The Canadian Journal of Statistics
Vol. 15, No. 2, 1987, Pages 159–167
La Revue Canadienne de Statistique

On the asymptotic efficiency of certain correlation tests of fit

C.G. McLAREN and Richard LOCKHART

Ministry of Agriculture, Zimbabwe and *Simon Fraser University*

Key words and phrases: Exponential distribution, extreme-value distribution, logistic distribution, probability plots.
AMS 1980 subject classifications: Primary 62F05; secondary 62G30.

ABSTRACT

Tests of fit based on correlation-type statistics are investigated for the exponential, extreme-value, and logistic distributions. The statistics are shown to be asymptotically normal at the rate $\log^{\frac{1}{2}} n$. The result is used to show that such tests have 0 asymptotic relative efficiency.

RÉSUMÉ

On examine le comportement des tests d'ajustement basés sur des statistiques ayant la forme d'un coefficient de corrélation lorsque les observations sont régies tour à tour par la loi exponentielle, la loi des valeurs extrêmes et la loi logistique. On montre que ces statistiques convergent en loi vers une normale à une vitesse de l'ordre de $\log^{\frac{1}{2}}(n)$. Ce résultat permet de conclure que les tests correspondants possèdent une efficacité asymptotique relative qui est nulle.

1. INTRODUCTION

Suppose $X_1 \leq \cdots \leq X_n$ are the order statistics for a sample from the distribution G. In order to test the hypothesis, H_0, that $G(x) = F\{(x - \alpha)/\beta\}$ for some specified distribution F with unknown location α and scale β, the order statistics are often plotted against their expected values under F or against an approximation to these values. If $W_i = (X_i - \alpha)/\beta$, then under H_0 the X_i follow a linear model $X_i = \alpha + \beta m_i + \epsilon_i$, where $m_i = \mathscr{E}(W_i)$ and the ϵ_i have mean 0 but are generally correlated. Thus formal tests of H_0 are based on assessing the linearity of the plot. A natural test statistic is $R(X, m)$, the Pearson correlation coefficient. See Stephens and D'Agostino (1986) for a discussion of this and related tests.

When F is the standard normal distribution, the resulting test is powerful; see Stephens and D'Agostino (1986). In this case $1 - R^2$ is the Shapiro-Francia statistic. Fotopolous, Leslie, and Stephens (1984) have shown that this statistic is equivalent to the statistic of De Wet and Venter (1972). Leslie et al. (1986) have shown that $R^2(X, m)$ is asymptotically equivalent to the Shapiro-Wilk statistic.

R^2 and analogous statistics have been suggested by Gerlach (1976) for the extreme-value distribution and by Smith and Bain (1976) for the exponential distribution. For the exponential distribution Spinelli and Stephens (1983) have shown that such statistics have much lower Monte Carlo powers than empirical distribution-function techniques. In this note we establish that for the exponential, logistic, and extreme-value distributions the tests have asymptotic power equal to their level against a wide variety of contiguous alternatives. In other words, $R(X, m)$ has 0 asymptotic efficiency relative to standard goodness-of-fit tests such as the usual empirical distribution-function techniques.

In Section 2 we establish that statistics of the form

$$T_n(\hat{\alpha}, \hat{\beta}, \tilde{\beta}) = \sum_1^n \frac{(X_i - \hat{\alpha} - \hat{\beta} m_i)^2}{\tilde{\beta}^2} \tag{1.1}$$

are asymptotically normally distributed at the rate $\log^{\frac{1}{2}} n$ provided $\hat{\alpha}$, $\hat{\beta}$, and $\tilde{\beta}$ are $n^{\frac{1}{2}}$-consistent estimates. The statistic $n\{1 - R^2(X, m)\}$ has this form with $\hat{\alpha}, \hat{\beta}$ the usual least-squares estimates, and $\tilde{\beta}^2 = \Sigma(X_i - \bar{X})^2/n$. The results extend to other null hypotheses where α or β or both are known, on replacing estimates with the known values. The proofs show that T_n is asymptotically equivalent to a statistic based only on the spacings between the tail order statistics from an exponential sample. We end Section 2 with a brief discussion of the effect of using approximations to the m_i.

