The Annals of Statistics
1990, Vol. 18, No. 3, 987–998

THE UNFATHOMABLE INFLUENCE OF KOLMOGOROV

BY NIKOLAI N. CHENTSOV

Keldysh Institute of Applied Mathematics

A. N. Kolmogorov contributed much to mathematical statistics, both through his own statistical works and through the researches of his students and disciples in statistics. But the case will be made that he influenced statistics still more through his general mathematical concepts and approaches. This paper gives some examples of unexpected applications of Kolmogorov's ideas in topics of nonparametric statistics.

Andrei N. Kolmogorov (1903–1987) was an outstanding mathematician of the twentieth century, one of the leaders in several branches of pure and applied mathematics and, especially, in probability theory, where he developed the set-theoretic basis of the subject and contributed to the origin and development of many new subfields.

Among the personal contributions of Kolmogorov to mathematical statistics one should recall first of all his famous paper [Kolmogorov (1933b)] where the limit distribution of the normed deviation $\sqrt{N}\sup_x|F_N(x) - F(x)|$ of the empirical distribution function $F_N(x)$ from a continuous theoretic one $F(x)$ was determined. Conversion of the problem statement, where the theoretical distribution is unknown and the empirical one is given, leads to the Kolmogorov goodness-of-fit test which is widely used now in nonparametric statistics, along with the test of Kolmogorov's disciple, N. V. Smirnov [see Smirnov (1939a, b)] and the Cramér–von Mises test [see von Mises (1931), Cramér (1928) and Smirnoff (1936, 1937)], where the above deviation is measured in a somewhat different way.

It is not so widely known that there were periods in Kolmogorov's life when applied statistics was his main occupation. I should mention first of all the years of the Second World War when he solved some statistical problems of the theory of artillery fire [Kolmogorov (1942, 1945a, b). Later, in 1948–1951, Kolmogorov worked on problems of quality control [see Kolmogorov (1950, 1951)] and queueing theory. It is said that it was Kolmogorov who advised, during the hard days of 1941, how the barrage balloons should be quasistochastically distributed to make it difficult for the Nazis to bomb Moscow. Just at that time an incident occurred that could have had a tragic end for Kolmogorov. One stuffy summer evening he got out of the local train dressed in an elegant white suit with a rucksack on his back and walked towards his country house. He was soon stopped by a patrol who assumed that a white spot against the asphalt background might serve as a reference point disclosing

Received June 1989; revised January 1990.

AMS 1980 *subject classifications*. Primary 01A60; secondary 62G99.

Key words and phrases. Kolmogorov–Smirnov test, density estimation, statistical point estimation problem, Kolmogorov width.

the highway to enemy aircraft. Then A. N. removed his suit, put it into his rucksack and went further in shorts.

The next patrol arrested him as a suspected saboteur–paratrooper. He was immediately transported to the Security Service Building. Fortunately, the security officer called to the duty official of the Academy of Sciences with the question: "Find out right away if Academician Kolmogorov is alive. A suspicious-looking man in shorts with the identity card of Academician Kolmogorov has just been arrested." Kolmogorov liked to hike. He usually went in shorts with a rucksack on his back, not only on summer walking tours but also during winter ski ones. Lazar A. Lusternik (member-correspondent of the USSR Academy of Sciences and the then deputy secretary of the Department of Physical and Mathematical Sciences) knew this very well. So he replied to the security officer: "Seems to me that man in shorts is Academician Kolmogorov himself... ." From time to time Kolmogorov and Lusternik told what had happened, the Lubyanka Building being usually left out of the story.

We may also mention an early work, Kolmogorov (1931), where he determined for which distribution types the sample median is preferable to the sample mean as an estimate of a location parameter. There are also several papers on the analysis of variance [Kolmogorov (1946, 1949b) and Kolmogorov, Petrov and Smirnov (1947)] that were unknown to Western experts and that anticipated parts of Scheffé (1959).

Kolmogorov was constantly interested in the usefulness of probability and statistics methods in other branches of knowledge: linguistics, physics, meteorology, geology and, especially, biology. He did not keep aloof in 1939 when the theory of Lysenko (which denied the theses of classical genetics, including Mendel's laws) was officially supported in Soviet biology. Instead, Kolmogorov analyzed by the χ^2-test the statistical data which were used by Lysenko supporters to refute the law of $1:3$ ratios. Kolmogorov (1940) showed the deviations of their observed data from this ratio to be nonsignificant. It was rather dangerous to make such statements at that time. I recall (though I am not sure) that in the fifties he carried out some other studies of the same kind but they were not published.

However, the influence of Kolmogorov on the development of mathematical statistics cannot be reduced to the effects of his statistical works or of the studies of his followers in applied statistics. First and foremost, it was the influence of the mathematician which determined both the trend of theoretical studies [see, for example, Kolmogorov (1949a)] and their mathematical level. His ideas and conceptions of a general mathematical nature seemed to have the most striking effect upon the development of our province of science. In fact, we were captivated by these ideas and conceptions. I might be mistaken but this is my personal perception of A. N. Kolmogorov and I should like to describe how some of my own work was related to his.

