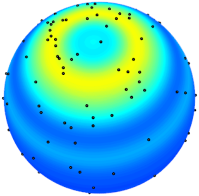
Kernel density estimation with directional data under rotational symmetry



Eduardo García-Portugués (egarcia@math.ku.dk)

Department of Mathematical Sciences and
The Bioinformatics Centre
University of Copenhagen

Joint work with C. Ley and T. Verdebout
(Université Libre de Bruxelles)

EYSM, Prague, 31st August 2015

Outline

1. **Introduction**
   - Rotasymmetry
   - KDE with directional data
2. **Density estimation under rotasymmetry**
   - The rotasymmetrizer
   - Rotasymmetric KDE
3. **Simulation study**

Outline

1. **Introduction**
   - Rotasymmetry
   - KDE with directional data
2. **Density estimation under rotasymmetry**
   - The rotasymmetrizer
   - Rotasymmetric KDE
3. **Simulation study**

Introduction

- Directional data are vectors whose support is the hypersphere

$$\Omega_q = \left\{ \mathbf{x} \in \mathbb{R}^{q+1} : ||\mathbf{x}|| = 1 \right\}.$$

- Particular cases are the circle ($q = 1$) and the sphere ($q = 2$).
- Statistical methods must account for the special nature of directional data.
- Present in different applied fields: corner stone in protein modelling.

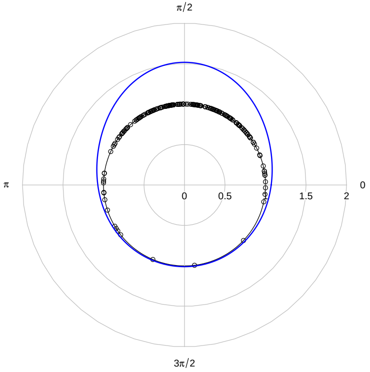


Figure: Circular von Mises density.

Introduction

- Directional data are vectors whose support is the hypersphere

$$\Omega_q = \left\{ \mathbf{x} \in \mathbb{R}^{q+1} : ||\mathbf{x}|| = 1 \right\}.$$

- Particular cases are the circle ($q = 1$) and the sphere ($q = 2$).
- Statistical methods must account for the special nature of directional data.
- Present in different applied fields: corner stone in protein modelling.

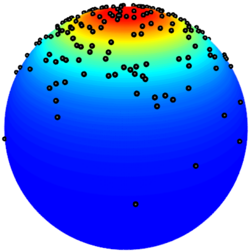


Figure: Spherical von Mises density.

Introduction

- Directional data are vectors whose support is the hypersphere

$$\Omega_q = \left\{ \mathbf{x} \in \mathbb{R}^{q+1} : ||\mathbf{x}|| = 1 \right\}.$$

- Particular cases are the circle ($q = 1$) and the sphere ($q = 2$).
- Statistical methods must account for the special nature of directional data.
- Present in different applied fields: corner stone in protein modelling.

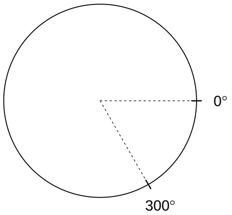


Figure: Circular mean.

Introduction

- Directional data are vectors whose support is the hypersphere

$$\Omega_q = \left\{\mathbf{x} \in \mathbb{R}^{q+1} : ||\mathbf{x}|| = 1\right\}.$$

- Particular cases are the circle ($q = 1$) and the sphere ($q = 2$).
- Statistical methods must account for the special nature of directional data.
- Present in different applied fields: corner stone in protein modelling.

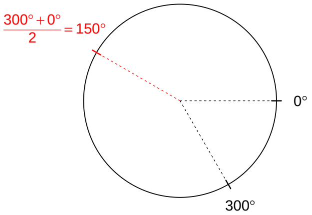


Figure: Circular mean.

Von Mises-Fisher distribution

- The most well known directional distribution is the von Mises-Fisher (vMF), with density:

$$f_{\mathrm{vMF}}(\mathbf{x};\boldsymbol{\mu},\kappa)=C_q(\kappa)\exp\left\{\kappa\mathbf{x}^T\boldsymbol{\mu}\right\},\quad C_q(\kappa)=\frac{\kappa^{\frac{q-1}{2}}}{(2\pi)^{\frac{q+1}{2}}\mathcal{I}_{\frac{q-1}{2}}(\kappa)}$$

  parametrized by a mean $\boldsymbol{\mu}\in\Omega_q$ and a concentration $\kappa\geq 0$.

- Density wrt the Lebesgue measure ω_q in Ω_q. ω_q denotes also the area surface of Ω_q:

$$\omega_q\equiv\omega_q(\Omega_q)=2\pi^{\frac{q+1}{2}}/\Gamma\left(\frac{q+1}{2}\right).$$

- **Gaussian analogue (isotropic)**:
  1. Same MLE characterization property.
  2. If $\mathbf{X}\sim\mathcal{N}_{q+1}\left(\boldsymbol{\mu},\sigma^2\mathbf{I}_{q+1}\right)$, with $\boldsymbol{\mu}\in\mathbb{R}^{q+1}\backslash\{\mathbf{0}\}$ and $\sigma^2>0$, then

$$\mathbf{Y}=\left(\mathbf{X}\,\middle|\,||\mathbf{X}||=1\right)\sim\mathrm{vM}\left(\frac{\boldsymbol{\mu}}{||\boldsymbol{\mu}||},\frac{||\boldsymbol{\mu}||}{\sigma^2}\right).$$

Rotasymmetry I

- A recurrent assumption about a directional rv $\mathbf{X}$ is **rotational symmetry** (or **rotasymmetry**) about some direction $\boldsymbol{\theta} \in \Omega_q$.
- In the circular case, rotasymmetry is **reflective symmetry**, a feature appearing in most of the distributions.
- In the high-dimensional situation, rotasymmetry is behind many **celebrated distributions** such as the vMF.

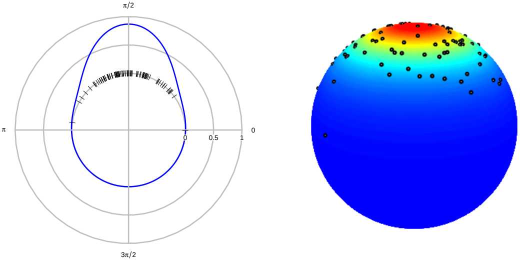


Figure: Rotasymmetry in the circular and spherical cases.

Rotasymmetry II

- It is a building block in numerous contributions: inference, simulation, descriptive statistics.

Saw, J. G. (1978). A family of distributions on the *m*-sphere and some hypothesis tests. *Biometrika*, 65:69–73.

Bingham, C. and Mardia, K. V. (1978). A small circle distribution on the sphere *Biometrika*, 65:379–389.

Wood, A. T. A (1994). Simulation of the von Mises Fisher distribution. *Commun. Stat. Simulat.*, 23:157–164.

Ley, C., Swan, Y., Thiam, B. and Verdebout, T. (2013). Optimal *R*-estimation of a spherical location. *Statist. Sinica*, 23:305–332.

Ley, C., Sabbah, C. and Verdebout, T. (2014). A new concept of quantiles for directional data and the angular Mahalanobis depth. *Electron. J. Stat.*, 8:795–816.

Rotasymmetry III

Proposition (Rotasymmetry characterization)

Let $\mathbf{X}$ *a directional rv with density* f*. These statements are equivalent:*

1. $\mathbf{X} \stackrel{d}{=} \mathbf{O}\mathbf{X}$, *where* $\mathbf{O} = \boldsymbol{\theta}\boldsymbol{\theta}^T + \sum_{i=1}^q \mathbf{b}_i\mathbf{b}_i^T$ *is a rotation matrix on* $\mathbb{R}^{q+1}$ *that fixes* $\boldsymbol{\theta} \in \Omega_q$.
2. $f(\mathbf{x}) = g\left(\mathbf{x}^T\boldsymbol{\theta}\right)$, $\forall \mathbf{x} \in \Omega_q$, *where* $g : [-1,1] \longrightarrow \mathbb{R}^+$ *is a* ***link*** *such that* $f^*(t) = \omega_{q-1} g(t)(1-t^2)^{\frac{q}{2}-1}$ *is a density in* $[-1,1]$.

