The Annals of Statistics
1995, Vol. 23, No. 5, 1594–1608

CONSISTENCY AND MONTE CARLO SIMULATION OF A DATA DRIVEN VERSION OF SMOOTH GOODNESS-OF-FIT TESTS

BY WILBERT C. M. KALLENBERG AND TERESA LEDWINA[1]

University of Twente and Technical University of Wrocław

The data driven method of selecting the number of components in Neyman's smooth test for uniformity, introduced by Ledwina, is extended. The resulting tests consist of a combination of Schwarz's Bayesian information criterion (BIC) procedure and smooth tests. The upper bound of the dimension of the exponential families in applying Schwarz's rule is allowed to grow with the number of observations to infinity.

Simulation results show that the data driven version of Neyman's test performs very well for a wide range of alternatives and is competitive with other recently introduced (data driven) procedures. It is shown that the data driven smooth tests are consistent against essentially all alternatives. In proving consistency, new results on Schwarz's selection rule are derived, which may be of independent interest.

1. Introduction. Let $X_1, X_2, \dots, X_n$ be i.i.d. r.v.'s. Consider the goodness-of-fit problem of testing the simple null hypothesis H_0 that the X_i's have distribution function F_0, where F_0 is a given continuous distribution function. Without loss of generality, assume that under H_0 the distribution of X_i is the uniform distribution on $[0, 1]$.

The so-called smooth test statistics [cf., e.g., Rayner and Best (1989)] form a well-known class of test statistics for testing H_0. They are given by

$$T_k = \sum_{j=1}^{k} \left\{ n^{-1/2} \sum_{i=1}^{n} \phi_j(X_i) \right\}^2, \qquad k = 1, 2, \dots, \tag{1.1}$$

where $\phi_0, \phi_1, \dots,$ is an orthonormal system in $L_2([0,1])$ with $\phi_0(x) \equiv 1$. Choosing for $\{\phi_j\}$ the orthonormal Legendre polynomials on $[0, 1]$, one gets the test introduced and investigated by Neyman (1937). Further on we refer to this test as Neyman's test with test statistic N_k. Another member of the class of special interest is obtained by taking $\phi_j(x) = \sqrt{2}\cos(j\pi x)$, the cosine system. This orthonormal system is used in the data driven testing procedure of Eubank and LaRiccia (1992) and plays a role in Bickel and Ritov (1992).

Received September 1993; revised October 1994.
[1]Research supported by Grant KBN 665/2/91.
AMS 1991 *subject classifications*. 62G10, 62E25, 62G20.
Key words and phrases. Goodness-of-fit, smooth test, Neyman's test, Schwarz's BIC criterion, exponential family, Monte Carlo power.

Recommendations in the literature for choosing the number k of components in (1.1) are often inconsistent with each other. An extensive discussion on this is given in Inglot, Kallenberg and Ledwina (1994), where also a criterion is developed for selecting k.

Recently, Ledwina (1994) introduced a new way of selecting k in Neyman's smooth test of uniformity. Unlike earlier proposals, depending on alternatives of special interest, the new procedure provides an automatic choice of k, based on the data. Roughly speaking it works as follows. First, Schwarz's (1978) selection rule is applied to find a suitable dimension S, say, of an exponential family model for the data. Then Neyman's test is applied within the fitted model, resulting in the test statistic N_S. So Schwarz's rule serves as a kind of first selection, followed by the more precise instrument, being Neyman's test in the "right" dimension.

There is now a lot of interest in this kind of procedure as is seen in the papers of Bickel and Ritov (1992), Eubank and Hart (1992), Eubank and LaRiccia (1992) and Eubank, Hart and LaRiccia (1993).

In this paper new simulation results show that the test N_S works very well for a wide range of alternatives. In most of the considered cases in the extensive simulation study, the data driven version of Neyman's test is competitive with the new tests of Bickel and Ritov (1992) and Eubank and LaRiccia (1992).

Theoretical support of the new test is obtained by proving its consistency. The results on N_S, given in Ledwina (1994), are extended in several ways. First of all, the fixed upper bound of the dimension in applying Schwarz's selection rule is replaced by the more realistic upper bound $d(n)$, which may tend to infinity as $n \to \infty$. Second, consistency at essentially all alternatives is proved. Finally, general orthonormal systems are considered.

The new results on Schwarz's selection rule in this paper may be of independent interest. They include an analysis for sequences of exponential family models with dimension $d(n)$ tending to infinity as $n \to \infty$.

As noted by Eubank and LaRiccia [(1992), page 2072], there are numerous empirical studies, where smooth tests have been shown to be more powerful than common omnibus test statistics over a wide range of realistic alternatives.

In view of the (simulation) results here and the earlier (simulation) results presented in Ledwina (1994), where a comparison was made with widely recommended tests proposed by Anderson and Darling (1952), Watson (1961) and Neuhaus (1988), the conclusion of Rayner and Best [(1990), page 9], "don't use those other methods—use a smooth test!," may be slightly sharpened to "use a data driven smooth test."

Extension of the method to goodness-of-fit problems with nuisance parameters will be discussed elsewhere.

2. Definition of the test statistics. In this section the test statistic introduced in Ledwina (1994) is extended in two ways. The upper bound of the dimension of the exponential families in applying Schwarz's selection rule is not fixed, but may tend to infinity as $n \to \infty$. Second, general orthonormal systems are allowed. While the second generalization is rather obvious, the first one leads to essential new problems and results.

