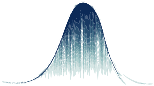
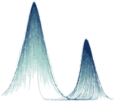
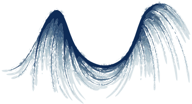


Universidad Autónoma de Madrid

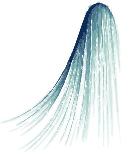


Estimation of differential and topological features of probability density functions

Javier Fernández Serrano



Departamento de Matemáticas
Universidad Autónoma de Madrid



PhD thesis defense, Madrid, 2026

FACULTAD DE CIENCIAS

DEPARTAMENTO DE MATEMÁTICAS
UNIVERSIDAD AUTÓNOMA DE MADRID

Thesis by publication

arXiv > a

Javier Fernández Serrano's articles on arXiv

See also ORCID https://orcid.org/0000-0001-5270-9941.

[1] **arXiv:2508.18909 [pdf, ps, other]**

Rejoinder to the discussion of "Mode-based estimation of the center of symmetry"
José E. Chacón, Javier Fernández Serrano
Comments: 3 pages. Related article: arXiv:2406.08241
Journal-ref: Annals of the Institute of Statistical Mathematics 77 (2025) 727-730
Subjects: **Methodology (stat.ME)**

[2] **arXiv:2406.08241 [pdf, ps, other]**

Mode-based estimation of the center of symmetry
José E. Chacón, Javier Fernández Serrano
Comments: 27 pages, 8 figures
Journal-ref: Annals of the Institute of Statistical Mathematics 77 (2025) 685-717
Subjects: **Methodology (stat.ME)**; Statistics Theory (math.ST)

[3] **arXiv:2307.05825 [pdf, other]**

Bayesian taut splines for estimating the number of modes
José E. Chacón, Javier Fernández Serrano
Comments: 21 pages, 8 figures (manuscript) + 22 pages, 17 figures (supplementary material)
Journal-ref: Computational Statistics and Data Analysis 196 (2024) 107961
Subjects: **Methodology (stat.ME)**; Machine Learning (cs.LG); Statistics Theory (math.ST); Machine Learning (stat.ML)

[4] **arXiv:2208.00174 [pdf, other]**

Bump hunting through density curvature features
José E. Chacón, Javier Fernández Serrano
Comments: 19 pages, 5 figures (manuscript) + 13 pages, 2 figures (supplementary material)
Journal-ref: TEST 32 (2023) 1251-1275
Subjects: **Methodology (stat.ME)**; Statistics Theory (math.ST); Machine Learning (stat.ML)

Outline

Overview 📌

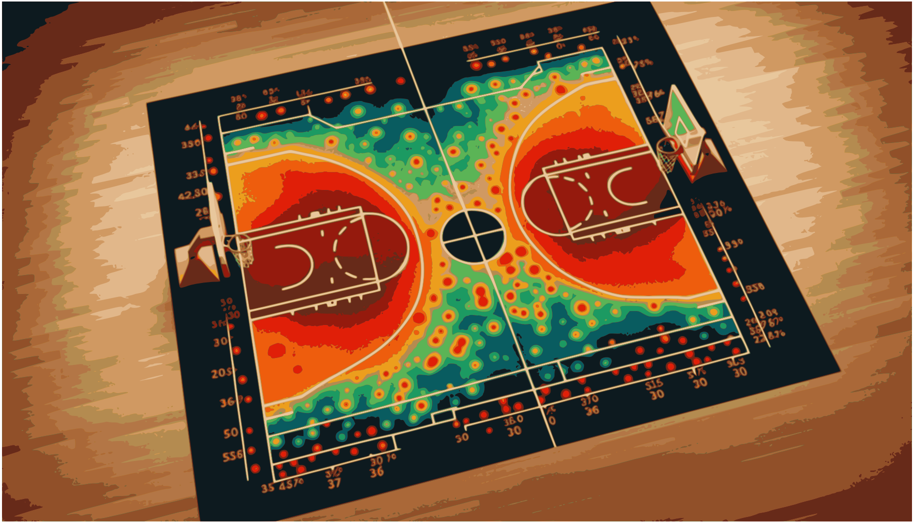
Data analysis & statistics 🧮→💻

- A relatively old discipline entering a new era.
- A joint venture between industry and academia.
- More complexity and sophistication:
  - Univariate → multivariate → infinite-dimensional
  - Parametric → nonparametric
  - Classic → Bayesian

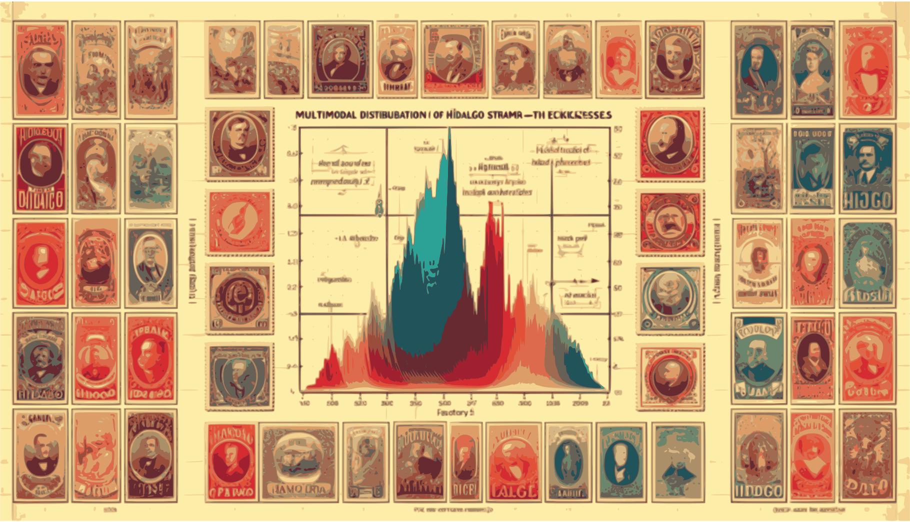
A classic thesis for modern times 🚂→🚄

- Exploiting the kernel density estimator (Chacón and Duong, 2018).
- Following the theses by Ameijeiras-Alonso (2017) and Casa (2019).
- Starring the concept of mode (Chacón, 2020).
- Honoring the spirit of Good and Gaskins (1980).

