Optimal Gridding of Visibility Data in Radio Interferometry

Article · January 1984

1 author:

Frederic R. Schwab
National Radio Astronomy Observatory

K-Sample Anderson-Darling Tests

F. W. Scholz; M. A. Stephens

Journal of the American Statistical Association, Vol. 82, No. 399. (Sep., 1987), pp. 918-924.

K-Sample Anderson–Darling Tests

F. W. SCHOLZ and M. A. STEPHENS*

Two k-sample versions of an Anderson–Darling rank statistic are proposed for testing the homogeneity of samples. Their asymptotic null distributions are derived for the continuous as well as the discrete case. In the continuous case the asymptotic distributions coincide with the $(k-1)$-fold convolution of the asymptotic distribution for the Anderson–Darling one-sample statistic. The quality of this large sample approximation is investigated for small samples through Monte Carlo simulation. This is done for both versions of the statistic under various degrees of data rounding and sample size imbalances. Tables for carrying out these tests are provided, and their usage in combining independent one- or k-sample Anderson–Darling tests is pointed out.

The test statistics are essentially based on a doubly weighted sum of integrated squared differences between the empirical distribution functions of the individual samples and that of the pooled sample. One weighting adjusts for the possibly different sample sizes, and the other is inside the integration placing more weight on tail differences of the compared distributions. The two versions differ mainly in the definition of the empirical distribution function. These tests are consistent against all alternatives. The use of these tests is two-fold: (a) in a one-way analysis of variance to establish differences in the sampled populations without making any restrictive parametric assumptions or (b) to justify the pooling of separate samples for increased sample size and power in further analyses. Exact finite sample mean and variance formulas for one of the two statistics are derived in the continuous case. It appears that the asymptotic standardized percentiles serve well as approximate critical points of the appropriately standardized statistics for individual sample sizes as low as 5. The application of the tests is illustrated with an example. Because of the convolution nature of the asymptotic distribution, a further use of these critical points is possible in combining independent Anderson–Darling tests by simply adding their test statistics.

KEY WORDS: Combining tests; Convolution; Consistency; Empirical processes; Pearson curves; Simulation.

1. INTRODUCTION AND SUMMARY

Anderson and Darling (1952, 1954) introduced the goodness-of-fit statistic

$$A_m^2 = m \int_{-\infty}^{\infty} \frac{\{F_m(x) - F_0(x)\}^2}{F_0(x)\{1 - F_0(x)\}} \, dF_0(x)$$

to test the hypothesis that a random sample $X_1, \ldots, X_m$, with empirical distribution $F_m(x)$, comes from a continuous population with completely specified distribution function $F_0(x)$. Here $F_m(x)$ is defined as the proportion of the sample $X_1, \ldots, X_m$ that is not greater than x. The corresponding two-sample version

$$A_{mn}^2 = \frac{mn}{N} \int_{-\infty}^{\infty} \frac{\{F_m(x) - G_n(x)\}^2}{H_N(x)\{1 - H_N(x)\}} \, dH_N(x) \qquad (1)$$

was proposed by Darling (1957) and studied in detail by Pettitt (1976). Here $G_n(x)$ is the empirical distribution function of the second (independent) sample $Y_1, \ldots, Y_n$ obtained from a continuous population with distribution function $G(x)$, and $H_N(x) = \{mF_m(x) + nG_n(x)\}/N$, with $N = m + n$, is the empirical distribution function of the pooled sample. The above integrand is appropriately defined to be zero whenever $H_N(x) = 1$. In the two-sample case A_{mn}^2 is used to test the hypothesis that $F = G$ without specifying the common continuous distribution function.

The k-sample Anderson–Darling test is a rank test and thus makes no restrictive parametric model assumptions. The need for a k-sample version is twofold. It can either be used to establish differences in several sampled populations with particular sensitivity toward the tails of the pooled sample or it may be used to judge whether several samples are sufficiently similar so that they may be pooled for further analysis. For example, one may be interested in testing whether several batches of data come from a common normal population. This could be done by testing first for the homogeneity of the batches using the Anderson–Darling k-sample test, and, in the case of acceptance, the order statistics of the pooled batches may then be used in a test of normality. Rejection at either stage more clearly pinpoints the reason for rejecting the hypothesis of a common normal distribution. The combination of the Type I error probabilities of such a two-stage procedure is facilitated by the fact that the Anderson–Darling rank test is statistically independent of the pooled order statistics on which the second test might be based, provided that the hypothesis accepted by the first test is correct. This independence follows through Basu's theorem (Lehmann 1983, p. 46) from the ancillarity of the ranks and the sufficiency and completeness of the order statistics when all observations come from a common (continuous) distribution (Bell, Blackwell, and Breiman 1960).

Similarly, other k-sample rank tests could satisfy the aforementioned needs. Popular tests are the Kruskal–Wallis and the Brown–Mood median tests. Of more theoretical interest is the normal scores k-sample test. For a description of these tests and further references we refer to Conover (1980), Hájek and Šidak (1967), Hollander and Wolfe (1973), and Lehmann (1975). The disadvantage of the aforementioned rank tests is that they are only consistent against a rather restricted set of alternatives and would thus be quite ineffective in certain situations, especially when the alternative is largely characterized by changes in scale.

