The Annals of Statistics
1990, Vol. 18, No. 3, 1011–1016

KOLMOGOROV'S CONTRIBUTIONS TO MATHEMATICAL STATISTICS[1]

BY ANDREW L. RUKHIN

University of Massachusetts

In this paper Kolmogorov's work in mathematical statistics is reviewed. The main areas under discussion are: the idea of sufficiency, linear models and unbiased estimators. The relationship of Kolmogorov's contributions with modern statistical theory, in particular, with Bayesian analysis, is analyzed.

1. Introduction. Andrei Nikolaevitch Kolmogorov was keenly interested in statistical theory and its applications all of his life. He remained the director of the Statistical Laboratory at Moscow State University until his last days and was known to lament about insufficient development of statistical methods in the USSR. His first statistical paper, "Method of Median in the Theory of Errors" was published in 1931; the last paper in this area on estimation of parameters of a complex stationary Gaussian Markov process appeared in 1962.

In this review I concentrate on the following three areas of Kolmogorov's interests and contributions: the sufficiency concept and Bayesian approach, estimation theory, in particular, unbiased estimation and linear statistical models.

2. Sufficiency and confidence intervals: was Kolmogorov a closet Bayesian? One of the most important contributions by Kolmogorov to theoretical statistics is his article "Determination of Dispersion Center and of Accuracy Measure from a Finite Number of Observations" published in prestigious and highly mathematically oriented Izvestia Akademii Nauk SSSR in 1942. Inexplicably, this landmark paper is missing in the volume of Kolmogorov's collected works in probability theory and mathematical statistics which appear in 1986 (Nauka Publishing House). One should note that in 1942 the Soviet Union was in the midst of World War II, during which Kolmogorov was heavily involved in defense projects. He explicitly states that one of the reasons for this article was his intention "to explain to artillerists some results of Student and Fisher related to small samples."

For methodological reasons Kolmogorov restricts his attention to random samples coming from one- or two-parameter normal distributions. In Section 1, which he calls "The Classical Approach," posterior densities for the parameter(s) unknown are discussed. Kolmogorov states that these formulae can

Received March 1989.
[1]Research supported by NSF Grant DMS-88-03259.
AMS 1980 *subject classifications*. 01A60, 62A15, 62B05, 62F10, 62G05.
Key words and phrases. Sufficiency, prior distribution, confidence interval, unbiased estimators, linear statistical model, least squares estimators, accuracy estimation.

hardly be used in practice *because prior densities in them are typically unknown*. This is of course a conclusion shared by most modern statisticians. However it took quite some time before this fact was recognized in western literature [cf. discussion in Chapter 3 in Berger (1985b)]. More importantly, Kolmogorov writes: "One has to realize that the assumption of the existence of a particular prior distribution of the mean a and the precision h can be justified only in some rather restricted cases. For instance, it is meaningful to speak about probability distributions for precision in rifle shooting under fixed conditions for a marksman chosen at random from a given troop. However it does not make any sense to discuss a prior probability distribution of precision for missiles' discharge in general (under all possible conditions and by using any firing devices of the past and the future)." Apparently Kolmogorov takes a pragmatic point of view here that a prior distribution above which can be interpreted as a mixture of existing priors, cannot be tractable.

Despite this remark, in the next section Kolmogorov gives a Bayesian definition of sufficient statistic: A statistic T is called sufficient if the posterior distribution of the unknown parameter depends only on the prior distribution of this parameter and the value of T.

This is a departure from the classical sufficiency definition of Fisher (1922). Kolmogorov was well aware of Fisher's definition and apparently believed that these two definitions are equivalent. This is indeed true under some regularity assumptions, see Heyer (1972), (1982). However, Blackwell and Ramamoorthi (1982) constructed an example of a Bayesian sufficient statistic which is not sufficient in Fisher's sense. (Fisher's sufficiency implies Bayesian sufficiency).

After this definition and some examples of sufficient statistics, Kolmogorov returns to posterior densities for the normal parameters under the assumption of noninformative constant prior density, which becomes an object of Kolmogorov's criticism. Kolmogorov notices that these formulae can be justified only as approximate ones when the sample size n is large, and he obtains limit theorems of the following form.

