Analyzing GPS datasets using density ranking

Yen-Chi Chen

Department of Statistics
University of Washington

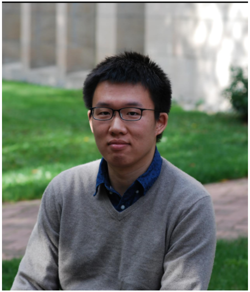
Outline

- Introduction
- Density Ranking
- Density Ranking: Multiple Datasets
- Summary

Introduction

Collaborators

Adrian Dobra

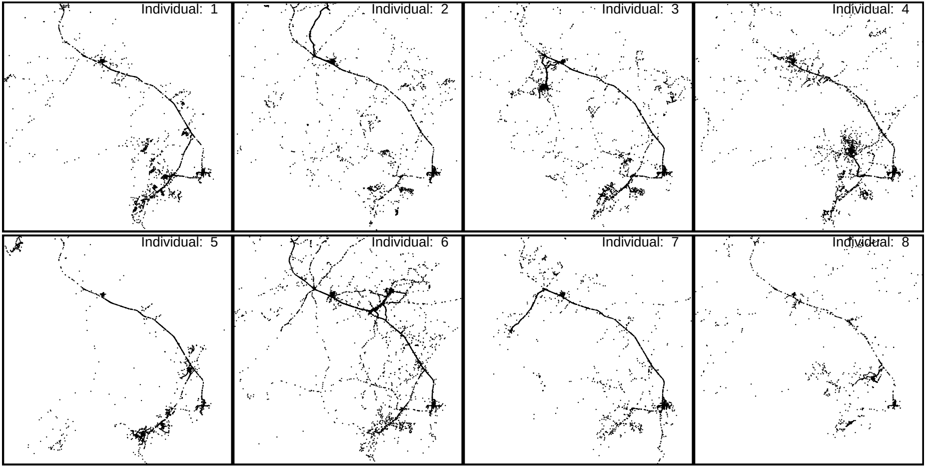


Zhihang Dong

Real Person Datasets

- This data is about 10 real person's GPS records from Chen and Dobra (2017).
- All these participants share the same work place.
- The ages of the study participants were between 34 and 48 years.
- Each person has around $3,500$ to $8,500$ GPS records.

Real Persons Datasets: Raw Data



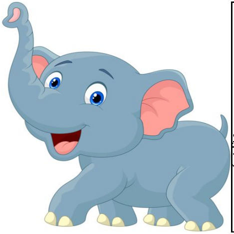
African Animal Datasets

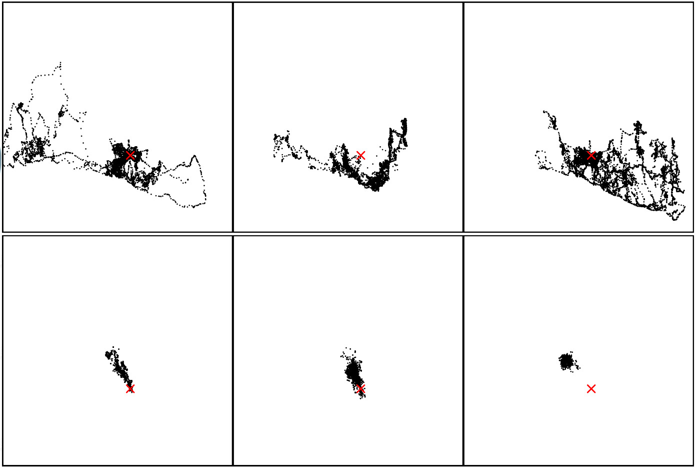
- This data is from the Movebank Data Repository[1] and was analyzed in Abrahms et al. (2017).
- Here we compare 4 different types of animals: elephants, jackals, vultures, and zebras.
- In this data, we have 8 elephants, 15 jackals, 10 vultures, and 9 zebras.
- Each animal has a set of GPS records with record size ranging from $1,000$ to $10,000$.

[1] https://www.datarepository.movebank.org/

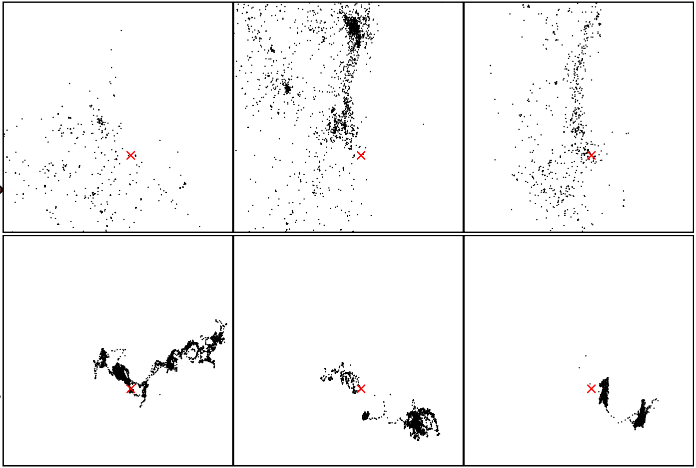
African Animal Datasets: Raw Data

African Animal Datasets: Raw Data





African Animal Datasets: Raw Data



Density Ranking

A Naive Idea: KDE

- Given a collection of points, a common statistical approach is to estimate the probability density function (PDF).

A Naive Idea: KDE

- Given a collection of points, a common statistical approach is to estimate the probability density function (PDF).
- Based on the estimated density function, we can then compare these datasets.

A Naive Idea: KDE

- Given a collection of points, a common statistical approach is to estimate the probability density function (PDF).
- Based on the estimated density function, we can then compare these datasets.
- Here we will use a simple approach called the kernel density estimation (KDE).

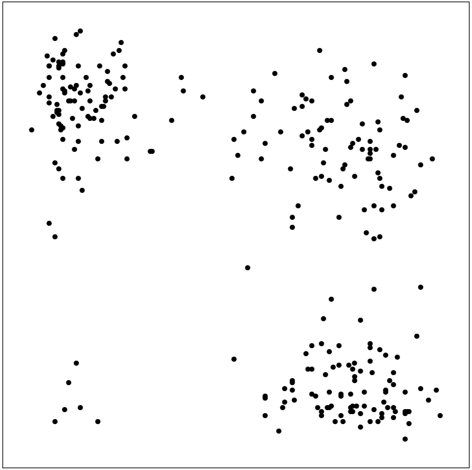
A Naive Idea: KDE

- Given a collection of points, a common statistical approach is to estimate the probability density function (PDF).
- Based on the estimated density function, we can then compare these datasets.
- Here we will use a simple approach called the kernel density estimation (KDE).
- At a given point x, KDE returns a density value

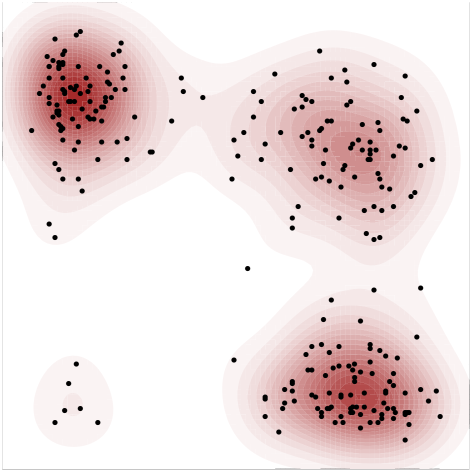
$$\widehat{p}(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right),$$

where $K(\cdot)$ is the kernel function that is often a smooth function like a Gaussian, and $h > 0$ is the smoothing bandwidth that controls the amount of smoothing.

