SUPPLEMENTARY MATERIAL

A. Backgrounds and Basic Definitions

First, we define the Hausdorff measure ((Pesin, 1997, Section 6), (Falconer, 2014, Section 2.2)), which is a generalization of the Lebesgue measure to lower dimensional subsets of $\mathbb{R}^d$. For a subset $A \subset \mathbb{R}^d$, we let $\mathrm{diam}(A)$ be its diameter, that is

$$\mathrm{diam}(A) = \sup\{\|x - y\| : x, y \in A\}.$$

Definition 2. *Fix $\nu > 0$ and $\delta > 0$. For any set $A \subset \mathbb{R}^d$, define H_δ^ν be*

$$H_\delta^\nu(A) := \inf\left\{\sum_{i=1}^{\infty}(\mathrm{diam}U_i)^\nu : A \subset \bigcup_{i=1}^{\infty} U_i \textit{ and } \mathrm{diam}(U_i) < \delta\right\},$$

where the infimum is over all countable covers of A by sets $U_i \subset \mathbb{R}^d$ satisfying $\mathrm{diam}(U_i) < \delta$. Then, let the ν-dimensional Hausdorff measure *H^ν be*

$$H^\nu(A) := \lim_{\delta \to 0} H_\delta^\nu(A).$$

Then, the Hausdorff dimension of a set is the infimum over dimensions that make the Hausdorff measure on that set to be 0.

Definition 3. *For any set $A \subset \mathbb{R}^d$, its* Hausdorff dimension *$d_H(A)$ is*

$$d_H(A) := \inf\left\{\nu : H^\nu(A) = 0\right\}.$$

We use the normalized ν-dimensional Hausdorff measure so that when ν is an integer, its measure on ν-dimensional unit cube is 1. This can be done by defining the normalized ν-dimensional Hausdorff measure λ_ν as

$$\lambda_\nu = \frac{\pi^{\frac{\nu}{2}}}{2^\nu \Gamma(\frac{\nu}{2}+1)} H^\nu.$$

Now, we define the reach, which is a regularity parameter in geometric measure theory. Given a closed subset $A \subset \mathbb{R}^d$, the medial axis of A, denoted by $\mathrm{Med}(A)$, is the subset of $\mathbb{R}^d$ composed of the points that have at least two nearest neighbors on A. Namely, denoting by $d(x, A) = \inf_{q \in A} ||q - x||$ the distance function of a generic point x to A,

$$\mathrm{Med}(A) = \left\{x \in \mathbb{R}^d \setminus A | \exists q_1 \neq q_2 \in A, ||q_1 - x|| = ||q_2 - x|| = d(x, A)\right\}. \tag{22}$$

The reach of A is then defined as the minimal distance from A to $\mathrm{Med}(A)$.

Definition 4. *The* reach *of a closed subset $A \subset \mathbb{R}^d$ is defined as*

$$\tau_A = \inf_{q \in A} d\left(q, \mathrm{Med}(A)\right) = \inf_{q \in A, x \in \mathrm{Med}(A)} ||q - x||. \tag{23}$$

B. Proof for Section 3

We show Lemma 4 first, which is a simple argument from the definition of d_{vol} in (4) in Definition 1.

Lemma 4. *Let P be a probability distribution on $\mathbb{R}^d$, and d_{vol} be its volume dimension. Then for any $\nu \in [0, d_{\mathrm{vol}})$, there exists a constant $C_{\nu,P}$ depending only on P and ν such that for all $x \in \mathbb{X}$ and $r > 0$,*

$$\frac{P(\mathbb{B}_{\mathbb{R}^d}(x, r))}{r^\nu} \leq C_{\nu,P}.$$

Proof of Lemma 4. From the definition of d_{vol} in (4) in Definition 1, $\nu \in [0, d_{\mathrm{vol}})$ implies that

$$\limsup_{r \to 0} \sup_{x \in \mathbb{X}} \frac{P(\mathbb{B}_{\mathbb{R}^d}(x, r))}{r^\nu} < \infty.$$

Then there exist $r_0 > 0$ and $C'_{\nu,P} > 0$ such that for all $r \leq r_0$ and for all $x \in \mathbb{X}$,

$$\frac{P(\mathbb{B}_{\mathbb{R}^d}(x,r))}{r^\nu} \leq C'_{\nu,P}. \tag{24}$$

And for all $r \geq r_0$ and for all $x \in \mathbb{X}$,

$$\frac{P(\mathbb{B}_{\mathbb{R}^d}(x,r))}{r^\nu} \leq \frac{1}{r_0^\nu}. \tag{25}$$

Hence combining (24) and (25) gives that for all $r > 0$ and for all $x \in \mathbb{X}$,

$$\frac{P(\mathbb{B}_{\mathbb{R}^d}(x,r))}{r^\nu} \leq \max\left\{C'_{\nu,P}, \frac{1}{r_0^\nu}\right\}.$$

□

Then we can show Proposition 1 by using Lemma 4 and the definition of Hausdorff dimension in Definition 3.

Proposition 1. *Let P be a probability distribution on $\mathbb{R}^d$, and d_{vol} be its volume dimension. Suppose there exists a set A satisfying $P(A \cap \mathbb{X}) > 0$ and with Hausdorff dimension d_H. Then $0 \leq d_{\mathrm{vol}} \leq d_H$. Hence if A is a d_M-dimensional manifold, then $0 \leq d_{\mathrm{vol}} \leq d_M$. In particular, for any probability distribution P on $\mathbb{R}^d$, $0 \leq d_{\mathrm{vol}} \leq d$. Also, if P has a point mass, i.e. there exists $x \in \mathbb{X}$ with $P(\{x\}) > 0$, then $d_{\mathrm{vol}} = 0$.*

Proof of Proposition 1. We first show $d_{\mathrm{vol}} \geq 0$. For any $x \in \mathbb{X}$ and $r \geq 0$,

$$\frac{P(\mathbb{B}_{\mathbb{R}^d}(x,r))}{r^0} \leq 1 < \infty.$$

Hence $d_{\mathrm{vol}} \geq 0$ holds.

Now we show $d_{\mathrm{vol}} \leq d_H = d_H(A)$. Fix any $\nu < d_{\mathrm{vol}}$, and we will show that $H^\nu(A \cap \mathbb{X}) > 0$. Let $\{U_i\}$ be a countable cover of $A \cap \mathbb{X}$, i.e. $A \cap \mathbb{X} \subset \bigcup_{i=1}^\infty U_i$, and let $r_i = \mathrm{diam}(U_i)$. For each i, we can assume that $U_i \cap (A \cap \mathbb{X}) \neq \emptyset$ and choose $x_i \in U_i \cap (A \cap \mathbb{X})$. Then $U_i \subset \overline{\mathbb{B}_{\mathbb{R}^d}(x_i, r_i)} \subset \mathbb{B}_{\mathbb{R}^d}(x_i, 2r_i)$, and hence

$$A \cap \mathbb{X} \subset \bigcup_{i=1}^\infty \mathbb{B}_{\mathbb{R}^d}(x_i, 2r_i).$$

Then with $x_i \in \mathbb{X}$, applying (5) from Lemma 4 gives

$$\begin{aligned} P(A \cap \mathbb{X}) &< P\left(\bigcup_{i=1}^\infty \mathbb{B}_{\mathbb{R}^d}(x_i, 2r_i)\right) = \sum_{i=1}^\infty P(\mathbb{B}_{\mathbb{R}^d}(x_i, 2r_i)) \\ &\leq \sum_{i=1}^\infty 2^\nu C_{\nu,P} r_i^\nu. \end{aligned}$$

Hence

$$\sum_{i=1}^\infty r_i^\nu \geq \frac{P(A \cap \mathbb{X})}{2^\nu C_{\nu,P}} > 0.$$

Since this holds for arbitrary covers of $A \cap \mathbb{X}$, $H_\delta^\nu(A \cap \mathbb{X}) \geq \frac{P(A\cap\mathbb{X})}{2^\nu C_{\nu,P}}$ for all $\delta > 0$. And $A \cap \mathbb{X} \subset A$ implies

$$H^\nu(A) \geq H^\nu(A \cap \mathbb{X}) = \lim_{\delta \to 0} H_\delta^\nu(A \cap \mathbb{X}) \geq \frac{P(A \cap \mathbb{X})}{2^\nu C_{\nu,P}} > 0.$$

Since this holds for arbitrary $\nu < d_{\mathrm{vol}}$, the definition of Hausdorff dimension in Definition 3 gives that

$$d_H = \inf\{\nu : H^\nu(A) = 0\} \geq d_{\mathrm{vol}}.$$

Now, if A is a d_M-dimensional manifold, then the Hausdorff dimension of A is d_M, and hence $0 \leq d_{\mathrm{vol}} \leq d_M$ holds. In particular, setting $A = \mathbb{R}^d$ gives $0 \leq d_{\mathrm{vol}} \leq d$ for all probability distributions. Also, if there exists $x \in \mathbb{X}$ with $P(\{x\}) > 0$, then setting $A = \{x\}$ gives $d_{\mathrm{vol}} = 0$. □

Proposition 2 is again a simple argument from the definition of d_{vol} in (4) in Definition 1.

Proposition 2. *Let $P_1, \ldots, P_m$ be probability distributions on $\mathbb{R}^d$, and $\lambda_1, \ldots, \lambda_m \in (0,1)$ with $\sum_{i=1}^m \lambda_i = 1$. Then*

$$d_{\mathrm{vol}}\left(\sum_{i=1}^m \lambda_i P_i\right) = \min\left\{d_{\mathrm{vol}}(P_i) : 1 \leq i \leq m\right\}.$$

In particular, when d_{vol} is understood as a real-valued function on the space of probability distributions, both its sublevel sets and superlevel sets are convex.

Proof of Proposition 2. It is enough to show for the case $m = 2$. Let $P := \lambda_1 P_1 + \lambda_2 P_2$.

We first show $d_{\mathrm{vol}}(P) \geq \min\{d_{\mathrm{vol}}(P_1), d_{\mathrm{vol}}(P_2)\}$. Fix $\nu < \min\{d_{\mathrm{vol}}(P_1), d_{\mathrm{vol}}(P_2)\}$, then Definition 1 gives that

$$\limsup_{r\to 0} \sup_{x\in\mathbb{X}} \frac{P_1(\mathbb{B}_{\mathbb{R}^d}(x,r))}{r^\nu}, \ \limsup_{r\to 0} \sup_{x\in\mathbb{X}} \frac{P_2(\mathbb{B}_{\mathbb{R}^d}(x,r))}{r^\nu} < \infty.$$

And hence

$$\begin{aligned} \limsup_{r\to 0} \sup_{x\in\mathbb{X}} \frac{P(\mathbb{B}_{\mathbb{R}^d}(x,r))}{r^\nu} &= \limsup_{r\to 0} \sup_{x\in\mathbb{X}} \left\{\frac{\lambda_1 P_1(\mathbb{B}_{\mathbb{R}^d}(x,r))}{r^\nu} + \frac{\lambda_2 P_2(\mathbb{B}_{\mathbb{R}^d}(x,r))}{r^\nu}\right\} \\ &\leq \lambda_1 \limsup_{r\to 0} \sup_{x\in\mathbb{X}} \frac{P_1(\mathbb{B}_{\mathbb{R}^d}(x,r))}{r^\nu} + \lambda_2 \limsup_{r\to 0} \sup_{x\in\mathbb{X}} \frac{P_2(\mathbb{B}_{\mathbb{R}^d}(x,r))}{r^\nu} < \infty. \end{aligned}$$

And hence $d_{\mathrm{vol}}(P) \geq \min\{d_{\mathrm{vol}}(P_1), d_{\mathrm{vol}}(P_2)\}$ holds.

Next, we show $d_{\mathrm{vol}}(P) \leq \min\{d_{\mathrm{vol}}(P_1), d_{\mathrm{vol}}(P_2)\}$. Without loss of generality, suppose $d_{\mathrm{vol}}(P_1) \leq d_{\mathrm{vol}}(P_2)$, and fix $\nu > d_{\mathrm{vol}}(P_1)$. Then Definition 1 gives that

$$\limsup_{r\to 0} \sup_{x\in\mathbb{X}} \frac{P_1(\mathbb{B}_{\mathbb{R}^d}(x,r))}{r^\nu} = \infty.$$

Then from $P \geq \lambda_1 P_1$,

$$\limsup_{r\to 0} \sup_{x\in\mathbb{X}} \frac{P(\mathbb{B}_{\mathbb{R}^d}(x,r))}{r^\nu} \geq \limsup_{r\to 0} \sup_{x\in\mathbb{X}} \frac{\lambda_1 P_1(\mathbb{B}_{\mathbb{R}^d}(x,r))}{r^\nu} = \infty.$$

And hence $d_{\mathrm{vol}}(P) \leq d_{\mathrm{vol}}(P_1) = \min\{d_{\mathrm{vol}}(P_1), d_{\mathrm{vol}}(P_2)\}$ holds.

□

For Proposition 3 and 5, we need to bound the volume of the ball on the manifold. The following is rephrased from Lemma 3 in Kim et al. (2019).

Lemma 24. *Let $M \subset \mathbb{R}^d$ be a d_M-dimensional submanifold with reach τ_M. For a subset $U \subset M$ and $r < \tau_M$, let $U_r := \{x \in \mathbb{R}^d : \operatorname{dist}(x, U) < r\}$ be an r-neighborhood of U in $\mathbb{R}^d$. Then*

$$\lambda_{d_M}(U) \leq \frac{d!}{d_M!} r^{d_M - d} \lambda_d(U_r).$$

Then, the following Lemma is by combining Lemma 5.3 in Niyogi et al. (2008) and Lemma 24.

Lemma 25. *Let $M \subset \mathbb{R}^d$ be a d_M-dimensional submanifold with reach τ_M. Then, for $x \in M$ and $r < \tau_M$,*

$$\left(1 - \frac{r^2}{4\tau_M^2}\right)^{\frac{d_M}{2}} r^{d_M} \omega_d \leq \lambda_{d_M}(M \cap \mathbb{B}_{\mathbb{R}^d}(x,r)) \leq \frac{d!}{d_M!} 2^d r^{d_M} \omega_d.$$

Proof of Lemma 25. The LHS inequality is from Lemma 5.3 in Niyogi et al. (2008). The RHS inequality is applying $U = M \cap \mathbb{B}_{\mathbb{R}^d}(x,r)$ to Lemma 24 and $\lambda_d(U_r) \leq \lambda_d(\mathbb{B}_{\mathbb{R}^d}(x, 2r)) = (2r)^d \omega_d$.

□

Now, we show Proposition 3 and 5 simultaneously via the following Proposition:

Proposition 26. *Let P be a probability distribution on $\mathbb{R}^d$, and $d_{\rm vol}$ be its volume dimension. Suppose there exists a d_M-dimensional manifold M with positive reach satisfying $P(M\cap\mathbb{X})>0$ and $\mathrm{supp}(P)\subset M$. If P has a bounded density p with respect to the normalized d_M-dimensional Hausdorff measure λ_{d_M}, then $d_{\rm vol}=d_M$, and Assumption 1 and 2 are satisfied. In particular, when P has a bounded density p with respect to the d-dimensional Lebesgue measure λ_d, then $d_{\rm vol}=d$, and Assumption 1 and 2 are satisfied.*

Proof for Proposition 26. Let τ_M be the reach of M.

