The Annals of Statistics
1985, Vol. 13, No. 1, 315–330

OPTIMAL GOODNESS-OF-FIT TESTS FOR LOCATION/SCALE FAMILIES OF DISTRIBUTIONS BASED ON THE SUM OF SQUARES OF L-STATISTICS[1]

BY VINCENT N. LARICCIA AND DAVID M. MASON

University of Delaware

A class of goodness-of-fit tests based on sums of squares of L-statistics is proposed for testing a composite parametric location and/or scale null hypothesis versus a general parametric alternative. It is shown that such tests can be constructed optimally to have the same asymptotic power against sequences of local alternatives as the generalized likelihood ratio statistics [G.L.R.S.] and, in fact, under suitable regularity conditions to be asymptotically equivalent to the G.L.R.S. One advantage of the proposed test statistic over the G.L.R.S. is that only an estimate of the scale parameter is needed in the computation of the statistic. No other parameter estimates are required. Also, an example of the practical implementation of the proposed hypothesis testing procedure is given.

1. Introduction. Let $X_1, \cdots, X_n$ be independent random variables with common distribution function $F((x - \mu)/\sigma, \vec{\theta})$, depending on location and scale parameters $-\infty < \mu < \infty$ and $0 < \sigma < \infty$ and a vector of additional parameters $\vec{\theta} = (\theta_1, \cdots, \theta_k) \in \Theta$, where Θ is an open subset of $\mathbb{R}^k$ which contains the zero vector $\vec{0}$. Write $\Omega = (-\infty, \infty) \times (0, \infty) \times \Theta$.

Suppose we are interested in testing the composite null hypothesis

$$\mathrm{H}_0\colon (\mu, \sigma, \vec{\theta}) \in (-\infty, \infty) \times (0, \infty) \times \{\vec{0}\} = \Omega_0$$

versus the composite alternative hypothesis

$$\mathrm{H}_a\colon (\mu, \sigma, \vec{\theta}) \in (-\infty, \infty) \times (0, \infty) \times \{\Theta - \{\vec{0}\}\} = \Omega_a.$$

Assume that $F((x - \mu)/\sigma, \vec{\theta})$ has a density function $\sigma^{-1}f((x - \mu)/\sigma, \vec{\theta})$ and set

$$L_n(\mu, \sigma, \vec{\theta}) = \sigma^{-n} \prod_{i=1}^{n} f((X_i - \mu)/\sigma, \vec{\theta}).$$

Let

$$L_{0,n} = \sup\{L_n(\mu, \sigma, \vec{\theta})\colon (\mu, \sigma, \vec{\theta}) \in \Omega_0\}$$

and

$$L_{a,n} = \sup\{L_n(\mu, \sigma, \vec{\theta})\colon (\mu, \sigma, \vec{\theta}) \in \Omega\}.$$

The classical test procedure of H_0 versus H_a is based on the generalized likelihood

Received April 1983; revised April 1984.

[1] This research was supported by University of Delaware Research Foundation Grant #8325530015.

AMS 1980 *subject classifications*. Primary 62F03, 62F05; secondary 62G30.

Key words and phrases. Goodness-of-fit tests, L-statistics, generalized likelihood ratio, power, continuous time regression.

ratio statistic [G.L.R.S.]

$$\lambda_n = L_{0,n}/L_{a,n},$$

where H_0 is rejected at significance level α if $-2 \ln \lambda_n$ is greater than its upper α-critical value. It is well known that under suitable regularity conditions, when H_0 is true,

$$-2 \ln \lambda_n \to_d \chi_k^2,$$

where χ_k^2 denotes a Chi-squared random variable with k degrees of freedom. (For this plus certain optimality properties and distribution results under local alternatives of the G.L.R.S., refer to Wald, 1943; Davidson and Lever, 1970; Hayakawa, 1975; and Dzhaparidze, 1977.) Two competitors to the G.L.R.S. are the Wald test and the Rao scores test. (Refer to Rao, 1973, pages 418–420.) Under suitable regularity conditions, these two statistics have the same asymptotic null distribution and asymptotic power against "local alternatives" as the G.L.R.S. (see Silvey, 1959).

For each $n \geq 1$ let $X_{1,n} \leq \cdots \leq X_{n,n}$ denote the order statistics based on $X_1, \cdots, X_n$. Let $\vec{w} = (w_1, \cdots, w_k)$ denote a vector of k real valued measurable functions defined on (0, 1). We will consider test procedures of H_0 versus H_a based on sums of squares of L-statistics of the form

$$T_n(\vec{w}) = \sum_{i=1}^k C_n^2(w_i)/\hat{\sigma}_n^2,$$

where for each $1 \leq i \leq k$, $C_n(w_i)$ is the L-statistic

$$C_n(w_i) = n^{-1/2} \sum_{j=1}^n w_i(j/(n+1))X_{j,n}$$

and $\hat{\sigma}_n$ is a consistent estimator of σ. Here, H_0 is rejected at significance level α if $T_n(\vec{w})$ is greater than its upper α-critical value. (We consider this version of the statistic here for the sake of mathematical convenience. In actual practice it must be adjusted slightly for finite samples. See Remark 7 below and the example in Section 4.)

We will show that under appropriate regularity conditions $\vec{w}$ and $\hat{\sigma}_n$ can be chosen optimally so that $T_n(\vec{w})$ has both the same asymptotic null distribution and the same asymptotic power against sequences of "local alternatives" as $-2 \ln \lambda_n$. We also discuss when our statistic is asymptotically equivalent to the G.L.R.S. Refer to Remark 5 below.

Typically, the G.L.R.S. is calculated by the formula,

$$\lambda_n = L_n(\hat{\mu}_0, \hat{\sigma}_0, \vec{0})/L_n(\hat{\mu}_a, \hat{\sigma}_a, \hat{\vec{\theta}}_a),$$

where $\hat{\mu}_0$ and $\hat{\sigma}_0$ are the maximum likelihood estimators [M.L.E.s] of μ and σ computed assuming that H_0 is true and $\hat{\mu}_a$, $\hat{\sigma}_a$ and $\hat{\vec{\theta}}_a$ are the M.L.E.s of μ, σ, and $\vec{\theta}$ determined subject to the constraint that $(\mu, \sigma, \vec{\theta}) \in \Omega$. In the computation of the Wald test only $\hat{\mu}_a$, $\hat{\sigma}_a$, and $\hat{\vec{\theta}}_a$ are needed, whereas the Rao scores test only requires $\hat{\mu}_0$ and $\hat{\sigma}_0$. One advantage of our proposed test statistic $T_n(\vec{w})$ over these three statistics is that the only parameter estimate required is a suitably consistent estimate of the unknown scale parameter σ.

