Uniform Convergence Rate of the Kernel Density Estimator Adaptive to Intrinsic Volume Dimension

Jisu KIM

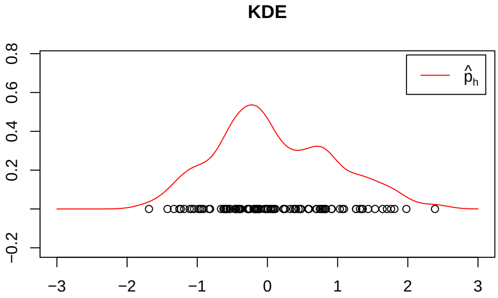
Inria Saclay
Joint work with Jaehyeok Shin, Alessandro Rinaldo, Larry Wasserman

2019-06-11

Kernel Density Estimator

- For $X_1, \ldots, X_n \sim P$, a given kernel function K, and a bandwidth $h > 0$, the Kernel Density Estimator (KDE) $\hat{p}_h : \mathbb{R}^d \to \mathbb{R}$ is

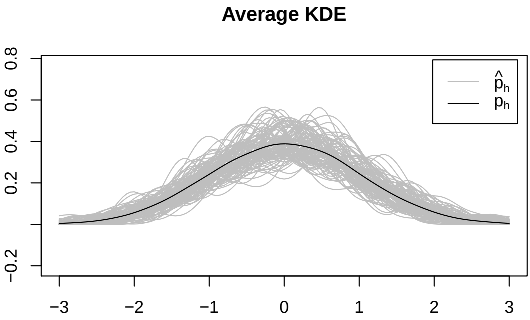
$$\hat{p}_h(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right).$$



Average Kernel Density Estimator

- The Average Kernel Density Estimator $p_h : \mathbb{R}^d \to \mathbb{R}$ is

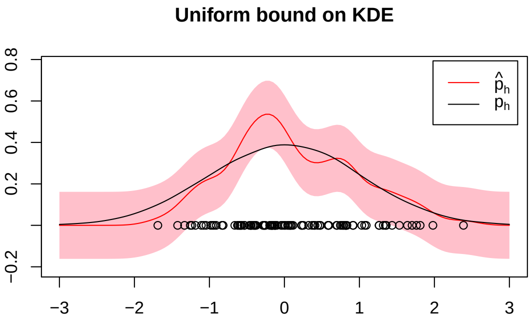
$$p_h(x) = \mathbb{E}_P\left[\hat{p}_h(x)\right] = \frac{1}{h^d}\mathbb{E}_P\left[K\left(\frac{x-X}{h}\right)\right].$$



We get the uniform convergence rate on Kernel Density Estimator.

- Fix a subset $\mathbb{X} \subset \mathbb{R}^d$, we need uniform control of the Kernel Density Estimator over $\mathbb{X}$, $\sup_{x \in \mathbb{X}} |\hat{p}_h(x) - p_h(x)|$, for various purposes.
- We get the concentration inequalities for the Kernel Density Estimator in the supremum norm that hold uniformly over the selection of the bandwidth, i.e.,

$$\sup_{h \geq l_n, x \in \mathbb{X}} \left| \hat{p}_h(x) - p_h(x) \right| .$$



The volume dimension characterizes the intrinsic dimension of the distribution related to the convergence rate of the Kernel Density Estimator.

- For a probability distribution P on $\mathbb{R}^d$, the volume dimension is

$$d_{\mathrm{vol}} := \sup\left\{\nu \geq 0 : \limsup_{r\to 0} \sup_{x\in\mathbb{X}} \frac{P(\mathbb{B}(x,r))}{r^{\nu}} < \infty\right\},$$

  where $\mathbb{B}(x,r) = \{y \in \mathbb{R}^d : \|x-y\| < r\}$.
- In other words, the volume dimension is the maximum possible exponent rate dominating the probability volume decay on balls.

The uniform convergence rate of the Kernel Density Estimator is derived in terms of the volume dimension.

Theorem

(Corollary 13, Corollary 17) Let P be a probability distribution on $\mathbb{R}^d$ satisfying weak assumptions and K be a kernel function satisfying weak assumptions. Suppose $l_n \to 0$ and $nl_n^{d_{\mathrm{vol}}} \to \infty$. Then with high probability,

$$\sqrt{\frac{1}{nl_n^{2d-d_{\mathrm{vol}}}}} \lesssim \sup_{h \geq l_n, x \in \mathbb{X}} |\hat{p}_h(x) - p_h(x)| \lesssim \sqrt{\frac{\log(1/l_n)}{nl_n^{2d-d_{\mathrm{vol}}}}},$$

for all large n.

Poster: Pacific Ballroom #188

- Poster: Tuesday Jun 11th 18:30 - 21:00 @ Pacific Ballroom #188
- Thank you!