SOME SMIRNOV TYPE THEOREMS OF PROBABILITY THEORY[1]

BY MIKLÓS CSÖRGŐ

Princeton University and McGill University

1. Introduction. Let $\xi_{11}, \xi_{12}, \cdots, \xi_{1n}$ and $\xi_{21}, \xi_{22}, \cdots, \xi_{2m}$ be two samples of mutually independent random variables having a continuous distribution function $F(t)$. Let $F_{1n}(t)$ and $F_{2m}(t)$ be the corresponding empirical distribution functions. In 1939 Smirnov [10] proved the following two theorems:

$$(1.1) \quad \lim_{N\to\infty} P\{N^{\frac{1}{2}} \sup_{-\infty<t<+\infty} (F_{1n}(t) - F_{2m}(t)) < y\} = 1 - e^{-2y^2},$$

if $y > 0$, zero otherwise, and

$$(1.2) \quad \lim_{N\to\infty} P\{N^{\frac{1}{2}} \sup_{-\infty<t<+\infty} |F_{1n}(t) - F_{2m}(t)| < y\} = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2y^2},$$

if $y > 0$, zero otherwise.

In both cases $N = nm/(n + m)$, and $N \to \infty$ is to mean that $n \to \infty$, $m \to \infty$ so that $m/n \to \rho$, where ρ is a constant. (The problem of determining the exact distributions of the respective random variables for finite values of n and m was solved by Koroljuk [6] on the assumption that $m = np$ where p is an integer.)

Results (1.1) and (1.2) are used to test the statistical hypothesis that two random samples come from the same unknown population. Even if $F(t)$, the hypothetical distribution function of the two random samples in question, was assumed to have a specific form, we would not get more information out of these theorems, for they consider the supremum of the difference $(F_{1n}(t) - F_{2m}(t))$ and that of its absolute value with the same weight 1, regardless of the value of $F(t)$. Thus in this way the idea arises of considering the limit distribution of the quotients $\{F_{1n}(t) - F_{2m}(t)\}/F(t)$ and $|F_{1n}(t) - F_{2m}(t)|/F(t)$, with the natural limitation on $F(t)$ that we restrict ourselves to an interval $t_a \leqq t < +\infty$, where $F(t_a) = a > 0$, when taking the supremum of these random variables. The value of a can be arbitrarily close to zero.

2. Statement, discussion, and consequences of theorems. Using the notation and assumptions of Section 1 we are going to prove the following theorems (for the definition of the distribution functions $\Phi(\cdot)$, $L(\cdot)$, $N(\cdot)$ and $R(\cdot)$ of Theorems 1, 2, 3 and 4 we refer the reader to (3.4), (3.5), (3.6) and (3.7) of [8] respectively).

THEOREM 1.

$$(2.1) \quad \lim_{N\to\infty} P\{N^{\frac{1}{2}} \sup_{a\leqq F(t)} (F_{1n}(t) - F_{2m}(t))/F(t) < y\} = \Phi(y\{a/[1 - a]\}^{\frac{1}{2}}),$$

if $y > 0$, $0 < a < 1$, *zero otherwise.*

Received 19 February 1965; revised 25 March 1965.

[1] This research was partially supported by the Canadian National Research Council at McGill University and by the Office of Naval Research 042-023, Contract Nonr 1858(05) to Princeton University.