- Rotasymmetry is related with the **tangent-normal decomposition**:

$$\mathbf{x} = t\boldsymbol{\theta} + (1-t^2)^{\frac{1}{2}}\mathbf{B}_{\boldsymbol{\theta}}\boldsymbol{\xi}, \quad \omega_q(d\mathbf{x}) = (1-t^2)^{\frac{q}{2}-1}\,dt\,\omega_{q-1}(d\boldsymbol{\xi}),$$

  with $t = \mathbf{x}^T\boldsymbol{\theta} \in [-1,1]$, $\boldsymbol{\xi} \in \Omega_{q-1}$ and $\mathbf{B}_{\boldsymbol{\theta}} = (\mathbf{b}_1, \ldots, \mathbf{b}_q)_{(q+1)\times q}$ such that $\mathbf{B}_{\boldsymbol{\theta}}^T\mathbf{B}_{\boldsymbol{\theta}} = \mathbf{I}_q$ and $\mathbf{B}_{\boldsymbol{\theta}}\mathbf{B}_{\boldsymbol{\theta}}^T = \mathbf{I}_{q+1} - \boldsymbol{\theta}\boldsymbol{\theta}^T$.

- **No monotonicity** required in g, axial variables are covered as well.

KDE with directional data

- For a sample $\mathbf{X}_1, \ldots, \mathbf{X}_n \sim f$, the Kernel Density Estimator (KDE) for directional data is

$$\hat{f}_h(\mathbf{x}) = \frac{c_{h,q}(L)}{n} \sum_{i=1}^{n} L\left(\frac{1 - \mathbf{x}^T \mathbf{X}_i}{h^2}\right) = \frac{1}{n} \sum_{i=1}^{n} L_h(\mathbf{x}, \mathbf{X}_i), \quad \mathbf{x} \in \Omega_q.$$

Bai, Z. D., Rao, C. R. and Zhao, L. C. (1988). Kernel estimators of density function of directional data. *J. Multivariate Anal.*, 27:24–39.

- Kernel: usually $L(r) = e^{-r}$, known as the *von Mises kernel*. In that case $c_{h,q}(L) = e^{1/h^2} C_q(1/h^2)$.
- Normalizing constant $c_{h,q}(L)^{-1} = \lambda_q(L) h^q (1 + o(1))$ with

$$\lambda_q(L) = 2^{\frac{q}{2}-1} \omega_{q-1} \int_0^\infty L(r) r^{\frac{q}{2}-1}\, dr.$$

- "Second moment" of L: $b_q(L) = \int_0^\infty L(r) r^{\frac{q}{2}}\, dr \big/ \int_0^\infty L(r) r^{\frac{q}{2}-1}\, dr$.
- Bandwidth: key parameter that controls the smoothness.

KDE construction: spherical case

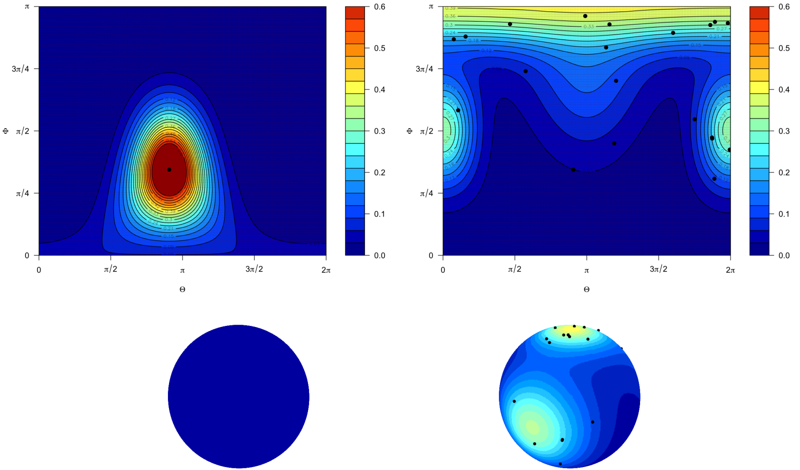


Figure: Left: KDE with $n = 1$. Right: true density.

KDE construction: spherical case

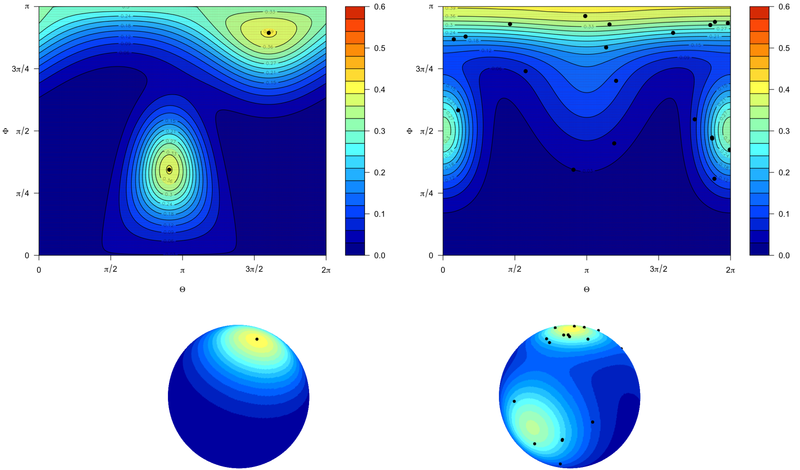


Figure: Left: KDE with $n = 2$. Right: true density.

KDE construction: spherical case

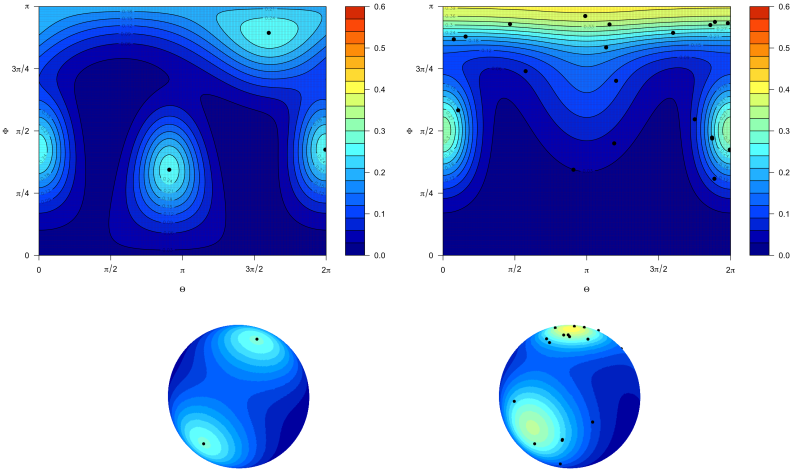


Figure: Left: KDE with $n = 3$. Right: true density.

KDE construction: spherical case

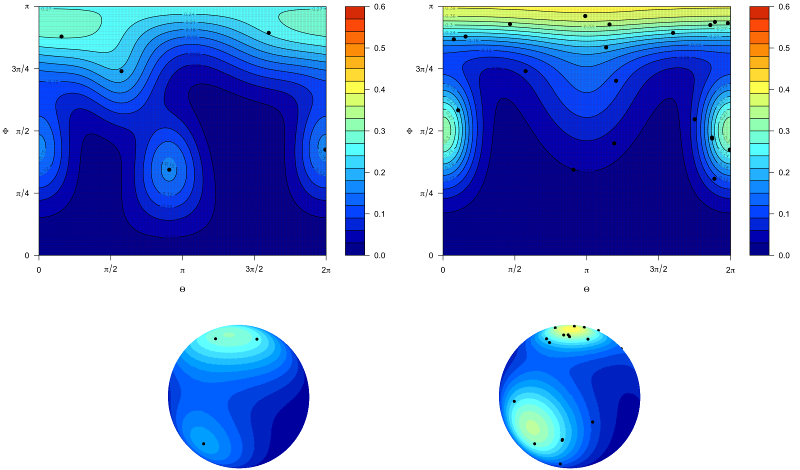


Figure: Left: KDE with $n = 5$. Right: true density.

KDE construction: spherical case

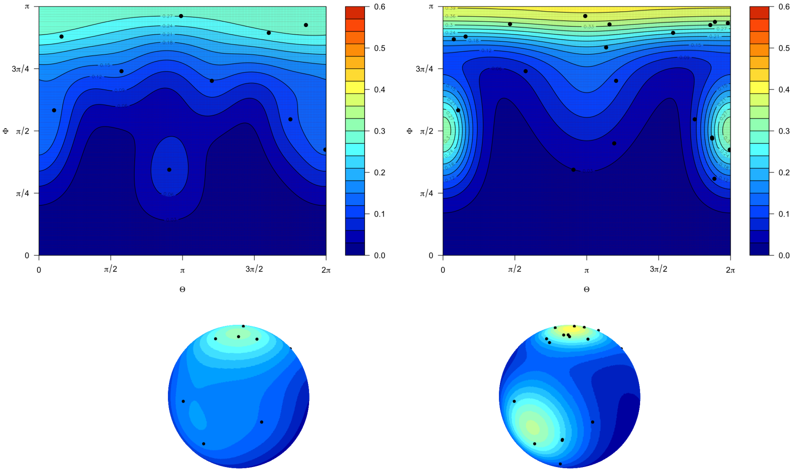


Figure: Left: KDE with $n = 10$. Right: true density.

