The Annals of Statistics
1996, Vol. 24, No. 5, 1982–2019

ASYMPTOTIC OPTIMALITY OF DATA-DRIVEN NEYMAN'S TESTS FOR UNIFORMITY[1]

BY TADEUSZ INGLOT AND TERESA LEDWINA

Technical University of Wrocław

Data-driven Neyman's tests resulting from a combination of Neyman's smooth tests for uniformity and Schwarz's selection procedure are investigated. Asymptotic intermediate efficiency of those tests with respect to the Neyman–Pearson test is shown to be 1 for a large set of converging alternatives. The result shows that data-driven Neyman's tests, contrary to classical goodness-of-fit tests, are indeed omnibus tests adapting well to the data at hand.

1. Introduction. Consider testing the hypothesis H: $P = P_0$, P_0 a known continuous distribution on R, against K: $P \neq P_0$, based on n independent observations distributed according to P.

There are many consistent tests for testing H against K. The most popular are Kolmogorov–Smirnov (KS) and Cramér–von Mises (CM) tests. Consistency here means that those tests are capable of detecting any deviation from P_0 provided that the sample size is large enough. However, there is now strong evidence that, in fact, for moderate sample sizes, only a few deviations can be detected by these tests with substantial frequency. First of all this feature is seen in simulations [cf. Quesenberry and Miller (1977), Miller and Quesenberry (1979) and Kim (1992)]. Moreover, there are some theoretical results explaining the behavior of the power function. We shall mention here two kinds of such explanations.

The first is based on the local asymptotic theory and is due to Neuhaus (1976) and Milbrodt and Strasser (1990). It is related to the Hájek and Šidák (1967) method of analysis of the asymptotic power around the hypothesis. The results obtained by Neuhaus (1976) and Milbrodt and Strasser (1990) show how the above-mentioned and some other tests distribute their powers in the space of all alternatives when the sample size is large. From these investigations it follows that there are only very few directions of deviations from the hypothesis for which the tests are of reasonable asymptotic power. The directions are also identified in the papers. Roughly speaking, these directions correspond to some very smooth departures from the null distribution (low-frequency alternatives). Moreover, from the "principal component representation" of the local asymptotic power, it follows that there is only one

---

Received March 1995; revised October 1995.
[1]Research supported by Grant KBN 350 044.
AMS 1991 *subject classifications*. 62G10, 62G20, 62G05, 62A10.
Key words and phrases. Goodness of fit, smooth test, Schwarz's criterion, exponential family, efficiency, log-density estimation, minimum relative entropy estimation, large deviations.

direction with highest asymptotic power that is possible. In each other direction the power is smaller. For "bad" directions the power is close to the significance level. Therefore, it is not surprising that one of the conclusions in Milbrodt and Strasser is that each of the KS and CM tests behaves very much like a parametric test for a one-dimensional alternative and not like a well-balanced test for higher-dimensional alternatives. Thus, at least from a local point of view, the tests do not have the omnibus property usually attributed to them.

The second approach we would like to mention is to investigate the relative efficiency of a given test with respect to the Neyman–Pearson test for an alternative of interest. Such an approach for the KS, CM and other goodness-of-fit tests has been developed by Nikitin (1984, 1995). He used the notion of Bahadur efficiency and has shown that if one restricts attention to one-parameter location or scale alternatives, then, for each of the above-mentioned tests, there is only a single alternative for which the test is locally (under location or scale close to 0) as efficient as the Neyman–Pearson test. Using some results of Inglot and Ledwina (1990) and exploiting the notion of intermediate efficiency, it can be shown that the same conclusion follows. For a justification see subsection 7.7 of the present paper. Some related results concerning the local equivalence of intermediate and Bahadur slopes of a class of goodness-of-fit tests can be found in Koning (1992, 1993).

The mentioned inability of commonly used goodness-of-fit tests to detect higher-frequency alternatives has caused renewed interest in Neyman's (1937) smooth test of fit. For details see Rayner and Best (1989, 1990), Milbrodt and Strasser (1990), Eubank and LaRiccia (1992) and Kaigh (1992). To enlarge the applicability of the original Neyman's test and to make the test consistent against any alternative, some data-driven versions of Neyman's test have recently been proposed by Bickel and Ritov (1992), Eubank and LaRiccia (1992), Kim (1996), Eubank, Hart and LaRiccia (1993), Ledwina (1994) and Kallenberg and Ledwina (1995a). After the first draft of this paper was submitted, we received a manuscript on a related topic by Fan (1996). Extensive simulations presented in Ledwina (1994), Kallenberg and Ledwina (1995a, 1995b), Bogdan (1995) and Bogdan and Ledwina (1996) show that the data-driven Neyman's test proposed in Ledwina (1994) and extended in Kallenberg and Ledwina (1995a) compares very well to classical tests and other competitors.

The present paper provides some theoretical results supporting the nice simulation results for the data-driven test proposed in Ledwina (1994). Namely, for the class of data-driven smooth tests described in Kallenberg and Ledwina (1995a), asymptotic power is investigated via the intermediate efficiency approach. More precisely, for a given member of the class of tests, a large set of alternatives, say P_n's, converging to the null distribution (as the sample size n increases) is found for which the data-driven Neyman's test is asymptotically as efficient as the Neyman–Pearson test for testing uniformity against P_n. It is shown that in this set the actual direction of approach to the null hypothesis is immaterial but the rate of convergence plays a role. So, the data-driven Neyman's tests considered in this paper have qualitatively different asymptotic properties from those of the KS and CM tests.

The outline of the paper is as follows. In Section 2 we state the assumptions, illustrate our main results on asymptotic efficiency and discuss some of their implications. Section 3 contains a modified version of a basic result from Kallenberg (1983) which is a useful tool in our investigation of efficiency. Section 4 presents the intermediate slope of T_S. In Section 5 the intermediate slope of the Neyman–Pearson test for P_0 against P_n is given and discussed. Section 6 contains results on the asymptotic optimality of T_S. The main result of Section 4 (Theorem 4.1) is proved in subsection 7.5. The proof is based on a series of auxiliary results which are collected in subsections 7.1, 7.3 and 7.4. In particular, in subsection 7.1 we develop new results related to Schwarz's rule. Subsection 7.3 deals with properties of information projection of alternatives onto increasing exponential families related to the considered test. In subsection 7.4 the asymptotic behavior of Schwarz's rule designed to select a member of an increasing (with n) exponential family is studied under general (as a rule nonexponential) alternatives P_n converging to the null hypothesis. Subsection 7.2 provides a large deviation result for T_S, which is exploited in Section 4. Subsection 7.7 contains a justification that, for KS and CM, local Bahadur efficiency investigated in Nikitin (1984) is equal to the intermediate efficiency.

2. Notation, assumptions and discussion of results. We start with a formal definition of the class of tests we shall investigate. To motivate further considerations, we find it useful to first mention Neyman's argument which resulted in the Neyman smooth test.

When testing H against K without loss of generality attention can be restricted to i.i.d. $X_1, \dots, X_n$ with values in $[0,1]$. The null hypothesis is then that the X_i are uniform on $[0,1]$. Suppose an alternative has the form

$$g_k(x;\theta) = \exp\left\{\sum_{j=1}^{k} \theta_j \phi_j(x) - \psi_k(\theta)\right\}, \tag{2.1}$$

where $\phi_0, \phi_1, \dots, \phi_k$ are orthonormal in $L_2[0,1]$ with $\phi_0 \equiv 1$ and $\psi_k(\theta)$ is a normalizing constant defined by

$$\psi_k(\theta) = \log \int_0^1 \exp\left\{\sum_{j=1}^{k} \theta_j \phi_j(x)\right\} dx. \tag{2.2}$$

The system of orthonormal Legendre polynomials was Neyman's original choice for the ϕ_j. Neyman's smooth test statistic with k components is then given by

$$T_k = \sum_{j=1}^{k} \left\{ n^{-1/2} \sum_{i=1}^{n} \phi_j(X_i) \right\}^2 \tag{2.3}$$

and is a locally (under θ_i's, $1 \le i \le k$, converging to 0 at the rate $n^{1/2}$) optimal (under some side conditions) test statistic for testing uniformity against $g_k(x;\theta)$. A consequence of the above is that rejecting the null hypothesis when T_k is large is a powerful test provided that the alternative $g_k(x;\theta)$ is close to the distribution of the data at hand. Otherwise the power of Neyman's test can be unsatisfactory. For some evidence, see Neyman (1937), Inglot, Jurlewicz and Ledwina (1990a), Inglot, Kallenberg and Ledwina (1994) and Kallenberg and Ledwina (1995a).

To overcome this difficulty, Ledwina (1994) and Kallenberg and Ledwina (1995a) proposed considering an increasing family of exponential models $g_k(x;\theta)$, $1 \le k \le m_n$, choosing first, via Schwarz's (1978) criterion, in that family the density (with dimension S, say), fitting the data at hand and then applying to the data the test statistic T_S.

Denoting by $\mathscr{L}_k(\theta)$ the log-likelihood based on $X_1, \dots, X_n$ with density $g_k(x;\theta)$ and defining

$$L_k = n \sup_{\theta \in R^k} \mathscr{L}_k(\theta) - \tfrac{1}{2} k \log n, \tag{2.4}$$

Schwarz's criterion produces the data-driven order

$$S = \min\{k : 1 \le k \le m_n,\ L_k \ge L_j,\ 1 \le j \le m_n\}. \tag{2.5}$$

Theorem 4.4 in Kallenberg and Ledwina (1995a) shows that if the growth of m_n is suitably related to oscillations of ϕ_j's $1 \le j \le m_n$, then T_S is consistent against essentially any alternative.

It should be recalled also that, under P_0, $\lim_{n\to\infty} P_0(S = 1) = 1$ [cf. Kallenberg and Ledwina (1995a), Theorem 3.2]. This result implies that if a sequence of alternatives $\{P_n\}$ is contiguous to P_0, then $\lim_{n\to\infty} P_n(S = 1) = 1$ as well. So, under contiguous alternatives, T_S degenerates to T_1 and no interesting conclusions on asymptotic power can follow. Hence, to get nontrivial results, the convergence of P_n to P_0 should be slightly slower than under contiguity. Due to the characterization of contiguity given by Oosterhoff and van Zwet (1979), the above implies we shall restrict attention to P_n's such that

$$\lim_{n\to\infty} H(P_n, P_0) = 0 \quad \text{and} \quad \lim_{n\to\infty} nH^2(P_n, P_0) = \infty, \tag{2.6}$$

where $H(P_n, P_0)$ is the Hellinger distance between P_n and P_0. Note that (2.6) is naturally related to Kallenberg's (1983) notion of intermediate efficiency.

A goal of the present paper is to find alternatives satisfying (2.6) for which T_S is asymptotically as efficient as the Neyman–Pearson test for P_0 against P_n. To this end, we shall show that if P_n is smooth enough to be well approximated by some exponential family $g_k(x;\theta)$, given by (2.1), with k and θ depending on n, and the number of models m_n that are allowed to be chosen with Schwarz's rule (2.5) is well balanced with the magnitude of $H(P_n, P_0)$, then, under such P_n's, the intermediate asymptotic efficiency of T_S to the Neyman–Pearson test is 1. If for some alternatives it holds that the efficiency of T_S to the Neyman–Pearson test is 1, we shall say that T_S is asymptotically optimal under such alternatives or, merely, asymptotically optimal.

Below we state the assumptions we impose on the orthonormal systems [(A1) and (A2)], the number m_n of models [(A3)], the smoothness of alternatives via approximation properties of exponential models (2.1) [(A5)] and the rate of convergence to 0 of $H(P_n, P_0)$ [(A6)]. Assumption (A4) is a technical one.

2.1. *Basic assumptions.* Let $\phi_1(x), \phi_2(x), \dots,\ x \in [0,1]$, be an orthonormal sequence in $L_2[0,1] = L_2([0,1], \lambda)$, λ the Lebesgue measure, satisfying

$$\int_0^1 \phi_j(x)\,dx = 0, \qquad j = 1,2,\dots. \tag{2.7}$$

Assume

$$\sup_x |\phi_j(x)| < \infty, \tag{A1}$$

$$V_k = O(k^{\omega}) \quad \text{for some } \omega \ge 0, \tag{A2}$$

where

$$V_k = \max_{1 \le j \le k} \sup_x |\phi_j(x)|. \tag{2.8}$$

For some r such that $r > \omega + \frac{3}{2}$, let $\{m_n\}$ be a sequence such that

$$m_n \to \infty \quad \text{as } n \to \infty \quad \text{and} \quad m_n = O(n^{1/(2r+1)}). \tag{A3}$$

Assume $\{P_n\}$ is a sequence of distributions on $[0,1]$ possessing densities p_n with respect to λ. By P_0 we shall denote throughout the uniform distribution on $[0,1]$, while $p_0(x) \equiv 1,\ x \in [0,1]$, shall stand for its density. We shall assume that p_n approaches p_0, as n is growing. Moreover, we shall assume

(A4) There exists M such that, for sufficiently large n,

$$e^{-M} \le p_n(x) \le e^M, \qquad x \in [0,1].$$

Now set $\phi = (\phi_1, \dots, \phi_{m_n})$ and let $\circ$ stand for the inner product in R^{m_n}.

(A5) There exists $\theta = (\theta_1, \dots, \theta_{m_n})$ such that

$$\gamma_{m_n} = \|\log p_n - \theta \circ \phi\|_\infty \text{ is bounded,}$$

$$\Delta_{m_n} = \|\log p_n - \theta \circ \phi\|_2 = O\left(\frac{1}{m_n^r}\right),$$

where $\|\cdot\|_\infty$ denotes the supremum norm and $\|\cdot\|_2$ is the $L_2[0,1]$ norm.

(A6) $H(p_n, p_0) \to 0$ as $n \to \infty$ and $m_n^r H(p_n, p_0) \to \infty$.

Recall that, for any two densities p and q with respect to λ, the Hellinger distance $H(p,q)$ is defined by

$$H(p,q) = \left\{\int_0^1 \left(\sqrt{p} - \sqrt{q}\right)^2 dx\right\}^{1/2}. \tag{2.9}$$

REMARK 2.1. The trigonometric basis satisfies (A1) and (A2) with $\omega = 0$. For Legendre polynomials (A1) and (A2) hold with $\omega = \frac{1}{2}$.

REMARK 2.2. Assume that $f_n = \log p_n \in W_2^r$, $r \geq 1$, where W_2^r denotes the Sobolev space of functions f on $[0,1]$ for which $f^{(r-1)}$ is absolutely continuous and $\int_0^1 (f^{(r)}(x))^2\,dx < \infty$. Then (A5) holds for Legendre polynomials [cf. Barron and Sheu (1991), Section 7]. (A5) holds also for a trigonometric basis provided that the extra boundary condition $f_n^{(j)}(0) = f_n^{(j)}(1)$, $1 \leq j \leq r$, is imposed [see Barron and Sheu (1991), Section 7]. θ is then the vector of the first m_n Fourier coefficients of $\log p_n$ with respect to a suitable basis. Moreover, $\gamma_{m_n} = O(m_n^{-r+1})$ for Legendre polynomials and $\gamma_{m_n} = O(m_n^{-r+1/2})$ for trigonometric functions.

REMARK 2.3. In what follows the subscript n in m_n shall be suppressed in the notation.

REMARK 2.4. A set of $\{p_n\}$'s satisfying (A4)–(A6) shall be denoted by $\mathscr{P}_{mr}$. This is the set of sequences of alternatives to uniformity under which we shall investigate asymptotic optimality of the test based on T_S (asymptotic optimality of T_S for short).

2.2. *Asymptotic optimality of T_S: an illustrative result.* In this section we shall present one theorem stating asymptotic optimality of T_S under a given choice of the growth rate of m in S. The result follows from Corollary 6.12. For simplicity, we will consider alternatives of the form

$$p_n(x) = 1 + n^{-\xi} g(x), \qquad x \in [0,1],\ \xi \geq 0, \tag{2.10}$$

where $\int_0^1 g(x)\,dx = 0$.

We shall compare T_S and the most powerful test under p_n via the intermediate efficiency approach. This approach can be considered as intermediate between the asymptotic Pitman and the exact Bahadur approach. That is, when calculating the intermediate efficiency, the significance level, say α_n, tends to 0 but not too fast, the alternative p_n tends to the hypothesis but also not too fast and α_n and p_n agree in such a way that the power at p_n stays away from 0 and 1. Moreover, similar to the Pitman and Bahadur concepts, the intermediate efficiency can be interpreted in terms of the sample sizes required to attain, with α_n level tests, the same power at an alternative p_n.

