Characterization Problems in Mathematical Statistics

A. M. KAGAN, YU. V. LINNIK,
C. RADHAKRISHNA RAO

Characterization Problems in Mathematical Statistics

A. M. KAGAN & YU. V. LINNIK
Steklov Institute of Mathematics, Leningrad

C. RADHAKRISHNA RAO
Indian Statistical Institute, Calcutta

Translated from Russian text by
B. RAMACHANDRAN
Indian Statistical Institute, Calcutta

A Wiley-Interscience Publication

JOHN WILEY & SONS
New York London Sydney Toronto

Library of Congress Cataloging in Publication Data

Kagan, Abram Meerovich.
Characterization problems in mathematical statistics.

(Probability and Mathematical statistics series. Probablity & statistics section)

Translation of Kharakterizatsionnye zadachi matematicheskoi statistiki.

"A Wiley-Interscience publication."
Bibliography: p.

1. Mathematical statistics. I. Linnik, IUrii Vladimirovich, 1915–1972, joint author. II. Rao, Calyampudi Radhakrishna, joint author. III. Title.

QA276.K213 519.5 73–9643

ISBN 0-471-45421-4

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Preface

In many problems of mathematical statistics, conclusions are based on the circumstance that certain special distributions possess important properties which permit the reduction of the original problem to a substantially simpler one.

The natural question as to how fully such a reduction utilizes the special nature of the parent distribution leads to the study of the characteristic properties of the principal distributions of mathematical statistics.

The present work is concerned with the analytical theory of characterization problems and their connection with various areas of mathematical statistics such as the theory of estimation, testing of hypotheses, and sequential analysis. Here are collected together many results (mostly of recent origin) and also formulated problems which appear to the authors to be of interest and importance.

The book is primarily intended for specialists in the areas of probability theory and mathematical statistics. A large part of it should also be accessible to students of advanced courses. It may also prove useful to workers in the area of applications of mathematical statistics.

A. M. Kagan
Yu. V. Linnik (deceased)
C. R. Rao

Contents

CHAPTER 4
Independant Nonlinear Statistics. Random Linear Statistics 101

CHAPTER 5
Regression Problems Connected with Linear Statistics 155

CHAPTER 12

CHAPTER 13

CHAPTER 14

Introduction

The standard procedure for making statistical inferences about a population under study is as follows: first of all, from certain *a priori* considerations, the class of possible parent distributions is fixed. This class determines the class of distributions P_s of the sample s and the class of distributions P_T of any chosen statistic T. The observed value of the statistic T and prior information consisting in the knowledge of the class P_s serve as the basis for statistical inferences concerning the population. It is clear that the choice of T must be coordinated in a definite manner with the class P_s fixed beforehand.

If, for instance, the statistic T is used for testing the hypothesis that the sample s comes from one of the distributions of the class P_s, then it is highly desirable that P_s can be reconstructed from P_T.

Or, if the statistic T possesses some desirable property when the sample s has one of the distributions P_s, then it is natural to enquire: what is the widest class P_s in which this property of the statistic T is preserved?

Let us turn to some examples.

For constructing goodness of fit tests for normal, Poisson, and other distributions, conditional distributions free of nuisance parameters are used. Clearly, tests based on conditional distributions are unsatisfactory, at least from a theoretical point of view, if only because of the lack of one-one correspondence between conditional and parent distributions. Thus it becomes necessary to study distributions which correspond to given conditional distributions for fixed values of some statistics.

It is well-known that if $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables (r.v.'s) following the normal law $N(0, \sigma^2)$, then the statistic $t = \sqrt{n}\,\bar{X}/s$ (in standard notation) follows a Student's t-distribution with $(n-1)$ degrees of freedom. It can be shown that for $n = 2$, the same is true if the density of the r.v.'s has one of the forms

$$\frac{\sqrt{2}}{\pi\sigma\left[1+\left(\frac{x}{\sigma}\right)^4\right]} \quad \text{or} \quad \frac{\sqrt{2}\,x^2}{\sigma^3\left[1+\left(\frac{x}{\sigma}\right)^4\right]}.$$

In fact, the statistic t will have the *t-distribution with one degree of freedom* if the distribution of the r.v.'s X_i belongs to a certain infinite class which includes the normal laws. Therefore, inferences based on the t-distribution are valid not only for normal variables. A complete description of the class of all distributions for which, for $n > 2$, the statistic t has the Student's t-distribution, remains an unsolved problem.

