A Statistical Perspective on Coreset Density Estimation

Paxton Turner (MIT)

Jingbo Liu (UIUC)

Philippe Rigollet (MIT)

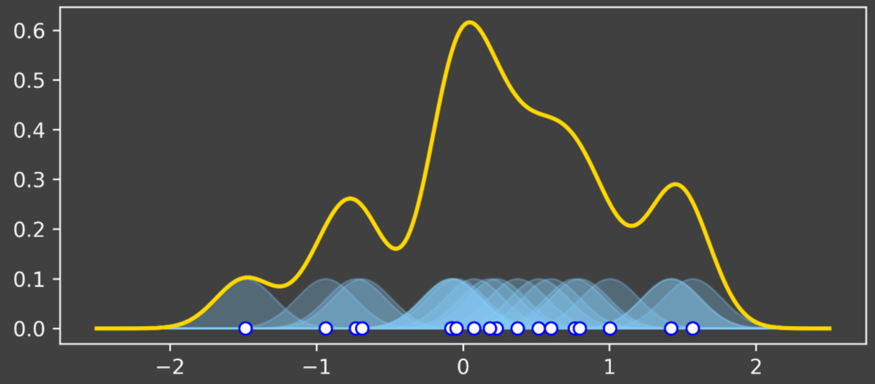
Density Estimation

- **Goal:** reconstruct a probability density function from data $X_1, \dots, X_n$

Density Estimation

- **Goal:** reconstruct a probability density function from data $X_1, \dots, X_n$
- **Kernel density estimator (KDE):**

$$\hat{f}(y) = \frac{1}{n} \sum_{j=1}^{n} K_h\left(X_j - y\right)$$

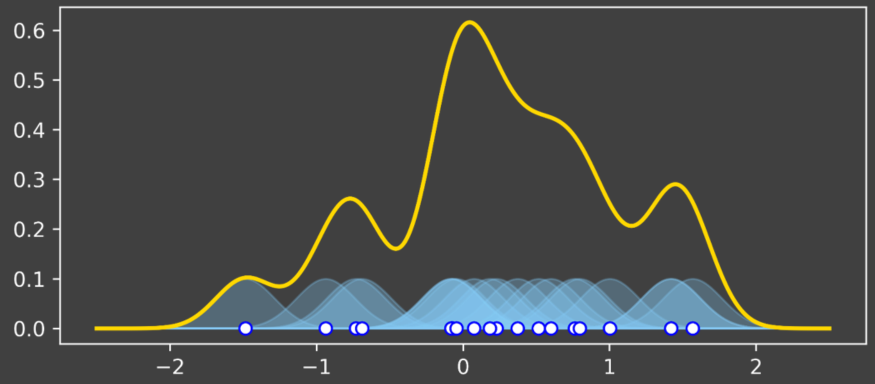


- **Fact:** KDE has rate of estimation $n^{-\frac{\beta}{2\beta+d}}$ over Hölder β densities.

Density Estimation

- **Goal:** reconstruct a probability density function from data $X_1, \dots, X_n$
- **Kernel density estimator (KDE):**

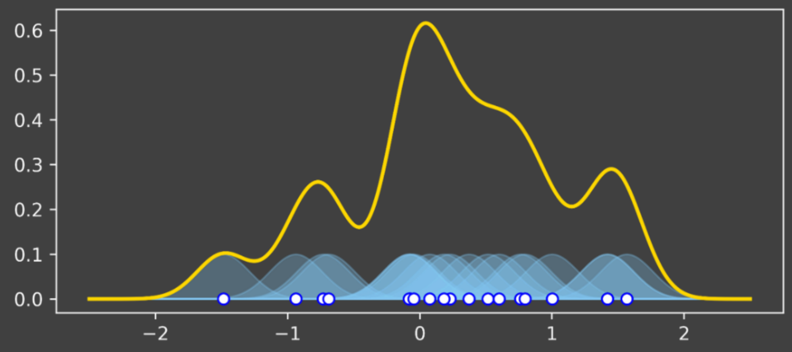
$$\hat{f}(y) = \frac{1}{n} \sum_{j=1}^{n} K_h\left(X_j - y\right)$$



- **Fact:** KDE has rate of estimation $n^{-\frac{\beta}{2\beta+d}}$ over Hölder β densities.
- **Question:** Can we improve on computational aspects?

