icm©

COLLOQUIUM MATHEMATICUM

VOL. VII 1960 FASC. 2

SOME NON-PARAMETRIC TESTS FOR THE k-SAMPLE PROBLEM

BY

M. FISZ (WARSAW)

1. Summary. This note is a summary of some methods that have been proposed for testing the hypothesis that k $(k > 2)$ independent samples have been drawn from populations with the same (unspecified) continuous distribution function. The methods discussed are generalizations to $k > 2$ of Smirnov's [1], [2] tests for $k = 2$, which have been proposed by Ozols [3], Fisz [4], Chang and Fisz [5], [6], Kiefer [7], [8], Gichman [9] and David [10].

2. Formulation of the problem. Let $(x_{11}, \ldots, x_{1n_1}), \ldots, (x_{k1}, \ldots, x_{kn_k})$ be k independent samples drawn from populations having the same continuous distribution function $F(x)$. Denote by $S_{jn_j}(x)$ $(j = 1, \ldots, k)$ the empirical distribution function of the j-th sample, i. e. $n_j S_{jn_j}(x)$ represents the number of those observations of the j-th sample which are smaller than x. Kolmogorov [11] has shown that for $k = 1$ (omitting here the subscript j) the relation

$$\lim_{n\to\infty} P\big(\sqrt{n}\sup_x |S_n(x) - F(x)| < \lambda\big) = K(\lambda) = \sum_{s=-\infty}^{\infty} (-1)^s \exp(-2\lambda^2 s^2) \tag{1}$$

holds for arbitrary $\lambda > 0$. Smirnov [1], [2] has shown for $k = 2$ that if $n_2/n_1 = a > 0$, the following relations hold for arbitrary $\lambda > 0$:

$$\lim_{n_1\to\infty} P\left(\sqrt{\frac{n_1 n_2}{n_1+n_2}} \max_x [S_{1n_1}(x) - S_{2n_2}(x)] < \lambda\right) = 1 - \exp(-2\lambda^2), \tag{2}$$

$$\lim_{n_1\to\infty} P\left(\sqrt{\frac{n_1 n_2}{n_1+n_2}} \max_x |S_{1n_1}(x) - S_{2n_2}(x)| < \lambda\right) = K(\lambda). \tag{3}$$

Smirnov's formulae are used as a basis for general tests of the hypothesis that 2 independent samples have been drawn from populations with the same unspecified continuous distribution function. Now a similar problem arises for $k > 2$ samples.

It is worthwhile to note that the construction of tests of practical importance which generalize Smirnov's tests to arbitrary $k > 2$ (methods 2 and 3 below) have been proposed very recently although Smirnov's tests were constructed about 20 years ago. It is an elegant idea of Doob [12] for proving the Kolmogorov-Smirnov limit theorems that stimulated investigations in the direction considered.

3. Method 1. A natural generalization of Smirnov's 2-sample procedure would be to consider tests based on the expressions:

$$(*)\qquad \max_{\substack{i,j\\ i\neq j}} \sqrt{\frac{n_i n_j}{n_i+n_j}} \max_x [S_{in_i}(x)-S_{jn_j}(x)],$$

$$(**)\qquad \max_{\substack{i,j\\ i\neq j}} \sqrt{\frac{n_i n_j}{n_i+n_j}} \max_x |S_{in_i}(x)-S_{jn_j}(x)|.$$

In a recent paper, David [10] obtains the exact and asymptotic distribution of $(*)$ for $k=3$ and $n_1=n_2=n_3=n$. We present David's theorem for the asymptotic case.

THEOREM 1. *For $\lambda\sqrt{n}$ integral the relation*

$$(4)\qquad \lim_{n\to\infty} P\Big(\sqrt{n}\max\big[\max_x\big(S_{2n}(x)-S_{1n}(x)\big), \max_x\big(S_{3n}(x)-S_{2n}(x)\big), \max\big(S_{1n}(x)-S_{3n}(x)\big)\big]\geqslant\lambda\Big)$$

$$= 3\sum_{i=1}^{\infty}\sum_{j\in J(i)}(\pm)\exp[-\lambda^2(i^2+j^2-ij)]$$

holds, where $J(i)$ consists of the integers $(2-i, 3-i, 5-i, 6-i, 8-i, 9-i, \ldots, 2i)$ and where $(\pm)$-sign indicates that for fixed i successive terms in the finite series indexed by j have alternating signs beginning with $+$ for $j=2-i$, $-$ for $j=3-i$, $+$ for $j=5-i$, and so on.

