PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 15 (1995), pp. 409–426

ON DATA DRIVEN NEYMAN'S TESTS

BY

WILBERT C. M. KALLENBERG (ENSCHEDE) AND TERESA LEDWINA* (WROCŁAW)

Abstract. The smooth tests for testing uniformity were introduced by Neyman [15]. The data driven method of selecting the number of components in a smooth test for uniformity is discussed, including the first-order asymptotic null distribution, consistency, empirical critical values and Monte Carlo powers. The first-order asymptotic null distribution is not sufficiently precise for approximation tools. A substantial improvement is made in this paper by deriving a second-order approximation of the null distribution, which turns out to be very accurate in numerical examples. The approximations are based on the second-order behaviour of Schwarz's selection rule S under uniformity. The new results on S are of independent interest.

1. Introduction. Let $X_1, X_2, \ldots, X_n$ 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 we assume that under H_0 the distribution of X_i is the uniform distribution on $[0, 1]$.

Neyman's [15] paper of 1937, called *'Smooth test' for goodness of fit*, is seen as the starting point of a subbranch of goodness-of-fit tests, the smooth tests of goodness-of-fit. In fact, Neyman's test is a particular example of the whole class of smooth goodness-of-fit tests, taking the Legendre polynomials as the orthonormal system. In [16], pp. 19–22, it is memorized that Neyman's test is intended as an "optimal" competitor to Pearson's chi-squared test. *Optimal* means here that the test is locally uniformly most powerful symmetric, unbiased for testing uniformity against alternatives of the form

$$c(\theta)\exp\Big\{\sum_{j=1}^{k}\theta_j\pi_j(x)\Big\}, \quad 0 \leqslant x \leqslant 1, \tag{1.1}$$

where $\{\pi_j\}$ are the orthonormal Legendre polynomials on $[0, 1]$ and $c(\theta)$ is a normalizing constant. Alternatives of the form (1.1) were vaguely described

* Research supported by Grant MEN 341114.

as *smooth*. The test statistic for testing H_0: $\theta_j = 0, j = 1, \ldots, k$, corresponding to uniformity, is given by

$$N_k = \sum_{j=1}^{k} \{n^{-1/2} \sum_{i=1}^{n} \pi_j(X_i)\}^2. \tag{1.2}$$

Replacing π_j in (1.2) by

$$p_j(x) = \{1_{(d_{j-1}, d_j]}(x) - (d_j - d_{j-1})\}(d_j - d_{j-1})^{-1/2}, \tag{1.3}$$

where $1_A(x)$ denotes the indicator function of the set A, we get Pearson's chi-square test

$$P_k = \sum_{j=1}^{k} \{n^{-1/2} \sum_{i=1}^{n} p_j(X_i)\}^2 = \sum_{j=1}^{k} (O_j - E_j)^2/E_j \tag{1.4}$$

with partition $0 = d_0 < d_1 < \ldots < d_k = 1$, O_j and E_j denoting the observed and expected (under uniformity) number of observations in the j-th interval, respectively.

Hamdan [6], [8] implemented a suggestion of Lancaster to use different orthonormal systems, leading to what nowadays is called the *class of smooth tests for goodness-of-fit*, given by the test statistics

$$T_k = \sum_{j=1}^{k} \{n^{-1/2} \sum_{i=1}^{n} \phi_j(X_i)\}^2, \tag{1.5}$$

where $\phi_0, \phi_1, \ldots$ is an orthonormal system in $L_2([0, 1])$ with $\phi_0(x) \equiv 1$. For more extensive and general information on the class of smooth tests we refer to Rayner and Best [18]. After the papers of Barton [2]–[4], Watson [21], and Hamdan [6]–[8] had appeared, there was no much progress in the field. A possible reason is that the required computations are quite heavy by hand. This is no longer a barrier with modern computers. Although smooth tests arose little interest for several years, nowadays Neyman's paper [15] is considered to be ingenious (cf. Le Cam and Lehmann [12], p. ix). Recently, there was a renewed interest in smooth tests. For an overview we refer to Rayner and Best [19], who conclude in reviewing several tests of fit: "don't use those other methods – use a smooth test!". The same conclusion is derived in Milbrodt and Strasser [14], p. 14. One of the important issues in applying T_k is the number k of components. We discuss three different approaches.

In the first one, k (and the orthonormal system) are selected such that the alternatives of particular interest are represented using only the first k components of the orthonormal system with k as small as possible. It requires that the user not automatically applies the test, but thinks why the goodness-of-fit test is executed and what types of alternatives are of particular interest. This imperative task is useful on itself, and therefore the method is more advantagous than disadvantagous. While especially focused on the selected (type of)

alternatives, the test keeps its omnibus character. A criterion of a simple structure to implement the idea is presented and extensively motivated in Inglot et al. [9]. This approach is closest to the classical application of smooth tests, since the test statistic is not modified and a formal test is executed.

The second approach is a more data-analytic approach. Obviously, it is not valid to apply a lot of significance tests, by varying k, to a data set and focus only on the most critical of them. Neyman was very clear about it as can be seen from the citation on p. 47 of [18]. Nevertheless, if the null hypothesis is rejected, the components may be used informally to suggest the nature of the departure from the null hypothesis. This second approach is described in detail in [18].

Returning to formal testing theory, the third approach concerns a data driven selection of k. This idea was introduced by Ledwina [13]. Roughly speaking, it works as follows. First, Schwarz's [20] 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 a more precise instrument, being Neyman's test in the "right" dimension. As with the original test of Neyman, the data driven version can also be applied using other orthonormal systems than the Legendre polynomials.

Observe that the "data driven approach" has also the extra advantage lying behind the idea of the second approach. Namely, when rejecting the null hypothesis, automatically a well-defined alternative model is provided for the data at hand.

The rest of this paper is mainly concerned with this third approach. In Section 2 the method is described and properties of the tests as obtained by Ledwina [13] and Kallenberg and Ledwina [11] are discussed. In Section 3 new results on Schwarz's selection rule are presented. The results may be of independent interest. They are applied in Section 4, where an accurate new approximation of the null distribution of the data driven version of smooth goodness-of-fit tests is derived.

2. A data driven version of smooth goodness-of-fit tests. Consider the class of test statistics given by (1.5). Assume that the functions $\phi_1, \phi_2, \ldots$ are bounded, but not necessarily uniformly bounded in $j = 1, 2, \ldots$ Define for $k = 1, 2, \ldots$ exponential families 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

(2.2)

$$\theta = (\theta_1, \ldots, \theta_k), \quad \phi = (\phi_1, \ldots, \phi_k), \quad \psi_k(\theta) = \log \int_0^1 \exp\{\theta \circ \phi(x)\}\, dx,$$

and $\circ$ stands for the inner product in $\mathbb{R}^k$. When there is no confusion, the dimension k is sometimes suppressed in the notation.

