Stochastic Convergence of Persistence Landscapes and Silhouettes

Frédéric Chazal
INRIA Saclay
France
frederic.chazal@inria.fr

Brittany Terese Fasy
Dept of Computer Science
Tulane University
New Orleans, Louisiana
brittany.fasy@alumni.duke.edu

Fabrizio Lecci
Department of Statistics
Carnegie Mellon University
Pittsburgh, Pennsylvania
lecci@cmu.edu

Alessandro Rinaldo
Department of Statistics
Carnegie Mellon University
Pittsburgh, Pennsylvania
arinaldo@cmu.edu

Larry Wasserman
Department of Statistics
Carnegie Mellon University
Pittsburgh, Pennsylvania
larry@stat.cmu.edu

ABSTRACT

Persistent homology is a widely used tool in Topological Data Analysis that encodes multiscale topological information as a multi-set of points in the plane called a persistence diagram. It is difficult to apply statistical theory directly to a random sample of diagrams. Instead, we can summarize the persistent homology with the persistence landscape, introduced by Bubenik, which converts a diagram into a well-behaved real-valued function. We investigate the statistical properties of landscapes, such as weak convergence of the average landscapes and convergence of the bootstrap. In addition, we introduce an alternate functional summary of persistent homology, which we call the silhouette, and derive an analogous statistical theory.

Categories and Subject Descriptors

G.2 [**Probability and statistics**]: Probabilistic Algorithms

General Terms

Algorithms, Theory, Measurement

Keywords

persistent homology, empirical processes, bootstrap

1. INTRODUCTION

Often, data can be represented as point clouds that carry specific topological and geometric structures. Identifying, extracting, and exploiting these underlying geometric structures has become a problem of fundamental importance for data analysis and statistical learning. With the emergence of new geometric inference and algebraic topology tools, computational topology has recently seen an important development toward data analysis, giving birth to the field of Topological Data Analysis, whose aim is to infer relevant, multiscale, qualitative, and quantitative topological structures directly from the data.

Persistent homology ([11, 19]) is a fundamental tool for providing multi-scale homology descriptors of data. More precisely, it provides a framework and efficient algorithms to quantify the evolution of the topology of a family of nested topological spaces, $\{\mathbb{X}(t)\}_{t\in\mathbb{R}}$, built on top of the data and indexed by a set of real numbers – that can be seen as scale parameters – such that $\mathbb{X}(t) \subseteq \mathbb{X}(s)$ for all $t \leq s$. At the homology level[1], such a filtration induces a family $\{H(\mathbb{X}(t))\}_{t\in\mathbb{R}}$ of homology groups and the inclusions $\mathbb{X}(t) \hookrightarrow \mathbb{X}(s)$ induce a family of homomorphisms $H(\mathbb{X}(t)) \rightarrow H(\mathbb{X}(s))$, for $t \leq s$, which is known as the persistence module associated to the filtration. When the rank of all the homomorphisms $H(\mathbb{X}(t)) \rightarrow H(\mathbb{X}(s))$, are finite the module is said to be q-tame ([3]) and it can be summarized as a set of real intervals (b_i, d_i) representing homological features that appear in the filtration at $t = b_i$ and disappear at $t = d_i$. Such a set of intervals can be represented as a multi-set of points in the real plane and is then called a persistence diagram. Thanks to their stability properties ([9, 3]), persistence diagrams provide relevant multi-scale topological information about the data.

In a more statistical framework, when several data sets are randomly generated or are coming from repeated experiments, one often has to deal with not only one persistence diagram but with a whole distribution of diagrams. Unfortunately, since the space of persistence diagrams is a general metric space, analyzing and quantifying the statistical properties of such a distribution is particularly difficult.

A few attempts have been made towards a statistical analysis of distributions of persistence diagrams. For example, the concentration and convergence properties of persistence diagrams obtained from point clouds randomly sampled on manifolds and from more general compact metric spaces are studied in [1] and [6]. Considering general distributions of

[1]We consider here homology with coefficient in a given field, so the homology groups are vector spaces.

SoCG'14, June 8–11, 2014, Kyoto, Japan.

persistence diagrams, [16] have suggested using the Fréchet average of the diagrams $D_1, \dots, D_n$. Unfortunately, the Fréchet average is unstable and not even unique. A solution that uses a probabilistic approach to define a unique Fréchet average can be found in [14], but its computation remains practically prohibitive.

In this paper, we also consider general distributions of persistence diagrams but we build on a completely different approach, proposed in [2], consisting of encoding persistence diagrams as a collection of real-valued one-Lipschitz functions that are called persistence landscapes; see Section 2. The advantage of landscapes — and, more generally, of any function-valued summaries of persistent homology — is that we can analyze them using existing techniques and theories from nonparametric statistics.

We have in mind two scenarios where multiple persistence diagrams arise:

Scenario 1: We have a random sample of compact sets $K_1, \dots, K_n$ drawn from a probability distribution on the space of compact sets. Each set K_i gives rise to a persistence diagram which in turn yields a persistence landscape function λ_i. An analogous sampling scenario is the one where we observe a sample of n random Morse functions $f_1, \dots, f_n$ from a common probability distribution. Each such function f_i induces a persistence diagram built from its sub-level set filtration, which can again be encoded by a landscape λ_i. The goal is to use the observed landscapes $\lambda_1, \dots, \lambda_n$ to infer the mean landscape $\mu = \mathbb{E}(\lambda_i)$.

Scenario 2: We have a very large dataset with N points. There is a diagram D and landscape λ corresponding to some filtration built on the data. When N is large, computing D is prohibitive. Instead, we draw n subsamples, each of size m. We compute a diagram and landscape for each subsample yielding landscapes $\lambda_1, \dots, \lambda_n$. (Assuming m is much smaller than N, these subsamples are essentially independent and identically distributed.) Then we are interested in estimating $\mu = \mathbb{E}(\lambda_i)$, which can be regarded as an approximation of λ. Two questions arise: how far are the λ_i's from their mean μ? How far is μ from λ? We focus on the first question in this paper.

In both sampling scenarios, we study the statistical behavior as the number of persistence diagrams n grows. We will then analyze the stochastic limiting behavior of the average landscape, as well as the speed of convergence to the limit. Specifically, the contributions of this paper are as follows:

1. We show that the average persistence landscape converges weakly to a Gaussian process and we find the rate of convergence of that process.
2. We show that a statistical procedure known as the bootstrap leads to valid confidence bands for the average landscape. We provide an algorithm to compute confidence bands and illustrate it on a few real and simulated examples.
3. We define a new functional summary of persistent homology, which we call the *silhouette*.

As the proofs are rather technical, we defer the interested reader to the appendices.

2. DIAGRAMS AND LANDSCAPES

A (finite) persistence diagram is a multiset of real intervals $\{(b_i, d_i)\}_{i \in I}$ where I is a finite set. We represent a

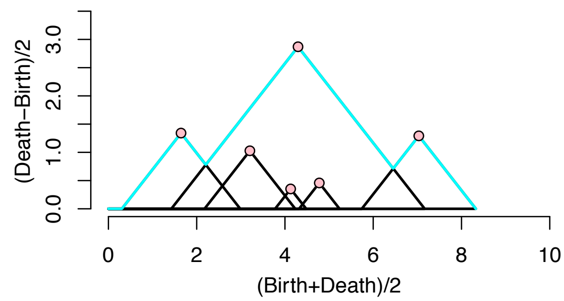

Figure 1: The pink circles are the points in a persistence diagram D. The landscape $\lambda(k, \cdot)$ is the k-th largest of the arrangement of the graphs of $\{\Lambda_p\}$. In particular, the cyan curve is the landscape $\lambda(1, \cdot)$.

persistence diagram as the finite multiset of points $D = \left\{\left(\frac{b_i+d_i}{2}, \frac{d_i-b_i}{2}\right)\right\}_{i \in I}$. Given a positive real number T, we say that D is T-bounded if for each point $(x, y) = \left(\frac{d+b}{2}, \frac{d-b}{2}\right) \in D$, we have $0 \le b \le d \le T$. We denote by $\mathcal{D}_T$ the space of all positive, finite, T-bounded persistence diagrams.

A persistence landscape, introduced in [2], is a sequence of continuous, piecewise linear functions $\lambda \colon \mathbb{Z}^+ \times \mathbb{R} \to \mathbb{R}$ which provides an encoding of a persistence diagram. To define the landscape, consider the set of functions created by tenting each persistence point $p = (x, y) = \left(\frac{b+d}{2}, \frac{d-b}{2}\right) \in D$ to the base line $x = 0$ as with the following function:

$$\Lambda_p(t) = \begin{cases} t - x + y & t \in [x-y, x] \\ x + y - t & t \in (x, x+y] \\ 0 & \text{otherwise} \end{cases} = \begin{cases} t - b & t \in [b, \frac{b+d}{2}] \\ d - t & t \in (\frac{b+d}{2}, d] \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Notice that p is itself on the graph of $\Lambda_p(t)$. We obtain an arrangement of curves by overlaying the graphs of the functions $\{\Lambda_p\}_{p \in D}$; see Figure 1.

The persistence landscape of D is just a summary of this arrangement. Formally, the persistence landscape of D is the collection of functions

$$\lambda_D(k, t) = \underset{p \in D}{\operatorname{kmax}}\, \Lambda_p(t), \quad t \in [0, T], k \in \mathbb{N}, \tag{2}$$

where kmax is the kth largest value in the set; in particular, 1max is the usual maximum function. We set $\lambda_D(k, t) = 0$ if the set $\{\Lambda_p(t), p \in D\}$ contains less than k points. From the definition of persistence landscape, we immediately observe that $\lambda_D(k, \cdot)$ is one-Lipschitz, since Λ_p is one-Lipschitz. We denote by $\mathcal{L}_T$ the space of persistence landscapes corresponding to $\mathcal{D}_T$.

For ease of exposition, in this paper we only focus on the case $k = 1$, and set $\lambda(t) = \lambda_D(1, t)$. However, the results we present hold for $k > 1$.