We first show $d_{\rm vol}=d_M$ and Assumption 1. Since the density p is bounded, for all $x\in\mathbb{X}$ and $r>0$, the probability on the ball $\mathbb{B}_{\mathbb{R}^d}(x,r)$ is bounded as

$$P(\mathbb{B}_{\mathbb{R}^d}(x,r))\leq\|p\|_\infty\,\lambda_{d_M}(M\cap\mathbb{B}_{\mathbb{R}^d}(0,r)). \tag{26}$$

Then for $r<\tau_M$, Lemma 25 implies $\lambda_{d_M}(M\cap\mathbb{B}_{\mathbb{R}^d}(x,r))\leq\frac{d!}{d_M!}2^d r^{d_M}\omega_d$, and hence

$$\limsup_{r\to 0}\sup_{x\in\mathbb{X}}\frac{P(\mathbb{B}_{\mathbb{R}^d}(x,r))}{r^{d_M}}\leq\|p\|_\infty\frac{d!}{d_M!}2^d\omega_d<\infty, \tag{27}$$

which implies

$$d_{\rm vol}\geq d_M.$$

Then from Proposition 1,

$$d_{\rm vol}=d_M.$$

Now, (27) shows that Assumption 1 is satisfied.

For Assumption 2, define a density $q:\mathbb{R}^d\to\mathbb{R}$ as

$$q(x)=\lim_{r\to 0}\frac{\Gamma(\frac{d_M}{2}+1)}{\pi^{\frac{d_M}{2}}}\frac{P(\mathbb{B}_{\mathbb{R}^d}(x,r))}{r^{d_M}}.$$

Since M is a submanifold with positive reach, P is λ_{d_M}-rectifiable. This imply that such limit $q(x)$ exists a.e. $[\lambda_{d_M}]$, and for any measurable set A,

$$P(A)=\int_{A\cap M}q(x)d\lambda_{d_M}(x).$$

See, for instance, Rinaldo & Wasserman (2010, Appendix), Mattila (1995, Corollary 17.9), or Ambrosio et al. (2000, Theorem 2.83). Then from

$$P(M\cap\mathbb{X})=\int_{M\cap\mathbb{X}}q(x)d\lambda_{d_M}(x)>0,$$

there exists $x_0\in M\cap\mathbb{X}$ with $q(x_0)>0$. And hence

$$\sup_{x\in\mathbb{X}}\liminf_{r\to 0}\frac{P(\mathbb{B}_{\mathbb{R}^d}(x,r))}{r^{d_M}}\geq q(x_0)>0,$$

and hence Assumption 2 is satisfied.

□

The proof of Proposition 6 is simply checking the convexities for Assumption 1 and Assumption 2.

Proposition 6. *The set of probability distributions satisfying Assumption 1 is convex. And so is the set of probability distributions satisfying Assumption 2.*

Proof of Proposition 6. Suppose P_1,P_2 are two probability distributions and $\lambda\in(0,1)$. Let $P:=\lambda P_1+(1-\lambda)P_2$. Then Proposition 2 implies that

$$d_{\rm vol}(P)=\min\{d_{\rm vol}(P_1),d_{\rm vol}(P_2)\}.$$

Consider Assumption 1 first. Suppose P_1 and P_2 satisfies Assumption 1. Then for all $x \in \mathbb{X}$ and $r \leq 1$, applying $d_{\rm vol}(P_1) \leq d_{\rm vol}(P_1), d_{\rm vol}(P_2)$ gives

$$\begin{aligned}\frac{P(\mathbb{B}_{\mathbb{R}^d}(x,r))}{r^{d_{\rm vol}(P)}} &= \lambda \frac{P_1(\mathbb{B}_{\mathbb{R}^d}(x,r))}{r^{d_{\rm vol}(P)}} + (1-\lambda)\frac{P_2(\mathbb{B}_{\mathbb{R}^d}(x,r))}{r^{d_{\rm vol}(P)}} \\ &\leq \lambda \frac{P_1(\mathbb{B}_{\mathbb{R}^d}(x,r))}{r^{d_{\rm vol}(P_1)}} + (1-\lambda)\frac{P_2(\mathbb{B}_{\mathbb{R}^d}(x,r))}{r^{d_{\rm vol}(P_2)}}.\end{aligned}$$

Hence,

$$\begin{aligned}\limsup_{r\to 0}\sup_{x\in\mathbb{X}} \frac{P(\mathbb{B}_{\mathbb{R}^d}(x,r))}{r^{d_{\rm vol}(P)}} &\leq \limsup_{r\to 0}\sup_{x\in\mathbb{X}} \left\{ \lambda \frac{P_1(\mathbb{B}_{\mathbb{R}^d}(x,r))}{r^{d_{\rm vol}(P_1)}} + (1-\lambda)\frac{P_2(\mathbb{B}_{\mathbb{R}^d}(x,r))}{r^{d_{\rm vol}(P_2)}} \right\} \\ &\leq \lambda \limsup_{r\to 0}\sup_{x\in\mathbb{X}} \frac{P_1(\mathbb{B}_{\mathbb{R}^d}(x,r))}{r^{d_{\rm vol}(P_1)}} + (1-\lambda)\limsup_{r\to 0}\sup_{x\in\mathbb{X}} \frac{P_2(\mathbb{B}_{\mathbb{R}^d}(x,r))}{r^{d_{\rm vol}(P_2)}} \\ &< \infty,\end{aligned}$$

and Assumption 1 is satisfied for $P = \lambda P_1 + (1-\lambda)P_2$.

Now, consider Assumption 2. Suppose P_1 and P_2 satisfies Assumption 1, and without loss of generality, assume $d_{\rm vol}(P_1) \leq d_{\rm vol}(P_2)$. Then there exists $x_0 \in \mathbb{X}$ such that

$$\liminf_{r\to 0} \frac{P_1(\mathbb{B}_{\mathbb{R}^d}(x_0,r))}{r^{d_{\rm vol}(P_1)}} > 0.$$

Then $P \geq \lambda P_1$ and $d_{\rm vol}(P) = d_{\rm vol}(P_1)$ give

$$\liminf_{r\to 0} \frac{P(\mathbb{B}_{\mathbb{R}^d}(x_0,r))}{r^{d_{\rm vol}(P)}} \geq \liminf_{r\to 0} \frac{\lambda P_1(\mathbb{B}_{\mathbb{R}^d}(x_0,r))}{r^{d_{\rm vol}(P_1)}} > \infty.$$

Hence

$$\sup_{x\in\mathbb{X}}\liminf_{r\to 0} \frac{P(\mathbb{B}_{\mathbb{R}^d}(x,r))}{r^{d_{\rm vol}(P)}} > 0,$$

and Assumption 2 is satisfied for $P = \lambda P_1 + (1-\lambda)P_2$.

□

C. Volume Dimension and Other Dimensions

In this section, we compare the volume dimension with other various dimensions.

For a set, one commonly used dimension other than the Hausdorff dimension is the box dimension ((Pesin, 1997, Section 6), (Falconer, 2014, Section 3.1)). This has various names as Kolmogorov entropy, entropy dimension, capacity dimension, metric dimension, logarithmic density or Minkowski dimension.

Definition 5. *For any set $A \subset \mathbb{R}^d$ and $\delta > 0$, let $N(A,\delta)$ be the smallest number of balls of radius δ to cover A. Then the* lower box dimension *of A is defined as*

$$d_B^-(A) := \liminf_{\delta\to 0} \frac{\log N(A,\delta)}{-\log \delta},$$

and the upper box dimension *of A is defined as*

$$d_B^+(A) := \limsup_{\delta\to 0} \frac{\log N(A,\delta)}{-\log \delta}.$$

The Hausdorff dimension and the lower and upper box dimensions are related as (Pesin, 1997, Theorem 6.2 (2)):

$$\text{forall } A \subset \mathbb{R}^d,\ d_H(A) \leq d_B^-(A) \leq d_B^+(A). \tag{28}$$

So far the Hausdorff dimension in Section A and the box dimension is defined for a set. For a probability distribution, there are two ways for natural extension. One way is to take the infimum of the set dimensions over all sets with positive probabilities ((Mattila et al., 2000, Section 2), (Falconer, 2014, Section 13.7)). We will use this as the definition of the Hausdorff dimension and the box dimension.

Definition 6. *Let P be a probability distribution on $\mathbb{R}^d$. Its* Hausdorff dimension $d_H(P)$ *is the infimum of the Hausdorff dimensions over a set with positive probability, i.e.,*

$$d_H(P) := \inf_{A:P(A)>0} d_H(A).$$

Similarly, the lower box dimension $d_B^-(P)$ *and the* upper box dimension $d_B^+(P)$ *is the infimum of the lower box dimensions and the upper box dimensions, respectively, over a set with positive probability, i.e.*

$$d_B^-(P) := \inf_{A:P(A)>0} d_B^-(A),$$
$$d_B^+(P) := \inf_{A:P(A)>0} d_B^+(A).$$

Another way is to take the infimum of the set dimensions over all sets with probabilities 1 (Pesin, 1997, Section 6). We will denote these dimensions as Hausdorff support dimension and the box support dimension to differentiate from the previous dimensions.

Definition 7. *Let P be a probability distribution on $\mathbb{R}^d$. Its* Hausdorff support dimension $d_{HS}(P)$ *is the infimum of the Hausdorff dimensions over a set with probability 1, i.e.,*

$$d_{HS}(P) := \inf_{A:P(A)=1} d_H(A).$$

Similarly, the lower box dimension $d_{BS}^-(P)$ *and the* upper box dimension $d_{BS}^+(P)$ *is the infimum of the lower box dimensions and the upper box dimensions, respectively, over a set with positive probability, i.e.*

$$d_{BS}^-(P) := \inf_{A:P(A)=1} d_B^-(A),$$
$$d_{BS}^+(P) := \inf_{A:P(A)=1} d_B^+(A).$$

The volume dimension, the Hausdorff dimension, and the lower and upper box dimensions have the following relations.

Proposition 27. *Let P be a probability distribution on $\mathbb{R}^d$ with $P(\mathbb{X}) > 0$. Then its volume dimension, Hausdorff dimension, lower and upper box dimension, Hausdorff support dimension, and lower and upper box support dimension satisfy the following inequality:*

$$d_{\mathrm{vol}}(P) \leq d_H(P) \leq d_B^-(P) \leq d_B^+(P),$$

and

$$d_{\mathrm{vol}}(P) \leq d_{HS}(P) \leq d_{BS}^-(P) \leq d_{BS}^+(P).$$

Proof. Since $P(\mathrm{supp}(P) \cap \mathbb{X}) = P(\mathbb{X}) > 0$, $d_{\mathrm{vol}}(P) \leq d_H(P)$ is direct from Proposition 1. Now, combining this with $d_H(P) \leq d_B^-(P) \leq d_B^+(P)$ and $d_{HS}(P) \leq d_{BS}^-(P) \leq d_{BS}^+(P)$ from (28) and that $d_H(P) \leq d_{HS}(P)$ gives the statement. □

Now, we introduce the q-dimension, which generalizes the box support dimension (Lee & Verleysen, 2007, Section 3.2.1).

Definition 8. *Let P be a probability distribution on $\mathbb{R}^d$. For $q \geq 0$ and $\delta > 0$, define $C_q(P, \delta)$ as*

$$C_q(P, \delta) := \int [P(\overline{\mathbb{B}_{\mathbb{R}^d}(x, \delta)})]^{q-1} dP(x).$$

Now for $q \geq 0$ and $q \neq 1$, the lower q-dimension *of P is*

$$d_q^-(P) := \liminf_{\delta \to 0} \frac{\log C_q(P, \delta)}{(q-1)\log \delta},$$

and the upper q-dimension *of P is*

$$d_q^+(P) := \limsup_{\delta \to 0} \frac{\log C_q(P, \delta)}{(q-1)\log \delta}.$$

For $q = 1$, we understand in the limit sense, i.e., $d_1^-(P) = \lim_{q \to 1} d_q^-(P)$ and $d_1^+(P) = \lim_{q \to 1} d_q^+(P)$.

This q-dimension is a generalization of the box support dimension in the sense that when $q = 0$, the lower and upper q-dimensions reduce to the lower and upper box support dimensions, respectively, i.e. $d_0^-(P) = d_{BS}^-(P)$ and $d_0^+(P) = d_{BS}^+(P)$ Pesin (1997, Section 8). When $q = 1$, the q-dimension is called the information dimension, and when $q = 2$, the q-dimension is called the correlation dimension.

The volume dimension and the q-dimension have the following relation.

Proposition 28. *Let P be a probability distribution on $\mathbb{R}^d$ with $P(\mathbb{X}) = 1$. Then for any $q \geq 0$, the volume dimension and the q-dimension has the following inequality:*

$$d_{\mathrm{vol}}(P) \leq d_q^-(P) \leq d_q^+(P).$$

Proof. Since $d_q^-(P) \leq d_q^+(P)$ is obvious, we only need to show $d_{\mathrm{vol}}(P) \leq d_q^-(P)$.

Fix any $\nu < d_{\mathrm{vol}}(P)$. Then from $P(\mathbb{X}) = 1$, $C_q(P, \delta)$ can be expressed as taking an integration over $\mathbb{X}$. Hence applying (5) from Lemma 4 gives

$$\begin{aligned} C_q(P,\delta) &= \int_{\mathbb{X}} [P(\overline{B_{\mathbb{R}^d}(x,\delta)})]^{q-1} dP(x) \\ &\leq \int_{\mathbb{X}} [P(B_{\mathbb{R}^d}(x,2\delta))]^{q-1} dP(x) \\ &\leq (2^\nu C_{\nu,P}\delta^\nu)^{q-1}. \end{aligned}$$

And hence $d_q^-(P)$ is lower bounded as

$$\begin{aligned} d_q^-(P) &= \liminf_{\delta\to 0} \frac{\log C_q(P,\delta)}{(q-1)\log\delta} \geq \liminf_{\delta\to 0} \frac{\log(2^\nu C_{\nu,P}\delta^\nu)}{\log\delta} \\ &= \nu + \liminf_{\delta\to 0} \frac{\log(2^\nu C_{\nu,P})}{\log\delta} = \nu. \end{aligned}$$

Since this holds for arbitrary $\nu < d_{\mathrm{vol}}(P)$, we have

$$d_{\mathrm{vol}}(P) \leq d_q^-(P).$$

□

We end this section by comparing the volume dimension and the Wasserstein dimension (Weed & Bach, 2017, Definition 4).

Definition 9. *Let P be a probability distribution on $\mathbb{R}^d$. For any $\delta > 0$ and $\tau \in [0,1]$, let the (δ,τ)-covering number of P be*

$$N(P,\delta,\tau) := \inf\{N(A,\delta) : P(A) \geq 1-\tau\},$$

and let the (δ,τ)-dimension be

$$d_\delta(P,\tau) := \frac{\log N(P,\delta,\tau)}{-\log\delta}.$$

Then for a fixed $p > 0$, the lower and upper Wasserstein dimensions are respectively,

$$\begin{aligned} d_*(P) &= \lim_{\tau\to 0}\liminf_{\delta\to 0} d_\delta(P,\tau) \\ d_p^*(P) &= \inf\{s \in (2p,\infty) : \limsup_{\delta\to 0} d_\epsilon(P, \delta^{\frac{sp}{s-2p}}) \leq s\}. \end{aligned}$$

Proposition 29. *Let P be a probability distribution on $\mathbb{R}^d$ with $P(\mathbb{X}) > 0$. Then its volume dimension and lower and upper Wasserstein dimensions satisfy the following inequality:*

$$d_{\mathrm{vol}}(P) \leq d_{HS}(P) \leq d_*(P) \leq d_p^*(P).$$

Proof. Since $P(\mathrm{supp}(P)\cap\mathbb{X}) = P(\mathbb{X}) > 0$, $d_{\mathrm{vol}}(P) \leq d_H(P)$ is direct from Proposition 1. The inequality $d_H(P) \leq d_*(P) \leq d_p^*(P)$ is from Weed & Bach (2017, Proposition 2).

□

D. Uniform convergence on a function class

As we have seen in (8) in Section 4, uniform bound on the kernel density estimator $\sup_{h\geq l_n, x\in\mathbb{X}} |\hat{p}_h(x) - p_h(x)|$ boils down to uniformly bounding on the function class $\sup_{f\in\tilde{\mathcal{F}}_{K,[l_n,\infty)}} \left|\frac{1}{n}\sum_{i=1}^n f(X_i) - \mathbb{E}[f(X)]\right|$. In this section, we derive a uniform convergence for a more general class of functions. Let $\mathcal{F}$ be a class of functions from $\mathbb{R}^d$ to $\mathbb{R}$, and consider a random variable

$$\sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_{i=1}^n f(X_i) - \mathbb{E}[f(X)]\right|. \tag{29}$$

As discussed in Section 4, we combine the Talagrand inequality (Theorem 8) and VC type bound (Theorem 9) to bound (29), which is generalizing the approach in Sriperumbudur & Steinwart (2012, Theorem 3.1).

