$$\frac{1}{243}\int_0^\infty xe^{-x/3}\left(\int_x^\infty ye^{-y/3}\left(\int_y^\infty e^{-z/3}\,dz\right)dy\right)dx$$

$$\frac{1}{243}\int_0^\infty xe^{-x/3}\left(\int_x^\infty ye^{-y/3}\cdot 3e^{-y/3}\,dy\right)dx$$

$$\frac{1}{81}\int_0^\infty xe^{-x/3}\left[\frac{ye^{-2y/3}}{-2/3}\Big|_x^\infty + \frac{3}{2}\int_x^\infty e^{-2y/3}\,dy\right]dx$$

STATISTICAL INFERENCE

$$\frac{1}{243}\int_0^\infty xe^{-x/3}\left(\int_x^\infty ye^{-y/3}\left(\int_y^\infty e^{-z/3}\,dz\right)dy\right)dx$$

$$\frac{1}{243}\int_0^\infty xe^{-x/3}\left(\int_x^\infty ye^{-y/3}\cdot 3e^{-y/3}\,dy\right)dx$$

$$\frac{1}{81}\int_0^\infty xe^{-x/3}\left[\frac{ye^{-2y/3}}{-2/3}\Big|_x^\infty + \frac{3}{2}\int_x^\infty e^{-2y/3}\,dy\right]dx$$

$$\frac{1}{81}\int_0^\infty xe^{-x/3}\left[\frac{3}{2}xe^{-2x/3} + \frac{9}{4}e^{-2x/3}\right]dx$$

$$\frac{1}{81}\int_0^\infty \left[\frac{3}{2}x^2e^{-x} + \frac{9}{4}xe^{-x}\right]dx$$

$$\frac{1}{81}\left[\frac{3}{2}(2) + \frac{9}{4}(1)\right]$$

$$(1/81)(21/4) = 7/108.$$

Vijay K. Rohatgi

A Volume in the Wiley Series in Probability and Mathematical Statistic
Vic Barnett, Ralph A. Bradley, J. Stuart Hunter, David G. Kendall,

STATISTICAL INFERENCE

VIJAY K. ROHATGI
Department of Mathematics and Statistics
Bowling Green State University
Bowling Green, Ohio

JOHN WILEY & SONS
New York · Chichester · Brisbane · Toronto · Singapore

Library of Congress Cataloging in Publication Data:

Rohatgi, V. K., 1939–
Statistical inference.

Includes index.
1. Mathematical statistics. I. Title.

QA276.R624 1984 519.5 83-21848
ISBN 0-471-87126-5

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

PREFACE

This course in statistical inference is designed for juniors and seniors in most disciplines (including mathematics). No prior knowledge of probability or statistics is assumed or required. For a class meeting four hours a week this is a two-semester or three-quarter course. The mathematical prerequisite is modest. The prospective student is expected to have had a three-semester or four-quarter course in calculus. Whenever it is felt that a certain topic is not covered in a calculus course at the stated prerequisite level, enough supplementary details are provided. For example, a section on gamma and beta functions is included, as are supplementary details on generating functions.

There are many fine books available at this level. Then why another? Their titles notwithstanding, almost all of these books are in probability and statistics. Roughly the first half of these texts is usually devoted to probability and the second half to statistics. Statistics almost always means parametric statistics, with a discussion of nonparametric statistics usually relegated to the last chapter.

This text is on statistical inference. My approach to the subject separates this text from the rest in several respects. First, probability is treated here from a modeling viewpoint with strong emphasis on applications. Second, statistics is not relegated to the second half of the course—indeed, statistical thinking is encouraged and emphasized from the beginning. Formal language of statistical inference is introduced as early as Chapter 4 (essentially the third chapter, since Chapter 1 is introductory), immediately after probability distributions have been introduced. Inferential questions are considered along with probabilistic models, in Chapters 6 and 7. This approach allows and facilitates an early introduction to parametric as well as nonparametric techniques. Indeed, every attempt has been made to integrate the two: Empirical distribution function is introduced in Chapter 4, in Chapter 5 we show that it is unbiased and consistent, and in Chapter 10 we show that it is the maximum likelihood estimate of the population distribution function. Sign test and Fisher–Irwin test are introduced in Chapter 6, and inference concerning quantiles is covered in Section 8.5. There is not even a separate chapter entitled nonparametric statistical inference.

