Journal of Multivariate Analysis **74**, 1–35 (2000)
doi:10.1006/jmva.1999.1879, available online at http://www.idealibrary.com on IDEAL®

Optimal Tests for the General Two-Sample Problem

Dietmar Ferger

Department of Mathematics, Dresden University of Technology, D-01062 Dresden, Germany

Received May 4, 1998

In this paper we consider the classical problem of testing whether two samples of observations are from the same distribution. Since in many situations the data are multivariate or even of functional type, classical methodology is not applicable. In our approach we conceive a difference in distribution as the occurrence of a change-point problem, where the change-point is known in advance. This point of view enables us to construct new tests which are distribution-free under the null hypothesis for general sample spaces. The power function of the tests is studied under local and global alternatives. Finally some optimality results are provided.

AMS 1991 subject classifications: primary 62G05, 62G10, 62G20; secondary 60F05.

Key words and phrases: two-sample test, contiguous alternatives, local power, optimality, Brownian bridge with trend, functional data.

0. INTRODUCTION

Let $\xi_1, ..., \xi_m$ and $\eta_1, ..., \eta_l$ be two samples of independent random variables with distribution ν_1 and ν_2, respectively. The classical two-sample problem then consists of testing the hypothesis $H_0 : \nu_1 = \nu_2$ versus the alternative $H_1 : \nu_1 \neq \nu_2$, where ν_1 and ν_2 are unknown. For real-valued observations many procedures are available and discussed in detail in monographs on nonparametric inference. To mention only a few we refer to Hájek and Šidák (1967), Hollander and Wolfe (1973), and Lehmann (1975). A common feature of most tests is that for real data the underlying test statistic is invariant under continuous monotone transformations, resulting in tests which are distribution-free under $H_{o,c} : \nu_1 = \nu_2$ continuous. These tests include the Kolmogorov–Smirnov (KS) and Cramér–von Mises (CvM) test as well as the broad class of tests based on linear rank statistics. The situation becomes completely different for multivariate rather than univariate data. Since there is no natural order on $\mathbb{R}^d$ for $d \geqslant 2$ it is not possible to carry over, for example, rank-procedures to the present setting. This may explain why the multivariate case has been treated in far less detail. For recent work on the multivariate two-sample problem see, for

0047-259X/00 $35.00

example, Ahmad and Cerrito (1993), Bahr (1996), Einmahl and Khmaladze (1998), Epps and Singleton (1986), Henze (1988), and Henze and Voigt (1992). A review on multivariate rank like tests is given in Hettmansperger and McKean (1998).

In many practical situations the observations are not only finite dimensional vectors but also functions. For instance longitudinal studies in continuous time provide such data. The recent interesting monograph by Ramsay and Silverman (1997) reports on many statistical experiments in various scientific fields where functional data appear. Among others the measurements may be the force exerted on a meter during a brief pinch by the thumb and forefinger, the angles formed by the hip and knee over a child's gait cycle, or the temperatures over one year for several Canadian weather stations.

Although there has been an increasing interest in functional data, no methodology for testing $\nu_1 = \nu_2$ in this general context seems to exist. In the present paper we propose and study in detail tests for equality in distribution when the sample space, $\mathcal{X}$, is arbitrary. In particular, our results apply to multivariate observations and functional data. For real-valued observations, we shall also compare the power of the new tests with that of the tests known from the literature.

In what follows let $\xi_1, ..., \xi_m, \eta_1, ..., \eta_l$ be independent random variables with values in a sample space $\mathcal{X}$. Consider the pooled sample

$$X_i := \begin{cases} \xi_i, & 1 \leqslant i \leqslant m \\ \eta_{i-m}, & m < i \leqslant n \end{cases}, \tag{0.1}$$

where $n = m + l$. If the alternative $H_1: \nu_1 \neq \nu_2$ holds, then

$$\mathcal{L}(X_1) = \cdots = \mathcal{L}(X_m) \neq \mathcal{L}(X_{m+1}) = \cdots = \mathcal{L}(X_n), \tag{0.2}$$

where $\mathcal{L}(X)$ denotes the distribution of a random variable X. In other words, (0.2) means that the distribution of the X_i changes from ν_1 to ν_2 at $i = m$. Thus the two-sample model is closely related to the so-called *change-point model*. Here one observes independent $\mathcal{X}$-valued random variables, $Y_1, ..., Y_n$ say, such that for some $\theta \in [0, 1]$

$$\mathcal{L}(Y_1) = \cdots = \mathcal{L}(Y_{[n\theta]}) \neq \mathcal{L}(Y_{[n\theta]+1}) = \cdots = \mathcal{L}(Y_n), \tag{0.3}$$

with $[a]$ denoting the integer part of $a \in [0, \infty)$. The parameter θ is called the *change-point* of the sequence (Y_i). Clearly, $\theta \in \{0, 1\}$ corresponds to $Y_1, ..., Y_n$ which are i.i.d. Usually, θ is the unknown parameter of interest and the problem, for example, may be one of estimating θ. Obviously the pooled sample (0.1) may be viewed as a sequence with change-point $\theta = \frac{m}{n}$ provided the alternative $H_1 : \nu_1 \neq \nu_2$ holds, while the hypothesis $H_0 : \nu_1 = \nu_2$

corresponds to $\theta \in \{0, 1\}$. The intrinsic difference between (0.2) and (0.3) lies in the fact that in (0.2) the possible position m of a change is precisely known in advance, while in (0.3) the most important question is to determine the unknown point $[n\theta]$.

In recent years many estimators for θ have been proposed and proved to be consistent as long as $0 < \theta < 1$. See, for example, Hinkley (1970), Bhattacharya and Brockwell (1976), Darkhovski (1976, 1994), Carlstein (1988), Dümbgen (1991), and Yao *et al.* (1994). In addition to the case of an actual change ($0 < \theta < 1$) Gombay and Horváth (1993), Lombard and Hart (1994), Ferger (1995), and Hušková (1996) also studied their estimators under the hypothesis of no change ($\theta \in \{0, 1\}$) and showed convergence in distribution to a nondegenerate limit variable.

Now, let θ_n denote a change-point estimator, which is consistent in the case of $0 < \theta < 1$ and convergent in law if $\theta \in \{0, 1\}$. The idea of our approach is to apply θ_n to the pooled sample (X_i). As we have already mentioned, under the alternative H_1, the sequence (X_i) has a *known* change-point $\theta = \frac{m}{n}$. Therefore the consistency of θ_n entails small values of the distance $|\theta_n - \frac{m}{n}|$ under H_1, whereas under the hypothesis H_0 it should be large. Thus the following decision rule suggests itself: *Reject the hypothesis H_0, if the distance $|\theta_n - \frac{m}{n}|$ between the change-point estimator θ_n pertaining to the pooled sample (X_i) and the possible change-point $\theta = \frac{m}{n}$ is too small.*

This paper is organized as follows. In Section 1 we present a precise formulation of our tests. Section 2 is devoted to a comprehensive study of the power. Our test statistics depend on a score function (called kernel), which has to be specified by the statistician in advance. The first part of Section 3 shows how one has to determine suitable kernels when the alternative to H_0 is specified. Two general methods of construction are presented and applied to many examples. In the second part of this section we introduce our notion of optimality. It uses the concept of maximizing the local power of the test. The optimal kernel is the minimizer of a certain functional. It is explicitly known and surprisingly simple. This enables us to determine it in many examples. In Section 4 special emphasis is given to functional data. In Section 5, a simulation study is reported on which confirms our theoretical findings. Finally, the proofs are given in the final section.

1. A NEW TWO-SAMPLE TEST

To properly formulate our results we let the subsample size $m = m_n$ and $l = l_n$ tend to infinity as n tends to infinity. In addition it is assumed that the fraction $\bar{\theta}_n = m_n n^{-1}$ of the first subsample converges to a limit fraction

$\theta \in (0, 1)$. More precisely let $(\xi_{in}: 1 \leqslant i \leqslant m_n, n \in \mathbb{N})$ and $(\eta_{jn}: 1 \leqslant j \leqslant l_n, n \in \mathbb{N})$ be two double-indexed schemes such that for all $n \in \mathbb{N}$,

$$\xi_{1n}, ..., \xi_{m_n n}, \eta_{1n}, ..., \eta_{l_n n}$$

are independent $\mathcal{X}$-valued random variables on a common probability space $(\Omega, \mathcal{A}, P)$ with distributions

$$\mathcal{L}(\xi_{in}) = P \circ \xi_{in}^{-1} = \nu_1, \qquad 1 \leqslant i \leqslant m_n$$

and

$$\mathcal{L}(\eta_{jn}) = P \circ \eta_{jn}^{-1} = \nu_2, \qquad 1 \leqslant j \leqslant l_n,$$

where ν_1 and ν_2 are unknown. Recall that for each $n \in \mathbb{N}$ the pooled sample

$$X_{in} = \begin{cases} \xi_{in}, & 1 \leqslant i \leqslant m_n \\ \eta_{i-m_n, n}, & m_n < i \leqslant n \end{cases}$$

is a sequence with change-point $\bar{\theta}_n = m_n/n$ provided H_1 holds. The test we propose requires a suitable change-point estimator θ_n. Ferger (1995) suggests

$$\theta_n = \frac{1}{n} \operatorname*{argmax}_{1 \leqslant k \leqslant n-1} w\left(\frac{k}{n}\right) \left| \sum_{i=k+1}^{n} \sum_{j=1}^{k} K(X_{in}, X_{jn}) \right| \tag{1.1}$$

with a one-sided version

$$\theta_n^+ = \frac{1}{n} \operatorname*{argmax}_{1 \leqslant k \leqslant n-1} w\left(\frac{k}{n}\right) \sum_{i=k+1}^{n} \sum_{j=1}^{k} K(X_{in}, X_{jn}).$$

Here the so-called kernel

$$K: \mathcal{X}^2 \to \mathbb{R}$$

is assumed to be antisymmetric, that is $K(x, y) = -K(y, x)$ for all $x, y \in \mathcal{X}$. The weight-function $w: (0, 1) \to (0, \infty)$ is assumed to be of the type

$$w(t) = t^{-a}(1-t)^{-b}, \qquad 0 < t < 1, \qquad 0 \leqslant a, b < 1/2. \tag{1.2}$$

The weight-function $w = 1$ plays a particular role. We will see that in this case it is advantageous to work with the following modification of θ_n^+:

$$\tilde{\theta}_n^+ = \frac{1}{n} \operatorname*{argmax}_{0 \leqslant k \leqslant n-1} \sum_{i=k+1}^{n} \sum_{j=1}^{k} K(X_{in}, X_{jn}). \tag{1.3}$$

The estimator $\theta_n^{(+)}$ for general weights w has been studied in detail, under a second-moment condition in K, in Ferger (1995), where among other things it was shown that

$$\theta_n^{(+)} - \theta = O(n^{-1}) \qquad \text{almost surely, if} \quad 0 < \theta < 1. \tag{1.4}$$

Here and in the rest of the paper the superscripts in brackets can be used optionally.

