Fréchet Analysis of Variance for Random Objects

Paromita Dubey[1] and Hans-Georg Müller[2]
Department of Statistics, University of California, Davis
Davis, CA 95616 USA

October 2017

ABSTRACT

Fréchet mean and variance provide a way of obtaining mean and variance for general metric space valued random variables and can be used for statistical analysis of data objects that lie in abstract spaces devoid of algebraic structure and operations. Examples of such spaces include covariance matrices, graph Laplacians of networks and univariate probability distribution functions. We derive a central limit theorem for Fréchet variance under mild regularity conditions, utilizing empirical process theory, and also provide a consistent estimator of the asymptotic variance. These results lead to a test to compare k populations based on Fréchet variance for general metric space valued data objects, with emphasis on comparing means and variances. We examine the finite sample performance of this inference procedure through simulation studies for several special cases that include probability distributions and graph Laplacians, which leads to tests to compare populations of networks. The proposed methodology has good finite sample performance in simulations for different kinds of random objects. We illustrate the proposed methods with data on mortality profiles of various countries and resting state Functional Magnetic Resonance Imaging data.

KEY WORDS: Functional Data Analysis; Fréchet mean; Fréchet variance; Central Limit Theorem; Two sample test; Samples of probability distributions; Wasserstein metric; Samples of networks; Graph Laplacians; fMRI

[1]Corresponding author; email: pdubey@ucdavis.edu
[2]Research supported by National Science Foundation grants DMS-1407852 and DMS-172864

1. INTRODUCTION

With an increasing abundance of complex non-Euclidean data in multiple disciplines there is a need for statisticians to develop techniques suitable for the analysis of such data. Settings where data objects take values in a metric space that is devoid of vector space structure are common, and occur for example when only pairwise distances between the observed data objects are available. It is then natural to assume that observed samples of random objects are drawn from a distribution in a metric space. The standard problem of K-sample testing, with the most important case concerning the two sample case with $K = 2$, is of basic interest in statistics and becomes more challenging when data objects lie in general metric spaces. For the case of Euclidean data this is the classical K sample comparison problem. For Gaussian data it corresponds to the classical analysis of variance (ANOVA), if comparing means is of primary interest, which uses comparisons of summary variation measures between and within groups.

For general metric space valued random variables, Fréchet (1948) provided a direct generalization of the mean which implies a corresponding generalization of variance that may be used to quantify the spread of the distribution of metric space valued random variables or random objects around their Fréchet mean. The Fréchet mean resides in the object space and thus is not amenable to algebraic operations, which implies that a central limit theorem cannot be directly applied to obtain limit distributions for Fréchet means. In contrast, the Fréchet variance is always a scalar, which makes it more tractable. A key result of this paper is a central limit theorem for the empirical Fréchet variance of data objects in general metric spaces under weak assumptions. While this result is of interest in itself, we demonstrate how it can be applied to derive a k-sample test based on groupwise Fréchet variances for testing the null hypothesis of equal population distributions for random objects, with emphasis on testing the equality of Fréchet means or of the Fréchet variances. In recent years the study of nonparametric tests for the equality of two distributions for Euclidean data has expanded to cover non-Euclidean data, which are increasingly encountered in frameworks that feature large and complex data. Tests that are suitable for random objects in metric spaces are typically based on pairwise distances.

Major approaches have been based on nearest neighbors, graphs, energy statistics and kernels.

Nearest neighbor based tests (Henze, 1988; Henze and Penrose, 1999; Schilling, 1986) count the number of K-nearest neighbor comparisons for which the observations and their neighbors belong to the same sample. The choice of the tuning parameter K impacts the resulting inference. For *graph-based* two sample tests (Friedman and Rafsky, 1979), pairwise distances of the pooled sample are used to form an edge-weighted graph and its minimum spanning tree. The test statistic is the number of subtrees formed by the removal of the edges for which the defining nodes belong to different samples. Recently, modified versions of this test were proposed (Chen and Friedman, 2017; Chen et al., 2017) also based on a similarity graph of the pooled observations and a weighted edge count test in order to deal with the problem of unequal sample sizes. Since it is more difficult to form within sample edges for samples with smaller sizes, weights proportional to the reciprocal of sample sizes are assigned to the within sample edges. A related approach is minimum distance non-bipartite pairing (Rosenbaum, 2005), where the dataset is split into disjoint pairs by minimizing the total distance within the pairs and the test statistic is the number of cross matches.

The class of statistics popularly known as *energy statistics*, initially proposed for the case of a multivariate two sample test (Baringhaus and Franz, 2004; Székely and Rizzo, 2004) and based on the fact that for independent d-dimensional random vectors X_1, X_2, Y_1 and Y_2 where X_1 and X_2 have the same distribution F and Y_1 and Y_2 have same distribution G, both with finite expectations, the inequality

$$E\|X_1 - Y_1\| - \frac{1}{2}E\|X_1 - X_2\| - \frac{1}{2}E\|Y_1 - Y_2\| \geq 0$$

holds, with equality if and only if $F = G$, where F and G are the two distributions to be compared. This framework was extended to the case of metric space valued random variables in Lyons (2013) for metric spaces of strong negative type, leading to the distinction of measures that have finite first moments with respect to the underlying metric. An extension of the energy statistics approach for spaces admitting a manifold stratification (Patrangenaru and Ellingson, 2015) has been explored recently (Guo and Patrangenaru, 2017). A recent review is Székely and Rizzo (2017).

Kernel based two sample testing (Gretton et al., 2012) aims at finding the the *maximum mean discrepancy (MMD)* which is the largest difference in expectations over functions in the unit ball of a reproducing kernel Hilbert space (RKHS) and corresponds to distances between embeddings of distributions in the RKHS. Distribution free tests can be obtained based on large deviation bounds of empirical estimates and also the asymptotic distribution of MMD, and the energy statistics approach can be linked to the kernel method for two sample testing, as the distance based energy statistic computed using a semi-metric of negative type is equivalent to the MMD statistic obtained from a special type of kernel induced by the semi-metric (Sejdinovic et al., 2013). For every positive definite kernel, the MMD statistic corresponds to the energy statistic obtained from a semi-metric derived from the kernel used to compute the MMD. Connections between kernel methods, energy statistics and Wasserstein distance between the distributions to be compared are explored in Ramdas et al. (2017). Commonly used kernels in the kernel framework, like the Gaussian kernel, might be sensitive to the choice of the tuning parameter needed to scale the kernel and determining a good scaling factor can be challenging when the goal is inference for complex data. We confirm in our simulations that this can be problematic when adopting the kernel method to obtain inference.

Empirically, these tests have good power performance for either location type alternatives or scale type alternatives, but usually not for both simultaneously. A major challenge for some of these tests is the choice of the required tuning parameters, which often has a major impact on the resulting inference. The challenges associated with existing inference procedures motivate our proposed test, which is simple and is based on a test statistic that mimics the statistics on which t tests and classical ANOVA are based, replacing between and within sums of squares with the corresponding Fréchet variances for separate and combined samples, and is easy to compute for k- sample comparisons.

Fréchet mean based testing and corresponding large sample theory including laws of large numbers and central limit theorems for empirical Fréchet means have been explored previously for data objects that lie in special metric spaces, such as smooth Riemannian manifolds (Bhattacharya and Patrangenaru, 2003, 2005; Bhattacharya and Bhattacharya, 2012) and topologically stratified spaces under certain restrictions, like phylogenetic trees (Kendall and Le, 2011; Barden et al., 2013; Bhattacharya and Lin, 2017). Virtually all of these results depend on local linear tangent or similar approximations that are specific to these manifold spaces but do not apply for random objects in more general metric spaces as they require local Euclidean approximation. The central limit theorem for Fréchet means was recently applied to the space of graph Laplacians (Ginestet et al., 2017), which are of interest to obtain inference for networks. This required choosing a high dimension for the approximating space, thus leading to problems with small sample high dimensional data and the ensuing complications for inference.

Since Fréchet mean based testing exploits the local Euclidean approximation property of the underlying space, it does not extend to data in general metric spaces such as the space of univariate probability density functions, which will serve as one of our examples to illustrate the proposed approach. Another drawback of these methods is that they lose power when the data is high dimensional as large covariance matrices and their inverses need to be estimated. For a network with as few as 20 nodes, the Fréchet mean of the graph Laplacians has a covariance matrix of dimension 36100. To address this issue, strong assumptions, such as sparsity, have been invoked for the estimation of these large covariance matrices and their inverses in order to implement the corresponding tests. This provides additional motivation for our approach, which is very simple and where dimension of the data enters only indirectly through properties of the metric. To obtain asymptotic properties, we draw on empirical process theory, in a similar spirit as a recent study of regression relationships between general metric space valued random objects as responses and real valued predictors (Petersen and Müller, 2017).

Our goal in this paper is to develop a simple and straightforward extension of ANOVA, which is one of the very basic tools for inference in statistics, to the case of metric space valued random objects. Our starting point is a totally bounded metric space Ω equipped with a metric d. We show that consistency of the sample Fréchet mean can be derived by using results of Petersen and Müller (2017) concerning Fréchet regression estimators under mild assumptions on Ω. We derive a central limit theorem for Fréchet variance under mild assumptions and provide a consistent estimator of its asymptotic variance. Our method is applicable to a wide class of objects including correlation matrices, univariate probability distributions, manifolds and also the space of graph Laplacians. Making use of the central limit theorem, we derive a new k-sample test for random objects and study the asymptotic distribution of the test statistic under the null hypothesis of equality of the population distributions, as well as its power function.

It is customary to test for heteroscedasticity of the population groups prior to the F-test of classical ANOVA to evaluate the assumption of equal variances across the populations that are compared. One popular test for this purpose was proposed by Levene (1960), where the test statistic is of the form of the usual ANOVA F-test, but applied to pseudo-observations which could in principle be any monotonic function of the absolute deviations of the observations from their group 'centers'. Our proposed test unifies Levene's test and the classical ANOVA for testing inequality of population means for data objects in general metric spaces and therefore aims at both location and scale type alternatives instead of only location alternatives which is the objective of classical ANOVA. Our test statistic is a sum of two components. One component is proportional to the squared difference of the pooled sample Fréchet variance and the weighted average of the groupwise Fréchet variances, with weights proportional to the sample sizes of the groups. For the special case of Euclidean data, this part of our statistic is proportional to the squared F-ratio as in the usual ANOVA and is useful to detect differences in the Fréchet means of the populations. A key auxiliary result is that this statistic converges to zero at rate $o_P(\frac{1}{n})$ under the null hypothesis of equality of Fréchet means of population distributions and is bounded away from zero otherwise. The other component of our test statistic accounts for differences in the Fréchet variances of the population groups and under the Euclidean setting simplifies to a generalization of the Levene's test applied to squared absolute deviations of the observations from their group Fréchet means. When the assumptions of the central limit theorem hold, the asymptotic distribution of this component of our test statistic is $\chi^2_{(k-1)}$ where k is the number of populations to be compared.

The paper is organized as follows: The basic set up is defined in Section 2 and the theory regarding the asymptotic behavior of Fréchet variance is provided in Section 3. The k-sample test is introduced in Section 4, followed by empirical studies to study its power performance in Section 5. Our data applications in Section 6 include samples of univariate probability distributions for which we employ the L^2 Wasserstein metric and samples of graph Laplacians of networks, for which we use the Frobenius metric. In Section 6.1 we illustrate the proposed tests with human mortality records of 31 countries over the time period 1960-2009 to analyze the evolution of age at death distributions of the countries over the years. In Section 6.2 we analyze a resting state fMRI dataset where we compare the probability distribution of positive correlations between fMRI signals in the posterior cingulate area of the brain of Alzheimer's Disease patients to those of similarly aged normal subjects. Finally, in Section 6.3 we analyze brain connectivity networks of patients with dementia that are derived from fMRI signals of certain brain regions (Buckner et al., 2009) and compare the connectivity networks of demented patients for three different age groups.

2. PRELIMINARIES

The Fréchet mean is a generalization of centroids to metric spaces and for the special case of Euclidean data it includes the arithmetic mean, median and geometric mean under different choices of distance functions. The Fréchet variance is the corresponding generalized measure of dispersion around the Fréchet mean. More formally, in all of the following, (Ω, d, P) is a totally bounded metric space with metric d and probability measure P. Random objects in the following are random variables Y that take values in Ω. The (population) Fréchet mean of Y is

$$\mu_F = \underset{\omega \in \Omega}{\operatorname{argmin}} E(d^2(\omega, Y)), \tag{1}$$

while for a random sample $Y_1, Y_2, \ldots, Y_n$ of i.i.d. random variables with the same distribution as Y, the corresponding sample Fréchet mean is

$$\hat{\mu}_F = \underset{\omega \in \Omega}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} d^2(\omega, Y_i). \tag{2}$$

The sample Fréchet mean is an M-estimator as it is obtained by minimizing a sum of functions of the data objects. The population Fréchet variance quantifies the spread of the random variable Y around its Fréchet mean μ_F,

$$V_F = E(d^2(\mu_F, Y)), \tag{3}$$

with corresponding sample based estimator

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

Note that $\mu_F, \hat{\mu}_F \in \Omega$, while $V_F, \hat{V}_F \in \mathcal{R}$.

The asymptotic consistency of the sample Fréchet mean $\hat{\mu}_F$ follows from results in Petersen and Müller (2017), under the following assumption:

(P0) The objects $\hat{\mu}_F$ and μ_F exist and are unique, and for any $\varepsilon > 0$, $\inf_{d(\omega,\mu_F)>\varepsilon} E(d^2(\omega, Y)) > E(d^2(\mu_F, Y))$.

Assumption (P0) is instrumental to establish the weak convergence of the empirical process H_n to the population process H, where

$$H_n(\omega) = \frac{1}{n} \sum_{i=1}^{n} d^2(\omega, Y_i), \quad H(\omega) = E(d^2(\omega, Y)),$$

which in turn implies the consistency of $\hat{\mu}_F$,

$$d(\hat{\mu}_F, \mu_F) = o_P(1). \tag{5}$$

Consistency of $\hat{\mu}_F$ then implies the consistency of $\hat{V}_F$ since

$$|\hat{V}_F - V_F| \leq 2 \operatorname{diam}(\Omega)\, d(\hat{\mu}_F, \mu_F), \tag{6}$$

where $\operatorname{diam}(\Omega) = \sup\{d(\omega_1, \omega_2) : \omega_1, \omega_2 \in \Omega\}$ is finite, since Ω is totally bounded. For the central limit theorem (CLT) to hold for empirical Fréchet variance we need an assumption on the entropy integral of the space Ω (Wellner and van der Vaart, 1996),

$$J(\delta) = \int_0^1 \sqrt{1 + \log(N(\varepsilon\delta/2, B_\delta(\mu_F), d))}\, d\varepsilon, \tag{7}$$

where $B_\delta(\mu_F)$ is the δ-ball in the metric d, centered at μ_F and $N(\varepsilon\delta/2, B_\delta(\mu_F), d)$ is the covering number for $B_\delta(\mu_F)$ using open balls of radius $\varepsilon\delta/2$. Specifically, for our CLT we assume that

(P1) $\delta\, J(\delta) \to 0$ as $\delta \to 0$.

