Biometrika (2019), **xx**, x, *pp*. 1–17
Printed in Great Britain

Supplementary Material to "Fréchet Analysis of Variance for Random Objects"

BY PAROMITA DUBEY
Department of Statistics, University of California, Davis, California 95616, U.S.A.
pdubey@ucdavis.edu

AND HANS-GEORG MÜLLER
Department of Statistics, University of California, Davis, California 95616, U.S.A.
hgmueller@ucdavis.edu

SUMMARY

Section S.1 contains the proofs of the main results in the paper. and a remark on the extension of Proposition 1 to more general M-estimators, Section S.2 contains the proof of Lemma 1, which is needed for the proof of Theorem 3, while Section S.3 contains theoretical justification for the construction of bootstrap confidence intervals for Fréchet variance, as discussed at the end of section 3 of the paper. Section S.4 provides additional simulations for samples of probability distributions and samples of networks of unequal sample sizes.

S.1. MAIN PROOFS

Proof of Proposition 1. Define

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

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 $J(\delta) = J(\delta, \mu_F)$ be the integral in Assumption 2 so that $\delta\ J(\delta) \to 0$ as $\delta \to 0$. Theorems 2.7.11 and 2.14.2 of Van der Vaart & Wellner (1996) and Assumption 2 imply that for small $\delta > 0$,

$$E\left\{\sup_{d(\omega,\mu_F)<\delta}|M_n(\omega)|\right\} \le n^{-1/2}J(\delta)G(\delta),$$

and therefore we have for some $a > 0$,

$$E\left\{\sup_{d(\omega,\mu_F)<\delta}|M_n(\omega)|\right\} \le n^{-1/2}a\delta J(\delta). \tag{1}$$

We want to show that for any $\epsilon > 0,\ \gamma > 0$, there exists $N = N(\epsilon, \gamma)$ such that for all $n \ge N$,

$$P\left[\left|n^{-1/2}\sum_{i=1}^{n}\{d^2(\hat{\mu}_F, Y_i) - d^2(\mu_F, Y_i)\}\right| > \epsilon\right] < \gamma. \tag{2}$$

© 2017 Biometrika Trust

For any small $\delta > 0$,

$$
\begin{aligned}
& P\left[\left| n^{-1/2} \sum_{i=1}^{n} \{d^2(\hat{\mu}_F, Y_i) - d^2(\mu_F, Y_i)\} \right| > \epsilon \right] \\
\leq & P\left[\frac{1}{n} \sum_{i=1}^{n} \{d^2(\hat{\mu}_F, Y_i) - d^2(\mu_F, Y_i)\} < -\epsilon n^{-1/2},\, d(\hat{\mu}_F, \mu_F) \leq \delta \right] + P\left\{ d(\hat{\mu}_F, \mu_F) > \delta \right\} \\
\leq & P\left[- \inf_{d(\omega, \mu_F) < \delta} \frac{1}{n} \sum_{i=1}^{n} \{d^2(\omega, Y_i) - d^2(\mu_F, Y_i)\} > \epsilon n^{-1/2} \right] + P\left\{ d(\hat{\mu}_F, \mu_F) > \delta \right\} \\
\leq & P\left\{ \sup_{d(\omega, \mu_F) < \delta} |M_n(\omega)| > \epsilon n^{-1/2} \right\} + P\left\{ d(\hat{\mu}_F, \mu_F) > \delta \right\}. \qquad (3)
\end{aligned}
$$

Here the last inequality is obtained by observing the following. By definition we have $\inf_{d(\omega,\mu_F)<\delta} \frac{1}{n} \sum_{i=1}^{n} \{d^2(\omega, Y_i) - d^2(\mu_F, Y_i)\} \leq 0$ and from Assumption 1 we have $\inf_{d(\omega,\mu_F)<\delta} [E\{d^2(\omega, Y_i)\} - E\{d^2(\mu_F, Y_i)\}] = 0$. Also note that $\frac{1}{n} \sum_{i=1}^{n} \{d^2(\omega, Y_i) - d^2(\mu_F, Y_i)\}$ and $[E\{d^2(\omega, Y_i)\} - E\{d^2(\mu_F, Y_i)\}]$ are bounded. This implies

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

By using Markov's inequality and the bound in equation (1), for any small $\delta > 0$ such that $\delta\, J(\delta) < \frac{\gamma \epsilon}{2a}$, the expression in (2) can be bounded above by

$$
\begin{aligned}
& P\left\{ \sup_{d(\omega, \mu_F) < \delta} |M_n(\omega)| > \epsilon n^{-1/2} \right\} + P\left\{ d(\hat{\mu}_F, \mu_F) > \delta \right\} \\
\leq & E\left\{ \sup_{d(\omega, \mu_F) < \delta} |M_n(\omega)| \right\} \frac{n^{1/2}}{\epsilon} + P\left\{ d(\hat{\mu}_F, \mu_F) > \delta \right\} \\
\leq & \frac{a \delta\, J(\delta)}{\epsilon} + P\left\{ d(\hat{\mu}_F, \mu_F) > \delta \right\}.
\end{aligned}
$$

