The Annals of Statistics
2003, Vol. 31, No. 6, 2036–2058
© Institute of Mathematical Statistics, 2003

EFFICIENT AND ADAPTIVE NONPARAMETRIC TEST FOR THE TWO-SAMPLE PROBLEM

BY GILLES R. DUCHARME AND TERESA LEDWINA

Université Montpellier II and Polish Academy of Sciences

The notion of efficient test for a Euclidean parameter in a semiparametric model was introduced by Stein [*Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** (1956) 187–195]. Such tests are locally most powerful for a wide class of infinite-dimensional nuisance parameters. The first formal application of this notion to a suitably parametrized two-sample problem was provided by Hájek [*Ann. Math. Statist.* **33** (1962) 1124–1147]. However, this and subsequent solutions appear to be not well-suited for practical applications. This article aims to show that an adaptive two-sample test introduced recently by Janic-Wróblewska and Ledwina [*Scand. J. Statist.* **27** (2000) 281–297] is locally most powerful under a more realistic setting.

1. Introduction. Efficient nonparametric tests were introduced by Stein (1956) for situations where, as expressed by Bickel, Klaassen, Ritov and Wellner (1993), "...we believe we have enough knowledge to model some features of the data parametrically but are unwilling to assume anything for other features."

One such situation is the two-sample pure shift model that is typically phrased as follows. We have two independent samples $X_1, \dots, X_m$ and $X_{m+1}, \dots, X_N$. We assume

$$X_i \sim \begin{cases} F(\cdot), & i = 1, \dots, m, \\ F(\cdot - \Delta), & i = m+1, \dots, N, \end{cases} \tag{1.1}$$

where F is an unknown absolutely continuous distribution function possessing a differentiable density f and $\Delta \in \mathbb{R}$. The testing problem is

$$H_0 : \Delta = 0 \quad \text{against} \quad H_1 : \Delta \neq 0 \tag{1.2}$$

or

$$H_0 : \Delta = 0 \quad \text{against} \quad H_2 : \Delta > 0. \tag{1.3}$$

In the context of (1.1) and (1.2) [or (1.3)], an efficient test according to Stein's notion of efficiency should have the same asymptotic power [under $\Delta = \Delta_N = O(N^{-1/2})$ as $N \to \infty$] when F is unknown as the best test for (1.2) [(1.3)] when F is known. Note that to get asymptotic results in a standard form, one considers

Received July 2001; revised July 2002.
AMS 2000 subject classifications. Primary 62G10, 62G20; secondary 62G99.
Key words and phrases. Adaptive test, efficient test, model selection, nonparametric test, two-sample problem.

an even more specialized pure shift model

$$X_i \sim \begin{cases} F\left(\cdot + \Delta\sqrt{\dfrac{1-\eta_N}{\eta_N}}\right), & i = 1, \dots, m, \\ F\left(\cdot - \Delta\sqrt{\dfrac{\eta_N}{1-\eta_N}}\right), & i = m+1, \dots, N, \end{cases} \tag{1.4}$$

where $\eta_N = m/N$ and the asymptotic analysis is again considered under $\Delta = \Delta_N = O(N^{-1/2})$ as $N \to \infty$.

Hájek (1962) provided the first formal efficient two-sample test. The construction of his test is related to the estimation of the score generating function $f'(x)/f(x)$. However, due to its very slow rate of convergence, this procedure has turned out to be of little practical value. Many researchers have proposed different, data-dependent approaches for choosing an adequate score generating function. In general, such choices can be interpreted as selecting a model for the data at hand. Therefore, the resulting test procedures have been called adaptive. For nice reviews of adaptive methods for (1.2) and (1.3), see Hušková (1984), Hogg and Lenth (1984) and references therein [see also Hušková and Sen (1985) for an alternative development].

The pure shift model (1.1) is, in many practical situations, unrealistic. Unfortunately, efficient and/or adaptive procedures derived under (1.1) break down when applied to more realistic situations. For evidence and some discussion, see Behnen (1975) and Behnen and Neuhaus (1989). Thus several investigators [Behnen (1975, 1981), Behnen and Neuhaus (1983), Neuhaus (1987), Bajorski (1992) and Fan (1996)] have attempted to provide sensitive two-sample tests for more realistic setups than (1.1). In particular, the test proposed by Neuhaus (1987) exhibits nice empirical power behavior. However, none of the above tests has been shown to be asymptotically efficient. More recently, Janic-Wróblewska and Ledwina (2000) proposed a new adaptive two-sample test which, in an extensive simulation study, has been shown to compare nicely with many other tests, including Neuhaus' (1987) test.

The aim of the present article is to prove that the test proposed by Janic-Wróblewska and Ledwina (2000), hereafter referred to as the J-WL test, is efficient under a general nonparametric two-sample model. To explain our result more precisely, we now describe the model under consideration and how Stein's notion of efficiency is adapted to this setup.

We assume throughout that we have two independent samples,

$$X_i \sim \begin{cases} F(\cdot), & i = 1, \dots, m, \\ G(\cdot), & i = m+1, \dots, N, \end{cases}$$

where F and G are unknown continuous distribution functions. The testing problem is

$$\mathcal{H}_0 : F = G \quad \text{against} \quad \mathcal{H}_1 : F \neq G. \tag{1.5}$$

A key point in our construction is a reparametrization of (1.5) that was introduced by Behnen (1981) and successfully exploited by Neuhaus (1987), among others. A similar reparametrization was used in some unpublished reports by Parzen [cf. Eubank, LaRiccia and Rosenstein (1987), Sections 2.1 and 4.1]. To write this reparametrization, some auxiliary notation is needed. Set

$$H(x) = \eta_N F(x) + (1 - \eta_N)G(x), \tag{1.6}$$

where $\eta_N = m/N$. We shall assume throughout that $\eta_N \to \eta \in (0, 1)$ as $N \to \infty$. Set also $U_i = H(X_i)$, $i = 1, \dots, N$. Each of these random variables has a density with respect to Lebesgue's measure λ on $[0, 1]$. In particular,

$$U_i \sim \begin{cases} p_1(u) = 1 + (1 - \eta_N)\bar{b}(u), & i = 1, \dots, m, \\ p_2(u) = 1 - \eta_N \bar{b}(u), & i = m + 1, \dots, N, \end{cases} \tag{1.7}$$

where, under our assumptions, it follows from (1.8) of Neuhaus (1987) that $\bar{b}$ is a bounded function from $L_2([0, 1], \lambda)$ [$L_2(\lambda)$ for short] satisfying $\int \bar{b}(u)\, d\lambda(u) = 0$. The connection between $\bar{b}$ and (F, G) is given by $\bar{b} = \bar{f} - \bar{g}$, where $\bar{f} = d(F \circ H^{-1})/d\lambda$ and $\bar{g} = d(G \circ H^{-1})/d\lambda$. We have $\bar{b} \equiv 0 \Leftrightarrow F \equiv G \Leftrightarrow p_1 \equiv p_2 \equiv 1$ so that $\bar{b}$ contains full information on whether F and G obey $\mathcal{H}_0$ or $\mathcal{H}_1$. In this setting, H is an unknown nuisance parameter.

To derive local asymptotic results, we require a setup similar to (1.4) and will consider the following type of density for U_1 and U_N, respectively,

$$p_{1N}(u) = 1 + \Delta_N \sqrt{\frac{1 - \eta_N}{\eta_N}} a(u),$$

$$p_{2N}(u) = 1 - \Delta_N \sqrt{\frac{\eta_N}{1 - \eta_N}} a(u),$$

where $a(\cdot)$ is a bounded function such that $\int a(u)\, d\lambda(u) = 0$ while $\Delta_N \to 0$ as $N \to \infty$. The case $\Delta_N = O(N^{-1/2})$ and $a(\cdot) = \bar{b}(\cdot)$ corresponds to Neuhaus' (1982) $\mathcal{H}_0$-contiguous alternatives further exploited in Neuhaus (1987).

Without loss of generality, we will assume $\int a^2(u)\, d\lambda(u) = 1$. Moreover, to give a limpid presentation of our result, we shall restrict attention to the alternatives

$$\begin{aligned} p_{1N}(u) &= 1 + \frac{\rho}{N^\xi} \sqrt{\frac{1 - \eta_N}{\eta_N}} a(u), \\ p_{2N}(u) &= 1 - \frac{\rho}{N^\xi} \sqrt{\frac{\eta_N}{1 - \eta_N}} a(u), \end{aligned} \tag{1.8}$$

where $\rho > 0$, $\xi \in (0, 1/2]$ while

$$a(\cdot) \in \mathcal{A} = \left\{ a(\cdot) \,\Big|\, \sup_{u \in [0,1]} |a(u)| < \infty, \int a(u)\, d\lambda(u) = 0, \int a^2(u)\, d\lambda(u) = 1 \right\}.$$

In (1.8), the function $a(\cdot)$ describes a type of deviation from the null $U[0, 1]$ density while ρ regulates the distance from this null density.

