Improved inference for vaccine-induced immune responses via shape-constrained methods[*]

Nilanjana Laha[1], Zoe Moodie[2], Ying Huang[2], and Alex Luedtke[2,3]

[1]*Department of Statistics, Texas A&M University, College Station, TX 77843, U.S.A*
e-mail: nlaha@tamu.edu

[2]*Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center,*
Seattle, WA 98109 e-mail: zoe@fredhutch.org; yhuang@fredhutch.org

[3]*Department of Biostatistics, University of Washington, Seattle, WA 98195 e-mail:*
aluedtke@uw.edu

Abstract: We study the performance of shape-constrained methods for evaluating immune response profiles from early-phase vaccine trials. The motivating problem for this work involves quantifying and comparing the IgG binding immune responses to the first and second variable loops (V1V2 region) arising in HVTN 097 and HVTN 100 HIV vaccine trials. We consider unimodal and log-concave shape-constrained methods to compare the immune profiles of the two vaccines, which is reasonable because the data support that the underlying densities of the immune responses could have these shapes. To this end, we develop novel shape-constrained tests of stochastic dominance and shape-constrained plug-in estimators of the squared Hellinger distance between two densities. Our techniques are either tuning parameter free, or rely on only one tuning parameter, but their performance is either better (the tests of stochastic dominance) or comparable with the nonparametric methods (the estimators of the squared Hellinger distance). The minimal dependence on tuning parameters is especially desirable in clinical contexts where analyses must be prespecified and reproducible. Our methods are supported by theoretical results and simulation studies.

MSC2020 subject classifications: 62G07, 62G10, 62G05, 62P10.
Keywords and phrases: Vaccine trial, shape-constraint, test of stochastic dominance, Hellinger distance, logconcave, unimodal.

1. Introduction

To date, the RV144 trial conducted in Thailand is the only vaccine efficacy trial to show a signal of efficacy (31%) against HIV infection (Rerks-Ngarm *et al.*, 2009). RV144 inspired a phase 1b trial, named HIV Vaccine Trials Network (HVTN) 097, which evaluated the safety and immunogenicity of the same

---

regimen in a South African population (Gray *et al.*, 2019). The predominant subtype of HIV in South Africa is Clade C. Therefore, in an effort to increase the potential for high efficacy against clade C infections, scientists modified the HVTN 097 regimen to include HIV strains matched to the South African clade C infections (Bekker *et al.*, 2018). A phase 1/2 trial named HVTN 100 assessed the safety and immunogenicity of the modified regimen in South Africa. However, the HVTN 702 phase 2B/3 trial of the HVTN 100 regimen in South Africa met its non-efficacy criteria in February 2020 at a planned interim analysis.

In light of the above, the comparison between the immune response profiles of HVTN 097 and HVTN 100 trials becomes important because the latter can shed some light on why the HVTN 100 regimen lacks efficacy. To fix ideas, here we focus on one class of immune responses, namely the binding of IgG antibodies to the first and second variable loops (V1V2 region) of the HIV envelope. This immune response is of particular interest because the RV144 trial revealed an inverse association between HIV infection and this immune response among vaccinees (Haynes *et al.*, 2012). Using data from HVTN 097 and HVTN 100, we focus on answering the following three questions:

Q1. How can we estimate the densities of the aggregated IgG binding immune responses (to HIV-1 envelope proteins)?

Q2. Is there any ordering between the distributions of the immune responses from the two trials?

Q3. How can we measure the discrepancy between the densities of the immune responses from the two trials?

Although Q1 is not directly related with comparison of the two vaccine trials, answering Q1 is important because the resulting density estimators can help in designing subsequent vaccines. To answer Q2, we resort to testing for stochastic dominance, which addresses the ordering of the underlying distribution functions. To answer Q3, we rely on the squared Hellinger distance as a measure of discrepancy between two densities, where, for densities f and g, the Hellinger distance $\mathcal{D}(f, g)$ is defined by

$$\mathcal{D}(f, g) = \sqrt{\frac{1}{2} \int_{\mathbb{R}} \left(\sqrt{f(x)} - \sqrt{g(x)} \right)^2 dx}.$$

The significance of Q3 may not be immediately obvious. However, the squared Hellinger distance between the immune responses of HVTN 097 and HVTN 100 trials can serve as a benchmark when new pairs of vaccines are compared in future vaccine trials. The reason behind choosing the squared Hellinger distance as the measure of discrepancy in particular is discussed in Section 5.

There are numerous nonparametric methods that can be implemented to carry out the aforementioned steps. However, traditional nonparametric methods do not exploit any information on the shape of the underlying densities. The uniformity of the trial population and exploratory analyses bear evidence that the underlying densities can be unimodal. The data are also consistent with the possibility that the densities are log-concave. The latter is an important subclass of unimodal densities, often advocated for use in modelling because it contains

most of the well-known subexponential unimodal densities and allows powerful density estimation tools (Walther *et al.*, 2009). Of late, shape-constrained density estimation has gained much attention. Among other reasons, the reduced burden of external tuning parameters (Samworth and Sen, 2018, Johnson *et al.*, 2018) makes shape-constrained density estimation an attractive alternative to traditional nonparametric approaches like kernel-based or basis expansion type methods, which are known to be sensitive to the choice of the tuning parameters (cf. pp. 327, Efromovich, 2008, Laha, 2021). Also, shape-constrained density estimation methods require weaker smoothness assumptions for asymptotic consistency than do the nonparametric methods. Shape-constrained techniques are widely used in economics and operations research (Johnson *et al.*, 2018), and have seen application in other domains such as circuit design (Hannah and Dunson, 2012). For a detailed account on the recent development of the shape constraint literature, we refer the survey articles Samworth and Sen (2018) and Samworth (2018).

Although leveraging shape information can potentially increase efficiency, there is little to no literature on the application of shape-constrained tools in vaccine trials. To answer our motivating questions, therefore, we develop new methods using shape-constrained tools. The application of the new methods is not limited to vaccine trials. For example, our tests can be applied to other areas of medical research where tests of stochastic dominance are relevant. See Leshno and Levy (2004) for an in-depth discussion of potential uses of tests of stochastic dominance in medical research. In particular, our methods are applicable to the mortality data considered in Leshno and Levy (2004) to infer whether a surgery increases the mortality of patients with abdominal aortic aneurysma. See also Stinnett and Mullahy (1998) and DeFauw (2011) for the use of stochastic dominance in cost effective analysis of healthcare. Outside medical research, our tests have substantial applicability in finance, economics, social welfare, and operations research, where stochastic dominance is a popular tool to compare portfolios, income, utility, poverty, opportunity etc.; see Levy (1992), Sriboonchita *et al.* (2009), Le Breton (1991), among others, for a detailed account. On the other hand, Hellinger distance has also seen successful application as a measure of discrepancy in various disciplines ranging from machine learning (Cieslak and Chawla, 2009, González-Castro *et al.*, 2013, 2010) to ecology (Rao, 1995) to fraud detection (Yamanishi *et al.*, 2004). However, for income data, log-concavity based methods should be used with caution since distributions with log-concave density are always sub-exponential (Cule and Samworth, 2010), whereas income data can have heavier tails. Diagnostic tools such as those in Asmussen and Lehtomaa (2017) can be used to enquire if the data has heavier tail, e.g. regular varying tails, in these cases.

1.1. Organization of article and main contributions

Although the methods developed in this paper are general, central to our application lies the HVTN data, which we describe in Section 2. Below we briefly

discuss our methods and main contributions. We write f_{097} and f_{100} for the densities of the immune responses in the HVTN 097 and HVTN 100 trials respectively, and F_{097} and F_{100} for the corresponding distribution functions.

Estimating f_{097} and f_{100}

In vaccine trials, traditionally a kernel density estimator (KDE) is used for the purpose of density estimation (cf. Miladinovic *et al.*, 2014). However, using cross-validation, in Section 3, we show that the log-concave maximum likelihood estimator (MLE) based estimators of Dümbgen and Rufibach (2009) and Chen and Samworth (2013) minimize the estimated mean integrated squared error (MISE) among a class of shape-constrained density estimators and KDEs.

Shape-constrained tests of stochastic dominance

The claim of a stochastic ordering between two samples is made stronger when it is backed by a test of stochastic dominance. We say a distribution function F stochastically dominates another distribution function G in first order ($F \succeq G$) if $F(x) \le G(x)$ for all $x \in \mathbb{R}$. If X and Y are two random variables with distribution functions F and G, respectively, then in this case, we say X stochastically dominates Y, and write $X \succeq Y$. The dominance is regarded as "strict" ($F \succ G$ or $X \succ Y$) if, in addition, there exists $x \in \mathbb{R}$ such that $F(x) < G(x)$. If F does not strictly stochastically dominate G, then this event is defined as the non-dominance ($F \nsucceq G$) of F over G (Whang, 2019, p. 25).

The rejection of the null of non-dominance against the alternative of stochastic dominance makes the strongest case for ranking one distribution over the other (Davidson and Duclos, 2013, Álvarez-Esteban *et al.*, 2016, Ledwina and Wyłupek, 2012). However, the resulting test suffers from lack of power because the overlap of distribution functions at the tails make unrestricted stochastic dominance almost impossible to establish via hypothesis testing (cf. Davidson and Duclos, 2013, Whang, 2019). There are many ways to deal with this difficulty. For example, Ledwina and Wyłupek (2012) exploits the fact that the dominance between two distribution functions employs a dominance between their Fourier coefficients under a carefully chosen basis. However, their method rejects the null of non-dominance in some scenarios when the distribution functions cross each other. Another remedy discussed in the literature (Kaur *et al.*, 1994, Davidson and Duclos, 2013, Álvarez-Esteban *et al.*, 2016) involves excluding the tail region because it does not contain enough reliable information for the problem at hand. This type of tests focus only on a compact set D inside the interior of the combined support $\{0 < F + G < 2\}$. In Section 4, we follow the latter strategy because it allows for many ways of incorporating shape-constraints with desirable asymptotic properties.

To the best of our knowledge, we are the first to introduce the use of shape constraints in the context of testing the null of non-dominance against stochastic dominance. Moreover, one of our nonparametric test statistics in Section 4.1

has not, to our knowledge, previously been studied in the context of testing the null of non-dominance. In Section 4.2, we show that the shape-constrained and nonparametric versions of our tests control the asymptotic type I error at any null configuration under reasonable conditions. We also show that our tests are asymptotically unbiased, and consistent against all alternatives lying in the interior of the class of alternative distributions. In Section 4.3, we empirically show that the shape-constrained tests have better power than their nonparametric counterparts although they have the same asymptotic critical values. Section 4.4 analyses the application of these tests to our data. The proofs are deferred to Appendix A.

Shape-constrained plug-in estimators of the squared Hellinger distance

In Section 5, we construct plug-in estimators of the squared Hellinger distance. It is well known that, unless bias-corrected, plug-in estimators based on the KDE generally have a first-order bias (cf. Section 2 of Robins *et al.*, 2009). In contrast, for some smooth functionals, shape constrained MLE based plug-in estimators do not require further bias correction (cf. Jankowski, 2010, Mukherjee *et al.*, 2019). The results in Lopuhaä and Musta (2019) indicate that the squared Hellinger distance is an example of such a smooth functional. When the underlying density is unimodal, we show that our unimodal density based plug-in estimator enjoys the same asymptotic guarantees as that of the bias corrected KDE based estimators (Kandasamy *et al.*, 2015) under some regularity conditions. The simulation studies in Section 5.3 suggest that similar results hold for our log-concave MLE based plug in estimators as well. In fact, our smooth log-concave MLE based estimator shows stable performance across all settings, where even the bias-corrected version of the KDE based estimator struggles in some cases.

In the process, we develop theoretical tools for analyzing the asymptotic behavior of plug-in estimators based on the unimodal density estimator of Birgé, which may of independent interest. We defer the latter analysis to Appendix B. The methods developed in this paper are implementable using the R package SDNNtests (Laha and Luedtke, 2020), which is available on GitHub.

1.2. Notations and terminologies

Before proceeding further, we introduce some notation that will be used throughout this paper. We consider two independent samples $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ drawn from distributions with densities f and g. We denote the corresponding distribution functions by F and G, respectively. The respective empirical distribution functions will be denoted by $\mathbb{F}_m$ and $\mathbb{G}_n$. The pooled sample $(X_1, \ldots, X_m, Y_1, \ldots, Y_n)$ has size $N = m + n$. We denote the corresponding empirical distribution function by $\mathbb{H}_N$.

For $k \geq 1$, we let $||\cdot||_k$ denote the usual L_k norm, i.e. $||\mu||_k = (\int_{\mathbb{R}} |\mu(x)|^k dx)^{1/k}$, where μ is a function supported on the real line. Also, we denote $||\mu||_{\infty} = \sup_{x \in \mathbb{R}} \mu(x)$. For a density f, denote by $\text{supp}(f)$ the set $\{x : f(x) > 0\}$. For a concave function $f : \mathbb{R} \mapsto \mathbb{R}$, the domain $\text{dom}(f)$ will be defined as in (Rockafellar, 1970, p. 40), that is, $\text{dom}(f) = \{x \in \mathbb{R} : f(x) > -\infty\}$. For a sequence of measures $\{P_n\}_{n \geq 1}$, we say P_n converges weakly to P, and write $P_n \to_d P$, if $\lim_{n \to \infty} \int \mu dP_n = \int \mu dP$ holds for any bounded continuous function $\mu : \mathbb{R} \mapsto \mathbb{R}$. For any two sets $A, B \subset \mathbb{R}$, we denote by $\text{dist}(A, B)$ the quantity $\min\{|x - y| : x \in A, y \in B\}$. We let $\text{int}(A)$ denote the interior of the set A.

2. Background: HVTN 097 and HVTN 100

This section presents an exploratory analysis of the dataset. For both trials, we consider the magnitude of IgG binding to the V1V2 region of seven clade C glycoprotein 70 antigens. The immune responses were measured by an HIV-1 binding antibody multiplex assay (BAMA). Following Haynes *et al.* (2012), Gray *et al.* (2019), and Bekker *et al.* (2018), we use the log-transformed net median fluorescence intensity (MFI) as the measure of immune response for statistical analysis.

In HVTN 097 and HVTN 100, four injections of HIV vaccines were given at months 0, 1, 3, and 6. In this study, we only consider the responses measured two weeks after the month six vaccination, which is considered to be the peak immune response time point. We include only those vaccinees in this study who (a) completed the first four scheduled vaccinations and provided samples at two weeks after the month six vaccination (known as vaccinated per-protocol participants), and (b) developed a positive immune response for at least one of the seven clade C V1V2 antigens. There are 68 and 180 vaccinees in the HVTN 097 and HVTN 100 trial, respectively, who satisfy the above criteria. We base our analysis on the aggregated response, averaged over the seven clade C antigens mentioned above, and refer only to the latter when we say "immune response". We let F_{097} and F_{100} denote the distribution functions corresponding to the aggregated response from the two trials.

Figure 1 illustrates the empirical CDFs, the histogram, the boxplot, and the KDES of the immune responses from the two trials. Figure 1i illustrates that F_{100} is always greater than F_{097} except at the tails, hinting at the stochastic dominance of the HVTN 097 immune response over the HVTN 100 immune response. The histogram in Figure 1ii, the boxplot in Figure 1iii, and the plot of the KDEs in Figure 1iv also suggest that the HVTN 097 trial induces higher immune response. Bekker *et al.* (2018) also indicated that the magnitude of positive responses in the HVTN 100 trial is lower than that of the RV144 trial, which, on the other hand, is reported to be slightly lower than the responses in the HVTN 097 trial (Gray *et al.*, 2019). Therefore, it makes sense to posit the null of non-dominance of F_{097} over F_{100} against the alternative that F_{097} stochastically dominates F_{100}. Note also that the comparisons between the two sets of immune responses is reasonable because the trials were conducted on

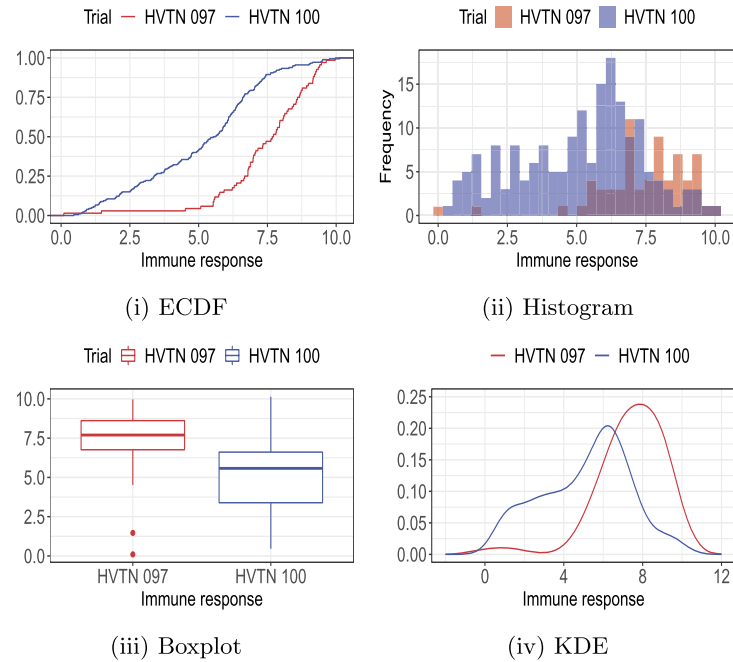(i) ECDF

(ii) Histogram

(iii) Boxplot

(iv) KDE

Fig 1: Plots (i), (ii), (iii), (iv) display the empirical distribution functions (ECDF), the histogram, the boxplot, and the KDEs of the average IgG binding responses corresponding to the HVTN 097 and HVTN 100 regimens.

similar populations and share approximately the same support (cf. Figure 1). See also Table 6 in Appendix D for a comparison between the two trials.

3. Density estimation

This section compares different estimators of f_{097} and f_{100}. Since our study includes some unimodal and log-concave density estimators, we begin by presenting some observations in support of the shape-restriction assumptions.

As mentioned previously, the unimodality assumption is not unreasonable owing to the homogeneity of the trial populations. The histogram in Figure 1ii and the KDEs displayed in Figure 1iv both support this claim. Although unimodality is a naturally occurring shape constraint, the class of all unimodal densities is too large to admit an MLE (Birgé, 1997). The class of log-concave densities is a subclass of the class of unimodal densities, which is small enough to admit an MLE (Pal *et al.*, 2007), but contains most of the commonly used subexponetial unimodal densities (Walther *et al.*, 2009). The log-concave MLE can be computed efficiently using the R package `logcondens`. Furthermore, there is a smoothed version of the log-concave MLE, which is also free of tuning parameters (Chen and Samworth, 2013), and thus can potentially replace the

smoothed unimodal density estimators, which generally depend crucially on external tuning parameters. In view of the above, many researchers, e.g. Walther (2002) and Walther *et al.* (2009), advocate opting for log-concavity shape constraint in situations where unimodality may seem plausible.

Although it is difficult to provide visual evidence in favor of the assumption of log-concavity, the KDE plot in Figure 1iv does not indicate a departure from log-concavity either. Using the test of log-concavity in Chen and Samworth, we test the null of log-concavity against the alternative of violation of log-concavity for f_{097} and f_{100}. The corresponding p-values for f_{097} and f_{100} are 0.4890 and 0.4631, respectively, which implies that our data does not have enough evidence for rejecting the null of log-concavity. We briefly describe below the density estimators that we consider in this section.

Grenander-type unimodal density estimator (Birgé, 1997): Although the class of all unimodal densities does not permit an MLE (Birgé, 1997), when the mode of a unimodal density f is known, the MLE $\widehat{f}_m^0$ exists (Rao, 1969), and it is a piecewise constant unimodal density with mode at the true mode. However, the MLE $\widehat{f}_m^0$ is generally not useful due to the lack of knowledge on the location of the true mode. The estimator $\widehat{f}_m$ presented in Birgé (1997) is a piecewise constant estimator of f, constructed in such a way so that the Kolmogorov-Smirnov distance between the corresponding distribution functions can be made arbitrarily small, in particular, smaller than a pre-fixed number $\eta > 0$. Although this $\eta > 0$ is an external parameter, unlike the kernel bandwidth, a smaller η always leads to a more accurate estimation (Birgé, 1997), and hence it does not require actual tuning. Indeed, if we choose the parameter $\eta = o(m^{-1})$, then our Lemma A.5 in Appendix A ensures that the total variation distance between $\widehat{f}_m$ and $\widehat{f}_m^0$ is $o(m^{-1/2})$ with probability one, where $\widehat{f}_m^0$ was defined to be the MLE of f had the true mode been known. Therefore, we choose η to be the inverse of the combined sample size of the two trials.

Smooth unimodal estimator: There is a substantial body of literature on smooth unimodal density estimators. See for instance, Eggermont and LaRiccia (2000), Mammen *et al.* (2001), Hall and Huang (2002), Wolters (2012), Meyer (2012), Turnbull and Ghosh (2014), and Wolters and Braun (2018), among others. The smooth unimodal density estimators generally depend crucially on external tuning parameters, as mentioned previously. Since we opt for shape constraints mainly to avoid tuning parameters and we already have at our disposal the tuning free smooth log-concave MLE estimator, the smooth unimodal density estimators are not particularly attractive to us. Still, since the current section solely focuses on density estimation, we include some smooth unimodal estimators for comparison here, namely the estimators of Turnbull and Ghosh (2014), Hall and Huang (2002), Wolters (2012), and Wolters and Braun (2018).

Turnbull and Ghosh (2014) approximates the unknown unimodal density using Bernstein polynomial. We use the condition number approach of Turnbull and Ghosh (2014) to select tuning parameters related to the degree of the polynomial. The estimators in Hall and Huang (2002), Wolters (2012), and Wolters and Braun (2018) are kernel based. To compute them, we use the R package
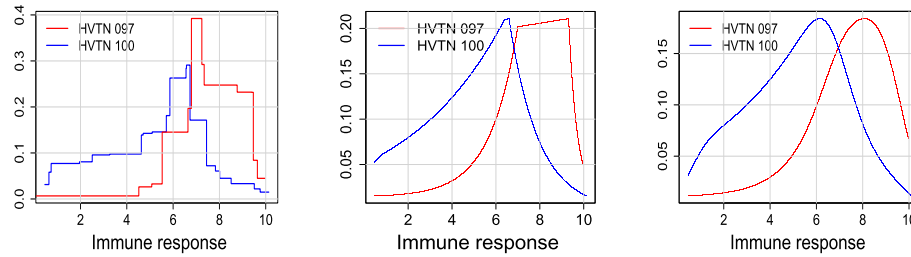
Fig 2: Birgé's estimator (left), log-concave MLE (middle), and the smoothed log-concave MLE (right) based on the immune responses in the HVTN 097 and the HVTN 100 trial.

`scdensity` with the default choice of bandwidth.

Log-concave estimators: Our first log-concave estimator is the MLE among the class of all log-concave densities. The MLE is continuous but non-smooth (Dümbgen and Rufibach, 2009). The second estimator is a smoothed version of the MLE (Dümbgen and Rufibach, 2009, Chen and Samworth, 2013). The smoothing parameter for the latter is data dependent and has a closed form formula, and hence it does not require external tuning. Figure 2 displays these two density estimators. For more on the properties of the log-concave density estimators, see, e.g., Balabdaoui *et al.* (2009), Dümbgen *et al.* (2011), Cule and Samworth (2010), and Doss and Wellner (2016).

Kernel density estimators: We consider kernel density estimators (KDE) with Gaussian kernel. The optimal bandwidth was chosen either by the univariate plug-in selector of Wand and Jones (1994), or the univariate least square cross-validation (LSCV) selector of Bowman (1984) and Rudemo (1982) (see Figure 1iv).

We prefer the estimator $\widehat{f}_m$ of a density f with the smallest mean integrated squared error (MISE), which is given by $E \int_{\mathbb{R}} (\widehat{f}_m(x) - f(x))^2 dx$. Noting minimizing the MISE with respect to $\widehat{f}_n$ is equivalent to minimizing

$$\text{MISE-err} = E \int_{\mathbb{R}} \widehat{f}_m^2(x)dx - 2E \int_{\mathbb{R}} \widehat{f}_m(x)f(x)dx, \qquad (3.1)$$

we estimate the latter quantity using a ten folds cross-validation. We also estimate the negative log-likelihood

$$-l_m = -m^{-1} \sum_{i=1}^{n} \log \widehat{f}_m(X_i). \qquad (3.2)$$

Based on these risks (see Table 1), our recommended estimators are the log-concave estimators which exhibit the lowest risk in an overall sense. Table 1 indicates that Birgé's estimator excels in minimizing the $-l_m$ risk but it has higher estimated MISE when compared to the other estimators. This can be

TABLE 1
Table of the estimated risks for different density estimators of the aggregated immune response in trials HVTN 097 and HVTN 100

Estimators	HVTN 097 MISE-err	$-l_m$	HVTN 100 MISE-err	$-l_m$
Unimodal (Birgé's estimator)	-0.172	1.662	-0.126	2.071
Unimodal (Bernstein)	-0.191	1.797	-0.130	2.133
Unimodal (Wolters, 2012)	-0.190	1.878	-0.102	2.549
Unimodal (Wolters and Braun, 2018)	-0.190	1.795	-0.124	2.300
Unimodal (Hall and Huang, 2002)	-0.190	1.965	-0.124	2.288
Smooth log-concave MLE	-0.196	1.758	-0.129	2.127
Log-concave MLE	-0.193	1.758	-0.130	2.127
KDE (plug-in bandwidth selector)	-0.189	1.750	-0.128	2.140
KDE (LSCV bandwidth selector)	-0.189	1.777	-0.128	2.140

attributed to its spikes at the mode (see Figure 2), which contributes large positive terms to l_n and MISE. Grenander type unimodal estimators are known to exhibit such "spike-problem" at the mode (Walther *et al.*, 2009), which is caused by the inconsistency of the density estimator at the mode (for more details, see Woodroofe and Sun, 1993, Balabdaoui *et al.*, 2009).

4. Test of stochastic dominance

To provide an answer to Q2, we construct tests for the null of non-dominance against that of stochastic dominance using the log-concave MLEs and the unimodal estimator of Birgé. We compare the resulting shape-constrained tests with their nonparametric counterparts.

Our shape-restricted methods rely on estimating the densities f and g. We denote the corresponding unimodal estimators of Birgé by $\widehat{f}_m$ and $\widehat{g}_n$, respectively. The construction of Birgé's estimators requires a tuning parameter η, which we set to be N^{-1} where $N = m + n$. We let $\tilde{f}_m$ and $\tilde{g}_n$ denote the log-concave MLEs of f and g (Dümbgen and Rufibach, 2009), and write $\tilde{f}_m^{sm}$ and $\tilde{g}_n^{sm}$ for their respective smoothed versions (Chen and Samworth, 2013). The corresponding distribution functions will be denoted by $\widehat{F}_m$, $\widehat{G}_n$, $\tilde{F}_m$, $\tilde{G}_n$, $\tilde{F}_m^{sm}$, and $\tilde{G}_n^{sm}$, respectively.

As m and n approach ∞, we assume that $m/N \to \lambda \in (0, 1)$.

Letting $H = \lambda F + (1 - \lambda)G$, for $p \in (0, 1/2)$, we also define the sets

$$D_p(F, G) := D_p = [H^{-1}(p), H^{-1}(1-p)] \quad \text{and} \quad D_{p,m,n} := [\mathcal{H}^{-1}(p), \mathcal{H}^{-1}(1-p)]. \tag{4.1}$$

4.1. Construction of the tests

Suppose $D \subset \text{supp}(f) \cup \text{supp}(g)$ is compact. Following Kaur *et al.* (1994) and Davidson and Duclos, we formulate the hypotheses as follows:

$$\mathcal{H}_0 : F(z) \geq G(z) \text{ for some } z \in D \quad \text{vs.} \quad \mathcal{H}_1 : F(z) < G(z) \text{ for all } z \in D. \tag{4.2}$$

The null configuration $\mathcal{H}_0$ occurs if $F = G$, or G stochastically dominates F, or if F and G touch or cross each other on D. Thus our formulation is unable to reject the null when F and G touch at a point in D even if $F \succ G$. This limitation seems to be unavoidable because such a configuration (F, G) lies on the common boundary shared by $\{(F, G) : F \succeq G\}$ and $\{(F, G) : F \npreceq G\}$, and hence can not be discriminated from the null without sacrificing control over the size of test. Notably, our exploratory analysis (see Fig 1i) suggests that it is unlikely that F_{097} and F_{100} fall in this category. See Figure 3 for examples of different scenarios associated with our hypotheses.

Regarding the choice of D, we need to ensure that D is inside the combined support of f and g because otherwise, $\inf_{z \in D}[G(z) - F(z)]$ will always be 0. The set D_p defined in (4.1) satisfies this criterion. In practice, we replace this unknown D_p by $D_{p,m,n}$ defined in (4.1), which always utilizes $100(1 - 2p)\%$ of the combined data. Naturally, if p is too small, rejection of the null will be difficult, where a large p will exclude a large portion of the data, which might be unnecessary. If only some particular interval of the data is of practical interest (e.g. some particular range of immune responses or biomarkers), we suggest setting D to be the smallest superset of that interval. In the absence of such prior knowledge, we suggest choosing the largest p so that $D_{p,m,n}$ excludes the tail region where empirical distribution functions overlap. We will return to this issue later in Section 4.4, with a demonstration on our motivating dataset.

Now we are in a position to introduce our test statistics.

Minimum t-statistic: This statistic was first introduced by Kaur *et al.* (1994) in context of second order stochastic dominance, and then extended to the first order by Davidson and Duclos. For distribution functions F_1 and F_2, this statistic is given by

$$T_{m,n}^{\min}(F_1, F_2) = \inf_{x \in D_{p,m,n}} \frac{\Big(F_2(x) - F_1(x)\Big)}{\sqrt{\dfrac{F_1(x)\Big(1 - F_1(x)\Big)}{m} + \dfrac{F_2(x)\Big(1 - F_2(x)\Big)}{n}}}. \tag{4.3}$$

Our tests reject the $\mathcal{H}_0$ for large values of $T_{m,n}^{\min}(\widehat{F}_m, \widehat{G}_n)$, $T_{m,n}^{\min}(\tilde{F}_m, \tilde{G}_n)$, and $T_{m,n}^{\min}(\mathbb{F}_m, \mathbb{G}_n)$. The last test-statistic, which is nonparametric, equals the minimum t-statistic of the Kaur *et al.* in context of first order stochastic dominance.

Two sample empirical process (TSEP) type test statistic: Our second test rejects the $\mathcal{H}_0$ for large values of $T_{m,n}^{\text{tsep}}(\widehat{F}_m, \widehat{G}_n)$, $T_{m,n}^{\text{tsep}}(\tilde{F}_m, \tilde{G}_n)$, or $T_{m,n}^{\text{tsep}}(\mathbb{F}_m, \mathbb{G}_n)$, where for distribution functions F_1 and F_2, $T_{m,n}^{\text{tsep}}$ is defined by

$$T_{m,n}^{\text{tsep}}(F_1, F_2) = \sqrt{\frac{mn}{N}} \inf_{z \in [p, 1-p]} \frac{F_2(\mathcal{H}^{-1}(z)) - F_1(\mathcal{H}^{-1}(z))}{\sqrt{z(1 - z)}}. \tag{4.4}$$

Ledwina and Wyłupek (2013) uses a test statistic similar to $T_{m,n}^{\text{tsep}}(\mathbb{F}_m, \mathbb{G}_n)$ (the second test statistic in Section 2.2 of Ledwina and Wyłupek, 2013) for testing the null of stochastic dominance against non-dominance. The pivotal distribution of their test statistic is completely different from ours because they based their

critical values on the configuration $F = G$, which is very different from what we will consider. We are not aware of any existing test which uses $T_{m,n}^{\mathrm{tsep}}(\mathbb{F}_m, \mathbb{G}_n)$ for testing non-dominance against stochastic dominance.

Wilcoxon rank sum (WRS) type test statistic: Wilcoxon rank sum (WRS) test is widely used for comparing two vaccines (cf. Miladinovic *et al.*, 2014) although the WRS test is actually designed for testing location shift. It is a popular choice for testing the null $F = G$ against the alternative $F = G(\cdot - \delta)$ for $\delta > 0$ (Lee and Wolfe, 1976). The WRS test is the most powerful nonparametric test for testing the following hypotheses (cf. Example 25.46 of Van der Vaart, 1998):

$$\mathcal{H}_0^a : \int_{\mathbb{R}} G(z)dF(z) \geq 1/2 \quad vs \quad \mathcal{H}_1^a : \int_{\mathbb{R}} G(z)dF(z) < 1/2. \qquad (4.5)$$

Although the WRS test is not designed to test the null of non-dominance, we include this test to demonstrate its failure to control the type I error at some null configurations.

The one-sided WRS test rejects $\mathcal{H}_0$ for large values of $T_{m,n}^{\mathrm{wrs}}(\mathbb{F}_m, \mathbb{G}_n)$, where, for distribution functions F_1 and F_2,

$$T_{m,n}^{\mathrm{wrs}}(F_1, F_2) = \sqrt{\frac{12mn}{N+1}} \left(\int_{\mathbb{R}} F_2(x)dF_1(x) - 1/2 \right). \qquad (4.6)$$

$T_{m,n}^{\mathrm{wrs}}(\mathbb{F}_m, \mathbb{G}_n)$ is the Mann-Whitney form of the two-sample WRS statistic. The corresponding shape-constraint versions are given by $T_{m,n}^{\mathrm{wrs}}(\widehat{F}_m, \widehat{G}_n)$ and $T_{m,n}^{\mathrm{wrs}}(\tilde{F}_m, \tilde{G}_n)$.

We excluded tests based on the smoothed log-concave MLE because rigorous asymptotic analysis of the corresponding tests is out of the scope of the present paper. However, our empirical study in Section 4.3 includes minimum t-test and TSEP test based on the smoothed log-concave MLE. Our simulations indicate that the asymptotic critical values of the tests based on $T_{m,n}^{\mathrm{min}}(\tilde{F}_m, \tilde{G}_n)$ and $T_{m,n}^{\mathrm{tsep}}(\tilde{F}_m, \tilde{G}_n)$ are valid for the corresponding smoothed log-concave tests. Our simulations also indicate that the finite sample performance of the tests based on the log-concave MLE and the smoothed log-concave MLE are quite similar. We leave the rigorous analysis of the tests based on the smoothed log-concave MLE for future study.

Remark 1. Since the nonparametric tests use the empirical distribution function, shape constrained methods do not gain any advantage in terms of tuning parameters. Also, we will see that the nonparametric and shape constrained tests are asymptotically equivalent. For moderate sized samples, however, our simulations in Section 4.3 show that the shape-constrained tests exhibit better performance.

Remark 2. Davidson and Duclos proposed an empirical likelihood ratio approach to test $\mathcal{H}_0$ vs $\mathcal{H}_1$. We did not appeal to this nonparametric approach in this paper because this approach does not extend easily to shape-constrained scenarios. Regardless, we point out that Davidson and Duclos showed that their

empirical likelihood ratio test is asymptotically equivalent to the minimum t-test.