In Section 3 we study the behaviour of T_n on sequences of contiguous alternatives by finding conditions under which T_n is asymptotically independent of the log-likelihood ratio. The conditions show that statistics of the form T_n are unable to separate contiguous Weibull or gamma alternatives from exponential null hypotheses, log-generalized gamma alternatives from extreme-value null hypotheses, or log-generalized logistic alternatives from logistic null hypotheses.

2. ASYMPTOTIC NULL DISTRIBUTIONS

We will need the following notation. By P and $\mathscr{E}$ we denote probability and expectation under H_0. For notational convenience we take $\alpha = 0$ and $\beta = 1$; for many estimators T is actually location and scale invariant. For $i = 1, \ldots, n$ set $Y_i = -\log\{1 - F(X_i)\}$, $Y_0 = 0$, and $D_i = (n - i + 1)(Y_i - Y_{i-1})$. On H_0, $Y_1, \ldots, Y_n$ are order statistics from the exponential distribution and $D_1, \ldots, D_n$ are independent standard exponentials. Let $k = [n/\log n]$, and set

$$Q_n = \Sigma\left\{\frac{(D_{n-i+1} - 1)(D_{n-j+1} - 1)}{\max(i,j)};\ 1 \le i, j \le k\right\}. \tag{2.1}$$

Let $Y_i^* = -\log F(X_{n-i+1})$. Note that $Y_1^*, \ldots, Y_n^*$ have the joint distribution of $Y_1, \ldots, Y_n$. Define D_i^* and Q_n^* from the Y_i^* analogously to D_i and Q_n, so that Q_n and Q_n^* are identically distributed. Put $p_i = \mathscr{E}(Y_i) = \Sigma_{n-i+1}^n 1/j$. Moment calculations show

$$\frac{\Sigma(Y_i - p_i)^2 - Q_n}{\sigma_n} \to 0 \quad \text{in P-probability} \tag{2.2}$$

where $\sigma_n^2 = 4 \log n$. Since Q_n is a function of $D_{n-k+1}, \ldots, D_n$ and Q_n^* is a function of $Y_1, \ldots, Y_k$ we see that Q_n^* and Q_n are independent. It follows from the work of Lockhart (1985) that $\{(Q_n^* - \mu_n)/\sigma_n, (Q_n - \mu_n)/\sigma_n\}$ is asymptotically standard bivariate normal, where $\mu_n = \log n$.

We are now able to give the asymptotic behaviour of T_n when F is exponential, extreme-value, or logistic.

PROPOSITION 1. *Suppose that, under* P, $n^{\frac{1}{2}}\hat{\alpha}$ *and* $n^{\frac{1}{2}}(\hat{\beta} - 1)$ *are bounded in probability and that* $(\log n)^{\frac{1}{2}}(\tilde{\beta} - 1) \to 0$ *in probability.*

(a) *If* $F(x) = 1 - e^{-x}$ *for* $x > 0$ *or* $F(x) = \exp\{-e^{-x}\}$, *then*
 (1) $(T_n - Q_n)/\sigma_n \to 0$ *in* P-*probability, and*
 (2) $(T_n - \mu_n)/\sigma_n$ *is asymptotically standard normal.*

(b) *If* $F(x) = 1/(1 + e^{-x})$, *then*
(1) $(T_n - Q_n - Q_n^*)/\sigma_n \to 0$ *in* P-*probability and*
(2) $(T_n - 2\mu_n)/(2^{\frac{1}{2}}\sigma_n)$ *is asymptotically standard normal.*

The proposition is proved for the extreme-value and logistic distributions in the Appendix.

Exact values of the m_i are not always used in plots or in the statistics used for formal testing. In particular, $F^{-1}\{i/(n+1)\}$ or $F^{-1}\{(i - 1/2)/n\}$ or Blom-type formulas are often used in place of the m_i. Let h_i denote some approximation to the m_i. For most approximations $m_i - h_i = O(1/n)$ for i/n bounded away from 0 or 1. The approximation is generally much worse for i near n or 1. For the distributions considered here the present theory shows that is is precisely these i which are important. However, the techniques used in proving the proposition can be used to show that the conclusions of the proposition remain valid for a variety of standard approximations.