Rank tests that do not share this weakness are typically based on some distance measure applied to the empirical distribution functions of the respective samples. Kiefer

* F. W. Scholz is Statistician, Boeing Computer Services, MS 7L-22, Seattle, WA 98124-0346, and Affiliate Associate Professor, Department of Statistics, University of Washington, Seattle, WA 98105. M. A. Stephens is Professor, Department of Mathematics and Statistics, Simon Fraser University, Burnaby, British Columbia, V5A 1S6, Canada. The authors thank Steve Rust for drawing their attention to this problem; Galen Shorack and Jon Wellner for several helpful conversations and the use of galleys of their book; and the referees for their suggestions. The variance formula (4) was derived with the partial aid of MACSYMA (1984), a large symbolic manipulation program. This research was supported in part by the Natural Sciences and Engineering Research Council of Canada (Grant A8561), and by the U.S. Office of Naval Research (Grant N00014-86-K-0156).

Journal of the American Statistical Association
September 1987, Vol. 82, No. 399, Theory and Methods

(1959) investigated several k-sample extensions of the Kolmogorov–Smirnov and Cramer–von Mises tests and provided tables for their asymptotic distributions, but the usefulness of these tables for small samples has not been investigated. Other such distance tests with tables for small samples are described in Conover (1980). These tables are applicable, however, only for samples of equal size. Since the asymptotic behavior of the Anderson–Darling tests manifests itself rather rapidly in the one- and two-sample case (Pettitt 1976; Stephens 1974) it is worthwhile to investigate its k-sample version.

In Section 2 a k-sample version of the Anderson–Darling test is proposed for the continuous case and a computational formula is given. In Section 3 the finite sample distribution of this statistic is discussed, and exact formulas for its mean and variance are given. The asymptotic null distribution of this statistic is described in Section 4, and tables of approximate percentiles are provided. Some remarks are made about the power of the test. In Section 5 we discuss the case of discrete parent populations; we give a computational formula for the proposed statistic as well as an alternate version of the k-sample Anderson–Darling statistic, which treats ties in a different manner. Monte Carlo simulation results are given in Section 6, to examine the adequacy of the approximate asymptotic percentiles for the continuous and discrete case. An example is discussed in Section 7. The tables can be used also for combining independent one- and k-sample Anderson–Darling tests; this is discussed in Section 8. The technical details of the asymptotic derivations are given in the Appendix.

2. THE k-SAMPLE ANDERSON–DARLING TEST: CONTINUOUS POPULATIONS

It is not immediately obvious how to extend the two-sample test to the k-sample situation. There are several reasonable possibilities, but not all are mathematically tractable as far as asymptotic theory is concerned. Kiefer's (1959) treatment of the k-sample analog of the Cramer–von Mises test shows the appropriate path. To set the stage the following notation is introduced. Let X_{ij} be the jth observation in the ith sample ($j = 1, \ldots, n_i$; $i = 1, \ldots, k$). All observations are independent. Suppose that the ith sample has continuous distribution function F_i. We wish to test the hypothesis

$$H_0: F_1 = \cdots = F_k$$

without specifying the common distribution F. Denote the empirical distribution function of the ith sample by $F_{in_i}(x)$ and that of the pooled sample of all $N = n_1 + \cdots + n_k$ observations by $H_N(x)$. The k-sample Anderson–Darling test statistic is then defined as

$$A_{kN}^2 = \sum_{i=1}^{k} n_i \int_{B_N} \frac{\{F_{in_i}(x) - H_N(x)\}^2}{H_N(x)\{1 - H_N(x)\}}\, dH_N(x), \qquad (2)$$

where $B_N = \{x \in R: H_N(x) < 1\}$. For $k = 2$ (2) reduces to (1). Under the continuity assumption on the F_i the probability of ties is zero. Hence the pooled ordered sample is $Z_1 < \cdots < Z_N$, and a straightforward evaluation of (2) yields the following computational formula for A_{kN}^2:

$$A_{kN}^2 = \frac{1}{N} \sum_{i=1}^{k} \frac{1}{n_i} \sum_{j=1}^{N-1} \frac{(NM_{ij} - jn_i)^2}{j(N - j)}, \qquad (3)$$

where M_{ij} is the number of observations in the ith sample that are not greater than Z_j. For the appropriate formula in the presence of ties see Section 5.

3. THE FINITE SAMPLE DISTRIBUTION UNDER H_0

Under H_0 and assuming continuity of the common distribution F the expected value of A_{kN}^2 is $k - 1$. This can be seen by conditioning on the order statistics $Z_1, \ldots, Z_N$ when taking the expectation of (2). This conditioning reduces the calculation to the evaluation and manipulation of simple hypergeometric moments.

Higher moments of A_{kN}^2 are very difficult to compute. Pettitt (1976) gave an approximate variance formula for A_{2N}^2 as $\mathrm{var}(A_{2N}^2) \approx \sigma^2(1 - 3.1/N)$, where $\sigma^2 = 2(\pi^2 - 9)/3$. This approximation does not account for any dependence on the individual sample sizes.