We aim to prove that, under local alternatives (1.8) with any $\rho > 0$ and almost arbitrary $a(\cdot) \in \mathcal{A}$, the J-WL test has the same asymptotic power as the Neyman–Pearson test for testing $p_{1N}(\cdot) \equiv p_{2N}(\cdot) \equiv 1$ against (1.8). This corresponds to a nonparametric version of Stein's postulate. However, there is one basic difference between Stein's approach and the one considered here: we are not considering $\xi = 1/2$, but instead $\xi \in (0, 1/2)$. There are two reasons for this choice. The first one is technical. Under $\xi = 1/2$, the limiting distribution of the J-WL test statistic is the same, except in some special cases, as under the null hypothesis, and thus no conclusion about $\mathcal{H}_0$, $\mathcal{H}_1$ can be drawn. For a discussion, see Inglot and Ledwina (2001a) and our comments in Section 2. The second reason is qualitative and can be phrased as follows. First, let us see what are the consequences of taking $\xi < 1/2$. To have, under such alternatives, asymptotic power converging to some value in (0, 1), the "classical" approach of fixing at some given value the significance level α, finding critical values and then computing powers must be abandoned. Instead, we must first find critical values tending to ∞ as $N \to \infty$ such that power under (1.8) converges to a $\beta \in (0, 1)$ and then compute the resulting significance levels. These will drift toward 0 at a rate related to the value of ξ. To describe an efficiency notion under such a setup, consider (1.8) with $\xi < 1/2$ fixed, $\rho > 0$ and all $a(\cdot)$ belonging to a given subset of $\mathcal{A}$. Denote by $\mathcal{P}_N$ the resulting class of alternatives. We shall say that a test φ_N for $\mathcal{H}_0$ against $\mathcal{H}_1$ is asymptotically efficient under $\mathcal{P}_N$ if for any given density from $\mathcal{P}_N$, the asymptotic power of φ_N converges to a $\beta \in (0, 1)$ and is the same as the asymptotic power of the Neyman–Pearson test for testing $p_{1N}(u) \equiv p_{2N}(u) \equiv 1$ against this alternative from $\mathcal{P}_N$. Simultaneously, both tests should have the same type I error probabilities $\alpha_N = \alpha_N(\xi, \rho)$ tending to 0 as $N \to \infty$. For more details, see Section 4.

A by-product of such an efficiency notion is that such a way of comparing tests is more appealing than the traditional approach in the sense that one has the guarantee that increasing the information contained in the sample increases the precision of the ensuing inference. Namely, the compared tests have the same asymptotic power (separated from 0 and 1) and the same (for each N) probability of the first kind of error vanishing as $N \to \infty$. Note also that there is evidence that in applying such an approach to some classical tests for which the standard asymptotics works as well, one is getting more informative results than those derived under $\xi = 1/2$ [see Inglot, Kallenberg and Ledwina (2000) and Inglot and Ledwina (2001a)].

Note also that Oosterhoff (1969) and Oosterhoff and van Zwet (1972) were the first to consider optimality of tests in the situation where, basically, asymptotic power is separated from 0 and 1 and the significance level tends to 0 as the sample size grows. Further results can be found in Kallenberg (1978), Inglot, Kallenberg and Ledwina (1998, 2000) and Inglot and Ledwina (2001a, b).

This article is organized as follows. In Section 2, we present the adaptive J-WL test. In Section 4, we describe explicitly how the critical values of this test as well as the Neyman–Pearson test have to be chosen and what are the related probabilities of the first kind of error. The basic tools needed to get this are central limit theorems (CLTs) for both statistics derived under (1.8) and moderate deviations derived under $\mathcal{H}_0$, which are presented in Section 3. The basic result on efficiency of the J-WL test is stated and proved in Section 4. The auxiliary results from Section 3 are proved in Section 5.

We close this section by emphasizing that the results we are presenting can be generalized in many directions. However, our primary goal is to explain and justify the efficiency of the J-WL test in the simplest possible setting. The technical tools we have developed to this end are adjusted to this goal.

2. Adaptive test. To define the test statistic, several auxiliary notations are needed. Let $R_i, i = 1, \dots, N$, be the rank of X_i in the pooled sample $X_1, \dots, X_N$. Set $n = N - m$ and define

$$c_{Ni} = \sqrt{\frac{mn}{N}} \begin{cases} \dfrac{1}{m}, & \text{as } i = 1, \dots, m, \\ -\dfrac{1}{n}, & \text{as } i = m+1, \dots, N. \end{cases} \tag{2.1}$$

Let $\phi_0(u) \equiv 1, \phi_1(u), \phi_2(u), \dots$ be the orthonormal Legendre polynomials on [0, 1]. Given $k = 1, 2, \dots$, define

$$T_k = \sum_{j=1}^{k} \left\{ \sum_{i=1}^{N} c_{Ni} \phi_j \left(\frac{R_i - 0.5}{N} \right) \right\}^2. \tag{2.2}$$

Throughout the article, $\{d(N)\}$ will denote a nondecreasing sequence of integers such that $d(N) \to \infty$ as $N \to \infty$. Define the selection rule

$$\begin{aligned} S2 = \min\bigl\{k : 1 \le k \le d(N) : T_k - k \log N \ge T_j - j \log N, \\ 1 \le j \le d(N)\bigr\}. \end{aligned} \tag{2.3}$$

The data driven, or adaptive in the terminology of the present article, test for $\mathcal{H}_0$ against $\mathcal{H}_1$ rejects $\mathcal{H}_0$ for large values of T_{S2}. For an interpretation, a relationship to model selection, basic properties and an extensive simulation study, see Janic-Wróblewska and Ledwina (2000). Here we discuss three aspects of these findings. First, note that by Theorem 1 of Janic-Wróblewska and Ledwina (2000), $S2 \to 1$ in probability under $\mathcal{H}_0$. This immediately implies that under (1.8) with $\xi = 1/2$, the same takes place. Hence, under such contiguous alternatives, T_{S2} behaves asymptotically as T_1, which is the squared Wilcoxon rank-sum statistic. The simulation results reported on Figures 1–3 of Janic-Wróblewska and Ledwina (2000) show, however, that the finite sample performances of T_{S2} and T_1 are completely different. This indicates that such a local approach can be noninformative. The second aspect of their simulation study which is worth mentioning is that the comparison density $\bar{b}$ introduced in (1.7) has, for a very large range of typical probability models for both samples, an expansion in the system $1, \phi_1, \phi_2, \ldots$ with few large Fourier coefficients. Moreover, these large coefficients are essentially concentrated on the first few terms of the expansion. Finally, observe that the empirical critical values and powers are very stable with respect to the choice of $d(N)$ [see Table 1 and Figure 4 of Janic-Wróblewska and Ledwina (2000)]. For more evidence on the empirical behavior of similar constructions, see Kallenberg and Ledwina (1997). The above remarks motivate the present study as well as the scope of the basic Theorem 4.1.

We close this section by noting that Albers, Kallenberg and Martini (2001) have elaborated a counterpart of T_{S2} for testing $\mathcal{H}_0$ against some restricted class of alternatives.

3. Notation and auxiliary results. Recall that we are considering the transformed observations $U_i = H(X_i)$ with H given in (1.6). The joint distributions of $U_1, \ldots, U_N$ under (1.8) and $\mathcal{H}_0$ will be denoted by P_N and P_0, respectively. Obviously under $\mathcal{H}_0$, $U_1, \ldots, U_N$ are i.i.d., each obeying the uniform distribution on $(0, 1)$.

Given a bounded function $a(\cdot) \in \mathcal{A}$, set

$$\begin{aligned} \|a\| &= \left\{ \int a^2(u)\, d\lambda(u) \right\}^{1/2}, \\ \|a\|_\infty &= \sup_{u \in [0,1]} |a(u)|, \\ \hat{a}_j &= \int a(u)\phi_j(u)\, d\lambda(u), \\ \hat{a} &= (\hat{a}_1, \hat{a}_2, \ldots), \\ |\hat{a}|_k &= \left\{ \sum_{j=1}^{k} \hat{a}_j^2 \right\}^{1/2}, \qquad k = 1, 2, \ldots. \end{aligned} \tag{3.1}$$

With this notation, we state the following results regarding the behavior of T_{S2}.

THEOREM 3.1. *Let P_N be defined via* (1.8) *with $\xi \in (0, 1/2)$ and $\rho > 0$. Assume that $d(N) \to \infty$, $N^{-1}[d(N)]^{20} \to 0$, $N^{-2\xi}[d(N)]^9 \to 0$ and $N^{\xi - 1/2} \times [d(N) \log N] \to 0$ as $N \to \infty$. Then*

$$\lim_{N \to \infty} P_N\left(\frac{T_{S2} - \rho^2 N^{1-2\xi} |\hat{a}|_{d(N)}^2}{2\rho N^{1/2-\xi} |\hat{a}|_{d(N)}} \le x \right) = \Phi(x), \qquad x \in \mathbb{R}, \tag{3.2}$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal random variable.

THEOREM 3.2. *Let $x_N \in \mathbb{R}_+$ be such that $x_N = O(N^{-\delta})$ for some $\delta \in (1/4, 1/2)$. Assume $d(N) \to \infty$ in such a way that $\lim_{N\to\infty} x_N^{\varepsilon}[d(N)]^3 = 0$ for some $\varepsilon \in (0,1)$ and $Nx_N^2 \geq d(N)$ for N sufficiently large. Then for any $\eta \in [\varepsilon, 1)$,*

$$\begin{aligned} &P_0(T_{S2} \geq Nx_N^2) \\ &\quad \leq \exp\{-\tfrac{1}{2}Nx_N^2 + 2Nx_N^{2+\eta}[d(N)]^3 + d(N)[\log(Nx_N^2)]\} \end{aligned} \tag{3.3}$$

for all sufficiently large N.

Set now

$$e_{iN}^{(0)} = E_{P_0}(\log p_{iN}(U)), \qquad v_{iN}^2 = E_{P_0}(\log^2 p_{iN}(U)), \qquad i = 1, 2,$$

and $v_N^2 = mv_{1N}^2 + nv_{2N}^2$, and define a standardized version of the Neyman–Pearson test statistic for testing $\mathcal{H}_0$ against (1.8) by

$$V_N^{(1)} = \frac{1}{v_N}\left\{\sum_{i=1}^{m} \log p_{1N}(U_i) - me_{1N}^{(0)} + \sum_{i=m+1}^{N} \log p_{2N}(U_i) - ne_{2N}^{(0)}\right\}.$$

The corresponding results about the behavior of $V_N^{(1)}$ are as follows.

