K-SAMPLE ANDERSON-DARLING TESTS OF FIT, FOR CONTINUOUS AND DISCRETE CASES

by

F.W. SCHOLZ
M.A. STEPHENS

TECHNICAL REPORT No. 81
May 1986

Department of Statistics, GN-22
University of Washington
Seattle, Washington 98155 USA

K-Sample Anderson-Darling Tests of Fit, for Continuous and Discrete Cases

F.W. Scholz[1] and M.A.Stephens[2]

[1]Boeing Computer Services, MS 9C-01,
Seattle WA 98124-0346, P.O. Box 24346, USA

[2]Department of Mathematics and Statistics,
Simon Fraser University, Burnaby, BC, V5A 1S6 Canada

Abstract

Two k-sample versions of the Anderson-Darling (AD) test of fit are proposed and 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 1-sample AD asymptotic distribution. Monte Carlo simulation is used to investigate the null distribution small sample behavior of the two versions under various degrees of data rounding and sample size imbalances. Tables for carrying out these tests are provided and their usage in combining independent 1- or k-sample AD-tests is pointed out.

Some key words: Combining tests, Convolution, Empirical processes, Midranks, Pearson Curves, Simulation.

May 20, 1986[1]

[1]re-typeset with minor corrections March 26, 2008
F.W. Scholz, Affiliate Professor, Department of Statistics, University of Washington, Seattle, WA 98105. and M.A. Stephens, Professor Emeritus, Department of Statistics and Actuarial Science, Simon Fraser University, 8888 University Drive, Burnaby, BC, V5A 1S6 Canada

K-Sample Anderson-Darling Tests of Fit, for Continuous and Discrete Cases

F.W. Scholz[1] and M.A.Stephens[2]

[1]Boeing Computer Services, MS 9C-01,
Seattle WA 98124-0346, P.O. Box 24346, USA

[2]Department of Mathematics and Statistics,
Simon Fraser University, Burnaby, BC, V5A 1S6 Canada

1 Introduction and Summary

Anderson and Darling (1952, 1954) introduce 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 distribution function $F(x)$ where $F(x) = F_0(x)$ for some completely specified distribution function $F_0(x)$. Here $F_m(x)$ is defined as the proportion of the sample $X_1, \ldots, X_m$ which is not greater than x. The corresponding two-sample version

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

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.

In Sections 2 and 3 two k-sample versions of the Anderson-Darling test are proposed for the continuous as well as discrete case and computational formulae are given. Section 4 discusses the finite sample distribution of these statistics and gives a variance formula for one of the two statistics. Section 5 derives the asymptotic null distribution of both versions which in the continuous case is the $(k-1)$-fold convolution of the

1-sample Anderson-Darling asymptotic distribution. Section 6 describes the proposed test procedures and gives a table for carrying out the tests. Section 7 reports the results of a Monte-Carlo simulation testing the adequacy of the table for the continuous and discrete case. Section 8 presents two examples and Section 9 points out another use of the above table for combining independent 1- and k-sample Anderson-Darling tests of fit.

2 The K-Sample Anderson-Darling Test

On the surface 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 analogue of the Cramer-v. Mises test shows the appropriate path. To set the stage the following notation is introduced. Let X_{ij} be the j^{th} observation in the i^{th} sample, $j = 1, \ldots, n_i, i = 1, \ldots, k$. All observations are independent. Suppose the i^{th} sample has distribution function F_i. We wish to test the hypothesis

$$H_0 : F_1 = \ldots = F_k$$

without specifying the common distribution F. Since rounding of data occurs routinely in practice we will not necessarily assume that the F_i, and hence the common F under H_0, are continuous. Denote the empirical distribution function of the i^{th} sample by $F_{n_i}(x)$ and that of the pooled sample of all $N = n_1 + \ldots + n_k$ observations by $H_N(x)$. The k-sample Anderson-Darling test statistic is then defined as

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

where $B_N = \{x \in R : H_N(x) < 1\}$. For $k = 2$ (2) reduces to (1). In the case of untied observations, i.e., the pooled ordered sample is $Z_1 < \ldots < Z_N$, a computational formula for A^2_{kN} is

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

where M_{ij} is the number of observations in the i^{th} sample which are not greater than Z_j.

