ASYMPTOTIC RESULTS FOR GOODNESS-OF-FIT STATISTICS WHEN PARAMETERS MUST BE ESTIMATED

by

M.A. STEPHENS

TECHNICAL REPORT NO. 159
AUGUST 10, 1970

PREPARED UNDER CONTRACT N00014-67-A-0112-0053
NR-042-267
OFFICE OF NAVAL RESEARCH

Reproduction in Whole or in Part is Permitted for any purpose of the United States Government

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ASYMPTOTIC RESULTS FOR GOODNESS-OF-FIT STATISTICS WHEN PARAMETERS MUST BE ESTIMATED

by

M.A. Stephens
McGill University and Stanford University

Summary.

The paper considers the asymptotic distribution of the goodness-of-fit statistics \( w^2 \) and \( u^2 \) when parameters of the tested distribution must be estimated from the sample. The asymptotic mean and variance are found for the cases when, in testing for normality, either \( \mu \) or \( \sigma^2 \) or both must be estimated, and when, in testing for exponentiality, the scale parameter must be estimated. Section 2 also gives a list of integrals connected with the normal distribution, which might have other uses.

1.1 Introduction.

Let \( x_1, x_2, \ldots, x_n \) be independent observed 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

\[ H_0 : G(x) = F(x; \theta) , \quad (1) \]

where \( F(x; \theta) \) contains a parameter \( \theta \) (a vector, each of whose components represents a separate scalar parameter) is based on the Cramer-von Mises statistic
\[ W_n^2 = n \int_{-\infty}^{\infty} (F_n(x) - F(x; \theta))^2 f(x; \theta) \, dx. \] \hspace{1cm} (2)

Large values of \( W^2 \) (the subscript will be dropped) will suggest a poor fit. When \( \theta \) is completely specified, the asymptotic null distribution of \( W^2 \) is known; (Smirnov (1936), Anderson and Darling (1952)); for small samples, some exact results are known, and excellent approximations exist, so that \( W^2 \) has been well tabulated (Marshall, 1958), Pearson and Stephens (1962), Stephens and Maag (1968)). The convergence to the asymptotic points is rapid; for small \( n \), modifications to the calculated value of \( W^2 \) have been given (Stephens, 1970) to make only the asymptotic percentage points necessary for the goodness-of-fit test.

1.2 Unknown Parameters.

When \( \theta \) is not known, but must be estimated from the data, the null distribution of \( W^2 \) is changed and the tables above cannot be used. Two simultaneous papers, by Darling (1955), and by Kac, Kiefer and Wolfowitz (1955) explored this situation. Of the many results in these papers, an important one is the following (Darling (1955)): provided \( \theta \) consists of location and/or scale parameters, and provided its estimator satisfies certain conditions, the asymptotic distribution of \( W^2 \) depends only on the family of distributions to which \( F(x; \theta) \) belongs, and not on the value of \( \theta \). In particular, if \( F(x; \theta) \) is the normal family, \( (\theta = (\mu, \sigma^2)) \), or the exponential family, \( F(x; \theta) = 1 - \exp(-\theta x), (\theta = \theta) \),
and if the parameter $\theta$ is estimated by maximum likelihood, the result applies; in fact, the distribution, for any $n$, does not depend on $\theta$ (see, e.g., David and Johnson (1948)). It thus becomes attractive to try to find the null distribution, at least in these two important special cases, so that the test for goodness-of-fit can still be applied in these practical situations. For small samples, the distributional problems are very difficult and have not been solved analytically, but for the asymptotic distribution, an analytic attack is possible by showing that the asymptotic distribution of $W^2$ is that of a functional of a Gaussian process. Both Darling (1955) and Kac, Kiefer and Wolfowitz (1955) show that the asymptotic distribution is that of a sum $S = \sum_{i=1}^{\infty} \lambda_i Z_i$, where the $Z_i$ are independent $\chi^2_1$ variables, and where $\lambda_i$ are weights. Darling shows how to find the characteristic function of the $W^2$ distribution when only one parameter is to be estimated, though in the particular problems of testing for normality and exponentiality it is almost impossible to invert the characteristic function to get the distribution; Kac, Kiefer and Wolfowitz, discussing normality only, give the Gaussian process to be considered when both $\mu$ and $\sigma^2$ are to be estimated, and give a technique to calculate the $\lambda_i$. This is also very difficult, though they calculate $\lambda_1$, $\lambda_2$, $\lambda_3$, $\lambda_4$, and approximate the $S$ distribution by the distribution of $S^* = \sum_{i=1}^{4} \lambda_i Z_i$. We examine this approximation later (section 3.6). Darling gives also explicit integrals to use for calculating the asymptotic mean and variance of $W^2$; in theory, this could be extended to get higher cumulants. As an illustration, Darling gives the asymptotic mean (0.092523) and variance