The same conditioning method, used for the first moment, also yields the second moment and hence the variance of A_{kN}^2. The calculations are very tedious and benefited greatly from the use of MACSYMA (1984). The variance formula of A_{kN}^2 is given for the continuous case as

$$\sigma_N^2 = \mathrm{var}(A_{kN}^2) = \frac{aN^3 + bN^2 + cN + d}{(N - 1)(N - 2)(N - 3)}, \qquad (4)$$

with

$$\begin{aligned} a &= (4g - 6)(k - 1) + (10 - 6g)H, \\ b &= (2g - 4)k^2 + 8hk + (2g - 14h - 4)H \\ &\quad - 8h + 4g - 6, \\ c &= (6h + 2g - 2)k^2 + (4h - 4g + 6)k \\ &\quad + (2h - 6)H + 4h, \\ d &= (2h + 6)k^2 - 4hk, \end{aligned}$$

where

$$H = \sum_{i=1}^{k} \frac{1}{n_i}, \qquad h = \sum_{i=1}^{N-1} \frac{1}{i},$$

and

$$g = \sum_{i=1}^{N-2} \sum_{j=i+1}^{N-1} \frac{1}{(N - i)j}.$$

Note that

$$g \to \int_0^1 \int_y^1 \frac{1}{x(1 - y)}\, dx\, dy = \frac{\pi^2}{6} \quad \text{as} \quad N \to \infty$$

and thus $\mathrm{var}(A_{kN}^2) \to (k - 1)\sigma^2$ as $\min(n_1, \ldots, n_k) \to \infty$. The effect of the individual sample sizes is reflected through H and is not negligible to order $1/N$.

From the definition of the M_{ij} in formula (3) it is clear that the statistic A_{kN}^2 depends on the sample values only through their ranks and the test based on it is thus a rank test. In principle, it is possible to derive the null distribution (under H_0) of (3) by recording the distribution of (3) as one traverses through all rank permutations. For small sample sizes it may be feasible to derive this distribution and tables could be constructed. The computational and tabulation effort, however, quickly grows prohibitive as k and N get larger. If $k = 4$ and $n_1 = \cdots = n_4 = 8$, as in the example of Section 7, it would be necessary to evaluate (3) for about 9.9 10^{16} different permutations.

A more pragmatic approach would be to record the relative frequency $\hat{p}$ with which the observed value a^2 of (3) is matched or exceeded when computing (3) for a large number M of random rank permutations. This was done, for example, to get the distribution of the two-sample Watson statistic U^2_{mn} in Watson (1962). This method is applicable equally well in small and large samples. $\hat{p}$ is an unbiased estimator of the true P value of a^2, and the variance of $\hat{p}$ can be controlled by the choice of M.

4. ASYMPTOTIC DISTRIBUTION UNDER H_0 AND CRITICAL POINTS

In the Appendix it is shown that the limiting distribution of A^2_{kN} under H_0 is the $(k - 1)$-fold convolution of the asymptotic distribution for the one-sample Anderson–Darling statistic. In particular, assuming a common continuous distribution, it follows that A^2_{kN} converges in distribution to

$$A^2_{k-1} \equiv \sum_{j=1}^{\infty} \frac{1}{j(j+1)} Y_j, \tag{5}$$

where Y_j are independent chi-squared random variables with $(k - 1)$ df. Since the cumulants and first four moments of (5) are easily calculated, approximate percentiles of the random variable A^2_{k-1} were obtained by fitting Pearson curves as in Stephens (1976) and Solomon and Stephens (1978).

For the one- and two-sample Anderson–Darling test statistics the use of the asymptotic percentiles as critical values works very well even in small samples (Pettitt 1976; Stephens 1974). This suggests the use of asymptotic percentiles here as well. These will be standardized to make the test. Table 1 contains the upper tail percentage points

$$t_{k-1}(\alpha) \text{ of } T_{k-1} = \{A^2_{k-1} - (k - 1)\}/(\sqrt{k-1}\,\sigma).$$

Following Pettitt (1976) the test statistic A^2_{kN} will be standardized using its exact finite sample mean and standard deviation. The hope that this latter form of standardization removes some of the dependence of the test on the sample size was confirmed through the Monte Carlo study described in Section 6.

The test procedure is then as follows.

1. Calculate A^2_{kN} from (3) and σ_N from (4).
2. Calculate

$$T_{kN} = \frac{A^2_{kN} - (k - 1)}{\sigma_N}.$$

3. Refer T_{kN} to the upper tail percentage points, $t_{k-1}(\alpha)$, given in Table 1; reject H_0 at significance level α if T_{kN} exceeds the given point $t_{k-1}(\alpha)$.

This use of asymptotic percentage points works very well in the case $k - 1 = 1$ and can be expected to improve as k increases.

