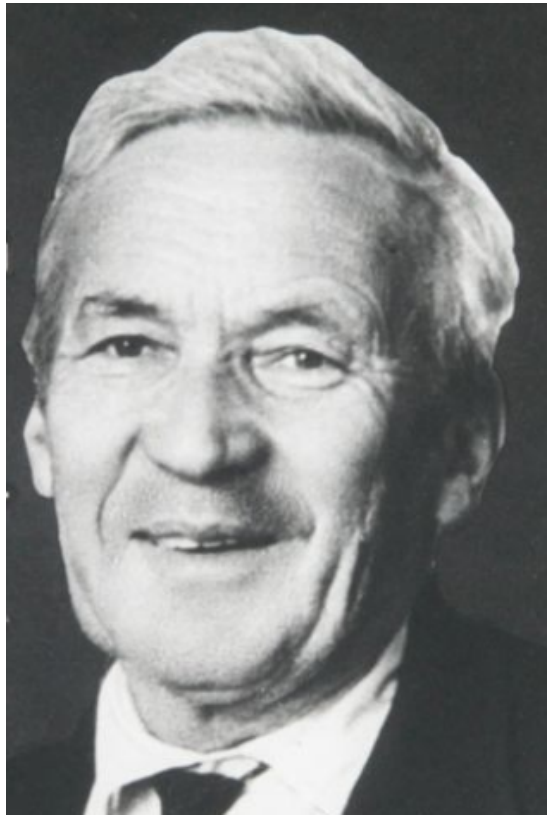
Statistician and Mathematician. Ronald Fisher and Andrey Kolmogorov: a distant strained relationship

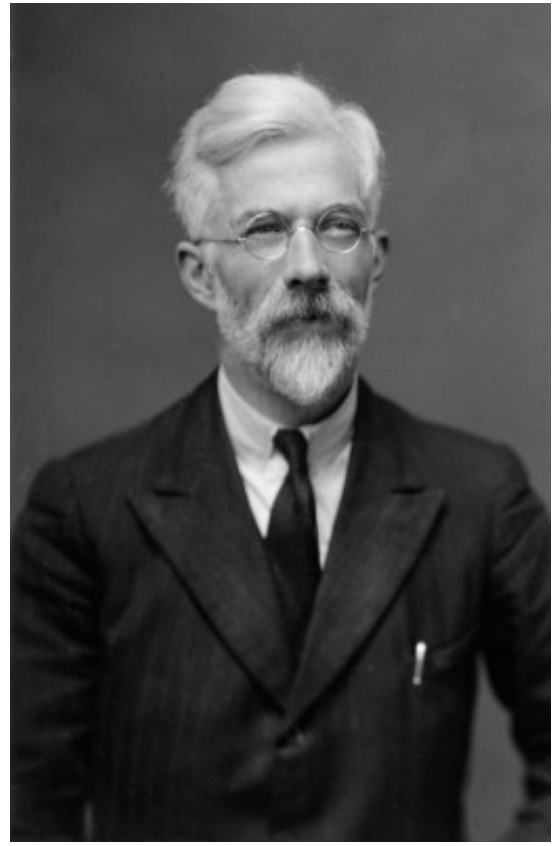
M. MALYUTOV

NORTHEASTERN UNIVERSITY, BOSTON

m.malioutov@neu.edu

HSE, Moscow, August 2018





Ouverture

In 1996 I was invited to present a talk on H. Chernoff's 75th birthday celebration at Harvard Statistics.
I started with acknowledging strong influence of Chernoff's papers to my development as a Mathematical statistician.

Suddenly, late Marvin Zelen, influential head of Harvard Biostatistics at that time, climbed the podium and interrupted me with yelling: 'Maths and Stats have nothing in common!'

My talk aims to show how these two disciplines can enrich each other on an example of distant relations of two giants R. Fisher and A. Kolmogorov. Both made revolutionary progress in theory and applications of their disciplines. Here I describe striking examples.

I touch on statistical education of a mathematician and problems of their uneasy coexistence in Academia, see 'Probability. and Statistics through the centuries', G. Roussas,, UC Davies 2009.

See also 'Statistics and Mathematics: tension and cooperation', D. S. Moore and G. W. Cobb, Am. Math. Monthly (2000), 615-630.

Contents

Outline

During my work as academic secretary of the Kolmogorov Statistical Lab in Moscow University and in subsequent years I had a privilege to hear Kolmogorov's comments on various issues. His first question to my doctoral thesis: do references include Fisher? This shows his high esteem to the Fisher's work.

I became interested in the history of Statistics in the former Soviet Union and the leading role of Kolmogorov in it as complemented to his world leadership in the Probability development.

My footnote remarks to Kolmogorov's papers on regression published in a **Probability and Statistics volume of his selected works** are omitted here.

