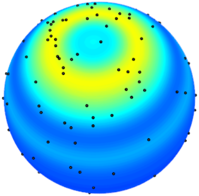


Smoothing-based inference with directional data

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

Department of Mathematical Sciences
University of Copenhagen
↓
Department of Statistics
Universidad Carlos III de Madrid

Wencesfest, June 8, 2016

Directional data: what, why, where?

- 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 bioinformatics

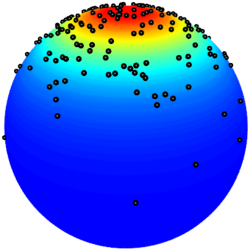


Figure: Spherical von Mises density

Directional data: what, why, where?

- 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 bioinformatics

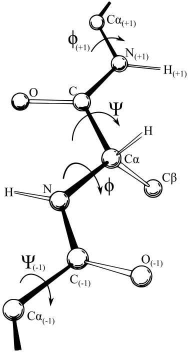


Figure: Schematic view of the protein's backbone

Von Mises-Fisher distribution

- The **von Mises-Fisher** (vMF) is the most well known directional 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:

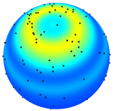
$$\omega_q \equiv \omega_q(\Omega_q) = 2\pi^{\frac{q+1}{2}}/\Gamma\left(\frac{q+1}{2}\right)$$

- (Isotropic) **Gaussian analogue**:
  1. Same MLE characterization property
  2. If $\mathbf{X}\sim\mathcal{N}_{q+1}\left(\boldsymbol{\mu},\sigma^2\mathbf{I}_{q+1}\right)$, then

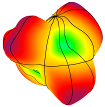
$$\mathbf{X}\,\big|\,||\mathbf{X}||=1 \sim \mathrm{vMF}\left(\frac{\boldsymbol{\mu}}{||\boldsymbol{\mu}||},\frac{||\boldsymbol{\mu}||}{\sigma^2}\right)$$

Contents of the talk

1. **Part I**. Kernel **density** estimation with directional data under rotational symmetry
   - Present a KDE under **rotational symmetry**
   - Study its main **asymptotic properties**
   - Illustrate **empirical performance** through simulations



2. **Part II**. Estimation and testing in linear-directional **regression**
   - Present a **local linear estimator** with directional predictor
   - Build a **goodness-of-fit** test for regression models
   - Apply both to test a common assumption in **bioinformatics**



Part I

Kernel density estimation with directional data under rotational symmetry

García-Portugués, E., Ley, C., Verdebout, T. (2016). Kernel density estimation for directional data under rotational symmetry. *Under preparation.*

Contents of Part I

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

- Note the h^2 because $2(1 - \mathbf{x}^T \mathbf{X}_i) = ||\mathbf{x} - \mathbf{X}_i||^2$
- 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: $b_q(L) = \int_0^\infty L(r) r^{\frac{q}{2}}\, dr \Big/ \int_0^\infty L(r) r^{\frac{q}{2}-1}\, dr$
- If $L(r) = e^{-r}$, the *vMF kernel*, $c_{h,q}(L) = e^{1/h^2} C_q(1/h^2)$

