Joint Third Congress of Czechoslovak Mathematicians and Seventh Congress of Polish Mathematicians

Aktuárské vědy, Vol. 8 (1948), No. 4, 152–162

Persistent URL: http://dml.cz/dmlcz/144735

JOINT THIRD CONGRESS OF CZECHOSLOVAK MATHEMATICIANS AND SEVENTH CONGRESS OF POLISH MATHEMATICIANS.

From 28th August till 4th September, 1949, the Mathematical Institute of the Czech Academy of Science and the Arts, the Union of Czechoslovak Mathematicians and Physicists and the Polish Mathematical Association organised a joint congress of Czechoslovak and Polish mathematicians at which were also present delegates from Hungary and France. There was a total of 179 participants. Fifteen main lectures were held and 109 communications were submitted in the five sections: 1. mathematical logic and the theory of sets, 2. algebra and the theory of numbers, 3. analysis, 4. geometry and topology, 5. probability and applied mathematics.

In our journal we shall confine our reports to the lectures and communications of the fifth section. The main lectures in this field were given by:

G. Alexits: The basis of a theory of the development of monopoly capitalism.
B. Hostinský: Modern work on Markoff chains and related problems.
J. Janko: Advances in the theory of non-parametric tests in statistical inference.
H. Steinhaus: Various forms of the law of large numbers.

Communications in the fifth section, as far as they concerned the theory of probability and statistics, were presented by:

H. Gruźewska: An approximation of the limit probability law.
J. Hájek: The cluster sampling by the two-stage method.
Z. Horák: The frequency law of errors in measurement.
A. Huťa: Concerning the function $T_k = \sum_{s=0}^{k-1} (-1)^s \binom{m}{s} (n - sq)^l$.
N. Krysicki: A limit theorem concerning expressions of higher order in Bayes' problem.
E. Marczewski: Note on the ergodic theorem and on the law of large numbers.
B. Pardubský: The estimation of structural parameters of partially consistent series.
J. Perkal: Concerning certain regional correlations.
H. Steinhaus: On the conception of the length of empirical curves and their measurement.
A. Špaček: A binomial test for which the maximum of a given risk function is a minimum.
M. Warmus: Estimation of the areas of plane regions by means of a plane network.

Now follow short summaries of the lectures and communications as far as they reached us in time for inclusion in this number.

B. Hostinský: Modern work on Markoff chains and related problems.

The lecture is connected with the articles which the author published in 1934 in the „Časopis pro p stování matematiky a fysiky" (Journal for the Advancement of Mathematics and Physics) 63, 167; 64. It surveys work on the subject of Markoff chains and related problems over the period of the last fifteen years. The theorems on the asymptotic values of probabilities for Markoff chains (ergodic principle) have been worked out in detail and extended to include more general assumptions. The relation between the chains and general and partial differential equations has been clarified from various points of view. Functional equations which satisfy the probability of transition in the case of continuous variables have been solved in various ways for very general assumptions. General „stochastic processes" have also been taken into consideration where it is a question of the probability of transition in the time history of a certain population; if it is assumed that the whole „prehistory" of a certain quantity varying with time has an effect on the probability of its further changes, then such processes are really Markoff chains of an infinitely high order. All these considerations can be explained as a generalisation of the theory of Brown movement. Justification of the ergodic principle on the basis of the theory of probability, where in fact nothing is known about the nature of the trajectories, has been compared with its derivation on the assumption that the trajectories are precisely defined by the differential equations.

J. Janko: Advances in the theory of non-parametric tests in statistical inference.

The theory of statistical tests has recently concentrated with increased measure on the problems where it is impossible to assume a certain functional form of the frequency distribution of the population and has tried to give a solution which would be valid for all populations with continous cumulative distribution functions. The problems of this kind are called non-parametric.