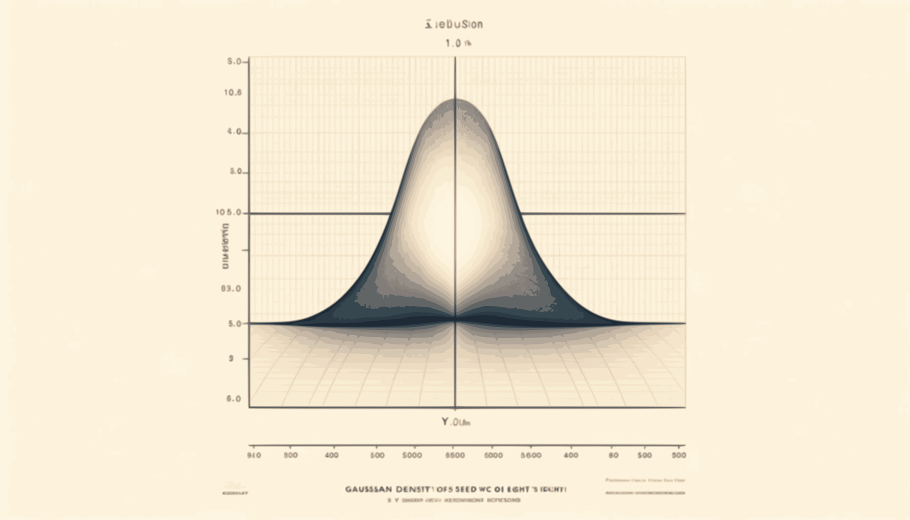
Curvature features ⛰️ (Chacón and FS, 2023)



Number of modes 🍀 (Chacón and FS, 2024)



Center of symmetry 🪞 (Chacón and FS, 2025a,b)



Outline

[4] arXiv:2208.00174 [pdf, other]

Bump hunting through density curvature features

José E. Chacón, Javier Fernández Serrano

Comments: 19 pages, 5 figures (manuscript) + 13 pages, 2 figures (supplementary material)

Journal-ref: TEST 32 (2023) 1251-1275

Subjects: **Methodology (stat.ME)**; Statistics Theory (math.ST); Machine Learning (stat.ML)

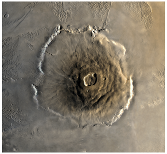
What is bump hunting (BH)?

"Bump hunting" (Good and Gaskins, 1980)

Discovering meaningful data regions, called bumps, in a sample space:

- "Literal" bumps → the core of modal regions
- The classic approach → highest density region (HDR)
- A broader scope, relatively unexplored → curvature regions

Usage in exploratory data analysis.


A brief history of BH 📽️

1980. Good and Gaskins (1980) started BH addressing concave bumps in the univariate case.

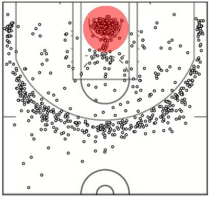
1996. Hyndman (1996) introduced HDRs.

1999. Chaudhuri and Marron (1999) presented SiZer, envisioning bumps as places where the first derivative becomes zero. They provided an extension with second derivatives in Chaudhuri and Marron (2002).

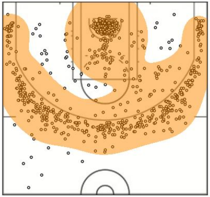
2002. Godtliebsen et al. (2002) explored curvature features from a pointwise perspective by assessing Hessian eigenvalue sign combinations in the bivariate case.

2008. Duong et al. (2008) continued the work on curvature in a multivariate setting, targeting the pointwise significance of nonzero Hessian determinants.

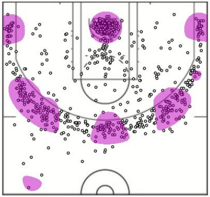
Curvature bumps VS HDRs



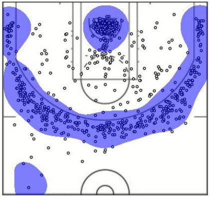
(a) 35%-HDR



(b) 95%-HDR



(c) Concave bumps



(d) Laplacian bumps

Abstract construction ☁️

Curvature bump 💡

Consider a d-variate probability density function (pdf) $f : \mathbb{R}^d \to [0, \infty)$. We define bumps as subsets of $\mathbb{R}^d$ of the form

$$\mathcal{B}^{\phi} = \{\mathrm{x} \in \mathbb{R}^d : (-1)^s \phi[f](\mathrm{x}) \geq 0\},$$

for some functional ϕ measuring the curvature of f at any point, and some sign selector $s \in \{0, 1\}$.

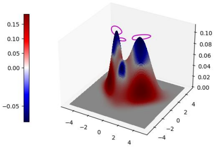
Solving arbitrariness 😌

Contrary to HDRs, ϕ carries an implicit threshold, say zero.

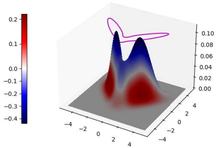
Geometry and derivatives 🔒

Our quest for curvature features is fulfilled with up to second derivatives.

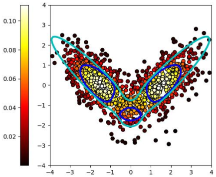
Curvature bumps illustrated



(a) Concave bumps



(b) Mean curvature bump



(c) Both kinds of bumps

Concavity and convexity ↕️ I

Given a smooth pdf f, define $\lambda_i[f]$, for $i \in \{1, 2, \dots, d\}$, as the function mapping $\mathrm{x} \in \mathbb{R}^d$ to the i-th largest possibly repeated eigenvalue of $\mathrm{D}^2 f(\mathrm{x})$, the Hessian matrix of f at x, i.e.,

$$\lambda_1[f](\mathrm{x}) \geq \lambda_2[f](\mathrm{x}) \geq \cdots \geq \lambda_d[f](\mathrm{x})\,.$$

Local concavity/convexity 🤓

Let us assume that $(-1)^s \lambda_i[f] > 0$, for all i on some subset $\mathcal{U} \subset \mathbb{R}^d$.

- $s = 1 \implies f$ locally concave on $\mathcal{U}$
- $s = 0 \implies f$ locally convex on $\mathcal{U}$

Concavity and convexity ↕️ II

Concave bump 🙃

We define the concave bump as

$$\mathcal{B}^{\lambda_1} = \{\mathrm{x} \in \mathbb{R}^d : \lambda_1[f](\mathrm{x}) \leq 0\} .$$

Convex dip 🙂

We define the convex dip as

$$\mathcal{B}^{\lambda_d} = \{\mathrm{x} \in \mathbb{R}^d : \lambda_d[f](\mathrm{x}) \geq 0\} .$$

Strong and weak modal regions 💪

- Concave bumps *typically* delineate local pdf modes.
- If not, they point out incipient modal regions.

Gradient divergence 🧭 I

Imagine the pdf graph as a landscape... 🏞️

- Concave bumps originate near mountain peaks.
- Gradient divergence allows the discovery of entire mountain chains.

Hessian & shape operator coincide in eigenvalue signature 📝

- The eigenvalues of the shape operator are known as principal curvatures. They reflect how the normal vector changes.
- The Hessian eigenvalues have the same interpretation but with a non-standardized normal vector field.

Combining eigenvalue signs ➕

- Concavity & convexity require that all curvatures have the same sign.
- The gradient divergence sums all the curvatures.

Gradient divergence 🧭 II

Mean curvature bump 🌍

Consider $\bar{\nabla} f = \nabla f / \sqrt{1 + \|\nabla f\|^2}$. We define the mean curvature bump as

$$\mathcal{B}^{\bar{\nabla}} = \{\mathrm{x} \in \mathbb{R}^d : \operatorname{div}(\bar{\nabla} f)(\mathrm{x}) \leq 0\}.$$

Laplacian bump 🗺️

Consider $\Delta f = \mathrm{tr}(\mathrm{D}^2 f) = \sum_{i=1}^{d} \partial^2 f / \partial x_i^2 = \operatorname{div}(\nabla f) \approx \operatorname{div}(\bar{\nabla} f)$. We define the Laplacian bump as

$$\mathcal{B}^{\Delta} = \{\mathrm{x} \in \mathbb{R}^d : \Delta f(\mathrm{x}) \leq 0\}.$$

Improved tractability 👍

The Laplacian is a second-order linear differential operator of f.

