The Annals of Statistics
1976, Vol. 4, No. 2, 357-369

ASYMPTOTIC RESULTS FOR GOODNESS-OF-FIT STATISTICS WITH UNKNOWN PARAMETERS

By M. A. Stephens

McMaster University

Percentage points are given for the asymptotic distributions of the goodness-of-fit statistics W^2, U^2 and A^2, for the cases where the distribution tested is

(a) normal, with mean or variance, or both, unknown;
(b) exponential, with scale parameter unknown.

Some exact means and variances are also given. The distributions can be expressed as a sum of weighted chi-square variables; the weights are calculated, and the higher cumulants can then be found. The first four cumulants are used to approximate the distributions and give the percentage points.

1. Introduction. Let $x_1, x_2, \cdots, x_n$ be independent random variables from a continuous distribution $G(x)$, and let $F_n(x)$ be the empirical distribution function. A well-known goodness-of-fit test, to test the null hypothesis

$$H_0: G(x) = F(x; \boldsymbol{\theta}) ,$$

where $F(x; \boldsymbol{\theta})$ contains a parameter $\boldsymbol{\theta}$ of several components, is based on the Cramér–von Mises statistic

$$W^2 = n \int_{-\infty}^{\infty} \{F_n(x) - F(x; \boldsymbol{\theta})\}^2 \, dF(x; \boldsymbol{\theta}) .$$

Two other goodness-of-fit statistics similar to W^2 are U^2, introduced by Watson (1961) and A^2, introduced by Anderson and Darling (1952, 1954):

$$U^2 = n \int_{-\infty}^{\infty} \{F_n(x) - F(x; \boldsymbol{\theta}) - \int_{-\infty}^{\infty} [F_n(t) - F(t; \boldsymbol{\theta})] \, dF(t; \boldsymbol{\theta})\}^2 \, dF(x; \boldsymbol{\theta})$$

and

$$A^2 = n \int_{-\infty}^{\infty} \{F_n(x) - F(x; \boldsymbol{\theta})\}^2 \Psi(x; \boldsymbol{\theta}) \, dF(x; \boldsymbol{\theta})$$

with

$$1/\Psi(x; \boldsymbol{\theta}) = F(x; \boldsymbol{\theta})\{1 - F(x; \boldsymbol{\theta})\} .$$

The statistic U^2 was introduced for observations on a circle, since its value is independent of the choice of origin, but it can also be used for observations on a line. The statistic A^2 modifies W^2 by giving greater weight to the tails of the distribution, and so can be expected to detect discrepancies in the tails better than W^2. These statistics are often written with a subscript n, which will here be omitted.

2. Contents of this paper. Most of the distribution theory hitherto found for

---

Received October 1971; revised August 1975.

AMS 1970 *subject classifications.* Primary 62E20, 62F05; Secondary 62E15; 62G30.

Key words and phrases. Goodness-of-fit tests, asymptotic distributions, empirical distribution function, order statistics.

these statistics is for the situation (which we call Case 0) when $F(x; \boldsymbol{\theta})$ is completely specified. In many practical situations, however, some or all of the components of $\boldsymbol{\theta}$ will be unknown, and must be estimated from the sample of x-values. The following important examples, arising in tests for normality or exponentiality, will be called Cases 1, 2, 3 and 4.

CASE 1. $F(x; \boldsymbol{\theta})$ is the normal distribution, with $\boldsymbol{\theta} = (\mu, \sigma^2)$; σ^2 is known and μ is to be estimated by $\bar{x}$;

CASE 2. $F(x; \boldsymbol{\theta})$ is normal, μ is known, and σ^2 is to be estimated by $\sum_i (x_i - \mu)^2/n$;

CASE 3. $F(x; \boldsymbol{\theta})$ is normal, and μ and σ^2 are to be estimated by $\bar{x}$ and $s^2 = \sum_i (x_i - \bar{x})^2/(n-1)$;

CASE 4. $F(x; \boldsymbol{\theta})$ is the exponential distribution, $\boldsymbol{\theta} = \theta$; thus $F(x; \theta) = 1 - \exp(-\theta x)$, $x \geqq 0$, and θ is to be estimated by $1/\bar{x}$.

For all these cases, the asymptotic distributions of W^2, U^2, and A^2 can be expressed as the weighted sum of χ_1^2 variables. In this paper the weights are found for Cases 1 to 4; the cumulants of the asymptotic distributions can then be calculated, and they are used, with curve-fitting techniques, to provide tables of asymptotic percentage points for the different statistics. Exact values are also found for the means and variances of W^2 and U^2. It is found that the null distributions of W^2, U^2, and A^2, in Cases 1 to 4, are drastically changed from those obtained in Case 0, and reliance on the published tables for Case 0 will introduce serious errors into the significance levels of the test statistics.

