The Annals of Statistics
1983, Vol. 11, No. 1, 13–24

AN INFINITE-DIMENSIONAL APPROXIMATION FOR NEAREST NEIGHBOR GOODNESS OF FIT TESTS[1]

BY MARK F. SCHILLING

University of Southern California

Let $X_1, \cdots, X_n$ be i.i.d. $\mathbb{R}^m$-valued observations from a bounded density $g(x)$ continuous on its support. Let $W_i = \exp\{-ng(X_i)V(R_i)\}$, $i = 1, \cdots, n$, where $V(R_i)$ is the volume of the nearest neighbor sphere around X_i, and let $w(x)$ be any bounded continuous weight function on $\mathbb{R}^m$. An infinite-dimensional approximation to the asymptotic form of the weighted empirical distribution function of the W_i's is presented. The distributions of quadratic functionals of the limiting normalized weighted e.d.f. are found and tabulated for $m = \infty$ and $m = 1$ and compared with finite $m > 1$. Monte Carlo results are given for $n, m < \infty$.

1. Introduction. Recently Bickel and Breiman (1983) have introduced goodness of fit tests for multidimensional densities based on quadratic functionals of the empirical distribution function of the variables

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

where $X_1, \cdots, X_n$ is a random sample from a bounded density in $\mathbb{R}^m$ continuous on its support (assumed open), $g(x)$ is the hypothesized density, $R_i = \min_{j \neq i} \| X_j - X_i \|$ is the distance from X_i to its nearest neighbor, and $V(r)$ is the volume of an m-sphere of radius r. When $g(x)$ is the true density the W_i's have an asymptotically uniform distribution. A more general version has been considered by the author (Schilling, 1983), which involves the *weighted* empirical process

$$F_n(t) = \frac{1}{n} \sum_{i=1}^n w(X_i) I(W_i \leq t), \qquad 0 \leq t \leq 1,$$

for some bounded continuous weight function $w(x)$, which can be chosen optimally for a prespecified sequence of contiguous alternatives to $g(x)$. Appropriately centered and scaled versions of these processes are shown in Bickel and Breiman (1983) and Schilling (1983), respectively, to converge weakly under the null hypothesis as n tends to infinity to zero mean Gaussian processes which are essentially independent of $g(x)$.

An obstacle to the implementation of these tests is the occurrence of a dimension-dependent term in the covariance kernel of the limiting processes which is intractable for $m > 1$. In this paper a simple approximating process is found for the case when $\| \cdot \|$ represents Euclidean distance by letting the dimension of the sample space tend to infinity. This rather unorthodox asymptotic technique is presented in Section 2. In Section 3 a representation for the null distribution of a quadratic functional of the approximating process is produced by the method of Kac and Siegert (1947); tables are given for two important weight functions. A corresponding tabulation for $m = 1$ is provided in Section 4 along with an examination of the adequacy of the $m = \infty$ and $m = 1$ distributions as approximations for situations in which $1 < m < \infty$. Section 5 provides some Monte Carlo

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 1970 *subject classifications.* Primary 62M99; secondary 62G10, 62H15, 62E20.

Key words and phrases. Nearest neighbor, infinite-dimensional approximation, empirical distribution function, quadratic functional.

comparisons for finite n and m. The main theoretical result is proven in Section 6. Finally, Section 7 gives details of the distributional calculations.

2. The $m = \infty$ approximation. Let

$$Z_n(t) = n^{1/2}\{F_n(t) - E_g F_n(t)\}, \qquad 0 \le t \le 1.$$

When $g(x)$ is the true density, the process $Z_n(t)$ converges weakly (see Bickel and Breiman, 1983, Schilling, 1983) to a Gaussian process $Z(t)$ with mean zero and covariance kernel

$$k(s, t) = \left[s(1 + t \log s) + st \int_{B(s,t)} \{\eta(s, t, \omega) - 1\}\, d\omega \right] E_g w^2(X_1)$$

$$- st(1 + \log st + \log s \log t)\{E_g w(X_1)\}^2, \qquad 0 \le s \le t \le 1,$$

where

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

and

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

where $r(\cdot)$ represents the radius of an m-sphere with volume $-\log(\cdot)$. The term

$$\int_{B(s,t)} \{\eta(s, t, \omega) - 1\}\, d\omega \tag{2.1}$$

can be transformed to a one-dimensional integral; however, unless $m = 1$ this integral is unmanageable—the integrand itself has an exponentiated term containing integrals which must be evaluated numerically for each s and t. This renders exact solutions to the integral equation (3.1)—a virtual necessity for solving the distribution problem—rather hopeless. It turns out, however, that the above term is rather small with respect to the other terms in $k(s, t)$, is relatively constant in m, and approaches a simple limit function. Notice that (2.1) is the only term in $k(s, t)$ which depends on m. The following theorem leads to a process which approximates $Z(t)$ in distribution for all m.

THEOREM 1.

$$\lim_{m\to\infty} \int_{B(s,t)} \{\eta(s, t, \omega) - 1\}\, d\omega = (\log s)(\log t) \qquad \forall s, t \in (0, 1]^2. \tag{2.2}$$

The rather lengthy proof is given in Section 6.