Consider, for example, the exponential distribution, and take $h_i = F^{-1}\{i/(n+1)\}$. Let S_n be defined by (1.1) with m_i replaced by h_i. The inequality $h_i \le m_i \le h_{i+1}$ and direct variance calculations can be used to show that $\{T_n(0,1,1) - S_n(0,1,1)\}/\sigma_n$ tends to 0 in probability. Arguments similar to those given in the appendix establish that Proposition 1 is valid with S_n replacing T_n.

3. CONTIGUOUS ALTERNATIVES

Now suppose that F_n is a sequence of alternative distributions contiguous to F. Let f_n and f be the corresponding densities, and let $\lambda_n(x) = \log f_n(x) - \log f(x)$ with the usual conventions concerning ∞. By P_n and $\mathscr{E}_n$ we denote probability and expectation under F_n for samples of size n. Contiguity means that statement (1) of (a) or (b) of Proposition 1 is valid under P_n. The log-likelihood ratio is $\Lambda_n = \Sigma\lambda_n(X_i)$.

We need the following property of contiguity. Let T_n be a sequence of statistics converging in distribution to G under P. Suppose there is a sequence of random variables L_n such that $\Lambda_n - L_n$ tends to 0 in P-probability and such that L_n and T_n are independent under P. Then T_n converges in distribution to G under P_n.

Consider first the exponential and extreme-value distributions; in this case the statistic T_n of this paper is asymptotically equivalent to Q_n, which is determined by $D_{n-k+1}, \ldots, D_n$. To apply the proposition it suffices to prove that Λ_n is asymptotically equivalent to a function of $D_1, \ldots, D_{n-k}$ or equivalently to a function of $X_1, \ldots, X_{n-k}$. In the logistic case we would need to show that Λ_n is asymptotically equivalent to a function of $D_{k+1}, \ldots, D_{n-k}$ or $D^*_{k+1}, \ldots, D^*_{n-k}$. We illustrate with several examples.

EXAMPLE 1. Let F be the exponential distribution, and F_n be the Gamma distribution with shape $1 + \gamma/n^{\frac{1}{2}}$. Then $\lambda_n(x) = (\gamma \log x)/n^{\frac{1}{2}} - \log\Gamma(1 + \gamma/n^{\frac{1}{2}})$. Under P, $\log X_1, \ldots, \log X_n$ are extreme-value order statistics. Take

$$L_n = \frac{\gamma\Sigma_1^{n-k}\log(X_i)}{n^{\frac{1}{2}}} + \frac{\gamma\Sigma_{n-k+1}^{n} m_i}{n^{\frac{1}{2}}} - n\log\Gamma\left(1 + \frac{\gamma}{n^{\frac{1}{2}}}\right),$$

where the m_i are those of the extreme-value distribution, and use Lemma 1(c) of the appendix to see that $(T_n - \mu_n)/\sigma_n$ is asymptotically standard normal under this sequence of alternatives. Since γ is arbitrary, the asymptotic relative efficiency of T_n is 0.

EXAMPLE 2. Let F be the extreme-value distribution, and let F_n be the log-generalized gamma distribution whose density is $f_n(x) = \exp\{-(1 - \delta)/n^{\frac{1}{2}})x - e^{-x}\}/\Gamma(1 - \delta/n^{\frac{1}{2}})$.

Then $\lambda_n(x) = \delta x/n^{\frac{1}{2}} - \log\{\Gamma(1 - \delta/n^{\frac{1}{2}})\}$. Take

$$L_n = \frac{\delta \sum_1^{n-k} X_i}{n^{\frac{1}{2}}} + \frac{\delta \sum_{n-k+1}^{n} m_i}{n^{\frac{1}{2}}} - n \log \Gamma\left(1 - \frac{\delta}{n^{\frac{1}{2}}}\right).$$

Again the m_i are those of extreme-value distribution and Lemma 1(c) establishes that T_n has 0 asymptotic relative efficiency.

