The Annals of Statistics
2000, Vol. 28, No. 1, 239–253

GLOBAL POWER FUNCTIONS OF GOODNESS OF FIT TESTS

BY ARNOLD JANSSEN

Heinrich-Heine-University of Düsseldorf

It is shown that the global power function of any nonparametric test is flat on balls of alternatives except for alternatives coming from a finite dimensional subspace. The present benchmark is here the upper one-sided (or two-sided) envelope power function. Every choice of a test fixes a priori a finite dimensional region with high power. It turns out that also the level points are far away from the corresponding Neyman–Pearson test level points except for a finite number of orthogonal directions of alternatives. For certain submodels the result is independent of the underlying sample size. In the last section the statistical consequences and special goodness of fit tests are discussed.

1. Introduction. Omnibus tests are commonly used if the specific structure of certain nonparametric alternatives is unknown. Among other justifications, it turns out that they typically are consistent against fixed alternatives and $\sqrt{n}$-consistent under sequences of local alternatives of sample size n. For these reasons, people often trust in goodness of fit tests and these are frequently applied to data of finite sample size.

On the other hand, every asymptotic approximation should be understood as an approximation of the underlying finite sample case. Thus the statistician likes to distinguish and to compare the power of different competing tests.

The present paper offers a concept for the comparison and justification of different tests by their power functions and level points. It is shown that under certain circumstances every test has a preference for a finite dimensional space of alternatives. Apart from this space, the power function is almost flat on balls of alternatives. There exists no test which pays equal attention to an infinite number of orthogonal alternatives. The results do not only hold for asymptotic models but they also hold for concrete alternatives on the real line at finite sample size and their level points uniformly for the sample size.

The results are not surprising. Every statistician knows that it is impossible to separate an infinite sequence of different parameters simultaneously if only a finite number of observations is available.

The conclusions of the results are two-fold.

1. The statistician should analyze the goodness of fit tests of his computer package in order to get some knowledge and an impression about their preferences.

Received May 1998; revised December 1999.

AMS 1991 *subject classifications*. 62G10, 62G20.

Key words and phrases. Goodness of fit test, Kolmogorov–Smirnov test, power function, envelope power function, curvature of power functions, level points, data driven Neyman's smooth test, Pitman efficiency, Bahadur efficiency, intermediate efficiency.

2. A well-reflected choice of tests requires some knowledge about preferences concerning alternatives which may come from the practical experiment. A guide (and summary) to the construction of tests is given in Section 3.

The present results will be explained for a one-sample goodness of fit problem.

EXAMPLE 1.1. Suppose that $X_1, \dots, X_n$ are real i.i.d. random variables with joint distribution $P \in \mathscr{P}$ and distribution function F_P where $\mathscr{P}$ is a nonparametric subset of all continuous distributions on $\mathbb{R}$. Suppose that the testing problem is given by the simple null hypothesis

$$H_0 = \{P_0\} \tag{1.1}$$

at sample size n. Two concrete cases should be kept in mind.

(a) *One-sided testing problem.* Suppose that the alternatives

$$\mathscr{P} \setminus \{P_0\} \subset \{P\colon F_P \le F_{P_0}\} \tag{1.2}$$

are stochastically larger than F_{P_0}.

(b) *Two-sided testing problem.* In the general case given by testing H_0 against all continuous alternatives, the tests are restricted to the class of unbiased level α tests ϕ_n with $E_{Q^n}(\phi_n) \ge \alpha$ for all $Q \in \mathscr{P}$, where ϕ_n is a test which is based on $X_1, \dots, X_n$.

In both situations we have an upper envelope power function at level α on $\mathscr{P}$,

$$Q \mapsto \beta(Q^n)\colon = \sup(E_{Q^n}(\phi_n)), \tag{1.3}$$

where the supremum is taken over the present class of level α tests. In the case of Example 1.1(a) this supremum is just the power of the Neyman–Pearson test of P_0^n against Q^n.

The power function (1.3) is now the benchmark which should be compared with the nonparametric power function of a given test.

Throughout, we will give a brief survey about related work dealing with power functions for nonparametric tests. A principle component decomposition of goodness of fit tests has been studied by Anderson and Darling (1952) and Durbin and Knott (1972), see also Shorack and Wellner (1986). Bounds for global power functions of two-sided tests were obtained by Strasser (1990). They rely on an extrapolation of their curvature at the null hypothesis; see Milbrodt and Strasser (1990) and Janssen (1995). Global power functions of one-sided Kolmogorov–Smirnov tests for a restricted class of alternatives were obtained by Anděl (1967) and Hájek and Šidák (1967).

For sparse sets of nonparametric alternatives [compared with the omnibus alternatives of Example 1.1(b)] minimax tests were established by various authors. We refer to Ingster (1993), Lepski and Spokoiny (1999) and references therein. They deal with the related signal detection problem for the

Wiener process. Detailed information about the Hodges–Lehmann and Bahadur asymptotic efficiency for goodness of fit tests can be found in Nikitin (1995).

Neyman's smooth tests are also very popular; see Neyman (1937). There is much interest in extensions and data driven tests of this kind which are then applied to omnibus alternatives, see Bickel and Ritov (1992), Eubank, Hart and LaRiccia (1993), Kallenberg and Ledwina (1995), Inglot and Ledwina (1996) and Inglot, Kallenberg and Ledwina (1998). However, the estimator of the dimension should not be too restrictive since otherwise these data driven tests are close to parametric procedures; see the detailed discussion in Section 3. It is also pointed out that the Pitman efficiency the local Bahadur efficiency, the Hodges–Lehmann efficiency and the intermediate efficiency may be different. This discussion may explain some paradoxial results for the Kolmogorov–Smirnov test and data driven tests.

