Consistency of Persistent Homology

김지수 (Jisu KIM)

통계이론세미나 - 위상구조의 통계적 추정, 2023 가을학기

We first recall the consistency:

Suppose we obtain a sample $X_1, \ldots, X_n \sim P$. Let $\theta(P)$ be a parameter, which is some function of P. Let $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ denote an estimator for $\hat{\theta}$, which is a function of a sample. Consistency is about, whether the estimator $\hat{\theta}$ converge in probability to θ, i.e. $\hat{\theta} \xrightarrow{P} \theta$. More precisely, can we find some function $f(n)$ of the sample size n such that $d(\hat{\theta}, \theta) = O_P(f(n))$? This is analogous to the Law of Large Number.

Let $\mathbb{X} \subset \mathbb{R}^d$ be the target geometric structure, and P be a distribution on $\mathbb{R}^d$ with $\text{supp}(P) = \mathbb{X}$. Let $X_1, \ldots, X_n$ be i.i.d. samples from P and $\mathcal{X} = \{X_1, \ldots, X_n\}$. For the consistency of persistent homology, the distance is the bottleneck distance d_B, and $\theta(P)$ and $\hat{\theta}(\mathcal{X})$ should be appropriate persistent homologies of P and $\mathcal{X}$, respectively. We consider two cases:

1. Persistent homologies from Čech complexes and Vietoris-Rips complexes. Let $\mathcal{PC}_{\mathbb{R}^d}(\mathbb{X})$ and $\mathcal{PC}_{\mathbb{R}^d}(\mathcal{X})$ be the k-th dimensional persistent homologies induced from Čech complexes $\left\{H_k \check{\text{C}}\text{ech}_{\mathbb{R}^d}(\mathbb{X}, r)\right\}_{r\in\mathbb{R}}$ and $\left\{H_k \check{\text{C}}\text{ech}_{\mathbb{R}^d}(\mathcal{X}, r)\right\}_{r\in\mathbb{R}}$, respectively. Similarly, let $\mathcal{PR}(\mathbb{X})$ and $\mathcal{PR}(\mathcal{X})$ be the k-th dimensional persistent homologies induced from Vietoris-Rips complexes $\{H_k \text{Rips}(\mathbb{X}, r)\}_{r\in\mathbb{R}}$ and $\{H_k \text{Rips}(\mathcal{X}, r)\}_{r\in\mathbb{R}}$, respectively. We would like to know $d_B(\mathcal{PC}_{\mathbb{R}^d}(\mathbb{X}), \mathcal{PC}_{\mathbb{R}^d}(\mathcal{X})) = O_P(f(n))$ and $d_B(\mathcal{PR}(\mathbb{X}), \mathcal{PR}(\mathcal{X})) = O_P(f(n))$.

2. Persistent homologies from the superlevel filtration of kernel density estimator (KDE). Consider the superlevel filtration $\left\{\hat{p}_h^{-1}[\lambda, \infty)\right\}_{\lambda\in\mathbb{R}}$, then the persistent homology consists of morphisms $\imath_k^{\lambda_1,\lambda_2} : H_k \hat{p}_h^{-1}[\lambda_1, \infty) \to H_k \hat{p}_h^{-1}[\lambda_2, \infty)$ for $\lambda_1 \geq \lambda_2$ induced from inclusions $\hat{p}_h^{-1}[\lambda_1, \infty) \subset \hat{p}_h^{-1}[\lambda_2, \infty)$. Let $\mathcal{P}(\hat{p}_h), \mathcal{P}(p_h), \mathcal{P}(p)$ be the k-th dimensional persistent homology induced from $\hat{p}_h, p_h, p$, respectively, where $p_h = \mathbb{E}[\hat{p}_h]$ and p is the density of P. We would like to know either $d_B(\mathcal{P}(p_h), \mathcal{P}(\hat{p}_h)) = O_P(f(n))$ or $d_B(\mathcal{P}(p), \mathcal{P}(\hat{p}_h)) = O_P(f(n))$.

Consistency of persistent homologies from Čech complexes and Vietoris-Rips complexes

Assume $\mathbb{X}$ is compact. Recall the stability theorem for Čech complexes and Vietoris-Rips complexes:

Corollary. *For a compact set $\mathbb{X} \subset \mathbb{R}^d$ and $\mathcal{X} \subset \mathbb{X}$,*

$$d_B(\mathcal{PC}_{\mathbb{R}^d}(\mathbb{X}), \mathcal{PC}_{\mathbb{R}^d}(\mathcal{X})) \leq d_H(\mathbb{X}, \mathcal{X}).$$
$$d_B(\mathcal{PR}(\mathbb{X}), \mathcal{PR}(\mathcal{X})) \leq d_H(\mathbb{X}, \mathcal{X}).$$

Hence bounding the bottleneck distance between persistent homologies from Čech complexes and Vietoris-Rips complexes can be sufficed by bounding Hausdorff distance.

