BOLYAI SOCIETY
MATHEMATICAL STUDIES, 14

A Panorama of Hungarian Mathematics in the Twentieth Century, pp. 491–521.

Mathematical Statistics

ENDRE CSÁKI*

1. Introduction

The word "statistics" originated from the Latin word "status" and according to Kendall and Stuart (*The Advanced Theory of Statistics. Vol. 1. Distribution Theory,* Hafner Publishing Co., New York (1958)) "Statistics is the branch of scientific method which deals with the data obtained by counting or measuring the properties of populations". So the main task of Statistics is to collect data and make conclusions, usually called Statistical Inference. Thus statistical methods have been used for a long time also in Hungary, e.g., in Hungarian Central Statistical Office founded in 1867 and also in other organizations. The data are usually subject to random fluctuations and so the theory of statistical inference should be based on rigorous mathematical concepts treating random phenomena, i.e., on the Theory of Probability. Mathematical Statistics is the theory of statistical methods based on rigorous mathematical concepts of Probability. In this way we can consider Károly (Charles) Jordan as the founder of the probability and statistics school in Hungary, who wrote the first book on Mathematical Statistics in Hungary.

K. Jordan was born in 1871 in Budapest. He started his activity in mathematics, probability and statistics in particular, around 1910. He wrote 5 books and 83 scientific papers. His book on Mathematical Statistics appeared in 1927 in Hungarian and also in extended form in French (*Statistique Mathématique,* Gauthier-Villars (1927)). His books "Calculus of Finite Dif-

*Supported by the Hungarian National Foundation for Scientific Research, Grant No. T 029621 and T 037886.

ferences", appeared in 1939, and "Chapters on Classical Probability Theory", appeared in 1956 in Hungarian, are very important frequently cited works in probability and in the theory of difference equations. He had a number of students and a great influence in developing the theory of probability and statistics in Hungary in the first half of twentieth century.

Another prominent Hungarian probabilist and statistician was Ábrahám Wald who made significant contributions in these subjects. He was born in Kolozsvár, Hungary in 1902. In 1938 he went to the United States where he turned his main interest toward Statistics. Perhaps he is most well-known as a founder of sequential analysis and also the theory of statistical decision functions, but his basic results in other areas such as hypothesis testing, goodness of fit tests, tolerance limits, analysis of variance, nonparametric statistics, sampling, etc. are also very important.

One of the most distinguished mathematicians of 20th the century, János (John von) Neumann has also contributed to statistics. We refer to Section 5 for his results which appeared in The Annals of Mathematical Statistics.

Mathematical Statistics in Hungary became a vigourous subject in the fifties when the Mathematical Institute of the Hungarian Academy of Sciences was founded, featuring also a Department of Mathematical Statistics. First of all, the works and school in probability and statistics of Alfréd Rényi should be mentioned. His main works are in Probability Theory and Applications (see the Probability Theory Section), but he has also important contributions in Statistics. István Vincze, Károly Sarkadi, Lajos Takács and their collaborators made also significant contributions. The works of Béla Gyires in statistics at the Kossuth Lajos University, Debrecen, should also be mentioned.

2. Early statistics in Hungary

In the first half of the 20th century the outstanding works of K. Jordan in both theoretical and applied statistics are to be mentioned. His contributions to applied statistics concern a number of subjects such as meteorology, chemistry, population statistics, industry, etc. Even in his applied works he was very careful to base his investigations on rigorous theoretical disciplines. Since his works started well before Kolmogorov's fundamental works to establish rigorous mathematical probability theory, Jordan himself had to work on theoretical foundations of statistics, i.e., he had to develop a rigorous probability theory needed for the application in statistics. In a series of papers {23}, {24} etc. he gave a rigorous definition of probability and proved some fundamental theorems. This can be considered as a forerunner of Kolmogorov's theory. His book {31} is based on the author's experience of fifty years of research and thirty years of teaching. It is written in his lucid style and reflects his profound knowledge of the history of probability and his significant contributions to probability theory. He contributed also to other subjects in mathematics such as geometry and, first of all, to the theory of difference equations which he also needed in his research in probability and statistics. His book [78] contains 109 sections and gives a detailed account of the theory and application of statistics, equipped also with a number of illuminating examples. In order to give a flavour of the content, here is a selection of section titles: Definition of mathematical probability — Theorem of total probability — Mathematical expectation — Theorem of Bernoulli — Poisson limit — Theorem of Tchebychef — Theory of least squares — Elements of calculus of differences — Statistical classifications — Mean values — Standard deviation — Construction of statistical functions — Normal distribution — Asymmetric distributions — Approximation of functions — Method of moments — Method of least squares — Interpolations — Correlations — Independence — Correlation for nonnormal distribution — Correlation ratio — Theory of sampling — Contingency tables — Rank correlations.

In his far-reaching statistical investigations K. Jordan had to develop certain numerical methods such as interpolation, least square methods, etc. An outline of his contributions in this area is based on nice accounts of Jordan's life and works by L. Takács {48} and by B. Gyires {18}.

In {25}, {28} and {29} he formulated and proved the following result: Let $A_1, \dots, A_n$ arbitrary events and put

$$B_j = \sum_{1 \le i_1 < \cdots < i_j \le n} \mathbf{P}(A_{i_1} \cap \cdots \cap A_{i_j}).$$

Then the probability P_k that exactly k events occur among them is given by

$$P_k = \sum_{j=k}^{n} (-1)^{j-k} \binom{j}{k} B_j, \qquad k = 0, 1, \dots, n.$$

In {26} Jordan gave an interpolation formula

$$f(a+xh) = \sum_{m=0}^{n-1} C_m(x) \sum_{k=1}^{m+1} B_{mk} I_k + R_{2n},$$

where

$$C_m(x) = (-1)^m \binom{x+m-1}{2m}$$

and

$$B_{mk} = (-1)^{k+1} \binom{2m+1}{k+m} \frac{2k-1}{2m+1}.$$

I_k is obtained by linear interpolation:

$$I_k = \frac{x+k-1}{2k-1} f(a+kh) + \frac{k-x}{2k-1} f(a-kh+h).$$

Moreover, R_{2n} is a remainder for which

$$|R_{2n}| \le h^{2n} \left| \binom{n-\frac{1}{2}}{2n} D^{2n} f(a+\xi h) \right|$$

with some $-n+1 < \xi < n$.

A further contribution of K. Jordan concerns the following least square problem. Let $Y_0, Y_1, \ldots, Y_{N-1}$ be observations corresponding to $x = 0, 1, \ldots, N-1$. Find polynomials $f_n(x)$ of degree n such that

$$S_n = \sum_{x=0}^{N-1} \big(Y_x - f_n(x)\big)^2$$

is minimum. The solution of this problem given by C. Jordan in {27} is as follows: Consider the expansion

$$f_n(x) = \sum_{m=0}^{n} a_m U_m(x),$$

where the polynomials $U_m(x)$ are orthogonal with respect to $x = 0, 1, \ldots, N-1$, i.e.,

$$\sum_{x=0}^{N-1} U_i(x) U_j(x) = 0, \qquad i \ne j.$$

The Newton expansion of $U_m(x)$ has the form

$$U_m(x) = C_m \sum_{i=1}^{m} (-1)^{m+i} \binom{m+i}{m} \binom{N-i-1}{m-i} \binom{x}{i},$$

where the coefficients C_m can be chosen as

$$C_m = \left[(m+1) \binom{N}{m+1} \right]^{-1}.$$

The values of a_m which minimize S_n are independent of n. The Newton expansion of $f_n(x)$ is given by

$$f_n(x) = \sum_{m=0}^{n} \sum_{i=0}^{m} C_{mi} \Theta_m \binom{x}{i},$$

where

$$C_{mi} = (-1)^{m+i}(2m+1) \binom{m+i}{m} \frac{\binom{N-i-1}{m-i}}{\binom{N+m}{m}}$$

and

$$\Theta_m = \sum_{x=0}^{N-1} U_m(x) Y_x.$$

The mean square deviation is

$$\sigma^2 = \frac{1}{N} \sum_{x=0}^{N-1} \big(Y_x - f_n(x) \big)^2 = \frac{1}{N} \sum_{x=0}^{N-1} Y_x^2 - \Theta_0^2 - |C_{10}|\Theta_1^2 - \cdots - |C_{n0}|\Theta_n^2.$$

In {30} he introduced the notion of surprisingness. If the events $A_1, A_2, \ldots, A_i$ occur respectively $k_1, k_2, \ldots, k_i$ times in n trials $(k_1 + k_2 + \cdots + k_i = n)$, its probability being $P_{k_1,k_2,\ldots,k_i}$, then define the surprise index by

$$S = 1 - \frac{P_{k_1,k_2,\ldots,k_i}}{P_{m_1,m_2,\ldots,m_i}},$$

where $P_{m_1,m_2,\ldots,m_i}$ is the probability of the most probable system $m_1, m_2, \ldots, m_i$. This can be used in hypothesis testing to control the type 1 error by constructing a critical region which contains the points of the sample space with small probabilities, i.e., high surprise index. This approach is used to introduce Pearson's chi-square and other tests.

