Translations of

MATHEMATICAL MONOGRAPHS

Volume 53

Statistical Decision Rules and Optimal Inference

N. N. Čencov

American Mathematical Society

Statistical Decision Rules and Optimal Inference

This page intentionally left blank

10.1090/mmono/053

Translations of

MATHEMATICAL MONOGRAPHS

Volume 53

Statistical Decision Rules and Optimal Inference

N. N. Čencov

American Mathematical Society
Providence, Rhode Island

СТАТИСТИЧЕСКИЕ
РЕШАЮЩИЕ ПРАВИЛА
И ОПТИМАЛЬНЫЕ
ВЫВОДЫ

Н. Н. ЧЕНЦОВ

ИЗДАТЕЛЬСТВО «НАУКА»
ГЛАВНАЯ РЕДАКЦИЯ
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ
МОСКВА 1972

Translated from the Russian
by the Israel Program for Scientific Translations

Translation edited by Lev J. Leifman

2000 *Mathematics Subject Classification.* Primary 62C05, 62E10;
Secondary 57R55.

ABSTRACT. This monograph is devoted to the general theory of statistical inference. The approaches developed here permit the author to consider, from a single point of view, the main concepts and laws of mathematical statistics, methods of constructing optimal statistical estimates, etc. The book is intended for those working in mathematical statistics, information theory, game theory, and also applications of probabilistic and statistical methods. Individual sections may be of interest to specialists in measure theory, differential geometry and nonlinear functional analysis.

Library of Congress Cataloging-in-Publication Data

Chentsov, N. N. (Nikolaĭ Nikolaevich)
Statistical decision rules and optimal inference.
(Translations of mathematical monographs; 53)
Translation of: Statisticheskie reshaiushchie pravila i optimal′nye vyvody.
Bibliography: p.
Includes index.
1. Statistical decision. 2. Distribution (Probability theory) I. Title. II. Series.

QA279.4.C4613 519.5′42 81-15039
ISBN 0-8218-4502-0 AACR2
ISSN 0065-9282

AMS softcover ISBN 978-0-8218-1347-8

10 9 8 7 6 5 4 3 13 12 11 10 09 08

TABLE OF CONTENTS

PREFACE

The general concepts of a statistical decision and a statistical decision rule are basic for all of modern statistical theory. According to Wald, every particular statistical problem is a problem of decision-making: the statistician, having processed certain observational material, must draw conclusions as to the observed phenomenon. Since the outcome of each observation is random, one cannot usually expect these conclusions to be absolutely accurate. It is a job for the theory to ascertain the minimal unavoidable uncertainty of the conclusions in the problem and to indicate an optimal decision rule.

In classical problems of mathematical statistics, one is required to determine the (unknown) probability distribution of the outcomes on the basis of independent observations and certain additional information. When the number of observations used in such cases increases, one can establish various simple and quite general asymptotic relationships.

In any theory, a general law should be amenable to equivalent formulations; that is to say, the statement of the law should not vary when a situation is replaced by another one, equivalent to the former (within the framework of that theory); otherwise it would not be a general law. In classical geometry, such "changes of situation" form a group. In mathematical statistics the description of the set of equivalent situations is more complicated.

The system of all statistical decision rules for all conceivable statistical problems, together with the natural operation of composition, forms an algebraic category. This category generates a uniform geometry of families of probability laws, in which the "figures" are the families and the "motions" are the decision rules. Two families are "congruent" if and only if they possess equivalent statistical properties.

An apt name for the subject of this monograph might be "geometrical statistics". The algebra of decision rules and the natural geometry that it generates are studied here from a statistical standpoint. The geometrical methods and language that we develop are then applied to the equivariant theory of optimal estimates.

The book is intended for specialists in mathematical statistics, information theory and game theory, and also for those interested in applications of probability-theoretical methods. The reader is expected to be familiar with probability theory and measure theory, to the extent of Kolmogorov's *Grundbegriffe der Wahrscheinlichkeitsrechnung* and Halmos' *Measure Theory*, or Neveu's *Bases Mathématiques du Calcul des Probabilités*. Formally speaking, a knowledge of statistics is not assumed; all the necessary concepts are introduced and explained in §1 of the Introduction. Also included in the Introduction is the requisite material from the theory of smooth manifolds and category theory.

My interest in the "uncertainty" of statistical estimates arose when I was engaged in the development of computer-oriented methods for the estimation of unknown densities; further encouragement came from my teacher N. V. Smirnov. Without his approval, I would probably not have risked taking my research so far from orthodox statistics.

My discussions with Ju. V. Linnik had considerable influence on my work. Many of the theorems proved here are answers to his clearly phrased questions. I am indebted to him for his unflagging interest in my research and his kind attention.

I had many useful discussions with experts concerning both the basic ideas of the theory and individual results; thanks are due in this respect to L. N. Bol′šev, B. N. Delone, I. M. Gel′fand, B. V. Gnedenko, A. M. Kagan, A. N. Kolmogorov, A. A. Ljapunov, Ju. V. Prohorov, Richard Sacksteder, Ju. M. Smirnov, Charles Stein and V. S. Vladimirov.

It is my pleasure to thank Elena Aleksandrovna Morozova for her great help and valuable advice on geometrical matters, and to Fridrih Izrailevič Karpelevič for numerous profound and important remarks.

Finally, I wish to thank the book's editor, A. V. Černavskiĭ, who read the manuscript attentively and helped to eliminate various shortcomings.

N. N. Čencov

NOTES AND COMMENTS

§1

A general description of equivalent statistical problems was given by Blackwell [2, 1951], [3, 1953]. At the same time, similar ideas were developed by Stein [1951, oral communication; see 3] and Sherman [1, 1951]. Earlier work had considered only equivalent reduction of statistical problems connected with the transition to sufficient statistics; see Fisher [3, 1925], and also Halmos and Savage [1, 1949], Bahadur [1, 1954], Burkholder [1], and LeCam [2, 1964], [3, 1969]. Čencov [5, 1965] and Morse and Sacksteder [1, 1966] noticed that this equivalence relation was, roughly speaking, generated by the category of statistical decision rules. This enabled them to "algebraicize" certain concepts of statistics. Further steps in this direction were taken by Romier and coworkers in their "*Introduction à la statistique mathématique*" (Romier [2], [3], Littaye-Petit et al. [1], Martin and Vaguelsy [1], Laurant et al. [1], Martin et al. [1]). The idea of considering families of probability distributions not as objects on their own but as "figures" in a Kleinian geometry with a category of transformations is due to the author [5], [7].

The systematic investigation of statistical properties which are invariant (equivariant) in the category of statistical decision rules was initiated by Sacksteder and Čencov. Attention was paid previously only to invariance determined by the (group) symmetry of a specific problem; see Lehmann [1]; also Berk [1] and Brillinger [1].

(1) The history of statistics as the science of statistical inference usually begins with the amusing episode recounted by Bertrand in the preface to his course "*Calcul des probabilités*" [1]:

> "One day in Naples the reverend Galiani saw a man from the Basilicata who, shaking three dice in a cup, wagered to throw three sixes .... Such luck is possible, you say. Yet the man succeeded a second time, and the bet was repeated. He put back the dice in the cup, three, four, five times, and each time he produced three sixes. "Sangue di Bacco", exclaimed the reverend, 'the dice are loaded!' And they were . . . " [Quoted from G. Pólya, *Patterns of Plausible Inference*.]

(2) This approach was first put forth explicitly by Neyman and Pearson [1, 1928] in their theory of hypothesis testing.

(3) Essentially, this is a special case of the game-theoretical approach. Long ago Laplace [1, 1820] likened the derivation of an estimate to a game of chance in which the statistician suffers defeat if his estimates are inferior.

A detailed description of the concept of "statistical problem" according to Wald [1], [3] was given by Lehmann [1]. See also Birnbaum [1], De Groot and Rao [1], and Hoeffding [1].

(4) Choice of action based on results of an auxiliary experiment was first systematically considered by von Neumann in game theory. See von Neumann [1, 1928], and also von Neumann and Morgenstern [1] and Blackwell and Girshick [1] (cf. footnote (3), above).

(5) The statistician deals directly with the outcomes of the experiment only in the simplest problems, involving a discrete sample space. The slightest additional complication in the problem leads to distortion of the experimental result, due to measurement errors and grouping (when of necessity the measurement is rounded off). In such situations the decision is described by a compound decision rule $\Pi_{\text{instrument}} \circ \Pi_{\text{processing}}$, where the first factor is independent of the statistician. In this book we shall consider only the ideal situation, in which the distortions due to $\Pi_{\text{instrument}}$ may be neglected.

§2

(1) Cap = Collection of All Probability distributions.

(2) In all topological concepts we follow Kelley [1].

(3) In §13 we shall introduce the Hellinger integrals of the Radon-Nikodým derivative, such as $\int [(dQ/dP)(\omega)]^2 P\{d\omega\}$, as the limits of integral sums of a special form, with specific conventions for resolution of the indeterminacy $\infty \cdot 0$, different from the usual.

(4) See the work of C. Ionescu Tulcea, under whose "lifting" $P\{\cdot\} \to p(\cdot)$, to a (finite) linear combination of laws corresponds the linear combination of the densities, and to a geodesic mean (see Definition 18.2) corresponds the normalized geometric mean of the densities. The correspondence may fail for countable linear combinations (see end of §9.4).

(5) The possibility of modifying almost proper conditional distributions to obtain proper distributions depends only on the properties of the mapping f. If there exists a proper conditional distribution for at least one Q, it will serve to modify any $P\{\cdot|f\}$.

Blackwell and Ryll-Nardzewski [1] proved that not every almost proper Borel conditional distribution can be modified by a Borel procedure to obtain a proper conditional distribution. Their arguments rely on the fact, established by Novikov [1], that it is not possible to **B**-uniformize an arbitrary Borel function (see subsection 14).

(6) This approach to the theory of stochastic processes was developed by Lévy [1]; see also Wiener [1]. For the connection with the theory of Monte Carlo methods, see Gel'fand, Frolov and Čencov [1], Rankin [1] and Čencov [11]. A similar approach was suggested by Blackwell [4] (see also Sazonov [1]).

(7) Provided that the initial function is not many-valued. Nevertheless, any **B**-function may be uniformized by an **A**(**B**)-function, where **A**(**B**) is the σ-algebra generated by the analytic sets (see Saks [1]) $\mathbf{B} \subset \mathbf{A}(\mathbf{B}) \subset \mathbf{B}^*$. See also Luzin [1], Luzin and Novikov [1], and Luzin and Sierpiński [1].

(8) The proofs of Lemma 5.11 and all other assertions of §5 rely only on general facts of measure theory and make no use of the concept of a constructive measure. Therefore the proof given here of Lemma 2.15 contains no vicious circle.

