Estimation and testing under shape constraints

Nilanjana Laha

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Jon Wellner, Chair

Alex Luedtke, Chair

Marina Meila

Program Authorized to Offer Degree:
Department of Statistics

University of Washington

Abstract

Estimation and testing under shape constraints

Nilanjana Laha

Co-Chairs of the Supervisory Committee:
Jon Wellner
Department of Statistics

Alex Luedtke
Department of Statistics

This thesis consists of three projects, the common thread to all of which is using shape-restricted densities in inference problems. In the first project, we revisit the problem of estimating the center of symmetry θ of an unknown symmetric density f. This problem dates back to Stone (1975), Van Eeden (1970), and Sacks (1975), who constructed adaptive estimators relying on tuning parameters. Our third project, which aims to compare the outcomes from two vaccine trials, focuses on developing methodologies for testing stochastic dominance and estimating the Hellinger distance between densities. In both of these projects, we impose an additional shape restriction of either log-concavity or unimodality on the underlying densities. We show that, in both cases, the introduction of shape restrictions lead to simpler inference procedures, relying on either only one tuning parameter or none. My other project introduces a new shape-constrained class of distribution functions on $\mathbb{R}$, the *bi-s^*-concave* class, which, in parallel to the results of Dümbgen *et al.* (2017), extends the class of s-concave densities to a class including possibly multi-modal densities.

TABLE OF CONTENTS

Page

List of Figures 3

Chapter 1: Introduction 10
1.1 Overview of the thesis 11

Chapter 2: Location estimation for symmetric log-concave densities 13
2.1 Introduction 13
2.2 Estimators 19
2.3 Asymptotic properties 31
2.4 Simulation study 55
2.5 Discussion 61
2.6 Proofs 64

Chapter 3: Bi-s^*-concave distributions 113
3.1 Introduction 113
3.2 Questions and extensions: the bi$-s^*$-concave class 115
3.3 s-concavity of f implies s^*-concavity of F and $1-F$ 118
3.4 Bi-s^*-concave is bigger than $s-$concave 123
3.5 The bi-s^*-concave analogue of Theorem 3.1 127
3.6 A consequence for Fisher information 141
3.7 Confidence Bands for $F \in \mathcal{F}_{b-s^*}$ 142
3.8 Proof of Theorem 3.2 when $s \in (0, \infty)$ 148

Chapter 4: Application of shape-constrained techniques to HIV vaccine trials . . . 156
4.1 Introduction 156
4.2 Background: HVTN 097 and HVTN 100 161
4.3 Density estimation 163

4.4 Test of stochastic dominance 173
4.5 Measures of discrepancy 199
4.6 Conclusion 212
4.7 Acknowledgment 214
4.8 Proofs 214

Bibliography 243

LIST OF FIGURES

Figure Number Page

2.1 Plot of $\Psi_n(\theta, \widehat{\psi}_\theta)$ vs θ generated from the standard Gaussian distribution when $n = 5$. Here the blue ticks, the green line segments, and the pink vertical line represent the data points, the knots of $\widehat{\psi}_{0,n}$ and $\widehat{\theta}_n$, and the point of maxima respectively. 26

2.2 Plot of $\Psi_n(\theta, \widehat{\psi}_\theta)$ vs θ generated from the standard Gaussian distribution when $n = 100$. Here the blue ticks, the green line segments, and the pink vertical line represent the data points, the knots of $\widehat{\psi}_{0,n}$ and $\widehat{\theta}_n$, and the point of maxima respectively. 27

2.3 Plot of the function h_Δ defined in (2.18) versus t for a standard Gaussian sample of size 50. The pink vertical lines represents the positive knots of $\widehat{\psi}_{0,n}$, and the blue ticks represent the $|X_i - \widehat{\theta}_n|$'s. 30

2.4 Plot of Fisher information $\mathcal{I}_{f_{0,r}}$ versus r, for $f_{0,r}$ defined in (2.36) 42

2.5 The left panel displays the densities of standard Laplace, normal and logistic distribution. The right panel displays the symmetrized beta (SB) densities $f_{0,r}$, which was defined in (2.36), for different values of r. 44

2.6 Plot of $\mathcal{I}_{f_0}(\eta)/\mathcal{I}_{f_0}$ versus η for different distributions 45

2.7 The plot in the left panel displays the density f (in blue) in (2.38) and its projection $f^*(F)$ (in green); the plot in the right panel displays the density f in (2.40) (drawn in blue) and the corresponding $f^*(F)$ (in green). 50

2.8 Plot of efficiency vs sample size (n) for standard Gaussian density. We plotted the efficiency of partial MLE one-step estimator in (2.13) (Partial MLE OE; left column), the smoothed symmetrized one step estimator in (2.10) (Smoothed OE; right column), and the MLE (in purple; present in both columns). The preliminary estimators $\bar{\theta}_n$ corresponding to the one-step estimators are the mean (top), median (middle), and trimmed mean (bottom). In context of the one step estimators, the preliminary estimator is drawn in red, the truncated and untruncated estimators are drawn in blue and green respectively. 58

2.9 Plot of efficiency vs sample size (n) when the underlying density is standard Logistic. Here we consider the partial MLE one-step estimator in (2.13) (Partial MLE OE; left column), the smoothed symmetrized one step estimator in (2.10) (Smoothed OE; right column), and the MLE (in purple; present in both columns). Here the underlying distribution is standard Logistic. The preliminary estimators $\bar{\theta}_n$ corresponding to the one-step estimators are the mean (first row), median (second row), trimmed mean (third row), and the parametric MLE (fourth MLE). In context of the one step estimators, the preliminary estimator is drawn in red, the truncated and untruncated estimators are drawn in blue and green respectively. 59

2.10 Plot of efficiency vs sample size (n) for the standard Laplace density. Here we consider the partial MLE one-step estimator in (2.13) (Partial MLE OE; left column), the smoothed symmetrized one step estimator in (2.10) (Smoothed OE; right column), and the MLE (in purple; present in both columns). Here the underlying distribution is standard Laplace. The preliminary estimators $\bar{\theta}_n$ corresponding to the one-step estimators are the mean (top), median (middle), and trimmed mean (bottom). In context of the one step estimators, the preliminary estimator is drawn in red, the truncated and untruncated estimators are drawn in blue and green respectively. 60

2.11 Plot of efficiency vs sample size (n) when the underlying density is symmetrized beta, which is defined in (2.36). We consider $r = 2.1$ (left), 2.5 (middle), and 4.5 (right). The MLE is drawn in purple. We calculated the partial MLE one-step estimator in (2.13) (Partial MLE OE; top), and the smoothed symmetrized one step estimator in (2.10) (Smoothed OE; bottom). The preliminary estimator for them, which is sample mean in this case, is drawn in red; and the truncated and untruncated versions are drawn in blue and green respectively. 63

3.1 The $s-$concave densities g_r of Example 6 with $s = 2/r \in (0, \infty)$: $s = 1/8$, magenta; $s = 1/2$, green; $s = 2$, black; $s = 4$, blue; $s = 8$, red; $s = 16$, purple. 118

3.2 The bi$-s^*-$concave t_1 mixture distribution function F (black) for $\delta = 1.3$ with its convex upper bound F_U (red) and concave lower bound F_L (blue) defined by (3.8) and (3.9). 124

3.3 The bi-s^*-concave t_1 mixture density function f (black), $\delta = 1.3$, with its bi-s^*-concave upper bounds F'_U (red) and F'_L (blue) defined by (3.11) and (3.10). 125

3.4 The bi-s^*-concave t_1 mixture density function derivative f' (black) for $\delta = 1.3$ with its bi-s^*-concave upper (blue) and lower (red) bounds as given in (iv) of Theorem 3.2. 126

3.5 The Csörgő-Révész functions CR (blue) and CR_{min} (red) for the mixed t_1 density with $\delta = 1.475$ 126

3.6 The bi$-s^*-$concave t_1 distribution function F (black) with its convex upper bound F_U (red) and concave lower bound F_L (blue), where F_U and F_L are given in (3.8) and (3.9). 136

3.7 The bi-s^*-concave t_1 density function f (black) with its bi-s^*-concave upper bounds F_U' (red) and F_L' (blue) as given by (3.10) and (3.11). 137

3.8 The bi-s^*-concave t_1 density function derivative f' (black) with its bi-s^*-concave lower (red) and upper (blue) bounds as given in (iv) of Theorem 3.2. 138

3.9 The bi-s^*-concave distribution function F (black) corresponding to $g(\cdot;1)$ of Example 6 with its convex upper bound F_U (red) and concave lower bound F_L (blue) (where F_U and F_L are given in (3.12) and (3.13)). 139

3.10 The bi-s^*-concave density function $g(\cdot;1)$ of Example 6 (black) with its bi-s^*-concave upper bounds F_L' and F_U' given in (3.15) and (3.14). 140

3.11 $F'' = f'$ (black) for the bi-s^*-concave function F corresponding to the density $g(\cdot;1)$ as in Example 6 with its bi-s^*-concave upper (blue) and lower (red) bounds as given in (iv) of Theorem 3.2. 140

4.1 Plots (i), (ii), and (iii) display the empirical distribution functions, the histogram, and the boxplot of the average IgG binding responses (measured in log (net MFI)) corresponding to the HVTN 097 and HVTN 100 regimens. . . 169

4.2 The KDEs corresponding to the immune responses from the trials HVTN 097 and HVTN 100, where the bandwidths were chosen using the LSCV selector (Bowman, 1984; Rudemo, 1982). 170

4.3 Log-concave density estimators based on the immune responses in the HVTN 097 and the HVTN 100 trials. 171

4.4 Unimodal density estimators based on the immune responses in the HVTN 097 and the HVTN 100 trials. 172

4.5 foo bar 176

4.6 Plots of the densities f_γ and g_γ corresponding to the the cases (a), (b), and (c), for $\gamma = 1, 1$, and 3, respectively. 190

4.7 Plots of the distribution functions F_γ and G_γ for some values of γ, in cases (c), (d), and (e). In all the plots, the distribution functions are shown only on the region $D_p(F_\gamma, G_\gamma)$. 191

4.8 This plot compares the powers of different tests for case (a) (top), and case (b) (bottom). The NP, UM, and LC tests are given by the left, middle, and the right panels, respectively. In each of the above panels, the power curves of WRS-type, unscaled RSD ($T_{1,m,n}$ based), and the scaled RSD ($T_{2,m,n}$ based) tests are drawn in red, green, and blue, respectively. The errorbars (± 2 SD) corresponding to these tests are drawn in the same colours. 193

4.9 This plot compares the powers of different tests for case (c) (top), case (d) (middle), and case (e) (bottom). The NP, UM, and LC tests are given by the left, middle, and the right panels, respectively. In each of the above panels, the power curves of WRS-type ($T_{3,m,n}$ based), unscaled RSD ($T_{1,m,n}$ based), and the scaled RSD ($T_{2,m,n}$ based) tests are drawn in red, green, and blue, respectively. The errorbars ($\pm 2SD$) corresponding to these tests are drawn in the same colours. The pink line represents γ^*. The yellow line represents the largest γ for which the RSD based UM test satisfies $\nu(\gamma) \leq 0.05$ (γ =2.1, 1.50 and 0.23 for cases (c), (d) and (e), respectively). 197

4.10 The power of the NP, the UM, and the LC tests are given by the left, middle, and the right panels, respectively. In each of the above panels, the power curves of WRS-type, unscaled RSD ($T_{1,m,n}$ based), and the scaled RSD ($T_{2,m,n}$ based) tests are drawn in red, green, and blue, respectively. The errorbars ($\pm$2SD) corresponding to these tests are drawn in the same colours. 198

4.11 Plots of the densities f and g for the cases (d) and (e). 208

4.12 Plots of the absolute values of the biases for the cases (a)–(f). For each estimator, the errorbars are given by ± 2 SD. Here SD denotes the standard deviation of the estimator, which is estimated from the 10000 Monte Carlo samples. 209

4.13 Plots of the MSE for cases (a)–(f). For each estimator, the errorbars are given by ± 2 SD. Here SD denotes the standard deviation of the estimator, which is estimated from the 10000 Monte Carlo samples. 210

4.14 Estimated coverage probability of the 95% confidence intervals based on the estimator $H^2(\widehat{f}_m, \widehat{g}_n)$ and Theorem 4.4 for cases (a)–(f). The errorbars are given by $\pm 2SD$. Here SD, i.e. the standard deviation of the estimated probability $\widehat{p}$, is given by $(\widehat{p}(1-\widehat{p})/n)^{1/2}$. 211

ACKNOWLEDGMENTS

First, I would like to express my deepest gratitude to my advisors Jon Wellner and Alex Luedtke for their guidance, encouragement, continuous support, and thoughtful insight. I am also thankful to Marina Meila for her endless support.

There are many in the Padelford Hall who made my journey of the last five years a memorable one. Specially I want to thank Moumanti Podder and Ranjini Grove for always trusting me. I thank Saonli Basu for her helpful advices which greatly enhanced my decision making skills. I thank Alex Tank, Sean Jewell, Jenny Choi, Samson Koelle, Sheridan Grant, Kitty Mohammed, and June Morita for all the great memories we made, and all the great conversations we had.

I thank my flatmate Bindita Chaudhuri for being the best flatmate I ever had. I also thank my friends Navonil De Sarkar, Anindita Chatterjee, Rahul Mallik, Amrita Basu, Sudipto Mukherjee, Soham Dutta, Ayoush Mukherjee, Anwesha Pan for making Seattle my second home. I want to specially thank Avijit Hazra for standing by me through thick and thin and playing an instrumental role in shaping my outlook towards research.

I want to thank Paromita Dubey for staying beside me for the past ten years. I want to thank Snehashis Chakraborty for sharing both my delights and sorrows with the same grace for the past decade. I want to thank my all other friends from ISI (Kolkata) for making my stay in USA for the past five years so enjoyable.

I want to thank my family for being the greatest pillar of support in my life and instilling in me my core values. Especially, I thank my aunts Rita Pal, Sulekha Laha, Hena Pal, Vanumati Laha, and Gopali Laha; my uncles Gautam Laha and Amal Pal.

Had it not been for them, I would not be here. I thank my cousins Ahana Pal, Priyam Laha, and Mousina Pal for always making me feel special. I thank my grandmother Hira Bala Dutta for her influences in my formative years. I thank my late grandfather Nanda Dulal Dutta for always believing in me, and showing me the values of honesty and simplicity. I thank my late father Pallab Laha, whose confidence in me is my greatest fuel. I thank my mother Munmun Laha, whose love is my biggest asset.

Finally, I thank God for his kindness and the gift of faith.

.

DEDICATION

To my mother, Munmun Laha
and my late father, Pallab Laha

Chapter 1

Introduction

Over the last few decades, shape constrained methods have increasingly gathered importance in statistical inference as attractive alternatives to more traditional smoothness constraints in nonparametric estimation. While non-parametric methods such as kernel density estimation are flexible and less restrictive on the underlying data-generating process, often they require tuning parameters. shape constrained methods, on the other hand, do not rely on tuning parameters and impose conditions mostly on the shape of certain functions. Also shape constraints such as monotonicity, convexity, unimodality and log-concavity arise naturally in many applications. For example, per capita income can be modeled as an increasing function of educational qualification. Another example is the height of each person in a homogeneous population, which generally has a unimodal density. The first paper revisits the symmetric location model (Example 3.4.1 of Bickel *et al.*, 1998) with an additional assumption of log-concavity. The second paper in this thesis proposes a new shape constrained class of densities. The third paper focuses on application of shape-constraints like unimodality and log-concavity to the analysis of vaccine trial data. In particular we propose some shape-constrained alternatives for non-parametric tests of stochastic dominance.

1.1 Overview of the thesis

My first project, in collaboration with Dr. Jon Wellner, revisits the location-estimation problem in symmetric models. The aim is to estimate the center of symmetry, θ, of an unknown distribution $f = g(\cdot - \theta)$, where g is symmetric about zero and has finite Fisher information for location. This problem dates back to Stein (1956). Although Stone (1975), Van Eeden (1970), and Sacks (1975) constructed adaptive estimators of θ, a major limitation of their methods is that they require tuning parameters for which there is currently no universally accepted data-adaptive selection method. In view of the fact that many common symmetric unimodal densities are log-concave, we make an additional assumption that g is log-concave, which results in tuning parameter free estimation of g. We estimate θ using only one or even zero tuning parameters. Moreover, we analytically show that our estimators of θ are robust to the misspecification of log-concavity.

In my second project with Dr. Jon Wellner, we introduce and study a new shape-constrained class of distribution functions on $\mathbb{R}$, which we name the **bi-s^*-concave class**. This is an extension of the class of s-concave densities, which includes many common continuous densities such as the class of log-concave densities ($s = 0$) and t-densities. Although this is a rich class, s-concave densities are necessarily unimodal, thereby excluding many types of mixture densities which naturally arise in many fields such as speech recognition, pattern recognition, climatology, just to name a few. The bi-s^*-concave class, on the other hand, permits distributions with bimodal and multimodal densities.

My third project, in collaboration with Dr. Alex Luedtke, seeks to compare the IgG binding immune responses arising in two HIV vaccine trials. The goal of this comparison is to help in explaining the role of the immune response invoked by IgG antibodies in preventing HIV. Our data indicates that the underlying densities of the immune responses are unimodal, which is unsurprising given that each trial is based on homogeneous population. Therefore, we consider shape-constrained methods to compare the immune profiles of the

two vaccines. To this end, we develop novel shape-constrained tests of stochastic dominance and shape-constrained plug-in estimators of the Hellinger distance between two densities. Our techniques are either tuning parameter free, or rely on only one tuning parameter, but their performance is comparable with nonparametric methods. The minimal dependence of tuning parameters is especially desirable in clinical contexts where analyses must be prespecified and reproducible. We also show that our tests and estimators have desirable asymptotic properties.

Chapter 2

Location estimation for symmetric log-concave densities

2.1 Introduction

Let $\mathcal{P}$ denote the class of all densities on the real line $\mathbb{R}$. For any $\theta \in \mathbb{R}$, denote by $\mathcal{S}_\theta$ the class of all densities that are symmetric about θ. We also denote

$$\mathcal{SC}_\theta := \{\phi : \mathbb{R} \mapsto \mathbb{R} \;\Big|\; \phi \text{ is concave and symmetric about } \theta\}.$$

We let $\mathcal{LC}$ be the set of all log-concave densities on $\mathbb{R}$, and denote the class of all log-concave densities symmetric about θ by

$$\mathcal{SLC}_\theta := \{f \in \mathcal{L} \;\Big|\; \phi = \log f \in \mathcal{SC}_\theta\}.$$

In this paper we focus on the log-concave symmetric location model

$$\mathcal{P}_0 = \left\{ f \in \mathcal{P} \;\Big|\; f(x;\theta) = g(x-\theta),\ \theta \in \mathbb{R},\ g \in \mathcal{SLC}_0,\ \mathcal{I}_f < \infty \right\}, \tag{2.1}$$

where $\mathcal{I}_f$ is the Fisher information for location. Our aim is to estimate θ in $\mathcal{P}_0$.

Estimation of θ in the full symmetric location model

$$\mathcal{P}_s = \left\{ f \in \mathcal{P} \;\Big|\; f(x;\theta) = g(x-\theta),\ \theta \in \mathbb{R},\ g \in \mathcal{S}_0,\ \mathcal{I}_f < \infty \right\} \tag{2.2}$$

is an old semi-parametric problem, dating back to Stein (1956). From then on, this problem has been considered by many early authors including, but not limited to, Stone (1975), Beran (1974), Sacks (1975), and Van Eeden (1970). There are two main reasons behind the assumption of symmetry in this model. First, as Stone (1975) has noted, if f is totally unrestricted, θ is not identifiable. Second, the definition of location becomes less clear in the absence of symmetry (Takeuchi, 1975).

This model is appealing because Stone (1975) showed that the restriction of symmetry and the finiteness of $\mathcal{I}_f$ are sufficient to guarantee adaptive estimation of θ. Therefore, in this case, it is possible to construct a consistent estimator of θ, whose asymptotic variance attains the parametric lower bound of $\mathcal{I}_f^{-1}$. From Theorem 3 of Huber (1964), it follows that $\mathcal{I}_f$ is finite if and only if f is an absolutely continuous density satisfying

$$\int_{-\infty}^{\infty} \left(\frac{f'(x)}{f(x)} \right)^2 f(x)dx < \infty,$$

where f' is an L_1-derivative of f. Also in this case $\mathcal{I}_f$ takes the form

$$\mathcal{I}_f = \int_{-\infty}^{\infty} \left(\frac{f'(x)}{f(x)} \right)^2 f(x)dx.$$

See section 3.2, 3.3, and 6.3 of Bickel *et al.* (1998) for more discussion on adaptive estimation in $\mathcal{P}_s$.

Now we will briefly discuss some existing methods of estimating θ in $\mathcal{P}_s$. Stone (1975) considers a one-step estimator. One-step estimators generally depend on the estimation of the score $-\phi' = f'/f$. Stone (1975) estimates the scores using symmetrized Gaussian kernels based on truncated data, thereby incorporating two tuning parameters, one for the truncation, and another for the scaling of the Gaussian Kernel. Although the tuning parameters are required to satisfy some asymptotic conditions (Stone, 1975, Theorem 5.3), no particular prescription was provided for choosing them. Sacks (1975) considers a subclass of $\mathcal{P}_s$ with some restrictions on ϕ' that are weaker than the assumption of log-concavity, and includes heavy-tailed distributions like Cauchy. His estimator is based on a linear function of scores,

which he estimates using the order statistics. The estimation procedure involves three tuning parameters. Sacks (1975) gives some examples of these tuning parameters, but does not provide any data dependent method for choosing them. Beran (1974) uses linearized rank estimate for adaptive estimation of θ in $\mathcal{P}_s$. But he makes the additional assumption that ϕ' is twice continuously differentiable, which excludes densities like Laplace. His method exploits the Fourier series expansion for estimating the score function ϕ'.

There has been substantial development in log-concave density estimation in recent years. But so far there has been very little (or no) use of shape - constraints in connection with semiparametric models. For a short list of exceptions, see below. In particular, shape constraints have not been considered in detail in connection with the one-sample symmetric location problem considered here, with perhaps one exception. To the best of our knowledge, the only work of location-estimation in $\mathcal{P}_0$ was presented by Van Eeden (1970), quite long ago. Although her paper did not mention log-concavity, Van Eeden (1970) restricted $\mathcal{P}_s$ by assuming $f'(F^{-1}(u))/f(F^{-1}(u))$ is non-increasing in u, or equivalently $f'(x)/f(x)$ is non-increasing in x, which is in turn equivalent to f being log-concave. Van Eeden (1970) considered data-partitioning to estimate scores from a small fraction of the data and construct the Hodges-Lehmann rank estimate of location (Hodges and Lehmann, 1963) from the remaining data. The partitioning of the data and the construction of the score function involves introduction of some tuning parameters. Van Eeden (1970) did not explicitly describe how to choose these tuning parameters in her proposed procedure. A recent work in a somewhat similar direction was accomplished by Bhattacharyya (2013) who consider both location and scale estimation in an elliptical symmetry model, which albeit bearing some resemblance, is quite different from $\mathcal{P}_0$. Also, Bhattacharyya (2013)'s estimation procedure is completely different from ours.

Now we provide some justification for imposing the shape restriction of log-concavity. The class of log-concave densities, $\mathcal{LC}$, belongs to the larger class of unimodal densities. Uni-

modality is a reasonable assumption in context of location estimation of symmetric densities because, as Takeuchi (1975) points out, in practice multimodal densities generally result from unimodal mixtures, for which separate procedures are available. However, the class of unimodal densities does not admit a maximum likelihood estimator (MLE) (Birgé, 1997). However, smaller and richer classes like $\mathcal{LC}$ and $\mathcal{SLC}_\theta$ (for a fixed $\theta \in \mathbb{R}$) are known to admit MLE (cf. Pal *et al.*, 2007; Dümbgen and Rufibach, 2009; Doss and Wellner, 2019; Xu and Samworth, 2017), which is a big reason to opt for log-concavity. Since maximum likelihood estimation does not depend on tuning parameters, it is possible to estimate f without relying on any tuning parameter, which can facilitate tuning parameter free estimation of θ. Finally, most common symmetric unimodal continuous densities are log-concave except the Cauchy and the t-densities.

In this article, we consider two types of estimators of θ: a truncated one-step estimator and the MLE of θ in $\mathcal{P}_0$. We propose four different types of one-step estimators. We estimate the corresponding scores using either the MLE of f in $\mathcal{LC}$, or in $\mathcal{SLC}_{\bar{\theta}_n}$, where $\bar{\theta}_n$ is a preliminary estimator of θ. When we use the MLE of f in $\mathcal{LC}$, we symmetrize it about $\bar{\theta}_n$ prior to computing the scores. The existence of the MLE of f in $\mathcal{SLC}_{\theta'}$ for any fixed θ' is central to the computation of the MLE of θ in $\mathcal{P}_0$ as well. Because of this fact, the joint MLE of (θ, f) can be easily computed using the profile likelihood method. All our estimators are easily computable because the MLE of f in $\mathcal{LC}$ or $\mathcal{SLC}_\theta$ is easily computable using R package "logcondens" or "logcondens.mode". Though the one step estimators rely on a truncation parameter η, our simulations show that the performance of the estimators is not very sensitive to the choice of η. The MLE does not depend any tuning parameter.

A question that naturally arises is whether our estimators of θ are robust to the violation of the log-concavity assumption. The log-concave projection theory developed by Dümbgen *et al.* (2011), Cule and Samworth (2010), and Xu and Samworth (2017) is able to answer this question providing interpretability of the estimators of f and θ under a general $f \in \mathcal{P}$.

This projection theory expresses the distribution function of the MLE of f in $\mathcal{LC}$ (or $\mathcal{SLC}_\theta$) as the projection (in some suitable sense) of the empirical distribution $\mathbb{F}_n$ on the space of all distribution functions with densities in $\mathcal{LC}$ (or $\mathcal{SLC}_\theta$). We show that our estimators of θ are consistent under very mild conditions, as long as the underlying density f is symmetric about θ. When f is not symmetric, or if the data generating distribution does not even possess a density, $\widehat{\theta}_n$ can still possess a limit, whose existence can be proved by the log-concave projection theory mentioned above. The early authors, Stone (1975), Beran (1974), Sacks (1975), and Van Eeden (1970), did not discuss the behavior of their estimators under model miss-specification (although Stone (1975) did discuss the behavior of his estimators when $\mathcal{I}_{f_0} = \infty$.

When the model is correctly specified, i.e. $f \in \mathcal{P}_0$, we establish that the one-step estimators are $\sqrt{n}-$consistent, and nearly achieve the asymptotic efficiency bound $\mathcal{I}_f^{-1}$. When f is not log-concave, we show that under some reasonable conditions, the one-step estimators are still $\sqrt{n}$-consistent. For the MLE $\widehat{\theta}_n$, we derive the rate of $O_p(n^{-2/5})$ when f is log-concave. We conjecture that the actual rate is of $O_p(n^{-1/2})$. Let $H(f,g)$ be the Hellinger distance between two densities f, g, i.e.

$$H^2(f,g) = 2^{-1} \int (\sqrt{f(x)} - \sqrt{g(x)})^2 dx.$$

We show that the Hellinger distance between f and its MLE under the model $\mathcal{P}_0$ is of order $O_p(n^{-2/5})$, which is analogous to the rate obtained by Doss and Wellner (2019) for the MLE of f in $\mathcal{SLC}_0$, or by Doss and Wellner (2016) for the MLE of f in $\mathcal{LC}$.

The article is organized as follows. In section 2.2 we introduce the estimators and discuss their asymptotic properties in section 2.3. We provide a simulation study in section 2.4. The proofs are deferred to section 2.6.

2.1.1 Preliminaries/ Notation and terminology

We assume that

$X_1, \ldots, X_n$ are independent and identically distributed (i.i.d.) random variables with density $f_0 \in \mathcal{P}_0$. Therefore from (2.1) it follows that $f_0 = g_0(\cdot - \theta_0)$, where $g_0 \in \mathcal{SLC}_0$, the class of all log-concave densities symmetric about θ_0. We denote the corresponding log-densities by $\phi_0 = \log f_0$, and $\psi_0 = \log g_0$. We let F_0 and G_0 denote the respective distribution functions of f_0 and g_0, and denote by P_0 the measure corresponding to F_0. We denote the empirical distribution function of the X_i's by $\mathbb{F}_n$, and write $\mathbb{P}_n$ for the corresponding empirical measure. The (classical) empirical process of the X_i's will be denoted by $\mathbb{Z}_n = \sqrt{n}(\mathbb{F}_n - F_0)$. For any integrable function $h : \mathbb{R} \mapsto \mathbb{R}$, we write

$$\mathbb{P}_n h := n^{-1} \sum_{i=1}^{n} h(X_i) \quad \text{and} \quad P_0 h = \int_{-\infty}^{\infty} h(x) f_0(x) dx.$$

As usual, we denote the order statistics of a sample $(Y_1, \ldots, Y_n)$ by

$$Y_{()} = (Y_{(1)}, \ldots, Y_{(n)}).$$

We denote the set of all concave functions on $\mathbb{R}$ by $\mathcal{C}$. We will follow the convention that any concave function ϕ takes the value $-\infty$ outside its effective domain where $\mathrm{dom}(\phi)$ is as defined in Rockafellar (1970) (page 40), that is, $\mathrm{dom}(\phi) = \{x \in \mathbb{R} | \phi(x) > -\infty\}$. For any concave function $\psi : \mathbb{R} \mapsto \mathbb{R}$, we say $x \in \mathbb{R}$ is a knot of ψ, if either $\psi'(x+) \neq \psi'(x-)$, or x is at the boundary of $\mathrm{dom}(\psi)$. We denote by $\mathcal{S}(\psi)$ the set of the knots of ψ. Unless otherwise mentioned, for a real valued function h, provided they exist, h' and $h'(\cdot-)$ will refer to the right and left derivatives of h respectively. For a distribution function F, we let $J(F)$ denote the set $\{0 < F < 1\}$. For any non-negative real-valued function h, by $\mathrm{supp}(h)$ we denote the set where h is positive. For two sets A and B, $A \times B$ will represent the Cartesian product. For any set $A \subset \mathbb{R}$, and $x \in \mathbb{R}$, we use the usual notation $A + x$ to denote the translated set $\{y + x \ : \ y \in A\}$. The notation $\overline{A}$ will refer to the closure of the set A. As usual, we denote the set of natural numbers by $\mathbb{N}$.

Next, we develop some notation for maximum likelihood estimation. For $\psi \in \mathcal{SC}_0$ and $\theta \in \mathbb{R}$, following Dümbgen *et al.* (2011) and Xu and Samworth (2017), we define the criterion function for maximum likelihood estimation by

$$\Psi(\theta, \psi, F) = \int_{-\infty}^{\infty} \psi(x-\theta) dF(x) - \int_{-\infty}^{\infty} e^{\psi(x-\theta)} dx. \tag{2.3}$$

Following the common practice (cf. Dümbgen and Rufibach, 2009; Doss and Wellner, 2019), we use the standard device of including a Lagrange term to get rid of the normalizing constant involved in density estimation. We use the notation $\Psi_n(\theta, \psi)$ to denote the sample version $\Psi(\theta, \psi, \mathbb{F}_n)$ of $\Psi(\theta, \psi, F)$. Thus,

$$\Psi_n(\theta, \psi) = \int_{-\infty}^{\infty} \psi(x-\theta) d\mathbb{F}_n(x) - \int_{-\infty}^{\infty} e^{\psi(x-\theta)} dx. \tag{2.4}$$

We denote the maximized criterion function by

$$L(F) = \sup_{\theta \in \mathbb{R}, \psi \in \mathcal{SC}_0} \Psi(\theta, \psi, F). \tag{2.5}$$

2.2 Estimators

2.2.1 One-step estimators

In this section, we discuss the construction of the one-step estimators of θ. The construction of one step estimators always starts with a preliminary estimator $\bar{\theta}_n$. We will impose two requirements on $\bar{\theta}_n$: $\sqrt{n}$-consistency and strong consistency. The $Z-$estimator of shift θ in the logistic location shift model satisfies these requirements under minimal regularity conditions (see Theorem 5.23 of Van der Vaart (1998)). The sample mean and the sample median also satisfy our requirements on $\bar{\theta}_n$ when $f_0 \in \mathcal{P}_0$.

Suppose $\tilde{g}_n$ is an estimator of the centered density $g_0 = f_0(\cdot + \theta_0)$. We let $\tilde{\psi}_n$ denote the respective log-density $\log \tilde{g}_n$, and write $\tilde{G}_n$ for the distribution function corresponding to $\tilde{g}_n$. Ideally, to construct a one-step estimator in this setting, one would first estimate the Fisher

information by

$$\widehat{\mathcal{I}}_n = \int_{-\infty}^{\infty} \tilde{\psi}_n'(x - \bar{\theta}_n)^2 d\mathbb{F}_n(x), \tag{2.6}$$

and set

$$\tilde{\theta}_n = \bar{\theta}_n - \int_{-\infty}^{\infty} \frac{\tilde{\psi}_n'(x - \bar{\theta}_n)}{\widehat{\mathcal{I}}_n} d\mathbb{F}_n(x). \tag{2.7}$$

It is not hard to see that the the choice of $\tilde{g}_n$ will heavily influence the performance of $\tilde{\theta}_n$. In the process of choosing $\tilde{g}_n$, we also want to ensure that $\tilde{g}_n$ is symmetric about 0, and if possible, log-concave.

A simple way to estimate $f_0 = g_0(\cdot - \theta_0)$ is to compute its MLE in $\mathcal{LC}$, the class of all log-concave densities. However, this log-concave MLE, which we will denote by $\bar{f}_m$ from now on, may not be symmetric about any $\theta \in \mathbb{R}$. Nevertheless we can still use this naive estimator $\bar{f}_m$ to construct reasonable choices of $\tilde{g}_n$. In what follows, we illustrate several estimators of $\tilde{g}_n$.

Symmetrized estimator $\tilde{g}_n^{sym}$

The simplest way of obtaining a symmetrized estimator of g_0 from $\bar{f}_m$ is by setting

$$\tilde{g}_n^{sym}(z) = \frac{1}{2}(\bar{f}_m(\bar{\theta}_n + z) + \bar{f}_m(\bar{\theta}_n - z)), \quad z \in \mathbb{R}. \tag{2.8}$$

Although $\tilde{g}_n^{sym}$ is symmetric, it is not, in general, log-concave. Additionally, the performance of $\tilde{g}_n^{sym}$ can suffer from the lack of smoothness of $\bar{f}_m$. As a consequence, the one-step estimator based on $\tilde{g}_n^{sym}$ has considerably less efficiency, especially for smaller sample sizes. We seek to improve the performance of $\tilde{g}_n^{sym}$ by incorporating smoothing, which leads to our next estimator of g_0.

Smoothed symmetrized estimator $(\tilde{g}_n^{sym})^{sm}$

To perform smoothing, our first task is to choose a data dependent smoothing parameter b_n. Towards this end, denoting the sample variance $\widehat{\sigma}_n^2$ by

$$\widehat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2,$$

and the variance corresponding to the density $\bar{f}_m$ by

$$\sqrt{\frac{12mn}{N+1}} = \int_{-\infty}^{\infty} z^2 \bar{f}_m(z)dz - \left(\int_{-\infty}^{\infty} z\bar{f}_m(z)dz\right)^2,$$

we note that $\widehat{\sigma}_n^2 - \sqrt{\frac{12mn}{N+1}} > 0$, which follows from (2.1) of Chen and Samworth (2013). This difference of variances is a reasonable choice of b_n. Therefore, setting

$$b_n^2 := \widehat{\sigma}_n^2 - \sqrt{\frac{12mn}{N+1}}, \tag{2.9}$$

we define the smoothed symmetrized estimator by

$$(\tilde{g}_n^{sym})^{sm}(z) := \frac{1}{b_n}\int_{-\infty}^{\infty} \tilde{g}_n^{sym}(t)\varphi\left(\frac{z-t}{b_n}\right)dt, \quad z \in \mathbb{R}, \tag{2.10}$$

where φ is the standard normal density. The estimator $(\widehat{f}_n^{sym})^{sm}$ can also be represented in terms of the smoothed version of $\bar{f}_m$, which we denote by

$$\bar{f}_m^{sm}(z) = \frac{1}{b_n}\int_{-\infty}^{\infty} \bar{f}_m(z-t)\varphi(t/b_n)dt, \quad z \in \mathbb{R}, \tag{2.11}$$

by noting that

$$(\widehat{f}_n^{sym})^{sm}(z) = \frac{\widehat{f}_n^{sm}(\bar{\theta}_n + z) + \widehat{f}_n^{sm}(\bar{\theta}_n - z)}{2}. \tag{2.12}$$

Though $(\tilde{g}_n^{sym})^{sm}$ leads to a more efficient one-step estimator of θ_0 in small samples, it is not necessarily log-concave, which leaves room for further improvement.

Partial MLE estimator $\widehat{g}_{\bar{\theta}_n}$

The following estimator of g_0 is obtained by maximizing the criterion function $\Psi_n(\bar{\theta}_n, \psi)$ in (2.4) over $\psi \in \mathcal{SC}_0$. From Theorem 2.1(C) of Doss and Wellner (2019) it follows that the maximizer $\widehat{\psi}_{\bar{\theta}_n}$ satisfies

$$\int_{-\infty}^{\infty} e^{\widehat{\psi}_{\bar{\theta}_n}(x)} dx = 1,$$

which leads to the density estimator

$$\widehat{g}_{\bar{\theta}_n} = e^{\widehat{\psi}_{\bar{\theta}_n}}. \tag{2.13}$$

We call this estimator a *Partial MLE estimator* to distinguish it from the traditional MLE of θ_0, which we discuss in Section 2.2.2. Notice that $\widehat{g}_{\bar{\theta}_n}$ satisfies both requirements of symmetry (about 0) and log-concavity.

From Dümbgen and Rufibach (2009) and Doss and Wellner (2019), it follows that the three density estimators $\tilde{g}_n$ we have considered so far share the same support

$$[-\max(X_{(n)} - \bar{\theta}_n, \bar{\theta}_n - X_{(1)}), \max(X_{(n)} - \bar{\theta}_n, \bar{\theta}_n - X_{(1)})].$$

Geometric mean type symmetrized estimator $\tilde{g}_n^{geo,sym}$

Similar to $\tilde{g}_n^{sym}$, this estimator is also a symmetrized version of $\bar{f}_m$. However, this time the symmetrization is based on the geometric mean of $\bar{f}_m(\bar{\theta}_n + \cdot)$ and $\bar{f}_m(\bar{\theta}_n - \cdot)$ instead of the arithmetic mean. For $z \in \mathbb{R}$, we define the geometric mean type symmetrized estimator by

$$\tilde{g}_n^{geo,sym}(z) := C_n^{geo} \left(\bar{f}_m(\bar{\theta}_n + z) \bar{f}_m(\bar{\theta}_n - z) \right)^{1/2}, \tag{2.14}$$

where C_n^{geo} is a random normalizing constant. Noting that the sum of two concave functions is again concave, we deduce that $\tilde{g}_n^{geo,sym}$ is log-concave. However, its support

$$\text{supp}(\tilde{g}_n^{geo,sym}) = [-\min(X_{(n)} - \bar{\theta}_n, \bar{\theta}_n - X_{(1)}), \min(X_{(n)} - \bar{\theta}_n, \bar{\theta}_n - X_{(1)})]$$

is smaller than that of the previous estimators of g_0.

Constructing the truncated one-step estimator

As pointed out earlier in this section, one would ideally like to consider the estimator suggested by (2.7), which does not rely on any truncation. However, the behavior of $\tilde{\psi}_n'$ near the boundary of its domain is hard to control because $\tilde{\psi}_n'$ is unbounded on its domain when ϕ_0' is unbounded (see Corollary 3 in Section 3.1).

To avoid the difficulties in the tails, we consider a truncated one-step estimator. Hendrickx and Groeneboom (2017) used a similar idea in a current status linear regression model where a similar truncated log-likelihood was considered to avoid problems in the tails of the unknown error distribution. Alhough our method introduces a tuning parameter for the truncation, the choice of this tuning parameter is less crucial compared to the nonparametric methods. Kuchibhotla *et al.* (2017), who considered convexity-constrained single index model, also used a tuning parameter in parallel to shape restriction.

Recall that we denoted the distribution function of $\tilde{g}_n$ by $\tilde{G}_n$. We take the truncation parameter η to be a small fixed positive number in the interval $(0, 1/2)$, and denote by ξ_n the $(1-\eta)$-th quantile of $\tilde{G}_n$. The symmetry of $\tilde{g}_n$ about 0 implies that $\tilde{G}_n^{-1}(\eta) = -\xi_n$.

We can estimate the Fisher information either by

$$\widehat{\mathcal{I}}_n^*(\eta) = \int_{\bar{\theta}_n - \xi_n}^{\bar{\theta}_n + \xi_n} \tilde{\psi}_n'(x - \bar{\theta}_n)^2 d\tilde{G}_n(x - \bar{\theta}_n),$$

or

$$\widehat{\mathcal{I}}_n(\eta) = \int_{\bar{\theta}_n - \xi_n}^{\bar{\theta}_n + \xi_n} \tilde{\psi}_n'(x - \bar{\theta}_n)^2 d\mathbb{F}_n(x). \tag{2.15}$$

Though it can be shown that both the above estimators converge to the same limit almost surely, our simulations indicate that the estimator $\widehat{\mathcal{I}}_n(\eta)$ leads to a more efficient one-step estimator of θ_0. Therefore, we take $\widehat{\mathcal{I}}_n(\eta)$ as our estimator of the Fisher information for the purpose of estimating θ.

In the same spirit as the estimator in (2.7), we now construct a truncated one-step

estimator of θ_0 as follows:

$$\tilde{\theta}_n = \bar{\theta}_n - \int_{\bar{\theta}_n - \xi_n}^{\bar{\theta}_n + \xi_n} \frac{\tilde{\psi}_n'(x - \bar{\theta}_n)}{\widehat{\mathcal{I}}_n(\eta)} d\mathbb{F}_n(x). \tag{2.16}$$

First note that we use all the data to estimate $\bar{\theta}_n$, $\tilde{\psi}_n'$, and $\tilde{g}_n$. The truncation comes only in the step where we are to set-up the one step estimator. Next, observe that $\widehat{\mathcal{I}}_n(\eta)$ is always smaller than the estimator of untruncated Fisher information $\widehat{\mathcal{I}}_n$ in (2.6). In Section 2.3.1, we show that the asymptotic variance of $\tilde{\theta}_n$ has an inverse relation with the Fisher information, which makes one expect that the truncated estimator in (2.16) will be asymptotically less efficient than the untruncated estimator in (2.7). However, our simulations in Section 2.4 show that for sufficiently small η, this loss in efficiency is negligible for most densities. It is also evident that $\widehat{\mathcal{I}}_n(\eta)$ increases in η, so smaller values of the truncation parameter η will lead to estimators with higher asymptotic efficiency.

2.2.2 Maximum likelihood estimator (MLE)

The main aim of this section is to study the existence and the basic properties of the MLE of (θ_0, g_0). We previously mentioned that the MLE of f_0 in $\mathcal{SLC}_\theta$ exists for fixed $\theta \in \mathbb{R}$. This fact plays a major role in maximizing the log-likelihood function in the model $\mathcal{P}_0$ via profile likelihood method.

Recall the definition of $\Psi_n(\theta, \psi)$ from (2.4). The MLE $(\widehat{\theta}_n, \widehat{\psi}_{0,n})$ of (θ_0, ψ_0) satisfies

$$(\widehat{\theta}_n, \widehat{\psi}_{0,n}) = \underset{\theta_0 \in \mathbb{R}, \psi \in \mathcal{SC}_0}{\arg\max} \Psi_n(\theta, \psi, F).$$

Theorem 2.1(c) of Doss and Wellner (2019) ensures the existence of a unique maximizer of $\Psi_n(\theta, \psi)$ in $\psi \in \mathcal{SC}_0$ for fixed $\theta \in \mathbb{R}$. This maximizer will be denoted by $\widehat{\psi}_\theta$ from now on. Theorem 2.1(c) of Doss and Wellner (2019) also implies that the function $e^{\widehat{\psi}_\theta}$ is a proper density. Maximizing $\Psi_n(\theta, \widehat{\psi}_\theta)$ as a function of θ yields $\widehat{\theta}_n$.

