# Confidence Set of Persistent Homology

김지수 (Jisu KIM)

통계이론세미나 - 위상구조의 통계적 추정, 2023 가을학기

Imagine a persistence diagram. In the persistence diagram, homological features whose lifetimes (the difference between death and birth) are short are informally considered to be "noise", since corresponding holes will be soon filled out right after they are born. Those features corresponds to points in a persistence diagram lying close to the diagonal. Meanwhile, homological features whose lifetimes are long are considered to be "signal"; those features corresponds to points in a persistence diagram lying far from the diagonal. To statistically separate the noise from the signal and provide statistical interpretation, we use the confidence set (or confidence band). See Figure .

We first recall the confidence set:

Suppose we have a statistical model (i.e. a collection of distributions) $\mathcal{P}$. Let $C_n(X_1, \ldots, X_n)$ be a set constructed using the observed data $X_1, \ldots, X_n$. This is a random set. $C_n$ is a $1-\alpha$ confidence set for a parameter $\theta$ if:

$$P\left(\theta \in C_n(X_1, \ldots, X_n)\right) \geq 1-\alpha.$$

And an asymptotic $1-\alpha$ confidence set for a parameter $\theta$ if

$$\liminf_{n\to\infty} P\left(\theta \in C_n(X_1, \ldots, X_n)\right) \geq 1-\alpha. \tag{1}$$

This means that no matter which distribution in $\mathcal{P}$ generated the data, the set guarantees the coverage property described above.

How should $C_n(X_1, \ldots, X_n)$ be like? A typical way to build the confidence set is to use a ball centered at your estimator: Let $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ denote an estimator for $\theta$, which is a function of a sample, and let $\delta_n = \delta_n(X_1, \ldots, X_n) > 0$. Sometimes $\delta_n$ is computed using bootstrap samples $X_1^*, \ldots, X_n^*$ as well. Then set

$$C_n(X_1, \ldots, X_n) = \overline{\mathcal{B}}_d(\hat{\theta}, \delta_n),$$

where $\overline{\mathcal{B}}_d(\hat{\theta}, \delta_n) = \left\{\theta : d(\theta, \hat{\theta}) \leq \delta_n\right\}$ is the closed ball centered at $\hat{\theta}$ and radius $\delta_n$. Then the above coverage condition becomes

$$\liminf_{n\to\infty} P\left(\theta \in \overline{\mathcal{B}}_d(\hat{\theta}, \delta_n)\right) \geq 1-\alpha, \tag{2}$$

and this is equivalent to

$$\liminf_{n\to\infty} P\left(d(\hat{\theta}, \theta) \leq \delta_n\right) \geq 1-\alpha. \tag{3}$$

In (3), $\delta_n$ is a random variable that upper bounds $d(\hat{\theta}, \theta)$ with probability (asymptotically) $1-\alpha$, and called confidence band.

Let $\mathbb{X} \subset \mathbb{R}^d$ be the target geometric structure, and $P$ be a distribution on $\mathbb{R}^d$ with $\mathrm{supp}(P) = \mathbb{X}$. Let $X_1, \ldots, X_n$ be i.i.d. samples from $P$ and $\mathcal{X} = \{X_1, \ldots, X_n\}$. For the confidence set of persistent homology, the distance is the bottleneck distance $d_B$, and $\theta(P)$ and $\hat{\theta}(\mathcal{X})$ should be appropriate persistent homologies (or persistence diagrams) of $P$ and $\mathcal{X}$, denoted as $\mathcal{D}(P)$ and $\mathcal{D}(\mathcal{X})$, respectively. Also see Figure . Then (2) and (3) become

$$\liminf_{n\to\infty} P\left(\mathcal{D}(P) \in \overline{\mathcal{B}}_{d_B}(\mathcal{D}(\mathcal{X}), \delta_n)\right) \geq 1-\alpha, \tag{4}$$

where $\overline{\mathcal{B}}_{d_B}(\mathcal{D}(\mathcal{X}), \delta_n) = \{\mathcal{D} : d_B(\mathcal{D}, \mathcal{D}(\mathcal{X})) \leq \delta_n\}$, and

$$\liminf_{n\to\infty} P\left(d(\mathcal{D}(\mathcal{X}), \mathcal{D}(P)) \leq \delta_n\right) \geq 1-\alpha. \tag{5}$$

We consider two cases:

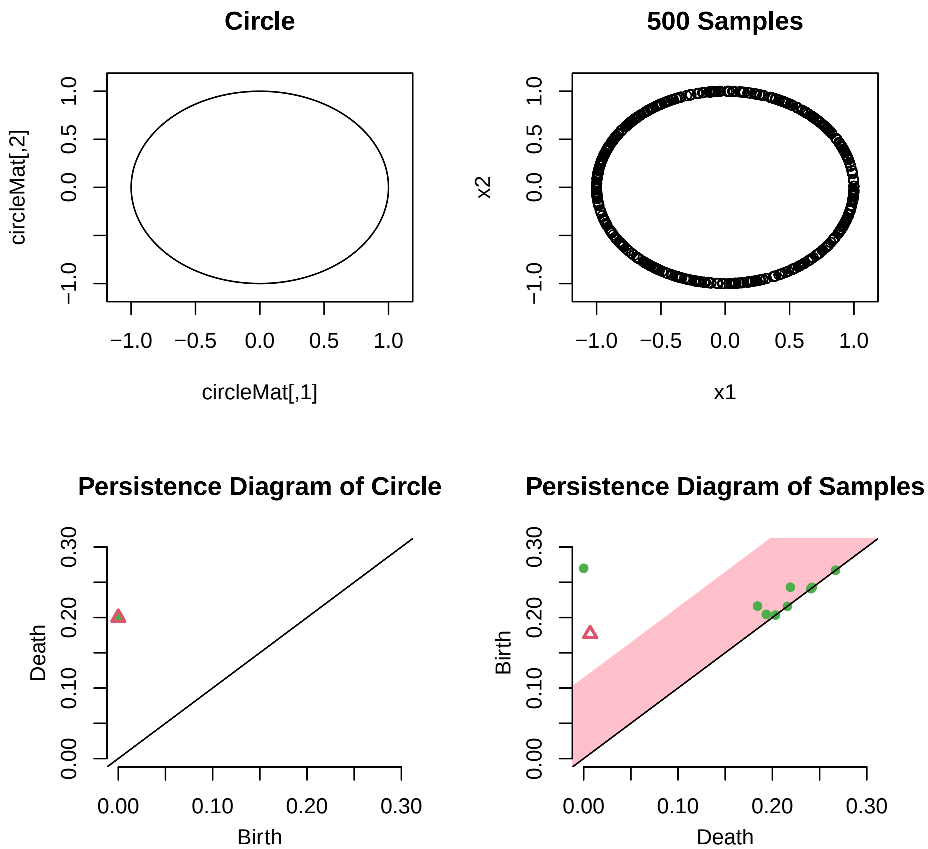


Figure 1: We use the confidence set / band to statistically separate the noise from the signals. In the persistence diagram (right), points above the pink band are topological signals, while points inside the pink band are noise.

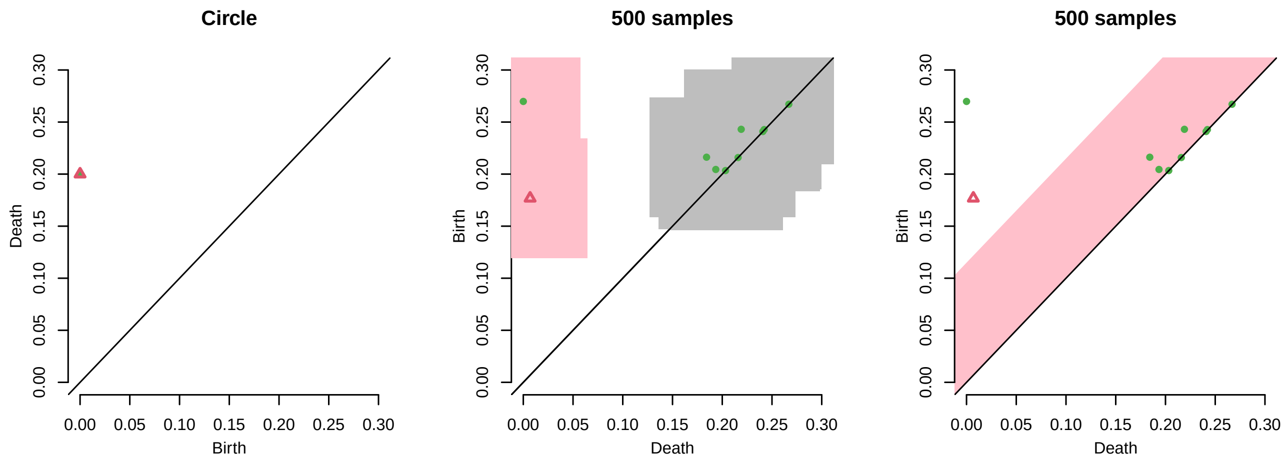


Figure 2: We use the confidence set / band to statistically separate the noise from the signals. In the persistence diagram (right), points above the pink band are topological signals, while points inside the pink band are noise.

1. Persistent homologies from Čech complexes and Vietoris-Rips complexes. Let $\mathcal{DC}_{\mathbb{R}^d}(\mathbb{X})$ and $\mathcal{DC}_{\mathbb{R}^d}(\mathcal{X})$ be the $k$-th dimensional persistence diagrams induced from Čech complexes $\left\{H_k \check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathbb{X}, r)\right\}_{r\in\mathbb{R}}$ and $\left\{H_k \check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathcal{X}, r)\right\}_{r\in\mathbb{R}}$, respectively. Similarly, let $\mathcal{DR}(\mathbb{X})$ and $\mathcal{DR}(\mathcal{X})$ be the $k$-th dimensional persistence diagrams induced from Vietoris-Rips complexes $\{H_k\mathrm{Rips}(\mathbb{X}, r)\}_{r\in\mathbb{R}}$ and $\{H_k\mathrm{Rips}(\mathcal{X}, r)\}_{r\in\mathbb{R}}$, respectively. We would like to find $\delta_n$ such that $\liminf_{n\to\infty} P\left(d_B\left(\mathcal{DC}_{\mathbb{R}^d}(\mathcal{X}), \mathcal{DC}_{\mathbb{R}^d}(\mathbb{X})\right) < \delta_n\right) \geq 1-\alpha$ and $\liminf_{n\to\infty} P\left(d_B\left(\mathcal{DR}(\mathcal{X}), \mathcal{DR}(\mathbb{X})\right) < \delta_n\right) \geq 1-\alpha$.