It is well known that

$$\lambda(\theta) = \psi'(\theta) = \mathrm{E}_\theta \phi(X), \tag{2.3}$$

where $'$ denotes derivative. Moreover, by orthonormality, we have

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

Writing

$$Y_n = (\bar{\phi}_1, \ldots, \bar{\phi}_k), \quad \bar{\phi}_j = n^{-1} \sum_{i=1}^{n} \phi_j(x_i) \tag{2.5}$$

we obtain the expression for the density of $X_1, \ldots, X_n$, each X_i having density (2.1), in the form

$$\exp\{n[Y_n \circ \theta - \psi_k(\theta)]\}. \tag{2.6}$$

Schwarz's [20] Bayesian information criterion (BIC) for choosing submodels corresponding to successive dimensions yields

$$S = \min\{k\colon\ 1 \leqslant k \leqslant d(n),\ n \sup_{\vartheta \in \mathbb{R}^k} \{Y_n \circ \vartheta - \psi_k(\vartheta)\} - \tfrac{1}{2} k \log n \geqslant n \sup_{t \in \mathbb{R}^j} \{Y_n \circ t - \psi_j(t)\} - \tfrac{1}{2} j \log n,\ j = 1, \ldots, d(n)\}. \tag{2.7}$$

Although it is not mentioned in the notation, S depends of course on the upper bound $d(n)$ of the exponential families under consideration. Schwarz's BIC is further discussed in Section 3.

The data driven smooth test statistic is defined by

$$T_S = \sum_{j=1}^{S} \{n^{-1/2} \sum_{i=1}^{n} \phi_j(X_i)\}^2 \tag{2.8}$$

with S given by (2.7). The null hypothesis of uniformity is rejected for large values of T_S.

The idea behind (2.8) is that Schwarz's rule gives a first indication about the true density of the observations by fitting an exponential family model to the data and that the finishing touch comes from the smooth test in the selected exponential family.

Numerical results in, e.g., [9] show that Neyman's test performs very well provided a good choice of k has been made. On the other hand, examples in Table 2 of [9] and in Tables 3–5 in [11] 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. Both by theoretical results and by the Monte Carlo simulation it will be argued that T_S is such a procedure. To do this we start by considering N_S, the data driven

version of Neyman's test, i.e., (2.8) with $\{\phi_j\} = \{\pi_j\}$, the orthonormal Legendre polynomials. The asymptotic null distribution of N_S is given by Corollary 3.5 of [11] and reads as follows. (By χ_1^2 we denote an r.v. with a chi-square distribution with 1 degree of freedom.)

THEOREM 2.1. *If*

$$\lim_{n\to\infty} d^3(n)n^{-1}\log n = 0, \tag{2.9}$$

then

$$N_S \xrightarrow{d} \chi_1^2 \text{ under } H_0. \tag{2.10}$$

Using this limit theorem as an approximation theorem we obtain the asymptotic 0.05 and 0.01 critical values equal to 3.841 and 6.635, respectively. To see how this approximation performs compared to the finite sample case, simulated 0.05 and 0.01 critical values are presented in Table 1 (partially from Table 2 of [13] and Table 2 of [11]).

The accuracy of the simulation results are seen in Table 2, where exemplary empirical sizes of N_S are calculated for empirical critical values from

TABLE 1. 5% and 1% critical values of N_S simulated for subsequent values of $d(n)$, from 10 000 samples in each case

$d(n)$	n	20		50		80		100		120	
	α	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01
1		3.705	6.363	3.896	6.762	3.889	6.408	3.836	6.472	3.811	6.587
2		5.367	8.451	5.350	8.592	5.260	8.511	5.269	8.449	5.032	8.967
3		6.537	10.216	5.865	10.447	5.508	9.853	5.499	9.941	5.240	10.272
4		7.213	11.332	6.015	11.191	5.581	10.239	5.557	10.227	5.314	10.522
5		7.617	12.183	6.094	11.827	5.592	10.386	5.571	10.352	5.321	10.603
6		7.854	12.997	6.096	12.128	5.592	10.388	5.581	10.352	5.321	10.603
7				6.117	12.217	5.592	10.388	5.581	10.352	5.321	10.615
8				6.121	12.465	5.592	10.388	5.586	10.352	5.321	10.615
9				6.121	12.465	5.592	10.388	5.586	10.352	5.321	10.615
10				6.121	12.495	5.592	10.388	5.586	10.352	5.321	10.615
11								5.586	10.352	5.321	10.615
12								5.586	10.352	5.321	10.615

TABLE 2. Empirical 5% sizes of N_S simulated for subsequent values of $d(n)$, from 10 000 samples in each case

n \ $d(n)$	1	2	3	4	5	6	7	8	9	10
20	0.0540	0.0491	0.0471	0.0470	0.0452	0.0469	–	–	–	–
50	0.0506	0.0489	0.0500	0.0499	0.0502	0.0504	0.0496	0.0497	0.0497	0.0497
80	0.0480	0.0477	0.0498	0.0499	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500

Table 1 on pseudorandom sequences different from those used to prepare Table 1.

It is clearly seen from Table 1 that the difference between the asymptotic 0.05 (0.01) critical value equal to 3.841 (6.635) and the simulated ones for $d(n) \geqslant 2$ is substantial, even for $n = 120$. Therefore, we may conclude that the first-order limiting theorem (Theorem 2.1) is not sufficiently precise for approximation tools, and hence there is a need for a second-order limiting theorem to improve the accuracy of the approximation of N_S under H_0. This is done, more generally, for T_S in Section 4 on the basis of the results of Section 3.

To show that N_S has good power properties under a wide class of alternatives we consider its consistency under the alternative distribution P on $[0, 1]$. Suppose that

$$\text{(2.11)} \qquad E_P\pi_1(X) = \ldots = E_P\pi_{K-1}(X) = 0, \quad E_P\pi_K(X) \neq 0$$

for some $K = K(P)$, X having distribution P. Essentially any alternative of interest satisfies (2.11). (Note that the orthonormal system $\{\pi_j\}$ is complete.) The consistency of N_S is given by Corollary 4.5 of [11] and reads as follows.

THEOREM 2.2. *If* (2.11) *holds, and if*

$$\text{(2.12)} \qquad \liminf_{n\to\infty} d(n) \geqslant K(P), \quad \lim_{n\to\infty} d^3(n)n^{-1}\log n = 0,$$

then the test based on N_S is consistent at P.