KDE construction: spherical case

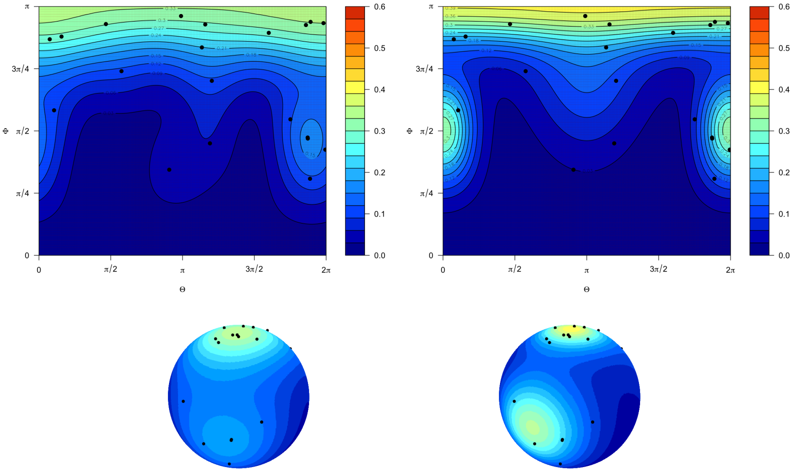


Figure: Left: KDE with $n = 20$. Right: true density.

Outline

Density estimation under rotasymmetry

- Suppose that $\mathbf{X}$ is rotasymmetric with density f.
- **Goal: estimate semiparametrically f under rotasymmetry**.
- Estimation approaches, sorted from weaker to stronger assumptions:
  1. Nonparametrically: KDE for directional data.
  2. **Semiparametrically, θ unknown**.
  3. **Semiparametrically, θ known**.
  4. Parametrically: assuming a parametric family.
- Related references in the Euclidean setting:

Stute, W. and Werner, U. (1991). Nonparametric estimation of elliptically contoured densities. *In G. Roussas (Ed.), Nonparametric Functional Estimation and Related Topics*, 173–190.

Liebscher, E. (2005). A semiparametric density estimator based on elliptical distributions. *J. Multivariate Anal.*, 92:205–225.

- The first step is to build an operator that ensures rotasymmetry.

The rotasymmetrizer

Definition (Rotasymmetrizer)

The ***rotasymmetrizer*** *around* $\boldsymbol{\theta}$, $R_{\boldsymbol{\theta}}$, *transforms a function* $f : \Omega_q \longrightarrow \mathbb{R}$ *into*

$$R_{\boldsymbol{\theta}} f(\mathbf{x}) = \frac{1}{\omega_{q-1}} \int_{\Omega_{q-1}} f\left(\mathbf{x}_{\boldsymbol{\theta},\boldsymbol{\xi}}\right) \omega_{q-1}(d\boldsymbol{\xi}),$$

with $\mathbf{x}_{\boldsymbol{\theta},\boldsymbol{\xi}} = (\mathbf{x}^T\boldsymbol{\theta})\boldsymbol{\theta} + (1 - (\mathbf{x}^T\boldsymbol{\theta})^2)^{\frac{1}{2}} \mathbf{B}_{\boldsymbol{\theta}}\boldsymbol{\xi}$.

- For point $\mathbf{x} \in \Omega_q$, the operator **averages out** the density along the points sharing the same **colatitude** (wrt $\boldsymbol{\theta}$).
- Intuitively: parallel redistribution of probability mass.

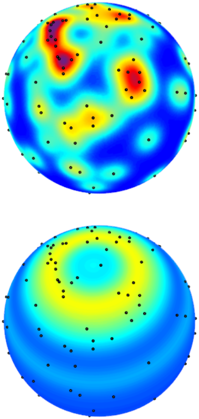


Figure: Input and output of $R_{\boldsymbol{\theta}}$ with $\boldsymbol{\theta} = (0, 0, 1)$.

Properties

Proposition (Rotasymmetrizer properties)

Let be $f, f_1, f_2 : \Omega_q \longrightarrow \mathbb{R}^+$ directional densities and $\boldsymbol{\theta} \in \Omega_q$.

1. ***Invariance*** *from different matrices $\mathbf{B}_{\boldsymbol{\theta}}$:*

$$\int_{\Omega_{q-1}} f\left(\mathbf{x}_{\boldsymbol{\theta},\boldsymbol{\xi},1}\right) \omega_{q-1}(d\boldsymbol{\xi}) = \int_{\Omega_{q-1}} f\left(\mathbf{x}_{\boldsymbol{\theta},\boldsymbol{\xi},2}\right) \omega_{q-1}(d\boldsymbol{\xi}),$$

with $\mathbf{x}_{\boldsymbol{\theta},\boldsymbol{\xi},k} = (\mathbf{x}^T\boldsymbol{\theta})\boldsymbol{\theta} + (1-(\mathbf{x}^T\boldsymbol{\theta})^2)^{\frac{1}{2}}\mathbf{B}_{\boldsymbol{\theta},k}\boldsymbol{\xi},\ k = 1, 2.$

2. ***Linearity***: $R_{\boldsymbol{\theta}}(\lambda_1 f_1 + \lambda_2 f_2)(\mathbf{x}) = \lambda_1 R_{\boldsymbol{\theta}} f_1(\mathbf{x}) + \lambda_2 R_{\boldsymbol{\theta}} f_2(\mathbf{x})$.
3. ***Density preservation***: $R_{\boldsymbol{\theta}} f$ *is a density.*
4. ***Rotasymmetry characterization***:

$$R_{\boldsymbol{\theta}} f = f \iff f \text{ is rotasymmetric around } \boldsymbol{\theta}.$$

5. *Particular expression for the* ***vMF density***:

$$R_{\boldsymbol{\theta}} f_{\mathrm{vMF}}(\mathbf{x}; \boldsymbol{\mu}, \kappa) = \frac{C_q(\kappa)\exp\left\{\kappa \mathbf{x}^T\boldsymbol{\theta}\boldsymbol{\theta}^T\boldsymbol{\mu}\right\}}{\omega_{q-1} C_{q-1}\left(\kappa\left[(1-(\mathbf{x}^T\boldsymbol{\theta})^2)(1-(\boldsymbol{\mu}^T\boldsymbol{\theta})^2)\right]^{\frac{1}{2}}\right)}.$$

Rotasymmetric KDE

Definition (Rotasymmetric KDE)

The ***rotasymmetric KDE (RKDE)*** *is the application of the rotasymmetrizer to the usual KDE:*

$$\hat{f}_{h,\boldsymbol{\theta}}(\mathbf{x}) = R_{\boldsymbol{\theta}} \hat{f}_h(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} L_{h,\boldsymbol{\theta}}\left(\mathbf{x}, \mathbf{X}_i\right),$$

with $L_{h,\boldsymbol{\theta}}\left(\mathbf{x}, \mathbf{X}_i\right) = \frac{c_{h,q}(L)}{\omega_{q-1}} \int_{\Omega_{q-1}} L\left(\frac{1 - \mathbf{x}_{\boldsymbol{\theta},\boldsymbol{\xi}}^T \mathbf{X}_i}{h^2}\right) \omega_{q-1}(d\boldsymbol{\xi}).$

- The rotasymmetric von Mises kernel has a closed expression:

$$L_{h,\boldsymbol{\theta}}(\mathbf{x}, \mathbf{X}_i) = \frac{C_q(1/h^2) \exp\left\{\mathbf{x}^T \boldsymbol{\theta}\boldsymbol{\theta}^T \mathbf{X}_i / h^2\right\}}{\omega_{q-1} C_{q-1}\left(\left[(1 - (\mathbf{x}^T\boldsymbol{\theta})^2)(1 - (\mathbf{X}_i^T\boldsymbol{\theta})^2)\right]^{\frac{1}{2}} / h^2\right)}.$$

- The order of the normalizing constant of the kernel is $\mathcal{O}\left(h^{-1}\right)$.

