The Annals of Statistics
1983, Vol. 11, No. 1, 1–12

GOODNESS OF FIT TESTING IN $\mathbb{R}^m$ BASED ON THE WEIGHTED EMPIRICAL DISTRIBUTION OF CERTAIN NEAREST NEIGHBOR STATISTICS[1]

BY MARK F. SCHILLING

University of Southern California

Let $X_1, \cdots, X_n$ be a random sample in $\mathbb{R}^m$ from an unknown density $f(x)$. Recently Bickel and Breiman have introduced a goodness of fit test for this situation based on the empirical distribution function of the variables $W_i = \exp\{-ng(X_i)V(R_i)\}$, $i = 1, \cdots, n$, where $g(x)$ is the hypothesized density and $V(R_i)$ represents the volume of the nearest neighbor sphere centered at X_i. Under the null hypothesis $H: f(x) = g(x)$, the empirical process is asymptotically independent of g. In this paper a weighted version of the empirical process is shown to produce tests which are still (essentially) distribution-free under H but in addition have asymptotic power against sequences of alternatives contiguous to g. The optimal weight function is obtained as a function of the particular sequence of alternatives chosen, and consistency behavior against fixed alternatives is determined. Monte Carlo results illustrate the power performance of the tests for various densities and weight functions.

1. Introduction. Let $X_1, \cdots, X_n$ be a random sample in $\mathbb{R}^m$ from an unknown density $f(x)$. For the hypothesis $H: f(x) = g(x)$ there are suitable tests available for the multivariate normal and a few other special models, but a dearth of effective procedures which can be applied for general g and arbitrary m. A recent paper by Bickel and Breiman (1983) (henceforth referred to as BB) proposes a new goodness of fit test for this situation based on the empirical distribution function $F_n(t)$ of the variables

$$W_i = \exp\{-ng(X_i)V(R_i)\}, \qquad i = 1, \cdots, n,$$

where $R_i = \min_{j \neq i} \| X_j - X_i \|$ is the distance from X_i to its nearest neighbor, and $V(r)$ represents the volume of an m-dimensional sphere of radius r. The term $g(X_i)V(R_i)$ is a first-order approximation to the coverage (probability content) under $g(x)$ of the nearest neighbor sphere centered at X_i.

In BB it is shown under the conditions outlined in Section 2 below that $F_n(t)$ approaches its asymptotic limit uniformly with probability one—under the null hypothesis $H: f(x) = g(x)$ this limit is the uniform distribution—and when appropriately centered and scaled converges weakly when H holds to a Gaussian process with mean zero and covariance function independent of $f = g$. Quadratic functionals of $F_n(t)$ are suggested for testing H and are shown to be consistent against all fixed alternatives.

In the sequel let E without subscript refer to expectation with respect to the true density $f(x)$. It can be shown by means of an asymptotic expansion of $EF_n(t)$ that tests based on $F_n(t)$ have no asymptotic power in the usual sense; i.e., against sequences of alternatives $\{f = f_n\}$ deviating in a particular direction from g by $O(n^{-1/2})$. In this paper a weighted version of the preceding process is considered. This process is defined in Section 2 and asymptotic results analogous to those above are obtained. It is shown that power against any prespecified contiguous sequence of alternatives can be achieved through a

Received May 1981; revised June 1982.
[1] Research supported in part by NSF Grants No. MCS76-10238 and MCS79-19141 and ONR Contract No. N00014-75-C-0444/NR 042-036.
AMS 1980 *subject classifications*. Primary 62G10; 62H15; secondary 62M99, 62E20.
Key words and phrases. Nearest neighbor, goodness of fit, empirical distribution function, contiguity, quadratic functional.

proper selection of the weight function, and the optimal weight function, which depends on the likelihood ratio, is given. In Section 3 quadratic functionals of the weighted process are considered. It is shown that the optimally weighted process leads to an optimal test based on the corresponding functional. Consistency behavior is studied. Finally, Section 4 describes how the asymptotic distributions of these quadratic functionals are produced and presents a Monte Carlo study which presents some promising results for certain situations.

2. The process. Assume throughout that $w(x)$ is a continuous bounded function in $\mathbb{R}^m$ and define on [0, 1] the function

$$\tilde{F}_n(t) = \frac{1}{n}\sum_{i=1}^n w(X_i)I(W_i \le t),$$

where $I(\cdot)$ is the indicator function. Let $\bar{w} = \sup_x |w(x)|$ and write X for X_1. The primary purpose of this section is to show that $\tilde{F}_n(t)$ when normalized appropriately converges weakly to a Gaussian process with covariance function depending only on m and the first two null moments of $w(X)$; this process has zero mean under H but, in general, nonzero mean under particular contiguous alternative sequences. The weight function providing the maximum shift in mean for a given alternative sequence is then derived. Several of the results below are similar to those in BB; for this reason, only outlines of proofs will be given in some cases. Various fourth moment inequalities due to BB which are used below are presented at the end of this section.

2.1 Asymptotic results. We shall draw from the following set of assumptions:

ASSUMPTION A. *A version of f can be chosen such that* (i) $\{f > 0\}$ *is open,* (ii) f *is continuous on* $\{f > 0\}$, (iii) f *is uniformly bounded.*

ASSUMPTION B. *The given function g is nonnegative, and* (i) $\{g = 0\} \subset \{f = 0\}$, (ii) g *is continuous on* $\{f > 0\}$.

