The Annals of Statistics
1986, Vol. 14, No. 4, 1497–1506

ASYMPTOTIC DISTRIBUTION OF THE SHAPIRO–WILK W FOR TESTING FOR NORMALITY[1]

BY J. R. LESLIE, M. A. STEPHENS AND S. FOTOPOULOS

University of London, Simon Fraser University and Washington State University

Twenty years have elapsed since the Shapiro–Wilk statistic W for testing the normality of a sample first appeared. In that time a number of statistics that are close relatives of W have been found to have a common (known) asymptotic distribution. It was assumed, therefore, that W must have that asymptotic distribution. We show this to be the case and examine the norming constants that are used with all the statistics. In addition the consistency of the W test is established.

1. Introduction. A popular test for the normality of a random sample is based on the Shapiro–Wilk statistic W. This statistic, which was presented in Shapiro and Wilk (1965), is the ratio of the square of the BLUE of σ to the sample variance, where σ^2 is the variance of the normal population from which the sample is assumed, under the null hypothesis, to have been drawn. For convenience we shall work with $W^{1/2}$, which has the form

$$W^{1/2} = \mathbf{X}'V_0^{-1}\mathbf{m}\Big/\left\{\sum_1^n (X_i - \overline{X})^2 \mathbf{m}'V_0^{-1}V_0^{-1}\mathbf{m}\right\}^{1/2},$$

where $\mathbf{X} = (X_1, \dots, X_n)'$, $X_1 < X_2 < \cdots < X_n$, is the vector of order statistics from the sample, $\overline{X}$ is the sample mean, and $\mathbf{m}$ is the mean vector and V_0 the covariance matrix of standard normal order statistics. As $W^{1/2}$ is location and scale invariant we can assume henceforth that $X_1, \dots, X_n$ are order statistics for a sample from a $N(0,1)$ population.

A number of authors (for example, Sarkadi (1975), (1977); Gregory (1977)) have (correctly) guessed at the form of the asymptotic distribution for W as well as predicting that the test should be consistent. However, no rigorous proofs have been possible due to the presence of V_0^{-1}. Neither V_0 nor V_0^{-1} can be found explicitly and until recently no reasonably accurate asymptotic approximation for V_0 was available. A paper by one of the authors (Leslie (1984)) has now remedied the situation. In that paper an approximation for V_0 together with a number of asymptotic properties of V_0 can be found, one of which is of particular importance to this work. It states that $\mathbf{m}$ is approximately an eigenvector of V_0^{-1} in the following sense:

$$\|V_0^{-1}\mathbf{m} - 2\mathbf{m}\| \le C(\log n)^{-1/2}, \tag{1}$$

Received March 1985; revised December 1985.

[1]This work was partly supported by the Natural Sciences and Engineering Research Council of Canada.

AMS 1980 *subject classifications.* Primary 62F05, 62E20; secondary 62G30.

Key words and phrases. Shapiro–Wilk statistic, goodness of fit, normal order scores, tests of normality.

where C is a constant independent of n, and $\|\mathbf{b}\|^2 = \Sigma b_i^2$ for $\mathbf{b} = (b_1, \ldots, b_n)'$. This latter result formalises a similar one appearing in Stephens (1975).

The asymptotic distribution of W, after appropriate normalising, has been assumed to be the same as that of the De Wet and Venter (1972) statistic

$$W^* = r^2(\mathbf{X}, \mathbf{H});$$

here $r(\mathbf{X}, \mathbf{Y})$ is the sample correlation coefficient between $\mathbf{X}$ and $\mathbf{Y}$, $\mathbf{H}$ is the $n \times 1$ vector whose ith element is $\Phi^{-1}\{i/(n+1)\}$, and $\Phi^{-1}(\cdot)$ is the inverse function for the standard normal distribution function $\Phi(\cdot)$, that is, $\Phi^{-1}(\Phi(x)) = x$.