KDE construction: spherical case

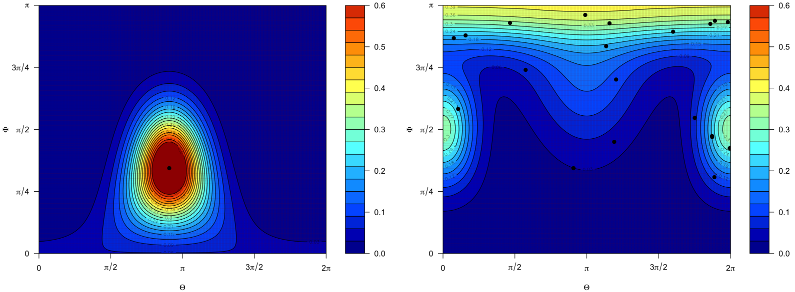


Figure: Left: KDE with $n = 1$. Right: true density

KDE construction: spherical case

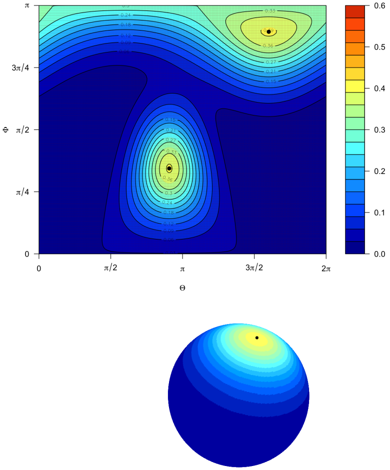


Figure: Left: KDE with $n = 2$. Right: true density

KDE construction: spherical case

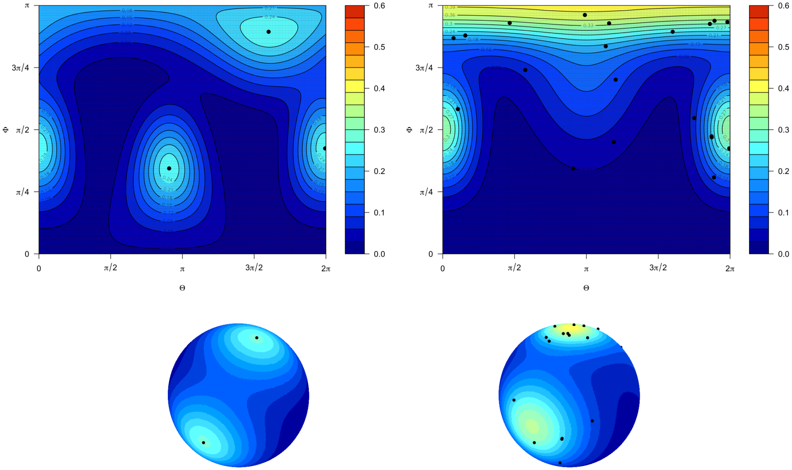


Figure: Left: KDE with $n = 3$. Right: true density

KDE construction: spherical case

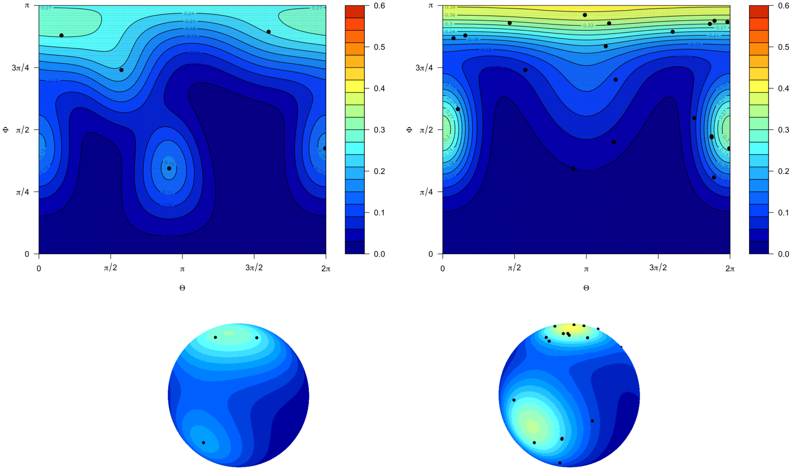


Figure: Left: KDE with $n = 5$. Right: true density

KDE construction: spherical case

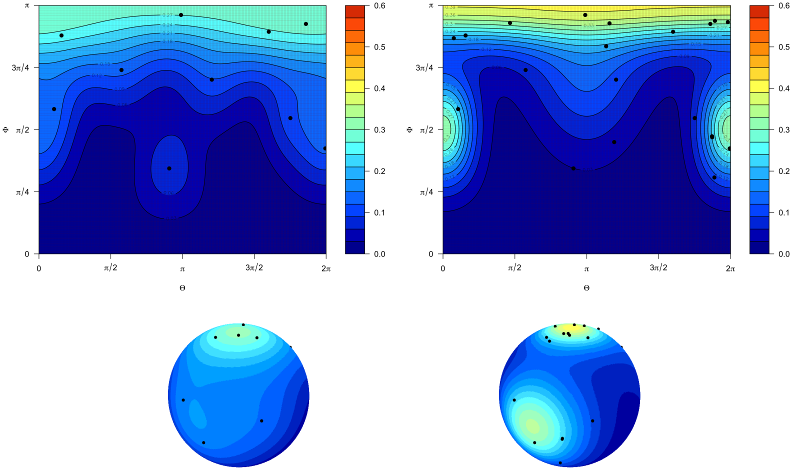


Figure: Left: KDE with $n = 10$. Right: true density

KDE construction: spherical case

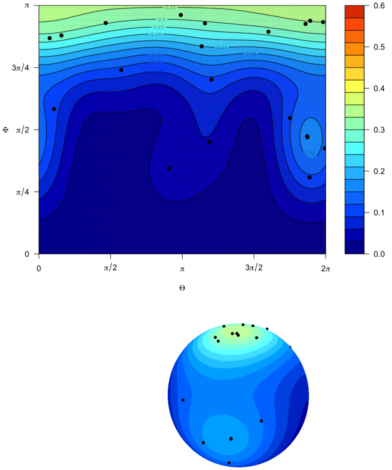
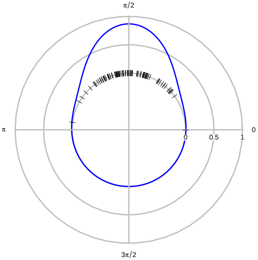


Figure: Left: KDE with $n = 20$. Right: true density

Rotasymmetry I

- Recurrent assumption: **X** is **rotational symmetric** (or **rotasymmetric**) about some direction $\boldsymbol{\theta} \in \Omega_q$
- Circular case: rotasymmetry is **reflective symmetry**
- High-dimensional situation: rotasymmetry is behind many **celebrated distributions**



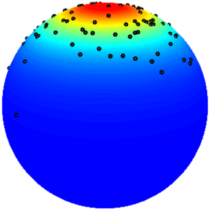


Figure: Rotasymmetry in the circular and spherical cases

Rotasymmetry II

Proposition (Rotasymmetry characterization)

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

1. $\mathbf{X} \stackrel{d}{=} \mathbf{OX}$, *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*
$$f^*(t) = \omega_{q-1} g(t)(1-t^2)^{\frac{q}{2}-1} \text{ is a density in } [-1,1]$$

- Rotasymmetry is related with the **tangent-normal decomposition**:
$$\mathbf{x} = t\boldsymbol{\theta} + (1-t^2)^{\frac{1}{2}}\mathbf{B}_{\boldsymbol{\theta}}\boldsymbol{\xi}$$
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

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

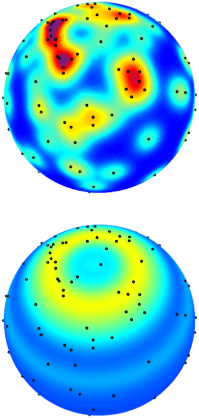


Figure: Input and output of $R_{\boldsymbol{\theta}}$ with $\boldsymbol{\theta} = (0, 0, 1)$

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. ***Characterization***: $R_{\boldsymbol\theta} f = f \iff f$ *is rotasymmetric around* $\boldsymbol\theta$
5. *Explicit expression for the* ***vMF density***:
$$R_{\boldsymbol\theta} f_{\mathrm{vMF}}(\mathbf{x};\boldsymbol\mu,\kappa) = \frac{C_q(\kappa)\exp\left\{\kappa\mathbf{x}^T\boldsymbol\theta\boldsymbol\mu^T\boldsymbol\theta\right\}}{\omega_{q-1}C_{q-1}\left(\kappa\left[(1-(\mathbf{x}^T\boldsymbol\theta)^2)(1-(\boldsymbol\mu^T\boldsymbol\theta)^2)\right]^{\frac{1}{2}}\right)}$$

Rotasymmetric KDE

- **Goal: estimate semiparametrically f under rotasymmetry**

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 vMF kernel has an explicit expression:

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

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

Comparison of kernels

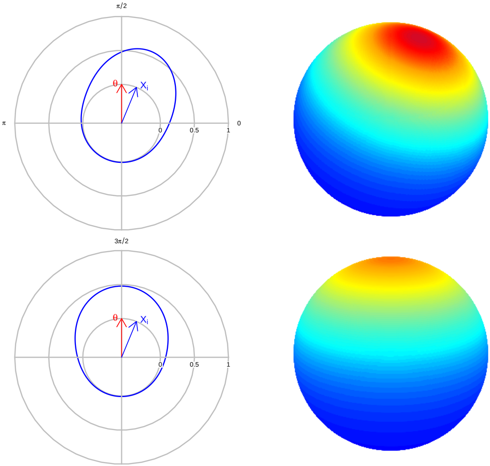


Figure: Kernels for the KDE (upper row) and their RKDE counterparts (lower), with $\boldsymbol{\theta} = (\mathbf{0}_q, 1)$. The kernels have the same bandwidth

Comparison of kernels

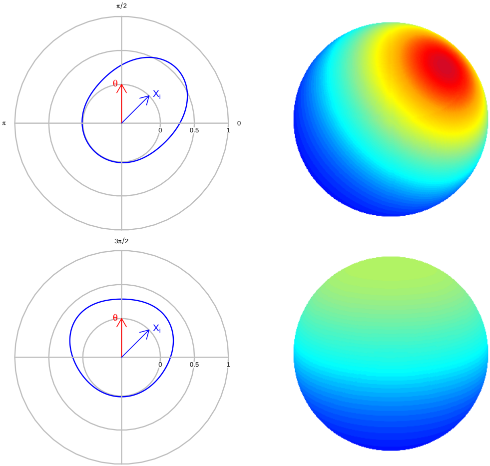


Figure: Kernels for the KDE (upper row) and their RKDE counterparts (lower), with $\boldsymbol{\theta} = (\mathbf{0}_q, 1)$. The kernels have the same bandwidth

Comparison of kernels

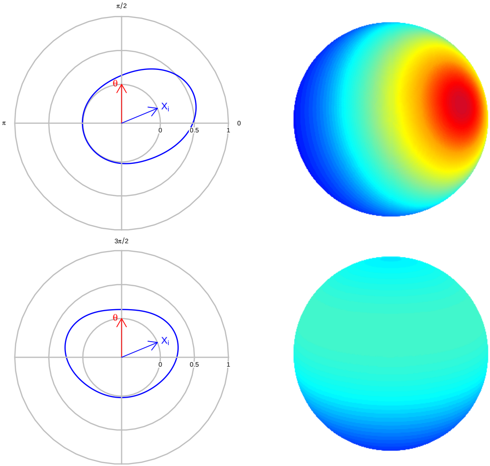


Figure: Kernels for the KDE (upper row) and their RKDE counterparts (lower), with $\boldsymbol{\theta} = (\mathbf{0}_q, 1)$. The kernels have the same bandwidth