Once $\widehat{\theta}_n$ is derived, $\widehat{\psi}_{0,n}$ can also be found by plugging $\widehat{\theta}_n$ into $\widehat{\psi}_\theta$. From $\widehat{\psi}_{0,n}$, we compute $\widehat{g}_n$, the MLE of g_0, by taking $\widehat{g}_n = e^{\widehat{\psi}_{0,n}}$. We will denote the distribution function corre-

sponding to $\widehat{g}_n$ by $\widehat{G}_m$. Note that the MLE of f_0 can be obtained by $\widehat{f}_n = \widehat{g}_n(\cdot - \widehat{\theta}_n)$. The corresponding distribution function will be denoted by $\bar{F}_m$. Further, we let $\widehat{\phi}_{0,n}$ denote the log-density $\log \widehat{f}_n$.

Before getting into the analysis of the existence of $(\widehat{\theta}_n, \widehat{\psi}_{0,n})$ for a general case, let us consider a special case first. Suppose $\mathbb{F}_n$ is degenerate, i.e. $\mathbb{F}_n\{x_0\} = 1$ for some $x_0 \in \mathbb{R}$. Intuition leads us to make the guess that in this case, $\widehat{\theta}_n = x_0$. Indeed, denoting A_k to be the set $[-1/(2k), 1/(2k)]$ for $k \geq 1$, and considering the sequence of functions $\{\psi_k\}_{k\geq 1} \in \mathcal{SC}_0$ defined by

$$\psi_k(x) = (\log k) 1_{A_k}(x) - \infty \cdot 1_{A_k^c}(x), \quad k \geq 1,$$

we observe that

$$\Psi_n(x_0, \psi_k) = \log k - 1 \to \infty, \quad \text{as } k \to \infty.$$

Therefore, x_0 is a candidate for the MLE of θ_0. However, the MLE of ψ_0, i.e. $\widehat{\psi}_{0,n}$, does not exist for this case. To verify, observe that if the MLE of ψ_0 does exist for some $\widehat{\theta}_n \in \mathbb{R}$, it follows that

$$\widehat{\psi}_{0,n}(x_0 - \widehat{\theta}_n) - \int_{-\infty}^{\infty} e^{\widehat{\psi}_{0,n}(x)} dx = \Psi_n(x_0, \widehat{\psi}_{0,n}) \geq \lim_{k\to\infty} \Psi_n(x_0, \psi_k) = \infty,$$

leading to

$$\widehat{\psi}_{0,n}(x_0 - \widehat{\theta}_n) = \infty,$$

which contradicts the fact that $\widehat{\psi}_{0,n}$ is a real valued function. Hence, we conclude that the MLE of (θ_0, ψ_0) does not exist when $\mathbb{F}_n$ is degenerate. So we only need to focus on the case when $\mathbb{F}_n$ is non-degenerate, i.e., $\mathbb{F}_n\{x\} < 1$ for all $x \in \mathbb{R}$. We will show that the map $\theta \mapsto \Psi_n(\theta, \widehat{\psi}_\theta)$ is continuous when $\mathbb{F}_n$ is non-degenerate. We also establish existence of the MLE in this case.

Lemma 2.1. *If $\mathbb{F}_n$ is non-degenerate, $\theta \mapsto \Psi_n(\theta, \widehat{\psi}_\theta)$ is a continuous map.*

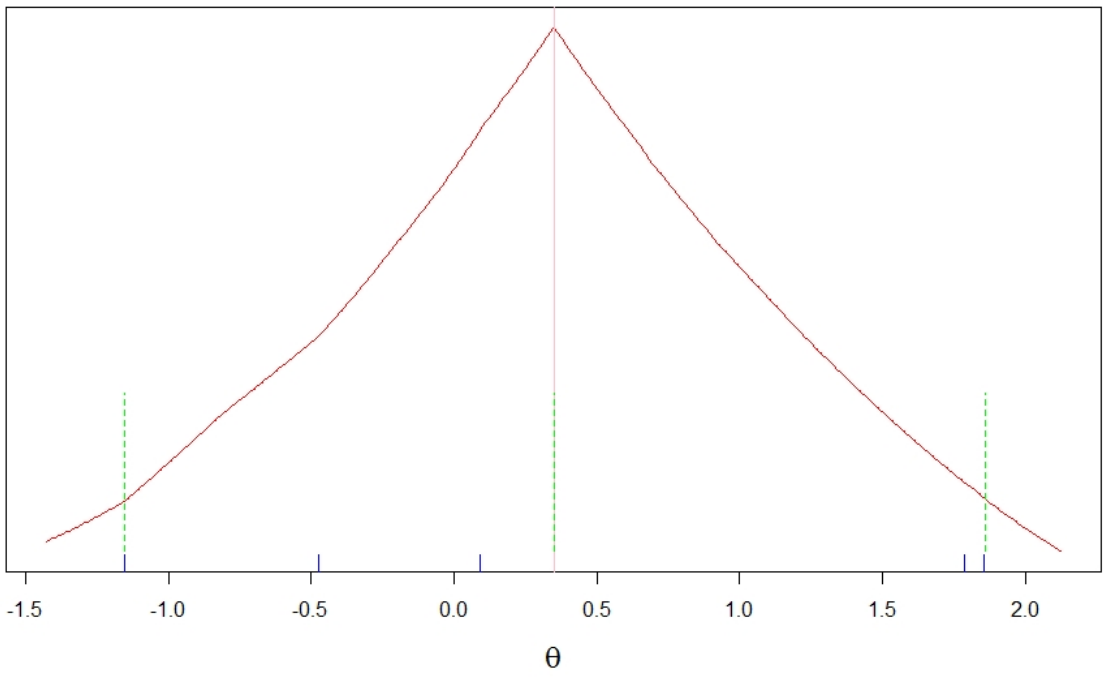


Figure 2.1: Plot of $\Psi_n(\theta, \widehat{\psi}_\theta)$ vs θ generated from the standard Gaussian distribution when $n = 5$. Here the blue ticks, the green line segments, and the pink vertical line represent the data points, the knots of $\widehat{\psi}_{0,n}$ and $\widehat{\theta}_n$, and the point of maxima respectively.

Figure 2.1 and Figure 2.2 illustrate the continuity of $\theta \mapsto \Psi_n(\theta, \widehat{\psi}_\theta)$ for two samples of size 5 and 100, generated from the standard Gaussian distribution.

Our next theorem establishes the existence of a maximizer of $\Psi_n(\theta, \widehat{\psi}_\theta)$ for non-degenerate $\mathbb{F}_n$.

Theorem 2.1. *When $\mathbb{F}_n$ is non-degenerate, the MLE $(\widehat{\theta}_n, \widehat{\psi}_n)$ of (θ_0, ψ_0) exists. Moreover, $\widehat{\theta}_n \in [X_{(1)}, X_{(n)}]$.*

Observe that Theorem 2.1 does not ensure the uniquenesss of $\widehat{\theta}_n$. Since $\Psi_n(\theta, \psi)$ may not

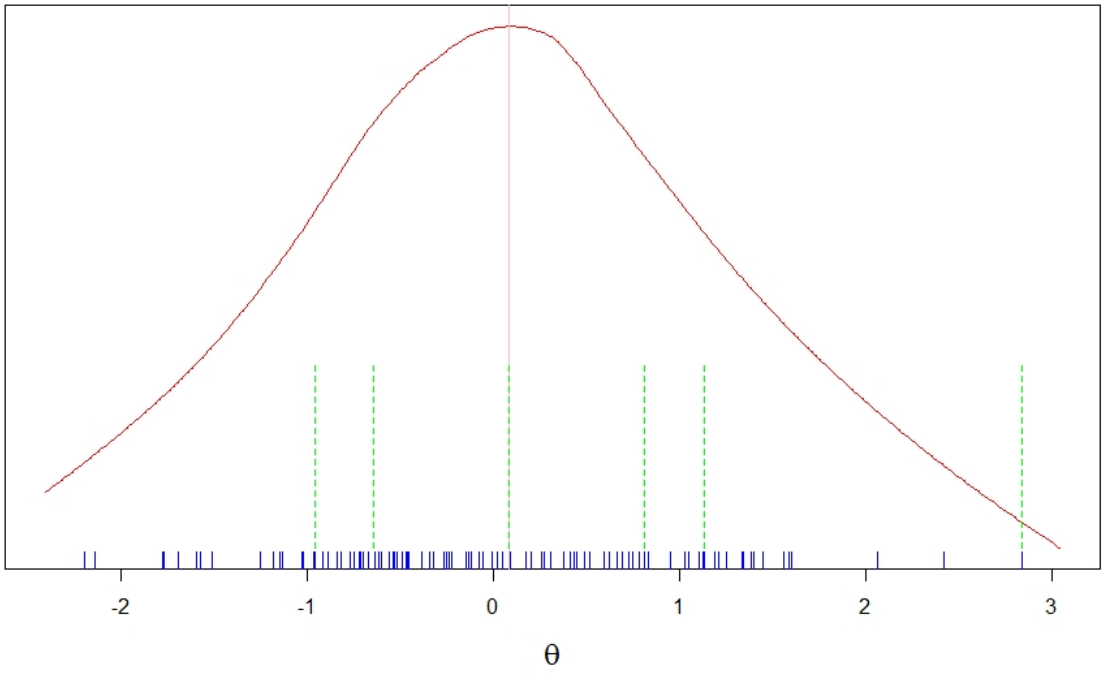


Figure 2.2: Plot of $\Psi_n(\theta, \widehat{\psi}_\theta)$ vs θ generated from the standard Gaussian distribution when $n = 100$. Here the blue ticks, the green line segments, and the pink vertical line represent the data points, the knots of $\widehat{\psi}_{0,n}$ and $\widehat{\theta}_n$, and the point of maxima respectively.

be jointly convex in θ and ψ, existence of a maximizer does not lead to uniqueness. However, for all our simulations, the MLE turned out to be unique, even when the underlying density f_0 was skewed or non-log-concave. Therefore we will refer to $\widehat{\theta}_n$ as "the MLE" instead of "an MLE". However, even if $\widehat{\theta}_n$ is not unique, all our theorems will still hold for each version of $\widehat{\theta}_n$. Also, it is important to note that by Theorem 2.1(c) of Doss and Wellner (2019), for a particular choice of $\widehat{\theta}_n$, the estimator $\widehat{\psi}_{\widehat{\theta}_n}$, i.e. $\widehat{\psi}_{0,n}$ is unique. In other words, if (θ, ψ_1) and (θ, ψ_2) both are MLEs of (θ_0, ψ_0), we must have $\psi_1 = \psi_2$.

Now we discuss some finite-sample properties of $\widehat{\psi}_{0,n}$. The following theorem sheds some light on the structure of $\widehat{\psi}_{0,n}$. We find that $\widehat{\psi}_{0,n}$ is piecewise linear, and $\mathcal{S}(\widehat{\psi}_{0,n})$, i.e. the set of the knots of $\widehat{\psi}_{0,n}$, is a subset of the dataset. This theorem is a direct consequence of Theorem 2.1(c) of Doss and Wellner (2019).

Theorem 2.2. *For $\mathbb{F}_n$ non-degenerate, the MLE $\widehat{\psi}_{0,n}$ of ψ_0 is piecewise linear with knots belonging to a subset of the set $\{0, \pm|X_1 - \widehat{\theta}_n|, \ldots, \pm|X_n - \widehat{\theta}_n|\}$. Also, for*

$$x \notin [-|X - \widehat{\theta}_n|_{(n)}, |X - \widehat{\theta}_n|_{(n)}],$$

$\widehat{\psi}_{0,n}(x) = -\infty$. Moreover if $0 \notin \{\pm|X_1 - \widehat{\theta}_n|, \ldots, \pm|X_n - \widehat{\theta}_n|\}$, we have $\widehat{\psi}'_{0,n}(0\pm) = 0$.

Theorem 2.2 implies that if $0 \notin \{\pm|X_1 - \widehat{\theta}_n|, \ldots, \pm|X_n - \widehat{\theta}_n|\}$, by our definition of a knot in section 2.1.1, 0 is not a knot of $\widehat{\psi}_{0,n}$.

There is no closed form for $\widehat{\psi}_{0,n}$, though $\widehat{\psi}_{0,n}$ can be characterized by a family of inequalities. The following two theorems provide characterizations of $\widehat{\psi}_{0,n}$.

Theorem 2.3. *Suppose $\widehat{\theta}_n$ is the MLE of θ_0. Consider $g = e^{\psi} \in \mathcal{SLC}_0$. Then $(\widehat{\theta}_n, g)$ is the MLE of (θ_0, g_0) if and only if*

$$\int_{-\infty}^{\infty} \Delta(x - \widehat{\theta}_n) d\mathbb{F}_n(x) \leq \int_{-\infty}^{\infty} \Delta(x) g(x) dx$$

for all $\Delta : \mathbb{R} \mapsto \mathbb{R}$ such that $\psi + t\Delta \in \mathcal{SC}_0$ for some $t > 0$.

Note that the MLE in $\mathcal{LC}$ (Dümbgen and Rufibach, 2009) or $\mathcal{SLC}_0$ (Doss and Wellner, 2019) also satisfy similar characterizations. The proof of Theorem 2.3 follows from Theorem 2.2(c) of Doss and Wellner (2019).

Our next theorem exhibits another characterization of $\widehat{\psi}_{0,n}$. This theorem also follows as a direct consequence of Theorem 2.4(C) of Doss and Wellner (2019).

Theorem 2.4. *Suppose $\widehat{\theta}_n$ is the MLE of θ_0. For $g \in \mathcal{SLC}_0$, define*

$$\widehat{G}^+_{0,n}(x) = 2\int_x^{|X-\widehat{\theta}_n|_{(n)}} g(y)dy,$$

and

$$\mathbb{F}^+_{0,n}(x) = n^{-1}\sum_{i=1}^{n} 1\{|X-\widehat{\theta}_n|_{(i)} \geq x\}.$$

Then $(\widehat{\theta}_n, g)$ is the MLE of (θ_0, g_0) if and only if

$$\int_t^{|X-\widehat{\theta}_n|_{(n)}} \widehat{G}^+_{0,n}(x)dx \begin{cases} \leq \displaystyle\int_t^{|X-\widehat{\theta}_n|_{(n)}} \mathbb{F}^+_{0,n}(x)dx, & \text{if } t \in [0, |X-\widehat{\theta}_n|_{(n)}], \\ = \displaystyle\int_t^{|X-\widehat{\theta}_n|_{(n)}} \mathbb{F}^+_{0,n}(x)dx, & \text{if } t \in \mathcal{S}(\psi) \cap [0, |X-\widehat{\theta}_n|_{(n)}], \end{cases} \tag{2.17}$$

where $\psi = \log g$.

Figure 2.3 displays a plot of the function

$$h_\Delta(t) = \int_t^{|X-\widehat{\theta}_n|_{(n)}} \mathbb{F}^+_{0,n}(x)dx - \int_t^{|X-\widehat{\theta}_n|_{(n)}} \widehat{G}^+_{0,n}(x)dx, \tag{2.18}$$

constructed from a standard Gaussian sample of size 50 with $\psi = \widehat{\psi}_{0,n}$. Observe that as implied in Theorem 2.4, the function always stays above 0, and hits 0 only at the positive knots of $\widehat{\psi}_{0,n}$.

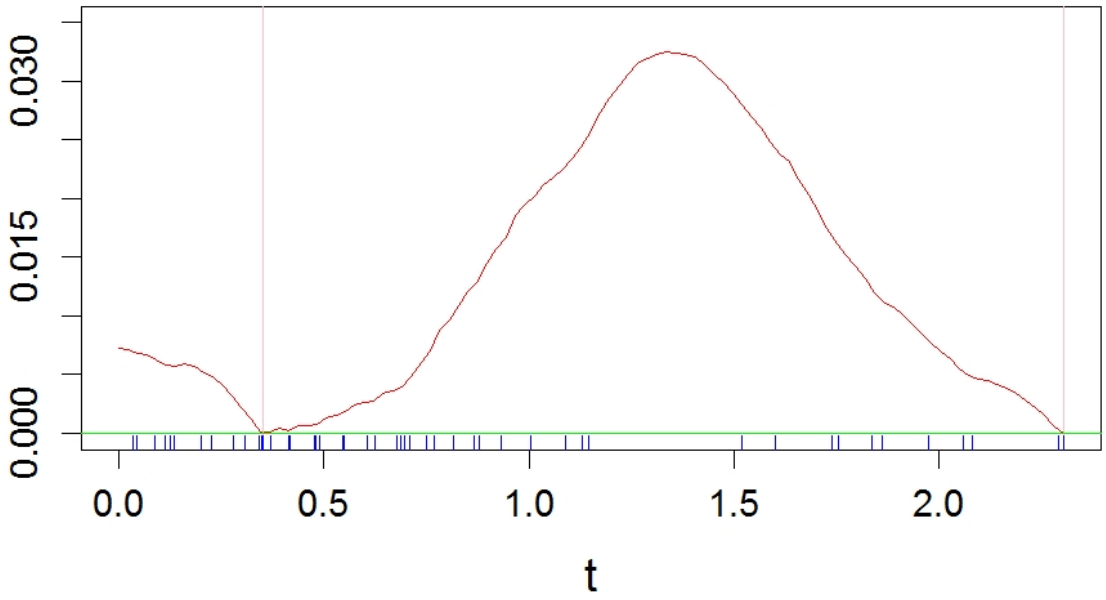


Figure 2.3: Plot of the function h_Δ defined in (2.18) versus t for a standard Gaussian sample of size 50. The pink vertical lines represents the positive knots of $\widehat{\psi}_{0,n}$, and the blue ticks represent the $|X_i - \widehat{\theta}_n|$'s.

The characterization theorems stated above lead to some interesting facts about the distribution function $\widehat{G}_m$ of $\widehat{g}_n$. In fact, we show that $\widehat{G}_m$ has a close connection with the empirical distribution function of the $|X_i - \widehat{\theta}_n|$'s.

Corollary 1. *Suppose $(\widehat{\theta}_n, \widehat{\psi}_{0,n})$ is the MLE of (θ_0, ψ_0). Let us denote the empirical distribution function of the $|X_i - \widehat{\theta}_n|$'s by $\mathbb{F}_{n,|X-\widehat{\theta}_n|}$. Then the following holds almost surely on $\mathcal{S}(\widehat{\psi}_{0,n}) \cap (0, \infty)$:*

$$\mathbb{F}_{n,|X-\widehat{\theta}_n|} - n^{-1} \leq 2\widehat{G}_m - 1 \leq \mathbb{F}_{n,|X-\widehat{\theta}_n|}.$$

The function $2\widehat{G}_m - 1$ can be interpreted as the distribution function of the random variable $|Y|$, where $Y \sim \widehat{G}_m$. Corollary 1 implies that at the knots, $\mathbb{F}_{n,|X-\widehat{\theta}_n|}$ and the distribution function $2\widehat{G}_m - 1$ of $|Y|$ differ at most by $1/n$. The proof of Corollary 1 follows from The proof of Corollary 2.7(C) of Doss and Wellner (2019).

Our next corollary provides an upper bound for the second central moment of $\widehat{G}_m$. This corollary is analogous to Corollary 2.8(C) of Doss and Wellner (2019).

Corollary 2. *Suppose $(\widehat{\theta}_n, \widehat{\psi}_{0,n})$ is the MLE of (θ_0, ψ_0). Then*

$$Var(\widehat{G}_m) = \int_{-\infty}^{\infty} x^2 d\widehat{G}_m(x) \leq \int_{-\infty}^{\infty} (x - \widehat{\theta}_n)^2 d\mathbb{F}_n(x),$$

where $\widehat{G}_m$ is the distribution function corresponding to the density $\widehat{g}_n = e^{\widehat{\psi}_{0,n}}$.

Since $\int x d\widehat{G}_m(x) = 0$, the proof follows by taking $\Delta = -x^2$ in Theorem 2.3.

2.3 Asymptotic properties

Having discussed how to construct the estimators of θ_0 and g_0, we move on to a discussion of the asymptotic properties of our estimators. In Sections 2.3.1 and 2.3.2 respectively, we study the asymptotic behavior of the estimators proposed in sections 2.2.1 and 2.2.2.

Before proceeding any further, we introduce some new notations. For any real valued function $h : \mathbb{R} \mapsto \mathbb{R}$, we let $||h||_p$ denote its L_p norm, i.e.

$$||h||_p = \left(\int_{-\infty}^{\infty} |h(x)|^p dx \right)^{1/p}, \quad p \geq 1.$$

We define the Wasserstein distance between two measures μ and ν on $\mathbb{R}$ by

$$d_W(\mu, \nu) = \int_{-\infty}^{\infty} |F(x) - G(x)| dx, \tag{2.19}$$

where F and G are the distribution functions corresponding to μ and ν respectively. This representation of $d_W(\mu, \nu)$ follows from Villani (2003), (2.48), page 75. By an abuse of

notation, sometime we will denote the above integral by $d_W(F, G)$ as well. For a sequence of distribution functions $\{F_n\}_{n\geq 1}$, we say F_n converges weakly to F, and write $F_n \to_d F$, if for all bounded continuous functions $h : \mathbb{R} \mapsto \mathbb{R}$, we have $\lim_{n\to\infty} \int h dF_n = \int h dF$.

2.3.1 Asymptotic properties of the one step estimators

Even though the primary objective of this section is to establish the almost asymptotic efficiency of $\tilde{\theta}_n$ for $f_0 \in \mathcal{P}_0$, we are equally interested in the asymptotic behavior of $\tilde{\theta}_n$ when $f_0 \notin \mathcal{P}_0$. Observe that even when f_0 violates the log-concavity assumption, as long as $f_0 \in \mathcal{S}_{\theta_0}$, the parameter θ_0 is still well-defined as the location parameter. Keeping that in mind, we consider the larger model $\mathcal{S}_{\theta_0} \cap \mathcal{P}$ where $\mathcal{P}$ was defined as the class of all densities on $\mathbb{R}$ in Section 2.1.1. Some questions immediately arise – what is the asymptotic behavior of $\tilde{\theta}_n$ when $f_0 \in \mathcal{S}_{\theta_0} \cap \mathcal{P}$, but may not be log-concave? Can $\tilde{\theta}_n$ still converge to θ_0, and if it does, under which conditions? In what follows, we try to address these questions. Following the terminology developed in Section 2.2.1, we let $\bar{f}_m$ denote the unconstrained log-concave MLE, and let $\tilde{g}_n$ denote any of the density estimators of g_0 developed in Section 2.2.1.

The log-concave projection theorem developed in Dümbgen *et al.* (2011), Xu and Samworth (2017), and Chen and Samworth (2013) plays a major role in our analysis. To be more precise, it provides us with tools to analyze the asymptotic behavior of $\tilde{g}_n$ when $f_0 \in \mathcal{S}_{\theta_0} \cap \mathcal{P}$, provided F_0 satisfies the following condition.

Condition A. *F_0 is a non-degenerate distribution function satisfying*

$$\int_{-\infty}^{\infty} |x| dF_0(x) < \infty.$$

In the special case when $f_0 \in \mathcal{P}_0$, Lemma 1 of Cule and Samworth (2010) implies that condition A is automatically satisfied. The restrictions in Condition A, which also feature in Dümbgen *et al.* (2011), Chen and Samworth (2013), and Xu and Samworth (2017), are required because otherwise $\bar{f}_m$ and $\widehat{g}_\theta$ may not be consistent for any density.

In view of the above, we restrict our attention to the following model

$$\mathcal{P}_1 = \left\{ f \in \mathcal{P} \;\middle|\; f \in \mathcal{S}_\theta \text{ for some } \theta \in \mathbb{R},\ \sup_x f(x) < \infty, \int_{-\infty}^{\infty} |x| f(x) dx < \infty \right\}. \tag{2.20}$$

We require the densities in $\mathcal{P}_1$ to be bounded to ensure the pointwise consistency of our density estimators. Note that any member of $\mathcal{P}_0$ is bounded by Lemma 1 of Cule and Samworth (2010). Also note that since $\mathcal{P}_1$ allows densities with infinite Fisher information, $\mathcal{P}_1 \subsetneqq \mathcal{P}_s$ defined in (2.2), which was considered by Stone (1975) and Beran (1974).

Before proceeding further, it is necessary to provide some background on the log-concave approximation theory mentioned above. For a distribution function F, denote by $\tilde{\phi}$ the maximizer of the criterion function

$$\omega(\phi, F) = \int_{-\infty}^{\infty} \phi(x) dF(x) - \int_{-\infty}^{\infty} e^{\phi(x)} dx \tag{2.21}$$

subject to $\phi \in \mathcal{C}$, the set of all concave functions on $\mathbb{R}$. Theorem 2.7 of Dümbgen *et al.* (2011) states that when F satisfies Condition A, a unique maximizer $\tilde{\phi}$ exists, and $\tilde{f} = \exp\left(\tilde{\phi}\right)$ integrates to 1. Also in the case when F has a density f, $\tilde{f}$ can be interpreted as the minimizer of the Kullback-Leiber divergence

$$d_{KL}(f, h) = \int_{-\infty}^{\infty} \log \frac{f(x)}{h(x)} f(x) dx$$

over all $h \in \mathcal{LC}$, the class of all log-concave densities (Dümbgen *et al.*, 2011). The distribution function $\tilde{F}$ of $\tilde{f}$ is regarded as the log-concave projection of F onto the space of all distributions with densities in $\mathcal{LC}$.

Now for $f_0 \in \mathcal{P}_1$ and ω as defined in (2.21), we denote

$$\tilde{\phi}_0 = \arg\max_{\phi \in \mathcal{C}} \omega(\phi, F_0), \quad \text{and} \quad \tilde{\psi}_0 = \tilde{\phi}_0(\cdot + \theta_0). \tag{2.22}$$

By our earlier arguments $e^{\tilde{\phi}_0}$ yields a density; which we denote by $\tilde{f}_0$, and write $\tilde{F}_0$ for the corresponding distribution function. We also set $\tilde{g}_0 = e^{\tilde{\psi}_0}$, and $\tilde{G}_0 = \tilde{F}_0(\cdot + \theta_0)$.

We are interested in $\tilde{f}_0$ because under Condition A, the unconstrained log-concave MLE $\bar{f}_m$ converges almost surely to $\tilde{f}_0$ in L_1. See the remark following Proposition 4 of Cule and Samworth (2010) or Theorem 2.15 of Dümbgen *et al.* (2011) for a proof. Therefore from (2.8) and (2.14), it is not hard to see that the limit of $\tilde{g}_n^{sym}$ and $\tilde{g}_n^{geo,sym}$ will be dependent on $\tilde{f}_0$. It turns out that for $f_0 \in \mathcal{P}_1$, both these density estimators are strongly L_1 consistent for $\tilde{g}_0$ when f_0 satisfies Condition A. For $f_0 \in \mathcal{P}_0$, the relations $\tilde{g}_0 = q_0$ and $\tilde{f}_0 = f_0$ follow from Dümbgen *et al.* (2011).

Next we consider the maximization of the criterion function $\omega(\phi, F)$ in ϕ over the constrained class $\mathcal{SC}_\theta$ for some fixed $\theta \in \mathbb{R}$. We denote

$$\overline{\phi}_\theta = \arg\max_{\phi \in \mathcal{SC}_\theta} \omega(\phi, F),$$

and let $\overline{\psi}_\theta = \overline{\phi}_\theta(\cdot + \theta)$. From (2.3) we note that

$$\overline{\psi}_\theta = \arg\max_{\psi \in \mathcal{SC}_0} \Psi(\theta, \psi, F).$$

Suppose Condition A holds. Then Proposition 2(iii) of Xu and Samworth (2017) entails that $\overline{\phi}_\theta$ exists, it is unique, and its exponential also yields a density just as $\tilde{\phi}$ does. Notice that when $F = \mathbb{F}_n$, exponentiating $\overline{\psi}_{\bar{\theta}_n}$ leads to the partial MLE estimator $\widehat{g}_{\bar{\theta}_n}$. Therefore we are interested in the latter maximizer; in particular we want to investigate its relation with the unconstrained maximizer $\tilde{\phi}$. Our next lemma shows that $\overline{\phi}_\theta$ is connected with the log-concave projection of the distribution function

$$F_\theta^{sym}(x) = 2^{-1}\Big(F(x) + 1 - F(2\theta - x)\Big). \tag{2.23}$$

Lemma 2.2. *Suppose F satisfies condition A. Then for any $\theta \in \mathbb{R}$,*

$$\arg\max_{\phi \in \mathcal{SC}_\theta} \omega(\phi, F) = \arg\max_{\phi \in \mathcal{C}} \omega(\phi, F_\theta^{sym}),$$

where is F_θ^{sym} is as defined in (2.23), and ω is the criterion function defined in (2.21).

For a general F satisfying condition A, $\overline{\phi}_\theta$ need not equal $\tilde{\phi}$ since $\overline{\phi}_\theta$ is symmetric while $\tilde{\phi}$ is not necessarily symmetric. However, when $F = F_\theta^{sym}$, or equivalently

$$F(\theta + x) + F(\theta - x) = 1, \tag{2.24}$$

we do have $\tilde{\phi} \in \mathcal{SC}_\theta$. The proof follows from either Lemma 2.2 or Lemma 2.12. Therefore if F satisfies (2.24), $\tilde{\phi}$ is still the maximizer of $\omega(\cdot, F)$ over the restricted class $\mathcal{SC}_\theta$. Note that if F has a density in $\mathcal{S}_\theta$, (2.24) holds. From the above discussion, we conclude that the projection of $f_0 \in \mathcal{P}_1$ onto $\mathcal{SLC}_{\theta_0}$ yields $\tilde{f}_0$.

Now let us return to the case of the partial MLE estimator $\widehat{g}_{\bar{\theta}_n}$. It turns out that if additionally $\bar{\theta}_n$ is strongly consistent, $\widehat{g}_{\bar{\theta}_n}$ converges almost surely to $\tilde{g}_0$ in L_1, where $\tilde{g}_0$ denotes the centered density $\tilde{f}_0(\cdot + \theta_0)$.

Indeed for $f_0 \in \mathcal{P}_1$, under the strong consistency of $\bar{\theta}_n$, all the estimators $\tilde{g}_n$ developed in section 2.3.1 converge to $\tilde{g}_0$ in L_1 except the smoothed symmetrized estimator $(\widehat{f}_n^{sym})^{sm}$. We discover that the L_1 limit of $(\widehat{f}_n^{sym})^{sm}$ may not equal $\tilde{g}_0$ unless $f_0 = \tilde{g}_0$ almost everywhere on $\mathbb{R}$. We will elaborate a little bit on this estimator.

The different behavior of $(\widehat{f}_n^{sym})^{sm}$ stems from the fact that the smoothing parameter b_n defined in (2.9) does not necessarily decrease to 0 for large n unless f_0 is log-concave. Also, Condition A is not sufficient for $\lim_{n\to\infty} b_n$ to exist, for which the finiteness of the second central moment of f_0 is required, i.e. we would need

$$\int_{-\infty}^{\infty} z^2 f_0(z) dz < \infty. \tag{2.25}$$

Now, note that the moments of $\bar{f}_m$ converge almost surely to those of $\tilde{f}_0$ by Theorem 4 of Cule and Samworth (2010), and the first moment of $\tilde{f}_0$ equals that of f_0 by Remark 2.3 of Dümbgen *et al.* (2011). Therefore (2.9) and (2.25) lead to

$$b_n^2 \to_{a.s.} \tilde{b}^2 = \int_{-\infty}^{\infty} z^2 (f_0(z) - \tilde{f}_0(z)) dz. \tag{2.26}$$

Remark 2.3 of Dümbgen *et al.* (2011) ensures that $\tilde{b}$ is non-negative. Denote by $\tilde{f}_0^{sm}$ the convolution of $\tilde{f}_0$ and a centered Gaussian density with variance $\tilde{b}^2$. This density $\tilde{f}_0^{sm}$ is relevant for us because Theorem 2 of Chen and Samworth (2013) implies that the smoothed log-concave MLE $\widehat{f}_n^{sm}$ in (2.11) is strongly consistent for $\tilde{f}_0^{sm}$ in L_1 metric. Using the connection between $(\widehat{f}_n^{sym})^{sm}$ and $\widehat{f}_n^{sm}$ shown in (2.12), we can show that $(\widehat{f}_n^{sym})^{sm}$ is also strongly L_1 consistent for $\tilde{g}_0^{sm} = \tilde{f}_0^{sm}(\cdot - \theta_0)$.

We denote the distribution function corresponding to $\tilde{g}_0^{sm}$ by $\tilde{G}_0^{sm}$, and denote by $\tilde{\psi}_0^{sm}$ and $\tilde{\phi}_0^{sm}$ the log-densities $\log \tilde{g}_0^{sm}$ and $\log \tilde{f}_0^{sm}$ respectively. In the special case when f_0 is log-concave, $\tilde{f}_0 = f_0$, thus leading to $\tilde{b} = 0$. As a result, all $\tilde{g}_n$'s are strongly L_1 consistent estimators of g_0 when the model is correctly specified. Lemma 2.3 summarizes the above discussion.

Lemma 2.3. *Assume $f_0 \in \mathcal{P}_1$. Suppose $\tilde{g}_n$ is one of the estimators of g_0 defined in section 2.2.1, and the preliminary estimator of location $\bar{\theta}_n$ is a consistent estimator of θ_0. Let $\tilde{g}_0 = \exp\Big(\tilde{\psi}_0\Big)$, where $\tilde{\psi}_0$ was defined in (2.22). Then for $\tilde{g}_n = \tilde{g}_n^{sym}$, $\widehat{g}_{\bar{\theta}_n}$, and $\bar{f}_m^{geo,sym}$, it follows that*

$$||\tilde{g}_n - \tilde{g}_0||_1 \to_p 0. \tag{2.27}$$

Suppose f_0 additionally satisfies (2.25). Then we also have

$$||(\widehat{f}_n^{sym})^{sm} - \tilde{g}_0^{sm}||_1 \to_p 0,$$

where $\tilde{g}_0^{sm}$ is the convolution of $\tilde{g}_0$ and a centered Gaussian density with variance $\tilde{b}^2$ defined in (2.9). If $\bar{\theta}_n$ is an strongly consistent estimator of θ_0, the above convergences hold almost surely.

Proving Lemma 2.3 is a key step for us because combined with Proposition 2 of Cule and Samworth (2010), Lemma 2.3 leads to further consistency results involving $\tilde{\psi}_n$, $\tilde{\psi}_n'$ etc. In particular, we find that when $\tilde{g}_n$ is logconcave, $\tilde{\psi}_n$ converges uniformly to $\tilde{\psi}_0$ over all compact sets in $\text{int}(\text{supp}(g_0))$ with probability one. We can derive a very similar result concerning the

convergence of $\tilde{\psi}_n'(x)$ to $\tilde{\psi}_0'(x)$ except that at the derivative level, the convergence can only be proved at the continuity points of $\tilde{\psi}_0'$. The above-mentioned results continue to hold for $\tilde{g}_n^{sym}$ or $(\widehat{f}_n^{sym})^{sm}$ despite their lack of log-concavity, though for the special case of $(\widehat{f}_n^{sym})^{sm}$, the limit changes to $\tilde{\psi}_0^{sm}$ from $\tilde{\psi}_0$. The above discussion can be formulated in the following consistency theorem.

Theorem 2.5. *Assume $f_0 \in \mathcal{P}_1$. Suppose $\tilde{g}_n$ is one of the estimators of g_0 defined in section 2.2.1, where $\bar{\theta}_n$ is a strongly consistent estimator of θ_0. Let $\{y_n\}_{n\geq 1}$ be any sequence of random variables converging to 0 in probability. Then for $\tilde{g}_n = \tilde{g}_n^{sym}$, $\widehat{g}_{\bar{\theta}_n}$, $\tilde{g}_n^{geo,sym}$, or $\bar{f}_m(\bar{\theta}_n \pm \cdot)$, on any compact set $K \subset int(supp(g_0))$ we have,*

(A)

$$\sup_{x\in K} |\tilde{g}_n(x + y_n) - \tilde{g}_0(x)| \to_p 0.$$

(B)

$$\sup_{x\in K} |\tilde{\psi}_n(x + y_n) - \tilde{\psi}_0(x)| \to_p 0.$$

(C) Suppose $x \in int(dom(\tilde{\psi}_0))$ is a continuity point of $\tilde{\psi}_0'$. Then

$$\tilde{\psi}_n'(x + y_n) \to_p \tilde{\psi}_0'(x).$$

Suppose f_0 also satisfies (2.25). Then for $\tilde{g}_n = (\widehat{f}_n^{sym})^{sm}$ or $\widehat{f}_n^{sm}(\bar{\theta}_n \pm \cdot)$, (A)-(C) hold for any compact set $K \subset \mathbb{R}$, with $\tilde{g}_0$ and $\tilde{\psi}_0$ replaced by $\tilde{g}_0^{sm}$ and $\tilde{\psi}_0^{sm}$ respectively. If $\bar{\theta}_n$ is a strongly consistent estimator of θ_0, and $y_n \to_{a.s.} 0$, then the above convergences hold almost surely.

Though the log-concave density $\tilde{g}_0$ is continuous on its support, it can have jump discontinuities at the endpoints of its support. In case $\tilde{g}_0$ is continuous on $\mathbb{R}$, using Proposition 2(c) of Cule and Samworth (2010), part A of Theorem 2.5 can be strengthened to yield

$$\sup_{x\in\mathbb{R}} |\tilde{g}_n(x + y_n) - \tilde{g}_0(x)| \to_{a.s.} 0.$$

Now we present a corollary to Theorem 2.5. Consider the case when the model is correctly specified, i.e. $f_0 \in \mathcal{P}_0$, implying $\tilde{f}_0 = \tilde{f}_0^{sm} = f_0$. We show that if ψ_0 is unbounded, $\tilde{\psi}_n'$ imitates ψ_0'. However, we do not know whether the sequence of concave functions $\tilde{\psi}_n'$ stays uniformly bounded inside the domain when ψ_0' is bounded.

Corollary 3. *Suppose $f_0 \in \mathcal{P}_0$. Then under the assumptions of Theorem 2.5, for $\tilde{g}_n = \tilde{g}_n^{sym}$, $\widehat{g}_{\bar{\theta}_n}$, or $\tilde{g}_n^{geo,sym}$, if ψ_0 satisfies*

$$\sup_{x \in int(dom(\psi_0))} |\psi_0'(x)| = \infty,$$

we have

$$\limsup_n \sup_{x \in int(dom(\tilde{\psi}_n))} |\tilde{\psi}_n'(x)| = \infty \quad a.s.$$

If additionally (2.25) holds, then the above holds for $\tilde{g}_n = (\widehat{f}_n^{sym})^{sm}$ as well.

Now we are in a position to address the consistency of $\widehat{\mathcal{I}}_n(\eta)$ defined in (2.15), which is an estimator of the Fisher information. The limit of

$$\widehat{\mathcal{I}}_n(\eta) = \int_{\bar{\theta}_n - \xi_n}^{\bar{\theta}_n + \xi_n} \tilde{\psi}_n'(x - \bar{\theta}_n)^2 d\mathbb{F}_n(x)$$

depends on the truncation parameter η via $\xi_n = \tilde{G}_n^{-1}(1-\eta)$. Let us denote $\xi_0 = \tilde{G}_0^{-1}(\eta)$. We find that for $\tilde{g}_n \neq (\widehat{f}_n^{sym})^{sm}$, $\widehat{\mathcal{I}}_n(\eta)$ is a consistent estimator of $\mathcal{I}_{f,g}(\eta)$, where

$$\mathcal{I}_{f,g}(\eta) = \int_{\tilde{F}_0^{-1}(\eta)}^{\tilde{F}_0^{-1}(1-\eta)} \tilde{\phi}_0'(x)^2 f_0(x) dx = \int_{-\xi_0}^{\xi_0} \tilde{\psi}_0'(x)^2 g_0(z) dz. \tag{2.28}$$

$\mathcal{I}_{f,g}(\eta)$ can be thought of as a truncated Fisher information under model miss-specification. On the other hand when $\tilde{g}_n = (\widehat{f}_n^{sym})^{sm}$, $\widehat{\mathcal{I}}_n(\eta)$ converges almost surely to

$$\begin{aligned} \mathcal{I}_{f,g}^{sm}(\eta) &= \int_{(\tilde{F}_0^{sm})^{-1}(\eta)}^{(\tilde{F}_0^{sm})^{-1}(1-\eta)} (\tilde{\phi}_0^{sm})'(x)^2 f_0(x) dx \\ &= \int_{-\xi_0^{sm}}^{\xi_0^{sm}} (\tilde{\psi}_0^{sm})'(z)^2 g_0(z) dz, \end{aligned} \tag{2.29}$$

where ξ_0^{sm} is the $(1-\eta)$-th quantile of $\tilde{G}_0^{sm}$. Note that for f_0 log-concave, $\mathcal{I}_{f,g}(\eta) = \mathcal{I}_{f,g}^{sm}(\eta)$, and they equal

$$\mathcal{I}_{f_0}(\eta) = \int_{F_0^{-1}(\eta)}^{F_0^{-1}(1-\eta)} \phi_0'(x)^2 dF_0(x). \tag{2.30}$$

The consistency of $\widehat{\mathcal{I}}_n(\eta)$ along with the consistency results in Lemma 2.5 lead to the desired consistency of $\tilde{\theta}_n$ for $f_0 \in \mathcal{P}_1$. Therefore we conclude that $\tilde{\theta}_n$ is robust to the violation of log-concavity of f_0 as long as f_0 is symmetric about some $\theta \in \mathbb{R}$, and has finite first moment.

Lemma 2.4. *Under the conditions of Theorem 2.5, for $\tilde{g}_n = \tilde{g}_n^{sym}$, $\widehat{g}_{\bar{\theta}_n}$, and $\tilde{g}_n^{geo,sym}$, we have $\widehat{\mathcal{I}}_n(\eta) \to_p \mathcal{I}_{f,g}(\eta)$, which was defined in (2.28). Also, $\widehat{\theta}_n \to_p \theta_0$ in these cases.*

When $\tilde{g}_n = (\widehat{f}_n^{sym})^{sm}$, if additionally (2.25) holds, we have $\widehat{\mathcal{I}}_n(\eta) \to_p \mathcal{I}_{f,g}^{sm}(\eta)$ (defined in (2.29)), and $\widehat{\theta}_n \to_p \theta_0$. When $\bar{\theta}_n$ is a strongly consistent estimator of θ_0, the above convergences hold almost surely.

Now we focus on the rate of convergence of $\tilde{\theta}_n - \theta_0$. In this case we make the assumption that $\bar{\theta}_n$ is $\sqrt{n}$-consistent. A similar assumption was made by Stone (1975). Theorem 2.6 provides an asymptotic representation of $\sqrt{n}(\tilde{\theta}_n - \theta_0)$ for $f_0 \in \mathcal{P}_1$ under $\sqrt{n}$-consistency of $\bar{\theta}_n$.