Similarly as in the parametrical case it is necessary to construct so called „similar" regions and among them to elect a certain region which has been done mostly up to now by means of a statistical coefficient chosen intuitively. In order to obtain the „similar" regions it is possible to use Fisher's method of randomization. A special case of general randomization methods is the method of ranks. One of the fundamental problems in the non-parametric statistical inference is the problem of estimation of cumulative distribution function $F(x)$ about which we only know that it is continous. It is solved by means of a sample cumulative distribution function $F_N(x)$ so that there are constructed the acceptance regions in which $F(x)$ lies with the certain probability α. In fact it is possible to construct practically the respective confidence region, for it is also determined, as has been shown by Wolfowitz through the belt of such two step-functions $F_N^{(1)}(x)$ and $F_N^{(2)}(x)$, that $F_N^{(2)}(x) \leqq F_N(x) \leqq F_N^{(1)}(x)$. It is useful to choose a belt of a constant width and equal on both sides, so that then $F_N^{(1)}(x) = F_N(x) + \Delta$ and $F_N^{(2)}(x) = F_N(x) - \Delta$ with the exception of a necessary correction consisting in it that the boundaries do not exceed the figure one and do not fall below nought. If we then have a table of values Δ as function of α, we can easily construct the confidence regions. Such a table has been constructed by Kolmogorov and was extended by Smirnov. How to determine the exact confidence belts with a small sample size has been shown by Wald and Wolfowitz who determined a method about finding a probability that $F_N(x)$ would be within an acceptance region which is usable for generally given widths of the belt and for finite sample sizes.

The method which has been used in estimating $F(x)$, can also be used in testing the hypotheses of distribution function of the population. If we test whether the two samples with the sizes n_1 and n_2 have come from the same population, we say that it is a problem of two samples. In its solution the development of the distribution theory of runs (Mood) was made valid, with success. This theory is of great importance with the solution of the question whether the sample values $x_1, x_2, \ldots, x_N$ in order sequence as drawn, are random. Important are the tests of independence consisting in ordered statistics which we obtain through arranging the observed sample values in increasing order from the lowest value to the highest one. Of great significance then are the tests using the runs above and below the median or runs from the binomial and multinomial population ((Bortkiewicz, Mises, Wishart and Hirschfeld, Cochran and Mood) and „runs up and down" which are of outstanding practical use especially in the quality control. Different criteria for testing randomness have been suggested by Young, Anderson, Hotelling and Pabst, Pitman. About the two tests i. e. Thompson's and Mathiesen's it has been proved that they are not consistent. Wald and Wolfowitz have defined the consistency so as to be able to be also used in the non-parametric problems. According to them a test is consistent, if the probability of rejecting the null hypothesis, when it is false, approaches one while the sample number increases indefinitely. Furthermore we should quote a Wald-Wolfowitz's U-test which was used with the solution of the two samples' problem. Its frequency distribution has been determined. If the sample sizes n_1 resp. n_2 are not big there is no difficulty in obtaining a table of critical points (Swed and Eisenhart). If the total of the sizes of these two samples $N = n_1 + n_2$ increases indefinitely with constant $\frac{n_1}{n_2} = k$ the U-distribution approaches the normal distribution asymptotically.

It is necessary to attach great importance to the Wald's general formulation of the problem of statistical inference which starts with a general theory of statistical decision functions. He expanded it first for the non-sequential case, where the number of observations, that is the basis of decision, had been determined in advance. In his next work he extended this theory on the sequential case, where the number of observations required for decision had not been determined in advance but had depended on the result of observation. The solution of the problems of statistical decisions has been issued by Wald in the year 1947 under certain restrictive assumptions; some of these restrictive assumptions have been removed in his work of the year 1949. There has also been started the application of this theory on practical problems.

H. Steinhaus: Various forms of the law of large numbers.

Let the function $x(t)$ be defined and measurable in the interval (0,1). We define the distribution function $X(\alpha)$ of the function $x(t)$ by the expression $X(\alpha) = |\underset{t}{\mathfrak{E}}\{x(t) < \alpha\}|$, and the distribution function $H(\alpha, \beta)$ of the bivariate population $x(t)$, $y(t)$ by the expression $H(\alpha, \beta) = |\mathfrak{E}\{x(t) < \alpha, y(t) < \beta\}|$. The independence of the two variables x, y is expressed by the identity $H(\alpha, \beta) = X(\alpha) \, . \, Y(\beta)$. In a similar way we define the independence of systems of two, three or more variables. „En bloc" independence means independence of each system of n variables for all n.