Comparison of kernels

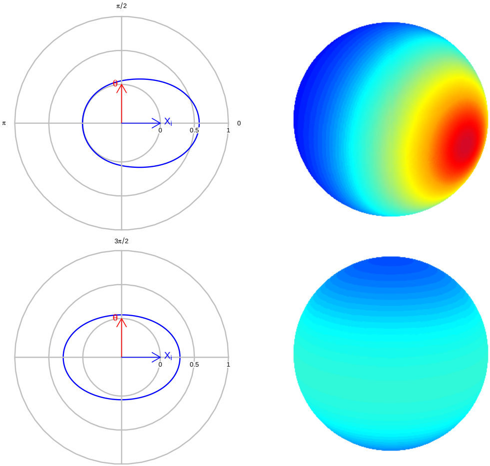


Figure: Kernels for the KDE (upper row) and their RKDE counterparts (lower), with $\boldsymbol{\theta} = (\mathbf{0}_q, 1)$. The kernels have the same bandwidth

Comparison of kernels

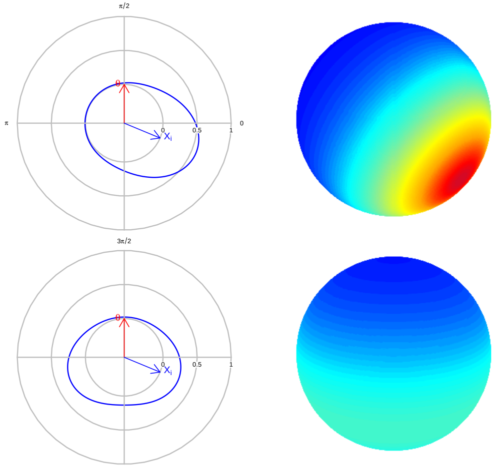


Figure: Kernels for the KDE (upper row) and their RKDE counterparts (lower), with $\boldsymbol{\theta} = (\mathbf{0}_q, 1)$. The kernels have the same bandwidth

Comparison of kernels

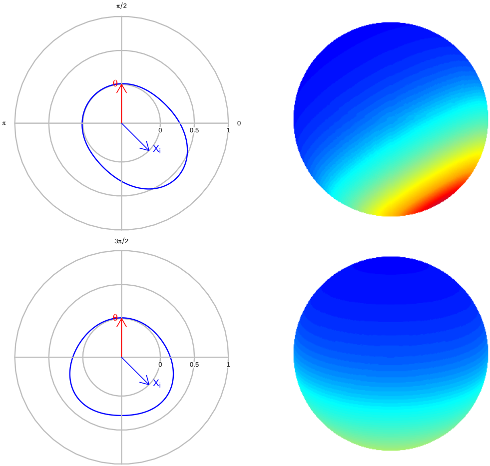


Figure: Kernels for the KDE (upper row) and their RKDE counterparts (lower), with $\boldsymbol{\theta} = (\mathbf{0}_q, 1)$. The kernels have the same bandwidth

Comparison of kernels

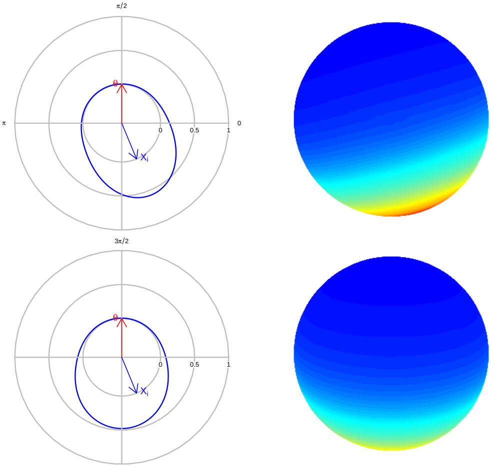


Figure: Kernels for the KDE (upper row) and their RKDE counterparts (lower), with $\boldsymbol{\theta} = (\mathbf{0}_q, 1)$. The kernels have the same bandwidth

Comparison of kernels

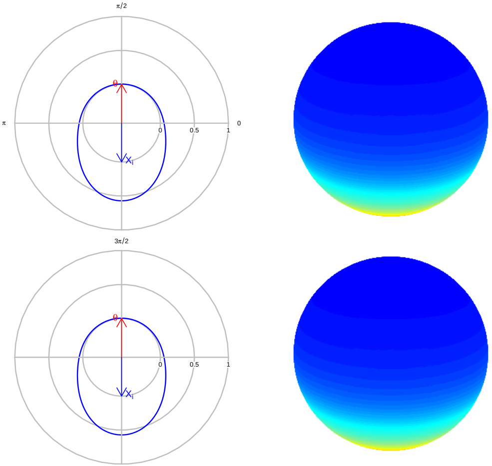


Figure: Kernels for the KDE (upper row) and their RKDE counterparts (lower), with $\boldsymbol{\theta} = (\mathbf{0}_q, 1)$. The kernels have the same bandwidth

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

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

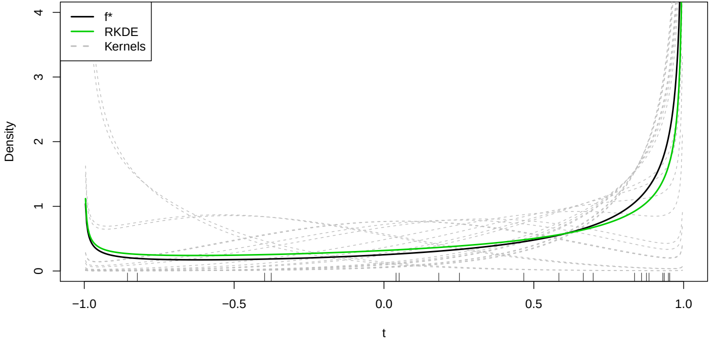


Figure: KDE of f^* with $g(t) = C_q(\kappa)\exp\{\kappa t\}$, $\kappa = 1$ and $q = 1$

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

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

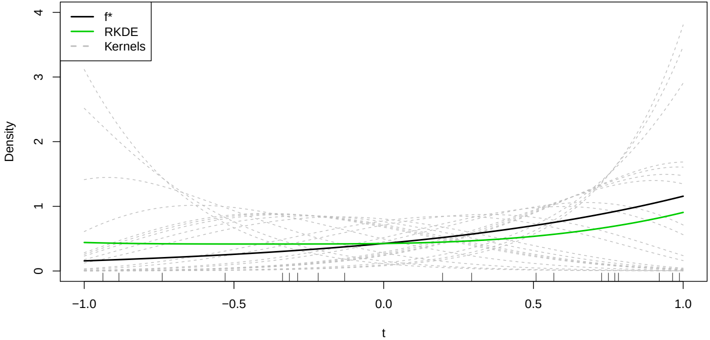


Figure: KDE of f^* with $g(t) = C_q(\kappa)\exp\{\kappa t\}$, $\kappa = 1$ and $q = 2$