Apart from the growing importance of and interest in statistics, there are several reasons for introducing statistics early in the text. The traditional approach in which statistics follows probability leaves the reader with the false notion that probability and statistics are the same and that statistics is the mathematics of computing certain probabilities. I do not believe that an appreciation for the utility of statistics should be withheld until a large dose of probability is digested. In a traditional course, students who leave the course after one semester or one quarter learn little or no statistics. They are left with little understanding of the important role that statistics plays in scientific research. A short course in probability becomes just another hurdle to pass before graduation, and the students are deprived of the chance to use statistics in their disciplines. I believe that the design of this text alleviates these problems and enables the student to acquire an outlook approaching that of a modern mathematical statistician and an ability to apply statistical methods in a variety of situations.

There appears to be a reasonable agreement on the topics to be included in a course at this level. I depart a little from the traditional coverage, choosing to exclude regression since it did not quite fit into the scheme of "one sample, two sample, many sample" problems. On the other hand, I include the Friedman test, Kendall's coefficient of concordance, and multiple comparison procedures, which are usually not done at this level.

The guiding principles in my selection have been usefulness, interrelationship, and continuity. The ordering of the selections is dictated by need rather than relative importance.

While the topics covered here are traditional, their order, coverage, and discussion are not. Many other features of this text separate it from previous texts. I mention a few here:

(i) An unusually large number of problems (about 1450) and examples (about 400) are included. Problems are included at the end of each section and are graded according to their degree of difficulty; more advanced (and usually more mathematical) problems are identified by an asterisk. A set of review problems is also provided at the end of each chapter to test the student's ability to choose relevant techniques. Every attempt has been made to avoid the annoying and time-consuming practice of creating new problems by referring to earlier problems (often scores of pages earlier). Either completely independent problems are given in each section or relevant details (with cross references) are restated whenever a problem is important enough to be continued in a later section. The amount of duplication, however, is minimal and improves readability.

(ii) Sections with a significant amount of mathematical content are also identified by an asterisk. These sections are aimed at the more mathematically inclined students and may be omitted at first reading. This procedure allows us to encompass a much wider audience without sacrificing mathematical rigor. Needless to say, this is not a recipe book. The emphasis is on the how and why of all the techniques introduced here, in the hope that the student is challenged to think like a statistician.

(iii) Applications are included from diverse disciplines. Most examples and problems are application oriented. It is true that no attempt has been made to include "real life data" in these problems and examples but I hope that the student will be motivated enough to follow up this course with an exploratory data analysis course.

(iv) A large number of figures (about 150) and remarks supplement the text. Summaries of main results are highlighted in boxed or tabular form.

In a two-semester course, meeting four times a week, my students have been able to cover the first twelve chapters of the text without much haste. In the first semester we cover the first five chapters with a great deal of emphasis on Chapter 4, the introduction to statistical inference. In the second semester we cover all of Chapters 6 and 7 on models (but at increased pace and with emphasis on inferential techniques), most of Chapter 8 on random variables and random vectors (usually excluding Section 6, depending on the class composition), and all of Chapter 9 on large-sample theory.

In Chapter 10 on point and interval estimation, more time is spent on sections on sufficiency, method of moments, maximum likelihood estimation, and confidence intervals than on other sections. In Chapter 11 on testing hypotheses, we emphasize sections on Wilcoxon signed rank test, two-sample tests, chi-square test of goodness of fit, and measures of association. The point is that if the introductory chapter on statistical inference (Chapter 4) is covered carefully, then one need not spend much time on unbiased estimation (Section 10.3), Neyman–Pearson Lemma (Section 11.2), composite hypotheses (Section 11.3), or likelihood ratio tests (Section 11.4). Chapter 12 on categorical data is covered completely.