3 Discrete Parent Population

If continuous data is grouped, or of the parent populations are discrete, tied observations can occur. To give the computational formula in the case of tied observations we introduce the following notation. Let $Z_1^\star < \ldots < Z_L^\star$ denote the $L(>1)$ distinct ordered observations in the pooled sample. Further let f_{ij} be the number of observations in the i^{th} sample coinciding with $Z_j^\star$ and let $\ell_j = \sum_{i=1}^k f_{ij}$ denote the multiplicity of $Z_j^\star$. Using (2) as the definition of A_{kN}^2 the computing formula in the case of ties becomes

$$A_{kN}^2 = \sum_{i=1}^{k} \frac{1}{n_i} \sum_{j=1}^{L-1} \frac{\ell_j}{N} \frac{(NM_{ij} - n_i B_j)^2}{B_j(N - B_j)} \,, \tag{4}$$

where $M_{ij} = f_{i1} + \ldots + f_{ij}$ and $B_j = \ell_1 + \ldots + \ell_j$.

An alternative 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, i.e.,

$$F_{an_i}(x) := \frac{1}{2}\left(F_{n_i}(x) + F_{n_i}(x-)\right)$$

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

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

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

$$A_{akN}^2 = \frac{N-1}{N} \sum_{i=1}^{k} \frac{1}{n_i} \sum_{j=1}^{L} \frac{\ell_j}{N} \frac{(NM_{aij} - n_i B_{aj})^2}{B_{aj}(N - B_{aj}) - N\ell_j/4} \,, \tag{5}$$

where $M_{aij} = f_{i1} + \ldots + f_{ij-1} + f_{ij}/2$ and $B_{aj} = \ell_1 + \ldots + \ell_{j-1} + \ell_j/2$. The formula (5) is valid provided not all the X_{ij} are the same. This latter way of dealing with ties corresponds to the treatment of ties through midranks in the case of the Wilcoxon two-sample and the Kruskal-Wallis k-sample tests, see Lehmann (1975).

4 The Finite Sample Distribution under H_0

Under H_0 the expected value of A^2_{kN} is

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

where $\psi(u) := F(F^{-1}(u))$ with $\psi(u) \geq u$ and $\psi(u) \equiv u$ if and only if F is continuous, see Section 5 for some details. In the continuous case the expected value becomes $k-1$. 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 expected value of A^2_{akN} under H_0 is

$$E\left(A^2_{akN}\right) = (k-1)\left\{1 - P(X_{11} = \ldots = X_{kn_k})\right\}$$

which is $k-1$ for continuous F and otherwise becomes $k-1$ in the nondegenerate case as $N \to \infty$.
Higher moments of A^2_{kN} and A^2_{akN} are very difficult to compute. Pettitt (1976) gives an approximate variance formula for A^2_{2N} as $\mathrm{var}(A^2_{2N}) \approx \sigma^2(1 - 3.1/N)$ where $\sigma^2 = 2(\pi^2 - 9)/3$ is the variance of A^2_1. This approximation does not account for any dependence on the individual sample sizes. Below a general variance formula of A^2_{kN} is given for the continuous case.

$$\sigma^2_N := \mathrm{var}(A^2_{kN}) = \frac{aN^3 + bN^2 + cN + d}{(N-1)(N-2)(N-3)}\ , \tag{6}$$

with

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

where

$$H = \sum_{i=1}^{n} \frac{1}{n_i}\ , \quad h = \sum_{i=1}^{N-1} \frac{1}{i} \quad \text{and} \quad g = \sum_{i=1}^{N-2} \sum_{j=i+1}^{N-1} \frac{1}{(N-i)j}\ .$$

Note that

$$g \longrightarrow \int_0^1 \int_y^1 \frac{1}{x(1-y)} dx\, dy = \frac{\pi^2}{6}$$

as $N \to \infty$ and thus $\mathrm{var}(A^2_{kN}) \to (k-1)\sigma^2$ 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 $1/N$. A variance formula for A^2_{akN} was not derived. However, simulations showed that the variances of A^2_{kN} and A^2_{akN} are very close to each other in the continuous case. Here closeness was judged by the discrepancy between the simulated variance of A^2_{akN} and that obtained by (6).

