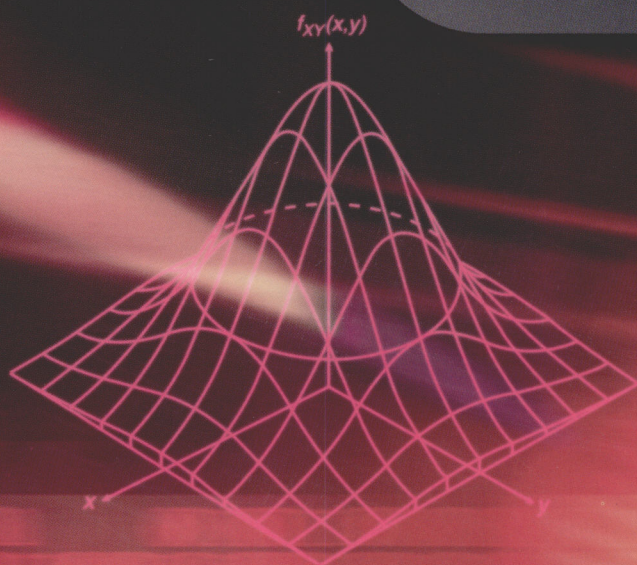
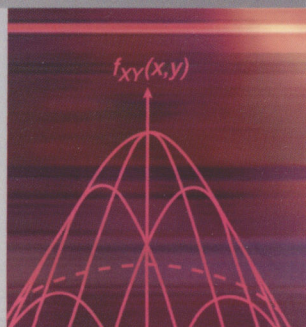


FUNDAMENTALS OF PROBABILITY AND STATISTICS FOR ENGINEERS



T. T. SOONG

 WILEY

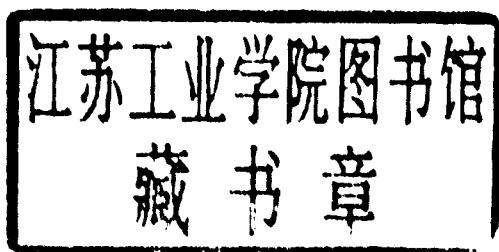


TB114
9

FUNDAMENTALS OF PROBABILITY AND STATISTICS FOR ENGINEERS

T.T. Soong

State University of New York at Buffalo, Buffalo, New York, USA



John Wiley & Sons, Ltd

Copyright © 2004 John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester,
West Sussex PO19 8SQ, England

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page www.wileyeurope.com or www.wiley.com

Reprinted April 2004

All Rights Reserved. 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 under the terms of the Copyright, Designs and Patents Act 1988 or under the terms of a licence issued by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London W1T 4LP, UK, without the permission in writing of the Publisher. Requests to the Publisher should be addressed to the Permissions Department, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, or emailed to permreq@wiley.co.uk, or faxed to (+44) 1243 770620.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the Publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark,
Singapore 129809

John Wiley & Sons (Canada) Ltd, 22 Worcester Road, Etobicoke, Ontario M9W 1L1

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

British Library Cataloguing in Publication Data

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

ISBN 0-470-86813-9 (Cloth)

ISBN 0-470-86814-7 (Paper)

Typeset in 10/12pt Times by LaTeX files supplied by the author, processed by
Integra Software Services Pvt. Ltd, Pondicherry, India

Printed and bound in Great Britain by Biddles Ltd, King's Lynn, Norfolk

This book is printed on acid-free paper responsibly manufactured from sustainable forestry,
in which at least two trees are planted for each one used for paper production.

FUNDAMENTALS OF PROBABILITY AND STATISTICS FOR ENGINEERS

To the memory of my parents

Preface

This book was written for an introductory one-semester or two-quarter course in probability and statistics for students in engineering and applied sciences. No previous knowledge of probability or statistics is presumed but a good understanding of calculus is a prerequisite for the material.