Nonnegativity is immediate here since g is assumed to be a density. Note that A subsumes B when H is true. Let

$$\tilde{F}(t) = \int t^{\lambda(x)}w(x)f(x)\,dx,$$

where

$$\lambda(x) = \begin{cases} f(x)/g(x) & \text{if } g(x) > 0 \\ \infty & \text{if } g(x) = 0. \end{cases}$$

THEOREM 2.1. *If Assumption* A(iii) *holds, then*

$$\lim_{n\to\infty}\sup_{t\in[0,1]}|\tilde{F}_n(t) - \tilde{F}(t)| = 0 \quad \text{a.s.}$$

PROOF. Decompose $w(x)$ into $w(x) = w^+(x) - w^-(x)$, where $w^\pm(x) = (\pm w(x)) \vee 0$. Then $w^+(x)$, $w^-(x)$ are nonnegative and bounded. Let $\tilde{F}_n^\pm(t) = (1/n), \sum_{i=1}^n w^\pm(X_i)I(W_i \le t)$. By conditioning on X_i it is easy to show that

$$EF_n^\pm(t) \to \tilde{F}^\pm(t) = \int t^{\lambda(x)}w^\pm(x)f(x)\,dx. \tag{2.1}$$

Furthermore, from Chebycheff's inequality and (2.5),

$$P[|\tilde{F}_n^\pm(t) - E\tilde{F}_n^\pm(t)| \ge \varepsilon] = O(n^{-2}) \tag{2.2}$$

for any $\varepsilon > 0$. Then (2.1) and (2.2) together with the Borel-Cantelli lemma imply that

$\tilde{F}_n^{\pm}(t) \to \tilde{F}^{\pm}(t)$ a.s. pointwise for $t < 1$. In addition, by the strong law of large numbers,

$$\tilde{F}_n^{\pm}(1^-) \to \tilde{F}^{\pm}(1^-) = \int_{\{g(x)>0\}} w^{\pm}(x) f(x)\, dx \quad \text{a.s.} \tag{2.3}$$

Now let

$$J_n^{\pm}(t) = \begin{cases} \tilde{F}_n^{\pm}(t)/\tilde{F}_n^{\pm}(1^-) & t < 1, \\ 1 & t = 1, \end{cases}$$

and define $J^{\pm}(t)$ similarly for $\tilde{F}^{\pm}(t)$. Then $J_n^{\pm}(t)$ are distribution functions on [0, 1] approaching the continuous limiting distribution functions $J^{\pm}(t)$; thus

$$\sup_{t\in[0,1]} | J_n^{\pm}(t) - J^{\pm}(t) | \to 0 \qquad \text{a.s.}$$

The theorem follows from $\tilde{F}_n(t) = \tilde{F}_n^{+}(t) - \tilde{F}_n^{-}(t)$ and (2.3).

Now form the normalized process

$$\tilde{Z}_n(t) = n^{1/2}\{\tilde{F}_n(t) - E_g\tilde{F}_n(t)\}$$

on [0, 1]. If $w(x) = 1$, then $\tilde{F}_n(t) = F_n(t)$ and the $\tilde{Z}_n$ process will be denoted by $Z_n(t)$. Consider the testing problem involving

$$H: f(x) = g(x)$$

and the sequence of alternatives

$$K_n: f = f_n(x) = g(x) + n^{-1/2}h(x) + \varepsilon_n(x),$$

where $\varepsilon_n(x) \sim o(n^{-1/2})$ for each x and $\{f_n\}$ and g satisfy

(i) $\displaystyle\int \frac{h^2(x)}{g(x)}\, dx < \infty,$

(ii) $\displaystyle\lim_{n\to\infty} \int \left[n^{1/2}\{f_n^{1/2}(x) - g^{1/2}(x)\} - \frac{h(x)}{2g^{1/2}(x)} \right]^2 dx = 0.$

A specialization of a result by LeCam (1960) shows that $\{f_n\}$ is then contiguous to g. A simple case in which (ii) is satisfied is when $n^{1/2}\varepsilon_n(x) \to 0$ uniformly in x.

Define

$$r(x) = h(x)/g(x)$$

on $\{x: g(x) > 0\}$. It will be assumed throughout that the densities $\{f_n(x)\}$ satisfy Assumption A and that $r(x)$ is bounded in absolute value. The contiguity condition (i) above is fulfilled by virtue of this latter assumption.

The limiting distribution of $\tilde{Z}_n(t)$ under both H and $\{K_n\}$, when the densities involved satisfy Assumptions A and B, is established by Theorems 2.2 and 2.3 below. Contiguity plays an essential part. Denote the log-likelihood ratio of f_n to g by

$$L_n^0 = \log \prod_{i=1}^n \{f_n(X_i)/g(X_i)\} = \sum_{i=1}^n \log\{1 + n^{-1/2}r(X_i) + \varepsilon_n(X_i)/g(X_i)\}.$$

The asymptotic joint behavior of L_n^0 and $\tilde{Z}_n(t)$ must be determined. Let

$$L_n = n^{-1/2} \sum_{i=1}^n r(X_i), \qquad \sigma_r^2 = E_g r^2(X).$$

LEMMA 2.1. *$L_n^0 - L_n \to -\sigma_r^2/2$ a.s. under H.*

This is a consequence of a proposition by LeCam (1971). The lemma allows L_n to be used in place of the less convenient log-likelihood ratio.

THEOREM 2.2. *Under Assumption A, $(L_n - \sigma_r^2/2, \tilde{Z}_n(t))$ converges weakly under H to $(L, \tilde{Z}(t))$, where (i) $L \sim N(-\sigma_r^2/2, \sigma_r^2)$; (ii) $\tilde{Z}(t)$ is a mean zero Gaussian process with* *covariance function*

$$k(s,t) = E\tilde{Z}(s)\tilde{Z}(t) = \left[s(1 + t\log s) + st \int_{B(s,t)} \{\eta(s,t,\omega) - 1\}\, d\omega \right] E_g w^2(X)$$

$$- st(1 + \log st + \log s \log t)\{E_g w(X)\}^2 \quad \text{for} \quad s \le t,$$

with

$$B(s,t) = \{\omega \in \mathbb{R}^m : r_1 \le \|\omega\| < r_1 + r_2\}$$

and

$$\log \eta(s,t,\omega) = \int_{\{z \in \mathbb{R}^m : \|z\| \le r_1,\ \|z-\omega\| \le r_2\}} dz,$$

where r_1 and r_2 are defined by $V(r_1) = -\log s$, $V(r_2) = -\log t$ and $\|\cdot\|$ represents the Euclidean norm; and (iii) $\mathrm{Cov}(L, \tilde{Z}(t)) = \sigma_{12} = t(1 + \log t)E_g r(X)w(X)$.

The convergence of $\tilde{Z}_n$ to $\tilde{Z}$ occurs in $D[0, 1]$ although $\tilde{Z}$ is in fact in $C[0, 1]$.

