FRÉCHET CHANGE-POINT DETECTION

BY PAROMITA DUBEY[*] AND HANS-GEORG MÜLLER[†]

Department of Statistics, University of California, Davis, [*]pdubey@ucdavis.edu; [†]hgmueller@ucdavis.edu

We propose a method to infer the presence and location of change-points in the distribution of a sequence of independent data taking values in a general metric space, where change-points are viewed as locations at which the distribution of the data sequence changes abruptly in terms of either its Fréchet mean, Fréchet variance or both. The proposed method is based on comparisons of Fréchet variances before and after putative change-point locations and does not require a tuning parameter, except for the specification of cut-off intervals near the endpoints where change-points are assumed not to occur. Our results include theoretical guarantees for consistency of the test under contiguous alternatives when a change-point exists and also for consistency of the estimated location of the change-point, if it exists, where, under the null hypothesis of no change-point, the limit distribution of the proposed scan function is the square of a standardized Brownian bridge. These consistency results are applicable for a broad class of metric spaces under mild entropy conditions. Examples include the space of univariate probability distributions and the space of graph Laplacians for networks. Simulation studies demonstrate the effectiveness of the proposed methods, both for inferring the presence of a change-point and estimating its location. We also develop theory that justifies bootstrap-based inference and illustrate the new approach with sequences of maternal fertility distributions and communication networks.

1. Introduction. Change-point detection has become a popular tool for identifying locations in a data sequence where an abrupt change occurs in the data distribution. A data sequence $\{Y_1, Y_2, \ldots, Y_n\}$ has a change-point at n_0 if $Y_1, \ldots, Y_{n_0}$ comes from a distribution P_1, while $Y_{n_0+1}, \ldots, Y_n$ comes from a different distribution P_2. In the classical framework $\{Y_1, \ldots, Y_n\}$ assume values in $\mathbb{R}$. Change-point detection is an important task in many areas and has been studied thoroughly for univariate data (Carlstein, Müller and Siegmund (1994), Niu, Hao and Zhang (2016)).

The multivariate setting, where the $\{Y_1, Y_2, \ldots, Y_n\}$ take values in $\mathbb{R}^d$, presents additional challenges over the univariate case and can be considered a precursor of the even more challenging case of metric-space valued objects that we study here; in both cases the Y_i cannot be ordered. Change-point detection in multivariate settings has been studied both using parametric (Srivastava and Worsley (1986), James, James and Siegmund (1987, 1992), Chen and Gupta (2012), Csörgő and Horváth (1997)) and nonparametric (Jirak (2015), Lung-Yut-Fong, Lévy-Leduc and Cappé (2015), Matteson and James (2014), Wang and Samworth (2018)) approaches in various settings.

With modern statistical applications moving toward the study of more complex phenomena, statistical analysts increasingly encounter data that correspond to random objects in general spaces that cannot be characterized as univariate or multivariate data. Often such data objects do not even reside in Euclidean spaces, for example, when one deals with observations that correspond to sequences of networks such as brain networks or communication

networks. Other types of such random objects include covariance matrices as encountered in brain connectivity in neuroscience (Ginestet et al. (2017)) and in the analysis of spoken language (Tavakoli et al. (2019)) as well as sequences of random probability distributions, which are quite common (Cazelles et al. (2018), Petersen and Müller (2019a)). We refer to such data as *random objects* which are random variables that take values in a general metric space, generalizing the notion of random vectors. The challenge when dealing with such data is that vector space operations are not available, and one does not have much more to go by than pairwise distances between the objects.

Existing results on change-point detection for the case when the data sequence takes values in a general metric space are quite limited. Parametric approaches exist for change detection in a sequence of networks (De Ridder, Vandermarliere and Ryckebusch (2016), Peel and Clauset (2015), Wang, Yu and Rinaldo (2018)). These approaches have been developed for special cases and are not applicable more generally. There are very few approaches that are more generally applicable to metric space valued random objects, and they include kernel based (Arlot, Celisse and Harchaoui (2012), Garreau and Arlot (2018)) and graph based (Chen and Zhang (2015), Chu and Chen (2019)) methods. A major disadvantage of the kernel based methods is that they heavily depend on the choice of the kernel function and its tuning parameters. Graph based methods are applicable in general metric spaces. However, their starting point is a similarity/dissimilarity graph of the data sequence, which then plays the role of a tuning parameter that critically impacts the resulting inference. Graph based methods also lack theoretical guarantees for the consistency of the estimated change-point when there is a true change-point in the data sequence; furthermore, inference for the presence of a change-point has not yet been developed.

In this paper, we introduce a tuning-free (except for the choice of the size of small intervals at the end-points where change-points are assumed not to occur) method for change-point detection and inference in a sequence of data objects taking values in a general metric space. We provide a test for the presence of a change-point in the sequence, obtain its consistency under contiguous alternatives, and derive a consistency result for the estimated change-point location. We consider the *offline* approach, where the entire sequence of data is available to conduct inference and assume that the data objects Y_i are independent and ordered with some meaningful ordering. The starting point of the proposed method is similar in spirit to classical analysis of variance and builds upon a recently proposed two sample test for random objects (Dubey and Müller (2019)), aiming at inference for the presence of differences in Fréchet means (Fréchet (1948)) and Fréchet variances for samples of random objects.

Fréchet means and variances provide a generalization of center and spread for metric space valued random variables. Our goal in this paper is to develop a method for the detection and, specifically, a test for the presence of a change-point in terms of Fréchet means and/or Fréchet variances of the distributions of the data sequence taking values in a general metric space. The needed assumptions are relatively weak and apply to a broad class of metric spaces, including the space of univariate probability distributions and the space of networks, under suitable metrics. We provide theoretical guarantees for type I error control under the null hypothesis of no change-point by deriving a distribution free limiting distribution of the proposed test statistic under the null, where we show that the proposed scan function converges weakly to the square of a standardized Brownian bridge (Theorem 1). When there is a change-point, we show that the proposed estimate of the location of the change-point is consistent (Proposition 1 and Theorem 2) and demonstrate that the proposed test is consistent and gains power under contiguous alternatives for increasing sample sizes (Theorem 3). We also provide theoretical support for bootstrap versions of the proposed test (Theorem 4).

The organization of the paper is as follows. In Section 2 we describe and explore the proposed test statistic and motivate the proposed estimate of the location of the change-point,

starting from an improved version of the two-sample version of the Fréchet test for random objects (Dubey and Müller (2019)). Theoretical results on the behavior of the proposed test under the null hypothesis of no change-point and the alternative that a change-point is present and on the estimated location of the change-point, if it is present, are in Section 3, while Section 4 is devoted to a study of the finite sample performance of the proposed test statistic under several simulation settings. The proposed inference method for change-points is illustrated in Section 5 with an analysis of Finnish fertility data that reflect the evolution of maternal age distributions over calendar years and also with change-point detection in dynamic networks for the Enron e-mail network data, where in both cases we find strong evidence for the presence of change-points.

2. Methodology.

2.1. *Model.* Let $Y_1, Y_2, \ldots, Y_n$ be a sequence of independent random objects taking values in a metric space (Ω, d) that is totally bounded, that is, for any $\epsilon > 0$ there is a finite number of open ϵ-balls, the union of which covers Ω. Given different probability measures P_1 and P_2 on (Ω, d), we are interested to test the null hypothesis,

(2.1) $$H_0 : Y_1, Y_2, \ldots, Y_n \sim P_1,$$

against the single change-point alternative,

(2.2) $\qquad H_1 :$ there exists $0 < \tau < 1$ such that $\begin{cases} Y_1, Y_2, \ldots, Y_{[n\tau]} \sim P_1 \\ Y_{[n\tau]+1}, Y_{[n\tau]+2}, \ldots, Y_n \sim P_2, \end{cases}$

where τ denotes the change-point. In most practical situations, differences in distributions arise either in location or in scale. Fréchet means and variances (Fréchet (1948)) provide a generalization of the notion of location and scale for probability measures on Euclidean spaces to the case of general metric spaces.

Our goal is to detect differences in Fréchet means or Fréchet variances of the distributions P_1 and P_2. Fréchet means μ_1 and μ_2 of P_1 and P_2, respectively, are defined as

$$\mu_1 = \operatorname*{argmin}_{\omega \in \Omega} E_1\big(d^2(Y, \omega)\big) \quad \text{where } Y \sim P_1 \quad \text{and}$$

$$\mu_2 = \operatorname*{argmin}_{\omega \in \Omega} E_2\big(d^2(Y, \omega)\big) \quad \text{where } Y \sim P_2,$$

and the corresponding Fréchet variances as

$$V_1 = \min_{\omega \in \Omega} E_1\big(d^2(Y, \omega)\big) \quad \text{where } Y \sim P_1 \quad \text{and}$$

$$V_2 = \min_{\omega \in \Omega} E_2\big(d^2(Y, \omega)\big) \quad \text{where } Y \sim P_2.$$

Our assumptions below will ensure that these notions are well defined when we employ them in the following. Here, $E_1(\cdot)$ and $E_2(\cdot)$ denote expectations taken with respect to P_1 and P_2, respectively, and we will use this notation throughout in what follows, which then implicitly specifies the assumed distribution of the generic random object Y. Under the null hypothesis H_0 of no change-point, we have $P_1 = P_2 = P$ and use the notations $\mu = \mu_1 = \mu_2$ and $V = V_1 = V_2$ to denote the Fréchet mean and Fréchet variance of P. While the alternative space is the entire complement, our methods and power analysis aim at alternatives H_1 for which at least one of $\mu_1 \neq \mu_2$ or $V_1 \neq V_2$ applies, that is, alternatives that entail changes in mean or scale.

It is easy to see that, for real valued random variables, Fréchet means and variances are the same as the expectation and variance of the distributions P_1 and P_2. For $\mathbb{R}^d$ valued random

variables, the Fréchet mean is the ordinary mean vector of the distribution and the Fréchet variance corresponds to the trace of the covariance matrix which captures the total variability of the data. For more general spaces, for example, the space of networks or the space of probability distributions, Fréchet means provide a notion of center of the probability distribution generating the random data objects, and the Fréchet variance measures the spread of the random objects around the Fréchet mean. However, the Fréchet mean lives in the object space and, therefore, is not amenable to operations like addition and multiplication, while the Fréchet variance as a scalar is easier to handle. Classical analysis of variance can provide intuition on how to harness Fréchet variances to determine whether data segments differ or not, which is essential when testing for the presence of change-points.

Like other change-point detection techniques, we partition the data sequence into two segments and then maximize differences between these segments to develop inference and change-point estimation. Both segments need to contain a minimum number of observations in order to represent their Fréchet means and variance adequately; hence, we assume that τ lies in a compact interval $\mathcal{I}_c = [c, 1 - c] \subset [0, 1]$ for some $c > 0$, an assumption that is commonly adopted in the Euclidean case.

2.2. *Test statistic for a single change-point alternative.* Here, we describe how we construct the new test statistic for testing the null hypothesis H_0 (2.1) of no change-point vs. the alternative H_1 (2.2). Our approach is inspired by a recent two-sample test for detecting differences in Fréchet means or Fréchet variances of metric space valued data samples (Dubey and Müller (2019)). Each $u \in \mathcal{I}_c$ could be a possible value of τ under the alternative. For each $u \in \mathcal{I}_c$, we consider two data segments: A first segment consisting of all observations that are located before $[nu]$ and a second segment consisting of the observations located after $[nu]$.