For values of $m = k - 1$ not covered by Table 1 the following interpolation formula should give satisfactory percentiles. It reproduces the entries in Table 1 to within half a percent of relative error. The general form of the interpolation formula is

$$t_m(\alpha) = b_0 + \frac{b_1}{\sqrt{m}} + \frac{b_2}{m},$$

where the coefficients for each α may be found in Table 2. Similarly, one could interpolate and even extrapolate in Table 1 with respect to α to establish an approximate P value for the observed Anderson–Darling statistic; see Section 7 for an example.

The power behavior of this test has not been studied except to show (see the Appendix) that A^2_{kN} is consistent against all alternatives. Being consistent against omnibus alternatives naturally entails some loss of power against specific alternatives. For some indication of what might be expected in the k-sample case we refer to Pettitt's (1976) limited power study of the two-sample Anderson–Darling test. There the power of A^2_{mn} was compared with that of several other two-sample competitors in a situation where both populations were normal but differed in location and scale. For this situation A^2_{mn} clearly dominated the two-sample Cramer–von Mises W^2_{mn}, but no clear dominance emerged in relation to any of the other tests. Of the competitors studied by Pettitt, only Watson's U^2_{mn} and Cramer–von Mises's W^2_{mn} are consistent against omnibus alternatives, and a k-sample version of Watson's U^2_{mn} is not available at present.

5. DISCRETE PARENT POPULATION

So far it has been assumed that the sampled parent distributions F_i are continuous. If continuous data are grouped, or if the parent populations are discrete, tied

Table 1. Upper α Percentiles $t_m(\alpha)$ of the T_m Distribution

m	.25	.10	.05	.025	.01
1	.326	1.225	1.960	2.719	3.752
2	.449	1.309	1.945	2.576	3.414
3	.498	1.324	1.915	2.493	3.246
4	.525	1.329	1.894	2.438	3.139
6	.557	1.332	1.859	2.365	3.005
8	.576	1.330	1.839	2.318	2.920
10	.590	1.329	1.823	2.284	2.862
∞	.674	1.282	1.645	1.960	2.326

Table 2. Interpolation Coefficients

α	b_0	b_1	b_2
.25	.675	−.245	−.105
.10	1.281	.250	−.305
.05	1.645	.678	−.362
.025	1.960	1.149	−.391
.01	2.326	1.822	−.396

observations can occur. To give the computational formula in the case of tied observations we introduce the following notation. Let $Z_1^* < \cdots < Z_L^*$ denote the L (>1) distinct ordered observations in the pooled sample. Further, let f_{ij} be the number of observations in the ith sample coinciding with Z_j^* and let $l_j = \sum_{i=1}^{k} f_{ij}$ denote the multiplicity of Z_j^*. Using (2) as the definition of A_{kN}^2 the computing formula in the case of ties becomes

$$A_{kN}^2 = \sum_{i=1}^{k} \frac{1}{n_i} \sum_{j=1}^{L-1} \frac{l_j}{N} \frac{(NM_{ij} - n_i B_j)^2}{B_j(N - B_j)}, \tag{6}$$

where $M_{ij} = f_{i1} + \cdots + f_{ij}$ and $B_j = l_1 + \cdots + l_j$.

Under H_0, without assuming continuity of the common F, the expected value of A_{kN}^2 is

$$E(A_{kN}^2) = (k - 1)\frac{N}{N - 1}\left[1 - \int_0^1 \{\psi(u)\}^{N-1}\,du\right],$$

where $\psi(u) \equiv F\{F^{-1}(u)\}$ with $\psi(u) \ge u$ and $\psi(u) \equiv u$ iff F is continuous. In the continuous case this expected value reduces to $k - 1$. In general, as $N \to \infty$, the expected value converges to $(k - 1)P(\psi(U) < 1)$, where $U \sim U(0, 1)$ is uniform. The formula for the expectation of A_{kN}^2 is given for theoretical interest only and not for use in the standardization of the test statistic.

An alternate way of dealing with ties is to change the definition of the empirical distribution function to the average of the left and right limit of the ordinary empirical distribution function, that is,

$$F_{ain_i}(x) = \tfrac{1}{2}\{F_{in_i}(x) + F_{in_i}(x-)\}$$

and, similarly, $H_{aN}(x)$. Using these modified distribution functions we modify (2) slightly to

$$A_{akN}^2 = \frac{N-1}{N}\int \frac{\sum_{i=1}^{k} n_i\{F_{ain_i}(x) - H_{aN}(x)\}^2}{H_{aN}(x)\{1 - H_{aN}(x)\} - \{H_N(x) - H_N(x-)\}/4} \times dH_N(x),$$

for (nondegenerate) samples whose observations do not all coincide. Otherwise let $A_{akN}^2 = 0$. The denominator of the integrand of A_{akN}^2 is chosen to simplify the mean of A_{akN}^2. For nondegenerate samples the computational formula for A_{akN}^2 becomes

$$A_{akN}^2 = \frac{N-1}{N}\sum_{i=1}^{k}\frac{1}{n_i}\sum_{j=1}^{L}\frac{l_j}{N}\frac{(NM_{aij} - n_i B_{aj})^2}{B_{aj}(N - B_{aj}) - Nl_j/4}, \tag{7}$$

where $M_{aij} = f_{i1} + \cdots + f_{ij-1} + f_{ij}/2$ and $B_{aj} = l_1 + \cdots + l_{j-1} + l_j/2$. This formula applies for a continuous population also, then all $l_j = 1$.

