Uniform Convergence of the Density Estimator Adaptive to Geometric Dimension

Jisu Kim (김지수)
joint work with Jaehyeok Shin, Alessandro Rinaldo, Larry Wasserman

2025 KMS Spring Meeting (2025년 대한수학회 봄 연구발표회)
2025-04-26

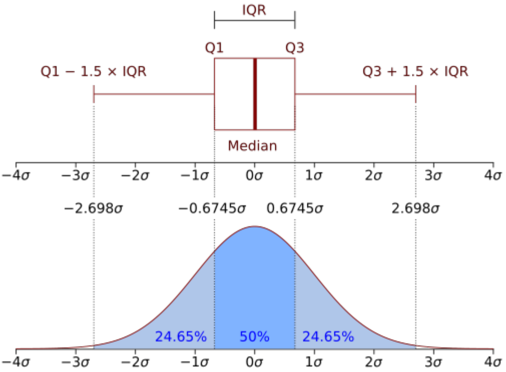
Uniform Convergence of the Density Estimator Adaptive to Geometric Dimension

- Jisu Kim, Jaehyeok Shin, Alessandro Rinaldo, and Larry Wasserman, Uniform Convergence of the Kernel Density Estimator Adaptive to Intrinsic Volume Dimension, 2019.
  - We derive uniform convergence bounds for asymptotic behavior of Kernel Density Estimator
  - We propose the volume dimension d_{vol}, a geometric dimension related to the convergence rate of the Kernel Density Estimator.
  - We will show

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

Probability density function

- A probability density function (or just density function) of a probability measure gives the probability per unit volume.
- For a probability measure P, its density function $p : \mathbb{R}^d \to \mathbb{R}$ is defined as a Radon-Nikodym derivative, $p(x) = \frac{dP}{d\mu}$, where μ is usally a Lebesgue measure on $\mathbb{R}^d$.
- A density function is of interest in statistics and machine learning.



[1]

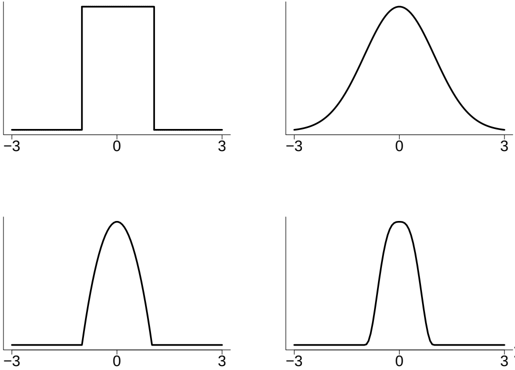
[1] https://en.wikipedia.org/wiki/Probability_density_function/

Kernel function

- A kernel function $K : \mathbb{R}^d \to \mathbb{R}$ is a function satisfying

$$\int K(t)\,dt = 1, \quad \int tK(t)\,dt = 0, \quad 0 < \int t^2 K(t)\,dt < \infty.$$

- Some examples are: boxcar (top left), Gaussian (top right), Epanechnikov (bottom left), Tricube (bottom right)

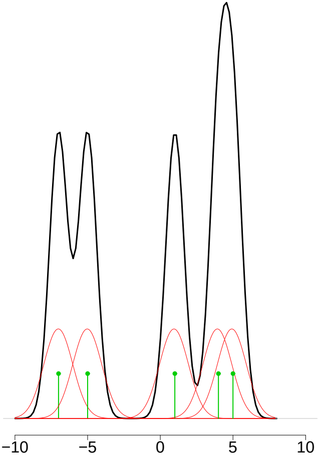


[2]

[2] **Larry Wasserman's lecture note**

We use Kernel Density Estimator to estimate probability density function.

- Kernel Density Estimator (KDE) $\hat{p}_h : \mathbb{R}^d \to \mathbb{R}$ is a weighted sum of kernel functions located at each data point.
- Kernel Density Estimator is an estimator for the probability density function $p : \mathbb{R}^d \to \mathbb{R}$.

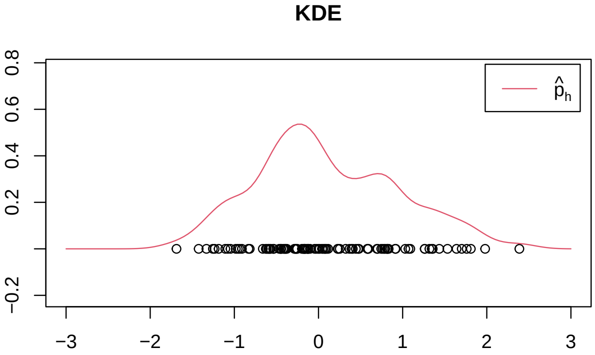


[3]

We use Kernel Density Estimator to estimate probability density function.