PROOF. It is first shown that the sequence of processes $\tilde{Z}_1(t), \tilde{Z}_2(t), \cdots$ is tight. Write $\tilde{Z}_n(t) = Z_n^+(t) - Z_n^-(t)$ with

$$Z_n^{\pm}(t) = n^{1/2}\{\tilde{F}_n^{\pm}(t) - E_g \tilde{F}_n^{\pm}(t)\},$$

where $\tilde{F}_n^{\pm}(t)$ are defined in Theorem 2.1. It is enough to prove that the sequence $\{Z_n^+\}$ (say) is tight. The procedure is briefly sketched below.

Define $Q_n(t) = 1 - H_n(-\log t/V(1))$, where $H_n(\cdot)$ is defined in (2.7), and for given $\delta > 0$ choose $t_1, \cdots, t_k$ such that $Q_n(t_i) = i\delta n^{-1/2}$, $i = 1, \cdots, k$ with $k\delta n^{-1/2} \le 1 < (k + 1)\delta n^{-1/2}$. Under Assumptions A(iii) and B, $Q_n(t)$ converges to a continuous distribution function $Q(t)$. Let

$$Z_n^0(t) = Z_n^+(t_i) + \frac{n^{1/2}}{\delta}\{Q_n(t) - Q_n(t_i)\}\{Z_n^+(t_{i+1}) - Z_n^+(t_i)\}$$

for $t_i \le t < t_{i+1}$, $0 \le i \le k$, where $t_0 = 0$, $t_{k+1} = 1$. Following a method of Shorack (1973) and using Billingsley (1968, Theorem 12.3), Corollary 2.3 to Theorem 2.4 implies that $\{Z_n^0\}$ satisfies the moment condition

$$E\{Z_n^0(t) - Z_n^0(s)\}^4 \le M(\delta)\{Q(t) - Q(s)\}^2 \qquad 0 \le s, t \le 1$$

for M independent of n and therefore $\{Z_n^0\}$ is tight with all limit points in $C[0, 1]$. The monotonicity of $F_n^+(t)$ and $E_g F_n^+(t)$ can then be used to show that the sequence $\{Z_n^+\}$ is properly "close" to $\{Z_n^0\}$ and is therefore also tight with a.s. continuous limits.

We now turn attention to the asymptotic behavior of the finite-dimensional distributions of $(\tilde{Z}_n(t_1), \cdots, \tilde{Z}_n(t_k), L_n)$ for arbitrary k and $t_1, \cdots, t_k$. The asymptotic normality and limiting covariance matrix of $(\tilde{Z}_n(t_1), \cdots, \tilde{Z}_n(t_k))$ are determined by specializations of results in Sections 3-4 of BB. What remains to be shown, restricting for simplicity to the case $k = 1$ and taking $t_1 = t$, is that $\tilde{Z}_n(t)$ and L_n are actually joint normal asymptotically with the indicated covariance. The argument t is suppressed.

Begin by constructing a sequence of compact sets $C_1 \subset C_2 \subset \cdots \subset \mathbb{R}^m$ having diameter$(C_N) \le N$ with $\sigma_N = \inf\{g(x) : x \in C_N\} > 0$ and $P(X_1 \notin C_N) \to 0$. For given N, introduce a cube D_N of sidelength N containing C_N. Partition D_N into L congruent subcubes whose closures intersected with C_N are labelled as $B_1, \cdots, B_L$. Put $B_0 = C_N^c$, and denote the number of points in B_ℓ by n_ℓ. Select $d_N > 0$ and let $E_N = \{x : x \in B_\ell \text{ for } \ell \ge 1, \|x - y\| > d_N\ \forall y \in \cup_{\ell' \ne \ell} B_{\ell'}\}$. For any random variable Y, let $Y^* = Y - E_g Y$. By choosing L and d_N in such a way that $\max_\ell P(X \in B_\ell) \to 0$ as $P(X \notin E_N) \to 0$ as $N \to \infty$, we obtain the desired result from the arguments below.

For $X_j \in B_\ell$, let R_j' represent the nearest neighbor distance to X_j among the remaining points in B_ℓ, if any, taking $R_j' = \infty$ otherwise. Let

$$Z_n' = n^{-1/2} \sum_{j=1}^n h^*(X_j, R_j'),$$

where $h(x, u) = I(x \in E_N)w(x)I(\exp\{-ng(x)V(u)\} \le t)$. By reindexing the X_j and matching R_j' according to cell membership, the quantities

$$T_\ell = \begin{cases} \sum_{j=1}^{n_\ell} h^*(X_j^{(\ell)}, R_j'^{(\ell)}) & n_\ell > 1, \\ 0 & n_\ell \le 1, \end{cases}$$

are obtained, which represent the contribution to Z_n' of each cell B_ℓ; i.e., $Z_n' = n^{-1/2} \cdot \sum_{\ell=1}^L T_\ell$. The advantage here is the independence of the T_ℓ given $\mathbf{n} = (n_0, \cdots, n_L)$. It follows from (2.5) that $E(\tilde{Z}_n - Z_n')^2 \to 0$ so that Z_n' may indeed by used in place of $\tilde{Z}_n$. Similarly, write

$$L_n = n^{-1/2} \sum_{\ell=1}^L Y_\ell$$

with

$$Y_\ell = \sum_{j=1}^{n_\ell} r(X_j^{(\ell)}), \qquad \ell = 1, \cdots, L.$$

For the asymptotic joint normality of Z_n' and L_n it is sufficient to show that $aZ_n' + bL_n$ converges to $N(0, a^2k(t, t) + 2ab\sigma_{12} + b^2\sigma_r^2)$ for all real numbers a and b not both zero. Take $a = 1$ without loss of generality since for $a = 0$ the result follows by the basic Lindeberg-Feller Central Limit Theorem. Make the decomposition $Z_n' = U_n + V_n$ where