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 $\mathcal{D}(\hat{p}_h), \mathcal{D}(p_h), \mathcal{D}(p)$ be the $k$-th dimensional persistence diagrams induced from $\hat{p}_h, p_h, p$, respectively, where $p_h = \mathbb{E}\left[\hat{p}_h\right]$ and $p$ is the density of $P$. We would like to know either $\liminf_{n\to\infty} P\left(d_B\left(\mathcal{D}(\hat{p}_h), \mathcal{D}(p_h)\right) < \delta_n\right) \geq 1-\alpha$ or $\liminf_{n\to\infty} P\left(d_B\left(\mathcal{D}(\hat{p}_h), \mathcal{D}(p)\right) < \delta_n\right) \geq 1-\alpha$.

## Confidence set 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 $\mathcal{X} \subset \mathbb{X}$,*

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

Hence bounding the bottleneck distance between persistent homologies from Čech complexes and Vietoris-Rips complexes can be sufficed by bounding Hausdorff distance. In other words, it suffices to find $\delta_n > 0$ such that

$$\liminf_{n\to\infty} P\left(d_H(\mathbb{X}, \mathcal{X}) \leq \delta_n\right) \geq 1-\alpha.$$

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 $x \in \mathrm{supp}(P)$ and for all $r < r_0$,

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

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

### Method I: Subsampling.

Subsampling can be used to construct estimators of the quantiles of the distribution that behave well uniformly over a large class of distributions. The usual approach to subsampling is based on the assumption that we have an estimator $\hat{\theta}$ of a parameter $\theta$ such that $f(n)(\hat{\theta} - \theta)$ converges in distribution to some fixed distribution $J$ for some $\xi > 0$. Unfortunately, our problem is not of this form. Nonetheless, we can still use subsampling as long as we are willing to have conservative confidence intervals.

I first explain the usual approach for subsampling for estimating quantiles of the distribution of $f(n)(\hat{\theta} - \theta)$. Denote by $J_n(x, P)$ the distribution of $f(n)(\hat{\theta} - \theta)$ at $x$, i.e., $J_n(x, P) = P\left(f(n)(\hat{\theta} - \theta) \leq x\right)$. In order to describe the subsampling approach to approximate $J_n(x, P)$, let $b = b_n < n$ be a sequence of positive integers tending to infinity, but satisfying $b/n \to 0$, and define $N_n = \begin{pmatrix} n \\ b \end{pmatrix}$. For $i = 1, \ldots, N_n$, denote by $\mathcal{X}_{n,b}^i$ the $i$th subset of data of size $b$. We consider a feasible subsampling-based estimator of the distribution of $f(n)(\hat{\theta} - \theta)$ as

$$\hat{L}_n(x) = \frac{1}{N_n}\sum_{i=1}^{N_n} I\left(f(n)(\hat{\theta}(\mathcal{X}_{n,b}^i) - \hat{\theta}(\mathcal{X})) \leq x\right).$$

**Theorem** ([13, Theorem 2.1, Corollary 2.1])**.** *Let $b = b_n < n$ be a sequence of positive integers tending to infinity, but satisfying $b/n \to 0$. then under the conditions that $\hat{L}_n(x)$ converges to $J_n(x, P)$ uniformly over $x \in \mathbb{R}$and $P \in \mathcal{P}$, then*

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} P\left\{\hat{L}_n^{-1}(\alpha_1) \leq f(n)(\hat{\theta} - \theta) \leq \hat{L}_n^{-1}(1-\alpha_2)\right\} \geq 1 - \alpha_1 - \alpha_2,$$

*for any $\alpha_1, \alpha_2 \geq 0$ with $0 \leq \alpha_1 + \alpha_2 < 1$.*

For our case, we want to estimate the quantiles of the distributions $d_H(\mathbb{X}, \mathcal{X})$. We consider a subsampling estimator of the distribution of $d_H(\mathbb{X}, \mathcal{X})$ as

$$\hat{L}_n(x) = \frac{1}{N_n} \sum_{i=1}^{N_n} I\left(d_H(\mathcal{X}, \mathcal{X}_{n,b}^i) \leq x\right),$$

and let $c_b = 2\hat{L}_n^{-1}(1-\alpha)$.

**Theorem** ([13, Theorem 2.1, Corollary 2.1]). *Let $P$ be a distribution on $\mathbb{R}^d$ with $\operatorname{supp}(P) = \mathbb{X}$, and assume $P$ satisfies $(a,k)$ assumption with $a, k > 0$. Let $X_1, \ldots, X_n$ be i.i.d. samples from $P$, and let $\mathcal{X} = \{X_1, \ldots, X_n\}$. Let $b = o\left(\frac{n}{\log n}\right)$ be a sequence of positive integers, and define $N_n = \begin{pmatrix} n \\ b \end{pmatrix}$. For $i = 1, \ldots, N_n$, denote by $\mathcal{X}_{n,b}^i$ the $i$th subset of data of size $b$. Then,*

$$\begin{aligned} &P\left(d_B(\mathcal{DC}_{\mathbb{R}^d}(\mathbb{X}), \mathcal{DC}_{\mathbb{R}^d}(\mathcal{X})) \leq c_b\right), P\left(d_B(\mathcal{DR}(\mathbb{X}), \mathcal{DR}(\mathcal{X})) \leq c_b\right) \\ &\geq P\left(d_H(\mathbb{X}, \mathcal{X}) \leq c_b\right) \geq 1 - \alpha + O\left((\frac{b}{n})^{1/4}\right). \end{aligned}$$