In a three-quarters course, the pace should be such that the first four chapters are covered in the first quarter, Chapters 5 to 9 in the second quarter, and the remaining chapters in the third quarter. If it is found necessary to cover Chapter 13 on k-sample problems in detail, we exclude the technical sections on transformations (Section 8.3) and generating functions (Section 8.6), and also sections on inference concerning quantiles (Section 9.8), Bayesian estimation (Section 10.6), and composite hypotheses (Section 11.3).

I take this opportunity to thank many colleagues, friends, and students who made suggestions for improvement. In particular, I am indebted to Dr. Humphrey Fong for drawing many diagrams, to Dr. Victor Norton for some numerical computations used in Chapter 4 and to my students, especially Lisa Killel and Barbara Christman, for checking many solutions to problems. I am grateful to the Literary Executor of the late Sir Ronald A. Fisher, F. R. S., to Dr. Frank Yates, F. R. S., and to Longman Group Ltd., London, for permission to reprint Tables 3 and 4 from their book *Statistical Tables for Biological, Agricultural and Medical Research* (6th edition, 1974). Thanks are also due to Macmillan Publishing Company, Harvard University Press, the Rand Corporation, Bell Laboratories, Iowa University Press, John Wiley & Sons, the Institute of Mathematical Statistics, Stanford University Press, Wadsworth Publishing Company, Biometrika Trustees, Statistica Neerlandica, Addison–Wesley Publishing Company, and the American Statistical Association for permission to use tables and to John Wiley & Sons for permission to use some diagrams.

I also thank the several anonymous reviewers whose constructive comments greatly improved presentation. Finally, I thank Mary Chambers for her excellent typing and Beatrice Shube, my editor at John Wiley & Sons, for her cooperation and support in this venture.

VIJAY K. ROHATGI

Bowling Green, Ohio
February 1984

CONTENTS

CHAPTER 1

Introduction

1.1 INTRODUCTION

Probabilistic statements are an integral part of our language. We use the expressions random, odds, chance, risk, likelihood, likely, plausible, credible, as likely as not, more often than not, almost certain, possible but not probable, and so on. All these words and phrases are used to convey a certain degree of uncertainty although their nontechnical usage does not permit sharp distinctions, say, between probable and likely or between improbable and impossible. One of our objectives in this course is to introduce (in Chapter 2) a numerical measure of uncertainty. Once this is done we can use the apparatus of mathematics to describe many physical or artificial phenomena involving uncertainty. In the process, we shall learn to use some of these words and phrases as technical terms.

The basic objective of this course, however, is to introduce techniques of statistical inference. It is hardly necessary to emphasize here the importance of statistics in today's world. Statistics is used in almost every field of activity. News media carry statistics on unemployment rate, inflation rate, batting averages, average rainfall, the money supply, crime rates. They carry the results of Gallop and Harris polls and many other polls. Most people associate statistics with a mass of numerical facts, or *data*. To be sure statistics does deal with the collection and description of data. But a statistician does much more. He or she is—or should be—involved in the planning and design of experiments, in collecting information, and in deciding how best to use the collected information to provide a basis for decision making. This text deals mostly with this latter aspect: the art of evaluating information to draw reliable inferences about the true nature of the phenomenon under study. This is called *statistical inference*.

1.2 STOCHASTIC MODELS

Frequently the objective of scientific research is to give an adequate mathematical description of some natural or artificial phenomenon. A *model* may be defined as a mathematical idealization used to approximate an observable phenomenon. In any such idealization, certain assumptions are made (and hence certain details are ignored as unimportant). The success of the model depends on whether or not these assumptions are valid (and on whether the details ignored actually are unimportant).

In order to check the validity of a model, that is, whether or not a model adequately describes the phenomenon being studied, we take observations. The process of taking observations (to discover something that is new or to demonstrate something that is already known) is called an *experiment*.

A *deterministic model* is one which stipulates that the conditions under which an experiment is performed determine the outcome of the experiment. Thus the distance d traveled by an automobile in time t hours at a constant speed s kilometers per hour is governed by the relation $d = st$. Knowledge of s and t precisely determines d. Similarly, gravitational laws describe precisely what happens to a falling object, and Kepler's laws describe the behavior of planets.

