Edward B. Manoukian

# Modern Concepts and Theorems of Mathematical Statistics

Edward B. Manoukian

# Modern Concepts and Theorems of Mathematical Statistics

Springer-Verlag
New York Berlin Heidelberg Tokyo

Edward B. Manoukian
Department of National Defence
Royal Military College of Canada
Kingston, Ontario K7L 2W3
Canada

AMS Classification: 62-01

Library of Congress Cataloging in Publication Data
Manoukian, Edward B.
Modern concepts and theorems of mathematical statistics.
(Springer series in statistics)
Bibliography: p.
Includes index.
1. Mathematical statistics. I. Title. II. Series.
QA276.M333 1985 519.5 85-14686



Typeset by Asco Trade Typesetting Ltd., Hong Kong.
Printed and bound by R. R. Donnelley & Sons, Harrisonburg, Virginia.
Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-96186-0 Springer-Verlag New York Berlin Heidelberg Tokyo
ISBN 3-540-96186-0 Springer-Verlag Berlin Heidelberg New York Tokyo

# Preface

With the rapid progress and development of mathematical statistical methods, it is becoming more and more important for the student, the instructor, and the researcher in this field to have at their disposal a quick, comprehensive, and compact reference source on a very wide range of the field of modern mathematical statistics. This book is an attempt to fulfill this need and is encyclopedic in nature. It is a useful reference for almost every learner involved with mathematical statistics at any level, and may supplement any textbook on the subject. As the primary audience of this book, we have in mind the beginning busy graduate student who finds it difficult to master basic modern concepts by an examination of a limited number of existing textbooks. To make the book more accessible to a wide range of readers I have kept the mathematical language at a level suitable for those who have had only an introductory undergraduate course on probability and statistics, and basic courses in calculus and linear algebra. No sacrifice, however, is made to dispense with rigor. In stating theorems I have not always done so under the weakest possible conditions. This allows the reader to readily verify if such conditions are indeed satisfied in most applications given in modern graduate courses without being lost in extra unnecessary mathematical intricacies. The book is not a mere dictionary of mathematical statistical terms. It is also expository in nature, providing examples and putting emphasis on theorems, limit theorems, comparison of different statistical procedures, and statistical distributions. The various topics are covered in appropriate details to give the reader enough confidence in himself (herself) which will then allow him (her) to consult the references given in the Bibliography for proofs, more details, and more applications. At the end of various sections of the book references are given where proofs and/or further details may be found. No attempt is made here to supply historical details on who did what, when. Accordingly, I apologize to any colleague whose name is not found in the list of references or whose name may not appear attached to a theorem or to a statistical procedure. All that should matter to the reader is to obtain quick and precise information on the technical aspects he or she is seeking. To benefit as much as possible from the book, it is advised to consult first the Contents on a given topic, then the Subject Index, and then the section on Notations. Both the Contents and the Subject Index are quite elaborate.

We hope this book will fill a gap, which we feel does exist, and will provide a useful reference to all those concerned with mathematical statistics.

E. B. M.

# Contents

PART 1

# FUNDAMENTALS OF MATHEMATICAL STATISTICS

CHAPTER 1

# Basic Definitions, Concepts, Results, and Theorems

## §1.1. Probability Concepts

The collection of elements under investigation is called the *population*. An experiment for which the outcome cannot *a priori* be determined, but is known to be one of a set of given possible outcomes, is called a *random experiment*. [Here it is assumed that if the same experiment is repeated any number of times, then the outcome each time is again one of the outcomes in the initial set of possible outcomes.] The set of all possible outcomes of a random experiment is called the *sample space* and will be denoted by $\mathscr{S}$. An element $\omega \in \mathscr{S}$ is called a *sample point*.