EXAMPLE 3. Let F be the logistic distribution, and let F_n be the generalized logistic distribution whose density is $f_n(x) = (1 - \delta/n^{\frac{1}{2}})e^{-x}/(1 + e^{-x})^{2-\delta/n^{\frac{1}{2}}}$. Then $\lambda_n(x) = \log(1 - \delta/n^{\frac{1}{2}}) + \delta \log(1 + e^{-x})/n^{\frac{1}{2}}$ and

$$\Lambda_n = n \log\left(1 - \frac{\delta}{n^{\frac{1}{2}}}\right) + \delta \sum \frac{Y_i^*}{n^{\frac{1}{2}}}.$$

Take

$$L_n = \frac{\delta \sum_{k+1}^{n-k} D_i^*}{n^{\frac{1}{2}}} + n \log\left(1 - \frac{\delta}{n^{\frac{1}{2}}}\right) + \frac{2\delta k}{n^{\frac{1}{2}}}.$$

In general, for the exponential and extreme-value distributions, the statistics T_n will have 0 asymptotic relative efficiency against any sequence of contiguous alternatives for which the extreme order statistics $X_{n-k+1}, \dots, X_n$ do not contribute asymptotically to the variance of Λ_n. (The situation for the logistic distribution is more complicated, and we will not consider that case here.) We have not been able to find good general conditions under which this holds; we present one approach which may help.

Suppose that $X_1, \dots, X_n$ are the order statistics for the sample $U_1, \dots, U_n$; that is, the U_i are independent and identically distributed according to F. We will try to approximate $\sum_1^k \lambda_n(X_{n-i+1})$ by

$$A_n = \sum_1^n \lambda_n(U_i) 1\{nF(U_i) > n - k\}$$

and then use the central limit theorem to control the variability of the latter sum. Define

$$\eta_n(M) = \sup\{|\lambda_n(x)| : |nF(x) - (n - k)| \leq Mk^{\frac{1}{2}}\}.$$

PROPOSITION 2. *Let F_n be a sequence of contiguous alternatives. Assume that $n^{\frac{1}{2}}\lambda_n(U_1)$ is uniformly square integrable under* P *and that $\tilde{\alpha}$, $\hat{\beta}$, and $\tilde{\beta}$ satisfy the hypotheses of Proposition 1. Assume that $\eta_n(M)$ is $o[\{(\log n)/n\}^{\frac{1}{2}}]$. Then the conclusions of Proposition 1 remain valid under F_n.*

EXAMPLE 4. Take F to be the exponential distribution and F_n to be the Weibull distribution with shape parameter $1 + \delta/n^{\frac{1}{2}}$. Proposition 2 can be applied.

APPENDIX. PROOFS OF THE PROPOSITIONS

Proof of Proposition 1. $F(x) = \exp(-e^{-x})$.

Let $j = n - i + 1$ define j in terms of i. Define $Z_j = Y_i - X_i$. Let $q_i = \mathscr{E}(Z_i)$. Write $T_n - Q_n = \sum_0^6 R_i$, where

$$R_0 = \sum_{n-k+1}^{n} (Y_i - p_i)^2 - Q_n,$$

$$R_1 = \sum_{n-k+1}^{n} (Z_j - q_j)^2 = \sum_{1}^{k} (Z_i - q_i)^2,$$

$$R_2 = -2 \sum_{n-k+1}^{n} (Y_i - p_i)(Z_j - q_j) = -2 \sum_{1}^{k} (Y_j - p_j)(Z_i - q_i),$$

$$R_3 = \sum_{k+1}^{n-k} (X_i - m_i)^2,$$

$$R_4 = \sum_{1}^{k} (X_i - m_i)^2,$$

$$R_5 = T_n(\hat{\alpha}, \hat{\beta}, 1) - T_n(0, 1, 1),$$

$$R_6 = T_n(\hat{\alpha}, \hat{\beta}, \tilde{\beta}) - T_n(\hat{\alpha}, \hat{\beta}, 1).$$

LEMMA 1. *For $i, l = 1, \ldots, n$,*

(a) $0 < \mathscr{E}(Z_i) = q_i < E(Y_i)/2 = p_i/2 = (\Sigma_{n-i+1}^{n} 1/l)/2$,
(b) $Var(Z_i) < Var(Y_i) = \Sigma_{n-i+1}^{n} 1/l^2$,
(c) $0 < Cov(Z_i, Z_l) < Cov(Y_i, Y_l) = Var(Y_i)$ for $i \leq l$.

