Geometric Data Analysis Reading Group

An Efficient and Continuous Voronoi Density Estimator

Paper Authors:
Giovanni Marchetti, Vladislav Polianskii, Anastasiia Varava, Florian Pokorny, Danica Kragic

Paper link: https://arxiv.org/abs/2210.03964
(AISTATS 2023 Oral)

February 13, 2023
Presented by *Yikun Zhang*

Main References

- *Vladislav Polianskii, Giovanni Luca Marchetti, Alexander Kravberg, Anastasiia Varava, Florian T. Pokorny, and Danica Kragic*. **"Voronoi density estimator for high-dimensional data: Computation, compactification and convergence."** In Uncertainty in Artificial Intelligence, pp. 1644-1653. PMLR (Polianskii et al., 2022).

- *Giovanni L. Marchetti, Vladislav Polianskii, Anastasiia Varava, Florian T. Pokorny, and Danica Kragic*. **"An Efficient and Continuous Voronoi Density Estimator."** arXiv preprint arXiv:2210.03964 (Marchetti et al., 2022).

Table of Contents

Background: Voronoi Density Estimator

Density Estimation

Goal: Given observed data $\mathcal{D} = \{X_1, ..., X_n\} \subset \mathcal{X}$, we want to estimate the underlying density ρ that generates the data.

- **Parametric methods:** Assume that $\mathcal{D} \sim f(x; \theta)$ and estimate the parameter θ by its maximum likelihood estimator or maximum a posteriori (MAP) estimator.

- **Nonparametric methods:** Make no distributional/model assumptions and learn the estimator $\widehat{f}(x)$ directly from $\mathcal{D}$.

- **Semiparametric methods:** Fit a combination of the parametric and nonparametric estimators (Olkin and Spiegelman, 1987):

$$g(x; \pi) = \pi \cdot f(x; \widehat{\theta}) + (1 - \pi) \cdot \widehat{f}(x),$$

where $\pi \in [0, 1]$ is unknown and should be estimated from $\mathcal{D}$.

Nonparametric Density Estimation: Histograms

Partition the data space $\mathcal{X} \subset \mathbb{R}^d$ into (uniform) grids and count the number n_k of observations falling into grid Δ_k:

$$\widehat{f}_{hist}(x) = \frac{n_k}{n \cdot |\Delta_k|} \quad \text{with} \quad x \in \Delta_k,$$

where $|\Delta_k|$ is the Lebesgue measure of Δ_k in $\mathbb{R}^d$.

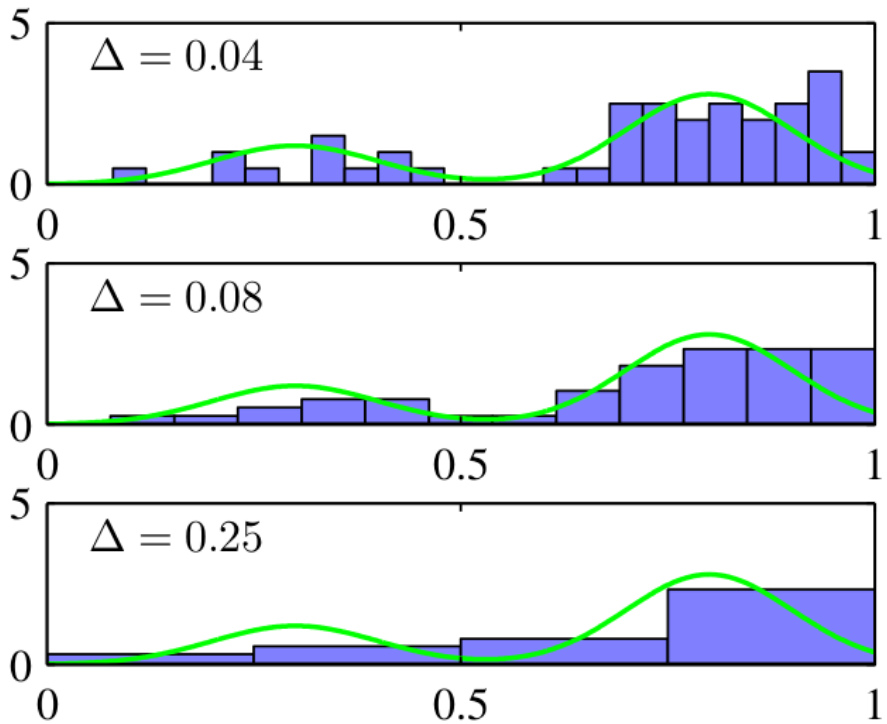


Figure 1: Illustrations of the histogram density estimation (Bishop, 2006).

For a fixed point $x \in \mathcal{X} \subset \mathbb{R}^d$ and an integer K,

$$\widehat{f}_{KNN}(x) = \frac{K}{n \cdot V_K(x)},$$

where $V_K(x) = \frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}+1\right)} \cdot R_K(x)$ is the volume of a d-dimensional ball centered at x with radius $R_K(x)$ as the distance from x to its K^{th} nearest neighbor in $\mathcal{D}$ (Loftsgaarden and Quesenberry, 1965; Zhao and Lai, 2022).

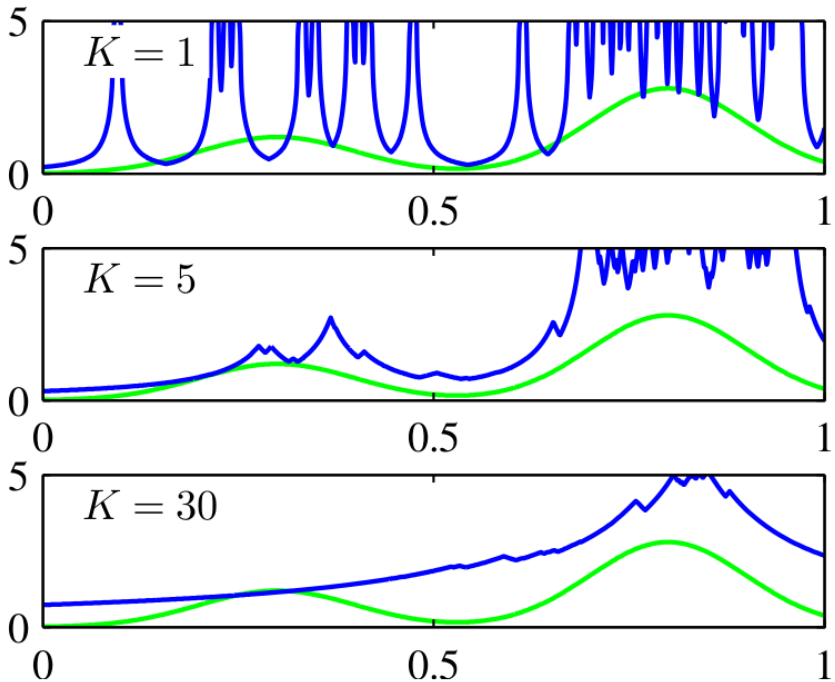


Figure 2: Illustrations of the KNN density estimation (Bishop, 2006).

Nonparametric Density Estimation: KDE

The kernel density estimator (KDE; Chen 2017) at point $x \in \mathcal{X} \subset \mathbb{R}^d$ is defined as:

$$\widehat{f}_h(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right),$$

where h is the bandwidth parameter and $K : \mathbb{R}^d \to \mathbb{R}_+$ is the kernel function satisfying

$$\int_{\mathbb{R}^d} K(x)\, dx = 1, \quad \int_{\mathbb{R}^d} x \cdot K(x)\, dx = 0, \quad \text{and} \quad \int_{\mathbb{R}^d} ||x||^2 K(x)\, dx < \infty.$$

Some examples of the kernel function includes

- Gaussian kernel: $K(x) = \frac{1}{(2\pi)^{\frac{d}{2}}} \exp\left(-\frac{||x||^2}{2}\right)$.
- Spherical kernel: $K(x) = \frac{\Gamma\left(\frac{d}{2}+1\right)}{\pi^{\frac{d}{2}}} \cdot \mathbb{1}_{\{||x|| \leq 1\}}$.
- Any other products of one-dimensional kernels: $K(x) = \prod_{j=1}^{d} K_j(x_j)$.

Nonparametric Density Estimation: KDE II

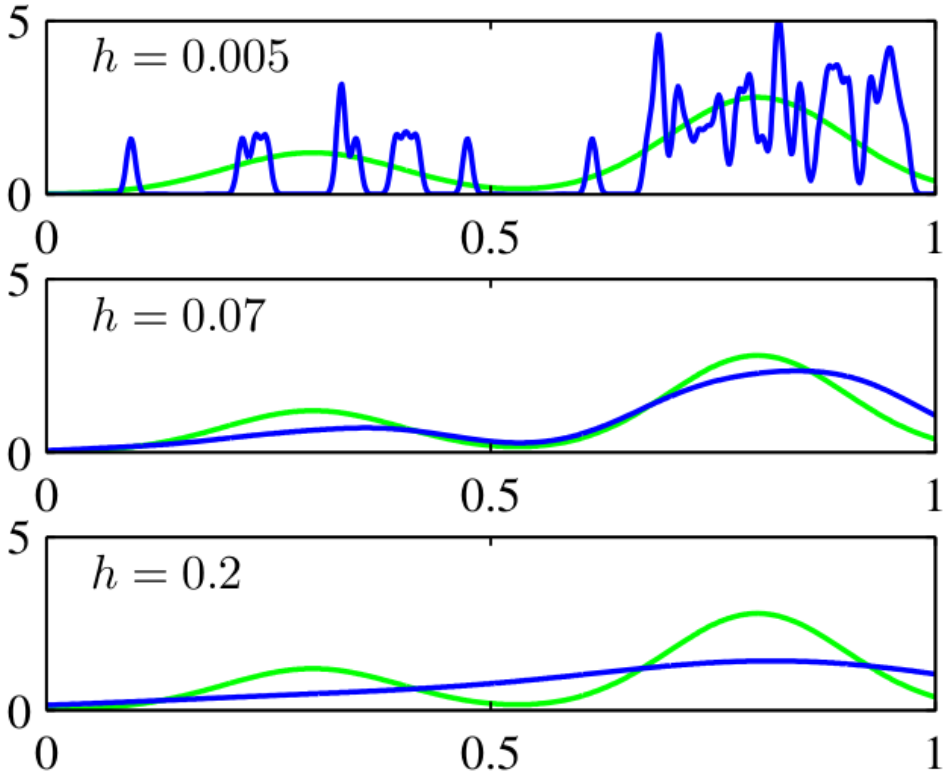


Figure 3: Illustrations of the kernel density estimator (Bishop, 2006).