3
\[ (0.004-3566) \text{ of } W^2 \text{ for the test for exponentiality. In this paper, we apply Darling's technique to the Gaussian process considered by Kac, Kiefer and Wolfowitz, and also to other Gaussian processes giving the asymptotic distribution of } U^2, \text{ an adaptation of } W^2 \text{ introduced by Watson (1961): the definition of } U^2, \text{ adapted for the case of an unknown parameter } \theta, \text{ is}
\]
\begin{equation}
U^2 = n \int_{-\infty}^{\infty} \left( F_n(x) - F(x; \theta) \right)^2 dF(x; \theta) + \int_{-\infty}^{\infty} \left( F_n(t) - F(t; \theta) \right)^2 dF(t; \theta), \tag{3}
\end{equation}

Watson (1961) introduced the statistic for observations on a circle, since its value does not depend on the choice of origin, but it may be used also for observations on a line. \( U^2 \) will tend to give significant values, when the null hypothesis completely specifies \( F(x; \theta) \), (the case considered by Watson), if the sample actually comes from a distribution with a different variance from the null, whereas \( W^2 \) will be stronger in detecting a shift in mean.

1.3 Contents.

In this paper are given:

(a) the mean and variance of the asymptotic distribution of \( W^2 \) (henceforth all means, variances and distributions of \( W^2 \) and \( U^2 \) refer to the asymptotic case), when \( F(x; \theta) \) is the normal distribution and when either \( \mu \) or \( \sigma^2 \) or both are to be estimated from the data;
(b) an examination of the approximation $S^*$ to the distribution of $W^2$;

(c) the mean and variance of $U^2$, when $F(x; \theta)$ is the normal distribution, parameters to be estimated, and also when $F(x; \theta)$ is the exponential distribution;

(d) some definite integrals connected with the normal distribution.

These are needed to calculate the values in (a) and (c) above, but might be useful also in other contexts.

The results have been used, together with Monte Carlo studies, to give a complete discussion of the tests for normality and for exponentiality (Stephens, 1969a, 1969b); in these papers modifications are again made to the calculated values of $W^2$ and $U^2$ (also, to those of the Kolmogorov statistic $D$ and the Kuiper statistic $V$), to make the actual test dependent only on the asymptotic percentage points; it is therefore important to have as much information on the asymptotic results as possible.

2. Integrals Connected with the Normal Distribution.

The integrals below are used in the sequel. They are similar to those occurring in the theory of normal order statistics and perhaps will be useful in other contexts.

2.1 Notation.

Throughout the paper, the following notation will be used,
(a) Integrals will be definite integrals; when limits are not shown, they are $-\infty$ or $\infty$.

(b) Symmetry is used to show that an integral over the half-plane (usually $y \geq x$) is set equal to half the integral over the whole plane because interchanging $y$ and $x$ does not change the integrand; order indicates that the order of integration has been changed.

(c) Define $d = 1/\sqrt{2x}$; $n(x) = \exp(-x^2/2)$; $N(x) = \int_n^x n(t) dt$.

When $s = N(x)$, let $x = J(s)$; i.e., $J$ is the inverse of $N$; this substitution will be used. Then $ds = n(x)dx$, and $n(x) = d \exp(-J^2(s)/2)$. Further notation is introduced in section 3.4.

2.2 Integrals.

The first equality is the definition of the integral.

$$A_1(k) \equiv \int n^k(x)dx = x^k / \sqrt{k}.$$  

$$A_2(k) \equiv \int n^k(x)N(x)dx = \frac{x^{k-1}}{2\sqrt{k}}.$$  

$$A_3(a,b) \equiv \int n^b(x)N(x)N(ax)dx = x^{b+1}(\tan^{-1}(a/\sqrt{b})(1+b+a^2)^{1/2})/\sqrt{b} + a^{b-1/4}\sqrt{b}.$$  

Proof.

$$\frac{dA_3}{ds} = \int n^b(x)N(x)N(ax)dx = \int x n^b(x)n(ax)\int n(t)dt dx$$  

$$= \int n(t)\int x n^b(x)n(ax)dx \quad \text{(order);}$$
\[ = \frac{\text{d}^{b+1}}{(b+a^2)^{1/2}} \int n(t) \exp(-t^2(b+a^2)/2) dt \]

Integration then gives \( A_j(a,b) \) as above.

\[ A_4(k,\gamma) = \int \gamma x n^k(x)N(x)dx = -d^kN(y)(\exp(-ky^2/2))/x \]

\[ + \frac{d^k}{k} \int \gamma \exp(-x^2(1+k)/2)dx \]

\[ = -d^kN(y)(\exp(-ky^2/2))/k + d^k(N((1+k)^{1/2}y))/k(1+k)^{1/2} \]

\[ \therefore A_4(k,\infty) = \frac{d^k}{k}(1+k)^{1/2} \]

\[ A_5(k,b) = \int x n^k(x)dx = \frac{(x-1)}{b} A_5(k-2,b) \]

Then

\[ A_5(0,b) = \frac{\text{d}^{b-1}}{\sqrt{b}} \quad A_5(1,b) = \frac{\text{d}^{b-1}}{b^{1/2}} \quad A_5(2,b) = \frac{\text{d}^{b-1}}{b^{3/2}} \]

\[ A_6(x,x) = \int x n^k(t)dt = (\text{d}^{k-1}/\sqrt{k})(1-N(\sqrt{k}x)) \]

3.1 The Statistic \( \hat{W}^2 \). Previous results.

(See, e.g. Anderson and Darling (1952); Darling (1955, particularly lemmas 3.1, 3.2); Kac, Kiefer and Wolfowitz (1955)).

