| 7.9 | Asymptotic Properties of Sample Quantiles,   | 451 |
|     | PART II: EXAMPLES,   | 454 |
|     | PART III: PROBLEMS,  | 475 |
|     | PART IV: SOLUTION OF SELECTED PROBLEMS,  | 479 |

## **8 Bayesian Analysis in Testing and Estimation 485**

|       |  |     |
|-------|--|-----|
|       | PART I: THEORY,                                  | 485 |
| 8.1   | The Bayesian Framework,                          | 486 |
| 8.1.1 | Prior, Posterior, and Predictive Distributions,  | 486 |
| 8.1.2 | Noninformative and Improper Prior Distributions, | 487 |
| 8.1.3 | Risk Functions and Bayes Procedures,             | 489 |
| 8.2   | Bayesian Testing of Hypothesis,                  | 491 |
| 8.2.1 | Testing Simple Hypothesis,                       | 491 |
| 8.2.2 | Testing Composite Hypotheses,                    | 493 |
| 8.2.3 | Bayes Sequential Testing of Hypotheses,          | 495 |
| 8.3   | Bayesian Credibility and Prediction Intervals,   | 501 |
| 8.3.1 | Credibility Intervals,                           | 501 |
| 8.3.2 | Prediction Intervals,                            | 501 |
| 8.4   | Bayesian Estimation,                             | 502 |
| 8.4.1 | General Discussion and Examples,                 | 502 |
| 8.4.2 | Hierarchical Models,                             | 502 |
| 8.4.3 | The Normal Dynamic Linear Model,                 | 504 |
| 8.5   | Approximation Methods,                           | 506 |
| 8.5.1 | Analytical Approximations,                       | 506 |
| 8.5.2 | Numerical Approximations,                        | 508 |
| 8.6   | Empirical Bayes Estimators,                      | 513 |
|       | PART II: EXAMPLES,                               | 514 |
|       | PART III: PROBLEMS,                              | 549 |
|       | PART IV: SOLUTIONS OF SELECTED PROBLEMS,         | 557 |

## **9 Advanced Topics in Estimation Theory 563**

|     |                     |     |
|-----|---------------------|-----|
|     | PART I: THEORY,     | 563 |
| 9.1 | Minimax Estimators, | 563 |

|  |   |            |
|--|---|------------|
| 9.2  | Minimum Risk Equivariant, Bayes Equivariant, and Structural Estimators, 565                                     |            |
| 9.2.1  | Formal Bayes Estimators for Invariant Priors, 566   |            |
| 9.2.2  | Equivariant Estimators Based on Structural Distributions, 568   |            |
| 9.3  | The Admissibility of Estimators, 570  |            |
| 9.3.1  | Some Basic Results, 570   |            |
| 9.3.2  | The Inadmissibility of Some Commonly Used Estimators, 575   |            |
| 9.3.3  | Minimax and Admissible Estimators of the Location Parameter, 582  |            |
| 9.3.4  | The Relationship of Empirical Bayes and Stein-Type Estimators of the Location Parameter in the Normal Case, 584 |            |
| PART II: EXAMPLES, 585                       |   |            |
| PART III: PROBLEMS, 592                      |   |            |
| PART IV: SOLUTIONS OF SELECTED PROBLEMS, 596 |   |            |
| <b>References</b>                            |   | <b>601</b> |
| <b>Author Index</b>                          |   | <b>613</b> |
| <b>Subject Index</b>                         |   | <b>617</b> |





# Preface

I have been teaching probability and mathematical statistics to graduate students for close to 50 years. In my career I realized that the most difficult task for students is solving problems. Bright students can generally grasp the theory easier than apply it. In order to overcome this hurdle, I used to write examples of solutions to problems and hand it to my students. I often wrote examples for the students based on my published research. Over the years I have accumulated a large number of such examples and problems. This book is aimed at sharing these examples and problems with the population of students, researchers, and teachers.

The book consists of nine chapters. Each chapter has four parts. The first part contains a short presentation of the theory. This is required especially for establishing notation and to provide a quick overview of the important results and references. The second part consists of examples. The examples follow the theoretical presentation. The third part consists of problems for solution, arranged by the corresponding sections of the theory part. The fourth part presents solutions to some selected problems. The solutions are generally not as detailed as the examples, but as such these are examples of solutions. I tried to demonstrate how to apply known results in order to solve problems elegantly. All together there are in the book 167 examples and 431 problems.

