HANDBOOK OF MATHEMATICAL ECONOMICS

VOLUME III

Edited by

KENNETH J. ARROW

and

MICHAEL D. INTRILIGATOR

HANDBOOK OF MATHEMATICAL ECONOMICS

VOLUME III

Edited by

KENNETH J. ARROW
Stanford University

and

MICHAEL D. INTRILIGATOR
University of California, Los Angeles

1986

NORTH-HOLLAND
AMSTERDAM • NEW YORK • OXFORD • TOKYO

ISBN North-Holland for this volume 0 444 86128 9
ISBN North-Holland for this set 0 444 86054 1

Publishers
ELSEVIER SCIENCE PUBLISHERS B.V.
P.O. Box 1991
1000 BZ Amsterdam
The Netherlands

Sole distributors for the USA and Canada
ELSEVIER SCIENCE PUBLISHING COMPANY INC
52 Vanderbilt Avenue
New York, NY 10017, USA

Library of Congress Cataloging in Publication Data

Main entry under title:

Handbook of mathematical economics.

(Handbooks in economics ; bk. 1)
Includes bibliographies and index.
1. Economics, Mathematical--Addresses, essays, lectures. I. Arrow, Kenneth Joseph, 1921- II. Intriligator, Michael D. III. Series.
HB135.H357 330'.01'51 81-2820
ISBN 0-444-86054-1 (U.S. : set)

PRINTED IN THE NETHERLANDS

PREFACE

Although the first appearance of statistics as a science dates to the seventeenth century, it was not until much later that the papers of F. Wilcoxon (1945) and H.B. Mann and D.R. Whitney (1947) provided the initial theoretical foundation for the discipline of nonparametric statistics. In the years that followed, further development took place at a fast pace. However, when a field matures this rapidly, it is often difficult for textbooks to keep pace. Although early texts on the methodology were completed by Kendall (1948) and Siegel (1956), and on the theory by Fraser (1957), it was not until the late 1960s and early 1970s that the newly developed general theory again reached the settled confines of a book. During that period methodological texts appeared by Conover (1971), Hollander and Wolfe (1973), Lehmann (1975), and Gibbons (1976). In addition, nonparametric theory books at the intermediate level were written by Noether (1967), Hájek (1969), and Gibbons (1971), while more advanced texts were written by Hájek and Šidák (1967), and Puri and Sen (1971).

The level of this text is intermediate. The reader is expected to have completed an introductory mathematical statistics sequence at the level of Hogg and Craig (1978), Mood, Graybill, and Boes (1974), or Dudewicz (1976), for example. This, of course, entails that the reader be familiar with the basic concepts of statistical inference and have a good knowledge of advanced calculus. The goal of our approach is to bring the reader to a basic understanding of the concepts and theory that are important to the field of nonparametric statistics. An equally important goal to us, however, is to develop this theory without sacrificing the intuitive flavor that is prevalent in most of the early work in nonparametric statistics and, indeed, remains an important force in current nonparametric usage and research. Thus, for example, we devote the entire Chapter 2 to the introduction of the most important basic approaches that lead to nonparametric distribution-free tests of hypotheses. The emphasis there is on the nature of nonparametric procedures, and the sign, Wilcoxon signed rank, and Mann–Whitney–Wilcoxon rank sum tests are presented as illustrations of the important counting and ranking techniques. In addition, we choose to introduce the reader to the theory behind the nonparametric procedures by way of U-statistics. This again serves to maintain the intuitive nature, and provides a lead-in to the more comprehensive theory of general linear rank statistics included later.

We touch on the most prominent hypothesis testing settings, but detailed coverage is provided only for the one- and two-sample problems. However, the book is not devoted solely to tests. Confidence intervals and point estimation are discussed in detail in Chapters 6 and 7, respectively. We feel their inclusion is vital since they play an increasingly important role in modern nonparametric development.