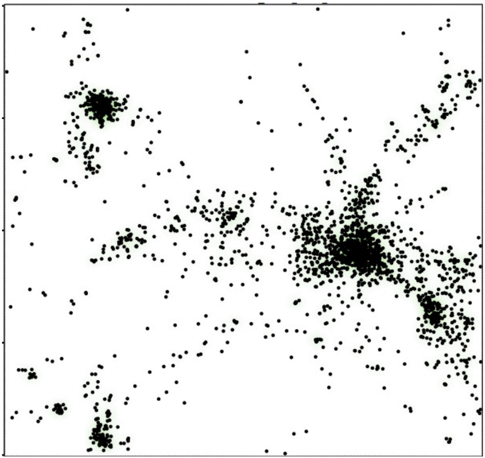
KDE: an Illustration



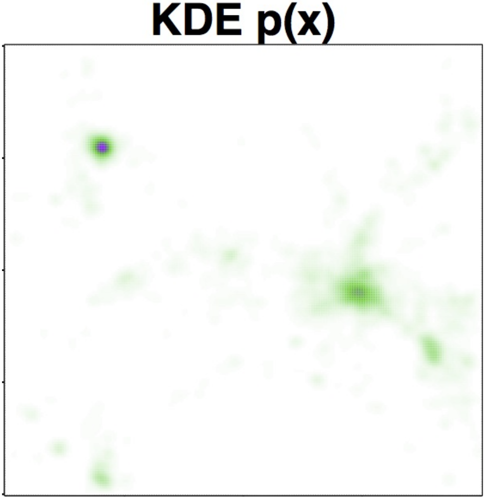
KDE: an Illustration



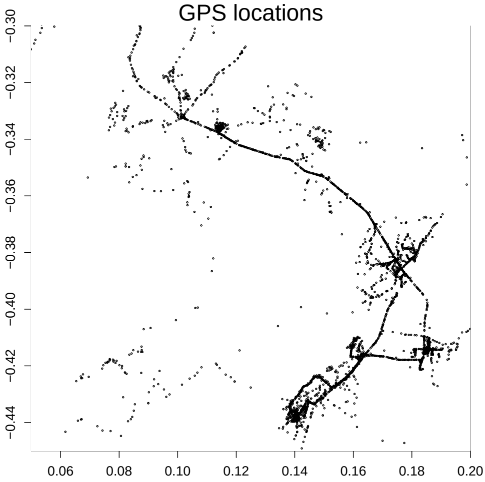
KDE: a More Complex Dataset (Astronomy)



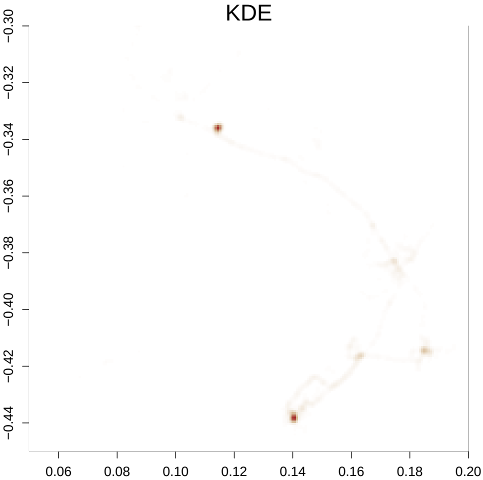
KDE: a More Complex Dataset (Astronomy)



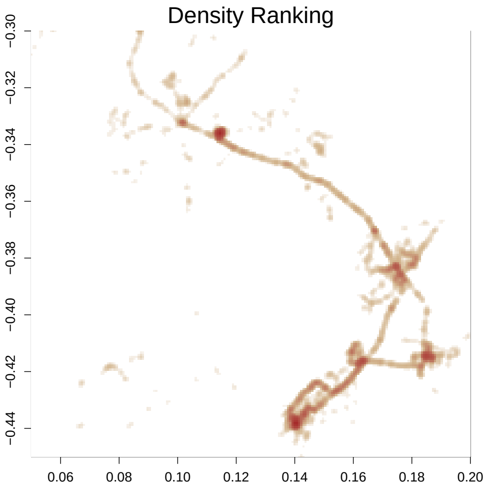
KDE and GPS data



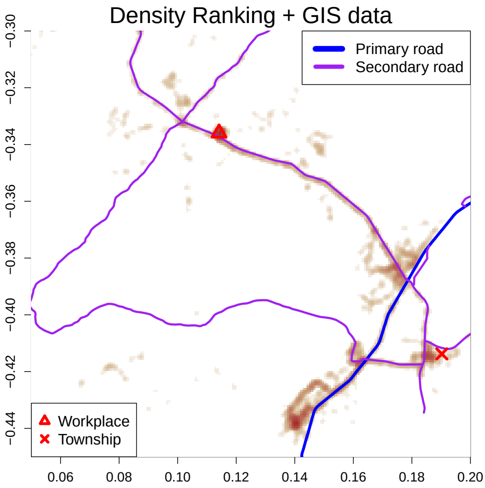
KDE and GPS data



KDE and GPS data



KDE and GPS data



Density Ranking: Introduction

- The KDE cannot detect intricate structures inside the GPS data.

Density Ranking: Introduction

- The KDE cannot detect intricate structures inside the GPS data.
- This is because the underlying PDF does not exist!
- Namely, our probability distribution function is singular.

Density Ranking: Introduction

- The KDE cannot detect intricate structures inside the GPS data.
- This is because the underlying PDF does not exist!
- Namely, our probability distribution function is singular.
- However, density ranking still works!

Definition of Density Ranking

- The density ranking (Chen 2016; Chen and Dobra 2017) is a transformed quantity/function from the KDE.
- Instead of using the density value, we focus on the *ranking* of it.

Definition of Density Ranking

- The density ranking (Chen 2016; Chen and Dobra 2017) is a transformed quantity/function from the KDE.
- Instead of using the density value, we focus on the *ranking* of it.
- The density ranking at point x is

$$\widehat{\alpha}(x) = \frac{1}{n}\sum_{i=1}^{n} I\left(\widehat{p}(x) \geq \widehat{p}(X_i)\right)$$

$$= \text{ratio of observations' density below the density of point } x.$$

Definition of Density Ranking

- The density ranking (Chen 2016; Chen and Dobra 2017) is a transformed quantity/function from the KDE.
- Instead of using the density value, we focus on the *ranking* of it.
- The density ranking at point x is

$$\widehat{\alpha}(x) = \frac{1}{n}\sum_{i=1}^{n} I\left(\widehat{p}(x) \geq \widehat{p}(X_i)\right)$$
$$= \text{ratio of observations' density below the density of point } x.$$

- Namely, $\widehat{\alpha}(x) = 0.3$ implies that the (estimated) density of point x is above the (estimated) density of 30% of all observations.

Property of Density Ranking

- For an observation $X_{\max}$ with $\widehat{\alpha}(X_{\max}) = 1$, then it means

$$\widehat{p}(X_{\max}) = \max\left\{\widehat{p}(X_1), \cdots, \widehat{p}(X_n)\right\}.$$

Property of Density Ranking

- For an observation $X_{\max}$ with $\widehat{\alpha}(X_{\max}) = 1$, then it means
$$\widehat{p}(X_{\max}) = \max\left\{\widehat{p}(X_1), \cdots, \widehat{p}(X_n)\right\}.$$
- Similarly, for an observation $X_{\min}$ with $\widehat{\alpha}(X_{\min}) = \frac{1}{n}$,
$$\widehat{p}(X_{\min}) = \min\left\{\widehat{p}(X_1), \cdots, \widehat{p}(X_n)\right\}.$$

Property of Density Ranking

- For an observation $X_{\max}$ with $\widehat{\alpha}(X_{\max}) = 1$, then it means
$$\widehat{p}(X_{\max}) = \max\left\{\widehat{p}(X_1), \cdots, \widehat{p}(X_n)\right\}.$$
- Similarly, for an observation $X_{\min}$ with $\widehat{\alpha}(X_{\min}) = \frac{1}{n}$,
$$\widehat{p}(X_{\min}) = \min\left\{\widehat{p}(X_1), \cdots, \widehat{p}(X_n)\right\}.$$
- If an observation X_ℓ satisfies $\widehat{\alpha}(X_\ell) = 0.25$, this means that the ranking of density at X_ℓ is the 25%.