(a) Subject to regularity conditions on \( F(x; \theta) \) and on the distribution of the estimator of \( \theta \), the asymptotic distribution of
\( W^2 \) is that of \( \int_0^1 (Y(t))^2 \, dt \), where \( Y(t) \) is a Gaussian process with covariance function \( \rho(s,t) \), depending on the parameters estimated, and the estimation procedure.

(b) When the parameter \( \theta \) in \( F(x; \theta) \) is completely specified, \( \rho(x,t) \) is \( r(s,t) \) where

\[
r(s,t) = \min(s,t) - st .
\]  

(c) The asymptotic cumulants of \( W^2 \) are given by

\[
k_j = 2^{j-1}(j-1)! \int_0^1 \rho_j(s,s) \, ds
\]

where \( \rho_j(s,t) \) is the \( j \)-th iterate of \( \rho(s,t) \). Then the mean and variance (these terms refer to the asymptotic mean and variance) are

\[
\mu = k_1 = \int_0^1 \rho(s,s) \, ds
\]

and

\[
\sigma^2 = k_2 = 2 \int_0^1 \int_0^1 \rho(s,t) \rho(t,s) \, ds \, dt .
\]

(d) In the applications to follow, \( \rho(s,t) \) has the form

\[
\rho(s,t) = \min(s,t) - st - Z(s,t) .
\]

Darling (1955, Theorems 4.1 and 5.1) shows that when \( \theta \) has only one component \( \theta \), estimated by an efficient estimator or by an asymptotically efficient maximum likelihood estimator, then \( Z(s,t) \) factors into

\[
Z(s,t) = \Phi(s) \Phi(t) .
\]
\( \phi(s) \) is found as follows. Let \( u = F(x; \theta) \) define \( x \) implicitly in terms of \( u \). Let \( f(x; \theta) = \frac{\partial}{\partial x} F(x; \theta) \); \( g(u) = \frac{\partial}{\partial \theta} F(x; \theta) \), and let \( k^2 \) be defined by

\[
\frac{1}{k^2} = \int \left( \frac{\partial}{\partial \theta} \ln f(x; \theta) \right)^2 f(x; \theta) dx ;
\]

then

\[
\phi(u) = k^2 g(u) ,
\]

and

\[
\phi'(u) = k^2 \frac{\partial}{\partial \theta} \ln f(x; \theta) .
\]

Note that the presence of \( k \) in the definition of \( \phi(u) \) ensures that

\[
\int_0^1 (\phi'(u))^2 du = 1 .
\]

Darling (1955, Theorem 7.1) also shows that when \( \theta \) is a location or scale parameter, \( \phi(u) \) will not depend on \( \theta \), and the asymptotic distribution of \( W^2 \) will also not depend on \( \theta \). The mean and variance then become

\[
u = \frac{1}{6} - \int_0^1 \phi^2(s) ds
\]

and

\[
\sigma^2 = \frac{1}{45} - 8 \int_0^1 \phi(t)(1-t) \int_t^1 \phi(s) ds dt + 2(\mu - \frac{1}{6})^2 .
\]

3.2 The Statistic \( W^2 \); Test for normality.

We now examine the asymptotic distribution of \( W^2 \), when \( F(x; \theta) \) is the cumulative normal distribution, mean \( \mu \), variance \( \sigma^2 \), in three cases; Case 1, when \( \sigma^2 \) is known and \( \mu \) is estimated from the
the sample observations \( x_1, x_2, \ldots, x_n \) by \( \bar{x} \); Case 2, when \( \mu \) is known, and \( \sigma^2 \) is estimated by \( s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \); Case 3, when both \( \mu \) and \( \sigma^2 \) are unknown, and are estimated by \( \bar{x} \) and \( s^2 \). The asymptotic mean and variance of \( \hat{w}^2 \), in Case 1, will be called \( \mu_1 \) and \( \sigma_1^2 \).