The text is not intended to be a comprehensive presentation of the existing nonparametric methodology. Such a need is satisfied by the texts by Conover (1971), Hollander and Wolfe (1973), Lehmann (1975), and Gibbons (1976). The purpose of the text is instead to present the important mathematical statistics tools that are fundamental to the development of nonparametric statistics. Thus the text is organized around these tools rather than around methodological topics. We emphasize (1) techniques for making a test distribution-free, (2) U-statistics, (3) asymptotic efficiency, (4) the Hodges–Lehmann technique for creating a confidence interval and a point estimator from a test, and (5) linear rank statistics, to name just a few. We have also included some currently developing areas, such as M-estimation, adaptive testing procedures, rank-like techniques, and partially sequential sampling schemes. We have not attempted to be complete on any of these topics—actually, completeness is impossible in such active arenas—but we have tried to acquaint the reader with the fundamental ideas. Our goal is to present a more comprehensive description of the basic tools of nonparametrics than is available in current intermediate level texts, while remaining well below the mathematical level of Hájek and Šidák (1967), and Puri and Sen (1971).

The material in the book can be used for either a one-semester or a two-quarter course. The underlying core for either of these is found in Chapters 1–10, inclusively. Any remaining time could be devoted to topics from Chapters 11 and 12 most appropriate to an individual instructor. In addition, Sections 3.5, 3.6, 4.3, 6.3, 7.3, and 7.4 could be omitted without affecting the coverage of later material. An instructor who wished to emphasize testing problems could skip Chapters 6 and 7 and include them later as time permitted. This would not destroy the flow of the topics.

The Appendix at the end of this book includes (1) a list of the major distribution types mentioned in the text, (2) a discussion of our representation for integrals, and some theorems for manipulating them, (3) statements of important results in mathematical statistics that are used but not proved in text, and (4) statements of selected mathematical results applied in text.

The reference system within this book lists chapter, section, and quantity within the section, in that order. Thus a citation of equation (11.4.8) refers to the eighth numbered quantity in Section 4 of Chapter 11. Within a given section all quantities assigned reference numbers are numbered sequentially. Thus equations, theorems, lemmas, etc., are included in the same numbering scheme. The only exception is for the exercises, which are numbered separately in a sequential fashion within each section. References to the Appendix use the letter A in place of the chapter number.

We have benefited from the help of numerous people. Michael Fligner read almost all of the manuscript and made many valuable suggestions. Tim Robertson also made a number of important contributions. Kathy Altobelli gave portions of the manuscript a hard line-by-line reading, and her suggestions led to an improved manuscript. James Kepner, Ping Lu, and Tie Hua Ng also contributed suggestions and exercise solutions.

We were able to collaborate on this text during the academic year 1976–1977 while Ronald Randles was a Visiting Associate Professor at Ohio State University. We wish to thank The University of Iowa and The Ohio State University, and particularly Robert V. Hogg and D. Ransom Whitney, for making this arrangement possible.

Preliminary versions of this book were used in five separate offerings of the theoretical nonparametric statistics sequences at The Ohio State University and The University of Iowa. We are appreciative of the contributions of the many students who corrected errors and cleared up ambiguities.

This book was written and typed at The Ohio State University and The University of Iowa. We thank Kathleen Narcross (who did most of the typing), Amy Bernstein, Ada Burns, Rozanne Huff, Barbara Meeder, Carolyn Randles, and Gloria Rudolph for their excellent preparation of various versions of the manuscript.

It was our pleasure to work with Beatrice Shube, editor in Wiley-Interscience, on this project.

Ronald H. Randles
Douglas A. Wolfe

Iowa City, Iowa
Columbus, Ohio
January 1979

CONTENTS

1 PRELIMINARIES

1.1. Introduction

The rudiments of nonparametric statistics originate in the latter part of the nineteenth century. Some early contributions are cited in a survey paper by H. Scheffé (1943). However, it was not until the pioneering papers of F. Wilcoxon (1945) and H. B. Mann and D. R. Whitney (1947) that the theoretical foundations of nonparametric statistics began to be assembled. In the ensuing years the advances in this field have been of major proportions, providing in many cases important techniques in mathematical statistics indigenous to nonparametric theory but finding usage in parametric settings as well. The concept of asymptotic relative efficiency proposed by Pitman (1948) and the development of the theory of U-statistics initiated by Hoeffding (1948) are prime examples of such contributions.