I was a disciple of Kolmogorov in the second generation: Eugen B. Dynkin was my mentor during all my years of studying at Moscow University; after 1952 when I started working at the Steklov Mathematical Institute I simultaneously took a postgraduate course under the guidance of N. V. Smirnov. My personal scientific contacts with Kolmogorov were minimal.

Functional approaches to the calculation of the limit distributions of tests [see the review in Kolmogorov and Prohorov (1956)] acquired popularity in the fifties. In particular, the justification of the heuristic approaches of Doob (1947, 1949) to the Kolmogorov–Smirnov theorems was extensively discussed [see Donsker (1952) and Anderson and Darling (1952)]. While working in the seminar of E. B. Dynkin, I learned that the finite dimensional distributions describe well the behaviour of the trajectories of a Markov process if the probability of two nearly simultaneous jumps of the trajectory is sufficiently small. Having examined several variants, I found that under the condition

$$(1) \qquad \sup_{N,t} \mathbf{E}|\xi_N(t+\tau) - \xi_N(t)|^2 |\xi_N(t) - \xi_N(t-\theta)|^2 < C|\tau+\theta|^2,$$

which restricts the possibility of repeated jumps, the convergence of probabilities of the trajectory $\xi_N(t)$ to stay in a band follows from the weak convergence of finite-dimensional distributions, $\xi_N(t) \Rightarrow \zeta(t)$, for a sequence of processes $\xi_N(t)$ whose trajectories have no discontinuities of the second kind.

The condition (1) is easily verified for $\xi_N(t) = \sqrt{N}[F_N(F^{-1}(t)) - t]$. My report at the "large" seminar on probability theory was greeted with much interest. However, by the end of the discussion, A. N. noted that the additional condition I introduced for the process trajectories $\xi_N(t)$ to have no discontinuities of the second kind appeared to be unnecessary. He said: "This property is sure to follow from (1). Look into my theorem in the paper of Slutsky (1937), where the continuity of the trajectory $\xi(t)$ with probability 1 is derived from the simpler condition

$$(2) \qquad \sup_{t} \mathbf{E}|\xi(t+\tau) - \xi(t)|^p < C|\tau|^{1+\varepsilon}, \qquad \varepsilon > 0, p > 0."$$

And he was right (while I was discouraged). My work was appropriately revised. Thanks to the initiative of E. B. Dynkin, it was published a year later in the first issue of the new Soviet journal *Probability Theory and its Applications* [Čencov (1956)]. Specifically, it contains a very short proof of the Kolmogorov (1933b) theorem mentioned previously from general theoretical results [see Billingsley (1968)]. Meanwhile, condition (1) and its generalizations to other powers in the manner of (2) were called Kolmogorov–Čencov conditions in Doob (1960). Statistical applications were not considered in subsequent studies [Skorohod (1956), Kolmogorov (1956b) and Cramér (1966)]. Nevertheless, I would like to note that the metrization of the Skorohod topology on the space of functions without discontinuities of the second kind introduced by Kolmogorov (1956b) and modified by Billingsley (1968), has been underestimated in other fields of applied mathematics. For example, the authors of some highly accurate difference methods in gas dynamics essentially try to approximate numerically a desired discontinuous solution in the

Kolmogorov metric without being aware of it [see Zabrodin, Sofronov and Chentsov (1988)].

In the years that followed I happened to be involved in the theory and practice of transport phenomena computation by the Monte Carlo method. In order to estimate the unknown density $p(x)$ of the observed data I suggested [Čencov (1962)] to estimate the coefficients

$$a_k = \int p(x)\varphi_k(x)r(x)\,dx = \mathbf{E}[\varphi_k(\xi)r(\xi)] \tag{3}$$

of a series expansion for the density using a suitable basis $\{\varphi_k(x)\}$ orthonormal with the weight $r(x)$. In principle, this method is more economical than the histogram method [which is obtained by a special choice of $\varphi_k(x)$], and the idea of it was then hanging in the air. Soon it was noted by a number of other statisticians, the first of them being Van Ryzin (1966) and Schwartz (1967).

