Probability & Statistical Inference

Robert V. Hogg
& Elliot A. Tanis

COLLIER
MACMILLAN
INTERNATIONAL
EDITIONS

PROBABILITY AND STATISTICAL INFERENCE

by

Robert V. Hogg

University of Iowa

and

Elliot A. Tanis

NEW YORK

COLLIER MACMILLAN PUBLISHERS

LONDON

Printed in the United States of America

Macmillan Publishing Co., Inc.
866 Third Avenue, New York, New York 10022

Collier Macmillan Canada, Ltd.

Library of Congress Cataloging in Publication Data

Hogg, Robert V
Probability and statistical inference.

Includes index.
1. Probabilities. 2. Mathematical statistics.
I. Tanis, Elliot, joint author. II. Title.
QA273.H694 519.2 75-35592
ISBN 0-02-355650-1

Printing: 1 2 3 4 5 6 7 8 Year: 7 8 9 0 1 2

preface

This book is designed for use in a course having from three to six semester hours of credit, such as a three-hour course for one semester, a two-quarter course, or a three-hour course for the full academic year. No previous study of statistics is assumed, and a standard two-semester course in calculus should provide an adequate mathematical background. For that matter, the material is organized in such a way that little multiple integration is needed until the last few chapters.

Our aim is to provide a book, at this mathematical level, that emphasizes fundamental concepts and presents them in a logical order. Probability and distributions of the discrete type are treated first. The usual descriptive statistics are found by computing the characteristics of a discrete-type empirical distribution. Histograms and ogives motivate the definitions of probability density and distribution functions of the continuous type. Certain basic sampling distribution theory is immediately used to make some elementary statistical inferences, the first of which is distribution-free and is based on the order statistics. The probabilities associated with these inferences are easy to determine using the binomial distribution and provide good applications of approximating distributions. After some of the standard parametric and nonparametric inferences involving one and two distributions, multivariate distributions are introduced. We feel that the student is better prepared to understand them at this stage, and they provide the necessary background for chi-square tests and the analysis of variance. The final chapter concerns certain interesting theoretical problems, most of which are treated more fully in an advanced course in mathematical statistics.

Although it is not necessary to have a computer available to study this text, we have included some computer output to make certain theories (like the central limit theorem) more plausible.

We are indebted to the *Biometrika* Trustees for permission to include Tables III and V, which are abridgments and adaptations of tables published in *Biometrika Tables for Statisticians.* We are also 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 use Table III from their book *Statistical Tables for Biological, Agricultural, and Medical Research* (6th Edition, 1974), reproduced as our Table VI.

Finally, we wish to thank our colleagues and friends for many suggestions, Mrs. Mary DeYoung for her help with the typing, and our families for their patience and understanding during the preparation of this manuscript.

R. V. H.
E. A. T.

contents

10 CHI–SQUARE TESTS OF MODELS 329

11 ANALYSIS OF VARIANCE 359

12 A BRIEF THEORY OF STATISTICAL INFERENCE 381

1
PROBABILITY

1.1 Random Experiments and Random Variables

Many decisions have to be made that involve uncertainties. In medical research, interest may center on the effectiveness of a new vaccine for mumps; an agronomist must decide if an increase in yield can be attributed to a new strain of wheat; a meteorologist is interested in predicting the probability of rain; the state legislature must decide whether decreasing speed limits will help prevent accidents; the admissions officer of a college must predict the college performance of an incoming freshman. Probability and statistics can provide the models that could help people make decisions such as these.

In the study of probability we shall consider *random experiments*. Each experiment ends in an *outcome* that cannot be determined with certainty before the performance of the experiment. However, the experiment is such that the collection of every possible outcome can be listed; and this collection of all outcomes is called the outcome space or, more frequently, the *sample space S*.

The following examples will help illustrate what we mean by random experiments, outcomes, and sample spaces:

Example 1.1–1. Consider the flip of an unbiased coin as a random experiment. The outcome is heads (H) or tails (T) and the collection $S = \{H, T\}$ is the sample space.

Example 1.1–2. Flip an unbiased coin three times and observe the sequence of heads and tails. Here the sample space is the collection of sequences

$$S = \{(H, H, H), (H, H, T), (H, T, H), (T, H, H), (H, T, T), (T, H, T), (T, T, H), (T, T, T)\}.$$

Example 1.1–3. From a bowl containing three red (R), two white (W), and five blue (B) chips, draw one at random and observe its color. Here $S = \{R, W, B\}$.

Example 1.1–4. Each of six students selects an integer at random from the first 52 positive integers. We are interested in whether at least two of these six integers match (M) or whether they are all different (D). Thus, $S = \{M, D\}$.

