Weierstraß-Institut
für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Preprint ISSN 0946 – 8633

Sharp-Optimal Adjustment for Multiple Testing in the Multivariate Two-Sample Problem

Angelika Rohde[1]

submitted: 27th August 2008

[1] Weierstraß-Institut Berlin
E-Mail: rohde@wias-berlin.de

No. 1356
Berlin 2008

2000 *Mathematics Subject Classification.* 62G10, 62G20.

Key words and phrases. Combinatorial process, exponential concentration bound, coupling, decoupling inequality, exact multiple test, nearest-neighbors, sharp asymptotic adaptivity.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Mohrenstraße 39
10117 Berlin
Germany

Fax: + 49 30 2044975
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Abstract

Based on two independent samples $X_1, ..., X_m$ and $X_{m+1}, ..., X_n$ drawn from multivariate distributions with unknown Lebesgue densities p and q respectively, we propose an exact multiple test in order to identify simultaneously regions of significant deviations between p and q. The construction is built from randomized nearest-neighbor statistics. It does not require any preliminary information about the multivariate densities such as compact support, strict positivity or smoothness and shape properties. The adjustment for multiple testing is sharp-optimal for typical arrangements of the observation values which appear with probability close to one, and it relies on a new coupling Bernstein type exponential inequality, reflecting the non-subgaussian tail behavior of the combinatorial process. For power investigation of the proposed method a reparametrized minimax set-up is introduced, reducing the composite hypothesis "$p = q$" to a simple one with the multivariate mixed density $(m/n)p + (1 - m/n)q$ as infinite dimensional nuisance parameter. Within this framework, the test is shown to be spatially and sharply asymptotically adaptive with respect to uniform loss on isotropic Hölder classes.

1 Introduction

Given two independent multivariate iid samples

$$X_1, ..., X_m \quad \text{and} \quad X_{m+1}, ..., X_n$$

with corresponding Lebesgue densities p and q respectively, we are interested in identifying simultaneously subregions of the densities support where p deviates significantly from q at prespecified but arbitrarily chosen level $\alpha \in (0, 1)$. For this aim a multiple test of the composite hypothesis $H_0 : p = q$ versus $H_A : p \neq q$ is proposed, built from a suitable combination of randomized nearest-neighbour statistics. The procedure does not require any preliminary information about the multivariate densities such as compact support, strict positivity or smoothness and shape properties, and it is valid for arbitrary finite sample sizes m and $n - m$. The hierarchical structure of p-values for subsets of deviation between p and q provides insight into the local power of nearest-neighbor classifiers, based on the training set $\{X_1, ..., X_n\}$. Thus our method is of interest in particular if the classification error depends strongly on the value of the feature vector, related to recent literature on classification procedures by Belomestny and Spokoiny (2007).

There is an extensive amount on literature concerning two-sample problems. Most of it is devoted to the one-dimensional case as there exists the simple but powerful "quantile transformation", allowing for distribution-freeness under the null hypothesis of several test statistics. Starting from the classical univariate mean shift problem (see e.g. Hájek and Šidák 1967), more flexible alternatives as stochastically larger or omnibus alternatives have been investigated for instance by Behnen, Neuhaus and Ruymgaart (1983), Neuhaus (1982, 1987), Fan (1996), Janic-Wróblewska and Ledwina (2000) and Ducharme and Ledwina (2003). Our approach is different in that it aims at spatially adaptive and simultaneous identification of local rather than global deviations. In the above cited literature asymptotic power is discussed against single directional alternatives tending to zero at a prespecified rate, typically formulated by means of the densities $\tilde{p}$ and $\tilde{q}$ corresponding to the transformed observations $\tilde{X}_i = H(X_i)$, where H denotes the mixed distribution function with density $h = (m/n)p + (1 - m/n)q$. Note that the mapping H coincides with the inverse quantile transformation under the null.

For power investigation of our procedure a specific two-sample minimax set-up is introduced. It is based on a reparametrization of (p, q) to a couple (ϕ, h), reducing the composite hypothesis "$p = q$" to the simple one "$\phi \equiv 0$" with the multivariate mixed density h as infinite dimensional nuisance parameter. The reparametrization conceptionally differs from the above described transformation for the univariate situation as it cannot rely on the inverse mixed distribution function. Nevertheless it leads under moderate additional assumptions in that case to the same notion of efficiency. In order to explore the power of our method, the alternative is assumed to be of the form

$$\Big\{(p, q) : \ (m/n)p + (1 - m/n)q = h, \ \phi \in \mathcal{F}, \ \|\phi\| \geq \delta\Big\} \tag{1}$$

for fixed but unknown h, some suitably chosen (semi-)norm $\|.\|$, a constant $\delta > 0$ and a given smoothness class $\mathcal{F}$. For any $\alpha \in (0, 1)$ the quality of a statistical level-α-test ψ is then quantified by its minimal power

$$\inf \mathbb{E}_{(p,q)}\psi,$$

where the infimum is running over all couples (p, q) which are contained in the set (1). It is a general problem that an optimal solution ψ may depend on $\mathcal{F}$ and h. Since the smoothness and shape of a potential difference $p - q$ are rarely known in practice, it is of interest to come up with a procedure which does not depend on these properties but is (almost) as good as if they were known, leading to the notion of minimax adaptive testing as introduced in Spokoiny (1996). Note that here we have however h as an additional infinte dimensional nuisance parameter.

The problem of data-driven testing a simple hypothesis is further investigated for instance by Eubank and Hart (1992), Ledwina (1994), Ledwina and Kallenberg (1995), Fan (1996) and Dümbgen and Spokoiny (2001) among others. The idea in common is to combine a family of test statistics corresponding to different values of the smoothing parameters, respectively. The closest in spirit to ours is the multiscale test developed in D'umbgen and Spokoiny (2001) within the continuous time Gaussian white noise model and further explored by Dümbgen (2002), Dümbgen and Walther (2008) and Rohde (2008), all concerned with univariate problems.

The paper is organized as follows. In the subsequent section, a multiple randomization test is introduced, built from a combination of suitably standardized nearest-neighbor statistics. Its calibration relies on a new coupling exponential bound and an appropriate extension of the multiscale empirical process theory. Asymptotic power investigations and adaptivity properties are studied in Section 3, where the reparametrized minimax setup is introduced. It is shown that our procedure is sharply asymptotically adaptive with respect to $\sup$-norm $\|\cdot\|$ on isotropic Hölder classes $\mathcal{F}$, i.e. minimax efficient over a broad range of Hölder smoothness classes simultaneously. The application to local classification is discussed in Section 4. The one-dimensional situation is considered separately in Section 5 where an alternative approach based on local pooled order statistics is proposed. In that case the statistic does not depend on the observations explicitly but only on their order which in contrast to nearest-neighbor relations is invariant under the quantile transformation. Section 6 is concerned with a decoupling inequality and the coupling exponential bounds which are essential for our construction. Both results are of independent theoretical interest. All proofs and auxiliary results about empirical processes are deferred to Section 7 and Section 8.

2 Combining randomized nearest-neighbor statistics

The procedure below is mainly designed for dimension $d \geq 2$. The univariate case contains a few special features and is considered separately in Section 5. Let $\underline{\mathrm{X}} := (X_1, ..., X_n)'$ and denote by $\mathcal{X}_n$ the pooled set of observations. For any $1 \leq k \leq n$, the k'th nearest-neighbor of $X \in \mathcal{X}_n$ with respect to the *Euclidean distance* is denoted by X^k; additionally define $X^0 := X$. Note that the nearest-neighbors are unique a.s. The weighted labels are defined as follows

$$\Lambda(X) := \begin{cases} \frac{n}{m} & \text{if } X \text{ is contained in the first sample} \\ -\frac{n}{n-m} & \text{otherwise.} \end{cases}$$

In order to judge about some possible deviation of p from q on a given set $B \in \mathcal{B}^d$, a natural statistic to look at is a standardized version of $\hat{\mathbb{P}}_n(B) - \hat{\mathbb{Q}}_n(B)$ or more sophisticated,

$$\int_B \psi_B(x)\Big(d\hat{\mathbb{P}}_n(x) - d\hat{\mathbb{Q}}_n(x)\Big)$$

for some kernel ψ_B supported by B, where $\hat{\mathbb{P}}_n$ and $\hat{\mathbb{Q}}_n$ denote the empirical measures corresponding to the first and second sample, respectively. Note that the statistic is not distribution-free, and in order to build up a multiple testing procedure several statistics corresponding to different sets B have to be combined in a certain way.

2.1 Local nearest-neighbor statistics

Let $\psi : [0,1] \to \mathbb{R}$ denote any kernel of bounded total variation with $\max_{x\in[0,1]} |\psi(x)| = \psi(0) = 1$. We introduce the local test statistics

$$\begin{aligned} T_{jkn} &:= \frac{\sqrt{(m/n)(1-m/n)}}{\gamma_{jkn}} \frac{1}{\sqrt{n}} \sum_{i=0}^{k} \psi\Big(\frac{\|X_j - X_j^i\|_2}{\|X_j - X_j^k\|_2}\Big) \Lambda(X_j^i) \\ &= \frac{\sqrt{(m/n)(1-m/n)}}{\gamma_{jkn}} \sqrt{n} \int \psi\Big(\frac{\|X_j - x\|_2}{\|X_j - X_j^k\|_2}\Big) \Big(d\hat{\mathbb{P}}_n(x) - d\hat{\mathbb{Q}}_n(x)\Big), \end{aligned}$$

where

$$\gamma_{jkn}{}^2 \;:=\; \frac{1}{n-1} \sum_{i=0}^{n-1} \left[\psi\Big(\frac{\|X_j - X_j^i\|_2}{\|X_j - X_j^k\|_2}\Big) - \frac{1}{n} \sum_{l=0}^{n-1} \psi\Big(\frac{\|X_j - X_j^l\|_2}{\|X_j - X_j^k\|_2}\Big) \right]^2 .$$

Every T_{jkn} is some in a certain sense standardized weighted average of the nearest-neighbor's labels and its absolute value should tend to be large whenever p is clearly larger or smaller than q within the random Euclidean ball with center X_j and radius $\|X_j - X_j^k\|_2$.

2.2 Adjustment for multiple testing

The idea is to build up a multiple test, combining all possible local statistics T_{jkn}. Precisely, we aim at a supremum type test statistic

$$T_n \;:=\; \sup_{1\le k\le n} \sup_{1\le j\le n} \Big\{ |T_{jkn}| - C_{jkn} \Big\},$$

where the constants C_{jkn} are appropriately chosen correction terms (independent of the label vector Λ) for adjustment of multiple testing within every "scale" k of k-nearest-neighbor statistics. Although the distribution of T_n under the null hypothesis depends on the unknown underlying distribution $p = q$, the conditional distribution $\mathcal{L}_0(T_n|\mathcal{X}_n)$ of the above statistic is invariant under permutation of the the components of the label vector $\underline{\Lambda}$. Here and subsequently, the index "0" indicates the null hypothesis, i.e. any couple (p, q) with $p = q$. Precisely, let the random variable Π be uniformly distributed on the symmetric group $\mathcal{S}_n$ of order n, independent of $\underline{\mathrm{X}}$. Then $\mathcal{L}_0\big(T_n \,\big|\, \mathcal{X}_n\big) \;=\; \mathcal{L}\big(T_n \circ \Pi \,\big|\, \mathcal{X}_n\big)$, where $\big(T_n \circ \Pi\big)(\underline{\Lambda}) := T_n\big(\Lambda_{\Pi_1}, ..., \Lambda_{\Pi_n}\big)$. Elementary calculation entails that

$$\mathbb{E}\Big(T_{jkn} \circ \Pi \,\Big|\, \mathcal{X}_n\Big) \;=\; 0 \quad \text{and} \quad \mathrm{Var}\Big(T_{jkn} \circ \Pi \,\Big|\, \mathcal{X}_n\Big) \;=\; 1.$$

Thus the null hypothesis is satisfied if, and only if, the hypothesis of permutation invariance (or complete randomness) conditional on $\mathcal{X}_n$ is satisfied.

An adequate calibration of the randomized nearest-neighbor statistics, i.e. the choice of smallest possible constants C_{jkn}, requires both, an exact understanding of their tail behavior and their dependency structure. Note that the randomized nearest-neighbor statistics have a geometrically involved dependency structure. Even in case of the rectangular kernel ψ it depends explicitly on the "random design" $\mathcal{X}_n$ which incomplicates the sharp-optimal calibration for multiple testing compared to univariate problems, where the dependency of the single test statistics remains typically invariant under monotone transformation of the design points. Also, the optimal correction originally designed for Gaussian tails in Dümbgen and Spokoiny (2001) does not carry over as only the subsequent Bernstein type exponential tail bound is available.

A coupling exponential inequality Based on an explicit coupling, the following proposition remarkably tightens the exponential bounds derived in Serfling (1974) in the present framework. If not stated otherwise, the random variable Π is uniformly distributed on $\mathcal{S}_n$, independent of $\underline{\mathrm{X}}$.

Proposition 1. *Let T_{jkn} be as introduced above and define*

$$\delta(m,n) \ := \ \Big(\mathbb{E}\min\Big(\frac{S}{m}, \frac{n-S}{n-m}\Big)\Big)^{-1} \quad \textit{with} \quad S \sim \mathrm{Bin}\big(n, m/n\big).$$

Then

$$\mathbb{P}\Big(\big| T_{jkn} \circ \Pi \big| \ > \ \delta(m,n)\eta \,\Big|\, \mathcal{X}_n \Big) \ \leq \ 2\, \exp\left(- \frac{\eta^2/2}{1 + \eta\, n^{-1/2}\gamma_{jkn}^{-1}\, R_\psi(m,n)} \right),$$

where

$$R_\psi(m,n) \ := \ \frac{\|\psi\|_{\sup}}{3} \frac{\max(m, n-m)}{\sqrt{m(n-m)}}.$$

REMARK The expression $\delta(m,n)$ is the payment for decoupling which appears by replacing the tail probability of an hypergeometric ensemble by that of the Binomial analogon. For details we refer to Section 6. In the typical case $0 < \liminf_n(m/n) \leq \limsup_n(m/n) < 1$ we obtain $\delta(m,n) = 1 + O(n^{-1/2})$. Compared to results obtained for weighted averages of standardized, independent Bernoulli's, the above Bernstein type appears to be nearly optimal, i.e. subgaussian tail behavior is actually not present.

Via inversion of the above exponential inequality, additive correction terms C_{jkn} for adjustment of multiple testing are constructed. The next Theorem motivates our approach. The construction is designed for typical arrangements of the observation values which appear with probability close to one. To avoid technical expenditure, we restrict our attention to compactly supported densities. d_w denotes the dual bounded Lipschitz metric which generates the topology of weak convergence. "$\rightarrow_{\mathbb{P}_n}$" refers to convergence in probability along the sequence of distributions $(\mathbb{P}_n)$.

Theorem 1. *Define the test statistic*

$$T_n \ := \ \sup_{1 \leq j,k \leq n} \Big\{ |T_{jkn}| \ - \ C_{jkn} \Big\}$$

with

$$C_{jkn} \ := \ 3\, R_n \gamma_{jkn}^{-1} \delta(m,n) \Gamma_{jkn} \ + \ \delta(m,n) \sqrt{2\,\Gamma_{jkn}},$$

where $R_n = n^{-1/2} R_\psi(m,n)$ *and* $\Gamma_{jkn} := \log\big(1/\gamma_{jkn}{}^2\big)$. *Assume that the sequence of mixed densities* $h_n := (m/n)p_n + (1-m/n)q_n$ *on* $[0,1]^d$ *is equicontinuous and uniformly bounded away from zero, while* $0 < \liminf_n m/n \leq \limsup_n m/n < 1$. *Then the sequence* $\mathcal{L}\big(T_n \circ \Pi \,\big|\, \mathcal{X}_n\big)$ *of conditional distributions is tight in* $\big(\mathbb{P}_n^{\otimes m} \otimes \mathbb{Q}_n^{\otimes(n-m)}\big)$-*probability. Additionally,*

$$d_w\Big(\mathcal{L}\big(T_n \circ \Pi \,\big|\, \mathcal{X}_n\big),\, \mathcal{L}(T_{\mathbb{H}_n})\Big) \;\longrightarrow_{\mathbb{P}_n^{\otimes m} \otimes \mathbb{Q}_n^{\otimes(n-m)}}\; 0,$$

where

$$T_{\mathbb{H}_n} \;:=\; \sup_{\substack{t \in [0,1]^d, \\ 0 < r \leq \max\limits_{x \in [0,1]^d} \|x-t\|_2}} \left\{ \frac{\big|\int_{[0,1]^d} \phi_{rt,n}(x)\, dW(x)\big|}{\gamma_{rt,n}} \;-\; \sqrt{2\log\big(1/\gamma_{rt,n}{}^2\big)} \right\}$$

with W *a standard Brownian sheet in* $[0,1]^d$, $\gamma_{rt,n} := \big(\int_{[0,1]^d} \phi_{rt,n}(x)^2 dx\big)^{1/2}$ *and*

$$\phi_{rt,n}(x) \;:=\; \left[\psi\Big(\frac{\|x-t\|_2}{r}\Big) \;-\; \int_{[0,1]^d} \psi\Big(\frac{\|z-t\|_2}{r}\Big) h_n(z)dz\right] \sqrt{h_n(x)}.$$

The extra-term $3\,R_n \gamma_{jkn}^{-1} \delta(m,n) \Gamma_{jkn}$ in the constant C_{jkn} results from the exponential inequality in Proposition 1 and can be viewed as an additional penalty for non-subgaussianity. The theorem entails in particular that the sequence $\mathcal{L}(T_n \circ \Pi \,|\mathcal{X}_n)$ is weakly approximated in probability by a tight sequence of *non-degenerate* distributions $\mathcal{L}(T_{\mathbb{H}_n})$ which indicates that our corrections C_{jkn} are appropriately defined and cannot be chosen essentially smaller. Note that the approximation $\mathcal{L}(T_{\mathbb{H}_n})$ depends on the unknown mixed distribution even under the null hypothesis. For non-compactly supported densities, the tightness may be shown using the coordinatewise quantile transformation (which however does not preserve the geometry) before applying the techniques of the proof for the compact case.