Since $K(P)$ is fixed, (2.9) together with $\lim_{n\to\infty} d(n) = \infty$ implies that the test based on N_S is consistent against any alternative of the form (2.11). Apart from the theoretical justification in Theorem 2.2, the good power properties of N_S for a wide range of alternatives are shown by simulation results in Tables 3 and 4 of [13] and Tables 3–5 of [11]. For illustration here we present in one table simulated powers of N_S, the widely recommended tests of Watson (W), Anderson and Darling (AD), Neuhaus (N) and the recently introduced procedures of Bickel and Ritov (BR), and Eubank and LaRiccia (ELR). For definitions of the several test statistics we refer to [13] and [11]. The alternatives under consideration are given by the following densities:

$$g_1(x) = \tfrac{1}{4}\{x^{-1/2} + (1-x)^{-1/2}\}, \quad g_2(x) = 1.5(1-x)^{1/2},$$

$$g_3(x) = (3\sqrt{2})\{x^{1/2}1_{[0,1/2]}(x) + (1-x)^{1/2}1_{(1/2,1]}(x)\}, \quad g_4(x) = 1 + \tfrac{1}{2}\cos(2\pi x),$$

$$g_5(x) = \exp\{-0.5\pi_2(x) - 0.2\pi_4(x) - \psi(0, -0.5, -0.2)\} \text{ (cf. (2.1))}.$$

Table 3 and other simulation results show that N_S has a stable and relatively high power for a broad range of alternatives.

TABLE 3. Estimated powers (in %) based on 10000 samples in each case; $\alpha = 0.05$, $n = 50$, $d(n) = 10$

Alternatives	N_S	W	AD	N	BR	ELR
g_1	81	65	69	64	63	63
g_2	58	39	71	62	60	24
g_3	61	59	16	50	43	24
g_4	57	61	33	56	48	27
g_5	63	55	15	46	40	25

The theoretical results on the asymptotic null distribution of N_S and its consistency may be extended to other orthonormal systems as well. To do this we make some assumptions. Define

$$V_k = \max_{1 \leq j \leq k} \sup_{x \in [0,1]} |\phi_j(x)|. \tag{2.13}$$

ASSUMPTION 1. *The following condition holds*:

$$\lim_{n \to \infty} d(n) V_{d(n)} (n^{-1} \log n)^{1/2} = 0. \tag{2.14}$$

For the orthonormal Legendre polynomials on [0, 1] we get $V_k = (2k+1)^{1/2}$. The asymptotic null distribution and the consistency of T_S are given by Theorems 3.4 and 4.4 in [11] and read as follows:

THEOREM 2.3. *If Assumption* 1 *holds, then under* H_0

$$T_S \xrightarrow{d} \chi_1^2, \tag{2.15}$$

and the test based on T_S *is consistent against any alternative* P *of the form* (2.11), *provided that* $\liminf_{n \to \infty} d(n) \geq K(P)$.

3. Schwarz's selection rule. It is seen in Section 2 that the first-order result $T_S \xrightarrow{d} \chi_1^2$ is not sharp enough to imply accurate approximations of the null distribution. To improve the approximations we need some second-order results. Both for technical reasons and for more insight in what is going on, we have to derive precise results on the behaviour of S under H_0. The notation P_0 refers to the null hypothesis, where X_i has the uniform distribution on [0, 1]. While at the first order only the event $\{S = 1\}$ plays a role (cf. Theorems 3.2 and 3.4 in [11]), for the second-order results both $\{S = 1\}$ and $\{S = 2\}$ come in

the picture. The event $\{S \geqslant 3\}$ may still be neglected under H_0 as is seen in the following theorem:

THEOREM 3.1. *If Assumption* 1 *holds, then*

$$P_0(S \geqslant 3) = O(n^{-1}(\log n)^{1/2}) \quad \text{as } n \to \infty. \tag{3.1}$$

Proof. Let K be fixed with $3 \leqslant K \leqslant d(n)$. Write

$$P_0(S \geqslant 3) = \sum_{k=3}^{K} P_0(S = k) + \sum_{k=K+1}^{d(n)} P_0(S = k). \tag{3.2}$$

By the definition of S and the application of Lemma 3.2 in [10], there exists for each $k \in \{3, \ldots, K\}$ a constant c_k^* and there exists a constant $\tilde{c}_K$ such that

$$\begin{aligned} P_0(S = k) &\leqslant P_0\Big(n \sup_{\vartheta \in R^k} \{Y_n \circ \vartheta - \psi_k(\vartheta)\} \geqslant \tfrac{1}{2}(k-1)\log n\Big) \\ &\leqslant c_k^*(\log n)^{(1/2)(k-2)} \exp\{-\tfrac{1}{2}(k-1)\log n\} \leqslant \tilde{c}_K(\log n)^{1/2} n^{-1}. \end{aligned} \tag{3.3}$$

For $k = K+1, \ldots, d(n)$ it follows from Lemma 3.1 in [11] that for

$$\varepsilon \in (0, \min\{1, \tfrac{2}{3} k V_k^2\}),$$

uniformly for $k \in \{1, \ldots, d(n)\}$,

$$P_0(S = k) \leqslant P_0\big(nY_n \circ Y_n \geqslant (1-\varepsilon)(k-1)\log n\big),$$

provided that n is sufficiently large. Note that here we use Assumption 1. Next we apply formula (2) of [17] with

$$\varrho = \{(1-\varepsilon)(k-1)\log n\}^{1/2}, \quad m = k, \quad \lambda = 1,$$

$$a = k^{1/2} V_k n^{-1/2} \{(1-\varepsilon)(k-1)\log n\}^{1/2},$$

which yields

$$P_0\big(nY_n \circ Y_n \geqslant (1-\varepsilon)(k-1)\log n\big) \leqslant c_1 \left(\frac{\varrho^2}{2}\right)^{(1/2)(k-1)} \frac{1}{\Gamma(k/2)} \exp\{-\tfrac{1}{2}\varrho^2(1-\eta(a))\},$$

where c_1 is an absolute constant, $\eta(a) \to 0$ as $a \to 0$, provided that $\varrho^2/2 \geqslant k$ and $a \leqslant 1$ ($\eta(a)$ is explicitly given on p. 188 of [17] and satisfies $\eta(a) \leqslant a$). In view of Assumption 1, $a \to 0$ as $n \to \infty$ (uniformly for $k \in \{1, \ldots, d(n)\}$). Moreover, $\varrho^2/2 \geqslant k$ for sufficiently large n (independent of k). By taking ε small enough we therefore get for any $\zeta > 0$ and $k \in \{1, \ldots, d(n)\}$

$$P_0(S = k) \leqslant n^{-(1/2)(1-\zeta)(k-1)} \tag{3.4}$$

for $n = n(\zeta)$ large enough, independent of $k \in \{1, \ldots, d(n)\}$. Taking, for instance, $K = 3$ and $\zeta = 0.2$ we get for n sufficiently large

$$P_0(S \geqslant 3) \leqslant \tilde{c}_3 (\log n)^{1/2} n^{-1} + \sum_{k=4}^{\infty} n^{-0.4(k-1)} = \tilde{c}_3 (\log n)^{1/2} n^{-1} + \frac{n^{-1.2}}{1-n^{-0.4}},$$

and hence (3.1) holds. ∎

Remark 3.1. It is clear from the proof that the main contribution comes from $S = 3$, as expected. In other words, by virtually the same proof (cf. (3.3) and (3.4)) it follows that for each fixed K

$$P_0(S \geqslant K) = O\big(n^{-(1/2)(K-1)} (\log n)^{(1/2)(K-2)}\big) \quad \text{as } n \to \infty. \tag{3.5}$$ ∎