3. CONVERGENCE OF LANDSCAPES

Let P be a probability distribution on $\mathcal{L}_T$, and let $\lambda_1, \dots, \lambda_n \overset{iid}{\sim} P$. We define the mean landscape as

$$\mu(t) = \mathbb{E}[\lambda_i(t)], \quad t \in [0, T].$$

The mean landscape is an unknown function that we would like to estimate. We estimate μ with the sample average

$$\overline{\lambda}_n(t) = \frac{1}{n}\sum_{i=1}^{n}\lambda_i(t), \quad t \in [0,T].$$

Note that since $\mathbb{E}[\overline{\lambda}_n(t)] = \mu(t)$, we have that $\overline{\lambda}_n$ is a pointwise unbiased estimator of the unknown function μ. Our goal is then quantify how close the resulting estimate is to the function μ. To do so, we first need to explore the statistical properties of $\overline{\lambda}_n$. [2] showed that $\overline{\lambda}_n$ converges pointwise to μ and that the pointwise Central Limit Theorem holds. In this section, we extend these results, proving the uniform convergence of the average landscape. In particular, we show that the process

$$\left\{\sqrt{n}\left(\overline{\lambda}_n(t) - \mu(t)\right)\right\}_{t\in[0,T]} \tag{3}$$

converges weakly (see below) to a Gaussian process on $[0,T]$ and we establish the rate of convergence. For more details on the theory of empirical processes, see [18].

Let

$$\mathcal{F} = \{f_t\}_{0\le t\le T} \tag{4}$$

where $f_t : \mathcal{L}_T \to \mathbb{R}$ is defined by $f_t(\lambda) = \lambda(t)$. Writing $P(f) = \int f dP$ and letting P_n be the empirical measure that puts mass $1/n$ at each λ_i, we can and will regard (3) as an empirical process indexed by $f_t \in \mathcal{F}$. Thus, for $t \in [0,T]$, we will write

$$\begin{aligned}\mathbb{G}_n(t) = \mathbb{G}_n(f_t) &:= \sqrt{n}\left(\overline{\lambda}_n(t) - \mu(t)\right) \\ &= \sqrt{n}(P_n - P)(f_t)\end{aligned} \tag{5}$$

We note that the function $F(\lambda) = T/2$ is a measurable envelope for $\mathcal{F}$.

A Brownian bridge is a Gaussian process on the set of bounded functions from $\mathcal{F}$ to $\mathbb{R}$ such that the process has mean zero and the covariance between any pair $f, g \in \mathcal{F}$ has the form $\int f(u)g(u)dP(u) - \left(\int f(u)dP(u)\right)\left(\int g(u)dP(u)\right)$. A sequence of random objects X_n converges weakly to X, written $X_n \rightsquigarrow X$, if $\mathbb{E}^*(f(X_n)) \to \mathbb{E}(f(X))$ for every bounded continuous function f. (The symbol $\mathbb{E}^*$ is an outer expectation, which is used for technical reasons; the reader can think of this as an expectation.) Thus, we arrive at the following theorem (see Theorem 2.4 in [5]):

Theorem 1 (Weak Convergence of Landscapes). *Let $\mathbb{G}$ be a Brownian bridge with covariance function $\kappa(t,s) = \int f_t(\lambda)f_s(\lambda)dP(\lambda) - \int f_t(\lambda)dP(\lambda)\int f_s(\lambda)dP(\lambda)$, for $t, s \in [0,T]$. Then $\mathbb{G}_n \rightsquigarrow \mathbb{G}$.*

Next, we describe the rate of convergence of the maximum of the normalized empirical process $\mathbb{G}_n$ to the maximum of the limiting distribution $\mathbb{G}$. The maximum is relevant for statistical inference as we shall see in the next section.

For each $t \in [0,T]$, let $\sigma(t)$ be the standard deviation of $\sqrt{n}\,\overline{\lambda}_n(t)$, i.e.

$$\sigma(t) = \sqrt{n\mathrm{Var}(\overline{\lambda}_n(t))} = \sqrt{\mathrm{Var}(f_t(\lambda_1))}. \tag{6}$$

Suppose that $\sigma(t) > c$ for every t in an interval $[t_*, t^*] \subset [0,T]$ and some constant $c > 0$.

Theorem 2 (Uniform CLT). *There exists a random variable $W \stackrel{d}{=} \sup_{t\in[t_*,t^*]}|\mathbb{G}(f_t)|$ such that*

$$\sup_{z\in\mathbb{R}}\left|\mathbb{P}\Big(\sup_{t\in[t_*,t^*]}|\mathbb{G}_n(t)| \le z\Big) - \mathbb{P}\left(W \le z\right)\right| = O\Big(\frac{(\log n)^{\frac{7}{8}}}{n^{\frac{1}{8}}}\Big).$$

Remarks: The assumption in Theorem 2 that the standard deviation function σ is positive over a subinterval of $[0,T]$ can be replaced with the weaker assumption of positivity of σ over a finite collection of sub-intervals without changing the result. We have stated the theorem in this simplified form for ease of readability. Furthermore, it may be possible to improve the term $n^{-1/8}$ in the rate using what is known as a "Hungarian embedding" (see Chapter 19 of [17]). We do not pursue this point further, however.

4. THE BOOTSTRAP FOR LANDSCAPES

Recall that our goal is to use the observed landscapes $(\lambda_1, \ldots, \lambda_n)$ to make inferences about $\mu(t) = \mathbb{E}[\lambda_i(t)]$, where $0 \le t \le T$. Specifically, in this paper we will seek to construct an asymptotic *confidence band* for μ. A pair of functions $\ell_n, u_n \colon \mathbb{R} \to \mathbb{R}$ is an asymptotic $(1-\alpha)$ confidence band for μ if, as $n \to \infty$,

$$\mathbb{P}\Big(\ell_n(t) \le \mu(t) \le u_n(t) \text{ for all } t\Big) \ge 1 - \alpha - O(r_n), \tag{7}$$

where $r_n = o(1)$. Confidence bands are valuable tools for statistical inference, as they allow to quantify and visualize the uncertainty about the mean persistence landscape function μ and to screen out topological noise.

Below, we will describe an algorithm for constructing the functions ℓ_n and u_n from the sample of landscapes $\lambda_1^n := (\lambda_1, \ldots, \lambda_n)$, and prove that it yields an asymptotic $(1-\alpha)$-confidence band for the unknown mean landscape function μ and determine its rate r_n. Our algorithm relies on the use of the *bootstrap*, a simulation-based statistical method for constructing confidence set under minimal assumptions on the data generating distribution P; see [12, 13, 17]. There are several different versions of the bootstrap. This paper uses the *multiplier bootstrap*.

Let $\xi_1^n = (\xi_1, \ldots, \xi_n)$ where $\xi_i \sim N(0,1)$ (Gaussian random variables with mean 0 and variance 1) for all i and define the multiplier bootstrap process

$$\begin{aligned}\tilde{\mathbb{G}}_n(f_t) &= \tilde{\mathbb{G}}_n(\lambda_1^n, \xi_1^n)(f_t) \\ &:= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i\left(f_t(\lambda_i) - \overline{\lambda}_n(t)\right),\ t \in [0,T].\end{aligned} \tag{8}$$

Let $\tilde{Z}(\alpha)$ be the unique value such that

$$\mathbb{P}\left(\sup_{t\in[t_*,t^*]}\left|\tilde{\mathbb{G}}_n(f_t)\right| > \tilde{Z}(\alpha)\ \middle|\ \lambda_1, \ldots, \lambda_n\right) = \alpha. \tag{9}$$

Note that the only random quantities in this definition are $\xi_1, \ldots, \xi_n \sim N(0,1)$. Hence, $\tilde{Z}(\alpha)$ can be approximated by Monte Carlo simulation to a great precision as follows: repeat the bootstrap B times, yielding B processes, $\{\tilde{\mathbb{G}}_n^{(j)}(\cdot), j = 1, \ldots, B\}$, and the corresponding values $\tilde{\theta}_j := \sup_{t\in[t_*,t^*]}|\tilde{\mathbb{G}}_n^{(j)}(f_t)|$, $j = 1, \ldots, B$. Then let

$$\tilde{Z}(\alpha) = \inf\left\{z : \frac{1}{B}\sum_{j=1}^{B} I(\tilde{\theta}_j > z) \le \alpha\right\}. \tag{10}$$

We may take B as large as we like so the Monte Carlo error arbitrarily small. Thus, when using bootstrap methods, one ignores the error in approximating $\tilde{Z}(\alpha)$ as defined in (9) with its simulation approximation as defined in (10). The multiplier bootstrap confidence band is $\{(\ell_n(t), u_n(t)) :\ t \in$

$[t_*, t^*]\}$, where

$$\ell_n(t) = \overline{\lambda}_n(t) - \frac{\tilde{Z}(\alpha)}{\sqrt{n}}, \quad u_n(t) = \overline{\lambda}_n(t) + \frac{\tilde{Z}(\alpha)}{\sqrt{n}}. \quad (11)$$

The steps of the algorithm are given in Algorithm 1.

Algorithm 1 The multipler bootstrap algorithm.

INPUT: Landscapes $\lambda_1, \ldots, \lambda_n$; confidence level $1-\alpha$; number of bootstrap samples B

OUTPUT: confidence functions $\ell_n, u_n : \mathbb{R} \to \mathbb{R}$

Compute the average $\overline{\lambda}_n(t) = \frac{1}{n}\sum_{i=1}^n \lambda_i(t)$, for all t

for $j = 1$ to B **do**

  Generate $\xi_1, \ldots, \xi_n \sim N(0,1)$

  Set $\tilde{\theta}_j = \sup_t n^{-1/2} |\sum_{i=1}^n \xi_i \ (\lambda_i(t) - \overline{\lambda}_n(t))|$

end for

Define $\tilde{Z}(\alpha) = \inf\{z : \frac{1}{B}\sum_{j=1}^B I(\tilde{\theta}_j > z) \leq \alpha\}$

Set $\ell_n(t) = \overline{\lambda}_n(t) - \frac{\tilde{Z}(\alpha)}{\sqrt{n}}$ and $u_n(t) = \overline{\lambda}_n(t) + \frac{\tilde{Z}(\alpha)}{\sqrt{n}}$

return $\ell_n(t), u_n(t)$

The accuracy of the coverage of the confidence band and the width of the band are described in the next result, which follows from Theorem 2 and Proposition 13 in Appendix B.