For the change-point problem weight-functions have been introduced to make the estimator more sensitive if θ is close to zero or one. Similarly for the two-sample problem weight-functions are recommended to compensate a loss of power, which usually occurs, if one sample is much larger than the other one. Even more important is the choice of the kernel K. This will be discussed in Section 3.

As pointed out in the Introduction, the consistency (1.4) of (θ_n) motivates the following two-sample test:

$$\Phi_n^{(+)} = 1_{\{|\theta_n^{(+)} - m/n| \leqslant c^{(+)}\}}. \tag{1.5}$$

In order to determine the critical value $c^{(+)}$ appropriately we make use of Theorem 3.1 of Ferger (1995), which states that under H_0,

$$\theta_n^{(+)} \xrightarrow{\mathscr{L}} \tau_w^{(+)}, \qquad n \to \infty. \tag{1.6}$$

Here

$$\tau_w = \operatorname*{argmax}_{0<t<1} w(t)\,|B_0(t)| \qquad \text{and} \qquad \tau_w^+ = \operatorname*{argmax}_{0<t<1} w(t)\,B_0(t),$$

where B_0 is a Brownian Bridge. Note that by Corollary 1.2, p. 189, of Csörgő and Horváth (1993) the random variables τ_w and τ_w^+ are well-defined. Let $\alpha \in (0, 1)$ denote the given significance level. Because of (1.6),

$$c^{(+)} = c_\alpha^{(+)} \qquad \text{is the } \alpha\text{-quantile of} \qquad |\tau_w^{(+)} - \theta|. \tag{1.7}$$

The values of $c_\alpha^{(+)}$ depending on w and α are easy to obtain by a Monte Carlo approximation.

When the subsample sizes m and l are of the same order of magnitude, weight-functions are not needed. That is, one can use $w = 1$. In this case the distribution functions of τ_1 and τ_1^+ are analytically known. See Proposition 3.2 of Ferger (1995). Especially $\mathscr{L}(\tau_1^+)$ is the uniform distribution on the unit interval $(0, 1)$, so that by (1.6), under H_0, the sequence (θ_n^+) asymptotically is uniformly distributed on $(0, 1)$. Clearly this is also true

for the minor modification $\tilde{\theta}_n^+$. But surprisingly even the finite sample distribution of $\tilde{\theta}_n^+$ under H_0 is known, namely

$$P_{H_0}(n\tilde{\theta}_n^+ = k) = \frac{1}{n} \qquad \text{for all integers} \qquad 0 \leqslant k \leqslant n-1. \tag{1.8}$$

In view of (1.8), we also introduce the exact test

$$\tilde{\Phi}_n^+ = 1_{\{|\tilde{\theta}_n^+ - m/n| \leqslant c_{m,l}\}}$$

with critical value

$$c_{m,l} = \begin{cases} 0, & 0 < \alpha < n^{-1} \\ [\tfrac{1}{2}(n\alpha - 1)]\, n^{-1}, & n^{-1} \leqslant \alpha \leqslant (2l-1)\, n^{-1}, \\ [n\alpha - l]\, n^{-1}, & (2l-1)\, n^{-1} < \alpha < 1 \end{cases} \tag{1.9}$$

where w.l.o.g. $m \geqslant l$.

2. POWER INVESTIGATIONS

Our first result ensures that Φ_n and Φ_n^+ are asymptotic level-α tests.

PROPOSITION 2.1. *Suppose* $v_1 = v_2 = v$ *and* K *is a kernel such that for some* $p > 2$,

$$\int |K|^p \, dv \otimes v < \infty \tag{2.1}$$

and

$$\sigma^2 = \int \left[\int K(x, y)\, v(dy) \right]^2 v(dx) > 0. \tag{2.2}$$

Then we have

$$\lim_{n \to \infty} P(\Phi_n^{(+)} \textit{ rejects } H_0) = \alpha.$$

Note that (2.1) is only a weak condition. Especially it is fulfilled for all bounded kernels K no matter how the distribution v looks like. As to the second condition (2.2) observe that most frequently K is of the type $K(x, y) = a(x) - a(y)$ with some mapping $a: \mathcal{X} \to \mathbb{R}$. Then (2.2) excludes the degenerate case that $K(X_i, X_j) = 0$ for all $1 \leqslant i, j \leqslant n$ P-almost surely.

For the one-sided test $\tilde{\Phi}_n^+$ we are actually able to give the exact probability of the Type I error when the sample sizes m and l are finite.

PROPOSITION 2.2. *Let $m \geqslant l$. Suppose $\nu_1 = \nu_2$ and K is a kernel such that*

$$P\left(\sum_{i=k+1}^{n} \sum_{j=1}^{k} K(X_i, X_j) = 0 \right) = 0 \qquad \text{for all} \quad 1 \leqslant k \leqslant n-1. \tag{2.3}$$

Then

$$P(\tilde{\Phi}_n^+ \text{ rejects } H_0) = \begin{cases} n^{-1}, & \alpha < \dfrac{1}{n} \\ \left(2 \left[\dfrac{1}{2}(n\alpha - 1) \right] + 1 \right) n^{-1}, & n^{-1} \leqslant \alpha \leqslant (2l-1)\, n^{-1}. \\ ([n\alpha - 1] + l)\, n^{-1}, & \alpha > (2l-1)\, n^{-1} \end{cases}$$

We see that in Proposition 2.2 the validity of the moment conditions (2.1) and (2.2) is not required. Condition (2.3) obviously is satisfied for all $n \in \mathbb{N}$ when $K(x, y) = a(x) - a(y)$ and $a(X_1)$ is a continuous random variable.

Next we investigate the behavior of our tests under the alternative. A sufficient condition for consistency is given in the following proposition.

PROPOSITION 2.3. *Suppose $\nu_1 \neq \nu_2$ with*

$$\int K^2 \, d\nu_1 \otimes \nu_2 < \infty. \tag{2.4}$$

By (2.4) *we can define*

$$\lambda = \lambda(K) = \int K \, d\nu_2 \otimes \nu_1.$$

If $\lambda \neq 0$, then

$$\lim_{n \to \infty} P(\Phi_n \text{ rejects } H_0) = 1. \tag{2.5}$$

If $\lambda > 0$, then

$$\lim_{n \to \infty} P(\Phi_n^+ \text{ rejects } H_0) = 1. \tag{2.6}$$

Under an additional (and very regular) assumption on $v_1 \neq v_2$ the condition $\lambda \neq 0$ or $\lambda > 0$, respectively, is not only sufficient but also necessary for the consistency of (Φ_n) or (Φ_n^+), respectively.

PROPOSITION 2.4. *Let $w = 1$ and suppose $v_1 \neq v_2$ with (2.4) also meets the requirement*

$$\int \left[\int K(x, y)\, v_1(dy) \right] \left[\int K(x, y)\, v_2(dy) \right] v_j(dy) = \tau^2 \in (0, \infty) \qquad \text{for} \quad j = 1, 2. \tag{2.7}$$

Then the sequence of tests (Φ_n) is consistent if and only if $\lambda \neq 0$:

$$\lim_{n \to \infty} P(\Phi_n \text{ rejects } H_0) = 1 \Leftrightarrow \lambda \neq 0. \tag{2.8}$$

Similarly, if $\frac{\alpha}{2} < \theta < 1 - \frac{\alpha}{2}$, then the sequence of one-sided tests (Φ_n^+) is consistent if and only if $\lambda > 0$:

$$\lim_{n \to \infty} P(\Phi_n^+ \text{ rejects } H_0) = 1 \Leftrightarrow \lambda > 0. \tag{2.9}$$

Since (Φ_n^+) and $(\tilde{\Phi}_n^+)$ are asymptotically not distinguishable the equivalence in (2.9) also holds for $(\tilde{\Phi}_n^+)$. Propositions 2.3 and 2.4 reveal the importance of the quantity λ. The part of the alternative H_1 where (Φ_n) and (Φ_n^+) are consistent is uniquely determined through the sign of λ. The interesting question of what happens on the rest of the alternative remains. More precisely, how do the tests behave in case of $\lambda = 0$? The answer to this question is given in the following proposition.

PROPOSITION 2.5. *Let $w = 1$ and suppose $v_1 \neq v_2$ with (2.4) and (2.7) are such that $\lambda = 0$. Then for all $\alpha \in (0, 1)$*

$$\lim_{n \to \infty} P(\Phi_n^{(+)} \text{ rejects } H_0) = \alpha.$$

Note that for kernels K of the type $K(x, y) = a(x) - a(y)$ the pertaining λ is equal to the difference $E(a(\eta_1)) - E(a(\xi_1))$ of the first moments. Thus, for example, $\lambda = 0$ means that the first moments coincide. Similarly, condition (2.7) is satisfied if $\mathrm{Var}(a(\xi_1)) = \mathrm{Var}(a(\eta_1))$.

By construction the test $\Phi_n^{(+)} = \Phi_n^{(+)}(K)$ depends on the kernel K. According to Propositions 2.3–2.5 the statistician has to choose K such that the corresponding $\lambda(K)$ does not vanish (or is strictly positive for the one-sided test). Looking at λ and K in this way we see that a kernel K with $\lambda(K) \neq 0$

($\lambda(K) > 0$) is able to distinguish between the underlying distributions ν_1 and ν_2. In Section 3 it is shown that to each pair $\nu_1 \neq \nu_2$ one can construct explicitly kernels K with $\lambda(K) \neq 0$. On the other hand, if $\nu_1 \neq \nu_2$ but $\lambda(K) = 0$ this means that K is unable to detect the difference between ν_1 and ν_2. Indeed, from Proposition 2.5 we know that in this case the test behaves as if ν_1 and ν_2 were equal. Finally, if actually $\nu_1 = \nu_2$ then $\lambda(K) = 0$ for each kernel K. This is a consequence of Fubini's theorem and the antisymmetry of K. Of course for K fixed $\lambda = \lambda(K, \nu_1, \nu_2)$ can be interpreted as a distance between ν_1 and ν_2. However, as far as consistency is concerned, it is not the absolute value of λ which is decisive but merely the sign of λ. This also becomes clear when observing that $\lambda(\alpha K) = \alpha\lambda(K)$ for all $\alpha \in \mathbb{R}$ but

$$\Phi_n^{(+)}(\alpha K) = \Phi_n^{(+)}(K) \qquad \text{for all} \quad \alpha \neq 0 (\alpha > 0). \tag{2.10}$$

Thus one can arbitrarily enlarge $|\lambda| = |\lambda(K)|$ by replacing K through αK but the test always remains the same. The following result, where rates of convergence for the probability of the Type II error are established, sheds more light on the role of λ.