The rationale behind this assumption was that first, $V_0^{-1}\mathbf{m}$ was known to behave like $2\mathbf{m}$ (see Stephens (1975)); second, $\Phi^{-1}\{i/(n+1)\}$ approximates the ith element of $\mathbf{m}$, and third, as V_0 is a doubly stochastic matrix (the sum along any row or column is 1), we may write

$$W = r^2\big(\mathbf{X}, V_0^{-1}\mathbf{m}\big).$$

De Wet and Venter (1972) showed that the asymptotic distribution of W^* has the form

$$2n(1 - W^{*1/2}) - a_n \to_D \zeta, \tag{2}$$

where $\zeta = \Sigma_3^\infty (Y_i^2 - 1)/i$, $\{Y_i, i \geq 1\}$ is a sequence of i.i.d. $N(0,1)$ variates,

$$a_n = (n+1)^{-1}\left\{ \sum_{i=1}^{n} j(1-j)\big(\phi\{\Phi^{-1}(j)\}\big)^{-2} \right\} - \tfrac{3}{2}, \tag{3}$$

$j = i/(n+1)$, and $\phi(\cdot)$ is the $N(0,1)$ density function.

Beyond the De Wet and Venter result the first step toward the asymptotic distribution for W was to show that the Shapiro–Francia (1972) statistic $W^\dagger$, given by

$$W^\dagger = r^2(\mathbf{X}, \mathbf{m}),$$

behaves in the same way as W^*. This was done independently and via different routes by Verrill and Johnson (1983) and by the authors in Fotopoulos, Leslie and Stephens (1984) (henceforth called FLS), where expression (2) was established with $W^\dagger$ in place of W^*. In fact we show in FLS the equivalent result that

$$n(W^{*1/2} - W^{\dagger 1/2}) \to 0 \quad \text{in probability.} \tag{4}$$

Our task in the present paper is to show that

$$n(W^{1/2} - W^{\dagger 1/2}) \to 0 \quad \text{in probability.} \tag{5}$$

We note that Verrill and Johnson (1983) contains a result (Theorem 3) that should eventually cover the asymptotic distribution of W. However, certain properties of $V_0^{-1}\mathbf{m}$ need to be established before it can be applied. Inequality (1) does not appear to be enough.

2. Asymptotic properties of W and a_n. The following theorem presents one version of the asymptotic distribution for W—in fact the asymptotic distribution for $W^{1/2}$—whilst the corollary offers the complementary form in terms of W.

THEOREM. *Under the hypothesis that the observed sample is from a normal population, the asymptotic distribution of the Shapiro–Wilk W takes the form*:

$$2n(1 - W^{1/2}) - 2En(1 - W^{1/2}) \to_D \zeta,$$

where $\zeta = \Sigma_3^\infty (Y_i^2 - 1)/i$, *and* $\{Y_i, i \geq 3\}$ *is a sequence of i.i.d.* $N(0,1)$ *variables.*

From the following lemma and the theorem we have $\sqrt{n}(1 - W^{1/2}) \to 0$ in probability, which leads to

$$2n(1 - W^{1/2}) - n(1 - W) = \left(\sqrt{n}(1 - W^{1/2})\right)^2 \to 0 \quad \text{in probability.}$$

Again applying the lemma below we obtain

COROLLARY. *An equivalent form for the asymptotic distribution of W is*

$$n(W - EW) \to_D -\zeta.$$

It is not obvious from their definition just how the constants a_n will behave as n gets large. The following lemma should shed some light on this matter.

LEMMA. *The constants* a_n *defined in* (3) *have the following properties*:

(i) $a_n - 2nE\{1 - r(\mathbf{X}, \mathbf{b})\} \to 0$, *where* $\mathbf{b}$ *can be any of* $\mathbf{m}$, $\frac{1}{2}V_0^{-1}\mathbf{m}$, *or* $\mathbf{H}$;
(ii) $a_n - nE(1 - W) \to 0$;
(iii) $|a_n - n(1 - n^{-1}\mathbf{m}'\mathbf{m}) + \frac{3}{2}| \leq C(\log n)^{-1}$; *and*
(iv) $C_1 \log\log(n) < a_n < C_2 \log\log(n)$, $0 < C_1 < C_2 < \infty$.

Note that (iii) implies that

$$\mathbf{m}'\mathbf{m} = n - a_n - \tfrac{3}{2} + o(1).$$

As far as we are aware, this property of $\mathbf{m}'\mathbf{m}$ has not appeared elsewhere; the behaviour of $\mathbf{m}'\mathbf{m}$ is of interest in other contexts and has been the subject of a number of papers (see for example, Balakrishnan (1984), Ruben (1956) and Saw and Chow (1966)).