The function which results from applying Theorem 1 to $k(s, t)$ is the continuous limit of covariance kernels, hence is positive definite and indeed the covariance function of some Gaussian process. In fact a combination of familiar processes can be exhibited which has this covariance function: Let $W_0(t)$ be the Brownian bridge on [0, 1], let $W(t)$ be the Wiener process on $[0, \infty)$ and take Z_0 to be a standard normal variable; assume $W_0(t)$, $W(t)$ and Z_0 are mutually independent. Then it can be easily verified that

$$\{W_0(t) + tW(-\log t)\}(E_g w^2(X_1))^{1/2} + \{t(1 + \log t)Z_0\}(\mathrm{Var}_g w(X_1))^{1/2}$$

is a zero-mean Gaussian process with covariance kernel $\lim_{m\to\infty} k(s, t)$. The Brownian bridge occurs because the W_i's are asymptotically uniform on (0, 1], with the remaining components arising from the nearest neighbor dependency structure.

3. Tabulation of quadratic functionals by the Kac-Siegert method. A natural test based on $Z_n(t)$ is to reject when the value of $S_n = \int_0^1 Z_n^2(t)\, dt$ is large. By Donsker's

TABLE 1
Density and Distribution of S for m = ∞

	$w(x) \equiv 1$		$E_g w(X_1) = 0$ and $E_g w^2(X_1) = 1$	
y	$f(y)$	$F(y)$	$f(y)$	$F(y)$
.05	2.514	.036	.904	.011
.10	3.814	.212	2.016	.091
.15	3.072	.386	2.081	.196
.20	2.296	.519	1.881	.295
.25	1.719	.619	1.646	.383
.30	1.309	.694	1.429	.460
.35	1.015	.752	1.238	.526
.40	.799	.797	1.074	.584
.45	.637	.833	.934	.634
.50	.514	.861	.813	.678
.60	.343	.904	.619	.749
.70	.234	.932	.475	.803
.80	.163	.9516	.367	.845
.90	.115	.9653	.285	.878
1.00	.081	.9750	.222	.903
1.10	.058	.9819	.174	.923
1.20	.042	.9869	.138	.938
1.30	.030	.9905	.109	.9504
1.40	.022	.9930	.087	.9601
1.50	.016	.9949	.069	.9678
1.60	.012	.9963	.055	.9740
1.70	.008	.9973	.044	.9790
1.80	.006	.9980	.036	.9830
1.90	.005	.9985	.029	.9862
2.00	.003	.9989	.023	.9887
2.10	.002	.9992	.019	.9908
2.25	.002	.9995	.014	.9933
2.50	.0007	.9998	.008	.9959
3.00	.00016	.99995	.003	.9985
4.00			.0004	.99979
5.00			.00006	.99997

Theorem (Billingsley, 1968), S_n converges weakly under the null hypothesis to $S = \int_0^1 Z^2(t)\,dt$. The distributional theory of functionals such as S has been studied extensively. The technique established by Kac and Siegert (1947) (see also Anderson and Darling, 1952) shows that $Z(t)$ can be represented equivalently by

$$\sum_{j=1}^{\infty} \lambda_j^{1/2} \phi_j(t) Y_j,$$

where $\{\lambda_j, \phi_j(t); j = 1, 2, \cdots\}$ is the set of all solutions to the Fredholm integral equation

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

and $Y_1, Y_2, \cdots$ are independent standard normal variables. The eigenfunctions $\{\phi_j(t); j = 1, 2, \cdots\}$ form a complete orthonormal set, hence it is easily seen that

$$\sum_{j=1}^{\infty} \lambda_j Y_j^2 \tag{3.2}$$

has the same distribution as S, and only the eigenvalues $\lambda_1, \lambda_2, \cdots$ are needed to fully characterize S. All eigenvalues are greater then zero due to the positive definiteness of $k(s, t)$.

TABLE 2
Density and Distribution of S for m = 1

	$w(x) \equiv 1$		$E_g w(X_1) = 0$ and $E_g w^2(X_1) = 1$	
y	$f(y)$	$F(y)$	$f(y)$	$F(y)$
.05	2.652	.034	.934	.011
.10	4.473	.236	2.336	.101
.15	3.548	.439	2.405	.222
.20	2.538	.590	2.115	.335
.25	1.800	.697	1.789	.433
.30	1.291	.764	1.501	.515
.35	.941	.829	1.259	.584
.40	.695	.870	1.060	.642
.45	.520	.900	.896	.691
.50	.393	.923	.760	.732
.60	.230	.9529	.554	.797
.70	.139	.9709	.410	.845
.80	.085	.9819	.307	.880
.90	.053	.9886	.233	.907
1.00	.033	.9928	.178	.928
1.10	.021	.9954	.137	.943
1.20	.013	.9971	.106	.9553
1.30	.008	.9981	.083	.9647
1.40	.005	.9988	.065	.9721
1.50	.003	.9992	.051	.9778
1.60			.040	.9823
1.70			.032	.9859
1.75	.001	.9998	.028	.9874
1.80			.025	.9888
1.90			.020	.9910
2.00	.0004	.99991	.016	.9928
2.10			.013	.9942
2.25	.0001	.99997	.009	.9959
2.50			.005	.9976
3.00			.002	.9992
4.00			.0002	.99990
5.00			.00002	.99999

For general m, solutions to (3.1) are impracticable due to the complexity of (2.1); however, using the result of Theorem 1 a solution can be found for the case $m = \infty$ by using power series for the eigenfunctions. Closed form solutions to equations such as (3.1) are generally quite difficult or impossible to obtain.