**Method II: Concentration of measure.**

Recall the probabilistic bound of Hausdorff distance $d_H(\mathbb{X}, \mathcal{X})$:

**Proposition** ([11, Proposition 7.2][3, Theorem 2]). *Let $P$ be a distribution on $\mathbb{R}^d$ with $\operatorname{supp}(P) = \mathbb{X}$, 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 $t_0 > 0$ such that for all $t < t_0$,*

$$P\left(d_H(\mathbb{X}, \mathcal{X}) < t\right) \geq 1 - a^{-1}t^{-b}\exp(-nat^b). \tag{6}$$

We just solve (6) numerically. Let $t_n(\alpha) < t_0$ be the solution to the equation

$$a^{-1}t^{-b}\exp(-nat^b) = \alpha,$$

then

$$P\left(d_H(\mathbb{X}, \mathcal{X}) < t_n(\alpha)\right) \geq 1 - \alpha.$$

For making a confidence set based on this, we need to know $a$ and $b$. $b$ can be estimated as well, but we regard $b$ as given. For e.g., $b$ can be the dimension of the manifold $\mathbb{X}$. Let $r_n$ be a positive small number, and then we consider the plug-in estimator of $a$,

$$\hat{a}_n = \min_i \left\{ r_n^{-b} \frac{1}{n} \sum_{j=1}^{n} I\left(X_j \in \mathcal{B}(X_i, r_n/2)\right) \right\}.$$

Then if $r_n$vanishes at an appropriate rate as $n \to \infty$, $\hat{a}_n$ is a consistent estimator of $a$.

**Proposition** ([4, Theorem 5]). *Let $P$ be a distribution on $\mathbb{R}^d$ satisfying that for all $x \in \operatorname{supp}(P)$ and for all $r < r_0$,*

$$ar^b \leq P\left(\mathcal{B}(x, r)\right) \leq a'r^b.$$

*Let $X_1, \ldots, X_n$ be i.i.d. samples from $P$, and $r_n \asymp \left(\frac{\log n}{n}\right)^{1/(b+2)}$. Then*

$$\hat{a}_n - a = O_P(r_n).$$

We now use $\hat{a}_n$ to estimate $t_n(\alpha)$ as follows. Assume that $n$ is even, and split the data randomly into two halves, $\mathcal{X} = \mathcal{X}_1 \sqcup \mathcal{X}_2$. Let $\hat{a}_n$ be the plug-in estimator of $a$ computed from $\mathcal{X}_1$, and define $\hat{t}_{1,n}$ to solve the equation

$$\hat{a}_n^{-1}t^{-b}\exp(-n\hat{a}_n t^b) = \alpha. \tag{7}$$

**Theorem** ([4, Theorem 5]). *Let $\mathcal{DC}_{\mathbb{R}^d}(\mathcal{X}_2)$ and be $\mathcal{DR}(\mathcal{X}_2)$ the k-th dimensional persistence diagrams induced from Čech complexes or Vietoris-Rips complexes, respectively, with the second halves $\mathcal{X}_2$. Then,*

$$\begin{aligned} &P\left(d_B(\mathcal{DC}_{\mathbb{R}^d}(\mathbb{X}), \mathcal{DC}_{\mathbb{R}^d}(\mathcal{X}_2)) \leq \hat{t}_{1,n}\right), P\left(d_B(\mathcal{DR}(\mathbb{X}), \mathcal{DR}(\mathcal{X}_2)) \leq \hat{t}_{1,n}\right) \\ &\geq P\left(d_H(\mathbb{X}, \mathcal{X}) \leq \hat{t}_{1,n}\right) \geq 1 - \alpha + O\left( (\frac{\log n}{n})^{1/(2+b)} \right). \end{aligned}$$

In practice, [4] has found that solving (7) for $\hat{t}_n$ without splitting the data also works well although they do not have a formal proof. Another way to define $\hat{t}_n$ which is simpler but more conservative, is to define

$$\hat{t}_n = \left( \frac{2}{n\hat{a}_n} \log\left(\frac{n}{\alpha}\right) \right)^{1/b}.$$

Then $\hat{t}_n = u_n\left(1 + O(\hat{a}_n - a)\right)$ where $u_n = \left(\frac{a}{n\hat{a}_n}\log\left(\frac{n}{\alpha}\right)\right)^{1/b}$, and so

$$\begin{aligned} P\left(d_H(\mathbb{X}, \mathcal{X}) \leq \hat{t}_n\right) &= P\left(d_H(\mathbb{X}, \mathcal{X}) \leq u_n\right) + O\left( (\frac{\log n}{n})^{1/(2+b)} \right) \\ &\geq 1 - \alpha + O\left( (\frac{\log n}{n})^{1/(2+b)} \right). \end{aligned}$$

## Confidence set of persistent homologies from kernel density estimators

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 $p_h$ can be sufficed by bounding their infinity distances $\|\hat{p}_h - p_h\|_\infty$ . In other words, it suffices to find $\delta_n > 0$ such that

$$\liminf_{n\to\infty} P\left(\|\hat{p}_h - p_h\|_\infty \leq \delta_n\right) \geq 1 - \alpha.$$