For our results on the power of the proposed test in Section 4, we need an additional assumption on the entropy integral of the whole space Ω,

(P2) The entropy integral of Ω is finite, $\int_0^1 \sqrt{1 + \log N(\varepsilon, \Omega, d)} d\varepsilon < \infty$.

Random objects that satisfy assumptions (P0)-(P2) include the space of univariate probability distributions on $\mathcal{R}$ with finite second moments equipped with the Wasserstein metric d_W and the spaces of correlation matrices and graph Laplacians of fixed dimensions equipped with the Frobenius metric d_F. For two univariate distributions F and G with finite variances, the L^2-Wasserstein distance, also known as earth movers distance and closely related to optimal transport (Villani, 2003),

$$d_W^2(F, G) = \int_0^1 (F^{-1}(t) - G^{-1}(t))^2 dt, \tag{8}$$

where F^{-1} and G^{-1} are the quantile functions corresponding to F and G, respectively. For two matrices A and B of the same dimension, we consider the Frobenius metric,

$$d_F^2(A,B) = \text{trace}((A-B)'(A-B)). \tag{9}$$

To better characterize the space of graph Laplacians, we denote a weighted undirected graph by $G=(V,E)$, where V is the set of its vertices and E the set of its edges. Given an adjacency matrix W, where $w_{ij}=w_{ji}\geq 0$ and equality with zero holds if and only if $\{i,j\}\notin E$, the graph Laplacian is defined as $L=D-W$, where D is the diagonal matrix of the degrees of the vertices, i.e. $d_{jj}=\sum_i w_{ij}$. Under the assumption that the graphs are simple (i.e., there are no self loops or multi edges), there is a one to one correspondence between the space of graphs and the graph Laplacians and therefore the graph Laplacians can be used to characterize the space of networks (Ginestet et al., 2017). The following results imply that the spaces described above provide examples of spaces that satisfy assumptions (P0) and (P1). Our inference methods therefore will apply to compare samples of univariate distributions, correlation matrices and networks.

Proposition 1. *The space (Ω, d_W) satisfies assumptions (P0)-(P2) when the set Ω consists of univariate probability distributions on $\mathcal{R}$ with finite second moments and d_W is the L^2-Wasserstein metric.*

Proposition 2. *The space (Ω, d_F) satisfies assumptions (P0)-(P2) when the set Ω consists of graph Laplacians of connected, undirected and simple graphs of a fixed dimension r or correlation matrices of a fixed dimension r and d_F is the Frobenius metric.*

All proofs are in the Supplementary Materials.

3. CENTRAL LIMIT THEOREM FOR FRÉCHET VARIANCE

The following Proposition lays the foundations for proving the Central Limit Theorem for the empirical Fréchet variance $\hat{V}_F$.

Proposition 3. *Suppose assumptions (P0)-(P1) hold. Then*

$$\frac{1}{n}\sum_{i=1}^{n}\{d^2(\hat{\mu}_F, Y_i) - d^2(\mu_F, Y_i)\} = o_P(\frac{1}{\sqrt{n}}).$$

Proposition 3 makes it possible to deal with the sum of dependent random variables $\sum_{i=1}^{n} d^2(\hat{\mu}_F, Y_i)$, by replacing it with the sum of i.i.d. random variables $\sum_{i=1}^{n} d^2(\mu_F, Y_i)$, a crucial step in the derivation of the Central Limit Theorem for $\hat{V}_F$. Since Ω is totally bounded, the population Fréchet variance $\mathrm{Var}(d^2(\mu_F, Y))$ is always finite. The Central Limit Theorem for Fréchet variance is as follows.

Theorem 1. *Under the assumptions of Proposition 3,*

$$\sqrt{n}(\hat{V}_F - V_F) \xrightarrow{D} N(0, \sigma_F^2),$$

where $\sigma_F^2 = \mathrm{Var}(d^2(\mu_F, Y))$.

An intuitive sample based estimator for σ_F^2 is

$$\hat{\sigma}_F^2 = \frac{1}{n}\sum_{i=1}^{n} d^4(\hat{\mu}_F, Y_i) - \left(\frac{1}{n}\sum_{i=1}^{n} d^2(\hat{\mu}_F, Y_i)\right)^2, \tag{10}$$

and this estimator is $\sqrt{n}$-consistent as the following result shows.

Proposition 4. *Under the assumptions of Proposition 3,*

$$\sqrt{n}\left(\widehat{\sigma}_F^2 - \sigma_F^2\right) \xrightarrow{D} N(0, A), \tag{11}$$

where $A = a'Da$ *with* $a' = \left(1, -2E\left(d^2(\mu_F, Y)\right)\right)$ *and*

$$D = \begin{pmatrix} \mathrm{Var}\left(d^4(\mu_F, Y)\right) & \mathrm{Cov}\left(d^4(\mu_F, Y), d^2(\mu_F, Y)\right) \\ \mathrm{Cov}\left(d^4(\mu_F, Y), d^2(\mu_F, Y)\right) & \mathrm{Var}\left(d^2(\mu_F, Y)\right). \end{pmatrix}$$

Therefore $\widehat{\sigma}_F^2$ is a $\sqrt{n}$-consistent estimator of σ_F^2.

Proposition 4 is a variant of Proposition 3 and follows from a simple application of the delta method. The quantity A is finite by the boundedness of Ω. Combining Theorem 1, Proposition 3 and Slutsky's Theorem leads to

$$\frac{1}{\hat{\sigma}_F}\sqrt{n}(\hat{V}_F - V_F) \xrightarrow{D} N(0,1). \tag{12}$$

A simple application of the delta method gives the asymptotic distribution of the Fréchet standard deviation which is the square root of Fréchet variance,

$$\sqrt{n}(\hat{V}_F^{1/2} - V_F{}^{1/2}) \xrightarrow{D} N(0, \frac{\sigma_F^2}{4V_F}) \tag{13}$$

and since both $\hat{\sigma}_F$ and $\hat{V}_F$ are consistent estimators,

$$\frac{2\hat{V}_F^{1/2}}{\hat{\sigma}_F}\sqrt{n}(\hat{V}_F^{1/2} - V_F{}^{1/2}) \xrightarrow{D} N(0,1). \tag{14}$$

One can use (12) and (14) to construct asymptotic confidence intervals for Fréchet variance and standard deviation, which depend on the quality of the large sample approximations. The bootstrap provides an alternative that often has better finite sample properties under weak assumptions (Bickel and Freedman, 1981; Beran, 2003). Under fairly general assumptions, resampling methods like bootstrapping and permutation tests work whenever a central limit theorem holds (Janssen and Pauls, 2003). A basic criterion for bootstrap confidence sets to have correct coverage probability asymptotically is convergence of the bootstrap distribution of the root, in our case $\sqrt{n}(\hat{V}_F - V)/\hat{\sigma}_F$. Then Monte Carlo approximations of the bootstrap distribution of the root provide approximate quantiles for the construction of confidence sets.

For the empirical measure P_n generated by $Y_1, Y_2, \ldots, Y_n$ which puts mass $\frac{1}{n}$ to each of

$Y_1,\ldots,Y_n$, where P is the underlying measure and any measurable set A we have

$$P_n(A) = \frac{1}{n}\sum_{i=1}^{n} 1_A(Y_i) \xrightarrow{P} P(A), \tag{15}$$

where $1_A(\cdot)$ is the indicator function for the set A, by the weak law of large numbers. Given a sample $Y_1^*, Y_2^*, \ldots, Y_m^*$ of size m drawn with replacement from $Y_1, Y_2, \ldots, Y_n$, the bootstrap approximation of the root is $R_{m,n}^* = \sqrt{m}(\hat{V}_m^* - \hat{V}_F)/\hat{\sigma}_m^*$, where $\hat{V}_m^*$ and $\hat{\sigma}_m^*$ are the sample based estimators of the Fréchet variance and its asymptotic variance, obtained from the bootstrap sample $Y_1^*, Y_2^*, \ldots, Y_m^*$. If the Fréchet mean $\hat{\mu}_m^*$ of the bootstrap sample $Y_1^*, Y_2^*, \ldots, Y_m^*$ exists and is unique almost surely conditionally on $Y_1, Y_2, \ldots, Y_n$, then by applying the central limit theorem in Theorem 1 conditionally on $Y_1, Y_2, \ldots, Y_n$, $R_{m,n}^* \xrightarrow{D} N(0,1) \quad \text{as} \quad m \to \infty$. Since $N(0,1)$ has a continuous distribution function on the real line we have for each $\varepsilon > 0$, as $m \to \infty$,

$$P_n\left(\sup_x |H_{R_{m,n}^*}(x) - \Phi(x)| > \varepsilon\right) \to 0 \tag{16}$$

where $H_{R_{m,n}^*}(\cdot)$ is the distribution function of $R_{m,n}^*$ conditional on $Y_1, Y_2, \ldots, Y_n$ and $\Phi(\cdot)$ is the standard normal distribution function. From (15) for each $\varepsilon > 0$ we have as both $m, n \to \infty$

$$P\left(\sup_x |H_{R_{m,n}^*}(x) - \Phi(x)| > \varepsilon\right) \to 0,$$

which establishes asymptotic consistency of the bootstrap distribution. Therefore non-parametric bootstrapping is a viable option for the construction of confidence intervals for Fréchet variance.

4. COMPARING POPULATIONS OF RANDOM OBJECTS

Assume we have a sample of n Ω-valued random data objects $Y_1, Y_2, \ldots, Y_n$ that belong to k different groups $G_1, G_2, \ldots, G_k$, each of size n_j, $j = 1, \ldots, k$, such that $\sum_{j=1}^{k} n_j = n$. We wish to test the null hypothesis that the population distributions of the k groups are identical versus the alternative that at least one of the groups has a different population distribution compared to the others. Consider sample Fréchet means $\hat{\mu}_j$ in Ω, which are random objects computed just from the data falling into group j and the corresponding real-valued sample Fréchet variances $\hat{V}_j$ and variance estimates (10) $\hat{\sigma}_j^2$, $j = 1, \ldots, k$,

$$\begin{aligned} \hat{\mu}_j &= \underset{\omega \in \Omega}{\operatorname{argmin}} \frac{1}{n_j} \sum_{i \in G_j} d^2(\omega, Y_i), \quad \hat{V}_j = \frac{1}{n_j} \sum_{i \in G_j} d^2(\hat{\mu}_j, Y_i), \\ \hat{\sigma}_j^2 &= \frac{1}{n_j} \sum_{i \in G_j} d^4(\hat{\mu}_j, Y_i) - \left(\frac{1}{n_j} \sum_{i \in G_j} d^2(\hat{\mu}_j, Y_i) \right)^2, \end{aligned}$$

as well as the pooled sample Fréchet mean $\hat{\mu}_p$ and the corresponding pooled sample Fréchet $\hat{V}_p$,

$$\hat{\mu}_p = \underset{\omega \in \Omega}{\operatorname{argmin}} \frac{1}{n} \sum_{j=1}^{k} \sum_{i \in G_j} d^2(\omega, Y_i), \quad \hat{V}_p = \frac{1}{n} \sum_{j=1}^{k} \sum_{i \in G_j} d^2(\hat{\mu}_p, Y_i), \tag{17}$$

as well as weights

$$\lambda_{j,n} = \frac{n_j}{n}, \quad j = 1, \ldots, k, \quad \text{such that} \quad \sum_{j=1}^{k} \lambda_{j,n} = 1.$$

We will base our inference procedures on the auxiliary statistics

$$F_n = \hat{V}_p - \sum_{j=1}^{k} \lambda_{j,n} \hat{V}_j, \tag{18}$$

and

$$U_n = \sum_{j<l} \frac{\lambda_{j,n} \lambda_{l,n}}{\hat{\sigma}_j^2 \hat{\sigma}_l^2} (\hat{V}_j - \hat{V}_l)^2, \tag{19}$$

where F_n is almost surely non-negative and is equal to the numerator of the F-ratio in classical Euclidean ANOVA. Specifically, F_n corresponds to the weighted variance of the group means, with weights proportional to the group sizes, and correspond to the between group variance in the classical ANOVA setting. Hence F_n can be regarded as a generalization of the F-ratio in classical ANOVA to the more general setting of metric space valued data. Under the Euclidean setting, simple algebra shows that in this special case F_n is also proportional to the weighted average of the squared pairwise distances between the group Fréchet means. Analogous to ANOVA, the numerator of F_n is expected to be small under the null hypothesis of equality of the population distributions, which is indeed the case as the following Proposition demonstrates.

Proposition 5. *Suppose $\hat{\mu}_p$ and $\hat{\mu}_j$ exist and are unique almost surely for all $j = 1, \dots, k$. Let $0 < \lambda_{j,n} < 1$ for all $j = 1, \dots, k$ and $\lambda_{j,n} \to \lambda_j$ as $n \to \infty$, where λ_j is such that $0 < \lambda_j < 1$ for each $j = 1, \dots, k$, with $\sum_{k=1}^{k} \lambda_j = 1$. Then under the null hypothesis of equality of population distributions and under assumptions (P0) and (P1) for each of the groups, as $n \to \infty$,*

$$\sqrt{n} F_n = o_p(1). \tag{20}$$

Inference in classical ANOVA requires Gaussianity and equality of the population variances and hence targets only differences in the group means to capture differences in the population distributions. These assumptions are obviously too restrictive. We employ the statistics U_n in (19) to account for differences among the population variances, where in the Euclidean case, U_n turns out to be a slightly modified version of the traditional Levene's test, substituting squared distances of the observations from their group Fréchet means instead of just distances. The following result provides the asymptotic distribution of U_n under the null hypothesis.

Proposition 6. *Under the assumptions of Proposition 5, we have under the null hypothesis, as $n \to \infty$,*

$$\frac{nU_n}{\sum_{j=1}^{k} \frac{\lambda_{j,n}}{\hat{\sigma}_j^2}} \xrightarrow{D} \chi^2_{(k-1)}. \tag{21}$$

We construct our test statistic T_n by combining F_n and U_n in such a way that the distribution of T_n under the null hypothesis is the same as that given in Proposition 6 while gaining power against alternatives by ensuring that the asymptotic mean of T_n diverges when departing from the null hypothesis,

$$T_n = \frac{nU_n}{\sum_{j=1}^{k} \frac{\lambda_{j,n}}{\hat{\sigma}_j^2}} + \frac{nF_n^2}{\sum_{j=1}^{k} \lambda_{j,n}^2 \hat{\sigma}_j^2}. \tag{22}$$

For constructing T_n, we scale F_n by the estimated standard deviation of $\sum_{j=1}^{k} \lambda_{j,n} \hat{V}_j$, which is equal to $\sqrt{\sum_{j=1}^{k} \lambda_{j,n}^2 \hat{\sigma}_j^2}$, so that F_n is suitably scaled with respect to the variability of $\sum_{j=1}^{k} \lambda_{j,n} \hat{V}_j$. The second term on the r.h.s. of (22) is scaled such that both terms in T_n are of the same order in n. Under the null hypothesis, consistency of $\hat{\sigma}_j^2$ for $j = 1, 2, \cdots, k$ and Proposition 5 imply $\frac{nF_n^2}{\sum_{j=1}^{k} \lambda_{j,n}^2 \hat{\sigma}_j^2} = o_P(1)$.