THEOREM 3.3. *Under P_N of (1.8) with $\xi \in (0, 1/2)$ and $\rho > 0$, we have*

$$\lim_{N\to\infty} P_N\big(V_N^{(1)} - \sqrt{N}b^{(1)}(P_N) \leq x\big) = \Phi(x), \qquad x \in \mathbb{R}, \tag{3.4}$$

where $b^{(1)}(P_N)$ is explicitly given by (5.25) and satisfies $b^{(1)}(P_N) = \rho N^{-\xi} + O(N^{-2\xi})$. In particular, for $\xi \in (1/4, 1/2)$, it holds that

$$\lim_{N\to\infty} P_N\big(V_N^{(1)} - N^{1/2-\xi}\rho \leq x\big) = \Phi(x), \qquad x \in \mathbb{R}.$$

THEOREM 3.4. *Let $x_N \in \mathbb{R}_+$ be such that $x_N \to 0$ and $Nx_N^2 \to \infty$. Set $\sigma_N^2 = \mathrm{Var}_{P_0} V_N^{(1)}$. Then*

$$P_0\big(V_N^{(1)} \geq \sqrt{N}x_N\big) = \exp\left\{-\frac{Nx_N^2}{2\sigma_N^2} + O\left(\frac{Nx_N^3}{\sigma_N^3}\right) + O\left(\log\left(\frac{Nx_N^2}{\sigma_N^2}\right)\right)\right\}. \tag{3.5}$$

In particular, if $x_N = O(N^{-\xi})$ with $\xi \in (1/4, 1/2)$, then

$$P_0\big(V_N^{(1)} \geq \sqrt{N}x_N\big) = \exp\left\{-\frac{Nx_N^2}{2} + o(N^{1/2-\xi})\right\}. \tag{3.6}$$

4. On efficiency of adaptive test. The techniques we are using enable us to get efficiency results for $\xi \in (1/4, 1/2)$ (see Remark 4.2). Therefore, in this section we restrict consideration to this case. By Theorem 3.3 we have

$$\lim_{N\to\infty} P_N\big(V_N^{(1)} - N^{1/2-\xi}\rho \le x\big) = \Phi(x), \qquad x \in \mathbb{R}.$$

Consequently, to have local power of the test rejecting for large values of $V_N^{(1)}$ lying in (0, 1), we define a critical region as

$$\mathcal{C}_{N,k_1}^{(1)} = \{V_N^{(1)} - N^{1/2-\xi}\rho \ge k_1\}, \qquad k_1 \in \mathbb{R}.$$

Then $P_N(\mathcal{C}_{N,k_1}^{(1)}) \to 1 - \Phi(k_1)$. The related probability of first kind of error is given by $P_0(\mathcal{C}_{N,k_1}^{(1)})$. By (3.6) of Theorem 3.4, we have

$$P_0\big(\mathcal{C}_{N,k_1}^{(1)}\big) = \exp\biggl\{-\frac{\rho^2}{2}N^{1-2\xi} - k_1 N^{1/2-\xi}\rho + o(N^{1/2-\xi})\biggr\}. \tag{4.1}$$

This gives the rate at which the size of $\mathcal{C}_{N,k_1}^{(1)}$ vanishes as $N \to \infty$.

We will apply a similar argument to T_{S2}. For ease of presentation, we focus herein on the special case where $d(N) = O(\log N)$ and the function $a(\cdot)$ appearing in (1.8) has a finite Fourier expansion in the system $1, \phi_1, \phi_2, \dots$. This means that for some $K \ge 1$,

$$a(u) = \sum_{j=1}^{K} \hat{a}_j \phi_j(u),$$

where $\hat{a}_j$ is defined in (3.1). The more general situation is briefly discussed in Remark 4.1.

Under the above restrictions, Theorem 3.1 reads as

$$\lim_{N\to\infty} P_N\biggl(\frac{T_{S2} - \rho^2 N^{1-2\xi}}{2\rho N^{1/2-\xi}} \le x\biggr) = \Phi(x), \qquad x \in \mathbb{R}.$$

Define a critical region for this standardized version of T_{S2} by

$$\mathcal{C}_{N,k_2}^{(2)} = \biggl\{\frac{T_{S2} - \rho^2 N^{1-2\xi}}{2\rho N^{1/2-\xi}} \ge k_2\biggr\}, \qquad k_2 \in \mathbb{R}.$$

Then $P_N(\mathcal{C}_{N,k_2}^{(2)}) \to 1 - \Phi(k_2)$ and the related probability of first kind of error is given by

$$\alpha_N^{(2)} = \alpha_N^{(2)}(\xi, \rho) = P_0\big(\mathcal{C}_{N,k_2}^{(2)}\big).$$

Obviously, to fulfill the first postulate of our efficiency notion presented in the Introduction, that is, to guarantee that the two tests yield the same asymptotic power, one must have $k_1 = k_1(N) = k_2 + o_N$, where $\{o_N\}$ stands for an arbitrary real sequence tending to 0 as $N \to \infty$ [see (4.3) below]. To fulfill the second postulate of efficiency, one has to show that there exists such a sequence $\{o_N\}$ for which $\alpha_N^{(2)} = P_0(\mathcal{C}^{(1)}_{N,k_2+o_N})$, that is, for each N the tests with critical regions $\mathcal{C}^{(2)}_{N,k_2}$ and $\mathcal{C}^{(1)}_{N,k_2+o_N}$ have the same size [see (4.2)].

THEOREM 4.1. *Assume P_N obeys* (1.8) *for some $\xi \in (1/4, 1/2)$, $\rho > 0$ and arbitrary $a(\cdot)$ having a finite expansion in the system $1, \phi_1, \phi_2, \ldots$. Suppose $d(N) \to \infty$ and $d(N) = O(\log N)$. Then, for any $k_2 \in \mathbb{R}$, there exists a real sequence $\{o_N\}$, $o_N \to 0$ as $N \to \infty$, such that for tests with critical regions $\mathcal{C}^{(2)}_{N,k_2}$ and $\mathcal{C}^{(1)}_{N,k_2+o_N}$ it holds for N sufficiently large that*

$$\begin{aligned} P_0(\mathcal{C}^{(2)}_{N,k_2}) &= P_0(\mathcal{C}^{(1)}_{N,k_2+o_N}) \\ &= \exp\left\{-\frac{\rho^2}{2}N^{1-2\xi} - k_2\rho N^{1/2-\xi} + o(N^{1/2-\xi})\right\} \end{aligned} \tag{4.2}$$

and

$$\lim_{N\to\infty} P_N(\mathcal{C}^{(2)}_{N,k_2}) = \lim_{N\to\infty} P_N(\mathcal{C}^{(1)}_{N,k_2+o_N}) = 1 - \Phi(k_2). \tag{4.3}$$

PROOF. The argument patterns the proof of Theorem 5.1 in Inglot, Kallenberg and Ledwina (1998) and Theorem 3.3(3) in Inglot and Ledwina (2001a). We provide it here to give an idea of this argument and to indicate possibilities of some variants of Theorem 4.1.

Take $x_N^2 = N^{-2\xi}\rho^2 + 2\rho N^{-1/2-\xi}k_2$. Then $x_N^2 = N^{-2\xi}\rho^2[1 + o(1)]$ and, by the definition, $\alpha_N^{(2)} = P_0(T_{S2} \geq N x_N^2)$. Applying Theorem 3.2 with $\eta = \varepsilon$ and $\varepsilon > (1/2 - \xi)/\xi$, one gets

$$\alpha_N^{(2)} \leq \exp\{-\tfrac{1}{2}N^{1-2\xi}\rho^2 - k_2\rho N^{1/2-\xi} + o(N^{1/2-\xi})\}. \tag{4.4}$$

To get information on the relationship between k_1 and k_2, take now $\bar{x}_N = b^{(1)}(P_N)/2 = \rho N^{-\xi}/2 + O(N^{-2\xi})$ and apply (3.6). This yields

$$P_0\big(V_N^{(1)} \geq \sqrt{N}\bar{x}_N\big) = \exp\{-\tfrac{1}{8}N^{1-2\xi}\rho^2 + o\big(N^{1/2-\xi}\big)\}. \tag{4.5}$$

Since for large N, (4.5) is greater than the right-hand side of (4.4), we infer that there exists $k_1 = k_1(N)$ such that $k_1(N) > -\sqrt{N}\bar{x}_N$ and

$$\begin{aligned} &P_0\big(V_N^{(1)} - 2\sqrt{N}\bar{x}_N \geq k_1(N)\big) \\ &\quad = \alpha_N^{(2)} \\ &\quad = P_0\big(V_N^{(1)} - \sqrt{N}b^{(1)}(P_N) \geq k_1(N)\big). \end{aligned} \tag{4.6}$$

Hence, we have the critical region $\mathcal{C}^{(1)}_{N,k_1(N)}$ of level $\alpha_N^{(2)}$. Moreover, by Theorem 3.3, $P_N(\mathcal{C}^{(1)}_{N,k_1(N)}) - [1 - \Phi(k_1(N))] \to 0$. Hence, for sufficiently large N, it has to hold that

$$k_1(N) \le k_2 + o(1), \tag{4.7}$$

as $\mathcal{C}^{(1)}_{N,k_1(N)}$ corresponds to the most powerful test. Now, we apply Theorem 3.4 to the last expression in (4.6). Take $x_N^* = N^{-1/2}k_1(N) + b^{(1)}(P_N) = N^{-1/2}k_1(N) + N^{-\xi}\rho + O(N^{-2\xi})$. By (4.7) and $k_1(N) > -\sqrt{N}\bar{x}_N$, $x_N^* = O(N^{-\xi})$. Hence, by (4.6), (3.6) and the assumption $\xi \in (1/4, 1/2)$,

$$\begin{aligned} &P_0(V_N^{(1)} \ge \sqrt{N}x_N^*) \\ &\quad = \alpha_N^{(2)} \\ &\quad = \exp\left\{-\frac{1}{2}N^{1-2\xi}\rho^2 - \rho k_1(N)N^{1/2-\xi} - \frac{[k_1(N)]^2}{2} + o(N^{1/2-\xi})\right\}. \end{aligned} \tag{4.8}$$

