

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

© 1986 by Springer-Verlag New York Inc.

All rights reserved. No part of this book may be translated or reproduced in any form without written permission from Springer-Verlag, 175 Fifth Avenue, New York, New York 10010, U.S.A.

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

Some Notations

xv

PART I

Fundamentals of Mathematical Statistics

CHAPTER 1

Basic Definitions, Concepts, Results, and Theorems	3
§1.1. Probability Concepts	3
§1.2. Random Samples	7
§1.3. Moments	7
§1.4. Some Inequalities Involving Probabilities and Moments	10
§1.5. Characteristic Functions	12
§1.6. Moment Generating Functions	14
§1.7. Determination of a Distribution from Its Moments	14
§1.8. Probability Integral Transform	14
§1.9. Unbiased and Asymptotically Unbiased Estimators	14
§1.10. Uniformly Minimum Variance Unbiased Estimators	15
§1.11. Consistency of an Estimator	15
§1.12. M -Estimators	15
§1.13. L -Estimators and the α -Trimmed Mean	16
§1.14. R -Estimators	16
§1.15. Hodges-Lehmann Estimator	16
§1.16. U -Statistics	18
§1.17. Cramér-Rao-Fr�chet Lower Bound	21
§1.18. Sufficient Statistics	22
§1.19. Fisher-Neyman Factorization Theorem for Sufficient Statistics	23
§1.20. Rao-Blackwell Theorem	24
§1.21. Completeness of Statistics and Their Families of Distributions	25
§1.22. Theorem on Completeness of Statistics with Sampling from the Exponential Family	25

§1.23. Lehmann–Scheffé Uniqueness Theorem	27
§1.24. Efficiency, Relative Efficiency, and Asymptotic Efficiency of Estimators	27
§1.25. Estimation by the Method of Moments	28
§1.26. Confidence Intervals	28
§1.27. Tolerance Intervals	30
§1.28. Simple and Composite Hypotheses, Type-I and Type-II Errors, Level of Significance or Size, Power of a Test and Consistency	30
§1.29. Randomized and Nonrandomized Test Functions	31
§1.30. Uniformly Most Powerful (UMP), Most Powerful (MP), Unbiased and Uniformly Most Powerful Unbiased (UMPU) Tests	31
§1.31. Neyman–Pearson Fundamental Lemma	32
§1.32. Monotone Likelihood Ratio Property of Family of Distributions and Related Theorems for UMP and UMPU Tests for Composite Hypotheses	32
§1.33. Locally Most Powerful Tests	34
§1.34. Locally Most Powerful Unbiased Tests	34
§1.35. Likelihood Ratio Test	35
§1.36. Theorems on Unbiasedness of Tests	36
§1.37. Relative Efficiency of Tests	36
§1.38. Sequential Probability Ratio Test (SPRT)	37
§1.39. Bayes and Decision-Theoretic Approach	39
§1.40. The Linear Hypothesis	41
§1.41. The Bootstrap and the Jackknife	50
§1.42. Robustness	54
§1.43. Pitman–Fisher Randomization Methods	59
§1.44. Nonparametric Methods	63

CHAPTER 2

Fundamental Limit Theorems	75
§2.1. Modes of Convergence of Random Variables	75
§2.2. Slutsky's Theorem	76
§2.3. Dominated Convergence Theorem	76
§2.4. Limits and Differentiation Under Expected Values with Respect to a Parameter t	76
§2.5. Helly–Bray Theorem	77
§2.6. Lévy–Cramér Theorem	77
§2.7. Functions of a Sequence of Random Variables	77
§2.8. Weak Laws of Large Numbers	77
§2.9. Strong Laws of Large Numbers	78
§2.10. Berry–Esséen Inequality	78
§2.11. de Moivre–Laplace Theorem	79
§2.12. Lindeberg–Lévy Theorem	79
§2.13. Liapounov Theorem	79
§2.14. Kendall–Rao Theorem	79
§2.15. Limit Theorems for Moments and Functions of Moments	80
§2.16. Edgeworth Expansions	82
§2.17. Quantiles	83
§2.18. Probability Integral Transform with Unknown Location and/or Scale Parameters	83

§2.19. α -Trimmed Mean	84
§2.20. Borel's Theorem	85
§2.21. Glivenko-Cantelli Theorem	85
§2.22. Kolmogorov-Smirnov Limit Theorems	85
§2.23. Chi-Square Test of Fit	86
§2.24. Maximum Likelihood Estimators	88
§2.25. M -Estimators	89
§2.26. Likelihood Ratio Statistic	91
§2.27. On Some Consistency Problems of Tests	91
§2.28. Pitman Asymptotic Efficiency	92
§2.29. Hodges-Lehmann Estimators	93
§2.30. Hoeffding's Theorems for U -Statistics	95
§2.31. Wald-Wolfowitz Theorem	96
§2.32. Chernoff-Savage's for R -Statistics	96
§2.33. Miller's for Jackknife Statistics	97