Property of Density Ranking

- For an observation $X_{\max}$ with $\widehat{\alpha}(X_{\max}) = 1$, then it means

$$\widehat{p}(X_{\max}) = \max\left\{\widehat{p}(X_1), \cdots, \widehat{p}(X_n)\right\}.$$

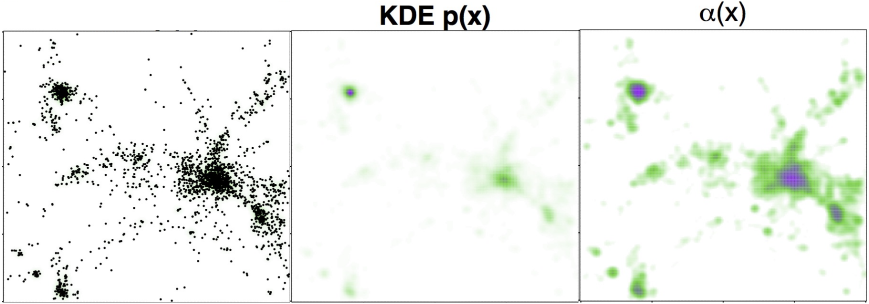
- Similarly, for an observation $X_{\min}$ with $\widehat{\alpha}(X_{\min}) = \frac{1}{n}$,

$$\widehat{p}(X_{\min}) = \min\left\{\widehat{p}(X_1), \cdots, \widehat{p}(X_n)\right\}.$$

- If an observation X_ℓ satisfies $\widehat{\alpha}(X_\ell) = 0.25$, this means that the ranking of density at X_ℓ is the 25%.
- Moreover, for any pairs of data points X_i, X_j,

$$\widehat{p}(X_i) > \widehat{p}(X_j) \Longrightarrow \widehat{\alpha}(X_i) > \widehat{\alpha}(X_j)$$
$$\widehat{p}(X_i) < \widehat{p}(X_j) \Longrightarrow \widehat{\alpha}(X_i) < \widehat{\alpha}(X_j)$$
$$\widehat{p}(X_i) = \widehat{p}(X_j) \Longrightarrow \widehat{\alpha}(X_i) = \widehat{\alpha}(X_j)$$

Density Ranking: Astronomy data



Density Ranking as an Estimator

- Density ranking $\widehat{\alpha}(x)$ can be viewed as an estimator to certain characteristics of the underlying population distribution.

Density Ranking as an Estimator

- Density ranking $\widehat{\alpha}(x)$ can be viewed as an estimator to certain characteristics of the underlying population distribution.
- When the distribution function has a PDF, the population version of density ranking is defined as:

$$\alpha(x) = P(p(x) \geq p(X_1)).$$

Density Ranking in Singular Measures

- Density ranking is still a consistent estimator *even when the density does not exist!*

Density Ranking in Singular Measures

- Density ranking is still a consistent estimator *even when the density does not exist!*
- To generalize population density ranking to a singular measure, we introduce the concept of the *Hausdorff (geometric) density*.

Density Ranking in Singular Measures

- Density ranking is still a consistent estimator *even when the density does not exist!*
- To generalize population density ranking to a singular measure, we introduce the concept of the *Hausdorff (geometric) density*.
- Let C_d be the volume of a d dimensional unit ball and $B(x, r) = \{y : \|x - y\| \leq r\}$.

Density Ranking in Singular Measures

- Density ranking is still a consistent estimator *even when the density does not exist!*
- To generalize population density ranking to a singular measure, we introduce the concept of the *Hausdorff (geometric) density*.
- Let C_d be the volume of a d dimensional unit ball and $B(x, r) = \{y : \|x - y\| \leq r\}$.
- For any integer s, we define

$$\mathscr{H}_s(x) = \lim_{r \to 0} \frac{P(B(x, r))}{C_s r^s}.$$

Density Ranking in Singular Measures

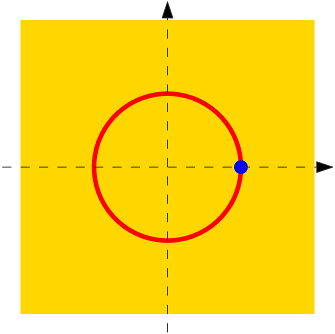
- Density ranking is still a consistent estimator *even when the density does not exist!*
- To generalize population density ranking to a singular measure, we introduce the concept of the *Hausdorff (geometric) density*.
- Let C_d be the volume of a d dimensional unit ball and $B(x,r) = \{y : \|x-y\| \leq r\}$.
- For any integer s, we define

$$\mathcal{H}_s(x) = \lim_{r\to 0} \frac{P(B(x,r))}{C_s r^s}.$$

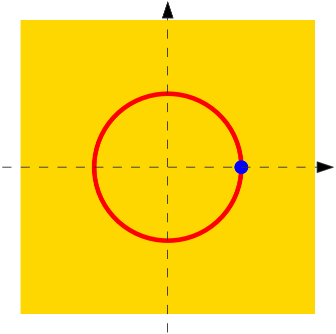
- For a point x, we then define

$$\tau(x) = \max\{s \leq d : \mathcal{H}_s(x) < \infty\}, \quad \rho(x) = \mathcal{H}_{\tau(x)}(x).$$

- Assume the distribution function P is a mixture of a $2D$ uniform distribution within $[-1, 1]^2$, a $1D$ uniform distribution over the ring $\{(x, y) : x^2 + y^2 = 0.5^2\}$, and a point mass at $(0.5, 0)$, then the support can be partitioned as follows:



Geometric Hausdorff: Example - 2



- Orange region: $\tau(x) = 2$.
- Red region: $\tau(x) = 1$.
- Blue region: $\tau(x) = 0$.

Hausdorff Density and Ranking

- The function $\tau(x)$ measures the dimension of P at point x.
- The function $\rho(x)$ describes the density of that corresponding dimension.

Hausdorff Density and Ranking

- The function $\tau(x)$ measures the dimension of P at point x.
- The function $\rho(x)$ describes the density of that corresponding dimension.
- We can use τ and ρ to compare any pairs of points and construct a ranking.

Hausdorff Density and Ranking

- The function $\tau(x)$ measures the dimension of P at point x.
- The function $\rho(x)$ describes the density of that corresponding dimension.
- We can use τ and ρ to compare any pairs of points and construct a ranking.
- For two points x_1, x_2, we define an ordering such that $x_1 \succ_{\tau,\rho} x_2$ if

$$\tau(x_1) < \tau(x_2), \qquad \text{or} \quad \tau(x_1) = \tau(x_2), \quad \rho(x_1) > \rho(x_2).$$

Hausdorff Density and Ranking

- The function $\tau(x)$ measures the dimension of P at point x.
- The function $\rho(x)$ describes the density of that corresponding dimension.
- We can use τ and ρ to compare any pairs of points and construct a ranking.
- For two points x_1, x_2, we define an ordering such that $x_1 \succ_{\tau,\rho} x_2$ if

$$\tau(x_1) < \tau(x_2), \qquad \text{or} \quad \tau(x_1) = \tau(x_2), \quad \rho(x_1) > \rho(x_2).$$

- Namely, we first compare the dimension of the two points, the lower dimensional structure wins. If they are on regions of the same dimension, we then compare the density of that dimension.

