HAL
open science

Supersmooth Testing on the Sphere over Analytic Classes

Peter S. Kim, Ja-Yong Koo, Thanh Mai Pham Ngoc

To cite this version:

Peter S. Kim, Ja-Yong Koo, Thanh Mai Pham Ngoc. Supersmooth Testing on the Sphere over Analytic Classes. Journal of Nonparametric Statistics, 2016, 28 (1). hal-01082263v3

HAL Id: hal-01082263

https://hal.science/hal-01082263v3

Submitted on 19 Oct 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Supersmooth Testing on the Sphere over Analytic Classes

Peter T. Kim*, Ja-Yong Koo†and Thanh Mai Pham Ngoc ‡

Saturday 5th September, 2015

Abstract

We consider the nonparametric goodness-of-fit test of the uniform density on the sphere when we have observations whose density is the convolution of an error density and the true underlying density. We will deal specifically with the supersmooth error case which includes the Gaussian distribution. Similar to deconvolution density estimation, the smoother the error density the harder is the rate recovery of the test problem. When considering nonparametric alternatives expressed over analytic classes, we show that it is possible to obtain original separation rates much faster than any logarithmic power of the sample size according to the ratio of the regularity index of the analytic class and the smoothness degree of the error. Furthermore, we show that our fully data driven statistical procedure attains these optimal rates.

Keywords : Spherical deconvolution, fully data driven procedure, minimax hypothesis testing, nonparametric alternatives, supersmooth error, analytic classes, rotational harmonics.
MSC 2010. Primary 62G10, secondary 62H11.

1 Introduction

In this paper, we deal with the nonparametric problem of testing the uniform density on the unit sphere when only corrupted observations are available. In particular, consider the following spherical convolution model:

$$Y_i = W_i X_i, \quad i = 1, \ldots, N \tag{1}$$

where W is an $\mathbb{SO}(3)$ random element, with $\mathbb{SO}(3)$ the rotation group in $\mathbb{R}^3$ and Y, X are $\mathbb{S}^2$ random elements where $\mathbb{S}^2$ denotes the unit sphere of $\mathbb{R}^3$. The elements W and X assumed independent. Let f, g, h denote the densities of X, Y, W respectively. We assume that the noise density h is known. Roughly speaking, one has a genuine directional measurement X_i which is perturbated by a small random rotation W_i. Statistically, model (1) is describing the spherical analogue of a convolution model on the real line where one would observe Y made up of the true measurement X, corrupted by an additive noise W, $Y = X + W$. As we are on the sphere, the additive noise turns naturally to be a rotation and we have the following relation:

$$g = h * f,$$

where convolution $(*)$ is defined below in (8).

*Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario, N1G 2W1, Canada. Supported in part by the NSERC (Canada), DG-2011-46204. Email: pkim@uoguelph.ca

†Department of Statistics, Korea University, Anam-Dong Sungbuk-Ku, Seoul, 136-701 Korea. Supported in part by NRF (Korea), 2009-0075827 and 2013R1A1A2008619. Email: jykoo@korea.ac.kr

‡Laboratoire de Mathématiques, UMR 8628, Université Paris-Sud, 91405 Orsay Cedex, France. Email: thanh.pham_ngoc@math.u-psud.fr.

Inverse problems are characterized by their degree of ill-posedness. Spherical convolution is no exception. The severity of the current problem is determined by the regularity of the noise matrix W. In this paper we will focus on extremely ill-posed problem: the errors are very regular i.e supersmooth and we would like to test the corrupted observations Y for the case where the density follows the uniform distribution against nonparametric alternatives over analytic class of functions. Supersmooth error means that the operator norm of the rotational Fourier transform of W (defined in (4)) decays exponentially for which the Gaussian distribution on $\mathbb{SO}(3)$ is the primary example. Of course, the more regular the noise is, the more difficult the testing problem becomes as the corrupted observations tend to spread more uniformly over the sphere making separation between the two hypotheses harder.

In practise, the spherical convolution model (1) may be viewed as a first attempt to describe the ultra high energy cosmic rays (UHECR) problem. The UHECR provide an extreme kinetic energy (of the order of 10^{19} eV) and are the rarest particles in the universe. Astrophysicists aim at highlighting the mechanisms which generate those cosmic rays. So far, it is not clear which processes produce them. To find out about this question, measurements of the incoming directions of the UHECR on Earth are collected in observatories such as the Pierre Auger one in Argentina. A key question lies in whether the incoming directions of the UHECR are uniform or not. Several hypotheses are made about the underlying probability of the incoming directions. A uniform density would suggest that the UHECR are generated by cosmological effects. On the contrary, if these UHECR are generated by astrophysical phenomena, then we should observe a density function which is highly non-uniform. The answer to this question would provide important information about the sources of those cosmic rays. But of course, measurements of the incoming directions are always made with a certain uncertainty due to the instruments. The noise in the instruments can be modelled through small random rotations hence the considered model (1).

There is a large literature concerning deconvolution for both the Euclidean and non-Euclidean cases. Among the many papers we cite the works which are closest to our approach. For the density estimation setting, among others they are Fan (1991), Pensky and Vidakovic (1999), Kalifa et al. (2003), Meister (2009), Huckemann et al. (2010), Healy et al. (1998), Kim and Koo (2002), Kim et al. (2004), Kerkyacharian et al. (2011), Comte and Lacour (2013). As for the literature concerning nonparametric goodness-of-fit testing with indirect observations one can consult the papers of Butucea (2007) and Butucea et al. (2009) where the authors dealt with convolution model on the real line. In Euclidean density deconvolution estimation, it is known that the smoother the error density, the slower the rate of convergence (see Comte et al. (2006), Comte and Lacour (2013)). For instance, if the unknown density belongs to Hölder or Sobolev classes the optimal rate of convergence is logarithmic when the error is supersmooth. Pensky and Vidakovic (1999) in the wavelets frame, Butucea and Tsybakov (2007), Comte et al. (2006) with kernel machinery however have shown that when considering analytic classes, the rate of convergence can be improved to near polynomial rates.

In our paper, we shall demonstrate that the same phenomenon is observed when testing uniformity from noisy directional data. Lacour and Pham Ngoc (2014) has recently treated the same model and problem as ours. But in their work, the authors deal with alternatives over Sobolev classes and when considering the supersmooth noise situation, they obtain logarithmic separation rates. To date, analytic classes for the present testing problem from a spherical convolution model have not been considered. It is the subject of the present work. Analytic classes are of high interest as they include the most popular families of spherical densities, the so-called von Mises Fisher distributions which has the form $f(x) = C(\kappa)e^{\kappa\mu^\top x}$ with κ a concentration parameter and μ a mean direction. Indeed, we propose a fully data driven testing procedure in the sense that no knowledge of the unknown regularity of the underlying density is required. Furthermore, we point out that similar to deconvolution density estimation, novel separation rates faster than any logarithmic power can be achieved. We identify three principal regions: the bias dominated case in which the variance is negligible with respect to the bias;

the variance dominated case; and, the equal case with the usual bias and variance trade-off. For each of these scenarios, the optimal separation rates are significantly different that exhibit lower bounds which as far as we are aware are unavailable. In addition, we prove that our test procedure achieves up to a logarithmic factor the optimal testing rates.

We now summarize our paper. In Section 2, we provide some basic background to harmonic analysis in $\mathbb{SO}(3)$ and $\mathbb{S}^2$. In Section 3, we describe the model and give assumptions about the density and the noise distribution. We also explain the construction of the test procedure. In Section 4, we provide asymptotic results for analytic classes namely lower and upper bounds for the separation rates in the different regimes. In section 5, we give a numerical study to illustrate our theoretical results. Proofs of the theorems are gathered in Section 6 and all the technical lemmas are proved in the Appendix.

2 Some preliminaries

This section provides the basic tools of harmonic analysis on the group $\mathbb{SO}(3)$ and the unit sphere $\mathbb{S}^2$. Most of the material can be found in expanded form in Vilenkin (1968), Talman (1968), Terras (1985), Healy et al. (1998) and Kim and Koo (2002).

Let $L^2(\mathbb{SO}(3))$ denote the space of square integrable functions on $\mathbb{SO}(3)$, that is, the set of measurable functions f on $\mathbb{SO}(3)$ for which

$$\|f\|_2 = \left(\int_{\mathbb{SO}(3)} |f(x)|^2 dx \right)^{\frac{1}{2}} < \infty,$$

where dx is the Haar measure on $\mathbb{SO}(3)$.

Let

$$u(\varphi) = \begin{pmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad v(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix},$$

where , $\varphi \in [0, 2\pi),\ \theta \in [0, \pi)$. It is well-known that any rotation matrix can be decomposed as a product of three elemental rotations, one around the z-axis first with angle ψ, followed by a rotation around the x-axis with angle θ, and finally another rotation again around the z-axis with angle φ. Indeed, the *Euler angle decomposition* says that any $g \in \mathbb{SO}(3)$ can almost surely be uniquely represented by three angles (φ, θ, ψ), with the following formula (see Talman (1968) chapter 9):

$$g = u(\varphi)v(\theta)u(\psi), \tag{2}$$

where $\varphi \in [0, 2\pi),\ \theta \in [0, \pi),\ \psi \in [0, 2\pi)$.

Now consider the *rotational harmonics*

$$D^c_{klm}(w) = D^c_{klm}(\varphi, \theta, \psi) = e^{-il\varphi} d^k_{lm}(\cos\theta) e^{-im\psi} \quad \text{for} \quad w \in \mathbb{SO}(3),$$

where (φ, θ, ψ) are the *Euler angles* for w and d^k_{lm} are related to the Jacobi polynomials:

$$d^k_{lm}(\cos\theta) = i^{l-m} \frac{\sin^{m-l}\theta (1 + \cos\theta b)^l}{2^k [(k+l)!(k-l)!]^{\frac{1}{2}}} \left[\frac{(k-m)}{(k+m)} \right]^{\frac{1}{2}} \times \frac{d^{k+m}}{d(\cos\theta)^{k+m}} (\cos\theta - 1)^{k+l} (\cos\theta + 1)^{k-l},$$

for $-k \le l, m \le k$, $k \in \mathbb{N}_0 := \{0, 1, 2, \ldots\}$.

Superscript "c" indicates that the rotational harmonics are complex. In addition, if we define a $d_k \times d_k$ matrix D^c_k by

$$D^c_k(w) = [D^c_{klm}(w)] \quad \text{for} \quad w \in \mathbb{SO}(3), \tag{3}$$

where $d_k = 2k+1$, $-k \leq l, m \leq k$, $k \in \mathbb{N}_0$, these constitute the collection of inequivalent irreducible representations of $\mathbb{SO}(3)$.

Let $h \in L^2(\mathbb{SO}(3))$. We define the *rotational Fourier transform* on $\mathbb{SO}(3)$

$$\hat{h}^c_{klm} = \int_{\mathbb{SO}(3)} h(w) D^c_{klm}(w) dw, \tag{4}$$

with dw the probability Haar measure on $\mathbb{SO}(3)$. We think of (4) as the matrix entry at line l and column m of the $d_k \times d_k$ matrix $\hat{h}^c_k = \left[\hat{h}^c_{klm}\right]$ where $d_k = 2k+1$, $-k \leq l, m \leq k$, $k \in \mathbb{N}_0$.

We shall do an analogue analysis for $\mathbb{S}^2$. Any point $x \in \mathbb{S}^2$ can be represented by its spherical coordinates

$$x = (\cos\varphi \sin\theta, \sin\varphi \sin\theta, \cos\theta)^\top,$$

where $\theta \in [0, \pi)$, $\varphi \in [0, 2\pi)$ and superscript "$\top$" denotes transpose.

The well-known *spherical harmonics* constitute an orthonormal basis of $L^2(\mathbb{S}^2)$. They have the following form

$$\phi^c_{kl}(x) = \phi^c_{kl}(\theta, \varphi) = (-1)^l \sqrt{\frac{(2k+1)(k-l)!}{4\pi(k+l)!}} P^k_l(\cos\theta) e^{il\varphi} \quad \text{for} \quad x \in \mathbb{S}^2, \tag{5}$$

where $\theta \in [0, \pi)$, $\varphi \in [0, 2\pi)$, $-k \leq l \leq k$, $k \in \mathbb{N}_0$ and P^k_l are the Legendre functions. Again, the superscript "c" indicates that the rotational harmonics are complex. We can think of (5) as the vector entry at position l to the d_k vector $\phi^c_k(x) = [\phi^c_{kl}(x)]$, $k \in \mathbb{N}_0$.

Here is a representation in spherical coordinates of a spherical harmonic

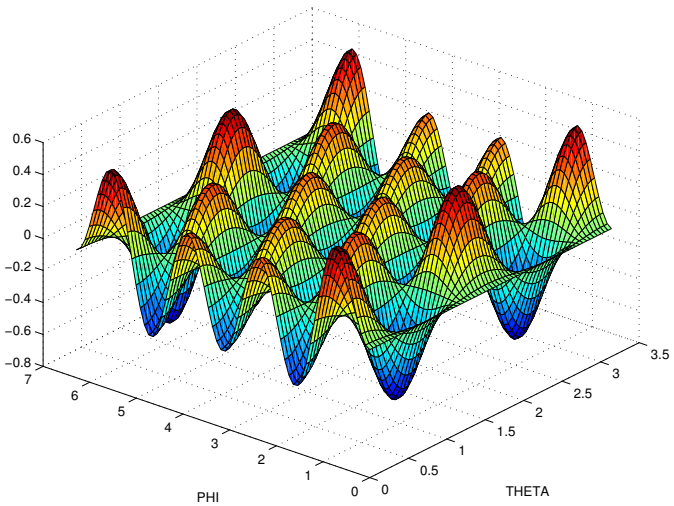


Figure 1: a/ Representation of a spherical harmonic of degree k=8.