To get further information on k_2 and $k_1(N)$, compare the right-hand side of (4.8) with the bound in (4.4). This yields the inequality

$$\tfrac{1}{2}N^{\xi-1/2}[k_1(N)]^2 + k_1(N)\rho - \rho k_2 + o(1) \ge 0. \tag{4.9}$$

The minimum of the parabola in (4.9) is attained at $-N^{1/2-\xi}\rho$ and, by the previous argument, $k_1(N) \ge -N^{1/2-\xi}\rho/2$ for large N. Hence we get $k_1(N) \ge k_2 + o(1)$. This and (4.7) prove the existence of a sequence $\{o_N\}$ that satisfies both requirements of the theorem. $\square$

REMARK 4.1. The assumption of a finite expansion for $a(\cdot)$ allows us to skip the term $|\hat{a}|^2_{d(N)}$ in the definition of $\mathcal{C}^{(2)}_{N,k_2}$. Another way to reach the same result is to assume that $|\hat{a}|^2_{d(N)}$ is sufficiently close to $|\hat{a}|^2_\infty = 1$. For this, one can use some results on the rate of decay of $1 - |\hat{a}|^2_{d(N)}$ for smooth functions given in Barron and Sheu [(1991), Section 7]. However, one then has to consider $d(N)$ increasing faster than $O(\log N)$ and, by the methods given here, optimality can be established only for ξ in a subinterval of $(1/4, 1/2)$, depending on the rate of growth of $d(N)$ and the smoothness of $a(\cdot)$. For an illustration of this in the case of testing uniformity, see Inglot and Ledwina [(2001a), Theorem 3.1(3)]. One can also leave $|\hat{a}|^2_{d(N)}$ in the definition of $\mathcal{C}^{(2)}_{N,k_2}$, but then, when comparing the resulting $\alpha_N^{(2)}$ and $\alpha_N^{(1)}$, some information on the magnitude of $|a|_{d(N)}$ is needed or, alternatively, we must allow the size of the test to depend on $|\hat{a}|_{d(N)}$. For an example of this in the case of testing uniformity, see Inglot, Kallenberg and Ledwina [(1998), Theorem 5.1].

The empirical behavior of $S2$ under moderate sample sizes indicates that T_{S2} can detect, with high probability, alternatives corresponding to $\bar{b}$ with large first few Fourier coefficients. In this context, it is worthwhile to observe that the results collected on pages 1365–1366 of Barron and Sheu (1991) show that smooth functions from $L_2(\lambda)$ can be very precisely approximated by linear combinations of the first few Legendre polynomials. Consequently, the setup of Theorem 4.1 is most natural for the present application.

REMARK 4.2. The condition $\xi \in (1/4, 1/2)$ in Theorem 4.1 is assumed to get the desired result without further technical work. Indeed, observe first that it is possible to have a more precise expression for (3.5) [see (5.27) and Book (1976)]. However, getting an exponential inequality more precise than (3.3) seems to be far more complicated. Therefore, when equating $\alpha_N^{(1)}$ and $\alpha_N^{(2)}$, we are comparing as few terms as possible. From (4.4), (4.8) and (4.9) it is seen that we can achieve the same asymptotic powers and the same rates of decay of sizes when the remainders arising in (4.4) and (4.8) are indeed $o(N^{1/2-\xi})$. For this, we need to assume $\xi > 1/4$. Obviously we cannot exclude that another line of argument could give a similar result for $\xi \leq 1/4$. In this context, note that better results can be achieved under another notion of optimality. Namely, one can consider optimality (or efficiency) in the sense of asymptotic equality of sample sizes of two tests with powers converging to a $\beta \in (0, 1)$ and levels tending to 0 at the same rate. Such a notion, called intermediate efficiency, was introduced by Kallenberg (1983) and extended by Inglot (1999). It has been applied by Inglot and Ledwina (1996) and Inglot (1999) to prove optimality of some data-driven tests for uniformity. For an easy exposition, see Inglot and Ledwina (2001a). Such an approach could be applied here leading, under $d(N) = O(\log N)$, to the optimality of T_{S2} for any $\xi \in (0, 1/2)$.

5. Proofs.

PROOF OF THEOREM 3.1. The proof of Theorem 3.1 will be given in steps. Two of these are stated below as separate theorems. They concern the behavior of $S2$ and a CLT for T_k under P_N. The proofs of these results exploit some ideas developed in Inglot and Ledwina (1996) coupled with Hájek's (1968) projection method.

We start by introducing some auxiliary rank statistics and their approximation as proposed by Hájek (1968). Throughout the rest of the article, we omit in the integrals the notation λ for the Lebesgue measure. For any $t = 1, 2, \ldots$, set

$$\mathcal{S}_t = \sum_{i=1}^{N} c_{Ni}\phi_t\left(\frac{R_i}{N+1}\right), \qquad s_t = E_{P_N}\mathcal{S}_t, \qquad s = (s_1, s_2, \ldots) \tag{5.1}$$

and introduce the independent random variables $\mathcal{Z}_{it}, i = 1, \ldots, N$, according to formula (4.28) of Hájek (1968). In our application, his X_i are replaced by our $U_i = H(X_i)$, his c_i correspond to our c_{Ni} of (2.1) and instead of his $F_i(x)$, we have

$$F_{1N}(x) = \int_0^x p_{1N}(u)\,du, \qquad i = 1, \ldots, m,$$

$$F_{2N}(x) = \int_0^x p_{2N}(u)\,du, \qquad i = m+1, \ldots, N,$$

where $p_{1N}(u)$ and $p_{2N}(u)$ are given in (1.8). Consequently, $N^{-1}\sum_{i=1}^{N} F_{iN}(x) = x$ and we have

$$\mathcal{Z}_{it} = \begin{cases} -c_{N1}\int_0^1 [I(U_i \le x) - F_{1N}(x)]\phi_t'(x)\,dF_{2N}(x), & \text{as } i = 1, \dots, m, \\ -c_{NN}\int_0^1 [I(U_i \le x) - F_{2N}(x)]\phi_t'(x)\,dF_{1N}(x), & \text{as } i = m+1, \dots, N, \end{cases} \tag{5.2}$$

where $I(A)$ denotes the indicator function of the event A. By (4.26) of Hájek (1968) we get

$$E_{P_N}\left(\mathcal{S}_t - s_t - \sum_{i=1}^{N}\mathcal{Z}_{it}\right)^2 \le M(\phi_t)N^{-1}, \tag{5.3}$$

where, by pages 339 and 340 of Hájek (1968) and the relationships [see Sansone (1959)] $\|\phi_t'\|_\infty = c_1 t^{5/2}$, $\|\phi_t''\|_\infty = c_2 t^{9/2}$ for some absolute constants c_1 and c_2, we have

$$M(\phi_t) \le ct^{18} \tag{5.4}$$

for some positive constant c. Moreover, (5.2) easily yields

$$\mathcal{Z}_{it} = c_{Ni}[\phi_t(U_i) + \rho N^{-\xi} t^{5/2}\mathcal{O}(1)], \qquad i = 1, \dots, N, \tag{5.5}$$

where $\mathcal{O}(1)$ stands for a bounded everywhere random variable.

Now we define a deterministic counterpart to $S2$ of (2.3) by

$$\begin{aligned} \ell(N) = \min\{k : 1 \le k \le d(N), |s|_k^2 - \mu^2 k \log N \ge |s|_j^2 - \mu^2 j \log N, \\ j = 1, \dots, d(N)\}, \end{aligned} \tag{5.6}$$

where s is defined in (5.1), $|s|_k^2 = \sum_{j=1}^{k} s_j^2$ and $\mu > 1$ is a constant defined in the proof of the following theorem.

THEOREM 5.1. *Let P_N be defined via* (1.8) *with $\rho > 0$ and $\xi \in (0, 1/2)$, and assume that $N^{-1}[d(N)]^{20} \to 0$ with $N^{-2\xi}[d(N)]^5 \to 0$. Then*

$$P_N(S2 < \ell(N)) \to 0 \qquad \text{as } N \to \infty.$$

PROOF. For simplicity, write ℓ instead of $\ell(N)$. Moreover, for any real sequence $x = (x_1, x_2, \dots)$ and any $1 \le j < k$, set

$$|x|_{jk} = \left\{\sum_{t=j+1}^{k} x_t^2\right\}^{1/2}. \tag{5.7}$$

Observe now that (2.2) and (2.3) yield

$$P_N(S2 < \ell) \le \sum_{j=1}^{\ell-1} P_N\big(T_\ell - T_j \le (\ell - j)\log N\big). \tag{5.8}$$

Put $\pi_j = P_N(T_\ell - T_j \le (\ell - j)\log N)$. Definition (5.6) implies $|s|_{j\ell} \ge \mu \times \sqrt{(\ell - j)\log N}$ for $j = 1, \dots, \ell$. This together with the triangle inequality implies

$$\begin{aligned}\pi_j \le P_N\Bigg(\Bigg\{\sum_{t=j+1}^{\ell}\Bigg[\sum_{i=1}^{N} c_{Ni}\phi_t\Big(\frac{R_i - 0.5}{N}\Big) - s_t\Bigg]^2\Bigg\}^{1/2} \\ \ge (\mu - 1)\sqrt{(\ell - j)\log N}\Bigg).\end{aligned} \tag{5.9}$$

The rest of the proof consists of three steps. First, we approximate $\sum_{i=1}^N c_{Ni} \times \phi_t((R_i - 0.5)/N)$ by $\mathcal{S}_t$. Then, to $\mathcal{S}_t - s_t$ we apply Hájek's approximation $\sum_{i=1}^N \mathcal{Z}_{it}$. Finally, to estimate the tails of $\sum_{i=1}^N \mathcal{Z}_{it}$, we apply Yurinskii's (1976) inequality.