For the special case of testing for normality against a particular class of alternatives, La Brecque (1977) has proposed a test statistic which turns out to be a special case of our class of statistics. While no asymptotic distribution theory or optimality results are presented in his paper, his test is shown to have good finite sample properties. Another example of the practical implementation of our hypothesis testing procedure is given in Section 4.

2. Preliminaries and motivation. First, we must introduce some notation and assumptions in order to properly motivate our optimal choice of $T_n(\vec{w})$.

For each $\vec{\theta} \in \Theta$ and $u \in (0, 1)$ let $Q(u, \vec{\theta}) = \inf\{x: F(x, \vec{\theta}) \geq u\}$ denote the inverse or quantile function of $F(x, \vec{\theta})$. Notice that for arbitrary μ, σ, and $\vec{\theta}$ the quantile function of $F((x - \mu)/\sigma, \vec{\theta})$ is given by

$$Q(\cdot, \mu, \sigma, \vec{\theta}) = \mu + \sigma Q(\cdot, \vec{\theta}).$$

Let $N \subset \Theta$ denote an open neighborhood of the zero vector $\vec{0}$.

(A) Assume that for each $u \in (0, 1)$

$$D_i(u, \vec{\theta}) = (\partial/\partial\theta_i)Q(u, \vec{\theta}) \quad \text{for} \quad i = 1, \cdots, k;$$

and

$$D_{i,j}(u, \vec{\theta}) = (\partial^2/\partial\theta_i\partial\theta_j)Q(u, \vec{\theta}) \quad \text{for} \quad 1 \leq i, \quad j \leq k$$

exist and are continuous for $\vec{\theta} \in N$.

Write $Q_0(\cdot) = Q(\cdot, \vec{0})$, $D_i(\cdot) = D_i(\cdot, \vec{0})$ for $i = 1, \cdots, k$, $D_{-1} = 1$ and $D_0 = Q_0$. Let F_0 denote the distribution function corresponding to Q_0.

For each $n \geq 1$ let $U_1, \cdots, U_n$ be independent uniform (0, 1) random variables and let $U_{1,n} \leq \cdots \leq U_{n,n}$ denote their order statistics. It is well known that if $X_1, \cdots, X_n$ are independent random variables with common quantile function $Q(\cdot, \mu, \sigma, \vec{\theta})$ then $X_{1,n}, \cdots, X_{n,n}$ have the same joint distribution as $Q(U_{1,n}, \mu, \sigma, \vec{\theta}), \cdots, Q(U_{n,n}, \mu, \sigma, \vec{\theta})$. *To simplify the presentation that follows, we will from now on use the latter distributionally equivalent version of the order statistics* $X_{i,n}$ *for* $i = 1, \cdots, n$.

For any $\vec{w}$ and choice of $-\infty < \mu < \infty$ and $0 < \sigma < \infty$ let

$$\vec{L}_{n,\mu,\sigma}(\vec{w}) = (L_{n,\mu,\sigma}(w_1), \cdots, L_{n,\mu,\sigma}(w_k))$$

denote the vector of linear combinations of order statistics, where $1 \leq i \leq k$

$$L_{n,\mu,\sigma}(w_i) = n^{-1/2} \textstyle\sum_{j=1}^{n} w_i(j/(n+1))(\mu + \sigma Q_0(U_{j,n})).$$

For any two measurable functions h_1 and h_2 defined on (0, 1), let

$$\langle h_1, h_2 \rangle = \int_0^1 h_1(u)h_2(u)\, du,$$

and

$$\langle\langle h_1, h_2 \rangle\rangle = \int_0^1 \int_0^1 (u \wedge v - uv)h_1(u)h_2(v)\, dQ_0(u)\, dQ_0(v).$$

We will use the convention that whenever we say that $\langle h_1, h_2 \rangle$, respectively

$\langle\langle h_1, h_2\rangle\rangle$, is finite, we will also mean that $\langle |h_1|, |h_2|\rangle$, respectively $\langle\langle |h_1|, |h_2|\rangle\rangle$, is finite.

Let $\mathscr{W}$ denote the class of $\vec{w}$ such that

(I) $$\langle w_i, 1\rangle = \langle w_i, Q_0\rangle = 0 \quad \text{for} \quad i = 1, \cdots, k;$$

(II) $$\langle\langle w_i, w_j\rangle\rangle = \delta_{i,j},$$

where $\delta_{i,j}$ equals one if $i = j$ and zero otherwise;

(III) for each $-\infty < \mu < \infty$ and $0 < \sigma < \infty$

$$\vec{L}_{n,\mu,\sigma}(\vec{w}) \to_d N(\vec{0}, \sigma^2 I_k),$$

where I_k denotes the $k \times k$ identity matrix;

(IV) for each $1 \le \alpha, \beta \le k$

$$n^{-1} \textstyle\sum_{j=1}^n w_\alpha(j/(n+1))D_\beta(U_{j,n}) \to_p \langle w_\alpha, D_\beta\rangle \quad \text{(finite)};$$

and

(V) for each $1 \le i \le k$

$$n^{-1} \textstyle\sum_{j=1}^n |w_i(j/(n+1))| M(U_{j,n}) = o_p(n^{1/2}),$$

where for each $u \in (0, 1)$

$$M(u) = \max_{1\le i,j\le k} \sup_{\vec{\theta}\in N} |D_{i,j}(u, \vec{\theta})|.$$

REMARK 1. Conditions on $\vec{w}$ and Q_0 which imply (III) can be found in Shorack (1972), Stigler (1974) and Mason (1981) and the references therein; while conditions on $\vec{w}$ and the D_i's which imply (IV) are given in Wellner (1977), van Zwet (1980) and Mason (1982).

For each $n \ge 1$ let σ_n denote a measurable function from $\mathbb{R}^n$ to $\mathbb{R}$. We will require the following definitions:

DEFINITION 1. Let $-\infty < \mu < \infty$, $0 < \sigma < \infty$, and $\vec{\beta} \in \mathbb{R}^k - \{\vec{0}\}$ be fixed. Suppose for each $n \ge 1$, it is assumed that $X_1^{(n)}, \cdots, X_n^{(n)}$ are independent random variables with common quantile function

$$Q(\cdot, \mu, \sigma, \beta_1/\sqrt{n}, \cdots, \beta_k/\sqrt{n}).$$

Any such sequence is called a *sequence of local alternatives.*

DEFINITION 2. σ_n will be called a *consistent estimator of scale under local alternatives* if for any choice of $-\infty < \mu < \infty$, $0 < \sigma < \infty$ and constants $\beta_1, \cdots, \beta_k$, $\hat{\sigma}_n$ converges in probability to σ, where for each $n \ge 1$, $\hat{\sigma}_n$ denotes the function σ_n evaluated at

$$Q(U_i, \mu, \sigma, \beta_1/\sqrt{n}, \cdots, \beta_k/\sqrt{n}) \quad \text{for} \quad i = 1, \cdots, n.$$

REMARK 2. Under assumption (A) such estimators of scale typically exist.