For a distribution P, we assume (a, b) assumption:

Definition. P satisfies (a, b) assumption if there exists $r_0 > 0$ such that for all $x \in \text{supp}(P)$ and for all $r < r_0$,

$$P(\mathcal{B}(x, r)) \geq ar^b.$$

Recall that under (a, b) assumption, we have probabilistic bound on the Hausdorff distance between $\mathbb{X}$ and $\mathcal{X}$:

Proposition ([8, Proposition 7.2][3, Theorem 2])**.** *Let P be a distribution on $\mathbb{R}^d$ with $\text{supp}(P) = \mathbb{X}$, and assume P satisfies (a, b) assumption with $a, b > 0$. Let $X_1, \ldots, X_n$ be i.i.d. samples from P, and let $\mathcal{X} = \{X_1, \ldots, X_n\}$. Then there exists $\epsilon_0 > 0$ such that for all $\epsilon < \epsilon_0$,*

$$P(d_H(\mathbb{X}, \mathcal{X}) < \epsilon) \geq 1 - a^{-1}\epsilon^{-b} \exp(-na\epsilon^b). \tag{1}$$

This directly implies that with probability $1-\delta$, with large enough n,

$$d_H(\mathbb{X},\mathcal{X}) < C\left(\frac{\log n}{n}\right)^{1/b},$$

and hence

$$d_H(\mathbb{X},\mathcal{X}) = O_P\left(\left(\frac{\log n}{n}\right)^{1/b}\right).$$

Then this implies both that

$$d_B(\mathcal{PC}_{\mathbb{R}^d}(\mathbb{X}),\mathcal{PC}_{\mathbb{R}^d}(\mathcal{X})) = O_P\left(\left(\frac{\log n}{n}\right)^{1/b}\right),$$
$$d_B(\mathcal{PR}(\mathbb{X}),\mathcal{PR}(\mathcal{X})) = O_P\left(\left(\frac{\log n}{n}\right)^{1/b}\right).$$

(1) not only gives the probabilistic bound as above, but this also gives the bound on the expectation as well. Roughly speaking, this is deduced from

$$\mathbb{E}\left[d_H(\mathbb{X},\mathcal{X})\right] = \int_0^\infty P\left(d_H(\mathbb{X},\mathcal{X}) > \epsilon\right) d\epsilon.$$

Theorem ([3, Theorem 4])**.** *Let P be a distribution on $\mathbb{R}^d$ with $\mathrm{supp}(P)=\mathbb{X}$, and assume P satisfies (a,b) assumption with $a,b>0$. Let $X_1,\ldots,X_n$ be i.i.d. samples from P, and let $\mathcal{X}=\{X_1,\ldots,X_n\}$. Then,*

$$\mathbb{E}\left[d_H(\mathbb{X},\mathcal{X})\right] \leq C\left(\frac{\log n}{n}\right)^{1/b},$$

where C only depends on a and b. And correspondingly,

$$\mathbb{E}\left[d_B(\mathcal{PC}_{\mathbb{R}^d}(\mathbb{X}),\mathcal{PC}_{\mathbb{R}^d}(\mathcal{X}))\right] \leq C\left(\frac{\log n}{n}\right)^{1/b},$$
$$\mathbb{E}\left[d_B(\mathcal{PR}(\mathbb{X}),\mathcal{PR}(\mathcal{X}))\right] \leq C\left(\frac{\log n}{n}\right)^{1/b}.$$

The convergence rate $\left(\frac{\log n}{n}\right)^{1/b}$ of Čech complexes and Vietoris-Rips complexes is in fact minimax up to a logarithmic term.

Theorem ([3, Theorem 4])**.** *Let $\mathcal{P}$ be a set of distributions P with $\mathrm{supp}(P)$ being compact and satisfying (a,b) assumption with fixed $a,b>0$. Then for any estimator $\hat{\mathrm{dgm}}_n$ (that is, a function of data $X_1,\ldots,X_n$),*

$$\sup_{P\in\mathcal{P}}\mathbb{E}_P\left[d_B(\mathcal{PC}_{\mathbb{R}^d}(\mathbb{X}),\mathcal{PC}_{\mathbb{R}^d}(\mathcal{X}))\right] \geq Cn^{-1/b},$$
$$\sup_{P\in\mathcal{P}}\mathbb{E}_P\left[d_B(\mathcal{PR}(\mathbb{X}),\mathcal{PR}(\mathcal{X}))\right] \geq Cn^{-1/b}.$$

Consistency of persistent homologies from kernel density estimators

This section is based on the results from [5, 7].