The present paper is organized as follows. Section 2 contains the treatment of the power function for the Brownian bridge shift experiment, which is just the limit model of Example 1.1. It is shown that only a sparse space of alternatives has sufficient high power. In Section 3 it is shown that these results uniformly hold for each sample size n. Thus the same gap shows up for finite sample size and the asymptotic model as well. Special attention is devoted to the behavior of level points which yield a good tool for the comparison of different tests. The practical consequences are also discussed in that section, which includes comments about data driven tests.

2. Tests for the Brownian bridge B_0. In a first step, asymptotic power functions are compared with their upper envelope power functions. Notice that the asymptotic considerations of (one- and two-sample) goodness of fit tests lead to a shift experiment,

$$B_0(t) + \int_0^t h(u)\, du, \qquad 0 \le t \le 1 \tag{2.1}$$

on $C[0, 1]$ for the Brownian bridge $B_0(\cdot)$ with parameter space

$$H = L_2^0(\boldsymbol{\lambda}_{|(0,1)}) := \{h \in L_2(\boldsymbol{\lambda}_{|(0,1)}) : \int h \, d\boldsymbol{\lambda}_{|(0,1)} = 0\} \tag{2.2}$$

which is endowed with the natural inner product $< h_1, h_2 >= \int h_1 h_2 \, d\boldsymbol{\lambda}_{|(0,1)}$, where $\boldsymbol{\lambda}_{|(0,1)}$ is the uniform distribution on the unit interval. We refer to Strasser [(1985), Chapters 11, 13 (82.23)], Shorack and Wellner (1986) and Milbrodt and Strasser (1990). More details about abstract Wiener spaces can be found in Bouleau and Hirsch (1991). The model (2.1) is a standard Gaussian shift $G = (\Omega, \mathscr{A}, \{P_h : h \in H\})$ [with distribution P_h of (2.1)] in the following sense.

Let $(H, < \cdot, \cdot >)$ be a real Hilbert space with distributions $P_h \ll P_0$ for each $h \in H$. The likelihood is given by

$$\log \frac{dP_h}{dP_0} = L(h) - \|h\|^2/2, \tag{2.3}$$

where $h \mapsto L(h)$ is a linear centered Gaussian process w.r.t. P_0 and covariance $\mathrm{Cov}(L(h_1), L(h_2)) =< h_1, h_2 >$; see Strasser (1985), Section 68. In the case of (2.1) the process $L(h)$ is just the stochastic integral $L(h) = \int h(u)B_0(du)$. Another example similar to (2.1) is the Gaussian white noise shift model which was analyzed by Drees and Milbrodt (1991, 1994), for instance.

Since the present results only use the likelihood structure (2.3) we will consider an arbitrary Gaussian shift G. The benchmark for testing the null hypothesis $\{P_0^n\}$ can easily be calculated. Notice that the Neyman–Pearson envelope power function at level α and sample size n for $\{P_0^n\}$ against $\{P_h^n\}$ is given by

$$\beta_1(h, n):= \phi(n^{1/2}\|h\| - u_{1-\alpha}) \tag{2.4}$$

and its two-sided counterpart for unbiased testing [see Example 1.1(b)] is just

$$\beta_2(h, n):= \Phi(n^{1/2}\|h\| - u_{1-\alpha/2}) + \Phi(-n^{1/2}\|h\| - u_{1-\alpha/2}), \tag{2.5}$$

where $u_{1-\alpha} = \Phi^{-1}(1-\alpha)$ denotes the $(1-\alpha)$-quantile of the standard normal distribution function Φ.

Within an arbitrary Gaussian shift we will now show that the power function of any test is almost flat except for certain directions given by a finite dimensional subspace of alternatives. Let $V^\perp \subset H$ denote the orthogonal complement of a linear subspace V of H. Due to the stability of the Gaussian shift we have $\beta_i(h, n) = \beta_i(n^{1/2}h, 1)$ for $i = 1, 2$ and we will restrict ourselves first to $n = 1$.

THEOREM 2.1. *Let ϕ be any test with $E_{P_0}(\phi) = \alpha$, $0 < \alpha < 1$, for the null hypothesis $\{P_0\}$ of the Gaussian shift G. For each $\varepsilon > 0$ and $K > 0$ there exists a linear subspace $V \subset H$ of finite dimension with*

$$\sup\left\{|E_{P_h}(\phi) - \alpha|: h \in V^\perp,\ \|h\| \le K\right\} \le \varepsilon. \tag{2.6}$$

Moreover the following upper bound:

$$dim(V) - 1 \le \varepsilon^{-1}\alpha(1-\alpha)\big(\exp\left(K^2\right) - 1\big) \tag{2.7}$$

holds for the dimension of V which is independent of the test ϕ.

REMARK 2.1. (a) Compared with (2.4) or [2.5] the power of ϕ is poor on $V^\perp \cap \{h: \|h\| \le K\}$. Notice that the envelope power functions (2.4) [(2.5)] are attained by one-sided (two-sided) Neyman–Pearson tests, respectively.

(b) By Lemma 2.1 below, the total amount of squares of the power higher than α of any test is limited for orthonormal directions. Thus the statistician has to decide how he can distribute his power on different underlying directions. The directions of alternatives and their subspaces stand for preferences which should reflect the priority of the given experiment. The restriction to some important preferences can be helpful in practice.

