Large Sample Theory of Spacings Statistics for Tests of Fit for the Composite Hypothesis

By MARTIN T. WELLS†,   S. RAO JAMMALAMADAKA   and   RAM C. TIWARI

Cornell University,	*University of California,*	*University of North Carolina,*
Ithaca, USA	*Santa Barbara, USA*	*Charlotte, USA*

SUMMARY
Let $X_1, \ldots, X_n$ be a sequence of independent and identically distributed random variables with an unknown underlying continuous cumulative distribution function F. Often we would like to test a null hypothesis concerning the goodness of fit of F to some distribution function which is fully specified or belongs to some parametric family, i.e. in some applications the null hypothesis is simple whereas in others it may be composite. In this paper we present the large sample theory of tests based on non-symmetric functions of sample spacings under composite null hypotheses as well as under contiguous alternatives. Goodness-of-fit tests which are optimal within this class are constructed.

Keywords: GOODNESS OF FIT; NUISANCE PARAMETERS; OPTIMAL TESTS; SPACINGS TESTS

1. INTRODUCTION

Suppose that we have a sample of independent observations $X_1, \ldots, X_{n-1}$ on $\mathbf{R}$ from the family of absolutely continuous distributions given by $\mathscr{F} = \{F(x; \beta_0, \theta) : \theta \in \Theta\}$ where β_0 is a p-dimensional column vector of *known or specified* parameters and θ is a q-dimensional column vector of *unknown* parameters, belonging to a given subset Θ of $\mathbf{R}^q$. We are often interested in testing whether the true distribution function F of the independent and identically distributed sequence $\{X_n\}$ belongs to the family $\mathscr{F}$, i.e. in testing the following composite null hypothesis: $H_0 \colon F \in \mathscr{F}$. Tests of H_0 based on the empirical distribution function type of statistics have been discussed, for example, by Durbin (1973) and for the χ^2-type statistics, for instance, by Moore and Spurill (1975). See D'Agostino and Stephens (1986) for an excellent review of goodness-of-fit tests for the composite null hypotheses. Our aim here is to discuss asymptotic distribution theory of test statistics based on functions of spacings, under H_0 as well as under a sequence of contiguous alternatives $\{A_n\}$. The asymptotic theory for spacings statistics for tests with or without nuisance parameters is given in this paper. Spacings tests of H_0 have been considered by Csörgő and Révész (1980) for unknown location and scale parameters. However, they gave no optimality theory. They discuss the asymptotic behaviour of statistics having a form similar to that of equation (4.6) later. We shall discuss the relationship between these two statistics in more detail later and give the optimality theory as well.

When $\theta = \theta_0$ is a specified value, we can define the one-step uniform spacings as $D_k = F(X_{(k)}; \beta_0, \theta_0) - F(X_{(k-1)}; \beta_0, \theta_0)$ $(k = 1, \ldots, n)$, with $F(X_{(0)}; \beta_0, \theta_0) = 0$ and $F(X_{(n)}; \beta_0, \theta_0) = 1$, where $X_{(k)}$ is the kth-order statistic from a sample of size $n - 1$.

†*Address for correspondence*: Department of Economic and Social Statistics, Cornell University, Ithaca, NY 14851-0952, USA.

Let $h(\ )$ and $\{h_k(\ ),\ k = 1, \ldots, n\}$ be real-valued functions satisfying some regularity conditions. Consider the spacings statistics

$$T_n^* = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} h(nD_k),$$

$$T_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} h_k(nD_k). \tag{1.1}$$

Although T_n^* is symmetric in $\{D_k\}_{k=1}^n$, T_n is not necessarily symmetric. See Pyke (1965) and D'Agostino and Stephens (1986) for good references on theory of spacings statistics.

Sethuraman and Rao (1970) and Rao and Sethuraman (1975) studied T_n^* through the weak convergence of the empirical spacings process. They show that the class of symmetric tests cannot discriminate between alternatives converging to H_0 at a rate faster than $n^{-1/4}$. Hence T_n^* has poor asymptotic performance compared with, say, the Kolmogorov–Smirnov, Cramér–von-Mises statistic for goodness of fit. It is also shown that $h(x) = x^2$, resulting in the so-called Greenwood test, is asymptotically most powerful in this class (see Rao and Kuo (1984)). Holst and Rao (1981) investigate non-symmetric spacings statistics and have found that the class of non-symmetric tests can discriminate between alternatives converging to H_0 at a rate of $n^{-1/2}$, as in the Kolmogorov–Smirnov, Cramér–von-Mises tests. Examples of T_n include symmetric functions of spacings and linear combinations of spacings.

We can also define higher order uniform spacings or m-step overlapping and non-overlapping uniform spacings and discuss the associated asymptotic distribution theory (see for example Cressie (1976), Vasicek (1976), Del Pino (1979) and Kuo and Rao (1981)). Jammalamadaka *et al.* (1989) and Hall (1986) consider m-step spacings and obtain the asymptotic distribution by letting the spacing step length tend to infinity along with the sample size. Although the results of this paper may be generalized to statistics based on higher order spacings, it will not be done here for simplicity.

In the next section, we discuss the limit theory for $\hat{T}_n$ under a sequence of contiguous alternatives $\{A_n\}$, where $\hat{T}_n$ is a version of T_n with the estimated parameters substituted for the unknown parameters. Section 3 derives the uniformly most powerful test, under $\{A_n\}$, for the composite hypothesis. In Section 4, we give several examples and provide a comparison of the Monte Carlo powers of the proposed statistics with some other well-known goodness-of-fit statistics. Appendix A contains most of the proofs.