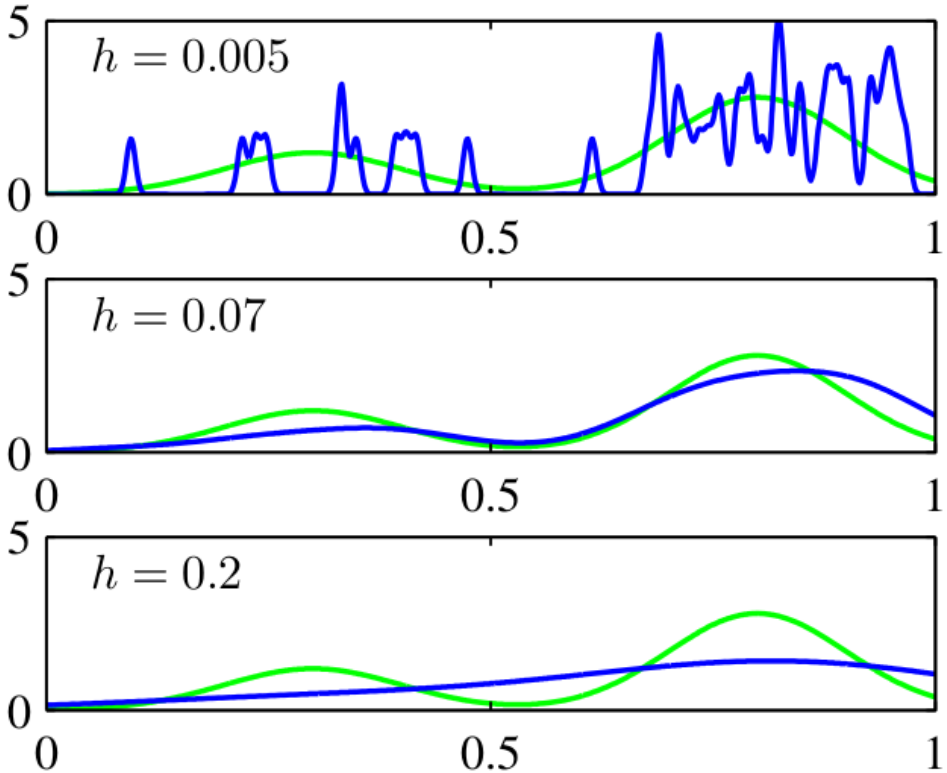


Figure 3: Illustrations of the kernel density estimator (Bishop, 2006).

To better capture the underlying distribution through KDE, we can also vary the bandwidth parameter h over the data space.

Nonparametric Density Estimation: Adaptive KDE I

Key idea: Use smaller bandwidths in regions with dense data and use larger bandwidths in regions with sparse data (Terrell and Scott, 1992):

Nonparametric Density Estimation: Adaptive KDE I

Key idea: Use smaller bandwidths in regions with dense data and use larger bandwidths in regions with sparse data (Terrell and Scott, 1992):

1. *Balloon estimator:*

$$\widehat{f}_1(x) = \frac{1}{n \cdot h(x)^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h(x)}\right),$$

where the bandwidth $h(x)$ depends on the query point $x \in \mathcal{X} \subset \mathbb{R}^d$.
Example: $h(x)$ is the distance from x to its K^{th} nearest neighbor in $\mathcal{D}$.

Nonparametric Density Estimation: Adaptive KDE I

Key idea: Use smaller bandwidths in regions with dense data and use larger bandwidths in regions with sparse data (Terrell and Scott, 1992):

1. *Balloon estimator:*

$$\widehat{f}_1(x) = \frac{1}{n \cdot h(x)^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h(x)}\right),$$

where the bandwidth $h(x)$ depends on the query point $x \in \mathcal{X} \subset \mathbb{R}^d$. **Example:** $h(x)$ is the distance from x to its K^{th} nearest neighbor in $\mathcal{D}$.

2. *Sample smoothing estimator:*

$$\widehat{f}_2(x) = \frac{1}{n \cdot h(X_i)^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h(X_i)}\right),$$

which is a mixture of individually scaled kernels centered at $X_1, ..., X_n$. **Example:** $h(X_i) \propto \widehat{f}_{pilot}(X_i)^{-\frac{1}{2}}$; see Abramson (1982).

Nonparametric Density Estimation: Adaptive KDE I

Another algorithm for constructing the sample smoothing (adaptive) KDE (Wang and Wang, 2007):

1. Find a pilot estimator $\widehat{f}_h(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)$ with a fixed bandwidth h selected by Silverman (1986); Stone (1984); Sheather and Jones (1991).

2. Define local bandwidth factors $\lambda_i, i = 1, ..., n$ by

$$\lambda_i = \left(\frac{g}{\widehat{f}_h(X_i)}\right)^{\alpha},$$

where $g = \prod_{i=1}^{n} \left[\widehat{f}_h(X_i)\right]^{\frac{1}{n}}$ and $\alpha \in [0, 1]$ is a fixed sensitivity parameter.

3. Define the adaptive KDE $\widehat{f}_{Ada}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{(h\lambda_i)^d} \cdot K\left(\frac{x-X_i}{h\lambda_i}\right)$.

Drawbacks of Histogram and KDE

- They suffer from bias due to the prior choice of the local geometric structures, *i.e.*, the shapes of grids or the contours of chosen kernels.
- This bias get worse in high-dimensional ambient space.

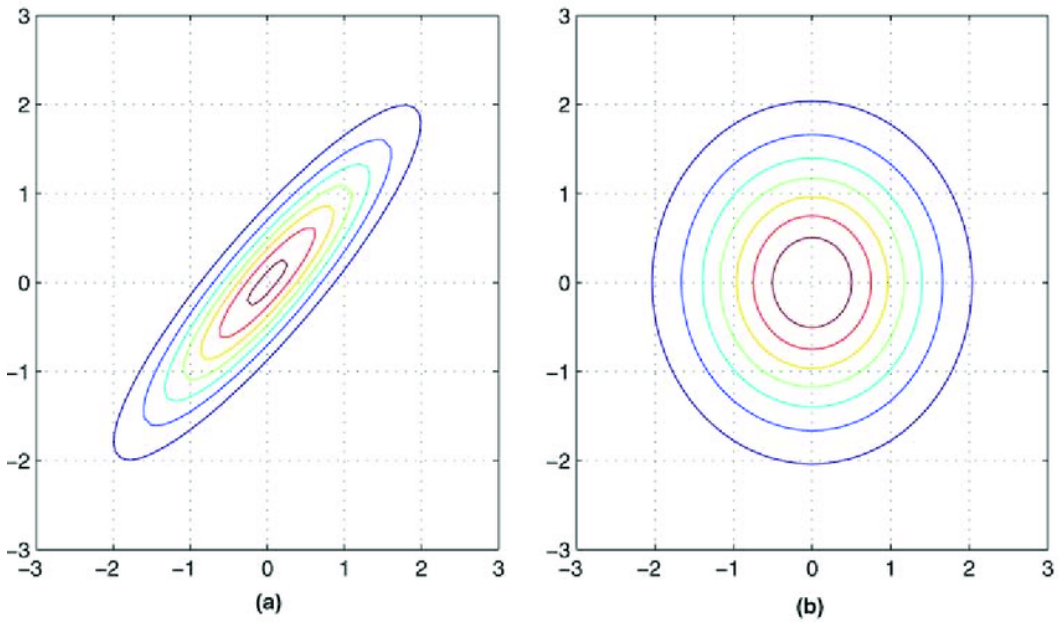


Figure 4: Contour plots of two Gaussian kernels (cited from Arasaratnam et al. 2007).

Voronoi Density Estimator

Voronoi cell: Given $\mathcal{D} = \{X_1, ..., X_n\}$, the Voronoi cell of X_i is defined as:

$$C(X_i) = \{x \in \mathbb{R}^d : d(x, X_i) \leq d(x, X_j), \forall X_j \in \mathcal{D}\}.$$

The collection $\{C(X_i)\}_{i=1}^n$ is called *Voronoi tessellation* generated by $\mathcal{D}$.

Math. Scientist, 1978, 3, 23-33

How Many Trees in a Forest?

Keith Ord

University of Warwick, England

(Received: 10 October 1975. Revision received 9 April 1976)

Abstract

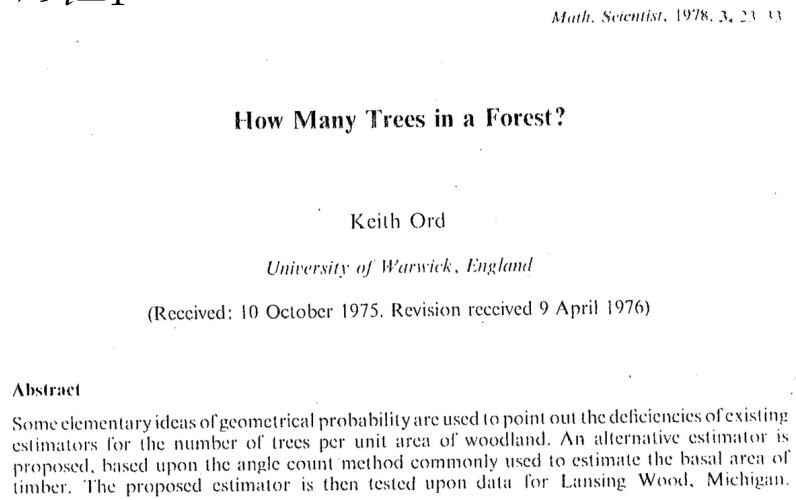
Some elementary ideas of geometrical probability are used to point out the deficiencies of existing estimators for the number of trees per unit area of woodland. An alternative estimator is proposed, based upon the angle count method commonly used to estimate the basal area of timber. The proposed estimator is then tested upon data for Lansing Wood, Michigan.

Voronoi Density Estimator

Voronoi cell: Given $\mathcal{D} = \{X_1, ..., X_n\}$, the Voronoi cell of X_i is defined as:

$$C(X_i) = \{x \in \mathbb{R}^d : d(x, X_i) \leq d(x, X_j), \forall X_j \in \mathcal{D}\}.$$

The collection $\{C(X_i)\}_{i=1}^n$ is called *Voronoi tessellation* generated by $\mathcal{D}$.



Math. Scientist. 1978, 3, 23-33

How Many Trees in a Forest?

Keith Ord

University of Warwick, England

(Received: 10 October 1975. Revision received 9 April 1976)

Abstract

Some elementary ideas of geometrical probability are used to point out the deficiencies of existing estimators for the number of trees per unit area of woodland. An alternative estimator is proposed, based upon the angle count method commonly used to estimate the basal area of timber. The proposed estimator is then tested upon data for Lansing Wood, Michigan.

Voronoi density estimator (VDE): It is defined (almost everywhere) on $\mathcal{X} \subset \mathbb{R}^d$ as (Ord, 1978):

$$\widehat{f}_{VDE}(x) = \frac{1}{n \cdot \text{Vol}(C(x))},$$

where $C(x)$ is the Voronoi cell to which the query point x belongs.

Drawbacks of the Standard VDE

1. Some Voronoi cells can have infinite volumes with respect to the Lebesgue measure.
   - A common solution is to restrict the measure to a fixed bounded region $A \subset \mathbb{R}^d$ that contains $\mathcal{D}$ (Moradi et al., 2019).

Drawbacks of the Standard VDE

1. Some Voronoi cells can have infinite volumes with respect to the Lebesgue measure.
   - A common solution is to restrict the measure to a fixed bounded region $A \subset \mathbb{R}^d$ that contains $\mathcal{D}$ (Moradi et al., 2019).

2. Data may concentrate around a submanifold in $\mathbb{R}^d$ with high codimension (Fefferman et al., 2016) and the performance of VDE becomes highly sensitive to the choice of A.