LEMMA 2.1. *Let $(h_i)_{i\in I}$ be an orthonormal system in the parameter space H. For each constant $K > 0$ we have*

$$\sum_{i\in I}(\sup\{|E_{th_i}(\phi)-\alpha|: |t| \le K\})^2 \le \alpha(1-\alpha)\big(\exp\big(K^2\big)-1\big). \tag{2.8}$$

PROOF. We will start with a family $K_i > 0$, $i \in I$, of positive reals. Since $t \mapsto |E_{th_i}(\phi)-\alpha|$ is continuous for each $i \in I$ the function attains its maximum at some point t_i on $[-K_i, K_i]$. Consider now $i \in I_0$: $= \{j \in I: t_j \ne 0\}$ and the $L_2(P_0)$ functions,

$$f_i: = \exp\big(t_i L(h_i) - t_i^2/2\big) - 1. \tag{2.9}$$

This definition leads to $E_{t_i h_i}(\phi) - \alpha = \int \phi f_i \, dP_0 = \mathrm{Cov}(\phi, f_i)$. Observe that $(L(h_i))_{i\in I_0}$ are i.i.d. standard normal random variables under P_0. Thus $(f_i)_{i\in I_0}$ is a system of independent random variables. By (2.3) we have $E_0(f_i) = 0$ and

$$0 < \int f_i^2 \, dP_0 = \mathrm{Var}_{P_0}(f_i) = \exp\big(t_i^2\big) - 1 \le \exp\big(K_i^2\big) - 1. \tag{2.10}$$

Let now $\beta_i = \mathrm{Cov}(\phi, f_i)/\int f_i^2 \, dP_0$ be the Fourier coefficient of ϕ in direction f_i. Thus there exists an $L_2(P_0)$ function ϕ_r with

$$\phi - \alpha = \sum_{i\in I_0} \beta_i f_i + \phi_r \tag{2.11}$$

in $L_2(P_0)$ where $\mathrm{Cov}(f_i, \phi_r) = 0$ holds for each $i \in I_0$.

By standard L_2-arguments we have

$$\begin{aligned}\sum_{i\in I_0} \beta_i^2 \int f_i^2 \, dP_0 &= \sum_{i\in I_0} \mathrm{Var}(\beta_i f_i)\\ &= \mathrm{Var}\Big(\sum_{i\in I_0} \beta_i f_i\Big) \le \mathrm{Var}(\phi-\alpha)\\ &= E_0(\phi^2) - \alpha^2 \le \alpha(1-\alpha).\end{aligned} \tag{2.12}$$

Statement (2.12) combined with (2.10) now implies the result if we put $K_i = K$. □

PROOF OF THEOREM 2.1. Let ε and K be fixed. Below we can construct by induction an orthonormal system $(h_i)_{i\in\mathbb{N}}$ in H, linear subspaces $V_n = sp(h_1, \dots, h_n)$, generated by $h_1, \dots, h_n$ with $V_0 = \{0\}$, and a sequence of reals $(t_i)_{i\in\mathbb{N}}$, $|t_i| \le K$, with the following properties: for each $n \ge 0$ we have

$$\begin{aligned}(\sup\{|E_{th}(\phi)-\alpha|: |t| \le K,\ h \in V_n^{\perp},\ \|h\| = 1\})^2&\\ \le (E_{t_{n+1}h_{n+1}}(\phi)-\alpha)^2 + \varepsilon/2^{n+1}&\\ =: a_{n+1} + \varepsilon/2^{n+1}.&\end{aligned} \tag{2.13}$$

According to Lemma 2.1 we have $a_n \to 0$ as $n \to \infty$. Let now m denote the smallest positive integer with $a_m + \varepsilon/2^m \le \varepsilon$. Thus we arrive at

$$\begin{aligned}(m-1)\varepsilon &\le \sum_{j=1}^{m-1} (a_j + \varepsilon/2^j)\\ &\le \alpha(\alpha-1)(\exp(K^2)-1) + \varepsilon.\end{aligned} \tag{2.14}$$

If we choose $V = V_{m-1}$ then our construction (2.13) implies (2.6) and (2.7) follows from (2.14). □

Different tests will now be compared by the examination of their level points along one-parametric subalternatives $(P_{th})_{t\in\mathbb{R}}$, where $h \in H$ is a normalized direction with $\|h\| = 1$. Let β be a level with $\alpha < \beta < 1$. For instance, $1-\beta$ may be the error probability of second kind which is just acceptable. The level point (lp) of a test ϕ in direction h is (up to the sign) the smallest distance $|s|$ from the null hypothesis zero where the power β is attained, namely

$$\mathrm{lp}(\phi, \beta, h) := \inf\{|s|: E_{sh}(\phi) \geq \beta\} \tag{2.15}$$

[and let (2.15) be infinite if the power is always below β]. In addition, let $\mathrm{lp}_1(\beta, n)$ and $\mathrm{lp}_2(\beta, n)$ be the level points of the envelope power function β_j of (2.4) and (2.5), respectively, which actually belong to one-sided or two-sided Neyman–Pearson tests for $j = 1, 2$. Theorem 2.1 has further consequences. We will only treat $\beta_2(\cdot)$ in Lemma 2.2(b) and their level points $\mathrm{lp}_2(\cdot)$. The results for one-sided tests are similar.

LEMMA 2.2. *Let ϕ: $\Omega \to [0, 1]$ be any test for the Gaussian shift with $E_{P_0}(\phi) = \alpha$.*

(a) *For each $\varepsilon > 0$ and $K > 0$ there exists a linear subspace $W \subset H$ of finite dimension with*

$$\frac{|E_h(\phi) - \alpha|}{\beta_1(h, 1) - \alpha} \leq \varepsilon \tag{2.16}$$

for all $h \in (W^\perp \setminus \{0\}) \cap \{h: \|h\| \leq K\}$.