2.   LIMIT THEORY FOR $\hat{T}_n$

Consider the parametric family of distribution functions given by $\mathscr{F}(\beta_0, \theta)$. It is desired to test the null hypothesis $H_0: F \in \mathscr{F}$ against the sequence of alternatives $A_n: F \in \mathscr{F}(\beta_n, \theta)$ satisfying assumption 3 below. We discuss the distribution theory of $\hat{T}_n$ under $\{A_n\}$. Since θ is unknown, the usual probability integral transform cannot be applied. In the presence of the unknown parameters, define $\hat{U}_{(k)} = F(X_{(k)}; \beta_0, \hat{\theta}_n)$, with $\hat{U}_{(0)} = 0$ *and* $\hat{U}_{(n)} = 1$, where $X_{(k)}$ denotes the kth-order statistic for the sample of the Xs and $\hat{\theta}_n$ is an estimate of θ (see assumption 4). Define $\hat{D}_k = \hat{U}_{(k)} - \hat{U}_{(k-1)}\ (k =$

$1, \ldots, n$) as the one-step spacings with the estimated parameters. Consider the non-symmetric spacings statistic

$$\hat{T}_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} h_k(n\hat{D}_k). \tag{2.1}$$

Throughout this paper we shall let W denote an exponential random variable with $E(W) = 1$, i.e. W has density $\exp(-w)$, $w \geqslant 0$. Let

$$\zeta_k = \zeta_{nk} = (k - \tfrac{1}{2})/n \qquad (k = 1, \ldots, n). \tag{2.2}$$

To deduce the large sample theory for $\hat{T}_n$ we shall need several assumptions. See Kuo and Rao (1981) for further details on the class of functions $\{h_k(\ )\}$. Assumptions 1 and 2 below are assumptions on the class of functions $\{h_k\}$, whereas assumptions 3–5 are the usual type of assumptions needed to study problems of goodness of fit with nuisance parameters (see, for instance, Durbin (1973) and Moore and Spurill (1975)).

Assumption 1. The function $h_k(\ )$ is of the form

$$h_k(x) = h(x, \zeta_k) \qquad (k = 1, \ldots, n) \tag{2.3}$$

where $h(x, y)$ is defined on $[0, \infty] \times [0, 1]$. For any fixed y, assume that $h_x = (\partial/\partial_x)h$ and $h_{xx} = (\partial^2/\partial_x^2)h$ are continuous with the following properties:

$$\int_0^1 E\{h(W, y)\}^3 \, dy < \infty; \tag{2.4}$$

$$\int_0^1 E\{h_{xx}(W, y)\} \, dy < \infty; \tag{2.5}$$

$$\text{var}\{h_x(W, y)\} < \infty, \qquad \text{for } 0 \leqslant y \leqslant 1. \tag{2.6}$$

Assumption 2. (Without loss of generality) $E\{h(W, y)\} = 0$ for all $0 \leqslant y \leqslant 1$.

Concerning the sequence of alternatives, we consider the contiguous alternatives studied by Durbin (1973) and Moore and Spurill (1975). For this we shall assume the following.

Assumption 3. For some p-dimensional vector γ, $\sqrt{n}\,(\beta_n - \beta_0) \to \gamma$ as $n \to \infty$.

A typical example of the set-up is testing the null hypothesis that a sample is normally distributed with unknown mean θ and specified variance β_0. The alternative of interest is $\beta_n = \beta_0 = \gamma/\sqrt{n}$, the shift on the variance parameter. The asymptotic distribution theory under $\{A_n\}$ will allow us to discuss asymptotic power and efficiency of the test procedure.

Let Λ denote the closure of a given neighbourhood of θ_0, the true unknown value of θ, and of β_0, the value of β specified under H_0. Define $\xi_n \equiv \sqrt{n}\,(\hat{\theta}_n - \theta_0)$. Concerning the sequence of estimators of θ we shall assume the following.

Assumption 4. Under the sequence of alternatives $\{A_n\}$ the estimator of the nuisance parameter has the form

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = n^{-1/2} \sum_{j=1}^{n} J(X_j; \beta_0, \theta_0) + A\gamma + \epsilon_n, \tag{2.7}$$

where A is a given finite matrix of order $q \times p$, J is a measurable q-dimensional vector-valued function and ϵ_n converges to 0 in probability as $n \to \infty$. Assume that

(a) $E\{J(X_j; \beta_0, \theta_0) | A_n\} = 0$ and
(b) $E\{J(X_j; \beta_0, \theta_0) J(X_j; \beta_0, \theta_0)^t | A_n\} = I^{-1}(\beta_n, \theta_0),$

where $I^{-1}(\beta_n, \theta_0)$ is the information matrix which converges to a finite non-negative definite matrix $I^{-1} = I^{-1}(\beta_0, \theta_0)$ as $n \to \infty$, i.e. $\hat{\theta}_n$ is an asymptotically efficient estimator of θ_0. Assume that $E|\xi_n|^3 < \infty$ for all n.

Assumption 5.

(a)   The vector-valued functions $g_\beta(u; \beta, \theta) = (\partial/\partial\beta) F(x; \beta, \theta)$ and $g_\theta(u; \beta, \theta) = (\partial/\partial\theta) F(x; \beta, \theta)$ are uniformly continuous in $u \in (0, 1)$ for all $(\beta, \theta) \in \Lambda$, where the right-hand sides of each of these functions are expressed as a function of u by means of the transformation $u = F(x; \beta, \theta)$.
(b)   The functions g_β and g_θ are uniformly bounded in u for $(\beta, \theta) \in \Lambda$. Also $(\partial/\partial\alpha) g_\alpha$ ($\alpha = \beta, \theta$) are uniformly bounded in u for $(\beta, \theta) \in \Lambda$.
(c)   The functions $\dot{g}_\alpha(u; \beta, \theta) = (d/du) g_\alpha(u; \beta, \theta)$, for $\alpha = \beta, \theta$, are uniformly continuous in u for $(\beta, \theta) \in \Lambda$.

In what follows, we use the notation $\sim$, $\to^D$ and $\to^P$ to denote distributed as, convergence in distribution and convergence in probability respectively. Also let $\gamma_n \equiv \sqrt{n}(\beta_n - \beta_0)$.

As a consequence of representation (2.7) and the central limit theorem the following result may be deduced.

Theorem 1. Under assumption 4 and the sequence of alternatives $\{A_n\}$,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} \xi \sim N_q(A\gamma, I^{-1}) \qquad \text{as } n \to \infty.$$

We need the following result of Holst and Rao (1981) concerning the spacings statistics T_n, when the distribution under H_0 is completely specified as $F(x; \beta_0, \theta_0)$. Let

$$D_k = F(X_{(k)}; \beta_0, \theta_0) - F(X_{(k-1)}; \beta_0, \theta_0) \qquad (k = 1, \ldots, n) \qquad (2.8)$$

and T_n be given by statistics (1.1).