Theorem 2. *Under the null hypothesis and the assumptions of Proposition 5,*

$$T_n \xrightarrow{D} \chi^2_{(k-1)}. \tag{23}$$

For a level α test, we accordingly reject the null hypothesis of equality of population distributions if the test statistic T_n turns out to be bigger than $\chi^2_{k-1,\alpha}$, which is the $(1-\alpha)^{th}$ quantile of the $\chi^2_{(k-1)}$ distribution, i.e., the rejection region that defines the test is

$$R_{n,\alpha} = \{T_n > \chi^2_{k-1,\alpha}\}. \tag{24}$$

To study the consistency of the proposed test (24) we consider contiguous alternatives that capture departures from the null hypothesis of equal population distributions in terms of differences in their Fréchet means and variances. We begin by defining the following population quantities,

$$\mu_p = \operatorname*{argmin}_{\omega \in \Omega} \sum_{j=1}^{k} \lambda_j E_j(d^2(\omega, Y_j)), \qquad V_p = \sum_{j=1}^{k} \lambda_j E_j(d^2(\mu_p, Y_j)) \tag{25}$$

and

$$F = V_p - \sum_{j=1}^{k} \lambda_j V_j, \qquad U = \sum_{j<l} \frac{\lambda_j \lambda_l}{\sigma_j^2 \sigma_l^2} (V_j - V_l)^2, \tag{26}$$

where $E_j(\cdot)$ denotes expectation under the probability distribution for the jth population and

$\lambda_{j,n}$ is as defined in Proposition 5, and the Y_j are random objects distributed according to the jth population distribution. In the Euclidean setting, μ_p and V_p are analogous to the pooled population Fréchet mean and pooled population Fréchet variance, respectively, and in the general case can be interpreted as generalizations of these quantities. While in the Euclidean case F is proportional to the weighted sum of the differences between the groupwise population Fréchet means and the pooled population Fréchet mean, by simple algebra it can be seen that this property still holds in the general metric case.

Proposition 7 below states that under mild assumptions on the existence and uniqueness of the pooled and the groupwise Fréchet means, the statistics $\hat{\mu}_p$ and F_n are consistent estimators of the population quantities μ_p and F. By our assumptions, F is zero only under the equality of the population Fréchet means and positive otherwise. The population quantity U is proportional to the weighted average of the pairwise differences between the groupwise Fréchet variances, which is nonnegative and is zero only if the population groupwise Fréchet variances are all equal.

Proposition 7. *Suppose $\hat{\mu}_p$, $\hat{\mu}_j$, μ_p and μ_j exist and are unique, the sample based estimators almost surely for all $j = 1,\dots,k$. Assume for any $\varepsilon > 0$, $\inf_{d(\omega,\mu_p)>\varepsilon} \sum_{j=1}^{k} \lambda_j E_j(d^2(\omega,Y_j)) > \sum_{j=1}^{k} \lambda_j E_j(d^2(\mu_p,Y_j))$ and also $\inf_{d(\omega,\mu_j)>\varepsilon} E_j(d^2(\omega,Y_j)) > E_j(d^2(\mu_j,Y_j))$ for all $j = 1,\dots,k$. Let $0 < \lambda_{j,n} < 1$ for all $j = 1,\dots,k$ and $\lambda_{j,n} \to \lambda_j$ as $n \to \infty$, where λ_j is such that $0 < \lambda_j < 1$ for each $j = 1,\dots,k$, with $\sum_{k=1}^{k} \lambda_j = 1$, as defined in Proposition 5. Then, as $n \to \infty$,*

$$\hat{\mu}_p \xrightarrow{P} \mu_p \quad \text{and} \quad F_n \xrightarrow{P} F. \tag{27}$$

The population quantity F is nonnegative and is zero iff the population Fréchet means μ_j are all equal.

To study the power performance of the proposed test (24), we consider sequences of alternatives H_n where $H_n = \{(U,F) : U \geq a_n \text{ or } F \geq b_n\}$ for non negative sequences $\{a_n\}$ or $\{b_n\}$ with either a_n or b_n strictly greater than 0. The case where either $a_n \to 0$ or $b_n \to 0$, as $n \to \infty$

reflects contiguous alternatives. Of interest is the asymptotic behavior of the power function β_{H_n}, where

$$\beta_{H_n} = \inf_{(U,F)\in H_n} P\left(R_{n,\alpha}\right). \tag{28}$$

For the following result we require assumption (P2). The examples considered here satisfy this assumptions, as shown in Propositions 1 and 2. The following result provides sufficient conditions for the consistency of the proposed test (24) under this family of contiguous alternatives, i.e., where its asymptotic power is 1 under these alternatives, for any choice of the level α.

Theorem 3. *Under the assumptions of Proposition 7 and assumption (P3), for sequences of contiguous alternatives $\{H_n\}$ for which either $F \geq a_n$ or $U \geq b_n$, where $a_n \to 0$ and $b_n \to 0$ as $n \to \infty$, for all $\alpha > 0$ the power function (28) satisfies*
(A) If $\sqrt{n}a_n \to \infty$, then $\beta_{H_n} \to 1$.
(B) If $nb_n \to \infty$, then $\beta_{H_n} \to 1$.

Asymptotic tests may not work very well for common situations where the group sample sizes are modest. By arguments similar to those provided at the end of Section 3, resampling methods like bootstrapping and permutation tests using the proposed test statistic T_n can be adopted to obtain more accurate level α tests. In our empirical experiments, which are presented in the next section, we found that when sample sizes of the groups are large the asymptotic test is stable and accurate, but for small sample sizes, using bootstrap or permutation based critical values for the test statistic T_n instead of the asymptotic critical values leads to more accurate inference.

5. SIMULATION STUDIES

In order to gauge the performance of the proposed test (24), we performed simulation experiments under various settings. The random objects we consider include samples of univariate distributions equipped with the L^2-Wasserstein metric, samples of graph Laplacians of scale free networks from the Barabási-Albert model (Barabási and Albert, 1999) with the Frobenius metric and samples of multivariate data with the usual Euclidean metric. In each case we considered two groups of equal size $n_1 = n_2 = 100$ and constructed the empirical power functions of the proposed test against departures from the null hypothesis of equality of population distributions of the two groups at level 0.05. The empirical power was computed by finding the proportion of rejections for 1000 Monte Carlo runs. For comparing the performance of the proposed test against existing tests we used the bootstrap version of the proposed test, where the critical value for the test was obtained from the bootstrap distribution of T_n in 1000 Monte Carlo simulations. The performance of the bootstrap version was found to have the correct level for all sample sizes while the asymptotic version of the test did not always produce the correct level of the test for very small sample sizes. We also investigated the finite sample power of the asymptotic test for increasing sample sizes by comparing power functions for group sizes $n_1 = n_2 = 100, 250, 450$.

In the simulations we explored not only location differences but also differences in shape and scale of the population distributions. We compared the proposed test (24) with the graph based test (Chen and Friedman, 2017), the energy test based on pairwise distances (Székely and Rizzo, 2004) and a kernel based test (Gretton et al., 2012). For the graph based test, we constructed the similarity graph of the pooled observations of the two groups 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 $1^{st}, \dots, k^{th}$ MSTs, where a k^{th} 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 $1^{st}, \dots, (k-1)^{th}$ MST. For computing the statistic of the energy test of Székely and Rizzo (2004), we used the pairwise distance matrix obtained from the specified metric in the space of random objects. For the kernel based method, we chose a Gaussian kernel with the kernel width as the median of the pairwise distances, as suggested in Gretton et al. (2012). In the multivariate Euclidean setting, for the two sample case we additionally compared the proposed test with Hotelling's T^2 test.

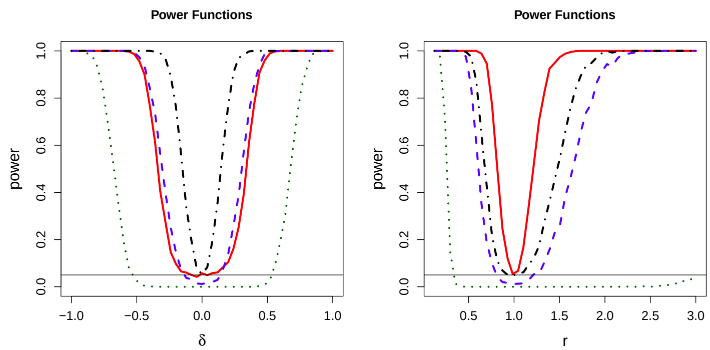

Figure 1: Empirical power as function of δ for $N(\mu, 1)$ probability distributions with μ from $N(0, 0.5)$ for group G_1 and $N(\delta, 0.5)$ for group G_2 (left), and empirical power as function of r for $N(\mu, 1)$ probability distributions with μ from $N(0, 0.2)$ for G_1 and $N(0, 0.2r)$ for G_2 (right). The solid red curve corresponds to the bootstrapped version of the proposed test with test statistic (22), the dashed blue curve to the graph based test of Chen and Friedman (2017) , the dot-dashed black curve to the energy test of Székely and Rizzo (2004) and the dotted green curve corresponds to the kernel test of Gretton et al. (2012). The level of the tests is $\alpha = 0.05$ and is indicated by the line parallel to the x-axis. Sample sizes of the groups are fixed at $n_1 = n_2 = 100$.

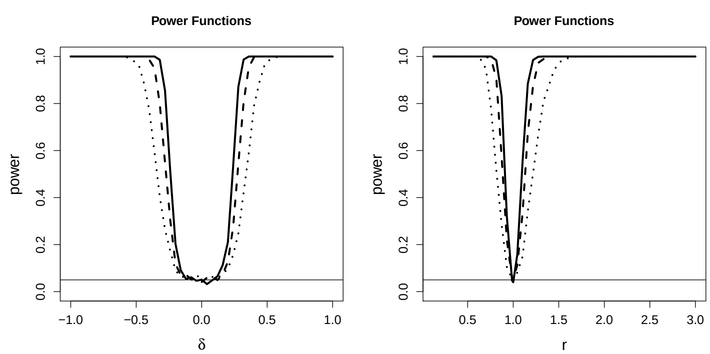

Figure 2: Empirical power as function of δ for $N(\mu, 1)$ probability distributions with μ from $N(0, 0.5)$ for group G_1 and $N(\delta, 0.5)$ for group G_2 (left), and empirical power as function of r for $N(\mu, 1)$ probability distributions with μ from $N(0, 0.2)$ for G_1 and $N(0, 0.2r)$ for G_2 (right), for the proposed test at different sample sizes. The tests are at level $\alpha = 0.05$, indicated by the line parallel to the x-axis. The solid curve corresponds to sample sizes $n_1 = n_2 = 450$, the dashed curve to $n_1 = n_2 = 250$ and the dotted curve to $n_1 = n_2 = 100$.

The first type of random 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 L^2-Wasserstein metric. In the first scenario, for group G_1, we generate μ to be distributed as $N(0,0.5)$ and for group G_2 as $N(\delta,0.5)$ and compute the empirical power function of the tests for $-1 \le \delta \le 1$. In the second scenario μ is drawn randomly from $N(0,0.2)$ for group G_1 and from $N(0,0.2r)$ for group G_2 and empirical power is evaluated for $0.125 \le r \le 3$. The first scenario emphasizes location differences between the populations and the second emphasizes scale differences. The results are presented in Figure 1. We find that in the first scenario of mean differences, the proposed test and the graph based test perform similarly. Both of them are outperformed by the energy test but perform better than the kernel based test. In the second scenario of scale differences the proposed test outperforms all other tests. Figure 2 indicates that the proposed test is consistent for large sample sizes in both scenarios, and Figure 3 indicates that for insufficient sample sizes, the bootstrap version of the proposed test has more stable rejection regions and overall is more reliable than the asymptotic version of the test (24). As sample sizes increase, the asymptotic test becomes more reliable and yields results that are similar to the bootstrap test.

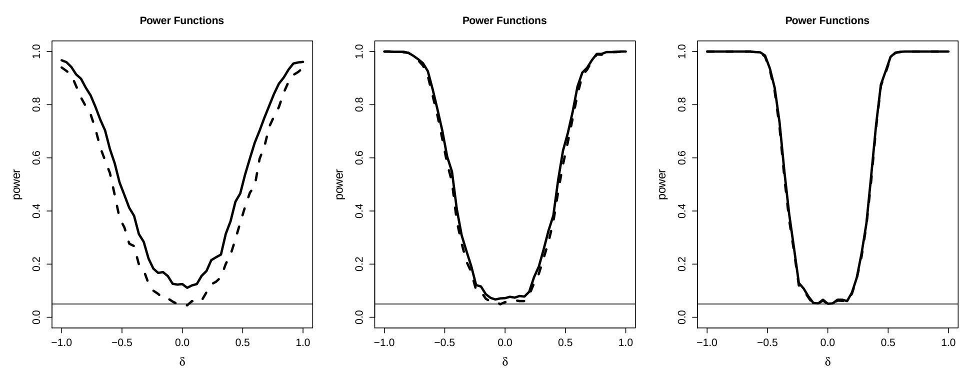

Figure 3: Empirical power as function of δ for $N(\mu,1)$ probability distributions with μ from $N(0,0.5)$ for group G_1 and $N(\delta,0.5)$ for group G_2. The leftmost panel corresponds to group sample sizes $n_1 = n_2 = 10$, the middle panel corresponds $n_1 = n_2 = 30$ and the rightmost panel to $n_1 = n_2 = 90$. The solid curve corresponds to the asymptotic version of the proposed test in (22) and the dashed curve corresponds to the bootstrap version. The level of the test is $\alpha = 0.05$ and is indicated by the line parallel to the x-axis.

Next we consider samples of 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 $2 \leq \gamma \leq 3$. Specifically, we used the Barabási-Albert algorithm to generate samples of scale free networks with 10 nodes, as one might encounter in brain networks. For group G_1, we set $\gamma = 2.5$ and for group G_2 we selected a fixed γ in the interval $2 \leq \gamma \leq 3$, studying the empirical power as a function of γ. The left panel in Figure 4 indicates that in this scenario the proposed test has better power behavior than both the graph based test and the kernel based test. The kernel based test with automatic scaling parameter choice (Gretton et al., 2012) in the published software has very low power, while the graph based test has a high false positive rate. The right panel in Figure 4 shows empirical evidence that the proposed test is also consistent in this scenario as sample size increases, and Figure 5 that especially for small samples, bootstrapping the test statistic leads to the correct empirical level of the test. With increasing sample size, the asymptotic and the bootstrap versions of the test perform similarly.

