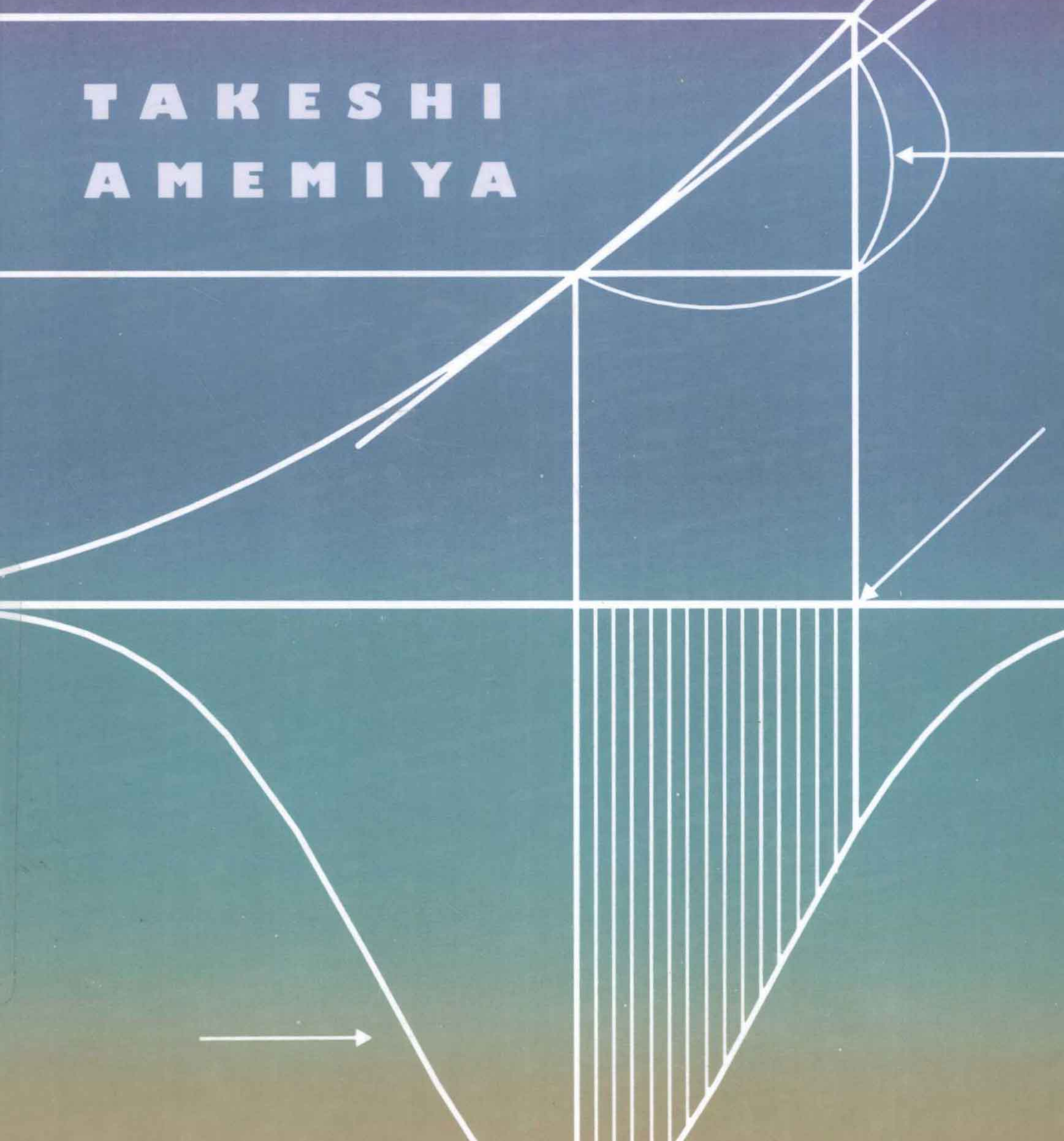
INTRODUCTION TO STATISTICS AND ECONOMETRICS

TAKESHI AMEMIYA



INTRODUCTION TO STATISTICS AND ECONOMETRICS

TAKESHI AMEMIYA

Harvard University Press
Cambridge, Massachusetts
London, England

Printed in the United States of America

This book has been digitally reprinted. The content remains identical to that of previous printings.

Library of Congress Cataloging-in-Publication Data

Amemiya, Takeshi.
Introduction to statistics and econometrics / Takeshi Amemiya.
p. cm.
Includes bibliographical references and index.
ISBN 0-674-46225-4 (alk. paper)
1. Econometrics. 2. Statistics. I. Title.
HB139.A513 1993
330'.01'5195—dc20

93-3777
CIP

knowledge of the basic facts about random variables than to have a superficial knowledge of the latest techniques.