To define the test statistic, consider exponential families generated by the uniform distribution and an orthonormal system

$$\phi_0, \phi_1, \phi_2, \dots$$

with bounded functions $\phi_1, \phi_2, \dots$ and $\phi_0(x) \equiv 1$. The functions $\phi_1, \phi_2, \dots$ are not necessarily uniformly bounded. For $k = 1, 2, \dots,$ the exponential families are defined by their densities $p_\theta(x)$ with respect to Lebesgue measure on $[0, 1]$ of the form

$$p_\theta(x) = \exp\{\theta \circ \phi(x) - \psi_k(\theta)\}, \tag{2.1}$$

where

$$\begin{aligned} \theta &= (\theta_1, \dots, \theta_k), \qquad \phi = (\phi_1, \dots, \phi_k), \\ \psi_k(\theta) &= \log \int_0^1 \exp\{\theta \circ \phi(x)\}\, dx \end{aligned} \tag{2.2}$$

and $\circ$ stands for the inner product in $\mathbb{R}^k$. Since the functions ϕ_j are bounded, p_θ is defined for every $\theta \in \mathbb{R}^k$. When there is no confusion, the dimension k is sometimes suppressed in the notation.

Let $\lambda(\theta) = E_\theta \phi(X)$. It is well known that

$$\lambda(\theta) = \psi'(\theta) \tag{2.3}$$

with the prime denoting derivative. Moreover, by orthonormality

$$\lambda(0) = 0, \qquad \psi''(0) = I, \quad \text{the identity matrix.} \tag{2.4}$$

Writing

$$\begin{aligned} Y_n &= (\bar{\phi}_1, \dots, \bar{\phi}_k), \qquad \bar{\phi}_j = n^{-1} \sum_{i=1}^{n} \phi_j(X_i), \\ L_k &= n \sup_{\theta \in \mathbb{R}^k} \{Y_n \circ \theta - \psi_k(\theta)\} - \tfrac{1}{2} k \log n, \end{aligned} \tag{2.5}$$

Schwarz's (1978) Bayesian information criterion (BIC) for choosing submodels corresponding to successive dimensions yields

$$S = \min\{k : 1 \le k \le d(n), L_k \ge L_j, j = 1, \dots, d(n)\}. \tag{2.6}$$

Although it is not mentioned in the notation, S depends of course on the upper bound $d(n)$ of the exponential families under consideration. The data driven smooth test statistic is defined by

$$T_S = \sum_{j=1}^{S} \left\{ n^{-1/2} \sum_{i=1}^{n} \phi_j(X_i) \right\}^2 \tag{2.7}$$

with S given by (2.6). The null hypothesis is rejected for large values of T_S.

As is seen from (2.6) and (2.7), first an exponential family model is fitted to the data and then the asymptotic local (θ close to 0) optimal solution of the testing problem $H_0: \theta = 0$ (corresponding to uniformity) against $\theta \neq 0$ is applied. More information on Schwarz's BIC is found in Schwarz (1978), Haughton (1988), Rissanen (1983, 1987), Barron and Cover (1991), Ledwina (1994) and in the following sections of this paper.

While some authors [cf. Eubank, Hart and LaRiccia (1993) and references mentioned there] include dimension 0 (the null hypothesis) as a candidate dimension, others [e.g., Bickel and Ritov (1992)] start from dimension 1. We prefer in our case the latter approach, because of the following reasons.

First of all, we like to select k in the class given by (1.1) and k simply starts with 1. From this point of view, S is only a (suitable) way to do the selection and consequently it should start with 1.

Second, the idea behind (2.7) is that Schwarz's rule gives a first indication about the true density of the observations while the finishing touch comes from the smooth test in the selected exponential family. The lowest dimension, where for the second step a testing problem can be formulated, equals 1.

Numerical results in, for example, Inglot, Kallenberg and Ledwina (1994) show that Neyman's test performs very well, provided a good choice of k has been made. Therefore, the data driven version of Neyman's smooth test (denoted by N_S) is especially considered here. Examples in Table 2 of Inglot, Kallenberg and Ledwina (1994) and the examples in Tables 3–5 in this paper show that a considerable loss of power may occur when a wrong choice of k is made. This illustrates that a good procedure for choosing k based on the data is very welcome.

3. Asymptotic behavior of S and T_S under H_0. To prove consistency of T_S, we need information on the behavior of T_S both under H_0 and the (fixed) alternative. Here we consider the null case and start with the asymptotic behavior of S, which may be of independent interest.

In view of (2.6) we have (P_0 denotes that X_i is uniformly distributed on $[0,1]$)

$$P_0(S = 1) = 1 - \sum_{k=2}^{d(n)} P_0(S = k) \tag{3.1}$$

and, using $\psi_1(0) = 0$,

$$\begin{aligned} P_0(S = k) &\le P_0(L_k \ge L_1) \\ &\le P_0\Big(n \sup_{\vartheta \in \mathbb{R}^k} \{Y_n \circ \vartheta - \psi_k(\vartheta)\} \ge \tfrac{1}{2}(k-1)\log n\Big). \end{aligned} \tag{3.2}$$

Next define

$$V_k = \max_{1 \le j \le k} \sup_{x \in [0,1]} |\phi_j(x)|. \tag{3.3}$$

For the orthonormal Legendre polynomials on $[0,1]$, we get

$$V_k = (2k+1)^{1/2}, \tag{3.4}$$

while V_k is bounded in the trigonometric case.

The proof of the following lemma is given in Inglot and Ledwina (1994).