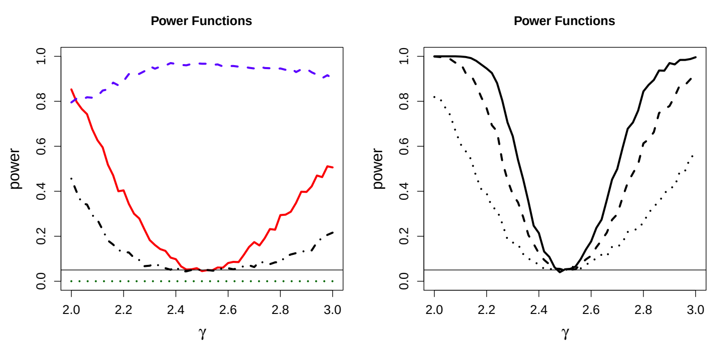

Figure 4: Empirical power functions of γ for scale-free networks from the Barabási-Albert model with parameter 2.5 for G_1 and γ for G_2. In the left panel, the solid red curve corresponds to the bootstrapped version of our proposed test with test statistic (22), the blue dashed curve to the graph based test in Chen and Friedman (2017), the dot-dashed black curve to the energy test of Székely and Rizzo (2004) and the green dotted curve to the kernel test in Gretton et al. (2012). Sample sizes are fixed at $n_1 = n_2 = 100$. In the right panel, the solid power function corresponds to the proposed asymptotic test (24) for $n_1 = n_2 = 450$, the dashed power function to the test for $n_1 = n_2 = 250$ and the dotted power function to the test for $n_1 = n_2 = 100$. The level of the tests is $\alpha = 0.05$ and is indicated by the line parallel to the x-axis.

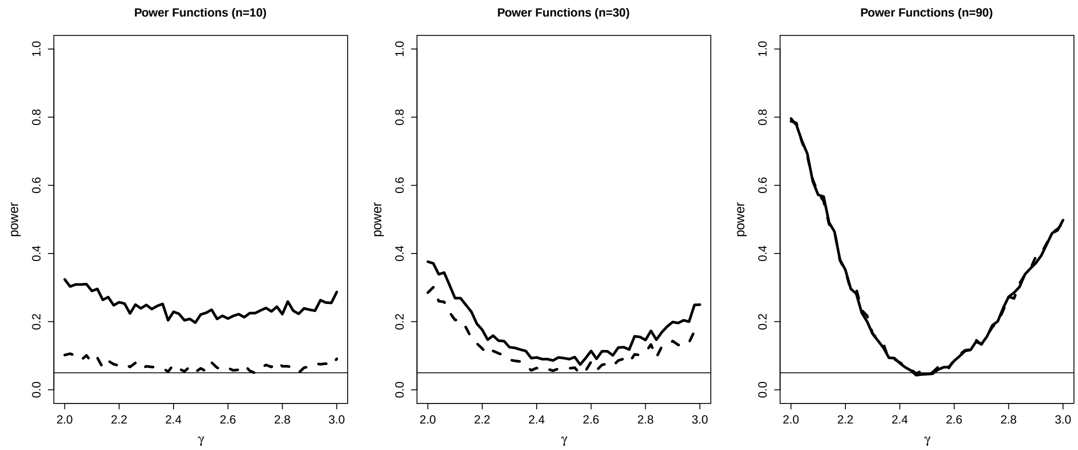

Figure 5: Empirical power as function of γ for scale-free networks from the Barabási-Albert model, with the model parameter 2.5 for G_1 and γ for G_2. The leftmost panel corresponds to group sample sizes $n_1 = n_2 = 10$, the middle panel to $n_1 = n_2 = 30$ and the rightmost panel to $n_1 = n_2 = 90$. The solid curve corresponds to the proposed asymptotic test (24) and the dashed curve to the bootstrap version of the proposed test at level $\alpha = 0.05$. The level is indicated by the line parallel to the x-axis.

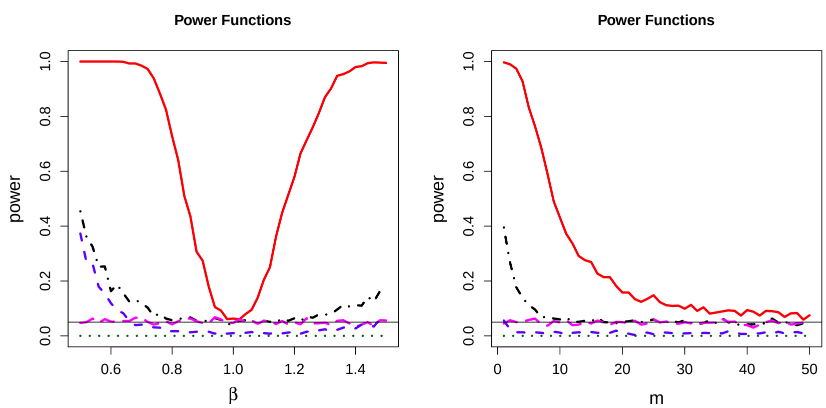

Figure 6: Empirical power as function of β for 5-dimensional vectors, where each component is distributed independently as $Beta(1,1)$ for group G_1 and as $Beta(\beta,\beta)$ with β varying between $0.5 \leq \beta \leq 1.5$ for group G_2 (left). Empirical power as function of degrees of freedom m for 5-dimensional vectors which are distributed independently as truncated $N(0, I_5)$ for group G_1 and as truncated multivariate t-distribution $t_m(0, I_5)$ with varying degrees of freedom between $1 \leq m \leq 50$ for group G_2 with each component of the vectors truncated to lie between $[-5, 5]$ (right). Sample sizes are $n_1 = n_2 = 100$ and level $\alpha = 0.05$. Solid red curves correspond to the bootstrapped version of our proposed test with test statistic (22), dashed blue curves to the graph based test of Chen and Friedman (2017), dotted green curves corresponds to the kernel test of Gretton et al. (2012), dot-dashed black curves to the energy test of Székely and Rizzo (2004) and long-dashed magenta curves to Hotelling's T^2 test.

In the multivariate setting we considered 5-dimensional vectors distributed as truncated multivariate normal distributions $N(0, I_5)$ for group G_1 where each of the components was truncated to lie between $[-5, 5]$. For group G_2, we chose a 5-dimensional t-distribution $t_m(0, I_5)$, m indicating the degrees of freedom. As the degrees of freedom m increases, the shape of the distribution of G_2 becomes more similar to that of group G_1. We obtained the empirical power as functions of m, for $1 \leq m \leq 50$. In a second scenario, we took the five components of the vectors to be distributed independently as $Beta(1, 1)$ for group G_1, while for group G_2, the five components were assumed to be distributed independently as $Beta(\beta, \beta)$. Empirical power was then obtained as a function of β for $0.5 \leq \beta \leq 1.5$. Figure 6 illustrates that for these cases, the proposed test overwhelmingly outperforms the comparison tests.

6. DATA ILLUSTRATIONS

6.1. Mortality Data

The Human Mortality Database provides data in the form of yearly lifetables differentiated by countries and gender. Presently it includes yearly mortality data for 37 countries, available at <www.mortality.org>. These can be converted to a density of age-at-death for each country, gender and calendar year, by first converting the available lifetables into histograms and then applying local least squares smoothing, for which we used the Hades package at <http://www.stat.ucdavis.edu/hades/> with bandwidth $h = 2$. The random objects we consider are the resulting densities of age-at-death. Considering the time period 1960-2009 and the 31 countries in the database for which records are available for this time period, we obtained the densities of age at death for the age interval $[0, 80]$.

From these densities we obtained quantile functions to compute the Wasserstein distance, which is the metric we choose for this distribution space. The Fréchet standard deviations were computed as a function of calendar year, and are shown along with pointwise 95 % bootstrap confidence bands in Figure 7, separately for males and females. One finds that there is a small peak in variance of mortality between 1980-1985 for males followed by a larger peak between

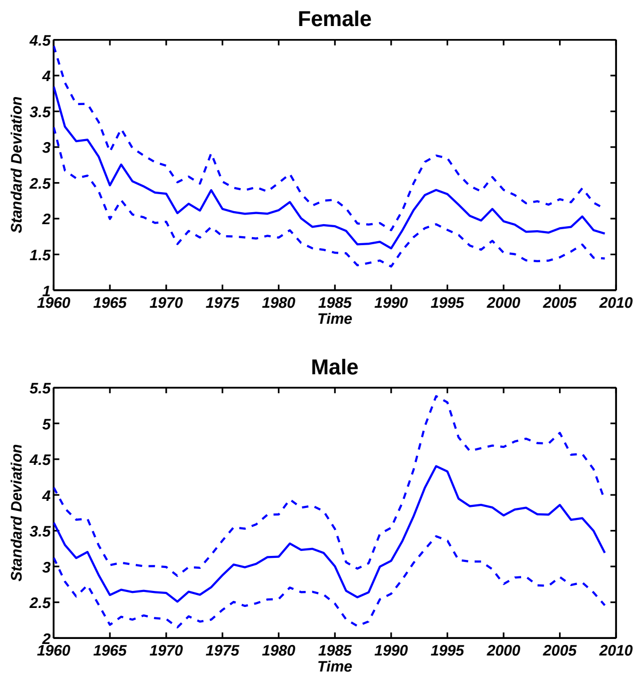

Figure 7: Yearly Fréchet standard deviations (solid line) along with 95 % pointwise bootstrap confidence limits (dashed lines) for females (top) and males (bottom).

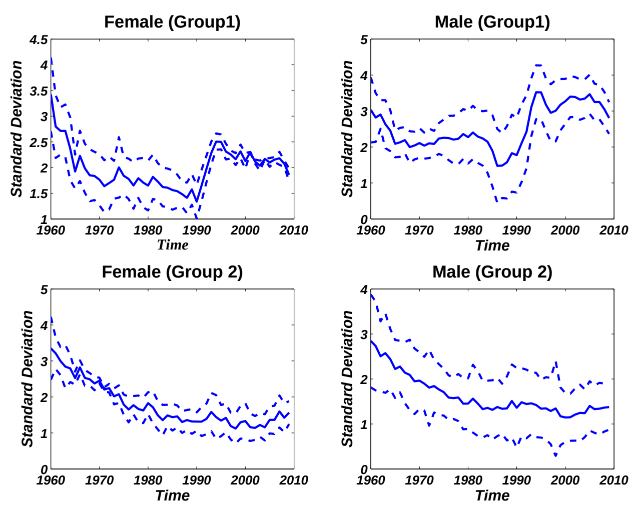

Figure 8: Yearwise Fréchet standard deviations (solid line) along with 95% pointwise bootstrap confidence limits (dashed lines) for females in group 1, females in group 2, males in group 2 and males in group 1 (counter clockwise starting from top left)

1993-1996. For females, this later peak is also quite prominent. These peaks might possibly be attributed to major political upheaval in Central and Eastern Europe during that period since several countries in the dataset belong to these regions. The countries in the dataset that experienced some turmoil associated with the end of Communist role are Belarus, Bulgaria, Czech Republic, Estonia, Hungary, Latvia, Poland, Lithuania, Russia, Slovakia and Ukraine.

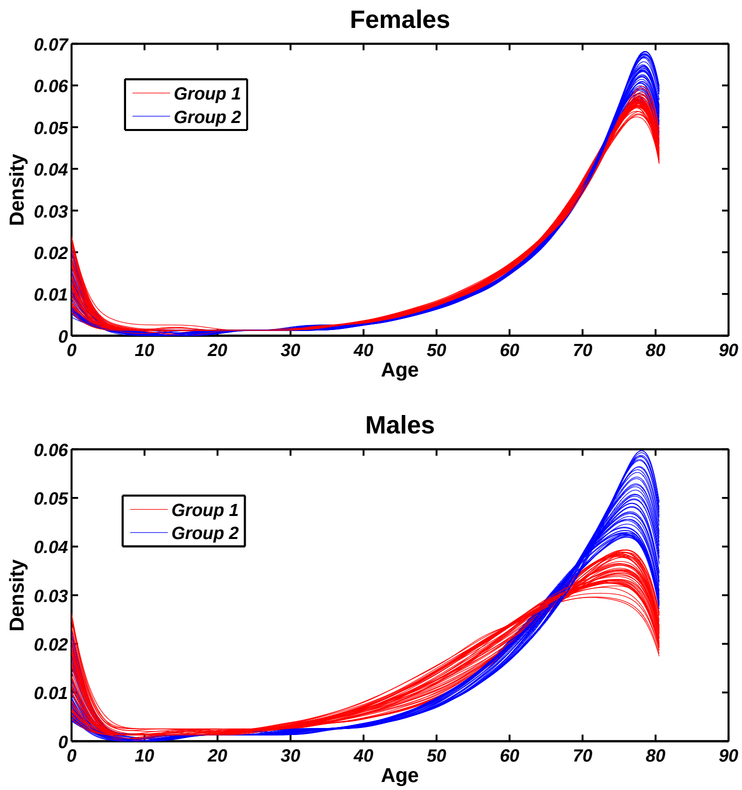

Figure 9: Wasserstein-Fréchet mean age-at-death densities for the years 1960-2009 for groups 1 (red) and 2 (blue) and for females (top) and males (bottom)

To check whether it is indeed these countries that are responsible for the variance peak around 1990-1995, we split our dataset into two groups, group 1 consisting of the above Eastern European countries and group 2 of all other countries and repeated the earlier analysis. Figure 8 shows that for group 2 the variance of age-at-death distributions indeed has a decreasing trend over the years for both males and females, while the variance shows distinct

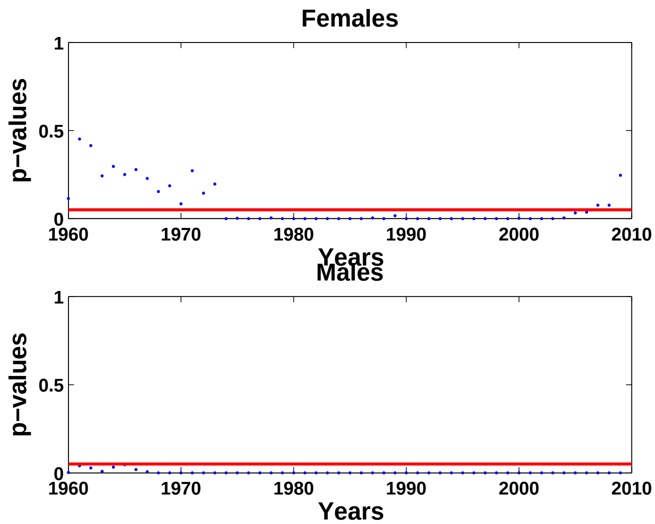

Figure 10: p-values for testing the differences in population age-at-death distributions of Groups 1 and 2 over the years for females (top) and males (bottom) with the proposed test.

fluctuations for both males and females in group 1. Figure 9 illustrates the group-wise Wasserstein Fréchet mean densities of the countries for the various calendar years. There seems to be a clear difference between the mean densities of the two groups for both males and females, and implementing the bootstrap version of the proposed test to compare the distributions of age-at-death between groups 1 and 2. For obtaining the p-values we carried out the bootstrap version of the proposed test due to the relatively small sample sizes. Figure 10 illustrates the p-values obtained for each year. The null hypothesis that the populations are identical is rejected for both males and females at the 5% level for most years between 1990-1995.