For topological data analysis, we often fix $h$: when the goal is to correctly estimate the density $p$, it is necessary to have $h \to 0$. However, when the goal is to estimate "topological information" of the distribution $P$, topological information carried by $p_h$ is often equivalent to $p$. For example, suppose the support of the kernel $K$ is $\mathrm{supp}(K) = \mathcal{B}(0,1)$. Then when $\mathrm{supp}(p) = \mathbb{X}$, then $\mathrm{supp}(p_h) = \overline{\mathbb{X}^h} = \{x \in \mathbb{R}^d : d(x, \mathbb{X}) \leq h\}$, (closed) $h$-offset of $\mathbb{X}$. And we have already seen that $\mathbb{X}^h$ and $\mathbb{X}$ are homotopy equivalent under suitable conditions. Further, $P$ might not have the density $p$ but $p_h$ always exists, and then $\hat{p}_h \to \infty$ if $h \to 0$ but $\hat{p_h} \to p_h$ if $h$ is fixed. Also, the $\hat{p}_h$'s convergence to $p_h$ is $\asymp \sqrt{\frac{1}{nh^d}}$ (when density $p$ exists) while to $p$ is $\asymp h^{2\beta} + \sqrt{\frac{1}{nh^d}}$ for some constant $\beta > 0$, so the convergence to $p_h$ is much faster if we fix $h$. See [4, Section 4.4] for more discussions.

**Finite sample band**

**Lemma** ([4, Lemma 9])**.** *Let $P$ be a distribution on $\mathbb{R}^d$ with $\mathrm{supp}(P) = \mathbb{X}$, with $\mathbb{X} \subset [-C, C]^d$. Let $X_1, \ldots, X_n$ be i.i.d. samples from $P$. Assume that $\sup_x K(x) = K(0)$ and that $K$ is L-Lipschitz, that is, $|K(x) - K(y)| \leq L \|x - y\|_2$. Then*

$$P\left(\|\hat{p}_h - p_h\|_\infty > \delta\right) \leq 2\left(\frac{4CL\sqrt{d}}{\delta h^{d+1}}\right)^d \exp\left(-\frac{n\delta^2 h^{2d}}{2K^2(0)}\right).$$

The proof of the above lemma uses Hoeffding's inequality. A sharper result can be obtained by using Bernstein's inequality; however, this introduces extra constants that need to be estimated.

We can use the above lemma to approximate the persistence diagram for $p_h$, denoted by $\mathcal{D}(p_h)$, with the diagram for $\hat{p}_h$, denoted by $\mathcal{D}(\hat{p}_h)$:

**Corollary** ([4, Corollary 10])**.** *Let $\delta_n$ solve*

$$\left(\frac{4CL\sqrt{d}}{\delta_n h^{d+1}}\right)^d \exp\left(-\frac{n\delta_n^2 h^{2d}}{2K^2(0)}\right) = \alpha.$$

*Then*

$$P\left(d_B(\mathcal{D}(p_h), \mathcal{D}(\hat{p}_h)) \leq \delta_n\right) \geq P\left(\|\hat{p}_h - p_h\|_\infty \leq \delta_n\right) \geq 1 - \alpha.$$

**Asymptotic bootstrap confidence band**

A tighter—albeit only asymptotic—bound can be obtained using large sample theory. The simplest approach is the bootstrap.

First, recall the pivotal bootstrap confidence interval:

et $\theta = T(P)$ and $\hat{\theta}_n = T(P_n)$ and define the pivot $R_n = \hat{\theta}_n - \theta$. Let $\hat{\theta}^*_{n,1}, \ldots, \hat{\theta}^*_{n,B}$ denote bootstrap replications of $\hat{\theta}_n$. Let $H(r)$ denote the cdf of the pivot:

$$H(r) = \mathbb{P}(R_n \leq r).$$

Define

$$C_n^* = \left(\hat{\theta}_n - H^{-1}\left(1 - \frac{\alpha}{2}\right), \hat{\theta}_n - H^{-1}\left(\frac{\alpha}{2}\right)\right).$$

Then

$$\mathbb{P}\left(\theta \in C_n^*\right) = 1 - \alpha.$$

Hence, $C_n^*$ is an exact $1 - \alpha$ confidence interval for $\theta$. Unfortunately, computing $C_n^*$ depends on the unknown distribution $H$ but we can form a bootstrap estimate of $H$:

$$\hat{H}(r) = \frac{1}{B}\sum_{b=1}^{B} I(R^*_{n,b} \leq r),$$

where $R^*_{n,b} = \hat{\theta}^*_{n,b} - \hat{\theta}_n$. Let $r^*_\beta$ denote the $\beta$ sample quantile of $(R^*_{n,1}, \ldots, R^*_{n,B})$. It follows that the $1 - \alpha$ bootstrap confidence interval is

$$C_n = \left(\hat{\theta}_n - r^*_{1-\alpha/2}, \hat{\theta}_n - r^*_{\alpha/2}\right).$$