The reader can easily construct examples of M-estimators, L-estimators, and maximum likelihood estimators that are consistent estimators under local alternatives.

Choose any fixed constants $\beta_1, \cdots, \beta_k$, $-\infty < \mu < \infty$, $0 < \sigma < \infty$ and $\vec{w} \in \mathscr{W}$. Set

$$\vec{A}_{n,\mu,\sigma}(\vec{w}) = (A_{n,\mu,\sigma}(w_1), \cdots, A_{n,\mu,\sigma}(w_k)),$$

and

$$\vec{B}_{n,\mu,\sigma}(\vec{w}) = (B_{n,\mu,\sigma}(w_1), \cdots, B_{n,\mu,\sigma}(w_k)),$$

where for each $1 \le i \le k$

$$A_{n,\mu,\sigma}(w_i) = n^{-1/2} \textstyle\sum_{j=1}^{n} w_i\left(\frac{j}{n+1}\right)\left\{\mu + \sigma Q\left(U_{j,n}, \frac{\beta_1}{\sqrt{n}}, \cdots, \frac{\beta_k}{\sqrt{n}}\right)\right\},$$

and

$$B_{n,\mu,\sigma}(w_i) = n^{-1/2} \textstyle\sum_{j=1}^{n} w_i\left(\frac{j}{n+1}\right)\{\mu + \sigma Q_0(U_{j,n}) + \sigma n^{-1/2} \sum_{\nu=1}^{k} \beta_\nu D_\nu(U_{j,n})\}.$$

Also, let $\vec{m}$ denote the $1 \times k$ vector with ith component m_i equal to

$$\textstyle\sum_{j=1}^{k} \beta_j \langle w_i, D_j\rangle.$$

The following lemma will be essential to our discussion later on.

LEMMA 1. *Choose any fixed constants* $\beta_1, \cdots, \beta_k$, $-\infty < \mu < \infty$, *and* $0 < \sigma < \infty$. *Whenever assumption* (A) *holds with* $\vec{w} \in \mathscr{W}$, *then*

$$\vec{A}_{n,\mu,\sigma}(\vec{w}) \to_d N(\sigma\vec{m}, \sigma^2 I_k).$$

PROOF. It is easy to see that conditions (III) and (IV) imply that

$$\vec{B}_{n,\mu,\sigma}(\vec{w}) \to_d N(\sigma\vec{m}, \sigma^2 I_k).$$

Therefore it is enough to show that for each $1 \le i \le k$

$$| A_{n,\mu,\sigma}(w_i) - B_{n,\mu,\sigma}(w_i) | \to_p 0.$$

Assumption (A), in combination with a two-term Taylor expansion and an elementary bound, implies that the above expression is less than or equal to

$$\sigma n^{-1/2} \textstyle\sum_{j=1}^{n} n^{-1} | w_i(j/(n+1)) | M(U_{j,n})(\sum_{\nu=1}^{k} | \beta_\nu |)^2,$$

which by (V) converges in probability to zero. □

Let σ_n be a consistent estimator of scale under local alternatives. Choose any $\vec{w} \in \mathscr{W}$. Observe that under H_0, $X_{i,n} = \mu + \sigma Q_0(U_{i,n})$ for $i = 1, \cdots, n$ and some $-\infty < \mu < \infty$ and $0 < \sigma < \infty$. Hence, by condition (III) and the fact that

$$C_n(w_i) = L_{n,\mu,\sigma}(w_i) \quad \text{for} \quad i = 1, \cdots, k,$$

we have that under H_0

$$T_n(\vec{w}) \rightarrow_d \chi_k^2.$$

Now consider any sequence of local alternatives determined by fixed $-\infty < \mu < \infty$, $0 < \sigma < \infty$ and $\vec{\beta} \in \mathbb{R}^k - \{\vec{0}\}$. Lemma 1 implies that

$$T_n(\vec{w}) \rightarrow_d \sum_{i=1}^k (Z_i + m_i)^2$$

where $Z_1, \cdots, Z_k$ are independent standard normal random variables; or in other words, the asymptotic distribution of $T_n(\vec{w})$ under any sequence of local alternatives as given by Definition 1 is that of a noncentral Chi-squared random variable with k degrees of freedom and noncentrality parameter $\Delta(\vec{w}, \vec{\beta})$ dependent on $\vec{\beta}$ and $\vec{w}$ given by

$$\Delta(\vec{w}, \vec{\beta}) = \sum_{i=1}^k \left(\sum_{j=1}^k \beta_j \langle w_i, D_j \rangle\right)^2.$$

Set the $k \times k$ matrix $R(\vec{w})$ equal to

$$\| \langle w_i, D_j \rangle \|_{i,j=1,\cdots,k}.$$

Observe that $\Delta(\vec{w}, \vec{\beta})$ can be written equivalently as

$$\Delta(\vec{w}, \vec{\beta}) = \vec{\beta} R'(\vec{w}) R(\vec{w}) \vec{\beta}',$$

where $'$ denotes transpose.

Notice that the asymptotic power of $T_n(\vec{w})$ at any significance level α against any sequence of local alternatives is a strictly increasing function of $\Delta(\vec{w}, \vec{\beta})$. Also $T_n(\vec{w})$ will have asymptotic power greater than level α for all $\vec{\beta} \in \mathbb{R}^k - \{\vec{0}\}$ if and only if $\Delta(\vec{w}, \vec{\beta}) > 0$ for all $\vec{\beta} \in \mathbb{R}^k - \{\vec{0}\}$, which in turn happens if and only if

(VI) $\qquad R(\vec{w})$ is nonsingular.