Theorem 2 (Holst and Rao, 1981). Let $h_k(\ )$ satisfy assumptions 1 and 2. Then, under hypothesis H_0,

$$T_n \xrightarrow{D} T$$

as $n \to \infty$ where $T \sim N(0, \sigma^2)$ with

$$\sigma^2 = \int_0^1 \text{var}\{h(W, y)\} dy - \left[\int_0^1 \text{cov}\{h(W, y), W\} dy \right]^2. \qquad (2.9)$$

A Taylor series expansion of $\hat{T}_n$ in equation (2.2) around θ_0 and β_0 gives

$$\hat{T}_n = T_n + \sqrt{n}(\hat{\theta}_n - \theta_0)^t \psi_{1n} + \sqrt{n}(\beta_n - \beta_0)^t \psi_{2n} + o_p(1) \qquad (2.10)$$

where by an application of the mean value theorem

$$\psi_{1n} = \frac{1}{n} \sum_{k=1}^{n} \frac{\partial}{\partial \theta} h(nD_k, \zeta_k)$$

$$= \frac{1}{n} \sum_{k=1}^{n} n \, h_x(nD_k, \zeta_k) \frac{\partial}{\partial \theta} \{F(X_{(k)}; \beta_0, \theta_0) - F(X_{(k-1)}; \beta_0, \theta_0)\}$$

$$= \frac{1}{n} \sum_{k=1}^{n} nD_k \, h_x(nD_k, \zeta_k) \dot{g}_\theta(\tilde{U}_{1k}), \qquad \text{for } U_{(k-1)} \leqslant \tilde{U}_{1k} \leqslant U_{(k)}, \qquad (2.11)$$

$$\psi_{2n} = \frac{1}{n} \sum_{k=1}^{n} nD_k \, h_x(nD_k, \zeta_k) \dot{g}_\beta(\tilde{U}_{2k}), \qquad \text{for } U_{(k-1)} \leqslant \tilde{U}_{2k} \leqslant U_{(k)}. \qquad (2.12)$$

Therefore equation (2.10) can be expressed as

$$\hat{T}_n = T_n + \xi_n^t \psi_{1n} + \gamma_n^t \psi_{2n} + o_p(1). \qquad (2.13)$$

In Appendix A we prove several preliminary lemmas to show the following main result with regard to the asymptotic distribution of $\hat{T}_n$.

Theorem 3. Under assumptions 1–5 and the sequence of alternatives $\{A_n\}$,

$$\hat{T}_n \xrightarrow{D} \hat{T} \sim N\{\gamma^t(A^t\psi_1 + \psi_2), \sigma^2 - \psi_1^t I^{-1} \psi_1\} \qquad \text{as } n \to \infty,$$

where

$$\psi_1 = \int_0^1 EW h_x(W, y) \dot{g}_\theta(y) dy$$

and

$$\psi_2 = \int_0^1 EW h_x(W, y) \dot{g}_\beta(y) dy.$$

Proof. By lemma 3 (Appendix A), $(T_n, \xi_n)^t$ has a limiting normal distribution; therefore, by representation (2.13), $\hat{T}_n$ will have a limiting normal distribution. Also, since $E(T) = 0, E(\xi) = A\gamma, \gamma_n \to \gamma$,

$$\psi_{1n} \xrightarrow{p} \psi_1$$

and

$$\psi_{2n} \xrightarrow{p} \psi_2,$$

as $n \to \infty$, $E(\hat{T})$ can easily be found by using representation (2.13). The variance term is given in lemma 4. $\square$

Setting $\gamma = 0$ in theorem 3 gives the asymptotic distribution of $\hat{T}_n$ under the null hypothesis. In addition, setting $\psi_1 \equiv 0$ in theorem 3 gives the results of Holst and Rao (1981) under the sequence $\{A_n\}$. Theorem 3 generalizes the asymptotic distribution for the shift of p parameters, not just a single parameter as discussed by Holst and Rao

(1981). If $\gamma = 0$ and $p = 1$ we are back to theorem 2. An application of theorem 3 yields the following two corollaries.

Corollary 1. Under H_0,

$$\hat{T}_n \overset{D}{\to} \hat{T} \sim N(0, \sigma^2 - \psi_1^t I^{-1} \psi_1) \qquad \text{as } n \to \infty.$$

Corollary 2. Under equations (A.1) and (A.2) and the alternatives $\{A_n\}$,

$$T_n \overset{D}{\to} T \sim N(\gamma^t \psi_2, \sigma^2) \qquad \text{as } n \to \infty.$$

It is clear that the asymptotic distributions discussed above may depend on the unknown parameters. However, since $\hat{\theta}_n - \theta_0 = O_p(n^{-1/2})$ we can replace θ_0 by the estimate $\hat{\theta}_n$ and apply Slutsky's theorem to yield the following.

Corollary 3. Under the conditions of theorem 3

$$\frac{\hat{T}_n - \mu(\beta_0, \hat{\theta}_n)}{\lambda(\hat{\theta}_n)} \overset{D}{\to} N(0, 1) \qquad \text{as } n \to \infty,$$

where

$$\mu(\beta_0, \hat{\theta}_n) = \gamma^t \{A^t \psi_1(\beta_0, \hat{\theta}_n) + \psi_2(\beta_0, \hat{\theta}_n)\}$$

and

$$\lambda^2(\hat{\theta}_n) = \sigma^2 - \psi_1^t(\beta_0, \hat{\theta}_n) I^{-1}(\beta_0, \hat{\theta}_n) \psi_1(\beta_0, \hat{\theta}_n).$$

3. LOCALLY MOST POWERFUL TESTS

We now turn our attention to the question of finding the asymptotically most powerful test, i.e. the test with the highest Pitman asymptotic relative efficiency (ARE). For more information on the ARE see Serfling (1980). The efficacy of the test statistic $\hat{T}_n$ is given by

$$e(h_k) = \mu^2(h_k)/\sigma^2(h_k), \qquad (3.1)$$

where $\mu(h_k) = \gamma^t(\psi_2 + A^t \psi_1)$ and $\sigma^2(h_k) = \sigma^2 - \psi_1^t I^{-1} \psi_1$ are the asymptotic mean and variance in theorem 3. The test with the maximum efficacy has asymptotically maximum local power. In the light of theorem 3, to find such a test, against the sequence of alternatives $\{A_n\}$, we need to find the optimal function $h_k(\ )$ that maximizes $e(h_k)$ in equation (3.1). This is given by the following (see Appendix A for the proof).