If $X_1, X_2, \ldots, X_n$ are independent and identically distributed according to the law $N(\mu, \sigma^2)$, then the vector

$$\mathbf{Y} = \left(\frac{X_1 - \bar{X}}{s}, \ldots, \frac{X_n - \bar{X}}{s}\right)$$

is uniformly distributed on the unit sphere. Suppose now that there are independent sets of observations $(x_1^{(i)}, \ldots, x_n^{(i)})$ where, within the i-th set, the observations follow a law with d.f. $F[(x - \mu_i)/\sigma_i]$, with μ_i, σ_i unknown. Then it is possible to construct a sequence $\{y^{(i)}\}$ and test the hypothesis that the sequence arises from a population with uniform distribution on the unit sphere. If this hypothesis is accepted, then does it follow that the hypothesis that the d.f. F is a normal law is also to be accepted? Clearly it would, if, from the uniformity of distribution of $\mathbf{Y}$ on the unit sphere, the normality of the r.v.'s X_i follows (concerning this characterization of the normal law, *vide* Section 13.5).

It is well-known that the sample mean $\bar{X}$ is an unbiased estimator with minimum variance for the population mean in samples from a normal population. Is this a property solely of the normal law? It is shown in Chapter 7 that, for any $n \geqq 3$, this is so.

We have so far cited typical examples of problems considered in this book. We shall now briefly describe the contents of the individual chapters.

In Chapter 1, we present general information from the theory of characteristic functions and also some apparatus for solving certain nonlinear differential and functional equations which will be encountered in the book. These equations appear to us to be of interest also from a purely analytical point of view, and some of the results of this chapter appear to be potentially useful for solving other problems.

Chapters 2, 3, and 5 are mostly concerned with characterizations of distributions through stochastic properties of pairs of linear statistics of independent observations. The property of identical distribution is considered in Chapter 2, that of independence in Chapter 3, and that of constancy of regression of one linear statistic on another in Chapter 5. The normal law plays a central role in all these chapters, but, at the same time, it will appear that certain other laws behave, in respect of the properties studied, similarly to the normal law.

The problem of conditions for the identical distribution of two linear statistics was studied by Yu. V. Linnik [81] several years ago. A full description of laws admitting the property of equidistribution was given there, and it was further shown that under certain supplementary conditions this property distinguishes the normal law. The proof of Yu. V. Linnik's was simplified by A. A. Zinger [37], and the latter's results are essentially used in Chapter 2.

The characterization of the normal law through the independence of two linear statistics was originally considered by Bernstein, Darmois, and Skitovich. The original proofs were simplified by Zinger and Linnik [38]. The corresponding theorems for linear statistics in a finite or denumerable number of scalar or vector r.v.'s are given in Chapter 3. Clearly, if in place of $X_1, \ldots, X_n$, we consider $\phi(X_1), \ldots, \phi(X_n)$, where ϕ is a measurable function, and require the independence of two linear functions in the arguments $\phi(X_1), \ldots, \phi(X_n)$, then we obtain a characterization of such distributions of the r.v.'s X_i for which $\phi(X_i)$ is normally distributed. Thus, for any ϕ, we obtain a "two-parameter" family of distributions.

In Chapter 4, we study the independence of linear statistics and also the related problem of independence of linear statistics whose coefficients are themselves r.v.'s. The analysis of this phenomenon requires the solving of some nonlinear integrofunctional equations. For the problem of the independence of a tube statistic with finite basis and the sample mean, the corresponding equation was solved by Anosov [1]. In some cases, we show that the parent distribution is normal; in others, we only establish some analytical properties of the distribution. We may remark that, in this circle of problems, only the simplest have been solved so far.

The problem of constancy of regression of one linear statistic on another arises as a natural generalization of the Kagan-Linnik-Rao theorem [56], which characterizes the normal law by the condition

$$E(\overline{X} \mid X_1 - \overline{X}, \ldots, X_n - \overline{X}) = \text{constant}, n \geqq 3.$$

In a certain sense, constancy of regression is a weaker requirement than independence. But independence is not linked to any condition on the moments, whereas in the regression problem it is necessary to assume beforehand the existence of the first moment. The results of Chapter 5 are based on the works of Rao [130], Ramachandran and Rao [122, 123], and Shimizu [164]. It is interesting to note that the functional equation arising in the regression problem is analogous to the one which arises in the study of the problem of identical distribution of two linear statistics. In special cases, the solutions constitute a subclass of infinitely divisible distributions, which we call "generalized stable laws."