THEOREM 3 (UNIFORM BAND). *Suppose that $\sigma(t) > c$ for each t in an interval $[t_*, t^*] \subset [0,T]$ and some some constant $c > 0$. Then*

$$\mathbb{P}\Big(\ell_n(t) \leq \mu(t) \leq u_n(t) \text{ for all } t \in [t_*, t^*]\Big) \geq 1 - \alpha - O\left(\frac{(\log n)^{7/8}}{n^{1/8}}\right).$$

Also, $\sup_t (u_n(t) - \ell_n(t)) = O_P\left(\sqrt{\frac{1}{n}}\right)$.

The second statement follows from the fact that $\tilde{Z}(\alpha) = O_P(1)$, where $\tilde{Z}(\alpha)$ is defined in (10). We remark that the randomness is with respect to the joint probabilities of the landscapes and of the ξ's. In [5] a similar asymptotic confidence band is computed for the whole interval $[0,T]$ (see Theorem 2.5), but the rate of convergence is not provided.

The confidence band above has constant width; that is, the width is the same for all t. However, the empirical estimate $\overline{\lambda}(t)$ might be a more accurate estimator of $\mu(t)$ for some t than others. This suggests that we may construct a more refined confidence band whose width varies with t. Hence, we construct a *variable width confidence band*. Consider the standard deviation function σ, defined in (6), and its estimate

$$\hat{\sigma}_n(t) := \sqrt{\frac{1}{n}\sum_{i=1}^n [f_t(\lambda_i)]^2 - [\overline{\lambda}_n(t))]^2}, \quad t \in [0,T]. \quad (12)$$

Define the standardized empirical process

$$\mathbb{H}_n(f_t) = \mathbb{H}_n(\lambda_1^n)(f_t) := \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{f_t(\lambda_i) - \mu(t)}{\sigma(t)}, \quad t \in [t_*, t^*] \quad (13)$$

and, for $\xi_1, \ldots, \xi_n \sim N(0,1)$, define its multiplier bootstrap version: for $\in [t_*, t^*]$,

$$\hat{\mathbb{H}}_n(f_t) = \hat{\mathbb{H}}_n(\lambda_1^n, \xi_1^n)(f_t) := \frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i \frac{f_t(\lambda_i) - \overline{\lambda}_n(t)}{\hat{\sigma}_n(t)}. \quad (14)$$

Just like in the construction of uniform bands, let $\hat{Q}(\alpha)$ be such that

$$\mathbb{P}\Big(\sup_{t \in [t_*, t^*]} \Big|\hat{\mathbb{H}}_n(\lambda_1^n, \xi_1^n)(f_t)\Big| > \hat{Q}(\alpha) \,\Big|\, \lambda_1, \ldots, \lambda_n\Big) = \alpha. \quad (15)$$

Again, $\hat{Q}(\alpha)$ can be computed by simulation to arbitrary precision. The variable width confidence band is $\{(\ell_{\sigma_n}(t), u_{\sigma_n}(t)) : t \in [t_*, t^*]\}$, where

$$\ell_{\sigma_n}(t) = \overline{\lambda}_n(t) - \frac{\hat{Q}(\alpha)\hat{\sigma}_n(t)}{\sqrt{n}}, \quad u_{\sigma_n}(t) = \overline{\lambda}_n(t) + \frac{\hat{Q}(\alpha)\hat{\sigma}_n(t)}{\sqrt{n}}. \quad (16)$$

THEOREM 4 (VARIABLE WIDTH BAND). *Suppose that $\sigma(t) > c > 0$ in an interval $[t_*, t^*] \subset [0,T]$, for some constant c. Then*

$$\mathbb{P}\Big(\ell_{\sigma_n}(t) \leq \mu(t) \leq u_{\sigma_n}(t) \text{ for all } t \in [t_*, t^*]\Big) \geq 1 - \alpha - O\left(\frac{(\log n)^{1/2}}{n^{1/8}}\right).$$

5. THE WEIGHTED SILHOUETTE

The kth persistence landscape $\lambda(k,t)$ can be interpreted as a summary function of the persistence diagrams. A *summary function* is a functor that takes a persistence diagram and outputs a real-valued continuous function. If the diagram corresponds to the distance function to a random set, then we have a probability distribution on the space of summary functions induced by a probability distribution on the original sample space. The persistence landscape is just one of many functions that could be used to summarize a persistence diagram. In this section, we introduce a new family of summary functions called *weighted silhouettes*.

Consider a persistence diagram with m off diagonal points. In this formulation, we take the weighted average of the triangle functions defined in (1):

$$\phi(t) = \frac{\sum_{j=1}^m w_j \Lambda_j(t)}{\sum_{j=1}^m w_j}. \quad (17)$$

Consider two points of the persistence diagram, representing the pairs (b_i, d_i) and (b_j, d_j). In general, we would like to have $w_j \geq w_i$ whenever $|d_j - b_j| \geq |d_i - b_i|$. In particular, let $\phi(t)$ have weights $w_j = |d_j - b_j|^p$, for $p > 0$.

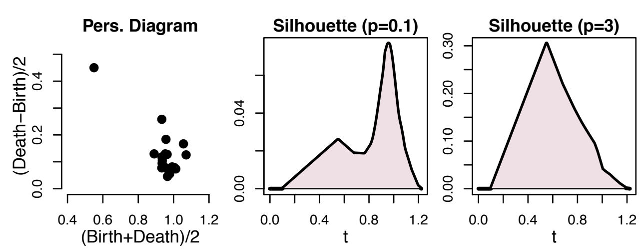

Figure 2: An example of power-weighted silhouettes for different choices of p. The axes are on different scales. The weighted silhouette is one-Lipschitz.

DEFINITION 5 (POWER-WEIGHTED SILHOUETTE). *For every $0 < p \leq \infty$ we define the power-weighted silhouette*

$$\phi^{(p)}(t) = \frac{\sum_{j=1}^m |d_j - b_j|^p \Lambda_j(t)}{\sum_{j=1}^m |d_j - b_j|^p}.$$

The value p can be though of as a trade-off parameter between uniformly treating all pairs in the persistence diagram and considering only the most persistent pairs. Specifically, when p is small, $\phi^{(p)}(t)$ is dominated by the effect of low persistence pairs. Conversely, when p is large, $\phi^{(p)}(t)$ is dominated by the most persistent pair; see Figure 2.

The power-weighted silhouette preserves the property of being one-Lipschitz. In fact, this is true for any choice of non-negative weights. Therefore all the result of Sections 3 and 4 hold for the weighted silhouette, by simply replacing λ with ϕ. In particular, consider $\phi_1, \ldots, \phi_n \sim P_\phi$. Applying theorems 1, 2, 3 and 4, we obtain:

Corollary 6. *The empirical process $\sqrt{n}\left(n^{-1}\sum_{i=1}^{n}\phi_i(t) - \mathbb{E}[\phi(t)]\right)$ converges weakly to a Brownian bridge. The rate of convergence of the maximum of this process to the maximum of the limiting distribution is $O\left(\frac{(\log n)^{7/8}}{n^{1/8}}\right)$.*

Corollary 7. *The multiplier bootstrap algorithm of Algorithm 1 can be used to construct a uniform confidence band for $\{\mathbb{E}[\phi(t)]\}_{t\in[t_*,t^*]}$ with coverage at least $1-\alpha-O\left(\frac{(\log n)^{7/8}}{n^{1/8}}\right)$ and a variable width confidence band with coverage at least $1-\alpha-O\left(\frac{(\log n)^{1/2}}{n^{1/8}}\right)$, where $[t_*,t^*]\subset[0,T]$ is such that $\sqrt{Var(\phi(t))} > c > 0$ for all $t\in[t_*,t^*]$ and some constant c.*

6. EXAMPLES

In Topological Data Analysis, persistent homology is classically used to encode the evolution of the homology of filtered simplicial complexes built on top of data sampled from a metric space; see [4]. For example, given a metric space $(\mathbb{X}, d_{\mathbb{X}})$ and a probability distribution $P_{\mathbb{X}}$ supported on $\mathbb{X}$, one can sample m points, $K = \{X_1,\ldots,X_m\}$, i.i.d. from $P_{\mathbb{X}}$ and consider the Vietoris-Rips (VR) filtration built on top of these points. The persistent homology of this filtration induces a persistence diagram D and a landscape λ. Sampling n such K, one obtains n persistence landscapes $\lambda_1,\ldots,\lambda_n$. In this section, we adopt this setting to illustrate our results on two examples, one real and one simulated.

6.1 Earthquake Data

Figure 3 (left) shows the epicenters of 8000 earthquakes in the latitude/longitude rectangle $[-75, 75]\times[-170, 10]$ of magnitude greater than 5.0 recorded between 1970 and 2009.[2] We randomly sample $m = 400$ epicenters, construct the VR filtration (using the Euclidean distance), compute the first persistence diagram using Dionysus[3] and the corresponding landscape function. We repeat this procedure $n = 30$ times and compute the mean landscape $\overline{\lambda}_n$. Using Algorithm 1, we obtain the uniform 95% confidence band of Theorem 3 and the variable width 95% confidence band of Theorem 4. See Figure 3 (middle). Both the confidence bands have coverage around 95% for the mean landscape $\mu(t)$ that is attached to the distribution induced by the sampling scheme. Similarly, using the same 30 persistence diagrams we construct the corresponding weighted silhouettes using $p = 0.01$ and construct uniform and variable width 95% confidence bands for the mean weighted silhouette $\mathbb{E}[\phi^{(0.01)}(t)]$; see Figure 3 (right). Notice that, for most $t \in [0, T]$, the variable width confidence band is tighter than the fixed-width confidence band.