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

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

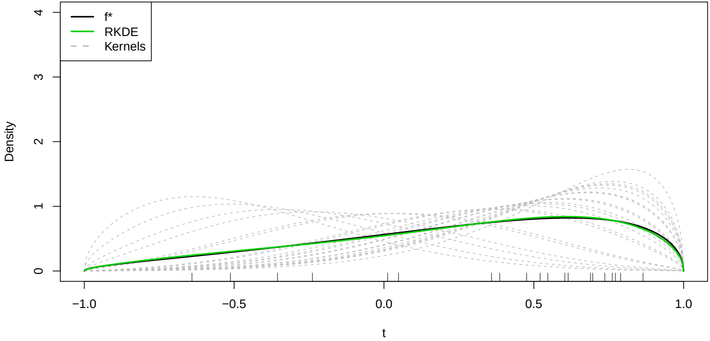


Figure: KDE of f^* with $g(t) = C_q(\kappa)\exp\{\kappa t\}$, $\kappa = 1$ and $q = 3$

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

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

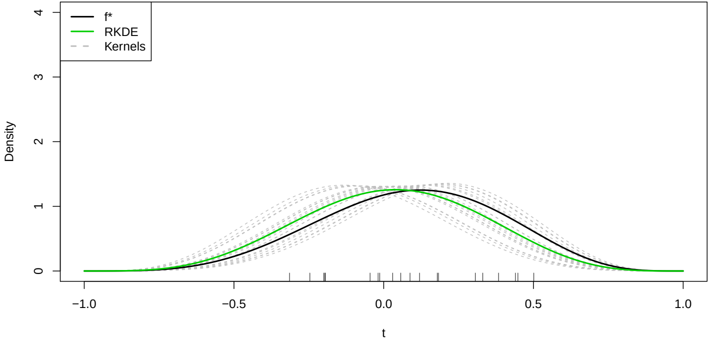


Figure: KDE of f^* with $g(t) = C_q(\kappa)\exp\{\kappa t\}$, $\kappa = 1$ and $q = 10$

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

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

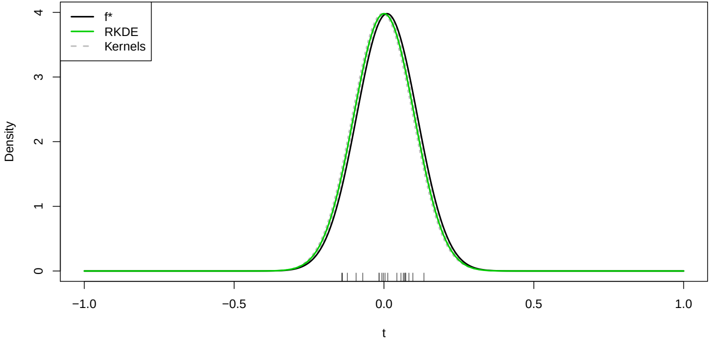


Figure: KDE of f^* with $g(t) = C_q(\kappa) \exp\{\kappa t\}$, $\kappa = 1$ and $q = 100$

Bias (θ known)

- **Assumptions:**

  A1 f is extended by $f\left(\mathbf{x}/\left|\left|\mathbf{x}\right|\right|\right)$ and is twice continuously differentiable
  A2 $L:\mathbb{R}^+\rightarrow\mathbb{R}^+$ is continuous, bounded and has exponential decay
  A3-1 The sequence $h=h_n$ satisfies $h\rightarrow 0$ and $nh\rightarrow\infty$
  A3-q The sequence $h=h_n$ satisfies $h\rightarrow 0$ and $nh^q\rightarrow\infty$

- **A3**-q is required for consistency at $\mathbf{x}=\pm\boldsymbol{\theta}$ (note **A3**-q $\Rightarrow$ **A3**-1)

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

*Under **A1**–**A3**-1 and uniformly in* $\mathbf{x}\in\Omega_q$,

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

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

Variance (θ known)

Proposition (Variance, θ 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

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

- The asymptotic constant of the variance increases with $q \to \infty$ since $\omega_{q-1} \to 0$! (but slowly than KDE's)

Spherical area surface

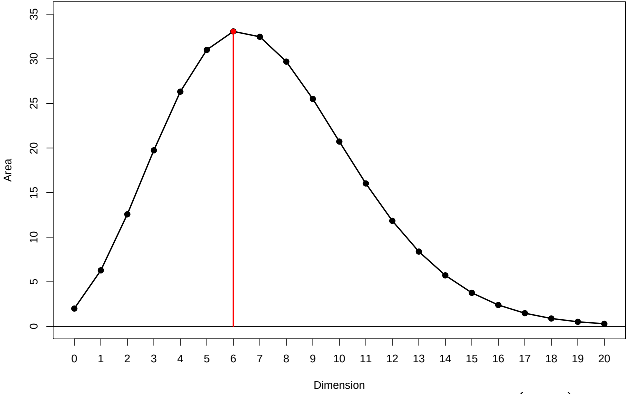


Figure: Spherical surface $\omega_q = 2\pi^{\frac{q+1}{2}} / \Gamma\left(\frac{q+1}{2}\right)$

- 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!

Key orders & asymptotic normality

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

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

What if θ is unknown?

- **Assumption:**

  A4 $\hat{\boldsymbol\theta}$ is a $\sqrt{n}$-consistent estimator: $\hat{\boldsymbol\theta} - \boldsymbol\theta = \mathcal{O}_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$

- Examples of $\hat{\boldsymbol\theta}$:
  - If $\mathbf{X}$ such that g is strictly monotone, $\sum_{i=1}^n \mathbf{X}_i \big/ \left|\left|\sum_{i=1}^n \mathbf{X}_i\right|\right|$
  - If $\mathbf{X}$ is an axial rv, the first eigenvector of $\frac{1}{n}\sum_{i=1}^n \mathbf{X}_i\mathbf{X}_i^T$

- **Work in progress**: under **A1–A2**, **A3-**1/**A3-**q and **A4**:

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

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

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

Simulation study

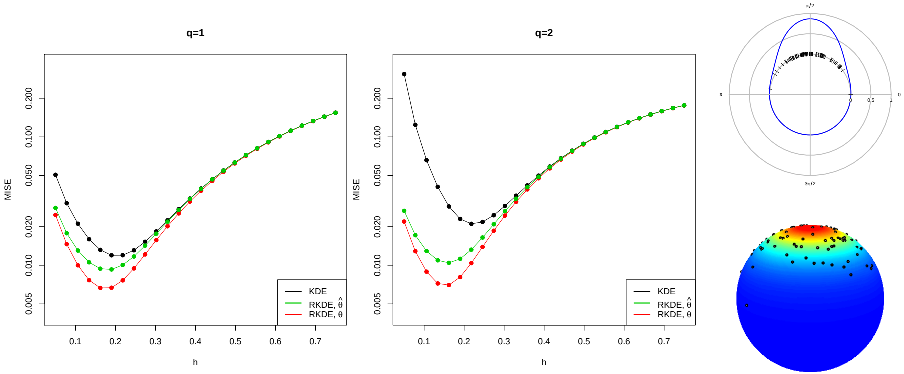
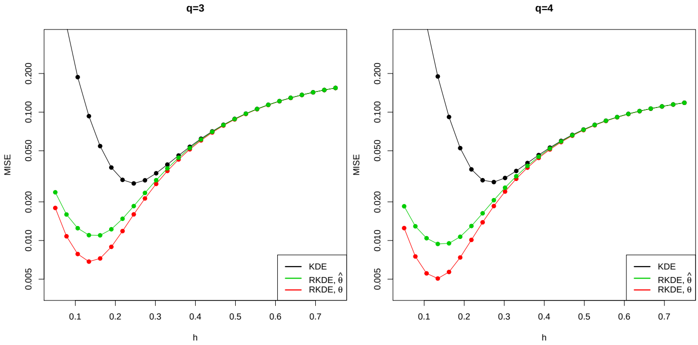