Let a be the normal mean and $h = 2^{-1/2}\sigma^{-1}$ be the normal precision. Then $\bar{x} = n^{-1}\sum_1^n x_j$ and $S^2 = \sum_1^n (x_j - \bar{x})^2$ form a sufficient statistic for (a, h) on the basis of the sample $x_1, \dots, x_n$. For the posterior density $\phi_1(a|x_1, \dots, x_n)$, the following approximation holds:

$$\phi_1(a|x_1, \dots, x_n) \sim n^{1/2} h \pi^{-1/2} \exp\{-nh^2(a - \bar{x})^2\}$$

and for the posterior density $\phi_2(a, h|x_1, \dots, x_n)$, one obtains

$$\phi_2(a, h|x_1, \dots, x_n) \sim (2n)^{1/2} S\bar{h}\pi^{-1/2} \exp\{-n\bar{h}^2(a - \bar{x})^2 - 2S(h - \bar{h})^2\},$$

where $\bar{h} = (n-1)^{1/2} 2^{-1/2} S^{-1}$ (Kolmogorov's notation is somewhat different). Results of such form have since been extensively studied and are known as theorems of the von Mises–Bernstein type [cf. for instance De Groot (1970), Chapter 10 or Johnson (1970) and the references there].

Kolmogorov uses these formulae to obtain approximate *credible* intervals for a and h. He also discusses the use of posterior variance as an accuracy measure of the posterior mean.

For small sample sites, Kolmogorov basically seems to agree with the widespread opinion (in the USSR) of S. N. Bernstein that without exact knowledge of the prior distribution one cannot use credible sets obtained from the posterior distribution. Kolmogorov's suggestion is, as a remedy, to use the idea of confidence intervals from J. Neyman. Kolmogorov discusses in some detail the form of unbiased confidence intervals for both normal parameters. His desire to see the practical implementation of these results leads to a table of percentiles of t-distribution and chi-square distribution. (It is worth repeating that the article was published in a highly theoretical mathematical journal. Tables like this were uncommon in such journals). Kolmogorov concludes this trenchant paper with a derivation of unbiased estimators of precision h. His interest in unbiased estimation lead him to another important paper discussed in the next section.

3. Unbiased estimators. In this section I shall discuss Kolmogorov's article "Unbiased Estimators" which was published in the same Izvestia Akademii Nauk SSSR in 1950 and which is more well-known than his previous paper. Kolmogorov starts this paper by proving what is known as the Rao–Blackwell theorem: conditional expected value of any estimator for a given value of a sufficient statistic is an estimator with the same expectation and smaller variance. (This result is sometimes referred to as the Kolmogorov–Rao–Blackwell theorem in Soviet literature.)

Kolmogorov derives equations determining unbiased estimators both in discrete and continuous cases. An application of unbiased estimation for discrete random variables of main interest to Kolmogorov is a quality control problem of sampling inspection in which hypergeometric or binomial distribution for the number of defective items X is assumed. The parametric function of interest is the probability $P(X \leq c)$, where c is a given constant. Kolmogorov reproduces a graph of this characteristic obtained by the Statistical Research Group at Columbia University and makes some practical recommendations for the choice of sample sizes. He also considers the problem of unbiased estimation of the variance of the unbiased estimator for the proportion of defective items in the sample.

Kolmogorov then goes on to unbiased estimation of a given function f of the normal mean a for which the sample mean $\bar{x}$ is a sufficient statistic. The integral equation defining the unbiased estimator $\phi(\bar{x})$ has the form

$$\int_{-\infty}^{\infty} \phi(\bar{x}) G(\bar{x} - a, T)\, d\bar{x} = f(a),$$

where

$$G(z, t) = (\pi t)^{-1/2} \exp\{-z^2/4t\}/2$$

and

$$T = \sigma^2/(2n).$$

Kolmogorov observes that this classical inverse heat problem has a unique solution if it exists, and in this case ϕ can be found from the heat equation

$$\frac{\partial^2}{\partial z^2}\phi = \frac{\partial}{\partial t}\phi,$$

where $t > -T$.

As a corollary to this result, Kolmogorov obtains unbiased density and distribution function estimators. The same estimation problem is of more practical interest when both normal parameters a and σ are unknown, in which case Kolmogorov uses a very ingenious device to derive the desired estimators.