In the mid fifties of the last century one of the main tasks of the Statistics Department of the Mathematical Institute was to introduce statistical applications in practice, industrial quality control in particular. This is well reflected in producing the book {53} edited by I. Vincze, the head of the department. This book is written for engineers and technicians who wish to acquire familiarity with the theoretical foundations and with the applications of statistical quality control. An introduction into the elements of probability theory and mathematical statistics is given; the viewpoint of the quality control engineer is stressed and this also determines the choice of examples. Seven authors participated in writing the book; Part I, Theoretical foundations was written by K. Sarkadi and I. Vincze. Part II, Chapter 1 was written by Ágnes Fontányi and Mrs. Éva Vas and deals with statistical methods for the control of a manufacturing process. An interesting method developed by Á. Fontányi, K. Sarkadi and Mrs. É. Vas which uses order statistics, is discussed in detail. Part II, Chapter 2 (written by Károly Kollár) treats the statistical methods of acceptance control. The theory is adequately discussed, and sampling plans, with due references to the American sources, are given. Part III has the title “Applications of statistical quality control”. Chapters 1 and 2 (written by Tibor Tallián) deal with problems of organizing quality control in a plant. Chapter 3 (written by M. Borbély) discusses specific applications in the textile industry. A mathematical appendix to part I, a collection of statistical tables and a bibliography conclude the book.

3. Sequential analysis

A. Wald in the mid forties of the 20th century developed a statistical procedure called sequential method. Here the number of observed elements are not specified in advance, in certain situations the experimentation should be continued until there is enough evidence as to which decision should be made. This method can be described as follows. (cf. {67} and [192]).

Let X be a random variable having a density function $f(x)$. Consider two hypotheses, $H_0 : f(x) = f_0(x)$ and $H_1 : f(x) = f_1(x)$, where $f_0(x)$ and $f_1(x)$ are two different density functions. The sequential probability ratio test for testing H_0 against H_1 is given as follows: Put

$$Z_i = \log \frac{f_1(X_i)}{f_0(X_i)},$$

where X_i denotes the i-th observation on X. Let $\log A > 0$ and $\log B < 0$ be two constants depending on error probabilities. At each stage of the experiment, at the m-th trial consider the partial sum

$$S_m = Z_1 + \cdots + Z_m$$

and continue experimentation as long as $\log B < S_m < \log A$. The first time $S_m \notin (\log B, \log A)$, the experimentation is terminated. Accept H_1, resp. H_0 if $S_m \geq \log A$, resp. $S_m \leq \log B$. It is proved that with probability one, this sequential probability ratio test terminates in a finite (random) number of steps. Let ν be the number of observations required by this test. Then ν is a random variable, for which Wald showed that

$$\mathbf{E}\Big(e^{S_\nu t}\varphi(t)^{-\nu}\Big) = 1$$

for all points t on the complex plane for which the moment generating function $\varphi(t) = \mathbf{E}(e^{Zt})$ exists and its absolute value is not less than 1. Here $\mathbf{E}$ stands for expectation and $Z = \log\big(f_1(X)/f_0(X)\big)$. This is a celebrated identity, called Wald's identity in the literature today.

In order to investigate the number of observations required by this test, Wald shows also that

$$\mathbf{E}(S_\nu) = \mathbf{E}(\nu)\mathbf{E}(Z) = \mathbf{E}(\nu)\mathbf{E}\left(\log\frac{f_1(X)}{f_0(X)}\right).$$

Based on this identity, it is shown that if both the absolute value of the expectation and the variance of Z are small, then the expectation of ν can be approximated by

$$\mathbf{E}(\nu) \sim \frac{(1-\gamma)\log B + \gamma\log A}{\mathbf{E}(Z)},$$

where γ is the probability that H_1 is accepted, i.e. $S_\nu \geq \log A$. Let $\mathbf{E}_i$, $i = 1, 2$ denote the expectation under H_i, and let α be the probability of an error of the first kind (H_1 is accepted when H_0 is true), β be the probability of an error of the second kind (H_0 is accepted when H_1 is true). Then Wald gives the following inequalities for the expectation of ν:

$$\mathbf{E}_0(\nu) \geq \frac{1}{\mathbf{E}_0(Z)}\left((1-\alpha)\log\frac{\beta}{1-\alpha} + \alpha\log\frac{1-\beta}{\alpha}\right),$$

$$\mathbf{E}_1(\nu) \geq \frac{1}{\mathbf{E}_1(Z)}\left(\beta\log\frac{\beta}{1-\alpha} + (1-\beta)\log\frac{1-\beta}{\alpha}\right).$$

The denominators

$$\mathbf{E}_0(Z) = \int f_0(x) \log \frac{f_1(x)}{f_0(x)}\, dx \qquad \text{and} \qquad \mathbf{E}_1(Z) = \int f_1(x) \log \frac{f_1(x)}{f_0(x)}\, dx$$

are the same quantities as introduced and called I-divergence a few years later by S. Kullback (*Information Theory and Statistics,* Wiley, New York (1959)). So this can be considered as the first statistical application of information theory (see the Information Theory section). Further investigations on the expectation $\mathbf{E}(\nu)$ was given by A. Wald {68} and {69}. In subsequent papers {47} and {72} used the sequential method also for estimation problems.

4. Statistical decision functions

In one of his first papers in Statistics {63} he presented some of the main concepts of decision theory developed later in his book (*Statistical Decision Functions,* John Wiley and Sons, 1950). The basic idea can be described as follows.

Assume that experimentations are carried out on a random phenomenon, i.e., we have random observations $X = (X_1, X_2, \dots)$ on a random variable having a distribution function F. Usually F is unknown, but it is assumed to be known that F is a member of a given class Ω of distribution functions. Moreover, there is a space D, called decision space, whose elements d represent the possible decisions that can be made by the statistician in the problem under consideration. Let $W(F, d, x)$ be the loss when F is the true distribution function, the decision d is made and x is the observed value of X. A distance on the space D can be defined by

$$\rho(d_1, d_2) = \sup_{F,x} \big| W(F, d_1, x) - W(F, d_2, x) \big| .$$

A decision function $\delta(x)$ is a function which associates with each x a probability measure on D. Usually, this is a randomized decision function. In the particular case, when $\delta(x)$ for each x assigns the probability one to a single point d in D, the decision function is called nonrandomized. The aim of the statistician is to choose d so that W is in some sense minimized. Practically all statistical problems, including estimation, testing hypotheses, and the design of experiments, can be formulated in this way.

Given the sample point x and given that $\delta(x)$ is the decision function adopted, the expected value of the loss is given by

$$W^*(F, \delta, x) = \int_D W(F, d, x)\, d\delta(x).$$

The function

$$r(F, \delta) = \int_R W^*(F, \delta, x)\, dF(x)$$

is called the risk when F is true and δ is adopted. Wald's fundamental idea consists in considering this risk as the outcome of a zero-sum two-person game played by the statistician against nature. The main theorems refer to conditions under which the decision problem is strictly determined or has a Bayes and/or minimax solution.

In {11} it is shown that when Ω and D are finite and each element of Ω is atomless, then for any decision function $\delta(x)$ there exists an equivalent nonrandomized decision function $\delta^*(x)$, i.e. $r(F, \delta^*) = r(F, \delta)$ for all $F \in \Omega$.

In a series of papers Wald and his collaborators investigated the properties of statistical decision functions.