A family $\mathscr{A}$ of subsets of $\mathscr{S}$ is called a *sigma-field* if: (i) $A \in \mathscr{A}$, then $A^c \equiv \mathscr{S} - A \in \mathscr{A}$; (ii) $A_1, A_2, \ldots$ are pairwise disjoint sets in $\mathscr{A}$, that is, $A_i \cap A_j = \varnothing$ (null set) for all $i \neq j$, then $\bigcup_{i=1}^{\infty} A_i \in \mathscr{A}$. An element $A$ in $\mathscr{A}$ is called an *event*. An element of the form $\{\omega\}$ is called an *elementary event*.

With each event $A \in \mathscr{A}$, we associate a nonnegative function $P(A)$, called a probability (or probability measure), which satisfies: (i) $0 \leqslant P(A) \leqslant 1$; (ii) $P(\mathscr{S}) = 1$; (iii) for pairwise disjoint events $A_1, A_2, \ldots$ in $\mathscr{A}$, $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$. In particular, we note that $\mathscr{S}$ and $\phi$ are in $\mathscr{A}$. $\mathscr{S}$ is called the *sure event* and $\phi (P(\phi) = 0)$ is called the *impossible event*. The triple $(\mathscr{S}, \mathscr{A}, P)$, consisting of the sample space $\mathscr{S}$, the sigma-field $\mathscr{A}$, and the probability measure $P$, is called a *probability space*.

$A$ and $B$ are said to be *independent events*, if $P(A \cap B) = P(A)P(B)$. In general, $A_1, \ldots, A_n$ are said to be *mutually independent* if for any subset $\{i_1, \ldots, i_t\}$ of $\{1, \ldots, n\}$ with any $2 \leqslant t \leqslant n$, $P(A_{i_1} \cap \cdots \cap A_{i_t}) = P(A_{i_1}) \ldots P(A_{i_t})$.

The *conditional probability* of $A$ *given* $B$, is defined by $P(A|B) = P(A \cap B)/P(B)$, $A, B \in \mathscr{A}$ for $P(B) \neq 0$.

$\{A_1, A_2, \ldots\}$ is said to constitute a *complete set of events* if $A_1, A_2, \ldots$ $(\in \mathscr{A})$ are mutually disjoint and $\bigcup_{i=1}^{\infty} A_i = \mathscr{S}$. The set $\{A_1, A_2, \ldots\}$ is also called a partition of the set $\mathscr{S}$.

**Poincaré's Additive Theorem.** *For any two events $A_1, A_2 \in \mathscr{A}$, $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$. The generalization of this formula for an arbitrary finite number of events is immediate. For example,*

$$P(A_1 \cup A_2 \cup A_3) = \sum_{i=1}^{3} P(A_i) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3).$$

**Bonferroni Inequality.** *Let $A_1, \ldots, A_k \in \mathscr{A}$, then*

$$P(A_1 \cap \cdots \cap A_k) \geqslant 1 - \sum_{i=1}^{k} P(A_i^c).$$

**Total Probability Theorem.** *Let $\{A_1, A_2, \ldots\}$ constitute a complete set of events (a partition of $\mathscr{S}$) in $\mathscr{A}$. Then for any $B \in \mathscr{A}$,*

$$P(B) = \sum_i P(B|A_i)P(A_i).$$

**Bayes' Formula.** *For $P(B) \neq 0$, we may write:*

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_i P(B|A_i)P(A_i)}.$$

The *Borel field* $\mathscr{B}$ in $R$ (real line) is the smallest sigma-field containing all open intervals $\{x: a < x < b\}$ in $R$. The Borel field $\mathscr{B}^k$ in $R^k$ ($k$-dimensional Euclidean space) is the smallest sigma-field containing all open $k$-rectangles $\{(x_1, \ldots, x_k): a_i < x_i < b_i, i = 1, \ldots, k\}$ in $R^k$.

A *random variable* $X$ is a function from $\mathscr{S}$ to $R$ such that the set $\{\omega: X(\omega) \in B\}$ is in $\mathscr{A}$ for every $B \in \mathscr{B}$. A vector $\mathbf{X} = (X_1, \ldots, X_k)'$ is called a *random k-vector* if $X_1, \ldots, X_k$ are random variables. [We note, conversely, that if $\{\omega: \mathbf{X}(\omega) \in B\}$ is in $\mathscr{A}$ for every $B \in \mathscr{B}^k$, then $X_1, \ldots, X_k$ are random variables.] The probability of the event $\{\omega: \mathbf{X}(\omega) \in B\}$ will be denoted by $P[\mathbf{X} \in B]$, and is called the *probability distribution* of $\mathbf{X}$.