Once the eigenvalues have been determined, the density and distribution function of S can be obtained by inverting the characteristic function of (3.2),

$$E(e^{iuS}) = \{\textstyle\prod_{j=1}^{\infty} (1 - 2i\lambda_j u)\}^{-1/2}. \tag{3.3}$$

Details on both the power series method for solving (3.1) and the inversion of (3.3) for $m = \infty$ are given in Section 7.

The distribution of S for given m depends only on the first two moments of $w(x)$. Table 1 gives the density and distribution function of S for $m = \infty$ for two important types of weight functions: $w(x) \equiv 1$ (the Bickel-Breiman test statistic), and any weight function with $E_g w(X_1) = 0$ and (without loss of generality) $E_g w^2(X_1) = 1$. The latter case is of

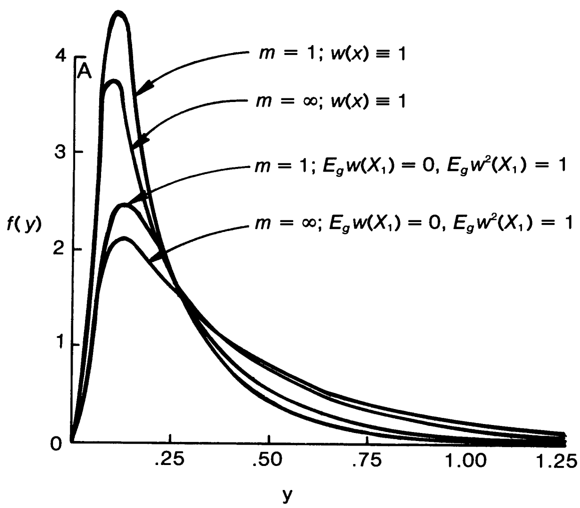

FIG. 1a. *Densities of S for m* = 1, ∞.

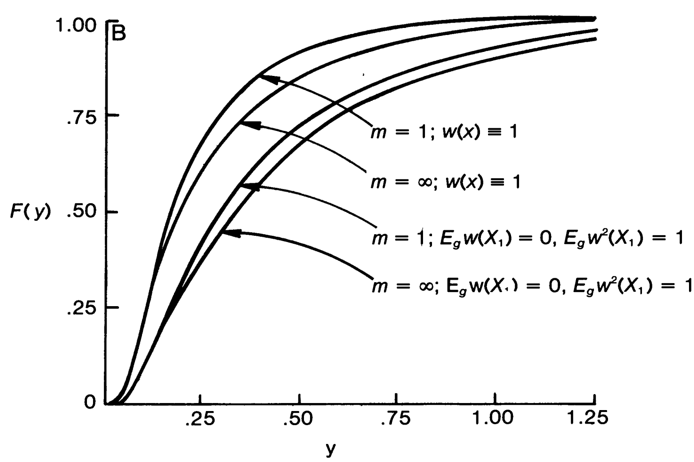

FIG. 1b. *Distributions of S for m* = 1, ∞.

interest because the optimal weight function against any prespecified contiguous sequence of alternatives to $g(x)$ is of this form (see Schilling, 1983).

4. The case $m = 1$ and a comparison with $1 < m < \infty$.

4.1. *The case* $m = 1$. The only finite value of m for which (2.1) takes a simple form is $m = 1$. We then have

$$\int_{B(s,t)} \{\eta(s, t, \omega) - 1\} \, d\omega = \log t + 2t^{-1/2} - 2,$$

and the Fredholm equation (3.1) may be solved with power series in a similar fashion to the $m = \infty$ case.

A tabulation of the density and distribution function of S for $m = 1$ is given in Table 2 for the same two weight functions used in Table 1. The $m = 1$ and $m = \infty$ densities and distributions are compared graphically in Figures 1a and 1b.

TABLE 3
Means of S for selected m and w(x)

	$m = 1$	3	5	7	9	∞
$w(x) \equiv 1$	.226	.247	.258	.264	.268	.278
$E_g w(X_1) = 0$ and $E_g w^2(X_1) = 1$	.411	.432	.443	.449	.453	.463

4.2. *Comparison with* $1 < m < \infty$. The primary usefulness of the S_n test is for (finite) values of m greater than one, for which very few distribution-free competitors exist. The distributions given in Tables 1 and 2 are of value only if they are close to the distributions of S for such m. While the convergence indicated in Theorem 1 has not been proved to be monotone, there are strong indications that this is the case.

After some manipulations, the following closed form for (2.1) can be obtained for the simplest case, $s = t$, m odd:

$$(4.1)\quad \int_{B(t,t)} \{\eta(t, t, \omega) - 1\}\, d\omega = -\log t \int_1^2 \exp\left[\frac{2K_m}{K_{m-1}} \sum_{j=0}^{(m-1)/2} \binom{(m-1)/2}{j} \frac{(-1)^j}{2j+1} \cdot \{1 - (u/2)^{2j+1}\}\log t\right] mu^{m-1}\, du - (2^m - 1)\log t,$$

where K_m denotes the volume of an m-sphere with radius 1. For m even, the upper limit of the summation is replaced by ∞.