Wald and Wolfowitz (Characterization of the minimal complete class of decision functions when the number of distributions and decisions is finite, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability,* University of California Press, Berkeley and Los Angeles (1951), 149–157) consider a statistical decision problem with the space of distributions consisting of a finite number m of distinct probability distributions on a Euclidean space, and the space of decisions being also finite. Then every admissible decision function is a Bayes solution with respect to some a priori probability distribution, but the converse is not true. The concept of a Bayes solution with respect to a sequence $(\xi_i)_{i=1}^h = (\xi_{i1}, \ldots, \xi_{im})_{i=1}^h$ of a priori probability distributions is introduced. This is defined as follows: When $h = 1$ it is a Bayes solution with respect to ξ_1. When $h > 1$ it is a Bayes solution with respect to ξ_h if one restricts oneself only to those decision functions which are Bayes solutions with respect to the sequence $\xi_1, \ldots, \xi_{h-1}$. The main result of the paper is the following: A decision function is admissible if and only if it is a Bayes solution with respect to a sequence of $h \leq m$ a priori probability distributions $\xi_1, \ldots, \xi_h$ such that $\sum_{i=1}^h \xi_{ij} > 0$ for $j = 1, \ldots, m$. The proof involves a rather elaborate study of the intersections of a convex body with its supporting planes.

5. Asymptotic theory of testing and estimation

In a series of papers Wald worked out a theory on the asymptotic properties of tests and estimations. In {64} Wald showed that, under certain regularity conditions, the test based on maximum likelihood estimation is asymptotically most powerful and gives some examples for the most powerful tests. The connection between most powerful tests and shortest confidence intervals is treated in {65}.

Wald's asymptotic theory was developed in {66} generalizing and extending his previous works on the subject. He considered random vectors and multidimensional parameter space. The main feature of this paper is to reduce the general problem to the normal case and show optimum properties for the normal distribution. The general model is as follows: Let

$$f(x_1, x_2, \ldots, x_r; \theta_1, \ldots, \theta_k)$$

be a density function involving k unknown parameters $\theta_1, \ldots, \theta_k$ lying in a parameter space Ω. For a subset ω of Ω denote by H_ω the hypothesis that the parameter point lies in ω. Consider independent observations $X_1, X_2, \ldots, X_n$, where each X_i is a vector in the r-dimensional space and has density f. The maximum likelihood estimator $(\hat{\theta}_{1n}, \ldots, \hat{\theta}_{kn})$ of the parameters is the values of $\theta_1, \ldots, \theta_k$ for which $\prod_{i=1}^n f(X_i; \theta_1, \ldots, \theta_k)$ becomes a maximum. Wald considers asymptotic properties of tests constructed by the help of maximum likelihood estimators. He introduced the idea of most stringent test, which can be described briefly as follows (cf. {73}): define the envelope power function of a family of tests as the supremum at each parameter point of the powers of the tests and define the "shortcoming" of a test at a parameter point as the amount by which the power of the test falls short of the envelope power there. We may then define the maximum shortcoming of the test as the supremum over the parameter values of its shortcoming. A sequence of tests is asymptotically most stringent if the maximum amount by which its maximum stringency can be reduced tends to zero as n increases. Note that an asymptotically most powerful test is asymptotically most stringent.

Concerning estimation problems in {70}, Wald studies the asymptotic properties of maximum likelihood estimators in the case of stochastically dependent observations. Let (X_i), $i = 1, 2, \ldots$, be a sequence of random variables. It is assumed that for any n the first n variables admit a joint probability density function $f(x_1, \ldots, x_n; \theta)$ involving an unknown parameter θ. It is shown that under certain restrictions on the joint probability distribution the maximum likelihood equation has at least one root which is a consistent estimate of θ, and any root of the maximum likelihood equation which is a consistent estimate of θ is shown to be asymptotically efficient. Therefore the consistency of the maximum likelihood estimate implies its asymptotic efficiency, since this estimate is always a root of the maximum likelihood equation.

6. Randomness

It is an important problem in Statistics to test whether a sequence $(X_1, X_2, \ldots, X_N)$ of variables is a random one, i.e. they are independent and identically distributed (abbreviated i.i.d.). Tests for randomness are important in the analysis of time series, its investigations usually based on serial correlation coefficient

$$R_h = \frac{\sum_{i=1}^N X_i X_{h+i} - \left(\sum_{i=1}^N X_i\right)^2/N}{\sum_{i=1}^N X_i^2 - \left(\sum_{i=1}^N X_i\right)^2/N}.$$

Here h is a given positive integer and for $h+i > N$ the term X_{h+i} is to be replaced by X_{h+i-N}.

Wald and Wolfowitz in {75} proposed the following procedure: Let a_i be the observed value of X_i, $i = 1, \ldots, N$. Consider the subpopulation where the set $(X_1, \ldots, X_N)$ is restricted to permutations of $a_1, \ldots, a_N$ and for any particular permutation assign the probability $1/N!$ This determines the probability distribution of R_h in the subpopulation. They propose a randomness test based on this distribution. It is shown moreover that under some mild restrictions, the limiting distribution is normal. Further investigations for tests based on permutations of the observations is given in {76}. They show that under some conditions the weighted sum

$$L_N = \sum_{i=1}^N d_i X_i$$

has normal limiting distribution. As consequences of this result, they conclude that a number of statistics such as rank correlation coefficient, Pitman's two-sample statistics, Hotelling's generalized T statistics, etc. are asymptotically normal.

Related statistics were investigated by J. von Neumann with R. H. Kent, H. R. Bellinson and B. I. Hart (The mean square successive difference, *Ann. Math. Statist.,* **12** (1941), 153–162). Minimizing the effect of the trend on dispersion they considered

$$\delta^2 = \frac{\sum_{i=1}^{N} (X_i - X_{i+1})^2}{N-1}$$

as a competitor of the sample variance

$$s^2 = \frac{\sum_{i=1}^{N} (X_i - \overline{X})^2}{N},$$

where

$$\overline{X} = \frac{\sum_{i=1}^{N} X_i}{N}.$$

It was shown that if the X_i are independent normal with mean μ and variance σ^2, then the density of δ^2 is given by

$$p(\delta^2) = \frac{1}{\sigma^2\sqrt{3}}\, e^{-2\delta^2/3\sigma^2} J_0\left(\frac{i\delta^2}{3\sigma^2}\right),$$

where J_0 is the Bessel function of order zero.

It was also shown that the efficiency of δ^2 compared to s^2, the best estimation of σ^2, is $2(n-1)/(3n-4)$. The advantage of using δ^2 instead of s^2, is that it is robust in the sense that it has small effect when the mean of the observation is not constant, but can be changed in time. The ratio $\eta = \delta^2/s^2$ can also be used in testing randomness and this was investigated in further papers of J. von Neumann, see {61} and {62}.

7. Nonparametric tests, order statistics

Consider a random sample

$$(X_1, X_2, \dots, X_n)$$

of size n, coming from a population with (theoretical) distribution function $F(x) = \mathbf{P}(X_1 < x)$. In this context, the above random variables are independent and identically distributed. The order statistics are the rearrangement of sample elements according to their magnitude:

$$X_1^* \le X_2^* \le \cdots \le X_n^*.$$

The empirical or sample distribution function is defined by

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}\{X_i < x\} = \begin{cases} 0, & \text{if } x \le X_1^*, \\ \dfrac{k}{n}, & \text{if } X_k^* < x \le X_{k+1}^*, \\ 1, & \text{if } X_n^* < x. \end{cases}$$

Here $\mathbf{I}\{A\}$ stands for the indicator of the event A. Order statistics and empirical distribution functions are widely used in statistics, nonparametric statistics, in particular. Basic results are due to V. Glivenko {13} and F. P. Cantelli {5}: with probability one

$$\lim_{n\to\infty} D_n = 0,$$

where

$$D_n = \sup_{x\in R} \left| F_n(x) - F(x) \right|.$$

This theorem expresses the important fact that with probability one, the empirical distribution tends uniformly to the theoretical distribution. Hence, it can be effectively used for goodness of fit problems, i.e., to test whether a sample comes from a population with given distribution. This test is applicable thanks to a result of A. N. Kolmogorov {32} who determined the limiting distribution of D_n:

$$\lim_{n\to\infty} \mathbf{P}\big(\sqrt{n} D_n < y\big) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2y^2}, \qquad y > 0,$$

provided $F(x)$ is continuous. Later N. V. Smirnov {46} proved a one-sided version:

$$\lim_{n\to\infty} \mathbf{P}\big(\sqrt{n} D_n^+ < y\big) = \lim_{n\to\infty} \mathbf{P}\big(\sqrt{n} D_n^- < y\big) = 1 - e^{-2y^2}, \qquad y > 0,$$

where

$$D_n^+ = \sup_{x\in R} \big(F_n(x) - F(x)\big), \quad D_n^- = \sup_{x\in R} \big(F(x) - F_n(x)\big).$$

Note that in the above theorems the distributions of D_n, $D_n^{\pm}$ do not depend on the underlying distribution function $F(x)$.