In principle it is possible to derive the conditional null distribution (under H_0) of (3), (4) or (5) given the pooled (ordered) vector $Z = (Z_1, \ldots, Z_N)$ of observations $Z_1 \leq \ldots \leq Z_N$ by recording the distribution of (3), (4) or (5) as one traverses through all possible ways of splitting Z into k samples of sizes $n_1, \ldots, n_k$ respectively. Thus the test is truly non-parametric in this sense. For small sample sizes it may be feasible to derive this distribution and tables could be constructed. However, the computational and tabulation effort quickly grows prohibitive as k gets larger. Not only will the null distribution be required for all possible combinations $(n_1, \ldots, n_k)$ but also for all combinations of ties.

A more pragmatic approach would be to record the relative frequency $\hat{p}$ with which the observed value a^2 of (3), (4) or (5) is matched or exceeded when computing (3), (4) or (5) for a large number Q of random partitions of Z into k samples of sizes $n_1, \ldots, n_k$ respectively. This was done, for example, to get the distribution of the two-sample Watson statistic U^2_{nm} in Watson (1962). This bootstrap like method is applicable equally well in small and large samples. $\hat{p}$ is an unbiased estimator of the true conditional as well as unconditional P-value of a^2, and the variance of $\hat{p}$ can be controlled by the choice of Q. In the next section the unconditional asymptotic distribution of (3), (4) or (5) will be derived.

5 Asymptotic Distribution of A^2_{kN} under H_0

Since the asymptotic distribution of (4) reduces to that of (3) in the case that the common distribution function F is assumed continuous, only the case of (4) will be treated in detail and the result in the case of (5) will only be stated with proof following similar lines. In deriving the asymptotic distribution of (4) 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 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, corresponding to $n_1, \ldots, n_k$, independent uniform smaples $U_{isN} \sim U(0,1), s =$

$1, \ldots, n_i,\ i = 1, \ldots, k$ and independent Brownian bridges $U_1, \ldots, U_k$ such that

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

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

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

is the empirical process corresponding to the i^{th} uniform sample. Let $X_{isN} := F^{-1}(U_{isN})$ and

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

so that $F_{iN}(x)$ is equal in distribution to $F_{n_i}(x)$. The empirical distribution function of the pooled sample of the X_{isN} is also denoted by $H_N(x)$ and that of the pooled uniform sample of the U_{isN} is denoted by $K_N(t)$ so that $H_N(x) = K_N(F(x))$. 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 (4) are drawn.

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

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

for all $x \in R$.

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

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

converges in distribution to

$$A_{k-1}^2 := \int_A \frac{\sum_{i=2}^k V_i^2\{\psi(u)\}}{\psi(u)\{1 - \psi(u)\}} du$$

as $M := \min(n_1, \ldots, n_k) \to \infty$. Here $A_N = \{u \in [0,1] : K_N(\psi(u)) < 1\}$ and $A = \{u \in [0,1] : \psi(u) < 1\}$. To make this rigorous we follow Pettitt and claim that for each $\delta \in (0, 1/2)$ and $S(\delta) = \{u \in [0,1] : \delta \leq u, \psi(u) \leq 1 - \delta\}$ we have (see Billingsley, theorem 5.2)

$$\int_{A_N \cap S(\delta)} \frac{\sum_{i=2}^k V_{iN}^2\{\psi(u)\}}{K_N\{\psi(u)\}[1 - K_N\{\psi(u)\}]} dK_N(u) - \int_{S(\delta)} \frac{\sum_{i=2}^k V_i^2\{\psi(u)\}}{\psi(u)\{1 - \psi(u)\}} dK_N(u)$$

$$+ \int_{S(\delta)} \frac{\sum_{i=2}^k V_i^2\{\psi(u)\}}{\psi(u)\{1 - \psi(u)\}} d\{K_N(u) - u\} \longrightarrow 0$$

for all $\omega \in \Omega$ as $M \to \infty$.

If $W = (W_1, \ldots, W_N)$ with $W_1 < \ldots < W_N$ denote the order statistics of the pooled sample of the U_{isN} then the conditional expectation of

$$\sum_{i=2}^k V_{iN}^2\{\psi(W_j)\} = \sum_{i=1}^k n_i \left[G_{iN}\{\psi(W_j)\} - K_N\{\psi(W_j)\}\right]^2$$

given W is $K_N\{\psi(W_j)\}\left[1 - K_N\{\psi(W_j)\}\right](k-1)N/(N-1)$. Thus the unconditional expectation of

$$D_{1N} = \int I_{[0,\delta)}(u) I_{A_N}(u) \frac{\sum_{i=2}^k V_{iN}^2\{\psi(u)\}}{K_N\{\psi(u)\}[1 - K_N\{\psi(u)\}]} dK_N(u)$$

is

$$\begin{aligned} E(D_{1N}) &= \sum_{j=1}^N \frac{(k-1)}{(N-1)} P[W_j < \delta, K_N\{\psi(W_j)\} < 1] \\ &= \frac{(k-1)N}{N-1} P[U_{11N} < \delta, K_N\{\psi(U_{11N})\} < 1] \leq \delta \frac{(k-1)N}{N-1} \ . \end{aligned}$$