(b) *For each constant $C > 0$ and $\alpha < \beta < 1$ there exists a linear subspace $U \subset H$ of finite dimension with*

$$\frac{\mathrm{lp}(\phi, \beta, h)}{\mathrm{lp}_2(\beta, 1)} \geq C \tag{2.17}$$

for all directions $h \in U^\perp \cap \{h: \|h\| = 1\}$. The dimension of the subspace U is bounded by some constant $k(\alpha, \beta, C)$ which is independent of the test ϕ.

PROOF. (a) The power function (2.4) admits a Taylor expansion,

$$\beta_1(th, 1) - \alpha = t\, \dot{\Phi}(-u_{1-\alpha}) + o(t) \tag{2.18}$$

for each h, $\|h\| = 1$, at $t = 0$. If the assertion (a) is violated then there exists an $\varepsilon > 0$ such that (2.16) does not hold for each subspace of finite dimension. There exists a sequence $g_i \neq 0$ of orthogonal elements in H, $\|g_i\| \leq K$, with

$$|E_{g_i}(\phi) - \alpha| \geq \varepsilon(\beta_1(g_i, 1) - \alpha). \tag{2.19}$$

Define $K_i = \|g_i\|$ and $h_i = K_i^{-1} g_i$. Obviously we have $K_i \to 0$ as $i \to \infty$ by (2.6). On the other hand we may apply (2.12) for $t_i = K_i$ which yields

$$\sum_{i=1}^{\infty} \left(|E_{g_i}(\phi) - \alpha| \left(\exp\left(K_i^2\right) - 1\right)^{-1/2}\right)^2 \leq \alpha(1-\alpha). \tag{2.20}$$

Since $(\beta_1(g_i, 1) - \alpha)(\exp(K_i^2) - 1)^{-1/2} \stackrel{\cdot}{\to} \Phi(-u_{1-\alpha})$ by (2.18) the statement (2.19) contradicts (2.20) and part (a) is proved.

(b) The result and (2.17) immediately follow from Theorem 2.1. □

A further Gaussian shift model is the Wiener process with noise $B(t) + \int_0^t h(u)\,du$, $0 \le t \le 1$, where in contrast to (2.1) all $L_2(\boldsymbol{\lambda}_{|(0,1)})$ functions h serve as parameters. For this model the minimax testing problem and the minimax rate of testing are discussed in detail; see Ingster (1993) and Lepski and Spokoiny (1999) and references therein, for instance. As conclusion it is well known that $H_0 = \{h = 0\}$ and alternatives which are too large cannot be distinguished in the minimax sense. Under extra conditions given by smoothness parameters for h, the minimax rate of testing is established in the literature.

The following condition is necessary for the minimax distinguishability of $\{h = 0\}$ and alternatives $\mathscr{K}$: *for each* $A > 0$ *the intersection* $\mathscr{K} \cap \{h \in L_2(\boldsymbol{\lambda}_{|(0,1)})\colon \|h\|^2 = A\}$ *includes only a finite number of orthogonal vectors*. We will point out that this result [see Burnashev (1979), page 114 and Ingster (1993), Theorem 2.3] is related to the present results. Notice that Lemma 2.1 yields a lower bound for the minimax bound for testing $\{h = 0\}$ versus $\{h_i\colon 1 \le i \le M\}$ where h_i are M mutually orthogonal elements of $L_2(\boldsymbol{\lambda}_{|(0,1)})$ with $\|h_i\|^2 = A$. Let $m = \min\{E_{P_{h_i}}(\phi)\colon 1 \le i \le M\}$ be the minimax power of a test ϕ with $E_{P_0}(\phi) = \alpha$. Then Lemma 2.1 implies

$$M(m - \alpha)^2 \le \alpha(1 - \alpha)(\exp(A) - 1). \tag{2.21}$$

Thus the minimax bound of the sum of error probabilities of first and second kind is given by

$$\alpha + (1 - m) \ge 1 - (\alpha(1 - \alpha))^{1/2}(M^{-1}(\exp(A) - 1))^{1/2}, \tag{2.22}$$

which is slighly sharper than the related bound of Burnashev [(1979), Corollary 2]. After the obviously missing brackets are inserted, the formula of Burnashev reads as

$$\alpha + \beta(\alpha, S) \ge 1 - \tfrac{1}{2} \sup_A \sqrt{M^{-1}(A)(\exp(A) - 1)}, \tag{2.23}$$

which follows his notation. The inequality (2.22) holds for every Gaussian shift. So far the present results are related to nonparametric minimax testing.

3. A discussion about global power functions. The results for the limit model (2.1) given in Section 2 have some important consequences for the goodness of fit testing problem on the real line for fixed sample size n; see Example 1.1. One might have the impression that the situation for the limit model (2.1) of Example 1.1 may be worse or better fixed sample size than the model. Our next example shows that it is here neither better nor worse in general, if all nonparametric alternatives have equal rights. Without loss of generality we will discuss goodness of fit tests for the uniform distribution on the unit interval $(0,1)$ for H_0 given in (1.1).

Throughout, let P_h be the distribution of (2.1) and let

$$G = (C[0,1], \mathscr{B}(C[0,1]), \{P_h\colon h \in H\}) \tag{3.1}$$

with $H = L_2^0(\boldsymbol{\lambda}_{|(0,1)})$ be the Brownian bridge shift model. Since $C[0,1]$ is a polish space we can find a Borel isomorphism $T\colon C[0,1] \to \mathbb{R}$ on the real line, see Parthasarathy [(1967), Section 2, Theorem 2.12].