Estimation: from population to sample 🛠️

Kernel density estimator (KDE) plug-in 🔌

Let $\mathrm{X}_1, \ldots, \mathrm{X}_n \sim f$ be an i.i.d. sample. Estimate $\mathcal{B}^\phi$ with

$$\tilde{\mathcal{B}}^\phi_{n,h} = \{\mathrm{x} \in \mathbb{R}^d : (-1)^s \phi[\hat{f}_{n,h}](\mathrm{x}) \geq 0\},$$

where

$$\hat{f}_{n,h}(\mathrm{x}) = \frac{1}{n} \sum_{i=1}^n K_h(\mathrm{x} - \mathrm{X}_i) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{\mathrm{x} - \mathrm{X}_i}{h}\right)$$

is the KDE with kernel K and bandwidth h.

Selecting K and h 🧐

- Our results are compatible with a *universally* preferred Gaussian K.
- For h, target $r = 2$ in one of the standard selectors.

Consistency 🧲

Theorem (Consistency of curvature bump boundaries) ♾️

Assume:

- $\phi[f]$ and $\phi[\hat{f}_{n,h}]$ are sufficiently smooth.
- ϕ depends on partial derivatives of f up to ℓ-th order.
- $\hat{f}_{n,h}$ is based on the Gaussian kernel.
- $h \equiv h_n \to 0$ and $n^{-1}h^{-(d+2\ell+2)} \log n \to 0$ as $n \to \infty$.
- There exist λ, δ such that $\|\nabla\phi[f](\mathrm{x})\| > \lambda$, for all $\mathrm{x} \in \partial\mathcal{B}^\phi \oplus \delta$.

Then,

$$\mathrm{Haus}(\partial\tilde{\mathcal{B}}^\phi_{n,h}, \partial\mathcal{B}^\phi) = O\left(h^2 + \sqrt{\frac{\log n}{nh^{d+2\ell}}}\right) \xrightarrow[n\to\infty]{a.s.} 0\,.$$

This result uses the stability theorem by Chen (2022).

Inference 🗜 I

Simplifying assumptions ⚠

- We will target $\bar{f}_h : \mathbb{R}^d \to [0, \infty)$, given by $\bar{f}_h(\mathrm{x}) = \mathbb{E}[\hat{f}_{n,h}(\mathrm{x})]$.
- We will focus on functionals ϕ deriving from the pdf Hessian $\mathrm{D}^2 f$.

Inference scheme 🥪

Given $\alpha \in (0, 1)$, a path for narrowing down a $(1 - \alpha)$-level confidence region for $\mathcal{B}_h^{\phi}$ is constructing two sets

$$\begin{aligned} \bar{\mathcal{B}}^{\phi}_{n,h}(\zeta^{\alpha}_{n,h}) &= \{x \in \Theta : (-1)^s \phi[\hat{f}_{n,h}](\mathrm{x}) \geq -\zeta^{\alpha}_{n,h}\} \\ \underline{\mathcal{B}}^{\phi}_{n,h}(\zeta^{\alpha}_{n,h}) &= \{x \in \Theta : (-1)^s \phi[\hat{f}_{n,h}](\mathrm{x}) \geq \zeta^{\alpha}_{n,h}\} \end{aligned},$$

for some margin $\zeta^{\alpha}_{n,h} \in [0, \infty)$, so that $\underline{\mathcal{B}}^{\phi}_{n,h}(\zeta^{\alpha}_{n,h}) \subset \tilde{\mathcal{B}}^{\phi}_{n,h} \subset \bar{\mathcal{B}}^{\phi}_{n,h}(\zeta^{\alpha}_{n,h})$.

Inference II

Asymptotic goal

Find a sequence $\{\tilde{\zeta}^{\alpha}_{n,h}\}_{n=1}^{\infty}$ such that, as $n \to \infty$,

$$\begin{cases} \mathbb{P}\left(\underline{\mathcal{B}}^{\phi}_{n,h}(\tilde{\zeta}^{\alpha}_{n,h}) \subset \mathcal{B}^{\phi}_{h} \subset \bar{\mathcal{B}}^{\phi}_{n,h}(\tilde{\zeta}^{\alpha}_{n,h})\right) \geq 1 - \alpha + o(1) \\ \tilde{\zeta}^{\alpha}_{n,h} = o(1) \end{cases}. \tag{1}$$

Bootstrap (Chen et al., 2017)

The key point is approximating the stochastic errors for second-order linear differential operators $\mathcal{D}$, $\mathcal{E}_{n,h}[\mathcal{D}] = \sup_{\mathrm{x}\in\Theta} |\mathcal{D}\hat{f}_{n,h}(\mathrm{x}) - \mathcal{D}\bar{f}_h(\mathrm{x})|$, with bootstrap estimates $\mathcal{E}^*_{n,h}[\mathcal{D}|\mathfrak{X}_n] = \sup_{\mathrm{x}\in\Theta} |\mathcal{D}\hat{f}^*_{n,h}(\mathrm{x}|\mathfrak{X}_n) - \mathcal{D}\hat{f}_{n,h}(\mathrm{x}|\mathfrak{X}_n)|$.

Inference III

Theorem (Inference for Laplacian bumps)

Let us fix $h \in (0,1)$. Let $\mathcal{E}^*_{n,h}[\cdot|\mathfrak{X}_n]$ be based on a kernel $K \in \mathcal{C}^2(\mathbb{R}^d)$ with bounded second derivatives. Suppose that the class

$$\mathcal{K} = \left\{ \mathrm{y} \in \mathbb{R}^d \mapsto \partial^{\boldsymbol{\beta}} K \left(\frac{\mathrm{x} - \mathrm{y}}{h} \right) : \mathrm{x} \in \Theta, h > 0, \boldsymbol{\beta} \in \mathbb{Z}^d_+, |\boldsymbol{\beta}| = 2 \right\} \quad (2)$$

is Vapnik–Chervonenkis (VC)-type. For any $\alpha \in (0,1)$, define the margin $\tilde{\zeta}^{\alpha}_{n,h} = \mathcal{Q}_{1-\alpha}\{\mathcal{E}^*_{n,h}[\Delta|\mathfrak{X}_n]\}$.

Then, for all $\alpha \in (0,1)$, the asymptotic validity of (1) holds a.s. for the *smoothed* version of the Laplacian bump.