Constructing Density Ranking using Hausdorff Density

- Using the ordering $\succ_{\tau,\rho}$, we then define the population density ranking as

$$\alpha(x) = P(x \succeq_{\tau,\rho} X_1)$$

Constructing Density Ranking using Hausdorff Density

- Using the ordering $\succ_{\tau,\rho}$, we then define the population density ranking as

$$\alpha(x) = P(x \succeq_{\tau,\rho} X_1)$$

- When the PDF exists, the ordering $\succ_{\tau,\rho}$ equals to $\succ_{d,p}$ so

$$\alpha(x) = P(x \succeq_{d,p} X_1) = P(p(x) \geq p(X_1)),$$

which recovers our original definition.

Convergence under Singular Measure: Density Ranking - 1

- When P is a singular distribution and satisfies certain regularity conditions,
$$\int \left|\widehat{\alpha}(x) - \alpha(x)\right|^2 dP(x) \xrightarrow{P} 0.$$
- Note that here $\widehat{\alpha}(x)$ is still the same estimator from the KDE.

Convergence under Singular Measure: Density Ranking - 1

- When P is a singular distribution and satisfies certain regularity conditions,
$$\int \left|\widehat{\alpha}(x) - \alpha(x)\right|^2 dP(x) \xrightarrow{P} 0.$$
- Note that here $\widehat{\alpha}(x)$ is still the same estimator from the KDE.
- Ideas: the KDE
$$\widehat{p}(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)$$
diverges when x is in a lower dimensional space $\tau(x) < d$ as $h \to 0$.

Convergence under Singular Measure: Density Ranking - 1

- When P is a singular distribution and satisfies certain regularity conditions,
$$\int \left|\widehat{\alpha}(x) - \alpha(x)\right|^2 dP(x) \xrightarrow{P} 0.$$
- Note that here $\widehat{\alpha}(x)$ is still the same estimator from the KDE.
- Ideas: the KDE
$$\widehat{p}(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)$$
diverges when x is in a lower dimensional space $\tau(x) < d$ as $h \to 0$.
- However, the speed of diverging depends on $\tau(x)$. The smaller $\tau(x)$, the faster (actually the diverging rate is $O(h^{\tau(x)-d})$).

Convergence under Singular Measure: Density Ranking - 1

- When P is a singular distribution and satisfies certain regularity conditions,
$$\int \left|\widehat{\alpha}(x) - \alpha(x)\right|^2 dP(x) \xrightarrow{P} 0.$$
- Note that here $\widehat{\alpha}(x)$ is still the same estimator from the KDE.
- Ideas: the KDE
$$\widehat{p}(x) = \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)$$
diverges when x is in a lower dimensional space $\tau(x) < d$ as $h \to 0$.
- However, the speed of diverging depends on $\tau(x)$. The smaller $\tau(x)$, the faster (actually the diverging rate is $O(h^{\tau(x)-d})$).
- So eventually, we can separate different dimensional structures.

Convergence under Singular Measure: Density Ranking - 2

- Despite the pointwise convergence and convergence in $L_2(P)$, there no guarantee for the uniform convergence $\sup_x \left|\widehat{\alpha}(x) - \alpha(x)\right|$.

Convergence under Singular Measure: Density Ranking - 2

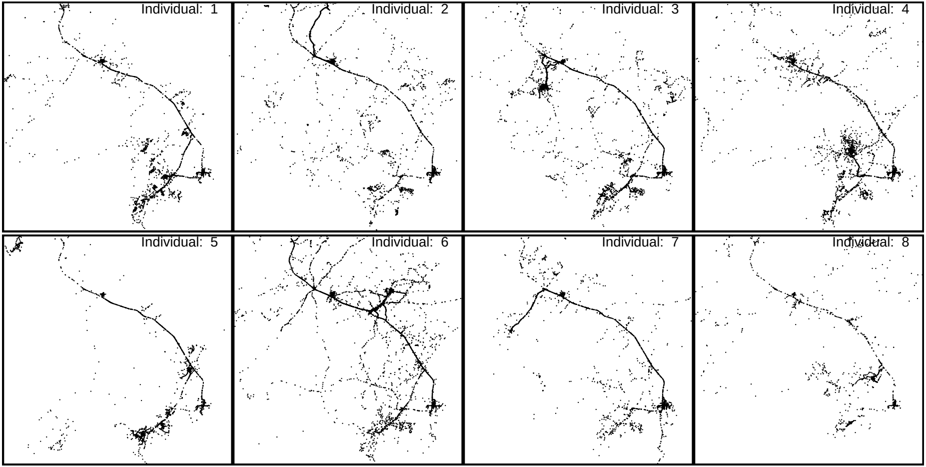
- Despite the pointwise convergence and convergence in $L_2(P)$, there no guarantee for the uniform convergence $\sup_x \left|\widehat{\alpha}(x) - \alpha(x)\right|$.
- Example of non-convergence of supreme norm: consider a sequence of points on a higher dimensional space but moving toward a lower dimensional space within distance $\frac{h}{2}$.

Convergence under Singular Measure: Density Ranking - 3

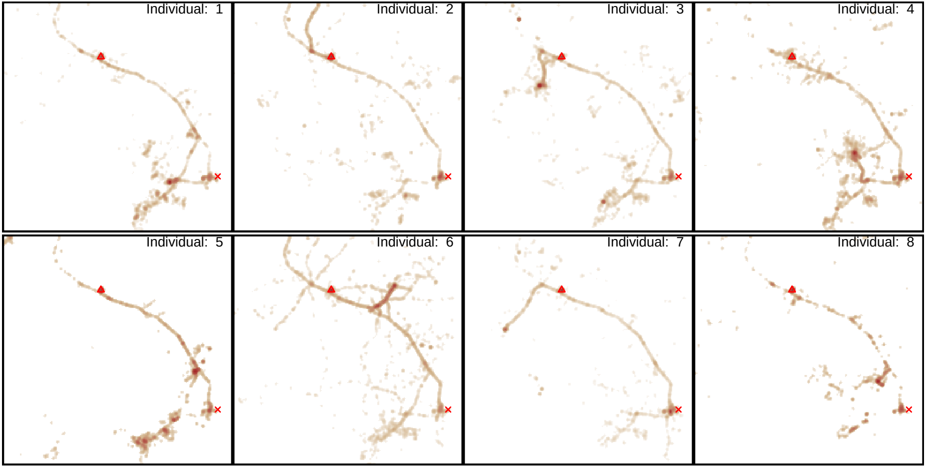
- A technical remark: $\widehat{\alpha}$ also converges to α is a 'topological' sense.
- More specifically, the cluster tree and persistent diagram of $\widehat{\alpha}$ both converge to the cluster tree and persistent diagram of α in certain metrics.

Density Ranking: Multiple Datasets

Application of Density Ranking: GPS dataset - 1



Application of Density Ranking: GPS dataset - 2



Summarizing Multiple Density Ranking: Level Plots - 1

- In the above example, we have multiple GPS datasets and each of them yields one density ranking.
- Thus, we have multiple density rankings.

Summarizing Multiple Density Ranking: Level Plots - 1

- In the above example, we have multiple GPS datasets and each of them yields one density ranking.
- Thus, we have multiple density rankings.
- To compare these density rankings, a simple approach is to overlap level plots.
- For a density ranking $\widehat{\alpha}$, let

$$\widehat{A}_\gamma = \{x : \widehat{\alpha}(x) \geq 1 - \gamma\}$$

be the (upper) level set.