PROPOSITION 2.6. *Assume $\nu_1 \neq \nu_2$ and K is a kernel with $\lambda \neq 0$ ($\lambda > 0$) and*

$$M_p = \int |K|^p \, d\nu_1 \otimes \nu_2 < \infty \qquad \text{for some} \quad p \geqslant 2.$$

Then an upper bound for the probability of the Type II error is given by

$$P(\Phi_n^{(+)} \text{ does not reject } H_0) \leqslant C_p M_p \, |\lambda|^{-p} \begin{cases} n^{-1} \log n, & \text{if} \quad p = 2 \\ n^{-p/2}, & \text{if} \quad p > 2 \end{cases}, \tag{2.11}$$

where $C_p = C_p(w, \theta)$ is a positive constant. If K is bounded ($p = \infty$), then we obtain an exponential rate

$$P(\Phi_n^{(+)} \text{ does not reject } H_0) \leqslant A_0 n \exp(-A_1 \lambda^2 n),$$

where A_0 is an absolute positive constant and $A_1 = A_1(w, \theta, \|K\|) > 0$ with $\|K\|$ denoting the sup-norm of K.

Notice that by (2.11) the power of $\Phi_n^{(+)}$ is controlled by the ratio

$$\Delta_p = \Delta_p(K) = \frac{|\lambda|^p}{M_p} = \frac{|\int K \, d\nu_1 \otimes \nu_2|^p}{\int |K|^p \, d\nu_1 \otimes \nu_2}$$

rather than by λ alone. Jensen's inequality ensures that $0 < \Delta_p \leqslant 1$. Moreover Δ_p is invariant w.r.t. scalar multiplication, i.e., $\Delta(\alpha K) = \Delta(K)$, which is not surprising in view of (2.10).

So far our test has been studied only under fixed alternatives $\nu_1 \neq \nu_2$. Next we consider local alternatives which approach the hypothesis in some specified sense. Two types of local alternatives are distinguished. As to the first we assume that

$$\begin{aligned} \mathscr{L}(\xi_{in}) &= P \circ \xi_{in}^{-1} = \nu_{1n}, \qquad 1 \leqslant i \leqslant m_n \\ \mathscr{L}(\eta_{jn}) &= P \circ \eta_{jn}^{-1} = \nu_{2n}, \qquad 1 \leqslant j \leqslant l_n, \end{aligned} \tag{2.12}$$

where ν_{1n} and ν_{2n} are probability distributions with the corresponding "distances"

$$\lambda_n = \int K \, d\nu_{2n} \otimes \nu_{1n}$$

converging to zero such that

$$\lambda_n n^{1/2} \to \gamma \in \bar{\mathbb{R}} \qquad \text{as} \qquad n \to \infty. \tag{2.13}$$

Note that the quantity γ depends on the kernel K and on the underlying local alternatives $\nu^{(1)} = (\nu_{1n})_{n \geqslant 1}$ and $\nu^{(2)} = (\nu_{2n})_{n \geqslant 1}$:

$$\gamma = \gamma(K, \nu^{(1)}, \nu^{(2)}).$$

For the sake of illustration consider the following example.

EXAMPLE 2.7 (Location-Model). Let $\mathscr{X} = \mathbb{R}$ and suppose that ν_{1n} and ν_{2n} have Lebesgue-densities f and $f(\cdot - d_n)$, respectively, where f is a smooth density on $\mathbb{R}$ and $d_n = rn^{-1/2}$, $r \neq 0$. Then $\gamma = r \int \int K(x, y) f(x) f'(y) \, dx \, dy$. For example, if $K(x, y) = x - y$ then we obtain $\gamma = r$.

The limit value $\lim_{n \to \infty} P(\Phi_n^{(+)} = 1)$ of the power function of $\Phi_n^{(+)}$ in the case of local alternatives (2.12) satisfying (2.13) is called *local power* of $\Phi_n^{(+)}$. In the next propositions we investigate the local power of $\Phi_n^{(+)}$ in dependence on the quantity γ.

PROPOSITION 2.8. *Let K be a kernel such that* (2.12) *and* (2.13) *hold. Assume also that*

$$m_p = \sup_{n \geqslant 1} \int |K|^p \, d\nu_{1n} \otimes \nu_{2n} < \infty, \qquad p > 2.$$

If $|\gamma| = \infty$ $(\gamma = \infty)$, *then*

$$\lim_{n \to \infty} P(\Phi_n^{(+)} \text{ rejects } H_0) = 1. \tag{2.14}$$

If $0 < |\gamma| < \infty$ $(0 < \gamma < \infty)$, *then*

$$\liminf_{n \to \infty} P(\Phi_n^{(+)} \text{ rejects } H_0) \geqslant 1 - C_p m_p |\gamma|^{-p} \tag{2.15}$$

with C_p *as in Proposition* 2.6.

In order to determine the exact local power of $\Phi_n^{(+)}$ we need the following assumption, which is the asymptotic counterpart of (2.7),

$$\int \left[\int K(x, y)\, v_{1,n}(dy) \right] \left[\int K(x, y)\, v_{2n}(dy) \right] v_{jn}(dx) \to \tau^2 \in (0, \infty) \qquad \text{for} \quad j = 1, 2 \tag{2.16}$$

as n tends to infinity. Similarly to (2.7), in the case of $K(x, y) = a(x) - a(y)$, condition (2.16) implies that the variances $\mathrm{Var}(a(\xi_{1n}))$ and $\mathrm{Var}(a(\eta_{1n}))$ are asympotically equal and not degenerated. Observe that τ^2 does not depend on j, but on K, $v^{(1)}$ and $v^{(2)}$ as it is the case for γ:

$$\tau^2 = \tau^2(K, v^{(1)}, v^{(2)}).$$

EXAMPLE 2.9. Let the sample space $\mathcal{X}$ be endowed with a topology and the induced Borel-σ-algebra. If K is bounded, then by the Portmanteau theorem the convergence of (v_{1n}) and (v_{2n}) in the weak topology to a common limit v ensures the validity of (2.16) with

$$\tau^2 = \int \left[\int K(x, y)\, v(dy) \right]^2 v(dx).$$

In the next proposition we compute the exact local power of $\Phi_n^{(+)}$. It involves the maximizer of a weighted Brownian bridge with a certain trend function.

PROPOSITION 2.10. *Suppose that the conditions of Proposition* 2.8 *are satisfied and that* (2.16) *holds. Additionally let* w *be truncated, that is,* w *is constant outside an interval* $[\delta_0, \delta_1]$ *for some arbitrary small numbers* $\delta_0, \delta_1 \in (0, \frac{1}{2})$. *Then for all* $\gamma \in \mathbb{R}$ $(\gamma \geqslant 0)$,

$$\lim_{n \to \infty} P(\Phi_n^{(+)} \text{ rejects } H_0) = P(|T^{(+)} - \theta| \leqslant c^{(+)}), \tag{2.17}$$

where

$$T = \operatorname*{argmax}_{0<t<1} w(t)\,|\tau B_0(t) + \gamma\,\Delta(t)|,$$
$$T^+ = \operatorname*{argmax}_{0<t<1} w(t)(\tau B_0(t) + \gamma\,\Delta(t))$$

and

$$\Delta(t) = \begin{cases} (1-\theta)\,t, & 0 < t \leqslant \theta \\ \theta(1-t), & \theta < t < 1 \end{cases}.$$

Especially, if $\gamma = 0$ *then*

$$\lim_{n\to\infty} P(\Phi_n^{(+)} \textit{ rejects } H_0) = \alpha. \tag{2.18}$$

The local power (2.17) is easy to approximate by a Monte Carlo simulation. From (2.14)–(2.18) we can infer that $n^{-1/2}$ is the exact rate at which the alternatives can approach the hypothesis in order to obtain a non-degenerate local power. What happens if the rate is slower or faster than $n^{-1/2}$ is reflected in (2.14) and (2.18). The formula (2.17) will serve as a means to characterize optimal kernels. See the next section.

As already mentioned there is a second type of contiguous alternatives. Here we consider real-valued observations. Assume (2.12) holds with ν_{1n} and ν_{2n} having distribution functions F_n and F, respectively, such that F_n is absolutely continuous with respect to F. The densities dF_n/dF are defined by

$$\frac{dF_n}{dF}(F^{-1}(u)) = 1 + \frac{1}{\sqrt{n}}\,g_n(u), \qquad 0 < u < 1, \tag{2.19}$$

where $g_n \in L_2[0,1]$, $\int g_n(u)\,du = 0$ and for some $g \in L_2[0,1]$

$$g_n \to g, \qquad n \to \infty, \qquad \text{almost everywhere.}$$

For F continuous and strictly monotone (2.19) equivalently can be rewritten as

$$F_n(x) = F(x) + \frac{1}{\sqrt{n}} \int_0^{F(x)} g_n(u)\,du, \qquad x \in \mathbb{R}.$$

If the Lebesgue density f of F exists, then F_n has a Lebesgue density f_n as well, which is given by

$$f_n(x) = f(x) + \frac{1}{\sqrt{n}}\,g_n(F(x))\,f(x), \qquad x \in \mathbb{R}. \tag{2.20}$$

Thus (2.19) is equivalent to a one-term expansion of order $n^{-1/2}$ of $F_n(x)$ or $f_n(x)$ at $F(x)$ or $f(x)$, respectively. The counterpart of Proposition 2.10 is provided by the following result.