Let $\hat{\mu}_{[0,u]}$ and $\hat{V}_{[0,u]}$ denote the estimated Fréchet mean and variance of all observations coming before $[nu]$,

$$\hat{\mu}_{[0,u]} = \underset{\omega \in \Omega}{\operatorname{argmin}} \frac{1}{[nu]} \sum_{i=1}^{[nu]} d^2(Y_i, \omega),$$

$$\hat{V}_{[0,u]} = \frac{1}{[nu]} \sum_{i=1}^{[nu]} d^2(Y_i, \hat{\mu}_{[0,u]}),$$

and, analogously, for observations coming after $[nu]$; set

$$\hat{\mu}_{[u,1]} = \underset{\omega \in \Omega}{\operatorname{argmin}} \frac{1}{(n - [nu])} \sum_{[nu]+1}^{n} d^2(Y_i, \omega),$$

$$\hat{V}_{[u,1]} = \underset{\omega \in \Omega}{\operatorname{argmin}} \frac{1}{(n - [nu])} \sum_{[nu]+1}^{n} d^2(Y_i, \hat{\mu}_{[u,1]}).$$

Next, we define "contaminated" versions of Fréchet variances of the data segments, obtained by plugging in the Fréchet mean from the complementary data segment,

$$(2.3) \qquad \hat{V}_{[0,u]}^C = \frac{1}{[nu]} \sum_{i=1}^{[nu]} d^2(Y_i, \hat{\mu}_{[u,1]}) \quad \text{and} \quad \hat{V}_{[u,1]}^C = \frac{1}{n - [nu]} \sum_{[nu]+1}^{n} d^2(Y_i, \hat{\mu}_{[0,u]}).$$

By definition, the contaminated Fréchet variances of the data segments are at least as large as the correct Fréchet variances.

Intuitively, under H_0 the population Fréchet means and variances of observations do not depend on u, and, therefore, we expect $\hat{\mu}_{[u,1]}$ to be close to $\hat{\mu}_{[0,u]}$, which in turn implies

that, under H_0, the differences $\hat{V}_{[0,u]}^C - \hat{V}_{[0,u]}$ and $\hat{V}_{[u,1]}^C - \hat{V}_{[u,1]}$ are small. Moreover, since the Fréchet variances of the two data segments are also the same under H_0, we expect the absolute difference $|\hat{V}_{[0,u]} - \hat{V}_{[u,1]}|$ to be small.

It is instructive to consider the special case $\Omega = \mathbb{R}$. Without loss of generality, assume $u > \tau$, and let $\bar{Y}_{0,\tau}$ denote the mean of all observations before $[n\tau]$, $\bar{Y}_{\tau,u}$ the mean of all observations coming after $[n\tau]$ but before $[nu]$ and $\bar{Y}_{u,1}$ the mean of all observations coming after $[nu]$. It is easy to see that

$$\hat{\mu}_{[0,u]} \approx \frac{\tau}{u}\bar{Y}_{0,\tau} + \frac{(u-\tau)}{u}\bar{Y}_{\tau,u} \quad \text{and} \quad \hat{\mu}_{[u,1]} = \bar{Y}_{u,1},$$

and a simple calculation leads to the approximation

$$\hat{V}_{[0,u]}^C - \hat{V}_{[0,u]} = \hat{V}_{[u,1]}^C - \hat{V}_{[u,1]} \approx \left(\frac{\tau}{u}\bar{Y}_{0,\tau} + \frac{(u-\tau)}{u}\bar{Y}_{\tau,u} - \bar{Y}_{u,1}\right)^2.$$

If $\mu_1 \neq \mu_2$, one expects that, for large n, $\bar{Y}_{0,\tau}$ is close to μ_1, and $\bar{Y}_{\tau,u}$ and $\bar{Y}_{u,1}$ are close to μ_2 which implies that $(\frac{\tau}{u}\bar{Y}_{0,\tau} + \frac{(u-\tau)}{u}\bar{Y}_{\tau,u} - \bar{Y}_{u,1})^2$ is close to $(\frac{\tau}{u}(\mu_1 - \mu_2))^2$. The latter is maximized when $u = \tau$. Here, one can view the terms $\hat{V}_{[0,u]}^C - \hat{V}_{[0,u]}$ and $\hat{V}_{[u,1]}^C - \hat{V}_{[u,1]}$ as the between group variance of the two data segments, the size of which is expected to reflect differences in the means.

Comparing the within group variances of the data segments, let $\hat{V}_{0,\tau}$ denote the variance of the observations located before $[n\tau]$, $\hat{V}_{\tau,u}$ the variance of the observations located after $[n\tau]$ but before $[nu]$ and $\hat{V}_{u,1}$ the variance of the observations located after $[nu]$. Again, by a simple calculation

$$\hat{V}_{[0,u]} \approx \frac{\tau}{u}\hat{V}_{0,\tau} + \frac{u-\tau}{u}\hat{V}_{\tau,u} + \frac{\tau}{u}\frac{u-\tau}{u}(\bar{Y}_{0,u} - \bar{Y}_{\tau,u})^2 \quad \text{and} \quad \hat{V}_{[u,1]} = \hat{V}_{u,1}.$$

If $\mu_1 = \mu_2$ and $V_1 \neq V_2$, then, for large n, $(\hat{V}_{[0,u]} - \hat{V}_{[u,1]})^2$ is approximately equal to $(\frac{\tau}{u}(V_1 - V_2))^2$ and is maximized at $u = \tau$. The term $(\hat{V}_{[0,u]} - \hat{V}_{[u,1]})^2$ can be related to between group variances of the data segments and is expected to account for variance differences between the two groups in the absence of a mean difference. It can be shown with a few steps of calculation that, when both $\mu_1 \neq \mu_2$ and $V_1 \neq V_2$, the term $(\hat{V}_{[0,u]} - \hat{V}_{[u,1]})^2 + (\hat{V}_{[0,u]}^C - \hat{V}_{[0,u]} + \hat{V}_{[u,1]}^C - \hat{V}_{[u,1]})^2$ is also maximized at $u = \tau$.

Now, considering the general case where the data sequence takes values in a general metric space, the term $(\hat{V}_{[0,u]}^C - \hat{V}_{[0,u]} + \hat{V}_{[u,1]}^C - \hat{V}_{[u,1]})^2$ reflects differences in Fréchet means of the data segments and the term $(\hat{V}_{[0,u]} - \hat{V}_{[u,1]})^2$ differences in Fréchet variances. For a fixed $u \in \mathcal{I}_c$, it can be shown that a central limit theorem for Fréchet variances (Dubey and Müller (2019)) implies that $\sqrt{u(1-u)}(\sqrt{n}/\sigma)(\hat{V}_{[0,u]} - \hat{V}_{[u,1]})$ has an asymptotic standard normal distribution under H_0, where σ is the asymptotic variance of the empirical Fréchet variance. Another key result, which holds under H_0 and for which we refer to Theorem 1 below and its proof in Appendix B in the online Supplementary Material (Dubey and Müller (2020)), is that, under some regularity assumptions and if $\mu_1 = \mu_2$,

$$\sup_{u \in \mathcal{I}_c} (\hat{V}_{[0,u]}^C - \hat{V}_{[0,u]} + \hat{V}_{[u,1]}^C - \hat{V}_{[u,1]})^2 = o_P(1/n).$$

We also require an estimate of σ^2, for which we use

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} d^4(Y_i, \hat{\mu}) - \hat{V}^2,$$

an estimator that is known to be consistent under H_0 (Dubey and Müller (2019)), where

$$(2.4) \qquad \hat{\mu} = \underset{\omega \in \Omega}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} d^2(Y_i, \omega) \quad \text{and} \quad \hat{V} = \frac{1}{n} \sum_{i=1}^{n} d^2(Y_i, \hat{\mu}).$$

The above considerations motivate the proposed statistics

$$(2.5) \qquad T_n(u) = \frac{u(1-u)}{\hat{\sigma}^2} \{ (\hat{V}_{[0,u]} - \hat{V}_{[u,1]})^2 + (\hat{V}_{[0,u]}^C - \hat{V}_{[0,u]} + \hat{V}_{[u,1]}^C - \hat{V}_{[u,1]})^2 \}$$

for $u \in \mathcal{I}_c$. We refer to $T_n(\cdot)$ when considered as a function of u as the scan function.

We will show in Theorem 1 in Section 3 that, under mild regularity conditions, $\{nT_n(u) : u \in \mathcal{I}_c\}$ converges weakly to the square of a tight standardized Brownian bridge on the interval $\mathcal{I}_c$ under H_0, which is a stochastic process given by

$$(2.6) \qquad \mathcal{G} = \left\{ \frac{\mathcal{B}(u)}{\sqrt{u(1-u)}} : u \in \mathcal{I}_c \right\},$$

where $\{\mathcal{B}(u) : u \in \mathcal{I}_c\}$ is a tight Brownian bridge on $\mathcal{I}_c$, that is, a tight Gaussian process indexed by $\mathcal{I}_c$ with zero mean and covariance structure given by $K(s, t) = \min(s, t) - st$. For the role of tightness in this connection, we refer to Section 2.1.2 of van der Vaart and Wellner (1996). Other versions of the Brownian bridge have surfaced in the context of the asymptotics of change-point detection in various previous studies (Chernoff and Zacks (1964), Csörgő and Horváth (1997), MacNeill (1974), Siegmund (1988)).

Specifically, for testing H_0 versus H_1, we use the statistic

$$(2.7) \qquad \sup_{u \in \mathcal{I}_c} nT_n(u) = \max_{[nc] \leq k \leq n-[nc]} nT_n\left(\frac{k}{n}\right).$$

Let $q_{1-\alpha}$ be the $(1 - \alpha)$th quantile of $\sup_{u \in \mathcal{I}_c} \mathcal{G}^2(u)$. Under H_0 and the regularity conditions described in Section 3, by Theorem 1 and the continuous mapping theorem, one obtains weak convergence (denoted here and in all of the following by $\Rightarrow$),

$$\sup_{u \in \mathcal{I}_c} nT_n(u) \Rightarrow \sup_{u \in \mathcal{I}_c} \mathcal{G}^2(u).$$

A level α significance test is then characterized by the rejection region

$$(2.8) \qquad R_{n,\alpha} = \left\{ \sup_{u \in \mathcal{I}_c} nT_n(u) > q_{1-\alpha} \right\}.$$

In Section 3.3 we describe how to obtain asymptotic and bootstrap critical values under H_0. We will show in Section 3 that, under H_1, when a change-point is present at $\tau \in \mathcal{I}_c$, the process $T_n(u)$ converges uniformly in probability to a limit process $T(u)$, which has a unique maximizer at $u = \tau$, that is,

$$\tau = \underset{u \in \mathcal{I}_c}{\operatorname{argmax}} T(u).$$

It is then natural to estimate the location of the change-point τ by

$$(2.9) \qquad \hat{\tau} = \underset{u \in \mathcal{I}_c}{\operatorname{argmax}} T_n(u) = \underset{[nc] \leq k \leq n-[nc]}{\operatorname{argmax}} T_n\left(\frac{k}{n}\right).$$

2.3. *Population limit of the test statistic under H_1.*   To obtain the limit $T(u)$ of the process $T_n(u)$ (2.5) under H_1 (2.2), we first derive the pointwise limits of the sample based estimators which form the components of $T_n(u)$. Starting with the Fréchet mean $\hat{\mu}_{[0,u]}$ (2.3), we observe that if $u \leq \tau$, $\hat{\mu}_{[0,u]}$ converges to μ_1 in probability pointwise in u, which is a consequence of results in Dubey and Müller (2019). If $u > \tau$, we have

$$\hat{\mu}_{[0,u]} = \operatorname*{argmin}_{\omega \in \Omega} \left\{ \frac{[n\tau]}{[nu]} \frac{1}{[n\tau]} \sum_{i=1}^{[n\tau]} d^2(Y_i, \omega) + \frac{([nu] - [n\tau])}{[nu]} \frac{1}{([nu] - [n\tau])} \sum_{[n\tau]+1}^{[nu]} d^2(Y_i, \omega) \right\},$$

which intuitively implies that the pointwise limit of $\hat{\mu}_{[0,u]}$ for $u > \tau$ is

$$\operatorname*{argmin}_{\omega \in \Omega} \left\{ \frac{\tau}{u} E_1(d^2(Y, \omega)) + \frac{u - \tau}{u} E_2(d^2(Y, \omega)) \right\}.$$