In view of Theorem 3.1 we may concentrate on the events $\{S = 1\}$ and $\{S = 2\}$. The event $\{S = 1\}$ means that dimension 1 is more important than dimension 2, dimension 3, dimension 4, etc. Under H_0 it turns out that the comparison of dimension 1 with dimension 2 is up to the second order the only comparison of interest. Similarly, on the event $\{S = 2\}$ the only important contribution under H_0 comes from the comparison of dimension 2 with dimensions 1 and 3. This is shown in the next two theorems.

THEOREM 3.2. *There exists a positive constant c_2 such that for each event A*

$$P_0(A, S = 1) \leqslant P_0\big(A, n\bar{\phi}_2^2 \leqslant \log n + c_2 n^{-1/2} (\log n)^{3/2}\big) + O\big(n^{-1} (\log n)^{-1/2}\big) \tag{3.6}$$

and

$$P_0(A, S = 1) \geqslant P_0\big(A, n\bar{\phi}_2^2 \leqslant \log n - c_2 n^{-1/2} (\log n)^{3/2}\big) + O\big(n^{-1} (\log n)^{-1/2}\big) - P_0(S \geqslant 3) \tag{3.7}$$

as $n \to \infty$, uniformly in A.

Proof. By the definition of S we get

$$P_0(A, S = 1) \leqslant P_0\Big(A, n \sup_{\vartheta \in \boldsymbol{R}} \{Y_n \vartheta - \psi_1(\vartheta)\} - \tfrac{1}{2} \log n \geqslant n \sup_{t \in \boldsymbol{R}^2} \{Y_n \circ t - \psi_2(t)\} - \log n\Big).$$

For each fixed k we have

$$\sup_{\theta \in \boldsymbol{R}^k} \{y \circ \theta - \psi_k(\theta)\} = \tfrac{1}{2} \|y\|^2 + O(\|y\|^3) \quad \text{as } y \to 0,$$

where $\|\cdot\|$ denotes the Euclidean norm in $\boldsymbol{R}^k$. Therefore, on the set

$$\{y\colon |y_j| \leqslant n^{-1/2} (2 \log n)^{1/2},\ j = 1, \ldots, k\}$$

we get

$$\Big|\sup_{\theta \in \boldsymbol{R}^k} \{y \circ \theta - \psi_k(\theta)\} - \tfrac{1}{2} \|y\|^2\Big| \leqslant \bar{c}_k n^{-3/2} (\log n)^{3/2}$$

for some constant $\bar{c}_k$. Hence, uniformly in A,

$$P_0(A, S=1) \leqslant P_0\big(A, n\{\tfrac{1}{2}\bar{\phi}_1^2+\bar{c}_1 n^{-3/2}(\log n)^{3/2}\}+\tfrac{1}{2}\log n \geqslant n\{\tfrac{1}{2}(\bar{\phi}_1^2+\bar{\phi}_2^2)-\bar{c}_2 n^{-3/2}(\log n)^{3/2}\}\big) + P_0\big(n^{1/2}|\bar{\phi}_1| > (2\log n)^{1/2}\big)+P_0\big(n^{1/2}|\bar{\phi}_2| > (2\log n)^{1/2}\big) = P_0\big(A, n\bar{\phi}_2^2 \leqslant \log n + c_2 n^{-1/2}(\log n)^{3/2}\big)+O\big(n^{-1}(\log n)^{-1/2}\big)$$

as $n \to \infty$, with $c_2 = 2(\bar{c}_1+\bar{c}_2)$, where for $j = 1, 2$ we have used

$$P_0\big(n^{1/2}|\bar{\phi}_j| > (2\log n)^{1/2}\big) = O\big(n^{-1}(\log n)^{-1/2}\big),$$

which follows by standard large deviation theory. This completes the proof of (3.6).

To prove (3.7) note that

$$P_0(A, S=1) \geqslant P_0\big(A, n\sup_{\vartheta\in R}\{Y_n\vartheta-\psi_1(\vartheta)\}-\tfrac{1}{2}\log n > n\sup_{t\in R^2}\{Y_n\circ t-\psi_2(t)\}-\log n\big)-P_0(S\geqslant 3).$$

The rest of the proof is similar to the proof of (3.6). ∎

THEOREM 3.3. *Let $d(n) \geqslant 2$. There exists a positive constant c_3 such that for each event A*

$$(3.8)\quad P_0(A, S=2) \leqslant P_0\big(A, n\bar{\phi}_2^2 \geqslant \log n - c_3 n^{-1/2}(\log n)^{3/2}, n\bar{\phi}_3^2 \leqslant \log n + c_3 n^{-1/2}(\log n)^{3/2}\big)+O\big(n^{-1}(\log n)^{-1/2}\big)$$

and

$$(3.9)\quad P_0(A, S=2) \geqslant P_0\big(A, n\bar{\phi}_2^2 \geqslant \log n + c_3 n^{-1/2}(\log n)^{3/2}, n\bar{\phi}_3^2 \leqslant \log n - c_3 n^{-1/2}(\log n)^{3/2}\big)+O\big(n^{-1}(\log n)^{-1/2}\big)-P_0(S\geqslant 3)$$

as $n \to \infty$, uniformly in A.