2.3 The multiple rerandomization test

Let $\kappa_\alpha(\underline{\mathrm{X}}) := \mathrm{argmin}_{C>0} \big\{\mathbb{P}\big(T_n \circ \Pi \leq C \big|\, \mathcal{X}_n\big) \geq 1-\alpha\big\}$ denote the generalized $(1-\alpha)$-quantile of $\mathcal{L}\big(T_n \circ \Pi \,\big|\, \mathcal{X}_n\big)$. Then we propose the conditional test

$$\phi_\alpha(\underline{\mathrm{X}}) \;:=\; \begin{cases} 0 & \text{if } T_n \leq \kappa_\alpha(\underline{\mathrm{X}}) \\ 1 & \text{if } T_n > \kappa_\alpha(\underline{\mathrm{X}}). \end{cases}$$

Our method can be viewed as a multiple testing procedure. For a given set of observations $\{X_1, ..., X_n\}$, the corresponding test statistic exceeds the $(1-\alpha)$-quantile if, and only if, the random set

$$\mathcal{D}_\alpha \;:=\; \Big\{ B_{X_j}\big(\|X_j^k - X_j\|_2\big) \;\Big|\; 1 \leq j,k \leq n;\; T_{jkn}(\underline{\mathrm{X}}) \;>\; C_{jkn}(\underline{\mathrm{X}}) \;+\; \kappa_\alpha(\underline{\mathrm{X}}) \Big\}$$

is nonempty, where $B_t(r)$ denotes the Euclidean ball in $\mathbb{R}^d$ with center t and radius r. Since the test is valid conditional on the set of observations, we may conclude that p

deviates from q at significance level α on *every* Euclidean ball $B_t(r) \in \mathcal{D}_\alpha$. In order to reduce the computational expenditure and to increase sensitivity on smaller scales, one may restrict one's attention to pairs (j, k) for $k \leq m$ for some integer $m \in \{1, ..., n\}$. Note the validity of the test does not require any assumption about the densities - even not Lebesgue continuity.

3 Minimax-efficiency and spatial adaptivity

In this section we show that the above introduced multiple testing procedure possesses optimality properties in a certain minimax sense. Let us first introduce some notation. For any set $J \subset [0,1]^d$ and function f from $[0,1]^d \to \mathbb{R}$, $\|f\|_J := \sup_{x\in J} |f(x)|$. For any pair of densities (p, q) on $[0,1]^d$, let $h(m, n, p, q)$ denote the corresponding mixed density $(m/n)p + (1 - m/n)q$. Fix a continuous density $h > 0$ and define $\mathcal{F}_h^{(m,n)}(\beta, L)$ to be the set of pairs of densities such that

$$\phi(m, n, p, q) \; := \; \frac{p - q}{\sqrt{h(m, n, p, q)}} \; \in \; \mathcal{H}_d\big(\beta, L; [0,1]^d\big) \; \text{ and } \; h(m, n, p, q) = h.$$

For any convex $I \subset \mathbb{R}^d$ let $\mathcal{H}_d(\beta, L; I)$ denote the isotropic Hölder smoothness class, which for $\beta \leq 1$ equals

$$\mathcal{H}_d(\beta, L; I) \; := \; \Big\{ \phi : I \to \mathbb{R} : \; \big|\phi(x) - \phi(y)\big| \; \leq L\|x - y\|_2^\beta \Big\}.$$

Let $\lfloor\beta\rfloor$ denote the largest integer strictly smaller than β. For $\beta > 1$, $\mathcal{H}_d(\beta, L; I)$ consists of all functions $f : I \to \mathbb{R}$ that are $\lfloor\beta\rfloor$ times continuously differentiable such that the following property is satisfied: if $P_y^{(f)}$ denotes the Taylor polynomial of f at the point $y \in I$ up to the $\lfloor\beta\rfloor$'th order,

$$\Big| f(x) - P_y^{(f)}(x) \Big| \;\; \leq \;\; L\|x - y\|_2^\beta \; \text{ for all } x, y \in I.$$

In particular the definition entails that $f \in \mathcal{H}_d(\beta, L; \mathbb{R}^d)$ implies $f \circ U \in \mathcal{H}_d(\beta, L; \mathbb{R}^d)$ for every orthonormal transformation $U \in \mathbb{R}^{d\times d}$.

Reparametrizing the composite hypothesis With the notation above,

$$p = h \cdot \Big(1 + (1 - m/n)\,\phi/\sqrt{h}\,\Big) \; \text{ and } \; q = h \cdot \Big(1 - (m/n)\,\phi/\sqrt{h}\,\Big).$$

Consequently "$p = q$" is equivalent to "$\phi \equiv 0$", and if $(m/n)p + (1 - m/n)q = h$ is kept fixed, the composite hypothesis "$p = q$" reduces to the simple hypothesis "$\phi \equiv 0$". In order to develop a meaningful notion of minimax-efficiency for the two-sample problem we treat subsequently the mixed density $h = h(m, n, p, q)$ as fixed but unknown infinite dimensional nuisance parameter for testing the hypothesis

$$H_0 : \; \phi = 0 \quad \text{versus} \quad H_A : \; \phi \neq 0.$$

Note that in case that h is uniformly bounded away from zero and p is close to q, ϕ coincides approximately with the difference $2\big(\sqrt{p} - \sqrt{q}\big)$, see also the explanation subsequent to Theorem 2.

REMARK It is worth being noticed that the optimal statistic for testing H_0 against any fixed alternative ϕ equals the likelihood ratio statistic

$$\frac{d\mathbb{P}_{(m,n,p,q)}}{d\mathbb{P}_{(m,n,h,h)}}(\underline{\mathrm{X}}) \;=\; \prod_{i=1}^{m}\Big(1-(m/n)\frac{\phi}{\sqrt{h}}(X_i)\Big)\prod_{j=m+1}^{n}\Big(1+(1-m/n)\frac{\phi}{\sqrt{h}}(X_j)\Big),$$

whose distribution still depends on h under the null. Here and subsequently, the subscript (m,n,p,q) indicates the distribution with density $\prod_{i=1}^{m} p \prod_{i=m+1}^{n} q$. The rational behind the reparametrization is to eliminate the dependency on the nuisance parameter h in the expectation under the null of the first and second order term of the $\log$-likelihood expansion, resulting in asymptotic independence of h for its distribution under the hypothesis for any local sequence (ϕ_n).

Theorem 2 (Minimax lower bound)**.** *Let*

$$\rho_{m,n} \;:=\; \Big(\frac{n\log n}{m(n-m)}\Big)^{\beta/(2\beta+d)} \quad \textit{and define} \quad c(\beta,L) \;:=\; \Bigg(\frac{2\,d\,L^{d/\beta}}{(2\beta+d)\|\gamma_\beta\|_2^2}\Bigg)^{\beta/(2\beta+d)},$$

where γ_β defines the solution to the optimal recovery problem (2) below. Assume that the sequence of mixed densities (h_n) on $[0,1]^d$ is equicontinuous and uniformly bounded away from zero. Then for any fixed $\delta > 0$ and every nondegenerate rectangle $J \subset [0,1]^d$,

$$\limsup_{n\to\infty} \inf_{\substack{(p,q)\in\mathcal{F}_{h_n}^{(m,n)}(\beta,L):\\ \|\phi\|_J\geq(1-\delta)c(\beta,L)\rho_{m,n}}} \mathbb{E}_{(m,n,p,q)}\,\psi_n \;\leq\; \alpha$$

for arbitrary tests ψ_n at significance level $\leq \alpha$.

Note that ψ_n may depend on (β, L) and even on the nuisance parameter h_n as already does the Neyman-Pearson test for testing H_0 against any one-point alternative.

We now turn to the investigation of the test introduced in Section 2. To motivate the choice of an optimal kernel for our test statistics and its relation to the optimal recovery problem, let us restrict our consideration to the Gaussian white noise context, leading in case of univariate Hölder continuous densities on $[0,1]$ with $\beta > 1/2$ to locally asymptotically equivalent experiments

$$dX_{1n}(t) = p_n(t)\,dt + \frac{\sqrt{h_n(t)}}{\sqrt{m}}\,dW_1(t) \quad\text{and}\quad dX_{2n}(t) = q_n(t)\,dt + \frac{\sqrt{h_n(t)}}{\sqrt{(n-m)}}\,dW_2(t)$$

for two independent Brownian motions W_1 and W_2 on the unit interval (Nussbaum 1996, Theorem 2.7 with $f_0 = h_n$ and Remark 2.8). A multiscale statistic built from standardized differences of kernel estimates

$$\frac{\sqrt{(m/n)(1-m/n)}}{\|\psi\sqrt{h_n}\|_2}\int \psi(t)\Big(dX_{1n}(t) - dX_{2n}(t)\Big)$$

(which is actually not admissible since h_n is unknown in general) then yields a distribution under the null close to ours in Theorem 1, up to the fact that our local integrals in dimension one are taken with respect to a Brownian bridge, reformulated to a Wiener process integrand by change of the kernel. Concerning the optimization of ψ, the quantity to be maximized within this Gaussian white noise context appears to be the expectation of the single test statistics under the least favorable alternatives as their variances do not depend on the mean. In case $h_n \equiv 1$ this expression equals

$$\inf_{\substack{\phi \in \mathcal{H}_1(\beta, L; [0,1]): \\ \|\phi\|_J \geq \delta}} \frac{\int \phi(t)\, \psi(t)\, dt}{\|\psi\|_2},$$

leading to the dual representation of the optimal recovery problem (see Donoho 1994a).

The optimal recovery problem in higher dimension In the framework of isotropic Hölder balls, the optimal recovery problem leads to the solution $\gamma = \gamma_\beta$ of the optimization problem

$$\text{Minimize } \|\gamma\|_2 \text{ over all } \gamma \in \mathcal{H}_d\big(\beta, 1; \mathbb{R}^d\big) \text{ with } \gamma(0) \geq 1. \tag{2}$$

The closedness of $\mathcal{H}_d(\beta, L; \mathbb{R}^d) \cap \big\{\gamma : \mathbb{R}^d \to \mathbb{R} \,\big|\, \gamma(0) \geq 1\big\}$ in L_2 entails that the solution exists, its convexity implies furthermore uniqueness whence by isotropy of the functional class $\mathcal{H}_d(\beta, 1; \mathbb{R}^d)$ it must be radially symmetric. In case $\beta \leq 1$, one easily verifies that $\gamma_\beta(x) = \psi_\beta\big(\|x\|_2\big) = \big(1 - \|x\|_2^\beta\big)_+$. In its generality, the optimal recovery problem in higher dimension has not yet been investigated. Considering the partial derivatives of γ_β at the origin entails that ψ_β is necessarily contained in $\mathcal{H}_1(\beta, L; \mathbb{R})$. However, the transferred optimization problem

$$\text{minimize} \int \psi(r)^2 |r|^{d-1} dr \ \text{ over all } \psi \text{ with } \psi\big(\|.\|_2\big) \in \mathcal{H}_d(\beta, 1; \mathbb{R}) \text{ and } \psi(0) \geq 1 \tag{3}$$

does not coincide with the univariate optimal recovery problem due to the additional weighting by $|r|^{d-1}$ which comes into play by polar coordinate transformation. Whether the solution of (3) for $\beta > 1$ is compactly supported or not is still open. For the case of univariate densities, it is known that the solution of the optimal recovery problem has compact support for any $\beta > 0$ (Leonov 1997), but an explicit solution in case $\beta > 1$ is known for $\beta = 2$ only. Concerning details and advice on its construction, see Donoho (1994b) and Leonov (1999).

The next Theorem is about the asymptotic power of the multiple test developed in Section 2. We restrict our attention to compact rectangles of $(0,1)^d$ to avoid boundary effects. This restriction may be relaxed by the use of suitable boundary kernels, extending those of Lepski and Tsybakov (2000) for the univariate regression case to higher dimension.

Theorem 3 (Adaptivity and minimax efficiency). *Let $\phi^*_{n,\alpha}$ denote the multiple rerandomization test at significance level α, based on the kernel $\psi_\beta I\{\cdot \geq 0\}$ rescaled to $[0,1]$. In case of unbounded support of ψ_β, we may use a truncated solution $\psi_{\beta,K} = \psi_\beta I\{0 \leq \cdot \leq K\}$. Let $0 < \liminf_n m/n \leq \limsup_n m/n < 1$. Assume that (h_n) is equicontinuous* *and uniformly bounded away from zero. Then for any fixed $\delta > 0$, there exists a $K > 0$ such that*

$$\liminf_{n\to\infty} \inf_{\substack{(p,q)\in\mathcal{F}_{h_n}^{(m,n)}(\beta,L):\\ \|\phi\|_J \geq (1+\delta)c(\beta,L)\rho_{m,n}}} \mathbb{P}_{(m,n,p,q)}\big(\phi^*_{n,\alpha} = 1\big) \;=\; 1$$

for any nondegenerate compact rectangle $J \subset (0,1)^d$.

Note at this point that in its origin, the question of optimal adjustment for multiple testing is connected to a fixed choice of local test statistics and does not involve any optimality considerations concerning the kernel. Theorem 3 shows however that the use of adequately chosen kernels in the local test statistics even leads to sharp-optimality in the above introduced minimax sense - which in retrospect shows optimality of the calibration for the multiple test with respect to $\sup$-norm loss.

REMARK It is worth being noticed that the procedure achieves the upper bound uniformly over a large class of possible mixed densities. The intrinsic reason is that conditioning on $\mathcal{X}_n$ is actually equivalent to conditioning on $\hat{\mathbb{H}}_n$, which indeed is a sufficient and complete statistic for the nuisance functional $\mathbb{H}_n$.

REMARK (Sharp adaptivity with respect to β and L) Our construction, including the procedure especially designed for the one-dimensional situation, involves one kernel, shifted and rescaled depending on location and volume of the nearest-neighbor cluster under consideration. Due to the dependency of the optimal recovery solution γ_β on β, the corresponding test statistic $T_n = T_n(\beta)$ achieves sharp adaptivity with respect to the second Hölder parameter L only. Taking in addition the supremum $\sup_{\beta\in[\beta_0,\beta_1]} T_n(\beta)$ over all kernels γ_β within a compact range $[\beta_0,\beta_1] \subset (0,\infty)$, one may check the proof of Theorem 3 to verify that sharp adaptivity with respect to both Hölder parameters can be attained, provided that the above supremum statistic still defines a tight sequence. We however omit the investigation to avoid the technical expenditure as the result is rather of theoretical interest than of practical relevance.

The next theorem shows however that our procedure simply based on the rectangular kernel is rate-adaptive with respect to both Hölder parameters (β, L). Due to the fact that it combines locally all nearest-neighbor scales at the same time, it even adapts to inhomogeneous smoothness of $p - q$, i.e. achieves *spatial adaptivity*.

Theorem 4 (Spatial rate-optimality). *Let $\phi^*_{n,\alpha}$ denote the multiple rerandomization test based on the rectangular kernel. Assume that $0 < \liminf_n m/n \leq \limsup_n m/n < 1$. Then for any fixed $k \in \mathbb{N}$ and parameters $(\beta_1,...,\beta_k,L_1,...,L_k)$, $K > 0$ and any collection of disjoint compact rectangles $J_i \subset [0,1]^d$, $i = 1,...,k$, there exist constants $d_i = d(\beta_i, L_i, K)$ with*

$$\liminf_{n\to\infty} \inf_{\substack{(p,q):\\ (p-q)_{|J_i}\in\mathcal{H}_d\left(\beta_i,L_i;J_i\right)\\ \|p-q\|_{J_i}\geq d_i\,\rho_{m,n}(\beta_i),\\ h(m,n,p,q)_{|J_i}\geq K}} \mathbb{P}_{(m,n,p,q)}\bigg(J_i \cap \mathcal{D}_\alpha(\mathcal{X}_n) \neq \emptyset \;\forall\, i = 1,...,k\bigg) \;=\; 1.$$

4 Application to classification

Suppose we are given an iid sample (X_i, Y_i), $i = 1, ..., n$, where the marginal distribution of X_i is assumed to be Lebesgue-continuous with density h on $\mathbb{R}^d$, and Y_i takes values in $\{0, 1\}$ with

$$\mathbb{P}\Big(Y_i = 1 \,\Big|\, X_i = x\Big) \;=\; \rho(x).$$

Then $M := \sum_{i=1}^n Y_i \;\sim\; \mathrm{Bin}\big(n, \lambda\big)$ with $\lambda := \int \rho(x)h(x)dx$. Assuming $\lambda \in (0, 1)$ to be known, the question of local classification is to identify simultaneously subregions in $\mathbb{R}^d$ where ρ deviates sifnificantly from λ which results in local testing the hypotheses

$$H_0 : \rho \;=\; \lambda \quad \text{versus} \quad H_A : \rho \;\neq\; \lambda.$$

Imitating our procedure introduced in Section 2, we may combine suitably standardized local weighted averages of labels, but the standardization differs due to the fact that the sum of (strictly) positive labels is random and not fixed, in particular $Y_1, ..., Y_n$ are stochastically independent. Consequently, we may then rely the procedure on the classical Bernstein exponential inequality for weighted averages of standardized Bernoullis. Of course, the optimal separation constant for testing "$\rho = \lambda$" within some Euclidean ball $B_t(r)$ and its complement depends on the amount of observations in $B_t(r)$, whence analogously to the consideration above for the two-sample problem we may use the reparametrization of (ρ, h) to (ϕ, h) with

$$\phi \;:=\; \frac{\rho - \lambda}{\lambda(1 - \lambda)}\sqrt{h}.$$

The power optimality results carry over to the classification context with similar arguments as used in the proof of Theorem 3. We omit its explicit formulation at this point.

5 Distribution-freeness via quantile transformation – the case d=1

The one-dimensional situation allows an alternative and more elegant approach based on order relations. For let $X_{(1)}, ..., X_{(n)}$ denote the order statistic built from the pooled sample and define for any $0 \leq j < k \leq n$ the local test statistics

$$U_{jkn} := \frac{\sqrt{(m/n)(1 - m/n)}}{\eta_{jkn}} \frac{1}{\sqrt{n}} \sum_{i=j+1}^{k} \psi\Big(\frac{i}{k - j}\Big)\Lambda(X_{(i)}),$$

where

$${\eta_{jkn}}^2 \;:=\; \frac{1}{n-1}\sum_{i=1}^{n}\Big(\psi\Big(\frac{i-j}{k-j}\Big) - \frac{1}{n}\sum_{l=1}^{n}\psi\Big(\frac{l-j}{k-j}\Big)\Big)^2.$$

Compared to the procedure described in the previous section, we omit the explicit dependence of the weights on the observed values. Note that in contrast to nearest-neighbor relations, the order remains invariant under quantile transformation, i.e. $\text{rank}(H_n(X_i)) = \text{rank}(X_i)$, resulting in distribution-freeness of the corresponding multiscale statistic under the null. Suppose the null hypothesis is satisfied for some Lebesgue continuous distribution on the real line. Then conditional on the order statistics as well as unconditional, the label vector is uniformly distributed on the set

$$\Big\{\Lambda \in \big\{n/m, -n/(n-m)\big\}^n : \sum_{i=1}^n \Lambda_i^{-1} = 0\Big\}.$$

The described test statistics are local versions of classical Wilcoxon rank sum statistics. We omit any further investigation as the calibration for multiple testing can be done analogously to that proved in Theorem 1 – but keep in mind that the approximating Gaussian multiscale statistic under the null hypothesis will be independent of the nuisance functional $\mathbb{H}_n$ due to the quantile transformation. Note that the use of typical mathematical tools for power investigation of rank statistics like Hoeffding's decomposition is getting involved because the kernel ψ_β for $\beta \le 1$ is not differentiable.