Theorem 2.6. *Suppose $f_0 \in \mathcal{P}_1$ is absolutely continuous. $\bar{\theta}_n$ is a $\sqrt{n}$-consistent estimator of θ_0. For some $\eta \in (0, 1/2)$, let $\tilde{\theta}_n$ be the corresponding truncated one-step estimator defined in (2.16). Suppose $\tilde{g}_n = \tilde{g}_n^{sym}$, $\widehat{g}_{\bar{\theta}_n}$ or $\tilde{g}_n^{geo,sym}$. Then with $\xi_0 = \tilde{G}_0^{-1}(1-\eta)$, under the conditions of Theorem 2.5, the following assertion holds:*

$$\sqrt{n}(\tilde{\theta}_n - \theta_0) = \int_{\theta_0-\xi_0}^{\theta_0+\xi_0} \frac{\tilde{\psi}_0'(x-\theta_0)}{\mathcal{I}_{f,g}(\eta)} d\mathbb{Z}_n(x) + \sqrt{n}(1-\gamma_\eta)(\bar{\theta}_n - \theta_0) + o_p(1), \tag{2.31}$$

where $\mathbb{Z}_n = \sqrt{n}(\mathbb{F}_n - F_0)$,

$$\gamma_\eta = 1 - \frac{\int_{-\xi_0}^{\xi_0} \tilde{\psi}_0'(z)\Big(\tilde{\psi}_0'(z) - \psi_0'(z)\Big) g_0(z) dz}{\mathcal{I}_{f,g}(\eta)}, \tag{2.32}$$

$\xi_0 = \tilde{G}_0^{-1}(\eta)$, *and* $\mathcal{I}_{f,g}(\eta)$ *is as defined in* (2.28). *When* $\tilde{g}_n = (\widehat{f}_n^{sym})^{sm}$, *we have*

$$\sqrt{n}(\tilde{\theta}_n - \theta_0) = \int_{\theta_0-\xi_0}^{\theta_0+\xi_0} \frac{(\tilde{\psi}_0^{sm})'(x-\theta_0)}{\mathcal{I}_{f,g}^{sm}(\eta)} d\mathbb{Z}_n(x) + \sqrt{n}(1-\gamma_\eta^{sm})(\bar{\theta}_n - \theta_0) + o_p(1), \tag{2.33}$$

where

$$\gamma_\eta^{sm} = 1 - \frac{\int_{-\xi_0^{sm}}^{\xi_0^{sm}} (\tilde{\psi}_0^{sm})'(z)\Big((\tilde{\psi}_0^{sm})'(z) - \psi_0'(z)\Big) g_0(z)dz}{\mathcal{I}_{f,g}^{sm}(\eta)}, \tag{2.34}$$

ξ_0^{sm} *is the* $(1-\eta)$*-th quantile of* $\tilde{G}_0^{sm}$, *and* $\mathcal{I}_{f,g}^{sm}(\eta)$ *is as defined in*(2.29). *In the special case when* f_0 *is log-concave, or more specifically when* $f_0 \in \mathcal{P}_0$, (2.32) *and* (2.34) *reduce to*

$$\sqrt{n}(\tilde{\theta}_n - \theta_0) \to_d N(0, \mathcal{I}_{f_0}^{-1}(\eta)) \tag{2.35}$$

where $\mathcal{I}_{f_0}(\eta)$ *is as defined in* (2.30).

Theorem 2.6 shows that if $\bar{\theta}_n$ is an asymptotically linear estimator of θ_0, then so is $\tilde{\theta}_n$. Moreover, the central limit Theorem, (2.30), and (2.29) entail that the first terms in the expansion of $\sqrt{n}(\tilde{\theta}_n - \theta_0)$ in (2.31) and (2.33) converge weakly to mean zero Gaussian random variables with variances $\mathcal{I}_{f,g}^{-1}$ and $(\mathcal{I}_{f,g}^{sm})^{-1}$ respectively. From (2.31) and (2.33) it is also clear that for $f_0 \notin \mathcal{P}_0$, the asymptotic distribution of $\sqrt{n}(\tilde{\theta}_n - \theta_0)$ can depend on $\bar{\theta}_n$. This dependence is quantified by γ_η or γ_η^{sm}. For $\tilde{g}_n \neq (\widehat{f}_n^{sym})^{sm}$, we observe that Cauchy-Schwartz inequality leads to

$$\begin{aligned}
|1-\gamma_\eta| &= \frac{\int_{-\xi_0}^{\xi_0} \tilde{\psi}_0'(z)(\tilde{\psi}_0'(z) - \psi_0'(z)) g_0(z)dz}{\mathcal{I}_{f,g}(\eta)} \\
&\leq \frac{\left(\int_{-\xi_0}^{\xi_0} \tilde{\psi}_0'(z)^2 g_0(z)dz\right)^{1/2}}{\mathcal{I}_{f,g}(\eta)} \left(\int_{-\xi_0}^{\xi_0} (\tilde{\psi}_0'(z) - \psi_0'(z))^2 g_0(z)dz\right)^{1/2} \\
&\leq \left(\mathcal{I}_{f,g}(\eta)^{-1} \int_{-\infty}^{\infty} (\tilde{\psi}_0'(z) - \psi_0'(z))^2 g_0(z)dz\right)^{1/2},
\end{aligned}$$

where the last inequality follows from (2.28). The above bound on $|1-\gamma_\eta|$ hints that when the $L_2(P_0)$ distance between $\tilde{\psi}_0'$ and ψ_0' is large, the dependence of the asymptotic distribution of $\sqrt{n}(\tilde{\theta}_n - \theta_0)$ on the choice of $\bar{\theta}_n$ is non-negligible. We can draw a similar conclusion for γ_η^{sm} replacing $\tilde{\psi}_0$ by $\tilde{\psi}_0^{sm}$.

Theorem 2.6 ensures that $\sqrt{n}(\tilde{\theta}_n - \theta_0)$ is asymptotically normal with variance $\mathcal{I}_{f_0}(\eta)^{-1}$. Thus $\mathcal{I}_{f_0}(\eta)$ is a crucial quantity for us since it is inversely proportional to the asymptotic efficiency of $\tilde{\theta}_n$. For the rest of this section, we restrict our attention to the model $\mathcal{P}_0$, and describe in greater detail how η effects the asymptotic efficiency of $\tilde{\theta}_n$.

It is natural to expect $\mathcal{I}_{f_0}(\eta)$ to be close to $\mathcal{I}_{f_0}$ if $\eta \in (0, 1/2)$ is small. However, the magnitude of this closeness is likely to be controlled by the underlying density f_0. Keeping that in mind, we seek to investigate how the ratio $\mathcal{I}_{f_0}(\eta)/\mathcal{I}_{f_0}$ behaves as a function of η for different choices of f_0. Our choices of f_0 include the standard normal, Laplace and logistic densities. Note that all these densities are members of $\mathcal{P}_0$ with center of symmetry $\theta_0 = 0$. We have plotted them in Figure 2.5 for convenience.

Another choice for f_0 is the standard symmetrized beta density with parameter $r > 2$ (see Figure 2.5), which we define by

$$f_0(x) \equiv f_{0,r}(x) = \frac{\Gamma((3+r)/2)}{\sqrt{\pi r}\Gamma(1+r/2)}\left(1 - \frac{x^2}{r}\right)^{r/2} 1_{[-\sqrt{r},\sqrt{r}]}(x), \quad r > 0, \tag{2.36}$$

where Γ is the usual Gamma function. It is straightforward to verify that in this case the score equals

$$\phi_0'(x) = \frac{-x}{1 - x^2/r} 1_{[-\sqrt{r},\sqrt{r}]}(x).$$

Some computation shows that $r \leq 2$ leads to $\mathcal{I}_{f_{0,r}} = \infty$. However for $r > 2$,

$$\mathcal{I}_{f_0} = \frac{r\Gamma\left(\frac{r}{2} - 1\right)\Gamma\left(\frac{3+r}{2}\right)}{2\Gamma\left(\frac{r}{2} + 1\right)\Gamma\left(\frac{1+r}{2}\right)} < \infty,$$

and $f_0 \in \mathcal{P}_0$. Figure 2.4 displays the plot of $\mathcal{I}_{f_{0,r}}$ versus r for the symmetrized beta density, which depicts that $\mathcal{I}_{f_{0,r}}$ decreases steeply for $r > 2$. This finding is consistent with $\mathcal{I}_{f_0}$ being ∞ when f_0 is the uniform density on $[-1, 1]$.

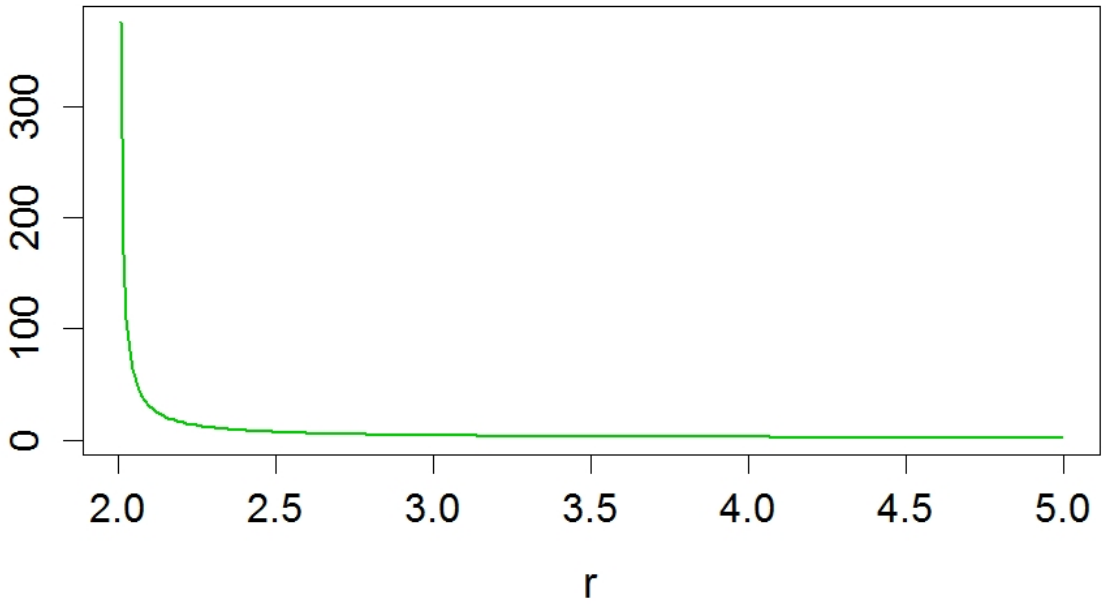


Figure 2.4: Plot of Fisher information $\mathcal{I}_{f_{0,r}}$ versus r, for $f_{0,r}$ defined in (2.36)

Now we examine the ratio $\mathcal{I}_{f_0}(\eta)/\mathcal{I}_{f_0}$, which gives the asymptotic efficiency of the truncated one step estimators in correctly specified models. Figure 2.6 illustrates the plot of $\mathcal{I}_{f_0}(\eta)/\mathcal{I}_{f_0}$ versus η for the above-mentioned densities. It may not be observable from the plot but we verified that for logistic, normal, and the Laplace distribution, the asymptotic efficiency is greater than 0.98 for $\eta \in (0, 0.001)$. However Figure 2.6 illustrates that for the symmetrized beta distribution though, the situation is somewhat different. We considered

symmetrized beta distributions with parameter $r = 2.1$, 2.5, 3.5, and 4.5. From Figure 2.6, we observe that for symmetrized beta densities, the asymptotic efficiency is quite small even for smaller values of η. This trend becomes more extreme as r decreases to 2.1. We calculated that for $r = 2.1$, as η goes down from 0.01 to 10^{-6}, the ratio $\mathcal{I}_{f_0}(\eta)/\mathcal{I}_{f_0}$ only increases from 0.05 to 0.233.

The aberration for the symmetrized beta case can be explained inspecting Figure 2.5, which elucidates that as r decreases, the slope of the density near the boundary of the support becomes sharper. As a result, the score, which equals f_0'/f_0, also becomes large. In particular,

$$\phi_0'(x) \sim \frac{\sqrt{r}/2}{1 - |x|/\sqrt{r}} \qquad \text{as } |x| \to \sqrt{r}.$$

Therefore $\mathcal{I}_{f_0}$ grows very fast as r decreases. Since $\mathcal{I}_{f_0}(\eta)$ does not depend on the value of the score near the boundary, it does not grow as fast as $\mathcal{I}_{f_0}$; the difference being more striking for smaller r. Therefore truncation leads to greater loss of asymptotic efficiency for symmetrized beta densities, especially for smaller values of r, which is reflected in Figure 2.6.

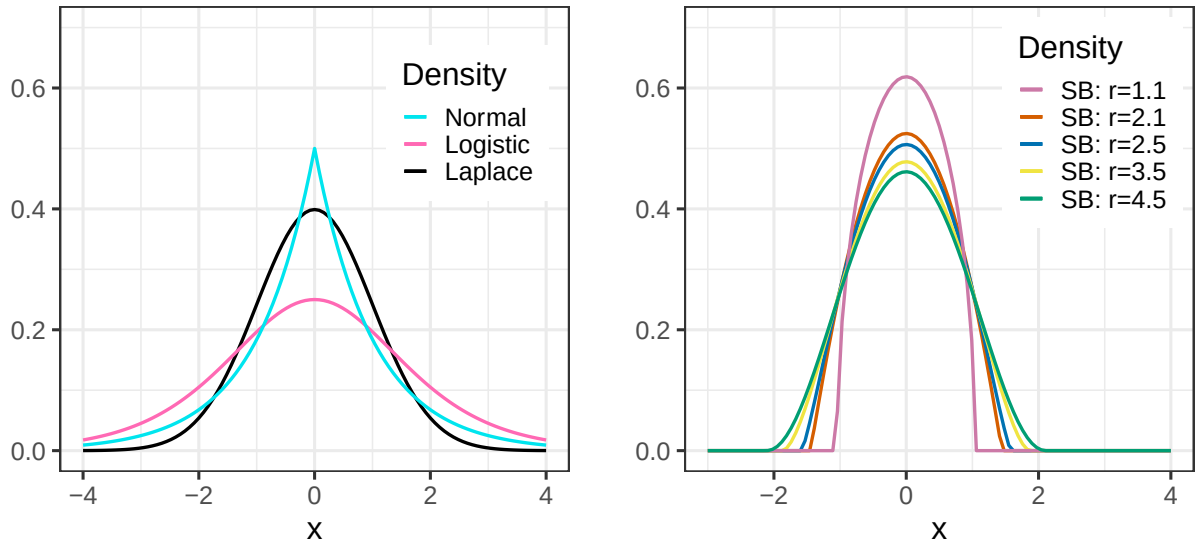


Figure 2.5: The left panel displays the densities of standard Laplace, normal and logistic distribution. The right panel displays the symmetrized beta (SB) densities $f_{0,r}$, which was defined in (2.36), for different values of r.

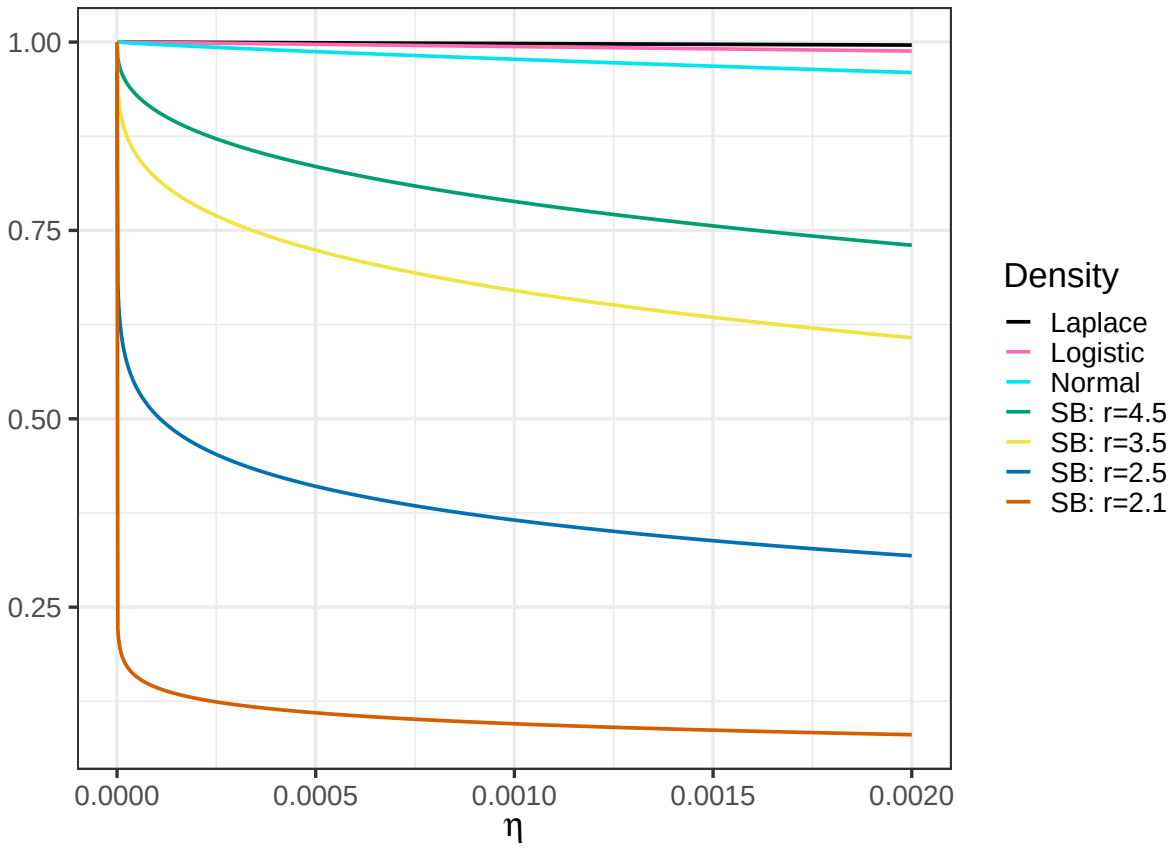


Figure 2.6: Plot of $\mathcal{I}_{f_0}(\eta)/\mathcal{I}_{f_0}$ versus η for different distributions

Finally, even though we do not know the rate of convergence of the untruncated estimator in (2.7), we conjecture that this estimator is fully efficient. Our simulation study in Section 2.4 also supports this conjecture.

2.3.2 Asymptotic properties of the MLE

In this section we explore the asymptotic properties of the MLE $\widehat{\theta}_n$ and $\widehat{g}_n$. When $f_0 \in \mathcal{P}_1$, both $\widehat{\theta}_n$ and $\widehat{g}_n$ enjoy consistency properties similar to $\tilde{\theta}_n$ and the $\tilde{g}_n$'s in Section 2.2.1. Therefore the robustness properties of $\widehat{\theta}_n$ and the one step estimator $\tilde{\theta}_n$ are comparable as long as $f_0 \in \mathcal{P}_1$. However in case of the MLE, even when f_0 is not symmetric, we can show that $\widehat{\theta}_n$ converges almost surely to some limit under some quite minimal conditions. Much as in Section 2.3.1, we again appeal to the log-concave projection theory developed by Dümbgen *et al.* (2011), Cule and Samworth (2010) and Xu and Samworth (2017) to analyze these asymptotic properties.

We introduce some new notation for the functions related to log-concave projection. For any distribution function F, let us denote

$$(\theta^*(F), \psi^*(F)) = \underset{\theta \in \mathbb{R}, \psi \in \mathcal{SC}_0}{\arg\max}\, \Psi(\theta, \psi, F), \tag{2.37}$$

where Ψ was defined in (2.3). This maximization problem is different from those considered in Section 2.3.1 because in the previous cases, the maximization was only in ψ. Proposition 1 below entails that Condition A is sufficient for the existence of the maximizers. Following the same argument as in Theorem 2.1(c) of Doss and Wellner (2019), one can show that when $\psi^*(F)$ exists, $g^*(F) = e^{\psi^*(F)}$ is a density. Denote by $G^*(F)$ the corresponding distribution function. We also denote $\phi^*(F) = \psi^*(F)(\cdot - \theta^*(F))$, and write $f^*(F)$ for $e^{\phi^*(F)}$. The corresponding distribution function will be denoted by $F^*(F)$, which has the interpretation as the projection of F onto the space of all distribution functions with densities in $\mathcal{SLC}_{\theta^*(F)}$.

Proposition 1. *Suppose F satisfies Condition A. Then $L(F) < \infty$, and $\theta^*(F)$ and $\psi^*(F)$ exist for F, where $\theta^*(F)$ and $\psi^*(F)$ were defined in (2.37).*

Our next lemma provides a representation of $\theta^*(F)$ and $\psi^*(F)$ in terms of the log-concave projection in (2.22). Lemma 2.5 is a direct consequence of Lemma 2.2.

Lemma 2.5. *Suppose a distribution function F satisfies condition A. Then*

$$\theta^*(F) = \arg\max_{\theta\in\mathbb{R}} \left(\max_{\phi\in\mathcal{C}} \omega(\phi, F_\theta^{sym}) \right),$$

and

$$\psi^*(F) = \arg\max_{\phi\in\mathcal{C}} \omega\left(\phi, F_{\theta^*(F)}^{sym}\right),$$

where F_θ^{sym} is defined in (2.23) for any $\theta \in \mathbb{R}$, and ω is the criterion function defined in (2.21).

The benefit of the representation of $\theta^*(F)$ and $\psi^*(F)$ in terms of the unrestricted log-concave projection defined in (2.22) lies in the fact that characterizations and tools for computing the latter are available (cf. Dümbgen *et al.*, 2011, Remark 2.3, Theorem 2.7, Remark 2.10, Remark 2.11).

Note that neither Proposition 1 nor Lemma 2.5 provides any information about the uniqueness of $\theta^*(F)$ and $\psi^*(F)$. However, the uniqueness of $\theta^*(F)$ is critical for the consistency of $\widehat{\theta}_n$. Therefore we are interested in those distribution functions which satisfy the following condition.

Condition B. *The maximizer $\theta^*(F)$ defined in (2.37) is unique for distribution function F.*

Though Condition B does not say anything about the uniqueness of $\psi^*(F)$, from Theorem 2.1(c) of Doss and Wellner (2019) or Proposition 2 of Xu and Samworth (2017), it follows that if $\theta^*(F)$ is unique, so is $\psi^*(F)$. We have not yet found an example of F which violates Condition B, but satisfies condition A. However, it is easy to find density-classes satisfying both Conditions A and B, a trivial example being distributions satisfying (2.24) for some $\theta \in \mathbb{R}$.

Lemma 2.6. *Suppose a distribution function F satisfies (2.24) for some $\theta \in \mathbb{R}$, and set $G = F(\cdot + \theta)$. Then condition B is satisfied with*

$$\theta^*(F) = \theta,$$

and

$$\psi^*(F) = \arg\max_{\psi \in \mathcal{SC}_0} \Psi(0, \psi, G) = \arg\max_{\phi \in \mathcal{C}} \omega(\phi, G),$$

where ω is the criterion function defined in (2.21).

The uniqueness of $\psi^*(F)$ in Lemma 2.6 follows from that of $\theta^*(F)$. The following example, which is taken from Example 2.9 of Dümbgen *et al.* (2011), helps in understanding the implications of Lemma 2.6.

Example 1. Consider a rescaled version of Student's t_2 distribution, whose density takes the form $f(x) = g(x - \theta)$, with g satisfying

$$g(x) = 2^{-1}(1 + x^2)^{-3/2}, \quad x \in \mathbb{R}.$$

Lemma 2.6 and (2.21), (2.22) indicate that $\theta^*(F) = 0$, and that $\psi^*(F)$ can be found using the unrestricted log-concave projection theorem developed in Dümbgen *et al.* (2011). Using Example 2.9 of Dümbgen *et al.* (2011), we derive that $\psi^*(F)(x) = -|x| - \log 2$, which corresponds to the Laplace density. From Lemma 2.6, $\theta^*(F) = \theta$ directly follows.

Example 2. Consider the bimodal density

$$f(x) = 2^{-1}(\varphi(x - 2) + \varphi(x + 2)), \tag{2.38}$$

where φ is the standard Gaussian density. Since $f \in \mathcal{S}_0$, Lemma 2.6 implies that $\theta^*(F) = 0$. Similar to Example 1, by Lemma 2.6 and (2.22), $\psi^*(F)$ equals $\arg\max_{\phi \in \mathcal{C}} \omega(\phi, F)$. Therefore the properties of the log-concave projection illustrated in Dümbgen *et al.* (2011) can be used here to calculate $\psi^*(F)$. Remark 2.11 of Dümbgen *et al.* (2011) implies that there exists

$z > 2$ such that the projection $f^*(F)$ equals f on $\mathbb{R} \setminus [-z, z]$, and is constant on $[-z, z]$. In addition, since $f^*(F) \in \mathcal{SLC}_0$, we have

$$\int_{-\infty}^{\infty} f^*(F)(x)dx = 2zf(z) + 2\Big(1 - F(z)\Big), \tag{2.39}$$

which implies

$$1 = z\Big(\varphi(z-2) + \varphi(z+2)\Big) + 2 - \Phi(z-2) - \Phi(z+2),$$

where Φ is the standard Gaussian distribution function. The solution to the above equation is approximately 2.83.

Similarly, consider the following mixture of Laplace densities:

$$f(x) = 4^{-1}\Big(e^{-|x-2|} + e^{-|x+2|}\Big). \tag{2.40}$$

Its projection $f^*(F)$ can also be found by Remark 2.11 of Dümbgen *et al.* (2011) which implies that that $f^*(F)$ equals f on $\mathbb{R} \setminus [-z, z]$, and is constant on $[-z, z]$, where $z > 2$. Using (2.39), which holds in this case as well, we find that $z = 2.61$. Figure 2.7 displays the above bimodal densities and their symmetric log-concave projections.

Our next lemma gives some insight into the domain of $\psi^*(F)$.

Lemma 2.7. *Suppose Condition A and B hold for a distribution function F. Then $\theta^*(F) \in \overline{J(F)}$, where $J(F) = \{0 < F < 1\}$. Moreover, letting a and b denote the left and right end-points of the set $J(F)$ respectively, we have,*

$$int(dom(\psi^*(F))) = (-d, d),$$

where $d = (b - \theta^*(F)) \wedge (\theta^*(F) - a)$.

Lemma 2.7 underscores that for a general F_0, the domain of ψ_0 is contained the domain of $\psi^*(F_0)$. In particular when $\psi_0 \in \mathcal{S}_0$, which is the case for $f_0 \in \mathcal{P}_1$, we have $b - \theta^*(F) = \theta^*(F) - a$, leading to

$$\text{int}(\text{dom}(\psi^*(F_0))) = \text{int}(\text{dom}(\psi_0)).$$

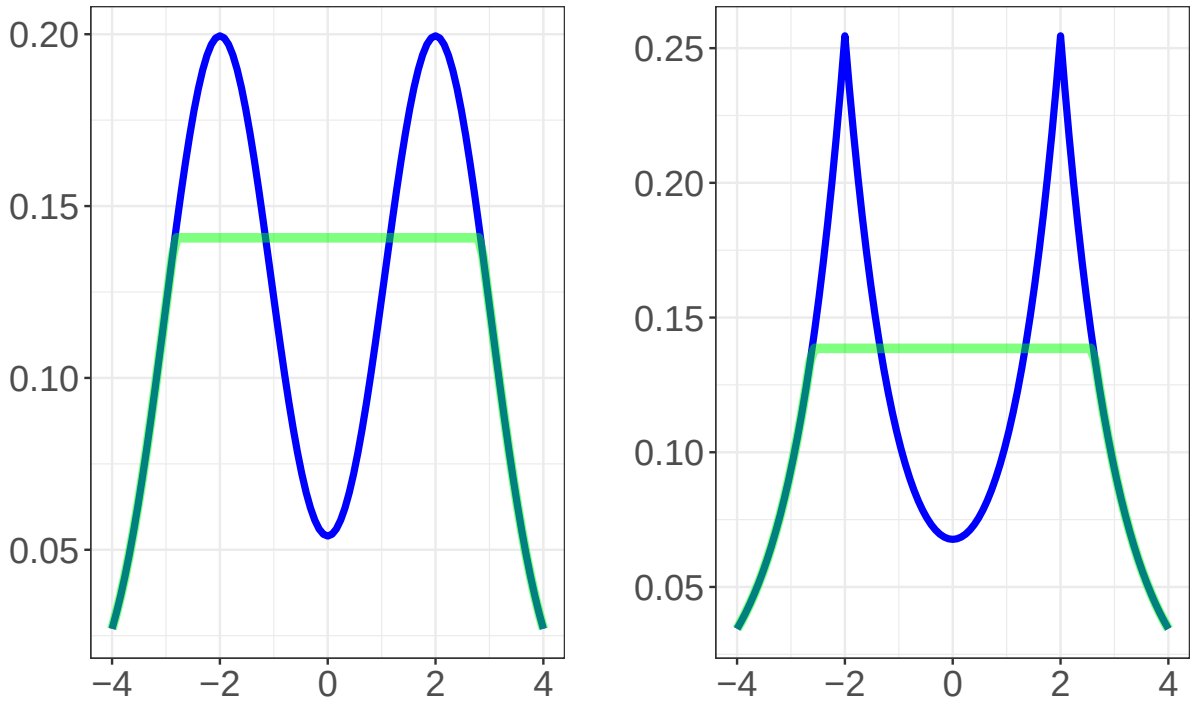


Figure 2.7: The plot in the left panel displays the density f (in blue) in (2.38) and its projection $f^*(F)$ (in green); the plot in the right panel displays the density f in (2.40) (drawn in blue) and the corresponding $f^*(F)$ (in green).

On the other hand, substituting $F = \mathbb{F}_n$ in Lemma 2.7, we obtain that $\widehat{\theta}_n \in [X_{(1)}, X_{(n)}]$, which was already stated in Theorem 2.1.

Now we move on to the consistency properties of the MLE. To motivate our results, we briefly revisit the log-concave projection theory mentioned in Section 2.3.1. As mentioned earlier in Section 2.3.1, Dümbgen *et al.* (2011) considered projection of F onto the class of distributions with density in $\mathcal{LC}$. Dümbgen *et al.* (2011) showed that under some regularity

conditions, the map

$$F \mapsto \sup_{\psi \in \mathcal{LC}} \omega(\psi, F)$$

is continuous with respect to the Wasserstein distance d_W, where ω was defined in (2.21). Xu and Samworth (2017) obtained similar results for ω, though they considered the class of densities $\mathcal{SLC}_\theta$, which is a subclass of $\mathcal{LC}$. Instead of $\omega(\psi, F)$, we are working with a slightly more complicated criterion function $\Psi(\theta, \psi, F)$, which also involves θ. Yet it is not unnatural to expect the same continuity result to hold in our set-up. Being led by this intuition, in Theorem 2.7, we establish the continuity of the map $F \mapsto L(F)$, where L was defined in (2.5). We can also show that when Condition A and B are satisfied, $f^*(F_n)$ converges pointwise to $f^*(F)$ when $d_W(F_n, F) \to 0$. Since $\theta^*(F_n)$ is the first moment of $f^*(F_n)$, its limit is found to be $\theta^*(F)$.

Theorem 2.7. *Let $\{F_n\}_{n\geq 1}$ be a sequence of distribution functions such that $d_W(F_n, F) \to 0$ for some distribution function F. Further suppose, F and F_n satisfy Conditions A and B for each $n \geq 1$. Then as $n \to \infty$,*

$$L(F_n) \to L(F), \tag{2.41}$$

and the following assertions hold:

(a) $\lim_n \theta^*(F_n) = \theta^*(F)$.

(b) $\lim_n \psi^*(F_n)(x) = \psi^*(F)(x)$, *for all* $x \in int(dom(\psi^*(F)))$.

(c) $\limsup_n \psi^*(F_n)(x) \leq \psi^*(F)(x)$, *for all* $x \in \mathbb{R}$.

(d) $\lim_n g^*(F_n)(x) = g^*(F)(x)$, *for all* $x \in \mathbb{R}$.

(e) There exists $\alpha > 0$ *such that for all* $\alpha_1 < \alpha$,

$$\lim_n \int_{-\infty}^{\infty} e^{\alpha_1 |x|} |g^*(F_n)(x) - g^*(F)(x)| dx = 0.$$

Moreover, if $g^*(F)$ is continuous on $\mathbb{R}$, then as $n \to \infty$,

$$\sup_{x \in \mathbb{R}} e^{\alpha_1 |x|} |g^*(F_n)(x) - g^*(F)(x)| \to 0.$$

Since $g^*(F)$ is a log-concave density, there exists α and β such that $\psi^*(F)(x) \leq -\alpha|x| + \beta$ for all $x \in \mathbb{R}$ (Cule and Samworth, 2010, Lemma 1). This is same as the α that appears in Theorem 2.7.

Now we are in a situation to address the convergence of $\widehat{\theta}_n$ and $\widehat{g}_n$. Suppose F_0 satisfies Condition A and B. Then Theorem 2.8 entails that $\widehat{\theta}_n \to_{a.s.} \theta^*(F_0)$, and $\widehat{g}_n$ converges to $e^{\psi^*(F_0)}$ almost surely, both pointwise and in L_1. Theorem 2.8 follows as a corollary of Theorem 2.7. This can be verified easily observing that under Condition A, $d_W(\mathbb{F}_n, F_0) \to_{a.s.} 0$ (Villani, 2009, Theorem 6.9), and $\mathbb{F}_n$ is non-degenerate with probability 1 for large n.

Theorem 2.8. *Suppose F_0 satisfies condition A and B. Then as $n \to \infty$, $L(\mathbb{F}_n) \to_{a.s.} L(F_0)$, and the following assertions hold:*

(a) $\widehat{\theta}_n \to_{a.s.} \theta^*(F_0)$.

(b) $\widehat{\psi}_{0,n}(x) \to_{a.s.} \psi^*(F_0)(x)$, *for all* $x \in int(dom(\psi^*(F_0)))$.

(c) $\limsup_n \widehat{\psi}_{0,n}(x) \leq \psi^*(F_0)(x)$, *for all* $x \in \mathbb{R}$ *a.s.*

(d) $\widehat{g}_n(x) \to_{a.s.} g^*(F_0)(x)$, *for all* $x \in \mathbb{R}$.

(e) There exists $\alpha > 0$ such that for all $\alpha_1 < \alpha$,

$$\int_{-\infty}^{\infty} e^{\alpha_1 |x|} |\widehat{g}_n(x) - g^*(F_0)(x)| dx \to_{a.s.} 0.$$

Moreover, if $g^*(F_0)$ is continuous on $\mathbb{R}$,

$$\sup_{x \in \mathbb{R}} e^{\alpha_1 |x|} |\widehat{g}_n(x) - g^*(F_0)(x)| \to_{a.s.} 0.$$

Theorem 2.8 has some important consequences for F_0 satisfying (2.24) with $\theta_0 \in \mathbb{R}$. Lemma 2.6 implies that Condition B holds F_0 satisfying (2.24) with $\theta_0 \in \mathbb{R}$, which leads to the following corollary.

Corollary 4. *If F_0 satisfies Condition A and (2.24) with $\theta_0 \in \mathbb{R}$, all conclusions of Theorem 2.8 hold with $\theta^*(F_0) = \theta_0$.*

In a more specific case when $f_0 \in \mathcal{SLC}_{\theta_0}$, we have both $\theta^*(F_0) = \theta_0$ and $\psi^*(F_0) = \psi_0^*$, entailing $f^*(F_0) = f_0$. As a result, not only $\widehat{\theta}_n$ is a consistent estimator of θ_0, but also $\widehat{g}_n$ converges to g_0 almost surely, both pointwise and in L_1. Since $\mathcal{P}_0 \subset \mathcal{SLC}_{\theta_0}$, the same conclusions hold for a correctly specified model.

When $f_0 \in \mathcal{P}_0$, we can derive even stronger results for the global convergence of the MLE $\widehat{g}_n$. For the unrestricted log-concave MLE $\bar{f}_m$, it can be shown that $H(\bar{f}_m, f_0)^2 = O_p(n^{-4/5})$ (Doss and Wellner, 2016, Theorem 3.2). Theorem 4.1(c) of Doss and Wellner (2019) derives the same rate of convergence for the Hellinger distance between a density $f \in \mathcal{SLC}_0$ and its MLE. Theorem 1 of Kim and Samworth (2016) states that this is the minimax rate of convergence for estimating an f in $\mathcal{LC}$. In particular, Theorem 1 and Theorem 5 of Kim and Samworth (2016) imply that there exist $0 < c_1 < c_2$ such that

$$c_1 n^{-4/5} \leq \inf_{\hat{f}_n} \sup_{f \in \mathcal{LC}} E_f H^2(\hat{f}_n, f) \leq c_2 n^{-4/5},$$

where the infimum is over all measurable estimators $\hat{f}_n$ of f. Doss and Wellner (2019) conjectured in their Remark 4.2 that $n^{-4/5}$ is the minimax rate of convergence for $\mathcal{SLC}_0$ as well. In Theorem 2.9 we derive that $H(\widehat{f}_n, f_0)^2$ and $H(\widehat{g}_n, g_0)^2$ is $O_p(n^{-4/5})$. We conjecture that the minimax rate of estimation of $\widehat{f}_n$ in $\mathcal{P}_0$ is the same. As a corollary to Theorem 2.9, it follows that $||\widehat{g}_n - g_0||_1 = O_p(n^{-2/5})$. For the MLE of θ_0, we could show that $|\widehat{\theta}_n - \theta_0|$ is $O_p(n^{-2/5})$.

Theorem 2.9. *Suppose $f_0 = g_0(\cdot - \theta_0)$ for some $g_0 \in \mathcal{SLC}_0$, and $\theta_0 \in \mathbb{R}$. Then the following assertions hold:*

(A) As $n \to \infty$,

$$H(\widehat{f}_n, f_0) \to_{a.s.} 0, \quad \textit{and} \quad H(\widehat{g}_n, g_0) \to_{a.s.} 0.$$

Moreover,

$$H^2(\widehat{f}_n, f) = O_p(n^{-4/5}).$$

(B) If $f_0 \in \mathcal{P}_0$, *we have* $|\widehat{\theta}_n - \theta_0| = O_p(n^{-2/5})$, *and*

$$H^2(\widehat{g}_n, g_0) = O_p(n^{-4/5}).$$

Lemma 2.8 is an analogue of Theorem 2.5 for one step estimators, and addresses the consistency of $\widehat{f}_n$, $\widehat{\phi}_{0,n}$ and their derivatives.

Lemma 2.8. *Suppose* $f_0 = g_0(\cdot - \theta_0)$, *for some* $g_0 \in \mathcal{SLC}_0$, *and* $\theta_0 \in \mathbb{R}$. *Denoting* $\phi_0 = \log f_0$, *on any compact set* $K \subset dom(\phi_0)$, *we have,*

$$\sup_{x \in K} |\widehat{\phi}_{0,n}(x) - \phi_0(x)| \to_{a.s.} 0.$$

Suppose ϕ_0 *is differentiable at* $x \in K$. *Then it also follows that*

$$\widehat{\phi}'_{0,n}(x) \to_{a.s.} \phi'_0(x),$$

and

$$\widehat{f}'_n(x) \to_{a.s.} f'_0(x).$$

Similarly for $\psi_0 = \log g_0$, *on any compact set* $K \subset dom(\psi_0)$, *we have,*

$$\sup_{x \in K} |\widehat{\psi}_{0,n}(x) - \psi_0(x)| \to_{a.s.} 0.$$

If ψ_0 *is differentiable at* $x \in K$, *then we also have*

$$\widehat{\psi}'_{0,n}(x) \to_{a.s.} \psi'_0(x),$$

and

$$\widehat{g}'_n(x) \to_{a.s.} g'_0(x).$$

We close this section by mentioning that for $f_0 \in \mathcal{P}_0$, we anticipate $n^{1/2}$-consistency for the MLE $\widehat{\theta}_n$. So we are pursuing further research in this direction. In particular, we expect to establish that $\widehat{\theta}_n$ is asymptotically efficient.

2.4 Simulation study

In this section, we study the performance of the one step estimators $\tilde{\theta}_n$ and the MLE $\widehat{\theta}_n$ for a number of log-concave densities symmetric about 0. These densities include the standard normal, the logistic, and the Laplace density, which are plotted in Figure 2.5. We also consider the symmetrized beta distribution discussed in Section 2.3.1 (see Figure 2.5), whose density is given by (2.36). Taking g_0 to be any of the above densities, we let $\theta_0 = 0$.

The general set-up of the simulations is as follows. We generate 3000 samples of size $n = 30$, 100, 200, and 500 from each of the above-mentioned densities. Then we construct the MLE $\widehat{\theta}_n$, and the one step estimators, both truncated and untruncated, defined in (2.16) and (2.7) respectively. The truncated one step estimator requires specification of the truncation parameter η. We set η to be 0.002 because Figure 2.5 indicates that the value of $\mathcal{I}_{f_0}(\eta)/\mathcal{I}_{f_0}$ does not vary significantly for $\eta \in (0.0001, 0.002)$, at least for normal, Laplace, and logistic distribution.

Different choices of the preliminary estimator $\bar{\theta}_n$ and the density estimator $\tilde{g}_n$ in (2.16) and (2.7) lead to different one-step estimators. The preliminary estimator $\bar{\theta}_n$ is generally taken to be the mean, median, or the trimmed mean (12.5% trimming from each tail). However, for the logistic distribution, we also consider the parametric MLE as an initial estimator. The $\tilde{g}_n$ in (2.16) and (2.7) is taken to be $\bar{f}_m(\cdot - \bar{\theta}_n)$, or the estimators of g_0 defined in (2.8), (2.10), (2.13), and (2.14). The performance of the resulting estimators are compared with the preliminary estimator $\bar{\theta}_n$.

Our simulations reveal that all our estimators are consistent for θ_0. The parametric MLE displays least bias in both small and large samples, followed by the smoothed symmetrized

one step estimator in (2.10), and the partial MLE estimator in (2.13). We conjecture that the untruncated one step estimator and the MLE have limiting variance given by $\mathcal{I}_{f_0}^{-1}$, and Theorem 2.6 states that the asymptotic variance of the truncated one step estimators is a truncated version of $\mathcal{I}_{f_0}^{-1}$. Therefore we consider a measure of efficiency defined as follows

$$\text{Efficiency}(\theta_n) = \frac{1/(n\mathcal{I}_{f_0})}{Var(\theta_n)}. \tag{2.42}$$

Since we can not directly compute $Var(\theta_n)$, we will replace it by its Monte Carlo estimate.

Our simulations suggest that in both small and large samples, the smoothed symmetrized one step estimator in (2.10) and the partial MLE estimator in (2.13) dominate the other one step estimators in terms of efficiency. Therefore in this section, alongside the MLE, we present the asymptotic efficiency of only these two one step estimators.

2.4.1 Normal density

In this case, f_0 is the standard Guassian density, which implies $\mathcal{I}_{f_0} = 1$, and $\theta_0 = 0$. For Gaussian location model, the sample mean is the parametric MLE, which is also (asymptotically) efficient. Figure 2.8 compares the efficiency of the MLE, the partial MLE one-step estimator in (2.13), and the smoothed symmetrized one step estimator in (2.10). From Figure 2.8, we observe that the smoothed symmetrized estimator has the best efficiency for small samples. It turns out that it also slightly outperforms the other estimators in large sample ($n = 500$) as well. As expected, the parametric MLE, i.e. the mean, has the highest efficiency. Our estimators, however, outperform all other preliminary estimators. Figure 2.8 also indicates that the untruncated one step estimators have higher efficiency than their truncated counterparts.

2.4.2 Logistic density

The standard logistic density has $\mathcal{I}_{f_0} = 1/3$. In this case along with the mean, the median, and the trimmed mean, we also consider the parametric MLE as a preliminary estimator.

Figure 2.9 compares the efficiency of the one step estimators and the MLE. We observe that the MLE has lower efficiency than both the one step estimators, the partial MLE one-step estimator and the smoothed symmetrized one step estimator. Figure 2.9 also indicates that the smoothed symmetrized one step estimator has the highest efficiency in smaller samples (sample size $n = 40, 100$). For this estimator, surprisingly, the truncated estimator outperforms the untruncated estimator when $n = 200$, and 500, regardless of the choice of the preliminary estimator. For all the other cases, the efficiency of the untruncated estimator dominates the efficiency of the truncated estimator. As usual, the parametric MLE has the highest efficiency.

2.4.3 Laplace density

In this subsection, we simulate observations from the standard Laplace density, which has $\mathcal{I}_{f_0} = 1$. In this case, the sample median is the parametric MLE of θ_0. Unlike the other densities we considered, this density is not smooth, in fact it has a kink at 0 (see Figure 2.5). Figure 2.10 indicates that the smoothed symmetrized one step estimator in (2.10) suffers from the lack of smoothness, and loses its edge to the MLE and the partial MLE one step estimator, both in small and large samples. When $n = 40$ and 500, the partial MLE one step estimator has the highest efficiency. For all other sample sizes, Figure 2.10 suggests that the MLE has the highest efficiency. In all cases, the untruncated one step estimator outperforms the truncated version.

