Approximations to the Distribution Function of the Anderson–Darling Test Statistic

C. D. SINCLAIR and B. D. SPURR*

A theoretical approximation to the upper tail area of the Anderson–Darling test statistic is derived by applying a result of Zolotarev (1961) to the characteristic function. This approximation is particularly good in the upper tail. An empirical distribution is given that is accurate over the complete domain of the distribution.

KEY WORDS: Characteristic function; Cumulants; Empirical approximation; Theoretical approximation.

1. INTRODUCTION

The Anderson–Darling test is one of the most powerful and important goodness-of-fit tests in the statistical literature. Although designed to be more sensitive to discrepancies between the hypothesized and empirical distribution functions in the tails of the distribution, it also performs well in other situations. [See D'Agostino and Stephens (1986, chap. 4) for further details and a full list of references.]

In Section 2 we give the first four cumulants, skewness, and kurtosis of the asymptotic distribution of the Anderson–Darling test statistic. In Section 3 we derive a theoretical approximation to the upper tail area and give an empirical approximation to the distribution function over the whole domain. The latter approximation, derived with values given by Lewis (1961), also provides an adequate fit to results of simulations.

2. CUMULANTS OF A^2

The Anderson–Darling test statistic A_n^2 is defined as

$$A_n^2 = n \int_{-\infty}^{\infty} \frac{[F_n(x) - F(x)]^2}{F(x)[1 - F(x)]}\, dF(x), \qquad (2.1)$$

where $F_n(x)$ is the empirical distribution function of the sample of n observations and $F(x)$ is the hypothesized continuous distribution function.

The first two cumulants of A^2 (the asymptotic case of A_n^2) are well known, but the third and fourth cumulants do not seem to be readily available.

The characteristic function of A^2 (Anderson and Darling 1952) is given by

$$\phi(t) = [(-2\pi it)/\cos(\pi/2\sqrt{1 + 8it})]^{1/2}. \qquad (2.2)$$

Now, either by differentiation of $\phi(t)$ or inspection of the coefficients of $\log \phi(t)$, the first four cumulants of the distribution are

$$\kappa_1 = 1$$

$$\kappa_2 = \frac{2}{3}(\pi^2 - 9) = .579726$$

$$\kappa_3 = 80 - 8\pi^2 = 1.04329$$

$$\kappa_4 = \frac{16}{15}\pi^4 + 160\pi^2 - 1{,}680 = 3.03699. \qquad (2.3)$$

The squared coefficient of skewness and the coefficient of kurtosis are then given by

$$\beta_1 = 5.5865, \qquad \beta_2 = 12.036. \qquad (2.4)$$

Thus, in relation to the Pearson and Johnson systems of curves, the distribution lies between the gamma and lognormal.

3. APPROXIMATING THE UPPER TAIL AREA

The characteristic function (2.2) can also be written as

$$\phi(t) = \prod_{j=1}^{\infty} \left(1 - \frac{2it}{j(j+1)}\right)^{-1/2} \qquad (3.1)$$

(see Anderson and Darling 1952, p. 204). This corresponds to the characteristic function of an infinite weighted sum of independent chi-squared random variables with weights

$$\lambda_j = 1/j(j+1). \qquad (3.2)$$

Consequently, we can use results on limiting distributions of quadratic forms of Zolotarev (1961). The general result is that if λ_j are positive and decreasing and $\sum_{j=1}^{\infty} \lambda_j < \infty$, then the upper tail probabilities of $\sum_{j=1}^{\infty} \lambda_j z_j^2$, where $z_j \sim N(0, 1)$, are given by

$$1 - F(x) \simeq \left[\prod_{j=2}^{\infty}\left[1 - \frac{\lambda_j}{\lambda_1}\right]^{-1/2} \Big/ \Gamma(\tfrac{1}{2})\right] \times \left[\frac{x}{2\lambda_1}\right]^{-1/2} \exp\left[\frac{-x}{2\lambda_1}\right][1 + \varepsilon(x)], \qquad (3.3)$$

with $\varepsilon(x) \to 0$ as $x \to \infty$.