Let $L^2(\mathbb{S}^2)$ denote the space of square integrable functions on $\mathbb{S}^2$, that is, the set of measurable functions f on $\mathbb{S}^2$ for which

$$\|f\|_2 = \left(\int_{\mathbb{S}^2} |f(x)|^2 dx\right)^{\frac{1}{2}} < \infty,$$

where dx is the Lebesgue measure on the sphere $\mathbb{S}^2$.

Let $f \in L^2(\mathbb{S}^2)$. We define the *spherical Fourier transform* on $\mathbb{S}^2$ by

$$\hat{f}^c_{kl} = \int_{\mathbb{S}^2} f(x) \overline{\phi^c_{kl}(x)} dx, \tag{6}$$

where dx is the spherical measure on $\mathbb{S}^2$ and the overbar denotes complex conjugation. Again we think of (6) as the vector entry at position l of the d_k vector $\hat{f}_k^c = \left[\hat{f}_{kl}^c\right]$, $-k \leq l \leq k$, $k \in \mathbb{N}_0$. The *spherical inversion* can be obtained by

$$f(x) = \sum_{k \geq 0} (\hat{f}_k^c)^\top \phi_k^c(x) \quad \text{for} \quad x \in \mathbb{S}^2. \tag{7}$$

Strictly speaking, (7) should be interpreted in the L^2-sense although with additional smoothness conditions, it can hold pointwise.

One of the most useful tools of Fourier analysis is the fact that convolution of two functions in the Fourier domain turns out to be ordinary matrix multiplication. Indeed, let $h \in L^2(\mathbb{SO}(3))$ and $f \in L^2(\mathbb{S}^2)$. Defining the convolution as

$$h * f(y) = \int_{\mathbb{SO}(3)} h(w) f(w^{-1}y) dw \quad \text{for} \quad y \in \mathbb{S}^2, \tag{8}$$

we have the following *convolution property*

$$(\widehat{h * f})_k^c = \hat{h}_k^c \hat{f}_k^c \quad \text{for} \quad k \in \mathbb{N}_0, \tag{9}$$

for the proof see Lemma 2.1 in Healy et al. (1998). This *convolution property* will be very useful in the sequel (section 3.2) to construct our test procedure.

The way the rotational and spherical harmonics are presented in (3) and (5), respectively, uses complex bases. We can convert the problem into a real basis by defining the $d_k \times d_k$ complex matrix

$$A_k = \frac{1}{\sqrt{2}} \begin{bmatrix} i & 0 & \cdots & 0 & \cdots & 0 & -i(-1)^k \\ 0 & i & \cdots & 0 & \cdots & -i(-1)^{k-1} & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \sqrt{2} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & \cdots & 0 & \cdots & (-1)^{k-1} & 0 \\ 1 & 0 & \cdots & 0 & \cdots & 0 & (-1)^k \end{bmatrix} \quad \text{for} \quad k \in \mathbb{N}_0. \tag{10}$$

Notice that (10) is a unitary matrix, $A_k A_k^\star = A_k^\star A_k = I_{d_k}$, where superscript "$\star$" denotes conjugate transpose. Define

$$\phi_k := \overline{A_k} \phi_k^c \quad \text{and} \quad D_k := A_k D_k^c A_k^\star \quad \text{for} \quad k \in \mathbb{N}_0. \tag{11}$$

One can see that the d_k-vector ϕ_k is real and that the $d_k \times d_k$ matrix D_k is also real, and because they are derived from a unitary transformation, they form real orthogonal bases for $L^2(\mathbb{S}^2)$ and $L^2(\mathbb{SO}(3))$, respectively. Hence contrary to Lacour and Pham Ngoc (2014), we will use a real basis which will lighten the proofs.

The supremum norm for functions on $\mathbb{S}^2$ is denoted as

$$\|f\|_\infty = \sup_{x \in \mathbb{S}^2} |f(x)|.$$

For $u, v \in \mathbb{R}^{d_k}$, let $\langle u, v \rangle = u^\top v$ and $\|v\| = \sqrt{\langle v, v \rangle}$.

3 Testing uniformity on the sphere

3.1 Model and assumptions

3.1.1 Statistical model

The uniform density f^0 on $\mathbb{S}^2$ is given by

$$f^0 = \mathfrak{C}_0 \mathbb{1}_{\mathbb{S}^2} \quad \text{for} \quad \mathfrak{C}_0 = \frac{1}{\text{area}(\mathbb{S}^2)} = \frac{1}{4\pi}.$$

Define a separation measure between f^0 and f by

$$\Delta_f := \left\| f - f^0 \right\|_2. \tag{12}$$

The uniform testing problem is

$$H_0 : f = f^0 \quad \text{versus} \quad H_a : f \in \mathcal{H}_a(\mathcal{F}, M, \varepsilon)$$

based on the random sample $Y_1, \ldots, Y_N$ from the distribution having the density g, where the alternative is

$$\mathcal{H}_a(\mathcal{F}, M, \varepsilon) = \{ f : f \in \mathcal{F} \quad \text{and} \quad \Delta_f \geq M\varepsilon \}$$

for a class $\mathcal{F}$ of densities, a testing rate ε and a constant M.

For any test T, the maximal misclassification error rate (or the risk for the zero-one loss) is defined by

$$R_N(T, \mathcal{F}, M, \varepsilon) = \mathbb{P}_{f^0}(T = 1) + \sup \left\{ \mathbb{P}_f(T = 0) : f \in \mathcal{H}_a(\mathcal{F}, M, \varepsilon) \right\} \tag{13}$$

where $\mathbb{P}_f$ denotes the probability measure with respect to the joint distribution of $Y_1, \ldots, Y_N$.

3.1.2 Analytic class for alternatives

We shall present more formally analytic densities. On the space $C^\infty(\mathbb{S}^2)$ of infinitely continuous differentiable function on $\mathbb{S}^2$, consider the analytic norm $\|\cdot\|_{\mathcal{A}_{p,r}}$ defined as

$$\|f\|^2_{\mathcal{A}_{p,r}} = \sum_{k \geq 0} \exp\left(2pk^r\right) \left\| \hat{f}_k \right\|^2.$$

Let the eigenvalues of the Laplace-Beltrami operator on $\mathbb{S}^2$ be $\lambda_k = k(k+1) \quad \text{for} \quad k \in \mathbb{N}_0$. One can show that there exist positive constants c_1 and c_2 such that

$$c_1 \|f\|^2_{\mathcal{A}_{p,r}} \leq \sum_{k \geq 0} \exp\left(2p\lambda_k^{r/2}\right) \left\| \hat{f}_k \right\|^2 \leq c_2 \|f\|^2_{\mathcal{A}_{p,r}}.$$

An analytic class of densities is

$$\mathcal{A}_{p,r}(Q) = \left\{ f : \mathbb{S}^2 \to \mathbb{R}^+, \quad \int_{\mathbb{S}^2} f = 1 \quad \text{and} \quad \|f\|^2_{\mathcal{A}_{p,r}} \leq \mathfrak{C}_0 + Q^2 \right\}.$$

where $p > 0$, $r > 0$, $Q > 0$ are finite constants and $\mathbb{R}^+ = \{ t \in \mathbb{R} : t \geq 0 \}$.

The smoothness classes described above are classically considered both in deconvolution and in direct density estimation. They have been studied in many papers in the Euclidean cases (see for instance

Belitser and Levit (2001), Comte et al. (2006), Butucea and Tsybakov (2007)). Note that all such densities are infinitely many times differentiable on $\mathbb{S}^2$.

This class of functions are of great interest as they include the most widely referenced distribution in directional data (see Kim and Koo (2002)), the von Mises Fisher distribution which has the following form on $\mathbb{S}^2$:

$$f_{VMF}(x|\mu,\kappa) = c_3(\kappa)e^{\kappa\mu^\top x},$$

where $\|\mu\| = 1$, $\kappa \geq 0$ and the normalizing constant is equal to $c_3(\kappa) = \frac{\kappa}{2\pi(e^\kappa - e^{-\kappa})}$. The parameter κ is the concentration parameter and μ the mean direction. The bigger κ is, the more concentrated the observations are around the direction μ. This distribution is unimodal and for $\kappa = 0$, one retrieves the uniform distribution.

3.1.3 Noise assumptions

As in Euclidean case, the difficulty of the deconvolution turns out to depend on the spectral properties of the error distribution. Following Fan (1991), Fan and Koo (2002), Healy et al. (1998), Kim and Koo (2002), Lacour and Pham Ngoc (2014), we will appropriately define the smoothness of the error spectrally but with a modification due to the fact that the error W belongs to $\mathbb{SO}(3)$. This can be done by considering convolution as an operator (see Kim and Koo (2002)).

Indeed let $\mathcal{E}_k$ be the d_k-dimensional vector space spanned by $\{\phi_{kl} : -k \leq l \leq k\}$ for each $k \in \mathbb{N}_0$. Now according to (9), $\hat{h}_k$ and $\hat{h}_k^{-1}$ are also understood as maps $\hat{h}_k : \mathcal{E}_k \to \mathcal{E}_k$ and $\hat{h}_k^{-1} : \mathcal{E}_k \to \mathcal{E}_k$ defined by

$$\hat{h}_k v = (\hat{h}_k \hat{v}_k)^\top \phi_k \quad \text{and} \quad \hat{h}_k^{-1} v = (\hat{h}_k^{-1}\hat{v}_k)^\top \phi_k \quad \text{for} \quad v = \hat{v}_k^\top \phi_k \in \mathcal{E}_k.$$

Again by Parseval's identity, $\|\hat{h}_k v\|_2^2 = \|\hat{h}_k \hat{v}_k\|$ for all $k \geq 0$. Consequently, we have the operator inequality:

$$\left\|\hat{h}_k v\right\|_2 \leq \left\|\hat{h}_k\right\|_{\text{op}} \|v\|_2 \text{ where } \left\|\hat{h}_k\right\|_{\text{op}} = \sup_{v\neq 0, v\in\mathcal{E}_k} \frac{\left\|\hat{h}_k v\right\|_2}{\|v\|_2} \text{ and } \left\|\hat{h}_k^{-1}\right\|_{\text{op}} = \sup_{v\neq 0, v\in\mathcal{E}_k} \frac{\left\|\hat{h}_k^{-1} v\right\|_2}{\|v\|_2}. \tag{14}$$

Assumption 1. *For all $k \in \mathbb{N}_0$, the matrix $\hat{h}_k$ is invertible and there exist constants $q > 0$, $s > 0$, $b_0 > 0$, $b_1 > 0$, $\nu_0, \nu_1 \in \mathbb{R}$ such that*

$$\left\|\hat{h}_k^{-1}\right\|_{\text{op}} \leq b_0 k^{\nu_0} \exp\left(qk^s\right) \tag{15}$$

and

$$\left\|\hat{h}_k\right\|_{\text{op}} \leq b_1 k^{-\nu_1} \exp\left(-qk^s\right). \tag{16}$$

The error density h is referred to as *supersmooth*.

A typical supersmooth noise distribution is the Gaussian distribution (see Kim and Koo (2002)). This latter can be solved on general Riemanian manifold by solving the appropriate heat equation. Its expression in the basis of rotational harmonics is:

$$h_G = \sum_k \sum_{l=-k}^{k} \exp(-\sigma^2 l(l+1)/2)(2l+1)D_{kl}. \tag{17}$$

Consequently, its rotational Fourier transform is given by:

$$\hat{h}_{G,klm} = \exp(-\sigma^2 l(l+1)/2)\delta_{mn},$$

with δ_{mn} the Kronecker delta. Hence, Gaussian distribution is supersmooth with parameters $b_0 = 1, b_1 = 1, \nu_0 = 0, \nu_1 = 0$, $s = 2$ and $q = \frac{\sigma^2}{2}$ in (15) and (16).

3.2 Test procedure

The test procedure relies on the construction of an estimator of the separation measure Δ_f.

Through (1), the relation among the densities can be described by convolution,

$$g = h * f$$

as seen by following the familiar corresponding Euclidean result. By (9) we can write,

$$\hat{f}_k = \hat{h}_k^{-1} \hat{g}_k,$$

provided Assumption 1.

Since f is unknown, g is also unknown, hence $\hat{g}_k$ is unknown. Nevertheless, we assume that a random sample $Y_1, \ldots, Y_N$ is available. This will allow us to construct an empirical version $\hat{g}_k^N$ as:

$$\hat{g}_k^N = \frac{1}{N} \sum_{j=1}^{N} \phi_k(Y_j).$$

An estimator for $\hat{f}_k$ is therefore

$$\hat{f}_k^N = \hat{h}_k^{-1} \hat{g}_k^N = \frac{1}{N} \sum_{j=1}^{N} \Phi_k(Y_j) \quad \text{for} \quad k \in \mathbb{N}_0, \tag{18}$$

where

$$\Phi_k = \hat{h}_k^{-1} \phi_k.$$

Define

$$G(y_1, y_2) := \sum_{k=1}^{K} \langle \Phi_k(y_1), \Phi_k(y_2) \rangle \quad \text{for} \quad y_1, y_2 \in \mathbb{S}^2 \tag{19}$$

and

$$U_K := \frac{1}{N(N-1)} \sum_{1 \le i_1 \ne i_2 \le N} G(Y_{i1}, Y_{i2}) \quad \text{for} \quad K \ge 1.$$

Since $\hat{g}_k = \hat{h}_k \hat{f}_k$, we obtain

$$\mathbb{E}_f \Phi_k(Y) = \int_{\mathbb{S}^2} g \hat{h}_k^{-1} \phi_k = \hat{h}_k^{-1} \int_{\mathbb{S}^2} g \phi_k = \hat{h}_k^{-1} \hat{g}_k = \hat{f}_k$$

and

$$\mathbb{E}_f U_K = \mathbb{E}_f G(Y_1, Y_2) = \sum_{k=1}^{K} \langle \mathbb{E}_f \Phi_k(Y_1), \mathbb{E}_f \Phi_k(Y_2) \rangle = \sum_{k=1}^{K} \left\langle \hat{f}_k, \hat{f}_k \right\rangle = \sum_{k=1}^{K} \left\| \hat{f}_k \right\|^2.$$

Hence, U_K is a natural estimator of

$$\Delta_f^2 := \sum_{k \ge 1} \left\| \hat{f}_k \right\|^2. \tag{20}$$

Given the estimator U_K of Δ_f^2, we can define a hard thresholding test procedure T_K by

$$T_K = \mathbb{1}\big(|U_K| > \tau_K \big) = \begin{cases} 1 & \text{if} \quad |U_K| > \tau_K \\ 0 & \text{if} \quad |U_K| \le \tau_K, \end{cases} \tag{21}$$

for a threshold τ_K to be made precise later. The way to choose K is important and will be highlighted in section 4.2.