Let $\mathscr{L}$ denote the subclass of $\vec{w} \in \mathscr{W}$, which in addition to (I) through (V), also satisfy (VI), and let $\mathscr{S} = \{T_n(\vec{w}): \vec{w} \in \mathscr{L}\}$. In the next section, we will describe how to choose a fixed vector of weight functions $\vec{w}_0 \in \mathscr{L}$ such that

$$\Delta(\vec{w}_0, \vec{\beta}) \geq \Delta(\vec{w}, \vec{\beta})$$

for every $\vec{w} \in \mathscr{L}$ and choice of $\vec{\beta}$. $T_n(\vec{w}_0)$ will then have optimal asymptotic power against local alternatives among all statistics in $\mathscr{S}$. Moreover, $T_n(\vec{w}_0)$ will be shown to have the same asymptotic power against local alternatives as the G.L.R.S.

3. The optimal choice of $\vec{w}$. In addition to (A), we require the following assumptions:

(B) Assume that F_0 has a density function f_0, which is strictly positive on the support of F_0.

For each $i = -1, 0, 1, \cdots, k$, let

$$g_i(\cdot) = D_i(\cdot) f_0(Q_0(\cdot)).$$

Assume that for each $i = -1, 0, 1, \cdots, k$

(C) g_i is absolutely continuous inside (0, 1) with derivative g_i', which is also absolutely continuous inside (0, 1) with derivative g_i'' (c.f. page 328 of Hájek, 1968);

(D) g_i' is square integrable on (0, 1) such that $\lim_{u\downarrow 0} u g_i'(u) = 0$ and $\lim_{u\uparrow 1}(1-u)g_i'(u) = 0$; and

(E) $g_i(0+) = g_i(1-) = 0$.

Also, assume that for each $-1 \le i, j \le k$

(F) $$\lim_{u\downarrow 0} g_j'(u)g_i(u) = \lim_{u\uparrow 1} g_j'(u)g_i(u) = 0.$$

Let $\mathscr{I}$ denote the $(k+2)\times(k+2)$ "information matrix"

$$\| \langle g_i', g_j' \rangle \|_{i,j=-1,\cdots,k}.$$

(G) Assume that $\mathscr{I}$ is nonsingular.

Finally, we assume that for each $-1 \le i, j \le k$

(H) $$\langle\langle \, | g_i''(\cdot) | f_0(Q_0(\cdot)), | g_j''(\cdot) | f_0(Q_0(\cdot)) \rangle\rangle < \infty.$$

OBSERVATION 1. Assumptions (B) through (F) and (H) allow us to apply Fubini's theorem to show that for each $-1 \le i, j \le k$

$$\langle\langle g_i''(\cdot)f_0(Q_0(\cdot)), g_j''(\cdot)f_0(Q_0(\cdot))\rangle\rangle = -\langle g_i, g_j''\rangle,$$

which, by applying integration by parts equals $\langle g_i', g_j' \rangle$ (finite).

We will now construct the optimal $\vec{w}$. Let A_k denote the $(k+2)\times k$ matrix

$$A_k = \left\| \begin{matrix} 0 \cdots 0 \\ 0 \cdots 0 \\ I_k \end{matrix} \right\|,$$

and let B_k denote the inverse matrix of $A_k'\mathscr{I}^{-1}A_k$. Since B_k is symmetric and positive definite, we can write $B_k = C_k'C_k$ for an appropriate nonsingular $k\times k$ matrix C_k. Let $\vec{v}$ denote the $(k+2)\times 1$ vector with ith component equal to $-g_{i-2}''(\cdot)f_0(Q_0(\cdot))$. Set

$$\vec{w}_0' = C_k A_k' \mathscr{I}^{-1}\vec{v}, \quad \text{and} \quad \vec{w}_0 = (w_{0,1}, \cdots, w_{0,k}).$$

Notice that by Observation 1,

(1) $$\| \langle v_i, D_j \rangle \|_{i,j=-1,\cdots,k} = \mathscr{I}.$$

Hence the $k\times(k+2)$ matrix

$$\| \langle w_{0,i}, D_j \rangle \|_{i=1,\cdots,k;\, j=-1,\cdots,k} = \left\| \begin{matrix} 0 & 0 & \\ \vdots & \vdots & C_k \\ 0 & 0 & \end{matrix} \right\|.$$

Thus, since $D_{-1} = 1$ and $D_0 = Q_0$

(2) $$\langle w_{0,i}, 1\rangle = \langle w_{0,i}, Q_0\rangle = 0 \quad \text{for} \quad i = 1, \cdots, k.$$

Observe that $R(\vec{w}_0) = C_k$, hence we have for any $\vec{\beta} \in \mathbb{R}^k$

(3) $$\Delta(\vec{w}_0, \vec{\beta}) = \textstyle\sum_{i=1}^k (\sum_{j=1}^k \beta_j \langle w_{0,i}, D_j\rangle)^2 = \vec{\beta} C_k' C_k \vec{\beta}' = \vec{\beta} B_k \vec{\beta}'.$$

Also, it is easily checked using Observation 1 that $\langle\langle w_{0,i}, w_{0,j}\rangle\rangle = \delta_{i,j}$. Therefore $\vec{w}_0$ satisfies conditions (I), (II) and (VI).

The following theorem shows that, in fact, $\vec{w}_0$ is the optimal choice of $\vec{w} \in \mathscr{L}$, whenever $\vec{w}_0 \in \mathscr{L}$.

THEOREM. *Whenever conditions* (A) *through* (H) *hold and* $\vec{w}_0 \in \mathscr{L}$ *then for any* $\vec{\beta} \in \mathbb{R}^k$

$$\Delta(\vec{w}_0, \vec{\beta}) = \vec{\beta} B_k \vec{\beta}' = \max\{\Delta(\vec{w}, \vec{\beta}) : \vec{w} \in \mathscr{L}\}.$$

PROOF. By (3), it is enough to verify that

(4) $$\vec{\beta} B_k \vec{\beta}' = \max\{\Delta(\vec{w}, \vec{\beta}) : \vec{w} \in \mathscr{L}\}.$$

Let $\mathscr{L}^*$ denote the class of all $\vec{w}$ such that conditions (I) and (VI) hold, each $\langle\langle w_i, w_j\rangle\rangle$ for $1 \le i, j \le k$ is finite, and the $k \times k$ matrix

$$V(\vec{w}) = \| \langle\langle w_i, w_j\rangle\rangle \|_{i,j=1,\cdots,k}$$

is nonsingular. (Recall the definition of $\langle\langle \cdot, \cdot\rangle\rangle$ given above.)