Case 1. Suppose \( \sigma^2 \) is known, equal to \( \sigma_0^2 \); substitute \( y_i = x_i / \sigma_0 \), and the test reduces to a test that the \( y_i \) are \( N(\mu, 1) \); so without loss of generality we may take \( \sigma_0 \) as 1. The unknown \( \theta \) is now \( \mu \), \( F(x; \theta) = N(x-\theta) \); \( \theta \) is a location parameter, so \( \Phi(u) \) may be found, independent of \( \theta \), as follows. We have \( \frac{\partial}{\partial \theta} F(x; \theta) = -n(x-\theta) \). From (10) and (11), \( k^2 = 1 \) and \( \Phi(u) = -n(x-\theta) \), i.e.,

\[
\hat{\phi}(u) = -d \exp(-J^2(u)/2)
\] (16)

Case 2. Suppose \( \mu \) is known, equal to \( \mu_0 \); substitute \( y_i = x_i - \mu_0 \) and the test becomes the test that the \( y_i \) are from \( N(0, \sigma^2) \), so we may take \( \mu_0 \) as zero. Now \( u = d \int x((\exp(-t^2/2\theta))/\sqrt{2\theta}) dt \), i.e.,

\[
x = \sqrt{2} \int \hat{\phi}(u) \cdot \exp(-J^2(u)/2) \cdot \sqrt{2}
\] (17)

3.3 Covariance Function \( \rho(s,t) \).

These results give, using (8) and (9),

Case 1: \( \rho(s,t) = r(s,t) - d^2 \exp(-(J^2(s)+J^2(t))/2) \). (18)

Case 2: \( \rho(s,t) = r(s,t) - (d^2/2)J(s)J(t) \exp(-(J^2(s)+J^2(t))/2) \). (19)

For Case 3, the covariance function was given by Mac, Kiefer and Wolfowitz (1955):
Case 3: \[ \rho(s,t) = r(s,t) - d^2(1 + J(s)J(t)/2) \exp(-(J^2(s) + J^2(t))/2). \] (20)

It will now be useful to introduce further notation. Let

\[ b(s,t) = -d^2 \exp(-(J^2(s) + J^2(t))/2); \]

and

\[ c(s,t) = -(J(s)J(t)/2). \]

Then the three covariances above are, respectively (omitting the arguments \( s, t \)) \( r+b, r+c, r+b+c. \)

3.4. Notation for definite integrals.

Let

\[ R^2 = \int_0^1 \int_0^1 r^2(s,t)dsdt; \quad RB = \int_0^1 \int_0^1 r(s,t)b(s,t)dsdt, \text{ etc.,} \]

the capital letters indicating which functions form the integrand.

3.5. Means and Variances, \( \nu^2 \), in testing for normality.

Case 1. The mean is, by (6),

\[ \mu_1 = \frac{1}{c} - \frac{d^2}{c} \int_0^1 \exp(-J^2(s))ds. \]

Let \( x = J(s); \)

\[ \mu_1 = \frac{1}{c} - \frac{d^2}{c^2} \int n^2(x)n(x)dx = \frac{1}{c} - \frac{d^2}{\sqrt{\frac{3}{c}}}, \quad .074778. \]

The variance is, from (7),

\[ \sigma^2 = 2(R^2 + 2RB + B^2) \] (21)
\[ 2R^2 = \frac{1}{4y}, \text{ and } 2B^2 = 2(\mu - \frac{1}{6})^2; \text{ 4RB must be found. Using the substitution } x = J(s), \ y = J(t), \text{ (here and throughout the rest of the paper),} \]

\[ RB = 2 \int_0^1 \int_0^t s(1-t) b(s,t) ds dt \]

\[ = -2 \int (1-N(y)) n^2(y) \int^y N(x) n^2(x) dx dy \]

\[ = -2 \int^y n^2(y) N(x) n^2(x) dx dy + 2 \int^y N(y) N(x) n^2(y) n^2(x) dx dy \]

\[ = I_1 + I_2, \text{ say.} \]

In \[ I_1, \] change the order of integration:

\[ I_1 = -2 \int n^2(x) N(x) \int_x^y n^2(y) dy, \]

\[ = -\sqrt{2} \int n^2(x) N(x) (1 - N(\sqrt{2} x)) dx, \text{ using } A_6(2, x); \]

\[ = -\sqrt{2} \left(A_2(2) - A_3(\sqrt{2}, 2)\right), \]

\[ = -\frac{d^2}{2} + d^4 \tan^{-1} \left(\frac{1}{\sqrt{5}}\right) + \frac{d^2}{4} = -\frac{d^2}{4} + d^4 \tan^{-1} \left(\frac{1}{\sqrt{5}}\right) \]

Also,

\[ I_2 = \int \int N(y) n^2(y) N(x) n^2(s) dx dy \quad \text{(Symmetry)}; \]

\[ = (A_2(2))^2 = \frac{d^2}{8}. \]
Thus

\[ 4RB = -\frac{1}{4\pi} + \frac{1}{2} \tan^{-1} \frac{1}{\sqrt{5}} = -0.036970; \]

substitution in (21) gives \( q^2_1 = 0.002139. \)

**Case 2.** The mean is

\[ \mu^2 = \frac{1}{6} - \frac{d^2}{2} \int_0^1 J^2(s) \exp(-J^2(s)) ds \]

\[ = \frac{1}{6} - \frac{1}{12 \sqrt{3} \pi} = 0.15135. \]

The variance is

\[ c^2_2 = 2(R^2 + 2RC + c^2); \quad (22) \]

\[ 2R^2 = \frac{1}{4\pi}, \quad \text{and} \quad c^2 = (\mu - \frac{1}{6})^2 = \frac{1}{1432 \pi^2}; \quad 4RC \text{ is needed.} \]

\[ RC = -d^2 \int_0^1 \int_0^t (1-t)J(s)J(t) \exp(-(J^2(s) + J^2(t))/2) ds dt. \]