At that time A. N. was participating in a round-the-world cruise aboard the scientific ship *Mendeleev*, and my paper approved by N. V. Smirnov, was communicated to *Doklady Akademii Nauk SSSR* by Academician M. V. Keldysh. It contained a formula for coefficient estimation,

$$\alpha^*_{N,k} = N^{-1}[\varphi_k(\xi^{(1)})r(\xi^{(1)}) + \cdots + \varphi_k(\xi^{(N)})r(\xi^{(N)})], \tag{4}$$

an estimate of the density itself

$$\pi^*_{N,n}(\cdot) = \alpha^*_{N,1}\varphi_1(\cdot) + \cdots + \alpha^*_{N,n}\varphi_n(\cdot), \tag{5}$$

and a bound for the mean norm of the error. Also posed and solved in this paper was the problem of almost optimal choice of the system $\{\varphi_k\}$ and approximation dimension. I used approaches that were being developed by A. N. not long before that time to solve the well-known Hilbert's 13th problem on representability of a function of several variables by superpositions of a function of a smaller number of variables [see Kolmogorov (1955)]. Kolmogorov's final result was so unexpected that it should be cited here: Any continuous function of several variables may be represented by means of superposition of continuous functions of one variable and addition [Kolmogorov (1957)]. A similar superposition theorem in terms of smooth functions is not valid. Investigating the problem, Kolmogorov gave some asymptotic characteristics of totally bounded sets in function spaces. They are now widely used in the theory of numerical methods. In particular, they are essential in solving the so-called ill-posed inverse problems. That was just what I needed. The problem under consideration was posed in the following way. Let $\xi^{(1)}, \dots, \xi^{(N)}$ be a sequence of independently observed values of some random variable ξ with unknown distribution in $(\Omega, \mathscr{A})$ and let $\mathscr{P}$ be the given a priori class of all the feasible densities $p(x)$ that could describe the distribution of the observable ξ. The problem is to construct a density estimator, $(\xi^{(1)}, \dots, \xi^{(N)}) \to \pi^*_N(\cdot)$, to be optimal on the class $\mathscr{P}$.

The quality of a numerical method is usually measured with a guaranteed upper bound of its error (i.e., the deviation of a calculated value from the true one). Such a confidence limit is given in our case by a $\mathscr{P}$-bound $b(N) = b(N, \delta)$

of some quantile $\zeta_{1-\delta}^{(N)}$ of a random variable $\|\pi_N^*(\cdot) - p(\cdot)\|$. As

$$\zeta_{1-\delta}^{(N)} \leq \delta^{-1/\gamma}\left[\mathbf{E}\left\|\pi_N^*(\cdot) - p(\cdot)\right\|^\gamma\right]^{1/\gamma}$$

from the well-known Chebyshev-type inequality, one may describe the accuracy of a decision rule (or, more precisely, the inaccuracy of it) in the framework of Wald's paradigm. The mean value $\mu_\gamma^{(N)} = \mathbf{E}\|\pi_N^*(\cdot) - p(\cdot)\|^\gamma$ is then treated as the risk corresponding to a loss function $L(p, \pi_N^*) = \|\pi_N^* - p\|^\gamma$. Thus, the asymptotic quality of the estimator $\pi_N^*(\cdot)$ may be measured by the order with which the mean value $\mathbf{E}\|\pi_N^*(\cdot) - p(\cdot)\|^\gamma$ decreases as $N \to \infty$.

For the $L_2(r)$-norm of the deviation $\pi_{N,n}^* - p$ we have

$$\mathbf{E}\|\pi_{N,n}^* - p\|^2 = \left\|F_n[p] - p\right\|^2 + \sum_{k=1}^{n} \mathbf{D}\alpha_{N,k}^*, \tag{6}$$

where $F_n[p](\cdot)$ is the orthogonal projection of $p(\cdot)$ on the subspace Φ_n determined in $L_2(r)$ by the basis vectors $\varphi_1(\cdot), \dots, \varphi_n(\cdot)$. Under the natural assumption

$$\forall\, p(\cdot) \in \mathscr{P}, \forall\, x \in X, \qquad |\ln[p(x)r(x)]| \leq A(\mathscr{P}), \tag{7}$$

where $A(p) < \infty$, the second summand in (6) is of order nN^{-1}. Therefore, the subspace Φ_n should be chosen to minimize the maximal deviation of the class $\mathscr{P}$ from Φ_n. This minimum $d_n(\mathscr{P}, r)$ is called the nth Kolmogorov linear width of the class $\mathscr{P}$ in the $L_2(r)$ metric. It first appeared in Kolmogoroff (1936) [see also Tihomirov (1960)]. As the dimension n increases the value $d_n(\mathscr{P}) = g(n)$ decreases. Thus the construction of an asymptotically optimal (in order) algorithm may be reduced for the given N to determine $n^* = \Gamma(N)$ minimizing $g(n) + CnN^{-1}$. The corresponding risk does not exceed

$$\min_n \left[g(n) + CnN^{-1}\right]. \tag{8}$$

It was well known [see Hodges and Lehmann (1951) and Girshick and Savage (1951) in the case of one-dimensional families $\mathscr{P}$] that lower bounds for the minimax and Bayes risk for quadratic loss functions follow from the Cramér–Rao inequality. Let $s_n(\mathscr{P}, r)$ be the radius of the maximal n-dimensional sphere Σ_n imbeddable into the class $\mathscr{P}$, and let $s_n(\mathscr{P}) = h(n)$. Then, integrating the information inequality over Σ_n, we find that the lower bound for the risk of the projection estimator (5) is not less than

$$\max_n \min\left\{\tfrac{1}{2}h(n), cnN^{-1}\right\} \tag{9}$$

for some c. If $g(n) \asymp h(n)$, the bounds (8) and (9) are weakly equivalent; so the estimator (5) based on the subspace Φ_n corresponding to the nth Kolmogorov width at $n = \Gamma(N)$ has the best order of decrease of the risk (6).