Figure: Performance of the three kernel estimators with $q = 1$ (left) and $q = 2$ (right), with $n = 100$

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

Simulation study



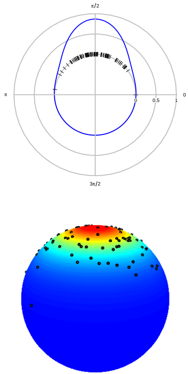


Figure: Performance of the three kernel estimators with $q = 3$ (left) and $q = 4$ (right), with $n = 100$

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

Simulation study

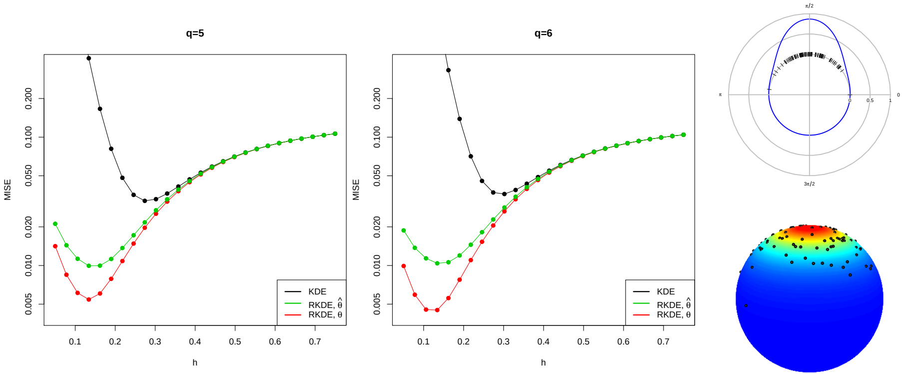


Figure: Performance of the three kernel estimators with $q = 5$ (left) and $q = 6$ (right), with $n = 100$

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

Part II

Estimation and testing in linear-directional regression

García-Portugués, E., Van Keilegom, I., Crujeiras, R. and González-Manteiga, W. (2016). Testing parametric models in linear-directional regression. *Scand. J. Stat. (to appear)*

Contents of Part II

Regression with directional data

- Let $(\mathbf{X}, Y)$ be a rv with support in $\Omega_q \times \mathbb{R}$ and $\mathbf{X}$ having density f
- Consider the regression model

$$Y = m(\mathbf{X}) + \sigma(\mathbf{X})\varepsilon \quad \text{with} \quad \begin{cases} m(\mathbf{x}) = \mathbb{E}\left[Y|\mathbf{X} = \mathbf{x}\right], \\ \sigma^2(\mathbf{x}) = \mathbb{V}\mathrm{ar}\left[Y|\mathbf{X} = \mathbf{x}\right], \end{cases}$$

with $\mathbb{E}\left[\varepsilon|\mathbf{X}\right] = 0$, $\mathbb{E}\left[\varepsilon^2|\mathbf{X}\right] = 1$ and $\mathbb{E}\left[|\varepsilon|^3|\mathbf{X}\right]$ and $\mathbb{E}\left[\varepsilon^4|\mathbf{X}\right]$ bounded rv's

- **Goal: estimate m nonparametrically from $\{(\mathbf{X}_i, Y_i)\}_{i=1}^n$**
- Taylor expansions are required, so the first condition is:

  A1 m and f ar extended as $m\left(\mathbf{x}/\left|\left|\mathbf{x}\right|\right|\right)$ and $f\left(\mathbf{x}/\left|\left|\mathbf{x}\right|\right|\right)$. m is third and f is twice continuously differentiable and f is bounded away from zero

Estimator

- Let $\mathbf{x}, \mathbf{X}_i \in \Omega_q$. The one term Taylor expansion of m is:

$$m(\mathbf{X}_i) = m(\mathbf{x}) + \boldsymbol{\nabla} m(\mathbf{x})^T (\mathbf{X}_i - \mathbf{x}) + \mathcal{O}\left(||\mathbf{X}_i - \mathbf{x}||^2\right)$$

Estimator

- Let $\mathbf{x}, \mathbf{X}_i \in \Omega_q$. The one term Taylor expansion of m is:

$$m(\mathbf{X}_i) = m(\mathbf{x}) + \boldsymbol{\nabla} m(\mathbf{x})^T \left(\mathbf{I}_{q+1} - \mathbf{x}\mathbf{x}^T\right)(\mathbf{X}_i - \mathbf{x}) + \mathcal{O}\left(||\mathbf{X}_i - \mathbf{x}||^2\right)$$

Estimator

- Let $\mathbf{x}, \mathbf{X}_i \in \Omega_q$. The one term Taylor expansion of m is:

$$m(\mathbf{X}_i) \;=\; m(\mathbf{x}) + \boldsymbol{\nabla} m(\mathbf{x})^T \mathbf{B}_{\mathbf{x}} \mathbf{B}_{\mathbf{x}}^T (\mathbf{X}_i - \mathbf{x}) + \mathcal{O}\left(||\mathbf{X}_i - \mathbf{x}||^2\right)$$

Estimator

- Let $\mathbf{x}, \mathbf{X}_i \in \Omega_q$. The one term Taylor expansion of m is:

$$\begin{aligned} m(\mathbf{X}_i) &= m(\mathbf{x}) + \boldsymbol{\nabla} m(\mathbf{x})^T \mathbf{B_x} \mathbf{B_x}^T (\mathbf{X}_i - \mathbf{x}) + \mathcal{O}\left(||\mathbf{X}_i - \mathbf{x}||^2\right) \\ &\approx \beta_0 + (\beta_1, \ldots, \beta_q)^T \mathbf{B_x}^T (\mathbf{X}_i - \mathbf{x}), \end{aligned}$$

  with $\mathbf{B_x} = (\mathbf{b}_1, \ldots, \mathbf{b}_q)_{(q+1)\times q}$ such that $\mathbf{B_x}\mathbf{B_x}^T = \mathbf{I}_{q+1} - \mathbf{x}\mathbf{x}^T$, $\beta_0 = m(\mathbf{x})$ and $(\beta_1, \ldots, \beta_q) = \mathbf{B_x}^T \boldsymbol{\nabla} m(\mathbf{x})$

Estimator

- Let $\mathbf{x}, \mathbf{X}_i \in \Omega_q$. The one term Taylor expansion of m is:

$$\begin{aligned} m(\mathbf{X}_i) &= m(\mathbf{x}) + \boldsymbol{\nabla} m(\mathbf{x})^T \mathbf{B}_{\mathbf{x}} \mathbf{B}_{\mathbf{x}}^T (\mathbf{X}_i - \mathbf{x}) + \mathcal{O}\left(\|\mathbf{X}_i - \mathbf{x}\|^2 \right) \\ &\approx \beta_0 + (\beta_1, \ldots, \beta_q)^T \, \mathbf{B}_{\mathbf{x}}^T (\mathbf{X}_i - \mathbf{x}), \end{aligned}$$

with $\mathbf{B}_{\mathbf{x}} = (\mathbf{b}_1, \ldots, \mathbf{b}_q)_{(q+1)\times q}$ such that $\mathbf{B}_{\mathbf{x}}\mathbf{B}_{\mathbf{x}}^T = \mathbf{I}_{q+1} - \mathbf{x}\mathbf{x}^T$, $\beta_0 = m(\mathbf{x})$ and $(\beta_1, \ldots, \beta_q) = \mathbf{B}_{\mathbf{x}}^T \boldsymbol{\nabla} m(\mathbf{x})$