Using \( x = J(s), y = J(t), \)

\[ RC = -\int_0^y xy n^2(x)n^2(y)N(x) dx dy + \int_0^y xy n^2(x)n^2(y)N(y)N(x) dx dy \]

\[ = -I_3 + I_4. \]

Each integral separately gives
\[ I_2 = \int y n^2(y) \int x n^2(x) N(x) dx dy = \int y n^2(y) A_{h}(2,y) dy \]

finally

\[ I_3 = 1/(32\pi^2 \sqrt{5}) , \]

using integration by parts.

\[ I_4 = \frac{1}{2} \int \int xy n^2(x) n^2(y) N(x) N(y) dx dy \quad \text{(symmetry)} \]

\[ = \frac{1}{2} (A_4(2))^2 = 1/(96 \pi^2) . \]

Then \( R_C = -0.00144 \) and \( \sigma_2^2 = 0.02222 - 0.00105 + 0.00047 \) so

\[ \sigma_2^2 = 0.02125 . \quad (23) \]

We now examine the most important situation of the three.

Case 3. Both mean and variance are estimated. Then

\[ \mu_3 = \frac{1}{6} - d^2 \int_0^1 (1 + J^2(s)/2) \exp(-J^2(s)) ds \]

\[ = \frac{1}{6} - \frac{7}{12 \sqrt{3} \pi} = 0.05946 \quad (24) \]

and

\[ \sigma_2^2 = 2(R^2 + 2RE + 2RC + (E-C)^2) . \quad (25) \]

The new integral needed is \( (E-C)^2 \). This is
\[(B+C)^2 = \frac{4}{3} \int_0^1 \int_0^1 (1+\frac{J(s)J(t)}{2})^2 \exp(-\frac{J^2(s)}{2})\exp(-\frac{J^2(t)}{2}) dsdt\]

\[= \int (1+xy/2)^2 n^3(x)n^3(y) dydx\]

\[= \iint n^3(x)n^3(y) dxdy + \iint xy n^3(x)n^3(y) dxdy\]

\[+ \frac{1}{4} \iint x^2y^2 n^3(x)n^3(y) dxdy\]

\[= (A_1(3))^2 + 0 + \frac{1}{4} (A_5(2,3))^2.\]

Then

\[C^2 = (A_1(3))^2 = 1/(12 \pi^2) ;\]

\[BC = 0\]

and

\[C^2 = (A_5(2,3))^2/4 = 1/(432 \pi^2) ;\]

so \[(B+C)^2 = 37/(432 \pi^2) \text{ and } \frac{c^2}{\lambda_3} = 0.00116.\]

3.6. The Approximation \(S^*\).

The accuracy of the approximation \(S^*\) (Section 1.2) can now be examined. The values of \(\lambda_1\) were computed (Kac, Kiefer and Wolfowitz, 1955) to be 0.01836, 0.01346, 0.00536, 0.00436. This gives a mean of \(S^*\) equal to 0.0415 and a variance 0.00113, both are, as expected, too low, and we can expect the significance points of the approximation.
to be too low. This is found to be so: Monte Carlo results (extrapolated
from results for finite \( n \), given in more detail in Stephens (1969a),
give 10, 5 and 1 percent points to be 0.103, 0.125, 0.177, while
the \( S^* \) approximation gives roughly 0.086, 0.109, 0.153; the 5% point of \( S^* \) is approximately at the 9% level given by empirical re-
results. The differences in the values are roughly the difference
between the \( S^* \) mean (0.0415) and the true mean (0.0595).

4.1. The Statistic \( U^2 \).

Watson (1951) 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

\[
Q(t) = Y(t) - \int_0^1 Y(u) du
\]

\( Y(t) \) is the same Gaussian process as that used for the limiting
distribution of \( W^2 \), with covariance function \( \rho(s,t) \). Let the
covariance function of \( Q(t) \) be \( q(s,t) \); then

\[
q(s,t) = \rho(s,t) + \int_0^1 \int_0^1 \rho(s,t) ds dt - \int_0^1 \rho(s,t) ds - \int_0^1 \rho(s,t) dt
\]

\[
= \rho(s,t) + E_1 + E_2 + E_3, \text{ say .}
\]

Cases 1, 2, and 3 will describe the same three situations as for
\( W^2 \), when \( F(x; \theta) \) is the cumulative normal distribution. We find
\( q(s,t) \) first for Case 3, the most general situation.
Case 3. From (20),

\[ \rho(s,t) = \min(s,t) - st - d^2(1 + J(s)J(t)/2) \exp(-(J^2(s) + J^2(t))/2) \cdot \]

\[ E_1 = 2 \int_0^1 \int_0^t \rho(s,t) \, ds \, dt ; \text{ using } x = J(s), y = J(t), \]

we have

\[ E_1 = \frac{1}{12} - \int \int (1 + xy/2) n^2(x) n^2(y) \, dx \, dy \]

\[ = \frac{1}{12} - (A_1(2))^2 = \frac{1}{12} - \frac{1}{4\pi}. \]

Similarly, \(-E_2\) is \(\frac{1}{2}(s-t^2) - (d^2/\sqrt{\pi}) \exp(-J^2(t)/2)\)

and \(-E_3\) is \(\frac{1}{2}(s-s^2) - (d^2/\sqrt{\pi}) \exp(-J^2(s)/2)\).