It should be pointed out that the convergence for (i) and (ii) in the lemma is extremely slow; for example $a_n - 2En(1 - r(\mathbf{X}, \mathbf{m})) \approx -0.1$ for $40 \leq n \leq 400$. It is, therefore, unclear which set of norming constants is best to use.

When Sarkadi (1975) established the consistency of the Shapiro–Francia test, it seemed likely that the Shapiro–Wilk test would share that property. That it is indeed consistent will follow from a straightforward application of a result in Sarkadi (1981).

3. Proofs.

NOTATION. We give some notation that will be used throughout the rest of the paper. With or without subscripts, C is a generic constant that is independent of i and n. Set $\mathbf{g} = \frac{1}{2}V_0^{-1}\mathbf{m}$, $nG_n^2 = \mathbf{g}'\mathbf{g}$, $nM_n^2 = \mathbf{m}'\mathbf{m}$, N as the integer part of $\frac{1}{2}(n+1)$, $S_n^2 = \Sigma_1^n(X_i - \bar{X})^2/n$, $\psi(v) = \Phi^{-1}\{\exp(-v)\}$, $s_i = \Sigma_i^n \nu^{-1}$, $\psi_i = \psi(s_i)$, and $W_i = -\log(\Phi(X_i)) - s_i$. Note that $W_i + s_i$ is the ith *largest* order statistic in a random sample from an exponential distribution; $EW_i = 0$, $EW_i^2 = d_{in}$, where $d_{in} = \Sigma_i^n \nu^{-2}$, $EW_i^3 = 2\Sigma_i^n \nu^{-3}$, and $|EW_i^r| \le Ci^{-2}$ for $r \ge 3$. Denote the ith element of $\mathbf{g}$, $\mathbf{m}$, and $\mathbf{H}$ by g_i, m_i, and H_i, respectively ($\mathbf{m}$ and $\mathbf{H}$ are given in Section 1). Further, as r is scale and location invariant we assume without loss of generality that our sample is from a $N(0,1)$ population.

PROOF OF CONSISTENCY. The consistency of W follows directly from Theorem 1 of Sarkadi (1981). There is a small difficulty in that whilst it appears to be the case that $V_0^{-1}\mathbf{m}$ is a vector whose elements, as you move down the vector, are monotonic increasing, we are unable to prove it. This means we cannot establish that $W^{1/2}$ is always positive. Sarkadi exploits the fact that $W^{\dagger 1/2}$ is always positive to argue that tests based on $W^{\dagger 1/2}$ are equivalent to those based on $W^{\dagger}$. We need to argue likewise for W. (Note: we distinguish between $W^{1/2}$, $W^{\dagger 1/2}$, etc., and the square roots of W, $W^{\dagger}$, etc.; it is true that $W = (W^{1/2})^2$, but in view of what has just been said, we are unable to say whether $W^{1/2}$ is the positive square root of W.) We overcome this difficulty by showing that

$$W^{1/2} \ge -C(\log n)^{-1/2}, \quad C \text{ independent of } n. \tag{6}$$

From the theorem and the lemma, the $100\alpha\%$ critical region for the test based on $W^{1/2}$ is $W^{1/2} < 1 - \frac{1}{2}(c(\alpha) + a_n)n^{-1}$. For the test based on W it is $W < 1 - (c(\alpha) + a_n)n^{-1}$. By (6) the two critical regions are asymptotically equivalent. We need only show, therefore, that $W^{1/2}$ is consistent. We establish (6) by setting $\mathbf{1}_n$ to be an $n \times 1$ vector of 1's and writing

$$W^{1/2} = \{(\mathbf{X} - \bar{X}\mathbf{1}_n)'(\mathbf{g} - \mathbf{m}) + \mathbf{X}'\mathbf{m}\}/(nS_nG_n).$$

As $\mathbf{X}'\mathbf{m} > 0$ provided only that the components of $\mathbf{X}$ are increasing (see Sarkadi (1975), Lemma 2), and from (1), $\max|g_i - m_i| < C/\sqrt{(\log n)}$, we have, with the help of (21),

$$W^{1/2} \ge -C\sum_1^n |X_i - \bar{X}|/\left\{nS_nG_n\sqrt{(\log n)}\right\} \ge -C/\sqrt{(\log n)}.$$

