# TESTS OF FIT BASED ON THE NUMBER OF OBSERVATIONS FALLING IN THE SHORTEST SAMPLE SPACINGS DETERMINED BY EARLIER OBSERVATIONS[1]

By Lionel Weiss

*Cornell University*

**1. Introduction.** Suppose the random variables $X_1, X_2, \cdots$ are known to be independent and identically distributed, with a continuous cumulative distribution function which is otherwise unknown. The problem which this paper discusses is the familiar one of testing the hypothesis that the cumulative distribution function is equal to a given completely specified cumulative distribution function $G(x)$. By using the random variables $G(X_1), G(X_2), \cdots$ in place of $X_1, X_2, \cdots$, the problem becomes that of testing the hypothesis that the common cumulative distribution function of $G(X_1), G(X_2), \cdots$ is the uniform distribution function $U(x)$, where $U(x) = x$ for $0 \leqq x \leqq 1$. For the remainder of the paper, it will be assumed that the problem has been reduced to this form, so that there is no loss of generality in assuming that $G(x) = U(x)$, and that all distributions considered assign probability one to the closed interval [0, 1].

Let $Y_1(n), Y_2(n), \cdots, Y_n(n)$ denote the ordered values of $X_1, \cdots, X_n$, where $0 \leqq Y_1(n) \leqq Y_2(n) \leqq \cdots \leqq Y_n(n) \leqq 1$. For convenience, $Y_0(n)$ is defined as 0 and $Y_{n+1}(n)$ is defined as 1. $T'_i(n)$ denotes the closed interval $[Y_{i-1}(n), Y_i(n)]$, and $T_i(n)$ denotes the length of this interval, for $i = 1, \cdots, n+1$. The $T'_i(n)$ are known as sample spacings.

Let $p$ be a fixed quantity in the open interval $(0, 1)$. The set $S_n(p)$ is defined as the union of the shortest sample spacing, the next shortest sample spacing, $\cdots$, until the total length of the sample spacings included in $S_n(p)$ is exactly equal to $p$. With probability one, this will require the use of a portion of the last sample spacing used, which for convenience will always be taken as the left-hand portion of the sample spacing broken up. The chance event $C_n(p)$ is defined as that event which occurs when and only when the random variable $X_{n+1}$ falls in the set $S_n(p)$.

If the hypothesis of a uniform distribution is true, the chance events $C_1(p)$, $C_2(p), \cdots$ are independent events, each with probability exactly equal to $p$. If the hypothesis is not true, the chance events are not independent and their probabilities are not all the same. However, the definition of the set $S_n(p)$ clearly favors the inclusion of those sections of the unit interval at which the true density function is relatively high, and it seems reasonable to suppose that the conditional probability of $C_n(p)$ given $X_1, \cdots, X_n$ has a high probability of approaching some limit greater than $p$. This conjecture is proved by the theorem of

Received November 27, 1959; revised December 29, 1960.

[1] Research on original version supported by the Office of Naval Research; revision supported by the National Science Foundation, Grant No. NSF-G11321.