- Weighted minimum least squares problem:

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^{q+1}} \sum_{i=1}^n \left(Y_i - \beta_0 - \delta_{p,1} \left(\beta_1, \ldots, \beta_q\right)^T \mathbf{B}_{\mathbf{x}}^T (\mathbf{X}_i - \mathbf{x}) \right)^2 L_h(\mathbf{x}, \mathbf{X}_i)$$

Estimator

- Let $\mathbf{x}, \mathbf{X}_i \in \Omega_q$. The one term Taylor expansion of m is:

$$\begin{aligned} m(\mathbf{X}_i) &= m(\mathbf{x}) + \boldsymbol{\nabla} m(\mathbf{x})^T \mathbf{B_x} \mathbf{B_x}^T (\mathbf{X}_i - \mathbf{x}) + \mathcal{O}\left(\|\mathbf{X}_i - \mathbf{x}\|^2\right) \\ &\approx \beta_0 + (\beta_1, \ldots, \beta_q)^T \mathbf{B_x}^T (\mathbf{X}_i - \mathbf{x}), \end{aligned}$$

  with $\mathbf{B_x} = (\mathbf{b}_1, \ldots, \mathbf{b}_q)_{(q+1)\times q}$ such that $\mathbf{B_x}\mathbf{B_x}^T = \mathbf{I}_{q+1} - \mathbf{x}\mathbf{x}^T$, $\beta_0 = m(\mathbf{x})$ and $(\beta_1, \ldots, \beta_q) = \mathbf{B_x}^T \boldsymbol{\nabla} m(\mathbf{x})$

- Weighted minimum least squares problem:

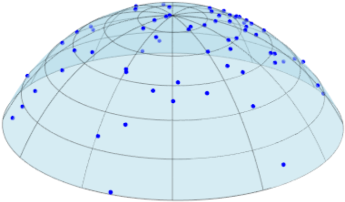
$$\min_{\boldsymbol{\beta} \in \mathbb{R}^{q+1}} \sum_{i=1}^n \left(Y_i - \beta_0 - \delta_{p,1} (\beta_1, \ldots, \beta_q)^T \mathbf{B_x}^T (\mathbf{X}_i - \mathbf{x})\right)^2 L_h(\mathbf{x}, \mathbf{X}_i)$$

- The solution is given by

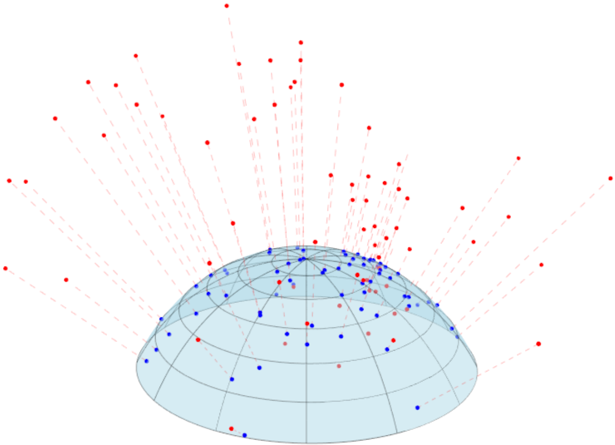
$$\hat{m}_{h,p}(\mathbf{x}) = \mathbf{e}_{1,p}^T \left(\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T \boldsymbol{\mathcal{W}}_{\mathbf{x}} \boldsymbol{\mathcal{X}}_{\mathbf{x},p}\right)^{-1} \boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T \boldsymbol{\mathcal{W}}_{\mathbf{x}} \mathbf{Y} = \sum_{i=1}^n W_p^n(\mathbf{x}, \mathbf{X}_i)\, Y_i,$$

$$\boldsymbol{\mathcal{X}}_{\mathbf{x},1} = \begin{pmatrix} 1 & (\mathbf{X}_1 - \mathbf{x})^T \mathbf{B_x} \\ \vdots & \vdots \\ 1 & (\mathbf{X}_n - \mathbf{x})^T \mathbf{B_x} \end{pmatrix}, \quad \boldsymbol{\mathcal{W}}_{\mathbf{x}} = \mathrm{diag}\left(L_h(\mathbf{x}, \mathbf{X}_1), \ldots, L_h(\mathbf{x}, \mathbf{X}_n)\right)$$

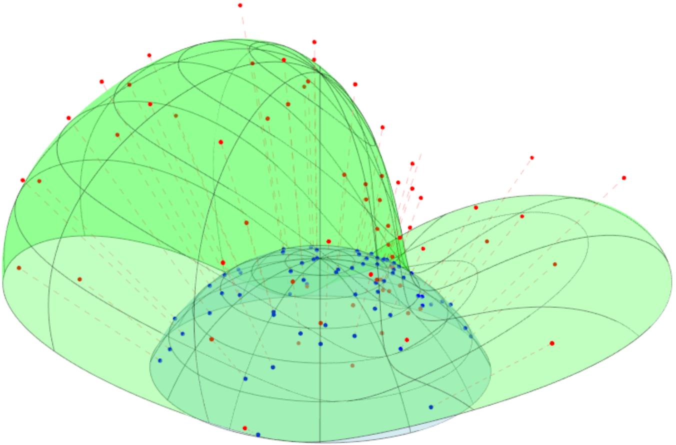
How does it work?



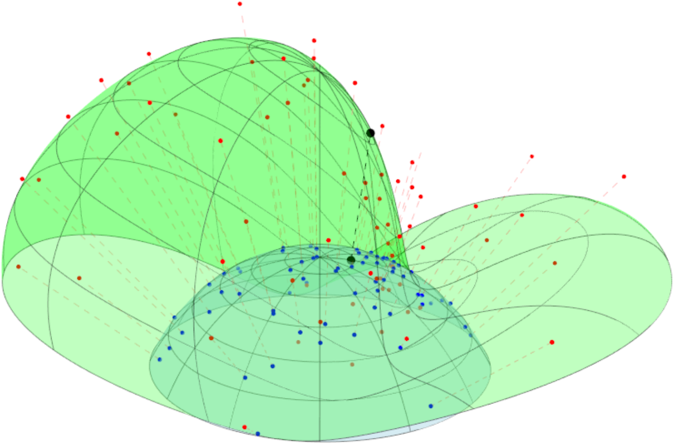
How does it work?



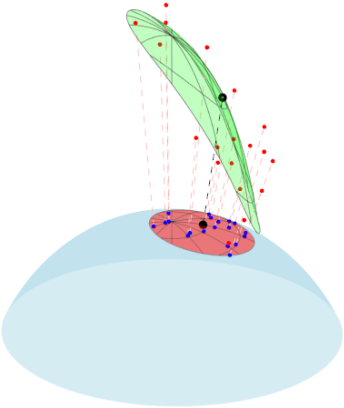
How does it work?



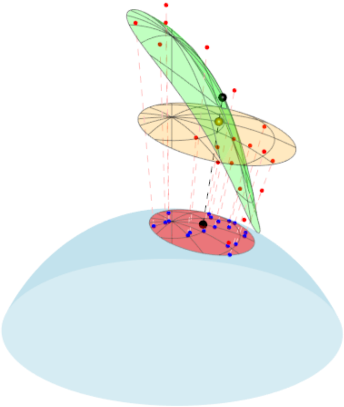
How does it work?



