4. Greenfield, C. C. and Tam, S. M. (1976). *J. R. Statist. Soc. A*, **139**, 96–103.
5. Macarthur, E. W. (1983). *J. R. Statist. Soc. A*, **146**, 85–86.
6. Nour, EI-S. (1982). *J. R. Statist. Soc. A*, **145**, 106–116.

See also Log-linear Models in Contingency Tables; Tetrachoric Correlation Coefficient; and Two-By-Two (2 × 2) Tables.

TWOING INDEX

An index of "distinctiveness" or diversity* used in some recursive partitioning* procedures. If a node is split into two nodes, say L and R, in proportions $p_L : p_R (p_L + p_R = 1)$ such that the proportion of items from class (j) in $L(R)$ is $P_{jL}(P_{jR})$, the *twoing index* of the split is

$$\tfrac{1}{4} p_L p_R \left\{ \textstyle\sum_j |p_{jL} - p_{jR}| \right\}^2 .$$

The greater the index, the more effective the split.

See also Diversity Indices and Recursive Partitioning.

TWO-PHASE SAMPLING. See Survey Sampling

TWO-SAMPLE MATCHING TEST

The two-sample matching test is a procedure to test whether two independent samples come from the same continuous distribution or not. Let X and Y be independent random variables with continuous cumulative distribution functions F and G respectively, and let $X_{(k)}$ and $Y_{(k)}$, $1 \leqslant k \leqslant n$, be the order statistics* of independent random samples from F and G. In inference with two samples, one issue of interest is whether the two samples indeed come from two different populations or not. The standard procedure is to test the null hypothesis $H_0 : G = F$.

Since comparison of two samples is a fundamental topic of statistics, interest in testing the above hypothesis goes back to the late thirties. Smirnov [7] proposed the test statistic $\sup_{-\infty<x<\infty} |F_n(x) - G_m(x)|$ for arbitrary sample sizes against the alternative $H_1 : G \neq F$, where $F_n(x)$ and $G_m(x)$ are the empirical distribution functions of F and G respectively (*see* Kolmogorov–Smirnov Statistics). The distribution of this test statistic and generalizations of it were studied later by several authors, including Takács [8], where related references can also be found.

However, if the main concern is a specific alternative, other nonparametric tests with greater power are preferable. In particular, the well-known nonparametric tests such as the Wilcoxon, Mann–Whitney*, and von Mises tests are developed against location shift alternatives that assume that $G(x) = F(x - \Delta)$, and the corresponding alternatives $H_1 : \Delta > (<, \neq) 0$ then induce a stochastic ordering between F and G [2,4].

The two-sample matching test to be described here is appropriate for any general alternative $H_1 : F(x) \neq G(x)$. In this entry, the exact distribution and the moments of the test statistic are presented and the limiting distribution is provided, together with the result that the test is consistent against a general alternative. More details are presented in refs. 3, 6.

Assume without loss of generality that the distributions F and G are concentrated on the unit interval [0, 1], and that $F(x) = x$, $0 \leqslant x \leqslant 1$. The test considered in Rao and Tiwari [5] and Gürler and Siddiqui [1] is based on the number of *matches* between the order statistics $X_{(k)}$ and $Y_{(k)}$, $k = 1, \ldots, n$, from F and G respectively. For the event A_k, a match occurs if $X_{(k)} \in (Y_{(k-1)}, Y_{(k)}]$ for any k, $1 \leqslant k \leqslant n$, with $Y_{(0)} = 0$. The *matching* test statistic is then $S_n = \sum_{k=1}^{n} \boldsymbol{I}_k$, where $\boldsymbol{I}_k$ is the indicator of A_k. S_n is a special case of a more general class of statistics considered in ref. 8. Later, Rao and Tiwari [5] provided a simpler proof for the special case of S_n defined above. Let P_0 and P_1 be the probability measures, and E_0 and E_1 the expectations, under H_0 and H_1, respectively. Clearly, for $k \geqslant 1$, $P_0(S_n \geqslant k) \geqslant P_1(S_n \geqslant k)$. Hence, for a significance level α, $0 < \alpha < 1$, the critical region for testing H_0 against H_1 is of the form $S_n \leqslant k_{n,\alpha}$

Table 1.