The emphasis in the book is on statistical inference. The first chapter on probability is especially important for students who have not had a course on advanced probability. Chapter Two is on the theory of distribution functions. This is basic to all developments in the book, and from my experience, it is important for all students to master this calculus of distributions. The chapter covers multivariate distributions, especially the multivariate normal; conditional distributions; techniques of determining variances and covariances of sample moments; the theory of exponential families; Edgeworth expansions and saddle-point approximations; and more. Chapter Three covers the theory of sufficient statistics, completeness of families of distributions, and the information in samples. In particular, it presents the Fisher information, the Kullback–Leibler information, and the Hellinger distance. Chapter Four provides a strong foundation in the theory of testing statistical hypotheses. The Wald SPRT is

discussed there too. Chapter Five is focused on optimal point estimation of different kinds. Pitman estimators and equivariant estimators are also discussed. Chapter Six covers problems of efficient confidence intervals, in particular the problem of determining fixed-width confidence intervals by two-stage or sequential sampling. Chapter Seven covers techniques of large sample approximations, useful in estimation and testing. Chapter Eight is devoted to Bayesian analysis, including empirical Bayes theory. It highlights computational approximations by numerical analysis and simulations. Finally, Chapter Nine presents a few more advanced topics, such as minimaxity, admissibility, structural distributions, and the Stein-type estimators.

I would like to acknowledge with gratitude the contributions of my many ex-students, who toiled through these examples and problems and gave me their important feedback. In particular, I am very grateful and indebted to my colleagues, Professors A. Schick, Q. Yu, S. De, and A. Polunchenko, who carefully read parts of this book and provided important comments. Mrs. Marge Pratt skillfully typed several drafts of this book with patience and grace. To her I extend my heartfelt thanks. Finally, I would like to thank my wife Hanna for giving me the conditions and encouragement to do research and engage in scholarly writing.

SHELEMYAHU ZACKS

# List of Random Variables

|                        |  |
|------------------------|--|
| $B(n, p)$              | Binomial, with parameters $n$ and $p$                        |
| $E(\mu)$               | Exponential with parameter $\mu$                             |
| $EV(\lambda, \alpha)$  | Extreme value with parameters $\lambda$ and $\alpha$         |
| $F(v_1, v_2)$          | Central $F$ with parameters $v_1$ and $v_2$                  |
| $F(n_1, n_2; \lambda)$ | Noncentral $F$ with parameters $v_1, v_2, \lambda$           |
| $G(\lambda, p)$        | Gamma with parameters $\lambda$ and $p$                      |
| $H(M, N, n)$           | Hyper-geometric with parameters $M, N, n$                    |
| $N(\mu, V)$            | Multinormal with mean vector $\mu$ and covariance matrix $V$ |
| $N(\mu, \sigma)$       | Normal with mean $\mu$ and $\sigma$                          |
| $NB(\psi, v)$          | Negative-binomial with parameters $\psi$ , and $v$           |
| $P(\lambda)$           | Poisson with parameter $\lambda$                             |
| $R(a, b)$              | Rectangular (uniform) with parameters $a$ and $b$            |
| $t[n; \lambda]$        | Noncentral Student's $t$ with parameters $n$ and $\lambda$   |
| $t[n; \xi, V]$         | Multivariate $t$ with parameters $n, \xi$ and $V$            |
| $t[n]$                 | Student's $t$ with $n$ degrees of freedom                    |
| $W(\lambda, \alpha)$   | Weibul with parameters $\lambda$ and $\alpha$                |
| $\beta(p, q)$          | Beta with parameters $p$ and $q$                             |
| $\chi^2[n, \lambda]$   | Noncentral chi-squared with parameters $n$ and $\lambda$     |
| $\chi^2[n]$            | Chi-squared with $n$ degrees of freedom                      |