§3

(1) Applying the Taylor formula with integral remainder to the function $f(x) = f(\varphi_\alpha^{-1}(x))$ of the local coordinates $(x^{(1)}, \ldots, x^{(n)})$, we can write (see Helgason [1])

$$f(x) = f(x_0) + \sum_{i=1}^{n} (x^{(i)} - x_0^{(i)}) g_i(x),$$

where $g_i(x)$ is a differentiable function, $g_i(x) \approx X_i f(x)$, $g_i(x_0) = (X_i f)_p$, and $x_0 = \varphi_\alpha(p)$. It follows at once from conditions (3.3) and (3.4) that

$$(Xf)_p = \sum_i (X_i f)_p \cdot (Xx^{(i)})(p).$$

This gives (3.10).

(2) The concept of affine (linear) connection is due to Levi-Civita [1]; see also Weyl [1]. Originally the definition of linear connection referred to surfaces in euclidean space, and the

concept was then carried over to Riemannian spaces. Any Riemannian differential metric defines a torsion-free linear connection (see, for example, Favard [1]).

§4

The invariants of a family of probability laws were first considered as invariants of an object of a category by Morse and Sacksteder [1]; see also Sacksteder [1], [2]; monotone invariants were first studied by the author in [7]. Covariants in geometry (with the group of motions) were investigated by Rozenfel'd [1].

(1) Yet another example of a category is the category of all measurable spaces (Ω, S) with all measurable mappings of one into the other. There is a closure operation $(\Omega, S) \to (\Omega, S^*)$ for the objects of this category. By Lemma 2.1 this operation is a functor.

§5

(1) A transition probability distribution describes not only a decision rule but also, for example, a communication channel with random noise, with "input alphabet" Ω and "output alphabet" $\mathfrak{E}$. Thus, as remarked by Dobrušin, the category of Markov morphisms is at the same time a category of statistical communication channels without memory etc. (see also Csiszár [5]). Note that Theorem 9.1 (on sufficient statistics) has a very graphic interpretation in terms of a two-way communication channel. In fact, it was shown by Wrighton [1] that in a certain sense the problem of statistical inference is a degenerate problem of communication in the presence of noise.

(2) It can be shown that the integrals (5.5) are equal to the double integral

$$\iint_{\Omega' \times \Omega''} P\{d\omega'\}\Pi\{\omega'; d\omega''\}f(\omega''),$$

understood as a limit of Darboux-Young integral sums

$$\sum_{(B \times A)} P\{B\}\Pi(\omega'_B, A)f(\omega_A)$$

with respect to the filter of finite partitions of the space $\Omega' \times \Omega''$ into measurable rectangles $B \times A$, $B \in S'$, $A \in S''$. To prove the existence of the limit, one must consider the lower and upper Lebesgue integral sums for the partition with

$$A_k = \{\omega'': kn^{-1} \leqslant f(\omega'') < (k+1)n^{-1}\}, \qquad k = 0, 1, \dots, n;$$
$$B_j^{(k)} = \{\omega': jn^{-2} \leqslant \Pi(\omega'; A_k) < (j+1)n^{-2}\}, \qquad j = 0, 1, \dots, n^2,$$

and then let $n \to \infty$.

(3) Incidentally, the usual δ-function defines a functional on continuous functions, while any measurable function can be integrated with respect to a δ-measure. Hence they define distinct functionals.

§6

(1) For example, let $\{P_1, Q_1\} \geqslant \{P_2, Q_2\}$ (Π_{12}) and, conversely, $\{P_1, Q_1\} \leqslant \{Q_2, P_2\}$ (Π_{21}). It is readily checked that then $\{P_1, Q_1\} \sim \{P_2, Q_2\}$ $(\Pi_{12}, \Pi_{21} \circ \Pi_{12} \circ \Pi_{21})$, and $\{P_1, Q_1\} \sim \{Q_1, P_1\}$ $(\Pi_{12} \circ \Pi_{21}, \Pi_{12} \circ \Pi_{21})$.

§7

(1) All the auxiliary lemmas of §5 were proved for bounded functions. They carry over trivially to nonnegative functions which may take the value $+\infty$.

(2) In Morse and Sacksteder [1] the integral invariants were not introduced, and their arguments were therefore a little more complicated.

(3) Or weakly continuous Markov chains with compact state space (see Bebutov [1]).

§8

(1) In contrast to the term "divergence" (see Kullback [3]), "deviation" reflects the asymmetric way in which the probability measures enter.

(2) A similar assertion was made (without proof) by Rosenblatt-Roth in the text of [1]. A more general statement, using slightly different terminology, was however proved earlier by Csiszár [5].

(3) Of course, for arbitrary pairs the equality $I[Q|P] = I[Q'|P']$ does not imply equivalence. We published (8.6) in [7].

(4) The functional $\chi^2(Q, P)$ was studied by Kagan [1] as a measure of the difference between Q and P (and called the *divergence* of P and Q). The functional (8.10) was studied by Perez [1] and Onicescu [1], under the name of information energy.

(5) In the same way as Shannon's entropy i_{Sh} describes the exponential growth $\exp[Ni_{\text{Sh}}]$ of the number of highly probable messages when the length N of the message is increased. Recall that for a discrete space $\Omega = \{\omega_1, \dots, \omega_m\}$, uniform distribution $Q \leftrightarrow (1/m, \dots, 1/m)$ of outcomes and distribution $P \leftrightarrow (p_1, \dots, p_m)$ we have

$$i_{\text{Sh}}[P] = -\sum_{k=1}^{m} p_k \ln p_k = \ln m - I[Q|P].$$

(6) The proof of (8.21) goes back to an unpublished paper of Stein [3] and the dissertation of Joshi [1] (in this connection Chernoff [2] and Kullback [3]). The relation (8.22) is due to Chernoff [1], whose proof relies on a rather delicate limit theorem of Cramér [3]. Formula (8.23) is due to N. P. Salihov. Salihov has recently proved that in testing several hypotheses $P_1, \dots, P_m$ the rate of exponential decrease of the maximum probability of error is given by

$$I = \min_{j,k} \inf_{R \in \text{Caph}} \max\left\{ I[P_j|R], I[P_k|R] \right\}.$$

(7) If the quantile falls on an atom of the distribution of $f_N(\varepsilon)$, $P_1^N\{\varepsilon: f_N(\varepsilon) = k_N\} > 0$, then, as usual, randomization is carried out (see Lehmann [1], Chapter 3), so that the hypothesis P_1 is rejected invariably if $f_N(\varepsilon) < k_N$ and with a certain probability q if $f_N(\varepsilon) = k_N$, where

$$P_1^N\{\varepsilon: f_N(\varepsilon) < k_N\} + qP_1^N\{\varepsilon: f_N(\varepsilon) = k_N\} = b.$$

(8) The convergence $k_N \to -I[P_0|P_1]$ also occurs if $-I[P_0|P_1] = -\infty$, since the difference $\ln p_1(\omega) - \ln p_0(\omega)$ is P_1-quasi-integrable. A rigorous proof of this version of the law of large numbers for the likelihood function is contained in the proof of Lemma 27.18.

(9) If $I = +\infty$, the estimation from below is trivial.

(10) The asymptotic behavior described by Mourier and Sakaguchi (see Kullback [3]) is apparently erroneous.

(11) Thus, Rényi [1] considered the invariants $\ln J_u$ and $d \ln J_u/du$, related to the information deviations by (8.25) and (8.26). The integral $J_{1/2}(P, Q)$ was, however, investigated by Bhattacharyya [1, 1943].

A concept closely related to information deviation is that of information (contained in one random variable relative to another); see Kolmogorov [5] and Gel'fand, Kolmogorov and Jaglom [1]. If F_{12} is the joint distribution law of two random variables $\xi = x(\omega)$ and $\eta = y(\omega)$, and F_1 and F_2 are the marginal laws, then $I = I[F_1 \times F_2|F_{12}]$.

It is worth mentioning that 2 arccos $J_{1/2}(P, Q)$ is the distance between P and Q in the natural Riemannian metric defined by the Fisher information tensor (see §11 and §12.9).

§9

(1) Here lies the essential difference between the semigroup of Markov transformations and the semigroup of all linear transformations; in the latter almost all transformations are invertible (within the semigroup).

(2) Lemma 9.4 has been known for a long time, and not only to specialists in Markov chains. See, for example, Kreĭn and Rutman [1] or Bebutov [1]. The sets ε of outcomes ω described in the

condition are (except for ε_0) sets of communicating recurrent states of the Markov chain Π. We have omitted the statement and proof of the fact that every such set has its own stationary distribution.

§10

(1) Let $C_y^+ = \{\omega: f(\omega) > y\} \cap A_j$, $C_y^- = A_j - C_y^+$. Then, since A_j is an atom, either C_y^+ or C_y^- is a Z-set. By induction, it follows that for any finite sequence $-\infty = y_0 < y_1 < \cdots < y_N = +\infty$ only one of the sets $\{\omega: y_{k-1} < f(\omega) \leq y_k\} \cap A_j$ is not a Z-set. Consider a countable sequence of numbers, dense in the real line, with $x_0 = -\infty$ and $x_1 = +\infty$. Subject to reordering, every finite subsequence $x_0, x_1, \ldots, x_N$ defines a partition of the real line into half-closed intervals. Applying the previous reasoning, we see that only one of the intervals does not correspond to a Z-set, and with increasing N this interval becomes smaller. As $N \to \infty$ these half-closed intervals $(t_N', t_N'']$ shrink to a point, which we denote by τ. Denoting $B_N = \{\omega: t_N' < f(\omega) \leq t_N''\} \cap A_j$, we have $A_j - B_N = C_N \in \mathbf{Z}$. Since $\mathbf{Z}$ is a σ-ideal, it follows that $\bigcup_1^\infty C_N = C \subseteq A_j$ and $C \in \mathbf{Z}$, whence $A_j \neq C$. Thus $\bigcap_1^\infty B_N = B = A_j - C$ is not empty. The function $f(\omega)$ takes the value τ on the set B, and $A_j - B = C \in \mathbf{Z}$.

(2) For any $(f^1, \ldots, f^m)$ there is a corresponding class of equivalent S-measurable functions, a suitable representative of which is the function $f(\omega) = f^j$ for $\omega \in A_j$, where $A_1, \ldots, A_m$ is a canonical partition of Ω into atoms. The function $\nu\{\cdot\}$ defined on the σ-algebra S by

$$\nu\{H\} = \int_H f(\omega)\mu\{d\omega\},$$

is a μ-dominated measure, independent of the specific choice of f from its equivalence class. Obviously,

$$\nu_j = \nu\{A_j\} = \int_{A_j} f(\omega)\mu\{d\omega\} = f^j \times \mu_j.$$

Thus any S-measurable function $f(\cdot)$ determines a linear transformation of the cone of measures, Z-equivalent functions determining the same transformation.

(3) In a finite-dimensional space all norms are equivalent (see (14.3)). Here, therefore, the expression $o(\tau)$ has an absolute meaning. The fact that this is not so in infinite-dimensional spaces makes it difficult to define smooth families (see §§13 and 14).

(4) A "function-measure" in the terminology of Bogoljubov [1].

