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

Jisu Kim, Jaehyeok Shin, Alessandro Rinaldo, Larry Wasserman

INVENTORS FOR THE DIGITAL WORLD

Abstract

We derive concentration inequalities for the supremum norm of the difference between a kernel density estimator (KDE) and its point-wise expectation that hold uniformly over the selection of the bandwidth. We first propose the volume dimension to measure the intrinsic dimension of the support of a probability distribution based on the rates of decay of the probability of vanishing Euclidean balls. Our bounds depend on the volume dimension and generalize the existing bounds derived in the literature. Analogous bounds are derived for the derivative of the KDE, of any order. Our results are generally applicable but are especially useful for problems in geometric inference and topological data analysis, including level set estimation, density-based clustering, modal clustering and mode hunting, ridge estimation and persistent homology.

Uniform Convergence Rate of Kernel Density Estimator

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

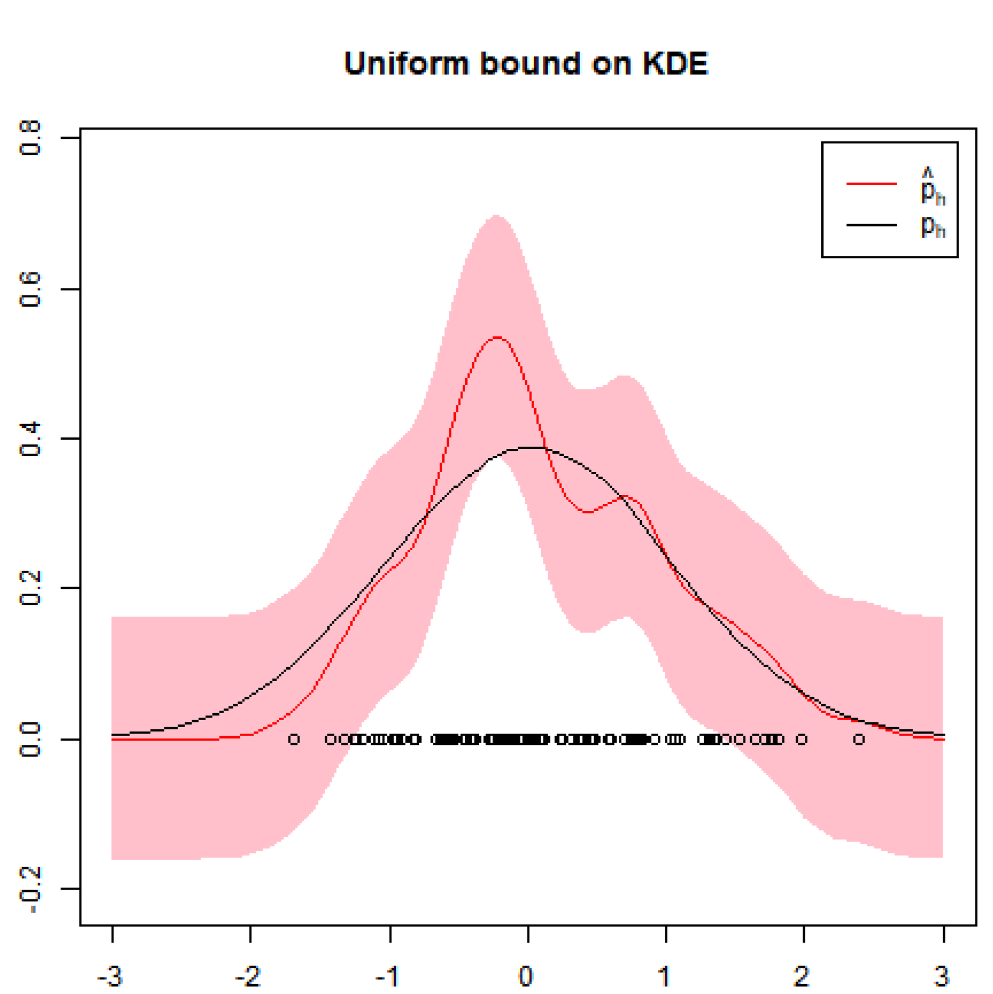
$$\hat{p}_h(x)=\frac{1}{nh^d}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right).$$

- The average kernel density estimator $p_h:\mathbb{R}^d\to\mathbb{R}$ is

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

- Fix a subset $\mathbb{X}\subset\mathbb{R}^d$, then 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 density-based clustering, modal clustering and mode hunting, mean-shift clustering, ridge estimation and inference for density level sets, cluster density trees, and persistence diagrams.
- The goal of this work is to get the concentration inequalities for the kernel density estimator in the supremum norm that hold uniformly over the selection of the bandwidth, i.e., the concentration inequalities for

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



Volume Dimension

- For a probability distribution P on $\mathbb{R}^d$, the volume dimension is the maximum possible exponent rate dominating the probability volume decay on balls, i.e. let $\mathbb{B}(x,r)=\{y\in\mathbb{R}^d:\|x-y\|<r\}$, then

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

- (Proposition 1) When there exists a set A satisfying $P(A\cap\mathbb{X})>0$ with Hausdorff dimension d_H, then $0\leq d_{vol}\leq d_H$.
- (Proposition 3) Suppose there exists a d_M-dimensional manifold with positive reach satisfying $P(M\cap\mathbb{X})>0$ and $\mathrm{supp}(P)\subset M$. If P has a bounded density with respect to the d_M-dimensional Hausdorff measure, then $d_{vol}=d_M$.

Assumptions

Let P be a probability distribution on $\mathbb{R}^d$, and d_{vol} be its volume dimension.