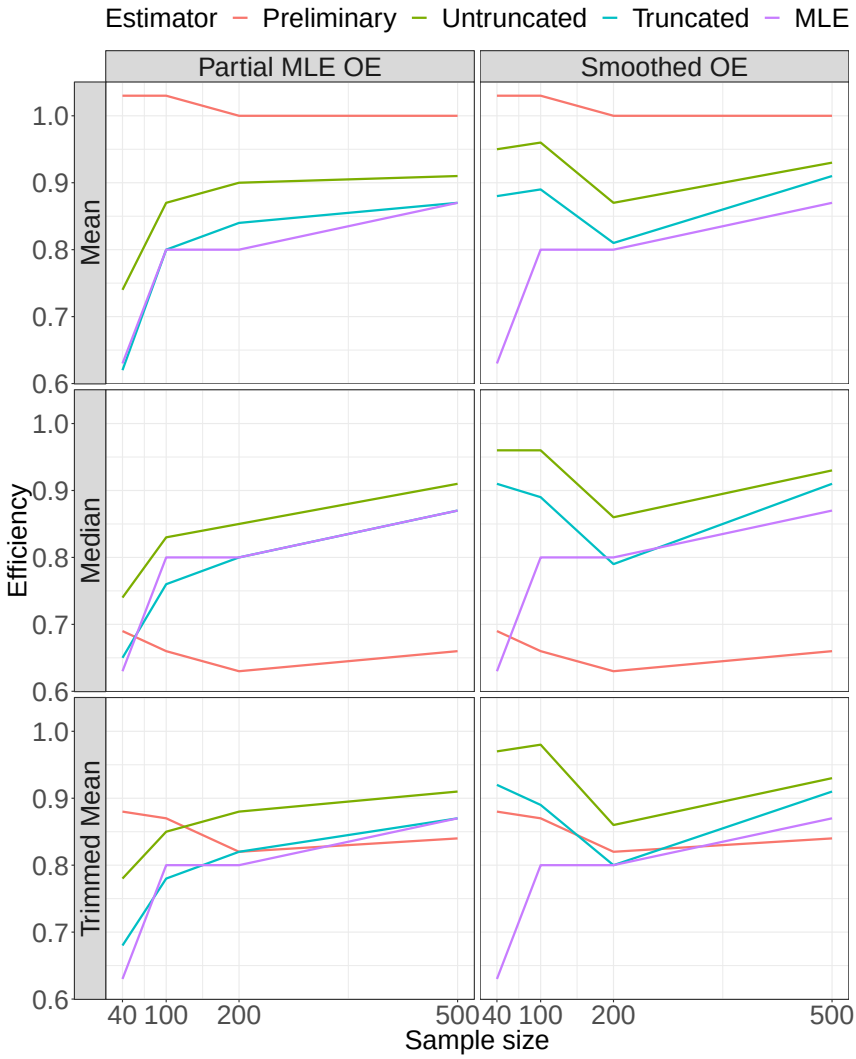


Figure 2.8: Plot of efficiency vs sample size (n) for standard Gaussian density. We plotted the efficiency of partial MLE one-step estimator in (2.13) (Partial MLE OE; left column), the smoothed symmetrized one step estimator in (2.10) (Smoothed OE; right column), and the MLE (in purple; present in both columns). The preliminary estimators $\bar{\theta}_n$ corresponding to the one-step estimators are the mean (top), median (middle), and trimmed mean (bottom). In context of the one step estimators, the preliminary estimator is drawn in red, the truncated and untruncated estimators are drawn in blue and green respectively.

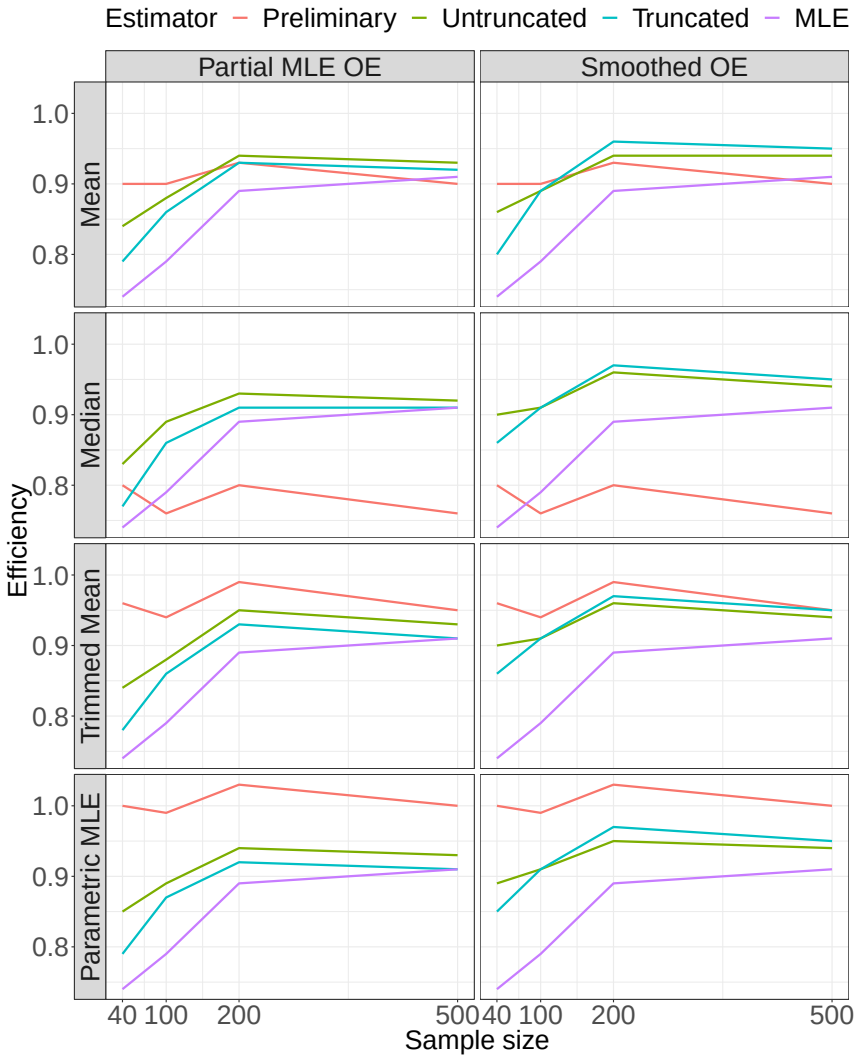


Figure 2.9: Plot of efficiency vs sample size (n) when the underlying density is standard Logistic. Here we consider the partial MLE one-step estimator in (2.13) (Partial MLE OE; left column), the smoothed symmetrized one step estimator in (2.10) (Smoothed OE; right column), and the MLE (in purple; present in both columns). Here the underlying distribution is standard Logistic. The preliminary estimators $\bar{\theta}_n$ corresponding to the one-step estimators are the mean (first row), median (second row), trimmed mean (third row), and the parametric MLE (fourth MLE). In context of the one step estimators, the preliminary estimator is drawn in red, the truncated and untruncated estimators are drawn in blue and green respectively.

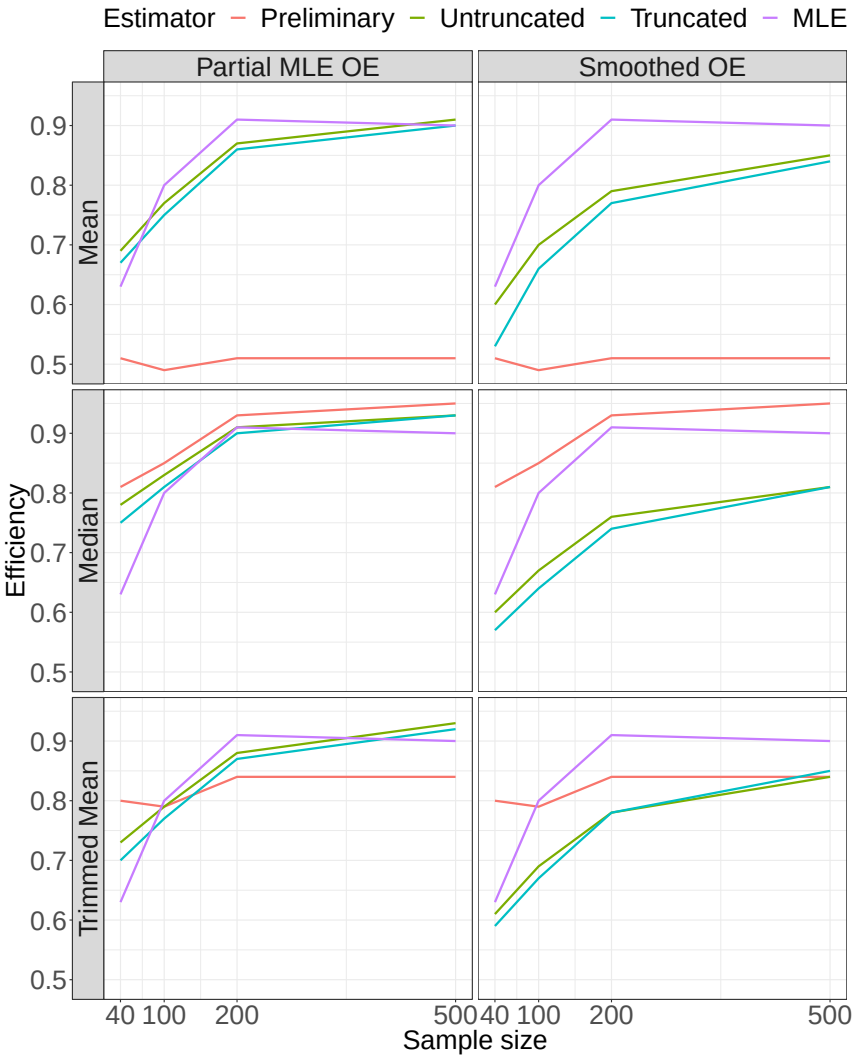


Figure 2.10: Plot of efficiency vs sample size (n) for the standard Laplace density. Here we consider the partial MLE one-step estimator in (2.13) (Partial MLE OE; left column), the smoothed symmetrized one step estimator in (2.10) (Smoothed OE; right column), and the MLE (in purple; present in both columns). Here the underlying distribution is standard Laplace. The preliminary estimators $\bar{\theta}_n$ corresponding to the one-step estimators are the mean (top), median (middle), and trimmed mean (bottom). In context of the one step estimators, the preliminary estimator is drawn in red, the truncated and untruncated estimators are drawn in blue and green respectively.

2.4.4 Symmetrized beta density

In this subsection, we consider the symmetrized beta density in (2.36), for $r = 2.1$, 2.5, and 4.5. Figure 2.11 compares the efficiency of the MLE and the one step estimators, where the latter has the mean as the preliminary estimator. The plots for the other preliminary estimators depict the same story, and hence we do not present them. The first thing to observe from Figure 2.11 is that the efficiency of the estimators increases with r. Also for $r = 2.1$ and 2.5, all the estimators, including the mean, exhibit poor efficiency even for sample size 500. Our simulations reveal that the smoothed untruncated one step estimator has the highest asymptotic efficiency among all the estimators. Figure 2.11 indicates that similar to the other distributions, the untruncated one step estimators display higher efficiency than the truncated one step estimator. We note that for $r = 2.5$ and 4.5, this gap in the efficiency is higher compared to the other densities.

Our simulations indicate that except for the case of standard Laplace, the smoothed symmetrized one step estimator has comparatively higher efficiency, only second to the parametric MLE. We also note that in most cases the untruncated one step estimator exhibits higher efficiency than its truncated counterpart.

2.5 Discussion

Here we revisit the problem of estimating the center of symmetry $\theta_0 \in \mathbb{R}$ of the density $f_0 = g_0(\cdot - \theta_0)$, where g_0 is symmetric about 0. All existing methods involve use of various tuning parameters. To avoid the dependence on tuning parameters, we impose an additional restriction of log-concavity on f_0, and consider two different estimation procedures.

First we consider truncated one-step estimators of θ_0, and prove that they are $\sqrt{n}$-consistent, nearly achieving the asymptotic efficiency bound $\mathcal{I}_{f_0}^{-1}$. Next we demonstrate that the MLE of (θ_0, g_0) exists for our model. Then we establish that the MLE of g_0 is strongly Hellinger consistent. We also show that the corresponding Hellinger distance has the rate

$O_p(n^{-2/5})$. Although we could only show that the MLE of θ is $O_p(n^{-2/5})$, we conjecture that the MLE is actually $\sqrt{n}$-consistent.

Our work shows that the assumption of log-concavity can ameliorate the dependence on tuning parameters. In particular the one-step estimators use only one additional parameter for truncation, and the MLE is tuning parameter free. Furthermore, we show that even if the log-concavity assumption is violated, our estimators are still consistent under fairly mild conditions as long as the assumption of symmetry holds. This is a consequence of the impressive stability of log-concave density estimators (Dümbgen *et al.*, 2011; Xu and Samworth, 2017). When g_0 is not log-concave but symmetric about 0, all our estimators of θ_0 are still consistent under very mild conditions; in fact, the one step estimators are still $\sqrt{n}$-consistent. This clarifies the usefulness of a log-concavity assumption in semiparametric models in facilitating a simplified and robust estimation procedure.

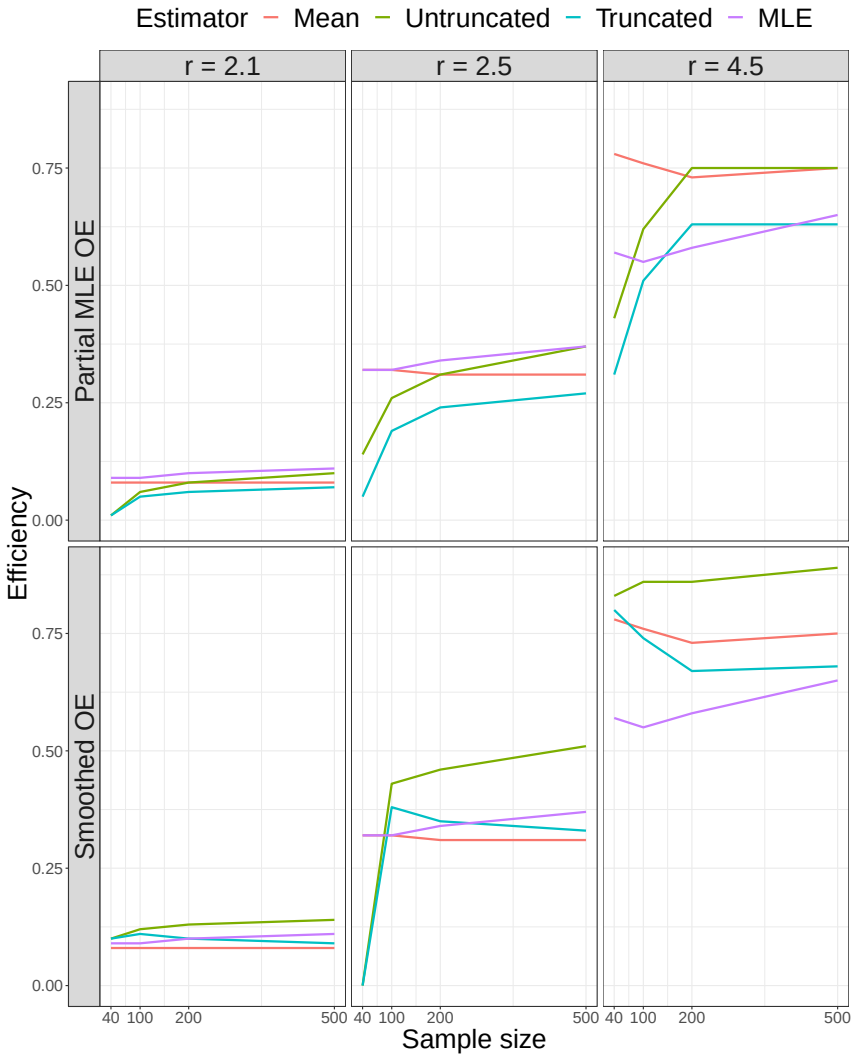


Figure 2.11: Plot of efficiency vs sample size (n) when the underlying density is symmetrized beta, which is defined in (2.36). We consider $r = 2.1$ (left), 2.5 (middle), and 4.5 (right). The MLE is drawn in purple. We calculated the partial MLE one-step estimator in (2.13) (Partial MLE OE; top), and the smoothed symmetrized one step estimator in (2.10) (Smoothed OE; bottom). The preliminary estimator for them, which is sample mean in this case, is drawn in red; and the truncated and untruncated versions are drawn in blue and green respectively.

2.6 Proofs

2.6.1 Proofs for the one-step estimators

Proof of Corollary 3

Suppose that

$$\sup_{x\in \text{int}(\text{dom}(\psi_0))} |\psi_0'(x)| = \infty,$$

but there exists $M > 0$ such that $\sup_{x\in \text{int}(\text{dom}(\tilde{\psi}_n))} |\tilde{\psi}_n'(x)| < M$ for all sufficiently large n. We claim that the probability of such an event is 0. It is straightforward to see that there exists $y \in \text{int}(\text{dom}(\phi_0))$ such that $|\psi_0'(y)| > 2M$. One can also choose y such that ψ_0' is continuous at y. Therefore pointwise convergence of $\tilde{\psi}_n'$ to ψ_0' fails at y. Hence the proof follows by Theorem 2.5. □

Proof of Lemma 2.2

Note that each $\phi \in \mathcal{SC}_\theta$ can be written as $\psi(\cdot - \theta)$ for some $\psi \in \mathcal{SC}_0$, and by (2.21) and (2.3), we also have

$$\omega(\phi, F) = \Psi(\theta, \psi, F). \tag{2.43}$$

For $\psi \in \mathcal{SC}_0$ we calculate,

$$\begin{aligned}
\Psi(0, \psi, F) &= \int_{-\infty}^{0} \psi(x) dF(x) + \int_{0}^{\infty} \psi(x) dF(x) - \int_{-\infty}^{\infty} e^{\psi(x)} dx \\
&= -\int_{0}^{\infty} \psi(-x) dF(-x) + \int_{0}^{\infty} \psi(x) dF(x) - \int_{-\infty}^{\infty} e^{\psi(x)} dx \\
&= \int_{0}^{\infty} \psi(x) d(F(x) - F(-x)) - \int_{-\infty}^{\infty} e^{\psi(x)} dx.
\end{aligned}$$

It also follows that

$$\int_{0}^{\infty} \psi(x) d(F(x) - F(-x)) = \int_{-\infty}^{0} \psi(x) d(F(x) - F(-x)).$$

Therefore

$$\Psi(0,\psi,F) = 2^{-1}\int_{-\infty}^{\infty}\psi(x)d(F(x)-F(-x)) - \int_{-\infty}^{\infty}e^{\psi(x)}dx,$$

implying

$$\begin{aligned}\Psi(\theta,\psi,F) &= 2^{-1}\int_{-\infty}^{\infty}\psi(x)d(F(\theta+x)-F(\theta-x)) - \int_{-\infty}^{\infty}e^{\psi(x)}dx \\ &= 2^{-1}\int_{-\infty}^{\infty}\psi(z-\theta)d(F(z)-F(2\theta-z)) - \int_{-\infty}^{\infty}e^{\psi(z)}dz \\ &= \int_{-\infty}^{\infty}\psi(z-\theta)dF_{\theta}^{sym}(z) - \int_{-\infty}^{\infty}e^{\psi(z)}dz \\ &= \Psi(\theta,\psi,F_{\theta}^{sym}).\end{aligned}$$

Thus (2.43) shows that for any $\phi \in \mathcal{SC}_{\theta}$,

$$\omega(\phi,F) = \omega(\phi,F_{\theta}^{sym})$$

Therefore

$$\arg\max_{\phi\in\mathcal{SC}_{\theta}}\omega(\phi,F) = \arg\max_{\phi\in\mathcal{SC}_{\theta}}\omega(\phi,F_{\theta}^{sym}),$$

but the distribution function

$$F_{\theta}^{sym}(x) = 2^{-1}\Big(F(x)+1-F(2\theta-x)\Big)$$

satisfies Condition A and (2.24). Hence applying Lemma 2.12, we obtain that

$$\arg\max_{\phi\in\mathcal{SC}_{\theta}}\omega(\phi,F_{\theta}^{sym}) = \arg\max_{\phi\in\mathcal{C}}\omega(\phi,F_{\theta}^{sym}),$$

which exists by Condition A, thus completing the proof. □

Proof of Lemma 2.3

First note that it suffices to prove the current lemma when $\bar{\theta}_n \to_{a.s.} \theta_0$. Suppose the strong consistency does not hold but $\bar{\theta}_n \to_p \theta_0$. Then given any subsequence of $\{\bar{\theta}_n\}_{n\geq 1}$, there

exists a further subsequence $\{\bar{\theta}_{n_k}\}_{k\geq 1}$ such that $\bar{\theta}_{n_k} \to_{a.s.} \theta_0$ as $k \to \infty$. Therefore along this subsequence $\{n_k\}_{k\geq 1}$, the L_1 distance between $\tilde{g}_n$ and $\tilde{f}_0$ or $\tilde{f}_0^{sym}$ (in case of $\tilde{g}_n = (\widehat{f}_n^{sym})^{sm}$) approaches 0 almost surely. Hence, if $||\tilde{g}_n - \tilde{g}_0||_1$ or $||\tilde{g}_n - \tilde{g}_0^{sm}||_1$ has any convergent subsequence at all, the subsequence converges to 0 almost surely. In that case Shorack (2000) (Theorem 5.7, page 57) implies that the whole sequence converges in probability to 0. Therefore in what follows, we assume that $\bar{\theta}_n \to_{a.s.} \theta_0$.

For our first estimator $\tilde{g}_n^{sym}$, the proof is motivated by Theorem 4 of Cule and Samworth (2010), which states that if $\int \log f_0(x) dF_0(x) < \infty$, and Condition A holds,

$$\int_{-\infty}^{\infty} |\bar{f}_m(x) - \tilde{f}_0(x)| dx \to_{a.s.} 0. \tag{2.44}$$

Note that $f_0 \in \mathcal{P}_1$ is bounded, so $\int \log_+ f_0(x) dF_0(x) < \infty$ trivially holds, where $\log_+(y) = \max\{\log y, 0\}$. To prove the L_1 consistency of $\tilde{g}_n^{sym}$, we now exploit the symmetry of $\tilde{f}_0$ and $\bar{f}_m$ about θ_0 and $\bar{\theta}_n$ respectively. Recalling the definition of $\tilde{g}_n^{sym}$ from (2.8), we calculate that

$$\begin{aligned}
&2\int_{-\infty}^{\infty} |\tilde{g}_n^{sym}(x) - \tilde{g}_0(x)| dx \\
&= 2\int_{-\infty}^{\infty} |\tilde{g}_n^{sym}(x) - \tilde{f}_0(\theta_0 + x)| dx \\
&\leq \int_{-\infty}^{\infty} |\bar{f}_m(\bar{\theta}_n + x) - \tilde{f}_0(\bar{\theta}_n + x)| dx + \int_{-\infty}^{\infty} |\bar{f}_m(\bar{\theta}_n - x) - \tilde{f}_0(\bar{\theta}_n - x)| dx \\
&\quad + \int_{-\infty}^{\infty} |\tilde{f}_0(\bar{\theta}_n + x) - \tilde{f}_0(\theta_0 + x)| dx + \int_{-\infty}^{\infty} |\tilde{f}_0(\bar{\theta}_n - x) - \tilde{f}_0(\theta_0 - x)| dx,
\end{aligned} \tag{2.45}$$

whose first two terms approach 0 almost surely by (2.44). Therefore we only need to take care of the last two terms of (2.45). Since $\tilde{f}_0$ is log-concave, it is continuous almost everywhere on $\mathbb{R}$ with respect to the Lebesgue measure. Therefore $|\tilde{f}_0(\bar{\theta}_n + x) - \tilde{f}_0(\theta_0 + x)| \to_{a.s.} 0$ almost everywhere with respect to the Lebesgue measure. Therefore application of Glick's Theorem (Theorem 2.6 Devroye, 1987) then ensures that the last two terms of (2.45) converge to 0

almost surely as well.

To establish the convergence of $(\widehat{f}_n^{sym})^{sm}$ to $\tilde{g}_0^{sm}$ in L_1, we will exploit the representation of $(\widehat{f}_n^{sym})^{sm}$ in terms of $(\bar{f}_m)^{sm}$ established in (2.12). Since $\tilde{g}_0^{sm}$ and $(\widehat{f}_n^{sym})^{sm}$ are symmetric about θ_0 and $\bar{\theta}_n$ respectively, similar algebra as in the case of $\tilde{g}_n = \tilde{g}_n^{sym}$ leads to

$$
\begin{aligned}
&2\int_{-\infty}^{\infty} |(\widehat{f}_n^{sym})^{sm}(x) - \tilde{g}_0^{sm}(x)|dx \\
&\quad \leq \int_{-\infty}^{\infty} |\widehat{f}_n^{sm}(\bar{\theta}_n + x) - \tilde{f}_0^{sm}(\bar{\theta}_n + x)|dx + \int_{-\infty}^{\infty} |\widehat{f}_n^{sm}(\bar{\theta}_n - x) - \tilde{f}_0^{sm}(\bar{\theta}_n - x)|dx \\
&\qquad + \int_{-\infty}^{\infty} |\tilde{f}_0^{sm}(\bar{\theta}_n + x) - \tilde{f}_0^{sm}(\theta_0 + x)|dx \\
&\qquad + \int_{-\infty}^{\infty} |\tilde{f}_0^{sm}(\bar{\theta}_n - x) - \tilde{f}_0^{sm}(\theta_0 - x)|dx.
\end{aligned}
\tag{2.46}
$$

The first two terms can be handled by using Theorem 1 of Chen and Samworth (2013), which states that when f_0 satisfies (2.25), we have

$$
\int_{-\infty}^{\infty} |\widehat{f}_n^{sm}(x) - \tilde{f}_0^{sm}(x)|dx \to_{a.s.} 0. \tag{2.47}
$$

Since $\tilde{f}_0^{sm}$ is the convolution of a log-concave density and a Gaussian density, it is continuous almost everywhere on $\mathbb{R}$ with respect to the Lebesgue measure. Therefore imitating the proof for the case of $\tilde{g}_n = \tilde{g}_n^{sym}$, we can show that the last two terms of (2.46) are negligible.

To establish the L_1 consistency of the partial MLE estimator $\widehat{g}_{\bar{\theta}_n}$, we appeal to the projection theory developed in Xu and Samworth (2017). In this case $\widehat{g}_{\bar{\theta}_n}$ is the density of the projection of F_{097}, the empirical distribution function of the $X_i - \bar{\theta}_n$'s, onto the space of the distribution functions with density in $\mathcal{SLC}_0$. By Proposition 6 of Xu and Samworth (2017), if

$$
d_W(F_{097}, G_0) \to_{a.s.} 0, \tag{2.48}
$$

we have $||\widehat{g}_{\bar{\theta}_n} - \tilde{g}_0||_1 \to_{a.s.} 0$ as long as Condition A holds. Hence we only need to show that $d_W(F_{097}, G_0) \to_{a.s.} 0$, for which, by Theorem 6.9 of Villani (2009), it suffices to show the

following. First, we need to show that

$$\int_{-\infty}^{\infty} |x| dF_{097}(x) \to_{a.s.} \int_{-\infty}^{\infty} |x| dG_0(x), \tag{2.49}$$

and that F_{097} converges to G_0 weakly with probability one. Since $\bar{\theta}_n$ is strongly consistent for θ_0, and

$$\int_{-\infty}^{\infty} |x| dF_{097}(x) = \int_{-\infty}^{\infty} |x - \bar{\theta}_n| d\mathbb{F}_n(x), \tag{2.50}$$

for any $d > 0$, an application of Glivenko-Cantelli Theorem (for example, see Theorem 2.4.1 of Van der Vaart and Wellner (1996)) yields

$$\sup_{\bar{\theta}_n \in [\theta_0 - d, \theta + d]} \left| \int_{-\infty}^{\infty} |x - \bar{\theta}_n| d(\mathbb{F}_n - F_0)(x) \right| \to_{a.s.} 0.$$

On the other hand, strong consistency of $\bar{\theta}_n$ implies $|x - \bar{\theta}_n| \leq |x - \theta_0| + 1$ with probability one for all sufficiently large n, where the latter is integrable. Therefore the dominated convergence theorem leads to

$$\int_{-\infty}^{\infty} |x - \bar{\theta}_n| dF_0(x) \to_{a.s.} \int_{-\infty}^{\infty} |x - \theta_0| dF_0(x) = \int_{-\infty}^{\infty} |x| dG_0(x),$$

which proves (2.49).

Our next step is to prove the weak convergence of $F_{097}$to G_0. To this end, we note that

$$F_{097}(x) = \mathbb{F}_n(x + \bar{\theta}_n) = \int_{-\infty}^{\infty} 1_{(-\infty, x + \bar{\theta}_n]}(z) d\mathbb{F}_n(z), \tag{2.51}$$

which converges almost surely to

$$\int_{-\infty}^{\infty} 1_{(-\infty, x + \theta_0]}(z) dF_0(z) = G_0(x)$$

by an application of basic Glivenko-Cantelli Theorem (see Theorem 2.4.1 of Van der Vaart and Wellner (1996)), and the fact that $F_0(x + \bar{\theta}_n) \to_{a.s.} F_0(x + \theta_0)$ for all $x \in \mathbb{R}$. This completes the proof of strong L_1 consistency of $\widehat{g}_{\bar{\theta}_n}$ for $\tilde{g}_0$.

Finally we consider the geometric mean estimator $\widehat{f}_n^{geo,sym}$. Recall that in the course of showing the consistency of $\tilde{g}_n^{sym}$, we have shown that the right hand side of (2.45) converges to 0 almost surely, from which one can derive that

$$\int_{-\infty}^{\infty} |\bar{f}_m(\bar{\theta}_n \pm x) - \tilde{g}_0(x)| dx \to_{a.s.} 0. \tag{2.52}$$

Proposition 2(b) of Cule and Samworth (2010) shows that (2.52) leads to almost sure convergence of $\bar{f}_m(\bar{\theta}_n \pm x)$ to $\tilde{g}_0(x)$ almost everywhere on $\mathbb{R}$ with respect to the Lebesgue measure. As a consequence, it follows that

$$\widehat{f}_n^{geo,sym}(x)/C_n^{geo} \to_{a.s.} \Big(\tilde{g}_0(x)\tilde{g}_0(-x)\Big)^{1/2} = \tilde{g}_0(x) \quad a.e. \quad x.$$

Recall from (2.14) that

$$C_n^{geo} = \int_{-\infty}^{\infty} \sqrt{\bar{f}_m(\bar{\theta}_n + x)\bar{f}_m(\bar{\theta}_n - x)} dx.$$

From Scheffé's Lemma it follows that $C_n^{geo} \to_{a.s.} \int d\tilde{G}_0 = 1$. We have thus established that $\widehat{f}_n^{geo,sym}$ converges pointwise to $\tilde{g}_0(x)$ almost surely. The desired L_1 consistency then follows from Proposition 2(c) of Cule and Samworth (2010). □

Proof of Theorem 2.5

As in the proof of Lemma 2.3, one can show that it suffices to prove Theorem 2.5 when $\bar{\theta}_n \to_{a.s.} \theta_0$, and $y_n \to_{a.s.} 0$. Hence in what follows, we assume that $\bar{\theta}_n \to_{a.s.} \theta_0$, and $y_n \to_{a.s.} 0$.

First note that by theorem 2.2 of Dümbgen *et al.* (2011),

$$\text{int}(\text{dom}(\tilde{\psi}_0)) = \text{int}(\text{supp}(g_0)).$$

Also since $\tilde{g}_0^{sm} = e^{\tilde{\psi}_0^{sm}} > 0$ on $\mathbb{R}$, we find that $\text{int}(\text{dom}(\tilde{\psi}_0^{sm})) = \mathbb{R}$. Hence,

$$K \subset \text{int}(\text{dom}(\tilde{\psi}_0)) \subset \text{int}(\text{dom}(\tilde{\psi}_0^{sm})) = \mathbb{R}.$$

When $\tilde{g}_n$ equals $\widehat{g}_{\bar{\theta}_n}$, $\tilde{g}_n^{geo,sym}$, $\bar{f}_m(\bar{\theta}_n \pm \cdot)$, or $\widehat{f}_n^{sm}(\bar{\theta}_n \pm \cdot)$, $\tilde{g}_n$ is log-concave, which simplifies the proof of part A and B. We first consider the cases when $\tilde{g}_n = \widehat{g}_{\bar{\theta}_n}$, $\tilde{g}_n^{geo,sym}$ or $\bar{f}_m(\bar{\theta}_n \pm \cdot)$. Now to prove part A, note that

$$\begin{aligned}\int_{-\infty}^{\infty} |\tilde{g}_n(x+y_n) - \tilde{g}_0(x)|dx \leq & \int_{-\infty}^{\infty} |\tilde{g}_n(x+y_n) - \tilde{g}_0(x+y_n)|dx \\ & + \int_{-\infty}^{\infty} |\tilde{g}_0(x+y_n) - \tilde{g}_0(x)|dx,\end{aligned}$$

whose first term

$$\int_{-\infty}^{\infty} |\tilde{g}_n(x+y_n) - \tilde{g}_0(x+y_n)|dx = \int_{-\infty}^{\infty} |\tilde{g}_n(x) - \tilde{g}_0(x)|dx \to_{a.s.} 0,$$

which follows from either Lemma 2.3 or (2.52). Also, noting

$$\tilde{g}_0(x+y_n) \to_{a.s.} \tilde{g}_0(x) \quad a.e.\ x.\ \text{Lebesgue},$$

and applying Glick's Theorem (Theorem 2.6 Devroye, 1987), we derive that

$$\int_{-\infty}^{\infty} |\tilde{g}_0(x+y_n) - \tilde{g}_0(x)|dx \to_{a.s.} 0.$$

Thus we obtain that

$$\int_{-\infty}^{\infty} |\tilde{g}_n(x+y_n) - \tilde{g}_0(x)|dx \to_{a.s.} 0,$$

which combined with Proposition 2(c) of Cule and Samworth (2010) yields that

$$\tilde{g}_n(x+y_n) \to_{a.s.} \tilde{g}_0(x) \quad a.e.\ x$$

with respect to Lebesgue measure. As a consequence,

$$\tilde{\psi}_n(x+y_n) = \log(\tilde{g}_n(x+y_n)) \to_{a.s.} \tilde{\psi}_0(x) \quad a.e.\ x.\ \text{Lebesgue}.$$

Since $\tilde{\psi}_n$ is concave for our choices of $\tilde{g}_n$, Theorem 10.8 of Rockafellar (1970) indicates that pointwise convergence on $\mathbb{R}$ translates to uniform convergence on all compact sets inside $\text{int}(\text{dom}(\tilde{\psi}_0))$, leading to

$$\sup_{x \in K} |\tilde{\psi}_n(x+y_n) - \tilde{\psi}_0(x)| \to_{a.s.} 0.$$

As a corollary,

$$\sup_{x\in K}|\tilde{g}_n(x+y_n)-\tilde{g}_0(x)|\to_{a.s.}0.$$

Thus we have established part A and part B of Theorem 2.5 for $\tilde{g}_n=\widehat{g}_{\bar{\theta}_n}$, $\tilde{g}_n^{geo,sym}$ and $\bar{f}_m(\bar{\theta}_n\pm\cdot)$. Now consider the case when $\tilde{g}_n=\widehat{f}_n^{sm}(\bar{\theta}_n\pm x)$. Note that (2.47) and continuity of $\tilde{g}_0^{sm}$ imply that

$$\int_{-\infty}^{\infty}|\widehat{f}_n^{sm}(\bar{\theta}_n\pm x)-\tilde{g}_0^{sm}(x)|dx\to_{a.s.}0.$$

Therefore, following the same arguments as in the case of $\tilde{g}_n=\widehat{g}_{\bar{\theta}_n}$, $\tilde{g}_n^{geo,sym}$ and $\bar{f}_m(\bar{\theta}_n\pm\cdot)$, we can prove part A and part B for $\tilde{g}_n=\widehat{f}_n^{sm}(\bar{\theta}_n\pm\cdot)$.

Now we focus on proving part A when $\tilde{g}_n$ is the mixture density $\tilde{g}_n^{sym}$. From the definition of $\tilde{g}_n^{sym}$ in (2.10) we obtain that

$$\begin{aligned}&2\sup_{x\in K}|\tilde{g}_n^{sym}(x+y_n)-\tilde{g}_0(x)|\\&\quad\le\sup_{x\in K}|\bar{f}_m(\bar{\theta}_n+x+y_n)-\tilde{g}_0(x)|+\sup_{x\in K}|\bar{f}_m(\bar{\theta}_n-x+y_n)-\tilde{g}_0(x)|.\end{aligned}$$

Since we have already proved part A for $\bar{f}_m(\bar{\theta}_n\pm\cdot)$, it is not hard to see that both terms on the right hand side of the above display converge to 0 almost surely, which proves part A for $\tilde{g}_n^{sym}$. Part A follows for $(\widehat{f}_n^{sym})^{sm}$ in a similar way noting that (2.12) implies

$$\begin{aligned}&2\sup_{x\in K}|(\widehat{f}_n^{sym})^{sm}(x+y_n)-\tilde{g}_0^{sm}(x)|\\&\quad\le\sup_{x\in K}|\bar{f}_m^{sm}(\bar{\theta}_n+x+y_n)-\tilde{g}_0^{sm}(x)|+\sup_{x\in K}|\bar{f}_m^{sm}(\bar{\theta}_n-x+y_n)-\tilde{g}_0^{sm}(x)|.\end{aligned}$$

Now we prove part B of Theorem 2.5 when $\tilde{g}_n=\tilde{g}_n^{sym}$. To this end, note that we can write

$$\begin{aligned}\sup_{x\in K}(\tilde{\psi}_n^{sym}(x+y_n)-\tilde{\psi}_0(x))&=\sup_{x\in K}\log\left(\frac{\tilde{g}_n^{sym}(x+y_n)-\tilde{g}_0(x)}{\tilde{g}_0(x)}+1\right)\\&\le\sup_{x\in K}\frac{\tilde{g}_n^{sym}(x+y_n)-\tilde{g}_0(x)}{\tilde{g}_0(x)},\end{aligned}$$

because $\log(1+x) \leq x$ for any $x > -1$. Similarly we can show that

$$\sup_{x \in K}(\tilde{\psi}_0(x) - \tilde{\psi}_n^{sym}(x+y_n)) \leq \sup_{x \in K} \frac{\tilde{g}_0(x) - \tilde{g}_n^{sym}(x+y_n)}{\tilde{g}_n^{sym}(x+y_n)},$$

leading to

$$\sup_{x \in K} \left| \tilde{\psi}_n^{sym}(x+y_n) - \tilde{\psi}_0(x) \right| \leq \frac{\sup_{x \in K} |\tilde{g}_n^{sym}(x+y_n) - \tilde{g}_0(x)|}{\min\left(\inf_{x \in K} \tilde{g}_n^{sym}(x+y_n), \inf_{x \in K} \tilde{g}_0(x) \right)},$$

whose numerator converges to 0 almost surely by part A of Theorem 2.5. Thus to prove part B for $\tilde{\psi}_n^{sym}$, we only need to show that the denominator of the term on the right hand side of last display is bounded away from 0. To this end, notice that since $\tilde{g}_0$ is unimodal, and $\tilde{g}_n^{sym}$ is a convex combination of two unimodal densities, they attain their infimum over K at either of the endpoints. By part A, it also follows that as $n \to \infty$, the respective minimum values approach $\inf_K \tilde{g}_0$, which is bounded away from 0 as K is a closed subset of $\text{int}(\text{dom}(\tilde{\psi}_0))$. Thus part B is proved for $\tilde{\psi}_n^{sym}$.

One can prove part B for $(\widehat{f}_n^{sym})^{sm}$ in the similar fashion because $\tilde{g}_0^{sm}$ is unimodal, and $(\widehat{f}_n^{sym})^{sm}$ is a convex combination of two unimodal densities.

Now we can proceed to prove part C of the Theorem 2.5. First, we consider the log-concave density estimators $\widehat{g}_{\bar{\theta}_n}$, $\tilde{g}_n^{geo,sym}$, and $\bar{f}_m(\bar{\theta}_n \pm \cdot)$, all of which are strongly L_1 consistent for $\tilde{g}_0$. Note that if $\tilde{\psi}_0'$ is continuous at some $x \in \text{int}(\text{dom}(\tilde{\psi}_0))$, then there exists a neighborhood $[x-\xi, x+\xi]$ where $\tilde{\psi}_0$ is continuous. It follows that for sufficiently large n, $x + y_n \in [x-\xi, x+\xi]$ with probability 1. Since $\tilde{\psi}_n$ is concave, and converges to $\tilde{\psi}_0$ uniformly over K by part B, Lemma 3.10 of Seijo and Sen (2011) entails that

$$\sup_{t \in [x-\xi, x+\xi]} |\tilde{\psi}_n'(t\pm) - \tilde{\psi}_0'(t)| \to_{a.s.} 0, \tag{2.53}$$

from which we obtain that

$$|\tilde{\psi}_n'(x+y_n) - \tilde{\psi}_0'(x+y_n)| \to_{a.s.} 0,$$

which results in

$$|\tilde{\psi}_n'(x+y_n) - \tilde{\psi}_0'(x)| \to_{a.s.} 0 \tag{2.54}$$

because $\tilde{\psi}_0'$ is continuous at x, thereby proving part C for $\tilde{g}_n = \widehat{g}_{\bar{\theta}_n}$, $\tilde{g}_n^{geo,sym}$, and $\bar{f}_m(\bar{\theta}_n \pm \cdot)$. For $\tilde{g}_n = \widehat{f}_n^{sm}(\bar{\theta}_n \pm x)$ part C follows in the similar manner and hence, we skip the proof here. To prove part C for the non-logconcave estimators, $\tilde{g}_n^{sym}$ and $(\widehat{f}_n^{sym})^{sm}$, we have to exploit their connection to $\bar{f}_m$ and $\bar{f}_m^{sm}$, respectively. Beginning with $\tilde{g}_n^{sym}$, we observe that

$$(\tilde{\psi}_n^{sym})'(x) = \varrho_n(x)\Big((\log \bar{f}_m)'(\bar{\theta}_n + x)\Big) - (1-\varrho_n(x))\Big((\log \bar{f}_m)'(\bar{\theta}_n - x)\Big) \tag{2.55}$$

where $\varrho_n(x) = \bar{f}_m(\bar{\theta}_n + x)/2\tilde{g}_n^{sym}(x) \leq 1$. Thus

$$\begin{aligned} |(\tilde{\psi}_n^{sym})'(x+y_n) - \tilde{\psi}_0'(x)| &\leq \varrho_n(x+y_n)|(\log \bar{f}_m)'(\bar{\theta}_n + x + y_n) - \tilde{\psi}_0'(x)| \\ &+ \ (1-\varrho_n(x+y_n))|(\log \bar{f}_m)'(\bar{\theta}_n - x - y_n) - \tilde{\psi}_0'(-x)|. \end{aligned}$$

Since ϱ_n is uniformly bounded above by 1, (2.54) (applied on $\log \bar{f}_m(\bar{\theta}_n \pm \cdot)$) completes the proof of part C for $\tilde{\psi}_n^{sym}$. Analogously using (2.55), for $\tilde{g}_n = (\widehat{f}_n^{sym})^{sm}$, one can show that

$$((\tilde{\psi}_n^{sym})^{sm})'(x) = \varrho_n^{sm}(x)\Big((\log \bar{f}_m^{sm})'(\bar{\theta}_n + x)\Big) - (1-\varrho_n^{sm}(x))\Big((\log \bar{f}_m^{sm})'(\bar{\theta}_n - x)\Big) \tag{2.56}$$

for some $\varrho_n^{sm}(x) < 1$. Therefore following the same steps as in the case of $\tilde{g}_n^{sym}$, we can establish part C for $(\widehat{f}_n^{sym})^{sm}$. □

Proof of Lemma 2.4

As in the proof of Lemma 2.3, one can show that it suffices to prove Lemma 2.4 when $\bar{\theta}_n$ is a strongly consistent estimator of θ_0.