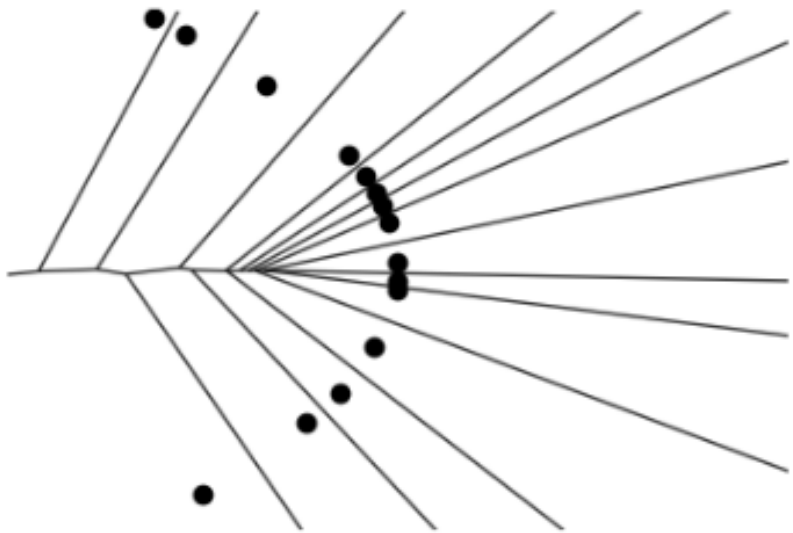


Figure 5: Voronoi tessellation for generators distributed on a parabola in $\mathbb{R}^2$.

Highlight of Today's Paper Discussion

We will discuss two new extensions for the traditional VDE to resolve the above issues.

1. Compactified Voronoi density estimator (CVDE):
   - Efficient computation for Voronoi volumes.
   - Feasible approach for sampling.

2. Radial Voronoi density estimator (RVDE):
   - Inherit the computational benefits from CVDE.
   - RVDE is continuous among the domain $\mathcal{X}$.

Compactified Voronoi Density Estimator

Compactified Voronoi Density Estimator

Key idea: Make the measure of each cell finite by a *local* kernel $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$ with its mode at each point in $\mathcal{D} = \{X_1, ..., X_n\}$.

Compactified Voronoi Density Estimator

Key idea: Make the measure of each cell finite by a *local* kernel $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$ with its mode at each point in $\mathcal{D} = \{X_1, ..., X_n\}$.

Compactified Voronoi Density Estimator (CVDE): It is defined (almost everywhere) on $\mathcal{X} \subset \mathbb{R}^d$ as (Polianskii et al., 2022):

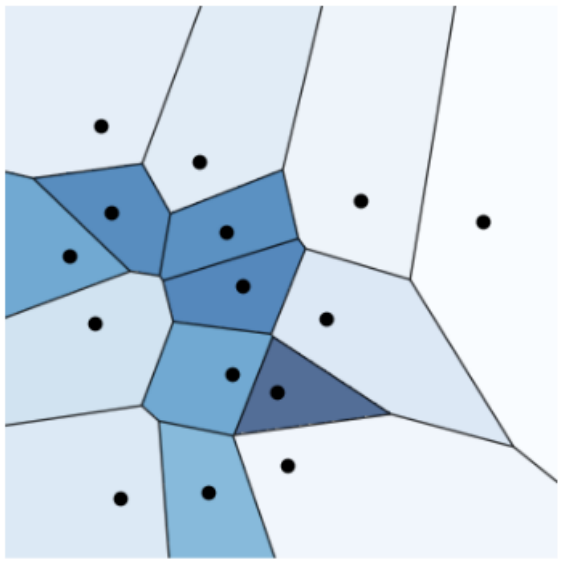
$$\widehat{f}_{CVDE}(x) = \frac{K(p, x)}{n \cdot \mathrm{Vol}_p(C(x))},$$

where $\mathrm{Vol}_p(C(x)) = \int_{C(x)} K(p, y)\, dy$ and $p \in \mathcal{D}$ is the generator of $C(x)$, *i.e.*, the point in $\mathcal{D}$ closest to x.

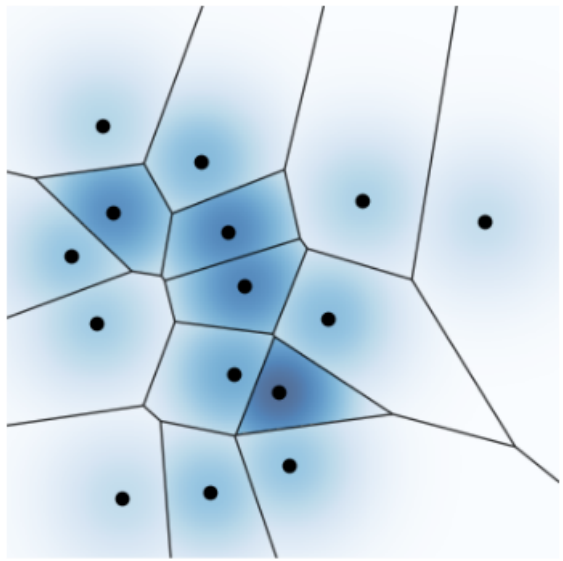
- A common choice of the kernel $K(p, x) \equiv K\left(\frac{p-x}{h}\right)$ is still the Gaussian one with a bandwidth parameter $h > 0$ as:

$$K(p, x) = \exp\left(-\frac{||p - x||^2}{2}\right).$$

Comparisons Between VDE and CVDE



VDE with bounding square A



CVDE with Gaussian kernel

Figure 6: Comparison between VDE and CVDE for data in $\mathbb{R}^2$.

Theoretical Properties of CVDE

Suppose that the true density ρ for generating $\{X_1, ..., X_n\}$ has its support as $\mathbb{R}^d$ and $K \in L^1(\mathbb{R}^d)$.

- $\widehat{f}_{CVDE}(x) \xrightarrow{P} f(x)$ for any fixed $x \in \mathbb{R}^d$ as $n \to \infty$.
- In the paper, it states that $\mathbb{P}_n(E) \xrightarrow{P} \mathbb{P}(E)$ for any measurable set $E \subset \mathbb{R}^d$ and $\mathbb{P}_n = \widehat{f}_{CVDE}\, dx, \mathbb{P} = \rho\, dx$.

Theoretical Properties of CVDE

Suppose that the true density ρ for generating $\{X_1, ..., X_n\}$ has its support as $\mathbb{R}^d$ and $K \in L^1(\mathbb{R}^d)$.

- $\widehat{f}_{CVDE}(x) \xrightarrow{P} f(x)$ for any fixed $x \in \mathbb{R}^d$ as $n \to \infty$.
- In the paper, it states that $\mathbb{P}_n(E) \xrightarrow{P} \mathbb{P}(E)$ for any measurable set $E \subset \mathbb{R}^d$ and $\mathbb{P}_n = \widehat{f}_{CVDE}\, dx, \mathbb{P} = \rho\, dx$.

Pros: Compared with KDE theory, it requires mild assumptions on K, no needs for h to vanish asymptotically, etc.

Cons: The proof follows from Portmanteau Lemma and cannot be modified to derive the rate of convergence.

If in addition $K : \mathbb{R}^d \to \mathbb{R}_+$ is continuous, we have

- $\widehat{f}_{CVDE} \xrightarrow{d} \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ as $h \to 0$, where δ_p is the Dirac's measure at p.
- $\widehat{f}_{CVDE} \xrightarrow{d} \widehat{f}_{VDE}$ as $h \to \infty$ when restricting K to a bounded region A.

Computations of CVDE (*Directional Radius*)

Setup: Consider an arbitrary unit vector $\sigma \in \mathbb{S}^{d-1}$ and point $z \in \mathbb{R}^d$. Define

$$\ell_z(\sigma) = \sup\{t \geq 0 : z + t\sigma \in C(z)\},$$

where $C(z)$ is again the Voronoi cell that contains z. If such t does not exist, then $\ell_z(\sigma) = \infty$.

Computations of CVDE (*Directional Radius*)

Setup: Consider an arbitrary unit vector $\sigma \in \mathbb{S}^{d-1}$ and point $z \in \mathbb{R}^d$. Define

$$\ell_z(\sigma) = \sup\{t \geq 0 : z + t\sigma \in C(z)\},$$

where $C(z)$ is again the Voronoi cell that contains z. If such t does not exist, then $\ell_z(\sigma) = \infty$. It can be computed as follows:

1. Denote by $p \in \mathcal{D}$ the generator closest to z and for $q \in \mathcal{D} \setminus \{p\}$, set

$$\ell_z^q(\sigma) = \frac{||q-p||^2}{2\langle \sigma, q-p\rangle}.$$

2. Take

$$\ell_z(\sigma) = \min_{q \neq p, \ell_z^q(\sigma) \geq 0} \ell_z^q(\sigma)$$

with $\ell_z(\sigma) = \infty$ if $\ell_z^q(\sigma) < 0$ for all $q \in \mathcal{D} \setminus \{p\}$; see Polianskii and Pokorny (2019) as well.

Correctness of Computing $\ell_z(\sigma)$

Recall that $\ell_z^q(\sigma) = \frac{||q-p||^2}{2\langle\sigma, q-p\rangle}$ and $\ell_z(\sigma) = \min_{q\neq p, \ell_z^q(\sigma)\geq 0} \ell_z^q(\sigma)$.

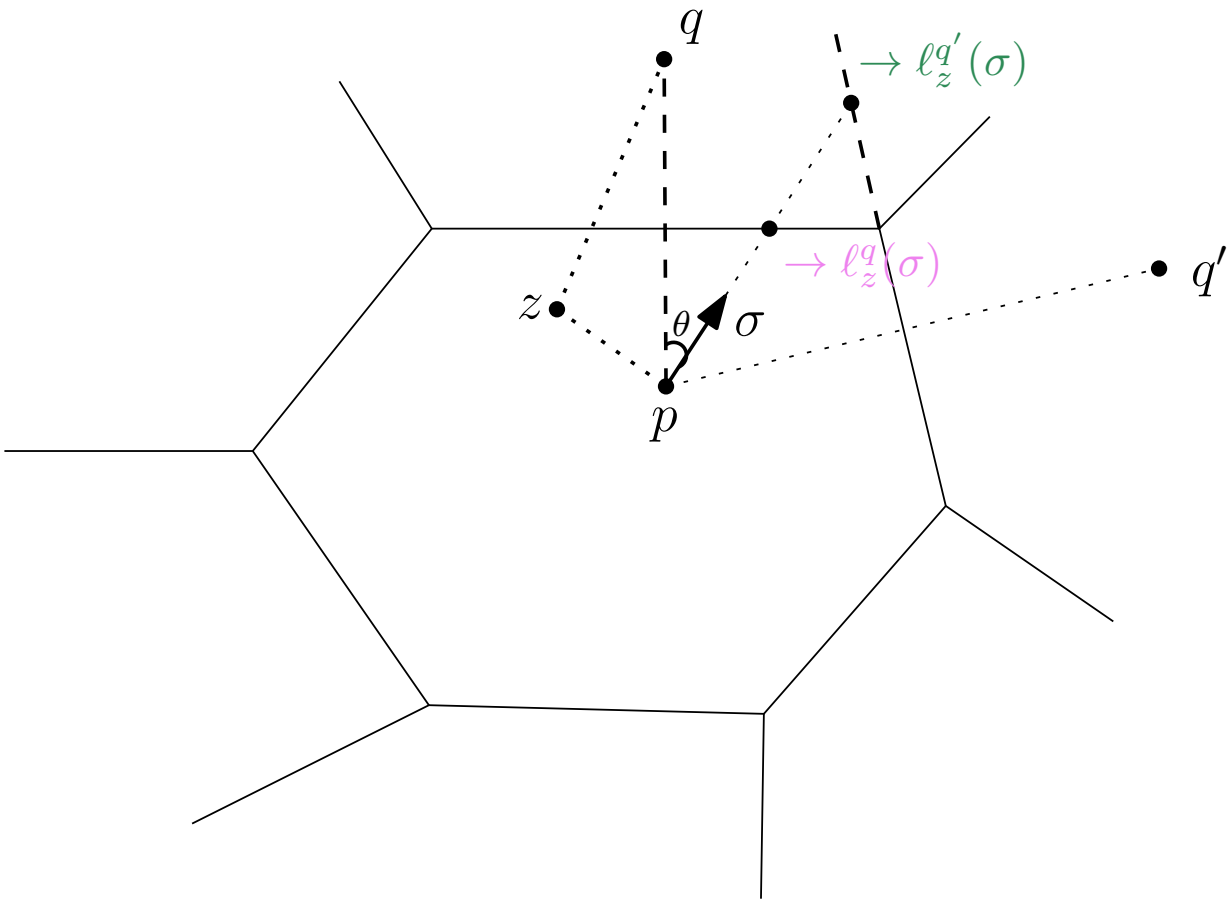


Figure 7: Illustration of the directional radius involved in the volume estimation and sampling.