Similarly the unconditional expectation of

$$D_{2N} = \int I_{[\psi(u) > 1-\delta]}(u) I_{A_N}(u) \frac{\sum_{i=2}^k V_{iN}^2\{\psi(u)\}}{K_N\{\psi(u)\}[1 - K_N\{\psi(u)\}]} dK_N(u)$$

is

$$E(D_{2N}) = \frac{(k-1)N}{N-1} \, P[\psi(U_{11N}) > 1 - \delta, K_N\{\psi(U_{11N})\} < 1] \ .$$

If $\psi(t) = 1$ for some $t < 1$ then $E(D_{2N}) = 0$ for δ sufficiently small, and if $\psi(t) < 1$ for all $t < 1$ then

$$E(D_{2N}) \leq \frac{(k-1)N}{N-1} \left\{1 - \psi^{-1}(1-\delta)\right\} \ .$$

In either case $E(D_{1N}+D_{2N}) \to 0$ as $\delta \to 0$, uniformly in N. Thus by Markov's inequality $P(D_{1N}+D_{2N} \geq \epsilon) \to 0$ as $M \to \infty$ and then $\delta \to 0$. Theorem 4.2 of Billingsley (1968) and the fact that for all $\omega \in \Omega$

$$A^2_{k-1}(\delta) = \int_{S(\delta)} \frac{\sum_{i=2}^k V_i^2\{\psi(u)\}}{\psi(u)\{1-\psi(u)\}}\, du \longrightarrow A^2_{k-1}$$

as $\delta \to 0$ (monotone convergence theorem) proves the claim that A^2_{kN} converges in distribution to A^2_{k-1}. The integral defining A^2_{k-1} exists and is finite for almost all $\omega \in \Omega$ by Fubini's theorem upon taking expectation of A^2_{k-1}.
Similarly one can show that under H_0 the modified version A^2_{akN} converges in distribution to

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

where $\psi_-(u) = F(F^{-1}(u)-)$ and $\bar{\psi}(u) = (\psi(u)+\psi_-(u))/2$ and $V_2,\ldots,V_k$ are the same Brownian bridges as before. Note that A^2_{k-1} and $A^2_{a(k-1)}$ coincide when F is continuous. Thus the limiting distribution in the continuous case can be considered an approximation to the limiting distributions of A^2_{kN} and A^2_{akN} under rounding of data provided 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. The fact that $\bar{\psi}$ approximates the diagonal better than ψ appears to point to A^2_{akN} as the better approximand. Only simulation can bear this out.

6 Table of Critical Points

Since the 1- and 2-sample Anderson-Darling tests of fit the use of asymptotic percentiles works very well even in small samples, Stephens (1974) and Pettitt (1976), the use of the asymptotic percentiles is suggested here as well. To obtain s somewhat better accuracy in the approximation we follow Pettitt (1976) and reject H_0 at significance level α whenever

$$\frac{A^2_{kN} - (k-1)}{\sigma_N} \geq z_{k-1}(1-\alpha)$$

where $z_{k-1}(1-\alpha)$ is the $(1-\alpha)$-percentile of the standardized asymptotic $Z_{k-1} = \{A^2_{k-1} - (k-1)\}/\sigma$ distribution and σ_N is given by (6).

If $Y_1, Y_2, \ldots$ are independent chi-square random variables with $k-1$ degrees of freedom then A^2_{k-1} has the same distribution as

$$\sum_{i=1}^{\infty} \frac{1}{i(i+1)} Y_i \ .$$

It cumulants and first four moments are easily calculated and approximate percentiles of this distribution were obtained by fitting Pearson curves as in Stephens (1976) and Solomon and Stephens (1978). This approximation works very well in the case $k-1=1$ and can be expected to improve as k increases. A limited number of standardized percentiles $z_m(1-\alpha)$ of Z_m are given in Table 1. The test using A^2_{akN} is carried out the same way by just replacing A^2_{kN} with A^2_{akN} above.