In what follows we denote by C constants with values in $\mathbb{R}$ which may differ on various occurences. Suppose a_N and b_N are positive real sequences. If $a_N = O(b_N)$, we write $a_N \ll b_N$. The notation $a_N \asymp b_N$ means that $a_N = b_N\big(1 + o(1)\big)$.

4 Asymptotic results for Analytic classes

In the same framework, Lacour and Pham Ngoc (2014) showed that considering supersmooth noise and Sobolev class alternatives leads to only logarithmic separation rates. Let us see how introducing analytic classes change notably that conclusion. Indeed, the following theorems show that it is possible to achieve polynomial testing rates. Furthermore, the asymptotic results on minimax misclassification error rate differ for three regions: the *variance dominated case* ($r > s$), the *equivalent case* ($r = s$) and the *bias dominated case* ($r < s$). In the sequel, we will focus more specifically on the cases $r < \frac{s}{2}$, $r = s$ and $s < \frac{r}{2}$ for which we have specific minimax rates. Accordingly, we define three regions :

$$\Theta_{s<\frac{r}{2}} := \Big\{(p,r,Q) : p > 0, \quad 0 < s < \frac{r}{2} \quad \text{and} \quad Q > 0\Big\},$$
$$\Theta_{s=r} := \Big\{(p,r,Q) : p > 0, \quad 0 < s = r \quad \text{and} \quad Q > 0\Big\},$$
$$\Theta_{r<\frac{s}{2}} := \Big\{(p,r,Q) : p > 0, \quad 0 < r < \frac{s}{2} \quad \text{and} \quad Q > 0\Big\}.$$

4.1 Lower bounds for separation rates

We will denote in the sequel

$$C(p,r,Q) := (\mathfrak{C}_0 + Q^2)^{\frac{1}{2}} \sqrt{\sum_{k\geq 0} d_k \exp\left(-2pk^r\right)} \quad \text{with} \quad d_k = 2k+1. \tag{22}$$

Theorem 1. *Let $(p,r,Q) \in \Theta_{s<\frac{r}{2}}$ and $0 < \eta < 1$. Let*

$$\varepsilon_{\mathtt{L}} = \exp\left[-\frac{1}{2}\log N + q\left(\frac{\log N}{2p}\right)^{s/r} + \frac{2\nu_0 + 1 - \frac{s}{2}}{2r}\log\frac{\log N}{2p}\right].$$

Suppose (15) and (16) hold and $\nu_1 > \nu_0 + (2-s)/4$. If $M < \min(Q, \frac{\mathfrak{C}_0}{QC(r,p,Q)})$ then

$$\liminf_{N\to\infty} \inf_T R_N(T, \mathcal{A}_{p,r}(Q), M, \varepsilon_{\mathtt{L}}) \geq \eta, \tag{23}$$

If $\nu_0 = \nu_1$, then we require $s > 2$ to guarantee (23).

For the case $0 < s < \frac{r}{2}$, Theorem 1 shows that we cannot test with a faster rate than $\varepsilon_{\mathtt{L}} = \exp\left[-\frac{1}{2}\log N + q\left(\frac{\log N}{2p}\right)^{s/r} + \frac{2\nu_0+1-\frac{s}{2}}{2r}\log\frac{\log N}{2p}\right]$ otherwise the maximal misclassification error would be close to 1.

If $s = r$, then we can trade-off the bias and the variance.

Theorem 2. *Let $(p,r,Q) \in \Theta_{s=r}$ and $0 < \eta < 1$. Let*

$$\varepsilon_{\mathtt{L}} = \exp\left(-\frac{p}{2(p+q)}\log N + \frac{2\bar{\nu}_1}{2(p+q)r}\log\frac{\log N}{2(p+q)}\right),$$

for $\bar{\nu_1} < \nu_1$. Under assumptions (15) and (16), as soon as $M < \min(Q, \frac{\mathfrak{C}_0}{QC(r,p,Q)})$ we have

$$\liminf_{N\to\infty} \inf_T R_N(T, \mathcal{A}_{p,r}(Q), M, \varepsilon_{\mathtt{L}}) \geq \eta. \tag{24}$$

When $s > r$, we have the *bias dominated case.*

Theorem 3. *Let $(p, r, Q) \in \Theta_{r<\frac{s}{2}}$ and $0 < \eta < 1$. Let*

$$\varepsilon_{\mathtt{L}} = \exp\left[-p\left(\frac{\log N}{2q}\right)^{r/s}\right](1 + o(1)).$$

Under (15), (16) and $\nu_1 > \nu_0 + (2-s)/4$, if $M < \min(Q, \frac{\mathfrak{C}_0}{QC(r,p,Q)})$ then

$$\liminf_{N\to\infty} \inf_{T} R_N(T, \mathcal{A}_{p,r}(Q), M, \varepsilon_{\mathtt{L}}) \geq \eta, \tag{25}$$

If $\nu_0 = \nu_1$, then we require $s > 2$ to guarantee (25).

To the best of our knowledge, these testing rates have never been showed before. Nonetheless, one may make some parallels with the results obtained by Butucea and Tsybakov (2007) in the context of estimating a density in a convolution model on the real line. Considering analytic densities and supersmooth noise, the authors also obtained different convergence rate regimes according to the ratio between the regularity of the density and the noise.

4.2 Upper bounds

Before defining our test procedure, let us start with an exponential inequality which will be useful to define our test.

Lemma 1. *Let $\xi' = \alpha(\mathbb{1}(0 \leq s \leq 1)K^{1-s/2} + \mathbb{1}(s > 1)K^{s/2})$, $\alpha > 0$. Under the null hypothesis, there exist positive constants M_1 and M_2 such that*

$$\mathbb{P}_{f^0}\left(|U_K| \geq \xi' \frac{V_{0K}}{N}\right) \leq M_1 \exp\left(-M_2 \xi'^2\right),$$

if $K^a N^{-1/3} = o(1)$, $a > 0$, with $V_{0K} = K^{2\nu_0+1-s/2}\exp\left(2qK^s\right)$.

Remark 1. The proof of Lemma 1 is given in the Appendix.

Now consider the test statistic given by

$$T_{\mathtt{A}} := \mathbb{1}\left(\max_{K\in\mathcal{K}} |U_K|/\tau_K > 1/\sqrt{M_2}\right), \tag{26}$$

with

$$\tau_K = \xi\frac{V_{0K}}{N}, \quad V_{0K} := K^{2\nu_0+1-s/2}\exp(2qK^s) \quad \text{and} \quad \xi := \left(\mathbb{1}_{(0\leq s\leq 1)}K^{(2-s)/2} + \mathbb{1}_{(s>1)}K^{s/2}\right), \tag{27}$$

where $\mathcal{K} = \left\{K : L \leq K \leq U\right\}$ is an index family that will be chosen appropriately depending on the ratio $\frac{r}{s}$.

First, consider the case when $s < \frac{r}{2}$ which is included in the *variance dominated* case. We denote $\lceil x \rceil$ the smaller integer larger than or equal to x and $\lfloor x \rfloor$ the largest integer smaller than or equal to x.

Theorem 4. *Suppose (15) holds. For $p \geq p_0 > 0$, $0 < s < \frac{r}{2}$ and $Q > 0$ let the test $T_{\mathtt{A}}$ be defined in (26) with*

$$L = \left\lceil \left(\frac{\log\log N}{s}\right)^{\mathbb{1}(0\leq s\leq 1)\frac{1}{2-s} + \mathbb{1}(s>1)\frac{1}{s}} \right\rceil \quad \textit{and} \quad U = \left\lceil \left(\frac{\log N}{2p_0}\right)^{\frac{1}{s}} \right\rceil.$$

Let

$$\varepsilon_{\mathtt{U}} = \exp\left[-\frac{1}{2}\log N + q\left(\frac{\log N}{2p}\right)^{s/r} + \frac{2\nu_0 + 1 + (1-s)\mathbb{1}_{(0\le s\le 1)}}{2r}\log\frac{\log N}{2p}\right].$$

Then if $M > (Q^2+2)^{\frac{1}{2}}$,

$$\lim_{N\to\infty} R_N(T_A, \mathcal{A}_{p,r}(Q), M, \varepsilon_{\mathtt{U}}) = 0.$$

Let us consider the case $r = s$.

Theorem 5. *Suppose (15) holds. For* $(p,Q) \in \Theta_{s=r}$ *let the test* $T_{\mathtt{A}}$ *be defined in (26) with*

$$L = \left\lceil \left(\frac{\log\log N}{s}\right)^{\mathbb{1}_{(0\le s\le 1)}\frac{1}{2-s} + \mathbb{1}_{(s>1)}\frac{1}{s}} \right\rceil \quad \textit{and} \quad U = \left\lceil \left(\frac{\log N}{2q}\right)^{\frac{1}{s}} \right\rceil.$$

Let

$$\varepsilon_{\mathtt{U}} = \exp\left[-\frac{p}{2(p+q)}\log N + \frac{2\nu_0 + 1 + (1-s)\mathbb{1}_{(0\le s\le 1)}}{2(p+q)r}\log\frac{\log N}{2(p+q)}\right].$$

Then if $M > (Q^2+2)^{\frac{1}{2}}$,

$$\lim_{N\to\infty} R_N(T_A, \mathcal{A}_{p,r}(Q), M, \varepsilon_{\mathtt{U}}) = 0.$$

Finally, for the *bias dominated case* $s > r$, we have the following result :

Theorem 6. *Suppose (15) holds. For* $(p,r,Q) \in \Theta_{r<s/2}$ *let the test* $T_{\mathtt{A}}$ *defined in (26) with*

$$L = \left\lceil \left(\frac{\log\log N}{s}\right)^{\mathbb{1}_{(0\le s\le 1)}\frac{1}{2-s} + \mathbb{1}_{(s>1)}\frac{1}{s}} \right\rceil \quad \textit{and} \quad U = \left\lceil \left(\frac{\log N}{2q}\right)^{\frac{1}{s}} \right\rceil.$$

Let

$$\varepsilon_{\mathtt{U}} = Q\exp\left[-p\left(\frac{\log N}{2q}\right)^{r/s}\right](1+o(1)).$$

Then if $M > (Q^2+2)^{\frac{1}{2}}$,

$$\lim_{N\to\infty} R_N(T_A, \mathcal{A}_{p,r}(Q), M, \varepsilon_{\mathtt{U}}) = 0.$$

We remark that the test $T_{\mathtt{A}}$ constructed does not require any knowledge of smoothness parameters (p,r) of the analytic class $\mathcal{A}_{p,r}(Q)$ which are not known in practice. Hence we have a fully data driven procedure.

The upper bound results show that our test procedure T_A achieves the optimal rates up to a logarithmic factor. In addition, one notes that the limiting distribution of the asymptotically minimax test statistic is degenerate in all regimes.

5 Numerical illustration

In this section, we will provide a simulation study to illustrate our theoretical results. We focus on the von Mises Fisher and Watson distributions which provide examples of analytic densities on the sphere. Concerning the noise, we will naturally deal with the Gaussian distribution. We compare the performances of our test with those of the Beran Giné test.

5.1 Test statistics

For the Gaussian distribution, $s = 2, \nu_0 = 0$ and $q = \frac{\sigma^2}{2}$ where σ^2 is the variance, hence the expression of our test $T_{\mathtt{A}}$ described in (26) has the following expression :

$$T_{\mathtt{A}} = \mathbb{1}\left(\max_{K\in\mathcal{K}} |U_K|/\tau_K > \frac{1}{\sqrt{M_2}}\right), \tag{28}$$

with

$$\tau_K = K\frac{\exp(\sigma^2 K^2)}{N} \quad \text{and} \quad \mathcal{K} = \left\{K : \left\lceil \sqrt{\log\log(N)} \right\rceil \leq K \leq \left\lceil \sqrt{\frac{\log(N)}{\sigma^2}} \right\rceil \right\}. \tag{29}$$

We compute the 5% quantile M_2 of the statistic $\max_{K\in\mathcal{K}} |U_K|/\tilde{\tau}_K$ by 1000 Monte Carlo simulations of N observations under the null hypothesis. We point out that our procedure does not require any tuning parameter.

The second procedure was introduced by Beran (1968) and Giné (1975). The test statistic is

$$F_N = \frac{3N}{2} - \frac{4}{N\pi}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} d(Z_i, Z_j) + \sin(d(Z_i, Z_j))$$

where $d(Z_i, Z_j) = \arccos\langle Z_i, Z_j\rangle$ is the spherical distance between Z_i and Z_j, and the quantiles are computed via simulations under H_0. This test was designed for non-noisy data.

5.2 Alternatives

We will deal with two alternatives. For the first alternative hypothesis, we have chosen the von Mises Fisher distribution with several cases of concentration parameters κ and the same mean direction $\mu = (-1, 0, 0)$. For instance, for $\kappa = 3$, the density has the expression in spherical coordinates : $f_{VMF}(\theta, \varphi) = 0.0238 \times e^{-3\sin\theta\cos\varphi}$. We recall that for $\kappa = 0$, the von Mises Fisher density turns to be the uniform density. So the smaller the concentration parameter κ is, the "flatter" is the density and the more severe the alternative is.