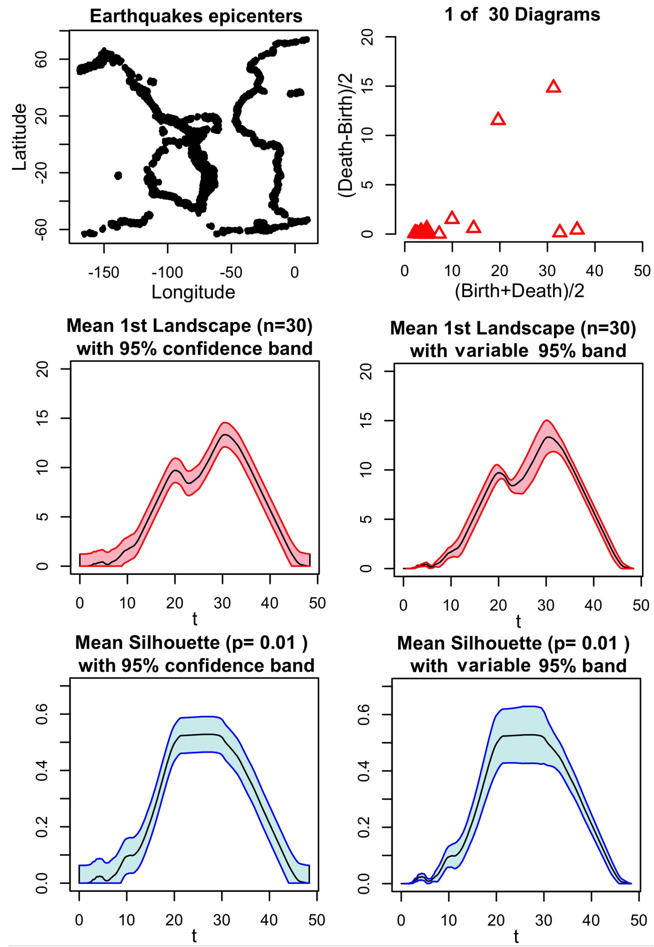

Figure 3: Top: Sample space of epicenters of 8000 earthquakes and one of the 30 persistence diagrams. Middle: uniform and variable width 95% confidence bands for the mean landscape $\mu(t)$. Bottom: uniform and variable width 95% confidence bands for the mean weighted silhouette $\mathbb{E}[\phi^{(0.01)}(t)]$.

6.2 Toy Example: Rings

In this example, we embed the torus $\mathbb{S}^1\times\mathbb{S}^1$ in $\mathbb{R}^3$ and we use the rejection sampling algorithm of [10] (R = 5, r = 1.8) to sample 10,000 points uniformly from the torus. Then, we link it with a circle of radius 5, from which we sample 1,800 points; see Figure 4 (top left). These $N = 11,800$ points constitute the sample space. We randomly sample $m = 600$ of these points, construct the VR filtration, compute the persistence diagram (Betti 1) and the corresponding first and third landscapes and the silhouettes for $p = 0.1$ and $p = 4$. We repeat this procedure $n = 30$ times to construct 95% variable width confidence bands for the mean landscapes $\mu_1(t)$, $\mu_3(t)$ and the mean silhouettes $\mathbb{E}[\phi^{(4)}(t)]$,

[2]USGS Earthquake Search.
http://earthquake.usgs.gov/earthquakes/search/.
[3]Dionysus is a C++ library for computing persistent homology, developed by Dmitriy Morozov. http://mrzv.org/software/dionysus/.

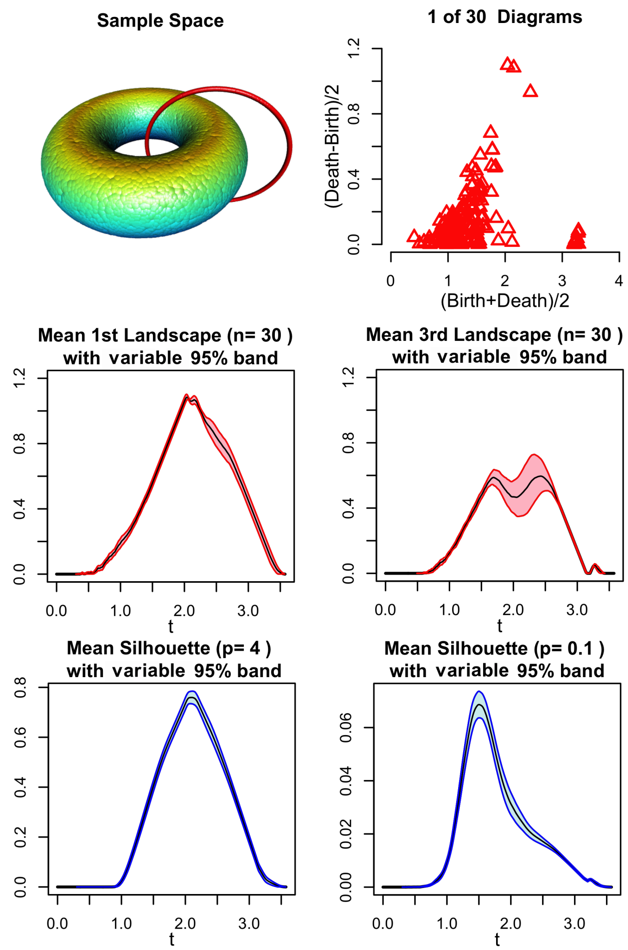

Figure 4: Top: Sample space and one of the 30 persistence diagrams. Middle: variable width 95% confidence bands for the mean first landscape $\mu_1(t)$ and mean third landscape $\mu_3(t)$. Bottom: variable width 95% confidence bands for the mean weighted silhouettes $\mathbb{E}[\phi^{(4)}(t)]$ and $\mathbb{E}[\phi^{(0.1)}(t)]$.

$\mathbb{E}[\phi^{(0.1)}(t)]$. Figure 4 (bottom left) shows one of the 30 persistence diagrams. In the persistence diagram, notice that three persistence pairs are more persistent than the rest. These correspond to the two nontrivial cycles of the torus and the cycle corresponding to the circle. We notice that many of the points in the persistence diagram are hidden by the first landscape. However, as shown in the figure, the third landscape function and the silhouette with parameter $p = 0.1$ are able to detect the presence of these features.

7. DISCUSSION

We have shown how the bootstrap can be used to give confidence bands for Bubenik's persistence landscape and for the persistence silhouette defined in this paper. We are currently working on several extensions to our work including the following: allowing persistence diagrams with countably many points, allowing T to be unbounded, and extending our results to new functional summaries of persistence diagrams. In the case of subsampling (scenario 2 defined in the introduction), we have provided accurate inferences for the mean function μ. We are investigating methods to estimate the difference between μ (the mean landscape from subsampling) and λ (the landscape from the original large dataset). Coupled with our confidence bands for μ, this could provide an efficient approach to approximating the persistent homology in cases where exact computations are prohibitive.

8. REFERENCES

[1] S. Balakrishnan, B. Fasy, F. Lecci, A. Rinaldo, A. Singh, and L. Wasserman. Statistical inference for persistent homology, 2013. arXiv 1303.7117.

[2] P. Bubenik. Statistical topology using persistence landscapes, 2012. arXiv 1207.6437.

[3] F. Chazal, V. de Silva, M. Glisse, and S. Oudot. The structure and stability of persistence modules, July 2012. arXiv 1207.3674.

[4] F. Chazal, V. de Silva, and S. Oudot. Persistence stability for geometric complexes. *Geom. Dedicata*, Dec. 2013.

[5] F. Chazal, B. T. Fasy, F. Lecci, A. Rinaldo, A. Singh, and L. Wasserman. On the bootstrap for persistence diagrams and landscapes, 2013. arXiv 1311.0376.

[6] F. Chazal, C. Labruère, M. Glisse, and B. Michel. Optimal rates of convergence for persistence diagrams in topological data analysis. *arXiv 1305.6239*, 2013.

[7] V. Chernozhukov, D. Chetverikov, and K. Kato. Gaussian approximation of suprema of empirical processes, 2012. arXiv 1212.6885.

[8] V. Chernozhukov, D. Chetverikov, and K. Kato. Anti-concentration and honest adaptive confidence bands, 2013. arXiv 1303.7152.

[9] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. *Discrete Comput. Geom.*, 37(1):103–120, 2007.

[10] P. Diaconis, S. Holmes, and M. Shahshahani. Sampling from a manifold, 2012. arXiv 1206.6913.

[11] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Disc. Comput. Geom.*, 28(4):511–533, July 2002.

[12] B. Efron. Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, pages 1–26, 1979.

[13] B. Efron and R. Tibshirani. *An introduction to the bootstrap*, volume 57. CRC press, 1993.

[14] E. Munch, P. Bendich, K. Turner, S. Mukherjee, J. Mattingly, and J. Harer. Probabilistic Fréchet means and statistics on vineyards, 2013. arXiv 1307.6530.

[15] M. Talagrand. Sharper bounds for Gaussian and empirical processes. *The Annals of Probability*, 22(1):28–76, 1994.

[16] K. Turner, Y. Mileyko, S. Mukherjee, and J. Harer. Fréchet means for distributions of persistence diagrams, 2012. arXiv 1206.2790.

[17] A. van der Vaart. *Asymptotic Statistics*, volume 3. Cambridge UP, 2000.

[18] A. van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Verlag, 1996.

[19] A. Zomorodian and G. Carlsson. Computing persistent homology. *Disc. Comp. Geom.*, 33(2):249–274, 2005.