I also believe that students should be trained to question the validity and reasonableness of conventional statistical techniques. Therefore, I give a thorough analysis of the problem of choosing estimators, including a comparison of various criteria for ranking estimators. I also present a critical evaluation of the classical method of hypothesis testing, especially in the realistic case of testing a composite null against a composite alternative. In discussing these issues as well as other problematic areas of classical statistics, I frequently have recourse to Bayesian statistics. I do so not because I believe it is superior (in fact, this book is written mainly from the classical point of view) but because it provides a pedagogically useful framework for consideration of many fundamental issues in statistical inference.

Chapter 10 presents the bivariate classical regression model in the conventional summation notation. Chapter 11 is a brief introduction to matrix analysis. By studying it in earnest, the reader should be able to understand Chapters 12 and 13 as well as the brief sections in Chapters 5 and 9 that use matrix notation. Chapter 12 gives the multiple classical regression model in matrix notation. In Chapters 10 and 12 the concepts and the methods studied in Chapters 1 through 9 in the framework of the i.i.d. (independent and identically distributed) sample are extended to the regression model. Finally, in Chapter 13, I discuss various generalizations of the classical regression model (Sections 13.1 through 13.4) and certain other statistical models extensively used in econometrics and other social science applications (13.5 through 13.7). The first part of the chapter is a quick overview of the topics. The second part, which discusses qualitative response models, censored and truncated regression models, and duration models, is a more extensive introduction to these important subjects.

Chapters 10 through 13 can be taught in the semester after the semester that covers Chapters 1 through 9. Under this plan, the material in Sections 13.1 through 13.4 needs to be supplemented by additional readings. Alternatively, for students with less background, Chapters 1 through 12 may be taught in a year, and Chapter 13 studied independently. At Stanford about half of the students who finish a year-long course in statistics and econometrics go on to take a year's course in advanced econometrics, for which I use my *Advanced Econometrics* (Harvard University Press, 1985).

It is expected that those who complete the present textbook will be able to understand my advanced textbook.

I am grateful to Gene Savin, Peter Robinson, and James Powell, who read all or part of the manuscript and gave me valuable comments. I am also indebted to my students Fumihiro Goto and Dongseok Kim for carefully checking the entire manuscript for typographical and more substantial errors. I alone, however, take responsibility for the remaining errors. Dongseok Kim also prepared all the figures in the book. I also thank Michael Aronson, general editor at Harvard University Press, for constant encouragement and guidance, and Elizabeth Gretz and Vivian Wheeler for carefully checking the manuscript and suggesting numerous stylistic changes that considerably enhanced its readability.

I dedicate this book to my wife, Yoshiko, who for over twenty years has made a steadfast effort to bridge the gap between two cultures. Her work, though perhaps not conspicuous in the short run, will, I am sure, have a long-lasting effect.

CONTENTS

1 INTRODUCTION

1.1 WHAT IS PROBABILITY?

As a common word in everyday usage, *probability* expresses a degree of belief a person has about an event or statement by a number between zero and one. Probability also has a philosophical meaning, which will not be discussed here. The two major schools of statistical inference—the Bayesian and the classical—hold two different interpretations of probability. The *Bayesian* (after Thomas Bayes, an eighteenth-century English clergyman and probabilist) interpretation of probability is essentially that of everyday usage. The *classical* school refers to an approach that originated at the beginning of the twentieth century under the leadership of R. A. Fisher and is still prevalent. The classical statistician uses the word *probability* only for an event which can be repeated, and interprets it as the limit of the empirical frequency of the event as the number of repetitions increases indefinitely. For example, suppose we toss a coin n times, and a head comes up r times. The classical statistician interprets the probability of heads as a limit (in the sense that will be defined later) of the empirical frequency r/n as n goes to infinity. Because a coin cannot be tossed infinitely many times, we will never know this probability exactly and can only guess (or estimate) it.

To consider the difference between the two interpretations, examine the following three events or statements:

(1) A head comes up when we toss a particular coin.

(2) Atlantis, described by Plato, actually existed.

(3) The probability of obtaining heads when we toss a particular coin is ½.

A Bayesian can talk about the probability of any one of these events or statements; a classical statistician can do so only for the event (1), because only (1) is concerned with a repeatable event. Note that (1) is sometimes true and sometimes false as it is repeatedly observed, whereas statement (2) or (3) is either true or false as it deals with a particular thing—one of a kind. It may be argued that a frequency interpretation of (2) is possible to the extent that some of Plato's assertions have been proved true by a later study and some false. But in that case we are considering any assertion of Plato's, rather than the particular one regarding Atlantis.

As we shall see in later chapters, these two interpretations of probability lead to two different methods of statistical inference. Although in this book I present mainly the classical method, I will present Bayesian method whenever I believe it offers more attractive solutions. The two methods are complementary, and different situations call for different methods.