$\mathbf{X}$ is said to be a *discrete random variable* if there exists a countable set of points $\{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$, $\mathbf{x}_i \in R^k$, $i = 1, 2, \ldots$, such that $\sum_i P[\mathbf{X} = \mathbf{x}_i] = 1$, and $P[\mathbf{X} = \mathbf{x}_i] \equiv f(\mathbf{x}_i)$ is called the *probability mass* function (or just probability mass) of $\mathbf{X}$. In particular, if $X$ is a random variable having values of the form $a + bk$, where $a$, and $b > 0$ are fixed real numbers, and $k$ runs through the set of values $\{0, 1, 2, \ldots, n\}$ or $\{0, \pm 1, \pm 2, \ldots, \pm n\}$ for $n$ finite *or* infinite, then $X$ is said to have a *lattice* probability *distribution*. For a random vector $\mathbf{X}$ if

$P[\mathbf{X} = \mathbf{x}] = 0$ for every $\mathbf{x} \in R^k$, then $\mathbf{X}$ is called a continuous random $k$-vector. If, in addition, there exists a nonnegative function $f$ from $(R^k, \mathscr{B}^k)$ to $(R, \mathscr{B})$ such that

$$P[\mathbf{X} \in B] = \int_B f(\mathbf{x})\, d\mathbf{x} \qquad \left(\text{and } P[\mathbf{X} \in R^k] = \int_{R^k} f(\mathbf{x})\, d\mathbf{x} = 1\right),$$

then $\mathbf{X}$ is called an *absolutely continuous random k-vector.* $f(\mathbf{x})$ is called the *probability density* function of $\mathbf{X}$. Throughout the bulk of this work, by a continuous random $k$-vector it will mean an absolutely continuous random $k$-vector.

Consider the interval $\{x \leqslant \alpha\}$ in $\mathscr{B}$. Then $P[X \leqslant a]$, denoted by $F(a)$, is called the *cumulative distribution function* (or just the distribution) of the random variable $X$ in question. Similarly, $P[X_1 \leqslant a_1, \ldots, X_k \leqslant a_k] \equiv F(\mathbf{a})$, $\mathbf{a} = (a_1, \ldots, a_k)'$, will be called the *joint distribution* of $X_1, \ldots, X_k$.

*Properties of $F(x)$*

(i) $0 \leqslant F(x) \leqslant 1$.

(ii) $F(x) \leqslant F(x')$ for $x \leqslant x'$.

(iii) $F(x + \varepsilon) = F(x)$, $\varepsilon \to +0$, that is, $F(x)$ is continuous from the right.

(iv) $$\lim_{x \to \infty} F(x) \equiv F(+\infty) = 1, \qquad \lim_{x \to -\infty} F(x) = F(-\infty) = 0.$$

(v) If $X$ is a continuous random variable, then

$$\frac{d}{dx} F(x) = f(x), \qquad dF(x) = f(x)\, dx, \qquad F(x) = \int_{-\infty}^{x} f(x')\, dx'.$$

(vi) If $X$ is a discrete random variable taking values $x_1 < x_2 < \cdots$, then $F(x_i) - F(x_{i-1}) = P[X = x_i]$.

*Properties of $F(x_1, x_2)$*

(i) $0 \leqslant F(x_1, x_2) \leqslant 1$.

(ii) For $x_1 \leqslant x_1'$, $x_2 \leqslant x_2'$, $F(x_1', x_2') + F(x_1, x_2) - F(x_1', x_2) - F(x_1, x_2') \geqslant 0$.

(iii) $F(x_1 + \varepsilon, x_2) = F(x_1, x_2)$, $F(x_1, x_2 + \varepsilon) = F(x_1, x_2)$ for $\varepsilon \to +0$.