LEMMA 2. *There are constants $c > 0$, $a > 0$, and $0 < s < 1$ such that for all $n > 0$ and all $i < ns$ we have*

(a) $-\log p_j < m_i < -\log p_j + Var(Y_j) + c\exp(-an^{\frac{1}{2}})$,
(b) $\mathscr{E}(X_i^2) < \log^2 p_j + ce^{-an}$,
(c) $Var(X_i) < 2\log p_j Var(Y_j)/p_j^2 + c\log p_n \exp(-an^{\frac{1}{2}})$.

Direct moment calculations show $R_0/\sigma_n \to 0$ in probability. Use Lemma 1(b) to check that $\mathscr{E}(R_1) = o(1)$ and hence R_1 tends to 0 in probability. Moreover,

$$|R_2| < 2\Big(\sum_{1}^{k} (Y_j - p_j)^2 \sum_{1}^{k} (Z_i - q_i)^2\Big)^{\frac{1}{2}}.$$

Since

$$\mathscr{E}\left\{\frac{\Sigma_{n-k+1}^{n}(Y_i - p_i)^2}{\sigma_n^2}\right\}$$

is bounded and R_1 tends to 0 in probability, we see that R_2/σ_n tends to 0 in probability. Use Lemma 1(b) and the inequality $(x + y)^2 < 2(x^2 + y^2)$ to check

$$\mathscr{E}(R_3) < 2 \sum_{k+1}^{n-k} \{Var(Y_i) + Var(Z_i)\} < 4 \sum_{k+1}^{n-k} Var(Y_i),$$

which is of the order $\log\log n$ by an integral comparison. Hence R_3/σ_n tends to 0 in probability. Use Lemma 2(c) and integral comparison to check that $\mathscr{E}(R_4)/\sigma_n$ tends to 0. Use $\Sigma m_i/n = \gamma = \int xF(dx)$ to write

$$R_5 = n\hat{\alpha}^2 + n(\hat{\beta} - 1)^2\frac{\Sigma m_i^2}{n} + 2\gamma n\hat{\alpha}(\hat{\beta} - 1) - 2n\hat{\alpha}(\bar{X} - \gamma) - 2(\hat{\beta} - 1)\Sigma m_i(X_i - m_i).$$

It suffices to show that R_5 is bounded in probability; this will follow from the hypotheses of the theorem if we check $\Sigma m_i^2/n = O(1)$ and $\Sigma m_i(X_i - m_i) = O_P(n^{\frac{1}{2}})$. First,

$\Sigma m_i^2/n \leq \Sigma \mathscr{E}(X_i^2)/n = \int x^2 F(dx)$. Finally,

$$Var\{\Sigma m_i(X_i - m_i)\} < 4[Var\{\Sigma p_i(Y_i - p_i)\} + Var\{\Sigma q_j(Y_i - p_i)\} + Var\{\Sigma p_i(Z_j - q_j)\} + Var\{\Sigma q_j(Z_j - q_j)\}],$$

which, using Lemma 1(a), (c), is bounded by

$$8[Var\{\Sigma p_i(Y_i - p_i)\} + Var\{\Sigma p_j(Y_i - p_i)\}].$$

The latter quantity is $O(n)$ by direct calculation.

Having established that $R_i/\sigma_n \to 0$ in probability for $i = 0, \ldots, 5$, we see that $\{T_n(\hat{\alpha}, \hat{\beta}, 1)\text{-}\mu_n\}/\sigma_n$ is bounded in probability. Then

$$\frac{R_6}{\sigma_n} = \left(\frac{T_n(\hat{\alpha}, \hat{\beta}, 1) - \mu_n}{\sigma_n} + \frac{\sigma_n}{4}\right)\frac{1 - \tilde{\beta}^2}{\tilde{\beta}^2},$$

which tends to 0 in probability by the hypothesis on $\tilde{\beta}$. Q.E.D.

Proof of Proposition 1. $F(x) = 1/(1 + e^{-x})$.