(5) Computational formulas of this type are used, for example, in Gel'fand, Frolov and Čencov [1].

(6) The category π CAP of collections Capf with the system of positive centrally-projective homomorphisms was mentioned in Čencov [5] in connection with the study of natural equivalences of families. The connected group of invertible positive centrally-projective transformations is the translation group.

§11

Various authors have repeatedly expressed the opinion that the Fisher information tensor defines a natural Riemannian metric in the manifold of mutually absolutely continuous probability distributions (see, for example, Kullback [2], [3]). However, serious results in this area were obtained only by Kozlov [1], whose work stimulated our own investigations.

(1) Congruent embeddings of the simplexes Capf are described by linear mappings in both natural and canonical coordinates. For this reason the statement of the lemma is true for any index values which vary according to a tensor law under (linear) transformation from one canonical (or natural) coordinate system to another. This argument is applied in Lemma 11.4 to the matrix of second derivatives in canonical coordinates, and it may be applied to fields of differential operators.

Note that since the geometry is almost homogeneous, equivariant scalar fields, (contra)vector fields and tensor contravalent fields vanish identically if they vanish at some point. This is not the case for covalent tensor fields (since they are "carried backward" by the mappings, not "taken out").

(2) The geometries of the simplex Caph in natural and canonical coordinates are very reminiscent of intrinsic geometries of the first and second kind of the simplex as hypersurface in the enveloping m-dimensional space $\mathrm{Var}(\Omega, \mathrm{S}_m)$ equipped (see Norden [1]) with a unique equivariant field of normals $\mathfrak{n}(P) = p \leftrightarrow P$.

It is highly important that the equivariant tensor g coincides for "tangent" measures with the Radon-Nikodým tensor, which converts a measure on S into an S-measurable function, i.e. into an element of the dual space.

(3) The gradients of an invariant function on different objects are represented by covectors of different dimensions. In the case of congruent embedding, however, the scalar product of any embedded tangent vector with the gradient remains unchanged, since it is equal to the derivative in the appropriate direction. This implies that the gradient is equivariant.

§12

I am indebted to E. A. Morozova for advising me to seek the natural geometry of a linear connection. Theorem 12.3 provides the answer to a question put to the author by Ju. V. Linnik. I am indebted to F. I. Karpelevič for sharpening its formulation.

(1) G. and G. G. Vranceanu [1] associate a certain generalized affine connection with every time-continuous Markov process having finitely many states. Their formulation of the problem differs from ours.

(2) This follows from the uniqueness theorem for the solution of system (12.18) or of the system of second-order equations obtained by eliminating the parameters.

(3) A geodesic is a trajectory of a one-parameter subgroup of the translation group. According to the supplement to Lemma 10.2, this group is simply transitive, while the correspondence between the coordinate s and the law P is unique in view of Lemma 10.3.

§13

The monotonicity of the approximating sums for the information deviation was apparently first pointed out by Sanov [1] (see also the proof in Kallianpur [1]). A general approximation theory for functionals of the form (13.18) has been developed by Csiszár [2], [3] and independently by Ghurye [1] and Čencov [12], [13]. Special subclasses of such functionals have also been investigated by other authors; see Martin and Oheix [1] and their subsequent publications.

(1) Translations, as projective transformations of a simplex, are also fractional-linear. Passage to a conditional distribution may be described as a limit of translations when the canonical parameter becomes infinite (see §21.8).

(2) For the definition of projective and inductive limits, see Palamodov [1] and also Scheffer [1].

(3) Recall that in view of the conventions (13.23) the second integral is a Lebesgue integral only when $Q \gg P$. But if

$$P\{\omega\colon (dQ/dP)(\omega) = 0\} > 0,$$

then the Q-integral of an infinite function over this set is put equal to $+\infty$.

§14

(1) Since the function f may not be one-to-one, the surface defined by its image may be self-intersecting. The tangent plane is a local concept. If $f(x_1) = f(x_2) = y_0$, then there are two tangents at the same point y_0 of the space, corresponding to different points (x_1, y_0) and (x_2, y_0) of the surface.

It should be noted that condition (14.5) is quite restrictive. It is certainly not satisfied when $\dim X > \dim Y$ (i.e. the dimension of a smooth surface cannot exceed the dimension of the enveloping space).

(2) However, this situation is typical even in linear topological spaces (see Averbuh and Smoljanov [1]). One then has to be content with functionals which are differentiable with respect to the subspace of increments of a finite "norm".

Note that the weakest $\mathfrak{L}^1(P)$-metric (which coincides with variation for dominated measures) is too weak for us (see above, subsection 3), while the strongest $\mathfrak{L}^\infty(P)$-metric defined by the essential supremum of the Radon-Nikodým derivative is far too strong–even the family of normal laws is not continuous in it.

(3) A definition similar to $\mathfrak{L}^2(P)$-differentiability, for families of one real parameter, was given by Schmetterer [1].

(4) This follows from well-known theorems of classical analysis. We shall not present the proofs, since the corresponding proposition is easier to state and to prove for continuously differentiable surfaces (see Lemma 16.3).

§15

(1) It is natural to confine the study of single-fold differentiability to differentiability in the field of $\mathfrak{L}^2(P)$-metrics. However, n-fold differentiability should be considered in the stronger $\mathfrak{L}^u(P)$-metrics, or, more precisely, in a hierarchy of metrics (see §26.5); otherwise such "analytic functions" as $I[Q|P]$ turn out to be nondifferentiable.

(2) That is, we obtain an estimate for the squared norm in the quotient space of $\mathbf{L}^2(P)$ by a line of constants (cf. Schmetterer [1] and Čencov [9]). The corresponding induced metric was introduced by Gel'fand (see Hille [1]).

(3) (15.12) and (15.14) are known as the (one-dimensional) information inequalities. Associated with their rigorous proofs are such names as Cramér [1], [2], C. R. Rao [1], Darmois [1], Fréchet [1], as well as Dugué [1] and many others (see van der Waerden [1]).

(4) See Cramér [1], and C. R. Rao [2].

(5) For example, P''' etc. (see Bhattacharyya [2], Bol'šev [1] or Seth [1]). The more the law P admits unbiased estimators $f(\omega)$ for zero, $\mathbf{M}_P f(\omega) = 0$, the larger the lower bound for the variance of the unbiased estimator (see Kagan [4] and C. R. Rao [3]).

(6) The definition of an efficient estimator goes back to Fisher [1], [2]. Regarding estimators for which the risk coincides with the bound (15.23), see De Groot and M. Rao [1].

(7) This theorem was considered with no clearly formulated smoothness conditions by Bhattacharyya [2]. A logical error has slipped into the proof given by Kullback [3, Chapter 3]. Fraser [1] gives the proof for one parameter only. In contradistinction to the authors just listed, we have not presumed the existence of smooth densities, and we allow the appearance of formal estimators. A stronger version of this theorem is Theorem 23.6.

§16

(1) But convergence (16.2) in the $\mathfrak{L}^2$-metric follows from the truth of condition (16.1) in the $\mathfrak{L}^4$-metric (cf. §26.5 and footnote 4 to §26).

(2) Or as the limit of Riemann sums in the local $\mathfrak{L}^u$-metric.

(3) A related fact was actually used in the proof of Lemma 16.1.

(4) A fortiori, according to Corollary 1 to Lemma 16.1, differentiable with uth moment in the sense of Definition 14.3.

(5) Continuity of $p_i'(\omega; t)$ is not required here. Neither is measurability of $\tau(\omega)$, since the measurability of

$$(p'_{(e)}(\omega; \tau(\omega))) = |t - \theta|^{-1}[p(\omega; t) - p(\omega; \theta)]$$

follows from the definition.

(6) Integral corollaries from the information inequality for one-parameter families were first obtained by Blyth [1], Girshick and Savage [1], and Hodges and Lehmann [1]; see also Karlin [1]. The examples of Stein [1], [2] show that these corollaries do not carry over to $n > 3$ parameters. This is why Kiefer [1], in his fundamental review of the theory of optimal multivariate estimators, expressed doubt as to the applicability of the information inequality as a tool for construction of optimal estimators. The theorem stated here was proved by the present author in [9].

Integral corollaries for the variance of an asymptotically normal law were obtained by LeCam [1] and Schmetterer [1], in the case $n = 1$.

§17

(1) This definition was apparently first given by Minty [1]. For further research on the subject, see Rockafellar [1].

(2) Any one-to-one continuous mapping of the real line into itself, with continuous inverse, is automatically monotone and takes monotone functions to monotone functions. This is not the case for multidimensional spaces.

§18

Exponent families are perhaps the most important and well-studied class of families. Systematic investigation of their general theory began with Koopman [1, 1936], who characterized them as families having finitely many sufficient statistics (see also Dynkin [1, 1951]). They appeared independently in problems of statistical physics (see Khintchine [1]).

An account of the general theory of exponent families was given in Kullback [3, 1959] and Lehmann [1, 1960]. General analytical approaches to the theory of parameter estimation and hypothesis discrimination for these families were proposed in monographs by Linnik [2], [3]. A number of fundamental problems of the theory were also touched upon in Blackwell and Girschick [1] and Robbins [1]. Our own research has dealt mainly with the "geometrical" aspect of the theory (see Čencov [4], [6]).

(1) The component of the identity of the group of centrally-projective transformations of Caph(Ω, S_m, $\mathbb{D}$) into itself, see §10.

(2) More precisely, the statistics $q_1(\omega), \ldots, q_n(\omega)$ and $q_0(\omega) = 1(\omega)$ are linearly independent.

(3) Though this density may determine a random functional on smooth functions, i.e. define a generalized random variable in Gel'fand's sense [2].

(4) Though the dimension of the family of generalized random variables is n.

(5) These limits form the Vorob'ev-Faddeev fine boundary of the simplex Caph(Ω, S_m, $\mathbb{D}$). Under these conditions, to each strictly dominated law corresponds a whole manifold of limits. The point is that the simplex Caph(Ω, S_m, $\mathbb{D}$), as a homogeneous manifold of zero curvature, has no natural boundary (see Gel'fand and Graev [1]). The boundary, obtained by Vorob'ev and Faddeev [1] from other considerations, may be obtained by the geometrical methods of Karpelevič [1].

(6) This means that the correspondence $s \to P_s$ determined by (18.1) and (18.2) defines a chart of the family.

(7) Recall that, in any semiordered space, first-order homogeneous functions are defined in an invariant manner (see Kantorovič, Vulih and Pinsker [1]).

(8) The desirability of presenting a direct proof for this theorem was pointed out to the author by Ju. V. Linnik.

The theorem may also be derived from the general results of §7.

(9) This equation is transformed to a form more convenient for our purposes (cf. Bernstein [1]). Regarding the equation $p'(x) = sg'(x)p(x)$, see also Mathai [1].

§19

(1) The canonical affine parameter of a geodesic (see Favard [1]), or the canonical variable (see Khintchine [1]). This terminology is more "canonical" than that adopted by Linnik [2].