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. □

Remark on general M-estimators. The line of arguments in the proof of Proposition 1 can be extended to cover M-estimators that satisfy certain conditions. Let $\hat{\theta}$ be an M-estimator maximizing an empirical criterion, say $\hat{\theta} = \operatorname{argmax}_{\theta \in \Theta} \rho_n(\theta)$, where $\rho_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \rho(Y_i, \theta)$, and let θ_0 be the maximizer of the population target criterion, i.e. $\theta_0 = \operatorname{argmax}_{\theta \in \Theta} \rho(\theta)$, where $\rho(\theta) = E\{\rho(Y, \theta)\}$. Analogous to the result in Proposition 1, one might be interested in obtaining a result of the form

$$
n^{1/2} \left\{ \rho_n(\hat{\theta}) - \rho_n(\theta_0) \right\} = o_P(1). \qquad (4)
$$

For a general result as in (4) by extending the arguments in the above proof of Proposition 1, one needs establish two things. First, one must establish consistency of $\hat{\theta}$, for which one needs an assumption similar to assumption 1, on the existence and uniqueness of $\hat{\theta}$ and θ_0, then applying Corollary 3.2.3 in Van der Vaart & Wellner (1996). By consistency, for large n, one has that $\hat{\theta}$ will be contained in a δ-ball around θ_0 for any $\delta > 0$. Second, one needs to show

$$E\left\{ n^{1/2} \sup_{d(\theta,\theta_0)<\delta} |(\rho_n - \rho)(\theta) - (\rho_n - \rho)(\theta_0)| \right\} \to 0 \quad \text{as} \quad \delta \to 0. \tag{5}$$

For achieving this, one could consider function classes of the form $\mathcal{F}_\delta = \{\rho(\cdot,\theta) - \rho(\cdot,\theta_0) : d(\theta,\theta_0) < \delta\}$. Theorem 2.14.2 in Van der Vaart & Wellner (1996) then provides an upper bound,

$$E\left\{ n^{1/2} \sup_{d(\theta,\theta_0)<\delta} |(\rho_n - \rho)(\theta) - (\rho_n - \rho)(\theta_0)| \right\} \leq ||G_\delta||_{L_2} J_{[]}(\delta),$$

where $G_\delta = G(\cdot,\delta)$ is an envelope function for the function class $\mathcal{F}_\delta$ and $J_{[]}(\delta) = \int_0^1 \left\{1 + \log N_{[]}(\varepsilon||G_\delta||_{L_2}, \mathcal{F}_\delta, L_2(P))\right\}^{1/2} d\varepsilon$ is the bracketing entropy integral of the function class $\mathcal{F}_\delta$ as defined in section 2.14.1 of Van der Vaart & Wellner (1996). If the functions in $\mathcal{F}_\delta$ are Lipschitz continuous in θ, then for some $F(\cdot)$,

$$|\rho(y,\theta_1) - \rho(y,\theta_2)| \leq F(y) d(\theta_1,\theta_2),$$

which implies that $G(y,\delta) = F(y)\delta$ is an envelope for $\mathcal{F}_\delta$ and therefore $||G_\delta||_{L_2} \leq ||F||_{L_2}\delta$. By using Theorem 2.7.11 in Van der Vaart & Wellner (1996), one can obtain an upper bound on $J_{[]}(\delta)$,

$$J_{[]}(\delta) \leq 1 + \int_0^1 \left\{\log N(||F||_{L_2}\varepsilon\delta, B_\delta(\theta_0), d)\right\}^{1/2} d\varepsilon.$$

Similar to assumption 3, if one assumes that $\delta \int_0^1 \left\{\log N(||F||_{L_2}\varepsilon\delta, B_\delta(\theta_0), d)\right\}^{1/2} d\varepsilon \to 0$ as $\delta \to 0$, then

$$||G_\delta||_{L_2} J_{[]}(\delta) \to 0$$

as $\delta \to 0$, which establishes (5).

Proof of Theorem 1.

$$n^{1/2}(\hat{V}_F - V_F) = A_1 + A_2,$$

where

$$A_1 = n^{1/2} \frac{1}{n} \sum_{i=1}^{n} \{d^2(\hat{\mu}_F, Y_i) - d^2(\mu_F, Y_i)\}$$

and

$$A_2 = n^{1/2} \frac{1}{n} \sum_{i=1}^{n} [d^2(\mu_F, Y_i) - E\{d^2(\mu_F, Y_1)\}].$$

The first term A_1 is $o_P(1)$ by Proposition 1 and the second term A_2 converges in distribution to $N(0, \sigma_F^2)$ by applying the Central Limit Theorem to the identically independently distributed 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 2. Observe that

$$\left|\frac{1}{n}\sum_{i=1}^{n} d^4(\hat{\mu}_F, Y_i) - \frac{1}{n}\sum_{i=1}^{n} d^4(\mu_F, Y_i)\right| \leq \ 2\text{diam}^2(\Omega)\left|\frac{1}{n}\sum_{i=1}^{n} d^2(\hat{\mu}_F, Y_i) - \frac{1}{n}\sum_{i=1}^{n} d^2(\mu_F, Y_i)\right| \tag{6}$$