A result for a somewhat related statistic has been obtained by Ozols [3], namely

THEOREM 2. *For arbitrary positive λ_1, λ_2 the relation*

$$(4')\qquad \lim_{n=\infty} P\left(\sqrt{\frac{n}{2}}\max_x[S_{1n}(x)-S_{2n}(x)]<\lambda_1, \sqrt{\frac{n}{2}}\max_x[S_{2n}(x)-S_{3n}(x)]<\lambda_2\right)$$

$$=1-\exp(-2\lambda_1^2)-\exp(-2\lambda_2^2)+ +2\exp\big(-[\lambda_1^2+\lambda_2^2+(\lambda_1+\lambda_2)^2]\big)-\exp\big(-2(\lambda_1+\lambda_2)^2\big)$$

holds.

In particular, for $\lambda_1=\lambda_2=\lambda$ the right-hand side of (4′) becomes

$$1-\exp(-2\lambda^2)+2\exp(-6\lambda^2)-\exp(-8\lambda^2).$$

Ozols has also found the exact distribution of the left side of (4′) with $n_1\neq n_2\neq n_3$.

The idea of the proofs of the theorems of David and Ozols consists in a straightforward generalization of the proof for the case $k=2$ given by Gnedenko and Koroluk [13]. In general, however, it seems impossible to treat the distribution for arbitrary k by this method.

4. Method 2. Write $N=(n_1, n_2, \ldots, n_k)$ and

$$(5)\qquad S_{N0}(x)=\frac{\sum_{j=1}^{k} n_j S_{jn_j}(x)}{n_1+n_2+\ldots+n_k},$$

$$(6)\qquad D_{Nk}^2=\max_x \sum_{j=1}^{k} n_j[S_{jn_j}(x)-S_{N0}(x)]^2.$$

Method 2 is based on the following theorem of Gichman [9] and Kiefer [7], [8].

THEOREM 3. *Let $S_{jn_j}(x)$ $(j=1,2,\ldots,k)$ be k empirical distribution functions of k independent samples drawn from populations having the same continuous distribution function, and let D_{Nk} be defined by (6). Then for arbitrary $\lambda>0$ the relation*

$$(7)\qquad \lim_{N\to\infty} P(D_{Nk}<\lambda)=\frac{4}{\Gamma\left(\frac{k-1}{2}\right)(2\lambda^2)^{(k-1)/2}}\sum_{s=1}^{\infty}\frac{p_s^{k-3}}{[J'_{(k-3)/2}(p_s)]^2}\exp\left(-\frac{p_s^2}{2\lambda^2}\right)$$

holds, where $N\to\infty$ denotes $n_1\to\infty, \ldots, n_k\to\infty$ and where p_s is the s-th positive root of the Bessel function $J_{(k-3)/2}(z)$.

The idea of the proof of theorem 3 is the following: write for $j=1,\ldots,k$

$$\xi_{Nj}(x)=\sqrt{n_j}\,[S_{jn_j}(x)-S_{N0}(x)].$$

Consider the vector-process $\big(\xi_{N1}(x), \ldots, \xi_{Nk}(x)\big)$. The processes $\xi_{Nj}(x)$ are linearly dependent since they satisfy the linear relation

$$\sum_{j=1}^{k}\sqrt{\frac{n_j}{n_1+\ldots+n_k}}\,\xi_{Nj}(x)=0.$$

Transform the vector process $(\xi_{N1}(x), \ldots, \xi_{Nk}(x))$ by using an arbitrary orthogonal $k \times k$ matrix (γ_{ij}) with

$$\gamma_{kj} = \sqrt{\frac{n_j}{n_j + \ldots + n_k}} \qquad (j = 1, \ldots, k).$$

The vector process $(\zeta_{N1}(x), \ldots, \zeta_{N(k-1)}(x))$ is then obtained, for which

$$D^2_{Nk} = \max_x \sum_{i=1}^{k-1} \zeta^2_{Ni}(x),$$

where $\zeta_{N1}(x), \ldots, \zeta_{N(k-1)}(x)$ are, as $N \to \infty$, asymptotically normal and asymptotically independent, and moreover the relations [1]

$$(8) \qquad \begin{aligned} E\zeta_{Ni}(x) &= 0 \qquad (0 \leqslant x \leqslant 1), \\ E\zeta_{Ni}(x_1)\zeta_{Ni}(x_2) &= x_1(1-x_2) \qquad (0 \leqslant x_1 \leqslant x_2 \leqslant 1) \end{aligned}$$

are satisfied. In virtue of Centsov's [14] theorem the problem is then reduced to that of finding the probability distribution of the maximal length of a vector $(\zeta_1(x), \ldots, \zeta_{k-1}(x))$, where the processes $\zeta_i(x)$ are Gaussian, independent and satisfy relations (8). It is shown then that this probability distribution is given by the right side of (7).