Proof. By the definition of S we get (as in the proof of Theorem 3.2)

$$P_0(A, S=2) \leqslant P_0\big(A, n\sup_{\vartheta\in R^2}\{Y_n\circ\vartheta-\psi_2(\vartheta)\}-\log n \geqslant n\sup_{t\in R}\{Y_n t-\psi_1(t)\}-\tfrac{1}{2}\log n, n\sup_{\vartheta\in R^2}\{Y_n\circ\vartheta-\psi_2(\vartheta)\}-\log n \geqslant n\sup_{\theta\in R^3}\{Y_n\circ\theta-\psi_3(\theta)\}-\tfrac{3}{2}\log n\big) \leqslant P_0\big(A, n\{\tfrac{1}{2}(\bar{\phi}_1^2+\bar{\phi}_2^2)+\bar{c}_2 n^{-3/2}(\log n)^{3/2}\} - \log n \geqslant n\{\tfrac{1}{2}\bar{\phi}_1^2-\bar{c}_1 n^{-3/2}(\log n)^{3/2}\} - \tfrac{1}{2}\log n, n\{\tfrac{1}{2}(\bar{\phi}_1^2+\bar{\phi}_2^2)+\bar{c}_2 n^{-3/2}(\log n)^{3/2}\} - \log n \geqslant n\{\tfrac{1}{2}(\bar{\phi}_1^2+\bar{\phi}_2^2+\bar{\phi}_3^2)-\bar{c}_3 n^{-3/2}(\log n)^{3/2}\} - \tfrac{3}{2}\log n\big)$$

$$+\sum_{j=1}^{3} P_0\left(n^{1/2}|\bar{\phi}_j| > (2\log n)^{1/2}\right)$$

$$= P_0\left(A,\ n\bar{\phi}_2^2 \geqslant \log n - 2(\bar{c}_1+\bar{c}_2)n^{-1/2}(\log n)^{3/2},\right.$$

$$\left. n\bar{\phi}_3^2 \leqslant \log n + 2(\bar{c}_2+\bar{c}_3)n^{-1/2}(\log n)^{3/2}\right) + O\left(n^{-1}(\log n)^{-1/2}\right).$$

By taking $c_3 = \max\{2(\bar{c}_1+\bar{c}_2),\ 2(\bar{c}_2+\bar{c}_3)\}$ the proof of (3.8) is completed. Inequality (3.9) is proved in a similar way. ∎

Remark 3.2. The $O(n^{-1}(\log n)^{1/2})$-terms in (3.6)–(3.9) may be replaced by $O(n^{-r}(\log n)^{-1/2})$ for arbitrary r. This is seen by replacing the set $\{y\colon |y_j| \leqslant n^{-1/2}(2\log n)^{1/2},\ j = 1, \ldots, k\}$ by the set $\{y\colon |y_j| \leqslant n^{-1/2}(2r\log n)^{1/2},\ j = 1, \ldots, k\}$ and noting that

$$P_0\left(n^{1/2}|\bar{\phi}_j| > (2r\log n)^{1/2}\right) = O\left(n^{-r}(\log n)^{-1/2}\right) \quad \text{as } n \to \infty. \ \blacksquare$$

4. Second-order approximation to the null distribution of T_S. As was seen from Theorem 2.1 and Table 1, the first-order χ_1^2-approximation of the null distribution of T_S is not very precise. Therefore, we consider the second-order approximation, based on the second-order limiting theorems for S under H_0 in the previous section. Denoting by U_1, U_2 two independent r.v.'s, each with a standard normal distribution, we shall prove the following result:

THEOREM 4.1. *Let $d(n) \geqslant 2$. If Assumption 1 holds, and Cramér's condition is satisfied, i.e.,*

$$\limsup_{|t|\to\infty} |\mathrm{E}_{P_0} \exp\{it\phi_1(X_1)\}| < 1, \tag{4.1}$$

then

$$P_0(T_S \leqslant x) = \Pr(U_1^2 \leqslant x)\Pr(U_2^2 \leqslant \log n) + \Pr(U_1^2 + U_2^2 \leqslant x,\ U_2^2 \geqslant \log n) + O(n^{-1}\log n) \tag{4.2}$$

as $n \to \infty$, uniformly in x.

It is easily seen that the approximation on the right-hand side of (4.2) may also be written as

$$\Pr(U_1^2 \leqslant x) - \Pr(U_1^2 \leqslant x,\ U_1^2 + U_2^2 \geqslant x,\ U_2^2 \geqslant \log n). \tag{4.3}$$

In this formulation the correction term in comparison to the first-order approximation $\Pr(U_1^2 \leqslant x)$ is clearly seen. Corresponding to the simulation results in Table 1, indeed, $P_0(T_S \leqslant x)$ is overestimated by its first-order approximation. The two terms on the right-hand side of (4.2) are the leading terms of $P_0(T_S \leqslant x,\ S = 1)$ and $P_0(T_S \leqslant x,\ S = 2)$, respectively. For $x \leqslant \log n$

the second term disappears and the second-order approximation in that case is simply obtained by multiplying the first-order approximation by $\Pr(U_2^2 \leqslant \log n)$.

Condition (4.1) is satisfied if $\phi_1(X_i)$ has an absolutely continuous component, and therefore it is fulfilled in many cases, e.g., for Neyman's test with the orthonormal Legendre polynomials. To prove Theorem 4.1 we state and prove first several lemmas.

LEMMA 4.1. *For each fixed k and for each constant $c_4 \in \boldsymbol{R}$ we have*

$$P_0\left(n\bar{\phi}_k^2 \leqslant \log n + c_4 n^{-1/2}(\log n)^{3/2}\right) - P_0\left(n\bar{\phi}_k^2 \leqslant \log n\right) = O(n^{-1}\log n). \tag{4.4}$$

Proof. By standard large deviation theory (cf., e.g., [5], p. 553) we have

$$\begin{aligned} &P_0\left(n\bar{\phi}_k^2 \leqslant \log n + c_4 n^{-1/2}(\log n)^{3/2}\right) - P_0\left(n\bar{\phi}_k^2 \leqslant \log n\right) \\ &\quad = P_0\left(n\bar{\phi}_k^2 > \log n\right) - P_0\left(n\bar{\phi}_k^2 > \log n + c_4 n^{-1/2}(\log n)^{3/2}\right) \\ &\quad = 2\Pr\left(U_1 > (\log n)^{1/2}\right)\left\{1 + O\left(n^{-1/2}(\log n)^{3/2}\right)\right\} \\ &\qquad - 2\Pr\left(U_1 > \left\{\log n + c_4 n^{-1/2}(\log n)^{3/2}\right\}^{1/2}\right)\left\{1 + O\left(n^{-1/2}(\log n)^{3/2}\right)\right\} \\ &\quad = O(n^{-1}\log n) \quad \text{as } n \to \infty. \ \blacksquare \end{aligned} \tag{4.5}$$

LEMMA 4.2. *If Assumption 1 holds, then*

$$P_0(T_S \leqslant x, S = 1) = P_0\left(n\bar{\phi}_1^2 \leqslant x, n\bar{\phi}_2^2 \leqslant \log n\right) + O(n^{-1}\log n) \tag{4.6}$$

as $n \to \infty$, uniformly in x.