Theorem 30. *Let $(\mathbb{R}^d, P)$ be a probability space and let $X_1,\ldots,X_n$ be i.i.d. from P. Let $\mathcal{F}$ be a class of functions from $\mathbb{R}^d$ to $\mathbb{R}$ that is uniformly bounded VC-class with dimension ν, i.e. there exists positive numbers A,B such that, for all $f\in\mathcal{F}$, $\|f\|_\infty \leq B$, and for every probability measure Q on $\mathbb{R}^d$ and for every $\epsilon\in(0,B)$, the covering number $\mathcal{N}(\mathcal{F}, L_2(Q), \epsilon)$ satisfies*

$$\mathcal{N}(\mathcal{F}, L_2(Q), \epsilon) \leq \left(\frac{AB}{\epsilon}\right)^\nu.$$

Let $\sigma > 0$ with $\mathbb{E}_P f^2 \leq \sigma^2$ for all $f\in\mathcal{F}$. Then there exists a universal constant C not depending on any parameters such that $\sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_{i=1}^n f(X_i) - \mathbb{E}[f(X)]\right|$ is upper bounded with probability at least $1-\delta$,

$$\sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_{i=1}^n f(X_i) - \mathbb{E}[f(X)]\right| \leq C\left(\frac{\nu B}{n}\log\left(\frac{2AB}{\sigma}\right) + \sqrt{\frac{\nu\sigma^2}{n}\log\left(\frac{2AB}{\sigma}\right)} + \sqrt{\frac{\sigma^2\log(\frac{1}{\delta})}{n}} + \frac{B\log(\frac{1}{\delta})}{n}\right).$$

Proof of Theorem 30. Let $\mathcal{G} := \{f - \mathbb{E}_P[f] :\ f\in\mathcal{F}\}$. Then it is immediate to check that for all $g\in\mathcal{G}$,

$$\begin{aligned}
&\mathbb{E}_P g = \mathbb{E}_P f - \mathbb{E}_P f = 0,\\
&\mathbb{E}_P g^2 = \mathbb{E}_P (f - \mathbb{E}_P f)^2 \leq \mathbb{E}_P f^2 \leq \sigma^2,\\
&\|g\|_\infty \leq \|f\|_\infty + \mathbb{E}_P f \leq 2B.
\end{aligned} \tag{30}$$

Now, $\sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_{i=1}^n f(X_i) - \mathbb{E}[f(X)]\right|$ is expanded as

$$\sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_{i=1}^n f(X_i) - \mathbb{E}[f(X)]\right| = \sup_{g\in\mathcal{G}} \left|\frac{1}{n}\sum_{i=1}^n g(X_i)\right|.$$

Hence from (30), applying Proposition 8 to above gives the probabilistic bound on $\sup_{g\in\mathcal{G}} \left|\frac{1}{n}\sum_{i=1}^n g(X_i)\right|$ as

$$P\left(\sup_{g\in\mathcal{G}} \left|\frac{1}{n}\sum_{i=1}^n g(X_i)\right| < 4\mathbb{E}_P \sup_{g\in\mathcal{G}} \left|\frac{1}{n}\sum_{i=1}^n g(X_i)\right| + \sqrt{\frac{2\sigma^2\log(\frac{1}{\delta})}{n}} + \frac{2B\log(\frac{1}{\delta})}{n}\right) \geq 1-\delta. \tag{31}$$

It thus remains to bound the term $\mathbb{E}_P \sup_{g\in\mathcal{G}} \left|\frac{1}{n}\sum_{i=1}^n g(X_i)\right|$. Let $\tilde{\mathcal{F}} := \{f - a :\ f\in\mathcal{F},\ a\in[-B,B]\}$. Then $\mathcal{F}$ being a uniform VC-class with dimension ν implies that for all $\epsilon\in(0,B)$,

$$\begin{aligned}
\sup_P \mathcal{N}\left(\tilde{\mathcal{F}}, L_2(P), \epsilon\right) &\leq \sup_P \mathcal{N}\left(\mathcal{F}, L_2(P), \frac{\epsilon}{2}\right) \sup_P \mathcal{N}\left([-B,B], |\cdot|, \frac{\epsilon}{2}\right)\\
&\leq \left(\frac{2AB}{\epsilon}\right)^{\nu+1}.
\end{aligned}$$

Hence from (30), applying Proposition 9 yields the upper bound for $\mathbb{E}_P \sup_{g\in\mathcal{G}} \left|\frac{1}{n}\sum_{i=1}^n g(X_i)\right|$ as

$$\mathbb{E}_{P^n} \sup_{g\in\mathcal{G}} \left|\frac{1}{n}\sum_{i=1}^n g(X_i)\right| \leq 2C\left(\frac{2(\nu+1)B}{n}\log\left(\frac{2AB}{\sigma}\right) + \sqrt{\frac{(\nu+1)\sigma^2}{n}\log\left(\frac{2AB}{\sigma}\right)}\right). \tag{32}$$

Hence applying (32) to (31) yields that, $\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}f(X_i)-\mathbb{E}[f(X)]\right|$ is upper bounded with probability at least $1-\delta$ as

$$\begin{aligned}\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}f(X_i)-\mathbb{E}[f(X)]\right| &\le 4C\left(\frac{2(\nu+1)B}{n}\log\left(\frac{2AB}{\sigma}\right)+\sqrt{\frac{(\nu+1)\sigma^2}{n}\log\left(\frac{2AB}{\sigma}\right)}\right)\\ &\quad+\sqrt{\frac{2\sigma^2\log(\frac{1}{\delta})}{n}}+\frac{2B\log(\frac{1}{\delta})}{n}\\ &\le 16C\left(\frac{\nu B}{n}\log\left(\frac{2AB}{\sigma}\right)+\sqrt{\frac{\nu\sigma^2}{n}\log\left(\frac{2AB}{\sigma}\right)}+\sqrt{\frac{\sigma^2\log(\frac{1}{\delta})}{n}}+\frac{B\log(\frac{1}{\delta})}{n}\right).\end{aligned}$$

□

E. Proof for Section 4

Lemma 11 is shown by the calculation using integral by parts and change of variables.

Lemma 11. *Let $(\mathbb{R}^d, P)$ be a probability space and let $X\sim P$. For any kernel K satisfying Assumption 3 with $k>0$, the expectation of the k-moment of the kernel is upper bounded as*

$$\mathbb{E}_P\left[\left|K\left(\frac{x-X}{h}\right)\right|^k\right]\le C_{k,P,K,\epsilon}h^{d_{\rm vol}-\epsilon},$$

for any $\epsilon\in(0,d_{\rm vol})$, where $C_{k,P,K,\epsilon}$ is a constant depending only on k, P, K, and ϵ. Further, if $d_{\rm vol}=0$ or under Assumption 1, ϵ can be 0 in (11).

Proof of Lemma 11. We first consider the case when $d_{\rm vol}=0$. Then $\mathbb{E}_P\left[\left|K\left(\frac{x-X}{h}\right)\right|^k\right]$ is simply bounded as

$$\mathbb{E}_P\left[\left|K\left(\frac{x-X}{h}\right)\right|^k\right]\le\|K\|_\infty^k h^0.$$

Now, we consider the case when $d_{\rm vol}>0$. Fix $\epsilon\in(0,d_{\rm vol})$. Under Assumption 1, ϵ can be chosen to be 0.

Let $C_{k,K,d_{\rm vol},\epsilon}:=\int_0^\infty t^{d_{\rm vol}-\epsilon-1}\sup_{\|x\|\le t}|K(x)|^k dt$, then it is finite from (10) and $\|K\|_\infty<\infty$ in Assumption 4 as

$$\begin{aligned}\int_0^\infty t^{d_{\rm vol}-\epsilon-1}\sup_{\|x\|\le t}|K(x)|^k dt &\le \int_0^1 t^{d_{\rm vol}-\epsilon-1}\|K\|_\infty dt+\int_1^\infty t^{d_{\rm vol}-1}\sup_{\|x\|\le t}|K(x)|^k dt\\ &\le\frac{\|K\|_\infty}{d_{\rm vol}-\epsilon}+\int_0^\infty t^{d_{\rm vol}-1}\sup_{\|x\|\le t}|K(x)|^k dt<\infty.\end{aligned}$$

Fix $\eta>0$, and let $\tilde{K}_\eta:[0,\infty)\to\mathbb{R}$ be a continuous and strictly decreasing function satisfying $\tilde{K}_\eta(t)>\sup_{\|x\|\ge t}|K(x)|^k$ for all $t\ge0$ and $\int_0^\infty t^{d_{\rm vol}-\epsilon-1}(\tilde{K}_\eta(t)-\sup_{\|x\|\ge t}|K(x)|^k)dt=\eta$. Such existence is possible since $t\mapsto\sup_{\|x\|\ge t}|K(x)|^k$ is nonincreasing function, so have at most countable discontinuous points, and $\int_0^\infty t^{d_{\rm vol}-\epsilon-1}\sup_{\|x\|\le t}|K(x)|^k dt<\infty$. Then it is immediate to check that

$$|K(x)|^k<\tilde{K}_\eta(\|x\|)\text{ for all }x\in\mathbb{R}. \tag{33}$$

Then $\int_0^\infty t^{d_{\rm vol}-\epsilon-1}\tilde{K}(t)dt$ can be expanded as

$$\begin{aligned}\int_0^\infty t^{d_{\rm vol}-\epsilon-1}\tilde{K}_\eta(t)dt &= \int_0^\infty t^{d_{\rm vol}-\epsilon-1}\sup_{\|x\|\le t}|K(x)|^k dt+\int_0^\infty t^{d_{\rm vol}-\epsilon-1}(\tilde{K}_\eta(t)-\sup_{\|x\|\ge t}|K(x)|^k)dt\\ &= C_{k,K,d_{\rm vol},\epsilon}+\eta<\infty.\end{aligned} \tag{34}$$

Now since $\tilde{K}_\eta$ is continuous and strictly decreasing, change of variables $t = \tilde{K}_\eta(u)$ is applicable, and then $\mathbb{E}_P\left[\left|K\left(\frac{x-X}{h}\right)\right|^k\right]$ can be expanded as

$$\begin{aligned}\mathbb{E}_P\left[\left|K\left(\frac{x-X}{h}\right)\right|^k\right] &= \int_0^\infty P\left(\left|K\left(\frac{x-X}{h}\right)\right|^k > t\right)dt \\ &= \int_\infty^0 P\left(\left|K\left(\frac{x-X}{h}\right)\right|^k > \tilde{K}_\eta(u)\right)d\tilde{K}_\eta(u).\end{aligned}$$

Now, from (33) and $\tilde{K}_\eta$ being a strictly decreasing, we can upper bound $\mathbb{E}_P\left[\left|K\left(\frac{x-X}{h}\right)\right|^k\right]$ as

$$\begin{aligned}\mathbb{E}_P\left[\left|K\left(\frac{x-X}{h}\right)\right|^k\right] &\leq \int_\infty^0 P\left(\tilde{K}_\eta\left(\frac{\|x-X\|}{h}\right) > \tilde{K}_\eta(u)\right)d\tilde{K}_\eta(u) \\ &= \int_\infty^0 P\left(\frac{\|x-X\|}{h} < u\right)d\tilde{K}_\eta(u) \\ &= \int_\infty^0 P\left(\mathbb{B}_{\mathbb{R}^d}(x, hu)\right)d\tilde{K}_\eta(u).\end{aligned}$$

Now, from Lemma 4 (and (6) for Assumption 1 case), there exists $C_{d_{\mathrm{vol}}-\epsilon,P} < \infty$ with $P\left(\mathbb{B}_{\mathbb{R}^d}(x,r)\right) \leq C_{d_{\mathrm{vol}}-\epsilon,P}\, r^{d_{\mathrm{vol}}-\epsilon}$ for all $x \in \mathbb{X}$ and $r > 0$. Then $\mathbb{E}_P\left[\left|K\left(\frac{x-X}{h}\right)\right|^k\right]$ is further upper bounded as

$$\begin{aligned}\mathbb{E}_P\left[\left|K\left(\frac{x-X}{h}\right)\right|^k\right] &\leq \int_\infty^0 C_{d_{\mathrm{vol}}-\epsilon,P}(hu)^{d_{\mathrm{vol}}-\epsilon}d\tilde{K}(u) \\ &= C_{d_{\mathrm{vol}}-\epsilon,P}h^{d_{\mathrm{vol}}-\epsilon}\int_\infty^0 u^{d_{\mathrm{vol}}-\epsilon}d\tilde{K}(u). \end{aligned} \tag{35}$$

Now, $\int_\infty^0 u^{d_{\mathrm{vol}}-\epsilon}d\tilde{K}(u)$ can be computed using integration by part. Note first that $\int_0^\infty t^{d_{\mathrm{vol}}-\epsilon-1}\tilde{K}(t)dt < \infty$ implies

$$\lim_{t\to\infty} t^{d_{\mathrm{vol}}-\epsilon}\tilde{K}(t) = 0.$$

To see this, note that $t^{d_{\mathrm{vol}}-\epsilon}\tilde{K}(t)$ is expanded as

$$t^{d_{\mathrm{vol}}-\epsilon}\tilde{K}(t) = \int_0^t u^{d_{\mathrm{vol}}-\epsilon}d\tilde{K}(u) + \int_0^t (d_{\mathrm{vol}}-\epsilon)u^{d_{\mathrm{vol}}-\epsilon-1}\tilde{K}(u)du,$$

then $\int_0^\infty (d_{\mathrm{vol}}-\epsilon)u^{d_{\mathrm{vol}}-\epsilon-1}\tilde{K}(u)du < \infty$ and $\int_0^t u^{d_{\mathrm{vol}}-\epsilon}d\tilde{K}(u)$ being monotone function of t imply that $\lim_{t\to\infty} t^{d_{\mathrm{vol}}-\epsilon}\tilde{K}(t)$ exists. Now, suppose $\lim_{t\to\infty} t^{d_{\mathrm{vol}}-\epsilon}\tilde{K}(t) = a > 0$, then we can choose $t_0 > 0$ such that $t^{d_{\mathrm{vol}}-\epsilon}\tilde{K}(t) > \frac{a}{2}$ for all $t \geq t_0$, and then

$$\infty > \int_0^\infty t^{d_{\mathrm{vol}}-\epsilon-1}\tilde{K}(t)dt \geq \int_{t_0}^\infty t^{d_{\mathrm{vol}}-\epsilon-1}\tilde{K}(t)dt \geq \frac{a}{2}\int_{t_0}^\infty t^{-1}dt = \infty,$$

which is a contradiction. Hence $\lim_{t\to\infty} t^{d_{\mathrm{vol}}-\epsilon}\tilde{K}(t) = 0$. Now, applying integration by part to $\int_\infty^0 u^{d_{\mathrm{vol}}-\epsilon}d\tilde{K}(u)$ with $d_{\mathrm{vol}}-\epsilon > 0$ gives

$$\begin{aligned}\int_\infty^0 u^{d_{\mathrm{vol}}-\epsilon}d\tilde{K}(u) &= \left[u^{d_{\mathrm{vol}}-\epsilon}\tilde{K}(u)\right]_\infty^0 - \int_\infty^0 (d_{\mathrm{vol}}-\epsilon)u^{d_{\mathrm{vol}}-\epsilon-1}\tilde{K}(u)du \\ &= \int_0^\infty (d_{\mathrm{vol}}-\epsilon)u^{d_{\mathrm{vol}}-\epsilon-1}\tilde{K}(u)du. \end{aligned} \tag{36}$$

Then applying (34) and (36) to (35) gives an upper bound for $\mathbb{E}_P\left[\left|K\left(\frac{x-X}{h}\right)\right|^k\right]$ as

$$\mathbb{E}_P\left[\left|K\left(\frac{x-X}{h}\right)\right|^k\right] \leq C_{d_{\mathrm{vol}}-\epsilon,P}(d_{\mathrm{vol}}-\epsilon)h^{d_{\mathrm{vol}}-\epsilon}(C_{k,K,d_{\mathrm{vol}},\epsilon}+\eta). \tag{37}$$

And then note that RHS of (37) holds for any $\eta > 0$, and hence $\mathbb{E}_P\left[\left|K\left(\frac{x-X}{h}\right)\right|^k\right]$ is further upper bounded as

$$\begin{aligned}\mathbb{E}_P\left[\left|K\left(\frac{x-X}{h}\right)\right|^k\right] &\leq \inf_{\eta>0}\left\{C_{d_{\mathrm{vol}}-\epsilon,P}(d_{\mathrm{vol}}-\epsilon)h^{d_{\mathrm{vol}}-\epsilon}(C_{k,K,d_{\mathrm{vol}},\epsilon}+\eta)\right\}\\ &= C_{d_{\mathrm{vol}}-\epsilon,P}(d_{\mathrm{vol}}-\epsilon)C_{k,K,d_{\mathrm{vol}},\epsilon}h^{d_{\mathrm{vol}}-\epsilon}\\ &= C_{k,P,K,\epsilon}h^{d_{\mathrm{vol}}-\epsilon},\end{aligned}$$

where $C_{k,P,K,\epsilon} = C_{d_{\mathrm{vol}}-\epsilon,P}(d_{\mathrm{vol}}-\epsilon)C_{k,K,d_{\mathrm{vol}},\epsilon}$.

