NOTES

CONSISTENCY OF CERTAIN TWO-SAMPLE TESTS

By J. R. Blum[1] and Lionel Weiss[2]

Indiana University and University of Oregon

1. Introduction and Summary. Let $X_1, \cdots, X_m; Y_1, \cdots, Y_n$ be independently distributed on the unit interval. Assume that the X's are uniformly distributed and that the Y's have an absolutely continuous distribution whose density $g(y)$ is bounded and has at most finitely many discontinuities. Let $Z_0 = 0$, $Z_{n+1} = 1$, and let $Z_1 < \cdots < Z_n$ be the values of the Y's arranged in increasing order. For each $i = 1, \cdots, n + 1$ let S_i be the number of X's which lie in the interval $[Z_{i-1}, Z_i]$.

For each nonnegative integer r, let $Q_n(r)$ be the proportion of values among $S_1, \cdots, S_{n+1}$ which are equal to r. Suppose m and n approach infinity in the ratio $(m/n) = \alpha > 0$. In Section 2 it is shown that

$$\lim_{n\to\infty} \sup_{r \geqq 0} |Q_n(r) - Q(r)| = 0$$

with probability one, where

$$Q(r) = \alpha^r \int_0^1 \frac{g^2(y)}{[\alpha + g(y)]^{r+1}} \, dy.$$

This result may be used to prove consistency of certain tests of the hypothesis that the two samples have the same continuous distribution. Several such examples are given in Section 3. A further property of one of these tests is briefly discussed in Section 4.

2. The convergence theorem. With $Q_n(r)$ and $Q(r)$ defined as in the previous section we have the

Theorem.

$$P\{\lim_{n\to\infty} \sup_{r \geqq 0} |Q_n(r) - Q(r)| = 0\} = 1.$$

Proof. We shall first prove that $\lim_{n\to\infty} Q_n(r) = Q(r)$ with probability one, where r is any positive integer. The proof for $r = 0$ is entirely analogous. To this end let $W_i = Z_i - Z_{i-1}, i = 1, \cdots, n + 1$ and define $V_i(r)$ by

$$V_i(r) = \begin{cases} 1, & \text{if } S_i = r \\ 0, & \text{otherwise} \end{cases} \qquad \text{for } i = 1, \cdots, n + 1.$$

Received November 4, 1955.

[1] Research supported in part by the United States Air Force under Contract No. AF18(600)–685 monitored by the Office of Scientific Research.

[2] Research sponsored by the Office of Naval Research.

Then

$$Q_n(r) = \frac{1}{n+1} \sum_{i=1}^{n+1} V_i(r)$$

and it follows that

$$P\{V_i(r) = 1 \mid W_i = w_i\} = \frac{m!}{r!(m-r)!} w_i^r (1 - w_i)^{m-r},$$

$$\text{(2.1)} \quad P\{V_i(r) = 1, V_j(r) = 1 \mid W_i = w_i, W_j = w_j\}$$

$$= \frac{m!}{r!^2(m-2r)!} w_i^r w_j^r (1 - w_i - w_j)^{m-2r} \qquad \text{for } i \neq j.$$

Let $\mu(w_1, \cdots, w_{n+1}; r) = E\{Q_n(r) \mid W_1 = w_1, \cdots, W_{n+1} = w_{n+1}\}$ and let $\mu(W_1, \cdots, W_{n+1}; r)$ be the corresponding random variable. From (2.1) we obtain

$$\text{(2.2)} \qquad \mu(W_1, \cdots, W_{n+1}; r) = \frac{1}{n+1} \sum_{i=1}^{n+1} \frac{m!}{r!(m-r)!} W_i^r (1 - W_i)^{m-r}.$$

For each $t \geqq 0$, let $R_n(t)$ be the proportion among $W_1, \cdots, W_{n+1}$ which do not exceed t/n, and let

$$R(t) = 1 - \int_0^1 e^{-tg(y)} g(y)\, dy.$$

It was shown by one of the authors [1] that

$$\text{(2.3)} \qquad P\{\lim_{n\to\infty} \sup_{t \geqq 0} |R_n(t) - R(t)| = 0\} = 1.$$

Setting $m = \alpha n$ we rewrite (2.2) as

$$\text{(2.4)} \qquad \begin{aligned} \mu(W_1, \cdots, W_{n+1}; r) &= \frac{1}{r!} (\alpha n)^r \prod_{j=0}^{r-1} \left[1 - \frac{j}{\alpha n}\right] \frac{1}{n+1} \sum_{i=1}^{n+1} W_i^r (1 - W_i)^{\alpha n - r} \\ &= \frac{\alpha^r}{r!} \prod_{j=0}^{r-1} \left[1 - \frac{j}{\alpha n}\right] \int_0^n t^r \left(1 - \frac{t}{n}\right)^{\alpha n - r} dR_n(t). \end{aligned}$$