In his fundamental paper, dedicated to Kolmogorov's fiftieth birthday, A. Rényi {38} proposed to modify the above statistics by considering the "relative error" of the empirical distribution. He defined the following statistics:

$$R_n^+(a) = \sup_{a \le F(x)} \frac{F_n(x) - F(x)}{F(x)}$$

and

$$R_n(a) = \sup_{a \le F(x)} \frac{\left| F_n(x) - F(x) \right|}{F(x)}.$$

Now these are called Rényi statistics in the literature. In the said paper, Rényi determined the limiting distributions:

$$\lim_{n\to\infty} \mathbf{P}\big(\sqrt{n}R_n^+(a) < y\big) = \sqrt{\frac{2}{\pi}} \int_0^{y\sqrt{\frac{a}{1-a}}} e^{-t^2/2}\, dt, \qquad y > 0, \tag{1}$$

$$\lim_{n\to\infty} \mathbf{P}\big(\sqrt{n}R_n(a) < y\big) = \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} e^{-\frac{(1-a)\pi^2(2k+1)^2}{8ay^2}}, \qquad y > 0. \tag{2}$$

Rényi considered also the more general statistics

$$R_n^+(a,b) = \sup_{a \le F(x) \le b} \frac{F_n(x) - F(x)}{F(x)}$$

and

$$R_n(a,b) = \sup_{a \le F(x) \le b} \frac{\left| F_n(x) - F(x) \right|}{F(x)}.$$

The proofs of (1) and (2) given by Rényi are based on his method presented in the same paper. Since, as remarked above, the statistics are distribution free, it is no loss of generality assuming that the sample comes from exponential distribution, i.e.

$$P(X_i < x) = 1 - e^{-x}, \qquad x > 0. \tag{3}$$

Rényi proves that in this case the variables X_k^* can be expressed in the form

$$X_k^* = \frac{\delta_1}{n} + \frac{\delta_2}{n-1} + \cdots + \frac{\delta_k}{n-k+1}, \qquad k = 1, 2, \ldots, n, \tag{4}$$

where the variables $\delta_1, \delta_2, \ldots, \delta_n$ are independent having exponential distribution as in (3). Based on this simple fact, with some clever manipulations, Rényi derives the above limiting distributions and also some other limit distributions such as the asymptotic normality of the sample quantile.

{22} based on (4), present elementary and simple method to derive certain distributions and conditional distributions concerning order statistics.

Another ingenious method concerning empirical distributions was developed by Lajos Takács, based on his celebrated ballot theorem. The classical ballot theorem due to M. J. Bertrand {3}, V. André {1} and É. Barbier {2} says that if in a ballot one candidate scores a votes, the other candidate scores b votes and $a \geq \mu b$ with a non-negative integer μ, then the probability that throughout the counting the number of votes registered for the first candidate is always greater than μ times the number of votes registered for the second candidate is given by

$$\mathbf{P} = \frac{a - \mu b}{a + b}.$$

L. Takács in a series of papers {49} and {50} etc. and in his book {51} presented a generalization of the ballot theorem and applied it in various problems, such as empirical distribution functions, queueing, dams, etc.

Let $\chi_1, \chi_2, \ldots, \chi_n$ be non-negative, cyclically interchangeable random variables and let $\tau_1 < \tau_2 < \cdots < \tau_n$ be the order statistics of a random sample, uniformly distributed on the interval $(0, t)$, and assume also that $\{\chi_r\}$ and $\{\tau_r\}$ are independent. Define

$$\chi(u) = \sum_{\tau_r \leq u} \chi_r, \qquad 0 \leq u \leq t.$$

Then

$$\mathbf{P}\big(\chi(u) \leq u,\ 0 \leq u \leq t \mid \chi(t) = y\big) = 1 - \frac{y}{t}, \qquad 0 \leq y \leq t.$$

Based on this extension of the ballot theorem, Takács (An application of a ballot theorem in order statistics, *Ann. Math. Statist.,* **35** (1964), 1356–1358) derives the exact distributions of the statistics

$$T_n^+(a, b, c) = \sup_{a \leq F(u) \leq b} \big(F_n(u) - cF(u)\big)$$

and

$$R_n^+(a,b,c) = \sup_{a \le F(u) \le b} \left(\frac{F_n(u) - cF(u)}{F(u)} \right).$$

Takács (On the comparison of a theoretical and an empirical distribution function, *J. Appl. Probab.,* **8** (1971), 321–330) proved a ballot-type theorem equivalent to the following: Let $X_1, \dots, X_n$ be i.i.d. random variables with distribution $P(X_1 = i) = q$, $i = 1, \dots, n$, $P(X_1 = n+1) = p$, where $p + nq = 1$, $0 < p < 1$. Let $X_1^* \le X_2^* \le \dots \le X_n^*$ be its order statistics. For $1 \le l \le n$, let A denote the event: There exist at least l distinct positive integers $k_1, k_2, \dots, k_l$ for which $X_{k_i}^* = k_i$ $(i = 1, 2, \dots, l)$. Then

$$\mathbf{P}(A) = q^l n(n-1) \dots (n-l+1).$$

Based on this result, the distribution of the number of intersections of $cF(x)$ with $F_n(x) + a/m$ was determined. K. Sarkadi {45} presented a simple elegant combinatorial proof of this theorem.

In the two-sample case B. V. Gnedenko and V. S. Korolyuk {14} developed a method based on random walk models. Let

$$(X_1, X_2, \dots, X_m) \quad \text{and} \quad (Y_1, Y_2, \dots, Y_n)$$

be two samples coming from continuous distributions. Let $F(x)$ and $G(x)$, resp. be their theoretical distribution functions and let $F_m(x)$ and $G_n(x)$, resp. be their empirical distribution functions. Testing the null hypothesis $H_0 : F(x) = G(x)$, a number of statistics has been investigated and their distributions, limiting distributions and other characteristics have been determined in the statistical literature. The idea of Gnedenko and Korolyuk was as follows: let

$$Z_1^* < Z_2^* < \dots < Z_{m+n}^*$$

denote the order statistics of the union of the two samples and define

$$\theta_i = \begin{cases} +1 & \text{if } Z_i^* = X_j \text{ for some } j, \\ -1 & \text{if } Z_i^* = Y_j \text{ for some } j \end{cases} \tag{5}$$

$i = 1, 2, \dots, m+n$. Put

$$S_0 = 0, \quad S_i = \theta_1 + \dots \theta_i, \quad i = 1, 2, \dots, 2n.$$

Then $(S_0, S_1, \dots, S_{m+n})$ is a random walk path with $S_{m+n} = m - n$ and under H_0 each of them has the same probability. This idea of Gnedenko and

Korolyuk enables one to determine the distributions of certain statistics by reducing the problems to combinatorial enumeration.