In the sequel, we may use the terms "log-concave" or "unimodal" to refer to the test statistics based on the log-concave or unimodal density estimators. For example, we may refer to $T_{m,n}^{\min}(\widehat{F}_m, \widehat{G}_n)$ and $T_{m,n}^{\min}(\tilde{F}_m, \tilde{G}_n)$ as the unimodal minimum t-statistic and the log-concave minimum t-statistic, respectively. Also, unless otherwise specified, the terms "null" and "alternative" will refer to the $\mathcal{H}_0$ and $\mathcal{H}_1$ defined in (4.2), respectively.

4.2. Asymptotic distribution

In this section, we explore the asymptotic distribution of our test statistics. We show that the minimum t-test and the two sample empirical process (TSEP) test asymptotically control type I error for each null configuration and they are asymptotically consistent against each $(F, G) \in \mathcal{H}_1$. We also show that, with the exception of the test based on log-concave MLE, the WRS type tests control the type I error for distributions in $\mathcal{H}_0^a$ and are consistent against $(F, G) \in \mathcal{H}_1^a$. We first prove the asymptotic results on the nonparametric test statistics. Then, we show that the shape-constrained test statistics are equivalent to their nonparametric counterparts up to a $o_p(N^{-1/2})$ term, which implies that the same critical values can be used for them.

We exclude the log-concave MLE based WRS statistic from our discussion because we are unable to infer on its asymptotic limit. The difficulty arises due to our inability to track the asymptotic behaviour of $\sqrt{m}||\tilde{F}_m - \mathbb{F}_m||_\infty$. In the remainder of this section, by "shape-constrained test statistics", we will therefore refer to $T_{m,n}^{\min}(\widehat{F}_m, \widehat{G}_n)$, $T_{m,n}^{\text{tsep}}(\widehat{F}_m, \widehat{G}_n)$, $T_{m,n}^{\text{wrs}}(\widehat{F}_m, \widehat{G}_n)$, $T_{m,n}^{\min}(\tilde{F}_m, \tilde{G}_n)$, and $T_{m,n}^{\text{tsep}}(\tilde{F}_m, \tilde{G}_n)$ only. For the sake of clarity, in Table 2, we summarize the current state of results on the different tests discussed in this paper.

Before going into further details, we state a technical condition that will be required by all of our Theorems.

Condition N. F and G are continuous. Also, F and G have densities f and g, respectively, such that $\text{supp}(f) \cup \text{supp}(g)$ contains an open neighborhood of D_p.

The first requirement of Condition N, i.e., the continuity of F and G, is necessary for the weak convergence of the empirical processes to Brownian bridges. The second requirement ensures that $\text{dist}(D_{p,m,n}, D_p)$ approaches 0 as $m, n \to \infty$ with probability one. These assumptions are likely to be satisfied by our immune response data provided p is not too small (see Figure 1i and Figure 1ii). For the rest of the paper, we restrict our attention to F and G that satisfy Condition N.

TABLE 2
Table summarizing different tests used in this paper and current results on them.

Test	Asymptotic results	Empirical results
	Nonparametric	
Minimum t-test	previously known (Kaur *et al.*, 1994)	included in this paper
TSEP	we derived	— — —″ — ——
WRS	previously known (Dwass, 1956)	— — —″ — ——
	Unimodal	
Minimum t-test	we derived	included in this paper
TSEP	— — —″ — ——	— — —″ — ——
WRS	— — —″ — ——	— — —″ — ——
	Log-concave	
Minimum t-test	we derived	included in this paper
TSEP	— — —″ — ——	— — —″ — ——
WRS	unknown	— — —″ — ——
	Smoothed log-concave	
Minimum t-test	unknown	included in this paper
TSEP	— — —″ — ——	— — —″ — ——
WRS	— — —″ — ——	not included

4.2.1. Asymptotic critical values of the nonparametric tests

We begin our discussion with the minimum t-statistic and the TSEP statistic. Our first objective is to identify the null configurations that lead to the highest asymptotic type I error. Here we remind the readers that (F, G) is a null configuration if there exists $x \in D_p$ so that $G(x) \leq F(x)$. One may guess that the interesting cases appear on the boundary of $\mathcal{H}_0$. However, to formally discuss the boundary of $\mathcal{H}_0$, we need to equip it with a suitable topology. To formalize our discussion, we consider the space $\mathcal{F}$ of all continuous distribution functions on $\mathbb{R}$, and equip it with the uniform metric $d(F, F') = \sup_{x \in \mathbb{R}} |F(x) - F'(x)|$. Consider the product space $\mathcal{F} \times \mathcal{F}$ with the metric

$$d_2\Big((F, G), (F', G')\Big) = \max\left\{d(F, F'), d(G, G')\right\}.$$

By an abuse of notation, we denote by $\mathcal{H}_0$ and $\mathcal{H}_1$ the set of all combinations $(F, G) \in \mathcal{F} \times \mathcal{F}$ that satisfy the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. For $i = 0, 1$, we denote the closure of $\mathcal{H}_i$ in $\mathcal{F} \times \mathcal{F}$ by $\text{cl}(\mathcal{H}_i)$. Then the boundary of $\mathcal{H}_i$ is given by $\text{cl}(\mathcal{H}_i) \setminus \text{int}(\mathcal{H}_i)$. The following lemma characterizes $\text{bd}(\mathcal{H}_0)$ and $\text{int}(\mathcal{H}_0)$.

Lemma 1. $\mathcal{H}_0$ *is a closed subset of* $\mathcal{F} \times \mathcal{F}$ *with boundary*

$$bd(\mathcal{H}_0) = \left\{(F, G) \in \mathcal{F} \times \mathcal{F} \ : \ \sup_{x \in D_p}\Big(F(x) - G(x)\Big) = 0\right\}.$$

Moreover, $bd(\mathcal{H}_0) = bd(\mathcal{H}_1)$. Also, the interior of $\mathcal{H}_0$ is given by

$$int(\mathcal{H}_0) = \left\{ (F, G) \in \mathcal{F} \times \mathcal{F} \ : \ F(z) > G(z) \text{ for some } z \in D_p \right\}.$$

Figure 3(ii) gives an example of an (F, G) pair in the interior of $\mathcal{H}_0$. The following lemma entails that the minimum t-statistic and the TSEP statistic are asymptotically degenerate on $int(\mathcal{H}_0)$.

Lemma 2. *Suppose $(F, G) \in int(\mathcal{H}_0)$ satisfies Condition N, $m/N \to \lambda$, and $supp(f) \cup supp(g)$ contains an open neighborhood of D_p. Then,*

$$T_{m,n}^{min}(\mathbb{F}_m, \mathbb{G}_n) \to_p -\infty \quad \text{and} \quad T_{m,n}^{tsep}(\mathbb{F}_m, \mathbb{G}_n) \to_p -\infty.$$

The proof of Lemma 2 for the minimum t-statistic can be found in Whang (2019) (see also Davidson and Duclos, 2013). However, we include it in Appendix 2 for the sake of completeness. Lemma 2 indicates that non-trivial type I errors can originate only at the boundary of $\mathcal{H}_0$, which is a subset of $\mathcal{H}_0$ because the latter is a closed set (see Lemma 1). Lemma 1 also implies that $bd(\mathcal{H}_0)$ consists of all those F and G that touch each other on D_p. To concretize this idea, we define the contact set C_p by

$$C_p = \{x \in D_p \ : \ F(x) = G(x)\}.$$

Note that if $(F, G) \in bd(\mathcal{H}_0)$, then $C_p \neq \emptyset$, where $F = G$ on C_p, and $F < G$ on $D_p \setminus C_p$. Let us also define

$$H(C_p) = \{t \in [p, 1-p] \mid H^{-1}(t) \in C_p\}. \tag{4.7}$$

Theorem 1 shows that the asymptotic distribution of the test statistics on $bd(\mathcal{H}_0)$ crucially depends on this contact set C_p and $H(C_p)$. The proof of Theorem 1 is given in Appendix A.0.2.
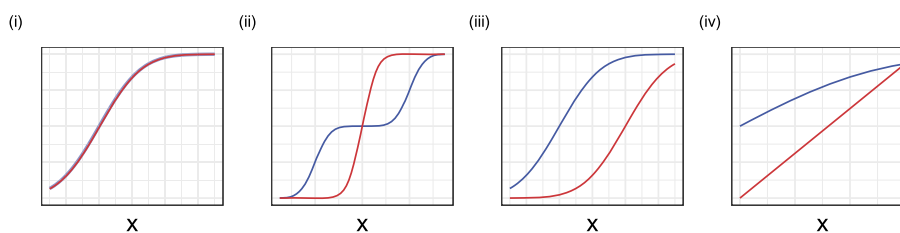

Fig 3: This figure displays plots of two distribution functions F (red) and G (blue). The range of x in these plots correspond to $D_p := D_p(F, G)$. (i) $F = G$ on D_p. This is a null configuration. (ii) F and G cross each other on D_p. This is also a null configuration. (iii) F strictly stochastically dominates G. In fact $G(x) > F(x)$ for all $x \in D_p$. This is an alternative configuration. (iv) F and G touch each other at the endpoint of D_p. This is a null configuration.

Theorem 1. *Suppose $(F, G) \in bd(\mathcal{H}_0)$, $m/N \to \lambda$, and F and G have continuous densities f and g satisfying*

$$\inf_{x \in D_p} \min\{f(x), g(x)\} > 0,$$

where $D_p = [H^{-1}(p), H^{-1}(1-p)]$. Let $\mathbb{U}$ denote a standard Brownian bridge. Then under the stated conditions, the following assertions hold:

A.

$$T_{m,n}^{min}(\mathbb{F}_m, \mathbb{G}_n) \to_d \inf_{x \in C_p} \frac{\mathbb{U} \circ F(x)}{\sqrt{F(x)(1 - F(x))}}, \tag{4.8}$$

where $C_p = \{x \in D_p \ : \ F(x) = G(x)\}$.

B.

$$T_{m,n}^{tsep}(\mathbb{F}_m, \mathbb{G}_n) \to_d \inf_{t \in H(C_p)} \sqrt{\frac{\lambda}{1 - \lambda}} \frac{\mathbb{L}_0(t)}{\sqrt{t(1 - t)}},$$

*where $\mathbb{L}_0$ is the centred Gaussian process given by (A.17) of Appendix A, and $H(C_p)$ is as in (4.7).*

C. *In particular, if $C_p \subset int(D_p)$, then*

$$T_{m,n}^{tsep}(\mathbb{F}_m, \mathbb{G}_n) \to_d \inf_{t \in H(C_p)} \frac{\mathbb{U}(t)}{\sqrt{t(1 - t)}}.$$

The Gaussian process $\mathbb{L}_0$, which is defined in (A.17) of Appendix A, depends on F and G. We postpone further discussion on the form of $\mathbb{L}_0$ till Appendix A. Next we discuss the implication of Theorem 1 on the minimum t-test. Then we will discuss the case of the TSEP test.

Asymptotic critical value of minimum t-test: Theorem 1 reveals an interesting fact: the length of C_p imposes a stochastic ordering among the limiting laws of the minimum t-statistic for the boundary configurations. To elaborate, let us consider (F_1, G_1) and $(F_2, G_2) \in bd(\mathcal{H}_0)$, with respective contact sets C_p^1 and C_p^2. Then, under the conditions of Theorem 1,

$$T_{m,n}^{\min}(\mathbb{F}_m, \mathbb{G}_n) \xrightarrow{F_i, G_i}_d \inf_{x \in C_p^i} \frac{\mathbb{U} \circ F(x)}{\sqrt{F(x)(1 - F(x))}}, \quad \text{for } i = 1, 2.$$
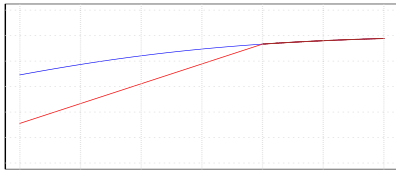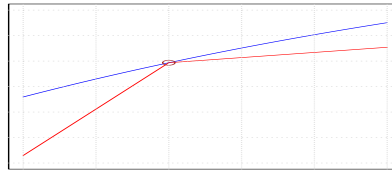
(i) $(F, G) \in bd(\mathcal{H}_0)$ but not LFC   (ii) (F, G) is LFC

Fig 4: Illustration of F (red) and G (blue) on $bd(\mathcal{H}_0)$. (i) The contact set C_p is an interval (ii) The contact set is singleton; so this is an LFC configuration.

If the contact sets satisfy the ordering $C_p^1 \subset C_p^2$, then

$$\inf_{z \in C_p^2} \frac{\mathbb{U} \circ F(z)}{\sqrt{F(z)(1 - F(z))}} \preceq \inf_{z \in C_p^1} \frac{\mathbb{U} \circ F(z)}{\sqrt{F(z)(1 - F(z))}},$$

implying that the limiting law of the minimum t-statistic under (F_1, G_1) stochastically dominates that under (F_2, G_2). The extreme cases for C_p^1 are the singleton sets $\{x\}$, where $x \in D_p$. In this case, the asymptotic distribution of both test statistics is standard Gaussian. Therefore, we set the critical value of our tests to be z_α, the $(1-\alpha)$-th quantile of the standard Gaussian distribution. The class of boundary configurations with a singleton contact set is referred to as "the least favorable class" (LFC) (Davidson and Duclos, 2013). Figure 4 illustrates the difference between an LFC and an ordinary non-LFC boundary combination. Theorem 2 of Davidson and Duclos shows that there is no null configuration under which the law of the minimum t-statistic strictly stochastically dominates that of the LFC configuration. This result, which holds for any m and n, implies that, among the null configurations, the LFC configurations lead to the greatest dominance of F over G. The above finding, in conjunction with our Theorem 1, imply that our tests, whose critical values are based on the LFC class, is likely to have asymptotic size α. This being a stronger assertion than the asymptotic control of type I error can be an interesting topic for further investigation.

Remark 3. The asymptotic behavior of the nonparametric minimum t-statistic has been previously studied (Davidson and Duclos, 2013, Kaur *et al.*, 1994). However, previous studies focus only on the asymptotic behaviour of the minimum t-statistic at the LFC configurations and the interior of $\mathcal{H}_0$, whereas our results show that there are other classes of boundary configurations with non-vanishing type I error. Although existing results are enough for the purpose of constructing critical values, our new results provide a more complete understanding of the scenario.

Asymptotic critical value of the TSEP test: The asymptotic distribution of the TSEP statistic under $\mathcal{H}_0$ also exhibits a monotonocity property similar to the minimum t-test. To elaborate, suppose the pairs (F_1, G_1) and (F_2, G_2) have respective contact sets C_p^1 and C_p^2 satisfying $C_p^1 \subset C_p^2$. Then

$$\inf_{t \in H(C_p^2)} \frac{\mathbb{L}_0(t)}{\sqrt{t(1 - t)}} \preceq \inf_{t \in H(C_p^1)} \frac{\mathbb{L}_0(t)}{\sqrt{t(1 - t)}}.$$

Therefore, similar to the case of the minimum t-test, the configurations with singleton $H(C_p)$ constitute the class of LFC configurations for the TSEP test. Under Condition N, H is continuous and strictly increasing on C_p. Therefore, $H(C_p)$ is singletone if and only if C_p is singletone, i.e., C_p is of the form $\{x\}$, where $x \in D_p$. To find the critical value, it suffices to study the asymptotics in such LFC cases. If $x \in \text{int}(D_p)$, the TSEP test statistic weakly converges to a standard Gaussian distribution by part C of Theorem 1. Till this point, there has been no difference between the TSEP and the minimum t-test statistic.

If, however, $x \in \mathrm{bd}(D_p)$ (see Figure 15iv), the asymptotic distribution of the TSEP test statistic can be different. To this end, first we state a lemma, and then using this lemma, we explain the asymptotics of the TSEP test when $C_p = \{x\} \subset \{p, 1-p\}$.

Lemma 3. *Suppose F and G are as in Theorem 1 and $H(C_p) = \{t\}$ for some $t \in [p, 1-p]$. Then*

$$\sqrt{\frac{\lambda}{1-\lambda}} \frac{\mathbb{L}_0(t)}{\sqrt{t(1-t)}} \sim N(0, \sigma^2_{TSEP}),$$

where

$$\sigma^2_{TSEP} = \frac{\lambda(f \circ H^{-1}(t))^2 + (1-\lambda)(g \circ H^{-1}(t))^2}{\left(\lambda f \circ H^{-1}(t) + (1-\lambda)g \circ H^{-1}(t)\right)^2}. \tag{4.9}$$

Moreover, the following assertions also hold:

A. *$\sigma^2_{TSEP} = 1$ if and only if $f \circ H^{-1}(t) = g \circ H^{-1}(t)$. Otherwise,*

$$1 < \sigma^2_{TSEP} \le \max\{(1-\lambda)^{-1}, \lambda^{-1}\}. \tag{4.10}$$

B. *Given any $\epsilon \in (0, 2(1-\lambda)/\lambda)$, we can find a constant $C_\lambda > 0$, depending only on $\lambda > 0$, so that whenever $g(H^{-1}(t))/f(H^{-1}(t)) < C_\lambda$, then $\sigma^2_{TSEP} > \lambda^{-1} - \epsilon$.*

C. *Given any $\epsilon \in (0, 2\lambda/(1-\lambda))$, we can find a constant $C'_\lambda > 0$, depending only on $\lambda > 0$, so that whenever $f(H^{-1}(t))/g(H^{-1}(t)) < C'_\lambda$, then $\sigma^2_{TSEP} > (1-\lambda)^{-1} - \epsilon$.*

Lemma 3 has some interesting consequences. First, if $C_p = \{H^{-1}(p)\}$ or $\{H^{-1}(1-p)\}$, then using Theorem 1B and Lemma 3 one can show that $T^{\text{tsep}}_{m,n}$ converges weakly to a centred Gaussian distribution with variance σ^2_{TSEP}. Part A of Lemma 3 implies that if F and G touch at C_p, i.e. if $f = g$ at the point of contact, then σ^2_{TSEP} is still one. Hence, the TSEP test statistic is asymptotically standard Gaussian for this case. However, if F and G cross at the point of contact instead of touching, i.e. if $f \ne g$ at the point of contact, then $\sigma^2_{TSEP} > 1$. The precise value of σ^2_{TSEP} is given by (4.9). Moreover, the value of σ^2_{TSEP} increases as the value of f and g diverges at the point of contact. On one hand, if f is much larger than g, then part B of Lemma 3 implies σ^2_{TSEP} is close to λ^{-1}. On the other hand, if g is much larger than f, then part C of Lemma 3 indicates that σ^2_{TSEP} is close to $(1-\lambda)^{-1}$. These bounds are tight because, by part B of Lemma 3, σ^2_{TSEP} can not be larger than $\max\{\lambda^{-1}, (1-\lambda)^{-1}\}$ under the set up of Theorem 1.

The above discussion leads to the following conclusion for the TSEP test. If we want to control the asymptotic type I error of the TSEP test at all null configurations, then we should use the critical value $C_{m,n}z_\alpha$ where $C_{m,n} = \max\{\sqrt{N/m}, \sqrt{N/n}\}$. There is a caveat, however. To see this, we begin by noting that $C_{m,n} \ge \sqrt{2}$, with equality holding only when $m = n$. For example, for our motivating dataset, $C_{m,n} \approx 1.9$. However, if $m/N \to 0$ or $n/N \to 0$,

then $C_{m,n} \to \infty$. Thus if m and n are not close to being equal, $C_{m,n}$ can be a large quantity. Therefore, using $C_{m,n}z_\alpha$ as critical value yields a conservative test. Hence, we will call the corresponding TSEP tests the conservative TSEP tests. Our simulations indicate that conservative TSEP tests have poorer power compared to the minimum t-test even when m and n are equal, and their power keeps degrading as $C_{m,n}$ increases.

If we use instead use the critical value z_α, then we control the asymptotic type I error at the null configurations with $C_p \subset \mathrm{int}(D_p)$ or those with $f = g$ at C_p. This only excludes the null cases where F and G may cross at $\mathrm{bd}(D_p)$ (see Figure 3 iv). These type of configurations can be considered pathological cases. Moreover, our simulations in Section 4.3 (see case b) show that even when F and G cross at $\mathrm{bd}(D_p)$, the TSEP tests with critical value z_α control the type I error. The TSEP tests with critical value z_α have decent power and their overall performance is comparable with the minimum t-tests. In view of the above, we recommend using the asymptotic critical value z_α when using the TSEP test.

Our final result on the minimum t-test and the TSEP test establishes their asymptotic consistency.

Theorem 2. *Suppose* $(F, G) \in \mathcal{H}_1$ *satisfy Condition N. Then if* $m/N \to \lambda$, *then*
$$\lim_{m,n\to\infty} T_{m,n}^{min}(\mathbb{F}_m, \mathbb{G}_n) \to_p \infty \quad and \quad T_{m,n}^{tsep}(\mathbb{F}_m, \mathbb{G}_n) \to_p \infty.$$

The asymptotic distribution of the WRS statistic is well-established in the literature. Suppose F and G are continuous distribution functions. In that case, it is well known that, when $F = G$, the WRS statistic is asymptotically distributed as a standard gaussian random variable, i.e. $T_{m,n}^{\mathrm{wrs}}(\mathbb{F}_m, \mathbb{G}_n) \to_d N(0, 1)$ (Dwass, 1956). If $(F, G) \in H_0^a$ satisfies $\int_\mathbb{R} G(z)dF(z) > 1/2$, however, $T_{m,n}^{\mathrm{wrs}}(\mathbb{F}_m, \mathbb{G}_n) \to_p -\infty$, whereas for $(F, G) \in \mathcal{H}_1^a$, we have $T_{m,n}^{\mathrm{wrs}}(\mathbb{F}_m, \mathbb{G}_n) \to_p \infty$.

4.2.2. Asymptotic critical values of the shape-constrained tests

We will show that under some additional conditions, the difference between the nonparametric and the shape-constrained test statistics is $o_p(1)$, which automatically implies that the shape-constrained tests enjoy the same asymptotic properties as the nonparametric tests.

For the unimodal case, the additional condition is a curvature condition, which requires f and g to be nowhere flat within their respective domains.

Condition A. For the density μ, the Lebesgue measure of the set $\{\mu' = 0, \mu > 0\}$ is 0, where μ' is the derivative of μ.

For densities satisfying Condition A, $\sqrt{m}(\widehat{F}_m^0 - F)$ almost surely weakly converges to $\mathbb{V} \circ F$, where $\mathbb{V}$ is a Brownian bridge. In this case, it can be shown that (Kiefer and Wolfowitz, 1976) $\sqrt{m}\|\widehat{F}_m^0 - \mathbb{F}_m\|_\infty = o_p(1)$. Our Lemma A.5 in Appendix A.1 states that $\sqrt{m}\|\widehat{F}_m^0 - \widehat{F}_m\|_\infty = o_p(1)$, which implies $\sqrt{m}\|\widehat{F}_m - \mathbb{F}_m\|_\infty = o_p(1)$ in this case. However, under the violation of Condition A, the process $\sqrt{m}(\widehat{F}_m^0 - F)$ no longer converges to $\mathbb{V} \circ F$ weakly. The densities that

violate Condition A form the boundary of the class of unimodal densities. For these densities, $\sqrt{m}\|\widehat{F}_m^0 - \mathbb{F}_m\|_\infty$, and hence, $\sqrt{m}\|\widehat{F}_m - \mathbb{F}_m\|_\infty$ is non-negligible. The limiting process of $\sqrt{m}(\widehat{F}_m^0 - F)$ in this case is slightly convoluted, and we refer to Beare *et al.* (2017), Carolan and Dykstra (1999), Carolan (2002) for more details on the limiting process. Just to give an example, if $f \sim U[0, 1]$, then the limiting process is the least concave majorant of a Brownian bridge (Carolan and Dykstra, 1999). Condition A is thus required to ensure that neither f nor g is one of these problematic boundary densities.

Lemma 4. *Suppose that f and g are unimodal densities satisfying Condition A. Further suppose that f and g are bounded away from 0 on an open set containing D_p, and m, n satisfy $m/N \to \lambda$. Then,*

$$|T_i(\mathbb{F}_m, \mathbb{G}_n) - T_i(\widehat{F}_m, \widehat{G}_n)| \to_{a.s.} 0, \quad for\ i = 1, 2, 3.$$

In case of the log-concave test statistics, however, we require a smoothness condition as well as a curvature condition. We will quantify smoothness via a Hölder condition. For a compact set $K \subset \mathbb{R}$, a function h is said to be in the Hölder class $\mathcal{H}^{\beta, L}(K)$ with exponent $\beta \in [1, 2]$ and constant $L > 0$ if, for all $x, y \in K$, $|h(x) - h(y)| \leq L|x - y|$ if $\beta = 1$ and $|h'(x) - h'(y)| \leq L|x - y|^{\beta-1}$ if $\beta > 1$.

We say that a density μ ($\mu = f$ or g) satisfies Condition B1 if the following holds.

Condition B1. There exists $\beta \in [1, 2]$, $L > 0$, and a compact $K \subset \mathbb{R}$ such that the density μ satisfies $\log \mu \in \mathcal{H}^{\beta, L}(K)$.

In addition to Condition B1, we also require f and g to satisfy a curvature condition.

Condition B2. μ is a log-concave density with log-density $\phi = \log \mu$. Suppose $K \subset \text{dom}(\phi)$ is compact. Then there exists $C > 0$ such that all $x, y \in K$ satisfying $x < y$ obeys

$$\phi'(x) - \phi'(y) \geq C(y - x),$$

where ϕ' is the left derivative or the right derivative of ϕ.

Note that, since ϕ is concave, its left and right derivatives always exist. If ϕ' is differentiable on K, Condition B2 reads as $\phi''(x) \leq -C$ for $x \in K$. Conditions of type B1 and B2 also appear in Dümbgen and Rufibach.

Lemma 5. *Suppose that f and g are log-concave densities satisfying Conditions B1 and B2. Suppose, further, f and g are bounded away from 0 on an open set containing D_p, and $m/N \to \lambda$. Then it follows that $|T_i(\mathbb{F}_m, \mathbb{G}_n) - T_i(\tilde{F}_m, \tilde{G}_n)| = o_p(1)$ for $i = 1, 2$.*

Remark 4. All results of Section 4.2 hold if we replace D_p by a compact set D as long as there is a $p > 0$ so that $D \subset D_p$ and D_p satisfies the conditions of the theorems stated in Section 4.2.
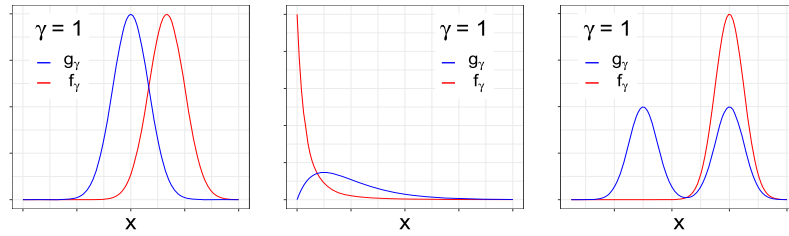
Fig 5: Plots of the densities f_γ and g_γ corresponding to the cases (a), (d), and (e) for $\gamma = 1$.

4.3. Simulations

This section compares the performance of the shape-constrained tests designed in Section 4.1 with their nonparametric counterparts. We let $m = n = 100$, which is reflective of the sample sizes anticipated in many phase 1b or phase 2 vaccine trials — for example, our motivating dataset has $m + n = 248$. For all TSEP tests in this section, we use the critical value z_α. See Appendix C for the simulations with TSEP tests that have critical value $C_{m,n}z_\alpha$. For the TSEP and the minimum t-test, we also include the tests based on the smoothed log-concave MLE. We will refer to the corresponding test as the smoothed log-concave test. Although we do not have any theoretical result for this test, we use the asymptotic critical value z_α.

By the design of our hypotheses, $\mathcal{H}_0$ encompasses a broad number of cases ranging from $G \succ F$, $F = G$ to cases where F and G touch or cross each other. We develop simulation schemes so that we can explore a wide range of scenarios. Our simulation schemes involves a parameter γ varying over the range $[0, 1]$. Here γ quantifies the difference between the data generating distribution functions F_γ and G_γ. We evaluate the power $\nu(\gamma)$ at a grid of equally spaced points in $[0, 1]$.

For our simulation study, we consider the following cases:

(a) $F_\gamma \sim N(\gamma, 1)$, and $G_\gamma \sim N(0, 1)$.

(b) $F_\gamma \sim N(3\gamma, 1)$, and $G_\gamma \sim N(0.5, 2)$.

(c) $F_\gamma \sim Gamma(2, 0.1+0.4\gamma)$ and $G_\gamma \sim Gamma(1, 0.5)$, where $Gamma(a, b)$ is a Gamma random variable with shape parameter a and scale parameter b.

(d) $F_\gamma \sim Gamma(2, 1)$ and $G_\gamma \sim Pareto(0.5+2\gamma, 1)$. Here $Pareto(a, b)$ is the Pareto distribution function with shape parameter a and scale parameter b.

(e) $F_\gamma \sim N(0, 1)$ and $G_\gamma \sim N(2\gamma + 4, 1)/2 + N(2\gamma - 2, 1)/2$.

Case (a) corresponds to the traditional setting of null of equity against a shift alternative. The last four cases cover $\mathcal{H}_0$ combinations that include crossing. In those cases, F_γ and G_γ cross each other on $\text{int}(D_p)$ at $\gamma = 0$. As γ increases, however, F_γ and G_γ eventually touch (cases c, d, and e), or cross (case b) each other at $\text{bd}(D_p)$, generating a LFC configuration. We denote the corresponding

γ by γ^*. Finally, at $\gamma = 1$, F_γ strictly dominates G_γ in the sense of $\mathcal{H}_1$. Figures 18i, 17i, 17ii, and 18ii in Appendix D display the plots of F_γ and G_γ for several values of γ in cases (b), (c), (d) and (e), respectively. Figure 5 illustrates the densities in cases (a), (d), and (e) for $\gamma = 1$. Cases (d) and (e) are chosen to reflect violations of the shape constraints that we consider. The Pareto density in (d) violates the log-concavity assumption and the normal mixture in (e) violates both the unimodality and the log-concavity assumptions. All other densities satisfy both shape constraints.

We evaluate the properties of the tests under consideration using 10,000 Monte Carlo replicates. We set the p in $D_{p,m,n}$ to be 0.05, where $D_{p,m,n}$ was defined to be the set $[\mathbb{H}_N^{-1}(p), \mathbb{H}_N^{-1}(1-p)]$. Also, the level of significance is 0.05 for all our tests. The tests based on the empirical cumulative distribution function will be referred to as NP (nonparametric) tests. For brevity, we will refer to the nonparametric, unimodal, log-concave and smoothed log-concave tests by NP, UM, LC, and smoothed LC tests, respectively.

Figure 7 displays the power curves for the minimum t-test and the TSEP test. In terms of power, the LC and smootheed LC tests generally outperform the UM tests, which generally outperform the NP tests. In their simulation study, Davidson and Duclos (2013) reported the NP minimum t-test to be conservative, which aligns with our observation. Also, the shape-constrained minimum t-tests have slightly higher power than the shape-constrained TSEP tests, although the difference is not always significant. For the NP tests, the rejection rates of the minimum t-tests and the TSEP tests are almost identical.

Except for in case (d), where G_γ is Pareto, all tests control the type I error in all null set-ups, including the LFC configuration where $\gamma = \gamma^*$. Since LFC configurations constitute the boundary of $\mathcal{H}_0$, this observation indicates that our tests have size 0.05 for all cases except case (d). Although Pareto density violates only the log-concavity assumption, apparently no test has the correct size, albeit the NP and UM tests performing the best in terms of size. Also, for case (d), the smoothed LC test has lower type I error than LC test. Surprisingly, in case (e), where both log-concavity and unimodality are violated, the shape-constrained tests have overall better performance than the NP tests although all tests exhibit poor power in this case.

All the WRS type tests, including the LC WRS test (whose asymptotic behavior is yet unknown), exhibit a much larger size than 0.05 in all cases except the null of equity type case (a), which is the ideal scenario for WRS type tests. Figure 8 illustrates the power curve $\nu(\gamma)$ in cases (b) and (c), where all three WRS tests exhibit very high type I error at several null configurations. Our findings are consistent with the observation in Ledwina and Wyłupek (2012) that the WRS test can be misleading for testing the null of non-dominance.

In summary, under a correctly specified model, the shape-constrained minimum t-test and TSEP test outperform their nonparametric counterparts, with the log-concave minimum t-test having the best power. When the shape constraints are violated, the nonparametric tests are not distinguishably better than the unimodal tests. On the other hand, The WRS tests do not control type I error for null configurations with crossing, as expected. Another impor-

tant takeaway from this section is that with the critical value z_α, the smoothed log-concave tests seem to perform as well as the log-concave tests under log-concavity.

The current article uses asymptotic critical values for performing the above-mentioned tests, but bootstrap critical values could also be an option. Bootstrap for the shape-constrained tests, however, is not straightforward. Generation of observations from shape-constrained LFC configuration poses some challenges, which involves solving non-trivial optimization problems. Further discussion in this direction is out of the scope of the present paper. Therefore, we leave bootstrap tests for future research.

4.4. Application to HVTN 097 and HVTN 100 data

Our first task is to select a p for choosing $D_{p,m,n}$. Figure 1 and some inspection show that the empirical distribution functions are very close on the sets $(-\infty, \mathcal{H}^{-1}(0.072)]$ and $[\mathcal{H}^{-1}(0.975), \infty)$ in that they either cross or touch each other on these regions. Therefore, this is the problematic region we wish to exclude from our $D_{p,m,n}$, because clearly there is not enough evidence of any dominance in this region. Therefore, we set the p in $D_{p,m,n}$ to be 0.075. We remark that it is ideal to choose p in a systematic way without looking at the data. However, constructing a rigorous procedure for choosing p is out of the scope of the present paper, and we leave it for future research. Table 3, which tabulates the p-values of the tests, displays that all tests reject the null at the level of significance 0.05. The highest p-value is observed for the NP TSEP test, which is approximately 0.042.

TABLE 3
Table of the p-values of different tests applied on our data. Here the p in $D_{p,m,n}$ is set to be 0.075. The critical value was z_α for all the tests.

Tests	Nonparametric	Unimodal	Log-concave	Smoothed log-concave
Minimum t-test	0.015	0.007	0.007	0.001
TSEP	0.042	0.037	0.035	0.030

It is natural to ask if we can estimate the power of our tests at (F_{097}, F_{100}). Because (F_{097}, F_{100}) is unavailable, we analyze the power in a neighborhood of $(\tilde{F}_{097}^{sm}, \tilde{F}_{100}^{sm})$ instead, where $\tilde{F}_{097}^{sm}$ and $\tilde{F}_{100}^{sm}$ correspond to the distributions of the smoothed log-concave MLE (Chen and Samworth, 2013) estimators of f_{100} and f_{097}, respectively.