LEMMA 3.1. *For every* $k \geq 1$, $0 < \varepsilon < \min(1, (2/3)kV_k^2)$ *and* $a \leq (2-\varepsilon)\varepsilon^2(16kV_k^2)^{-1}$, *we have*

$$(3.5) \qquad \Big\{x \in \mathbb{R}^k : \sup_{\theta \in \mathbb{R}^k} [\theta \circ x - \psi_k(\theta)] \geq a\Big\} \subset \big\{x : \|x\|^2 \geq (2-\varepsilon)a\big\}.$$

The asymptotic behavior of S is given in the following theorem.

THEOREM 3.2. *Assume*

$$(3.6) \qquad \lim_{n\to\infty} d(n)V_{d(n)}(n^{-1}\log n)^{1/2} = 0.$$

Then

$$(3.7) \qquad \lim_{n\to\infty} P_0(S = 1) = 1.$$

PROOF. Let $0 < \varepsilon < 2/3$. By (3.5) and (3.6) we have for all sufficiently large n, uniformly for $k \in \{1, \dots, d(n)\}$,

$$(3.8) \qquad \begin{aligned} &P_0\Big(n \sup_{\vartheta \in \mathbb{R}^k} \{Y_n \circ \vartheta - \psi_k(\vartheta)\} \geq \tfrac{1}{2}(k-1)\log n\Big) \\ &\quad \leq P_0\big(\|Y_n\|^2 \geq (2-\varepsilon)\tfrac{1}{2}n^{-1}(k-1)\log n\big). \end{aligned}$$

Application of formula (2) of Prohorov (1973) [with, in the notation of that paper, $\rho = \{(2-\varepsilon)\frac{1}{2}(k-1)\log n\}^{1/2}$, $m = k$, $\lambda = 1$, $a = k^{1/2}V_k n^{-1/2}\{(2-\varepsilon)\frac{1}{2}(k-1)\log n\}^{1/2}$] yields

$$(3.9) \qquad \begin{aligned} &P_0\Big(\|Y_n\|^2 \geq (2-\varepsilon)\frac{1}{2}n^{-1}(k-1)\log n\Big) \\ &\quad \leq c_1\Big(\frac{\rho^2}{2}\Big)^{(k-1)/2}\Big\{\Gamma\Big(\frac{k}{2}\Big)\Big\}^{-1}\exp\Big\{-\frac{1}{2}\rho^2(1-\eta(a))\Big\} \end{aligned}$$

with c_1 an absolute constant, $\eta(a) \to 0$ as $a \to 0$, provided that $\rho^2/2 \geq k$ and $a \leq 1$ [$\eta(a)$ is explicitly given on page 188 of Prohorov (1973) and satisfies $\eta(a) \leq a$]. By (3.6), $a \to 0$ as $n \to \infty$, uniformly for $k \in \{1, \dots, d(n)\}$. Moreover, $\rho^2/2 \geq k$ for n sufficiently large, independent of k.

By taking ε small enough, combination of (3.2), (3.8) and (3.9) gives, for any $\zeta > 0$ and $k \in \{1, \dots, d(n)\}$,

$$P_0(S = k) \leq n^{-(1-\zeta)(k-1)/2},$$

for $n = n(\zeta)$ large enough, uniformly for $k \in \{1, \dots, d(n)\}$. Taking, for instance, $\zeta = 0.2$ we get by (3.1), for n sufficiently large,

$$P_0(S = 1) \geq 1 - \sum_{k=2}^{d(n)} n^{-0.4(k-1)} \geq 1 - \sum_{k=2}^{\infty} n^{-0.4(k-1)}$$

and the result follows. □

COROLLARY 3.3. *If* $\{\phi_j\}$ *are the orthonormal Legendre polynomials and if* $\lim_{n\to\infty} d^3(n)n^{-1}\log n = 0$, *then* $\lim_{n\to\infty} P_0(S = 1) = 1$. *If* $\{\phi_j\}$ *is the cosine system and if* $\lim_{n\to\infty} d^2(n)n^{-1}\log n = 0$, *then* $\lim_{n\to\infty} P_0(S = 1) = 1$.

The asymptotic null distribution of T_S is given in the next theorem. Let χ_1^2 denote a r.v. with a chi-square distribution with 1 degree of freedom.

THEOREM 3.4. *If* (3.6) *holds, then under* H_0,

$$T_S \to_d \chi_1^2.$$

PROOF. By Theorem 3.2 we have $P_0(S = 1) \to 1$ as $n \to \infty$. Since under $H_0, T_1 \to_d \chi_1^2$, the result follows. □

COROLLARY 3.5. *If* $\lim_{n\to\infty} d^3(n)n^{-1}\log n = 0$, *then* $N_S \to_d \chi_1^2$ *under* H_0. *If* $\{\phi_j\}$ *is the cosine system and if* $\lim_{n\to\infty} d^2(n)n^{-1}\log n = 0$, *then* $T_S \to_d \chi_1^2$ *under* H_0.

PROOF. Combine Corollary 3.3 and Theorem 3.4. □

In Ledwina (1994), simulation results are presented on the null distribution of S and 5% critical points of N_S. The finite sample results in Table 1 of Ledwina (1994) show that indeed S is concentrating its probability mass on 1 as n becomes large. Nevertheless, the implied chi-square distribution with 1 degree of freedom for N_S does not work very well as approximation to establish accurate critical values [cf. Table 2 in Ledwina (1994) and Table 2 herein]. The first obvious remark is that $T_k \le T_{k+1}$ and hence $T_1 \le T_S$ with probability 1. Moreover, although the cases where $S > 1$ are relatively rare, they have special influence, since $n\sup\{Y_n \circ \vartheta - \psi_k(\vartheta): \vartheta \in \mathbb{R}^k\}$ and T_k are strongly related. Therefore, if $S = 2$, say, then as a rule T_2 is much larger than T_1 (due to the penalty for higher dimension in Schwarz's rule). This explains why the simulated critical values of N_S are substantially larger than the asymptotic critical values based on the chi-square-one approximation. Results on a more accurate approximation of the null distribution of N_S will be reported elsewhere.

