ON THE THEORY OF ORDER STATISTICS.

ALFRÉD RÉNYI

The author developed a new method, which reduces the problems concerning limiting distributions of order statistics to the study of sums of independent random variables. Many results of the theory of order statistics can be obtained by this method with surprising simplicity. The method is based on the fact, that if $\xi_1^* \leqq \xi_2^* \leqq \ldots \leqq \xi_n^*$ are the order statistics from a sample taken from a population having the continuous cumulative distribution function (c.d.f.) $F(x)$, we have

$$(1) \qquad \xi_k^* = F^{-1}\left(\exp\left(-\sum_{j=1}^{n-k+1} \frac{\delta_j}{n-j+1}\right)\right) \quad (k = 1, 2, \ldots, n)$$

where the random variabes δ_j $(j = 1, 2, \ldots, n)$ are independent and have the same c.d.f. $1 - e^{-x}$ $(x \geqq 0)$ and $x = F^{-1}(y)$ denotes the inverse function of $y = F(x)$.

This fact is equivalent with the statement that the n random variables $\{F(\xi_k^*)/F(\xi_{k+1}^*)\}^k$ are independent and homogeneously distributed between 0 and 1, which is contained in the stencilled lectures (1947—1949) by D. van Dantzig [3], who kindly called the attention of the author to this fact.

One of the new results which have been obtained by the author by means of his method is as follows. Let $F_n(x)$ denote the c.d.f. of a sample of size n taken from a population having the continuous c.d.f. $F(x)$. A well known theorem of A. N. Kolmogorov gives a confidence band for the unknown c.d.f. $F(x)$ which band has the same breadth for all x. From a practical point of view a confidence band the breadth of which is proportional to the value of $F(x)$ has some advantages. Such a confidence band can be (approximately) given by means of the following theorem:
for any a $(0 < a < 1)$ we have for $y > 0$

$$(2) \quad \lim_{n\to\infty} P\left(\sqrt{n} \sup_{a \leqq F(x)} \left|\frac{F_n(x) - F(x)}{F(x)}\right| < y\right) = \frac{4}{\pi} \sum_{k=0}^{\infty} (-1)^k e^{-\frac{(2k+1)^2\pi^2(1-a)}{8ay^2}}.$$

The proof of (2) and of other similar results was published in [1].

In the second part of his lecture the author mentioned a new two sample test, of Wilcoxon's type, which is similar to a test proposed by E. Lehmann [2], but somewhat simpler. Let us denote by $(\xi_1, \xi_2, \ldots, \xi_m)$ a sample taken from a population having the continuous c.d.f. $F(x)$ and $(\eta_1, \eta_2, \ldots, \eta_n)$ a sample taken from an other population having the continuous c.d.f. $G(x)$. Let $\mathbf{W}_1$ denote the number of valid pairs of inequalities $\eta_j < \xi_i$, $\eta_k < \xi_i$ $(i = 1, 2, \ldots, m;$

$j, k = 1, 2, \ldots, n; j \neq k$) and $\mathbf{W}_2$ the number of valid pairs of inequalities $\xi_j < \eta_i$, $\xi_k < \eta_i$ ($i = 1, 2, \ldots, n$; $j, k = 1, 2, \ldots m, j \neq k$). The test proposed by the author is based on the statistic

$$W = \frac{W_1}{m\binom{n}{2}} + \frac{W_2}{n\binom{m}{2}}. \tag{3}$$

This test is consistent against any alternative hypothesis $G(x) \not\equiv F(x)$. This is the consequence of the fact that if $F(x)$ and $G(x)$ are continuous c.d.f. we have

$$\int_{-\infty}^{+\infty} F^2(x)dG(x) + \int_{-\infty}^{+\infty} G^2(x)dF(x) \geqq \tfrac{2}{3} \tag{4}$$

and equality stands in (4) if and only if $F(x) \equiv G(x)$.

Finally the author mentioned some unsolved problems, for instance the following one: which systems of curves on the (x, y)-plane can be a system of level curves $F(x, y) =$ const. of a two dimensional c.d.f. $F(x, y)$?

References

[1] A. RENYI, On the theory of order statistics, Acta Math. Acad. Sci. Hung. 4(1953) pp. 191—231.

[2] E. L. LEHMANN, Ann. Math. Statistics 22 (1951) pp. 165—180.

[3] D. VAN DANTZIG, Kadercursus Mathematische Statistiek (1947—1949), Mathematical Centre, Ch. 6—§1—p. 371.

INSTITUTE FOR APPLIED MATHEMATICS OF
THE HUNGARIAN ACADEMY OF SCIENCES
BUDAPEST VI. SZTÁLIN UT 31.