For finite n, Monte Carlo studies show that, as for Case 0, the percentage points converge rapidly to the asymptotic points. A slight modification of the calculated statistic enables a goodness-of-fit test to be made with finite samples, using only the asymptotic points. Details of these practical aspects of the tests, including power studies which show the statistics to be very effective in goodness-of-fit tests, are given in Stephens (1974).

3. Basic results. In this section we give results which will be needed in the sequel. Basic asymptotic theory was given by Anderson and Darling (1952) for W^2 and A^2, and by Watson (1961) for U^2. Darling (1955) considered problems with one parameter to be estimated; this was extended to multiparameter situations by Sukhatme (1972), and, for Case 3, by Stephens (1971). Case 3 was also examined, from a different point of view, by Kac, Kiefer, and Wolfowitz (1955). A comprehensive treatment of the basic theory has recently been given by Durbin (1973). The results following are taken from these references. For any of the three statistics, the asymptotic distribution is that of $\int_0^1 Y^2(t)\,dt$, where $Y(t)$ is an appropriate Gaussian process; $Y(0) = Y(1) = 0$, the mean is zero, and the covariance function $\rho(s, t)$ depends on the statistic, on $F(x, \boldsymbol{\theta})$, and on the parameter(s) to be estimated. When $\rho(s, t)$ is known, the characteristic function of the distribution (the word *asymptotic* will be dropped) is given by

$(D(2it))^{-\frac{1}{2}}$, where $D(\lambda)$ is the Fredholm determinant associated with $\rho(s, t)$. The eigenvalues λ_i of the integral equation

$$f(x) = \lambda \int_0^1 \rho(x, y) f(y)\, dy \tag{1}$$

are the solutions of $D(\lambda) = 0$. Given a statistic and the case considered, the λ_i must be found for the appropriate $\rho(s, t)$. The asymptotic distribution is then the same as that of

$$S = \sum_{i=1}^{\infty} z_i/\lambda_i\,, \tag{2}$$

where the z_i are independent χ_1^2 variables. Further, the cumulants of the distribution are given by

$$K_j = s^{j-1}(j-1)!\int_0^1 \rho_j(s, s)\, ds \tag{3}$$

where $\rho_j(s, t)$, the jth iterate of $\rho(s, t)$, is given by

$$\rho_j(s, t) = \int_0^1 \rho_{j-1}(s, u)\rho(u, t)\, du\,, \qquad j \geqq 2\,;$$
$$\rho_1(s, t) = \rho(s, t)\,.$$

From the representation (2) we have also the cumulants in the form

$$K_j = 2^{j-1}(j-1)!\sum_{i=1}^{\infty}(1/\lambda_i)^j\,. \tag{4}$$

In order to use these results, we must find $\rho(s, t)$. Let $\rho_0(s, t)$ be the covariance of the Gaussian process, for the statistic considered, in the Case 0 situation. If only one parameter θ is unknown, and maximum likelihood is used to give an efficient estimator, $\rho(s, t)$ becomes

$$\rho(s, t) = \rho_0(s, t) - \phi(s)\phi(t) \tag{5}$$

where $\phi(s)$ is a function found as follows.

Let $s = F(x; \theta)$ define x implicitly in terms of s. Let $f(x; \theta) = (\partial/\partial x)F(x; \theta)$; $g(s) = (\partial/\partial\theta)F(x; \theta)$; and let k^2 be defined by

$$\frac{1}{k^2} = \int_{-\infty}^{\infty}\left\{\frac{\partial}{\partial\theta}\ln f(x; \theta)\right\}^2 f(x; \theta)\, dx\,.$$

Then $\phi(s) = kg(s)$. If θ is a location or scale parameter, $\phi(s)$ is independent of θ. In Case 3 there are location and scale parameters, both unknown. If $\phi_1(s)$ and $\phi_2(s)$ are the $\phi(s)$ for Cases 1 and 2 respectively, the covariance for Case 3 becomes

$$\rho(s, t) = \rho_0(s, t) - \phi_1(s)\phi_1(t) - \phi_2(s)\phi_2(t)\,. \tag{6}$$

4. Solution for $D(\lambda)$. When $\rho(s, t)$ is known, equation (1) must be solved; the fact that the covariance, for Cases 1 to 4, is obtained from that for Case 0 by subtracting terms which factor into two parts, one containing only s and the other only t, enables a straightforward solution based on that for Case 0.