The mapping T is one-to-one with measurable inverse. After an obvious manipulation with inverse distribution functions we may assume that

$$Q_0\colon = \boldsymbol{\lambda}_{|(0,1)} = \mathscr{L}(T|P_0) \tag{3.2}$$

is the uniform distribution on $(0,1)$. Setting

$$Q_h = \mathscr{L}(T|P_h) \tag{3.3}$$

we will now consider the submodel $F = \{\mathbb{R}, \mathscr{B}, \{Q_h\colon h \in H\}\}$ with

$$\frac{dQ_h}{dQ_0} = \exp(L(h) \circ T^{-1} - \|h\|^2/2)$$

of all continuous distributions on the real line and goodness of fit tests for (3.2) at sample size n. In this connection the familiar normalization of alternatives with scale factors $n^{-1/2}$ is appropriate since the Neyman–Pearson power bound for testing Q_0^n versus $Q^n_{h/\sqrt{n}}$ is still $\Phi(\|h\| - u_{1-\alpha})$ which is independent of n. Notice that $t \mapsto Q_{th}$, $t \in \mathbb{R}$, is an ordinary exponential family of distributions on the real line for fixed direction h. It is remarkable that in this case the results of Section 2 hold uniformly w.r.t. the sample size n.

THEOREM 3.1. (a) *Let $\phi_n\colon \mathbb{R}^n \to [0,1]$ be any test with $E_{Q_0^n}(\phi_n) = \alpha$. For each constant $K > 0$ the inequality*

$$\sum_{i\in I}(\sup\{|E_{Q^n_{th_i/\sqrt{n}}}(\phi_n) - \alpha|\colon |t| \le K\})^2 \le \alpha(1-\alpha)(\exp(K^2) - 1). \tag{3.4}$$

holds uniformly in n and ϕ_n. The dimension restriction of Theorem 2.1 *holds uniformly in n and ϕ_n for the rescaled family of order $n^{-1/2}$.*

(b) *For each $\alpha < \beta < 1$ and $C > 0$ the following results hold for the level points* $\mathrm{lp}(\phi_n, \beta, h)$ *of ϕ_n* (2.15) *with respect to $\{Q^n_{th}\colon t \in \mathbb{R}\}$ and the level points* $\mathrm{lp}_2(\beta, n)$ *of their upper envelope power function at sample size n. There exists a linear subspace $U \subset H$ of finite dimension with*

$$\frac{\mathrm{lp}(\phi_n, \beta, h)}{\mathrm{lp}_2(\beta, n)} \ge C \tag{3.5}$$

for all directions $h \in U^\perp \cap \{h\colon \|h\| = 1\}$. In this case the bound $k(\alpha, \beta, C)$ of the dimension of U is the same as in Lemma 2.2(b) *and it is now independent of ϕ_n and the sample size n.*

PROOF. The proof follows from Lemma 2.1 and Lemma 2.2(b). We will show that the bound (3.4) and the bound on the dimension of U can be made independent of the sample size. Notice first that $(y_1, \dots, y_n) \mapsto (T(y_1), \dots, T(y_n))$, defined on the space $C[0,1]^n \to \mathbb{R}^n$, is a sufficient statistic. Thus the new tests

$$\psi_n(y_1, \dots, y_n) = \phi_n(T(y_1), \dots, T(y_n)) \tag{3.6}$$

on $C[0,1]^n$ have the same power w.r.t. P_h^n as ϕ_n w.r.t. Q_h^n and they have the same level points. On the other hand the dimension can be reduced by the statistic S: $C[0,1]^n \to C[0,1]$,

$$S(y_1, \dots, y_n) = n^{-1/2} \sum_{i=1}^{n} y_i \tag{3.7}$$

which is sufficient for $\{P_h^n: h \in H\}$, refer to (2.1)–(2.3). Notice that if

$$B_0^{(1)}(\cdot), \dots, B_0^{(n)}(\cdot)$$

are independent Brownian bridges we have

$$n^{-1/2} \sum_{i=1}^{n} \left(B_0^{(i)}(t) + \int_0^t h(u)\, du\right) = n^{-1/2} \sum_{i=1}^{n} B_0^{(i)}(t) + n^{1/2} \int_0^t h(u)\, du \tag{3.8}$$

and consequently

$$\mathscr{L}(S | P_h^n) = P_{n^{1/2}h}. \tag{3.9}$$

(This is just the stability of the Gaussian shift G.)

By the sufficiency of S we may choose versions of the conditional expectations $\psi_0 = E_{\cdot}(\psi_n | S)$ of ψ_n which are independent of the parameter h. Hence

$$E_{P_h^n}(\psi_n) = E_{P_{n^{1/2}h}}(\psi_0) \tag{3.10}$$

follows. This equation together with Lemma 2.1 establishes our assertion (a).