# List of Abbreviations

|                    |   |
|--------------------|---|
| a.s.               | Almost surely                                 |
| ANOVA              | Analysis of variance                          |
| c.d.f.             | Cumulative distribution function              |
| $\text{cov}(x, y)$ | Covariance of $X$ and $Y$                     |
| CI                 | Confidence interval                           |
| CLT                | Central limit theorem                         |
| CP                 | Coverage probability                          |
| CR                 | Cramer Rao regularity conditions              |
| $E\{X   Y\}$       | Conditional expected value of $X$ , given $Y$ |
| $E\{X\}$           | Expected value of $X$                         |
| FIM                | Fisher information matrix                     |
| i.i.d.             | Independent identically distributed           |
| LBUE               | Linear best unbiased estimate                 |
| LCL                | Lower confidence limit                        |
| m.g.f.             | Moment generating function                    |
| m.s.s.             | Minimal sufficient statistics                 |
| MEE                | Moments equations estimator                   |
| MLE                | Maximum likelihood estimator                  |
| MLR                | Monotone likelihood ratio                     |
| MP                 | Most powerful                                 |
| MSE                | Mean squared error                            |
| MVU                | Minimum variance unbiased                     |
| OC                 | Operating characteristic                      |
| p.d.f.             | Probability density function                  |
| p.g.f.             | Probability generating function               |
| $P\{E   A\}$       | Conditional probability of $E$ , given $A$    |
| $P\{E\}$           | Probability of $E$                            |
| PTE                | Pre-test estimator                            |
| r.v.               | Random variable                               |
| RHS                | Right-hand side                               |
| s.v.               | Stopping variable                             |

|                  |   |
|------------------|---|
| SE               | Standard error                          |
| SLLN             | Strong law of large numbers             |
| SPRT             | Sequential probability ratio test       |
| $\text{tr}\{A\}$ | trace of the matrix $A$                 |
| UCL              | Upper control limit                     |
| UMP              | Uniformly most powerful                 |
| UMPI             | Uniformly most powerful invariant       |
| UMPU             | Uniformly most powerful unbiased        |
| UMVU             | Uniformly minimum variance unbiased     |
| $V\{X \mid Y\}$  | Conditional variance of $X$ , given $Y$ |
| $V\{X\}$         | Variance of $X$                         |
| w.r.t.           | With respect to                         |
| WLLN             | Weak law of large numbers               |

## CHAPTER 1

# Basic Probability Theory

### PART I: THEORY

It is assumed that the reader has had a course in elementary probability. In this chapter we discuss more advanced material, which is required for further developments.

#### 1.1 OPERATIONS ON SETS

Let  $S$  denote a **sample space**. Let  $E_1, E_2$  be subsets of  $S$ . We denote the **union** by  $E_1 \cup E_2$  and the intersection by  $E_1 \cap E_2$ .  $\bar{E} = S - E$  denotes the **complement** of  $E$ . By DeMorgan's laws  $\overline{E_1 \cup E_2} = \bar{E}_1 \cap \bar{E}_2$  and  $\overline{E_1 \cap E_2} = \bar{E}_1 \cup \bar{E}_2$ .

Given a sequence of sets  $\{E_n, n \geq 1\}$  (finite or infinite), we define

$$\sup_{n \geq 1} E_n = \bigcup_{n \geq 1} E_n, \quad \inf_{n \geq 1} E_n = \bigcap_{n \geq 1} E_n. \quad (1.1.1)$$

Furthermore,  $\liminf_{n \rightarrow \infty}$  and  $\limsup_{n \rightarrow \infty}$  are defined as

$$\liminf_{n \rightarrow \infty} E_n = \bigcup_{n \geq 1} \bigcap_{k \geq n} E_k, \quad \limsup_{n \rightarrow \infty} E_n = \bigcap_{n \geq 1} \bigcup_{k \geq n} E_k. \quad (1.1.2)$$

If a point of  $S$  belongs to  $\limsup_{n \rightarrow \infty} E_n$ , it belongs to infinitely many sets  $E_n$ . The sets  $\liminf_{n \rightarrow \infty} E_n$  and  $\limsup_{n \rightarrow \infty} E_n$  always exist and

$$\liminf_{n \rightarrow \infty} E_n \subset \limsup_{n \rightarrow \infty} E_n. \quad (1.1.3)$$

If  $\liminf_{n \rightarrow \infty} E_n = \limsup_{n \rightarrow \infty} E_n$ , we say that a limit of  $\{E_n, n \geq 1\}$  exists. In this case,

$$\lim_{n \rightarrow \infty} E_n = \liminf_{n \rightarrow \infty} E_n = \limsup_{n \rightarrow \infty} E_n. \quad (1.1.4)$$