Suppose first that $\rho_0(s, t)$ has Fredholm determinant $D_0(\lambda)$, roots $0 \leqq \lambda_1 \leqq \lambda_2 \cdots$, and corresponding eigenfunctions $f_1(x), f_2(x), \cdots$. Define

$$a_j = \int_0^1 f_j(x)\phi_1(x)\,dx\,;$$
$$b_j = \int_0^1 f_j(x)\phi_2(x)\,dx\,;$$
$$S_a(\lambda) = 1 + \lambda \sum_{i=1}^{\infty} a_i^2/(1 - \lambda/\lambda_i)\,;$$
$$S_b(\lambda) = 1 + \lambda \sum_{i=1}^{\infty} b_i^2/(1 - \lambda/\lambda_i)\,;$$
$$S_{ab}(\lambda) = \lambda \sum_{i=1}^{\infty} a_i b_i/(1 - \lambda/\lambda_i)\,.$$

Further, let $c_i(g) = \int_0^1 g(x)\phi_i(x)\,dx$, $i = 1, 2$. When $\rho(s, t)$ takes the form $\rho_0(s, t) - \phi_1(s)\phi_1(t)$, i.e., when a location parameter only is estimated, Darling shows that the Fredholm determinant is

$$D_1(\lambda) = D_0(\lambda)S_a(\lambda)\,; \tag{7}$$

when a scale parameter is estimated, $\phi_1(s)$ is replaced by $\phi_2(s)$, and the new determinant is

$$D_2(\lambda) = D_0(\lambda)S_b(\lambda)\,. \tag{8}$$

For the kernel of equation (6), the determinant is (Sukhatme (1972) with $k = 2$; or Stephens (1971))

$$D_3(\lambda) = D_0(\lambda)\{S_a(\lambda)S_b(\lambda) - S_{ab}^2(\lambda)\}\,. \tag{9}$$

5. Two useful lemmas. Suppose $C_r(t)$ is the characteristic function, for Case r, of the asymptotic distribution of one of the statistics, and let K_{rj} be the jth cumulant of the distribution.

LEMMA 1. *If $S_{ab}(\lambda) = 0$, then*

$$D_0(\lambda)D_3(\lambda) = D_0^2(\lambda)S_a(\lambda)S_b(\lambda) = D_1(\lambda)D_2(\lambda)\,,$$

and

$$C_0(t)C_3(t) = C_1(t)C_2(t)\,.$$

Also

$$K_{0j} + K_{3j} = K_{1j} + K_{2j} \tag{10}$$

and equivalently

$$K_{0j} - K_{3j} = K_{0j} - K_{1j} + K_{0j} - K_{2j}\,.$$

PROOF. Follows immediately from (7), (8), and (9).

The condition $S_{ab}(\lambda) = 0$ will be satisfied if $\phi_1(s)$, $\phi_2(s)$ are such that either a_i or b_i is zero for all i; this can easily be shown to occur in Cases 1 to 3. Then we have the interesting result that the decrease in value of any cumulant in proceeding from Case 0 to Case 3 is the sum of the decreases between Case 0 and Case 1 and between Case 0 and Case 2. Lemma 1 applies also to nonnormal distributions containing location and/or scale parameters to be estimated; then, even if $S_{ab}(\lambda)$ is not zero, a weaker result may nevertheless hold.

LEMMA 2. *If the functions $\phi_1(x)$, $\phi_2(x)$ are orthogonal*, i.e., $\int_0^1 \phi_1(x)\phi_2(x)\,dx = 0$, *the means and variances of Cases* 0–3 *satisfy* (10) *for $j = 1$ and* 2.

PROOF. Follows from the cumulant formulae in (3). The proof is obvious for the mean; for the more difficult variance,

$$\begin{aligned} K_{02} + K_{32} &= \textstyle\int_0^1 \int_0^1 \{\rho_0(s, t)\}^2 + \{\rho_0(s, t) - \phi_1(s)\phi_1(t) - \phi_2(s)\phi_2(t)\}^2 \, ds\, dt \\ &= \textstyle\int_0^1 \int_0^1 \{\rho_0(s, t) - \phi_1(s)\phi_1(t)\}^2 \, ds\, dt \\ &\quad + \textstyle\int_0^1 \int_0^1 \{\rho_0(s, t) - \phi_2(s)\phi_2(t)\}^2 \, ds\, dt \\ &= K_{12} + K_{22} , \qquad \text{if} \quad \textstyle\int_0^1 \phi_1(s)\phi_2(s)\, ds = 0 . \end{aligned}$$

Note that the orthogonality condition implies only $\sum_i a_i b_i = 0$, not $a_i b_i = 0$ for all i. The weaker result is that Lemma 2 holds only for two cumulants.

6. Related results. In this paragraph we use the notation of Kac, Kiefer and Wolfowitz (1955). They solve (1) above for W^2 by solving the related differential equation, in their Section 2.6. They point out that the Fredholm determinant for Case 3, there called $D(\mu)$, can be factored: $D(\mu) = -2D_1(\mu)D_2(\mu)$, where $D_1(\mu)$, $D_2(\mu)$ are given by their (2.18) and (2.19), and where μ^2 is λ in the present paper. $D_1(\mu)$ and $D_2(\mu)$ are not the Case 1 and Case 2 Fredholm determinants, but are closely related to them. If the Case 1 and Case 2 integral equations are solved by the method of Kac et al., to give Fredholm determinants $D_1^*(\mu)$, and $D_2^*(\mu)$, we obtain

$$\begin{aligned} D_1^*(\mu) &= 2(\sin \tfrac{1}{2}\mu)D_1(\mu)/\mu ; \\ D_2^*(\mu) &= -2(\cos \tfrac{1}{2}\mu)D_2(\mu) . \end{aligned}$$

The solution follows their Section 2.6, and the long algebraic details will be omitted. The product $D_1^*(\mu)D_2^*(\mu)$ is then $(\sin \mu)D(\mu)/\mu$. Since $D_0(\mu)$, for W^2, is $(\sin \mu)/\mu$, we have

$$D_1^*(\mu)D_2^*(\mu) = D_0(\mu)D(\mu) ,$$

which is Lemma 1 above, in the notation of Kac et al. for the particular example of the normal distribution.