(iv) $F(-\infty, x_2) = 0$, $F(x_1, -\infty) = 0$ for all $x_1, x_2 \in R$.

(v) $F(x_1, \infty) \equiv F_{X_1}(x_1)$ is called the *marginal* distribution of $X_1$, and $F(\infty, x_2) \equiv F_{X_2}(x_2)$ is called the marginal distribution of $X_2$. $X_1$ and $X_2$ are said to be independent if $F(x_1, x_2) = F_{X_1}(x_1)F_{X_2}(x_2)$ for all $x_1, x_2 \in R$, and with an obvious generalization for more than two random variables: $X_1, \ldots, X_n$.

(vi) Let $\mathbf{X} = (X_1, X_2)'$ be a random vector such that $X_1$ and $X_2$ are discrete random variables taking, respectively, values $x_{11}, x_{12}, \ldots; x_{21}, x_{22}, \ldots$. We set $P[\{X_1 = x_{1i}\} \cap \{X_2 = x_{2j}\}] = p_{ij}$. Then $\sum_i \sum_j p_{ij} = 1$ and

$p_{i.} = \sum_j p_{ij}$, $p_{.j} = \sum_i p_{ij}$ denote the marginal probability mass functions of $X_1$ and $X_2$, respectively. We note that the conditional probability $P[\{X_1 = x_{1i}\}|\{X_2 = x_{2j}\}]$ may be written as $p_{ij}/p_{.j}$. Also formally,

$$E[X_1] = \sum_i x_{1i}p_{i.} \equiv \mu_{X_1}, \qquad \sigma^2(X_1) = \sum_i x_{1i}^2 p_{i.} - \mu_{X_1}^2,$$

$$\mathrm{Cov}(X_1, X_2) = \sum_i \sum_j x_{1i}x_{2j}p_{ij} - \mu_{X_1}\mu_{X_2}.$$

(vii) Let $\mathbf{X} = (X_1, X_2)'$ be a continuous random vector. Then

$$F(a, b) = \int_{-\infty}^{b}\int_{-\infty}^{a} f(x_1, x_2)\,dx_1\,dx_2, \qquad \frac{\partial}{\partial x_1}\frac{\partial}{\partial x_2}F(x_1, x_2) = f(x_1, x_2).$$

Also

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2)\,dx_2 \quad \text{and} \quad f_{X_2}(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2)\,dx_1,$$

are called the marginal densities of $X_1$ and $X_2$. The conditional probability density of $X_1$ given that $X_2 = x_2$ is defined by $f(x_1|x_2) = f(x_1, x_2)/f_{X_2}(x_2)$ for $f_{X_2}(x_2) \neq 0$. We note that formally,

$$E[X_1] = \int_{-\infty}^{\infty} x_1 f_{X_1}(x_1)\,dx \equiv \mu_{X_1}, \quad \sigma^2(X_1) = \int_{-\infty}^{\infty} x_1^2 f_{X_1}(x_1)\,dx_1 - \mu_{X_1}^2,$$

$$E[X_1|X_2] = \int_{-\infty}^{\infty} x_1 f(x_1|x_2)\,dx_1,$$

$$P[a \leqslant X_1 \leqslant b|X_2 = x_2] = \int_a^b dx_1 f(x_1|x_2),$$

$$\mathrm{Cov}(X_1, X_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 f(x_1, x_2)\,dx_1\,dx_2 - \mu_{X_1}\mu_{X_2}.$$

**Double Expectation.** *If* $E[|X|] < \infty$, *then* $E[E[X|Y]] = E[X]$ *for two random variables.*

A number which maximizes the probability density or the probability mass function of a random variable $X$ is called a *mode* of the distribution of $X$. A number $Q_\alpha$, called the $\alpha$th quantile $(0 < \alpha < 1)$ of the distribution of a random variable, is defined by $P[X < Q_\alpha] \leqslant \alpha \leqslant P[X \leqslant Q_\alpha]$. In particular, for $\alpha = 0.5$, $Q_{0.5}$ is called a *median* of the distribution in question.