The expected value of A_{akN}^2 under H_0 is

$$E(A_{akN}^2) = (k - 1)\{1 - \Pr(X_{11} = \cdots = X_{kn_k})\},$$

which is $k - 1$ for continuous F and otherwise rapidly becomes $k - 1$ in the nondegenerate case as $N \to \infty$. Again this expected value is given for theoretical interest. Here, however, it turns out that its value does not differ much from the value $k - 1$ as it is used in the standardization later.

The test based on A_{akN}^2 is carried out by rejecting H_0 at significance level α whenever

$$T_{akN} \equiv \frac{A_{akN}^2 - (k - 1)}{\sigma_N} \ge t_{k-1}(\alpha).$$

Note that σ_N represents the exact standard deviation of A_{kN}^2 and not of A_{akN}^2. Simulation results suggest that this procedure is reasonable.

A variance formula was derived only for A_{kN}^2 and that only for the continuous case [Eq. (4)]. The corresponding calculation for A_{akN}^2 has not been attempted. Simulations show, however, that the variances of A_{kN}^2 and A_{akN}^2 are very close to each other in the continuous case. Here closeness is judged by the discrepancy between the simulated variance of A_{kN}^2 and that obtained by (4).

The asymptotic distribution of A_{kN}^2 is derived in the Appendix for the general case without continuity assumptions. The corresponding limiting distribution of A_{akN}^2 is also given there, without derivation. It is again a $(k - 1)$-fold convolution. In both cases the limiting distribution depends on F through ψ. This dependence on F disappears and the two limiting distributions coincide when F is continuous. Thus the limiting distribution in the continuous case can be considered an approximation to the limiting distributions of A_{kN}^2 and A_{akN}^2 under rounding of data provided that the rounding is not too severe. Analytically it appears difficult to decide which of the two discrete case limiting distributions is better approximated by the continuous case. Only a simulation may provide some answers.

6. MONTE CARLO SIMULATION

To see how well the percentiles given in Table 1 perform in small samples a number of Monte Carlo simulations were performed. For the continuous case samples were generated from a Weibull distribution, with scale parameter $a = 1$ and shape $b = 3.6$, to approximate a normal distribution reasonably well. The underlying uniform random numbers were generated using Schrage's (1979) portable random number generator. Selected typical simulation results taken from a more extensive simulation study are summarized in Table 3. For each of these simulations 5,000 pooled samples were generated. Each pooled sample was then broken down into the indicated number of subsamples with the given sample sizes. The observed probabilities of Type I error are recorded in columns 2 and 3 for the two versions of the statistic, given by (2) and (7) for continuous populations. Next, to examine discrete populations, for each pooled sample created previously the scale parameter was changed to $a = 150$, $a = 100$, and $a = 30$ and the sample values were rounded to the nearest integer. The observed Type I error probabilities are given in columns (4, 5), (6, 7), and (8, 9), respectively. On top of these columns the degree of rounding is expressed in terms of the average proportion of distinct observations in the pooled sample.

Table 3. Observed Significance Levels of A^2_{kN} and A^2_{akN}

Nominal significance level α	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a
	Sample sizes: 5, 5, 5							
	Average proportion of distinct observations							
	1.0000		.9555		.9346		.8034	
.250	.2654	.2656	.2656	.2714	.2614	.2702	.2632	.2770
.100	.1000	.1040	.0998	.1062	.0994	.1046	.1034	.1146
.050	.0476	.0502	.0488	.0526	.0486	.0532	.0500	.0586
.025	.0228	.0252	.0230	.0256	.0224	.0262	.0220	.0298
.010	.0070	.0086	.0058	.0086	.0062	.0084	.0054	.0096
	Sample sizes: 5, 10, 15, 20, 25							
	Average proportion of distinct observations							
	1.0000		.7938		.7152		.4035	
.250	.2522	.2552	.2538	.2582	.2528	.2572	.2544	.2632
.100	.1016	.1026	.1020	.1044	.1024	.1056	.1016	.1090
.050	.0494	.0510	.0496	.0516	.0496	.0526	.0492	.0566
.025	.0228	.0234	.0222	.0248	.0234	.0252	.0230	.0302
.010	.0110	.0110	.0114	.0114	.0114	.0118	.0108	.0142
	Sample sizes: 5, 5, 5, 5, 25							
	Average proportion of distinct observations							
	1.0000		.8688		.8124		.5432	
.250	.2490	.2512	.2496	.2550	.2464	.2554	.2488	.2608
.100	.0968	.0998	.0980	.1022	.0966	.1024	.0942	.1076
.050	.0466	.0488	.0468	.0506	.0462	.0502	.0458	.0544
.025	.0206	.0222	.0198	.0224	.0208	.0232	.0196	.0266
.010	.0072	.0088	.0072	.0094	.0070	.0094	.0066	.0106

NOTE: The number of replications is 5,000.