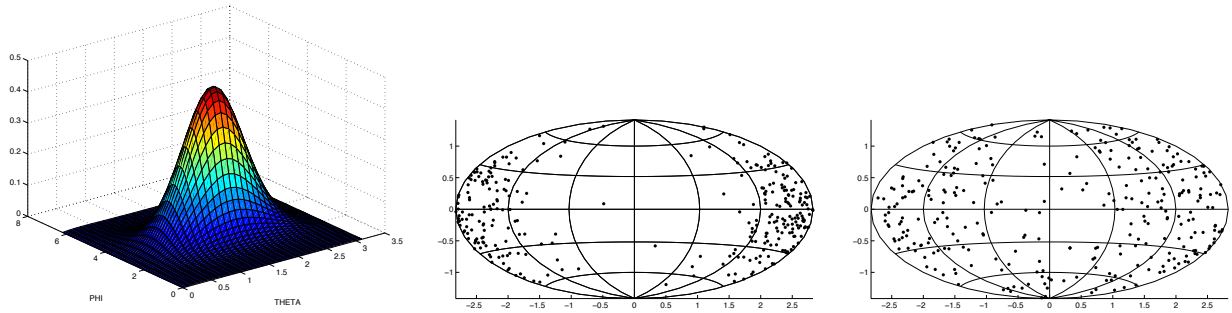
For the second alternative, we focus on the Watson distribution which provides a way to model axially symmetric data i.e for which $\pm x \in \mathbb{S}^2$ are equivalent. Its general form is

$$f_W(x) = c(\kappa)e^{\kappa(\mu^\top x)^2}, \quad x \in \mathbb{S}^2,$$

where $\kappa \in \mathbb{R}$ is a concentration parameter, μ a mean direction and $c(\kappa)$ a normalizing constant such as $\int_0^{2\pi}\int_0^{\pi} f_W(\theta, \varphi)\sin(\theta)d\theta d\varphi = 1$. For our simulations, we will consider $\mu = (0, 0, 1)$ and $\kappa < 0$, hence $f_W(\theta, \varphi) = c(\kappa)e^{\kappa\cos^2\theta}$. For negative κ, the data are concentrated around the great circle orthogonal to μ which amounts in our case to be concentrated around the equator.

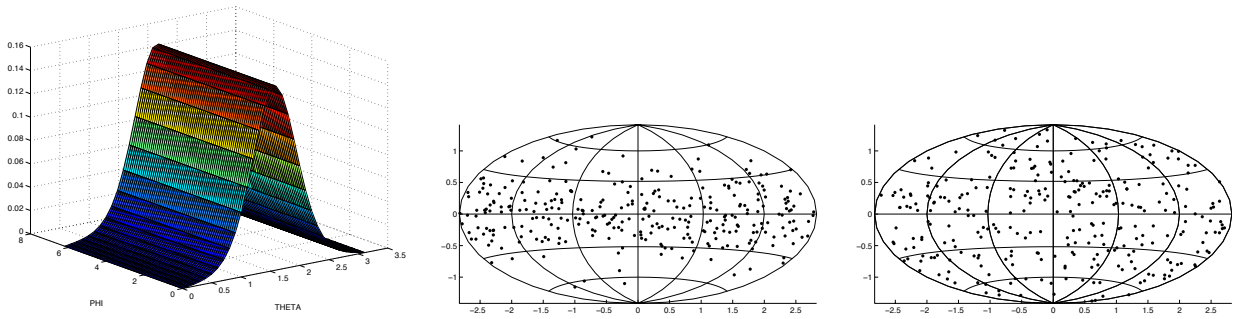
Figure 2 b. shows 300 random draws X_i from the von Mises Fisher distribution with $\kappa = 3$ and $\mu = (-1, 0, 0)$ and Figure 2 c. the Y_i (obtained by corrupting the X_i by a Gaussian noise with variance 0.4). One can notice that for the choice $\kappa = 3$, original observations are clearly concentrated in a zone, so obviously far from the uniform distribution. But once perturbated, the observations tend to be spread all over (Figure 3 c.), making the separation between the uniform hypothesis H_0 and the alternative hypothesis H_a difficult.

The same phenomenon appears with the Watson distribution (see Figure 3) but one can notice that the effect of the noise seems to be more severe in the case of the Watson distribution.



a b c

Figure 2: a/ Representation of the f_{VMF} density with $\kappa = 3, \mu = (-1, 0, 0)$. b/ 300 random draws from the f_{VMF} density with $\kappa = 3, \mu = (-1, 0, 0)$. c/ These 300 observations perturbated by a Gaussian noise with $\sigma^2 = 0.4$.



a b c

Figure 3: a/ Representation of the f_W density with $\kappa = -3, \mu = (0, 0, 1)$. a/ 300 random draws from the f_W density with $\kappa = -3, \mu = (0, 0, 1)$. b/ These 300 observations perturbated by a Gaussian noise with $\sigma^2 = 0.4$.

5.3 Gaussian noise perturbation

The genuine observations are perturbated by a Gaussian noise which has been defined analytically in (17). Let us make a brief remark on how to generate random rotation matrices following a Gaussian distribution on $\mathbb{SO}(3)$. After rewriting carefully the density expression (17) in terms of rotational harmonics, it turned out that h_G only depends on the angle of the rotation say γ. Then the simulation of a rotation following h_G amounts to pick at random an axis and perform a rotation around this axis by an angle following the law $h_G(\gamma)\frac{1-\cos(\gamma)}{\pi}$.

5.4 Numerical results

We generate N random variables X_i following the alternative density and we perturbate each of them with a random rotation following a Gaussian distribution to get the Y_i. Then we implement the two test procedures $T_{\mathtt{A}}$ and the Beran Giné test defined above and assess their performances by computing their power.

For a prescribed 5% level, we compute the power (in percentages) for different values of von Mises Fisher and Watson concentration parameters κ, noise variances and number of observations.

concentration parameter κ	Gaussian variance 0.1	0.2	0.4
1	99.5	90	40.50
2	100	100	94
3	100	100	100

concentration parameter κ	Gaussian variance 0.1	0.2	0.4
1	99.5	91	62
2	100	100	98
3	100	100	100

Table 1: Test powers for the f_{VMF} alternative, $N = 100$: on the left the test T_A, on the right the Beran Giné test.

concentration parameter κ	Gaussian variance 0.1	0.2	0.4
1	100	100	81
2	100	100	100
3	100	100	100

concentration parameter κ	Gaussian variance 0.1	0.2	0.4
1	100	100	90
2	100	100	100
3	100	100	100

Table 2: Test powers for the f_{VMF} alternative, $N = 300$: on the left the test T_A, on the right the Beran Giné test.

concentration parameter κ	Gaussian variance 0.1	0.2	0.4
-1	24.75	9	6
-2	49.50	11.25	6.25
-3	79.25	19	6

concentration parameter κ	Gaussian variance 0.1	0.2	0.4
-1	6.25	10	7.75
-2	19	5.75	6.50
-3	27.50	8.50	3.25

Table 3: Test powers for the f_W alternative, $N = 100$: on the left the test T_A, on the right the Beran Giné test.

concentration parameter κ	Gaussian variance 0.1	0.2	0.4
- 1	85.50	18.50	5.25
-2	98.50	40.25	5.50
-3	100	64.50	5.75

concentration parameter κ	Gaussian variance 0.1	0.2	0.4
-1	42	12	4.75
-2	81.50	17.50	6.50
-3	96.75	20.50	7.75

Table 4: Test powers for the f_W alternative, $N = 300$: on the left the test T_A, on the right the Beran Giné test.

Generally, the performances are much better with the von Mises Fisher distribution which is a less severe alternative than the Watson distribution. Visually, a clue can be found when comparing Figure 2 and Figure 3. Indeed, in Figure 3 c. it is hard to detect a signal any more. As expected, for both alternatives and the two tests, the bigger the concentration parameter $|\kappa|$ is, the better the performances are since the data tend to be more concentrated in a zone. On the contrary, the bigger the noise variances are, the worse the test powers are as the data get closer to the uniform distribution. And of course, increasing the number of observations improves the performances. When compared the two test procedures, for the von Mises Fisher alternative, one remarks that the Beran Giné test has slightly better performances but when considering the Watson distribution, the performances of our test, except for $\sigma^2 = 0.4$ for which both procedures are very poor, $T_{\mathtt{A}}$ behaves much better.

6 Proofs

Proof of Theorem 1

Proof. We shall follow the usual scheme to prove lower bounds rates (see for example Ingster (1997) or Tsybakov (2009)). Let us consider

$$K_{\mathrm{L}} = \left\lfloor \frac{\log N}{2p} - \frac{q}{p}\left(\frac{\log N}{2p}\right)^{\frac{s}{r}} - \frac{2\nu_0 + 1 - \frac{s}{2}}{2pr}\log\frac{\log N}{2p}\right\rfloor^{\frac{1}{r}}$$

and the following function

$$f^1 := f^0 + C_1\Psi,$$

with $\Psi = \delta\phi_{K_{\mathrm{L}}0}$, $\delta = \exp(-pK_{\mathrm{L}}^r)$ and $C_1 = \min(Q, \frac{\mathfrak{C}_0}{C(p,r,Q)})$. We remind that $\phi_{K_{\mathrm{L}}0}$ denotes the spherical harmonic indexed by $(K_{\mathrm{L}}, 0)$. If we prove the four following conditions

- f^1 belongs to $\mathcal{A}_{p,r}(Q)$,
- f^1 is a density function,
- $\|f^1 - f^0\|_2 \geq M\varepsilon$,
- $N\chi^2(g^0, g^1) \leq (1-\eta)^2$,

then,

$$\begin{aligned}
\inf_T R_N(T, \mathcal{A}_{p,r}(Q), M, \varepsilon) &= \inf_T \left[\mathbb{P}_{f^0}(T=1) + \sup\left\{\mathbb{P}_f(T=0) : f \in \mathcal{H}_a(\mathcal{A}_{p,r}(Q), M, \varepsilon)\right\}\right] \\
&\geq \inf_T \left[\mathbb{P}_{f^0}(T=1) + \mathbb{P}_{f^1}(T=0)\right] \\
&\geq \int \min(d\mathbb{P}_{f^0}, d\mathbb{P}_{f^1}) \\
&\geq 1 - \sqrt{N\chi^2(g^0, g^1)} \\
&\geq \eta.
\end{aligned}$$

Now let us turn to the proof of the theorem.

- First let us prove that f^1 belongs to $\mathcal{A}_{p,r}(Q)$. We have

$$\begin{aligned}
\left\|f^1\right\|_{\mathcal{A}_{p,r}} &= \sum_{k\geq 0}\exp(2pk^r)\|\hat{f}_k^1\|^2 \\
&= \mathfrak{C}_0 + \sum_{k>0}\exp(2pk^r)\|\hat{f}_k^1\|^2 \\
&= \mathfrak{C}_0 + C_1^2\delta^2\exp\left(2pK^r\right) \\
&= \mathfrak{C}_0 + C_1^2.
\end{aligned}$$

Since $C_1 \leq Q$, this ensures that f^1 belongs to $\mathcal{A}_{p,r}(Q)$.

- Secondly, let us prove that f^1 is a density. We have

$$\int_{\mathbb{S}^2} f^1 = 1 + C_1\delta\int_{\mathbb{S}^2}\phi_{K0} = 1,$$

since $\int_{\mathbb{S}^2}\phi_{K0} = 0$.

Because $\|\Psi\|_{\mathcal{A}_{p,r}} = \delta^2 \exp(2pK^r) = 1$, we get that $\Psi \in \mathcal{A}_{p,r}(Q)$ and using Lemma 2, we get $\|\Psi\|_\infty \leq C(p,r,Q)$ and thus $f^1 \geq \mathfrak{C}_0 - C_1\|\Psi\|_\infty \geq \mathfrak{C}_0 - C_1 \times C(p,r,Q)$. Since $C_1 \leq \frac{\mathfrak{C}_0}{C(p,r,Q)}$ we obtain that $f^1 \geq 0$.

- Thirdly, let us assess the separation distance between f^1 and f^0. We have

$$\Delta_{f^1} = C_1\delta = C_1 \exp(-pK_{\mathtt{L}}^r) \geq M\varepsilon_{\mathtt{L}} \quad \text{for} \quad 0 < M < C_1,$$

and thus $f^1 \in \mathcal{H}_a(\mathcal{A}_{p,r}(Q), M, \varepsilon)$.

- χ^2 distance

Using (16), observe

$$\begin{aligned} N\chi^2(g^0,g^1) &= N\int \frac{\left(g^1-g^0\right)^2}{g^0} = \frac{N}{\mathfrak{C}_0}\|g^1-g^0\|_2^2 = \frac{N\delta^2C_1^2}{\mathfrak{C}_0}\left\|\hat{h}_{K_{\mathtt{L}}}e_0\right\|^2 \\ &\leq \frac{Nb_1^2}{\mathfrak{C}_0}\delta^2C_1^2K_{\mathtt{L}}^{-2\nu_1}\exp(-2qK_{\mathtt{L}}^s) = \frac{b_1^2C_1^2}{\mathfrak{C}_0}\exp(\log N - 2pK_{\mathtt{L}}^r - 2qK_{\mathtt{L}}^s - 2\nu_1\log K_{\mathtt{L}}) \end{aligned}$$

where $e_0 = (0,\ldots,0,1,0,\ldots,0) \in \mathbb{R}^{d_{K_{\mathtt{L}}}}$, is a $2K_{\mathtt{L}}+1$ vector with zeros at $K_{\mathtt{L}}$ first positions, 1 at the $K_{\mathtt{L}}+1$ position, then zeros at $K_{\mathtt{L}}$ last positions. But

$$\begin{aligned} &2pK_{\mathtt{L}}^r + 2qK_{\mathtt{L}}^s \\ &= \log N - 2q\left(\frac{\log N}{2p}\right)^{s/r} - \frac{2\nu_0+1-\frac{s}{2}}{r}\log\frac{\log N}{2p} + 2q\left[\left(\frac{\log N}{2p}\right)^{s/r} + o(1)\right] \\ &= \log N - (2\nu_0+1-s/2)\log\left(\frac{\log N}{2p}\right)^{1/r} + o(1) \\ &= \log N - (2\nu_0+1-s/2)\log K_{\mathtt{L}} + o(1), \end{aligned}$$

we obtain

$$\begin{aligned} N\chi^2(g^0,g^1) &\ll \exp(\log N - 2pK_{\mathtt{L}}^r - 2qK_{\mathtt{L}}^s - 2\nu_1\log K_{\mathtt{L}}) \\ &\ll \exp[-(2\nu_1 - 2\nu_0 - 1 + s/2)\log K_{\mathtt{L}}] = o(1). \end{aligned}$$

Therefore as soon as $2\nu_1 - 2\nu_0 - 1 + s/2 > 0$

$$\begin{aligned} \inf_T R_N(T, \mathcal{A}_{p,r}(Q), M, \varepsilon) &= \inf_T\left[\mathbb{P}_{f^0}(T=1) + \sup\left\{\mathbb{P}_f(T=0) : f \in \mathcal{H}_a(\mathcal{A}_{p,r}(Q), M, \varepsilon)\right\}\right] \\ &\geq \inf_T\left[\mathbb{P}_{f^0}(T=1) + \mathbb{P}_{f^1}(T=0)\right] \\ &\geq \int \min(d\mathbb{P}_{f^0}, d\mathbb{P}_{f^1}) \\ &\geq 1 - \sqrt{N\chi^2(g^0,g^1)} \\ &\geq \eta. \end{aligned}$$

The proof is complete.