Proof. Taking $A = \{T_1 \leqslant x\}$ we see by Theorem 3.2 that there exists a positive constant c_2 such that

$$\begin{aligned} P_0(T_S \leqslant x, S = 1) &= P_0(T_1 \leqslant x, S = 1) \\ &\leqslant P_0\left(T_1 \leqslant x, n\bar{\phi}_2^2 \leqslant \log n + c_2 n^{-1/2}(\log n)^{3/2}\right) + O\left(n^{-1}(\log n)^{-1/2}\right) \\ &\leqslant P_0\left(T_1 \leqslant x, n\bar{\phi}_2^2 \leqslant \log n\right) + P_0\left(n\bar{\phi}_2^2 \leqslant \log n + c_2 n^{-1/2}(\log n)^{3/2}\right) \\ &\quad - P_0\left(n\bar{\phi}_2^2 \leqslant \log n\right) + O\left(n^{-1}(\log n)^{-1/2}\right) \end{aligned}$$

as $n \to \infty$, uniformly in x. The application of Lemma 4.1 and noting that $T_1 = n\bar{\phi}_1^2$ yield

$$P_0(T_S \leqslant x, S = 1) \leqslant P_0\left(n\bar{\phi}_1^2 \leqslant x, n\bar{\phi}_2^2 \leqslant \log n\right) + O(n^{-1}\log n)$$

as $n \to \infty$, uniformly in x. Similarly, the converse inequality is obtained by using (3.7) in combination with Theorem 3.1. $\blacksquare$

LEMMA 4.3. *For each constant $c_5 \in \boldsymbol{R}$ we have*

$$P_0\left(n\bar{\phi}_2^2 \geqslant \log n, n\bar{\phi}_3^2 > \log n - c_5 n^{-1/2}(\log n)^{3/2}\right) = O\left(n^{-1}(\log n)^{-1}\right) \tag{4.7}$$

as $n \to \infty$.

Proof. Write

$$D_n = \{(y_1, y_2)\colon\ y_1 \geqslant (\log n)^{1/2},\ y_2 \geqslant \{\log n - c_5 n^{-1/2}(\log n)^{3/2}\}^{1/2}\}.$$

The set D_n is a convex Borel set and the point

$$a = \left((\log n)^{1/2},\ \{\log n - c_5 n^{-1/2}(\log n)^{3/2}\}^{1/2}\right)$$

is the nearest point to the origin. The application of Assertion 2 in Remark 1 on p. 67 of [1] yields

$$\begin{aligned} P_0\left((n^{1/2}\bar{\phi}_2, n^{1/2}\bar{\phi}_3) \in D_n\right) &= \Pr\left((U_1, U_2) \in D_n\right)\{1 + O(n^{-1/2}(\log n)^{3/2})\} \\ &= O\left(n^{-1}(\log n)^{-1}\right) \quad \text{as } n \to \infty. \end{aligned}$$

By the same argument for the other three regions of similar form, which together with D_n constitute the event in (4.7), the proof is completed. ∎

LEMMA 4.4. *If Assumption* 1 *holds, then*

$$(4.8)\quad P_0(T_S \leqslant x,\ S = 2) = P_0\left(n(\bar{\phi}_1^2 + \bar{\phi}_2^2) \leqslant x,\ n\bar{\phi}_2^2 \geqslant \log n\right) + O(n^{-1}\log n)$$

as $n \to \infty$, *uniformly in* x.

Proof. Taking $A = \{T_2 \leqslant x\}$ we see by Theorem 3.3 that there exists a positive constant c_3 such that

$$\begin{aligned} &P_0(T_S \leqslant x,\ S = 2) \\ &\leqslant P_0\left(T_2 \leqslant x,\ n\bar{\phi}_2^2 \geqslant \log n - c_3 n^{-1/2}(\log n)^{3/2}\right) + O\left(n^{-1}(\log n)^{-1/2}\right) \\ &\leqslant P_0\left(T_2 \leqslant x,\ n\bar{\phi}_2^2 \geqslant \log n\right) + P_0\left(\log n - c_3 n^{-1/2}(\log n)^{3/2} \leqslant n\bar{\phi}_2^2 < \log n\right) \\ &\quad + O\left(n^{-1}(\log n)^{-1/2}\right) \end{aligned}$$

as $n \to \infty$, uniformly in x. The application of Lemma 4.1 and noting that $T_2 = n(\bar{\phi}_1^2 + \bar{\phi}_2^2)$ yield

$$P_0(T_S \leqslant x,\ S = 2) \leqslant P_0\left(n(\bar{\phi}_1^2 + \bar{\phi}_2^2) \leqslant x,\ n\bar{\phi}_2^2 \geqslant \log n\right) + O(n^{-1}\log n)$$

as $n \to \infty$, uniformly in x.

For the converse inequality we consider (3.9) in combination with Theorem 3.1 and we obtain

$$\begin{aligned} &P_0(T_S \leqslant x,\ S = 2) \\ &\geqslant P_0\left(T_2 \leqslant x,\ n\bar{\phi}_2^2 \geqslant \log n\right) - P_0\left(\log n \leqslant n\bar{\phi}_2^2 < \log n + c_3 n^{-1/2}(\log n)^{3/2}\right) \\ &\quad - P_0\left(n\bar{\phi}_2^2 \geqslant \log n,\ n\bar{\phi}_3^2 > \log n - c_3 n^{-1/2}(\log n)^{3/2}\right) + O\left(n^{-1}(\log n)^{1/2}\right) \\ &\geqslant P_0\left(T_2 \leqslant x,\ n\bar{\phi}_2^2 \geqslant \log n\right) + O(n^{-1}\log n) \end{aligned}$$

by Lemmas 4.1 and 4.3. This completes the proof of the lemma. ∎

LEMMA 4.5. *If* (4.1) *holds, then, uniformly for* x, *as* $n \to \infty$

$$P_0(n\bar{\phi}_1^2 \leqslant x) = \Pr(U_1^2 \leqslant x) + O(n^{-1}). \tag{4.9}$$

Proof. By the standard Edgeworth expansion (cf., e.g., [5], Theorem 3 on p. 541) it follows that, uniformly in x,

$$\begin{aligned} P_0(n\bar{\phi}_1^2 \leqslant x) &= P_0(n^{1/2}\bar{\phi}_1 \leqslant x^{1/2}) - P_0(n^{1/2}\bar{\phi}_1 < -x^{1/2}) \\ &= \Pr(U_1 \leqslant x^{1/2}) - \left\{\frac{\mathrm{E}\phi_1^3}{6n^{1/2}}\right\}\varphi(x^{1/2})(x-1) - \Pr(U_1 \leqslant -x^{1/2}) \\ &\quad + \left\{\frac{\mathrm{E}\phi_1^3}{6n^{1/2}}\right\}\varphi(-x^{1/2})(x-1) + O(n^{-1}) \\ &= \Pr(U_1^2 \leqslant x) + O(n^{-1}), \end{aligned}$$

where φ denotes the standard normal density. ∎

LEMMA 4.6. *Uniformly for* x *as* $n \to \infty$ *we have*

$$P_0(n\bar{\phi}_1^2 \leqslant x,\ n\bar{\phi}_2^2 > \log n) = \Pr(U_1^2 \leqslant x)\Pr(U_2^2 > \log n) + O(n^{-1}\log n) \tag{4.10}$$

and

$$\begin{aligned} P_0(n(\bar{\phi}_1^2 + \bar{\phi}_2^2) \leqslant x,\ n\bar{\phi}_2^2 \geqslant \log n) \\ = \Pr(U_1^2 + U_2^2 \leqslant x,\ U_2^2 \geqslant \log n) + O(n^{-1}\log n). \end{aligned} \tag{4.11}$$