For our case, $\theta = p_h$ and $\hat{\theta} = \hat{p}_h$. Let $X_1^*, \ldots, X_n^*$ be a sample from the empirical distribution $P_n$. Then $\hat{\theta}_n^* = \hat{p}_h^*$, the kernel density estimator constructed from $X_1^*, \ldots, X_n^*$ . We use the pivot as $\sqrt{nh^d}\,\|\hat{p}_h - p_h\|_\infty$ instead of $\|\hat{p}_h - p_h\|_\infty$, due to the reason which will be clarified later. Let $H(r)$ denote the cdf of the pivot:

$$H(r) = P\left(\sqrt{nh^d}\,\|\hat{p}_h - p_h\|_\infty \leq r\right).$$

And then we let

$$C_n^* = \left(\hat{p}_h - \frac{H^{-1}(1-\alpha)}{\sqrt{nh^d}}, \hat{p}_h + \frac{H^{-1}(1-\alpha)}{\sqrt{nh^d}}\right),$$

where $f \in (g, h)$ is understood as $g(x) \leq f(x) \leq h(x)$ for all $x \in \mathbb{R}^d$. Then $p \in C_n^*$ if and only if $\sqrt{nh^d}\,\|\hat{p}_h - p_h\|_\infty \leq H^{-1}(1-\alpha)$, so

$$P\left(p_h \in C_n^*\right) = 1 - \alpha.$$

As above, we form a bootstrap estimate of $H$:

$$\hat{H}(r) = \frac{1}{B}\sum_{b=1}^{B} I\left(\sqrt{nh^d}\left\|\hat{p}_h^{(b)} - \hat{p}_h\right\|_\infty \le r\right),$$

where $\hat{p}^*_{h,b}$ is the kernel density estimator computed from the $b$-th bootstrap sample $X_1^{(b)}, \ldots, X_n^{(b)}$. And let

$$Z_\alpha \coloneqq \hat{H}^{-1}(1-\alpha) = \inf\left\{r : \frac{1}{B}\sum_{b=1}^{B} I\left(\sqrt{nh^d}\left\|\hat{p}_h^{(b)} - \hat{p}_h\right\|_\infty \le r\right) \ge 1-\alpha\right\}.$$

Then the $1-\alpha$ bootstrap confidence interval is

$$C_n = \left(\hat{p}_h - \frac{Z_\alpha}{\sqrt{nh^d}}, \hat{p}_h + \frac{Z_\alpha}{\sqrt{nh^d}}\right).$$

**Theorem** ([4, Theorem 12])**.** *As $n \to \infty$ and $B$ sufficiently large with respect to $n$,*

$$P\left(d_B(\mathcal{D}(p_h), \mathcal{D}(\hat{p}_h)) \le \frac{Z_\alpha}{\sqrt{nh^d}}\right) \ge P\left(\sqrt{nh^d}\,\|\hat{p}_h - p_h\|_\infty \le \frac{Z_\alpha}{\sqrt{nh^d}}\right) = 1-\alpha + O\left(\sqrt{\frac{1}{n}}\right).$$

The algorithm for computing the confidence set $C_n$ can be summarized as below:

1. Given a sample $X = \{X_1, \ldots, X_n\}$, compute the kernel density estimator $\hat{p}_h$.
2. Draw $X^* = \{X_1^*, \ldots, X_n^*\}$ from $X = \{X_1, \ldots, X_n\}$ (with replacement), and compute $\theta^* = \sqrt{nh^d}\,\|\hat{p}_h^* - \hat{p}_h\|_\infty$, where $\hat{p}_h^*$ is the density estimator computed using $X^*$.
3. Repeat the previous step $B$ times to obtain $\theta_1^*, \ldots, \theta_B^*$.
4. Compute $Z_\alpha = \inf\left\{r : \frac{1}{B}\sum_{j=1}^{B} I(\theta_j^* \le r) \ge 1-\alpha\right\}$.
5. The $(1-\alpha)$ confidence band for $p_h$ is $\left[\hat{p}_h - \frac{Z_\alpha}{\sqrt{nh^d}}, \hat{p}_h + \frac{Z_\alpha}{\sqrt{nh^d}}\right]$.

*Remark.* We have emphasized fixed h asymptotics since, for topological inference, it is not necessary to let $h \to 0$ as $n \to \infty$. However, it is possible to let $h \to 0$ if one wants. Suppose $h \equiv h_n$ and $h \to 0$ as $n \to \infty$. We require that $nh^d/\log n \to \infty$ as $n \to \infty$. As before, let $Z_\alpha$ be the bootstrap quantile. It follows from [10, Theorem 3.4] that

$$P\left(d_B(\mathcal{D}(p_h), \mathcal{D}(\hat{p}_h)) \le \frac{Z_\alpha}{\sqrt{nh^d}}\right) \ge P\left(\sqrt{nh^d}\,\|\hat{p}_h - p_h\|_\infty \le \frac{Z_\alpha}{\sqrt{nh^d}}\right) = 1-\alpha + O\left(\left(\frac{\log n}{nh^d}\right)^{(4+d)/(4+2d)}\right).$$

**Bottleneck bootstrap**

The previous bootstrap confidence band is by bootstrapping on the distance $\|\hat{p}_h - p_h\|_\infty$ and by using the stability theorem. However, more precise inferences can be obtained by directly bootstrapping the persistence diagram. Let $\hat{t}_\alpha$ be

$$\hat{t}_\alpha = \inf\left\{r : \frac{1}{B}\sum_{b=1}^{B} I\left(\sqrt{n}\,d_B(\mathcal{D}(\hat{p}_h^{(b)}), \mathcal{D}(\hat{p}_h)) \le r\right) \ge 1-\alpha\right\}.$$