(2) The transformation from canonical to natural parameter in statistical physics is associated with the introduction of the notion of temperature (Khintchine [1]). A natural parameter was considered as an ancillary tool by Bhattacharyya [2] and Kullback [3]. The term "natural parameter" itself is due to the author [6]. The main result of the section were published in the indicated papers.

(3) See, for example, Dieudonné [1].

(4) Since for $s^{(2)} > 0$ the integral of the positive function $\exp[s^{(2)}x^2 + s^{(1)}x]$ from $-\infty$ to $+\infty$ is identically equal to $+\infty$.

(5) Unique with probability one, since both $q(\omega)$ and $p(\omega; s)$ are defined up to values on any Z-set.

§20

(1) Formula (20.8) is due to Huzurbazar [1]. (20.9) was proved by Kullback [3], who also provides references to the work of other authors. In the multidimensional case the Legendre conjugacy of the parameters, and, accordingly, (20.10), was first explicitly considered in Čencov [6]. The Young inequality (20.12) was proved much earlier by Sanov, by minimization of a suitable expression.

(2) Since $I[P|R] = +\infty$ when P does not dominate R, formulas (20.15) and (20.16) permit a description of the geodesic hull spanned by mutually absolutely continuous laws P_i, as the family of probability distributions $\{P_A\}$ where each $P_A = R$ minimizes a sum $\Sigma_i\, \alpha_i I[P_i|R]$. Formula (20.15) recalls the expression for the moment of inertia about an arbitrary point in terms of the moment of inertia about the center of gravity. (And when the "masses" of points vary, the center of gravity runs over the entire linear span.)

(3) For nonconstructive families, all that one can assert is weak equivalence in the sense of Morse and Sacksteder (see above, Definition 6.4).

§21

(1) This simple example shows that the natural chart of an infinite-dimensional exponent family has a more intricate structure than one might expect at first glance. Example 6 (below) is due to the author [12].

(2) See also Basu [1], Doss [1] and Kale [1].

(3) This boundary is the most economical. The finer and therefore more massive boundary of Vorob'ev and Faddeev [1] converts the simplex Caph(Ω, S_m, $\mathbb{D}$) into a compact set on which the conditional probabilities of each event relative to any nonempty hypothesis are everywhere continuous.

§22

(1) The theorem may be slightly generalized by allowing the law to be a point of the boundary at infinity.

(2) The fact that the maximum likelihood method coincides with the method of minimum information deviation has been noticed by many authors, among them Kullback and Čencov; see also Kriz and Talacko [1], and Hartigan [1].

§23

(1) It is convex by Lemma 19.1. Thus, for exponent families γ with domain G^γ satisfying (23.4), the domain of the natural parameter is convex.

(2) Definition 27.4 of a smooth family is a substitute, since it demands that the densities, not the measures themselves, be smooth.

(3) By Lemma 28.5, it is smooth in the sense of Definition 27.5. The definitions themselves were sought so that geodesic (exponent) families satisfy them (and for differentiable families, by analogy with finite-dimensional manifolds Caph, the information inequality be automatically valid).

(4) The inequality between the first and last members in (23.12) follows at once from (26.20). The direct proof is also quite easy.

(5) If the functional $\int \varphi[(dQ/dP)(\omega)]P\{d\omega\}$, where φ is a convex function, is not continuous on Caph(Ω, S, Z) in variation $|\mu|$, then according to Csiszár [2], [4] it does not define a uniform topology (evidently, a topology can be defined only by a hierarchy of such functionals; see §26.5). The terms used here are therefore conditional.

§24

(1) By §27.5, estimators of the form P_α, where α is a parameter estimator, for an exponent family $\mathfrak{N}$ with Gaussian loss function, form a complete class. In regard to this problem our formulation is close to that of De Groot and M. M. Rao [1] and M. M. Rao [1]; see also De Groot [1].

(2) Lemma 24.1 and its corollary are essentially rephrased versions of a well-known inequality of Blackwell [1], Kolmogorov [4] and C. R. Rao [1]; see also M. M Rao [1].

(3) The proof given in our paper [9] is based not on Theorem 16.1 but on a simpler version, dealing specifically with the family $\mathfrak{N}$.

(4) The author arrived at the "nonsymmetric Pythagorean geometry" of §§22 and 23 in [10] when he tried to generalize the geometry of the Gaussian method of least squares (see Gauss [1] and Linnik [1]) to arbitrary geodesic (exponent) families. The actual statement of Lemma 24.4, in some form or another, has frequently been used by statisticians to improve decision rules (see Thompson [1]).

(5) We put aside the very important and interesting (but much more complicated) Bayesian formulation of Robbins [1], [2] (see also Neyman [2]).

(6) Note that if $L(P, Q)$ is monotone and admits both expansions (24.24) and (26.10), then the constants $c_0[L]$ in both formulas are equal, but without the additional assumptions of uniform differentiability one cannot assert that the constants $c[L]$ are equal.

§25

(1) See also Frolov and Čencov [1], and the following publications: Van Ryzin [1], Schwartz [1], Kronmal and Tarter [1], Bosq [1], [2], [3], Sesan et al. [1], and Sizova [1]. Estimators of type (25.1) have been discussed earlier by Rosenblatt [1]; see also Schuster [1].

(2) It is easy to see that $\Sigma_j \mathbf{M}[\alpha_j(\xi)]^2$ is independent of the choice of the orthonormal basis $\varphi_1(x), \ldots, \varphi_n(x)$ in E_n, since for each x the sum $\Sigma_k[\varphi_k(x)]^2$ is invariant under orthonormal changes of basis.

(3) The integrand in (25.31) is also independent of the choice of basis (cf. footnote 2, above).

(4) The norm of the deviation of the histogram from the graph of the density, in the metric of the space C, has been estimated by Smirnov [1], [2] (see also Tumanjan [1]). If the grouping is optimal, it decreases almost like $N^{-1/3}$ (up to a logarithmic factor).

(5) See also Rosenblatt [1].

(6) These constraints may be relaxed somewhat by making them closer to quasi-homogeneity conditions (see Definition 28.1).

(7) We used this argument in [3]. It leads to a less precise accuracy bound than the inequality of Lemma 16.6, and is applicable only when the quality of the estimator is measured by maximum risk. On the other hand, it works under weaker constraints, and this is essential in the case of infinite-dimensional families.

(8) We are assuming here that all the moments $\int \varphi_k(x)p^*(x)\mu\{dx\}$ of the initial estimator p^* are measurable, which implies that p_n^* is measurable and hence (by Lemma 17.14) that π_n^* is measurable.

(9) Under very weak restrictions (cf. Tumanjan [1]) the squared norm (25.10) of the deviation of the estimator is asymptotically normal (see Bosq [2], also [1]). Hence one can obtain sharper results for asymptotic confidence limits than we obtained in Cencov [3], [12].

(10) The question of statistical estimation of a smooth curve is very timely (see Tukey [2] and Whittle [1]).

§26

(1) Random measures have been introduced in many publications as random functionals (i.e. random generalized functions); see Prohorov [1] and Gel'fand [2]. Discrete empirical random measures (25.9) have also been considered. Both concepts admit an effective definition.

(2) At first sight it might seem natural, following Laplace [1], to adopt as loss function some invariant metric or a substitute thereof (see the review of Adhikari and Joshi [1], where more than ten such functions are listed). However, as Gauss noted [1], the theory is much simplified if one takes a quadratic loss function (see LeCam [1]). This is why our loss function is a nonsymmetric analog $2I[P|P^*]$ of the squared euclidean distance–an analytic, "approximately quadratic" functional for which the nonsymmetric Pythagorean theorem and various other geometrical theorems are valid (see subsection 5).

An interesting unsolved problem is to describe all natural matrix loss functions (see Kagan [4]).

(3) This follows from a result of Csiszár [2], [4] (see footnote 5 to §23).

(4) For example,

$$\|R - P\|_{(P,2)} \leqslant \|R - Q\|_{(P,2)} + \|Q - P\|_{(P,2)}$$
$$\leqslant [\|Q\|_{(P,3)}]^{3/4}\|R - Q\|_{(Q,4)} + \|Q - P\|_{(P,2)}.$$

(5) Moreover, $P\{A'\}I[P'|R'] \nearrow I[P|R]$ as $A' \nearrow \Omega$.

§27

(1) This is a way of using an argument of Blackwell, Kolmogorov and Rao (see footnote 2 to §24).

(2) Under our assumptions, we can no longer assert that the majorant

$$g_1(\omega)\{h(\omega) + 3[h(\omega)]^2 + [h(\omega)]^3\}$$

for the third derivatives is integrable.

(3) The statement of the corollary follows from the theory of §16, since the smooth families are contained in the class of continuously differentiable families.

(4) This class contains all compact exponent families corresponding to a compact subdomain of the canonical parameter (see Lemma 28.5).

(5) If the dimension n is large, these estimators become nontrivial only for large N. If one assumes that the higher moments of the majorant are bounded, the estimators become efficient at values as low as $n \sim N^{2+\delta}$.

(6) Another possible procedure for localizing the root is used in the proof of Theorem 27.3.

(7) Unless restrictions are imposed on the compactness of the family and of the maximization domain, the optimality and even consistency of the maximum likelihood estimator become problematic. Linnik and Mitrofanova [1] give a very delicate proof that the maximum likelihood estimator is efficient for a certain subclass of the exponent families.

§28

(1) A brief account of the contents of this section was given in our note [8] (see also [12]).

(2) Of course, only the rate of decrease of the quantity $n \asymp \Gamma(N)$ is determined; n itself may be fixed arbitrarily within certain bounds.

(3) Recall that the indices for the information matrix in canonical coordinates s^j are written as subscripts: $\mathbf{v}_{jk}$. The notation $\mathbf{w}^{jk}$ is reserved for the information matrix in natural coordinates t_j.

§29

The definition of a random probability measure (on the unit interval or on the whole real line) by its random distribution function goes back to Kolmogorov [1]. Dubins and Freedman [1] systematically considered the definition of a random distribution function $F(t)$ as a monotone stochastic process on the real line, i.e. in terms of probabilities of quasi-intervals:

$$\mathfrak{P}\left\{a_i' < F(t_i) < a_i'', i = 1, \ldots, n\right\}.$$

These probabilities define a Baire distribution and, according to Kakutani [1] (see Halmos [1]), a unique regular Borel distribution, which is easily shown to be concentrated on monotone functions $F(t)$. However, as follows from a well-known result of the author [1], this distribution is generally not concentrated on distribution functions, since the set of all distribution functions is neither a Borel set nor absolutely measurable (although the set of continuous distribution functions has these properties). We therefore have to prove the existence of a non-Borel extension $\mathfrak{P}$, concentrated on the distribution functions, just as in the theory of Markov processes one has to prove the existence of a canonical modification of the process (see Itô [1], and also Doob [2]).

(1) The class of $\mathbf{B}_0(\mathbf{S})$-sets is fairly small. If the algebra $\mathbf{S}$ is uncountable, it does not even contain "singleton subfamilies". Since every family consisting of a single set function is closed in the product topology, they are all $\mathbf{B}(\mathbf{S})$-sets.