PROPOSITION 2.11. *Let $\mathcal{X} = \mathbb{R}$ and let w be a weight-function of the type* (1.2). *If K is a kernel with*

$$\int |K|^p \, dF \otimes F < \infty, \qquad p > 2$$

and

$$\tau^* = \int \left[\int K(x, y) \, F(dy) \right]^2 F(dx) > 0,$$

then under (2.19),

$$\lim_{n \to \infty} P(\Phi_n^{(+)} \text{ rejects } H_0) = P(|U^{(+)} - \theta| \leqslant c^{(+)}), \tag{2.21}$$

where

$$U = \operatorname*{argmax}_{0 < t < 1} w(t) \, |\tau^* B_0(t) + \gamma^* \, \Delta(t)|,$$

$$U^+ = \operatorname*{argmax}_{0 < t < 1} w(t)(\tau^* B_0(t) + \gamma^* \, \Delta(t)),$$

and

$$\gamma^* = -\int_0^1 g(u) \int K(x, F^{-1}(u)) \, F(dx) \, du.$$

EXAMPLE 2.7 (Continuation). Clearly the location-model can also be described by means of (2.19). Using (2.20) and the Mean Value Theorem we obtain

$$g(u) = -r \, \frac{f'(F^{-1}(u))}{f(F^{-1}(u))}, \qquad u \in (0, 1),$$

and thus

$$\gamma^* = r \int \int K(x, u) \, f(x) \, f'(u) \, dx \, du = \gamma.$$

Consequently the random variables $T^{(+)}$ and $U^{(+)}$ coincide, so that Proposition 2.11 is consistent with its counterpart Proposition 2.10.

3. CONSTRUCTION AND OPTIMALITY OF KERNELS

Recall that the test $\Phi_n^{(+)} = \Phi_n^{(+)}(K)$ involves an antisymmetric kernel K which has to satisfy the *separation property*

$$\lambda = \lambda(K) = \int K \, dv_2 \otimes v_1 \neq 0. \tag{3.1}$$

For the one-sided test $\Phi_n^+(K)$ one has to require

$$\lambda(K) > 0. \tag{3.2}$$

This leads to the following questions: Is it always possible to find kernels K with the separation property? What do these kernels look like? Do optimal kernels exist and how is the notion of optimality to be understood?

As to the second question many examples are given in Ferger (1994b, 1995), if some additional information about the underlying distribution is available. The next proposition gives a complete answer to the first and the second questions.

PROPOSITION 3.1. *Let $v_1 \neq v_2$ be dominated by some σ-finite measure μ with Radon–Nikodym derivatives $f_i = dv_i/d\mu$, $i = 1, 2$.*

(1) *Put for some fixed $p \in (0, 1)$*

$$h = \frac{f_2 - f_1}{pf_1 + (1-p) f_2}$$

and define

$$K(x, y) = h(x) - h(y), \qquad x, y \in \mathcal{X}. \tag{3.3}$$

(2) *Assume that $\{f_1 = 0\} = \{f_2 = 0\}$ and that the Kulback–Leibler Informationnumbers of v_1 and v_2 are finite. Set*

$$K(x, y) = l(x) - l(y), \qquad x, y \in \mathcal{X}. \tag{3.4}$$

with $l(x) = \log(f_2(x)/f_1(x))$.

Then the kernel K in both (3.3) *and* (3.4) *is $v_1 \otimes v_2$-integrable with $\lambda(K) > 0$. Moreover K is invariant with respect to μ, that is passing to another dominating σ-finite measure does not change K.*

Observe that $\mu := v_1 + v_2$ dominates v_1 and v_2 and is σ-finite. Thus Proposition 3.1 simultaneously provides the existence and the construction of an appropriate kernel, which not only has the separation property (3.1)

but even meets the stronger requirement (3.2), so that the one-sided test $\Phi_n^+(K)$ can be used. As to the parameter p in (3.3) we recommend $p=\theta$ since the pertaining change-point estimator $\theta_n=\theta_n(K)$ has the minimal asymptotic mean-squared error among all admissible kernels (cf. Ferger, 1997).

Of course the kernels in Proposition 3.1, depending on the densities f_1 and f_2, are typically not known to the statistician. In a complete nonparametric framework one has to replace f_1 and f_2 by their estimators, for which a rich literature is available. In parametric models very often the estimation even can be dropped.

EXAMPLE 3.2 (Exponential Family). Suppose

$$f_i(x)=\exp\left\{\sum_{j=1}^{k} Q_j(\gamma_i)\, T_j(x)+Q_0(\gamma_i)+T_0(x)\right\}, \qquad x\in\mathscr{X}, \qquad i=1,2,$$

where $\gamma_1\neq\gamma_2$ belong to some parameter space Γ and $Q_j\colon \Gamma\to\mathbb{R}$ and $T_j\colon \mathscr{X}\to\mathbb{R}$ are certain mappings. Then Proposition 3.1 (2) yields

$$K(x,\, y)=\sum_{j=1}^{k} (Q_j(\gamma_2)-Q_j(\gamma_1))(T_j(x)-T_j(y)).$$

Example 3.2 is the source of many other examples; cf. Ferger (1998). Next, we treat the third question concerning the *optimal choice of a kernel*. First the notion of optimality has to be made precise. Here we use the concept of maximizing the local power (2.17) and (2.21), respectively. As to the first, in the nondegenerate case $0<|\gamma|<\infty$, the basic random variable T can be rewritten as

$$T=\operatorname*{argmax}_{0<t<1} w(t)\left|\frac{\tau}{\gamma}B_0(t)+\Delta(t)\right|.$$

One can prove (see (4.24) in Ferger, 1995) that if the ratio τ/γ tends to zero then $T\to\theta$ almost surely and therefore $P(|T-\theta|\leqslant c)\to 1$. This means, the smaller the ratio $|\tau/\gamma|$, the larger is the corresponding local power (2.17). Recall that $\tau^2=\tau^2(K, v^{(1)}, v^{(2)})$ and $\gamma=\gamma(K, v^{(1)}, v^{(2)})$ depend on the kernel K and on the underlying local alternatives $v^{(1)}=(v_{1n})$ and $v^{(2)}=(v_{2n})$. This motivates the following definition.

DEFINITION 3.3 (Optimality). Let $v^{(1)}=(v_{1n})$ and $v^{(2)}=(v_{2n})$ be local alternatives and assume the set

$$\mathscr{K}=\mathscr{K}(v^{(1)}, v^{(2)})=\{K\neq 0 : 0<\tau^2(K, v^{(1)}, v^{(2)})<\infty,\, 0\neq\gamma(K, v^{(1)}, v^{(2)})\in\mathbb{R}\}$$

of *admissible* kernels is nonempty. Then

$$K_{\text{opt}} = K_{\text{opt}}(v^{(1)}, v^{2)}) = \operatorname{argmin} \left\{ \frac{\tau^2(K, v^{(1)}, v^{(2)})}{\gamma^2(K, v^{(1)}, v^{(2)})} : K \in \mathcal{K}(v^{(1)}, v^{(2)}) \right\}$$

is called and *optimal kernel* with respect to the local alternatives $v^{(1)}$ and $v^{(2)}$ (provided the minimizer exists).

Note that the ratio

$$\rho(K) = \rho(K, v^{(1)}, v^{(2)}) = \frac{\tau^2(K, v^{(1)}, v^{(2)})}{\gamma^2(K, v^{(1)}, v^{(2)})}$$

satisfies

$$\rho(\alpha K) = \rho(K) \qquad \text{for all} \quad \alpha \neq 0, \tag{3.5}$$

whence αK_{opt} is also an optimal kernel for all $\alpha \neq 0$. This is in accordance with property (2.10). The determiniation of $K_{\text{opt}} = \operatorname{argmin}\{\rho(K) : K \in \mathcal{K}\}$ is possible in a very general setting. Let $v_{in} = f_{in}\, d\mu$, $i = 1, 2$, for some measure μ on $\mathcal{X}$. Assume that $f_{1n} \to f$ and $n^{1/2}(f_{1n} - f_{2n}) \to g$, $n \to \infty$, μ-a.e. By the antisymmetry of K it is

$$\sqrt{n}\, \lambda_n = \int\int K(x, y)\, f_{2n}(x) \sqrt{n}(f_{1n}(y) - f_{2n}(y))\, \mu(dx)\, \mu(dy).$$

So under an integrability assumption which ensures the application of, for example, the Dominated Convergence Theorem we can infer that

$$\rho(K) = \frac{\int [\int K(x, y)\, f(x)\, \mu(dx)]^2\, f(y)\, \mu(dy)}{(\int\int K(x, y)\, f(x)\, g(y)\, \mu(dx)\, \mu(dy))^2}. \tag{3.6}$$

In our next result we present a minimizer of the functional ρ on the set $\mathcal{K}$ where only minimal requirements are needed.

PROPOSITION 3.4. *Assume μ is an arbitrary measure and g and g^2/f are μ-integrable. If $\mu(g \neq 0) > 0$ and $\mu(g \neq df) > 0\ \forall d \neq 0$, then the set of admissible kernels*

$$\mathcal{K} = \left\{ K \neq 0 : 0 < \int \left[\int K(x, y)\, f(x)\, \mu(dx) \right]^2 f(y)\, \mu(dy) < \infty, \right.$$

$$\left. 0 \neq \int\int K(x, y)\, f(x)\, g(y)\, \mu(dx)\, \mu(dy) \in \mathbb{R} \right\}$$

is nonempty and the optimal kernel is given by

$$K_{\text{opt}}(x, y) = a_{\text{opt}}(x) - a_{\text{opt}}(y), \qquad x, y \in \mathcal{X},$$

where

$$a_{\text{opt}}(x) = -\frac{g(x)}{f(x)}, \qquad x \in \mathcal{X}.$$

Furthermore

$$\rho(K_{\text{opt}}) = \left(\int \frac{g^2}{f} \, d\mu - \left(\int g \, d\mu \right)^2 \right)^{-1}, \tag{3.7}$$

where

$$\int \frac{g^2}{f} \, d\mu - \left(\int g \, d\mu \right)^2 > 0. \tag{3.8}$$

By (3.5) any affine transformation $\alpha a_{\text{opt}} + \beta$ of α_{opt} with $\alpha \neq 0$ and $\beta \in \mathbb{R}$ also yields an optimal kernel. The functional ρ is well-defined, since then a division by zero in (3.6) is excluded. Observe that the optimal kernel K_{opt} has a very simple shape. This enables us to compute it in many situations, which are of practical relevance. Consider the following example and in particular the next section on functional data.