(b) As further consequence of (3.10) one obtains the identity

$$\operatorname{lp}(\psi_n, \beta, h) = n^{1/2} \operatorname{lp}(\psi_0, \beta, h) \tag{3.11}$$

for the level points in direction h. On the other hand, the level points of the upper envelope functions are obviously given by

$$\operatorname{lp}_2(\beta, n) = n^{1/2} \operatorname{lp}_2(\beta, 1). \tag{3.12}$$

Now we may choose the subspace U according to Lemma 2.2(b) for the test ψ_0. For $h \in U^\perp$, $\|h\| = 1$, we see that the factor $n^{1/2}$ of (3.11) and (3.12) can be cancelled and we have

$$\frac{\operatorname{lp}(\psi_n, \beta, h)}{\operatorname{lp}_2(\beta, n)} = \frac{\operatorname{lp}(\psi_0, \beta, h)}{\operatorname{lp}_2(\beta, 1)} \geq C, \tag{3.13}$$

which implies the desired result (3.5). □

The present result gives rise to various comments and conclusions. Following the labels 1 and 2 of our introduction we will now summarize proposals and results which are cited from the literature.

1. Since the asymptotic Brownian bridge model (2.1) is much the same as the (suitably normalized) finite sample nonparametric testing problem we will restrict ourselves to the limit model (3.1). Below, let ϕ be a goodness of fit test on $C[0, 1]$. In general it is hard analytic work to get information about those finite dimensional linear subspaces which are prefered by ϕ. However, many results are known for various concrete tests.

   (a) If ϕ is an integral test of Cramér–von Mises or Anderson–Darling type, then often a global principle component decomposition of the test statistic and its power function is available; see Anderson and Darling (1952), Durbin and Knott (1972), Shorack and Wellner [(1986), Chapter 5] and see also Neuhaus (1976), Milbrodt and Strasser (1990), Drees and Milbrodt (1994).

   (b) Two-sided goodness of fit tests ϕ with centrally symmetric and convex acceptance regions have a more general structure than integral tests. Since there is no principle component decomposition of their test statistics available, Milbrodt and Strasser (1990) proposed a principle decomposition of the curvature of the power function at $h = 0$ in H. A Taylor expansion of the power function along the present exponential family is given by

$$E_{P_{th}}(\phi) = \alpha + \langle h, Th\rangle t^2/2 + o(t^2), \qquad t \in \mathbb{R},\ h \in H \tag{3.14}$$

   at $t = 0$ where T: $H \to H$ is a Hilbert–Schmidt operator with

$$T(g) = \sum_{i=1}^{\infty} \lambda_i < h_i, g > h_i\ . \tag{3.15}$$

   See Janssen (1995) for the most general result. In various cases the spectral decomposition and their eigenvalues $\lambda_i \downarrow 0$ can be derived (at least approximately or by numerical methods). In his 1990 paper Strasser obtained global extrapolations for power functions

$$t \mapsto E_{P_{th}}(\phi) \tag{3.16}$$

   of tests ϕ with centrally symmetric acceptance regions which are based on the curvature $< h, Th >$, $\|h\| = 1$. This procedure yields sharp upper bounds for (3.16) given the curvature $< h, Th >$ which are attained in the class of tests with centrally symmetric acceptance regions. Using his bounds it is easy to see that within this class of goodness of fit tests the global power function becomes flat if the curvature is small. More precisely, let $g_n \in H$ be a sequence of parameters with $\|g_n\| = 1$ and let $t_n \to t$, $t > 0$, be convergent. Then $< g_n, Tg_n > \to 0$ implies here

$$E_{P_{t_n g_n}}(\phi) \to \alpha \quad \text{as } n \to \infty. \tag{3.17}$$

   The program works, for example, for Kolmogorov–Smirnov tests. The curvature of the two-sided Kolmogorov–Smirnov test was calculated in Milbrodt and Strasser (1990) by numerical methods and it was analytically treated by Janssen (1995). Together with the global extrapolations, it is now known that the tests are roughly speaking most sensitive to deviations of the median, which has a nice practical interpretation. Its first principal component dominates all other directions and it does not pay equal attention to further (finitely many) directions as Neyman's smooth tests do by their construction. The method works for one- or two-sample testing problems. A similar interpretation is true for one-sided Kolmogorov–Smirnov tests, see Anděl (1967) and Hájek and Šidák (1967) who obtained its gradient. If one likes to apply a goodness of fit test of Kolmogorov–Smirnov type then an adjustment of principle components may be of interest. An adjustment by weight functions was proposed by Janssen and Milbrodt (1993) for survival tests. However, the argument (3.17) does not hold for arbitrary tests and it cannot be used to prove Theorem 2.1. General questions and problems concerning the extrapolation of local quantities of tests to global power functions will be considered in a forthcoming paper elsewhere.

In order to be concrete, consider again the two-sided Kolmogorov–Smirnov (KS) goodness of fit test $\phi_{\mathrm{KS}}^{(n)}$ of asymptotic level α for Example 1.1(b). Let

$$\phi_{\mathrm{KS}}^{(n)} = 1_{\{\sup_{0\le t\le 1} |n^{1/2}(\hat{F}_n(t) - t)| > c_\alpha\}} \tag{3.18}$$

denote this test at sample size n for $\boldsymbol{\lambda}_{|(0,1)}$ versus unspecified continuous alternatives, where $\hat{F}_n$ is the empirical distribution function. Under contiguous local alternatives of order $n^{-1/2}$, given by alternatives with tangent $h \in L_2^0(\boldsymbol{\lambda}_{|(0,1)})$, the asymptotic power function of $\phi_{\mathrm{KS}}^{(n)}$ is just

$$P\Big(\sup_{0\le t\le 1} |B_0(t) + \int_0^t h(u)\,du| > c_\alpha\Big), \tag{3.19}$$

see Milbrodt and Strasser (1990), page 3, for details. The curvature $< h, Th >$ of the power function is greater than zero and the power is strictly larger than the level $\alpha = P(\sup_{0\le t\le 1} |B_o(t)| > c_\alpha)$ for all nontrivial directions $h \neq 0$. This fact is labeled as $\sqrt{n}$-consistency of the sequence of tests. Since (3.19) is close to one for tangents sh with $s \in \mathbb{R}$ large enough we see that

$$\limsup_{n\to\infty} \frac{\mathrm{lp}(\phi_{KS}^{(n)}, \beta, h)}{\mathrm{lp}_2(\beta, n)} < \infty \tag{3.20}$$

holds for the level points for each $\beta < 1$ and each direction $h \neq 0$; see also Lemma 3.1 below.