A numerical integration program was used to evaluate (4.1); the resulting figures were substituted into $k(t, t)$ to produce values of Var $Z(t)$ for selected m and t. For each t these quantities increased monotonically in m, converging quite rapidly to the $m = \infty$ values. This convergence is indicated in Table 3, which contains estimated means of S for selected m and $w(x)$, obtained from the formula (see Section 7) $ES = \int_0^1 \text{Var}\, Z(t)$.

5. Monte Carlo results for $n, m < \infty$. A Monte Carlo examination of the S_n test for small to moderate sample sizes and numbers of dimensions suggests that the S distributions given in Tables 1 and 2 are reasonably adequate for actual experiments.

Samples of size 25 and 100 were generated from selected densities in $\mathbb{R}^m$ for $m = 1$, 3 and 5. Although data sets of 100 or less points might not generally be considered adequate for assessing models in as many as five dimensions, the trade-off between sample size and cost deemed investigation of larger samples unprofitable. Furthermore, given the asymptotic behavior of any statistical procedure, one generally has a better notion of how the procedure will perform for large samples based on its behavior for small samples than vice versa.

Figure 2a gives a comparison of the estimated upper tails of the unweighted ($w(x) \equiv 1$) S_n distribution (the primary region of interest for testing hypotheses) with those of the S distributions for $m = 1$ and $m = \infty$, for samples of size 100 from the multivariate standard normal density in 1, 3 and 5 dimensions. 1000 samples were used for $m = 1$, 200 each for $m = 3$ and 5.

In Figure 2b a corresponding comparison is given for weighted S_n test statistics with weight functions satisfying $E_g w(X_1) = 0$, $E_g w^2(X_1) = 1$. For $m = 1$, the optimal weight function for contiguous normal shift alternatives ($w(x) = x$) was used. For $m = 3$ and 5 three different weight functions were tried with 200 replications of each; one optimal against location alternatives as for $m = 1$, one optimal for scale alternatives and one optimal against a mixture of two multivariate normal distributions with common covariance matrix **I** but different means. Figure 2b plots averages for these 600 test samples.

The centering quantity $E_g F_n(t)$ of S_n is not in general computable exactly. It is shown

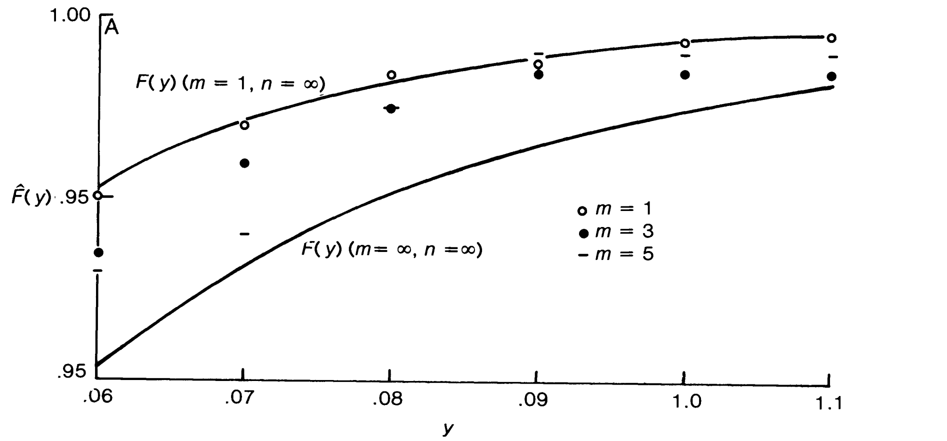

FIG. 2a. *Monte Carlo upper tail values of S_n distribution for $w(\boldsymbol{x}) \equiv 1$; $g(x) = N(\boldsymbol{O}, \boldsymbol{I})$; $m = 1, 3, 5$; $n = 100$.*

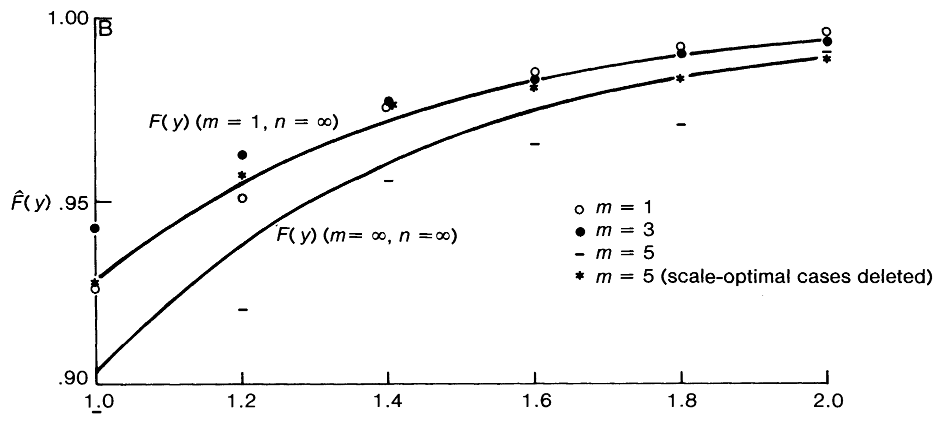

FIG. 2b. *Monte Carlo upper tail values of S_n distribution for $E_g w(\boldsymbol{X}_1) = 0$, $E_g w^2(\boldsymbol{X}_1) = 1$; $g = N(\boldsymbol{O}, \boldsymbol{I})$; $m = 1, 3, 5$; $n = 100$.*