Theorem 4. The value of $e(h_k)$ in equation (3.1) is maximized by taking

$$h(x, y) = \{\gamma^t \dot{g}_\beta(y) + (A\gamma)^t \dot{g}_\theta(y)\}x.$$

If there is no estimation involved, then $\dot{g}_\theta(y) \equiv 0$ and the choice of the optimal function is $h(x, y) = \gamma^t \dot{g}_\beta(y)x$. This extends lemma 3.1 of Holst and Rao (1981) which gives the optimal function for a shift of one parameter. It is also interesting that in the case of no estimation the optimal test statistic is a linear combination of spacings (hence, also a linear combination of order statistics) with the weights being the derivatives of the log-likelihood function with respect to the parameter shifted in $\{A_n\}$, i.e. the weights are the functions used in obtaining the 'locally most powerful tests'. In the estimated parameters case again the optimal test statistic is a linear

combination of spacings. However, now the weights are similar to the functions used in the 'Neyman $C(\alpha)$' tests for locally most powerful tests of a composite hypothesis. Note that the expression $(A\gamma)^t \dot{g}_\theta(Y)$ is the regression coefficient of $\dot{g}_\theta(Y)$ onto $\dot{g}_\beta(Y)$ and that $E[\{\gamma^t \dot{g}_\beta(Y) + (A\gamma)^t \dot{g}_\theta(Y)\} \dot{g}_\theta(Y)] = 0$.

As remarked in Section 2, when the optimal test statistic and asymptotic distributions discussed above depend on the unknown parameters, we can replace θ_0 by the estimate $\hat{\theta}_n$ and apply corollary 3.

4.   EXAMPLES AND MONTE CARLO COMPARISONS

In this section we provide some examples of the results obtained in the previous sections. We are primarily concerned with examples of location–scale families where the estimates of the unknown parameters are maximum likelihood estimates. We also give a Monte Carlo study of the power of competing test procedures.

4.1.   *Example 1*

Let X be a random variable with a fixed known distribution G. Let $\theta = (\theta_1, \theta_2)$ and define $G(x; \theta_1, \theta_2) = F\{(x - \theta_1)/\theta_2\}$.

4.1.1.   *Location unknown and scale shifted under $\{A_n\}$*

Suppose that we wish to test $H_0:G = F\{(x - \theta_1)/\theta_{20}\}$ versus $A_n:G = F\{(x - \theta_1)/\theta_{2n}\}$ where $\theta_{2n} = \theta_{20} + \gamma/\sqrt{n}$ and $\theta_1 \in \mathbf{R}$. We reject H_0 if

$$\hat{T}_n = \frac{1}{n} \sum_{k=1}^{n} h(nD_k, \zeta_k) > c_\alpha$$

where c_α is a critical value to be determined from a table of the normal distribution. By theorem 1, the optimal score function is given by (where prime denotes the derivative)

$$h(x, u) = -\frac{\gamma}{\theta_{20}} \left[1 + \frac{F^{-1}(u) f'\{F^{-1}(u)\}}{f\{F^{-1}(u)\}} + \frac{i_{12}}{i_{11}} \frac{f'\{F^{-1}(u)\}}{f\{F^{-1}(u)\}} \right] x, \qquad (4.1)$$

where

$$i_{11} = \int_{-\infty}^{+\infty} \left\{ \frac{f'(x)}{f(x)} \right\}^2 f(x) \, dx$$

and

$$i_{12} = \int_{-\infty}^{+\infty} x \left\{ \frac{f'(x)}{f(x)} \right\}^2 f(x) \, dx.$$

(a)   If F is the normal distribution, the optimal function is given by

$$h(x, u) = -\frac{\gamma}{\theta_{20}} [1 - \{\Phi^{-1}(u)\}^2] x. \qquad (4.2)$$

(b)   If F is the logistic distribution, the optimal function is given by

$$h(x, u) = -\frac{\gamma}{\theta_{20}} \left\{ 1 - (1 - 2u) \ln \left(\frac{u}{1 - u} \right) \right\} x.$$

(c)   If F is the extreme value distribution,

$$F(x) = 1 - \exp\left\{-\exp\left(\frac{x - \theta_1}{\theta_2}\right)\right\}, \qquad -\infty < x < \infty,$$

then the information matrix is given by

$$I = \frac{1}{\theta_{20}^2}\begin{pmatrix} 1 & \tilde{\gamma} - 1 \\ \tilde{\gamma} - 1 & c^2 \end{pmatrix}, \tag{4.3}$$

where $\tilde{\gamma} = 0.57712$ is Euler's constant, $c^2 = \Gamma''(3) + \Gamma''(1) - 2\Gamma''(2) - 1 = \pi^2/6 + (\tilde{\gamma} - 1)^2 = 1.82368$ giving $c = 1.350437$ and $\Gamma^{(k)}(x)$ is the kth-derivative of the gamma function $\Gamma(x)$. In this case the optimal function is given by

$$h(x, u) = \frac{1}{\theta_{20}}\{1 + (1 + \ln u)\ln(-\ln u) - (\tilde{\gamma} - 1)(1 + \ln u)\}x. \tag{4.4}$$

In examples (a) and (b) the distributions are symmetric about θ_1 and hence $i_{12} = 0$, i.e. the parameters are orthogonal. In orthogonality, the optimal score function reduces to that corresponding to no estimation.

4.1.2. *Scale unknown and location shifted under* $\{A_n\}$.