We will show that the pointwise limit $\mu_{[0,u]}$ of $\hat{\mu}_{[0,u]}$ is indeed

$$(2.10) \qquad \mu_{[0,u]} = \begin{cases} \mu_1 & u \leq \tau, \\ \operatorname*{argmin}_{\omega \in \Omega} \left\{ \dfrac{\tau}{u} E_1(d^2(Y, \omega)) + \dfrac{u - \tau}{u} E_2(d^2(Y, \omega)) \right\} & u > \tau \end{cases}$$

and that, likewise, the pointwise limit $\mu_{[u,1]}$ of $\hat{\mu}_{[u,1]}$ (2.3) is

$$\mu_{[u,1]} = \begin{cases} \operatorname*{argmin}_{\omega \in \Omega} \left\{ \dfrac{\tau - u}{1 - u} E_1(d^2(Y, \omega)) + \dfrac{1 - \tau}{1 - u} E_2(d^2(Y, \omega)) \right\} & u \leq \tau, \\ \mu_2 & u > \tau. \end{cases}$$

We denote the pointwise limits of the Fréchet variances $\hat{V}_{[0,u]}$ (2.3) and $\hat{V}_{[u,1]}$ (2.3) by $V_{[0,u]}$ and $V_{[u,1]}$, respectively. They are given by

$$V_{[0,u]} = \begin{cases} V_1 & u \leq \tau, \\ \dfrac{\tau}{u} E_1(d^2(Y, \mu_{[0,u]})) + \dfrac{u - \tau}{u} E_2(d^2(Y, \mu_{[0,u]})) & u > \tau; \end{cases}$$

$$V_{[u,1]} = \begin{cases} \dfrac{\tau - u}{1 - u} E_1(d^2(Y, \mu_{[u,1]})) + \dfrac{1 - \tau}{1 - u} E_2(d^2(Y, \mu_{[u,1]})) & u \leq \tau, \\ V_2, & u > \tau. \end{cases}$$

Similarly, the pointwise limits of the contaminated Fréchet variances $\hat{V}^C_{[0,u]}$ and $\hat{V}^C_{[u,1]}$ (2.3), denoted by $V^C_{[0,u]}$ and $V^C_{[0,1]}$, are given by

$$V^C_{[0,u]} = \begin{cases} E_1(d^2(Y, \mu_{[u,1]})) & u \leq \tau, \\ \dfrac{\tau}{u} E_1(d^2(Y, \mu_2)) + \dfrac{u - \tau}{u} E_2(d^2(Y, \mu_2)) & u > \tau; \end{cases}$$

$$V^C_{[u,1]} = \begin{cases} \dfrac{\tau - u}{1 - u} E_1(d^2(Y, \mu_1)) + \dfrac{1 - \tau}{1 - u} E_2(d^2(Y, \mu_1)) & u \leq \tau, \\ E_2(d^2(Y, \mu_{[0,u]})) & u > \tau. \end{cases}$$

Under H_1, the pooled sample Fréchet mean $\hat{\mu}$ (2.4) and Fréchet variance $\hat{V}$ have the pointwise limits

$$\tilde{\mu} = \operatorname*{argmin}_{\omega \in \Omega} \tau E_1(d^2(Y, \omega)) + (1 - \tau) E_2(d^2(Y, \omega));$$

$$\tilde{V} = \tau E_1(d^2(Y, \tilde{\mu})) + (1 - \tau) E_2(d^2(Y, \tilde{\mu})).$$

Key steps are that, under H_1, $\hat{\sigma}^2$ converges in probability to

$$\sigma^2 = \tau E_1(d^4(Y, \tilde{\mu})) + (1 - \tau)E_2(d^4(Y, \tilde{\mu})) - \tilde{V}^2,$$

while under H_0, where $P_1 = P_2$ and, therefore, $E_1 = E_2$, the limit is $\sigma^2 = E(d^4(Y, \tilde{\mu})) - \tilde{V}^2$ and that, under regularity conditions and H_1, $T_n(u)$ converges uniformly in probability to

$$(2.11) \qquad T(u) = \frac{u(1 - u)}{\sigma^2}\{(V_{[0,u]} - V_{[u,1]})^2 + (V_{[0,u]}^C - V_{[0,u]} + V_{[u,1]}^C - V_{[u,1]})^2\},$$

as stated in Theorem 2 below, with detailed arguments provided in the proof in Appendix B. At $u = \tau$ we have with

$$\Delta_1 = E_1(d^2(Y, \mu_2)) - E_1(d^2(Y, \mu_1)), \qquad \Delta_2 = E_2(d^2(Y, \mu_1)) - E_2(d^2(Y, \mu_2))$$

that

$$T(\tau) = \frac{\tau(1 - \tau)}{\sigma^2}\{(V_1 - V_2)^2 + (\Delta_1 + \Delta_2)^2\}.$$

Note that, under the assumption of uniqueness of Fréchet means of the two populations under H_1, $\Delta_1 \geq 0$ and $\Delta_1 = 0$ if and only if $\mu_1 = \mu_2$. Similarly, $\Delta_2 \geq 0$ and $\Delta_2 = 0$ if and only if $\mu_1 = \mu_2$. This implies that $T(\tau) = 0$ if and only if either $V_1 = V_2$ or $\mu_1 = \mu_2$. Hence, $T(\tau)$ quantifies the divergence between P_1 and P_2, where the term $(V_1 - V_2)^2$ accounts for scale differences and the term $(\Delta_1 + \Delta_2)^2$ accounts for location differences. Formally, we will show in Proposition 1 in Section 3 that, under H_1 and mild assumptions on existence and uniqueness of Fréchet means, $T(u)$ is maximized uniquely at $u = \tau$, that is,

$$T(\tau) = \sup_{u \in \mathcal{I}_c} T(u)$$

which is the justification for the proposed estimate of τ in (2.9).

3. Theory.

3.1. *Assumptions.*
For any $\alpha = \{\alpha_1, \alpha_2, \ldots, \alpha_n : 0 \leq \alpha_i < 1, \sum_{i=1}^{n} \alpha_i = 1\}$, let

$$(3.1) \qquad R_n(\omega, \alpha) = \sum_{i=1}^{n} \alpha_i d^2(Y_i, \omega) \quad \text{and} \quad \hat{\mu}_\alpha = \underset{\omega \in \Omega}{\operatorname{argmin}}\, R_n(\omega, \alpha).$$

Furthermore, for any $0 \leq \gamma \leq 1$, let

$$(3.2) \qquad S(\omega, \gamma) = \gamma E_1(d^2(Y, \omega)) + (1 - \gamma)E_2(d^2(Y, \omega)) \quad \text{and} \quad \mu_\gamma = \underset{\omega \in \Omega}{\operatorname{argmin}}\, S(\omega, \gamma).$$

We need the following assumptions, which are assumed to hold irrespective of whether H_0 is true, and we show that this is possible for various salient examples, as discussed below after the listing of the assumptions.

(A1) For any $0 \leq \gamma \leq 1$, μ_γ exists and is unique. Additionally, there exists $\zeta > 0$ and $C > 0$ such that

$$\inf_{0 \leq \gamma \leq 1} \inf_{d(\omega, \mu_\gamma) < \zeta} \{S(\omega, \gamma) - S(\mu_\gamma, \gamma) - Cd^2(\omega, \mu_\gamma)\} \geq 0.$$

Note that this implies that under H_0,

$$(3.3) \qquad \inf_{d(\omega, \mu) < \zeta} \{E(d^2(Y, \omega)) - E(d^2(Y, \mu)) - Cd^2(\omega, \mu)\} \geq 0.$$

(A2) For any $\alpha = \{\alpha_1, \alpha_2, \ldots, \alpha_n : 0 \le \alpha_i \le 1, \sum_{i=1}^n \alpha_i = 1\}$, $\hat{\mu}_\alpha$ exists and is unique almost surely. Additionally, for any $\varepsilon > 0$, there exists $\kappa_0 = \kappa_0(\varepsilon) > 0$ such that as $n \to \infty$,

$$P\left(\inf_\alpha \inf_{d(\omega, \hat{\mu}_\alpha) > \varepsilon} \{R_n(\omega, \alpha) - R_n(\hat{\mu}_\alpha, \alpha)\} \ge \kappa_0\right) \to 1.$$

(A3) Let $B_\delta(\omega) \subset \Omega$ be a ball of radius δ centered at ω, which is any arbitrary element in Ω, and let $N(\varepsilon, B_\delta(\omega), d)$ be its covering number which is defined as the minimum number of balls of radius $\varepsilon > 0$ needed to cover $B_\delta(\omega)$ (see Section 2.1.1 of van der Vaart and Wellner (1996) for the definition and further details). Then, for any $\omega \in \Omega$,

$$\int_0^1 \sqrt{\log N(\varepsilon\delta, B_\delta(\omega), d)}\, d\varepsilon = O(1) \quad \text{as } \delta \to 0.$$

Moreover, there exist constants $K > 0$ and $0 < \beta < 2$ such that the covering number of Ω satisfies that for any $\varepsilon > 0$,

$$(3.4) \qquad\qquad\qquad\qquad \log N(\varepsilon, \Omega, d) \le \frac{K}{\varepsilon^\beta}.$$

(A4) There exist $\delta > 0$ and $C > 0$ such that for all $\omega \in B_\delta(\mu_j)$ one has

$$E_1(d^2(Y, \omega)) - E_1(d^2(Y, \mu_1)) = Cd^2(\omega, \mu_1) + O(\delta^2) \quad \text{as } \delta \to 0,$$

and, analogously, for all $\omega \in B_\delta(\mu_2)$,

$$E_2(d^2(Y, \omega)) - E_2(d^2(Y, \mu_2)) = Cd^2(\omega, \mu_2) + O(\delta^2) \quad \text{as } \delta \to 0.$$

We note that the existence *and uniqueness* of the minimizers in (A1) and (A2) is guaranteed for the case of Hadamard spaces; the curvature of which is bounded above by zero (Sturm (2003, 2006)). However, for positively curved manifolds, such as spheres, the existence of minimizers depends intrinsically on the probability distribution on the space and is not guaranteed; see Ahidar-Coutrix, Gouic and Paris (2018) where the authors study conditions on probability distributions in positively curved metric spaces which imply existence and uniqueness of Fréchet means along with the condition in (3.3). Therefore, a case by case analysis is inevitable to ascertain whether these assumptions are satisfied. In the following we provide examples of spaces that satisfy assumptions (A1)–(A4) and that we use in simulations and real data applications; see, also, Petersen and Müller (2019b):

1. Let Ω be the set of univariate probability distributions F defined on $\mathbb{R}$ such that $\int_\mathbb{R} x^2 \, dF(x) \le M$ for all $F \in \Omega$ and some $M > 0$ equipped with the 2-Wasserstein metric d_W. For two univariate probability distributions with quantile functions $G_1(\cdot)$ and $G_2(\cdot)$, the metric d_W is

$$(3.5) \qquad\qquad\qquad d_W^2(G_1, G_2) = \int_0^1 (G_1(t) - G_2(t))^2 \, dt.$$

2. Let Ω be the set of covariance matrices of a fixed dimension r, with bounded variances (uniformly bounded diagonals) equipped with the Frobenius metric d_F. For two $r \times r$ matrices Σ_1 and Σ_2,

$$(3.6) \qquad\qquad\qquad d_F^2(\Sigma_1, \Sigma_2) = \text{trace}(\Sigma_1 - \Sigma_2)'(\Sigma_1 - \Sigma_2).$$

3. Let Ω be the space of network adjacency matrices or graph Laplacians of a fixed dimension r of undirected weighted graphs with bounded edge weights equipped with the Frobenius metric d_F. The space of graph Laplacians of simple graphs can be used to characterize the space of networks (Ginestet et al. (2017)).

More details on these spaces and proofs that they satisfy (A1)–(A4) are provided in Appendix C in the Supplementary Material, specifically in Propositions C.1 and C.2. Assumptions (A1) and (A2) are commonly used to establish consistency of M-estimators, such as $\hat{\mu}_{[0,u]}$ and $\hat{\mu}_{[u,1]}$, for $u \in \mathcal{I}_c$; see Chapter 3.2 in van der Vaart and Wellner (1996). In particular, one needs to ensure weak convergence of processes $\frac{1}{[nu]} \sum_{i=1}^{[nu]} d^2(Y_i, \omega)$ and $\frac{1}{n-[nu]} \sum_{[nu]+1}^{n} d^2(Y_i, \omega)$ to their population counterparts under H_0 (2.1) and H_1 (2.2) in order to obtain convergence of their minimizers.