$$U_n = n^{-1/2} \sum_{\ell=1}^L \{T_\ell - E(T_\ell \mid n_\ell)\},$$

$$V_n = n^{-1/2} \sum_{\ell=1}^L \{E(T_\ell \mid n_\ell) - ET_\ell\}.$$

The characteristic function of $Z_n' + bL_n$ is

$$E\{e^{it(Z_n'+bL_n)}\} = E[e^{itV_n}E\{e^{it(U_n+bL_n)} \mid \mathbf{n}\}].$$

It is proved in BB that $V_n \to_{\mathscr{L}} N(0, \sigma^2)$ with $\sigma^2 < \infty$ and $E(U_n^2 \mid \mathbf{n}) \to s^2$ a.s. in L_1 where $s^2 = k(t, t) - \sigma^2$ independently of $\mathbf{n}$. Write

$$U_n + bL_n = \sum_{\ell=1}^L \xi_\ell,$$

where

$$\xi_\ell = n^{-1/2}\{T_\ell - E(T_\ell \mid n_\ell) + bY_\ell\}, \qquad \ell = 1, \cdots, L,$$

and note that the ξ_ℓ are mutually independent given $\mathbf{n}$. Thus

$$E\{e^{it(U_n+bL_n)} \mid \mathbf{n}\} = \prod_{\ell=1}^L f_\ell(t)$$

where

$$f_\ell(t) = E(e^{it\xi_\ell} \mid n_\ell), \qquad \ell = 1, \cdots, L.$$

The $f_\ell(t)$ can be expanded by means of appropriate bounds on $E(|\xi_\ell^p| \mid n_\ell)$ $(p = 2, 3)$ obtained from (2.5) to yield

$$\prod_{\ell=1}^L f_\ell(t) = \exp\left\{-\frac{t}{2} \sum_{\ell=1}^L E(\xi_\ell^2 \mid n_\ell) + \Delta_n\right\}$$

for t in a restricted range $R(t)$, where under appropriate limiting operations Δ_n becomes arbitrarily small and $R(t)$ grows to $(-\infty, \infty)$. Finally, the result

$$\sum_{\ell=1}^L E(\xi_\ell^2 \mid n_\ell) \to s^2 + 2b\sigma_{12} + b^2\sigma_r^2 \text{ a.s. in } L_1$$

can be established by showing that

(i) $\frac{1}{n}\sum_{\ell=1}^{L}\operatorname{Cov}(T_\ell, Y_\ell \mid n_\ell) \to \sigma_{12}$ a.s. in L_1, and

(ii) $\frac{1}{n}\sum_{\ell=1}^{L}\operatorname{Var}(Y_\ell \mid n_\ell) \to \sigma_r^2$ a.s. in L_1

Equation (ii) follows immediately from the independence of the terms comprising Y_ℓ, and (i) can be shown through minor modifications of the ensuing argument, which establishes the overall covariance of $\tilde{Z}_n$ and L_n.

PROPOSITION 2.1. $\lim_{n\to\infty}\operatorname{Cov}_g(\tilde{Z}_n(t), L_n) = \sigma_{12}$.

PROOF. Begin with

$$\begin{aligned}\operatorname{Cov}_g(L_n, \tilde{Z}_n(t)) &= E_g L_n \tilde{Z}_n(t)\\ &= E_g r(X_1)w(X_1)I(W_1 \le t) + (n-1)E_g r(X_2)w(X_1)I(W_1 \le t).\end{aligned}$$

By conditioning on X_1, it is easy to obtain $tE_g r(X_1)w(X_1)$ as the limit of the first term on the right. For the second term condition on both X_1 and X_2 to obtain

$$(n-1)E_g\{r(X_2)w(X_1)P_g(W_1 \le t \mid X_1, X_2)\}. \tag{2.4}$$

Set

$$\sigma_n(x) = \left\{\frac{-\log t}{V(1)ng(x)}\right\}^{1/m}$$

and let $\rho_n(X_1)$ represent $P_g(W_1 \le t \mid X_1)$ with X_2 deleted from the sample. Then

$$P_g(W_1 \le t \mid X_1, X_2) = \begin{cases}\rho_n(X_1) & \|X_2 - X_1\| \ge \sigma_n(X_1),\\ 0 & \|X_2 - X_1\| < \sigma_n(X_1),\end{cases}$$

and (2.4) may be written as

$$\begin{aligned}&(n-1)\int w(x_1)\rho_n(x_1)g(x_1)\int_{\{\|x_2-x_1\|\ge\sigma_n(x_1)\}} r(x_2)g(x_2)\,dx_2\,dx_1\\ &\qquad = -(n-1)\int w(x_1)\rho_n(x_1)g(x_1)\int_{\{\|x_2-x_1\|<\sigma_n(x_1)\}} r(x_2)g(x_2)\,dx_2\,dx_1\end{aligned}$$

using $E_g r(X_2) = 0$. Now letting $u = (x_2 - x_1)/\sigma_n(x_1)$ for the inner integral we obtain

$$-(n-1)\int w(x_1)\rho_n(x_1)\,g(x_1)\int_{\{\|u\|<1\}} r(x_1 + \sigma_n u)g(x_1 + \sigma_n u)\sigma_n^m(x_1)\,du\,dx_1.$$

Using $\lim_{n\to\infty}\rho_n(x_1) \equiv t$, continuity of r and dominated convergence, the desired result follows.

The limiting distribution of $\tilde{Z}_n(t)$ under $\{K_n\}$ can now be easily determined.

THEOREM 2.3. *If Assumptions A and B hold, then under* $\{K_n\}$, $\tilde{Z}_n(t)$ *converges weakly to* $\tilde{Z}(t) + t(1 + \log t)E_g r(X)w(X)$.

PROOF. Convergence of the finite-dimensional distributions follows immediately from Theorem 2.2 and LeCam's third lemma (Hájek and Sǐdák, 1967, page 208). The proof of tightness comes directly from the definition of contiguous probability measures; for given γ and ε,

$$P[\max\{|\tilde{Z}_n(t_2) - \tilde{Z}_n(t_1)| : |t_2 - t_1| \le \gamma\} \ge \varepsilon\,] \to 0$$

under H implies that the same holds under $\{K_n\}$ as well.

Theorem 2.3 verifies the assertion that if $w(x) \equiv 1$, tests based on $\tilde{Z}_n(t)$ have no power against contiguous alternatives. Interestingly, this is also the case for tests based on $\tilde{Z}_n(e^{-1})$ for any weight function.