Suppose that we wish to test $H_0: G = F\{(x - \theta_{10})/\theta_2\}$ *versus* $A_n: G = F\{(x - \theta_{1n})/\theta_2\}$ where $\theta_{1n} = \theta_{10} + \gamma/\sqrt{n}$ and $\theta_2 > 0$. We reject H_0 if

$$\hat{T}_n = \frac{1}{n}\sum_{k=1}^{n} h(nD_k, \zeta_k) > c_\alpha, \tag{4.5}$$

where c_α is the critical value to be determined from a table of the normal distribution. By theorem 4, the optimal function is given by

$$h(x, u) = -\frac{\gamma}{\theta_2}\left[\frac{f'\{F^{-1}(u)\}}{f\{F^{-1}(u)\}} + \frac{i_{12}}{i_{11}}\left(1 + \frac{F^{-1}(u)f'\{F^{-1}(u)\}}{f\{F^{-1}(u)\}}\right)\right]x. \tag{4.6}$$

Since θ_2 is unknown we shall replace θ_2 by a consistent estimate $\hat{\theta}_2$ (see corollary 3).

(a)   If F is normal the optimal h is given by the 'normal scores statistic'

$$h(x, u) = \gamma\Phi^{-1}(u)x/\hat{\theta}_2. \tag{4.7}$$

(b)   If F is the logistic distribution then the optimal h is given by $h(x, u) = -\gamma(1 - 2u)x/\hat{\theta}_2$.

(c)   If F is the extreme value distribution, the information matrix is given by equation (4.4) and the optimal h is

$$h(x, u) = -\frac{\gamma}{\hat{\theta}_2}\left\{(1 - \ln u) + \frac{1 - \tilde{\gamma}}{c^2}(1 + \ln u)\ln(-\ln u)x\right\}. \tag{4.8}$$

A convenient estimator of θ_2, which satisfies assumption 4, proposed by Weiss (1961) is $\hat{\theta}_2 = \Sigma_{k=1}^{n} f\{F^{-1}(\zeta_k)\}D_k$. Csörgő and Révész (1980) proposed using the statistic constructed from equation (4.6) with the second term in the square brackets equal to 0 and with Weiss's scale estimate in the denominator. Similar to the unknown location case, $i_{12} = 0$ if F is a symmetric distribution. Hence the optimal score

function is the same in the case of no estimation, apart from the estimate of θ_2. Therefore, in the location–scale problem with $i_{12} = 0$, Csörgő and Révész's statistic will have the same optimality properties.

4.2. *Example 2*

In this example we discuss the goodness of fit of a simple hypothesis when more than one parameter is shifted under $\{A_n\}$. Through a probability integral transformation, the simple goodness-of-fit problem is equivalent to testing whether a sample is from the uniform distribution on [0, 1], i.e. $H_0:F(x) = x, 0 \leqslant x \leqslant 1$. The local alternatives usually discussed are of the form $A_n:F(x) = x + M(x)/\sqrt{n}, 0 \leqslant x \leqslant 1$, where $M(0) = M(1) = 0$ and M satisfies certain regularity conditions. It was shown by Holst and Rao (1981) that the asymptotically locally most powerful test of H_0 against A_n is based on the statistic

$$T_n = \frac{1}{n} \sum_{k=1}^{n} m(\zeta_k)D_k$$

where $m(u) = (d/du)M(u)$. We can use theorem 4 to deduce this result. In our case we have p parameters shifted under the alternatives, and the asymptotically locally most powerful test is given by

$$T_n = \frac{1}{n} \sum_{k=1}^{n} \gamma^t \dot{g}_\beta(\zeta_k)D_k. \tag{4.9}$$

In the one-parameter case, $\gamma^t \dot{g}_\beta(u) = m(u)$. However, if we are interested in guarding against more than one parameter shift we would use test (4.9).

We shall now report the results of a Monte Carlo study on the power of the proposed test statistic. A power comparison of our statistic ($\hat{T}$, with h given in equation (4.7)) with the Anderson–Darling (A^2), Cramér–von-Mises (W^2), Rao–Robson (χ^2) and Shapiro–Francia (W') statistics for testing normality (location and scale unknown) is given. See D'Agostino and Stephens (1986) for the details on A^2, W^2, χ^2 and W'. The simulation study is in the style of Stephens (1974) who compared several goodness-of-fit statistics for testing normality against various distributions. The critical points for A^2 and W^2 are found in Stephens (1974), for χ^2 in the χ^2-table and for W' in Shapiro and Francia (1972). We omit the Kolmogorov–Smirnov statistic from our comparison since Stephens (1974) demonstrated that A^2 and W^2 generally perform better. The proposed spacing statistic we chose is the 'normal scores' statistic based on equation (4.7) which is optimal for shifts in location for the normal distribution. Note that the normal scores functional is a major component in the optimal test for scale shift in equation (4.2). As proposed previously, we shall use the estimate suggested by Weiss (1961) as an auxiliary estimate of the scale parameter needed in equation (4.7). Table 1 gives the percentage of 2000 Monte Carlo samples each of size 50 or 90 drawn from the given population which were declared significant by the reported statistics. The size of the tests is set at $\alpha = 0.10$. We have chosen the alternative distributions that were used by Stephens (1974). The results are given in Table 1.

The results of the simulation are as follows. The statistics $\hat{T}$ and W' appear to be

TABLE 1
Power comparisons and tests for normality

Population (θ_1, θ_2)	n	A^2	W^2	χ^2	W'	$\hat{T}$
Uniform (0, 1.8)	50	71	65	61	88	89
	90	92	84	76	99	99
Exponential (4, 9)	50	99	99	96	99	99
	90	100	100	100	100	100
Laplace (0, 6)	50	72	69	64	63	75
	90	83	84	73	82	86
Weibull, $k = 5$ (0.63, 3.25)	50	42	38	33	48	46
	90	76	72	67	78	77
Tukey, $\lambda = 5$ (0, 2.9)	50	35	37	33	29	42
	90	59	61	48	39	65
Student t_4	50	49	47	42	41	52
	90	67	66	59	65	71

slightly better than A^2, the best competitor. Overall, $\hat{T}$ and W' perform nearly equally; however, $\hat{T}$ has a slight edge. $\hat{T}$ is much easier to calculate. Clearly, A^2 and W^2 are superior to χ^2. These results virtually tell the same story as the results in Stephens (1974).