In a series of papers Vincze and his collaborators presented a number of results in this subject. His first result concerns the joint distribution of the maximum and its location in the case $m = n$: Let $x_0^{(n)}$ be the first point where $F_n(x) - G_n(x)$ takes its (one-sided) maximum for the first time. Then, under H_0, I. Vincze {54} showed that

$$\mathbf{P}\left(\max_{-\infty<x<\infty}\big(F_n(x)-G_n(x)\big)=\frac{k}{n},\ \frac{1}{2}\Big(F_n\big(x_0^{(n)}\big)+G_n\big(x_0^{(n)}\big)=\frac{r}{2n}\right)$$

$$=\mathbf{P}\left(\max_{1\le i\le r-1}S_i<k,\ S_r=k,\ \max_{r+1\le i\le 2n}S_i\le k\right)$$

$$=\frac{k(k+1)}{r(2n-r+1)}\frac{\binom{r}{\frac{r+k}{2}}\binom{2n-r+1}{n-\frac{r+k}{2}}}{\binom{2n}{n}},$$

$k = 1, 2, \ldots, n$; $r = k, k+2, \ldots, 2n-k$.

$$\lim_{n\to\infty}\mathbf{P}\left(\sqrt{\frac{n}{2}}\max_{-\infty<x<\infty}\big(F_n(x)-G_n(x)\big)<y,\ \frac{1}{2}\Big(F_n\big(x_0^{(n)}\big)+G_n\big(x_0^{(n)}\big)<z\right)$$

$$=\sqrt{\frac{2}{\pi}}\int_0^y\int_0^z\frac{u^2}{\big(v(1-v)\big)^{3/2}}\exp\left(-\frac{u^2}{v(1-v)}\right)du\,dv.$$

Similar results were given for the absolute maximum and its location.

Using his extension of the ballot theorem, L. Takács {52} gives joint distributions of the maximum and its location for different sample sizes. In the case when m divides n, Takács gives the joint exact distribution of the statistics

$$T^-(m,n)=\sup_{x\in R}\big(G_n(x)-F_m(x)\big)=\max_{1\le r\le n}\big(G_n(Y_r^*)-F_m(Y_r^*)\big),$$

and of $\rho^-(m,n)$, the smallest $1 \le r \le n$ for which the maximum is attained.

In a subsequent paper {55} I. Vincze proposed to use generating functions to determine distributions and joint distributions. He determined, e.g., the generating function of the above joint distribution.

Further results in this topic (in the case $m = n$):

Let

$$(X_1^* < X_2^* < \cdots < X_n^*), \qquad (Y_1^* < Y_2^* < \cdots < Y_n^*)$$

denote the ordered samples. Then

$$\gamma_n = \sum_{i=1}^{n} \mathbf{I}\{X_i^* > Y_i^*\}$$

is the so-called Galton statistics. For random walk paths defined above, γ_n is the number of i's such that $S_{2i-1} > 0$, $i = 1, \ldots, n$. K. L. Chung and W. Feller {6} showed that γ_n is uniformly distributed, i.e.,

$$\mathbf{P}(\gamma_n = g) = \frac{1}{n+1}, \qquad g = 0, 1, 2, \ldots, n.$$

The proof of Chung and Feller was based on generating function, while {40} gave a combinatorial proof by showing that there exists a bijection between random walk paths with $\gamma_n = 0$ and $\gamma_n = g$.

E. Csáki and I. Vincze in {8} considered the number of times the random walk crosses zero (number of intersections):

$$\lambda_n = \sum_{i=1}^{n-1} \mathbf{I}\{S_i = 0,\ S_{i-1}S_{i+1} < 0\}$$

and showed

$$\mathbf{P}(\lambda_n = \ell - 1) = \frac{2\ell}{n} \frac{\binom{2n}{n-\ell}}{\binom{2n}{n}}, \qquad \ell = 1, 2, \ldots, n.$$

The joint exact and limiting distribution of (γ_n, λ_n) was also given:

$$\mathbf{P}(\gamma_n = g,\ \lambda_n = \ell - 1)$$

$$= \frac{1}{\binom{2n}{n}} \frac{\ell^2}{2g(n-g)} \binom{2g}{g - \ell/2} \binom{2n-2g}{n-g-\ell/2}$$

for ℓ even. A similar result was given for ℓ odd. For the limiting distribution it was shown that

$$\lim_{n\to\infty} \mathbf{P}\big(\gamma_n \le zn, \lambda_n \le y\sqrt{2n}\,\big)$$

$$= \sqrt{\frac{2}{\pi}} \int_0^y \int_0^z \frac{u^2}{\big(v(1-v)\big)^{3/2}} \exp\left(-\frac{u^2}{2v(1-v)}\right) du\, dv.$$

Another use of the generating function method is found in {9} where the joint distribution of the maximum and the number of intersections was given in the form

$$\sum_{n=1}^{\infty}\binom{2n}{n}\mathbf{P}\left(\max_x \bigl|F_n(x)-G_n(x)\bigr| = \frac{k}{n},\ \lambda_n=\ell-1\right)z^n$$

$$=2\left(\frac{w-w^k}{1-w^{k+1}}\right)^{\ell}, \qquad \ell,k=1,2,\dots,$$

where

$$w=\frac{1-\sqrt{1-4z}}{1+\sqrt{1-4z}}, \qquad |z|<\frac{1}{4}.$$

Vincze's idea in determining joint distributions was to construct tests based on a pair of statistics (instead of one single statistic) in order to improve the power of the tests. For details see {56}. This idea however deserves further investigations even today.

In {37} the two-sample problem is treated for different sample sizes by investigating

$$\max_x \bigl(nF_n(x)-mG_m(x)\bigr)$$

and related quantities.

The power of the Kolmogorov–Smirnov two-sample test is treated in {57}. In {58} the analogues of Gnedenko–Korolyuk distribution is given both for discontinuous random variables and for the two-dimensional case.

K. Sarkadi in {43}, by using the well-known inclusion-exclusion principle in combinatorics, gives an alternative method of deriving the exact distribution of the Kolmogorov–Smirnov statistics for both the one-sample and the two-sample cases.

In the two-sample case an important contribution was made by A. Wald and J. Wolfowitz {74}, who constructed a test based on the number of runs. Consider two samples $(X_1, X_2, \dots, X_m)$ and $(Y_1, Y_2, \dots, Y_n)$ as before, and the variables θ_i defined by (5). A subsequence $\theta_{s+1}, \theta_{s+2}, \dots, \theta_{s+r}$ is called a run, if $\theta_{s+1} = \theta_{s+2} = \dots \theta_{s+r}$ but $\theta_s \neq \theta_{s+1}$ when $s > 0$ and $\theta_{s+r} \neq \theta_{s+r+1}$ when $s + r < m + n$. Let U be the number of runs in the sequence $(\theta_1, \theta_2, \dots, \theta_{m+n})$. The exact distribution, mean and variance of

U under the null hypothesis $F(x) = G(x)$ was given for continuous F and it was shown that U is asymptotically normal with mean and variance

$$E(U) = \frac{2mn}{m+n} + 1,$$

$$\operatorname{Var}(U) = \frac{2mn(2mn - m - n)}{(m+n)^2(m+n-1)}.$$

Hence using either the exact (for small sample sizes) or the asymptotic (for large sample sizes) distribution, a test can be constructed with critical region $U < u_0$, so that $\mathbf{P}(U < u_0) = \beta$, where β is a predetermined level of significance. In other words the null hypothesis $H_0 : F(x) = G(x)$ is rejected if the number of runs in the combined sample is too small. Wald and Wolfowitz have also shown that the test is consistent against any alternatives $F(x) \neq G(x)$.

Z. W. Birnbaum and I. Vincze {4} proposed a test based on order statistics, which can replace Student's t test. Let $X_1, \ldots, X_n$ be a random sample from a population with continuous distribution function $F(x)$. Let $X_1^* < X_2^* < \ldots X_n^*$ be their order statistics. For a given $0 < q < 1$ the q-quantile is defined by

$$\mu_q = \inf\left\{x : F(x) = q\right\}$$

and the corresponding sample quantile is defined as the order statistic X_k^* such that

$$\left|\frac{k}{n} - q\right| \leq \frac{1}{2n}.$$

Consider the statistic

$$S_{n,k,r,s} = \frac{X_k^* - \mu_q}{X_{k+s}^* - X_{k-r}^*}$$

that can be used for testing the location parameter when the scale parameter is unknown for a general distribution. Exact and limiting distributions are derived for this statistic under some mild conditions on the distribution function F.