Example 1.1–5. A light bulb is turned on continuously and we observe the time t until it burns out. Here $S = \{t: \ 0 \le t\}$.

Example 1.1–6. A three-month-old chicken is selected from a flock of chickens and weighed. The sample space is $S = \{w: \ 0 < w \le 9\}$. Note that 9 pounds is, perhaps, too large. In addition we are victims of the accuracy with which we can weigh chickens; thus, we might more realistically describe S as $S = \{w: \ w = 0.5, 0.6, \ldots, 5.0\}$. However, it is often easier to work with a mathematically "idealized" sample space rather than a more realistic one.

Let S denote a sample space, and let A be part of this collection S. Suppose the performance of the random experiment terminates so that the outcome is in A. Then we shall say that *event* A has occurred. Now consider the possibility of repeating the experiment a large number of times, say n. Then we can count the number of times that event A actually occurred throughout n performances; this number is called the frequency of event A and is denoted $N(A)$. The ratio $N(A)/n$ is called the *relative frequency* of event A in these n experiments. A relative frequency is usually very unstable for small values of n, but it tends to stabilize as n increases. Possibly you should check this by tossing a coin a large number of times, computing the relative frequency after each toss. This suggests that we associate with event A a number, say p, that is equal or approximately equal to the number about which the relative frequency seems to stabilize. This number p can then be taken as that number which the relative frequency of event A will be near in future performances of the experiment. Thus, although we cannot predict the outcome of a random experiment with certainty, we can, for a large value of n, predict fairly accurately the relative frequency associated with event A. The number p assigned to event A is called the *probability* of event A, and it is denoted by $P(A)$.

To illustrate some of these ideas, we performed the following two random experiments.

Example 1.1–7. Four unbiased coins are to be tossed and the number of heads observed. Here the sample space $S = \{0, 1, 2, 3, 4\}$. Let the event $A = \{0, 4\}$; that is, A occurs when the coins are all heads or all tails. This experiment was actually repeated a large number of times, and each time we recorded whether or not A had occurred. After 50 trials, the frequency of A was $N(A) = 4$; after 100 trials, $N(A) = 11$; after 500 trials, $N(A) = 49$; and after 1000 trials, $N(A) = 118$. These results provide the relative frequencies of 0.080, 0.110, 0.098, and 0.118, respectively. Accordingly, we believe that the probability $P(A)$ of the event A is close to 0.118. Later we will learn, if certain assumptions are fulfilled, that $P(A)$ equals 0.125.

Example 1.1–8. A pair of fair dice is to be cast and the sum of the dots on the top of the dice observed. Here

$$S = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}.$$

Let $A = \{7\}$; that is, A occurs when the sum of the dots equals 7. After repeating the experiment a large number of times, we observed these combinations of the number n of trials and the frequency $N(A)$ of A:

$$\begin{aligned} n &= 50, & N(A) &= 10; \\ n &= 100, & N(A) &= 17; \\ n &= 500, & N(A) &= 81; \\ n &= 1000, & N(A) &= 175. \end{aligned}$$

From these observations, the calculated respective relative frequencies $N(A)/n$ are 0.200, 0.170, 0.162, and 0.175. Other considerations later in the book allow us to assign the probability $P(A) = 1/6 = 0.167$ to the event A; this is near the relative frequency 0.175.

Note that a sample space S may be difficult to describe if the elements of S are not numbers. We shall now discuss how we can use a rule by which an element s of S may be associated with a number x. We begin the discussion with a simple example.

Example 1.1–9. In Example 1.1–1, we had the sample space $S = \{\text{H}, \text{T}\}$. Let X be a function defined on S such that $X(\text{H}) = 0$ and $X(\text{T}) = 1$. Thus, X is a real-valued function that has the sample space S as its domain and the space of real numbers $\{x: \; x = 0, 1\}$ as its range. We call X a random variable and, in this example, the space associated with X is $\{x: \; x = 0, 1\}$.

We now formulate the definition of a random variable.

DEFINITION 1.1–1. *Given a random experiment with a sample space S, a function X that assigns to each element s in S one and only one real number $X(s) = x$ is called a* random variable. *The space of X is the set of real numbers $\{x: \ x = X(s), s \in S\}$, where $s \in S$ means the element s belongs to the set S.*

It may be that the set S has elements that are themselves real numbers. In such an instance we could write $X(s) = s$ so that X is the identity function and the space of X is also S. This is illustrated in Example 1.1–10.

Example 1.1–10. Let the random experiment be the cast of a die. The sample space associated with this experiment is $S = \{1, 2, 3, 4, 5, 6\}$. For each $s \in S$, let $Y(s) = s$. The space of the random variable Y is then $\{1, 2, 3, 4, 5, 6\}$.