I. Bernoulli's theorem states:

If the functions $x_1, x_2 \ldots, x_n, \ldots$ are independent in pairs, if they have the same distribution function and if each of them possesses first and second moments, then

$$\text{(1)} \qquad \lim \text{ as } s_n(t)/n = \text{const., where } s_n = \sum_{i=1}^{n} x_i.$$

For the proof it is possible to use Banach's theorem that the mean value of normalised and orthogonal functions converges almost everywhere to zero. This proof can be used even without the assumption of the existence of the second moment; in particular it is possible to replace the functions $x_i(t)$ by limited functions plus a remainder, concerning which it can be shown that its mean absolute value has an integral as small as we please. On the other hand, omission of the assumption of the existence of the second moment permits us to weaken the assumption of independence and to replace it by the assumption of zero correlation between the pairs x_i, x_k. It is not known whether the simultaneous weakening of both assumptions still gives relation (1), i. e. the so-called weak law of large numbers.

II. The strong law of large numbers is expressed by the relation

$$\text{(2)} \qquad \lim s_n(t)/n = \text{const. almost everywhere.}$$

Cantelli proved this in 1916; on the assumption of the independence in pairs and of the existence of both moments (and naturally also of a common distribution function), we can here again make use of Banach's theorem (Bull. Ac. Pol. 1919).

Kolmogorov proved this theorem assuming the existence of only the first moment, and of course on the assumption of „en bloc" independence. It is not known whether it is here possible to weaken the „en bloc" independence.

Just as it is possible to define the distribution function of a function, so also we can define the distribution function of a sequence of numbers: it is the frequency of the members of the sequence less than α. In the same way we can define the conception of the independence of two sequences. The law of large numbers can then be formulated thus:

III. If the functions $x_i(t)$ are independent in pairs and if they have the same distribution function $F(\alpha)$, then the sequence $\{x_i(t)\}$ has the distribution function $F(\alpha)$ for almost all t. The proof consists of Cantelli's law; the existence of moments is not assumed.

From this law follows the explanation of the Petrograd paradox: if a player wins 2^{n-1} money units provided that the coin does not fall head up untill the n-th throw, then before each throw he ought to wager the sums:

$$1, 2, 1, 4, 1, 2, 1, 8, 1, 2, 1, 4, 1, 2, 1, 16, \ldots$$

which have the same distribution function as the variable $x(t)$; hence, according to law III, the game can be regarded as just. On the assumptions made in law III, the sequences $\{x_i(s)\}$ and $\{x_i(t)\}$ are independent of one another for almost all points (s, t) of the unit square.

IV. From the ergodic theorem, E. Marczewski obtained the following form of the law of large numbers: If the functions $x_i(t)$ are equivalent, i. e. if the distribution function of each system of n variables is equal to the distribution function of the n-variable system obtained by altering the subscripts by one, then

(3) $\lim s_n(t)/n$ exists almost everywhere.

This form of the law of large numbers was expressed by Khintchin already in the year 1937 and later by Doob. It should be noted that this paper does not deal with the authorship of the individual theorems. This would indeed be a difficult task due to the incompleteness of bibliographical reports on work in the theory of probability.

The sequence $\{a_n\}$ is said to be random if its distribution function assumes at least three different values and if the sequence is independent of the alteration $\{a_{n+k}\}$. If $x_i(t)$ are independent in sets of four and if they have a common non-trivial distribution function, then the sequence $\{x_i(t)\}$ is random for almost all t.

Halina Milicer-Grużewska: An approximation of the limit probability law.

The author examines the conditions under which the asymptotic expansion theorem is fulfilled[1]) if the random variables are equivalent.[2]) It is proved among other things that the central limit theorem is false, when the equivalent variables are strongly correlated and uniformly bounded, although, as we know,[3]) this theorem is true when the moments of these variables fulfil some special conditions (1). Even if we suppose that the conditions (1) are satisfied, the asymptotic expansion theorem may be false if further moments of the variables do not fulfil some additional con-

---

[1]) H. Cramer: Random Variables and Probability Distributions (Cambridge 1937, ch. VII Th. 2, 5. p. 81—83).