6.2. Analyzing intra-hub connectivities using fMRI data for Alzheimer's Disease

Alzheimers disease (AD) is an irreversible, progressive neuro-degenerative brain disorder that slowly destroys memory and thinking skills, eventually leading to severe dementia. AD has been found to have associations with abnormalities in functional integration of brain regions. Recent studies as in Sui et al. (2015) have indicated that AD selectively targets regions of high-connectivity (so-called hubs) in the brain. The posterior midline, in particular the posterior cingulate/precuneus (PCP) as described in Buckner et al. (2009) is a nexus or hub of high cortical connectivity and functional connectivity in this region could be a potential biomarker for AD. For each hub, a so-called seed voxel is identified as the voxel with the signal that has the highest correlation with the signals of nearby voxels. To quantify intra-hub connectivity, following Petersen and Müller (2016), we analyze the distribution of the correlations between the signal at the seed voxel of the PCP hub and the signals of all other voxels within an $11 \times 11 \times 11$ cube of voxels that is centered at the seed voxel.

The subjects in our analysis consisted of cognitively normal elderly patients and demented elderly patients diagnosed with AD (after removal of outliers), each of whom underwent an fMRI scan at the UC Davis Imaging Research Center. Preprocessing of the recorded BOLD (blood-oxygenation-level-dependent) signals was implemented by adopting the standard procedures of slice-timing correction, head motion correction and normalization, in addition to linear detrending to account for signal drift and band-pass filtering to include only frequencies between 0.01 and 0.08 Hz. The signals for each subject were recorded over the interval [0, 470] (in seconds), with 236 measurements available at 2 second intervals. The study actually had 171 normal subjects but since AD is a disease that is known to progress with age, for fair comparison, only 87 out of these 171 were selected, by matching their ages with that of the demented patients. To check that the age matching worked, the age distributions of the 87 normal elderly subjects in our analysis and the 65 AD patients were compared with the Wilcoxon rank sum test for the null hypothesis of equal age distributions of the two groups, which yielded a p-value of 0.84. For each subject, the target is the density function of positive correlations within the PCP hub, where this density was estimated from the observed correlations with a kernel density estimator, utilizing the standard Gaussian kernel and bandwidth $h = 0.08$. As negative correlations are commonly ignored in connectivity analyses, the densities were estimated on [0, 1]. The resulting sample of densities is then an i.i.d. sample across subjects. Figure 11 shows the Wasserstein Fréchet mean probability distribution functions. To compare the two populations of distributions, we applied the asymptotic and the bootstrap version of the proposed test to these

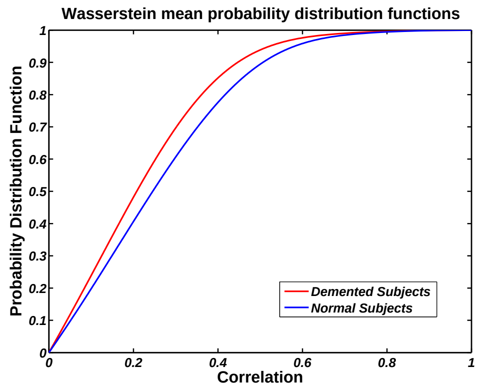

Figure 11: Wasserstein mean probability distribution functions of positive correlations in the PCP hub for normal subjects (blue) and demented patients (red)

samples of density functions, which yielded a p-value of $p = 0.002$ (bootstrap p-value=0.001), indicating that significant differences exist between in terms of intra-hub connectivity between AD patients and age-matched normal subjects.

6.3. Comparing brain networks of Alzheimer's patients

Recent advances in neurological studies have revealed that brain hubs, being regions of high connectivity in the brain, interconnect with each other for functional integration of their specialized roles. Studying interconnections between hubs can reveal important insights about brain diseases like AD. Disorders of cognition can be associated with disrupted connectivity between cortical hubs as discussed in Buckner et al. (2009). One question of interest is whether the interconnections change with aging in patients having dementia. We consider connections between 10 cortical hubs that are listed in Table 3 of Buckner et al. (2009).

In order to analyze the cognitively impaired patients we considered the 65 demented subjects that were discussed in the preceding subsection. For each subject, a 10×10 connectivity matrix was obtained whose entries are the correlations between average fMRI signals (with the same pre-processing as described in subsection 6.2) from $3 \times 3 \times 3$ cubes around the seed voxels of the 10 hubs. These subject-specific connectivity matrices were thresholded at 0.25 as discussed in Buckner et al. (2009) to obtain adjacency matrices of networks with the hubs as the nodes, so that the presence of an edge indicates a correlation greater than 0.25. Subject-specific graph Laplacians were then formed from these adjacency matrices.

These cognitively impaired patients were split into three groups based on their ages. Subjects were assigned to groups G_1, G_2 or G_3 based on whether they were aged 70 or below, between age 70 and 80 or 80 and above. The left panel in Figure 12 shows the difference of the average graph Laplacians of subjects in group G_2 and subjects in group G_1 and the right panel the difference of the average graph Laplacians of subjects in group G_3 and subjects in group G_1.

Since the group sample sizes of G_1, G_2 and G_3 are small we applied the bootstrap version of the proposed test to see if the differences are significant in the populations of graph Laplacians of the three different groups of cognitively impaired patients. The null hypothesis of equality of population distributions of the graph Laplacians was rejected with a bootstrap p-value of 0.032. The conclusion of the test indicates that there is evidence to support that the inter hub connections do show changes with age for patients having Alzheimer's disease.

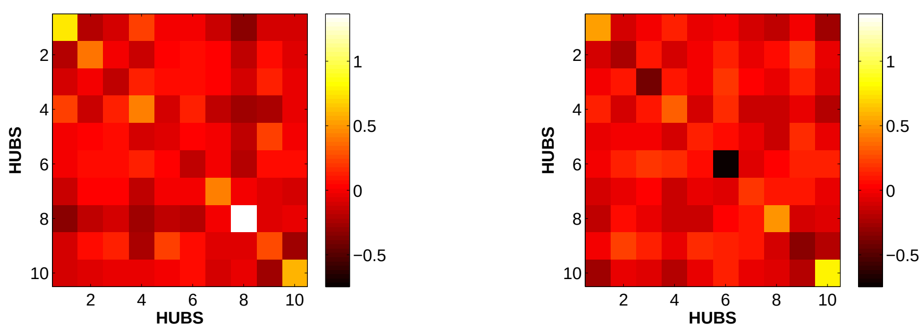

Figure 12: The left panel shows the difference between the average graph Laplacians of demented patients aged between 70 and 80 and demented patients aged 70 or below. The right panel shows the difference between the average graph Laplacians of demented patients aged 80 or above and demented patients aged 70 or below.

REFERENCES

Barabási, A.-L. and Albert, R. (1999), "Emergence of scaling in random networks," *Science*, 286, 509–512.

Barden, D., Le, H., and Owen, M. (2013), "Central limit theorems for Fréchet means in the space of phylogenetic trees," *Electronic Journal of Probability*, 18, 1–25.

Baringhaus, L. and Franz, C. (2004), "On a new multivariate two-sample test," *Journal of Multivariate Analysis*, 88, 190–206.

Beran, R. (2003), "The impact of the bootstrap on statistical algorithms and theory," *Statistical Science*, 18, 175–184.

Bhattacharya, A. and Bhattacharya, R. (2012), *Nonparametric Inference on Manifolds: With Applications to Shape Spaces*, no. 2, Cambridge University Press.

Bhattacharya, R. and Lin, L. (2017), "Omnibus CLTs for Fréchet means and nonparametric inference on non-Euclidean spaces," *Proceedings of the American Mathematical Society*, 145, 413–428.

Bhattacharya, R. and Patrangenaru, V. (2003), "Large sample theory of intrinsic and extrinsic sample means on manifolds. I," *Annals of Statistics*, 31, 1–29.

— (2005), "Large sample theory of intrinsic and extrinsic sample means on manifolds: II," *Annals of Statistics*, 33, 1225–1259.

Bickel, P. J. and Freedman, D. A. (1981), "Some asymptotic theory for the bootstrap," *Annals of Statistics*, 9, 1196–1217.

Buckner, R. L., Sepulcre, J., Talukdar, T., Krienen, F. M., Liu, H., Hedden, T., Andrews-Hanna, J. R., Sperling, R. A., and Johnson, K. A. (2009), "Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer's disease," *The Journal of Neuroscience*, 29, 1860–1873.

Chen, H., Chen, X., and Su, Y. (2017), "A weighted edge-count two-sample test for multivariate and object data," *Journal of the American Statistical Association*.

Chen, H. and Friedman, J. H. (2017), "A new graph-based two-sample test for multivariate and object data," *Journal of the American Statistical Association*, 112, 397–409.

Fréchet, M. (1948), "Les éléments aléatoires de nature quelconque dans un espace distancié," 10, 215–310.

Friedman, J. H. and Rafsky, L. C. (1979), "Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests," *Annals of Statistics*, 7, 697–717.

Ginestet, C. E., Li, J., Balachandran, P., Rosenberg, S., Kolaczyk, E. D., et al. (2017), "Hypothesis testing for network data in functional neuroimaging," *The Annals of Applied Statistics*, 11, 725–750.

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., and Smola, A. (2012), "A kernel two-sample test," *Journal of Machine Learning Research*, 13, 723–773.

Guo, R. and Patrangenaru, V. (2017), "Testing for the Equality of two Distributions on High

Dimensional Object Spaces," *arXiv preprint arXiv:1703.07856*.

Henze, N. (1988), "A multivariate two-sample test based on the number of nearest neighbor type coincidences," *Annals of Statistics*, 16, 772–783.

Henze, N. and Penrose, M. D. (1999), "On the multivariate runs test," *Annals of Statistics*, 27, 290–298.

Janssen, A. and Pauls, T. (2003), "How do bootstrap and permutation tests work?" *Annals of Statistics*, 31, 768–806.

Kendall, W. S. and Le, H. (2011), "Limit theorems for empirical Fréchet means of independent and non-identically distributed manifold-valued random variables," *Brazilian Journal of Probability and Statistics*, 25, 323–352.

Levene, H. (1960), "Robust tests for equality of variances," *Contributions to Probability and Statistics*, 1, 278–292.

Lyons, R. (2013), "Distance covariance in metric spaces," *Annals of Probability*, 41, 3284–3305.

Patrangenaru, V. and Ellingson, L. (2015), *Nonparametric statistics on manifolds and their applications to object data analysis*, CRC Press.

Petersen, A. and Müller, H.-G. (2016), "Functional data analysis for density functions by transformation to a Hilbert space," *Annals of Statistics*, 44, 183–218.

— (2017), "Fréchet Regression for Random Objects with Euclidean predictors," *Annals of Statistics, to appear (arXiv preprint arXiv:1608.03012)*, xx, xxx–xxx.

Ramdas, A., Trillos, N. G., and Cuturi, M. (2017), "On Wasserstein Two-Sample Testing and Related Families of Nonparametric Tests," *Entropy*, 19, 47.

Rosenbaum, P. R. (2005), "An exact distribution-free test for comparing two multivariate distributions based on adjacency," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67, 515–530.

Schilling, M. F. (1986), "Multivariate two-sample tests based on nearest neighbors," *Journal of the American Statistical Association*, 81, 799–806.

Sejdinovic, D., Sriperumbudur, B., Gretton, A., and Fukumizu, K. (2013), "Equivalence of distance-based and RKHS-based statistics in hypothesis testing," *Annals of Statistics*, 41, 2263–2291.

Sui, X., Zhu, M., Cui, Y., Yu, C., Sui, J., Zhang, X., Liu, J., Duan, Y., Zhang, Z., Wang, L., Zhang, X., and Jiang, T. (2015), "Functional connectivity hubs could serve as a potential biomarker in Alzheimers disease: a reproducible study," *Current Alzheimer Research*, 12, 974–983.

Szarek, S. J. (1998), "Metric Entropy of Homogeneous Spaces," *Quantum Probability*, 43, 395–

410.
Székely, G. J. and Rizzo, M. L. (2004), “Testing for equal distributions in high dimension,” *InterStat*, 5, 1–6.
— (2017), “The Energy of Data,” *Annual Review of Statistics and Its Application*, 4, 447–479.
Villani, C. (2003), *Topics in Optimal Transportation*, American Mathematical Society.
Wellner, J. A. and van der Vaart, A. W. (1996), *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer Series in Statistics.

SUPPLEMENTARY MATERIALS

S.1 Main Proofs

Proof of Proposition 1. Let Q_Ω be the space of quantile functions corresponding to the space Ω of univariate distribution functions on $\mathcal{R}$ with finite second moments. For a random object Y taking values in Ω, let Q_Y denote the corresponding quantile function. By the convexity of the space Q_Ω and the properties of the L^2-Wasserstein distance, the population and sample Fréchet means exist and are unique and given by $\mu_F = \left(E(Q_Y(\cdot))\right)^{-1}$ and $\hat{\mu}_F = \left(\frac{1}{n}\sum_{i=1}^n Q_{Y_i}(\cdot)\right)^{-1}$ thereby proving (P0). For proving (P1) and (P2) observe that quantile functions are a part of a bigger class M comprising of monotone functions and under L^2 metric d_{L^2} (the Wasserstein metric corresponds to the L^2 metric on the space of quantile functions) the metric entropy of the space M is upper bounded as $\log N(\varepsilon, M, d_{L^2}) \leq \frac{a}{\varepsilon}$ for some constant $a > 0$ as per Theorem 2.7.5 in Wellner and van der Vaart (1996). Therefore,

$$J(\delta) \leq \int_0^1 \sqrt{1+\frac{1}{\varepsilon\delta}} d\varepsilon \leq \int_0^1 \left(1+\frac{1}{\sqrt{\varepsilon\delta}}\right) d\varepsilon = 1+\frac{2}{\sqrt{\delta}},$$

and so $\delta\, J(\delta) \to 0$ as $\delta \to 0$, thus establishing (P1). The entropy integral of the whole space is

$$\int_0^1 \sqrt{1+\log N(\varepsilon, \Omega, d_W)} \leq \int_0^1 \sqrt{1+\frac{1}{\varepsilon}} d\varepsilon = 3,$$

which establishes (P2). $\square$

Proof of Proposition 2. The space of graph Laplacians for graphs considered in this Proposition is convex (Ginestet et al., 2017) and so is the space of correlation matrices. The properties of Frobenius distance imply that $\mu_F = E(Y)$ and $\hat{\mu}_F = n^{-1}\sum_{i=1}^n Y_i$, which exist and are unique by the convexity of Ω proving (P0). (P1) and P(2) hold as both the space of graph Laplacians and the space of correlation matrices are bounded subsets of a finite-dimensional Euclidean space R^{r^2} of $r\times r$ matrices and under the Euclidean metric d_E, which is equivalent to the Frobenius metric for the matrices. The metric entropy of this space is seen to be bounded above by $\log N(\varepsilon, R^{r^2}, d_E) \leq ar^2 \log\left(1+\frac{1}{\varepsilon}\right)$ for some constant $a>0$, due to a simple volume comparison argument (Szarek, 1998), so that

$$J(\delta) \leq \int_0^1 \sqrt{1+ar^2\log\left(1+\frac{1}{\varepsilon\delta}\right)}d\varepsilon \leq \int_0^1 \left(1+\sqrt{a}r\sqrt{\log\left(1+\frac{1}{\varepsilon\delta}\right)}\right)d\varepsilon \leq 1+\sqrt{a}r(1+\frac{2}{\sqrt{\delta}}),$$

implying $\delta J(\delta)\to 0$ as $\delta\to 0$ and establishing (P1). Using similar upper bounds for the entropy integral of the whole space,

$$\int_0^1 \sqrt{1+\log N(\varepsilon,\Omega,d_F)} \leq \int_0^1 \sqrt{1+ar^2\log\left(1+\frac{1}{\varepsilon}\right)}d\varepsilon \leq 1+3\sqrt{a}r,$$

establishing (P2). □

Proof of Proposition 3. Define

$$M_n(\omega) = \frac{1}{n}\sum_{i=1}^n \{d^2(\omega,Y_i) - d^2(\mu_F,Y_i) - E(d^2(\omega,Y_i)) + E(d^2(\mu_F,Y_i))\}.$$