1.2 WHAT IS STATISTICS?

In our everyday life we must continuously make decisions in the face of uncertainty, and in making decisions it is useful for us to know the probability of certain events. For example, before deciding to gamble, we would want to know the probability of winning. We want to know the probability of rain when we decide whether or not to take an umbrella in the morning. In determining the discount rate, the Federal Reserve Board needs to assess the probabilistic impact of a change in the rate on the unemployment rate and on inflation. It is advisable to determine these probabilities in a reasonable way; otherwise we will lose in the long run, although in the short run we may be lucky and avoid the consequences of a haphazard decision. A reasonable way to determine a probability should take into account the past record of an event in question or, whenever possible, the results of a deliberate experiment.

We are ready for our first working definition of statistics: *Statistics is the science of assigning a probability to an event on the basis of experiments.*

Consider estimating the probability of heads by tossing a particular coin many times. Most people will think it reasonable to use the ratio of heads over tosses as an estimate. In statistics we study whether it is indeed reasonable and, if so, in what sense.

Tossing a coin with the probability of heads equal to p is identical to choosing a ball at random from a box that contains two types of balls, one of which corresponds to heads and the other to tails, with p being the proportion of heads balls. The statistician regards every event whose outcome is unknown to be like drawing a ball at random from a box that contains various types of balls in certain proportions.

For example, consider the question of whether or not cigarette smoking is associated with lung cancer. First, we need to paraphrase the question to make it more readily accessible to statistical analysis. One way is to ask, What is the probability that a person who smokes more than ten cigarettes a day will contract lung cancer? (This may not be the optimal way, but we choose it for the sake of illustration.) To apply the box-ball analogy to this example, we should imagine a box that contains balls, corresponding to cigarette smokers; some of the balls have lung cancer marked on them and the rest do not. Drawing a ball at random corresponds to choosing a cigarette smoker at random and observing him until he dies to see whether or not he contracts lung cancer. (Such an experiment would be a costly one. If we asked a related but different question—what is the probability that a man who died of lung cancer was a cigarette smoker?—the experiment would be simpler.)

This example differs from the example of coin tossing in that in coin tossing we create our own sample, whereas in this example it is as though God (or a god) has tossed a coin and we simply observe the outcome. This is not an essential difference. Its only significance is that we can toss a coin as many times as we wish, whereas in the present example the statistician must work with whatever sample God has provided. In the physical sciences we are often able to conduct our own experiments, but in economics or other behavioral sciences we must often work with a limited sample, which may require specific tools of analysis.

A statistician looks at the world as full of balls that have been drawn by God and examines the balls in order to estimate the characteristics ("proportion") of boxes from which the balls have been drawn. This mode of thinking is indispensable for a statistician. Thus we state a second working definition of statistics: *Statistics is the science of observing data and making inferences about the characteristics of a random mechanism that has generated the data.*

Coin tossing is an example of a random mechanism whose outcomes are objects called heads and tails. In order to facilitate mathematical analysis, the statistician assigns numbers to objects: for example, 1 to heads and 0 to tails. A random mechanism whose outcomes are real numbers is called a *random variable.* The random mechanism whose outcome is the height (measured in feet) of a Stanford student is another random variable. The first is called a *discrete random variable,* and the second, a *continuous random variable* (assuming hypothetically that height can be measured to an infinite degree of accuracy). A discrete random variable is characterized by the probabilities of the outcomes. The characteristics of a continuous random variable are captured by a *density function,* which is defined in such a way that the probability of any interval is equal to the area under the density function over that interval. We use the term *probability distribution* as a broader concept which refers to either a set of discrete probabilities or a density function. Now we can compose a third and final definition: *Statistics is the science of estimating the probability distribution of a random variable on the basis of repeated observations drawn from the same random variable.*

2 PROBABILITY

2.1 INTRODUCTION

In this chapter we shall define probability mathematically and learn how to calculate the probability of a complex event when the probabilities of simple events are given. For example, what is the probability that a head comes up twice in a row when we toss an unbiased coin? We shall learn that the answer is $\frac{1}{4}$. As a more complicated example, what is the probability that a student will be accepted by at least one graduate school if she applies to ten schools for each of which the probability of acceptance is 0.1? The answer is $1 - 0.9^{10} \cong 0.65$. (The answer is derived under the assumption that the ten schools make independent decisions.) Or what is the probability a person will win a game in tennis if the probability of his winning a point is p? The answer is

$$p^4\{1 + 4(1 - p) + 10(1 - p)^2 + 20p(1 - p)^3/[1 - 2p(1 - p)]\}.$$

For example, if $p = 0.51$, the formula gives 0.525.

In these calculations we have not engaged in any statistical inference. Probability is a subject which can be studied independently of statistics; it forms the foundation of statistics.