in Schilling (1979) that $E_g F_n(t) = tE_g w(X_1) + R_n(t)$, with

$$R_n(t) = c(t, m)E_g\left[\frac{w(X_1)\operatorname{tr}\{g(X_1)\}}{\{g(X_1)\}^{1+2/m}}\right]n^{-2/m} + O(n^{-4/m}),$$

where c depends only on t and m, and $\operatorname{tr}\{g(X_1)\}$ represents the trace of the Hessian matrix of g, evaluated at X_1. For the case when g is the multivariate standard normal density, $\operatorname{tr}\{g(X_1)\}/g(X_1) = \sum_{i=1}^m (X_{1i}^2 - 1)$, which corresponds directly with the optimal weight function for scale alternatives. Hence $R_n(t)$ is rather large in this case, leading to larger values for S_n. This is evident in Figure 2b, where the deletion of samples using the scale-optimal weight function yields a marked improvement in fit to the asymptotics.

The sample S_n distributions matched the corresponding S distributions equally well in the lower portions of their range. The S_n distributions for samples of size 25 also followed the asymptotic distributions fairly well on the whole considering the smallness of n. The best fits were for $w(x) \equiv 1$.

Other densities and weight functions were tried for $\mathbb{R}^1$ with very similar results. Complete information on all Monte Carlo experiments performed may be found in Schilling (1979).

REMARK. Caution should be exercised when using Tables 1 and 2 for certain high-dimensional models due to possible boundary effects. For example, the multivariate uniform distribution will not provide a close fit to the asymptotics for $m > 3$ without many thousands of sample points.

6. Proof of Theorem 1. Expression (2.2) clearly holds when $s = 1$ or $t = 1$. For given m, s and t such $0 < s \le t < 1$, let $w = \| \omega \|$ and transform to spherical coordinates to obtain for (2.1)

$$(6.1) \qquad mK_m \int_{r(s)}^{r(s)+r(t)} w^{m-1}\{\eta(s, t, (w, 0, \cdots, 0)) - 1\}\, dw$$

where $K_m = \pi^{m/2}/\Gamma(m/2 + 1)$ is the volume of a sphere in $\mathbb{R}^m$ with radius 1. The factor mK_m is the surface area of such a sphere. To evaluate (6.1) the method for computing volumes by revolution may be employed to yield

$$(6.2) \qquad \begin{aligned} \log \eta(s, t, (w, 0, \cdots, 0)) = K_{m-1} &\int_{(r^2(s)-r^2(t)+w^2)/2w}^{r(s)} \{r^2(s) - z^2\}^{(m-1)/2}\, dz \\ &+ K_{m-1} \int_{(r^2(t)-r^2(s)+w^2)/2w}^{r(t)} \{r^2(t) - z^2\}^{(m-1)/2}\, dz. \end{aligned}$$

The first term in (6.2) equals

$$\frac{K_{m-1}}{K_m} (-\log s) \int_{(r^2(s)-r^2(t)+w^2)/2wr(s)}^{1} (1 - z^2)^{(m-1)/2}\, dz \le \frac{K_{m-1}}{K_m} (-\log s) \int_{1/2}^{1} (1 - z^2)^{(m-1)/2}\, dz$$

upon minimizing $(r^2(s) - r^2(t) + w^2)/2wr(s)$ in $(r(s), r(s) + r(t))$. Now

$$\frac{K_{m-1}}{K_m} = \frac{\Gamma\left(\frac{m}{2} + 1\right)}{\pi^{1/2}\Gamma\left(\frac{m+1}{2}\right)} \sim (m/2\pi)^{1/2}$$

by Stirling's approximation; since

$$\int_{1/2}^{1} (1 - z^2)^{(m-1)/2} < \left(\frac{3}{4}\right)^{m(-1)/2},$$

the first term in (6.2) tends to zero uniformly in w. A similar argument works for the second term of (6.2). Hence $\log \eta(s, t, (w, 0, \cdots, 0)) \to 0$ uniformly in w. Combining this result with the inequality $x \le e^x - 1 \le xe^x$ produces

$$\begin{aligned} &\lim_{m\to\infty} \int_{B(s,t)} \{\eta(s, t, \omega) - 1\}\, d\omega \\ &= \lim_{m\to\infty} mK_m \int_{r(s)}^{r(s)+r(t)} w^{m-1}\{\log \eta(s, t, (w, 0, \cdots, 0))\}\, dw \end{aligned}$$

$$(6.3) \qquad \begin{aligned} &= \lim_{m\to\infty} K_{m-1}\Big\{ (-\log s) \int_{r(s)}^{r(s)+r(t)} mw^{m-1} \int_{(r^2(s)-r^2(t)+w^2)/2wr(s)}^{1} (1 - z^2)^{(m-1)/2}\, dz\, dw \\ &\quad + (-\log t) \int_{r(s)}^{r(s)+r(t)} mw^{m-1} \int_{(r^2(t)-r^2(s)+w^2)/2wr(t)}^{1} (1 - z^2)^{(m-1/2}\, dz\, dw\Big\}. \end{aligned}$$