In a first step we control the behavior of $M_n(\omega)$ uniformly for small $d(\omega,\mu_F)$. Define functions $g_\omega : \Omega \to R$ as $g_\omega(y) = d^2(y,\omega)$ and the function class $M_\delta = \{g_\omega - g_{\mu_F} : d(\omega,\mu_F) < \delta\}$. An envelope function for M_δ is $G(\delta) = 2\text{diam}(\Omega)\delta$. Let $J = J(\delta)$ be the integral in (P1) so that $\delta J(\delta)\to 0$ as $\delta\to 0$. Theorems 2.7.11 and 2.14.2 of Wellner and van der Vaart (1996) and (P1) imply that for small $\delta > 0$,

$$E\left(\sup_{d(\omega,\mu_F)<\delta} |M_n(\omega)|\right) \leq \frac{J(\delta)G(\delta)}{\sqrt{n}},$$

therefore

$$E\left(\sup_{d(\omega,\mu_F)<\delta} |M_n(\omega)|\right) \leq \frac{a\delta\, J(\delta)}{\sqrt{n}} \tag{29}$$

for some $a > 0$. We want to show that for any $\varepsilon > 0$, $\gamma > 0$, there exists $N = N(\varepsilon, \gamma)$ such that for all $n \geq N$,

$$P\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{d^2(\hat{\mu}_F, Y_i) - d^2(\mu_F, Y_i)\}\right| > \varepsilon\right) < \gamma. \tag{30}$$

For any small $\delta > 0$,

$$\begin{aligned}
&P\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{d^2(\hat{\mu}_F, Y_i) - d^2(\mu_F, Y_i)\}\right| > \varepsilon\right) \\
\leq &P\left(\frac{1}{n}\sum_{i=1}^{n}\{d^2(\hat{\mu}_F, Y_i) - d^2(\mu_F, Y_i)\} < -\frac{\varepsilon}{\sqrt{n}}, d(\hat{\mu}_F, \mu_F) \leq \delta\right) + P\left(d(\hat{\mu}_F, \mu_F) > \delta\right) \\
\leq &P\left(-\inf_{d(\omega,\mu_F)<\delta}\frac{1}{n}\sum_{i=1}^{n}\{d^2(\omega, Y_i) - d^2(\mu_F, Y_i)\} > \frac{\varepsilon}{\sqrt{n}}\right) + P\left(d(\hat{\mu}_F, \mu_F) > \delta\right) \\
\leq &P\left(\sup_{d(\omega,\mu_F)<\delta} |M_n(\omega)| > \frac{\varepsilon}{\sqrt{n}}\right) + P\left(d(\hat{\mu}_F, \mu_F) > \delta\right),
\end{aligned}$$

since

$$\begin{aligned}
&\sup_{d(\omega,\mu_F)<\delta} |M_n(\omega)| \\
= &\sup_{d(\omega,\mu_F)<\delta}\left|\frac{1}{n}\sum_{i=1}^{n}\{d^2(\omega, Y_i) - d^2(\mu_F, Y_i) - E(d^2(\omega, Y_i)) + E(d^2(\mu_F, Y_i))\}\right| \\
\geq &\inf_{d(\omega,\mu_F)<\delta} E(d^2(\omega, Y_i)) - E(d^2(\mu_F, Y_i)) - \inf_{d(\omega,\mu_F)<\delta}\frac{1}{n}\sum_{i=1}^{n}\{d^2(\omega, Y_i) - d^2(\mu_F, Y_i)\} \\
= &-\inf_{d(\omega,\mu_F)<\delta}\frac{1}{n}\sum_{i=1}^{n}\{d^2(\omega, Y_i) - d^2(\mu_F, Y_i)\},
\end{aligned}$$

using

$$\inf_{d(\omega,\mu_F)<\delta}\{E(d^2(\omega,Y_i))-E(d^2(\mu_F,Y_i))\}=0 \quad \text{from assumption } (P0).$$

By using Markov's inequality and the bound in equation (29), for any small $\delta>0$ such that $\delta\, J(\delta)<\frac{\gamma\varepsilon}{2a}$, the expression in (30) can be bounded above by

$$\begin{aligned}
&P\left(\sup_{d(\omega,\mu_F)<\delta}|M_n(\omega)|>\frac{\varepsilon}{\sqrt{n}}\right)+P\left(d(\hat{\mu}_F,\mu_F)>\delta\right)\\
\leq &E\left(\sup_{d(\omega,\mu_F)<\delta}|M_n(\omega)|\right)\frac{\sqrt{n}}{\varepsilon}+P\left(d(\hat{\mu}_F,\mu_F)>\delta\right)\\
\leq &\frac{a\delta\, J(\delta)}{\varepsilon}+P\left(d(\hat{\mu}_F,\mu_F)>\delta\right)<\frac{\gamma}{2}+P\left(d(\hat{\mu}_F,\mu_F)>\delta\right).
\end{aligned}$$

For any such δ, using the consistency of Fréchet mean $\hat{\mu}_F$ it is possible to choose N such that $P\left(d(\hat{\mu}_F,\mu_F)>\delta\right)<\frac{\gamma}{2}$ for all $n\geq N$. This completes the proof. □

Proof of Theorem 1.

$$\begin{aligned}
&\sqrt{n}(\hat{V}_F-V_F)\\
&=\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\{d^2(\hat{\mu}_F,Y_i)-d^2(\mu_F,Y_i)\}\right)+\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\{d^2(\mu_F,Y_i)-E(d^2(\mu_F,Y_1))\}\right),
\end{aligned}$$

where the first term is $o_P(1)$ by Proposition 3 and the second term converges in distribution to $N(0,\sigma_F^2)$ by applying Central Limit Theorem to i.i.d random variables $d^2(\mu_F,Y_1),\ldots,d^2(\mu_F,Y_n)$. Theorem 1 then follows directly from Slutsky's Theorem. □

Proof of Proposition 4. Observe that

$$\left(\frac{1}{n}\sum_{i=1}^{n}d^4(\hat{\mu}_F,Y_i)-\frac{1}{n}\sum_{i=1}^{n}d^4(\mu_F,Y_i)\right)\leq\ 2\mathrm{diam}^2(\Omega)\left(\frac{1}{n}\sum_{i=1}^{n}d^2(\hat{\mu}_F,Y_i)-\frac{1}{n}\sum_{i=1}^{n}d^2(\mu_F,Y_i)\right) \quad (31)$$

which is $o_p\left(\frac{1}{\sqrt{n}}\right)$ by Proposition 3. Hence $\sqrt{n}\left[\begin{pmatrix}\frac{1}{n}\sum_{i=1}^n d^4(\hat{\mu}_F, Y_i)\\ \frac{1}{n}\sum_{i=1}^n d^2(\hat{\mu}_F, Y_i)\end{pmatrix} - \begin{pmatrix}E\left(d^4(\mu_F, Y)\right)\\ E\left(d^2(\mu_F, Y)\right)\end{pmatrix}\right]$ can be decomposed into two components, $\sqrt{n}\left[\begin{pmatrix}\frac{1}{n}\sum_{i=1}^n d^4(\hat{\mu}_F, Y_i)\\ \frac{1}{n}\sum_{i=1}^n d^2(\hat{\mu}_F, Y_i)\end{pmatrix} - \begin{pmatrix}\frac{1}{n}\sum_{i=1}^n d^4(\mu_F, Y_i)\\ \frac{1}{n}\sum_{i=1}^n d^2(\mu_F, Y_i)\end{pmatrix}\right]$ which is $o_P(1)$ by equation (31) and Proposition 3 and $\sqrt{n}\left[\begin{pmatrix}\frac{1}{n}\sum_{i=1}^n d^4(\mu_F, Y_i)\\ \frac{1}{n}\sum_{i=1}^n d^2(\mu_F, Y_i)\end{pmatrix} - \begin{pmatrix}E\left(d^4(\mu_F, Y)\right)\\ E\left(d^2(\mu_F, Y)\right)\end{pmatrix}\right]$ which converges in distribution to $N(0, D)$ by applying Central Limit Theorem to i.i.d random vectors $\begin{pmatrix}d^4(\mu_F, Y_i)\\ d^2(\mu_F, Y_i)\end{pmatrix}$, $i = 1, 2, \ldots, n$, with

$$D = \begin{pmatrix}\operatorname{Var}\left(d^4(\mu_F, Y)\right) & \operatorname{Cov}\left(d^4(\mu_F, Y), d^2(\mu_F, Y)\right)\\ \operatorname{Cov}\left(d^4(\mu_F, Y), d^2(\mu_F, Y)\right) & \operatorname{Var}\left(d^2(\mu_F, Y)\right)\end{pmatrix}.$$

Observing

$$\hat{\sigma}_F^2 = g\left(\frac{1}{n}\sum_{i=1}^n d^4(\hat{\mu}_F, Y_i), \frac{1}{n}\sum_{i=1}^n d^2(\hat{\mu}_F, Y_i)\right),$$

where $g(x_1, x_2) = x_1 - x_2^2$ is a differentiable function with gradient function $\triangledown g = (1, -2x_2)$, a simple application of the delta method establishes the asymptotic normality of $\hat{\sigma}_F^2$. The asymptotic variance is given by $a'Da$, where a is the gradient function $\triangledown g$ evaluated at $\begin{pmatrix}E\left(d^4(\mu_F, Y)\right)\\ E\left(d^2(\mu_F, Y)\right)\end{pmatrix}$. □

Proof of Proposition 5. Under the null hypothesis the groupwise means are all equal,

$$\mu_1 = \mu_2 = \cdots = \mu_k = \mu.$$

Then, under assumptions (P0) and (P1),

$$
\begin{aligned}
&\sqrt{n}F_n = \sqrt{n}(\hat{V} - \sum_{j=1}^{k} \lambda_{j,n}\hat{V}_j) \\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{d^2(\hat{\mu},Y_i) - d^2(\mu,Y_i)\} - \sum_{j=1}^{k}\frac{\sqrt{n_j}}{\sqrt{n}}\frac{1}{\sqrt{n_j}}\sum_{i\in G_j}\{d^2(\hat{\mu}_j,Y_i) - d^2(\mu,Y_i)\} \qquad (32) \\
=&o_p(1), \qquad (33)
\end{aligned}
$$

where (33) follows from (32) by applying Proposition 3 to all observations and also for the individual groups and noting that $\frac{\sqrt{n_j}}{\sqrt{n}} \to \sqrt{\lambda_j}$ for all $j = 1,2,...,k$. Slutsky's theorem completes the proof. □

Proof of Proposition 6. Under the null hypothesis, let $\mu_1 = \mu_2 = \cdots = \mu_k = \mu$ and $V_1 = V_2 = \cdots = V_k = V$. Using Proposition 4 and since $\lambda_{j,n} \to \lambda_j$ as $n \to \infty$ we find for the denominator

$$
\sum_{j=1}^{k}\frac{\lambda_{j,n}}{\hat{\sigma}_j^2} \xrightarrow{P} \sum_{j=1}^{k}\frac{\lambda_j}{\sigma_j^2}. \qquad (34)
$$

Simple algebraic manipulation shows that the numerator of nU_n is

$$
\begin{aligned}
&n\sum_{j<l}\frac{\lambda_{j,n}\lambda_{l,n}}{\hat{\sigma}_j^2\hat{\sigma}_l^2}(\hat{V}_j - \hat{V}_l)^2 \\
=&n\left(\sum_{j=1}^{k}\frac{\lambda_{j,n}}{\hat{\sigma}_j^2}\tilde{V}_j^2\sum_{j=1}^{k}\frac{\lambda_{j,n}}{\hat{\sigma}_j^2} - \left(\sum_{j=1}^{k}\frac{\lambda_{j,n}}{\hat{\sigma}_j^2}\tilde{V}_j\right)^2\right) \\
=&n\left(\sum_{j=1}^{k}\frac{\lambda_{j,n}}{\hat{\sigma}_j^2}\right)\ \tilde{V}'\left(\Lambda_n - \frac{\tilde{\lambda}_n\tilde{\lambda}_n'}{\sum_{j=1}^{k}\frac{\lambda_{j,n}}{\hat{\sigma}_j^2}}\right)\tilde{V} \qquad (35)
\end{aligned}
$$

under the null hypothesis. Here $\tilde{V}_j = \hat{V}_j - V$ and

$$\tilde{V} = \begin{pmatrix} \tilde{V}_1 \\ \tilde{V}_2 \\ \vdots \\ \tilde{V}_k \end{pmatrix}, \Lambda_n = \begin{pmatrix} \frac{\lambda_{1,n}}{\hat{\sigma}_1^2} & 0 & \cdots & 0 \\ 0 & \frac{\lambda_{2,n}}{\hat{\sigma}_2^2} & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\lambda_{k,n}}{\hat{\sigma}_k^2} \end{pmatrix}, \tilde{\lambda}_n = \begin{pmatrix} \frac{\lambda_{1,n}}{\hat{\sigma}_1^2} \\ \frac{\lambda_{2,n}}{\hat{\sigma}_2^2} \\ \vdots \\ \frac{\lambda_{k,n}}{\hat{\sigma}_k^2} \end{pmatrix}, s_n = \begin{pmatrix} \frac{\sqrt{\lambda_{1,n}}}{\hat{\sigma}_1} \\ \frac{\sqrt{\lambda_{2,n}}}{\hat{\sigma}_2} \\ \vdots \\ \frac{\sqrt{\lambda_{k,n}}}{\hat{\sigma}_k} \end{pmatrix},$$

$$\Lambda = \begin{pmatrix} \frac{\lambda_1}{\sigma_1^2} & 0 & \cdots & 0 \\ 0 & \frac{\lambda_2}{\sigma_2^2} & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\lambda_k}{\sigma_k^2} \end{pmatrix}, \tilde{\lambda} = \begin{pmatrix} \frac{\lambda_1}{\sigma_1^2} \\ \frac{\lambda_2}{\sigma_2^2} \\ \vdots \\ \frac{\lambda_k}{\sigma_k^2} \end{pmatrix}, s = \begin{pmatrix} \frac{\sqrt{\lambda_1}}{\sigma_1} \\ \frac{\sqrt{\lambda_2}}{\sigma_2} \\ \vdots \\ \frac{\sqrt{\lambda_k}}{\sigma_k} \end{pmatrix}.$$