□

E.1. Proof for Section 4.1

Theorem 12 follows from applying Theorem 30.

Theorem 12. *Let P be a probability distribution and let K be a kernel function satisfying Assumption 3 and 4. Then, with probability at least $1-\delta$,*

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

for any $\epsilon\in(0,d_{\mathrm{vol}})$, where C is a constant depending only on A, $\|K\|_\infty$, d, ν, d_{vol}, $C_{k=2,P,K,\epsilon}$, ϵ. Further, if $d_{\mathrm{vol}}=0$ or under Assumption 1, ϵ can be 0 in (12).

Proof of Theorem 12. For $x\in\mathbb{X}$ and $h\geq l_n$, let $K_{x,h}:\mathbb{R}^d\to\mathbb{R}$ be $K_{x,h}(\cdot)=K\left(\frac{x-\cdot}{h}\right)$, and let $\tilde{\mathcal{F}}_{K,[l_n,\infty)} := \left\{\frac{1}{h^d}K_{x,h}: x\in\mathbb{X}, h\geq l_n\right\}$ be a class of normalized kernel functions centered on $\mathbb{X}$ and bandwidth in $[l_n,\infty)$. Note that $\hat{p}_h(x)-p_h(x)$ can be expanded as

$$\hat{p}_h(x)-p_h(x)=\frac{1}{nh^d}\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)-\mathbb{E}_P\left[\frac{1}{h^d}K\left(\frac{x-X_i}{h}\right)\right]=\frac{1}{n}\sum_{i=1}^n\frac{1}{h^d}K_{x,h}(X_i)-\mathbb{E}_P\left[\frac{1}{h^d}K_{x,h}\right].$$

Hence $\sup_{h\geq l_n,x\in\mathbb{X}}|\hat{p}_h(x)-p_h(x)|$ can be expanded as

$$\sup_{h\geq l_n,x\in\mathbb{X}}|\hat{p}_h(x)-p_h(x)| = \sup_{f\in\tilde{\mathcal{F}}_{K,[l_n,\infty)}}\left|\frac{1}{n}\sum_{i=1}^n f(X_i)-\mathbb{E}_P[f(X)]\right|. \tag{38}$$

Now, it is immediate to check that

$$\|f\|_\infty \leq l_n^{-d}\|K\|_\infty. \tag{39}$$

For bounding the VC dimension of $\tilde{\mathcal{F}}_{K,[l_n,\infty)}$, consider $\mathcal{F}_{K,[l_n,\infty)} := \{K_{x,h}: x\in\mathbb{X}, h\geq l_n\}$ be a class of unnormalized kernel functions centered on $\mathbb{X}$ and bandwidth in $[l_n,\infty)$. Fix $\eta < l_n^{-d}\|K\|_\infty$ and a probability measure Q on $\mathbb{R}^d$. Suppose $\left[l_n, \left(\frac{\eta}{2\|K\|_\infty}\right)^{-1/d}\right]$ is covered by balls $\left\{\left(h_i-\frac{l_n^{d+1}\eta}{2d\|K\|_\infty}, h_i+\frac{l_n^{d+1}\eta}{2d\|K\|_\infty}\right): 1\leq i\leq N_1\right\}$ and $(\mathcal{F}_{K,[l_n,\infty)}, L_2(Q))$ is covered by balls $\left\{\mathbb{B}_{L_2(Q)}\left(f_j, \frac{l_n^d\eta}{2}\right): 1\leq j\leq N_2\right\}$, and let $f_{i,j} := h_i^{-d}f_j$ for $1\leq i\leq N_1$ and $1\leq j\leq N_2$. Also, choose $h_0 > \left(\frac{\eta}{2\|K\|_\infty}\right)^{-1/d}$, $x_0\in\mathbb{X}$, and let $f_0=\frac{1}{h_0^d}K_{x_0,h_0}$. We will show that

$$\left\{\mathbb{B}_{L_2(Q)}(f_{i,j},\eta): 1\leq i\leq N_1,\, 1\leq j\leq N_2\right\}\cup\left\{\mathbb{B}_{L_2(Q)}(f_0,\eta)\right\} \text{ covers } \tilde{\mathcal{F}}_{K,[l_n,\infty)}. \tag{40}$$

For the first case when $h \leq \left(\frac{\eta}{\|K\|_\infty}\right)^{-1/d}$, find h_i and f_j with $h \in \left(h_i - \frac{l_n^{d+1}\eta}{2d\|K\|_\infty}, h_i + \frac{l_n^{d+1}\eta}{2d\|K\|_\infty}\right)$ and $K_{x,h} \in \mathbb{B}_{L_2(Q)}\left(f_j, \frac{l_n^d \eta}{2}\right)$. Then the distance between $\frac{1}{h^d}K_{x,h}$ and $\frac{1}{h_i^d}f_j$ is upper bounded as

$$\left\|\frac{1}{h^d}K_{x,h} - \frac{1}{h_i^d}f_j\right\|_{L_2(Q)} \leq \left\|\frac{1}{h^d}K_{x,h} - \frac{1}{h_i^d}K_{x,h}\right\|_{L_2(Q)} + \left\|\frac{1}{h_i^d}K_{x,h} - \frac{1}{h_i^d}f_j\right\|_{L_2(Q)}. \tag{41}$$

Now, the first term of (41) is upper bounded as

$$\begin{aligned}\left\|\frac{1}{h^d}K_{x,h} - \frac{1}{h_i^d}K_{x,h}\right\|_{L_2(Q)} &= \left|\frac{1}{h^d} - \frac{1}{h_i^d}\right| \|K_{x,h}\|_{L_2(Q)} \\ &= |h_i - h| \sum_{k=0}^{d-1} h_i^{k-d} h^{-1-k} \|K_{x,h}\|_{L_2(Q)} \\ &\leq |h_i - h|\, d l_n^{-d-1} \|K\|_\infty < \frac{\eta}{2}. \end{aligned} \tag{42}$$

Also, the second term of (41) is upper bounded as

$$\begin{aligned}\left\|\frac{1}{h_i^d}K_{x,h} - \frac{1}{h_i^d}f_j\right\|_{L_2(Q)} &= \frac{1}{h_i^d}\|K_{x,h} - f_j\|_{L_2(Q)} \\ &\leq l_n^{-d}\|K_{x,h} - f_j\|_{L_2(Q)} < \frac{\eta}{2}. \end{aligned} \tag{43}$$

Hence applying (42) and (43) to (41) gives

$$\left\|\frac{1}{h^d}K_{x,h} - \frac{1}{h_i^d}f_j\right\|_{L_2(Q)} < \eta.$$

For the second case when $h > \left(\frac{\eta}{2\|K\|_\infty}\right)^{-1/d}$, $\left\|\frac{1}{h^d}K_{x,h}\right\|_{L_2(Q)} \leq \left\|\frac{1}{h^d}K_{x,h}\right\|_\infty < \frac{\eta}{2}$ holds, and hence

$$\left\|\frac{1}{h^d}K_{x,h} - f_0\right\|_{L_2(Q)} \leq \left\|\frac{1}{h^d}K_{x,h}\right\|_{L_2(Q)} + \|f_0\|_{L_2(Q)} < \eta.$$

Therefore, (40) is shown. Hence combined with Assumption 4 gives that for every probability measure Q on $\mathbb{R}^d$ and for every $\eta \in (0, h^{-d}\|K\|_\infty)$, the covering number $\mathcal{N}(\tilde{\mathcal{F}}_{K,[l_n,\infty)}, L_2(Q), \eta)$ is upper bounded as

$$\begin{aligned}\sup_Q \mathcal{N}(\tilde{\mathcal{F}}_{K,[l_n,\infty)}, L_2(Q), \eta) &\leq \mathcal{N}\left(\left[l_n, \left(\frac{\eta}{2\|K\|_\infty}\right)^{-1/d}\right], |\cdot|, \frac{l_n^{d+1}\eta}{2d\|K\|_\infty}\right) \sup_Q \mathcal{N}\left(\mathcal{F}_{K,[l_n,\infty)}, L_2(Q), \frac{l_n^d \eta}{2}\right) + 1 \\ &\leq \frac{2d\|K\|_\infty}{l_n^{d+1}\eta}\left(\frac{2\|K\|_\infty}{\eta}\right)^{1/d}\left(\frac{2A\|K\|_\infty}{l_n^d\eta}\right)^{\nu} + 1 \\ &\leq \left(\frac{2Ad\|K\|_\infty}{l_n^d \eta}\right)^{\nu+2}. \end{aligned} \tag{44}$$

Also, Lemma 11 implies that under Assumption 3, for any $\epsilon \in (0, d_{\mathrm{vol}})$ (and ϵ can be 0 if $d_{\mathrm{vol}} = 0$ or under Assumption 1),

$$\mathbb{E}_P\left[\left(\frac{1}{h^d}K_{x,h}\right)^2\right] \leq C_{k=2,P,K,\epsilon} l_n^{-2d+d_{\mathrm{vol}}-\epsilon}. \tag{45}$$

Hence from (39), (44), and (45), applying Theorem 30 to (38) gives that $\sup_{h \geq l_n, x \in \mathbb{X}} |\hat{p}_h(x) - p_h(x)|$ is upper bounded with probability at least $1-\delta$ as

$$\begin{aligned}
&\sup_{h\geq l_n,x\in\mathbb{X}}|\hat{p}_h(x)-p_h(x)| \\
&\leq C\left(\frac{2(\nu+2)\left\|K\right\|_{\infty}\log\left(\frac{2Ad\|K\|_{\infty}}{\sqrt{C_{k=2,P,K,\epsilon}}l_n^{(d_{\mathrm{vol}}-\epsilon)/2}}\right)}{nl_n^d}+\sqrt{\frac{2(\nu+2)C_{k=2,P,K,\epsilon}\log\left(\frac{2Ad\|K\|_{\infty}}{\sqrt{C_{k=2,P,K,\epsilon}}l_n^{(d_{\mathrm{vol}}-\epsilon)/2}}\right)}{nl_n^{2d-d_{\mathrm{vol}}+\epsilon}}}\right. \\
&\qquad\left.+\sqrt{\frac{C_{k=2,P,K,\epsilon}\log(\frac{1}{\delta})}{nl_n^{2d-d_{\mathrm{vol}}+\epsilon}}}+\frac{\|K\|_{\infty}\log(\frac{1}{\delta})}{nl_n^d}\right) \\
&\leq C_{A,\|K\|_{\infty},d,\nu,d_{\mathrm{vol}},C_{k=2,P,K,\epsilon}}\left(\frac{\left(\log\left(\frac{1}{l_n}\right)\right)_+}{nl_n^d}+\sqrt{\frac{\left(\log\left(\frac{1}{l_n}\right)\right)_+}{nl_n^{2d-d_{\mathrm{vol}}+\epsilon}}}+\sqrt{\frac{\log\left(\frac{2}{\delta}\right)}{nl_n^{2d-d_{\mathrm{vol}}+\epsilon}}}+\frac{\log\left(\frac{2}{\delta}\right)}{nl_n^d}\right),
\end{aligned}$$

where $C_{A,\|K\|_{\infty},d,\nu,d_{\mathrm{vol}},C_{k=2,P,K,\epsilon},\epsilon}$ depends only on A, $\|K\|_{\infty}$, d, ν, d_{vol}, $C_{k=2,P,K,\epsilon}$, ϵ.

□

Then Corollary 13 is just simplifying the result in Theorem 12.

Corollary 13. *Let P be a probability distribution and let K be a kernel function satisfying Assumption 3 and 4. Fix $\epsilon\in(0,d_{\mathrm{vol}})$. Further, if $d_{\mathrm{vol}}=0$ or under Assumption 1, ϵ can be 0. Suppose*

$$\limsup_n\frac{(\log(1/\ell_n))_+ +\log(2/\delta)}{n\ell_n^{d_{\mathrm{vol}}-\epsilon}}<\infty.$$

Then, with probability at least $1-\delta$,

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

where C' depending only on A, $\|K\|_{\infty}$, d, ν, d_{vol}, $C_{k=2,P,K,\epsilon}$, ϵ.

Proof of Corollary 13. From (12) in Theorem 12, $\sup_{h\geq l_n,x\in\mathbb{X}}|\hat{p}_h(x)-p_h(x)|$ is upper bounded with probability at least $1-\delta$ as

$$\begin{aligned}
&\sup_{h\geq l_n,x\in\mathbb{X}}|\hat{p}_h(x)-p_h(x)| \\
&\leq C_{A,\|K\|_{\infty},d,\nu,d_{\mathrm{vol}},C_{k=2,P,K,\epsilon},\epsilon}\left(\frac{\left(\log\left(\frac{1}{l_n}\right)\right)_+}{nl_n^d}+\sqrt{\frac{\left(\log\left(\frac{1}{l_n}\right)\right)_+}{nl_n^{2d-d_{\mathrm{vol}}+\epsilon}}}+\sqrt{\frac{\log\left(\frac{2}{\delta}\right)}{nl_n^{2d-d_{\mathrm{vol}+\epsilon}+\epsilon}}}+\frac{\log\left(\frac{2}{\delta}\right)}{nl_n^d}\right) \\
&=C_{A,\|K\|_{\infty},d,\nu,d_{\mathrm{vol}},C_{k=2,P,K,\epsilon},\epsilon}\left(\sqrt{\frac{\left(\log\left(\frac{1}{l_n}\right)\right)_+}{nl_n^{2d-d_{\mathrm{vol}}+\epsilon}}}\left(\sqrt{\frac{\left(\log\left(\frac{1}{l_n}\right)\right)_+}{nl_n^{d_{\mathrm{vol}}-\epsilon}}}+1\right)+\sqrt{\frac{\log\left(\frac{2}{\delta}\right)}{nl_n^{2d-d_{\mathrm{vol}}+\epsilon}}}\left(\sqrt{\frac{\log\left(\frac{2}{\delta}\right)}{nl_n^{d_{\mathrm{vol}}-\epsilon}}}+1\right)\right).
\end{aligned}$$

Then from $\limsup_n\frac{\left(\log\left(\frac{1}{l_n}\right)\right)_+ +\log\left(\frac{2}{\delta}\right)}{nl_n^{d_{\mathrm{vol}}-\epsilon}}<\infty$, there exists some constant C' with $\left(\log\left(\frac{1}{l_n}\right)\right)_+ +\log\left(\frac{2}{\delta}\right)\leq C'nl_n^{d_{\mathrm{vol}}+\epsilon}$.