B. Gyires in {20} investigated asymptotic results for linear rank statistic defined as

$$S = \sum_{j=1}^{m} f_j\left(x_{R(X_j)}^{(j)}\right),$$

where $X_1, \ldots, X_n$ are i.i.d. continuous random variables, $R(X_j)$ denotes the rank of X_j, $N = m + n$, $x_i^{(j)}$ $(i = 1, \ldots, N,\ j = 1, \ldots, m)$ are real numbers in $(0, 1)$, and f_j are continuous functions on $[0, 1]$ with bounded variation. Let

$$V = \sum_{j=1}^{m} f_j(\eta_j),$$

where the η_j are i.i.d. uniform $(0, 1)$ random variables. An upper bound of

$$\left| \varphi(t) - N^m \prod_{j=1}^{m} (n + j)^{-1} \varphi_V(t) \right|$$

is given, where $\varphi(t)$ and $\varphi_V(t)$ are the characteristic functions of S and V, respectively. The bound is then exploited to prove that, as $n \to \infty$ with m remaining fixed, S converges weakly to V if and only if the discrepancy of the sequence $\left(x_i^{(j)}\right)_{i=1}^{n}$ from what is called a uniform sequence tends to zero. Application of this result to certain two-sample rank tests are also given.

Further results on asymptotic properties for linear rank and order statistics can be found in {16}, {17}, {19} and {21}. In these papers Gyires gives a necessary and sufficient condition for linear order statistics to have a limit distribution and he studies the case when the limit distribution is normal in particular. Limit distributions are also given for linear order statistics in the case when the observations are not necessarily independent. A doubly ordered linear rank statistic is also investigated. The methods employed by Gyires uses matrix theory, in particular Gábor Szegő's result concerning the eigenvalues of Toeplitz and Hankel matrices. For further comments in this regard we refer to the Section on Probability Theory.

8. Goodness of fit tests

An important problem in Mathematical Statistics is to test whether a random sample comes from a well-defined family of distributions. E.g., tests for normality or other goodness of fit tests are aimed to decide whether a sample comes from normal, or other distributions usually involving nuisance parameters, i.e., we are faced with a composite hypothesis. The most commonly used goodness of fit tests are Pearson's χ^2-tests. In the case of simple hypothesis $H_0 : F(x) = F_0(x)$ with given F_0 this is based on the statistics

$$\chi^2 = \sum_{i=1}^{k} \frac{(\nu_i - Np_i)^2}{Np_i},$$

where the range of the variable is divided into a number k of class intervals, N is the sample size, ν_i stands for the number of sample elements in ith class and p_i is the probability that a sample element falls into the ith class. H. B. Mann and A. Wald {34} investigated the problem of optimal choice of class intervals. They show that

$$k = k_N = 4\left(\frac{2(N-1)^2}{c^2}\right)^{1/5}$$

and $p_i = 1/k$, $i = 1, \ldots, k$ is in certain sense optimal, where c is a constant depending on the probability of the critical region.

E. Csáki and I. Vincze in {10} proposed a modification of the Pearson's χ^2-statistic:

$$\overline{\chi}^2 = \sum_{i=1}^{k} \left(\frac{\overline{X}_{(i)} - E_i}{\sigma_i}\right)^2 \nu_i,$$

where E_i and σ_i^2, resp. are the expectation and variance, resp. of the observations in the ith class and $\overline{X}_{(i)}$ are the mean value of the observations in the ith class. It was shown that (for fixed k) the limiting distribution of $\overline{\chi}^2$ statistic is $chi - square$ with k degrees of freedom (instead of $k-1$ degrees of freedom of Pearson's χ^2).

For simple hypotheses the one-sample Kolmogorov–Smirnov type tests discussed in Section 6 are also applicable for goodness of fit problems. In case of composite hypotheses, i.e., when parameters are unknown, a usual procedure is to estimate the parameters and apply a modified χ^2-test. But in some cases this has disadvantages. K. Sarkadi {39}, {41} in the case of normality test, presented a method which reduces the problem of composite hypothesis to a simple one. Assume first that we want to test normality based on the sample $(X_1, \ldots, X_n, X_{n+1})$ in the case when the expectation is unknown and the variance is known. Define

$$\overline{Y} = \frac{\overline{X} - X_{n+1}}{\sqrt{n+1}},$$

where

$$\overline{X} = \frac{X_1 + \cdots + X_n}{n}.$$

Put

$$Y_i = \overline{Y} + X_i - \overline{X}, \qquad i = 1, \ldots, n.$$

If $X_1, \ldots, X_{n+1}$ are independent random variables having normal distribution with expectation μ and variance σ^2, then $Y_1, \ldots, Y_n$ are independent random variables, each normally distributed with expectation zero and variance σ^2. This way the normality test with unknown expectation reduces to the normality test with expectation 0.

Similarly, if the variance is unknown and the expectation is known (assuming to be equal to zero without loss of generality), so that $(X_1, \ldots, X_{n+1})$ are i.i.d. mean zero normal random variables with unknown variance, Sarkadi gives the following transformation:

$$Y_i = X_i \frac{s'}{s}, \qquad i = 1, \ldots, n,$$

where

$$s = \sqrt{\frac{X_1^2 + \cdots + X_n^2}{n}},$$

$$s' = \psi_n \left(\frac{|X_{n+1}|}{s} \right),$$

and the function $\psi_n(t)$ is defined by the following relation:

$$\int_{\psi_n^2}^{\infty} u^{(n-1)/2} \exp(-u/2)\, du = \frac{2^{n/2+1}\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}} \int_{-\infty}^{t} \left(1 + \frac{u^2}{n}\right)^{-(n+1)/2} du.$$

It is shown that $(Y_1, \ldots, Y_n)$ are independent standard normal variables. Hence testing normality in the case of composite hypothesis is reduced to that of simple hypothesis.

Similarly, if $(X_1, X_2, \ldots, X_{n+2})$ are independent random variables each having normal distribution with expectation μ and variance σ^2, Sarkadi gives a transformation based on this sample, resulting in $(Y_1, Y_2, \ldots, Y_n)$, independent standard normal variables. The advantage of Sarkadi's transformation is that random numbers are avoided and he also shows that the transformation is optimal in some sense.

Sarkadi (The asymptotic distribution of certain goodness of fit test statistics, *Lecture Notes in Statistics* **8**, Springer, New York (1981), 245–253) investigated goodness of fit statistics of the form

$$W_n = \frac{\left(\sum_{i=1}^n a_{in} X_i^*\right)^2}{\sum_{i=1}^n \left(X_i - \overline{X}\right)^2},$$

where $X_1^* < \cdots < X_n^*$ are order statistics, $a_{1n}, \ldots, a_{nn}$ are appropriately chosen constants, and $\overline{X}$ is the sample mean. Sufficient conditions are given for W_n to have asymptotically normal distribution. It is shown that many statistics proposed for testing goodness of fit are of the above type with different values of a_{in}. Asymptotic properties of these tests are discussed and some of the tests are shown to be inconsistent for specific alternatives. In the case when $a_{in} = m_{in}/\left(\sum_{j=1}^n m_{jn}^2\right)^{1/2}$, with $m_{in} = E(X_i^*)$, this is the Shapiro–Francia test for which K. Sarkadi {44} proved consistency.

In {33} a goodness of fit test is proposed for testing uniformity. The test statistic is

$$J = n^2 \sum_{i=1}^n d_i^2 - n\left(\sum_{i=1}^n d_i\right)^2,$$

where $d_i = \left(X_i^* - i/(n+1)\right)/i(n-i+1)$ and the X_i^* are order statistics from a sample of size n. The Monte Carlo method is used to compare the test with some competitors.

9. Cramér–Fréchet–Rao inequality

Let $X = (X_1, X_2, \ldots, X_n)$ be a sample from a distribution having (joint) density $p(x;\theta) = p(x_1, x_2, \ldots, x_n; \theta)$ with respect to a measure μ, where θ is a parameter and we want to estimate its function $g(\theta)$. Let $t(X)$ be an unbiased estimator of $g(\theta)$, i.e. $\mathbf{E}_\theta\big(t(X)\big) = g(\theta)$. M. Fréchet {12}, C. R. Rao {36} and H. Cramér {7} gave the following inequality:

$$\mathrm{Var}_\theta\left(t(X)\right) \geq \frac{\left(g'(\theta)\right)^2}{I(\theta)},$$

with

$$I(\theta) = \int \left(\frac{\partial p}{\partial \theta}\right)^2 p(x;\theta)\, dx.$$

I. Vincze {59} and {60} for fixed θ, θ' considered the mixture

$$p_\alpha = p_\alpha(x;\theta,\theta') = (1-\alpha)p(x;\theta) + \alpha p(x;\theta'), \quad 0<\alpha<1$$

with α being a new parameter. Then

$$\hat{\alpha} = \frac{t(X)-g(\theta)}{g(\theta')-g(\theta)}$$

is an unbiased estimator of α.