For notational purposes we shall denote the event $\{s: \ s \in S \text{ and } X(s) = a\}$ by $\{X = a\}$. That is, the event $\{X = a\}$ is the set of points in the sample space that are mapped onto the real number a by the function X. Similarly, $\{s: \ s \in S \text{ and } a < X(s) < b\}$ will be denoted by $\{a < X < b\}$. Now if we want to find the probabilities associated with events described in terms of X, such as $\{X = a\}$ and $\{a < X < b\}$, we use the probabilities of those events in the original space S, if they are known. That is, when we define the probability of these events, we shall let

$$P(X = a) = P(\{s: \ s \in S \text{ and } X(s) = a\})$$

and

$$P(a < X < b) = P(\{s: \ s \in S \text{ and } a < X(s) < b\}).$$

We say that probabilities are *induced* on the points of the space of X by the probabilities assigned to outcomes of the sample space S through the function X.

Example 1.1–11. If, in Example 1.1–10, we associate a probability of 1/6 with each outcome, then, for example, $P(Y = 5) = 1/6$, $P(2 \leq Y \leq 5) = 4/6$, and $P(Y \leq 2) = 2/6$ seem to be reasonable assignments.

The student will no doubt recognize two major difficulties here:

(1) In many practical situations the probabilities assigned to the events A of the sample space S are unknown.
(2) Since there are many ways of defining a function X on S, which function do we want to use?

As a matter of fact, the solutions to these problems in particular cases are major concerns in applied statistics. In considering (2), statisticians try to determine what "measurement" (or measurements) should be taken on an outcome; that is, how best do we "mathematize" the outcome (which, for the anthropologist, might be a skull)? These measurement problems are most difficult and can only be answered by getting involved in a practical project. For (1), we need, through repeated observations (called sampling), to estimate these probabilities or "percentages." For example, what percentage of newborn girls in the University of Iowa Hospital weigh less than 7 pounds. Here a newborn baby girl is the outcome, and we have measured her one way (by weight); but obviously there are many other ways of measuring her. If we let X be the weight in pounds, we are interested in the probability $P(X < 7)$ and we can only estimate this by repeated observations. One obvious way of estimating this is by use of the relative frequency of $\{X < 7\}$ after a number of observations. If additional assumptions can be made, we will study, in this text, other ways of estimating this probability. It is this latter aspect with which mathematical statistics is concerned. That is, if we assume certain models, we find that the theory of statistics can explain how best to draw conclusions or make predictions. Now the construction of such a model does require some knowledge of probability, and most theories of probabilities are based on the concept of sets (or events). Accordingly, a basic review of the algebra of sets is given in Section 1.2.

One final remark should be made. In many instances, it is clear exactly what function X the experimenter wants to define on the sample space. For example, the caster in a dice game is concerned about the sum of the spots, say X, that are up on the pair of dice. Hence, we go directly to the space of X and sometimes even call this the sample space S, if there is no confusion. After all, in the dice game, the caster is directly concerned only with the probabilities associated with X. Hence, the reader can, in many instances, think of the space of X as being the sample space.

Exercises

1.1–1. In each of the following random experiments describe the sample space S. Use your intuition or any experience you may have had to assign a value to the probability p of each of the events A.

(a) The toss of an unbiased coin where the event A is heads.

(b) The cast of an honest die where the event A occurs if we observe a three, four, five, or six.

(c) The draw of a card from an ordinary deck of playing cards where the event A is a club.

(d) The choice of a point from a square with opposite vertices (0, 0) and (1, 1) where the event A occurs if the sum of the coordinates of the point is less than 3/4.

1.1–2. Describe the sample space for each of the following experiments.

(a) Toss a coin seven times and observe the number of heads.

(b) Toss a coin five times and observe the sequence of heads and tails.

(c) Observe the number of tosses of a coin until the first head appears.

(d) Draw five cards at random from a standard deck of cards and record each card in that five-card hand (order of drawing is not important).

1.1–3. If a disk two inches in diameter is thrown at random on a tiled floor, where each tile is a square with sides four inches in length, assign a probability to the event that the disk will land entirely on one tile.

1.1–4. Divide a line segment into two parts by selecting a point at random. Assign a probability to the event that the larger segment is at least two times longer than the shorter segment.

1.1–5. Let the interval $[-r, r]$ be the base of a semicircle. If a point is selected at random from this interval, assign a probability to the event that the length of the perpendicular segment from this point to the semicircle is less than $r/2$.