Summarizing Multiple Density Ranking: Level Plots - 1

- In the above example, we have multiple GPS datasets and each of them yields one density ranking.
- Thus, we have multiple density rankings.
- To compare these density rankings, a simple approach is to overlap level plots.
- For a density ranking $\widehat{\alpha}$, let

$$\widehat{A}_{\gamma} = \{x : \widehat{\alpha}(x) \geq 1 - \gamma\}$$

  be the (upper) level set.
- We can compare the density ranking of each individual by overlapping their level sets at different levels.

Summarizing Multiple Density Ranking: Level Plots - 2

- Note that we use $1-\gamma$ as the level in the set $\widehat{A}_\gamma$.
- This is because such a set has a natural interpretation in activity space.
- Activity space: the spatial regions where an individual undertakes his/her daily life.

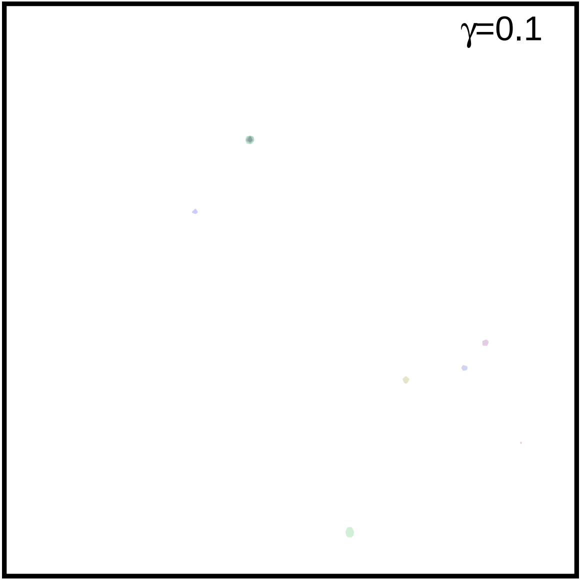
Summarizing Multiple Density Ranking: Level Plots - 2

- Note that we use $1-\gamma$ as the level in the set $\widehat{A}_\gamma$.
- This is because such a set has a natural interpretation in activity space.
- Activity space: the spatial regions where an individual undertakes his/her daily life.
- We can interpret $\widehat{A}_\gamma$ as the (top) $\gamma \cdot 100\%$ activity space because it contains at least $\gamma \cdot 100\%$ GPS records.

Summarizing Multiple Density Ranking: Level Plots - 2

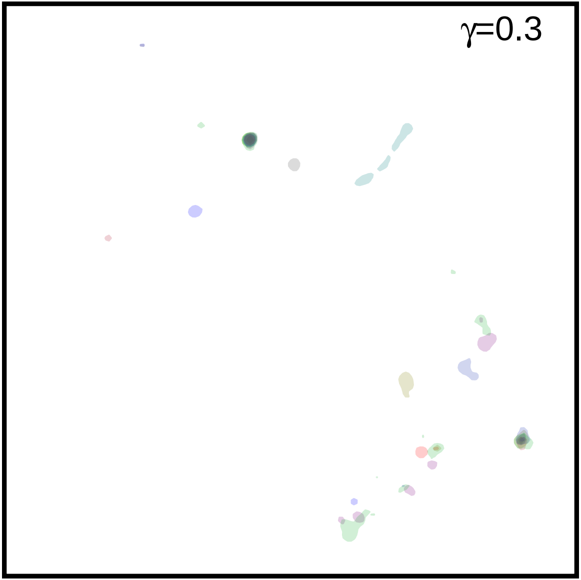
- Note that we use $1 - \gamma$ as the level in the set $\widehat{A}_\gamma$.
- This is because such a set has a natural interpretation in activity space.
- Activity space: the spatial regions where an individual undertakes his/her daily life.
- We can interpret $\widehat{A}_\gamma$ as the (top) $\gamma \cdot 100\%$ activity space because it contains at least $\gamma \cdot 100\%$ GPS records.
- Namely, $\widehat{A}_{\gamma=0.3}$ is the (top) 30% activity space.

