ON SEVERAL STATISTICS RELATED TO EMPIRICAL DISTRIBUTION FUNCTIONS[1]

By Meyer Dwass

Northwestern University

1. Introduction. Let $X_1, \cdots, X_n$ be n independent random variables, each with the same continuous c.d.f., $F(x)$. Let $F_n(x)$ be the empirical c.d.f. of the X_i's. We consider the following random variables,

$$U_n = \mu\{F(t):F_n(t) - F(t) > 0\},$$

$$D_n = \sup_{-\infty<t<\infty} (F_n(t) - F(t)),$$

$$V_n = \inf_{-\infty<t<\infty} \{F(t):F_n(t) - F(t) = D_n\},$$

where $\{F(t): \quad \}$ denotes the set of values of $F(t)$, for which t satisfies the condition after the colon. These are sets in the interval (0, 1). In the definition of U_n, $\mu\{ \quad \}$ means Lebesgue measure. Obviously, there is no loss of generality in supposing that the X_i's are uniformly distributed over (0, 1) and hence

$$(1) \qquad \begin{cases} U_n = \mu\{t:F_n(t) - t > 0\}, \\ D_n = \sup_{0\leqq t\leqq 1} (F_n(t) - t), \\ V_n = \inf_{-\infty<t<\infty} \{t:F_n(t) - t = D_r\}. \end{cases}$$

In [5], Kac showed that as $n \to \infty$, U_n has an asymptotic distribution which is uniform over (0, 1). A stronger result was recently obtained by Gnedenko and Mihalevič in [4] in which they showed that *for every* n, U_n is uniformly distributed. Birnbaum and Pyke in a forthcoming paper [2] show that for every n, V_n is also distributed uniformly over (0.1). The methods of [2] and [4] are computational and the purpose of this note is to derive the uniform distribution of U_n and V_n by a short method which employs results of E. S. Andersen and a well-known relationship between the Poisson process and uniformly distributed random variables. In Sec. 3, a generalization of these results is given.

2. Proof of uniform distribution of U_n and V_n. The proof depends on two sets of facts. The first refers to the Poisson process. By this we mean the stochastic process, $X(t)$, with independent and stationary Poisson distributed increments, defined for $t \geqq 0$ and such that $X(0) = 0$. For this process, it is well known that given that $X(1) = n$, a positive integer, then the conditional distribution of the discontinuity (jump) points, $t_1 \leqq t_2 \leqq \cdots \leqq t_n$ of $X(t)$, $0 \leqq t \leqq 1$, is that

Received February 11, 1957; revised June 18, 1957.

[1] Sponsored by Air Force Office of Scientific Research AFOSR-TN-57-784, AD148015 Contract No. AF 49(638)-151.