Table 1
Percentiles $z_m(\gamma)$ of the Z_m-Distribution

	γ				
m	.75	.90	.95	.975	.99
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

For values of m 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 relative error. The general form of the interpolation formula is

$$z_m(\gamma) = b_0 + \frac{b_1}{\sqrt{m}} + \frac{b_2}{m} \ ,$$

where the coefficients for each γ may be found in Table 2. Similarly one could interpolate and even extrapolate in Table 1 with respect to γ in order to establish an approximate P-value for the observed Anderson-Darling statistic, see Section 8 for examples.

Table 2			
Interpolation Coefficients			
γ	b_0	b_1	b_2
.75	.675	-.245	-.105
.90	1.281	.250	-.305
.95	1.645	.678	-3.62
.975	1.960	1,149	-.391
.99	2.326	1.822	-.396

7 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. 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. The results of the simulations are summarized in Tables 3-17 of the Appendix. For each of these tables 5000 pooled samples were generated. Each pooled sample was then broken down into the indicated number of subsamples with the given sample sizes. The observed false alarm rates are recorded in columns 2 and 3 for the two versions of the statistic. Next, for each pooled sample created above the scale was changed to $a = 150, a = 100$ and $a = 30$ and the sample values were rounded to the nearest integer. The observed false alarm rates are given respectively in columns (4,5), (6,7) and (8,9). On top of these columns the degree of rounding is expressed in terms of the average proportion of distinct observations in the pooled sample.

It appears that the proposed tests maintain their levels quite well even for samples as small as $n_i = 5$. 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.

The power behavior of these tests has not been studied but one can expect that the good behavior of the 2-sample test A^2_{mn}, as demonstrated by Pettitt (1976), carries over to the k-sample case as well.

8 Two Examples

As a first example consider the paper smoothness data used by Lehmann (1968, p. 209, Example 3 and reproduced in Table 18 of the Appendix) as an illustration of the Kruskal-Wallis test adjusted for ties. According to 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 Z-scores of 4.449 and 4.480 respectively, which are outside the range of Table 1. Plotting the log-odds of γ versus $z_3(\gamma)$ a strong linear pattern indicates that simple linear extrapolation should give good approximate P-values. They are .0023 and .0022 respectively, somewhat smaller than that of the Kruskal-Wallis test.

As a second example consider the air conditioning failure data analyzed by Proschan (1963) who showed that the data sets are significantly different (P-value $\approx .007$) under the assumption that the data sets are individually exponentially distributed. The data consists of operating hours between failures of the air conditioning system on a fleet of Boeing 720 jet airplanes. For some of the airplanes the sequence of intervals between failures is interrupted by a major overhaul. For this reason segments of data from separate airplanes, which are not separated by a major overhaul, are treated as separate samples. Also, only segments of length at least 5 are considered. These are reproduced in Table 19 of the Appendix.

Applying the Anderson-Darling tests to these 14 data sets yields $A^2_{kN} = 21.6948$ and $A^2_{akN} = 21.7116$. Together with $\sigma_N = 2.6448$ this yields standardized Z-scores of 3.288 and 3.294 respectively, which are outside the range of Table 1. Using Table 2, the interpolation formula for $z_m(\gamma)$ yields appropriate percentiles for $m = 13$. Plotting the log-odds of γ versus $z_{13}(\gamma)$ suggests that a cubic extrapolation should provide good approximate P-values. These are .0042 and .0043 respectively. Hence the evidence against homogeneity appears stronger here even without the exponentiality assumption.

9 Combining Independent Anderson-Darling Tests

Due to the convolution nature of the asymptotic distribution of the k-sample Anderson-Darling tests of fit the following additional use of Table 1 is possible. If m independent 1-sample Anderson-Darling tests of fit, see Section 1, are performed for various hypotheses then the joint statement of these hypotheses may be tested by using the sum S of the m 1-sample test statistics as the new test statistic and by comparing the appropriately standardized S against the row corresponding to m in Table 1. To standardize S note that the variance of a 1-sample Anderson-Darling test based on n_i observations can either be computed directly or can be deduced from the variance formula (6) for $k = 2$ by letting the other sample size go to infinity as:

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

It should be noted that these 1-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 by Stephens (1986), where tables are given for combining tests of normality or of exponentiality.