The purpose of this text is to introduce the important theoretical foundations of nonparametric statistics, both classical and current. Consequently, the text is organized around them rather than around methodological topics. We are not able to present a complete accounting of these subjects, since many are currently active research arenas, but we have tried to provide the reader with the fundamental concepts and tools.

We begin with a brief description of some of the notation and terms which are used throughout the text. All random variables will assume values on the real line (which is denoted by R). The distribution of a random variable, say X, is most often described in terms of its **cumulative distribution function** (abbreviated **c.d.f.**), say $F(\cdot)$, which is defined to be

$$F(x) = P[X \leqslant x], \qquad -\infty < x < \infty.$$

In this text we emphasize **continuous random variables**, each of which is characterized by a nonnegative **density function**, say $f(x)$, that satisfies

$\int_{-\infty}^{\infty} f(x)\,dx = 1$. The density is related to the c.d.f. in that $F'(x)$ exists and equals $f(x)$ at all but at most a countable number of x-values. We then have

$$P[a < X < b] = P[a \leq X \leq b] = \int_a^b f(x)\,dx, \text{ for any constants } a < b.$$

The **support** of a continuous random variable is defined to be the closure of the set

$$\{x \mid f(x) > 0\}.$$

On occasion we discuss a **discrete random variable**, say, X. Its distribution is described by the **probability function**

$$p(x) = P[X = x], \qquad -\infty < x < \infty,$$

which is related to its c.d.f. through

$$F(x) = P[X \leq x] = \sum_{t \leq x} p(t),$$

where the latter sum is over all $t \leq x$ such that $p(t) > 0$. The **support** of this discrete random variable is defined as

$$\{t \mid p(t) > 0\}.$$

If $X_1, \ldots, X_n$ denotes a random sample from some underlying population, we say that these random variables are **independent and identically distributed** (abbreviated **i.i.d.**). In such contexts, the **sample mean** refers to

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

and the **sample variance** is

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

A vector of n random variables (not necessarily i.i.d.) is represented by $\mathbf{X} = (X_1, \ldots, X_n)$.

It is also useful to note the definition of an indicator function that occurs repeatedly in this text, namely,

$$\Psi(t) = 0, \qquad \text{if } t \leq 0,$$
$$= 1. \qquad \text{if } t > 0.$$

The symbol $\Psi(\cdot)$ is reserved for this purpose. (The only exception occurs in Section 7.4.) One final notation that is widely used in this text is that for the **greatest integer function**. For any real number x we let $[\![x]\!]$ represent the greatest integer less than or equal to x.

1.2. Order Statistics

In many statistical analyses the information from a random sample is utilized through the ordered values of the sample. These ordered sample observations are referred to as **order statistics**, and they play a fundamental role in the development of nonparametric statistics, both for hypothesis tests and for estimators. In this section we develop some of the basic properties of order statistics that are used throughout the rest of the book. For a more detailed accounting, see Sarhan and Greenberg (1962) or David (1970).

The Joint Distribution

Let the continuous random variables $X_1,\ldots,X_n$ denote a random sample from a population with c.d.f. $F(x)$ and density $f(x)$. Let $X_{(i)}$, $i=1,\ldots,n$, be the ith smallest of these sample observations. We refer to $X_{(1)}<\cdots<X_{(n)}$ as the **order statistics** for the random sample $X_1,\ldots,X_n$. Unlike the Xs themselves, the order statistics are neither mutually independent nor identically distributed. We obtain their joint distribution by a change of variable.

Theorem 1.2.1. *Let $X_{(1)}<\cdots<X_{(n)}$ be the order statistics for a random sample of continuous random variables from a distribution with c.d.f. $F(x)$ and density $f(x)$. The joint density for the order statistics is then*

$$
\begin{aligned}
g(x_{(1)},\ldots,x_{(n)}) &= n!\prod_{i=1}^{n} f(x_{(i)}), \qquad -\infty < x_{(1)} < \cdots < x_{(n)} < \infty \\
&= 0, \qquad \text{elsewhere.} \qquad\qquad (1.2.2)
\end{aligned}
$$