Since $\mathscr{L} \subset \mathscr{L}^*$, to show (4) it is sufficient to establish that for any $\vec{\beta} \in \mathbb{R}^k$

(5) $$\vec{\beta} B_k \vec{\beta}' = \max\{\vec{\beta} R'(\vec{w}) V^{-1}(\vec{w}) R(\vec{w}) \vec{\beta}' : \vec{w} \in \mathscr{L}^*\}.$$

Let $\mathscr{L}_0^*$ denote the subclass of $\mathscr{L}^*$ such that

(6) $$R(\vec{w}) = I_k.$$

An elementary argument shows that the right-hand side of expression (5) equals

$$\max\{\vec{\beta} V^{-1}(\vec{w}) \vec{\beta}' : \vec{w} \in \mathscr{L}_0^*\}.$$

Notice that for any $\vec{\beta} \in \mathbb{R}^k$ and $\vec{w} \in \mathscr{L}_0^*$

(7) $$\sup_{\vec{x} \in \mathbb{R}^k} (\vec{\beta} \vec{x}')^2 / (\vec{x} V(\vec{w}) \vec{x}') = \vec{\beta} V^{-1}(\vec{w}) \vec{\beta}'.$$

(See page 60 of Rao, 1973.) It is convenient at this point to introduce the following continuous time regression problem:

For any $-\infty < \mu < \infty$, $0 < \sigma < \infty$ and $\vec{\beta} \in \mathbb{R}^k$ consider the Gaussian process defined on (0, 1) by

$$Y(\cdot, \mu, \sigma, \vec{\beta}) = \mu + \sigma Q_0(\cdot) + \textstyle\sum_{i=1}^k \beta_i D_i(\cdot) + B(\cdot)/f_0(Q_0(\cdot)),$$

where B is a Brownian bridge defined on (0, 1). Unless there is a possibility of confusion, we will write

$$Y(\cdot) = Y(\cdot, \mu, \sigma, \vec{\beta}).$$

For any $\vec{x} \in \mathbb{R}^k$, $(\mu, \sigma, \vec{\beta})$ as above, and $\vec{w} \in \mathscr{L}_0^*$ set

$$b(\vec{x}, \vec{w}, Y) = \sum_{i=1}^k x_i\langle w_i, Y\rangle.$$

We will use the convention that the random variables $\langle w_i, Y\rangle$ are to be interpreted as the limit in expected mean square as $\varepsilon \downarrow 0$ of the normal random variables

$$\int_\varepsilon^{1-\varepsilon} w_i(u)Y(u, \mu, \sigma, \vec{\beta})\, du.$$

Note that the conditions of C_0^* imply that for each $1 \le i \le k$

$$E\langle w_i, Y\rangle = \lim_{\varepsilon\downarrow 0} E\left(\int_\varepsilon^{1-\varepsilon} w_i(u)Y(u)\, du\right) = \beta_i.$$

Hence $b(\vec{x}, \vec{w}, Y)$ is an unbiased estimator of

$$\phi(\vec{x}, \vec{\beta}) = \sum_{i=1}^k x_i\beta_i.$$

Keeping in mind our above convention, a straightforward computation shows that

$$\operatorname{Var}(b(\vec{x}, \vec{w}, Y)) = \lim_{\varepsilon\downarrow 0} \operatorname{Var}\left(\sum_{i=1}^k x_i \int_\varepsilon^{1-\varepsilon} w_i(s)B(s)\, dQ_0(s)\right) = \vec{x}V(\vec{w})\vec{x}'.$$

Set

$$\tilde{\vec{w}}_0' = A_k'\mathscr{I}^{-1}\vec{v}.$$

Observe that by (1)

$$\| \langle \tilde{w}_{0,i}, D_j\rangle \|_{i=1,\cdots,k;j=-1,\cdots,k} = A_k'. \tag{8}$$

It is easy to see that $\tilde{\vec{w}}_0 \in \mathscr{L}_0^*$, with

$$V(\tilde{\vec{w}}_0) = A_k'\mathscr{I}^{-1}A_k.$$

We will now establish that for any $\vec{x} \in \mathbb{R}^k$ and $\vec{w} \in \mathscr{L}_0^*$

$$\vec{x}V(\tilde{\vec{w}}_0)\vec{x}' \le \vec{x}V(\vec{w})\vec{x}',$$

that is $b(\vec{x}, \tilde{\vec{w}}_0, Y)$ is the best linear unbiased estimator of $\phi(\vec{x}, \vec{\beta})$, which by (7), will imply (4).

A standard computation shows that

$$\vec{x}V(\vec{w})\vec{x}' = E(b(\vec{x}, \vec{w}, Y) - b(\vec{x}, \tilde{\vec{w}}_0, Y))^2 + 2\operatorname{cov}((b(\vec{x}, \vec{w}, Y) - b(\vec{x}, \tilde{\vec{w}}_0, Y)), b(\vec{x}, \tilde{\vec{w}}_0, Y)) + \vec{x}V(\tilde{\vec{w}}_0)\vec{x}'.$$

To complete the proof, we need only show that the above covariance term equals zero, or in other words that

$$\sum_{i=1}^k \sum_{j=1}^k x_i\langle\langle w_i - \tilde{w}_{0,i}, \tilde{w}_{0,j}\rangle\rangle x_j = 0. \tag{9}$$

Since each $\tilde{w}_{0,j}$ can be written as

$$\tilde{w}_{0,j} = \textstyle\sum_{s=1}^{k+2} \alpha_{j,s} v_s$$

for appropriate $\alpha_{j,s}$'s, verifying (9) is equivalent to showing that for each $1 \le i \le k$ and $1 \le j \le k + 2$

$$\langle\langle w_i - \tilde{w}_{0,i}, v_j \rangle\rangle = 0. \tag{10}$$

The left side of (10) equals

$$-\int_0^1 (w_i(u) - \tilde{w}_{0,i}(u)) \left[\int_0^1 (u \wedge v - uv) g''_{j-2}(v)\, dv \right] dQ_0(u),$$

which by integration of the inner integral equals

$$\langle w_i - \tilde{w}_{0,i}, D_{j-2} \rangle = \langle w_i, D_{j-2} \rangle - \langle \tilde{w}_{0,i}, D_{j-2} \rangle.$$

This last expression equals zero by condition (I), (6) and (8). □

REMARK 3. The optimal choice of the $\vec{w}$ was motivated by the continuous time regression ideas of Parzen (1961a, b). In fact, in the second half of the proof of our theorem, we are essentially proving a special case of Theorem 7A of Parzen (1961b). However, to use his theorem directly would have entailed introducing more concepts and notation, which would have taken more space than the present proof requires. Another method of motivating $\vec{w}_0$ is outlined in Remark 5.

