The Annals of Statistics
1975, Vol. 3, No. 2, 518-520

ARGUMENTS FOR FISHER'S PERMUTATION TEST

BY ANDERS ODÉN AND HANS WEDEL

University of Göteborg

The problem of statistical comparison of two distributions, continuous as well as discrete, is considered. Very slight and reasonable modifications of traditional parametric models, e.g. 'normal distributions with equal variances', are shown to result in permutation tests, only. Fisher's permutation test is shown to have optimum properties which mean a good merit for its practical use. Further, an accurate method of determining the p-value of Fisher's test is proposed.

1. Introduction and summary. This paper concerns one of the most common problems in statistical applications, the comparison of two distributions. A new justification based on the principle of unbiasedness for permutation tests is given. Good reasons are given for the transition from the t-test to a permutation test (Fisher's test).

Assume that $(X_1, \cdots, X_r)$ and $(X_{r+1}, \cdots, X_N)$ are independent samples from distributions F and G, respectively, where F and G belong to a class $\mathscr{F}$ of one-dimensional distribution functions. In the present discussion we consider the set of stochastically ordered pairs of distribution functions, that is

$$\{(F, G) \in \mathscr{F} \times \mathscr{F}; F \geqq G \vee F \leqq G\} .$$

Consider the problem of testing

$$H_0: \{(F, G) \in \mathscr{F} \times \mathscr{F}; F \leqq G, F \neq G\}$$

against

$$H_1: \{(F, G) \in \mathscr{F} \times \mathscr{F}; F \geqq G, F \neq G\} .$$

2. Unbiasedness and conditioning. In the following it is assumed that all distributions of $\mathscr{F}$ are absolutely continuous with respect to a measure μ. We denote the densities $dF/d\mu$ and $dG/d\mu$ by f and g, respectively, and specify a distribution by F or f.

We shall require unbiasedness in testing H_0 against H_1. Further, assume that to every $F \in \mathscr{F}$ there exist two sequences $\{G_i\}$ and $\{\tilde{G}_i\}$ such that $G_i \leqq F \leqq \tilde{G}_i$, $G_i, \tilde{G}_i \in \mathscr{F}$, $G_i, \tilde{G}_i \neq F$,

$$\lim_{i\to\infty} g_i = \lim_{i\to\infty} \tilde{g}_i = f . \tag{1}$$

Then a necessary condition for unbiasedness is that the test is similar with respect to $\{(F, G); F = G \in \mathscr{F}\}$.

Next we shall deal with conditions imposed on $\mathscr{F}$ in order to make the set of order statistics $X^{(1)}, \cdots, X^{(N)}$ complete with respect to $\{(F, G); F = G \in \mathscr{F}\}$.

Received August 1973; revised February 1974.

AMS 1970 *subject classifications.* Primary 62G10.

Key words and phrases. Permutation test, unbiasedness, conditioning, Fisher's test.

It is a known result that the set of order statistics is complete with respect to the family of distributions with continuous densities. A proof is given by Lehmann (1959), page 133, example 6. By a similar technique it can be proved that the corresponding statement is true for essentially smaller families of distributions. This is shown in the following theorem.

DEFINITION. Let F_0 be a distribution function and $\varepsilon > 0$ a number. Denote by $\mathscr{G}(F_0, \varepsilon)$ the family of all distribution functions F with densities such that

$$(1 - \varepsilon)f_0(x) \leqq f(x) \leqq (1 + \varepsilon)f_0(x) \qquad \text{for} \quad |x| < \frac{1}{\varepsilon}$$

and $f(x) = 0$ if $f_0(x) = 0$.

THEOREM. *The set of order statistics is complete with respect to* $\mathscr{G}(F_0, \varepsilon)$.

REMARK. With a fixed sample size and ε small enough it is impossible to distinguish between distributions in $\mathscr{G}(F_0, \varepsilon)$ by statistical methods.