Let us denote the smoothed log-concave MLE of the pooled sample by $\tilde{f}_{m,n}^0$. Letting $\tilde{F}_{m,n}^0$ denote the corresponding distribution function, we consider the mixture distributions

$$\tilde{F}_{097}^{sm}(\gamma) = (1-\gamma)\tilde{F}_{m,n}^0 + \gamma\tilde{F}_{097}^{sm}, \quad \text{and} \quad \tilde{F}_{100}^{sm}(\gamma) = (1-\gamma)\tilde{F}_{m,n}^0 + \gamma\tilde{F}_{100}^{sm}, \ (4.11)$$

where $\gamma \in [0, 1]$. Note that, similar to case (a) in Section 4.3, here also γ quantifies the departure of the configuration $(\tilde{F}_{097}^{sm}(\gamma), \tilde{F}_{100}^{sm}(\gamma))$ from the null of equality

of distributions. Also, the distance between $\tilde{F}_{097}^{sm}(\gamma)$ and $\tilde{F}_{100}^{sm}(\gamma)$ increases as γ approaches 1. We denote the densities of $\tilde{F}_{097}^{sm}(\gamma)$ and $\tilde{F}_{100}^{sm}(\gamma)$ by $\tilde{f}_{097}^{sm}(\gamma)$ and $\tilde{f}_{100}^{sm}(\gamma)$, respectively.

Now observe that when $\gamma \in (0,1)$, the mixture densities may not be log-concave or even unimodal. Hence, we compute the log-concave projections (Dümbgen *et al.*, 2011) of $\tilde{f}_{097}^{sm}(\gamma)$ and $\tilde{f}_{100}^{sm}(\gamma)$, respectively. Log-concave projection of a density f is the log-concave density closest to f in Kullback-Leibler (KL) distance. Since the log-concave projection of any arbritrary density is not directly computable, we adopt a two step approach to approximate the log-concave projections. In the first step, we simulate 1000 observations from each of $\tilde{f}_{097}^{sm}(\gamma)$ and $\tilde{f}_{100}^{sm}(\gamma)$. In the second step, we calculate the smoothed log-concave MLE density estimators of Chen and Samworth based on the simulated samples in the last step. The resulting densities are the approximate smoothed log-concave projections of $\tilde{f}_{097}^{sm}(\gamma)$ and $\tilde{f}_{100}^{sm}(\gamma)$. Finally, we generate two samples of size 68 and 180 from the projected densities using the methods in Dümbgen and Rufibach (2010) and the R package *logcondens*, and replicate this process 10,000 times.

Figure 9 entails that all the tests exhibit decent power for higher values of γ. The LC minimum t-test exhibits the highest power, which is unsurprising since the underlying data is generated from log-concave densities. Also, since $\tilde{F}_{097}^{sm}(\gamma) = \tilde{F}_{100}^{sm}(\gamma)$ at $\gamma = 0$, the power curves resemble that of case (a) in our simulation schemes.

5. Measures of discrepancy

This section describes an approach for quantifying the difference between f and g via estimates of the squared Hellinger distance $\mathcal{D}^2(f,g)$, which provides complementary insights to the tests of stochastic dominance presented in Section 4. The Hellinger distance is an example of f-divergence. These divergences are widely used to measure the similarity or dissimilarity between two probability measures. Compared to other commonly used f-divergences such as the KL divergence or the χ^2 divergence (Nielsen and Nock, 2014), the Hellinger distance is appealing due to its symmetry in its arguments, which is a desirable property for a measure of discrepancy. Another crucial advantage of the Hellinger distance is that its value is finite for every pair of densities (Gibbs and Su, 2002). In contrast, the KL and χ^2 divergences can both be infinite if the densities under consideration do not share the same support. It is not always reasonable to assume that the underlying densities of responses collected from different vaccine trials will have the same support. Therefore, the Hellinger distance appeals to us more than the KL divergence or the χ^2 divergence. The Hellinger distance has also seen successful application in various disciplines ranging from machine learning (Cieslak and Chawla, 2009, González-Castro *et al.*, 2013, 2010), to ecology (Rao, 1995), to fraud detection (Yamanishi *et al.*, 2004). Finally, we choose to work with the squared Hellinger distance instead of the Hellinger distance because due to its simpler form, the squared version easily lends itself to efficient

estimation procedures. Regardless, a 95% confidence interval for the Hellinger distance can always be constructed using that of its squared version.

We estimate the squared Hellinger distance between f and g using the same density estimators involved in the construction of tests of stochastic dominance, that is, the log-concave MLE (Dümbgen and Rufibach, 2009) and its smooth version (Chen and Samworth, 2013), or Birgé's estimator. Recalling the definition of the density estimators $\widehat{f}_m$, $\widehat{g}_n$, $\tilde{f}_m$, $\tilde{g}_n$, $\tilde{f}_m^{sm}$, and $\tilde{g}_n^{sm}$ from Section 4.1, we propose the plug-in estimators $\mathcal{D}^2(\widehat{f}_m, \widehat{g}_n)$, $\mathcal{D}^2(\tilde{f}_m, \tilde{g}_n)$, and $\mathcal{D}^2(\tilde{f}_m^{sm}, \tilde{g}_n^{sm})$ for the purpose of estimating $\mathcal{D}^2(f, g)$. We refer to the resulting estimators as the "unimodal", "log-concave", and the "smoothed log-concave" estimator, respectively.

5.1. Asymptotic properties of the unimodal estimator

We will show that, under some regularity conditions, $\mathcal{D}^2(\widehat{f}_m, \widehat{g}_n)$ is a $\sqrt{N}$-consistent estimator of $\mathcal{D}^2(f, g)$ with asymptotic variance

$$\sigma_{f,g}^2 = \frac{2\mathcal{D}^2(f,g) - \mathcal{D}^4(f,g)}{4\lambda(1-\lambda)}. \tag{5.1}$$

Letting $b, B > 0$, we denote by $\mathcal{P}(b, B)$ the class of densities that are bounded below and above by b and B on their support, that is,

$$\mathcal{P}(b, B) = \left\{ f \in \mathcal{P} \ : \ b \leq f(x) \leq B, \ \text{ for } x \in \text{supp}(f) \right\}. \tag{5.2}$$

We will assume that $f, g \in \mathcal{P}(b, B)$. Simulations suggest that this condition may not be necessary. However, this condition is required for technical reasons in our proof. Such technical condition is quite common in the literature, and has appeared in the analysis of plug-in estimators (Kandasamy *et al.*, 2015) and functionals of Grenander estimators (Mukherjee *et al.*, 2019).

Theorem 3. *Suppose f and g are unimodal densities in $\mathcal{P}(b, B)$, where b, $B > 0$. Further suppose that f and g satisfy condition A and $m/N \to \lambda$. Then,*

$$\sqrt{N}[\mathcal{D}^2(\widehat{f}_m, \widehat{g}_n) - \mathcal{D}^2(f,g)] \to_d N(0, \sigma_{f,g}^2), \tag{5.3}$$

where $\sigma_{f,g}^2$ is as defined in (5.1).

Note that, since $\sigma_{f,g}^2$ can be consistently estimated plugging in the estimator $\mathcal{D}^2(\widehat{f}_m, \widehat{g}_n)$, a Wald type confidence interval is readily available for $\mathcal{D}^2(f, g)$. Also, the asymptotic variance $\sigma_{f,g}^2$ equals the lower bound of the asymptotic variance on a regular estimator of the squared Hellinger distance under the nonparametric model (Kandasamy *et al.*, 2015). See Van der Vaart (1998), Birgé and Massart (1995) for more detail on the lower bound and related theory.

We now give a high-level explanation of the idea behind Theorem 3. It can be shown that the squared Hellinger distance is a smooth functional of the

underlying distribution functions in some suitable sense. Specifically, we will show that it allows a first order Von Mises expansion (Fernholz, 2012), which has the same essence as the Taylor series expansion (Kandasamy *et al.*, 2015). On the other hand, for $\widehat{F}_m^0$, the unimodal MLE of F based on the true mode, we show that $\sqrt{m}(\widehat{F}_m^0 - F)$ converges weakly to a Brownian process almost surely under Condition A. It can then be shown via a delta-method type argument, applied on the squared Hellinger distance functional, that (5.3) holds for the unimodal MLEs of F and G. The final step in proving Theorem 3 is showing that the squared Hellinger distance between Birgé's estimator and the MLE is small.

5.2. Asymptotic properties of the log-concave estimators

Recall that our log-concave estimators of the squared Hellinger distance are given by $\mathcal{D}^2(\tilde{f}_m, \tilde{g}_n)$ and $\mathcal{D}^2(\tilde{f}_m^{sm}, \tilde{g}_n^{sm})$, where we remind the reader that $\tilde{f}_m, \tilde{g}_n$ are the log-concave MLEs and $\tilde{f}_m^{sm}$, $\tilde{g}_n^{sm}$ are the smoothed log-concave MLEs. Lemma 6 implies that these estimators are strongly consistent for $\mathcal{D}^2(f, g)$ provided the shape constraint holds.

Lemma 6. *Suppose that the densities f and g are log-concave and continuous. Then, as $m, n \to \infty$, $\mathcal{D}^2(\tilde{f}_m, \tilde{g}_n) \to_{a.s.} \mathcal{D}^2(f, g)$ and $\mathcal{D}^2(\tilde{f}_m^{sm}, \tilde{g}_n^{sm}) \to_{a.s.} \mathcal{D}^2(f, g)$.*

Our simulations indicate that both the log-concave and the smoothed log-concave plug-in estimators are $\sqrt{N}$-consistent with asymptotic variance $\sigma_{f,g}^2$ when f and g are continuous log-concave densities. Therefore, in our empirical study, we include Wald type confidence intervals based on these log-concave plug-in estimators as well. Our simulations in Section 5.3 indicate that under the violation of the continuity assumption, the log-concave plug-in estimator may still remain $\sqrt{N}$-consistent, but the $\sqrt{N}$-consistency of the the smoothed log-concave plug-in estimator may fail to hold.

It may be possible to analyze the $\sqrt{N}$-consistency of the log-concave estimators working along the lines of Kulikov and Lopuhaä (2006), Groeneboom (1984, 1989), which pertain to the shape restriction of monotonicity. However, we leave this investigation for future research because a detailed treatment of the $\sqrt{N}$-consistency of the log-concave estimators is out of scope of the present paper. We remark in passing that proving the $\sqrt{N}$-consistency of the log-concave plug-in estimator may be easier for some special cases, e.g., when the logarithm of f and g are linear or piece-wise affine, by using the results of Kim *et al.* (2018). We do not pursue this direction because it is unlikely that, for our data, f and g belong to such restricted classes.

Remark 5. The case of model misspecification is of natural interest in the study of shape-constrained estimators. Suppose f and g are not log-concave, but they are bounded continuous densities with finite first moments. Then it follows that there exist unique log-concave densities f^* and g^*, which are almost sure limits of $\tilde{f}_m$ and $\tilde{g}_n$ in both uniform and L_1 metric (Theorem 4 Cule and

Samworth, 2010). Here f^* and g^* are also the log-concave projections of f and g, respectively, in the sense of Dümbgen *et al.* (2011). Using the same arguments as in the proof of Lemma 6, it can be shown that the log-concave estimator of the squared Hellinger distance converges almost surely to $\mathcal{D}^2(f^*, g^*)$. If f and g additionally have finite second moments, a similar phenomenon takes place for the smoothed log-concave MLEs as well. In this case, however, $\tilde{f}_m^{sm}$ and $\tilde{g}_n^{sm}$ converge uniformly (and also in L_1) to different limits f^{**} and g^{**}, which can be interpreted as the respective smoothed versions of the log-concave projections f^* and g^* (cf. Theorem 1, Chen and Samworth, 2013). In this case also, the smoothed log-concave estimator of the squared Hellinger distance converges almost surely to $\mathcal{D}^2(f^{**}, g^{**})$. In summary, if the log-concavity assumption is violated, the log-concave plug-in estimators converge to a different limit, whose distance from $\mathcal{D}^2(f, g)$ depends on the departure of f and g from log-concavity.

KDE based plug-in estimators: The natural non-parametric comparators of the shape constrained plug-in estimators are the KDE based plug-in estimators. Though the latter is simple to implement, it can have a bias of order $||\widehat{f}_{m,k} - f||_2 + ||\widehat{g}_{n,k} - g||_2$ (cf. Section 2 of Robins *et al.*, 2009), where $\widehat{f}_{m,k}$ and $\widehat{g}_{n,k}$ are the KDEs of f and g, respectively. The above bias decreases to zero at a rate slower than $N^{-1/2}$ (Stupfler, 2014), thereby leading to a suboptimal performance. See Section 5 of Kandasamy *et al.* (2015) for more discussion on the disadvantages of the KDE based naïve plug-in estimators.

One can improve the naïve plug-in estimator, however, using a one-step Newton-Raphson procedure (cf. Van der Vaart, 1998, Pfanzagl and Wefelmeyer, 1985), which leads to a bias corrected plug-in estimator. In context of the Hellinger distance, the bias-corrected estimator $(\mathcal{D}^2)^*(\widehat{f}_{m,k}, \widehat{g}_{n,k})$ takes the form (Kandasamy *et al.*, 2015)

$$\mathcal{D}^2(\widehat{f}_{m,k}, \widehat{g}_{n,k}) + \int_{\mathbb{R}} \psi_f(x; \widehat{f}_{m,k}, \widehat{g}_{n,k}) d\mathbb{F}_m(x) + \int_{\mathbb{R}} \psi_g(y; \widehat{f}_{m,k}, \widehat{g}_{n,k}) d\mathbb{G}_n(y)$$
$$= 1 - \frac{1}{2} \left(\int_{\mathbb{R}} \sqrt{\widehat{g}_{n,k}(x)/\widehat{f}_{m,k}(x)} d\mathbb{F}_m(x) + \int_{\mathbb{R}} \sqrt{\widehat{f}_{m,k}(y)/\widehat{g}_{n,k}(y)} d\mathbb{G}_n(y) \right),$$

where ψ_f and ψ_g are the influence functions corresponding to the functional $(f, g) \mapsto \mathcal{D}^2(f, g)$. We will formally introduce the influence functions in Appendix B. Note that learning the form of the bias-corrected estimator thus requires the explicit computation of the influence functions ψ_f and ψ_g. From a broader prospective, each time one tries to estimate a functional of the underlying distributions using a bias-corrected plug-in estimator, they have to carry out some extra analytical calculations that depend on the functional of interest.

In contrast, Theorem 3 shows that when the shape constraint is satisfied, the unimodal estimator of the squared Hellinger distance does not require any bias correction for $\sqrt{N}$-consistency. Our analysis in Appendix B indicates that this is not an artifact of the Hellinger distance, but rather the result of Birgé's estimator's proximity to the unimodal MLE, i.e. the Grenander estimator based on the true mode. In fact, Theorem 4 in Appendix B shows that, if T is a smooth functional of the distribution functions F, then $T(\widehat{f}_m^0)$ is $\sqrt{m}$-consistent for $T(F)$

TABLE 4

Table summarizing the current results on different plug-in estimators of the squared Hellinger distance. The consistency and $\sqrt{N}$-consistency of the shape constrained estimators are proved under the correct specification of the shape constraints. Empirical studies are provided on all the estimators listed above.

* Indicates whether we included confidence intervals for the corresponding estimator in the empirical study of Section 5.3.

Estimator	Consistency	$\sqrt{N}$-consistency	Showed in	Confidence interval*
Naïve KDE	yes	no	Kandasamy *et al.* (2015)	no
Bias corrected KDE	yes	yes	Kandasamy *et al.* (2015)	yes
Unimodal	yes	yes	current paper	yes
Log-concave	yes	unknown, but suggested by simulations	current paper	yes
Smoothed log-concave	yes	$---''---$	current paper	yes

under mild conditions. Because the total variation distance between $\widehat{f}_m$ and $\widehat{f}_m^0$ is $o_p(m^{-1/2})$ (see our Lemma A.5 in Appendix A), it is then natural to expect that $T(\widehat{f}_m)$ would be $\sqrt{m}$-consistent if T is sufficiently smooth.

Although we do not have any such theoretical evidence for the log-concave estimators, our simulations and the simulations of Cule *et al.* (2010) on plug-in estimators based on log-concave MLEs (see Figure 18 therein) indicate that these estimators do not have any $O_p(N^{-1/2})$ bias term either. This allows users of correctly specified shape-constrained estimators to avoid analytic calculation of the influence functions entirely.

In our upcoming simulation study, we use both $\mathcal{D}^2(\widehat{f}_{m,k}, \widehat{g}_{n,k})$ and $(\mathcal{D}^2)^*(\widehat{f}_{m,k}, \widehat{g}_{n,k})$ as comparators, where the KDEs are based on the Gaussian kernel. We refer to these estimators as the KDE estimator and the bias-corrected KDE estimator, respectively. As in Section 3, the kernel bandwidth is chosen using the univariate least square cross-validation (LSCV) selector of Bowman (1984) and Rudemo (1982). Generally, kernel-based bias corrected estimators satisfy $\sqrt{N}$ consistency results of the type (5.3) (cf. Theorem 6, Kandasamy *et al.*, 2015). Therefore, we use the bias-corrected KDE based confidence intervals to benchmark the performance of our shape constrained estimators. However, we do not report any confidence interval based on the naïve plug-in estimator because our simulations indicate that usually its coverage is much less than the nominal level. For the sake of clarity, in Table 4, we summarize the current state of results on the above-mentioned estimators of the squared Hellinger distance.

5.3. Simulations

To compare the performance of different estimators of the squared Hellinger distance, we consider the following combinations of f and g:
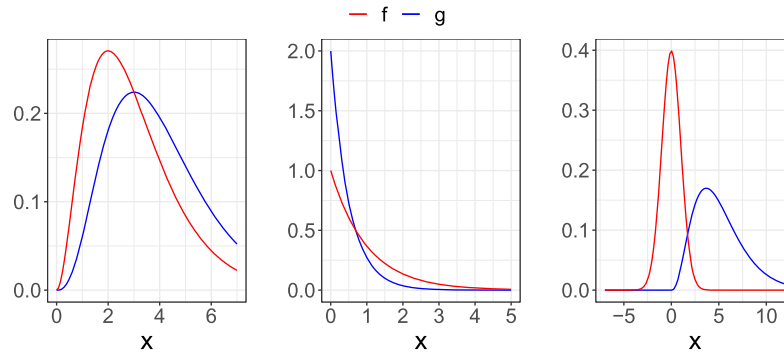
Fig 6: Plots of the densities f and g for the cases (a) (left), (e) (middle), and (f) (right) in the simulation scheme of Section 5.3

(a) $f \sim Gamma(4,1)$ and $g \sim Gamma(3,1)$.

(b) $f \sim N(1,1)$ and $g \sim N(0,1)$, which corresponds to case (a) in Section 4.3 with $\gamma = 1$.

(c) $f = \tilde{f}^{sm}_{097}$ and $g = \tilde{f}^{sm}_{100}$, where $\tilde{f}^{sm}_{097}$ and $\tilde{f}^{sm}_{100}$ are the smoothed log-concave MLEs of f_{097} and f_{100}, respectively (see Figure 2).

(d) $f \sim \text{Exp}(1)$ and $g \sim \text{Exp}(2)$ where $\text{Exp}(t)$ is the exponential distribution with rate t.

(e) (f,g) corresponds to case (e) in Section 4.3 with $\gamma = 1$.

(f) $f \sim N(0,1)$, and $g \sim Gamma(3.61, 1.41)$.

The plots of the above schemes can be found in Figure 5 and Figure 6. In cases (a), (b), (c), (d), and (f), both shape constraints are satisfied, where in case (e), both shape constraints are violated. We will refer to case (a), (b), (c), (d), and (f), therefore, as the "correctly specified" cases and case (e), as the "misspecified" case. In case (d), the logarithm of the densities are linear on their respective supports. This is also the only case where the densities are discontinuous (see Figure 6; discontinuity at zero). In case (f), g was chosen so as to resemble the density estimate of f_{100} (see Figure 2 and Figure 6). This is the only correctly specified case where the densities are known to come from different families, and the densities also have very different shapes.

We generate two samples of same size from each simulation setting. We vary the common sample size n from 50 to 500 in increments of 50. We do not consider larger values of n because in our motivating phase 1b and phase 2 vaccine trial applications, the sample sizes are generally no larger than 500. We consider 10,000 Monte Carlo replications for each sample size.

Figures 10 and 11 plot $\sqrt{n}$ times the absolute bias, and Figure 12 plots n times the mean squared error (MSE). Figure 13i and Figure 13ii display, respectively, the coverage and the average length of the confidence intervals, both of which are estimated using 10,000 Monte Carlo samples. In the misspecified case (e), the performance of the shape-constrained estimators deteriorate sharply with n, which is unsurprising due to the violation of the shape constraints in this case.
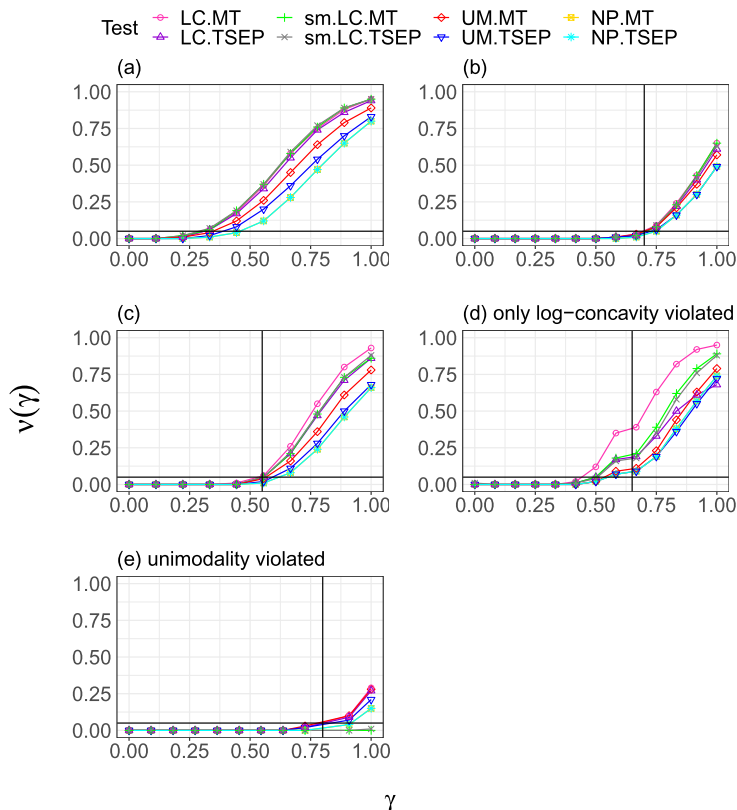
Fig 7: Plot of estimated power $\nu(\gamma)$ vs γ: the labels (a)-(e) correspond to simulation schemes (a)-(e). Here MT and TSEP correspond to the minimum t-test and the TSEP test, respectively. The standard deviation of the $\nu(\gamma)$ estimate in each case is less than 0.005. The black horizontal line corresponds to the level of the test, $\alpha = 0.05$. For cases (b)-(e), the black vertical line represents the LFC configuration γ^*, taking value 0.70 (b), 0.55 (c), 0.65 (d), and 0.80 (e).

In case (d), where the densities are exponential, the scaled bias of the smoothed log-concave plug-in estimator increases with n, which implies $\sqrt{N}$-consistency does not hold for this estimator in this case. Closer inspection reveals that although it is not $\sqrt{N}$-consistent, the smoothed log-concave plug-in estimator is still consistent in case (d). We found out that the smoothed log-concave density estimator fails to approximate the exponential densities near zero, their point of discontinuity (see Figure 19 in Appendix D). It is worth mentioning that although the curvature Condition B2 is violated in case (d), it does not affect the performance of the log-concave plug-in estimator.

For all other cases, the smoothed log-concave estimator exhibits the best performance among the shape-constrained estimators. In all cases, the unimodal
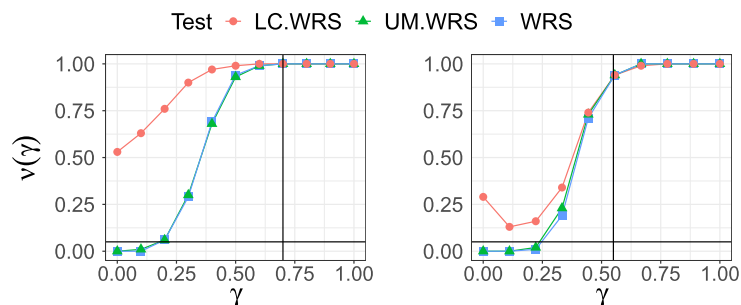
Fig 8: Estimated power curves of WRS type tests for case (b) (left) and case (c) (right). Here LC.WRS and UM.WRS correspond to the tests based on $T_{m,n}^{\mathrm{wrs}}(\tilde{F}_m, \tilde{G}_n)$ and $T_{m,n}^{\mathrm{wrs}}(\widehat{F}_m, \widehat{G}_n)$, respectively. The standard errors of the estimated powers are less than 0.005. The black horizontal line corresponds to the level of the test, $\alpha = 0.05$. The black vertical lines represent the LFC configuration γ^*, which are 0.70 (b) and 0.55 (c).
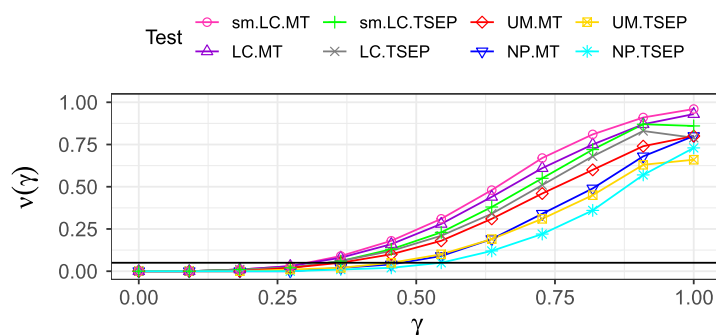
Fig 9: Estimated power curve applied on our datasets: Here MT and TSEP correspond to the minimum t-test and the TSEP test, respectively. The black horizontal line corresponds to the level of the test, which is 0.05. The standard deviation of the estimated power does not exceed 0.005 in any case.

estimator underperforms. Importantly, Figure 13i indicates that the unimodal estimator would require sample size larger than 500 for the Wald type confidence interval to be valid, where for the other confidence intervals, this sample size is sufficient for the asymptotics to kick in.

The KDE-based plug-in estimator experiences an increase in the bias and the MSE with n, which agrees with our previous discussion on KDE based plug-in estimators. Its bias corrected version performs comparably to the shape-constrained estimators in all cases except case (f). In case (f), however, the bias corrected estimator yields a confidence interval with poor coverage, which is probably due to the large bias incurred by the original KDE-based plug-in estimator in this case (see panel (f) of Figure 11). Case (f) is a case where the
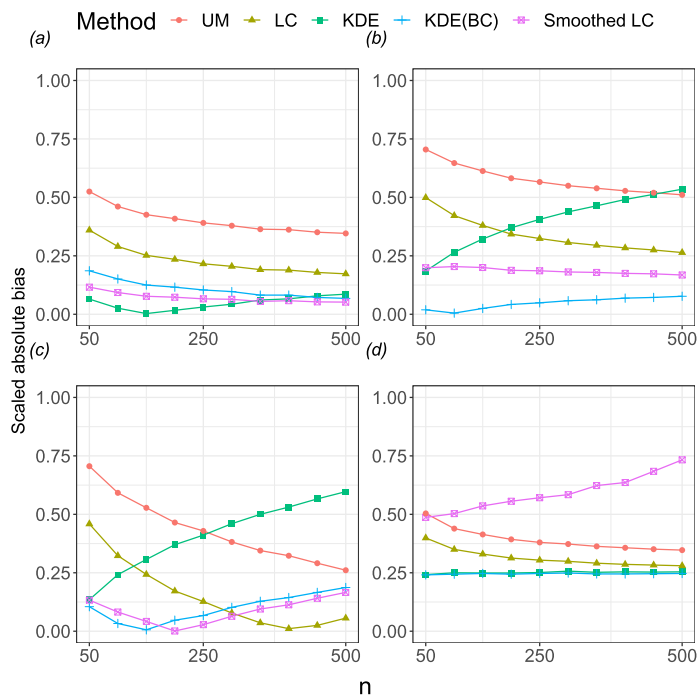
Fig 10: Plots of the absolute values of the bias (scaled by $\sqrt{n}$) for cases (a)–(d). Here "KDE (BC)" and "smoothed LC" stand for the bias corrected KDE estimator and the smoothed log-concave estimator, respectively. For each estimator, either the standard error is less than 0.006 or the relative standard error is less than 2%.

underlying densities differ by both shape and scale. We suspect that the cross-validated bandwidth for the KDE estimator does not work in this case even with bias correction. Finally, the average length of the confidence intervals do not vary noticeably across different methods.

In summary, the log-concave plug-in estimator exhibits reliable performance when the shape restriction holds. The smoothed log-concave plug-in estimator may also require the underlying densities to be continuous, but otherwise it performs comparably with the bias-corrected KDE-based plug-in estimator. The latter performs well in many settings, but it is not always reliable. The lacking performance of the KDE based methods in some settings is probably due to the variable nature of the optimal bandwidth under different settings.

5.4. Application to HVTN 097 and HVTN 100 data

Table 5 tabulates the point estimates of the squared Hellinger distance (between f_{100} and f_{097}) and the corresponding 95% intervals. Table 5 also displays the
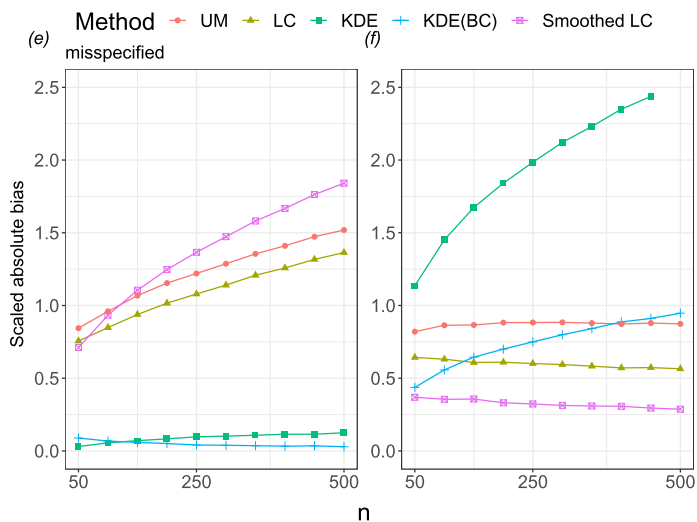
Fig 11: Plots of the bias (scaled by n) for cases (e) and (f). Here "KDE (BC)" and "smoothed LC" stand for the bias corrected KDE estimator and the smoothed log-concave estimator, respectively. For each estimator, either the standard error is less than 0.006 or the relative standard error is less than 2%.

95% confidence intervals of the Hellinger distance, which are obtained by taking square root of the upper and lower bounds of the previous confidence intervals.

TABLE 5
Table of the point estimates of $\mathcal{D}^2(f_{097}, f_{100})$, and 95% confidence intervals of $\mathcal{D}^2(f_{097}, f_{100})$ and $\mathcal{D}(f_{097}, f_{100})$.

Estimator	Naïve KDE	Bias -corrected KDE	UM	LC	Smoothed LC
Point estimate of $\mathcal{D}^2(f_{097}, f_{100})$	0.16	0.19	0.18	0.15	0.21
95% CI for $\mathcal{D}^2(f_{097}, f_{100})$	not available	$(0.110, 0.274)$	$(0.128, 0.300)$	$(0.098, 0.256)$	$(0.079, 0.228)$
95% CI for $\mathcal{D}(f_{097}, f_{100})$	$--''--$	$(0.332, 0.523)$	$(0.358, 0.548)$	$(0.313, 0.506)$	$(0.281, 0.447)$

To give the reader some perspective, if f_0 is a $N(0, 1)$ distribution and f_μ is a $N(\mu, 1)$ distribution, then, when μ is equal to 1.00, 1.25, 1.50, 1.75, and 2.00, $\mathcal{D}(f_0, f_\mu)$ is equal to 0.346, 0.424, 0.500, 0.566, and 0.624, respectively. Also, in Figure 14, we display some density pairs (f, g) satisfying $F \succeq G$ with Hellinger distance in the range $0.35 - 0.60$, which is similar to our data.

(i) Plots of the MSE (scaled by n) for cases (a)–(c).