[2]) B. N. Finetti: Classi di numeri aleatori equivalenti (Rendiconti della Reale Accademia Naz. d. Lincei, ser. 6a vl. XVIII, pp. 203-7, Roma).

[3]) H. Milicer-Grużewska: Sulla legge limite d. variabili casuali equivalenti (Atti d. Acc. Naz. d. Lincei s. VIII, v. II, ser. I, fase 2, Teorema I, p. 28-9).

ditions (2). However if we suppose that the conditions (1) and (2) are satisfied, we receive the following approximation for the distribution function $F_n(x)$ of the standardised sum of n equivalent variables:

$$F_n(x) = \varphi(x) + \mathfrak{O}(n^{-\alpha} \lg n^{-1}) \quad 0 < \alpha < 1$$

where $\varphi(x)$ is the Gauss-Laplace integral. The properties of the distribution function $F_n(x)$ and the corresponding characteristic function $\varphi_n(t)$ are subject to the properties of the joint distribution of the equivalent variables: $X_1, X_2, \ldots, X_n : F_n(x_1, x_2, \ldots, x_n)$ which the author describes. The conditions (1) and (2) differ essentially from the analogous ones for the case of independent random variables. Not only the existence of the moments and the estimation of their gradual increase is supposed but also the fulfilment of an enumerable number of equalities.

A more detailed report will appear soon in the Proceedings of the Third Class of the Warsaw Science Society under the title „On the distribution law and the characteristic function of the standardised sum of equivalent variables".

Jaroslav Hájek: The cluster sampling by the two-stage method.

During applications of the sampling method, it is often impossible to select single elements, but only whole groups. J. Neyman presents a solution in which each combination of groups has the same chance of being selected. Then, however, the sample mean

$$\bar{y} = \frac{r_1 y_1 + \ldots + r_k y_k}{r_1 + \ldots + r_k}$$

(r_i and y_i denote the size resp. mean of the selected groups) cannot be used as an estimate of the population mean since it has a bias, and it becomes necessary to choose some other quantity which has usually, however, a larger dispersion. In order that the bias of the sample mean should vanish, it suffices to alter the sampling procedure: we choose a method which the author has called the two-stage method. The first stage consists of selecting one of the groups in such a way that the probability of selection will be, for each group, proportional to its size. (This can be easily arranged with the help of tables of random numbers.) In the second stage, we select the $K - 1$ remaining groups in the usual way, where each combination of groups has the same chance.

Z. Horák: The frequency law of errors in measurement.

With a view to verifying the normal (Gauss) law of errors, the distribution of errors was determined for two series of bridge measurements, each of 500 readings. The obtained frequency curves displayed systematic deviations from the Gauss curve and they can be well represented by the equation

$$\eta = \frac{a}{\pi \cosh(a\varepsilon)} \quad \text{or} \quad \eta = \frac{a}{2\cosh^2(a\varepsilon)} \qquad (1), (2)$$

The author was thus led to study a new type of frequency curve,

$$\eta(\varepsilon) = \frac{a}{\cosh^n(b\varepsilon)}$$

which for whole number $n > 0$ can be easily integrated.

From a theoretical point of view, however, it would be more convenient to keep to the Gauss law of frequency for the simplest cases and to explain deviations of the observed distribution from the normal distribution by the assumption that, when dealing with physical measurements, several simultaneous effects are superimposed with different weights for individual measurings. The resultant frequency curve is then given by the superposition of two (or more) simple Gauss curves, even in cases of symmetrical distribution of errors. The author thus arrived at frequency curves of the type

$$\eta = a\eta_1 + (1 - a)\,\eta_2 \tag{3}$$

where η_1 and η_2 represent Gauss curves with different standard deviations, and by a superposition of an infinite number of Gauss curves, whose standard errors decrease in proportion to the square roots of whole numbers, he obtained a frequency curve of the type

$$\eta(\varepsilon) = \frac{a}{e^{h^2\varepsilon^2} - b} \tag{4}$$

which differs from the Gauss curve in the same way as the basic equation of quantum statistics differs from the Maxwell-Boltzmann equations of classical statistics. Both functions (3) and (4) contain more than one constant and can therefore better represent an actual distribution of errors. The author has derived simple relations for determining the constants of the functions (1), (3) and (4).