Applying Theorem 1 to the individual groups we find

$$Z_n = \sqrt{n}\Lambda_n^{\frac{1}{2}}\tilde{V} \xrightarrow{D} N(0, I_k).$$

Continuing from (35), we see that $\frac{nU_n}{\sum_{j=1}^k \frac{\lambda_{j,n}}{\hat{\sigma}_j^2}}$ is $Z_n' A_n Z_n$ with $A_n = I_k - s_n(s_n' s_n)^{-1} s_n'$, and

$$A_n \xrightarrow{P} A = I_k - s(s's)^{-1}s'. \tag{36}$$

Here A is a symmetric idempotent matrix and is an orthogonal projection into the space orthogonal to the column space of s. The rank of A is same as its trace and equals $k-1$ by the property of orthogonal projector matrices. Applying the continuous mapping theorem, Slutsky's theorem and (36),

$$\frac{nU_n}{\sum_{j=1}^k \frac{\lambda_{j,n}}{\hat{\sigma}_j^2}} \xrightarrow{D} Z'AZ, \tag{37}$$

where the limiting distribution is a quadratic form of normal random variables and is therefore distributed as a χ^2 distribution with degrees of freedom equal to rank of A, which is $k-1$. $\square$

Proof of Proposition 7. For proving consistency of the pooled Fréchet mean $\hat{\mu}_p$, the arguments are essentially the same as those in the proof of Lemma 1 in Petersen and Müller (2017). Consider

$$M_n(\omega) = \frac{1}{n}\sum_{j=1}^{k}\sum_{i\in G_j} d^2(\omega, Y_{ij}), \quad M(\omega) = \sum_{j=1}^{k} \lambda_j E_j(d^2(\omega, Y_j)).$$

For each $\omega \in \Omega$,

$$M_n(\omega) = \sum_{j=1}^{k} \lambda_{j,n} \frac{1}{n_j} \sum_{i\in G_j} d^2(\omega, Y_{ij}) \xrightarrow{P} M(\omega),$$

by the weak law of large numbers applied to the individual groups, using $\lim_{n\to\infty} \lambda_{j,n} = \lambda_j$ for each $j = 1,\ldots,k$. From $|M_n(\omega_1) - M_n(\omega_2)| \leq 2\,\mathrm{diam}(\Omega) d(\omega_1, \omega_2)$ we find that M_n is asymptotically equicontinuous in probability, as

$$\sup_{d(\omega_1,\omega_2)<\delta} |M_n(\omega_1) - M_n(\omega_2)| = O_p(\delta),$$

which allows us to use Theorem 1.5.4 in Wellner and van der Vaart (1996) to conclude that M_n converges weakly to M in $l^\infty(\Omega)$. By applying 1.3.6 of Wellner and van der Vaart (1996) we have that $\sup_{\omega\in\Omega}|M_n(\omega) - M(\omega)|$ converges to zero in probability. By our assumptions and Corollary 3.2.3 in Wellner and van der Vaart (1996) this implies that

$$\hat{\mu}_p \xrightarrow{P} \mu_p. \tag{38}$$

For proving consistency of F_n it is enough to prove the consistency of $\hat{V}_p$ as the consistency of the groupwise Fréchet variances follows from our earlier results. Observe that

$$\left| \hat{V}_p - \frac{1}{n}\sum_{j=1}^{k}\sum_{i\in G_j} d^2(\mu_p, Y_{ij}) \right| \leq 2\,\mathrm{diam}(\Omega)\, d(\hat{\mu}_p, \mu_p) = o_P(1), \tag{39}$$

which implies

$$\left|\hat{V}_p - V_p\right| \le \left|\hat{V}_p - \frac{1}{n}\sum_{j=1}^{k}\sum_{i\in G_j} d^2(\mu_p, Y_{ij})\right| + \left|\frac{1}{n}\sum_{j=1}^{k}\sum_{i\in G_j} d^2(\mu_p, Y_{ij}) - V_p\right| = o_P(1).$$

Here the first term is $o_P(1)$ by (39) and the second term is also $o_P(1)$ by the weak law of large numbers applied to the individual groups, whence $\hat{V}$ converges in probability to V_p. Clearly $V_p - \sum_{j=1}^{k}\lambda_j V_j = \sum_{j=1}^{k}\lambda_j\left\{E_j(d^2(\mu_p, Y_j)) - E_j(d^2(\mu_j, Y_j))\right\}$ is nonnegative, as for each individual group we have

$$E_j(d^2(\mu_p, Y_j)) - E_j(d^2(\mu_j, Y_j)) \ge 0,$$

with equality with zero holding only if $\mu_p = \mu_j$ for all $j = 1, \ldots, k$. Therefore F is always nonnegative and is zero if and only of $\mu_p = \mu_j$ for all $j = 1, \ldots, k$. □

Proof of Theorem 3. This proof relies on the following auxiliary result on uniform consistency of estimators $\hat{V}_p$, $\hat{V}_j$ and $\hat{\sigma}_j^2$ for all $j = 1, 2, \ldots, k$, under the assumption of boundedness of the entropy integral for the space Ω.

Lemma 1. *Under the assumptions of Theorem 3, it holds for all $\varepsilon > 0$ and for all $j = 1, 2, \ldots, k$, where we denote any of the $\hat{V}_j$ and V_j by $\hat{V}$ and V respectively and any of the $\hat{\sigma}_j$ and σ_j by $\hat{\sigma}$ and σ respectively, that*

(A) $\lim_{n\to\infty}\left\{\sup_{P\in\mathcal{P}} P(\left|\hat{V} - V\right| > \varepsilon)\right\} = 0;$

(B) $\lim_{n\to\infty}\left\{\sup_{P\in\mathcal{P}} P(\left|\hat{\sigma}^2 - \sigma^2\right| > \varepsilon)\right\} = 0;$

(C) $\lim_{n\to\infty}\left\{\sup_{P\in\mathcal{P}} P(\left|\hat{V}_p - V_p\right| > \varepsilon)\right\} = 0.$

In all of the above statements the supremum is taken with respect to the underlying true probability measure P of $Y_1, Y_2, \ldots, Y_n$, over the class $\mathcal{P}$ of possible probability measures which generate random observations from Ω.

The proof of Lemma 1 can be found in the following subsection titled Additional Proofs. With similar notation as in the proof of Proposition 6, define

$$\tilde{V}_j = \hat{V}_j - V_j$$

for $j = 1, \dots, k$. The statistic U_n then can be represented as

$$U_n = \sum_{j<l} \frac{\lambda_{j,n}\lambda_{l,n}}{\hat{\sigma}_j^2\hat{\sigma}_l^2}(\hat{V}_j - \hat{V}_l)^2 = \tilde{U}_n + \Delta_n,$$

where $\tilde{U}_n = \sum_{j<l} \frac{\lambda_{j,n}\lambda_{l,n}}{\hat{\sigma}_j^2\hat{\sigma}_l^2}(\tilde{V}_j - \tilde{V}_l)^2$ and $\Delta_n = \sum_{j<l} \frac{\lambda_{j,n}\lambda_{l,n}}{\hat{\sigma}_j^2\hat{\sigma}_l^2}(V_j - V_l)^2 + 2\sum_{j<l} \frac{\lambda_{j,n}\lambda_{l,n}}{\hat{\sigma}_j^2\hat{\sigma}_l^2}(V_j - V_l)(\tilde{V}_j - \tilde{V}_l)$. By replicating the steps in the proof of Proposition 6 we find that $\frac{n\tilde{U}_n}{\sum_{j=1}^k \frac{\lambda_{j,n}}{\hat{\sigma}_j^2}}$ converges in distribution to $\chi^2_{(k-1)}$ asymptotically. Moreover as a consequence of Lemma 1 and by continuity we have that $\frac{\Delta_n}{\sum_{j=1}^k \frac{\lambda_{j,n}}{\hat{\sigma}_j^2}} + \frac{F_n^2}{\sum_{j=1}^k \lambda_{j,n}^2\hat{\sigma}_j^2}$ is a uniformly consistent estimator of $\frac{U}{\sum_{j=1}^k \frac{\lambda_j}{\sigma_j^2}} + \frac{F^2}{\sum_{j=1}^k \lambda_j^2\sigma_j^2}$.

For sets $\{A_n\}$ defined as

$$A_n = \left\{ \frac{\Delta_n}{\sum_{j=1}^k \frac{\lambda_{j,n}}{\hat{\sigma}_j^2}} + \frac{F_n^2}{\sum_{j=1}^k \lambda_{j,n}^2\hat{\sigma}_j^2} < \frac{1}{2}\left(\frac{U}{\sum_{j=1}^k \frac{\lambda_j}{\sigma_j^2}} + \frac{F^2}{\sum_{j=1}^k \lambda_j^2\sigma_j^2} \right) \right\}$$

the uniform consistency of $\frac{\Delta_n}{\sum_{j=1}^k \frac{\lambda_{j,n}}{\hat{\sigma}_j^2}} + \frac{F_n^2}{\sum_{j=1}^k \lambda_{j,n}^2\hat{\sigma}_j^2}$ implies that as $n \to \infty$,

$$\begin{aligned}
&\sup_{P\in\mathcal{P}} P(A_n)\\
&\le \sup_{P\in\mathcal{P}} P\left(\left| \frac{\Delta_n}{\sum_{j=1}^k \frac{\lambda_{j,n}}{\hat{\sigma}_j^2}} + \frac{F_n^2}{\sum_{j=1}^k \lambda_{j,n}^2\hat{\sigma}_j^2} - \frac{U}{\sum_{j=1}^k \frac{\lambda_j}{\sigma_j^2}} - \frac{F^2}{\sum_{j=1}^k \lambda_j^2\sigma_j^2} \right| > \frac{1}{2}\left(\frac{U}{\sum_{j=1}^k \frac{\lambda_j}{\sigma_j^2}} + \frac{F^2}{\sum_{j=1}^k \lambda_j^2\sigma_j^2} \right) \right) \to 0.
\end{aligned}$$

Writing c_α for the $(1-\alpha)$-th quantile of $\chi^2_{(k-1)}$ distribution, we can now represent the limiting power function as

$$
\begin{aligned}
P(R_{n,\alpha}) =& P\left(\frac{\tilde{U}_n+\Delta_n}{\sum_{j=1}^k \frac{\lambda_{j,n}}{\hat{\sigma}_j^2}}+\frac{F_n^2}{\sum_{j=1}^k \lambda_{j,n}^2\hat{\sigma}_j^2}>\frac{c_\alpha}{n}\right)\\
\geq& P\left(\frac{\tilde{U}_n}{\sum_{j=1}^k \frac{\lambda_{j,n}}{\hat{\sigma}_j^2}}+\frac{1}{2}\left(\frac{U}{\sum_{j=1}^k \frac{\lambda_j}{\sigma_j^2}}+\frac{F^2}{\sum_{j=1}^k \lambda_j^2\sigma_j^2}\right)>\frac{c_\alpha}{n},A_n^C\right)\\
\geq& P\left(\frac{n\tilde{U}_n}{\sum_{j=1}^k \frac{\lambda_{j,n}}{\hat{\sigma}_j^2}}>c_\alpha-\frac{n}{2}\left(\frac{U}{\sum_{j=1}^k \frac{\lambda_j}{\sigma_j^2}}+\frac{F^2}{\sum_{j=1}^k \lambda_j^2\sigma_j^2}\right)\right)-P\left(A_n\right)\\
\geq& P\left(\frac{n\tilde{U}_n}{\sum_{j=1}^k \frac{\lambda_{j,n}}{\hat{\sigma}_j^2}}>c_\alpha-\frac{n}{2}\left(\frac{U}{\sum_{j=1}^k \frac{\lambda_j}{\sigma_j^2}}+\frac{F^2}{\sum_{j=1}^k \lambda_j^2\sigma_j^2}\right)\right)-\sup_{P\in\mathcal{P}} P\left(A_n\right).
\end{aligned}
$$

This implies for the sequence of hypotheses $\{H_n\}$,

$$
\begin{aligned}
\lim_{n\to\infty}\beta_{H_n} &= \lim_{n\to\infty}\left\{\inf_{H_n} P\left(R_{n,\alpha}\right)\right\}\\
&\geq \lim_{n\to\infty} P\left(\frac{n\tilde{U}_n}{\sum_{j=1}^k \frac{\lambda_{j,n}}{\hat{\sigma}_j^2}}>c_\alpha-\frac{n}{2}\left(\frac{b_n}{\sum_{j=1}^k \frac{\lambda_j}{\sigma_j^2}}+a_n^2\right)\right)-\lim_{n\to\infty}\left\{\sup_{P\in\mathcal{P}} P\left(A_n\right)\right\}\\
&= \lim_{n\to\infty} P\left(\frac{n\tilde{U}_n}{\sum_{j=1}^k \frac{\lambda_{j,n}}{\hat{\sigma}_j^2}}>c_\alpha-\frac{n}{2}\left(\frac{b_n}{\sum_{j=1}^k \frac{\lambda_j}{\sigma_j^2}}+a_n^2\right)\right).
\end{aligned}
$$

Since $\frac{n\tilde{U}_n}{\sum_{j=1}^k \frac{\lambda_{j,n}}{\hat{\sigma}_j^2}}$ converges in distribution to a $\chi^2_{(k-1)}$ random variable, we find that if a_n is such that $\sqrt{n}a_n\to\infty$ or if b_n is such that $nb_n\to\infty$, then $\lim_{n\to\infty}\beta_{H_n}=1$, completing the proof. □

Proof of Lemma 1. (A). Since the proof is similar for all $j = 1,2,\ldots,k$, we ignore the index j for the proof. Observe that

$$P(\left|\hat{V} - V\right| > \varepsilon) = P\left(\left|\frac{1}{n}\sum_{i=1}^{n} d^2(\hat{\mu}, Y_i) - \frac{1}{n}\sum_{i=1}^{n} d^2(\mu, Y_i) + \frac{1}{n}\sum_{i=1}^{n} d^2(\mu, Y_i) - E(d^2(\mu, Y))\right| > \varepsilon\right)$$

$\leq A_n + B_n$

with A_n and B_n being respectively equal to $P\left(\left|\frac{1}{n}\sum_{i=1}^{n} d^2(\hat{\mu}, Y_i) - \frac{1}{n}\sum_{i=1}^{n} d^2(\mu, Y_i)\right| > \varepsilon/2\right)$ and $P\left(\left|\frac{1}{n}\sum_{i=1}^{n} d^2(\mu, Y_i) - E(d^2(\mu, Y))\right| > \varepsilon/2\right)$.