It follows that

$$\mathrm{Var}_\alpha(\hat{\alpha}) \ge \frac{1}{J_\alpha(\theta,\theta')},$$

where

$$J_\alpha(\theta,\theta') = \int \frac{\big(p(x;\theta')-p(x;\theta)\big)^2}{p_\alpha(x;\theta,\theta')}\,d\mu.$$

Then

$$(1-\alpha)\,\mathrm{Var}_\theta\big(t(X)\big) + \alpha\,\mathrm{Var}_{\theta'}\big(t(X)\big) \ge \frac{1}{J_\alpha(\theta,\theta')} - \alpha(1-\alpha)$$

and in the case when $\mathrm{Var}_\theta\big(t(x)\big)$ does not depend on θ, Vincze concluded the following lower bound:

$$\mathrm{Var}\big(t(X)\big) \ge \sup_\alpha \sup_{\theta'} \alpha(1-\alpha)\big(g(\theta')-g(\theta)\big)^2 \left(\frac{1}{\alpha(1-\alpha)J_\alpha} - 1\right).$$

In certain cases this gives a reasonably good bound. This problem was further investigated by M. L. Puri and I. Vincze {35} and Z. Govindarajulu and I. Vincze {15}. It was shown among others that for the translation parameter of the uniform distribution this lower bound is of order n^{-2}, which is attainable.

10. Estimation problems

An interesting estimation problem is treated in {71}. Let $X_1, X_2, \dots$ be an infinite sequence of random variables, such that for each n the variables $X_1, \dots, X_n$ admit a continuous joint probability density $f_n(x_1, \dots, x_n \mid \theta, \xi_1, \dots, \xi_n)$, where $\theta, \xi_1, \dots, \xi_n$ are unknown parameters, all of which are restricted to finite intervals. Then $t_n(X_1, \dots, X_n)$ is said to be a uniformly consistent estimate of θ if $P\left(|t_n - \theta| < \delta\right) \to 1$ as $n \to \infty$, for any $\delta > 0$, uniformly in θ and the ξ's. Necessary and sufficient conditions for the existence of a uniformly consistent estimate are given. An information function is defined for the present case, and a sufficient condition for the nonexistence of a uniformly consistent estimate is given in terms of the information function. In the particular case when the X_i are independent, the total information contained in the first n observations is equal to the sum of the amounts of information contained in each observation separately.

Another interesting estimation problem is treated by K. Sarkadi. In statistics the following selection procedure often occurs. Let $\mu_1, \dots, \mu_n$ be parameters characterizing different populations. The parameters are unknown but we know their unbiased estimators $X_1, X_2, \dots, X_n$, i.e., $E(X_i) = \mu_i$. Given these estimators, one population is selected according to some predetermined decision rule. Suppose, e.g., the population of the lowest value X_i is selected, because this proves to be the highest quality among the possible choices. In this case $\min_i X_i$ as the estimator of the parameter of the corresponding population is obviously biased. This problem was treated by K. Sarkadi {42} who proved that though no unbiased estimation with finite variance exists in general, he suggests randomized estimations with arbitrarily small bias. The variance of the estimator however tends to infinity when the bias tends to zero.

References

[78] Jordán, Károly, *Matematikai Statisztika (Mathematical Statistics),* Természet és Technika, Volume 4, Athenaeum (Budapest, 1927, in Hungarian), and *Statistique mathématique,* Gauthier-Villars (Paris, 1927, in French).

[192] Wald, Ábrahám, *Sequential Analysis,* John Wiley and Sons (New York) – Chapman and Hall (London, 1947).

{1} V. André, Solution directe du problème résolu par M. J. Bertrand, *C.R. Acad. Sci. Paris,* **105** (1887), 436–437.

{2} É. Barbier, Généralisation du problème résolu par M. J. Bertrand, *C.R. Acad. Sci. Paris,* **105** (1887), 407.

{3} M. J. Bertrand, Solution d'un problème, *C.R. Acad. Sci. Paris,* **105** (1887), 369.

{4} Z. W. Birnbaum and I. Vincze, Limiting distributions of statistics similar to Student's t, *Ann. Statist.,* **1** (1973), 958–963.

{5} F. P. Cantelli, Sulla determinazione empirica delle leggi di probabilità, *Giorn. Ist. Ital. Attuari,* **4** (1933), 421–424.

{6} K. L. Chung and W. Feller, Fluctuations in coin tossing, *Proc. Nat. Acad. Sci. USA,* **35** (1949), 605–608.

{7} H. Cramér, *Mathematical Methods of Statistics,* Princeton University Press (Princeton, 1946).

{8} E. Csáki and I. Vincze, On some problems connected with the Galton test, *Publications of the Math. Inst. Hung. Acad. Sci.,* **6** (1961), 97–109.

{9} E. Csáki and I. Vincze, Two joint distribution laws in the theory of order statistics, *Mathematica, Cluj* **5** (1963), 27–37.

{10} E. Csáki and I. Vincze, On limiting distribution laws of statistics analogous to Pearson's chi-square, *Mathematische Operationsforschung und Statistik, Ser. Statistics,* **9** (1978), 531–548.

{11} A. Dvoretzky, A. Wald and J. Wolfowitz, Elimination of randomization in certain statistical decision procedures and zero-sum two-person games, *Ann. Math. Statist.,* **22** (1951), 1–21.

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

{13} V. Glivenko, Sulla determinazione empirica delle leggi di probabilità, *Giorn. Ist. Ital. Attuari,* **4** (1933), 92–99.

{14} B. V. Gnedenko and V. S. Korolyuk, On the maximum discrepancy between two empirical distribution functions, *Dokl. Akad. Nauk SSSR,* **80** (1951), 525–528; English translation: *Selected Transl. Math. Statist. Probab., Amer. Math. Soc.,* **1** (1951), 13–16.

{15} Z. Govindarajulu and I. Vincze, The Cramér–Fréchet–Rao inequality for sequential estimation in non-regular case, *Statistical Data Analysis and Inference* (Ed. Y. Dodge), North Holland (Amsterdam, 1989), 257–268.

{16} B. Gyires, On limit distribution theorems of linear order statistics, *Publ. Math. Debrecen,* **21** (1974), 95–112.

{17} B. Gyires, Linear order statistics in the case of samples with non-independent elements, *Publ. Math. Debrecen,* **22** (1975), 47–63.

{18} B. Gyires, Jordan Károly élete és munkássága, *Alkalmazott Matematikai Lapok,* **1** (1975), 274–298.

{19} B. Gyires, Normal limit-distributed linear order statistics, *Sankhyā,* Ser. A **39** (1977), 11–20.

{20} B. Gyires, Linear rank statistics generated by uniformly distributed sequences, *Colloq. Math. Soc. János Bolyai* **32**, North-Holland (Amsterdam, 1982), 391–400.

{21} B. Gyires, Doubly ordered linear rank statistics, *Acta Math. Acad. Sci. Hungar.,* **40** (1982), 55–63.

{22} G. Hajós and A. Rényi, Elementary proofs of some basic facts in the theory of order statistics, *Acta Math. Acad. Sci. Hung.,* **5** (1954), 1–6.

{23} K. Jordan, A valószínűség a tudományban és az életben (Probability in science and life, in Hungarian), *Természettudományi Közlöny,* **53** (1921), 337–349.

{24} K. Jordan, On probability, *Proceedings of the Physico-Mathematical Society of Japan,* **7** (1925), 96–109.

{25} K. Jordan, A valószínűségszámítás alapfogalmai (Fundamental concepts of the theory of probability) *Mathematikai és Physikai Lapok,* **34** (1927), 109–136.

{26} K. Jordan, Sur une formule d'interpolation *Atti del Congresso Internazionale dei Matematici,* Bologna, Vol. **6** (1928), 157–177.

{27} C. Jordan, Approximation and graduation according to the principle of least squares by orthogonal polynomials, *Ann. Math. Statist.,* **3** (1932), 257–357.

{28} Ch. Jordan, Le théorème de probabilité de Poincaré, généralisé au cas de plusieurs variables indépendantes, *Acta Scientiarum Mathematicarum (Szeged),* **7** (1934–35), 103–111.