In order to find the zeros of $D(\mu)$, those of $D_1(\mu)$ and $D_2(\mu)$ are required. Expressions for these functions are given by Kac et al., but the techniques for finding the zeros were computationally difficult, and so only the first eight μ_i were found; the first four were used in an approximation of the form of equation (2), to be discussed in Section 10.

Returning to the notation of Section 4, we find weights λ_i by finding the zeros of $D_k(\lambda)$ for Case k. This can be done, from (7), (8) and (9), very quickly, and to a large order i: values of a_i and b_i involve only single integrals and the solutions of $S_a(\lambda) = 0$, $S_b(\lambda) = 0$ are straightforward. This will be done for Cases 1 to 3 in Section 8 and Case 4 in Section 9.

7. Cases 1, 2, 3; covariance functions, and means and variances.

Notation. The following notation will be used in the calculations. Let $d = (2\pi)^{-\frac{1}{2}}$; $n(x) = d\exp(-x^2/2)$; $N(x) = \int_{-\infty}^{x} n(t)\, dt$. When $s = N(x)$, let $x = J(s)$, i.e., $J(\cdot)$ is the inverse of $N(\cdot)$. Define functions:

$$r(s, t) = \min(s, t) - st$$

$$B(s) = d\exp(-J^2(s)/2)\,; \qquad b(s, t) = -B(s)B(t)$$

$$C(s) = d(J(s)/2^{\frac{1}{2}})\exp(-J^2(s)/2)\,; \qquad c(s, t) = -C(s)C(t)$$

$$w(s, t) \doteq \{(s - s^2)(t - t^2)\}^{-\frac{1}{2}}\,; \qquad a(s, t) = \tfrac{1}{12} - \{(s - s^2) + (t - t^2)\}/2$$

$$E(s) = B(s) - 1/(2\pi^{\frac{1}{2}})\,; \qquad e(s, t) = -E(s)E(t)\,.$$

When no ambiguity can arise, the arguments s, t will be omitted. To express the value of a double integral, with limits 0 and 1, capital letters will denote the functions in the integrand; e.g.

$$R^2 = \int_0^1 \int_0^1 r^2(s, t)\, ds\, dt\,; \qquad RB = \int_0^1 \int_0^1 r(s, t)b(s, t)\, ds\, dt\,.$$

Covariance functions for W^2. For Case 0, $\rho_0(s, t) = \min(s, t) - st = r$ above. For Case 1, take the known variance to be 1 (if it were σ^2, values of x/σ would be tested to come from a normal distribution with variance 1). The unknown θ is the mean μ, and $\phi_1(s)$, calculated as in Section 3, becomes $-B$ above. Similarly, for Case 2, take μ as zero; the unknown θ is the variance, and $\phi_2(s)$ is $-C$. Covariances for Cases 1 and 2 are then given by (5) and for Case 3 by (6). They can be summarized in a table.

Case:	0	1	2	3
ρ:	r	$r + b$	$r + c$	$r + b + c$

Equation (3) can now be used to find cumulants of W^2; in practice, the calculations become extremely long and only the means and variances have been calculated, as shown below.

Means and variances: W^2. *Case* 1. Use of (3) with ρ equal to $r + b$ gives, for Case 1,

$$\mu = K_1 = \int_0^1 s(1 - s)\, ds - d^2 \int_0^1 \exp(-J^2(s))\, ds\,.$$

Let $x = J(s)$; then

$$\mu = \tfrac{1}{6} - \int_{-\infty}^{\infty} n^3(x)\, dx = \tfrac{1}{6} - \frac{d^2}{3^{\frac{1}{2}}} = .074778\,.$$

Terms involving r alone, such as the first term in μ above, arise in Case 0 and we shall use known values for this case, taken from Anderson and Darling (1952) or Watson (1961).

The variance is

$$\sigma^2 = K_2 = 2 \int_0^1 \int_0^1 (r + b)^2\, ds\, dt = 2(R^2 + 2RB + B^2)\,,$$

where (from Case 0, known result) $2R^2 = \frac{1}{45}$ and $2B^2 = 2(\mu - \frac{1}{6})^2$; $4RB$ must be found. The substitution $x = J(s)$, $y = J(t)$ gives, after much algebra,

$$RB = -\frac{d^2}{8} + d^4 \tan^{-1}\left(\frac{1}{5^{\frac{1}{2}}}\right) = -0.0092425\,;$$

finally we have $\sigma^2 = 0.002139$. These and subsequent calculations are given in more detail in Stephens (1971).