Comparison of kernels

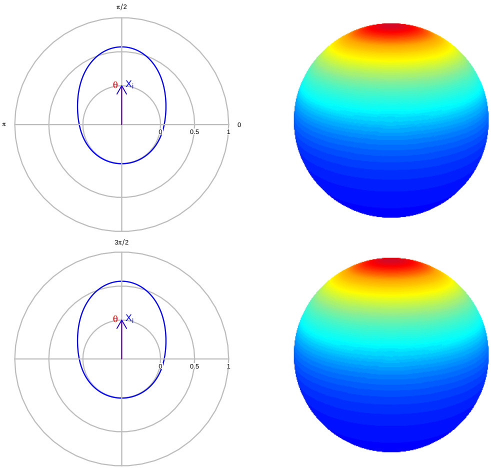


Figure: Kernels for the KDE (upper row) and their RKDE counterparts (lower), with $\boldsymbol{\theta} = (\mathbf{0}_q, 1)$. The kernels have the same bandwidth.

Comparison of kernels

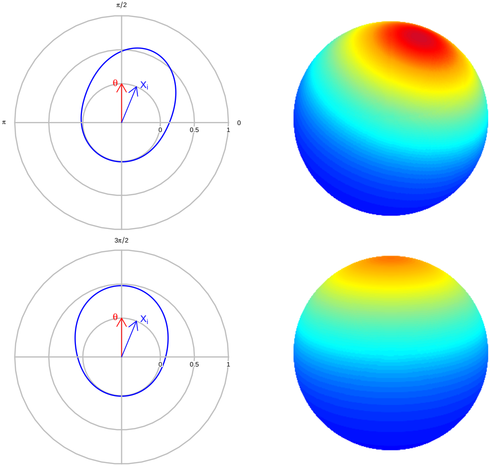


Figure: Kernels for the KDE (upper row) and their RKDE counterparts (lower), with $\boldsymbol{\theta} = (\mathbf{0}_q, 1)$. The kernels have the same bandwidth.

Comparison of kernels

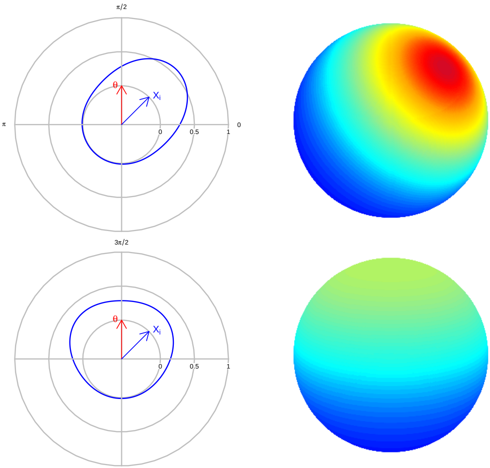


Figure: Kernels for the KDE (upper row) and their RKDE counterparts (lower), with $\boldsymbol{\theta} = (\mathbf{0}_q, 1)$. The kernels have the same bandwidth.

Comparison of kernels

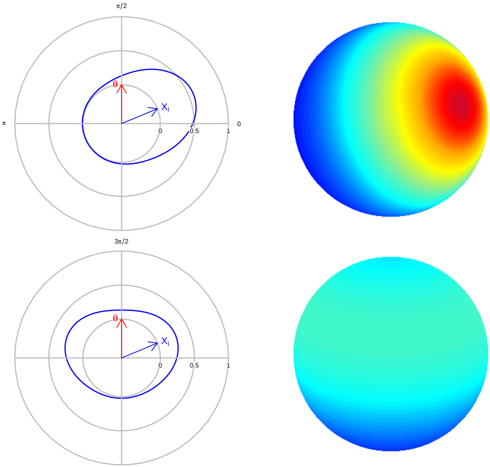


Figure: Kernels for the KDE (upper row) and their RKDE counterparts (lower), with $\boldsymbol{\theta} = (\mathbf{0}_q, 1)$. The kernels have the same bandwidth.

Comparison of kernels

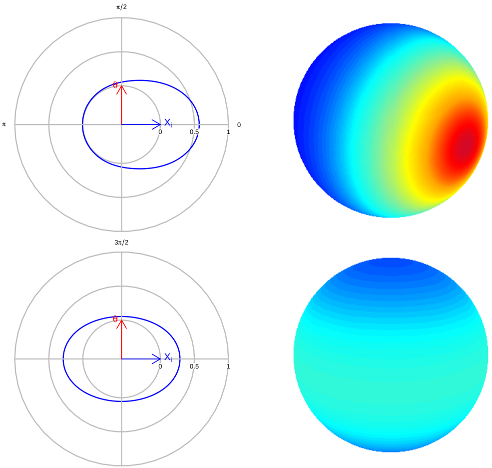


Figure: Kernels for the KDE (upper row) and their RKDE counterparts (lower), with $\boldsymbol{\theta} = (\mathbf{0}_q, 1)$. The kernels have the same bandwidth.

Comparison of kernels

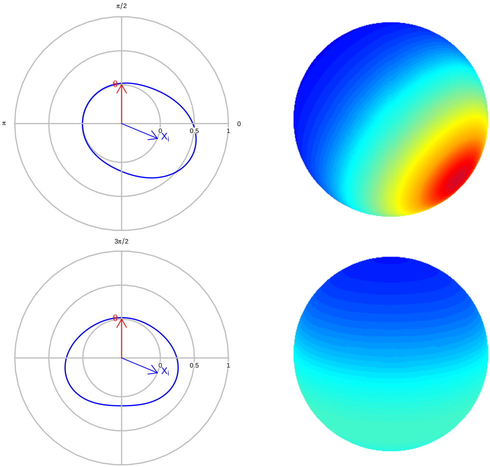


Figure: Kernels for the KDE (upper row) and their RKDE counterparts (lower), with $\boldsymbol{\theta} = (\mathbf{0}_q, 1)$. The kernels have the same bandwidth.

Comparison of kernels

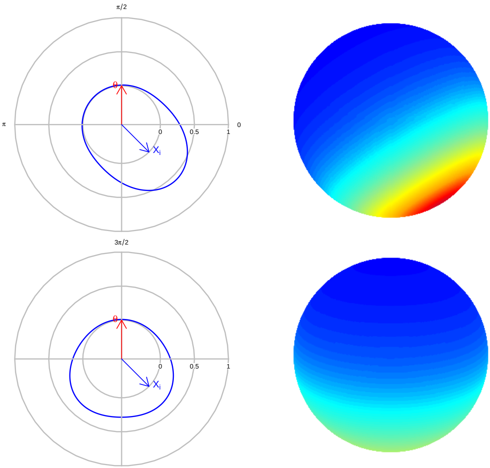


Figure: Kernels for the KDE (upper row) and their RKDE counterparts (lower), with $\boldsymbol{\theta} = (\mathbf{0}_q, 1)$. The kernels have the same bandwidth.

Comparison of kernels

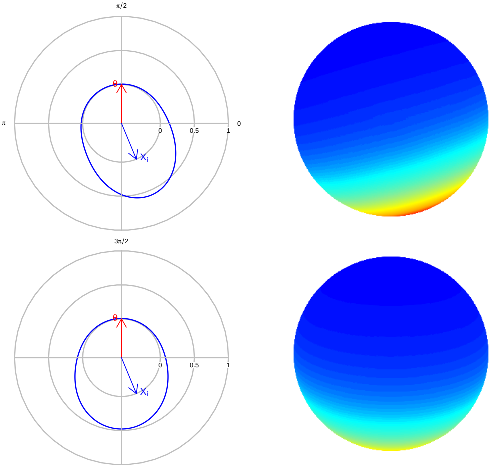


Figure: Kernels for the KDE (upper row) and their RKDE counterparts (lower), with $\boldsymbol{\theta} = (\mathbf{0}_q, 1)$. The kernels have the same bandwidth.

Comparison of kernels

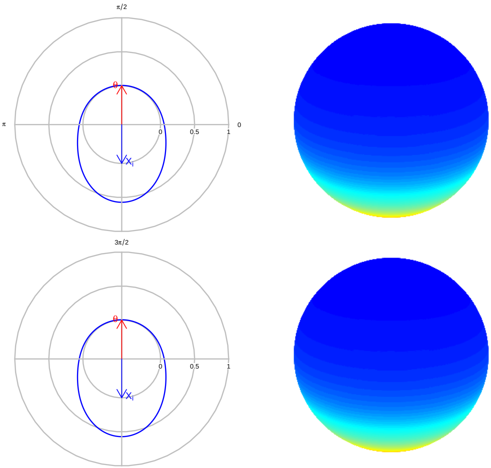


Figure: Kernels for the KDE (upper row) and their RKDE counterparts (lower), with $\boldsymbol{\theta} = (\mathbf{0}_q, 1)$. The kernels have the same bandwidth.

Connections with KDE in $[-1, 1]$

- The kernels of the RKDE **only** depend on the **projected sample** $T_i = \mathbf{X}_i^T\boldsymbol{\theta}$, $i = 1, \ldots, n$, and the **projected point** $t = \mathbf{x}^T\boldsymbol{\theta}$.
- RKDE is equivalent to KDE on the projected sample in $[-1, 1]$ with bounded kernels adapted to capture the possible spikes of f^*.
- Boundary bias is $\mathcal{O}\left(h^2\right)$ without any corrections.