We turn now to Theorem 1 of Sarkadi (1981). Applied to our context, it states that $W^{1/2}$ will determine a consistent test of H_0: that the random sample is normal, versus H_1: that the observations are not normal (Sarkadi also allows the observations under H_1 to be m dependent with common nonnormal marginal) providing

$$\sum_1^n g_i G_n^{-1} \int I(i - 1 < nu < i)\Phi^{-1}(u)\,du = 1 + o(1), \tag{7}$$

where $I(A)$ is the indicator function of A. Note that Sarkadi's theorem is framed in terms of a statistic T_n, which here takes the form

$$T_n = \sum_1^n \{(X_i - \bar{X})n^{-1/2}S_n^{-1} - c_{in}\}^2 = 2(1 - W^{1/2}),$$

where $c_{in}\sqrt{n} = g_i/G_n$. To establish (7) we require results contained in the proof of both our lemma and theorem; therefore, we will leave the derivation of (7) until the end of the article.

PROOF OF LEMMA. We start by showing (iii). Observe that

$$n(1 - M_n^2) = 2\left\{\sum_1^N \operatorname{Var}(X_i)\right\} - (2N - n)\operatorname{Var}(X_N).$$

We can write

$$\operatorname{Var}(X_i) = E\{\psi(s_i + W_i) - E\psi(s_i + W_i)\}^2.$$

Expanding ψ in W_i up to third order terms, using the properties of W_i given in the section on notation and, together with results in Leslie (1984) (in particular, Lemma 6 and the properties of ψ given in Section 3), we can show that

$$\left|\operatorname{Var}(X_i) - \{\psi'(s_i)\}^2 d_{in}\right| \le C\{i(\log(n/i))\}^{-2},$$

where $\psi'(s_i) = \{\exp(-s_i)\}/\phi(\Phi^{-1}(\exp(-s_i)))$ and $d_{in} = \sum_i^n \nu^{-2}$. This yields

$$\left|\sum_1^N \operatorname{Var}(X_i) - \sum_1^N \{\psi'(s_i)\}^2 d_{in}\right| \le C(\log n)^{-2}. \tag{8}$$

Using the Euler–Maclaurin summation formula (Knopp (1951), page 534)

$$0 < s_i - \log((n+1)/i) - \tfrac{1}{2}\left(i^{-1} - (n+1)^{-1}\right) < \left\{i^{-2} - (n+1)^{-2}\right\}\Big/12 \tag{9}$$

and

$$0 < d_{in} - i^{-1}\{1 - (i/(n+1))\} - \tfrac{1}{2}\left\{i^{-2} - (n+1)^{-2}\right\} < 1/(6i^3). \tag{10}$$

In FLS we show that $|\psi'(v)|$ and $|\psi''(v)|$ are monotonic decreasing in v; also in Lemmas 1 and 4 in Leslie (1984) it is shown that

$$|\psi'\{\log((n+1)/i)\}| < C\{\log(n/i)\}^{-1/2} \tag{11}$$

and

$$|\psi''\{\log((n+1)/i)\}| < C\{\log(n/i)\}^{-3/2}. \tag{12}$$

With (9), (10), and (11) we have

$$\left|\{\psi'(s_i)\}^2\left(d_{in} - i^{-1}\{1 - (i/(n+1))\}\right)\right| < Ci^{-2}/\log(n/i), \tag{13}$$

$$\left|\{\psi'(s_i)\}^2 - \left(\psi'\{\log((n+1)/i)\}\right)^2\right| < C|\psi'(\alpha_i)\psi''(\alpha_i)|/i, \tag{14}$$

where $\log\{(n+1)/i\} < \alpha_i < s_i$. Expressions (11)–(14) taken together imply that

$$\left|\sum_1^N \{\psi'(s_i)\}^2 d_{in} - (n+1)^{-2} \times \sum_1^N \left(\phi\{\Phi^{-1}(i/(n+1))\}\right)^{-2} i\left\{1 - i(n+1)^{-1}\right\}\right| < C(\log n)^{-1}. \tag{15}$$

From the definition of a_n and with (8) and (15) we obtain (iii).