APPENDIX

A. CHERNOZHUKOV ET AL. (2013)

In this appendix, we summarize the results from [8] that are used in this paper. Given a set of functions $\mathcal{G}$ and a probability measure Q, define the covering number $N(\mathcal{G}, L_2(Q), \varepsilon)$ as the smallest number of balls of size ε needed to cover $\mathcal{G}$, where the balls are defined with respect to the norm $\|g\|^2 = \int g^2(u)dQ(u)$. Let $X_1, \ldots, X_n$ be i.i.d. random variables taking values in a measurable space $(S, \mathcal{S})$. Let $\mathcal{G}$ be a class of functions defined on S and uniformly bounded by a constant b, such that the covering numbers of $\mathcal{G}$ satisfy

$$\sup_Q N(\mathcal{G}, L_2(Q), b\tau) \le (a/\tau)^v \;,\; 0 < \tau < 1 \qquad (18)$$

for some $a \ge e$ and $v \ge 1$ and where the supremum is taken over all probability measures Q on $(S, \mathcal{S})$. The set $\mathcal{G}$ is said to be of VC type, with constants a and v and envelope b. Let σ^2 be a constant such that $\sup_{g\in\mathcal{G}} E[g(X_i)^2] \le \sigma^2 \le b^2$ and for some sufficiently large constant C_1, denote $K_n := C_1 v(\log n \vee \log(ab/\sigma))$. Finally, define

$$\mathbb{G}_n(g) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (g(X_i) - \mathbb{E}[g(X_i)]), g \in \mathcal{G},$$

and let $W_n := \|\mathbb{G}_n\|_{\mathcal{G}} = \sup_{g\in\mathcal{G}} |\mathbb{G}_n(g)|$ denote the supremum of the empirical process $\mathbb{G}_n$.

Theorem 8 (Theorem A.1 in [8]). *Consider the setting specified above. For any $\gamma \in (0,1)$, there is a random variable $W \overset{d}{=} \|\mathbb{G}\|_{\mathcal{G}}$ such that*

$$\mathbb{P}\left(|W_n - W| > \frac{bK_n}{\gamma^{1/2} n^{1/2}} + \frac{\sigma^{1/2} K_n^{3/4}}{\gamma^{1/2} n^{1/4}} + \frac{b^{1/3}\sigma^{2/3} K_n^{2/3}}{\gamma^{1/3} n^{1/6}}\right) \le C_2\left(\gamma + \frac{\log n}{n}\right)$$

for some constant C_2.

Let $\xi_1, \ldots, \xi_n$ be i.i.d. $N(0,1)$ random variables independent of $X_1^n := \{X_1, \ldots, X_n\}$. Let $\xi_1^n := \{\xi_1, \ldots, \xi_n\}$. Define the Gaussian multiplier process

$$\begin{aligned}\tilde{\mathbb{G}}_n(g) &= \tilde{\mathbb{G}}_n(X_1^n, \xi_1^n)(g) \\ &:= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i \left(g(X_i) - \frac{1}{n}\sum_{i=1}^{n}[g(X_i)]\right), \quad g \in \mathcal{G}.\end{aligned}$$

Lastly, for fixed x_1^n, let $\tilde{W}_n(x_1^n) := \sup_{g\in\mathcal{G}} |\tilde{\mathbb{G}}_n(x_1^n, \xi_1^n)(g)|$ denote the supremum of this process.

Theorem 9 (Theorem A.2 in [8]). *Consider the setting specified above. Assume that $b^2 K_n \le n\sigma^2$. For any $\delta > 0$ there exists a set $S_n \in \mathcal{S}^n$ such that $\mathbb{P}(S_n) \ge 1 - 3/n$ and for any $x_1^n \in S_n$ there is a random variable $W \overset{d}{=} \sup_{g\in\mathcal{G}} |\mathbb{G}|$ such that*

$$\mathbb{P}\left(|\tilde{W}_n(x_1^n) - W| > \frac{\sigma K_n^{1/2}}{n^{1/2}} + \frac{b^{1/2}\sigma^{1/2}K_n^{3/4}}{n^{1/4}} + \delta\right) \le C_3\left(\frac{b^{1/2}\sigma^{1/2}K_n^{3/4}}{\delta n^{1/4}} + \frac{1}{n}\right)$$

for some constant C_3.

Theorem 10 (Corollary 2.1 in [8]). *Let $W = (W_t)_{t\in T}$ be a separable Gaussian process indexed by a semimetric space T such that $E[W_t] = 0$ and $E[W_t^2] = 1$ for all $t \in T$. Assume that $\sup_{t\in T} W_t < \infty$ a.s. Then, $a(|W|) := E[\sup_{t\in T}|W_t|] \in [\sqrt{2/\pi}, \infty)$ and*

$$\sup_{x\in\mathbb{R}} \mathbb{P}\left(\left|\sup_{t\in T}|W_t| - x\right| \le \varepsilon\right) \le A\varepsilon a(|W|)$$

for all $\varepsilon \ge 0$ and some constant A.

Theorem 11 (Lemma 6.1 in [7]). *Let $(S, \mathcal{S}, P)$ be a probability space, and let $\mathcal{F} \subset L^2(P)$ be a P-pre-Gaussian class of functions. Denote by $\mathbb{G}$ a tight Gaussian random element in $\ell^\infty(\mathcal{F})$ with mean zero and covariance function $\mathbb{E}[\mathbb{G}(f)\mathbb{G}(g)] = Cov_P(f,g)$ for all $f, g \in \mathcal{F}$. Suppose that there exist constants $\underline{\sigma}$, $\overline{\sigma} > 0$ such that $\underline{\sigma}^2 \le Var_P(f) \le \overline{\sigma}^2$ for all $f \in \mathcal{F}$. Then for every $\varepsilon > 0$,*

$$\sup_{x\in\mathbb{R}} \mathbb{P}\left(\left|\sup_{f\in\mathcal{F}} \mathbb{G}f - x\right| \le \varepsilon\right) \le C_\sigma \varepsilon\left(\mathbb{E}\left[\sup_{f\in\mathcal{F}} \mathbb{G}f\right] + \sqrt{1 \vee \log(\underline{\sigma}/\epsilon)}\right),$$

where C_σ is a constant depending only on $\underline{\sigma}$ and $\overline{\sigma}$.

Theorem 12 (Talagrand's ineq., Th. B.1 in [8]). *Let $\xi_1, \ldots, \xi_n$ be i.i.d. random variables taking values in a measurable space $(S, \mathcal{S})$. Suppose that $\mathcal{G}$ is a measurable class of functions on S uniformly bounded by a constant b such that there exist constants $a \ge e$ and $v > 1$ with $\sup_Q N(\mathcal{G}, L_2(Q), b\varepsilon) \le (a/\varepsilon)^v$ for all $0 < \varepsilon < 1$. Let σ^2 be a constant such that $\sup_{g\in\mathcal{G}} Var(g) \le \sigma^2 \le b^2$. If $b^2 v \log(ab(\sigma)) \le n\sigma^2$, then for all $t \le n\sigma^2/b^2$,*

$$\mathbb{P}\left(\sup_{g\in\mathcal{G}}\left|\sum_{i=1}^{n}\{g(\xi_i) - \mathbb{E}[g(\xi_1)]\}\right| > A\sqrt{n\sigma^2\left[t \vee \left(v \log\frac{ab}{\sigma}\right)\right]}\right) \le e^{-t},$$

where A is an absolute constant.

B. TECHNICAL TOOLS

In this section, we prove some results that will be used in the proofs of Appendix C. Some of our techniques are an adaptation of the strategy used in [8] to construct adaptive confidence bands.

Consider the class of functions $\mathcal{F} = \{f_t\}_{0\le t\le T}$, defined in (4) and let $\lambda_1^n = (\lambda_1, \ldots, \lambda_n)$ be an i.i.d. sample from a probability P on the measurable space $(\mathcal{L}_T, \mathcal{S})$ of persistence landscapes. We summarize the processes used in the analysis of persistence landscapes, given in Sections 3 and 4:

- $\mathbb{G}(f_t)$ is a Brownian Bridge described in Theorem 1,
- $\mathbb{G}_n(f_t) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(f_t(\lambda_i) - \mu(t))$,
- $\tilde{\mathbb{G}}_n(f_t) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i\left(f_t(\lambda_i) - \overline{\lambda}_n(t)\right)$.

For $\sigma(t) > c > 0$, we also defined

- $\mathbb{H}_n(f_t) = \mathbb{H}_n(\lambda_1^n)(f_t) := \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{f_t(B_i)-\mu(t)}{\sigma(t)}$,
- $\hat{\mathbb{H}}_n(f_t) = \tilde{\mathbb{H}}_n(\lambda_1^n, \xi_1^n)(f_t) := \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i\frac{f_t(\lambda_i)-\overline{\lambda}_n(t)}{\hat{\sigma}_n(t)}$,

and for completeness we introduce

- $\mathbb{H}(f_t)$, the standardized Brownian Bridge with covariance function
$$\kappa(t,u)=\int \frac{f_t(\lambda)f_u(\lambda)}{\sigma(t)\sigma(u)}dP(\lambda)-\int \frac{f_t(\lambda)}{\sigma(t)}dP(\lambda)\int \frac{f_u(\lambda)}{\sigma(u)}dP(\lambda) \tag{19}$$
- The process
$$\tilde{\mathbb{H}}_n(f_t):=\hat{\mathbb{H}}_n(\lambda_1^n,\xi_1^n)(f_t):=\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i\frac{f_t(\lambda_i)-\overline{\lambda}_n(t)}{\sigma(t)}, \tag{20}$$
which differs from $\hat{\mathbb{H}}_n(f_t)$ in the use of the standard deviation $\sigma(t)$ that replace its estimate $\hat{\sigma}_n(t)$.

Proposition 13 (Bootstrap Convergence). *Suppose that $\sigma(t)>c>0$ in an interval $[t_*,t^*]\subset[0,T]$, for some constant c. Then, for large n, there exists a random variable $W\overset{d}{=}\sup_{t\in[t_*,t^*]}|\mathbb{G}(f_t)|$ and a set $S_n\in\mathcal{S}^n$ such that $\mathbb{P}(\lambda_1^n\in S_n)\geq 1-3/n$ and, for any fixed $\breve{\lambda}_1^n:=(\breve{\lambda}_1,\ldots,\breve{\lambda}_n)\in S_n$,*
$$\sup_{z\in\mathbb{R}}\left|\mathbb{P}\Big(\sup_{t\in[t_*,t^*]}|\tilde{\mathbb{G}}_n(\breve{\lambda}_1^n,\xi_1^n)(f_t)|\leq z\Big)-\mathbb{P}\left(W\leq z\right)\right| \leq C_6\Big(\frac{(\log n)^{5/8}}{n^{1/8}}\Big),$$
for some constant $C_6>0$.