EXAMPLE 3.5 (Multivariate Location-Scale Model). Let f be a smooth density on $\mathbb{R}^k$. We assume that each vector η of the second sample results from an affine transformation of a vector ξ of the first sample. More precisely

$$f_{1n}(x) = f(x) \qquad \text{and} \qquad f_{2n}(x) = \frac{f(A_n^{-1}(x - d_n))}{|\det(A_n)|},$$

where $A_n = I + rn^{-1/2}A$, $d_n = rn^{-1/2}d$, $r \neq 0$, A is a $k \times k$-matrix, I is the identity matrix, and $d \in \mathbb{R}^k$, where $A \neq 0$ or $d \neq 0$. A Taylor expansion shows that in this case

$$g(x) = r(\text{trace}(A) \, f(x) + \langle Ax + d, \nabla f(x) \rangle)$$

so that by (3.5),

$$a_{\text{opt}}(x) = -\frac{\langle Ax + d, \nabla f(x) \rangle}{f(x)}$$

and

$$\rho(K_{\text{opt}}) = r^{-2}\left[\int \frac{(\text{trace}(A)\, f(x) + \langle Ax + d, \nabla f(x)\rangle)^2}{f(x)}\, dx\right]^{-1}.$$

Here $\langle\ ,\ \rangle$ denotes the usual inner product on $\mathbb{R}^k$ and $\nabla f(x)$ is the gradient of f.

In the location model ($A = 0$, $d \neq 0$) and in the scale model ($A \neq 0$, $d = 0$) the optimal kernel simplifies accordingly. In particular in the univariate case ($k = 1$) one obtains

$$a_{\text{opt}}(x) = -\frac{f'(x)}{f(x)} \qquad \text{with} \qquad \rho(K_{\text{opt}}) = r^{-2} I(f)^{-1}$$

and

$$a_{\text{opt}}(x) = -x\,\frac{f'(x)}{f(x)} \qquad \text{with} \qquad \rho(K_{\text{opt}}) = r^{-2} I_1(f)^{-1},$$

respectively. Here $I(f)$ and $I_1(f)$ denote the corresponding Fisher informations. Notice the close relationship between a_{opt} and the φ-functions of Hájek and Šidák (1967), which bring forth the optimal scores of the rank-test.

Beside of maximizing the local power (2.17) with respect to K one can also try to maximize the local power (2.21). The same arguments which led to Definition 3.3 now lead to

DEFINITION AND PROPOSITION 3.6. *Let $\mathscr{X} = \mathbb{R}$ and assume the local alternatives satisfy* (2.19). *Define*

$$\mathscr{K}^* = \left\{K \neq 0 : 0 < \int\left[\int K(x, y)\, F(dy)\right]^2 F(dx) < \infty,\right.$$

$$\left. 0 \neq \int_0^1 g(u) \int K(x, F^{-1}(u))\, F(dx)\, du < \infty\right\}.$$

If F is continuous then the minimizer

$$K^*_{\text{opt}} = \text{argmin}\{\rho^*(K) : K \in \mathscr{K}^*\}$$

with

$$\rho^*(K) = \frac{\int [\int K(x, y)\, F(dy)]^2\, F(dx)}{(\int_0^1 g(u) \int K(x, F^{-1}(u))\, F(dx)\, du)^2}$$

exists and is given by

$$K^*_{\text{opt}}(x, y) = g(F(x)) - g(F(y)), \qquad x, y \in \mathbb{R}.$$

Moreover,

$$\rho^*(K^*_{\text{opt}}) = \left[\int_0^1 g^2(u)\, du - \left(\int_0^1 g(u)\, du \right)^2 \right]^{-1}.$$

EXAMPLE 3.7 (Cauchy Distribution). Let

$$f(x) = \frac{1}{\pi} \frac{1}{1+x^2} \qquad \text{and} \qquad f_n(x) = \frac{1}{\pi} \frac{1}{1+(x-\mu_n)^2}, \qquad x \in \mathbb{R},$$

be the Lebesgue densities of F and F_n, respectively, where $(\mu_n) \subseteq \mathbb{R}$ converges to zero such that $n^{1/2}\mu_n \to 1$. Then

$$K^*_{\text{opt}}(x, y) = \frac{x}{1+x^2} - \frac{y}{1+y^2} \qquad \text{and} \qquad \rho^*(K^*_{\text{opt}}) = \left(\frac{\pi}{2} - 1 - (\log 2)^2 \right)^{-1}.$$

EXAMPLE 3.8 (Laplace Distribution). Let

$$f(x) = \frac{1}{2} \exp(-|x|) \qquad \text{and} \qquad f_n(x) = \frac{1}{2\lambda_n} \exp\left(-\frac{|x-\mu_n|}{\lambda_n} \right), \qquad x \in \mathbb{R}.$$

If $(\lambda_n) = 1$ and $n^{1/2}\mu_n \to 1$, then

$$K^*_{\text{opt}}(x, y) = \text{sign}(x) - \text{sign}(y) \qquad \text{and} \qquad \rho^*(K^*_{\text{opt}}) = 1.$$

If $(\mu_n) = 0$ and $n^{1/2}(1-\lambda_n) \to 1$, then

$$K^*_{\text{opt}}(x, y) = |x| - |y| \qquad \text{and} \qquad \rho^*(K^*_{\text{opt}}) = 1.$$

For the local alternatives of the last two examples also Proposition 3.4 is applicable. As required we obtain that K_{opt} and K^*_{opt} coincide.

4. APPLICATIONS TO FUNCTIONAL DATA

As was already pointed out at the beginning of this article, in many statistical applications the experimenter observes random functions. For this reason we now consider the sample space $\mathcal{X} = L_2[0, 1]$, which denotes the collection of all real measurable functions g on $[0, 1]$ satisfying $\int g^2(t)\, dt < \infty$. Recall that $L_2 = L_2[0, 1]$ endowed with the usual inner

product $\langle f, g\rangle = \int f(t)\, g(t)\, dt$, $f, g \in L_2$ and the induced norm $\|f\| = \sqrt{\langle f, f\rangle}$ is a Hilbert space.

Let $\xi_1, ..., \xi_m$ be independent copies of a process $\xi = \{\xi(t) : 0 \leqslant t \leqslant 1\}$ with trajectories in L_2 and let $\eta_1, ..., \eta_l$ be independent copies of η, where

$$\eta \stackrel{\mathscr{L}}{=} \xi + \alpha \Delta, \qquad 0 \neq \alpha \in \mathbb{R}, \tag{4.1}$$

and $0 \neq \Delta \in L_2$ is a deterministic drift function. Throughout this section we assume

$$\xi \text{ is a nontrivial zero-mean Gaussian process with covariance function } C \text{ satisfying } \int C(t, t)\, dt < \infty; \tag{4.2}$$

$$\text{the eigenfunctions} \{f_j : j \geqslant 1\} \text{ of } C \text{ are an orthonormal basis of } L_2; \tag{4.3}$$

$$C(s, t) = \sum_{j \geqslant 1} \lambda_j f_j(s)\, f_j(t)\ \forall 0 \leqslant s, t \leqslant 1, \text{ where } \lambda_j \text{ is the eigenvalue associated with the eigenfunction } f_j,\ j \geqslant 1. \tag{4.4}$$

We mention that (4.2)–(4.4) are automatically fulfilled whenever C is continuous and positive definite; cf. Shorack and Wellner (1986, p. 207). Recall that for the construction and optimality of kernels we need densities of $\nu_1 = \mathscr{L}(\xi)$ and $\nu_2 = \mathscr{L}(\eta)$ w.r.t. a common dominating measure μ. The following lemma is due to Grenander (1952).

LEMMA 4.1. *Suppose* (4.1)–(4.4) *holds and*

$$\sum_{i \geqslant 1} \frac{\langle \Delta, f_i\rangle^2}{\lambda_i} < \infty. \tag{4.5}$$

Then

$$\varphi_\alpha(x) = \exp\left\{ -\frac{\alpha^2}{2} \sum_{i \geqslant 1} \frac{\langle \Delta, f_i\rangle^2}{\lambda_i} + \alpha \sum_{i \geqslant 1} \frac{\langle \Delta, f_i\rangle}{\lambda_i} \langle x, f_i\rangle \right\}, \qquad x \in L_2,$$

is a ν_1*-density of* ν_2*:* $\varphi_\alpha = d\nu_2/d\nu_1$.

Note that by Kolmogorov's series theorem φ_α is well-defined ν_1-a.e. Grenander's lemma makes it possible for us to apply the general results of the last section with $\mathscr{X} = L_2$, $\mu = \nu_1$, $f_1 = 1$ and $f_2 = \varphi_\alpha$. It is not difficult to check the conditions of Proposition 3.1, whence we obtain

COROLLARY 4.2. *If* (4.1)–(4.5) *hold, then*

$$K_0(x, y) = h(x) - h(y), \qquad x, y \in L_2$$

with

$$h(x) = \frac{1 - \varphi_\alpha(x)}{\theta + (1-\theta)\,\varphi_\alpha(x)}, \qquad x \in L_2,$$

and

$$K_1(x, y) = l(x) - l(y)$$

with

$$l(x) = \sum_{i \geqslant 1} \frac{\langle \Delta, f_i \rangle}{\lambda_i} \langle x, f_i \rangle, \qquad x \in L_2,$$

entail strictly positive $\lambda(K_0)$ *and* $\lambda(K_1)$.

Next we specify the optimal kernel K_{opt} of Proposition 3.4 for the following local alternatives:

$$\eta \overset{\mathscr{L}}{=} \xi + rn^{-1/2}\,\Delta, \qquad r \neq 0, \qquad n \in \mathbb{N}. \tag{4.6}$$

So with $\mu = \nu_1, f_{1n} = 1, f_{2n} = \varphi_\alpha, \alpha = rn^{-1/2}$ an easy computation shows that

$$n^{1/2}(f_{1n} - f_{2n}) \to g = -rl, \qquad n \to \infty, \qquad \nu_1 - \text{a.e.}$$

Again it is not hard to verify the conditions of Proposition 3.4. Thus we get

COROLLARY 4.3. *If* (4.2)–(4.6) *hold, then*

$$K_{\text{opt}}(x, y) = K_1(x, y) = l(x) - l(y)$$

and

$$\rho(K_{\text{opt}}) = \frac{1}{r^2}\left(\sum_{i \geqslant 1} \frac{\langle \Delta, f_i \rangle^2}{\lambda_i} \right)^{-1} \in (0, \infty). \tag{4.7}$$

Formula (4.7) precisely shows the manner in which the Fourier coefficients $\langle \Delta, f_i \rangle$ of Δ have an effect on the local power of $\Phi_n(K_{\text{opt}})$. The squares of the coefficients can be regarded as weights of the summands $1/\lambda_i$ which tend to infinity as $i \to \infty$. Especially high-frequency functions Δ are detected rather than smooth ones. As a simple example consider $\Delta = f_k$, where $\rho(K_{\text{opt}}) = r^{-2}\lambda_k$ which tends to zero as $k \to \infty$.