(2) If $\mathbf{S}$ has countably many generators, there is a substantial difference between the algebras Б$_0(\mathbf{S})$ and $\mathbf{B}_0(\mathbf{S})$. For example, all "singleton" families are measurable in Б$_0(\mathbf{S})$, since every σ-additive measure is completely determined by its values on countably many generators, while an arbitrary set function is determined by its values on the whole algebra $\mathbf{S}$.

(3) Each of conditions 1°–3° of system (2.5) (for fixed sets H, H_1 and H_2) describes a closed $\mathbf{B}_0(\mathbf{S})$-subset of the space $X^{\mathbf{S}}$. Hence axioms 1°–3°, which define the collection of all normalized finitely-additive set functions, describe a closed $\mathbf{B}(\mathbf{S})$-set as the intersection of the above-mentioned closed $\mathbf{B}_0(\mathbf{S})$-sets. The descriptive nature of the set of all normalized countably-additive set functions (i.e. probability measures) is as yet unknown, notwithstanding several interesting studies (see Freedman [1]).

By virtue of the well-known continuity property of any finite measure relative to a dominating measure, and the σ-additivity of any finitely-additive measure which is continuous relative to a σ-additive measure (cf. the proof of Lemma 14.1), the collection Capd(Ω, $\mathbf{S}$, $\mathbf{Z}$) is a family of the type $F_{\sigma\delta}$ in $X^{\mathbf{S}}$, and hence $\mathbf{B}(\mathbf{S})$-measurable, although it is not $\mathbf{B}_0(\mathbf{S})$-measurable.

(4) In view of conditions 1°–3°, condition 4° (convergence with probability one) may be replaced by convergence in probability:

For any $\varepsilon > 0$ *and every fixed sequence* $H_k \searrow \varnothing$, $H_k \in \mathbf{S}$,

$$\mathcal{P}\left\{\Phi\{H_k\} > \varepsilon\right\} \to 0.$$

(5) Note that Lebesgue collections Capd(Ω, $\mathbf{S}$, $\mathbf{Z}$) are not only $\mathbf{B}(\mathbf{S})$-measurable in $X^{\mathbf{S}}$ (see footnote 3, above) but also $\mathbf{K}$-measurable, also corresponding to an $F_{\sigma\delta}$-set in the sector W.

(6) An analogous statement is readily proved for other functionals (13.18) satisfying the conditions of Theorem 13.1.

BIBLIOGRAPHY

B. P. Adhikari and D. D. Joshi

1. *Distance, discrimination et résumé exhaustif*, Publ. Inst. Statist. Univ. Paris **5** (1956), 57–74.

P. S. Aleksandrov

1. *Combinatorial topology*, OGIZ, Moscow, 1947; English transl., vols. 1, 2, 3, Graylock Press, Albany, N. Y., 1956, 1957, 1960.
2. *Über die Urysohnschen Konstanten*, Fund. Math. **20** (1933), 140–150.

V. I. Averbuh and O. G. Smoljanov

1. *Differentiation theory in linear topological spaces*, Uspehi Mat. Nauk **22** (1967), no. 6 (138), 201–260; English transl. in Russian Math. Surveys **22** (1967).

R. R. Bahadur

1. *Sufficiency and statistical decision functions*, Ann. Math. Statist. **25** (1954), 423–462.

Stefan Banach and Casimir Kuratowski

1. *Sur une généralisation du problème de la mesure*, Fund. Math. **14** (1929), 127–130.

A. P. Basu

1. *Effect of truncation on a test for the scale parameter of the exponential distribution*, Ann. Math. Statist. **35** (1964), 209–213.

M. Bebutoff [M. V. Bebutov]

1. *Markov chains with a compact state space*, C. R. (Dokl.) Acad. Sci. URSS **30** (1941), 482–483.

Edwin F. Beckenbach and Richard Bellman

1. *Inequalities*, Springer-Verlag, Berlin, 1961.

Robert H. Berk

1. *A special group structure and equivariant estimation*, Ann. Math. Statist. **38** (1967), 1436–1445.

S. N. Bernstein

1. *Theory of probability*, 4th ed., Gostehizdat, Moscow, 1946. (Russian)

J. Bertrand

1. *Calcul des probabilités*, Gauthier-Villars, Paris, 1889.

A. Bhattacharyya

1. *On a measure of divergence between two statistical populations defined by their probability distributions*, Bull. Calcutta Math. Soc. **35** (1943), 99–109.
2. *On some analogues of the amount of information and their use in statistical estimation*. I, II, Sankhyā **8** (1946), 1–14, 201–218.

Herluf Bidstrup

1. *Gewitztes und Verschmitztes*, Eulenspiegel-Verlag, Berlin, 1955.

Garrett Birkhoff

1. *Lattice theory*, 2nd rev. ed., Amer. Math. Soc., Providence, R. I., 1948.

Allan Birnbaum

1. *On the foundations of statistical inference*, J. Amer. Statist. Assoc. **57** (1962), 269–306; discussion, 307–326.

Richard L. Bishop and Richard J. Crittenden

1. *Geometry of manifolds*, Academic Press, New York, 1964.

David Blackwell

1. *Conditional expectation and unbiased sequential estimation*, Ann. Math. Statist. **18** (1947), 105–110.
2. *Comparison of experiments*, Proc. Second Berkeley Sympos. Math. Statist. and Probability (1950), Univ. of California Press, Berkeley, Cal., 1951, pp. 93–102.
3. *Equivalent comparisons of experiments*, Ann. Math. Statist. **24** (1953), 265–272.
4. *On a class of probability spaces*, Proc. Third Berkeley Sympos. Math. Statist. and Probability (1954/55), Vol. II, Univ. of California Press, Berkeley, Cal., 1956, pp. 1–6.

David Blackwell and M. A. Girshick

1. *Theory of games and statistical decisions*, Wiley, New York; Chapman & Hall, London, 1954.

D. Blackwell and C. Ryll-Nardzewski

1. *Non-existence of everywhere proper conditional distributions*, Ann. Math. Statist. **34** (1963), 223–225.

J. R. Blum and Judah Rosenblatt

1. *On partial a priori information in statistical inference*, Ann. Math. Statist. **38** (1967), 1671–1678.

Colin R. Blyth

1. *On minimax statistical decision procedures and their admissibility*, Ann. Math. Statist. **22** (1951), 22–42.

Salomon Bochner and William Ted Martin

1. *Several complex variables*, Princeton Univ. Press, Princeton, N. J., 1948.

N. N. Bogoljubov

1. *Selected works in three volumes*, Vol. I, "Naukova Dumka", Kiev, 1969. (Russian)

L. N. Bol′šev

1. *A refinement of the Cramér-Rao inequality*, Teor. Verojatnost. i Primenen. **6** (1961), 319–326; English transl. in Theor. Probability Appl. **6** (1961).

Denis Bosq

1. *Sur l'estimation d'une densité multivariée par une série de fonctions orthogonales*, C. R. Acad. Sci. Paris Sér. A-B **268** (1969), A555–A557.
2. *Estimation non paramétrique de la densité et de ses dérivées*, C. R. Acad. Sci. Paris Sér. A-B **269** (1969), A1010–A1012.
3. *Complément à deux notes sur l'estimation de la densité et de ses dérivées*, C. R. Acad. Sci. Paris Sér. A-B **271** (1970), A45.

N. Bourbaki

1. *Fonctions d'une variable réelle (théorie élémentaire)*, Chaps. 1–3, Actualités Sci. Indust., no. 1074, Hermann, Paris, 1949.
2. *Espaces vectoriels topologiques*, Chaps. 1, 2, Actualités Sci. Indust., no. 1189, Hermann, Paris, 1953.

David R. Brillinger

1. *Necessary and sufficient conditions for a statistical problem to be invariant under a Lie group*, Ann. Math. Statist. **34** (1963), 492–500.

D. L. Burkholder

1. *On the order structure of the set of sufficient subfields*, Ann. Math. Statist. **33** (1962), 596–599.

N. N. Čencov

1. *Topological measures and the theory of random functions*, Proc. Fifth All-Union Conf. Theor. Probability and Math. Statist. (Erevan, 1958), Izdat. Akad. Nauk Armjan. SSR, Erevan, 1960, pp. 83–87. (Russian)
2. *On the asymptotic efficiency of the maximum likelihood estimator*, Proc. Sixth All-Union Conf. Theor. Probability and Math. Statist. (Vilnius, 1960), Gos. Izdat. Političesk. i Naučn. Lit. Litovsk. SSR, Vilnius, 1962, pp. 399–402. (Russian)

3. *Estimation of an unknown distribution density from observations*, Dokl. Akad. Nauk SSSR **147** (1962), 45–48; English transl. in Soviet Math. Dokl. **3** (1962); erratum in Soviet Math. Dokl. **4** (1963), no. 3, p. vi.
4. *The geometry of a "manifold" of probability distributions*, Dokl. Akad. Nauk SSSR **158** (1964), 543–546; English transl. in Soviet Math. Dokl. **5** (1964).
5. *The categories of mathematical statistics*, Dokl. Akad. Nauk SSSR **164** (1965), 511–514; English transl. in Soviet Math. Dokl. **6** (1965).
6. *Towards a systematic theory of exponential families of probability distributions*, Teor. Verojatnost. i Primenen. **11** (1966), 483–494; English transl. in Theor. Probability Appl. **11** (1966).
7. *Infinitesimal methods of mathematical statistics*, ms. no. 85–66, deposited at VINITI, 1966. (Russian) RŽ Mat. **1966** #10B79.
8. *Invariant loss functions in problems of mathematical statistics*, Uspehi Mat. Nauk **22** (1967), no. 1 (133), 178–180. (Russian)
9. *On estimating an unknown mean in a multivariate normal distribution*, Teor. Verojatnost. i Primenen. **12** (1967), 619–634; English transl. in Theor. Probability Appl. **12** (1967).
10. *A nonsymmetric distance between probability distributions; entropy and the Pithagorean theorem*, Mat. Zametki **4** (1968), 323–332; English transl. in Math. Notes **4** (1968).
11. *Pseudorandom numbers for the simulation of Markov chains*, Ž. Vyčisl. Mat. i Mat. Fiz. **7** (1967), 632–643; English transl. in USSR Comput. Math. and Math. Phys. **7** (1967).
12. *A general theory of statistical inference*, Doctoral Dissertation, Inst. Appl. Math. Acad. Sci. USSR, Moscow, 1968. (Russian)
13. *A general theory of statistical inference*, Mat. Zametki **5** (1969), 635–648; English transl. in Math. Notes **5** (1969).

Herman Chernoff

1. *A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations*, Ann. Math. Statist. **23** (1952), 493–507.
2. *Large-sample theory: parametric case*, Ann. Math. Statist. **27** (1956), 1–22.