5. Method 3. Method 3 has been formulated in Fisz's paper [4] for $k = 3$. Theorem 4 below is a generalization of this result to arbitrary k (Chang and Fisz [5], Kiefer [8]).

Define for $i = 1, \ldots, k-1$

$$(9) \qquad \eta_{Ni}(x) = \sum_{j=1}^{k} \beta_{Nij}\sqrt{n_j}\, S_{jn_j}(x),$$

$$(9') \qquad A^+_{Ni} = \max_x \eta_{Ni}(x); \qquad A_{Ni} = \max_x |\eta_{Ni}(x)|,$$

where $N = (n_1, \ldots, n_k)$ and β_{Nij} are real constants.

THEOREM 4. *Let $S_{jn_j}(x)$ $(j = 1, \ldots, k)$ be empirical distribution functions of k independent samples drawn from populations having the same continuous distribution function. Assume that*

$$(10) \qquad \sum_{j=1}^{k} \beta_{Nij}\sqrt{n_j} = 0 \qquad (i = 1, \ldots, k-1),$$

$$(11) \qquad \lim_{N\to\infty} \beta_{Nij} = \beta_{ij} \qquad (i = 1, \ldots, k-1;\ j = 1, \ldots, k),$$

(1) We make the unrestrictive assumption that the theoretical distribution considered is uniform in the interval $[0, 1]$.

where

$$\sum_{j=1}^{k} \beta_{hj}\beta_{ij} = \delta_{hi} \ (^2) \qquad (h, i = 1, \ldots, k-1).$$

Then the following relations hold for arbitrary positive $\lambda_1, \ldots, \lambda_{k-1}$:

$$(12) \qquad \lim_{N\to\infty} P(A^+_{Ni} < \lambda_i,\ i = 1, \ldots, k-1) = \prod_{i=1}^{k-1} [1 - \exp(-2\lambda_i^2)],$$

$$(13) \qquad \lim_{N\to\infty} P(A_{Ni} < \lambda_i,\ i = 1, \ldots, k-1) = \prod_{i=1}^{k-1} K(\lambda_i) = \prod_{i=1}^{k-1} \sum_{s=-\infty}^{\infty} (-1)^s \exp(-2\lambda_i^2 s^2).$$

In particular, for arbitrary positive λ

$$(14) \qquad \lim_{N\to\infty} P(\max_{1\leqslant i\leqslant k-1} A^+_{Ni} < \lambda) = [1 - \exp(-2\lambda^2)]^{k-1},$$

$$(15) \qquad \lim_{N\to\infty} P(\max_{1\leqslant i\leqslant k-1} A_{Ni} < \lambda) = [K(\lambda)]^{k-1}.$$

It follows from relations (10) and (11) that the $k(k-1)$ unknown β_{ij}'s must satisfy $2(k-1) + \binom{k-1}{2}$ equations. This permits an arbitrary choice of values for $\binom{k-1}{2}$ β_{ij}'s. A particularly interesting set of β_{Nij}'s arises by assuming that

$$(16) \qquad \lim_{n_1\to\infty} \frac{n_j}{n_1} = a_j > 0 \qquad (j = 1, \ldots, k)$$

and by setting

$$\beta_{i(i+2)} = \ldots = \beta_{ik} = 0 \qquad (i = 1, \ldots, k-1).$$

This choice gives rise to

$$(17) \qquad \beta_{Nij} = \begin{cases} \sqrt{\dfrac{n_j n_{i+1}}{(n_1 + \ldots + n_i)(n_1 + \ldots + n_{i+1})}} & (j = 1, \ldots, i), \\ -\sqrt{\dfrac{n_1 + \ldots + n_i}{n_1 + \ldots + n_{i+1}}} & (j = i+1), \\ 0 & (j = i+2, \ldots, k). \end{cases}$$

(2) δ_{hi} denotes the Kronecker delta.