We may write $X_i = Y_i - Y_j^*$. Write $T_n - Q_n^* - Q_n = \Sigma_0^6 R_i$, where R_0, R_3, R_5, and R_6 are the same as in the extreme-value case and

$$R_1 = \sum_1^k (Y_i^* - p_i)^2,$$

$$R_2 = -2\sum_1^k (Y_j^* - p_j)(Y_i^* - p_i),$$

and

$$R_4 = \sum_1^k (X_i - m_i)^2 - Q_n^*.$$

Arguments similar to, but simpler than, those given for the extreme-value distribution show that $R_i/\sigma_n \to 0$ in probability for $i \neq 4$. By symmetry R_4/σ_n has the same as $(R_0 + R_1 + R_2)/\sigma_n$; R_4/σ_n then tends to 0 in probability. Q.E.D.

Proof of Lemma 1. Let $\phi(y) = \log\{y/(1 - e^{-y})\}$. Then $Z_i = \phi(Y_i^*)$. Statement (a) follows from the inequality $0 < \phi(y) < y/2$. To prove (b) and (c) write, for $i \leq l$,

$$Cov(Y_i^*, Y_l^*) = Cov(Y_i^*, Y_l^* - Z_l) + Cov(Y_i^* - Z_i, Z_l) + Cov(Z_i, Z_l).$$

Each term on the right-hand side of this equation is positive; we prove this only for the first. Let

$$\eta(y) = \mathscr{E}\{y + Y_l^* - Y_i^* - \phi(y + Y_l^* - Y_i^*)\}.$$

Condition on Y_i^* to see that

$$Cov(Y_i^*, Y_l^* - Z_l) = Cov\{Y_i^*, \eta(Y_i^*)\}.$$

This is positive, since $\eta(y)$ is an increasing function of y. Similar arguments work for the other terms. Q.E.D.

Proof of Lemma 2. The first inequality in (a) is just Jensen's inequality applied to the relation $X_i = -\log Y_j^*$. Statement (c) is an immediate consequence of (a) and (b). The density of Y_j^* is

$$g(y) = (1 - e^{-y})^{n-i} e^{-yi} B(i, n - i + 1),$$

where $B(x,y) = \Gamma(x+y)/\{\Gamma(x)\Gamma(y)\}$. Temporarily letting $p = p_j$, define $\zeta(y) = \log p + (y-p)/p - (y-p)^2/p^2$. Note that $\zeta(y) \le \log y$ for $y \ge p/2$ and $\zeta(y) \le \log p$ for $y < p/2$. Thus

$$\begin{aligned} -m_i &= \int_0^\infty \zeta(y) g(y)\,dy + \int_0^\infty \{\log y - \zeta(y)\} g(y)\,dy \\ &> \log p - \frac{Var(Y_j^*)}{p^2} - I_1 - \log(p) I_2, \end{aligned}$$

where

$$I_1 = \int_0^{p/2} \{-(\log y) g(y)\}\,dy$$

and

$$I_2 = \int_0^{p/2} g(y)\,dy.$$

Now $\{-(\log y)(1 - e^{-y})\}$ is positive only on $(0,1)$, where it is bounded by 1. A change of variables then shows that

$$\begin{aligned} I_1 &\le \int_{1/e}^1 u^{i-1}(1-u)^{n-i-1} B(i, n-i+1)\,du \\ &\le \int_{1/e}^1 (1-u)^{n-i-1} B(i, n-i+1)\,du \\ &= \frac{(1-e^{-1})^{n-i} B(i, n-i+1)}{n-i}. \end{aligned}$$

The latter is a monotone increasing function of i for $i < n/2$. Hence for each $s < \frac{1}{2}$

$$I_1 < \exp\{-cn(1-s)\} B(ns, n - ns + 1),$$

where $c = -\log(1 - e^{-1})$. Use Stirling's formula to show that for some s this decreases geometrically fast.

An integral comparison shows that

$$-\log \frac{i}{n+1} < p_j < \log t(i,n)$$

where $t(i,n) = (i - \frac{1}{2})/(n + \frac{1}{2})$. A change of variables shows that

$$I_2 < \{1 - t^{\frac{1}{2}}(i,n)\}^{n-i+1} B(i, n-i+1).$$

Let $\delta(x) = \{1 - t^{\frac{1}{2}}(x,n)\}^{n-x+1} B(x, n-x+1)$. Choose s so small that

$$(1-u)\log(1 - u^{\frac{1}{2}}) - u \log u - (1-u)\log(1-u) < 0$$

for all $0 < u < s$. Fix $0 < a < 2^{-\frac{1}{2}}$. Use Stirling's formula to show that

$$\limsup \frac{\log \delta(x_n)}{n^{\frac{1}{2}}} < -a$$

for any sequence x_n such that $1 \le x_n \le ns$. Statement (a) follows.