$X_{(i)}$	$Y_{(i)}$
0.094	0.0039
0.168	0.0041
0.229	0.0064
0.265	0.0116
0.384	0.0706
0.460	0.0997
0.482	0.1028
0.511	0.1069
0.523	0.5792
0.710	0.6155

where $k_{n,\alpha}$ is the largest integer k such that $P_0(S_n \leqslant k) \leqslant \alpha$.

Example. Suppose an X-sample of size 10 is observed from a uniform distribution on (0, 1), and a Y-sample of the same size is observed from a beta distribution* with parameters $\alpha = 0.5$ and $\beta = 1.5$, with the order statistics shown in Table 1. Then $S_n = 1$, indicating that there is only one matching between the order statistics of the two samples at $X_{(9)} = 0.523$, which lies between 0.1069 and 0.5792. We will see shortly that, according to the asymptotic distribution of S_n, we reject H_0 for this particular sample.

DISTRIBUTION, MEAN, AND VARIANCE OF S_N

An exact expression for $P_0(S_n \geqslant k)$, given in ref. 8, uses lattice-path counting methods. The alternative proof in ref. 5 is based on the number of *crossings* in the Kolmogorov–Smirnov statistics. The exact distribution of S_n under H_0, for $k = 1, \ldots, n$, is given by

$$P(S_n \geqslant k) = \binom{2n}{n+k} \bigg/ \binom{2n}{n}. \qquad (1)$$

The mean and the variance of S_n, under the hypothesis $H_0 : F = G$, are given by

$$E[S_n] = \frac{2^{2n-1}}{\binom{2n}{n}} - \frac{1}{2},$$

$$\operatorname{Var}[S_n] = n + \frac{1}{4} - \frac{2^{4n-2}}{\binom{2n}{n}^2}.$$

For small n, the exact distribution of S_n can be tabulated using (1). However, when the sample size is large, it is desirable to exhibit the rejection region in a simpler form for practical purposes, via the limiting distribution of S_n. Under the null hypothesis of identical distributions for X and Y, and for $x > 0$ [1],

$$\lim_{n\to\infty} P_0(n^{-1/2} S_n \geqslant x) = e^{-x^2}.$$

From this result, approximate critical regions of the test can be found. For a size-α test, we reject if $S_n \geqslant k_{n,\alpha}$, where $k_{n,\alpha} = [-n \ln(1 - \alpha)]^{1/2}$. Then, if we fix $\alpha = 0.05$, for $n = 5$, 10, 15, 20, 30, the $k_{n,\alpha}$-values are obtained as $k_{n,0.05} = 0.506$, 0.716, 0.877, 1.01, 1.24 respectively. In the example above, since $S_{10} = 1 > 0.716$, we reject H_0.

Consistency

A test is *consistent* if its power tends to one as n tends to infinity, against a fixed alternative. If the measure of the interval on which $F(x)$ and $G(x)$ do not cross each other is one (that is, if they cross each other only at countably many points), then the test based on S_n is consistent for any significance level α, $0 < \alpha < 1$ [1]. This result guarantees that as the sample size increases, the power of the matching test converges to one against a general alternative and therefore the test is consistent. However, the result does not indicate anything about the small-sample behavior of the test and the rate of convergence for the power.

REFERENCES

1. Gürler, Ü. and Siddiqui, M. M. (1996). On the consistency of a two-sample matching test. *Nonparametric Statist.*, **7**, 69–73.
2. Hájek, J. and Šidák, Z. (1967). *Theory of Ranked Tests*. Academic Press, New York.
3. Khidr, A. M. (1981). Matching of order statistics with intervals. *Indian J. Pure Appl. Math.*, **12**, 1402–1407.
4. Randles, R. H. and Wolfe, D. A. (1979). *Introduction to the Theory of Nonparametric Statistics*. Wiley, New York.
5. Rao, J. S. and Tiwari, R. C. (1983). One- and two-sample match statistics. *Statist. Probab. Lett.*, **1**, 129–135.

6. Siddiqui, M. M. (1982). The consistency of a matching test. *J. Statist. Plann. Inference*, **6**, 227–233.
7. Smirnov, N. V. (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. (In Russian.) *Bull. Math. Univ. Moscow A*, **2**, 3–14.
8. Takács, L. (1971). On the comparison of two empirical distribution functions. *Ann. Math. Statist.*, **42**, 1157–1166.