Using (2.3), (2.4), and a straightforward analytic argument we find that

$$\text{(2.5)} \qquad \begin{aligned} \lim_{n\to\infty} \mu(W_1, \cdots, W_{n+1}; r) &= \frac{\alpha^r}{r!} \int_0^\infty t^r e^{-\alpha t}\, dR(t) \\ &= \frac{\alpha^r}{r!} \int_0^\infty \int_0^1 t^r e^{-t[\alpha + g(y)]} g^2(y)\, dy\, dt \\ &= \frac{\alpha^r}{r!} \int_0^1 \Gamma(r+1) \frac{g^2(y)}{[\alpha + g(y)]^{r+1}}\, dy \\ &= \alpha^r \int_0^1 \frac{g^2(y)}{[\alpha + g(y)]^{r+1}}\, dy = Q(r), \end{aligned}$$

with probability one.

Next we use Chebyshev's inequality and (2.1) to obtain

$$
\begin{aligned}
&P\{\, | \, Q_n(r) - \mu(W_1, \cdots, W_{n+1}; r) \, | > \epsilon \} \\
(2.6) \quad &\leqq \frac{1}{\epsilon^2} \frac{1}{(n+1)^2} \sum_{i=1}^{n+1} E\left\{\left[V_i(r) - \frac{m!}{r!(m-r)!} W_i^r(1 - W_i)^{m-r}\right]^2\right\} \\
&+ \frac{1}{\epsilon^2} \frac{1}{(n+1)^2} \sum_{i \neq j} E\left\{\left[V_i(r) - \frac{m!}{r!(m-r)!} W_i^r(1 - W_i)^{m-r}\right] \cdot \right. \\
&\qquad \left. \left[V_j(r) - \frac{m!}{r!(m-r)!} W_j^r(1 - W_j)^{m-r}\right]\right\}
\end{aligned}
$$

for every $\epsilon > 0$.

On examining the right-hand side of (2.6) we find that the first sum is $0[1/(n + 1)]$ since each term is nonnegative and bounded by one.

The same holds true for the second sum. This can be seen by rewriting each term in the form

$$
E\left\{\frac{m!}{r!^2(m - 2r)!} W_i^r W_j^r(1 - W_i - W_j)^{m-2r} - \frac{m!^2}{r!^2(m - r)!^2} W_i^r W_j^r(1 - W_i)^{m-r}(1 - W_j)^{m-r}\right\}
$$

and maximizing the sum subject to the conditions $0 \leqq W_i \leqq 1$, $\sum_{i=1}^{n+1} W_i = 1$. Now we use (2.5), (2.6), and the Borel-Cantelli lemma to obtain

$$
P\{\lim_{n \to \infty} Q_{n^2}(r) = Q(r)\} = 1. \tag{2.7}
$$

For each positive integer n, let $K(n)$ be the positive integer satisfying $[k(n) - 1]^2 < n \leqq k^2(n)$. From the definition of $Q_n(r)$ it follows immediately that $|[k^2(n) + 1]Q_{k^2(n)}(r) - (n + 1)Q_n(r)| \leqq \alpha|k^2(n) - n|$. From this and the fact that $\lim_{n\to\infty} [k^2(n)/n] = 1$ we see that (2.7) implies

$$
P\{\lim_{n \to \infty} Q_n(r) = Q(r)\} = 1.
$$

To complete the proof of the theorem we merely note that the uniformity of the convergence is an immediate consequence of the fact that for each integer n we have

$$
\sum_{r=0}^{\infty} Q_n(r) = 1 = \sum_{r=0}^{\infty} Q(r).
$$

3. Applications. Most tests proposed for testing the hypothesis that two samples come from the same continuous distribution are based on the ranks of one set of observations in the combined ordered sample. In our terminology these are the tests based on the statistics $S_1, \cdots, S_{n+1}$. In this section we give several examples of such tests which can be shown to be consistent against wide classes of alternatives by applying the theorem of the previous section. Throughout this section we make the usual assumption, valid for rank tests, that the X's are uniformly distributed on the unit interval and that the range of the Y's is also the unit interval.

