Osaka Mathematical Journal
Vol. 4, No. 1, May, 1952

On a Non-Parametric Test

By Masashi OKAMOTO

1. Introduction. Let X be a random variable having the distribution function ($d.f.$) $F(x)$. We want to test the hypothesis H_0 that $F(x)$ is identical with a specified continuous $d.f. F_0(x)$. F. N. David [1] has recently proposed the following test (though this is slightly modified in comparison with the original one):

Let $x_1, x_2, \dots, x_N$ be N independent observations of X. As $F_0(x)$ is continuous, there are real numbers $\{a_i\}$, $i=1, \dots, n-1$, such that $F_0(a_i) - F_0(a_{i-1}) = 1/n$, $i=1, \dots, n$, where $a_0 = -\infty$, $a_n = +\infty$. Let C be the set of intervals on the real line on each of which $F_0(x)$ is constant and C' be its complementary set. The intersection of $(a_{i-1}, a_i]$ with C' will be called "part". Let v be the number of parts which contain no x's and w be the number of x's which fall in C. If either w is positive or v is too large we reject H_0.

David conjectured that under the null hypothesis H_0 v is asymptotically normally distributed when $n, N \to \infty$, $N/n \to$ const. This can be proved by the method of B. Sherman [2]. Furthermore this test is consistent and unbiased against a rather general class of alternative hypotheses. As Lehmann [3] says, very little work has been done on the existence of unbiased tests for non-parametric problems. It is remarkable that David's test has this property.

2. Distribution of v under H_0. Put $u=n-v$, i.e., u is the number of parts which contain at least one x. First we shall determine the distribution of u under H_0.

Denote by P_k the probability that N x's "fill" k given parts (i.e., every x_i falls in some of them and each of them contains at least one x). The probability that N x's fall into k given parts is

$$\left(\frac{k}{n}\right)^N = \sum_{i=1}^{k} \binom{k}{i} P_i \ .$$

Therefore, for every positive integer ν,

$$\begin{aligned}\sum_{k=1}^{\nu}(-1)^{\nu-k}\binom{\nu}{k}\left(\frac{k}{n}\right)^N &= \sum_{k=1}^{\nu}(-1)^{\nu-k}\binom{\nu}{k}\sum_{i=1}^{k}\binom{k}{i}P_i \\ &= \sum_{i=1}^{\nu}\binom{\nu}{i}P_i\sum_{k=i}^{\nu}(-1)^{\nu-k}\binom{\nu-i}{k-i} \\ &= P_\nu \ ,\end{aligned}$$