And hence $\sup_{h\geq l_n, x\in\mathbb{X}} |\hat{p}_h(x)-p_h(x)|$ is upper bounded with probability $1-\delta$ as

$$
\begin{aligned}
&\sup_{h\geq l_n, x\in\mathbb{X}} |\hat{p}_h(x)-p_h(x)| \\
&\leq C_{A,\|K\|_\infty,d,\nu,d_{\mathrm{vol}},C_{k=2,P,K,\epsilon},\epsilon}\left(\sqrt{\frac{\left(\log\left(\frac{1}{l_n}\right)\right)_+}{n l_n^{2d-d_{\mathrm{vol}}+\epsilon}}}\left(\sqrt{C'}+1\right)+\sqrt{\frac{\log\left(\frac{1}{\delta}\right)}{n l_n^{2d-d_{\mathrm{vol}}+\epsilon}}}\left(\sqrt{C'}+1\right)\right) \\
&\leq C'_{A,\|K\|_\infty,d,\nu,d_{\mathrm{vol}},C_{k=2,P,K,\epsilon},\epsilon}\sqrt{\frac{\left(\log\left(\frac{1}{l_n}\right)\right)_+ +\log\left(\frac{1}{\delta}\right)}{n l_n^{2d-d_{\mathrm{vol}}+\epsilon}}},
\end{aligned}
$$

where $C'_{A,\|K\|_\infty,d,\nu,d_{\mathrm{vol}},C_{k=2,P,K,\epsilon},\epsilon}$ depending only on A, $\|K\|_\infty$, d, ν, d_{vol}, $C_{k=2,P,K,\epsilon}$, ϵ.

□

E.2. Proof for Section 4.2

Lemma 14 is by covering $\mathbb{X}$ and then using the Lipschitz property of the kernel function K.

Lemma 14. *Suppose there exists $R>0$ with $\mathbb{X}\subset\mathbb{B}_{\mathbb{R}^d}(0,R)$. Let the kernel K is M_K-Lipschitz continuous. Then for all $\eta\in(0,\|K\|_\infty)$, the supremum of the η-covering number $\mathcal{N}(\mathcal{F}_{K,h},L_2(Q),\eta)$ over all measure Q is upper bounded as*

$$
\sup_Q \mathcal{N}(\mathcal{F}_{K,h},L_2(Q),\eta)\leq\left(\frac{2RM_K h^{-1}+\|K\|_\infty}{\eta}\right)^d.
$$

Proof of Lemma 14. For fixed $\eta>0$, let $x_1,\ldots,x_M$ be the maximal η-covering of $\mathbb{B}_{\mathbb{R}^d}(0,R)$, with $M=\mathcal{M}\left(\mathbb{B}_{\mathbb{R}^d}(0,R),\|\cdot\|_2,\eta\right)$ being the packing number of $\mathbb{B}_{\mathbb{R}^d}(0,R)$. Then $\mathbb{B}_{\mathbb{R}^d}(x_i,\eta)$ and $\mathbb{B}_{\mathbb{R}^d}(x_j,\eta)$ do not intersect for any i,j and $\bigcup_{i=1}^M \mathbb{B}_{\mathbb{R}^d}(x_i,\eta)\subset\mathbb{B}_{\mathbb{R}^d}(x_i,R+\eta)$, and hence

$$
\sum_{i=1}^M \lambda_d\left(\mathbb{B}_{\mathbb{R}^d}(x_i,\eta)\right)\leq\lambda_d\left(\mathbb{B}_{\mathbb{R}^d}(x_i,R+\eta)\right). \tag{46}
$$

Then $\lambda_d\left(\mathbb{B}_{\mathbb{R}^d}(x,r)\right)=r^d\lambda_d\left(\mathbb{B}_{\mathbb{R}^d}(0,1)\right)$ gives the upper bound on $\mathcal{M}(\mathbb{B}_{\mathbb{R}^d}(0,R),\|\cdot\|_2,\eta)$ as

$$
\mathcal{M}\left(\mathbb{B}_{\mathbb{R}^d}(0,R),\|\cdot\|_2,\eta\right)\leq\left(1+\frac{R}{\eta}\right)^d.
$$

Then $\mathbb{X}\subset\mathbb{B}_{\mathbb{R}^d}(0,R)$ and the relationship between covering number and packing number gives the upper bound on the covering number $\mathcal{N}\left(\mathbb{X},\|\cdot\|_2,\eta\right)$ as

$$
\mathcal{N}\left(\mathbb{X},\|\cdot\|_2,\eta\right)\leq\mathcal{N}\left(\mathbb{B}_{\mathbb{R}^d}(0,R),\|\cdot\|_2,\eta\right)\leq\mathcal{M}\left(\mathbb{B}_{\mathbb{R}^d}(0,R),\|\cdot\|_2,\frac{\eta}{2}\right)\leq\left(1+\frac{2R}{\eta}\right)^d. \tag{47}
$$

Now, note that for all $x,y\in\mathbb{X}$ and for all $z\in\mathbb{R}^d$, $|K_{x,h}(z)-K_{y,h}(z)|$ is upper bounded as

$$
|K_{x,h}(z)-K_{y,h}(z)|=\left|K\left(\frac{x-z}{h}\right)-K\left(\frac{y-z}{h}\right)\right|\leq\frac{M_K}{h}\|(x-z)-(y-z)\|_2=\frac{M_K}{h}\|x-y\|_2.
$$

Hence for any measure Q on $\mathbb{R}^d$, $\|K_{x,h}-K_{y,h}\|_{L_2(Q)}$ is upper bounded as

$$
\|K_{x,h}-K_{y,h}\|_{L_2(Q)}=\sqrt{\int(K_{x,h}(z)-K_{y,h}(z))^2dQ(z)}\leq\frac{M_K}{h}\|x-y\|_2.
$$

Hence applying this to (47) implies that for all $\eta > 0$, the supremum of the covering number $\mathcal{N}(\mathcal{F}_{K,h}, L_2(Q), \eta)$ over all measure Q is upper bounded as

$$\sup_Q \mathcal{N}(\mathcal{F}_{K,h}, L_2(Q), \eta) \leq \mathcal{N}\left(\mathbb{X}, \|\cdot\|_2, \frac{h\eta}{M_K}\right) \leq \left(1 + \frac{2RM_K}{h\eta}\right)^d.$$

Hence for all $\eta \in (0, \|K\|_\infty)$,

$$\sup_Q \mathcal{N}(\mathcal{F}_{K,h}, L_2(Q), \eta) \leq \left(\frac{2RM_K h^{-1} + \|K\|_\infty}{\eta}\right)^d.$$

□

Then Corollary 15 follows from applying Theorem 30 with bounding the covering number from Lemma 14.

Corollary 15. *Suppose there exists $R > 0$ with $\mathbb{X} \subset \mathbb{B}_{\mathbb{R}^d}(0, R)$. Let K be a M_K-Lipschitz continuous kernel function satisfying Assumption 3. Fix $\epsilon \in (0, d_{\mathrm{vol}})$. Further, if $d_{\mathrm{vol}} = 0$ or under Assumption 1, ϵ can be 0. Suppose*

$$\limsup_n \frac{(\log(1/h_n))_+ + \log(2/\delta)}{nh_n^{d_{\mathrm{vol}}-\epsilon}} < \infty.$$

Then with probability at least $1 - \delta$,

$$\sup_{x \in \mathbb{X}} |\hat{p}_{h_n}(x) - p_{h_n}(x)| \leq C'' \sqrt{\frac{(\log(\frac{1}{h_n}))_+ + \log(\frac{2}{\delta})}{nh_n^{2d - d_{\mathrm{vol}} + \epsilon}}},$$

where C'' is a constant depending only on R, M_K, $\|K\|_\infty$, d, ν, d_{vol}, $C_{k=2,P,K,\epsilon}$, ϵ.

Proof of Corollary 15. For $x \in \mathbb{X}$, let $K_{x,h} : \mathbb{R}^d \to \mathbb{R}$ be $K_{x,h}(\cdot) = K\left(\frac{x - \cdot}{h}\right)$, and let $\tilde{\mathcal{F}}_{K,h} := \left\{\frac{1}{h^d} K_{x,h} : x \in \mathbb{X}\right\}$ be a class of normalized kernel functions centered on $\mathbb{X}$ and bandwidth h. Note that $\hat{p}_h(x) - p_h(x)$ can be expanded as

$$\hat{p}_h(x) - p_h(x) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) - \mathbb{E}_P\left[\frac{1}{h^d} K\left(\frac{x - X_i}{h}\right)\right] = \frac{1}{n} \sum_{i=1}^n \frac{1}{h^d} K_{x,h}(X_i) - \mathbb{E}_P\left[\frac{1}{h^d} K_{x,h}\right].$$

Hence $\sup_{x \in \mathbb{X}} |\hat{p}_h(x) - p_h(x)|$ can be expanded as

$$\sup_{x \in \mathbb{X}} |\hat{p}_h(x) - p_h(x)| = \sup_{f \in \tilde{\mathcal{F}}_{K,h}} \left|\frac{1}{n} \sum_{i=1}^n f(X_i) - \mathbb{E}_P[f(X)]\right|. \tag{48}$$

Now, it is immediate to check that

$$\|f\|_\infty \leq h^{-d} \|K\|_\infty. \tag{49}$$

Also, Since $\tilde{\mathcal{F}}_{K,h} = h^{-d}\mathcal{F}_{K,h}$, VC dimension is uniformly bounded as Lemma 14 gives that for every probability measure Q on $\mathbb{R}^d$ and for every $\eta \in (0, h^{-d}\|K\|_\infty)$, the covering number $\mathcal{N}(\tilde{\mathcal{F}}_{K,h}, L_2(Q), \eta)$ is upper bounded as

$$\begin{aligned}
\sup_Q \mathcal{N}(\tilde{\mathcal{F}}_{K,h}, L_2(Q), \eta) &= \sup_Q \mathcal{N}(\mathcal{F}_{K,h}, L_2(Q), h^d\eta) \\
&\leq \left(\frac{2RM_K h^{-1} + \|K\|_\infty}{h^d \eta}\right)^d \\
&\leq \left(\frac{2RM_K \|K\|_\infty}{h^{d+1}\eta}\right)^d.
\end{aligned} \tag{50}$$

Also, Lemma 11 implies that under Assumption 3, for any $\epsilon \in (0, d_{\mathrm{vol}})$ (and ϵ can be 0 if $d_{\mathrm{vol}} = 0$ or under Assumption 1),

$$\mathbb{E}_P\left[\left(\frac{1}{h^d} K_{x,h}\right)^2\right] \leq C_{k=2,P,K,\epsilon} h^{-2d + d_{\mathrm{vol}} - \epsilon}. \tag{51}$$

Hence from (49), (50), and (51), applying Theorem 30 to (48) gives that $\sup_{x\in\mathbb{X}}|\hat{p}_h(x)-p_h(x)|$ is upper bounded with probability at least $1-\delta$ as

$$\begin{aligned}
&\sup_{x\in\mathbb{X}}|\hat{p}_h(x)-p_h(x)| \\
&\leq C\left(\frac{2d\left\|K\right\|_\infty \log\left(\frac{2RM_K\|K\|_\infty}{\sqrt{C_{k=2,P,K,\epsilon}}h^{1+(d_{\mathrm{vol}}-\epsilon)/2}}\right)}{nh^d}+\sqrt{\frac{2dC_{k=2,P,K,\epsilon}\log\left(\frac{2RM_K\|K\|_\infty}{\sqrt{C_{k=2,P,K,\epsilon}}h^{1+(d_{\mathrm{vol}}-\epsilon)/2}}\right)}{nh^{2d-d_{\mathrm{vol}}+\epsilon}}}\right. \\
&\qquad\left.+\sqrt{\frac{C_{k=2,P,K,\epsilon}\log(\frac{1}{\delta})}{nh^{2d-d_{\mathrm{vol}}+\epsilon}}}+\frac{\left\|K\right\|_\infty\log(\frac{1}{\delta})}{nh^d}\right) \\
&\leq C_{R,M_K,\|K\|_\infty,d,\nu,d_{\mathrm{vol}},C_{k=2,P,K,\epsilon},\epsilon}\left(\frac{\left(\log\left(\frac{1}{h}\right)\right)_+}{nh^d}+\sqrt{\frac{\left(\log\left(\frac{1}{h}\right)\right)_+}{nh^{2d-d_{\mathrm{vol}}+\epsilon}}}+\sqrt{\frac{\log\left(\frac{2}{\delta}\right)}{nh^{2d-d_{\mathrm{vol}}+\epsilon}}}+\frac{\log\left(\frac{2}{\delta}\right)}{nh^d}\right),
\end{aligned}$$

where $C_{R,M_K,\|K\|_\infty,d,\nu,d_{\mathrm{vol}},C_{k=2,P,K,\epsilon},\epsilon}$ depends only on R, M_K, $\|K\|_\infty$, d, ν, d_{vol}, $C_{k=2,P,K,\epsilon}$, ϵ. □

F. Proof for Section 5

Proposition 16 is shown by finding $x_0\in\mathbb{X}$ where the volume dimension is obtained, and analyzing the behavior of $|\hat{p}_{h_n}(x_0)-p_{h_n}(x_0)|$ by applying Central Limit Theorem.

Proposition 16. *Suppose P is a distribution satisfying Assumption 2 and with positive volume dimension $d_{\mathrm{vol}}>0$. 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 $\lim_n nh_n^{d_{\mathrm{vol}}}=\infty$. Then, with probability $1-\delta$, the following holds for all large enough n and small enough h_n:*

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

where $C_{P,K,\delta}$ is a constant depending only on P, K,and δ.