Inference IV

Theorem (Inference for concave bumps & convex dips)

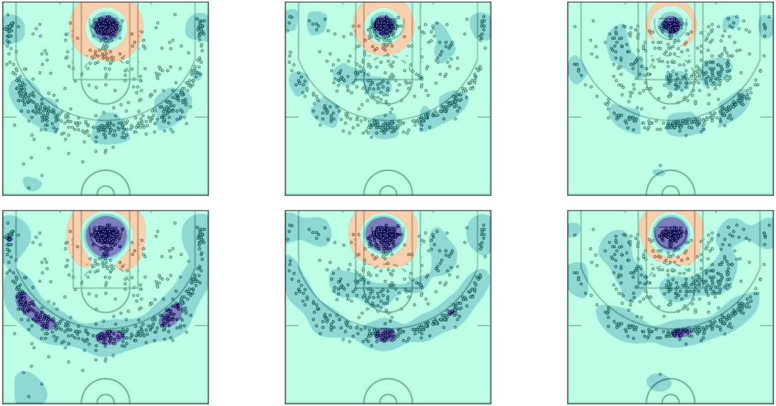
Let us fix $h \in (0,1)$. Let $\mathcal{E}_{n,h}[\cdot]$ and $\mathcal{E}^*_{n,h}[\cdot|\mathfrak{X}_n]$ be based on the same kernel $K \in \mathcal{C}^2(\mathbb{R}^d)$ with bounded second derivatives. Suppose that $\mathcal{K}$ in (2) is VC-type. For any $\alpha \in (0,1)$, define the margin

$$\tilde{\zeta}^{\alpha}_{n,h} = \sum_{i=1}^{d} \sum_{j=1}^{d} \mathrm{TVaR}_{1-\alpha} \left\{ \mathcal{E}^*_{n,h}[\mathrm{D}_{ij}|\mathfrak{X}_n] \right\} ,$$

where D_{ij} is a second-order partial derivative.

Under some extra mild regularity assumptions, for all $\alpha \in (0,1)$, the asymptotic validity of (1) holds a.s. for the *smoothed* version of the concave bump and the convex dip.

NBA application



(a) Stephen Curry (b) James Harden (c) Kevin Durant

Outline

[3] arXiv:2307.05825 [pdf, other]

Bayesian taut splines for estimating the number of modes

José E. Chacón, Javier Fernández Serrano

Comments: 21 pages, 8 figures (manuscript) + 22 pages, 17 figures (supplementary material)

Journal-ref: Computational Statistics and Data Analysis 196 (2024) 107961

Subjects: **Methodology (stat.ME)**; Machine Learning (cs.LG); Statistics Theory (math.ST); Machine Learning (stat.ML)

The problem of multimodality 🎢

Modes 🔝

- Local maxima of a probability density function (pdf).
- Two main *tasks*: locating them & counting them.
- Usage in exploratory data analysis, pointing out new phenomena.

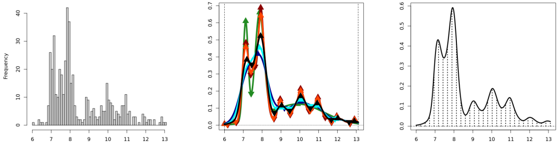
The number of modes (NoM) 🔢

- Representative of the pdf's complexity.
- It can be viewed as the number of existing subpopulations.

Theoretically, a complicated problem ⛔

The NoM k is impossible to bound from above (with confidence) if the sample is *genuinely nonparametric* (Donoho, 1988).

Classic problem, new *Bayesian taut spline* (BTS) ®



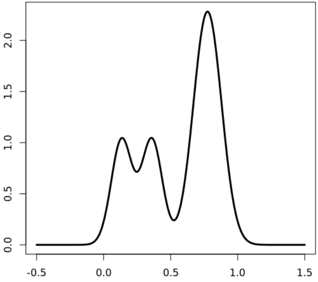
(a) Hidalgo sample (b) Several pdfs (c) BTS solution

Three overlooked aspects of the NoM & three pillars for BTS 🏛

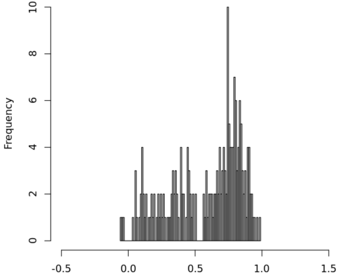
- Structure: Balancing *data fitting* and *model complexity*.
- Bayesian inference: *Subjective* and *uncertain* nature.
- Holism: Dual *local* and *global* nature.

Reminiscent of the penalized likelihood by Good and Gaskins (1980).

Let us illustrate it with an example



(a) Theoretical pdf



(b) Random sample ($n = 200$)

1. Exploration 🔭

Mixing kernel density estimator (KDE) & compositional spline (CS) ☯️

- The KDE is a flexible, nonparametric tool for pdf estimation.
- CS are d-dimensional pdf vectors, with operations $\oplus$ & $\odot$.

"KDEs are a *scaffold* for CSs & CSs are a *harness* for KDEs" 🏗️

1. Given a bandwidth $h > 0$, build the KDE $\hat{f}_h$ from data $\mathcal{D}$.
2. Fit a CS $[\hat{f}_h]_\alpha$ to $\hat{f}_h$ via *least squares* with curvature penalty $\alpha \in (0, 1)$.
3. Define a pdf for $\mathcal{D}$ conditioning on (h, α) as $\Pr(\cdot|h, \alpha) = [\hat{f}_h]_\alpha$.

Two-dimensional exploration with *proxy* parameters 🕹️

After imposing a prior $\Pr(h, \alpha)$, exploring the *CS space* and the subsequent *modes* comes down to evaluating the posterior $\Pr(h, \alpha|\mathcal{D})$.

Mode exploration by simulation 🎲

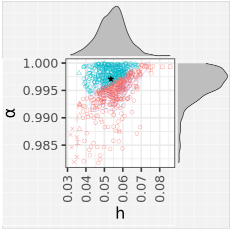
The *raw* BTS estimator 🥚

Via Markov chain Monte Carlo (MCMC) simulation, secure a posterior sample $\mathcal{S} = \{(h_i, \alpha_i)\}_{i=1}^{\nu}$ from $\Pr(h, \alpha | \mathcal{D})$. Then, define

$$\hat{k}_{\mathrm{BTS},0} = \underset{k \in \mathbb{N}}{\arg\max} \left| \{ i \in \{1, \ldots, \nu\} : \Pr(\cdot | h = h_i, \alpha = \alpha_i) \text{ has } k \text{ modes} \} \right| .$$

Visualization tool 👁

- NoM of $[\hat{f}_h]_\alpha$ (colour) vs NoM of $\hat{f}_h$ (shape).
- The counts can be normalized to assess probabilities.



2. Analysis 🔬

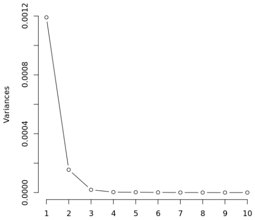
Issues to address 🚧

- Robust approximations require long, time-consuming MCMC chains.
- Having one parameter, *mode trees* are easier to interpret.

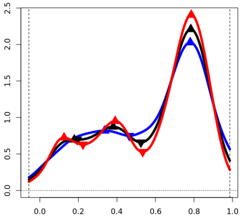
Simplicial functional principal component analysis (SFPCA) ⚗️

- Let $\boldsymbol{\theta}_i \in \mathbb{R}^d$ be the CS coordinates underlying $(h_i, \alpha_i) \in \mathcal{S}$.
- SFPCA works on $\{\boldsymbol{\theta}_i\}_{i=1}^{\nu}$, similarly to vanilla PCA.
- Orthonormal CS axes $\mathrm{PC}_1, \ldots, \mathrm{PC}_d$ with variances $\lambda_1 \geq \cdots \geq \lambda_d$.
- Define a pdf model $\Pr(\cdot|\delta) = \mu \oplus \delta \odot \sigma$, where μ is a sample average CS pdf with coordinates $(1/\nu) \sum_{i=1}^{\nu} \boldsymbol{\theta}_i$, and $\sigma = \sqrt{\lambda_1} \odot \mathrm{PC}_1$.
- We can empirically build a sensible *support* $\Delta = [\delta_{\min}, \delta_{\max}]$ for δ.
- Jeffreys prior $\Pr(\delta) \propto \sqrt{\mathrm{Var}(\log \sigma(X_\delta))}$, where $X_\delta \sim \Pr(\cdot|\delta)$.