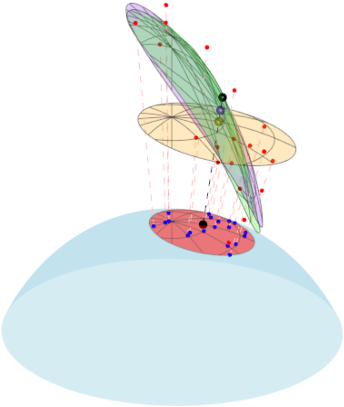
How does it work?



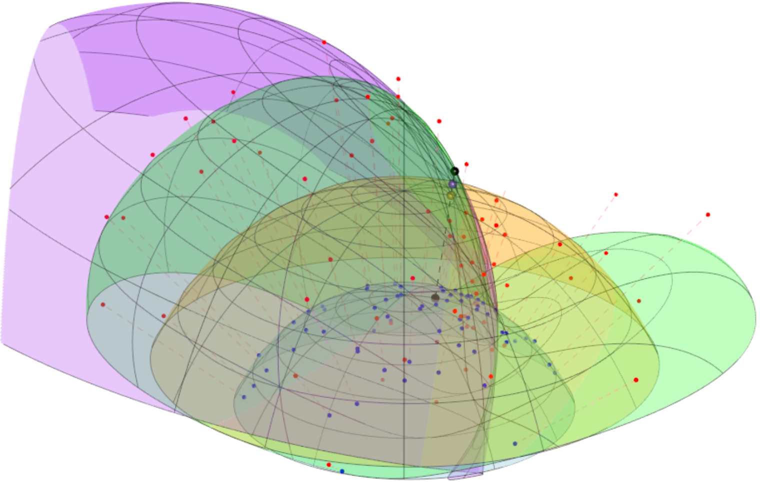
How does it work?



How does it work?



How does it work?



Output

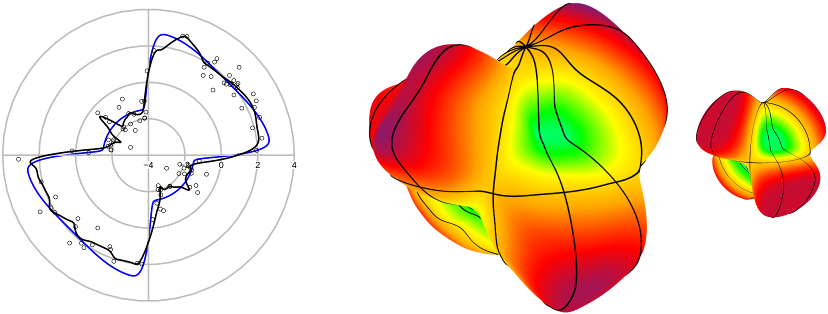


Figure: Local linear estimator with $n = 100$ for the circle and the sphere

Testing a parametric model

- **Goal: check nonparametrically** $H_0 : m \in \mathcal{M}_\Theta = \{m_{\boldsymbol\theta} : \boldsymbol\theta \in \Theta \subset \mathbb{R}^s\}$
- The statistic is the weighted L^2-distance between $\hat m_{h,p}$ and the smoothed $m_{\hat{\boldsymbol\theta}}$:

$$T_n = \int_{\Omega_q} \left(\hat m_{h,p}(\mathbf{x}) - \mathcal{L}_{h,p} m_{\hat{\boldsymbol\theta}}(\mathbf{x})\right)^2 \hat f_h(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x}),$$

with $\mathcal{L}_{h,p} m_{\hat{\boldsymbol\theta}}(\mathbf{x}) = \sum_{i=1}^n W_n^p(\mathbf{x}, \mathbf{X}_i)\, m_{\hat{\boldsymbol\theta}}(\mathbf{X}_i)$ the smoothing operator and $w : \Omega_q \to \mathbb{R}^+$ a weight function (useful for removing possible boundary effects)

Alcalá, J. T., Cristóbal, J. A., and González-Manteiga, W. (1999). Goodness-of-fit test for linear models based on local polynomials. *Statist. Probab. Lett.*, 42(1):39–46

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.*, 21(4):1926–1947

Asymptotic distribution

Theorem (Goodness-of-fit for linear-directional models)

Under A1–A6 and $H_0 : m \in \mathcal{M}_\Theta$ (i.e., $m = m_{\boldsymbol{\theta}_0}$),

$$nh^{\frac{q}{2}}\left(T_n - \frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{nh^q}\int_{\Omega_q}\sigma^2_{\boldsymbol{\theta}_0}(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x})\right)\xrightarrow{d}\mathcal{N}\left(0,2\nu^2_{\boldsymbol{\theta}_0}\right),$$

where $\sigma^2_{\boldsymbol{\theta}_0}(\mathbf{x}) = \mathbb{E}\left[(Y - m_{\boldsymbol{\theta}_0}(\mathbf{X}))^2|\mathbf{X} = \mathbf{x}\right]$ and

$$\nu^2_{\boldsymbol{\theta}_0} = \int_{\Omega_q}\sigma^4_{\boldsymbol{\theta}_0}(\mathbf{x})w(\mathbf{x})^2\,\omega_q(d\mathbf{x})$$

$$\times\,\gamma_q\lambda_q(L)^{-4}\int_0^\infty r^{\frac{q}{2}-1}\left\{\int_0^\infty \rho^{\frac{q}{2}-1}L(\rho)\varphi_q(r,\rho)\,d\rho\right\}^2 dr$$

- Conditions:
  - A5 $\hat{\boldsymbol{\theta}}$ is such that $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_1 = \mathcal{O}_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$, with $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_0$ if H_0 holds
  - A6 $m_{\boldsymbol{\theta}}$ is continuously differentiable as a function of $\boldsymbol{\theta}$, being this derivative also continuous for $\mathbf{x} \in \Omega_q$
- If L is the von Mises kernel, $\nu^2_{\boldsymbol{\theta}_0} = \int_{\Omega_q}\sigma^4_{\boldsymbol{\theta}_0}(\mathbf{x})w(\mathbf{x})^2\,\omega_q(d\mathbf{x}) \times (8\pi)^{-\frac{q}{2}}$

Empirical evidence

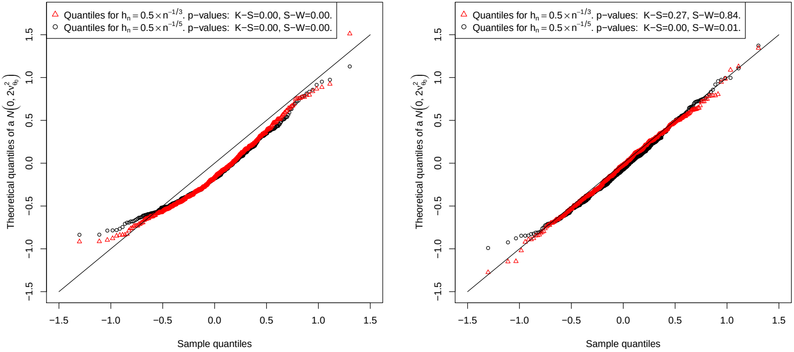


Figure: QQ-plot comparing the quantiles of the asymptotic distribution with the sample quantiles for $\left\{nh^{\frac{1}{2}}\left(T_n^j - \frac{\sqrt{\pi}}{4}nh\right)\right\}_{j=1}^{500}$ with $n = 10^2$ (left) and $n = 5 \times 10^5$ (right)

Calibration in practice

Algorithm (Calibration in practice)

To test $H_0 : m \in \mathcal{M}_{\mathbf{\Theta}}$ from the sample $\{(\mathbf{X}_i, Y_i)\}_{i=1}^n$:

1. *Obtain $\hat{\boldsymbol{\theta}}$, set $\hat{\varepsilon}_i = Y_i - m_{\hat{\boldsymbol{\theta}}}(\mathbf{X}_i)$, $i = 1, \ldots, n$ and compute T_n*
2. *Bootstrap resampling. For $b = 1, \ldots, B$:*
   - *Set $Y_i^* = m_{\hat{\boldsymbol{\theta}}}(\mathbf{X}_i) + \hat{\varepsilon}_i V_i^*$, where V_i^* are iid rv's such that $\mathbb{E}^*[V_i^*] = 0$ and $\mathbb{E}^*[(V_i^*)^2] = 1$, $i = 1, \ldots, n$*
   - *Compute $\hat{\boldsymbol{\theta}}^*$ from $\{(\mathbf{X}_i, Y_i^*)\}_{i=1}^n$ and T_n^{*b}*
3. *Estimate the p-value by $\frac{1}{B}\sum_{b=1}^B \mathbf{1}_{\{T_n \leq T_n^{*b}\}}$*

Theorem (Bootstrap consistency)

Under A1–A4, A5–A6 and A9, conditionally on the sample,

$$nh^{\frac{q}{2}}\left(T_n^* - \frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{nh^q}\int_{\Omega_q}\sigma_{\boldsymbol{\theta}_1}^2(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x})\right) \xrightarrow{d} \mathcal{N}\left(0, 2\nu_{\boldsymbol{\theta}_1}^2\right)$$

in probability. If H_0 holds, then $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_0$ and $T_n^* \stackrel{d}{=} T_n$ asymptotically

Protein structure modelling

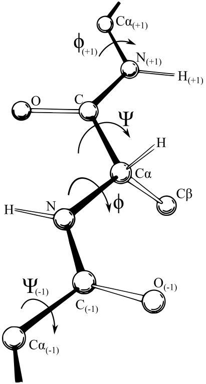


Figure: Backbone and C_α representation

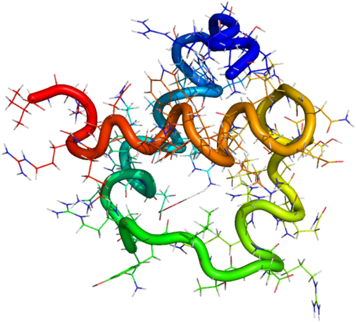


Figure: Cartoon view of a protein

Boomsma, W., Mardia, K. V., Taylor, C. C., Ferkinghoff-Borg, J., Krogh, A. and Hamelryck, T. A generative, probabilistic model of local protein structure. *PNAS*, 105(26):8932-8937

Protein structure modelling

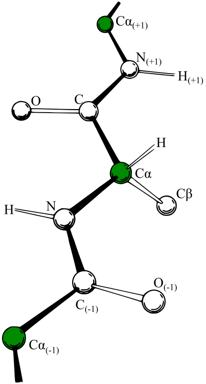


Figure: Backbone and C_α representation

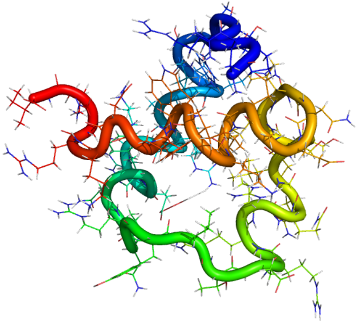


Figure: Cartoon view of a protein

Boomsma, W., Mardia, K. V., Taylor, C. C., Ferkinghoff-Borg, J., Krogh, A. and Hamelryck, T. A generative, probabilistic model of local protein structure. *PNAS*, 105(26):8932-8937

Protein structure modelling

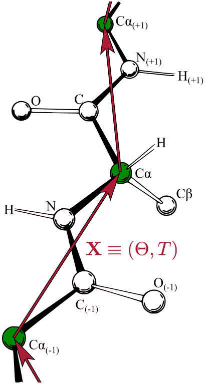


Figure: Backbone and C_α representation

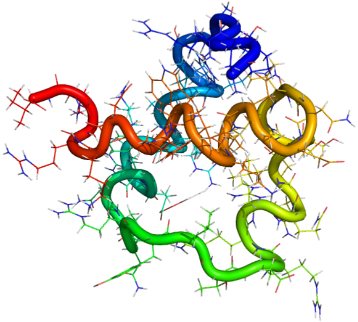


Figure: Cartoon view of a protein

Boomsma, W., Mardia, K. V., Taylor, C. C., Ferkinghoff-Borg, J., Krogh, A. and Hamelryck, T. A generative, probabilistic model of local protein structure. *PNAS*, 105(26):8932-8937

Testing in the C_α representation

- **Goal: test the constant pseudo- -bond length assumption**:

$$H_0 : m(\mathbf{x}) = c,\ c \in \mathbb{R}$$

- Data: $n = 18030$ pseudo-angles ($\mathbf{X} \equiv (\Theta, T)$) and pseudo-lengths ($Y$) extracted from 100 high precision protein structures
- Grid of 10 bandwidths, $B = 1000$ bootstrap replicates and weight $w(\theta,\tau) = \mathbf{1}_{\{80 \leq \frac{180}{\pi}\theta \leq 150\}}$
- **Emphatically rejection of H_0**
- Exploration of $m(\theta,\tau)$ by local linear estimator $\hat{m}_{h_{\mathrm{CV}},1}(\theta,\tau)$

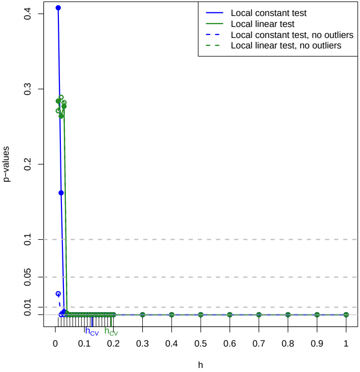


Figure: Significance trace of the goodness-of-fit tests

Testing in the C_α representation

- **Goal: test the constant pseudo--bond length assumption**:

$$H_0 : m(\mathbf{x}) = c, \; c \in \mathbb{R}$$

- Data: $n = 18030$ pseudo-angles ($\mathbf{X} \equiv (\Theta, T)$) and pseudo-lengths ($Y$) extracted from 100 high precision protein structures
- Grid of 10 bandwidths, $B = 1000$ bootstrap replicates and weight $w(\theta, \tau) = \mathbf{1}_{\left\{80 \leq \frac{180}{\pi}\theta \leq 150\right\}}$
- **Emphatically rejection of H_0**
- Exploration of $m(\theta, \tau)$ by local linear estimator $\hat{m}_{h_{\mathrm{CV}},1}(\theta, \tau)$

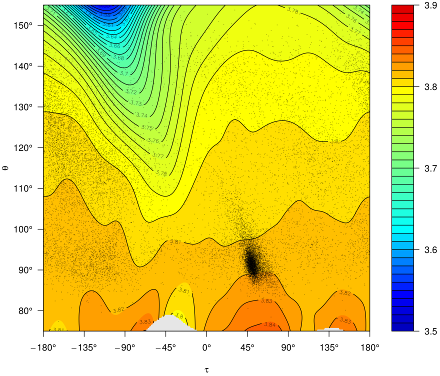


Figure: Contourplot of $\hat{m}_{h_{\mathrm{CV}},1}(\theta, \tau)$ and pseudo-angles sample

Thanks for your attention!