Anton Huťa: Concerning the function

$$T_k = \sum_{s=0}^{k-1} (-1)^s \binom{m}{s} (n - sq)^l \tag{1}$$

It is a well-known fact that in practical work and in the theory of applied mathematics, especially in mathematical statistics, there very often occur expressions of the form

$$P = \frac{1}{l!}\sum_{i=0}^{l} (-1)^i \binom{l}{i} \left(\frac{n}{q} - i\right)^l \quad \text{(de Moivre's problem)}$$

$$\nu(1-g)^{\nu-1} - \binom{\nu}{2}(1-2g)^{\nu-1} + \ldots + (-1)^{m-1}\binom{\nu}{m}(1-mg)^{\nu-1} \quad \text{(„Fisher's test")}$$

$$\frac{n^n}{(n-1)!}\sum_{r=0}^{K} (-1)^r \binom{n}{r}\left(m - \frac{r}{n}\right)^{n-1} \quad \text{(distribution of sample means } m \text{ for a rectangular distribution)}$$

All these expressions are special cases of the relation T_K. We wish to investigate the properties of this function. If we introduce into Equ. (1) the substitution $n = (k - 1)q + r$, we can write (1) in the form

$$T_k = T(k, m, q, r, l) = \sum_{s=0}^{k-1} (-1)^s \binom{m}{s} [(k - 1 - s)\,q + r]^l \tag{2}$$

The values of T_k are the coefficients of the members of the series of a certain generating function $f(x)$ which, after several transformations and rearrangements, can be written

$$f(x) = (1 - x)^m\,\varphi_l(x) = (1 - x)^{m-l-1} \sum_{k=0}^{l+1} x^k \sum_{s=0}^{k-1} (-1)^s \binom{l+1}{s} [(k - 1 - s)\,q + r]^l \tag{3}$$

The function $f(x)$ can be expanded into the form of a Laurent series whose coefficient

T_e of x^e will be given by the expression

$$T_e = \frac{1}{2\pi i} \int\limits_{(0)} \frac{1}{x^{e+1}} (1-x)^{m-l-1} \sum_{k=1}^{l} x^l \sum_{s=0}^{k-1} (-1)^s \binom{l}{s} [(k-1-s)\, q + r]^{l-1} \quad (4)$$

From Equ. (3) it follows

$$\varphi_e(x) = \frac{P_{e+1}(x)}{(1-x)^{l+1}} \quad (5)$$

where P_{l+1} is a polynomial of degree $(l+1)$ in x. The polynomials P display a certain symmetry which can be demonstrated by means of the expressions

$$P_{2r}(x) = \sum_{t=1}^{r} g_t(u, r)\, x^{2r-t+1} + g_t(r, u)\, x^t \quad (6)$$

$$P_{2r+1}(x) = g_{r+1}(u, r)\, x^{r+1} + \sum_{t=1}^{r} g_t(u, r)\, x^{2r-t+2} + g_t(r, u)\, x^t. \quad (7)$$

For special cases of these expressions we obtain the individual properties of the expressions occurring in practical work. As an example, let us consider the function

$$T_n^{(l)} = \sum_{t=0}^{N} (-1)^t \binom{n}{t} (n-2t)^l \qquad \frac{n}{2} - 1 \leqq N < \frac{n}{2} \quad (8)$$

From Equ. (4) we can obtain the values of this function. If n and l are simultaneously odd or even, then $T_n^{(l)} = 0$; for special cases, we obtain from (6) and (7) reciprocal polynomials, etc.

E. Marczewski: Note on the ergodic theorem and on the law of large numbers.

The author discusses the analogy between Birkoff's ergodic theorem and the strong law of large numbers. Using this analogy he proves, with the help of the ergodic theorem, a certain, hitherto unknown, version of the law of large numbers which does not require the assumption of independence.

Bohumil Pardubský: The estimation of structural parameters of partially consistent series.

We start out from the assumption that we are given a certain random sample from a given basic population, possessing a cumulative distribution function whose mathematical form is known, but which contains a certain number of parameters which are unknown. The problem is to estimate these parameters as functions of the values of the random sample. The number of such functions is infinitely large. The question arises which of these estimates is the best.