Given p_n, let α_n be defined by

$$\alpha_n = \exp\{-[nD(p_n \| p_0)][1 + o(1)]\}, \tag{2.11}$$

where $D(p_n \| p_0)$ is the Kullback–Leibler distance between p_n and p_0. Recall that, restricting attention to (2.10), the Pitman efficiency is calculated under $\xi = \frac{1}{2}$, while the Bahadur efficiency deals with $\xi = 0$. Then $D(p_n \| p_0) = O(n^{-1})$ and $D(p_n \| p_0) = \text{constant}$, respectively. This illustrates the fact that (2.11) defines any intermediate significance level in between an exponentially small one (the Bahadur efficiency) and a constant one (the Pitman efficiency).

Let

$$(2.12)\quad V_n^{(0)} = \left\{ n \operatorname{Var}_{P_0} \log p_n(X) \right\}^{-1/2} \left\{ \sum_{i=1}^{n} \left[\log p_n(X_i) - E_{P_0} \log p_n(X) \right] \right\}$$

be a standardized version of the logarithm of the Neyman–Pearson test statistic for P_0 against P_n.

For given n and α_n as in (2.11), let $\mathscr{T}_{n;\,\alpha_n}$ and $\Omega_{n;\,\alpha_n}$ be test functions on the significance level α_n rejecting P_0 for large values of T_S and $V_n^{(0)}$, respectively. The critical value of T_S, defining precisely α_n, is given in Corollary 6.11.

Finally, set

$$N_{T,V}(n, p_n) = \inf\{N\colon E_{P_n} \Omega_{N+k;\,\alpha_n} \ge E_{P_n} \mathscr{T}_{n;\,\alpha_n} \text{ for all } k = 0, 1, 2, \dots\}.$$

THEOREM 2.5. *Assume that g has finite number of non-zero Fourier coefficients in the cosine basis. For $r \ge 2$ let $m = cn^{1/(2r+2)}$, where c is an arbitrary constant. Then, for any*

$$\xi \in \left(\frac{3}{2(2r+2)}, \frac{r}{2r+2} \right)$$

and related p_n given by (2.10), *it holds that*

$$(2.13)\quad \lim_{n \to \infty} \frac{N_{T,V}(n, p_n)}{n} = 1.$$

The same statement holds for the basis of Legendre polynomials provided $r \ge 3$ and

$$\xi \in \left(\frac{2}{2r+2}, \frac{r}{2r+2} \right).$$

REMARK 2.6. The left-hand side of (2.13) is the intermediate efficiency of T_S with respect to $V_n^{(0)}$. If the efficiency equals 1 we say that T_S is asymptotically efficient or asymptotically optimal. So, for any g satisfying the assumptions of Theorem 2.5, T_S is asymptotically optimal. The alternatives (2.10) are called the contamination model, while the function g is called the direction of the alternative (2.10). For instance, taking as g succeeding functions $\phi_1, \phi_2, \dots$ from the cosine or Legendre systems, (2.13) holds. So, approaching p_0 in any of directions ϕ_j, $j = 1, 2, \dots,$ T_S is asymptotically optimal. Recall that goodness-of-fit tests like KS or CM are asymptotically optimal in one direction, only. T_S does not have this drawback. This implies, for example, also that T_S is more efficient than KS (or CM) in all directions ϕ_j's but possibly one in which the tests are equally efficient. Moreover, the above holds for a large range of rates of convergence of p_n to p_0, provided that the α_n's are related to these rates via (2.11). In particular, for large r the interval $(n^{-r/(2r+2)}, n^{-(3+2\omega)/2(2r+2)})$ [with $\omega = 0$ and $\omega = \frac{1}{2}$ for cosine and Legendre basis, respectively] corresponds to almost all intermediate ranges in between the Pitman and Bahadur efficiency. Similar statements hold true also for other standard alternative models considered in the literature (cf. Corollary 6.12).

3. Intermediate efficiency and asymptotic optimality: definitions and tools. In this section we present a generalization of the notion of intermediate efficiency introduced and investigated by Kallenberg (1983). We also give a related version of Kallenberg's (1983) result on calculating intermediate efficiency.

Let $\{\mathbf{P}_\pi,\ \pi \in \Pi\}$ be a family of distributions on a sample space $(\mathscr{X}, \mathscr{B})$. Suppose the hypothesis H_0: $\pi \in \Pi_0$ has to be tested against H_1: $\pi \in \Pi_1 \subset \Pi - \Pi_0$, where Π_0 and Π_1 are given subsets of Π.

Let α_n be a sequence of levels such that $\alpha_n > 0$ and

$$\lim_{n\to\infty} \alpha_n = \lim_{n\to\infty} n^{-\tau} \log \alpha_n = 0 \quad \text{for some } 0 < \tau \le 1. \tag{3.1}$$

Let $\{\pi_n\}$ be a sequence of alternatives with

$$\lim_{n\to\infty} H(\pi_n, \Pi_0) = 0, \qquad \lim_{n\to\infty} nH^2(\pi_n, \Pi_0) = \infty. \tag{3.2}$$

Here $H(\pi, \Pi_0) = \inf_{\pi_0 \in \Pi_0} H(\pi, \pi_0)$ and $H(\pi, \pi_0)$ denotes the Hellinger distance between the probability measures $\mathbf{P}_\pi$ and $\mathbf{P}_{\pi_0}$.

Let $V_n^{(i)}$, $i = 1, 2$, be some test statistics and let $\mathscr{V}_{n;\alpha_n}^{(i)}$ be a sequence of test functions for H_0 against H_1, rejecting H_0 for large values of $V_n^{(i)}$. Assume

$$\sup_{\pi_0 \in \Pi_0} E_{\pi_0} \mathscr{V}_{n;\alpha_n}^{(i)} \le \alpha_n, \qquad i = 1, 2, \tag{3.3}$$

and

$$0 < \liminf_{n\to\infty} E_{\pi_n} \mathscr{V}_{n;\alpha_n}^{(2)} \le \limsup_{n\to\infty} E_{\pi_n} \mathscr{V}_{n;\alpha_n}^{(2)} < 1. \tag{3.4}$$

Define

$$\begin{aligned} &N_{V^{(2)}, V^{(1)}}(n, \pi_n) \\ &\quad = \inf\Big\{N\colon E_{\pi_n} \mathscr{V}_{N+k;\alpha_n}^{(1)} \ge E_{\pi_n} \mathscr{V}_{n;\alpha_n}^{(2)} \ \text{ for all } k = 0, 1, 2, \dots\Big\} \end{aligned} \tag{3.5}$$

and set

$$e_{V^{(2)}, V^{(1)}} = \lim_{n\to\infty} \frac{N_{V^{(2)}, V^{(1)}}(n, \pi_n)}{n}. \tag{3.6}$$

REMARK 3.1. Recall that elements of $\{\pi_n\}$ come from Π_1, the given set of alternatives. All notions below are related to the set Π_1. However, to simplify the notation and formulations, we shall not repeat it throughout. In most applications the form of Π_1 shall follow from the context. Moreover, in this paper the limit (3.6) shall be calculated only for some family $\mathscr{P}$ of sequences $\{\pi_n\}$, $\pi_n \in \Pi_1$, which obey (3.2). Therefore we shall introduce the following notion of intermediate efficiency.

DEFINITION. If the limit (3.6) exists and does not depend on the special sequences $\{\pi_n\}$ and $\{\alpha_n\}$ such that (3.1)–(3.4) hold, we say that the asymptotic τ intermediate efficiency of $\mathscr{V}^{(2)}_{n;\alpha_n}$ (or $V^{(2)}$) with respect to $\mathscr{V}^{(1)}_{n;\alpha_n}$ (or $V^{(1)}$) in the family $\mathscr{P}$ equals $e_{V^{(2)},V^{(1)}}$.

The following lemma is an immediate generalization of Lemma 2.1 and Remark 2.1 in Kallenberg (1983).

LEMMA 3.2. *Let* $V_n^{(i)}$, $i = 1, 2$, *be test statistics rejecting the hypothesis for large values of* $V_n^{(i)}$. *Suppose that there exist positive constants* $c^{(i)}$ *and positive functions* $b^{(i)}(\pi)$, $\pi \in \Pi_1$ $i = 1, 2$, *such that,*

(i)(a) *for all sequences* $\{\mathbf{P}_{\pi_n}\}$, $\pi_n \in \Pi_1$, *of alternatives satisfying* (3.2),

$$\lim_{n\to\infty} \mathbf{P}_{\pi_n}\big(1-\varepsilon \le n^{-1/2}V_n^{(1)}/b^{(1)}(\pi_n) \le 1+\varepsilon\big) = 1 \text{ for each } \varepsilon > 0,$$

(i)(b) *for all sequences* $\{\pi_n\} \in \mathscr{P}$

$$\lim_{n\to\infty} \mathbf{P}_{\pi_n}\big(1-\varepsilon \le n^{-1/2}V_n^{(2)}/b^{(2)}(\pi_n) \le 1+\varepsilon\big) = 1 \quad \textit{for each } \varepsilon > 0,$$

$$\text{(ii)} \qquad \lim_{n\to\infty}\left\{-\frac{1}{nx_n^2}\log \sup_{\pi_0\in\Pi_0} \mathbf{P}_{\pi_0}\big(V_n^{(i)} \ge \sqrt{n}\,x_n\big)\right\} = c^{(i)}$$

$$\textit{if } \sqrt{n}\,x_n \to \infty \textit{ and } x_n = o(n^{-(1-\tau)/2})$$

and for all sequences $\{\pi_n\} \in \mathscr{P}$

$$\text{(iii)} \qquad \lim_{n\to\infty} \frac{c^{(2)}}{c^{(1)}}\left\{\frac{b^{(2)}(\pi_n)}{b^{(1)}(\pi_n)}\right\}^2 = \mathscr{E} \textit{ exists}, \qquad 0 < \mathscr{E} < \infty.$$

Then the asymptotic τ *intermediate efficiency of* $V_n^{(2)}$ *with respect to* $V_n^{(1)}$ *in the family* $\mathscr{P}$ *exists and is given by* $e_{V^{(2)},V^{(1)}} = \mathscr{E}$.

REMARK 3.3. Making an analogy to the Bahadur slope, for any statistic, say $V_n^{(*)}$, satisfying (i) [(a) or (b)] and (ii) of Lemma 3.2 with a positive $c^{(*)}$ and a positive $b^{(*)}$, we shall call $c^{(*)}\{b^{(*)}(\cdot)\}^2$ its intermediate slope.

REMARK 3.4. A proof of Lemma 3.2 consists of rewriting line by line the proof of Lemma 2.1 of Kallenberg (1983) and is here omitted. For details see Inglot (1996). However, we find it useful to mention here that the proof relies on checking that for any particular sequence of alternatives under consideration, under (3.1)–(3.4), conditions (i)–(iii) imply that (3.6) exists and that

$$\text{(3.7)} \qquad \lim_{n\to\infty} \frac{N_{V^{(2)},V^{(1)}}(n,\pi_n)}{n} = \lim_{n\to\infty} \frac{c^{(2)}}{c^{(1)}}\left\{\frac{b^{(2)}(\pi_n)}{b^{(1)}(\pi_n)}\right\}^2.$$

Moreover, to guarantee that (3.4) holds, the rates of convergence of $\{p_n\}$ and α_n have to be related. To this end, in Kallenberg's paper first the significance level is chosen. Consequently, the rate τ and the related critical value are determined. Then attention is focused on those alternatives which satisfy

(3.2) and (3.4). So, in such an approach, the rate of convergence in (3.2) depends on the choice of τ. Obviously, one can reverse the order of the above steps. Namely, first fix a rate of convergence in (3.2) and find a critical value for which (3.4) holds. Then calculate the related significance level and determine the value of τ. In subsections 6.2 and 6.3 of the present paper just the second approach shall be applied.

REMARK 3.5. Letting $\mathscr{P}$ consist of all $\{\pi_n\}$ satisfying (3.2), choosing $\tau = \frac{1}{3}$ and $\tau = 1$ corresponds to Kallenberg's notion of asymptotic i-efficiency and strong asymptotic i-efficiency, respectively.

We shall say that a test statistic V_n is asymptotically τ efficient if its asymptotic τ intermediate efficiency in a family $\mathscr{P}$ with respect to the Neyman–Pearson test exists and equals 1.

If V_n is τ efficient for some τ, and some family $\mathscr{P}$ we shall say that V_n is asymptotically optimal.

4. Intermediate slope of T_S. Recall that $\mathscr{P}_{mr}$ is a set of sequences $\{p_n\}$'s satisfying (A4)–(A6). In addition to (A4)–(A6), assume now that for each sequence $\{p_n\} \in \mathscr{P}_{mr}$ there exists $\beta \in (\omega + \frac{3}{2}, r)$ such that

(A7) $$m^{\beta} H(p_n, p_0) = O(1) \quad \text{as } n \to \infty$$

and

(A8) $$(n^{-1} \log n)^{-\beta/(2\beta+1)} H(p_n, p_0) \to \infty \quad \text{as } n \to \infty.$$

The following key theorem shows that T_S fulfills condition (i)(b) of Lemma 3.2.

THEOREM 4.1. *Under* (A1)–(A8) *for any* $\varepsilon \in (0,1)$,

(4.1) $$\lim_{n\to\infty} P_n\left(\left|\frac{n^{-1}T_S}{2D(p_n\|p_0)} - 1\right| \le \varepsilon\right) = 1.$$

REMARK 4.2. In view of the famous Bahadur–Raghavachari theorem on exact slopes [cf. Theorem 7.5 in Bahadur (1971)], the function $b^{(2)}(\cdot) = 2D(\cdot\|p_0)$ is the most obvious candidate to satisfy (i)(b) for T_S, if one aims at proving the asymptotic optimality of T_S. On the other hand, the structure of T_S is not explicitly related to such a choice. A naive choice would be $b^{(\bullet)}(p_n) = \sum_{j=1}^{S} a_j^2$, where $a_j = E_{P_n}\phi_j(X)$. But the function $b^{(\bullet)}(\cdot)$ is not a deterministic one! Anyway, in subsection 7.4, using some results of subsection 7.1, we show that it is possible to find a deterministic dimension l [cf. (7.32) and (7.33)] such that $b^{(\diamond)}(p_n) = \sum_{j=1}^{l} a_j^2$ satisfies (i)(b) for T_S (cf. Theorem 7.20). However, a relation in between $b^{(2)}(\cdot)$ and $b^{(\diamond)}(\cdot)$ is not obvious, also. To show that $b^{(\diamond)}(\cdot)$ and $b^{(2)}(\cdot)$ differ by a negligible quantity (cf. Theorem 7.21), we express both functions via Euclidean norms of the vector θ^*, where θ^* is the vector of parameters of the information projection of the alternative p_n into the exponential family $g_m(x;\theta)$ [cf. (7.19), (7.51) and (7.52)]. To get the last result, several properties of the projection are elaborated in subsection 7.3. A formal proof of Theorem 4.1 is given in subsection 7.5.

The following corollary to Theorem 7.9 shows that condition (ii) of Lemma 3.2 is satisfied.

PROPOSITION 4.3. *Suppose* (A1)–(A8) *hold. Let* $\{x_n\}$ *be a sequence of positive numbers such that* $x_n \to 0$, $nx_n^2 \to \infty$ *and* $x_n = o(m^{-(1+2\omega)/2})$. *Then*

$$\lim_{n\to\infty} \frac{1}{nx_n^2} \log P_0\big(T_S \ge nx_n^2\big) = -\frac{1}{2}. \tag{4.2}$$

PROOF. To get (4.2), it is enough to check the assumptions of Theorem 7.9 for $d(n) = m_n$. Under the choice of x_n it suffices to prove $n^{-1}(\log n) m u_m^2 \to 0$ as $n \to \infty$. However, a much stronger statement is true. Namely, $n^{-1}(\log n) m^{2(\beta-\omega)} u_m^2 \to 0$. Indeed, since $n^{-1}(\log n) m^{2(\beta-\omega)} u_m^2 = O(1)\{m^{\beta} H(p_n, p_0)\}^{2+1/\beta} \log n \{n[H(p_n, p_0)]^{2+1/\beta}\}^{-1}$, by (A7) and (A8) the conclusion follows. □

Now set

$$V_n^{(2)} = \{T_S\}^{1/2}. \tag{4.3}$$

Theorem 4.1 and Proposition 4.3 yield the following result.

COROLLARY 4.4. *If* (A1)–(A8) *hold, then* $V_n^{(2)}$ *satisfies* (i)(b) *and* (ii) *of Lemma* 3.2 *with* $c^{(2)} = \frac{1}{2}$ *and* $b^{(2)}(\cdot) = \{2D(\cdot\|p_0)\}^{1/2}$.