The outlines of this theory connected with Kolmogorov's width concept can be extended to estimation in other norms with loss functions of the form $\|\pi_N^*(\cdot) - p(\cdot)\|^\gamma$. Since the Pythagorean equality (6) no longer holds and the error cannot be so easily decomposed into the sum of systematic and random errors, we should take into account that the norm of the sum of two terms exceeds the larger norm for the two terms by no more than a factor of 2, which yields, instead of (8), the upper bound

$$\min_n \max\{2g(n), 2CnN^{-1}\},$$

coinciding with (8) within a factor of 2.

The estimation scheme (6) is reasonable for convex classes satisfying the condition

$$\sup_{p,q \in \mathscr{P}} \sup_{x \in X} |\ln p(x) - \ln q(x)| \le 2A(\mathscr{P}), \tag{10}$$

which follows from (7). Meanwhile, for classes $\mathscr{P}$ of more intricate configuration, nonlinear methods for estimating $p \in \mathscr{P}$ are preferable. Upper bounds for their mean error may be obtained in terms of the ε-entropy of the metric compact $\mathscr{P}$, the logarithm of the minimal number of elements in an ε-net in $\mathscr{P}$. This notion was introduced by Kolmogorov in the 1950s [Kolmogorov (1955, 1956a)]. A lower bound for the mean error may be obtained in terms of Kolmogorov's ε-capacity—i.e., the logarithm of the maximal number of disjoint spheres of radius ε with their centers in the set $\mathscr{P}$ [Kolmogorov and Tihomirov (1959)]. Such bounds were used, for example, by Čencov (1981) and Ibragimov and Has'minskii (1980).

The metric $L_2(r)$ in the theory of the projection estimate $\pi^*_{N,n}$ described previously seems somewhat unnatural. Even the inequality (7), which requires that $[r(\cdot)]^{-1} \asymp p(\cdot) \in \mathscr{P}$, allows too much arbitrariness in the choice of metric measuring the error and determining the loss function. Kolmogorov was always interested in finding "information" distances between probability distributions on a measurable space $(\Omega, \mathscr{A})$ more adequate to the essence of the problems of statistics. He discussed the properties of the quantity

$$d_H(P,Q) = 1 - \int_\Omega \sqrt{P(d\omega)Q(d\omega)} \tag{11}$$

as of a measure of unlikeness of P and Q in a lecture at the Institut Henri Poincaré (November, 1955). The lectures were not published but a reference to (11) can be found in the review of Adhikari and Joshi (1956). We should also note that the integral in the right-hand side of (11) was previously studied by Bhattacharyya (1943) as the characteristic of likeness of P and Q. In his lectures Kolmogorov also stressed the theoretical importance of the total variation of the difference $P - Q$,

$$s_V(P,Q) = \int_\Omega |P(d\omega) - Q(d\omega)|, \tag{12}$$

as a metric (in this paper Kolmogorov's notations are modernized).

It is interesting that the distance

$$s_F(P,Q) = 2 \arccos \int_\Omega \sqrt{P(d\omega)Q(d\omega)} \tag{13}$$

in the spherical Riemannian metric generated by the quadratic differential form with the Fisher information matrix was not regarded at that time as a value of information unlikeness.

For several years I tried unsuccessfully to find an information metric which would give a natural differential structure on smooth families of probability distributions. Then, giving up all hope I followed the advice of E. A. Morozova and turned my attention to a search for natural information linear connections. In no time a nontrivial but very simple flat connection was found [Čencov (1964)], where exponential families of one or several parameters having densities

$$p(\omega; \mathbf{s}) = p(\omega; \mathbf{0}) \exp\left[s^j q_j(\omega; \mathbf{0}) - I(P_0 : P_s) \right] \tag{14}$$

are, respectively, geodesic lines and completely geodesic surfaces, i.e., they play the roles of "straight lines" and "linear subspaces." In (14) $\mathbf{s}$ denotes a canonical geodesic parameter; the basic tangent vectors at a point P_0

$$q_j(\omega; \mathbf{0}) = (\partial/\partial s^j) \ln p(\omega; \mathbf{s})\big|_{s=0}$$

are taken for directional sufficient statistics, while the normalizing divisor

$$\exp[I(P_0 : P_s)] = \int_\Omega \exp\left[s^j q_j(\omega; \mathbf{0}) \right] P_0(d\omega)$$

coincides with the exponential of the Kullback information,

$$I(P_0 : P_s) = \int_\Omega [\ln P_0(d\omega) - \ln P_s(d\omega)] P_0(d\omega).$$