SFPCA magic ✨



(a) Variances scree plot



(b) Representative pdfs

Visualization tool 👁

- The sudden drop is characteristic, justifying keeping *one dimension*.
- Scatter plot summarized through μ, $\mu \oplus \delta_{\min} \odot \sigma$ and $\mu \oplus \delta_{\max} \odot \sigma$.

3. Selection ⚖️

The *processed* BTS estimator 🥚→🍳

- Estimate the NoM via Bayesian model selection:

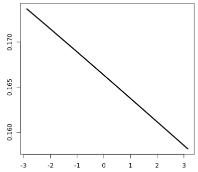
$$\hat{k}_{\mathrm{BTS},1} = \arg\max_{k \in \mathbb{N}} \Pr(k|\mathcal{D}),$$

where, letting $\Delta_k = \{\delta \in \Delta : \Pr(\cdot|\delta) \text{ has } k \text{ modes}\}$,

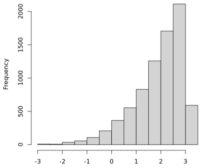
$$\underbrace{\boxed{\Pr(k|\mathcal{D})}}_{\text{Posterior}} \propto \underbrace{\boxed{\frac{\int_{\Delta_k} \Pr(\delta|\mathcal{D})\ d\delta}{\int_{\Delta_k} \Pr(\delta)\ d\delta}}}_{\text{Bayes factor}} \times \underbrace{\boxed{\Pr(k)}}_{\text{Prior}}.$$

- Alternatives for the prior: (i) $\Pr(k) = \int_{\Delta_k} \Pr(\delta)\ d\delta$, (ii) $\Pr(k) \propto 1$, and (iii) $\Pr(k)$ derived from the $\mathcal{S}$ obtained in the exploration phase.

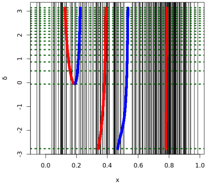
Selection *debugging*



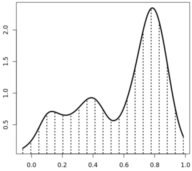
(a) Jeffreys prior $\mathrm{Pr}(\delta)$



(b) Sample from posterior $\mathrm{Pr}(\delta|\mathcal{D})$



(c) Mode tree



(d) Posterior median CS ($k = 3$)

4. Testing

The NoM in BTS so far: *global* views

- It measured model complexity in the *exploration* phase.
- It encompassed a range of parameter values in the *selection* phase.

Testing mode significance *locally*

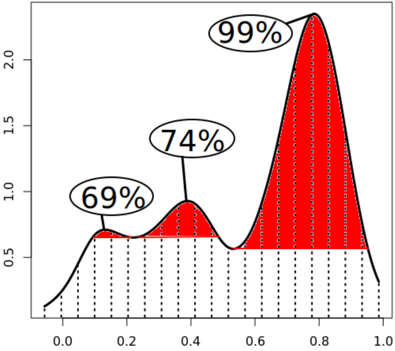
- Let $\tilde{f}_k$ be the median k-modal pdf selected with probability $\Pr(k|\mathcal{D})$.
- For the i-th mode of $\tilde{f}_k$, consider its *excess mass region** $\mathcal{M}_{k,i}$.
- Define $\Pr(x|\tau, k, i) \propto (\tau \odot \tilde{f}_k)(x) \cdot \mathbb{1}_{\mathcal{M}_{k,i}}(x)$, where $\tau \in [0, \infty)$.
- $[\mathcal{H}_0 : \tau = 0 \equiv$ "there *isn't* mode"$]$ vs $[\mathcal{H}_1 : \tau > 0 \equiv$ "there *is* mode"$]$
- Imposing $\tau|k, i \sim \text{Exponential}(\lambda = 1)$, we have

$$\underbrace{\boxed{\Pr(\mathcal{H}_1|\mathcal{D}, k, i)}}_{\text{Posterior}} = \Big[1 + \underbrace{\boxed{\Pr(\tau = 0|\mathcal{D}, k, i)}}_{\text{Bayes factor}} \Big/ \underbrace{\boxed{\text{Odds}(\mathcal{H}_1|k, i)}}_{\text{Prior}}\Big]^{-1}.$$

*Please, wait for the next slide...

Excess mass regions

Posterior significance probabilities† based on nearby data:



† Prior odds equal to one.

Almost finished 🏁

Mode significance & thresholding 🛑

- Kass and Raftery (1995) offer a scale to interpret the probabilities.
- Setting a decision *threshold* might be too rigid & arbitrary.
- It *overlooks* the global results that inform the local analysis.

The *refined* BTS estimator 🥚→🍳→🧂

- Let $\mathrm{Pr}(\mathcal{H}_1|k,\mathcal{D})$ be the *overall* significance of the k hypothesis.
- E.g., the harmonic mean: $\mathrm{Pr}(\mathcal{H}_1|k,\mathcal{D}) = k/\sum_{i=1}^{k}\mathrm{Pr}(\mathcal{H}_1|\mathcal{D},k,i)^{-1}$.
- Then, "add" probabilities: $\mathrm{Pr}(k|\mathcal{H}_1,\mathcal{D}) \propto \mathrm{Pr}(\mathcal{H}_1|k,\mathcal{D}) \times \mathrm{Pr}(k|\mathcal{D})$.
- Finally, estimate the NoM as

$$\hat{k}_{\mathrm{BTS},2} = \arg\max_{k\in\mathbb{N}} \mathrm{Pr}(k|\mathcal{H}_1,\mathcal{D}) .$$

Results 💯

Simulation study 🧪

- Target: Prediction *accuracy* of the NoM (0-1 loss).
- Methods: *Refined* BTS ($\Pr(k) \propto 1$) plus KDEs, modality tests, etc.
- Test-beds: $T = 5$ mixtures from Ameijeiras-Alonso (2017).
- Sample sizes: $S = 3$ scenarios ($n = 100$, $n = 400$ & $n = 1600$).
- Significance: 200 replications, McNemar's test & ranking aggregation.

Global ranking 🥇

Aggregating results from $T \times S = 15$ sampling configurations:

PI0	PI1	PI2	SCV	STE	LSCV0	LSCV1	LSCV2	GM	TS	SI	FM	EIG	BTS2U
1	2	5	**1**	**1**	2	2	3	**1**	6	6	4	6	**1**

Remarks ✏️

PI0	PI1	PI2	SCV	STE	LSCV0	LSCV1	LSCV2	GM	TS	SI	FM	EIG	BTS2U
1	2	5	**1**	**1**	2	2	3	**1**	6	6	4	6	**1**

Counterintuitive results 🤨

- Traditional methods for modality *fall deeply* in the rankings:
  - Hypothesis tests underperform if sequentially employed.
- Generic KDE-based estimators typically work *better*:
  - Our overall recommendation for estimating the NoM is `PI0`.
  - BTS offers valuable resources with similar accuracy to `PI0`.