EXAMPLE 4.4 (Brownian Motion with Linear Drift). Let ξ be a standard Brownian motion B on $[0, 1]$. Then the eigenfunctions and associated eigenvalues of $C(s, t) = \min(s, t)$ are given by (cf., e.g., Todorovic, 1992, p. 143)

$$f_j(t) = \sqrt{2} \sin\left(j - \frac{1}{2}\right)\pi t \qquad \text{and} \qquad \lambda_j = \frac{1}{\pi^2 (j - 1/2)^2}, \qquad j \geqslant 1.$$

If $\Delta(t) = t$, then straightforward calculations show that $\langle \Delta, f_i \rangle = \sqrt{2}\,(-1)^{i-1}\,\lambda_i$, $i \geqslant 1$, whence (4.5) holds and $l(x) = \sum_{i \geqslant 1} \langle x, f_i \rangle f_i(1) = x(1)$ for all $x \in L_0$, where

$$L_0 := \{x \in L_2 : x(0) = 0,\, x(1-) = x(1)\}.$$

Here, the last equality can be proved with Féjer's theorem. Since $\nu_1(L_0) = P(\xi \in L_0) = 1$, the optimal kernel admits the representation

$$K_{\text{opt}}(x, y) = x(1) - y(1), \qquad x, y \in L_2.$$

Moreover,

$$\rho(K_{\text{opt}}) = \frac{1}{r^2}. \tag{4.8}$$

Our example is an extreme case, where the optimal kernel only uses the value at the endpoint one per observed function. At first sight this might be surprising, but obviously this is due to the drift-function $\alpha\,\Delta$ being the straight line starting at the origin with slope $\alpha = rn^{-1/2}$. Formula (4.8) reflects the large sample performance of the optimal test in dependence on the amount of the "slope" r of the underlying drift $rn^{-1/2}t$. However, note that in general as can be seen from the shape of $a_{\text{opt}} = l$ the optimal test processes the whole information of each single observation point $x \in L_2$ via the inner products $\langle x, f_i \rangle$, $i \geqslant 1$, in $l(x)$. For instance consider

EXAMPLE 4.5 (Brownian Motion with Sinusoidal Drift). Let $\xi = B$ and $\Delta = f_k$ the kth eigenfunction of B. Then

$$a_{\text{opt}}(x) = \int_0^1 x(t) \sin\left(j - \frac{1}{2}\right)\pi t\, dt$$

$$\text{and} \qquad \rho(K_{\text{opt}}) = \frac{1}{r^2}\,\lambda_k = \frac{1}{r^2}\,\frac{4}{\pi^2 (2k-1)^2}.$$

The Brownian motion is a well-known and very important stochastic process. Its well-developed theory allows a nice simplification of the

analytical expressions of K_{opt} and $\rho(K_{\text{opt}})$. The key to this improvement is the following counterpart to Grenander's Lemma 4.1.

LEMMA 4.6. *Suppose ξ is a Brownian motion B on $[0, 1]$, and Δ is twice continuously differentiable on $[0, 1]$ with $\Delta' \neq 0$. If (4.1) holds, then*

$$\varphi_\alpha(x) = \exp\left\{ -\frac{\alpha^2}{2} \int_0^1 (\Delta'(t))^2 \, dt \right.$$
$$\left. + \alpha \left(x(1)\, \Delta'(1) - \int_0^1 x(t)\, \Delta''(t)\, dt \right) \right\}, \qquad x \in C[0, 1]$$

is a density of η w.r.t. the Wiener measure on $C[0, 1]$.

Lemma 4.6 is a consequence of the Corollary to Theorem 7.1 of Liptser and Shiryayev (1977). Analogously to the derivation of the Corollaries 4.2 and 4.3 we obtain

COROLLARY 4.7. *Suppose $\xi = B$ and $\Delta \in C^{(2)}[0, 1]$ with $\Delta' \neq 0$. If (4.1) holds, then K_0 and K_1 of Corollary 4.2 induce strictly positive $\lambda(K_0)$ and $\lambda(K_1)$. Especially, the mapping l which determines K_1 is given by*

$$l(x) = x(1)\, \Delta'(1) - \int_0^1 x(t)\, \Delta''(t)\, dt, \qquad x \in C[0, 1].$$

If (4.6) holds then $K_{\text{opt}} = K_1$ and

$$\rho(K_{\text{opt}}) = \frac{1}{r^2} \left(\int_0^1 (\Delta'(t))^2 \, dt \right)^{-1}. \tag{4.9}$$

In order to deduce the equality (4.9) from the general formula (3.7) note that $g = -rl$, whence

$$\rho(K_{\text{opt}}) = \frac{1}{r^2} \frac{1}{\operatorname{Var}(l(B))}.$$

Since by Itô's formula (cf. Theorem 3.6 in Karatzas and Shreve (1988))

$$l(B) = \int_0^1 \Delta'(t)\, B(dt),$$

(4.9) follows from the elementary properties of stochastic integrals (cf. Karatzas and Shreve, 1988, p. 137). Formula (4.9) describes how the local power of the optimal test is influenced by the non-smoothness of the drift Δ.

If for example the graph of Δ is very wiggly, then the Type II error is very small.

Finally we consider a model which extends the situation of Example 4.4 in many respects. Let $W = \{ W(t) : 0 \leqslant t \leqslant T \}$ be the stochastic process with

$$W(t) = B(t) + Y(t-\tau)^+, \qquad 0 \leqslant t \leqslant T,$$

where B is a standard Brownian motion on $[0, T]$, $T > 0$ is a finite time horizon, and Y and τ are random variables such that B, Y and τ are independent. Moreover, Y is normally distributed with expectation $\mu \in \mathbb{R}$ and variance $\sigma^2 \in [0, \infty)$, whereas τ is any nonnegative random variable with distribution $\mathscr{L}(\tau) = \rho$ and support $[0, T]$. This means that up to the random point τ the process $W(t)$ is a standard Brownian motion and after the time τ a linear drift-function with random slope Y is added. Before τ the process $W(t)$ is regarded as to be "under control" and after τ as to be "out of control." Beibel (1997) investigates the process W in a sequential framework. He proves the following lemma.

LEMMA 4.8. *Suppose ξ is a standard Brownian motion on $[0, T]$, $T > 0$, and $\eta \overset{\mathscr{L}}{=} W$. If $\sigma^2 > 0$, then*

$$\psi(x) = \frac{1}{\sigma} e^{-\mu^2/2\sigma^2} \int_0^T \frac{1}{\sqrt{T - s + \sigma^{-2}}} \times \exp\left\{ \frac{(x(T) - x(s) + \mu\sigma^{-2})^2}{2(T - s + \sigma^{-2})} \right\} \rho(ds), \qquad x \in C[0, T]$$

is a density of $\eta \overset{\mathscr{L}}{=} W$ with respect to the Wiener measure $v_1 = \mathscr{L}(B)$.

In the limit case $\sigma^2 = 0$, i.e., $Y = \mu$ almost surely, the density is given by

$$\psi(x) = \int_0^T \exp\{ -\tfrac{1}{2}(T-s)\mu^2 + \mu(x(T) - x(s)) \} \rho(ds), \qquad x \in C[0, T]. \tag{4.10}$$

The above lemma can be used in the same fashion for the construction of suitable and optimal kernels as its counterparts Lemmas 4.1 and 4.6.

EXAMPLE 4.9. Let $T = 1$ and $\tau = 0$. If $\mu = 0$ and $\sigma^2 = 1$, then

$$\psi(x) = \frac{1}{\sqrt{2}} \exp\left(\frac{x(1)^2}{4} \right), \qquad \mathscr{L}(B)\text{-a.e.},$$

which coincides with formula (7.11) in Liptser and Shiryayev (1977). If $\mu = \alpha$ and $\sigma^2 = 0$, then

$$\psi(x) = \exp\left\{-\frac{\alpha^2}{2} + \alpha x(1)\right\}, \qquad \mathscr{L}(B)\text{-a.e.},$$

which could also be deduced from Lemma 4.1 or Lemma 4.6.

In conclusion we consider the following contiguous alternatives:

$$\eta \stackrel{\mathscr{L}}{=} \{B(t) + rn^{-1/2}(t-\tau)^+, 0 \leqslant t \leqslant T\}, \qquad r \neq 0, \qquad n \in \mathbb{N}. \quad (4.11)$$

With $f_{1n} = 1$ and f_{2n} equal to ψ in (4.10) with $\mu = rn^{-1/2}$ it follows from Lemma 4.8 that

$$n^{1/2}(f_{1n} - f_{2n}) \to g \qquad \text{as} \qquad n \to \infty,$$

where

$$g(x) = -r\left[x(T) - \int_0^T x(s)\,\rho(ds)\right], \qquad x \in C[0, T].$$

If τ is absolutely continuous, Itô's formula implies that

$$g(B) = -r\int_0^T R(s)\,B(ds),$$

where R is the distribution function of τ. Thus by Proposition 3.4 the optimal kernel K_{opt} is determined by

$$a_{\text{opt}}(x) = x(T) - \int_0^T x(s)\,\rho(ds), \qquad x \in C[0, T]$$

and

$$\rho(K_{\text{opt}}) = r^{-2}\left(\int_0^T R^2(s)\,ds\right)^{-1}.$$

If $T = 1$ and $\tau = 0$ almost surely this is consistent with Example 4.4.

5. SIMULATIONS

In this section we report on the results of a small simulation study. The power of our test is computed on certain parametrized parts of the alternative H_1 for several kernels K. This is done for functional data ($\mathscr{X} = L_2$)

as well as for vector-valued observations ($\mathcal{X} = \mathbb{R}^4$). In both cases the optimal kernel is compared with ad hoc kernels.

First, we consider

$$\xi_1, ..., \xi_m \qquad \text{and} \qquad \eta_1, ..., \eta_l, \tag{5.1}$$

where $\xi_1, ..., \xi_l$ are independent copies of a Brownian motion B on $[0, 1]$ and the $\eta_1, ..., \eta_l$ are independent copies of

$$\eta \overset{\mathcal{L}}{=} B + r\Delta, \qquad r \geqslant 0,$$

with Δ specified below. Fix $m = l = 50$ and $\alpha = 0.05$ (level of significance). Let $K_i(x, y) = a_i(x) - a_i(y)$, $i = 1, 2, 3$, be the kernels defined by $a_1 = a_{\text{opt}}$,

$$a_2(x) = \langle x, 1 \rangle = \int_0^1 x(t)\, dt \qquad \text{and} \qquad a_3(x) = \langle x, x \rangle = \int_0^1 x^2(t)\, dt$$

and let

$$p_i(r) := P_r(\Phi_n^+(K_i) \text{ rejects } H_0), \qquad r \geqslant 0,$$

denote the power function of $\Phi_n^+(K_i)$, $i = 1, 2, 3$. Tables I and II contain Monte Carlo approximations of the power $p_i(r)$. Each single value has been obtained from 10,000 replicates.