As far as consistency is concerned, the lack of accuracy in the critical values is no problem at all, since for consistency only $o_{P_0}(n)$ of T_S under H_0 is required (cf. the remark after Theorem 4.3) and this is certainly true, not only in the formal sense (Theorem 3.4), but also in the approximation sense as is seen from Table 2 in Ledwina (1994) and Table 2 herein.

4. Asymptotic behavior of S under alternatives and consistency of T_S. Let $X_1, X_2, \dots, X_n$ be i.i.d. r.v.'s each with distribution P on $[0, 1]$. Suppose that

$$E_P\phi_1(X) = \cdots = E_P\phi_{K-1}(X) = 0, \qquad E_P\phi_K(X) \ne 0, \tag{4.1}$$

for some $K = K(P)$. Consistency of T_S will be proved for any alternative of the form (4.1), thus including essentially any alternative of interest. It will be assumed in this section that

$$\liminf_{n\to\infty} d(n) \geq K, \tag{4.2}$$

which is certainly the case if $\lim_{n\to\infty} d(n) = \infty$, since K is fixed. The asymptotic behavior of S is described in the following theorem.

THEOREM 4.1. *If* (4.1) *holds, then*

$$\lim_{n\to\infty} P(S \geq K) = 1. \tag{4.3}$$

PROOF. Since K is fixed, it suffices to prove

$$\lim_{n\to\infty} P(S = k) = 0 \quad \text{for } k = 1,\dots,K-1. \tag{4.4}$$

Consider a fixed $k \in \{1,\dots,K-1\}$. In view of (2.6) we have

$$P(S = k) \leq P(L_k \geq L_K). \tag{4.5}$$

Since $(d/dt)\psi_K(0,\dots,0,t)|_{t=0} = E_0\phi_K(X) = 0$, it holds that, for every $a \neq 0$,

$$\sup_{t\in\mathbb{R}}\{at - \psi_K(0,\dots,0,t)\} > 0. \tag{4.6}$$

Further we have

$$\sup_{\theta\in\mathbb{R}^K}\{Y_{n,K}\circ\theta - \psi_K(\theta)\} \geq \sup_{t\in\mathbb{R}}\{\bar{\phi}_K t - \psi_K(0,\dots,0,t)\} \tag{4.7}$$

and, by the law of large numbers,

$$\bar{\phi}_K \to_P E_P\phi_K(X) \neq 0 \quad \text{as } n\to\infty. \tag{4.8}$$

In view of (4.6), (4.7) and (4.8) we obtain

$$L_K \to_P \infty. \tag{4.9}$$

On the other hand, since $k \in \{1,\dots,K-1\}$ and hence $E_P\phi_1(X) = \cdots = E_P\phi_k(X) = 0$, it is easily seen [cf. (3.8)] that

$$L_k \to_P -\infty \quad \text{as } n\to\infty. \tag{4.10}$$

Combination of (4.5), (4.9) and (4.10) yields (4.4) and this completes the proof of the theorem. □

Theorem 4.1 is now applied to prove the following proposition, which in turn is the key for proving consistency of T_S.

PROPOSITION 4.2. *If* (4.1) *holds, then*

$$T_S \to_P \infty \quad \text{as } n\to\infty.$$

PROOF. Let $x \in \mathbb{R}^+$. By Theorem 4.1 we get

$$P(T_S \le x) = P(T_S \le x, S \ge K) + o(1)$$
$$\le P\left(\left\{n^{-1/2} \sum_{i=1}^{n} \phi_K(X_i)\right\}^2 \le x\right) + o(1)$$

as $n \to \infty$. Since $E_P \phi_K(X) \neq 0$, the result follows by the law of large numbers. □

THEOREM 4.3. *If T_S is bounded in probability under H_0 as $n \to \infty$, then for each (fixed) alternative of the form* (4.1), *the power of the test based on T_S tends to* 1 *as $n \to \infty$.*

PROOF. The result follows immediately from Proposition 4.2. □

In view of the proof of Proposition 4.2, it is seen that Theorem 4.3 continues to hold if $T_S = o_{P_0}(n)$ under H_0. Combination of Theorems 3.4 and 4.3 now yields the consistency.

THEOREM 4.4. *If* (3.6) *holds, the test based on T_S is consistent against any alternative of the form* (4.1).

In particular, we get by Corollary 3.5 and Theorem 4.3 for the data driven version of Neyman's test and for the data driven version of the test based on the cosine system the following result.

COROLLARY 4.5. *If $\lim_{n\to\infty} d^3(n)n^{-1}\log n = 0$, then the test based on N_S is consistent against any alternative of the form* (4.1).

If $\{\phi_j\}$ is the cosine system and if $\lim_{n\to\infty} d^2(n)n^{-1}\log n = 0$, then the test based on T_S is consistent against any alternative of the form (4.1).

To prove consistency it was sufficient to show that $\lim_{n\to\infty} P(S \ge K) = 1$ as far as the asymptotic behavior of S was concerned. By inspection of the proof of Theorem 4.1, we can get more information about the behavior of S under alternatives. Denote by

$$\mathscr{E}_k = (E_P\phi_1(X), \dots, E_P\phi_k(X)) \tag{4.11}$$

for $k = 1, 2, \dots$.