Harald Cramér

1. *Mathematical methods of statistics*, Princeton Univ. Press, Princeton, N. J., 1946.
2. *A contribution to the theory of statistical estimation*, Skand. Aktuarietidskr. **29** (1946), 85–94.
3. *Sur un nouveau théorème-limite de la théorie des probabilités*, Actualités Sci. Indust., no. 736, Hermann, Paris, 1938, pp. 5–23.

Imre Csiszár

1. *On certain measures of divergence of probability distributions*, Abstracts of Brief Scientific Communications, Internat. Congr. Math., Moscow, 1966, Section 11, p. 7.
2. *Information-type indices of the divergence of distributions*. I, II, Magyar Tud. Akad. Mat. Fiz. Oszt. Közl. **17** (1967), 123–149, 267–291. (Hungarian; English summaries)
3. *Information-type measures of difference of probability distributions and indirect observations*, Studia Sci. Math. Hungar. **2** (1967), 299–318.
4. *On topological properties of f-divergences*, Studia Sci. Math. Hungar. **2** (1967), 329–339.
5. *Eine informationstheoretische Ungleichung und ihre Anwendung auf en Beweis der Ergodizität von Markoffschen Ketten*, Magyar Tud. Akad. Mat. Kutató Int. Közl. **8** (1963), 85–108.

Georges Darmois

1. *Sur les limites de la dispersion de certaines estimations*, Rev. Inst. Internat. Statist. **13** (1945), 9–15.

Morris H. DeGroot

1. *Sufficient experiments and the optimal allocation of observations*, Tech. Rep. no. 10, Carnegie Institute of Technology, Pittsburgh, Pa., 1964.

M. H. DeGroot and M. M. Rao

1. *Multidimensional information inequalities and prediction*, Multivariate Analysis (Proc. Internat. Sympos., Dayton, Ohio, 1965), Academic Press, New York, 1966, pp. 287–313.

J. Dieudonné

1. *Foundations of modern analysis*, Academic Press, New York, 1960.

J. L. Doob

1. *Stochastic processes*, Wiley, New York; Chapman & Hall, London, 1953.
2. *Application of the theory of martingales*, Colloques Internat. CNRS, no. 13, Centre Nat. Recherche Sci., Paris, 1949, pp. 23–27.

S. D. A. C. Doss

1. *On uniqueness and maxima of the roots of likelihood equations under truncated and censored sampling from normal populations*, Sankhyā Ser. A **24** (1962), 355–362.

Lester E. Dubins and David A. Freedman

1. *Random distribution functions*, Bull. Amer. Math. Soc. **69** (1963), 548–551.

Daniel Dugué

1. *Application des propriétés de la limite au sens du calcul des probabilités à l'étude de diverses questions d'estimation*, J. École Polytech. (3) **1937**, 305–373.

E. B. Dynkin

1. *Necessary and sufficient statistics for a family of probability distributions*, Uspehi Mat. Nauk **6** (1951), no. 1 (41), 68–90; English transl. in Selected Transl. Math. Statist. and Probability, Vol. 1, Amer. Math. Soc., Providence, R. I., 1961.

Samuel Eilenberg and Saunders Mac Lane

1. *General theory of natural equivalences*, Trans. Amer. Math. Soc. **58** (1945), 231–294.

Samuel Eilenberg and Norman Steenrod

1. *Foundations of algebraic topology*, Princeton Univ. Press, Princeton, N. J., 1952.

J. Favard

1. *Cours de géométrie différentielle locale*, Gauthier-Villars, Paris, 1957.

Werner Fenchel

1. *On conjugate convex functions*, Canad. J. Math. **1** (1949), 73–77.

Ronald Aylmer Fisher

1. *On the mathematical foundations of theoretical statistics*, Philos. Trans. Roy. Soc. London Ser. A **222** (1922), 309–368.
2. *Theory of statistical estimation*, Proc. Cambridge Philos. Soc. **22** (1924/25), 700–725.
3. *Statistical methods for research workers*, 10th ed., Oliver and Boyd, London, 1946.

D. A. S. Fraser

1. *On local unbiased estimation*, J. Roy. Statist. Soc. Ser. B **26** (1964), 46–51.
2. *Statistical models and invariance*, Ann. Math. Statist. **38** (1967), 1061–1067.

Maurice Fréchet

1. *Sur l'extension de certaines évaluations statistiques au cas de petits échantillons*, Rev. Inst. Internat. Statist. **11** (1943), 182–205.

David A. Freedman

1. *On two equivalence relations between measures*, Ann. Math. Statist. **37** (1966), 686–689.

A. S. Frolov and N. N. Čencov

1. *Use of dependent tests in the Monte Carlo method for obtaining smooth curves*, Proc. Sixth All-Union Conf. Theor. Probability and Math. Statist. (Vilnius, 1960), Gos. Izdat. Političesk. i Naučn. Lit. Litovsk. SSR, Vilnius, 1962, pp. 425–437. (Russian)

B. A. Fuks

1. *Introduction to the theory of analytic functions of several complex variables*, Fizmatgiz, Moscow, 1962; English transl., Amer. Math. Soc., Providence, R. I., 1963.

C. F. Gauss

1. *Abhandlungen zur Methode der kleinsten Quadrate*, P. Stavkiewicz, Berlin, 1887.

I. M. Gel'fand

1. *Lectures on linear algebra*, 3rd ed., "Nauka", Moscow, 1966; English transl. of 2nd ed., Interscience, New York, 1961.
2. *Generalized random processes*, Dokl. Akad. Nauk SSSR **100** (1955), 853–856. (Russian)

I. M. Gel'fand and S. V. Fomin

1. *Calculus of variations*, Fizmatgiz, Moscow, 1961; English transl., Prentice-Hall, Englewood Cliffs, N. J., 1963.

I. M. Gel'fand, A. S. Frolov and N. N. Čencov

1. *The computation of functional integrals by the Monte Carlo method*, Izv. Vysš. Učebn. Zaved. Matematika **1958**, no. 5(6), 32–45. (Russian)

I. M. Gel'fand and M. I. Graev

1. *Geometry of homogeneous spaces, representations of groups in homogeneous spaces and related questions of integral geometry*. I, Trudy Moskov. Mat. Obšč. **8** (1959), 321–390; English transl. in Amer. Math. Soc. Transl. (2) **37** (1964).

I. M. Gel'fand, A. N. Kolmogorov and A. M. Jaglom

1. *On the general definition of the amount of information*, Dokl. Akad. Nauk SSSR **111** (1956), 745–748; German transl. in Arbeiten zur Informationstheorie. II, VEB Deutscher Verlag der Wissenschaften, Berlin, 1958, pp. 57–60.

S. G. Ghurye

1. *Information and sufficient sub-fields*, Ann. Math. Statist. **39** (1968), 2056–2066.

M. A. Girshick and L. J. Savage

1. *Bayes and minimax estimates for quadratic loss functions*, Proc. Second Berkeley Sympos. Math. Statist. and Probability (1950), Univ. of California Press, Berkeley, Cal., 1951, pp. 53–73.

B. V. Gnedenko and A. N. Kolmogorov

1. *Limit distributions for sums of independent random variables*, GITTL, Moscow, 1949; English transl., Addison-Wesley, Reading, Mass., 1954; rev. ed., 1968.

Paul R. Halmos

1. *Measure theory*, Van Nostrand, Princeton, N. J., 1950.

Paul R. Halmos and L. J. Savage

1. *Application of the Radon-Nikodým theorem to the theory of sufficient statistics*, Ann. Math. Statist. **20** (1949), 225–241.

J. A. Hartigan

1. *The likelihood and invariance principles*, J. Roy. Statist. Soc. Ser. B **29** (1967), 533–539.

Felix Hausdorff

1. *Mengenlehre*, Veit, Leipzig, 1914; English transl. of 3rd ed., Chelsea, New York, 1957, 1962.

Sigurdur Helgason

1. *Differential geometry and symmetric spaces*, Academic Press, New York, 1962.

Einar Hille

1. *Functional analysis and semi-groups*, Amer. Math. Soc., Providence, R. I., 1948.

J. L. Hodges, Jr., and E. L. Lehmann

1. *Some problems in minimax point estimation*, Ann. Math. Statist. **21** (1950), 182–197.
2. *Some applications of the Cramér-Rao inequality*, Proc. Second Berkeley Sympos. Math. Statist. and Probability (1950), Univ. of California Press, Berkeley, Cal., 1951, pp. 13–22.

Wassily Hoeffding

1. *The role of assumptions in statistical decisions*, Proc. Third Berkeley Sympos. Math. Statist. and Probability (1954/55), vol. I, Univ. of California Press, Berkeley, Cal., 1956, pp. 105–114.

V. S. Huzurbazar

1. *Exact forms of some invariants for distributions admitting sufficient statistics*, Biometrika **42** (1955), 533–537.

A. D. Ioffe and V. M. Tihomirov

1. *The duality of convex functions and extremal problems*, Uspehi Mat. Nauk **23** (1968), no. 6 (144), 51–116; English transl. in Russian Math. Surveys **23** (1968).

Alexander Ionescu Tulcea and Cassius Ionescu Tulcea

1. *On the lifting property*. I, J. Math. Anal. Appl. **3** (1961), 537–546.

C. Ionescu Tulcea

1. *On the lifting property and disintegration of measures*, Bull. Amer. Math. Soc. **71** (1965), 829–842.

C. T. Ireland and S. Kullback

1. *Minimum discrimination information estimation*, Biometrics **24** (1968), 707–713.

Kiyoshi Itô

1. *The canonical modification of stochastic processes*, J. Math. Soc. Japan **20** (1968), 130–150.

W. James and Charles Stein

1. *Estimation with quadratic loss*, Proc. Fourth Berkeley Sympos. Math. Statist. and Probability (1960), vol. I, Univ. of California Press, Berkeley, Cal., 1961, pp. 361–379.

Devi Datt Joshi

1. *L'information en statistique mathématique et dans la théorie des communications*, Publ. Inst. Statist. Univ. Paris **8** (1959), 81–159.

A. M. Kagan

1. *On the theory of Fisher's amount of information*, Dokl. Akad. Nauk SSSR **151** (1963), 277–278; English transl. in Soviet Math. Dokl. **4** (1963).
2. *Families of distributions and separating partitions*, Dokl. Akad. Nauk SSSR **153** (1963), 522–525; English transl. in Soviet Math. Dokl. **4** (1963).
3. *Remarks on separating partitions*, Trudy Mat. Inst. Steklov. **79** (1965), 26–31; English transl. in Proc. Steklov Inst. Math. **79** (1965).
4. *Estimation theory for families with shift or scale parameters and for exponent families*, Dissertation, Leningrad State Univ., Leningrad, 1967. (Russian)

Shizuo Kakutani

1. *Construction of a non-separable extension of the Lebesgue measure space*, Proc. Imp. Acad. Tokyo **20** (1944), 115–119.

B. K. Kale

1. *Maximum likelihood estimation for truncated exponential family*, J. Indian Statist. Assoc. **1** (1963), 86–90.