Kolmogorov viewed Fisher's work as a **beacon** showing directions in Statistics (notably the early notion of information) and Population Genetics to develop MATHEMATICALLY.

Statistical education of a mathematician

I joined the Kolmogorov Statistical Lab in Moscow University in 1966 being already an accomplished pure mathematician with completed PhD dissertation on PDE and Diffusion processes. Due to a standard dull Statistics course contrasted to the brilliant Dynkin's Probability courses, I had felt disgust to Statistics regarding it as a recipes list and had no interest in studying it.

In my entrance interview, Kolmogorov insisted that my duties include report to him on the Math activity in the V.V. Nalimov's Design of Experiment Dept of the Lab (which he never requested).

After spending considerable time on learning brilliant Linear models and optimal experimental design and especially Functional approach to Statistical Asymptotics, I understood that these topics were spectacular chapters of Mathematics.

My opinion: Statistics courses **for mathematicians** by qualified lecturers should emphasize profound beautiful **Math theories behind Statistical methods** in an **adequately** rigorous way.

Functional approach to Statistical Asymptotics

Apparently, the Kolmogorov-Smirnov **nonparametric** goodness of fit and homogeneity tests initiated the groundbreaking **functional approach** in Mathematical Statistics as the first fundamental Kolmogorov's contribution to the field following his fundamental Probability book published that same year.

Fisher regarded his Permutation and the so-called exact tests as more practically relevant and showed no interest in Kolmogorov's theory and results.
This apparently damaged feelings of his young rival. In Fisher's polemics with more rigorous J. Neyman and E. Pearson, Kolmogorov was on the latter's side, while the Fisher's polemics against the Bayesian statistics was supported by Kolmogorov.

The Functional approach to Statistical Asymptotics and **cross-entropy** have become later the key instrument in works by A. Wald (1947, 1949), Le Cam (1960, 1974) followed by fundamental book by I. Ibragimov and R. Khasminski (1981).

Formal Mathematical proficiency does not preclude errors

S. Wilks, a President of the American Statistical Association and Institute of Mathematical Statistics, editor of 'Annals of Mathematical Statistics', 1938-1949, pushed on the measure-theoretic language in papers submitted to this central journal. He postponed publication of S. Kullback, a giant of military cryptology, on a cross-entropy use in statistical inference over a **finite sample space with trivial sigma-algebras** until an ignorant in the Information theory coauthor R. A. Leibler competent in sigma-algebras was added. As a result, the cross entropy bears two names, those of Kullback-Leibler (1951). This pushing for formal rigor does not replace understanding of more advanced Mathematical theories such as functional approach to statistics. Ironically, S. Wilks in his monumental textbook on Mathematical Statistics, (1947, 1962) erroneously interchanged the order of $P(\sup ...)$ and $\sup P(...)$ in a key auxiliary Lemma 4.3.8. repeatedly used in subsequent proofs of asymptotic results. This made his conditions of their validity unwarrantly mild. This mistake continued to appear in subsequent books and papers.

ANOVA

Kolmogorov was one of the first to present a LINEAR ALGEBRA BASED intro to Linear models which he extended to groundbreaking **Hilbert space-based linear methods of Time Series prediction**.

Fisher's intuitive style of writing was addressed mostly to practical statisticians and Fisher's gaps in Mathematical rigor had to irritate Kolmogorov. Kolmogorov published a 35-pages long critical survey of Fisher's ANOVA methods published in the **Proceedings of the all-union conference on Statistics in Tashkent (PT-1949)**, and reproduced with minor changes in the fundamental textbook of **Mathematical Statistics by V. Romanovsky (R61)**.

It must be emphasized that 1949 was critical for the survival of the Soviet Mathematical Statistics which has been under attacks of Stalin's henchmen, who had succeeded previously in crushing applied Statistics (economic, sociological and biological, including the genetic applications developed by Kolmogorov). All of them had been already practically prohibited in the USSR.

Due to the severe political pressure, Kolmogorov had to agree to provocative Lysenko's slogan '**Science is an enemy of random**', making a witty remark that enemies should be carefully studied!

The leading Soviet Probabilists (with exception of Kolmogorov and few others) published political and philosophical insinuations addressed to prominent western Statisticians in the Proceedings mentioned above. Contrary to their style, the Kolmogorov criticism dealt mostly with inaccurate mathematical formulations admitting excessively broad interpretation while praising profound Fisher's inventions, notably introduction of the **INFORMATION** notion and its efficient applications.
This remarkable critical discussion of ANOVA appeared only in those hardly accessible PT-49 and R61, both **in Russian**. Its **English translation** obviously deserves publication.
Due to the prohibition of large scale statistical projects under Stalin, the implementation of the Fisher's pioneering Optimal multivariate Experimental Design theory had to wait until the influence of Stalin's henchmen disappeared.