Proof of Proposition 16. Note that $\lim_{t\to 0}\inf_{\|x\|\leq t}K(x)>0$ implies that there exists $t_0,K_0\in(0,\infty)$ such that

$$K(x)\geq K_0 I(\|x\|\leq t_0). \tag{52}$$

Also, from $\sup_{x\in\mathbb{X}}\liminf_{r\to 0}\frac{P(\mathbb{B}_{\mathbb{R}^d}(x,r))}{r^{d_{\mathrm{vol}}}}>0$, we can choose $x_0\in\mathbb{X}$ such that $\liminf_{r\to 0}\frac{P(\mathbb{B}_{\mathbb{R}^d}(x_0,r))}{r^{d_{\mathrm{vol}}}}>0$. From $\{h_n\}_{n\in\mathbb{N}}$ bounded, there exists $r_0>0$ and $p_0>0$ such that $r_0\geq h_n t_0$ for all $n\in\mathbb{N}$ and for all $r\leq r_0$,

$$P(\mathbb{B}_{\mathbb{R}^d}(x_0,r))\geq p_0 r^{d_{\mathrm{vol}}}. \tag{53}$$

For $x\in\mathbb{X}$ and $h>0$, let $f_{x,h}:\mathbb{R}^d\to\mathbb{R}$ be $f_{x,h}=\frac{1}{h^d}\left(K_{x,h}-\mathbb{E}_P[K_{x,h}]\right)$, so that at $x_0\in\mathbb{X}$, $\hat{p}_{h_n}(x_0)-p_{h_n}(x_0)$ is expanded as

$$\hat{p}_{h_n}(x_0)-p_{h_n}(x_0)=\frac{1}{n}\sum_{i=1}^n f_{x_0,h_n}(X_i).$$

Below we get a lower bound for $\mathbb{E}_P[f_{x_0,h_n}^2]$. First, fix $\epsilon<\frac{d_{\mathrm{vol}}}{2}$. Then from Lemma 11,

$$\mathbb{E}_P\left[|K_{x_0,h}|\right]\leq C_{k=1,P,K,\epsilon}h^{d_{\mathrm{vol}}-\epsilon}. \tag{54}$$

Now, we lower bound $\mathbb{E}_P[K_{x_0,h}^2]$. By applying (52), $\mathbb{E}_P[K_{x_0,h}^2]$ is lower bounded as

$$\begin{aligned}
\mathbb{E}_P\left[K_{x_0,h}^2\right]&\geq\mathbb{E}_P\left[K_0 I\left(\left\|\frac{x_0-X_i}{h}\right\|\geq t_0\right)\right]\\
&=K_0^2 P(\mathbb{B}_{\mathbb{R}^d}(x_0,ht_0)).
\end{aligned}$$

Then applying (53) gives a further lower bound as

$$\mathbb{E}_P\left[K_{x_0,h}^2\right] \geq K_0^2 p_0 t_0^{d_{\mathrm{vol}}} h^{d_{\mathrm{vol}}}. \tag{55}$$

Then combining (54) and (55) gives a lower bound of $\mathbb{E}_P[f_{x_0,h}^2]$ as

$$\begin{aligned}\mathbb{E}_P\left[f_{x_0,h}^2\right] &= \frac{1}{h^{2d}}\left(\mathbb{E}_P\left[K_{x_0,h}^2\right] - \left(\mathbb{E}_P\left[K_{x_0,h}\right]\right)^2\right)\\ &\geq h^{d_{\mathrm{vol}}-2d}(K_0^2 p_0 t_0^{d_{\mathrm{vol}}} - C_{k=1,P,K,\epsilon}^2 h^{d_{\mathrm{vol}}-2\epsilon}).\end{aligned}$$

Hence from $d_{\mathrm{vol}} - 2\epsilon > 0$, there exists $h_{P,K}$ and $C'_{P,K}$ depending only on P and K such that $h_n \leq h_{P,K}$ implies

$$\mathbb{E}_P\left[f_{x_0,h_n}^2\right] \geq C'_{P,K} h_n^{d_{\mathrm{vol}}-2d}. \tag{56}$$

Now, let $s_n := \sqrt{\sum_{i=1}^n \mathbb{E}_P[f_{x_0,h_n}^2(X_i)]}$. Then (56) gives

$$s_n \geq \sqrt{C'_{P,K} n h_n^{d_{\mathrm{vol}}-2d}}.$$

Then for any $\epsilon > 0$, when n is large enough so that $nh_n^{d_{\mathrm{vol}}} > \frac{\|K\|_\infty^2}{\epsilon^2 C'_{P,K}}$, then

$$\|f_{x_0,h_n}\|_\infty \leq h^{-d}\|K\|_\infty < \epsilon\sqrt{C'_{P,K} n h_n^{d_{\mathrm{vol}}-2d}} \leq s_n.$$

Hence Lindeberg condition holds as for n large enough so that $nh_n^{d_{\mathrm{vol}}} > \frac{\|K\|_\infty^2}{\epsilon^2 C'_{P,K}}$, then

$$\frac{1}{s_n^2}\sum_{i=1}^n \mathbb{E}\left[f_{x_0,h_n}^2(X_i) I\left(|f_{x_0,h_n}(X_i)| \geq \epsilon s_n\right)\right] = 0.$$

Hence, Lindeberg-Feller Central Limit Theorem gives

$$\sqrt{\frac{n}{\mathbb{E}_P[f_{x_0,h_n}^2]}}(\hat{p}_{h_n}(x_0) - p_{h_n}(x_0)) \xrightarrow{d} N\left(0,1\right).$$

Hence, for fixed $\delta \in (0,1)$, let $q_{\delta/2} \in \mathbb{R}$ be such that $P(|Z| \leq q_{\delta/2}) = \frac{\delta}{2}$ for $Z \sim N(0,1)$, then

$$\lim_{n\to\infty} P\left(\left|\sqrt{\frac{n}{\mathbb{E}_P[f_{x_0,h_n}^2]}}(\hat{p}_{h_n}(x_0) - p_{h_n}(x_0))\right| \geq q_{\delta/2}\right) = 1 - \frac{\delta}{2}.$$

And hence there exists $N < \infty$ that for all $n \geq N$,

$$P\left(|\hat{p}_{h_n}(x_0) - p_{h_n}(x_0)| \geq q_{\delta/2}\sqrt{\frac{\mathbb{E}_P[f_{x_0,h_n}^2]}{n}}\right) \geq 1 - \delta.$$

Then applying (56) implies that with probability at least $1-\delta$,

$$|\hat{p}_{h_n}(x_0) - p_{h_n}(x_0)| \geq \sqrt{\frac{q_{\delta/2}^2 C'_{P,K}}{nh_n^{2d-d_{\mathrm{vol}}}}} = C_{P,K,\delta}\sqrt{\frac{1}{nh_n^{2d-d_{\mathrm{vol}}}}},$$

where $C_{P,K,\delta} = q_{\delta/2}\sqrt{C'_{P,K}}$ depends only on P, K, and δ. Then from

$$\sup_{x\in\mathbb{X}} |\hat{p}_{h_n}(x) - p_{h_n}(x)| \geq |\hat{p}_{h_n}(x_0) - p_{h_n}(x_0)|,$$

we get the same lower bound for $\sup_{x\in\mathbb{X}} |\hat{p}_{h_n}(x) - p_{h_n}(x)|$ with probability at least $1-\delta$ as

$$\sup_{x\in\mathbb{X}} |\hat{p}_{h_n}(x) - p_{h_n}(x)| \geq \sqrt{\frac{q_{\delta/2}^2 C'_{P,K}}{nh_n^{2d-d_{\mathrm{vol}}}}} = C_{P,K,\delta}\sqrt{\frac{1}{nh_n^{2d-d_{\mathrm{vol}}}}}.$$

$\square$

G. Proof for Section 6

For showing Lemma 19, we proceed similarly to proof of Lemma 11, where we plug in $D^s K$ in the place of K.

Lemma 19. *Let $(\mathbb{R}^d, P)$ be a probability space and let $X \sim P$. For any kernel K satisfying Assumption 6, the expectation of the square of the derivative of the kernel is upper bounded as*

$$\mathbb{E}_P\left[\left(D^s K\left(\frac{x-X}{h}\right)\right)^2\right] \leq C_{s,P,K,\epsilon} h^{d_{\mathrm{vol}}-\epsilon},$$

for any $\epsilon \in (0, d_{\mathrm{vol}})$, where $C_{s,P,K,\epsilon}$ is a constant depending only on s, P, K, ϵ. Further, if $d_{\mathrm{vol}} = 0$ or under Assumption 1, ϵ can be 0 in (18).

Proof of Lemma 19. We first consider the case when $d_{\mathrm{vol}} = 0$. Then $\mathbb{E}_P\left[\left(D^s K\left(\frac{x-X}{h}\right)\right)^2\right]$ is simply bounded as

$$\mathbb{E}_P\left[\left(D^s K\left(\frac{x-X}{h}\right)\right)^2\right] \leq \|D^s K\|_\infty^2 h^0.$$

Now, we consider the case when $d_{\mathrm{vol}} > 0$. Fix $\epsilon \in (0, d_{\mathrm{vol}})$. Under Assumption 1, ϵ can be chosen to be 0.

Let $C_{s,K,d_{\mathrm{vol}},\epsilon} := \int_0^\infty t^{d_{\mathrm{vol}}-\epsilon-1} \sup_{\|x\|\leq t}(D^s K(x))^2 dt$, then it is finite from (17) and $\|D^s K\|_\infty < \infty$ in Assumption 7 as

$$\begin{aligned}\int_0^\infty t^{d_{\mathrm{vol}}-\epsilon-1} \sup_{\|x\|\leq t}(D^s K(x))^2 dt &\leq \int_0^1 t^{d_{\mathrm{vol}}-\epsilon-1} \|D^s K\|_\infty dt + \int_1^\infty t^{d_{\mathrm{vol}}-1} \sup_{\|x\|\leq t}(D^s K(x))^2 dt \\ &\leq \frac{\|D^s K\|_\infty}{d_{\mathrm{vol}}-\epsilon} + \int_0^\infty t^{d_{\mathrm{vol}}-1} \sup_{\|x\|\leq t}(D^s K(x))^2 dt < \infty.\end{aligned}$$

Fix $\eta > 0$, and let $\tilde{K}_\eta : [0,\infty) \to \mathbb{R}$ be a continuous and strictly decreasing function satisfying $\tilde{K}_\eta(t) > \sup_{\|x\|\geq t}(D^s K(x))^2$ for all $t \geq 0$ and $\int_0^\infty t^{d_{\mathrm{vol}}-\epsilon-1}(\tilde{K}_\eta(t) - \sup_{\|x\|\geq t}(D^s K(x))^2)dt = \eta$. Such existence is possible since $t \mapsto \sup_{\|x\|\geq t}(D^s K(x))^2$ is nonincreasing function, so have at most countable discontinuous points, and $\int_0^\infty t^{d_{\mathrm{vol}}-\epsilon-1} \sup_{\|x\|\leq t}(D^s K(x))^2 dt < \infty$. Then it is immediate to check that

$$(D^s K(x))^2 < \tilde{K}_\eta(\|x\|) \text{ for all } x \in \mathbb{R}. \tag{57}$$

Then $\int_0^\infty t^{d_{\mathrm{vol}}-\epsilon-1}\tilde{K}(t)dt$ can be expanded as

$$\begin{aligned}\int_0^\infty t^{d_{\mathrm{vol}}-\epsilon-1}\tilde{K}_\eta(t)dt &= \int_0^\infty t^{d_{\mathrm{vol}}-\epsilon-1} \sup_{\|x\|\leq t}(D^s K(x))^2 dt + \int_0^\infty t^{d_{\mathrm{vol}}-\epsilon-1}(\tilde{K}_\eta(t) - \sup_{\|x\|\geq t}(D^s K(x))^2)dt \\ &= C_{s,K,d_{\mathrm{vol}},\epsilon} + \eta < \infty.\end{aligned} \tag{58}$$

Now since $\tilde{K}_\eta$ is continuous and strictly decreasing, change of variables $t = \tilde{K}_\eta(u)$ is applicable, and then $\mathbb{E}_P\left[\left(D^s K\left(\frac{x-X}{h}\right)\right)^2\right]$ can be expanded as

$$\begin{aligned}\mathbb{E}_P\left[\left(D^s K\left(\frac{x-X}{h}\right)\right)^2\right] &= \int_0^\infty P\left(\left(D^s K\left(\frac{x-X}{h}\right)\right)^2 > t\right) dt \\ &= \int_\infty^0 P\left(\left(D^s K\left(\frac{x-X}{h}\right)\right)^2 > \tilde{K}_\eta(u)\right) d\tilde{K}_\eta(u).\end{aligned}$$

Now, from (57) and $\tilde{K}_\eta$ being a strictly decreasing, we can upper bound $\mathbb{E}_P\left[\left(D^s K\left(\frac{x-X}{h}\right)\right)^2\right]$ as

$$\begin{aligned}\mathbb{E}_P\left[\left(D^s K\left(\frac{x-X}{h}\right)\right)^2\right] &\leq \int_\infty^0 P\left(\tilde{K}_\eta\left(\frac{\|x-X\|}{h}\right) > \tilde{K}_\eta(u)\right) d\tilde{K}_\eta(u) \\ &= \int_\infty^0 P\left(\frac{\|x-X\|}{h} < u\right) d\tilde{K}_\eta(u) \\ &= \int_\infty^0 P\left(\mathbb{B}_{\mathbb{R}^d}(x, hu)\right) d\tilde{K}_\eta(u).\end{aligned}$$

Now, from Lemma 4 (and (6) for Assumption 1 case), there exists $C_{d_{\mathrm{vol}}-\epsilon,P} < \infty$ with $P\left(\mathbb{B}_{\mathbb{R}^d}(x,r)\right) \leq C_{d_{\mathrm{vol}}-\epsilon,P}\, r^{d_{\mathrm{vol}}-\epsilon}$ for all $x \in \mathbb{X}$ and $r > 0$. Then $\mathbb{E}_P\left[\left(D^s K\left(\frac{x-X}{h}\right)\right)^2\right]$ is further upper bounded as

$$\begin{aligned}\mathbb{E}_P\left[\left(D^s K\left(\frac{x-X}{h}\right)\right)^2\right] &\leq \int_\infty^0 C_{d_{\mathrm{vol}}-\epsilon,P}(hu)^{d_{\mathrm{vol}}-\epsilon} d\tilde{K}(u) \\ &= C_{d_{\mathrm{vol}}-\epsilon,P}\, h^{d_{\mathrm{vol}}-\epsilon} \int_\infty^0 u^{d_{\mathrm{vol}}-\epsilon} d\tilde{K}(u).\end{aligned} \tag{59}$$

Now, $\int_\infty^0 u^{d_{\mathrm{vol}}-\epsilon} d\tilde{K}(u)$ can be computed using integration by part. Note first that $\int_0^\infty t^{d_{\mathrm{vol}}-\epsilon-1}\tilde{K}(t)dt < \infty$ implies

$$\lim_{t\to\infty} t^{d_{\mathrm{vol}}-\epsilon}\tilde{K}(t) = 0.$$

To see this, note that $t^{d_{\mathrm{vol}}-\epsilon}\tilde{K}(t)$ is expanded as

$$t^{d_{\mathrm{vol}}-\epsilon}\tilde{K}(t) = \int_0^t u^{d_{\mathrm{vol}}-\epsilon} d\tilde{K}(u) + \int_0^t (d_{\mathrm{vol}}-\epsilon) u^{d_{\mathrm{vol}}-\epsilon-1}\tilde{K}(u)du,$$

then $\int_0^\infty (d_{\mathrm{vol}}-\epsilon)u^{d_{\mathrm{vol}}-\epsilon-1}\tilde{K}(u)du < \infty$ and $\int_0^t u^{d_{\mathrm{vol}}-\epsilon}d\tilde{K}(u)$ being monotone function of t imply that $\lim_{t\to\infty} t^{d_{\mathrm{vol}}-\epsilon}\tilde{K}(t)$ exists. Now, suppose $\lim_{t\to\infty} t^{d_{\mathrm{vol}}-\epsilon}\tilde{K}(t) = a > 0$, then we can choose $t_0 > 0$ such that $t^{d_{\mathrm{vol}}-\epsilon}\tilde{K}(t) > \frac{a}{2}$ for all $t \geq t_0$, and then

$$\infty > \int_0^\infty t^{d_{\mathrm{vol}}-\epsilon-1}\tilde{K}(t)dt \geq \int_{t_0}^\infty t^{d_{\mathrm{vol}}-\epsilon-1}\tilde{K}(t)dt \geq \frac{a}{2}\int_{t_0}^\infty t^{-1}dt = \infty,$$

which is a contradiction. Hence $\lim_{t\to\infty} t^{d_{\mathrm{vol}}-\epsilon}\tilde{K}(t) = 0$. Now, applying integration by part to $\int_\infty^0 u^{d_{\mathrm{vol}}-\epsilon}d\tilde{K}(u)$ with $d_{\mathrm{vol}} - \epsilon > 0$ gives

$$\begin{aligned}\int_\infty^0 u^{d_{\mathrm{vol}}-\epsilon}d\tilde{K}(u) &= \left[u^{d_{\mathrm{vol}}-\epsilon}\tilde{K}(u)\right]_\infty^0 - \int_\infty^0 (d_{\mathrm{vol}}-\epsilon)u^{d_{\mathrm{vol}}-\epsilon-1}\tilde{K}(u)du \\ &= \int_0^\infty (d_{\mathrm{vol}}-\epsilon)u^{d_{\mathrm{vol}}-\epsilon-1}\tilde{K}(u)du.\end{aligned} \tag{60}$$

Then applying (58) and (60) to (59) gives an upper bound for $\mathbb{E}_P\left[\left(D^s K\left(\frac{x-X}{h}\right)\right)^2\right]$ as

$$\mathbb{E}_P\left[\left(D^s K\left(\frac{x-X}{h}\right)\right)^2\right] \leq C_{d_{\mathrm{vol}}-\epsilon,P}(d_{\mathrm{vol}}-\epsilon)h^{d_{\mathrm{vol}}-\epsilon}(C_{s,K,d_{\mathrm{vol}},\epsilon} + \eta). \tag{61}$$

And then note that RHS of (61) holds for any $\eta > 0$, and hence $\mathbb{E}_P\left[\left(D^s K\left(\frac{x-X}{h}\right)\right)^2\right]$ is further upper bounded as

$$\begin{aligned}\mathbb{E}_P\left[\left(D^s K\left(\frac{x-X}{h}\right)\right)^2\right] &\leq \inf_{\eta>0}\left\{C_{d_{\mathrm{vol}}-\epsilon,P}(d_{\mathrm{vol}}-\epsilon)h^{d_{\mathrm{vol}}-\epsilon}(C_{s,K,d_{\mathrm{vol}},\epsilon}+\eta)\right\} \\ &= C_{d_{\mathrm{vol}}-\epsilon,P}(d_{\mathrm{vol}}-\epsilon)C_{s,K,d_{\mathrm{vol}},\epsilon}h^{d_{\mathrm{vol}}-\epsilon} \\ &= C_{s,P,K,\epsilon}h^{d_{\mathrm{vol}}-\epsilon},\end{aligned}$$

where $C_{k,P,K,\epsilon} = C_{d_{\mathrm{vol}}-\epsilon,P}(d_{\mathrm{vol}} - \epsilon)C_{s,K,d_{\mathrm{vol}},\epsilon}$.