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.

Acknowledgements

We would like to thank Steve Rust for drawing our attention to this problem and Galen Shorack and Jon Wellner for several helpful conversations and the use of galleys of their book. The variance formula (6) 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, and by the U.S. Office of Naval Research.

References

Anderson, T.W. and Darling, D.A. (1952), Asymptotic theory of certain ”goodness of fit” criteria based on stochastic processes. *Ann. Math. Statist.*, 23,193-212

Anderson, T.W. and Darling, D.A. (1954), A test of goodness of fit, *J. Amer. Stat. Assn.*, 49, 765-769

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

Darling, D.A. (1957), The Kolmogorov-Smirnov, Cramer-yon Mises Tests, *Ann. Math. Statist.*, 28, 823-838

Kiefer, I. (1959), K-sample analogues of the Kolmogorov-Smirnov and Cramer-v. Mises tests, *Ann. Math. Statist.*, 30, 420-447

Lehmann, E.L. (1968), *Nonparametrics, Statistical Methods Based on Ranks*, Holden Day; San Francisco

MACSYMA (1984), *Symbolics Inc.*, Eleven Cambridge Center, Cambridge, MA.O2142

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, *Ann. Math. Statist.*, 39, 755-771

Schrage, L. (1979), A more portable Fortran random number generator, *ACM Trans. Math. Software*, 5,132-138

Shorack, G.R. and Wellner I.A. (1986), *Empirical Processes with Applications to Statistics, Wiley*; New York

Solomon, H. and Stephens, M.A. (1978), Approximations to Density Functions Using Pearson Curves, *J. Amer. Stat. Assoc.*, 73, 153-160

Stephens, M.A. (1974), EDF statistics for goodness of fit and some comparisons *J. Amer. Stat. Assoc.*, 69, 730-737

Stephens, M.A. (1976), Asymptotic Results for Goodness-of-Fit Statistics with Unknown Parameters, *Ann. Statist.*, Vol. 4, 2, 357-369

Stephens, M.A. (1986), Tests based on EDF Statistics, Chap. 4 of *Goodness-of-Fit Techniques* (ed. R.B. d'Agostino and M.A. Stephens), Marcel Dekker; New York.

Watson, G.S. (1962), Goodness-of-fit tests on a circle II, *Biometrika*, 49, 57-63

Appendix

Results of the Monte Carlo Simulations

Table 3 Observed Significance Levels of A^2_{kN} and A^2_{akN}
Number of Replications = 5000
Sample Sizes 5 5 5

nominal significance level α	average proportion of distinct observations 1.0000		.9555		.9346		.8034	
	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a
.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

Table 4 Observed Significance Levels of A^2_{kN} and A^2_{akN}
Number of Replications = 5000
Sample Sizes 10 10 10

nominal significance level α	average proportion of distinct observations 1.0000		.9099		.8700		.6520	
	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a
.250	.2416	.2438	.2418	.2464	.2438	.2482	.2432	.2534
.100	.1044	.1066	.1040	.1086	.1050	.1078	.1026	.1126
.050	.0496	.0524	.0502	.0542	.0490	.0530	.0512	.0588
.025	.0238	.0244	.0240	.0256	.0242	.0254	.0236	.0314
.010	.0102	.0108	.0102	.0112	.0100	.0114	.0110	.0128

Table 5 Observed Significance Levels of A^2_{kN} and A^2_{akN}
Number of Replications = 5000
Sample Sizes 30 30 30

nominal significance level α	average proportion of distinct observations 1.0000		.7594		.6719		.3552	
	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a
.250	.2504	.2508	.2488	.2532	.2502	.2556	.2478	.2572
.100	.0964	.0972	.0946	.0978	.0950	.0998	.0956	.1048
.050	.0466	.0474	.0474	.0482	.0480	.0500	.0462	.0532
.025	.0228	.0236	.0238	.0244	.0242	.0252	.0228	.0268
.010	.0092	.0092	.0094	.0096	.0090	.0094	.0086	.0104