A *nondeterministic* (or *stochastic*) model, on the other hand, is one in which past information, no matter how voluminous, does not permit the formulation of a rule to determine the precise outcome of an experiment. Many natural or artificial phenomena are *random* in the sense that the exact outcome cannot be predicted, and yet there is a predictable long-term pattern. Stochastic models may be used to describe such phenomena. Consider, for example, the sexes of newborns in a certain County Hospital. Let B denote a boy and G a girl. Suppose sexes are recorded in order of birth. Then we observe a sequence of letters B and G, such as

$$G\,B\,G\,G\,B\,G\,B\,B\,B\,G\,G\ldots.$$

This sequence exhibits no apparent regularity. Moreover, one cannot predict the sex of the next newborn, and yet one can predict that in the long run the proportion of girls (or boys) in this sequence will settle down near 1/2. This long-run behavior is called *statistical regularity* and is noticeable, for example, in all games of chance.

In this text we are interested only in experiments that exhibit the phenomena of randomness and statistical regularity. Probability models are used to describe such phenomena.

We consider some examples.

Example 1. Measuring Gravity. Consider a simple pendulum with unit mass suspended from a fixed point O which swings only under the effect of gravity. Assume that the string is of unit length and is weightless. Let $t = t(\theta)$ be the period of oscillation when θ is the angle between the pendulum and the vertical (see Figure 1). It is shown in calculus[†] that when the pendulum goes from $\theta = \theta_0$ to $\theta = 0$ (corresponding to one fourth of a period)

$$\frac{t}{4} = \sqrt{\frac{1}{2g}} \int_0^{\theta_0} \frac{d\theta}{\sqrt{\cos\theta - \cos\theta_0}}$$

[†]Al Shenk, *Calculus and Analytic Geometry*, Scott-Foresman, Glenview Illinois, 1979, p. 544.

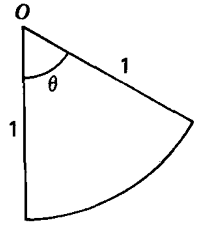

Figure 1

so that

$$t = 4\sqrt{\frac{1}{g}} \int_0^{\pi/2} \frac{dx}{\sqrt{1 - k^2 \sin x}}, \qquad k = \sin\frac{\theta_0}{2}.$$

Hence for small oscillations $k = 0$ and $t = 2\pi/\sqrt{g}$ gives approximately the period of oscillation. This gives a deterministic model giving g as a function of t, namely,

$$g = \frac{4\pi^2}{t^2}.$$

If t can be measured accurately, this formula gives the value of g. If repeated readings on t are taken, they will all be found different (randomness) and yet there will be a long-run pattern in that these readings will all concentrate near the true value of t (statistical regularity). The randomness may be due to limitations of our measuring device and the ability of the person taking the readings. To take into account these nondeterministic factors we may assume that t varies in some random manner such that

$$t = \frac{2\pi}{\sqrt{g}} + \varepsilon$$

where $2\pi/\sqrt{g}$ is the true value of t and ε is a random error that varies from reading to reading. A stochastic model will postulate that $t = 2\pi/\sqrt{g} + \varepsilon$ along with some assumptions about the random error ε. A statistician's job, then, is to estimate g based on, say, n readings $t_1, t_2, \ldots, t_n$ on t. □

Example 2. Ohm's Law. According to Ohm's law for a simple circuit, the voltage V is related to the current I and the resistance R according to the formula $V = IR$. If the conditions underlying this deterministic relationship are met, this model predicts precisely the value of V given those of I and R. Such a description may be adequate for most practical purposes.

If, on the other hand, repeated readings of either I or R or both are found to vary, then V will also vary in a random manner. Given a pair of readings on I and R, V is still determined by $V = IR$. Since all pairs of readings on I and R are different, so will be the values of V. In a stochastic model the assumptions we make concerning the randomness in I and/or R determine the random behavior of V through the relation $V = IR$. □

Example 3. Radioactive Disintegration. The deterministic model for radioactive disintegration postulates that the rate of decay of a quantity of radioactive element is proportional to the mass of the element. That is,

$$\frac{dm}{dt} = -\lambda m, \tag{1}$$

where $\lambda > 0$ is a constant giving the rate of decay and m is the mass of the element. Integrating (1) with respect to t we get:

$$\ln m = -\lambda t + c$$

where c is a constant. If $m = m_0$ at time $t = 0$, then $c = \ln m_0$ and we have:

$$m = m_0 \exp(-\lambda t) \tag{2}$$

for $t \geq 0$. Given m_0 and λ, we know m as a function of t.