2.2. Optimal selection of the weight function. Tests which use $\tilde{Z}_n(t)$ for some $t \in (0, 1]$, $t \neq e^{-1}$, as the test statistic will have asymptotic power under $\{K_n\}$ which depends on the normalized shift

$$S(t) = t(1 + \log t)\{\text{Var } \tilde{Z}(t)\}^{-1/2} E_g r(X) w(X);$$

maximizing $|S(t)|$ will make the limiting power as large as possible.

PROPOSITION 2.2. *Among all possible weight functions, the choice $w(x) = ar(x)$, $a \neq 0$, maximizes the limiting power of the test based on $\tilde{Z}_n(t)$ for fixed $t \in (0, 1]$, $t \neq e^{-1}$.*

PROOF. $\text{Var } \tilde{Z}(t) = k(t, t) = c_1(t)\text{Var}_g w(X) + c_2(t)\{E_g w(X)\}^2$, where $c_1(t)$, $c_2(t)$ depend only on t. By considering the unweighted statistic $Z(t)$ it is clear that $c_2(t) > 0$. Thus, given any weight function $w(x)$ with $E_g w(X) \neq 0$, the centered version $w_1(x) = w(x) - E_g w(X)$ yields the same covariance with $r(X)$ but makes $\text{Var } \tilde{Z}(t)$ smaller, thereby producing a larger shift than does $w(x)$. Among weight functions $w(x)$ with expectation zero,

$$|S(t)| = t|1 + \log t| c_1^{-1/2}(t)\{E_g r^2(X)\}^{1/2} |\rho_g(r(X), w(X))|,$$

where ρ is the correlation; this is a maximum for $w(x) = ar(x) + b$, $a \neq 0$, whence $b = E_g w(X) = 0$.

Denote the statistic $\tilde{Z}_n(t)$ with weight function $r(x)$ by $Z_n^*(t)$. Note that $Z_n^*(1) = L_n$, so that the test based on $Z_n^*(1)$ is asymptotically equivalent to the likelihood ratio test of H versus $\{K_n\}$.

2.3. Moment bounds. This subsection presents some fourth moment inequalities which are utilized in several of the preceding proofs. Within the model of Section 1, write

$$F(A) = \int_A f(x)\,dx; \; S(x, r) = \{y : \|y - x\| \leq r\}; \; D_i = n^{1/m} R_i, \qquad i = 1, \cdots, n.$$

Let h be a bounded, measurable function from $\mathbb{R}^m \times [0, \infty)$ to $\mathbb{R}^1$ and put $h_i = h(X_i, D_i)$ and $h_i^* = h_i - Eh_i$, $i = 1, \cdots n$. The theorem and first two corollaries given below represent Theorem 2.1 and Corollaries 2.3, 2.5 of BB respectively.

THEOREM 2.4. *Let h be bounded with $\bar{h} = \sup_{x,d}|h(x, d)|$. Then*

$$E(\textstyle\sum_{i=1}^n h_i^*)^4 \leq Mn^2\bar{h}^2[(E|h_1|)^2 + n^2\{E|h_1|F^2(S(X_1, R_1))\}^2 + n^{-1}\bar{h}^2]$$

for some $M < \infty$ depending only on m.

COROLLARY 2.1. *Suppose $h(x, d) = u(x)v(x, d)$ for u, v bounded. Then*

$$E(\textstyle\sum_{i=1}^n h_i^*)^4 \leq M\{n^2(E|u(X)|)^2 + n\} \tag{2.5}$$

for some $M < \infty$ depending only on m, $\sup_x|u(x)|$ and $\sup_{x,d}|v(x, d)|$.

COROLLARY 2.2. *If $h(x, d) = I(a \leq g(x)\, d^m \leq b)$ then*

$$E(\textstyle\sum_{i=1}^n h_i^*)^4 \leq M[n^2\{H_n(b) - H_n(a)\}^2 + n] \tag{2.6}$$

for $M < \infty$ depending only on m, where

$$H_n(y) = (1 - e^{-n/2})^{-1} \int f(x)\left[1 - \exp\left\{-\frac{n}{2} F(S(x, (y/ng(x))^{1/m}))\right\}\right] dx. \tag{2.7}$$

A slight extension of Corollary 2.2 is needed for our purposes. The proof is omitted.

COROLLARY 2.3. *Let $h(x, d) = u(x)I(a \leq g(x)d^m \leq b)$ for $u(x)$ bounded. Then* (2.6) *holds for some $M < \infty$ depending only on M and $\sup_x|u(x)|$.*

3. The quadratic functional. Define the functional

$$\tilde{S}_n = \int_0^1 \tilde{Z}_n^2(t)\, dt$$

on $D[0,1]$. The asymptotic null distribution of $\tilde{S}_n$ is indicated in the following corollary to Theorem 2.2:

COROLLARY 3.1. *If $f = g$ and A holds, $\tilde{S}_n$ converges in distribution to*

$$\tilde{S} = \int_0^1 \tilde{Z}^2(t)\, dt.$$

PROOF. Apply Donsker's Theorem (Billingsley, 1968, Theorem 5.1).

3.1. Distribution of $\tilde{S}$. The distribution of functionals such as $\tilde{S}$ has been studied extensively. The main result, due to Kac and Siegert (1947), is that $\tilde{Z}(t)$ is equivalent in distribution to $\sum_{j=1}^{\infty} \lambda_j^{1/2}\phi_j(t)V_j$, where $V_1, V_2, \cdots$ are i.i.d. standard normal variables and $\{(\lambda_j, \phi_j(t)); j = 1, \cdots, \infty\}$ is the set of all eigenvalues and corresponding eigenfunctions of the integral equation

$$\int_0^1 k(s, t)\phi(s)\, ds = \lambda\phi(t). \tag{3.1}$$

Each λ_j is positive as a result of the positive definiteness of $k(s, t)$. The eigenfunctions $\phi_1(t), \phi_2(t), \cdots$ form a complete orthonormal set; thus $\tilde{S}$ is equivalent to $\sum_{j=1}^{\infty} \lambda_j V_j^2$.