The development of this book was guided by a number of considerations observed over many years of teaching courses in this subject area, including the following:

- As an introductory course, a sound and rigorous treatment of the basic principles is imperative for a proper understanding of the subject matter and for confidence in applying these principles to practical problem solving. A student, depending upon his or her major field of study, will no doubt pursue advanced work in this area in one or more of the many possible directions. How well is he or she prepared to do this strongly depends on his or her mastery of the fundamentals.
- It is important that the student develop an early appreciation for applications. Demonstrations of the utility of this material in nonsuperficial applications not only sustain student interest but also provide the student with stimulation to delve more deeply into the fundamentals.
- Most of the students in engineering and applied sciences can only devote one semester or two quarters to a course of this nature in their programs. Recognizing that the coverage is time limited, it is important that the material be self-contained, representing a reasonably complete and applicable body of knowledge.

The choice of the contents for this book is in line with the foregoing observations. The major objective is to give a careful presentation of the fundamentals in probability and statistics, the concept of probabilistic modeling, and the process of model selection, verification, and analysis. In this text, definitions and theorems are carefully stated and topics rigorously treated but care is taken not to become entangled in excessive mathematical details.

Practical examples are emphasized; they are purposely selected from many different fields and not slanted toward any particular applied area. The same objective is observed in making up the exercises at the back of each chapter.

Because of the self-imposed criterion of writing a comprehensive text and presenting it within a limited time frame, there is a tight continuity from one topic to the next. Some flexibility exists in Chapters 6 and 7 that include discussions on more specialized distributions used in practice. For example, extreme-value distributions may be bypassed, if it is deemed necessary, without serious loss of continuity. Also, Chapter 11 on linear models may be deferred to a follow-up course if time does not allow its full coverage.

It is a pleasure to acknowledge the substantial help I received from students in my courses over many years and from my colleagues and friends. Their constructive comments on preliminary versions of this book led to many improvements. My sincere thanks go to Mrs. Carmella Gosden, who efficiently typed several drafts of this book. As in all my undertakings, my wife, Dottie, cared about this project and gave me her loving support for which I am deeply grateful.

T.T. Soong
Buffalo, New York

Contents

PREFACE	xiii
1 INTRODUCTION	1
1.1 Organization of Text	2
1.2 Probability Tables and Computer Software	3
1.3 Prerequisites	3
PART A: PROBABILITY AND RANDOM VARIABLES	5
2 BASIC PROBABILITY CONCEPTS	7
2.1 Elements of Set Theory	8
2.1.1 Set Operations	9
2.2 Sample Space and Probability Measure	12
2.2.1 Axioms of Probability	13
2.2.2 Assignment of Probability	16
2.3 Statistical Independence	17
2.4 Conditional Probability	20
Reference	28
Further Reading	28
Problems	28
3 RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS	37
3.1 Random Variables	37
3.2 Probability Distributions	39
3.2.1 Probability Distribution Function	39
3.2.2 Probability Mass Function for Discrete Random Variables	41

3.2.3 Probability Density Function for Continuous Random Variables	44
3.2.4 Mixed-Type Distribution	46
3.3 Two or More Random Variables	49
3.3.1 Joint Probability Distribution Function	49
3.3.2 Joint Probability Mass Function	51
3.3.3 Joint Probability Density Function	55
3.4 Conditional Distribution and Independence	61
Further Reading and Comments	66
Problems	67
4 EXPECTATIONS AND MOMENTS	75
4.1 Moments of a Single Random Variable	76
4.1.1 Mean, Median, and Mode	76
4.1.2 Central Moments, Variance, and Standard Deviation	79
4.1.3 Conditional Expectation	83
4.2 Chebyshev Inequality	86
4.3 Moments of Two or More Random Variables	87
4.3.1 Covariance and Correlation Coefficient	88
4.3.2 Schwarz Inequality	92
4.3.3 The Case of Three or More Random Variables	92
4.4 Moments of Sums of Random Variables	93
4.5 Characteristic Functions	98
4.5.1 Generation of Moments	99
4.5.2 Inversion Formulae	101
4.5.3 Joint Characteristic Functions	108
Further Reading and Comments	112
Problems	112
5 FUNCTIONS OF RANDOM VARIABLES	119
5.1 Functions of One Random Variable	119
5.1.1 Probability Distribution	120
5.1.2 Moments	134
5.2 Functions of Two or More Random Variables	137
5.2.1 Sums of Random Variables	145
5.3 m Functions of n Random Variables	147
Reference	153
Problems	154
6 SOME IMPORTANT DISCRETE DISTRIBUTIONS	161
6.1 Bernoulli Trials	161
6.1.1 Binomial Distribution	162