Let $T_i(x_1, \ldots, x_n)$ for $i = 1, 2, \ldots, \nu$ be a system of estimates of the parameters Θ_i $(i = 1, 2, \ldots, \nu)$. We shall consider the best system of estimates to be the one whose simultaneous distribution of probability density of the estimates has maximum concentration. On the assumption that certain conditions of regularity are fulfilled, Harold Cramér established in the Skandinavisk aktuarietidskrift in 1946 the equation of the ellipsoid of maximum concentration belonging to a given simultaneous distribution of probability density of estimates, in the form

$$\sum_{i,j}^{r} \varkappa_{ij}(T_i - \Theta_i)(T_j - \Theta_j) = r + 2$$

where

$$\varkappa_{ij} = E\left[\frac{\partial \log f}{\partial \Theta_i} \frac{\partial \log f}{\partial \Theta_j}\right], \quad f = f(x_1, \ldots, x_n; \Theta_1, \ldots, \Theta_n)$$

is the simultaneous distribution of probability density of the variables $x_1, \ldots, x_n$. The random variables of a sequence $\{\xi_i\}$ are said to be consistent if the number of parameters involved in their distribution functions of probability density is finite, and if each parameter occurs in all or an infinite number of distributions of probability density. The random variables of a sequence $\{\xi_i\}$ are said to be partially consistent if the parameters involved in their distributions of probability density can be divided into two groups. The first group consists of a finite number of parameters and these parameters are contained in all distributions of probability density. The second group of parameters is composed of an infinite number of parameters of which each occurs in only a finite number of distributions of probability density. The parameters of the first group are called structural parameters; the parameters of the second group are called incidental parameters.

Because estimates of structural parameters of partially consistent sequences determined by the method of maximum likelihood are not necessarily consistent, new systematic methods were sought which would give consistent estimates. J. Neyman and Elizabeth Scott established five conditions which the system of equations

$$F_{si}(x_1, \ldots, x_s, T_1, \ldots, T_r) = 0, \quad i = 1, 2, \ldots, r$$

must satisfy i. e. the system of estimates of the structural parameters Θ_i $(i = 1, 2, \ldots, r)$.

Because of the difficulty of fulfilling these conditions it was decided to abandon this general method and to look for a method for a more special, but frequently occurring, case. The method may be called the method of modified equations of maximum likelihood, where we need not limit ourselves to only one incidental parameter, always occurring in the function of one random variable, but where there may be any number. A necessary condition is that the mathematical chance of the partial derivative of the logarithm of the appropriate distribution of probability density, after elimination of the incidental parameters, should be a constant, or should be a function only of the structural parameters.

The squares of the standard deviations of the estimates of individual parameters have lower limits given by the coefficients of the quadratic form which is the inverse to the quadratic form of the corresponding ellipsoid of maximum concentration.

J. Perkal: Concerning certain regional correlations.

In this paper the author deals with the problem: Suppose that the result of a series of observations is n numbers. The results of N such series of observations can be arranged in a matrix with N rows and n columns. If we interpret each row of this matrix as a point in an n-dimensional space, we obtain a system of N points in this space.

In mathematical statistics we are familiar with the conception of the ellipsoid of concentration, belonging to a given distribution system of probability density. Explaining the relations involved in this conception, the author demonstrates methods of working out results from observations of the above-mentioned type.

As an example he presents new tables of the dependence between the weight and the height (with regard to age) of school children in Lower Silesia.

H. Steinhaus: On the length of empirical curves.

A paradox in the conception of length arises from the fact that length is an unbounded functional in the neighbourhood of every arc. If we wish to eliminate this paradox, we can use the following method of measuring length: On transparent paper we have drawn a system of equidistant, parallel lines, distance d apart. We place this paper over the arc S and count the number of intersections of the system with the curve S, then we turn the paper through an angle of $1/k$ of a straight angle and again count the number of intersections; in this way we obtain the numbers $n_1, n_2, \ldots, n_k$ and $N = \sum_{i=1}^{k} n_i$; the approximate length of the arc S is then

$$L = Nd \frac{\pi}{2k} \tag{1}$$

and the accuracy depends on d and k. Equation (1) alone does not eliminate the paradox of length, but can be modified by taking, in place of the numbers $n_1, n_2, \ldots, n_k$, the numbers $n_1', n_2', \ldots, n_k'$, defined by the equations

$$n_i' = n_i \tag{2}$$

for $n_i \leq p$, $n_i' = p$ for $n_i > p$.