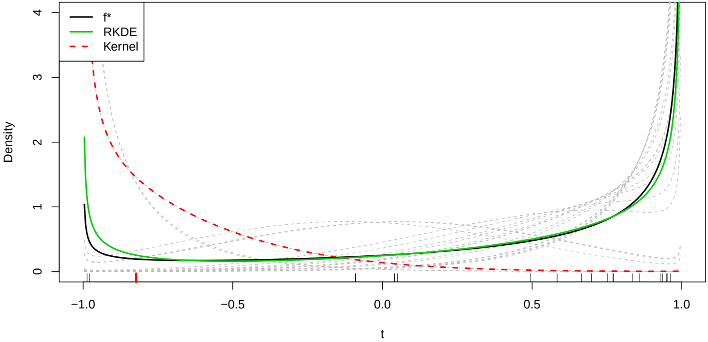


Figure: KDE of f^* with $g(t) = C_q(\kappa)\exp\{\kappa t\}$, $\kappa = 1$ and $q = 1$.

Connections with KDE in $[-1, 1]$

- The kernels of the RKDE **only** depend on the **projected sample** $T_i = \mathbf{X}_i^T \boldsymbol{\theta}$, $i = 1, \ldots, n$, and the **projected point** $t = \mathbf{x}^T \boldsymbol{\theta}$.
- RKDE is equivalent to KDE on the projected sample in $[-1, 1]$ with bounded kernels adapted to capture the possible spikes of f^*.
- Boundary bias is $\mathcal{O}\left(h^2\right)$ without any corrections.

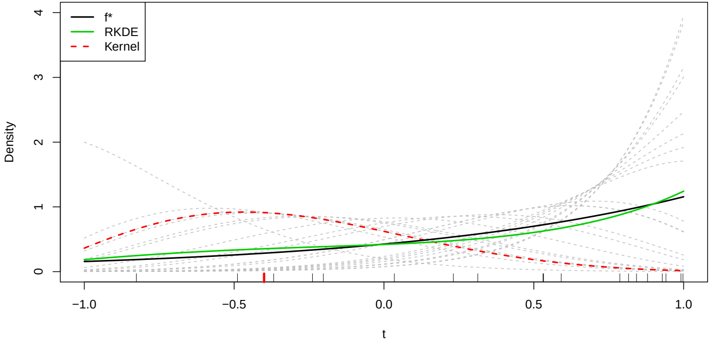


Figure: KDE of f^* with $g(t) = C_q(\kappa) \exp\{\kappa t\}$, $\kappa = 1$ and $q = 2$.

Connections with KDE in $[-1, 1]$

- The kernels of the RKDE **only** depend on the **projected sample** $T_i = \mathbf{X}_i^T \boldsymbol{\theta}$, $i = 1, \ldots, n$, and the **projected point** $t = \mathbf{x}^T \boldsymbol{\theta}$.
- RKDE is equivalent to KDE on the projected sample in $[-1, 1]$ with bounded kernels adapted to capture the possible spikes of f^*.
- Boundary bias is $\mathcal{O}\left(h^2\right)$ without any corrections.

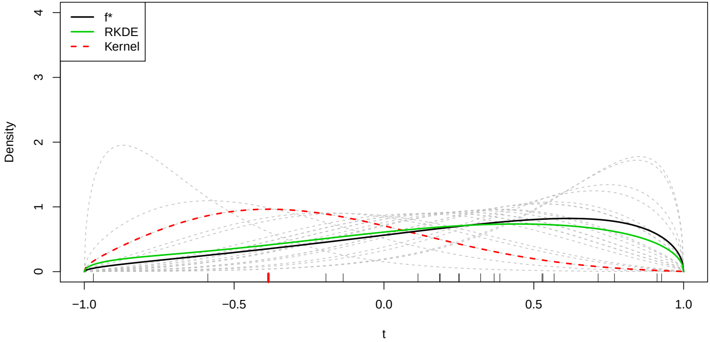


Figure: KDE of f^* with $g(t) = C_q(\kappa) \exp\{\kappa t\}$, $\kappa = 1$ and $q = 3$.

Connections with KDE in $[-1, 1]$

- The kernels of the RKDE **only** depend on the **projected sample** $T_i = \mathbf{X}_i^T \boldsymbol{\theta}$, $i = 1, \ldots, n$, and the **projected point** $t = \mathbf{x}^T \boldsymbol{\theta}$.
- RKDE is equivalent to KDE on the projected sample in $[-1, 1]$ with bounded kernels adapted to capture the possible spikes of f^*.
- Boundary bias is $\mathcal{O}\left(h^2\right)$ without any corrections.

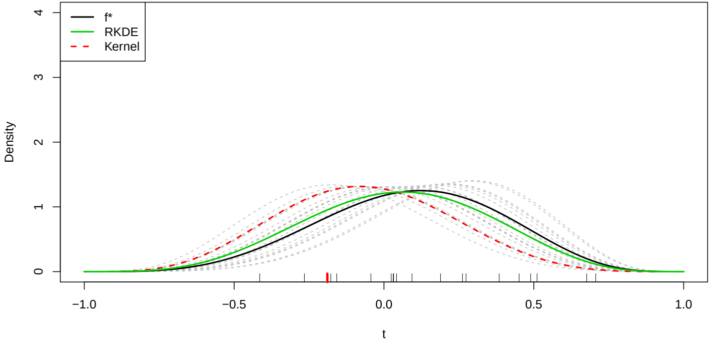


Figure: KDE of f^* with $g(t) = C_q(\kappa) \exp\{\kappa t\}$, $\kappa = 1$ and $q = 10$.

Connections with KDE in $[-1,1]$

- The kernels of the RKDE **only** depend on the **projected sample** $T_i = \mathbf{X}_i^T\boldsymbol{\theta}$, $i = 1,\ldots,n$, and the **projected point** $t = \mathbf{x}^T\boldsymbol{\theta}$.
- RKDE is equivalent to KDE on the projected sample in $[-1,1]$ with bounded kernels adapted to capture the possible spikes of f^*.
- Boundary bias is $\mathcal{O}\left(h^2\right)$ without any corrections.

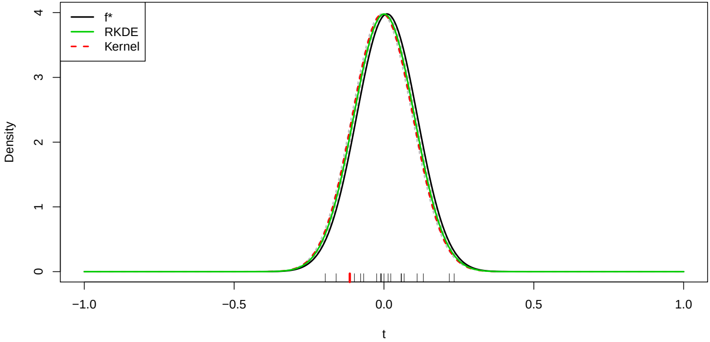


Figure: KDE of f^* with $g(t) = C_q(\kappa)\exp\{\kappa t\}$, $\kappa = 1$ and $q = 100$.

Bias of the RKDE

- **Assumptions:**

  A1 f is extended from Ω_q to $\mathbb{R}^{q+1} \setminus \{\mathbf{0}\}$ by $f(\mathbf{x}) \equiv f(\mathbf{x}/\|\mathbf{x}\|)$. f is twice continuously differentiable with Hessian $\boldsymbol{\mathcal{H}} f(\mathbf{x})$.

  A2 L is a continuous and bounded function $L : [0,\infty) \to [0,\infty)$ with exponential decay: $L(r) \leq Me^{-\alpha r}$, $M, \alpha > 0$.

  A3 The sequence $h = h_n$ satisfies $h \to 0$ and $nh \to \infty$.

  A4 The sequence $h = h_n$ satisfies $h \to 0$ and $nh^q \to \infty$.

- **A4** is required for consistency at $\mathbf{x} = \pm\boldsymbol{\theta}$. Of course, **A4** $\implies$ **A3**.

Proposition (Bias, $\boldsymbol{\theta}$ known)

*Under **A1–A3** and uniformly in $\mathbf{x} \in \Omega_q$,*

$$\mathbb{E}\left[\hat{f}_{h,\boldsymbol{\theta}}(\mathbf{x})\right] = R_{\boldsymbol{\theta}} f(\mathbf{x}) + \frac{b_q(L)}{q} \mathrm{tr}\left[\boldsymbol{\mathcal{H}} R_{\boldsymbol{\theta}} f(\mathbf{x})\right] h^2 + o\left(h^2\right).$$

If rotasymmetry holds, then $R_{\boldsymbol{\theta}} f = f$ and the bias is KDE's one.