Putting $a = w/r(s)$ in the first term and $a = w/r(t)$ in the second, then using integration by parts on each double integral, (6.3) becomes

$$\lim_{m\to\infty} \frac{K_{m-1}}{K_m}\left\{\log^2 s\left[-\int_{1-(r(t)/r(s))^2/2}^{1} (1-z^2)^{(m-1)/2}\,dz\right.\right.$$

$$+\left.\int_1^{1+r(t)/r(s)} \left\{1-\left[\frac{1-(r(t)/r(s))^2+a^2}{2a}\right]^2\right\}^{(m-1)/2}\left\{\frac{(r(s)/r(t))^2-1}{2a^2}+\frac{1}{2}\right\}a^m\,da\right]$$

(6.4) $$+\log^2 t\left[-(r(s)/r(t))^m\int_{r(t)/2r(s)}^{1}(1-z^2)^{(m-1)/2}\,dz\right.$$

$$\left.\left.+\int_{r(s)/r(t)}^{1+r(s)/r(t)}\left\{1-\left[\frac{1-(r(s)/r(t))^2+a^2}{2a}\right]^2\right\}^{(m-1)/2}\left\{\frac{(r(t)/r(s))^2-1}{2a^2}+\frac{1}{2}\right\}a^m\,da\right]\right\}.$$

The terms $\int (1-z^2)^{(m-1)/2}\,dz$ in (6.3) are again less than $(3/4)^{(m-1)/2}$ in magnitude and therefore contribute nothing to the limit.

For the remainder of the proof the property

$$\lim_{m\to\infty} r(t)/r(s) = 1$$

is employed. Using this fact and making the substitution $u = a^2/4$ for each integral, (6.4) can be seen to be equivalent in the limit to

(6.5) $$\lim_{m\to\infty}(m/2\pi)^{1/2}2^{m-1}\left(\log^2 s\int_{1/4}^{(1+r(t)/r(s))^2/4}\left\{1-\left[\frac{1-(r(t)/r(s))^2+4u}{4u^{1/2}}\right]^2\right\}^{(m-1)/2} u^{(m-1)/2}\,du\right.$$

$$\left.+\log^2 t\int_{(r(s)/r(t))^2/4}^{(1+r(s)/r(t))^2/4}\left\{1-\left[\frac{1-(r(s)/r(t))^2+4u}{4u^{1/2}}\right]^2\right\}^{(m-1)/2} u^{(m-1)/2}\,du\right).$$

The limits of each integral can be replaced by zero and one; this can be seen by verifying that for sufficiently large m each integrand is nonnegative and less than some constant times $u^{(m-1)/2}(1-u)^{(m-1)/2} \sim o(b^m)$ for some $b < 1/2$ when u is bounded away from $1/2$. In fact one can express (6.5) as an expectation against a beta density of the form $u^{(m-1)/2}(1-u)^{(m-1)/2}$. Observe that the normalizing constant for such a density is

$$\frac{\left\{\Gamma\left(\frac{m+1}{2}\right)\right\}^2}{\Gamma(m+1)} \sim (2\pi/m)^{1/2}2^{-m},$$

using Stirling's approximation again. Consequently, (6.5) is equivalent to

(6.6) $$\tfrac{1}{2}\lim_{m\to\infty}(\log^2 s)EV_1^{(m)} + \tfrac{1}{2}\lim_{m\to\infty}(\log^2 t)EV_2^{(m)},$$

where

$$V_i^{(m)} = f_i^{(m)}(U_m), \qquad i = 1, 2$$

with

$$U_m \sim \text{Beta}\left(\frac{m+1}{2}, \frac{m+1}{2}\right),$$

$$f_1^{(m)}(u) = \{(1-[1-\{r(t)/r(s)\}^2+4u]^2/16u)/(1-u)\}^{(m-1)/2},$$

$$f_2^{(m)}(u) = \{(1-[1-\{r(s)/r(t)\}^2+4u]^2/16u)/(1-u)\}^{(m-1)/2}.$$

Note that $\{U_m\} \to_p \tfrac{1}{2}$. The limiting behavior of $f_1^{(m)}(u)$ and $f_2^{(m)}(u)$ must be found. Some algebraic manipulation produces

$$f_1^{(m)}(u) = \left\{1 - \frac{1 - p^{1/m}}{2(1-u)} - \frac{(1-p^{1/m})^2}{16u(1-u)}\right\}^{(m-1)/2},$$

where $p = \{r(t)/r(s)\}^m = \log^2 t/\log^2 s$. It is enough to check convergence behavior for $\log f_1^{(m)}(u)$. Since $p^{1/m} \to 1$, $\log f_1^{(m)}(u)$ can be expanded to

$$\frac{-(m-1)}{2}\left\{\frac{1-p^{1/m}}{2(1-u)} + \frac{(1-p^{1/m})^2}{16u(1-u)}\right\} + R_m(u), \tag{6.7}$$

where R_m is of uniformly smaller order than the first term in (6.7) for u bounded away from zero and one. Upon using L'Hôpital's rule for (6.7) we find that $f_1^{(m)}(u)$ converges uniformly in a neighborhood of $\tfrac{1}{2}$ to

$$f_1(u) = p^{1/4(1-u)} = (\log t/\log s)^{1/2(1-u)}.$$

The function $f_1(u)$ is continuous at $u = \tfrac{1}{2}$ and equals $\log t/\log s$ there.