3.2 Optimality. One would expect the functional $\int_0^1 Z_n^{*2}(t)\, dt$ to possess a similar optimality property to that of $Z_n^*(t)$ for a specified sequence $\{K_n\}$. The Kac-Siegert representation can be used to verify this.

PROPOSITION 3.1. *Among all weight functions with $E_g w(X) = 0$, the choice $w(x) = ar(x)$, for any $a \neq 0$, maximizes the limiting power of the test which rejects for large values of $\tilde{S}_n$.*

PROOF. Without loss of generality restrict to weight functions satisfying $E_g w(X) = 0$, $E_g w^2(X) = 1$. Recall from Theorem 2.3 that under the sequence $\{K_n\}$

$$\tilde{Z}_n(t) \to \tilde{Z}(t) + t(1 + \log t)E_g r(X)w(X). \tag{3.2}$$

The completeness of the $\phi_j(t)$ sequence allows the shift term in (3.2) to be expressed as $\sum_{j=1}^{\infty} \lambda_j^{1/2}\alpha_j(w)\phi_j(t)$ for some constants $\alpha_1(w), \alpha_2(w), \cdots$. Hence the right-hand side of (3.2) is equivalent to $\sum_{j=1}^{\infty} \lambda_j^{1/2}\phi_j(t)\{V_j + \alpha_j(w)\}$ and thus $\tilde{S}$ is representable as the sum of the positively weighted independent non-central Chi squared variables $\{V_1 + \alpha_1(w)\}^2$, $\{V_2 + \alpha_2(w)\}^2, \cdots$. Since each $\alpha_j(w)$ is clearly proportional to $E_g r(X)w(X)$ independently of t, the $\alpha_j(w)$'s are simultaneously maximized in absolute value by that $w(x)$ which maximizes $E_g r(X)w(X)$, namely $w(x) = r(x)$. This makes all of the variables $\{V_j + \alpha_j(w)\}^2$—and hence $\tilde{S}$—stochastically as large as possible.

3.3. Consistency. The following theorem describes the consistency behavior of the $\tilde{S}_n$ test for fixed alternatives.

THEOREM 3.1. *If Assumptions A and B hold, and if* (i) $\{x : f(x) \neq g(x)\} \cap \{x : w(x) \neq 0\}$ *has positive probability,* (ii) $w(x) = \psi(\lambda(x))$ *for some ψ, then $\tilde{S}_n \to_{\mathscr{L}} \infty$.*

PROOF. $$\tilde{S}_n = n\int_0^1 \{\tilde{F}_n(t) - E_f\tilde{F}_n(t)\}^2\,dt$$
$$+2n\int_0^1 \{\tilde{F}_n(t) - E_f\tilde{F}_n(t)\}\{E_f\tilde{F}_n(t) - E_g\tilde{F}_n(t)\}\,dt$$
$$+n\int_0^1 \{E_f\tilde{F}_n(t) - E_g\tilde{F}_n(t)\}^2\,dt$$
$$\le n\int_0^1 \{E_f\tilde{F}_n(t) - E_g\tilde{F}_n(t)\}^2\,dt$$
$$-2n\left[\int_0^1 \{\tilde{F}_n(t) - E_f\tilde{F}_n(t)\}^2\,dt\int_0^1 \{E_f\tilde{F}_n(t) - E_g\tilde{F}_n(t)\}^2\,dt\right]^{1/2}$$

It is enough to show that $\lim_{n\to\infty}\int_0^1 \{E_f\tilde{F}_n(t) - E_g\tilde{F}_n(t)\}^2\,dt > 0$. From Theorem 2.1 we have

$$\zeta(t) = \lim_{n\to\infty}\{E_f\tilde{F}_n(t) - E_g\tilde{F}_n(t)\} = \int w(x)\{f(x)t^{\lambda(x)} - tg(x)\}\,dx.$$

Define $\zeta_1(t) = \zeta(t)/t$. It suffices to show that $t^2(d\zeta_1(t)/dt) \not\equiv 0$ in $(0, 1]$; the result then follows from continuity of $\zeta(t)$ and dominated convergence. Now upon passing the derivative,

$$t^2\frac{d}{dt}\zeta_1(t) = \int w(x)\{\lambda(x) - 1\}t^{\lambda(x)}f(x)\,dx.$$

This can be viewed as the expectation of a constant times $w(X)\{\lambda(X) - 1\}$ with respect to the probability measure

$$dP_t(x) = \frac{t^{\lambda(x)}f(x)\,dx}{\int t^{\lambda(x)}f(x)\,dx}.$$

On considering t as a parameter, $\{P_t(x)\}$ defines an exponential family having $\sum_{i=1}^n \lambda(X_i)$ as a sufficient statistic for t. Hence $\{P_t(x)\}$ is complete (Lehmann, 1959, page 132) and the result follows.

COROLLARY 3.2. *If Assumptions A and B hold, the test which rejects for large values of $S_n = \int_0^1 Z_n^2(t)\,dt$ is consistent against all $f(x) \ne g(x)$.*

PROOF. Choose $\psi(\cdot) \equiv 1$. This result can also be proven by utilizing Laplace transforms.

COROLLARY 3.3. *If Assumptions A and B hold, the $\tilde{S}_n$ test with optimal weight function for testing f against g is consistent against f provided $f(x) \ne g(x)$.*

PROOF. Choose $\psi(u) = u - 1$ in Theorem 3.1.

Unfortunately the weighted test may be consistent against alternatives for which the weight function is not optimal. For example if $E_g\{w(X)\,|\,\lambda(X)\} \equiv 0$ then $\zeta(t) \equiv 0$ in Theorem 3.1, and at least for $m < 3$ it can be shown that the $\tilde{S}_n$ test is not consistent. However, it is easy to obtain tests which have both asymptotic power against a particular sequence $\{K_n\}$ and consistency against all fixed alternatives by merely combining the S_n and $\tilde{S}_n$ tests. Proposition 3.2 validates this approach.