6 Decoupling inequality and coupling exponential bounds

This section contains the coupling exponential bounds, i.e. in this context for weighted averages from a hypergeometric ensemble. Using a different technique, namely an explicit coupling construction, the subsequent proposition extends results of Hoeffding (1963) on decoupling of expectations of convex functions in the arithmetic mean of a sample without replacement. Whereas in the latter case decoupling with constant 1 is actually correct, a simple counterexample for an ensemble of two elements already shows that the result does not extend to arbitrary weighted averages, and some payment for decoupling appears to be necessary.

Proposition 2 (Decoupling inequality)**.** *Let $Z_1, Z_2, ..., Z_n$ be iid with*

$$\mathbb{P}\big(Z_i = 1\big) \;=\; \frac{m}{n} \quad \textit{and} \quad \mathbb{P}\big(Z_i = 0\big) \;=\; 1 - \frac{m}{n}, \quad 0 < m < n.$$

Let $a \in \mathbb{R}^n$ with $\sum_{i=1}^n a_i = 0$ and $\Psi : \mathbb{R} \to \mathbb{R}$ be convex. Then

$$\mathbb{E}\bigg(\Psi\Big(\sum_{i=1}^n a_i Z_i\Big) \,\bigg|\, \sum_{i=1}^n Z_i = m\bigg) \;\le\; \mathbb{E}\,\Psi\Big(\delta(m,n)\sum_{i=1}^n a_i Z_i\Big),$$

with

$$\delta(m,n)^{-1} \;:=\; \mathbb{E}\min\Big(\frac{S}{m}, \frac{n-S}{n-m}\Big), \quad S \sim \mathrm{Bin}\Big(n, \frac{m}{n}\Big).$$

In particular, $\delta(m,n)^{-1} = 1 + O(n^{-1/2})$ for $m/n \to \lambda \in (0,1)$.

PROOF Let X be uniformly distributed on the set

$$\Big\{x \in \{0,1\}^n : \sum_{i=1}^n x_i \ = \ m\Big\}$$

and let $S \sim \mathrm{Bin}(n, m/n)$ such that X and S are independent. Define

$$M \ := \ \big\{i : \ X_i \ = \ 1\big\}.$$

Conditional on X and S, the random vector $Z \in \{0,1\}^n$ is constructed as follows:

If $S > m$, let $Z_i = 1$ for all $i \in M$ and let $(Z_i)_{i \in M^c}$ be uniformly distributed on the set

$$\Big\{z \in \{0,1\}^{M^c} : \sum_{i \in M^c} z_i = S - m\Big\}.$$

For $S \leq m$, let $Z_i = 0$ for all $i \in M^c$ and let $(Z_i)_{i \in M}$ be uniformly distributed on

$$\Big\{z \in \{0,1\}^M : \sum_{i \in M} z_i = S\Big\}.$$

Note that $Z_1, ..., Z_n$ are iid $\mathrm{Bin}(1, m/n)$. Then

$$\begin{aligned}
\mathbb{E}\,\Psi\Big(\sum_{i=1}^n a_i Z_i\Big) \ &= \ \mathbb{E}\mathbb{E}\Big(\Psi\Big(\sum_{i=1}^n a_i Z_i\Big)\,\Big|\, X, S\Big) \\
&\geq \ \mathbb{E}\,\Psi\Big(\mathbb{E}\Big(\sum_{i=1}^n a_i Z_i \,\Big|\, X, S\Big)\Big) \qquad \text{(Jensen inequality)} \\
&= \ \mathbb{E}\,\Psi\Big(I\{S \leq m\}\frac{S}{m}\sum_{i \in M} a_i \ + \ I\{S > m\}\Big(\sum_{i \in M} a_i + \frac{S-m}{n-m}\sum_{i \in M^c} a_i\Big)\Big) \\
&= \ \mathbb{E}\,\Psi\Big(I\{S \leq m\}\frac{S}{m}\sum_{i \in M} a_i \ + \ I\{S > m\}\frac{n-S}{n-m}\sum_{i \in M} a_i\Big) \qquad \big(\sum_{i=1}^n a_i \ = \ 0\big) \\
&= \ \mathbb{E}\,\Psi\Big(\min\Big(\frac{S}{m}, \frac{n-S}{n-m}\Big)\sum_{i=1}^n a_i X_i\Big) \\
&= \ \mathbb{E}\mathbb{E}\Big[\Psi\Big(\min\Big(\frac{S}{m}, \frac{n-S}{n-m}\Big)\sum_{i=1}^n a_i X_i\Big)\,\Big|\, X\Big] \\
&\geq \ \mathbb{E}\,\Psi\Big(\mathbb{E}\Big\{\min\Big(\frac{S}{m}, \frac{n-S}{n-m}\Big)\Big\}\sum_{i=1}^n a_i X_i\Big) \qquad \text{(Jensen inequality).}
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\mathbb{E}\,\min\Big(\frac{S}{m}, \frac{n-S}{n-m}\Big) \ &= \ 1 - \mathbb{E}\Big(\frac{(S-m)_-}{m} + \frac{(S-m)_+}{n-m}\Big) \\
&\geq \ 1 - \mathbb{E}\Big(\frac{|S-m|}{\min(m, n-m)}\Big) \\
&\geq \ 1 - \frac{\lambda(m,n)}{\sqrt{n}}
\end{aligned}$$

with $\lambda(m,n) := \sqrt{m(n-m)}/\min(m,n-m)$, which is uniformly bounded for $m/n \to \lambda \in (0,1)$. □

Using the decoupling above, the next proposition presents the exponential bounds which are essential for our construction. It implies Proposition 1 in particular and remarkably tightens related exponential bounds of Serfling (1974) for the present context. The results may also be compared with the decoupling based exponential tail bounds in de la Peña (1994, 1999).

Proposition 3 (Coupling exponential inequalities)**.** *Let $Z_1, ..., Z_n$ be iid with*

$$\mathbb{P}\big(Z_i = 1\big) \;=\; \frac{m}{n} \quad \textit{and} \quad \mathbb{P}\big(Z_i = 0\big) \;=\; 1 - \frac{m}{n}, \;\; 0 < m < n.$$

Let $\psi_1, ..., \psi_n$ real valued numbers with $\bar{\psi}$ its arithmetic mean and denote

$$\gamma_{m,n}{}^2 \;:=\; \operatorname{Var}\bigg(\sum_{i=1}^n \psi_i Z_i \,\bigg|\, \sum_{i=1}^n Z_i = m\bigg) \;=\; \frac{m(n-m)}{n(n-1)} \sum_{i=1}^n \big(\psi_i - \bar{\psi}\big)^2.$$

Then in case of $\gamma_{m,n} \neq 0$,

$$\begin{aligned}\mathbb{P}\left(\left|\frac{1}{\gamma_{m,n}}\sum_{i=1}^n \psi_i\Big(Z_i - \frac{m}{n}\Big)\right| > \delta(m,n)\eta \,\middle|\, \sum_{i=1}^n Z_i = m\right) \;&\leq\; 2\,\exp\left(-\frac{\eta^2/2}{1 \;+\; \eta\, R(\psi,m,n)}\right)\\ &\leq\; 2\,\exp\left(-\frac{3\eta}{2c(m,n)} + \frac{9}{2c(m,n)^2}\right),\end{aligned}$$

where

$$R(\psi,m,n) \;:=\; \frac{\max_i |\psi_i - \bar{\psi}|}{3\,\gamma_{m,n}} \max\Big(\frac{m}{n},\, 1 - \frac{m}{n}\Big) \quad \textit{and} \quad c(m,n) \;:=\; \frac{\max(m, n-m)}{\sqrt{m(n-m)}}.$$

PROOF With

$$M \;:=\; \frac{\max_i |\psi_i - \bar{\psi}|}{\gamma_{m,n}} \max\Big(\frac{m}{n},\, 1 - \frac{m}{n}\Big)$$

we obtain for any $t > 0$

$$\begin{aligned}&\mathbb{P}\left(\frac{1}{\gamma_{m,n}}\sum_{i=1}^n \psi_i\Big(Z_i - \frac{m}{n}\Big) \;>\; \delta(m,n)\eta \,\middle|\, \sum_{i=1}^n Z_i = m\right)\\
&\quad=\; \mathbb{P}\left(\frac{1}{\gamma_{m,n}}\sum_{i=1}^n \big(\psi_i - \bar{\psi}\big)\Big(Z_i - \frac{m}{n}\Big) \;>\; \delta(m,n)\eta \,\middle|\, \sum_{i=1}^n Z_i = m\right)\\
&\quad\leq\; \exp\Big(-t\,\frac{\eta}{M}\Big)\,\mathbb{E}\left\{\exp\left(\frac{t\,\delta(m,n)^{-1}}{M\,\gamma_{m,n}}\sum_{i=1}^n \big(\psi_i - \bar{\psi}\big)\Big(Z_i - \frac{m}{n}\Big)\right) \,\middle|\, \sum_{i=1}^n Z_i = m\right\}\\
&\quad\leq\; \exp\Big(-t\,\frac{\eta}{M}\Big)\,\mathbb{E}\exp\left(\frac{t}{M\,\gamma_{m,n}}\sum_{i=1}^n \big(\psi_i - \bar{\psi}\big)\Big(Z_i - \frac{m}{n}\Big)\right) \quad \text{(Proposition 2)}\\
&\quad\leq\; \exp\left(\frac{1}{M^2}\big(e^t - 1 - t\big) \;-\; t\,\frac{\eta}{M}\right), \qquad (4)\end{aligned}$$

whereby the last inequality follows from the fact that for any random variable Y with $|Y| \leq 1$, $\mathbb{E}Y = 0$ and $\mathrm{Var}(Y) = \sigma^2$,

$$\mathbb{E}\exp(tY) \ \leq \ 1 + \sigma^2(e^t - 1 - t) \ \leq \ \exp\Big(\sigma^2(e^t - 1 - t)\Big).$$

Elementary algebra shows that (4) is minimized with the choice $t := \log\big(1+\eta M\big)$, which yields first a Bennett (1962) exponential bound by Chebychef's inequality and because of $(1 + x)\log(1 + x) - 1 \geq 1/(1 + x/3)$ consequently the Bernstein type

$$\mathbb{P}\left(\frac{1}{\gamma_{m,n}}\sum_{i=1}^{n}\psi_i\Big(Z_i - \frac{m}{n}\Big) \ > \ \delta(m,n)\eta \ \middle| \ \sum_{i=1}^{n} Z_i = m\right) \ \leq \ \exp\left(-\frac{\eta^2/2}{1 + \eta\, M/3}\right).$$

A symmetry argument provides the same bound for ψ_i replaced by $-\psi_i$, which completes the proof of the first inequality. Using that $\gamma_{m,n} \geq \sqrt{(m/n)(1 - m/n)}\max_i|\psi_i - \bar{\psi}|$, we obtain the second asserted inequality from

$$\begin{aligned}
\frac{\eta^2/2}{1 + \eta\, M/3} \ &\geq \ \frac{\eta^2/2}{1 + \eta\, c(m,n)/3} \\
&= \ \frac{\eta}{2c(m,n)/3} \ - \ \frac{\eta}{2c(m,n)/3(1 + \eta\, c(m,n)/3)} \\
&\geq \ \frac{\eta}{2c(m,n)/3} \ - \ \frac{1}{2c(m,n)^2/9}.
\end{aligned}$$

□

7 Auxiliary results about empirical processes

This section collects results in the context of empirical processes which are essential for the next section. For any totally-bounded pseudo-metric space $(\mathcal{T}, \rho)$, we define the covering number

$$N\big(\varepsilon, \mathcal{T}, \rho\big) \ := \ \min\Big\{\sharp\mathcal{T}_0 : \ \mathcal{T}_0 \subset \mathcal{T}, \inf_{t_0 \in \mathcal{T}_0}\rho(t, t_0) \leq \varepsilon \text{ for all } t \in \mathcal{T}\Big\}.$$

Let $\mathcal{F} \subset [0,1]^{\mathcal{T}}$. For any probability measure P on $\mathcal{T}$, consider the pseudo-distance $d_P(f,g)^2 := \int(f - g)^2\, dP$ for $f, g \in \mathcal{F}$. Then the uniform covering numbers of $\mathcal{F}$ are defined as

$$\mathcal{N}(u, \mathcal{F}) := \sup_P N(u, \mathcal{F}, d_P)$$

for $u > 0$, where the supremum is running over all probability measures P on $\mathcal{T}$.

Theorem 5. (Dümbgen and Walther (2008, technical report)) *Let $Z = (Z(t))_{t\in\mathcal{T}}$ be a stochastic process on a totally bounded pseudo-metric space $(\mathcal{T}, \rho)$. Let K be some positive constant, and for $\delta > 0$ let $G(\cdot, \delta)$ a nondecreasing function on $[0, \infty)$ such that for all $\eta \geq 0$ and $s, t \in \mathcal{T}$,*

$$\mathbb{P}\Big\{\frac{|Z(s) - Z(t)|}{\rho(s,t)} > G(\eta, \delta)\Big\} \ \leq \ K\exp(-\eta) \quad \text{if } \rho(s,t) \geq \delta. \tag{5}$$

Then for arbitrary $\delta > 0$ *and* $a \geq 1$,

$$\mathbb{P}\Big\{|Z(s) - Z(t)| \geq 12J(\rho(s,t), a) \text{ for some } s, t \in \mathcal{T}_* \text{ with } \rho(s,t) \leq \delta\Big\} \ \leq\ \frac{K\delta}{2a},$$

where $\mathcal{T}_*$ *is a dense subset of* $\mathcal{T}$, *and*

$$\begin{aligned} J(\epsilon, a) &:= \int_0^\epsilon G(\log(aD(u)^2/u), u)\, du, \\ D(u) = D(u, \mathcal{T}, \rho) &:= \max\Big\{\#\mathcal{T}_o : \mathcal{T}_o \subset \mathcal{T}, \rho(s,t) > u \text{ for different } s, t \in \mathcal{T}_o\Big\}. \end{aligned}$$

Remark. Suppose that $G(\eta, \delta) = \widetilde{q}\,\eta^q$ for some constants $\widetilde{q}, q > 0$. In addition let $D(u) \leq Au^{-B}$ for $0 < u \leq 1$ with constants $A \geq 1$ and $B > 0$. Then elementary calculations show that for $0 < \epsilon \leq 1$ and $a \geq 1$, $J(\epsilon, a) \leq C\,\epsilon \log(e/\epsilon)^q$ with $C = \widetilde{q}\max\big(1 + 2B, \log(aA^2)\big)^q \int_0^1 \log(e/z)^q\, dz$.

For the proof of Theorem 1 the subsequent extension of the Chaining Lemma VII.9 in Pollard (1984) and Theorem 8 in the technical report to Dümbgen and Walther (2008) will be used. It complements in particular the existing multiscale theory by a uniform tightness result and to a situation where only a sufficiently sharp *uniform stochastic* bound on local covering numbers is available, which typically involves additional logarithmic terms. The situation arises for example in the multivariate random design case where a non-stochastic bound obtained via uniform covering numbers and VC-theory may be too rough.

Theorem 6 (Chaining). *Let* $(Y_n)_{n\in\mathbb{N}}$ *a sequence of random variables such that* Y_n *takes values in some polish space* $\mathcal{Y}_n$. *For any* $y_n \in \mathcal{Y}_n$, *let* $(Z_n(t; y_n))_{t\in\mathcal{T}_{y_n}}$ *be a stochastic process on some countable, metric space* $\big(\mathcal{T}_{y_n}, \rho_n(.,.;y_n)\big)$, *where* $\rho_n(.,.;y_n) \leq 1$. *Suppose that the following conditions are satisfied:*

(i) *There are measurable functions* $\sigma_n(.;Y_n) : \mathcal{T}_{Y_n} \to (0,1]$ *and* $G_n : [0,\infty) \to [0,\infty)$ *such that for arbitrary* $s, t \in \mathcal{T}_{Y_n}$, $\eta \geq 0$ *and* $\delta > 0$,

$$\mathbb{P}\Big(|Z_n(t,Y_n)| \geq \sigma_n(t;Y_n)G_n(\eta,\delta)\,\Big|\,Y_n\Big) \ \leq\ 2\exp(-\eta) \quad \text{if } \sigma_n(t;Y_n) \geq \delta,$$

$$\sup_{s,t\in\mathcal{T}_{Y_n}} \frac{|\sigma_n(t;Y_n) - \sigma_n(s;Y_n)|}{\rho_n(s,t;Y_n)} \ \leq\ C < \infty \text{ for some constant } C > 0,$$

$$\big\{t \in \mathcal{T}_{Y_n} : \sigma_n(t;Y_n) \geq \delta\big\} \text{ is compact, and } G_o := \sup_{n\in\mathbb{N}} \sup_{\eta\geq 0, 0<\delta\leq 1} \frac{G_n(\eta,\delta)}{1+\eta} \ <\ \infty\,.$$

(ii) *There exists a sequence* $(\mathcal{C}_n)_{n\in\mathbb{N}}$ *of measurable sets and positive constants* A, B, W, α *such that*

$$N\Big(u\delta, \{t \in \mathcal{T}_{Y_n} : \sigma_n(t;Y_n) \leq \delta\}, \rho_n(.,.;Y_n)\Big) \ \leq\ Au^{-B}\delta^{-W}\log\big(e/(u\delta)\big)^\alpha \quad \text{for } u,\ \delta \in (0,1]$$

whenever $Y_n \in \mathcal{C}_n$.

For constants $q, Q > 0$ *define*

$$\mathcal{A}_n(\delta, q, Q; Y_n) \ := \ \left\{ \sup_{s,t \in \mathcal{T}_{Y_n} : \rho_n(s,t;Y_n) \le \delta} \frac{|Z_n(s;Y_n) - Z_n(t;Y_n)|}{\rho_n(s,t;Y_n) \log(e/\rho_n(s,t;Y_n))^q} \ \le \ Q \right\}.$$

Then there exists a constant $C = C(G_o, A, B, W, \alpha, q, Q) > 0$ *such that for* $0 < \delta \le 1$

$$\mathbb{P}\Bigg(\frac{|Z_n(t;Y_n)|}{\sigma_n(t;Y_n)} \ \le \ G_n\Big(W \log\big(1/\sigma_n(t;Y_n)\big) + C \log\log\big(e/\sigma_n(t;Y_n)\big), \ \sigma_n(t;Y_n) \Big) \\ + C \log(e/\sigma_n(t;Y_n))^{-1} \text{ on } \big\{ t : \sigma_n(t;Y_n) \le \delta \big\} \ \Bigg| \ Y_n \Bigg)$$

is at least $\mathbb{P}\Big(\mathcal{A}_n(2\delta, q, Q; Y_n) \ \Big| \ Y_n \Big) - C \log(e/\delta)^{-1}$ *whenever* $Y_n \in \mathcal{C}_n$.