which is $o_P\left(n^{-1/2}\right)$ by Proposition 1. Hence

$$n^{1/2}\left[\begin{pmatrix}\frac{1}{n}\sum_{i=1}^{n} d^4(\hat{\mu}_F, Y_i)\\ \frac{1}{n}\sum_{i=1}^{n} d^2(\hat{\mu}_F, Y_i)\end{pmatrix} - \begin{pmatrix}E\left\{d^4(\mu_F, Y)\right\}\\ E\left\{d^2(\mu_F, Y)\right\}\end{pmatrix}\right]$$

can be decomposed into two components,

$$n^{1/2}\left[\begin{pmatrix}\frac{1}{n}\sum_{i=1}^{n} d^4(\hat{\mu}_F, Y_i)\\ \frac{1}{n}\sum_{i=1}^{n} d^2(\hat{\mu}_F, Y_i)\end{pmatrix} - \begin{pmatrix}\frac{1}{n}\sum_{i=1}^{n} d^4(\mu_F, Y_i)\\ \frac{1}{n}\sum_{i=1}^{n} d^2(\mu_F, Y_i)\end{pmatrix}\right],$$

which is $o_P(1)$ by equation (6) and Proposition 1 and

$$n^{1/2}\left[\begin{pmatrix}\frac{1}{n}\sum_{i=1}^{n} d^4(\mu_F, Y_i)\\ \frac{1}{n}\sum_{i=1}^{n} d^2(\mu_F, Y_i)\end{pmatrix} - \begin{pmatrix}E\left\{d^4(\mu_F, Y)\right\}\\ E\left\{d^2(\mu_F, Y)\right\}\end{pmatrix}\right],$$

which converges in distribution to $N(0, D)$ by applying the Central Limit Theorem to the identically independently distributed random vectors

$$\begin{bmatrix}d^4(\mu_F, Y_i)\\ d^2(\mu_F, Y_i)\end{bmatrix} \ (i = 1, 2, \ldots, n),$$

with

$$D = \begin{bmatrix}\text{var}\left\{d^4(\mu_F, Y)\right\} & \text{cov}\left\{d^4(\mu_F, Y), d^2(\mu_F, Y)\right\}\\ \text{cov}\left\{d^4(\mu_F, Y), d^2(\mu_F, Y)\right\} & \text{var}\left\{d^2(\mu_F, Y)\right\}\end{bmatrix}.$$

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 $\begin{bmatrix}E\left\{d^4(\mu_F, Y)\right\}\\ E\left\{d^2(\mu_F, Y)\right\}\end{bmatrix}$. □

Proof of Proposition 3. Under the null hypothesis the groupwise means are all equal,

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

Then, under Assumptions 1 and 2,

$$\begin{aligned} n^{1/2}F_n &= n^{1/2}(\hat{V}_p - \sum_{j=1}^{k}\lambda_{j,n}\hat{V}_j)\\ &= n^{-1/2}\sum_{i=1}^{n}\{d^2(\hat{\mu}, Y_i) - d^2(\mu, Y_i)\} - \sum_{j=1}^{k}(n_j/n)^{1/2}\frac{1}{{n_j}^{1/2}}\sum_{i\in G_j}\{d^2(\hat{\mu}_j, Y_i) - d^2(\mu, Y_i)\}\\ &= o_p(1), \end{aligned} \tag{7}$$

where the last equality follows by applying Proposition 1 to all observations and also for the individual groups and noting that $(n_j/n)^{1/2} \to \lambda_j^{1/2}$ for all $j = 1, 2, ..., k$. Slutsky's theorem completes the proof. □

Proof of Proposition 4. Under the null hypothesis, let $\mu_1 = \mu_2 = \cdots = \mu_k = \mu$ and $V_1 = V_2 = \cdots = V_k = V$. Using Proposition 2 and since $\lambda_{j,n} \to \lambda_j$ as $n \to \infty$ we find for the denominator

$$\sum_{j=1}^{k} \frac{\lambda_{j,n}}{\hat{\sigma}_j^2} \to \sum_{j=1}^{k} \frac{\lambda_j}{\sigma_j^2} \quad \text{in probability.} \tag{8}$$

Simple algebraic manipulation shows that the numerator of nU_n is

$$\begin{aligned} & n \sum_{j<l} \frac{\lambda_{j,n}\lambda_{l,n}}{\hat{\sigma}_j^2 \hat{\sigma}_l^2} (\hat{V}_j - \hat{V}_l)^2 \\ = & n \left\{ \sum_{j=1}^{k} \frac{\lambda_{j,n}}{\hat{\sigma}_j^2} \tilde{V}_j^2 \sum_{j=1}^{k} \frac{\lambda_{j,n}}{\hat{\sigma}_j^2} - \left(\sum_{j=1}^{k} \frac{\lambda_{j,n}}{\hat{\sigma}_j^2} \tilde{V}_j \right)^2 \right\} \\ = & n \left(\sum_{j=1}^{k} \frac{\lambda_{j,n}}{\hat{\sigma}_j^2} \right) \tilde{V}' \left(\Lambda_n - \frac{\tilde{\lambda}_n \tilde{\lambda}_n'}{\sum_{j=1}^{k} \frac{\lambda_{j,n}}{\hat{\sigma}_j^2}} \right) \tilde{V} \end{aligned} \tag{9}$$