The λ_j given by (3.2) are obviously positive and decreasing, and $\sum_{j=1}^{\infty} \lambda_j = \sum_{j=1}^{\infty} 1/j(j+1) = 1$. Hence, in this case (3.3) reduces to

$$1 - F(x) \simeq \left[\prod_{j=2}^{\infty}\left[1 - \frac{2}{j(j+1)}\right]^{-1/2} \Big/ \Gamma(\tfrac{1}{2})\right] \times (x)^{-1/2} \exp(-x)[1 + \varepsilon(x)]. \qquad (3.4)$$

To four significant figures the constant term $K = \prod_{j=2}^{\infty} (1 - 2/j(j+1))^{-1/2} = 1.732$. Equation (3.4) then becomes

$$1 - F(x) \simeq [1.732/\Gamma(\tfrac{1}{2})]x^{-1/2} \exp(-x)[1 + \varepsilon(x)]. \qquad (3.5)$$

* C. D. Sinclair and B. D. Spurr are Lecturers, Department of Mathematical Sciences, University of St. Andrews, Fife, KY16 9SS, Scotland.

Journal of the American Statistical Association
December 1988, Vol. 83, No. 404, Theory and Methods

Table 1. Tail Area Comparisons Using the Theoretical Approximation (3.5)

Critical value	*Nominal significance level*	*Tail area from (3.5)*
.283	.95	1.3841
.346	.90	1.1754
.399	.85	1.0380
.774	.50	.5122
1.249	.25	.2508
1.933	.10	.1017
2.492	.05	.0512
3.079	.025	.0256
3.880	.01	.010

If $\varepsilon(x) = 0$ in (3.5), the resulting approximation is in good agreement with the percentage points of A^2 given by Lewis (1961) for values above the median. It is particularly good in the upper tail of the distribution. Table 1 compares the tail areas obtained from (3.5) with the nominal levels for the critical values obtained by interpolation by Lewis (1961).

The previous approximation is poor in the lower tail. The following empirical approximation is adequate over the whole range (see Tables 2 and 3):

$$F(x) = 1 - 1/[1 + \exp(1.784 + .9936x + .03287/x - (2.018 + .2029/x)x^{-1/2})]. \quad (3.6)$$

This approximation was obtained by fitting a generalization of the logistic distribution to the values given by Lewis (1961), omitting the three points at .65, 7.0, and 8.0, whose accuracy is put in doubt by the results of simulations. Tables 2 and 3 compare tail areas obtained from (3.6) with those given by Lewis (1961).

Table 2. Upper Tail Area Comparisons Using the Empirical Approximation (3.6)

	x					
	1.25	*1.65*	*1.95*	*2.50*	*3.05*	*3.85*
$P = 1 - F(x)$	.2497	.1444	.0979	.0496	.0258	.0103
$\hat{P}$ from (3.6)	.2493	.1447	.0981	.0496	.0258	.0103
$100(\hat{P} - P)/P$	−.1	.2	.2	.1	−.0	−.3

Table 3. Lower Tail Area Comparisons Using the Empirical Approximation (3.6)

	x				
	.50	*.40*	*.35*	*.275*	*.20*
$P = F(x)$	.2532	.1513	.1036	.0443	.0096
$\hat{P}$ from (3.6)	.2533	.1507	.1029	.0440	.0096
$100(\hat{P} - P)/P$	.0	−.4	−.7	−.7	.3

4. CONCLUSION

From Table 1 we note that the approximation (3.5) consistently overestimates the tail area. Nevertheless, it is simple to compute, and involves only one coefficient. Theoretical considerations suggest that it will be increasingly accurate in the extreme tail. We know of no other formula that is applicable to the extreme tail.

Comparisons of absolute and relative errors reveals that (3.6) is more accurate than (3.5) to the right of the median. Moreover, (3.6) is also accurate to the left of the median. It involves five coefficients, however.

We suggest that for practical purposes (3.5) is sufficiently accurate for testing the goodness of fit of a single data set, whereas (3.6) is a useful complement to Lewis's table and may be helpful in assessing the combined goodness of fit of several data sets.

[*Received September 1987. Revised March 1988.*]

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.

D'Agostino, R. B., and Stephens, M. A. (eds.) (1986), *Goodness of Fit Techniques*, New York: Marcel Dekker.

Lewis, P. A. W. (1961), "Distribution of the Anderson–Darling Statistic," *Annals of Mathematical Statistics*, 32, 1118–1124.

Zolotarev, V. M. (1961), "Concerning a Certain Probability Problem," *Theory of Probability and Its Applications*, 6, 201–204.