In Chapter 6 we consider the problem of constancy of regression for some nonlinear statistics. Various characterizations of the gamma distribution are obtained here.

In Chapter 7, we study conditions of admissibility and optimality of some estimators commonly used in statistics. The original result (cf. [56]), which stimulated subsequent research in the area, is as follows: let the observations have the form $x_i = \theta + \varepsilon_i$, where θ is an unknown parameter and ε_i are errors of observation. Then, for $n \geqq 3$, the sample mean is an admissible estimator for θ, under quadratic loss, if and only if the ε_i are normal; otherwise it is possible to suggest an estimator better than $\bar{X}$ for θ. This chapter contains many results of this type, relating to various schemes of observation and to various loss-functions. From an analytical standpoint, Chapter 7 borders closely on Chapters 5 and 6. An important result in estimation theory is established that *linear estimators of location parameters are admissible iff the r.v.'s are normally distributed.*

Chapter 8 concerns the exponential family of distributions, which are the only solutions of the functional equation

$$\prod_{i=1}^{n} f(x_i, \theta) = R[T(x_1, \ldots, x_n); \theta] \cdot r(x_1, \ldots, x_n)$$

arising in the theory of sufficient statistics. A general method of solving this equation is described, which is essentially due to Dynkin [26]; other approaches are also presented, applicable in the case of special parameters (shift and scale). Some characteristic properties of sufficient statistics are mentioned, and some extensions of the concept of sufficiency are also studied.

In Chapter 9, the "stability" of certain characteristic properties discussed in the preceding chapters are analysed. For example, if, instead of the independence of $X_1 - X_2$ and $X_1 + X_2$, is assumed only their "ε-independence" (in a suitable sense), then what can be said about the distribution of the random variables under study? It will be shown that X_1 and X_2 will be "$h(\varepsilon)$-normal" with a certain $h(\varepsilon)$. Analogous stability holds for characterizations of the normal law through the admissibility of $\bar{X}$ as an estimator for the shift parameter.

The problem of characterizing the distribution of the structure r.v.'s in linear models studied extensively by Rao [128, 175, 176] is considered in Chapter 10. It is shown that the normal random vector is characterized by the nonuniqueness of linear structure, both for a given number of structure r.v.'s and with respect to their number.

Conditions for the uniqueness of linear structure are examined. It is shown that every random vector possessing a linear structure can be expressed as the sum of two random vectors, of which one has unique structure and the other is normal. These results have important significance in some problems of biological and psychological measurement.

In Chapter 11, we study polynomial and rational statistics of a normal sample, and algebraic transformations preserving normality. It is shown that "almost all" (in a certain sense) pairs of independent polynomials can be "unlinked" provided only that the degree of one is significantly larger than that of the other.

Some isolated results constitute Chapter 12. Characterization problems from the theory of sequential estimation are studied there. We consider Markovian stopping times of the type of "first passage" for Binomial, Multinomial, Poisson, and Wienerian parameters; the problem consists in characterizing sequential estimation plans through some properties of the Markovian stopping times corresponding to them.

In Chapter 13, some results of a miscellaneous character are collected together. The normal, gamma, and other distributions are characterized here through the properties of "minimum Fisherian information," "maximum entropy," etc. Various properties of order statistics are used for characterizing the exponential and geometric distributions. Some characteristic properties of the important discrete distributions are also cited. The set of results given here are important in problems of testing of hypotheses.

Finally, some unsolved problems, which appear to the authors to be of interest and importance, are given in Chapter 14.

The chapters of the book bear the Indo-Arabic numbers: 1, 2, . . . ; sections within chapters bear double numbers (for example, 7.1 denotes the first section of Chapter 7): articles within sections (if there be such) bear triple numbers (1.1.1 denotes the first article of Section 1.1). Formulas, theorems and lemmas in each section have been numbered sequentially, independently of the presence or otherwise of articles or subsections: for example, formula (7.7.1.), Theorem 7.5.1., etc. The same practice is followed in making references.

We express our thanks to A. A. Zinger, A. L. Rukhin, and V. N. Sudakov for help in the compilation of the various sections of the book and also to A. P. Khusu, S. I. Chirkunova and T. M. Gryaznova for help in preparing the manuscript.