Variance

Proposition (Variance, $\boldsymbol{\theta}$ known)

*Under **A1–A2**, **A3** if $(\mathbf{x}^T\boldsymbol{\theta})^2 < 1$ and **A4** otherwise,*

$$\mathbb{V}\mathrm{ar}\left[\hat{f}_{h,\boldsymbol{\theta}}(\mathbf{x})\right] = C_{\mathbf{x}^T\boldsymbol{\theta},q,L}(h)\frac{R_{\boldsymbol{\theta}}f(\mathbf{x})}{n}(1+o(1)) - \frac{(R_{\boldsymbol{\theta}}f(\mathbf{x}))^2}{n}$$

uniformly in $\mathbf{x} \in \Omega_q$, where

$$C_{\mathbf{x}^T\boldsymbol{\theta},q,L}(h) = \begin{cases} \dfrac{\lambda_q(L^2)\lambda_q(L)^{-2}}{h^q}, & (\mathbf{x}^T\boldsymbol{\theta})^2 = 1,\ q \geq 1, \\ \dfrac{\lambda_1(L^2)\lambda_1(L)^{-2}}{2h}, & (\mathbf{x}^T\boldsymbol{\theta})^2 < 1,\ q = 1, \\ \dfrac{\lambda_{q-1}(L)^2\lambda_q(L)^{-2}}{\omega_{q-1}\left(1-(\mathbf{x}^T\boldsymbol{\theta})^2\right)^{\frac{1}{2}} h}, & (\mathbf{x}^T\boldsymbol{\theta})^2 < 1,\ q \geq 2. \end{cases}$$

- The variance increases when $q \to \infty$ since $\omega_{q-1} \to 0$!

Spherical area surface

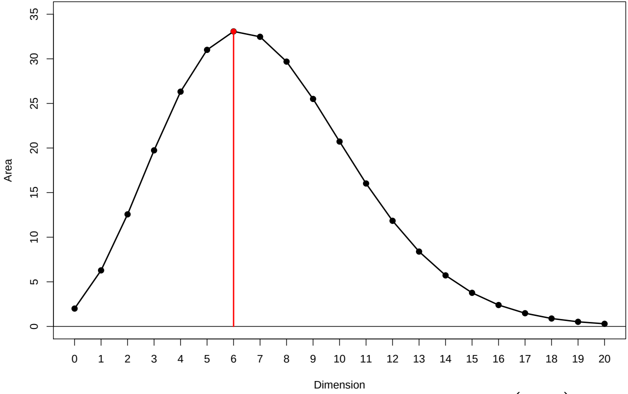


Figure: Spherical surface $\omega_q = 2\pi^{\frac{q+1}{2}} / \Gamma\left(\frac{q+1}{2}\right)$.

- The area of Ω_q tends to zero, but not monotonically.
- *Weird* maximum at dimension $q = 6$.
- $[-1,1]^q$ touches Ω_q in 2^q points, yet its area tends to infinity.

Asymptotic normality

Corollary (Pointwise asymptotic normality, $\boldsymbol{\theta}$ known)

Under ***A1–A2***, ***A3*** *if* $(\mathbf{x}^T\boldsymbol{\theta})^2 < 1$ *and* ***A4*** *otherwise,*

$$a_n\left(\hat{f}_{h,\boldsymbol{\theta}}(\mathbf{x}) - f(\mathbf{x})\right) \xrightarrow{d} \mathcal{N}\left(R_{\boldsymbol{\theta}}f(\mathbf{x}) - f(\mathbf{x}), C_{\mathbf{x}^T\boldsymbol{\theta},q,L}(1)\right),$$

where $a_n = \sqrt{nh}$ *if* $(\mathbf{x}^T\boldsymbol{\theta})^2 < 1$ *and* $a_n = \sqrt{nh^q}$ *otherwise.*

Concept	KDE (✓/× rotasym.)	RKDE (✓ rotasym.)	RKDE (× rotasym.)
Bias	$\mathcal{O}\left(h^2\right)$	$\mathcal{O}\left(h^2\right)$	$\mathcal{O}\left(R_{\boldsymbol{\theta}}f(\mathbf{x}) - f(\mathbf{x})\right)$
Variance	$\mathcal{O}\left((nh^q)^{-1}\right)$	$\mathcal{O}\left((nh)^{-1}\right)$	$\mathcal{O}\left((nh)^{-1}\right)$
Optimal AMISE	$\mathcal{O}\left(n^{-\frac{4}{4+q}}\right)$	$\mathcal{O}\left(n^{-\frac{4}{5}}\right)$	$\mathcal{O}\left(\int (R_{\boldsymbol{\theta}}f - f)^2\right)$

Table: Summary of the KDE and RKDE key orders.

Properties with unknown θ

- **Assumption:**

  A5 $\hat{\boldsymbol{\theta}}$ is a $\sqrt{n}$-consistent estimator: $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \mathcal{O}_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$.

- Examples:
  - If g is strictly monotone, $\hat{\boldsymbol{\theta}} = \frac{\sum_{i=1}^n \mathbf{X}_i}{\left|\left|\sum_{i=1}^n \mathbf{X}_i\right|\right|}$ satisfies **A5**.
  - If g is symmetric wrt 0 and strictly monotone in $[0,1]$, then the first eigenvector of $\frac{1}{n}\sum_{i=1}^n \mathbf{X}_i\mathbf{X}_i^T$ satisfies **A5**.

- Bias and variance: under **A1–A2**, **A3** or **A4** and **A5**,

$$\mathbb{E}\left[\hat{f}_{h,\hat{\boldsymbol{\theta}}}(\mathbf{x})\right] = R_{\boldsymbol{\theta}} f(\mathbf{x}) + \frac{b_q(L)}{q}\mathrm{tr}\left[\boldsymbol{\mathcal{H}} R_{\boldsymbol{\theta}} f(\mathbf{x})\right] h^2 + o\left(h^2\right) + \mathcal{O}\left(n^{-\frac{1}{2}}\right),$$

$$\mathbb{V}\mathrm{ar}\left[\hat{f}_{h,\hat{\boldsymbol{\theta}}}(\mathbf{x})\right] = C_{\mathbf{x}^T\boldsymbol{\theta},q,L}(h)\frac{R_{\boldsymbol{\theta}} f(\mathbf{x})}{n}(1 + o(1)) - \frac{(R_{\boldsymbol{\theta}} f(\mathbf{x}))^2}{n}.$$

- Asymptotic normality: under **A1–A5**,

$$a_n\left(\hat{f}_{h,\hat{\boldsymbol{\theta}}}(\mathbf{x}) - f(\mathbf{x})\right) \xrightarrow{d} \mathcal{N}\left(R_{\boldsymbol{\theta}} f(\mathbf{x}) - f(\mathbf{x}), C_{\mathbf{x}^T\boldsymbol{\theta},q,L}(1)\right).$$

Outline

1. **Introduction**
   - Rotasymmetry
   - KDE with directional data
2. **Density estimation under rotasymmetry**
   - The rotasymmetrizer
   - Rotasymmetric KDE
3. **Simulation study**

Simulation study

Goal:

- Compare for a grid of bandwidths h the performance of the estimators.

Settings:

- Estimators: KDE, RKDE with $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$ (directional mean). All with von Mises kernel.
- Error measurement: log of the Mean Integrated Squared Error (MISE):

$$\log \mathrm{MISE} = \log \mathbb{E}\left[\int_{\Omega_q} \left(\hat{f}(\mathbf{x}) - f(\mathbf{x})\right)^2 \omega_q(d\mathbf{x})\right].$$

- Target density: $\mathrm{vMF}((\mathbf{0}_q, 1), 5)$. Dimensions: $q = 1, 2, 3, 4, 5, 6$.
- Sample size: $n = 100$. Monte Carlo replicates: $M = 1000$.