Throughout, we use c to denote various positive constants, the values of which might change from line to line. By (5.9),

$$\begin{aligned}\pi_j \le{}& P_N\Bigg(\Bigg\{\sum_{t=j+1}^{\ell}\Bigg[\sum_{i=1}^{N} c_{Ni}\phi_t\Big(\frac{R_i - 0.5}{N}\Big) - \mathcal{S}_t\Bigg]^2\Bigg\}^{1/2} \ge \frac{1}{2}(\mu - 1)\sqrt{(\ell - j)\log N}\Bigg) \\ &+ P_N\Bigg(\Bigg\{\sum_{t=j+1}^{\ell}[\mathcal{S}_t - s_t]^2\Bigg\}^{1/2} \ge \frac{1}{2}(\mu - 1)\sqrt{(\ell - j)\log N}\Bigg).\end{aligned}$$

Since $|\phi_t(x) - \phi_t(y)| \le ct^{5/2}|x - y|$, $\eta_N \to \eta \in (0, 1)$ and $[d(N)]^{20} = o(N)$, the first component in the above sum is 0 for N large enough.

Set $y_j = \frac{1}{2}(\mu - 1)\sqrt{(\ell - j)\log N}$. By (5.8), (5.9) and the above, for N large enough,

$$\begin{aligned}P_N(S2 < \ell) &\le \sum_{j=1}^{\ell-1} P_N\Bigg(\Bigg\{\sum_{t=j+1}^{\ell}[\mathcal{S}_t - s_t]^2\Bigg\}^{1/2} \ge y_j\Bigg) \\ &\le \sum_{j=1}^{\ell-1}\sum_{t=j+1}^{\ell} P_N\Bigg(\Bigg[\mathcal{S}_t - s_t - \sum_{i=1}^{N}\mathcal{Z}_{it}\Bigg]^2 \ge \Bigg[\frac{y_j}{2\sqrt{d(N)}}\Bigg]^2\Bigg) \\ &\quad + \sum_{j=1}^{\ell-1} P_N\Bigg(\Bigg\{\sum_{t=j+1}^{\ell}\Bigg[\sum_{i=1}^{N}\mathcal{Z}_{it}\Bigg]^2\Bigg\}^{1/2} \ge \frac{1}{2}y_j\Bigg).\end{aligned} \tag{5.10}$$

Markov's inequality, (5.3) and (5.4), the definition of y_j and the assumption $N^{-1}[d(N)]^{20} \to 0$ imply that the first component of (5.10) tends to 0 as $N \to \infty$.

Introduce now the vectors $\mathcal{Z}_i(\ell) = (\mathcal{Z}_{i1}, \dots, \mathcal{Z}_{i\ell})$ and $\mathcal{Z} = \sum_{i=1}^N \mathcal{Z}_i(\ell)$. By construction, $\mathcal{Z}_1(\ell), \dots, \mathcal{Z}_N(\ell)$ are independent with $E_{P_N}\mathcal{Z}_i(\ell) = 0$ [see Hájek (1968), page 340]. The second component of (5.10) can be rewritten as

$$\sum_{j=1}^{\ell-1} P_N\big(|\mathcal{Z}|_{j\ell} \geq \tfrac{1}{2} y_j\big). \tag{5.11}$$

To estimate the terms in (5.11), we apply Yurinskii's [(1976), Corollary, page 491] inequality. To this end, observe that by (5.5) and the assumption $N^{-2\xi}[d(N)]^5 \to 0$, we get, for large N,

$$E_{P_N}\big(|\mathcal{Z}_i(\ell)|_{j\ell}\big)^2 \leq \begin{cases} 2c_{N1}^2(\ell - j), & i = 1, \dots, m, \\ 2c_{NN}^2(\ell - j), & i = m+1, \dots, N. \end{cases}$$

Analogously, for large N,

$$|\mathcal{Z}_i(\ell)|_{j\ell} \leq \begin{cases} c\big\{c_{N1}^2(\ell - j)\ell\big\}^{1/2}, & i = 1, \dots, m, \\ c\big\{c_{NN}^2(\ell - j)\ell\big\}^{1/2}, & i = m+1, \dots, N. \end{cases}$$

Set $\mathcal{H} = c[\max\{|c_{N1}|, |c_{NN}|\}]\, d(N), b_1^2 = \cdots = b_m^2 = c_{N1}^2(\ell - j), b_{m+1}^2 = \cdots = b_N^2 = c_{NN}^2(\ell - j)$. Applying Yurinskii's (1976) inequality with this $\mathcal{H}$ and these b_i^2's, we get $B_N^2 = b_1^2 + \cdots + b_N^2 = (\ell - j)$ and

$$P_N\bigg(|\mathcal{Z}|_{j\ell} \geq \frac{1}{2} y_j\bigg) \leq 2\exp\bigg\{-\frac{1}{16}(\mu - 1)^2(\log N)\bigg(1 - c\frac{d(N)}{\sqrt{N}}\sqrt{\log N}\bigg)\bigg\}.$$

So there exists $\mu > 1$ such that, for N sufficiently large,

$$\sum_{j=1}^{\ell-1} P_N\bigg(|\mathcal{Z}|_{j\ell} \geq \frac{1}{2} y_j\bigg) \leq \frac{2}{N}.$$

This concludes the proof of Theorem 5.1. $\square$

Our next auxiliary result is as follows.

THEOREM 5.2. *Let P_N be defined via* (1.8) *with $\rho > 0$ and $\xi \in (0, 1/2)$. For $\hat{a}$ given in* (3.1), *let $k(N)$ be a deterministic sequence such that, for N large enough, $|\hat{a}|_{k(N)} > \delta$ for some positive δ. Assume $N^{-1}[k(N)]^{20} \to 0$ while $N^{\xi - 1/2}[k(N)] \to 0$ and $N^{-2\xi}[k(N)]^9 \to 0$. Then,*

$$\lim_{N\to\infty} P_N\bigg(\frac{T_{k(N)} - |s|_{k(N)}^2}{2|s|_{k(N)}} \leq x\bigg) = \Phi(x), \qquad x \in \mathbb{R}.$$

PROOF. In what follows, we supress the dependence of $k(N)$ on N. Also throughout, as before, c is a running positive absolute constant. Set now

$$\mathcal{D}_t = \sum_{i=1}^{N} c_{Ni}\left[\phi_t\left(\frac{R_i - 0.5}{N}\right) - \phi_t\left(\frac{R_i}{N+1}\right)\right] \quad \text{and} \quad \mathcal{Z}_t = \sum_{i=1}^{N} \mathcal{Z}_{it},$$

where $\mathcal{Z}_{1t}, \dots, \mathcal{Z}_{Nt}$ are given by (5.2). We split $T_k = T_{k(N)}$ [see (2.2)] into several terms. For brevity, we consider at once the standardized version of T_k. Simple algebra yields

$$\frac{T_k - |s|_k^2}{2|s|_k} = \mathcal{L}_k + \sum_{j=1}^{4} \mathcal{R}_{jk},$$

where

$$\mathcal{L}_k = (|s|_k)^{-1} \sum_{t=1}^{k} s_t \mathcal{Z}_t, \qquad \mathcal{R}_{1k} = (2|s|_k)^{-1} \sum_{t=1}^{k} (\mathcal{D}_t)^2,$$

$$\mathcal{R}_{2k} = (|s|_k)^{-1} \sum_{t=1}^{k} \mathcal{D}_t \mathcal{S}_t, \qquad \mathcal{R}_{3k} = (2|s|_k)^{-1} \sum_{t=1}^{k} [\mathcal{S}_t - s_t]^2,$$

$$\mathcal{R}_{4k} = (|s|_k)^{-1} \sum_{t=1}^{k} s_t [\mathcal{S}_t - s_t - \mathcal{Z}_t].$$

The rest of the proof consists of showing that, under P_N, $\mathcal{L}_k \to_D N(0,1)$ while $\mathcal{R}_{jk} \to 0$ in P_N, $j = 1, \dots, 4$. We start by proving that the remainders are negligible. To this end, we need some information on the magnitude of $|s|_k$. This is provided by the following approximation of s_t [see Hájek (1968), Theorem 4.2 and page 340]. For $t = 1, 2, \dots,$ set

$$r_t = \rho N^{1/2-\xi} \hat{a}_t \qquad \text{with } \hat{a}_t = \int_0^1 a(u)\phi_t(u)\,du. \tag{5.12}$$

Then by (5.4) and (4.27) of Hájek (1968), with $\mu = r_t$,

$$(s_t - r_t)^2 \le M(\phi_t)/N \le c t^{18}/N. \tag{5.13}$$

Hence, for $r = (r_1, r_2, \dots)$ it holds that $||s|_k - |r|_k| \le \{ck^{19}/N\}^{1/2}$. So under our assumptions, $\lim_{N\to\infty} ||s|_k - |r|_k| = 0$.

Consider first $\mathcal{R}_{1k}$. By $|\phi_t(x) - \phi_t(y)| \le c t^{5/2}|x - y|$, for large N we get

$$|\mathcal{D}_t| \le c t^{5/2}/\sqrt{N}. \tag{5.14}$$

Hence, $\mathcal{R}_{1k} \to 0$ in P_N.