{29} Ch. Jordan, Problèmes de la probabilité des épreuves répétées dans le cas général, *Bull. de la Société Mathématique de France,* **67** (1939), 223–242.

{30} K. Jordan, Következtetések statisztikai észlelésekből (Statistical inference, in Hungarian), *Magyar Tud. Akad. Mat. Fiz. Oszt. Közleményei,* **1** (1951), 218–227.

{31} K. Jordan, *Fejezetek a klasszikus valószínűségszámításból,* Akadémiai Kiadó (Budapest, 1956, in Hungarian), *Chapters on the classical calculus of probability* (1972, in English).

{32} A. N. Kolmogorov, Sulla determinazione empirica di una legge di distribuzione, *Giorn. Ist. Ital. Attuari,* **4** (1933), 83–91.

{33} P. Kosik and K. Sarkadi, A new goodness-of-fit test, *Probability Theory and Mathematical Statistics with Applications* (Visegrád, 1985), Reidel (Dordrecht, 1988), 267–272.

{34} H. B. Mann and A. Wald, On the choice of the number of class intervals in the application of the chi square test, *Ann. Math. Statist.,* **13** (1942), 306–317.

{35} M. L. Puri and I. Vincze, On the Cramér–Fréchet–Rao inequality for translation parameter in the case of finite support, *Statistics,* **16** (1985), 495–506.

{36} C. R. Rao, Information and accuracy attainable in the estimation of statistical parameter, *Bull. Calcutta Math. Soc.,* **37** (1945), 81–91.

{37} J. Reimann and I. Vincze, On the comparison of two samples with slightly different sizes, *Publications of Math. Inst. Hung. Acad. Sci.,* **5** (1960), 293–309.

{38} A. Rényi, On the theory of order statistics, *Acta Math. Acad. Sci. Hung.,* **4** (1953), 191–231.

{39} K. Sarkadi, On testing for normality, *Magyar Tud. Akad. Mat. Kutató Int. Közl.,* **5** (1960), 269–275.

{40} K. Sarkadi, On Galton's rank order test, *Magyar Tud. Akad. Mat. Kutató Int. Közl.,* **6** (1961), 127–131.

{41} K. Sarkadi, *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability* **I**, Univ. California Press (Berkeley, Calif., 1967), 373–387.

{42} K. Sarkadi, Estimation after selection, *Studia Sci. Math. Hung.,* **2** (1967), 341–350.

{43} K. Sarkadi, On the exact distributions of statistics of Kolmogorov–Smirnov type, *Collection of articles dedicated to the memory of Alfréd Rényi, II. Period. Math. Hungar.,* **3** (1973), 9–12.

{44} K. Sarkadi, The consistency of the Shapiro–Francia test, *Biometrika,* **62** (1975), 445–450.

{45} K. Sarkadi, A direct proof for a ballot type theorem, *Colloq. Math. Soc. János Bolyai,* **32**, North-Holland (Amsterdam, 1982), 785–794.

{46} N. V. Smirnov, On the empirical distribution function (in Russian), *Mat. Sbornik,* **6 (48)** (1939), 3–26.

{47} C. Stein and A. Wald, Sequential confidence intervals for the mean of a normal distribution with known variance, *Ann. Math. Statist.,* **18** (1947), 427–433.

{48} L. Takács, Charles Jordan, 1871–1959, *Ann. Math. Statist.,* **32** (1961), 1–11.

{49} L. Takács, Ballot problems, *Z. Wahrsch. verw. Geb.,* **1** (1962), 154–158.

{50} L. Takács, Fluctuations in the ratio of scores in counting a ballot, *J. Appl. Probab.,* **1** (1964), 393–396.

{51} L. Takács, *Combinatorial Methods in the Theory of Stochastic Processes,* Wiley (New York, 1967).

{52} L. Takács, On maximal deviation between two empirical distribution functions, *Studia Sci. Math. Hung.,* **10** (1975), 117–121.

{53} I. Vincze, *Statisztikai minőségellenőrzés: az ipari minőségellenőrzés matematikai statisztikai módszerei* (*Statistical quality control: The mathematical-statistical methods of industrial quality control,* in Hungarian), Közgazdasági és Jogi Könyvkiadó, Budapest (1958).

{54} I. Vincze, Einige zweidimensionale Verteilungs- und Grenzverteilungssätze in der Theorie der geordneten Stichproben, *Publications of the Math. Inst. Hung. Acad. Sci.,* **2** (1958), 183–209.

{55} I. Vincze, On some joint distribution and joint limiting distribution in the theory of order statistics, II, *Publications of the Math. Inst. Hung. Acad. Sci.,* **4** (1959), 29–47.

{56} I. Vincze, Some questions connected with two sample tests of Smirnov type, *Proc. of the Fifth Berkeley Symp. on Math. Stat. and Prob.,* Univ. of Calif. Press Vol. 1 (1967), 654–666.

{57} I. Vincze, On the power of the Kolmogorov–Smirnov two-sample test and related nonparametric tests, *Studies in Mathematical Statistics,* Publishing House of the Hung. Acad. Sci. (Budapest, 1968), 201–210.

{58} I. Vincze, On some results and problems in connection with statistics of the Kolmogorov–Smirnov type, *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability,* **I**, Univ. California Press (1972), 459–470.

{59} I. Vincze, On the Cramér–Fréchet–Rao inequality in the non–regular case, *Contributions to Statistics. Hájek Memorial Volume,* Academia (Prague, 1979), 253–262.

{60} I. Vincze, On nonparametric Cramér–Rao inequalities, *Order Statistics and Nonparametrics* (Ed. P. K. Sen and I. A. Salama) North Holland (Amsterdam, 1981), 439–454.

{61} J. von Neumann, Distribution of the ratio of the mean square successive difference to the variance, *Ann. Math. Statist.,* **12** (1941), 367–395.

{62} J. von Neumann, A further remark concerning the distribution of the ratio of the mean square successive difference to the variance, *Ann. Math. Statist.,* **13** (1942), 86–88.

{63} A. Wald, Contributions to the theory of statistical estimation and testing hypotheses, *Ann. Math. Statist.,* **10** (1939), 299–326.

{64} A. Wald, Asymptotically most powerful tests of statistical hypotheses, *Ann. Math. Statist.,* **12** (1941), 1–19.

{65} A. Wald, Asymptotically shortest confidence intervals, *Ann. Math. Statist.,* **13** (1942), 127–137.

{66} A. Wald, Tests of statistical hypotheses concerning several parameters when the number of observations is large, *Trans. Amer. Math. Soc.,* **54** (1943), 426–482.

{67} A. Wald, Sequential tests of statistical hypothesis, *Ann. Math. Statist.,* **16** (1945), 117–186.

{68} A. Wald, Some improvements in setting limits for the expected number of observations required by a sequential probability ratio test, *Ann. Math. Statist.,* **17** (1946), 466–474.

{69} A. Wald, Differentiation under the expectation sign in the fundamental identity of sequential analysis, *Ann. Math. Statist.,* **17** (1946), 493–497.

{70} A. Wald, Asymptotic properties of the maximum likelihood estimate of an unknown parameter of a discrete stochastic process, *Ann. Math. Statist.,* **19** (1948), 40–46.

{71} A. Wald, Estimation of a parameter when the number of unknown parameters increases indefinitely with the number of observations, *Ann. Math. Statist.,* **19** (1948), 220–227.

{72} A. Wald, Asymptotic minimax solutions of sequential point estimation problems, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability,* University of California Press (1951), 1–11.

{73} A. Wald, *Selected Papers in Statistics and Probability by Abraham Wald,* McGraw-Hill Bokk Company, Inc. (New York–Toronto–London, 1955).

{74} A. Wald and J. Wolfowitz, On a test whether two samples are from the same population, *Ann. Math. Statist.,* **11** (1940), 147–162.

{75} A. Wald and J. Wolfowitz, An exact test for randomness in the nonparametric case based on serial correlation, *Ann. Math. Statist.,* **14** (1943), 378–388.

{76} A. Wald and J. Wolfowitz, Statistical tests based on permutations of the observations, *Ann. Math. Statist.,* **15** (1944), 358–372.

Endre Csáki

Alfréd Rényi Institute of Mathematics
Hungarian Academy of Sciences
P.O.B. 127
1364 Budapest
Hungary

csaki@renyi.hu