It appears that the proposed tests maintain their levels quite well even for samples as small as $n_i = 5$. For small sample sizes the observed levels tend to be slightly conservative, that is, smaller than nominal, for extreme tail probabilities. Another simulation implementing the tests without the finite sample variance adjustment did not perform quite as well, although the results were good once the individual sample sizes reached 30. It is not clear whether A^2_{akN} has any clear advantage over A^2_{kN} as far as data rounding is concerned. At level .01 A^2_{akN} seems to perform better than A^2_{kN}, although that is somewhat offset at level .25.

7. AN EXAMPLE

As an example consider the paper smoothness data used by Lehmann (1975, p. 209, example 3; reproduced in Table 4) as an illustration of the Kruskal–Wallis test adjusted for ties. By use of this test the four sets of eight laboratory measurements show significant differences with P value $\approx$ .005.

Applying the two versions of the Anderson–Darling k-sample test to this set of data yields $A^2_{kN} = 8.3559$ and $A^2_{akN} = 8.3926$. Together with $\sigma_N = 1.2038$ this yields standardized T values of 4.449 and 4.480, respectively, which are outside the range of Table 1. Plotting the log-odds of α versus $t_3(\alpha)$, a strong linear pattern indicates that simple linear extrapolation should give good approximate P values. They are .0023 and .0022, respectively, somewhat smaller than the aforementioned .005 of the Kruskal–Wallis test.

Table 4. Four Sets of Eight Measurements Each of the Smoothness of a Certain Type of Paper, Obtained in Four Laboratories

Laboratory	Smoothness							
A	38.7	41.5	43.8	44.5	45.5	46.0	47.7	58.0
B	39.2	39.3	39.7	41.4	41.8	42.9	43.3	45.8
C	34.0	35.0	39.0	40.0	43.0	43.0	44.0	45.0
D	34.0	34.8	34.8	35.4	37.2	37.8	41.2	42.8

As a check, the simulation-based evaluation of the P values, described at the end of Section 3, yielded estimated P values of .00150 and .00155 for A^2_{kN} and A^2_{akN}, respectively. This simulation was based on 20,000 random permutations. The same kind of simulation yielded estimated P values of .00185 and .00165 for the Kruskal–Wallis and the permutation version of the analysis of variance F test, respectively. The simulation-based P values agree fairly well for these three types of test, and it appears that for the Kruskal–Wallis test the P value of .005, which was obtained by the chi-squared approximation, is not too accurate in this case. To confirm this, the simulation for the Kruskal–Wallis test P value was repeated with 10^6 random permutations, resulting in an estimated P value of .002092.

8. COMBINING INDEPENDENT ANDERSON–DARLING TESTS

Due to the convolution nature of the asymptotic distribution of the k-sample Anderson–Darling rank tests the following additional use of Table 1 is possible. If m independent one-sample Anderson–Darling tests of fit (Sec.

1) are performed for various hypotheses, then the joint statement, that all of these hypotheses are true together, may be tested by using the sum S of the m one-sample test statistics as the new test statistic and by comparing the appropriately standardized S with the row corresponding to m in Table 1. To standardize S note that the variance of a one-sample Anderson–Darling test based on n_i observations can either be computed directly or can be deduced from the variance formula (4) for $k = 2$ by letting the other sample size go to infinity as

$$\mathrm{var}(A^2_{n_i}) = 2(\pi^2 - 9)/3 + (10 - \pi^2)/n_i.$$

It seems quite natural to combine independent Anderson–Darling tests in this fashion, but it is not clear whether this procedure is optimal in any sense. It should be noted that these one-sample tests can only be combined this way if no unknown parameters are estimated. In that case different tables would be required. This problem is discussed further in Stephens (1986), where tables are given for combining tests of normality or of exponentiality with parameters estimated.

Similarly, independent k-sample Anderson–Darling tests can be combined. Here the value of k may change from one group of samples to the next, and the common distribution function may also be different from group to group. The objective in combining one-sample Anderson–Darling tests is typically not that of pooling data. In the case of combining m independent k-sample Anderson–Darling tests, however, one would naturally make a joint statement about the pooling of m groups of data sets into m separately pooled samples.

APPENDIX: ASYMPTOTIC DERIVATIONS

Asymptotic Distribution of A^2_{kN} Under H_0

The asymptotic distribution of (6) will be derived without any continuity assumptions on F. The asymptotic distribution of (3) for continuous F represents a special case. The result in the case of (7) will only be stated, with the proof following similar lines. In deriving the asymptotic distribution of (6) we combine the techniques of Kiefer (1959) and Pettitt (1976), with a slight shortening in the argument of the latter, and track the effect of a possibly discontinuous F.