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

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

$$\Lambda = \begin{pmatrix} \frac{\lambda_1}{\sigma_1^2} & 0 & \cdots & 0 \\ 0 & \frac{\lambda_2}{\sigma_2^2} & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\lambda_k}{\sigma_k^2} \end{pmatrix}, \ \tilde{\lambda} = \begin{pmatrix} \frac{\lambda_1}{\sigma_1^2} \\ \frac{\lambda_2}{\sigma_2^2} \\ \vdots \\ \frac{\lambda_k}{\sigma_k^2} \end{pmatrix}, \ s = \begin{pmatrix} \frac{{\lambda_1}^{1/2}}{\sigma_1} \\ \frac{{\lambda_2}^{1/2}}{\sigma_2} \\ \vdots \\ \frac{{\lambda_k}^{1/2}}{\sigma_k} \end{pmatrix}.$$

Applying Theorem 1 to the individual groups we find

$$Z_n = n^{1/2} \Lambda_n^{\frac{1}{2}} \tilde{V} \to N(0, I_k) \quad \text{in distribution.}$$

Continuing from (9), we see that

$$\frac{nU_n}{\sum_{j=1}^{k} \frac{\lambda_{j,n}}{\hat{\sigma}_j^2}} = Z_n' A_n Z_n$$

with $A_n = I_k - s_n(s_n' s_n)^{-1} s_n'$, and

$$A_n \to A = I_k - s(s's)^{-1} s' \quad \text{in probability.} \tag{10}$$

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 (10),

$$\frac{nU_n}{\sum_{j=1}^{k} \frac{\lambda_{j,n}}{\hat{\sigma}_j^2}} \to Z'AZ \quad \text{in distribution,} \tag{11}$$

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 $k-1$. □

Proof of Proposition 5. 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 & Müller (2018). Consider

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

For each $\omega \in \Omega$,

$$M_n(\omega) = \sum_{j=1}^{k} \lambda_{j,n} \frac{1}{n_j} \sum_{i\in G_j} d^2(\omega, Y_i) \to M(\omega) \quad \text{in probability,}$$

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

$$|M_n(\omega_1) - M_n(\omega_2)| \leq 2 \text{ diam}(\Omega) d(\omega_1, \omega_2)$$

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 Van der Vaart & Wellner (1996) to conclude that M_n converges weakly to M in $l^\infty(\Omega)$. By applying 1.3.6 of Van der Vaart & Wellner (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 Van der Vaart & Wellner (1996) this implies that

$$d(\hat{\mu}_p, \mu_p) = o_P(1). \tag{12}$$

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

$$\left| \hat{V}_p - \frac{1}{n}\sum_{j=1}^{k}\sum_{i\in G_j} d^2(\mu_p, Y_i) \right| \leq 2 \text{ diam}(\Omega)\, d(\hat{\mu}_p, \mu_p) = o_P(1), \tag{13}$$

which implies

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

Here the first term is $o_P(1)$ by (13) and the second term is also $o_P(1)$ by the weak law of large numbers applied to the individual groups, whence $\hat{V}_p$ 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

$$E_j\{d^2(\mu_p, Y_j)\} - E_j\{d^2(\mu_j, Y_j)\} \geq 0,$$

where equality with zero holds 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 $\epsilon > 0$ 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) $\lim_{n\to\infty} \left\{ \sup_{P\in\mathcal{P}} \; P\left(\left| \hat{V} - V \right| > \epsilon \right) \right\} = 0;$
(B) $\lim_{n\to\infty} \left\{ \sup_{P\in\mathcal{P}} \; P\left(\left| \hat{\sigma}^2 - \sigma^2 \right| > \epsilon \right) \right\} = 0;$
(C) $\lim_{n\to\infty} \left\{ \sup_{P\in\mathcal{P}} \; P\left(\left| \hat{V}_p - V_p \right| > \epsilon \right) \right\} = 0.$

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 section S.2.

With similar notation as in the proof of Proposition 4, define

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

for $j = 1, \ldots, k$. 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 4 we find that

$$\frac{n\tilde{U}_n}{\sum_{j=1}^{k} \frac{\lambda_{j,n}}{\hat{\sigma}_j^2}} \to \chi^2_{(k-1)} \quad \text{in distribution.}$$

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 $n \to \infty$,

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

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

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

In the above argument A_n^C is the complement of the set A_n that was defined earlier in the proof. 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

$$\frac{n\tilde{U}_n}{\sum_{j=1}^k \frac{\lambda_{j,n}}{\hat{\sigma}_j^2}} \to \chi^2_{(k-1)} \quad \text{in distribution,}$$

we find that if a_n is such that $n^{1/2}a_n \to \infty$ or if b_n is such that $nb_n \to \infty$, then $\lim_{n\to\infty}\beta_{H_n} = 1$, completing the proof. □