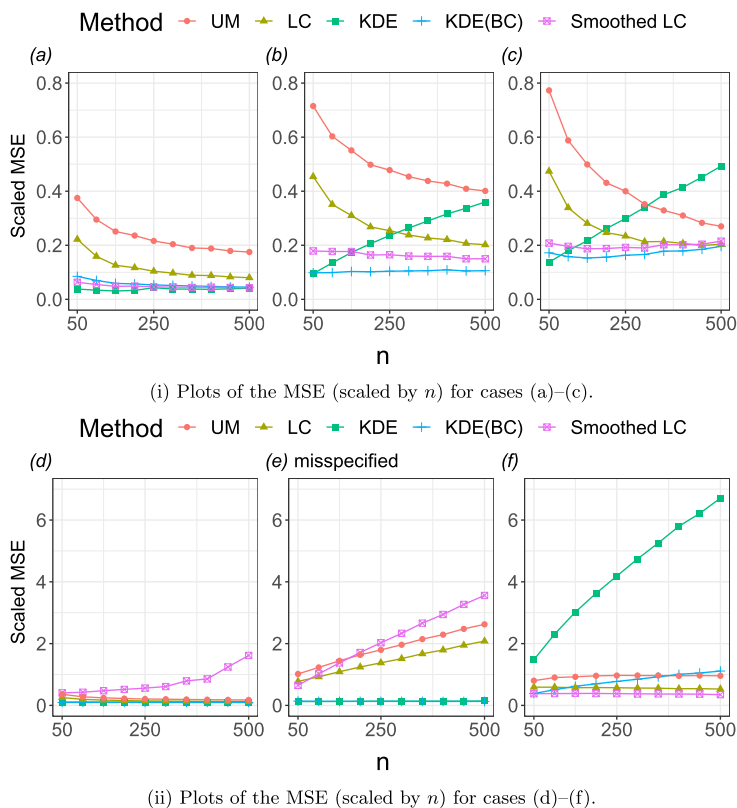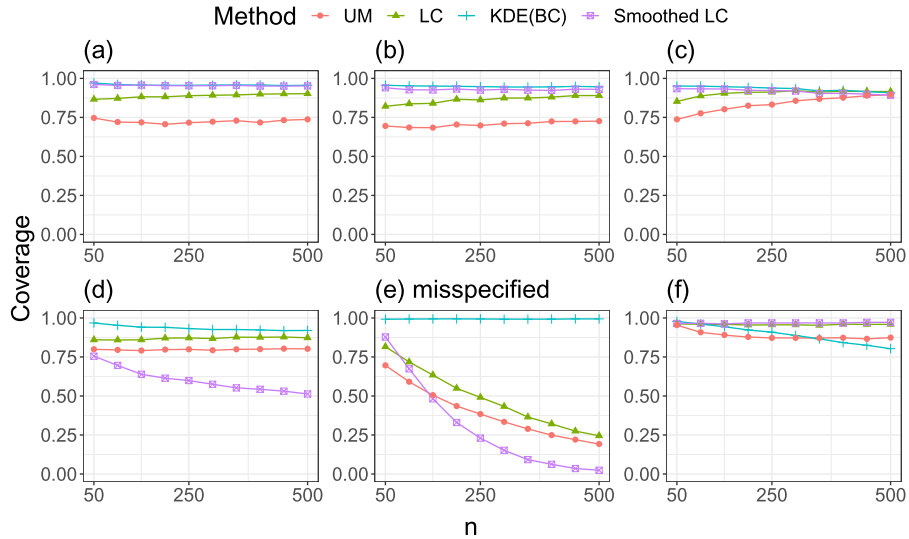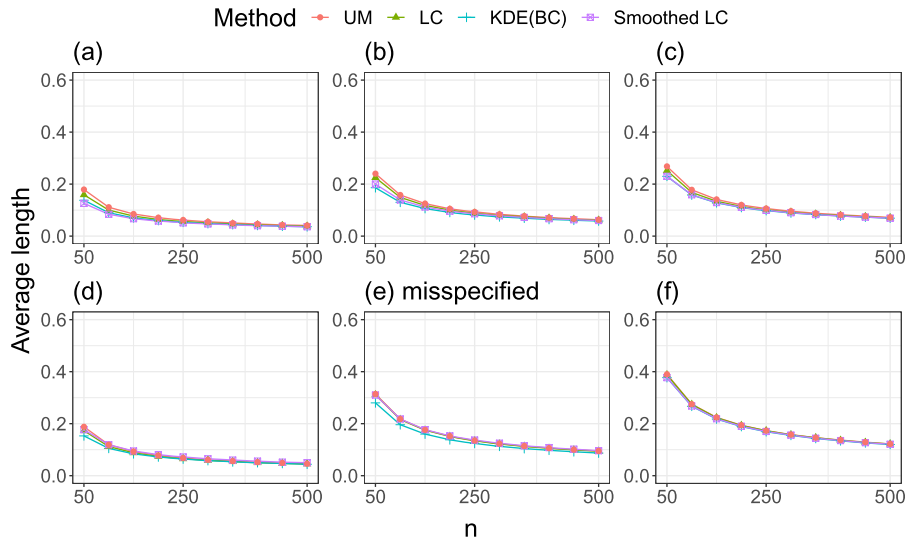(ii) Plots of the MSE (scaled by n) for cases (d)–(f).

Fig 12: Plot of MSE: Here "KDE (BC)" and "smoothed LC" stand for the bias corrected KDE estimator and the smoothed log-concave estimator, respectively. For each estimator, either the standard error is less than 0.006 or the relative standard error is less than 2%.

6. Discussion

The first contribution of our work is a novel analysis of the data from the HVTN 097 and HVTN 100 trials. All of our tests reject the null of non-dominance in favor of the strict stochastic dominance of F_{097} over F_{100}. To provide further insight into the discrepancy between the two IgG binding response distributions, we estimated the squared Hellinger distance between the corresponding densities, which turns out to be approximately 0.20 (95% CI 0.10-0.30). We remark that our findings are consistent with those of Bekker *et al.* (2018), who found that the average magnitude of IgG binding to V1V2 antigens observed in the HVTN 100 trial is lower than that in the RV 144 trial. Although the latter used the same regimen as HVTN 097, it was conducted in a different population (Thailand). Our findings indicate that the difference in the magnitude of IgG binding response between HVTN 100 and RV 144 regimen may be attributable

(i) Coverage probability of the 95% confidence intervals for cases (a)–(f).

(ii) Average length of the 95% confidence intervals for cases (a)–(f).

Fig 13: Here "KDE (BC)" and "smoothed LC" stand for the bias corrected KDE estimator and the smoothed log-concave estimator, respectively. In all the cases, the standard error is less than 0.005, and hence, not plotted.

to the HIV clade difference rather than to the difference in populations.

The outcome of our tests become meaningful when viewed against the lack of efficacy observed in the phase 2b/3 HVTN 702 trial, which evaluated the safety
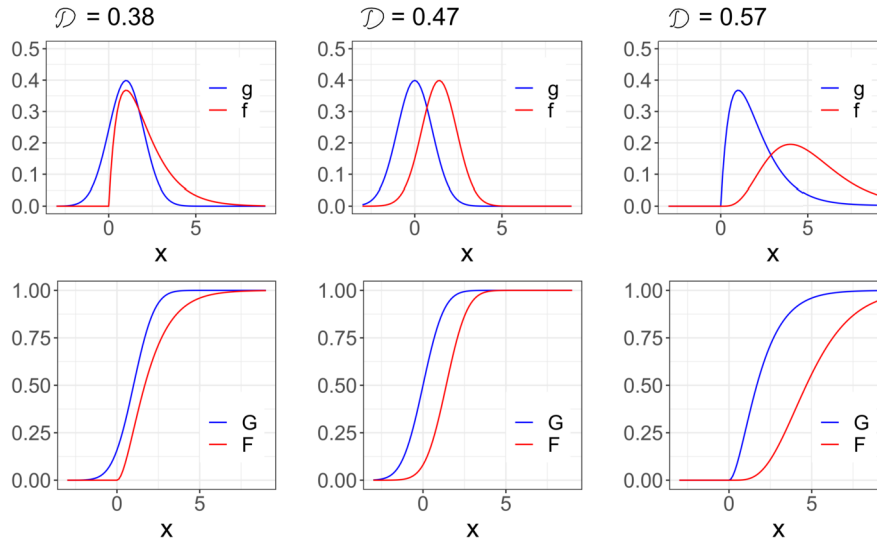
Fig 14: Plot of different density pairs and their distribution functions. The Hellinger distance ($\mathcal{D}$) is given on top of each pair. Left: $f \sim Gamma(25,1)$, $g \sim N(0.5,1)$; middle: $f \sim N(1.40,1)$, $g \sim N(0,1)$; right: $f \sim Gamma(5,1)$, $g \sim Gamma(2,1)$.

and efficacy of the HVTN 100 regimen in South Africa. A possible hypothesis for why no efficacy was observed when the HVTN 100 regimen was evaluated in this trial, whereas efficacy was observed when the HVTN 097 regimen was evaluated in the RV144 trial, is that the HVTN 100 regimen leads to a lower magnitude of IgG binding to the V1V2 region. This possibility is supported by the observation made by Haynes *et al.* (2012) regarding the negative correlation between rate of infection and the magnitude of IgG binding to V1V2 region. This hypothesis can be tested when the immune profile of the participants in HVTN 702 trial becomes available.

Another contribution of our work relates to density estimation in the context of vaccine trials. Based on a cross-validated analysis of the HVTN 097 and HVTN 100 data, we believe that the log-concave density estimators of Dümbgen and Rufibach (2009) and Chen and Samworth (2013) may yield improved density estimation in vaccine studies. In future work, it would be worth further validating this claim on other vaccine trial datasets.

We also made several methodological contributions. In Section 4, we introduce three novel shape-constrained tests. These tests have the desirable asymptotic properties of nonparametric tests, and simulations illustrate that the shape-constrained tests have better overall performance than the nonparametric tests. Moreover, even under the violation of the shape-constraints, their performance is not much worse than the nonparametric tests. We also introduce shape-constrained plug-in estimators of the squared Hellinger distance and provide

asymptotic consistency and distributional results. Our simulations suggest that, when the shape constraint is satisfied, the log-concave plug-in estimators exhibit overall lower MSE and absolute bias than the KDE based plug-in estimator. In fact, they perform comparably with the bias-corrected KDE based estimator. However, unlike the bias-corrected KDE plug-in estimator, the log-concave plug-in estimators require neither selecting a tuning parameter nor carrying out the analytic calculations needed to derive the bias-correction term. Therefore, the log-concave plug-in estimators may be preferred in settings where this shape constraint is plausible.

Appendix A: Proofs for Section 4

Before proceeding any further, we introduce some new notations. We let l^∞ denote the collection of all bounded functions on $\mathbb{R}$, which we equip with the uniform metric $\|\cdot\|_\infty$. For any function $\mu : \mathbb{R} \mapsto \mathbb{R}$, we define the norm $\|\cdot\|_a^b$ by $\|\mu\|_a^b = \sup_{x \in [a,b]} |\mu(x)|$. Also, we denote the boundary of a set A by $\mathrm{bd}(A)$.

Suppose $(\Omega, \mathrm{int}(\mathcal{H}_0), \mathcal{P})$ is the common probability space corresponding to the X_i's and Y_j's. For the rest of this section, we let $\to_p$ and $\to_{a.s.}$ correspond to this probability space. Since F and G are continuous, using the construction of Section 1.1 of Shorack (1984) (see also p.93 of Shorack and Wellner, 2009), we can show that there exist two independent Brownian bridges $\mathbb{V}_1$ and $\mathbb{V}_2$ on $(\Omega, \mathrm{int}(\mathcal{H}_0), \mathcal{P})$ such that

$$\left\| \sqrt{m}(\mathbb{F}_m - F) - \mathbb{V}_1 \circ F \right\|_\infty \to_{a.s.} 0 \quad \text{as } m \to \infty, \tag{A.1}$$

and

$$\left\| \sqrt{n}(\mathbb{G}_n - G) - \mathbb{V}_2 \circ G \right\|_\infty \to_{a.s.} 0 \quad \text{as } n \to \infty. \tag{A.2}$$

Let us denote

$$\mathbb{U} = \lambda^{1/2} \mathbb{V}_2 - (1 - \lambda)^{1/2} \mathbb{V}_1, \tag{A.3}$$

where λ is so that $m/N \to \lambda$. Note that $\mathbb{U}$ is also distributed as a Brownian bridge. We will show that the asymptotic distributions of our test statistics depend on $\mathbb{U}$. Also, for ours F and G,

$$\|\mathbb{F}_m - F\|_\infty \to_{a.s.} 0, \quad \|\mathbb{G}_n - G\|_\infty \to_{a.s.} 0, \quad \|\mathbb{H}_N - H\|_\infty \to_{a.s.} 0.$$

In the sequel, we will also use the following fact on the convergence of the quantiles of $\mathbb{H}_N$ and $D_{p,m,n}$, often without mentioning. As a corollary to Fact 3.1A, it follows that $\mathrm{dist}(D_{p,m,n}, D_p) \to_{a.s.} 0$ under Condition N.

Fact 3.1. *Suppose Condition N holds and U is the open set such that $D_p \subset U \subset supp(f) \cup supp(g)$. Further suppose $p' < p$ such that $D_{p'} \subset U$. Then*

- *A. For any $t \in [p', 1 - p']$, $\mathbb{H}_N^{-1}(t) \to_{a.s.} H^{-1}(t)$.*
- *B. $D_{p,m,n} \subset D_{p'}$ for all sufficiently large m and n with probability one.*
- *C. For any $q > p$, $D_q \subset D_{p,m,n}$ for all sufficiently large m and n with probability one.*

Proof of Fact 3.1. Since $h = \lambda f + (1 - \lambda)g > 0$ on U, it follows that H^{-1} is continuous on U (cf. Proposition A.7, pp. 98 of Bobkov and Ledoux, 2016). Also, under Condition N, $\|\mathbb{H}_N - H\|_\infty \to_{a.s.} 0$. Since $D_{p'} \subset U$, $H^{-1}(t) \in U$ for any $t \in [p', 1 - p']$. Hence, $\mathbb{H}_N^{-1}(t) \to_{a.s.} H^{-1}(t)$ by Lemma A5, pp. 96 of Bobkov and Ledoux (2016). Hence part (A) of Fact 3.1 is proved.

Part (A) of Fact 3.1 implies $\mathbb{H}_N^{-1}(p) \to_{a.s.} H^{-1}(p)$ and $\mathbb{H}_N^{-1}(1 - p) \to_{a.s.} H^{-1}(1 - p)$. To prove part B and C of Fact 3.1, therefore, it suffices to show that $H^{-1}(p') < H^{-1}(p)$, $H^{-1}(1 - p') > H^{-1}(1 - p)$, $H^{-1}(p) < H^{-1}(q)$, and $H^{-1}(1 - q) < H^{-1}(1 - p)$. Since $q < p < p'$ and $D_q \subset D_p \subset D_{p'} \subset U$, it is enough to show that H^{-1} is continuous and strictly increasing on U. The latter holds if H s continuous and strictly increasing on U. The proof now follows since we already showed that H has a positive density on U. $\qquad\square$

A.0.1. Proof of Lemma 1

To prove Lemma 1, we will use an alternative definition of $\mathcal{H}_0$ and $\mathcal{H}_1$. Note that continuous distribution functions $(F, G) \in \mathcal{H}_0$ if and only if $\sup_{x \in D_p}[F(x) - G(x)] \geq 0$. Because D_p is compact, such a pair is in $\mathcal{H}_1$ if and only if $\sup_{x \in D_p}[F(x) - G(x)] < 0$. We will prove the current lemma in some steps.

Step 1: closure of $\mathcal{H}_0$ Suppose $(F_n, G_n) \in \mathcal{H}_0$ converges to $(F, G) \in \mathcal{F} \times \mathcal{F}$ with respect to the metric d_2. Then

$$
\sup_{x \in D_p} (F(x) - G(x)) \geq - \sup_{x \in D_p} (F_n(x) - F(x))
$$
$$
+ \sup_{x \in D_p} (F_n(x) - G_n(x)) - \sup_{x \in D_p} (G(x) - G_n(x)),
$$

which is bounded below by $-\|F_n - F\|_\infty - \|G_n - G\|_\infty$. Taking $n \to \infty$ yields $\sup_{x \in D_p}(F(x) - G(x)) \geq 0$. Therefore,

$$
\mathrm{cl}(\mathcal{H}_0) \subset \left\{ (F, G) \in \mathcal{F} \times \mathcal{F} : \sup_{x \in D_p} (F(x) - G(x)) \geq 0 \right\} = \mathcal{H}_0.
$$

Hence, $\mathrm{cl}(\mathcal{H}_0) = \mathcal{H}_0$, i.e. $\mathcal{H}_0$ is closed in $\mathcal{F} \times \mathcal{F}$.

Step 2: interior and boundary of $\mathcal{H}_0$

If $(F, G) \in \mathcal{H}_0$, either $\sup_{x \in D_p}(F(x) - G(x)) > 0$ or $\sup_{x \in D_p}(F(x) - G(x)) = 0$. Let us denote

$$
\mathcal{A} = \{(F, G) \in \mathcal{F} \times \mathcal{F} : \sup_{x \in D_p} (F(x) - G(x)) = 0\}.
$$

By Lemma A.1, $\mathcal{A} \subset \mathrm{bd}(\mathcal{H}_0)$ because this set is not a part of the interior. Hence,

$$
\mathrm{int}(\mathcal{H}_0) \subset \{(F, G) \in \mathcal{F} \times \mathcal{F} : \sup_{x \in D_p} (F(x) - G(x)) > 0\}.
$$

Therefore, to prove

$$
\mathrm{int}(\mathcal{H}_0) = \{(F, G) \in \mathcal{F} \times \mathcal{F} : \sup_{x \in D_p} (F(x) - G(x)) > 0\}, \qquad (\mathrm{A.4})
$$

it suffices to prove

$$\{(F, G) \in \mathcal{F} \times \mathcal{F} : \sup_{x \in D_p} (F(x) - G(x)) > 0\} \subset \text{int}(\mathcal{H}_0). \qquad (A.5)$$

Suppose $(F, G) \in \mathcal{H}_0$ satisfies $\sup_{x \in D_p}(F(x) - G(x)) > \epsilon$ for some $\epsilon > 0$. Consider any $(\tilde{F}, \tilde{G})$ such that $d_2((\tilde{F}, \tilde{G}), (F, G)) < \epsilon/3$, which means $\|\tilde{F} - F\|_\infty < \epsilon/3$ and $\|\tilde{G} - G\|_\infty < \epsilon/3$. Since $F - G$ is continuous and D_p is compact, the supremum of $F - G$ over D_p is attained at some $z \in D_p$. We have

$$\tilde{F}(z) - \tilde{G}(z) \geq F(z) - G(z) - \|\tilde{F} - F\|_\infty - \|\tilde{G} - G\|_\infty,$$

which is greater than $\epsilon/3$. Thus $\sup_{x \in D_p}(\tilde{F}(x) - \tilde{G}(x)) > 0$ which implies $(F, G) \in \text{int}(\mathcal{H}_0)$. Therefore, we have established (A.4). Because $\text{cl}(\mathcal{H}_0) = \mathcal{H}_0$, (A.4) also implies $\text{bd}(\mathcal{H}_0) = \mathcal{A}$.

Step 3: boundary of $\mathcal{H}_1$

Our first step is to show

$$\text{cl}(\mathcal{H}_1) \subset \left\{(F, G) \in \mathcal{F} \times \mathcal{F} : \sup_{x \in D_p} (F(x) - G(x)) \leq 0\right\}. \qquad (A.6)$$

To this end, consider $(F_n, G_n) \in \mathcal{H}_1$ converging to $(F, G) \in \mathcal{F} \times \mathcal{F}$ in d_2. Then

$$\sup_{x \in D_p} (F(x) - G(x)) \leq \sup_{x \in D_p} (F(x) - F_n(x))$$
$$+ \sup_{x \in D_p} (F_n(x) - G_n(x)) + \sup_{x \in D_p} (G_n(x) - G(x)),$$

which is bounded above by $\|F_n - F\|_\infty + \|G_n - G\|_\infty$. Letting $n \to \infty$, we obtain $\sup_{x \in D_p}(F(x) - G(x)) \leq 0$ implying $(F, G) \in \mathcal{H}_1$. Hence, (A.6) holds, which implies $\text{cl}(\mathcal{H}_1) \subset \mathcal{A} \cup \mathcal{H}_1$. Using Lemma A.1 we obtain

$$\text{cl}(\mathcal{H}_1) \subset \mathcal{A} \cup \mathcal{H}_1 \subset \text{bd}(\mathcal{H}_1) \cup \mathcal{H}_1 = \text{cl}(\mathcal{H}_1).$$

Hence, the inclusion in (A.6) is actually an equality and $\mathcal{A} \cup \mathcal{H}_1 = \text{bd}(\mathcal{H}_1) \cup \mathcal{H}_1$. The proof will be complete if we can show that $\text{int}(\mathcal{H}_1) = \mathcal{H}_1$, because then $\text{bd}(\mathcal{H}_1) = \mathcal{A}$ follows. To that end, consider $(F, G) \in \mathcal{H}_1$. Suppose $\sup_{x \in D_p}(F(x) - G(x)) < -\epsilon$. Let $(\tilde{F}, \tilde{G}) \in \mathcal{F} \times \mathcal{F}$ be such that $d_2((\tilde{F}, \tilde{G}), (F, G)) < \epsilon/3$. Then

$$\sup_{x \in D_p} (\tilde{F}(x) - \tilde{G}(x)) \leq \sup_{x \in D_p} (F(x) - G(x)) + \|\tilde{F} - F\|_\infty + \|\tilde{G} - G\|_\infty,$$

which is less than $-\epsilon/3$. Therefore, $(\tilde{F}, \tilde{G}) \in \mathcal{H}_1$, which completes the proof. $\square$

Lemma A.1. *Under the set-up of Lemma 1,*

$$\mathcal{A} := \left\{(F, G) \in \mathcal{F} \times \mathcal{F} : \sup_{x \in D_p} (F(x) - G(x)) = 0\right\} \subset bd(\mathcal{H}_1).$$

Proof. Consider $(F, G) \in \mathcal{A}$. Because $\mathcal{A} \cap \mathcal{H}_1 = \emptyset$, it suffices to show that (F, G) is a limit point of $\mathcal{H}_1$. We will show that given $\delta > 0$, there exists $(\tilde{F}, \tilde{G}) \in \mathcal{H}_1$ so that $d_2((F, G), (\tilde{F}, \tilde{G})) < \delta$.

Since $F - G$ is continuous, its supremum over D_p is attained. Thus, there exists $C_p \subset D_p$ so that $F = G$ on C_p. Because $p \leq \lambda F + (1 - \lambda)G \leq 1 - p$ on D_p, $\inf_{z \in C_p} F(z) \geq p$ and $\sup_{z \in C_p} F(z) \leq 1 - p$. Suppose $a' = \inf C_p$ and $\beta' = \sup C_p$. Since F is continuous, there exists closed interval $L_p = [a, \beta] \supset C_p$ such that $L_p \subset D_p$ and $p/2 < F < 1 - p/2$ on $[a, \beta]$. We choose a and β so that additionally the followings hold:

1. If $a' > H^{-1}(p)$ then $a < a'$. Thus $a = a'$ only if $a' = H^{-1}(p)$, in which case, the only choice for a is a' because $[a', \beta'] \subset [a, \beta] \subset D_p$.
2. If $\beta' < H^{-1}(1 - p)$ then $\beta' < \beta$. When $\beta' = H^{-1}(1 - p)$, the only choice for β is β' because $[a', \beta'] \subset [a, \beta] \subset D_p$.

Letting $\epsilon' = \min(\delta/2, p/4)$, $t_1 = F^{-1}(1 - \epsilon')$ and $t_2 = t_1 + 1$, we define

$$\tilde{F}(x) = \begin{cases} \max(0, F(x) - \epsilon') & x < a \\ F(x) - \epsilon' & x \in [a, t_1] \\ (x - t_1)\frac{F(t_2) - \tilde{F}(t_1)}{t_2 - t_1} + \tilde{F}(t_1) & x \in (t_1, t_2] \\ F(x) & x > t_2 \end{cases} \tag{A.7}$$

That $\tilde{F}$ is a distribution function is clear from the definition. We claim that (a) $\tilde{F}$ is continuous, (b) $\|\tilde{F} - F\|_\infty < \delta$, and finally (c) $(\tilde{F}, G) \in \mathcal{H}_1$. Taking $\tilde{G} = G$, the proof of the current lemma follows. Hence, it remains to prove Claim (a), (b), and (c).

First, we prove Claim (a). Because F is continuous, $\max(0, F(x) - \epsilon')$ is continuous in x on $(-\infty, a)$. BSince $F(a) > p/2 > \epsilon'$, continuity of F also implies $\tilde{F}(a-) = F(a) - \epsilon' = F(a+)$. Therefore, $\tilde{F}$ is continuous on $(-\infty, a]$. Therefore (A.7) implies $\tilde{F}$ is continuous on $(-\infty, a]$. The continuity of $F(x) - \epsilon'$ on $(a, t_1]$ follows because F is continuous. $\tilde{F}$ is linear on $(t_1, t_2]$, and equals F on (t_2, ∞) Moreover, the left and right limits of $\tilde{F}$ agree at t_1 and t_2. Therefore $\tilde{F}$ is continuous.

To prove Claim (b), it suffices to show that $|\tilde{F}(x) - F(x)| \leq \epsilon'$ for any $x \in (-\infty, a) \cup (t_1, t_2]$. Note that if $x < a$ and $F(x) \leq \epsilon'$, then $\tilde{F}(x) = 0$. Thus, $\tilde{F}(x) - F(x) \leq \epsilon'$ in this case. If $x < a$ but $F(x) > \epsilon'$, $\tilde{F}(x) - F(x) = \epsilon'$. Therefore, $|\tilde{F}(x) - F(x)| \leq \epsilon'$ for $x \in (-\infty, a)$. On the other hand, since $t_1 = F^{-1}(1 - \epsilon')$, $F(x)$ is greater than $1 - \epsilon'$ on (t_1, t_2). Also for $x \in [t_1, t_2]$,

$$\tilde{F}(x) \geq \tilde{F}(t_1) = F(t_1) - \epsilon' = F(F^{-1}(1 - \epsilon')) - \epsilon' = 1 - 2\epsilon'.$$

Thus on $[t_1, t_2]$, $|\tilde{F}(x) - F(x)|$ is bounded by $2\epsilon'$, which is not greater than δ by our choice of ϵ'. Hence, we have shown that $\|\tilde{F} - F\|_\infty < \delta$.

To prove Claim (c), first note that $\tilde{F} \leq F$ on $\mathbb{R}$. Let us partition $D_p = [H^{-1}(p), a) \cup [a, \beta] \cup (\beta, H^{-1}(1 - p)]$. If $a = H^{-1}(p)$, then we define $[H^{-1}(p), a)$ to be the empty set. Similarly, $(\beta, H^{-1}(1 - p)]$ is non-empty only if $\beta < H^{-1}(1 - p)$.

Note that $\beta \le t_1$, implying $[a, \beta] \subset [a, t_1]$. Therefore, by (A.7),

$$\sup_{x \in [a,\beta]} (\tilde{F}(x) - G(x)) = \sup_{x \in [a,\beta]} (F(x) - G(x)) - \epsilon' \le \sup_{x \in D_p} (F(x) - G(x)) - \epsilon',$$

which equals $-\epsilon'$. Now let us consider the set $[H^{-1}(p), a)$. Of course if $a = H^{-1}(p)$, there is nothing to prove. So suppose $a > H^{-1}(p)$. Because $C_p \cap [H^{-1}(p), a) = \emptyset$, there is no $x \in [H^{-1}(p), a)$ such that $F(x) - G(x) = 0$. Therefore, if $\sup_{x \in [H^{-1}(p),a)} (F(x) - G(x)) = 0$, the set $[H^{-1}(p), a]$ must contain a limit point of C_p. However, $\mathrm{cl}(C_p) \subset [a', \beta']$ where $a < a'$ because $a > H^{-1}(p)$. Therefore, $[H^{-1}(p), a]$ can not contain any limit point of C_p either. Thus we must have $\sup_{x \in [H^{-1}(p),a)} (F(x) - G(x)) < 0$. Therefore, $\sup_{x \in [H^{-1}(p),a)} (\tilde{F}(x) - G(x)) < 0$ as well because $\tilde{F} \le F$. On the other hand, using $H^{-1}(1 - p) < t_1$, we have

$$\sup_{x \in (\beta, H^{-1}(1-p)]} (\tilde{F}(x) - G(x))$$
$$= \sup_{x \in (\beta, H^{-1}(1-p)]} (F(x) - G(x)) - \epsilon \le \sup_{x \in D_p} (F(x) - G(x)) - \epsilon',$$

which equals $-\epsilon'$. Combining the above pieces, we obtain $\sup_{x \in D_p} (\tilde{F}(x) - G(x)) < 0$, which completes the proof of $(\tilde{F}, G) \in \mathcal{H}_1$. □

Proof of Lemma 2

Since $F, G \in \mathrm{int}(\mathcal{H}_0)$, there exists $x_0' \in D_p$ and $\delta > 0$ such that $G(x_0') - F(x_0') < -3\delta$. Because F and G are continuous, we can find $x_0 \in \mathrm{int}(D_p)$ such that $G(x_0) - F(x_0) < -2\delta$. Hence, (A.41) indicates that with probability one, $\mathbb{G}_n(x_0) - \mathbb{F}_m(x_0) < -\delta$ for all sufficiently large m and n, which leads to

$$\frac{\mathbb{G}_n(x_0) - \mathbb{F}_m(x_0)}{\sqrt{\dfrac{\mathbb{F}_m(x_0)(1 - \mathbb{F}_m(x_0))}{m} + \dfrac{\mathbb{G}_n(x_0)(1 - \mathbb{G}_n(x_0))}{n}}} \le \frac{-\delta}{\sqrt{1/m + 1/n}},$$

which approaches $-\infty$ as $m, n \to \infty$. Here we used the fact that

$$\frac{\mathbb{F}_m(x_0)(1 - \mathbb{F}_m(x_0))}{m} + \frac{\mathbb{G}_n(x_0)(1 - \mathbb{G}_n(x_0))}{n} \le \frac{1}{m} + \frac{1}{n}.$$

Because $x_0 \in \mathrm{int}(D_p)$, Fact 3.1 implies $x_0 \in D_{p,m,n}$ almost surely. Therefore, the proof follows for the minimum t-statistic.

Now we will prove the result for for $T_{m,n}^{\mathrm{tsep}}$. Let us denote $q = H(x_0)$ where H was denoted to be $\lambda F + (1 - \lambda)G$. Since either $f > 0$ or $g > 0$ on D_p under Condition N, it follows that $h > 0$ on D_p. Thus $H^{-1}(q) = x_0$ (see Lemma A.3.7, pp.94, Bobkov and Ledoux, 2016). Hence,
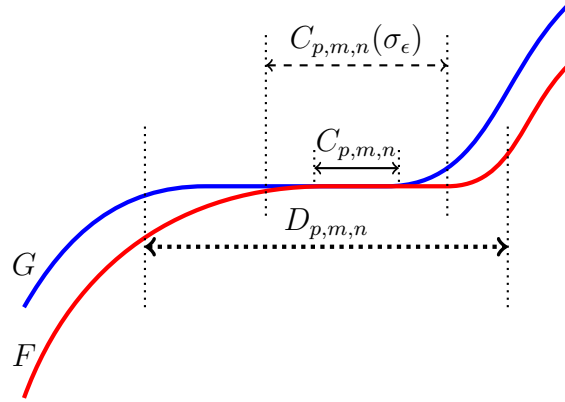
$$G(H^{-1}(q)) - F(H^{-1}(q)) < -2\delta.$$

Fig 15: An illustration of $D_{p,m,n}$, $C_{p,m,n}$, and $C_{p,m,n}(\sigma_\epsilon)$ as in the Proof of Theorem 1 for $(F, G) \in \mathcal{B}$.

Since $H^{-1}(q) \in \text{int}(D_p)$, it follows that $q \in [p, 1-p]$. Fact 3.1 also implies that $\mathbb{H}_N^{-1}(q) \to_{a.s.} H^{-1}(q)$. Combined with the fact that F and G are continuous, we obtain that

$$\limsup_{m,n}\{G(\mathbb{H}_N^{-1}(q)) - F(\mathbb{H}_N^{-1}(q))\} < -\delta$$

with probability one. Since $\|\mathbb{F}_m - F\|_\infty$ and $\|\mathbb{G}_n - G\|_\infty$ converges to zero almost surely, the above implies

$$\limsup_{m,n}\{\mathbb{G}_n(\mathbb{H}_N^{-1}(q)) - \mathbb{F}_m(\mathbb{H}_N^{-1}(q))\} < -\delta/2$$

almost surely. Note that

$$\limsup_{m,n} T_{m,n}^{\text{tsep}} \leq \limsup_{m,n} \sqrt{\frac{mn}{N}} \frac{G(\mathbb{H}_N^{-1}(q)) - F(\mathbb{H}_N^{-1}(q))}{q(1-q)} \leq -\lim_{m,n} \sqrt{\frac{mn}{N}} \frac{\delta}{2q(1-q)},$$

which equals $-\infty$ because $m/N \to \lambda$. Hence, the proof follows.

A.0.2. Proof of Theorem 1

Before proving Theorem 1, we introduce some notations and two lemmas. Recall that we defined the underlying probability space to be $(\Omega, \text{int}(\mathcal{H}_0), P)$. There exists $A \subset \Omega$ with $P(A) = 1$ such that as $m, n \to \infty$, the following assertions hold on A:

(a) $$\|\mathbb{F}_m - F\|_\infty \to 0, \quad \|\mathbb{G}_n - G\|_\infty \to 0, \tag{A.8}$$

(b) $$\left\|\sqrt{m}(\mathbb{F}_m - F) - \mathbb{V}_1 \circ F\right\|_\infty \to 0, \quad \left\|\sqrt{n}(\widehat{G}_n - G) - \mathbb{V}_2 \circ G\right\|_\infty \to 0, \tag{A.9}$$

where $\mathbb{V}_1$ and $\mathbb{V}_2$ are the Brownian bridges defined in (A.1) and (A.2), respectively.

(c) The trajectories of the Brownian bridges $\mathbb{V}_1$ and $\mathbb{V}_2$ are continuous on $\mathbb{R}$.

Define

$$\omega_{m,n}(x) = \frac{\mathbb{G}_n(x) - \mathbb{F}_m(x)}{\sqrt{\dfrac{\mathbb{F}_m(x)\left(1 - \mathbb{F}_m(x)\right)}{m} + \dfrac{\mathbb{G}_n(x)\left(1 - \mathbb{G}_n(x)\right)}{n}}} \tag{A.10}$$

and

$$\omega_0(x) = \frac{\sqrt{\lambda}\mathbb{V}_2(G) - \sqrt{1 - \lambda}\mathbb{V}_1(F)}{\sqrt{(1 - \lambda)F(x)\left(1 - F(x)\right) + \lambda G(x)\left(1 - G(x)\right)}}. \tag{A.11}$$

Let us denote $C_{p,m,n} = D_{p,m,n} \cap \{x : F(x) = G(x)\}$. Proceeding as in the proof of Fact 3.1, we can show that

$$\mathrm{Dist}(D_{p,m,n}, D_p) \to 0 \quad \text{and} \quad \mathrm{Dist}(C_{p,m,n}, C_p) \to 0 \quad \text{as} \quad m, n \to \infty \tag{A.12}$$

on A, where C_p is the contact set $D_p \cap \{x \in \mathbb{R} : F(x) = G(x)\}$ discussed in Section 4. Now we state the first lemma, which we require for proving part (A).