□

Proof of Theorem 2

Proof. The proof is similar to Theorem 1.

Let

$$K_{\mathtt{L}} = \left\lfloor \frac{\log N}{2(p+q)} - \frac{2\bar{\nu}_1}{2(p+q)r} \log \frac{\log N}{2(p+q)} \right\rfloor^{1/r} \quad \text{for} \quad \bar{\nu}_1 < \nu_1$$

Define f^1 as in Theorem 1:

$$f^1 := f^0 + C_1 \Psi,$$

with $\Psi = \delta \phi_{K_{\mathtt{L}}0}$, $\delta = \exp(-pK_{\mathtt{L}}^r)$ and $C_1 = \min(Q, \frac{\mathfrak{C}_0}{C(p,r,Q)})$.

To prove that f_1 belongs to the space $\mathcal{A}_{p,r}(Q)$, the density condition and the separation distance criteria, the arguments are identical to those of Theorem 1.

It remains to check the χ^2 distance condition between the two models.

We have

$$\begin{aligned} N\chi^2(g^0, g^1) &= \frac{b_1^2 C_1^2}{\mathfrak{C}_0} \exp\left(\log N - 2pK_{\mathtt{L}}^r - 2qK_{\mathtt{L}}^s - 2\nu_1 \log K_{\mathtt{L}}\right) \\ &\ll \exp\left(\log N - 2(p+q)K_{\mathtt{L}}^r - 2\nu_1 \log K_{\mathtt{L}}\right) \\ &\ll \exp\left(\log N - \log N - 2(\nu_1 - \bar{\nu}_1) \log K_{\mathtt{L}}\right) \\ &= o(1), \end{aligned}$$

as soon as $\nu_1 - \bar{\nu}_1 > 0$, which concludes the proof.

□

Proof of Theorem 3

Proof. We again follow the lines of the proof of Theorem 1. This time, let

$$K_{\mathtt{L}} = \left\lfloor \frac{\log N}{2q} - \frac{2\nu_0 + 1 - s/2}{2qs} \log \frac{\log N}{2q} - \frac{p}{q} \left(\frac{\log N}{2q}\right)^{r/s} \right\rfloor^{1/s}$$

and

$$\varepsilon_{\mathtt{L}} = \exp(-pK_{\mathtt{L}}^r).$$

Observe

$$K_{\mathtt{L}}^r \leq \left(\frac{\log N}{2q}\right)^{r/s} + C\left(\log N\right)^{2r/s-1}\big(1 + o(1)\big) + C\left(\log N\right)^{r/s-1} \log\log N\big(1 + o(1)\big).$$

Hence

$$\varepsilon_{\mathtt{L}} = \exp(-pK_{\mathtt{L}}^r) = \exp\left[-p\left(\frac{\log N}{2q}\right)^{r/s}\right]\big(1 + o(1)\big).$$

Define f^1 as in Theorem 1:

$$f^1 := f^0 + C_1 \Psi,$$

with $\Psi = \delta \phi_{K_{\mathtt{L}}0}$, $\delta = \exp(-pK_{\mathtt{L}}^r)$ and $C_1 = \min(Q, \frac{\mathfrak{C}_0}{C(p,r,Q)})$.

To prove that f_1 belongs to the space $\mathcal{A}_{p,r}(Q)$, the density condition and the separation distance criteria, the arguments are identical to those of Theorem 1.

Concerning the χ^2 distance between the two models we have:

$$\begin{aligned}
&2pK_{\mathtt{L}}^r + 2qK_{\mathtt{L}}^s \\
&= 2p\left(\frac{\log N}{2q}\right)^{r/s} + o(1) + \log N - \frac{2\nu_0 + 1 - s/2}{s}\log\frac{\log N}{2q} - 2p\left(\frac{\log N}{2q}\right)^{r/s} \\
&= \log N - (2\nu_0 + 1 - s/2)\log\left(\frac{\log N}{2q}\right)^{1/s} + o(1) \\
&= \log N - (2\nu_0 + 1 - s/2)\log K_{\mathtt{L}} + o(1).
\end{aligned}$$

Then we obtain

$$N\chi^2(g^0, g^1) \ll \exp\left(\log N - 2pK_{\mathtt{L}}^r - 2qK_{\mathtt{L}}^s - 2\nu_1 \log K_{\mathtt{L}}\right) \ll \exp\left[-\left(2\nu_1 - 2\nu_0 - 1 + s/2\right)\log K_{\mathtt{L}}\right] = o(1)$$

as soon as $\nu_1 > \nu_0 + (2-s)/4$ which completes the proof.

□

Proof of Theorem 4

Proof. • Suppose $s > 1$. We have

$$L = \left\lceil \left(\frac{\log\log N}{s}\right)^{1/s} \right\rceil,$$

and note that $|\mathcal{K}| \le C(\log N)^{1/s}$.

Let us control the risk of first kind. Using Lemma 1 implies

$$\begin{aligned}
\mathbb{P}_{f^0}\left(T_{\mathtt{A}} = 1\right) &\le \sum_{\mathcal{K}} \mathbb{P}_{f^0}\left(|U_K| > \frac{\tau_K}{\sqrt{M_2}}\right) = \sum_{\mathcal{K}} \mathbb{P}_{f^0}\left(|U_K| > \xi\frac{V_{0K}}{\sqrt{M_2}N}\right) \\
&\le \sum_{\mathcal{K}} M_1 \exp\left(-M_2\xi^2/M_2\right) = \sum_{\mathcal{K}} M_1 \exp\left(-K^s\right) \\
&\le M_1 \exp\left(-L^s\right)|\mathcal{K}| \ll (\log N)^{-1/s} = o(1).
\end{aligned}$$

We shall control the risk of second kind. Let the optimal cutoff point be

$$K_\star := \left\lfloor \frac{\log N}{2p} - \frac{2\nu_0 + 1}{pr}\log\frac{\log N}{2p} - \frac{q}{p}\left(\frac{\log N}{2p}\right)^{s/r} \right\rfloor^{1/r}.$$

When N is sufficiently large,

$$\left(\frac{\log\log N}{s}\right)^{\frac{1}{s}} \le K_\star \le \left(\frac{\log N}{2p_0}\right)^{1/s},$$

so that $K_\star \in \mathcal{K}$.

Define $B_K = Q^2\exp\left(-2pK^r\right)$. Observe

$$B_{K_\star} = Q^2\exp(-2pK_\star^r) \le Q^2\exp\left[-\log N + \frac{2\nu_0 + 1}{r}\log\frac{\log N}{2p} + 2q\left(\frac{\log N}{2p}\right)^{s/r}\right] = Q^2\varepsilon_{\mathtt{U}}^2$$

and

$$\begin{aligned}
\tau_{K_\star} &= \frac{\exp(2qK_\star^s)K_\star^{2\nu_0+1-s/2}K_\star^{s/2}}{N} = \frac{\exp(2qK_\star^s)K_\star^{2\nu_0+1}}{N} \\
&\leq \exp\left[-\log N + 2q\left(\frac{\log N}{2p}\right)^{s/r} + \frac{2\nu_0+1}{r}\log\left(\frac{\log N}{2p}\right)\right] \\
&\qquad \times \exp\left[C_1(\log N)^{2s/r-1}\big(1+o(1)\big) + C_2(\log N)^{s/r-1}\log\log N\big(1+o(1)\big)\right] \\
&\asymp \varepsilon_{\mathtt{U}}^2.
\end{aligned}$$

Let

$$E_K := \Delta_f^2 - B_K - \tau_K.$$

Hence we have for $E_{K_\star}$ using that $\Delta_f \geq M\varepsilon_{\mathtt{U}}$

$$\begin{aligned}
E_{K_\star} &= \Delta_f^2 - B_{K_\star} - \tau_{K_\star} \geq \Delta_f^2 - Q^2\varepsilon_{\mathtt{U}}^2 - \varepsilon_{\mathtt{U}}^2(1+o(1)) \\
&\geq \Delta_f^2\left(1 - \frac{Q^2+1+o(1)}{M^2}\right).
\end{aligned}$$

We want $1 - \frac{Q^2+1+o(1)}{M^2} > 0$. Hence we choose $M > M_{1,\mathtt{U}}^\star := (Q^2+2)^{\frac{1}{2}}$. We can now apply Lemma 8. We get

$$\mathbb{P}_f(T_{K_\star} = 0) \leq \frac{M_{\mathtt{V}}}{\left(M_{1,\mathtt{U}}^\star\right)^2}\left[\frac{N^{-1}V_{0K_\star}}{\Delta_f^2} + \frac{N^{-1}V_{1K_\star}}{\Delta_f} + \left(\frac{N^{-1}V_{0K_\star}}{\Delta_f^2}\right)^2 + \left(\frac{N^{-1}V_{1K_\star}}{\Delta_f}\right)^2 + \frac{1}{N}\right], \quad \text{for} \quad M_{\mathtt{V}} > 0,$$

where

$$V_{0K_\star} = K_\star^{2\nu_0+1-s/2}\exp\left(2qK_\star^s\right) \quad \text{and} \quad V_{1K_\star} = K_\star^{2\nu_0+2-s}\exp\left(2qK_\star^s\right).$$

Moreover,

$$\frac{N^{-1}V_{0K_\star}}{\Delta_f^2} \ll \frac{K_\star^{-s/2}\varepsilon_{\mathtt{U}}^2}{\varepsilon_{\mathtt{U}}^2} \ll K_\star^{-s/2} = o(1)$$

and

$$\frac{N^{-1}V_{1K_\star}}{\Delta_f} \leq \frac{N^{-1}V_{0K}K_\star^{1-s/2}}{M\varepsilon_{\mathtt{U}}} \ll \frac{\varepsilon_{\mathtt{U}}^2K_\star^{1-s/2}}{\varepsilon_{\mathtt{U}}} = \varepsilon_{\mathtt{U}}K_\star^{1-s/2} = o(1).$$

The desired result follows from

$$\mathbb{P}_f\left(T_{\mathtt{A}} = 0\right) = \mathbb{P}_f\left(|U_K| \leq \tau_K, \forall K \in \mathcal{K}\right) \leq \mathbb{P}_f\left(|U_{K_\star}| \leq \tau_{K_\star}\right) = \mathbb{P}_f(T_{K^\star} = 0) = o(1).$$

- Suppose $0 \leq s \leq 1$. Then

$$L = \left\lceil\left(\frac{\log\log N}{s}\right)^{1/(2-s)}\right\rceil,$$

and $|\mathcal{K}| \leq C(\log N)^{1/s}$. Lemma 1 implies

$$\begin{aligned}
\mathbb{P}_{f^0}(T=1) &\leq \sum_{\mathcal{K}}\mathbb{P}_{f^0}\left(|U_K| > \frac{\tau_K}{\sqrt{M_2}}\right) \leq \sum_{\mathcal{K}} M_1\exp\left(-M_2\xi^2/M_2\right) = \sum_{\mathcal{K}} M_1\exp\left(-K^{2-s}\right) \\
&\leq M_1\exp\left(-L^{2-s}\right)|\mathcal{K}| \ll (\log N)^{-1/s} = o(1).
\end{aligned}$$

Now, we control the risk of second kind. Let the optimal cutoff point be

$$K_\star := \left\lfloor\frac{\log N}{2p} - \frac{2\nu_0+1}{pr}\log\frac{\log N}{2p} - \frac{q}{p}\left(\frac{\log N}{2p}\right)^{s/r}\right\rfloor^{1/r}.$$

When N is sufficiently large,

$$\left(\frac{\log\log N}{s}\right)^{\frac{1}{2-s}} \leq K_\star \leq \left(\frac{\log N}{2p_0}\right)^{1/s},$$

so that $K_\star \in \mathcal{K}$.