S.2. Additional Proofs

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

$$\begin{aligned}&P\left(\left|\hat{V} - V\right| > \epsilon\right)\\ &= P\left[\left|\frac{1}{n}\sum_{i=1}^n d^2(\hat{\mu}, Y_i) - \frac{1}{n}\sum_{i=1}^n d^2(\mu, Y_i) + \frac{1}{n}\sum_{i=1}^n d^2(\mu, Y_i) - E\{d^2(\mu, Y)\}\right| > \epsilon\right]\\ &\leq\ A_n + B_n,\end{aligned}$$

where

$$A_n = P\left\{\left|\frac{1}{n}\sum_{i=1}^n d^2(\hat{\mu}, Y_i) - \frac{1}{n}\sum_{i=1}^n d^2(\mu, Y_i)\right| > \epsilon/2\right\}$$

and

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

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

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

where

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

Replicating the steps in proof of Proposition 1, one obtains $E\left(\sup_{\omega\in\Omega}|M_n(\omega)|\right) \leq n^{-1/2}2J\text{diam}^2(\Omega)$, where $J = \int_0^1\{1+\log N(\epsilon,\Omega,d)\}^{1/2}d\epsilon$ is the finite entropy integral of Ω and $2\text{diam}^2(\Omega)$ 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,

$$A_n \leq 4n^{-1/2}\epsilon^{-1}J\text{diam}^2(\Omega). \tag{14}$$

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 one finds that

$$E\left(\sup_{\omega\in\Omega}|H_n(\omega)|\right) \leq n^{-1/2}J\text{diam}^2(\Omega),$$

where J is the finite entropy integral of Ω and $\text{diam}^2(\Omega)$ 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,

$$B_n \leq 2n^{-1/2}\varepsilon^{-1}J\text{diam}^2(\Omega). \tag{15}$$

From equations (14) and (15),

$$\sup_{P\in\mathcal{P}} P(|\hat{V} - V| > \epsilon) \leq 6n^{-1/2}\varepsilon^{-1}J\text{diam}^2(\Omega) \to 0 \quad \text{as } n\to\infty, \tag{16}$$

completing the proof of (A).

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

the proof of (A), we find

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

where

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

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

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

implies that

$$A_n \leq 8n^{-1/2}\varepsilon^{-1}J\text{diam}^4(\Omega). \tag{17}$$

Next we observe

$$\left|\frac{1}{n}\sum_{i=1}^{n} d^4(\mu, Y_i) - E\{d^4(\mu, Y)\}\right| \leq \sup_{\omega\in\Omega}\left|\frac{1}{n}\sum_{i=1}^{n} d^4(\omega, Y_i) - E\{d^4(\omega, Y)\}\right| = \sup_{\omega\in\Omega}|K_n(\omega)|, \tag{18}$$

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

$$E\left(\sup_{\omega\in\Omega}|K_n(\omega)|\right) \leq Jn^{-1/2}\text{diam}^4(\Omega),$$

where J is the finite entropy integral of Ω and $\text{diam}^4(\Omega)$ 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,

$$B_n \leq 2Jn^{-1/2}\varepsilon^{-1}\text{diam}^4(\Omega), \tag{19}$$

and from equations (17) and (19),

$$\sup_{P\in\mathcal{P}} P\left[\left|\frac{1}{n}\sum_{i=1}^{n} d^4(\hat{\mu}, Y_i) - E\{d^4(\mu, Y)\}\right| > \epsilon\right] \leq 10Jn^{-1/2}\varepsilon^{-1}\text{diam}^4(\Omega) \to 0 \quad \text{as } n \to \infty. \tag{20}$$

This completes the proof.

(C). Note that

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

with

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

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

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

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 $\{d^2(\omega, y) - d^2(\mu_p, y) : \omega \in \Omega\}$, obtaining

for some constants $a_1, a_2 > 0$,

$$E\left(\sup_{\omega\in\Omega}|H_n(\omega)|\right)$$

$$\leq \sum_{j=1}^{k}\lambda_{j,n}E\left(\sup_{\omega\in\Omega}|M_{n_j}(\omega)|\right) + \sum_{j=1}^{k}|\lambda_{j,n}-\lambda_j|\left|E_j\{d^2(\omega,Y_j)\} - E_j\{d^2(\mu_p,Y_j)\}\right|$$

$$\leq \sum_{j=1}^{k}\lambda_{j,n}2Jn_j^{-1/2}\text{diam}^2(\Omega) + 2\text{diam}^2(\Omega)\sum_{j=1}^{k}|\lambda_{j,n}-\lambda_j| \leq a_1 n^{-1/2} + a_2\sum_{j=1}^{k}|\lambda_{j,n}-\lambda_j|.$$

By Markov's inequality,

$$A_n \leq 2\epsilon^{-1}\left(a_1 n^{-1/2} + a_2\sum_{j=1}^{k}|\lambda_{j,n}-\lambda_j|\right). \tag{21}$$

Next observe that

$$\left|\frac{1}{n}\sum_{j=1}^{k}\sum_{i\in G_j}d^2(\mu_p,Y_i) - \sum_{j=1}^{k}\lambda_j E_j\{d^2(\mu_p,Y_j)\}\right|$$

$$\leq \left|\sum_{j=1}^{k}\lambda_{j,n}K_{n_j}(\mu_p)\right| + \left|\sum_{j=1}^{k}(\lambda_{j,n}-\lambda_j)E_j\{d^2(\mu_p,Y_j)\}\right|$$

$$\leq \sum_{j=1}^{k}\lambda_{j,n}\sup_{\omega\in\Omega}\left|K_{n_j}(\omega)\right| + \sum_{j=1}^{k}|(\lambda_{j,n}-\lambda_j)|\,\text{diam}^2(\Omega),$$