Next we establish (iv). A well known inequality is useful here (see Rényi (1970), page 164); for $x < 0$,

$$\phi(x)(1 - x^{-2})/|x| < \Phi(x) < \phi(x)/|x|. \tag{16}$$

From this we obtain, for $1 \le i \le N$ and with $x = H_i$,

$$1 - H_i^{-2} \le i|H_i|/\{(n+1)\phi(H_i)\} \le 1. \tag{17}$$

In view of the symmetry in the summands in a_n, we need consider only $1 \le i \le N$. We use (17) over the range $1 \le i \le [\frac{1}{2}N]$ and for $[\frac{1}{2}N] < i \le N$ we use

$$C_1 < \phi(H_i)(i/(n+1))\{1 - (i/(n+1))\} < C_2, \tag{18}$$

where C_1, C_2 do not depend on i or n. Based on (16) we show in Lemma 3 of FLS that for any c_0 $(0 < c_0 < 1)$ there is a $\gamma(c_0)$ such that when $0 < u < \gamma(c_0) < \frac{1}{2}$,

$$-\left\{-\log(2\pi u^2)\right\}^{1/2} < \Phi^{-1}(u) < -\left\{-c_0 \log(2\pi u^2)\right\}^{1/2}. \tag{19}$$

This yields, for $1 \le i \le N$,

$$C_3\{\log(n/i)\}^{1/2} < |H_i| < C_4\{\log(n/i)\}^{1/2}. \tag{20}$$

Applying (17), (18), and (20) we find

$$C_5 \sum_1^{[N/2]} \{i \log(n/i)\}^{-1} + C_6 < a_n + \tfrac{3}{2} < C_7 \sum_1^{[N/2]} \{i \log(n/i)\}^{-1} + C_8,$$

which, after approximating the sum by an integral, establishes (iv).

To complete the lemma we prove (i) and (ii). First, however, we need two results that will be used here and in the proof of the theorem:

$$G_n \to 1 \quad \text{as } n \to \infty \tag{21}$$

and

$$0 \le \|\mathbf{m}\|\,\|\mathbf{g}\| - \mathbf{m}'\mathbf{g} \le M_n G_n^{-1} \|\mathbf{g} - \mathbf{m}\|^2. \tag{22}$$

It is well known that $M_n \to 1$ as $n \to \infty$ (see Hoeffding (1953)). On writing $G_n^2 = M_n^2 + 2\mathbf{m}'(\mathbf{g} - \mathbf{m})n^{-1} + \|\mathbf{g} - \mathbf{m}\|^2 n^{-1}$, from (1) and the Schwarz inequality we obtain (21). We demonstrate (22) by exploiting an idea in Sarkadi (1972). First note that $\mathbf{m}'\mathbf{g} > 0$, for $\mathbf{m}'\mathbf{g} = \mathbf{m}'V_0^{-1}\mathbf{m}$ and V_0, being a covariance matrix, is positive definite. Set θ to be the angle between $\mathbf{m}$ and $\mathbf{g}$; then $\cos\theta > 0$ and $0 < \theta < \frac{1}{2}\pi$. Consider the triangle formed by vectors $\mathbf{m}$, $\mathbf{g}$, and

$\mathbf{a} = \mathbf{m} - \mathbf{g}$, lines AB, AC, and CB, respectively. Let CD be the perpendicular from C to AB. Then $\|\mathbf{a}\|^2 \geq (CD)^2 = \|\mathbf{g}\|^2 \sin^2\theta = \|\mathbf{g}\|^2(1 + \cos\theta)(1 - \cos\theta) > \|\mathbf{g}\|^2(1 - \cos\theta) \geq 0$. As $\cos\theta = \mathbf{m}'\mathbf{g}/\{\|\mathbf{m}\|\|\mathbf{g}\|\}$, (22) follows.