The function $\log^2(v)$ is concave on (e, ∞); by Jensen's inequality

$$\mathscr{E}(X_i^2) < \log^2(p - I_3) + I_4,$$

where

$$I_3 = \int_0^e yg(y)\,dy$$

and

$$I_4 = \int_0^e \log^2(y)g(y)\,dy.$$

Arguments similar to that given for I_1 show that both I_3 and I_4 decrease uniformly geometrically fast. Q.E.D.

Proof of Proposition 2.

LEMMA 3. *Suppose* $\eta_n(M) = o[\{(\log n)/(n+1)\}^{\frac{1}{2}}]$ *for each* M. *Then* $|\Sigma_{n-k+1}^{n}\lambda_n(X_i) - A_n| \to 0$ *in* P-*probability*.

Since the sequence $n^{\frac{1}{2}}\lambda_n(U_1)$ is uniformly square integrable under P, a variance calculation shows $A_n - \mathscr{E}(A_n)$ tends to 0 in P-probability, establishing the proposition. Q.E.D.

Proof of Lemma 3. The number of U_i such that $nF(U_i) > n - k + Mk^{\frac{1}{2}}$ is binomially distributed with mean asymptotic to k and standard deviation asymptotic to $k^{\frac{1}{2}}$. Hence if $M(n)$ is any sequence tending to ∞, then with probability approaching 1 we have

$$\left|A_n - \sum_{n-k+1}^{n} \lambda_n(X_i)\right| \leq B_n(M), \tag{3.1}$$

where $B_n(M) = \Sigma|\lambda_n(U_i)|1\{|nF(U_i) - (n-k)| < Mk^{\frac{1}{2}}\}$. For any fixed M, $\mathscr{E}\{B_n(M)\} \leq \eta_n(M)Mk^{\frac{1}{2}}$. This tends to 0 for each fixed M by assumption. There is thus a sequence $M(n)$ tending to ∞ for which this conclusion continues to hold. Markov's inequality and (3.1) establish the lemma. Q.E.D.

ACKNOWLEDGEMENTS

The results of Section 2 were originally obtained in the Ph.D. thesis of the first author, written under the supervision of Dr. M.A. Stephens at Simon Fraser University. Both authors thank Dr. Stephens for his inspiration and encouragement. The authors are grateful for the comments of referees which made the work more accurate and more concise.

REFERENCES

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

Fotopolous, S.; Leslie, J.R., and Stephens, M.A. (1984). Approximations for expected values of normal order statistics with an application to goodness-of-fit. Technical Report, Department of Statistics, Stanford University, Stanford, California.

Gerlach, B. (1979). A consistent correlation-type goodness-of-fit test; with application to the two-parameter Weibull distribution. *Math. Operationsforsch. Statist. Ser. Statist.*, 10, 427–452.

Leslie, J.R., Stephens, M.A. and Fotopolous, S. (1986) Asymptotic distribution of the Shapiro-Wilk W test for normality. Ann. Statist., 14, 1497–1506.

Lockhart, R.A. (1985). The asymptotic distribution of the correlation coefficient in testing fit to the exponential distribution. *Canad. J. Statist.*, 13, 253–256.

Smith, R.M., and Bain, L.J. (1976). Correlation-type goodness-of-fit statistics with censored sampling. *Comm. Statist. A—Theory Methods*, 5, 119–132.

Spinelli, J.J., and Stephens, M.A. (1983). Tests for exponentiality with unknown origin. Technical Report, Department of Statistics, Stanford University, Stanford, California.
Stephens, M.A., and D'Agostino, R.B. (1986). *Goodness-of-Fit Techniques*. Marcel Dekker, New York.

Received 1 November 1985
Revised 2 June 1986
Accepted 19 February 1987

Department of Research and Specialist Services
Ministry of Agriculture
P.O. Box 8108, Causeway
Fifth Street extension
Harare
Zimbabwe

Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, Ontario N2L 3G1