Using (1) we obtain a length of p-th order, the functional of which we denote by $L_p(S)$. This functional is bounded. We can calculate the value $L_p(S)$ with any chosen degree of accuracy. We can compare the lengths of the boundaries of two regions using any chosen order p, e. g. $p = 10$. If we are given two maps of different accuracy and if we wish to use the information contained in them, e. g. compare the lengths of two rivers, one of which appears on the one map and the other river on the other map, we must adjust the order p to the map of rougher accuracy. Even if we are investigating objects which in fact have infinite length, we can define the relation of their lengths s by the expression

$$s = \lim_{p=\infty} L_p(A)/L_p(B)$$

where A, B are these objects, e. g. the left and the right bank of the River Vltava. In cases where we do not wish to use this method, it is better to avoid the conception of length.

Antonín Špaček: Sampling plans for percent defective, which minimize the maximum of a given risk function. (Tesla Electronic National Corporation, Prague.)

Given a lot of size N of manufactured product which contains an *unknown* fraction defective p. The acceptance or rejection of this lot is to be decided on the basis of a single random sample of size n with acceptance number k. The decision procedure is determined as follows: *If the random sample of size n from this lot contains k or less defective items, the lot will be accepted, and otherwise rejected.* Evidently, according to this rule, the two integers $n \leq N$ and $k \leq n$ completely determine the decision procedure and with each N there is associated the set S_N of all decision procedures of this type, i. e. the set of all pairs of integers (n, k), where $0 \leq k \leq n \leq N$. The choice of an element of S_N corresponds to the choice of a decision procedure. The probability of acceptance is

$$P(p, n, k) = \frac{1}{\binom{N}{n}} \sum_{j=0}^{k} \binom{Np}{j} \binom{N(1-p)}{n-j} \tag{1}$$

If n/N is sufficiently small (for example $\leqq 1/10$), then (1) may be approximated by

$$L(p, n, k) = \sum_{j=0}^{k} \binom{n}{j} p^j (1-p)^{n-j} \tag{2}$$

With each $(n, k) \in S_N$ and each fraction defective p there is associated a risk $R(p, n, k)$, which expresses the economic outlay involved by the given decision procedure. If rejected lots are returned by the consumer to the producer, it seems natural to establish the risk-function as follows: $cn + a(N-n)\, pL(p, n, k)$, where c is the cost of inspection of one item and a is the loss incurred by accepting one defective. Since cN is the cost of making 100% inspection, we may write

$$R(p, n, k) = \frac{n}{N} + \alpha \left(1 - \frac{n}{N}\right) pL(p, n, k) \tag{3}$$

where $\alpha = a/c$.

A question of fundamental importance is that of the proper choice of sampling plans. Following an idea of A. Wald we shall now establish a principle, on the basis of which the sampling plan will be chosen.

The decision procedure $(\nu, \varkappa) \in S_N$ is optimum if and only if simultaneously

$$L(p_0, \nu, \varkappa) \geqq 1 - \varepsilon, \quad 0 < \varepsilon < 1, \tag{4}$$

the maximum of (3) with respect to p is minimum for $n = \nu$, $k = \varkappa$. (5)

The condition (4) excludes the unreasonable consumer's specifications, which cannot be met by the producer, i. e. there is given a $p = p_0$ and a value $0 < \varepsilon < 1$, such that (4) holds. It is easy to show, that $\min_n \max_p R(p, n, k)$ always exists and that

$$\min_n \max_p R(p, n, k) = \max_p \min_n R(p, n, k) \tag{6}$$

for each $k = 0, 1, 2, \ldots$ Clearly, (6) corresponds to the condition of strict determinateness in the sense of von Neumann's theory of zero-sum two-person games.