Returning to the proof of (i) and (ii) of the lemma, we show first that

$$(23) \qquad \left|nE(1 - r(\mathbf{X}, \mathbf{b})) - n(1 - M_n) + \tfrac{3}{4}\right| \leq C(\log n)^{-1}.$$

As r is scale invariant and as S_n^2 is sufficient for the scale parameter σ, we can use Theorem 7 of Hogg and Craig ((1970), page 243) to yield

$$nEr(\mathbf{X}, \mathbf{b}) = \mathbf{m}'\mathbf{b}/\{ES_n\|\mathbf{b}\|n^{-1/2}\}.$$

With nS_n^2 distributed as χ^2 on $n - 1$ degrees of freedom it is elementary to show that

$$ES_n = (2/n)^{1/2}\Gamma(n/2)/\Gamma((n-1)/2).$$

By Stirling's formula this reduces to $1 - (\frac{3}{4})n^{-1} + O(n^{-2})$. As $n^{-1/2}\|\mathbf{b}\| \to 1$ (the case $\mathbf{b} = \mathbf{H}$ is shown in Lemma 2 of De Wet and Venter (1972)), and using (1), (22), and an analogue of (22) with $\mathbf{g}$ and G_n replaced by $\mathbf{H}$ and $H_n = \sqrt{\{(\mathbf{H}'\mathbf{H})/n\}}$ (this analogue holds because $m_i H_i > 0$ for all i; m_i and H_i always having the same sign), we have

$$\begin{aligned} En\{1 - r(\mathbf{X}, \mathbf{b})\} &= n(ES_n)^{-1}\{1 - (\tfrac{3}{4})n^{-1} - n^{-1/2}\mathbf{m}'\mathbf{b}\|\mathbf{b}\|^{-1}\} + O(n^{-1}) \\ &= n(ES_n)^{-1}(1 - M_n) - (\tfrac{3}{4}) + O(\log n)^{-1} \\ &= n(1 - M_n) - (\tfrac{3}{4}) + O(\log n)^{-1}, \end{aligned}$$

the latter expression resulting from the fact that $n(1 - M_n) = O(\log\log n)$ (using (iii) and (iv) of the lemma and recalling that $M_n \to 1$). This establishes (23). Analogous to (23) for $\mathbf{b} = \mathbf{g}$ we have

$$(24) \qquad \left|nE\{1 - r^2(\mathbf{X}, \mathbf{g})\} - n(1 - M_n^2) + \tfrac{3}{2}\right| \leq C(\log n)^{-1}.$$

To show this we note that as $nES_n^2 = n - 1$, we can write

$$nEr^2(\mathbf{X}, \mathbf{g}) = E(\mathbf{X}'\mathbf{g})^2/\{(n-1)G_n^2\}$$

with

$$\begin{aligned} E(\mathbf{X}'\mathbf{g})^2 &= \mathbf{g}'V_0\mathbf{g} + (\mathbf{g}'\mathbf{m})^2 = \tfrac{1}{2}\mathbf{m}'\mathbf{g} + (\mathbf{g}'\mathbf{m})^2 \\ &= \tfrac{1}{2}nM_nG_n + (nG_nM_n)^2 + O(n/\log n), \end{aligned}$$

using (1) and (22). Again using the property that $n(1 - M_n) = O(\log\log n)$, we obtain (24). As

$$(25) \quad 2n(1 - M_n) - n(1 - M_n^2) = \{\sqrt{n}(1 - M_n)\}^2 = O\{(\log\log n)^2/n\},$$

it is clear from (23), (24), and (iii) of the lemma that (i) and (ii) hold. □

Undoubtedly it is true that $a_n - En\{1 - r^2(\mathbf{X}, \mathbf{b})\} \to 0$, for $\mathbf{b} = \mathbf{m}$ and $\mathbf{H}$. However, this entails showing that $\|V_0\mathbf{H} - 2\mathbf{H}\| \to 0$ and $\|V_0\mathbf{m} - 2\mathbf{m}\| \to 0$, both of which will follow once V_0 is replaced by the approximation V given in Leslie (1984): Corollary 1 in Leslie (1984) permits this. These two results involve a quantity of tedious analysis and it seems unnecessary to set it down here.