KDE

$$\hat{f}(y) = \frac{1}{n}\sum_{j=1}^{n} K_h\left(X_j - y\right)$$

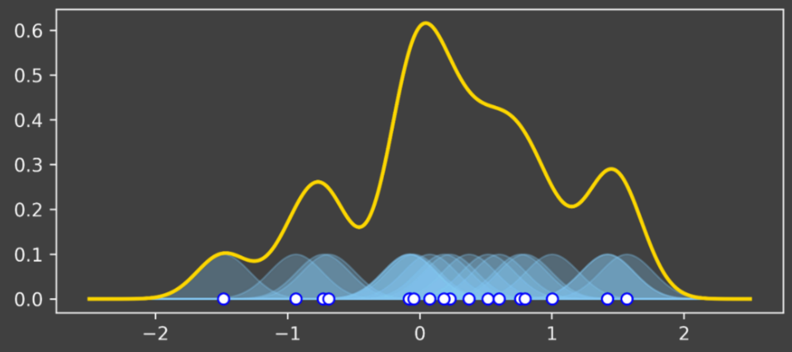


KDE

$$\hat{f}(y) = \frac{1}{n}\sum_{j=1}^{n} K_h\left(X_j - y\right)$$

$$= \mathbb{E}_{X\sim\mathbb{P}_n}\left[K_h(X - y)\right]$$

empirical measure

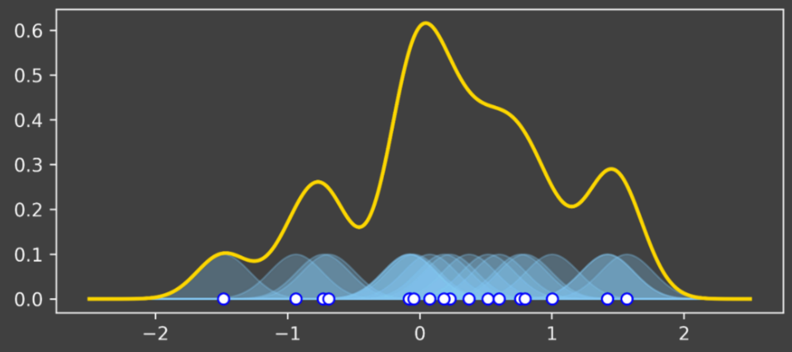


KDE

$$\hat{f}(y) = \frac{1}{n}\sum_{j=1}^{n} K_h\left(X_j - y\right)$$

$$= \mathbb{E}_{X\sim\mathbb{P}_n}[K_h(X - y)] \longrightarrow \mathbb{E}_{X\sim\mathbb{P}_c}[K_h(X - y)]$$