$\square$

For proving Theorem 20, we proceed similarly to the proof of Theorem 12. Analogous to bounding $\mathbb{E}_P[K_{x,h}^2]$ by Lemma 11, we bound $\mathbb{E}_P[(D^s K_{x,h})^2]$ by Lemma 19.

Theorem 20. *Let P be a distribution and K be a kernel function satisfying Assumption 5, 6, and 7. Then, with probability at least $1-\delta$,*

$$\sup_{h\geq l_n, x\in\mathbb{X}} |D^s\hat{p}_h(x) - D^s p_h(x)| \leq C\left(\frac{(\log(1/l_n))_+}{nl_n^{d+|s|}} + \sqrt{\frac{(\log(1/l_n))_+}{nl_n^{2d+2|s|-d_{\mathrm{vol}}+\epsilon}}} + \sqrt{\frac{\log(2/\delta)}{nl_n^{2d+2|s|-d_{\mathrm{vol}}+\epsilon}}} + \frac{\log(2/\delta)}{nl_n^{d+|s|}}\right),$$

for any $\epsilon \in (0, d_{\mathrm{vol}})$, where C is a constant depending only on A, $\|D^s K\|_\infty$, d, ν, d_{vol}, $C_{s,P,K,\epsilon}$, ϵ. Further, if $d_{\mathrm{vol}} = 0$ or under Assumption 1, ϵ can be 0 in (19).

Proof of Theorem 20. For $x \in \mathbb{X}$ and $h \geq l_n$, let $D^s K_{x,h} : \mathbb{R}^d \to \mathbb{R}$ be $D^s K_{x,h}(\cdot) = D^s K\left(\frac{x-\cdot}{h}\right)$, and let $\tilde{\mathcal{F}}_{K,[l_n,\infty)}^s := \left\{\frac{1}{h^{d+|s|}} D^s K_{x,h} : x \in \mathbb{X}, h \geq l_n\right\}$ be a class of normalized kernel functions centered on $\mathbb{X}$ and bandwidth in $[l_n, \infty)$. Note that $D^s\hat{p}_h(x) - D^s p_h(x)$ can be expanded as

$$\begin{aligned} D^s\hat{p}_h(x) - D^s p_h(x) &= \frac{1}{nh^{d+|s|}}\sum_{i=1}^n D^s K\left(\frac{x-X_i}{h}\right) - \mathbb{E}_P\left[\frac{1}{h^{d+|s|}} D^s K\left(\frac{x-X_i}{h}\right)\right] \\ &= \frac{1}{n}\sum_{i=1}^n \frac{1}{h^{d+|s|}} D^s K_{x,h}(X_i) - \mathbb{E}_P\left[\frac{1}{h^{d+|s|}} D^s K_{x,h}\right]. \end{aligned}$$

Hence $\sup_{h\geq l_n, x\in\mathbb{X}} |D^s\hat{p}_h(x) - D^s p_h(x)|$ can be expanded as

$$\sup_{h\geq l_n, x\in\mathbb{X}} |D^s\hat{p}_h(x) - D^s p_h(x)| = \sup_{f\in\tilde{\mathcal{F}}_{K,[l_n,\infty)}^s}\left|\frac{1}{n}\sum_{i=1}^n f(X_i) - \mathbb{E}_P[f(X)]\right|. \tag{62}$$

Now, it is immediate to check that

$$\|f\|_\infty \leq l_n^{-d-|s|}\|D^s K\|_\infty. \tag{63}$$

For bounding the VC dimension of $\tilde{\mathcal{F}}_{K,[l_n,\infty)}^s$, consider $\mathcal{F}_{K,[l_n,\infty)}^s := \{D^s K_{x,h} : x \in \mathbb{X}, h \geq l_n\}$ be a class of unnormalized kernel functions centered on $\mathbb{X}$ and bandwidth in $[l_n, \infty)$. Fix $\eta < l_n^{-d-|s|}\|D^s K\|_\infty$ and a probability measure Q on $\mathbb{R}^d$. Suppose $\left[l_n, \left(\frac{\eta}{2\|D^s K\|_\infty}\right)^{-1/(d+|s|)}\right]$ is covered by balls $\left\{\left(h_i - \frac{l_n^{d+|s|+1}\eta}{2(d+|s|)\|D^s K\|_\infty}, h_i + \frac{l_n^{d+|s|+1}\eta}{2(d+|s|)\|D^s K\|_\infty}\right) : 1 \leq i \leq N_1\right\}$ and $(\mathcal{F}_{K,[l_n,\infty)}^s, L_2(Q))$ is covered by balls $\left\{\mathbb{B}_{L_2(Q)}\left(f_j, \frac{l_n^{d+|s|}\eta}{2}\right) : 1 \leq j \leq N_2\right\}$, and let $f_{i,j} := h_i^{-d-|s|} f_j$ for $1 \leq i \leq N_1$ and $1 \leq j \leq N_2$. Also, choose $h_0 > \left(\frac{\eta}{2\|D^s K\|_\infty}\right)^{-1/(d+|s|)}$, $x_0 \in \mathbb{X}$, and let $f_0 = \frac{1}{h_0^{d+|s|}} D^s K_{x_0,h_0}$. We will show that

$$\left\{\mathbb{B}_{L_2(Q)}(f_{i,j}, \eta) : 1 \leq i \leq N_1, 1 \leq j \leq N_2\right\} \cup \left\{\mathbb{B}_{L_2(Q)}(f_0, \eta)\right\} \text{ covers } \tilde{\mathcal{F}}_{K,[l_n,\infty)}^s. \tag{64}$$

For the first case when $h \leq \left(\frac{\eta}{2\|D^s K\|_\infty}\right)^{-1/(d+|s|)}$, find h_i and f_j with $h \in \left(h_i - \frac{l_n^{d+|s|+1}\eta}{2(d+|s|)\|D^s K\|_\infty}, h_i + \frac{l_n^{d+|s|+1}\eta}{2(d+|s|)\|D^s K\|_\infty}\right)$ and $K_{x,h} \in \mathbb{B}_{L_2(Q)}\left(f_j, \frac{l_n^{d+|s|}\eta}{2}\right)$. Then the distance between $\frac{1}{h^{d+|s|}} D^s K_{x,h}$ and $\frac{1}{h_i^{d+|s|}} f_j$ is upper bounded as

$$\left\|\frac{1}{h^{d+|s|}} D^s K_{x,h} - \frac{1}{h_i^{d+|s|}} f_j\right\|_{L_2(Q)} \leq \left\|\frac{1}{h^{d+|s|}} D^s K_{x,h} - \frac{1}{h_i^{d+|s|}} D^s K_{x,h}\right\|_{L_2(Q)} + \left\|\frac{1}{h_i^{d+|s|}} D^s K_{x,h} - \frac{1}{h_i^{d+|s|}} f_j\right\|_{L_2(Q)}. \tag{65}$$

Now, the first term of (65) is upper bounded as

$$\begin{aligned}\left\|\frac{1}{h^{d+|s|}}D^sK_{x,h}-\frac{1}{h_i^{d+|s|}}D^sK_{x,h}\right\|_{L_2(Q)}&=\left|\frac{1}{h^{d+|s|}}-\frac{1}{h_i^{d+|s|}}\right|\left\|D^sK_{x,h}\right\|_{L_2(Q)}\\&=|h_i-h|\sum_{k=0}^{d+|s|-1}h_i^{k-d-|s|}h^{-1-k}\left\|D^sK_{x,h}\right\|_{L_2(Q)}\\&\leq|h_i-h|\,(d+|s|)l_n^{-d-|s|-1}\left\|D^sK\right\|_\infty<\frac{\eta}{2}.\end{aligned}\tag{66}$$

Also, the second term of (65) is upper bounded as

$$\begin{aligned}\left\|\frac{1}{h_i^{d+|s|}}D^sK_{x,h}-\frac{1}{h_i^{d+|s|}}f\right\|_{L_2(Q)}&=\frac{1}{h_i^{d+|s|}}\left\|D^sK_{x,h}-f\right\|_{L_2(Q)}\\&\leq l_n^{-d-|s|}\left\|D^sK_{x,h}-f\right\|_{L_2(Q)}<\frac{\eta}{2}.\end{aligned}\tag{67}$$

Hence applying (66) and (67) to (65) gives

$$\left\|\frac{1}{h^{d+|s|}}D^sK_{x,h}-\frac{1}{h_i^{d+|s|}}f_j\right\|_{L_2(Q)}<\eta.$$

For the second case when $h>\left(\frac{\eta}{2\|D^sK\|_\infty}\right)^{-1/(d+|s|)}$, $\left\|\frac{1}{h^{d+|s|}}D^sK_{x,h}\right\|_{L_2(Q)}\leq\left\|\frac{1}{h^{d+|s|}}D^sK_{x,h}\right\|_\infty<\frac{\eta}{2}$ holds, and hence

$$\left\|\frac{1}{h^{d+|s|}}D^sK_{x,h}-f_0\right\|_{L_2(Q)}\leq\left\|\frac{1}{h^{d+|s|}}D^sK_{x,h}\right\|_{L_2(Q)}+\left\|f_0\right\|_{L_2(Q)}<\eta.$$

Therefore, (64) is shown. Hence combined with Assumption 7 gives that for every probability measure Q on $\mathbb{R}^d$ and for every $\eta\in(0,h^{-d}\left\|D^sK\right\|_\infty)$, the covering number $\mathcal{N}(\tilde{\mathcal{F}}_{K,[l_n,\infty)},L_2(Q),\eta)$ is upper bounded as

$$\begin{aligned}&\sup_Q\mathcal{N}(\tilde{\mathcal{F}}_{K,[l_n,\infty)},L_2(Q),\eta)\\&\leq\mathcal{N}\left(\left[l_n,\left(\frac{\eta}{2\left\|D^sK\right\|_\infty}\right)^{-1/(d+|s|)}\right],|\cdot|,\frac{l_n^{d+|s|+1}\eta}{2(d+|s|)\left\|D^sK\right\|_\infty}\right)\sup_Q\mathcal{N}\left(\mathcal{F}_{K,[l_n,\infty)},L_2(Q),\frac{l_n^{d+|s|}\eta}{2}\right)+1\\&\leq\frac{2(d+|s|)\left\|D^sK\right\|_\infty}{l_n^{d+|s|+1}\eta}\left(\frac{2\left\|D^sK\right\|_\infty}{\eta}\right)^{1/(d+|s|)}\left(\frac{2A\left\|D^sK\right\|_\infty}{l_n^{d+|s|}\eta}\right)^\nu+1\\&\leq\left(\frac{2A(d+|s|)\left\|D^sK\right\|_\infty}{l_n^{d+|s|}\eta}\right)^{\nu+2}.\end{aligned}\tag{68}$$

Also, Lemma 19 implies that under Assumption 6, for any $\epsilon\in(0,d_{\rm vol})$ (and ϵ can be 0 if $d_{\rm vol}=0$ or under Assumption 1),

$$\mathbb{E}_P\left[\left(\frac{1}{h^{d+|s|}}D^sK_{x,h}\right)^2\right]\leq C_{s,P,K,\epsilon}l_n^{-2d-2|s|+d_{\rm vol}-\epsilon}.\tag{69}$$

Hence from (63), (68), and (69), applying Theorem 30 to (62) gives that $\sup_{h\geq l_n,x\in\mathbb{X}}|D^s\hat{p}_h(x)-D^sp_h(x)|$ is upper bounded with probability at least $1-\delta$ as

$$\begin{aligned}
&\sup_{h\geq l_n, x\in\mathbb{X}} \left|D^s\hat{p}_h(x)-D^s p_h(x)\right| \\
&\leq C\left(\frac{2(\nu+2)\left\|D^sK\right\|_\infty \log\left(\frac{2A(d+|s|)\|D^sK\|_\infty}{\sqrt{C_{s,P,K,\epsilon}}l_n^{(d_{\rm vol}-\epsilon)/2}}\right)}{nl_n^{d+|s|}}+\sqrt{\frac{2(\nu+2)C_{s,P,K,\epsilon}\log\left(\frac{2A(d+|s|)\|D^sK\|_\infty}{\sqrt{C_{s,P,K,\epsilon}}l_n^{(d_{\rm vol}-\epsilon)/2}}\right)}{nl_n^{2d+2|s|-d_{\rm vol}+\epsilon}}}\right. \\
&\qquad\left.+\sqrt{\frac{C_{s,P,K,\epsilon}\log(\frac{1}{\delta})}{nl_n^{2d+2|s|-d_{\rm vol}+\epsilon}}}+\frac{\|D_sK\|_\infty\log(\frac{1}{\delta})}{nl_n^{d+|s|}}\right) \\
&\leq C_{A,\|D^sK\|_\infty,d,\nu,d_{\rm vol},C_{s,P,K,\epsilon}}\left(\frac{\left(\log\left(\frac{1}{l_n}\right)\right)_+}{nl_n^{d+|s|}}+\sqrt{\frac{\left(\log\left(\frac{1}{l_n}\right)\right)_+}{nl_n^{2d+2|s|-d_{\rm vol}+\epsilon}}}+\sqrt{\frac{\log\left(\frac{2}{\delta}\right)}{nl_n^{2d+2|s|-d_{\rm vol}+\epsilon}}}+\frac{\log\left(\frac{2}{\delta}\right)}{nl_n^{d+|s|}}\right),
\end{aligned}$$

where $C_{A,\|D^sK\|_\infty,d,\nu,d_{\rm vol},C_{s,P,K,\epsilon},\epsilon}$ depends only on A, $\|D^sK\|_\infty$, d, ν, $d_{\rm vol}$, $C_{s,P,K,\epsilon}$, ϵ.

□

For showing Corollary 21, we proceed similarly to the proof of Corollary 13, where we plug in D^sK in the place of K.

Corollary 21. *Let P be a distribution and K be a kernel function satisfying Assumption 5, 6, and 7. Suppose*

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

for fixed $\epsilon\in(0,d_{\rm vol})$. Then, with probability at least $1-\delta$,

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

where C' is a constant depending only on A, $\|D^sK\|_\infty$, d, ν, $d_{\rm vol}$, $C_{s,P,K,\epsilon}$, ϵ. Further, if $d_{\rm vol}=0$ or under Assumption 1, ϵ can be 0.