Proof. We give the proof of (4.11); the relation (4.10) can be shown in a quite similar way. Without loss of generality assume that $x \geqslant \log n$. Write

$$\begin{aligned} D_{n1} &= \{(y_1, y_2):\ y_2 \geqslant (\log n)^{1/2},\ y_1^2 + y_2^2 \leqslant x\}, \\ D_{n2} &= \{(y_1, y_2):\ y_2 \leqslant -(\log n)^{1/2},\ y_1^2 + y_2^2 \leqslant x\}. \end{aligned}$$

The sets D_{n1} and D_{n2} are convex. It follows by Assertion 2 in Remark 1 on p. 67 of [1] that for $i = 1, 2$, with $a = (0, (\log n)^{1/2})$ if $i = 1$ and $a = (0, -(\log n)^{1/2})$ if $i = 2$, we have

$$\begin{aligned} P_0((n^{1/2}\bar{\phi}_1, n^{1/2}\bar{\phi}_2) \in D_{ni}) &= \Pr((U_1, U_2) \in D_{ni})\{1 + O(n^{-1/2}(\log n)^{3/2})\} \\ &= \Pr((U_1, U_2) \in D_{ni}) + O(n^{-1}\log n), \end{aligned}$$

which implies (4.11). ∎

Proof of Theorem 4.1. By (4.6), (4.8)–(4.11) and (3.1) we have, uniformly in x, as $n \to \infty$

$$\begin{aligned} P_0(T_S \leqslant x) &= P_0(T_S \leqslant x,\ S = 1) + P_0(T_S \leqslant x,\ S = 2) + P_0(T_S \leqslant x,\ S \geqslant 3) \\ &= P_0(n\bar{\phi}_1^2 \leqslant x) - P_0(n\bar{\phi}_1^2 \leqslant x,\ n\bar{\phi}_2^2 > \log n) \\ &\quad + P_0(n(\bar{\phi}_1^2 + \bar{\phi}_2^2) \leqslant x,\ n\bar{\phi}_2^2 \geqslant \log n) + O(n^{-1}\log n) \end{aligned}$$

$$= \Pr(U_1^2 \leqslant x) - \Pr(U_1^2 \leqslant x,\ U_2^2 > \log n) + \Pr(U_1^2 + U_2^2 \leqslant x,\ U_2^2 \geqslant \log n) + O(n^{-1}\log n),$$

and hence (4.2) holds. ∎

The approximation on the right-hand side of (4.2) is surprisingly simple and very accurate, especially when compared to the first-order approximation. The structure of the approximation clarifies that under H_0 up to the second order only $\{S = 1\}$ and $\{S = 2\}$ are important. Moreover, $\{S = 1\}$ and $\{S = 2\}$ may be further restricted to dimension 1 beats dimension 2, corresponding to $U_2^2 \leqslant \log n$, and dimension 2 beats dimension 1, corresponding to $U_2^2 > \log n$, respectively. (As a side-note, remark that here also dimension 2 beats dimension 3 is cancelled out.)

Both to investigate the accuracy of the approximation and to facilitate its evaluation we derive an upper and lower bound of $\Pr(U_1^2 + U_2^2 \leqslant x,\ U_2^2 \geqslant \log n)$. Let for $x \geqslant \log n$

(4.12)
$$D^* = \{(y_1, y_2):\ y_1^2 + y_2^2 \geqslant x,\ 0 \leqslant y_1 \leqslant (x - \log n)^{1/2},\ (\log n)^{1/2} \leqslant y_2 \leqslant x^{1/2}\}.$$

The area $a(x)$ of this set equals

$$a(x) = \{x^{1/2} - (\log n)^{1/2}\}(x - \log n)^{1/2} - \tfrac{1}{2}x \arccos\left(\frac{\log n}{x}\right)^{1/2} + \tfrac{1}{2}(\log n)^{1/2}(x - \log n)^{1/2}, \tag{4.13}$$

while for $(y_1, y_2) \in D^*$ we have

$$n^{1/2}\exp(-x) \leqslant \exp\{-\tfrac{1}{2}(y_1^2 + y_2^2)\} \leqslant \exp(-\tfrac{1}{2}x). \tag{4.14}$$

Using symmetry we therefore get

$$\begin{aligned} &\Pr(U_1^2 + U_2^2 \leqslant x,\ U_2^2 \geqslant \log n) \qquad (4.15)\\ &= 4\Pr(U_1^2 + U_2^2 \leqslant x,\ 0 \leqslant U_1 \leqslant (x - \log n)^{1/2},\ (\log n)^{1/2} \leqslant U_2 \leqslant x^{1/2})\\ &= 4\{\Pr(0 \leqslant U_1 \leqslant (x - \log n)^{1/2})\Pr((\log n)^{1/2} \leqslant U_2 \leqslant x^{1/2}) - (2\pi)^{-1}a(x)b(x)\} \end{aligned}$$

with

$$n^{1/2}\exp(-x) \leqslant b(x) \leqslant \exp(-\tfrac{1}{2}x). \tag{4.16}$$

Replacing $\Pr(U_1^2 + U_2^2 \leqslant x,\ U_2^2 \geqslant \log n)$ on the right-hand side of (4.2) by its lower and upper bounds obtained from (4.15) and (4.16) and inserting for x the empirical critical values of Table 1 we get the results compiled in Table 4. It is seen that for $d(n) = 2$ the approximation is excellent, even for small values of n. For $d(n) \geqslant 3$ we have a slight overestimation, but the approximation is

TABLE 4. Lower and upper bounds of the second-order approximation of $P_0(T_S \leqslant x)$ for empirical critical values x of Table 1 (probabilities $\times 10^5$)

	$\alpha = 0.05$			$\alpha = 0.01$		
n	x	Lower bounds	Upper bounds	x	Lower bounds	Upper bounds
20	5.367	94055	94912	8.451	98376	99095
	6.537	96377	97341	10.216	99226	99662
	7.213	97273	98182	11.332	99518	99817
	7.617	97698	98552	12.183	99664	99885
	7.854	97915	98733	12.997	99763	99926
50	5.350	94758	95050	8.592	98654	99152
	5.865	95784	96204	10.447	99378	99698
	6.015	96043	96493	11.191	99545	99800
	6.094	96173	96637	11.827	99652	99859
	6.096	96176	96641	12.128	99693	99880
	6.117	96210	96678	12.217	99704	99886
	6.121	96217	96685	12.465	99734	99900
				12.495	99737	99902
80	5.260	94928	95028	8.511	98697	99109
	5.508	95426	95579	9.853	99253	99578
	5.581	95563	95732	10.239	99364	99659
	5.592	95584	95755	10.386	99401	99686
				10.388	99402	99686
100	5.269	95122	95174	8.449	98704	99076
	5.499	95562	95652	9.941	99301	99597
	5.557	95667	95768	10.227	99380	99656
	5.571	95692	95796	10.352	99411	99679
	5.581	95710	95815			
	5.586	95719	95825			
120	5.032	94813	94819	8.967	98981	99306
	5.240	95214	95235	10.272	99406	99664
	5.314	95353	95382	10.522	99464	99707
	5.321	95366	95396	10.603	99482	99720
				10.615	99485	99722

also in that case rather accurate, and far more accurate than the first-order approximation.