6.1.2 Geometric Distribution	167
6.1.3 Negative Binomial Distribution	169
6.2 Multinomial Distribution	172
6.3 Poisson Distribution	173
6.3.1 Spatial Distributions	181
6.3.2 The Poisson Approximation to the Binomial Distribution	182
6.4 Summary	183
Further Reading	184
Problems	185
7 SOME IMPORTANT CONTINUOUS DISTRIBUTIONS	191
7.1 Uniform Distribution	191
7.1.1 Bivariate Uniform Distribution	193
7.2 Gaussian or Normal Distribution	196
7.2.1 The Central Limit Theorem	199
7.2.2 Probability Tabulations	201
7.2.3 Multivariate Normal Distribution	205
7.2.4 Sums of Normal Random Variables	207
7.3 Lognormal Distribution	209
7.3.1 Probability Tabulations	211
7.4 Gamma and Related Distributions	212
7.4.1 Exponential Distribution	215
7.4.2 Chi-Squared Distribution	219
7.5 Beta and Related Distributions	221
7.5.1 Probability Tabulations	223
7.5.2 Generalized Beta Distribution	225
7.6 Extreme-Value Distributions	226
7.6.1 Type-I Asymptotic Distributions of Extreme Values	228
7.6.2 Type-II Asymptotic Distributions of Extreme Values	233
7.6.3 Type-III Asymptotic Distributions of Extreme Values	234
7.7 Summary	238
References	238
Further Reading and Comments	238
Problems	239
PART B: STATISTICAL INFERENCE, PARAMETER ESTIMATION, AND MODEL VERIFICATION	245
8 OBSERVED DATA AND GRAPHICAL REPRESENTATION	247
8.1 Histogram and Frequency Diagrams	248
References	252
Problems	253

9	PARAMETER ESTIMATION	259
9.1	Samples and Statistics	259
9.1.1	Sample Mean	261
9.1.2	Sample Variance	262
9.1.3	Sample Moments	263
9.1.4	Order Statistics	264
9.2	Quality Criteria for Estimates	264
9.2.1	Unbiasedness	265
9.2.2	Minimum Variance	266
9.2.3	Consistency	274
9.2.4	Sufficiency	275
9.3	Methods of Estimation	277
9.3.1	Point Estimation	277
9.3.2	Interval Estimation	294
	References	306
	Further Reading and Comments	306
	Problems	307
10	MODEL VERIFICATION	315
10.1	Preliminaries	315
10.1.1	Type-I and Type-II Errors	316
10.2	Chi-Squared Goodness-of-Fit Test	316
10.2.1	The Case of Known Parameters	317
10.2.2	The Case of Estimated Parameters	322
10.3	Kolmogorov-Smirnov Test	327
	References	330
	Further Reading and Comments	330
	Problems	330
11	LINEAR MODELS AND LINEAR REGRESSION	335
11.1	Simple Linear Regression	335
11.1.1	Least Squares Method of Estimation	336
11.1.2	Properties of Least-Square Estimators	342
11.1.3	Unbiased Estimator for σ^2	345
11.1.4	Confidence Intervals for Regression Coefficients	347
11.1.5	Significance Tests	351
11.2	Multiple Linear Regression	354
11.2.1	Least Squares Method of Estimation	354
11.3	Other Regression Models	357
	Reference	359
	Further Reading	359
	Problems	359