REMARK 4. For any $\vec{\theta} \in \Theta$ and $1 \le i \le k$ set

$$g_{i,\vec{\theta}}(\cdot) = D_i(\cdot, \vec{\theta}) f(Q(\cdot, \vec{\theta}), \vec{\theta}),$$

and let

$$g_{0,\vec{\theta}}(\cdot) = Q(\cdot, \vec{\theta}) f(Q(\cdot, \vec{\theta}), \vec{\theta}),$$

and

$$g_{-1,\vec{\theta}}(\cdot) = f(Q(\cdot, \vec{\theta}), \vec{\theta}).$$

The information matrix of $\sigma^{-1} f((x - \mu)/\sigma, \vec{\theta})$ is given by

$$\mathscr{I}(\mu, \sigma, \vec{\theta}) = \left\| \begin{matrix} \sigma^{-2} \mathscr{I}_{1,1}(\vec{\theta}) & \sigma^{-1} \mathscr{I}_{1,2}(\vec{\theta}) \\ \sigma^{-1} \mathscr{I}'_{1,2}(\vec{\theta}) & \mathscr{I}_{2,2}(\vec{\theta}) \end{matrix} \right\|,$$

where

$$\mathscr{I}_{1,1}(\vec{\theta}) = \| \langle g'_{i,\vec{\theta}}, g'_{j,\vec{\theta}} \rangle \|_{i,j=-1,0}$$

$$\mathscr{I}_{1,2}(\vec{\theta}) = \| \langle g'_{i,\vec{\theta}}, g'_{j,\vec{\theta}} \rangle \|_{i=-1,0;j=1,\cdots,k}$$

and

$$\mathscr{I}_{2,2}(\vec{\theta}) = \| \langle g'_{i,\vec{\theta}}, g'_{j,\vec{\theta}} \rangle \|_{i,j=1,\cdots,k}$$

$\mathscr{I}^{-1}(\mu, \sigma, \vec{\theta})$ may be written as

$$\left\| \begin{matrix} \sigma^2 \mathscr{I}^{1,1}(\vec{\theta}) & \sigma \mathscr{I}^{1,2}(\vec{\theta}) \\ \sigma \mathscr{I}^{1,2'}(\vec{\theta}) & \mathscr{I}^{2,2}(\vec{\theta}) \end{matrix} \right\|,$$

where in particular

$$\mathscr{I}^{2,2}(\vec{\theta}) = (\mathscr{I}_{2,2}(\vec{\theta}) - \mathscr{I}'_{1,2}(\vec{\theta})\mathscr{I}_{1,1}^{-1}(\vec{\theta})\mathscr{I}_{1,2}(\vec{\theta}))^{-1}. \tag{11}$$

Note that

$$\mathscr{I}(0, 1, \vec{0}) = \mathscr{I},$$

and

$$\mathscr{I}^{-1}(0, 1, \vec{0}) = \mathscr{I}^{-1}, \tag{12}$$

where $\mathscr{I}$ is defined as in (G).

Modulo regularity conditions, using the results of Wald (1943) pages 481 to 482, routine arguments show that under any sequence of local alternatives depending on fixed $-\infty < \mu < \infty$, $0 < \sigma < \infty$ and $\vec{\beta} \in \mathbb{R}^k$ that $-2 \ln \lambda_n$ converges in distribution to a noncentral Chi-squared random variable with k degrees of freedom and noncentrality parameter

$$\Delta(\vec{\beta}) = \vec{\beta}(A'_k\mathscr{I}^{-1}(\mu, \sigma, \vec{0})A_k)^{-1}\vec{\beta}'.$$

However, by (11) and (12)

$$A'_k\mathscr{I}^{-1}(\mu, \sigma, \vec{0})A_k = A'_k\mathscr{I}^{-1}A_k.$$

Hence

$$\Delta(\vec{\beta}) = \Delta(\vec{w}_0, \vec{\beta}).$$

This says that subject to regularity conditions that our proposed test procedure and the G.L.R.S. have the same asymptotic distribution under H_0 and under sequences of local alternatives, and hence have the same asymptotic power against local alternatives.

In the following remark, we discuss when the G.L.R.S. and $T_n(\vec{w}_0)$ are asymptotically equivalent in a stronger sense.

REMARK 5. Under certain regularity conditions, our test statistic is asymptotically equivalent to the G.L.R.S. in the sense described in Dzhaparidze (1977). See especially page 110 of this paper. We shall briefly outline the details of showing this.

Set

$$p_1(\mu, \sigma, x) = (\partial/\partial\mu) \log(\sigma^{-1}f((x - \mu)/\sigma, \vec{\theta}))\,|_{(\mu,\sigma,\vec{0})},$$

$$p_2(\mu, \sigma, x) = (\partial/\partial\sigma) \log(\sigma^{-1}f((x - \mu)/\sigma, \theta))\,|_{(\mu,\sigma,\vec{0})},$$

and for $i = 3, \cdots, k + 2$ set

$$p_i(\mu, \sigma, x) = (\partial/\partial\theta_{i-2}) \log(\sigma^{-1}f((x - \mu)/\sigma, \vec{\theta}))\,|_{(\mu,\sigma,\vec{0})}.$$

Let $\vec{P}_n(\mu, \sigma)$ denote the $1 \times (k + 2)$ vector with ith component equal to

$$P_{i,n}(\mu, \sigma) = n^{-1/2} \textstyle\sum_{j=1}^{n} p_i(\mu, \sigma, X_j),$$

and $\vec{P}_n^{(1)}(\mu, \sigma)$ denote the 1×2 vector with ith component equal to $P_{i,n}(\mu, \sigma)$ for

$i = 1, 2$. Also let $\vec{V}_n$ denote the $1 \times (k+2)$ vector with ith component equal to

$$V_{i,n} = n^{-1/2} \sum_{j=1}^{n} v_i(j/(n+1))X_{j,n},$$

where the v_i functions are as above; and $\vec{V}_n^{(1)}$ denotes the 1×2 vector with ith component equal to $V_{i,n}$ for $i = 1, 2$.