For each $k = 0, 1, 2, \ldots$ let n_k denote the value of n which minimizes the maximum with respect to p of (3). We have $n_0 \sim \sqrt{(\alpha N)]e}$ and n_k for $k = 1, 2, 3, \ldots$ may be easily computed by successive approximations. Since $R(p, n, k) \leqq R(p, n, k+1)$ for each p, n, k we have $\min_n \max_p R(p, n, k) \leqq \min_n \max_p R(p, n, k+1)$ for each $k = 0, 1, 2, \ldots$ and the optimum decision procedure $(\nu, \varkappa)$ can be found as follows:

For $k = 0, 1, 2, \ldots$ *we compute $n_0, n_1, n_2, \ldots$ as long as*

$$L(p_0, n_k, k) < 1 - \varepsilon.$$

Then

$$L(p_0, n_{k+1}, k+1) \geqq 1 - \varepsilon$$

and $(\nu, \varkappa) \in S_N$, where $\nu = n_{k+1}$, $\varkappa = k + 1$ is the optimum decision procedure. If ν/N is not sufficiently small, then (2) must be replaced by (1) and the computations are obviously much more laborious.

M. Warmus: Estimation of the areas of plane regions by means of a parallelogram network.

In a plane parallelogram network W we are given a q-connected region A of area a and diameter b, within which lie w nodes of the network W. Region A is bounded by q closed Jordan curves which are capable of rectification.

Let l_i denote the length and b_i the diameter of the curve L_i. Let $\Sigma l_i = 1$.

Let m denote the distance between the two nearest nodes of the network W, α the area of one panel of this network. The number φ is given by the equation

$$m^2(\varphi + \sin\varphi\cos\varphi) - 2\alpha\sin^2\varphi = 0 \quad 0 < \varphi < \tfrac{1}{2}$$

and the number c by the equation $c = \dfrac{1}{\sqrt{2\varphi + \sin 2\varphi}}$.

Theorem: If $l \geqq l^* = 2\left(\pi c - \sqrt{\pi^2c^2 - \pi}\right).\sqrt{\alpha}$ or $b \geqq m$ and if for q_0 various natural numbers k_j ($j = 1, 2, \dots, q_0$), where $1 \leqq k_j \leqq q$, the relation $l_{k_j} \geqq l^*$ or $b_{k_j} \geqq m$ holds, then the inequality

$$-cl\sqrt{\alpha} + (q_0 - 2)\,\alpha < a - w\alpha \leqq cl\sqrt{\alpha} + (q - 2)\,\alpha$$

holds true.

ZPRÁVY

POLSKO-ČESKOSLOVENSKÁ PRACOVNÍ KONFERENCE MATEMATICKÝCH STATISTIKŮ VE VRATISLAVĚ se bude konat ve dnech od 13. do 18. června 1950. Předmětem konference bude matematická statistika a její užití v průmyslu a v problémech hospodářsko-sociálních.

Práce bude rozdělena do tří sekcí:

A. Užití matematické statistiky ve výrobě a obchodu.

B. Užití matematické statistiky v problémech demologických, zvláště užití methody výběrové (representativní) v problémech úmrtnosti a nemocnosti.

C. Prostředky numerického počítání, hlavně matematické stroje elektronkové a jejich užití v badáních statistických.

V těchto třech sekcích se budou projednávati jednak sdělení s časovým vymezením 15minutovým, jednak referáty, stanovené na 30 minut. Jednacími jazyky bude čeština, slovenština, polština a jazyky, jichž se užívá na mezinárodních sjezdech matematických.

Přihlášené referáty a sdělení československých účastníků budou pojednávat o problémech péče o jakost s hlediska matematické statistiky, o rozhodovacích funkcích, pro které maximum jistých risikových funkcí je minimální, o jejich aplikaci na výběrové postupy a kontrolní diagramy, o statistickém sledování a kontrole plánovaných jevů ve všech úsecích socialistického hospodářství, o prvcích kontrolních diagramů s hlediska hospodárnosti. V dalších pracích se pojednává o užití výběrových method při zpracování sčítání lidu, o theorii kontroly přejímací i plynulého procesu, o otázkách statistiky v národních podnicích obchodních a jejím vztahu k plánování a podnikovému početnictví, o souvislosti řetězových testů výběrových se stochastickými procesy a o užití nomografických pomůcek při numerickém počítání.