Observe

$$\begin{aligned} B_{K_\star} &= Q^2 \exp(-2pK_\star^r) \\ &= Q^2 \exp\left[-\log N + \frac{2\nu_0+1}{r}\log\frac{\log N}{2p} + 2q\left(\frac{\log N}{2p}\right)^{s/r}\right] \\ &= Q^2 \exp\left(\frac{s-1}{r}\log\frac{\log N}{2p}\right)\varepsilon_{\mathtt{U}}^2 \\ &= a_N \varepsilon_{\mathtt{U}}^2, \end{aligned}$$

with $\lim_{N\to\infty} a_N = 0$. Besides we have

$$\begin{aligned} \tau_{K_\star} &= \frac{\exp(2qK_\star^s)K_\star^{2\nu_0+1-s/2}K_\star^{1-s/2}}{N} = \frac{\exp(2qK_\star^s)K_\star^{2\nu_0+2-s}}{N} \\ &\leq \exp\left[-\log N + 2q\left(\frac{\log N}{2p}\right)^{s/r} + \frac{2\nu_0+2-s}{r}\log\left(\frac{\log N}{2p}\right)\right] \\ &\quad \times \exp\left[C\left(\frac{\log N}{2p}\right)^{s/r-1}\log\frac{\log N}{2p}\big(1+o(1)\big)\right] \\ &\asymp \varepsilon_{\mathtt{U}}^2. \end{aligned}$$

Hence we have for $E_{K_\star}$ using that $\Delta_f \geq M\varepsilon_{\mathtt{U}}$

$$\begin{aligned} E_{K_\star} &= \Delta_f^2 - B_{K_\star} - \tau_{K_\star} \geq \Delta_f^2 - a_N\varepsilon_{\mathtt{U}}^2 - \varepsilon_{\mathtt{U}}^2(1+o(1)) \\ &\geq \Delta_f^2\left(1 - \frac{2}{M^2}\right). \end{aligned}$$

We want $1 - \frac{2}{M^2} > 0$. Hence we choose $M > \sqrt{2}$. We now apply Lemma 8. We get

$$\mathbb{P}_f(T_{K^\star} = 0) \leq \frac{M_{\mathtt{V}}}{2}\left[\frac{N^{-1}V_{0K^\star}}{\Delta_f^2} + \frac{N^{-1}V_{1K^\star}}{\Delta_f} + \left(\frac{N^{-1}V_{0K^\star}}{\Delta_f^2}\right)^2 + \left(\frac{N^{-1}V_{1K^\star}}{\Delta_f}\right)^2 + \frac{1}{N}\right].$$

Moreover,

$$\frac{N^{-1}V_{0K_\star}}{\Delta_f^2} \ll \frac{K_\star^{-(1-s/2)}\varepsilon_{\mathtt{U}}^2}{\varepsilon_{\mathtt{U}}^2} = K_\star^{-(1-s/2)} = o(1)$$

and

$$\frac{N^{-1}V_{1K_\star}}{\Delta_f} \leq \frac{N^{-1}V_{0K}K_\star^{1-s/2}}{M\varepsilon_{\mathtt{U}}} \ll \frac{\varepsilon_{\mathtt{U}}^2 K_\star^{1-s/2}}{\varepsilon_{\mathtt{U}}} = \varepsilon_{\mathtt{U}}K_\star^{1-s/2} = o(1).$$

And consequently similarly to the case $s > 1$

$$\mathbb{P}_f\left(T_{\mathtt{A}} = 0\right) = o(1).$$

□

Proof of Theorem 5

Proof. We recall that we are in the case $r = s$.

- Suppose $r > 1$. We have

$$L = \left\lceil \left(\frac{\log \log N}{r} \right)^{1/r} \right\rceil,$$

and $|\mathcal{K}| \leq C(\log N)^{1/r}$. Lemma 1 implies

$$\begin{aligned} \mathbb{P}_{f^0}(T_A = 1) &\leq \sum_{\mathcal{K}} \mathbb{P}_{f^0}\left(|U_K| > \frac{\tau_K}{\sqrt{M_2}} \right) \leq \sum_{\mathcal{K}} M_1 \exp\left(-\xi^2\right) = \sum_{\mathcal{K}} M_1 \exp\left(-K^r\right) \\ &\leq M_1 \exp\left(-L^r\right) |\mathcal{K}| \ll (\log N)^{-1/r} = o(1). \end{aligned}$$

Let us focus on the second kind risk error. Let the optimal cutoff point be

$$K_\star = \left\lfloor \frac{\log N}{2(p+q)} - \frac{2\nu_0 + 1 - \frac{s}{2}}{2(p+q)r} \log \frac{\log N}{2(p+q)} \right\rfloor^{1/r}.$$

When N is sufficiently large,

$$\left(\frac{\log \log N}{s} \right)^{\frac{1}{r}} \leq K_\star \leq \left(\frac{\log N}{2q} \right)^{1/r},$$

so that $K_\star \in \mathcal{K}$.

Observe

$$B_{K_\star} = Q^2 \exp(-2pK_\star^r) \leq Q^2 \exp\left[-\frac{p}{p+q} \left(\log N - \frac{2\nu_0 + 1}{r} \log \frac{\log N}{2(p+q)} \right) \right] = Q^2 \varepsilon_{\mathtt{U}}^2,$$

and

$$\begin{aligned} \tau_{K_\star} &= \frac{\exp(2qK_\star^r) K_\star^{2\nu_0 + 1 - r/2} K_\star^{r/2}}{N} = \frac{\exp(2qK_\star^r) K_\star^{2\nu_0+1}}{N} \\ &\leq \exp\left[-\log N + \frac{2\nu_0 + 1}{r} \log \frac{\log N}{2(p+q)} + o(1) + 2q \left(\frac{\log N}{2(p+q)} - \frac{2\nu_0 + 1}{2(p+q)r} \log \frac{\log N}{2(p+q)} \right) \right] \\ &\leq \exp\left[-\left(1 - \frac{q}{p+q}\right) \log N + \left(1 - \frac{q}{p+q}\right) \frac{2\nu_0 + 1}{r} \log \frac{\log N}{2(p+q)} + o(1) \right] \\ &\leq \exp\left[-\frac{p}{p+q} \left(\log N - \frac{2\nu_0 + 1}{r} \log \frac{\log N}{2(p+q)} \right) \right] \left(1 + o(1)\right) \\ &\asymp \varepsilon_{\mathtt{U}}^2. \end{aligned}$$

Hence we have for $E_{K_\star}$ using that $\Delta_f \geq M\varepsilon_{\mathtt{U}}$

$$\begin{aligned} E_{K_\star} &= \Delta_f^2 - B_{K_\star} - \tau_{K_\star} \geq \Delta_f^2 - Q^2 \varepsilon_{\mathtt{U}}^2 - \varepsilon_{\mathtt{U}}^2 (1 + o(1)) \\ &\geq \Delta_f^2 \left(1 - \frac{Q^2 + 1 + o(1)}{M^2} \right). \end{aligned}$$

We want $1 - \frac{Q^2+1+o(1)}{M^2} > 0$. Hence we choose again $M > M_{1,\mathtt{U}}^\star$. We can now apply Lemma 8. We get

$$\mathbb{P}_f(T_{K^\star} = 0) \leq \frac{M_{\mathtt{V}}}{\left(M_{1,\mathtt{U}}^\star\right)^2} \left[\frac{N^{-1} V_{0K^\star}}{\Delta_f^2} + \frac{N^{-1} V_{1K^\star}}{\Delta_f} + \left(\frac{N^{-1} V_{0K^\star}}{\Delta_f^2} \right)^2 + \left(\frac{N^{-1} V_{1K^\star}}{\Delta_f} \right)^2 + \frac{1}{N} \right].$$

Moreover,

$$\frac{N^{-1} V_{0K_\star}}{\Delta_f^2} \ll \frac{K_\star^{-r/2} \varepsilon_{\mathtt{U}}^2}{\varepsilon_{\mathtt{U}}^2} \leq K_\star^{-r/2} = o(1)$$

and

$$\frac{N^{-1}V_{1K_\star}}{\Delta_f} \le \frac{N^{-1}V_{0K}K_\star^{1-r/2}}{M\varepsilon_{\mathtt{U}}} \ll \frac{\varepsilon_{\mathtt{U}}^2 K_\star^{1-r/2}}{\varepsilon_{\mathtt{U}}} = \varepsilon_{\mathtt{U}} K_\star^{1-r/2} = o(1).$$

The desired result follows from

$$\mathbb{P}_f\left(T_{\mathtt{A}}=0\right) = \mathbb{P}_f\left(|U_K| \le \tau_K,\, \forall K \in \mathcal{K}\right) \le \mathbb{P}_f\left(|U_{K_\star}| \le \tau_{K_\star}\right) = \mathbb{P}_f(T_{K^\star}=0) = o(1).$$

- Suppose $0 \le r \le 1$. Then

$$L = \left\lceil \left(\frac{\log\log N}{r}\right)^{1/(2-r)} \right\rceil,$$

and $|\mathcal{K}| \le C(\log N)^{1/r}$.

Lemma 1 implies

$$\begin{aligned}
\mathbb{P}_{f^0}\left(T_{\mathtt{A}}=1\right) &\le \sum_{\mathcal{K}} \mathbb{P}_{f^0}\left(|U_K| > \frac{\tau_K}{\sqrt{M_2}}\right) \le \sum_{\mathcal{K}} M_1 \exp\left(-\xi^2\right) = \sum_{\mathcal{K}} M_1 \exp\left(-K^{2-r}\right) \\
&\le M_1 \exp\left(-L^{2-r}\right)|\mathcal{K}| \ll (\log N)^{-1/s} = o(1).
\end{aligned}$$

We shall deal with the error of second kind. Let the optimal cutoff point be

$$K_\star = \left\lfloor \frac{\log N}{2(p+q)} - \frac{2\nu_0+1-\frac{s}{2}}{2(p+q)r}\log\frac{\log N}{2(p+q)} \right\rfloor^{1/r}.$$

When N is sufficiently large,

$$\left(\frac{\log\log N}{s}\right)^{\frac{1}{2-r}} \le K_\star \le \left(\frac{\log N}{2q}\right)^{1/r},$$

so that $K_\star \in \mathcal{K}$.

We still have

$$B_{K_\star} \le Q^2 \exp\left[-\frac{p}{p+q}\left(\log N - \frac{2\nu_0+1}{r}\log\frac{\log N}{2(p+q)}\right)\right] = Q^2\varepsilon_{\mathtt{U}}^2.$$

Observe

$$\begin{aligned}
\tau_{K_\star} &= \frac{\exp(2qK_\star^s)K_\star^{2\nu_0+1-r/2}K_\star^{1-r/2}}{N} = \frac{\exp(2qK_\star^r)K_\star^{2\nu_0+2-s}}{N} \\
&\le \exp\left[-\log N + \frac{2\nu_0+2-s}{r}\log\frac{\log N}{2(p+q)} + o(1) + 2q\left(\frac{\log N}{2(p+q)} - \frac{2\nu_0+2-s}{(p+q)r}\log\frac{\log N}{2(p+q)}\right)\right] \\
&\le \exp\left[-\left(1-\frac{q}{p+q}\right)\log N + \left(1-\frac{q}{p+q}\right)\frac{2\nu_0+2-s}{r}\log\frac{\log N}{2(p+q)} + o(1)\right] \\
&\le \exp\left[-\frac{p}{p+q}\left(\log N - \frac{2\nu_0+2-s}{r}\log\frac{\log N}{2(p+q)}\right)\right]\left(1+o(1)\right) \\
&\asymp \varepsilon_{\mathtt{U}}^2.
\end{aligned}$$

Choosing $M > M_{1,\mathtt{U}}^\star$, we again apply Lemma 8. Since

$$\frac{N^{-1}V_{0K_\star}}{\Delta_f^2} \ll \frac{K_\star^{-(1-s/2)}\varepsilon_{\mathtt{U}}^2}{\varepsilon_{\mathtt{U}}^2} = K_\star^{-(1-s/2)} = o(1)$$

and

$$\frac{N^{-1}V_{1K_\star}}{\Delta_f} \le \frac{N^{-1}V_{0K}K_\star^{1-s/2}}{M\varepsilon_{\mathtt{U}}} \ll \frac{\varepsilon_{\mathtt{U}}^2 K_\star^{1-s/2}}{\varepsilon_{\mathtt{U}}} = \varepsilon_{\mathtt{U}} K_\star^{1-s/2} = o(1),$$

hence

$$\mathbb{P}_f\left(T_{\mathtt{A}}=0\right)=o(1),$$

which completes the proof.

□

Proof of Theorem 6

Proof. • Suppose $s>1$. Then

$$L=\left\lceil\left(\frac{\log\log N}{s}\right)^{1/s}\right\rceil,$$

and $|\mathcal{K}|\le C(\log N)^{1/s}$. For the risk of first kind, Lemma 1 implies

$$\begin{aligned}\mathbb{P}_{f^0}\left(T_{\mathtt{A}}=1\right) &\le \sum_{\mathcal{K}}\mathbb{P}_{f^0}\left(|U_K|>\frac{\tau_K}{\sqrt{M_2}}\right)\le\sum_{\mathcal{K}}M_1\exp\left(-\xi^2\right)=\sum_{\mathcal{K}}M_1\exp\left(-K^s\right)\\ &\le M_1\exp\left(-L^s\right)|\mathcal{K}|\ll(\log N)^{-1/s}=o(1).\end{aligned}$$

Let us deal with the risk of second kind. Let the optimal cutoff point be

$$K_\star=\left\lfloor\frac{\log N}{2q}-\frac{2\nu_0+1-\frac{s}{2}}{2qs}\log\frac{\log N}{2q}-\frac{p}{q}\left(\frac{\log N}{2q}\right)^{r/s}\right\rfloor^{1/s}.$$

When N is sufficiently large,

$$\left(\frac{\log\log N}{s}\right)^{\frac{1}{s}}\le K_\star\le\left(\frac{\log N}{2q}\right)^{1/s},$$

so that $K_\star\in\mathcal{K}$.