PROOF OF THEOREM. The theorem follows from (2), (4), and (5) together with the lemma; therefore, to prove the theorem it remains to establish (5). Now

$$\begin{aligned} nS_n(W^{1/2} - W^{\dagger 1/2}) &= \sum_1^n X_i(g_i G_n^{-1} - m_i M_n^{-1}) \\ &= \sum (X_i - m_i)(g_i - m_i) G_n^{-1} \\ &\quad + \sum (X_i - m_i) m_i (G_n^{-1} - M_n^{-1}) + (\mathbf{m}'\mathbf{g} - \|\mathbf{m}\|\|\mathbf{g}\|) G_n^{-1}. \end{aligned}$$

As $S_n \to 1$ a.s. and with (21) and (22), expression (5) will follow from Markov's inequality once we demonstrate that

$$E\left|\sum (X_i - m_i)(g_i - m_i)\right| \to 0 \tag{26}$$

and

$$E\left|\sum (X_i - m_i) m_i (G_n^{-1} - M_n^{-1})\right| \to 0. \tag{27}$$

Result (26) follows from the Schwarz inequality:

$$E\left|\sum (X_i - m_i)(g_i - m_i)\right| \le \left\{n\left(1 - M_n^2\right)\right\}^{1/2} \|\mathbf{g} - \mathbf{m}\|.$$

With (1) and with (iii) and (iv) of the lemma we have (26).

To deal with (27) we note that in Lemma 11 of FLS

$$E|X_i - \psi_i| < C/\sqrt{\{i \log(n/i)\}} \tag{28}$$

and in Theorem 1 in FLS

$$|\psi_i - m_i| < C i^{-1} \{\log(n/i)\}^{-3/2}; \tag{29}$$

both of these bounds hold provided $1 \le i \le N$. As $|\|\mathbf{m}\| - \|\mathbf{g}\|| \le \|\mathbf{m} - \mathbf{g}\|$,

$$\left|G_n^{-1} - M_n^{-1}\right| \le \|\mathbf{m} - \mathbf{g}\| / \left(M_n G_n \sqrt{n}\right) \le C(n \log n)^{-1/2} \tag{30}$$

and

$$E\left|\sum (X_i - m_i) m_i\right| \le 2 \sum_1^N (E|X_i - \psi_i||m_i| + |\psi_i - m_i||m_i|). \tag{31}$$

From (29), (9), (20), and the monotonicity (decreasing) of $|\psi(v)|$,

$$|m_i| \le C\{\log(n/i)\}^{1/2}, \qquad 1 \le i \le N, \tag{32}$$

so by combining results (28) to (32) we find

$$E\left|\sum (X_i - m_i) m_i \{G_n^{-1} - M_n^{-1}\}\right| \le C(\log n)^{-1/2}.$$

This establishes (27) and hence the theorem. □

DERIVATION OF EXPRESSION (7). Denote the integral in (7) by $J(i, n)$; then

$$J(i,n) = \begin{cases} -\phi\{\Phi^{-1}(1/n)\}, & \text{for } i = 1, \\ \Phi^{-1}((i-1)/n)n^{-1} + \tfrac{1}{2}n^{-2}\left(\phi\{\Phi^{-1}((i-\theta)/n)\}\right)^{-1}, & \\ \qquad\qquad 0 < \theta < 1, 1 < i < n, & \\ \phi\{\Phi^{-1}(1-n^{-1})\}, & \text{for } i = n. \end{cases}$$

Without loss of generality, assume n is even. Then

$$\sum_1^n g_i J(i,n)/G_n = 2n^{-1}\sum_{i=2}^{n/2} g_i\Big(\Phi^{-1}\{(i-1)/n\} + \tfrac{1}{2}n^{-1}\{\phi\{\Phi^{-1}((i-\theta)/n)\}\}^{-1}\Big) + 2g_n\phi\{\Phi^{-1}(1/n)\}.$$

By (16), for $1 < i \le \frac{1}{2}n$,

$$\phi\{\Phi^{-1}((i-\theta)/n)\} \ge \phi\{\Phi^{-1}((i-1)/n)\} \ge \begin{cases} (i-1)\Phi^{-1}\{(i-1)/n\}/n, & 2 \le i \le kn,\ k < \frac{1}{2} \\ C(k), & kn < i \le \frac{1}{2}n. \end{cases}$$