It is really not surprising that $\hat{T}$ and W' have similar properties since W' may be expressed as a square of a linear combination of order statistics divided by a scale estimate (see D'Agostino and Stephens (1986), p. 399) and, on rearranging the terms, W' equals the square of a linear combination of spacings divided by a scale estimate. However, the weights of the linear combination of order statistics for W' are the expected values of normal order statistics, not the optimal weights in theorem 4 (asymptotically these weights are quite similar; see D'Agostino and Stephens (1986), p. 400). Therefore, $\hat{T}$ ought to be more powerful, as was shown in the simulation study. It would be of interest to pin down the exact relationship between these two statistics.

For distributions other than the normal distribution, theorem 4 can be used to derive the optimal test statistic. In the case of no estimation, the optimal function is a linear combination of spacings with the weights being the derivatives of the log-likelihood function with respect to the parameter shifted in $\{A_n\}$. The weights are the functions used in obtaining the locally most powerful tests and optimal score functions. In the estimated parameters case, again, the optimal function is a linear combination of spacings. However, now the weights are similar to the functions used in the Neyman $C(\alpha)$ tests for locally most powerful tests of a composite hypothesis.

As pointed out to us by a referee, the forms of other test statistics have a more intuitive rationale. For instance, when testing normality the Shapiro–Francia statistic may be interpreted as a correlation coefficient on the quantile plot and Vasicek's spacing test has an interpretation as a measure of entropy. Goodness-of-fit tests which have a nice intuitive rationale are usually based on a particular characterization of the distribution being tested. However, our goal was to derive the form of the optimal tests under a sequence of contiguous alternatives. As a result, the forms of our test statistics do not have as nice an interpretation as other more intuitive procedures. There is a trade-off with intuitive tests for procedures which are optimal. It also

turns out that tests which have a nice intuitive rationale for a particular distribution may not have for another distribution. Such is the case with the Shapiro–Francia and Vasicek's tests and hence these tests may not be appropriate for testing the goodness of fit for a variety of distributions. However, this is not the case when the procedure is constructed on the basis of an optimality criterion.

ACKNOWLEDGEMENTS

APPENDIX A

Lemma 1. let $L(\ , \ )$ satisfy assumption 1. Then

$$\frac{1}{n} \sum_{k=1}^{n} L(nD_k, \zeta_k) \xrightarrow{\text{P}} \int_0^1 E L(W, u) \, du \qquad \text{as } n \to \infty.$$

Proof. Since $\{nD_i\}_{i=1}^n$ are distributed as independent exponential random variables divided by their mean, on a common probability space we can construct independent and identically distributed random variables $W_i \sim \exp(1)$ with $n\overline{W}_n = \Sigma_{k=1}^n W_k$, so that

$$\frac{1}{n} \sum_{k=1}^{n} L(nD_k, \zeta_k) = \frac{1}{n} \sum_{k=1}^{n} L\left(\frac{W_k}{\overline{W}_n}, \zeta_k\right) \tag{A.1}$$

$$= \frac{1}{n} \sum_{k=1}^{n} \left\{ L(W_k, \zeta_k) - L_x(W_k, \zeta_k) \frac{W_k(\overline{W}_n - 1)}{\overline{W}_n} \right\} + O_p(1), \tag{A.2}$$

By the weak law of large numbers

$$\overline{W}_n - 1 \xrightarrow{\text{P}} 0.$$

Hence the second term of equation (A.2) tends to 0 in probability. Note that

$$\left| \frac{1}{n} \sum_{k=1}^{n} L(W_k, \zeta_k) - \int_0^1 E L(W, u) \, du \right| \leqslant \left| \frac{1}{n} \sum_{k=1}^{n} \{ L(W_k, \zeta_k) - E L(W, \zeta_k) \} \right|$$

$$+ \left| \frac{1}{n} \sum_{k=1}^{n} \{ E L(W, \zeta_k) - \int_0^1 E L(W, u) \, du \} \right|. \tag{A.3}$$

Since $L(x, \ )$ is continuous, the first term of inequality (A.3) tends to 0 in probability by the weak law of large numbers. The second term in inequality (A.3) tends to 0 by the Lebesgue dominated convergence theorem.   □

Lemma 2.

$$\psi_{1n} \xrightarrow{\text{P}} \psi_1 = \int_0^1 E W h_x(W, y) \dot{g}_\bullet(y) \, dy$$

and

$$\psi_{2n} \xrightarrow{\text{p}} \psi_2 = \int_0^1 EW h_x(W, y) \dot{g}_\beta(y) \, dy \qquad \text{as } n \to \infty.$$

Proof. From the existence of the limiting distribution of the Kolmogorov–Smirnov statistic we have

$$\sup_{1 \leqslant k \leqslant n} (n^{1/2} | U_{(k)} - \zeta_k |) = O_p(1).$$

Since $\dot{g}_\theta(\ )$ is assumed to be continuous,

$$\sup_{1 \leqslant k \leqslant n} \{ | \dot{g}_\theta(\tilde{U}_{1k}) - \dot{g}_\theta(\zeta_k) | \} = o_p(1),$$

where $\dot{g}_\theta(\tilde{U}_{1k})$ is defined in equation (2.11), it follows that

$$\left| \frac{1}{n} \sum_{k=1}^n nD_k h_x(nD_k, \zeta_k) \dot{g}_\theta(\tilde{U}_{1k}) - \frac{1}{n} \sum_{k=1}^n nD_k h_x(nD_k, \zeta_k) \dot{g}_\theta(\zeta_k) \right|$$

$$\leqslant \left| \frac{1}{n} \sum_{k=1}^n nD_k h_x(nD_k, \zeta_k) \right| \sup_{1 \leqslant k \leqslant n} \{ | \dot{g}_\theta(\tilde{U}_{1k}) - \dot{g}_\theta(\zeta_k) | \} = o_p(1).$$

Thus, by lemma 1 with $L(x, y) = x h_x(x, y) \dot{g}_\theta(y)$, we have

$$\psi_{1n} \xrightarrow{\text{p}} \psi_1 \qquad \text{as } n \to \infty.$$

By the same argument with $\dot{g}_\theta(\tilde{U}_{1k})$ replaced by $\dot{g}_\beta(\tilde{U}_{2k})$ we have

$$\psi_{2n} \xrightarrow{\text{p}} \psi_2 \qquad \text{as } n \to \infty. \qquad \square$$

Lemma 3.

$$(T_n, \xi_n)^t \xrightarrow{\text{D}} (T, \xi)^t \sim N_{q+1}(\mu, \Sigma) \qquad \text{as } n \to \infty,$$

where $\mu^t = (0 : A\gamma)$ and

$$\Sigma = \begin{pmatrix} \sigma^2 & \Sigma_{12} \\ \Sigma_{21} & I^{-1} \end{pmatrix},$$

and Σ_{12} will be defined in lemma 4.