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

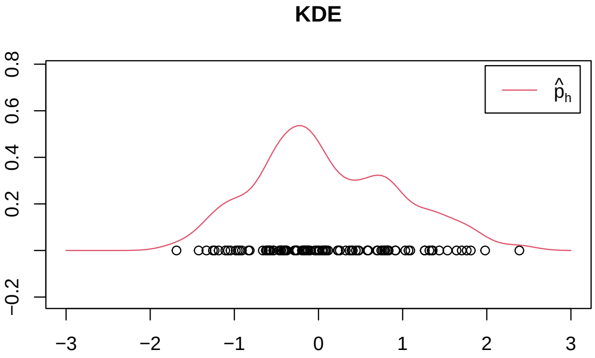
$$\hat{p}_h(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right).$$



We study asymptotic behavior of Kernel Density Estimator.

- Asymptotic behavior of the Kernel Density Estimator is a classic and fundamental problem in non-parametric statistics.
- When p is a density function with $\sup_{x\in\mathbb{R}^d}|p(x)|<\infty$, classical results have

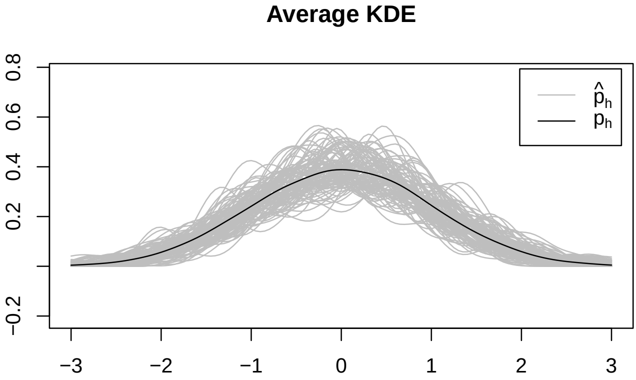
$$\|\hat{p}_h - p\|_\infty = O(h^2) + O_P\left(\sqrt{\frac{\log(1/h)}{nh^d}}\right).$$



Average Kernel Density Estimator

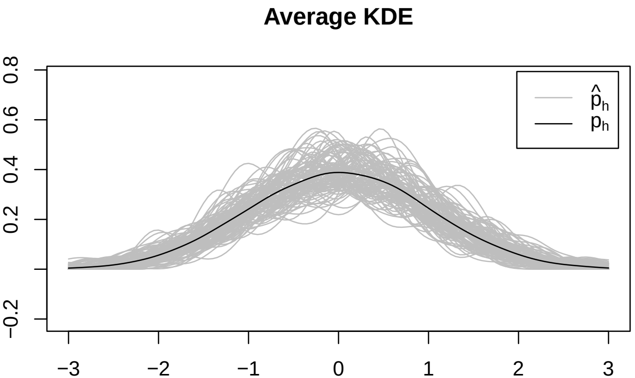
- The Average Kernel Density Estimator $p_h : \mathbb{R}^d \to \mathbb{R}$ is

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



Average Kernel Density Estimator conveys geometric and topological information of data

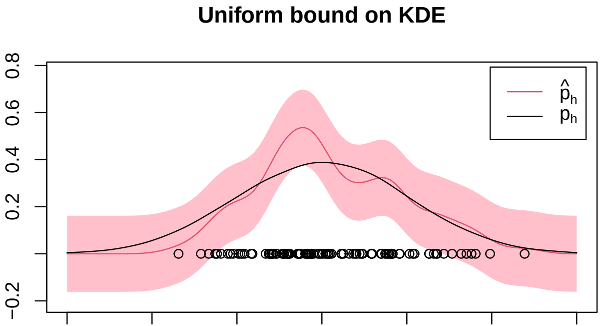
- The Average Kernel Density Estimator p_h can be viewed as a smoothed version of the density function p.
- The Average KDE p_h exists even if the density function p does not exist, in particular when the data are supported on a manifold.
- The Average KDE p_h itself is an important target of interest, in particular it conveys geometirc and topological information of data.



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

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



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 $\mathbb{B}(x, r) = \left\{ y \in \mathbb{R}^d : \|x - y\|_2 < r \right\}$.
- In other words, the volume dimension is the maximum possible exponent rate dominating the probability volume decay on balls.

The volume dimension has relations with Hausdorff dimension.