5. Intermediate slope of the Neyman–Pearson test for $\mathbf{P_0}$ against $\mathbf{P_n}$. For a given alternative P_n with a density p_n, set

$$e_{0n} = E_{P_0} \log p_n(X), \qquad \sigma_{0n}^2 = \mathrm{Var}_{P_0} \log p_n(X) \tag{5.1}$$

and

$$V_n^{(0)} = \big(\sqrt{n}\,\sigma_{0n}\big)^{-1} \sum_{i=1}^{n} \{\log p_n(X_i) - e_{0n}\}. \tag{5.2}$$

$V_n^{(0)}$ is the standardized version of the logarithm of the Neyman–Pearson test statistic for P_0 against P_n.

First we shall prove that $V_n^{(0)}$ obeys (i)(a) of Lemma 3.2 with $b^{(0)}(\cdot)$ such that

$$b^{(0)}(p_n) = \sigma_{0n}^{-1}\{D(p_n\|p_0) - e_{0n}\}. \tag{5.3}$$

To this end, we shall use the bounds given in the following two lemmas.

LEMMA 5.1. *Assume that* (A4) *holds. Then, for sufficiently large* n,

$$H^2(p_n, p_0) \le D(p_n\|p_0) - e_{0n} \le C_1(M) H^2(p_n, p_0), \tag{5.4}$$

where $C_1(M) = \{8 + 8\cosh M\}^{1/2}$.

PROOF. Since $D(p_n \| p_0) - e_{0n} = \int_0^1 (p_n(x) - 1)\log p_n(x)\,dx$, it is enough to show

$$(x-1)\log x \le C_1(M)\left(\sqrt{x}-1\right)^2 \tag{5.5}$$

for $x \in (e^{-M}, e^M)$. To get (5.5), it is enough to check that the function $g(x) = C_1(M)(\sqrt{x}-1)^2 - (x-1)\log x$ obeys $g(1) = g'(1) = 0$ and $g''(x) > 0$ for $x \in (e^{-M}, e^M)$. Since the functions in (5.5) are continuous and $e_{0n} \le 0$, the proof is completed. □

By a similar argument one obtains the following result.

LEMMA 5.2. *Assume* (A4) *is satisfied. Then, for sufficiently large* n,

$$\int_0^1 p_n(x)\log^2 p_n(x)\,dx \le C_2(M)H^2(p_n, p_0), \tag{5.6}$$

where $C_2(M) = 4\exp\{3M/2\}$.

THEOREM 5.3. *Suppose* $\{p_n\}$ *obeys* (A4), $H^2(p_n, p_0) \to 0$ *and* $nH^2(p_n, p_0) \to \infty$ *as* $n \to \infty$. *Then, for any* $\varepsilon \in (0,1)$,

$$\lim_{n\to\infty} P_n\left(\left|\frac{V_n^{(0)}}{\sqrt{n}\,b^{(0)}(p_n)} - 1\right| \le \varepsilon\right) = 1. \tag{5.7}$$

PROOF. To prove (5.7), it is enough to show that

$$\begin{aligned}\Lambda_n = P_n\Bigg(\Bigg|\frac{1}{n}\sum_{i=1}^n \{\log p_n(X_i) - D(p_n\|p_0)\}\Bigg| \\ \ge \varepsilon\{D(p_n\|p_0) - e_{0n}\}\Bigg) \to 0\end{aligned} \tag{5.8}$$

as $n \to \infty$. To this end, we shall apply Proposition 7.17 with $Y_i = \log p_n(X_i) - D(p_n\|p_0)$, $\mathscr{K} = 2M$, $b_i^2 = E_{P_n}Y_i^2$ and $B_n^2 = nb_1^2$. Hence

$$\Lambda_n \le 2\exp\left\{\frac{-\varepsilon^2 n\left[D(p_n\|p_0) - e_{0n}\right]^2}{2\left[b_1^2 + 1.62\,\varepsilon M(D(p_n\|p_0) - e_{0n})\right]}\right\}. \tag{5.9}$$

Since, by (5.6), $b_1^2 \le C_2(M)H^2(p_n, p_0)$, an application of (5.4) and the assumption $nH^2(p_n, p_0) \to \infty$ proves (5.8). □

The next result allows us to replace σ_{0n} in $V_n^{(0)}$ and $b^{(0)}(\cdot)$ by a slightly simpler expression.

LEMMA 5.4. *If* (A4) *holds and* $H(p_n, p_0) \to 0$ *as* $n \to \infty$, *then*

$$\sigma_{0n}^2\left\{\int_0^1 \log^2 p_n(x)\,dx\right\}^{-1} \to 1 \quad \text{as } n \to \infty. \tag{5.10}$$

PROOF. It is enough to show

$$\left\{\int_0^1 \log p_n(x)\,dx\right\}^2 \left\{\int_0^1 \log^2 p_n(x)\,dx\right\}^{-1} \to 0 \quad \text{as } n \to \infty. \tag{5.11}$$

To get (5.11), we shall apply the inequalities

$$\log y \geq (y-1) - C_3(M)\left(\sqrt{y}-1\right)^2, \tag{5.12}$$

$$\log^2 y \geq C_4(M)\left(\sqrt{y}-1\right)^2, \tag{5.13}$$

where $y \in [e^{-M}, e^M]$, $C_3(M) = 2e^{M/2}$, $C_4(M) = 2e^{-M}$, which can be proved similarly to (5.5).

By (5.12) and Jensen's inequality,

$$-C_3(M)H^2(p_n, p_0) \leq \int_0^1 \log p_n(x)\,dx \leq 0. \tag{5.14}$$

Moreover, (5.13) implies

$$\int_0^1 \log^2 p_n(x)\,dx \geq C_4(M)H^2(p_n, p_0). \tag{5.15}$$

Hence the left-hand side of (5.11) does not exceed $2e^{2M}H^2(p_n, p_0)$ and, by $H(p_n, p_0) \to 0$, the conclusion follows. □

Set

$$v_{0n}^2 = \int_0^1 \log^2 p_n(x)\,dx. \tag{5.16}$$

REMARK 5.5. In view of (5.10), under the assumptions of Theorem 5.3, an equivalent form of the Neyman–Pearson statistic $V_n^{(0)}$ given by

$$V_n^{(1)} = \left(\sqrt{n}\,v_{0n}\right)^{-1} \sum_{i=1}^n \{\log p_n(X_i) - e_{0n}\} \tag{5.17}$$

satisfies (5.7) with

$$b^{(1)}(p_n) = v_{0n}^{-1}\{D(p_n \| p_0) - e_{0n}\}. \tag{5.18}$$

Now we shall establish some conditions under which $b^{(1)}(p_n)/b^{(2)}(p_n) \to 1$ as $n \to \infty$, where $b^{(2)}(p_n) = \{2D(p_n \| p_0)\}^{1/2}$ (cf. also Corollary 4.4).

THEOREM 5.6. *Assume* (A4) *holds and* $\|p_n - p_0\|_\infty \to 0$ *as* $n \to \infty$. *Then*

$$b^{(1)}(p_n)\{2D(p_n \| p_0)\}^{-1/2} \to 1 \quad \textit{as } n \to \infty. \tag{5.19}$$

PROOF. Observe that $D(p_n \| p_0) - e_{0n} = \int_0^1 (p_n(x) - 1)^2 \psi_1(p_n(x))\,dx$, $2D(p_n \| p_0) = \int_0^1 (p_n(x) - 1)^2 \psi_2(p_n(x))\,dx$ and $\int_0^1 \log^2 p_n(x)\,dx = \int_0^1 (p_n(x) - 1)^2 \psi_3(p_n(x))\,dx$, where $\psi_1(y) = (y-1)^{-1}\log y$, $\psi_2(y) = 2(y-1)^{-2}[y \log y - y + 1]$, $\psi_3(y) = (y-1)^{-2}\log^2 y$ and $y \in [e^{-M}, e^M]$. Setting $\psi_i(1) = 1$, $i = 1, 2, 3$, we see that the ψ_i's are continuous with continuous derivatives.

Therefore, the uniform convergence of p_n to p_0 implies that each of the considered expressions can be written as $\{1 + o(1)\}\int_0^1 (p_n(x) - 1)^2\,dx$. Hence (5.19) follows. Moreover,

$$\begin{aligned} D(p_n \| p_0) - e_{0n} &= O\big(H^2(p_n, p_0)\big), \\ D(p_n \| p_0) &= O\big(H^2(p_n, p_0)\big), \\ v_{0n}^2 &= O\big(H^2(p_n, p_0)\big). \end{aligned} \tag{5.20}$$

□

REMARK 5.7. Using some results collected in subsection 7.3 (cf. Lemmas 7.11 and 7.12), it follows easily that $\|p_n - p_0\|_\infty \to 0$ provided that $\gamma_{m_n} \to 0$. This, however, is satisfied in standard situations (cf. Remark 2.2). Hence Theorem 5.6 holds if (A1)–(A5) are satisfied and we additionally assume that in (A5) $\gamma_{m_n} \to 0$ as $n \to \infty$.

Having checked (i)(a), we shall state some conditions ensuring that (ii) of Lemma 3.2 holds for $V_n^{(1)}$, also. For this purpose we shall use Theorem 5.8 below, being an immediate consequence of a Cramér-type large deviations result obtained by Book (1976) [cf. Lemma 4.1 in Jurečková, Kallenberg and Veraverbeke (1988)].

Set

$$Y_{ni} = v_{0n}^{-1}\{\log p_n(X_i) - e_{0n}\}. \tag{5.21}$$

Then $V_n^{(1)} = n^{-1/2}\sum_{i=1}^n Y_{ni}$.

THEOREM 5.8. *If there exist positive constants* B, C', C'' *such that, for all complex* h, $|h| < B$,

$$C' \le |E_{P_0} e^{hY_{n1}}| \le C'' \quad \textit{for all } n, \tag{5.22}$$

then for all sequences $\{x_n\}$ *of positive numbers such that* $x_n \to 0$ *and* $nx_n^2 \to \infty$ *as* $n \to \infty$ *we have*

$$\lim_{n\to\infty} \left(nx_n^2\right)^{-1} \log P_0\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n Y_{ni} \ge x_n\sqrt{n}\right) = -\frac{1}{2}. \tag{5.23}$$

Theorem 5.8 and Remark 5.5 imply the following result.

COROLLARY 5.9. *Suppose* $H^2(p_n, p_0) \to 0$, $nH^2(p_n, p_0) \to \infty$, $\{p_n\}$ *obeys* (A4) *and* (5.22). *Then* $V_n^{(1)}$ *satisfies* (i)(a) *and* (ii) *of Lemma* 3.2 *with* $c^{(1)} = \frac{1}{2}$ *and* $b^{(1)}(\cdot)$ *as in* (5.18).

The rest of this section deals with sufficient conditions for (5.22) to hold.

Set

$$k_n(x) = |\log p_n(x)|\{H(p_n, p_0)\}^{-1}.$$

PROPOSITION 5.10. *Assume* (A4) *holds,* $H(p_n, p_0) \to 0$ *and* $nH^2(p_n, p_0) \to \infty$ *as* $n \to \infty$. *Further assume there exists* $\delta > 0$ *such that*

$$\sup_n \int_0^1 \exp\{\delta k_n(x)\}\,dx < \infty. \tag{5.24}$$

Then (5.22) *holds. In particular*, (5.22) *is satisfied if*

$$\sup_n E_{P_0} e^{\delta Y_{n1}} < \infty$$

for some positive δ.

PROOF. Since $\exp\{tk_n(x)\} \geq i_n(x) = \{1 + \frac{1}{2}tk_n(x)\}\exp\{\frac{1}{2}tk_n(x)\}$, $t > 0$, therefore $\int_0^1 i_n(x)\,dx < \infty$. Put $I = \sup_n \int \delta k_n(x)\exp\{\delta k_n(x)\}\,dx$, $J = \sup_n \int \exp\{\delta k_n(x)\}\,dx$ and $B = \delta\sqrt{C_4(M)}\min\{1, 1/2IJ\}$, $C_4(M) = 2e^{-M}$. Then take $h = h_1 + ih_2$ with $|h| < B$ and set $g_1(x) = v_{0n}^{-1}h_1 \log p_n(x)$. It follows that

$$\begin{aligned} |E_{P_0}\exp(hY_{n1})| &= |\exp(-hv_{0n}^{-1}e_{0n})| \\ &\quad\times\left|\int_0^1 \exp(g_1(x))\,dx + \int_0^1 \exp(g_1(x))\right. \\ &\qquad\left.\times\left(\exp\left(iv_{0n}^{-1}h_2 \log p_n(x)\right) - 1\right)dx\right|. \end{aligned} \tag{5.25}$$

Using the triangle inequality, (5.11), (5.14), (5.15) and the relation $|e^{it} - 1| \leq |t|$, $t \in R$, we get

$$\begin{aligned} |E_{P_0}&\exp(Y_{n1})| \\ &\geq C\left(\int_0^1 \exp(-|g_1(x)|)\,dx\right. \\ &\qquad\left.-\int_0^1 \exp(|g_1(x)|)v_{0n}^{-1}|h_2|\,|\log p_n(x)|\,dx\right), \end{aligned} \tag{5.26}$$

where $C = \exp\{-2\delta\sqrt{2}\,e^{M/2}\}$.

By choice of B, (5.15), (5.16) and Jensen's inequality applied to the first integral in (5.26), we get $|E_{P_0}\exp\{hY_{n1}\}| \geq 1/4J$ and the lower estimate in (5.22) follows.

For $|h| < B$, by (5.14) and (5.15), the upper estimate in (5.25) can be easily derived. □

Another sufficient condition for (5.22) follows from the following result.

PROPOSITION 5.11. *Suppose* (A4) *holds*, $H(p_n, p_0) \to 0$ *and* $nH^2(p_n, p_0) \to \infty$ *as* $n \to \infty$. *Further assume there exists a positive constant C such that*

$$\|\log p_n\|_\infty \{H(p_n, p_0)\}^{-1} \leq C \quad \text{for all } n. \tag{5.29}$$

Then (5.24) *holds*.

Finally, consider some known families of distributions which are used in the statistical literature to describe the path along which one goes from alternative to hypothesis. Different aspects of using such local families are discussed in Kallenberg and Drost (1993).

PROPOSITION 5.12. *Suppose that alternative densities have the form*

$$p_n(x) = 1 + c_n g_n(x), \qquad x \in [0, 1], \tag{5.30}$$

where $\int_0^1 g_n(x)\,dx = 0$. *Assume* $c_n \to 0$ *and* $0 < \inf_n \|g_n\|_2 \le \sup_n \|g_n\|_\infty < \infty$. *Then* $H(p_n, p_0) = O(c_n)$, (5.29) *is satisfied and* (5.22) *holds.*

PROPOSITION 5.13. *Let*

$$p_n(x) = \exp\{c_n w_n(x) - \delta_n\} \quad \text{for } x \in [0,1], \tag{5.31}$$

where δ_n *is a normalizing constant. Assume* $\sup_n \|w_n\|_\infty < \infty$ *and* $\inf_n \mathrm{Var}_{P_0} w_n(X) > 0$. *Then* $H(p_n, p_0) = O(c_n)$ *and* (5.22) *holds.*

PROOF. We have $p_n(x) = 1 + c_n g_n(x)$, where $g_n(x) = c_n^{-1}(\exp(c_n w_n(x)) - \exp(\delta_n))\exp(-\delta_n)$. Notice that $\delta_n \to 0$ and, by Lemma 7.1, g_n satisfies the assumptions of Proposition 5.12. □

PROPOSITION 5.14. *Assume that* p_n *is given by*

$$p_n(x) = \gamma_n\Big\{1 + c_n\big([z_n(x)]^\sigma - 1\big)\Big\}^{1/\sigma}, \qquad x \in [0,1],\ \sigma > 0, \tag{5.32}$$

where γ_n *is a normalizing constant. Assume that* $c_n \to 0$ *as* $n \to \infty$, $z_n(x) \ge 0$ *a.e.*, $\sup_n \|z_n\|_\infty < \infty$ *and* $\inf_n \mathrm{Var}_{P_0}[z_n(X)]^\sigma > 0$. *Then* $H(p_n, p_0) = O(c_n)$ *and* (5.22) *holds.*

PROOF. We have $p_n = 1 + c_n g_n(x)$, where $g_n(x) = \gamma_n\{[1 + c_n(z_n^\sigma(x) - 1)]^{1/\sigma} - \gamma_n^{-1}\}c_n^{-1}$. Using the inequality $1 + \xi y + l(\xi)y^2 \le (1+y)^\xi \le 1 + \xi y + r(\xi)y^2$, $y \in [-\frac{1}{2}, \frac{1}{2}]$, $\xi > 0$ [where the form of $l(\xi)$ and $r(\xi)$ changes as $\xi \in (0,1]$, $\xi \in [1,2]$ and $\xi \ge 2$], we infer that $\gamma_n \to 1$ as $n \to \infty$ and g_n satisfies the assumptions of Proposition 5.12. □

REMARK 5.15. The family (5.30) is called the contamination family, (5.31) is known as the exponential family, while (5.32) is called the Rényi-type family. For $\sigma = 1$, (5.32) reduces to (5.30), while $\sigma = \frac{1}{2}$ leads to the Hellinger-type family.

