ON DISTRIBUTION-FREE STATISTICS[1]

By Z. W. Birnbaum and H. Rubin

University of Washington and Stanford University

1. Introduction. Let $X_1, X_2, \cdots, X_n$ be a sample of a one-dimensional random variable X which has the continuous cumulative probability function F. It has been observed [1] that, to the authors' knowledge, all distribution-free statistics considered in the past can be written in the form $\Phi[F(X_1), F(X_2), \cdots, F(X_n)]$ where Φ is a measurable symmetric function defined on the unit-cube $\{U: 0 \leqq U_i \leqq 1, i = 1, 2, \cdots, n\}$. It is the purpose of this paper to study the relationship between the class of statistics which can be written in this particular form and the class of distribution-free statistics.

2. Distribution-free statistics and statistics of structure (*d*). Let Ω and Ω' be two families of cumulative probability functions. A real quantity $W = S(X_1, X_2, \cdots, X_n, G)$ will be called a *statistic in* Ω *with regard to* Ω' if, for any $G \varepsilon \Omega$, $F \varepsilon \Omega'$, and $X_1, X_2, \cdots, X_n$ in the n-dimensional sample-space for a random variable X which has the cumulative probability function F,

(i) $S(X_1, X_2, \cdots, X_n, G)$ is defined almost everywhere in the sample-space $X_1, X_2, \cdots, X_n$ (i.e. with the possible exception of a set of probability zero), and

(ii) $W = S(X_1, X_2, \cdots, X_n, G)$ has a probability distribution; this probability distribution will be denoted by $\mathcal{P}(W; F) = \mathcal{P}[S(X_1, X_2, \cdots, X_n, G); F]$.

For example, Kolmogorov's statistic

$$D_n = \sup_{-\infty < x < \infty} | F_n(x) - G(x) |, \tag{2.1}$$

where F_n is the empirical cumulative distribution function determined by the sample $X_1, X_2, \cdots, X_n$, satisfies (i) and (ii) when $\Omega = \Omega' = \Omega_1$, the class of all nondegenerate cumulative probability functions[2], hence D_n is a statistic in Ω_1 with regard to Ω_1.

If for a statistic $S(X_1, X_2, \cdots, X_n, G)$ in Ω with regard to Ω' there exists a function Φ defined on the n-dimensional unit cube and symmetric in its arguments, such that for any $G \varepsilon \Omega$, $F \varepsilon \Omega'$ we have

$$S(X_1, X_2, \cdots, X_n, G) = \Phi[G(X_1), G(X_2), \cdots, G(X_n)]$$

Received 8/6/53.

[1] Work done under the sponsorship of the Office of Naval Research.

[2] The notations for various classes of cumulative probability functions are those introduced by Scheffé [2].