A sequence  $\{E_n, n \geq 1\}$  is called **monotone increasing** if  $E_n \subset E_{n+1}$  for all  $n \geq 1$ . In this case  $\lim_{n \rightarrow \infty} E_n = \bigcup_{n=1}^{\infty} E_n$ . The sequence is **monotone decreasing** if  $E_n \supset E_{n+1}$ , for

all  $n \geq 1$ . In this case  $\lim_{n \rightarrow \infty} E_n = \bigcap_{n=1}^{\infty} E_n$ . We conclude this section with the definition

of a **partition** of the sample space. A collection of sets  $\mathcal{D} = \{E_1, \dots, E_k\}$  is called a finite **partition** of  $\mathcal{S}$  if all elements of  $\mathcal{D}$  are **pairwise disjoint** and their union is  $\mathcal{S}$ , i.e.,  $E_i \cap E_j = \emptyset$  for all  $i \neq j$ ;  $E_i, E_j \in \mathcal{D}$ ; and  $\bigcup_{i=1}^k E_i = \mathcal{S}$ . If  $\mathcal{D}$  contains a countable number of sets that are mutually exclusive and  $\bigcup_{i=1}^{\infty} E_i = \mathcal{S}$ , we say that  $\mathcal{D}$  is a countable partition.

## 1.2 ALGEBRA AND $\sigma$ -FIELDS

Let  $\mathcal{S}$  be a sample space. An algebra  $\mathcal{A}$  is a collection of subsets of  $\mathcal{S}$  satisfying

- (i)  $\mathcal{S} \in \mathcal{A}$ ;
  - (ii) if  $E \in \mathcal{A}$  then  $\bar{E} \in \mathcal{A}$ ;
  - (iii) if  $E_1, E_2 \in \mathcal{A}$  then  $E_1 \cup E_2 \in \mathcal{A}$ .
- (1.2.1)

We consider  $\emptyset = \bar{\mathcal{S}}$ . Thus, (i) and (ii) imply that  $\emptyset \in \mathcal{A}$ . Also, if  $E_1, E_2 \in \mathcal{A}$  then  $E_1 \cap E_2 \in \mathcal{A}$ .

The **trivial algebra** is  $\mathcal{A}_0 = \{\emptyset, \mathcal{S}\}$ . An algebra  $\mathcal{A}_1$  is a subalgebra of  $\mathcal{A}_2$  if all sets of  $\mathcal{A}_1$  are contained in  $\mathcal{A}_2$ . We denote this inclusion by  $\mathcal{A}_1 \subset \mathcal{A}_2$ . Thus, the trivial algebra  $\mathcal{A}_0$  is a subalgebra of every algebra  $\mathcal{A}$ . We will denote by  $\mathcal{A}(\mathcal{S})$ , the algebra generated by all subsets of  $\mathcal{S}$  (see Example 1.1).

If a sample space  $\mathcal{S}$  has a finite number of points  $n$ , say  $1 \leq n < \infty$ , then the collection of all subsets of  $\mathcal{S}$  is called the **discrete algebra** generated by the elementary events of  $\mathcal{S}$ . It contains  $2^n$  events.

Let  $\mathcal{D}$  be a partition of  $\mathcal{S}$  having  $k$ ,  $2 \leq k$ , disjoint sets. Then, the algebra generated by  $\mathcal{D}$ ,  $\mathcal{A}(\mathcal{D})$ , is the algebra containing all the  $2^k - 1$  unions of the elements of  $\mathcal{D}$  and the empty set.



An algebra on  $\mathcal{S}$  is called a  $\sigma$ -**field** if, in addition to being an algebra, the following holds.

(iv) If  $E_n \in \mathcal{A}$ ,  $n \geq 1$ , then  $\bigcup_{n=1}^{\infty} E_n \in \mathcal{A}$ .

We will denote a  $\sigma$ -field by  $\mathcal{F}$ . In a  $\sigma$ -field  $\mathcal{F}$  the supremum, infimum, limsup, and liminf of any sequence of events belong to  $\mathcal{F}$ . If  $\mathcal{S}$  is finite, the discrete algebra  $\mathcal{A}(\mathcal{S})$  is a  $\sigma$ -field. In Example 1.3 we show an algebra that is not a  $\sigma$ -field.