Thus by the Schwarz inequality,

$$\left| n^{-2}\sum_2^{n/2} g_i/\phi\{\Phi^{-1}((i-\theta)/n)\}\right| \le n^{-1/2}G_n\left\{\sum_1^{kn-1}\{i\Phi^{-1}(i/n)\}^{-2} + \sum_{kn+1}^{n/2} C(k)^{-2}n^{-2}\right\},$$

which in turn is bounded by $C\{n\log(n)\}^{-1/2}$, in view of (19). Further, by (16) and (19), $\phi(\Phi^{-1}(1/n)) \sim O((\log n)/n)$; by (1), $g_n \sim m_n$; and with (32) and finally (22) we can argue that

$$2n^{-1}\sum_2^{n/2} g_i\Phi^{-1}\{(i-1)/n\} \sim n^{-1}\sum_1^n g_i\Phi^{-1}\{i/(n+1)\} \sim n^{-1}\mathbf{m}'\mathbf{g} \sim M_nG_n.$$

These ensure that (7) holds. □

Acknowledgments. We wish to thank the referees for their helpful comments and to thank one referee in particular whose very careful reading of the manuscript led to many improvements. We are also grateful to W. C. M. Kallenberg for drawing our attention to deficiencies in the proof of consistency in an early draft of this paper.

REFERENCES

BALAKRISHNAN, N. (1984). Approximating the sum of squares of normal scores. *Appl. Statist.* **33** 242–245.

DAVID, H. A. (1981). *Order Statistics*, 2nd ed. Wiley, New York.

DE WET, T. and VENTER, J. H. (1972). Asymptotic distributions of certain test criteria of normality. *South African Statist. J.* **6** 135–149.

FOTOPOULOS, S., LESLIE, J. R. and STEPHENS, M. A. (1984). Errors in approximations for expected normal order statistics with an application to goodness-of-fit. Technical Report No. 342, Dept. of Statistics, Stanford Univ.

GREGORY, G. G. (1977). Functions of order statistics and tests of fit. *South African Statist. J.* **11** 99–118.

HOEFFDING, W. (1953). On the distribution of the expected values of the order statistics. *Ann. Math. Statist.* **24** 93–100.

HOGG, R. V. and CRAIG, A. T. (1970). *Introduction to Mathematical Statistics*, 3rd ed. Macmillan, New York.

KNOPP, K. (1951). *Theory and Application of Infinite Series*, 2nd English ed. Blackie, London.

LESLIE, J. R. (1984). Asymptotic properties and a new approximation for both the covariance matrix of normal order statistics and its inverse. To appear in *Colloq. Math. Soc. J. Bolyai on Goodness of Fit* **45**.

RÉNYI, A. (1970). *Probability Theory*. North-Holland, Amsterdam.

RUBEN, H. (1956). On the sum of squares of normal scores. *Biometrika* **43** 456–458.

SARKADI, K. (1975). The consistency of the Shapiro–Francia test. *Biometrika* **62** 445–450.

SARKADI, K. (1977). Characterization and testing for normality. *Colloq. Math. Soc. J. Bolyai* **21** 317–329.

SARKADI, K. (1981). On the consistency of some goodness of fit tests. *Proc. Sixth Conf. Probab. Theory, Brasov, 1979*, Bucuresti 195–204. Ed. Acad. R. S. Romania.

SAW, J. G. and CHOW, B. (1966). The curve through the expected values of ordered variates and the sum of squares of normal scores. *Biometrika* **53** 252–255.

SHAPIRO, S. S. and FRANCIA, R. S. (1972). An approximate analysis of variance test for normality. *J. Amer. Statist. Assoc.* **67** 215–216.

SHAPIRO, S. S. and WILK, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika* **52** 591–611.

STEPHENS, M. A. (1975). Asymptotic properties for covariance matrices of order statistics. *Biometrika* **62** 23-28.

VERRILL, S. P. and JOHNSON, R. A. (1983). The asymptotic distributions of censored data versions of the Shapiro–Wilk test for normality statistic. Technical Report No. 702, Dept. of Statistics, Univ. of Wisconsin.

J. R. LESLIE
DEPARTMENT OF STATISTICS
UNIVERSITY OF LONDON
MALET STREET
LONDON WC1E 7HX
ENGLAND

M. A. STEPHENS
DEPARTMENT OF MATHEMATICS
SIMON FRASER UNIVERSITY
BURNABY, BRITISH COLUMBIA V5A 1S6
CANADA

S. FOTOPOULOS
DEPARTMENT OF MANAGEMENT SYSTEMS
WASHINGTON STATE UNIVERSITY
PULLMAN, WASHINGTON 99164