APPENDIX A: TABLES	365
A.1 Binomial Mass Function	365
A.2 Poisson Mass Function	367
A.3 Standardized Normal Distribution Function	369
A.4 Student's t Distribution with n Degrees of Freedom	370
A.5 Chi-Squared Distribution with n Degrees of Freedom	371
A.6 D_2 Distribution with Sample Size n	372
References	373
 APPENDIX B: COMPUTER SOFTWARE	 375
 APPENDIX C: ANSWERS TO SELECTED PROBLEMS	 379
Chapter 2	379
Chapter 3	380
Chapter 4	381
Chapter 5	382
Chapter 6	384
Chapter 7	385
Chapter 8	385
Chapter 9	385
Chapter 10	386
Chapter 11	386
 SUBJECT INDEX	 389

Introduction

At present, almost all undergraduate curricula in engineering and applied sciences contain at least one basic course in probability and statistical inference. The recognition of this need for introducing the ideas of probability theory in a wide variety of scientific fields today reflects in part some of the profound changes in science and engineering education over the past 25 years.

One of the most significant is the greater emphasis that has been placed upon complexity and precision. A scientist now recognizes the importance of studying scientific phenomena having complex interrelations among their components; these components are often not only mechanical or electrical parts but also ‘soft-science’ in nature, such as those stemming from behavioral and social sciences. The design of a comprehensive transportation system, for example, requires a good understanding of technological aspects of the problem as well as of the behavior patterns of the user, land-use regulations, environmental requirements, pricing policies, and so on.

Moreover, precision is stressed – precision in describing interrelationships among factors involved in a scientific phenomenon and precision in predicting its behavior. This, coupled with increasing complexity in the problems we face, leads to the recognition that a great deal of uncertainty and variability are inevitably present in problem formulation, and one of the mathematical tools that is effective in dealing with them is probability and statistics.

Probabilistic ideas are used in a wide variety of scientific investigations involving randomness. Randomness is an empirical phenomenon characterized by the property that the quantities in which we are interested do not have a predictable outcome under a given set of circumstances, but instead there is a statistical regularity associated with different possible outcomes. Loosely speaking, statistical regularity means that, in observing outcomes of an experiment a large number of times (say n), the ratio m/n , where m is the number of observed occurrences of a specific outcome, tends to a unique limit as n becomes large. For example, the outcome of flipping a coin is not predictable but there is statistical regularity in that the ratio m/n approaches $\frac{1}{2}$ for either

heads or tails. Random phenomena in scientific areas abound: noise in radio signals, intensity of wind gusts, mechanical vibration due to atmospheric disturbances, Brownian motion of particles in a liquid, number of telephone calls made by a given population, length of queues at a ticket counter, choice of transportation modes by a group of individuals, and countless others. It is not inaccurate to say that randomness is present in any realistic conceptual model of a real-world phenomenon.

1.1 ORGANIZATION OF TEXT

This book is concerned with the development of basic principles in constructing probability models and the subsequent analysis of these models. As in other scientific modeling procedures, the basic cycle of this undertaking consists of a number of fundamental steps; these are schematically presented in Figure 1.1. A basic understanding of probability theory and random variables is central to the whole modeling process as they provide the required mathematical machinery with which the modeling process is carried out and consequences deduced. The step from B to C in Figure 1.1 is the induction step by which the structure of the model is formed from factual observations of the scientific phenomenon under study. Model verification and parameter estimation (E) on the basis of observed data (D) fall within the framework of statistical inference. A model