The minimal  $\sigma$ -field containing the algebra generated by  $\{(-\infty, x], -\infty < x < \infty\}$  is called the **Borel  $\sigma$ -field** on the real line  $\mathbb{R}$ .

A sample space  $\mathcal{S}$ , with a  $\sigma$ -field  $\mathcal{F}$ ,  $(\mathcal{S}, \mathcal{F})$  is called a **measurable space**.

The following lemmas establish the existence of smallest  $\sigma$ -field containing a given collection of sets.

**Lemma 1.2.1.** *Let  $\mathcal{E}$  be a collection of subsets of a sample space  $\mathcal{S}$ . Then, there exists a smallest  $\sigma$ -field  $\mathcal{F}(\mathcal{E})$ , containing the elements of  $\mathcal{E}$ .*

*Proof.* The algebra of all subsets of  $\mathcal{S}$ ,  $\mathcal{A}(\mathcal{S})$  obviously contains all elements of  $\mathcal{E}$ . Similarly, the  $\sigma$ -field  $\mathcal{F}$  containing all subsets of  $\mathcal{S}$ , contains all elements of  $\mathcal{E}$ . Define the  $\sigma$ -field  $\mathcal{F}(\mathcal{E})$  to be the **intersection** of all  $\sigma$ -fields, which contain all elements of  $\mathcal{E}$ . Obviously,  $\mathcal{F}(\mathcal{E})$  is an algebra. QED

A collection  $\mathcal{M}$  of subsets of  $\mathcal{S}$  is called a **monotonic class** if the limit of any monotone sequence in  $\mathcal{M}$  belongs to  $\mathcal{M}$ .

If  $\mathcal{E}$  is a collection of subsets of  $\mathcal{S}$ , let  $\mathcal{M}^*(\mathcal{E})$  denote the smallest monotonic class containing  $\mathcal{E}$ .

**Lemma 1.2.2.** *A necessary and sufficient condition of an algebra  $\mathcal{A}$  to be a  $\sigma$ -field is that it is a monotonic class.*

*Proof.* (i) Obviously, if  $\mathcal{A}$  is a  $\sigma$ -field, it is a monotonic class.

(ii) Let  $\mathcal{A}$  be a monotonic class.

Let  $E_n \in \mathcal{A}$ ,  $n \geq 1$ . Define  $B_n = \bigcup_{i=1}^n E_i$ . Obviously  $B_n \subset B_{n+1}$  for all  $n \geq 1$ . Hence  $\lim_{n \rightarrow \infty} B_n = \bigcup_{n=1}^{\infty} B_n \in \mathcal{A}$ . But  $\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} E_n$ . Thus,  $\sup_{n \geq 1} E_n \in \mathcal{A}$ . Similarly,  $\bigcap_{n=1}^{\infty} E_n \in \mathcal{A}$ . Thus,  $\mathcal{A}$  is a  $\sigma$ -field. QED

**Theorem 1.2.1.** *Let  $\mathcal{A}$  be an algebra. Then  $\mathcal{M}^*(\mathcal{A}) = \mathcal{F}(\mathcal{A})$ , where  $\mathcal{F}(\mathcal{A})$  is the smallest  $\sigma$ -field containing  $\mathcal{A}$ .*

*Proof.* See Shiriyayev (1984, p. 139).

The measurable space  $(\mathbb{R}, \mathcal{B})$ , where  $\mathbb{R}$  is the real line and  $\mathcal{B} = \mathcal{F}(\mathbb{R})$ , called the **Borel measurable space**, plays a most important role in the theory of statistics. Another important measurable space is  $(\mathbb{R}^n, \mathcal{B}^n)$ ,  $n \geq 2$ , where  $\mathbb{R}^n = \mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}$  is the Euclidean  $n$ -space, and  $\mathcal{B}^n = \mathcal{B} \times \cdots \times \mathcal{B}$  is the smallest  $\sigma$ -field containing  $\mathbb{R}^n$ ,  $\emptyset$ , and all  $n$ -dimensional rectangles  $I = I_1 \times \cdots \times I_n$ , where