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,

$$E\left(\sup_{\omega\in\Omega}|K_{n_j}(\omega)|\right) \leq Jn_j^{-1/2}\text{diam}^2(\Omega).$$

By Markov's inequality, for some constants $b_1, b_2 > 0$,

$$B_n \leq 2\epsilon^{-1}\left(b_1 n^{-1/2} + b_2\sum_{j=1}^{k}|\lambda_{j,n}-\lambda_j|\right). \tag{22}$$

From equations (21) and (22),

$$\sup_{P\in\mathcal{P}}P\left(\left|\hat{V}_p - V_p\right| > \epsilon\right) \leq 2\epsilon^{-1}\left\{(a_1+b_1)n^{-1/2} + (a_2+b_2)\sum_{j=1}^{k}|\lambda_{j,n}-\lambda_j|\right\} \to 0$$

as $n\to\infty$, which completes the proof of (C). □

S.3. JUSTIFICATION OF BOOTSTRAP

Let P_n denote the empirical measure generated by $Y_1, Y_2, \ldots, Y_n$, which puts mass $\frac{1}{n}$ to each of $Y_1, Y_2, \ldots, Y_n$, and P be the underlying measure of $Y_1, Y_2, \ldots, Y_n$. 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 $W_n = n^{1/2}(\hat{V}_F - V_F)/\hat{\sigma}_F$ is $W_{m,n}^* = m^{1/2}(\hat{V}_m^* - \hat{V}_F)/\hat{\sigma}_m^*$, 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 $Y_1^*, Y_2^*, \ldots, Y_m^*$. Due to the sampling scheme, the "population" Fréchet mean μ_n, Fréchet variance V_n and its asymptotic variance σ_n^2 of Y^* in the bootstrap world are given by, respectively,

$$\mu_n = \operatorname*{argmin}_{\omega \in \Omega} E_{P_n}\{d^2(Y^*, \omega)\},\ V_n = \min_{\omega \in \Omega} E_{P_n}\{d^2(Y^*, \omega)\},\ \text{and}\ \sigma_n^2 = \mathrm{Var}_{P_n}\{d^2(Y^*, \mu_n)\},$$

where $E_{P_n}(\cdot)$ and $\mathrm{Var}_{P_n}(\cdot)$ denote expectation and variance under P_n. It is easy to see that μ_n, V_n and σ_n^2 correspond to $\hat{\mu}_F$, $\hat{V}_F$ and $\hat{\sigma}_F^2$ respectively.

Analogous to assumption 1, 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 $Y_1, Y_2, \ldots, Y_n$, and conditionally on $Y_1, Y_2, \ldots Y_n$, $\hat{\mu}_F$ satisfies the following: For any $\varepsilon > 0$,

$$\inf_{d(\omega, \hat{\mu}_F) > \varepsilon} \frac{1}{n}\sum_{i=1}^n d^2(Y_i, \omega) - \frac{1}{n}\sum_{i=1}^n d^2(Y_i, \hat{\mu}_F) > 0 \tag{23}$$

then the following holds: As $m \to \infty$,

$$d(\hat{\mu}_m^*, \hat{\mu}_F) = o_{P_n}(1), \quad |\hat{V}_m^* - \hat{V}_F| = o_{P_n}(1) \quad \text{and} \quad |\hat{\sigma}_m^* - \hat{\sigma}_F| = o_{P_n}(1).$$

The existence and uniqueness of the bootstrap Fréchet mean $\hat{\mu}_m^*$ is a mild assumption and can be established in the same way as the existence and uniqueness of $\hat{\mu}_F$ which is satisfied by the examples of random objects we consider in this paper. The condition in (23) holds in P-probability under assumption 1 and total boundedness of Ω due to the following arguments,

$$\begin{aligned}
&\left| \inf_{d(\omega, \hat{\mu}_F) > \varepsilon} \frac{1}{n}\sum_{i=1}^n d^2(Y_i, \omega) - \frac{1}{n}\sum_{i=1}^n d^2(Y_i, \hat{\mu}_F) - \inf_{d(\omega, \hat{\mu}_F) > \varepsilon} [E\{d^2(Y, \omega)\} - E\{d^2(Y, \mu_F)\}] \right| \\
&\le \sup_{d(\omega, \hat{\mu}_F) > \varepsilon} \left| \frac{1}{n}\sum_{i=1}^n \left\{ d^2(Y_i, \omega) - d^2(Y_i, \hat{\mu}_F) \right\} - E\{d^2(Y, \omega)\} + E\{d^2(Y, \mu_F)\} \right| \\
&\le \sup_{\omega \in \Omega} \left| \frac{1}{n}\sum_{i=1}^n \left\{ d^2(Y_i, \omega) - d^2(Y_i, \mu_F) \right\} - E\{d^2(Y, \omega)\} + E\{d^2(Y, \mu_F)\} \right| \\
&+ \left| \frac{1}{n}\sum_{i=1}^n \left\{ d^2(Y_i, \hat{\mu}_F) - d^2(Y_i, \mu_F) \right\} \right| \\
&= o_P(1),
\end{aligned}$$

where the second term is $o_P(1)$ by Proposition 1 and the first term is $o_P(1)$ as the class of functions $\mathcal{F} = \{d^2(y, \omega) - d^2(y, \mu_F) : \omega \in \Omega\}$ is Lipschitz in ω, which implies $\mathcal{F}$ is Glivenko-Cantelli by Theorems 2.4.1 and 2.7.11 of Van der Vaart & Wellner (1996), provided that Ω is totally bounded. Note also that for sufficiently large n, $\hat{\mu}_F$ is contained in $B_{\varepsilon/2}(\mu_F)$ and therefore