Later, it was found [Čencov (1968)] that in this geometry the quantity $I(P' : P'')$ is an asymmetric analogue of half of the squared Euclidian distance between points. In particular,

$$P_\sigma = \arg \min_{P_s \in \mathscr{P}} I(P : P_s)$$

determines the projection of P on the completely geodesic family $\mathscr{P}$ with regular boundary (as an analogue of a linear subspace) for which nonsymmetric variant of the Pythagorean equality is valid:

$$\forall\, P_s \in \mathscr{P}, \qquad I(P : P_s) = I(P : P_\sigma) + I(P_\sigma : P_s).$$

All this allowed an extension of the Kolmogorov concept of widths to collections of probability distributions equipped with the Kullback information quasimetric and to construct (according to the scheme proposed previously for the $L_2(r)$ metric) optimal algorithms for estimating the density, where a linear aggregate approximates not the density but its logarithm, for a wider range of a priori classes $\mathscr{P}$ [Čencov (1967, 1972), Stratonovich (1969)]. In particular, instead of restriction (10) upon the density ratio one can manage with restrictions on the second moments of its logarithm

$$\sup_{P,\,R\in\mathscr{P}} \left|\ln D_{R'}(P',P'') - \ln D_{R''}(P',P'')\right| \le A(\mathscr{P}), \tag{15}$$
$$D_R(P',P'') = \mathbf{D}_R[\ln(dP'/dP'')(\omega)].$$

The condition (15) has a simple geometric sense. It is known [see Čencov (1972)] that in mathematical statistics the spherical Fisher–Bhattacharyya–Rao metric (13) is the unique (up to a constant multiple) Riemannian metric invariant under the algebraic category of statistical decision rules (the Markov map category).

However, there exists a whole family ∇^γ of invariant linear connections, where the index $\gamma \in \mathbf{R}$ shows the degree $[p(\omega)]^\gamma$ of the distribution density, which is affine (up to a normalizing divisor) in a suitable parameter along the ∇^γ-geodesic family. In the case $\gamma \neq 0$, including $\gamma = \frac{1}{2}$, which responds to the Riemannian case, the connections ∇^γ are not hereditary with respect to the (tensor) powers of probability spaces corresponding to the scheme of independent observations. Thus, condition (15) requires that for the natural ($\gamma = 0$) parallel translation of a vector tangent to $\mathscr{P}$, its length measured at one point in the natural Riemannian metric ($\gamma = \frac{1}{2}$) will not differ by more than a factor $\exp[\frac{1}{2}A(\mathscr{P})]$ from the result of a similar measurement at any other point in $\mathscr{P}$.

The category of statistical decision rules was introduced in Čencov (1965) and in Morse and Sacksteder (1966) to formalize some known concepts of Wald and Blackwell. Many natural objects of mathematical statistics are invariants of the category. But some of the new conclusions, such as that the simplex of all probability distributions on a finite set proved to be a curved space with a nonsymmetric metrization, seemed slightly strange at first. Kolmogorov, writing a report on my works, had then pointed out the methodological interest of category approaches. He had also underlined the Kullback information metrization [see, for example, Kullback (1958)] to be of great importance in many subfields of mathematical statistics. Now, at the period of intensive geometrization of statistical theory (according to an expression of Yu. V. Prohorov), the trend of research just discussed has advanced substantially both in the development of geometrical approaches and in the range of statistical problems to be solved [Amari, Barndorff-Nielsen, Kass, Lauritzen and Rao (1987) and Has'minskii and Ibragimov (1983)].

Another fundamental problem of statistical theory, the solution of which has been obtained on the basis of Kolmogorov's ideas and conceptions, concerns the impossibility of localizing [in the metric (12) or any invariant metric] the probability distribution P of an unknown random variable ξ by using a sequence of independent observations of it under the restriction that there is no additional information about the measure P. From the Glivenko (1933) theorem on the convergence of the empirical distribution function to the theoretical one as well as from Kolmogorov's (1933b) result follows only that the localization of P in weak metrics is feasible.

To treat the P-localization problem as an inverse problem of probability theory, a formal description of the decision space is needed. In the classical theory of parameter estimation the localization problem is replaced by the problem of estimating the "true" values of parameters so that our estimates are random vectors. In such problems as density estimation or regression, estimating a function is again reduced to estimation of smooth parameters, but now there is an infinite sequence of parameters; for example, the coefficients (3) of an orthogonal expansion where the dependence of function values on parameters a_n decreases as n increases. In estimation theory of that kind, the information inequality and other approaches of finite-dimensional theory are valid (for this reason the class of problems mentioned previously should have been distinguished long ago from nonparametric statistics as an independent countable-parametric statistic).