Room for improvement 📈

- No method performs uniformly better across all settings.
- BTS might improve using, e.g., bootstrapping or ensembles.

Outline

1 Introduction

2 Curvature

3 Number of modes

4 Center of symmetry

5 Conclusions

[2] arXiv:2406.08241 [pdf, ps, other]

Mode-based estimation of the center of symmetry

José E. Chacón, Javier Fernández Serrano

Comments: 27 pages, 8 figures

Journal-ref: Annals of the Institute of Statistical Mathematics 77 (2025) 685-717

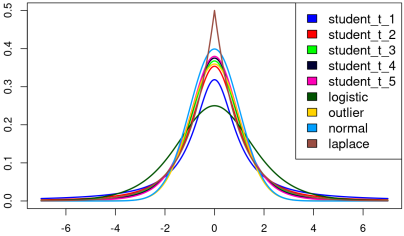
Subjects: **Methodology (stat.ME)**; Statistics Theory (math.ST)

The classic triad of centrality measures 🏛

- Let X be a random variable with pdf f.
- Let $X_1, \ldots, X_n$ be an i.i.d. random sample from f.
- Let us call, whenever defined:
  - $\mu = \mathbb{E}[X] \equiv$ "Mean of X" (assuming finite first moment)
  - $M \equiv$ "Median of X"
  - $\mathrm{m} = \arg\max_{x\in\mathbb{R}} f(x) \equiv$ "Mode of X" (assuming a unique maximizer)
- Let $\theta \in \mathbb{R}$ be the center of symmetry of X (or f), satisfying:
  - $f(\theta + x) = f(\theta - x)$ for all $x \in \mathbb{R}$.
  - $f_0(x) = f(\theta + x)$ is a strictly decreasing function of $|x|$ over its support.
- Under the previous assumptions, we have $\mu = M = \mathrm{m} = \theta$, so there are several estimators for θ based on $X_1, \ldots, X_n$:
  - The sample mean $\hat{\mu}_n$ (well-known).
  - The sample median $\hat{M}_n$ (well-known).
  - But what about the "sample mode" $\hat{\mathrm{m}}_{n,h}$? (far less known)

Symmetric, unimodal pdfs in action

- They naturally arise when modeling physical measurements without systematic errors (also called *biases*).
- If $X \sim f$, then $X = \theta + \varepsilon$, where $\varepsilon \sim f_0$ is a random error such that:
  - ε and $-\varepsilon$ have the same distribution (ε is symmetric about 0).
  - Larger values of $|\varepsilon|$ are increasingly less likely (ε is unimodal about 0).
- Many household parametric pdfs are symmetric and unimodal.



Test-bed pdfs in the simulation study.

A *plug-in* estimator: the kernel mode estimator (KME)

- The KDE is a classic estimator of f:

$$\hat{f}_{n,h}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right).$$

- The kernel mode estimator (KME) $\hat{\mathrm{m}}_{n,h}$ of the mode m of f is any random variable satisfying $\hat{f}_{n,h}(\hat{\mathrm{m}}_{n,h}) = \max_{x\in\mathbb{R}} \hat{f}_{n,h}(x)$. Ties are usually broken uniformly at random.
- For a fixed $h > 0$, the population target of $\hat{\mathrm{m}}_{n,h}$, as $n \to \infty$, is the value $\bar{\mathrm{m}}_h \in \mathbb{R}$ that maximizes the smoothed pdf $\bar{f}_h(x) = \mathbb{E}[\hat{f}_{n,h}(x)]$.
- In general, $\bar{\mathrm{m}}_h \neq \mathrm{m}$. Only as $h \to 0^+$ the bias goes to zero.
- However, if f and K are unimodal and symmetric, then, for all $h > 0$, $\bar{\mathrm{m}}_h = \mathrm{m}$.

M-estimators: a theory to rule them all 

- M-estimators of a location parameter $\theta \in \mathbb{R}$ of f are defined as sample random variables $\hat{\theta}_n$ minimizing $\sum_{i=1}^{n} \rho(X_i - \hat{\theta}_n)$, for some fixed function $\rho : \mathbb{R} \to \mathbb{R}$.
- Alternatively, if $\psi = \rho'$, an M-estimator $\hat{\theta}_n$ is one of the solutions of $\sum_{i=1}^{n} \psi(X_i - \hat{\theta}_n) = 0$. Care should be taken to avoid "bad solutions".
- M-estimators subsume many estimators of the center of symmetry:
  - $\hat{\theta}_n = \hat{\mu}_n$ when $\rho(x) = x^2$.
  - $\hat{\theta}_n = \hat{M}_n$ when $\rho(x) = |x|$.
  - $\hat{\theta}_n = \hat{\mathrm{m}}_{n,h}$ when $\rho(x) = -K_h(x) = -K(x/h)/h$.
- A relevant subclass is redescending M-estimators, which satisfy $\psi(x) \to 0^+$ as $|x| \to \infty$. For the most common kernels K, particularly kernels with compact support, KMEs are redescending.
- Redescending M-estimators are robust against outliers.

Efficiency of KMEs: asymptotic variance 🪗

Definition (Bell-shaped kernel) 🔔

Consider a twice continuously differentiable pdf kernel K that is symmetric and unimodal about zero, with bounded K' and K''. We say that K is bell-shaped if K has exactly two inflection points at $\pm a$, for some $a > 0$. That is, $\pm a$ are the only points where K'' changes its sign.

Theorem (Asymptotic variance of KMEs) ⚡

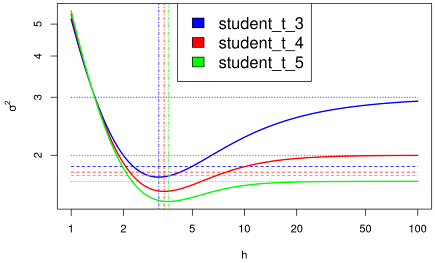
Let K be bell-shaped. Then, for any $h > 0$, $\sqrt{n}(\hat{\mathrm{m}}_{n,h} - \theta) \rightsquigarrow \mathcal{N}(0, \sigma^2_{f,K,h})$ as $n \to \infty$, with variance

$$\sigma^2_{f,K,h} = \frac{\mathbb{E}[K_h'(X-\theta)^2]}{\mathbb{E}[K_h''(X-\theta)]^2} = \frac{\int_{-\infty}^{\infty} K_h'(x-\theta)^2 f(x)\ dx}{[\int_{-\infty}^{\infty} K_h''(x-\theta) f(x)\ dx]^2} = \frac{[(K_h')^2 * f](\theta)}{(K_h'' * f)(\theta)^2}.$$

Efficiency of KMEs: limit behavior ↔️

Theorem (Limit behavior of the asymptotic variance of KMEs) ♊

1. Let f be twice continuously differentiable with bounded f' and f'' integrable. Let K be such that $\lim_{x\to\infty} xK(x) = 0$ and $\mathcal{R}(K') = \int_{-\infty}^{\infty} K'(x)^2\, dx$ is finite. Then, $\lim_{h\to 0^+} \sigma^2_{f,K,h} = \infty$.
2. If f has finite variance σ^2, and $K''(0) < 0$, then $\lim_{h\to\infty} \sigma^2_{f,K,h} = \sigma^2$.