For $\mathcal{R}_{2k}$, any positive γ and N large enough, by (5.14) we have

$$P_N\big(|\mathcal{R}_{2k}| \geq 2\gamma\big) \leq P_N\left(\left|\sum_{t=1}^{k} \mathcal{D}_t s_t\right| \geq \gamma |s|_k\right) + \sum_{t=1}^{k} P_N\Big(|\mathcal{S}_t - s_t| \geq \gamma c |s|_k \sqrt{N/k^7}\Big). \tag{5.15}$$

Again using (5.14) and the Schwarz inequality, we infer that the first component on the right-hand side of (5.15) tends to 0.

To get an upper bound for the second component, set $v_t^2 = \mathrm{Var}_{P_N} \mathcal{S}_t$ and $\tilde{\sigma}_t^2 = \mathrm{Var}_{P_N} \mathcal{Z}_t$. By (2.9), (4.28) and (5.6) of Hájek (1968), we have, for N large, $|\tilde{\sigma}_t - v_t| \leq \{M(\phi_t)\}^{1/2}\{\max_{1\leq i\leq N} |c_{Ni}|\} \leq ct^9/\sqrt{N}$.

On the other hand, the constant d defined in (2.14) of Hájek (1968) equals 1 in our application. Hence, (5.13) and (5.17) of Hájek (1968) yield

$$|\tilde{\sigma}_t - 1| \leq ct^{9/2}\|F_{1N} - F_{2N}\|_\infty \leq ct^{9/2}/N^{\xi}. \tag{5.16}$$

So under our assumptions, $|v_t - 1| = o(1)$ and an application of Markov's inequality yields $\mathcal{R}_{2k} \to 0$ in P_N.

Markov's inequality together with $|v_t - 1| = o(1)$, the relationship between $|s|_k$ and $|r|_k$, and the assumptions $N^{\xi-1/2}[k(N)] \to 0$, $|\hat{a}|_{k(N)} > \delta$ imply that $\mathcal{R}_{3k} \to 0$ in P_N. Finally, Schwarz's inequality and (5.3) yield the same conclusion for $\mathcal{R}_{4k}$.

Now we shall prove the asymptotic normality of $\mathcal{L}_k$. Observe that

$$\mathcal{L}_k = \sum_{i=1}^{N} \mathcal{X}_{Ni}, \qquad \text{where } \mathcal{X}_{Ni} = \sum_{t=1}^{k} \frac{s_t}{|s|_k}\mathcal{Z}_{it}.$$

Since by construction, $E_{P_N}\mathcal{Z}_{it} = 0, t = 1, \ldots, k$, we have $E_{P_N}\mathcal{X}_{Ni} = 0$ as well. Moreover, each $\mathcal{Z}_{it}$ is a function of U_i, solely. This implies $\mathcal{X}_{N1}, \ldots, \mathcal{X}_{NN}$ are independent. By (5.8) of Hájek (1968), we have, for N large enough, $|\mathcal{Z}_{it}| \leq c\{\max_{1\leq i\leq N} |c_{Ni}|t^{5/2}\} \leq ct^{5/2}/\sqrt{N}$. Hence, by Schwarz's inequality, $|\mathcal{X}_{Ni}| \leq ck^3/\sqrt{N}$. Set now $\mathcal{B}_N^2 = \mathrm{Var}_{P_N} \sum_{i=1}^{N} \mathcal{X}_{Ni}$. By the above, $\mathcal{G}_N = \sum_{i=1}^{N} E_{P_N}|\mathcal{X}_{Ni}|^3 \leq ck^3\mathcal{B}_N^2/\sqrt{N}$. By Lyapunov's theorem [see Serfling (1980), Section 1.9.3], it is enough to show that $\mathcal{G}_N = o(\mathcal{B}_N^3)$. We shall show that $\mathcal{B}_N^2 = 1 + o(1)$. Since $k^3/\sqrt{N} \to 0$, this will conclude the proof. We have

$$\mathcal{B}_N^2 = \sum_{t=1}^{k} \frac{s_t^2}{|s|_k^2}\tilde{\sigma}_t^2 + \sum_{i=1}^{N} \sum_{\substack{r,t=1\\ r\neq t}}^{k} \frac{s_r s_t}{|s|_k^2} E_{P_N}\mathcal{Z}_{ir}\mathcal{Z}_{it}. \tag{5.17}$$

Due to (5.16) and the assumption $k^9/N^{2\xi} \to 0$, the first component of (5.17) behaves like $1 + o(1)$. Moreover, by (5.5), properties of Legendre polynomials and Schwarz's inequality, we infer $E_{P_N}\mathcal{Z}_{ri}\mathcal{Z}_{ti} \leq c[N^{-2\xi-1}k^5 + N^{-\xi-1}k + N^{-2\xi-1}k^{5/2}]$. Simultaneously, by (5.13) and the imposed assumptions, $|s_r s_t|/|s|_k^2$

is bounded for N large enough. Hence, by $\xi \in (0, 1/2)$ and the assumption $N^{-1}k^{20} \to 0$, the second component of (5.17) is $o(1)$. $\square$

Now, using Theorems 5.1 and 5.2, we shall establish (3.2) of Theorem 3.1. Set

$$\mathcal{E}_x = \left\{ \frac{T_{S2} - |s|^2_{d(N)}}{2|s|_{d(N)}} \le x \right\}.$$

Since $d(N) \to \infty$ and $\|a\| = 1$, Theorem 5.2 can be applied to $T_{d(N)}$, but since $S2 \le d(N)$, we get $\liminf_{N\to\infty} P_N(\mathcal{E}_x) \ge \Phi(x)$. On the other hand, by Theorem 5.1,

$$P_N(\mathcal{E}_x) \le P_N\left(\frac{T_{\ell(N)} - |s|^2_{d(N)}}{2|s|_{d(N)}} \le x \right) + o(1).$$

So, to show that $P_N(\mathcal{E}_x) \to \Phi(x)$, it is enough to prove that $|s|_{d(N)}/|s|_{\ell(N)} \to 1$ and $|s|^2_{\ell(N)d(N)}/|s|_{d(N)} \to 0$. By the definition of $\ell(N)$ it follows that $|s|^2_{\ell(N)d(N)} \le \mu^2 d(N) \log N$. The assumption $N^{\xi - 1/2}[d(N) \log N] \to 0$ implies $d(N) \log N / |s|_{d(N)} \to 0$. This concludes the proof of $\lim_{N\to\infty} P_N(\mathcal{E}_x) = \Phi(x)$. Note that by (5.12) and (5.13), it follows that under our assumptions $|s|_{d(N)}/|r|_{d(N)} \to 1$. By this relationship we get

$$\lim_{N\to\infty} P_N\left(\frac{T_{S2} - |r|^2_{d(N)}}{2|r|_{d(N)}} \le x \right) = \Phi(x), \qquad x \in \mathbb{R}.$$

This is just (3.2). $\square$

PROOF OF THEOREM 3.2. Under $\mathcal{H}_0, U_1 = H(X_1), \dots, U_N = H(X_N)$ [cf. (1.6)] are i.i.d. uniformly distributed on (0, 1). Since $T_{S2} \le T_{d(N)}$, we have

$$P_0(T_{S2} \ge N x_N^2) \le P_0(T_{d(N)} \ge N x_N^2).$$

Set now $\gamma_N = x_N^\eta [d(N)]^3$,

$$\mathfrak{C}_{d(N)} = \sum_{j=1}^{d(N)} \left[\sum_{i=1}^{N} c_{Ni} \phi_j(U_i) \right]^2$$

and

$$\mathcal{R}_{d(N)} = \left\{ \sum_{j=1}^{d(N)} \left[\sum_{i=1}^{N} c_{Ni} \phi_j\left(\frac{R_i - 0.5}{N} \right) - \sum_{i=1}^{N} c_{Ni} \phi_j(U_i) \right]^2 \right\}^{1/2}.$$

The triangle inequality yields, for large N,

$$\begin{aligned} &P_0(T_{d(N)} \ge N x_N^2) \\ &\quad \le P_0(\mathfrak{C}^{1/2}_{d(N)} \ge (1 - \gamma_N) x_N \sqrt{N}) + P_0(\mathcal{R}_{d(N)} \ge \gamma_N x_N \sqrt{N}). \end{aligned} \tag{5.18}$$