Proof. By theorem 1,

$$\xi_n \xrightarrow{\text{D}} N_q(A\gamma, I^{-1})$$

and by theorem 2

$$T_n \xrightarrow{\text{D}} N(0, \sigma^2).$$

To prove that (T_n, ξ_n) has a limiting joint normal distribution it is sufficient to show that the linear combination

$$\tilde{Y}_k = h_k(nD_k) + \sum_{j=1}^q \alpha_j J_j(X_k; \beta_0, \theta_0), \qquad \alpha_j \in \mathbf{R},$$

satisfies Liapunov's condition (see Chung (1968), p. 201). Construct a sequence $\{Y_k\}$ on the same probability space as $\{\tilde{Y}_k\}$ where

$$Y_k = h_k(W_k) + \sum_{j=1}^{q} \alpha_j J_j(X_k; \beta_0, \theta_0) \qquad (A.4)$$

and $J = (J_1, \ldots, J_q)^t$ is the vector-valued function discussed in equation (2.7). Expanding $h_k(nD_k)$ in terms of exponential random variables $\{W_k\}_{k=1}^{n}$ and their mean $\bar{W}_n$ as in equation (A.1) it is easily seen that

$$|\tilde{Y}_k - Y_k| \overset{p}{\to} 0$$

uniformly in k. Hence it suffices to show that

$$\left(\sum_{k=1}^{n} E|Y_k|^3\right)^{1/3} \bigg/ \sum_{k=1}^{n} (E|Y_k|^2)^{1/2} \overset{p}{\to} 0 \qquad \text{as } n \to \infty \qquad (A.5)$$

for all real numbers $(\alpha_1, \ldots, \alpha_q)$. Since $E|h_k(W)|^3 < \infty$ (by property (2.4)) and

$$E|\sum_{j=1}^{q} \alpha_j J_j(X_k; \beta_0, \theta_0)|^3 < \infty,$$

it follows from Minkowski's inequality (see Chung (1968), p. 41) that $E|Y_k|^3 < \infty$. Hence the numerator in expression (A.5) is of order $n^{1/3}$. Also, since

$$\frac{1}{n} \sum_{k=1}^{n} \text{var}(Y_k)$$

tends to a non-zero constant, it follows that $\Sigma_{k=1}^{n} \text{var}(Y_k)$ is of order n. Hence expression (A.5) tends to 0 in probability as $n \to \infty$. $\qquad\square$

Lemma 4. The covariance matrix Σ_{12} of lemma 3 equals $-\psi_1^t I^{-1}$. Hence $\text{var}(\hat{T}) = \sigma^2 - \psi_1^t I^{-1} \psi_1$.

Proof. We first show that $\hat{T}$ and ξ are uncorrelated. Since $\hat{T}_n = T_n + \xi_n^t \psi_{1n} + \gamma_n^t \psi_{2n} + o_p(1)$, $E(T) = 0$ and $E(\xi) = A\gamma$, we see that $E\{\hat{T} - (A\gamma)^t \psi_1 - \gamma^t \psi_2\} = 0$. By the asymptotic efficiency of $\hat{\theta}_n$, we have

$$\text{var}(\xi^t) \leqslant \text{var}[\xi^t + \lambda\{\hat{T} - (A\gamma)^t\psi_1 - \gamma^t\psi_2\}] = \text{var}(\xi^t) + \lambda^2 \text{var}(\hat{T}) + 2\lambda \text{cov}(\xi^t, \hat{T}),$$

for all λ. However, for λ in the interval $[-2\text{cov}(\xi, \hat{T})/\text{var}(\hat{T}), 0]$, $\text{var}(\xi^t) \geqslant \text{var}[\xi^t + \lambda\{\hat{T} - (A\gamma)^t \psi_1 - \gamma^t\psi_2\}]$ unless $\text{cov}(\xi^t, \hat{T}) = 0$. By representation (2.13) we have $\text{var}(\hat{T}) = \sigma^2 + \psi_1^t I^{-1}\psi_1 + \psi_1^t\Sigma_{21} + \Sigma_{12}\psi_1$. However, since $\hat{T}$ and ξ are uncorrelated $\text{var}(\hat{T} - \xi^t\psi) = \text{var}(\hat{T}) + \psi_1^t I^{-1}\psi_1 = \sigma^2$. Therefore, $\text{var}(\hat{T}) = \sigma^2 - \psi_1^t I^{-1}\psi_1$. By equating these two representations for $\text{var}(\hat{T})$ we have $\Sigma_{12} = -\psi_1^t I^{-1}$. $\qquad\square$

Lemma 5.

$$\psi_1 = \int_0^1 \text{cov}\{h(W, y), W\} \dot{g}_\theta(y) \, dy$$

and

$$\psi_2 = \int_0^1 \text{cov}\{h(W, y), W\} \dot{g}_\beta(y) \, dy.$$

Proof. By assumption 2, $E\{h(W, y)\} = 0$, for $0 \leqslant y \leqslant 1$; hence by integration by parts

$$\psi_1 = \int_0^1 EW h_x(W, y)\, \dot{g}_\theta(y)\, \mathrm{d}y = \int_0^1 \int_0^\infty w\, h_x(w, y)\, \dot{g}_\theta(y) \exp(-w)\, \mathrm{d}w\, \mathrm{d}y$$

$$= \int_0^1 \int_0^\infty (w-1)\, h(w, y)\, \dot{g}_\theta(y) \exp(-w)\, \mathrm{d}w\, \mathrm{d}y = \int_0^1 \mathrm{cov}\{W, h(W, y)\}\, \dot{g}_\theta(y)\, \mathrm{d}y.$$