Lemma A.2. *Suppose $D \subset \mathbb{R}$ such that F and G are bounded away from 0 and 1 on D. Then, under the conditions of Theorem 1, the following holds on A:*

$$\lim_{m,n\to\infty} \left\| \omega_{m,n} - \omega_0 - \sqrt{\frac{mn}{N}}\vartheta_{m,n}^{-1/2}(G - F) \right\|_D = 0,$$

where

$$\vartheta_{m,n}(x) = \frac{n\mathbb{F}_m(x)\left(1 - \mathbb{F}_m(x)\right)}{N} + \frac{m\mathbb{G}_n(x)\left(1 - \mathbb{G}_n(x)\right)}{N}. \tag{A.13}$$

Moreover, on A,

$$\lim_{m,n\to\infty} \|\vartheta_{m,n}^{-1/2} - \vartheta_0^{-1/2}\|_D = 0, \tag{A.14}$$

where

$$\vartheta_0(x) = (1 - \lambda)F(x)(1 - F(x)) + \lambda G(x)(1 - G(x)).$$

Proof. Note that

$$\omega_{m,n}(x) - \omega_0(x) - \sqrt{\frac{mn}{N}}\vartheta_{m,n}(x)^{-1/2}\left(G(x) - F(x)\right)$$

$$= \sqrt{\frac{mn}{N}}\frac{\mathbb{G}_n(x) - G(x) - (\mathbb{F}_m(x) - F(x))}{\vartheta_{m,n}(x)^{1/2}} - \frac{\sqrt{\lambda}\mathbb{V}_2(G(x)) - \sqrt{1 - \lambda}\mathbb{V}_1(F(x))}{\vartheta_{m,n}(x)^{1/2}}$$

$$+ \left(\sqrt{\lambda}\mathbb{V}_2(G(x)) - \sqrt{1 - \lambda}\mathbb{V}_1(F(x))\right)\left(\frac{1}{\vartheta_{m,n}(x)^{1/2}} - \frac{1}{\vartheta_0(x)^{1/2}}\right).$$

Because F and G are bounded away from 0 and 1 on D, there exist $c > 0$ and $c' < 1$ such that $F(x), G(x) \in (c, c')$ for $x \in D$. We see that (A.8) and $m/N \to \lambda$ imply on A, the following hold:

$$\limsup_{m,n\to\infty} \|\vartheta_{m,n}\|_D, \|\vartheta_0\|_D < c'(1-c), \tag{A.15}$$

$$\limsup_{m,n\to\infty} \|\vartheta_{m,n}^{-1/2}\|_D, \|\vartheta_0^{-1/2}\|_D < c^{-1/2}(1-c')^{-1/2}. \tag{A.16}$$

Since on the probability one set A, $\vartheta_{m,n}$ converges to ϑ_0 uniformly, and both functions are bounded below on D, the following also holds:

$$\lim_{m,n\to\infty} \|\vartheta_{m,n}^{-1/2} - \vartheta_0^{-1/2}\|_D = 0.$$

Therefore,

$$\limsup_{m,n\to\infty} \left\| \vartheta_{m,n}^{-1/2} \left\{ \sqrt{\frac{mn}{N}} \left(\mathbb{G}_n - G - (\mathbb{F}_m - F) \right) - \sqrt{\lambda}\mathbb{V}_2(G) + \sqrt{1-\lambda}\mathbb{V}_1(F) \right\} \right\|_D$$

$$\leq \limsup_{m,n\to\infty} \|\vartheta_{m,n}^{-1/2}\|_D$$

$$\times \limsup_{m,n\to\infty} \left\| \sqrt{\frac{mn}{N}} \left(\mathbb{G}_n - G - (\mathbb{F}_m - F) \right) - \sqrt{\lambda}\mathbb{V}_2(G) + \sqrt{1-\lambda}\mathbb{V}_1(F) \right\|_D,$$

which, since $\limsup_{m,n\to\infty} \|\vartheta_{m,n}^{-1/2}\|_D$ is bounded, converges to 0 on A since $m/N \to \lambda$ and (A.9) holds. Also on A, the Brownian bridges $\mathbb{V}_1$ and $\mathbb{V}_2$ have continuous trajectories, which indicates $\sqrt{\lambda}\mathbb{V}_2(G) - \sqrt{1-\lambda}\mathbb{V}_1(F)$ is a continuous function, and hence bounded on D. Therefore, on A,

$$\limsup_{m,n\to\infty} \left\| \left(\sqrt{\lambda}\mathbb{V}_2(G(x)) - \sqrt{1-\lambda}\mathbb{V}_1(F(x)) \right) \left(\frac{1}{\vartheta_{m,n}(x)^{1/2}} - \frac{1}{\vartheta_0(x)^{1/2}} \right) \right\|_D$$

$$\leq \left\| \sqrt{\lambda}\mathbb{V}_2(G) - \sqrt{1-\lambda}\mathbb{V}_1(F) \right\|_D \limsup_{m,n\to\infty} \|\vartheta_{m,n}^{-1/2} - \vartheta_0^{-1/2}\|_D$$

equals zero, which completes the proof. $\qquad\square$

The second lemma, which is required for proving part (B), relies on the objects $s_1 : \tilde{t} \mapsto \frac{d}{dt}F \circ H^{-1}(t)\big|_{t=\tilde{t}}$, $s_2 : \tilde{t} \mapsto \frac{d}{dt}G \circ H^{-1}(t)\big|_{t=\tilde{t}}$, and

$$\mathbb{L}_0(t) = (1-\lambda)\left(\lambda^{-1/2}s_2(t)\mathbb{V}_1(F \circ H^{-1}(t)) - (1-\lambda)^{-1/2}s_1(t)\mathbb{V}_2(G \circ H^{-1}(t)) \right). \tag{A.17}$$

Note that

$$s_1(t) = \frac{f \circ H^{-1}(t)}{\lambda f \circ H^{-1}(t) + (1-\lambda)g \circ H^{-1}(t)}$$

$$\text{and} \quad s_2(t) = \frac{g \circ H^{-1}(t)}{\lambda f \circ H^{-1}(t) + (1-\lambda)g \circ H^{-1}(t)}. \tag{A.18}$$

Because $\mathbb{V}_1$ and $\mathbb{V}_2$ are independent Brownian bridges, it is immediate that $\mathbb{L}_0$ is a Gaussian process.

We now state a lemma that concerns the two sample empirical process $\sqrt{N}(\mathbb{F}_m \circ \mathcal{H}^{-1}(t) - F \circ H^{-1}(t))$.

Lemma A.3. *Under the setting of Theorem 1,*

$$\sup_{t \in [p', 1-p']} \left| \frac{\sqrt{N}\left(\mathbb{F}_m \circ \mathcal{H}^{-1}(t) - F \circ H^{-1}(t)\right)}{[t(1-t)]^{1/2}} - \frac{\mathbb{L}_0(t)}{[t(1-t)]^{1/2}} \right| \to_p 0 \qquad (A.19)$$

for any $p' \in (0, 1/2)$. Moreover, $\mathbb{L}_0$ has continuous trajectories almost surely.

Proof of Lemma A.3. Theorem 4.1 and Corollary 4.1 of Pyke and Shorack (1968) indicate (A.19), where here we emphasize that (A.17) represents the corrected formula for $\mathbb{L}_0$, given by (30) of Ledwina and Wyłupek (2012), rather than the original formula for this quantity given in (3.8) of Pyke and Shorack (1968).

Now observe that, because $h > 0$ on D_p under Condition N, H is strictly increasing, which implies that H^{-1} is a continuous function. Since f, g and H^{-1} are continuous, (A.18) implies that s_1 and s_2 are both continuous. Therefore, from (A.17) and the fact that $\mathbb{V}_1$ and $\mathbb{V}_2$ have continuous trajectories almost surely, it is not hard to see that $\mathbb{L}_0$ is continuous almost surely. $\square$

Our next lemma characterizes the Gaussian process $\mathbb{L}_0$ on the set C_p when $C_p \subset \mathrm{int}(D_p)$.

Lemma A.4. *Suppose F and G are as in Theorem 1 and $C_p \subset int(D_p)$. Then the Gaussian process*

$$\left\{ \sqrt{\frac{\lambda}{1-\lambda}} \mathbb{L}_0(t) : t \in H(C_p) \right\}$$

is distributed as $\{\mathbb{U}(t) : t \in H(C_p)\}$ where $\mathbb{U}$ is a Brownian bridge.

Proof of Lemma A.4. If $C_p = \emptyset$ then the statement is vacuously true. So we will assume that $C_p \neq \emptyset$. We first claim that if $C_p \subset \mathrm{int}(D_p)$, then $s_1(t) = s_2(t) = 1$ for all $t \in H(C_p)$. If the claim is true, then from (A.17) it follows that

$$\mathbb{L}_0(t) = (1-\lambda)\left(\lambda^{-1/2}\mathbb{V}_1(F \circ H^{-1}(t)) - (1-\lambda)^{-1/2}\mathbb{V}_2(G \circ H^{-1}(t))\right)$$

for $t \in H(C_p)$. Also for $t \in H(C_p)$,

$$F(H^{-1}(t)) = G(H^{-1}(t)) = \lambda F(H^{-1}(t)) + (1-\lambda)G(H^{-1}(t)) = H(H^{-1}(t)).$$

Under the set up of Theorem 1, H has positive and continuous density on a open neighborhood of D_p. Therefore, H is strictly increasing on this open set, which implies H^{-1} is also continuous on this open set. Hence, $H(H^{-1}(t)) = t$ for all $t \in H(C_p)$, leading to

$$\mathbb{V}_1(F \circ H^{-1}(t)) = \mathbb{V}_1(t) \quad \text{and} \quad \mathbb{V}_2(G \circ H^{-1}(t)) = \mathbb{V}_2(t),$$

which implies

$$\mathbb{L}_0(t) = (1 - \lambda)\left(\lambda^{-1/2}\mathbb{V}_1(t) - (1 - \lambda)^{-1/2}\mathbb{V}_2(t)\right) \quad \text{for all } t \in H(C_p).$$

Since $\mathbb{V}_1$ and $\mathbb{V}_2$ are independent Brownian bridge processes, it follows that the Gaussian process $\{\mathbb{L}'(t) : t \in [0, 1]\}$ defined by

$$\mathbb{L}'(t) = (1 - \lambda)\left(\lambda^{-1/2}\mathbb{V}_1(t) - (1 - \lambda)^{-1/2}\mathbb{V}_2(t)\right)$$

has variance

$$\text{var}(\mathbb{L}'(t)) = (1 - \lambda)^2\left(\lambda^{-1} + (1 - \lambda)^{-1}\right)(t - t^2) = \frac{1 - \lambda}{\lambda}(t - t^2).$$

In particular, it can be seen that $\sqrt{\lambda/(1 - \lambda)}\mathbb{L}'$ is a Brownian bridge. Hence the proof follows if we can prove the claim that $s_1(t) = s_2(t) = 1$ for all $t \in H(C_p)$.

Since $(F, G) \in \text{bd}(\mathcal{H}_0)$, Lemma 1 implies that $F - G$ attains maxima at C_p whenever $C_p \neq \emptyset$. Because $F - G$ is continuously differentiable and $C_p \subset \text{int}(D_p)$, $f - g = 0$ on C_p. Therefore, $f(H^{-1}(t)) = g(H^{-1}(t))$ for all $t \in H(C_p)$. The claim now follows from (A.18). $\qquad\square$

A.0.3. Proof of Lemma 3

Proof of Lemma 3. Since $t \in H(C_p)$, $H^{-1}(t) \in C_p$, implying

$$F(H^{-1}(t)) = G(H^{-1}(t)) = H(H^{-1}(t)).$$

Arguing as in the proof of Lemma A.4, we can show that $H(H^{-1}(t)) = t$ under the set up of Theorem 1. Thus (A.17) implies

$$\mathbb{L}_0(t) = (1 - \lambda)\left(\lambda^{-1/2}s_2(t)\mathbb{V}_1(t) - (1 - \lambda)^{-1/2}s_1(t)\mathbb{V}_2(t)\right).$$

Because $\mathbb{V}_1$ and $\mathbb{V}_2$ are independent Brownian bridges, $\mathbb{L}_0$ is distributed as a centered normal variable with variance

$$\text{var}(\mathbb{L}_0(t)) = (1 - \lambda)^2\left(\lambda^{-1}s_2(t)^2\text{var}(\mathbb{V}_1(t)) + (1 - \lambda)^{-1}s_1(t)^2\text{var}(\mathbb{V}_2(t))\right)$$

$$= (1 - \lambda)^2 t(1 - t)\left(\lambda^{-1}s_2(t)^2 + (1 - \lambda)^{-1}s_1(t)^2\right) \qquad (A.20)$$

because

$$\text{var}(\mathbb{V}_1(t)) = \text{var}(\mathbb{V}_2(t)) = t(1 - t).$$

Thus it follows that

$$\sigma^2_{TSEP} = \frac{\lambda}{1 - \lambda}\frac{\text{var}(\mathbb{L}_0(t))}{t(1 - t)} = \lambda s_1(t)^2 + (1 - \lambda)s_2(t)^2$$

$$= \frac{\lambda(f \circ H^{-1}(t))^2 + (1 - \lambda)(g \circ H^{-1}(t))^2}{\left(\lambda f \circ H^{-1}(t) + (1 - \lambda)g \circ H^{-1}(t)\right)^2}$$

where the last step follows from (A.18). This completes the proof of (4.9).

Next, we will establish the lower bound on σ_{TSEP}^2. Since $x \mapsto x^2$ is convex, by Jensen's inequality,

$$(1 - \lambda)s_2(t)^2 + \lambda_1 s_1(t)^2 \geq \left((1 - \lambda)s_1(t) + \lambda s_1(t)\right)^2 = 1 \qquad (A.21)$$

since $\lambda s_1 + (1 - \lambda)s_2 = 1$ by (A.18). Therefore, (A.20) implies

$$\text{var}(\mathbb{L}_0(t)) \geq \frac{1 - \lambda}{\lambda} t(1 - t)$$

which implies

$$1 \leq \frac{\lambda}{1 - \lambda} \frac{\text{var}(\mathbb{L}_0(t))}{t(1 - t)} = \sigma_{TSEP}^2. \qquad (A.22)$$

Since $x \mapsto x^2$ is strictly convex, the inequality in (A.21) is an equality if and only if $s_1(t) = s_2(t)$. Because f and g are positive on D_p, (A.18) implies that the latter occurs if and only if $f(H^{-1}(t)) = g(H^{-1}(t))$.

Now we will establish the upper bound on σ_{TSEP}^2. Because $\lambda s_1 + (1-\lambda)s_2 = 1$, we also have $s_2 = (1 - \lambda s_1)/(1 - \lambda)$. Therefore, using (A.20), we derive that $\text{var}(\mathbb{L}_0(t))$ equals

$$(1 - \lambda)^2 t(1 - t)\left(\lambda^{-1}\left(\frac{1 - \lambda s_1(t)}{(1 - \lambda)}\right)^2 + \frac{s_1(t)^2}{1 - \lambda}\right)$$

$$= t(1 - t)\left(\lambda^{-1} - 2s_1(t) + \lambda s_1(t)^2 + (1 - \lambda)s_1(t)^2\right)$$

$$= t(1 - t)\left(\lambda^{-1} - 2s_1(t) + s_1(t)^2\right)$$

$$= t(1 - t)\left(\lambda^{-1} - 1 + (1 - s_1(t))^2\right). \qquad (A.23)$$

Note that (A.18) implies $s_1(t) \in [0, \lambda^{-1}]$. On any interval, the convex function $\lambda^{-1} - 2x + x^2$ attains maxima at either endpoints of the interval. Therefore,

$$\frac{\text{var}(\mathbb{L}_0(t))}{t(1 - t)} \leq \max\{\lambda^{-1}, \lambda^{-1} - 1 + (1 - \lambda^{-1})^2\}$$

$$= \max\left\{\lambda^{-1}, \lambda^{-1}(\lambda^{-1} - 1)\right\}.$$

Therefore,

$$\sigma_{TSEP}^2 = \frac{\lambda}{1 - \lambda} \frac{\text{var}(\mathbb{L}_0(t))}{t(1 - t)} \leq \max\{(1 - \lambda)^{-1}, \lambda^{-1}\}, \qquad (A.24)$$

which, combined with (A.22), completes the proof of part (A) of the current lemma.

Now if $s_1(t) = \lambda^{-1}$, then $\sigma^2_{TSEP} = \lambda^{-1}$. Using (A.18), we can write

$$s_1(t) = \frac{1}{\lambda + (1-\lambda)g \circ H^{-1}(t)/f \circ H^{-1}(t)}.$$

Thus $s_1(t) = \lambda^{-1}$ if and only if $g \circ H^{-1}(t)/f \circ H^{-1}(t) = 0$. However, under our set up, f and g are positive on D_p, implying $g \circ H^{-1}(t)/f \circ H^{-1}(t) > 0$ for any $t \in [p, 1-p]$. Regardless, since the function $x \mapsto 1/(\lambda + (1-\lambda)x)$ is right continuous at 0, given any $\epsilon > 0$, there exists $C_\lambda > 0$, depending only on $\lambda > 0$, so that if $g \circ H^{-1}(t)/f \circ H^{-1}(t) < C_\lambda$, then $s_1(t) > \lambda^{-1} - \epsilon/2$. Suppose ϵ is so small such that $\epsilon/2 < \lambda^{-1} - 1$. Then from (A.23) it also follows that

$$\begin{aligned}
\sigma^2_{TSEP} &\geq \frac{\lambda}{1-\lambda}\left(\lambda^{-1} - 1 + (\lambda^{-1} - \epsilon/2 - 1)^2\right) \\
&\geq \frac{\lambda}{1-\lambda}\left(\lambda^{-1} - 1 + (\lambda^{-1} - 1)^2 - \epsilon(\lambda^{-1} - 1)\right) \\
&= \lambda^{-1} - \epsilon,
\end{aligned}$$

which completes the proof of part B of the current lemma.

Similarly, we can show that if $s_1(t) = 0$, then $\sigma^2_{TSEP} = (1-\lambda)^{-1}$. Using (A.18) again, we can write

$$s_1(t) = \frac{f \circ H^{-1}(t)/g \circ H^{-1}(t)}{\lambda f \circ H^{-1}(t)/g \circ H^{-1}(t) + 1 - \lambda}.$$

Therefore $s_1(t) = 0$ if and only if $f \circ H^{-1}(t)/g \circ H^{-1}(t) = 0$, which is impossible since $f, g > 0$ on D_p under our set up. However, since the map $x \mapsto x/(\lambda x + 1 - \lambda)$ is right continuous at 0, given any $\epsilon > 0$, we can find $C'_\lambda > 0$, depending only on $\lambda > 0$, so that if $f \circ H^{-1}(t)/g \circ H^{-1}(t) < C'_\lambda$, then $s_1(t) < (1-\lambda)\epsilon/(2\lambda)$. If $\epsilon < 2\lambda/(1-\lambda)$, then from (A.23), it also follows that

$$\begin{aligned}
\sigma^2_{TSEP} &\geq \frac{\lambda}{1-\lambda}\left(\lambda^{-1} - 1 + (1 - (1-\lambda)\epsilon/(2\lambda))^2\right) \\
&\geq \frac{\lambda}{1-\lambda}\left(\lambda^{-1} - (1-\lambda)\epsilon/\lambda\right) \\
&= (1-\lambda)^{-1} - \epsilon,
\end{aligned}$$

which completes the proof of part C of the current lemma. $\square$

A.0.4. *Proof of part (A)*

First we will consider the case when $C_p \neq D_p$. The main steps of the proof are as follows:

(a) We fix $\epsilon > 0$, and choose some set $C_{p,m,n}(\sigma_\epsilon)$ satisfying $C_{p,m,n}(\sigma_\epsilon) \subset D_{p,m,n}$. Here $C_{p,m,n}(\sigma_\epsilon)$ depends on σ_ϵ, which is a random positive number

that will be chosen appropriately. Next we partition $D_{p,m,n}$ as follows (see Figure 15):

$$D_{p,m,n} = \underbrace{[D_{p,m,n} \setminus C_{p,m,n}(\sigma_\epsilon)]}_{B_{p,m,n}(\sigma_\epsilon)} \cup \underbrace{[C_{p,m,n}(\sigma_\epsilon) \setminus C_{p,m,n}]}_{E_{p,m,n}(\sigma_\epsilon)} \cup C_{p,m,n}.$$

Therefore, we can rewrite $T_{m,n}^{\min}(\mathbb{F}_m, \mathbb{G}_n)$ as

$$\min\left\{ \inf_{x \in B_{p,m,n}(\sigma_\epsilon)} \omega_{m,n}(x), \inf_{x \in E_{p,m,n}(\sigma_\epsilon)} \omega_{m,n}(x), \inf_{x \in C_{p,m,n}} \omega_{m,n}(x) \right\}.$$
(A.25)

(b) We show that on A,

$$\liminf_{m,n \to \infty} \inf_{x \in B_{p,m,n}(\sigma_\epsilon)} \omega_{m,n}(x) = \infty.$$

(c) We show that the following holds on A:

$$\liminf_{m,n \to \infty} \inf_{x \in E_{p,m,n}(\sigma_\epsilon)} \omega_{m,n}(x) > \limsup_{m,n \to \infty} \inf_{x \in C_{p,m,n}} \omega_{m,n}(x) - \epsilon.$$

(d) Finally we show that on A,

$$\lim_{m,n \to \infty} \inf_{x \in C_{p,m,n}} \omega_{m,n}(x) = \inf_{x \in C_p} \omega_0(x),$$

from which, we show that, (4.8) follows.

We will restrict our attention only to the set A for this part of the proof. However, because $P(A) = 1$, this serves our purpose.

Proof of step (a)

For $\sigma > 0$ and $p \in [0, 1]$, let us define

$$C_p(\sigma) = \{x \in D_p : Dist(x, C_p) < \sigma\}.$$
(A.26)

Because H is continuous, there exists $p' \in (0, p)$ such that $H^{-1}(p') < H^{-1}(p)$, which implies $D_{p'} \supset D_p$. For any $\sigma > 0$, we define

$$C_{p,m,n}(\sigma) = C_{p'}(\sigma) \cap D_{p,m,n}.$$
(A.27)

The first task is to properly choose a σ_ϵ so that certain properties hold on $C_{p'}(\sigma_\epsilon)$.

Since $A \subset \Omega$ was chosen so that on this set $\mathbb{V}_1$ and $\mathbb{V}_2$ have continuous trajectories, ω_0 also has continuous trajectory, which is also uniformly continuous on $D_{p'}$ because the latter is a compact set. Hence, for each such trajectory,

$$\sigma_\epsilon^1 = \sup\left\{ \sigma > 0 : |x-y| < \sigma \text{ implies } |\omega_0(x) - \omega_0(y)| < \epsilon/2 \quad \text{for all } x, y \in D_{p'} \right\}$$

is well defined and positive. Note that σ_ϵ^1 is a random quantity, which can take the value 0 on the set A^c but $\sigma_\epsilon^1 > 0$ on A.

On the other hand, since $C_p \neq D_p$ and $p' < p$, we have $C_{p'} \neq D_{p'}$. Therefore there exists $x \in D_{p'}$ so that $G(x) - F(x) > 0$. Suppose $\delta_p = (G(x) - F(x))/4$. Because $G - F$ is continuous and $D_{p'}$ is compact, $D_{p'} \setminus C_{p'}$ contains at least one interval where $G - F > 2\delta_p$. Using the continuity of $G - F$, we can choose σ_2 so small such that $D_{p'} \setminus C_{p'}(\sigma_2)$ contains at least an interval where $G - F > \delta_p$, i.e. $D_{p'} \setminus C_{p'}(\sigma_2) \neq \emptyset$. We will take

$$\sigma_\epsilon = \min(\sigma_\epsilon^1, \sigma_2).$$

Since σ_ϵ^1 is random, σ_ϵ is also random. Moreover, $\sigma_\epsilon > 0$ on A.

Note that, $B_{p,m,n}(\sigma_\epsilon)$ and $E_{p,m,n}(\sigma_\epsilon)$ can be empty for small m and n. In that case, we define the infimum of $\omega_{m,n}$ over those set to be ∞. Also since

$$\inf_{x \in D_{p'}} \min\{f(x), g(x)\} > 0,$$

H^{-1} is continuous on p'.

A.0.5. Proof of step (b)

For any $\sigma > 0$, let us denote $B_{p'}(\sigma) = D_{p'} \setminus C_{p'}(\sigma)$. We claim that $B_{p'}(\sigma_\epsilon)$, is non-empty which follows because our choice of σ_ϵ in step (a) implies $B_{p'}(\sigma_\epsilon) \supset B_{p'}(\sigma_2)$ where $B_{p'}(\sigma_2) \neq \emptyset$ by definition of σ_2. By the continuity of $G - F$ it follows that there exists a random quantity $\delta_{\sigma_\epsilon} > 0$ so that $G - F > \delta_{\sigma_\epsilon}$ on $B_{p'}(\sigma_\epsilon)$. Recall the definition of $\vartheta_{m,n}$ from (A.13). Since f and g are bounded away from 0 on $D_{p'}$, Lemma A.2 implies that on A,

$$\inf_{x \in B_{p'}(\sigma_\epsilon)} \omega_{m,n}(x) \geq \inf_{x \in B_{p'}(\sigma_\epsilon)} \omega_0(x) + \sqrt{\frac{mn}{N}} \frac{\delta_p}{\sup_{x \in B_{p'}(\sigma_\epsilon)} \vartheta_{m,n}(x)^{1/2}} + o(1).$$

where the $o(1)$ term approaches zero as $m, n \to \infty$. However, by (A.16), $\vartheta_{m,n}^{-1/2}$ is bounded above by constant depending only on p', F, and G on $D_{p'}$. Noting that the continuous function ω_0 is bounded on $D_{p'}$, and using $m/N \to \lambda$, we deduce that

$$\liminf_{m,n \to \infty} \inf_{x \in B_{p'}(\sigma_\epsilon)} \omega_{m,n}(x) = \infty$$

on A. Note that (A.27) implies

$$B_{p,m,n}(\sigma_\epsilon) = D_{p,m,n} \cap (C_{p'}(\sigma_\epsilon) \cap D_{p,m,n})^c = D_{p,m,n} \setminus C_{p'}(\sigma_\epsilon).$$

On the set A, we thus have $D_{p,m,n} \subset D_{p'}$ eventually as $m, n \to \infty$, which implies

$$B_{p,m,n}(\sigma_\epsilon) = D_{p,m,n} \setminus C_{p'}(\sigma_\epsilon) \subset D_{p'} \setminus C_{p'}(\sigma_\epsilon) = B_{p'}(\sigma_\epsilon)$$

eventually as $m, n \to \infty$, leading to

$$\liminf_{m,n \to \infty} \inf_{x \in B_{p,m,n}(\sigma_\epsilon)} \omega_{m,n}(x) = \infty. \tag{A.28}$$

Proof of step (c)

We denote the boundary of the set C_p by $\mathrm{bd}(C_p)$. We can show that C_p is a closed set, which implies $\mathrm{bd}(C_p) \subset C_p$. Therefore, for all $y \in \mathrm{bd}(C_p)$, we have $G(y) = F(y)$. Therefore, by Lemma A.2, for any $x \in E_{p,m,n}(\sigma_\epsilon)$, and $y \in \mathrm{bd}(C_p) \cap D_{p,m,n}$, the following holds on A for all sufficiently large m and n:

$$\left| \omega_{m,n}(x) - \omega_{m,n}(y) - \sqrt{\frac{mn}{N}} \vartheta_{m,n}(x)^{-1/2} \Big(G(x) - F(x) \Big) \right|$$

$$\leq \epsilon/2 + \sup_{\substack{x \in E_{p,m,n}(\sigma_\epsilon), \\ y \in \mathrm{bd}(C_p) \cap D_{p,m,n}}} |\omega_0(x) - \omega_0(y)|$$

$$\overset{(a)}{\leq} \epsilon/2 + \sup_{|x-y| < \sigma_\epsilon, \ x,y \in D_{p'}} |\omega_0(x) - \omega_0(y)|,$$

which, by our choice of σ_ϵ, is not larger than $\epsilon/2$. Here (a) follows because $E_{p,m,n}(\sigma_\epsilon) \subset C_{p,m,n}(\sigma_\epsilon) \subset C_{p'}(\sigma_\epsilon)$ for sufficiently large m and n. The above leads to

$$\inf_{x \in E_{p,m,n}(\sigma_\epsilon)} \omega_{m,n}(x) \geq \sqrt{\frac{mn}{N}} \inf_{x \in E_{p,m,n}(\sigma_\epsilon)} \frac{G(x) - F(x)}{\vartheta_{m,n}(x)^{1/2}}$$

$$+ \inf_{y \in D_{p,m,n} \cap \mathrm{bd}(C_p)} \omega_{m,n}(y) - \epsilon.$$

Since $(F, G) \in \mathrm{bd}(\mathcal{H}_0)$, we have $G - F \geq 0$, which yields

$$\liminf_{m,n \to \infty} \inf_{x \in E_{p,m,n}(\sigma_\epsilon)} \omega_{m,n}(x) \geq \limsup_{m,n \to \infty} \inf_{y \in D_{p,m,n} \cap C_p} \omega_{m,n}(y) - \epsilon.$$

Noting $C_{p,m,n} = D_{p,m,n} \cap C_p$, we conclude this step.

Proof of step (d)

This step follows from Lemma A.2. To see this, note that, Lemma A.2 implies that $\|\omega_{m,n} - \omega_0\|_{C_{p'}} \to 0$ on A as $m, n \to \infty$. Because $F = G$ on $C_{p'}$, it also follows that

$$\omega_0(x) = \frac{\sqrt{\lambda} \mathbb{V}_2(F(x)) - \sqrt{1-\lambda} \mathbb{V}_1(F(x))}{\sqrt{F(x)\Big(1 - F(x)\Big)}} = \frac{\mathbb{U} \circ F(x)}{\sqrt{F(x)\Big(1 - F(x)\Big)}}$$

where $\mathbb{U} = \sqrt{\lambda}\mathbb{V}_2 - \sqrt{1-\lambda}\mathbb{V}_1$ is a Brownian bridge. Since on A, $D_{p,m,n} \subset D_p \subset D_{p'}$ for sufficiently large m and n, it entails that $C_{p,m,n} \subset C_p \subset C_{p'}$ for sufficiently large m and n as well. Therefore,

$$\left\| \omega_{m,n} - \frac{\mathbb{U} \circ F}{\sqrt{F(1-F)}} \right\|_{C_{p,m,n}} \to 0$$

as well. Since $Dist(C_{p,m,n}, C_p) \to 0$ on A by (A.12), the above readily yields that on A,

$$\lim_{m,n\to\infty} \inf_{x\in C_{p,m,n}} \omega_{m,n}(x) = \inf_{x\in C_p} \frac{\mathbb{U}\circ F(x)}{\left(F(x)\left(1-F(x)\right)\right)^{1/2}}. \tag{A.29}$$

Combining steps (a)–(d) with (A.25) yields on A,

$$\limsup_{m,n\to\infty} \inf_{x\in C_{p,m,n}} \omega_{m,n}(x) - \epsilon \le \liminf_{m,n\to\infty} T_{m,n}^{\min}(\mathbb{F}_m, \mathbb{G}_n)$$
$$\le \limsup_{m,n\to\infty} T_{m,n}^{\min}(\mathbb{F}_m, \mathbb{G}_n) \le \liminf_{m,n\to\infty} \inf_{x\in C_{p,m,n}} \omega_{m,n}(x).$$

Letting $\epsilon \to 0$, we have

$$\lim_{m,n\to\infty} T_{m,n}^{\min}(\mathbb{F}_m, \mathbb{G}_n) = \lim_{m,n\to\infty} \inf_{x\in C_{p,m,n}} \omega_{m,n}(x) = \inf_{x\in C_p} \frac{\mathbb{U}\circ F(x)}{\sqrt{F(x)\left(1-F(x)\right)}}$$

on A, where the last step follows from (A.29). The above concludes the proof of (4.8) when $C_p \ne D_p$.

Now suppose $C_p = D_p$. In this case, we will only use Step C and D. Let us restrict our attention to only A. We define $\sigma_\epsilon = \sigma_{\mathbb{E}}^1$. Letting $C_{p'}(\sigma_\epsilon)$ be as in (A.26), we have $D_p \subset \text{int}(C_{p'}(\sigma_\epsilon))$, and also, $D_{p,m,n} \subset \text{int}(C_{p'}(\sigma_\epsilon))$ for sufficiently large m and n. Let us also denote $C_{p,m,n}$ as in (A.27) and $E_{p,m,n} = C_{p,m,n}(\sigma_\epsilon) \backslash C_{p,m,n}$ as in step C. Then for large m and n, the partition $D_{p,m,n} = E_{p,m,n}(\sigma_\epsilon) \cup C_{p,m,n}$ is valid. Therefore, the proof follows from combining Step C and D.

Proof of part (B) of Theorem 1

We now prove part (B). For the ease of reference, we let

$$\nu_{m,n}(t) = \frac{\sqrt{\dfrac{mn}{N}}\left(\mathbb{G}_n \circ \mathcal{H}^{-1}(t) - \mathbb{F}_m \circ \mathcal{H}^{-1}(t)\right)}{[t(1-t)]^{1/2}}.$$

Note that $T_{m,n}^{\text{tsep}}(\mathbb{F}_m, \mathbb{G}_n) = \inf_{t\in[p,1-p]} \nu_{m,n}(t)$.

We start by studying the numerator of the above display. Noting that $N\mathcal{H} = m\mathbb{F}_m + n\mathbb{G}_n$, we derive that

$$\sqrt{\frac{mn}{N}}\left(\mathbb{G}_n \circ \mathcal{H}^{-1}(t) - \mathbb{F}_m \circ \mathcal{H}^{-1}(t)\right)$$
$$= -\sqrt{\frac{mN}{n}}\left(\mathbb{F}_m \circ \mathcal{H}^{-1}(t) - t\right) + \sqrt{\frac{mN}{n}}\left(\mathcal{H}\circ \mathcal{H}^{-1}(t) - t\right).$$

Combining the fact that $\sup_{t \in [0,1]} \left| \mathcal{H} \circ \mathcal{H}^{-1}(t) - t \right| \leq 1/N$ (p. 762 of Pyke and Shorack, 1968) with the fact that $m/N \to \lambda$, we obtain that

$$\sup_{t \in [0,1]} \left| \sqrt{\frac{mn}{N}} \left(\mathbb{G}_n \circ \mathcal{H}^{-1}(t) - \mathbb{F}_m \circ \mathcal{H}^{-1}(t) \right) + \sqrt{\frac{mN}{n}} \left(\mathbb{F}_m \circ \mathcal{H}^{-1}(t) - t \right) \right| \to 0 \tag{A.30}$$

with probability one. The above readily shows that

$$\sup_{t \in [p,1-p]} \left| \nu_{m,n}(t) + \sqrt{\frac{mN}{n}} \frac{\mathbb{F}_m \circ \mathcal{H}^{-1}(t) - t}{[t(1-t)]^{1/2}} \right| \to_{a.s.} 0.$$

Combining the above with (A.19), we see that

$$\sup_{t \in [p,1-p]} \left| \nu_{m,n}(t) + \sqrt{\frac{\lambda}{1-\lambda}} \frac{\mathbb{L}_0(t)}{[t(1-t)]^{1/2}} + \sqrt{\frac{mN}{n}} \frac{(F \circ H^{-1}(t) - t)}{[t(1-t)]^{1/2}} \right| \to_p 0.$$

Upon noting that

$$F \circ H^{-1}(t) - t = F \circ H^{-1}(t) - H \circ H^{-1}(t) = (1-\lambda)\left(F \circ H^{-1}(t) - G \circ H^{-1}(t) \right),$$

the preceding limit reduces to

$$\sup_{t \in [p,1-p]} \left| \nu_{m,n}(t) + \sqrt{\frac{\lambda}{1-\lambda}} \frac{\mathbb{L}_0(t)}{[t(1-t)]^{1/2}} - (1-\lambda)\sqrt{\frac{mN}{n}} \nu(t) \right| \to_p 0, \tag{A.31}$$

where

$$\nu(t) = [G \circ H^{-1}(t) - F \circ H^{-1}(t)]/[t(1-t)]^{1/2}. \tag{A.32}$$

If we take any subsequence of the random sequence on the left side of the above, we can find a further subsequence that approaches zero almost surely. Suppose that we can show, along the latter subsequence, that $T_{m,n}^{\text{tsep}}(\mathbb{F}_m, \mathbb{G}_n) - \inf_{t \in H(C_p)} \mathbb{U}(t)/[t(1-t)]^{1/2}$ converges almost surely to zero. In light of the fact that the limit does not depend on the choice of sequence or subsequence, Theorem 5.7 of Shorack (2000) would then imply that the whole sequence converges weakly to the same limit, namely zero. Since weak convergence to a constant is equivalent to convergence in probability to that constant, this would complete the proof. Therefore, in what follows, we use m', n' to denote members of a subsequence along which (A.31) holds almost surely and set out to prove that, as $m', n' \to \infty$,

$$T_{2,m',n'}(\mathbb{F}_{m'}, \mathbb{G}_{n'}) + \sqrt{\frac{\lambda}{1-\lambda}} \inf_{t \in H(C_p)} \frac{\mathbb{L}_0(t)}{\sqrt{t(1-t)}} \to_{a.s.} 0. \tag{A.33}$$

Hence we assume that there exists $A' \subset \Omega$ such that $P(A') = 1$ and, as $m', n' \to \infty$,

$$\sup_{t \in [p,1-p]} \left| \nu_{m',n'}(t) + \sqrt{\frac{\lambda}{1-\lambda}} \frac{\mathbb{L}_0(t)}{[t(1-t)]^{1/2}} - (1-\lambda)\sqrt{\frac{m'N'}{n'}} \nu(t) \right| \to 0 \tag{A.34}$$

on A', where $N' = m' + n'$. We choose A' so that $\mathbb{L}_0$ has continuous trajectories on A', which Lemma A.3 shows is possible.