**Theorem** ([1, Corollary 20])**.** *Let $P$ be a distribution on $\mathbb{R}^d$ with $\mathrm{supp}(P) = \mathbb{X}$, and let $X_1, \ldots, X_n$ be i.i.d. samples from $P$. Suppose $\mathbb{X}$ is a compact manifold with boundary. Let $K : \mathbb{R}^d \to \mathbb{R}$ be a kernel function satisfying that $p_h = \mathbb{E}[\hat{p}_h]$ is Morse and has finitely many critical points. Then*

$$P\left(d_B(\mathcal{D}(p_h), \mathcal{D}(\hat{p}_h)) \le \frac{\hat{t}_\alpha}{\sqrt{n}}\right) = 1-\alpha + O\left(\frac{\log n}{\sqrt{n}}\right).$$

## Bootstrap Empirical Process of kernel density estimators

Bootstrap empirical process can be used to find a confidence band for a function $h(t)$; that is, we find a pair of functions $a(t)$ and $b(t)$ such that the probability that $h(t) \in [a(t), b(t)]$ for all $t$ is at least $1-\alpha$. I refer the reader to [2], Van der Vaart and Wellner [1996], and [9] for more details.

An empirical process is a stochastic process based on a random sample. Let $X_1, \ldots, X_n$ be independent and identically distributed random variables taking values in the measure space $(\mathbb{X}, P)$. For a measurable function $f : \mathbb{X} \to \mathbb{R}$, we denote $Pf = \int f dP$ and $P_n f = \int f dP_n = \frac{1}{n}\sum_{i=1}^n f(X_i)$. By the law of large numbers $P_n f$ converges almost surely to $Pf$. Given a class $\mathcal{F}$ of measurable functions, we define the empirical process $\mathbb{G}_n$ indexed by $\mathcal{F}$ as

$$\{\mathbb{G}_n f\}_{f\in\mathcal{F}} = \{\sqrt{n}(P_n f - Pf)\}_{f\in\mathcal{F}}.$$

$\ell^\infty(\mathcal{F})$ is the collection of all bounded functions $\phi : \mathcal{F} \to \mathbb{R}$, equipped with the sup norm. We say $\{\mathbb{G}_n f\}_{f\in\mathcal{F}}$ converges in distribution (or converges weakly) to $\{\mathbb{G} f\}_{f\in\mathcal{F}}$ in the space $\ell^\infty(\mathcal{F})$ if, for any bounded continuous function $H : \ell^\infty(\mathcal{F}) \to \mathbb{R}$, $\mathbb{E}H(\{\mathbb{G}_n f\}_{f\in\mathcal{F}}) \to \mathbb{E}H(\{\mathbb{G} f\}_{f\in\mathcal{F}})$ holds.

**Definition** ([2, Definition 1.3][9, Section 2.1])**.** A class $\mathcal{F}$ of measurable functions $f : \mathbb{X} \to \mathbb{R}$ is called $P$-Donsker if the process $\{\mathbb{G}_n f\}_{f\in\mathcal{F}}$ converges in distribution to a limit process in the space $\ell^\infty(\mathcal{F})$. The limit process is a Gaussian process $\mathbb{G}$ with zero mean and covariance function $\mathbb{E}[\mathbb{G}f\mathbb{G}g] := Pfg - PfPg$; this process is known as a Brownian Bridge.

One sufficient condition for Donsker class is to assume bound on the covering number: a set $\mathcal{C} = \{f_1, \ldots, f_N\}$ is an $\epsilon$-cover of $\mathcal{F}$ if, for every $f \in \mathcal{F}$ there exists a $f_j \in \mathcal{C}$ such that $\|f - f_j\|_{L_2(Q)} < \epsilon$, and the size of the smallest $\epsilon$-cover is called the covering numberand is denoted by $N_p(\mathcal{F}, L_2(Q), \epsilon)$.

**Theorem** ([2, Lemma 2.3][9, Theorem 2.5])**.** *Let $\mathcal{F}$ be an appropriately measurable class of measurable functions with $F$ satisfying $f(x) \le F(x)$ for all $f \in \mathcal{F}$ with $PF^2 < \infty$. Suppose*

$$\int_0^1 \sqrt{\log \sup_Q \mathcal{N}(\mathcal{F}, L_2(Q), \epsilon\,\|F\|_{Q,2})}d\epsilon < \infty,$$

*then $\mathcal{F}$ is $P$-Donsker.*

Let $P_n^* f = \frac{1}{n}\sum_{i=1}^n f(X_i^*)$ where $\{X_1^*, \ldots, X_n^*\}$ is a bootstrap sample from $P_n$. the measure that puts mass $1/n$ on each element of the sample $\{X_1, \ldots, X_n\}$. The bootstrap empirical process $\mathbb{G}_n^*$ indexed by $\mathcal{F}$ is defined as

$$\{\mathbb{G}_n^* f\}_{f\in\mathcal{F}} = \{\sqrt{n}(P_n^* f - P_n f)\}_{f\in\mathcal{F}}.$$

**Theorem** ([2, Theorem 1.4][9, Theorem 2.6, Theorem 2.7])**.** *$\mathcal{F}$ is $P$-Donsker if and only if $\mathbb{G}_n$ converges in distribution to $\mathbb{G}$ in $\ell^\infty(\mathcal{F})$.*