The exact number of decays in a given time interval, however, cannot be predicted with certainty because of the random nature of the time at which an element disintegrates. This forces us to consider a stochastic model. In a stochastic model one makes certain assumptions concerning the probability that a given element will decay in time interval $[0, t]$. These lead to an adequate description of the probability of exactly k decays in $[0, t]$. □

Example 4. Rolling a Die. A die is rolled. Let X be the number of points (face value) on the upper face. No deterministic model can predict which of the six face values 1, 2, 3, 4, 5, or 6 will show up on any particular roll of the die (randomness) and yet there is a predictable long-run pattern. The proportion of any particular face value in a sequence of rolls will be about $\frac{1}{6}$ provided the die is not loaded. □

It should be clear by now that in some experiments a deterministic model is adequate. In other experiments (Examples 3 and 4) we must use a stochastic model. In still other experiments, a stochastic model may be more appropriate than a deterministic model. Experiments for which a stochastic model is more appropriate are called *statistical* or *random* experiments.

DEFINITION 1. (RANDOM OR STATISTICAL EXPERIMENT). An experiment that has the following features is called a random or statistical experiment.

(i) All possible outcomes of the experiment are known in advance.
(ii) The exact outcome of any specific performance of the experiment is unpredictable (randomness).
(iii) The experiment can be repeated under (more or less) identical conditions.
(iv) There is a predictable long-run pattern (statistical regularity).

Example 5. Some Typical Random Experiments. We list here some typical examples of random experiments.

(i) Toss a coin and observe the up face.
(ii) A light bulb manufactured at a certain plant is put to a lifetime test and the time at which it fails is recorded.
(iii) A pair of dice is rolled and the face values that show up are recorded.
(iv) A lot consisting of N items containing D defectives ($D \leq N$) is sampled. An item sampled is not replaced, and we record whether the item selected is defective or nondefective. The process continues until all defective items are found.
(v) The three components of velocity of an orbital satellite are recorded continuously for a 24-hour period.

(vi) A manufacturer of refrigerators inspects its refrigerators for 10 types of defects. The number of defects found in each refrigerator inspected is recorded.

(vii) The number of girls in every family with five children is recorded for a certain town.

Problems for Section 1.2

In the following problems, state whether a deterministic or a nondeterministic model is more appropriate. Identify the sources of randomness. (In each case there is no clearly right or wrong answer; the decision you make is subjective.)

1. The time (in seconds) elapsed is measured between the end of a question asked of a person and the start of her or his response.
2. In order to estimate the average size of a car pool, a selection of cars is stopped on a suburban highway and the number of riders recorded.
3. On a graph paper, a line with equation $y = 3x + 5$ is drawn and the values of y for $x = 1, 3, 5, 7, 9$ are recorded.
4. Consider a binary communication channel that transmits coded messages consisting of a sequence of 0's and 1's. Due to noise, a transmitted 0 might be received as a 1. The experiment consists of recording the transmitted symbol (0 or 1) and the corresponding received symbol (0 or 1).
5. In order to estimate the average time patients spent at the emergency room of a county hospital from arrival to departure after service, the service times of patients are recorded.
6. A coin is dropped from a fixed height, and the time it takes to reach the ground is measured.
7. A roulette wheel is spun and a ball is rolled on its edge. The color (black or red) of the sector in which the ball comes to rest is recorded. (A roulette wheel consists of 38 equal sectors, marked 0, 00, 1, 2,...,36. The sectors 0 and 00 are green. Half of the remaining 36 sectors are red, the other half black.)