$\{\omega : d(\omega, \hat{\mu}_F) > \varepsilon\}$ is a subset of $\{\omega : d(\omega, \mu_F) > \varepsilon/2\}$. This leads to

$$\begin{aligned}
&\inf_{d(\omega,\hat{\mu}_F)>\varepsilon} \frac{1}{n}\sum_{i=1}^{n} d^2(Y_i,\omega) - \frac{1}{n}\sum_{i=1}^{n} d^2(Y_i,\hat{\mu}_F) \\
&\geq \inf_{d(\omega,\hat{\mu}_F)>\varepsilon} [E\{d^2(Y,\omega)\} - E\{d^2(Y,\mu_F)\}] + o_P(1) \\
&\geq \inf_{d(\omega,\mu_F)>\varepsilon/2} [E\{d^2(Y,\omega)\} - E\{d^2(Y,\mu_F)\}] + o_P(1).
\end{aligned}$$

The condition in (23) holds in P-probability by invoking assumption 1. Additionally under Assumption 2, $\delta J(\delta, \hat{\mu}_F) \to 0$ as $\delta \to 0$. Repeating the arguments for the proof of Proposition 1 and the central limit theorem in Theorem 1, conditionally on $Y_1, Y_2, \ldots, Y_n$, one has $W^*_{m,n} \to N(0,1)$ in distribution as $m \to \infty$. Since $N(0,1)$ has a continuous distribution function on the real line,

$$\sup_x \left| P_n(W^*_{m,n} \leq x) - \Phi(x) \right| = o_P(1)$$

as $m \to \infty$, where $\Phi(\cdot)$ is the standard normal distribution function. Theorem 1 implies

$$\sup_x \left| P(W_n \leq x) - P_n(W^*_{m,n} \leq x) \right| = o_P(1) \tag{24}$$

as $m, n \to \infty$, which establishes the asymptotic consistency of the bootstrap distribution, as discussed at the end of section 3 of the paper. Let $x^*_{1-\alpha}$ be such that,

$$P_n(W^*_{m,n} \leq x^*_{1-\alpha}) = \alpha.$$

In practice is $x^*_{1-\alpha}$ is obtained by Monte Carlo approximations of the bootstrap distribution of the root $R^*_{m,n}$. By (24) ,

$$P(W_n \leq x^*_{1-\alpha}) \approx \alpha.$$

S.4. Additional Simulations

S.4·1. *Distributions*

We replicated the simulation setting for testing distributions in section 5 of the paper and repeated the analysis for $n_1 = 100$ and $n_2 = 200$. Figure 5 shows that our conclusion remains the same for the case where $n_1 \neq n_2$. Figure 6 indicates that for insufficient sample sizes, the bootstrap version of the proposed test based on the approximated bootstrap distribution of the test statistic T_n has more stable rejection regions and overall is more reliable than the asymptotic version of the test. As sample sizes increase, the asymptotic test becomes more reliable and yields results that are similar to the bootstrap test.

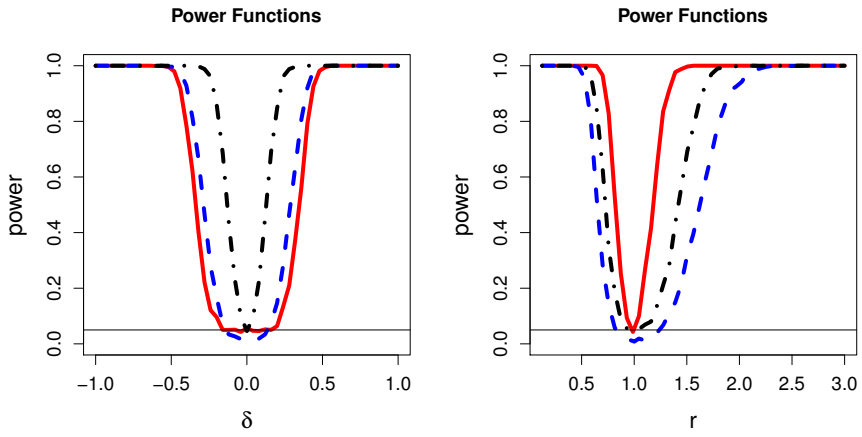


Fig. 5: Empirical power as function of δ for $N(\mu, 1)$ probability distributions with μ from $N(0, 0.5)$ for group G_1 and $N(\delta, 0.5)$ for group G_2, truncated to lie in $[-10, 10]$ for both groups (left) and empirical power as function of r for $N(\mu, 1)$ probability distributions with μ from $N(0, 0.2)$ for G_1 and $N(0, 0.2r)$ for G_2, also truncated to lie in $[-10, 10]$ for both groups (right). The solid red curve corresponds to the bootstrapped version of the proposed test , the dashed blue curve to the graph based test (Chen & Friedman, 2017) and the dot-dashed black curve to the energy test (Székely & Rizzo, 2004). 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 $n_1 = 100$ and $n_2 = 200$.