We see that the optimal test outperforms the ad hoc procedures. Also note that in all cases the actual probability of the type I error is very close to the given level of significance $\alpha = 0.05$.

TABLE I

$\Delta(t) = \sqrt{2} \sin \frac{5}{2} \pi t$

r	$p_1(r)$	$p_2(r)$	$p_3(r)$
0.0	0.0483	0.0477	0.0498
0.1	0.5356	0.0583	0.0515
0.2	0.8780	0.0697	0.0739
0.3	0.9766	0.0860	0.1229
0.4	0.9971	0.1035	0.1914
0.5	0.9999	0.1205	0.3083
0.6	1.0000	0.1330	0.4470
0.7	1.0000	0.1534	0.6039
0.8	1.0000	0.1691	0.7545
0.9	1.0000	0.1945	0.8464
1.0	1.0000	0.2099	0.9115

TABLE II

$\Delta(t) = (t-\tau)^+$, $\mathscr{L}(\tau) = U(0, 1)$

r	$p_1(r)$	$p_2(r)$	$p_3(r)$
0.0	0.0498	0.0482	0.0505
0.2	0.0976	0.0697	0.0921
0.4	0.1668	0.0981	0.0976
0.6	0.2371	0.1200	0.0981
0.8	0.3154	0.1578	0.0988
1.0	0.3888	0.1860	0.0967
1.2	0.4732	0.2325	0.0980
1.4	0.5530	0.2657	0.0950
1.6	0.6043	0.3046	0.0932
1.8	0.6640	0.3454	0.1037
2.0	0.7124	0.3664	0.0959

Second, we considered (5.1) where $\xi_1, ..., \xi_m$ are independent copies of the normal vector

$$\xi = \begin{bmatrix} 1 & -1 & 0 & 3 \\ 0 & 2 & 1 & -4 \\ 1 & 4 & -6 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \end{bmatrix}$$

and $N_1, ..., N_4$ are i.i.d. standard normal random variables. The variables $\eta_1, ..., \eta_l$ of the second sample are independent copies of

$$\eta \overset{\mathscr{L}}{=} \xi + r(2, -3, -5, 7)', \qquad r \geqslant 0.$$

According to Example 3.5 the optimal kernel is given by

$$K_{\text{opt}}(x, y) = d'S^{-1}(x-y), \qquad x, y \in \mathbb{R}^4,$$

where $d = (2, -3, -5, 7)'$ is the shift-vector and S^{-1} is the inverse of the covariance matrix S of ξ. As a competing kernel we took

$$K(x, y) = d'(x-y), \qquad x, y \in \mathbb{R}^4.$$

Table III lists the power of the pertaining tests $\Phi_n^+(K_{\text{opt}})$ and $\Phi_n^+(K)$. We approximated each single value by 100,000 Monte Carlo replicates.

Note the considerable difference of the two power functions. This all the more is remarkable since K uses the true shift vector d as part of the information.

TABLE III

$m = 40, l = 60, \alpha = 0.05$

r	K_{opt}	K
0.00	0.05078	0.04962
0.01	0.13716	0.05597
0.02	0.26344	0.06257
0.03	0.40385	0.06913
0.04	0.53076	0.07746
0.05	0.64381	0.08393
0.06	0.73343	0.09268
0.07	0.80377	0.09925
0.08	0.85837	0.10676
0.09	0.89868	0.11595
0.10	0.92867	0.12811
0.11	0.95042	0.13696
0.12	0.96483	0.14608
0.13	0.97614	0.15554
0.14	0.98430	0.16828
0.15	0.98896	0.18033

Finally, we compared our test with the Kolmogorov–Smirnov test. For that purpose we generated samples (5.1) of ξ's and η's with ξ is uniformly distributed on $(-1/2, 1/2)$ and either $\eta \stackrel{\mathscr{L}}{=} \xi + d$ (location model) or $\eta \stackrel{\mathscr{L}}{=} (1+d)\,\xi$ (scale model), where $d \geqslant 0$. Tables IV and V show the power of the KS-test and of $\Phi_n^+(K)$, where the ad hoc kernels $K(x, y) = x - y$ (location model) and $K(x, y) = x^2 - y^2$ (scale model) have been used.

TABLE IV

Uniform Location Model

d	KS	Φ_n^+
0.00	0.0381	0.0494
0.01	0.0409	0.0610
0.02	0.0456	0.0757
0.03	0.0526	0.0902
0.04	0.0641	0.1064
0.05	0.0789	0.1227
0.06	0.0939	0.1425
0.07	0.1168	0.1631
0.08	0.1400	0.1866
0.09	0.1702	0.2067
0.10	0.2037	0.2287
0.20	0.7138	0.4692

TABLE V

Uniform Scale Model

d	KS	Φ_n^+
0.00	0.0381	0.0462
0.04	0.0401	0.0773
0.08	0.0418	0.1154
0.12	0.0449	0.1626
0.16	0.0504	0.2143
0.20	0.0581	0.2627
0.24	0.0675	0.3124
0.28	0.0826	0.3550
0.32	0.0985	0.3981
0.36	0.1208	0.4362
0.40	0.1487	0.4721
0.80	0.5935	0.6715

Furthermore we specified: $\alpha = 0.05$, $m = l = 50$ and 10,000 Monte Carlo replicates per single value.

Tables VI and VII show the corresponding power values in case of standard normal ξ's.

In the scale model the Φ^+-test is uniformly better than the KS-test. In the location model this also holds but only for small deviations $d > 0$ from the hypothesis $d = 0$. For larger shifts d the performance of the KS-test is superior.

TABLE VI

Normal Location Model

d	KS	Φ_n^+
0.00	0.0373	0.0472
0.02	0.0380	0.0539
0.04	0.0398	0.0616
0.06	0.0432	0.0689
0.08	0.0471	0.0767
0.10	0.0536	0.0851
0.20	0.1093	0.1352
0.26	0.1643	0.1720
0.30	0.2129	0.1983
0.34	0.2673	0.2268
0.38	0.3300	0.2536
0.40	0.3646	0.2681

TABLE VII

Normal Scale Model

d	KS	Φ_n^+
0.00	0.0373	0.0514
0.04	0.0374	0.0698
0.08	0.0387	0.0920
0.12	0.0415	0.1134
0.16	0.0440	0.1386
0.20	0.0481	0.1651
0.40	0.0891	0.2881
0.52	0.1311	0.3507
0.60	0.1607	0.3880
0.72	0.2135	0.4329
0.80	0.2562	0.4550

6. PROOFS

Proof of Propositon 2.1. By Proposition 3.1 of Ferger (1995) we know that $\theta_n^{(+)} \xrightarrow{\mathscr{L}} \tau_w^{(+)}$ as $n \to \infty$. Since

$$P(\Phi_n^{(+)} \text{ rejects } H_0) = P\left(\left|\theta_n^{(+)} - \frac{m}{n}\right| \leqslant c^{(+)}\right)$$

the assertion follows from (1.7) upon noticing that $\tau_w^{(+)}$ is a continuous random variable. ∎

Proof of Proposition 2.2. By Proposition 3.7 of Ferger (1995) $\tilde{\theta}_n^+$ is uniformly distributed on the grid $\{k/n : 0 \leqslant k \leqslant n-1\}$, which proves the theorem. ∎

Proof of Proposition 2.3. By Proposition 2.1 of Ferger (1995) $\theta_n^{(+)} \to \theta$, $n \to \infty$, P-stochastically. Recall that $m/n \to \theta$ as $n \to \infty$, which implies (2.5) and (2.6), respectively. ∎

Proof of Proposition 2.4. According to Proposition 2.3 the condition $\lambda \neq 0 (\lambda > 0)$ is sufficient for the consistency of Φ_n (Φ_n^+). By Proposition 2.5, $\lambda \neq 0$ is also neccessary for the consistency of Φ_n. This shows (2.8). It remains to prove the if-part of (2.9). For this assume that

$$\lim_{n \to \infty} P(\Phi_n^+ \text{ rejects } H_0) = 1, \tag{6.1}$$

but $\lambda \leqslant 0$. In case of $\lambda = 0$ Proposition 2.5 yield the desired contradiction. If $\lambda < 0$, then

$$\lim_{n \to \infty} P(\theta_n^+ \notin [\varepsilon, 1-\varepsilon]) = 1 \qquad \text{for all} \quad \varepsilon > 0. \tag{6.2}$$

To see this we first introduce the stochastic process

$$r_n(t) = n^{-2} \sum_{i=[nt]+1}^{n} \sum_{j=1}^{[nt]} K(X_i, X_j), \qquad 0 \leqslant t \leqslant 1.$$

Then with the function Δ defined in Proposition 2.10 the following inclusion holds for all $\varepsilon > 0$:

$$\left\{ \max_{0 \leqslant t \leqslant 1} |r_n(t) - \lambda \Delta(t)| \leqslant -\frac{\lambda}{2} \min(\theta, 1-\theta)\, \varepsilon \right\} \subseteq \{\theta_n^+ \notin [\varepsilon, 1-\varepsilon]\}.$$

Since $\lambda < 0$, Lemma 2.1 of Ferger (1995) immediately yields (6.2). By assumption $\theta \in (\frac{\alpha}{2}, 1-\frac{\alpha}{2})$ so that by (1.7) $c^+ = \alpha/2$, whence $\varepsilon^* := \frac{1}{2}(\min(\theta, 1-\theta) - \frac{\alpha}{2})$ is strictly positive. Thus with (6.1) and (6.2) we obtain that

$$1 = \lim_{n \to \infty} P\left(\left| \theta_n^+ - \frac{m}{n} \right| \leqslant c^+ \right) \leqslant \lim P(\theta_n^+ \in [\varepsilon^*, 1-\varepsilon^*]) = 0,$$

which is a contradiction. ∎

Proof of Proposition 2.5. If we put

$$\Gamma_n(t) := n^{1/2} r_n(t), \qquad 0 \leqslant t \leqslant 1,$$

then by Theorem 1.1 of Ferger (1994a),

$$\Gamma_n \xrightarrow{\mathscr{L}} \tau B_0, \qquad n \to \infty,$$

where $\xrightarrow{\mathscr{L}}$ denotes convergence in distribution in the Skorokhod space $D = D[0, 1]$. Bhattacharya and Brockwell (1976) proved that the argmax-functional is continuous on the set $\hat{C}$ of continuous functions on $[0, 1]$ with unique maximizer. Since $B_0 \in \hat{C}$ and $|B_0| \in \hat{C}$ with probability one (as shown in the proof of Theorem 3.1 of Ferger, 1995), the assertion follows with the Continuous Mapping Theorem. ∎