If in particular $\mathbb{P}^{Y_n}(\mathcal{C}_n) \to 1$ *and* $\lim_{\delta \searrow 0} \inf_n \mathbb{P}\Big(\mathcal{A}_n(\delta, q, Q; Y_n) \ \Big| \ Y_n \Big) = 1$ *a.s., then the sequence*

$$\mathcal{L}\Bigg(\sup_{t \in \mathcal{T}_n} \bigg\{ \frac{|Z_n(t;Y_n)|}{\sigma_n(t;Y_n)} \ - \ G_n\Big(W \log\big(1/\sigma_n(t;Y_n)\big) \\ + \ C \log\log\big(e/\sigma_n(t;Y_n)\big), \ \sigma_n(t;Y_n) \Big) \bigg\} \ \Bigg| \ Y_n \Bigg)$$

is tight in $\big(\mathbb{P}^{Y_n}\big)$*-probability, provided that* $\inf_n \sup_{t \in \mathcal{T}_{Y_n}} \sigma_n(t;Y_n) > 0$ *a.s.*

REMARK Note that in case of $G(\eta, \delta) = (\kappa\eta)^{1/\kappa}$ with $\kappa > 1$,

$$\begin{aligned} & G\Big(W \log(1/\delta) + C \log\log(e/\delta), \delta \Big) + C \log(e/\delta)^{-1} \\ = \ & (\kappa W \log(1/\delta))^{1/\kappa} + O\Big(\log\log(e/\delta) \log(e\delta)^{1/\kappa - 1} \Big) \\ = \ & (\kappa W \log(1/\delta))^{1/\kappa} + o(1) \quad \text{as } \delta \searrow 0. \end{aligned}$$

PROOF Due to the factorization lemma, the conditional probability and expectation factorize under the above conditions, i.e. we may consider a sequence $(y_n)_{n \in \mathbb{N}}$ and work with the sequence of conditional laws $\mathcal{L}(Z_n(.,Y_n)|Y_n = y_n)$, but note that we do not assume equality of $\mathcal{L}(Z_n(.;Y_n)|Y_n = y_n)$ and $\mathcal{L}(Z_n(.;y_n))$ in general. The first part of the proof is a modification of the Chaining in Dümbgen and Walther (2008, technical report) applied to the conditional distribution $\mathcal{L}(Z_n(.,Y_n)|Y_n = y_n)$ for $y_n \in \mathcal{C}_n$. Here we need however to define their additive correction function H_1 in a different way, taking into account the additional logarithmic terms in the bound of the covering numbers. Lining up with their arguments, a suitable choice for the correction function appears to be

$$G_n\Big\{W\log\Big(\frac{1}{\sigma_n(t;y_n)}\Big)\ +\ (B+\alpha)\log u\big(\sigma_n(t;y_n)\big)\ +\ (2+\alpha)\log\log\Big(\frac{e}{\sigma_n(t;y_n)}\Big),\ \sigma_n(t;y_n)\Big\}$$
$$=\ G_n\Big\{W\log\Big(\frac{1}{\sigma_n(t;y_n)}\Big)+\ \Big((B+\alpha)\gamma+(2+\alpha)\Big)\log\log\Big(\frac{e}{\sigma_n(t;y_n)}\Big),\ \sigma_n(t;y_n)\Big\}.$$

This term is essential for our proof of efficiency. It is important that the constant α does not influence the leading term. Concerning the tightness in probability as stated in the second part of Theorem 6, notice that it does not follow by an immediate continuity argument because the metric (and the metric space) change with both, Y_n and n, hence some additional uniformity is required. For $0\le\delta<\delta'\le 1$ let $U_n(\delta,\delta';Y_n)$ be defined by

$$\sup_{\substack{\sigma_n(t;Y_n)\in(\delta,\delta']\\ t\in\mathcal{T}_n}}\left\{\frac{|Z_n(t;Y_n)|}{\sigma_n(t;Y_n)}-G_n\Big(W\log\big(1/\sigma_n(t;Y_n)\big)+C\log\log\big(e/\sigma_n(t;Y_n)\big),\ \sigma_n(t;Y_n)\Big)\right\}.$$

First observe that for any fixed $K>0$,

$$\mathbb{P}\Big(U_n(0,1;Y_n)>K\,\Big|\,Y_n\Big)\ \le\ \mathbb{P}\Big(U_n(0,\delta;Y_n)>K/2\,\Big|\,Y_n\Big)+\mathbb{P}\Big(U_n(\delta,1;Y_n)>K/2\,\Big|\,Y_n\Big). \tag{6}$$

The first part of Theorem 6 implies that the first term on the right-hand-side in (6) is bounded by $1-\mathbb{P}\big(\mathcal{A}_n(2\delta,q,Q;Y_n)\,\big|\,Y_n\big)+C\log(e/\delta)^{-1}$ for $K>2C\log(e/\delta)^{-1}$ whenever $Y_n\in\mathcal{C}_n$. Concerning the second term in (6), note that

$$U_n(\delta,1;Y_n)\ \le\ -\inf_{\delta'\in[\delta,1]}H_n(\delta';Y_n)\ +\ \frac{1}{\delta}\sup_{\substack{t\in\mathcal{T}_{Y_n}:\\ \sigma_n(t;Y_n)\ge\delta}}\Big|Z_n(t;Y_n)\Big|.$$

Then the conclusion follows if we establish that

$$\lim_{K\to\infty}\limsup_{n\to\infty}\mathbb{P}\Big(\sup_{t\in\mathcal{T}_{Y_n}}\big|Z_n(t;Y_n)\big|\ >\ K;\ Y_n\in\mathcal{C}_n\,\Big|\,Y_n\Big)\ =\ 0\ \ a.s.$$

For $\varepsilon>0$ and $Y_n\in\mathcal{C}_n$, let $t_1(Y_n),...,t_{m(Y_n)}(Y_n)$ be a maximal subset of $\mathcal{T}_{Y_n}$ with $\rho_n(t_i,t_j;Y_n)>\varepsilon$ for arbitrary different indices $i,j\in\{1,...,m(Y_n)\}$. Note that $m(Y_n)\le A\varepsilon^{-B}\log(e/\varepsilon)^{\alpha}$ by assumption (ii). Then condition (i) implies that

$$\lim_{K\to\infty}\limsup_{n\to\infty}\mathbb{P}\Big(\sup_{i=1,...,m(Y_n)}\big|Z_n\big(t_i(Y_n);Y_n\big)\big|>K;\ Y_n\in\mathcal{C}_n\,\Big|\,Y_n\Big)\ =\ 0\ \ a.s. \tag{7}$$

On the other hand, we have on the set $\mathcal{A}_n(\varepsilon,q,Q;Y_n)$ the bound

$$\sup_{t\in\mathcal{T}_{Y_n}}|Z_n(t;Y_n)|\ \le\ Q\varepsilon\log(e/\varepsilon)^q\ +\sup_{i=1,...,m(Y_n)}\big|Z_n(t_i(Y_n);Y_n)\big|. \tag{8}$$

With ε tending to zero sufficiently slowly, (7) and (8) show together with the stochastic equicontinuity condition $\lim_{\delta\searrow 0}\inf_n \mathbb{P}\Big(\mathcal{A}_n(\delta,q,Q;Y_n)\,\Big|\,Y_n\Big) = 1$ a.s.

$$\lim_{K\to\infty}\limsup_{n\to\infty}\mathbb{P}\Bigg(\sup_{t\in\mathcal{T}_{Y_n}}\,\big|\,Z_n(t;Y_n)\,\big|\,>K;\,Y_n\in\mathcal{C}_n\,\Bigg|\,Y_n\Bigg)\;=\;0\;\;a.s.$$

Since the assumption $\inf_n\sup_{t\in\mathcal{T}_{Y_n}}\sigma_n(t;Y_n)>0$ a.s. guarantees

$$\lim_{K\to\infty}\sup_n\mathbb{P}\Big(U_n(Y_n)<-K\,\Big|\,Y_n\Big)\;=\;0\;\;a.s.,$$

the tightness in $(\mathbb{P}^{Y_n})$-probability is proved. □

8 Proofs of the main results

PROOF OF THEOREM 1 Let $\lambda_n := m/n$. In view of the T_{jkn}'s, the behavior of the process

$$\Bigg(\frac{\sqrt{\lambda_n(1-\lambda_n)}}{\sqrt{n}}\sum_{i=0}^{k}\psi\bigg(\frac{\|X_j-X_j^i\|_2}{\|X_j-X_j^k\|_2}\bigg)\big(\Lambda\circ\Pi\big)(X_j^i)\Bigg)_{1\le j,k\le n}$$

conditional on $\mathcal{X}_n$ needs to be investigated, where $\Lambda\circ\Pi|\mathcal{X}_n$ is uniformly distributed on the set

$$\Big\{\lambda:\mathcal{X}_n\to\big\{1/\lambda_n,-1/(1-\lambda_n)\big\}:\sum_{x\in\mathcal{X}_n}\lambda(x)=0\Big\}.$$

For notational convenience it seems useful to redefine the process on the random index set

$$\hat{\mathcal{T}}_n\;:=\;\Big\{\big(X_j,\|X_j-X_j^k\|_2\big)\,:\,1\le j,k\le n\Big\}$$

via the map $(j,k)\;\mapsto\;\big(X_j,\;\big\|\,X_j-X_j^k\,\big\|_2\big)$ and extend it to a process $\big(Y_n(t,r)\big)_{(t,r)\in\mathcal{T}}$ with $\mathcal{T}:=\big\{(t,r):\,t\in[0,1]^d,\,0<r\le\max_{x\in[0,1]^d}\|x-t\|_2\big\}$ by the definition

$$Y_n(t,r)\;:=\;\sqrt{n}\sqrt{\lambda_n(1-\lambda_n)}\int\psi\bigg(\frac{\|t-x\|_2}{r}\bigg)\Big(d\hat{\mathbb{P}}_n^{\Pi}(x)-d\hat{\mathbb{Q}}_n^{\Pi}(x)\Big),$$

where $\hat{\mathbb{P}}_n^{\Pi}$ and $\hat{\mathbb{Q}}_n^{\Pi}$ denote the empirical measures based on the permutated variables $X_{\Pi(1)},...,X_{\Pi(m)}$ and $X_{\Pi(m+1)},...,X_{\Pi(n)}$, respectively. Let

$$\begin{aligned}\hat{\gamma}_n(t,r)^2\;:&=\;\mathrm{Var}\Big(Y_n(t,r)\,\Big|\,\mathcal{X}_n\Big)\\&=\;\frac{n}{n-1}\int\bigg[\psi\bigg(\frac{\|t-x\|_2}{r}\bigg)-\int\psi\bigg(\frac{\|t-z\|_2}{r}\bigg)d\hat{\mathbb{H}}_n(z)\bigg]^2d\hat{\mathbb{H}}_n(x),\end{aligned}$$

with $\hat{\mathbb{H}}_n$ the empirical measure of the observations $X_1,...,X_n$.

In the sequel we make use of the results in the previous section twice - in order to prove the tightness and weak approximation in probability of the sequence of conditional test statistics and within the "loop" we use the chaining arguments again to establish a sufficiently tightened uniform stochastic bound for the covering numbers below.

I. (SUBEXPONENTIAL INCREMENTS AND BERNSTEIN TYPE TAIL BEHAVIOR) The inversion of the conditional Bernstein type exponential inequality in Proposition 3 shows that for any $\eta > 0$,

$$\mathbb{P}\Big(\Big| \frac{Y_n(t,r)}{\hat{\gamma}_n(t,r)} \Big| > G_n\big(\eta, \hat{\gamma}_n(t,r)\big) \Big| \mathcal{X}_n \Big) \;\le\; 2\,\exp(-\eta),$$

where

$$G_n\big(\eta, \hat{\gamma}_n(t,r)\big) \;:=\; R_n\big(\hat{\gamma}_n(t,r)\big)\eta \;+\; \Big(\big(R_n\big(\hat{\gamma}_n(t,r)\big)\eta\big)^2 + 2\delta(m,n)^2\eta\Big)^{1/2}$$

with

$$R_n(\tau) \;:=\; \delta(m,n)\frac{\|\psi\|_{\sup}\sqrt{\lambda_n(1-\lambda_n)}}{3\,\min(\lambda_n, 1-\lambda_n)\sqrt{n}\,\tau}.$$

Let the random metric $\hat{\rho}_n$ on $\mathcal{T}$ be defined by

$$\begin{aligned}\hat{\rho}_n\big((t,r),(t',r')\big)^2 &:= \mathrm{Var}\Big(Y_n(t,r) - Y_n(t',r')\,\Big|\,\mathcal{X}_n\Big)\\ &= \int \Big(\psi_{tr}(x) - \psi_{t'r'}(x)\Big)^2 d\hat{\mathbb{H}}_n(x) \;-\; \Big(\int \Big(\psi_{tr}(x) - \psi_{t'r'}(x)\Big)d\hat{\mathbb{H}}_n(x)\Big)^2,\end{aligned}$$

with $\psi_{tr}(x) \;:=\; \psi\Big(\frac{\|t-x\|_2}{r}\Big)$. Then the application of the second exponential inequality of Proposition 3 implies for any fixed $(t,r), (t',r') \in \mathcal{T}$ that

$$\mathbb{P}\Big(\big| Y_n(t,r) - Y_n(t',r') \big| > \hat{\rho}_n\big((t,r),(t',r')\big)\,q\,\eta \,\Big|\, \mathcal{X}_n \Big) \le\; 2\,\exp(-\eta),$$

where

$$q \;:=\; 2\Big(1 \;+\; \frac{9\lambda_n(1-\lambda_n)}{2\max(\lambda_n, 1-\lambda_n)^2}\big(\log 2\big)^{-1}\Big).$$

II. (RANDOM LOCAL COVERING NUMBERS) We need a bound for the local random covering numbers $N\big((u\delta)^{1/2}, \big\{(t,r) \in \hat{\mathcal{T}}_n : \hat{\gamma}_n(t,r)^2 \;\le\; \delta\big\}, \hat{\rho}_n\big)$. This is the most involved part of the proof. In order to establish a sufficiently sharp upper bound, the following two claims are established:

(i) Let

$$\hat{\rho}_{2,n}\big((t,r),(t',r')\big)^2 \;:=\; \int \Big(\psi_{tr}(x) - \psi_{t'r'}(x)\Big)^2 d\hat{\mathbb{H}}_n(x)$$

and define d_n for arbitrary different points in $\hat{\mathcal{T}}_n$ via

$$d_n{}^2 := \max\big[\mathbb{E}\,\hat{\rho}_{2,n}^2, 4/n\big]\Big(1 + C\log\Big(4\,e\Big/\max\big[\mathbb{E}\,\hat{\rho}_{2,n}^2, 4/n\big]\Big)\Big),$$

with C a constant to be chosen later. Note that the map $x \mapsto x\sqrt{1+2C\log(\sqrt{e}/x)}$ is subadditive for $x \in (0,1]$, hence d_n defines a metric. Furthermore let ${\gamma_n}^2 := \mathbb{E}\,\hat{\gamma}_{2,n}^2 - \big(\mathbb{E}\hat{\gamma}_{1,n}\big)^2$, where

$$\hat{\gamma}_{1,n}(t,r)^2 \ := \ \Big(\int \psi_{tr}(x) d\hat{\mathbb{H}}_n(x)\Big)^2 \ \text{ and } \ \hat{\gamma}_{2,n}(t,r)^2 \ := \ \int \psi_{tr}(x)^2 d\hat{\mathbb{H}}_n(x).$$

Then there exist a constant $C' > 0$ and a sequence $(\mathcal{C}_n)_{n\in\mathbb{N}}$ of measurable sets with $\mathbb{P}_n^{\otimes m} \otimes \mathbb{Q}_n^{\otimes(n-m)}(\mathcal{C}_n) \to 1$, such that for any $\delta > 0$, $u \in (0,1]$ with $u\delta \geq 1/n$ and any realization $(X_1, ..., X_n) \in \mathcal{C}_n$

$$\begin{aligned} &N\Big((u\delta)^{1/2}, \Big\{(t,r) \in \hat{\mathcal{T}}_n : \hat{\gamma}_n(t,r)^2 \ \leq \ \delta\Big\}, \hat{\rho}_n\Big) \\ &\qquad\qquad \leq \ N\Big((u\delta)^{1/2}, \Big\{(t,r) \in \hat{\mathcal{T}}_n : \gamma_{2,n}(t,r)^2 \ \leq \ C'\delta \log(e/\delta)^4\Big\}, d_n\Big), \end{aligned}$$

if ψ is not rectangular. In case of the rectangular kernel, the set

$$\Big\{(t,r) \in \hat{\mathcal{T}}_n : \gamma_{2,n}(t,r)^2 \ \leq \ C'\delta \log(e/\delta)^4\Big\}$$

in the covering number has to be replaced by

$$\Big\{(t,r) \in \hat{\mathcal{T}}_n : \gamma_{2,n}^2 \ \leq \ C'\delta \log\big(e/\delta\big)^4\Big\} \cup \Big\{(t,r) \in \hat{\mathcal{T}}_n : \gamma_{2,n}^2 \ \geq \ 1 - C'\delta \log\big(e/\delta\big)^4\Big\}.$$

(ii) There exists a constant $A > 0$, independent of u, δ and n, such that whenever $u\delta \geq 1/n$, the upper bound given in (i) is again bounded from above by $Au^{-d-1}\delta^{-1}\log\big(e/(u\delta)\big)^5$. Moreover, the latter bound remains valid with $\mathcal{T}$ in place of $\hat{\mathcal{T}}_n$.

Note that we cannot rely our bound directly on uniform covering numbers and Vapnik-Cervonenkis (VC) theory as the envelope $I\{X \in \mathcal{X}_n\}$ only allows for a bound of order $u^{-2}\delta^{-2}$, which would result in the loss of efficiency of the procedure.