PROPOSITION 4.6. *If for some $m > k$ the distribution of $(\phi_1(X), \dots, \phi_m(X))$ under H_0 is absolutely continuous w.r.t. Lebesgue measure on $\mathbb{R}^m$ and*

$$\sup_{\vartheta\in\mathbb{R}^k} \{\mathscr{E}_k \circ \vartheta - \psi_k(\vartheta)\} < \sup_{\theta\in\mathbb{R}^m} \{\mathscr{E}_m \circ \theta - \psi_m(\theta)\} < \infty, \tag{4.12}$$

then

$$\lim_{n\to\infty} P(S = k) = 0.$$

For the proof, see Kallenberg and Ledwina (1993).

For each $\vartheta = (\vartheta_1, \dots, \vartheta_k) \in \mathbb{R}^k$ we may take $\theta = (\vartheta_1, \dots, \vartheta_k, 0, \dots, 0) \in \mathbb{R}^m$, yielding $\mathscr{E}_m \circ \theta - \psi_m(\theta) = \mathscr{E}_k \circ \vartheta - \psi_k(\vartheta)$ and, therefore,

$$\sup_{\vartheta \in \mathbb{R}^k} \{\mathscr{E}_k \circ \vartheta - \psi_k(\vartheta)\} \le \sup_{\theta \in \mathbb{R}^m} \{\mathscr{E}_m \circ \theta - \psi_m(\theta)\}. \tag{4.13}$$

In fact, the quantity

$$\sup_{\vartheta \in \mathbb{R}^k} \{\mathscr{E}_k \circ \vartheta - \psi_k(\vartheta)\} \tag{4.14}$$

is the Kullback–Leibler information number of the probability measures $P_{\lambda^{-1}(\mathscr{E}_k)}$ and P_0. The quantity (4.14) may be seen as a kind of distance between P and P_0 when considered in the k-parameter exponential family (2.1): $p_{\lambda^{-1}(\mathscr{E}_k)}$ is called the information projection [cf. Csiszár (1975) and Barron and Sheu (1991)].

Since the k-parameter exponential family is embedded in the m-parameter exponential family, P can be described at least as informative in the m-parameter exponential family as in the k-parameter exponential family [inequality (4.13)]. If really more information becomes available [i.e., if (4.12) holds], then it is picked up for large n by S, in the sense that dimension m is preferred to dimension k. If P is close to the uniform distribution in the sense that $\mathscr{E}_k$ is close to 0, (4.14) behaves like the Euclidean distance between $\mathscr{E}_k$ and 0, since $\lambda(0) = 0$ and $\psi''(0) = I$. In that case we may expect that if $E_P \phi_{k+i}(X)$ differs sufficiently from 0 for some $i \ge 1$, (4.12) holds with $m = k + i$.

On the other hand, if $P = P_\vartheta$ for some $\vartheta = (\vartheta_1, \dots, \vartheta_k)$ with $\vartheta_k \ne 0$, then for each $l < k$,

$$\begin{aligned} &\sup_{\theta \in \mathbb{R}^l} \{\theta \circ (E_\vartheta \phi_1, \dots, E_\vartheta \phi_l) - \psi_l(\theta)\} \\ &\quad = \sup_{(\theta, 0) \in \mathbb{R}^k} \{(\theta, 0) \circ \lambda_k(\vartheta) - \psi_k((\theta, 0))\} \\ &\quad < \sup_{t \in \mathbb{R}^k} \{t \circ \lambda_k(\vartheta) - \psi_k(t)\}, \end{aligned}$$

where $(\theta, 0) = (\theta_1, \dots, \theta_l, 0, \dots, 0)$. (The last inequality follows from the fact that the last supremum is uniquely attained at $t = \vartheta$ and $\vartheta_k \ne 0$.) Hence it easily follows by the law of large numbers that $P_\vartheta(S < k) \to 0$ as $n \to \infty$. Similar arguments as in the proof of Theorem 3.2 lead to $P_\vartheta(S > k) \to 0$ as $n \to \infty$. We omit the details. So in this case, we get $\lim_{n\to\infty} P_0(S = k) = 1$, that is, S selects the right dimension. [Note that in this case, for all $m \ge k$, equality holds in (4.13).]

5. General information on the simulation study and simulated distribution of S under alternatives. All programs used in this paper were written by Krzysztof Bogdan under MEN Grant 341 046 and KBN Grant 665/2/91. For a detailed description of the simulations see Ledwina (1994). Finite sample results on the null distribution of S are presented in Ledwina (1994) and Kallenberg and Ledwina (1995). Here we consider the distribution of S under alternatives.

A broad range of alternatives is investigated with different patterns of the density (pushed toward one or two ends of [0, 1], clustered near the center or a few points, multimodal etc.). Parameters are chosen to yield moderate powers (not too close to the significance level or to 1) for the tests under consideration. Also heavy-tailed alternatives are included. Such a wide class is provided by (contamination of the uniform distribution with) beta distributions. Moreover, cosine alternatives are investigated, since the test of Eubank and LaRiccia and the considered version of Bickel and Ritov's test are based on the cosine system. As a counterpart also similar Legendre-type alternatives are involved. Finally, alternatives from the exponential family (2.1) are discussed.

To be specific, the list of alternatives is as follows:

$$g_1(x) = 1 - \varepsilon + \varepsilon\beta_{p,q}(x),$$

with the Beta density $\beta_{p,q}(x) = \{B(p,q)\}^{-1}x^{p-1}(1-x)^{q-1}$

$$g_2(x) = 1 + \rho\cos(j\pi x),$$

$$g_3(x) = 1 + \rho\pi_j(x),$$

with $\{\pi_j\}$ the orthonormal Legendre polynomials on [0, 1],

$$g_4(x) = \exp\left\{\sum_{j=1}^{k}\theta_j\pi_j(x) - \psi_k(\theta)\right\}.$$

In view of Theorem 4.1 and the discussion at the end of Section 4, especially alternatives of the form g_4 are of interest when considering the distribution of S under alternatives. Some simulation results are given in Table 1.