- Assumption 1. We assume that

$$\limsup_{r\to 0}\sup_{x\in\mathbb{X}}\frac{P(\mathbb{B}(x,r))}{r^{d_{vol}}}<\infty.$$

- Assumption 2. We assume that

$$\sup_{x\in\mathbb{X}}\liminf_{r\to 0}\frac{P(\mathbb{B}(x,r))}{r^{d_{vol}}}>\infty.$$

Let $K:\mathbb{R}^d\to\mathbb{R}$ be a kernel function with $\|K\|_\infty<\infty$.

- Assumption 3. Fix $k>0$ with its default to be $k=2$. Then we assume that either $d_{vol}=0$ or

$$\int_0^\infty t^{d_{vol}-1}\sup_{\|x\|\geq t}|K(x)|^k\,dt<\infty.$$

- Assumption 4. Let K satisfy $\|K\|_2<\infty$. We assume that

$$\mathcal{F}_{K,[l_n,\infty)}:=\left\{K\left(\frac{x-\cdot}{h}\right):x\in\mathbb{X},h\geq l_n\right\}$$

is a uniformly bounded VC-class with dimension ν, i.e., there exist positive numbers A and ν such that, for every $\epsilon\in(0,\|K\|_\infty)$, the covering number $\mathcal{N}(\mathcal{F}_{K,[l_n,\infty)},L_2(Q),\epsilon)$ satisfies

$$\mathcal{N}(\mathcal{F}_{K,[l_n,\infty)},L_2(Q),\epsilon)\leq\left(\frac{A\|K\|_\infty}{\epsilon}\right)^\nu,$$

where the covering number is the minimal number of open balls of radius ϵ with respect to distance $L_2(Q)$ whose centers are in $\mathcal{F}_{K,[l_n,\infty)}$ to cover $\mathcal{F}_{K,[l_n,\infty)}$.

Uniform Convergence of the Kernel Density Estimator

Uniformity on a ray of bandwidths

- (Corollary 13) Let P be a probability distribution on $\mathbb{R}^d$ and K be a kernel function satisfying Assumption 3 and 4. Fix $\epsilon\in(0,d_{vol})$. Further, if $\epsilon=0$ or under Assumption 1, ϵ can be 0. Suppose $l_n<1$ and

$$\limsup_n\frac{\log(1/l_n)+\log(2/\delta)}{nl_n^{d_{vol}-\epsilon}}<\infty.$$

Then, with probability $1-\delta$,

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

Fixed bandwidth

- (Corollary 15) For fixed bandwidth, i.e. bounding $\sup_{x\in\mathbb{X}}|\hat{p}_{h_n}(x)-p_{h_n}(x)|$, then Assumption 4 can be dropped when $\mathbb{X}$ is bounded and K is Lipschitz continuous.

Derivatives of the Kernel Density Estimator

- For $s\in(\{0\}\cup\mathbb{N})^d$, let $|s|=s_1+\cdots+s_d$ and $D^s:=\frac{\partial^{|s|}}{\partial x_1^{s_1}\cdots\partial x_d^{s_d}}$.
- With analogous assumptions on P and D^sK , with probability $1-\delta$,

$$\sup_{h\geq l_n,x\in\mathbb{X}}|D^s\hat{p}_h(x)-D^sp_h(x)|\lesssim\sqrt{\frac{\log(1/l_n)+\log(2/\delta)}{nl_n^{2d+2|s|-d_{vol}+\epsilon}}}.$$

Lower bound for the convergence of the KDE

- (Proposition 16) Let P be a probability distribution on $\mathbb{R}^d$ satisfying Assumption 2 with positive volume dimension. Let K be a kernel function satisfying Assumption 3 with $k=1$ and $\lim_{t\to 0}\inf_{\|x\|\leq t}K(x)>0$. Suppose $\lim_n nh_n^{d_{vol}}=\infty$. Then, with probability $1-\delta$,

$$\sup_{x\in\mathbb{X}}|\hat{p}_{h_n}(x)-p_{h_n}(x)|\gtrsim\sqrt{\frac{1}{nh_n^{2d-d_{vol}}}}.$$

- (Corollary 17) The same lower bound holds for a ray of bandwidths as well.
- By combining the upper and lower bounds together, with high probability,

$$\sqrt{\frac{1}{nh_n^{2d-d_{vol}}}}\lessapprox\sup_{x\in\mathbb{X}}|\hat{p}_{h_n}(x)-p_{h_n}(x)|\lesssim\sqrt{\frac{\log(1/h_n)}{nh_n^{2d-d_{vol}}}}.$$

Example

Let P be a probability distribution on $\mathbb{R}^d$ with density p with respect to d-dimensional Lebesgue measure. Fix $\beta<d$, and suppose $p:\mathbb{R}^d\to\mathbb{R}$ is defined as

$$p(x)=\frac{(d-\beta)\Gamma(d/2)}{2\pi^{d/2}}\|x\|^{-\beta}I(\|x\|\leq 1).$$

Then, the volume dimension is $d_{vol}=d-\beta$. And Assumption 1,2,3 are satisfied, so with high probability,

$$\sqrt{1/nh_n^{d+\beta}}\lessapprox\sup_{x\in\mathbb{X}}|\hat{p}_{h_n}(x)-p_{h_n}(x)|\lesssim\sqrt{\log(1/h_n)/nh_n^{d+\beta}}.$$

Hence although it has Lebesgue density, its L_∞ convergence rate of the KDE $\sqrt{1/nh_n^{d+\beta}}$ (up to $\log(1/h_n)$ term) is different from $\sqrt{1/nh_n^{d}}$, which is the usual rate for probability distributions with bounded Lebesgue density.