Proof of (i): We first derive a uniform stochastic bound for the random metric $\hat{\rho}_{2,n}$. Recall that every function ψ of bounded total variation is representable as a difference of isotonic functions $\psi^{(1)}$ and $\psi^{(2)}$. With the definition of the subgraphs

$$\mathrm{sgr}\big(\psi_{tr}^{(i)}\big) \ := \ \Big\{(x,y) \in [0,1]^d \times \mathbb{R} : \ y \leq \psi_{tr}^{(i)}(x)\Big\}, \ \ i = 1,2,$$

the set $\big\{\mathrm{sgr}\big(\psi_{tr}^{(i)}\big) \, : \, (t,r) \in \mathcal{T}\big\}$ has a VC-dimension bounded by $d+3$ (van der Vaart and Wellner 1996) with envelope $TV(\psi)$. Consequently, the uniform covering numbers $N(\varepsilon, \mathcal{F})$ with

$$\mathcal{F} \ := \ \Big\{\big(\psi_{tr} - \psi_{t'r'}\big)^2 : \ (t,r), (t',r') \in \mathcal{T}\Big\}$$

are bounded by $C\varepsilon^{-\alpha}$ for some realvalued $\alpha > 0$ and some constant $C > 0$. The boundedness of ψ shows that $\mathcal{F}$ is uniform Glivenko-Cantelli in particular (Dudley, Giné and Zinn 1991). As an immediate consequence,

$$\lim_{n\to\infty} \mathbb{P}\Big(\Big\| \hat\rho_{2,n}\big((t,r),(t',r')\big)^2 - \mathbb{E}\hat\rho_{2,n}\big((t,r),(t',r')\big)^2 \Big\|_{\mathcal{T}\times\mathcal{T}} > \delta \Big) \; = \; 0, \tag{9}$$

for any $\delta > 0$. However such a bound is not sufficient for our purposes. Because of $\|\psi\|_{\sup} \le 1$, the squared random metric $\hat\rho_{2,n}^2$ is $1/n$ times the sum of n independent random variables with absolute values ≤ 4, hence

$$\mathrm{Var}\Big(\hat\rho_{2,n}\big((t,r),(t',r')\big)^2\Big) \;\le\; \frac{4}{n}\,\mathbb{E}\Big(\hat\rho_{2,n}\big((t,r),(t',r')\big)^2\Big) \;\le\; \max\Big\{\frac{4}{n}, \mathbb{E}\Big(\hat\rho_{2,n}\big((t,r),(t',r')\big)^2\Big)\Big\}^2.$$

Now the application of Bernstein's exponential inequality (see Shorack and Wellner 1986) entails

$$\begin{aligned}\mathbb{P}\Bigg(\Bigg| \frac{\hat\rho_{2,n}\big((t,r),(t',r')\big)^2 - \mathbb{E}\hat\rho_{2,n}\big((t,r),(t',r')\big)^2}{\max[4/n, \mathbb{E}\hat\rho_{2,n}\big((t,r),(t',r')\big)^2]} \Bigg| \;>\; \eta \Bigg) \;&\le\; 2\exp\Big(- \frac{\eta^2/2}{1+\eta/3}\Big) \\ &\le\; 2\exp\Big(- \frac{3}{2}\eta + \frac{9}{2}\Big)\end{aligned}$$

for arbitrary points $(t,r),(t',r') \in \mathcal{T}$. I.e. $\hat\rho_{2,n}^2 - \mathbb{E}\hat\rho_{2,n}^2$, standardized by

$$\max\Big\{4/n, \mathbb{E}\hat\rho_{2,n}\big((t,r),(t',r')\big)^2\Big\},$$

has (uniformly) subexponential tails. Analogously, the process $\hat\rho_{2,n}^2 - \mathbb{E}\hat\rho_{2,n}^2$ has subexponential increments with respect to the metric $\hat D_n$ given by

$$\hat D_n\big(a,b\big)^2 := \max\big[1/n, \mathbb{E}\big(\hat\rho_{2,n}^2(a) - \hat\rho_{2,n}^2(b)\big)^2\big]^2 I\big\{a \ne b\big\}, \;\; a,b \in \mathcal{T}\times\mathcal{T}.$$

Note that $\max[4/n, \mathbb{E}\hat\rho_{2,n}^2]$ is Lipschitz continuous with respect to $\hat D_n$. Theorem 5 shows that the above ingredients imply that $\lim_{\delta\searrow 0}\inf_n \mathbb{P}\big(\mathcal{A}_n(\delta,1,Q;\mathcal{X}_n)\,\big|\,\mathcal{X}_n\big) = 1$ for some adequately chosen $Q > 0$, where we use the definition of $\mathcal{A}_n$ from Theorem 6 with $Y_n = \mathcal{X}_n$ and $Z_n = \hat\rho_{2,n}^2 - \mathbb{E}\hat\rho_{2,n}^2$. Now we may apply the latter to conclude that there exists some universal constant $C > 0$ such that the probability of the event

$$\begin{aligned}\Big\{ \Big| \hat\rho_{2,n}\big((t,r),(t',r')\big)^2 - \mathbb{E}\hat\rho_{2,n}\big((t,r),(t',r')\big)^2 \Big| > \qquad\qquad\qquad\qquad & \\ C\,\max\big[4/n, \mathbb{E}\hat\rho_{2,n}\big((t,r),(t',r')\big)^2\big] \log\Big(4\,e\Big/\max\big[4/n, \mathbb{E}\hat\rho_{2,n}\big((t,r),(t',r')\big)^2\big]\Big) & \\ \text{for some } (t,r),(t',r') \text{ with } \mathbb{E}\hat\rho_{2,n}\big((t,r),(t',r')\big)^2 \le \delta \Big\} &\end{aligned} \tag{10}$$

is bounded by some function $\varepsilon(\delta)$ independent of n with $\lim_{\delta\searrow 0}\varepsilon(\delta) = 0$. Combining (9) and (10) for a sequence $\delta = \delta_n \searrow 0$ sufficiently slowly implies the existence of a sequence of sets $(\mathcal{A}_n)_{n\in\mathbb{N}}$ with $\mathbb{P}^{\otimes m}\otimes\mathbb{Q}^{\otimes(n-m)}(\mathcal{A}_n) \to 1$ such that

$$\hat\rho_{2,n} \;\le\; \max\big[4/n, \mathbb{E}\hat\rho_{2,n}^2\big]^{1/2}\Big(1 + C\log\big(4\,e\big/\max\big[4/n, \mathbb{E}\hat\rho_{2,n}^2\big]\big)\Big)^{1/2} \quad \text{whenever } \underline{\mathrm{X}} \in \mathcal{A}_n.$$

The treatment of the random set

$$\hat{\mathcal{B}}_\delta \;:=\; \Big\{(t,r) \in \hat{\mathcal{T}}_n : \hat{\gamma}_n(t,r)^2 \;\leq\; \delta\Big\}$$

is similar in spirit but more involved because the random quantity $\hat{\gamma}_n^2$ is not representable as a sum of independent variables. However we can use the decomposition $[n/(n-1)]\hat{\gamma}_n^2 = \hat{\gamma}_{2,n}^2 - \hat{\gamma}_{1,n}^2$. Before deriving a stochastic bound, we notice the following: If ψ describes the rectangular kernel, we have $\hat{\gamma}_{2,n}^2 = \hat{\gamma}_{1,n}$, i.e.

$$\hat{\gamma}_{2,n}^2 - \hat{\gamma}_{1,n}^2 \;=\; \hat{\gamma}_{2,n}^2\big(1 - \hat{\gamma}_{2,n}^2\big).$$

In this case, the random set $\hat{\mathcal{B}}_\delta$ is consequently contained in the union

$$\Big\{\hat{\gamma}_{2,n}^2 \leq 4\delta\Big\} \cup \Big\{\hat{\gamma}_{2,n}^2 \geq 1 - 4\delta\Big\} \quad \text{for } \delta \geq 1/n. \tag{11}$$

Consider the general case. Using that

$$\mathrm{Var}\Big(\hat{\gamma}_{1,n}(t,r)\Big) \;=\; \frac{1}{n^2}\sum_{i=1}^{n}\Big(\mathbb{E}\psi_{tr}(X_i)^2 - \big(\mathbb{E}\psi_{tr}(X_i)\big)^2\Big) \;\leq\; \frac{1}{n}\mathbb{E}\,\hat{\gamma}_{2,n}^2 \tag{12}$$

and

$$\mathrm{Var}\Big(\hat{\gamma}_{2,n}(t,r)^2\Big) \;=\; \frac{1}{n^2}\sum_{i=1}^{n}\Big(\mathbb{E}\psi_{tr}(X_i)^4 - \big(\mathbb{E}\psi_{tr}(X_i)^2\big)^2\Big) \;\leq\; \frac{1}{n}\mathbb{E}\,\hat{\gamma}_{2,n}^2, \tag{13}$$

we may apply the above chain of arguments for $\hat{\rho}_{2,n}^2$ to $\hat{\gamma}_{1,n}$ and $\hat{\gamma}_{2,n}^2$ together with the upper bounds in (12) and (13) for the standardization respectively and obtain the existence of a constant $C_1 > 0$ such that

$$\begin{aligned}\gamma_{1,n} - \frac{C_1 \max\big[1/n, \gamma_{2,n}^2\big]^{1/2}}{\sqrt{n}} \log\Big(e\sqrt{n}\Big/\max\big[1/n, \gamma_{2,n}\big]^{1/2}\Big) \\ \leq\; \hat{\gamma}_{1,n} \;\leq\; \gamma_{1,n} \;+\; \frac{C_1 \max\big[1/n, \gamma_{2,n}^2\big]^{1/2}}{\sqrt{n}} \log\Big(e\sqrt{n}\Big/\max\big[1/n, \gamma_{2,n}^2\big]^{1/2}\Big)\end{aligned}$$

whenever $\underline{\mathbf{X}} \in \mathcal{D}_n$ for some sequence $(\mathcal{D}_n)_{n\in\mathbb{N}}$ with asymptotic probability 1, uniformly evaluated at $(t,r) \in \hat{\mathcal{T}}_n$. Note that $\hat{\gamma}_{1,n} \geq 1/n$, $\hat{\gamma}_{2,n}^2 \geq 1/n$ for all $(t,r) \in \hat{\mathcal{T}}_n$. The same holds true with a constant $C_2 > 0$ and a sequence $(\mathcal{D}_n')_{n\in\mathbb{N}}$ with asymptotic probability 1 and $\hat{\gamma}_{1,n}$ and $\gamma_{1,n}$ replaced by $\hat{\gamma}_{2,n}^2$ and $\gamma_{2,n}^2$. Using the lower bound for $\hat{\gamma}_{2,n}^2$ and the upper bound for $\hat{\gamma}_{1,n}$, a bit of algebra yields

$$\hat{\mathcal{B}}_\delta \;\subset\; \left\{\gamma_{2,n}^2 - \gamma_{1,n}^2 \;\leq\; \delta \;+\; \max\big[1/n, \gamma_{2,n}^2\big]^{1/2}\frac{K}{\sqrt{n}}\log\Big(e\sqrt{n}\Big/\max\big[1/n, \gamma_{2,n}^2\big]^{1/2}\Big)^2\right\}$$

whenever $\underline{\mathbf{X}} \in \mathcal{D}_n \cap \mathcal{D}_n',\ \delta \geq 1/n$. Here and from now on, K denotes some universal constant, not dependent on n. Its value may be different in different expressions. Now we first consider the case

$$\sup_{n\in\mathbb{N}}\sup_{(t,r)\in\mathcal{T}}\Big(\gamma_{1,n}^2\Big/\gamma_{2,n}^2\Big) \;\leq\; C' \;<\; 1.$$

Then the above condition shows that

$$\begin{aligned}\gamma_{2,n}^2(1-C') \;&\leq\; \delta \;+\max\big[1/n,\gamma_{2,n}^2\big]^{1/2}\frac{K}{\sqrt{n}}\log\Big(e\sqrt{n}\Big/\max\big[1/n,\gamma_{2,n}^2\big]^{1/2}\Big)^2\\ &\leq\; 2\max\Big\{\delta,\;\max\big[1/n,\gamma_{2,n}^2\big]^{1/2}\frac{K}{\sqrt{n}}\log\Big(e\sqrt{n}\Big/\max\big[1/n,\gamma_{2,n}^2\big]^{1/2}\Big)^2\Big\},\end{aligned}$$

which entails that $\gamma_{2,n}^2 \;\leq\; K\,\delta\log\big(e/\delta\big)^4$ for $\delta \geq 1/n$ by the isotonicity of $x \mapsto x\log(e/x)^4$ on $(0,1]$. On the other hand, the case

$$\sup_{n\in\mathbb{N}}\sup_{(t,r)\in\mathcal{T}}\Big(\gamma_{1,n}^2\Big/\gamma_{2,n}^2\Big) \;=\; 1 \tag{14}$$

implies already that ψ is equal to the rectangular kernel: If the sup is attained it is obvious. The equicontinuity of $(h_n)_{n\in\mathbb{N}}$ and its uniformly bounded L_1-norm $\|h_n\|_1 = 1$ imply its uniform boundedness, hence relative compactness in the topology of uniform convergence by the Arzelà-Ascoli-Theorem. There therefore exists at least a uniformly convergent subsequence $(h_{m(n)})$ with (uniformly) continuous limit, say h, along this result holds true as well, because $\max_{(t,r)\in\mathcal{T}}\big(\gamma_{1,n}^2/\gamma_{2,n}^2\big)$ depends continuously on the mixed density. This however implies that ψ describes the rectangular kernel, because the uniform limit h of that subsequence is bounded away from zero. Hence in case of (14), we consequently obtain by (11)

$$\hat{\mathcal{B}}_\delta \;\subset\; \Big\{\gamma_{2,n}^2 \;\leq\; K\delta\log\big(e/\delta\big)^4\Big\}\cup\Big\{\gamma_{2,n}^2 \;\geq\; 1-K\delta\log\big(e/\delta\big)^4\Big\}\text{ whenever }\underline{\mathrm{X}}\in\mathcal{D}_n\cap\mathcal{D}_n',\;\delta\geq 1/n.$$

Proof of (ii): Since ψ is of bounded total variation, there exists some finite measure μ such that for any $0 \leq z_1 < z_2 \leq 1$, $|\psi(z_1)-\psi(z_2)| \leq \mu[z_1,z_2]$. With

$$M_x(t,t',r,r') \;:=\; \Big[0,\frac{\|t-x\|_2}{r}\Big]\Delta\Big[0,\frac{\|t'-x\|_2}{r'}\Big]$$

we obtain

$$\begin{aligned}\mathbb{E}\hat{\rho}_{2,n}\big((t,r),(t',r')\big)^2 \;&\leq\; \int\big(\psi_{tr}(x)-\psi_{t'r'}(x)\big)^2 d\mathbb{H}_n(x)\\ &\leq\; K\int\mu\big(M_x(t,t',r,r')\big)\,d\mathbb{H}_n(x)\\ &=\; K\int I\big\{y\in M_x(t,t',r,r')\big\}\,d\mathbb{H}_n(x)d\mu(y)\\ &\leq\; K\sup_{y\in[0,1]}\int I\big\{y\in M_x(t,t',r,r')\big\}\,d\mathbb{H}_n(x).\end{aligned}\tag{15}$$

Then $y \in M_x(t,t',r,r')$ implies that $x \in B_t\big(ry\big)\Delta B_{t'}\big(r'y\big)$. Since h_n is uniformly bounded from above, we obtain that (15) is not greater than $C\lambda\big(B_t(r)\Delta B_{t'}(r')\big)$. Because of $\int_{[0,1]^d}\psi(x)dx = 1$ with maximum attained at 0, there exists some compact ball $B_0(r^*)$ with $\psi(x)\geq 1/2$ for all $x\in B_0(r^*)$. Using in addition the uniform boundedness of h_n away from zero we obtain $\gamma_{2,n}(t,r)^2 \geq K \cdot r^d \;\; (t,r) \in \mathcal{T}$. We now start bounding the covering numbers

$$N\bigg((u\delta)^{1/2},\, \Big\{(t,r) \in \mathcal{T} : \gamma_{2,n}(t,r)^2 \leq 2\delta \log(e/\delta)^4\Big\},\, d\bigg),$$

where the metric d on $\mathcal{T} \times \mathcal{T}$ is pointwise defined by

$$d\big((t,r),(t',r')\big)^2 \;:=\; \lambda\big(B_t(r)\Delta B_{t'}(r')\big)\bigg(1 + C\log\Big[e\Big/\lambda\big(B_t(r)\Delta B_{t'}(r')\big)\Big]\bigg).$$

Because of the isotonicity of $x \mapsto x\log(e/x)$ for $x \in (0,1]$, the inequality $\tilde{d}\big((t,r),(t',r')\big) := \lambda\big(B_t(r)\Delta B_{t'}(r')\big)^{1/2} \;\leq\; \varepsilon/(\log(e/\varepsilon))$ implies that $d\big((t,r),(t',r')\big)$ is not greater that $(C+1)^{1/2}\varepsilon$. Thus it is sufficient to bound

$$N\bigg(\frac{(u\delta)^{1/2}}{\log(e/(u\delta))},\, \Big\{(t,r) \in \mathcal{T} : r^d \leq \delta\log(e/\delta)^4\Big\},\, \tilde{d}\bigg). \tag{16}$$

First note that there exists a finite collection of at most $m \;\leq\; K/(\delta\log(e/\delta)^4)$ points $t_1,...,t_m$ such that the set $\Big\{(t,r) \;\in\; \mathcal{T} \;:\; r^d \;\leq\; \delta\log(e/\delta)^4\Big\}$ is contained in the union $\cup_{i=1}^m \mathcal{A}_i$ with

$$\mathcal{A}_i \;:=\; \bigg\{(t,r) \in \mathcal{T} : B_t(r) \subset B_{t_i}\Big([K\delta\log(e/\delta)^4]^{1/d}\Big)\bigg\}$$

for some universal $K > 0$. The rotation and translatation invariance of the Lebesgue measure leads to the rescaling invariance for the covering numbers

$$N\Big(\varepsilon^{1/2},\, \big\{(t,r) :\; B_t(r) \subset B_0(R)\big\},\, \tilde{d}\Big) \;=\; N\Big((\varepsilon/R^d)^{1/2},\, \big\{(t,r) :\; B_t(r) \subset B_0(1)\big\},\, \tilde{d}\Big).$$

Now it remains being noticed that the latter quantity is bounded by $K(R^d/\varepsilon)^{d+1}$ uniformly in ε and R. Analogously for $N\big((u\delta)^{1/2}\big/\log[e/(u\delta)], \mathcal{A}_i, \tilde{d}\big)$, hence the covering number (16) is bounded by $A\delta^{-1}u^{-d+1}\log(e/u\delta)^5$ for some universal constant $A > 0$. An analogous bound holds for $\hat{\mathcal{T}}_n$ in place of $\mathcal{T}$: If $(t_1,r_1),...,(t_k,r_k)$ denotes an ε-net with respect to d in $B \subset \mathcal{T}$, we may define a $2\,\varepsilon$-net $(\hat{t}_1,\hat{r}_1),...,(\hat{t}_k,\hat{r}_k)$ in $\hat{\mathcal{T}}_n \cap B$ via the definition $(\hat{t}_i,\hat{r}_i) := \operatorname{argmin}_{(t,r)\in\hat{\mathcal{T}}_n\cap B}\, d\big((t,r),(t_i,r_i)\big)$. The corresponding covering numbers in case of the rectangular kernel for the sets $\big\{\gamma_{2,n}^2 \;\geq\; 1 - K\delta\log\big(e/\delta\big)^4\big\}$ can be treated with similar arguments, which concludes the proof of (ii).