The rest of the proof is similar to the proof of part (A) because the asymptotics of the infimum of $\nu_{m',n'}$ over $[p, 1-p]$ are largely governed by its numerator. Indeed, replacing the denominator of (A.10) from part (A) by the denominator of $[t(1-t)]^{1/2}$ for part (B) changes little since this new denominator is also bounded away from 0 on $[p, 1-p]$. Nonetheless, there are some differences, which we detail below.

Fix $\epsilon > 0$. We replace σ_ϵ from the proof of part (A) by σ'_ϵ, where we define σ'_ϵ as follows. If $t, t' \in [p, 1-p]$ satisfy $|t - t'| < \sigma'_\epsilon$, then, on A',

$$\sqrt{\frac{\lambda}{1-\lambda}} \left| \frac{\mathbb{L}_0(t)}{[t(1-t)]^{1/2}} - \frac{\mathbb{L}_0(t')}{(t'(1-t'))^{1/2}} \right| < \epsilon/2. \tag{A.35}$$

Note that since $[p, 1-p]$ is a compact set, and the function $\mathbb{L}_0$ has continuous trajectories on A', the random quantity $\sigma'_\epsilon > 0$ on A'.

Recalling $h = \lambda f + (1-\lambda)g$, we let

$$b_p = \sup_{z \in [H^{-1}(p), H^{-1}(1-p)]} h(z), \tag{A.36}$$

which is clearly positive. Because f and g are continuous, h is also continuous, and therefore $b_p < \infty$. Taking $\tilde{\sigma}_\epsilon = \sigma'_\epsilon / b_p$, and similar to (A.26), defining

$$C_p(z) = \{x \in D_{p'} \ : \ Dist(x, C_{p'}) < z\}, \quad \text{for all } z > 0,$$

we observe that $[p, 1-p]$ can be written as the union of the following three sets:

$$\begin{aligned} \tilde{B}_p(\tilde{\sigma}_\epsilon) &= \{t \in [p, 1-p] \ : \ H^{-1}(t) \in D_p \setminus C_p(\tilde{\sigma}_\epsilon)\}, \\ \tilde{E}_p(\tilde{\sigma}_\epsilon) &= \{t \in [p, 1-p] \ : \ H^{-1}(t) \in C_p(\tilde{\sigma}_\epsilon) \setminus C_p\}, \\ H(C_p) &= \{t \in [p, 1-p] \ : \ H^{-1}(t) \in C_p\}. \end{aligned}$$

Note that, the $D_{p'}$ used in the proof of part (a) of the current theorem is replaced by D_p in the above partitioning. Part (B) differs from part (A) in that $T_{m,n}^{\text{tsep}}(\mathbb{F}_m, \mathbb{G}_n)$ is the infimum of a random quantity, namely $\nu_{m,n}$, over a fixed set $[t, 1-t]$, whereas $T_{m,n}^{\min}(\mathbb{F}_m, \mathbb{G}_n)$ calculates the infimum of $\omega_{m,n}$ over a random set $D_{p,m,n}$. To deal with this randomness, the asymptotics in part (A) were analyzed on a set $D_{p'} \supset D_p$ constructed so as to ensure $D_{p'} \supset D_{p,m,n}$ for sufficiently large m and n almost surely. Since part (B) does not have this additional difficulty, it suffices to study the behavior of $\nu_{m,n}$ on D_p, circumventing the need to consider $D_{p'}$.

We now show that

$$|t - t'| \le \sigma'_\epsilon \quad \text{for all } t \in \tilde{E}_p(\tilde{\sigma}_\epsilon) \text{ and } t' \in \text{bd}(H(C_p)). \tag{A.37}$$

To prove this, we first fix $t \in \tilde{E}_p(\tilde{\sigma}_\epsilon)$ and $t' \in \text{bd}(H(C_p))$. Because H^{-1} is continuous, it holds that $H^{-1}(t') \in \text{bd}(C_p)$. Thus, $|H^{-1}(t) - H^{-1}(t')| \le \tilde{\sigma}_\epsilon$.

Note that the continuity of H also implies that $H(H^{-1}(t_0)) = t_0$ for all $t_0 \in [p, 1-p]$. By the mean value theorem applied to the function H, there exists an a between $H^{-1}(t)$ and $H^{-1}(t')$ such that $t - t' = h(a)[H^{-1}(t) - H^{-1}(t')]$. Hence, $|t - t'| \leq b_p |H^{-1}(t) - H^{-1}(t')|$ where b_p is as defined in (A.36). Combining this display with the fact that $|H^{-1}(t) - H^{-1}(t')| \leq \tilde{\sigma}_\epsilon$ and plugging in $\tilde{\sigma}_\epsilon = \sigma'_\epsilon / b_p$ shows that (A.37) indeed holds.

Recall the definition of $\nu(t)$ from (A.32). Because $\nu(t) \geq 0$, one can show that the infimum of ν over the random set $\tilde{B}_p(\tilde{\sigma}_\epsilon)$ is bounded below by some random number $\delta_{\tilde{\sigma}_\epsilon} > 0$. Therefore, using (A.34) and imitating the proof of (A.28), we can show that

$$\liminf_{m',n' \to \infty} \inf_{t \in \tilde{B}_p(\tilde{\sigma}_\epsilon)} \nu_{m',n'}(t) \to \infty \quad \text{on } A'. \tag{A.38}$$

Next let us consider $t \in \tilde{E}_p(\tilde{\sigma}_\epsilon)$ and $t' \in \mathrm{bd}(H(C_p))$. Note that (A.31) and (A.37) imply that the following holds on A':

$$\left| \nu_{m',n'}(t) - \nu_{m',n'}(t') - (1 - \lambda) \sqrt{\frac{m'N'}{n'}} \nu(t) \right|$$

$$\leq o(1) - \sqrt{\frac{\lambda}{1 - \lambda}} \sup_{t,t' \in S_t} \left| \frac{\mathbb{L}_0(t)}{[t(1-t)]^{1/2}} - \frac{\mathbb{L}_0(t')}{(t'(1-t'))^{1/2}} \right|,$$

where $S_t = \{t, t' \in [p, 1-p] : |t - t'| < \sigma'_\epsilon\}$. Equation A.35 yields that

$$\sqrt{\frac{\lambda}{1 - \lambda}} \sup_{t,t' \in S_t} \left| \frac{\mathbb{L}_0(t)}{[t(1-t)]^{1/2}} - \frac{\mathbb{L}_0(t')}{(t'(1-t'))^{1/2}} \right| < \epsilon/2$$

on A', which indicates that, on this set, for all sufficiently large m' and n',

$$\left| \nu_{m',n'}(t) - \nu_{m',n'}(t') - (1 - \lambda) \sqrt{\frac{m'N'}{n'}} \nu(t) \right| < \epsilon.$$

Because the above holds for any $t \in \tilde{E}_p(\tilde{\sigma}_\epsilon)$ and $t' \in \mathrm{bd}(H(C_p))$, we obtain that

$$\inf_{t \in \tilde{E}_p(\tilde{\sigma}_\epsilon)} \nu_{m',n'}(t) \geq \inf_{t' \in H(C_p)} \nu_{m',n'}(t') + \sqrt{\frac{m'N'}{n'}} \inf_{t \in \tilde{E}_p(\tilde{\sigma}_\epsilon)} \nu(t) - \epsilon.$$

The fact that ν is non-negative on $[p, 1-p]$ yields that $\inf_{t \in \tilde{E}_p(\tilde{\sigma}_\epsilon)} \nu(t) \geq 0$. Thus, the above shows that $\inf_{t \in \tilde{E}_p(\tilde{\sigma}_\epsilon)} \nu_{m',n'}(t) \geq \inf_{t \in H(C_p)} \nu_{m',n'}(t) - \epsilon$. Therefore, using (A.38), we derive that on A', for sufficiently large m' and n',

$$\left| \inf_{t \in [p, 1-p]} \nu_{m',n'}(t) - \inf_{t \in H(C_p)} \nu_{m',n'}(t) \right| \leq \epsilon.$$

As $\epsilon > 0$ was arbitrary, the above shows that on A', $\inf_{t \in [p, 1-p]} \nu_{m',n'}(t) - \inf_{t \in H(C_p)} \nu_{m',n'}(t)$ converges to zero as $m', n' \to \infty$. Finally, (A.17) and (A.34) yield that on A',

$$\limsup_{m',n' \to \infty} \sup_{t \in H(C_p)} \left| \nu_{m',n'}(t) - \sqrt{\frac{\lambda}{1 - \lambda}} \frac{\mathbb{L}_0(t)}{\sqrt{t(1-t)}} \right| = 0.$$

Recall that A' was chosen to satisfy $P(A') = 1$. Thus, the above convergence holds with probability one, from which, (A.33) follows. As was discussed above (A.33), the fact that this equation holds completes the proof of part (B). Part (C) follows from Lemma A.4. □

The proof of Theorem 2 for the minimum t-statistic can be found in Whang (2019) (see also Davidson and Duclos, 2013). However, we still include it here for the sake of completeness.

Proof of Theorem 2

If $(F, G) \in \mathcal{H}_1$, we have $G(x) > F(x)$ for all $x \in D_p$. Since D_p is compact, the continuous function $G - F$ attains its minima at some $x_0 \in D_p$. Therefore, it follows that

$$\inf_{x \in D_p} (G(x) - F(x)) > 3\delta$$

for some $\delta > 0$. We will show that this implies that

$$\liminf_{m,n \to \infty} \inf_{x \in D_{p,m,n}} \left(\mathbb{G}_n(x) - \mathbb{F}_m(x) \right) > \delta \qquad (A.39)$$

with probability one. As a result, $T_{m,n}^{\min} \to_p \infty$ follows because (A.39) indicates that with probability one, for all $x \in D_{p,m,n}$,

$$\frac{\mathbb{G}_n(x) - \mathbb{F}_m(x)}{\sqrt{\dfrac{\mathbb{F}_m(x)(1 - \mathbb{F}_m(x))}{m} + \dfrac{\mathbb{G}_n(x)(1 - \mathbb{G}_n(x))}{n}}}$$
$$\geq \frac{\delta}{\sqrt{\dfrac{\mathbb{F}_m(x)(1 - \mathbb{F}_m(x))}{m} + \dfrac{\mathbb{G}_n(x)(1 - \mathbb{G}_n(x))}{n}}}$$

for all large m and n. However, the right hand side of the last display is bounded below by $\sqrt{mn/N}\delta$. Since $m/N \to \lambda$, it follows that $T_{m,n}^{\min} \to_p \infty$. Hence, it suffices to prove (A.39). To this end, note that, since $\max(f, g) > 0$ on an open neighborhood of D_p, $H = \lambda F + (1 - \lambda)G$ has positive density on this neighborhood. Therefore, H is a continuous and strictly increasing function on this neighborhood. Therefore H^{-1} is continuous and strictly increasing in this neighborhood as well. Hence, we can choose $p' < p$ such that $H^{-1}(p') < H^{-1}(p)$, and

$$\inf_{x \in D_{p'}} (G(x) - F(x)) > 2\delta.$$

For all sufficiently large m and n, $D_{p,m,n} \subset D_{p'}$ almost surely by Fact 3.1. Therefore, with probability one,

$$\inf_{x \in D_{p,m,n}} \left(G(x) - F(x) \right) \geq 2\delta \qquad (A.40)$$

as $m, n \to \infty$. On the other hand, note that

$$\sup_{x \in D_{p,m,n}} \left| \left(\mathbb{G}_n(x) - \mathbb{F}_m(x) \right) - \left(G(x) - F(x) \right) \right|$$
$$\leq \sup_{x \in \mathbb{R}} \left| \mathbb{G}_n(x) - G(x) \right| + \sup_{x \in \mathbb{R}} \left| \mathbb{F}_m(x) - F(x) \right|, \qquad (A.41)$$

which converges to 0 almost surely. Therefore, (A.39) follows, which completes the proof for $T_{m,n}^{\min}$.

For $T_{m,n}^{\text{tsep}}$, note that (A.39) implies

$$T_{m,n}^{\text{tsep}}(\mathbb{F}_m, \mathbb{G}_n) \geq \sqrt{\frac{mn}{N}} \inf_{x \in D_{p,m,n}} \frac{\mathbb{G}_n(x) - \mathbb{F}_m(x)}{\sqrt{t(1-t)}} > \sqrt{\frac{mn}{N}} \delta \inf_{t \in [p, 1-p]} (t(1-t))^{-1/2},$$

which diverges to $+\infty$, thus completing the proof. $\qquad \square$

A.1. Proofs for the shape-constrained test statistics

Before going into the proof for the shape-constrained test statistics, we state and prove a useful lemma.

Lemma A.5. *Suppose that f is a unimodal density satisfying Condition A. Let $\widehat{f}_m$ be the unimodal density estimator of Birgé, based on the independent observations $X_1, \ldots, X_m$ with density f. Here we take $\eta = o(m^{-1})$, where η is the tuning parameter in Section 3. Denote by $\widehat{F}_m$ the distribution function of $\widehat{f}_m$. Further suppose that $\widehat{f}_m^0$ is the Grenander estimator of f based on the true mode M. Then the following assertions hold:*

(A) $\sqrt{m} \int_{\mathbb{R}} |\widehat{f}_m(x) - \widehat{f}_m^0(x)| dx \to_{a.s.} 0$.
(B) $\sqrt{m} \|\widehat{F}_m - \mathbb{F}_m\|_\infty \to_{a.s.} 0$.
(C) $\left\| \sqrt{m}(\widehat{F}_m - F) - \mathbb{V}_1 \circ F \right\|_\infty \to_{a.s.} 0$, where $\mathbb{V}_1$ is as defined in (A.1).

Proof. Suppose that M is the true mode of the density f. In this case F can be written as (Rao, 1969)

$$F = \alpha F^+ + (1-\alpha)F^-,$$

where $\alpha = P_F(X \leq M)$, and F^+ and F^- are the conditional distributions on $(-\infty, M]$ and $[M, \infty)$, respectively, i.e.

$$F^+(x) = \frac{F(x)1[x \leq M]}{F(M)} \quad \text{and} \quad F^-(x) = \frac{\left(F(x) - F(M) \right)1[x > M]}{1 - F(M)}.$$

Let us denote the distribution function of $\widehat{f}_m^0$ by $\widehat{F}_m^0$. From Rao (1969) it follows that $\widehat{F}_m^0$ can be expressed as

$$\widehat{F}_m^0 = \hat{\alpha}_m \widehat{F}_m^{0,+} + (1 - \hat{\alpha}_m)\widehat{F}_m^{0,-} \qquad (A.42)$$

where $\hat{\alpha}_m$ is the sample proportion on $(-\infty.M]$, and $\widehat{F}_m^{0,+}$ and $\widehat{F}_m^{0,-}$ are the monotonoe Grenander estimates of F^+ and F^-, respectively. Denote by $\mathbb{F}_m^+$ and $\mathbb{F}_m^-$, respectively, the empirical distribution functions corresponding to the observations in $(-\infty, M]$ and $[M, \infty)$. Since F is continuous, the probability that $X_i = M$ for some $i \in \{1, \ldots, m\}$ is 0. Hence, there is no ambiguity in the above definition of $\widehat{F}_m^0$. Also, the empirical distribution of the X_i' writes as

$$\mathbb{F}_m = \hat{\alpha}_m \mathbb{F}_m^+ + (1 - \hat{\alpha}_m) \mathbb{F}_m^-.$$

It is well known that under Condition A, the Grenander estimator $\widehat{F}_m^{0,+}$ satisfies $\|\sqrt{m}(\widehat{F}_m^{0,+} - F^+) - \mathbb{V}_1 \circ F^+\|_\infty \to_{a.s.} 0$ and $\sqrt{m}\|\widehat{F}_m^{0,+} - \mathbb{F}_m^+\|_\infty \to_{a.s.} 0$ (see Theorem 2.1 of Beare *et al.* (2017), the original result dates back to Kiefer and Wolfowitz (1976)). Similar results hold for $\widehat{F}_m^{0,-}$ and $\mathbb{F}_m^-$.

Since $\hat{\alpha}_m \to_{a.s.} \alpha$ with probability one, we conclude that

$$\sqrt{m}\|\widehat{F}_m^0 - \mathbb{F}_m\|_\infty \le \hat{\alpha}_m \sqrt{m}\|\widehat{F}_m^{0,-} - \mathbb{F}_m^-\|_\infty + (1-\hat{\alpha}_m)\sqrt{m}\|\widehat{F}_m^{0,+} - \mathbb{F}_m^+\|_\infty \to_{a.s.} 0. \tag{A.43}$$

To prove part (A) of the current lemma, now we invoke Theorem 1 of Birgé, which states that

$$\frac{\sqrt{m}}{2} \int_{\mathbb{R}} |\widehat{f}_m(x) - \widehat{f}_m^0(x)| dx \le \sqrt{m}\eta + \sqrt{m}\|\widehat{F}_m^0 - \mathbb{F}_m\|_\infty,$$

where η is as defined in Section 3, which implies that, in our case, $\eta = O(1/m)$. This, combined with (A.43), proves that the right hand side of the above display approaches 0 almost surely. Thus part (A) of the current lemma is proved.

Now note that since $\sqrt{m}\|\widehat{F}_m - \mathbb{F}_m\|_\infty \le \sqrt{m}\|\widehat{F}_m - \widehat{F}_m^0\|_\infty + \sqrt{m}\|\widehat{F}_m^0 - \mathbb{F}_m\|_\infty$, and

$$\|\widehat{F}_m - \widehat{F}_m^0\|_\infty \le \int_{\mathbb{R}} |\widehat{f}_m(x) - \widehat{f}_m^0(x)| dx,$$

part (B) of the current lemma follows by part (A) and (A.43).

Finally, part (C) of the current lemma follows by noting that

$$\left\|\sqrt{m}(\widehat{F}_m - F) - \mathbb{V}_1 \circ F\right\|_\infty \le \left\|\sqrt{m}(\widehat{F}_m - \mathbb{F}_m)\right\|_\infty + \left\|\sqrt{m}(\mathbb{F}_m - F) - \mathbb{V}_1 \circ F\right\|_\infty,$$

which converges to 0 as $m \to \infty$ by part (B) of the current lemma and (A.1). $\square$

A.1.1. Proof of Lemma 4

Let us define

$$\widehat{\vartheta}_{m,n}(x) = \frac{n}{N}\widehat{F}_m(x)\Big(1 - \widehat{F}_m(x)\Big) + \frac{m}{N}\widehat{G}_n(x)\Big(1 - \widehat{G}_n(x)\Big).$$

Recalling the definition of $\vartheta_{m,n}$ from (A.13), we obtain that

$$T_{m,n}^{\min}(\widehat{F}_m, \widehat{G}_n) - T_{m,n}^{\min}(\mathbb{F}_m, \mathbb{G}_n)$$

$$\leq \left(\frac{mn}{N}\right)^{1/2}\left\{\inf_{x\in D_{p,m,n}}\frac{\widehat{F}_m(x)-\widehat{G}_n(x)}{\widehat{\vartheta}_{m,n}(x)^{1/2}}-\inf_{x\in D_{p,m,n}}\frac{\mathbb{F}_m(x)-\mathbb{G}_n(x)}{\vartheta_{m,n}(x)^{1/2}}\right\}$$

$$\leq \left(\frac{mn}{N}\right)^{1/2}\sup_{x\in D_{p,m,n}}\left|\frac{\widehat{F}_m(x)-\mathbb{F}_m(x)-\left(\widehat{G}_n(x)-\mathbb{G}_n(x)\right)}{\vartheta_{m,n}(x)^{1/2}}\right.$$

$$+(\widehat{F}_m(x)-\widehat{G}_n(x))\left(\widehat{\vartheta}_{m,n}(x)^{-1/2}-\vartheta_{m,n}(x)^{-1/2}\right)\bigg|$$

$$\leq \left(\frac{mn}{N}\right)^{1/2}\left\{\sup_{x\in D_{p,m,n}}|\widehat{F}_m(x)-\mathbb{F}_m(x)|+\sup_{x\in D_{p,m,n}}|\widehat{G}_n(x)-\mathbb{G}_n(x)|\right\}\|\vartheta_{m,n}^{-1/2}\|_{D_{p,m,n}}$$

$$+\left(\frac{mn}{N}\right)^{1/2}\sup_{x\in D_{p,m,n}}\left|\frac{\vartheta_{m,n}(x)^{1/2}-\widehat{\vartheta}_{m,n}(x)^{1/2}}{\vartheta_{m,n}(x)^{1/2}\widehat{\vartheta}_{m,n}(x)^{1/2}}\right|. \tag{A.44}$$

Since f,g are bounded away from 0 on an open interval that includes D_p, we can find $p' < p$ so that f and g are bounded away from 0 on $D_{p'}$. Lemma A.2 then indicates that $\|\vartheta_{m,n}^{-1/2}\|_{D_{p'}}$ is bounded away from 0 with probability one. Since $Dist(D_{p,m,n}, D_p) \to_{a.s.} 0$ by (A.12), $D_{p,m,n} \subset D_{p'}$ almost surely for all sufficiently large m and n. As a result, $\|\vartheta_{m,n}^{-1/2}\|_{D_{p,m,n}}$ is also bounded with probability one.

Therefore, using $m/N \to \lambda$ and Part (B) of Lemma A.5, we conclude that

$$\left(\frac{mn}{N}\right)^{1/2}\sup_{x\in D_{p,m,n}}\left\{|\widehat{F}_m(x)-\mathbb{F}_m(x)|+|\widehat{G}_n(x)-\mathbb{G}_n(x)|\right\}\|\vartheta_{m,n}^{-1/2}\|_{D_{p,m,n}} \to_{a.s.} 0. \tag{A.45}$$

Now, observe that we can write

$$\frac{\vartheta_{m,n}(x)^{1/2}-\widehat{\vartheta}_{m,n}(x)^{1/2}}{\vartheta_{m,n}(x)^{1/2}\widehat{\vartheta}_{m,n}(x)^{1/2}}=\frac{\vartheta_{m,n}(x)-\widehat{\vartheta}_{m,n}(x)}{\vartheta_{m,n}(x)^{1/2}\widehat{\vartheta}_{m,n}(x)^{1/2}\left(\vartheta_{m,n}^{1/2}(x)+\widehat{\vartheta}_{m,n}^{1/2}(x)\right)}.$$

Since $\mathbb{F}_m, \widehat{F}_m, \mathbb{G}_n,$ and $\widehat{G}_n$ take values between 0 and 1, it follows that

$$|\vartheta_{m,n}(x)-\widehat{\vartheta}_{m,n}(x)|\leq\frac{2n}{N}|\widehat{F}_m(x)-\mathbb{F}_m(x)|+\frac{2m}{N}|\widehat{G}_n(x)-\mathbb{G}_n(x)|.$$

Therefore, another application of Part (B) of Lemma A.5 combined with the fact that $m/N \to \lambda$ entails that

$$\left(\frac{mn}{N}\right)^{1/2}\sup_{x\in D_{p,m,n}}|\widehat{\vartheta}_{m,n}(x)-\vartheta_{m,n}(x)| \to_{a.s.} 0.$$

Since $\|\vartheta_{m,n}^{-1/2}\|_{D_{p,m,n}}$ is bounded on $D_{p,m,n}$ almost surely, the above implies that $\|\widehat{\vartheta}_{m,n}^{-1/2}\|_{D_{p,m,n}}$ is also bounded on $D_{p,m.n}$ almost surely. Hence,

$$\left(\frac{mn}{N}\right)^{1/2}\sup_{x\in D_{p,m,n}}\frac{\vartheta_{m,n}(x)-\widehat{\vartheta}_{m,n}(x)}{\vartheta_{m,n}(x)^{1/2}\widehat{\vartheta}_{m,n}(x)^{1/2}\left(\vartheta_{m,n}^{1/2}(x)+\widehat{\vartheta}_{m,n}^{1/2}(x)\right)} \to_{a.s.} 0,$$

which combined with (A.44) and (A.45), implies that

$$T_{m,n}^{\min}(\widehat{F}_m, \widehat{G}_n) - T_{m,n}^{\min}(\mathbb{F}_m, \mathbb{G}_n) \to_{a.s.} 0.$$

Similarly one can show that $T_{m,n}^{\min}(\mathbb{F}_m, \mathbb{G}_n) - T_{m,n}^{\min}(\widehat{F}_m, \widehat{G}_n) \to_{a.s.} 0$, leading to

$$|T_{m,n}^{\min}(\mathbb{F}_m, \mathbb{G}_n) - T_{m,n}^{\min}(\widehat{F}_m, \widehat{G}_n)| \to_{a.s.} 0.$$

Using part (B) of Lemma A.5 and $m/N \to \lambda$ in the second step, we also deduce that

$$|T_{m,n}^{\text{tsep}}(\widehat{F}_m, \widehat{G}_n) - T_{m,n}^{\text{tsep}}(\mathbb{F}_m, \mathbb{G}_n)|$$
$$\leq \left(\frac{mn}{N}\right)^{1/2} \frac{\sup_{x \in \mathbb{R}} |\widehat{F}_m(x) - \mathbb{F}_m(x)| + \sup_{x \in \mathbb{R}} |\widehat{G}_n(x) - \mathbb{G}_n(x)|}{\inf_{z \in [p, 1-p]} \sqrt{z(1-z)}},$$

which converges to zero almost surely.

It remains to prove that $|T_{m,n}^{\text{wrs}}(\widehat{F}_m, \widehat{G}_n) - T_{m,n}^{\text{wrs}}(\mathbb{F}_m, \mathbb{G}_n)|$ converges to 0 almost surely. To this end, we first note that $\xi(F, G) = \int_{\mathbb{R}} F dG$ is Hadamard differentiable with respect to the norm $\|\cdot\|_\infty$ at every pair of distribution functions (F, G) (see Section 5, pages 362–371 Lehmann, 1975), where the derivative at (F, G) is given by

$$\dot{\xi}(F, G; \mu_X, \mu_Y) = \int_{\mathbb{R}} \mu_X dG - \int_{\mathbb{R}} \mu_Y dF,$$

where $\mu_X : \mathbb{R} \mapsto \mathbb{R}$ and $\mu_Y : \mathbb{R} \mapsto \mathbb{R}$ are bounded continuous functions. Observe that we can write

$$\left(\frac{mn}{N+1}\right)^{1/2} \left(\xi(\widehat{F}_m, \widehat{G}_n) - \xi(F, G)\right)$$
$$= \left(\frac{mn}{N(N+1)}\right)^{1/2} \frac{\xi\left(F + N^{-1/2}\widehat{\Delta}_{m,F}, G + N^{-1/2}\widehat{\Delta}_{n,G}\right) - \xi(F, G)}{N^{-1/2}},$$

where

$$\widehat{\Delta}_{m,F} = \sqrt{N}(\widehat{F}_m - F); \quad \widehat{\Delta}_{n,G} = \sqrt{N}(\widehat{G}_n - G).$$

Note that part (C) of Lemma A.5 and the fact that $m/N \to \lambda$ imply that as $m, n \to \infty$,

$$\left\|\widehat{\Delta}_{m,F} - \lambda^{-1/2}\mathbb{V}_1 \circ F\right\|_\infty \to_{a.s.} 0; \quad [\left\|\widehat{\Delta}_{n,G} - (1-\lambda)^{-1/2}\mathbb{V}_2 \circ G\right\|_\infty \to_{a.s.} 0,$$

and $\sqrt{mn/\{N(1+N)\}} \to \sqrt{\lambda(1-\lambda)}$. Therefore, the Hadamard differentiability of ξ implies that

$$\left|\left(\frac{mn}{N(N+1)}\right)^{1/2} \frac{\xi\left(F + N^{-1/2}\widehat{\Delta}_{m,F}, G + N^{-1/2}\widehat{\Delta}_{n,G}\right) - \xi(F, G)}{N^{-1/2}} - \mathbb{Y}\right| \to_{a.s.} 0,$$

where $\mathbb{Y}$ is the random variable $\dot{\xi}(\mu_X, \mu_Y; F, G)$ with $\mu_X = \sqrt{1-\lambda}\mathbb{V}_1 \circ F$, and $\mu_Y = \sqrt{\lambda}\mathbb{V}_2 \circ G$. Thus, we have established

$$\left| \left(\frac{mn}{N+1} \right)^{1/2} \left(\xi(\widehat{F}_m, \widehat{G}_n) - \xi(F, G) \right) - \mathbb{Y} \right| \to_{a.s.} 0.$$

Similarly using (A.1), (A.2), and $m/N \to \lambda$, one can show that

$$\left| \left(\frac{mn}{N+1} \right)^{1/2} \left(\xi(\mathbb{F}_m, \mathbb{G}_n) - \xi(F, G) \right) - \mathbb{Y} \right| \to_{a.s.} 0.$$

Then the proof for $T_{m,n}^{\mathrm{wrs}}(\widehat{F}_m, \widehat{G}_n)$ follows noting

$$12^{-1/2}|T_{m,n}^{\mathrm{wrs}}(\widehat{F}_m, \widehat{G}_n) - T_{m,n}^{\mathrm{wrs}}(\mathbb{F}_m, \mathbb{G}_n)|$$

$$\leq \left| \left(\frac{mn}{N+1} \right)^{1/2} \left(\xi(\widehat{F}_m, \widehat{G}_n) - \xi(F, G) \right) - \mathbb{Y} \right|$$

$$+ \left| \left(\frac{mn}{N+1} \right)^{1/2} \left(\xi(\mathbb{F}_m, \mathbb{G}_n) - \xi(F, G) \right) - \mathbb{Y} \right|,$$

which converges to 0 almost surely. □

A.1.2. Proof of Lemma 5

We can find $p' < p$ such that f and g are positive on $D_{p'}$. Theorem 4.4 of Dümbgen and Rufibach, Condition B1, and Condition B2 imply that

$$\sup_{z \in D_{p'}} |\tilde{F}_m(z) - \mathbb{F}_m(z)| = o_p(m^{-1/2}),$$

and

$$\sup_{z \in D_{p'}} |\tilde{G}_n(z) - \mathbb{G}_n(z)| = o_p(n^{-1/2}).$$

Recall the set $D_{p,m,n}$ defined in (4.1). We note that, for sufficiently large N, $D_{p,m,n} \subset D_{p'}$ with probability one, indicating

$$\sup_{u \in D_{p,m,n}} |\tilde{F}_m(u) - \mathbb{F}_m(u)| + \sup_{x \in D_{p,m,n}} |\tilde{G}_n(x) - \mathbb{G}_n(x)| = o_p(m^{-1/2}) + o_p(n^{-1/2}),$$

$$\text{(A.46)}$$

which is $o_p(N^{-1/2})$ since $m/N \to \lambda$. This result is similar to Lemma A.5(B) for unimodal densities, which is critical to proving the asymptotic equivalence between $T_{m,n}^{\mathrm{min}}(\widehat{F}_m, \widehat{G}_n)$ and $T_{m,n}^{\mathrm{min}}(\mathbb{F}_m, \mathbb{G}_n)$, and between $T_{m,n}^{\mathrm{tsep}}(\widehat{F}_m, \widehat{G}_n)$ and $T_{m,n}^{\mathrm{tsep}}(\mathbb{F}_m, \mathbb{G}_n)$ in Lemma 4. As a consequence, the rest of the proof will be nearly identical to the proof of Lemma 4. Hence, we only highlight the main steps of the proof.