Case 2. For Case 2,

$$\mu = \tfrac{1}{6} - \frac{d^2}{8} \int_0^1 J^2(s) \exp(-J^2(s))\, ds = \tfrac{1}{6} - \frac{1}{12\pi 3^{\frac{1}{2}}} = 0.15135 ,$$

and

$$\sigma^2 = 2(R^2 + 2RC + C^2) ;$$

here

$$2R^2 = \tfrac{1}{45}$$

and

$$C^2 = (\pi - \tfrac{1}{6})^2 = 1/(432\pi^2) ;$$

considerable algebra gives

$$RC = 1/(32\pi^2 5^{\frac{1}{2}}) + 1/(96\pi^2) ,$$

and finally $\sigma^2 = .02125$.

Case 3. For Case 3, the mean and variance are obtained by applying (10): $\mu = K_{31} = 0.07478 + 0.15135 - 0.16667 = 0.05946$, and $\sigma^2 = K_{32} = 0.00214 + 0.02125 - 0.02222 = 0.00117$. All these means and variances are recorded in Table 1.

TABLE 1
Asymptotic means and variances of W^2, U^2 and A^2: in Cases 1, 2, 3 *the test is for normality; in Case* 4 *the test is for exponentiality*

Case No.	Test	W^2		U^2		A^2	
		Mean	Variance	Mean	Variance	Mean	Variance
0	parameters given	0.16666	0.02222	0.08333	0.002777	1.0000	0.5797
1	σ^2 specified μ estimated	.0748	.00214	.0710	.00195	.5194	
2	μ specified σ^2 estimated	.1514	.02125	.0680	.00181	.8649	
3	μ, σ^2 estimated	.0595	.00116	.0557	.00097	.3843	
4	θ estimated	.0926	.00436	.0718	.00198	.5959	

Covariance function for U^2. Watson (1961) has shown that the limiting distribution of U^2 is that of $\int_0^1 Q^2(t)\, dt$, where $Q(t)$ is the Gaussian process $Y(t) - \int_0^1 Y(u)\, du$; here $Y(t)$ is the Gaussian process, for the appropriate case, used to find the limiting distribution of W^2. For a particular case, let the covariance function of $Y(t)$ be $\rho(s, t)$ and let that of $Q(t)$ be $\rho^*(s, t)$; then

$$\begin{aligned}\rho^*(s, t) &= \rho(s, t) + \int_0^1 \int_0^1 \rho(s, t)\, ds\, dt - \int_0^1 \rho(s, t)\, ds - \int_0^1 \rho(s, t)\, dt \\ &= \rho(s, t) + E_1 + E_2 + E_3 , \quad \text{say.}\end{aligned}$$

For each case it is necessary to find the integrals E_1, E_2 and E_3, corresponding to the $\rho(s, t)$ given in Section 7. This again needs only straightforward algebra, and, with the notation $a(s, t)$ and $e(s, t)$ introduced in Section 7, the covariance function $\rho^*(s, t)$ for U^2 can be tabulated as follows.

Case:	0	1	2	3
ρ^*:	$r + a$	$r + a + e$	$r + a + c$	$r + a + c + e$

Note that the Case 3 covariance is of type (6), i.e. the terms added to ρ^* for Case 0 are the sum of those added for Cases 1 and 2. When these covariances are used in (3), extensive algebra (details in Stephens, 1971) gives the moments recorded in Table 1.

Covariance functions for A^2. For the Anderson–Darling statistic A^2, and for any particular case, the Gaussian process appropriate for finding the asymptotic distribution has a covariance function which is the function for W^2, given above, multiplied by $w(s, t) = \{(s - s^2)(t - t^2)\}^{-\frac{1}{2}}$. The resulting integrals do not seem to be tractable by analytic methods, but since this particular statistic gives good results in goodness-of-fit testing (Stephens, 1974) the means have been calculated numerically and are given in Table 1.

8. Calculation of weights; Cases 1, 2 and 3. In this section are found the weights λ_i in the representation of equation (2). For each case, these are the solutions of the relevant $D_r(\lambda) = 0$, $r = 1, 2, 3, 4$. For each statistic let the weights λ_i obtained for Case 0, i.e. the solutions of $D_0(\lambda) = 0$, be called the *standard* weights; for all other cases the weights consist of a subset of the standards, plus a new set λ_i^* labelled with an asterisk. The standards are given below as functions of i; the values of $1/\lambda_i^*$ are given in Table 2, for i from 1 to 10.