Definition. Define

\[ a(s,t) = \frac{1}{12} - \frac{1}{2}(s-s^2) - \frac{1}{2}(t-t^2) \]

and

\[ d(s,t) = -\frac{1}{4\pi} + \frac{d^2}{2} \left( \exp(-J^2(s)/2) + \exp(-J^2(t)/2) \right). \]

Then, for \(U^2\), Case 3, the covariance function is

\[ q(s,t) = \rho(s,t) + a(s,t) + d(s,t); \]

omitting arguments \(s,t\) on the right hand side, and using the notation of section 3.3, we may write

\[ q(s,t) = r + b + c + a + d. \]
Case 1. For Case 1, $\rho(s,t) = r + b$, and

$$ q(s,t) = r + b + a + d $$

Case 2. For Case 2, $\rho(s,t) = r + c$ and

$$ q(s,t) = r + c + a. $$


The calculations for $U^2$ follow the same lines as for $W^2$, and we use the same notation for definite integrals. For the most difficult variance, $\sigma^2_2$, we have

$$ \sigma^2_2 = (2r+A+B+C+D)^2; $$

$2(R+A)^2$ is the variance for the case when no parameters need be estimated, and is $1/180$; new integrals required, not yet found for $W^2$, are $D^2$, $RD$, $AB$, $AD$, $BD$, $AC$, $CD$. These will now be found.

$$ D^2 = \int \int \left( \frac{d^2}{2} \right) (n(x)+n(y))^2 n(x)n(y) dx dy $$

$$ = \frac{d^2}{4} \int \int (d^2-2\sqrt{2}d(n(x)+n(y)) + 2n^2(x)+n^2(y)+2n(x)n(y))) n(x)n(y) dx dy $$

$$ = \frac{d^2}{4} \left( \frac{1}{4} - \frac{3}{2} (A_1(2)+A_1(2)) + \frac{d^2}{2} (A_1(3)+A_1(3)+2(A_1(2))^2) $$

$$ = d^4 \left( \frac{1}{4} - \frac{1}{4} \right). $$

$$ RD = -2 \int_0^1 \int_0^t s(1-t) \frac{d^2}{2} - \frac{d^2}{\sqrt{2}} \left( \exp(-J^2(s)/2)+\exp(-J^2(t)/2) \right) ds dt $$

18
\[= - \iint_{x-y} N(x)(1-N(y)) \left( d^2 - \sqrt{2}dn(x) - \sqrt{2}dn(y) \right) n(x)n(y) \, dx \, dy \]

\[= I_5 + I_6 \quad \text{where} \]

\[I_5 = \iint_{y=0}^{y=\infty} N(y)N(x) \left( d^2 - \sqrt{2}dn(x) - \sqrt{2}dn(y) \right) n(x)n(y) \, dx \, dy \]

\[I_5 = \frac{1}{2} \iint \text{integrand} \, dx \, dy \quad \text{(symmetry);} \]

\[= \frac{1}{2} \left( A_2'(1) \right)^2 - \sqrt{2} d A_2(2) A_2(1) - \sqrt{2} d A_2(2) A_2(1) \]

\[= -\frac{d^2}{8} = -1/16\pi. \]

\[I_6 = - \iint_{x-y} N(x)(d^2 - \sqrt{2}dn(x) - \sqrt{2}dn(y)) n(x)n(y) \, dx \, dy \]

\[= - \iint \n(x)n(x)(d^2 - \sqrt{2}dn(x)) n(y) \, dy \, dx + \sqrt{2} d \iint N(x)n(x)n^2(y) \, dx \, dy \]

\[= - \iint \n(x)n(x)(d^2 - \sqrt{2}dn(x))(1-N(x)) \, dx + \frac{d}{\sqrt{2}} \int n^2(y) \, dy \]

\[= -d^2 A_2(1) + \frac{d^3}{3} + \sqrt{2} d A_2(2) - \sqrt{2} d A_2(1,2) + d A_2(1,2); \]

\[= -\frac{d^2}{2} + \frac{d^3}{3} + \frac{d^2}{2} - \frac{d}{\sqrt{2}} \left( \frac{d^3}{3} \tan^{-1} \left( \frac{1}{2\sqrt{2}} \right) + \frac{d}{4\sqrt{2}} \right). \]

So

\[RD = \frac{d^2}{12} - \frac{d^3}{2} \tan^{-1} \left( \frac{1}{2\sqrt{2}} \right). \]