Volume Estimation in CVDE

Recall that $\widehat{f}_{CVDE}(x) = \frac{K(p,x)}{n \cdot \mathrm{Vol}_p(C(x))}$ with

$$\mathrm{Vol}_p(C(x)) = \int_{\mathbb{S}^{d-1}} \int_0^{\ell_p(\sigma)} K\left(\frac{t\sigma}{h}\right) \cdot t^{d-1}\, dt\, d\sigma,$$

where $\ell_p(\sigma)$ is the directional radius of $C(x)$ starting from its generator p.

Volume Estimation in CVDE

Recall that $\widehat{f}_{CVDE}(x) = \frac{K(p,x)}{n \cdot \mathrm{Vol}_p(C(x))}$ with

$$\mathrm{Vol}_p(C(x)) = \int_{\mathbb{S}^{d-1}} \int_0^{\ell_p(\sigma)} K\left(\frac{t\sigma}{h}\right) \cdot t^{d-1}\, dt\, d\sigma,$$

where $\ell_p(\sigma)$ is the directional radius of $C(x)$ starting from its generator p.

The spherical integral is approximated through a *Monte Carlo method* by pre-sampling a finite set of unit vectors $\Sigma_p \subset \mathbb{S}^{d-1}$ and computing

$$\frac{2\pi^{\frac{d}{2}}}{|\Sigma_p| \cdot \Gamma\left(\frac{d}{2}\right)} \sum_{\sigma \in \Sigma_p} \int_0^{\ell_p(\sigma)} K\left(\frac{t\sigma}{h}\right) \cdot t^{d-1}\, dt$$

$$= (2\pi h^2)^{\frac{d}{2}} \cdot \bar{\gamma}\left(\frac{d}{2}, \ell_p(\sigma)\right) \qquad \text{when } K \text{ is the Gaussian kernel,}$$

where $\bar{\gamma}(a,z) = \frac{1}{\Gamma(a)} \int_0^z t^{a-1} e^{-t} dt$ is the regularized lower incomplete Gamma function.

Sampling Procedure for CVDE

A version of the *hit-and-run sampling* as:

1. Choose $p = z^{(0)}$ uniformly from $\mathcal{D} = \{X_1, ..., X_n\}$.
2. Construct a Markov chain $\{z^{(i)}\}$ as follows.
   - Sample $\sigma^{(i+1)} \in \mathbb{S}^{d-1}$ uniformly.
   - Sample $z^{(i+1)}$ from $\frac{1}{\mathrm{Vol}_p(C(p))} \cdot K(p, \cdot)$ restricted to the line segment

$$\left\{ z^{(i)} + t\sigma^{(i+1)} : t \in \left[-\ell_{z^{(i)}}(-\sigma^{(i+1)}),\ \ell_{z^{(i)}}(\sigma^{(i+1)}) \right] \right\}.$$

3. Output the last point $z^{(I)}$ as one sample point from $\widehat{f}_{CVDE}$ after a number I of steps.

Computational Complexity of CVDE

Pre-sample $\Sigma \subset \mathbb{R}^{d-1}$ and pre-compute all $\langle \sigma, p \rangle, \langle q, p \rangle$ in time complexity $\mathcal{O}\left(n|\mathcal{D}|^2 + n|\Sigma||\mathcal{D}|\right)$.

- Volume Estimation: $\mathcal{O}\left(|\Sigma||\mathcal{D}|^2\right)$.
- Sampling procedure (hit-and-run Markov chain): $\mathcal{O}\left(I(|\Sigma| + |\mathcal{D}|)\right)$.

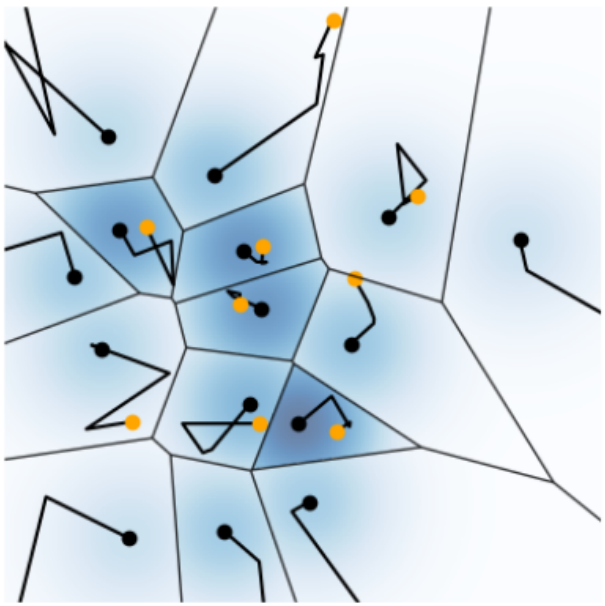


Figure 8: Illustration of the hit-and-run sampling procedure with a trajectory length as $I = 4$.

Sampling Comparison Between CVDE and VDE

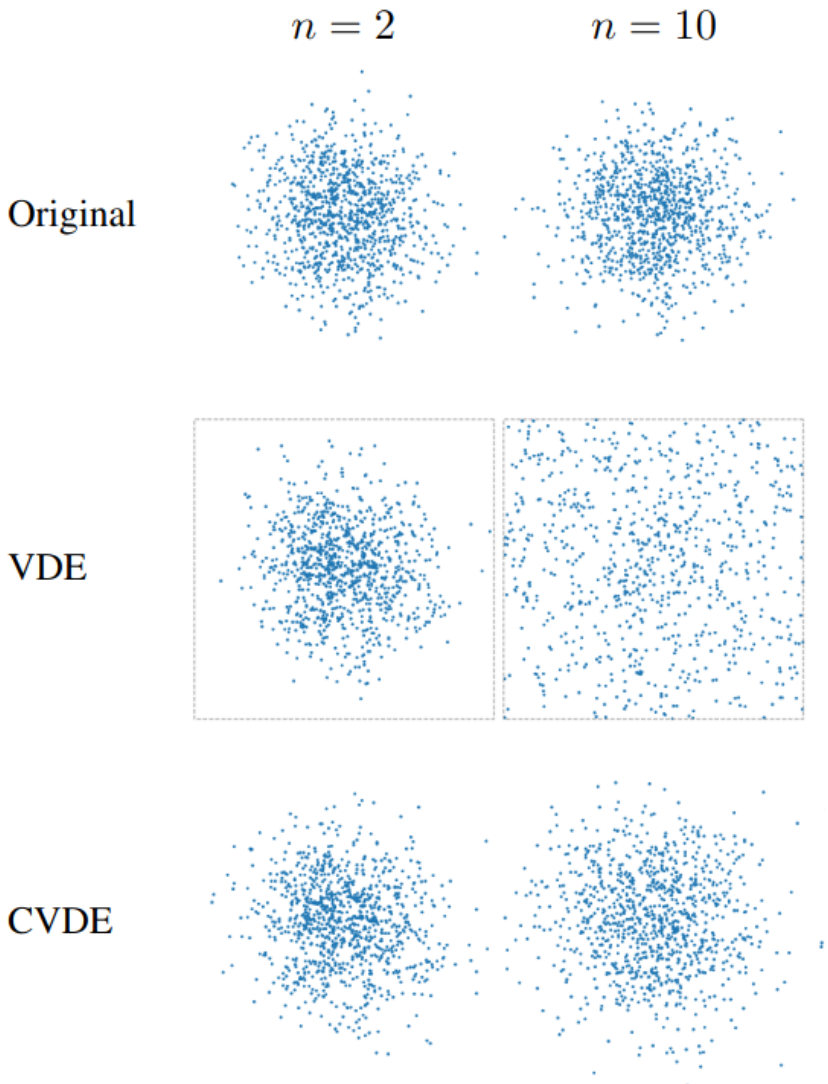


Figure 9: Visual Comparison between samples from the CVDE and VDE when estimating d-dimensional Gaussian distributions with $d = 2, 10$.

Drawbacks of CVDE

- The Monte Carlo method for approximating the volume $\mathrm{Vol}_p(C(x))$ is slow (and even infeasible) as the dimensions grow.
- The CVDE still jumps *discontinuously* when crossing the boundary of Voronoi cells.

Radial Voronoi Density Estimator

Radial Voronoi Density Estimator I

Recall from the definition of CVDE $\widehat{f}_{CVDE}(x) = \frac{K(p,x)}{n \cdot \mathrm{Vol}_p(C(x))}$ that

$$\int_{C(x)} \widehat{f}_{CVDE}(x)\, dx = \int_{\mathbb{S}^{d-1}} \int_0^{\ell_p(\sigma)} t^{d-1} \widehat{f}_{CVDE}(p + t\sigma)\, dt\, d\sigma,$$

where p is the generator of $C(x)$ (so that $C(x) = C(p)$) and $\ell_p(\sigma) = \sup\{t \geq 0 : p + t\sigma \in C(p)\}$.

Radial Voronoi Density Estimator I

Recall from the definition of CVDE $\widehat{f}_{CVDE}(x) = \frac{K(p,x)}{n \cdot \mathrm{Vol}_p(C(x))}$ that

$$\int_{C(x)} \widehat{f}_{CVDE}(x)\, dx = \int_{\mathbb{S}^{d-1}} \int_0^{\ell_p(\sigma)} t^{d-1} \widehat{f}_{CVDE}(p + t\sigma)\, dt\, d\sigma,$$

where p is the generator of $C(x)$ (so that $C(x) = C(p)$) and $\ell_p(\sigma) = \sup\{t \geq 0 : p + t\sigma \in C(p)\}$.

Now, define a similar quantity

$$\ell(x) = \sup\left\{t \geq 0 : p + t \cdot \frac{x-p}{||x-p||} \in C(p)\right\} \equiv \ell_p\left(\frac{x-p}{||x-p||}\right)$$

with $x \in C(p)$ and p being the generator of $C(p)$.