1.1–6. Consider the sequence of heads (H) and tails (T) if an unbiased coin is flipped four times.

(a) List the 16 points in the sample space.

If X equals the number of observed heads, list which of these sample points correspond to

(b) $\{X = 3\}$, (c) $\{0 \leq X \leq 1\}$.

1.1–7. Let X equal the number of observed heads in two flips of an unbiased coin. If each point in the original sample space $S = \{\text{HH, HT, TH, TT}\}$ has probability 1/4, assign values to

(a) $P(X = 0)$, (b) $P(X = 2)$,

(c) $P(X \leq 1)$.

1.1–8. Let the random variable W equal a number selected at random from the closed interval from zero to one, that is $[0, 1]$. Describe the sample space S of W. Assign values to

(a) $P(0 \leq W \leq 1/3)$, (b) $P(1/3 \leq W \leq 1)$,

(c) $P(W = 1/3)$, (d) $P(1/2 < W < 5)$.

1.1–9. Each of the numbers 1, 2, 3, 4, and 5 is written on a disk and placed in a hat. Two disks are drawn without replacement from the hat.

(a) List the 10 possible outcomes for this experiment.

(b) If the random variable Y is defined to be the sum of the two drawn numbers and each of the 10 outcomes has probability 1/10, assign values to $P(Y = 3)$, $P(Y = 5)$, and $P(6 \leq Y \leq 8)$.

1.2 Algebra of Sets

Before defining a probability set function in Section 1.3, we give some basic rules and definitions associated with set algebra. In addition, some terminology used in probability will be explained.

The totality of objects under consideration is called the *universal set* and is denoted S. Each object in S is called an *element* of S. If a set A is a collection of elements that are also in S, then A is said to be a *subset* of S. In applications in probability, S will usually denote the *sample space*. An *event* A will be a collection of possible outcomes of the experiment and will be a subset of S. We say that event A *has occurred* if the outcome of the experiment is an element of A. The set or event A may be described by listing all of its elements or by defining the properties that its elements must satisfy.

Example 1.2–1. Let $S = \{1, 2, 3, 4, 5, 6\}$. If A is the subset of S consisting of the even integers, we may write $A = \{2, 4, 6\}$ or $A = \{x: \; x \text{ is even}\}$. In order to emphasize that x is also in S we could write

$$A = \{x: \quad x \text{ is in } S \text{ and } x \text{ is even}\}.$$

When a is an element in A, we write $a \in A$. When a is not an element in A, we write $a \notin A$. So, in Example 1.2–1, we have $2 \in A$ and $3 \notin A$. If every element of a set A is also an element in a set B, then A is a *subset* of B. We write $A \subset B$. In probability, if event B occurs whenever event A occurs, then $A \subset B$. The two sets A and B are equal, $A = B$, if $A \subset B$ and $B \subset A$. Note that it is always true that $A \subset A$ and $A \subset S$, where S is the universal set. We denote the subset that contains no elements by $\varnothing$. This set is called the *null* or *empty* set. For all sets A, $\varnothing \subset A$.

The set of elements in either A or B or possibly in both A and B is called the *union* of A and B and is denoted $A \cup B$. The set of elements in both A and B is called the *intersection* of A and B and is denoted $A \cap B$. The *complement* of a set A is the set of elements in the universal set S that are not in the set A and is denoted A'. In probability, if A and B are two events, the event that at least one of the two events has occurred is denoted by $A \cup B$, or the event that both events have occurred is denoted by $A \cap B$. The event that A has not occurred is denoted by A', and the event that A has not occurred but B has occurred is denoted by $A' \cap B$. If $A \cap B = \varnothing$, we say that A and B are *mutually exclusive*.

The operations of union and intersection may be extended to more than two sets. Let $A_1, A_2, \ldots, A_n$ be a finite collection of sets. Then the *union*

$$A_1 \cup A_2 \cup \cdots \cup A_n = \bigcup_{k=1}^{n} A_k$$

is the set of all elements that belong to at least one A_k, $\; k = 1, 2, \ldots, n$. The *intersection*

$$A_1 \cap A_2 \cap \cdots \cap A_n = \bigcap_{k=1}^{n} A_k$$

is the set of all elements that belong to every A_k, $\; k = 1, 2, \ldots, n$. Similarly, let $A_1, A_2, \ldots, A_n, \ldots$ be a denumerable collection of sets. Then x belongs to the *union*

$$A_1 \cup A_2 \cup A_3 \cup \cdots = \bigcup_{k=1}^{\infty} A_k$$

if x belongs to at least one A_k, $k = 1, 2, 3, \ldots$. Also x belongs to the *intersection*

$$A_1 \cap A_2 \cap A_3 \cap \cdots = \bigcap_{k=1}^{\infty} A_k$$

if x belongs to every A_k, $k = 1, 2, 3, \ldots$.