Recall that a kernel function $K:\mathbb{R}^d\to\mathbb{R}$ is a function satisfying $\int K(x)dx=1$.Given a kernel K and a bandwidth h, the kernel density estimator (KDE) is defined to be

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

Then the average KDE $p_h:\mathbb{R}^d\to\mathbb{R}$ is

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

Recall the stability theorem for the persistent homology induced from functions:

Corollary. *For two functions $f, g : \mathbb{X} \to \mathbb{R}$, if $\mathcal{P}(f)$ and $\mathcal{P}(g)$ are q-tame, then*

$$d_B(\mathcal{P}(f), \mathcal{P}(g)) \leq \|f - g\|_\infty .$$

Hence bounding the bottleneck distance between persistent homologies of $\hat{p}_h$ and p_h (or p) can be sufficed by bounding their infinity distances $\|\hat{p}_h - p_h\|_\infty$ or $\|\hat{p}_h - p\|_\infty$.

Suppose P is a distribution on $\mathbb{R}^d$ with $\mathrm{supp}(P) = \mathbb{X}$. Suppose $X_1, \ldots, X_n$ are i.i.d. samples from P. We will get the concentration inequalities for the kernel density estimator in the supremum norm that hold uniformly over the selection of the bandwidth, i.e., the concentration inequalities for

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

We rewrite this as a supremum over a function class. For $x \in \mathbb{X}$ and $h \geq l_n$, let $K_{x,h} : \mathbb{R}^d \to \mathbb{R}$ be $K_{x,h}(\cdot) = K\left(\frac{x - \cdot}{h}\right)$. Let $\mathcal{F}_{K,[l_n,\infty)}$ be defined as

$$\mathcal{F}_{K,[l_n,\infty)} := \{K_{x,h} : x \in \mathbb{X}, h \geq l_n\} ,$$

i.e., $\mathcal{F}_{K,[l_n,\infty)}$ is a class of unnormalized kernel functions centered on each element in $\mathbb{X}$ and bandwidth greater than or equal to l_n. And similarly define the normalized version

$$\tilde{\mathcal{F}}_{K,[l_n,\infty)} := \left\{(1/h^d)K_{x,h} : x \in \mathbb{X},\, h \geq l_n\right\} .$$

Then $\sup_{h \geq l_n, x \in \mathbb{X}} |\hat{p}_h(x) - p_h(x)|$ can be rewritten as a supremum of an empirical process indexed by $\tilde{\mathcal{F}}$, that is,

$$\begin{aligned} &\sup_{h \geq l_n, x \in \mathbb{X}} |\hat{p}_h(x) - p_h(x)| \\ &= \sup_{f \in \tilde{\mathcal{F}}_{K,[l_n,\infty)}} \left| \frac{1}{n} \sum_{i=1}^n f(X_i) - \mathbb{E}[f(X)] \right| . \end{aligned} \tag{2}$$

Talagrand's inequality and a VC type bound

One of the most important developments in concentration of measure is Talagrand's inequality which can be thought of as a uniform version of Bernstein's inequality. We combine Talagrand's inequality and a VC type bound to bound (2), following the approach of [10, Theorem 3.1]. The following version of Talagrand's inequality is from [2, Theorem 2.3] and simplified in [11, Theorem 7.5].

Proposition ([2, Theorem 2.3][11, Theorem 7.5, Theorem A.9.1][7, Proposition 8])**.** *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 separable in $L_\infty(\mathbb{R}^d)$. Suppose all functions $f \in \mathcal{F}$ are P-measurable, and there exists $B, \sigma > 0$ such that $\mathbb{E}_P f = 0$, $\mathbb{E}_P f^2 \leq \sigma^2$, and $\|f\|_\infty \leq B$, for all $f \in \mathcal{F}$. Let*

$$Z := \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(X_i) \right| ,$$

Then for any $\delta > 0$,

$$P\left(Z \geq \mathbb{E}_P[Z] + \sqrt{\left(\frac{2}{n} \log \frac{1}{\delta}\right)(\sigma^2 + 2B\mathbb{E}_P[Z])} + \frac{2B \log \frac{1}{\delta}}{3n}\right) \leq \delta.$$

By applying Talagrand's inequality to (2), $\sup_{h \geq l_n, x \in \mathbb{X}} |\hat{p}_h(x) - p_h(x)|$ can be upper bounded in terms of n, $\|K_{x,h}\|_\infty$, $\mathbb{E}_P[K_{x,h}^2]$, and

$$\mathbb{E}_P\left[\sup_{f \in \tilde{\mathcal{F}}_{K,[l_n,\infty)}} \left| \frac{1}{n} \sum_{i=1}^n f(X_i) - \mathbb{E}[f(X)] \right|\right] . \tag{3}$$

To bound the last term, we use the uniformly bounded VC class assumption on the kernel. The following bound on the expected suprema of empirical processes of VC classes of functions is from [4, Proposition 2.1].