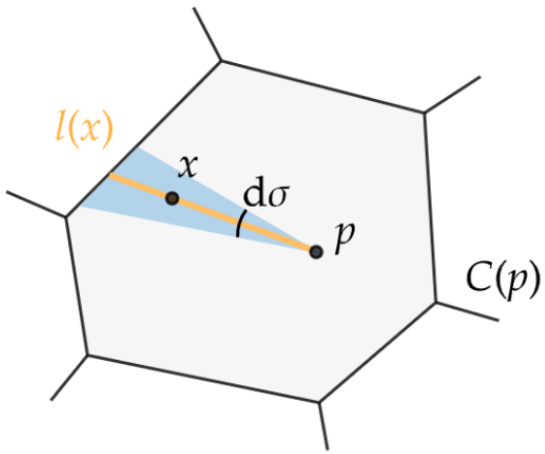
- $\ell(x)$ is defined for $x \neq p$.
- $\ell(x)$ is continuous since $\ell(x) = ||x-p|| = ||x-q||$ for $x \in C(p) \cap C(q)$.

Radial Voronoi Density Estimator II

Given a continuous and strictly decreasing kernel $\widetilde{K} : [-\delta, \infty) \to [0, \infty)$ with some $\delta > 0$ and $\int_0^\infty t^{d-1}\widetilde{K}(t)\,dt < \infty$, the radial Voronoi density estimator (RVDE) is defined as:

$$\widehat{f}_{RVDE}(x) = \frac{\widetilde{K}\left(\beta(\ell(x)) \cdot ||x - p||\right)}{\alpha n \cdot \mathrm{Vol}_p(\mathbb{S}^{d-1})} = \frac{2\pi^{\frac{d}{2}} \cdot \widetilde{K}\left(\beta(\ell(x)) \cdot ||x - p||\right)}{\alpha n \cdot \Gamma\left(\frac{d}{2}\right)},$$

where $\alpha > 0$ is a hyperparameter and $\beta : \mathbb{R}_+ \to \mathbb{R}$ is a continuous function to be determined.

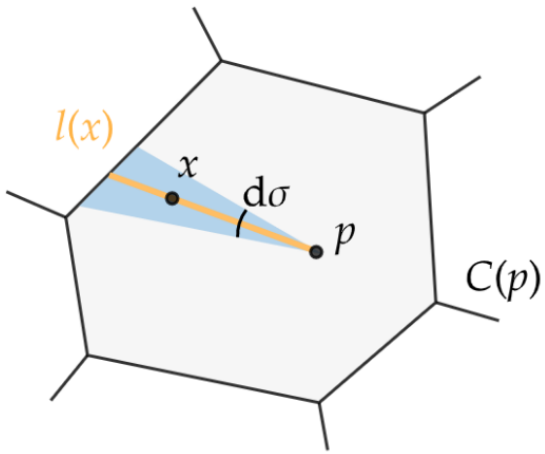


Radial Voronoi Density Estimator II

Given a continuous and strictly decreasing kernel $\widetilde{K} : [-\delta, \infty) \to [0, \infty)$ with some $\delta > 0$ and $\int_0^\infty t^{d-1}\widetilde{K}(t)\,dt < \infty$, the radial Voronoi density estimator (RVDE) is defined as:

$$\widehat{f}_{RVDE}(x) = \frac{\widetilde{K}\left(\beta(\ell(x)) \cdot ||x - p||\right)}{\alpha n \cdot \mathrm{Vol}_p(\mathbb{S}^{d-1})} = \frac{2\pi^{\frac{d}{2}} \cdot \widetilde{K}\left(\beta(\ell(x)) \cdot ||x - p||\right)}{\alpha n \cdot \Gamma\left(\frac{d}{2}\right)},$$

where $\alpha > 0$ is a hyperparameter and $\beta : \mathbb{R}_+ \to \mathbb{R}$ is a continuous function to be determined.



The idea is quite similar to the Face Density for an edge weights in Wei and Chen (2023).

Comparisons Between KDE, VDE, CVDE, and RVDE

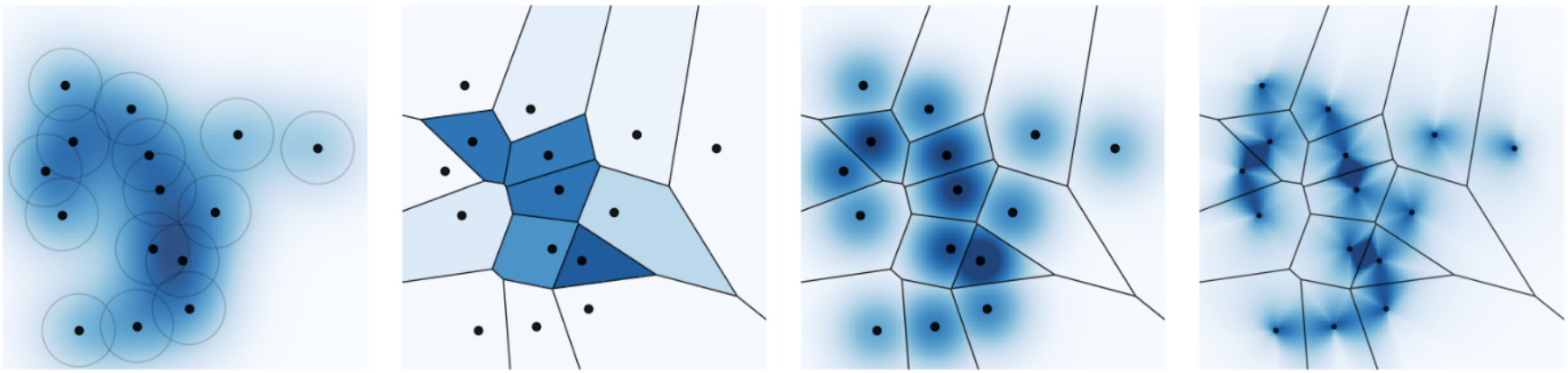


Figure 10: Comparisons between KDE, VDE, CVDE, and RVDE for data in $\mathbb{R}^2$.

Choices of Kernel $\widetilde{K}$ and β for RVDE

Recall that the kernel $\widetilde{K} : [-\delta, \infty) \to [0, \infty)$ for RVDE is required to be continuous and strictly decreasing with $\int_0^\infty t^{d-1}\widetilde{K}(t)\,dt < \infty$.

- *Exponential*: $\widetilde{K}(t) = e^{-t}$ with $\delta < \infty$.
- *Rational*: $\widetilde{K}(t) = \frac{1}{(t+1)^k}$ with $k > d$ and $\delta < 1$.

Choices of Kernel $\widetilde{K}$ and β for RVDE

Recall that the kernel $\widetilde{K} : [-\delta, \infty) \to [0, \infty)$ for RVDE is required to be continuous and strictly decreasing with $\int_0^\infty t^{d-1}\widetilde{K}(t)\,dt < \infty$.

- *Exponential*: $\widetilde{K}(t) = e^{-t}$ with $\delta < \infty$.
- *Rational*: $\widetilde{K}(t) = \frac{1}{(t+1)^k}$ with $k > d$ and $\delta < 1$.

To identify the function β, we constrain $\int_{C(x)} \widehat{f}_{RVDE}(x)\,dx = \frac{1}{n}$ so that

$$\int_0^\ell t^{d-1} \cdot \widetilde{K}\left(\beta(\ell) \cdot t\right)\, dt = \alpha \tag{1}$$

for every $\ell > 0$. When $\widetilde{K}$ is strictly decreasing, Eq.(1) has a unique solution $\beta(\ell) > -\frac{\delta}{\ell}$.

Choices of Kernel $\widetilde{K}$ and β for RVDE

Recall that the kernel $\widetilde{K} : [-\delta, \infty) \to [0, \infty)$ for RVDE is required to be continuous and strictly decreasing with $\int_0^\infty t^{d-1}\widetilde{K}(t)\,dt < \infty$.

- *Exponential*: $\widetilde{K}(t) = e^{-t}$ with $\delta < \infty$.
- *Rational*: $\widetilde{K}(t) = \frac{1}{(t+1)^k}$ with $k > d$ and $\delta < 1$.

To identify the function β, we constrain $\int_{C(x)} \widehat{f}_{RVDE}(x)\,dx = \frac{1}{n}$ so that

$$\int_0^\ell t^{d-1} \cdot \widetilde{K}\left(\beta(\ell) \cdot t\right)\, dt = \alpha \tag{1}$$

for every $\ell > 0$. When $\widetilde{K}$ is strictly decreasing, Eq.(1) has a unique solution $\beta(\ell) > -\frac{\delta}{\ell}$.

Example: When $d = 1$ and $\widetilde{K}(t) = e^{-t}$, $\beta(\ell) = \frac{1}{\alpha} + W\left(-\frac{\ell}{\alpha} \cdot e^{-\frac{\ell}{\alpha}}\right)$ is closely related to the Lambert W function (Corless et al., 1996).

Computing β In Practice

For any $\ell > 0$ and $\widetilde{K} : [-\delta, \infty) \to [0, \infty)$ being continuously differentiable, $\beta(\ell)$ is computed through a Newton-Raphson method with the iterative formula as:

$$\beta_{m+1} \leftarrow \beta_m + \frac{\beta_m}{d}\left(1 - \frac{\ell^d \cdot \widetilde{K}(\beta_m \cdot \ell) - n\alpha}{\ell^d \cdot \widetilde{K}(\beta_m \ell) - d\int_0^\ell t^{d-1}\widetilde{K}(\beta_m t)\,dt}\right)$$

for $m = 0, 1, \ldots$. If in addition $\widetilde{K}$ is convex, then $\lim_{m\to\infty} \beta_m = \beta(\ell)$.

Computing β In Practice

For any $\ell > 0$ and $\widetilde{K} : [-\delta, \infty) \to [0, \infty)$ being continuously differentiable, $\beta(\ell)$ is computed through a Newton-Raphson method with the iterative formula as:

$$\beta_{m+1} \leftarrow \beta_m + \frac{\beta_m}{d}\left(1 - \frac{\ell^d \cdot \widetilde{K}(\beta_m \cdot \ell) - n\alpha}{\ell^d \cdot \widetilde{K}(\beta_m \ell) - d\int_0^\ell t^{d-1}\widetilde{K}(\beta_m t)\,dt}\right)$$

for $m = 0, 1, \ldots$. If in addition $\widetilde{K}$ is convex, then $\lim_{m\to\infty} \beta_m = \beta(\ell)$.

Proof (Sketch).

Consider $F(\beta) = \int_0^\ell t^{d-1}\widetilde{K}(\beta \cdot t)\,dt - \alpha$ and solve for $F(\beta) = 0$ using

$$\beta_{m+1} \leftarrow \beta_m - \frac{F(\beta_m)}{\frac{d}{d\beta}F(\beta_m)}.$$

Use the integration by part for explicating $\frac{d}{d\beta}F(\beta_m)$. □

Other Properties for β

Recall that $\int_0^\ell t^{d-1} \cdot \widetilde{K}\left(\beta(\ell) \cdot t\right) \, dt = \alpha$.

- $\beta : \mathbb{R}_+ \to \mathbb{R}$ is increasing due to the increasing property of $\widetilde{K}$.
- $\beta(\ell) = 0$ when $\ell = (d \cdot \alpha)^{\frac{1}{d}}$.
- β has a horizontal asymptote: $\lim_{\ell \to \infty} \beta(\ell) = \left(\frac{1}{\alpha} \int_0^\infty t^{d-1} \widetilde{K}(t) \, dt \right)^{\frac{1}{d}}$.
- If, in addition, $\widetilde{K}$ is continuously differentiable, so does β and it satisfies the differential equation:

$$\left(\ell - \frac{d\alpha}{\ell^{d-1} \cdot \widetilde{K}\left(\beta(\ell) \cdot \ell\right)} \right) \frac{d}{d\ell} \beta(\ell) = -\beta(\ell).$$

So, β generalizes the Lambert W function (Corless et al., 1996).