Using the special construction of Pyke and Shorack (1968) (see also Shorack and Wellner 1986), we can assume that on a common probability space Ω there exist for each N, and corresponding $n_1, \ldots, n_k$, independent uniform samples $U_{isN} \sim U(0, 1)$ ($s = 1, \ldots, n_i$; $i = 1, \ldots, k$) and independent Brownian bridges $U_1, \ldots, U_k$ such that

$$\|U_{iN} - U_i\| \equiv \sup_{t\in[0,1]} |U_{iN}(t) - U_i(t)| \to 0$$

for every $\omega \in \Omega$ as $n_i \to \infty$. Here

$$U_{iN}(t) = n_i^{1/2}\{G_{iN}(t) - t\} \quad \text{with} \quad G_{iN}(t) = \frac{1}{n_i}\sum_{s=1}^{n_i} I_{[U_{isN}\le t]}$$

is the empirical process corresponding to the ith uniform sample. Let $X_{isN} \equiv F^{-1}(U_{isN})$ and

$$U_{iN}\{F(x)\} = n_i^{1/2}\{F_{iN}(x) - F(x)\} \quad \text{with} \quad F_{iN}(x) = \frac{1}{n_i}\sum_{s=1}^{n_i} I_{[X_{isN}\le x]}$$

so that $F_{iN}(x)$ and $F_{in_i}(x)$ have the same distribution. The empirical distribution function of the pooled sample of the X_{isN} is called $H_N(x)$ and that of the pooled uniform sample of the U_{isN} is called $K_N(t)$, so $H_N(x) = K_N\{F(x)\}$. This double use of H_N as empirical distribution of the X_{isN} and of the X_{is} should cause no confusion as long as only distributional conclusions concerning (6) are drawn.

Following Kiefer (1959), let $C = (c_{ij})$ denote a $k \times k$ orthonormal matrix with $c_{1j} = (n_j/N)^{1/2}$ ($j = 1, \ldots, k$). If $U = (U_1, \ldots, U_k)'$, then the components of $V = (V_1, \ldots, V_k)' = CU$ are again independent Brownian bridges. Further, if $U_N = (U_{1N}, \ldots, U_{kN})'$ and $V_N = (V_{1N}, \ldots, V_{kN})' = CU_N$, then $\|V_{iN} - V_i\| \to 0$ for all $\omega \in \Omega$ ($i = 1, \ldots, k$) and

$$\sum_{i=1}^k n_i\{F_{iN}(x) - H_N(x)\}^2 = \sum_{i=1}^k U^2_{iN}\{F(x)\} - V^2_{1N}\{F(x)\} = \sum_{i=2}^k V^2_{iN}\{F(x)\}$$

for all $x \in R$.

This suggests that A^2_{kN}, which is equal in distribution to

$$\int_{B_N} \frac{\sum_{i=2}^k V^2_{iN}\{F(x)\}}{H_N(x)\{1 - H_N(x)\}}\, dH_N(x) = \int_{A_N} \frac{\sum_{i=2}^k V^2_{iN}\{\psi(u)\}}{K_N\{\psi(u)\}[1 - K_N\{\psi(u)\}]}\, dK_N(u),$$

converges in distribution to

$$A^2_{k-1} \equiv \int_A \frac{\sum_{i=2}^k V_i^2\{\psi(u)\}}{\psi(u)\{1 - \psi(u)\}}\, du$$

as $q = \min(n_1, \ldots, n_k) \to \infty$. Here $A_N = \{u \in [0, 1]: K_N(\psi(u)) < 1\}$, $A = \{u \in [0, 1]: \psi(u) < 1\}$, and $\psi(u) = F\{F^{-1}(u)\}$.

The argument for this convergence can be made rigorous by splitting the integral into a central portion and the remainder. The convergence of the central portion follows from theorem 5.2 of Billingsley (1968) and the remainder is shown to be negligible by a combination of Markov's inequality [not Chebychev's as in Pettitt (1976)] and theorem 4.2 of Billingsley (1968).

Similarly, one can show that under H_0 the modified version A^2_{akN} converges in distribution to

$$A^2_{a(k-1)} \equiv \int_0^1 \frac{\sum_{i=2}^k [V_i\{\psi(u)\} + V_i\{\psi_-(u)\}]^2}{4\overline{\psi}(u)\{1 - \overline{\psi}(u)\} - \{\psi(u) - \psi_-(u)\}}\, du,$$

where $\psi_-(u) = F\{F^{-1}(u)-\}$ and $\overline{\psi}(u) = \{\psi(u) + \psi_-(u)\}/2$ and $V_2, \ldots, V_k$ are the same independent Brownian bridges as before.

When F is continuous the distributions of A^2_{k-1} and $A^2_{a(k-1)}$ coincide (Shorack and Wellner 1986, p. 225) with that of

$$\sum_{i=1}^k \sum_{j=1}^\infty \frac{1}{j(j+1)} Q^2_{ij} = \sum_{j=1}^\infty \frac{1}{j(j+1)} Y_j,$$

where the Q_{ij} are independent standard normal random variables and the Y_j are independent chi-squared random variables with $k - 1$ df.