Observing that $\inf_{\omega\in\Omega} E(d^2(\omega, Y) - d^2(\mu, Y)) = 0$ we have,

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^{n} d^2(\hat{\mu}, Y_i) - \frac{1}{n}\sum_{i=1}^{n} d^2(\mu, Y_i)\right| = \left|\inf_{\omega\in\Omega}\left(\frac{1}{n}\sum_{i=1}^{n} d^2(\omega, Y_i) - \frac{1}{n}\sum_{i=1}^{n} d^2(\mu, Y_i)\right)\right| \\
&= \left|\inf_{\omega\in\Omega}\left(\frac{1}{n}\sum_{i=1}^{n} d^2(\omega, Y_i) - d^2(\mu, Y_i)\right) - \inf_{\omega\in\Omega} E(d^2(\omega, Y) - d^2(\mu, Y))\right| \\
&\leq \sup_{\omega\in\Omega}\left|\frac{1}{n}\sum_{i=1}^{n} d^2(\omega, Y_i) - d^2(\mu, Y_i) - E(d^2(\omega, Y) + E(d^2(\mu, Y))\right| = \sup_{\omega\in\Omega}\left|M_n(\omega)\right|,
\end{aligned}$$

with $M_n(\omega) = \frac{1}{n}\sum_{i=1}^{n} d^2(\omega, Y_i) - d^2(\mu, Y_i) - E(d^2(\omega, Y) + E(d^2(\mu, Y))$. Replicating the steps in proof of Proposition 3, one obtains $E\left(\sup_{\omega\in\Omega}|M_n(\omega)|\right) \leq \frac{2J\mathrm{diam}^2(\Omega)}{\sqrt{n}}$, where J as given by $J = \int_0^1 \sqrt{1 + \log N(\varepsilon, \Omega, d)}\,d\varepsilon$ is the finite entropy integral of Ω and $2\mathrm{diam}^2(\Omega)$ is the envelope for the function class $\{d^2(\omega,\cdot) - d^2(\mu,\cdot) : \omega \in \Omega\}$ which indexes the empirical process $M_n(\omega)$. By Markov's inequality,

$$A_n \leq \frac{4J\mathrm{diam}^2(\Omega)}{\sqrt{n}\varepsilon}. \tag{40}$$

Next we observe that

$$\left|\frac{1}{n}\sum_{i=1}^{n} d^2(\mu,Y_i)-E(d^2(\mu,Y))\right| \leq \sup_{\omega\in\Omega}\left|\frac{1}{n}\sum_{i=1}^{n} d^2(\omega,Y_i)-E(d^2(\omega,Y))\right| = \sup_{\omega\in\Omega}|H_n(\omega)|,$$

where $H_n(\omega)=\frac{1}{n}\sum_{i=1}^{n} d^2(\omega,Y_i)-E(d^2(\omega,Y))$. By similar arguments as before we can see that $E\left(\sup_{\omega\in\Omega}|H_n(\omega)|\right) \leq \frac{J\mathrm{diam}^2(\Omega)}{\sqrt{n}}$, where J is the finite entropy integral of Ω and $\mathrm{diam}^2(\Omega)$ is the envelope of the function class $\{d^2(\omega,\cdot):\omega\in\Omega\}$, which indexes the empirical process $H_n(\omega)$. Again by Markov's inequality,

$$B \leq \frac{2J\mathrm{diam}^2(\Omega)}{\sqrt{n}\varepsilon}. \tag{41}$$

From equations (40) and (41),

$$\sup_{P\in\mathcal{P}} P(|\hat{V}-V|>\varepsilon) \leq \frac{6J\mathrm{diam}^2(\Omega)}{\sqrt{n}\varepsilon} \to 0 \quad \text{as } n\to\infty, \tag{42}$$

completing the proof of (A).

(B). Since the proof is similar for all $j=1,2,\ldots,k$, we ignore the index j for the proof. For proving the uniform consistency of $\hat{\sigma}^2$ it is enough to prove just the uniform consistency of $\frac{1}{n}\sum_{i=1}^{n} d^4(\hat{\mu},Y_i)$ to $E(d^4(\mu,Y))$ and the rest follows from Lemma 1 (A) by continuity. Similarly to the proof of (A), we find

$$\begin{aligned}
&P\left(\left|\frac{1}{n}\sum_{i=1}^{n} d^4(\hat{\mu},Y_i)-E(d^4(\mu,Y))\right|>\varepsilon\right)\\
=&P\left(\left|\frac{1}{n}\sum_{i=1}^{n} d^4(\hat{\mu},Y_i)-\frac{1}{n}\sum_{i=1}^{n} d^4(\mu,Y_i)+\frac{1}{n}\sum_{i=1}^{n} d^4(\mu,Y_i)-E(d^4(\mu,Y))\right|>\varepsilon\right)\\
\leq& A_n+B_n, \quad \text{where}
\end{aligned}$$

$$A_n = P\left(\left|\frac{1}{n}\sum_{i=1}^{n} d^4(\hat{\mu},Y_i) - \frac{1}{n}\sum_{i=1}^{n} d^4(\mu,Y_i)\right| > \varepsilon/2\right),$$
$$B_n = P\left(\left|\frac{1}{n}\sum_{i=1}^{n} d^4(\mu,Y_i) - E(d^4(\mu,Y))\right| > \varepsilon/2\right).$$

Observe that in analogy to the proof of (A),

$$\left|\frac{1}{n}\sum_{i=1}^{n} d^4(\hat{\mu},Y_i) - \frac{1}{n}\sum_{i=1}^{n} d^4(\mu,Y_i)\right| \leq 2\mathrm{diam}^2(\Omega)\left|\frac{1}{n}\sum_{i=1}^{n} d^2(\hat{\mu},Y_i) - \frac{1}{n}\sum_{i=1}^{n} d^2(\mu,Y_i)\right|$$

implies

$$A_n \leq \frac{8J\mathrm{diam}^4(\Omega)}{\sqrt{n}\varepsilon}. \tag{43}$$

Next we observe

$$\left|\frac{1}{n}\sum_{i=1}^{n} d^4(\mu,Y_i) - E(d^4(\mu,Y))\right| \leq \sup_{\omega\in\Omega}\left|\frac{1}{n}\sum_{i=1}^{n} d^4(\omega,Y_i) - E(d^4(\omega,Y))\right| = \sup_{\omega\in\Omega}|K_n(\omega)|, \tag{44}$$

where $K_n(\omega) = \frac{1}{n}\sum_{i=1}^{n} d^4(\omega,Y_i) - E(d^4(\omega,Y))$. Similar arguments as in proof of (A) imply $E(\sup_{\omega\in\Omega}|K_n(\omega)|) \leq \frac{J\mathrm{diam}^4(\Omega)}{\sqrt{n}}$, where J is the finite entropy integral of Ω and $\mathrm{diam}^4(\Omega)$ is the envelope of the function class $\{d^4(\omega,\cdot) : \omega \in \Omega\}$, which indexes the empirical process $K_n(\omega)$. By Markov's inequality,

$$B_n \leq \frac{2J\mathrm{diam}^4(\Omega)}{\sqrt{n}\varepsilon}, \tag{45}$$

and from equations (43) and (45),

$$\sup_{P\in\mathcal{P}} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} d^4(\hat{\mu},Y_i) - E(d^4(\mu,Y))\right| > \varepsilon\right) \leq \frac{10J\mathrm{diam}^4(\Omega)}{\sqrt{n}\varepsilon} \to 0 \quad \text{as } n\to\infty. \tag{46}$$

This completes the proof.

(C). Note that

$$
\begin{aligned}
&P\left(\left|\hat{V}_p - V_p\right| > \varepsilon\right) \\
=&P\left(\left|\frac{1}{n}\sum_{j=1}^{k}\sum_{i\in G_j}\left\{d^2(\hat{\mu}_p, Y_i) - d^2(\mu_p, Y_i)\right\} + \frac{1}{n}\sum_{j=1}^{k}\sum_{i\in G_j} d^2(\mu_p, Y_i) - \sum_{j=1}^{k}\lambda_j E_j\left(d^2(\mu_p, Y_j)\right)\right| > \varepsilon\right) \\
&\leq\ A_n + B_n, \quad \text{with}
\end{aligned}
$$

$$
\begin{aligned}
A_n &= P\left(\left|\frac{1}{n}\sum_{j=1}^{k}\sum_{i\in G_j} d^2(\hat{\mu}_p, Y_i) - \frac{1}{n}\sum_{j=1}^{k}\sum_{i\in G_j} d^2(\mu_p, Y_i)\right| > \frac{\varepsilon}{2}\right), \\
B_n &= P\left(\left|\frac{1}{n}\sum_{j=1}^{k}\sum_{i\in G_j} d^2(\mu_p, Y_i) - \sum_{j=1}^{k}\lambda_j E_j\left(d^2(\mu_p, Y_j)\right)\right| > \frac{\varepsilon}{2}\right).
\end{aligned}
$$

Since $\inf_{\omega\in\Omega}\sum_{j=1}^{k}\lambda_j E_j(d^2(\omega, Y_j) - d^2(\mu_p, Y_j)) = 0$,

$$
\begin{aligned}
&\left|\frac{1}{n}\sum_{j=1}^{k}\sum_{i\in G_j} d^2(\hat{\mu}_p, Y_i) - d^2(\mu_p, Y_i)\right| = \left|\inf_{\omega\in\Omega}\left(\frac{1}{n}\sum_{j=1}^{k}\sum_{i\in G_j} d^2(\omega, Y_i) - d^2(\mu_p, Y_i)\right)\right| \\
=&\left|\inf_{\omega\in\Omega}\left(\frac{1}{n}\sum_{j=1}^{k}\sum_{i\in G_j} d^2(\omega, Y_i) - d^2(\mu_p, Y_i)\right) - \inf_{\omega\in\Omega}\sum_{j=1}^{k}\lambda_j E_j(d^2(\omega, Y_j) - d^2(\mu_p, Y_j))\right| \\
&\leq \sup_{\omega\in\Omega}\left|H_n(\omega)\right|,
\end{aligned}
$$

where

$$
\begin{aligned}
&H_n(\omega) = \frac{1}{n}\sum_{j=1}^{k}\sum_{i\in G_j}\{d^2(\omega, Y_i) - d^2(\mu_p, Y_i)\} - \sum_{j=1}^{k}\lambda_j\{E_j(d^2(\omega, Y_j)) - E_j(d^2(\mu_p, Y_j))\} \\
=&\sum_{j=1}^{k}\lambda_{j,n} M_{n_j}(\omega) + \sum_{j=1}^{k}\left(\lambda_{j,n} - \lambda_j\right)\{E_j(d^2(\omega, Y_j)) - E_j(d^2(\mu_p, Y_j))\}
\end{aligned}
$$

and

$$M_{n_j}(\omega) = \frac{1}{n_j} \sum_{i \in G_j} \{d^2(\omega, Y_i) - d^2(\mu_p, Y_i) - E_j(d^2(\omega, Y_j)) + E_j(d^2(\mu_p, Y_j))\}.$$

Using similar arguments as in the proofs of (A) and (B) for the individual groups, we control the behavior of $H_n(\omega)$ by defining function classes $\{d^2(\omega, y) - d^2(\mu_p, y) : \omega \in \Omega\}$, to obtain for some constants $a_1, a_2 > 0$,

$$\begin{aligned}
&E\left(\sup_{\omega \in \Omega} |H_n(\omega)|\right) \\
&\leq \sum_{j=1}^{k} \lambda_{j,n} E\left(\sup_{\omega \in \Omega} |M_{n_j}(\omega)|\right) + \sum_{j=1}^{k} \left|\lambda_{j,n} - \lambda_j\right| \left|E_j(d^2(\omega, Y_j)) - E_j(d^2(\mu_p, Y_j))\right| \\
&\leq \sum_{j=1}^{k} \lambda_{j,n} \frac{2J\mathrm{diam}^2(\Omega)}{\sqrt{n_j}} + 2\mathrm{diam}^2(\Omega) \sum_{j=1}^{k} \left|\lambda_{j,n} - \lambda_j\right| \leq \frac{a_1}{\sqrt{n}} + a_2 \sum_{j=1}^{k} \left|\lambda_{j,n} - \lambda_j\right|.
\end{aligned}$$

By Markov's inequality,

$$A_n \leq \frac{2}{\varepsilon} \left(\frac{a_1}{\sqrt{n}} + a_2 \sum_{j=1}^{k} \left|\lambda_{j,n} - \lambda_j\right|\right). \tag{47}$$

Next observe that

$$\begin{aligned}
&\left|\frac{1}{n} \sum_{j=1}^{k} \sum_{i \in G_j} d^2(\mu_p, Y_i) - \sum_{j=1}^{k} \lambda_j E_j\left(d^2(\mu_p, Y_j)\right)\right| \\
&\leq \left|\sum_{j=1}^{k} \lambda_{j,n} K_{n_j}(\mu_p)\right| + \left|\sum_{j=1}^{k} (\lambda_{j,n} - \lambda_j) E_j\left(d^2(\mu_p, Y_j)\right)\right| \\
&\leq \sum_{j=1}^{k} \lambda_{j,n} \sup_{\omega \in \Omega} \left|K_{n_j}(\omega)\right| + \sum_{j=1}^{k} \left|(\lambda_{j,n} - \lambda_j)\right| \mathrm{diam}^2(\Omega),
\end{aligned}$$

where $K_{n_j}(\omega) = \frac{1}{n_j} \sum_{i \in G_j} d^2(\omega, Y_i) - E_j(d^2(\omega, Y_j))$.

By similar arguments as before, $E\left(\sup_{\omega \in \Omega} |K_{n_j}(\omega)|\right) \leq \frac{J\mathrm{diam}^2(\Omega)}{\sqrt{n_j}}$. By Markov's inequality, for some constants $b_1, b_2 > 0$,

$$B_n \le \frac{2}{\varepsilon}\left(\frac{b_1}{\sqrt{n}} + b_2 \sum_{j=1}^{k} \left|\lambda_{j,n} - \lambda_j\right|\right). \tag{48}$$

From equations (47) and (48),

$$\sup_{P\in\mathcal{P}} P\left(\left|\hat{V}_p - V_p\right| > \varepsilon\right) \le \frac{2}{\varepsilon}\left(\frac{(a_1+b_1)}{\sqrt{n}} + (a_2+b_2)\sum_{j=1}^{k}\left|\lambda_{j,n} - \lambda_j\right|\right) \to 0 \quad \text{as} \quad n \to \infty,$$

which completes the proof of (C).

□