Proposition. *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 the covering number $\mathcal{N}(\mathcal{F}, L_2(Q), \epsilon)$ satisfies*

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

for every probability measure Q on $\mathbb{R}^d$ and for every $\epsilon \in (0, B)$. Let $\sigma > 0$ be a positive number such that $\mathbb{E}_P f^2 \leq \sigma^2$ for all $f \in \mathcal{F}$. Then there exists a universal constant C not depending on any parameters such that

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

By applying above Propositions to $\tilde{\mathcal{F}}_{K,[l_n,\infty)}$, it can be shown that the upper bound of

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

can be written as a function of $\|K_{x,h}\|_\infty$ and $\mathbb{E}_P[K_{x,h}^2]$. When the lower bound on the interval l_n is not too small, the terms relating to $\mathbb{E}_P[K_{x,h}^2]$ are more dominant. Hence, to get a good upper bound with respect to both n and h, it is important to get a tight upper bound for $\mathbb{E}_P[K_{x,h}^2]$.

Lemma ([10, proof of Proposition A.5]). *When P has a Lebesgue density p and $\|K\|_2, \|p\|_\infty < \infty$,*

$$\mathbb{E}_P \left[K_{x,h}^2 \right] \leq \|K\|_2^2 \|p\|_\infty h^d.$$

Proof.

$$\begin{aligned} \mathbb{E}_P \left[K_{x,h}^2 \right] &= \int_{-\infty}^{\infty} K^2 \left(\frac{x-y}{h} \right) p(y) dy = h^d \int_{-\infty}^{\infty} K^2(t) p(x - ht) dt \\ &\leq \|p\|_\infty h^d \int_{-\infty}^{\infty} K^2(t) dt = \|K\|_2^2 \|p\|_\infty h^d. \end{aligned}$$

□

Assumptions

Assumption. *We first assume that the kernel function $K : \mathbb{R}^d \to \mathbb{R}$ satisfies $\|K\|_2, \|K\|_\infty < \infty$ and $\operatorname{supp}(K)$ is compact. Also, we assume that the density p exists for the distribution P, i.e., $p : \mathbb{R}^d \to \mathbb{R}$ is a function satisfying $P(A) = \int_A p(x) dx$ for all Borel sets $A \in \mathcal{B}(\mathbb{R}^d)$, and $\|p\|_\infty < \infty$.*

And we assume uniformly bounded VC dimension:
To apply the VC type bound, the function class,

$$\mathcal{F}_{K,[l_n,\infty)} := \{ K_{x,h} : x \in \mathbb{X}, h \geq l_n \},$$

should be not too complex. One common approach is to assume that $\mathcal{F}_{K,[l_n,\infty)}$ is a uniformly bounded VC-class, which is defined imposing appropriate bounds on the metric entropy of the function class [5, 10, 7]:

Assumption. *We assume $\mathcal{F}_{K,[l_n,\infty)} := \{K_{x,h} : x \in \mathbb{X}, h \geq l_n\}$ is a uniformly bounded VC-class with dimension ν, i.e. there exists positive numbers A and v such that, for every probability measure Q on $\mathbb{R}^d$ and for every $\epsilon \in (0, \|K\|_\infty)$, the covering numbers $\mathcal{N}(\mathcal{F}_{K,[l_n,\infty)}, L_2(Q), \epsilon)$ satisfy*

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

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

Since $[l_n, \infty) \subset (0, \infty)$, one sufficient condition is to impose uniformly bounded VC class condition on a larger function class,

$$\mathcal{F}_{K,(0,\infty)} = \{ K_{x,h} : x \in \mathbb{X}, h > 0 \}.$$

This is implied by condition (K) in [6] or condition (K_1) in [4], which are standard conditions to assume for the uniform bound on the KDE. In particular, the condition is satisfied when $K(x) = \phi(p(x))$, where p is a polynomial and ϕ is a bounded real function of bounded variation as in [9], which covers commonly used kernels, such as Gaussian, Epanechnikov, Uniform, etc.

Uniformity on a ray of bandwidths

Under Assumptions, we now have our main concentration inequality for $\sup_{h\geq l_n, x\in\mathbb{X}}\left|\hat{p}_h(x)-p_h(x)\right|$.