Next,
\[ AB = -d^2 \int_0^1 \int_0^1 \left( \frac{1}{12} - \frac{1}{2}(s-s^2) - \frac{1}{2}(t-t^2) \right) \exp(-\frac{j^2(s)+j^2(t)}{2}) \, ds \, dt \]

\[ = -\iint \left( \frac{1}{12} + \frac{1}{2}(N^2(x)+N^2(y)) \right) n^2(x)n^2(y) \, dx \, dy , \]

\[ = -\frac{1}{12} \left( A_1(2) \right)^2 - \frac{1}{2}(A_1(t) \{2A_2(1,2)) \right) \]

\[ = -\frac{d^2}{24} - \frac{d}{\sqrt{2}} \left[ \frac{d^2}{\sqrt{2}} \tan^{-1} \left( \frac{1}{2} \frac{1}{\sqrt{2}} + \frac{d}{4} \frac{1}{\sqrt{2}} \right) \right] ; \]

\[ AD = -\frac{d^2}{2} \int_0^1 \int_0^1 \left( \frac{1}{12} - \frac{1}{2}(s-s^2) - \frac{1}{2}(t-t^2) \right) (d^- \sqrt{2} \exp(-j^2(s)/2) - \sqrt{2} \exp(j^2(t)/2)) \, ds \, dt \]

\[ = -\frac{d}{2} \iint \left( \frac{1}{12} + \frac{1}{2}(N^2(x)+N^2(y)) \right) (d^- \sqrt{2} n(x) - \sqrt{2} n(y)) n(x) n(y) \, dx \, dy \]

\[ = -\frac{d^2}{24} + \frac{d^2}{12} - \frac{d}{4} \iint dN^2(x) n(x) n(y) \, dx \, dy - \iint \sqrt{2} N^2(x) n^2(x) n(y) \, dx \, dy \]

\[ -\iint 2 \sqrt{2} N^2(x) n^2(y) n(x) \, dx \, dy ; \text{ using symmetry} ; \]

\[ = +\frac{d^2}{24} - \frac{d}{4} \frac{2a}{3} - 2 \sqrt{2} \frac{A_2(1,2)}{-2} \frac{2d}{3} \]

\[ = \frac{d^2}{6} + \frac{d}{2} \tan^{-1} \left( \frac{1}{2} \frac{1}{\sqrt{2}} \right) . \]

\[ BD = \frac{d}{2} \iint \left( \exp(-\frac{j^2(s)+j^2(t)}{2}) \right) (1 - \sqrt{2} \exp(-j^2(s)/2) - \sqrt{2} \exp(-j^2(t)/2)) \, ds \, dt \]

\[ = \frac{d}{2} \iint n^2(x) n^2(y) (d^- \sqrt{2} n(x) - \sqrt{2} n(y)) \, dx \, dy \]

\[ = \frac{d}{2} (d(A_1(2)) - 2 \sqrt{2} A_1(3) A_1(2)) \]

\[ = \frac{d}{4} - \frac{d}{\sqrt{2}} . \]
Integrals AC and CD may be shown to be zero, and we omit the proofs.

4.3. Asymptotic means and variances of $U^2$, in testing for normality.

Case 1. We have, using $q(s,t) = r + b + a + d$,

$$
\nu_1 = \frac{1}{12} - \frac{1}{2\sqrt{3}} - \frac{1}{4\pi} - \frac{1}{2\pi} = 0.07102 \; ;
$$

and

$$
\sigma_1^2 = 2(R+A+B+D)^2 \; ;
$$

use of the integrals previously found gives

$$
\nu_2 = \frac{1}{360} + \frac{1}{6\pi} + \frac{1}{8\pi^2} \left( \frac{1}{2\sqrt{3}} \pi - \frac{1}{\sqrt{2}} \right) + \left( - \frac{1}{4\pi} + \frac{1}{2\pi} \tan^{-1} \frac{1}{\sqrt{5}} \right)
$$

$$
+ \left( - \frac{1}{2\pi} \tan^{-1} \frac{1}{2\sqrt{2}} + \frac{1}{6\pi} \right) = 0.001948 \; .
$$

The successive terms in this expression correspond to $2(R+A)^2$, $2B^2$, $4RD+D^2$, $4RB$, $4RD$, $A(B+C)$ is zero.

Case 2. We have

$$
\nu_2 = \frac{1}{12} - \frac{x^2}{2} n^3(x) = \frac{1}{12} + \frac{1}{2} A_2(2,3)
$$

$$
= \frac{1}{12} - \frac{1}{2\sqrt{3}} = \frac{1}{12} - \frac{1}{12\sqrt{3} \pi} = 0.06802
$$

21
\[ \sigma_2^2 = 2(R+A+C)^2 \]
\[ = 2(R+A)^2 + 2C^2 + 4(R+A)C \]
\[ = \frac{1}{360} + \frac{1}{216\pi^2} + \left( \frac{1}{24\pi^2} - \frac{1}{8\pi^2 \sqrt{5}} \right) = 0.001305 . \]

The last bracket is \( 4RC \), since \( AC \) is zero.