Example 1.2–2. Let S be the set of positive real numbers less than or equal to 6. Thus $S = \{x: \; 0 < x \le 6\}$. Let $A = \{x: \; 1 \le x \le 3\}$, $B = \{x: \; 2 \le x \le 6\}$, $C = \{x: \; 3 \le x < 5\}$, and $D = \{x: \; 0 < x < 2\}$. Then

$$\begin{aligned}
A \cup B &= \{x: \; 1 \le x \le 6\},\\
B \cup D &= S,\\
B \cap D &= \varnothing,\\
A \cap B &= \{x: \; 2 \le x \le 3\},\\
B \cap C &= C,\\
A' &= \{x: \; 0 < x < 1 \text{ or } 3 < x \le 6\},\\
B' &= \{x: \; 0 < x < 2\} = D.
\end{aligned}$$

Also

$$A \cup C \cup D = \{x: \; 0 < x < 5\}$$

and

$$A \cap B \cap C = \{x: \; x = 3\}.$$

Example 1.2–3. Let

$$A_k = \left\{x: \; \frac{10}{k+1} \le x \le 10\right\}, \qquad k = 1, 2, 3, \ldots.$$

Then

$$\bigcup_{k=1}^{\infty} A_k = \{x: \; 0 < x \le 10\};$$

note that the number zero is not in the union since it is not in one of the sets $A_1, A_2, A_3, \ldots$. Of course,

$$\bigcap_{k=1}^{\infty} A_k = \{x: \; 5 \le x \le 10\} = A_1$$

since $A_1 \subset A_k$, $k = 1, 2, 3, \ldots$.

A convenient way to illustrate operations on sets is with a *Venn* diagram. In Figure 1.2–1 the universal set S is represented by the rectangle and its interior and the subsets of S by the points enclosed by the circles. The sets under consideration are the shaded regions.

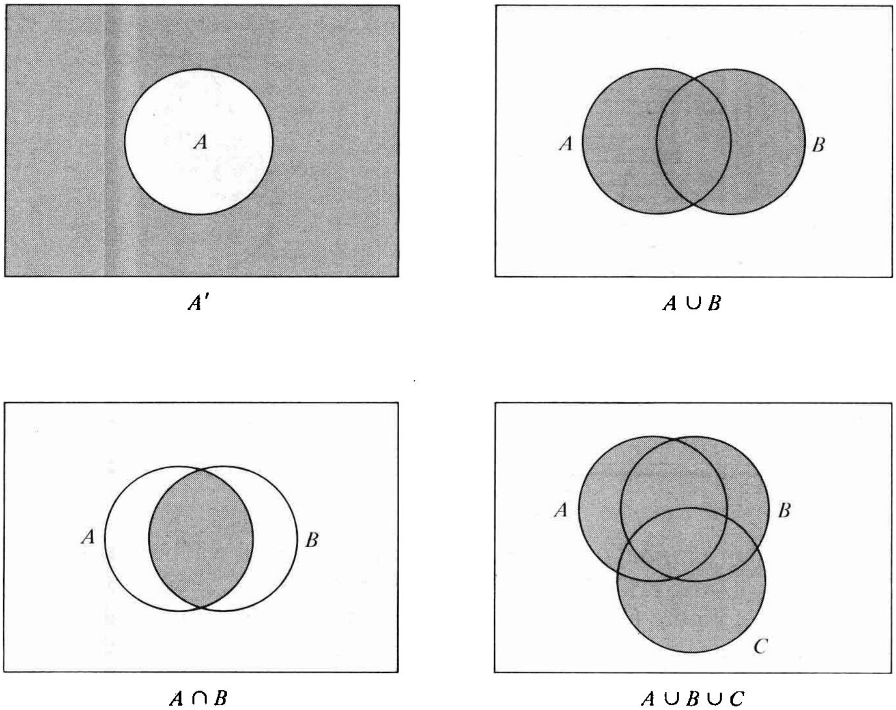

FIGURE 1.2–1

Set operations satisfy several properties. For example, if A, B, and C are subsets of S, we have the following:

Commutative laws:

$$A \cup B = B \cup A$$
$$A \cap B = B \cap A.$$

Associative laws:

$$(A \cup B) \cup C = A \cup (B \cup C)$$
$$(A \cap B) \cap C = A \cap (B \cap C).$$

Distributive laws:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

De Morgan's laws:

$$(A \cup B)' = A' \cap B'$$
$$(A \cap B)' = A' \cup B'.$$