Mode Estimation for RVDE

Let $\epsilon = (d \cdot \alpha)^{\frac{1}{d}}$. Recall that $\widehat{f}_{RVDE}(x) = \frac{2\pi^{\frac{d}{2}} \cdot \widetilde{K}(\beta(\ell(x)) \cdot ||x-p||)}{\alpha n \cdot \Gamma\left(\frac{d}{2}\right)}$.

Observation: $\widehat{f}_{RVDE}$ decreases radially with respect to p in the direction of x if $\ell(x) > \epsilon$ and increases otherwise.

Mode Estimation for RVDE

Let $\epsilon = (d \cdot \alpha)^{\frac{1}{d}}$. Recall that $\widehat{f}_{RVDE}(x) = \frac{2\pi^{\frac{d}{2}} \cdot \widetilde{K}(\beta(\ell(x)) \cdot ||x-p||)}{\alpha n \cdot \Gamma\left(\frac{d}{2}\right)}$.

Observation: $\widehat{f}_{RVDE}$ decreases radially with respect to p in the direction of x if $\ell(x) > \epsilon$ and increases otherwise.

The modes of $\widehat{f}_{RVDE}$ are classified as follows:

1. $p \in \mathcal{D}$ if $||p - q|| > 2\epsilon$ for every Voronoi cell $C(q)$ adjacent to $C(p)$.
2. $\frac{p+q}{2}$ for $p, q \in \mathcal{D}$ if $\frac{p+q}{2} \in C(p) \cap C(q)$ and $||p - q|| < 2\epsilon$.
3. all the points in the segment $[p, q]$ for $p, q \in \mathcal{D}$ if $\frac{p+q}{2} \in C(p) \cap C(q)$ and $||p - q|| = 2\epsilon$.

Mode Estimation for RVDE

Let $\epsilon = (d \cdot \alpha)^{\frac{1}{d}}$. Recall that $\widehat{f}_{RVDE}(x) = \frac{2\pi^{\frac{d}{2}} \cdot \widetilde{K}(\beta(\ell(x)) \cdot ||x-p||)}{\alpha n \cdot \Gamma\left(\frac{d}{2}\right)}$.

Observation: $\widehat{f}_{RVDE}$ decreases radially with respect to p in the direction of x if $\ell(x) > \epsilon$ and increases otherwise.

The modes of $\widehat{f}_{RVDE}$ are classified as follows:

1. $p \in \mathcal{D}$ if $||p - q|| > 2\epsilon$ for every Voronoi cell $C(q)$ adjacent to $C(p)$.
2. $\frac{p+q}{2}$ for $p, q \in \mathcal{D}$ if $\frac{p+q}{2} \in C(p) \cap C(q)$ and $||p - q|| < 2\epsilon$.
3. all the points in the segment $[p, q]$ for $p, q \in \mathcal{D}$ if $\frac{p+q}{2} \in C(p) \cap C(q)$ and $||p - q|| = 2\epsilon$.

Thus, α determines the extent by which points in $\mathcal{D}$ are "isolated" (*i.e.*, modes) or otherwise get "merged" by placing a mode between them!

Modes of RVDE and Gabriel Graph I

Gabriel graph (Gabriel and Sokal, 1969): A graph $G = (V, E)$ satisfies that $(p, q) \in E$ when the closed ball having (p, q) as a diameter contains no other points.

$\Longrightarrow$

Modes of RVDE and Gabriel Graph I

Gabriel graph (Gabriel and Sokal, 1969): A graph $G = (V, E)$ satisfies that $(p, q) \in E$ when the closed ball having (p, q) as a diameter contains no other points.

$\Longrightarrow$

The modes of $\widehat{f}_{RVDE}$ are

1. All isolated vertices in G;
2. Midpoints of edges in G;
3. Entire edges of length 2ϵ in G.

Modes of RVDE and Gabriel Graph I

Gabriel graph (Gabriel and Sokal, 1969): A graph $G = (V, E)$ satisfies that $(p, q) \in E$ when the closed ball having (p, q) as a diameter contains no other points.

$\Longrightarrow$

The modes of $\widehat{f}_{RVDE}$ are

1. All isolated vertices in G;
2. Midpoints of edges in G;
3. Entire edges of length 2ϵ in G.

Fact: The number of cycles in G is $|V| - |E| + 1$.

Heuristic Selection Rule for α: Select $2\epsilon = 2(d \cdot \alpha)^{\frac{1}{d}}$ as $\frac{|V|-1}{|E|}$-percentile for the lengths of edges in the Gabriel graph $G = (V, E)$.

Modes of RVDE and Gabriel Graph I

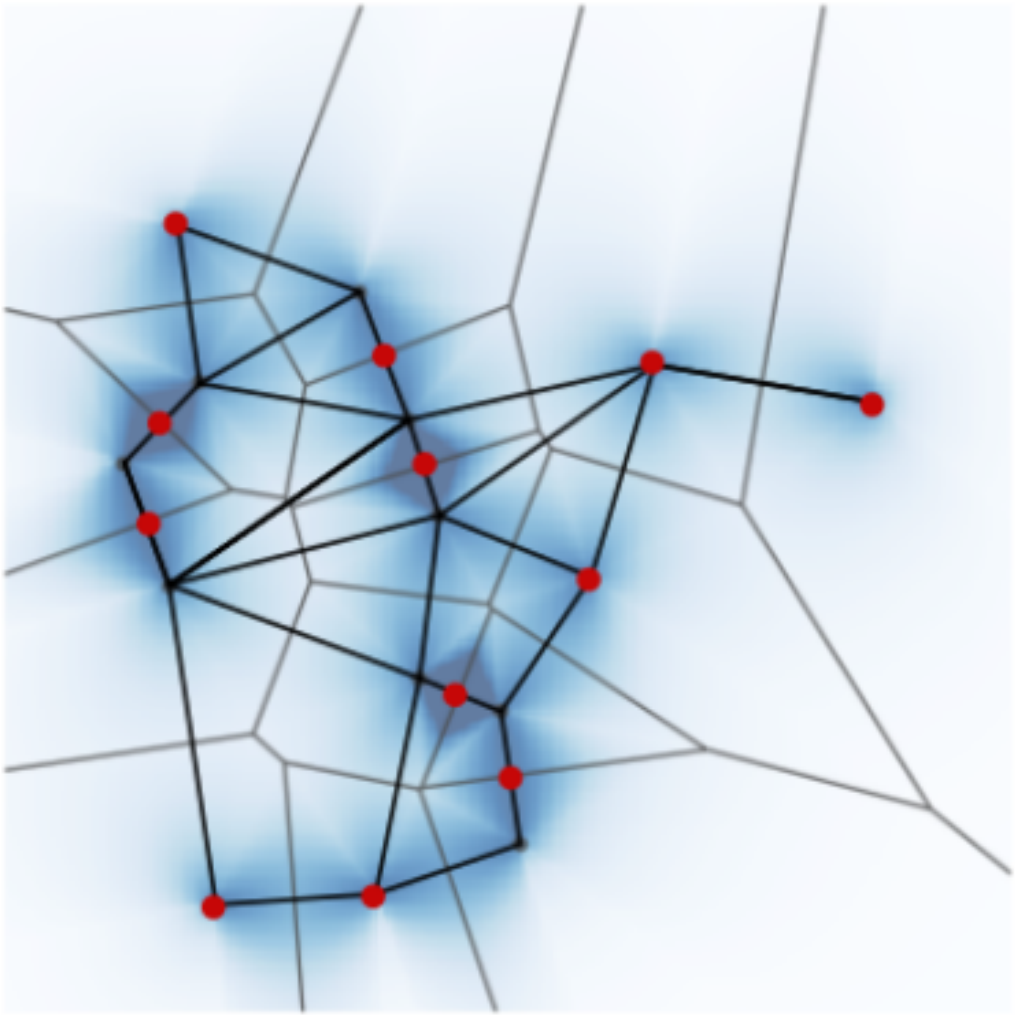


Figure 11: Illustration of the modes of RVDE (red) together with the Gabriel graph (black).

Experimental Results

Evaluation Metrics and Datasets

Evaluation Metrics: Average log-likelihood on a test set P_{test} as:

$$\frac{1}{|P_{\text{test}}|} \sum_{x \in P_{\text{test}}} \log \widehat{f}(x).$$

Datasets:

- *Synthetic datasets:* $d = 10$, $|\mathcal{D}| = 1000$, and $|P_{\text{test}}| = 1000$.
  - Standard Gaussian;
  - Dirichlet distribution with parameters $\alpha_i = \frac{1}{d+1}$;
  - Mixture of two Gaussians with means $\mu_1 = -\mu_2 = \left(-\frac{1}{2}, 0, ..., 0\right)$ and $\sigma_1 = 0.1, \sigma_2 = 10$.

- *MNIST:* Downscale the original resolution of 28×28 gray scale images to 10×10 so that $x \in [0,1]^{10 \times 10}$. Training set size: 30,000; Test set size: 10,000.

- *Anuran calls:* 7195 calls from 10 species of frogs with $x \in [0,1]^{21}$. Test set size: 10%.

Comparisons of Different Kernels in RVDE

- *Exponential*: $\widetilde{K}(t) = e^{-t}$ with $\delta < \infty$.
- *Rational*: $\widetilde{K}(t) = \frac{1}{(t+1)^k}$ with $k > d$ and $\delta < 1$.

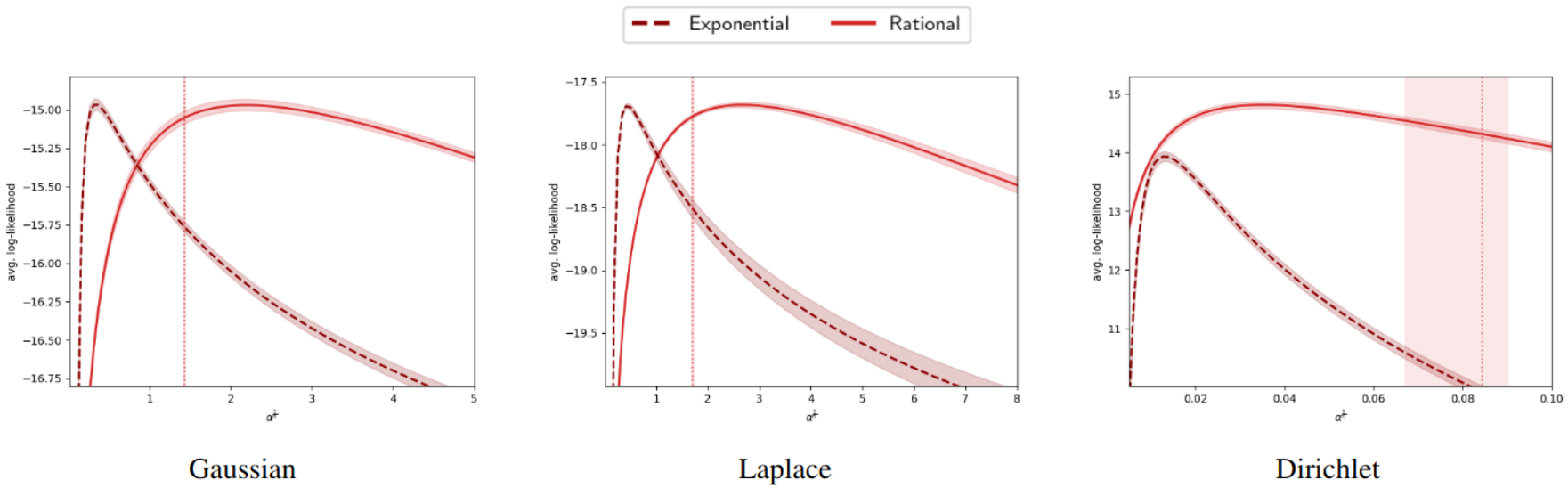


Figure 12: Comparisons between two kernels for RVDE. (The horizontal line corresponds to the heuristic selection rule for α in each panel.)