Assume that $P(x_1 < a) = \phi((a - \xi)/\sigma) = \theta$ is to be estimated on the basis of a normal random sample $x_1, \ldots, x_n$. Then by the Rao–Blackwell theorem,

$$P(x_1 < a|\bar{x}, S) = \delta(\bar{x}, S)$$

is an unbiased estimator of θ.

Kolmogorov knew that the ratio $Z = (x_1 - \bar{x})/S$ is independent of $(\bar{x}, S)$. This fact easily follows from the Basu lemma according to which a similar (ancillary) statistic, in our case Z, is independent of a complete sufficient statistic, in our case $(\bar{x}, S)$. Therefore

$$\begin{aligned}\delta(\bar{x}, S) &= P\{Z < (a - \bar{x})S^{-1}|\bar{x}, S\}\\ &= P\{Z < (a - \bar{x})S^{-1}\}\\ &= I_w(n/2 - 1, n/2 - 1),\end{aligned}$$

where I_w denotes the complete beta function and for $0 \le w \le 1$,

$$w = 0.5 - n^{1/2}(\bar{x} - a)\Big/\left[2S(n-1)^{1/2}\right].$$

Also $\delta(\bar{x}, S) = 0$ if $w < 0$ and $\delta(\bar{x}, S) = 1$ if $w > 1$. These formulae were obtained from the distribution of Z. The unbiased estimator of the normal density at a point can be derived now by differentiation δ in a.

Notice that this estimator vanishes outside the interval $|\omega - 0.5| < 0.5$ and δ takes extreme values 0 and 1 outside this interval. In particular, neither of these estimators is an analytic function of ω, so it seems that these estimators cannot be (generalized) Bayes procedures. It turns out however that the estimator δ can be interpreted as a generalized Bayes rule with respect to an improper prior density of the form $\exp\{a^2/(2\sigma^2)\}\, da\, d\sigma$. The explanation of the extreme values 0 and 1 for such a generalized Bayes estimator comes from the fact that the marginal density of $(\bar{x}, S)$ under this prior distribution is finite if and only if $|\omega - 0.5| < 0.5$.

After Kolmogorov's work, a large number of papers have been devoted to construction of unbiased estimators [cf. Lehmann (1983), Chapter 2 or a recent Russian monograph by Nikulin and Voinov (1989)]. However, the realization that unbiased estimators of positive parameters can take negative values and often have excessively large risk functions, especially in multivariate and nonsymmetric distributions, lead to gradual antiquation of this principle.

4. Linear statistical methods. In two papers, "Towards Justification of the Least Squares Method" [Uspekhi Matematiches kikh Nauk (1946)] and "A Formula of Gauss from the Theory of Least Squares Method", jointly with A. A. Petrov and Yu. M. Simirnov [Izvestia Nauk SSSR (1947)], Kolmogorov considers a general linear statistical model. In the first of these papers he gives a geometric interpretation of the least squares estimator as a projection of the corresponding subspace and discusses the problem of replacing the unknown value of the error variance by its unbiased estimate. He gives a systematic presentation of the chi-squared distribution and its use for obtaining confidence intervals for the variance. The same confidence estimation problem for regression coefficients is also carefully explored and a detailed table of quantiles of the t-distribution is given. In several places Kolmogorov warns against using normal percentiles as a substitute for t-percentiles for small and moderate sample sizes. Apparently this was a common mistake committed by practical users of least squares techniques. In fact the exposition of linear statistical methods in Russia before this paper of Kolmogorov's was not much different from the original presentation of Gauss with which Kolmogorov apparently was very familiar.

Kolmogorov spends some time in finding what properties of the least squares estimator are due to the normality assumption and what properties are related to the noncorrelation assumption only.

Kolmogorov was also interested in the problem of the analysis of variance. At the Second USSR Conference on Mathematical Statistics in 1951, Kolmogorov gave a paper, "The Real Meaning of the Results of the Analysis of Variance," which gave rigorous mathematical analysis of basic formulae of analysis of variance.

This work of Kolmogorov strongly influenced studies in the least squares method in the USSR. For instance, the monograph of Yu. V. Linnik, "Method of Least Squares and Principles of the Theory of Observations," whose first Russian edition appeared in 1952 is partly based on Kolmogorov's paper. However this work remained unnoticed in the west and the first geometric approach to linear statistical models has been given in Scheffé (1959).