Case 3. When both parameters are to be estimated, we have

\[ \mu_2 = \frac{1}{12} - \frac{1}{2\pi \sqrt{3}} + \frac{1}{4\pi} - \frac{1}{12\pi \sqrt{3}} = 0.05571 \]

and

\[ \sigma_2^2 = 2(R+A+B+C+D)^2 \]
\[ = \sigma_1^2 + 2C^2 + 4C(R+A+B+C) \]
\[ = \sigma_1^2 + 2C^2 + 4RC, \text{ other terms being zero;} \]
\[ = \sigma_1^2 + \frac{1}{216\pi^2} + \left( \frac{1}{24\pi^2} - \frac{1}{8\pi^2 \sqrt{5}} \right) = 0.000975 . \]

1.4 Table of Means and Variances.

Table 1 gives the asymptotic means and variance, for \( W^2 \) and \( U^2 \), in the various cases. For the test for normality, there is a very large drop in mean and variance, from the standard case when \( F(x; \theta) \) is completely specified, to Case 3, when both mean \( \mu \) and variance \( \sigma^2 \) must be estimated, and almost as large a drop when only \( \mu \) must be estimated; the drop is much less when only the variance \( \sigma^2 \) is estimated.
4.5 Asymptotic Means and Variances, \( W^2 \), in Testing for Exponentiality.

The null hypothesis is \( H_0 \): a sample \( x_1, x_2, \ldots, x_n \) comes from
\[
F(x) = 1 - \exp(-6x), \quad x \geq 0; \quad \theta \text{ is estimated by } 1/\bar{x}.
\]
Then, (Darling (1955)) \( \Phi(u) = u \log u \), and the covariance function for \( W^2 \) is
\[
r(s,t) = st \log s \log t.
\]
Following section 4.1 we have for the covariance
\[
g(s,t) \text{ of } W^2
\]
\[
g(s,t) = r(s,t) + g(s,t) + a(s,t) + m(s,t)
\]
where \( r(s,t), a(s,t) \) are as before, and
\[
g(s,t) = -st \log s \log t
\]
and
\[
m(s,t) = (0.25 - s \ln s - t \ln t)/4.
\]
The asymptotic mean and variance of \( W^2 \) then become
\[
\mu = 0.07176 \quad \text{and} \quad \sigma^2 = 0.00198.
\]

5. Further Remarks.

An examination of the calculations above, for both \( X^2 \) and \( W^2 \), shows that the change in mean, from the standard case (\( \theta \) specified) to case 3, is the sum of the changes for cases 1 and 2; the same is true for the variance. This illustrates an interesting result, stated below as a Lemma, without proof. It follows straightforwardly by applying Darling's (1955) Lemmas 3.1 and 3.2 to two parameters rather than one, and then carrying through his argument.
Preliminaries. In $F(x; \theta)$, let $\theta$ be $(\theta_1, \theta_2)$ where $\theta_1, \theta_2$ are respectively location and scale parameters; let these be estimated by maximum likelihood and let the conditions of Darling (1955, Lemmas 3.1 and 3.2) be satisfied. Let $\phi_1(u), \phi_2(u)$ be the functions $\Phi(u)$ obtained when $\theta_1, \theta_2$ are separately estimated; let the corresponding asymptotic mean of $W^2$ be then respectively $\mu_1 = 1/6 - c_1$, and $\mu_2 = 1/6 - c_2$; and let the corresponding variances be $\sigma_1^2 = 1/45 - v_1$ and $\sigma_2^2 = 1/45 - v_2$.

Lemma. The covariance function when both $\theta_1$ and $\theta_2$ are estimated is $\rho(s, t) = \min(s, t) - st\iint_1 \phi_1(s)\phi_1(t) - \phi_2(s)\phi_2(t)$; the asymptotic mean is

$$\mu_3 = 1/6 - c_1 - c_2;$$

if, further, the orthogonality condition

$$\int_0^1 \phi_1(s)\phi_2(s)ds = 0$$

is satisfied, the asymptotic variance is

$$\sigma_3^2 = 1/45 - v_1 - v_2.$$

A parallel result holds for $U^2$. 

24
TABLE 1.

Asymptotic means and variances of $W^2$ and $U^2$. N stands for the test for normality; E for exponentiality.

<table>
<thead>
<tr>
<th>Test</th>
<th>$W^2$</th>
<th>$U^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Variance</td>
</tr>
<tr>
<td>$F(x; g)$ specified</td>
<td>0.16666</td>
<td>0.02222</td>
</tr>
<tr>
<td>$N; \mu$ specified</td>
<td>0.1514</td>
<td>0.02125</td>
</tr>
<tr>
<td>$N; \sigma^2$ estimated</td>
<td>0.0748</td>
<td>0.0214</td>
</tr>
<tr>
<td>$N; \sigma^2$ specified</td>
<td>0.0595</td>
<td>0.0116</td>
</tr>
<tr>
<td>$N; \mu, \sigma^2$ estimated</td>
<td>0.0926</td>
<td>0.00436</td>
</tr>
</tbody>
</table>
REFERENCES