III. (Tightness and weak approximation in probability) As a consequence of the above exponential inequalities in step I and the bound for the uniform covering numbers $N(\delta,\mathcal{T})$, Theorem 5 shows

$$\lim_{\delta\searrow 0}\limsup_{n\to\infty}\mathbb{P}\left(\sup_{\hat{\rho}_n((t,r),(t',r'))\leq\delta}\frac{|Y_n(t,r) - Y_n(t',r')|}{\hat{\rho}_n((t,r),(t',r'))\log\big(e/\hat{\rho}_n((t,r),(t',r'))\big)} \;>\; \varepsilon\;\middle|\;\mathcal{X}_n\right) \;=\; 0, \tag{17}$$

where the $\sup$ within the brackets is even running over elements of $\mathcal{T} \times \mathcal{T}$. Now the application of Theorem 6 entails that $\mathcal{L}\big(T_n \circ \Pi \,\big|\, \mathcal{X}_n\big)$ is tight in $\big(\mathbb{P}_n^{\otimes m} \otimes \mathbb{Q}_n^{\otimes(n-m)}\big)$-probability. What remains being proved is the weak approximation. Starting from (17), the uniform convergence (9) implies in particular the asymptotic stochastic equicontinuity

$$\lim_{\delta\searrow 0} \limsup_{n\to\infty} \mathbb{E}_{(p_n,q_n,\lambda_n)} \mathbb{P}\Bigg(\sup_{\rho_n((t,r),(t',r'))\le\delta} \big| Y_n(t,r) - Y_n(t',r') \big| > \varepsilon \,\Bigg|\, \mathcal{X}_n \Bigg) \;=\; 0 \text{ for all } \varepsilon > 0.$$

Since to any subsequence of the metric ρ_n there exists some uniformly convergent subsubsequence as a consequence of the relative compactness of $(h_n)_{n\in\mathbb{N}}$ in the uniform topology, it suffices (via proof of contradiction) for the weak approximation in probability

$$d_w\Big\{\mathcal{L}\Big(\big(Y_n(t,r)\big)_{(t,r)\in\mathcal{T}} \,\Big|\, \mathcal{X}_n\Big),\; \mathcal{L}\Big(\big(Z_n(t,r)\big)_{(t,r)\in\mathcal{T}}\Big)\Big\} \longrightarrow_{\mathbb{P}_n^{\otimes m}\otimes\mathbb{Q}_n^{\otimes(n-m)}} 0$$

to establish the convergence of finite dimensional distributions. For let $\mathcal{S} := \big\{(t_1,r_1),\dots,(t_k,r_k)\big\}$ be a collection of points from $\mathcal{T}$. Denote furthermore $a_{rt}(X_i) := n^{-1/2}\sqrt{\lambda_n(1-\lambda_n)}\,\psi\Big(\frac{\|t-X_i\|_2}{r}\Big)$. Then

$$\mathcal{L}\Big(\big(Y_n(t,r)\big)_{(t,r)\in\mathcal{T}_n} \,\Big|\, \mathcal{X}_n\Big) \;=\; \mathcal{L}\Bigg(\sum_{i=1}^{n} a_{rt}(X_i)\Lambda(t^i) \,\Bigg|\, \mathcal{X}_n\Bigg).$$

Let $\big(Z_n(t,r)\big)_{(t,r)\in\mathcal{T}}$ be pointwise be defined by $Z_n(t,r) := \sqrt{\lambda_n(1-\lambda_n)}\int \phi_{rt}^{(n)}(x)\,dW(x)$. Using that $2\,\mathrm{cov}\,(X_1,X_2)$ equals $\mathrm{Var}(X_1+X_2) - \mathrm{Var}(X_1) - \mathrm{Var}(X_2)$ for two random variables X_1 and X_2, one finds

$$\begin{aligned}
&\frac{n}{n+1}\,\mathrm{cov}\Big(Y_n(t,r), Y_n(t',r') \,\Big|\, \mathcal{X}_n\Big) \\
&= \frac{1}{2}\int\Big(\psi_{tr}(x) - \psi_{t'r'}(x)\Big)^2 d\hat{\mathbb{H}}_n(x) - \frac{1}{2}\bigg(\int\Big(\psi_{tr}(x) - \psi_{t'r'}(x)\Big)d\hat{\mathbb{H}}_n(x)\bigg)^2 \qquad (18) \\
&- \frac{1}{2}\int \psi_{tr}(x)^2 d\hat{\mathbb{H}}_n(x) + \frac{1}{2}\bigg(\int \psi_{tr}(x)d\hat{\mathbb{H}}_n(x)\bigg)^2 - \frac{1}{2}\int \psi_{t'r'}(x)^2 d\hat{\mathbb{H}}_n(x) + \frac{1}{2}\bigg(\int \psi_{t'r'}(x)d\hat{\mathbb{H}}_n(x)\bigg)^2.
\end{aligned}$$

Replacing the empirical measure $\hat{\mathbb{H}}_n$ by its expectation $\mathbb{H}_n$, the above six expressions in (18) coincide with the covariance $\mathrm{cov}\,\big(Z_n(t,r), Z_n(t',r')\big)$ of the limiting process Z_n. Since

$$\sum_{j=1}^{k} \frac{\max_i\big(a_{r_jt_j}^{(n)}(X_i) - \bar{a}_{r_jt_j}^{(n)}\big)^2}{\sum_{i=1}^{n}\big(a_{r_jt_j}^{(n)}(X_i) - \bar{a}_{r_jt_j}^{(n)}\big)^2} \;\longrightarrow_{\mathbb{P}_n^{\otimes m}\otimes\mathbb{Q}_n^{\otimes(n-m)}} 0 \quad (n\to\infty)$$

and $\big|\,\mathrm{cov}\big(Y_n(t,r),Y_n(t',r') \,\big|\, \mathcal{X}_n\big) - \mathrm{cov}\big(Z_n(t,r),Z_n(t',r')\big)\,\big| \longrightarrow_{\mathbb{P}_n^{\otimes m}\otimes\mathbb{Q}_n^{\otimes(n-m)}} 0$ by an application of the weak law of large numbers for triangular arrays to each of the expressions in (18) separately, Hájek's multivariate Central Limit Theorem for permutation statistics yields the desired weak convergence in probability of the finite dimensional distributions. For notational convenience, define

$$T_n^{\Pi}(\delta,\delta') \;:=\; \sup_{\substack{(j,k):\\ \delta<\gamma_n(j,k)\le\delta'}} \Big\{\,\big|\,T_{jkn}\circ\Pi\,\big| \;-\; C_{jkn}\Big\}$$

and

$$S_n(\delta,\delta') \;:=\; \sup_{\substack{(t,r):\\ \delta<\gamma_n(t,r)\le\delta'}} \left\{ \frac{\bigl|\int \phi_{rt}^{(n)}(x)\,dW(x)\bigr|}{\gamma_n(t,r)} \;-\; \sqrt{2\log\bigl(1/\gamma_n(t,r)^2\bigr)} \right\}.$$

Since $\cup_{n\in\mathbb{N}}\hat{\mathcal{T}}_n$ is a.s. dense in $\mathcal{T}$ and $\sup_{(j,k):\,\gamma_n(j,k)\ge\delta}\bigl|\,C_{jkn}-(2\,\Gamma_{jkn})^{1/2}\,\bigr| \longrightarrow_{\mathbb{P}_n^{\otimes m}\otimes\mathbb{Q}_n^{\otimes(n-m)}} 0$ as $n\to\infty$, it follows from the above established weak approximation and tightness that

$$d_w\Bigl(\mathcal{L}\bigl(T_n^{\Pi}(\delta,1)\,\big|\,\mathcal{X}_n\bigr),\;\mathcal{L}\bigl(S_n(\delta,1)\bigr)\Bigr) \;\longrightarrow_{\mathbb{P}_n^{\otimes m}\otimes\mathbb{Q}_n^{\otimes(n-m)}}\; 0$$

for any fixed $\delta\in(0,1]$. An application of Theorem 6 as well as its subsequent Remark imply that

$$\lim_{\delta\searrow 0}\limsup_{n\to\infty}\mathbb{P}^*\bigl(T_n^{\Pi}(0,\delta)\ge\varepsilon\,\big|\,\mathcal{X}_n\bigr) \;=\; 0 \;\; a.s. \;\text{ and }\; \lim_{\delta\searrow 0}\limsup_{n\to\infty}\mathbb{P}\bigl(S_n(0,\delta)\ge\varepsilon\bigr) \;=\; 0$$

for any $\varepsilon>0$. Thus, because obviously $\lim_{\delta\searrow 0}\liminf_{n\to\infty}\mathbb{P}\bigl(S_n(\delta,1)\le-\varepsilon\bigr) \;=\; 0$, we obtain

$$d_w\Bigl(\mathcal{L}\bigl(T_n^{\Pi}(0,1)\,\big|\,\mathcal{X}_n\bigr),\;\mathcal{L}\bigl(S_n(0,1)\bigr)\Bigr) \;\longrightarrow_{\mathbb{P}_n^{\otimes m}\otimes\mathbb{Q}_n^{\otimes(n-m)}}\; 0. \qquad \square$$

PROOF OF THEOREM 2 Let $\mathcal{C}$ be some compact rectangle of J. Fix $\beta>0$. For any integer $k>1$ let $\mathcal{C}_{n,k}\subset\mathcal{C}$ be some maximal subset of points such that $\|x-y\|_2\ge 2k\delta_n$ and $B_x(k\delta_n)\subset\mathcal{C}$ for arbitrary different points $x,y\in\mathcal{C}_{n,k}$. Then $\sharp\mathcal{C}_{n,k}\sim(k\delta_n)^{-d}$. Now the let $\phi_{x,n}$ be the solution of the be the subsequent optimization problem:

$(*)$ Minimize $\|g\|_2$ under the constraints

$$g\in\mathcal{H}_d(\beta,L;\mathbb{R}^d),\;\; \mathrm{supp}(g)\subseteq B_x(k\delta_n),\;\; g(x)=L\delta_n^{\beta},\;\; \int g(z)\sqrt{h_n(z)}dz=0.$$

These constraints define a closed and convex set in $L_2\bigl([0,1]^d\bigr)$ which is non-empty for k sufficiently large. Consequently in the latter case, the argmin $\phi_{x,n}$ exists and is unique. The resulting density candidates

$$p_{x,n} \;=\; h_n\cdot\Bigl(1-(m/n)\phi_{x,n}/\sqrt{h_n}\Bigr) \;\text{ and }\; q_{x,n} \;=\; h_n\cdot\Bigl(1+(1-m/n)\phi_{x,n}/\sqrt{h_n}\Bigr)$$

are non-negative and thus contained in $\mathcal{F}_{h_n}^{(m,n)}$ as soon as additionally

$$-\frac{\sqrt{h_n(.)}}{1-m/n} \;\le\; \phi_{x,n}(.) \;\le\; \frac{\sqrt{h_n(.)}}{m/n} \quad \text{for all } x\in\mathcal{C}_n.$$

This is guaranteed for sufficiently large n when sequence $(\delta_n)_{n\in\mathbb{N}}$ tends to zero. For any statistical level-α-test $\psi=\psi(\beta,L,h_n):\mathbb{R}^{d\times n}\to[0,1]$ for testing the hypothesis "$\phi=0$"

it holds true that

$$\begin{aligned}\min_{x\in\mathcal{C}_n}\mathbb{E}_{(m,n,p_{x,n},q_{x,n})}\psi-\ \alpha\ &\le\ \min_{x\in\mathcal{C}_n}\mathbb{E}_{(m,n,p_{x,n},q_{x,n})}\psi-\mathbb{E}_{(m,n,h_n,h_n)}\psi\\ &\le\ \frac{1}{\sharp\mathcal{C}_n}\sum_{x\in\mathcal{C}_n}\mathbb{E}_{(m,n,p_{x,n},q_{x,n})}\psi-\mathbb{E}_{(m,n,h_n,h_n)}\psi\\ &\le\ \mathbb{E}_{(m,n,h_n,h_n)}\left|\frac{1}{\sharp\mathcal{C}_n}\sum_{x\in\mathcal{C}_n}\frac{d\mathbb{P}_{(m,n,p_{x,n},q_{x,n})}}{d\mathbb{P}_{(m,n,h_n,h_n)}}(\underline{\mathrm{X}})-1\right|. \qquad (19)\end{aligned}$$

For short we write $\mathbb{E}_0$ for $\mathbb{E}_{(m,n,h_n,h_n)}$ in the sequel. Note that the test is allowed to depend on the nuisance functional h_n (in fact the $\log$-likelihood and its distribution do). Now we aim at determing δ_n such that the right-hand-side tends to zero as n goes to infinity. Although $\lambda\big(\mathrm{supp}(\phi_{x,n})\cap\mathrm{supp}(\phi_{y,n})\big)\ =0$ for any different $x,y\in\mathcal{C}_{k,n}$, the likelihood-ratios

$$L_{x,n}\ :=\frac{d\mathbb{P}_{(m,n,p_{x,n},q_{x,n})}}{d\mathbb{P}_{(m,n,h_n,h_n)}}(\underline{\mathrm{X}})\ =\ \prod_{i=1}^{m}\Big(1-(m/n)\frac{\phi_{x,n}}{\sqrt{h_n}}(X_i)\Big)\prod_{i=m+1}^{n}\Big(1+(1-m/n)\frac{\phi_{x,n}}{\sqrt{h_n}}(X_i)\Big),$$

are not independent. However, they are independent conditional on the random vector $\Delta_n=(\Delta_{x,n})_{x\in\mathcal{C}_{k,n}}$ with entries

$$\Delta_{x,n}\ :=\ \Big(\sharp\big\{i\le m:\|X_i-x\|_2\le k\delta_n\ \big\},\ \sharp\big\{i>m:\|X_i-x\|_2\le k\delta_n\ \big\}\Big).$$

Note that $\mathbb{E}_0(L_{x,n}|\Delta_n)\ =\ \mathbb{E}_0\,L_{x,n}\ =\ 1$. Following at this point standard truncation arguments, it turns out to be sufficient for the convergence to zero of (19) to find δ_n and $\gamma=\gamma_n\in(0,1]$ such that the ratio

$$\max_{x\in\mathcal{C}_n}\frac{1}{(\sharp\,\mathcal{C}_n)^{\gamma}}\mathbb{E}_0\,L_{x,n}^{1+\gamma} \qquad (20)$$

tends to zero as n goes to infinity. But

$$\begin{aligned}\mathbb{E}_0L_{x,n}^{1+\gamma}&=\Big\{\int h_n(z)\Big(1+(1-m/n)\frac{\phi_{x,n}(z)}{\sqrt{h_n(z)}}\Big)^{1+\gamma}dz\Big\}^{m}\Big\{\int h_n(z)\Big(1-(m/n)\frac{\phi_{x,n}(z)}{\sqrt{h_n(z)}}\Big)^{1+\gamma}dz\Big\}^{n-m}\\ &=\Big\{1+\frac12\gamma(1+\gamma)\Big(1+O\big(\delta_n^{\beta}\big)\Big)(1-(m/n))^2\int_0^1\phi_{x,n}(z)^2dz\Big\}^{m}\times \qquad (21)\\ &\qquad\Big\{1+\frac12\gamma(1+\gamma)\Big(1+O\big(\delta_n^{\beta}\big)\Big)(m/n)^2\int_0^1\phi_{x,n}(z)^2dz\Big\}^{n-m},\end{aligned}$$

using the bound $(1+\Delta)^{1+\gamma}\le 1+(1+\gamma)\Delta+2^{-1}\gamma(1+\gamma)\Delta^2+3\gamma\Delta^2|\Delta|$ for $|\Delta|\le 1$. Now let $\tilde{\phi}_k$ be the solution to the following optimization problem

($**$) Minimize $\|g\|_2$ subject to

$$g\in\mathcal{H}_d(\beta,L;\mathbb{R}^d),\ \ \mathrm{supp}(g)\subseteq B_0(k),\ \ g(0)=1,\ \ \int g(x)dx=0. \qquad (22)$$

Notice the rescaling property $L\delta_n^\beta g(./\delta_n) \in \mathcal{H}_d(\beta, L; \mathbb{R}^d)$ with $\mathrm{supp}(g) = B_0(k\delta_n)$ and $g(0) = L\delta_n^\beta \Leftrightarrow g \in \mathcal{H}_d(\beta, L; \mathbb{R}^d)$ with $\mathrm{supp}(g) = B_0(k)$ and $g(0) = 1$. Recall from the previous proof that the sequence $(h_n)_{n\in\mathbb{N}}$ is relatively compact in the uniform topology, in particular we have (proof via contradiction)

$$\lim_{\delta\searrow 0} \sup_{x\in B_z(\delta)} \sup_n \big| h_n(x) - h_n(z) \big| = 0,$$

whence

$$\int \phi_{x,n}(z)^2\, dz \; = \; \big(1 + o(1)\big) L^2 \delta_n^{2\beta+d} \|\tilde{\phi}_k\|_2^2 \tag{23}$$

because the minimum in $(*)$ depends continuously on the mixed density h_n as can be seen using a Lagrange multiplier for the centering constraint. Note that the $o(1)$-term is uniformly in $x \in \mathcal{C}_{k,n}$. Now the combination of (21) and (23) shows that for δ_n sufficiently small, (20) is bounded by

$$\exp\bigg(n(m/n)(1-m/n)\frac{1}{2}\gamma(1+\gamma)L^2\delta_n^{2\beta+d}\|\tilde{\phi}_k\|_2^2\big(1+o(1)\big) - \gamma\log(\sharp\,\mathcal{C}_{k,n})\bigg).$$

By construction, $\sharp\,\mathcal{C}_{k,n} \geq d_k \cdot \delta_n^{-d}$ for some constant $d_k > 0$. Now fix $\delta > 0$ and define

$$c_k(\beta, L) \; := \; \left(\frac{2\,d\,L^{d/\beta}}{(2\beta+d)\|\tilde{\phi}_k\|_2^2}\right)^{\beta/(2\beta+d)}.$$

Observe that the sequence $c_k(\beta, L)$ is increasing in k. We need to check that $\lim_{k\to\infty}\|\tilde{\phi}_k\|_2 = \|\gamma_\beta\|_2$. Note that in contrast to (22), the solution of (2) does not integrate to zero in general and it remains still open if γ_β is compactly supported for $d \geq 2$ and $\beta > 1$. Starting from γ_β, it is sufficient to construct a sequence $\tilde{\gamma}_{\beta,k}$ satisfying the constraints of the optimization problem $(**)$ such that $\lim_{k\to\infty}\|\tilde{\gamma}_{\beta,k}\|_2 = \|\gamma_\beta\|_2$. Then the equality $\lim_{k\to\infty}\|\tilde{\phi}_k\|_2 = \|\gamma_\beta\|_2$ follows from $\|\tilde{\gamma}_{\beta,k}\|_2 \geq \|\tilde{\phi}_k\|_2$. The existence is sketched in the appendix. As a consequence there exists some $k' \in \mathbb{N}$ such that $c(\beta, L)(1-\delta) > c_{k'}(\beta, L)$. Now one verifies that the lower bound is established with the choice

$$\delta_n \; := \; \left(\frac{c_{k'}(\beta, L)\rho_n}{L}\right)^{1/\beta}.$$

and some sequence $\gamma = \gamma_n \to 0$ with $\lim_n \gamma_n(\log n)^{1/2} = \infty$. □