PART 2

Statistical Distributions

CHAPTER 3

Distributions	101
§3.1. Binomial	101
§3.2. Multinomial	101
§3.3. Geometric	102
§3.4. Pascal Negative Binomial	102
§3.5. Hypergeometric	103
§3.6. Poisson	103
§3.7. Wilcoxon's Null (One-Sample)	103
§3.8. Wilcoxon-(Mann-Whitney)'s Null (Two-Sample)	104
§3.9. Runs	104
§3.10. Pitman-Fisher Randomization (One-Sample)	105
§3.11. Pitman's Permutation Test of the Correlation Coefficient	105
§3.12. Pitman's Randomization (Two-Sample)	106
§3.13. Pitman's Randomization (k -Sample)	107
§3.14. Kolmogorov-Smirnov's Null (One-Sample)	108
§3.15. Kolmogorov-Smirnov's Null (Two-Sample)	109
§3.16. Uniform (Rectangular)	110
§3.17. Triangular	110
§3.18. Pareto	110
§3.19. Exponential	110
§3.20. Erlang and Gamma	111
§3.21. Weibull and Rayleigh	111
§3.22. Beta	112
§3.23. Half-Normal	112
§3.24. Normal (Gauss)	112
§3.25. Cauchy	113
§3.26. Lognormal	113

§3.27. Logistic	113
§3.28. Double-Exponential	114
§3.29. Hyperbolic-Secant	114
§3.30. Slash	114
§3.31. Tukey's Lambda	115
§3.32. Exponential Family	115
§3.33. Exponential Power	116
§3.34. Pearson Types	117
§3.35. Chi-Square χ^2	119
§3.36. Student's T	119
§3.37. Fisher's F	120
§3.38. Noncentral Chi-Square	120
§3.39. Noncentral Student	120
§3.40. Noncentral Fisher's F	121
§3.41. Order Statistics	121
§3.42. Sample Range	121
§3.43. Median of a Sample	122
§3.44. Extremes of a Sample	122
§3.45. Studenized Range	122
§3.46. Probability Integral Transform	123
§3.47. \bar{X} , $\bar{X} - \bar{Y}$	123
§3.48. S_1^2 , S_1^2/S_2^2 , and Bartlett's M	125
§3.49. Bivariate Normal	127
§3.50. Sample Correlation Coefficient	128
§3.51. Multivariate Normal	129
§3.52. Wishart	130
§3.53. Hotelling's T^2	130
§3.54. Dirichlet	131

CHAPTER 4

Some Relations Between Distributions	132
§4.1. Binomial and Binomial	132
§4.2. Binomial and Multinomial	132
§4.3. Binomial and Beta	132
§4.4. Binomial and Fisher's F	133
§4.5. Binomial and Hypergeometric	133
§4.6. Binomial and Poisson	133
§4.7. Binomial and Normal	133
§4.8. Geometric and Pascal	134
§4.9. Beta and Beta	134
§4.10. Beta and Fisher's F	134
§4.11. Beta and Chi-Square	134
§4.12. Beta and Uniform	134
§4.13. Poisson and Poisson	135
§4.14. Poisson and Chi-Square	135
§4.15. Poisson and Exponential	135
§4.16. Poisson and Normal	135
§4.17. Exponential and Exponential	136

§4.18. Exponential and Erlang	136
§4.19. Exponential and Weibull	136
§4.20. Exponential and Uniform	136
§4.21. Cauchy and Normal	136
§4.22. Cauchy and Cauchy	136
§4.23. Normal and Lognormal	137
§4.24. Normal and Normal	137
§4.25. Normal and Chi-Square	137
§4.26. Normal and Multivariate Normal	137
§4.27. Normal and Other Distributions	137
§4.28. Exponential Family and Other Distributions	138
§4.29. Exponential Power and Other Distributions	139
§4.30. Pearson Types and Other Distributions	140
§4.31. Chi-Square and Chi-Square	140
§4.32. Chi-Square and Gamma	140
§4.33. Chi-Square and Fisher's F	140
§4.34. Student, Normal, and Chi-Square	140
§4.35. Student and Cauchy	140
§4.36. Student and Hyperbolic-Secant	141
§4.37. Student and Fisher's F	141
§4.38. Student and Normal	141
§4.39. Student and Beta	141
§4.40. Student and Sample Correlation Coefficient	141
§4.41. Fisher's F and Logistic	141
§4.42. Fisher's F and Fisher's Z -Transform	141
§4.43. Noncentral Chi-Square and Normal	142
§4.44. Noncentral Chi-Square and Noncentral Chi-Square	142
§4.45. Noncentral Student, Normal, and Chi-Square	142
§4.46. Noncentral Fisher's F , Noncentral Chi-Square, and Chi-Square	142
§4.47. Multivariate Normal and Multivariate Normal	142
§4.48. Multivariate Normal and Chi-Square	143
§4.49. Multivariate Normal and Noncentral Chi-Square	143
§4.50. Multivariate Normal and Fisher's F	143
§4.51. Multivariate Normal and Noncentral Fisher's F	144
§4.52. Dirichlet and Dirichlet	144
§4.53. Dirichlet and Beta	144
 Bibliography	 145
Author Index	151
Subject Index	153

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 \mathcal{S} . An element $\omega \in \mathcal{S}$ is called a *sample point*.