Comparison with KDE

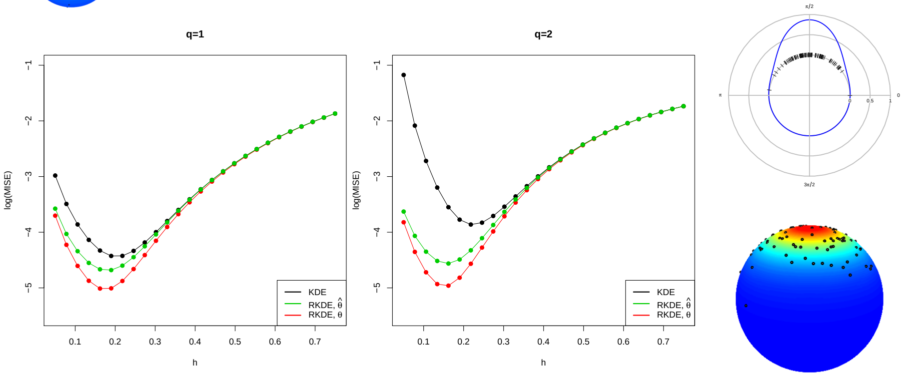


Figure: Performance of the three kernel estimators with $q = 1$ (left) and $q = 2$ (right), with $n = 100$.

Ratios optimal MISEs	$q = 1$	$q = 2$	$q = 3$	$q = 4$	$q = 5$	$q = 6$
KDE/RKDE, $\boldsymbol{\theta}$	1.796	2.999	4.065	5.643	5.871	8.019
KDE/RKDE, $\hat{\boldsymbol{\theta}}$	1.289	2.014	2.537	3.035	3.207	3.467

Comparison with KDE

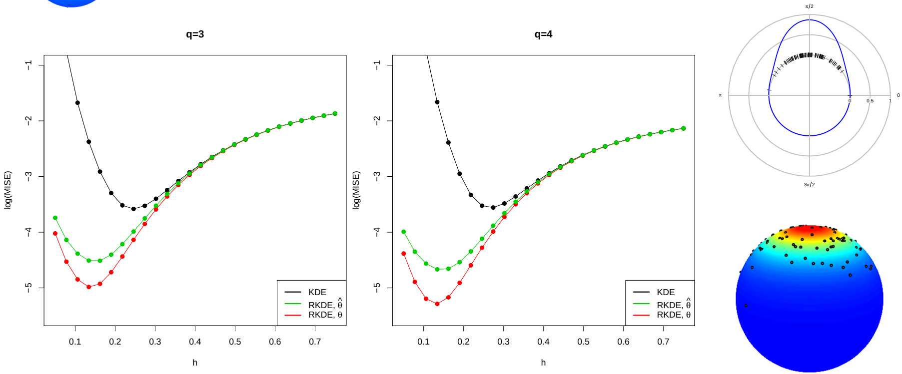


Figure: Performance of the three kernel estimators with $q = 3$ (left) and $q = 4$ (right), with $n = 100$.

Ratios optimal MISEs	$q = 1$	$q = 2$	$q = 3$	$q = 4$	$q = 5$	$q = 6$
KDE/RKDE, $\boldsymbol{\theta}$	1.796	2.999	4.065	5.643	5.871	8.019
KDE/RKDE, $\hat{\boldsymbol{\theta}}$	1.289	2.014	2.537	3.035	3.207	3.467

Comparison with KDE

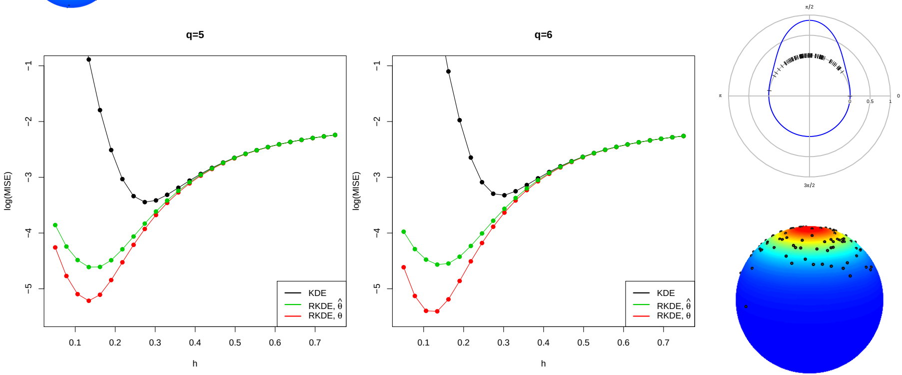


Figure: Performance of the three kernel estimators with $q = 5$ (left) and $q = 6$ (right), with $n = 100$.

Ratios optimal MISEs	$q = 1$	$q = 2$	$q = 3$	$q = 4$	$q = 5$	$q = 6$
KDE/RKDE, $\boldsymbol{\theta}$	1.796	2.999	4.065	5.643	5.871	8.019
KDE/RKDE, $\hat{\boldsymbol{\theta}}$	1.289	2.014	2.537	3.035	3.207	3.467

Wrapping up

We have seen that...

1. The **rotasymmetrizer** enforces rotasymmetry naturally.
2. The RKDE has the **same bias** as the KDE but **lower variance**.
3. The **variance still depends on** q and increases if $q \to \infty$.
4. **Improvements** on the **MISE** are notable in practise.

Appendix

Empirical variance

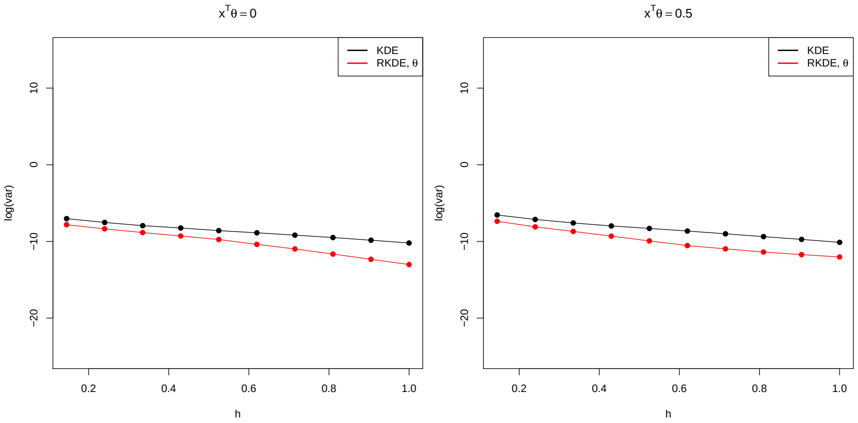


Figure: Log-variances of the KDE and RKDE at $\mathbf{x}$ such that $\mathbf{x}^T\boldsymbol{\theta} = 0$ (left) and $\mathbf{x}^T\boldsymbol{\theta} = 0.5$ (right), for a grid of bandwidths h and dimension $q = 1$.

Empirical variance

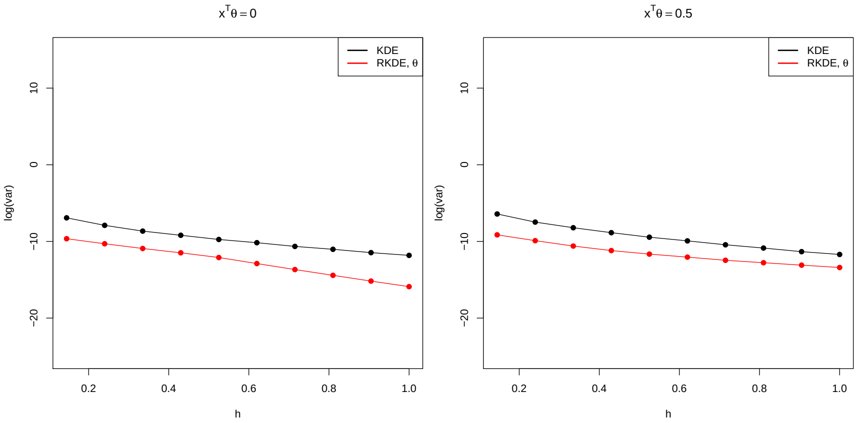


Figure: Log-variances of the KDE and RKDE at $\mathbf{x}$ such that $\mathbf{x}^T\boldsymbol{\theta} = 0$ (left) and $\mathbf{x}^T\boldsymbol{\theta} = 0.5$ (right), for a grid of bandwidths h and dimension $q = 2$.

Empirical variance

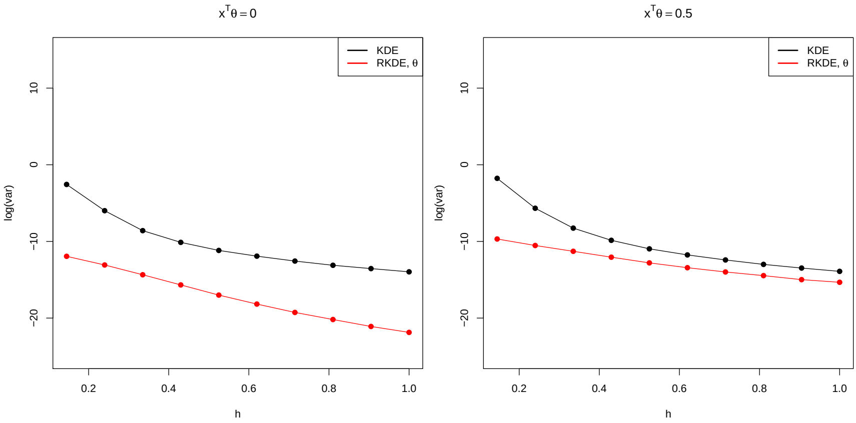


Figure: Log-variances of the KDE and RKDE at $\mathbf{x}$ such that $\mathbf{x}^T\boldsymbol{\theta} = 0$ (left) and $\mathbf{x}^T\boldsymbol{\theta} = 0.5$ (right), for a grid of bandwidths h and dimension $q = 10$.