W^2: *Case* 1. For W^2, $D_0(\lambda) = \sin \lambda^{\frac{1}{2}}/\lambda^{\frac{1}{2}}$, and the standards are $\lambda_i = \pi^2 i^2$, with $f_i(x) = 2^{\frac{1}{2}} \sin \pi i x$ (Anderson and Darling, 1952). For Case 1, $D_1(\lambda) = D_0(\lambda)S_a(\lambda)$; the zeros λ_i of $D_0(\lambda)$ are simple zeros, and will not be zeros of $D_1(\lambda)$ unless, in

TABLE 2
Values of $100/\lambda_i^*$ *for Cases* 1, 2 *and* 4;
for Case 3, *the values are given by Cases* 1 *and* 2 *combined*

i	W^2			U^2			A^2		
	Case 1	Case 2	Case 4	Case 1	Case 2	Case 4	Case 1	Case 2	Case 4
1	1.834	1.344	4.202	1.573	1.345	1.626	9.836	7.206	23.130
2	.535	.436	1.712	.477	.436	.488	3.593	2.897	9.964
3	.252	.216	.815	.230	.216	.234	1.810	1.584	5.635
4	.146	.129	.509	.135	.129	.137	1.148	1.002	3.641
5	.095	.085	.333	.089	.086	.090	.777	.692	2.552
6	.067	.061	.242	.063	.061	.063	.561	.508	1.890
7	.049	.045	.179	.047	.046	.047	.424	.388	1.457
8	.038	.035	.141	.036	.035	.036	.332	.307	1.158
9	.030	.028	.112	.029	.028	.029	.267	.248	.942
10	.024	.022	.092	.024	.023	.024	.213	.197	.782

$S_a(\lambda)$, the corresponding a_i is zero. For Case 1, $\phi_1(s) = -B(s)$, from Section 7, and is symmetric around $s = 0.5$; then the coefficients a_i are zero for i even, and the subset of the standards given by $\lambda_i = 4\pi^2 i^2$, $i = 1, 2, \cdots$ are zeros of $D_1(\lambda)$. The other zeros λ_i^* are solutions of $S_a(\lambda) = 0$; the first twenty have been calculated, and the reciprocals of the first ten are tabulated in Table 2. A plot of $1/\lambda_i^*$ against i^{-2} is a smooth curve asymptotic to the line through the origin with slope $1/(4\pi^2)$. Values of $1/\lambda_i^*$ for $i > 20$ have been found from this curve. Since convergence is slow, many values of λ_i are required to obtain the mean from (4); thus it is important to have the exact calculation. However, higher cumulants, calculated from (4), converge much faster, and the variance, given in Table 3, converges exactly to the value in Table 1. This provides a check on the values of $1/\lambda_i^*$. The third and fourth cumulants and values of $\beta_1 = K_3^2/K_2^3$ and $\beta_2 = K_4/K_2^2 + 3$ are also given in Table 3.

TABLE 3
Cumulants and shape parameters of asymptotic distributions calculated from weights

Statistic	Case	σ^2	10^3K_3	10^4K_4	β_1	β_2
W^2	1	.00214	.183	.253	3.418	8.528
	2	.02125	8.35	50.61	7.273	14.209
	3	.00117	.0709	.0705	3.170	8.186
	4	.00436	.639	1.54	4.949	11.117
U^2	1	.00195	.164	.228	3.662	9.015
	2	.00181	.152	.214	3.957	9.580
	3	.00097	.052	.045	2.966	7.795
	4	.00198	.168	.232	3.619	8.918
A^2	1	.08560	46.2	419.3	3.407	8.723
	2	.53029	1008.	30037.	6.817	13.681
	3	.03616	11.3	59.1	2.693	7.521
	4	.1393	109.0	1427.	4.396	10.358

W^2: *Case* 2. For Case 2, with $\phi_2(s) = -C(s)$, the b_i are zero for odd i; then in the solution of $D_2(\lambda) = 0$, the subset of λ_i given by $\lambda_i = \pi^2(2i - 1)^2$, $i = 1, 2, \cdots$ are included; the other solutions are λ_i^* satisfying $S_b(\lambda_i^*) = 0$, and their reciprocals are given in Table 2. For large i, these are asymptotic to $1/\pi^2(2i - 1)^2$. The cumulants, calculated from (4), are given in Table 3; the variances check perfectly with the exact value given in Table 1.

W^2: *Case* 3. For Case 3, $D_3(\lambda) = D_0(\lambda)S_a(\lambda)S_b(\lambda)$, and none of the standard λ_i can be a solution of $D_0(\lambda) = 0$. The solutions are given by the two sets of λ_i^* arising in Cases 1 and 2. These weights agree with those given by Kac et al., with a very slight numerical discrepancy in the fifth decimal place for the largest values.