6. Asymptotic optimality of T_S. We start this section with some results concerning conditions on sequences of alternatives $\{p_n\}$ and maximal allowable τ for which the slopes of $V_n^{(1)}$ and $V_n^{(2)} = \{T_S\}^{1/2}$ exist [(i)(a), (i)(b) and (ii) hold] and are asymptotically equivalent, i.e.

$$\mathscr{E} = \lim_{n\to\infty} \frac{c^{(2)}}{c^{(1)}} \left\{ \frac{b^{(2)}(p_n)}{b^{(1)}(p_n)} \right\}^2 \text{ exists and } \mathscr{E} = 1. \tag{6.1}$$

Next, in subsection 6.2, for any sequence of alternatives for which we established the above, leading terms of critical values of $V_n^{(1)}$ and the related sequence $\{\alpha_n\}$ of significance levels are given. The resulting τ, satisfying (3.1) and not exceeding the maximal value for τ obtained in Section 6.1, is given also. Finally in Section 6.3, we give α_n and state conditions under which (3.4) is fulfilled for $V_{n;\alpha_n}^{(2)}$. This guarantees that the result $\mathscr{E} = 1$ can be interpreted in terms of (3.7), that is, in the traditional meaning as the limit of sample sizes needed to achieve the same power by the two considered α_n level tests. The conditions determine also a family $\mathscr{P}$ in which the τ efficiency exists.

6.1. *Equivalence of slopes.* Theorem 4.1, Proposition 4.3, Theorem 5.8, Remark 5.7 and Theorem 5.6 imply the following results.

THEOREM 6.1. *Assume* (A1) *and* (A2) *hold and* r *is a fixed number such that* $r > \omega + \frac{3}{2}$. *Take* $m = cn^{1/\eta}$, *where* $\eta \geq 2r + 1$ *and* c *is an arbitrary constant. Let* μ, ν *be any numbers satisfying* $(3 + 2\omega)/\eta < \mu < \nu < 2r/\eta$. *Let* $\mathscr{P}_{mr}$ *be the set of all sequences of alternatives for which* (A4)–(A6) *are fulfilled and define* $\mathscr{P}_{mr}^{*}$ *to be the subset of* $\mathscr{P}_{mr}$ *for which* $\gamma_m \to 0$ *as* $n \to \infty$ *and* (5.22) *is satisfied. Further set*

$$\mathscr{D}(\mu, \nu) = \big\{\{p_n\} \in \mathscr{P}_{mr}^{*}\colon n^{\mu}H^2(p_n, p_0) \to 0 \text{ and } n^{\nu}H^2(p_n, p_0) \to \infty\big\}.$$

Then, for all sequences of alternatives from $\mathscr{D}(\mu, \nu)$ *and* $\tau = (\eta - 2\omega - 1)/\eta$, $\mathscr{E}$ *exists and* $\mathscr{E} = 1$.

COROLLARY 6.2. *If, in addition to the assumptions of Theorem* 6.1, *it is known that the* p_n's *come from the Sobolev space* $W_2^{r(*)}$ *with* $r(*) > \omega + \frac{3}{2}$, *then, taking in Theorem* 6.1 $r = r(*)$ *the number m of exponential models, the rule S is looking through as well as the range of* μ *and* ν *are naturally related to the degree of smoothness of alternatives. In particular, if the* p_n's *are from* C^{∞}, *then choosing* $m = cn^{1/(2r+1)}$ *with r large enough, any intermediate range of convergence* (*in between the Pitman and Bahadur cases*) *can be attained.*

PROOF OF THEOREM 6.1. Let $\{p_n\} \in \mathscr{P}_{mr}^{*}$ be such that for some $\mu < \nu$ with $(3 + 2\omega)/\eta < \mu < \nu < 2r/\eta$ it holds that $n^{\mu}H^2(p_n, p_0) \to 0$ and $n^{\nu}H^2(p_n, p_0) \to \infty$. First, notice that the definition of $\mathscr{D}(\mu, \nu)$ implies that (A6) holds. Indeed, $m^{2r}H^2(p_n, p_0) = c^{2r}n^{\nu}H^2(p_n, p_0)n^{2r/\eta - \nu} \to \infty$.

Next, we show that (A7) and (A8) are satisfied. To this end, denote $\xi_0 = \sup\{\xi\colon n^{\xi}H^2(p_n, p_0) \to 0\}$. Of course, $\mu \leq \xi_0 \leq \nu$. Then we get immediately $\eta\xi_0/2 > \omega + \frac{3}{2}$, $\xi_0/2(1 - \xi_0) < r$ and $\xi_0/2(1 - \xi_0) < \eta\xi_0/2$. So, choose β such that

$$\max\left\{\omega + \frac{3}{2}, \frac{\xi_0}{2(1 - \xi_0)}\right\} < \beta < \min\left\{r, \frac{\eta\xi_0}{2}\right\}.$$

Then, by the definition of ξ_0 and the choice of β,

$$m^{2\beta}H^2(p_n, p_0) = c^{2\beta}n^{\xi_0}H^2(p_n, p_0)n^{-(\xi_0 - 2\beta/\eta)} \to 0.$$

Moreover, since $2\beta/(2\beta + 1) > \xi_0$, the definition of ξ_0 implies that

$$\left\{\frac{n}{\log n}\right\}^{2\beta/(2\beta+1)} H^2(p_n, p_0) = n^{\xi_0}H^2(p_n, p_0)n^{2\beta/(2\beta+1) - \xi_0}\{\log n\}^{-2\beta/(2\beta+1)} \to \infty.$$

Now, by the results listed at the beginning of this section, Theorem 6.1 follows. □

REMARK 6.3. If $m = n^{1/\eta}\delta_n$, where δ_n is any sequence such that $\delta_n n^{\varepsilon}$ tends to ∞ for every $\varepsilon > 0$ and tends to 0 for every $\varepsilon < 0$, then Theorem 6.1 remains true.

The next result is a counterpart of Theorem 6.1 for the case of a slowly increasing sequence $m = m_n$.

THEOREM 6.4. *Assume* (A1) *and* (A2) *hold and r is some fixed number satisfying $r > \omega + \frac{3}{2}$. Let $m = m_n$ be any sequence increasing to ∞ such that $n^{-\varepsilon}m \to 0$ for every $\varepsilon > 0$. Let μ be any number satisfying $2\omega + 3 < \mu < 2r$. Let $\mathscr{P}^*_{mr}$ be as in Theorem* 6.1 *and define*

$$\mathscr{D}(\mu) = \big\{\{p_n\} \in \mathscr{P}^*_{mr}: m^{\mu}H^2(p_n, p_0) \to 0,\ m^{2r}H^2(p_n, p_0) \to \infty\big\}.$$

Then, for each $\tau < 1$ and any sequence $\{p_n\} \in \mathscr{D}(\mu)$, (6.1) *holds.*

PROOF. Let $\{p_n\} \in \mathscr{P}^*_{mr}$ be such that, for some $\mu > \omega + \frac{3}{2}$, $m^{\mu}H^2(p_n, p_0) \to 0$ and $m^{2r}H^2(p_n, p_0) \to \infty$. First, notice that (A6) is contained in the definition of $\mathscr{D}(\mu)$. Next, observe that (A8) holds for every $\beta > \omega + \frac{3}{2}$. Indeed, by the assumptions,

$$\left(\frac{n}{\log n}\right)^{2\beta/(2\beta+1)} H^2(p_n, p_0)$$
$$= m^{2r}H^2(p_n, p_0)\left\{m\left(\frac{n}{\log n}\right)^{-\beta/r(2\beta+1)}\right\}^{-2r} \to \infty.$$

Now set $\xi_0 = \sup\{\xi: m^{\xi}H^2(p_n, p_0) \to 0\}$. Then $\mu \le \xi_0 \le 2r$. Choose a β satisfying $\omega + \frac{3}{2} < \beta < \xi_0/2$. Hence (A7) follows since $m^{2\beta}H^2(p_n, p_0) = m^{\xi_0}H^2(p_n, p_0)m^{-(\xi_0 - 2\beta)} \to 0$. By the results listed at the beginning of this section, the proof is concluded. □

REMARK 6.5. Faster growth of m_n in (A3) is possible. However, at least using the methods of proof employed in this paper, any other choice of m_n leads to a more narrow range of μ and ν under which (6.1) holds. In particular, if $m_n n^{-1/(2\omega+4)} \to \infty$, as $n \to \infty$, we are not able to find a sequence of alternatives under which (6.1) is fulfilled.

6.2. *Allowable $\{\alpha_n\}$ and some examples.* Denote by $\mathscr{V}^{(1)}_{n;\alpha_n}$ the test having the significance level α_n and rejecting P_0 for large values of $V^{(1)}_n$. First we shall identify those alternatives satisfying (6.1) for which

$$0 < \liminf_{n\to\infty} E_{P_n}\mathscr{V}^{(1)}_{n;\alpha_n} \le \limsup_{n\to\infty} E_{P_n}\mathscr{V}^{(1)}_{n;\alpha_n} < 1. \tag{6.2}$$

For each such sequence we find an asymptotic formula for corresponding significance level and the related value of τ fulfilling (3.1).

To this end we shall use the following result.

PROPOSITION 6.6. *If* (A4) *holds*, $\|p_n - p_0\|_\infty \to 0$ *and* $nH^2(p_n, p_0) \to \infty$, *then*

$$P_n\big(V_n^{(1)} - \sqrt{n}\, b^{(1)}(p_n) \le x\big) \to \boldsymbol{\Phi}(x), \qquad x \in R, \tag{6.3}$$

where $\boldsymbol{\Phi}$ *stands for the standard normal distribution function.*

PROOF. Set [cf. (5.21)]

$$X_{ni} = n^{-1/2} Y_{ni} = \big(\sqrt{n}\, v_{0n}\big)^{-1}\{\log p_n(X_i) - e_{0n}\}$$

and

$$\mathbf{B}_n^2 = n \operatorname{Var}_{P_n} X_{n1}.$$

Observe that $E_{P_n} X_{ni} = n^{-1/2} b^{(1)}(p_n)$ and $V_n^{(1)} - n^{1/2} b^{(1)}(p_n) = \sum_{i=1}^n (X_{ni} - E_{P_n} X_{ni})$. Moreover $\mathbf{B}_n^2 = v_{0n}^{-2}\{\int_0^1 (p_n - 1)\log^2 p_n + \int_0^1 \log^2 p_n - [D(p_n\|p_0)]^2\}$. The argument from the proof of Theorem 5.6 and the relations (5.20) yield $\mathbf{B}_n^2 = 1 + o(1)$. Therefore, by (A4) and (5.20),

$$\sum_{i=1}^n E_{P_n} |X_{ni} - E_{P_n} X_{ni}|^3 \le \mathbf{B}_n^2 \Big\| \big(\sqrt{n}\, v_{0n}\big)^{-1}\{\log p_n - D(p_n\|p_0)\}\Big\|_\infty$$
$$\le O(1)\big\{\sqrt{n}\, H(p_n, p_0)\big\}^{-1}.$$

Hence, Liapounov's central limit theorem for double arrays works [cf. Serfling (1980), page 32] and $\{V_n^{(1)} - n^{1/2} b^{(1)}(p_n)\}\mathbf{B}_n^{-1}$ is asymptotically (under P_n) standard normal. Since $\mathbf{B}_n^2 = 1 + o(1)$, the proof is concluded. □

Let c be an arbitrary constant. Proposition 6.6 implies that the test with critical region

$$\big\{V_n^{(1)} \ge \sqrt{n}\, b^{(1)}(p_n) + c\big\} \tag{6.4}$$

has asymptotic power staying away from 0 and 1. Since all of the results of subsection 6.1 concern p_n's such that $x_n = b^{(1)}(p_n)$ satisfies the assumptions of Theorem 5.8 [cf. (5.20), (A6) and (5.22)], therefore the sequence of related significance levels of (6.4) is defined via

$$-\log \alpha_n = \tfrac{1}{2}\big\{\sqrt{n}\, b^{(1)}(p_n)\big\}^2(1 + o(1)). \tag{6.5}$$

Moreover, since throughout subsection 6.1 it is assumed that $\|p_n - p_0\|_\infty \to 0$ (cf. Remark 5.7), by (5.19) we get

$$b^{(1)}(p_n) = \big\{2D(p_n\|p_0)\big\}^{1/2}(1 + o(1)). \tag{6.6}$$

So

$$-\log \alpha_n = \big\{nD(p_n\|p_0)\big\}(1 + o(1)). \tag{6.7}$$

Hence (3.1) leads to the following condition on τ: $n^{1-\tau} D(p_n\|p_0) = o(1)$. By (5.20) an equivalent condition is $n^{1-\tau} H^2(p_n, p_0) = o(1)$. So, for instance, for any sequence of alternatives $\{p_n\}$ from the set $\mathscr{D}(\mu, \nu)$, defined in Theorem 6.1, we can determine related τ defining α_n's under which the tests based on $V^{(2)} = \{T_S\}^{1/2}$ and $V_n^{(1)}$, respectively, satisfy (3.1) when observations obey the distribution P_n. Namely, setting $\tau_0 = 1 - \mu$, we get $n^{1-\tau_0}H^2(p_n, p_0) = o(1)$ and $\tau_0 < (\eta - 2\omega - 1)/\eta$ (cf. the formula for the maximal allowable τ given in Theorem 6.1). Hence we get the following result.

COROLLARY 6.7. *Suppose the assumptions of Theorem* 6.1 *are satisfied. Then for any sequence* $\{p_n\} \in \mathscr{D}(\mu, \nu)$, $\{\alpha_n\}$ *defined by* (6.7) *and* $\tau = 1 - \mu$ *the condition* (3.1) *is satisfied. Obviously*, $\mathscr{E} = 1$ *still holds.*

EXAMPLE 6.8. Consider the following alternatives:

(a) Contamination model:

$$p_n(x) = 1 + n^{-\xi}g(x), \qquad x \in [0,1],\ \xi > 0, \tag{6.8}$$

where $\int_0^1 g(x)\,dx = 0$ and $g \in W_2^r$ for some $r > \omega + \frac{3}{2}$.

(b) Exponential model:

$$p_n(x) = \exp\{n^{-\xi}w(x) - \delta_n\}, \qquad x \in [0,1],\ \xi > 0, \tag{6.9}$$

where $w \in W_2^r$ for some $r > \omega + \frac{3}{2}$ and δ_n is a normalizing constant.

(c) Rényi-type model:

$$p_n(x) = \gamma_n\{1 + n^{-\xi}([z(x)]^{\sigma} - 1)\}^{1/\sigma}, \qquad x \in [0,1],\ \sigma > 0,\ \xi > 0, \tag{6.10}$$

where $z(x) \geq 0$ a.e., $z \in W_2^r$ and γ_n is a normalizing constant.

Note that (6.8), (6.9) and (6.10) are special cases of (5.30), (5.31) and (5.32), respectively. Moreover, for each of the models $H(p_n, p_0) = O(n^{-\xi})$. Take α_n as in (6.7), $\eta \geq 2r + 1$ and set $m = cn^{1/(\eta)}$, c an arbitrary constant. If $\xi \in (\mu/2, \nu/2)$, where $(3 + 2\omega)/\eta < \mu < \nu < 2r/\eta$ then $\{p_n\} \in \mathscr{D}(\mu, \nu)$ for each of (a)–(c). Take $\tau = 1 - \mu$. Then the relations (3.1) and $\mathscr{E} = 1$ hold for alternatives given by (a), (b) and (c).