Observe

$$B_{K_\star}\asymp Q^2\exp\left[-2p\left(\frac{\log N}{2q}\right)^{r/s}\right]=Q^2\varepsilon_{\mathtt{U}}^2$$

and

$$\begin{aligned}\tau_{K_\star} &= \frac{\exp(2qK_\star^s)K_\star^{2\nu_0+1-s/2}K_\star^{s/2}}{N}=\frac{\exp(2qK_\star^s)K_\star^{2\nu_0+1}}{N}\\ &\le \exp\left[-\log N+\frac{2\nu_0+1}{s}\log\frac{\log N}{2q}\big(1+o(1)\big)+\log N-\frac{2\nu_0+1}{s}\log\frac{\log N}{2p}-2p\left(\frac{\log N}{2q}\right)^{r/s}\right]\\ &\quad\times\exp\left[C_1\left(\log N\right)^{2s/r-1}\big(1+o(1)\big)+C_2\left(\log N\right)^{s/r-1}\log\log N\big(1+o(1)\big)\right]\\ &\le \exp\left[-2p\left(\frac{\log N}{2q}\right)^{r/s}\right]\big(1+o(1)\big)\\ &\asymp \varepsilon_{\mathtt{U}}^2.\end{aligned}$$

Hence we have for $E_{K_\star}$ using that $\Delta_f\ge M\varepsilon_{\mathtt{U}}$

$$\begin{aligned}E_{K_\star} &= \Delta_f^2-B_{K_\star}-\tau_{K_\star}\ge\Delta_f^2-Q^2\varepsilon_{\mathtt{U}}^2(1+o(1))-\varepsilon_{\mathtt{U}}^2(1+o(1))\\ &\ge \Delta_f^2\left(1-\frac{Q^2+1+o(1)}{M^2}\right).\end{aligned}$$

We want $1-\frac{Q^2+1+o(1)}{M^2}>0$. Hence we choose $M>M_{1,\mathtt{U}}^\star$. We apply Lemma 8. Since

$$\frac{N^{-1}V_{0K_\star}}{\Delta_f^2} \ll \frac{K_\star^{-s/2}\varepsilon_{\mathtt{U}}^2}{\varepsilon_{\mathtt{U}}^2} \leq K_\star^{-s/2} = o(1)$$

and

$$\frac{N^{-1}V_{1K_\star}}{\Delta_f} \leq \frac{N^{-1}V_{0K}K_\star^{1-s/2}}{M\varepsilon_{\mathtt{U}}} \ll \frac{\varepsilon_{\mathtt{U}}^2 K_\star^{1-s/2}}{\varepsilon_{\mathtt{U}}} = \varepsilon_{\mathtt{U}} K_\star^{1-s/2} = o(1).$$

The desired result follows from

$$\mathbb{P}_f\left(T_{\mathtt{A}} = 0\right) = \mathbb{P}_f\left(|U_K| \leq \tau_K, \forall K \in \mathcal{K}\right) \leq \mathbb{P}_f\left(|U_{K_\star}| \leq \tau_{K_\star}\right) = o(1).$$

- Suppose $0 \leq s \leq 1$. Then

$$L = \left\lceil \left(\frac{\log\log N}{s}\right)^{1/(2-s)} \right\rceil,$$

and $|\mathcal{K}| \leq C(\log N)^{1/s}$.

For the risk of first kind, Lemma 1 implies

$$\begin{aligned}
\mathbb{P}_{f^0}\left(T = 1\right) &\leq \sum_{\mathcal{K}} \mathbb{P}_{f^0}\left(|U_K| > \frac{\tau_K}{\sqrt{M_2}}\right) \leq \sum_{\mathcal{K}} M_1 \exp\left(-\xi^2\right) = \sum_{\mathcal{K}} M_1 \exp\left(-K^{2-s}\right) \\
&\leq M_1 \exp\left(-L^{2-s}\right)|\mathcal{K}| \ll (\log N)^{-1/s} = o(1).
\end{aligned}$$

Let the optimal cutoff point be again

$$K_\star = \left\lfloor \frac{\log N}{2q} - \frac{\mu_0}{qs}\log\frac{\log N}{2q} - \frac{p}{q}\left(\frac{\log N}{2q}\right)^{r/s} \right\rfloor^{1/s}.$$

When N is sufficiently large,

$$\left(\frac{\log\log N}{s}\right)^{\frac{1}{2-s}} \leq K_\star \leq \left(\frac{\log N}{2q}\right)^{1/s},$$

so that $K_\star \in \mathcal{K}$ and $|\mathcal{K}| \leq C(\log N)^{1/s}$.

Observe

$$B_{K_\star} \asymp Q^2 \exp\left[-2p\left(\frac{\log N}{2q}\right)^{r/s}\right] = Q^2\varepsilon_{\mathtt{U}}^2$$

and

$$\tau_{K_\star} = \frac{\exp(2qK_\star^s)K_\star^{2\nu_0+1-s/2}K_\star^{1-s/2}}{N} = \frac{\exp(2qK_\star^s)K_\star^{2\nu_0+2-s}}{N} \asymp \varepsilon_{\mathtt{U}}^2.$$

Similarly to the case $s > 1$, we take $M > M_{1,\mathtt{U}}^\star$.

Moreover we apply Lemma 8,

$$\frac{N^{-1}V_{0K_\star}}{\Delta_f^2} \ll \frac{K_\star^{-(1-s/2)}\varepsilon_{\mathtt{U}}^2}{\varepsilon_{\mathtt{U}}^2} = K_\star^{-(1-s/2)} = o(1)$$

and

$$\frac{N^{-1}V_{1K_\star}}{\Delta_f} \leq \frac{N^{-1}V_{0K}K_\star^{1-s/2}}{M\varepsilon_{\mathtt{U}}} \ll \frac{\varepsilon_{\mathtt{U}}^2 K_\star^{1-s/2}}{\varepsilon_{\mathtt{U}}} = \varepsilon_{\mathtt{U}} K_\star^{1-s/2} = o(1).$$

The desired result follows from

$$\mathbb{P}_f\left(T_{\mathtt{A}} = 0\right) = o(1).$$

□

7 Appendix

Lemma 2. *For $p > 0$, $r > 0$ and $Q > 0$,*

$$\sup_{f \in \mathcal{A}_{p,r}(Q)} \|f\|_\infty \leq C(p,r,Q),$$

$$C(p,r,Q) = (\mathfrak{C}_0 + Q^2)^{\frac{1}{2}} \sqrt{\sum_{k \geq 0} d_k \exp\left(-2pk^r\right)}, \tag{30}$$

with $d_k = 2k+1$.

Proof. Choose $y \in \mathbb{S}^2$. The addition formula says

$$\|\phi_k(y)\|^2 = \frac{d_k}{4\pi}. \tag{31}$$

Cauchy-Schwarz inequality implies

$$|f(y)| \leq \sqrt{\sum_k \exp\left(2pk^r\right) \left\|\hat{f}_k\right\|^2} \sqrt{\sum_k \exp\left(-2pk^r\right) \|\phi_k(y)\|^2} \leq C(p,r,Q).$$

Note that $\sum_{k \geq 0} d_k \exp\left(-2pk^r\right)$ is convergent. □

Lemma 3. *Under Assumption 1. we have the following inequalities :*

$$\begin{aligned}
\sum_{k=1}^{K} d_k \left\|\hat{h}_k^{-1}\right\|_{\text{op}}^2 &\ll K^{2\nu_0+2-s} \exp\left(2qK^s\right), \\
\sum_{k=1}^{K} d_k^{3/2} \left\|\hat{h}_k^{-1}\right\|_{\text{op}}^4 &\ll K^{4\nu_0+5/2-s} \exp\left(4qK^s\right), \\
\sum_{k=1}^{K} d_k \left\|\hat{h}_k^{-1}\right\|_{\text{op}}^4 &\ll K^{4\nu_0+2-s} \exp\left(4qK^s\right), \\
\sum_{k=1}^{K} \left\|\hat{h}_k^{-1}\right\|_{\text{op}}^2 &\ll K^{2\nu_0+1-s} \exp\left(2qK^s\right).
\end{aligned}$$

Proof. Using Assumption 1, we have

$$\sum_{k=1}^{K} d_k \left\|\hat{h}_k^{-1}\right\|_{\text{op}}^2 \leq \sum_{k=1}^{K} (2k+1) b_0^2 k^{2\nu_0} \exp(2qk^s).$$

But standard integrals evaluation yields for any real α

$$\sum_{k=1}^{K} k^\alpha \exp(k^\beta/\delta) \leq CK^{\alpha+1-\beta} \exp(K^\beta/\delta).$$

Consequently

$$\sum_{k=1}^{K} d_k \left\|\hat{h}_k^{-1}\right\|_{\text{op}}^2 \ll K^{2\nu_0+2-s} \exp\left(2qK^s\right).$$

The proof of the other inequalities are similar. □

The following Lemma controls the bias of U_K.

Lemma 4. *Suppose $f \in \mathcal{A}_{p,r}(Q)$. Then the bias of U_K which is an estimate of Δ_f^2 satisfies*

$$\Delta_f^2 - \mathbb{E}_f U_K \leq B_K \quad \textit{for} \quad K \geq 1,$$

where $B_K = Q^2 \exp(-2pK^r)$.

Proof. Observe

$$\begin{aligned}
\Delta_f^2 - \mathbb{E}_f U_K &= \sum_{k=1}^{\infty} \left\|\hat{f}_k\right\|^2 - \sum_{k=1}^{K} \left\|\hat{f}_k\right\|^2 \\
&= \sum_{k>K}^{\infty} \left\|\hat{f}_k\right\|^2 \leq \exp(-2pK^r) \sum_{k>K}^{\infty} \exp(2pk^r) \left\|\hat{f}_k\right\|^2 \\
&\leq Q^2 \exp(-2pK^r) = B_K.
\end{aligned}$$

□

The next result gives a bound for the variance of U_K.

Lemma 5. *Suppose (15) holds. For arbitrary density f, the variance of U_K satisfies*

$$\mathrm{Var}_f(U_K) \leq M_{\mathtt{V}} \left[\frac{1}{N}\left(\Delta_f^4 + \Delta_f^2 V_{0K} + \Delta_f^3 V_{1K}\right) + \frac{1}{N^2}\left(V_{0K}^2 + \Delta_f^2 V_{1K}^2\right)\right] \quad \textit{for} \quad M_{\mathtt{V}} > 0,$$

with

$$V_{0K} = K^{2\nu_0 + 1 - s/2} \exp(2qK^s) \quad \textit{and} \quad V_{1K} = K^{2\nu_0 + 2 - s} \exp(2qK^s).$$

Proof. The proof of this lemma is identical to the proof of Lemma 4 in Lacour and Pham Ngoc (2014). Note that we deal with real valued quantities and do not need complex conjugaison. Using equation (18) in Lacour and Pham Ngoc (2014) which gives the expression of the variance of U_K we get

$$\begin{aligned}
&\mathrm{Var}_f(U_K) \\
&= \left[\frac{(N-2)(N-3)}{N(N-1)} - 1\right] \left(\sum_{k=1}^{K} \left\|\hat{f}_k\right\|^2\right)^2 \\
&\quad + \frac{2}{N(N-1)} \sum_{1 \leq k_1, k_2 \leq K} \mathbb{E}_f \langle \Phi_{k_1}(Y_1), \Phi_{k_1}(Y_2)\rangle \langle \Phi_{k_2}(Y_1), \Phi_{k_2}(Y_2)\rangle \\
&\quad + \frac{4(N-2)}{N(N-1)} \sum_{1 \leq k_1, k_2 \leq K} \mathbb{E}_f \left\langle \Phi_{k_1}(Y), \hat{f}_{k_1}\right\rangle \left\langle \Phi_{k_2}(Y), \hat{f}_{k_2}\right\rangle.
\end{aligned}$$

We control each of the three terms of the previous equality by following the lines of Lemma 4 in Lacour and Pham Ngoc (2014). In the sequel we denote $S_{0K} = \sqrt{\sum_{k=1}^{K} d_k \left\|\hat{h}_k^{-1}\right\|_{\mathrm{op}}^4}$ and $S_{1K} = \sum_{k=1}^{K} d_k \left\|\hat{h}_k^{-1}\right\|_{\mathrm{op}}^2$, we recall that $d_k = 2k+1$.

For the first term we get

$$\left[\frac{(N-2)(N-3)}{N(N-1)} - 1\right] \left(\sum_{k=1}^{K} \left\|\hat{f}_k\right\|^2\right)^2 \leq C \frac{\Delta_f^4}{N}.$$

For the second term we have

$$\sum_{1 \leq k_1, k_2 \leq K} \mathbb{E}_f \langle \Phi_{k_1}(Y_1), \Phi_{k_1}(Y_2)\rangle \langle \Phi_{k_2}(Y_1), \Phi_{k_2}(Y_2)\rangle = \sum_{1 \leq k_1, k_2 \leq K} \sum_{l_1=-k_1}^{k_1} \sum_{l_2=-k_2}^{k_2} \left|\mathbb{E}_f \Phi_{k_1 l_1}(Y) \Phi_{k_2 l_2}(Y)\right|^2$$

$$|\mathbb{E}_f \Phi_{k_1 l_1}(Y)\Phi_{k_2 l_2}(Y)|^2 \;=\; \left|\mathbb{E}_{f^0} \Phi_{k_1 l_1}(Y)\Phi_{k_2 l_2}(Y)\right|^2 + \|\Phi_{k_1 l_1}\Phi_{k_2 l_2}\|_2^2 \, \|f_Y - f^0\|.$$

We have

$$\sum_{1\le k_1,k_2\le K} \|\Phi_{k_1 l_1}\Phi_{k_2 l_2}\|_2^2 \, \|f_Y - f^0\| \le \Delta_f^2 \left(\sum_k d_k \left\|\hat{h}_k^{-1}\right\|_{\mathrm{op}}^2 \right)^2 .$$

And using Lemma 3 in Lacour and Pham Ngoc (2014)

$$\sum_{1\le k_1,k_2\le K} \left|\mathbb{E}_{f^0} \Phi_{k_1 l_1}(Y)\Phi_{k_1 l_1}(Y)\right|^2 \le \sum_{1\le k\le K} \left\|\hat{h}_k^{-1}\right\|_{\mathrm{op}}^4 d_k$$

So we get

$$\frac{2}{N(N-1)} \sum_{1\le k_1,k_2\le K} \mathbb{E}_f \langle \Phi_{k_1}(Y_1), \Phi_{k_1}(Y_2)\rangle \langle \Phi_{k_2}(Y_1), \Phi_{k_2}(Y_2)\rangle \le \frac{C}{N^2}\left[S_{0K}^2 + \Delta_f^2 S_{1K}^2\right].$$

