Osaka Mathematical Journal
Vol. 4, No. 2, December, 1952

Some Combinatorial Tests of Goodness of Fit

By Masashi OKAMOTO

1. Introduction. We have recently [1] considered a test of goodness of fit, i.e., a test whether a random sample has come from the population with the specified continuous distribution. We now present a new approach to the same problem.

Let $X_1, \dots, X_N$ be random variables distributed independently and identically according to the *d.f.* $F(x)$. To simplify the situation it is assumed that X's range from 0 to 1. The hypothesis H_0 to be tested is that $F(x)$ is identical with the *d.f.* $F_0(x)$ of uniform distribution on the interval $(0, 1]$. We divide the interval in n small intervals $((i-1)/n, i/n]$, $i=1, \dots, n$. In the sequel the word "interval" means if not stated otherwise any of these small intervals. Among $\binom{N}{k}$ k-tuples $(X_{\alpha_1}, \dots, X_{\alpha_k})$, $1 \leq \alpha_1 < \cdots < \alpha_k \leq N$, we denote by M_k the number of those such that $X_{\alpha_1}, \dots, X_{\alpha_k}$ fall in the same interval. When we consider one observation, the more uniformly are $X_1, \dots, X_N$ (observed values) distributed among the n intervals, the smaller becomes M_k, as shown in section 7. On account of this the following test (called M_k-test) of H_0 will be useful: we accept H_0 when M_k is sufficiently small.

It is proved in this paper that when the population distribution satisfies a certain condition M_k is asymptotically normally distributed as N and n tend to infinity (Theorems 1, 2, 1′, 2′). Furthermore M_k-test is shown to be consistent (Theorem 3) and unbiased (Theorem 4) against a rather general class of alternatives. The statistics M_k are closely related with David's test (cf. [1], [2]) and can be considered as a generalisation of the chi-square test in the case of equal probabitity.

2. Definition of U_k. For real numbers $t_1, \dots, t_k$ such that $0 < t_i \leq 1$, $i=1, \dots, k$, we define

$$\Theta_k(t_1, \dots, t_k) = 1, \text{ if } t_1, \dots, t_k \text{ fall in the same interval,}$$
$$= 0, \text{ otherwise,}$$

where the word "interval" means by convention any of intervals $((i-1)/n, i/n]$, $i=1, \dots, n$). Then