U^2: *Cases* 1 *and* 2. The discussion of U^2 is slightly more complicated. The roots λ_i of $D_0(\lambda) = 0$ are double roots, given by $\lambda_i = 4\pi^2 i^2$ (Watson, 1961), and the corresponding eigenfunctions are $f_i(x) = 2^{\frac{1}{2}} \sin 2\pi ix$ and $f_i^*(x) = 2^{\frac{1}{2}} \cos 2\pi ix$. Suppose a_1 and a_i^*, b_i and b_i^* are the coefficients obtained using $f_i(x)$ and $f_i^*(x)$ respectively. Then $S_a(\lambda)$ becomes

$$S_a(\lambda) = 1 + \lambda \sum_i \frac{a_i^2}{1 - \lambda/\lambda_i} + \lambda \sum_i \frac{a_i^{*2}}{1 - \lambda/\lambda_i};$$

there is a similar expression for $S_b(\lambda)$, and

$$S_{ab}(\lambda) = \lambda \sum_i \frac{a_i b_i}{1 - \lambda/\lambda_i} + \lambda \sum_i \frac{a_i^* b_i^*}{1 - \lambda/\lambda_i}.$$

When the coefficients are calculated, it is found that either a_i or a_i^* is zero, but not both, and b_i or b_i^* is zero, but not both, such that $a_i b_i = a_i^* b_i^* = 0$ for all i. Thus $S_{ab}(\lambda) = 0$ and Lemma 1 applies. In Case 1, $D_1(\lambda) = D_0(\lambda) S_a(\lambda) = 0$ gives $\lambda_i = 4\pi^2 i^2$ as a solution, but not repeated; similarly when $D_2(\lambda) = 0$. In each case another set λ_i^* is found from $S_a(\lambda^*) = 0$, or $S_b(\lambda^*) = 0$. Values of $1/\lambda^*$ from these two sets are given in Table 2, Columns 4 and 5.

U^2: *Case* 3. For Case 3, the standard set $\lambda_i = 4\pi^2 i^2$ cannot be included, and the zeros of $D_3(\lambda)$ are those of $S_a(\lambda)$ and of $S_b(\lambda)$, given already for Cases 1 and 2. For all three cases, cumulants, calculated from (4), are in Table 3; the variances agree exactly with those in Table 1.

A^2: *Cases* 1 *and* 2. For A^2, the standard λ_i are $i(i + 1)$, $i = 1, 2, \cdots$; the functions $f_i(x)$ are $P_i^1(2x - 1)$, where $P_i^1(t)$ are Ferrer associated Legendre functions (Anderson and Darling (1952); note that the λ_i and $f_i(x)$ are misprinted). For Cases 1 and 2, $a_i = 0$ for i even, and $b_i = 0$ for i odd, and the solution is similar to that of W^2. Thus in Case 1, the standard $\lambda_i = 2i(2i + 1)$, $i = 1, 2, \cdots$ are solutions, and the other solutions are those of $S_a(\lambda_i^*) = 0$; reciprocals of these are given in Table 2, Column 7. For Case 2, $\lambda_i = 2i(2i - 1)$ are solutions, and the values of $1/\lambda_i^*$, where $S_b(\lambda_i^*) = 0$, are in Table 2, Column 8.

A^2: *Case* 3. As with W^2 and U^2, the weights for A^2, Case 3, are the λ_i^* of Cases 1 and 2. The cumulants, calculated from the weights, are given in Table 3. The variances cannot now be checked against exact calculations; calculations of the variance involving numerical integration of double integrals, were found from (3), and differed only in the third decimal place. The values in Table 3 are considered more reliable since they involve only single integrals calculated numerically (the a_i and b_i).

9. Tests for exponentiality. For Case 4, $F(x, \theta) = 1 - \exp(-\theta x)$; θ is estimated by $1/\bar{x}$. Then $\phi(s)$ for W^2 is $(1 - s) \ln (1 - s)$; since the Gaussian process is symmetrical on (0, 1) we can make calculations easier by substituting $1 - s$ for s, and using $\phi(s) = s \ln s$ (Darling, 1955). Covariance functions for the three statistics then become

$$\text{for} \quad W^2: \quad \rho(s, t) = r - st \ln s \ln t ,$$
$$\text{for} \quad U^2: \quad \rho(s, t) = r + a + m ,$$
$$\text{for} \quad A^2: \quad \rho(s, t) = w(r - st \ln s \ln t) ,$$

where a and w are given in Section 3.1 and m is $-(s \ln s + 0.25)(t \ln t + 0.25)$. Straightforward calculations give the mean and variance for W^2 (first given by Darling) and those for U^2. For A^2 the necessary integrals are somewhat more complicated; details are given in Stephens (1971). The results for means and variances are given in Table 1.

Calculation of weights. The parameter θ is a scale parameter, and Case 4 is similar to Case 2. Let the Fredholm determinant be $D_4(\lambda)$; it equals $D_0(\lambda)S_b(\lambda)$, using the notation of Section 7, with the b_i calculated from the appropriate $\phi(s)$ for Case 4. These functions are such that no b_i is now zero. For W^2 and A^2, this means that no standard weight is a zero of $D_4(\lambda)$; the zeros λ_i^* are those of $S_b(\lambda)$. For U^2 the standards occur once, in addition to the λ_i^*. Reciprocals of λ_i^* are included in Table 2, and the cumulants and β_1, β_2 values in Table 3. Once again the variances check perfectly with the results of Table 1.