2.2 AXIOMS OF PROBABILITY

Definitions of a few commonly used terms follow. These terms inevitably remain vague until they are illustrated; see Examples 2.2.1 and 2.2.2.

Sample space. The set of all the possible outcomes of an experiment.

Event. A subset of the sample space.

Simple event. An event which cannot be a union of other events.

Composite event. An event which is not a simple event.

EXAMPLE 2.2.1

Experiment: Tossing a coin twice.

Sample space: {HH, HT, TH, TT}.

The event that a head occurs at least once: HH ∪ HT ∪ TH.

EXAMPLE 2.2.2

Experiment: Reading the temperature (F) at Stanford at noon on October 1.

Sample space: Real interval (0, 100).

Events of interest are intervals contained in the above interval.

A probability is a nonnegative number we assign to every event. The axioms of probability are the rules we agree to follow when we assign probabilities.

Axioms of Probability

(1) $P(A) \geq 0$ for any event A.

(2) $P(S) = 1$, where S is the sample space.

(3) If $\{A_i\}$, $i = 1, 2, \ldots$, are mutually exclusive (that is, $A_i \cap A_j = \varnothing$ for all $i \neq j$), then $P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$.

The first two rules are reasonable and consistent with the everyday use of the word probability. The third rule is consistent with the frequency interpretation of probability, for relative frequency follows the same rule. If, at the roll of a die, A is the event that the die shows 1 and B the event that it shows 2, the relative frequency of $A \cup B$ (either 1 or 2) is clearly the sum of the relative frequencies of A and B. We want probability to follow the same rule.

When the sample space is discrete, as in Example 2.2.1, it is possible to assign probability to every event (that is, every possible subset of the sample space) in a way which is consistent with the probability axioms. When the sample space is continuous, however, as in Example 2.2.2, it is not possible to do so. In such a case we restrict our attention to a smaller class of events to which we can assign probabilities in a manner consistent with the axioms. For example, the class of all the intervals contained in (0, 100) and their unions satisfies the condition. In the subsequent discussion we shall implicitly be dealing with such a class. The reader who wishes to study this problem is advised to consult a book on the theory of probability, such as Chung (1974).

2.3 COUNTING TECHNIQUES

2.3.1 Simple Events with Equal Probabilities

Axiom (3) suggests that often the easiest way to calculate the probability of a composite event is to sum the probabilities of all the simple events that constitute it. The calculation is especially easy when the sample space consists of a finite number of simple events with equal probabilities, a situation which often occurs in practice. Let $n(A)$ be the number of the simple events contained in subset A of the sample space S. Then we have

$$P(A) = \frac{n(A)}{n(S)} .$$

Two examples of this rule follow.

EXAMPLE 2.3.1 What is the probability that an even number will show in a roll of a fair die?

We have $n(S) = 6$; $A = \{2, 4, 6\}$ and hence $n(A) = 3$. Therefore, $P(A) = 0.5$.

EXAMPLE 2.3.2 A pair of fair dice are rolled once. Compute the probability that the sum of the two numbers is equal to each of the integers 2 through 12.

Let the ordered pair (i, j) represent the event that i shows on the first die and j on the second. Then $S = \{(i, j) \mid i, j = 1, 2, \ldots, 6\}$, so that $n(S) = 36$. We have

$$n(i + j = 2) = n[(1, 1)] = 1,$$

$$n(i + j = 3) = n[(1, 2), (2, 1)] = 2,$$

$$n(i + j = 4) = n[(1, 3), (3, 1), (2, 2)] = 3,$$

and so on. See Exercise 2.

2.3.2 Permutations and Combinations

The formulae for the numbers of permutations and combinations are useful for counting the number of simple events in many practical problems.

DEFINITION 2.3.1 The number of *permutations* of taking r elements from n elements is the number of distinct ordered sets consisting of r distinct elements which can be formed out of a set of n distinct elements and is denoted by P_r^n.

For example, the permutations of taking two numbers from the three numbers 1, 2, and 3 are (1, 2), (1, 3), (2, 1), (2, 3), (3, 1), (3, 2); therefore, we have $P_2^3 = 6$.

THEOREM 2.3.1 $P_r^n = n!/(n - r)!$, where $n!$ reads n factorial and denotes $n(n - 1)(n - 2) \cdots 2 \cdot 1$. (We define $0! = 1$.)

Proof. In the first position we can choose any one of n elements and in the second position $n - 1$ and so on, and finally in the rth position we can choose one of $n - r + 1$ elements. Therefore, the number of permutations is the product of all these numbers. ❑

DEFINITION 2.3.2 The number of *combinations* of taking r elements from n elements is the number of distinct sets consisting of r distinct elements which can be formed out of a set of n distinct elements and is denoted by C_r^n.

Note that the order of the elements matters in permutation but not in combination. Thus in the example of taking two numbers from three, (1, 2) and (2, 1) make the same combination.