A random variable $X_1$ with distribution $F_1(x)$ is said to be *stochastically smaller* than a random variable $X_2$ with distribution $F_2(x)$ if $F_1(x) \geqslant F_2(x)$ for all $x$, and with strict inequality holding for at least one $x$.

Consider two discrete random variables $X_1$, $X_2$, if $P[X_1 \neq X_2] = 0$, then we say that $X_1 = X_2$ with *probability one*. Similarly, if $g(x)$ is a continuous function with the exception on a set $S$, such that $P[X \in S] = 0$, then $g(x)$ is said to be continuous with probability one. Quite generally, if a certain relation holds identically with the exception for points in a set $S$ with probability equal to zero, then the relation is said to hold $P$-almost everywhere or to hold with probability one.

[Cf., Bickel and Doksum (1977), Billingsley (1979), Burrill (1972), Cramér (1974), Fisz (1980), Fourgeaud and Fuchs (1967), Gnedenko (1966), Milton and Tsokos (1976), Rényi (1970), Roussas (1973), Tucker (1967), Wilks (1962).]

## §1.2. Random Samples

*Finite Population.* Suppose every sample of size $n$ that may selected from the population has the same probability of being selected, then any such a sample is called a *random sample.*

*Infinite Population.* A set $\{X_1, \ldots, X_n\}$ of independent random variables each having the same population distribution is called a *random sample* of size $n$. Throughout by a sample we mean a random sample.

A random variable $g(X_1, \ldots, X_n)$ which is a function of independent identically distributed random variables such that if the sample yields $X_1 = x_1, \ldots, X_n = x_n$, then $g(x_1, \ldots, x_n)$ is uniquely determined is called a *statistic.* [We note that a statistic cannot depend on any unknown parameters which may characterize the distribution of the population.]

Let $X_1, \ldots, X_n$ be independent identically distributed random variables. Then $Y_k$ is called the $k$th *order statistic* of $X_1, \ldots, X_n$ if $Y_k$ is the $k$th smallest of the $X_1, \ldots, X_n$ observations. We note that $Y_1 \leqslant Y_2 \leqslant \cdots \leqslant Y_n$, where $Y_1 = \min_i X_i$, $Y_n = \max_i X_i$. The random variables $Y_1, \ldots, Y_n$ are called the *order statistics* of the sample $X_1, \ldots, X_n$. $Y_n - Y_1$ is called the *sample range* and $Y_1$, $Y_n$ are called the *extremes* of the sample. If $X_i \neq X_j$ for all $i \neq j$ in $(1, \ldots, n)$, then the *rank* $R_i$ of $X_i$ is defined by the number $N(i)$ of the $X_j$ less than $X_i$ plus one: $R_i = N(i) + 1$. If $X_{1i}, \ldots, X_{mi}$ denote the $X_j$ equal to $X_i$ (including $X_i$ itself), then the *average rank* or *mid rank* of $X_i$ is defined by $\sum_{j=1}^{m} (N(i) + j)/m$, where $N(i)$ is the number of the $X_j$ less than $X_i$.

A random variable $Q_\alpha$ is an $\alpha$th *quantile* of the sample $X_1, \ldots, X_n$, if

$$[\text{number of } X_j < Q_\alpha]/n \leqslant \alpha \leqslant [\text{number of } X_j \leqslant Q_\alpha]/n.$$

In particular, for $\alpha = 0.5$, $Q_{0.5} \equiv M$ is called the *median* of the sample. An equivalent definition for the median is the following: $M = [Y_{n/2} + Y_{(n+2)/2}]/2$ if $n = \text{even}$, and $M = Y_{(n+1)/2}$ if $n = \text{odd}$, where $Y_1, \ldots, Y_n$ are the order statistics of the sample $X_1, \ldots, X_n$.

The function $\hat{F}_n(x) = [\text{number of } X_i \leqslant x]/n$ is called the sample distribution or empirical distribution.