PROPOSITION 3.2. *Let $w(x)$ be chosen so that $E_g w(X) = 0$. Then under the conditions of Theorem 2.2, S_n and $\tilde{S}_n$ are asymptotically independent.*

TABLE 1

Monte Carlo powers of the $\tilde{S}_n$ test and competitors for $m = 1$, size $\alpha = 0.10$. Asterisk indicates cases where w is optimal for f. Entries in parentheses are estimates described in Section 3.

† Density f	Sample size n	Statistic††			
(a) g is $U[0, 1]$					
		S_n	$\tilde{S}_n, w_1$	KS	χ^2
f_1	25	.17	.86(.77)*	.89	.78
	100	.10	.82(.68)*	.86	.49
f_2	25	.19	.02(.12)	.28	.67
	100	.12	.08(.11)	.40	.56
f_3	25	.13	.56(.44)	.71	.52
	100	.10	.56(.46)	.70	.36
f_4	25	.12	.34(.28)	.42	.36
	100	.11	.32(.24)	.43	.22

† Density f	Sample size n					
(b) g is $N(0, 1)$						
		S_n	$\tilde{S}_n, w_2$	$\tilde{S}_n, w_3$	KS	χ^2
f_5	25	.12	.34(.24)*	.06(.07)	.38	.26
	100	.10	.41(.33)*	.08(.10)	.43	.21
f_6	25	.10	.09(.09)	.28(.26)*	.14	.18
	100	.07	.10(.09)	.31(.21)*	.16	.22

† Density f	Sample size n				
(c) g is Exp(1)					
		S_n	$\tilde{S}_n, w_4$	KS	χ^2
f_7	25	.10	.37(.28)*	.41	.30
	100	.09	.49(.40)*	.42	.29

† $f_1(x) = 1 + 10n^{-1/2}(x - ½)$; $f_2(x) = 1 + n^{-1/2}(5 - 20\,|x - ½|)$; $f_3(x) = 1 + 2.5n^{-1/2}\text{sgn}(x - ½)$; $f_4(x) = 0.6 - 0.4 \log x$; f_5 is $N(1.645n^{-1/2}, 1)$; f_6 is $N(0, 1 + 1.163n^{-1/2})$; $f_7(x) = \theta_n^{-1}\exp(-x/\theta_n)$, where $\theta_{25} = 1.447$, $\theta_{100} = 1.196$.

†† Weight functions for $\tilde{S}_n$ are: $w_1(x) = \sqrt{3}(2x - 1)$; $w_2(x) = x$; $w_3(x) = (x^2 - 1)/\sqrt{2}$; $w_4(x) = x - 1$.

PROOF. Following the proof of Theorem 2.2 in a straightforward manner it is apparent that $Z_n(t)$ and $\tilde{Z}_n(t)$ are joint normal for each t. A calculation similar to that for the covariance kernel of $\tilde{Z}_n(t)$ yields, for $s \le t$,

$$\lim_{n\to\infty}\text{Cov}_g(Z_n(s), \tilde{Z}_n(t)) = \lim_{n\to\infty}\text{Cov}_g(\tilde{Z}_n(s), Z_n(t)) = k(s, t)E_g w(X),$$

and the proposition follows.

A simple test which combines the S_n and $\tilde{S}_n$ tests is the test which rejects if either the S_n or the $\tilde{S}_n$ test rejects at level $\alpha/2$. This test has asymptotic level $\alpha - \alpha^2/4$ and limiting power approximately equal to that of the $\tilde{S}_n$ test with level $\alpha/2$.

4. Power performance. To table the distribution of $\tilde{S}$ requires solving (3.1) in order to obtain the eigenvalues $\lambda_1, \lambda_2, \cdots$. Except in the one-dimensional case, the intractability of the term $\int_{B(s,t)} \{\eta(s, t, \omega) - 1\}\, d\omega$ prevents this. It is possible, however to produce a simple form by letting m tend to infinity. This preserves the positive definiteness of $k(s, t)$; thus we may think of an infinite-dimensional version of $\tilde{S}$ having covariance function equal to the limit in m of (3.1). The distributions of $\tilde{S}$ for $m = 1$ and $m = \infty$ are quite close and there is strong evidence that the $\tilde{S}$ distributions decrease monotonically (pointwise) as m varies from one to infinity. Hence the $m = 1$ and $m = \infty$ distributions appear to provide adequate approximations for experiments with $1 < m < \infty$.

TABLE 2

Monte Carlo powers of the $\tilde{S}_n$ test and a competitor for $m = 3, 5$, $g = N(0, I)$, and size $\alpha = 0.10$. Entries in parentheses are estimates described in Section 3. Asterisk indicates cases where w is optimal for f.

† Density f	Sample size n	Statistic††				
(a) $m = 3$						
		S_n	$\tilde{S}_n, w_5$	$\tilde{S}_n, w_6$	$\tilde{S}_n, w_7$	SR
f_8	25	.23	.60(.56)*	.06(.15)	.06(.17)	.14
	100	.21	.59(.55)*	.16(.20)	.08(.16)	.06
f_9	25	.17	.06(.12)	.35(.31)*	.04(.08)	.39
	100	.13	.10(.14)	.44(.35)*	.07(.12)	.39
		S_n	$\tilde{S}_n, w_7$	$\tilde{S}_n, w_8$	$\tilde{S}_n, w_9$	SR
f_{10}	25	.12	.39(.31)*	.35(.30)	.08(.10)	.12
	100	.15	.46(.39)*	.41(.38)	.15(.15)	.12
(b) $m = 5$						
		S_n	$\tilde{S}_n, w_5$	$\tilde{S}_n, w_6$	$\tilde{S}_n, w_7$	SR
f_8	25	.16	.59(.51)*	.10(.12)	.04(.09)	.13
	100	.16	.64(.57)*	.22(.22)	.06(.12)	.12
f_9	25	.18	.12(.14)	.38(.33)*	.03(.07)	.33
	100	.15	.11(.17)	.47(.38)*	.06(.14)	.35
		S_n	$\tilde{S}_n, w_7$	$\tilde{S}_n, w_8$	$\tilde{S}_n, w_9$	SR
f_{10}	25	.13	.35(.30)*	.33(.28)	.08(.12)	.10
	100	.13	.34(.29)*	.39(.33)	.19(.18)	.12

† $f_8(x) = g(x) + (0, 2.487n^{-1/2}, 0, \cdots, 0)$; $f_9(x) = g(x_1, (1 + .75n^{-1/2})x_2, (1 + 1.5n^{-1/2})x_3)$; f_{10} is $(1 - 2.5n^{-1/2})\cdot N(0, I) + 2.5n^{-1/2}N((0.75, 0, \cdots, 0), I)$.