Theory states that $P(S = k) \to 1$ as $n \to \infty$. Indeed, it is seen that for $n = 100$ the event $\{S = k\}$ has high probability and as a rule much higher than for $n = 50$, thus showing the convergence. [The case $k = 4$, $\theta = (0, -0.5, 0, -0.2)$, is an exception; even for $n = 100$ the polynomial of second degree is still dominating.]

6. Simulated powers. Recently Eubank and LaRiccia (1992) proposed the data driven test statistic $T_{n\hat{m}}$ for testing uniformity, where

$$T_{nm} = n\sum_{j=1}^{m}\tilde{a}_{jn}^2, \qquad (6.1)$$

TABLE 1

Estimated $P(S = s)$ (%) *based on* 10,000 *samples in each case*; $d(50) = 10$, $d(100) = 12$; *alternative* g_4

Parameters			s											
k	θ	n	1	2	3	4	5	6	7	8	9	10	11	12
1	0.3	50	94	5	1									
		100	96	3	1									
2	(0, −0.4)	50	23	71	4	1								
		100	5	91	3	1								
2	(0.25, −0.35)	50	34	60	4	1								
		100	12	84	3	1								
3	(0, 0, 0.4)	50	41	1	53	4	1							
		100	13		83	3	1							
4	(0, −0.5, 0, −0.2)	50	19	70	2	8	1							
		100	3	83	1	12	1							
4	(0.1, 0.15, −0.25, −0.35)	50	47	3	6	40	3	1						
		100	17	1	4	74	3	1						
5	(0, 0, 0, 0, 0.5)	50	49	2	1		45	3	1					
		100	15				81	3	1					
8	(0, 0, 0, 0, 0, 0, 0, −0.7)	50	52	4	1					40	2	1		
		100	11							84	3	1		

with

$$\tilde{a}_{jn} = n^{-1} \sum_{i=1}^{n} \sqrt{2} \cos(j\pi X_i) \tag{6.2}$$

and where $\hat{m}$ is the minimizer of

$$-(n+1) \sum_{j=1}^{m} n^{-1} \tilde{a}_{jn}^2 + m(n-1)^{-1} + \sum_{j=1}^{m} (n-1)^{-1} \tilde{a}_{2j,n}. \tag{6.3}$$

Another goodness-of-fit statistic is proposed by Bickel and Ritov [(1992), page 55], which with $a_{jn} = a\sqrt{2j}$, $j_n = d(n)$ and based on the cosine system reads as

$$T_{\mathrm{BR}} = \max_{1 \le j \le d(n)} \left\{ \frac{T_{nj} - j}{\sqrt{2j}} \right\},$$

with T_{nj} given by (6.1). Uniformity is rejected for large values of T_{BR}. [Here $a = a(n, \alpha)$ in a_{jn} is the critical value of T_{BR}.]

Before presenting power results, in Table 2 we give for illustration the estimated critical values of the test statistics under $n = 100$ based on 10,000 samples. Also the null distribution of S and $\hat{m}$ are shown.

As already noted in Ledwina (1994), the difference between the asymptotic 0.05 critical value being equal to 3.841 and the simulated ones for $d(100) \ge 2$ is substantial. A deeper study, which will be reported elsewhere, is needed to get a better null approximation (cf. also the discussion at the end of Section 3). It turns out that a second-order approximation is very accurate in numeri-

TABLE 2
Empirical critical values of T_{BR}, N_S and $T_{n\hat{m}}$, $\alpha = 0.05$, and empirical distributions of S and $\hat{m}$ under uniformity and $d(n) = 12$, $n = 100$, based on 10,000 samples in each case

Critical values of	$d(n)$											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
T_{BR}	1.989	2.535	2.720	2.830	2.903	2.989	3.023	3.060	3.081	3.108	3.117	3.141
N_S	3.836	5.269	5.499	5.557	5.571	5.581	5.581	5.586	5.586	5.586	5.586	5.586
$T_{n\hat{m}}$	3.946	5.895	7.749	9.348	10.810	12.243	13.712	14.940	16.244	17.770	19.150	20.479
Estimated probab.	s											
(%)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
$(S = s)$	96	3	1									
$(\hat{m} = s)$	26	9	7	6	6	6	6	5	6	7	7	10

cal examples. Further, it is seen that $\hat{m}$ is much less concentrated in 1 than S and that the critical values of $T_{n\hat{m}}$ are much different from each other.

Power comparison of N_S, T_{BR} and $T_{n\hat{m}}$ is now made in Tables 3–5. Throughout, the significance level α equals 0.05. (All Monte Carlo experiments are based on 10,000 samples and hence the estimation error is smaller than 0.01 with confidence 0.95.)