It remains to bound the third term. We have

$$\begin{aligned}\frac{4(N-2)}{N(N-1)} \sum_{1\le k_1,k_2\le K} \mathbb{E}_f \left\langle \Phi_{k_1}(Y), \hat{f}_{k_1}\right\rangle \left\langle \Phi_{k_2}(Y), \hat{f}_{k_2}\right\rangle \;&=\; \sum_{1\le k_1,k_2\le K} \sum_{l_1,l_2} \mathbb{E}_f (\Phi_{k_1 l_1}(Y)\Phi_{k_1 l_1}(Y)) \hat{f}_{k_1 l_1} \hat{f}_{k_2 l_2} \\ &\le\; \frac{C}{N}\Delta_f^2 (S_{0K} + \Delta_f S_{1K});\end{aligned}$$

Finally we get

$$\mathrm{Var}_f\left(U_K\right) \;\ll\; \frac{1}{N}\left[\Delta_f^4 + \Delta_f^2 S_{0K} + \Delta_f^3 S_{1K}\right] + \frac{1}{N^2}\left[S_{0K}^2 + \Delta_f^2 S_{1K}^2\right].$$

Now using Lemma 3 which gives upperbounds for S_{0K} and S_{1K} conclude the proofs of Lemma 5. □

In the next lemma, we recall the exponential inequality for U-statistics due to Giné et al. (2000)

Lemma 6. *Let G denote a bounded canonical kernel, completely degenerate of the i.i.d variables $Y_1, \dots, Y_N$. There exist universal constants $\tilde{C}_1, \tilde{C}_2 > 0$ such that, for all $x > 0$,*

$$\mathbb{P}\left(\left| \sum_{1\le i_1\ne i_2\le N} G(Y_{i1}, Y_{i2}) \right| \ge x \right) \le \tilde{C}_1 \exp\left[-\tilde{C}_2 \min\left(\frac{x^{1/2}}{A_G^{1/2}}, \frac{x^{2/3}}{B_G^{2/3}}, \frac{x^2}{C_G^2}, \frac{x}{D_G} \right) \right]$$

where A_G, B_G, C_G, D_G are defined as

$$A_G := \|G\|_\infty, \quad B_G^2 := 2N \left\| \mathbb{E}(|G|^2(Y, \cdot) \right\|_\infty), \quad C_G^2 := N^2 \mathbb{E}(|G|^2(Y_1, Y_2))$$

and

$$D_G := N \sup\left\{ \mathbb{E}(G(Y_1, Y_2) u_1(Y_1) u_2(Y_2)) : \; \mathbb{E}u_1^2(Y) \le 1 \quad \textit{and} \quad \mathbb{E}u_2^2(Y) \le 1 \right\}.$$

We apply this lemma to the kernel G defined in (19):

$$G(y_1, y_2) = \sum_{k=1}^K \langle \Phi_k(y_1), \Phi_k(y_2)\rangle \quad \text{for} \quad y_1, y_2 \in \mathbb{S}^2$$

with $\Phi_k = \hat{h}_k^{-1}\phi_k$. The kernel G is degenerate for the Y_i under H_0.

Lemma 7. *Suppose (15) holds. Then,*

$$A_G \ll K^{2\nu_0 + 2 - s} \exp\left(2qK^s\right), \quad B_G^2 \ll N K^{4\nu_0 + 5/2 - s} \exp\left(4qK^s\right),$$

$$C_G^2 \ll N^2 K^{4\nu_0 + 2 - s} \exp\left(4qK^s\right) \quad \textit{and} \quad D_G \ll N\left[(\mathbb{1}_{(0\le s\le 1)} K^{2\nu_0} + \mathbb{1}_{(s>1)} K^{2\nu_0 + 1 - s}) \exp\left(2qK^s\right) \right].$$

Proof. Applying the arguments used to prove the case of ordinary smooth error in Lacour and Pham Ngoc (2014) (Lemma 2) to the case for super smooth error and using Lemma 3, we obtain bounds on A_G, B_G^2 and C_G^2. Indeed we have

$$A_G \ll \sum_{k=1}^{K} d_k \left\| \hat{h}_k^{-1} \right\|_{\text{op}}^2 \ll K^{2\nu_0+2-s} \exp\left(2qK^s\right).$$

Moreover

$$B_G^2 \ll N \sum_{k=1}^{K} d_k^{3/2} \left\| \hat{h}_k^{-1} \right\|_{\text{op}}^4 \ll NK^{4\nu_0+5/2-s} \exp\left(4qK^s\right),$$

and

$$C_G^2 \ll \sum_{k=1}^{K} d_k \left\| \hat{h}_k^{-1} \right\|_{\text{op}}^4 \ll NK^{4\nu_0+5/2-s} \exp\left(4qK^s\right).$$

For D_G, observe

$$\mathbb{E}_{f^0} G(Y_1, Y_2) u_1(Y_1) u_2(Y_2) \leq \alpha_K := \sum_{k=1}^{K} \left\| \hat{h}_k^{-1} \right\|_{\text{op}} \|\hat{u}_{1k}\| \left\| \hat{h}_k^{-1} \right\|_{\text{op}} \|\hat{u}_{2k}\|.$$

When $0 \leq s \leq 1$, we bound α_K as

$$\alpha_K \ll \left\| \hat{h}_K^{-1} \right\|_{\text{op}}^2 \sum_{k=1}^{K} \|\hat{u}_{1k}\| \|\hat{u}_{2k}\| \ll \left\| \hat{h}_K^{-1} \right\|_{\text{op}}^2 \sqrt{\mathbb{E}_{f^0} u_1^2(Y) \mathbb{E}_{f^0} u_2^2(Y)} \ll \left\| \hat{h}_K^{-1} \right\|_{\text{op}}^2 \leq b_0^2 K^{2\nu_0} \exp(2qK^s),$$

using Assumption 1. When $s > 1$, we bound α_K as

$$\alpha_K \ll \sqrt{\mathbb{E}_{f^0} u_1(Y)^2 \mathbb{E}_{f^0} u_2(Y)^2} \sum_{k=1}^{K} \left\| \hat{h}_k^{-1} \right\|_{\text{op}}^2 \ll \sum_{k=1}^{K} \left\| \hat{h}_k^{-1} \right\|_{\text{op}}^2.,$$

now using the last point of Lemma 3 proves the result. □

Proof of Lemma 1

Proof. By Lemma 6, we obtain

$$\mathbb{P}\left(|U_K| \geq x\right) \leq \tilde{C}_1 \exp\left[-\tilde{C}_2 \min\left(\frac{x^2}{C_G^2}, \frac{x}{D_G}, \frac{x^{1/2}}{A_G^{1/2}}, \frac{x^{2/3}}{B_G^{2/3}}\right)\right]$$

where $\tau = x/N(N-1)$ and $x = \tau N(N-1)$. Now let

$$\tau = \xi \frac{V_{0K}}{N} = \xi \frac{K^{2\nu_0+1-s/2} \exp\left(2qK^s\right)}{N}.$$

From Lemma 3 and Lemma 7, we have

$$\begin{aligned}
&\tilde{C}_2 \min\left(\frac{x^2}{C_G^2}, \frac{x}{D_G}, \frac{x^{1/2}}{A_G^{1/2}}, \frac{x^{2/3}}{B_G^{2/3}}\right) \\
&\geq M_2 \min\left(\xi^2,\ \xi^2,\ \ \xi^{1/2} N^{1/2} K^{(s-2)/4},\ \xi^{2/3} N^{1/3} K^{-1/6}\right) \\
&\geq M_2 \xi^2,
\end{aligned}$$

provided that $K^a N^{-1/3} = o(1)$, for any constant $a > 0$. □

Lemma 8. *Let the density $f \in \mathcal{H}_a(\mathcal{A}_{p,r}(Q), M, \varepsilon)$. Consider the test T_K defined in (21). If there exists $M_{\mathtt{U}}^\star > 0$ such that*

$$E_K := \Delta_f^2 - B_K - \tau_K \geq M_{\mathtt{U}}^\star \Delta_f^2, \tag{32}$$

with $B_K = Q^2 \exp\left(-2pK^r\right)$ then

$$\mathbb{P}_f\left(T_K = 0\right) \leq \frac{M_{\mathtt{V}}}{(M_{\mathtt{U}}^\star)^2}\left[\frac{N^{-1}V_{0K}}{\Delta_f^2} + \frac{N^{-1}V_{1K}}{\Delta_f} + \left(\frac{N^{-1}V_{0K}}{\Delta_f^2}\right)^2 + \left(\frac{N^{-1}V_{1K}}{\Delta_f}\right)^2 + \frac{1}{N}\right].$$

Proof. Suppose

$$E_K := \Delta_f^2 - B_K - \tau_K \geq M_{\mathtt{U}}^\star \Delta_f^2 \quad \text{for} \quad M_{\mathtt{U}}^\star > 0, \tag{33}$$

then Lemma 4 entails

$$\mathbb{E}_f\left(U_K\right) - \tau_K \geq \Delta_f^2 - B_K - \tau_K \geq M_{\mathtt{U}}^\star \Delta_f^2.$$

Using Chebyshev's inequality and Lemma 5, we obtain

$$\begin{aligned}
\mathbb{P}_f\left(T_K = 0\right) &= \mathbb{P}_f\Big(|U_K| \leq \tau_K\Big) \\
&\leq \mathbb{P}\Big(|U_K - \mathbb{E}_f\left(U_K\right)| \geq \mathbb{E}_f\big(U_K\big) - \tau_K\Big) \\
&\leq \mathbb{P}\Big(|U_K - \mathbb{E}_f\left(U_K\right)| \geq M_{\mathtt{U}}^\star \Delta_f^2\Big) \\
&\leq \frac{\mathrm{Var}_f\left(U_K\right)}{\left(M_{\mathtt{U}}^\star\right)^2 \Delta_f^4} \\
&\leq \frac{M_{\mathtt{V}}}{(M_{\mathtt{U}}^\star)^2}\left[\frac{N^{-1}V_{0K}}{\Delta_f^2} + \frac{N^{-1}V_{1K}}{\Delta_f} + \left(\frac{N^{-1}V_{0K}}{\Delta_f^2}\right)^2 + \left(\frac{N^{-1}V_{1K}}{\Delta_f}\right)^2 + \frac{1}{N}\right].
\end{aligned}$$

□

References

Belitser, E. and Levit, B. (2001). Asymptotically local minimax estimation of infinitely smooth density with censored data. *Ann. Inst. Statist. Math.*, 53(2):289–306.

Beran, R. J. (1968). Testing for uniformity on a compact homogeneous space. *J. Appl. Probability*, 5:177–195.

Butucea, C. (2007). Goodness-of-fit testing and quadratic functional estimation from indirect observations. *Ann. Statist.*, 35(5):1907–1930.

Butucea, C., Matias, C., and Pouet, C. (2009). Adaptive goodness-of-fit testing from indirect observations. *Ann. Inst. Henri Poincaré Probab. Stat.*, 45(2):352–372.

Butucea, C. and Tsybakov, A. B. (2007). Sharp optimality in density deconvolution with dominating bias. I. *Teor. Veroyatn. Primen.*, 52(1):111–128.

Comte, F. and Lacour, C. (2013). Anisotropic adaptive kernel deconvolution. *Ann. Inst. Henri Poincaré Probab. Stat.*, 49(2):569–609.

Comte, F., Rozenholc, Y., and Taupin, M.-L. (2006). Penalized contrast estimator for adaptive density deconvolution. *Canad. J. Statist.*, 34(3):431–452.

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, 19(3):1257–1272.

Fan, J. and Koo, J.-Y. (2002). Wavelet deconvolution. *IEEE Trans. Inform. Theory*, 48(3):734–747.

Giné, E. (1975). Invariant tests for uniformity on compact Riemannian manifolds based on Sobolev norms. *Ann. Statist.*, 3(6):1243–1266.

Giné, E., Latała, R., and Zinn, J. (2000). Exponential and moment inequalities for U-statistics. In *High dimensional probability, II (Seattle, WA, 1999)*, volume 47 of *Progr. Probab.*, pages 13–38. Birkhäuser Boston, Boston, MA.

Healy, Jr., D. M., Hendriks, H., and Kim, P. T. (1998). Spherical deconvolution. *J. Multivariate Anal.*, 67(1):1–22.

Huckemann, S. F., Kim, P. T., Koo, J.-Y., and Munk, A. (2010). Möbius deconvolution on the hyperbolic plane with application to impedance density estimation. *Ann. Statist.*, 38(4):2465–2498.

Ingster, Y. I. (1997). Adaptive chi-square tests. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)*, 244(Veroyatn. i Stat. 2):150–166, 333.

Kalifa, J., Mallat, S., and Rougé, B. (2003). Deconvolution by thresholding in mirror wavelet bases. *IEEE Trans. Image Process.*, 12(4):446–457.

Kerkyacharian, G., Pham Ngoc, T. M., and Picard, D. (2011). Localized spherical deconvolution. *Ann. Statist.*, 39(2):1042–1068.

Kim, P. T. and Koo, J.-Y. (2002). Optimal spherical deconvolution. *J. Multivariate Anal.*, 80(1):21–42.

Kim, P. T., Koo, J.-Y., and Park, H. J. (2004). Sharp minimaxity and spherical deconvolution for super-smooth error distributions. *J. Multivariate Anal.*, 90(2):384–392.

Lacour, C. and Pham Ngoc, T. (2014). Goodness-of-fit testing for noisy directional data. *Bernoulli*, 20(4):2131–2168.

Meister, A. (2009). *Deconvolution problems in nonparametric statistics*, volume 193 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin.

Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.*, 27(6):2033–2053.

Talman, J. D. (1968). *Special functions: A group theoretic approach.* Based on lectures by Eugene P. Wigner. With an introduction by Eugene P. Wigner. W. A. Benjamin, Inc., New York-Amsterdam.

Terras, A. (1985). *Harmonic analysis on symmetric spaces and applications. I.* Springer-Verlag, New York.

Tsybakov, A. B. (2009). *Introduction to nonparametric estimation.* Springer Series in Statistics. Springer, New York. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.

Vilenkin, N. J. (1968). *Special functions and the theory of group representations.* Translated from the Russian by V. N. Singh. Translations of Mathematical Monographs, Vol. 22. American Mathematical Society, Providence, R. I.