Table 6 Observed Significance Levels of A^2_{kN} and A^2_{akN}
Number of Replications = 5000
Sample Sizes 5 5 5 5 5

nominal significance level α	average proportion of distinct observations 1.0000		.9249		.8904		.6971	
	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a
.250	.2576	.2618	.2580	.2636	.2600	.2646	.2568	.2704
.100	.1068	.1100	.1062	.1130	.1044	.1128	.1036	.1210
.050	.0498	.0528	.0492	.0544	.0490	.0550	.0476	.0592
.025	.0202	.0226	.0188	.0234	.0196	.0232	.0214	.0280
.010	.0068	.0074	.0064	.0074	.0070	.0078	.0060	.0094

Table 7 Observed Significance Levels of A^2_{kN} and A^2_{akN}
Number of Replications = 5000
Sample Sizes 10 10 10 10 10

nominal significance level α	average proportion of distinct observations 1.0000		.8551		.7951		.5134	
	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a
.250	.2502	.2526	.2524	.2548	.2518	.2562	.2522	.2604
.100	.1028	.1042	.1034	.1070	.1022	.1066	.1018	.1114
.050	.0492	.0512	.0498	.0524	.0488	.0528	.0478	.0566
.025	.0246	.0268	.0250	.0272	.0250	.0284	.0250	.0302
.010	.0080	.0088	.0080	.0086	.0076	.0086	.0092	.0100

Table 8 Observed Significance Levels of A^2_{kN} and A^2_{akN}
Number of Replications = 5000
Sample Sizes 30 30 30 30 30

nominal significance level α	average proportion of distinct observations 1.0000		.6461		.5402		.2417	
	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a
.250	.2524	.2532	.2516	.2542	.2526	.2552	.2522	.2596
.100	.1036	.1042	.1022	.1056	.1034	.1068	.1028	.1138
.050	.0520	.0522	.0518	.0532	.0514	.0550	.0512	.0588
.025	.0256	.0260	.0266	.0266	.0262	.0266	.0264	.0306
.010	.0114	.0114	.0116	.0120	.0112	.0122	.0108	.0136

Table 9 Observed Significance Levels of A^2_{kN} and A^2_{akN}
Number of Replications = 5000
Sample Sizes 5 5 5 5 5 5 5 5 5

nominal significance level α	average proportion of distinct observations 1.0000		.8678		.8126		.5438	
	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a
.250	.2610	.2650	.2618	.2674	.2616	.2692	.2584	.2722
.100	.1108	.1128	.1114	.1142	.1096	.1150	.1056	.1234
.050	.0470	.0486	.0460	.0504	.0462	.0510	.0470	.0562
.025	.0206	.0226	.0204	.0242	.0198	.0246	.0204	.0290
.010	.0064	.0070	.0064	.0076	.0068	.0078	.0058	.0080

Table 10 Observed Significance Levels of A^2_{kN} and A^2_{akN}
Number of Replications = 5000
Sample Sizes 10 10 10 10 10 10 10 10 10

nominal significance level α	average proportion of distinct observations 1.0000		.7598		.6721		.3551	
	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a
.250	.2516	.2532	.2498	.2542	.2512	.2590	.2464	.2600
.100	.0944	.0952	.0950	.0966	.0958	.0976	.0964	.1024
.050	.0494	.0500	.0498	.0514	.0496	.0528	.0492	.0580
.025	.0256	.0258	.0252	.0268	.0248	.0268	.0256	.0306
.010	.0122	.0122	.0118	.0128	.0118	.0128	.0118	.0136

Table 11 Observed Significance Levels of A^2_{kN} and A^2_{akN}
Number of Replications = 5000
Sample Sizes 30 30 30 30 30 30 30 30 30

nominal significance level α	average proportion of distinct observations 1.0000		.4900		.3816		.1486	
	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a
.250	.2494	.2498	.2522	.2534	.2520	.2542	.2512	.2612
.100	.0956	.0958	.0964	.0978	.0956	.0994	.0968	.1058
.050	.0454	.0456	.0450	.0462	.0448	.0468	.0438	.0504
.025	.0222	.0224	.0220	.0230	.0224	.0230	.0214	.0260
.010	.0076	.0076	.0078	.0080	.0076	.0078	.0078	.0098

Table 12 Observed Significance Levels of A^2_{kN} and A^2_{akN}
Number of Replications = 5000
Sample Sizes 5 10 15 20 25

nominal significance level α	average proportion of distinct observations 1.0000		.7938		.7152		.4035	
	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a
.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