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

With each event $A \in \mathcal{A}$, we associate a nonnegative function $P(A)$, called a *probability* (or *probability measure*), which satisfies: (i) $0 \leq P(A) \leq 1$; (ii) $P(\mathcal{S}) = 1$; (iii) for pairwise disjoint events A_1, A_2, \dots in \mathcal{A} , $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$. In particular, we note that \mathcal{S} and \emptyset are in \mathcal{A} . \mathcal{S} is called the *sure event* and \emptyset ($P(\emptyset) = 0$) is called the *impossible event*. The triple $(\mathcal{S}, \mathcal{A}, P)$, consisting of the sample space \mathcal{S} , the sigma-field \mathcal{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, \dots, A_n are said to be *mutually independent* if for any subset $\{i_1, \dots, i_t\}$ of $\{1, \dots, n\}$ with any $2 \leq t \leq n$, $P(A_{i_1} \cap \dots \cap A_{i_t}) = P(A_{i_1}) \dots P(A_{i_t})$.

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

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

Poincaré's Additive Theorem. For any two events $A_1, A_2 \in \mathcal{A}$, $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$. 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 $A_1, \dots, A_k \in \mathcal{A}$, then

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

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

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

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

A random variable X is a function from \mathcal{S} to R such that the set $\{\omega: X(\omega) \in B\}$ is in \mathcal{A} for every $B \in \mathcal{B}$. A vector $\mathbf{X} = (X_1, \dots, X_k)$ is called a random k -vector if X_1, \dots, X_k are random variables. [We note, conversely, that if $\{\omega: \mathbf{X}(\omega) \in B\}$ is in \mathcal{A} for every $B \in \mathcal{B}^k$, then X_1, \dots, 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 $\{x_1, x_2, \dots\}$, $x_i \in R^k$, $i = 1, 2, \dots$, such that $\sum_i P[\mathbf{X} = x_i] = 1$, and $P[\mathbf{X} = x_i] \equiv f(x_i)$ 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 $\{0, 1, 2, \dots, n\}$ or $\{0, \pm 1, \pm 2, \dots, \pm n\}$ for n finite or infinite, then X is said to have a *lattice probability distribution*. For a random vector \mathbf{X} if

$P[X = \mathbf{x}] = 0$ 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, \mathcal{B}^k) to (R, \mathcal{B}) such that

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

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

Properties of $F(x)$

- (i) $0 \leq F(x) \leq 1$.
- (ii) $F(x) \leq F(x')$ for $x \leq x'$.
- (iii) $F(x + \varepsilon) = F(x)$, $\varepsilon \rightarrow +0$, that is, $F(x)$ is continuous from the right.
- (iv) $\lim_{x \rightarrow \infty} F(x) \equiv F(+\infty) = 1$, $\lim_{x \rightarrow -\infty} F(x) = F(-\infty) = 0$.
- (v) If X is a continuous random variable, then

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

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

Properties of $F(x_1, x_2)$

- (i) $0 \leq F(x_1, x_2) \leq 1$.
- (ii) For $x_1 \leq x'_1$, $x_2 \leq x'_2$, $F(x'_1, x'_2) + F(x_1, x_2) - F(x'_1, x_2) - F(x_1, x'_2) \geq 0$.
- (iii) $F(x_1 + \varepsilon, x_2) = F(x_1, x_2)$, $F(x_1, x_2 + \varepsilon) = F(x_1, x_2)$ for $\varepsilon \rightarrow +0$.
- (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, \dots, 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 $x_{11}, x_{12}, \dots; x_{21}, x_{22}, \dots$. 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,

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

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

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

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

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,

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

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

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

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

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_α , called the α th quantile ($0 < \alpha < 1$) of the distribution of a random variable, is defined by $P[X < Q_\alpha] \leq \alpha \leq P[X \leq Q_\alpha]$. 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 $F_1(x) \geq F_2(x)$ 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 be 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 $\{X_1, \dots, X_n\}$ 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 $g(X_1, \dots, X_n)$ which is a function of independent identically distributed random variables such that if the sample yields $X_1 = x_1, \dots, X_n = x_n$, then $g(x_1, \dots, x_n)$ 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, \dots, X_n be independent identically distributed random variables. Then Y_k is called the k th *order statistic* of X_1, \dots, X_n if Y_k is the k th smallest of the X_1, \dots, X_n observations. We note that $Y_1 \leq Y_2 \leq \dots \leq Y_n$, where $Y_1 = \min_i X_i$, $Y_n = \max_i X_i$. The random variables Y_1, \dots, Y_n are called the *order statistics* of the sample X_1, \dots, 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 $(1, \dots, n)$, 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}, \dots, 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_α is an α th *quantile* of the sample X_1, \dots, X_n , if

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

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 is even, and $M = Y_{(n+1)/2}$ if n is odd, where Y_1, \dots, Y_n are the order statistics of the sample X_1, \dots, X_n .

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