We will show that $\widehat{\mathcal{I}}_n(\eta) \to_{a.s.} \mathcal{I}_{f,g}(\eta)$ and $\tilde{\theta}_n \to_{a.s.} \theta_0$ when $\tilde{g}_n$ equals $\tilde{g}_n^{sym}$, $\widehat{g}_{\bar{\theta}_n}$ or $\tilde{g}_n^{geo,sym}$. The proof in the case of $\tilde{g}_n = (\widehat{f}_n^{sym})^{sm}$ follows in a similar way.

We will prove the consistency of $\widehat{\mathcal{I}}_n(\eta)$ first. Since Lemma 2.3 indicates that $||\tilde{g}_n - \tilde{g}_0||_1 \to_{a.s.} 0$, the corresponding distribution functions satisfy $||\tilde{G}_n - \tilde{G}_0(\eta)||_\infty \to_{a.s.} 0$, which

combined with the continuity of $\tilde{G}_0$ entails that

$$-\xi_n = \tilde{G}_n^{-1}(\eta) \to_{a.s.} \tilde{G}_0^{-1}(\eta) = -\xi_0; \tag{2.57}$$

and

$$\tilde{G}_n^{-1}(1-\eta) \to_{a.s.} \tilde{G}_0^{-1}(1-\eta) = \xi_0. \tag{2.58}$$

Combined with strong consistency of $\bar{\theta}_n$, the above leads to

$$\bar{\theta}_n + \xi_n = \bar{\theta}_n + \tilde{G}_n^{-1}(1-\eta) \to_{a.s.} \theta_0 + \xi_0 = \tilde{F}_0^{-1}(1-\eta), \tag{2.59}$$

and

$$\bar{\theta}_n - \xi_n = \bar{\theta}_n + \tilde{G}_n^{-1}(\eta) \to_{a.s.} \theta_0 - \xi_0 = \tilde{F}_0^{-1}(\eta). \tag{2.60}$$

Observe that

$$\begin{aligned} |\widehat{\mathcal{I}}_n(\eta) - \mathcal{I}_{f,g}(\eta)| \leq & \left| \int_{\bar{\theta}_n - \xi_n}^{\bar{\theta}_n + \xi_n} \tilde{\psi}_n'(x - \bar{\theta}_n)^2 d(\mathbb{F}_n - F_0)(x) \right| \\ & + \left| \int_{\bar{\theta}_n - \xi_n}^{\bar{\theta}_n + \xi_n} \tilde{\psi}_n'(x - \bar{\theta}_n)^2 dF_0(x) - \mathcal{I}_{f,g}(\eta) \right|. \end{aligned} \tag{2.61}$$

Let us consider the first term on the right hand side of (2.61) first. To this end, note that Lemma 2.9 combined with (2.57) and (2.58) imply that

$$\limsup_n \sup_{z \in [-\xi_n, \xi_n]} |\tilde{\psi}_{0n}'(z)| \leq C_{\xi_0} \quad a.s., \tag{2.62}$$

where $C(-\xi_0, \xi_0) > 0$ is a finite constant depending on the truncation parameter η via $\xi_0 = \tilde{G}_0^{-1}(1-\eta)$. For the sake of simplicity, we denote $C(-\xi_0, \xi_0)$ by M_η. Suppose $\mathcal{U}_\eta$ is the class of non-increasing functions bounded by M_η, i.e.,

$$\mathcal{U}_\eta = \left\{ h : \mathbb{R} \mapsto [-M_\eta, M_\eta] \;\middle|\; h \text{ is non-increasing} \right\},$$

and denote by $\mathcal{F}_\eta$ the class

$$\begin{aligned} \mathcal{F}_\eta = \Bigg\{ h^2 : \mathbb{R} \mapsto & [0, M_\eta^2] \;\Bigg|\; h(x) = u(x) 1_{[r_1, r_2]}(x), \\ & u \in \mathcal{U}_\eta,\ [r_1, r_2] \subset \left(\tilde{F}_0^{-1}(\eta/2), \tilde{F}_0^{-1}(1-\eta/2) \right) \Bigg\}. \end{aligned}$$

Theorem 2.7.5 of Van der Vaart and Wellner (1996) (pp. 159) indicates that for any $\epsilon > 0$, the bracketing entropy

$$\log N_{[\,]}(\epsilon, \mathcal{U}_\eta, L_2(P_0)) \lesssim \epsilon^{-1}. \tag{2.63}$$

Also by Example 2.5.4 of Van der Vaart and Wellner (1996), the class $\mathcal{F}_I$ of all indicator functions of the form $1_{[z_1,z_2]}$, where $z_1 \leq z_2$ with $z_1, z_2 \in \mathbb{R}$, satisfies

$$\log N_{[\,]}(\epsilon, \mathcal{F}_I, L_2(P_0)) \lesssim 2/\epsilon^2, \tag{2.64}$$

implying that

$$\log N_{[\,]}(\epsilon, \mathcal{F}_\eta, L_2(P_0)) \lesssim \epsilon^{-1}.$$

Since $\tilde{\psi}_{0n}'$ is non-increasing, (2.62) implies that $(\tilde{\psi}_{0n}')^2(\cdot - \bar{\theta}_n)$ restricted to $[\bar{\theta}_n - \xi_n, \bar{\theta}_n + \xi_n]$ belongs to $\mathcal{F}_\eta$. Therefore by Glivenko-Cantelli Theorem (see Theorem 2.4.1 of Van der Vaart and Wellner (1996)),

$$\int_{\bar{\theta}_n - \xi_n}^{\bar{\theta}_n + \xi_n} \tilde{\psi}_n'(x - \bar{\theta}_n)^2 d(\mathbb{F}_n - F_0)(x) \to_{a.s.} 0. \tag{2.65}$$

Now we claim that the second term in (2.61) also approaches 0 almost surely. This follows by Theorem 2.5(C) and the dominated convergence theorem since $\tilde{\psi}_{0n}'$ is bounded on $[-\xi_n, \xi_n]$, completing the proof of $\widehat{\mathcal{I}}_n(\eta) \to_{a.s.} \mathcal{I}_{f,g}(\eta)$.

Our next aim is to prove the consistency of $\tilde{\theta}_n$. Observe that from (2.16) it follows that

$$\begin{aligned} |\tilde{\theta}_n - \bar{\theta}_n| \leq \widehat{\mathcal{I}}_n(\eta)^{-1} &\left| \int_{\bar{\theta}_n - \xi_n}^{\bar{\theta}_n + \xi_n} \tilde{\psi}_n'(x - \bar{\theta}_n) d(\mathbb{F}_n - F_0)(x) \right| \\ &+ \widehat{\mathcal{I}}_n(\eta)^{-1} \left| \int_{\bar{\theta}_n - \xi_n}^{\bar{\theta}_n + \xi_n} \tilde{\psi}_n'(x - \bar{\theta}_n) dF_0(x) \right|. \end{aligned}$$

Note that $\widehat{\mathcal{I}}_n(\eta) \to_{a.s.} \mathcal{I}_{f,g}(\eta)$ by the first part of the current lemma. The proof of

$$\widehat{\mathcal{I}}_n(\eta)^{-1} \left| \int_{\bar{\theta}_n - \xi_n}^{\bar{\theta}_n + \xi_n} \tilde{\psi}_n'(x - \bar{\theta}_n) d(\mathbb{F}_n - F_0)(x) \right| \to_{a.s.} 0$$

is very similar to the proof of (2.65), and follows by Glivenko-Cantelli Theorem (see Theorem 2.4.1 of Van der Vaart and Wellner (1996)) noting that $\tilde{\psi}_{0n}'(\cdot - \bar{\theta}_n)$ restricted to $[-\xi_n, \xi_n]$ is a member of $\mathcal{U}_\eta$, which has finite bracketing entropy by (2.64).

Since $\tilde{\psi}_{0n}'$ is bounded on $[-\xi_n, \xi_n]$, (2.59), (2.60), and Theorem 2.5(C) combined with the dominated convergence theorem further entail that

$$\int_{\bar{\theta}_n - \xi_n}^{\bar{\theta}_n + \xi_n} \tilde{\psi}_n'(x - \bar{\theta}_n) dF_0(x) \to_{a.s.} \int_{\tilde{G}_0^{-1}(\eta)}^{\tilde{G}_0^{-1}(1-\eta)} \tilde{\psi}_0'(x) g_0(x) dx,$$

which is 0 because $\tilde{\psi}_0'$ is an odd function while g_0 is an even function. Therefore strong consistency of $\tilde{\theta}_n$ follows from that of $\bar{\theta}_n$. □

Proof of Theorem 2.6

As in the proof of Lemma 2.3, one can show that it suffices to prove Theorem 2.6 when $\bar{\theta}_n \to_{a.s.} \theta_0$. Therefore in what follows, we assume that $\bar{\theta}_n \to_{a.s.} \theta_0$.

First we consider the case when $\tilde{g}_n$ equals $\tilde{g}_n^{sym}$, $\widehat{g}_{\bar{\theta}_n}$ or $\tilde{g}_n^{geo,sym}$, so that by Lemma 2.3, the density estimator $\tilde{g}_n$ converges to $\tilde{g}_0$ in L_1 almost surely. From (2.16) we deduce that

$$-(\tilde{\theta}_n - \bar{\theta}_n) = \int_{\bar{\theta}_n - \xi_n}^{\bar{\theta}_n + \xi_n} \frac{\tilde{\psi}_n'(x - \bar{\theta}_n)}{\widehat{\mathcal{I}}_n(\eta)} d\mathbb{F}_n(x) = \int_{-\xi_n}^{\xi_n} \frac{\tilde{\psi}_n'(z)}{\widehat{\mathcal{I}}_n(\eta)} d\mathbb{F}_n(z + \bar{\theta}_n).$$

Denoting $\tilde{\delta}_n = \theta_0 - \bar{\theta}_n$, we observe that the expression in the above display can also be written as

$$\begin{aligned}
&\int_{-\xi_n}^{\xi_n} \frac{\tilde{\psi}_{0n}'(z) - \tilde{\psi}_0'(z - \tilde{\delta}_n)}{\widehat{\mathcal{I}}_n(\eta)} d(\mathbb{F}_n(z + \bar{\theta}_n) - F_0(z + \bar{\theta}_n)) \\
&\quad + \int_{-\xi_n}^{\xi_n} \frac{\tilde{\psi}_{0n}'(z)}{\widehat{\mathcal{I}}_n(\eta)} \Big(f_0(z + \bar{\theta}_n) - g_0(z) \Big) dz + \int_{-\xi_n}^{\xi_n} \frac{\tilde{\psi}_{0n}'(z) - \tilde{\psi}_0'(z)}{\widehat{\mathcal{I}}_n(\eta)} g_0(z) dz \\
&\quad + \int_{-\xi_n}^{\xi_n} \frac{\tilde{\psi}_0'(z)}{\widehat{\mathcal{I}}_n(\eta)} g_0(z) dz + \int_{-\xi_n}^{\xi_n} \frac{\tilde{\psi}_0'(z - \tilde{\delta}_n)}{\widehat{\mathcal{I}}_n(\eta)} d(\mathbb{F}_n(z + \bar{\theta}_n) - F_0(z + \bar{\theta}_n)) \\
&= T_{1n} + T_{2n} + T_{3n} + T_{4n} + T_{5n}.
\end{aligned}$$

Observe that T_{3n} and T_{4n} vanish since $\tilde{\psi}_{0_n}'$ and $\tilde{\psi}_0'$ are odd functions while $\tilde{f}_0$ is an even function. Now note that to prove (2.31), it suffices to show that $\sqrt{n}T_{1n} = o_p(1)$,

$$\frac{T_{2n}}{\tilde{\delta}_n} \to_{a.s.} \int_{-\xi_0}^{\xi_0} \frac{\tilde{\psi}_0'(z)\psi_0'(z)}{\mathcal{I}_{f,g}(\eta)} g_0(z)dz, \tag{2.66}$$

and that $\sqrt{n}T_{5n}$ converges weakly to a centered Gaussian random variable with variance $\mathcal{I}_{f,g}(\eta)^{-1}$. When $\tilde{g}_n = (\widehat{f}_n^{sym})^{sm}$, (2.33) can also be proved in a similar way, and hence we skip the proof. In the special case when f_0, or equivalently g_0 itself is log-concave, the log-concave approximation $\tilde{g}_0$ equals g_0. Hence $\tilde{\psi}_0 = \tilde{\psi}_0^{sm} = \psi_0$ follows, indicating $\mathcal{I}_{f,g}(\eta) = \mathcal{I}_{f,g}^{sm}(\eta) = \mathcal{I}_{f_0}(\eta)$, and $\gamma_\eta = 1$, which establishes (2.35), and completes the proof of the current Theorem.

We will first show that $\sqrt{n}T_{1n} = o_p(1)$. Recall that in section 2.1.1 we denoted the empirical process $\sqrt{n}(\mathbb{F}_n - F_0)$ by $\mathbb{Z}_n$. Denote by h_n the function

$$h_n(x) = (\tilde{\psi}_{0_n}'(x - \bar{\theta}_n) - \tilde{\phi}_0'(x))1_{[\bar{\theta}_n - \xi_n, \bar{\theta}_n + \xi_n]}(x),$$

where $\tilde{\phi}_0'(x)$ was previously defined as $\psi_0'(x - \theta_0)$. It follows that

$$\begin{aligned} \sqrt{n}T_{1n} &= \sqrt{n} \int_{\bar{\theta}_n - \xi_n}^{\bar{\theta}_n + \xi_n} \frac{\tilde{\psi}_{0_n}'(x - \bar{\theta}_n) - \tilde{\phi}_0'(x)}{\widehat{\mathcal{I}}_n(\eta)} d(\mathbb{F}_n - F_0)(x) \\ &= \int_{-\infty}^{\infty} \frac{h_n(x)}{\widehat{\mathcal{I}}_n(\eta)} d\mathbb{Z}_n(x). \end{aligned} \tag{2.67}$$

Lemma 2.9 combined with (2.57) and (2.58) imply that

$$\limsup_n \sup_{z \in [-\xi_n, \xi_n]} |\tilde{\psi}_{0_n}'(z)| \le C_{\xi_0} \quad a.s.,$$

where C_{ξ_0} is a positive constant depending on ξ_0. It is not hard to see that since $\tilde{\phi}_0'$ is monotone, $|\tilde{\phi}_0'|$ attains its maxima over $[\bar{\theta}_n - \xi_n, \bar{\theta}_n + \xi_n]$ at either of the endpoints. Though $[\bar{\theta}_n - \xi_n, \bar{\theta}_n + \xi_n]$ is random, (2.58) and (2.59) indicate that with probability one, this interval is a subset of $[\tilde{F}_0^{-1}(\eta/2), \tilde{F}_0^{-1}(1 - \eta/2)]$ for all sufficiently large n. Hence it follows that

$$\limsup_n \sup_{x \in [\bar{\theta}_n - \xi_n, \bar{\theta}_n + \xi_n]} |\tilde{\psi}_0'(x)| < M_\eta \quad a.s. \tag{2.68}$$

for some $M_\eta > 0$, which can be chosen in such a way so that $C_{\xi_0} \leq M_\eta$. Therefore, we obtain that

$$\limsup_n \sup_{x \in [\bar{\theta}_n - \xi_n, \bar{\theta}_n + \xi_n]} |h_n(x)| < M_\eta \quad a.s. \tag{2.69}$$

Now define the class $\mathcal{H}_\eta$ by

$$\mathcal{H}_\eta = \Bigg\{ h : \mathbb{R} \mapsto [-M_\eta, M_\eta] \;\Bigg|\; h(x) = (u(x) - \tilde{\phi}_0'(x)) 1_{[r_1, r_2]}(x), \\ u \in \mathcal{U}_\eta,\ [r_1, r_2] \subset \left(\tilde{F}_0^{-1}(\eta/2), \tilde{F}_0^{-1}(1 - \eta/2) \right) \Bigg\},$$

where $\mathcal{U}_\eta$ is the class of non-increasing functions bounded by $M_\eta > 0$, i.e.

$$\mathcal{U}_\eta = \Big\{ h : \mathbb{R} \mapsto [-M_\eta, M_\eta] \;\Big|\; h \text{ is non-increasing} \Big\}.$$

Observe that (2.59), (2.60), and (2.69) imply $h_n \in \mathcal{H}_\eta$ almost surely for all sufficiently large n.

From (2.63) and (2.64) we derive that

$$\log N_{[\,]}(\epsilon, \mathcal{H}_\eta, L_2(P_0)) \lesssim \frac{1}{\epsilon}.$$

Therefore we conclude that the bracketing integral

$$\mathcal{J}_{[\,]}(1, \mathcal{H}_\eta, L_2(P_0)) = \int_0^1 \sqrt{1 + \log N_{[]}(M_\eta \epsilon, \mathcal{H}_\eta, L_2(P_0))} d\epsilon \lesssim \int_0^1 \epsilon^{-1/2} d\epsilon,$$

which is finite, thereby establishing that $\mathcal{H}_\eta$ is P_0-Donsker. Also notice that $|h_n|$ is bounded above by M_η, and Theorem 2.5(C) indicates that h_n converges to 0 at the continuity points of $\tilde{\phi}_0'$. Since $\tilde{\phi}_0$ is concave, $\tilde{\phi}_0'$ is continuous almost everywhere with respect to Lebesgue measure on $\mathrm{dom}(\tilde{\phi}_0)$. Hence the dominated convergence theorem yields

$$\int_{-\infty}^{\infty} h_n(x)^2 dF_0(x) \to_{a.s.} 0.$$

Since $\mathcal{H}_\eta$ is Donsker, Theorem 2.1 of van der Vaart and Wellner (2007) then implies that

$$\int_{-\infty}^{\infty} h_n(x) d\mathbb{Z}_n(x) = o_p(1).$$

Finally, an application of Lemma 2.4 leads to $\widehat{\mathcal{I}}_n(\eta) \to_{a.s.} \mathcal{I}_{f_0}(\eta)$, and thus from (2.67) we conclude that $\sqrt{n}T_{1n} = o_p(1)$.

Now we turn our attention to T_{2n}. Here we aim to prove (2.66). Observe that $T_{2n}/\tilde{\delta}_n$ can be written as

$$\begin{aligned}\int_{-\xi_n}^{\xi_n} \frac{\tilde{\psi}_{0n}'(z)}{\widehat{\mathcal{I}}_n(\eta)} \frac{\Big(g_0(z-\tilde{\delta}_n) - g_0(z)\Big)}{\tilde{\delta}_n} dz &= \int_{-\xi_n}^{\xi_n} \frac{\tilde{\psi}_{0n}'(z)}{\widehat{\mathcal{I}}_n(\eta)} \frac{\int_z^{z-\tilde{\delta}_n} g_0'(t)dt}{\tilde{\delta}_n} dz \\ &= \int_{-\xi_n-\tilde{\delta}_n}^{\xi_n} g_0'(t) \frac{\int_t^{t+\tilde{\delta}_n} \tilde{\psi}_{0n}'(z)dz}{\tilde{\delta}_n \widehat{\mathcal{I}}_n(\eta)} dt, \end{aligned} \tag{2.70}$$

where the equality follows by Fubini's theorem since g_0 is absolutely continuous. The concavity of $\tilde{\psi}_n$ implies that $\tilde{\psi}_{0n}'$ is non-increasing. Hence for any $t \in \text{int}(\text{dom}(\tilde{\psi}_0))$, we have,

$$\min\Big\{\tilde{\psi}_{0n}'(t), \tilde{\psi}_{0n}'(t+\tilde{\delta}_n)\Big\} \le \frac{\int_t^{t+\tilde{\delta}_n} \tilde{\psi}_{0n}'(z)dz}{\tilde{\delta}_n} \le \max\Big\{\tilde{\psi}_{0n}'(t), \tilde{\psi}_{0n}'(t+\tilde{\delta}_n)\Big\}. \tag{2.71}$$

Therefore if $\tilde{\psi}_0'$ is continuous at $t \in \text{int}(\text{dom}(\tilde{\psi}_0))$, Theorem 2.5(C) leads to

$$\frac{\int_t^{t+\tilde{\delta}_n} \tilde{\psi}_{0n}'(z)dz}{\tilde{\delta}_n} \to_{a.s.} \tilde{\psi}_0'(t) \quad \text{as } n \to \infty.$$

Since $\tilde{\psi}_0$ is concave, $\tilde{\psi}_0'$ is continuous almost everywhere with respect to Lebesgue measure on $\text{dom}(\tilde{\psi}_0)$. Since $\tilde{\psi}_{0n}'$ is non-increasing, (2.71) further entails that

$$\sup_{t \in [-\xi_n - \tilde{\delta}_n, \xi_n]} \frac{\Big|\int_t^{t+\tilde{\delta}_n} \tilde{\psi}_{0n}'(z)dz\Big|}{|\tilde{\delta}_n|} \le \tilde{\psi}_{0n}'(-\xi_n - |\tilde{\delta}_n|),$$

which can be bounded above by $\tilde{\psi}_0'(\tilde{G}_0^{-1}(\eta/2))$ using (2.57) and the fact that $\tilde{\delta}_n \to_{a.s.} 0$. Hence, we deduce that

$$|g_0'(t)| \frac{\Big|\int_t^{t+\tilde{\delta}_n} \tilde{\psi}_{0n}'(z)dz\Big|}{|\tilde{\delta}_n|} 1_{[-\xi_n-\tilde{\delta}_n, \xi_n]}(z) \le |g_0'(t)| \Big(\tilde{\psi}_0'(\tilde{G}_0^{-1}(\eta/2))\Big).$$

Note that the symmetry of g_0 about 0 indicates that

$$\int_{-\infty}^{\infty} |g_0'(t)|dt = 2\int_{-\infty}^{0} g_0'(t)dt = 2g(0),$$

which is finite because of (2.20). Therefore (2.70), Lemma 2.4, and the dominated convergence theorem yield

$$\frac{T_{2n}}{\tilde{\delta}_n} \to_{a.s.} \int_{-\xi_0}^{\xi_0} \frac{\tilde{\psi}_0'(z)g_0'(z)}{\mathcal{I}_{f,g}(\eta)}dz,$$

which completes the proof of (2.66) since $g_0'(z) = \psi_0'(z)g_0(z)$.

Now it only remains to show that the last term T_{5n} converges weakly to a Gaussian random variable with variance $\mathcal{I}_{f,g}^{-1}$. Observe that a change of variable leads to

$$\begin{aligned}\sqrt{n}T_{5n} &= \sqrt{n}\int_{\bar{\theta}_n-\xi_n}^{\bar{\theta}_n+\xi_n} \frac{\tilde{\psi}_0'(x-\theta_0)}{\widehat{\mathcal{I}}_n(\eta)}d(\mathbb{F}_n - F_0)(x) \\ &= \int_{\theta_0-\xi_0}^{\theta_0+\xi_0} \frac{\tilde{\psi}_0'(x-\theta_0)}{\widehat{\mathcal{I}}_n(\eta)}d\mathbb{Z}_n(x) \\ &\quad + \int_{-\infty}^{\infty} 1_{C_n}(x)\frac{\tilde{\psi}_0'(x-\theta_0)}{\widehat{\mathcal{I}}_n(\eta)}d\mathbb{Z}_n(x), \end{aligned} \tag{2.72}$$

where

$$C_n = [\bar{\theta}_n - \xi_n, \bar{\theta}_n + \xi_n] \setminus [\theta_0 - \xi_0, \theta_0 + \xi_0].$$

From Lemma 2.4 and Slutsky's theorem it follows that

$$\int_{\theta_0-\xi_0}^{\theta_0+\xi_0} \frac{\tilde{\psi}_0'(x-\theta_0)}{\widehat{\mathcal{I}}_n(\eta)}d\mathbb{Z}_n(x) = \int_{\theta_0-\xi_0}^{\theta_0+\xi_0} \frac{\tilde{\psi}_0'(x-\theta_0)}{\mathcal{I}_{f,g}(\eta)}d\mathbb{Z}_n(x)(1+o_p(1)),$$

which equals

$$\int_{\theta_0-\xi_0}^{\theta_0+\xi_0} \frac{\tilde{\psi}_0'(x-\theta_0)}{\mathcal{I}_{f,g}(\eta)}d\mathbb{Z}_n(x) + o_p(1),$$

because the Central limit theorem and (2.30) imply that the first term on the above display is $O_p(1)$.

To deal with the second term, observe that the indicator function 1_{C_n} belongss to $\mathcal{F}_1$, the class of all indicator functions of the form $1_{[z_1,z_2]\setminus[\theta_0-\xi_0,\theta_0+\xi_0]}$, where $z_1 \le z_2$ with $z_1, z_2 \in \mathbb{R}$.

Since the latter class is Donsker by (2.64), Theorem 2.1 of van der Vaart and Wellner (2007) entails that the second term on the right hand side of (2.72) is of order $o_p(1)$ provided the following holds:

$$\int_{-\infty}^{\infty} 1_{C_n}(x) \frac{\tilde{\psi}_0'(x-\theta_0)^2}{\widehat{\mathcal{I}}_n(\eta)^2} dF_0(x) \to_{a.s.} 0.$$

To this end note that since $\tilde{\psi}_0'$ is non-increasing, (2.59), (2.60), (2.68), and Lemma 2.4 entail that for all sufficiently large n,

$$\int_{-\infty}^{\infty} 1_{C_n}(x) \frac{\tilde{\psi}_0'(x-\theta_0)^2}{\widehat{\mathcal{I}}_n(\eta)^2} dF_0(x) \leq \frac{\tilde{\psi}_0'(\tilde{G}_0^{-1}(\eta/2))^2}{\mathcal{I}_{f,g}^2(\eta)} |\bar{\theta}_n - \theta_0| \left(\sup_x f_0(x) \right) \quad a.s.$$

Since $f_0 \in \mathcal{P}_1$, from the definition of $\mathcal{P}_1$ in (2.20) and the strong consistency of $\bar{\theta}_n$ we obtain that the term on the right hand side of the last display converges to 0 almost surely. Thus we conclude that $\sqrt{n}T_{5n}$ is asymptotically distributed as a centered Gaussian random variable with variance $\mathcal{I}_{f,g}(\eta)^{-1}$, which completes the proof. □

2.6.2 Proofs for the MLE

We first state and prove a useful theorem which plays a crucial role in proving Proposition 1 and Lemma 2.1. This theorem concerns the continuity of the function $L(\cdot\,; F) : \mathbb{R} \mapsto \mathbb{R}$ given by

$$L(\theta; F) = \sup_{\psi \in \mathcal{SC}_0} \Psi(\theta, \psi, F) \tag{2.73}$$

where F is a fixed distribution function, and Ψ is as defined in (2.3).

Theorem 2.10. *Suppose the distribution function F satisfies Condition A. Then the map $\theta \mapsto L(\theta; F)$ is continuous on $\mathbb{R}$, where the function $L(\theta; F)$ was defined in (2.73).*

Proof. Suppose $\theta_k \to \theta$ as $k \to \infty$. Observe that $\Psi(\theta, \psi, F)$ can also be written as

$$\Psi(\theta, \psi, F) = \Psi(0, \psi, F(\cdot + \theta)).$$

Hence to prove Theorem 2.10, it suffices to show that as $k \to \infty$,

$$L(\theta_k; F) = \sup_{\psi \in \mathcal{SC}_0} \Psi(0, \psi, F(\cdot + \theta_k)) \to \sup_{\psi \in \mathcal{SC}_0} \Psi(0, \psi, F(\cdot + \theta)) = L(\theta, F).$$

Proposition 6 of Xu and Samworth (2017) implies that under Condition A, the convergence in the above display holds if the Wasserstein distance

$$d_W(F(\cdot + \theta_k), F(\cdot + \theta)) \to 0, \quad \text{as } k \to \infty. \tag{2.74}$$

Now by Theorem 6.9 of Villani (2009) (see also Theorem 7.12 of Villani, 2003), (2.74) follows if the followings hold as $k \to \infty$:

$$F(\cdot + \theta_k) \to_d F(\cdot + \theta), \tag{2.75}$$

and

$$\int_{-\infty}^{\infty} |x| dF(x + \theta_k) \to \int_{-\infty}^{\infty} |x| dF(x + \theta). \tag{2.76}$$

To prove (2.75), note that for any bounded continuous function h,

$$\int_{-\infty}^{\infty} h(x - \theta_k) dF(x) \to \int_{-\infty}^{\infty} h(x - \theta) dF(x)$$

by the dominated convergence theorem since $\theta_k \to \theta$ as $k \to \infty$. For proving (2.76), first notice that F has finite first moment by Condition A. Therefore another application of the dominated convergence yields that as $\theta_k \to \theta$,

$$\int_{-\infty}^{\infty} |x| dF(x + \theta_k) = \int_{-\infty}^{\infty} |x - \theta_k| dF(x) \to \int_{-\infty}^{\infty} |x| dF(x + \theta),$$

which, combined with (2.75), leads to (2.74), and completes the proof.

□

Proof of Proposition 1

Our first step is to show that $L(F)$ is finite. From the definition of Ψ in (2.3), it is not hard to see that

$$L(F) \leq \sup_{\psi \in \mathcal{C}} \left(\int_{-\infty}^{\infty} \psi(x) dF(x) - \int_{-\infty}^{\infty} e^{\psi(x)} dx \right),$$

where $\mathcal{C}$ denotes the set of all real-valued concave functions. Theorem 2.2 of Dümbgen *et al.* (2011) entails that under condition A, the term on the right hand side of the above display is finite. Therefore $L(F) < \infty$ follows. To show that $L(F) > -\infty$, we show that the map $x \mapsto -|x| \in \mathcal{SC}_0$. Therefore (2.3) and (2.5) lead to

$$L(F) > -\int_{-\infty}^{\infty} |x| dF(x) - \int_{-\infty}^{\infty} e^{-|x|} dx > -\infty,$$

which follows from Condition A. Therefore we conclude that $L(F) \in \mathbb{R}$.

Now we have to show that there exist $\theta^*(F) \in \mathbb{R}$ and $\psi^*(F) \in \mathcal{SLC}_0$ such that

$$\Psi(\theta^*(F), \psi^*(F), F) = \sup_{\theta \in \mathbb{R}, \psi \in \mathcal{SLC}_0} \Psi(\theta, \psi, F) = \sup_{\theta \in \mathbb{R}} L(\theta; F) = L(F).$$

Now there exists a sequence $\{\theta_k\}_{k \geq 1}$ such that $L(\theta_k; F) \uparrow L(F)$ as $k \to \infty$. Suppose the sequence $\{\theta_k\}_{k \geq 1}$ is bounded. Then we can find a subsequence $\{\theta_{k_r}\}_{r \geq 1}$ converging to some $\theta' \in \mathbb{R}$. Since the map $L(\theta; F)$ is continuous in θ by Theorem 2.10, we also have

$$L(\theta'; F) = \lim_{r \to \infty} L(\theta_{k_r}; F) = L(F),$$

which implies that θ' is a maximizer of $L(\theta; F)$. The rest of the proof then follows from Proposition 4(iii) of Xu and Samworth (2017), which states that for each $\theta \in \mathbb{R}$, there exists a unique log-density ψ_θ, which maximizes $\Psi(\theta, \psi, F)$ in $\psi \in \mathcal{SC}_0$ when F satisfies condition A. Hence $(\theta', \psi_{\theta'})$ is a candidate for $(\theta^*(F), \psi^*(F))$. Thus to complete the proof, it remains to show that $\{\theta_k\}_{k \geq 1}$ is bounded. We will show that $\theta_k \to \pm\infty$ leads to $L(\theta_k; F) \to -\infty$, which contradicts the fact that $L(\theta_k; F) \to L(F) \in \mathbb{R}$, thus completing the proof.

Consider $\theta_k \to \pm\infty$. We have already mentioned that by Proposition 4(iii) of Xu and Samworth (2017), for each θ_k, there exists a log-density $\psi_{\theta_k} \in \mathcal{SC}_0$ such that

$$L(\theta_k; F) = \Psi(\theta, \psi_{\theta_k}, F),$$

which equals

$$\int_{-\infty}^{\infty} \psi_{\theta_k}(x - \theta) dF(x) - 1 \leq -\int_{-\infty}^{\infty} \log\Big(2|x - \theta_k|\Big) dF(x) - 1,$$

where the last inequality follows from Lemma 2.10. Now if $\theta_k \to \pm\infty$, using Fatou's Lemma, we derive that

$$\limsup_{k\to\infty} L(\theta_k; F) \leq -\int_{-\infty}^{\infty} \liminf_{k\to\infty} \Big(\log|x-\theta_k| \Big) dF(x) - \log 2 = -\infty,$$

which leads to the desired contradiction. □

Proof of Lemma 2.1

The proof follows from (2.4) and Theorem 2.10 noting that $\mathbb{F}_n$ satisfies Condition A. □

Proof of Lemma 2.6

First we will show that if F satisfies (2.24) and $\psi \in \mathcal{SC}_0$, there exists $\psi_1 \in \mathcal{SC}_0$ such that

$$\Psi(\theta', \psi, F) \leq \Psi(\theta, \psi_1, F), \tag{2.77}$$

leading to

$$L(\theta'; F) \leq L(\theta; F),$$

which implies $\theta^*(F) = \theta$. Note that $G = F(\cdot + \theta)$ satisfies

$$G(x) + G(-x) = 1. \tag{2.78}$$

Using the symmetry of ψ about 0 and (2.78) in the fourth step we deduce that

$$\begin{aligned}
&\Psi(\theta', \psi, F) \\
&= \int_{-\infty}^{\infty} \psi(x-\theta')dG(x-\theta) - \int_{-\infty}^{\infty} e^{\psi(x)}dx \\
&= \int_{-\infty}^{0} \psi(y+\theta-\theta')dG(y) + \int_{0}^{\infty} \psi(y+\theta-\theta')dG(y) - \int_{-\infty}^{\infty} e^{\psi(x)}dx \\
&= \int_{-\infty}^{0} \psi(y+\theta-\theta')dG(y) - \int_{-\infty}^{0} \psi(-y+\theta-\theta')dG(-y) - \int_{-\infty}^{\infty} e^{\psi(x)}dx \\
&= \int_{-\infty}^{0} \psi(y+\theta-\theta')dG(y) + \int_{-\infty}^{0} \psi(y-\theta+\theta')dG(y) - \int_{-\infty}^{\infty} e^{\psi(x)}dx \\
&= 2\int_{-\infty}^{0} \psi_1(y)dG(y) - \int_{-\infty}^{\infty} e^{\psi(x)}dx,
\end{aligned}$$

where

$$\psi_1(x) \;=\; 2^{-1}\Big(\psi(x+\theta-\theta') + \psi(x-\theta+\theta')\Big), \quad x \in \mathbb{R}.$$

Since $\psi_1 \in \mathcal{SC}_0$, we obtain that

$$\begin{aligned}
\Psi(\theta', \psi, F) \;&=\; \int_{-\infty}^{\infty} \psi_1(x)dG(x) - \int_{-\infty}^{\infty} e^{\psi(x)}dx \\
&= \int_{-\infty}^{\infty} \psi_1(x-\theta)dF(x) - \int_{-\infty}^{\infty} e^{\psi(x)}dx \\
&= \Psi(\theta, \psi_1, F) + \int_{-\infty}^{\infty} e^{\psi_1(x)}dx - \int_{-\infty}^{\infty} e^{\psi(x)}dx.
\end{aligned}$$

Also by the arithmetic mean-geometric mean inequality,

$$\int_{-\infty}^{\infty} e^{\psi_1(x)}dx \le \int_{-\infty}^{\infty} 2^{-1}\Big(e^{\psi(x+\theta-\theta')} + e^{\psi(x-\theta+\theta')}\Big)dx = \int_{-\infty}^{\infty} e^{\psi(x)}dx,$$

where equality holds if and only if $\psi(x+\theta-\theta') = \psi(x-\theta+\theta')$ almost everywhere with respect to the Lebesgue measure, which only takes place when $\theta' = \theta$. Therefore we have established the inequality in (2.77) with equality holding if and only if $\theta' = \theta$, which proves $\theta^*(F) = \theta$. Observe that

$$\Psi(\theta, \psi, F) = \int_{-\infty}^{\infty} \psi(x)dG(x) - \int_{-\infty}^{\infty} e^{\psi(x)}dx = \Psi(0, \psi, G).$$

Finally, Lemma 2.12 combined with (2.78) and (2.43) entail that

$$\arg\max_{\psi\in\mathcal{SC}_0}\Psi(0,\psi,G)=\arg\max_{\phi\in\mathcal{SC}_0}\omega(\phi,G)=\arg\max_{\phi\in\mathcal{C}}\omega(\phi,G),$$

which completes the proof of Lemma 2.6. □

Proof of Lemma 2.5

Lemma 2.5 follows directly from Lemma 2.2. □

Proof of Lemma 2.7

Suppose $\overline{J(F)}=\mathbb{R}$. Then $\theta^*(F)\in\overline{J(F)}$ by Proposition 1. Otherwise, since F is non-decreasing, $\overline{J(F)}$ takes one of the following three forms: $[a,b]$, $[a,\infty)$, or $(-\infty,b]$, where $a,b\in\mathbb{R}$. Suppose $\overline{J(F)}=[a,b]$. We will show that in this case $L(\theta;F)$ defined in (2.73) is non-decreasing in θ on $(-\infty,a]$, and non-increasing in θ on $[b,\infty)$. Since $\theta^*(F)$ equals $\arg\max_{\theta\in\mathbb{R}}L(\theta;F)$, the above implies $\theta^*(F)\in[a,b]$. To show that $L(\theta;F)$ is non-decreasing in θ on $(-\infty,a]$, we first note that for $\theta<\theta'\le a$, and $\psi\in\mathcal{SC}_0$,

$$\int_a^b\psi(x-\theta)dF(x)\le\int_a^b\psi(x-\theta')dF(x),$$

since ψ is non-increasing on $[0,\infty)$, and

$$0\le x-\theta'<x-\theta$$

for $x\ge a$. Therefore from (2.73), it is not hard to see that $L(\theta;F)\le L(\theta';F)$. Similarly we can show that for $\theta>\theta'\ge b$,

$$\int_a^b\psi(x-\theta)dF(x)\le\int_a^b\psi(x-\theta')dF(x),$$

since ψ is non-decreasing on $(-\infty,0]$, and $x-\theta<x-\theta'\le 0$ for $x\le b$. Therefore $L(\theta;F)\le L(\theta';F)$, which completes the proof of $\theta^*(F)\in[a,b]$. When $\overline{J(F)}$ is of the form $[a,\infty)$ or $(-\infty,b]$, we can prove $\theta^*(F)\in\overline{J(F)}$ in a similar way. Hence the first part of Lemma 2.7 is proved. The second part of Lemma 2.7 follows from Lemma 2.11. □

Proof of Theorem 2.1

First part follows from (2.4) and Proposition 1 noting that $\mathbb{F}_n$ satisfies Condition A. The second part follows from Lemma 2.7. □

Proof of Theorem 2.7

Since $d_W(F_n, F) \to 0$, from Theorem 6.9 of Villani (2009) or Theorem 7.12 of Villani (2003) it follows that F_n converges to F weakly, and

$$\int_{-\infty}^{\infty} |x| dF_n(x) \to \int_{-\infty}^{\infty} |x| dF(x) < \infty. \tag{2.79}$$

Let us denote the measure corresponding to F_n by P_n.

For the sake of simplicity, we use the abbreviated notations ψ_n^*, ϕ_n^*, and θ_n^* in place of $\psi^*(F_n)$, $\phi^*(F_n)$, and $\theta^*(F_n)$ respectively.

Our proof is motivated by the proof of Proposition 6 of Xu and Samworth (2017). Since F_n satisfies condition A, $L(F_n) < \infty$ by Proposition 1. Therefore starting with any subsequence of $\{F_n\}_{n\geq 1}$, we can extract a further subsequence $\{F_{n_k}\}_{k\geq 1}$, such that $L(F_{n_k}) \to \vartheta \in [-\infty, \infty]$, and $\theta_{n_k}^* \to \theta' \in [-\infty, \infty]$.

The proof of Theorem 2.7 involves the following steps:

(A) Show that $\vartheta > -\infty$.

(B) Show that $\vartheta < \infty$, and there exists $M > \vartheta + 1$ such that

$$\limsup_{k\uparrow\infty} \sup_{x\in\mathbb{R}} \psi_{n_k}^*(x) < M.$$

(C) Show that $\theta' \in \mathbb{R}$.

(D) Taking a and b to be the left and right endpoints of $J(F)$ respectively, from Lemma 2.7, we observe that

$$\text{int}(\text{dom}(\psi^*(F))) = (-d, d), \tag{2.80}$$

where $d = \max(b - \theta^*(F), \theta^*(F) - a)$. We now define

$$D_{\theta'} = \begin{cases} \mathbb{R}, & \text{if } \ a = -\infty, b = \infty, \\ \text{dom}(\psi^*(F)) + \theta^*(F) - \theta', & \text{if } b - \theta^*(F) > \theta^*(F) - a, \\ \text{dom}(\psi^*(F)) - (\theta^*(F) - \theta'), & \text{if } b - \theta^*(F) \leq \theta^*(F) - a. \end{cases} \tag{2.81}$$

We Show that $\liminf_k \psi^*_{n_k}(x) > -\infty$ for $x \in \text{int}(D_{\theta'})$, leading to $\liminf_k \psi^*_{n_k}(0) > -\infty$.

(E) Show that there exist α', $\beta' > 0$ such that for all sufficiently large k,

$$\psi^*_{n_k}(x) \leq -\alpha'|x| + \beta', \quad \text{for all } x \in \mathbb{R}.$$

(F) Show that $\vartheta = L(F)$, $\theta' = \theta$, and $\psi^*_{n_k}$ converges pointwise to $\psi^*(F)$ on $\text{int}(\text{dom}(\psi^*(F))$.

The proof of (A) is quite straightforward. To see this, observe that the function $\psi(x) = -|x|$ is a member of $\mathcal{SC}_0$. Therefore

$$\vartheta = \lim_{k\to\infty} L(F_{n_k}) \geq \lim_{k\to\infty} \Psi(0, \psi, F_{n_k}) \geq -\lim_{k\to\infty} \int_{-\infty}^{\infty} |x| dF_{n_k}(x) - \int_{-\infty}^{\infty} e^{-|x|} dx,$$

which equals $-\int |x| dF(x) - 2$ by (2.79). Since $\int |x| dF(x) < \infty$ by Condition A, step (A) follows.

The proof of (B) hinges on the proof of Theorem 2.2 of Dümbgen *et al.* (2011). Suppose $\limsup_k \sup_{x\in\mathbb{R}} \psi^*_{n_k}(x) = \infty$. In that case we can replace $\{n_k\}$ by a further subsequence such that $\lim_k \sup_{x\in\mathbb{R}} \psi^*_{n_k}(x) = \infty$. Hence, without loss of generality, we assume that $\lim_k \sup_{x\in\mathbb{R}} \psi^*_{n_k}(x) = \infty$. Also note that $\sup_x \psi^*_{n_k}(x) = \psi^*_{n_k}(0) = M_k$, say.