Optimal Experimental Design

Decades of experiments of V. V.Nalimov on optimization of industrial output started in research facilities under the bars (the ill famed 'sharashka') supported by relevant literature lavishly were supplied from abroad. This form of prisoner's activity was widespread in Soviet Gulag Archipelago camps.

As soon as Kolmogorov was permitted to organize in 1966 his huge Interdisciplinary Statistical Laboratory inside the Moscow University, he pushed authorities to allow V. Nalimov to join with a team of followers for developing the Fisher's Design of Experiment. This initiative proved fruitful. V. Nalimov popularized optimization of experimental design over numerous applied fields.

Apart from applications, a major contribution was made in the **theory of optimal design**. The most spectacular in my opinion were:

i. initiated by S. N. Sokolov construction of iterative procedures for approximating optimal designs and

ii. extension of the Fisher's Response Surface methodology to multivariate factorial models under natural assumption of **sparsity**.

Population Genetics

Fisher's 'The Genetical Theory of Natural Selection'(1930), his theory of random genetic drift and the partial differential equation (1922) for time and place of gene fixation were beacons for Kolmogorov in his Population Genetics studies.

In addition, he published a popular discussion paper in 1940 in the Proceedings of Soviet Academy on elementary statistical confirmation of the Mendel Law based on an experimental material. In an immediate response (next issue), T. Lysenko claimed that biologists are not interested in Math formulas. Moreover, 'Correctness of the Stats depends on what theory CONTROLS its application'.
How close is it to the Stalin's remark:'It is not important how people votes, what matters is who counts!'

In his scandalous book 'Unknown Lysenko, 2014', L. Zhivotovsky tries to revive this Herostratus, whose crushing damage to the Soviet science and agriculture will never be forgotten and forgiven!

Population Genetics cont-d

In discussing Kolmogorov's contribution, we cannot loose from sight the grave political situation which forced him to abandon all defense attempts of the Genetics after 1948. Moreover, as a translation editor of the first edition of famous W. Feller's textbook on Probability, he had to abandon all its many parts devoted to genetical applications.

Kolmogorov was initially cautious in evaluating our first Phylogenetic applications of genetic drift published in Oxford University, where in particular, we corrected a **Fisher's mathematical (several times!) mistake** in multi-locus case. After our subsequent fitting parameters in the Kolmogorov derived stationary distribution of population frequencies under **selection** presence he fully acknowledged our achievements.

A spectacular extension of the Fisher's partial differential equation for time and place of gene fixation is presently called the Kolmogorov, Petrovsky and Piskunov reaction-diffusion equation.

Information and Statistics

The Fisher information introduced in 1925 in relation to ANOVA has been among the first groundbreaking achievements highlighting the limits of statistical experiments. It opened the way to famous Cramer-Rao inequalities, Le Cam theory, etc.
Kolmogorov immediately recognized its importance and has become the first mathematician who understood a similar revolution in Mathematics due to the Engineering C. Shannon's discoveries of the limits of the communication capacities.
Its importance for Mathematics development was immediately recognized by Kolmogorov, who was the first mathematician ever to understand Shannon's engineering style. He applied it for groundbreaking revolution in Dynamical systems, Approximation of functions, etc.
This led to numerous mathematical discoveries: from problems of functions approximation to dynamical systems and finally to the so-called Kolmogorov complexity and relation of complexity and randomness!

Information and Statistics cont-d

He wrote in his Editor's foreword to the Russian translation of S. Kullback, Information and Statistics(1965) presented here in my translation into my fancy English:

Development of methods for obtaining as complete as possible information from restricted data is a central statistics goal. It is natural that the first essential steps to make the notion of information precise were made by the creator of the principal part of contemporary Mathematical Statistics—R. Fisher...

The thinking style and analytic methods of the Information Theory should apparently reconstruct the building of Mathematical Statistics.

Both giants Fisher and Kolmogorov were fascinated by the phenomenon and notion of randomness. The fundamental Fisher's contribution to the analysis of relationships is insistence on **randomization** of explanatory variables which became a cornerstone of applied statistics. This reduces the **bias in estimates (due to lurking variables)** and increase variance.

Randomness and Complexity

For Kolmogorov, the randomness was rather a theoretical issue as it was the root of all theories which he developed in Probability Theory. He recognized as erroneous the early attempts made by von Mises to distinguish between random and nonrandom objects and worked his whole life correcting it.
The solution came from an unexpected side, when he started to develop universal compressors originated as such in Shannon's works. The fundamental step was to replace the Maximal Likelihood solution with that based on **complexity of the sample**, i.e obstacles to compress it to as short string as possible. It turned out that random strings had the maximal complexity.
An application of this Kolmogorov's discovery: Suppose a compressor is universal, i.e. it compresses a very long binary stationary string to a shorter one in the optimal way i.e. with length coinciding with the Shannon's lower bound. Let this binary stationary string be continued a bit obeying the same statistical law. Then the additional items of the extended compressed string have asymptotically the equally likely joint Bernoulli distribution.