The same arguments applied to $f_2^{(m)}(u)$ produce a limiting function $f_2(u) = 1/f_1(u)$ with the same convergence and continuity properties. Using Corollary 2 to Theorem 5.1 of Billingsley (1968) we obtain

$$V_1^{(m)} \to_p \log t/\log s, \qquad V_2^{(m)} \to_p \log s/\log t.$$

It has been mentioned in the arguments following (6.5) that the tail contributions to $EV_1^{(m)}$ and $EV_2^{(m)}$ are asymptotically negligible, i.e., the sequences $\{V_1^{(m)}\}$, $\{V_2^{(m)}\}$ are each uniformly integrable. Thus (Billingsley, 1968, Theorem 5.4) $\{EV_1^{(m)}\}$ and $\{EV_2^{(m)}\}$ have the same respective limits as $\{V_1^{(m)}\}$, $\{V_2^{(m)}\}$. Upon insertion of these limits into (6.6), the theorem is proved.

7. Distribution of S.

7.1. *Solution of the Fredholm equation* (3.1) *for* $m = \infty$. The kernel $k(s, t)$ is singular at 0 but analytic in (0, 1]; this suggests a power series expansion around the point $t = 1$. It is not difficult to show that all eigenfunction-eigenvalue pairs are generated by this method.

Successive differentiations of (3.1) yields the following set of equations for $\phi(t)$ and λ:

$$\begin{aligned} &t^{-1}\int_0^t s\phi(s)\,ds + \int_0^t s(\log s)\phi(s)\,ds \\ &\qquad + \int_t^1 \phi(s)\,ds + \log t\int_t^1 s\phi(s)\,ds \\ &\qquad + \log t\int_0^1 s(\log s)\phi(s)\,ds = \lambda\phi(t)/t, \end{aligned} \tag{7.1}$$

$$-\int_0^t s\phi(s)\,ds + t\int_0^1 s\phi(s)\,ds + t\int_0^1 s(\log s)\phi(s)\,ds = \lambda(t\phi'(t) - \phi(t)), \tag{7.2}$$

$$\int_0^t s\phi(s)\,ds - t^2\phi(t) - t^3\phi'(t) = \lambda(t^2\phi''(t) - t\phi'(t) + \phi(t)), \tag{7.3}$$

$$\begin{aligned} &\lambda(t\phi^{(n+3)}(t) + (n+1)\phi^{(n+2)}(t)) + (t^2 + t)\phi^{(n+1)}(t) \\ &\qquad + \{(n+1)(1+2t) + t\}\phi^{(n)}(t) \\ &\qquad + n(n+2)\phi^{(n-1)}(t) = 0, \qquad n = 0, 1, 2, \cdots, \end{aligned} \tag{7.4}$$

where $\phi^{(-1)}(t) \equiv 0$, $\phi^{(0)}(t) = \phi(t)$, and $\phi^{(n)}(t) = d^n\phi(t)/dt^n$ for $n \geq 1$. It is helpful to let $u =$

$1 - t$ and $\phi(t) = \psi(u) = \sum_{n=0}^{\infty} a_n u^n$, where $a_n = (-1)^n \phi^{(n)}(1)/n!$ for each n. A recursive formula for the a_n's can then be obtained from (7.4) for $n \geq 3$. It can be shown that a power series solution to (7.4) is convergent on $|u| < 1$ regardless of $\lambda > 0$. Series expansion of the log function and term by term integration yield for (7.1)–(7.3), respectively,

$$a_1 = \frac{2}{\lambda} \sum_{m=1}^{\infty} \frac{1}{m} \sum_{n=0}^{\infty} a_n/\{(m+n+1)(m+n+2)\}, \tag{7.5}$$

$$a_0 = -a_1/2 + \frac{1}{\lambda} \sum_{n=0}^{\infty} a_n/\{(n+1)(n+2)\}, \tag{7.6}$$

$$a_2 = -a_0/\lambda - a_1/4. \tag{7.7}$$

Consequently the eigenvalues of (3.1) are precisely those values of λ which yield a simultaneous solution to the recursive equation and (7.5)–(7.7). A possible approach to obtaining these is to try initial values of a_0, a_1 and λ and then iterate these quantities. Since $\phi(t)$ in (3.1) is determined only up to a multiplicative constant, just two unknowns are involved.

An algorithm for this procedure was successfully employed with the aid of quite accurate initial values, which we obtained by considering a discrete analogue to (3.1), namely the determination of eigenvalues and eigenvectors for the $n \times n$ matrix $(1/n)\{K_{ij}\}$ where

$$K_{ij} = K_{ji} = t_i(1 + t_j \log t_i) + t_i t_j \log(t_i)\log(t_j), \qquad 1 \leq i \leq j \leq n,$$

with $t_i = (i - \frac{1}{2})/n$, $i = 1, \cdots, n$. This was solved with an IMSL subroutine for $n = 100$. The five largest eigenvalues obtained from the power series convergence algorithm agreed with those for the discrete problem to within a relative error of 0.000003; consequently the remaining eigenvalues of the matrix were taken as correct for (3.1).