1.3 PROBABILITY, STATISTICS, AND INFERENCE†

There are three essential components of a stochastic model:

(i) Indentification of all possible outcomes of the experiment.

(ii) Identification of all events of interest.

(iii) Assignment of probabilities to these events of interest.

The most important as well as most interesting and difficult part of model building is the assignment of probabilities. Consequently a lot of attention will be devoted to it.

† This section uses some technical terms that are defined in Chapter 2. It may be read in conjunction with or after Chapter 2.

Consider a random experiment and suppose we have agreed on a stochastic model for it. This means that we have identified all the outcomes and relevant events and made an assignment of probabilities to these events. The word *population*, in statistics, refers to the collection of all outcomes along with the assignment of probabilities to events. The object in statistics is to say something about this population. This is done on the basis of a *sample*, which is simply a part of the population. It is clear that a sample is not just any part of the population. In order for our inferences to be meaningful, randomness should somehow be incorporated in the process of sampling. More will be said about this in later chapters.

At this stage let us distinguish between probability and statistics. In probability, we make certain assumptions about the population and then say something about the sample. That is, the problem in probability is: *Given a stochastic model, what can we say about the outcomes*?

In statistics, the process is reversed. The problem in statistics is: *Given a sample* (*set of outcomes*), *what can we say about the population* (*or the model*)?

Example 1. Coin Tossing. Suppose the random experiment consists of tossing a coin and observing the outcome. There are two possible outcomes, namely, heads or tails. The stochastic model may be that the coin is fair, that is, not fraudulently weighted. We will see that this completely specifies the probability of a head ($= 1/2$) and hence also of tails ($= 1/2$). In probability we ask questions such as: Given that the coin is fair, what is the chance of observing 10 heads in 25 tosses of the coin? In statistics, on the other hand we ask: Given that 25 tosses of a coin resulted in 10 heads, can we assert that the coin is fair? □

Example 2. Gasoline Mileage. When a new car model is introduced, the automobile company advertises (an estimated) Environmental Protection Agency rating of fuel consumption (miles per gallon) for comparison purposes. The initial problem of determining the probability distribution of fuel consumption for this model is a statistical problem. Once this has been solved, the computation of the probability that a particular car will give, say, at least 38 miles per gallon is a problem in probability. Similarly, estimating the average gas mileage for the model is a statistical problem. □

Example 3. Number of Telephone Calls. The number of telephone calls initiated in a time interval of length t hours is recorded at a certain exchange. The initial problem of estimating the probability that k calls are initiated in an interval of length t hours is a problem in statistics. Once these probabilities have been well established for each $k = 0, 1, 2, \dots$, the computation of the probability that more than j calls are initiated in a one-hour period is a probability problem. □

Probability is basic to the study of statistics, and we devote the next two chapters to the fundamental ideas of probability theory. Some basic notions of statistics are introduced in Chapter 4. Beginning with Chapter 5, the two topics are integrated.

Statistical inference depends on the laws of probability. In order to ensure that these laws apply to the problem at hand, we insist that the sample be random in a certain sense (to be specified later). Our conclusions, which are based on the sample outcomes, are therefore as good as the stochastic model we use to represent the experiment. If we observe an event that has a small probability of occurring under the model, there are two possibilities. Either an event of such a small probability has actually occurred or the postulated model is not valid. Rather than accept the explanation that a rare event has happened, statisticians look for alternative explanations. They argue that events with low probabilities cast doubt on the validity of the postulated model. If, therefore, such an event is observed in spite of its low probability, then it provides evidence against the model. Suppose, for example, we assume that a coin is fair and that heads and tails are equally likely on any toss. If the coin is then tossed and we observe five heads in a row, we begin to wonder about our assumption. If the tossing continues and we observe 10 heads in a row, hardly anyone would argue against our conclusion that the coin is loaded. Probability provides the basis for this conclusion. The chance of observing 10 heads in a row in 10 tosses of a fair coin, as we shall see, is 1 in $2^{10} = 1024$, or less than .001. This is evidence against the model assumption that the coin is fair. We may be wrong in this conclusion, but the chance of being wrong is 1 in 1024. And that is a chance worth taking for most practical purposes.