$$I_i = (a_i, b_i], \quad i = 1, \dots, n, \quad -\infty < a_i < b_i < \infty.$$

The measurable space  $(\mathbb{R}^\infty, \mathcal{B}^\infty)$  is used as a basis for probability models of experiments with infinitely many trials.  $\mathbb{R}^\infty$  is the space of ordered sequences  $\mathbf{x} = (x_1, x_2, \dots)$ ,  $-\infty < x_n < \infty$ ,  $n = 1, 2, \dots$ . Consider the cylinder sets

$$\mathcal{C}(I_1 \times \cdots \times I_n) = \{\mathbf{x} : x_i \in I_i, i = 1, \dots, n\}$$

and

$$\mathcal{C}(B_1 \times \cdots \times B_n) = \{\mathbf{x} : x_i \in B_i, i = 1, \dots, n\}$$

where  $B_i$  are Borel sets, i.e.,  $B_i \in \mathcal{B}$ . The smallest  $\sigma$ -field containing all these cylinder sets,  $n \geq 1$ , is  $\mathcal{B}(\mathbb{R}^\infty)$ . Examples of Borel sets in  $\mathcal{B}(\mathbb{R}^\infty)$  are

$$(a) \quad \{\mathbf{x} : \mathbf{x} \in \mathbb{R}^\infty, \sup_{n \geq 1} x_n > a\}$$

or

$$(b) \quad \{\mathbf{x} : \mathbf{x} \in \mathbb{R}^\infty, \limsup_{n \rightarrow \infty} x_n \leq a\}.$$

### 1.3 PROBABILITY SPACES

Given a measurable space  $(\mathcal{S}, \mathcal{F})$ , a **probability model** ascribes a countably additive function  $P$  on  $\mathcal{F}$ , which assigns a probability  $P\{A\}$  to all sets  $A \in \mathcal{F}$ . This function should satisfy the following properties.

$$(A.1) \quad \text{If } A \in \mathcal{F} \text{ then } 0 \leq P\{A\} \leq 1.$$

$$(A.2) \quad P\{\mathcal{S}\} = 1. \tag{1.3.1}$$

$$(A.3) \quad \text{If } \{E_n, n \geq 1\} \in \mathcal{F} \text{ is a sequence of disjoint}$$

$$\text{sets in } \mathcal{F}, \text{ then } P\left\{\bigcup_{n=1}^{\infty} E_n\right\} = \sum_{n=1}^{\infty} P\{E_n\}. \tag{1.3.2}$$

Recall that if  $A \subset B$  then  $P\{A\} \leq P\{B\}$ , and  $P\{\bar{A}\} = 1 - P\{A\}$ . Other properties will be given in the examples and problems. In the sequel we often write  $AB$  for  $A \cap B$ .

**Theorem 1.3.1.** *Let  $(\mathcal{S}, \mathcal{F}, P)$  be a probability space, where  $\mathcal{F}$  is a  $\sigma$ -field of subsets of  $\mathcal{S}$  and  $P$  a probability function. Then*

(i) *if  $B_n \subset B_{n+1}$ ,  $n \geq 1$ ,  $B_n \in \mathcal{F}$ , then*

$$P\left\{\lim_{n \rightarrow \infty} B_n\right\} = \lim_{n \rightarrow \infty} P\{B_n\}. \quad (1.3.3)$$

(ii) *if  $B_n \supset B_{n+1}$ ,  $n \geq 1$ ,  $B_n \in \mathcal{F}$ , then*

$$P\left\{\lim_{n \rightarrow \infty} B_n\right\} = \lim_{n \rightarrow \infty} P\{B_n\}. \quad (1.3.4)$$

*Proof.* (i) Since  $B_n \subset B_{n+1}$ ,  $\lim_{n \rightarrow \infty} B_n = \bigcup_{n=1}^{\infty} B_n$ . Moreover,

$$P\left\{\bigcup_{n=1}^{\infty} B_n\right\} = P\{B_1\} + \sum_{n=2}^{\infty} P\{B_n - B_{n-1}\}. \quad (1.3.5)$$

Notice that for  $n \geq 2$ , since  $\bar{B}_n B_{n-1} = \emptyset$ ,

$$\begin{aligned} P\{B_n - B_{n-1}\} &= P\{B_n \bar{B}_{n-1}\} \\ &= P\{B_n\} - P\{B_n B_{n-1}\} = P\{B_n\} - P\{B_{n-1}\}. \end{aligned} \quad (1.3.6)$$