For convenience, also the maximal available power $\beta^*_{1,j}$ and the minimal power $\beta^{2,j}_*$ of Neyman's tests based on subsequent $N_1, N_2, \ldots, N_j$ is pre-

TABLE 3
Estimated powers (%) of N_S, T_{BR} and $T_{n\hat{m}}$ for beta alternatives (g_1) based on 10,000 samples in each case; $\alpha = 0.05$

Parameters			$n = 50$, $d(n) = 10$						$n = 50$, $d(n) = 6$	$n = 100$, $d(n) = 12$						$n = 100$, $d(n) = 8$
p	q	ε	$\beta^*_{1,10}$	k^*	$\beta^{2,10}_*$	N_S	T_{BR}	$T_{n\hat{m}}$	$T_{n\hat{m}}$	$\beta^*_{1,12}$	k^*	$\beta^{2,12}_*$	N_S	T_{BR}	$T_{n\hat{m}}$	$T_{n\hat{m}}$
3	3	0.5	52	2	18	53	40	22	31	86	2	46	88	76	42	56
2	2	0.8	62	2	19	63	44	24	37	94	2	51	95	82	50	67
1.5	1.5	1.0	37	2	10	38	23	15	21	73	2	24	76	52	29	40
0.5	0.5	0.6	69	10	53	57	40	41	42	91	12	80	85	70	69	71
2	3	0.7	76	2	37	75	66	38	53	99	2	79	98	96	74	87
1.5	2	0.9	68	2	22	65	52	28	39	96	2	61	95	89	56	73
0.8	1.5	0.5	46	1	24	32	36	17	23	75	1	42	63	64	30	39
1	0.5	0.5	64	9	60	52	52	42	47	89	4	87	82	82	71	76
0.8	0.5	0.5	54	10	45	40	36	32	34	78	12	71	65	62	55	59
0.2	0.2	0.3	93	10	63	74	51	59	53	100	12	89	95	86	90	87
2	4	0.5	60	3	32	55	58	31	44	92	3	68	88	90	60	74
2	10	0.25	54	1	34	41	49	28	35	82	1	60	74	80	50	60
10	20	0.25	40	6	23	36	41	32	43	74	5	46	62	75	57	70

TABLE 4

Estimated powers (%) of N_S, T_{BR} and $T_{n\hat{m}}$ for cosine alternatives (g_2) and Legendre alternatives (g_3), based on 10,000 samples in each case; $\alpha = 0.05$

Parameters			$n = 50$, $d(n) = 10$						$n = 50$, $d(n) = 6$	$n = 100$, $d(n) = 12$						$n = 100$, $d(n) = 8$
	ρ	j	$\beta^*_{1,10}$	k^*	$\beta^{2,10}_*$	N_S	T_{BR}	$T_{n\hat{m}}$	$T_{n\hat{m}}$	$\beta^*_{1,12}$	k^*	$\beta^{2,12}_*$	N_S	T_{BR}	$T_{n\hat{m}}$	$T_{n\hat{m}}$
g_2	0.4	1	51	1	21	34	41	17	22	81	1	39	69	72	30	39
	0.5	2	57	2	41	56	48	27	39	88	2	70	87	84	50	65
	0.7	4	79	6	14	50	71	60	78	99	6	20	83	98	91	97
	0.7	5	71	7	5	33	63	60	78	98	7	6	65	97	92	98
	0.7	6	71	10	8	23	60	65	81	97	10	9	46	95	93	98
g_3	0.35	1	71	1	31	54	60	25	35	95	1	60	90	90	47	61
	0.40	2	72	2	53	70	59	36	49	95	2	81	95	90	64	77
	0.37	3	57	3	5	53	37	30	41	90	3	5	83	74	57	71
	0.33	4	45	4	7	25	24	26	33	76	4	7	45	51	48	60
	0.30	5	32	5	5	13	14	20	18	63	5	5	23	33	38	43

sented as well as the index k^* for which the maximal power is attained. (In determining the minimal power, N_1 is excluded, since otherwise in many cases the minimum is attained for $k = 1$ and the minimal power is close to α.)

It is seen from the great difference between $\beta^*_{1,j}$ and $\beta^{2,j}_*$ that one may lose much power by making an unfortunate choice of the index k.

The power of $T_{n\hat{m}}$ is given for different values of $d(n)$, since it is very sensitive for the choice of $d(n)$, in contrast to N_S and T_{BR}. For instance, taking alternative g_1 with $p = 2$, $q = 3$ and $\varepsilon = 0.7$, we get for $n = 50$ and $d(50) = 6$ estimated power 0.53, while for $d(50) = 10$ the estimated power

TABLE 5

Estimated powers (%) of N_S, T_{BR} and $T_{n\hat{m}}$ for the exponential family (g_4), based on 10,000 samples in each case; $\alpha = 0.05$

Parameters		$n = 50$, $d(n) = 10$						$n = 50$, $d(n) = 6$	$n = 100$, $d(n) = 12$						$n = 100$, $d(n) = 8$
s	$\theta = (\theta_1, \ldots, \theta_s)$	$\beta^*_{1,10}$	k^*	$\beta^{2,10}_*$	N_S	T_{BR}	$T_{n\hat{m}}$	$T_{n\hat{m}}$	$\beta^*_{1,12}$	k^*	$\beta^{2,12}_*$	N_S	T_{BR}	$T_{n\hat{m}}$	$T_{n\hat{m}}$
1	0.3	56	1	24	38	44	19	25	84	1	44	74	75	33	44
2	(0, −0.4)	58	2	20	59	44	23	34	91	2	50	93	82	46	62
2	(0.25, −0.35)	59	2	21	57	46	24	34	92	2	51	90	83	46	61
3	(0, 0, 0.4)	65	3	7	60	44	37	47	93	3	8	87	80	67	78
4	(0, −0.5, 0, −0.2)	60	2	19	63	40	25	38	94	2	54	96	81	53	73
4	(0.1, 0.15, −0.25, −0.35)	73	4	22	55	56	51	64	97	4	42	86	90	83	91
5	(0, 0, 0, 0, 0.5)	79	5	9	51	45	56	47	98	5	12	85	86	89	91
8	(0, 0, 0, 0, 0, 0, 0, −0.7)	90	8	7	48	55	67	76	100	8	10	89	95	97	97

equals 0.38. Taking g_4 with $s = 8$ and $\theta = (0, 0, 0, 0, 0, 0, 0, -0.7)$, we get for $n = 50$ and $d(50) = 6$ estimated power 0.76, while for $d(50) = 5$ the estimated power equals 0.08. The unstability makes $T_{n\hat{m}}$ less attractive as a general testing procedure.