Assumption (A3) is a restriction on the metric entropy of Ω and δ-balls in Ω which controls the complexity of Ω; in our case it controls uniform bracketing entropy integrals that are used to provide tail bounds as in Theorem 2.14.2 in van der Vaart and Wellner (1996). It is common to assume that the uniform bracketing entropy integral is bounded; for more details see Section 3.2.2 in van der Vaart and Wellner (1996). The condition in (3.4) is satisfied by a wide class of metric spaces, including the three examples discussed above, where $\beta = 1$ for the space of distributions with the 2-Wasserstein metric, since it can be represented as a special case of the space of monotone functions from $\mathbb{R}$ to a compact subset of $\mathbb{R}$ (van der Vaart and Wellner (1996)). For the other two examples, the condition in (3.4) is satisfied for any $\beta > 0$ since these spaces are subsets of a bounded Euclidean space. Other examples include the space of all Lipschitz functions of degree $1/2 < \gamma \le 1$ on the unit interval $[0, 1]$ with the L^2 metric, where $\beta = \frac{1}{\gamma}$, as well as the class of bounded convex functions on a compact, convex subset of $\mathbb{R}^d$ under certain restrictions, where $\beta = d/2$ (Guntuboyina and Sen (2013)).

Assumption (A4) appears in empirical process theory when one deals with approximate M-estimators (Arcones (1998)). It ensures that, under H_0, the contaminated Fréchet variances $\hat{V}_{[0,u]}^C$ and $\hat{V}_{[u,1]}^C$ are sufficiently close to the correct Fréchet variances $\hat{V}_{[0,u]}$ and $\hat{V}_{[u,1]}$.

For the special case of $\mathbb{R}^d$ with the Euclidean metric, assumptions (A1)–(A4) are satisfied for bounded convex subsets of $\mathbb{R}^d$. The proof follows the exact line of arguments as in the proof of Proposition C.2 in Appendix C of the online Supplementary Material. The required entropy condition (3.4) in assumption (A3) and some of the other proof techniques do not continue to hold for unbounded subsets of $\mathbb{R}^d$ under the Euclidean metric. Extensions to unbounded convex subsets of $\mathbb{R}$ might be possible in special cases. For nonconvex subsets of $\mathbb{R}^d$, existence and uniqueness of Fréchet means is not guaranteed.

3.2. *Main results.*

In this section we state the main results, including the asymptotic limit distribution of the proposed test statistic under H_0 (2.1), consistency of the test under contiguous alternatives and consistency of the estimated location of the change-point. Detailed proofs for all results can be found in the Appendix in the online Supplementary Material. We start with consistency of the estimated Fréchet means and variances to their population targets under H_0 using tools from Montgomery-Smith and Pruss (2001) and van der Vaart and Wellner (1996). Note that, under H_0, $\mu = \mu_1 = \mu_2$ and $V = V_1 = V_2$.

LEMMA 1. *Under H_0 (2.1) and assumptions* (A1)–(A3),

$$\sqrt{n} \sup_{u \in \mathcal{I}_c} d(\hat{\mu}_{[0,u]}, \mu) = O_P(1), \qquad \sqrt{n} \sup_{u \in \mathcal{I}_c} d(\hat{\mu}_{[u,1]}, \mu) = O_P(1)$$

and

$$|\hat{\sigma}^2 - \sigma^2| = o_P(1).$$

The proof of this and of most of the following results uses empirical process theory. To obtain the first statement of the lemma, we introduce processes

$$M_n(\omega, u) = \frac{1}{[nu]} \sum_{i=1}^{[nu]} d^2(Y_i, \omega) \quad \text{and} \quad M(\omega) = E(d^2(Y, \omega))$$

and show the weak convergence of the process $\{M_n(\omega, u) - M(\omega, u) : \omega \in \Omega\}$ to zero and the asymptotic equicontinuity of the processes $\{d(\hat{\mu}_{[0,u]}, \mu)\}_{u \in \mathcal{I}_c}$, which then implies uniform convergence. To obtain the rate of uniform convergence, we find deviation bounds over slices $S_{j,n} = \{\omega \in \Omega : 2^j \leq \sqrt{n}d(\omega, \mu) < 2^{j+1}\}$. Details are in the Supplementary Material.

Observe that Fréchet variances are sums of dependent random variables, because each summand contains the Fréchet mean, which is estimated using all other data objects. A key step in obtaining the limiting distribution of the test statistic in (2.7) under H_0 is to replace the estimated Fréchet means in the estimated Fréchet variances with the true mean μ under H_0, for which the following lemma is instrumental. Let

$$\tilde{V}_{[0,u]} = \frac{1}{[nu]} \sum_{i=1}^{[nu]} d^2(Y_i, \mu) \quad \text{and} \quad \tilde{V}_{[u,1]} = \frac{1}{n - [nu]} \sum_{i=[nu]+1}^{n} d^2(Y_i, \mu),$$

where μ is the true Fréchet mean of the data sequence under H_0. Note that $\tilde{V}_{[0,u]}$ and $\tilde{V}_{[u,1]}$ are sums of independent random variables and can be thought of as oracle versions of $\hat{V}_{[0,u]}$ and $\hat{V}_{[u,1]}$ with the true Fréchet means plugged in.

LEMMA 2. *Under H_0* (2.1) *and assumptions* (A1)–(A3),

$$\sup_{u \in \mathcal{I}_c} \sqrt{n} |\hat{V}_{[0,u]} - \tilde{V}_{[0,u]}| = o_P(1) \quad \text{and} \quad \sup_{u \in \mathcal{I}_c} \sqrt{n} |\hat{V}_{[u,1]} - \tilde{V}_{[u,1]}| = o_P(1).$$

The following result states that the contaminated Fréchet variances of the data segments are close to the correct Fréchet variances of the data segments under H_0.

LEMMA 3. *Under H_0* (2.1) *and assumptions* (A1)–(A4),

$$\sup_{u \in \mathcal{I}_c} \sqrt{n} |\hat{V}_{[0,u]}^C - \hat{V}_{[0,u]}| = o_P(1) \quad \text{and} \quad \sup_{u \in \mathcal{I}_c} \sqrt{n} |\hat{V}_{[u,1]}^C - \hat{V}_{[u,1]}| = o_P(1).$$

Recalling that $\mathcal{G} = \{\mathcal{B}(u)/\sqrt{u(1-u)} : u \in \mathcal{I}_c\}$ is a standardized Brownian bridge process on $\mathcal{I}_c$, where $\{\mathcal{B}(u) : u \in \mathcal{I}_c\}$ is a tight Brownian bridge process on $\mathcal{I}_c$, Theorem 1 gives the weak convergence of the scan function $T_n(\cdot)$ in the absence of a change-point and provides the asymptotic justification of the proposed test. The weak convergence is in $l^\infty([c, 1-c])$, the set of all uniformly bounded real functions on $[c, 1-c]$.

THEOREM 1. *Under H_0* (2.1) *and assumptions* (A1)–(A4),

$$\{nT_n(u) : u \in \mathcal{I}_c\} \Rightarrow \{\mathcal{G}^2(u) : u \in \mathcal{I}_c\},$$

where the covariance function $C(u, v)$ of $\mathcal{G}$ is $C(u, v) = [u(1-v)/v(1-u)]^{1/2}$ for $u \leq v$.

For the proof of this key result, we decompose $T_n(u) = T_n^I(u) + T_n^{II}(u)$, where

$$T_n^I(u) = \frac{(\hat{V}_{[0,u]} - \hat{V}_{[u,1]})^2}{\hat{\sigma}^2(1/u + 1/(1-u))}$$

and

$$T_n^{\mathrm{II}}(u) = \frac{(\hat{V}_{[0,u]}^C - \hat{V}_{[0,u]}) + \hat{V}_{[u,1]}^C - \hat{V}_{[u,1]})^2}{\hat{\sigma}^2(1/u + 1/(1-u))}.$$

We then show that $nT_n^I = \{nT_n^I(u) : u \in \mathcal{I}_c\}$ converges weakly to $\mathcal{G}^2$ and $nT_n^{\mathrm{II}} = \{nT_n^{\mathrm{II}}(u) : u \in \mathcal{I}_c\}$ converges weakly to zero. Once the weak convergence of nT_n is established, the continuous mapping theorem implies

$$\max_{u \in \mathcal{I}_c} nT_n(u) \Rightarrow \max_{u \in \mathcal{I}_c} \mathcal{G}^2(u).$$

Next, we move away from assuming the null hypothesis H_0 (2.1) of no change-point and establish the convergence of $\hat{\mu}_{[0,u]}$, $\hat{\mu}_{[u,1]}$ and $\hat{\sigma}^2$ in Lemma 4 and the convergence of the correct and the contaminated Fréchet variances of the data segments to their population targets in Lemma 5, using results from Ossiander (1987). Note that many of the following results are only valid under H_1, as the change-point location τ is defined only under H_1 but not under H_0. Two relevant results are as follows:

LEMMA 4. *Under assumptions* (A1)–(A3) *and with β as in assumption* (A3),

$$\sup_{u \in \mathcal{I}_c} d(\hat{\mu}_{[0,u]}, \mu_{[0,u]}) = O_P(n^{-\frac{1}{2+\beta}}), \qquad \sup_{u \in \mathcal{I}_c} d(\hat{\mu}_{[u,1]}, \mu_{[u,1]}) = O_P(n^{-\frac{1}{2+\beta}})$$

and

$$|\hat{\sigma}^2 - \sigma^2| = o_P(1).$$

LEMMA 5. *Under assumptions* (A1)–(A3) *and with β as in assumption* (A3),

$$(3.7) \quad \sup_{u \in \mathcal{I}_c} |\hat{V}_{[0,u]} - V_{[0,u]}| = O_P\left(\frac{1}{\sqrt{n}}\right) \quad and \quad \sup_{u \in \mathcal{I}_c} |\hat{V}_{[u,1]} - V_{[u,1]}| = O_P\left(\frac{1}{\sqrt{n}}\right),$$

$$(3.8) \quad \sup_{u \in \mathcal{I}_c} |\hat{V}_{[0,u]}^C - V_{[0,u]}^C| = O_P(n^{-\frac{1}{2+\beta}}) \quad and \quad \sup_{u \in \mathcal{I}_c} |\hat{V}_{[u,1]}^C - V_{[u,1]}^C| = O_P(n^{-\frac{1}{2+\beta}}).$$

Using these results, we show in the following proposition that the limit process $T(u)$ (2.11) of $T_n(u)$ (2.5) indeed has its maximum at $u = \tau$.

PROPOSITION 1. *Assume μ_γ (3.2) exists and is unique for any $0 \le \gamma \le 1$. Then, one has under H_1 that*

$$T(\tau) = \sup_{u \in \mathcal{I}_c} T(u)$$

and $T(\tau) > T(u)$ when $u \neq \tau$.

Consistency of the estimated location of the change-point is a consequence of the uniform convergence of the process $T_n(u)$ to the limit process $T(u)$ under H_1 and the argmax theorem which guarantees convergence of the maximizers.

THEOREM 2. *Under H_1 (2.2) and assumptions* (A1)–(A3),

$$\sup_{u \in \mathcal{I}_c} |T_n(u) - T(u)| = o_P(1) \quad and \quad |\hat{\tau} - \tau| = o_P(1).$$

In order to show consistency of the test (2.8) in a neighborhood of H_0, we construct a sequence of local alternatives

$$(3.9) \qquad H_{1,n} = \{d^2(\mu_1, \mu_2) = a_n, |V_1 - V_2| = b_n : a_n > 0 \text{ or } b_n > 0\}.$$

When $a_n \to 0$ and $b_n \to 0$ as $n \to \infty$, $\{H_{1,n}\}_{n \geq 1}$ is a sequence of contiguous alternatives. Let ψ_n be the power function for the test (2.8) when $H_{1,n}$ holds, that is,

$$(3.10) \qquad \psi_n = P(R_{n,\alpha}),$$

where $R_{n,\alpha}$, as defined in equation (2.8), is the rejection region for a level α test. The following result demonstrates the consistency of the test under contiguous alternatives $H_{1,n}$:

THEOREM 3. *Under H_1 (2.2) and assumptions* (A1)–(A3), *it holds for sequences $a_n \to 0$ and $b_n \to 0$, as in* (3.9), *that*

$$\psi_n \to 1$$

as $n \to \infty$, if either $\sqrt{n}a_n \to \infty$ or $\sqrt{n}b_n \to \infty$.