PROOF. Without loss of generality we can assume that $\varepsilon < \frac{1}{3}$. Let u be a real number such that $u > 1/\varepsilon$ and $F_0(u) - F_0(-u) > 1 - \varepsilon/2$. Then by choosing the real number C and the rectangle w in $R^{(N)}$ properly we obtain

$$\begin{aligned} \exp(-Cx^{2N} + \textstyle\sum_{i=1}^N \Theta_i x^i) &< 1 + \varepsilon/8 \qquad \text{for all} \quad x \\ &> 1 - \varepsilon/8 \qquad \text{for all} \quad |x| < u \end{aligned}$$

and for all $(\Theta_1, \cdots, \Theta_N) \in w$. Now consider the family of distributions such that

$$f(x) = c(\Theta_1, \cdots, \Theta_N)f_0(x)\exp(-Cx^{2N} + \textstyle\sum_{i=1}^N \Theta_i x^i) \tag{2}$$

for $(\Theta_1, \cdots, \Theta_N) \in w$. Every distribution of this form satisfies

$$(1 - \varepsilon)f_0(x) \leqq \frac{1 - \varepsilon/8}{1 + \varepsilon/8} f_0(x) \leqq f(x) \leqq \frac{1 + \varepsilon/8}{(1 - \varepsilon/8)(1 - \varepsilon/2)} f_0(x) \leqq (1 + \varepsilon)f_0(x)$$

for $|x| < u$ and thus it belongs to $\mathscr{G}(F_0, \varepsilon)$. By Lehmann's arguments it follows that the set of order statistics is complete with respect to the family of distributions with densities of the form (2). If a statistic is complete with respect to a family C of probability measures it is also complete with respect to every set containing C provided that none of the added measures assign positive probability to sets having zero probability for all measures in C. Thus the set of order statistics is complete with respect to the larger family $\mathscr{G}(F_0, \varepsilon)$.

We consider $\mathscr{F}$ such that for some F and ε

$$\mathscr{G}(F, \varepsilon) \subset \mathscr{F} \tag{3}$$

and none of the measures in $\mathscr{F}$ assign positive probability to sets having zero probability for all measures in $\mathscr{G}(F, \varepsilon)$. Then by the theorem $(X^{(1)}, \cdots, X^{(N)})$ is complete with respect to $\{(F, G); F = G \in \mathscr{F}\}$. Obviously $(X^{(1)}, \cdots, X^{(N)})$ is sufficient with respect to $\{(F, G); F = G \in \mathscr{F}\}$. In view of the similarity (stated in the beginning of this section), the completeness, and the sufficiency, it follows that the test must be carried out conditionally with $X^{(1)}, \cdots, X^{(N)}$ considered as given constants. The requirement of unbiasedness results in permutation tests, only. The unbiasedness of a certain class of permutation tests is evident from comments given in the next section.

Our results are based on the theory of unbiased tests. However, if $\mathscr{F}$ consists of all distribution functions the restriction to permutation tests could be justified by conditional inference principles, also. Such a principle of sufficient generality is formulated by Barndorff-Nielsen (1971).

In fact, the imposition of condition (3) is inevitable to make the theory well adapted to practice. For example, we can never state that a random variable X, inspected by measurements, has exactly a normal distribution. Sometimes we can a priori state that the departure from a normal distribution is small. More generally, if F is any candidate for the true distribution of X we have to accept as candidates all elements of $\mathscr{G}(F, \varepsilon)$, provided that ε is sufficiently small. That means a rejection of the restricted assumptions under which the t-test has any optimum properties.

3. Final remarks. For testing H_0 against H_1 we suggest the permutation test with a rejection region of the form

$$\sum_{j=1}^{r} X^{(i_j)} \leqq C \tag{4}$$

where i_j is the rank of X_j in $(X_1, \cdots, X_N)$. The unconditional unbiasedness of (4) follows from Lemma 2, page 187, of Lehmann (1959). Fisher suggested the test (4) and it is often referred to as Fisher's test.

Among all unbiased tests for testing H_0 against H_1 the test (4) is uniformly most powerful for the subclass of H_1 with elements (f, g) such that $\ln(f/g)$ is linear, including e.g. the case of 'normality and equal variances'. For using Fisher's test one must approximate

$$H(x) = P_{f=g}(\sum_{i=1}^{r} X_i \leqq x \mid X^{(1)}, \cdots, X^{(N)}) \tag{5}$$

for given $X^{(1)}, \cdots, X^{(N)}$ and x. Pitman (1937) considered this problem. An alternative method is obtained by approximating H by the initial part of the Edgeworth expansion.

REFERENCES

[1] BARNDORFF-NIELSEN, O. (1971). On conditional statistical inference. Matematisk Institut Aarhus Universitet.

[2] LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.

[3] PITMAN, E. J. G. (1937). Significance tests which may be applied to samples from any populations. *Suppl. J. Roy. Statist. Soc.* **4** 119–130.

DEPARTMENT OF MATHEMATICS
CHALMERS INSTITUTE OF TECHNOLOGY
UNIVERSITY OF GÖTEBORG
GÖTEBORG S, SWEDEN