COROLLARY 6.9. *Suppose the assumptions of Theorem* 6.4 *are satisfied. To get* $\mathscr{E} = 1$ *we need* $\tau < 1$. *However, to fulfill* (3.1) *in this case, by* (6.7), $\tau = 1$ *is required. Therefore the notion of* τ *efficiency in* $\mathscr{D}(\mu)$ *cannot be applied.*

6.3. *Condition* (3.4) *for* $V_n^{(2)}$. In this section we give the asymptotic distribution of T_S under $\{p_n\}$. To this end set $a_j = \int_0^1 \phi_j(x)p_n(x)\,dx$ and $|a|_k = \{\sum_{j=1}^k a_j^2\}^{1/2}$, $k \geq 1$. The previous assumptions (A1)–(A8) are strengthened by imposing additionally

(A9) For each j, ϕ_j' exists and $\sup_x|\phi_j'(x)| \leq cj^h$ for some positive h and c,

$$r > h, \tag{A10}$$

$$m = 0(n^{1/(2r+2)}). \tag{A11}$$

Moreover, using l^* defined in Section 7.6, our last assumption is

$$\text{(A12)} \qquad \sqrt{n}\left(|a|_m^2 - |a|_{l^*}^2\right)/|a|_m \to 0 \text{ as } n \to \infty.$$

THEOREM 6.10. *Under* (A1)–(A12)

$$\left(T_S - n|a|_m^2\right)/\left(2\sqrt{n}\,|a|_m\right) \to N(0,1).$$

Some remarks on the proof of Theorem 6.10 and a discussion of (A12) are given in Section 7.6. Note that for the cosine basis (A9) holds with $h = 1$ while for the Legendre system $h = 5/2$ [cf. Sansone (1959) page 251]. Therefore (A10) holds in standard cases. (A11) strengthens slightly (A3).

COROLLARY 6.11. *Let c be an arbitrary constant. Theorem* 6.10 *implies that the test with critical region* $\{T_S \geq n|a|_m^2 + c\sqrt{n}\,|a|_m\}$ *has asymptotic power staying away from* 0 *and* 1. *Set* $\alpha_n = P_0(T_S \geq n|a|_m^2 + c\sqrt{n}\,|a|_m)$. *Therefore the related test* $\mathscr{V}_{n;\alpha_n}^{(2)}$ *obeys* (3.4). *Moreover, since* $|a|_m = 0(H(p_n, p_o))$ [*cf.* (7.24), (7.26)], *by Proposition* 4.3, α_n *satisfies* (6.7) *and* (3.1).

COROLLARY 6.12. *Assume the* ϕ_j*'s satisfy* (*A*9). *Consider models* (a), (b) *and* (c) *of Example* 6.8. *Additionally suppose the functions* g, w *and* z^σ *have finite numbers of non-zero Fourier coefficients with respect to the system of* ϕ_j*'s. Take* $r > \max\{\omega + \frac{3}{2}, h\}$, $\eta \geq 2r + 2$ *and* $\xi \in (\mu/2, \nu/2)$, *where* $(3 + 2\omega)/\eta < \mu \not< < 2r/\eta$. *Denote respective sets of sequences of* $\{p_n\}$ *by* $\mathscr{P}_{(a)}$, $\mathscr{P}_{(b)}$ *and* $\mathscr{P}_{(c)}$. *If* $m = cn^{1/\eta}$, *c an arbitrary constant, then the test* $\mathscr{V}_{n;\alpha_n}^{(2)}$, *based on* T_S *and defined in Corollary* 6.11, *is asymptotically* $\tau = 1 - \mu$ *efficient in each of* $\mathscr{P}_{(a)}$, $\mathscr{P}_{(b)}$ *and* $\mathscr{P}_{(c)}$. *In particular, Theorem* 2.5 *follows.*

7. Some auxiliary results and proof of Theorem 4.1.

7.1. *Schwarz's rule S and related results.* Let us start with some notation. Throughout we shall operate on vectors from spaces of different dimensions. To simplify the formulas, the dimension is not incorporated into a notation of the vector and usually depends on the context. Moreover, to avoid embedding vectors into a common space, we use the following rule: if $x \in R^j$ and $y \in R^k$ for some $k \geq j$, k possibly not given explicitly in the text, we set $x \circ y = \sum_{l=1}^{j} x_l y_l$. Moreover, for $x \in R^k$ and $1 \leq i < j \leq k$, we define $|x|_j = \{\sum_{l=1}^{j} x_l^2\}^{1/2}$ and $|x|_{ij} = \{\sum_{l=i+1}^{j} x_l^2\}^{1/2}$.

Subsections 7.1 and 7.2 are self-contained with only orthonormality of the ϕ_j's and assumption (A1) being exploited. Therefore, we find it useful to replace the sequence m_n in the definition of S by an arbitrary increasing sequence $d(n)$.

For $k = 1, 2, \ldots, d(n)$, $\theta \in R^k$ and $\phi(x) = (\phi_1(x), \phi_2(x), \ldots, \phi_k(x))$, consider exponential families (2.1) defined by

$$g_k(x;\theta) = \exp\{\theta \circ \phi(x) - \psi_k(\theta)\},$$

with $\psi_k(\theta)$ defined as in (2.2). Given $X_1, \dots, X_n$ i.i.d. r.v.'s with density $g_k(x; \theta)$ set

$$\overline{\phi} = \left(\overline{\phi}_1, \overline{\phi}_2, \dots, \overline{\phi}_k\right) \quad \text{where } \overline{\phi}_j = \frac{1}{n} \sum_{i=1}^{n} \phi_j(X_i).$$

Then [cf. (2.4) and (2.5)]

$$L_k = n \sup_{\theta \in R^k} \left\{\overline{\phi} \circ \theta - \psi_k(\theta)\right\} - \tfrac{1}{2} k \log n$$

and

$$S = \min\left\{k : 1 \le k \le d(n),\, L_k \ge L_j,\, 1 \le j \le d(n)\right\}. \tag{7.1}$$

Below, we state and prove three theorems which prove to be useful in the analysis of the distribution of S. We start, however, with some facts concerning the exponential function and $\psi_k(\theta)$ which shall be used in this and succeeding sections.

LEMMA 7.1. *For any positive c and $y \in [-c, c]$, it holds*

$$1 + y + c^{-2}(e^{-c} + c - 1)y^2 \le e^y \le 1 + y + c^{-2}(e^c - c - 1)y^2. \tag{7.2}$$

For any $c \in (0, 3]$ and all $y \in [-c, c]$, it holds

$$1 + y + y^2(3 - c)/6 \le e^y \le 1 + y + y^2(1 + c)/2. \tag{7.3}$$

Let

$$u_k = k^{1/2} V_k, \tag{7.4}$$

with $V_k = \max_{1 \le j \le k} \sup_x |\phi_j(x)|$ [cf. (2.8)]. Then, for orthonormal ϕ_j's satisfying (A1) *and $\theta \in R^k$ such that $c = c_k(\theta) = u_k |\theta|_k \in (0, 3]$,*

$$\frac{3 - c}{6} |\theta|_k^2 - \frac{(3 - c)^2}{72} |\theta|_k^4 \le \psi_k(\theta) \le \frac{1 + c}{2} |\theta|_k^2. \tag{7.5}$$

PROOF. Since $e^y = 1 + y + y^2[y^{-2}(e^y - y - 1)]$ and the function $y^{-2}(e^y - y - 1)$ is increasing on R, (7.2) follows. For $c \in (0, 3]$ it holds that $c^{-2}(e^{-c} + c - 1) \ge (3 - c)/6$ and $c^{-2}(e^c - c - 1) \le (1 + c)/2$. Hence (7.2) implies (7.3). Since $|\sum_{j=1}^{k} \theta_j \phi_j(x)| \le u_k |\theta|_k$ and for any $z > 0$ it holds that $-z^2/2 + z \le \log(1 + z) \le z$, (7.5) follows. □

REMARK 7.2. The function $\psi_k(\theta)$ plays a central role in likelihood methods for exponential families. In the statistical literature a behavior of $\psi_k(\theta)$ in a vicinity of 0 is usually investigated via Taylor expansions [cf. Kallenberg (1981), Section 3, and Portnoy (1988), Section 2]. However, then an analysis of the remainder term of the expansion requires as a rule some extra assumptions (especially when considering exponential families with the number of parameters tending to ∞). On the contrary, (7.5) provides explicit bounds for the remainder.

Now we shall present the three main results of this subsection.

THEOREM 7.3. *For every* $k \geq 1$, $\varepsilon \in (0,1]$ *and* $0 < \delta \leq \varepsilon^2(2+\varepsilon)^{-3}u_k^{-2}$,

$$\left\{x \in R^k\colon \sup_{t\in R^k}\big(x\circ t - \psi_k(t)\big) \geq \delta\right\} \supset \left\{x \in R^k\colon |x|_k^2 \geq (2+\varepsilon)\delta\right\}. \tag{7.6}$$

PROOF. Consider first $x \in R^k$ such that $|x|_k^2 = (2+\varepsilon)\delta$ and set $c(x) = u_k|x|_k$. Then, by the assumption on δ, $c(x) < 3$. Obviously, $\sup_{t\in R^k}(x\circ t - \psi_k(t)) \geq |x|_k^2 - \psi_k(x)$, and, by (7.5), $\psi_k(x) \leq \frac{1}{2}\{1 + c(x)\}|x|_k^2$. The relation $|x|_k^2 = (2+\varepsilon)\delta$ and the assumption on δ then yield $\sup_{t\in R^k}(x\circ t - \psi_k(t)) \geq \delta$.

Now, let $x \in R^k$ be such that $|x|_k^2 > (2+\varepsilon)\delta$. Take $x_0 = \{(2+\varepsilon)\delta\}^{1/2}x|x|_k^{-1}$ and observe that

$$\begin{aligned}\sup_{t\in R^k}\big(x\circ t - \psi_k(t)\big) &\geq x\circ x_0 - \psi_k(x_0) = |x|_k\{(2+\varepsilon)\delta\}^{-1/2}|x_0|_k^2 - \psi_k(x_0)\\ &\geq |x_0|_k^2 - \psi_k(x_0) \geq \delta,\end{aligned}$$

where the last inequality follows by the previous argument. □

THEOREM 7.4. *For every* $k \geq 1$, $\varepsilon \in (0, \min\{1, 2u_k^2/3\})$ *and* $0 < \delta \leq \delta_0(k)$, *where* $\delta_0(k) = (2-\varepsilon)\varepsilon^2u_k^{-2}/16$,

$$\left\{x \in R^k\colon \sup_{t\in R^k}\big(x\circ t - \psi_k(t)\big) \geq \delta\right\} \subset \left\{x \in R^k\colon |x|_k^2 \geq (2-\varepsilon)\delta\right\}. \tag{7.7}$$

PROOF. Observe that to prove (7.7) it is enough to show that, for an arbitrary $t \in R^k$ and $\delta \in (0, \delta_0(k)]$,

$$\psi_k(t) + \delta \geq \sqrt{(2-\varepsilon)\delta}\,|t|_k. \tag{7.8}$$

Indeed, if $x \in R^k$ satisfies $|x|_k^2 < (2-\varepsilon)\delta$, then $\sup_{t\in R^k}(x\circ t - \psi_k(t)) \leq \sup_{t\in R^k}(|x|_k|t|_k - \psi_k(t)) < \sup_{t\in R^k}(\sqrt{(2-\varepsilon)\delta}\,|t|_k - \psi_k(t)) \leq \sup_{t\in R^k}(\psi_k(t) + \delta - \psi_k(t)) = \delta$.

The proof of (7.8) shall be given in two parts. First assume additionally that t is such that $c(t) = u_k|t|_k \in (0, \varepsilon]$. By (7.5) we get

$$\psi_k(t) \geq \tfrac{1}{2}|t|_k^2 - \tfrac{1}{6}u_k|t|_k^3 - \tfrac{1}{8}|t|_k^4\big(1 - \tfrac{1}{3}u_k|t|_k\big)^2.$$

An elementary calculation shows that from this and the assumption $|t|_k \leq \varepsilon u_k^{-1}$ it follows that

$$\psi_k(t) \geq \tfrac{1}{4}(2-\varepsilon)|t|_k^2. \tag{7.9}$$

Hence $\psi_k(t) + \delta \geq \frac{1}{4}(2-\varepsilon)|t|_k^2 + (\sqrt{\delta})^2 \geq \sqrt{(2-\varepsilon)\delta}\,|t|_k$ and (7.8) is proved in the case $u_k|t|_k \leq \varepsilon$.

Assume now $u_k|t|_k > \varepsilon$. Put $v = t/b$, where $b = u_k|t|_k\varepsilon^{-1}$. Then $b > 1$ and $|v|_k = \varepsilon u_k^{-1}$. By Jensen's inequality and (7.9),

$$\begin{aligned}\psi_k(t) + \delta &= \psi_k(bv) + \delta \geq b\psi_k(v) + \delta \geq \tfrac{1}{4}b(2-\varepsilon)|v|_k^2 + \delta\\ &= \sqrt{(2-\varepsilon)\delta}\,|t|_k + z,\end{aligned}$$

where $z = (\delta - b\sqrt{(2-\varepsilon)\delta}\,|v|_k + b(2-\varepsilon)|v|_k^2/4)$. Since for $\delta \in (0, \delta_0(k)]$ it holds that $z > 0$, the proof is concluded. □

REMARK 7.5. The logarithm of the likelihood ratio test statistic for testing $\theta = 0$ against $\theta \neq 0$ in the model $g_k(x;\theta)$ has the form $n \sup_{\theta \in R^k}\{\overline{\phi} \circ \theta - \psi_k(\theta)\}$. Theorems 7.3 and 7.4 show that some large deviation probabilities for this statistic can be obtained by analyzing a much simpler statistic $\frac{1}{2}n|\overline{\phi}|_k^2$. Besides, results for norms of random vectors are well developed. For example, nice exponential bounds are available (cf. Theorem 7.7 and Proposition 7.17 below). A variety of large deviation results can be found in Inglot, Kallenberg and Ledwina (1993). Finally, note that an idea of approximating the set $\{x \in R^k\colon \sup_{t \in R^k}(x \circ t - \psi_k(t)) \geq \delta\}$ by some easier tractable sets has been used earlier in Kourouklis (1984).

Our next theorem will play a key role in investigating the events of the form $\{S < l\}$ for some natural l.

THEOREM 7.6. *For all* $k \geq 1$, *every* $\varepsilon \in (0, \min\{1, 2u_k^2/3\}]$, *every* $c \geq 0$ *and all* $j < k$,

$$
\begin{aligned}
&\Big\{x \in R^k\colon \sup_{t \in R^k}\big(x \circ t - \psi_k(t)\big) < c + \sup_{s \in R^j}\big(x \circ s - \psi_j(s)\big)\Big\} \\
&\qquad \subset \Big\{x \in R^k\colon |x|_{jk}^2 \leq (2-\varepsilon)c + \varepsilon\delta_*\Big\} \cup \Big\{x \in R^k\colon |x|_k^2 \geq \delta_*\Big\},
\end{aligned} \tag{7.10}
$$

where $\delta_* = (2-\varepsilon)\varepsilon^2/12u_k^2$.

PROOF. Observe that if for some $x \in R^k$ we have $|x|_k^2 < \delta_*$ and $|x|_{jk}^2 > (2-\varepsilon)c + \varepsilon\delta_*$, then

$$
\begin{aligned}
|x|_j^2 &= |x|_k^2 - |x|_{jk}^2 < |x|_k^2 - (2-\varepsilon)c - \varepsilon\delta_* \\
&\leq |x|_k^2 - (2-\varepsilon)c - \varepsilon|x|_k^2 \\
&\leq (2-\varepsilon)(2+\varepsilon)^{-1}|x|_k^2 - (2-\varepsilon)c \\
&= (2-\varepsilon)\big[(2+\varepsilon)^{-1}|x|_k^2 - c\big].
\end{aligned} \tag{7.11}
$$

Set $\delta = (2+\varepsilon)^{-1}|x|_k^2 - c$. Then, due to $|x|_k^2 < \delta_*$ and $u_j < u_k$, it holds that $\delta < \delta_0(j)$, where $\delta_0(\cdot)$ is as in Theorem 7.4. Application of (7.11) and Theorem 7.4 with $k = j$ and the above-defined δ leads to $\sup_{s \in R^j}(x \circ s - \psi_j(s)) < \delta$ or equivalently $(2+\varepsilon)^{-1}|x|_k^2 > c + \sup_{s \in R^j}(x \circ s - \psi_j(s))$. On the other hand, by $|x|_k^2 < \delta_*$, $(2+\varepsilon)^{-1}|x|_k^2 < (2-\varepsilon)\varepsilon^2/12(2+\varepsilon)u_k^2 < \varepsilon^2/(2+\varepsilon)^3u_k^2$. Hence $b = c + \sup_{s \in R^j}(x \circ s - \psi_j(s)) < \varepsilon^2(2+\varepsilon)^{-3}u_k^{-2}$. Thus using Theorem 7.3 with the given k and $\delta = b$ yields $\sup_{t \in R^k}(x \circ t - \psi_k(t)) > c + \sup_{s \in R^j}(x \circ s - \psi_j(s))$, which completes the proof. $\square$

7.2. *Large deviations for* T_S *under* P_0. Let $X_1, \ldots, X_n$ be i.i.d. random variables taking values in $[0,1]$. The Neyman smooth test statistic (for testing uniformity) with a fixed number of components k has the form [cf.