A little algebra shows that $T_n(\vec{w}_0)$ equals

$$\sigma_n^{-2}\{\vec{V}_n \mathscr{I}^{-1}(0, 1, \vec{0})\vec{V}_n' - \vec{V}_n^{(1)}\mathscr{I}_{11}^{-1}(\vec{0})\vec{V}_n^{(1)\prime}\}.$$

Now assume that $X_1, \cdots, X_n$ are i.i.d. $F_0((x-\mu)/\sigma)$. Typically, the conditions on the v_i functions are such that for each $1 \le i \le k+2$ we can write

$$V_{i,n} = -n^{-1/2} \sum_{j=1}^{n} \int_{-\infty}^{\infty} v_i\left(F_0\left(\frac{x-\mu}{\sigma}\right)\right)\left\{I(X_j \le x) - F_0\left(\frac{x-\mu}{\sigma}\right)\right\} dx + o_p(1),$$

(refer for instance to Govindarajulu and Mason (1983)) which after integration by parts equals $\sigma^2 P_{i,n}(\mu, \sigma)$ when $i = 1, 2$ and $\sigma P_{i,n}(\mu, \sigma)$ when $3 \le i \le k+2$. The above representation for $V_{i,n}$ in combination with the assumption that σ_n is a consistent estimator of σ yields after some manipulation that $T_n(\vec{w}_0)$ equals

$$\vec{P}_n(\mu, \sigma)\mathscr{I}^{-1}(\mu, \sigma, \vec{0})\vec{P}_n'(\mu, \sigma) - \sigma^{-2}\vec{P}_n^{(1)}(\mu, \sigma)\mathscr{I}_{11}^{-1}(\vec{0})P_n^{(1)\prime}(\mu, \sigma) + o_p(1).$$

If in addition we assume that the conditions of Theorem 1 of Dzhaparidze (1977) hold (also see the discussion in Section 3 of his paper), when applied to our particular hypothesis testing situation, we can conclude that the G.L.R.S. has the same in probability representation as given above for $T_n(\vec{w}_0)$ when H_0 is true. Refer in particular to his equation (2.17). Hence, we can infer that $T_n(\vec{w}_0)$ and the G.L.R.S. are not only asymptotically equivalent under H_0, but by a contiguity argument that they are also asymptotically equivalent under sequences of local alternatives, and thus share the same asymptotically optimal properties. See the comment immediately following the proof of Dzhaparidze's Theorem 1. Finally, since it is shown in Dzhaparidze (1977) that subject to regularity conditions the generalized C_α-test for this problem has the above in probability representation, we can conclude more generally that $T_n(\vec{w}_0)$ is also asymptotically equivalent to the C_α-test under suitable conditions.

It is apparent that we could have used the above procedure to derive our statistic. We chose the continuous time regression approach to motivate our statistic, since it can be easily adapted to construct optimal sums of squares of L-statistics tests based on type II and randomly censored data, or based on a finite number of optimally spaced quantiles. (These tests will be developed elsewhere.) It is much less evident how to construct such statistics using the second approach. Also, we believe that our approach gives a clearer picture of how the class of optimal sums of squares of L-statistics tests arise. Finally, we note that the two approaches lead to two sets of regularity conditions. It is not clear how the two sets relate to each other.

REMARK 6. Our procedure can be modified to find optimal sums of squares of L-statistics tests for location only, $Q(\cdot, \mu, \vec{\theta}) = \mu + Q(\cdot, \vec{\theta})$, or scale only,

$Q(\cdot, \sigma, \vec{\theta}) = \sigma Q(\cdot, \vec{\theta})$ models. In the location only model repeat the above proofs without an estimate for scale, whereas in the scale only model repeat the proofs with the condition $\langle w_i, 1 \rangle = 0$ suppressed. With obvious changes of notation, the analogues of the above theorem and remarks remain valid for these models.

REMARK 7. Since for each $1 \le i \le k$ and $n \ge 1$

$$n^{-1} \sum_{j=1}^{n} w_{0,i}(j/(n+1))$$

need not be equal to zero, the statistic $T_n(\vec{w}_0)$ will not always be location and scale invariant. Also, in the special cases that we have looked at so far, for moderate size n the asymptotic null distribution of $T_n(\vec{w}_0)$ has not been a useful approximation to its finite sample distribution. However, we have found that with minor finite sample adjustments, a modified version of $T_n(\vec{w}_0)$ can often be easily constructed, which is location and scale invariant and converges rapidly to its asymptotic null distribution. An example of how this is done for a special case is given in the next section.

4. An example.

Consider the following hypothesis testing situation:

H_0: (Exponential) $Q(\cdot, \sigma) = \sigma Q_0(\cdot)$, where $Q_0(u) = -\ln(1-u)$,

with $0 < \sigma < \infty$ unknown versus

H_a: (Weibull) $Q(\cdot, \sigma, \theta) = \sigma(Q_0(\cdot))^{1/(1+\theta)}$ with both $0 < \sigma < \infty$ and $-1 < \theta < \infty$ unknown.

Notice that this is a special case of the scale only situation described in Remark 6. Standard calculations show that the optimal weight function for the squared L-statistic for testing the above H_0 versus H_a is given by

$$w_0(u) = -(\sqrt{6}/\pi)(C + \ln Q_0(u) - 1/Q_0(u)),$$

where $C = .577216$ is the Euler constant. Since it is easily established that the sample mean $\bar{X}$ is a consistent estimator of scale under local alternatives, an optimal squared L-statistic for this hypothesis testing situation is given by

$$T_n(w_0) = (n^{-1/2} \sum_{i=1}^{n} w_0(i/(n+1))X_{i,n})^2/\bar{X}^2.$$

Practical Implementation of the $T_n(w_0)$ statistic. The results of a Monte Carlo study indicated that for sample sizes $n \le 100$, the distribution of $T_n(w_0)$ is poorly approximated by that of a χ_1^2 random variable. However, we found that with the following small sample corrections, the modified statistic A_n given below converges quite rapidly in distribution to a χ_1^2 random variable:

Let $E_{1,n}, \cdots, E_{n,n}$ denote the order statistics based on n independent exponential random variables with mean one. Set

$$b_n = n^{-1} \sum_{i=1}^{n} w_0(i/(n+1))E(E_{i,n}),$$

and

$$\tau_n^2 = \mathrm{Var}(n^{-1/2} \sum_{i=1}^{n} (w_0(i/(n+1)) - b_n)E_{i,n}).$$

Let

$$L_n = n^{-1/2} \sum_{i=1}^{n} (w_0(i/(n + 1)) - b_n)X_{i,n}.$$

It is easy to verify that under H_0, $EL_n = 0$ and Var $L_n = \sigma^2\tau_n^2$. Consider the modified statistic

$$A_n = L_n^2/(\tau_n^2\bar{X}^2).$$

A_n and $T_n(w_0)$ can be shown to have the same asymptotic distribution both under H_0 and under local alternatives. For the sake of brevity, we do not provide a proof of this fact here. Also note that A_n is scale invariant.