Gopinath Kallianpur

1. *On the amount of information contained in a σ-field*, Contributions to Probability and Statistics (Essays in Honor of Harold Hotelling), Stanford Univ. Press, Stanford, Cal., 1963, pp. 265–273.

L V. Kantorovič, B. Z. Vulih and A. G. Pinsker

1. *Functional analysis in partially ordered spaces*, GITTL, Moscow, 1950. (Russian)

Samuel Karlin

1. *Admissibility for estimation with quadratic loss*, Ann. Math. Statist. **29** (1958), 406–436.

F. I. Karpelevič

1. *Geometry of geodesics and eigenfunctions of the Beltrami-Laplace operator on symmetric spaces*, Doctoral Dissertation, Inst. Appl. Math. Acad. Sci. USSR, Moscow, 1963. (Russian)

John L. Kelley

1. *General topology*, Van Nostrand, Princeton, N. J., 1955.

A. Ja. Khintchine

1. *Mathematical foundations of statistical mechanics*, OGIZ, Moscow, 1943; English transl., Dover, New York, 1949.

J. Kiefer

1. *Multivariate optimality results*, Multivariate Analysis (Proc. Internat. Sympos., Dayton, Ohio, 1965), Academic Press, New York, 1966, pp. 255–274.

Felix Klein

1. *Das Erlanger Programm: Vergleichende Betrachtungen über neuere geometrische Forschungen*, A. Deichert, Erlangen, 1872; reprint, Ostwalds Klassiker Exakt. Wiss., no. 253, Akademische Verlag; Geest & Portiz, Leipzig, 1974; French transl., Gauthier-Villars, Paris, 1974.

A. N. Kolmogorov

1. *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Springer-Verlag, Berlin, 1933; reprint, 1973; English transl., Chelsea, New York, 1948; 2nd ed., 1956.
2. *Über die beste Annäherung von Funktionen einer gegebenen Funktionenklasse*, Ann. of Math. (2) **37** (1936), 107–110.
3. *On the proof of the method of least squares*, Uspehi Mat. Nauk **1** (1946), no. 1 (11), 57–70. (Russian)
4. *Unbiased estimates*, Izv. Akad. Nauk SSSR Ser. Mat. **14** (1950), 303–326; English transl. in Amer. Math. Soc. Transl. (1) **11** (1962).
5. *Three approaches to the definition of the concept of the "amount of information"*, Problemy Peredači Informacii **1** (1965), no. 1, 3–11; English transl. in Selected Transl. Math. Statist. and Probability, vol. 7, Amer. Math. Soc., Providence, R. I., 1968.

B. O. Koopman

1. *On distributions admitting a sufficient statistic*, Trans. Amer. Math. Soc. **39** (1936), 399–409.

V. P. Kozlov

1. *Set capacity in signal space and the Riemannian metric*, Dokl. Akad. Nauk SSSR **166** (1966), 779–782; English transl. in Soviet Math. Dokl. 7 (1966).

M. G. Kreĭn and M. A. Rutman

1. *Linear operators leaving invariant a cone in a Banach space*, Uspehi Mat. Nauk **3** (1948), no. 1 (23), 3–95; English transl. in Amer. Math. Soc. Transl. (1) **10** (1962).

T. A. Kriz and J. V. Talacko

1. *Equivalence of the maximum likelihood estimator to a minimum entropy estimator*, Trabajos Estadíst. **19** (1968), 55–65.

Richard Kronmal and Michael Tarter

1. *The estimation of probability densities and cumulatives by Fourier series methods*, J. Amer. Statist. Assoc. **63** (1968), 925–952.

Solomon Kullback

1. *An application of information theory to multivariate analysis*. I, II, Amer. Math. Statist. **23** (1952), 88–102; **27** (1956), 122–146.
2. *Certain inequalities in information theory and the Cramér-Rao inequality*, Ann. Math. Statist. **25** (1954), 745–751.
3. *Information theory and statistics*, Wiley, New York; Chapman & Hall, London, 1959.

S. Kullback and M. A. Khairat

1. *A note on minimum discrimination information*, Ann. Math. Statist. **37** (1966), 279–280.

S. Kullback and R. A. Leibler

1. *On information and sufficiency*, Ann. Math. Statist. **22** (1951), 79–86.

A. G. Kuroš, A. H. Livšic and E. G. Šul'geifer

1. *Foundations of the theory of categories*, Uspehi Mat. Nauk **15** (1960), no. 6 (96), 3–52; English transl. in Russian Math. Surveys **15** (1960).

Serge Lang

1. *Algebra*, Addison-Wesley, Reading, Mass., 1965.

Pierre Simon, Marquis de Laplace

1. *Théorie analytique des probabilités*, 3rd ed., V. Courcier, Paris, 1820; reprint of 1812 ed., Culture et Civilisation, Brussels, 1967.

F. Laurent, M. Oheix and J.-P. Raoult

1. *Tests d'hypothèses*, Ann. Inst. H. Poincaré Sect. B **5** (1969), 385–414.

Henri Lebesgue

1. *Leçons sur l'intégration et la recherche des fonctions primitives*, 2nd ed., Gauthier-Villars, Paris, 1928.

Lucien LeCam

1. *On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates*, Univ. California Publ. Statist. **1** (1953), 277–329.
2. *Sufficiency and approximate sufficiency*, Ann. Math. Statist. **35** (1964), 1419–1455.
3. *Théorie asymptotique de la décision statistique*, Sem. Math. Sup., No. 33 (Été, 1968), Presses Univ. Montréal, Montréal, 1969.

E. L. Lehmann

1. *Testing statistical hypotheses*, Wiley, New York; Chapman & Hall, London, 1959.

Tullio Levi-Civita

1. *Nozione di parallelismo in una varietà qualunque e conseguente specificazione geometrica della curvatura Riemanniana*, Rend. Circ. Mat. Palermo **42** (1917), 173–205.

Paul Lévy

1. *Processus stochastiques et mouvement brownien*, 2nd ed., Gauthier-Villars, Paris, 1965.

Ju. V. Linnik

1. *The method of least squares and the foundations of the mathematico-statistical theory of reduction of observations*, Fizmatgiz, Moscow, 1958; English transl., Pergamon Press, New York, 1961.
2. *Statistical problems with nuisance parameters*, "Nauka", Moscow, 1966; English transl., Amer. Math. Soc., Providence, R. I., 1968.
3. *Leçons sur les problèmes de statistique analytique*, Gauthier-Villars, Paris, 1967.

Ju. V. Linnik and N. M. Mitrofanova

1. *On the asymptotic behavior of the maximum likelihood distribution*, Dokl. Akad. Nauk SSSR **149** (1963), 518–520; English transl. in Soviet Math. Dokl. **4** (1963).

M. Littaye-Petit, J.-L. Piednoir and B. Van Cutsem

1. *Exhaustivité*, Ann. Inst. H. Poincaré Sect. B **5** (1969), 289–322.

A. A. Ljapunov

1. *On choosing from a finite number of distribution laws*, Uspehi Mat. Nauk **6** (1951), no. 1 (41), 178–186. (Russian)

Michel Loève

1. *Probability theory*, Van Nostrand, Princeton, N. J., 1955; 2nd rev. ed., 1960.

N. N. Lusin [Luzin]

1. *Sur les propriétés de fonctions mesurables*, C. R. Acad. Sci. Paris **154** (1912), 1688–1690.

N. Lusin and P. Novikoff [N. N. Luzin and P. S. Novikov]

1. *Choix effectif d'un point dans un complémentaire analytique arbitraire, donné par un crible*, Fund. Math. **25** (1935), 559–560.

N. Lusin [N. N. Luzin] and W. Sierpiński

1. *Sur quelques propriétés des ensembles* (A), Bull. Internat. Acad. Sci. Cracovie Cl. Sci. Math. Nat. Ser. A: Sci. Math. **1918**, 35–48.

George W. Mackey

1. *The mathematical foundations of quantum mechanics*, Benjamin, New York, 1963.

Françoise Martin and Madeleine Oheix

1. *Sur la notion d'information d'un modèle statistique bayésien*, C. R. Acad. Sci. Paris Sér. A-B **268** (1969), A735–A737.

Françoise Martin, Jean-Luc Petit and Monique Petit-Littaye [Littaye-Petit]

1. *Comparison des expériences*, Ann. Inst. H. Poincaré Sect. B. **7** (1971), 145–176.

Françoise Martin and Daniel Vaguelsy

1. *Propriétés asymptotiques du modèle statistique*, Ann. Inst. H. Poincaré Sect. B. **5** (1969), 357–384.

A. M. Mathai

1. *Some characterizations of the one-parameter family of probability distributions*, Canad. Math. Bull. **9** (1966), 95–102.

George J. Minty

1. *On the monotonicity of the gradient of a convex function*, Pacific J. Math. **14** (1964), 243–247.

Norman Morse and Richard Sacksteder

1. *Statistical isomorphism*, Ann. Math. Statist. **37** (1966), 203–214.

John von Neumann

1. *Zur Theorie der Gesellschaftsspiele*, Math. Ann. **100** (1928), 295–320; English transl. in Contributions to the Theory of Games, Vol. IV, Ann. of Math. Studies, no. 40, Princeton Univ. Press, Princeton, N. J., 1959.

John von Neumann and Oskar Morgenstern

1. *Theory of games and economic behavior*, Princeton Univ. Press, Princeton, N. J., 1944; 3rd ed., 1953.

Jacques Neveu

1. *Bases mathématiques du calcul des probabilités*, Masson, Paris, 1964; English transl., Holden-Day, San Francisco, Cal., 1965.

J. Neyman

1. *L'estimation statistique traitée comme un problème classique de probabilité*, Actualités Sci. Indust., no. 793, Hermann, Paris, 1938, pp. 25–57.
2. *Two breakthroughs in the theory of statistical decision making*, Rev. Inst. Internat. Statist. **30** (1962), 11–27.

J. Neyman and E. S. Pearson

1. *On the use and interpretation of certain test criteria for purposes of statistical inference*, Biometrika **20A** (1928), 175–240.

A. P. Norden

1. *On intrinsic geometry of the second kind on a hypersurface of affine space*, Appendix to the book of P. A. and A. P. Širokov, *Affine differential geometry*, Fizmatgiz, Moscow, 1959, pp. 275–291; German transl., Teubner, Leipzig, 1962.

Pierre Novikoff [P. S. Novikov]

1. *Sur les fonctions implicites mesurables* B, Fund. Math. **17** (1931), 8–25.

Octav Onicescu

1. *Énergie informationnelle*, C. R. Acad. Sci. Paris Sér. A-B **263** (1966), A841–A842.

V. P. Palamodov

1. *Linear differential operators with constant coefficients*, "Nauka", Moscow, 1967; English transl., Springer-Verlag, Berlin and New York, 1970.

Emanuel Parzen

1. *On estimation of a probability density function and mode*, Ann. Math. Statist. **33** (1962), 1065–1076.

Albert Perez

1. *Sur l'énergie informationnelle de M. Octav Onicescu*, Rev. Roumaine Math. Pures Appl. **12** (1967), 1341–1347.