(2.3)] $T_k = \sum_{j=1}^{k}\{n^{-1/2}\sum_{i=1}^{n}\phi_j(X_i)\}^2 = n|\overline{\phi}|_k^2$. The data-driven smooth test statistic is then defined by $T_S = \sum_{j=1}^{S}\{n^{-1/2}\sum_{i=1}^{n}\phi_j(X_i)\}^2$, where S is given by (7.1). Recall that also in this section S chooses one out the $d(n)$ candidate exponential models (2.1).

The large deviation result for T_S, which we state and prove below, is based on the following version of inequality (2) in Prohorov (1973).

THEOREM 7.7 [Prohorov (1973)]. *Let $Y_1, \dots, Y_n$ be i.i.d. random vectors in R^k with uncorrelated components. Assume $EY_1 = 0$ and $|Y_1|_k \le L$, P_0 a.e. Then, for $y^2 \in [2k, nL^{-2}]$,*

$$P_0\left(\left|n^{-1/2}\sum_{i=1}^{n} Y_i\right|_k \ge y\right) \le \frac{C}{\Gamma(k/2)}\left(\frac{y^2}{2}\right)^{(k-1)/2} \exp\left\{-\frac{y^2}{2}(1-\eta_n)\right\}, \tag{7.12}$$

where C is an absolute constant while $0 \le \eta_n < Lyn^{-1/2}$.

Also we shall need the following result.

PROPOSITION 7.8. *Assume $E_{P_0}\phi_1(X) = 0$ and $\sup_x|\phi_1(x)| < \infty$. Then, for any sequence $\{x_n\}$ such that $x_n \to 0$ and $nx_n^2 \to \infty$ as $n \to \infty$,*

$$\lim_{n\to\infty}\left(nx_n^2\right)^{-1}\log P_0\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_1(X_i) \ge x_n\sqrt{n}\right) = -\frac{1}{2}.$$

PROOF. Proposition 7.8 follows from Theorem 1 of Chapter 8 in Petrov (1975). □

THEOREM 7.9. *Assume* (A1) *and recall that $u_k = \sqrt{k}\,V_k$. Assume that $d(n) \to \infty$ and $n^{-1}(\log n)\, d(n)u_{d(n)}^2 \to 0$. Let $\{x_n\}$ satisfy $x_n \to 0$, $nx_n^2 \to \infty$ and $x_n u_{d(n)} = o(1)$. Then*

$$\lim_{n\to\infty}\left(nx_n^2\right)^{-1}\log P_0\left(T_S \ge nx_n^2\right) = -\tfrac{1}{2}. \tag{7.13}$$

PROOF. By (7.1), $S \ge 1$. Hence $T_S \ge T_1$ and, by Proposition 7.8,

$$\liminf_{n\to\infty}\left(nx_n^2\right)^{-1}\log P_0\left(T_S \ge nx_n^2\right) \ge \lim_{n\to\infty}\left(nx_n^2\right)^{-1}\log P_0\left(T_1 \ge nx_n^2\right) = -\tfrac{1}{2}.$$

To majorize $P_0(T_S \ge nx_n^2)$, observe that

$$P_0\left(T_S \ge nx_n^2\right) = \sum_{k=1}^{d(n)} P_0\left(T_k \ge nx_n^2,\, S = k\right)$$

and the definition of S implies

$$\{S = k\} \subset \left\{\sup_{t\in R^k}\left(\phi\circ t - \psi_k(t)\right) \ge \tfrac{1}{2}(k-1)n^{-1}\log n\right\}.$$

The assumption $n^{-1}(\log n)\, d(n) u^2_{d(n)} \to 0$ implies that $\delta = \frac{1}{2}(k-1)n^{-1}\log n$ satisfies $\delta \le \delta_0(k) = (2-\varepsilon)\varepsilon^2/16u_k^2$ for any $k \le d(n)$ and $\varepsilon \in (0,1)$, provided that n is sufficiently large. Consequently, by Theorem 7.4, for $k = 1,\dots,d(n)$ and any $\varepsilon \in (0, \min\{1, 2u_1^2/3\})$, $\{S = k\} \subset \{|\overline{\phi}|_k^2 \ge (1/2n)(2-\varepsilon) \times (k-1)\log n\}$ for sufficiently large n. Since however $n|\overline{\phi}|_k^2 = T_k$, the above yields

$$P_0(T_S \ge nx_n^2) \le P_0(T_1 \ge nx_n^2) + \sum_{k=2}^{d(n)} P_0\left(T_k \ge \max\left\{nx_n^2, \frac{1}{2}(2-\varepsilon)(k-1)\log n\right\}\right). \tag{7.14}$$

To majorize the second component in (7.14), for each $k = 2,\dots,d(n)$ we shall apply Theorem 7.7 with $y = (\max\{nx_n^2, (1-\varepsilon/2)(k-1)\log n\})^{1/2}$ and $L = u_k$. To this end, observe that for any $k \ge 2$ and sufficiently large n it holds that $y^2 \ge 2k$. Moreover, since $x_n d(n) = o(1)$ it follows that $y^2L^2 \le n$, if n is sufficiently large. Thus (7.12) implies that

$$P_0(T_k \ge y^2) \le C\left(\frac{1}{2}\right)^{(k-1)/2} \exp\left\{\frac{k-1}{2}\log y^2 - \log\Gamma\left(\frac{k}{2}\right) - \frac{y^2}{2}(1-\eta_n)\right\}.$$

To estimate the first two components in the exponent, we shall use the facts that $\log\Gamma(x) \ge -x + (x-\frac{1}{2})\log x$, $2y^2(k-1)^{-1} \ge (2-\varepsilon)\log n$ and $x^{-1}\log x$ is decreasing for $x > e$. Hence $\frac{1}{2}(k-1)\log y^2 - \log\Gamma(k/2) \le \frac{1}{2}y^2c_n$, where $c_n = 2(\log n)^{-1}[\log\log(n^2) + 2]$. Therefore, $P_0(T_k \ge y^2) \le C2^{(1-k)/2}\exp\{-\frac{1}{2}y^2(1-\eta_n-c_n)\}$ and $\sum_{k=2}^{d(n)} P_0(T_k \ge y^2) \le 3C\exp\{-\frac{1}{2}y^2(1-\eta_n-c_n)\}$. Proposition 7.8 implies that there exists a sequence $\{\delta_n\}$, $\delta_n \to 0$, such that $P_0(T_1 \ge nx_n^2) \le \exp\{-\frac{1}{2}nx_n^2(1-\delta_n)\}$. We conclude from (7.14) and the last two bounds that $P_0(T_S \ge nx_n^2) \le (3C+1)\exp\{-\frac{1}{2}nx_n^2\zeta_n\}$, where $\zeta_n = \max\{1-\delta_n, 1-\eta_n-c_n\}$. As $\zeta_n \to 1$ the proof is concluded. □

7.3. *Information projection of P_n's.*

7.3.1. *Preliminaries.* In this section we shall study in detail some properties of the information projection. We shall start by recalling some notions and basic results derived by Barron and Sheu (1991). Throughout we restrict attention to the densities p_n introduced in subsection 2.1.

As in Barron and Sheu (1991), let us denote by p_m^* the density in the m-dimensional exponential family $g_m(\cdot;\theta)$ given by [cf. (2.1)]

$$g_m(x;\theta) = \exp\left\{\sum_{j=1}^{m}\theta_j\phi_j(x) - \psi_m(\theta)\right\}, \tag{7.15}$$

with $\theta \in R^m$, which is closest to p_n in the relative entropy sense. More precisely, let

$$a_i = \int_0^1 \phi_i(x)p_n(x)\,dx, \qquad 1 \le i \le m. \tag{7.16}$$

Then p_m^* is the unique (provided it exists) density in the family (7.15) for which

$$\int_0^1 \phi_i(x) p_m^*(x)\, dx = a_i, \qquad 1 \le i \le m. \tag{7.17}$$

Denoting the relative entropy (Kullback–Leibler distance) between two probability densities p and q by

$$D(p\|q) = \int_0^1 p(x) \log \frac{p(x)}{q(x)}\, dx,$$

p_m^* can also be defined via the relation

$$D(p_n\|g_m) = D(p_n\|p_m^*) + D(p_m^*\|g_m) \tag{7.18}$$

valid for all densities g_m in (7.15). So, in particular, (7.18) holds for $g_m = p_0$. In view of (7.18), p_m^* is called the information projection.

For further considerations the following remark is useful.

REMARK 7.10. As in Barron and Sheu (1991), let A_m be such that, for all f in the linear space spanned by $\phi_0, \phi_1, \ldots, \phi_m$, $\|f\|_\infty \le A_m \|f\|_2$. Observe that (A1) and the orthonormality of the ϕ_j's imply that we can take $A_m = u_m$, where $u_m = \sqrt{m}\, V_m$ [cf. (7.4)].

Set

$$f_m^* = \log p_m^* = \theta^* \circ \phi - \psi_m(\theta^*). \tag{7.19}$$

The following properties of p_m^* and f_m^* can be deduced from the formulation and proof of Theorem 3 of Barron and Sheu (1991).

LEMMA 7.11. *Assume* (A1)–(A5) *and set* $\varepsilon_m = 4 u_m \Delta_m \exp\{2M + 4\gamma_m + 1\}$. *Then the following conclusions are valid*:

(a) *If* n *is so large that* $\varepsilon_m < 1$, *then the information projection* p_m^* *exists*.
(b) $D(p_n\|p_m^*) = O(\Delta_m^2)$.
(c) $\|\log p_n - f_m^*\|_\infty \le 2\gamma_m + \varepsilon_m$.

7.3.2. *Properties of the information projection.*

LEMMA 7.12. *Assume* (A1)–(A5). *If* $m^\zeta H(p_n, p_0) \to 0$ *for some* $0 \le \zeta < r$, *then* $m^\zeta |\theta^*|_m \to 0$. *In particular*,

(a) *if* $H(p_n, p_0) \to 0$, *then* $|\theta^*|_m \to 0$ *as* $n \to \infty$,
(b) *if* $m^{\omega+1/2} H(p_n, p_0) \to 0$, *then* $u_m |\theta^*|_m \to 0$ *as* $n \to \infty$.

PROOF. It is well known that $H^2(p, q) \le D(p\|q)$ for any two densities p and q. Hence $H(p_m^*, p_0) \le H(p_n, p_0) + \{D(p_n\|p_m^*)\}^{1/2}$. By the assumption $m^\zeta H(p_n, p_0) \to 0$ and Lemma 7.11(b), this implies that $m^\zeta H(p_m^*, p_0) \to 0$ as $n \to \infty$ or equivalently $m^{2\zeta} \int_0^1 (\exp\{\frac{1}{2} f_m^*(x)\} - 1)^2\, dx \to 0$ as $n \to \infty$. Now, (A4) and Lemma 7.11(c) yield $\|f_m^*\|_\infty \le \|\log p_n - f_m^*\|_\infty + \|\log p_n\|_\infty \le 2M_1$, where $2M_1 = 2\gamma_m + \varepsilon_m + M$. Since $(e^t - 1)^2 \ge ([1 - \exp\{-M_1\}]/M_1)^2 t^2$ for all $t \in [-M_1, M_1]$, then $m^{2\zeta}\int_0^1 [f_m^*(x)]^2\,dx = m^{2\zeta}(|\theta^*|_m^2 + \psi_m^2(\theta^*)) \to 0$ and the conclusion follows. □

LEMMA 7.13. *Assume that* $m^{\omega+1/2}H(p_n, p_0) \to 0$ *and* (A1)–(A5) *are fulfilled. Then, for any* $\delta > 0$ *and sufficiently large* n,

$$(7.20)\qquad \left(\tfrac{1}{4} - \delta\right)|\theta^*|_m^2 \le H^2(p_m^*, p_0) \le \left(\tfrac{1}{4} + \delta\right)|\theta^*|_m^2.$$

PROOF. By the definition of H we get

$$H^2(p_m^*, p_0) = 2 - 2\exp\left\{\psi_m\left(\tfrac{1}{2}\theta^*\right) - \tfrac{1}{2}\psi_m(\theta^*)\right\}.$$

Since $\exp\{\psi_m(\theta^*)\} = \int_0^1 \exp\{\theta^* \circ \phi\}\,dx$ we shall apply Lemma 7.1 with $y = \theta^* \circ \phi$ and $c = c^* = u_m|\theta^*|_m$. By Lemma 7.12, $c^* \to 0$. Set $A = (1 + c^*)/2$ and $B = (3 - c^*)/6$. With this notation Lemma 7.1 yields

$$1 + B|\theta^*|_m^2 \le \exp\{\psi_m(\theta^*)\} \le 1 + A|\theta^*|_m^2.$$

Using similar bounds for $\exp\{\psi_m(\frac{1}{2}\theta^*)\}$ and applying the inequality $(1 + x)^{1/2} \le 1 + x/2$, we get

$$\frac{1}{2}(2B - A)|\theta^*|_m^2 \frac{1 - C|\theta^*|_m^2}{1 + B|\theta^*|_m^2} \le H^2(p_m^*, p_0) \le \frac{2A - B}{2 + A|\theta^*|_m^2}|\theta^*|_m^2,$$

where $C = AB/4(B - A/2)$. Hence (7.20) follows easily. □

Observe that (A7) and (A8), imposed in Section 4, imply there exists a sequence $\{\kappa_n\}$, $\kappa_n \to \infty$, such that

$$(7.21)\qquad n(\kappa_n \log n)^{-1}\{H(p_n, p_0)\}^{2+1/\beta} \to \infty$$

and

$$(7.22)\qquad \kappa_n\{H(p_n, p_0)\}^{(2\beta - 2\omega - 3)/2\beta} \to 0 \quad \text{as } n \to \infty.$$

Now put

$$(7.23)\qquad \rho_n = n(\kappa_n \log n)^{-1}\{H(p_n, p_0)\}^{2+1/\beta}.$$

COROLLARY 7.14. *Under* (A1)–(A7) *it holds that*

$$(7.24)\qquad \sqrt{3} \le |\theta^*|_m / H(p_n, p_0) \le \sqrt{5} \quad \textit{for sufficiently large } n$$

and

$$D(p_n \| p_m^*) = o\left(H^2(p_n, p_0)\right) = o\left(|\theta^*|_m^2\right).$$

PROOF. By (A6) and Lemma 7.11(b), $\{D(p_n \| p_m^*)\}^{1/2} = o(H(p_n, p_0))$. Hence (7.20) and the beginning of the proof of Lemma 7.12 give $|\theta^*|_m/\sqrt{5} \le H(p_m^*, p_0) \le H(p_n, p_0)(1 + o(1))$ which proves the right-hand side of (7.24).

Similarly, as $H(p_m^*, p_0) \geq H(p_n, p_0) - \{D(p_n \| p_m^*)\}^{1/2}$, then by (7.20) we get $\sqrt{\frac{1}{3}}\,|\theta^*|_m \geq H(p_m^*, p_0) \geq H(p_n, p_0)(1 + o(1))$ which proves the left-hand side of (7.24). Now the second part of the corollary follows immediately. □

COROLLARY 7.15. *Assume* (A1)–(A8) *hold. Then*

$$u_m|\theta^*|_m^3 = o\big(n^{-1}\rho_n \log n\big). \tag{7.25}$$

PROOF. (7.24) and (A2) yield

$$u_m|\theta^*|_m^3 \leq O(1)m^{\omega+1/2}H^3(p_n, p_0).$$

Hence, due to (A7),

$$\begin{aligned} &u_m|\theta^*|_m^3\big(n^{-1}\rho_n \log n\big)^{-1} \\ &\quad\leq O(1)\big\{m^{\omega+1/2}\kappa_n H^3(p_n, p_0)\big\}\big\{H(p_n, p_0)\big\}^{-(2+1/\beta)} \\ &\quad= O(1)\,\kappa_n\big\{H(p_n, p_0)\big\}^{(2\beta-2\omega-3)/2\beta}. \end{aligned}$$

By (7.22) the proof is concluded. □

Before we state the last result of this subsection recall that $a_j = \int_0^1 \phi_j(x)p_m^*(x)\,dx$ [cf. (7.17)] and set $a = (a_1, \dots, a_m)$.