PROOF OF THEOREM 3 By virtue of Theorem 1, the sequence $\mathcal{L}\big(T_n \circ \Pi \,\big|\, \mathcal{X}_n\big)$ is tight in $\big(\mathbb{P}_n^{\otimes m} \otimes \mathbb{Q}_n^{\otimes(n-m)}\big)$-probability, resulting in stochastic boundedness of the sequence of random quantiles $\big(\kappa_\alpha(\underline{\mathrm{X}})\big)_{n\in\mathbb{N}}$. The bounded total variation of the kernel for $\beta \leq 1$ is a consequence of its monotonicity, for $\beta > 1$ it results from the continuous differentiability of $\psi_{\beta,K}$ and its compact support. For notational convenience the dependency on β and K is suppressed. They are arbitrary but fixed unless stated otherwise. First note that for any random couple $(\hat{j}_n, \hat{k}_n)$ it holds true that

$$\mathbb{P}_{(m,n,p_n,q_n)}\Big(T_n > \kappa_\alpha(\underline{\mathrm{X}})\Big) \;\geq\; \mathbb{P}_{(m,n,p_n,q_n)}\Big(T_{\hat{j}_n\hat{k}_n n} - C_{\hat{j}_n\hat{k}_n n} \;>\; \kappa_\alpha(\underline{\mathrm{X}})\Big).$$

Hence it is sufficient to prove that for any sequence $(\phi_n)_{n\in\mathbb{N}}$ of admissible alternatives there exists a random sequence of $(\hat{j}_n, \hat{k}_n)_{n\in\mathbb{N}}$ with $T_{\hat{j}_n\hat{k}_n n} - C_{\hat{j}_n\hat{k}_n n} \longrightarrow_{\mathbb{P}^{\otimes m}\otimes\mathbb{Q}^{\otimes(n-m)}} \infty$. As in the proof of Theorem 1 define $\gamma_n(t,r)^2 := \mathbb{E}\hat{\gamma}_{2,n}(t,r)^2 - \big(\mathbb{E}\hat{\gamma}_{1,n}(t,r)\big)^2$, $(t,r) \in \mathcal{T}$. Let $t_n := \mathrm{argmax}_{x\in J}\,|\phi_n(x)|$ and $r_n := \big(\|\phi_n\|_{\sup}/L\big)^{1/\beta}$. Define $\big(\hat{t}_n, \hat{r}_n\big) := \big(X_{\hat{j}_n}, \big\|X_{\hat{j}_n} - X_{\hat{k}_n}\big\|_2\big)$ with

$$(\hat{j}_n, \hat{k}_n) \;:=\; \operatorname*{argmin}_{j,k=1,\dots,n} \lambda\Big(B_{t_n}(r_n)\,\Delta\, B_{X_j}\big(\|X_j - X_k\|_2\big)\Big).$$

Now let the process S_n on $\mathcal{T}$ pointwise be defined by

$$S_n\big(t,r\big) \;:=\; \frac{\sqrt{\lambda_n(1-\lambda_n)}}{\sqrt{n}} \sum_{i=1}^{n} \psi\Big(\frac{\|X_i - t\|_2}{r}\Big)\Lambda(X_i).$$

Furthermore, let us introduce the random variables $(\hat{t}_{ni}, \hat{r}_{ni})$, based on the indices $(\hat{j}_{ni}, \hat{k}_{ni})$ which are defined analogously to $(\hat{j}_n, \hat{k}_n)$ but with the minimum running over the set $j,k \in \{1,...,n\} \setminus \{i\}$ only. Then

$$\begin{aligned}
&\frac{1}{\gamma_n(t_n,r_n)}\Big|\,\mathbb{E}\Big(S_n(\hat{t}_n,\hat{r}_n) - S_n(t_n,r_n)\Big)\,\Big| \\
&\quad= \frac{1}{\gamma_n(t_n,r_n)}\frac{1}{\sqrt{n}}\,\Bigg|\,\frac{n}{m}\sum_{i=1}^{m}\mathbb{E}\Big(\psi_{\hat{t}_n\hat{r}_n}(X_i) - \psi_{t_n r_n}(X_i)\Big) \\
&\qquad\qquad - \frac{n}{n-m}\sum_{i=m+1}^{n}\mathbb{E}\Big(\psi_{\hat{t}_n\hat{r}_n}(X_i) - \psi_{t_n r_n}(X_i)\Big)\Bigg| \\
&\quad\le \frac{1}{\gamma_n(t_n,r_n)}\frac{1}{\sqrt{n}}\,\Bigg|\,\frac{n}{m}\sum_{i=1}^{m}\mathbb{E}\Big(\psi_{\hat{t}_n\hat{r}_n}(X_i) - \psi_{\hat{t}_{ni}\hat{r}_{ni}}(X_i)\Big) \\
&\qquad\qquad - \frac{n}{n-m}\sum_{i=m+1}^{n}\mathbb{E}\Big(\psi_{\hat{t}_n\hat{r}_n}(X_i) - \psi_{\hat{t}_{ni}\hat{r}_{ni}}(X_i)\Big)\Bigg| \\
&\qquad + \frac{1}{\gamma_n(t_n,r_n)}\frac{1}{\sqrt{n}}\,\Bigg|\,\frac{n}{m}\sum_{i=1}^{m}\mathbb{E}\Big(\psi_{\hat{t}_{ni}\hat{r}_{ni}}(X_i) - \psi_{t_n r_n}(X_i)\Big) \\
&\qquad\qquad - \frac{n}{n-m}\sum_{i=m+1}^{n}\mathbb{E}\Big(\psi_{\hat{t}_{ni}\hat{r}_{ni}}(X_i) \;-\; \psi_{t_n r_n}(X_i)\Big)\Bigg| \\
&\quad\le \frac{1}{\gamma_n(t_n r_n)}\frac{4}{\sqrt{n}}\|\psi\|_{\sup}\max\Big(\frac{n}{m},\frac{n}{n-m}\Big) \\
&\qquad + \frac{1}{\gamma_n(t_n,r_n)}\frac{1}{\sqrt{n}}\,\Bigg|\,\mathbb{E}\Bigg\{\frac{n}{m}\sum_{i=1}^{m}\int\Big(\psi_{\hat{t}_{ni}\hat{r}_{ni}}(x) - \psi_{t_n r_n}(x)\Big)p_n(x)dx \qquad (24) \\
&\qquad\qquad - \frac{n}{n-m}\sum_{i=m+1}^{n}\int\Big(\psi_{\hat{t}_{ni}\hat{r}_{ni}}(x) - \psi_{t_n r_n}(x)\Big)q_n(x)dx\Bigg\}\Bigg|,
\end{aligned}$$

whereby we used for the first term in the last inequality that $(\hat{t}_{ni}, \hat{r}_{ni})$ differs from $(\hat{t}_n, \hat{r}_n)$ for at most two indices $i, j \in \{1, ..., n\}$; the second term follows by including and evaluating the conditional expectation given $(\hat{t}_{ni}, \hat{r}_{ni})$ as X_i is independent of $(\hat{t}_{ni}, \hat{r}_{ni})$. Replacing again $(\hat{t}_{ni}, \hat{r}_{ni})$ by $(\hat{t}_n, \hat{r}_n)$, the second expression behind the inequality in formula (24) is bounded by

$$\frac{1}{\gamma_n(t_n r_n)} \frac{4}{\sqrt{n}} \|\psi\|_{\sup} \max\Big(\frac{n}{m}, \frac{n}{n-m}\Big) \\ + \frac{\sqrt{n}}{\gamma_n(t_n, r_n)} \left| \mathbb{E}\Big[\int \Big(\psi_{\hat{t}_n \hat{r}_n}(x) - \psi_{t_n r_n}(x)\Big)\Big(p_n(x) - q_n(x)\Big) dx \Big] \right|. \quad (25)$$

Now we can make use of the fact that $\big| p_n(x) - q_n(x) \big| = \big| \phi_n(x) \sqrt{h_n(x)} \big| \le C \|\phi_n\|_{\sup}$ with $C := \sup_n \sup_x \big| \sqrt{h_n(x)} \big|$. Recall that $\|h_n\|_{\sup}$ is uniformly bounded due to the equicontinuity assumption on $(h_n)_{n \in \mathbb{N}}$ and the constraint on the L_1-norm $\|h_n\|_1 = 1$, whence the term in (25) is not greater than

$$C \frac{\sqrt{n} \|\phi_n\|_{\sup}}{\gamma_n(t_n, r_n)} \mathbb{E}\Big(\int \Big| \psi_{\hat{t}_n \hat{r}_n}(x) - \psi_{t_n r_n}(x) \Big| dx \Big). \quad (26)$$

Using the bounded total variation $TV(\psi)$ of ψ and M_x and μ as defined in the proof of Theorem 1, the integral which appears in (26) can be bounded by

$$\begin{aligned} &\mathbb{E}\Big(\int \Big| \psi_{\hat{t}_n \hat{r}_n}(x) - \psi_{t_n r_n}(x) \Big| dx \Big) \\ &\le \mathbb{E}\Big(\int \mu\big(M_x(t_n, r_n, \hat{t}_n, \hat{r}_n)\big) dx \Big) \\ &= \mathbb{E}\Big(\int \int I\big\{ y \in M_x(t_n, r_n, \hat{t}_n, \hat{r}_n) \big\} dx d\mu(y) \Big) \qquad \text{(Fubini)} \\ &\le TV(\psi)\, \mathbb{E} \sup_{y \in [0,1]} \Big(\int I\big\{ y \in M_x(t_n, r_n, \hat{t}_n, \hat{r}_n) \big\} dx \Big) \\ &\le TV(\psi)\, \mathbb{E}\, \lambda\Big(B_{t_n}(r_n) \Delta B_{\hat{t}_n}(\hat{r}_n) \Big) \\ &\le TV(\psi)\, \mathbb{E}\, \lambda\Big(B_{t_n}(r_n) \Delta B_{\hat{t}_n}(r_n) \Big) + TV(\psi)\, \mathbb{E}\, \lambda\Big(B_{\hat{t}_n}(r_n) \Delta B_{\hat{t}_n}(\hat{r}_n) \Big) \\ &= O\big(r_n^{d-1} n^{-1/d} \big), \end{aligned} \quad (27)$$

using in the last inequality besides the stochastic convergence rates of $\hat{t}_n$ and $\hat{r}_n$ the uniform integrability of the sequences $\big(n^{1/d} \|\hat{t}_n - t_n\|_2\big)$, $\big(n^{1/d} |\hat{r}_n - r_n|\big)$ which result from $\mathbb{P}\big(\|\hat{t}_n - t_n\|_2 > x\big) \sim (1 - x^d)^n$ and $\mathbb{P}\big(|\hat{r}_n - r_n| > x\big) \le 2\, \mathbb{P}\big(\|\hat{t}_n - t_n\|_2 > x\big)$. Together with (24) - (26) this shows that for any sequence of admissible alternatives $(\phi_n)_{n \in \mathbb{N}}$

$$\frac{\big| \mathbb{E}\big(S_n(\hat{t}_n, \hat{r}_n) - S_n(t_n, r_n) \big) \big|}{\gamma_n(t_n, r_n)} = O\Big(r_n^{d/2 - 1 + \beta} n^{-1/d + 1/2} \Big). \quad (28)$$

If in particular $\|\phi_n\|_{\sup} = O\Big(\big((\log n)/n \big)^{\beta/(2\beta + d)} \Big)$, (28) is $O\Big((\log n)^{(\beta + d/2 - 1)/(2\beta + d)} n^{-(2\beta/d)/(2\beta + d)} \Big)$. Compared to (26), note at this point that $\sqrt{n}\, \gamma_n(t_n, r_n)^{-1} \mathbb{E} \int \big| \psi_{\hat{t}_n \hat{r}_n}(x) - \psi_{t_n r_n}(x) \big| dx$ is not even of order $\sqrt{\log n}$ if $\|\phi_n\|_{\sup}$ decreases to zero at the fastest possible rate as soon as $d > 2$. We need to check that

$$\frac{\gamma_n(t_n, r_n)}{\hat{\gamma}_n(\hat{t}_n, \hat{r}_n)} \longrightarrow_{\mathbb{P}^{\otimes m} \otimes Q^{\otimes(n-m)}} 1. \tag{29}$$

For this we use the decomposition $[(n+1)/n]\hat{\gamma}_n(t,r)^2 = \hat{\gamma}_{n,2}(t,r)^2 - \hat{\gamma}_{n,1}(t,r)^2$ and consider subsequently $i = 1$ only, the orther case is done analogously (taking the square). To this end note first that

$$\begin{aligned}
&\big|\, \hat{\gamma}_{n,1}(\hat{t}_n, \hat{r}_n) - \hat{\gamma}_{n,1}(t_n, r_n) \,\big| \\
&\quad\leq \big\| \psi_{\hat{t}_n \hat{r}_n} - \psi_{t_n r_n} \big\|_{\sup} \frac{1}{n} \sum_{i=1}^{n} I\Big\{ X_i \in B_{\hat{t}_n}\big(\hat{r}_n\big) \cap B_{t_n}\big(r_n\big) \Big\} \\
&\qquad + 2\|\psi\|_{\sup} \frac{1}{n} \sum_{i=1}^{n} I\Big\{ X_i \in B_{\hat{t}_n}\big(\hat{r}_n\big) \Delta B_{t_n}\big(r_n\big) \Big\} \\
&\quad\leq \big\| \psi_{\hat{t}_n \hat{r}_n} - \psi_{t_n r_n} \big\|_{\sup} \frac{1}{n} \sum_{i=1}^{n} I\Big\{ X_i \in B_{t_n}\big(r_n\big) \Big\} \\
&\qquad + 2\|\psi\|_{\sup} \frac{1}{n} \sum_{i=1}^{n} I\Big\{ X_i \in B_{\hat{t}_n}\big(\hat{r}_n\big) \Delta B_{t_n}\big(r_n\big) \Big\} \\
&\quad= o_p(1) O_p(r_n^d) + O_p\big(r_n^{d-1} n^{-1/d}\big) = o_p\Big(\gamma_{n,1}(t_n, r_n)\Big).
\end{aligned}$$

The "$o_p(1)$"-term results from the Hölder continuity of ψ (for $\beta > 1$ the first derivative of ψ is uniformly bounded on $[-K, K]$) and the fact that $r_n > \big(c(\beta, L)\rho_{m,n}/L\big)^{1/\beta}$ while $\hat{t}_n - t_n \sim n^{-1/d}$, $\hat{r}_n - r_n \sim n^{-1/d}$. To verify (29) it remains to be shown that $\hat{\gamma}_n(t_n, r_n)/\gamma_n(t_n, r_n) - 1 = o_p(1)$ which however is a simple consequence of Chebychef's inequality since for any $\beta > 0$ and any sequence of admissible alternatives $(\phi_n)_{n \in \mathbb{N}}$, the sequence $\gamma_n(t_n, r_n) \sim r_n^{d/2}$ decreases (if it decreases) at a slower rate than $n^{-1/2}$. The above considerations show in particular that

$$\begin{aligned}
C_{\hat{j}_n \hat{k}_n n} &= \frac{3\, R_\psi(m,n)}{\sqrt{n}\, \hat{\gamma}_n(\hat{t}_n, \hat{r}_n)} \delta(m,n) \log\Big(\hat{\gamma}_n(\hat{t}_n, \hat{r}_n)^{-2}\Big) + \delta(m,n) \sqrt{2 \log\Big(\hat{\gamma}_n(\hat{t}_n, \hat{r}_n)^{-2}\Big)} \\
&= \sqrt{2 \log\Big(\gamma_n(t_n, r_n)^{-2}\Big)} + o_p(1),
\end{aligned}$$

using in addition that $\delta(m,n) = 1 + O(n^{-1/2})$. Consequently,

$$T_{\hat{j}_n \hat{k}_n n} - C_{\hat{j}_n \hat{k}_n n} = O_p(1) + \frac{\mathbb{E} S_n(t_n, r_n)}{\gamma_n(t_n, r_n)} \Big(1 + o_p(1)\Big) - \sqrt{2 \log\Big(\gamma_n(t_n, r_n)^{-2}\Big)}, \tag{30}$$

and it has to be verified that the latter quantity goes to infinity. Recall that

$$\begin{aligned}
\mathbb{E} S_n(t_n, r_n) &= \sqrt{n} \sqrt{\lambda_n(1 - \lambda_n)} \int_{[0,1]^d} \psi_{t_n r_n}(x) \Big(p_n(x) - q_n(x)\Big) dx \\
&= \sqrt{n} \sqrt{\lambda_n(1 - \lambda_n)} \int_{[0,1]^d} \psi_{t_n r_n}(x) \phi_n(x) \sqrt{h_n(x)} dx
\end{aligned}$$

and analogously

$$\gamma_n(t_n,r_n)^2 \;=\; \int_{[0,1]^d} \psi_{t_n r_n}(x)^2 h_n(x)dx \;-\; \Big(\int_{[0,1]^d} \psi_{t_n r_n}(x) h_n(x)dx\Big)^2$$
$$\;=\; \Big(1+O(r_n^d)\Big)\int_{[0,1]^d} \psi_{t_n r_n}(x)^2 h_n(x)dx. \tag{31}$$

We first assume that $r_n = o(1)$, i.e. $\|\phi_n\|_{\sup} = o(1)$. Using that

$$\lim_{\delta\searrow 0}\sup_n \sup_{t\in[0,1]^d}\sup_{x\in B_t(r_n)} \big|\,\phi_n(x)-\phi_n(t)\,\big| \;=\; 0,$$

which follows by the same argument as used in Theorem 2 and the fact that any sequence of centers $(t_n)_{n\in\mathbb{N}}$ has a convergent subsequence by the compactness of $[0,1]^d$,

$$\frac{\mathbb{E}S_n(t_n,r_n)}{\gamma_n(t_n,r_n)} \;=\; \sqrt{n}\sqrt{\lambda_n(1-\lambda_n)}\frac{\int_{[0,1]^d}\psi_{t_n r_n}(x)\phi_n(x)dx}{\big[\int_{[0,1]^d}\psi_{t_n r_n}(x)^2 dx\big]^{1/2}}\big(1+o(1)\big). \tag{32}$$