2. Since every test has a preference for some finite dimensional subspace, one may be interested in the construction of tests which have good performance on a given finite dimensional linear subspace of alternatives U; see also Remark 2.1. In this case Milbrodt and Strasser (1990) proposed Neyman's smooth tests; see Neyman (1937) and the discussion below. Tests based on density estimators were established by Neuhaus (1988). One-sided non-

parametric tests for given cones of alternatives were proposed by Behnen and Neuhaus (1989). All these tests are typically admissible within full nonparametric models and no test will majorize the power functions of another tests for all directions. Results of this kind follow from the modern decision theory of Le Cam.

Data driven tests. As mentioned above, Neyman's smooth tests can be recommended for two-sided testing problems when $\mathscr{P}$ is an exponential family with k-dimensional parameter space. These tests are asymptotically minimax and likelihood ratio tests and they are therefore well motivated. The problem is now the choice of the dimension k in nonparametrics; see, for instance, Bickel and Ritov (1992) for a recent discussion of this subject. There exists a series of papers about data driven versions of Neyman's smooth tests with estimated order k; see Kallenberg and Ledwina (1995), Inglot and Ledwina (1996), Inglot, Kallenberg and Ledwina (1998) and earlier references therein. These authors show that their data driven tests are asymptotically equivalent to some Neyman–Pearson tests (given by the first direction of their sequence of orthogonal directions of alternatives) under contiguous local alternatives and that they are intermediate efficient under noncontiguous alternatives within their concept of efficiency. Moreover, the shortcoming (difference between power and the power of the most powerful tests) vanishes for intermediate alternatives even if the level α tends to zero. The convergence of the level $\alpha \to 0$ is part of the notion of intermediate efficiency of tests. According to Theorem 3.1 also each adaptive test distributes the "total amount of power" on orthogonal directions of the present alternatives. This is not a mathematical contradiction since nothing is said about noncontiguous alternatives in the preceeding sections.

In a brief discussion about data driven tests, and Kolmogorov–Smirnov tests, I will try to clarify this point again for the situation of Example 1.1(b) and $P_0 = \boldsymbol{\lambda}_{|(0,1)}$. Roughly speaking, different definitions of efficiency yield different answers. Let us start with the following concept.

(i) *Pitman efficiency* (fixed level α and local alternatives of order $n^{-1/2}$) is just the point of view of Theorem 3.1; see also Hájek and Šidák (1967). Consider a sequence ϕ_n of (perhaps data driven) tests which is asymptotically equivalent to a sequence of Neyman–Pearson tests $\phi_{\mathrm{NP}}^{(n)}$, in the sense that $\phi_n - \phi_{\mathrm{NP}}^{(n)} \to 0$ holds in $\boldsymbol{\lambda}^n_{|(0,1)}$-probability as $n \to \infty$. Let $\phi_{\mathrm{NP}}^{(n)}$ be $n^{-1/2}$ Pitman-efficient for a parametric family with tangent $h_0 \neq 0$. In contrast to (3.19) the asymptotic power function of ϕ_n is just α under all local alternatives of order $n^{-1/2}$ given by h_0-orthogonal tangents h with $\int h_0 h \, d\boldsymbol{\lambda}_{|(0,1)} = 0$, since the asymptotic relative Pitman efficiency of ϕ_n in direction h is then zero; see Hájek and Šidák (1967). Consequently, we have in comparision with (3.20),

$$\lim_{n\to\infty} \frac{\mathrm{lp}(\phi_n, \beta, h)}{\mathrm{lp}_2(\beta, n)} = \infty \tag{3.21}$$

for all $o < \alpha < \beta < 1$ and h_0-orthogonal tangents h.

Thus the KS test is asymptotically superior in the Pitman sense for these types of directions and tests of type ϕ_n cannot be supported by the present asymptotics. Also the shortcoming of KS-type tests vanishes for fixed α and all intermediate alternatives since the power function then converges to one; see Lemma 3.1 below.

(ii) *Intermediate efficiency* (level $\alpha_n \to 0$ and alternatives with $\beta < 1$). In this case, data driven Neyman's tests reach the optimum power for certain non-contiguous alternatives given by infinitely many different directions. The KS-test is only intermediate efficient (and local Bahadur efficient, respectively) for regression models in location with double exponential error variables, see Inglot and Ledwina (1996) and Nikitin (1995), Section 6.3. Note that this family is just the least favorable one parameter submodel for the median functional.

(iii) *Hodges–Lehmann efficiency* (fixed level α and fixed alternatives with $\beta = \beta_n \to 1$). Nikitin [(1995), Section 2.7,] pointed out that somewhat unexpectedly the KS-test is overall efficient.

The present results (i)–(iii) look like a paradox. First, I will comment the different results for the KS-test. The key is the level α which heavily influences the quality of the KS-test. For moderate values of α ($\alpha = 0.1$ or 0.05) it looks more like a goodness of fit test than for very small α. If α remains bounded away from zero, then everything is fine for intermediate alternatives; see also Lemma 3.1. If $\alpha \to 0$ holds, the curvature operator T of (3.15) collapses and behaves like a one-dimensional projection, see Janssen (1995). Together with the global extrapolation of Strasser (1990) this phenomenon confirms the dominating role of the first principal component of the power function and it may explain the poor intermediate and local Bahadur efficiency on a space with codimension one.