3.3. *Approximation of critical values by bootstrap.* For obtaining a level α asymptotic test as defined in (2.8), one needs to approximate the critical value $q_{1-\alpha}$. This can be done in practice by taking a large number of Monte Carlo simulations of $\mathcal{G}^2(\cdot)$ on the interval $[[nc], [n(1-c)]]$, computing $\sup_{[nc] \leq k \leq [n(1-c)]} \mathcal{G}^2(\frac{k}{n})$ in each simulation and then finding the $(1-\alpha)$th quantile across all simulations.

The testing procedure described above is based on the weak convergence of the maximum of a scan function to the maximum of a squared standardized Brownian bridge as established in Theorem 1. It is well known that the speed of convergence in limit theorems of this type is often slow, and the problem is magnified when the dimension of the data is moderate to high. Consequently, the approximation of the level of the test may not be accurate for finite sample sizes. This is borne out in simulations that we report in Section 4. In such situations it may be preferable to obtain critical values using a bootstrap approximation to ensure the accuracy of the test. We find that the bootstrap test, implemented according to the following specifications, tends to have larger critical values, compared with the asymptotic test, and is thus more conservative. While it typically has a level that is very close to the nominal level, this also means that the bootstrap test also often has lower power against alternatives, especially those that are close to the null.

A scheme for approximating $q_{1-\alpha}$ with the bootstrap distribution of the test statistic conditional on the observations instead of the asymptotic distribution is as follows:

Bootstrap Scheme:

Step I. Obtain a random sample of size m, $Y_1^\star, Y_2^\star, \ldots, Y_m^\star$ with replacement from the observations $Y_1, Y_2, \ldots, Y_n$.

Step II. For $u \in \mathcal{I}_c$, define the bootstrap quantities

$$\mu_{[0,u]}^\star = \operatorname*{argmin}_{\omega \in \Omega} \frac{1}{[mu]} \sum_{j=1}^{[mu]} d^2(Y_j^\star, \omega), \qquad \mu_{[u,1]}^\star = \operatorname*{argmin}_{\omega \in \Omega} \frac{1}{(m-[mu])} \sum_{j=[mu]+1}^{m} d^2(Y_j^\star, \omega),$$

$$V_{[0,u]}^\star = \frac{1}{[mu]} \sum_{j=1}^{[mu]} d^2(Y_j^\star, \mu_{[0,u]}^\star), \qquad V_{[u,1]}^\star = \frac{1}{(m-[mu])} \sum_{j=[mu]+1}^{m} d^2(Y_j^\star, \mu_{[u,1]}^\star),$$

$$V_{[0,u]}^{C,\star} = \frac{1}{[mu]} \sum_{j=1}^{[mu]} d^2(Y_j^\star, \mu_{[u,1]}^\star), \qquad V_{[u,1]}^{C,\star} = \frac{1}{(m-[mu])} \sum_{j=[mu]+1}^{m} d^2(Y_j^\star, \mu_{[0,u]}^\star),$$

$$\mu^\star = \operatorname*{argmin}_{\omega \in \Omega} \frac{1}{m} \sum_{j=1}^{m} d^2(Y_j^\star, \omega), \qquad (\sigma^\star)^2 = \frac{1}{m} \sum_{j=1}^{m} d^4(Y_j^\star, \mu^\star) - \left(\frac{1}{m} \sum_{j=1}^{m} d^2(Y_j^\star, \mu^\star)\right)^2$$

and

$$T_{m,n}^{\star}(u) = \frac{u(1-u)}{(\sigma^{\star})^2}\{(V_{[0,u]}^{\star} - V_{[u,1]}^{\star})^2 + (V_{[0,u]}^{C,\star} - V_{[0,u]}^{\star} + V_{[u,1]}^{C,\star} - V_{[u,1]}^{\star})^2\}.$$

Step III. Calculate $\tilde{T}_{m,n}^{\star} = \sup_{u \in \mathcal{I}_c} m T_{m,n}^{\star}(u) = \max_{[mc] \leq k \leq [m(1-c)]} m T_{m,n}^{\star}(\frac{k}{m})$.

We iterate Steps I–III B times. For each iteration, indexed by b, we obtain a bootstrap version $\tilde{T}_{m,n}^{\star,b}$ of $\tilde{T}_{m,n}^{\star}$ for $b = 1, 2, \ldots, B$. The following result shows that, under very mild assumptions, the distribution of $\sup_{u \in \mathcal{I}_c} n T_n(u)$ under H_0 can be approximated in probability by the conditional distribution of $\sup_{u \in \mathcal{I}_c} m T_{m,n}^{\star}(u)$, given the observations $Y_1, Y_2, \ldots, Y_n$. This implies that the conditional distribution of the bootstrap statistics is consistent under H_0. We can then use Monte Carlo approximations as described in Steps I–III for approximating the conditional distribution by the empirical distribution of $\tilde{T}_{m,n}^{\star,b}$, $b = 1, 2, \ldots, B$, where $q_{1-\alpha}$ is approximated as the empirical quantile of $\tilde{T}_{m,n}^{\star,b}$, $b = 1, 2, \ldots, B$.

Let $Y^{\star} \sim P_{|Y}$, which denotes the measure generated by resampling from $Y_1, Y_2, \ldots, Y_n$ uniformly with replacement. The following Theorem 4 provides the asymptotic consistency of the bootstrap distribution of the test statistic. Its proof in the Supplementary Material makes use of key results of Jirak (2015):

THEOREM 4.   *Under H_0 and assumptions* (A1)–(A4), *as $m \to \infty$ and $n \to \infty$,*

(3.11) $$\sup_x \left| P_{|Y}\left(\sup_{u \in \mathcal{I}_c} m T_{m,n}^{\star}(u) \leq x\right) - P\left(\sup_{u \in \mathcal{I}_c} \mathcal{G}^2(u) \leq x\right)\right| = o_P(1).$$

We note that, in principle, one could also use analytic approximations of the asymptotic distribution of the test statistics under the null hypothesis, which has some tradition in change-point analysis, with recent examples provided in Chen and Zhang (2015), Chu and Chen (2019). However, given that the bootstrap approximation can be theoretically justified and is seen to work well in practice, such approximations are less relevant for our purposes.

4. Simulations.   In order to study the power of the test (2.8) and the accuracy of the estimated change-point location, we report here the results of simulations for various settings. The random objects we consider in simulations are univariate probability distributions equipped with the 2-Wasserstein metric, graph Laplacians of scale free networks from the Barabási–Albert model (Barabási and Albert (1999)) with the Frobenius metric and multivariate data with the usual Euclidean metric.

It is usually harder to detect change-points that are located close to the endpoints of a data sequence. Each data sequence in the simulation is of length 300 and is generated such that the first and the second segments contain 100 and 200 data objects, respectively, under the alternative where a change-point is present. Hence, the change-point is placed at one-third of the data sequence under the alternative, that is, $\tau = 1/3$ under H_1 (2.2). We select the interval $\mathcal{I}_c$ as $[0.1, 0.9]$.

We construct power functions of the proposed test with certain parameters that we use to generate the data as argument, and we also quantify the accuracy of the estimated change-point location. We compare the results for the new test with the generalized edge count scan function in Chu and Chen (2019) which was shown in this previous work to dominate other graph based change-point detection approaches. For the graph based test we used these authors' implementation in the R package *gSeg* and constructed similarity graphs of the data sequences by constructing a 5-MST (minimal spanning tree) graph from the pooled pairwise distance matrix, following the suggestion in Chen and Friedman (2017). Here, a k-MST is the union of the 1st, . . . , kth MSTs, where a kth MST is a spanning tree connecting all observations that minimizes the sum of distances across edges, subject to the constraint that this spanning tree does not contain any edge in the 1st, . . . , $(k-1)$th MST.

In the simulations we explored not only location differences but also differences in shape and scale of the population segments. We carried out all tests at level $\alpha = 0.05$. Under both H_0 and H_1, the power of the tests was computed by finding the proportion of rejections from 1000 runs of the simulation. The critical value for rejection of the test (2.8) was obtained from the bootstrap scheme as described in Section 3.3. We quantified the accuracy of the estimated change-point location by mean absolute deviation (MAE) which was computed as follows: Denoting by $\hat{\tau}_i$ the estimated change-point in the ith simulation run,

$$\text{MAE} = \frac{1}{1000} \sum_{i=1}^{1000} |\hat{\tau}_i - \tau|,$$

where lower values of MAE indicate greater accuracy in the estimate of the change-point.

The first type of data objects we study are random samples of univariate probability distributions. Each datum is a $N(\mu, 1)$ distribution, where μ is random. As distance between two probability distributions, we choose the 2-Wasserstein metric. For investigating location differences, for the first data segment we generated μ as a truncated normal distribution $N(\delta, 0.75)$, constrained to lie in $[-10, 10]$ and for the second data segment as $N(0, 0.75)$, truncated within $[-10, 10]$ and then computed the empirical power function of the test and MAE for the estimated change point for $0 \leq \delta \leq 1$, where $\delta = 0$ represents H_0 (2.1). For investigating scale differences, μ was drawn randomly from $N(0, \delta)$ for the first data segment and from $N(0, 1)$ for the second data segment, truncated within $[-10, 10]$ in both cases; empirical power and MAE for the estimated change point were evaluated for $0.4 \leq \delta \leq 1$, where in this case $\delta = 1$ represents H_0. The results are presented in Figures 1 and 2. It is seen that the proposed test outperforms the graph based test in both cases, both in terms of power and accuracy of change-point detection.

Next, we consider sequences where the data objects are graph Laplacians of scale free networks from the Barabási–Albert model with the Frobenius metric. These popular networks have power law degree distributions and are commonly used for networks related to the world wide web, social networks and brain connectivity networks. For scale-free networks the fraction $P(c)$ of nodes in the network having c connections to other nodes for large values of c is approximately $c^{-\gamma}$, with γ typically in the range $1 \leq \gamma \leq 3$. Specifically, we used the
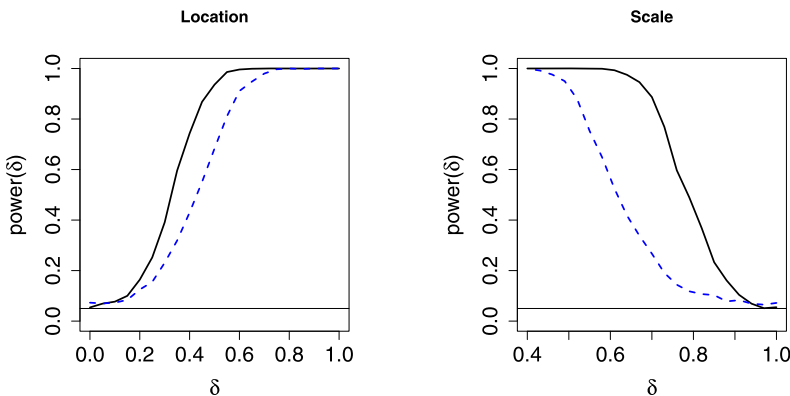

FIG. 1. *Empirical power as function of δ for $N(\mu, 1)$ probability distributions with μ generated from a truncated $N(\delta, 0.75)$ distributions for observations before $\tau = 1/3$ and truncated $N(0, 0.75)$ distributions for observations after τ (left panel) and empirical power as function of δ for $N(\mu, 1)$ probability distributions with μ generated from truncated $N(0, \delta)$ for observations before τ and from truncated $N(0, 1)$ distributions for observations after τ (right panel). The solid black curves correspond to the proposed test (2.8) and the dashed blue curves to the generalized scan function graph based test (Chu and Chen (2019)). The line parallel to the x axis indicates the level of the tests, which is 0.05.*
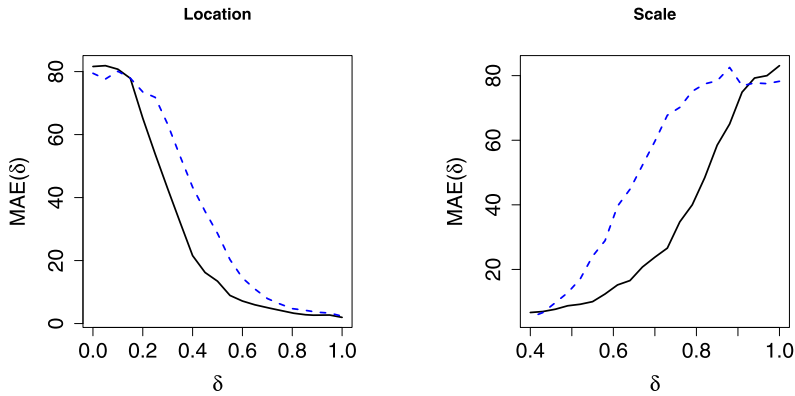
FIG. 2.   *MAE as function of δ for $N(\mu, 1)$ probability distributions, which are generated as described in Figure* 1, *with Fréchet mean change (left panel) and Fréchet variance change (right panel). The solid black curves correspond to the new estimates* (2.9) *and the dashed blue curves to those proposed in* Chu and Chen (2019).