See also EMPIRICAL DISTRIBUTION FUNCTION (EDF) STATISTICS; GOODNESS-OF-FIT DISTRIBUTION, TAKÁCS; KOLMOGOROV–SMIRNOV STATISTICS; and ORDER STATISTICS.

ÜLKÜ GÜRLER

TWO-SAMPLE PROBLEM.

See LOCATION TESTS

TWO-SAMPLE PROBLEM, BAUMGARTNER–WEISS–SCHINDLER TEST

Tests that two independent samples $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_m$ belong to the *same continuous* population (two-sample problem) abound in the statistical literature. Among them, the Kolmogorov-Smirnov* test, the Cramér-von Mises* test, and the Mann-Whitney–Wilcoxon* test are perhaps the most popular in applications.

These tests are based on comparisons of the two empirical* distribution functions or edfs

$$F_n(x) = \begin{cases} 0 & \text{for } x < X_1, \\ i/n & \text{for } X_i < x < X_{i+1} \\ 1 & \text{for } x > X_n, \end{cases}$$

and $F_m(x)$, defined analogously for the sample $Y_1, Y_2, \ldots, Y_m$. (Observations here are ordered in increasing order of magnitude.)

The Kolmogorov-Smirnov test uses the maximum of $|F_n(x) - F_m(x)|$, the Wilcoxon test utilizes the integral of $[F_n(x) - F_m(x)]$, and the Cramér-von Mises test employs the squared norm of $[F_n(x) - F_m(x)]$ as test values.

The Baumgartner–Weiss–Schindler (BWS) test [2] uses the squared norm of the difference $[F_n(x) - F_m(x)]$ weighted by its variance, an idea borrowed from the Anderson-Darling test for the one-sample problem [1].

Specifically, for testing the null hypothesis that both samples are drawn from the same population with continuous but unknown cdf $F(x)$, the integral transformation

$$Z = F(X)$$

is introduced, which results in Z being almost surely uniformly distributed on [0,1] if X has the cdf $F(x)$. Applying this transformation to the two sets of data, we arrive at the edfs $\hat{F}_n(z)$ and $\hat{F}_m(z)$. The difference $\hat{F}_n(z) - \hat{F}_m(z)$ is weighted by $[z(1-z)]^{-1}$, yielding a test value

$$\tilde{B} = \frac{mn}{m+n} \int_0^1 [z(1-z)]^{-1} (\hat{F}_n(z) - \hat{F}_m(z))^2 dz \tag{1}$$

Since $F(x)$ is unknown, BWS 2 propose to approximate Equation (1) using ranks. The rank G_i (resp. H_i) of each element X_i (resp. Y_j) is defined as the number of data values in both samples smaller than or equal to X_i (resp. Y_j). Then $\tilde{B}$ is approximated by

$$B_X = \frac{1}{n} \sum_{i=1}^{n} \frac{(G_i - \frac{m+n}{n} i)^2}{\frac{i}{n+1}(1 - \frac{i}{n+1}) \frac{m(m+n)}{n}} \tag{2}$$

or by

$$B_Y = \frac{1}{m} \sum_{j=1}^{m} \frac{(H_j - \frac{m+n}{m} j)^2}{\frac{j}{m+1}(1 - \frac{j}{m+1}) \frac{n(m+n)}{m}} \tag{3}$$

and new test statistic B is defined to be

$$B = (B_X + B_Y)/2.$$

The processes in Equations (2) and (3) converge to a standard Brownian bridge* as $n, m \to \infty$ and $n/m \to a$ (where a is a finite constant). Applying the Anderson-Darling technique [1], BWS derive the asymptotic distribution of B to be

$$\begin{aligned} \Psi(b) &= \lim_{n,m\to\infty} \Pr(B < b) \\ &= \sqrt{\frac{\pi}{2}} \frac{1}{b} \sum_{j=0}^{\infty} \left\{ \binom{-\frac{1}{2}}{j} (4j+1) \int_0^1 \frac{1}{\sqrt{r^3(1-r)}} \right. \\ &\quad \left. \times \exp\left(\frac{rb}{8} - \frac{\pi^2(4j+1)^2}{8rb} \right) dr \right\}, \end{aligned}$$