Theorem ([7, Theorem 12, Corollary 13])**.** *Let P be a probability distribution and let K be a kernel function satisfying Assumptions above. Suppose $l_n \leq 1$ and*

$$\limsup_n \frac{\log(1/\ell_n)+\log(2/\delta)}{nl_n^d} < \infty.$$

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

$$\sup_{h\geq l_n, x\in\mathbb{X}} |\hat{p}_h(x)-p_h(x)| \leq C\sqrt{\frac{\log(1/\ell_n)+\log(2/\delta)}{nl_n^d}}, \tag{4}$$

where C is a constant depending only on A, $\|K\|_\infty$, d, ν.

Sketch of proof.

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

where $\mathcal{G} \coloneqq \{f-\mathbb{E}_P[f] : f\in\tilde{\mathcal{F}}_{K,[l_n,\infty)}\}$. And then with probability $1-\delta$,

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

And then further show that $\mathcal{G}$ is also uniformly bounded VC dimension, and then

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

Then combining two gives

$$\sup_{h\geq l_n, x\in\mathbb{X}} |\hat{p}_h(x)-p_h(x)| \leq C\left(\frac{\nu B}{n}\log\left(\frac{2AB}{\sigma}\right)+\sqrt{\frac{\nu\sigma^2}{n}\log\left(\frac{2AB}{\sigma}\right)}+\sqrt{\frac{\sigma^2\log(\frac{1}{\delta})}{n}}+\frac{B\log(\frac{1}{\delta})}{n}\right).$$

Now we apply $B = l_n^{-d}\|K\|_\infty$ and $\sigma^2 = Cl_n^{-d}$ to get

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

$\square$

For fixed bandwidth case, there exists a high probability lower bound of matching order (up to log term), so the above upper bound is not improvable and is therefore optimal up to a $\log(1/h_n)$ term.

Proposition ([7, Proposition 16])**.** *Suppose P is a distribution and K is a kernel function satisfying Assumptions above, and further, $K(0)>0$ and K is continuous at 0. Suppose $\lim_n nh_n^d=\infty$. Then, with probability $1-\delta$, the following holds for all large enough n and small enough h_n:*

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

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

By combining the lower and upper bounds together, we conclude that, with high probability,

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

for all large enough n. Similar holds for a fixed bandwidth as well. They imply that the uniform convergence KDE bounds are optimal up to $\log(1/l_n)$ terms for both the fixed bandwidth and the ray on bandwidths cases.

Now, we come back to the persistent homology. We need to check that both $\hat{p}_h$ and p_h are q-tame. One sufficient condition is that $p_h(x)$ is going to 0 uniformly as $x \to \infty$:

Lemma. *If $f : \mathbb{R}^d \to [0, \infty)$ is a nonnegative and continuous function such that $\|f\|_\infty < \infty$ and $\lim_{r\to\infty} \sup\{f(x) : \|x\| \geq r\} = 0$, then the persistent homology $\mathcal{P}(f)$ of the superlevel filtrations $\left\{f^{-1}[\lambda, \infty)\right\}_{\lambda \in \mathbb{R}}$ is q-tame.*

Suppose the distribution P and the kernel K satisfy the above assumptions, and K and p satisfy the above condition in the Lemma. Then $\hat{p}_h$ and p_h also satisfy the above condition, and $\hat{p}_h$ and p_h are q-tame. Then we have the corresponding result: with probability $1 - \delta$,

$$d_B(\mathcal{P}_k(\hat{p}_{h_n}), \mathcal{P}_k(p_{h_n})) \leq \sqrt{\frac{\log(1/h_n)}{nh_n^d}}.$$

Consistency of persistent homologies from kernel density estimators with Čech complex or Vietoris-Rips complex

A slightly different version of computing the persistent homology from kernel density estimator is by [1].

We are targeting the persistent homology $\mathcal{P}(p)$ of the density p, induced from the superlevel filtration $\left\{p^{-1}[\lambda, \infty)\right\}_{\lambda \in \mathbb{R}}$. For estimating $\mathcal{P}(p)$, instead of using $\mathcal{P}(\hat{p}_h)$, we use a mixture of KDE filtration and Čech complex: for $\lambda \in \mathbb{R}$, define

$$\mathcal{X}_\lambda := \{X_i : \hat{p}_r(X_i) \geq \lambda; 1 \leq i \leq n\} = \hat{p}_r^{-1}[\lambda, \infty) \cap \mathcal{X},$$

and then

$$\hat{D}_\lambda(\mathcal{X}, r) := \mathcal{X}_\lambda^r = \left\{x \in \mathbb{R}^d : d(x, \mathcal{X}_\lambda) < r\right\}.$$