10. Calculation of percentage points. Where possible, the first four cumulants have been used to fit Pearson curves to the distributions, and to give the uppertail percentage points. For smooth curves such as these, with long tails, the points can be expected, from past experience, to be very accurate. A check is provided by plotting points for finite n, found from Monte Carlo studies,

TABLE 4
Upper tail percentage points for asymptotic distributions

Statistic	Case	α (%) 15.0	10.0	5.0	2.5	1.0
W^2	0	.284	.347	.461	.581	.743
	1	.118	.135	.165	.196	.237
	2	.265	.329	.443	.562	.723
	3	.091	.104	.126	.148	.178
	4	.149	.177	.224	.273	.337
U^2	0	.131	.152	.187	.221	.267
	1	.111	.128	.157	.187	.227
	2	.106	.123	.152	.182	.221
	3	.085	.096	.116	.136	.163
	4	.112	.130	.160	.191	.230
A^2	0	1.610	1.933	2.492	3.070	3.857
	1	.784	.897	1.088	1.281	1.541
	2	1.443	1.761	2.315	2.890	3.682
	3	.560	.632	.751	.870	1.029
	4	.918	1.070	1.326	1.587	1.943

against $1/n$ (or $1/n^{\frac{1}{2}}$) and extrapolating to $1/n = 0$. This was done before the Pearson curve fits were made, and the results agree extremely well. Where the values of β_1, β_2 are beyond the range of the Pearson curve tables (for W^2 and A^2, Cases 2 and 4), an approximation of the form $a + b\chi_p{}^2$ has been used; the values of a, b, p are chosen to match the first three cumulants. These are expected also to be quite accurate. The upper 10, 5, 2.5, and 1 percent points for all statistics, all cases, are given in Table 4. Some of these points differ slightly from those given in Stephens (1974), where incorrect parameters were sometimes used in fitting Pearson curves; a correction is planned.

As described in Section 6, Kac et al. used only the first four λ_i in the approximation $S^* = \sum_{i=1}^{4} z_i/\lambda_i$ for the asymptotic distribution for W^2, Case 3. This gives a mean 0.0415 and a variance 0.00113; both are too low, as must be expected, and they give significance points which are too low. The S^* approximation gives roughly 0.086, 0.109 and 0.153 for the 10%, 5%, and 1% points. These differ from the values in Table 4 by roughly the difference in the S^* mean and the true mean.

11. Comments on Table 4. The points for Case 0, calculated from known asymptotic distributions, are included in Table 4 for comparison. It is clear that when one is allowed to improve the fit by estimating one or more parameters, the values of the goodness-of-fit statistics, even asymptotically, become stochastically much smaller. In testing for normality, estimating the mean makes a much greater improvement to the fit, in general, than estimating the variance, particularly as measured by the statistics W^2 and A^2. Fuller details of the associated tests are in Stephens (1974); in particular modifications are given to enable the asymptotic points to be used in practical tests with n finite.

Acknowledgments. This work was partly supported by the National Research Council of Canada, and by the U. S. Office of Naval Research Contract N00014-67-A-0112-0053 (NR-042-267). The author thanks both these agencies for their support. He also acknowledges with thanks helpful conversations with Dr. J. Farina, and some detailed comments from the Editor.

REFERENCES

[1] 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.

[2] ANDERSON, T. W. and DARLING, D. A. (1954). A test for goodness-of-fit. *J. Amer. Statist. Assoc.* **49** 300–310.

[3] DARLING, D. A. (1955). The Cramér-Smirnov test in the parametric case. *Ann. Math. Statist.* **26** 1–20.

[4] DURBIN, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.* **1** 279–290.

[5] KAC, M., KIEFER, J. and WOLFOWITZ, J. (1955). On tests of normality and other tests of goodness-of-fit based on distance methods. *Ann. Math. Statist.* **26** 189–211.

[6] STEPHENS, M. A. (1971). Asymptotic results for goodness-of-fit statistics when parameters must be estimated. Stanford Research Report No. 180, Dept. of Statistics, Stanford Univ.

[7] STEPHENS, M. A. (1974). EDF statistics for goodness-of-fit and some comparisons. *J. Amer. Statist. Assoc.* **69** 730–737.

[8] SUKHATME, SHASHIKALA (1972). Fredholm determinant of a positive definite kernel of a special type and its application. *Ann. Math. Statist.* **43** 1914–1926.

[9] WATSON, G. S. (1961). Goodness-of-fit tests on a circle. *Biometrika* **48** 109–114.

DEPARTMENT OF MATHEMATICS
MCMASTER UNIVERSITY
HAMILTON, ONTARIO
CANADA