Level Plots: Example

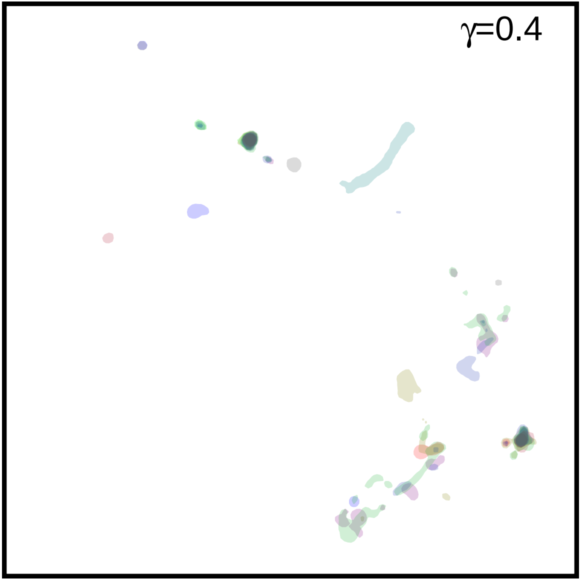


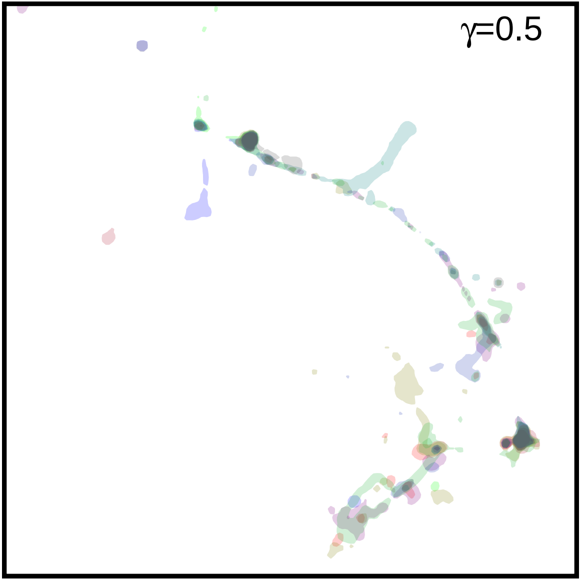
Level Plots: Example

Level Plots: Example

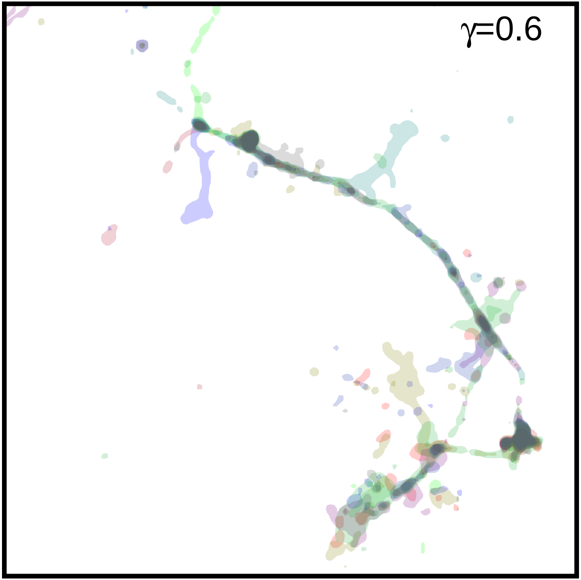


Level Plots: Example

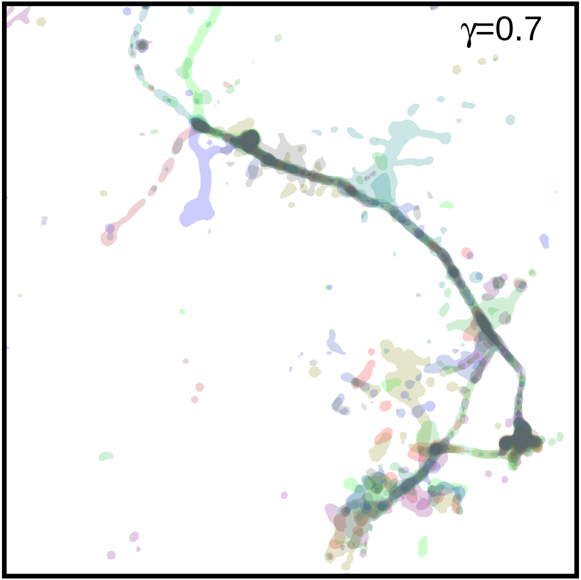



γ=0.5

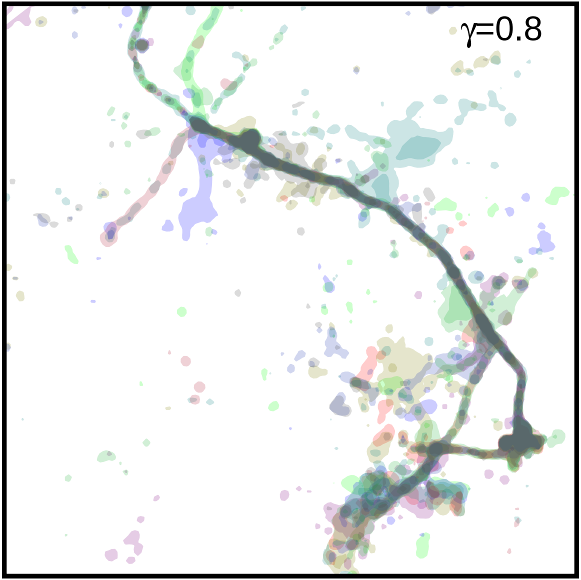
Level Plots: Example



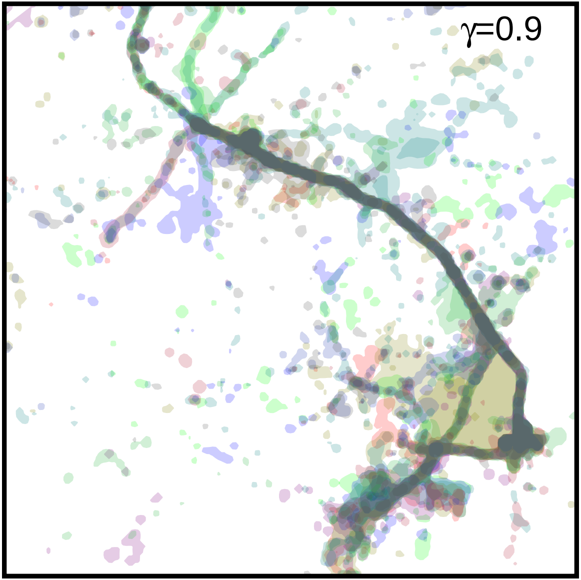
Level Plots: Example



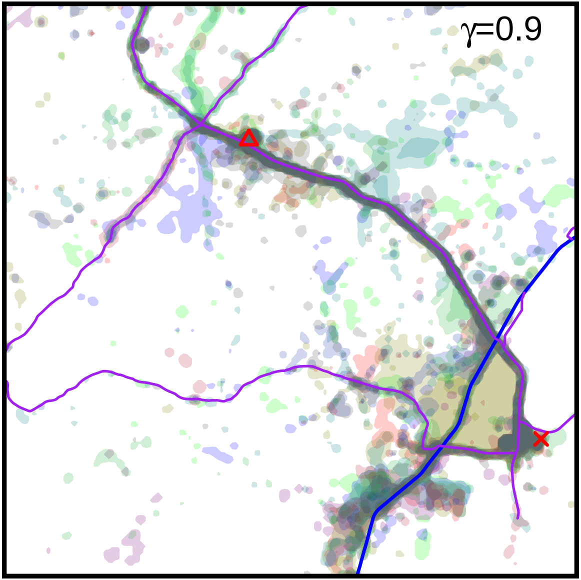
Level Plots: Example



Level Plots: Example



Level Plots: Example



Summary Curves of Density Ranking

- The level plot allows us to compare GPS datasets from different individuals.

Summary Curves of Density Ranking

- The level plot allows us to compare GPS datasets from different individuals.
- However, it has two drawbacks:
  - When we have many individuals, this approach might not work (too many contours).
  - We often need to choose a level γ to show the plot but which level to be chosen is unclear.

Summary Curves of Density Ranking

- The level plot allows us to compare GPS datasets from different individuals.
- However, it has two drawbacks:
  - When we have many individuals, this approach might not work (too many contours).
  - We often need to choose a level γ to show the plot but which level to be chosen is unclear.
- Here we introduce a few curves to summarize geometric and topological features of density ranking.

Mass-Volume Curve

- Recall that $\widehat{A}_\gamma = \{x : \widehat{\alpha}(x) \geq 1 - \gamma\}$ is the level set of density ranking.

Mass-Volume Curve

- Recall that $\widehat{A}_\gamma = \{x : \widehat{\alpha}(x) \geq 1 - \gamma\}$ is the level set of density ranking.
- The mass-volume curve is a curve of

$$\left(\gamma, \mathsf{Vol}(\widehat{A}_\gamma)\right) : \gamma \in [0, 1].$$

- Namely, we are plotting the size of set $\widehat{A}_\gamma$ at various level.

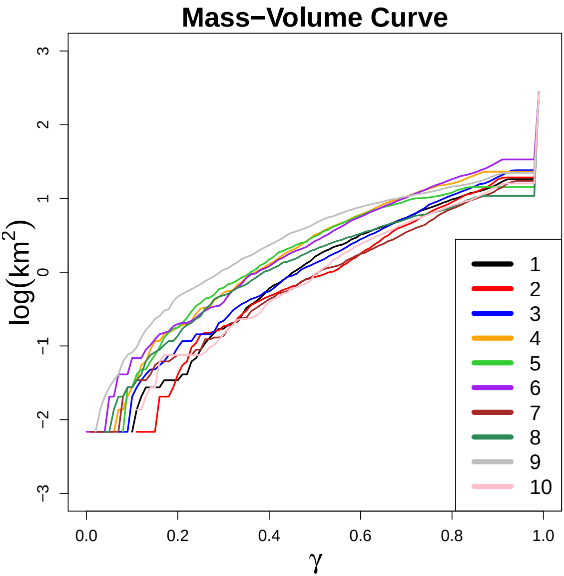
Mass-Volume Curve

- Recall that $\widehat{A}_\gamma = \{x : \widehat{\alpha}(x) \geq 1 - \gamma\}$ is the level set of density ranking.
- The mass-volume curve is a curve of

$$\left(\gamma, \mathsf{Vol}(\widehat{A}_\gamma)\right) : \gamma \in [0, 1].$$

- Namely, we are plotting the size of set $\widehat{A}_\gamma$ at various level.
- In practice, we often plot γ versus $\log \mathsf{Vol}(\widehat{A}_\gamma)$.

Mass-Volume Curve: Example



Betti Number Curve

- The Betti number curve is a curve quantifying topological features of the density ranking.
- It counts the number of connected components of $\widehat{A}_\gamma$ at various level γ.