LEMMA 7.16. *Assume that* (A1)–(A7) *are satisfied. Set* $c^* = u_m|\theta^*|_m$. *Then, for sufficiently large* n,

$$|a_j - \theta_j^*| \leq \tfrac{1}{2}(1 + c^*)V_j|\theta^*|_m^2 \quad \text{for } 1 \leq j \leq m \tag{7.26}$$

and

$$|a - \theta^*|_{jk} = o\big(u_m^{1/2}|\theta^*|_m^{3/2}\big) \quad \text{for } 1 \leq j < k \leq m. \tag{7.27}$$

PROOF. Since $\theta_j^* = \int_0^1 \phi_j(x)(\phi \circ \theta^*)\,dx$, we can write

$$a_j - \theta_j^* = \int_0^1 \phi_j(x)\big[\exp\{\phi \circ \theta^* - \psi_m(\theta^*)\} - 1 - (\phi \circ \theta^* - \psi_m(\theta^*))\big]\,dx.$$

Since $e^y \geq 1 + y$ for $y \in R$, it follows that $|a_j - \theta_j^*| \leq V_j\psi_m(\theta^*)$. By Lemma 7.12 and (7.5) we get for sufficiently large n that $\psi_m(\theta^*) \leq \frac{1}{2}(1 + c^*)|\theta^*|_m^2$. This establishes (7.26). The estimate (7.26) then yields (7.27). □

7.4. *Asymptotic behavior of S under P_n.* We start with a technical result we shall exploit to investigate the distribution of S under P_n. In subsection 7.2 we used a version of Prohorov's inequality to analyze the distribution of $\sqrt{n}\,|\overline{\phi}|_k$ under P_0. Since under P_n the components of $\overline{\phi}$ are, as a rule, correlated, the original Prohorov inequality is less convenient for application in this situation. However, the following result of Yurinskii [(1976), page 491], is more easily applicable.

PROPOSITION 7.17 [Yurinskii (1976)]. *Let independent random vectors* $Y_1, \dots, Y_n$ *with values in* R^k *satisfy the following conditions*:

$$EY_i = 0,$$

$$E|Y_i|_k^s \le \frac{s!}{2} b_i^2 \mathscr{K}^{s-2}, \qquad s = 2, 3, \dots, \tag{7.28}$$

$$B_n^2 = b_1^2 + \cdots + b_n^2.$$

Then, for $x > 0$,

$$P\left(\left|\sum_{i=1}^n Y_i\right|_k \ge x B_n\right) \le 2 \exp\left\{-\frac{x^2}{2}\left(1 + 1.62\frac{x\mathscr{K}}{B_n}\right)^{-1}\right\}. \tag{7.29}$$

In the sequel we shall use the following result.

COROLLARY 7.18. *Let assumptions* (A1)–(A4) *and* (A7) *hold. Then, for any* j, k *such that* $0 < j < k \le m^{\beta/(\omega+1/2)}$ *and arbitrary* $c_n > 0$ *such that*

$$c_n V_k (k-j)^{-1/2} \to 0 \quad \text{as } n \to \infty, \tag{7.30}$$

we have that

$$P_n\big(|\overline{\phi} - a|_{jk} \ge c_n\big) \le 2 \exp\left\{-\frac{n c_n^2}{2(k-j)}(1 + o(1))\right\}, \tag{7.31}$$

where $a = \int_0^1 \phi(x) p_n(x)\, dx$ [*cf.* (7.16)].

PROOF. Note that $|\phi - a|_{jk} \le 2\sqrt{k-j}\, V_k \le 2u_k$. Moreover, $E_{P_n}(\phi_j(X) - a_j)^2 \le E_{P_n}\phi_j^2(X) = 1 + \int_0^1 \phi_j^2(x)(p_n(x) - 1)\, dx \le 1 + \int_0^1 \phi_j^2(x)|\sqrt{p_n(x)} - 1|(1 + e^{M/2})\, dx \le 1 + (1 + e^{M/2}) H(p_n, p_0) \{\int_0^1 \phi_j^4(x)\, dx\}^{1/2}$. Since $\{\int_0^1 \phi_j^4(x)\, dx\}^{1/2} \le V_j \{\int_0^1 \phi_j^2(x)\, dx\}^{1/2} = V_j$, $V_k = O(k^{\omega})$ and $m^{\beta} H(p_n, p_0) = O(1)$, the above yields $E_{P_n}(\phi_j(X) - a_j)^2 \le 1 + o(1)$. Hence $E_{P_n}|\phi(X) - a|_{jk}^2 \le (k-j)(1 + o(1))$ and

$$E_{P_n}|\phi(X) - a|_{jk}^s \le (k-j)(1 + o(1))\big(2\sqrt{k-j}\, V_k\big)^{s-2}, \qquad s \ge 2.$$

Thus Proposition 7.17 is applicable for the vector $(\phi_{j+1}(X), \dots, \phi_k(X))$ with $b_i^2 = (k-j)(1 + o(1))$, $\mathscr{K} = 2\sqrt{k-j}\, V_k$ and $B_n^2 = n(k-j)(1 + o(1))$. Taking $x = n c_n / B_n = c_n\{n/(k-j)\}^{1/2}(1 + o(1))$ and using (7.30), Proposition 7.17 implies (7.31). □

Below we assume that conditions (A1)–(A8) are satisfied. So, we shall consider the exponential family (2.1) and related notions (in particular S) under $m = m_n$.

For $\theta \in R^m$ and ρ_n given by (7.23), we set

$$\begin{aligned} l_n(\theta) = \min\bigg\{ & k : 1 \le k \le m, \\ & |\theta|_k^2 - k\rho_n \frac{\log n}{n} \ge |\theta|_j^2 - j\rho_n \frac{\log n}{n}, 1 \le j \le m \bigg\}. \end{aligned} \tag{7.32}$$

Let θ^* be given by (7.19). Define

$$l = l_n(\theta^*). \tag{7.33}$$

THEOREM 7.19. *Under assumptions* (A1)–(A8),

$$P_n(S < l) \to 0. \tag{7.34}$$

PROOF. By the definition of S,

$$\begin{aligned}\{S < l\} \subset \bigcup_{j=1}^{l-1} \Big\{ n \sup_{s \in R^j} \big(\overline{\phi} \circ s - \psi_j(s)\big) - \frac{j}{2} \log n \\ \geq n \sup_{t \in R^l} \big(\overline{\phi} \circ t - \psi_l(t)\big) - \frac{l}{2} \log n \Big\}.\end{aligned} \tag{7.35}$$

Theorem 7.6 and (7.35) imply that, for $\varepsilon \in (0, \min\{1, 2u_l^2/3\}]$ and sufficiently large n,

$$\{S < l\} \subset \bigcup_{j=1}^{l-1} \Big\{ |\overline{\phi}|_{jl}^2 \leq \frac{1}{2n}(2-\varepsilon)(l-j)\log n + \varepsilon\delta_* \Big\} \cup \big\{ |\overline{\phi}|_l^2 \geq \delta_* \big\},$$

where $\delta_* = (2-\varepsilon)\varepsilon^2/12u_l^2$. Now take $\varepsilon = (24)^{1/3} u_l |\theta^*|_l$. By Lemma 7.12 it follows that $\varepsilon \to 0$.

Define also

$$t_{jl}^2 = \frac{1}{2n}(2-\varepsilon)(l-j)\log n + 2(2-\varepsilon)u_l|\theta^*|_l^3.$$

By the above

$$\{S < l\} \subset \bigcup_{j=1}^{l-1} \big\{ |\overline{\phi}|_{jl} \leq t_{jl} \big\} \cup \big\{ |\overline{\phi}|_l^2 \geq \sigma^2 |\theta^*|_l^2 \big\}, \tag{7.36}$$

where $\sigma^2 = (2-\varepsilon)/3^{1/3}$. The triangle inequality and (7.16) then imply that $\{|\overline{\phi}|_{jl} \leq t_{jl}\} \subset \{|\overline{\phi} - a|_{jl} \geq |a|_{jl} - t_{jl}\} \subset \{|\overline{\phi} - a|_{jl} \geq |\theta^*|_{jl} - |a - \theta^*|_{jl} - t_{jl}\}$.

By $(x+y)^{1/2} \leq x^{1/2} + y^{1/2}$ we get $t_{jl} \leq \{n^{-1}(l-j)\log n\}^{1/2} + 2u_m^{1/2}|\theta^*|_m^{3/2}$. Moreover, (7.27) implies that $t_{jl} + |a - \theta^*|_{jl} \leq \{n^{-1}(l-j)\log n\}^{1/2} + (2 + o(1))u_m^{1/2}|\theta^*|_m^{3/2}$. On the other hand, by (7.33) and (7.32), $|\theta^*|_{jl}^2 \geq n^{-1}(l-j)\rho_n \log n$. So, these two inequalities, (7.23) and Corollary 7.15 yield $|\theta^*|_{jl} - |a - \theta^*|_{jl} - t_{jl} \geq c_{jl}$, where $c_{jl} = \{n^{-1}(l-j)\rho_n \log n\}^{1/2}(1 + o(1))$. Since

$$\begin{aligned} c_{jl} V_l (l-j)^{-1/2} &\leq \{\rho_n V_m^2 n^{-1} \log n\}^{1/2} \\ &\leq \Big\{ O(1)\kappa_n^{-1} \{m^\beta H(p_n, p_0)\}^{\omega/\beta} \Big\} \{H(p_n, p_0)\}^{(2\beta+1-2\omega)/2\beta}, \end{aligned}$$

the assumption (A7) and the definition of κ_n [see (7.21)] imply that c_{jl} satisfies (7.30) with $k = l$. Applying Corollary 7.18 with $c_n = c_{jl}$ and $k = l$, we get $P_n(|\overline{\phi}|_{jl} \leq t_{jl}) \leq P_n(|\overline{\phi} - a|_{jl} \geq c_{jl}) \leq 2\exp\{-(1 + o(1))\rho_n \log n\}$. Thus

$$P_n\Bigg(\bigcup_{j=1}^{l-1} \big\{ |\overline{\phi}|_{jl} \leq t_{jl} \big\} \Bigg) \leq 2\exp\big\{ -(1 + o(1))\rho_n \log n + \log l \big\}.$$

By (A8), $nH^2(p_n, p_0) \to \infty$, while, by (A7), $m^{2\beta}H^2(p_n, p_0) = O(1)$. Hence $m^{2\beta} < n$ for n sufficiently large. However, $\log l \le \log m \le (2\beta)^{-1} \log n$. This, together with (7.23), implies that the first term in (7.36) tends to 0.

To majorize the second term in (7.36), observe that

$$\{|\bar{\phi}|_l \ge \sigma|\theta^*|_l\} \subset \{|\bar{\phi} - a|_l \ge \sigma|\theta^*|_l - |\theta^*|_l - |a - \theta^*|_l\}.$$

As $\varepsilon \to 0$, using (7.32) and (7.21), for sufficiently large n we see that

$$(\sigma - 1)|\theta^*|_l \ge \frac{1}{8}\left\{|\theta^*|_m^2 - m\rho_n \frac{\log n}{n}\right\}^{1/2}$$
$$= \frac{1}{8}\left\{|\theta^*|_m^2 - \kappa_n^{-1} m\big(H(p_n, p_0)\big)^{2+1/\beta}\right\}^{1/2}.$$

This, together with (A7), (7.24) and the definition of κ_n, implies that $(\sigma - 1)|\theta^*|_l \ge \frac{1}{8}(1 + o(1))H(p_n, p_0)$. Thus it follows that

$$|\theta^*|_l \ge |\theta^*|_m + o(|\theta^*|_m). \tag{7.37}$$

In addition, by (7.24) and Lemma 7.16, $|a - \theta^*|_l = o(|\theta^*|_m) = o(H(p_n, p_0))$ and hence $\{|\bar{\phi}|_l \ge \sigma|\theta^*|_l\} \subset \{|\bar{\phi} - a|_l \ge \frac{1}{8}(1 + o(1))H(p_n, p_0)\}$.

Now set $c_n = \frac{1}{8}(1 + o(1))H(p_n, p_0)$. Due to (A7) and Lemma 7.12, c_n satisfies (7.30). Thus, by Corollary 7.18,

$$P_n\big(|\bar{\phi}|_l^2 \ge \sigma|\theta^*|_l^2\big) \le 2\exp\left\{-\frac{n}{128l}H^2(p_n, p_0)(1 + o(1))\right\}. \tag{7.38}$$

Since (A7) and (A8) imply

$$\frac{nH^2(p_n, p_0)}{l} \ge \frac{1}{m\{H(p_n, p_0)\}^{1/\beta}} \frac{n\{H(p_n, p_0)\}^{2+1/\beta}}{\log n} \log n > \log n$$

for sufficiently large n, the proof of Theorem 7.19 is concluded. □

Note that the above argument yields

$$\frac{n}{m}H^2(p_n, p_0) > \log n \tag{7.39}$$

for sufficiently large n.

7.5. *Proof of Theorem* 4.1. Theorem 4.1 follows from the two following results.

THEOREM 7.20. *Under* (A1)–(A8) *for any* $\varepsilon \in (0, 1)$,

$$\lim_{n\to\infty} P_n\left(\left|\frac{T_S}{n|a|_l^2} - 1\right| \le \varepsilon\right) = 1, \tag{7.40}$$

where l is defined by (7.33).

PROOF. We have

$$(7.41)\quad P_n\big(|n^{-1}|a|_l^{-2}T_S - 1| \ge \varepsilon\big) \le P_n(S < l) + \sum_{k=l}^{m} P_n\big(|\,|a|_l^{-2}|\bar{\phi}|_k^2 - 1| \ge \varepsilon\big).$$

By Theorem 7.19, $P_n(S < l) \to 0$. So, consider the second component on the right-hand side of (7.41). Notice that $l \le k \le m$ and that

$$(7.42)\quad \begin{aligned} &P_n\big(|\,|a|_l^{-2}|\bar{\phi}|_k^2 - 1| \ge \varepsilon\big) \\ &\qquad = P_n\big(|\bar{\phi}|_k^2 \ge (1+\varepsilon)|a|_l^2\big) + P_n\big(|\bar{\phi}|_k^2 \le (1-\varepsilon)|a|_l^2\big). \end{aligned}$$

Using the triangle inequality, the first component in (7.42) can be majorized as follows:

$$(7.43)\quad \begin{aligned} &P_n\big(|\bar{\phi}|_k \ge \sqrt{1+\varepsilon}|a|_l\big) \\ &\qquad \le P_n\big(|\bar{\phi}|_m \ge \sqrt{1+\varepsilon}|a|_l\big) \\ &\qquad \le P_n\big(|\bar{\phi} - a|_m \ge \sqrt{1+\varepsilon}|a|_l - |a|_m\big) \\ &\qquad \le P_n\left(|\bar{\phi} - a|_m \ge \sqrt{1+\varepsilon}|\theta^*|_l - |\theta^*|_m - \left(2 + \frac{\varepsilon}{2}\right)|a - \theta^*|_m\right). \end{aligned}$$

The definitions of l and ρ_n [see (7.33) and (7.23)], (A7) and the fact $\kappa_n \to \infty$ imply

$$(7.44)\quad |\theta^*|_m^2 - |\theta^*|_l^2 \le \frac{1}{n}\rho_n(m-l)\log n = o\big(H^2(p_n, p_0)\big).$$

By (7.24) and Lemma 7.16, $|a - \theta^*|_m = o(|\theta^*|_m) = o(H(p_n, p_0))$. This last fact, (7.43) and (7.44) give

$$(7.45)\quad P_n\big(|\bar{\phi}|_k \ge \sqrt{1+\varepsilon}|a|_l\big) \le P_n\big(|\bar{\phi} - a|_m \ge \big(\tfrac{1}{3}\varepsilon + o(1)\big)H(p_n, p_0)\big).$$

Arguing as in (7.38), we get by Corollary 7.18 that the first component of (7.42) is majorized by

$$(7.46)\quad 2\exp\left\{-\frac{1}{18m}n\varepsilon^2 H^2(p_n, p_0)(1 + o(1))\right\}$$

and, by (7.39), goes to 0 for every $\varepsilon > 0$.