Consistency of A^2_{kN}

To show consistency of the test based on A^2_{kN}, it is sufficient to show that $A^2_{kN} \to \infty$ (a.s.) as $\min(n_1, \ldots, n_k) \to \infty$. Since

$$A_{kN} \ge 4 \sum_{i=1}^k n_i \int \{F_{in_i}(x) - H_N(x)\}^2\, dH_N(x) = 4W^2_{kN},$$

the consistency of the Cramer–von Mises statistic W^2_{kN} implies that of A^2_{kN}.

Assuming that $n_i/N \to \lambda_i > 0$ ($i = 1, \ldots, k$) as $N \to \infty$ and setting

$$\overline{F}(x) = \sum_{i=1}^{k} \lambda_i F_i(x),$$

it follows from the Glivenko–Cantelli theorem that

$$\int \{F_{in_i}(x) - H_N(x)\}^2 - \{F_i(x) - \overline{F}(x)\}^2 \, dH_N(x) \to 0 \text{ (a.s.)}$$

and by the law of large numbers that

$$\begin{aligned} \int \{F_i(x) - \overline{F}(x)\}^2 \, dH_N(x) &= \sum_{i=1}^{k} \frac{n_i}{N} \frac{1}{n_i} \sum_{j=1}^{n_i} \{F_i(X_{ij}) - \overline{F}(X_{ij})\}^2 \\ &\to \sum_{i=1}^{k} \lambda_i \int \{F_i(x) - \overline{F}(x)\}^2 \, dF_i(x) \\ &= \int \{F_i(x) - \overline{F}(x)\}^2 \, d\overline{F}(x) \text{ (a.s.).} \end{aligned}$$

Combining these two limits we have

$$\sum_{i=1}^{k} \frac{n_i}{N} \int \{F_{in_i}(x) - H_N(x)\}^2 \, dH_N(x) \to D \equiv \sum_{i=1}^{k} \lambda_i \int \{F_i(x) - \overline{F}(x)\}^2 \, d\overline{F}(x)$$

and positivity of the discrepancy measure D entails consistency of W^2_{kN} and A^2_{kN}. This shows the consistency against all alternatives to H_0 provided that $\lambda_i > 0$ ($i = 1, \ldots, k$).

[Received June 1986. Revised February 1987.]

REFERENCES

Anderson, T. W., and Darling, D. A. (1952), "Asymptotic Theory of Certain Goodness of Fit Criteria Based on Stochastic Processes," *Annals of Mathematical Statistics*, 23, 193–212.

——— (1954), "A Test of Goodness of Fit," *Journal of the American Statistical Association*, 49, 765–769.

Bell, C. B., Blackwell, D., and Breiman, L. (1960), "On the Completeness of Order Statistics," *Annals of Mathematical Statistics*, 31, 794–797.

Billingsley, P. (1968), *Convergence of Probability Measures*, New York: John Wiley.

Conover, W. J. (1980), *Practical Nonparametric Statistics* (2nd ed.), New York: John Wiley.

Darling, D. A. (1957), "The Kolmogorov–Smirnov, Cramer–von Mises Tests," *Annals of Mathematical Statistics*, 28, 823–838.

Hájek, J., and Šidak, Z. (1967), *Theory of Rank Tests*, New York: Academic Press.

Hollander, M., and Wolfe, D. A. (1973), *Nonparametric Statistical Methods*, New York: John Wiley.

Kiefer, J. (1959), "*K*-Sample Analogues of the Kolmogorov–Smirnov and Cramer–v. Mises Tests," *Annals of Mathematical Statistics*, 30, 420–447.

Lehmann, E. L. (1975), *Nonparametrics, Statistical Methods Based on Ranks*, San Francisco: Holden-Day.

——— (1983), *Theory of Point Estimation*, New York: John Wiley.

MACSYMA (1984), Symbolics Inc., Eleven Cambridge Center, Cambridge, MA 02142.

Pettitt, A. N. (1976), "A Two-Sample Anderson–Darling Rank Statistic," *Biometrika*, 63, 161–168.

Pyke, R., and Shorack, G. R. (1968), "Weak Convergence of a Two-Sample Empirical Process and a New Approach to Chernoff–Savage Theorems," *Annals of Mathematical Statistics*, 39, 755–771.

Schrage, L. (1979), "A More Portable Fortran Random Number Generator," *ACM Transactions on Mathematical Software*, 5, 132–138.

Shorack, G. R., and Wellner, J. A. (1986), *Empirical Processes With Applications to Statistics*, New York: John Wiley.

Solomon, H., and Stephens, M. A. (1978), "Approximations to Density Functions Using Pearson Curves," *Journal of the American Statistical Association*, 73, 153–160.

Stephens, M. A. (1974), "EDF Statistics for Goodness of Fit and Some Comparisons," *Journal of the American Statistical Association*, 69, 730–737.

——— (1976), "Asymptotic Results for Goodness-of-Fit Statistics With Unknown Parameters," *The Annals of Statistics*, 4, 357–369.

——— (1986), "Tests Based on EDF Statistics," in *Goodness-of-Fit Techniques*, ed. R. B. d'Agostino and M. A. Stephens, New York: Marcel Dekker, pp. 97–193.

Watson, G. S. (1962), "Goodness-of-Fit Tests on a Circle II," *Biometrika*, 49, 57–63.