empirical measure

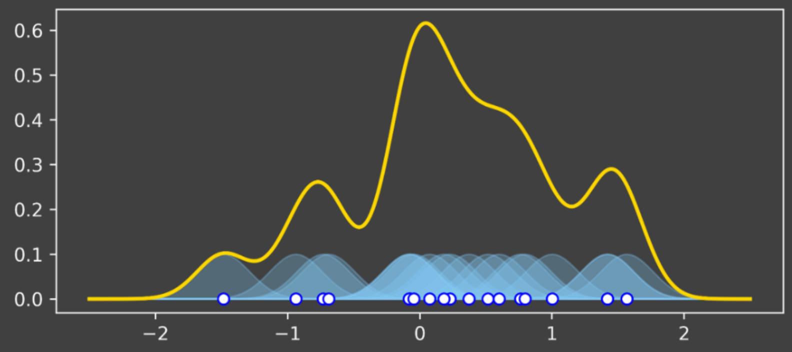


KDE

$$\hat{f}(y) = \frac{1}{n}\sum_{j=1}^{n} K_h\left(X_j - y\right)$$

$$= \mathbb{E}_{X\sim\mathbb{P}_n}[K_h(X - y)]$$

empirical measure



Coreset: a weighted subset of the data

$$\mathbb{E}_{X\sim\mathbb{P}_C}[K_h(X - y)]$$

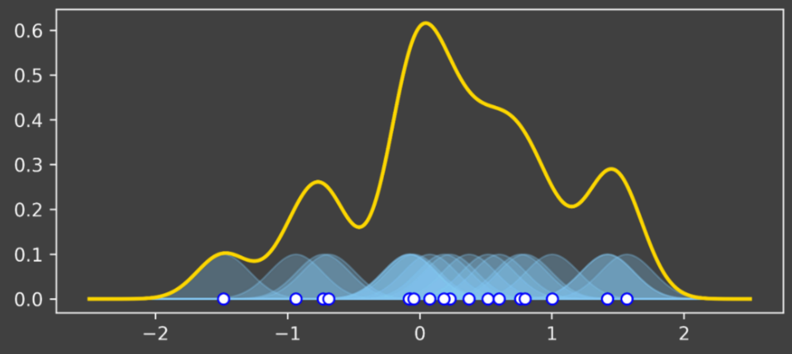
coreset measure

KDE

$$\hat{f}(y) = \frac{1}{n}\sum_{j=1}^{n} K_h\left(X_j - y\right)$$

$$= \mathbb{E}_{X\sim\mathbb{P}_n}[K_h(X - y)]$$

empirical measure



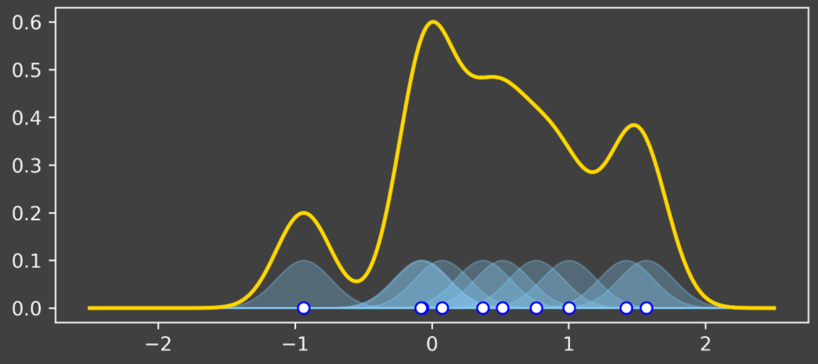
Coreset KDE

Coreset: a weighted subset of the data

$$\hat{f}_C(y) = \sum_{X_j\in C} \lambda_j K_h\left(X_j - y\right)$$

$$= \mathbb{E}_{X\sim\mathbb{P}_C}[K_h(X - y)]$$

coreset measure



What is the rate of estimation of coreset KDEs?

$$\hat{f}_C(y) = \sum_{X_j \in C} \lambda_j \, K_h\left(X_j - y\right)$$

Carathéodory Coresets

- Idea: Match Fourier coefficients of $\mathbb{P}_n$ and $\mathbb{P}_C$

Carathéodory Coresets

- Idea: Match Fourier coefficients of $\mathbb{P}_n$ and $\mathbb{P}_{\boldsymbol{C}}$
- Goal: Find the largest T so that

$$\frac{1}{n}\sum_{j=1}^{n} e^{i\,\omega\,X_j} = \sum_{X_j\,\epsilon\,\boldsymbol{C}} \lambda_j e^{i\,\omega\,X_j} \qquad \forall|\omega| < T$$