The acceptability of the substitution of eigenvalues from the discrete problem for those of the continuous case can be measured by an independent check on the moments of S. From Anderson and Darling (1952, page 200) we have

$$ES = \int_0^1 k(t, t)\, dt, \qquad \operatorname{Var} S = 2 \int_0^1 \int_0^1 k^2(s, t)\, ds\, dt, \tag{7.8}$$

whereas from (3.2) we obtain

$$ES = \sum_{j=1}^{\infty} \lambda_j, \qquad \operatorname{Var} S = 2 \sum_{j=1}^{\infty} \lambda_j^2. \tag{7.9}$$

The errors in using the 100 estimated eigenvalues for (7.9) as compared with the exact moments computed from (7.8) were 0.000004 and 0.000028 for ES and $\operatorname{Var} S$ respectively.

7.2 *Inverting the characteristic function.* It is assumed subsequently that $\lambda_1 \geq \lambda_2 \geq \cdots$. Seventy eigenvalues were retained for inverting the characteristic function of S. That is, we are actually finding the distribution of $S^{(70)} = \sum_{j=1}^{70} \lambda_j Y_j^2$. The density function of $S^{(70)}$ is

$$f_{S^{(70)}}(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{iuy} \{D(2iu)\}^{-1/2}\, du$$

where

$$D(y) = \prod_{j=1}^{70} (1 - \lambda_j y).$$

A procedure for complex integration due to Slepian (1958)—although the form first appears in Smirnov (1937)—yields the computational form

$$f_{S^{(70)}}(y) = \frac{1}{2\pi} \sum_{k=1}^{35} (-1)^k I_k,$$

where

$$I_k = \int_{1/2\lambda_{2k-1}}^{1/2\lambda_{2k}} e^{-yu} \,|\, D(2u)|^{-1/2}\, du, \qquad k = 1, 2, \cdots, 35.$$

It is easily seen (Fubini's Theorem) that the distribution function of $S^{(70)}$ can be expressed as

$$F_{S^{(70)}}(y) = 1 - \pi^{-1} \textstyle\sum_{k=1}^{35} (-1)^k J_k,$$

where

$$J_k = \int_{1/2\lambda_{2k-1}}^{1/2\lambda_{2k}} |\, D(2u)|^{-1/2} u^{-1}\, du, \qquad k = 1, 2, \cdots, 35.$$

The I_k's and J_k's were found by numerical integration, and for each value of y used decreased rapidly with k.

The distribution of $S^{(70)}$ clearly provides an upper bound for that of S. A lower bound is of greater importance since it allows conservative level α tests based on S_n, at least asymptotically. Letting $R^{(70)} = \sum_{j=71}^{\infty} \lambda_j Y_j^2$ we have, for any $\varepsilon > 0$,

$$F_S(y) \geq F_{S^{(70)}}(y - \varepsilon) F_{R^{(70)}}(\varepsilon). \tag{7.10}$$

A simple procedure due to Blum, Kiefer and Rosenblatt (1961) yields

$$F_{R^{(70)}}(\varepsilon) \geq 1 - \exp[-\{\varepsilon/\mu_R - \log(\varepsilon/\mu_R) - 1\}\mu_R/2\lambda_{71}] \tag{7.11}$$

where

$$\mu_R = ER^{(70)} = \textstyle\sum_{j=71}^{\infty} \lambda_j.$$

Upon choosing ε appropriately, it results that the lower bound for $F_S(y)$ obtained from (7.10) and (7.11) is never more than 0.006 below the tabled value; for the upper tail (of greatest interest for testing) the difference is much less. For example, using the 95th percentile of the lower bound as critical value will result in a true asymptotic level between 0.0489 and 0.0500.

Acknowledgment. The author is most grateful to Professor P. J. Bickel for his guidance and valuable advice during this research.

REFERENCES

ANDERSON, T. W. and DARLING, D. A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *Ann. Math. Statist.* **23** 193–212.

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.

BLUM, J. R., KIEFER, J. and ROSENBLATT, M. (1961). Distribution-free tests of independence. *Ann. Math. Statist.* **32** 485–498.

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

SCHILLING, M. F. (1979). Testing for goodness of fit based on nearest neighbors. Ph.D. dissertation, University of California at Berkeley.

SCHILLING, M. F. (1983). Goodness of fit testing in $\mathbb{R}^m$ based on the weighted empirical distribution of certain nearest neighbor statistics. *Ann. Statist.* **11** 1–12.

SLEPIAN, D. (1958). Fluctuations of random noise power. *Bell System Tech. J.* **37** 163–184.

SMIRNOV, N. V. (1937). On the distribution of the ω^2-criterion. *Mat. Sbornik (NS)* **2** 973–993.

DEPARTMENT OF MATHEMATICS/DRB 250
UNIVERSITY OF SOUTHERN CALIFORNIA
UNIVERSITY PARK-MC 1113
LOS ANGELES, CALIFORNIA 90089-1113