Let us define

$$\tilde{\vartheta}_{m,n}(x) = \frac{n}{N}\tilde{F}_m(x)\left(1 - \tilde{F}_m(x)\right) + \frac{m}{N}\tilde{G}_n(x)\left(1 - \tilde{G}_n(x)\right).$$

Recalling the definition of $\vartheta_{m,n}$ from (A.13), and proceeding like the proof of Lemma 4, we can prove a log-concave analogue of (A.44), that is

$$
\begin{aligned}
& T_{m,n}^{\min}(\tilde{F}_m, \tilde{G}_n) - T_{m,n}^{\min}(\mathbb{F}_m, \mathbb{G}_n) \\
& \leq \left(\frac{mn}{N}\right)^{1/2} \left\{ \inf_{x \in D_{p,m,n}} \frac{\tilde{F}_m(x) - \tilde{G}_n(x)}{\tilde{\vartheta}_{m,n}(x)^{1/2}} - \inf_{x \in D_{p,m,n}} \frac{\tilde{F}_m(x) - \tilde{G}_n(x)}{\vartheta_{m,n}(x)^{1/2}} \right\} \\
& \leq \left(\frac{mn}{N}\right)^{1/2} \left\{ \sup_{x \in D_{p,m,n}} |\tilde{F}_m(x) - \mathbb{F}_m(x)| + \sup_{x \in D_{p,m,n}} |\tilde{G}_n(x) - \mathbb{G}_n(x)| \right\} \|\vartheta_{m,n}^{-1/2}\|_{D_{p,m,n}} \\
& \quad + \left(\frac{mn}{N}\right)^{1/2} \sup_{x \in D_{p,m,n}} \left| \frac{\vartheta_{m,n}(x)^{1/2} - \tilde{\vartheta}_{m,n}(x)^{1/2}}{\vartheta_{m,n}(x)^{1/2} \tilde{\vartheta}_{m,n}(x)^{1/2}} \right|.
\end{aligned}
$$

Since f, g are bounded away from 0 on an open interval that includes D_p, using Lemma A.2, we can show that $\|\vartheta_{m,n}^{-1/2}\|_{D_{p,m,n}}$ is bounded with probability one. Since $m/N \to \lambda$, by (A.46) it follows that

$$
\left(\frac{mn}{N}\right)^{1/2} \sup_{x \in D_{p,m,n}} \left\{ |\tilde{F}_m(x) - \mathbb{F}_m(x)| + |\tilde{G}_n(x) - \mathbb{G}_n(x)| \right\} \|\vartheta_{m,n}^{-1/2}\|_{D_{p,m,n}} = o_p(1).
$$

Note that

$$
\frac{\vartheta_{m,n}(x)^{1/2} - \tilde{\vartheta}_{m,n}(x)^{1/2}}{\vartheta_{m,n}(x)^{1/2} \tilde{\vartheta}_{m,n}(x)^{1/2}} = \frac{\vartheta_{m,n}(x) - \tilde{\vartheta}_{m,n}(x)}{\vartheta_{m,n}(x)^{1/2} \tilde{\vartheta}_{m,n}(x)^{1/2} \left(\vartheta_{m,n}^{1/2}(x) + \tilde{\vartheta}_{m,n}^{1/2}(x) \right)}.
$$

Analogous to the proof of lemma 4, using (A.46), we can show that

$$
\left(\frac{mn}{N}\right)^{1/2} \sup_{x \in D_{p,m,n}} |\tilde{\vartheta}_{m,n}(x) - \vartheta_{m,n}(x)| = o_p(1).
$$

Since $\|\vartheta_{m,n}^{-1/2}\|_{D_{p,m,n}}$ is bounded almost surely, the above implies that $\|\tilde{\vartheta}_{m,n}^{-1/2}\|_{D_{p,m,n}}$ is also bounded almost surely. Therefore another application of (A.46) yields that

$$
T_{m,n}^{\min}(\tilde{F}_m, \tilde{G}_n) - T_{m,n}^{\min}(\mathbb{F}_m, \mathbb{G}_n) = o_p(1).
$$

Similarly one can show that $T_{m,n}^{\min}(\mathbb{F}_m, \mathbb{G}_n) - T_{m,n}^{\min}(\tilde{F}_m, \tilde{G}_n)$ is $o_p(1)$, which implies $|T_{m,n}^{\min}(\tilde{F}_m, \tilde{G}_n) - T_{m,n}^{\min}(\mathbb{F}_m, \mathbb{G}_n)|$ converges to 0 in probability. The proof of

$$
|T_{m,n}^{\text{tsep}}(\tilde{F}_m, \tilde{G}_n) - T_{m,n}^{\text{tsep}}(\mathbb{F}_m, \mathbb{G}_n)| \to_p 0
$$

is analogous to the proof of $|T_{m,n}^{\text{tsep}}(\widehat{F}_m, \widehat{G}_n) - T_{m,n}^{\text{tsep}}(\mathbb{F}_m, \mathbb{G}_n)| = o_p(1)$ in Lemma 4. $\square$

Appendix B: Proofs for Section 5

Although the aim of the current section is to derive the asymptotic distribution of $\mathcal{D}^2(\widehat{f}_m, \widehat{g}_n)$, we will prove a more general result on plug-in estimators of integrated functionals, which may be of independent interest. Theorem 3 then follows as a special case.

We keep using the notations and terminologies developed in Appendix A. Recall that we defined the set of all densities on $\mathbb{R}$ by $\mathcal{P}$. Let $\mathcal{P}_1 \subset \mathcal{P}$. Suppose that $T : \mathcal{P}_1^2 \mapsto \mathbb{R}$ is a functional of the form

$$T(f,g) = \int_{\mathbb{R}} v\Big(f(x), g(x)\Big) dx, \tag{B.1}$$

where $v : \mathbb{R}^2 \mapsto \mathbb{R}$ is a known function. In our case, $T(f,g)$ equals $H(f,g)^2$, leading to

$$v\Big(f(x), g(x)\Big) = 2^{-1}\Big(\sqrt{f(x)} - \sqrt{g(x)}\Big)^2.$$

Now we provide a brief background on a needed concept, namely on influence functions. Define the set of all densities on $\mathbb{R}$ by $\mathcal{P}$. Consider a functional $T : \mathcal{P}^2 \mapsto \mathbb{R}$. Suppose that f and g belong to $\mathcal{P}$, and denote the corresponding distribution functions by F and G, respectively. Suppose the functions $x \mapsto \psi_f(x; f, g)$ and $x \mapsto \psi_g(x; f, g)$ satisfy the following display for all f_1 and g_1 in $\mathcal{P}$:

$$\frac{\partial}{\partial t} T(F + t(F_1 - F), G)\Big|_{t=0} = \int_{\mathbb{R}} \psi_f(x; f, g) f_1(x) dx, \tag{B.2}$$

$$\frac{\partial}{\partial t} T(F, G + t(G_1 - G))\Big|_{t=0} = \int_{\mathbb{R}} \psi_g(x; f, g) g_1(x) dx, \tag{B.3}$$

where above F_1 and G_1 represent the cumulative distribution functions corresponding to f_1 and g_1. Then ψ_f and ψ_g represent the influence functions of T (respectively F- and G-almost surely unique) under the nonparametric model (Van der Vaart, 1998, p. 292). When $T(f,g)$ equals the Hellinger distance $\mathcal{D}^2(f,g)$, it follows that

$$\psi_f(x; f, g) = 2^{-1}\left(1 - \sqrt{\frac{g(x)}{f(x)}} - \mathcal{D}^2(f,g)\right) 1_{\text{supp}(f)}(x), \tag{B.4}$$

$$\psi_g(x; f, g) = 2^{-1}\left(1 - \sqrt{\frac{f(x)}{g(x)}} - \mathcal{D}^2(f,g)\right) 1_{\text{supp}(g)}(x). \tag{B.5}$$

We have already mentioned in Section 5 that the Von Mises Expansion (VME) plays a critical role in the proofs of this section. We define the first order VME of T in the same lines as Kandasamy *et al.* (2015). Suppose that T is Gateaûx differentiable, and the corresponding influence functions ψ_f and ψ_g (see (B.2)

in Section 5) exist. Then we say that $T : \mathcal{P}_1^2 \mapsto \mathbb{R}$ has a first order VME if it satisfies the following for all $f_1, f_2, g_1, g_2 \in \mathcal{P}_1$:

$$T(f_2, g_2) = T(f_1, g_1) + \int_{\mathbb{R}} \psi_f(x; f_1, g_1) f_2(x) dx + \int_{\mathbb{R}} \psi_g(x; f_1, g_1) g_2(x) dx$$
$$+ O(\|f_1 - f_2\|_2^2) + O(\|g_1 - g_2\|_2^2). \tag{B.6}$$

The first order VME implies that T can be written as a linear term plus second order bias term, i.e. T is sufficiently smooth. Kandasamy *et al.* (2015) gives examples of many T which has first order VME.

Let $\overline{f}_m$ and $\overline{g}_n$ be estimators of f and g based on samples of size m and n, respectively. We denote the corresponding distribution functions by $\overline{F}_m$ and $\overline{G}_n$. We aim to show that under some regularity conditions, the plug-in estimator $T(\overline{f}_m, \overline{g}_n)$ is $\sqrt{N}$-consistent for estimating $T(f, g)$.

The first condition we require is related to the weak convergence of the processes $\sqrt{m}(\overline{F}_m - F)$ and $\sqrt{n}(\overline{G}_n - G)$ to Brownian processes.

Condition C1. The distribution functions $\overline{F}_m$ and $\overline{G}_n$ corresponding to density estimators $\overline{f}_m$ and $\overline{g}_n$ satisfy $\sqrt{m}(\overline{F}_m - F) \to_d \mathbb{V}_1(F)$ and $\sqrt{n}(\overline{G}_n - G) \to_d \mathbb{V}_2(G)$, where $\mathbb{V}_1$ and $\mathbb{V}_2$ are Brownian bridges.

The second condition involves the order of the L_2 error in estimating f and g. In particular, we require $\|\overline{f}_m - f\|_2^2$ and $\|\overline{g}_n - g\|_2^2$ to be of order $o_p(m^{-1/2})$ and $o_p(n^{-1/2})$, respectively.

Condition C2. The density estimators $\overline{f}_m$ and $\overline{g}_n$ of f and g satisfy

$$O_p(\|\overline{f}_m - f\|_2^2) = o_p(m^{-1/2}) \quad \text{and} \quad O_p(\|\overline{g}_n - g\|_2^2) = o_p(n^{-1/2}). \tag{B.7}$$

If the model is correctly specified, and f is bounded, many density estimators $\overline{f}_m$ are also bounded with high probability, leading to

$$O_p(\|\overline{f}_m - f\|_2^2) = \mathcal{D}^2(\overline{f}_m, f) O_p(1).$$

Note that if $\overline{f}_m$ also satisfies $\mathcal{D}^2(\overline{f}_m, f) = o_p(m^{-1/2})$, Condition C2 follows.

Our next condition requires the influence functions $\psi_f(\cdot; f, g)$ and $\psi_g(\cdot; f, g)$ to be of bounded total variation on $\mathbb{R}$. We say a function $\mu : \mathbb{R} \mapsto \mathbb{R}$ is of bounded total variation on $\mathbb{R}$, if there exists a generalized derivative (in the sense of distribution) μ' of μ (cf. Section 3.2 of Ambrosio *et al.*, 2000) so that $\int_{\mathbb{R}} |\mu'(x)| dx < \infty$. If μ is of bounded total variation on $\mathbb{R}$, then μ is also of bounded variation on $\mathbb{R}$.

Condition I. The maps $x \mapsto \psi_f(x; f, g)$ and $x \mapsto \psi_g(x; f, g)$ are of bounded total variation.

We are now ready to state the main theorem of this section, which gives the asymptotic distribution of $T(\overline{f}_m, \overline{g}_n)$ under the above-stated conditions. Later we will show that when $T(f, g) = \mathcal{D}^2(f, g)$, the conditions are satisfied. Thus Theorem 3 will follow as a corollary to Theorem 4. Related literature (cf.

Kandasamy *et al.*, 2015) implies that the asymptotic variance of $T(\overline{f}_m, \overline{g}_n)$ as given by Theorem 4 agrees with the asymptotic lower bound for this case under the nonparametric model.

Theorem 4. *Suppose $\mathcal{P}_1 \subset \mathcal{P}$. Let $T : \mathcal{P}_1^2 \mapsto \mathbb{R}$ be a functional of the form* (B.1) *satisfying the first order VME in* (B.6)*. Consider $f, g \in \mathcal{P}_1$. We assume that the influence functions ψ_f and ψ_g defined in* (B.2) *satisfy Condition I. Let $\overline{f}_m$ and $\overline{g}_n$ be estimators of f and g based on two samples of size m and n, respectively, where m and n satisfy $m/N \to \lambda$. Let us denote $N = m + n$. Further suppose $\overline{f}_m, \overline{g}_n \in \mathcal{P}_1$ satisfy Conditions C1 and C2. Then we have*

$$\sqrt{N}\left(T(\overline{f}_m, \overline{g}_n) - T(f, g) \right) \to_d N(0, \sigma_{f,g}^2),$$

where

$$\sigma_{f,g}^2 = \lambda^{-1} \int_{\mathbb{R}} \psi_f(x; f, g)^2 f(x)dx + (1 - \lambda)^{-1} \int_{\mathbb{R}} \psi_g(x; f, g)^2 g(x)dx.$$

Proof. Since T satisfies the first order VME, (B.6) indicates that

$$
\begin{aligned}
T(\overline{f}_m, \overline{g}_n) &- T(f, g) \\
&= \int_{\mathbb{R}} \psi_f(x; f, g)\overline{f}_m(x)dx + \int_{\mathbb{R}} \psi_g(x; f, g)\overline{g}_n(x)dx \\
&\quad + O_p(\|\overline{f}_m - f\|_2^2) + O_p(\|\overline{g}_n - g\|_2^2) \\
&= \int_{\mathbb{R}} \psi_f(x; f, g)\overline{f}_m(x)dx + \int_{\mathbb{R}} \psi_g(x; f, g)\overline{g}_n(x)dx + o_p(N^{-1/2}),
\end{aligned}
$$

where the last step follows from Condition C2. Denote by F, $\overline{F}_m$, G, and $\overline{G}_n$ the distribution functions corresponding to f, $\overline{f}_m$, g, and $\overline{g}_n$, respectively. Since $\psi_f(x; f, g)$ is an influence function with respect to f, it satisfies

$$\int_{\mathbb{R}} \psi_f(x; f, g) f(x)dx = 0.$$

Hence we can write

$$\int_{\mathbb{R}} \psi_f(x; f, g)\overline{f}_m(x)dx = \int_{\mathbb{R}} \psi_f(x; f, g)d(\overline{F}_m(x) - F(x)).$$

Now note that $\psi_f(\cdot; f, g)$ is of bounded total variation on $\mathbb{R}$ by Condition I. Therefore, integration by parts yields that

$$
\begin{aligned}
\int_{\mathbb{R}} &\psi_f(x; f, g)d(\overline{F}_m(x) - F(x)) \\
&= \psi_f(x; f, g)(\overline{F}_m(x) - F(x))\Big|_{-\infty}^{\infty} - \int_{\mathbb{R}} (\overline{F}_m(x) - F(x))d\psi_f(x; f, g).
\end{aligned}
$$

The Riemann-Stieltjes integral in the second term on the right hand side of the last display exists because ψ_f is of bounded total variation and $\overline{F}_m - F$ is continuous. Since $\psi_f(\cdot; f, g)$ is of bounded total variation, it is also bounded, leading to

$$\lim_{x \to \pm\infty} \psi_f(x; f, g)(\overline{F}_m(x) - F(x)) = 0.$$

Therefore, we deduce that

$$\int_{\mathbb{R}} \psi_f(x; f, g)\overline{f}_m(x)dx = -\int_{\mathbb{R}} (\overline{F}_m(x) - F(x))d\psi_f(x; f, g).$$

Similarly we can show that

$$\int_{\mathbb{R}} \psi_g(x; f, g)\overline{g}_n(x)dx = -\int_{\mathbb{R}} (\overline{G}_n(x) - G(x))d\psi_g(x; f, g).$$

Since $\overline{F}_m$ and $\overline{G}_n$ satisfy Condition C1, it follows that

$$\left(\sqrt{m}(\overline{F}_m - F), \sqrt{n}(\overline{G}_n - G) \right) \to_d \left(\mathbb{V}_1(F), \mathbb{V}_2(G) \right),$$

where $\mathbb{V}_1$ and $\mathbb{V}_2$ are independent standard Brownian bridges. Here the underlying metric space corresponding to the weak convergence is $(l^\infty, \|\cdot\|_\infty) \times (l^\infty, \|\cdot\|_\infty)$, where l^∞ was defined to be the set of all bounded functions on $\mathbb{R}$. Since $m/N \to \lambda$, Slutsky's Theorem yields

$$\left(\sqrt{N}(\overline{F}_m - F), \sqrt{N}(\overline{G}_n - G) \right) \to_d \left(\lambda^{-1/2}\mathbb{V}_1(F), (1-\lambda)^{-1/2}\mathbb{V}_2(G) \right).$$

Since Condition I holds, it follows that, for $\mu_1, \mu_2 \in l^\infty$, the map

$$(\mu_1, \mu_2) \mapsto \int_{\mathbb{R}} \mu_1(x)\, d\psi_f(x; f, g) + \int_{\mathbb{R}} \mu_2(x)\, d\psi_g(x; f, g)$$

is continuous with respect to the uniform metric $\|\cdot\|_\infty$. Therefore, invoking the continuous mapping theorem we obtain that

$$\int_{\mathbb{R}} \sqrt{N}(\overline{F}_m(x) - F(x))d\psi_f(x; f, g) + \int_{\mathbb{R}} \sqrt{N}(\overline{G}_n(x) - G(x))d\psi_g(x; f, g)$$

$$\to_d \lambda^{-1/2} \int_{\mathbb{R}} \mathbb{V}_1(F(x))d\psi_f(x; f, g) + (1-\lambda)^{-1/2} \int_{\mathbb{R}} \mathbb{V}_2(G(x))d\psi_g(x; f, g).$$

Now for any continuous distribution function F, any Brownian bridge $\mathbb{V}$, and any function μ with finite total variation, the random variable

$$\mathbb{Y} = \int_{\mathbb{R}} \mathbb{V}(F(x))d\mu(x) \sim N(0, \sigma_\mu^2),$$

where

$$\sigma_\mu^2 = \int_{\mathbb{R}} \mu(x)^2 f(x) - \left(\int_{\mathbb{R}} \mu(x)f(x)dx \right)^2.$$

The above follows from the proof of Theorem 2.3 of Mukherjee *et al.* (2019). Therefore,

$$\lambda^{-1/2} \int_{\mathbb{R}} \mathbb{V}_1(F) d\psi_f(x; f, g) + (1 - \lambda)^{-1/2} \int_{\mathbb{R}} \mathbb{V}_2(G) d\psi_g(x; f, g),$$

which is distributed as a Gaussian random variable with variance $\sigma_{f,g}^2$, thus completing the proof. □

Now we focus on the special case at hand, i.e. $T(f, g) = \mathcal{D}^2(f, g)$. Towards this end, our first task is to show the existence of the first order VME. We take $\mathcal{P}_1$ to be $\mathcal{P}(b, B)$, where f $\mathcal{P}(b, B)$ is as defined in (5.2).

Lemma B.1. *Let* $0 < b < B < \infty$. *Define the map* $T : \mathcal{P}(b, B)^2 \mapsto \mathbb{R}$ *by*

$$T(f, g) = \mathcal{D}^2(f, g).$$

Then the first order VME in (B.6) *holds for* T *for any* $b, B > 0$.

Proof. Follows from Lemma 10 of Kandasamy *et al.* (2015). □

Recall that we defined $\widehat{f}_m^0$ to be the Grenander estimator of f based on the true mode of f. Denote by $\widehat{g}_n^0$ the the Grenander estimator of g based on the true mode of g. Our next step is to obtain the asymptotic distribution of $\mathcal{D}^2(\widehat{f}_m^0, \widehat{g}_n^0)$.

Corollary 1. *Let* f *and* g *be continuous unimodal densities in* $\mathcal{P}(b, B)$ *for some* $b, B > 0$. *Suppose* f *and* g *satisfy condition* A. *We let* $\widehat{f}_m^0$ *and* $\widehat{g}_n^0$ *be the Grenander estimators of* f *and* g *based on the true modes, constructed from samples of size* m *and* n, *respectively. Suppose* m *and* n *satisfy* $m/N \to \lambda$. *Then*

$$\sqrt{N}(\mathcal{D}^2(\widehat{f}_m^0, \widehat{g}_n^0) - \mathcal{D}^2(f, g)) \to_d N(0, \sigma_{f,g}^2),$$

where $\sigma_{f,g}^2$ *is as in* (5.1).

Proof. First we will show that the conditions of Theorem 4 are satisfied. Then we will show that the $\sigma_{f,g}$ of Theorem 4 takes the form of (5.1) when $T(f, g) = \mathcal{D}^2(f, g)$. Suppose M is the mode of f. Since $f(M) < B$, and $\widehat{f}_m^0$ satisfies (A.42), the behavior of $\widehat{f}_m^0$ at M is similar to that of the Grenander estimator of a monotone density at its maxima. Therefore, using Corollary 1.2(i) of Balabdaoui *et al.* (2009) (see also Woodroofe and Sun, 1993), we obtain that $\widehat{f}_m^0(M) \to_d f(M)/\mathbb{U}$, where $\mathbb{U} \sim \text{Uniform}(0, 1)$, which implies $\sup_{x \in \mathbb{R}} \widehat{f}_m^0(x)$ is $O_p(1)$. On the other hand, with probability one, $\widehat{f}_m^0$ converges to f uniformly over any interval of the from $[M + c, \infty)$ or $(-\infty, M - c]$ where $c > 0$ (cf. Balabdaoui *et al.*, 2009). Therefore it can be shown that

$$P\left(\liminf_n \inf_{x \in \text{supp}(\widehat{f}_m^0)} \widehat{f}_m^0(x) > b/2 \right) = 1. \tag{B.8}$$

Similar results hold for $\widehat{g}_n^0$ as well. Thus given any $\epsilon > 0$, we can find $B_\epsilon > B$ so that

$$\liminf_{m \to \infty} P\left(\widehat{f}_m^0, \widehat{g}_n^0 \in \mathcal{P}(b/2, B_\epsilon) \right) > 1 - \epsilon.$$

Since $\mathcal{P}(b, B) \subset \mathcal{P}(b/2, B_\epsilon)$, $f, g \in \mathcal{P}(b/2, B_\epsilon)$ as well. Thus it suffices to show that the conditions of Theorem 4 are satisfied when $\widehat{f}_m^0, \widehat{g}_n^0 \in \mathcal{P}_1 \equiv \mathcal{P}(b/2, B_\epsilon)$.

Notice that Lemma B.1 implies that the first order VME holds for the functional $\mathcal{D}^2 : \mathcal{P}_1 \mapsto \mathbb{R}$ when $\mathcal{P}_1 = \mathcal{P}(b/2, B_\epsilon)$. Condition I also follows in a straightforward way once we note that, when $f, g \in \mathcal{P}(b/2, B_\epsilon)$, (B.4) and (B.5) indicate that $\psi_f(\cdot; f, g)$ and $\psi_g(\cdot; f, g)$ are differentiable functions with integrable derivatives. Condition C1 follows from (A.1), (A.2), and (A.43).

It remains to verify only Condition C2, which we will do only for $\widehat{f}_m$, because the calculations for $\widehat{g}_n$ will be identical. Observe that

$$\|\widehat{f}_m^0 - f\|_2^2 \lesssim \left\|\left(\sqrt{\widehat{f}_m^0} + \sqrt{f}\right)^2\right\|_\infty \mathcal{D}^2(\widehat{f}_m^0, f) \lesssim \left(\|\widehat{f}_m^0\|_\infty + \|f\|_\infty\right)\mathcal{D}^2(\widehat{f}_m^0, f).$$

Using Theorem 7.12 of van de Geer (2000) one can show that $\mathcal{D}^2(\widehat{f}_m^0, f) = O_p(n^{-2/3})$, and we have already established that $\|\widehat{f}_m^0\|_\infty = O_p(1)$. Thus Condition C2 also follows. Now the proof will follow if we can show that

$$\frac{1}{\lambda}\int_{-\infty}^\infty \psi_f(x; f, g)^2 f(x)dx + \frac{1}{1-\lambda}\int_{-\infty}^\infty \psi_g(y; f, g)^2 g(x)dx = \frac{2\mathcal{D}^2(f, g) - \mathcal{D}^4(f, g)}{4\lambda(1-\lambda)},$$
$$\text{(B.9)}$$

where ψ_f and ψ_g are as in (B.2). To that end, note that

$$\mathcal{D}^2(f, g) = \frac{1}{2}\int_{-\infty}^\infty (\sqrt{f(x)} - \sqrt{g(x)})^2 dx = 1 - \underbrace{\int_{-\infty}^\infty \sqrt{f(x)g(x)}dx}_{\rho(f,g)}.$$

We calculate

$$4\int_{-\infty}^\infty \psi_f(x; f, g)^2 f(x)dx = \int_{-\infty}^\infty \left(1 - \frac{\sqrt{g(x)}}{\sqrt{f(x)}} - \mathcal{D}^2(f, g)\right)^2 f(x)dx$$
$$= \int_{-\infty}^\infty \left(\sqrt{f(x)} - \sqrt{g(x)} - \sqrt{f(x)}\mathcal{D}^2(f, g)\right)^2 dx,$$

which equals

$$\int_{-\infty}^\infty \left(\sqrt{f(x)} - \sqrt{g(x)}\right)^2 dx + \mathcal{D}^4(f, g) - 2\mathcal{D}^2(f, g)\int_{-\infty}^\infty \left(\sqrt{f(x)} - \sqrt{g(x)}\right)\sqrt{f(x)}dx$$
$$= 2\mathcal{D}^2(f, g) + \mathcal{D}^4(f, g) - 2\mathcal{D}^2(f, g)(1 - \rho(f, g))$$
$$= 2\mathcal{D}^2(f, g) + \mathcal{D}^4(f, g) - 2\mathcal{D}^4(f, g)$$

because $1 - \rho(f, g) = \mathcal{D}^2(f, g)$. Therefore,

$$\int_{-\infty}^\infty \psi_f(x; f, g)^2 f(x)dx = \frac{2\mathcal{D}^2(f, g) - \mathcal{D}^4(f, g)}{4}.$$

By symmetry, (B.9) follows, thus completing the proof. $\qquad\square$

Our next lemma establishes that $\mathcal{D}^2(\widehat{f}_m, \widehat{g}_n)$ and $\mathcal{D}^2(\widehat{f}_m^0, \widehat{g}_n^0)$ differ by an $o_p(N^{-1/2})$ term. Observe that Corollary 1 combined with Lemma B.2 implies Theorem 3, and thus establishes the asymptotic distribution of $\mathcal{D}^2(\widehat{f}_m, \widehat{g}_n)$ as well.

Lemma B.2. *Under the set up of Corollary 1,*

$$\left| \mathcal{D}^2(\widehat{f}_m, \widehat{g}_n) - \mathcal{D}^2(\widehat{f}_m^0, \widehat{g}_n^0) \right| = o_p(N^{-1/2}).$$

Proof. Adding and subtracting terms shows that

$$\begin{aligned}
\mathcal{D}^2(\widehat{f}_m, \widehat{g}_n) &- \mathcal{D}^2(\widehat{f}_m^0, \widehat{g}_n^0) \\
&= [\mathcal{D}^2(f, \widehat{g}_n) - \mathcal{D}^2(f, \widehat{g}_n^0)] + [\mathcal{D}^2(\widehat{f}_m, g) - \mathcal{D}^2(\widehat{f}_m^0, g)] \\
&\quad + \left[\mathcal{D}^2(\widehat{f}_m, \widehat{g}_n) - \mathcal{D}^2(\widehat{f}_m, g) - \mathcal{D}^2(f, \widehat{g}_n) + \mathcal{D}^2(f, g) \right] \\
&\quad - [\mathcal{D}^2(\widehat{f}_m^0, \widehat{g}_n^0) - \mathcal{D}^2(\widehat{f}_m^0, g) - \mathcal{D}^2(f, \widehat{g}_n^0) + \mathcal{D}^2(f, g)].
\end{aligned}$$

Now observe that

$$\begin{aligned}
\mathcal{D}^2(\widehat{f}_m^0, \widehat{g}_n^0) - \mathcal{D}^2(\widehat{f}_m^0, g) - \mathcal{D}^2(f, \widehat{g}_n^0) + \mathcal{D}^2(f, g) &= \int_{\mathbb{R}} [\sqrt{\widehat{f}_m^0} - \sqrt{f}][\sqrt{\widehat{g}_n^0} - \sqrt{g}] \\
&\leq 2H(\widehat{f}_m^0, f) H(\widehat{g}_n^0, g), \quad \text{(B.10)}
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{D}^2(\widehat{f}_m, \widehat{g}_n) - \mathcal{D}^2(\widehat{f}_m, g) - \mathcal{D}^2(f, \widehat{g}_n) + \mathcal{D}^2(f, g) &= \int_{\mathbb{R}} [\sqrt{\widehat{f}_m} - \sqrt{f}][\sqrt{\widehat{g}_n} - \sqrt{g}] \\
&\leq 2H(\widehat{f}_m, f) H(\widehat{g}_n, g). \quad \text{(B.11)}
\end{aligned}$$

Since squared Hellinger distance is smaller than the L_1 distance, using Lemma A.5 and the fact that $m/N \to \lambda$, we obtain

$$\mathcal{D}^2(\widehat{f}_m^0, f) \leq \|\widehat{f}_m^0 - f\|_1 = o_p(m^{-1/2}) = o_p(N^{-1/2}).$$

A similar result holds for $\mathcal{D}^2(\widehat{g}_n^0, g)$ as well. Thus it follows that the right hand side of (B.10) is $o_p(N^{-1/2})$. Lemma B.3 in Section B.1 implies $H(\widehat{f}_m^0, \widehat{f}_m)$ and $H(\widehat{f}_m^0, \widehat{f}_m)$ are $o_p(N^{-1/4})$. Therefore, using triangle inequality, we can show that $H(\widehat{f}_m, f)$ and $H(\widehat{g}_n, g)$ are $o_p(N^{-1/4})$ as well, which establishes that the right hand side of (B.11) is also $o_p(N^{-1/2})$.

Therefore, we have shown that

$$\begin{aligned}
\mathcal{D}^2(\widehat{f}_m, \widehat{g}_n) &- \mathcal{D}^2(\widehat{f}_m^0, \widehat{g}_n^0) \\
&= [\mathcal{D}^2(f, \widehat{g}_n) - \mathcal{D}^2(f, \widehat{g}_n^0)] + [\mathcal{D}^2(\widehat{f}_m, g) - \mathcal{D}^2(\widehat{f}_m^0, g)] + o_p(N^{-1/2}).
\end{aligned}$$

Hence, the proof will be complete if we can show that

$$\mathcal{D}^2(f, \widehat{g}_n) - \mathcal{D}^2(f, \widehat{g}_n^0) = o_p(N^{-1/2}),$$

$$\mathcal{D}^2(\widehat{f}_m, g) - \mathcal{D}^2(\widehat{f}_m^0, g) = o_p(N^{-1/2}).$$

We will only prove the first line of the above display because the argument for the second line is similar.

First, we denote $\mathcal{X}_n = \text{supp}(\widehat{g}_n^0) \cup \text{supp}(\widehat{g}_n)$ and $\mathcal{Y}_n = \mathcal{X}_n \setminus \text{supp}(\widehat{g}_n^0) = \text{supp}(\widehat{g}_n) \setminus \text{supp}(\widehat{g}_n^0)$. Also denote $b' = b/2$ and $B' = B + b/2$. Since $\sqrt{\widehat{g}_n} + \sqrt{\widehat{g}_n^0} > 0$ on $\mathcal{X}_n$, it follows that

$$
\begin{aligned}
|\mathcal{D}^2(f, \widehat{g}_n^0) - \mathcal{D}^2(f, \widehat{g}_n)| &= \left| \int_{\mathcal{X}_n} \sqrt{f(x)} (\sqrt{\widehat{g}_n(x)} - \sqrt{\widehat{g}_n^0(x)}) dx \right| \\
&= \left| \int_{\mathcal{X}_n} \frac{\sqrt{f(x)}}{(\sqrt{\widehat{g}_n^0(x)} + \sqrt{\widehat{g}_n(x)})} (\widehat{g}_n^0(x) - \widehat{g}_n(x)) dx \right| \\
&\leq \sup_{x \in \text{supp}(\widehat{g}_n^0)} \frac{\sqrt{f(x)}}{\sqrt{\widehat{g}_n^0(x)}} \|\widehat{g}_n^0 - \widehat{g}_n\|_1 + \sup_{x \in \mathcal{Y}_n} \frac{\sqrt{f(x)}}{\sqrt{\widehat{g}_n(x)}} \|\widehat{g}_n^0 - \widehat{g}_n\|_1
\end{aligned}
$$

where we used the fact that $\mathcal{X}_n = \text{supp}(\widehat{g}_n^0) \cup \mathcal{Y}_n$. Since (a) $f \in \mathcal{P}(b, B)$, (b) $\widehat{g}_n^0(x) > b/2$ for $x \in \text{supp}(\widehat{g}_n^0)$ with probability one by (B.8), and (c) $\|\widehat{g}_n^0 - \widehat{g}_n\|_1$ is $o_p(N^{-1/2})$ by Lemma A.5(A), we have

$$|\mathcal{D}^2(f, \widehat{g}_n^0) - \mathcal{D}^2(f, \widehat{g}_n)| = o_p(N^{-1/2}) O_p\left(\sup_{x \in \mathcal{Y}_n} \frac{1}{\sqrt{\widehat{g}_n(x)}} \right).$$

Hence, it only remains to show that $\sup_{x \in \mathcal{Y}_n} \widehat{g}_n(x)^{-1/2} = O_p(1)$.

Let us denote the mode of $\widehat{g}_n$ by $\widehat{M}_n$. First we show that it suffices to only consider the case when $\widehat{M}_n \notin [Y_{(1)}, Y_{(n)}]$. To that end, we will mainly use the following property of $\mathbb{G}_n$ that follows from (2.7) of Birgé (1997):

$$\widehat{G}_n(x) \leq \mathbb{G}_n(x) \quad \text{for} \quad x \in (-\infty, \widehat{M}_n] \quad \text{and} \quad \mathbb{G}_n(x) \leq \widehat{G}_n(x) \quad \text{for} \quad x \in (\widehat{M}_n, \infty). \tag{B.12}$$

Suppose $\widehat{M}_n \in [Y_{(1)}, Y_{(n)}]$. Then by (B.12), any $x < Y_{(1)}$ and $y > Y_{(n)}$ satisfy $\widehat{G}_n(x) = 0$ and $\widehat{G}_n(y) = 1$, respectively. Therefore, the support of $\widehat{g}_n$ is contained in $[Y_{(1)}, Y_{(n)}]$. However, $[Y_{(1)}, Y_{(n)}] \subset \text{supp}(\widehat{g}_n^0)$, which implies $\mathcal{Y}_n = \emptyset$ in this case. Therefore, we only consider the case when $\widehat{M}_n \notin [Y_{(1)}, Y_{(n)}]$.

First consider the case when $\widehat{M}_n > Y_{(n)}$. By (B.12), any $y \geq \widehat{M}_n$ satisfies $\widehat{G}_n(y) \geq \mathbb{G}_n(y) = 1$, and any $x < Y_{(1)}$ satisfies $\widehat{G}_n(x) \leq \mathbb{G}_n(x) = 0$. Therefore, we have $\text{supp}(\widehat{g}_n) \subset [Y_{(1)}, \widehat{M}_n]$. Hence, $\widehat{G}_n(Y_{(n)}) = \int_{Y_{(1)}}^{Y_{(n)}} \widehat{g}_n(x) dx$. Since $\widehat{g}_n$ is non-decreasing on $[Y_{(1)}, Y_{(n)}]$, we have

$$\widehat{G}_n(Y_{(n)}) \leq (Y_{(n)} - Y_{(1)}) \widehat{g}_n(Y_{(n)}).$$

Note that Because $g(y) > b$, this density has a bounded support, implying $(Y_{(n)} - Y_{(1)}) < \text{Diam}(Y)$ for some $\text{Diam}(Y) > 0$. On the other hand,

$$|\widehat{G}_n(Y_{(n)}) - 1| \leq \|\widehat{G}_n - \mathbb{G}_n\|_\infty = o_p(N^{-1/2})$$

by Lemma A.5(B). Therefore,

$$\widehat{g}_n(Y_{(n)}) \geq \frac{1 - o_p(n^{-1/2})}{\mathrm{Diam}(Y)}.$$

Because $\mathrm{supp}(\widehat{g}_n) \subset [Y_{(1)}, \widehat{M}_n]$ and $\mathrm{supp}(\widehat{g}_n^0) \supset [Y_{(1)}, Y_{(n)}]$, we have $\mathcal{Y}_n \subset [Y_{(n)}, \widehat{M}_n]$, indicating

$$\sup_{x \in \mathcal{Y}_n} \widehat{g}_n(x)^{-1/2} \leq \widehat{g}_n(Y_{(n)})^{-1/2} \leq \frac{\sqrt{\mathrm{Diam}(Y)}}{1 + o_p(n^{-1/2})} = O_p(1).$$

Now suppose $\widehat{M}_n < Y_{(1)}$. Then using (B.12), we deduce that $\widehat{G}_n(Y_{(n)}) \geq \mathbb{G}_n(Y_{(n)}) = 1$ and $\widehat{G}_n(x) \leq \mathbb{G}_n(x) = 0$ for any $x < Y_{(1)}$. Therefore, $\mathrm{supp}(\widehat{g}_n) \subset [\widehat{M}_n, Y_{(n)}]$ and $\mathcal{Y}_n \subset [\widehat{M}_n, Y_{(1)}]$. Because $\widehat{g}_n$ is non-increasing on $\mathcal{Y}_n$ in this case, $\widehat{g}_n(x) > \widehat{g}_n(Y_{(1)})$ for any $x \in \mathcal{Y}_n$. Therefore, if we can show that $\widehat{g}_n(Y_{(1)})$ is bounded away from 0, the rest of the proof will follow similar to the case of $\widehat{M}_n > Y_{(n)}$.