Comparisons Between Density Estimators I

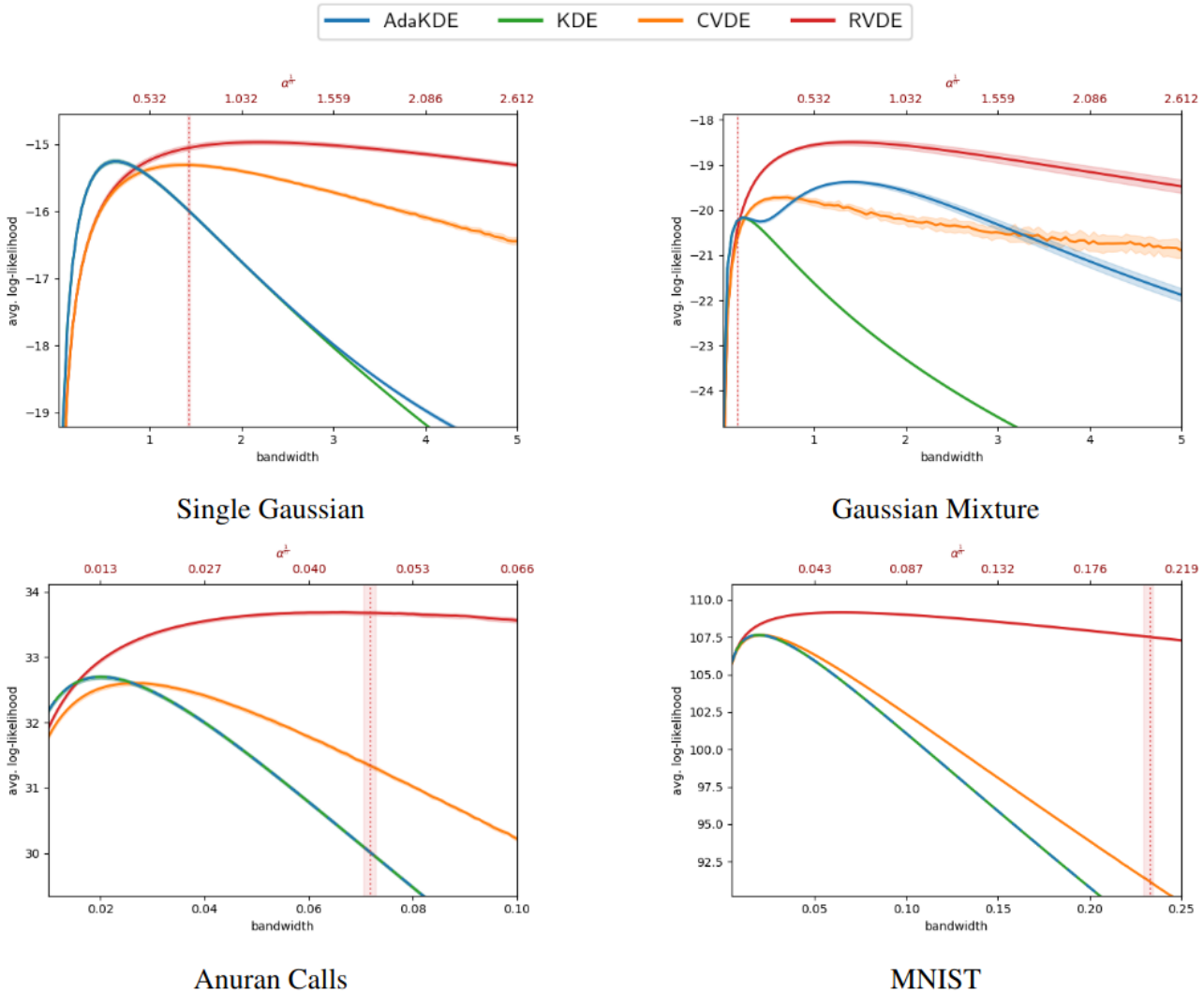


Figure 13: Comparisons of different estimators as the bandwidth $h = \alpha^{\frac{1}{d}} / \left(\int_0^\infty K(t)\,dt\right)^{\frac{1}{d}}$ varies. All the estimators implement the rational kernel.

Comparisons Between Density Estimators II

(Empirical) *Hellinger distance*: $\frac{1}{2|P_{\text{test}}|}\sum_{x\in P_{\text{test}}}\left[\widehat{f}(x)^{\frac{1}{2}}-\rho(x)^{\frac{1}{2}}\right]^2$.

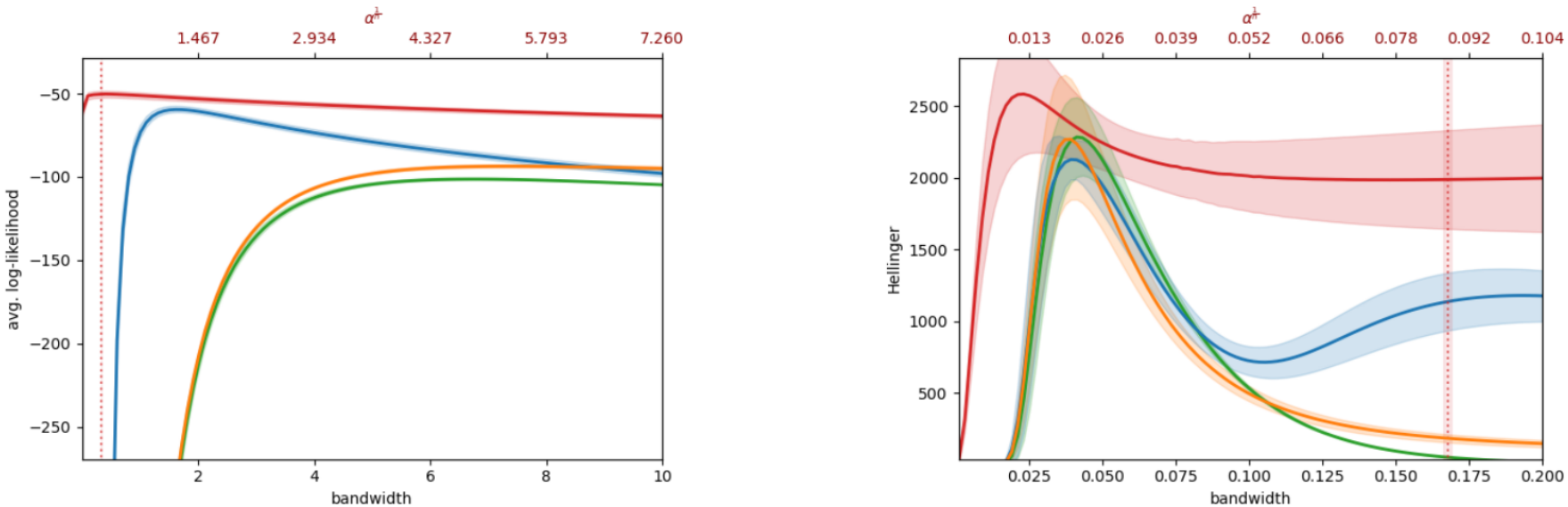


Figure 14: Comparison of different estimators on a 30-dimensional Gaussian mixture (left) and on a 10-dimensional Gaussian mixture with the Hellinger distance as a metric (right).

Comparisons Between Density Estimators III

	RVDE	CVDE	KDE	AdaKDE
Gaussian	0.0376	0.265	0.0340	0.266
Anuran Calls	0.0581	0.490	0.0787	0.870
MNIST	17.4	408	12.5	75.0

Figure 15: Average running times (in seconds) per one full train-test run with fixed bandwidths.

	RVDE	CVDE	KDE	AdaKDE
Gaussian	0.788	0.843	0.572	0.572
Anuran Calls	1.170	1.253	1.152	1.152
MNIST	5.507	5.767	5.735	5.735

Figure 16: Standard deviations of the (test) log-likelihood over 5 experimental runs. Each estimator is considered with its best bandwidth.

Conclusions and Future Directions

Conclusions and Future Directions

We discuss two new extensions for the traditional Voronoi density estimator (VDE) based on "kernel tricks" and rays from the generator p in a Voronoi cell $C(p)$.

1. Compactified Voronoi density estimator (CVDE):
   - Monte Carlo method for computing the volume of $V(p)$.
   - Hit-and-run method for sampling from $\widehat{f}_{CVDE}$.
2. Radial Voronoi density estimator (RVDE):
   - Inherit the computational benefits from CVDE.
   - $\widehat{f}_{RVDE}$ is continuous.

Conclusions and Future Directions

We discuss two new extensions for the traditional Voronoi density estimator (VDE) based on "kernel tricks" and rays from the generator p in a Voronoi cell $C(p)$.

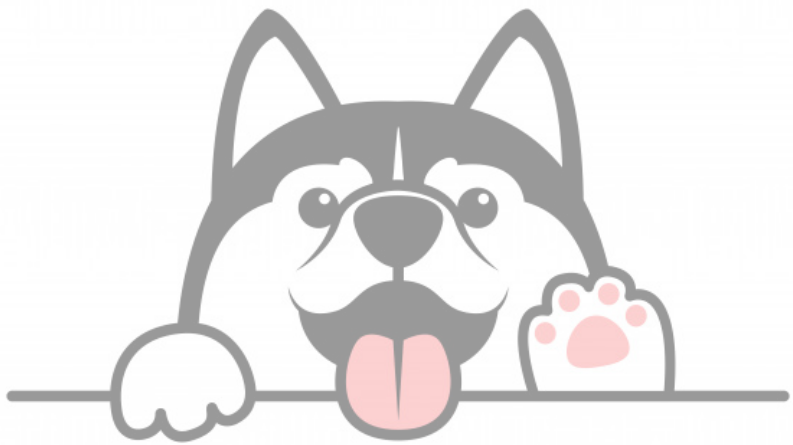
1. Compactified Voronoi density estimator (CVDE):
   - Monte Carlo method for computing the volume of $V(p)$.
   - Hit-and-run method for sampling from $\widehat{f}_{CVDE}$.

2. Radial Voronoi density estimator (RVDE):
   - Inherit the computational benefits from CVDE.
   - $\widehat{f}_{RVDE}$ is continuous.

Future directions: Extend CVDE or RVDE to Riemannian manifolds, such as directional data (Mardia et al., 2000), hyperbolic spaces (Nickel and Kiela, 2017), etc.

- Define the rays through the exponential map of a given manifold.
- ...

Thank you!



Reference I

I. S. Abramson. On bandwidth variation in kernel estimates-a square root law. *The annals of Statistics*, pages 1217–1223, 1982.

I. Arasaratnam, S. Haykin, and R. J. Elliott. Discrete-time nonlinear filtering algorithms using gauss–hermite quadrature. *Proceedings of the IEEE*, 95(5):953–977, 2007.