Efficiency of KMEs: existence of an optimal h 🪟

Definition (Regularly varying function) 🐌

A function $f : [x_0, \infty) \to (0, \infty)$, for $x_0 > 0$, is *regularly varying* with index $\alpha \in \mathbb{R}$ if, for all $\lambda > 0$, $\lim_{x\to\infty} f(\lambda x)/f(x) = \lambda^\alpha$. If $\alpha = 0$, f is said to be *slowly varying*. If f is regularly varying with index α, then we have the universal representation $f(x) = x^\alpha L(x)$, where L is slowly varying.

Theorem (Existence of an optimal h) 🦄

Assume that $f_0(x) = f(\theta + x)$ is regularly varying with $\alpha < -3$ and bounded L satisfying $\lim_{h\to\infty} L(h) = \ell > 0$. Suppose that K satisfies: (i) $I_{K,\alpha} = \int_0^\infty x^\alpha [|K''(0)|^{-2} K'(x)^2 - x^2]\, dx$ is finite and negative, and (ii) $\int_0^\infty x^\alpha [1 + |K''(0)|^{-1} K''(x)]\, dx$ is finite. Then, for every sufficiently large $h > 0$, we have $\sigma^2_{f,K,h} < \sigma^2$.

Efficiency of KMEs: a 🆕 parametric family of kernels

Theorem (Compliant family of kernels) 👌

Define the family of bump-like functions B_β, with $\beta > 0$,

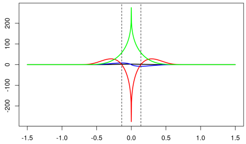
$$B_\beta(x) = \begin{cases} e^{-1/(1-|x|^\beta)}, & \text{if } |x| < 1 \\ 0, & \text{otherwise} \end{cases}.$$

Then, consider the family of pdf kernels K_β whose derivative is given by $K'_\beta(x) \propto -xB_\beta(x)$. If $|\alpha + 1| < \beta$, then K_β is compliant.

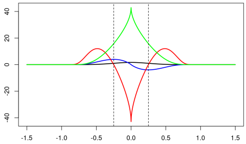
Limit behavior of the family 🪆

K_β asymptotically approaches, as $\beta \to \infty$, the Epanechnikov kernel, which is optimal for Gaussian data among compactly supported kernels.

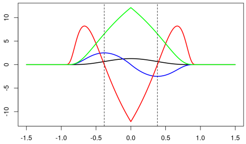
The flexibility of the new family K_β 🎭



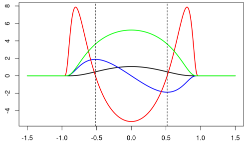
(a) $\beta = 1/4$



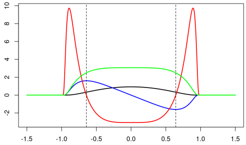
(b) $\beta = 1/2$



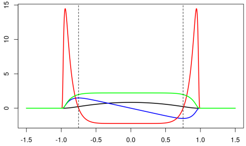
(c) $\beta = 1$



(d) $\beta = 2$



(e) $\beta = 4$



(f) $\beta = 8$

Computation of the KME ⚙️

Iterative reweighting (also known as *mean shift*) 🔗

Let us assume that a choice of β and h has been made. Define

$$\mathcal{Z}_\beta(x) = \log B_\beta(x) = \begin{cases} -(1-|x|^\beta)^{-1}, & \text{if } |x| < 1 \\ -\infty, & \text{otherwise} \end{cases} .$$

Starting from $\hat{\mathrm{m}}_0 = \hat{M}_n$, the transition from $\hat{\mathrm{m}}_k$ to $\hat{\mathrm{m}}_{k+1}$, for $k \geq 0$, is given by $\hat{\mathrm{m}}_{k+1} = \sum_{i=1}^n \mathbf{w}_i^{(k)} X_i$, where $\mathbf{w}^{(k)}$ depends on $\hat{\mathrm{m}}_k$ through

$$\mathbf{w}^{(k)} = \operatorname{softmax}\left[\mathcal{Z}_\beta\left(\frac{X_1 - \hat{\mathrm{m}}_k}{h}\right), \ldots, \mathcal{Z}_\beta\left(\frac{X_n - \hat{\mathrm{m}}_k}{h}\right)\right] .$$

Iterations stop when $|\hat{\mathrm{m}}_{k+1} - \hat{\mathrm{m}}_k| \leq \varepsilon h$, for some small $\varepsilon > 0$.

Optimization of the KME I

Joint optimization of β and h ✌

Let us denote $\sigma^2_{f,\beta,h}$ the asymptotic variance of the KME with kernel K_β. The optimal parameters are $(\beta^*, h^*) = \arg\min_{\beta>0, h>0} \sigma^2_{f,\beta,h}$.

Let us define $Z_h = (X - \theta)/h$, and, setting $\psi(x;\beta) = -xB_\beta(x)$, define the auxiliary functions $\Psi_1(x;\beta) = \psi(x;\beta)^2$ and $\Psi_2(x;\beta) = \partial\psi(x;\beta)/\partial x$. Then,

$$\sigma^2_{f,\beta,h} = h^2 \, \frac{\mathbb{E}[\Psi_1(Z_h;\beta)]}{\mathbb{E}[\Psi_2(Z_h;\beta)]^2}\,,$$

where, for $\eta \in \{1,2\}$,

$$\mathbb{E}[\Psi_\eta(Z_h;\beta)] = 2h \int_0^1 \Psi_\eta(x;\beta)\; f_0(hx)\; dx\,.$$

Optimization of the KME II

Estimation of f_0 from data

Take $f_0(x) \approx \tilde{f}_0(x; g) = \hat{f}^*_{n,g}(\hat{M}_n + x)$, where

$$\hat{f}^*_{n,g}(x) = \frac{\hat{f}_{n,g}(x) + \hat{f}_{n,g}(2\hat{M}_n - x)}{2},$$

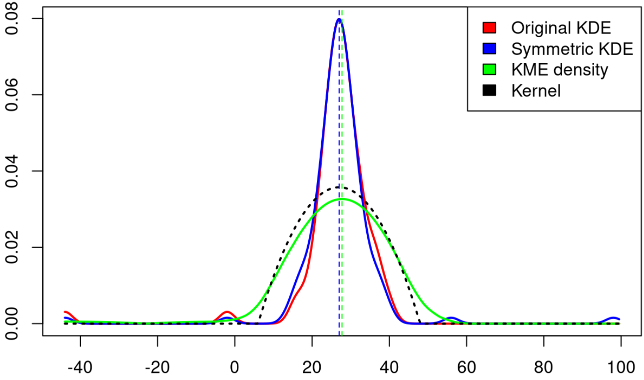
and $\hat{f}_{n,g}$ is a Gaussian KDE with g chosen via Silverman's *rule of thumb*.

The symmetric KDE $\hat{f}^*_{n,g}$ can be seen as the KDE with an *augmented* sample $(X_1, \ldots, X_n, 2\hat{M}_n - X_1, \ldots, 2\hat{M}_n - X_n)$.