Proof. Let $\mathcal{F}^*=\{f_t\in\mathcal{F}:t\in[t_*,t^*]\}$. Consider the covering number $N(\mathcal{F}^*,L_2(Q),||F||_2\varepsilon)$ of the class $\mathcal{F}^*$, as defined in Appendix A, with $F=T/2$. In the proof of Theorem 2 we show that
$$\sup_Q N(\mathcal{F}^*,L_2(Q),||F||_2\varepsilon)\leq 2/\varepsilon,$$
where the supremum is taken over all measures Q on $\mathcal{L}_T$. For $n>2$, $b=\sigma=T/2$, $v=1$, $K_n=A\,(\log n\vee 1)$, Theorem 9 implies that there exists a set S_n such that $\mathbb{P}(\lambda_1^n\in S_n)\geq 1-3/n$ and, for any fixed $\breve{\lambda}_1^n:=(\breve{\lambda}_1,\ldots,\breve{\lambda}_n)\in S_n$ and $\delta>0$,
$$\mathbb{P}\left(\Big|\sup_{t\in[t_*,t^*]}|\tilde{\mathbb{G}}_n|-W\Big|>\frac{T\sqrt{A\log n}}{2n^{1/2}}+\frac{T(A\log n)^{3/4}}{2n^{1/4}}+\delta\right) \leq C_3\left(\frac{T(A\log n)^{3/4}}{2\delta n^{1/4}}+\frac{1}{n}\right).$$
Define
$$g(n,\delta,T):=\frac{T(A\log n)^{1/2}}{2n^{1/2}}+\frac{T(A\log n)^{3/4}}{2n^{1/4}}+\delta.$$
Using the strategy of Theorem 2 and applying the anti-concentration inequality of Theorem 11, it follows that for large n and $\breve{\lambda}_1^n:=(\breve{\lambda}_1,\ldots,\breve{\lambda}_n)\in S_n$,
$$\sup_z\left|\mathbb{P}\left(\sup_{t\in[t_*,t^*]}|\tilde{\mathbb{G}}_n(\breve{\lambda}_1^n,\xi_1^n)|\leq z\right)-\mathbb{P}(W\leq z)\right| \leq C_5\,g(n,\delta,T)\sqrt{\log\frac{c}{g(n,\delta,T)}}+C_3\left(\frac{T(A\log n)^{3/4}}{2\delta n^{1/4}}+\frac{1}{n}\right) \tag{21}$$
for some constant $C_5>0$. Choosing $\delta=\frac{(A\log n)^{1/8}}{n^{1/8}}$, we have
$$g(n,\delta,T)=\frac{T(A\log n)^{1/2}}{2n^{1/2}}+\frac{T(A\log n)^{3/4}}{2n^{1/4}}+\frac{(A\log n)^{1/8}}{n^{1/8}}.$$
The result follows by noticing that, $g(n,\delta,T)=O\left(\frac{(\log n)^{1/8}}{n^{1/8}}\right)$ and $\sqrt{\log\frac{c}{g(n,\delta,T)}}=O\left((\log n)^{1/2}\right)$. $\square$

In the following lemma we consider the class $\mathcal{G}_c=\{g_t: g_t=f_t/\sigma(t),\ t_*\leq t\leq t^*\}$ where $f_t\in\mathcal{F}$ is defined in (4) and we bound the corresponding covering number, as in (18).

Lemma 14. *Consider the assumptions of Theorem 4 and consider the class of functions $\mathcal{G}_c=\{g_t:g_t=f_t/\sigma(t),\ t_*\leq t\leq t^*\}$, where $f_t\in\mathcal{F}$. Note that $T/(2c)$ is a measurable envelope for $\mathcal{G}_c$. Then*
$$\sup_Q N(\mathcal{G}_c,L_2(Q),\varepsilon\|T/(2c)\|_{Q,2})\leq(a/\varepsilon)^v,\ 0<\varepsilon<1$$
for $a=(T^2+2c^2)/c^2$ and $v=1$, where the supremum is taken over all measures Q on $\mathcal{L}_T$. $\mathcal{G}_c$ is of VC type, with constants a and v and envelope $T/(2c)$.

Proof. First, using the definition of $\sigma(t)$ given in (6) for $t>u$, we have
$$\begin{aligned}\sigma^2(t)-\sigma^2(u)&=\mathrm{Var}(f_t(\lambda_1))-\mathrm{Var}(f_u(\lambda_1))\\&=\mathbb{E}[f_t^2(\lambda_1)]-(\mathbb{E}[f_t(\lambda_1)])^2-\mathbb{E}[f_u^2(\lambda_1)]+(\mathbb{E}[f_u(\lambda_1)])^2\\&=\mathbb{E}\left[(f_t(\lambda_1)-f_u(\lambda_1))\,(f_t(\lambda_1)+f_u(\lambda_1))\right]+\\&\quad(\mathbb{E}[f_u(\lambda_1)]-\mathbb{E}[f_t(\lambda_1)])\,(\mathbb{E}[f_u(\lambda_1)]+\mathbb{E}[f_t(\lambda_1)])\\&\leq(t-u)\big(\mathbb{E}[f_t(\lambda_1)+f_u(\lambda_1)]+\mathbb{E}[f_u(\lambda_1)]+\mathbb{E}[f_t(\lambda_1)]\big)\\&\leq 2(t-u)T.\end{aligned}$$
Note that we used the fact that $f_t(\lambda)$ is 1-Lipschitz in t and $T/2$ is an envelope of $\mathcal{F}$. Therefore
$$|\sigma(t)-\sigma(u)|=\frac{|\sigma^2(t)-\sigma^2(u)|}{\sigma(t)+\sigma(u)}\leq\frac{|t-u|T}{c}.$$
Using that $f_t(\lambda)$ is one-Lipschitz, we also have that $|\sigma(t)g_t(\lambda)-\sigma(u)g(u)|\leq|t-u|$, for $t,u\in[t_*,t^*]$. Construct a grid $t_*\equiv t_0<t_1<\cdots<t_N\equiv t^*$ such that $t_{j+1}-t_j=\frac{\varepsilon Tc^2}{T^2+2c^2}$. We claim that $\{g_{t_j}:1\leq j\leq N\}$ is an $\varepsilon T/(2c)$-net of $\mathcal{G}_c$. If g_t in $\mathcal{G}_c$, then there exists a j so that $t_j\leq t\leq t_{j+1}$ and
$$\begin{aligned}\|g_{t_{j+1}}-g_t\|_{Q,2}&=\left\|\frac{\sigma(t_{j+1})g_{t_{j+1}}}{\sigma(t_{j+1})}-\frac{\sigma(t)g_t}{\sigma(t)}\right\|_{Q,2}\\&=\left\|\frac{\sigma(t_{j+1})\sigma(t)g_{t_{j+1}}-\sigma(t_{j+1})\sigma(t)g_t}{\sigma(t_{j+1})\sigma(t)}\right\|_{Q,2}.\end{aligned}$$
By subtracting and adding $\sigma^2(t_{j+1})g_{t_{j+1}}$ in the numerator the last quantity becomes
$$\begin{aligned}&\left\|\frac{\sigma(t_{j+1})g_{t_{j+1}}[\sigma(t)-\sigma(t_{j+1})]+\sigma(t_{j+1})[\sigma(t_{j+1})g_{t_{j+1}}-\sigma(t)g_t]}{\sigma(t_{j+1})\sigma(t)}\right\|_{Q,2}\\&\leq\left\|\frac{T[\sigma(t)-\sigma(t_{j+1})]}{2c^2}\right\|_{Q,2}+\frac{t_{j+1}-t}{c}\\&\leq\frac{(t_{j+1}-t)T^2}{2c^3}+\frac{t_{j+1}-t}{c}\leq(t_{j+1}-t_j)\frac{T^2+2c^2}{2c^3}\\&=\frac{\varepsilon Tc^2}{T^2+2c^2}\,\frac{T^2+2c^2}{2c^3}=\frac{\varepsilon T}{2c}.\end{aligned}$$
Thus,
$\sup_Q N(\mathcal{G}_c,L_2(Q),\varepsilon T/(2c))\leq\frac{(T^2+2c^2)(t^*-t_*)}{\varepsilon Tc^2}\leq\frac{T^2+2c^2}{\varepsilon c^2}$. $\square$

Let $\mathbb{H}$ be a Brownian bridge with covariance function given in (19). Then, combining Lemma 14 and Theorem 8, with $\gamma=\frac{(\log n)^{1/2}}{n^{1/8}}$, we obtain:

LEMMA 15. *One can construct a random variable* $Y \stackrel{d}{=} \sup_{t\in[t_*,t^*]} |\mathbb{H}|$ *such that for large* n,

$$\mathbb{P}\left(\Big|\sup_{t\in[t_*,t^*]} |\mathbb{H}_n(f_t)| - Y\Big| > C_7 \frac{(\log n)^{1/2}}{n^{1/8}}\right) \le C_8 \frac{(\log n)^{1/2}}{n^{1/8}}.$$

for some absolute constants C_7 *and* C_8.

Consider $\sigma(t)$ and $\hat\sigma(t)$, defined in (6) and (12).

LEMMA 16. *For large* n *and some constant* C_9,

$$\mathbb{P}\left(\sup_{t\in[t_*,t^*]}\left|\frac{\hat\sigma_n(t)}{\sigma(t)} - 1\right| \ge C_9 \frac{(\log n)^{1/2}}{n^{1/2}}\right) \le \frac{2}{n}. \tag{22}$$

PROOF. Let $\mathcal{G}_c = \{g_t : g_t = f_t/\sigma(t),\ t_* \le t \le t^*\}$ and $\mathcal{G}_c^2 := \{g^2 : g \in \mathcal{G}_c\}$.