Letting $D_{k,t}$ be the level set $\{\psi^*_{n_k} \geq t\}$, for any $c > 0$, we can bound $L(F_{n_k})$ in the

following way:

$$\begin{aligned}
L(F_{n_k}) = & \int_{-\infty}^{\infty} \psi^*_{n_k}(x-\theta^*_{n_k}) dF_{n_k}(x) - 1 \\
= & \int_{-\infty}^{\infty} \psi^*_{n_k}(x) dF_{n_k}(x+\theta^*_{n_k}) - 1 \\
\leq & - cM_k P_{n_k}\left(\mathbb{R} \setminus D_{k,-cM_k} + \theta^*_{n_k}\right) + M_k P_{n_k}(D_{k,-cM_k} + \theta^*_{n_k}) - 1 \\
= & - cM_k\left(1 - P_{n_k}(D_{k,-cM_k} + \theta^*_{n_k})\right) + M_k P_{n_k}(D_{k,-cM_k} + \theta^*_{n_k}) - 1 \\
= & - (c+1)M_k\left(\frac{c}{c+1} - P_{n_k}(D_{k,-cM_k} + \theta^*_{n_k})\right) - 1. \qquad (2.82)
\end{aligned}$$

Let us denote the Lebesue measure on $\mathbb{R}$ by Leb. We intend to show that $\text{Leb}(D_{k,-cM_k}) \to 0$ as $k \to \infty$. Applying Lemma 4.1 of Dümbgen *et al.* (2011), we obtain that as $M_k \to \infty$,

$$\begin{aligned}
\text{Leb}(D_{k,-cM_k}) &\leq (1+c)M_k e^{-M_k} / \int_0^{(1+c)M_k} t e^{-t} dt \\
&= (1+c)M_k e^{-M_k}/(1+o(1)) \to 0. \qquad (2.83)
\end{aligned}$$

Now we will show that $P_{n_k}(D_{k,-cM_k} + \theta^*_{n_k}) \to 0$. We consider two cases, $|\theta'| < \infty$ and $\theta' = \pm\infty$. Let us focus on the case $|\theta'| < \infty$ first. Now the concavity of $\psi^*_{n_k}$ indicates that sets of the form $D_{k,-cM_k}$ are intervals including 0. Since $\text{Leb}(D_{k,-cM_k}) \to 0$ and $\theta^*_{n_k} \to \theta'$ as $k \to \infty$, for sufficiently large k, the set $D_{k,-cM_k} + \theta^*_{n_k}$ is an interval around θ' whose length shrinks to 0 as k increases. Now since F is non-degenerate, one can certainly find a $x_0 \neq \theta'$ such that $x_0 \in \text{int}(J(F))$. As a result, $x_0 \notin D_{k,-cM_k} + \theta^*_{n_k}$ for all sufficiently large k. Thus for all sufficiently large k,

$$D_{k,-cM_k} + \theta^*_{n_k} \in \{C \subset \mathbb{R} \text{ closed and convex}, \ \ x_0 \notin \text{int}(C)\} = D_{x_0},$$

say. However, since F_{n_k} weakly converges to F, Lemma 2.13 of Dümbgen *et al.* (2011) implies that

$$\limsup_k \sup_{C \in D_{x_0}} P_{n_k}(C) \leq \sup_{C \in D_{x_0}} P_0(C). \qquad (2.84)$$

By Lemma 2.13 of Dümbgen *et al.* (2011), $\sup_{C \in D_{x_0}} P(C)$ is bounded away from 1 if

$$x_0 \in \bigcap_{C \in \mathcal{D}} C,$$

where

$$\mathcal{D} = \left\{ C \;\middle|\; C \text{ is closed and convex, } P(C) = 1 \right\}.$$

Since $x_0 \in \text{int}(J(F))$, the above holds. Therefore we conclude that the term in the left hand side of (2.84) is less than $c/(c+1)$ for sufficiently large c. Since $D_{k,-cM_k} + \theta^*_{n_k} \in D_{x_0}$, the above implies that $P_{n_k}(D_{k,-cM_k} + \theta^*_{n_k}) < c/(c+1)$ for sufficiently large k. Combined with (2.82), this result leads to

$$L(F_{n_k}) \to -\infty \quad \text{as } k \to \infty,$$

which contradicts the fact that $\vartheta > -\infty$. Therefore for this case, we see that $\lim_k M_k = \infty$ leads to a contradiction.

At this point it is clear that for the other case also, i.e. when $\theta' = \pm\infty$, it is enough to show that $\limsup_k P_{n_k}(D_{k,-cM_k} + \theta^*_{n_k}) < c/(c+1)$ for sufficiently large c. We will consider the case $\theta' = \infty$. The proof for $\theta' = -\infty$ will follow similarly. Consider y such that $1 - F(y) < c/(c+1)$. Since $\text{Leb}(D_{k,-cM_k}) \to 0$, we observe that as $\theta^*_{n_k} \to \infty$, for all sufficiently large k,

$$D_{k,-cM_k} + \theta^*_{n_k} \subset (y, \infty).$$

Noting $F_{n_k} \to_w F$, we obtain that

$$\limsup_k P_{n_k}(D_{k,-cM_k} + \theta^*_{n_k}) \le \limsup_k (1 - F_{n_k}(y)) = 1 - F(y) < c/(c+1),$$

which completes the proof of (B). Therefore we assume that $\{M_k\}_{k \ge 1}$ is bounded above by a number, say M, which can be chosen to be greater than $\vartheta + 1$.

Our next step is step (C), which follows from Lemma 2.10. To see this, observe that Lemma 2.10 implies

$$\phi^*_{n_k}(x) \le -\log|2(x - \theta^*_{n_k})|.$$

Suppose $\limsup_k |\theta^*_{n_k}| = \infty$. Then Fatou's Lemma implies that

$$\begin{aligned}\liminf_k \int_{-\infty}^{\infty} e^{\phi^*_{n_k}(x)} dx \leq & \liminf_k \int_{-\infty}^{\infty} e^{-\log|2(x-\theta^*_{n_k})|} dx \\ \leq & \int_{-\infty}^{\infty} \liminf_k e^{-\log|2(x-\theta^*_{n_k})|} dx = 0.\end{aligned}$$

However, the left hand side of the above display equals 1 since $e^{\phi^*_{n_k}}$ is a density, which leads to a contradiction. Therefore step (C) follows.

Next we will prove (D). For $x \in D_{\theta'}$, denote $A_{n_k} = (-x + \theta^*_{n_k}, x + \theta^*_{n_k})$. We calculate

$$\begin{aligned}L(F_{n_k}) = & \int_{-\infty}^{\infty} \psi^*_{n_k}(z) dF_{n_k}(z + \theta^*_{n_k}) - 1 \\ = & \int_{-x}^{x} \psi^*_{n_k}(z) dF_{n_k}(z + \theta^*_{n_k}) + \int_{-\infty}^{-x} \psi^*_{n_k}(z) dF_{n_k}(z + \theta^*_{n_k}) \\ & + \int_{x}^{\infty} \psi^*_{n_k}(z) dF_{n_k}(z + \theta^*_{n_k}) - 1 \\ \leq & \psi^*_{n_k}(0) P_{n_k}(A_{n_k}) + \psi^*_{n_k}(x) P_{n_k}(\mathbb{R} \setminus A_{n_k}) - 1,\end{aligned}$$

which follows since $\psi^*_{n_k} \in \mathcal{SC}_0$. Now observe that the above implies

$$\begin{aligned}\liminf_{k\to\infty} \psi^*_{n_k}(x) \geq & \liminf_{k\to\infty} \frac{L(F_{n_k}) + 1 - \psi^*_{n_k}(0) P_{n_k}(A_{n_k})}{P_{n_k}(\mathbb{R} \setminus A_{n_k})} \\ \geq & -\frac{M - 1 - \vartheta}{\liminf_{k\to\infty} P_{n_k}(\mathbb{R} \setminus A_{n_k})}. \end{aligned} \tag{2.85}$$

Note that for every $\epsilon > 0$,

$$\theta' - \epsilon < \theta^*_{n_k} < \theta' + \epsilon, \quad \text{for all sufficiently large } k.$$

Therefore we can write

$$\begin{aligned}& \liminf_{k\to\infty} P_{n_k}(\mathbb{R} \setminus A_{n_k}) \\ & = \liminf_{k\to\infty} F_{n_k}(-x + \theta^*_{n_k}) + 1 - \limsup_{k\to\infty} F_{n_k}\Big((x + \theta^*_{n_k}) - \Big) \\ & \geq \liminf_{k\to\infty} F_{n_k}(-x + \theta' - \epsilon) + 1 - \limsup_{k\to\infty} F_{n_k}\Big((x + \theta' + \epsilon) - \Big) \\ & \geq F\Big((-x + \theta' - \epsilon) - \Big) + 1 - F\Big((x + \theta' + \epsilon) - \Big).\end{aligned}$$

If $J(F) = \mathbb{R}$, it is clear that the right hand side of the last display is positive. Suppose $J(F) \neq \mathbb{R}$. Then at least one of a and b, the endpoints of $J(F)$, is finite. Now suppose $b - \theta^*(F) > \theta^*(F) - a$. Noting $x \in \text{int}(D_{\theta'})$, we argue that for small enough $\epsilon > 0$, $x + \epsilon \in \text{int}(D_{\theta'})$. Then from (2.80) and (2.81) it follows that

$$x + \epsilon < b - \theta^*(F) + \theta^*(F) - \theta',$$

implying that $x + \epsilon + \theta' < b$, leading to $F((x + \theta' + \epsilon)-) < 1$ since b is the right endpoint of $J(F)$. Now consider the case when $b - \theta^*(F) \leq \theta^*(F) - a$. Analogous to the previous case, (2.80) and (2.81) yield

$$x + \epsilon < \theta^*(F) - a - \theta^*(F) + \theta',$$

or equivalently,

$$a < -x + \theta' - \epsilon,$$

implying $F((-x + \theta' - \epsilon)-) > 0$, which leads to

$$\limsup_{k\to\infty} \mathbb{P}_n(\mathbb{R} \setminus A_{n_k}) > 0.$$

Since in step (B), M was chosen such that $M > \vartheta + 1$, (2.85) entails that

$$\liminf_{k\to\infty} \widehat{\psi}_{\theta_{r_k}}(x) > -\infty$$

which proves (D).

To prove (E), we first observe that since $\psi^*_{n_k} \in \mathcal{SC}_0$, Lemma 3 of Pal *et al.* (2007) implies that

$$\psi^*_{n_k}(x) \leq \psi^*_{n_k}(0) + 1 - e^{\psi^*_{n_k}(0)}|x|, \quad x \in \mathbb{R}.$$

Observe that (B) and (D), combined with the above, imply that there exist $\alpha' > 0$ and $\beta' \in \mathbb{R}$ such that for all sufficiently large k,

$$\psi^*_{n_k}(x) \leq \beta' - \alpha'|x| \quad \text{for all } x \in \mathbb{R}. \tag{2.86}$$

Now we move on to step (F). Since we have established (2.86) and showed that $\liminf_{k\to\infty} \widehat{\psi}_{\theta_{r_k}}(x) > -\infty$ for $x \in \text{int}(D_{\theta'})$, by Lemma 4.2 of Dümbgen *et al.* (2011), we can replace $\{\psi^*_{n_k}\}_{k\geq 1}$, if necessary, by a subsequence such that for a concave function $\bar{\psi}$ the following conditions are met:

$$\begin{aligned} &\text{int}(D_{\theta'}) \subset \text{dom}(\bar{\psi}), \\ &\bar{\psi}(x) \leq \beta' - \alpha'|x| \qquad \text{for all } x \in \mathbb{R}, k \in \mathbb{N}, \\ &\lim_{k\to\infty, y\to x} \psi^*_{n_k}(y) = \bar{\psi}(x) \qquad \text{for all } x \in \text{int}(\text{dom}(\bar{\psi})), \\ &\limsup_{k\to\infty, y\to x} \psi^*_{n_k}(y) \leq \bar{\psi}(x) \qquad \text{for all } x \in \mathbb{R}. \end{aligned} \tag{2.87}$$

We will now establish that $\bar{\psi} \in \mathcal{SC}_0$. To see this, first notice that since $\psi^*_{n_k} \in \mathcal{SC}_0$, for $x \notin \text{dom}(\bar{\psi})$,

$$\lim_{k\to\infty} \psi^*_{n_k}(-x) = \lim_{k\to\infty} \psi^*_{n_k}(x) \leq \bar{\psi}(x) = -\infty.$$

Hence, $-x \notin \text{int}(\text{dom}(\bar{\psi}))$, suggesting that either $-x \notin \text{dom}(\bar{\psi})$, or $-x \in \text{dom}(\bar{\psi})$ is a boundary point of $\text{dom}(\bar{\psi})$. Therefore $\text{dom}(\bar{\psi})$ is of the form $(-c, c)$, $[-c, c]$, $(-c, c]$, or $[-c, c)$ for some $c > 0$. For the last two cases, if we extend $\bar{\psi}$ to $[-c, c]$ so that $\bar{\psi}$ is continuous on $[-c, c]$ and $-\infty$ everywhere else, $\bar{\psi}$ still satisfies (2.87). Therefore without loss of generality, we assume that $\text{dom}(\bar{\psi})$ is of the form $(-c, c)$ or $[-c, c]$. For $y \in (-c, c)$ we obtain that

$$\bar{\psi}(y) = \lim_{k\to\infty} \psi^*_{n_k}(y) = \lim_{k\to\infty} \psi^*_{n_k}(-y) = \bar{\psi}(-y).$$

It remains to show that when $\text{dom}(\bar{\psi})$ is the closed interval $[-c, c]$, $\bar{\psi}(-c)$ and $\bar{\psi}(c)$ agree. Since ψ^* is concave, it is right continuous at $-c$, and left continuous at c. Therefore the symmetry of ψ^* on $(-c, c)$ implies symmetry on $[-c, c]$, which establishes that $\bar{\psi} \in \mathcal{SC}_0$. Now we claim that $e^{\bar{\psi}} \in \mathcal{SLC}_0$. To see this observe that since $\psi^*_{n_k}(x) \leq \beta' - \alpha'|x|$, application of the dominated convergence theorem yields

$$\int_{\text{int}(\text{dom}(\bar{\psi}))} e^{\bar{\psi}(x)} dx = \int_{\text{int}(\text{dom}(\bar{\psi}))} e^{\lim_k \psi^*_{n_k}(x)} dx = \lim_k \int_{\text{dom}(\bar{\psi})} e^{\psi^*_{n_k}(x)} dx = 1.$$

Since Condition B holds, note that if we can show

$$\vartheta = \Psi(\theta', \bar{\psi}, F) = L(F),$$

we have $\theta' = \theta^*(F)$, $\bar{\psi} = \psi^*(F)$, and (2.41) also follows. Therefore we have $\theta_n^* \to \theta^*(F)$, and (2.87) leads to the pointwise convergence of ψ_n^* to $\psi^*(F)$. The pointwise convergence of ψ_n^*'s implies pointwise convergence of $e^{\psi_n^*}$, which, in its turn leads to the weak convergence of the corresponding distribution functions. Then the rest of the proof of Theorem 2.10 follows from Proposition 2 of Cule and Samworth (2010). We prove the equations in the last display by first establishing that

$$\vartheta \leq \Psi(\theta', \bar{\psi}, F) \leq L(F), \tag{2.88}$$

and then showing $\vartheta \geq L(F)$.

Observe that since F_{n_k} converges to F weakly, by Skorohod's theorem, there exists a probability space $(\Omega, \mathcal{A}, P^*)$ with random variable $X_k \sim F_{n_k}$, and $X \sim F$ such that $X_k \to_{a.s.} X$ as $k \to \infty$. Denote by E^* the expectation operator with respect to the probability space $(\Omega, \mathcal{A}, P^*)$. Denoting the positive random variable $\beta' - \alpha'|X_k| - \phi_{n_k}^*(X_k)$ by H_k, we calculate

$$\begin{aligned}
\vartheta &= \limsup_{k\to\infty} \int_{-\infty}^{\infty} \phi_{n_k}^*(x) dF_{n_k}(x) - 1 \\
&= \lim_{k\to\infty} \int_{-\infty}^{\infty} (\beta' - \alpha'|x|) dF_{n_k}(x) - \liminf_{k\to\infty} E^*[H_k] - 1 \\
&\leq \beta' - \alpha' \int_{-\infty}^{\infty} |x| dF(x) - E^*\left[\liminf_{k\to\infty} H_k\right] - 1,
\end{aligned}$$

which follows using (2.79) and Fatou's Lemma. As $\theta_{n_k}^* \to \theta' \in \mathbb{R}$, and $X_k \to_{a.s.} X$, (2.87) indicates that the following holds with probability 1:

$$\liminf_{k\to\infty} H_k = \alpha' - \beta'|X| - \limsup_k \psi_{n_k}^*(X_k - \theta_{n_k}^*) \geq \beta' - \alpha'|X| - \bar{\psi}(X - \theta').$$

Since $X \sim F$, the above leads to

$$\begin{aligned}\vartheta &\leq \beta' - \alpha' \int_{-\infty}^{\infty} |x| dF(x) - E^*\left[\beta' - \alpha'|X| - \bar{\psi}(X - \theta')\right] - 1 \\ &= \int_{-\infty}^{\infty} \bar{\psi}(x - \theta') dF(x) - 1 \\ &= \Psi(\theta', \bar{\psi}, F) \leq L(F),\end{aligned}$$

which completes the proof of (2.88). Hence it only remains to prove that $\vartheta \geq L(F)$.

Let us denote $c_0 = \psi^*(F)(0)$. From Lemma 2.13 it follows that for each $\epsilon > 0$, there exist $\psi^{(\epsilon)} \in \mathcal{SC}_0$ such that

$$-|x|/\epsilon + c_0 \leq \psi^{(\epsilon)}(x) \quad \text{for all } x \in \mathbb{R}, \tag{2.89}$$

and

$$\psi^*(F) \leq \psi^{(\epsilon)} \leq c_0. \tag{2.90}$$

Further, $\psi^{(\epsilon)}$ decreases to $\psi^*(F)$ pointwise as $\epsilon \downarrow 0$. Using the random variables $X_k \sim F_{n_k}$ and $X \sim F$ constructed in the proof of (2.88), we write

$$\begin{aligned}\vartheta = \lim_{k\to\infty} L(F_{n_k}) &\geq \lim_{k\to\infty} \Psi(\theta^*(F), \psi^{(\epsilon)}, F_{n_k}) \\ &= \lim_{k\to\infty} E^*\left[\psi^{(\epsilon)}(X_k - \theta^*(F))\right] - \int_{-\infty}^{\infty} e^{\psi^{(\epsilon)}(x)} dx.\end{aligned} \tag{2.91}$$

Note that (2.89) and (2.90) indicate that $\psi^{(\epsilon)}(x)$ is bounded above and below by c_0 and $-|x|/\epsilon + c_0$ respectively. From (2.79) it follows that

$$E^*|X_k| = \int_{-\infty}^{\infty} |x| dF_{n_k}(x) \to \int_{-\infty}^{\infty} |x| dF(x), \quad \text{as } k \to \infty.$$

Moreover, as $k \to \infty$, we also have $X_k \to_{a.s.} X$. Observe that $\psi^{(\epsilon)}$ is concave, and by (2.89), $\text{dom}(\psi^{(\epsilon)}) = \mathbb{R}$, which implies that $\psi^{(\epsilon)}$ is continuous. Therefore

$$\psi^{(\epsilon)}(X_k - \theta^*(F)) \to_{a.s.} \psi^{(\epsilon)}(X - \theta^*(F)).$$

Therefore using Lemma 2.14, from (2.91) we obtain that

$$\begin{aligned}\vartheta &\geq E\left[\lim_{k\to\infty}\psi^{(\epsilon)}(X_k-\theta^*(F))\right]-\int_{-\infty}^{\infty}e^{\psi^{(\epsilon)}(x)}dx\\ &=\int_{-\infty}^{\infty}\psi^{(\epsilon)}(x-\theta^*(F))dF(x)-\int_{-\infty}^{\infty}e^{\psi^{(\epsilon)}(x)}dx.\end{aligned}$$

Since $\psi^{(\epsilon)}$ decreases to $\psi^*(F)$ pointwise as $\epsilon\downarrow 0$, the monotone convergence theorem and (2.90) lead to

$$\lim_{\epsilon\downarrow 0}\int_{-\infty}^{\infty}\Big(c_0-\psi^{(\epsilon)}(x-\theta^*(F))\Big)dF(x)=\int_{-\infty}^{\infty}\Big(c_0-\psi^*(F)(x-\theta^*(F))\Big)dF(x),$$

which can be rewritten as

$$\lim_{\epsilon\downarrow 0}\int_{-\infty}^{\infty}\psi^{(\epsilon)}(x-\theta^*(F))dF(x)=\int_{-\infty}^{\infty}\psi^*(F)(x-\theta^*(F))dF(x). \tag{2.92}$$

On the other hand, note that (2.90) implies that $g^*(F)\leq e^{\psi^{(\epsilon)}}\leq g^*(F)(0)$. Since $e^{\psi^{(\epsilon)}}$ decreases to $g^*(F)$ pointwise as ϵ decreases to 0, the monotone convergence theorem implies that

$$\lim_{\epsilon\downarrow 0}\int_{-\infty}^{\infty}\Big(g^*(F)(0)-e^{\psi^{(\epsilon)}(x)}\Big)dx=\int_{-\infty}^{\infty}\Big(g^*(F)(0)-g^*(F)(x)\Big)dx,$$

or equivalently,

$$\lim_{\epsilon\downarrow 0}\int_{-\infty}^{\infty}e^{\psi^{(\epsilon)}(x)}dx=\int_{-\infty}^{\infty}g^*(F)(x)dx=1. \tag{2.93}$$

Therefore from (2.92) and (2.93), we conclude that

$$\begin{aligned}\vartheta &\geq \lim_{\epsilon\downarrow 0}\int_{-\infty}^{\infty}\psi^{(\epsilon)}(x-\theta^*(F))\;dF(x)-\lim_{\epsilon\downarrow 0}\int_{-\infty}^{\infty}e^{\psi^{(\epsilon)}(x)}dx\\ &=\int_{-\infty}^{\infty}\psi^*(F)(x-\theta^*(F))\;dF(x)-1=L(F),\end{aligned}$$

which completes the proof. □

Proof of Theorem 2.9

In their proof of Theorem 3.1, Pal *et al.* (2007) show that if a sequence of log-concave functions $\{f_n\}_{n\geq 1}$ (which can be stochastic as well) satisfies

$$\sum_{i=1}^{n} \log f_n(X_i) \geq \sum_{i=1}^{n} \log f_0(X_i) \quad a.s., \tag{2.94}$$

we have $H(f_n, f_0) \to_{a.s.} 0$, provided

$$P\left(\sup_x \log f_n(x) = o\left(\frac{\sqrt{n}}{\log n} \right) \right) = 1.$$

If we take $f_n = \widehat{f}_n$, we have

$$\sup_x \log \widehat{f}_n(x) = \widehat{\phi}_{0,n}(\widehat{\theta}_n) = \widehat{\psi}_{0,n}(0).$$

Theorem 2.8(c) and Corollary 4 entail that

$$P\left(\limsup_n \widehat{\psi}_{0,n}(0) < \infty \right) = 1.$$

Also note that being the MLE of f_0, $\widehat{f}_n$ automatically satisfies (2.94), which implies

$$H(\widehat{f}_n, f_0) \to_{a.s.} 0. \tag{2.95}$$

Since $f_0 \in \mathcal{P}_0$ is log-concave and continuous, Proposition 2(c) of Cule and Samworth (2010) and (2.95) entail that

$$\sup_{z\in\mathbb{R}} |\widehat{f}_n(z) - f_0(z)| \to_{a.s.} 0. \tag{2.96}$$

Note that

$$\begin{aligned} 2H^2(\widehat{g}_n, g_0) &= \int_{-\infty}^{\infty} \left(\sqrt{\widehat{g}_n(z-\widehat{\theta}_n)} - \sqrt{g_0(z-\widehat{\theta}_n)} \right)^2 dz \\ &\leq 2\int_{-\infty}^{\infty} \left(\sqrt{\widehat{g}_n(z-\widehat{\theta}_n)} - \sqrt{g_0(z-\theta_0)} \right)^2 dz \\ &\quad + 2\int_{-\infty}^{\infty} \left(\sqrt{g_0(z-\widehat{\theta}_n)} - \sqrt{g_0(z-\theta_0)} \right)^2 dz, \end{aligned}$$

where the first term on the right hand side of the last display approaches 0 almost surely by (2.95). The second term is also bounded above by a constant multiple of $g_0(z-\widehat{\theta}_n)+g_0(z-\theta_0)$, which is integrable. Therefore using Lemma 2.14, Theorem 2.8(a), and Corollary 4, we deduce that this term also converge to 0 almost surely. Hence $H^2(\widehat{g}_n, g_0) \to_{a.s.} 0$ follows.

Our next step is to find the rate of convergence of $H(\widehat{f}_n, f_0)$. To do so, we first introduce the class of functions

$$\mathcal{P}_{M,0} = \Big\{ f \in \mathcal{LC} \;\Big|\; \sup_{x\in\mathbb{R}} f(x) < M,\ f(x) > 1/M \text{ for all } |x| < 1, \\ \text{Supp}(f) \subset \text{Supp}(f_0) \Big\}.$$

We will show that without loss of generality, one can assume that $f_0 \in \mathcal{P}_{M,0}$ for some $M > 0$. To this end, we translate and rescale the data letting $\tilde{X}_i = \alpha X_i + \beta$, where $\alpha > 0$ and $\beta \in \mathbb{R}$. Observe that the rescaled data has density $\tilde{f}_0(x) = \alpha^{-1} f_0((x-\beta)/\alpha)$. Denote by $\tilde{f}_{0,n}$ the MLE of f_0 based on the rescaled data. Note that the MLE is affine-equivalent, which entails that $\tilde{f}_{0,n}(x) = \alpha^{-1}\widehat{f}_n((x-\beta)/\alpha)$. Noting Hellinger distance is invariant under affine transformations, we observe that $H(\widehat{f}_n, f_0) = H(\tilde{f}_{0,n}, \tilde{f}_0)$. Therefore it suffices to show that $H(\tilde{f}_{0,n}, \tilde{f}_0) \to_{a.s.} 0$. Note that since f_0 is log-concave, $\text{int}(\text{dom}(f_0))$ contains an interval. We can choose α and β in a way such that $(x-\beta)/\alpha$ lie inside that interval for $x = \pm 1$. Then it is possible to find $M > 0$ large enough such that

$$f_0((x-\beta)/\alpha) > \alpha/M, \quad x = \pm 1,$$

or

$$\min(\tilde{f}_0(-1), \tilde{f}_0(1)) > 1/M,$$

leading to

$$\inf_{x\in[-1,1]} \tilde{f}_0(x) > 1/M,$$

since f_0, or equivalently $\tilde{f}_0$ is unimodal. Hence without loss of generality, we can assume that there exists $M > 0$ such that $f_0(x) > 1/M$ for $x \in [-1, 1]$. We can choose M large enough

such that additionally, $\sup_{x\in\mathbb{R}} f_0(x) < M$. On the other hand, (2.96) implies that the following inequalities hold with probability one:

$$\limsup_n \sup_{x\in\mathbb{R}} \widehat{f}_n(x) < M,$$

and

$$\lim_n \widehat{f}_n(\pm 1) > 1/M.$$

Therefore $f_0 \in \mathcal{P}_{M,0}$, and with probability one, $\widehat{f}_n \in \mathcal{P}_{M,0}$ as well for all sufficiently large n. Doss and Wellner (2016) obtained the bracketing entropy of the class $\mathcal{P}_{M,0}$. They showed that for any $\epsilon > 0$,

$$\log N_{[\,]}(\epsilon, \mathcal{P}_{M,0}, H) \lesssim \epsilon^{-1/2}.$$

The rest of the proof now follows from an application of Theorem 3.4.1 and 3.4.4 of Van der Vaart and Wellner (1996). To this end, fixing $\epsilon > 0$, for $f \in \mathcal{P}_{M,0}$ we define the function

$$m_f(x) = \log\left(\frac{f(x) + f_0(x)}{2f_0(x)}\right),$$

and we let $\mathcal{M}_\epsilon$ denote the class

$$\mathcal{M}_\epsilon = \{m_f - m_{f_0} \ : \ H(f, f_0) \le \epsilon\}.$$

Also we denote

$$||\mathbb{Z}_n||_{\mathcal{M}_\epsilon} = \sup_{h\in\mathcal{M}_\epsilon} \int h d\mathbb{Z}_n.$$

Then Theorem 3.4.4 of Van der Vaart and Wellner (1996) yields

$$P(m_f - m_{f_0}) \lesssim -H^2(f, f_0),$$

and

$$E_P||\mathbb{Z}_n||_{\mathcal{M}_\epsilon} \lesssim \tilde{J}_{[\,]}(\epsilon, \mathcal{P}_{M,0}, H)\left(1 + \frac{\tilde{J}_{[\,]}(\epsilon, \mathcal{P}_{M,0}, H)}{\epsilon^2\sqrt{n}}\right),$$

where

$$\tilde{J}_{[\,]}(\epsilon, \mathcal{P}_{M,0}, H) = \int_0^\epsilon \sqrt{\log N_{[\,]}(t, \mathcal{P}_{M,0}, H)}dt \lesssim \epsilon^{3/4}.$$

Therefore it is easy to see that

$$E_P||\mathbb{Z}_n||_{\mathcal{M}_\epsilon} \lesssim \epsilon^{3/4} + \epsilon^{-1/2}n^{-1/2} = g(\epsilon),$$

where

$$g(t) = t^{3/4} + t^{-1/2}n^{-1/2}, \quad t > 0.$$

Notice that $g(t)/t^{3/4}$ is decreasing in t. Also, for any constant $c > 0$, we have

$$n^{4/5}g(cn^{-2/5}) = (c^{3/4} + c^{-1/2})n^{1/2}.$$

Therefore an application of Theorem 3.4.1 of Van der Vaart and Wellner (1996) yields

$$H(\widehat{f}_n, f_0) = O_p(n^{-2/5}),$$

which completes the proof of part A of Theorem 2.9.

Now we turn to the proof of part B. If $x - \theta_0$ is a continuity point of g_0', noting $\widehat{\theta}_n \to_{a.s.} \theta_0$ by Theorem 2.8 and Corollary 4, we obtain that

$$\frac{\sqrt{g_0(x-\widehat{\theta}_n)} - \sqrt{g_0(x-\theta_0)}}{(\widehat{\theta}_n - \theta_0)} \to_{a.s.} \frac{g_0'(x-\theta_0)}{2\sqrt{g_0(x-\theta_0)}}.$$

Noting g_0' is continuous almost everywhere with respect to Lebesgue measure, and using Fatou's lemma and part A of the current theorem, we obtain that

$$\begin{aligned} &\liminf_n \frac{\int_{-\infty}^{\infty}\left(\sqrt{g_0(x-\widehat{\theta}_n)} - \sqrt{g_0(x-\theta_0)}\right)^2 dx}{(\widehat{\theta}_n - \theta_0)^2} \\ &\geq \int_{-\infty}^{\infty}\left(\frac{g_0'(x-\theta_0)}{2\sqrt{g_0(x-\theta_0)}}\right)^2 dx = \frac{\mathcal{I}_{f_0}}{4} \ a.s. \end{aligned} \tag{2.97}$$

Now observe that

$$\begin{aligned}
&2H(\widehat{f}_n, f_0)^2 \\
&\quad = \int_{-\infty}^{\infty} \left(\sqrt{\widehat{g}_n(x-\widehat{\theta}_n)} - \sqrt{g_0(x-\theta_0)} \right)^2 dx \\
&\quad = \int_{-\infty}^{\infty} \left(\sqrt{\widehat{g}_n(x-\widehat{\theta}_n)} - \sqrt{g_0(x-\widehat{\theta}_n)} \right)^2 dx \\
&\qquad + \int_{-\infty}^{\infty} \left(\sqrt{g_0(x-\widehat{\theta}_n)} - \sqrt{g_0(x-\theta_0)} \right)^2 dx + T_c,
\end{aligned}$$

where

$$T_c = 2 \int_{-\infty}^{\infty} \left(\sqrt{\widehat{g}_n(x-\widehat{\theta}_n)} - \sqrt{g_0(x-\widehat{\theta}_n)} \right) \left(\sqrt{g_0(x-\widehat{\theta}_n)} - \sqrt{g_0(x-\theta_0)} \right) dx.$$

The inequality in (2.97) entails that for all sufficiently large n,

$$2H(\widehat{f}_n, f_0)^2 \geq 2H(\widehat{g}_n, g_0)^2 + \frac{(\widehat{\theta}_n - \theta_0)^2 \mathcal{I}_{f_0}}{4} - |T_c| \quad a.s.$$

We aim to show that the cross-term $|T_c|$ is small. In fact, we show that

$$\frac{|T_c|}{|\widehat{\theta}_n - \theta_0|^2 + H(\widehat{g}_n, g_0)^2} = o_p(1). \tag{2.98}$$

Suppose (2.98) holds. Then it follows that

$$\begin{aligned}
2H(\widehat{f}_n, f_0)^2 \geq\ & 2H(\widehat{g}_n, g_0)^2 \\
& + \frac{(\widehat{\theta}_n - \theta_0)^2 \mathcal{I}_{f_0}}{4} - o_p(1) H(\widehat{g}_n, g_0)^2 - o_p(1)(\widehat{\theta}_n - \theta_0)^2,
\end{aligned}$$

which completes the proof because $\mathcal{I}_{f_0} > 0$.

Hence it remains to prove (2.98). To this end, notice that T_c can be written as

$$T_c = 2 \int_{-\infty}^{\infty} \left(\sqrt{\widehat{g}_n(x)} - \sqrt{g_0(x)} \right) \left(\sqrt{g_0(x)} - \sqrt{g_0(x + \widehat{\theta}_n - \theta_0)} \right) dx.$$

For the sake of simplicity, we will denote $\theta_0 - \widehat{\theta}_n$ by δ_n from now on. Since g_0 is absolutely continuous, we can write

$$|T_c| = \left| 2 \int_{-\infty}^{\infty} \left(\sqrt{\widehat{g}_n(x)} - \sqrt{g_0(x)} \right) \left(\int_{-\delta_n}^{0} \frac{g_0'(x+t)}{2\sqrt{g_0(x+t)}} dt \right) dx \right|.$$

Since $g_0 \in \mathcal{S}_0$, we have

$$\begin{aligned} |T_c| &= 2 \left| \int_{0}^{\infty} \left(\sqrt{\widehat{g}_n(x)} - \sqrt{g_0(x)} \right) \left(\int_{-\delta_n}^{0} \frac{g_0'(x+t)}{2\sqrt{g_0(x+t)}} dt \right) dx \right. \\ &\quad \left. - \int_{-\infty}^{0} \left(\sqrt{\widehat{g}_n(-x)} - \sqrt{g_0(-x)} \right) \left(\int_{-\delta_n}^{0} \frac{g_0'(-x-t)}{2\sqrt{g_0(-x-t)}} dt \right) dx \right| \\ &= 2 \left| \int_{0}^{\infty} \left(\sqrt{\widehat{g}_n(x)} - \sqrt{g_0(x)} \right) \left(\int_{-\delta_n}^{0} \frac{g_0'(x+t)}{2\sqrt{g_0(x+t)}} dt \right) dx \right. \\ &\quad \left. - \int_{0}^{\infty} \left(\sqrt{\widehat{g}_n(x)} - \sqrt{g_0(x)} \right) \left(\int_{-\delta_n}^{0} \frac{g_0'(x-t)}{2\sqrt{g_0(x-t)}} dt \right) dx \right|, \end{aligned}$$

yielding

$$|T_c| = 2 \left| \int_{0}^{\infty} \left(\sqrt{\widehat{g}_n(x)} - \sqrt{g_0(x)} \right) \left(\int_{-\delta_n}^{0} \left(\frac{g_0'(x+t)}{2\sqrt{g_0(x+t)}} - \frac{g_0'(x-t)}{2\sqrt{g_0(x-t)}} \right) dt \right) dx \right|. \quad (2.99)$$

Using the Cauchy-Schwarz inequality, we obtain that

$$\begin{aligned} \frac{|T_c|}{2|\delta_n|} &\leq \left(\int_{0}^{\infty} \left(\sqrt{\widehat{g}_n(x)} - \sqrt{g_0(x)} \right)^2 dx \right)^{1/2} \\ &\quad \left(\int_{0}^{\infty} \left(\int_{-\delta_n}^{0} \frac{1}{|\delta_n|} \left(\frac{g_0'(x+t)}{2\sqrt{g_0(x+t)}} - \frac{g_0'(x-t)}{2\sqrt{g_0(x-t)}} \right) dt \right)^2 dx \right)^{1/2}. \end{aligned}$$

Since $\sqrt{\widehat{g}_n(x)} - \sqrt{g_0(x)}$ is an even function, the first term on the right hand side of the last inequality is $\sqrt{2} H(\widehat{g}_n, g_0)$. Hence,

$$\frac{T_c^2}{8 H(\widehat{g}_n, g_0)^2 \delta_n^2} \leq \int_{0}^{\infty} \left(\int_{-\delta_n}^{0} \frac{1}{|\delta_n|} \left(\frac{g_0'(x+t)}{2\sqrt{g_0(x+t)}} - \frac{g_0'(x-t)}{2\sqrt{g_0(x-t)}} \right) dt \right)^2 dx,$$

which, noting that

$$t \mapsto \frac{g_0'(x+t)}{2\sqrt{g_0(x+t)}} - \frac{g_0'(x-t)}{2\sqrt{g_0(x-t)}}$$

is an even function for each $x > 0$, can be bounded above by

$$\int_0^\infty |\delta_n| \Bigg(\int_0^{|\delta_n|} \frac{1}{(\delta_n)^2} \left(\frac{g_0'(x+t)}{2\sqrt{g_0(x+t)}} - \frac{g_0'(x-t)}{2\sqrt{g_0(x-t)}} \right)^2 dt \Bigg) dx$$

using the Cauchy-Schwarz inequality. Therefore we obtain

$$\begin{aligned} \frac{T_c^2}{2H(\widehat{g}_n, g_0)^2 \delta_n^2} \leq \frac{1}{|\delta_n|} \int_0^{|\delta_n|} \Bigg[& \int_0^\infty \left(\frac{g_0'(x+t)}{\sqrt{g_0(x+t)}} \right)^2 dx + \int_0^\infty \left(\frac{g_0'(x-t)}{\sqrt{g_0(x-t)}} \right)^2 dx \\ & - 2 \int_0^\infty \frac{g_0'(x-t)}{\sqrt{g_0(x-t)}} \frac{g_0'(x+t)}{\sqrt{g_0(x+t)}} dx \Bigg] dt. \end{aligned} \tag{2.100}$$

For $t \geq 0$,

$$\int_0^\infty \left(\frac{g_0'(x+t)}{\sqrt{g_0(x+t)}} \right)^2 dx = \int_t^\infty \left(\frac{g_0'(x)}{\sqrt{g_0(x)}} \right)^2 dx \leq \frac{\mathcal{I}_{f_0}}{2}.$$

Now observe that for $z \in (-|\delta_n|, 0)$,

$$|g_0'(z)/\sqrt{g_0(z)}| = |\psi_0'(z)| \sqrt{g_0(z)} \leq |\psi_0'(\delta_n)| \sqrt{g_0(0)} = O_p(1), \tag{2.101}$$

since $\psi_0 \in \mathcal{SC}_0$, and $\delta_n \to_{a.s.} 0$. Hence, for $t \in (0, |\delta_n|)$,

$$\begin{aligned} \int_0^\infty \left(\frac{g_0'(x-t)}{\sqrt{g_0(x-t)}} \right)^2 dx &= \int_{-t}^\infty \left(\frac{g_0'(z)}{\sqrt{g_0(z)}} \right)^2 dz \\ &= \int_{-t}^0 \left(\frac{g_0'(x)}{\sqrt{g_0(z)}} \right)^2 dz + \int_0^\infty \left(\frac{g_0'(z)}{\sqrt{g_0(z)}} \right)^2 dz \\ &\leq |\delta_n| \psi_0'(\delta_n)^2 \sqrt{g_0(0)} + \mathcal{I}_{f_0}/2 \\ &= |\delta_n| O_p(1) + \mathcal{I}_{f_0}/2, \end{aligned}$$

where the last step follows from (2.101). Hence for any $t \in (0, |\delta_n|)$,

$$\int_0^\infty \left(\frac{g_0'(x+t)}{\sqrt{g_0(x+t)}} \right)^2 dx + \int_0^\infty \left(\frac{g_0'(x-t)}{\sqrt{g_0(x-t)}} \right)^2 dx = |\delta_n| O_p(1) + \mathcal{I}_{f_0}. \tag{2.102}$$

Our objective is to apply Fatou's lemma on the third term on the right hand side of (2.100). Therefore we want to ensure that the integrand is non-negative. Note that when $x \geq |\delta_n|$ and $t \in (0, |\delta_n|)$, we have $x > t$, which leads to

$$g_0'(x-t) g_0'(x+t) \geq 0. \tag{2.103}$$

Keeping that in mind, we partition the term

$$\begin{aligned}
&-\int_0^\infty \frac{g_0'(x-t)}{\sqrt{g_0(x-t)}}\frac{g_0'(x+t)}{\sqrt{g_0(x+t)}}dx \\
&= -\int_{|\delta_n|}^\infty \frac{g_0'(x-t)}{\sqrt{g_0(x-t)}}\frac{g_0'(x+t)}{\sqrt{g_0(x+t)}}dx - \int_0^{|\delta_n|} \frac{g_0'(x-t)}{\sqrt{g_0(x-t)}}\frac{g_0'(x+t)}{\sqrt{g_0(x+t)}}dx \\
&\leq -\int_{|\delta_n|}^\infty \frac{g_0'(x-t)}{\sqrt{g_0(x-t)}}\frac{g_0'(x+t)}{\sqrt{g_0(x+t)}}dx + |\delta_n|O_p(1),
\end{aligned}$$

where the last step follows from (2.101). The above combined with (2.100) and (2.102) leads to

$$\begin{aligned}
&\limsup_n \frac{T_c^2}{2H(\widehat{g}_n, g_0)^2\delta_n^2} \\
&\leq \limsup_n \frac{1}{|\delta_n|}\int_0^{|\delta_n|}\left[|\delta_n|O_p(1) + \mathcal{I}_{f_0} - 2\int_{|\delta_n|}^\infty \frac{g_0'(x-t)}{\sqrt{g_0(x-t)}}\frac{g_0'(x+t)}{\sqrt{g_0(x+t)}}dx\right]dt \\
&= O_p(1)\limsup_n |\delta_n| + \mathcal{I}_{f_0} \\
&\quad - 2\liminf_n \frac{1}{|\delta_n|}\int_0^{|\delta_n|}\int_{|\delta_n|}^\infty \frac{g_0'(x-t)}{\sqrt{g_0(x-t)}}\frac{g_0'(x+t)}{\sqrt{g_0(x+t)}}dxdt \\
&= 0 + \mathcal{I}_{f_0} - 2\liminf_n \int_{|\delta_n|}^\infty \frac{\displaystyle\int_0^{|\delta_n|}\frac{g_0'(x+t)}{\sqrt{g_0(x+t)}}\frac{g_0'(x-t)}{\sqrt{g_0(x-t)}}dt}{|\delta_n|}dx. \qquad (2.104)
\end{aligned}$$

Therefore an application of Fatou's Lemma and (2.103) yield

$$\liminf_n \int_{|\delta_n|}^\infty \frac{\displaystyle\int_0^{|\delta_n|}\frac{g_0'(x+t)}{\sqrt{g_0(x+t)}}\frac{g_0'(x-t)}{\sqrt{g_0(x-t)}}dt}{|\delta_n|}dx \geq \int_0^\infty \frac{g_0'(x)^2}{g_0(x)}dx = \frac{\mathcal{I}_{f_0}}{2}.$$

Thus (2.104) leads to

$$\frac{2T_c^2}{4H(\widehat{g}_n, g_0)^2\delta_n^2} = o_p(1).$$

from which it is obvious that

$$\frac{\sqrt{2}|T_c|}{|\delta_n|^2 + H(\widehat{g}_n, g_0)^2} \leq \frac{\sqrt{2}|T_c|}{2H(\widehat{g}_n, g_0)|\delta_n|} = o_p(1),$$

which proves (2.98) and thus completes the proof of part B of Theorem 2.9. □

Proof of Lemma 2.8

From part A of Theorem 2.9, we obtain that $\widehat{f}_n$ is strongly Hellinger consistent for f_0. As a consequence, $||\widehat{f}_n - f_0||_1 \to_p 0$. Therefore from proposition 2(c) of Cule and Samworth (2010), it follows that,

$$\sup_{z\in\mathbb{R}} |\widehat{f}_n(x) - f_0(x)| \to_{a.s.} 0. \tag{2.105}$$

Now consider any compact set $K = [b_1, b_2] \subset \text{dom}(\phi_0)$. Observe that

$$\begin{aligned}
&\sup_{x\in K} |\widehat{\phi}_{0,n}(x) - \phi_0(x)| \\
&= \sup_{x\in K} \log \frac{\max(\widehat{f}_n(x), f_0(x))}{\min(\widehat{f}_n(x), f_0(x))} \\
&= \sup_{x\in K} \log \left(\frac{|\widehat{f}_n(x) - f_0(x)|}{\min(\widehat{f}_n(x), f_0(x))} + 1 \right) \\
&\leq \log \left(\frac{\sup_{x\in K} |\widehat{f}_n(x) - f_0(x)|}{\min(\widehat{f}_n(b_1), \widehat{f}_n(b_2), f_0(b_1), f_0(b_2))} + 1 \right),
\end{aligned}$$

since $\widehat{f}_n$ and f_0 are unimodal. From (2.105) it follows that $\sup_{x\in K} |\widehat{f}_n(x) - f_0(x)| \to_{a.s.} 0$. Also since $b_1 \in \text{dom}(\phi_0)$, $\widehat{f}_n(b_1) \to_{a.s.} f_0(b_1) > 0$. The same holds for $\widehat{f}_n(b_2)$, which establishes that

$$\sup_{x\in K} |\widehat{\phi}_{0,n}(x) - \phi_0(x)| \to_{a.s.} 0.$$

Now from Lemma F.10 of Kuchibhotla *et al.* (2017), it follows that if ϕ_0 is differentiable at $x \in K$, leading to

$$\widehat{\phi}'_{0,n}(x) \to_{a.s.} \phi'_0(x),$$

which combined with (2.105) yields

$$\widehat{f}'_n(x) \to_{a.s.} f'_0(x).$$

Part A of Theorem 2.9 also indicates that $H(\widehat{g}_n, g_0) \to_{a.s.} 0$, from which, proceeding like above, one can show that similar results hold for $\widehat{g}_n$ and $\widehat{\psi}_{0,n}$ as well, which completes the proof. □

2.6.3 Additional lemmas

2.6.4 Additional lemmas for the one-step estimators

Lemma 2.9. *Consider $\zeta, \zeta' \in int(dom(\tilde{\psi}_0))$ such that $\zeta \leq \zeta'$. Suppose $\zeta_n \to_{a.s.} \zeta$, and $\zeta_n' \to_{a.s.} \zeta'$. Letting $\tilde{\psi}_{0_n}'$ denote $\log \tilde{g}_n$, for $\tilde{g}_n = \tilde{g}_n^{sym}$, $\widehat{g}_{\bar{\theta}_n}$, $\tilde{g}_n^{geo,sym}$, or $\bar{f}_m(\bar{\theta}_n \pm \cdot)$, under the conditions of Theorem 2.5, we have*

$$\limsup_n \sup_{z \in [\zeta_n, \zeta_n']} |\tilde{\psi}_{0_n}'(z)| \leq C_{\zeta,\zeta'} \quad a.s.,$$

where $C_{\zeta,\zeta'}$ is a constant depending on ζ and ζ'. When $\tilde{g}_n = (\widehat{f}_n^{sym})^{sm}$ or $\widehat{f}_n^{sm}(\bar{\theta}_n \pm \cdot)$, the above conclusion holds for any $\zeta, \zeta' \in \mathbb{R}$.