Then we let $\hat{\mathcal{P}}(p)$ be the persistent homology induced from the superlevel filtration

$$\left\{\hat{D}_\lambda(\mathcal{X}, r)\right\}_{\lambda \in I_\epsilon},$$

where $I_\epsilon \subset [0, \infty)$ is a finite grid with grid size ϵ and $|I_\epsilon| = N_\epsilon$. In other words, $\hat{\mathcal{P}}(p)$ consists of morphisms $\imath_k^{\lambda_1,\lambda_2} : H_k\hat{D}_{\lambda_1}(\mathcal{X}, r) \to H_k\hat{D}_{\lambda_2}(\mathcal{X}, r)$ for $\lambda_1 \geq \lambda_2$ and $\lambda_1, \lambda_2 \in I_\epsilon$ induced from inclusions $\hat{D}_{\lambda_1}(\mathcal{X}, r) \subset \hat{D}_{\lambda_2}(\mathcal{X}, r)$. From the Nerve Theorem, it is also equivalent to consider Čech complexes as well.

Then we have the consistency result as follows:

Theorem. *[1, Theorem 3.7] If $r \to 0$ and $nr^d \to \infty$ then*

$$P\left(d_B(\hat{\mathcal{P}}(p), \mathcal{P}(p)) \leq 5\epsilon\right) \geq 1 - 3Nn \exp\left(-Cnr^d\right).$$

And with further appropriate conditions,

$$\lim_{n\to\infty} P\left(d_B(\hat{\mathcal{P}}(p), \mathcal{P}(p)) \leq 5\epsilon\right) = 1.$$

VC dimension

For those who are interested in VC dimension, I put the lecture notes of VC dimension.

To develop uniform bounds we need to introduce some complexity measures. More specifically, given a class of functions $\mathcal{F}$, we need some way to measure how complex the class $\mathcal{F}$ is. If $\mathcal{F} = \{f_1, \ldots, f_N\}$ is finite then an obvious measure of complexity is the size of the set, N. The more challenging case is when $\mathcal{F}$ is infinite.

Shattering Numbers. Let $\mathcal{Z}$ be a set and let $\mathcal{F}$ is a class of binary functions on $\mathcal{Z}$. Thus, each $f \in \mathcal{F}$ maps $\mathcal{Z}$ to $\{0,1\}$. For any $z_1, \ldots, z_n$ define

$$\mathcal{F}_{z_1,\ldots,z_n} = \Big\{ (f(z_1), \ldots, f(z_n)) : \ f \in \mathcal{F} \Big\}. \tag{5}$$

Note that $\mathcal{F}_{z_1,\ldots,z_n}$ is a finite collection of vectors and that $|\mathcal{F}_{z_1,\ldots,z_n}| \leq 2^n$. The set $\mathcal{F}_{z_1,\ldots,z_n}$ is called *the projection of $\mathcal{F}$ onto $z_1, \ldots, z_n$*.

Example. Let $\mathcal{F} = \{f_t : \ t \in \mathbb{R}\}$ where $f_t(z) = 1$ if $z > t$ and $f_t(z) = 0$ of $z \leq t$. Consider three real numbers $z_1 < z_2 < z_3$. Then

$$\mathcal{F}_{z_1,z_2,z_3} = \Big\{ (0,0,0), \ (0,0,1), \ (0,1,1), \ (1,1,1) \Big\}.$$

Define the *growth function* or *shattering number* by

$$s(\mathcal{F}, n) = \sup_{z_1,\ldots,z_n} \Big| \mathcal{F}_{z_1,\ldots,z_n} \Big|. \tag{6}$$

A binary function f can be thought of as an indicator function for a set, namely, $A = \{z : \ f(x) = 1\}$. Conversely, any set can be thought of as a binary function, namely, its indicator function $I_A(z)$. We can therefore re-express the growth function in terms of sets. If $\mathcal{A}$ is a class of subsets of $\mathbb{R}^d$ then $s(\mathcal{A}, n)$ is defined to be $s(\mathcal{F}, n)$ where $\mathcal{F} = \{I_A : \ A \in \mathcal{A}\}|$ is the set of indicator functions and then $s(\mathcal{A}, n)$ is again called the *shattering number*. It follows that

$$s(\mathcal{A}, n) = \max_F s(\mathcal{A}, F)$$

where the maximum is over all finite sets of size n and $s(\mathcal{A}, F) = |\{A \cap F : \ A \in \mathcal{A}\}|$ denotes the number of subsets of F picked out by $\mathcal{A}$. We say that a finite set F of size n is *shattered* by $\mathcal{A}$ if $s(\mathcal{A}, F) = 2^n$.