This system has, as has been shown by Chang and Fisz [6] and Kiefer [8], the remarkable feature that the functionals A^{+}_{Ni} $(i = 1, \ldots, k-1)$ resp. A_{Ni} defined by (9′) are exactly independent.

An alternative system for $k = 3$ (if it is assumed that (16) holds) is given by

$$(18)\qquad \begin{aligned} \beta_{N11} &= \frac{b_2+b_3}{B}, & \beta_{N12} &= \frac{-b_2}{B}\sqrt{\frac{n_1}{n_2}}, & \beta_{N13} &= \frac{-b_3}{B}\sqrt{\frac{n_1}{n_3}}, \\ \beta_{N21} &= \frac{b_2-b_3}{B}, & \beta_{N22} &= \frac{-b_2}{B}\sqrt{\frac{n_1}{n_2}}, & \beta_{N23} &= \frac{b_3}{B}\sqrt{\frac{n_1}{n_3}}, \end{aligned}$$

where

$$b_j = \sqrt{\frac{n_j}{n_1+n_j}} \qquad (j = 2, 3)$$

and $B = \sqrt{2+2b_2b_3}$.

The power of the tests considered with different systems $\{\beta_{Nij}\}$ is of course not known and consequently it is difficult to say which of them is better.

We now present the idea of the proof of theorem 4.

Consider the sequence $\{Q_N\}$ of measures induced by the vector-processes $\{\eta_{N1}(x), \ldots, \eta_{N(k-1)}(x)\}$ in the Cartesian product-space

$$\vartheta = D_1[0,1]\times\ldots\times D_{k-1}[0,1],$$

where $D[0,1]$ is the space of real functions defined on $[0,1]$, having right-hand and left-hand limits at each point and continuous on the left with Prohorov's [15] distance d. Applying some results of Donsker [16] and Prohorov [15], the relation

$$(19)\qquad Q \Rightarrow Q_0$$

is obtained, where Q_0 is the measure induced in ϑ by the vector-process $\big(\eta_1(x), \ldots, \eta_{k-1}(x)\big)$ with $\eta_i(x)$ $(i = 1, \ldots, k-1)$ independent and Gaussian, satisfying the equalities

$$(20)\qquad \begin{aligned} &E\eta_i(x) = 0 \qquad (0 \leqslant x \leqslant 1), \\ &E\eta_i(x_1)\eta_i(x_2) = x_1(1-x_2) \qquad (0 \leqslant x_1 \leqslant x_2 \leqslant 1). \end{aligned}$$

Taking into account (19), the independence of $\eta_i(x)$ and the probability distributions of $\max_x \eta_i(x)$, resp. $\max_x |\eta_i(x)|$ (Doob [12]) the assertion of theorem 4 is obtained.

6. Concluding remarks. Let us first remark that for $k = 2$ the methods 2 and 3 are identical since in this case both coincide with Smirnov's method. On the other hand, relation (7) holds for any D_{Nk} defined by the formula

$$D^2_{Nk} = \max_x \sum_{i=1}^{k-1} \eta^2_{Ni}(x),$$

where $\eta_{Ni}(x)$ $(i = 1, \ldots, k-1)$ are given by (9) and β_{Nij} satisfy the assumptions of theorem 4. The essential difference between methods 2 and 3 is the following: Method 3 recommends the use of the limiting joint distribution of the A^{+}_{Ni} (resp. A_{Ni}) or that of the largest of them as a basis of the tests considered, and all calculations may be carried out by using Smirnov's [1] tables. Method 2 recommends the use of the limiting probability distribution of the maximal length of the vector $\big(\eta_{N1}(x), \ldots, \eta_{N(k-1)}(x)\big)$. Formula (7) has its own merits, but simplicity is the merit of method 3. Nevertheless it is only the knowledge of the power functions that can give a correct answer to the question which of these methods should be used. It is no doubt worthwhile to make considerable efforts in order to find a reasonable general solution to the problem of the power of the tests of Kolmogorov-Smirnov and of tests related to them.

REFERENCES

[1] Н. В. Смирнов, *Оценка расхожденння между эмпирическими кривыми распеделения в двух независимых выдорках*, Бюллетень МГУ 2 (1939), No 2, p. 3-16.

[2] — *Приближение законов распределения случайных величин по эмпирическим данным*, Успехи Математических Наук 10 (1944), No 2, p. 179 - 206.