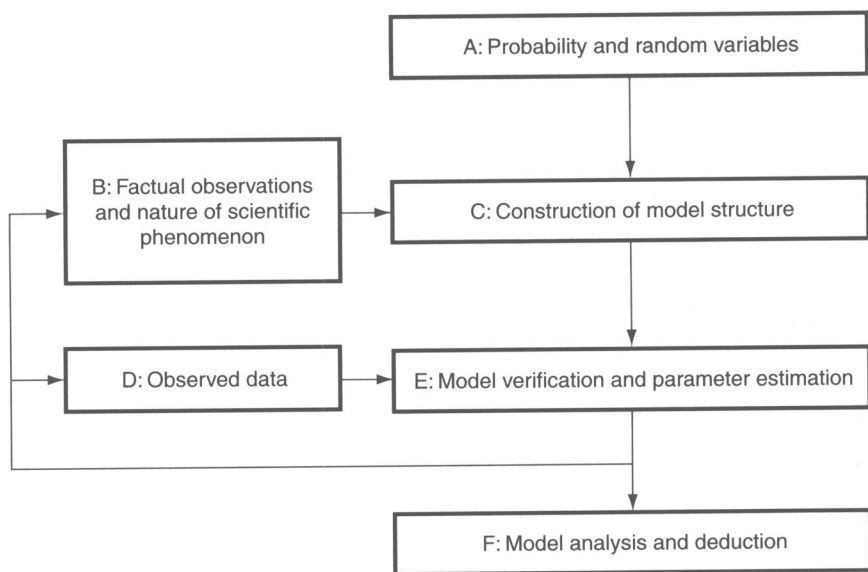


Figure 1.1 Basic cycle of probabilistic modeling and analysis

may be rejected at this stage as a result of inadequate inductive reasoning or insufficient or deficient data. A reexamination of factual observations or additional data may be required here. Finally, model analysis and deduction are made to yield desired answers upon model substantiation.

In line with this outline of the basic steps, the book is divided into two parts. Part A (Chapters 2–7) addresses probability fundamentals involved in steps $A \rightarrow C$, $B \rightarrow C$, and $E \rightarrow F$ (Figure 1.1). Chapters 2–5 provide these fundamentals, which constitute the foundation of all subsequent development. Some important probability distributions are introduced in Chapters 6 and 7. The nature and applications of these distributions are discussed. An understanding of the situations in which these distributions arise enables us to choose an appropriate distribution, or model, for a scientific phenomenon.

Part B (Chapters 8–11) is concerned principally with step $D \rightarrow E$ (Figure 1.1), the statistical inference portion of the text. Starting with data and data representation in Chapter 8, parameter estimation techniques are carefully developed in Chapter 9, followed by a detailed discussion in Chapter 10 of a number of selected statistical tests that are useful for the purpose of model verification. In Chapter 11, the tools developed in Chapters 9 and 10 for parameter estimation and model verification are applied to the study of linear regression models, a very useful class of models encountered in science and engineering.

The topics covered in Part B are somewhat selective, but much of the foundation in statistical inference is laid. This foundation should help the reader to pursue further studies in related and more advanced areas.

1.2 PROBABILITY TABLES AND COMPUTER SOFTWARE

The application of the materials in this book to practical problems will require calculations of various probabilities and statistical functions, which can be time consuming. To facilitate these calculations, some of the probability tables are provided in Appendix A. It should be pointed out, however, that a large number of computer software packages and spreadsheets are now available that provide this information as well as perform a host of other statistical calculations. As an example, some statistical functions available in Microsoft® Excel™ 2000 are listed in Appendix B.

1.3 PREREQUISITES

The material presented in this book is calculus-based. The mathematical prerequisite for a course using this book is a good understanding of differential and integral calculus, including partial differentiation and multidimensional integrals. Familiarity in linear algebra, vectors, and matrices is also required.



Part A

Probability and Random Variables

The

and

random

is not

to be

of pro

prob

In

situ

class

tain

of us

Con

tract

a re

ther

ma

2.1

mo

in

ing

var

var

to

pe

2.1

all

are

an

u

2.1

2.1

2.1

2.1

2.1

2.1