Proof: We provide the structure of the proof for general n, with a detailed illustration for the specific case $n=2$. Define the sets $A=\{(x_1,\ldots,x_n)|-\infty<x_i<\infty,\ x_i\neq x_j \text{ for } i\neq j\}$ and $B=\{(x_{(1)},\ldots,x_{(n)})|-\infty<x_{(1)}<\cdots<x_{(n)}<\infty\}$. The transformation defining the order statistics maps A onto B, but not in a one-to-one fashion, since each of the $n!$ permutations of the observed values yields the same value for the order statistics. Thus, for example, when $n=2$, $(x_1,x_2)=(1.4,6.9)$ and $(x_1,x_2)=(6.9,1.4)$ both yield $(x_{(1)},x_{(2)})=(1.4,6.9)$. If we partition A into $n!$ subsets, each one corresponding to a particular ordering of the sample observations, we see that the order statistics transformation now maps each of these partitioned sets onto B in a one-to-one fashion. For illustration, when $n=2$, A is partitioned into A_1 and A_2, where $A_1=\{(x_1,x_2)|-\infty<x_1<x_2<\infty\}$ and $A_2=\{(x_1,x_2)|-\infty<x_2<x_1<\infty\}$. On A_1 the order statistics set $x_1=x_{(1)}$ and $x_2=x_{(2)}$. The absolute value of the Jacobian of this one-to-one transformation is

$$|J_1| = \left\|\begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}\right\| = 1.$$

On A_2 the order statistics set $x_1=x_{(2)}$ and $x_2=x_{(1)}$, yielding the Jacobian with absolute value

$$|J_2| = \left\|\begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix}\right\| = 1.$$

For general n, it follows similarly that the Jacobian of each of the one-to-one transformations from one of the $n!$ partitions of A onto B has an absolute value equal to 1. The joint density of the order statistics is then simply the sum of the contributions from each of the partitions. Specifically, when $n=2$ we have

$$\begin{aligned} g(x_{(1)},x_{(2)}) &= f(x_{(1)})f(x_{(2)})|J_1| + f(x_{(2)})f(x_{(1)})|J_2| \\ &= 2f(x_{(1)})f(x_{(2)}), \qquad -\infty < x_{(1)} < x_{(2)} < \infty. \end{aligned}$$

For general n, the joint density in (1.2.2) follows from the fact that the contribution of each of the $n!$ partitions is $\prod_{i=1}^{n} f(x_{(i)})$. ■

A special case of the joint density of the order statistics that is important to nonparametric statistics corresponds to an underlying uniform distribution on the interval (0, 1). For this case, the joint density (1.2.2) of the order statistics is

$$\begin{aligned} g(x_{(1)},\ldots,x_{(n)}) &= n!, \qquad 0 < x_{(1)} < \cdots < x_{(n)} < 1, \\ &= 0, \qquad \text{elsewhere.} \end{aligned} \tag{1.2.3}$$

Marginal Distributions

Although the marginal distribution of a single order statistic $X_{(j)}$, $1<j<n$, can be obtained directly by the proper integration of (1.2.2) (see Exercise 1.2.1), we will develop these formulas by a more indirect, but instructive,

approach based on binomial probabilities. (An expression for the joint density of two order statistics is given in Exercise 1.2.2.)

Theorem 1.2.4. *The marginal density for the jth order statistic* $X_{(j)}$, $1 \leq j \leq n$, *under the conditions of Theorem* 1.2.1 *is*

$$g_j(t) = \frac{n!}{(j-1)!(n-j)!}[F(t)]^{j-1}[1-F(t)]^{n-j}f(t), \qquad -\infty < t < \infty. \tag{1.2.5}$$

Proof: Let $G_j(t)$ be the distribution function for $X_{(j)}$. Then, for any t,

$$\begin{aligned} G_j(t) &= P(X_{(j)} \leq t) = P(\text{at least } j\ X\text{s are} \leq t) \\ &= \sum_{i=j}^{n} P(\text{exactly } i\ X\text{s are} \leq t) \\ &= \sum_{i=j}^{n} \binom{n}{i}[F(t)]^i[1-F(t)]^{n-i}, \end{aligned} \tag{1.2.6}$$

since whether any particular X is $\leq t$ is a Bernoulli event with probability $F(t)$ and the n Bernoulli events in question are mutually independent. From the well-known relation between binomial sums and the incomplete beta function (see Exercise 1.2.8), we can write

$$G_j(t) = \frac{n!}{(j-1)!(n-j)!}\int_0^{F(t)} x^{j-1}(1-x)^{n-j}\,dx. \tag{1.2.7}$$