[3] V. Ozols, *Gnedenko-Koroluka teoremas vispãrinãjums uz tris izlasem pie divãm vienpuzigãn robezãm*, Latvijas PSR Zinatum Akademijas Vestis 10 (1956), p. 141-152.

[4] M. Fisz, *A limit theorem for empirical distribution functions*, Bulletin de l'Académie Polonaise des Sciences, Classe III, 5 (1957), p. 695-698; Studia Mathematica 17 (1958), p. 71-77.

[5] L. C. Chang, M. Fisz, *Asymptotically independent linear functions of empirical distribution functions*, Science Record 1 (1957), p. 335-340.

[6] — *Exact distributions of the maximal values of some functions of empirical distribution functions*, Science Record 1 (1957), p. 341-346.

[7] J. Kiefer, *Limiting distributions of k-sample test criteria of Kolmogorov-Smirnov-Mises type*, Annals of Mathematical Statistics 29 (1958), p. 614.

[8] — *k-sample analogues of the Kolmogorov-Smirnov and Cramér-v. Mises tests*, ibidem 30 (1959), p. 420-447.

[9] I. I. Gichman, *Über ein nichtparametrisches Kriterium der Homogenität der k-Stichproben*, Теория Вероятностей и ее Применения 2 (1957), p. 380-384.

[10] H. T. David, *A three-sample Kolmogorov-Smirnov test*, Annals of Mathematical Statistics 29 (1958), p. 842-851.

[11] A. Kolmogorov, *Sulla determinazione empirica di una legge di distribuzione*, Giornale dell Istituto degli Attuari 4 (1933), p. 83-91.

[12] J. L. Doob, *Heuristic approach to the Kolmogorov-Smirnov theorems*, Annals of Mathematical Statistics 20 (1949), p. 393-403.

[13] Б. В. Гнеденко, В. С. Королюк, *О максимальном расхождении двух эмпирических распределений*, Доклады Академии Наук СССР 80 (1951), p. 525-528.

[14] N. Centsov, *Weak convergence of stochastic processes whose realizations have no discontinuities of second kind*, Теория Вероятностей и ее Применения 1 (1956), p. 155-161.

[15] Yu. V. Prohorov, *Convergence of random processes and limit theorems in probability theory*, Теория Вероятностей и ее Применения 1 (1956), p. 177-238.

[16] M. L. Donsker, *Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems*, Annals of Mathematical Statistics 23 (1952), p. 277-281.

MATHEMATICAL INSTITUTE OF THE POLSIH ACADEMY OF SCIENCES

Reçu par la Rédaction le 17. 3. 1959

COLLOQUIUM MATHEMATICUM

VOL. VII 1960 FASC. 2

ON SOME LOSS FUNCTIONS

BY

S. TRYBUŁA (WROCŁAW)

In this paper we shall deal with some questions concerning the Wald theory of decision functions. For some known distributions depending on a parameter we shall find a loss function such that the minimax estimate of that parameter is unbiased. We shall see that the least favourable prior distribution of the estimated parameter is the uniform one.

1. Definitions. Let $F(x|\omega)$ be a distribution function defined on a Euclidean space $\mathcal{X}$ which depends on a parameter $\omega \in \Omega$. In the sequel we shall assume that ω is a vector. Each estimate of ω is a measurable function $f(x)$ with values belonging to Ω. Let $L[f(x), \omega_0]$ be the loss to the statistician if he applies the estimate $f(x)$ when x is the observed value of X, and ω_0 is the value of the parameter ω. If we establish the function $f(x)$ and the value of ω, then we can find the expected value of the loss L, i. e.

$$(1.1)\qquad R(f, \omega) = \int_{\mathcal{X}} L[f(x), \omega]\,dF(x|\omega) \overset{\mathrm{df}}{=} E\{L[f(X), \omega]|\omega\};$$

here X is a random variable with distribution function $F(x|\omega)$.

The function $R(f, \omega)$ will be called the *risk*.

The estimate f^0 is called *minimax* if

$$(1.2)\qquad \sup_{\omega\in\Omega} R(f^0, \omega) = \inf_f \sup_\omega R(f, \omega).$$

Let the prior distribution of the parameter ω be given by a distribution function $G(\omega)$. The expected risk $r(f, G)$ is

$$(1.3)\qquad r(f, G) = \int_{\Omega} R(f, \omega)\,dG(\omega) \overset{\mathrm{df}}{=} E_G[R(f, \omega)].$$