RE: Prof. Berrendero's inquiry during the thesis *pre-defense*

Numerical limitations aside, *this* KME is shift- and scale-equivariant.

Case study: estimating the speed of light 🔦



The sample estimators are $\hat{\mu}_n = 26.2$, $\hat{M}_n = 27$, and $\hat{\mathrm{m}}_{n,h} = 27.75$.
The optimal parameters are $(\beta^*, h^*) = (97.03537, 21.23523)$.

Simulation study: main conclusions ✅

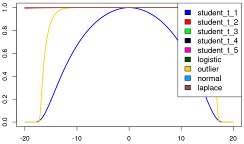
Head-to-head results summary (across 27 sampling configurations) 🏆

- KME vs. classic estimators:
  - Sample mean: 25/27 (23/27 statistically significant)
  - Sample median: 24/27 (24/27)
- KME vs. (non-adaptive) redescending M-estimators:
  - Tukey's biweight: 22/27 (16/27)
  - Andrew's sine: 23/27 (23/27)
- KME vs. (adaptive) L-estimators:
  - Trimmed mean: 18/27 (13/27)
  - Winsorized mean: 24/27 (17/27)

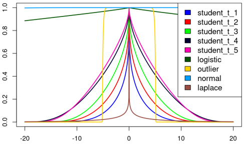
A crucial remark 👀

Optimizing the shape β is as important as optimizing the bandwidth h.

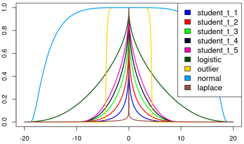
Simulation study: *debugging* $B_{\beta^*}(\cdot/h^*)$ 🔍



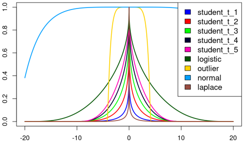
(a) $n = 100$



(b) $n = 1,000$



(c) $n = 10,000$



(d) True pdfs

Outline

Conclusions 💼 I

Density features & exploratory data analysis 📊

- Summarization questions: what?, how many?, center?
- The different faces of outliers: meaningful, spurious & inefficient.
- Local features: Hyndman (1996), Donoho (1988) & robustness.

Starring the kernel density estimator ⭐

- Flexibility for a diffuse problem like BH.
- Tempered with the structure of splines for NoM estimation.
- Universality reaffirmed with its connection to M-estimators.

Conclusions 💼 II

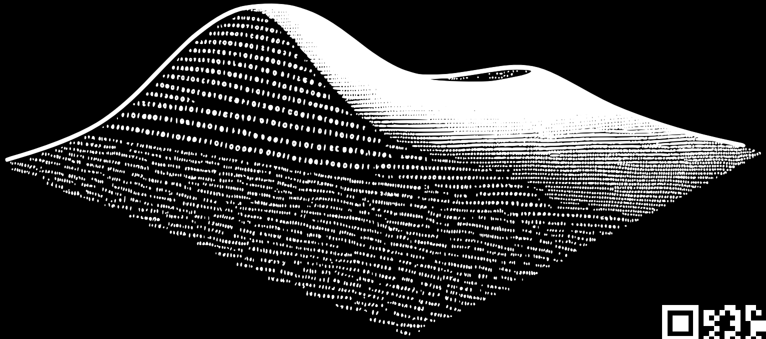
Contributions 🎁

- Connections: geometry, Bayesian statistics, robust statistics.
- Innovations: splines, likelihood-based estimation, kernel shape.
- Computing: simulations, bootstrap, MCMC (Berrendero, 2015).
- Multifaceted: logical/disruptive step forward vs filling a gap.

Future work 🤠

- Conservatism of inference for BH (Mammen and Polonik, 2013).
- Extending BTS to circular/bivariate data.
- The suggestions by Hino (2025) and Pardo-Fernández (2025).

The End! Thank you for coming! :)



References 📚 I

- Ameijeiras-Alonso, J. (2017). Assessing simplifying hypotheses in density estimation. PhD thesis. Universidade de Santiago de Compostela.
- Berrendero, J. R. (2015). Simulación e inferencia estadística. *La Gaceta de la RSME* **18**, 45–65.
- Casa, A. (2019). Climbing modes and exploring mixtures: a journey in density-based clustering. PhD thesis. Università degli studi di Padova.
- Chacón, J. E. (2020). The modal age of statistics. *International Statistical Review* **88**, 122–41.
- Chacón, J. E. and Duong, T. (2018). *Multivariate kernel smoothing and its applications*. Chapman and Hall/CRC.
- Chacón, J. E. and Fernández Serrano, J. (2023). Bump hunting through density curvature features. *TEST* **32**, 1251–75.

References 📚 II

Chacón, J. E. and Fernández Serrano, J. (2024). Bayesian taut splines for estimating the number of modes. *Computational Statistics & Data Analysis* **196**, 107961.

Chacón, J. E. and Fernández Serrano, J. (2025a). Mode-based estimation of the center of symmetry. *Annals of the Institute of Statistical Mathematics* **77**, 685–717.

Chacón, J. E. and Fernández Serrano, J. (2025b). Rejoinder to the discussion of "Mode-based estimation of the center of symmetry". *Annals of the Institute of Statistical Mathematics* **77**, 727–30.

Chaudhuri, P. and Marron, J. S. (1999). SiZer for exploration of structures in curves. *Journal of the American Statistical Association* **94**, 807–23.

Chaudhuri, P. and Marron, J. S. (2002). Curvature vs. slope inference for features in nonparametric curve estimates.

References 📚 III

Chen, Y.-C. (2022). Solution manifold and its statistical applications. *Electronic Journal of Statistics* **16**, 408–50.

Chen, Y.-C., Genovese, C. R., and Wasserman, L. (2017). Density level sets: asymptotics, inference, and visualization. *Journal of the American Statistical Association* **112**, 1684–96.

Donoho, D. L. (1988). One-sided inference about functionals of a density. *The Annals of Statistics* **16**, 1390–420.

Duong, T., Cowling, A., Koch, I., and Wand, M. P. (2008). Feature significance for multivariate kernel density estimation. *Computational Statistics & Data Analysis* **52**, 4225–42.

Godtliebsen, F., Marron, J. S., and Chaudhuri, P. (2002). Significance in scale space for bivariate density estimation. *Journal of Computational and Graphical Statistics* **11**, 1–21.

References 📚 IV

- Good, I. J. and Gaskins, R. A. (1980). Density estimation and bump-hunting by the penalized likelihood method exemplified by scattering and meteorite data. *Journal of the American Statistical Association* **75**, 42–56.
- Hino, H. (2025). Discussion of "Mode-based estimation of the center of symmetry". *Annals of the Institute of Statistical Mathematics* **77**, 719–21.
- Hyndman, R. J. (1996). Computing and graphing highest density regions. *The American Statistician* **50**, 120–6.
- Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association* **90**, 773–95.
- Mammen, E. and Polonik, W. (2013). Confidence regions for level sets. *Journal of Multivariate Analysis* **122**, 202–14.

References 📚 V

Pardo-Fernández, J. C. (2025). Discussion of "Mode-based estimation of the center of symmetry". *Annals of the Institute of Statistical Mathematics* **77**, 723–5.