Theory and applications

Both giants followed the famous J. Maxwell's principle: Nothing is more practical than a good theory. Fisher had a University degree in Earth Physics and started his research career in Eugenics. He came to Statistics occasionally being hired at the Rothamsted agricultural station for optimizing use of fertilizers. His main theoretical Statistics and Design of Experiment progress appeared logically in attempts to formalize his applied activity.

Kolmogorov repeatedly advised to employees of his Laboratory: try in consulting work to solve the problem of clients applying for help with adequate tools not necessarily close to those which you master. But finally, make your best in constructing an *appropriate mathematical model* of that problem. This program was accomplished in many of his applications: genetics, turbulence, statistics of verse language, artillery, tracking air targets, artificial sediment program, etc.

He managed to delegate continuation of his applied work to able people inspired by him, who converted these applications into new fields of science.

Conclusion

The displayed examples show: a great mathematician raises achievements of a great statistician to a new height! Even hard times could not stop this!
The unity of mathematical sciences as a Kolmogorov's goal, is now even more remote than ever. Statistics is now even more despised by pure mathematicians than before. Statistics courses are taught neither in HSE Dept of Math, nor in the Moscow Open University. In the US, situation is not better. Either the Departments of Mathematics and Statistics are separated by an 'Icy curtain', or mathematicians isolate statisticians inside Mathematics Departments preventing influx of new faculty members and graduate students unless they pay for education.

Specialists like me are viewed as don Quixote's type in spite of new challenges like big data, curse of dimensionality that urge development of principally new Math ideas and methods.
This leads inevitably to slowing the Statistics and Data Science progress down!

References

Fisher R. A. The correlation between relatives ... *Trans. R. Soc. Edinburgh*, **52**, 1918, 399–433.

Fisher R.A. On the dominance ratio, *Proc. Roy. Soc. Edinb.* 42, 1922, 321–341.

Fisher R. A. *Statistical methods for research workers*, Oliver and Boyd, London, 1925.

Fisher R. A. *The Genetical Theory of Natural Selection*, Clarendon Press, Oxford, 1930.

References cont-d

Kolmogorov A.N. *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Berlin: Springer- Verlag, 1933.
Kolmogorov A.N. Sulla determinazione empirica di una legge di distribuzione, *G. Inst. ital. attuar.*,. **4:1**, 1933, 83–91.
Kolmogorov A.N., Petrovsky I.G. and Piskunov N. S. Diffusion equation on mass growth and its application to a biological problem, *Bull. MSU, Math. and Mech.*, 1937.
Kolmogorov A. N. On the analytic methods of probability theory, *Uspekhi Mat. Nauk*, **5**, 1938, 5–41.
Kolmogorov A. N. On a new confirmation of Mendel laws. *Proceedings of the Soviet Academy of Sciences*,**27**, 1940, 38–42.

References cont-d

Kolmogorov A. N. On the proof of the method of least squares, *Uspekhi Mat. Nauk*, 1946, **1:1**, 1946, 57–70.
Kolmogorov A. N. Real meaning of the results of ANOVA, *Proceedings of the second all-union conference on mathematical statistics, 09/27-10/02, 1948.*, Academy of Uzbek. SSR, Tashkent, 1949, 240-268 (In Russian).
Kolmogorov A. N. Three ways of defining quantity of information, *Problems of Information transmission*, **1:1**, 1965, 3–11.

References cont-d

Lysenko T. D. On the paper by academician A.N. Kolmogorov, *Proceedings of the Soviet Academy of Sciences*, XXVII:1, 1940.
Zhivotovsky L. *Unknown Lysenko*, KMK Press, M. 2014.
Ryabko B., Astola J. and Malyutov M. *Compression-Based Methods of Prediction and Statistical analysis of Time Series: Theory and Appllications*, Springer International, 2016.

References cont-d

Romanovsky V. I. Mathematical Statistics, v.2, Academy of Uzbek. SSR, Tashkent, 1961 (In Russian).
Kullback S. *Information theory and Statistics*, N.Y., Wiley, 1959, Russian translation, Nauka, 1967.
Malyutov M., Passekov V., and Rychkov Yu. On the reconstruction of evolutionary trees of human populations resulting from random genetic drift. In: *Weiner, J.S., and J. Huizinga (eds.), The Assesment of Population Affinities in Man*, Clarendon Press, Oxford, 1972, 48–71.
Wilks S. *Mathematical Statistics*, Wiley, 1962.