Theorem. *Let $\mathcal{A}$ and $\mathcal{B}$ be classes of subsets of $\mathbb{R}^d$.*

1. $s(\mathcal{A}, n+m) \leq s(\mathcal{A}, n) s(\mathcal{A}, m)$.
2. *If* $\mathcal{C} = \mathcal{A} \bigcup \mathcal{B}$ *then* $s(\mathcal{C}, n) \leq s(\mathcal{A}, n) + s(\mathcal{B}, n)$
3. *If* $\mathcal{C} = \{A \bigcup B : \ A \in \mathcal{A}, B \in \mathcal{B}\}$ *then* $s(\mathcal{C}, n) \leq s(\mathcal{A}, n) s(\mathcal{B}, n)$.
4. *If* $\mathcal{C} = \{A \bigcap B : \ A \in \mathcal{A}, B \in \mathcal{B}\}$ *then* $s(\mathcal{C}, n) \leq s(\mathcal{A}, n) s(\mathcal{B}, n)$.

Proof. Exercise. □

VC Dimension. Recall that a finite set F of size n is *shattered* by $\mathcal{A}$ if $s(\mathcal{A}, F) = 2^n$. The VC dimension (named after Vapnik and Chervonenkis) of $\mathcal{A}$ is the size of the largest set that can be shattered by $\mathcal{A}$.

The *VC dimension* of a class of set $\mathcal{A}$ is

$$\mathsf{VC}(\mathcal{A}) = \sup \Big\{ n : \ s(\mathcal{A}, n) = 2^n \Big\}. \tag{7}$$

Similarly, the *VC dimension* of a class of binary functions $\mathcal{F}$ is

$$\mathsf{VC}(\mathcal{F}) = \sup \Big\{ n : \ s(\mathcal{F}, n) = 2^n \Big\}. \tag{8}$$

If the VC dimension is finite, then the growth function cannot grow too quickly. In fact, there is a phase transition: $s(\mathcal{F}, n) = 2^n$ for $n < d$ and then the growth switches to polynomial.

Theorem. *Suppose that $\mathcal{F}$ has finite VC dimension d. Then,*

$$s(\mathcal{F}, n) \leq \sum_{i=0}^{d} \binom{n}{i}, \tag{9}$$

and for all $n \geq d$,

$$s(\mathcal{F}, n) \leq \left(\frac{en}{d} \right)^d. \tag{10}$$

Class $\mathcal{A}$	VC dimension $V_{\mathcal{A}}$
$\mathcal{A} = \{A_1, \ldots, A_N\}$	$\leq \log_2 N$
Intervals $[a, b]$ on the real line	2
Discs in $\mathbb{R}^2$	3
Closed balls in $\mathbb{R}^d$	$\leq d + 2$
Rectangles in $\mathbb{R}^d$	$2d$
Half-spaces in $\mathbb{R}^d$	$d + 1$
Convex polygons in $\mathbb{R}^2$	∞

Table 1: The VC dimension of some classes $\mathcal{A}$.

Proof. When $n = d = 1$, (9) clearly holds. We show that now proceed by induction. Suppose that (9) holds for $n-1$ and $d-1$ and also that it holds for $n-1$ and d. We will show that it holds for n and d. Let $h(n, d) = \sum_{i=0}^{d} \binom{n}{i}$. We need to show that $\mathsf{VC}(\mathcal{F}) \leq d$ implies that $s(\mathcal{F}, n) \leq h(n, d)$. Let $F_1 = \{z_1, \ldots, z_n\}$ and $F_2 = \{z_2, \ldots, z_n\}$. Let $\mathcal{F}_1 = \{(f(z_1), \ldots, f(z_n) : \ f \in \mathcal{F}\}$ and $\mathcal{F}_2 = \{(f(z_2), \ldots, f(z_n) : \ f \in \mathcal{F}\}$. For $f, g \in \mathcal{F}$, write $f \sim g$ if $g(z_1) = 1 - f(z_1)$ and $g(z_j) = f(z_j)$ for $j = 2, \ldots, n$. Let

$$\mathcal{G} = \Big\{ f \in \mathcal{F} : \text{ there exists } g \in \mathcal{F} \text{ such that } g \sim f \Big\}.$$

Define $\mathcal{F}_3 = \{(f(z_2), \ldots, f(z_n)) : \ f \in \mathcal{G}\}$. Then $|\mathcal{F}_1| = |\mathcal{F}_2| + |\mathcal{F}_3|$. Note that $\mathsf{VC}(\mathcal{F}_2) \leq d$ and $\mathsf{VC}(\mathcal{F}_3) \leq d - 1$. The latter follows since, if $\mathcal{F}_3$ shatters a set, then we can add z_1 to create a set that is shattered by $\mathcal{F}_1$. By assumption $|\mathcal{F}_2| \leq h(n-1, d)$ and $|\mathcal{F}_3| \leq h(n-1, d-1)$. Hence,

$$|\mathcal{F}_1| \leq h(n-1, d) + h(n-1, d-1) = h(n, d).$$

Thus, $s(\mathcal{F}, n) \leq h(n, d)$ which proves (9).