Empirical variance

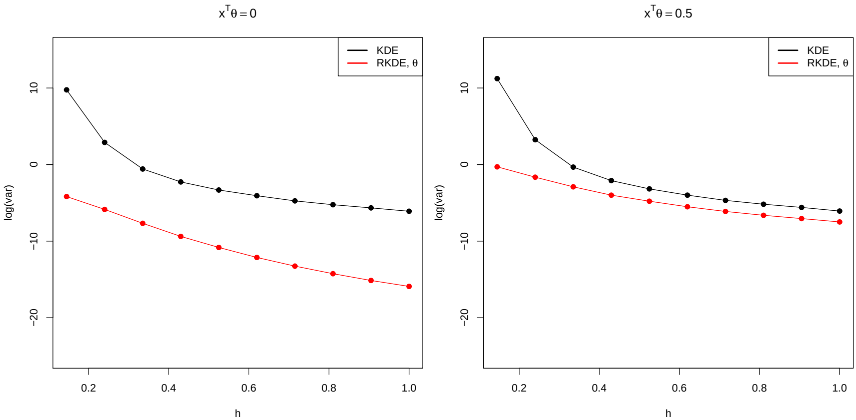


Figure: Log-variances of the KDE and RKDE at $\mathbf{x}$ such that $\mathbf{x}^T\boldsymbol{\theta} = 0$ (left) and $\mathbf{x}^T\boldsymbol{\theta} = 0.5$ (right), for a grid of bandwidths h and dimension $q = 20$.

Bandwidth selection

- Plug-in rules are possible, but more complex than in the usual KDE.
- Cross-validation rules apply as expected:

$$h_{\mathrm{CV}} = \arg\min_{h>0} \left[\frac{2}{n} \sum_{i=1}^{n} \hat{f}_{h,\boldsymbol{\theta}}^{-i}(\mathbf{X}_i) - \int_{\Omega_q} \hat{f}_{h,\boldsymbol{\theta}}(\mathbf{x})^2 \, \omega_q(d\mathbf{x}) \right],$$

$$h_{\mathrm{LCV}} = \arg\max_{h>0} \sum_{i=1}^{n} \log \hat{f}_{h,\boldsymbol{\theta}}^{-i}(\mathbf{X}_i).$$

- The integral is one dimensional:

$$\int_{\Omega_q} \hat{f}_{h,\boldsymbol{\theta}}(\mathbf{x})^2 \, \omega_q(d\mathbf{x}) = \omega_{q-1} \int_{-1}^{1} \left(\frac{1}{n} \sum_{i=1}^{n} L_h^*(t, T_i) \right)^2 (1-t^2)^{\frac{q}{2}-1} \, dt.$$

- If $\boldsymbol{\theta}$ is unknown, then we can opt for:
  1. Plug-in a consistent estimate $\hat{\boldsymbol{\theta}}$ (sample mean if g is monotonic).
  2. Joint optimization of the LCV loss, for example with

$$(\boldsymbol{\theta}, h)_{\mathrm{LCV}} = \arg\max_{\substack{h>0 \\ \boldsymbol{\theta}\in\Omega_q}} \sum_{i=1}^{n} \log \hat{f}_{h,\boldsymbol{\theta}}^{-i}(\mathbf{X}_i).$$

LCV bandwidth comparison

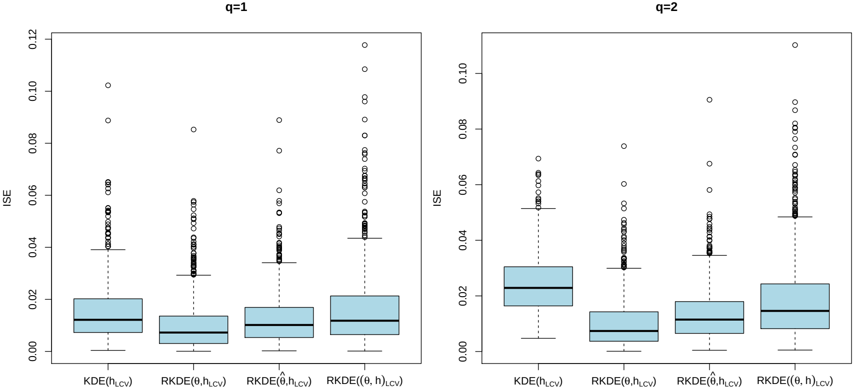
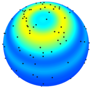


Figure: Distribution of the ISEs for the estimators KDE(h_{LCV}), RKDE($\boldsymbol{\theta}, h_{\mathrm{LCV}}$), RKDE($\hat{\boldsymbol{\theta}}, h_{\mathrm{LCV}}$) and RKDE($(\boldsymbol{\theta}, h)_{\mathrm{LCV}}$).



Hints on the estimation of θ

- The estimators for $\boldsymbol{\theta}$ are based on the eigenvector of the outer matrix that has multiplicity one:

$$\mathbb{E}\left[\mathbf{X}\mathbf{X}^T\right] = \left\{\int_{-1}^{1} t^2 f^*(t)\,dt \times \boldsymbol{\theta}\boldsymbol{\theta}^T + \frac{1}{q}\left(1 - \int_{-1}^{1} t^2 f^*(t)\,dt\right) \times (\mathbf{I}_{q+1} - \boldsymbol{\theta}\boldsymbol{\theta}^T)\right\}.$$

- Problems if all the eigenvalues are similar!
- **New estimator** based on the characterization

$$\{\mathbf{X}_i\}_{i=1}^n \text{ is rotasymmetric } \iff \left\{\frac{\mathbf{B}_{\boldsymbol{\theta}}^T\mathbf{X}_i}{\sqrt{1-(\mathbf{X}_i^T\boldsymbol{\theta})^2}}\right\}_{i=1}^n \text{ is uniform in } \Omega_q.$$

- The estimator minimizes discrepancy wrt uniformity, measured by an statistic T_n (consistent against all alternatives!).

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\Omega_q} T_n\left(\frac{\mathbf{B}_{\boldsymbol{\theta}}^T\mathbf{X}_1}{\sqrt{1-(\mathbf{X}_1^T\boldsymbol{\theta})^2}}, \ldots, \frac{\mathbf{B}_{\boldsymbol{\theta}}^T\mathbf{X}_n}{\sqrt{1-(\mathbf{X}_n^T\boldsymbol{\theta})^2}}\right),$$

- For example, Ajne's statistic:

$$T_n(\mathbf{Y}_1, \ldots, \mathbf{Y}_n) = \frac{n}{4} - \frac{1}{n\pi}\sum_{i<j}\cos^{-1}(\mathbf{Y}_i^T\mathbf{Y}_j).$$

Applications in testing

- The RKDE can be employed for nonparametric testing:
  1. Test **for rotational symmetry** comparing KDE and RKDE:

$$T_n = \int_{\Omega_q} \left(\hat{f}_{h,\hat{\boldsymbol{\theta}}}(\mathbf{x}) - \hat{f}_h(\mathbf{x})\right)^2 \omega_q(d\mathbf{x}).$$

  2. Goodness-of-fit test for parametric models **under rotasymmetry**, *i.e.* testing of $H_0 : f \in \mathcal{F}_\Lambda = \{f_{\boldsymbol{\lambda}} : \boldsymbol{\lambda} \in \Lambda\}$:

$$R_n = \int_{\Omega_q} \left(\hat{f}_{h,\hat{\boldsymbol{\theta}}}(\mathbf{x}) - L_h f_{\hat{\boldsymbol{\lambda}}}(\mathbf{x})\right)^2 \omega_q(d\mathbf{x}).$$

  Expected to be more powerful (under rotasymmetry) than:

Boente, G., Rodríguez, D. and González-Manteiga, W. (2014). Goodness-of-fit test for directional data. *Scand. J. Stat.*, 41:259–275.

- Resampling strategy: using the tangent-normal decomposition.