But here we choose as estimates random probability measures on an algebra $\mathscr{B}$ of events, i.e., random elements of the uncountable-dimensional space $\mathbf{R}^{\mathscr{B}}$ satisfying the axiomatic requirements of nonnegativity, total mass 1, additivity and countable additivity. They distinguish in $\mathbf{R}^{\mathscr{B}}$ a nonmeasurable set $W(\mathscr{B})$ to which the measurability theory of Kolmogoroff (1933a) cannot be applied directly, nor can the Kakutani–Halmos theory [Kakutani (1944) and Halmos (1950)]. However, there is a roundabout way outlined by Kolmogorov and implemented by Prohorov (1961) and Čencov (1972): a random measure is first determined on a countable set of generators of the algebra $\mathscr{B}$, then it is extended step-by-step without contradictions to the whole algebra [the method proposed by Doob (1947) in a simpler situation]. Thus, we may treat $W(\mathscr{B})$ as a measurable set, and the whole body of measure theory, including the inversion of the order of integration, may be used on it.

The final result of Čencov (1981) is as follows. Let Π_N be a statistical decision rule for processing the sequential independently observed values $\xi^{(1)}, \dots, \xi^{(N)}$ of a random variable ξ in $(\mathbf{R}, \mathscr{B})$ with a collection $W(\mathscr{B})$ as the space of decisions P_N^*. For any sequence $\{\Pi_N\}$ there exists a distribution $Q \in W(\mathscr{B})$ such that $\{\Pi_N\}$ is inconsistent to localize Q by a sequence of independent observations of the Q-distributed random variable η:

$$\text{as } N \to \infty, \qquad \mathbf{E}_{Q^N} s_V(P_N^*, Q) \nrightarrow 0.$$

Instead of s_V we may take any other metric ρ, invariant under the category of statistical decision rules, because for them we always have $\rho \geq c(\rho) s_V$ with a suitable constant $c(\rho)$ [see Morozova and Čencov (1987)].

I shall not list here the ideas of A. N. used in the proof. It is sufficient to note

that the number of references to Kolmogorov in Čencov (1981) exceeds the number of references to Prohorov and to the author taken together.

Thus, an elementary problem of mathematical statistics—the inverse problem of probability theory—turns out to be ill-posed. The degree of ill-posedness is at most $\frac{1}{2}$ if the ill-posedness degree of differentiation is taken to be unity. This follows from Kolmogorov (1933b).

REFERENCES

ADHIKARI, B. P. and JOSHI, D. D. (1956). Distance, discrimination et resumé exaustif. *Publ Inst. Statist. Univ. Paris* **5** 57–74.

AMARI, S.-I., BARNDORFF-NIELSEN, O. E., KASS, R. E., LAURITZEN, S. L. and RAO, C. R. (1987). *Differential Geometry in Statistical Inference*. IMS, Hayward, Calif.

ANDERSON, T. W. and DARLING, D. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic process. *Ann. Math. Statist.* **23** 193–212.

BHATTACHARYYA, A. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Calcutta Math. Soc.* **35** 99–109.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

ČENCOV (CHENTSOV), N. N. (1956). Weak convergence of stochastic processes with trajectories without discontinuities of second kind and so called "heuristic" approach to the goodness of fit tests of Kolmogorov–Smirnov type. *Teor. Verojatnost. i Primenen.* **1** 156–161. (In Russian.)

ČENCOV (CHENTSOV), N. N. (1962). Estimation of unknown distribution density from observations. *Dokl. Akad. Nauk SSSR* **147** 45–48. (English translation in *Soviet Math. Dokl.* **3** 1559–1562.)

ČENCOV, N. N. (1964). The geometry of a "manifold" of probability distributions. *Dokl. Akad. Nauk SSSR* **158** 543–546. (English translation in *Soviet Math. Dokl.* **5** 1282–1286.)

ČENCOV, N. N. (1965). The categories of mathematical statistics. *Dokl. Akad. Nauk SSSR* **164** 511–514. (English translation in *Soviet Math. Dokl.* **6** 1268–1272.)

ČENCOV, N. N. (1967). Invariant loss functions in problems of mathematical statistics. *Uspekhi Mat. Nauk* **22** 178–180. (In Russian.)

ČENCOV, N. N. (1968). A nonsymmetric distance between probability distributions, entropy and the Pythagorean theorem. *Mat. Zametki* **4** 323–332. (English translation in *Math. Notes* **4**.)

ČENCOV, N. N. (1972). *Statistical Decision Rules and Optimal Inference*. Nauka, Moscow. English translation (1982), Providence.

ČENCOV, N. N. (1981). On the correctness of statistical point estimation problem. *Teor. Verojatnost. i Primenen.* **26** 15–31. (English translation in *Theory Probab. Appl.* **26** 13–26.)

CRAMÉR, H. (1928). On the composition of elementary errors. Second paper: Statistical applications. *Skand. Aktuarietidskr.* **1/2** 141–180.

CRAMÉR, H. (1966). On stochastic processes whose trajectories have no discontinuities of the second kind. *Ann. Mat.* (*IV*) **71** 85–92.

DONSKER, M. D. (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov–Smirnov theorems. *Ann. Math. Statist.* **23** 277–281.

DOOB, J. L. (1947). Probability in function space *Bull. Amer. Math. Soc.* **53** 15–30.