To prove (10), we use the fact that $n \geq d$ and so:

$$\begin{aligned}
\sum_{i=0}^{d} \binom{n}{i} \quad &\leq \quad \left(\frac{n}{d}\right)^d \sum_{i=0}^{d} \binom{n}{i} \left(\frac{d}{n}\right)^i \leq \left(\frac{n}{d}\right)^d \sum_{i=0}^{n} \binom{n}{i} \left(\frac{d}{n}\right)^i \\
&\leq \quad \left(\frac{n}{d}\right)^d \left(1 + \frac{d}{n}\right)^n \leq \left(\frac{n}{d}\right)^d e^d.
\end{aligned}$$

□

The VC dimensions of some common examples are summarized in Table 1.

Uniform Bounds using VC dimension

Now we extend the concentration inequalities to hold uniformly over sets of functions. We consider results for the case where $\mathcal{F}$ is infinite. We begin with an important result due to Vapnik and Chervonenkis.

Theorem (Vapnik and Chervonenkis)**.** *Let $\mathcal{F}$ be a class of binary functions. For any $t > \sqrt{2/n}$,*

$$\mathbb{P}\left(\sup_{f \in \mathcal{F}} |(P_n - P)f| > t \right) \leq 4\, s(\mathcal{F}, 2n) e^{-nt^2/8},$$

and hence, with probability at least $1 - \delta$,

$$\sup_{f \in \mathcal{F}} |P_n(f) - P(f)| \leq \sqrt{\frac{8}{n} \log\left(\frac{4s(\mathcal{F}, 2n)}{\delta}\right)}. \tag{11}$$

Combining this theorem with uniform bounds from Rademacher complexity gives the following result.

Corollary. *With probability at least $1 - \delta$,*

$$\sup_{f \in \mathcal{F}} |P_n(f) - P(f)| \leq \sqrt{\frac{8 \log s(\mathcal{F}, n)}{n}} + \sqrt{\frac{1}{2n} \log\left(\frac{2}{\delta}\right)}.$$

If $\mathcal{F}$ has finite VC dimension d then, with probability at least $1 - \delta$,

$$\sup_{f \in \mathcal{F}} |P_n(f) - P(f)| \leq 2C\sqrt{\frac{d}{n}} + \sqrt{\frac{1}{2n} \log\left(\frac{2}{\delta}\right)}.$$

References

[1] Omer Bobrowski, Sayan Mukherjee, and Jonathan E. Taylor. Topological consistency via kernel estimation. *Bernoulli*, 23(1):288–328, 2017.

[2] O. Bousquet. A bennett concentration inequality and its application to suprema of empirical processes. *C. R. Acad. Sci. Paris, Ser. I*, 334:495–500, 2002.

[3] Frédéric Chazal, Marc Glisse, Catherine Labruère, and Bertrand Michel. Convergence rates for persistence diagram estimation in topological data analysis. *J. Mach. Learn. Res.*, 16:3603–3635, 2015.

[4] Evarist Giné and Armelle Guillou. On consistency of kernel density estimators for randomly censored data: rates holding uniformly over adaptive intervals. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, 37(4):503 – 522, 2001.

[5] Evarist Giné and Armelle Guillou. Rates of strong uniform consistency for multivariate kernel density estimators. volume 38, pages 907–921. 2002. En l'honneur de J. Bretagnolle, D. Dacunha-Castelle, I. Ibragimov.

[6] Evarist Giné, Vladimir Koltchinskii, and Joel Zinn. Weighted uniform consistency of kernel density estimators. *Ann. Probab.*, 32(3B):2570–2605, 07 2004.

[7] 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.

[8] Partha Niyogi, Stephen Smale, and Shmuel Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete & Computational Geometry*, 39(1-3):419–441, 2008.

[9] Deborah Nolan and David Pollard. u-processes: Rates of convergence. *Ann. Statist.*, 15(2):780–799, 06 1987.

[10] Bharath Sriperumbudur and Ingo Steinwart. Consistency and rates for clustering with dbscan. In Neil D. Lawrence and Mark Girolami, editors, *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, volume 22 of *Proceedings of Machine Learning Research*, pages 1090–1098, La Palma, Canary Islands, 21–23 Apr 2012. PMLR.

[11] Ingo Steinwart and Andreas Christmann. *Support Vector Machines*. Springer Publishing Company, Incorporated, 1st edition, 2008.