Similarly we can establish the identity for ψ_2.   □

Proof of theorem 4. Consider the non-degenerate statistic $\hat{T}(h_{1k})$ with $\mathrm{var}\{\hat{T}(h_k)\} = \sigma^2(h_{1k})$. A simple calculation shows that $e(h_{1k}) = e(h_{2k})$ if $h_{2k} = ah_{1k} + b$ where $a \neq 0$ and b are real numbers. By taking $a = \{\sigma^2(h_{1k})\}^{-1/2}$ and $b = 0$, we can find h_{2k} such that $e(h_{2k}) = e(h_{1k})$ with $\sigma^2(h_{2k}) = 1$. Therefore without loss of generality we may set $\sigma^2(h_k) \equiv 1$ for all $h_k( ) \in \mathscr{A}$, where the class of functions $\mathscr{A}$ is the set of all h_k satisfying assumption 1. Then for any $h_k( ) \in \mathscr{A}$, from equation (3.1) and lemma 5

$$e(h_k) = \int_0^1 \mathrm{cov}\{h(W, y), W\}\gamma^t \dot{g}_\beta(y)\, \mathrm{d}y + (A\gamma)^t \int_0^1 \mathrm{cov}\{h(W, y), W\}\, \dot{g}_\theta(y)\, \mathrm{d}y.$$

Applying the Cauchy–Schwartz inequality (see Chung (1968), p. 47) we have

$$\mathrm{cov}\{h(W, y)\, W\} \leqslant \mathrm{var}\{h(W, y)\}^{1/2}\, \mathrm{var}(W)^{1/2}$$

with equality only when $h(x, y) = K(y)x$ where $K(y)$ is a continuous function on $[0, 1]$. Thus $e(h_k)$ attains a maximum when $h(x, y) = K(y)x$, with $K(y)\dot{g}_\beta(y) \geqslant 0$ and $K(y)\dot{g}_\theta(y) \geqslant 0$ for $0 \leqslant y \leqslant 1$. Since $\mathrm{var}(W) = 1$, it is easy to verify that

$$e(h_k) = \int_0^1 K(y)\, \{\gamma^t \dot{g}_\beta(y)\, \mathrm{d}y + (A\gamma)^t \dot{g}_\theta(y)\}\, \mathrm{d}y$$

$$= \mathrm{cov}\{K(Y), \gamma^t \dot{g}_\beta(Y) + (A\gamma)^t \dot{g}_\theta(Y)\}$$

$$\leqslant [\mathrm{var}\{K(Y)\}^{1/2}[\mathrm{var}\{\gamma^t \dot{g}_\beta(Y)\} + (A\gamma)^t \dot{g}_\theta(Y)]$$

with equality if $K(Y) = \gamma^t \dot{g}_\beta(Y) + (A\gamma)^t \dot{g}_\theta(Y)$ where Y is a uniform random variable on $[0, 1]$. Therefore $e(h_k)$ is maximized if

$$h(x, y) = \{\gamma^t \dot{g}_\beta(y) + (A\gamma)^t \dot{g}_\theta(y)\}x. \qquad \qquad □$$

REFERENCES

Chung, K. L. (1968) *A Course in Probability Theory*, 2nd edn. New York: Academic Press.
Cressie, N. (1976) On the logarithms of high-order spacings. *Biometrika*, **63**, 343–355.
Csörgő, M. and Révész, P. (1980) Quantile processes and sums of weighted spacings for composite goodness-of-fit: I, II. *Carltn Math. Lect. Notes*, **27**, 1–76.
D'Agostino, R. B. and Stephens, M. A. (1986) *Goodness-of-fit Techniques*. New York: Dekker.
Del Pino, G. E. (1979) On the asymptotic distribution of k-spacings with applications to goodness-of-fit tests. *Ann. Statist.*, **7**, 1058–1065.

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

Hall, P. (1986) On powerful distributional tests based on sample spacings. *J. Multiv. Anal.*, **19**, 201–224.

Holst, L. and Rao, J. S. (1981) Asymptotic spacings theory with applications to the two sample problem. *Can. J. Statist.*, **9**, 79–89.

Jammalamadaka, S. R., Zhou, X. and Tiwari, R. C. (1989) Asymptotic efficiencies of spacings tests of goodness-of-fit. *Metrika*, **36**, 355–377.

Kuo, M. and Rao, J. S. (1981) Limit theory and efficiencies for tests based on higher order spacings. In *Statistics: Applications and New Directions: Proc. Golden Jubilee Conf. Indian Statistical Institute*, pp. 333–352. Calcutta: Statistical Publishing Society.

Moore, D. S. and Spurill, M. C. (1975) Unified large sample theory of general chi-squared statistics for tests of fit. *Ann. Statist.*, **3**, 599–616.

Pyke, R. (1965) Spacings (with discussion). *J. R. Statist. Soc.* B, **27**, 395–449.

Rao, C. R. (1973) *Linear Statistical Inference and Its Application*, 2nd edn. New York: Wiley.

Rao, J. S. and Kuo, M. (1984) Asymptotic results on the Greenwood statistic and some of its generalizations. *J. R. Statist. Soc.* B, **46**, 228–237.

Rao, J. S. and Sethuraman, J. (1975) Weak convergence of empirical distribution functions of random variables subject to perturbations and scale factors. *Ann. Statist.*, **3**, 299–313.

Serfling, R. (1980) *Approximation Theorems for Mathematical Statistics*. New York: Wiley.

Sethuraman, J. and Rao, J. S. (1970) Pitman efficiencies of tests based on spacings. In *Nonparametric Techniques in Statistical Inference*, pp. 405–416. Cambridge: Cambridge University Press.

Shapiro, S. S. and Francia, R. S. (1972) An approximate analysis of variance normality. *J. Am. Statist. Ass.*, **67**, 215–216.

Stephens, M. A. (1974) EDF statistics for goodness of fit and some comparisons. *J. Am. Statist. Ass.*, **69**, 730–737.

Vasicek, O. (1976) A test for normality based on sample entropy. *J. R. Statist. Soc.* B, **38**, 54–59.

Weiss, L. (1961) On the estimation of scale parameters. *Nav. Res. Logist. Q.*, **8**, 245–256.