Proposition

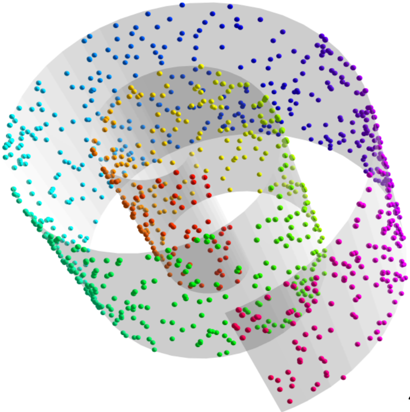
(Proposition 1) When there exists a set A satisfying $P(A \cap \mathbb{X}) > 0$ with Hausdorff dimension d_H, then

$$0 \leq d_{\mathrm{vol}} \leq d_H.$$

The volume dimension equals the geometric dimension under suitable conditions.

Proposition

(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 $d_{\mathrm{vol}} = d_M$.



4

The usual uniform convergence rate of Kernel Density Estimator

- For a probability distribution P with bounded Lebesgue density p, i.e., $\sup_{x\in\mathbb{R}^d}|p(x)|<\infty$, then with appropriate assumptions and choice of h_n, with high probability, (see e.g., Giné and Guillou, 2002)

$$\sup_{x\in\mathbb{R}^d}|\hat{p}_{h_n}(x)-p_{h_n}(x)|\precsim\sqrt{\frac{\log(1/h_n)}{nh_n^d}}.$$

The uniform convergence rate of Kernel Density Estimator is derived in terms of the volume dimension: uniform on a ray of bandwidths

Theorem

(Corollary 13) Let P be a probability distribution on $\mathbb{R}^d$, and K be a kernel function, and suppose P and K satisfy weak assumptions. Fix $\epsilon \in (0, d_{vol})$, or ϵ can be set to 0 as well under weak assumption. Suppose $l_n < 1$ and

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

Then with high probability,

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

The uniform convergence rate of the Kernel Density Estimator is derived in terms of the volume dimension: fixed bandwidth

Theorem

(Corollary 15) Let P be a probability distribution on $\mathbb{R}^d$, and K be a kernel function, and suppose P and K satisfy weak assumptions. Fix $\epsilon \in (0, d_{vol})$, or ϵ can be set to 0 as well under weak assumption. Suppose $h_n < 1$ and

$$\limsup_n \frac{\log(1/h_n)}{nh_n^{d_{vol}-\epsilon}} < \infty.$$

Then with high probability,

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

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

Theorem

(Corollary 21) Let $s \in (\{0\} \cup \mathbb{N})^d$, $|s| := s_1 + \cdots + s_d$ *and* $D^s := \frac{\partial^{|s|}}{\partial x_1^{s_1} \cdots \partial x_d^{s_d}}$. *Let* P *be a probability distribution on* $\mathbb{R}^d$, *and* K *be a kernel function, and suppose* P *and* $D^s K$ *satisfy weak assumptions. Fix* $\epsilon \in (0, d_{vol})$, *or* ϵ *can be set to* 0 *as well under weak assumption. Suppose* $l_n < 1$ *and*

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

Then with high probability,

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

The uniform convergence rate of the Kernel Density Estimator is derived in terms of the volume dimension: lower bound

Theorem

(Proposition 16) Let P be a probability distribution on $\mathbb{R}^d$, and K be a kernel function, and suppose P and K satisfy weak assumptions. Suppose $nh_n^{d_{vol}} \to \infty$. Then with high probability,

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

Example with nontrivial convergence rate of the Kernel Density Estimator

Example

(Example 18) Fix $\beta < d$, and let $p : \mathbb{R}^d \to \mathbb{R}$ be a Lebesgue density as

$$p(x) = \frac{(d-\beta)\Gamma(d/2)}{2\pi^{d/2}} \|x\|_2^{-\beta} I(\|x\|_2 \leq 1).$$

The volume dimension is $d_{vol} = d - \beta$, and with high probability,

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

Although p is a Lebesgue density, the convergence rate is different from $\sqrt{1/(nh_n^d)}$, which is a usual rate for probability distributions with bounded Lebesgue density.

Introduction

Volume Dimension

Uniform Convergence of the Kernel Density Estimator

Reference

Reference I

Evarist Giné and Armelle Guillou. Rates of strong uniform consistency for multivariate kernel density estimators. volume 38, pages 907–921. 2002. doi: 10.1016/S0246-0203(02)01128-7. URL https://doi.org/10.1016/S0246-0203(02)01128-7. En l'honneur de J. Bretagnolle, D. Dacunha-Castelle, I. Ibragimov.

Jisu Kim, Jaehyeok Shin, Alessandro Rinaldo, and Larry A. Wasserman. Uniform convergence rate of the kernel density estimator adaptive to intrinsic volume dimension. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pages 3398–3407. PMLR, 2019. URL http://proceedings.mlr.press/v97/kim19e.html.

Thank you!