Proof of Proposition 2.6. Observe that

$$P(\Phi_n^{(+)} \text{ does not reject } H_0) = P\left(n \left| \theta_n^{(+)} - \frac{m}{n} \right| > n c^{(+)} \right).$$

From the continuity of $\tau_w^{(+)}$ we can conclude that $c^{(+)}$ is strictly positive, whence the assertion follows from the tail-inequalities for the change-point estimator given in Proposition 2.1 of Ferger (1995). ▮

Proof of Proposition 2.8. Apply Proposition 2.1 of Ferger (1995) as in the above proof. ▮

Proof of Proposition 2.10. Let $D = D[0, 1]$ be the Skorokhod space and put

$$M[f] := \sup_{0 \leqslant u \leqslant 1} f(u), \qquad f \in D,$$

and

$$S[f] := \{0 \leqslant u \leqslant 1 : f(u) = M[f] \text{ or } f(u-) = M[f]\}.$$

By Lemma 5.5 of Ferger (1995) $S[f]$ is a nonempty closed subset of $[0, 1]$. Therefore the mapping $\psi: D \to [0, 1]$ defined by

$$\psi[f] := \min S[f]$$

is well-defined. The mapping $\psi_n: D \to [0, 1]$ defined by

$$\psi_n[f] := \operatorname*{argmax}_{t \in G_n} f(t), \qquad f \in D,$$

where $G_n := \{kn^{-1} : 1 \leqslant k \leqslant n-1\}$ entails the representation

$$\theta_n = \psi_n[w\,|\Gamma_n|]. \tag{6.3}$$

The mean functions $\bar{\Gamma}_n(t) := E\Gamma_n(t)$, $0 \leqslant t \leqslant 1$, converge uniformly on $[0, 1]$ to $\gamma\,\Delta(t)$. An application of Theorem 1.1 of Ferger (1994a) and a Cramér–Slutsky argument yields

$$\Gamma_n \xrightarrow{\mathscr{L}} \tau B_0 + \gamma\,\Delta, \qquad n \to \infty.$$

For truncated w this implies

$$w|\Gamma_n| \xrightarrow{\mathscr{L}} B^* = w\,|\tau B_0 + \gamma\,\Delta|, \qquad n \to \infty, \tag{6.4}$$

where $B^* \in \hat{C}$ with probability one. In view of (6.3) and (6.4) the next step is to check whether the extended Continuous Mapping Theorem 5.5 in Billingsley (1968) is applicable. For that purpose we consider

$$E := \{f \in D : \exists (f_n) \subseteq D,\ f_n \to_s f,\ \psi_n[f_n] \not\to \psi[f]\},$$

where $\to_s$ denotes convergence in the Skorokhod topology. According to Lemma 5.5 of Ferger (1995),

$$\hat{C} \subseteq D \backslash E.$$

But as already mentioned $P(B^* \in \hat{C}) = 1$. Thus we can infer that

$$\theta_n \xrightarrow{\mathscr{L}} \psi[B^*] = T.$$

Since T is a continuous random variable, (2.17) follows for the two-sided test. The proof of (2.17) for the one-sided test Φ_n^+ is completely analoguous. ▮

Proof of Proposition 2.11. Since

$$w\Gamma_n \xrightarrow{\mathscr{L}} w(\tau^* B_0 + \gamma^* \Delta)$$

by Theorem 3.4 of Szyszkowicz (1991) we can proceed as in the previous proof. ▮

Proof of Proposition 3.1. Since

$$\int |K| \, dv_1 \otimes v_2 \leqslant \int |h| \, dv_1 + \int |h| \, dv_2 \leqslant \left(\frac{1}{p} + \frac{1}{1-p} \right) \int |f_1 - f_2| \, d\mu,$$

it follows, that K in (3.3) is $(v_1 \otimes v_2)$-integrable. Consequently $\lambda(K)$ exists and is given by

$$\lambda(K) = \int \frac{(f_1 - f_2)^2}{pf_1 + (1-p) f_2} \, d\mu,$$

which is strictly positive, because $v_1 \neq v_2$. As to the invariance of $K = K_\mu$ with respect to μ, we first assume that $\mu \ll \tilde{\mu}$, where $\tilde{\mu}$ is σ-finite. Then by the chain rule

$$f_i \frac{d\mu}{d\tilde{\mu}} = \frac{dv_i}{d\tilde{\mu}}, \qquad i = 1, 2,$$

and thus $K_{\tilde{\mu}} = K_\mu$, since the factor $d\mu/d\tilde{\mu}$ cancels. For the same reason $K_{\tilde{\tau}} = K_\tau$, if $\tau \ll \tilde{\tau}$ and $\tilde{\tau}$ is σ-finite. Especially for $\tilde{\mu} = \tilde{\tau} := \mu + \tau$ we obtain that $K_\mu = K_{\tilde{\mu}} = K_{\tilde{\tau}} = K_\tau$. This completes the proof of the first part. The second part is proved similarly. ▮

The proofs of Propositions 3.4 and 3.6 are given in Ferger (1997).

REFERENCES

I. A. Ahmad and P. B. Cerrito, Goodness of fit tests based on the L_2-norm of multivariate probability density functions, *Nonparametric Statist.* **2** (1993), 169–181.

R. Bahr, "A New Test for the Multi-Dimensional Two-Sample Problem with General Alternatives," Doctoral thesis, University of Hannover, 1996. [In German]

M. Beibel, Sequential change-point detection in continuous time when the post-change drift is unknown, *Bernoulli* **3** (1997), 457–478.

P. K. Bhattacharya and P. J. Brockwell, The minimum of an additive process with applications to signal estimation and storage theory, *Z. Wahrsch. Verw. Gebiete* **37** (1976), 51–75.

P. Billingsley, "Convergence of Probability Measures," Wiley, New York, 1968.

E. Carlstein, Nonparametric change-point estimation, *Ann. Statist.* **16** (1988), 188–197.

M. Csörgő and L. Horváth, "Weighted Approximations in Probability and Statistics," Wiley, New York, 1993.

B. S. Darkhovski, A nonparametric method for the a posteriori detection of the "disorder" time of a sequence of independent random variables, *Theory Probab. Appl.* **21** (1976), 178–183.

B. S. Darkhovski, Nonparametric methods in change-point problems: A general approach and some concrete algorithms, *in* "IMS Lecture Notes Monographs Series," Vol. 23, pp. 99–107, 1994.

L. Dümbgen, The asymptotic behavior of some nonparametric change-point estimators, *Ann. Statist.* **19** (1991), 1471–1495.

J. H. J. Einmahl and E. V. Khmaladze, The two-sample problem in $\mathbb{R}^m$ and measure-valued martingales, technical report S 98-2, University of New South Wales, Australia, 1998.

T. W. Epps and K. J. Singleton, An omnibus test for the two-sample problem using the empirical characteristic function, *J. Statist. Comput. Simul.* **26** (1986), 177–203.

D. Ferger, An extension of the Csörgő–Horváth functional limit theorem and its applications to change-point problems, *J. Multivariate Anal.* **51** (1994a), 338–351.

D. Ferger, On the power of nonparametric change-point tests, *Metrika* **41** (1994b), 277–297.

D. Ferger, "Change-Point Estimators Based on Weighted Empirical Processes with Applications to the Two-Sample Problem in General Measurable Spaces," Habilitationsschrift, University of Giessen, 1995. [in German]

D. Ferger, Optimal change-point detection, unpublished manuscript, Dresden University of Technology, (1997).

D. Ferger, Testing for the existence of a Change-Point in a specified time interval, *in* "Advances in Stochastic Models for Reliability, Quality and Safety," pp. 277–289, Birkhäuser, Boston/Basel/Berlin, 1998.

E. Gombay and L. Horáth, Approximations for the time of change and the powerfunction in change-point models, preprint, 1993.

U. Grenander, Stochastic processes and statistical inferences, *Ark. Mat.* (1952), 195–277.

J. Hájek and Z. Šidák, "Theory of Rank Tests," Academic Press, New York, 1967.

N. Henze, A multivariate two-sample test based on the number of nearest neighbor type coincidences, *Ann. Statist.* **16** (1988), 772–783.

N. Henze and B. Voigt, Almost sure convergence of certain slowly changing symmetric one- and multi-sample statistics, *Ann. Probab.* **20** (1992), 1086–1098.

T. P. Hettmansperger and J. W. McKean, "Robust Nonparametric Statistical Methods," Kendall's Library of Statistics, Vol. 5, Wiley, New York, 1998.

C. V. Hinkley, Inference about the change-point in a sequence of random variables, *Biometrika* **57** (1970), 1–17.

M. Hollander and D. A. Wolfe, "Nonparametric Statistical Methods," Wiley, New York, 1973.

M. Hušková, Tests and estimators for the change-point problem based on M-statistics, preprint, Charles University of Praha, 1996.
I. Karatzas and S. E. Shreve, "Brownian Motion and Stochastic Calculus," Graduate Texts in Mathematics, Vol. 113, 2nd ed., Springer-Verlag, New York/Berlin/Heidelberg/London/Paris/Tokyo, 1988.
E. L. Lehmann, "Nonparametrics: Statistical Methods Based on Ranks," Holden–Day, San Francisco, 1975.
R. S. Liptser and A. N. Shiryayev, "Statistics of Random Processes I," Springer-Verlag, Berlin, 1977.
F. Lombard and J. D. Hart, The analysis of change-point data with dependent errors, *in* "IMS Lecture Notes Monograph Series," Vol. 23, pp. 194–209, 1994.
J. O. Ramsay and B. W. Silverman, "Functional Data Analysis," Springer-Verlag, New York, 1997.
G. R. Shorack and J. A. Wellner, "Empirical Processes with Applications to Statistics," Wiley, New York, 1986.
B. Szyszkowicz, Change-point problems and contiguous alternatives, *Statist. Probab. Lett.* **11** (1991), 299–308.
P. Todorivic, "An Introduction to Stochastic Processes and Their Applications," Springer-Verlag, New York, 1992.
Y.-C. Yao, D. Huang, and R. Davis, On almost sure behavior of change-point estimators, *in* "IMS Lecture Notes Monograph Series," Vol. 23, pp. 359–371, 1994.