The second component of (7.42) can be treated analogously to the first one. Indeed,

$$(7.47)\quad \begin{aligned} &P_n\big(|\bar{\phi}|_k \le \sqrt{1-\varepsilon}|a|_l\big) \le P_n\big(|\bar{\phi}|_l \le \sqrt{1-\varepsilon}|a|_l\big) \\ &\qquad \le P_n\left(|\bar{\phi}|_l \le \left(1 - \frac{\varepsilon}{2}\right)|a|_l\right) \le P_n\left(|\bar{\phi} - a|_l \ge \frac{\varepsilon}{2}|a|_l\right) \\ &\qquad \le P_n\left(|\bar{\phi} - a|_l \ge \frac{\varepsilon}{2}|\theta^*|_l - \frac{\varepsilon}{2}|a - \theta^*|_l\right). \end{aligned}$$

Arguing as before, we get the estimate

$$P_n\left(|\overline{\phi} - a|_l \geq \frac{1}{2}\varepsilon H(p_n, p_0)(1 + o(1))\right) \leq 2\exp\left\{-\frac{1}{8l}n\varepsilon^2 H^2(p_n, p_0)(1 + o(1))\right\}, \tag{7.48}$$

which goes to 0 by (7.39). Thus, by (A7) and (A8), it follows that the whole sum in (7.41) tends to 0 and the proof is concluded. □

Now we shall show that $|a|_l^2$ in (7.40) can be replaced by $2D(p_n \| p_0)$.

THEOREM 7.21. *Let assumptions* (A1)–(A8) *hold. Then*

$$\lim_{n\to\infty} \frac{|a|_l^2}{2D(p_n \| p_0)} = 1. \tag{7.49}$$

PROOF. The properties of p_m^* imply that

$$\begin{aligned} D(p_m^* \| p_0) &= \int_0^1 (\theta^* \circ \phi(x) - \psi_m(\theta^*)) p_m^*(x)\, dx = a \circ \theta^* - \psi_m(\theta^*) \\ &= |\theta^*|_m^2 + (a - \theta^*) \circ \theta^* - \psi_m(\theta^*). \end{aligned} \tag{7.50}$$

By (7.5) with $\theta = \theta^*$ and $c = c^* = u_m|\theta^*|_m$, (7.26) and Lemma 7.12, relation (7.50) yields

$$\begin{aligned} D(p_m^* \| p_0) &\leq |\theta^*|_m^2 + |\theta^*|_m |a - \theta^*|_m - \frac{1}{2}|\theta^*|_m^2 + \frac{c^*}{6}|\theta^*|_m^2 + \frac{(3 - c^*)^2}{72}|\theta^*|_m^4 \\ &= \frac{1}{2}|\theta^*|_m^2 + o\big(|\theta^*|_m^2\big) \end{aligned}$$

and

$$\begin{aligned} D(p_m^* \| p_0) &\geq |\theta^*|_m^2 - |\theta^*|_m |a - \theta^*|_m - \frac{1}{2}|\theta^*|_m^2 - \frac{c^*}{2}|\theta^*|_m^2 \\ &= \frac{1}{2}|\theta^*|_m^2 + o\big(|\theta^*|_m^2\big). \end{aligned}$$

This, together with (7.18) and Corollary 7.14, then implies that

$$2D(p_n \| p_0) = |\theta^*|_m^2 + o\big(|\theta^*|_m^2\big). \tag{7.51}$$

On the other hand, applying again Lemmas 7.16 and 7.12 and (7.37), we get

$$|a|_l \geq |\theta^*|_l - |a - \theta^*|_l \geq |\theta^*|_m + o(|\theta^*|_m)$$

and

$$|a|_l \leq |\theta^*|_l + |a - \theta^*|_l \leq |\theta^*|_m + o(|\theta^*|_m).$$

Hence

$$|a|_l^2 = |\theta^*|_m^2 + o\big(|\theta^*|_m^2\big), \tag{7.52}$$

and the proof of (7.49) is concluded. □

REMARK 7.22. By (7.51) and (7.52), relations (7.40) and (7.49) imply the statement of Theorem 4.1.

7.6. *Remarks on Thoerem* 6.10. Theorem 6.10 is based on:

THEOREM 7.23. *Let* $\{k_n\}$ *be a sequence of natural numbers satisfying* $1 \le k_n \le m = m_n$, $n \ge 1$. *Assume* $\liminf_n\{|a|_{k_n}/|a|_m\} > 0$. *Then under* (A1)–(A11), *it holds that*

$$\left(T_{k_n} - n|a|_{k_n}^2\right)\Big/\left(2\sqrt{n}\,|a|_{k_n}\right) \to N(0,1).$$

Theorem 7.23 can be proved via an application of Hungarian embedding in a way similar to that proposed in the proof of Theorem 3.1 in Inglot, Jurlewicz and Ledwina (1990a) [cf. also Inglot, Jurlewicz and Ledwina (1990b)]. Therefore details are omitted here.

To get Theorem 6.10 set

$$l^* = \min\left\{i: 1 \le i \le m: |a|_i^2 - 2i\rho_n\frac{\log n}{n} \ge |a|_j^2 - 2j\rho_n\frac{\log n}{n}, i = 1,\dots,m\right\}.$$

It can be seen that, under the imposed assumptions and n large enough, $l^* \le l$, where l is given by (7.33). Applying now Theorem 7.23 to $k_n = l$ and $k_n = m$ and using Theorem 7.19, Theorem 6.10 follows.

REMARK 7.24. Consider the model (a) of Example 6.8. Suppose that $g(x) = \sum_{j=1}^s \omega_j\phi_j(x)$ for some fixed s and $\omega_1,\dots,\omega_s$. Since $|a|_m = 0(H(p_n, p_o)) = 0(n^{-\xi})$ it follows that $l^* = s$ and $|a|_{l^*m} = 0$ for n large enough. Therefore (A12) holds. The same can be shown for models (b) and (c) of Example 6.8 provided $w(x)$ and $[z(x)]^\sigma$ have the same form as $g(x)$.

For more general analysis of (A12), see Inglot (1996).

7.7. *Local equivalence of efficiencies*. The purpose of this subsection is to show that the Nikitin (1984) result mentioned in Section 1 can be extended to other notions of efficiency.

Let T stand for one of the statistics KS or CM. Nikitin (1984) assumed the alternative distribution function is $G = G(x;\theta)$ for $\theta \in \Theta$ and G smooth in both arguments. In particular, it was assumed that

$$G(x;\theta) = G(x;0) + \theta G_\theta'(x;0) + o(\theta), \tag{7.53}$$

where $G_\theta'(x;\theta) = (\partial/\partial\theta)G(x;\theta)$ and that $G_\theta'(x;0) + \theta^{-1}o(\theta)$ is bounded for $x \in R$. The null hypothesis corresponds to $\theta = 0$. Denoting by $c_T(\theta)$ the exact Bahadur slope of T under an alternative $G(x;\theta)$ and by $2\mathscr{K}(\theta)$ the slope of the Neyman–Pearson test for $G(x;0)$ against $G(x;\theta)$, Nikitin (1984) in fact has shown that

$$\lim_{\theta\to 0}\frac{c_T(\theta)}{2\mathscr{K}(\theta)} = h(T,G) \tag{7.54}$$

for some h depending on the statistic T and the distribution G. Restricting attention to a location or scale family, he proved that the equation $h(T,G) = 1$ has a unique solution (in G) for given T.

A look at Theorem 7.2 in Bahadur (1971) and Lemma 2.1 and Remark 2.1 in Kallenberg (1983) shows that (7.54) shall be equal to any notion of Kallenberg's i-efficiency provided that condition (i) of Kallenberg's Lemma 2.1 is fulfilled and T satisfies

$$\lim_{t\to 0}\lim_{n\to\infty}(nt^2)^{-1}\log P_0\big(T\ge t\sqrt{n}\big) = \lim_{n\to\infty}\big(nt_n^2\big)^{-1}\log P_0\big(T\ge t_n\sqrt{n}\big) \tag{7.55}$$

for any $t_n \to 0$ in such a way that $t_n\sqrt{n}\to\infty$.

The equality (7.55) for KS and CM has been established in Inglot and Ledwina (1990). So, the argument will be concluded if we check that under (7.53) (with the imposed regularity conditions) assumption (i) is fulfilled for KS and CM. To this end, consider first (7.53) with $\theta=\theta_n\to 0$. Using the transformation $G^{-1}(x;0)$ and Nikitin's regularity conditions, we can restrict attention to distributions concentrated on $(0,1)$ and alternatives having densities

$$q(x;\theta_n) = 1+\theta_n g_n(x),\qquad x\in(0,1), \tag{7.56}$$

with $g_n(x)$ uniformly bounded and satisfying $\int_0^1 g_n(x)\,dx = 0$ and $\inf_n\int_0^1 g_n^2(x)\,dx > 0$.

Next, observe that both KS and CM have the form $T=\sqrt{n}\|F_n - F_0\|$, where F_n is the empirical distribution function, $F_0(x)=x$, $x\in(0,1)$, and $\|\cdot\|$ is a suitable norm. Moreover, in both cases it holds that $\|\cdot\|\le c\|\cdot\|_\infty$ for some positive c. Set $b(\theta)=\|Q_\theta - F_0\|$, where $Q_\theta(x)=\int_0^x q(y;\theta)\,dy$. Let $H(Q_\theta, F_0)$ stand for the Hellinger distance between densities of Q_θ and F_0.

To verify condition (i), we have to show that, for all sequences $\{\theta_n\}$ with

$$H\big(Q_{\theta_n},F_0\big)\to 0 \quad\text{and}\quad nH^2\big(Q_{\theta_n},F_0\big)\to\infty \quad \text{as } n\to\infty, \tag{7.57}$$

it holds that

$$\lim_{n\to\infty} P_{\theta_n}\left\{\left|\frac{T}{\sqrt{n}\,b(\theta_n)} - 1\right| < \varepsilon\right\} = 1 \quad\text{for each } \varepsilon>0. \tag{7.58}$$

The structure of T and the Dvoretzky–Kiefer–Wolfowitz exponential inequality yield

$$\begin{aligned} P_{\theta_n}\left\{\left|\frac{T}{\sqrt{n}\,b(\theta_n)} - 1\right| < \varepsilon\right\} &\ge P_{\theta_n}\big\{\|F_n - Q_{\theta_n}\| < \varepsilon\|Q_{\theta_n} - F_0\|\big\}\\ &\ge P_{\theta_n}\left\{\sqrt{n}\|F_n - Q_{\theta_n}\|_\infty < \frac{\varepsilon}{c}\sqrt{n}\|Q_{\theta_n}-F_0\|\right\}\\ &\ge 1 - 58\exp\left\{-2\left(\frac{n\varepsilon^2}{c^2}\right)\|Q_{\theta_n}-F_0\|^2\right\}. \end{aligned}$$

So, (7.58) follows provided that, under (7.56) and (7.57),

$$n\|Q_{\theta_n} - F_0\|^2\to\infty. \tag{7.59}$$

However, under (7.56), $H^2(Q_{\theta_n}, F_0) \le C\theta_n^2$ for some positive constant C and it is enough to show that the condition $n\theta_n^2 \to \infty$ implies (7.59). Check first the last implication for CM. Since in this case $\|Q_{\theta_n} - F_0\| = \theta_n C_n$, where C_n is bounded, the conclusion follows. A similar argument works for KS.

Acknowledgments. We would like to thank three referees for constructive comments on the preliminary version of this paper. These have led to significant improvements in the presentation. The second author is very grateful to R. L. Eubank and W. C. M. Kallenberg, who kindly sent her a copy of the dissertation by Kim (1992) and the report by Book (1976), respectively. She thanks also W. C. M. Kallenberg for helpful discussions.

REFERENCES

BAHADUR, R. R. (1971). *Some Limit Theorems in Statistics*. SIAM, Philadelphia.

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

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

BOGDAN, M. (1995). Data driven versions of Pearson's chi-square test for uniformity. *J. Statist. Comput. Simulation* **52** 217–237.

BOGDAN, M. and LEDWINA, T. (1996). Testing uniformity via log-spline modeling. *Statistics*. **28** 131–157.

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

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

FAN, J. (1996). Tests of significance based on wavelet thresholding and Neyman's truncation. *J. Amer. Statist. Assoc.* **91** 647–688.

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

INGLOT, T. (1996). Generalized intermiediate efficiency of goodness of fit tests. Technical Report No. 18, Institute of Mathematics, Technical Univ. Wrocław.

INGLOT, T., JURLEWICZ, T. and LEDWINA, T. (1990a). On Neyman-type smooth tests of fit. *Statistics* **21** 549–568.

INGLOT, T., JURLEWICZ, T. and LEDWINA, T. (1990b). Asymptotics for multinomial goodness of fit tests for simple hypothesis. *Theory Probab. Appl.* **35** 797–803.

INGLOT, T., KALLENBERG, W. C. M. and LEDWINA, T. (1993). Asymptotic behavior of some bilinear functionals of the empirical process. *Math. Methods Statist.* **2** 316–336.

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

INGLOT, T. and LEDWINA, T. (1990). On probabilities of excessive deviations for Kolmogorov–Smirnov, Cramér–von Mises and chi-square statistics. *Ann. Statist.* **18** 1491–1495.

JUREČKOVÁ, J., KALLENBERG, W. C. M. and VERAVERBEKE, N. (1988). Moderate and Cramér-type large deviation theorems for *M*-estimators. *Statist. Probab. Lett.* **6** 191–199.

KAIGH, W. D. (1992). EDF and EQF orthogonal component decompositions and tests of uniformity. *J. Nonparametric Statist.* **1** 313–334.

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

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

KALLENBERG, W. C. M. and DROST, F. C. (1993). Comparison of tests, and local families. In *Statistics and Probability: A Raghu Raj Bahadur Festschrift* (J. K. Ghosh, S. K. Mitra, K. R. Parthasarathy and B. L. S. Prakasa Rao, eds.) 303–324. Wiley Eastern, New Delhi.

KALLENBERG, W. C. M. and LEDWINA, T. (1995a). Consistency and Monte Carlo simulation of a data driven version of smooth goodness-of-fit tests. *Ann. Statist.* **23** 1594–1608.

KALLENBERG, W. C. M. and LEDWINA, T. (1995b). On data driven Neyman's tests. *Probab. Math. Statist.* **15** 409–426.

KIM, J. (1992). Testing goodness-of-fit via order selection criteria. Ph.D. dissertation, Texas A & M Univ.

KONING, A. J. (1992). Approximation of stochastic integrals with applications to goodness-of-fit tests. *Ann. Statist.* **20** 428–454.

KONING, A. J. (1993). Stochastic integrals and goodness-of-fit tests. *Math. Center Tracts* **98**. Math. Centrum, Amsterdam.

KOUROUKLIS, S. (1984). A large deviation result for the likelihood ratio statistic in exponential families. *Ann. Statist.* **12** 1510–1521.

LEDWINA, T. (1994). Data driven version of Neyman's smooth test of fit. *J. Amer. Statist. Assoc.* **89** 1000–1005.

MILBRODT, H. and STRASSER, H. (1990). On the asymptotic power of the two-sided Kolmogorov–Smirnov test. *J. Statist. Plann. Inference* **26** 1–23.

MILLER, F. L. and QUESENBERRY, C. P. (1979). Power studies of tests for uniformity. II. *Comm. Statist. B—Simulation Comput.* **8** 271–290.

NEUHAUS, G. (1976). Asymptotic power properties of the Cramér–von Mises test under contiguous alternatives. *J. Multivariate Anal.* **6** 95–110.

NEYMAN, J. (1937). "Smooth test" for goodness of fit. *Skand. Aktuarietidskr.* **20** 149–199.

NIKITIN, YA. YU. (1984). Local Bahadur optimality and characterization problems. *Theory Probab. Appl.* **29** 79–92. (In Russian).

NIKITIN, YA. YU. (1995). *Asymptotic Efficiency of Nonparametric Tests.* Cambridge Univ. Press.

OOSTERHOFF, J. and VAN ZWET, W. R. (1979). A note on contiguity and Hellinger distance. In *Contributions to Statistics, J. Hájek Memorial Volume* (J. Jurečkova, ed.) 157–166. Academia, Prague.

PETROV, V. V. (1975). *Sums of Independent Random Variables*. Akademie, Berlin.

PORTNOY, S. (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *Ann. Statist.* **16** 356–366.

PROHOROV, A. V. (1973). On sums of random vectors. *Theory Probab. Appl.* **18** 186–188.

QUESENBERRY, C. P. and MILLER, F. L. (1977). Power studies of some tests for uniformity. *J. Statist. Comput. Simulation* **5** 169–191.

RAYNER, J. C. W. and BEST, D. J. (1989). *Smooth Tests of Goodness of Fit.* Oxford Univ. Press.

RAYNER, J. C. W. and BEST, D. J. (1990). Smooth tests of goodness of fit: an overview. *Internat. Statist. Rev.* **58** 9–17.

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

SCHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461–464.

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

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

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