By definition $\hat\sigma_n^2(t) = \frac{1}{n}\sum_{i=1}^n f_t^2(\lambda_i) - [\overline{\lambda}_n(t)]^2$ and $\sigma^2(t) = \mathbb{E}[f_t^2(\lambda_1)] - (\mathbb{E}[f_t(\lambda_1)])^2$. Thus

$$\begin{aligned}\left|\frac{\hat\sigma_n(t)}{\sigma(t)} - 1\right| &\le \left|\frac{\hat\sigma_n^2(t)}{\sigma^2(t)} - 1\right| = \left|\frac{\hat\sigma_n^2(t) - \sigma^2(t)}{\sigma^2(t)}\right| \\ &\le \sup_{t\in[t_*,t^*]}\left|\frac{1}{n}\frac{\sum_{i=1}^n f_t^2(\lambda_i)}{\sigma^2(t)} - \frac{\mathbb{E}[f_t^2(\lambda_1)]}{\sigma^2(t)}\right| \\ &\quad + \sup_{t\in[t_*,t^*]}\left|\left[\frac{1}{n}\frac{\sum_{i=1}^n f_t(\lambda_i)}{\sigma(t)}\right]^2 - \left[\frac{\mathbb{E}[f_t(\lambda_1)]}{\sigma(t)}\right]^2\right| \\ &= \sup_{g\in\mathcal{G}_c^2}\left|\frac{1}{n}\sum_{i=1}^n g(\lambda) - \mathbb{E}[g(\lambda)]\right| \\ &\quad + \sup_{g\in\mathcal{G}_c}\left|\left[\frac{1}{n}\sum_{i=1}^n g(\lambda)\right]^2 - (\mathbb{E}[g(\lambda)])^2\right| \end{aligned} \tag{23}$$

Using the same strategy of Lemma 14, it can be shown that $\mathcal{G}_c^2$ is VC type with some constants A and $V \ge 1$ and envelope $T^2/(4c^2)$. Therefore, by Theorem 12, with $t = \log n$ and for large n,

$$\mathbb{P}\left(\sup_{g\in\mathcal{G}_c^2}\left|\frac{1}{n}\sum_{i=1}^n g(\lambda) - \mathbb{E}[g(\lambda)]\right| > C_{10}\frac{(\log n)^{1/2}}{n^{1/2}}\right) \le \frac{1}{n}. \tag{24}$$

Note that
$$\begin{aligned}\sup_{g\in\mathcal{G}_c}&\left|\left[\frac{1}{n}\sum_{i=1}^n g(\lambda)\right]^2 - (\mathbb{E}[g(\lambda)])^2\right| \\ &\le \frac{T}{c}\sup_{g\in\mathcal{G}_c}\left|\frac{1}{n}\sum_{i=1}^n g(\lambda) - \mathbb{E}[g(\lambda)]\right|\end{aligned}$$
and, applying again Theorem 12 to the right hand side, we obtain

$$\mathbb{P}\Big(\sup_{g\in\mathcal{G}_c}\left|\left[\frac{1}{n}\sum_{i=1}^n g(\lambda)\right]^2 - (\mathbb{E}[g(\lambda)])^2\right| > C_{11}\frac{(\log n)^{1/2}}{n^{1/2}}\Big) \le \frac{1}{n}. \tag{25}$$

The inequality of (22) follows from (23), (24) and (25). □

LEMMA 17 (ESTIMATION ERROR OF $\hat Q(\alpha)$). *Let* $Q(\alpha)$ *be the* $(1-\alpha)$*-quantile of the random variable* $Y \stackrel{d}{=} \sup_{t\in[t_*,t^*]} |\mathbb{H}|$ *and* $\hat Q(\alpha)$ *be the* $(1-\alpha)$*-quantile of the random variable* $\sup_{t\in[t_*,t^*]} |\hat{\mathbb{H}}_n|$. *There exist positive constants* C_{12} *and* C_{13} *such that for large* n:

(i) $\mathbb{P}\left[\hat Q(\alpha) < Q\left(\alpha + C_{12}\frac{(\log n)^{3/8}}{n^{1/8}}\right) - C_{13}\frac{(\log n)^{3/8}}{n^{1/8}}\right] \le \frac{5}{n}$,

(ii) $\mathbb{P}\left[\hat Q(\alpha) > Q\left(\alpha - C_{12}\frac{(\log n)^{3/8}}{n^{1/8}}\right) + C_{13}\frac{(\log n)^{3/8}}{n^{1/8}}\right] \le \frac{5}{n}$.

PROOF. Define $\Delta\mathbb{H}_n(f_t) := \hat{\mathbb{H}}_n(f_t) - \tilde{\mathbb{H}}_n(f_t)$. Consider the set $S_{n,1} \in \mathcal{S}^n$ of values $\breve\lambda_1^n$ such that, if $\lambda_1^n \in S_{n,1}$, then

$$\left|\frac{\hat\sigma(t)}{\sigma(t)} - 1\right| \le C_9\frac{(\log n)^{1/2}}{n^{1/2}} \quad \text{for all } t \in [t_*, t^*].$$

By Lemma 16, $\mathbb{P}(\lambda_1^n \in S_{n,1}) \ge 1 - 2/n$. Fix $\breve\lambda_1^n \in S_{n,1}$. Then

$$\Delta\mathbb{H}_n(\breve\lambda_1^n, \xi_1^n)(f_t) := \frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i \frac{f_t(\breve\lambda_i) - \overline{\lambda}_n(t)}{\sigma(t)}\left(\frac{\sigma(t)}{\hat\sigma_n(t)} - 1\right)$$

is a zero-mean Gaussian process with variance

$$\frac{\hat\sigma_n^2(t)}{\sigma^2(t)}\left(\frac{\sigma(t)}{\hat\sigma_n(t)} - 1\right)^2 \le C_9^2\frac{\log n}{n}.$$

Let $\tilde{\mathcal{G}}_c = \{ag : a \in (0,1], g \in \mathcal{G}_c\}$. $\tilde{\mathcal{G}}_c$ is VC type with some constants A and $V \ge 1$ and envelope $T^2/(4c^2)$. Moreover, the uniform covering number of the process $\Delta\mathbb{H}_n(\breve\lambda_1^n, \xi_1^n)(f_t)$ with respect to the natural semimetric (standard deviation) is bounded by the uniform covering number of $\tilde{\mathcal{G}}_c$. Therefore we can apply Theorem 2.4 in [15] (see also Section A.2.2 in [18]) and obtain

$$\begin{aligned}\mathbb{P}&\left(\left|\sup_{t\in[t_*,t^*]}|\hat{\mathbb{H}}(\breve\lambda_1^n)(f_t)| - \sup_{t\in[t_*,t^*]}|\tilde{\mathbb{H}}(\breve\lambda_1^n)(f_t)|\right| \ge \beta_n\right) \\ &\le \mathbb{P}\left(\sup_{t\in[t_*,t^*]}|\Delta\mathbb{H}_n(\breve\lambda_1^n, \xi_1^n)(f_t)| \ge \beta_n\right) \\ &\le D\left(\frac{\beta_n n}{C_9^2\log n}\right)^V \frac{C_9\sqrt{\log n}}{\beta_n\sqrt{n}}\exp\left(-\frac{\beta_n^2 n}{2C_9^2\log n}\right),\end{aligned} \tag{26}$$

for some constant D. For $C_{14} = \sqrt{2}C_9(1+V/2)^{1/2}$ and $\beta_n = C_{14}(\log n)/n^{1/2}$, the last quantity is bounded by $C_{15}/[n(\log n)^{1/2}]$, for some constant C_{15}. Therefore, for large n,

$$\begin{aligned}\mathbb{P}&\left(\left|\sup_t|\hat{\mathbb{H}}(\breve\lambda_1^n)(f_t)| - \sup_t|\tilde{\mathbb{H}}(\breve\lambda_1^n)(f_t)|\right| \ge C_{14}\frac{(\log n)^{3/8}}{n^{1/8}}\right) \\ &\le \mathbb{P}\left(\left|\sup_t|\hat{\mathbb{H}}(\breve\lambda_1^n)(f_t)| - \sup_t|\tilde{\mathbb{H}}(\breve\lambda_1^n)(f_t)|\right| \ge C_{14}\frac{(\log n)}{n^{1/2}}\right) \\ &\le C_{15}\frac{1}{n(\log n)^{1/2}} \le C_{15}\frac{(\log n)^{3/8}}{n^{1/8}}.\end{aligned} \tag{27}$$

By Theorem 9 with $\delta = \frac{(\log n)^{3/8}}{n^{1/8}}$, for large n, there exists a set $S_{n,2} \in \mathcal{S}^n$ such that $\mathbb{P}(\lambda_1^n \in S_{n,2}) \ge 1 - 3/n$, and for any $\breve\lambda_1^n \in S_{n,2}$, one can construct a random variable $Y \stackrel{d}{=} \sup_{t\in[t_*,t^*]} |\mathbb{H}|$ such that

$$\mathbb{P}\Big(\Big|\sup_{t\in[t_*,t^*]}|\tilde{\mathbb{H}}(\breve\lambda_1^n)(f_t)| - Y\Big| \ge C_{16}\frac{(\log n)^{\frac{3}{8}}}{n^{1/8}}\Big) \le C_{17}\frac{(\log n)^{\frac{3}{8}}}{n^{1/8}}. \tag{28}$$

Combining (27) and (28), we have that, for large n and $\breve\lambda_1^n \in S_{n,0} := S_{n,1} \cap S_{n,2}$,

$$\mathbb{P}\Big(\Big|\sup_{t\in[t_*,t^*]}|\hat{\mathbb{H}}(\breve\lambda_1^n)(f_t)| - Y\Big| \ge C_{13}\frac{(\log n)^{\frac{3}{8}}}{n^{1/8}}\Big) \le C_{12}\frac{(\log n)^{\frac{3}{8}}}{n^{1/8}}, \tag{29}$$

for some constants C_{12}, C_{13}.