Using the approximation in (31) we obtain analogously

$$\sqrt{2\log\Big(\gamma_n(t_n,r_n)^{-2}\Big)} \;=\; \Big[2\log\Big(1\Big/O(1)\int_{[0,1]^d}\psi_{t_n r_n}(x)^2 dx\Big)\Big]^{1/2}. \tag{33}$$

Recall that $\psi=\psi_{\beta,K}$ with K the bound of the support. Standard calculation shows that the bounded L_2-norm of γ_β implies

$$\frac{\big|\int\psi_{t_n r_n;\beta,K}(x)\phi_n(x)dx\,\big|}{\big[\int\psi_{t_n r_n;\beta,K}(x)^2dx\big]^{1/2}} \;=\; \frac{\big|\int\psi_{t_n r_n;\beta}(x)\phi_n(x)dx\,\big|}{\big[\int\psi_{t_n r_n;\beta}(x)^2dx\big]^{1/2}}\big(1+c_k\big)\;\text{ with } c_K\to 0 \text{ as } K\to\infty,$$

but note that the total variation $TV(\psi_{\beta,K})$ is increasing in K. Define now $\delta_n := (1+\delta)c(\beta,L)\rho_{m,n}$. Then by its construction, $\delta_n\psi_{t_n r_n;\beta}\in\mathcal{H}_d\big(\beta,L;\mathbb{R}^d\big)$. Moreover, by the closedness in L_2 and the convexity of the sets $\big\{\phi\in\mathcal{H}_d(\beta,L;\mathbb{R}^d)\,:\,\phi(t_n)\geq\delta_n\big\}$ and $\big\{\phi\in\mathcal{H}_d(\beta,L;\mathbb{R}^d)\,:\,\phi(t_n)\leq -\delta_n\big\}$, it results finally from convex analysis and the definition of γ_β that

$$\frac{\big|\int\psi_{t_n r_n;\beta}(x)\phi_n(x)dx\,\big|}{\big[\int\psi_{t_n r_n;\beta}(x)^2dx\big]^{1/2}} \;\geq\; \frac{\delta_n^{-1}\|\delta_n\psi_{t_n r_n;\beta}\|_2^2}{\|\psi_{t_n r_n;\beta}\|_2} \;=\; \delta_n r_n^{d/2}\|\gamma_\beta\|_2.$$

Combining (31) – (33), one verifies for the expression of the right hand side in (30) that it is possesses the approximation

$$(30) \;=\; O_p(1) \;+\; \sqrt{n}\sqrt{\lambda_n(1-\lambda_n)}\delta_n r_n^{d/2}\|\gamma_\beta\|_2\big(1+c_K\big) \;-\; \Big(\frac{2d}{2\beta+d}\Big)^{1/2}\sqrt{\log\big(n/\log n\big)}$$
$$\;=\; O_p(1) \;+\; \sqrt{\log n}\Big(\frac{2dL^{d/\beta}}{(2\beta+d)\|\gamma_\beta\|_2^2}\Big)^{1/2} L^{-d/(2\beta)}\|\gamma_\beta\|_2(1+c_K)(1+\delta)^{d/(2\beta)+1}$$
$$-\;\Big(\frac{2d}{2\beta+d}\Big)^{1/2}\sqrt{\log\big(n/\log n\big)},$$

which goes to infinity for K sufficiently large. If there exists a sequence $(\phi_n)_{n\in\mathbb{N}}$ of admissible alternatives such that $\limsup_{n\to\infty} \mathbb{P}_{(m,n,p_n,q_n)}\big(T_n > \kappa_\alpha(\underline{\mathrm{X}})\big) < 1$, there exists by the considerations above a subsequence (for simplicity also denoted by (n)) along which $\|\phi_n\|_{\sup}$ stays uniformly bounded away from zero. But the bounds (28) and (29) show that

$$\frac{\mathbb{E}S_n(\hat{t}_n, \hat{r}_n) - \mathbb{E}S_n(t_n, r_n)}{\gamma_n(\hat{t}_n, \hat{r}_n)} \;=\; O\Big(n^{-1/d+1/2}\Big)\big(1 + o_p(1)\big),$$

as well as the logarithmic correction term $C_{\hat{\jmath}_n\hat{k}_n n}$ are in this case of smaller order than $|\mathbb{E}S_n(t_n, r_n)|$, which concludes the proof by contradiction. □

PROOF OF THEOREM 4 Following the considerations of the proof of Theorem 3, it has to be established that there exist random sequences $\big(\hat{\jmath}_{ni}, \hat{k}_{ni}\big)_{n\in\mathbb{N}}$ with $B_{X_{\hat{\jmath}_{ni}}}\big(\big\| X_{\hat{\jmath}_{ni}} - X_{\hat{k}_{ni}}\big\|_2\big) \subset J_i$, $i = 1, ..., k$, such that for any sequence of alternatives as formulated in Theorem 4 and any fixed $K > 0$

$$\liminf_{n\to\infty} \mathbb{P}_{(m,n,p_n,q_n)}\Big(T_{\hat{\jmath}_{ni}\hat{k}_{ni}} - C_{\hat{\jmath}_{ni}\hat{k}_{ni}} > \kappa_\alpha(\underline{\mathrm{X}})\Big) \;=\; 1,\;\; i = 1, ..., k.$$

Then the result follows because the finite intersection of sets with asymptotic probability equal to 1 has asymptotically mass 1 as well. Inspired by the arguments in Rohde (2008) for the univariate regression context, we first establish the following:

For $\phi_n \in \mathcal{H}_d\big(\beta, L; [0,1]^d\big)$ with $\|\phi_n\|_{\sup} \le 1$ and $x^* = \mathrm{argmax}_{x\in[0,1]^d} |\phi_n(x)|$, there exists some constant $c = c(\beta, L) > 0$ and a compact ball $B = B(\phi_n) \subset \mathbb{R}^d$ with center x^* such that

$$\lambda\big(B \cap [0,1]^d\big) \ge\; c|\phi_n(x^*)|^{d/\beta} \;\text{ and }\; \big|\,\phi_n(x)\,\big| \;\ge\; \frac{1}{2}\,\big|\,\phi_n(x^*)\,\big| \;\text{ for all } x \in B \cap [0,1]^d. \tag{34}$$

For let us assume that $\beta > 1$ (the above inequality is trivial in case $\beta \le 1$). For let $\lfloor\beta\rfloor$ denote the largest integer strictly smaller than β and $\phi \in \mathcal{H}_d\big(\beta, L; [0,1]^d\big)$ with $\|\phi\|_{\sup} = D > 0$. With $j = (j_1, ..., j_d) \in \mathbb{N}_0^d$ we denote subsequently some multi-index, where $|j| = j_1 + ... + j_d$ defines its length, $j! := \prod_{i=1}^d j_i!$ the product of faculties, $(x-y)^j := \prod_{i=1}^d (x_i - y_i)^{j_i}$ and

$$D^j \;:=\; \frac{\partial^{|j|}}{\partial x_1^{j_1} \cdot ... \cdot \partial x_m^{j_m}}$$

the partial derivative operator. Taylor expansion around any point $y \in [0,1]^d$ provides the the approximation

$$\phi(x) \;=\; \sum_{|j|\le\lfloor\beta\rfloor} \frac{D^j\phi(y)}{j!}(x-y)^j \;+\; R_\phi(x,y)$$

with remainder term $|R_\phi(x,y)| \;\le\; L\|x-y\|_2^\beta$ by definition of $\mathcal{H}_d\big(\beta, L; [0,1]^d\big)$. In particular, these considerations entail that the polynomial in x for any fixed y

$$\sum_{|j|\le\lfloor\beta\rfloor} \frac{D^j\phi(y)}{j!}(x-y)^j \tag{35}$$

is bounded in sup-norm over $[0,1]^d$ by $2D+L\sqrt{d}^{\,\beta}$. In order to establish (34), note that for any polynomial $P = \sum_{|j|\le\lfloor\beta\rfloor} a_j x^j$, the topology induced by the metrics corresponding to the two norms $\|P\|_{(1)} = \sup_{x\in[0,1]^d}|P(x)|$ and $\|P\|_{(2)} := \max_j |a_j|$ respectively on the ring of polynomials of total degree $\lfloor\beta\rfloor$ on $[0,1]^d$ is the topology of uniform convergence, hence these two norms are equivalent. Consequently, the boundedness of the polynomial in (35) by $2D+L\sqrt{d}^{\,\beta}$ uniformly in y implies that there exists some constant $C=C(\beta)$ such that $\|D^j\phi\|_{\sup} \le C\big(2D+L\big)$ for all multi-indices j with $|j|\le\lfloor\beta\rfloor$. Now the Mean Value Theorem implies for some intermediate point $z\in\big\{x+t(x^*-x); 0\le t\le 1\big\}$

$$
\begin{aligned}
\big|\,\phi(x)-\phi(x^*)\,\big| \;&=\; \big|\,\big(\nabla\phi(z)\big)^T\big(x-x^*\big)\,\big| \\
&\le\; \sqrt{d}\sup_{j:\,|j|=1}\big\|\,D^j\phi\,\big\|_{\sup}\|x-x^*\|_2 \\
&\le\; \sqrt{d}\,C\,\big(2D+L\big)\|x-x^*\|_2.
\end{aligned}
$$

Thus,

$$
|\phi(x)| \;\ge\; \frac{1}{2}|\phi(x^*)| \;\text{ for all } x \text{ in } B_{x^*}\bigg(\frac{D}{2\sqrt{d}C(2D+L)}\bigg)\cap[0,1]^d.
$$

If $\phi\in\mathcal{H}_d\big(\beta,L;[0,1]^d\big)$ with $\|\phi\|_{\sup}=\delta\le 1$, then the function g_δ, for $x\in[0,1]^d$ pointwise defined by $g_\delta(x) := \delta^{-1}\phi\big(\delta^{1/\beta}x+x^*\big)\cdot I\big\{\delta^{1/\beta}x+x^*\in[0,1]^d\big\}$ is element of $\mathcal{H}_d(\beta,L;\mathrm{supp}(g_\delta))$ with $\|g_\delta\|_{\sup}=1$. Note that $\mathrm{supp}(g_\delta)$ is a convex set. Therefore, the above considerations imply that $|\phi(x)|\ge\delta/2$ on

$$
B_{x^*}\bigg(\frac{\delta^{1/\beta}}{2\sqrt{d}C(2+L)}\bigg)\;\cap\;[0,1]^d.
$$

But then its Lebesgue measure is always greater than $c|\delta|^{d/\beta}$ for some constant $c = c(\beta,L)$, independent of δ and x^*, hence (34) is established.

Let now $\beta_i, L_i\in(0,\infty)$ fixed but arbitrary, $J_i\subset[0,1]^d$ some nondegenerate rectangle, ϕ_n a sequence of functions with $\phi_{n|J_i}\in\mathcal{H}_d\big(\beta_i,L_i;J_i\big)$. It has to be shown that there exists a universal constant $k_i=k_i(\beta_i,L_i,c)$ such that $T_{\hat{j}_n\hat{k}_n n}-C_{\hat{j}_n\hat{k}_n n}\rightarrow_{\mathbb{P}^{\otimes m}\otimes\mathbb{Q}^{\otimes(n-m)}}\infty$ whenever $\|\phi_n\|_{J_i}\ge k_i\rho_{m,n}$. First, we choose a compact ball $B_i(\phi_n)$ with center $x_i^* := \mathrm{argmax}_{t\in J_i}|\phi_n(t)|$ satisfying $\lambda(B_i(\phi_n)\cap J_i)\ge c|\phi_n(x_i^*)|^{d/\beta}$ and (34). Let the couple $(\hat{t}_n,\hat{r}_n) := \big(X_{\hat{j}_n}, \|X_{\hat{j}_n}-X_{\hat{k}_n}\|_2\big)$ be defined by

$$
(\hat{j}_n,\hat{k}_n) \;:=\; \operatorname*{argmin}_{j,k\in\{1,\ldots,n\}}\lambda\Big(B_{X_j}\Big(\|X_j-X_k\|_2\Big)\,\Delta\,B_i(\phi_n)\Big).
$$

Consulting the proof of Theorem 3, this definition of $(\hat{t}_n,\hat{r}_n)$ allows for an approximation as in (30). Since $|\phi_n(x)|\ge 2^{-1}\|\phi_n\|_{J_i}$ for all $x\in B_i(\phi_n)\cap B_{\hat{t}_n}(\hat{r}_n)\cap J_i$,

$$
\frac{\mathbb{E}S_n(t_n,r_n)}{\gamma_n(t_n,r_n)} \;\ge\; \frac{1}{2}\|\phi_n\|_{J_i}\frac{\min_x h_n(x)}{\max_x h_n(x)}\mathbb{E}\lambda\Big(B_i(\phi_n)\cap B_{\hat{t}_n}(\hat{r}_n)\cap[0,1]^d\Big)^{1/2} \;\ge\; C\|\phi_n\|_{J_i}^{(\beta+d/2)/\beta}\big(1+o(1)\big).
$$

Now the asserted result is easily deduced for a sufficiently large constant k_i. $\square$

9 Appendix

We start with a basic but useful property of the solution to (3).

Lemma 4. *If the solution to (3) is not of bounded support, it has infinitely many crossings of zero. In that case, the lower isotonic and upper antitonic envelopes of $\psi_\beta I\{\cdot \geq 0\}$ are strictly monotone and vanishing in $+\infty$.*

PROOF The first part is obvious: from any local extremal point, we may extend the function in a monotonic way by minimizing its absolute value pointwise under the constraint that $\psi\big(\|.\|_2\big)$ belongs to $\mathcal{H}_d(\beta, L; \mathbb{R})$ and end finally up in zero. Since the L_2-norm of the solution (3) is finite and if there exists a sequence of local extrema of ψ_β which stays uniformly bounded away from zero, their width must be bounded by a zero sequence. But now the result follows via contradiction of (34), which, of course, is also applicable for local extrema. $\square$

Let $\varepsilon > 0$ be fixed. Define t_ε to be a positive real number such that the following conditions are satisfied: t_ε is a local extremal point, $\int_{B_{t_\varepsilon}(0)} \gamma_\beta(x)^2 d(x) \geq (1-\varepsilon/2)\|\gamma_\beta\|_2^2$, $\|\psi_\beta\|_{[t_\varepsilon,\infty)} \leq \varepsilon/2$. Now extend the function $\psi_\beta I\{\cdot \leq t_\varepsilon\}$ to a compactly supported function G_ε such that $G_\varepsilon \in \mathcal{H}_1(\beta, 1; \mathbb{R})$, G_ε crosses zero at most once for $t > t_\varepsilon$, $\int G_\varepsilon\big(\|x\|_2\big)dx = 0$ and $\int_{\mathbb{R}^d \setminus B_{t_\varepsilon}(0)} G_\varepsilon\big(\|x\|_2\big)^2 dx$ smaller than $\varepsilon\|\gamma_\beta\|_2^2$ (which is possible for t_ε sufficiently large, we omit an explicit construction at this point). With ε sufficiently small, this construction leads to what was required in the proof of Theorem 2.

Acknowledgments Lutz Dümbgen's contribution to the decoupling subsequent to an extended discussion in Bern is gratefully acknowledged. I would also like to thank Sasha Tsybakov for his interest in this article and his encouragement when I visited him in Paris in March 2008.

References

BEHNEN, K., NEUHAUS, G. AND RUYMGAART, F. (1983). Two sample rank estimators of optimal nonparametric score-functions and corresponding adaptive rank statistics. *Ann. Statist.* ***11***, 588–599.

BELOMESTNY, D. AND SPOKOINY, V. (2007). Spatial aggregation of local likelihood estimates with application to classification. *Ann. Statist.* ***35***, 2287–2311.

BENNETT, G. (1962). Probability inequalities for sums of independent random variables. *J. Amer. Statist. Assoc.* ***57***, 33–45.

DONOHO, D. (1994a). Statistical estimation and optimal recovery. *Ann. Statist.* ***22***, 238–270.

DONOHO, D. (1994b). Asymptotic minimax risk for sup-norm loss – solution via optimal recovery. *Probab. Theory and Related Fields.* ***99***, 145–170.

DUCHARME, G.R. AND LEDWINA. T. (2003). Efficient and adaptive nonparametric test for the two-sample problem. *Ann. Statist.* ***31***, 2036–2058.

DUDLEY, R.M., GINÉ, E. AND ZINN, J. (1991) Uniform and universal Glivenko-Cantelli classes. *J. Theoretical Probability.* ***4***, 485–510.

DÜMBGEN, L. (2002). Application of local rank tests to nonparametric regression. *J. Nonpar. Statist.* ***14***, 511–537.

DÜMBGEN, L. AND SPOKOINY, V.G. (2001). Multiscale testing of qualitative hypotheses. *Ann. Statist.* ***29***, 124–152.

DÜMBGEN, L. AND WALTHER, G. (2008). Multiscale inference about a density. *Ann. Statist.* ***36***, to appear; *accompagnying tecnical report available at* http://arxiv.org/abs/0706.3968

EUBANK, R.L. AND HART, J.D. (1992). Testing goodness-of-fit in regression via order selection criteria. *Ann. Statist.* ***20***, 1412–1425.

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

HÁJEK, J. AND ŠIDAK, Z. (1967). *Theory of rank tests.* Academic press.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* ***58***, 13–30.

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

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

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

LEONOV, S.L. (1997). On the solution of an optimal recovery problem and its applications in nonparametric statistics. *Math. Methods Statist.* ***4***, 476–490.

LEONOV, S.L. (1999). Remarks on extremal problems in nonparametric curve estimation. *Statist. Probab. Lett.* ***43***, 169–178.

LEPSKI, O. AND TSYBAKOV, A. (2000). Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probab. Theory Rel. Fields* ***117***, 17–48.

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

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

NUSSBAUM, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* ***24***, 2399-2430.

DE LA PEÑA, V.H. (1994). A bound on the moment generationg function of a sum of dependent variables with an application to simple sampling without replacement. *Ann. Inst. H. Poincaré Probab. Statist.* ***30***, 197–211.

DE LA PEÑA, V.H. (1999). A general class of exponential inequalities for martingales and ratios. *Ann. Prob.* ***27***, 537–564.

POLLARD, D. (1984). *Convergence of Stochastic Processes.* Springer.

ROHDE, A. (2008). Adaptive goodness-of-fit tests based on signed ranks. *Ann. Statist.* ***36***, 1346–1374.

SERFLING, R.J. (1974). Probability inequalities for the sum of sampling without replacement. *Ann. Statist.* ***2***, 39-48.

SHORACK, G.R. AND WELLNER, J.A. (1986). *Empirical processes with applications to statistics.* Wiley, New York.

SPOKOINY, V. (1996). Adaptive hypothesis testing using wavelets. *Ann. Statist.* ***24***, 2477–2498.

VAN DER VAART, A.W. AND WELLNER, J.A. (1986). *Weak convergence and Empirical processes.* Springer.