Table 1 gives simulated critical values for the statistic A_n for sample sizes $n = 25$, 50 and 100 with corresponding correction constants b_n and τ_n^2. Each critical value is based on 5000 replications. The numbers in parentheses indicate the empirical α-level if χ_1^2 critical values are used instead of the finite sample critical values. Note that the asymptotic critical values provide a useful approximation to the small sample critical values even for $n = 25$. We also mention in passing that some recent work by Helmers and Hušková (1983) on Berry-Esseen theorems for L-statistics with unbounded weight functions gives some theoretical justification for the apparent rapid rate at which our statistic converges in distribution under H_0 to a χ_1^2 random variable.

For testing an exponential composite null hypothesis versus a general alternative, two among many omnibus goodness-of-fit tests are the statistics

$$E_n = \sum_{i=1}^{n} [X_{i,n}/\bar{X} - Q_0(i/(n + 1))]^2/Q_0(i/(n + 1))$$

proposed by de Wet and Venter (1973), and

$$M_n = \sum_{i=1}^{n} [X_{i,n}/\bar{X} - Q_0(i/(n + 1))]^2(n + 1 - i)/(n + 1),$$

recommended by Csörgő and Révész (1981). Table 2 displays the results of a small sample simulation study comparing the power of the A_n statistic against Weibull alternatives to that of the statistics λ_n, E_n and M_n at sample size $n = 100$ and significance level $\alpha = .10$. The critical values of each of these test statistics for sample size $n = 100$ were determined by a Monte Carlo simulation consisting of 5000 replications, and their power was estimated on the basis of 1000 simulations for each value of the shape parameter θ considered. The numbers

TABLE 1
Simulated Critical Values for A_n

Critical Level	.10	.05	.01	b_n	τ_n^2
$n = 25$	2.6280 (.0952)	4.0132 (.0544)	7.0894 (.0130)	.1144	.8416
$n = 50$	2.7366 (.1014)	3.8438 (.0502)	7.0577 (.0126)	.0647	.8877
$n = 100$	2.6440 (.0940)	3.8340 (.0492)	6.8600 (.0122)	.0360	.9319
$n = \infty$	2.706	3.843	6.637	.0000	1.0000

TABLE 2
Simulated Power Comparisons for $n = 100$ and $\alpha = .10$

θ	A_n	λ_n	M_n	E_n
−.30	.995	.996	.981	.985
−.20	.835	.833	.783	.782
−.10	.379	.356	.356	.347
−.08	.273	.264	.281	.280
−.05	.169	.164	.191	.182
−.03	.136	.124	.139	.125
.05	.198	.218	.102	.073
.10	.351	.372	.188	.117
.20	.744	.777	.478	.397
.30	.951	.962	.785	.736
.40	.995	.998	.950	.934

in the columns indicate the fraction of rejections out of 1000 trials. The computations were performed on a Burroughs 7700 system, using I.M.S.L. subroutines (I.M.S.L. Library, 1978).

As expected, both A_n and λ_n generally outperformed the two omnibus goodness-of-fit tests E_n and M_n. On the other hand, there was little difference between the power of A_n and λ_n. The numerical evidence indicates that λ_n did a little better than A_n for values of $\theta > 0$ and A_n a little better than λ_n for values of $-1 < \theta < 0$, though these apparent small differences could possibly be due to random variation in the simulation study. Finally, note that only one estimate of σ is needed in the computation of A_n, whereas λ_n requires two maximum likelihood estimates of σ and one of θ.

Acknowledgment. The authors would like to thank the associate editor and the referees for pointing out a number of minor technical flaws in the original manuscript. Also, comments by one of the referees led us to the discussion in Remark 5.

REFERENCES

CSÖRGŐ, M. and RÉVÉSZ, P. (1981). *Strong Approximations in Probability and Statistics.* Academic, New York.

DAVIDSON, R. and LEVER, W. (1970). Limiting distribution of the likelihood ratio statistic under a class of local alternatives. *Sankhyā A* **32** 209–224.

DE WET, T. and VENTER, J. H. (1973). A goodness-of-fit test for a scale parameter family of distributions. *South African Statist.* **7** 380–387.

DZHAPARIDZE, K. O. (1977). Tests of composite hypotheses for random variables and stochastic processes. *Theory Prob. Applications* **22** 104–118.

GOVINDARAJULU, Z. and MASON, D. M. (1983). A strong representation for linear combinations of order statistics with application to fixed-width confidence intervals for location and scale parameters. *Scand. J. Statist.* **10** 97–115.

HÁJEK, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives. *Ann. Math. Statist.* **39** 325–346.

HAYAKAWA, T. (1975). The likelihood ratio criterion for a composite hypothesis under local alternatives. *Biometrika* **62** 451–460.

HELMERS, R. and HUŠKOVÁ, M. (1983). A Berry-Esseen bound for L-statistics with unbounded weight functions. To appear in *Proceedings of the Third Prague Symposium on Asymptotic Statistics.*

LABRECQUE, J. (1977). Goodness-of-fit tests based on nonlinearity in probability plots. *Technometrics* **19** 293–306.

MASON, D. M. (1981). Asymptotic normality of linear combinations of order statistics with a smooth score function. *Ann. Statist.* **9** 899–904.

MASON, D. M. (1982). Some characterizations of strong laws for linear functions of order statistics. *Ann. Probab.* **10** 1051–1057.

PARZEN, E. (1961a). An approach to time series analysis. *Ann. Math. Statist.* **32** 951–989.

PARZEN, E. (1961b). Regression analysis for continuous parameter time series. *Proc. 4th Berkeley Symp. Math. Statist. Prob., Vol. I*, Univ. California Press, Berkeley, CA, 469–489.

RAO, C. R. (1973). *Linear Statistical Inference and Its Applications.* Wiley, New York.

SHORACK, G. R. (1972). Functions of order statistics. *Ann. Math. Statist.* **43** 412–427.

SILVEY, S. D. (1959). The Lagrangian multiplier test. *Ann. Math. Statist.* **30** 389–407.

STIGLER, S. M. (1974). Linear functions of order statistics with smooth weight functions. *Ann. Statist.* **2** 676–693.

VAN ZWET, W. R. (1980). A strong law for linear functions of order statistics. *Ann. Probab.* **8** 986–990.

WALD, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.* **54** 426–482.

WELLNER, J. A. (1977). A Glivenko-Cantelli theorem and strong laws of large numbers for functions of order statistics. *Ann. Statist.* **5** 473–480.

DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF DELAWARE
NEWARK, DELAWARE 19716