As a first example we consider the run test proposed by Wald and Wolfowitz [2]. Let U be the number of runs of X's and Y's in the combined ordered sample. The hypothesis is rejected when U/m is too small. From the definition of $Q_n(r)$ it follows easily that

$$\left| \frac{U}{m} - 2\frac{n+1}{m}[1 - Q_n(0)] \right| \leqq \frac{1}{m}.$$

Thus if $g(y)$ is any density satisfying the assumptions of Section 1 we find that U/m converges with probability one to

$$\frac{2}{\alpha}\left[1 - \int_0^1 \frac{g^2(y)}{\alpha + g(y)}\,dy\right] = 2\int_0^1 \frac{g(y)}{\alpha + g(y)}\,dy$$

and the test is consistent against all such densities for which

$$\frac{1}{\alpha + 1} > \int_0^1 \frac{g(y)}{\alpha + g(y)}\,dy.$$

From a simple variational argument it follows that this holds for all such densities which are positive almost everywhere and differ from one on a set of positive measure. This result was obtained in [2].

Let k be a positive integer and let U_k be the number of intervals $[Z_{i-1}, Z_i]$ containing at most k X's. Consider the class of alternative densities $g(y)$ for which

$$\int_0^1 \frac{g(y)}{[\alpha + g(y)]^{k+1}}\,dy < \frac{1}{(\alpha + 1)^{k+1}}.$$

By an argument similar to the one given in the last paragraph it follows that the test which rejects when U_k is too large is consistent against alternative densities in this class.

As a third example we consider the test which rejects when $V^2 = [1/(n+1)]\sum_{i=1}^{n+1} S_i^2 = \sum_{r=0}^{\infty} r^2 Q_n(r)$ is too large. We note that $V^2 = [m^2/(n+1)]\{C^2 + [1/(n+1)]\}$ where C^2 is the statistic first proposed by Dixon [3]. Dixon computed the mean and variance of C^2 under the assumption that the hypothesis is true. Using these results we find that

$$E\{V^2\} = \frac{m(n + 2m)}{(n+1)(n+2)}$$

and

$$\sigma_{V^2}^2 = \frac{4mn(m-1)(m+n+1)(m+n+2)}{(n+1)^2(n+2)^2(n+3)(n+4)}$$

when the hypothesis is true. It follows that under this assumption V^2 converges stochastically to $2\alpha^2 + \alpha$. Now if $g(y)$ is any density satisfying the hypothesis of our convergence theorem it is easily verified that

$$\liminf_{n\to\infty} \sum_{r=0}^{\infty} r^2 Q_n(r) \geqq 2\alpha^2 \int_0^1 \frac{dy}{g(y)} + \alpha$$

with probability one. Thus the test is consistent for any such density for which $1 < \int_0^1 dy / g(y)$.

4. A further property of a test. In this section we discuss briefly another aspect of the last test considered in the previous section. Let $\{g_c(y)\}$ be the class of alternatives defined by $g_c(y) = 1 + c(y - \frac{1}{2})$ where $0 < |c| \leqq 2$. Let $(t_1, \cdots, t_{n+1})$ be $n + 1$ integers with $0 \leqq t_1 \leqq \cdots \leqq t_{n+1}$ and $\sum_{i=1}^{n+1} t_i = m$, and let U be the set of $(n + 1)$-tuplets $(s_1, \cdots, s_{n+1})$ of nonnegative integers which, when reordered according to size, yield the numbers $t_1, \cdots, t_{n+1}$. Further let $P_c(U)$ be the probability of the set U computed under the assumption that $g_c(y)$ is the density of the Y's. $P_c(U)$ can be written down in the form of an integral over the n-dimensional unit cube. After appropriate integration it turns out that

$$\left.\frac{dP_c(U)}{dc}\right|_{c=0} = 0 \quad \text{and} \quad \left.\frac{d^2P_c(U)}{dc^2}\right|_{c=0} = a \sum_{i=1}^{n+1} t_i^2 + b,$$

where a and b are positive numbers depending only on m and n. As a consequence we find that if we restrict ourselves to tests which are symmetric in the variables $S_1, \cdots, S_{n+1}$ then the test which rejects the hypothesis when V^2 is too large maximizes the slope of the power function at $c = 0$. In the Neyman-Pearson terminology, the test is of type A among the class of symmetric tests of the hypothesis.

REFERENCES

[1] L. WEISS, "The stochastic convergence of a function of sample successive differences," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 532–536.

[2] A. WALD AND J. WOLFOWITZ, "On a test whether two samples are from the same population," *Ann. Math. Stat.*, Vol. 11 (1940), pp. 147–162.

[3] W. J. DIXON, "A criterion for testing the hypothesis that two samples are from the same population," *Ann. Math. Stat.*, Vol. 11 (1940), pp. 199–204.