Barabási–Albert algorithm to generate samples of scale-free networks with 10 nodes. For the first segment of 100 observations, we set $\gamma = 3$, and for the second segment of 200 observations we varied γ in the interval $1 \le \gamma \le 3$ with $\gamma = 3$ representing H_0. We computed the empirical power and MAE as a function of γ. The left panel in Figure 3 indicates that in this scenario the proposed test has better power behavior than the graph-based test. The graph-based test has a high false positive rate near and at H_0. The right panel in Figure 3 shows that, in terms of accuracy of the estimated change-point, the proposed test in most parts of the alternative outperforms the graph-based test.

For the multivariate case we assume a Gaussian setting. We consider $\Omega = [-10, 10]^{50}$. Let $\tilde{\delta} = (\delta, \delta, \delta, 0, 0, \ldots, 0)$ be the random vector whose first three components are δ; the remaining components are all 0. Let I_d be the $d \times d$ identity matrix and J_d be a a $d \times d$ matrix with all entries equal to one. For location alternatives we generated the first segment of the data sequence from $N(0, I_{50})$, truncated to lie in $[-10, 10]^{50}$, and the second segment from $N(\tilde{\delta}, I_{50})$. To study the power of the test and MAE of change-point estimates, we varied δ between $0 \le \delta \le 1$, where $\delta = 0$ represents H_0. For scale alternatives we generated the first segment of the data sequence from $N(0, \delta I_{50})$, truncated to lie in $[-10, 10]^{50}$, and the second segment from $N(0, I_{50})$ and varied δ between $0.75 \le \delta \le 1$, where $\delta = 1$ represents H_0.
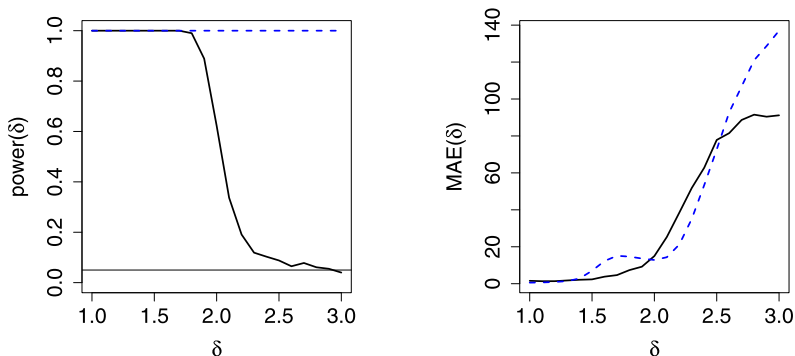
FIG. 3.   *Empirical power (left panel) and MAE (right panel) as a function of γ for scale-free networks from the Barabási–Albert model, with $\gamma = 3$ for the first data segment and $1 \le \gamma \le 3$ for the second data segment. The solid black curve corresponds to the proposed approach* (2.8) (2.9) *and the blue dashed curve to the graph based approach.*
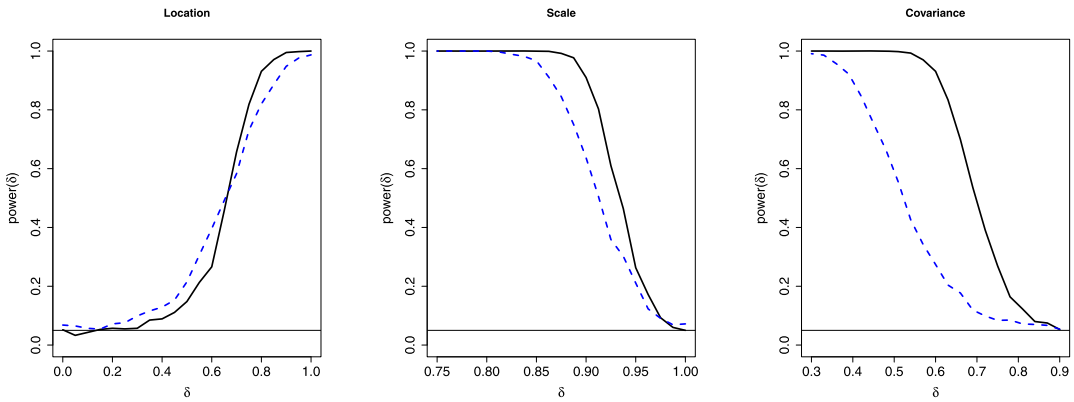
FIG. 4. *Empirical power as function of δ: (a) when the data segment is generated from truncated $N(0, I_{50})$ for observations before $\tau = 1/3$ and truncated $N(\tilde{\delta}, I_{50})$ for observations after τ (left panel); (b) for truncated $N(0, \delta I_{50})$ for observations before $\tau = 1/3$ and truncated $N(0, I_{50})$ for observations after τ (middle panel); (c) for $N(0, 0.9I_{50} + \delta^2 J_{50})$ for observations before $\tau = 1/3$ and truncated $N(0, 0.9I_{50} + 0.9^2 J_{50})$ for observations after τ (right panel). The solid black curves correspond to the proposed test (2.8) and the dashed blue curves to the graph based test.*

For detecting change-points in data correlation, we generated the first segment of the data sequence from $N(0, 0.9I_{50} + \delta^2 J_{50})$, truncated to lie in $[-10, 10]^{50}$, and the second segment from $N(0, 0.9I_{50} + 0.9^2 J_{50})$ and varied δ as $0.3 \leq \delta \leq 0.9$ for studying power and MAE, where $\delta = 0.9$ represents H_0. Figure 4 illustrates that in terms of power performance the proposed approach outperforms the graph-based approach for scale alternatives and has similar performance for location alternatives.

From Figure 5 we find that, for location alternatives, the proposed method has larger MAE closer to H_0, but when moving away from H_0, the proposed method has lower MAE than the graph based-approach. For scale alternatives the proposed method outperforms the graph-based approach. We also present a scenario in the multivariate Gaussian setting where the change in the vectors at the change-point is not reflected in mean or shape changes, and the proposed test, as expected, has no power and is outperformed by the graph-based test which still performs reasonably well. We provide the details in Appendix D in the online Supplementary Material. Interestingly, when choosing a different metric where the change is reflected in scale change, both tests perform equally well.
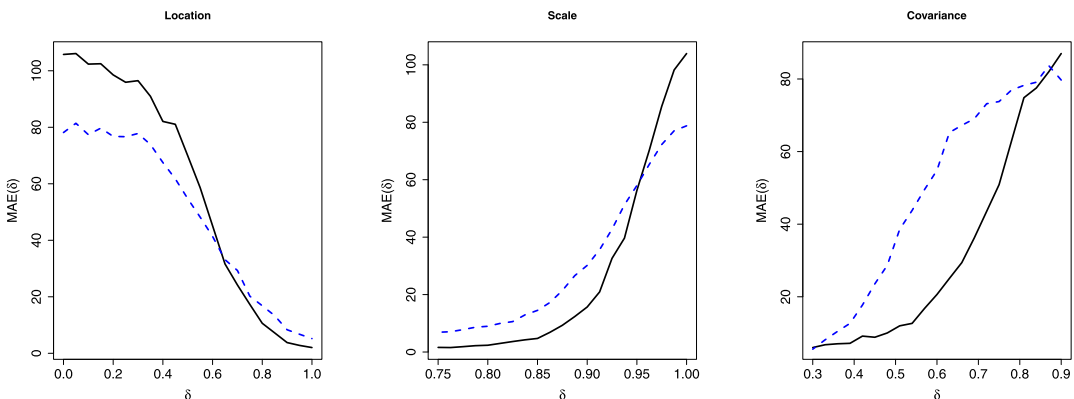

FIG. 5. *MAE as function of δ when the data are generated from the settings (a) (left panel), (b) (middle panel) and (c) (right panel) as described in Figure 4. In each panel the solid black curve corresponds to the proposed estimates (2.9) and the dashed blue curve to those for the graph based method.*
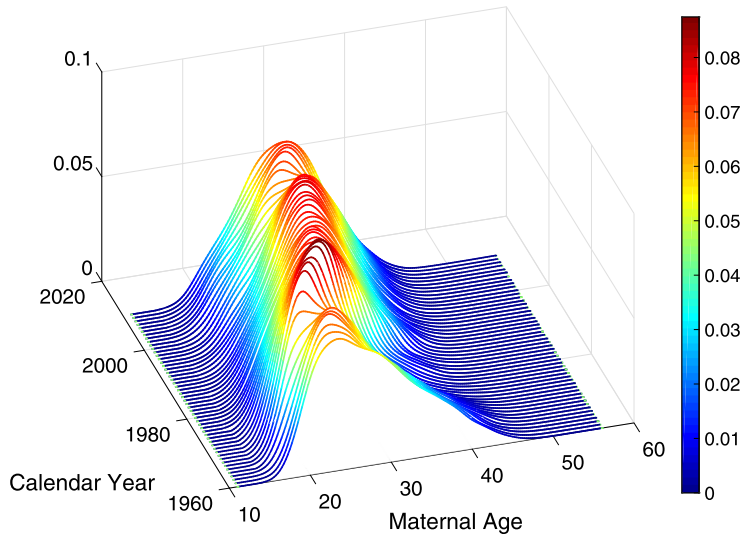
FIG. 6. *Yearly maternal age distributions represented as density functions for the age interval* [12, 55] *during the time period* 1960 *to* 2010 *for the Finnish fertility data.*

5. Data examples.

5.1. *Finnish fertility data.* The Human Fertility Database provides cohort fertility data for various countries and calendar years which are available at www.humanfertility.org. These data facilitate research on the evolution and intercountry differences in fertility over a period spanning more than 30 calendar years. For any country and year, the raw data consist of age-specific total live birth counts, aggregated per year. We treat these data as histograms of maternal age when giving birth, with bins representing age of birth, bin widths one year and the bin frequencies being equal to the total live births corresponding to that age. These histograms are then smoothed (for which we employ local least squares smoothing using the Hades package available at https://stat.ucdavis.edu/hades/) to obtain smooth probability density functions for maternal age, where we consider the age interval [12, 55].