M. M. Postnikov

1. *The variational theory of geodesics*, "Nauka", Moscow, 1965; English transl., Saunders, Philadelphia, Pa., 1967.

Ju. V. Prohorov

1. *Random measures on a compactum*, Dokl. Akad. Nauk SSSR **138** (1961), 53–55; English transl. in Soviet Math. Dokl. **2** (1961).

Bayard Rankin

1. *Computable probability spaces*, Acta Math. **103** (1960), 89–122.

C. Radhakrishna Rao

1. *Information and the accuracy attainable in the estimation of statistical parameters*, Bull. Calcutta Math. Soc. **37** (1945), 81–91.
2. *Minimum variance and the estimation of several parameters*, Proc. Cambridge Philos. Soc. **43** (1947), 280–283.
3. *Efficient estimates and optimal inference procedures in large samples*, J. Roy. Statist. Soc. Ser. B **24** (1962), 46–63; discussion, 63–72.

M. M. Rao

1. *Theory of lower bounds for risk functions in estimation*, Math. Ann. **143** (1961), 379–398.

Alfréd Rényi

1. *On measures of entropy and information*, Proc. Fourth Berkeley Sympos. Math. Statist. and Probability (1960), Vol. I, Univ. of California Press, Berkeley, Cal., 1961, pp. 547–561.

Herbert E. Robbins

1. *An empirical Bayes approach to statistics*, Proc. Third Berkeley Sympos. Math. Statist. and Probability (1954/55), Vol. I, Univ. of California Press, Berkeley, Cal., 1956, pp. 157–163.
2. *The empirical Bayes approach to statistical decision problems*, Ann. Math. Statist. **35** (1964), 1–20.

R. Tyrrell Rockafellar

1. *Characterization of the subdifferentials of convex functions*, Pacific J. Math. **17** (1966), 497–510.

V. A. Rohlin

1. *On the fundamental concepts of measure theory*, Mat. Sb. **25 (67)** (1949), 107–150; English transl. in Amer. Math. Soc. Transl. (1) **10** (1962).

Guy Romier

1. *Exhaustivité et équivalence des objets statistiques*, C. R. Acad. Sci. Paris Sér. A-B **267** (1968), A828–A831.
2. *Modèle d'expérimentation statistique*, Ann. Inst. H. Poincaré Sect. B **5** (1969), 275–288.
3. *Décision statistique*, Ann. Inst. H. Poincaré Sect. B **5** (1969), 323–355.

Murray Rosenblatt

1. *Remarks on some nonparametric estimates of a density function*, Ann. Math. Statist. **27** (1956), 832–837.

M. Rozenblat-Rot [Millu Rosenblatt-Roth]

1. *The concept of entropy in probability theory and its applications in the theory of transmission of information over communication channels*, Teor. Verojatnost. i Primenen. **9** (1964), 238–261; English transl. in Theor. Probability Appl. **9** (1964).

B. A. Rozenfel'd

1. *Non-Euclidean geometries*, GITTL, Moscow, 1955. (Russian)

Richard Sacksteder

1. *A note on statistical equivalence*, Ann. Math. Statist. **38** (1967), 787–794.
2. *On products of experiments*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **10** (1968), 203–211.

Stanislaw Saks

1. *Theory of the integral*, 2nd rev. ed., PWN, Warsaw; Stechert, New York, 1937; reprint, Dover, New York, 1964.

O. N. Šalaevskiĭ

1. *A short proof of the Cramér-Rao inequality*, Teor. Verojatnost. i Primenen. **6** (1961), 352–353; English transl. in Theor. Probability Appl. **6** (1961).

I. N. Sanov

1. *On the probability of large deviations of random variables*, Mat. Sb. **42** **(84)** (1957), 11–44; English transl. in Selected Transl. Math. Statist. and Probability, vol. 1, Amer. Math. Soc., Providence, R. I., 1961.

Leonard J. Savage

1. *The foundations of statistics*, Wiley, New York; Chapman & Hall, London, 1954.

V. V. Sazonov

1. *On perfect measures*, Izv. Akad. Nauk SSSR Ser. Mat. **26** (1962), 391–414; English transl. in Amer. Math. Soc. Transl. (2) **48** (1965).

Carel L. Scheffer

1. *Limits of directed projective systems of probability spaces*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **13** (1969), 60–80.

Leopold Schmetterer

1. *On superefficiency*, Sympos. Probability Methods in Analysis (Loutraki, Greece, 1966), Lecture Notes in Math., vol. 31, Springer-Verlag, Berlin and New York, 1967, pp. 291–295.

Eugene F. Schuster

1. *Estimation of a probability density function and its derivatives*, Ann. Math. Statist. **40** (1969), 1187–1195.

Stuart C. Schwartz

1. *Estimation of probability density by an orthogonal series*, Ann. Math. Statist. **38** (1967), 1261–1265.

Anton Şesan, Silvia Covali, N. Gheorghiu and Adrian Vulpe

1. *Problèmes de probabilité dans le calcul aux états limites*. I, Bul. Inst. Politehn. Iaşi **6** **(10)** (1960), 375–380. (Romanian; French summary)

G. R. Seth

1. *On the variance of estimates*, Ann. Math. Statist. **20** (1949), 1–27.

Seymour Sherman

1. *On a theorem of Hardy, Littlewood, Pólya, and Blackwell*, Proc. Nat. Acad. Sci. U.S.A. **37** (1951), 826–831; erratum, **38** (1952), 382.

A. F. Sizova

1. *Estimating the fission density in a spherical reactor*, The Monte Carlo Method in Problems of Radiation Transfer, Atomizdat, Moscow, 1967, pp. 228–231. (Russian)

A. V. Skorohod
1. *Constructive methods of defining random processes*, Uspehi Mat. Nauk **20** (1965), no. 3 (123), 67–87; English transl. in Russian Math. Surveys **20** (1965).

N. V. Smirnov
1. *On the construction of a confidence region for the density of distribution of random variables*, Dokl. Akad. Nauk SSSR **74** (1950), 189–191. (Russian)
2. *On approximating the density of the distribution of a random variable*, Učen. Zap. Moskov. Gorod. Ped. Inst. im. Potemkin. **16** (1951), 69–96. (Russian)

Charles M. Stein
1. *Inadmissibility of the usual estimator for the means of a multivariate normal distribution*, Proc. Third Berkeley Sympos. Math. Statist. and Probability (1954/55), Vol. I, Univ. of California Press, Berkeley, Cal., 1956, pp. 197–206.
2. *Confidence sets for the mean of a multivariate normal distribution*, J. Roy. Statist. Soc. Ser. B **24** (1962), 265–296.
3. *Comparison of experiments*, Mimeographed Notes, Univ. of Chicago, Chicago, Ill., 1951.

P. L. Tchebycheff
1. *Attempt at elementary analysis of probability theory*, Master's Dissertation, Moscow Univ., Moscow, 1845; reprinted in his *Selected works*, Izdat. Akad. Nauk SSSR, Moscow, 1955, pp. 111–189, and his *Complete collected works*, vol. V, Izdat. Akad. Nauk SSSR, Moscow, 1951, pp. 26–87. (Russian)

James R. Thompson
1. *Accuracy borrowing in the estimation of the mean by shrinkage to an interval*, J. Amer. Statist. Assoc. **63** (1968), 953–963.

V. M. Tihomirov
1. *Widths of sets in functional spaces and the theory of best approximations*, Uspehi Mat. Nauk **15** (1960), no. 3 (93), 81–120; English transl. in Russian Math. Surveys **15** (1960).

John W. T. Tukey
1. *Sufficiency, truncation and selection*, Ann. Math. Statist. **20** (1949), 309–311.
2. *Curves as parameters, and touch estimation*, Proc. Fourth Berkeley Sympos. Math. Statist. and Probability (1960), Vol. I, Univ. of California Press, Berkeley, Cal., 1961, pp. 681–694.

S. H. Tumanjan
1. *On the maximum deviation of the empirical density of a distribution*, Ěrevan. Gos. Univ. Naučn. Trudy **48** (Ser. Fiz.-Mat. Nauk no. 2) (1955), 3–48. (Russian) RŽ Mat. **1957** #692.

J. R. Van Ryzin
1. *Bayes risk consistency of classification procedures using density estimation*, Sankhyā Ser. A **28** (1966), 261–270.

V. G. Vinokurov
1. *Generalized Lebesgue spaces*, Dokl. Akad. Nauk SSSR **129** (1959), 9–11. (Russian)

D. A. Vladimirov

1. *Boolean algebras*, "Nauka", Moscow, 1969; German transl., Akademie-Verlag, Berlin, 1972.

V. S. Vladimirov

1. *Methods of the theory of functions of many complex variables*, "Nauka", Moscow, 1964; English transl., MIT Press, Cambridge, Mass., 1966.
2. *The Legendre transformation of convex functions*, Mat. Zametki **1** (1967), 675–682; English transl. in Math. Notes **1** (1967).

N. N. Vorob'ev and D. K. Faddeev

1. *Continualization of conditional probabilities*, Teor. Verojatnost. i Primenen. **6** (1961), 116–118; English transl. in Theor. Probability Appl. **6** (1961).

Georges Vranceanu and Georges-Georges Vranceanu

1. *Probabilités et transport parallèle*, C. R. Acad. Sci. Paris **255** (1962), 40–41.

B. L. van der Waerden

1. *Mathematische Statistik*, Springer-Verlag, Berlin, 1957; English transl. of 2nd ed., Die Grundlehren der Math. Wiss., Bd. 156, Springer-Verlag, New York-Heidelberg, 1969.

Abraham Wald

1. *Contributions to the theory of statistical estimation and testing hypotheses*, Ann. Math. Statist. **10** (1939), 299–326.
2. *Note on the consistency of the maximum likelihood estimate*, Ann. Math. Statist. **20** (1949), 595–601.
3. *Statistical decision functions*, Wiley, New York; Chapman & Hall, London, 1950.

Geoffrey S. Watson

1. *Density estimation by orthogonal series*, Ann. Math. Statist. **40** (1969), 1496–1498.

Hermann Weyl

1. *Reine Infinitesimalgeometrie*, Math. Z. **2** (1918), 384–411.

Peter Whittle

1. *On the smoothing of probability density functions*, J. Roy. Statist. Soc. Ser. B **20** (1958), 334–343.

Norbert Wiener

1. *Nonlinear problems in random theory*, Wiley, New York; Chapman & Hall, London, 1958.

J. Wolfowitz

1. *Minimax estimates of the mean of a normal distribution with known variance*, Ann. Math. Statist. **21** (1950), 218–230.
2. *The minimum distance method*, Ann. Math. Statist. **28** (1957), 75–88.

R. F. Wrighton

1. *The problem of statistical inference*. III, Acta Genetica et Statistica Medica **18** (1968), 84–96.

This page intentionally left blank

INDEX

ISBN 978-0-8218-1347-8
9 780821 813478
MMONO/53.S