To simplify notation, in what follows we skip the dependence of $d(N)$ on N and write simply d. The first component on the right-hand side of (5.18) will be estimated using Yurinskii's (1976) Theorem 3.1. To this end, set $\mathcal{Y}_{ij} = c_{Ni}\phi_j(U_i)$ and $\mathcal{Y}_i(d) = (\mathcal{Y}_{i1}, \dots, \mathcal{Y}_{id})$. Obviously, $\mathfrak{C}_d^{1/2} = |\sum_{i=1}^N \mathcal{Y}_i(d)|_d$. For any $j = 1, \dots, d$, we have $E_{P_0}\mathcal{Y}_{ij} = 0, i = 1, \dots, N, \operatorname{Var}_{P_0} \mathcal{Y}_{ij} = c_{N1}^2$ for $i = 1, \dots, m$ and $\operatorname{Var}_{P_0} \mathcal{Y}_{ij} = c_{NN}^2$ for $i = m+1, \dots, N$. Moreover, $\operatorname{Cov}_{P_0}(\mathcal{Y}_{ij}, \mathcal{Y}_{i\ell}) = 0$ for $j \neq \ell, i = 1, \dots, N$. Hence we get, for any $h \in \mathbb{R}^d$ and $i = 1, \dots, N$,

$$E_{P_0}\big(h^{\mathrm{T}}\mathcal{Y}_i(d)\big)^2 = c_{Ni}^2|h|_d^2 = h^{\mathrm{T}}b_ih \qquad \text{with } b_i = c_{Ni}^2\mathbf{I}_d,$$

where $\mathbf{I}_d$ is the identity matrix of order d. Consequently, by $\|\phi_j\|_\infty \le \sqrt{2j+1}$, for any integer $s \ge 2$ and all $h \in \mathbb{R}^d$, it holds that

$$\big|E_{P_0}\big(h^{\mathrm{T}}\mathcal{Y}_i(d)\big)^s\big| \le (h^{\mathrm{T}}b_ih)\mathcal{H}^{s-2}(|h|_d)^{s-2},$$

where $\mathcal{H} = 2d\max\{|c_{N1}|, |c_{NN}|\}$. Since $B_N = b_1 + \cdots + b_N = \mathbf{I}_d$, we apply Theorem 3.1 of Yurinskii (1976) with $\sigma_1^2 = \cdots = \sigma_d^2 = 1$, $K = 1$ and the above $\mathcal{H}$. To shorten the notation, set for a moment $y_N = (1-\gamma_N)x_N\sqrt{N}$. Assume N is large enough to have $d \ge 4$, $y_N^2 \ge 3d-3$ and $\alpha = \mathcal{H}y_N < 1$. Then for both d even and odd we infer, for N large enough,

$$\begin{aligned} &P_0\left(\left|\sum_{i=1}^N \mathcal{Y}_i(d)\right|_d \ge y_N\right) \\ &\qquad \le \frac{8.731}{\Gamma(d/2)2^{(d-1)/2}}\exp\left\{-\frac{y_N^2}{2}(1+\alpha)^{-1} + (d-1)\log y_N\right\}. \end{aligned}$$

Using the fact that $\log\Gamma(x) \ge -x + (x - 1/2)\log(x)$, the definition of γ_N and a simple argument yields, for N large enough,

$$\begin{aligned} &P_0\big(\mathfrak{C}_d^{1/2} \ge (1-\gamma_N)x_N\sqrt{N}\big) \\ &\qquad \le \exp\left\{-\frac{1}{2}Nx_N^2 + \big(1+o(1)\big)Nx_N^{2+\eta}d^3 + \frac{d}{2}\log(Nx_N^2)\right\}. \end{aligned} \tag{5.19}$$

To deal with the remainder $\mathcal{R}_d$, set

$$\mathcal{U}_{iN}^{(j)} = \phi_j\left(\frac{R_i - 0.5}{N}\right) - \phi_j(U_i).$$

Then

$$\begin{aligned} P_0\big(\mathcal{R}_d \ge \gamma_N x_N\sqrt{N}\big) &\le \sum_{j=1}^d P_0\left(\left|\sum_{i=1}^N c_{Ni}\mathcal{U}_{iN}^{(j)}\right| \ge \frac{\gamma_N x_N\sqrt{N}}{\sqrt{d}}\right) \\ &\le \left\{\frac{\sqrt{d}}{\gamma_N x_N\sqrt{N}}\right\}^{2p}\sum_{j=1}^d E_{P_0}\left|\sum_{i=1}^N c_{Ni}\mathcal{U}_{iN}^{(j)}\right|^{2p} \end{aligned} \tag{5.20}$$

for any positive p. Assume for the moment that p is an integer and satisfies $1 \le p \le \sqrt{N}/2$. By Lemma 2.5 of Hušková (1977), with the correction and implementation given by Kallenberg [(1982), see formula (2.3)], we have

$$E_{P_0}\left|\sum_{i=1}^{N} c_{Ni}\mathcal{U}_{iN}^{(j)}\right|^{2p} \le p^p(4e)^{2p+1}A(m,n,p)E_{P_0}[\mathcal{U}_{1N}^{(j)}]^{2p}, \tag{5.21}$$

where

$$A(m,n,p) = \left[\max\left(1, 2p\sqrt{\frac{mn}{N}}\max\left\{\frac{1}{m},\frac{1}{n}\right\}\right)\right]^{2p}.$$

Since $\|\phi_j'\|_\infty \le cj^{5/2}$, a Taylor expansion yields, for some V_1 between $(R_1 - 0.5)/N$ and U_1,

$$\begin{aligned} E_{P_0}[\mathcal{U}_{1N}^{(j)}]^{2p} &= E_{P_0}\left[\left(\frac{R_1}{N} - \frac{1}{2N} - U_1\right)\phi_j'(V_1)\right]^{2p} \\ &\le [cd^{5/2}]^{2p} E_{P_0}\left(\frac{R_1}{N} - \frac{1}{2N} - U_1\right)^{2p}. \end{aligned}$$

Set now $\omega_1 = N^{-1}(1/2 - U_1)$, $\omega_i = N^{-1}[I(U_i \le U_1) - U_1]$, $i = 2,\dots,N$. Then

$$\frac{R_1}{N} - \frac{1}{2N} - U_1 = \sum_{i=1}^{N}\omega_i,$$

and $\omega_1,\dots,\omega_N$ satisfy the assumption of Lemma 6.1 of Bickel (1974). Hence we get

$$E_{P_0}\left(\sum_{i=1}^{N}\omega_i\right)^{2p} \le N^p(4ep)^p \max_{1\le i\le N} E_{P_0}(\omega_i^{2p}) \le (4ep)^p\frac{2^{2p}}{N^p}. \tag{5.22}$$

Using Hölder's inequality, (5.21) and (5.22) yield, for any real $1 \le p \le \sqrt{N}/2$,

$$E_{P_0}\left|\sum_{i=1}^{N} c_{Ni}\mathcal{U}_{iN}^{(j)}\right|^{2p} \le A(m,n,p)(4e)(16e^{3/2}c)^{2p}p^{2p}\frac{d^{5p}}{N^p}. \tag{5.23}$$

Set $c_1 = 4e$ and $c_2 = 16e^{3/2}c$. By (5.20), (5.23) and the choice of γ_N,

$$P_0(\mathcal{R}_d \ge \gamma_N x_N\sqrt{N}) \le A(m,n,p)c_1 d\left\{\frac{c_2 p}{Nx_N^{1+\eta}}\right\}^{2p}.$$

Take now $p = Nx_N^2/2$. By our choice of x_N, we have $p < \sqrt{N}/2$ and $A(m,n,p) = 1$ for N large enough. This yields

$$P_0(\mathcal{R}_d \ge \gamma_N x_N\sqrt{N}) \le c_1 d\left\{\frac{c_2 x_N^{1-\eta}}{2}\right\}^{Nx_N^2}. \tag{5.24}$$

Since $\eta \in (0,1)$ and $d(N) \le Nx_N^2$, then $P_0(\mathcal{R}_d \ge \gamma_N x_N \sqrt{N}) \le \exp\{-Nx_N^2/2\}$ for N sufficiently large and an application of (5.18), (5.19) and (5.24) concludes the proof of Theorem 3.2. $\square$

PROOF OF THEOREM 3.3. To shorten the argument, we relate (3.4) to an earlier result of Inglot and Ledwina (1996). Set $e_{iN}^{(a)} = E_{P_{iN}} \log p_{iN}(U)$, $i = 1, 2$, $b^{(\mathrm{I})}(p_{1N}) = (e_{1N}^{(a)} - e_{1N}^{(0)})/v_{1N}$, $b^{(\mathrm{II})}(p_{2N}) = (e_{2N}^{(a)} - e_{2N}^{(0)})/v_{2N}$ and define

$$b^{(1)}(P_N) = \frac{mv_{1N}}{\sqrt{N}v_N} b^{(\mathrm{I})}(p_{1N}) + \frac{nv_{2N}}{\sqrt{N}v_N} b^{(\mathrm{II})}(p_{2N}). \tag{5.25}$$

With this notation, we have

$$\begin{aligned} V_N^{(1)} &= \frac{\sqrt{m}v_{1N}}{v_N} \left\{ \sum_{i=1}^{m} \frac{\log p_{1N}(U_i) - e_{1N}^{(a)}}{\sqrt{m}v_{1N}} \right\} \\ &\quad + \frac{\sqrt{n}v_{2N}}{v_N} \left\{ \sum_{i=m+1}^{N} \frac{\log p_{2N}(U_i) - e_{2N}^{(a)}}{\sqrt{n}v_{2N}} \right\} + \sqrt{N} b^{(1)}(P_N). \end{aligned} \tag{5.26}$$

By Proposition 6.6 of Inglot and Ledwina [(1996), page 2000] and using the notation on page 1994, the random variables appearing in braces in (5.26) are asymptotically standard normal. Therefore, to conclude the proof it is enough to study the deterministic expressions that appear in (5.26). To this end, introduce $c_N^{(1)} = \rho N^{-\xi} \sqrt{(1-\eta_N)/\eta_N}$ and $c_N^{(2)} = -\rho N^{-\xi} \sqrt{\eta_N/(1-\eta_N)}$. By Proposition 2.10 of Inglot [(1999), page 494; see also page 503], we have $[b^{(\mathrm{I})}(p_{1N})]^2 = [c_N^{(1)}]^2(1 + O(c_N^{(1)}))$, $[b^{(\mathrm{II})}(p_{2N})]^2 = [c_N^{(2)}]^2(1 + O(c_N^{(2)}))$ and $v_{iN}^2 = [c_N^{(i)}]^2(1 + O(c_{(N)}^{(i)}))$, $i = 1, 2$. Consequently, $v_{1N}^2/v_N^2 = (n/mN)(1 + O(c_N^{(1)}))$, $v_{2N}^2/v_N^2 = (m/nN)(1 + O(c_N^{(2)}))$ and $b^{(1)}(P_N) = \rho N^{-\xi} + O(N^{-2\xi})$. The theorem follows. $\square$

PROOF OF THEOREM 3.4. Statement (3.5) can be deduced from an unpublished article by Book (1976). For convenience, we quote here a version of this result suitable for our setting. Its applicability to the present context follows from some conditions given in Inglot and Ledwina (1996).