Figure 6 displays the evolving densities for Finland over a period spanning 51 years from 1960 to 2010. One can see that the mode of the distribution of maternal age shows a shift between the early and later years with more abrupt changes taking place near the end of the 1970s and beginning of the 1980s. These changes likely are attributable to increasingly delayed childbearing age, a phenomenon that has been observed since the 1960s in many European countries and has been much studied in demography (Hellstrand et al. (2019), Kohler, Billari and Ortega (2002, 2006)), where improved contraception, more women opting for higher education and increased labor force participation of women are discussed as possible causes. We chose the 2-Wasserstein metric as distance between probability distributions, which has proved to be well suited for many applications (Bolstad et al. (2003)) and applied the proposed Fréchet variance based change-point detection method, with calendar year serving as index for the sequence of fertility distributions. Figure 7 supports the intuition that there is a change-point in the data. The estimated location of the change-point is the year 1982. The test (2.8) for the presence of a change-point rejects the null hypothesis of no change-point in the sequence with a bootstrap p-value indistinguishable from zero. Figure 8 shows clear differences between the estimated Fréchet mean distributions before and after 1982. We see that the mode of the pre-1982 Fréchet mean distributions is before age 25 which shifts rapidly around 1982 to after age 25, indicating significant changes in the pattern of fertility.
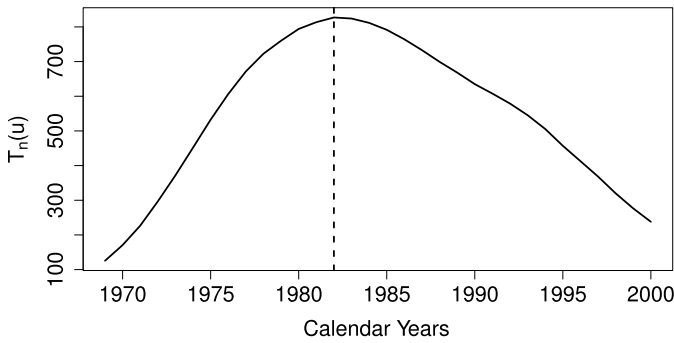
FIG. 7. *Scan function $T_n(u)$ during the time period* 1970 *to* 2000 *for the Finnish fertility data. The dotted line indicates the location of the estimated change-point in the year* 1982. *The proposed test for the presence of a change-point has a bootstrap p-value that is indistinguishable from zero for the null hypothesis of no change-point.*

5.2. *Enron e-mail network data.*   Enron Corporation was an American energy company which became notorious in 2001 for accounting fraud that eventually led to the company's bankruptcy. During the investigation after the company's collapse, data containing e-mail exchanges of the company's employees were made public by the Federal Energy Regulatory Commission. One of the versions is available at http://www.cis.jhu.edu/~parky/Enron/. We study whether changes in e-mail patterns reflect important events in the timeline of the company's downfall. In our analysis we consider the time period between November 1998 to June 2002. The weekly total e-mail activity between the 150 employees of the company during this time period is displayed in Figure 9. Peel and Clauset (2015) previously analyzed the Enron e-mail network data and identified a total of 16 change-points corresponding to 25 events of interest during the three year timeline.

The Enron e-mail network includes 184 e-mail addresses. Since the time period of three years is quite long, we break it into weekly intervals and then generate a network data sequence of length 183 corresponding to the 183 weeks between November 1998 to June 2002, where the 184 e-mail addresses were treated as nodes. For week i, from-to pairs extracted from the e-mails are used to calculate the total number of e-mails exchanged between the
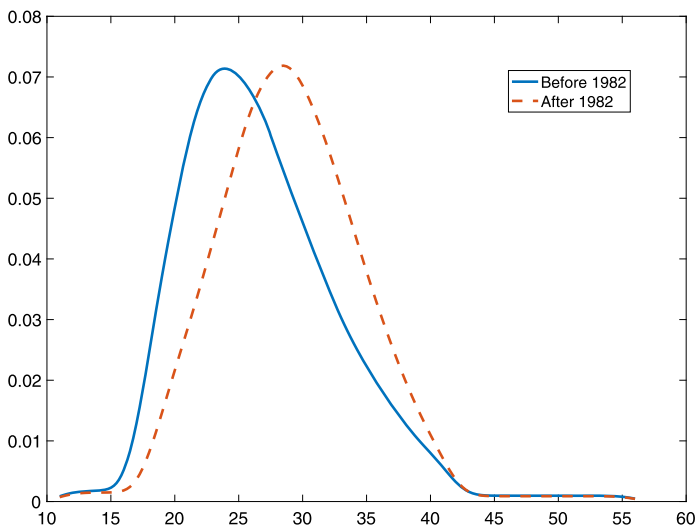

FIG. 8. *Estimated Fréchet mean distributions of maternal age, represented as densities, before the estimated change-point in* 1982 (*blue*) *and after* (*red*) *for the Finnish fertility data.*
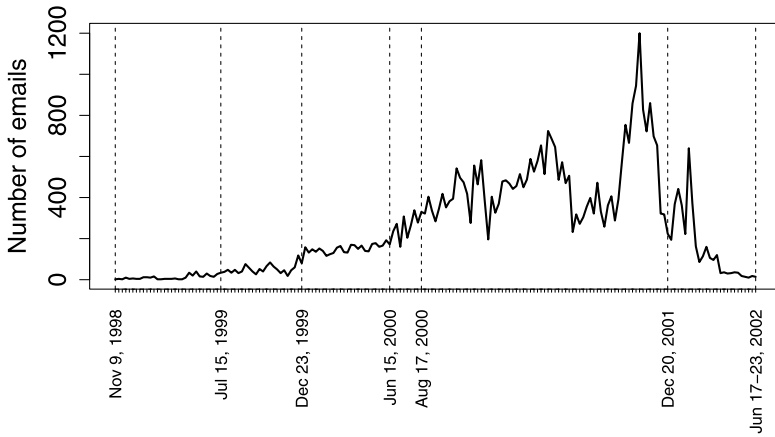
FIG. 9. *Weekly cumulative e-mail activity of Enron employees from November* 5,1998 *to June* 23, 2002. *Some relevant weeks are indicated on the x-axis by their mid-week dates.*

e-mail addresses *j* and *k* which then is considered the edge weight in the network adjacency matrix for that week. We used the Frobenius metric between network adjacency matrices and applied the proposed method to this sequence of networks.

The scan function in the left panel of Figure 10 indicates that a change-point might be present in week 88 with mid-week date August 17, 2000. The bootstrap version of the proposed test confirms the significance with a p-value of p ≈ 0. This date is located just before an important event in the timeline of Enron when its stock prices hit an all time high on August 23, 2000; for more details on the overall timeline of events we refer to http://www.agsm.edu.au/bobm/teaching/BE/Enron/timeline.html. The Fréchet mean networks also show clear distinctions before and after week 88, as illustrated in subfigures (a) and (b) of Figure 11.
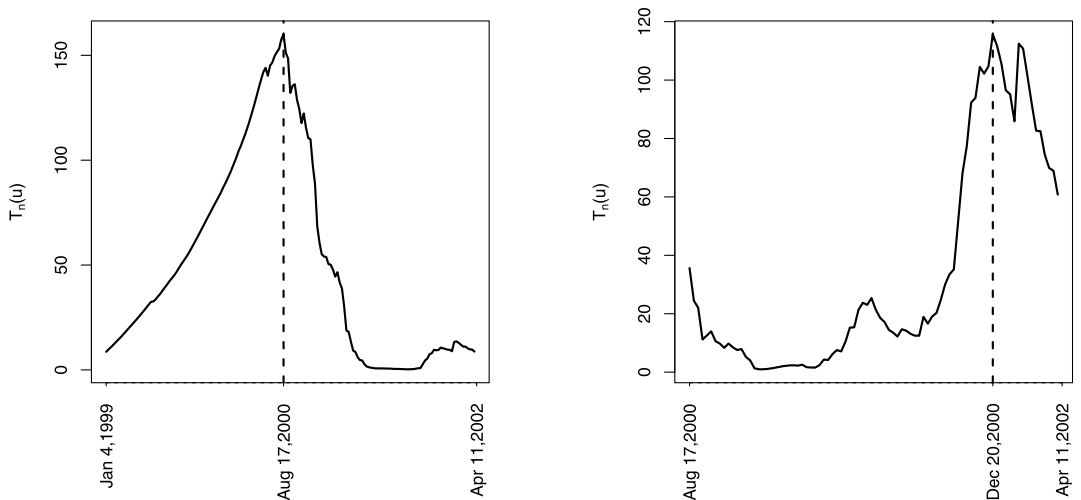
FIG. 10. *Weekly scan function* $T_n(u)$ *for the entire network data sequence between January* 4, 1999 *to April* 11, 2002 (*left panel*) *and the network data sequence after the first change-point* (*week* 88 *with midpoint August* 17, 2000, *as detected in the left panel*) *between August* 17, 2000 *to April* 11, 2002 (*right panel*) *for the Enron e-mail data. In the right panel a change-point is located on December* 20, 2000. *Dotted lines indicate the location of the estimated change-points. The test for the presence of a change-point has a bootstrap p-value indistinguishable from zero for both panels, indicating the significance of the presence of these change-points.*
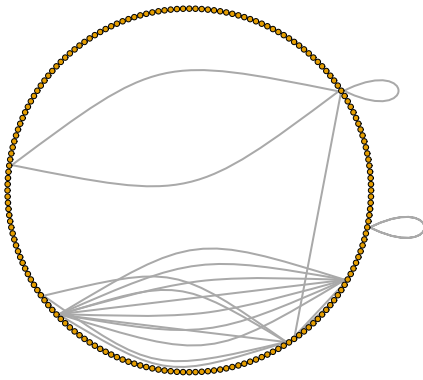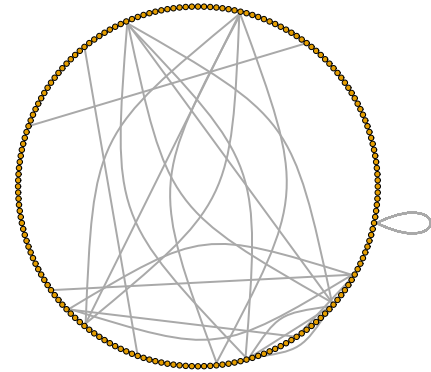
(a) Nov 9, 1998 to Aug 20, 2000
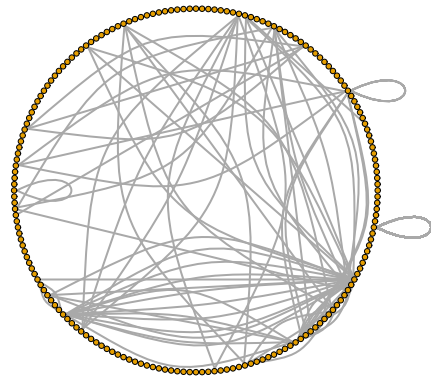
(b) Aug 21, 2000 to June 23, 2002
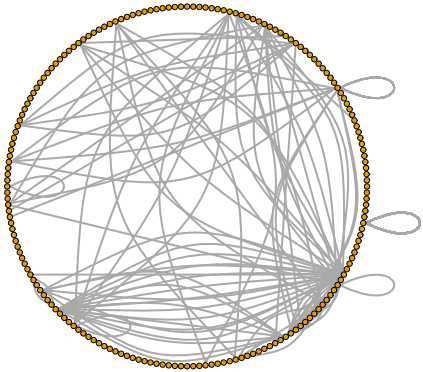
(c) Aug 21, 2000 to Dec 23, 2001

(d) Dec 24, 2001 to June 23, 2002

FIG. 11.   *Estimated Fréchet means of weekly network adjacency matrices for the time periods as indicated.*

On December 2, 2001, Enron filed for Chapter 11 bankruptcy protection, and on January 9, 2002, it was confirmed that a criminal investigation was started against the company. We separately analyzed the network sequence starting from week 89 to week 183, which is the period starting right after week 88 which we detected to be a change-point, as discussed above. The right panel of Figure 10 shows that the proposed test discovered a change-point in week 158, which is is the week December 17 to December 23, 2001, with mid-week date December 20, 2001. This is shortly after Enron filed for bankruptcy and just before the criminal investigation began. The Fréchet mean networks are again seen to clearly differ, as illustrated in the subfigures of Figure 11. When testing for change points in the above time intervals with the proposed test (2.2), the null hypothesis of no change-point is rejected at significance level 0.05 with bootstrap p-values indistinguishable from zero.

To investigate whether there are any additional relevant change-points, we carried out a binary segmentation approach and identified the following weeks to be candidates for potential change-points: July 12 to 18, 1999 (bootstrap p value = 0.04), December 20 to 26, 1999 (bootstrap p-value = 0) and June 12 to 18, 2000 (bootstrap p-value = 0). The first of

these corresponds to the period right after June 28, 1999, which is when Enron's CFO was allowed to run a private equity fund LJM1 which later became one of Enron's key tools to maintain a public balance sheet; for details see http://www.agsm.edu.au/bobm/teaching/BE/Enron/timeline.html. The second of these weeks is sandwiched between the launch of Enron Online in December 1999, and the launch of Enron Broadband Services (EBS) in January 1999; the third week with a change-point is right before a partnership was launched between EBS and Blockbuster to provide video on demand.