DOOB, J. L. (1949). Heuristic approach to the Kolmogorov–Smirnov theorems. *Ann. Math. Statist.* **20** 393–403.

DOOB, J. L. (1960). Some problems concerning the consistency of mathematical models. In *Information and Decision Processes* (R. E. Machol, ed.) 27–33. McGraw-Hill, New York.

GIRSHICK, M. A. and SAVAGE, L. J. (1951). Bayes and minimax estimates for quadratic loss functions. *Proc. Second Berkeley Symp., Math. Statist. Prob.* 53–73. Univ. California Press.

GLIVENKO, V. (1933). Sulla determinazione empirica della leggi di probabilità. *Giorn. Istit. Ital. Attuari* **4** 92–99.

HALMOS, P. R. (1950). *Measure Theory*. Van Nostrand, Princeton.

HAS'MINSKII, R. Z. and IBRAGIMOV, I. A. (1983). Efficient estimation in the presence of infinite dimensional incidental parameters. In *Probability Theory and Mathematical Statistics. Lecture Notes in Math.* **1021** 195–229. Springer, New York.

HODGES, J. L. and LEHMANN, E. L. (1951). Some applications of the Cramér–Rao inequality. *Proc. Second Berkeley Symp. Math. Statist. Prob.* 13–22. Univ. California Press.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1980). On estimation of the density function. In

Investigation in Mathematical Statistics IV Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov (LOMI) **98** 61–85. (In Russian.)

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

KOLMOGOROV, A. N. (1931). Median method in error theory. *Math. Sb.* **38** 47–50. (In Russian.)

KOLMOGOROFF (KOLMOGOROV), A. N. (1933a). *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Springer, Berlin.

KOLMOGOROV, A. (1933b). Sulla determinazione empirica di una legge di distribuzione. *Giorn. Istit. Ital. Attuari* **4** 83–91.

KOLMOGOROV, A. N. (1933c). To the question of fitness of prediction formulae obtained in statistical way. *Zh. Geofiz.* **3** 78–82. (In Russian.)

KOLMOGOROFF, A. N. (1934). Zufällige Bewegungen (Zur Theorie der Brownschen Bewegung). *Ann. of Math.* **35** 116–117.

KOLMOGOROFF, A. N. (1936). Ueber die beste Annäherung von Funktionen einer gegebenen Funktionklasse. *Ann. of Math.* **37** 107–110.

KOLMOGOROV, A. N. (1937). On the statistical theory of metal crystallization. *Izv. Akad. Nauk SSSR Ser. Mat.* No. 3 335–360. (In Russian.)

KOLMOGOROV, A. N. (1940). On a new confirmation of Mendel's laws. *Dokl. Akad. Nauk SSSR* **27** 38–42. (In Russian.)

KOLMOGOROV, A. N. (1941a). Confidence limits for an unknown distribution function. *Ann. Math. Statist.* **12** 461–463.

KOLMOGOROV, A. N. (1941b). On log normal distribution of sizes of particles in a crusher. *Dokl. Akad. Nauk SSSR* **31** 90–101. (In Russian.)

KOLMOGOROV, A. N. (1942). Determination of the center of scattering and the measure of accuracy by a limited number of observations. *Izv. Akad. Nauk SSSR Ser. Mat.* **6** 3–32. (In Russian.)

KOLMOGOROV, A. N. (1945a). Hit number for several firings and general principles of estimating the efficiency of a firing system. *Trudy Mat. Inst. Steklov* **12** 7–25. (In Russian.)

KOLMOGOROV, A. N. (1945b). Artificial dispersion in the case of destruction by a single hit and dispersion in one dimension. *Trudy Mat. Inst. Steklov* **12** 26–45. (In Russian.)

KOLMOGOROV, A. N. (1946). On the foundations of the method of least squares. *Uspekhi Mat. Nauk* **1** 57–70. (In Russian.)

KOLMOGOROV, A. N. (1949a). Basic problems of theoretical statistics. In *Trudy Vtorogo Vsesoyuzn. Soveshch. po Matem. Statistike Sept. 27–Oct. 2 1948* 216–220. Uzbekgosizdat, Tashkent. (In Russian.)

KOLMOGOROV, A. N. (1949b). The real meaning of results of analysis of variance. In *Trudy Vtorogo Vsesoyuzn. Soveshch. po Matem. Statistike* 240–268. (In Russian.)

KOLMOGOROV, A. N. (1949c). Solution of a problem of probability theory connected with a question on stratum-formation mechanisms. *Dokl. Akad. Nauk SSSR* **65** 793–796. (In Russian.)

KOLMOGOROV, A. N. (1950). Unbiased estimates. *Isv. Akad. Nauk SSSR Ser. Mat.* **14** 303–326. (In Russian.)

KOLMOGOROV, A. N. (1951). Statistical quality control if the acceptable number of defectives is equal to zero. Leningradsk. Dom Nauchno-Techn. Propagandy, Leningrad 1–24. (In Russian.)