In words, above theorem states that $\mathcal{F}$ is $P$-Donsker if and only if the bootstrap empirical process converges in distribution to the limit process $\mathbb{G}$. Suppose we are interested in constructing a condence band of level $1-\alpha$ for $\{Pf\}_{f\in\mathcal{F}}$, where $\mathcal{F}$ is $P$-Donsker. Let $\hat{\theta} = \sup_{f\in\mathcal{F}} |\mathbb{G}_n f|$. We proceed as follows:

1. Draw $X_1^*, \ldots, X_n^* \sim P_n$ and compute $\hat{\theta}^* = \sup_{f\in\mathcal{F}} |\mathbb{G}_n^* f|$.

2. Repeat the previous step $B$ times to obtain $\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$.

3. Compute $Z_\alpha = \inf\left\{r : \frac{1}{B}\sum_{j=1}^B I(\hat{\theta}_j^* \le r) \ge 1-\alpha\right\}$.

4. For $f \in \mathcal{F}$ define the confidence band $C_n(f) = \left[P_n f - \frac{Z_\alpha}{\sqrt{n}}, P_n f + \frac{Z_\alpha}{\sqrt{n}}\right]$.

Now we turn to the kernel density estimator. For fixed $h > 0$, let $\mathcal{F} = \{K_{x,h} : x \in \mathbb{X}\}$ and $\tilde{\mathcal{F}} = \left\{\tilde{K}_{x,h} : x \in \mathbb{X}\right\}$, where $K_{x,h}, \tilde{K}_{x,h} : \mathbb{R}^d \to \mathbb{R}$ is $K_{x,h}(\cdot) = K\left(\frac{x-\cdot}{h}\right)$ and $\tilde{K}_{x,h} = h^{-d}K_{x,h}$. Then it follows that $P\tilde{K}_{x,h} = p_h$, $P_n\tilde{K}_{x,h} = \hat{p}_h$, and $\hat{\theta} = \sup_{K_{x,h}\in\mathcal{F}} \left|\mathbb{G}_n\tilde{K}_{x,h}\right| = \sqrt{n}\,\|\hat{p}_h - p_h\|_\infty$. Then, the validity of the bootstrap empirical process is sufficed by whether $\tilde{\mathcal{F}}$, or equivalently $\mathcal{F}$, is $P$-Donsker. One sufficient condition is that $\mathcal{F}$ is a uniformly bounded VC-class, which is defined imposing appropriate bounds on the metric entropy of the function class [6, 14, 8]:

**Assumption.** *We assume $\mathcal{F} := \{K_{x,h} : x \in \mathbb{X}\}$ is a uniformly bounded VC-class with dimension $\nu$, 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 $\mathcal{N}(\mathcal{F}, L_2(Q), \epsilon)$ satisfy*

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

*where the covering numbers is the minimal number of open balls of radius $\epsilon$ with respect to $L_2(Q)$ distance whose centers are in $\mathcal{F}$ to cover $\mathcal{F}$.*

Note that $K_{x,h}(x) \le \|K\|_\infty$, so this assumption implies

$$\int_0^1 \sqrt{\log \sup_Q \mathcal{N}(\mathcal{F}, L_2(Q), \epsilon \|F\|_{Q,2})} d\epsilon \le \int_0^1 \sqrt{\nu \log(A/\epsilon)} d\epsilon < \infty,$$

and this implies that $\mathcal{F}$ is $P$-Donsker.

One sufficient condition is to impose uniformly bounded VC class condition on a larger function class,

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

This is implied by condition $(K)$ in [7] or condition $(K_1)$ in [5], 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 $\phi$ is a bounded real function of bounded variation as in [12], which covers commonly used kernels, such as Gaussian, Epanechnikov, Uniform, etc.

However, this is not equivalent to having that

$$\tilde{\mathcal{F}}_{(0,\infty)} = \left\{h^{-d} K_{x,h} : x \in \mathbb{X}, h > 0\right\}$$

is a uniformly bounded VC class. In fact, when we allow $h$ to vary among $(0, \infty)$, then $\tilde{\mathcal{F}}_{(0,\infty)}$ is not $P$-Donsker anymore.

# References


[1] Frédéric Chazal, Brittany Fasy, Fabrizio Lecci, Bertrand Michel, Alessandro Rinaldo, and Larry Wasserman. Robust topological inference: distance to a measure and kernel distance. *J. Mach. Learn. Res.*, 18:Paper No. 159, 40, 2017.

[2] Frédéric Chazal, Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, Aarti Singh, and Larry A. Wasserman. On the bootstrap for persistence diagrams and landscapes. *CoRR*, abs/1311.0376, 2013.

[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] Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, Larry Wasserman, Sivaraman Balakrishnan, and Aarti Singh. Confidence sets for persistence diagrams. *Ann. Statist.*, 42(6):2301–2339, 2014.

[5] 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.

[6] 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.

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

[8] 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.

[9] Michael R. Kosorok. *Introduction to empirical processes and semiparametric inference*. Springer Series in Statistics. Springer, New York, 2008.

[10] Michael H. Neumann. Strong approximation of density estimators from weakly dependent observations by density estimators from independent observations. *Ann. Statist.*, 26(5):2014–2048, 1998.

[11] 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.

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

[13] Joseph P. Romano and Azeem M. Shaikh. On the uniform asymptotic validity of subsampling and the bootstrap. *Ann. Statist.*, 40(6):2798–2822, 2012.

[14] 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.