C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag, Berlin, Heidelberg, 2006.

Y.-C. Chen. A tutorial on kernel density estimation and recent advances. *Biostatistics & Epidemiology*, 1 (1):161–187, 2017.

R. M. Corless, G. H. Gonnet, D. E. Hare, D. J. Jeffrey, and D. E. Knuth. On the lambert w function. *Advances in Computational mathematics*, 5:329–359, 1996.

C. Fefferman, S. Mitter, and H. Narayanan. Testing the manifold hypothesis. *Journal of the American Mathematical Society*, 29(4):983–1049, 2016.

K. R. Gabriel and R. R. Sokal. A new statistical approach to geographic variation analysis. *Systematic zoology*, 18(3):259–278, 1969.

D. O. Loftsgaarden and C. P. Quesenberry. A nonparametric estimate of a multivariate density function. *The Annals of Mathematical Statistics*, 36(3):1049–1051, 1965.

G. L. Marchetti, V. Polianskii, A. Varava, F. T. Pokorny, and D. Kragic. An efficient and continuous voronoi density estimator. *arXiv preprint arXiv:2210.03964*, 2022. URL https://arxiv.org/abs/2210.03964.

K. V. Mardia, P. E. Jupp, and K. Mardia. *Directional statistics*, volume 2. Wiley Online Library, 2000.

M. M. Moradi, O. Cronie, E. Rubak, R. Lachieze-Rey, J. Mateu, and A. Baddeley. Resample-smoothing of voronoi intensity estimators. *Statistics and computing*, 29(5):995–1010, 2019.

M. Nickel and D. Kiela. Poincaré embeddings for learning hierarchical representations. *Advances in neural information processing systems*, 30, 2017.

Reference II

I. Olkin and C. H. Spiegelman. A semiparametric approach to density estimation. *Journal of the American Statistical Association*, 82(399):858–865, 1987.

J. Ord. How many trees in a forest. *Mathematical Scientist*, 3:23–33, 1978.

E. Platen, R. van de Weygaert, B. J. Jones, G. Vegter, and M. A. A. Calvo. Structural analysis of the sdss cosmic web–i. non-linear density field reconstructions. *Monthly Notices of the Royal Astronomical Society*, 416(4):2494–2526, 2011.

V. Polianskii and F. T. Pokorny. Voronoi boundary classification: A high-dimensional geometric approach via weighted monte carlo integration. In *International Conference on Machine Learning*, pages 5162–5170. PMLR, 2019.

V. Polianskii, G. L. Marchetti, A. Kravberg, A. Varava, F. T. Pokorny, and D. Kragic. Voronoi density estimator for high-dimensional data: Computation, compactification and convergence. In *Uncertainty in Artificial Intelligence*, pages 1644–1653. PMLR, 2022. URL https://arxiv.org/abs/2206.08051.

S. J. Sheather and M. C. Jones. A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(3):683–690, 1991.

B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London, 1986.

C. J. Stone. An asymptotically optimal window selection rule for kernel density estimates. *The Annals of Statistics*, pages 1285–1297, 1984.

G. R. Terrell and D. W. Scott. Variable kernel density estimation. *The Annals of Statistics*, pages 1236–1265, 1992.

B. Wang and X. Wang. Bandwidth selection for weighted kernel density estimation. *arXiv preprint arXiv:0709.1616*, 2007.

Reference III

Z. Wei and Y.-C. Chen. Skeleton clustering: Dimension-free density-aided clustering. *Journal of the American Statistical Association*, (just-accepted):1–30, 2023.

Y. Zhang, R. S. de Souza, and Y.-C. Chen. sconce: a cosmic web finder for spherical and conic geometries. *Monthly Notices of the Royal Astronomical Society*, 517(1):1197–1217, 2022.

P. Zhao and L. Lai. Analysis of knn density estimation. *IEEE Transactions on Information Theory*, 68(12): 7971–7995, 2022.

VDE / Delaunay Tessellation Field Estimator

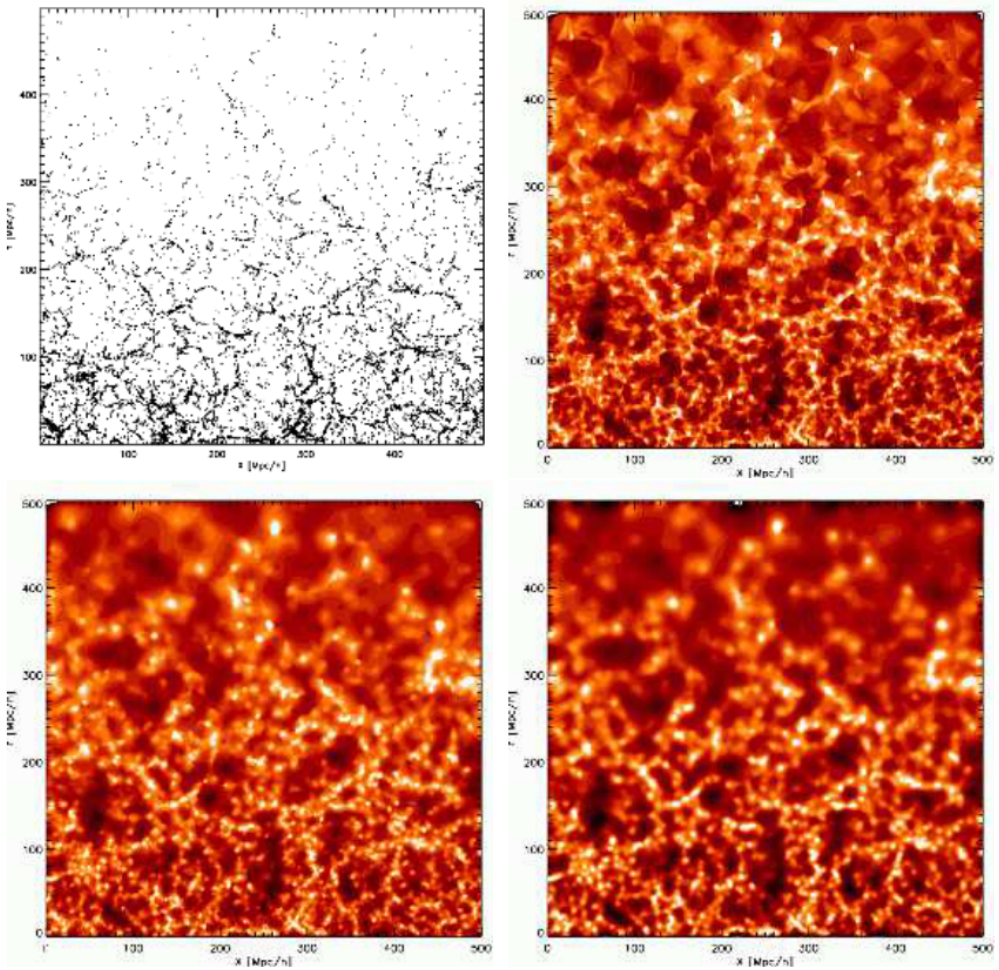


Figure 17: Density Reconstructions of SDSS mock catalogue. Top left: Millennium galaxy mock sample. Top right: DTFE reconstruction. Bottom left: NNFE reconstruction. Bottom right: lognormal kriging reconstruction (Platen et al., 2011).

Comparisons of DisPerSE with Our DirSCMS Algorithm

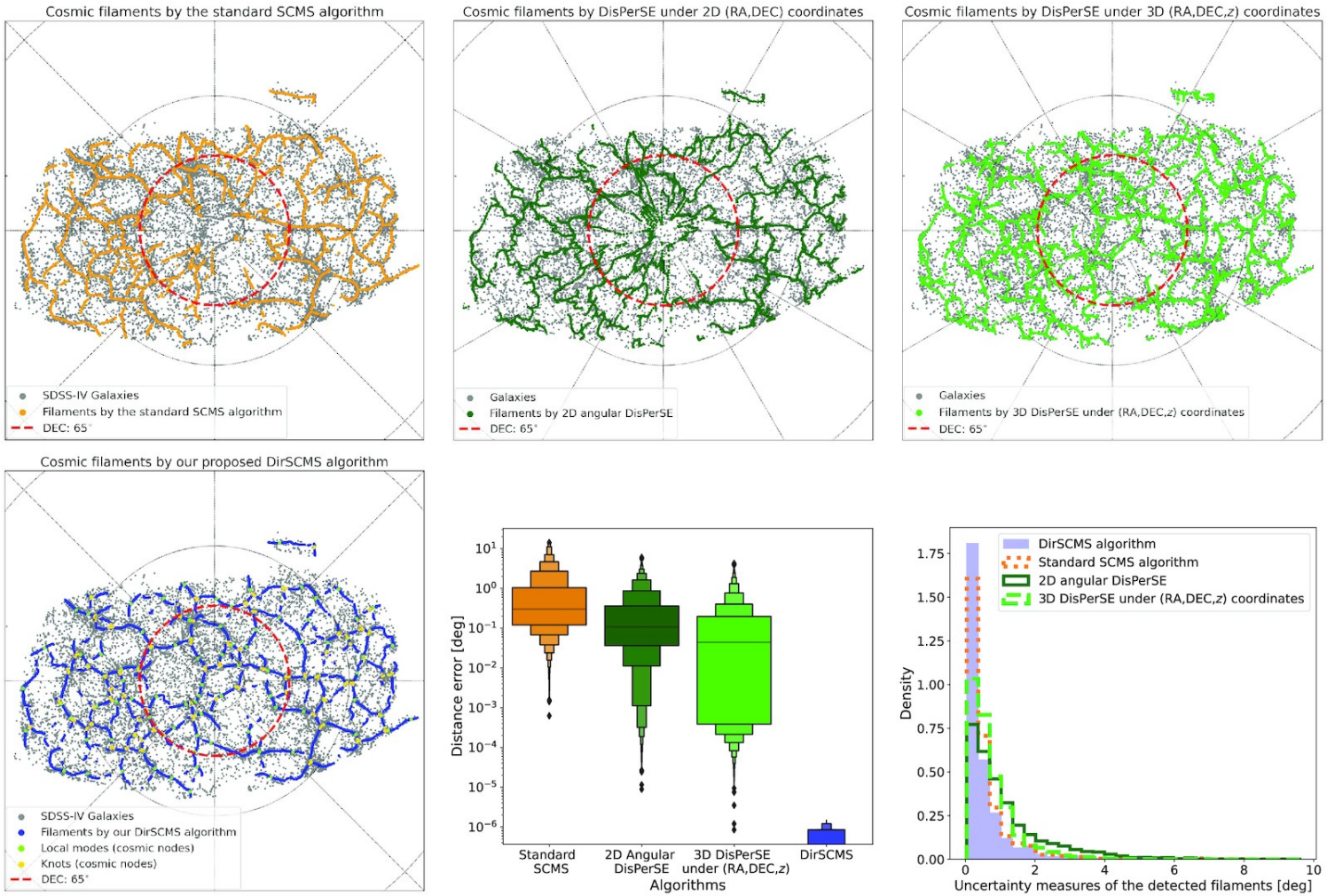


Figure 18: Comparisons between cosmic filaments detected by the standard SCMS, DisPerSE, and our proposed DirSCMS algorithms on the mock galaxy sample of SDSS-IV in the redshift slice $0.05 \leq z < 0.055$ (Zhang et al., 2022).