As was noted before, if $x \leqslant \log n$, the second term in the approximation (4.2) disappears. On the other hand, if $x \geqslant 2\log n$, then

$$(4.17)\quad 0 \leqslant \Pr(U_2^2 \geqslant \log n) - \Pr(U_1^2 + U_2^2 \leqslant x,\ U_2^2 \geqslant \log n)$$

$$= \Pr(U_1^2 + U_2^2 > x,\ U_2^2 \geqslant \log n) \leqslant \Pr(U_1^2 + U_2^2 > x) = \exp(-\tfrac{1}{2}x) \leqslant n^{-1},$$

which implies for $x \geqslant 2\log n$

$$P_0(T_S \leqslant x) = \Pr(U_1^2 \leqslant x)\Pr(U_2^2 \leqslant \log n) + \Pr(U_2^2 \geqslant \log n) + O(n^{-1}\log n).$$

Therefore we may take the following very simple approximation of $P_0(T_S \leqslant x)$:

$$\text{(4.18)}\quad \begin{cases} \Pr(U_1^2 \leqslant x)\Pr(U_2^2 \leqslant \log n) & \text{if } x \leqslant \log n, \\ \Pr(U_1^2 \leqslant x)\Pr(U_2^2 \leqslant \log n) + \Pr(U_2^2 \geqslant \log n) & \text{if } x \geqslant 2\log n, \\ \text{linearize} & \text{if } \log n \leqslant x \leqslant 2\log n. \end{cases}$$

To illustrate the approximation (4.18) consider $n = 50$ and the empirical 5% critical values x of Table 1. We obtain the following:

x	5.350	5.865	6.015	6.094	6.096	6.117	6.121
[approximation (4.18)] $\times 10^5$	93901	95068	95408	95587	95592	95639	95649

It is seen that the approximation works quite well in this case. As is seen from (4.17) the approximation (4.18) is larger than the right-hand side of (4.2) if $x \geqslant 2\log n$. It was seen in Table 4 that for the 1% critical values the right-hand side of (4.2) gives already a (slight) overestimation. Since as a rule in Table 4 the 1% critical values are larger than $2\log n$, the approximation (4.18) gives a (higher) overestimation of $P_0(T_S \leqslant x)$. A possible remedy is to replace the interval $[\log n, 2\log n]$ in (4.18) by, for instance, $[\log n, 3\log n]$ if small levels are concerned.

REFERENCES

[1] A. K. Aleshkyavichene, *Multidimensional integral limit theorems for large deviation probabilities*, Theory Probab. Appl. 28 (1983), pp. 65–88.

[2] D. E. Barton, *On Neyman's smooth test of goodness of fit and its power with respect to a particular system of alternatives*, Skand. Aktuarietidskr. 36 (1953), pp. 24–63.

[3] — *A form of Neyman's ψ^2 test of goodness of fit applicable to grouped and discrete data*, ibidem 38 (1955), pp. 1–16.

[4] — *Neyman's ψ^2 test of goodness of fit when the null hypothesis is composite*, ibidem 39 (1956), pp. 216–245.

[5] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. II, 2nd edition, Wiley, New York 1971.

[6] M. A. Hamdan, *The powers of certain smooth tests of goodness of fit*, Austral. J. Statist. 4 (1962), pp. 25–40.

[7] — *The number and width of classes in the chi-square test*, J. Amer. Statist. Assoc. 58 (1963), pp. 678–689.

[8] — *A smooth test of goodness of fit based on the Walsh functions*, Austral. J. Statist. 6 (1964), pp. 130–136.

[9] T. Inglot, W. C. M. Kallenberg and T. Ledwina, *Power approximations to and power comparison of smooth goodness-of-fit tests*, Scand. J. Statist. 21 (1994), pp. 131–145.

[10] W. C. M. Kallenberg, *Bahadur deficiency of likelihood ratio tests in exponential families*, J. Multivariate Anal. 11 (1981), pp. 506–531.

[11] — and T. Ledwina, *Consistency and Monte Carlo simulation of a data driven version of smooth goodness-of-fit tests*, Ann. Statist. (to appear).
[12] L. LeCam and E. L. Lehmann, *J. Neyman — On the occasion of his 80th birthday*, ibidem 2 (1974), pp. vii–xiii.
[13] T. Ledwina, *Data driven version of Neyman's smooth test of fit*, J. Amer. Statist. Assoc. 89 (1994), pp. 1000–1005.
[14] H. Milbrodt and H. Strasser, *On the asymptotic power of the two-sided Kolmogorov-Smirnov test*, J. Statist. Plann. Inference 26 (1990), pp. 1–23.
[15] J. Neyman, *'Smooth test' for goodness of fit*, Skand. Aktuarietidskr. 20 (1937), pp. 149–199.
[16] — *Some memorable incidents in probabilistic/statistical studies*, in: *Asymptotic Theory of Statistical Tests and Estimation*, I. M. Chakravarti (Ed.), Academic Press, New York 1980, pp. 1–32.
[17] A. V. Prohorov, *On sums of random vectors*, Theory Probab. Appl. 18 (1973), pp. 186–188.
[18] J. C. W. Rayner and D. J. Best, *Smooth Tests of Goodness of Fit*, Oxford University Press, New York 1989.
[19] — *Smooth tests of goodness of fit: An overview*, Internat. Statist. Rev. 58 (1990), pp. 9–17.
[20] G. Schwarz, *Estimating the dimension of a model*, Ann. Statist. 6 (1978), pp. 461–464.
[21] G. S. Watson, *Some recent results in chi-square goodness-of-fit tests*, Biometrics 15 (1959), pp. 440–468.

Faculty of Applied Mathematics
University of Twente
P.O. Box 217
7500 AE Enschede, The Netherlands

Institute of Mathematics
Technical University of Wrocław
Wybrzeże Wyspiańskiego 27
50-370 Wrocław, Poland

Received on 14.12.1993