Also, in (1.3.5)

$$\begin{aligned} P\{B_1\} + \sum_{n=2}^{\infty} P\{B_n - B_{n-1}\} &= \lim_{N \rightarrow \infty} \left( P\{B_1\} + \sum_{n=2}^N (P\{B_n\} - P\{B_{n-1}\}) \right) \\ &= \lim_{N \rightarrow \infty} P\{B_N\}. \end{aligned} \quad (1.3.7)$$

Thus, Equation (1.3.3) is proven.

(ii) Since  $B_n \supset B_{n+1}$ ,  $n \geq 1$ ,  $\bar{B}_n \subset \bar{B}_{n+1}$ ,  $n \geq 1$ .  $\lim_{n \rightarrow \infty} B_n = \bigcap_{n=1}^{\infty} B_n$ . Hence,

$$\begin{aligned} P\left(\lim_{n \rightarrow \infty} B_n\right) &= 1 - P\left\{\overline{\bigcap_{n=1}^{\infty} B_n}\right\} \\ &= 1 - P\left\{\bigcup_{n=1}^{\infty} \bar{B}_n\right\} \\ &= 1 - \lim_{n \rightarrow \infty} P\{\bar{B}_n\} = \lim_{n \rightarrow \infty} P\{B_n\}. \end{aligned}$$

QED

Sets in a probability space are called events.

## 1.4 CONDITIONAL PROBABILITIES AND INDEPENDENCE

The conditional probability of an event  $A \in \mathcal{F}$  given an event  $B \in \mathcal{F}$  such that  $P\{B\} > 0$ , is defined as

$$P\{A \mid B\} = \frac{P\{A \cap B\}}{P\{B\}}. \quad (1.4.1)$$

We see first that  $P\{\cdot \mid B\}$  is a probability function on  $\mathcal{F}$ . Indeed, for every  $A \in \mathcal{F}$ ,  $0 \leq P\{A \mid B\} \leq 1$ . Moreover,  $P\{\mathcal{S} \mid B\} = 1$  and if  $A_1$  and  $A_2$  are disjoint events in  $\mathcal{F}$ , then

$$\begin{aligned} P\{A_1 \cup A_2 \mid B\} &= \frac{P\{(A_1 \cup A_2)B\}}{P\{B\}} \\ &= \frac{P\{A_1 B\} + P\{A_2 B\}}{P\{B\}} = P\{A_1 \mid B\} + P\{A_2 \mid B\}. \end{aligned} \quad (1.4.2)$$

If  $P\{B\} > 0$  and  $P\{A\} \neq P\{A \mid B\}$ , we say that the events  $A$  and  $B$  are **dependent**. On the other hand, if  $P\{A\} = P\{A \mid B\}$  we say that  $A$  and  $B$  are **independent** events. Notice that two events are independent if and only if

$$P\{AB\} = P\{A\}P\{B\}. \quad (1.4.3)$$

Given  $n$  events in  $\mathcal{A}$ , namely  $A_1, \dots, A_n$ , we say that they are **pairwise** independent if  $P\{A_i A_j\} = P\{A_i\}P\{A_j\}$  for any  $i \neq j$ . The events are said to be independent in triplets if

$$P\{A_i A_j A_k\} = P\{A_i\}P\{A_j\}P\{A_k\}$$

for any  $i \neq j \neq k$ . Example 1.4 shows that pairwise independence does not imply independence in triplets.

Given  $n$  events  $A_1, \dots, A_n$  of  $\mathcal{F}$ , we say that they are **independent** if, for any  $2 \leq k \leq n$  and any  $k$ -tuple  $(1 \leq i_1 < i_2 < \dots < i_k \leq n)$ ,

$$P\left\{\bigcap_{j=1}^k A_{i_j}\right\} = \prod_{j=1}^k P\{A_{i_j}\}. \quad (1.4.4)$$

Events in an infinite sequence  $\{A_1, A_2, \dots\}$  are said to be **independent** if  $\{A_1, \dots, A_n\}$  are independent, for each  $n \geq 2$ . Given a sequence of events  $A_1, A_2, \dots$  of a  $\sigma$ -field  $\mathcal{F}$ , we have seen that

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

This event means that points  $\omega$  in  $\limsup_{n \rightarrow \infty} A_n$  belong to infinitely many of the events  $\{A_n\}$ . Thus, the event  $\limsup_{n \rightarrow \infty} A_n$  is denoted also as  $\{A_n, \text{i.o.}\}$ , where i.o. stands for “infinitely often.”