Letting $H(\cdot)$ be the c.d.f. for a beta distribution with parameters $\alpha = j$ and $\beta = n - j + 1$, we have $G_j(t) = H(F(t))$. Differentiating $G_j(t) = H(F(t))$ with respect to t via the chain rule yields $g_j(t) = h(F(t))f(t)$, the desired expression for the marginal density of $X_{(j)}$. ■

When computing cumulative probabilities for $X_{(j)}$, we note that we can use either cumulative binomial sums as in (1.2.6) or tables of the incomplete beta function in (1.2.7). Moreover, although the density form in (1.2.5) depends upon the underlying distribution being continuous, the two cumulative distribution expressions are valid for either continuous or discrete variables.

The Probability Integral Transformation and Uniform Order Statistics

As mentioned previously, the uniform distribution on (0, 1) plays a special role in nonparametric statistics. This is primarily due to a result referred to as the probability integral transformation. For a random variable X with c.d.f. $F(x)$, we define the **inverse distribution function** $F^{-1}(\cdot)$ by

$$F^{-1}(y) = \inf\{x \mid F(x) \geq y\}, \qquad 0 < y < 1. \tag{1.2.8}$$

We note that if $F(x)$ is strictly increasing between 0 and 1, then there is only one x such that $F(x)=y$. In this case the infimum is unnecessary, and $F^{-1}(y)=x$ without ambiguity.

Suppose there is some x such that $F(x)=y$. Since $F(\cdot)$ is continuous from the right, $F(F^{-1}(y))=y$. In particular, this shows that if $F(\cdot)$ is continuous, then $F(F^{-1}(y))=y$ for every y satisfying $0<y<1$. However, if $F(\cdot)$ is the c.d.f. of a discrete distribution, then for a given y there may be no x for which $F(x)=y$. In such cases $F^{-1}(y)$ is the smallest x yielding an $F(\cdot)$ value larger than y, and hence, in general, we have the relationship

$$y \leq F(F^{-1}(y)) \qquad \text{for } 0 < y < 1.$$

Theorem 1.2.9. (Probability Integral Transformation) *Let X be a continuous random variable with distribution function $F(x)$. The random variable $Y = F(X)$ has a uniform distribution on* $(0, 1)$.

Proof: As noted above, since $F(\cdot)$ is continuous, $F(F^{-1}(y))=y$ for $0<y<1$. Using the monotonicity of $F(\cdot)$ we see that $\{X \leq F^{-1}(y)\}$ implies $\{F(X) \leq F(F^{-1}(y))=y\}$. Also,

$$\{F(X) \leq y\} = \{X \leq F^{-1}(y)\} \cup \{X > F^{-1}(y) \text{ and } F(X)=y\}.$$

The continuous distribution of X implies that $P(F(X)=y)=0$. Thus

$$P(F(X) \leq y) = P(X \leq F^{-1}(y)).$$

Let $H(y)$ be the distribution function for Y. Since Y assumes values only in $[0, 1]$, we know that

$$\begin{aligned} H(y) &= 0, \qquad y < 0 \\ &= 1, \qquad y \geq 1. \end{aligned} \tag{1.2.10}$$

Also,

$$\begin{aligned} H(y) &= P(Y \leq y) = P(F(X) \leq y) = P(X \leq F^{-1}(y)) \\ &= F(F^{-1}(y)) = y, \qquad 0 < y < 1. \end{aligned} \tag{1.2.11}$$

Using (1.2.10), (1.2.11) and the nondecreasing nature of $H(\cdot)$, we see that $H(y)$ is the distribution function for a uniform distribution on $(0, 1)$, as desired. ■