The simulations show that N_S adapts well to the alternative at hand. For the beta alternatives (g_1), N_S has high power, also compared to the maximal available power $\beta^*_{1,j}$. In many cases it beats T_{BR} and, far more, $T_{n\hat{m}}$. Schwarz's rule is oriented to avoid overparametrizing the model. Therefore, one might think that the penalty for higher dimension causes a lower power of N_S in cases where k^* is large (heavy-tailed alternatives). It turns out that in such cases (see Table 3) nevertheless N_S has reasonable power and as a rule substantially higher than T_{BR} and $T_{n\hat{m}}$.

Since T_{BR} and $T_{n\hat{m}}$ are based on the cosine system, they perform, as expected, better for cosine alternatives (g_2), while for Legendre alternatives (g_3) the situation is reversed. Finally, also for the alternatives from the exponential family (g_4), N_S again performs very well.

Therefore, the conclusion in Ledwina (1994) that the data driven version of Neyman's test performs well in comparison to the tests of Anderson and Darling and Watson and Neuhaus can be extended to Eubank and LaRiccia's test and to Bickel and Ritov's test, unless the alternative is highly oscillating.

Acknowledgment. The authors are grateful to K. Bogdan for programming assistance.

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.

BARRON, A. R. and COVER, T. M. (1991). Minimum complexity density estimation. *IEEE Trans. Inform. Theory* **37** 1034–1054.

BARRON, A. R. and SHEU, C. (1991). Approximation of density functions by sequences of exponential families. *Ann. Statist.* **19** 1347–1369.

BICKEL, P. J. and RITOV, Y. (1992). Testing for goodness of fit: a new approach. In *Nonparametric Statistics and Related Topics* (A. K. Md. E. Saleh, ed.) 51–57. North-Holland, Amsterdam.

CSISZÁR, I. (1975). *I*-divergence geometry of probability distributions and minimization problems. *Ann. Probab.* **3** 146–158.

EUBANK, R. L. and HART, J. D. (1992). Testing goodness-of-fit in regression via order selection criteria. *Ann. Statist.* **20** 1412–1425.

EUBANK, R. L., HART, J. D. and LARICCIA, V. N. (1993). Testing goodness of fit via nonparametric function estimation techniques. *Comm. Statist. Theory Methods* **22** 3327–3354.

EUBANK, R. L. and LARICCIA, V. N. (1992). Asymptotic comparison of Cramér–von Mises and nonparametric function estimation techniques for testing goodness-of-fit. *Ann. Statist.* **20** 2071–2086.

HAUGHTON, D. M. A. (1988). On the choice of a model to fit data from an exponential family. *Ann Statist.* **16** 342–355.

INGLOT, T., KALLENBERG, W. C. M. and LEDWINA, T. (1994). Power approximations to and power comparison of certain goodness-of-fit tests. *Scand. J. Statist.* **21** 131–145.

INGLOT, T. and LEDWINA, T. (1994). Asymptotic optimality of data driven Neyman's tests. Preprint 43, Institute of Mathematics, Technical Univ. Wrocław.

KALLENBERG, W. C. M. and LEDWINA, T. (1993). Consistency and Monte Carlo simulation of a data driven version of smooth goodness-of-fit tests. Memorandum 1158, Univ. Twente.
KALLENBERG, W. C. M and LEDWINA, T. (1995). On data driven Neyman's tests. *Probab. Math. Statist.* To appear.
LEDWINA, T. (1994). Data driven version of Neyman's smooth test of fit. *J. Amer. Statist. Assoc.* **89** 1000–1005.
NEUHAUS, G. (1988). Addendum to "Local asymptotics for linear rank statistics with estimated score functions." *Ann. Statist.* **16** 1342–1343.
NEYMAN, J. (1937). "Smooth" test for goodness of fit. *Skand. Aktuarietidskr.* **20** 150–199.
PROHOROV, A. V. (1973). On sums of random vectors. *Theory Probab. Appl.* **18** 186–188.
RAYNER, J. C. W. and BEST, D. J. (1989). *Smooth Tests of Goodness of Fit.* Oxford Univ. Press.
RAYNER, J. C. W. and BEST, D. J. (1990). Smooth tests of goodness of fit: an overview. *Internat. Statist. Rev.* **58** 9–17.
RISSANEN, J. (1983). A universal prior for integers and estimation by minimum description length. *Ann. Statist.* **11** 416–431.
RISSANEN, J. (1987). Stochastic complexity. *J. Roy. Statist. Soc. Ser. B* **49** 223–239.
SCHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461–464.
WATSON, G. S. (1961). Goodness-of-fit tests on a circle. *Biometrika* **48** 109–114.

FACULTY OF APPLIED MATHEMATICS
UNIVERSITY OF TWENTE
P.O. BOX 217
7500 AE ENSCHEDE
THE NETHERLANDS

INSTITUTE OF MATHEMATICS
TECHNICAL UNIVERSITY OF WROCŁAW
50-370 WROCŁAW
WYBRZEŻE WYSPIAŃSKIEGO 27
POLAND