KOLMOGOROV, A. N. (1955). Estimation of the minimal number of elements in ε-nets in various functional classes and their application to the problem of representability of functions of several variables by superpositions of functions of a smaller number of variables. *Dokl. Akad. Nauk SSSR* **101** 192–194. (In Russian.)

KOLMOGOROV, A. N. (1956a). On some asymptotic characteristics of totally bounded metric spaces. *Dokl. Akad. Nauk SSSR* **108** 385–388. (In Russian.)

KOLMOGOROV, A. N. (1956b). On the convergence of A. V. Skorohod. *Teor. Verojatnost. i Primenen.* **1** 239–247. (In Russian.)

KOLMOGOROV, A. N. (1957). On the representation of continuous functions of several variables by means of superpositions of continuous functions of one variable and addition. *Dokl.*

Akad. Nauk SSSR **114** 953–956. (In Russian.)

KOLMOGOROV, A. N., PETROV, A. A. and SMIRNOV, YU. M. (1947). A formula of Gauss from the theory of the method of least squares. *Izv. Akad. Nauk SSSR Ser. Mat.* **11** 561–566. (In Russian.)

KOLMOGOROV, A. N. and PROHOROV, YU. V. (1956). Zufällige Funktionen und Grenzverteilungssätze. In *Bericht über die Tagung Wahrscheinlichkeitsrechnung und mathematische Statistik in Berlin, October, 1954* 113–126. Deutscher Verlag der Wissenschaften, Berlin.

KOLMOGOROV, A. N. and TIHOMIROV, V. M. (1959). ε-entropy and ε-capacity of sets in function spaces. *Uspekhi Mat. Nauk* **14** 3–86. (English translation in *Russian Math. Surveys* **14** 277–364.)

KULLBACK, S. (1958). *Information Theory and Statistics*. Wiley, New York.

MOROZOVA, E. A. and ČENCOV, N. N. (1987). Markov maps in noncommutative probability theory and mathematical statistics. In *Probability Theory and Mathematical Statistics* (Yu. V. Prohorov, V. A. Statulevicius, V. V. Sazonov, and B. GRIGELIONIS, eds.). *Proc. Fourth Internat. Vilnius Conf. on Probab. Theory and Math. Statist.* **2** 287–310. VNU Sci. Press, Utrecht.

MORSE, N. and SACKSTEDER, R. (1966). Statistical isomorphism. *Ann. Math. Statist.* **37** 203–214.

PROHOROV, YU. V. (1961). Random measures on a compactum. *Dokl. Akad. Nauk.* **138** 53–55. (English translation in *Soviet Math. Dokl.* **2** 539–541.)

SCHEFFÉ, H. (1959). *The Analysis of Variance*. Wiley, New York.

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

SKOROHOD, A. V. (1956). Limit theorems for random processes. *Teor. Verojatnost. i Primenen.* **1** 289–319. (In Russian.)

SLUTSKY, E. E. (1937). Qualche preposizione relativa alla teoria delle funzioni aleatori. *Giorn. Istit. Ital. Attuari* **8** 183–198.

SMIRNOFF (SMIRNOV), N. V. (1936). Sur la distribution de ω^2. *C. R. Acad. Sci. (Paris)* **202** 449–452.

SMIRNOV, N. V. (1937). On the distribution of Mises' ω^2-test. *Mat. Sb.* **2** 973–994. (In Russian.)

SMIRNOV, N. V. (1939a). On deviations of the empirical distribution function. *Mat. Sb.* **6** 3–26. (In Russian.)

SMIRNOV, N. V. (1939b). Evaluation of the divergence between empirical distribution functions in two independent samples. *Moscow, Bull. Univ.* **2** 3–14. (In Russian.)

STRATONOVICH, R. L. (1969). The speed of convergence of distribution density estimation algorithms. *Izv. Akad. Nauk SSSR Ser. Techn. Kibern.* N 6 3–15. (In Russian.)

TIHOMIROV, V. M. (1960). Widths of sets in functional spaces and the theory of best approximations. *Uspekhi Mat. Nauk.* **15** 81–120. (English translation in *Russian Math. Surveys* **15** 75–111.)

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

VON MISES, R. (1931). Wahrscheinlichkeitsrechnung und ihre Anwendungen in der Statistik und theoretische Physik. Franz Deuticke, Wien.

ZABRODIN, A. V., SOFRONOV, I. D. and CHENTSOV, N. N. (1988). Adaptive difference method for mathematical simulation of nonstationary gas dynamical flows. *VANT/Quest. Atom. Sci. Techn./Ser. Methods and Codes for Numerical Solution of Mathematical Physics Problems.* No. 4 3–22. (In Russian.)

KELDYSH INSTITUTE OF APPLIED MATHEMATICS
ACADEMY OF SCIENCES OF THE USSR
MIUSSKAYA SQ., 4
125047 MOSCOW
UNION OF SOVIET SOCIALIST REPUBLICS