can be solved for some $\{\lambda_j\}$ and $\boldsymbol{C}$

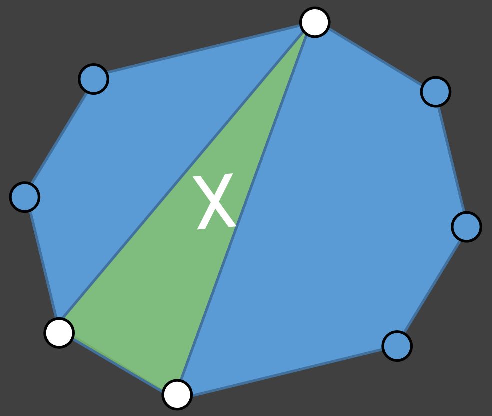
Carathéodory Coresets

- Idea: Match Fourier coefficients of $\mathbb{P}_n$ and $\mathbb{P}_{\boldsymbol{C}}$
- Goal: Find the largest T so that

$$\frac{1}{n}\sum_{j=1}^{n} e^{i\,\omega\,X_j} = \sum_{X_j\,\epsilon\,\boldsymbol{C}} \lambda_j e^{i\,\omega\,X_j} \qquad \forall|\omega| < T$$

can be solved for some $\{\lambda_j\}$ and $\boldsymbol{C}$

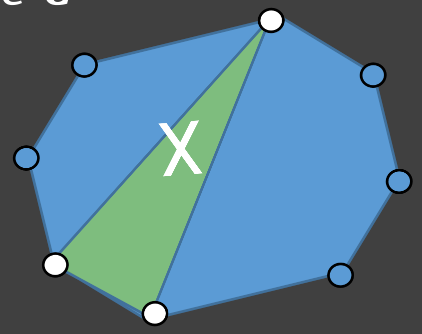
- By Carathéodory's theorem, can take $T = \Omega(|\boldsymbol{C}|)$



Analysis

- By Carathéodory, can find $\{\lambda_j\}$ and $\boldsymbol{C}$ such that

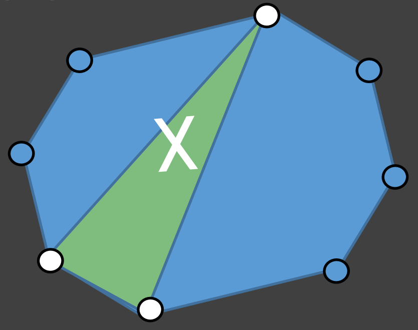
$$\frac{1}{n}\sum_{j=1}^{n} e^{i\,\omega\,X_j} = \sum_{X_j\,\epsilon\,\boldsymbol{C}} \lambda_j e^{i\,\omega\,X_j} \qquad \forall|\omega| < T = \Omega(|\boldsymbol{C}|)$$



Analysis

- By Carathéodory, can find $\{\lambda_j\}$ and $\boldsymbol{C}$ such that

$$\frac{1}{n}\sum_{j=1}^{n} e^{i\,\omega\,X_j} = \sum_{X_j\,\epsilon\,\boldsymbol{C}} \lambda_j e^{i\,\omega\,X_j} \qquad \forall|\omega| < T = \Omega(|\boldsymbol{C}|)$$



$$\left|\hat{f}(y_0) - \hat{f}_{\boldsymbol{C}}(y_0)\right|$$

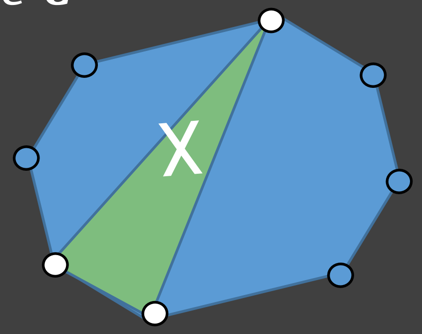
KDE

coreset
KDE

Analysis

- By Carathéodory, can find $\{\lambda_j\}$ and $\boldsymbol{C}$ such that

$$\frac{1}{n}\sum_{j=1}^{n} e^{i\,\omega\,X_j} = \sum_{X_j\,\epsilon\,\boldsymbol{C}} \lambda_j e^{i\,\omega\,X_j} \qquad \forall|\omega| < T = \Omega(|\boldsymbol{C}|)$$



$$\left|\hat{f}(y_0) - \hat{f}_{\boldsymbol{C}}(y_0)\right| \lesssim \left|\sum_{\omega}\left(\frac{1}{n}\sum_{j=1}^{n} e^{i\omega X_j} - \sum_{X_j\in\boldsymbol{C}} \lambda_j e^{i\omega X_j}\right) F[K_h](\omega)\right|$$