We now use this result in connection with the earlier results on order statistics. Let $X_{(1)} < \cdots < X_{(n)}$ be the order statistics for a random sample from a continuous distribution with c.d.f. $F(x)$. Then, in view of Theorem 1.2.9, $F(X_{(1)}) < \cdots < F(X_{(n)})$ are distributed as the order statistics from a uniform distribution on $(0, 1)$. Hence the joint density of $V_i = F(X_{(i)})$, $i = 1, \ldots, n$, is given in (1.2.3), and the marginal density for each V_j, $1 < j < n$, follows from (1.2.5) and has form

$$\begin{aligned} g_j(t) &= \frac{n!}{(j-1)!(n-j)!} t^{j-1}(1-t)^{n-j}, \qquad 0 < t < 1 \\ &= 0, \qquad \text{elsewhere.} \end{aligned} \tag{1.2.12}$$

Thus V_j has the beta distribution with parameters $\alpha = j$ and $\beta = n - j + 1$.

The moments for this beta distribution are used in later chapters and are easily calculated from (1.2.12). For any positive number r, we have

$$\begin{aligned} E(V_j^r) &= \frac{n!}{(j-1)!(n-j)!} \int_0^1 t^{r+j-1}(1-t)^{n-j}\,dt \\ &= \frac{n!\Gamma(r+j)}{(j-1)!\Gamma(n+r+1)} \int_0^1 \frac{\Gamma(n+r+1)}{\Gamma(r+j)(n-j)!} t^{r+j-1}(1-t)^{n-j}\,dt \\ &= \frac{n!\Gamma(r+j)}{(j-1)!\Gamma(n+r+1)}, \end{aligned} \tag{1.2.13}$$

where $\Gamma(k) = (k-1)!$ whenever k is a positive integer. Thus when V_j is the jth order statistic from a uniform distribution

$$E(V_j) = \frac{j}{n+1}$$

and (1.2.14)

$$\text{Var}(V_j) = \frac{j(n-j+1)}{(n+1)^2(n+2)}.$$

Expected values for order statistics from distributions other than the uniform are also useful in nonparametric testing procedures, but for many important distributions no simple expressions can be given for these expectations. (One notable exception is the exponential distribution—see Exercise 1.2.7.) In particular, such an explicit formula is not available for the normal distribution if n is larger than 3. However, because of the importance of the expectations of order statistics from a standard normal distribution (i.e., mean zero and variance one), these values have been calculated via numerical integration and have been tabulated by several people, including Harter (1961).

Another important result that, in a sense, is a converse to the probability integral transformation, is stated in the following theorem, which, unlike Theorem 1.2.9, is valid for c.d.f.'s $F(\cdot)$ of both discrete and continuous random variables.

Theorem 1.2.15. *Let U be a random variable with a uniform distribution on* $(0,1)$. *Define $X=F^{-1}(U)$, where $F^{-1}(\cdot)$ denotes the inverse of the distribution function $F(x)$, as given in expression* (1.2.8). *Then X has c.d.f. $F(x)$.*

Proof: Using the monotonicity of $F(\cdot)$, we see that $\{F^{-1}(U)\leq x\}$ implies $\{U\leq F(F^{-1}(U))\leq F(x)\}$. Also, the definition of $F^{-1}(\cdot)$ shows that $\{U\leq F(x)\}$ implies $\{F^{-1}(U)\leq x\}$, and hence

$$P[U\leq F(x)] = P[F^{-1}(U)\leq x].$$

Let $G(x)$ denote the c.d.f. of X. Then

$$\begin{aligned} G(x) &= P[X\leq x] \\ &= P[F^{-1}(U)\leq x] \\ &= P[U\leq F(x)] = F(x). \end{aligned}$$

A random number generator on a computer produces values $U_1,\ldots,U_n$ that are approximately independent uniform $(0,1)$ variates. If the c.d.f. $F(\cdot)$ has an inverse function $F^{-1}(\cdot)$, then $F^{-1}(U_1),\ldots,F^{-1}(U_n)$ will be approximately distributed as independent variables with c.d.f. $F(x)$. Note also that $F^{-1}(U_{(1)}),\ldots,F^{-1}(U_{(n)})$ behave like the order statistics for a sample of size n from $F(\cdot)$. Thus to generate order statistics from any specific distribution, we can first generate ordered uniform variates and then transform them using the inverse function of the population c.d.f. In Section 1 of Chapter 4 we illustrate the use of Theorem 1.2.15 in a Monte Carlo study of the relative performances of several test statistics.