On the other hand, data driven Neyman's tests heavily rely on the assumption $\alpha_n \to 0$ which produces the intermediate efficiency. In contrast to these different results the efficiency concepts coincide for instance for linear rank tests since here the Pitman efficiency is independent of the level α; see Nikitin (1995).

What can be done in practice? I think that data driven tests are worthwhile also if α is bounded away from zero. However, it is my feeling that in the asymptotic set-up one should then use estimators $S_n\colon \Omega^n \to \mathbb{N}$ of the dimension k of Neyman's smooth test so that $\boldsymbol{\lambda}^n_{|(0,1)}(S_n = 1)$ does not converge to one under the null hypothesis. This stands in accordance to the work (mostly about two-sample testing) of Neuhaus (1988) and Behnen and Neuhaus [(1989), page 120] who proposed a bandwidth of the kernel density estimators for data driven tests which does not converge to zero. They mentioned that otherwise their theory collapses. It is my feeling that level points and the global approach provide good tools for discriminating between and comparing different procedures.

LEMMA 3.1. *Let P_t be distributions on $(0,1)$ with distribution functions F_t and $P_0 = \boldsymbol{\lambda}_{|(0,1)}$. Suppose that there exists some $x \in (0,1)$ and $K_x \neq 0$*

[*typically* $K_x = \int_0^x h(u)\,du$] *with* $F_t(x) = F_0(x) + tK_x + o(t)$ as $t \to 0$. *Let* $t_n = a_n/\sqrt{n}$ *be real sequences with* $t_n \to 0$ *and* $|a_n| \to \infty$. *Then for fixed level* α,

$$E_{P^n_{t_n}}(\phi^{(n)}_{\mathrm{KS}}) \to 1$$

holds as $n \to \infty$. *The same result is true for* $\alpha_n \to 0$ *whenever* $c_{\alpha_n}/a_n \to 0$. *Recall from Hájek and Šidák* [(1967), *page* 182] *that* $\alpha_n \sim 2\exp(-2c^2_{\alpha_n})$ *holds.*

PROOF. By the central limit theorem for arrays of binomial variables we have asymptotic normality of $\sqrt{n}(\hat{F}_n(x) - F_{t_n}(x))$ with mean zero and variance $x(1-x)$ under $P^n_{t_n}$. Thus $t_n^{-1}(\hat{F}_n(x) - F_{t_n}(x)) \to 0$ follows in probability. This implies $t_n^{-1}(\hat{F}_n(x) - x) \to K_x$ and $|\sqrt{n}(\hat{F}_n(x) - x)| \to \infty$, both in probability. □

Acknowledgments. I thank the Editor and the referees for their suggestions and comments concerning the literature.

REFERENCES

ANDĚL, J. (1967). Local asymptotic power and efficiency of tests of Kolmogorov–Smirnov type. *Ann. Math. Statist.* **38** 1705–1725.

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

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

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.

BOULEAU, N. and HIRSCH, F. (1991). *Dirichlet Forms and Analysis on Wiener Spaces*. de Gruyter, Berlin.

BURNASHEV, M. V. (1979). On the minimax detection of an inaccurately known signal in a white Gaussian noise background. *Theory Probab. Appl.* **24** 107–119.

DREES, H. and MILBRODT, H. (1991). Components of the two-sided Kolmogorov–Smirnov test in signal detection problems with Gaussian white noise. *J. Statist. Plann. Inference* **29** 325–335.

DREES, H. and MILBRODT, H. (1994). The one-sided Kolmogorov–Smirnov test in signal detection problems with Gaussian white noise. *Statist. Neerlandica* **28** 103–116.

DURBIN, J. and KNOTT, M. (1972). Components of Cramér–von Mises statistics I. *J. Roy. Statist. Soc. Ser. B* **34** 290–307.

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.

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

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

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

INGSTER, Y. I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives. I, II, III. *Math. Methods Statist.* **2** 85–114, 171–189, 249–268.

JANSSEN, A. (1995). Principal component decomposition of non-parametric tests. *Probab. Theory Related Fields* **101** 193–209.

JANSSEN, A. and MILBRODT, H. (1993). Rényi type goodness of fit tests with adjusted principal direction of alternatives. *Scand. J. Statist.* **20** 177–194.

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

LEPSKI, O. V. and SPOKOINY, V. G. (1999). Minimax nonparametric hypothesis testing: the case of an inhomogenous alternative. *Bernoulli* **5** 333–358.

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

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

NEUHAUS, G. (1988). Addendum to Local asymptotics for linear rank statistics with estimated score functions. *Ann. Statist.* **16** 1342–1343.

NEYMAN, J. (1937). "Smooth test" for goodness of fit. *Skand. Aktuarie Tidskv.* **20** 150–199.

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

PARTHASARATHY, K. R. (1967). *Probability Measures on Metric Spaces*. Academic Press, New York.

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

STRASSER, H. (1985). *Mathematical Theory of Statistics*. de Gruyter, Berlin.

STRASSER, H. (1990). Global extrapolations of local efficiency. *Statist. Decisions* **8** 11–26.

MATHEMATICAL INSTITUTE
UNIVERSITY OF DÜSSELDORF
D-40225 DÜSSELDORF
GERMANY
E-MAIL: janssena@uni-duesseldorf.de