Proof. When $\tilde{g}_n$ equals the densities $\widehat{g}_{\bar{\theta}_n}$, $\widehat{f}_n^{geo,sym}$, $\bar{f}_m(\bar{\theta}_n \pm \cdot)$, or $\widehat{f}_n^{sm}(\cdot + \bar{\theta}_n)$, it is clear that $\tilde{\psi}_n$ is concave. We will consider these choices of $\tilde{g}_n$ first. For the first three cases, the pointwise limit of $\tilde{g}_n$ is $\tilde{g}_0$ by Theorem 2.5. Note that we can choose $\zeta_1, \zeta_1' \in \text{dom}(\tilde{\psi}_0)$ such that $\tilde{\psi}_0'$ is continuous at ζ_1 and ζ_1', and $[\zeta - \epsilon, \zeta' + \epsilon] \subset [\zeta_1, \zeta_1']$ for some sufficiently small $\epsilon > 0$. Now the conditions on ζ_n and ζ_n' underscore that for sufficiently large n,

$$[\zeta_n, \zeta_n'] \subset [\zeta_1, \zeta_1'] \quad a.s.$$

Now note that a concave $\tilde{\psi}_n$ leads to a non-increasing $\tilde{\psi}_{0_n}'$. Therefore on any interval, $|\tilde{\psi}_{0_n}'|$ attains its maxima on either of the endpoints. As a consequence, the supremum of $|\tilde{\psi}_{0_n}'|$ on the interval $[\zeta_1, \zeta_1']$ does not exceed $|\tilde{\psi}_{0_n}'(\zeta_1)| + |\tilde{\psi}_{0_n}'(\zeta_1')|$, which converges almost surely to $|\tilde{\psi}_0'(\zeta_1)| + |\tilde{\psi}_0'(\zeta_1')|$ by Theorem 2.5. Therefore, we conclude that

$$\limsup_n \sup_{z \in [\zeta_n, \zeta_n']} |\tilde{\psi}_{0_n}'(z)| < |\tilde{\psi}_0'(\zeta_1)| + |\tilde{\psi}_0'(\zeta_1')| \quad a.s., \tag{2.106}$$

when $\tilde{g}_n = \widehat{g}_{\bar{\theta}_n}$, $\widehat{f}_n^{geo,sym}$, or $\bar{f}_m(\cdot + \bar{\theta}_n)$. Now consider $\tilde{g}_n = \widehat{f}_n^{sm}(\bar{\theta}_n \pm \cdot)$. Note that the L_1 limit of this $\tilde{g}_n$ is $\tilde{g}_0^{sm}$, and $\text{dom}(\tilde{\psi}_0^{sm}) = \mathbb{R}$. Therefore for any $\zeta, \zeta' \in \mathbb{R}$, the same arguments as the previous case will lead to

$$\limsup_n \sup_{z \in [\zeta_n, \zeta_n']} |\tilde{\psi}_{0_n}'(z)| < |(\tilde{\psi}_0^{sm})'(\zeta_1)| + |(\tilde{\psi}_0^{sm})'(\zeta_1')| \quad a.s. \tag{2.107}$$

Now let us consider the non-log-concave cases. Assume $\tilde{g}_n = \tilde{g}_n^{sym}$. Then from (2.55) we obtain that

$$\tilde{\psi}_{0n}'(x) = \varrho_n(x)\Big((\log \bar{f}_m)'(\bar{\theta}_n + x)\Big) - (1 - \varrho_n(x))\Big((\log \bar{f}_m)'(\bar{\theta}_n - x)\Big),$$

where $\bar{f}_m$ is the log-concave MLE and $\varrho_n(x) < 1$, leading to

$$\limsup_n \sup_{x\in[\zeta_n,\zeta_n']} |\tilde{\psi}_{0n}'(x)| \le \sup_{x\in[\zeta_n,\zeta_n']} |(\log \bar{f}_m)'(\bar{\theta}_n + x)| + \sup_{x\in[\zeta_n,\zeta_n']} |(\log \bar{f}_m)'(\bar{\theta}_n - x)|,$$

which is bounded above by (2.106). Finally when $\tilde{g}_n = (\widehat{f}_n^{sym})^{sm}$, from (2.56) we observe that $\tilde{\psi}_{0n}'$ can be expressed in terms of $\widehat{f}_n^{sm}$ as

$$\tilde{\psi}_{0n}'(x) = \varrho_n^{sm}(x)\Big((\log \widehat{f}_n^{sm})'(\bar{\theta}_n + x)\Big) - (1 - \varrho_n^{sm}(x))\Big((\log \widehat{f}_n^{sm})'(\bar{\theta}_n - x)\Big),$$

for some $\varrho_n^{sm}(x) < 1$. Boundedness of $\tilde{\psi}_{0n}'$ then follows from (2.107). □

2.6.5 Additional lemmas for the MLE

Lemma 2.10. *Suppose $e^\psi \in \mathcal{SLC}_0$. Then $\psi(x) \le -\log|2x|$ for all $x \in \mathbb{R}$.*

Proof. Note that since $e^\psi \in \mathcal{SLC}_0$, for $x > 0$, we have,

$$1 = \int_{-\infty}^{\infty} e^{\psi(z)}dz \ge \int_{-x}^{x} e^{\psi(z)}dz,$$

which is bounded below by $2e^{\psi(x)}x$ because $\psi \in \mathcal{SC}_0$. Hence the result follows. □

Lemma 2.11. *Suppose F is a distribution function satisfying Condition A. Denote*

$$int(J(F)) = (a, b),$$

for some $a, b \in [-\infty, \infty]$. Suppose ψ_θ is the maximizer of $\Psi(\theta, \psi, F)$ over $\psi \in \mathcal{SC}_0$ for some $\theta \in \mathbb{R}$. Then

$$int(dom(\psi_\theta)) = (-d, d),$$

where $d = (b - \theta) \vee (\theta - a)$.

Note that a or b can be ∞, in which case $d = \infty$, and $\text{int}(\text{dom}(\psi_\theta)) = \mathbb{R}$.

Proof. Note that $\text{int}(\text{dom}(\psi_\theta)) = (-c, c)$ for some $c > 0$. We first show that $c \geq d$. Consider the set $A = \overline{(-d, d)} \setminus \text{dom}(\psi_\theta)$. Note that unless $d = \infty$, $\overline{(-d, d)} = [-d, d]$. Letting Δ denote the indicator function 1_A, we observe that $\psi_\theta + \Delta = \psi_\theta$. Also note that

$$\psi_\theta = \underset{\psi \in \mathcal{SC}_0}{\arg\max}\, \Psi(0, \psi, F(\cdot + \theta)) = \underset{\psi \in \mathcal{SC}_0}{\arg\max}\, \omega(\psi, F(\cdot + \theta)),$$

which follows from (2.3) and (2.21). Hence Proposition 5(ii) of Xu and Samworth (2017) leads to

$$\int_{-\infty}^{\infty} \Delta(x) dF(x + \theta) \leq \int_{-\infty}^{\infty} \Delta(x) e^{\psi_\theta(x)} dx.$$

Since $\Delta = 0$ on $\text{dom}(\psi_\theta)$, it follows that

$$\int_A dF(x + \theta) = 0,$$

implying $(a, b) \cap (A + \theta) = \emptyset$.

It is easy to see that if either a or b equals ∞, $d = \infty$, and $A + \theta = \mathbb{R} \setminus \text{dom}(\psi_\theta)$. Therefore

$$(a, b) \cap (A + \theta) = (a, b) \setminus \Big(\text{dom}(\psi_\theta) + \theta \Big) = \emptyset,$$

which indicates that

$$(a, b) \subset (-c, c) + \theta, \tag{2.108}$$

implying $(-c, c) = \mathbb{R}$. Hence $(-c, c) = (-d, d)$, which completes the proof for this case. Therefore it only remains to prove Lemma 2.11 for the case when both a and b are finite.

When $a, b < \infty$, there can be two sub-cases. In the first case, $b - \theta \geq \theta - a$, which indicates $d = b - \theta$. In this case, $[-d, d] + \theta = [2\theta - b, b]$, and

$$[a, b] \subset [2\theta - b, b] = [-d, d] + \theta,$$

leading to

$$(a, b) \setminus \Big(\text{dom}(\psi_\theta) + \theta \Big) = (a, b) \cap (A + \theta) = \emptyset,$$

from which (2.108) follows. Therefore $b \le c + \theta$, or $d \le c$. In the second case $b - \theta \le \theta - a$, $d = \theta - a$ and $[-d, d] + \theta = [a, 2\theta - a]$. Proceeding similarly as in the first case, we can again show that $c \ge d$. Hence we have established that $c \ge d$ in general.

Now suppose $c > d$, which implies that $a, b < \infty$. Observe that in this case, there exists a $\psi \in \mathcal{SC}_0$, which equals ψ_θ on $(-d, d)$ and $-\infty$ everywhere else. It is easy to verify that when $c > d$,

$$\begin{aligned} &\Psi(\theta, \psi_\theta, F) - \Psi(\theta, \psi, F) \\ &= \int_{-d}^{d} e^{\psi_\theta(x)} dx - \int_{-c}^{c} e^{\psi_\theta(x)} dx < 0. \end{aligned}$$

However, ψ_θ satisfies $\Psi(\theta, \psi_\theta, F) \ge \Psi(\theta, \psi, F)$ for all $\psi \in \mathcal{SC}_0$, which leads to a contradiction. Hence we conclude that $c = d$, which completes the proof. □

Lemma 2.12. *Suppose F satisfies Condition A and (2.24). Then*

$$\arg\max_{\phi \in \mathcal{C}} \omega(\phi, F) = \arg\max_{\phi \in \mathcal{SC}_\theta} \omega(\phi, F),$$

where the criterion function ω is defined in (2.21).

Proof. First we consider the special case $\theta = 0$. For $\phi \in \mathcal{C}$ and F symmetric about 0, using (2.24) with $\theta = 0$, the criterion function $\omega(\phi, F)$ can be written as

$$\begin{aligned} &\int_{-\infty}^{0} \phi(x) dF(x) + \int_{0}^{\infty} \phi(x) dF(x) - \int_{-\infty}^{\infty} e^{\phi(x)} dx \\ &= -\int_{0}^{\infty} \phi(-x) dF(-x) + \int_{0}^{\infty} \phi(x) dF(x) - \int_{-\infty}^{\infty} e^{\phi(x)} dx \\ &= \int_{0}^{\infty} \phi(-x) dF(x) + \int_{0}^{\infty} \phi(x) dF(x) - \int_{-\infty}^{\infty} e^{\phi(x)} dx \\ &= \int_{-\infty}^{\infty} \frac{\phi(-x) + \phi(x)}{2} dF(x) - \int_{-\infty}^{\infty} e^{\phi(x)} dx, \end{aligned}$$

which is not larger than

$$\int_{-\infty}^{\infty} \frac{\phi(x) + \phi(-x)}{2} dF(x) - \int_{-\infty}^{\infty} e^{(\phi(x) + \phi(-x))/2} dx,$$

by the convexity of the exponential function. This proves that a $\phi \in \mathcal{SC}_0$ maximizes $\omega(\phi, F)$ over $\mathcal{C}$ when F is symmetric about 0. The proof for a general θ follows in a similar way. □

Lemma 2.13. *For any $\psi \in \mathcal{SC}_0$ with nonempty domain, and each $\epsilon > 0$, there exists a function $\psi^{(\epsilon)} \in \mathcal{SC}_0$ for each $\epsilon > 0$ such that the following conditions are met:*

(A) $\psi(x) \leq \psi^{(\epsilon)}(x) \leq \psi(0)$ for all $x \in \mathbb{R}$.

(B)

$$\psi^{(\epsilon)}(x) \geq -|x|/\epsilon + \psi(0),$$

for all $x \in \mathbb{R}$.

(B) $\psi^{(\epsilon)}(x) \downarrow \psi(x)$ for each $x \in \mathbb{R}$ as $\epsilon \downarrow 0$.

Proof. Following the construction in Lemma 4.3 of Dümbgen *et al.* (2011) we set

$$\psi^{(\epsilon)}(x) = \inf_{(t,c)\in A}(tx + c), \quad x \in \mathbb{R}, \tag{2.109}$$

where

$$A = \left\{(t, c) \in \mathbb{R} \times \mathbb{R} \;\middle|\; |t| \leq 1/\epsilon,\ tx + c \geq \psi(x) \text{ for all } x \in \mathbb{R}\right\}.$$

Note that $\psi^{(\epsilon)}(x)$ is concave by Lemma 4.3 of Dümbgen *et al.* (2011). To show that $\psi^{(\epsilon)}$ is symmetric about 0, observe that if $tx + c \geq \psi(x)$ for all $x \in \mathbb{R}$, the symmetry of ψ about 0 leads to

$$-t(-x) + c \geq \psi(-x) \text{ for all } x \in \mathbb{R},$$

which can be rewritten as

$$-tx + c \geq \psi(x) \text{ for all } x \in \mathbb{R}.$$

Hence $(t, c) \in A$ implies that $(-t, c) \in A$. As a result,

$$\inf_{(t,c)\in A}(tx + c) = \inf_{(-t,c)\in A}(tx + c) = \inf_{(t,c)\in A}(-tx + c),$$

which implies that $\psi^{(\epsilon)} \in \mathcal{SC}_0$. That $\psi \leq \psi^{(\epsilon)}$ follows from Lemma 4.3 of Dümbgen *et al.* (2011). Hence to prove part A, we need to show that $\psi^{(\epsilon)}$ is bounded above by $\psi(0)$.

Observe that since $\psi^{(\epsilon)} \in \mathcal{SC}_0$, it is bounded above by $\psi^{(\epsilon)}(0)$, which, by (2.109), equals the infrimum of the set

$$A' = \{c \ : \ (t, c) \in A \text{ for some } t \in \mathbb{R}\}.$$

However, since $c \in A'$ satisfies $tx + c \geq \psi(x)$ for all $x \in \mathbb{R}$, and some t, we conclude that Thus we obtain

$$\psi^{(\epsilon)}(0) = \inf A' \geq \psi(0). \tag{2.110}$$

In light of (2.110), to prove part A of the current lemma, it suffices to show that $\psi(0) \in A'$. Since $\psi \in \mathcal{SC}_0$, for $x > 0$, we have $\psi(x) \leq \psi'(0+)x + \psi(0)$. When $x \leq 0$, using the fact that $\psi'(0+) \leq 0$, we calculate

$$\psi(x) = \psi(-x) \leq -\psi'(0+)x + \psi(0) \leq \psi'(0+)x + \psi(0).$$

Therefore for all $x \in \mathbb{R}$, the following holds:

$$\psi(x) \leq \psi'(0+)x + \psi(0).$$

When $-\psi'(0+) \leq 1/\epsilon$, (2.109) thus indicates that $(-\psi'(0+), \psi(0)) \in A$. In case $-\psi'(0+) > 1/\epsilon$, we still have

$$\psi(x) \leq \psi'(0+)x + \psi(0) < -x/\epsilon + \psi(0), \qquad \text{for all } x > 0.$$

For $x \leq 0$,

$$\psi(x) = \psi(-x) \leq x/\epsilon + \psi(0) \leq -x/\epsilon + \psi(0).$$

Therefore, it turns out that if $-\psi(0+) > 1/\epsilon$,

$$\psi(x) \leq -x/\epsilon + \psi(0) \text{ for all } x \in \mathbb{R},$$

which implies that $(1/\epsilon, \psi(0)) \in A$. The above calculations indicate that there always exists a $t \in \mathbb{R}$ such that $(t, \psi(0)) \in A$, entailing that $\psi(0) \in A'$. This fact, combined with (2.110), yields that $\psi^{(\epsilon)}(0) = \psi(0)$. Part A follows from (2.110).

Note that (2.109) and (2.110) imply that

$$\psi^{(\epsilon)}(x) \geq -|x|/\epsilon + \psi(0), \quad x \in \mathbb{R},$$

which settles the proof of part B.

Lemma 2.13(C) follows directly from Lemma 4.3 of Dümbgen *et al.* (2011). □

The next lemma is Pratt's lemma (Pratt, 1960, Theorem 1). We state it here for convenience.

Lemma 2.14. *Suppose $(\Omega, \mathcal{F}, \mu)$ is a measure space and a_n, b_n, c_n are sequences of functions on Ω converging almost everywhere to functions a, b, c respectively. Also all functions are integrable and $\int a_n d\mu \to \int a d\mu$ and $\int c_n d\mu \to \int c d\mu$. Moreover, $a_n \leq b_n \leq c_n$. Then*

$$\int b_n d\mu \to \int b d\mu.$$

Chapter 3

Bi-s^*-concave distributions

3.1 Introduction

Dümbgen *et al.* (2017) investigated a shape constraint they called "bi-log-concavity" for distribution functions F on $\mathbb{R}$: a distribution function F is *bi-log-concave* if both $x \mapsto \log F(x)$ and $x \mapsto \log(1 - F(x))$ are concave functions of x. They noted that Bagnoli and Bergstrom (2005) showed that any log-concave distribution with density f has a bi-log-concave distribution function F, but that the inclusion is proper: there are many bi-log-concave distributions that are not log-concave, and in fact bi-log-concave distributions may not be unimodal. Dümbgen *et al.* (2017) proved the following interesting theorem characterizing the class of bi-log-concave distributions.

First a bit of notation:

$$J(F) \equiv \{x \in \mathbb{R} : \ 0 < F(x) < 1\}.$$

A distribution function F is non-degenerate if $J(F) \neq \emptyset$.

Theorem 3.1. *(DKW, 2017) For a non-degenerate distribution function F the following four statements are equivalent:*

(i) F is bi-log-concave.

(ii) F is continuous on $\mathbb{R}$ and differentiable on $J(F)$ with derivative $f = F'$ such that

$$F(x+t) \begin{cases} \leq F(x) \exp\left(\frac{f(x)}{F(x)} t\right) \\ \geq 1 - (1 - F(x)) \exp\left(-\frac{f(x)}{(1-F(x))} t\right) \end{cases}$$

for all $x \in J(F)$ and $t \in \mathbb{R}$.

(iii) F is continuous on $\mathbb{R}$ and differentiable on $J(F)$ with derivative $f = F'$ such that the hazard function $f/(1-F)$ is non-decreasing and reverse hazard function f/F is non-increasing on $J(F)$.

(iv) F is continuous on $\mathbb{R}$ and differentiable on $J(F)$ with bounded and strictly positive derivative $f = F'$. Furthermore, f is locally Lipschitz-continuous on $J(F)$ with L^1-derivative $f' = F''$ satisfying

$$\frac{-f^2}{1-F} \leq f' \leq \frac{f^2}{F}.$$

An important implication of (iv) of Theorem 3.1 is that the inequalities can be rewritten as follows:

$$-1 \leq -F(x) \leq F(x)(1-F(x))\frac{f'(x)}{f^2(x)} \leq 1 - F(x) \leq 1.$$

This implies that the bi-log-concave family of distributions satisfies

$$\gamma(F) \equiv \sup_{x \in J(F)} F(x)(1-F(x))\frac{|f'(x)|}{f^2(x)} \leq 1. \tag{3.1}$$

The parameter $\gamma(F)$ arises in the study of quantile processes and transportation distances between empirical distributions and true distributions on $\mathbb{R}$: see e.g. Csörgő and Révész (1978), Shorack and Wellner (1986, 2009b) Chapter 18, page 643, Bobkov and Ledoux (2017), and del Barrio *et al.* (2005).

3.2 Questions and extensions: the bi$-s^*$-concave class

This immediately raises several questions:

Question 1 What about distributions in classes larger than the log-concave class? In particular what happens for the $s-$concave classes described by Borell (1975)? See Dharmadhikari and Joag-Dev (1988), and Brascamp and Lieb (1976).

Question 2 Is there a class of bi-s^*-concave distributions with the property that if f is $s-$concave, then F is bi-s-concave (or perhaps bi-s^*-concave with s^* related to s)?

Question 3 Is there a class of bi-s^*-concave distributions with a theorem analogous to Theorem 3.1 with an analogue of Theorem 3.1(iv) implying that $\gamma(F)$ is bounded by some function of s for all bi-s^*-concave distributions F?

We provide positive answers to Questions 1-3 when $s \in (-1, \infty)$, beginning with the following definition of *bi-s^*-concavity* of a distribution function F.

Definition 1. For $s \in (-1, \infty]$ we let $s^* \equiv s/(1+s) \in (-\infty, 1]$. For $s \in (-1, 0)$ we say that a distribution function F on $\mathbb{R}$ is bi$-s^*-$concave if both $x \mapsto F^{s^*}(x)$ and $x \mapsto (1-F(x))^{s^*}$ are convex functions of $x \in J(F)$. For $s \in (0, \infty)$ we say that F is bi$-s^*-$concave if $x \mapsto F^{s^*}(x)$ is concave for $x \in (\inf J(F), \infty)$ and $x \mapsto (1 - F(x))^{s^*}$ is concave for $x \in (-\infty, \sup J(F))$. For $s = 0$ we say that F is bi-0-concave (or bi-log-concave) if both $x \mapsto \log F(x)$ and $x \mapsto \log(1 - F(x))$ are concave functions of $x \in J(F)$. Note that this definition of bi-log-concavity is equivalent to the definition of bi-log-concavity given by Dümbgen *et al.* (2017).

To briefly explain this definition, recall that a density function f (or just a non-negative function f) on $\mathbb{R}$ (or even on $\mathbb{R}^d$) is $s-$concave for $s < 0$ if f^s is convex, while f is $s-$concave for $s > 0$ if f^s is concave on $J(F)$. Furthermore, from the theory of concave measures due

to Borell (1975), Brascamp and Lieb (1976), and Rinott (1976), if f is $s-$concave, the probability measure P on $(\mathbb{R}, \mathcal{B})$ defined by $P(B) = \int_B f(x)dx$ for Borel sets B, is $t-$concave with $t = s/(1+s) \equiv s^*$ if $s > -1$; see Dharmadhikari and Joag-Dev (1988) for an introduction, and Gardner (2002) for a comprehensive review. From the basic theory of Borell, Brascamp and Lieb, and Rinott, it follows easily that if f is $s-$concave with $s \in (-1, \infty]$, then F and $1-F$ are s^*-concave; i.e. the distribution function F corresponding to f is bi$-s^*-$concave. This proof, as well as a simpler calculus type proof assuming that derivatives exist, is given in Section 3.3. The same argument also establishes the corresponding implication in the log-concave case since, in the log-concave case, $s = 0$ and $s^* = 0$ as well. In Section 3.5 we provide a complete characterization of the class of bi-s^*-concave distributions on $\mathbb{R}$, answering Question 3.

For the moment we illustrate the definition with several examples.

Example 3. Suppose f_r is the $t-$density with $r > 0$ "degrees of freedom":

$$f_r(x) = \frac{C_r}{\left(1 + \frac{x^2}{r}\right)^{(r+1)/2}} \quad \text{for} \quad x \in \mathbb{R}.$$

Here $C_r = \Gamma((r+1)/2)/(\sqrt{\pi}\Gamma(r/2))$. It is well-known (see e.g. Borell (1975)) that $f_r \in \mathcal{P}_s$, the class of $s-$concave densities, if $s \leq -1/(1+r)$. Note that s takes values in $(-1, 0)$ since $r \in (0, \infty)$. From the Borell-Brascamp-Lieb inequality we guess that the "right" transformation h of F and $1-F$ to define the Bi$-s^*-$concave class is $h(u) = u^{s^*} = u^{s/(1+s)}$ where $s = -1/(1+r)$, the largest possible value of s. This leads directly to Definition 1. Note that s^* in the Borell-Brascamp-Lieb inequality is well-defined since $s > -1$. Since $s = -(1+r)^{-1}$ we see that we can take $s^* = s/(1+s) = -1/r$ for the t_r family. Then we want to know if $F_r^{-1/r}$ and $(1-F_r)^{-1/r}$ are convex. Direct computation shows that these are convex functions of x. Plotting these for $r \in \{1/2, 1, 4\}$ we see that they are indeed convex. Moreover we find that $\gamma(F_r) = 1+1/r = 1/(1+s)$; this agrees nicely with the log-concave and bi-log-concave picture when $r = \infty$ (so $\gamma(F_\infty) = \gamma(N(0,1)) = 1$), and it yields distributions

with arbitrarily large values of $\gamma(F)$ by considering $\gamma(F_r)$ with r arbitrarily small. Note, in particular, that this yields $\gamma(F_1) = \gamma(Cauchy) = 2$. Also note that this suggests the conjecture $\gamma(F) \leq 1/(1+s)$ for all bi-s^*-concave distribution functions F where $1/(1+s)$ varies from 1 to ∞ as s varies from 0 to -1.

Example 4. Suppose that $f_{a,b}$ is the family of $F-$distributions with "degrees of freedom" $a > 0$ and $b > 0$. (In statistical practice, if T has the density $f_{a,b}$, this would usually be denoted by $T \sim F_{b,a}$ where b is the "numerator degrees of freedom" and a is the "denominator degrees of freedom".) The density is given by

$$f_{a,b}(x) = C_{a,b} \frac{x^{(b/2)-1}}{(a+bx)^{(a+b)/2}} \quad \text{for} \quad x \geq 0.$$

(In fact, $C(a,b) = a^{a/2}b^{b/2}/\text{Beta}(a/2, b/2)$, and $f_{a,b}(x) \to g_b(x)$ as $a \to \infty$ where g_b is the Gamma density with parameters $b/2$ and $b/2$.) It is well-known (see e.g. Borell (1975)) that $f_{a,b} \in \mathcal{P}_s$, the class of $s-$concave densities, if $s \leq -1/(1+\frac{a}{2})$ when $a \geq 2$ and $b \geq 2$. This implies that $s \in [-1/2, 0)$, and the resulting $s^* = s/(1+s)$ is in $[-1, 0)$. By Proposition 3 it follows that F^{s^*} and $(1-F)^{s^*}$ are convex; i.e. F and $1-F$ are s^*- concave. This is confirmed by numerical computation.

Example 5. Suppose that $f_{a,b}(x) \equiv f(x; a, b) = (a/b)(x/b)^{-(a+1)} 1_{[b,\infty)}(x)$, the Pareto distribution with parameters a and b. In this case $f_{a,b}$ is $s-$concave for each $s \leq -1/(1+a)$. Thus we take $s = -1/(1+a) \in (-1, 0)$ for $a \in (0, \infty)$. Note that $s^* = s/(1+s) = -1/a$. Note that $f_{a,b}^s(x) = (x/b) \cdot (b/a)^{1/(1+a)}$ is certainly convex. Furthermore, it is easily seen that

$$CR_R(x) \equiv (1 - F(x)) \frac{f'(x)}{f^2(x)} = 1 - s^* = 1 + 1/a \quad \text{for all} \quad x > b.$$

Thus the Pareto distribution is analogous to the exponential distribution in the log-concave case in the sense that it is exactly on the convex *and* concave boundary.

Example 6. Suppose that $f_r(x) = C_r(1 - x^2/r)^{r/2} 1_{[-\sqrt{r},\sqrt{r}]}(x)$ where $r \in (0, \infty)$. Here

$$C_r = \Gamma((3+r)/2)/(\sqrt{\pi r}\Gamma(1+r/2)).$$

Note that f_r is $s-$concave with $s = 2/r \in (0, \infty)$ since $f_r^{2/r}(x) = C_r^{2/r}(1-x^2/r)1_{[-\sqrt{r},\sqrt{r}]}(x)$ is concave. As $r \to \infty$ it is easily seen that $f_r(x) \to (2\pi)^{-1/2} \exp(-x^2/2)$, the standard normal density. Thus $r = \infty$ corresponds to $s = 0$. On the other hand,

$$\begin{aligned} g_r(x) &\equiv \sqrt{r} f_r(\sqrt{r}x) = \sqrt{r} C_r (1 - x^2)^{r/2} 1_{[-1,1]}(x) \\ &\to 2^{-1} 1_{[-1,1]}(x) \quad \text{as } r \to 0. \end{aligned}$$

Thus $r = 0$ corresponds to $s = +\infty$.

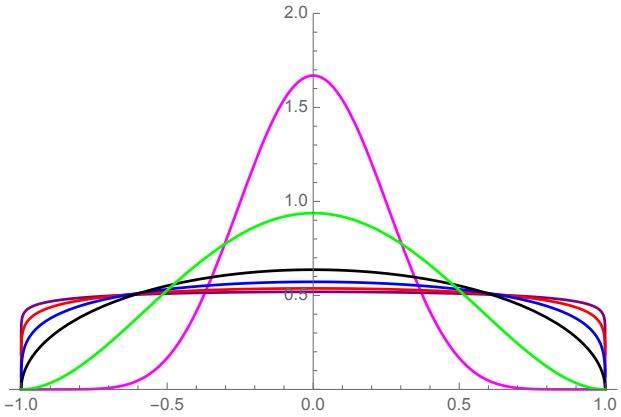


Figure 3.1: The $s-$concave densities g_r of Example 6 with $s = 2/r \in (0, \infty)$: $s = 1/8$, magenta; $s = 1/2$, green; $s = 2$, black; $s = 4$, blue; $s = 8$, red; $s = 16$, purple.

3.3 s-concavity of f implies s^*-concavity of F and $1 - F$

Motivated by Examples 3-6, we first give an extension of the log-concave preservation result of Bagnoli and Bergstrom (2005); also see Lemma 3 of An (1998).

Proposition 2. *(Bagnoli and Bergstrom; An; Barlow and Proschan)*
If f is log-concave then both F and $1-F$ are log-concave; i.e. $\log F$ and $\log(1-F)$ are concave.

Proposition 3. *If f is $s-$concave with $s \in (-1, \infty)$, then both F and $1-F$ are $s^* = s/(1+s)$ concave; i.e. F^{s^*} and $(1-F)^{s^*}$ are convex when $s < 0$; and $\log F$ and $\log(1-F)$ are concave when $s = 0$; and F^{s^*} and $(1-F)^{s^*}$ are convex when $s > 0$. Equivalently, F is bi-s^*-concave.*

Remark 1. Results related to Proposition 2 have a long history in reliability theory and econometrics. Barlow and Proschan (1975) (Lemma 5.8, page 77) showed that if f is log-concave (i.e. PF_2, or Polya frequency of order 2), then $f/(1-F)$ is non-decreasing (or "Increasing Failure Rate" in their terminology); they also noted that the IFR property is equivalent to $1-F$ being log-concave. Their proof of the IFR property using the equivalence of log-concavity of f and $f \in PF_2$ is delightfully short and does not rely on existence of f'. An (1998) also proves Proposition 2 using PF_2 equivalences to log-concavity without requiring existence of f'. The simple "calculus based" proof given here and taken from Bagnoli and Bergstrom (2005), which relies on the classical "second-order conditions" for convexity (see e.g. Boyd and Vandenberghe (2004), section 3.1.4), was apparently given by Dierker (1991), but is likely to have a much longer history.

In the modern theory of convexity, Proposition 2 is an immediate consequence of the results of Prekopa (1973). As we will see in the second proof, Proposition 3 is an immediate consequence of the results of Borell (1975), Brascamp and Lieb (1976), and Rinott (1976).

Proof of Proposition 2

First Proof, assuming f' exists:

Fact 1: First note that f is log-concave if and only if f'/f is non-increasing.

Fact 2: Note that $F(x) = \int_a^x f(y)dy$ is log-concave if and only if $f'(x)F(x) - f^2(x) \leq 0$. To

see this, note that

$$(\log F)'(x) = \frac{f}{F}(x), \quad \text{and}$$
$$(\log F)''(x) = \frac{f''}{F} - \frac{f^2}{F^2} = \frac{f'F - f^2}{F^2} \leq 0.$$

Now if f is log-concave we can use fact 1 to write

$$\begin{aligned} \frac{f'}{f}(x)F(x) &= \frac{f'}{f}(x)\int_a^x f(y)dy \\ &\leq \int_a^x \frac{f'(y)}{f(y)}f(y)dy = \int_a^x f'(y)dy \\ &= f(x) - f(a) = f(x). \end{aligned}$$

Rearranging this inequality yields $f'(x)F(x) - f^2(x) \leq 0$, and by Fact 2 we conclude that F is log-concave. Note that this inequality also can be rewritten as $\frac{f'(x)}{f^2(x)}F(x) \leq 1$, and hence we conclude that

$$\frac{f'(x)}{f^2(x)}F(x)(1 - F(x)) \leq 1 - F(x) \leq 1$$

The argument for $1 - F$ is analogous and yields the inequality $\frac{f'(x)}{f^2(x)}(1 - F(x)) \geq -1$, and hence we conclude that

$$\frac{f'(x)}{f^2(x)}F(x)(1 - F(x)) \geq -F(x) \geq -1$$

Thus both F and $1 - F$ are log-concave, and $\gamma(F) \leq 1$. □

Second Proof, general (without assuming f' exists): See the second proof of Proposition 3 below.

Proof of Proposition 3

First Proof, assuming f' exists:

Suppose $s \in (-1, 0)$; the proof for $s > 0$ is similar.

Fact 1-s: First note that f is $s-$concave for $s<0$ if and only if $\varphi \equiv f^s$ is convex on $J(F)$, which is equivalent to φ' being non-decreasing. But we find

$$\varphi'(x) = (f^s)'(x) = sf^{s-1}(x)f'(x) = sf^s(x)(f'(x)/f(x)).$$

Fact 2-s: Note that $F(x) = \int_a^x f(y)dy$ is s^*-concave for $s^*<0$ if and only if

$$(s^*-1)f^2 + Ff' \leq 0 \text{ on } J(F).$$

To see this, note that for $x \in J(F)$

$$\begin{aligned}
(F^{s^*})'(x) &= s^*F^{s^*-1}(x)f(x), \quad \text{and} \\
(F^{s^*})''(x) &= s^*(s^*-1)F^{s^*}(x)\left(\frac{f}{F}(x)\right)^2 + s^*\frac{f'(x)}{F(x)}F^{s^*}(x) \\
&= s^*\frac{F^{s^*}(x)}{F^2(x)}\left\{(s^*-1)f^2(x) + F(x)f'(x)\right\} \\
&\geq 0
\end{aligned}$$

if and only if (since $s^* = s/(1+s) < 0$)

$$(s^*-1)f^2(x) + F(x)f'(x) \leq 0.$$

Now if f is $s-$concave and $x \in J(F)$ we can use fact 1-s to write

$$\begin{aligned}
sf^s(x)\frac{f'}{f}(x)F(x) &= sf^s(x)\frac{f'}{f}(x)\int_a^x f(y)dy \\
&\geq \int_a^x sf^s(y)\frac{f'(y)}{f(y)}f(y)dy = \int_a^x sf^s(y)f'(y)dy \\
&= \frac{s}{s+1}\left(f^{s+1}(x) - f^{s+1}(a)\right) = \frac{s}{s+1}f^{s+1}(x).
\end{aligned}$$

Rearranging this inequality (and noting that $s<0$) yields $(s^*-1)f^2 + Ff' \leq 0$, and by Fact 2-s we conclude that F is s^*-concave. Note that for $x \in J(F)$ this inequality can also be rewritten as $\frac{f'(x)}{f^2(x)}F(x) \leq \frac{1}{1+s}$, and hence we conclude that

$$\frac{f'(x)}{f^2(x)}F(x)(1-F(x)) \leq \frac{1}{1+s}(1-F(x)) \leq \frac{1}{1+s} = 1 - s^*.$$

The argument for $1-F$ is analogous and yields the inequality $\frac{f'(x)}{f^2(x)}(1-F(x)) \geq -\frac{1}{1+s} = -(1-s^*)$, and hence we conclude that

$$\frac{f'(x)}{f^2(x)}F(x)(1-F(x)) \geq -\frac{1}{1+s}F(x) \geq -\frac{1}{1+s}$$

Thus both F and $1-F$ are s^*-concave, and $\gamma(F) \leq 1/(1+s)$. □

Proof of Proposition 3

Second Proof, general (without assuming f' exists): First some background and definitions:

- Let $a, b \geq 0$ and $\theta \in (0,1)$. The generalized mean of order $s \in \mathbb{R}$ is defined by

$$M_s(a,b;\theta) = \begin{cases} ((1-\theta)a^s + \theta b^s)^{1/s}, & \text{if } \pm s \in (0,\infty), \\ a^{1-\theta}b^{\theta}, & \text{if } s = 0, \\ \max\{a,b\}, & \text{if } s = \infty, \\ \min\{a,b\}, & \text{if } s = -\infty. \end{cases}$$

- Let (M,d) be a metric space with Borel $\sigma-$field $\mathcal{M}$. A measure μ on $\mathcal{M}$ is called *$t-$concave* if for nonempty sets $A, B \in \mathcal{M}$ and $0 < \theta < 1$ we have

$$\mu_*((1-\theta)A + \theta B) \geq M_t(\mu_*(A), \mu_*(B); \theta).$$

- A non-negative real-valued function h on (M,d) is called *$s-$concave* if for $x, y \in M$ and $0 < \theta < 1$ we have

$$h((1-\theta)x + \theta y) \geq M_s(h(x), h(y); \theta).$$

- Suppose $(M,d) = (\mathbb{R}^k, |\cdot|)$, $k-$dimensional Euclidean space with the usual Euclidean metric and suppose that f is an $s-$concave density function with respect to Lebesgue measure λ on $\mathcal{B}_k$, and consider the probability measure μ on $\mathcal{B}_k$ defined by

$$\mu(B) = \int_B f d\lambda \quad \text{for all} \quad B \in \mathcal{B}_k.$$

Then by a theorem of Borell (1975), Brascamp and Lieb (1976), and Rinott (1976), the measure μ is s^* concave where $s^* = s/(1+ks)$ if $s \in (-1/k, \infty)$ and $s^* = 0$ if $s = 0$.

- Here we are in the case $k = 1$. Thus for $s \in (-1, \infty)$ the measure μ is s^* concave: for $s \in (-1, \infty)$, $A, B \in \mathcal{B}_1$, and $0 < \theta < 1$,

$$\mu_*((1-\theta)A + \theta B) \geq M_{s^*}(\mu_*(A), \mu_*(B); \theta); \tag{3.2}$$

here μ_* denotes inner measure (which is needed in general in view of examples noted by Erdős and Stone (1970)). With this preparation we can give our second proof of Proposition 3: if $A = (-\infty, x]$ and $B = (-\infty, y]$ for $x, y \in J(F)$, it is easily seen that

$$\begin{aligned}(1-\theta)A + \theta B &= \{(1-\theta)x' + \theta y' : \ x' \leq x, \ y' \leq y\} \\ &\subset \{(1-\theta)x' + \theta y' : \ (1-\theta)x' + \theta y' \leq (1-\theta)x + \theta y\} \\ &= (-\infty, (1-\theta)x + \theta y].\end{aligned}$$

Therefore, with the second inequality following from (3.2)

$$\begin{aligned}F((1-\theta)x + \theta y) &= \mu((-\infty, (1-\theta)x + \theta y]) \\ &\geq \mu((1-\theta)(-\infty, x] + \theta(-\infty, y]) \\ &\geq M_{s^*}(\mu((-\infty, x]), \mu((-\infty, y]); \theta) = M_{s^*}(F(x), F(y); \theta);\end{aligned}$$

i.e. F is s^*-concave. Similarly, taking $A = (x, \infty)$ and $B = (y, \infty)$ it follows that $1 - F$ is s^*-concave.

Note that this argument contains a second proof of Proposition 2 when $s = 0$. □

3.4 Bi-s^*-concave is bigger than $s-$concave

Here we note that just as the class of bi-log-concave distributions is considerably larger than the class of log-concave distributions (as shown by Dümbgen *et al.* (2017)), the class of bi$-s^*-$concave distributions is considerably larger than the class of $s-$concave distributions.

In particular, multimodal distributions are allowed in both the bi-log-concave and the bi-s-concave classes.

Example 7. (Dümbgen *et al.* (2017), pages 2-3) Suppose that f is the mixture $(1/2)N(-\delta, 1)+(1/2)N(\delta, 1)$. Dümbgen *et al.* (2017) showed (numerically) that the corresponding distribution function F is bi-log-concave for $\delta \leq 1.34$ but not for $\delta \geq 1.35$. This distribution has a bi-modal density for $\delta = 1.34$.

Example 8. Now suppose that f is the mixture $(1/2)t_1(\cdot - \delta) + (1/2)t_1(\cdot + \delta)$ with $\delta > 0$ where t_r is the standard t density with r degrees of freedom as in Example 7. By numerical calculation, this density is bi$-s^*-$concave for $\delta = 1.4$, but fails to be bi$-s^*-$concave for $\delta = 1.5$. Again by numerical calculations the t_1 mixture density with $\delta = 1.475$ is bi-$(-1/2)^*$-concave, but with $\delta = 1.48$ it is *not* bi-$(-1/2)^*$-concave; see Figure 3.5.

The following plots illustrate the bounds in Section 3.5.