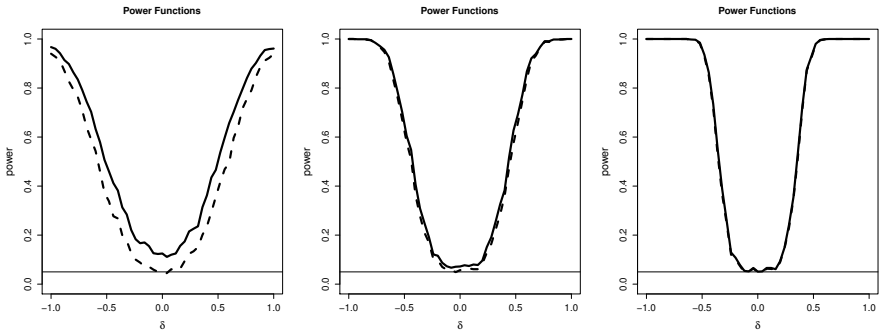


Fig. 6: Empirical power as function of δ for $N(\mu, 1)$ probability distributions with μ from $N(0, 0.5)$ for group G_1 and $N(\delta, 0.5)$ for group G_2. The leftmost panel corresponds to group sample sizes $n_1 = n_2 = 10$, the middle panel corresponds $n_1 = n_2 = 30$ and the rightmost panel to $n_1 = n_2 = 90$. The solid curve corresponds to the asymptotic version of the proposed test in 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.

S.4·2. *Networks*

We repeated our simulation comparison in the network setting as described in section 5 of the paper for $n_1 = 100$ and $n_2 = 200$. Figure 7 shows that the proposed test continues to outperform the other tests and the findings stated in the paper do not change when $n_1 \neq n_2$. Figure 8 shows 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.

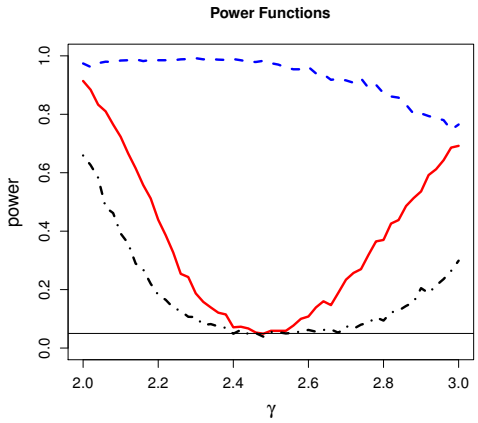


Fig. 7: Empirical power functions of γ for scale-free networks from the Barabási-Albert model, equipped with the Frobeinus metric, with parameter 2.5 for G_1 and γ for G_2. The solid red curve corresponds to the bootstrapped version of the proposed test, the blue dashed curve to the graph based test (Chen & Friedman, 2017) and the dot-dashed black curve to the energy test (Székely & Rizzo, 2004). Sample sizes are fixed at $n_1 = 100$ and $n_2 = 200$. The level of the tests is $\alpha = 0.05$ and is indicated by the line parallel to the x-axis.

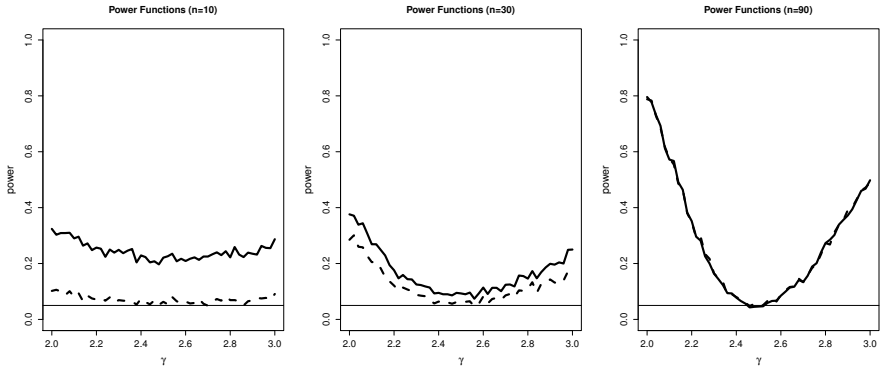


Fig. 8: Empirical power as function of γ for scale-free networks from the Barabási-Albert model, with the model parameter 2.5 for G_1 and γ for G_2. The leftmost panel corresponds to group sample sizes $n_1 = n_2 = 10$, the middle panel to $n_1 = n_2 = 30$ and the rightmost panel to $n_1 = n_2 = 90$. The solid curve corresponds to the proposed asymptotic test 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.

REFERENCES

CHEN, H. & 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.

PETERSEN, A. & MÜLLER, H.-G. (2018). Fréchet regression for random objects with Euclidean predictors. *Annals of Statistics, in press (arXiv preprint arXiv:1608.03012)* **xx**, xxx–xxx.

SZÉKELY, G. J. & RIZZO, M. L. (2004). Testing for equal distributions in high dimension. *InterStat* **5**, 1–6.

VAN DER VAART, A. & WELLNER, J. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.