In the second of the above-mentioned articles, Kolmogorov returns to a linear model with independent but not necessarily normal errors. Gauss, in his "Theoria Combinationis Observationum Erroribus Minimis Obnoxiae" [published in (1823)], obtained a formula for the variance of the traditional variance estimator. This formula involves the kurtosis of error distribution $\varkappa = m_4 - 3m_2^2$ and a sum of squares of diagonal elements of a matrix determined by the design matrix. Kolmogorov and his coauthors used some algebraic results to obtain unimprovable bounds on the mentioned sum of squares which translates into lower and upper bounds for the variance in question depending on the sign of $\varkappa$.

This paper is another sign of Kolmogorov's interest in measuring not only values of unknown parameters but also their degree of accuracy. The problem of the estimation of unknown variance was extensively studied afterwards. Improved point estimators of normal variance have been developed by Stein

(1964) and Brewster and Zidek (1974); confidence intervals have been investigated by Cohen (1972) and Maatta and Casella (1987). Estimation of the variance in linear models has been studied in Drygas (1982) and Gelfand and Dey (1988).

A general problem of loss (or accuracy) estimation was posed by Kiefer (1977) with subsequent work by Berger (1985a) and Rukhin (1988).

REFERENCES

BERGER, J. O. (1985a). The frequentist viewpoint and conditioning. In *Proc. Berkeley Conf. in Honor of Jerzy Neyman and Jack Kiefer* (L. Le Cam and R. Olshen, eds.) **1** 15–44. Wadsworth, Monterey, Calif.

BERGER, J. O. (1985b). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer, New York.

BLACKWELL, D. and RAMAMOORTHI, R. V. (1982). A Bayes but not classically sufficient statistic. *Ann. Statist.* **10** 1025–1026.

BREWSTER, J. F. and ZIDEK, J. V. (1974). Improving on equivariant estimators. *Ann. Statist.* **2** 21–38.

COHEN, A. (1972). Improved confidence intervals for the variance of a normal distribution. *J. Amer. Statist. Assoc.* **67** 382–387.

DEGROOT, M. H. (1970). *Optimal Statistical Decisions*. McGraw–Hill, New York.

DRYGAS, H. (1982). Asymptotic confidence intervals for the variance in the linear regression model. In *Statistics and Probability: Essays in Honor of C. R. Rao* (G. Kallianpur, P. R. Krishnaiah and J. K. Ghosh, eds.) 233–239. North-Holland, Amsterdam.

FISHER, R. A. (1922). On the mathematical foundations of theoretical statistics. *Phil. Trans. Roy. Soc. London (A)* **222** 309–368.

GELFAND, A. and DEY, D. (1988). Improved estimation of the disturbance variance in a linear regression model. *J. Econometrics* **39** 387–395.

HEYER, H. (1972). Zum Erschöpftheits Begriff von D. Blackwell. *Metrika* **19** 54–67.

HEYER, H. (1982). *Theory of Statistical Experiments*. Springer, New York.

JOHNSON, R. A. (1970). Asymptotic expansions associated with posterior distributions. *Ann. Math. Statist.* **41** 851–864.

KIEFER, J. (1977). Conditional confidence statements and confidence estimators. *J. Amer. Statist. Assoc.* **72** 789–827.

LEHMANN, E. L. (1983). *Theory of Point Estimation*. Wiley, New York.

LINNIK, YU. V. (1961). *Method of Least Squares and Principles of the Theory of Observations*. Pergamon, New York.

MAATTA, J. M. and CASELLA, G. (1987). Conditional properties of interval estimators of normal variance. *Ann. Statist.* **15** 1372–1388.

NIKULIN, M. S. and VOINOV, V. G. (1989). *Unbiased Estimators* (in Russian). Nauka, Moscow.

RUKHIN, A. L. (1988). Loss functions for loss estimation. *Ann. Statist.* **16** 1262–1269.

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

STEIN, C. (1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Ann. Inst. Statist. Math.* **16** 155–160.

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF MASSACHUSETTS
AMHERST, MASSACHUSETTS 01003