Because $\widehat{g}_n$ is non-increasing on its support,

$$\widehat{g}_n(Y_{(1)})(Y_{(n)} - Y_{(1)}) \geq \widehat{G}_n(Y_{(n)}) - \widehat{G}_n(Y_{(1)}) \overset{(a)}{=} 1 - o_p(n^{-1/2}),$$

where (a) follows from Lemma A.5(B). Since $Y_{(n)} - Y_{(1)} = \mathrm{Diam}(Y) < \infty$, we have $\widehat{g}_n(Y_{(1)})^{-1} = O_p(1)$, which completes the proof. □

Proof of Theorem 3

Theorem 5 follows from Corollary 1 and Lemma B.2. □

Proof of Lemma 6

Theorem 4 of Cule and Samworth implies that $\tilde{f}_m$ uniformly converges to f almost surely provided (i) f has finite first moment, (ii) the support of f has nonempty interior, and (iii) $\int \max\{\log f(x), 0\} f(x) dx < \infty$. For log-concave f, (i) follows from Lemma 1 of Cule and Samworth, (ii) follows from the continuity of f, and (iii) follows because f is bounded (cf. Lemma 5, Cule and Samworth, 2010). The similar result holds for $\tilde{g}_n$ as well. Because uniform convergence implies pointwise convergence, the above implies $\tilde{f}_m \tilde{g}_n$ pointwise converges to fg almost surely. Therefore, an application of Scheffé's Theorem (cf. Theorem 16.12, Billingsley, 2013) yields that as $m, n \to \infty$,

$$\int_{\mathbb{R}} \left(\tilde{f}_m(x) \tilde{g}_n(x) \right)^{1/2} dx \to_{a.s.} \int_{\mathbb{R}} \left(f(x) g(x) \right)^{1/2} dx,$$

which indicates

$$\mathcal{D}^2(\tilde{f}_m, \tilde{g}_n) = 1 - \int_{\mathbb{R}} \left(\tilde{f}_m(x) \tilde{g}_n(x) \right)^{1/2} dx \to_{a.s.} = \mathcal{D}^2(f, g).$$

For the smoothed log-concave MLE, the uniform convergence of $\tilde{f}_m$ to f follows from Theorem 1 of Chen and Samworth provided f has finite second moment, which follows trivially because all moments of a log-concave density are finite (Lemma 1, Cule and Samworth, 2010). The rest of the proof then follows from Scheffé's Theorem as in the case of the log-concave MLE. $\qquad\square$

B.1. Additional lemma

Lemma B.3. *Under the set up of Lemma B.2,*

$$\mathcal{D}^2(\widehat{f}_m, \widehat{f}_m^0) = o_p(m^{-1/2}) \quad and \quad \mathcal{D}^2(\widehat{g}_n, \widehat{g}_n^0) = o_p(n^{-1/2}).$$

Proof. We will prove the result only for $\mathcal{D}^2(\widehat{f}_m, \widehat{f}_m^0)$ because the proof of the other case will be identical.

Let us denote the mode of f and $\widehat{f}_m$ by M and $\widehat{M}_m$, respectively. First we consider the case when $\widehat{M}_m > M$. The proof of Lemma 1 of Birgé entails that, in this case, there exist $a, \beta, c \in \mathbb{R}$ such that

1. $a \leq M$, $\beta \geq \widehat{M}_m$, and $c \in [M, \widehat{M}_m]$.
2. $\widehat{f}_m^0(x) = \widehat{f}_m(x)$ for $x < a$ and $x > \beta$. Therefore, $\widehat{F}_m^0(x) = \widehat{F}_m(x)$ for $x \leq a$ and $x \geq \beta$.
3. $\widehat{f}_m^0(x) \geq \widehat{f}_m(x)$ for $x \in (a, c)$, and $\widehat{f}_m^0(x) \leq \widehat{f}_m(x)$ for $x \in (c, \beta)$.

Using the above relations, and denoting the distribution functions of $\widehat{f}_m$ and $\widehat{f}_m^0$ by $\widehat{F}_m$ and $\widehat{F}_m^0$, respectively, we deduce that

$$\int_{\mathbb{R}} \left(\sqrt{\widehat{f}_m(x)} - \sqrt{\widehat{f}_m^0(x)} \right)^2 dx$$

$$= \int_a^\beta \left(\sqrt{\widehat{f}_m(x)} - \sqrt{\widehat{f}_m^0(x)} \right)^2 dx$$

$$= \int_a^\beta \widehat{f}_m(x)dx + \int_a^\beta \widehat{f}_m^0(x)dx - 2\int_a^\beta \sqrt{\widehat{f}_m(x)\widehat{f}_m^0(x)}dx$$

$$\leq \widehat{F}_m(\beta) - \widehat{F}_m(a) + \widehat{F}_m^0(\beta) - \widehat{F}_m^0(a) - 2\int_a^c \widehat{f}_m(x)dx - 2\int_c^\beta \widehat{f}_m^0(x)dx$$

$$= \widehat{F}_m(\beta) - \widehat{F}_m(a) + \widehat{F}_m^0(\beta) - \widehat{F}_m^0(a) - 2\Big(\widehat{F}_m(c) - \widehat{F}_m(a) + \widehat{F}_m^0(\beta) - \widehat{F}_m^0(c)\Big)$$

which is $2\{\widehat{F}_m^0(c) - \widehat{F}_m(c)\}$ because $\widehat{F}_m^0(x) = \widehat{F}_m(x)$ for $x = a, \beta$. Hence, we observe that

$$\mathcal{D}^2(\widehat{f}_m, \widehat{f}_m^0) \leq 2\|\widehat{F}_m^0 - \widehat{F}_m\|_\infty,$$

which is less than $\eta = 1/N$ by the construction of Birgé's estimator (see Section 3). Hence the proof of Lemma B.2 follows for this case.

Now suppose that $\widehat{M}_m < M$. Then from the proof of Lemma 1 in Birgé, one can prove the existence of $a, \beta, c \in \mathbb{R}$ such that

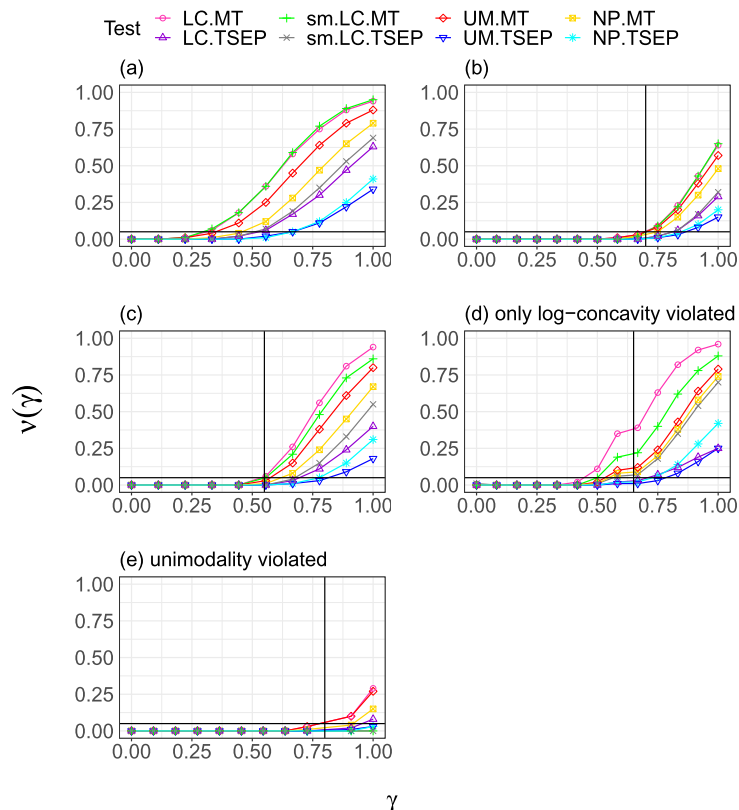1. $a \leq \widehat{M}_m$, $\beta \geq M$, and $c \in [\widehat{M}_m, M]$.

Fig 16: Plot of estimated power $\nu(\gamma)$ vs γ when $m = n = 100$ for simulation schemes (a)-(e); here we use the critical value $C_{m,n}z_\alpha$ for thee TSEP tests. Here MT and TSEP correspond to the minimum t-test and the TSEP test, respectively. The standard deviation of the $\nu(\gamma)$ estimate in each case is less than 0.005. The black horizontal line corresponds to the level of the test, $\alpha = 0.05$. For cases (b)-(e), the black vertical line represents the LFC configuration γ^*, taking value 0.70 (b), 0.55 (c), 0.65 (d), and 0.80 (e).

2. $\widehat{f}_m^0(x) = \widehat{f}_m(x)$ for $x < a$ and $x > \beta$. Therefore, $\widehat{F}_m^0(x) = \widehat{F}_m(x)$ for $x \leq a$ and $x \geq \beta$.
3. $\widehat{f}_m(x) \geq \widehat{f}_m^0(x)$ for $x \in (a, c)$, and $\widehat{f}_m(x) \leq \widehat{f}_m^0(x)$ for $x \in (c, \beta)$.

Then in the same way as in the case of $\widehat{M}_m > M$, we can show that $\mathcal{D}^2(\widehat{f}_m, \widehat{f}_m^0) \leq \|\widehat{F}_m^0 - \widehat{F}_m\|_\infty$, which completes the proof of the current lemma. □

Appendix C: Additional simulations

In this section, we perform simulations on the exact same settings as in Section 4.3, but we use the critical value $C_{m,n}z_\alpha$ for the TSEP tests. We remind

the readers that $C_{m,n}$ was set to be $\max\{\sqrt{N/m}, \sqrt{N/n}\}$. Since in this case $m = n$, we have $N = 2n$, which implies $C_{m,n} = \sqrt{2}$. The power curves are given by Figure 16, which implies that the resulting TSEP tests, which we will refer to as the conservative TSEP tests, have inferior power compared to the minimum t-tests. Moreover, a comparison between Figure 7 and Figure 16 indicates that the power of the conservative TSEP tests is much less compared to that of the ordinary TSEP tests, which use the critical value z_α. However, Figure 16 implies that in case (d), where one distribution is the heavy-tailed Pareto distribution, the conservative TSEP tests succeed to control the type I error at the LFC configuration. All other tests, including the nonparametric minimum t-test and the nonparametric ordinary TSEP test with critical value z_α, have type I error slightly higher than 0.05 at the LFC configuration in case (d); see Figure 7. The nonparametric tests control the type I error at all other cases, however. Also, all TSEP tests control type I error in case (b), where the distributions cross each other at the boundary. To summarize, the conservative TSEP tests might have a slight advantage over the ordinary counterparts in terms of type I error in some boundary cases, but this advantage comes at the cost of a drastic power-loss. In view of the above, we do not recommend the conservative TSEP tests for implementation.

## Appendix D: Additional tables and figures

TABLE 6

Summary of the trial HVTN 097 and trial HVTN 100. By positive respondents, we refer to vaccinees who developed immune response for at least one of the seven clade C V1V2 antigens under consideration.

| Trials	HVTN 097	HVTN 100
Phase	1b	1/2
Site	3 towns in South Africa	6 towns in South Africa
Study design	placebo controlled, randomized, double-blind	placebo controlled, randomized, double-blind
Enrollment	100	252
Vaccinee : Placebo ratio	4:1	4:1
Per protocol vaccinees	73	185
Positive respondents	68	180
Age-range	18-40	18-40
Enrollment period	June-December 2013	February-May 2015
Clade of HIV-1 insert strains used in vaccines	B and E	C
Products used	ALVAC and AIDSVAX	ALVAC and gp120
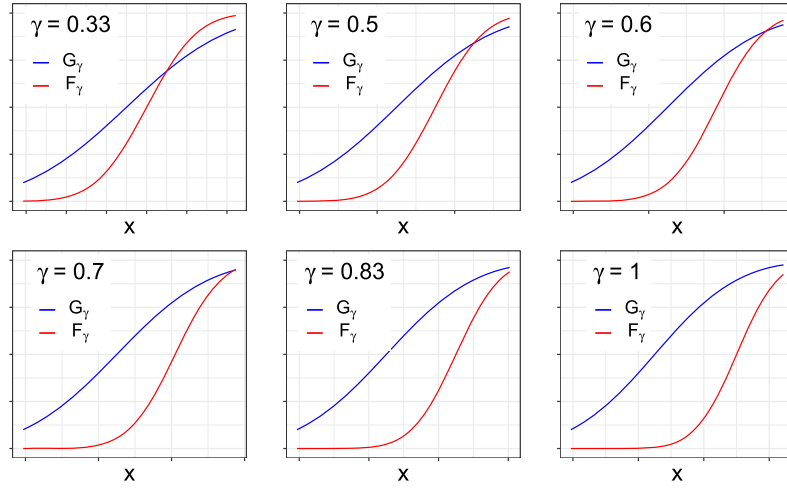
(i) Plots for case (b).
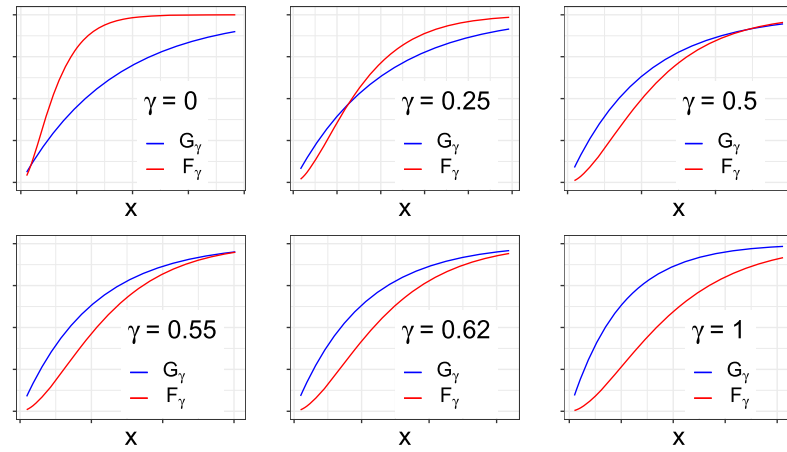
(ii) Plots for case (c).

Fig 17: Plots of the distribution functions F_γ and G_γ for several values of γ in cases (b) and (c). The distribution functions are shown on the region $D_p(F_\gamma, G_\gamma)$.
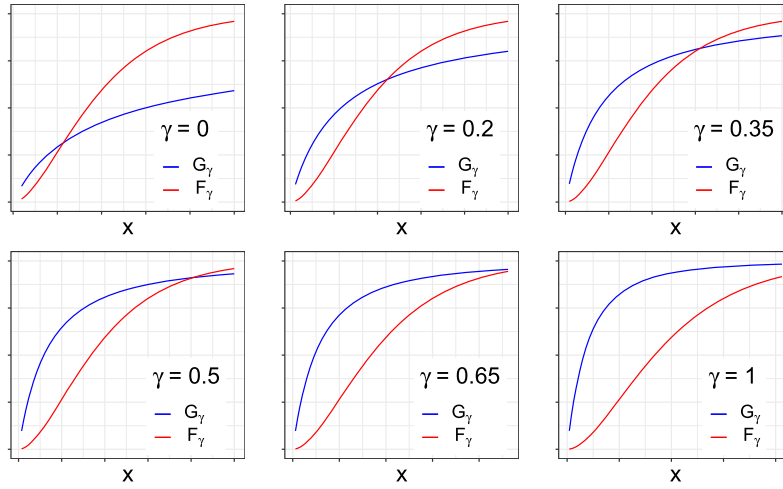
(i) Plots for case (d).
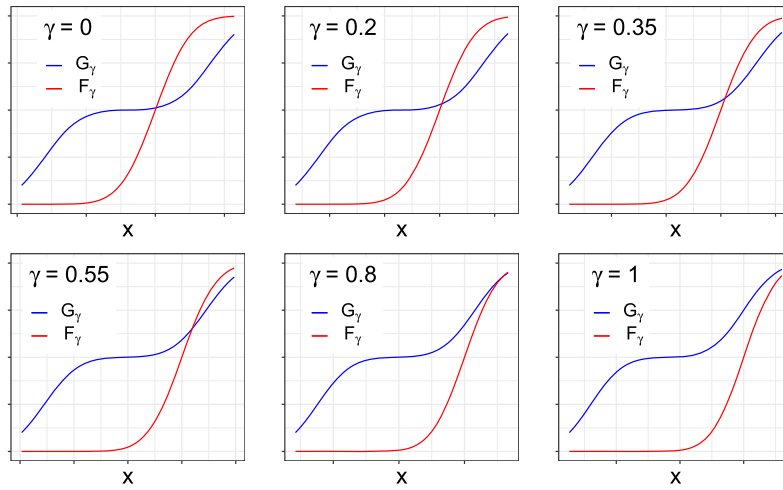
(ii) Plots for case (e).

Fig 18: Plots of the distribution functions F_γ and G_γ for several values of γ in cases (d) and (e). The distribution functions are shown on the region $D_p(F_\gamma, G_\gamma)$.
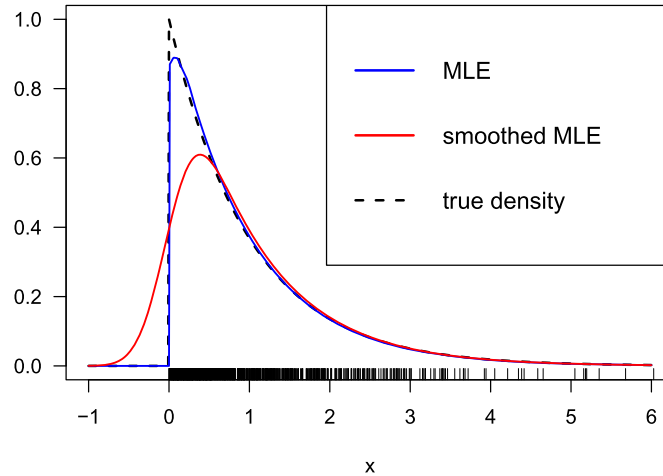
Fig 19: Plot of the log-concave density estimators based on a sample of size 1000 from standard exponential distribution. Observe that the smoothed log-concave MLE does not approximate the true density well near the origin, which is also the point of discontinuity.

Acknowledgments

References

[1] Álvarez-Esteban, P., Del Barrio, E., Cuesta-Albertos, J., and Matrán, C. (2016). A contamination model for the stochastic order. *Test*, **25**, 751–774.

[2] Ambrosio, L., Fusco, N., and Pallara, D. (2000). *Functions of bounded variation and free discontinuity problems*. Oxford Science Publications.

[3] Asmussen, S. and Lehtomaa, J. (2017). Distinguishing log-concavity from heavy tails. *Risks*, **5**(1), 10.

[4] Balabdaoui, F., Rufibach, K., and Wellner, J. A. (2009). Limit distribution theory for maximum likelihood estimation of a log-concave density. *Ann. Statist.*, **37**, 1299.

[5] Beare, B. K., Fang, Z., *et al.* (2017). Weak convergence of the least concave majorant of estimators for a concave distribution function. *Electron. J. Stat.*, **11**, 3841–3870.

[6] Bekker, L.-G., Moodie, Z., Grunenberg, N., Laher, F., Tomaras, G. D., Cohen, K. W., Allen, M., Malahleha, M., Mngadi, K., Daniels, B., *et al.* (2018). Subtype c alvac-hiv and bivalent subtype c gp120/mf59 hiv-1 vac-

cine in low-risk, hiv-uninfected, south african adults: a phase 1/2 trial. *The Lancet HIV*, **5**(7), e366–e378.

[7] Billingsley, P. (2013). *Convergence of probability measures*. John Wiley & Sons.

[8] Birgé, L. (1997). Estimation of unimodal densities without smoothness assumptions. *Ann. Statist.*, **25**, 970–981.

[9] Birgé, L. and Massart, P. (1995). Estimation of integral functionals of a density. *Ann. Statist.*, **23**, 11–29.

[10] Bobkov, S. and Ledoux, M. (2016). One-dimensional empirical measures, order statistics and Kantorovich transport distances. *preprint*.

[11] Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, **71**, 353–360.

[12] Carolan, C. and Dykstra, R. (1999). Asymptotic behavior of the grenander estimator at density flat regions. *Can. J. Stat.*, **27**, 557–566.

[13] Carolan, C. A. (2002). The least concave majorant of the empirical distribution function. *Canadian Journal of Statistics*, **30**(2), 317–328.

[14] Chen, Y. and Samworth, R. J. (2013). Smoothed log-concave maximum likelihood estimation with applications. *Statist. Sinica*, **23**, 1373–1398.

[15] Cieslak, D. A. and Chawla, N. V. (2009). A framework for monitoring classifiers' performance: when and why failure occurs? *Knowl. Inf. Syst.*, **18**, 83–108.

[16] Cule, M. and Samworth, R. (2010). Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density. *Electron. J. Statist.*, **4**, 254–270.

[17] Cule, M., Samworth, R., and Stewart, M. (2010). Maximum likelihood estimation of a multidimensional log-concave density. *J. Royal Stat. Soc: Series B*, **72**, 545–607.

[18] Davidson, R. and Duclos, J.-Y. (2013). Testing for restricted stochastic dominance. *Econometric Rev.*, **32**, 84–125.

[19] DeFauw, M. C. (2011). *Cost Effectiveness Analysis in Healthcare Decision-Making: Stochastic Modeling and Statistical Inference*. Ph.D. thesis.

[20] Doss, C. R. and Wellner, J. A. (2016). Global rates of convergence of the MLEs of log-concave and *s*-concave densities. *Ann. Statist.*, **44**, 954–981.

[21] Dümbgen, L. and Rufibach, K. (2009). Maximum likelihood estimation of a log-concave density and its distribution function: Basic properties and uniform consistency. *Bernoulli*, **15**, 40–68.

[22] Dümbgen, L. and Rufibach, K. (2010). logcondens: Computations related to univariate log-concave density estimation. *J. Stat. Softw.*, **39**, 1–28.

[23] Dümbgen, L., Samworth, R., and Schuhmacher, D. (2011). Approximation by log-concave distributions, with applications to regression. *Ann. Statist.*, **39**, 702–730.

[24] Dwass, M. (1956). The large-sample power of rank order tests in the two-sample problem. *Ann. Math. Statist.*, pages 352–374.

[25] Efromovich, S. (2008). *Nonparametric curve estimation: methods, theory, and applications*. Springer Science & Business Media.

[26] Eggermont, P. P. B. and LaRiccia, V. N. (2000). Maximum likelihood es-

timation of smooth monotone and unimodal densities. *Ann. Statist.*, **28**, 922–947.

[27] Fernholz, L. T. (2012). *Von Mises calculus for statistical functionals*, volume 19. Springer Science & Business Media.

[28] Gibbs, A. L. and Su, F. E. (2002). On choosing and bounding probability metrics. *International statistical review*, **70**(3), 419–435.

[29] González-Castro, V., Alaiz-Rodríguez, R., Fernández-Robles, L., Guzmán-Martínez, R., and Alegre, E. (2010). Estimating class proportions in boar semen analysis using the Hellinger distance. *IEA/AIE*, pages 284–293.

[30] González-Castro, V., Alaiz-Rodríguez, R., and Alegre, E. (2013). Class distribution estimation based on the Hellinger distance. *Inf. Sci.*, **218**, 146–164.

[31] Gray, G. E., Huang, Y., Grunenberg, N., Laher, F., Roux, S., Andersen-Nissen, E., De Rosa, S. C., Flach, B., Randhawa, A. K., Jensen, R., Swann, E. M., Bekker, L.-G., Innes, C., Lazarus, E., Morris, L., Mkhize, N. N., Ferrari, G., Montefiori, D. C., Shen, X., Sawant, S., Yates, N., Hural, J., Isaacs, A., Phogat, S., DiazGranados, C. A., Lee, C., Sinangil, F., Michael, N. L., Robb, M. L., Kublin, J. G., Gilbert, P. B., McElrath, M. J., Tomaras, G. D., and Corey, L. (2019). Immune correlates of the thai rv144 hiv vaccine regimen in south africa. *Sci. Transl. Med.*, **11**.

[32] Groeneboom, P. (1984). Estimating a monotone density. *Department of Mathematical Statistics*.

[33] Groeneboom, P. (1989). Brownian-motion with a parabolic drift and airy functions. *Probability theory and related fields*, **81**(1), 79–109.

[34] Hall, P. and Huang, L.-S. (2002). Unimodal density estimation using kernel methods. *Statist. Sinica*, pages 965–990.

[35] Hannah, L. and Dunson, D. (2012). Ensemble methods for convex regression with applications to geometric programming based circuit design. *arXiv:1206.4645*.

[36] Haynes, B. F., Gilbert, P. B., McElrath, M. J., Zolla-Pazner, S., Tomaras, G. D., Alam, S. M., Evans, D. T., Montefiori, D. C., Karnasuta, C., Sutthent, R., Liao, H.-X., DeVico, A. L., Lewis, G. K., Williams, C., Pinter, A., Fong, Y., Janes, H., DeCamp, A., Huang, Y., Rao, M., Billings, E., Karasavvas, N., Robb, M. L., Ngauy, V., de Souza, M. S., Paris, R., Ferrari, G., Bailer, R. T., Soderberg, K. A., Andrews, C., Berman, P. W., Frahm, N., De Rosa, S. C., Alpert, M. D., Yates, N. L., Shen, X., Koup, R. A., Pitisuttithum, P., Kaewkungwal, J., Nitayaphan, S., Rerks-Ngarm, S., Michael, N. L., and Kim, J. H. (2012). Immune-correlates analysis of an HIV-1 vaccine efficacy trial. *N. Engl. J. Med.*, **366**, 1275–1286.

[37] Jankowski, H. K. (2010). Discussion of "estimation of multi-dimensional log-concave density" by Cule, Samworth, and Stewart.

[38] Johnson, A. L., Jiang, D. R., *et al.* (2018). Shape constraints in economics and operations research. *Statist. Sci.*, **33**, 527–546.

[39] Kandasamy, K., Krishnamurthy, A., Póczos, B., Wasserman, L. A., and Robins, J. M. (2015). Nonparametric von mises estimators for entropies, divergences and mutual informations. In *Adv. Neural. Inf. Process. Syst.*,

pages 397–405.

[40] Kaur, A., Prakasa Rao, B., and Singh, H. (1994). Testing for second-order stochastic dominance of two distributions. *Econometric Theory*, **10**, 849–866.

[41] Kiefer, J. and Wolfowitz, J. (1976). Asymptotically minimax estimation of concave and convex distribution functions. *Z. Wahr. verw. Gabiete*, **34**, 73–85.

[42] Kim, A. K., Guntuboyina, A., Samworth, R. J., *et al.* (2018). Adaptation in log-concave density estimation. *Annals of Statistics*, **46**(5), 2279–2306.

[43] Kulikov, V. and Lopuhaä, H. (2006). The limit process of the difference between the empirical distribution function and its concave majorant. *Statistics & Probability Letters*, **76**, 1781–1786.

[44] Laha, N. (2021). Adaptive estimation in symmetric location model under log-concavity constraint. *Electron. J. Stat.*, **15**, 2939–3014.

[45] Laha, N. and Luedtke, A. (2020). SDNNtests. https://github.com/nilanjanalaha/SDNNtests/.

[46] Le Breton, M. (1991). Stochastic orders in welfare economics. *Lecture Notes-Monograph Series*, pages 190–206.

[47] Ledwina, T. and Wyłupek, G. (2012). Two-sample test against one-sided alternatives. *Scand. J. Stat.*, **39**, 358–381.

[48] Ledwina, T. and Wyłupek, G. (2013). Tests for first-order stochastic dominance. *preprint*.

[49] Lee, Y. J. and Wolfe, D. A. (1976). A distribution-free test for stochastic ordering. *J. Amer. Statist. Assoc.*, **71**, 722–727.

[50] Lehmann, E. L. (1975). *Nonparametrics: Statistical methods based on ranks*. San Francisco: Holden-Day.

[51] Leshno, M. and Levy, H. (2004). Stochastic dominance and medical decision making. *Health. Care. Manag. Sci.*, **7**, 207–215.

[52] Levy, H. (1992). Stochastic dominance and expected utility: Survey and analysis. *J. Manag. Sci.*, **38**(4), 555–593.

[53] Lopuhaä, H. P. and Musta, E. (2019). A central limit theorem for the hellinger loss of grenander-type estimators. *Statistica Neerlandica*, **73**(2), 180–196.

[54] Mammen, E., Marron, J. S., Turlach, B. A., and Wand, M. P. (2001). A general projection framework for constrained smoothing. *Stat. Sci.*, **16**(3), 232–248.

[55] Meyer, M. C. (2012). Nonparametric estimation of a smooth density with shape restrictions. *Stat. Sin.*, pages 681–701.

[56] Miladinovic, B., Kumar, A., Mhaskar, R., and Djulbegovic, B. (2014). Benchmarks for detecting 'breakthroughs' in clinical trials: empirical assessment of the probability of large treatment effects using kernel density estimation. *BMJ open*, **4**, e005249.

[57] Mukherjee, R., Sen, B., *et al.* (2019). On efficiency of the plug-in principle for estimating smooth integrated functionals of a nonincreasing density. *Electron. J. Stat.*, **13**, 4416–4448.

[58] Nielsen, F. and Nock, R. (2014). On the chi square and higher-order chi

distances for approximating f-divergences. *IEEE Signal Process. Lett.*, **21**, 10–13.

[59] Pal, J. K., Woodroofe, M., and Meyer, M. (2007). Estimating a pólya frequency function$_2$. *Lecture Notes-Monograph Series*, **54**, 239–249.

[60] Pfanzagl, J. and Wefelmeyer, W. (1985). *Asymptotic expansions for general statistical models*. Lecture notes in statistics. Springer-Verlag.

[61] Pyke, R. and Shorack, G. R. (1968). Weak convergence of a two-sample empirical process and a new approach to Chernoff-Savage theorems. *Ann. Math. Statist.*, **39**, 755–771.

[62] Rao, B. P. (1969). Estimation of a unimodal density. *Sankhya Ser. A*, **31**, 23–36.

[63] Rao, C. R. (1995). A review of canonical coordinates and an alternative to correspondence analysis using Hellinger distance. *Qüestiió: quaderns d'estadística i investigació operativa*, **19**.

[64] Rerks-Ngarm, S., Pitisuttithum, P., Nitayaphan, S., Kaewkungwal, J., Chiu, J., Paris, R., Premsri, N., Namwat, C., de Souza, M., Adams, E., *et al.* (2009). Vaccination with alvac and aidsvax to prevent hiv-1 infection in thailand. *N. Engl. J. Med.*, **361**, 2209–2220.

[65] Robins, J., Li, L., Tchetgen, E., and van der Vaart, A. W. (2009). Quadratic semiparametric von mises calculus. *Metrika*, **69**, 227–247.

[66] Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press.

[67] Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scand. J. Stat.*, **9**, 65–78.

[68] Samworth, R. J. (2018). Recent progress in log-concave density estimation. *Statist. Sci.*, **33**, 493–509.

[69] Samworth, R. J. and Sen, B. (2018). Editorial: Special issue on "nonparametric inference under shape constraints". *Statist. Sci.*, **33**, 469–472.

[70] Shorack, G. and Wellner, J. (2009). *Empirical Processes with Applications to Statistics*. Society for Industrial and Applied Mathematics.

[71] Shorack, G. R. (1984). Empirical and rank processes of observations and residuals. *Canad. J. Statist.*, **12**, 319–332.

[72] Shorack, G. R. (2000). *Probability for Statisticians*. Springer.

[73] Sriboonchita, S., Nguyen, H. T., Wong, W.-K., and Dhompongsa, S. (2009). *Stochastic dominance and applications to finance, risk and economics*. Chapman and Hall/CRC.

[74] Stinnett, A. A. and Mullahy, J. (1998). Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Med. Decis. Making.*, **18**, S68–S80.

[75] Stupfler, G. (2014). On the weak convergence of kernel density estimators in lp spaces. *J. Nonparametr. Stat.*, **26**, 721–735.

[76] Turnbull, B. C. and Ghosh, S. K. (2014). Unimodal density estimation using bernstein polynomials. *Comput. Statist. Data Anal.*, **72**, 13–29.

[77] van de Geer, S. (2000). *Empirical Processes in M-Estimation*. Cambridge University Press.

[78] Van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge University Press.

[79] Walther, G. (2002). Detecting the presence of mixing with multiscale maximum likelihood. *J. Amer. Statist. Assoc.*, **97**, 508–513.

[80] Walther, G. *et al.* (2009). Inference and modeling with log-concave distributions. *Statist. Sci.*, **24**(3), 319–327.

[81] Wand, M. P. and Jones, M. C. (1994). Multivariate plug-in bandwidth selection. *Comput. Statist.*, **9**, 97–116.

[82] Whang, Y. J. (2019). *Econometric Analysis of Stochastic Dominance: Concepts, Methods, Tools, and Applications*. Cambridge University Press.

[83] Wolters, M. (2012). A greedy algorithm for unimodal kernel density estimation by data sharpening. *J. Stat. Softw.*, **47**, 1–26.

[84] Wolters, M. A. and Braun, W. J. (2018). Enforcing shape constraints on a probability density estimate using an additive adjustment curve. *Comm. Statist. Simulation Comput.*, **47**, 672–691.

[85] Woodroofe, M. and Sun, J. (1993). A penalized maximum likelihood estimate of f (0+) when f is non-increasing. *Statistica Sinica*, **3**.

[86] Yamanishi, K., Takeuchi, J. I., Williams, G., and Milne, P. (2004). Online unsupervised outlier detection using finite mixtures with discounting learning algorithms. *Data Min. Knowl. Discov.*, **8**, 275–300.