6. Discussion. Change-point detection is challenging for data sequences that take values in a general metric space. Existing approaches like Chen and Zhang (2015), Chu and Chen (2019) address this challenge by considering similarity/dissimilarity graphs between the data objects as the starting point. In this approach the choice of the graph is a tuning parameter that affects the conclusions; generally, the presence of a tuning parameter is undesirable in inference problems. Both the graph based and the proposed methods require a cut-off parameter to determine intervals near both left and right endpoints where a change-point cannot occur. This could be potentially circumvented by introducing a suitable weight function, as pointed out by a reviewer; however, the power of the test will invariably suffer if a change-point occurs close to the endpoints. Neither the proposed test for the presence of a change-point nor the proposed estimate of the location of the change-point require additional tuning parameters. They possess several additional desirable features that are not available for alternative approaches for change-point detection in general metric spaces, including consistency of the test for the presence of a change-point and consistency of the estimated change-point location.

The proposed test does not always work well; in Appendix D in the Supplementary Material, we present a situation where a change in distribution occurs and consider the analysis under two metrics. For the first of these metrics the change in distributions is not reflected in terms of mean or scale differences; for the second metric the change in distributions is associated with a scale difference. This has the consequence that, under the first metric, the proposed test is unable to detect the alternative, whereas the graph-based test of Chen and Zhang (2015) performs well. Under the second metric the same differences in distributions are reflected in a scale change. Under this second metric both tests perform equally well. This demonstrates that in some situations the choice of the metric can play a critical role for inference with random objects.

Overall, it is an advantage of the proposed test statistic that, in spite of the complexity of metric space valued objects, it has an intuitive interpretation as it mimics change-points that feature location-scale alternatives. The theoretical guarantees for the consistency of the test under contiguous alternatives with such features (3.9) within small departures from H_0 (2.1) and the consistency of the estimated change-point location under H_1 (2.2) are distinguishing features of the proposed methods. Simulations and applications demonstrate that the proposed theoretically justified bootstrap version is very helpful to obtain improved inference in finite sample situations.

Lastly, the proposed approach is applicable for a wide class of random objects, as the necessary assumptions are satisfied by various metric spaces of interest. There are many open problems in this area, including sequential versions of the test or metric selection.

SUPPLEMENTARY MATERIAL

Supplementary material for "Fréchet change-point detection" (DOI: [10.1214/19-AOS1930SUPP](); .pdf). The supplementary material contain all proofs of the main results and also various additional auxiliary lemmas and their proofs. It also features a discussion of spaces (Ω, d), which satisfy assumptions (A1) to (A4) in the paper, with formal results stated as Proposition C.1 and C.2 in Appendix C, and also additional simulation results in Appendix D.

REFERENCES

AHIDAR-COUTRIX, A., GOUIC, T. L. and PARIS, Q. (2018). Convergence rates for empirical barycenters in metric spaces: Curvature, convexity and extendible geodesics. arXiv preprint [arXiv:1806.02740]().

ARCONES, M. A. (1998). A remark on approximate M-estimators. *Statist. Probab. Lett.* **38** 311–321. [MR1631213]() [https://doi.org/10.1016/S0167-7152(98)00038-8]()

ARLOT, S., CELISSE, A. and HARCHAOUI, Z. (2012). Kernel change-point detection. arXiv preprint [arXiv:1202.3878]().

BARABÁSI, A.-L. and ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286** 509–512. [MR2091634]() [https://doi.org/10.1126/science.286.5439.509]()

BOLSTAD, B. M., IRIZARRY, R. A., ÅSTRAND, M. and SPEED, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19** 185–193.

CARLSTEIN, E., MÜLLER, H.-G. and SIEGMUND, D., eds. (1994). *Change-Point Problems. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **23**. IMS, Hayward, CA. Papers from the AMS-IMS-SIAM Summer Research Conference held at Mt. Holyoke College, South Hadley, MA, July 11–16, 1992. [MR1477909]()

CAZELLES, E., SEGUY, V., BIGOT, J., CUTURI, M. and PAPADAKIS, N. (2018). Geodesic PCA versus log-PCA of histograms in the Wasserstein space. *SIAM J. Sci. Comput.* **40** B429–B456. [MR3780753]() [https://doi.org/10.1137/17M1143459]()

CHEN, H. and FRIEDMAN, J. H. (2017). A new graph-based two-sample test for multivariate and object data. *J. Amer. Statist. Assoc.* **112** 397–409. [MR3646580]() [https://doi.org/10.1080/01621459.2016.1147356]()

CHEN, J. and GUPTA, A. K. (2012). *Parametric Statistical Change Point Analysis*: *With Applications to Genetics*, *Medicine*, *and Finance*, 2nd ed. Birkhäuser/Springer, New York. [MR3025631]() [https://doi.org/10.1007/978-0-8176-4801-5]()

CHEN, H. and ZHANG, N. (2015). Graph-based change-point detection. *Ann. Statist.* **43** 139–176. [MR3285603]() [https://doi.org/10.1214/14-AOS1269]()

CHERNOFF, H. and ZACKS, S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. *Ann. Math. Stat.* **35** 999–1018. [MR0179874]() [https://doi.org/10.1214/aoms/1177700517]()

CHU, L. and CHEN, H. (2019). Asymptotic distribution-free change-point detection for multivariate and non-Euclidean data. *Ann. Statist.* **47** 382–414. [MR3910545]() [https://doi.org/10.1214/18-AOS1691]()

CSÖRGŐ, M. and HORVÁTH, L. (1997). *Limit Theorems in Change-Point Analysis*. *Wiley Series in Probability and Statistics*. Wiley, Chichester. With a foreword by David Kendall. [MR2743035]()

DE RIDDER, S., VANDERMARLIERE, B. and RYCKEBUSCH, J. (2016). Detection and localization of change points in temporal networks with the aid of stochastic block models. *J. Stat. Mech. Theory Exp.* **2016** 113302.

DUBEY, P. and MÜLLER, H.-G. (2019). Fréchet analysis of variance for random objects. *Biometrika* **106** 803–821. [MR4031200]() [https://doi.org/10.1093/biomet/asz052]()

DUBEY, P. and MÜLLER, H.-G. (2020). Supplement to "Fréchet change-point detection." [https://doi.org/10.1214/19-AOS1930SUPP]()

FRÉCHET, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. Henri Poincaré* **10** 215–310. [MR0027464]()

GARREAU, D. and ARLOT, S. (2018). Consistent change-point detection with kernels. *Electron. J. Stat.* **12** 4440–4486. [MR3892345]() [https://doi.org/10.1214/18-ejs1513]()

GINESTET, C. E., LI, J., BALACHANDRAN, P., ROSENBERG, S. and KOLACZYK, E. D. (2017). Hypothesis testing for network data in functional neuroimaging. *Ann. Appl. Stat.* **11** 725–750. [MR3693544]() [https://doi.org/10.1214/16-AOAS1015]()

GUNTUBOYINA, A. and SEN, B. (2013). Covering numbers for convex functions. *IEEE Trans. Inf. Theory* **59** 1957–1965. [MR3043776]() [https://doi.org/10.1109/TIT.2012.2235172]()

HELLSTRAND, J., MYRSKYLÄ, M., NISÉN, J. et al. (2019). All-time low period fertility in Finland: Drivers, tempo effects, and cohort implications. Technical Report, Max Planck Institute for Demographic Research, Rostock, Germany.

JAMES, B., JAMES, K. L. and SIEGMUND, D. (1987). Tests for a change-point. *Biometrika* **74** 71–83. MR0885920 https://doi.org/10.1093/biomet/74.1.71

JAMES, B., JAMES, K. L. and SIEGMUND, D. (1992). Asymptotic approximations for likelihood ratio tests and confidence regions for a change-point in the mean of a multivariate normal distribution. *Statist. Sinica* **2** 69–90. MR1152298

JIRAK, M. (2015). Uniform change point tests in high dimension. *Ann. Statist.* **43** 2451–2483. MR3405600 https://doi.org/10.1214/15-AOS1347

KOHLER, H.-P., BILLARI, F. C. and ORTEGA, J. A. (2002). The emergence of lowest-low fertility in Europe during the 1990s. *Popul. Dev. Rev.* **28** 641–680.

KOHLER, H.-P., BILLARI, F. C. and ORTEGA, J. A. (2006). Low fertility in Europe: Causes, implications and policy options. In *The Baby Bust*: *Who Will do the Work* 48–109.

LUNG-YUT-FONG, A., LÉVY-LEDUC, C. and CAPPÉ, O. (2015). Homogeneity and change-point detection tests for multivariate data using rank statistics. *J. SFdS* **156** 133–162. MR3436651

MACNEILL, I. B. (1974). Tests for change of parameter at unknown times and distributions of some related functionals on Brownian motion. *Ann. Statist.* **2** 950–962. MR0426253

MATTESON, D. S. and JAMES, N. A. (2014). A nonparametric approach for multiple change point analysis of multivariate data. *J. Amer. Statist. Assoc.* **109** 334–345. MR3180567 https://doi.org/10.1080/01621459.2013.849605

MONTGOMERY-SMITH, S. J. and PRUSS, A. R. (2001). A comparison inequality for sums of independent random variables. *J. Math. Anal. Appl.* **254** 35–42. MR1807885 https://doi.org/10.1006/jmaa.2000.7200

NIU, Y. S., HAO, N. and ZHANG, H. (2016). Multiple change-point detection: A selective overview. *Statist. Sci.* **31** 611–623. MR3598742 https://doi.org/10.1214/16-STS587

OSSIANDER, M. (1987). A central limit theorem under metric entropy with L_2 bracketing. *Ann. Probab.* **15** 897–919. MR0893905

PEEL, L. and CLAUSET, A. (2015). Detecting change points in the large scale structure of evolving networks. In *AAAI* **15** 1–11.

PETERSEN, A. and MÜLLER, H.-G. (2019a). Wasserstein covariance for multiple random densities. *Biometrika* **106** 339–351. MR3949307 https://doi.org/10.1093/biomet/asz005

PETERSEN, A. and MÜLLER, H.-G. (2019b). Fréchet regression for random objects with Euclidean predictors. *Ann. Statist.* **47** 691–719. MR3909947 https://doi.org/10.1214/17-AOS1624

SIEGMUND, D. (1988). Confidence sets in change-point problems. *Int. Stat. Rev.* **56** 31–48. MR0963139 https://doi.org/10.2307/1403360

SRIVASTAVA, M. S. and WORSLEY, K. J. (1986). Likelihood ratio tests for a change in the multivariate normal mean. *J. Amer. Statist. Assoc.* **81** 199–204. MR0830581

STURM, K.-T. (2003). Probability measures on metric spaces of nonpositive curvature. In *Heat Kernels and Analysis on Manifolds*, *Graphs*, *and Metric Spaces* (*Paris*, 2002). *Contemp. Math.* **338** 357–390. Amer. Math. Soc., Providence, RI. MR2039961 https://doi.org/10.1090/conm/338/06080

STURM, K.-T. (2006). On the geometry of metric measure spaces. I. *Acta Math.* **196** 65–131. MR2237206 https://doi.org/10.1007/s11511-006-0002-8

TAVAKOLI, S., PIGOLI, D., ASTON, J. A. D. and COLEMAN, J. S. (2019). A spatial modeling approach for linguistic object data: Analyzing dialect sound variations across Great Britain. *J. Amer. Statist. Assoc.* **114** 1081–1096. MR4011760 https://doi.org/10.1080/01621459.2019.1607357

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*: *With Applications to Statistics*. *Springer Series in Statistics*. Springer, New York. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2

WANG, T. and SAMWORTH, R. J. (2018). High dimensional change point estimation via sparse projection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 57–83. MR3744712 https://doi.org/10.1111/rssb.12243

WANG, D., YU, Y. and RINALDO, A. (2018). Optimal change point detection and localization in sparse dynamic networks. arXiv preprint arXiv:1809.09602.