The following important theorem, known as the **Borel–Cantelli Lemma**, gives conditions under which  $P\{A_n, \text{i.o.}\}$  is either 0 or 1.

**Theorem 1.4.1 (Borel–Cantelli).** *Let  $\{A_n\}$  be a sequence of sets in  $\mathcal{F}$ .*

- (i) *If  $\sum_{n=1}^{\infty} P\{A_n\} < \infty$ , then  $P\{A_n, \text{i.o.}\} = 0$ .*
- (ii) *If  $\sum_{n=1}^{\infty} P\{A_n\} = \infty$  and  $\{A_n\}$  are independent, then  $P\{A_n, \text{i.o.}\} = 1$ .*

*Proof.* (i) Notice that  $B_n = \bigcup_{k=n}^{\infty} A_k$  is a decreasing sequence. Thus

$$P\{A_n, \text{i.o.}\} = P\left\{\bigcap_{n=1}^{\infty} B_n\right\} = \lim_{n \rightarrow \infty} P\{B_n\}.$$

But

$$P\{B_n\} = P\left\{\bigcup_{k=n}^{\infty} A_k\right\} \leq \sum_{k=n}^{\infty} P\{A_k\}.$$

The assumption that  $\sum_{n=1}^{\infty} P\{A_n\} < \infty$  implies that  $\lim_{n \rightarrow \infty} \sum_{k=n}^{\infty} P\{A_k\} = 0$ .

(ii) Since  $A_1, A_2, \dots$  are independent,  $\bar{A}_1, \bar{A}_2, \dots$  are independent. This implies that

$$P\left\{\bigcap_{k=1}^{\infty} \bar{A}_k\right\} = \prod_{k=1}^{\infty} P\{\bar{A}_k\} = \prod_{k=1}^{\infty} (1 - P\{A_k\}).$$

If  $0 < x \leq 1$  then  $\log(1 - x) \leq -x$ . Thus,

$$\begin{aligned} \log \prod_{k=1}^{\infty} (1 - P\{A_k\}) &= \sum_{k=1}^{\infty} \log(1 - P\{A_k\}) \\ &\leq -\sum_{k=1}^{\infty} P\{A_k\} = -\infty \end{aligned}$$

since  $\sum_{n=1}^{\infty} P\{A_n\} = \infty$ . Thus  $P\left\{\bigcap_{k=1}^{\infty} \bar{A}_k\right\} = 0$  for all  $n \geq 1$ . This implies that  $P\{A_n, \text{i.o.}\} = 1$ . QED

We conclude this section with the celebrated **Bayes Theorem**.

Let  $\mathcal{D} = \{B_i, i \in J\}$  be a partition of  $\mathcal{S}$ , where  $J$  is an index set having a finite or countable number of elements. Let  $B_j \in \mathcal{F}$  and  $P\{B_j\} > 0$  for all  $j \in J$ . Let  $A \in \mathcal{F}$ ,  $P\{A\} > 0$ . We are interested in the conditional probabilities  $P\{B_j \mid A\}$ ,  $j \in J$ .

**Theorem 1.4.2 (Bayes).**

$$P\{B_j \mid A\} = \frac{P\{B_j\}P\{A \mid B_j\}}{\sum_{j' \in J} P\{B_{j'}\}P\{A \mid B_{j'}\}}. \quad (1.4.5)$$

*Proof.* Left as an exercise. QED

Bayes Theorem is widely used in scientific inference. Examples of the application of Bayes Theorem are given in many elementary books. Advanced examples of Bayesian inference will be given in later chapters.

## 1.5 RANDOM VARIABLES AND THEIR DISTRIBUTIONS

Random variables are finite real value functions on the sample space  $\mathcal{S}$ , such that measurable subsets of  $\mathcal{F}$  are mapped into Borel sets on the real line and thus can be