THEOREM 5.3 [Book (1976)]. *Let $\{W_{Ni}, 1 \le i \le N, 1 \le N < \infty\}$ be a triangular array of rowwise independent random variables. Assume $EW_{Ni} = 0$ and $EW_{Ni}^2 < \infty$ for all N and i. Put*

$$W_N = \sum_{i=1}^{N} W_{Ni} \quad \text{and} \quad \tau_N^2 = \operatorname{Var} W_N.$$

Further suppose there exist positive constants C', C'' and $0 < B \le \infty$ such that, for all $h, |h| < B, C' \le E\exp\{hW_{Ni}\} \le C''$ for all N and i. Then, for all sequences $\{z_N\}$ of positive numbers such that $z_N \to \infty$ and $z_N/\sqrt{N} \to 0$ as $N \to \infty$,

$$\begin{aligned} P\left(\frac{W_N}{\tau_N} \ge z_N\right) \\ = \varphi(z_N) \\ = (2\pi z_N^2)^{-1/2}\left[1 + O\left(\frac{z_N}{\sqrt{N}}\right)\right]\exp\left\{-\frac{z_N^2}{2} + \frac{z_N^3}{\sqrt{N}}\lambda_N\left(\frac{z_N}{\sqrt{N}}\right)\right\} \end{aligned} \tag{5.27}$$

as $N \to \infty$, where $\lambda_N(z)$ is a power series in z convergent for all sufficiently small values of z, uniformly for all N.

Observe now that

$$V_N^{(1)} = \frac{v_{1N}}{v_N}\sum_{i=1}^{m} Y_{Ni} + \frac{v_{2N}}{v_N}\sum_{i=m+1}^{N} Y_{Ni},$$

where $Y_{Ni} = \{\log p_{1N}(U_i) - e_{1N}^{(0)}\}/v_{1N}$ as $i = 1, \dots, m$ and $Y_{Ni} = \{\log p_{2N}(U_i) - e_{2N}^{(0)}\}/v_{2N}$ as $i = m+1, \dots, N$. Set $\sigma_{iN}^2 = \mathrm{Var}_{P_0} \log p_{iN}(U_1), i = 1, 2$. By Lemma 5.4 of Inglot and Ledwina [(1996); see (5.11) and the last sentence of the proof], $\sigma_{iN}^2/v_{iN}^2 = 1 + O(N^{-2\xi})$. The above and (5.26) imply $\sigma_N^2 = \mathrm{Var}_{P_0} V_N^{(1)} = 1 + O(N^{-2\xi})$. Finally, observe that by Proposition 5.12 of Inglot and Ledwina (1996), the variables $Y_{Ni}, i = 1, \dots, N$, satisfy Book's assumptions. Hence the same holds for the components of $V_N^{(1)}$. Therefore, (5.27) is applicable and $P_0(V_N^{(1)} \ge \sigma_N x_N \sqrt{N}) = \varphi(x_N\sqrt{N})$. Taking in particular $x_N = O(N^{-\xi})$, we have $\sigma_N = 1 + O(x_N^2)$ and (5.27) implies (3.5). $\square$

Acknowledgments. We are indebted to the referee and an Associate Editor for constructive comments. This article was partially written while the second author was visiting the Laboratoire de Probabilités et Statistique, Université Montpellier II. The hospitality of the university is gratefully acknowledged. The second author also thanks Marie Hušková, Tadeusz Inglot and Wilbert Kallenberg for useful discussions.

REFERENCES

ALBERS, W., KALLENBERG, W. C. M. and MARTINI, F. (2001). Data driven rank tests for classes of tail alternatives. *J. Amer. Statist. Assoc.* **96** 685–696.

BAJORSKI, P. (1992). Max-type rank tests in the two-sample problem. *Zastos. Mat*. **21** 371–385.

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

BEHNEN, K. (1975). The Randles–Hogg test and an alternative proposal. *Comm. Statist*. **4** 203–238.

BEHNEN, K. (1981). Nichtparametrische Statistik: Zweistichproben Rangtests. *Z. Angew. Math. Mech.* **61** T203–T212.

BEHNEN, K. and NEUHAUS, G. (1983). Galton's test as a linear rank test with estimated scores and its local asymptotic efficiency. *Ann. Statist.* **11** 588–599.

BEHNEN, K. and NEUHAUS, G. (1989). *Rank Tests with Estimated Scores and Their Application*. Teubner, Stuttgart.

BICKEL, P. J. (1974). Edgeworth expansions in nonparametric statistics. *Ann. Statist.* **2** 1–20.

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Johns Hopkins Univ. Press.

BOOK, S. A. (1976). The Cramér–Feller–Petrov large deviation theorem for triangular arrays. Technical report, Dept. Mathematics, California State College, Dominguez Hills.

EUBANK, R. L., LARICCIA, V. N. and ROSENSTEIN, R. B. (1987). Test statistics derived as components of Pearson's phi-squared distance measure. *J. Amer. Statist. Assoc.* **82** 816–825.

FAN, J. (1996). Test of significance based on wavelet thresholding and Neyman's truncation. *J. Amer. Statist. Assoc.* **91** 674–688.

HÁJEK, J. (1962). Asymptotically most powerful rank-order tests. *Ann. Math. Statist.* **33** 1124–1147.

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

HOGG, R. V. and LENTH, R. V. (1984). A review of some adaptive statistical techniques. *Comm. Statist. A—Theory Methods* **13** 1551–1579.

HUŠKOVÁ, M. (1977). The rate of convergence of simple linear rank statistics under hypothesis and alternatives. *Ann. Statist.* **5** 658–670.

HUŠKOVÁ, M. (1984). Adaptive methods. In *Handbook of Statistics* **4**. *Nonparametric Methods* (P. R. Krishnaiah and P. K. Sen, eds.) 347–358. North-Holland, Amsterdam.

HUŠKOVÁ, M. and SEN, P. K. (1985). On sequentially adaptive asymptotically efficient rank statistics. *Sequential Anal.* **4** 125–151.

INGLOT, T. (1999). Generalized intermediate efficiency of goodness-of-fit tests. *Math. Methods Statist*. **8** 487–509.

INGLOT, T., KALLENBERG, W. C. M. and LEDWINA, T. (1998). Vanishing shortcoming of data driven Neyman's tests. In *Asymptotic Methods in Probability and Statistics* (B. Szyszkowicz, ed.) 811–829. North-Holland, Amsterdam.

INGLOT, T., KALLENBERG, W. C. M. and LEDWINA, T. (2000). Vanishing shortcoming and asymptotic relative efficiency. *Ann. Statist.* **28** 215–238. [Correction (2000) **28** 1795.]

INGLOT, T. and LEDWINA, T. (1996). Asymptotic optimality of data-driven Neyman's tests for uniformity. *Ann. Statist.* **24** 1982–2019.

INGLOT, T. and LEDWINA, T. (2001a). Intermediate approach to comparison of some goodness-of-fit tests. *Ann. Inst. Statist. Math.* **53** 810–834.

INGLOT, T. and LEDWINA, T. (2001b). Asymptotic optimality of data driven smooth tests for location–scale family. *Sankhyā Ser. A* **63** 41–71.

JANIC-WRÓBLEWSKA, A. and LEDWINA, T. (2000). Data driven rank test for two-sample problem. *Scand. J. Statist.* **27** 281–297.

KALLENBERG, W. C. M. (1978). *Asymptotic Optimality of Likelihood Ratio Tests in Exponential Families*. Mathematical Centre Tracts **77**. Math. Centrum, Amsterdam.

KALLENBERG, W. C. M. (1982). Cramér type large deviations for simple linear rank statistics. *Z. Wahrsch. Verw. Gebiete* **60** 403–409.

KALLENBERG, W. C. M. (1983). Intermediate efficiency, theory and examples. *Ann. Statist.* **11** 170–182.

KALLENBERG, W. C. M. and LEDWINA, T. (1997). Data-driven smooth tests when the hypothesis is composite. *J. Amer. Statist. Assoc.* **92** 1094–1104.

NEUHAUS, G. (1982). H_0-contiguity in nonparametric testing problems and sample Pitman efficiency. *Ann. Statist.* **10** 575–582.

NEUHAUS, G. (1987). Local asymptotics for linear rank statistics with estimated score functions. *Ann. Statist.* **15** 491–512.

OOSTERHOFF, J. (1969). *Combination of One-sided Statistical Tests.* Mathematical Centre Tracts **28**. Math. Centrum, Amsterdam.

OOSTERHOFF, J. and VAN ZWET, W. R. (1972). The likelihood ratio test for the multinomial distribution. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 31–50. Univ. California Press, Berkeley.

SANSONE, G. (1959). *Orthogonal Functions*. Interscience, New York.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

STEIN, C. (1956). Efficient nonparametric testing and estimation. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 187–195. Univ. California Press, Berkeley.

YURINSKII, V. V. (1976). Exponential inequalities for sums of random vectors. *J. Multivariate Anal.* **6** 473–499.

LABORATOIRE DE PROBABILITÉS ET STATISTIQUE
UNIVERSITÉ MONTPELLIER II
PLACE EUGÈNE BATAILLON
34095, MONTPELLIER, CEDEX 5
FRANCE
E-MAIL: ducharme@math.univ-montp2.fr

INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
UL. KOPERNIKA 18
51-617 WROCŁAW
POLAND
E-MAIL: ledwina@impan.pan.wroc.pl