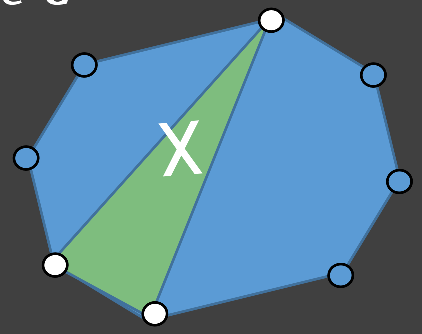
KDE

coreset KDE

Analysis

- By Carathéodory, can find $\{\lambda_j\}$ and $\boldsymbol{C}$ such that

$$\frac{1}{n}\sum_{j=1}^{n} e^{i\,\omega\,X_j} = \sum_{X_j\,\epsilon\,\boldsymbol{C}} \lambda_j e^{i\,\omega\,X_j} \qquad \forall|\omega| < T = \Omega(|\boldsymbol{C}|)$$



$$\left|\hat{f}(y_0) - \hat{f}_{\boldsymbol{C}}(y_0)\right| \lesssim \left|\sum_\omega \left(\frac{1}{n}\sum_{j=1}^{n} e^{i\omega X_j} - \sum_{X_j \in \boldsymbol{C}} \lambda_j e^{i\omega X_j}\right) F[K_h](\omega)\right|$$

$$\lesssim \sum_{|\omega|>T} |F[K](\omega h)|$$

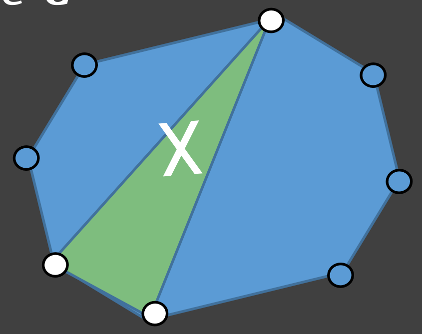
KDE

coreset KDE

Analysis

- By Carathéodory, can find $\{\lambda_j\}$ and $\boldsymbol{C}$ such that

$$\frac{1}{n}\sum_{j=1}^{n} e^{i\,\omega\,X_j} = \sum_{X_j\,\epsilon\,\boldsymbol{C}} \lambda_j e^{i\,\omega\,X_j} \qquad \forall|\omega| < T = \Omega(|\boldsymbol{C}|)$$



$$\left|\hat{f}(y_0) - \hat{f}_{\boldsymbol{C}}(y_0)\right| \lesssim \left|\sum_\omega \left(\frac{1}{n}\sum_{j=1}^{n} e^{i\omega X_j} - \sum_{X_j\in\boldsymbol{C}} \lambda_j e^{i\omega X_j}\right) F[K_h](\omega)\right|$$

$$\lesssim \sum_{|\omega|>T} |F[K](\omega h)|$$

KDE

coreset KDE

Small if K is smooth and $T > h^{-1-\varepsilon}$

Main Result

Theorem For appropriate kernel, the Carathéodory KDE achieves the minimax rate $n^{-\frac{\beta}{2\beta+d}}$ with a coreset of size

$$|\boldsymbol{C}| = n^{\frac{d}{2\beta+d}+\varepsilon}, \qquad \forall\, \varepsilon > 0$$

Main Result

Theorem For appropriate kernel, the Carathéodory KDE achieves the minimax rate $n^{-\frac{\beta}{2\beta+d}}$ with a coreset of size

$$|\boldsymbol{C}| = n^{\frac{d}{2\beta+d}+\varepsilon}, \qquad \forall\, \varepsilon > 0$$

We also show *any coreset procedure* requires at least $n^{\frac{d}{2\beta+d}}$ points to achieve minimax rate