Betti Number Curve

- The Betti number curve is a curve quantifying topological features of the density ranking.
- It counts the number of connected components of $\widehat{A}_\gamma$ at various level γ.
- Formally, the Betti number curve is

$$\left(\gamma, \mathsf{Betti}_0(\widehat{A}_\gamma)\right) : \gamma \in [0,1],$$

where for a set A

$$\mathsf{Betti}_0(A) = \text{ number of connected components inside } A.$$

Betti Number Curve

- The Betti number curve is a curve quantifying topological features of the density ranking.
- It counts the number of connected components of $\widehat{A}_\gamma$ at various level γ.
- Formally, the Betti number curve is

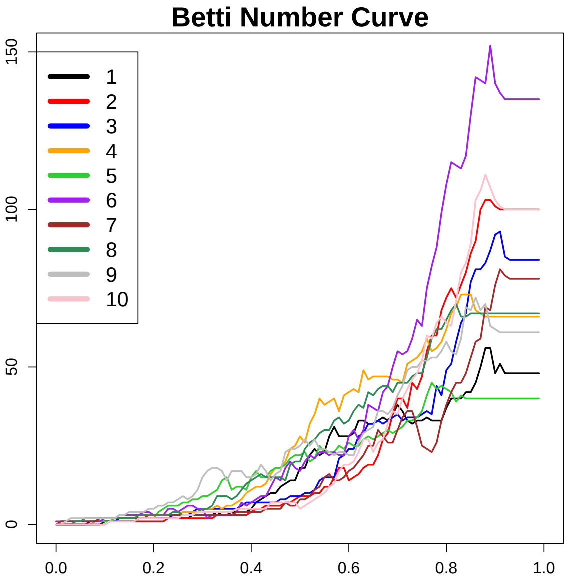
$$\left(\gamma, \mathsf{Betti}_0(\widehat{A}_\gamma)\right) : \gamma \in [0, 1],$$

where for a set A

$$\mathsf{Betti}_0(A) = \text{ number of connected components inside } A.$$

- Note that the number of connected component is called the 0th order Betti number (0th order topological structure); one can generalize this idea to higher order topological structures.

Betti Number Curve: Example

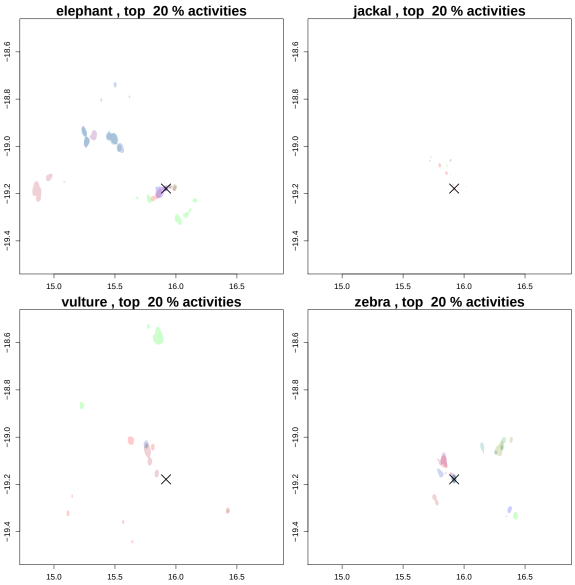


Applying to African Animal Datasets

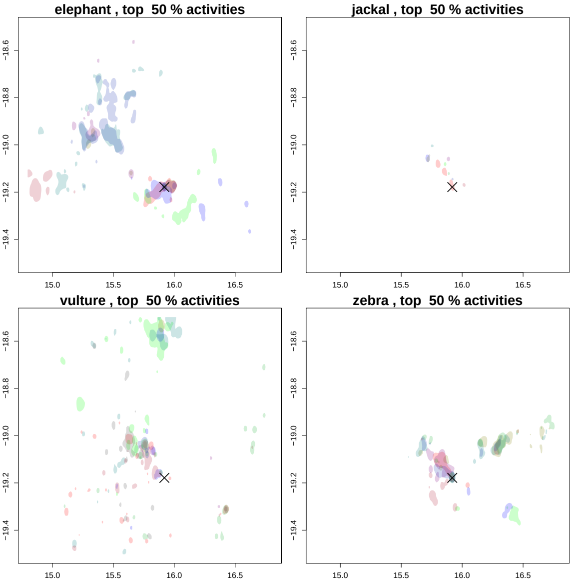
- Now we come back to our African Animal Datasets.
- We compare their level plots and summary curves.

Level Plots: Animal Example

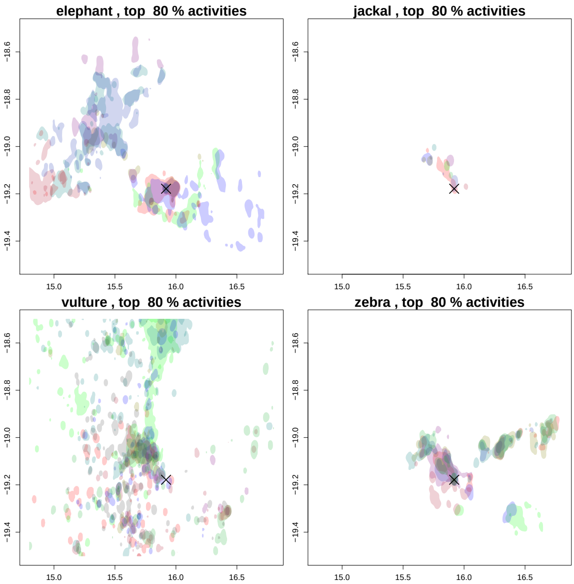
Level Plots: Animal Example



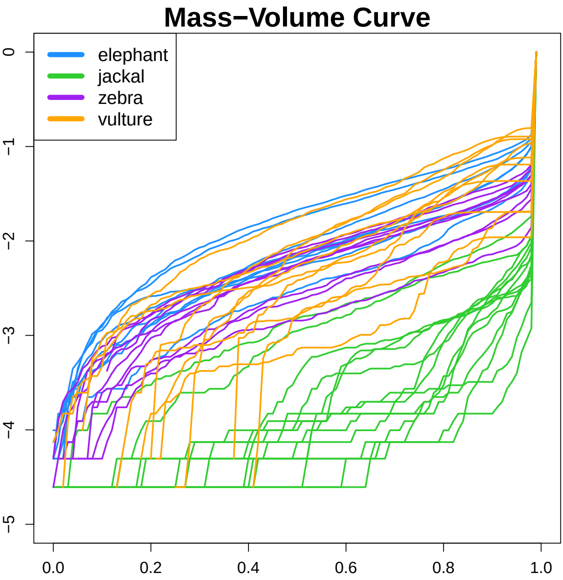
Level Plots: Animal Example



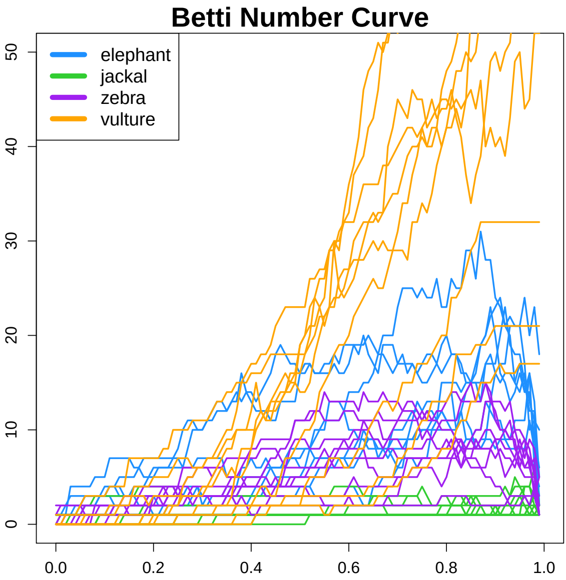
Level Plots: Animal Example



Mass-Volume Curve: Animal Example



Betti Number Curve: Animal Example



Summary

Summary

- For complex datasets such as GPS data, we cannot use the usual density estimation approach because the PDF does not exist.