The English edition closely follows the Russian text except for some remarks, notes and stylistic changes, and the new material given at the end of the book following Chapter 14, as addenda A and B.

CHAPTER 1

Preliminary Information and Auxiliary Tools

1.1 SOME LEMMAS ON CHARACTERISTIC FUNCTIONS

1.1.1 Characteristic Functions

For any given one-dimensional *distribution function* (d.f.), F, consider the function f of the real variable t given by

$$f(t) = \int e^{itx}\, dF(x). \tag{1.1.1}$$

Since $e^{itx} = \cos tx + i \sin tx$ has bounded real and imaginary parts, the integral (1.1.1) exists. The complex-valued function f of the real variable t is called the *characteristic function* (c.f.) of the d.f. F. The function f is uniformly continuous on the entire real line. It is also referred to as the Fourier-Stieltjes transform of the d.f. F. Some important results concerning c.f.'s are stated below without proofs—which are available in standard textbooks.

(i) $f(0) = 1, f(-t) = \overline{f(t)}, \ |f(t)| \leqq f(0) = 1.$

(ii) If $f_1, \ldots, f_n$ are c.f.'s, the a_j are nonnegative real numbers with their sum equal to unity, then $f = \Sigma\, a_j f_j$ is also a c.f.

(iii) (Uniqueness theorem) Two d.f.'s coincide if and only if their c.f.'s coincide.

(iv) (Inversion theorem) Let the d.f. F have f as its c.f. Then, if a and b are continuity points of F, we have

$$F(b) - F(a) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-iat} - e^{-ibt}}{it} f(t)\, dt. \tag{1.1.2}$$

It is possible to check whether the point x is a continuity point of F or not by means of the formula

$$F(x) - F(x-0) = \lim_{T\to\infty} (1/2T) \int_{-T}^{T} e^{-itx} f(t)\, dt$$

(v) (A sufficient condition for the existence of a density function) If f is the c.f. of the d.f. F, and f is absolutely integrable over $(-\infty, \infty)$, then F is absolutely continuous and a version of the density function is given by

$$F'(x) = (1/2\pi) \int e^{-itx} f(t)\, dt. \tag{1.1.3}$$

(vi) (The continuity theorem) A sequence $\{F_n\}$ of d.f.'s converges weakly to a d.f. F if and only if the corresponding sequence of c.f.'s $\{f_n\}$ converges pointwise to a function f which is continuous at the origin. f is then the c.f. of F.

(vii) (A variant of the continuity theorem) A sequence $\{F_n\}$ of d.f.'s converges weakly to a d.f. F if and only if the corresponding sequence of c.f.'s $\{f_n\}$ converges to a function f uniformly on every finite real interval $[-T, T]$. f is then the c.f. of the limiting d.f. F.

(viii) (Existence of moments) If f, the c.f. of the d.f. F, has the k-th order derivative at the origin, then F has moments of all orders up to k if k is even, and up to $(k-1)$ if k is odd.

Conversely, if the s-th moment of the d.f. F exists, then f can be differentiated s times for all real t, and

$$f^{(s)}(t) = i^s \int x^s e^{itx}\, dF(x) \tag{1.1.4}$$

so that, by the bounded convergence theorem, $f^{(s)}$ is a continuous function. In particular, the s-th moment of F is given by: $\mu_s = i^{-s} f^{(s)}(0)$. The existence of the derivatives of all orders for f at the origin is equivalent to the existence of the moments of all orders for F.

(ix) (Symmetric d.f.'s) A d.f. F is called *symmetric* if

$$\bar{F}(x) \equiv 1 - F(-x-0) = F(x) \qquad \text{for all real } x. \tag{1.1.5}$$

If f is the c.f. of such a d.f., then $f(-t) = f(t)$ and hence f is real-valued; the converse is also true.

1.1.2 Infinitely Divisible Distributions

The c.f. f will be called *infinitely divisible* (i.d.) if, for every natural number n, there exists a c.f. f_n such that

$$f(t) = [f_n(t)]^n \qquad \text{for all real } t.$$

We state below some properties of i.d.c.f.'s.

(i) An i.d.c.f. never vanishes.

(ii) The product of two, and hence of any finite number of, i.d.c.f.'s is again an i.d.c.f.

(iii) A c.f. which is the pointwise limit of a sequence of i.d.c.f.'s is i.d.