Let $\hat{Q}(\alpha, \breve{\lambda}_1^n)$ be the conditional $(1-\alpha)$-quantile of $\sup_{t\in[t_*,t^*]} |\hat{\mathbb{H}}(\breve{\lambda}_1^n)(f_t)|$. Then $\hat{Q}(\alpha) = \hat{Q}(\alpha, \breve{\lambda}_1^n)$ is a random quantity and for $\breve{\lambda}_1^n \in S_{n,0}$, we have that

$$\begin{aligned}
&\mathbb{P}\left(Y \le \hat{Q}(\alpha, \breve{\lambda}_1^n) + C_{13}\frac{(\log n)^{3/8}}{n^{1/8}}\right) \\
&\ge \mathbb{P}\left(\left\{Y \le \hat{Q}(\alpha, \breve{\lambda}_1^n) + C_{13}\frac{(\log n)^{3/8}}{n^{1/8}}\right\} \bigcap \right. \\
&\qquad \left. \left\{\left|\sup_{t\in[t_*,t^*]} |\hat{\mathbb{H}}(\breve{\lambda}_1^n)(f_t)| - Y\right| \le C_{13}\frac{(\log n)^{3/8}}{n^{1/8}}\right\}\right) \\
&\ge \mathbb{P}\left(\sup_{t\in[t_*,t^*]} |\hat{\mathbb{H}}(\breve{\lambda}_1^n)(f_t)| \le \hat{Q}(\alpha, \breve{\lambda}_1^n)\right) - C_{12}\frac{(\log n)^{3/8}}{n^{1/8}} \\
&\ge 1 - \alpha - C_{12}\frac{(\log n)^{3/8}}{n^{1/8}}.
\end{aligned}$$

Therefore $Q\left(\alpha + C_{12}\frac{(\log n)^{3/8}}{n^{1/8}}\right) \le \hat{Q}(\alpha) + C_{13}\frac{(\log n)^{3/8}}{n^{1/8}}$ whenever $\lambda_1^n \in S_{n,0}$, which happens with probability at least $1 - 5/n$. This proves part (i) of the theorem. The proof of part (ii) is similar and therefore is omitted. $\square$

C. MAIN PROOFS

Proof of Theorem 2. Let $\mathcal{F}^* = \{f_t \in \mathcal{F} : t \in [t_*, t^*]\}$. The Lipschitz property implies that for every $\lambda \in \mathcal{L}_T$, $|f_t(\lambda) - f_u(\lambda)| = |\lambda(t) - \lambda(u)| \le |t-u|$ and hence $\|f_t - f_u\|_{Q,2} \le |t-u|$. Construct a grid, $0 \equiv t_0 < t_1 < \cdots < t_N \equiv T$ where $t_{j+1} - t_j := \varepsilon \|F\|_{Q,2} = \varepsilon T/2$. In the last equality, we used the constant envelope $F(\lambda) = T/2$. We claim that $\{f_{t_j} : 1 \le j \le N\}$ is an $(\varepsilon T/2)-$net of $\mathcal{F}^*$: choosing $f_t \in \mathcal{F}^*$, then there exists a j so that $t_j \le t \le t_{j+1}$ and

$$\|f_{t_{j+1}} - f_t\|_{Q,2} \le |t_{j+1} - t| \le |t_{j+1} - t_j| = \varepsilon T/2.$$

Thus, we can bound the covering number of $\mathcal{F}^*$, as in (18):

$$\sup_Q N(\mathcal{F}^*, L_2(Q), ||F||_2 \varepsilon) \le \frac{T}{\varepsilon \|F\|_{Q,2}} = 2/\varepsilon,$$

where the supremum is taken over all measures Q on $\mathcal{L}_T$. By Theorem 8, with $b = \sigma = T/2$, $v = 1$, $K_n = A(\log n \vee 1)$, there exists $W \overset{d}{=} \sup_{f\in\mathcal{F}^*} \mathbb{G}$ such that, for $n > 2$,

$$\begin{aligned}
&\mathbb{P}\left(\big|\sup_{t\in[t_*,t^*]} |\mathbb{G}_n| - W\big| > \frac{TA\log n}{2\gamma^{1/2}n^{1/2}} + \frac{T^{1/2}(A\log n)^{3/4}}{\gamma^{1/2}n^{1/4}} + \right. \\
&\qquad \left. + \frac{T(A\log n)^{2/3}}{2\gamma^{1/3}n^{1/6}}\right) \le C_2\left(\gamma + \frac{\log n}{n}\right)
\end{aligned}$$

holds for $n > 2$ and for some constant C_2. Define the event $E := \left\{\big|\sup_{t\in[t_*,t^*]} |\mathbb{G}_n| - W\big| > g(n,\gamma,T)\right\}$, where

$$g(n,\gamma,T) = \frac{TA\log n}{2\gamma^{1/2}n^{1/2}} + \frac{T^{1/2}(A\log n)^{3/4}}{\gamma^{1/2}n^{1/4}} + \frac{T(A\log n)^{2/3}}{2\gamma^{1/3}n^{1/6}}.$$

Then, for any z and large n,

$$\begin{aligned}
&\mathbb{P}\left(\sup_{t\in[t_*,t^*]} |\mathbb{G}_n| \le z\right) - \mathbb{P}(W \le z) \\
&\le \mathbb{P}(W \le z + g(n,\gamma,T)) - \mathbb{P}(W \le z) + \mathbb{P}(E^c) \\
&\le C_4\, g(n,\gamma,T)\sqrt{\log\frac{c}{g(n,\gamma,T)}} + C_2\left(\gamma + \frac{\log n}{n}\right),
\end{aligned}$$

where in the last step we used the anti-concentration inequality of Theorem 11. Similarly,

$$\begin{aligned}
&\mathbb{P}(W \le z) - \mathbb{P}\left(\sup_{t\in[t_*,t^*]} |\mathbb{G}_n| \le z\right) \\
&\le \mathbb{P}(W \le z, E) - \mathbb{P}\left(\sup_{t\in[t_*,t^*]} |\mathbb{G}_n| \le z, E\right) + P(E^c) \\
&\le \mathbb{P}(z - g(n,\gamma,T) \le W \le z, E) + P(E^c) \\
&\le C_4\, g(n,\gamma,T)\sqrt{\log\frac{c}{g(n,\gamma,T)}} + C_2\left(\gamma + \frac{\log n}{n}\right).
\end{aligned}$$

It follows that

$$\begin{aligned}
&\sup_z \left|\mathbb{P}\left(\sup_{t\in[t_*,t^*]} |\mathbb{G}_n| \le z\right) - \mathbb{P}(W \le z)\right| \\
&\le C_4\, g(n,\gamma,T)\sqrt{\log\frac{c}{g(n,\gamma,T)}} + C_2\left(\gamma + \frac{\log n}{n}\right). \quad (30)
\end{aligned}$$

Choosing $\gamma = \frac{(A\log n)^{7/8}}{n^{1/8}}$, we have $g(n,\gamma,T) = \frac{T(A\log n)^{9/16}}{2n^{7/16}} + \frac{T^{1/2}(A\log n)^{5/16}}{n^{3/16}} + \frac{T(A\log n)^{3/8}}{2n^{1/8}}$. The result follows by noticing that, $g(n,\gamma,T) = O\left(\frac{(\log n)^{3/8}}{n^{1/8}}\right)$ and $\sqrt{\log\frac{c}{g(n,\gamma,T)}} = O\left((\log n)^{1/2}\right)$. $\square$

Proof of Theorem 4 (Variable Width Band). Let $\mathbb{H}(f_t)$ be the Brownian bridge with covariance function given in (19). Consider $Y \overset{d}{=} \sup_{t\in[t_*,t^*]} |\mathbb{H}|$. Let $Q(\alpha)$ be the $(1-\alpha)$-quantile of Y and $\hat{Q}(\alpha)$ be the $(1-\alpha)$-quantile of the random variable $\sup_{t\in[t_*,t^*]} |\hat{\mathbb{H}}_n|$.
Let $\varepsilon_1(n) = C_7(\log n)^{1/2}/n^{1/8}$, $\varepsilon_2(n) = C_{13}(\log n)^{3/8}/n^{1/8}$, $\varepsilon_3(n) = C_9(\log n)^{1/2}/n^{1/2}$, and define $\varepsilon(n) = \varepsilon_1(n) + \varepsilon_2(n) + \varepsilon_3(n)Q(\alpha)$.

Similarly let $\delta_1(n) = C_8(\log n)^{1/2}/n^{1/8}$, $\delta_2(n) = 5/n$, $\delta_3(n) = 2/n$, and define $\delta(n) = \delta_1(n) + \delta_2(n) + \delta_3(n)$. Define $\tau(n) = C_{12}(\log n)^{3/8}/n^{1/8}$. Then, for large n,

$$\begin{aligned}
&\mathbb{P}\Big(\ell_\sigma(t) \le \mu(t) \le u_\sigma(t) \text{ for all } t \in [t_*, t^*]\Big) \\
&= \mathbb{P}\left(\sup_{t\in[t_*,t^*]} \left|\mathbb{H}_n(f_t)\frac{\sigma(t)}{\hat{\sigma}_n(t)}\right| \le \hat{Q}(\alpha)\right) \\
&\ge \mathbb{P}\left[\sup_{t\in[t_*,t^*]} |\mathbb{H}_n(f_t)| \le (1 - \varepsilon_3(n))\, Q(\alpha + \tau(n)) - \varepsilon_2(n)\right] \\
&\quad - \delta_2(n) - \delta_3(n),
\end{aligned}$$

where we applied Lemmas 16 and 17. Using Lemma 15, the last quantity is no smaller than

$$\begin{aligned}
&\mathbb{P}\big[Y \le (1 - \varepsilon_3(n))\, Q(\alpha + \tau(n)) - \varepsilon_2(n) - \varepsilon_1(n)\big] \\
&\quad - \delta_1(n) - \delta_2(n) - \delta_3(n) \\
&\ge \mathbb{P}\big[Y \le Q(\alpha + \tau(n)) - \varepsilon(n)\big] - \delta(n) \\
&\ge \mathbb{P}\big[Y \le Q(\alpha + \tau(n))\big] - \sup_{x\in\mathbb{R}} \mathbb{P}\left(\Big|Y - x\Big| \le \varepsilon(n)\right) - \delta(n) \\
&\ge 1 - \alpha - \tau(n) - \delta(n) - \sup_{x\in\mathbb{R}} \mathbb{P}\left(\Big|Y - x\Big| \le \varepsilon(n)\right) \\
&\ge 1 - \alpha - \tau(n) - \delta(n) - A\varepsilon(n),
\end{aligned}$$

where in the last step we applied the anti-concentration inequality of Theorem 10. $\square$