Summary

- For complex datasets such as GPS data, we cannot use the usual density estimation approach because the PDF does not exist.
- But we can use density ranking to analyze data.

Summary

- For complex datasets such as GPS data, we cannot use the usual density estimation approach because the PDF does not exist.
- But we can use density ranking to analyze data.
- This quantity is stable and its estimator is statistically consistent.

Summary

- For complex datasets such as GPS data, we cannot use the usual density estimation approach because the PDF does not exist.
- But we can use density ranking to analyze data.
- This quantity is stable and its estimator is statistically consistent.
- Moreover, the contours of density ranking is related to activity spaces.

Summary

- For complex datasets such as GPS data, we cannot use the usual density estimation approach because the PDF does not exist.
- But we can use density ranking to analyze data.
- This quantity is stable and its estimator is statistically consistent.
- Moreover, the contours of density ranking is related to activity spaces.
- When multiple GPS datasets are available, we can summarize them by functional summaries of density ranking.

Summary

- For complex datasets such as GPS data, we cannot use the usual density estimation approach because the PDF does not exist.
- But we can use density ranking to analyze data.
- This quantity is stable and its estimator is statistically consistent.
- Moreover, the contours of density ranking is related to activity spaces.
- When multiple GPS datasets are available, we can summarize them by functional summaries of density ranking.
- Comparing these summaries gives us more insights about the data.

Future Work

- Sample size needed for GPS data analysis.
- Analysis of multiple summary curves.
- Analyzing activity spaces and covariates using density ranking.
- Measuring environmental exposure using density ranking.

Thank You!

An R script for density ranking:

`https://github.com/yenchic/density_ranking`

More details can be found in

`http://faculty.washington.edu/yenchic/`

References

1. Chen, Yen-Chi, Christopher R. Genovese, and Larry Wasserman. "Density level sets: Asymptotics, inference, and visualization." Journal of the American Statistical Association (2017): 1-13.
2. Jisu, K. I. M., Yen-Chi Chen, Sivaraman Balakrishnan, Alessandro Rinaldo, and Larry Wasserman. “Statistical inference for cluster trees." In Advances In Neural Information Processing Systems, pp. 1839-1847. 2016.
3. Chen, Yen-Chi. “Generalized Cluster Trees and Singular Measures." arXiv preprint arXiv:1611.02762 (2016).
4. Chen, Yen-Chi, and Adrian Dobra. “Measuring Human Activity Spaces With Density Ranking Based on GPS Data." arXiv preprint arXiv:1708.05017 (2017).
5. Stuetzle, Werner. “Estimating the cluster tree of a density by analyzing the minimal spanning tree of a sample." Journal of classification 20, no. 1 (2003): 025-047.
6. Klemelä, Jussi. “Visualization of multivariate density estimates with level set trees." Journal of Computational and Graphical Statistics 13, no. 3 (2004): 599-620.
7. Chaudhuri, Kamalika, and Sanjoy Dasgupta. “Rates of convergence for the cluster tree." In Advances in Neural Information Processing Systems, pp. 343-351. 2010.
8. Chaudhuri, Kamalika, Sanjoy Dasgupta, Samory Kpotufe, and Ulrike von Luxburg. “Consistent procedures for cluster tree estimation and pruning." IEEE Transactions on Information Theory 60, no. 12 (2014): 7900-7912.
9. Eldridge, Justin, Mikhail Belkin, and Yusu Wang. “Beyond hartigan consistency: Merge distortion metric for hierarchical clustering." In Conference on Learning Theory, pp. 588-606. 2015.
10. Balakrishnan, Sivaraman, Srivatsan Narayanan, Alessandro Rinaldo, Aarti Singh, and Larry Wasserman. “Cluster trees on manifolds." In Advances in Neural Information Processing Systems, pp. 2679-2687. 2013.
11. Abrahms B, Seidel DP, Dougherty E, Hazen EL, Bograd SJ, Wilson AM, McNutt JW, Costa DP, Blake S, Brashares JS, Getz WM (2017) “Suite of simple metrics reveals common movement syndromes across vertebrate taxa." Movement Ecology 5:12. doi:10.1186/s40462-017-0104-2

Assumptions for Regular Distributions

- **(R1)** The density function p has a compact support $\mathbb{K}$.
- **(R2)** The density function is a Morse function and is in $\mathbf{BC}^3$.
- **(K1)** The kernel function K is in $\mathbf{BC}^2$ and integrable.
- **(K2)** K satisfies the VC-type class condition.

(K2) Let

$$\mathcal{K}_r = \left\{ y \mapsto K^{(\alpha)}\left(\frac{x-y}{h}\right) : x \in \mathbb{R}^d, |\alpha| = r \right\},$$

where $K^{(\alpha)}$ is the α-th derivative and let $\mathcal{K}_l^* = \bigcup_{r=0}^{l} \mathcal{K}_r$. We assume that $\mathcal{K}_2^*$ is a VC-type class. i.e. there exists constants A, v and a constant envelope b_0 such that

$$\sup_Q N(\mathcal{K}_2^*, \mathcal{L}^2(Q), b_0\epsilon) \leq \left(\frac{A}{\epsilon}\right)^v, \tag{1}$$

where $N(T, d_T, \epsilon)$ is the ϵ-covering number for an semi-metric set T with metric d_T and $\mathcal{L}^2(Q)$ is the L_2 norm with respect to the probability measure Q.

Assumptions for Singular Distributions

(S1) The support can be partitioned into

$$K = K_0 \bigcup K_1 \bigcup \cdots \bigcup K_d,$$

where $K_\ell = \{x \in \mathbb{K} : \tau(x) = \ell\}$.

(S2) There exist $\rho_{\min}, \rho_{\max}$ such that $0 < \rho_{\min} \leq \rho(x) \leq \rho_{\max} < \infty$ for every $x \in \mathbb{K}$.

(S3) Restricted to each $\mathbb{K}_\ell$ where $\ell > 0$, $\rho(x)$ is a Morse function.

(K1') The kernel function K is in $\mathbf{BC}^2$, integrable, and supported in $[-1, 1]$.

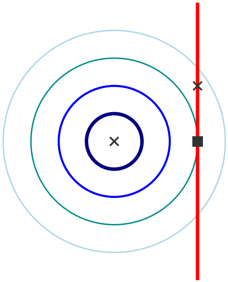
(K2) K satisfies the VC-type class condition.

Dimensional Critical Points - 1

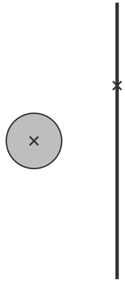
- In singular measure, there is a new type of critical points. We call them the *dimensional critical points*.
- These critical points contribute to the change of topology of level sets as the usual critical points but they cannot be defined by setting gradient to be 0.

Dimensional Critical Points - 2

- The box in the following figure is a dimensional critical point.
- Note: this is a mixture of 2D distribution and a 1D distribution on the black line (maximum value occurs at the cross).

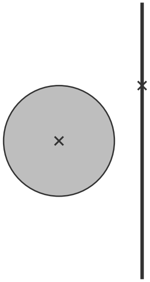


- The box in the following figure is a dimensional critical point.
- Note: this is a mixture of 2D distribution and a 1D distribution on the black line (maximum value occurs at the cross).



Dimensional Critical Points - 2

- The box in the following figure is a dimensional critical point.
- Note: this is a mixture of 2D distribution and a 1D distribution on the black line (maximum value occurs at the cross).



Dimensional Critical Points - 2

- The box in the following figure is a dimensional critical point.
- Note: this is a mixture of 2D distribution and a 1D distribution on the black line (maximum value occurs at the cross).