Table 13 Observed Significance Levels of A^2_{kN} and A^2_{akN}
Number of Replications = 5000
Sample Sizes 5 5 5 5 25

nominal significance level α	average proportion of distinct observations 1.0000		.8688		.8124		.5432	
	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a
.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

Table 14 Observed Significance Levels of A^2_{kN} and A^2_{akN}
Number of Replications = 5000
Sample Sizes 5 25 25 25 25

nominal significance level α	average proportion of distinct observations 1.0000		.7286		.6347		.3180	
	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a
.250	.2620	.2626	.2606	.2640	.2618	.2662	.2580	.2686
.100	.1040	.1056	.1028	.1068	.1028	.1084	.1042	.1144
.050	.0512	.0520	.0514	.0530	.0522	.0538	.0494	.0588
.025	.0248	.0252	.0252	.0260	.0250	.0268	.0242	.0290
.010	.0090	.0090	.0090	.0092	.0090	.0096	.0088	.0102

Table 15 Observed Significance Levels of A^2_{kN} and A^2_{akN}
Number of Replications = 5000
Sample Sizes 10 20 30 40 50

nominal significance level α	average proportion of distinct observations 1.0000		.6465		.5407		.2417	
	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a
.250	.2596	.2610	.2592	.2644	.2590	.2638	.2602	.2682
.100	.1052	.1062	.1054	.1072	.1056	.1078	.1040	.1136
.050	.0526	.0528	.0528	.0554	.0530	.0566	.0540	.0606
.025	.0266	.0270	.0268	.0280	.0258	.0282	.0240	.0312
.010	.0074	.0078	.0078	.0084	.0078	.0086	.0076	.0098

Table 16 Observed Significance Levels of A^2_{kN} and A^2_{akN}
Number of Replications = 5000
Sample Sizes 10 10 10 10 50

nominal significance level α	average proportion of distinct observations 1.0000		.7607		.6733		.3561	
	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a
.250	.2612	.2634	.2630	.2640	.2630	.2678	.2608	.2700
.100	.1094	.1104	.1098	.1112	.1086	.1124	.1074	.1168
.050	.0520	.0534	.0530	.0552	.0530	.0560	.0522	.0608
.025	.0262	.0270	.0262	.0278	.0266	.0278	.0280	.0316
.010	.0112	.0118	.0110	.0122	.0116	.0122	.0114	.0144

Table 17	Observed Significance Levels of A^2_{kN} and A^2_{akN} Number of Replications = 5000 Sample Sizes 10 50 50 50 50							
	average proportion of distinct observations							
nominal significance level α	1.0000		.5594		.4483		.1837	
	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a	A^2	A^2_a
.250	.2480	.2490	.2478	.2486	.2468	.2500	.2490	.2544
.100	.1016	.1012	.1010	.1030	.1018	.1040	.1010	.1074
.050	.0506	.0510	.0508	.0532	.0514	.0540	.0500	.0582
.025	.0254	.0258	.0254	.0266	.0262	.0272	.0268	.0290
.010	.0110	.0110	.0104	.0110	.0102	.0112	.0108	.0126

Data Sets

Table 18	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

Table 19 Operating hours between failures of air conditioning systems for separate airplanes and major overhaul (MO) segments

airplane	operating hours								
7907	194	15	41	29	33	181				
7908	413	14	58	37	100	65	9	169	447	184
	36	201	118							
7908 MO	34	31	18	18	67	57	62	7	22	34
7909	90	10	60	186	61	49	14	24	56	20
	79	84	44	59	29	118	25	156	310	76
	26	44	23	62						
7909 MO	130	208	70	101	208					
7910	74	57	48	29	502	12	70	21	29	386
	59	27								
7911	55	320	56	104	220	239	47	246	176	182
	33									
7912	23	261	87	7	120	14	62	47	225	71
	246	21	42	20	5	12	120	11	3	14
	71	11	14	11	16	90	1	16	52	95
7913	97	51	11	4	141	18	142	68	77	80
	1	16	106	206	82	54	31	216	46	111
	39	63	18	191	18	163	24			
7914	50	44	102	72	22	39	3	15	197	188
	79	88	46	5	5	36	22	139	210	97
	30	23	13	14						
7915	359	9	12	270	603	3	104	2	438	
7916	50	254	5	283	35	12				
8044	487	18	100	7	98	5	85	91	43	230
	3	130								
8045	102	209	14	57	54	32	67	59	134	152
	27	14	230	66	61	34				