†† weight functions for $\tilde{S}_n$ are: $w_5(x) = x_2$; $w_6(x) = 10^{-1/2}\{x_2^2 - 1 + 2(x_3^2 - 1)\}$; $w_7(x) = 1.151\{\exp(.75x_1 - .281) - 1\}$; $w_8(x) = x_1$; $w_9(x) = 10^{-1/2}\{x_2^2 - 1 + 2(x_1^2 - 1)\}$.

The foregoing is treated fully in Schilling (1983); Monte Carlo experiments discussed there indicate a reasonable fit under the null hypothesis to the distributions mentioned above for $m = 1, 3, 5$; $n = 25, 100$. Performance under the alternative is presented in this section for the same choices of m and n and various $g(x)$, $f(x)$ and $w(x)$. The nonconstant weight functions are normalized to satisfy $E_g w(X) = 0$, $E_g w^2(X) = 1$.

The univariate goodness of fit problem is treated first. Several distribution-free tests exist for this situation; Kolmogorov-Smirnov distance and the Chi squared test with equiprobable cells were selected as competitors for the $\tilde{S}_n$ test, which in one dimension becomes a test based on "spacings"; see Pyke (1965) for an overview. The results are shown in Table 1. Each entry is based on 200 trials. The uniform, standard normal and exponential with parameter 1 were the null densities used. Alternatives were chosen in such a way as to move in towards g according to particular $\{K_n\}$ sequences. Thus the sample powers should tend to stabilize for properly weighted tests and decrease for the unweighted test S_n. Alternative $f = f_4(x)$ (see Table 1) was studied in spite of its violation of Assumption A(iii). Values in parentheses represent the estimated powers of the test discussed in Section 3 which combines S_n and $\tilde{S}_n$. Finally, asterisks mark those entries for which the corresponding weight function is optimal.

Table 2 gives an indication of the power of $\tilde{S}_n$ for testing the multivariate standard normal against each of three alternatives. These include both location and scale departures as well as a mixture of the standard normal with a shifted normal distribution. The functions $w_5(x)$, $w_6(x)$ and $w_7(x)$ (see Table 2) represent the respective optimal weight functions for these alternatives. Again each case is based on 200 samples.

There is a distinct absence of competitors with similar properties to those of the multidimensional $\tilde{S}_n$ test. For testing the standard normal density, a simple procedure is to compute the squared distances from each point to the origin. Since under the null hypothesis these values follow a Chi squared distribution with m degrees of freedom, ordinary one-dimensional Kolmogrov-Smirnov distance can be used. This test is denoted by SR (squared radii) in Table 2.

The critical values used for the data in Table 2 are those of the asymptotic distributions of S_n and $\tilde{S}_n$ for $m = 1$, given in Schilling (1983), since these distributions fit the empirical null distributions ($m = 3, 5$) somewhat more closely than did the asymptotic distributions for $m = \infty$.

The power performance of the S_n test is rather disappointing, whereas the weighted test does quite well when the weight function is designed for a deviation from the null density at least roughly in the direction of the actual alternative. In one dimension the $\tilde{S}_n$ test with optimal weight function performs comparably to Kolmogorov-Smirnov distance and is superior to Chi square. The multidimensional version of $\tilde{S}_n$ having optimal weighting outperforms the SR test for both shift and mixture alternatives, and does about as well as the SR test for the scale alternative.

Examples in which $E_g\{w(X) \mid \lambda(X)\} \equiv 0$ include $f = f_2(x)$, $w = w_1(x)$ and $f = f_6(x)$, $w = w_2(x)$ in Table 1 and $f = f_9(x)$, $w = w_5(x)$ as well as $f = f_8(x)$, $w = w_6(x)$ or $w_7(x)$ in Table 2. As expected there is no apparent detection of the alternative by the weighted test in any of these cases; however, the entries of $f = f_3(x)$ or $f_4(x)$, $w = w_1(x)$ in Table 1 and $f = f_{10}(x)$, $w = w_8(x)$ in Table 2 give a measure of how well the $\tilde{S}_n$ test does against alternatives only crudely similar to that for which it has been designed. Thus in situations where the experimenter has at least a vague idea of the probable form an alternative to his hypothesized density may take, the $\tilde{S}_n$ test could provide a useful tool. He might also elect to use a test combining $\tilde{S}_n$ and S_n, as Tables 1 and 2 indicate that a fairly small price is paid in terms of loss of power in return for the assurance of consistency against all fixed alternatives.

Acknowledgment. The author is deeply indebted to Dr. Peter J. Bickel for his invaluable assistance and encouragement during the course of this research.

REFERENCES

BICKEL, P. J. and BREIMAN, L. (1983). Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test. *Ann. Probability* **11** 185–214.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

HÁJEK, J. and SĬDÁK, Z. (1967). *Theory of Rank Tests.* Academic Press, New York.

KAC, M. and SIEGERT, A. J. F. (1947). An explicit representation of a stationary Gaussian process. *Ann. Math. Statist.* **18** 438–442.

LECAM, L. (1960). Locally asymptotically normal families of distributions. *University of California Publ. in Statist.* **3** 37–98.

LECAM, L. (1971). Limits of experiments. *Proc. Sixth Berkeley Symposium on Statist. and Probab.* **1** 245–261.

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

PYKE, R. (1965). Spacings. *J. Roy. Statist. Soc. Ser. B.* **27** 395–436.

SCHILLING, M. F. (1983). An infinite-dimensional approximation for nearest neighbor goodness of fit tests. *Ann. Statist.* **11** 13–24.

SHORACK, G. (1973). Convergence of reduced empirical and quantile processes, etc. *Ann. Statist.* **1** 146–152.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SOUTHERN CALIFORNIA
UNIVERSITY PARK—MC 1113
LOS ANGELES, CA 90089-1113