Proof of Corollary 21. From (19) in Theorem 20, $\sup_{h\geq l_n,x\in\mathbb{X}}\left|D^s\hat{p}_h(x)-D^sp_h(x)\right|$ is upper bounded with probability at least $1-\delta$ as

$$\begin{aligned}
&\sup_{h\geq l_n, x\in\mathbb{X}} \left|D^s\hat{p}_h(x)-D^s p_h(x)\right| \\
&\leq C_{A,\|D^sK\|_\infty,d,\nu,d_{\rm vol},C_{s,P,K,\epsilon}}\left(\frac{\left(\log\left(\frac{1}{l_n}\right)\right)_+}{nl_n^{d+|s|}}+\sqrt{\frac{\left(\log\left(\frac{1}{l_n}\right)\right)_+}{nl_n^{2d+2|s|-d_{\rm vol}+\epsilon}}}+\sqrt{\frac{\log\left(\frac{2}{\delta}\right)}{nl_n^{2d+2|s|-d_{\rm vol}+\epsilon}}}+\frac{\log\left(\frac{2}{\delta}\right)}{nl_n^{d+|s|}}\right) \\
&= C_{A,\|D^sK\|_\infty,d,\nu,d_{\rm vol},C_{s,P,K,\epsilon}}\left(\sqrt{\frac{\left(\log\left(\frac{1}{l_n}\right)\right)_+}{nl_n^{2d+2|s|-d_{\rm vol}+\epsilon}}}\left(\sqrt{\frac{\left(\log\left(\frac{1}{l_n}\right)\right)_+}{nl_n^{d_{\rm vol}-\epsilon}}}+1\right)+\sqrt{\frac{\log\left(\frac{2}{\delta}\right)}{nl_n^{2d+2|s|-d_{\rm vol}+\epsilon}}}\left(\sqrt{\frac{\log\left(\frac{2}{\delta}\right)}{nl_n^{d_{\rm vol}-\epsilon}}}+1\right)\right).
\end{aligned}$$

Then from $\limsup_n \frac{\left(\log\left(\frac{1}{l_n}\right)\right)_+ +\log\left(\frac{2}{\delta}\right)}{nl_n^{d_{\rm vol}-\epsilon}}<\infty$, there exists some constant C' with $\left(\log\left(\frac{1}{l_n}\right)\right)_+ + \log\left(\frac{2}{\delta}\right)\leq C'nl_n^{d_{\rm vol}+\epsilon}$.

And hence $\sup_{h\geq l_n,x\in\mathbb{X}}|D^s\hat{p}_h(x)-D^sp_h(x)|$ is upper bounded with probability $1-\delta$ as

$$
\begin{aligned}
&\sup_{h\geq l_n,x\in\mathbb{X}}|D^s\hat{p}_h(x)-D^sp_h(x)| \\
&\leq C_{A,\|D^sK\|_\infty,d,\nu,d_{\mathrm{vol}},C_{s,P,K,\epsilon}}\left(\sqrt{\frac{\left(\log\left(\frac{1}{l_n}\right)\right)_+}{nl_n^{2d+2|s|-d_{\mathrm{vol}}+\epsilon}}}\left(\sqrt{C'}+1\right)+\sqrt{\frac{\log\left(\frac{1}{\delta}\right)}{nl_n^{2d+2|s|-d_{\mathrm{vol}}+\epsilon}}}\left(\sqrt{C'}+1\right)\right) \\
&\leq C'_{A,\|D^sK\|_\infty,d,\nu,d_{\mathrm{vol}},C_{s,P,K,\epsilon}}\sqrt{\frac{\left(\log\left(\frac{1}{l_n}\right)\right)_++\log\left(\frac{1}{\delta}\right)}{nl_n^{2d+2|s|-d_{\mathrm{vol}}+\epsilon}}},
\end{aligned}
$$

where $C'_{A,\|D^sK\|_\infty,d,\nu,d_{\mathrm{vol}},C_{s,P,K,\epsilon}}$ depending only on A, $\|D^sK\|_\infty$, d, ν, d_{vol}, $C_{s,P,K,\epsilon}$, ϵ.

□

For proving Lemma 22, we proceed similarly to the proof of Lemma 14, where we plug in D^sK in the place of K.

Lemma 22. *Suppose there exists $R>0$ with $\mathbb{X}\subset\mathbb{B}_{\mathbb{R}^d}(0,R)$. Also, suppose that D^sK is M_K-Lipschitz, i.e.*

$$\|D^sK(x)-D^sK(y)\|_2\leq M_K\|x-y\|_2.$$

Then for all $\eta\in(0,\|D^sK\|_\infty)$, the supremum of the η-covering number $\mathcal{N}(\mathcal{F}_{K,h}^s,L_2(Q),\eta)$ over all measure Q is upper bounded as

$$\sup_Q\mathcal{N}(\mathcal{F}_{K,h}^s,L_2(Q),\eta)\leq\left(\frac{2RM_Kh^{-1}+\|D^sK\|_\infty}{\eta}\right)^d.$$

Proof of Lemma 22. For fixed $\eta>0$, let $x_1,\ldots,x_M$ be the maximal η-covering of $\mathbb{B}_{\mathbb{R}^d}(0,R)$, with $M=\mathcal{M}\left(\mathbb{B}_{\mathbb{R}^d}(0,R),\|\cdot\|_2,\eta\right)$ being the packing number of $\mathbb{B}_{\mathbb{R}^d}(0,R)$. Then $\mathbb{B}_{\mathbb{R}^d}(x_i,\eta)$ and $\mathbb{B}_{\mathbb{R}^d}(x_j,\eta)$ do not intersect for any i,j and $\bigcup_{i=1}^M\mathbb{B}_{\mathbb{R}^d}(x_i,\eta)\subset\mathbb{B}_{\mathbb{R}^d}(x_i,R+\eta)$, and hence

$$\sum_{i=1}^M\lambda_d\left(\mathbb{B}_{\mathbb{R}^d}(x_i,\eta)\right)\leq\lambda_d\left(\mathbb{B}_{\mathbb{R}^d}(x_i,R+\eta)\right). \tag{70}$$

Then $\lambda_d\left(\mathbb{B}_{\mathbb{R}^d}(x,r)\right)=r^d\lambda_d\left(\mathbb{B}_{\mathbb{R}^d}(0,1)\right)$ gives the upper bound on $\mathcal{M}(\mathbb{B}_{\mathbb{R}^d}(0,R),\|\cdot\|_2,\eta)$ as

$$\mathcal{M}\left(\mathbb{B}_{\mathbb{R}^d}(0,R),\|\cdot\|_2,\eta\right)\leq\left(1+\frac{R}{\eta}\right)^d.$$

Then $\mathbb{X}\subset\mathbb{B}_{\mathbb{R}^d}(0,R)$ and the relationship between covering number and packing number gives the upper bound on the covering number $\mathcal{N}\left(\mathbb{X},\|\cdot\|_2,\eta\right)$ as

$$\mathcal{N}\left(\mathbb{X},\|\cdot\|_2,\eta\right)\leq\mathcal{N}\left(\mathbb{B}_{\mathbb{R}^d}(0,R),\|\cdot\|_2,\eta\right)\leq\mathcal{M}\left(\mathbb{B}_{\mathbb{R}^d}(0,R),\|\cdot\|_2,\frac{\eta}{2}\right)\leq\left(1+\frac{2R}{\eta}\right)^d. \tag{71}$$

Now, note that for all $x,y\in\mathbb{X}$ and for all $z\in\mathbb{R}^d$, $|D^sK_{x,h}(z)-D^sK_{y,h}(z)|$ is upper bounded as

$$|D^sK_{x,h}(z)-D^sK_{y,h}(z)|=\left|D^sK\left(\frac{x-z}{h}\right)-D^sK\left(\frac{y-z}{h}\right)\right|\leq\frac{M_K}{h}\|(x-z)-(y-z)\|_2=\frac{M_K}{h}\|x-y\|_2.$$

Hence for any measure Q on $\mathbb{R}^d$, $\|D^sK_{x,h}-D^sK_{y,h}\|_{L_2(Q)}$ is upper bounded as

$$\|D^sK_{x,h}-D^sK_{y,h}\|_{L_2(Q)}=\sqrt{\int(D^sK_{x,h}(z)-D^sK_{y,h}(z))^2dQ(z)}\leq\frac{M_K}{h}\|x-y\|_2.$$

Hence applying this to (71) implies that for all $\eta > 0$, the supremum of the covering number $\mathcal{N}(\mathcal{F}_{K,h}, L_2(Q), \eta)$ over all measure Q is upper bounded as

$$\sup_Q \mathcal{N}(\mathcal{F}^s_{K,h}, L_2(Q), \eta) \leq \mathcal{N}\left(\mathbb{X}, \|\cdot\|_2, \frac{h\eta}{M_K}\right) \leq \left(1 + \frac{2RM_K}{h\eta}\right)^d.$$

Hence for all $\eta \in (0, \|D^s K\|_\infty)$,

$$\sup_Q \mathcal{N}(\mathcal{F}^s_{K,h}, L_2(Q), \eta) \leq \left(\frac{2RM_K h^{-1} + \|D^s K\|_\infty}{\eta}\right)^d.$$

□

For Corollary 23, we proceed similarly to the proof of Corollary 15, where we plug in $D^s K$ in the place of K.

Corollary 23. *Suppose there exists $R > 0$ with $\mathrm{supp}(P) = \mathbb{X} \subset \mathbb{B}_{\mathbb{R}^d}(0, R)$. Let K be a kernel function with M_K-Lipschitz continuous derivative satisfying Assumption 6. If*

$$\limsup_n \frac{(\log(1/h_n))_+ + \log(2/\delta)}{n h_n^{d_{\mathrm{vol}} - \epsilon}} < \infty,$$

for fixed $\epsilon \in (0, d_{\mathrm{vol}})$. Then, with probability at least $1 - \delta$,

$$\sup_{x \in \mathbb{X}} |D^s \hat{p}_h(x) - D^s p_h(x)| \leq C'' \sqrt{\frac{(\log(\frac{1}{h_n}))_+ + \log(\frac{2}{\delta})}{n h_n^{2d + 2|s| - d_{\mathrm{vol}+\epsilon}}}},$$

where C'' is a constant depending only on A, $\|D^s K\|_\infty$, d, M_k, d_{vol}, $C_{s,P,K,\epsilon}$, ϵ. Further, if $d_{\mathrm{vol}} = 0$ or under Assumption 1, ϵ can be 0.

Proof of Corollary 23. For $x \in \mathbb{X}$, let $D^s K_{x,h} : \mathbb{R}^d \to \mathbb{R}$ be $D^s K_{x,h}(\cdot) = D^s K\left(\frac{x - \cdot}{h}\right)$, and let $\tilde{\mathcal{F}}^s_{K,h} := \left\{\frac{1}{h^{d+|s|}} D^s K_{x,h} : x \in \mathbb{X}\right\}$ be a class of normalized kernel functions centered on $\mathbb{X}$ and bandwidth h. Note that $D^s \hat{p}_h(x) - D^s p_h(x)$ can be expanded as

$$\begin{aligned} D^s \hat{p}_h(x) - D^s p_h(x) &= \frac{1}{nh^{d+|s|}} \sum_{i=1}^n D^s K\left(\frac{x - X_i}{h}\right) - \mathbb{E}_P\left[\frac{1}{h^{d+|s|}} D^s K\left(\frac{x - X_i}{h}\right)\right] \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{h^{d+|s|}} D^s K_{x,h}(X_i) - \mathbb{E}_P\left[\frac{1}{h^{d+|s|}} D^s K_{x,h}\right]. \end{aligned}$$

Hence $\sup_{x \in \mathbb{X}} |D^s \hat{p}_h(x) - D^s p_h(x)|$ can be expanded as

$$\sup_{x \in \mathbb{X}} |D^s \hat{p}_h(x) - D^s p_h(x)| = \sup_{f \in \tilde{\mathcal{F}}_{K,h}} \left|\frac{1}{n} \sum_{i=1}^n f(X_i) - \mathbb{E}_P[f(X)]\right|. \tag{72}$$

Now, it is immediate to check that

$$\|f\|_\infty \leq h^{-d-|s|} \|D^s K\|_\infty. \tag{73}$$

Also, since $\tilde{\mathcal{F}}^s_{K,h} = h^{-d-|s|} \mathcal{F}^s_{K,h}$, VC dimension is uniformly bounded as Lemma 22 gives that for every probability measure Q on $\mathbb{R}^d$ and for every $\eta \in (0, h^{-d-|s|} \|D^s K\|_\infty)$, the covering number $\mathcal{N}(\tilde{\mathcal{F}}^s_{K,h}, L_2(Q), \eta)$ is upper bounded as

$$\begin{aligned} \sup_Q \mathcal{N}(\tilde{\mathcal{F}}^s_{K,h}, L_2(Q), \eta) &= \sup_Q \mathcal{N}(\mathcal{F}_{K,h}, L_2(Q), h^{d+|s|}\eta) \\ &\leq \left(\frac{2RM_K h^{-1} + \|D^s K\|_\infty}{h^{d+|s|}\eta}\right)^d \\ &\leq \left(\frac{2RM_K \|D^s K\|_\infty}{h^{d+|s|+1}\eta}\right)^d. \end{aligned} \tag{74}$$

Also, Lemma 19 implies that under Assumption 3, for any $\epsilon \in (0, d_{\mathrm{vol}})$ (and ϵ can be 0 if $d_{\mathrm{vol}} = 0$ or under Assumption 1),

$$\mathbb{E}_P\left[\left(\frac{1}{h^{d+|s|}}D^s K_{x,h}\right)^2\right] \leq C_{s,P,K,\epsilon} h^{-2d-2|s|+d_{\mathrm{vol}}-\epsilon}. \tag{75}$$

Hence from (73), (74), and (75), applying Theorem 30 to (72) gives that $\sup_{x\in\mathbb{X}} |D^s\hat{p}_h(x) - D^s p_h(x)|$ is upper bounded with probability at least $1-\delta$ as

$$\begin{aligned}
&\sup_{x\in\mathbb{X}} |\hat{p}_h(x) - p_h(x)| \\
&\leq C\left(\frac{2d\left\|D^s K\right\|_\infty \log\left(\frac{2RM_K\|D^s K\|_\infty}{\sqrt{C_{s,P,K,\epsilon}}h^{1+(d_{\mathrm{vol}}-\epsilon)/2}}\right)}{nh^{d+|s|}} + \sqrt{\frac{2dC_{s,P,K,\epsilon}\log\left(\frac{2RM_K\|D^s K\|_\infty}{\sqrt{C_{s,P,K,\epsilon}}h^{1+(d_{\mathrm{vol}}-\epsilon)/2}}\right)}{nh^{2d+2|s|-d_{\mathrm{vol}}+\epsilon}}}\right. \\
&\qquad \left. + \sqrt{\frac{C_{s,P,K,\epsilon}\log(\frac{1}{\delta})}{nh^{2d+2|s|-d_{\mathrm{vol}}+\epsilon}}} + \frac{\|D^s K\|_\infty \log(\frac{1}{\delta})}{nh^d}\right) \\
&\leq C_{R,M_K,\|D^s K\|_\infty,d,\nu,d_{\mathrm{vol}},C_{s,P,K,\epsilon},\epsilon}\left(\frac{\left(\log\left(\frac{1}{h}\right)\right)_+}{nh^d} + \sqrt{\frac{\left(\log\left(\frac{1}{h}\right)\right)_+}{nh^{2d+2|s|-d_{\mathrm{vol}}+\epsilon}}} + \sqrt{\frac{\log\left(\frac{2}{\delta}\right)}{nh^{2d+2|s|-d_{\mathrm{vol}}+\epsilon}}} + \frac{\log\left(\frac{2}{\delta}\right)}{nh^d}\right),
\end{aligned}$$

where $C_{R,M_K,\|D^s K\|_\infty,d,\nu,d_{\mathrm{vol}},C_{s,P,K,\epsilon},\epsilon}$ depends only on R, M_K, $\|D^s K\|_\infty$, d, ν, d_{vol}, $C_{s,P,K,\epsilon}$, ϵ.

$\square$