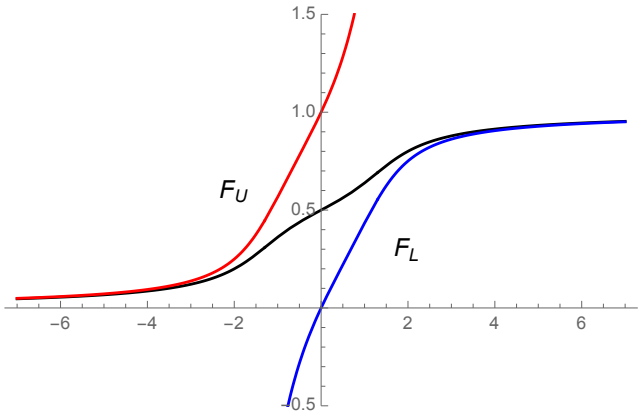


Figure 3.2: The bi$-s^*-$concave t_1 mixture distribution function F (black) for $\delta = 1.3$ with its convex upper bound F_U (red) and concave lower bound F_L (blue) defined by (3.8) and (3.9).

Upper and lower bounds for the density $f = F'$ of F follow from (iii) of Theorem 3.2.

These bounds are illustrated for the bi$-s^*-$concave distribution t_1 mixture with $\delta = 1.3$ in Figure 3.3.

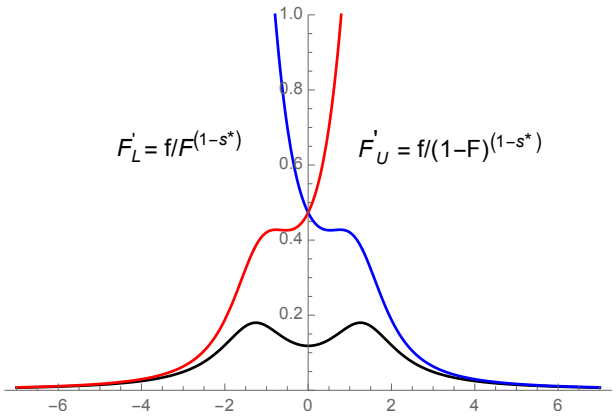


Figure 3.3: The bi-s^*-concave t_1 mixture density function f (black), $\delta = 1.3$, with its bi-s^*-concave upper bounds F'_U (red) and F'_L (blue) defined by (3.11) and (3.10).

To get some feeling for what is happening with the Csörgő - Révész condition, Figure 3.5 gives plots of the two functions

$$\begin{aligned} CR(x) &\equiv F(x)(1-F(x))\frac{f'(x)}{f^2(x)}, \\ CR_{min}(x) &= \min\{F(x), 1-F(x)\}\frac{f'(x)}{f^2(x)}. \end{aligned}$$

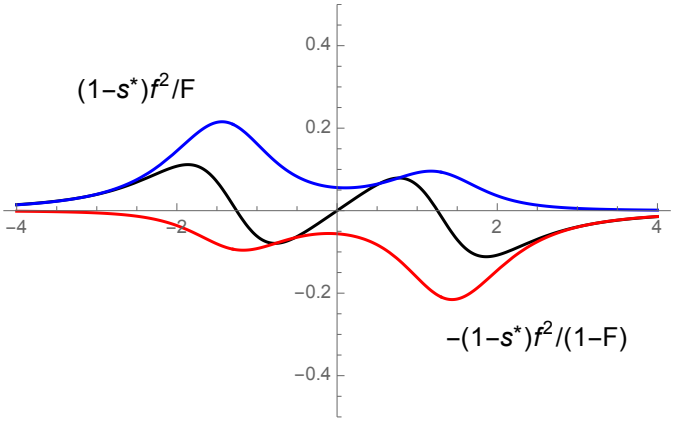


Figure 3.4: The bi-s^*-concave t_1 mixture density function derivative f' (black) for $\delta = 1.3$ with its bi-s^*-concave upper (blue) and lower (red) bounds as given in (iv) of Theorem 3.2.

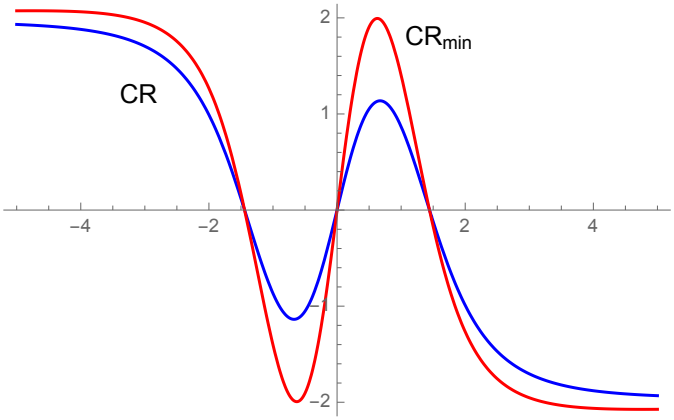


Figure 3.5: The Csörgő-Révész functions CR (blue) and CR_{min} (red) for the mixed t_1 density with $\delta = 1.475$.

3.5 The bi-s^*-concave analogue of Theorem 3.1

3.5.1 Characterization theorem, bi-s^*-concave class

Now we can formulate the natural bi-s^*-concave analogue of Theorem 3.1.

Theorem 3.2. *Let $s \in (-1, \infty]$. For a non-degenerate distribution function F the following four statements are equivalent:*

(i) F is bi-s^*-concave.

(ii) F is continuous on $\mathbb{R}$ and differentiable on $J(F)$ with derivative $f = F'$. Moreover when $s \le 0$,

$$F(x+t) \begin{cases} \le F(x) \cdot \left(1 + s^* \frac{f(x)}{F(x)} t\right)_+^{1/s^*} \\ \ge 1 - (1 - F(x)) \cdot \left(1 - s^* \frac{f(x)}{1-F(x)} t\right)_+^{1/s^*} \end{cases} \tag{3.3}$$

for all $x \in \mathbb{R}$ and $t \in \mathbb{R}$. When $s > 0$,

$$F(x+t) \begin{cases} \le F(x) \cdot \left(1 + s^* \frac{f(x)}{F(x)} t\right)_+^{1/s^*}, & \text{for } t \in (a - x, \infty) \\ \ge 1 - (1 - F(x)) \cdot \left(1 - s^* \frac{f(x)}{1-F(x)} t\right)_+^{1/s^*}, & \text{for } t \in (-\infty, b - x) \end{cases} \tag{3.4}$$

for all $x \in J(F)$.

(iii) F is continuous on $\mathbb{R}$ and differentiable on $J(F)$ with derivative $f = F'$ such that the s^*-hazard function $f/(1-F)^{1-s^*}$ is non-decreasing, and the reverse s^*-hazard function f/F^{1-s^*} is non-increasing on $J(F)$.

(iv) F is continuous on $\mathbb{R}$ and differentiable on $J(F)$ with bounded and strictly positive derivative $f = F'$. Furthermore, f is locally Lipschitz-continuous on $J(F)$ with L^1-derivative $f' = F''$ satisfying

$$-(1 - s^*) \frac{f^2}{1 - F} \le f' \le (1 - s^*) \frac{f^2}{F}. \tag{3.5}$$

Recall that $s^* = s/(1+s) \in (-\infty, 1]$ and $(1 - s^*) = 1/(1+s) \in [0, \infty)$. Alternatively,

$$-\frac{f^2}{1 - F} \le (1 + s) f' \le \frac{f^2}{F}.$$

This yields the following corollary extending (3.1) from $s = 0$ to $s \in (-1, \infty]$.

Corollary 5. *Suppose that F is bi-s^*-concave for $s \in (-1, \infty]$. Then*

$$\gamma(F) = \sup_{x \in J(F)} F(x)(1 - F(x)) \frac{|f'(x)|}{f^2(x)} \quad \leq \quad 1 - s^* = \frac{1}{1+s},$$

and

$$\tilde{\gamma}(F) = \sup_{x \in J(F)} \min\{F(x), 1 - F(x)\} \frac{|f'(x)|}{f^2(x)} \quad \leq \quad 1 - s^* = \frac{1}{1+s}.$$

Remark 2. The three distribution functions F considered by Shorack and Wellner (1986, 2009b) page 644 all involved log-concave densities with the resulting bound for $\gamma(F)$ being 1. Theorem 3.2 and Corollary 5 give a rather complete description of how the values of $\gamma(F)$ and $\tilde{\gamma}(F)$ depend on the index s^* of bi-s^*-concavity.

Proof of Theorem 3.2. If $s = 0$, the proof follows from Theorem 3.1 of Dümbgen *et al.* (2017). When $s = \infty$, $s^* = 1$ and $1 - s^* = 0$. In this case $f' = 0$ almost everywhere (Lebesgue) and f is a uniform density on (a, b). When $s \in (0, \infty)$ the proof is essentially the same as for $s = 0$ with only two minor modifications (in the proof of (i) implies (ii) and in the proof of (iii) implies (iv)); see Section 3.8 for complete details. It remains to consider the case when $s \in (-1, 0)$. Our proof closely parallels the proof for the case $s = 0$ given by Dümbgen *et al.* (2017). Throughout our proof we will denote $\inf J(F)$ and $\sup J(F)$ by a and b respectively. Notice that if F is continuous, $J(F) = (a, b)$.

Proof of (i) implies (ii): Since F is bi-s^*-concave with $s^* < 0$, $\psi = F^{1/s^*}$ is convex on $J(F)$. Since $\psi(x) = 1$ and ∞ for $x \geq \sup J(F)$ and $x \leq \inf J(F)$ respectively, ψ is convex on $\mathbb{R}$. By the convex version of Lemma 6 of Dümbgen *et al.* (2017) ψ is continuous on the interior of $\{\psi < \infty\}$. Therefore ψ and hence F is continuous on the interior of the set $\{F > 0\}$ or (a, ∞). Similarly, the s^*-concavity of $1 - F$ implies continuity of $1 - F$ on the interior of the set $\{1 - F > 0\} := (-\infty, b)$ where $b := \sup\{F < 1\}$. However unless $a < b$, F would be degenerate. Hence, $a < b$ and F is continuous on $\mathbb{R}$. More precisely $J(F) = (a, b)$.

Let $x \in (a, b)$. Convexity of ψ implies that

$$F'(x\pm) = \lim_{t \to 0, \pm t > 0} \frac{\psi^{1/s^*}(x+t) - \psi^{1/s^*}(t)}{t} = \frac{1}{s^*} \psi(x)^{1/s^* - 1} \psi'(x\pm)$$

exist and satisfy

$$F'(x-) \leq F'(x+).$$

Similarly, convexity of $(1-F)^{s^*}$ yields

$$(1-F)'(x+) \geq (1-F)'(x-)$$

which implies that

$$-F'(x+) \geq -F'(x-).$$

Therefore $F'(x-) = F'(x+)$ which proves the differentiability of F. It also shows that $\psi'(x+) = \psi'(x-) = \psi'(x)$ on (a, b).

By Lemma 6 (convex version) of Dümbgen *et al.* (2017) for each $x \in (a, b)$ and $c \in [\psi'(x-), \psi'(x+)]$ one has

$$\psi(x+t) - \psi(x) \geq ct \quad \text{for all } t \in \mathbb{R}.$$

Therefore

$$\psi(x+t) - \psi(x) \geq t\psi'(x).$$

Hence,

$$F^{s^*}(x+t) - F^{s^*}(x) \geq t s^* f(x) F(x)^{s^* - 1},$$

or, with $x_+ = \max\{x, 0\}$,

$$\frac{F^{s^*}(x+t)}{F^{s^*}(x)} \geq \left(1 + s^* \frac{f(x)}{F(x)} t\right)_+.$$

Hence,

$$\frac{F(x+t)}{F(x)} \leq \left(1 + s^* \frac{f(x)}{F(x)} t\right)_+^{1/s^*}.$$

Analogously it follows that

$$(1 - F(x+t))^{s^*} - (1 - F(x))^{s^*} \geq -ts^* f(x)(1 - F(x))^{s^*-1}$$

which yields

$$\left(\frac{1 - F(x+t)}{1 - F(x)}\right)^{s^*} \geq \left(1 - ts^* \frac{f(x)}{1 - F(x)}\right)_+$$

or

$$F(x+t) \geq 1 - (1 - F(x)) \cdot \left(1 - ts^* \frac{f(x)}{1 - F(x)}\right)_+^{1/s^*}.$$

Hence (3.3) is proved.

Since (ii) holds, F is continuous and differentiable on $J(F)$ with derivative $f = F'$ and satisfies (3.3). Now let $x, y \in J(F)$ with $x < y$. Let

$$h = f/F^{1-s^*}. \tag{3.6}$$

Then applying (3.3) we obtain that

$$\frac{F^{s^*}(x)}{F^{s^*}(y)} \geq 1 + s^* \frac{f(y)}{F(y)}(x - y).$$

Hence,

$$\begin{aligned} F^{s^*}(x) &\geq F^{s^*}(y) + s^* \frac{f(y)}{F(y)^{1-s^*}}(x - y) \\ &= F^{s^*}(y) + s^* h(y)(x - y) \\ &\geq F^{s^*}(x) + s^* h(x)(y - x) + s^* h(y)(x - y). \end{aligned}$$

Therefore

$$s^*(x - y)(h(y) - h(x)) \leq 0$$

where $s^*(x-y) > 0$, implying that $h(y) \leq h(x)$. Therefore h is non-increasing. Now let

$$\tilde{h} = f/(1-F)^{1-s^*}. \tag{3.7}$$

From (3.3) we also obtain that

$$(1-F(x))^{s^*} - (1-F(y))^{s^*} \geq -ts^* \frac{f(y)}{(1-F(y))^{1-s^*}} = -ts^* \tilde{h}(y)$$

or

$$\begin{aligned}
(1-F(x))^{s^*} &\geq (1-F(y))^{s^*} - (x-y)s^*\tilde{h}(y) \\
&= (1-F(x))^{s^*} - (y-x)s^*\tilde{h}(x) - (x-y)s^*\tilde{h}(y) \\
&= (1-F(x))^{s^*} - s^*(y-x)(\tilde{h}(y) - \tilde{h}(x)).
\end{aligned}$$

Since $s^*(y-x) < 0$, the last inequality leads to

$$0 \leq \tilde{h}(y) - \tilde{h}(x),$$

implying that $\tilde{h}$ is non-decreasing.

Proof of (iii) implies (iv): If the conditions of (iii) hold, then it immediately follows that $f > 0$ on $J(F)$. If not, suppose that $f(x_0) = 0$ for some $x_0 \in J(F)$. Now $J(F) = (a, b)$ since F is continuous. Since $f(x)/F(x)^{1-s^*}$ is non-increasing, $f(x) = 0$ for $x \in [x_0, b)$. Similarly since $f(x)/(1-F(x))^{1-s^*}$ is non-decreasing we obtain $f(x) = 0$ for $x \in (a, x_0]$. Therefore, $F' = 0$ or F is constant on $J(F)$. Then F violates the continuity condition of (iii). Hence $f > 0$ on $J(F)$.

Suppose h and $\tilde{h}$ are as defined in (3.6) and (3.7). Then the monotonicities of h and $\tilde{h}$ imply that for any $x, x_0 \in J(F)$,

$$f(x) = \begin{cases} F^{1-s^*}(x)h(x) \leq h(x_0) & \text{if } \ x \geq x_0, \\ (1-F(x))^{1-s^*}\tilde{h}(x) \leq \tilde{h}(x_0) & \text{if } \ x \leq x_0. \end{cases}$$

Next, let $c, d \in J(F)$ with $c < d$. We will bound $(f(y) - f(x))/(y - x)$ for $x, y \in J(F)$ such that $x, y \in (c, d)$ with $x \neq y$. This will yield local Lipschitz-continuity of f on $J(F)$. To this end, note that

$$\begin{aligned}
\frac{f(y) - f(x)}{y - x} &= \frac{F^{1-s^*}(y)h(y) - F^{1-s^*}(x)h(x)}{y - x} \\
&= h(y)\frac{F^{1-s^*}(y) - F^{1-s^*}(x)}{y - x} + F^{1-s^*}(x)\frac{h(y) - h(x)}{y - x} \\
&\leq h(c)\frac{F^{1-s^*}(y) - F^{1-s^*}(x)}{y - x} \\
&\to h(c)(1 - s^*)f(x)F^{-s^*}(x) = (1 - s^*)h(c)h(x)F^{1-2s^*}(x)
\end{aligned}$$

as $y \to x$. Here the inequality followed from the fact that

$$\frac{h(y) - h(x)}{y - x} \leq 0$$

which holds since h is non-increasing. Now since $h(x) \leq h(c)$, $1 - 2s^* > 0$, $1 - s^* > 0$, and $F(x) \leq F(d)$, we find that

$$\limsup_{y \to x} \frac{f(y) - f(x)}{y - x} \leq (1 - s^*)h(c)^2 F^{1-2s^*}(d)$$

for all $x \in (c, d)$. Analogously with $\bar{F} = 1 - F$ we obtain that

$$\begin{aligned}
\frac{f(y) - f(x)}{y - x} &= \frac{\bar{F}^{1-s^*}(y)\tilde{h}(y) - \bar{F}^{1-s^*}(x)\tilde{h}(x)}{y - x} \\
&= \tilde{h}(y)\frac{\bar{F}^{1-s^*}(y) - \bar{F}^{1-s^*}(x)}{y - x} + \bar{F}^{1-s^*}(x)\frac{\tilde{h}(y) - \tilde{h}(x)}{y - x} \\
&\geq \tilde{h}(y)\frac{\bar{F}^{1-s^*}(y) - \bar{F}^{1-s^*}(x)}{y - x}
\end{aligned}$$

since, by the non-decreasing property of $\tilde{h}$, for any $x, y \in J(F)$,

$$\frac{\tilde{h}(y) - \tilde{h}(x)}{y - x} > 0.$$

Next observe that since $1 - s^* = 1/(1 + s) > 0$, and $\bar{F}$ is nonincreasing,

$$\tilde{h}(y)\frac{\bar{F}^{1-s^*}(y) - \bar{F}^{1-s^*}(x)}{y - x} \geq \tilde{h}(d)\frac{\bar{F}^{1-s^*}(y) - \bar{F}^{1-s^*}(x)}{y - x}.$$

Hence as $y \to x$ it follows that

$$\begin{aligned}\liminf_{y\to x} \frac{f(y)-f(x)}{y-x} &\geq -\tilde{h}(d)(1-s^*)f(x)\bar{F}^{-s^*}(x) \\ &= -\tilde{h}(d)\tilde{h}(x)(1-s^*)\bar{F}^{1-2s^*}(x).\end{aligned}$$

Therefore using the fact that $\tilde{h}(x) \leq \tilde{h}(d)$ and $1-2s^* > 0$ we conclude that

$$\liminf_{y\to x} \frac{f(y)-f(x)}{y-x} \geq -\tilde{h}(d)^2(1-s^*)\bar{F}^{1-2s^*}(c).$$

Combining the above with (3.8) we find that f is Lipschitz-continuous on (c,d) with Lipschitz-constant

$$\max\{(1-s^*)h(c)^2F^{1-2s^*}(d), (1-s^*)\tilde{h}(d)^2\bar{F}^{1-2s^*}(c)\}.$$

This proves that f is locally Lipschitz continuous on $J(F)$. Hence, f is also locally absolutely continuous with L^1-derivative f' such that

$$f(y)-f(x) = \int_x^y f'(t)dt \quad \text{for all } x,y \in J(F);$$

hence $f'(x)$ can be chosen so that

$$f'(x) \in \left[\liminf_{y\to x} \frac{f(y)-f(x)}{y-x}, \limsup_{y\to x} \frac{f(y)-f(x)}{y-x}\right].$$

However (3.8) and (3.8) imply that for $c < x < d$,

$$\begin{aligned}&\left[\liminf_{y\to x} \frac{f(y)-f(x)}{y-x}, \limsup_{y\to x} \frac{f(y)-f(x)}{y-x}\right] \\ &\quad \subset \left[-(1-s^*)\tilde{h}(d)^2\bar{F}^{1-2s^*}(c), (1-s^*)h(c)^2F^{1-2s^*}(d)\right]\end{aligned}$$

Now since f and F are continuous and $F > 0$ on $J(F)$, so are h and $\tilde{h}$. Therefore, letting $c,d \to x$ it follows that

$$\frac{-(1-s^*)f(x)^2\bar{F}^{1-2s^*}(x)}{\bar{F}^{2-2s^*}(x)} \leq f'(x) \leq (1-s^*)\frac{f(x)^2F^{1-2s^*}(x)}{F^{2-2s^*}(x)};$$

and this implies (3.5).

Proof of (iv) implies (i): The fact that (iii) implies (i) can be easily proved since f/F^{1-s^*} non-increasing on $J(F)$ implies that F^{s^*} is convex on $J(F)$. Also $1 < F^{s^*} < \infty$ on $J(F)$. Now $F^{s^*}(x) = \infty$ for $x < \inf J(F)$ and $F^{s^*}(x) = 1$ for $x > \sup J(F)$. Therefore F^{s^*} is convex on $\mathbb{R}$. Similarly one can show that $(1-F)^{s^*}$ is convex on $\mathbb{R}$. Hence F is bi-s^*-concave. Therefore it is enough to prove that (iv) implies (iii).

By Lemma 7 of Dümbgen *et al.* (2017) h is non-increasing on $J(F)$ if and only if for any $x \in J(F)$ the following holds:

$$\limsup_{y \to x} \frac{h(y) - h(x)}{y - x} \leq 0.$$

Suppose $x \neq y \in J(F)$ and $r := \min(x, y)$ and $s := \max(x, y)$. Then it follows that

$$\begin{aligned}
\frac{h(y) - h(x)}{y - x} &= \frac{f(y)/F^{1-s^*}(y) - f(x)/F^{1-s^*}(x)}{y - x} \\
&= \frac{1}{F^{1-s^*}(y)} \frac{f(y) - f(x)}{y - x} - \frac{f(x)}{F^{1-s^*}(x)F^{1-s^*}(y)} \frac{F^{1-s^*}(y) - F^{1-s^*}(x)}{y - x} \\
&= \frac{1}{F^{1-s^*}(y)} \frac{\int_r^s f'(t)dt}{s - r} - \frac{f(x)}{F^{1-s^*}(x)F^{1-s^*}(y)} \frac{F^{1-s^*}(y) - F^{1-s^*}(x)}{y - x} \\
&\leq \frac{(1-s^*)}{F^{1-s^*}(y)(s-r)} \int_r^s \frac{f(t)^2}{F(t)} dt - \frac{f(x)}{F^{1-s^*}(x)F^{1-s^*}(y)} \frac{F^{1-s^*}(y) - F^{1-s^*}(x)}{y - x}.
\end{aligned}$$

by (3.5). Since F is continuous by (iv), $J(F)$ must be an interval. Also since $x, y \in J(F)$, $[r, s] \subset J(F)$. Since f and F are continuous on $J(F)$ and $F > 0$ on $J(F)$, f^2/F is continuous and integrable on $J(F)$ and hence also on $[r, s]$. Letting $y \to x$ we obtain that

$$\limsup_{y \to x} \frac{h(y) - h(x)}{y - x} \leq \frac{(1-s^*)f(x)^2}{F^{2-s^*}(x)} - \frac{(1-s^*)f(x)^2}{F^{2-s^*}(x)} = 0.$$

Analogously by Lemma 7 of Dümbgen *et al.* (2017), to show $\tilde{h}$ is non-decreasing it is enough to show that

$$\liminf_{y \to x} \frac{\tilde{h}(y) - \tilde{h}(x)}{y - x} \geq 0.$$

To verify this suppose $x \neq y \in J(F)$ and $r := \min(x,y)$ and $s := \max(x,y)$. As before we calculate

$$\begin{aligned}
\frac{\tilde{h}(y)-\tilde{h}(x)}{y-x} &= \frac{f(y)/\bar{F}^{1-s^*}(y) - f(x)/\bar{F}^{1-s^*}(x)}{y-x} \\
&= \frac{1}{F^{1-s^*}(y)}\frac{f(y)-f(x)}{y-x} - \frac{f(x)}{\bar{F}^{1-s^*}(x)\bar{F}^{1-s^*}(y)}\frac{\bar{F}^{1-s^*}(y)-\bar{F}^{1-s^*}(x)}{y-x} \\
&= \frac{1}{\bar{F}^{1-s^*}(y)}\frac{\int_r^s f'(t)dt}{s-r} - \frac{f(x)}{\bar{F}^{1-s^*}(x)\bar{F}^{1-s^*}(y)}\frac{\bar{F}^{1-s^*}(y)-\bar{F}^{1-s^*}(x)}{y-x} \\
&\geq -\frac{(1-s^*)}{\bar{F}^{1-s^*}(y)(s-r)}\int_r^s \frac{f(t)^2}{\bar{F}(t)}dt - \frac{f(x)}{\bar{F}^{1-s^*}(x)\bar{F}^{1-s^*}(y)}\frac{\bar{F}^{1-s^*}(y)-\bar{F}^{1-s^*}(x)}{y-x}
\end{aligned}$$

by (3.5). Since f and $\bar{F}$ are continuous on $J(F)$, letting $y \to x$ it follows that

$$\liminf_{y\to x} \frac{\tilde{h}(y)-\tilde{h}(x)}{y-x} \geq -\frac{(1-s^*)f(x)^2}{\bar{F}^{2-s^*}(x)} + \frac{(1-s^*)f(x)^2}{\bar{F}^{2-s^*}(x)} = 0.$$

□

3.5.2 Bounds for F bi-s^*-concave when $s < 0$.

First, upper and lower bounds on F: Note that $(1+y)^r \geq 1+ry$ for any $r < 0$ and $y \geq -1$. Taking $y = -F(x)$ and $r = s^*$ yields

$$(1-F(x))^{s^*} \geq 1 - s^*F(x)$$

or, by rearranging,

$$F(x) \leq \frac{1}{-s^*}\left\{(1-F(x))^{s^*} - 1\right\} \equiv F_{U,s}(x) \equiv F_U(x) \tag{3.8}$$

where F_U is a convex function if F is bi$-s^*-$concave. Similarly, taking $y = -(1-F(x))$ and $r = s^*$ yields, by rearranging terms

$$F(x) \geq \frac{1}{-s^*}\left\{(1-s^*) - F(x)^{s^*}\right\} \equiv F_{L,s}(x) \equiv F_L(x) \tag{3.9}$$

where F_L is a concave function if F is bi$-s^*-$concave. Note that

$$F_U'(x) = \frac{f(x)}{(1-F(x))^{1-s^*}} = \frac{f(x)}{(1-F(x))^{1/(1+s)}} \tag{3.10}$$

is monotone non-decreasing, while

$$F_L'(x) = \frac{f(x)}{F^{1-s^*}(x)} = \frac{f(x)}{F^{1/(1+s)}(x)} \tag{3.11}$$

is monotone non-increasing. Therefore

$$\begin{aligned} 0 \le F_U''(x) &= (1-F(x))^{s^*-2}\left\{(1-s^*)f^2(x) + (1-F(x))f'(x)\right\}, \\ 0 \ge F_L''(x) &= F(x)^{s^*-2}\left\{(s^*-1)f^2(x) + F(x)f'(x)\right\}. \end{aligned}$$

The upper and lower bounds in (iv) of Theorem 3.2 follow by rearranging these inequalities.

Taking F to be the distribution function of t_1 and plotting the bounds for F, $F' = f$ and $F'' = f'$ yields the following three figures.

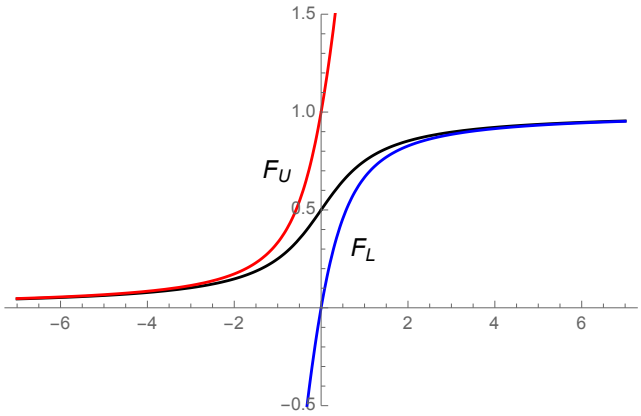


Figure 3.6: The bi$-s^*-$concave t_1 distribution function F (black) with its convex upper bound F_U (red) and concave lower bound F_L (blue), where F_U and F_L are given in (3.8) and (3.9).

Upper bounds for the density $f = F'$ of F follow from (iii) of Theorem 3.2: These bounds are illustrated for the bi-s^*-concave distribution t_1 in Figure 3.7.

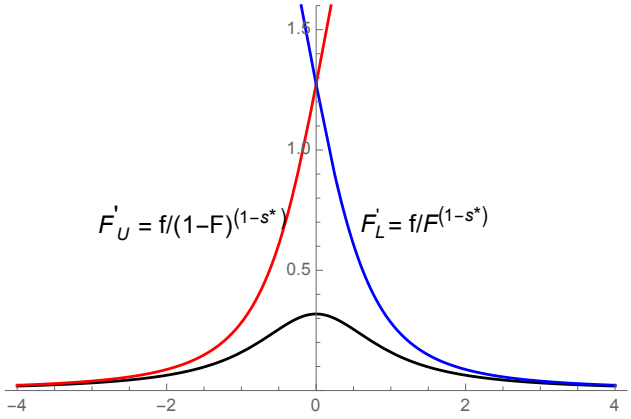


Figure 3.7: The bi-s^*-concave t_1 density function f (black) with its bi-s^*-concave upper bounds F_U' (red) and F_L' (blue) as given by (3.10) and (3.11).

Upper and lower bounds for the derivative f' of f are given in (iv) of Theorem 3.2: These bounds are illustrated for the bi$-s^*-$concave distribution t_1 in Figure 3.8.

3.5.3 Bounds for F bi-s^*-concave when $s > 0$.

Upper and lower bounds on F: Note that now $(1+y)^r \leq 1+ry$ for any $r \in (0,1]$ and $y \geq -1$ by concavity of $(1+y)^r$. Taking $y = -F(x)$ and $r = s^* > 0$ (since $s > 0$) yields

$$(1 - F(x))^{s^*} \leq 1 - s^* F(x).$$

By rearranging,

$$\begin{aligned} F(x) &\leq \frac{1}{-s^*}\left\{(1-F(x))^{s*} - 1\right\} \\ &= \frac{1}{s^*}\left\{1 - (1-F(x))^{s^*}\right\} \equiv F_{U,s}(x) \equiv F_U(x) \end{aligned} \tag{3.12}$$

where F_U is a convex function if F is bi$-s^*-$concave. Similarly, taking $y = -(1-F(x))$ and $r = s^*$ yields, by rearranging terms

$$F(x) \geq \frac{1}{s^*}\left\{F(x)^{s^*} - (1-s^*)\right\} \equiv F_{L,s}(x) \equiv F_L(x) \tag{3.13}$$

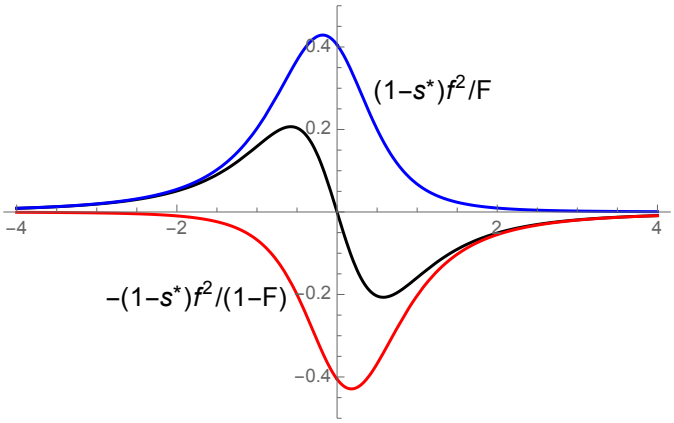


Figure 3.8: The bi-s^*-concave t_1 density function derivative f' (black) with its bi-s^*-concave lower (red) and upper (blue) bounds as given in (iv) of Theorem 3.2.

where F_L is a concave function if F is bi$-s^*-$concave. Note that

$$F'_U(x) = \frac{f(x)}{(1-F(x))^{1-s^*}} = \frac{f(x)}{(1-F(x))^{1/(1+s)}} \tag{3.14}$$

is monotone non-decreasing, while

$$F'_L(x) = \frac{f(x)}{F^{1-s^*}(x)} = \frac{f(x)}{F^{1/(1+s)}(x)} \tag{3.15}$$

is monotone non-increasing. Therefore

$$\begin{aligned} 0 \leq F''_U(x) &= (1-F(x))^{s^*-2}\left\{(1-s^*)f^2(x) + (1-F(x))f'(x)\right\}, \\ 0 \geq F''_L(x) &= F(x)^{s^*-2}\left\{(s^*-1)f^2(x) + F(x)f'(x)\right\}. \end{aligned}$$

Again note that the upper and lower bounds in (iv) of Theorem 3.2 follow by rearranging these inequalities.

Taking F to be the distribution function of $g(\cdot, r)$ with $r = 1$ as in Example 6 and plotting the bounds for F, $F' = f$ and $F'' = f'$ yields the following three figures.

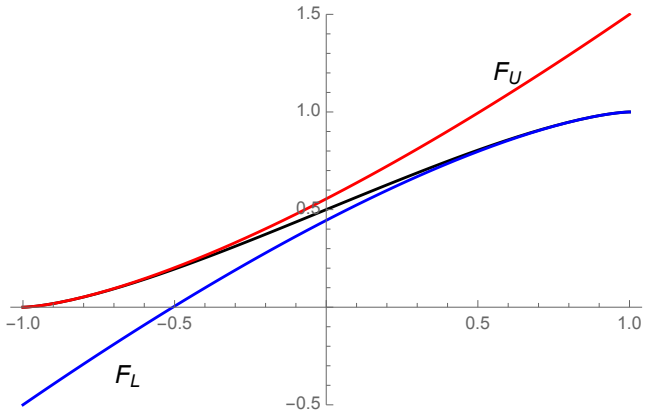


Figure 3.9: The bi-s^*-concave distribution function F (black) corresponding to $g(\cdot;1)$ of Example 6 with its convex upper bound F_U (red) and concave lower bound F_L (blue) (where F_U and F_L are given in (3.12) and (3.13)).

Upper and lower bounds for the density $f = F'$ of F follow from (iii) of Theorem 3.2. These bounds are illustrated for the bi-s^*-concave distribution F corresponding to $g(\cdot;1)$ of Example 6 in Figure 3.10.

Upper and lower bounds for the derivative f' of f are given in (iv) of Theorem 3.2 These bounds are illustrated for the bi-s^*-concave distribution function F with density $g(\cdot;1)$ as in Example 6 in Figure 3.11.

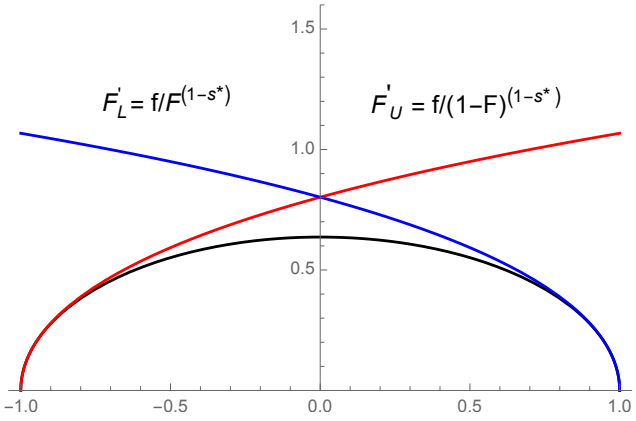


Figure 3.10: The bi-s^*-concave density function $g(\cdot;1)$ of Example 6 (black) with its bi-s^*-concave upper bounds F_L' and F_U' given in (3.15) and (3.14).

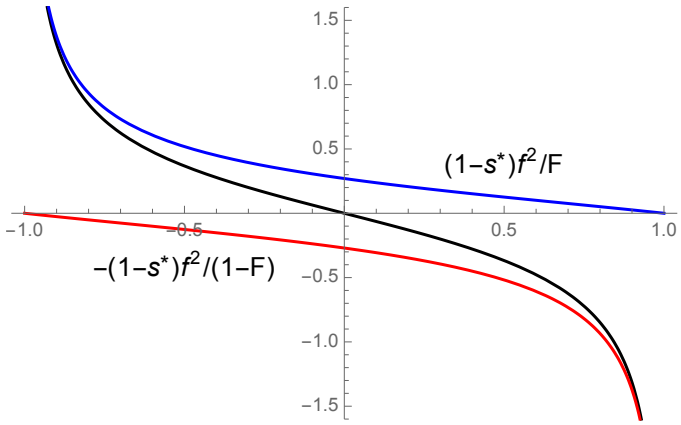


Figure 3.11: $F'' = f'$ (black) for the bi-s^*-concave function F corresponding to the density $g(\cdot;1)$ as in Example 6 with its bi-s^*-concave upper (blue) and lower (red) bounds as given in (iv) of Theorem 3.2.

3.6 *A consequence for Fisher information*

In this section we suppose that F is a bi-s^*-concave distribution function with absolutely continuous density f with respect to Lebesgue measure. Then from (3.5) of Theorem 3.2 it follows that

$$\frac{|f'(x)|}{f(x)} \leq \frac{1}{1+s}\frac{f(x)}{F(x) \wedge (1-F(x))} \quad \text{for all} \quad x \in J(F),$$

and hence that

$$\begin{aligned} I_f &\equiv \int_{\mathbb{R}} \left(\frac{|f'(x)|}{f(x)}\right)^2 f(x)dx \\ &\leq \frac{1}{(1+s)^2}\int_{\mathbb{R}} \frac{f^2(x)}{(F(x) \wedge (1-F(x)))^2} dF(x) \\ &\leq \frac{1}{(1+s)^2}\left\{\int_{\mathbb{R}} \frac{f^2(x)}{F^2(x)} dF(x) + \int_{\mathbb{R}} \frac{f^2(x)}{(1-F(x))^2} dF(x)\right\} \\ &\leq \frac{2}{(1+s)^2} \max\left\{\int_{\mathbb{R}} \left(\frac{f}{F}\right)^2 dF, \int_{\mathbb{R}} \left(\frac{f}{1-F}\right)^2 dF\right\}. \end{aligned} \tag{3.16}$$

But with $h = f'/f$, we find that

$$\int_{-\infty}^x h dF = \int_{-\infty}^x (f'(y)/f(y))f(y)dy = f(x) \quad \text{and} \quad \frac{f(x)}{F(x)} = \frac{\int_{-\infty}^x h dF}{F(x)},$$

while

$$\int_x^{\infty} h dF = \int_x^{\infty} (f'/f)f dy = -f(x), \quad \text{and} \quad \frac{-f(x)}{1-F(x)} = \frac{\int_x^{\infty} h dF}{1-F(x)}.$$

Thus by the L_2 version of Hardy's inequality

$$\begin{aligned} \int_{\mathbb{R}} \left(\frac{f(x)}{F(x)}\right)^2 dF(x) &\leq 4\int_{\mathbb{R}} \left(\frac{|f'(x)|}{f(x)}\right)^2 f(x)dx = 4I_f, \quad \text{and} \\ \int_{\mathbb{R}} \left(\frac{f(x)}{1-F(x)}\right)^2 dF(x) &\leq 4\int_{\mathbb{R}} \left(\frac{|f'(x)|}{f(x)}\right)^2 f(x)dx = 4I_f. \end{aligned} \tag{3.17}$$

Combining the inequalities in (3.16) and (3.17) yields

$$\begin{aligned} I_f &\leq \frac{2}{(1+s)^2} \max\left\{ \int_{\mathbb{R}} \left(\frac{f(x)}{F(x)}\right)^2 dF(x),\ \int_{\mathbb{R}} \left(\frac{f(x)}{1-F(x)}\right)^2 dF(x) \right\} \\ &\leq \frac{8}{(1+s)^2} I_f. \end{aligned} \tag{3.18}$$

But we note that the densities f_r in Example 4 have

$$I_{f_r} = \frac{r}{2} \cdot \frac{\Gamma\left(\frac{r}{2}-1\right)\Gamma\left(\frac{r+3}{2}\right)}{\Gamma\left(\frac{r}{2}+1\right)^2} \nearrow \infty$$

as $r \searrow 2$, and $I_{f_r} = \infty$ for $0 < r \leq 2$. In this latter case all the integrals in (3.18) are infinite.

3.7 Confidence Bands for $F \in \mathcal{F}_{b-s^*}$

Suppose that (L_n, U_n) is a $1-\alpha$ confidence band for an arbitrary distribution function F; that is,

$$P_F(L_n \leq F \leq U_n) = 1 - \alpha.$$

If we may assume, in addition, that $F \in \mathcal{F}_{bi-s^*}$, then we can refine this initial confidence band as follows: let

$$\begin{aligned} L_n^*(x) &\equiv \inf\{G(x) :\ G \in \mathcal{F}_{bi-s^*},\ L_n \leq G \leq U_n\}, \\ U_n^*(x) &\equiv \sup\{G(x) :\ G \in \mathcal{F}_{bi-s^*},\ L_n \leq G \leq U_n\}. \end{aligned}$$

If no bi-s*-concave distribution function F is contained in the band (L_n, U_n), then set $L_n^* \equiv 1$ and $U_n^* \equiv 0$ and conclude that with confidence $1-\alpha$ $F \notin \mathcal{F}_{bi-s^*}$. But if $F \in \mathcal{F}_{bi-s^*}$ this happens with probability at most α since

$$P(L_n^* \leq F \leq U_n^*) = P(L_n \leq F \leq U_n) \quad \text{if} \quad F \in \mathcal{F}_{bi-s^*}.$$

Now we use the procedure ConcInt$(\cdot,\cdot)$ developed in Dümbgen *et al.* (2017). As explained by DKW (2017), given any finite set $\mathcal{T} = \{t_0, \ldots, t_m\}$ of real numbers $t_0 < t_1 < \cdots < t_m$

and any pair (l, u) of functions $l, u : \mathcal{T} \to [-\infty, \infty)$ with $l < u$ pointwise and $l(t) > -\infty$ for at least two different points $t \in \mathcal{T}$, this procedure computes the pair (l^0, u^0) where

$$\begin{aligned} l^0(x) &\equiv \inf\{g(x) : \ g \text{ concave on } \mathbb{R}, \ l \le g \le u \ \text{ on } \mathcal{T}\}, \\ u^0(x) &\equiv \sup\{g(x) : \ g \text{ concave on } \mathbb{R}, \ l \le g \le u \ \text{ on } \mathcal{T}\}. \end{aligned}$$

First note that l^0 is the smallest concave majorant of l on $\mathcal{T}$; thus it may be computed by a version of the pool-adjacent - violators algorithm; see for example Robertson et al. (1988), page 8ff. [Use *lcm(·)* in fdrtool R)[1].]

Then we obtain indices $0 \le j(0) < j(1) < \cdots < j(b) \le m$ such that

$$l^0 \begin{cases} \equiv -\infty \ \text{ on } \mathbb{R} \setminus [t_{j(0)}, t_{j(b)}], \\ \text{is linear on } \ [t_{j(a-1)}, t_{j(a)}] \ \text{ for } 1 \le a \le b, \\ \text{changes slope at } t_{j(a)} \ \text{ if } 1 \le a < b. \end{cases}$$

With l^0 in hand, we then check to see if $l^0 \le u$ on $\mathcal{T}$. If this fails, then there is no concave function lying between l and u, and the procedure returns an error message. If this test succeeds, then we compute $u^0(x)$ as

$$\min \left\{ u(s) + \frac{u(s) - l^0(r)}{s - r}(x - s) : \ r \in \mathcal{T}_0, \ r < s \le x \text{ or } x \le s < r \right\}$$

where $\mathcal{T}_0 = \{t_{j(1)}, t_{j(2)}, \ldots, t_{j(b)}\}$. [The rest of the description of the procedure ConcInt$(\cdot, \cdot)$ is just as in DKW (2017).]

When $s < 0$ and hence $s^* < 0$, let $g(v; s^*) \equiv g(v) \equiv -v^{s*}$ and $h(v; s^*) \equiv h(v) \equiv (-v)^{1/s^*}$. [This is the most important new case. When $s = s^* = 0$, $g(v) \equiv \log(v)$, $h(v) \equiv \exp(v)$. When $s > 0$ and hence $s^* > 0$, $g(v) = v^{s^*}$ and $h(v) = v^{1/s^*}$.] Here is pseudocode for the

[1] https://rdrr.io/cran/fdrtool/man/gcmlcm.html

computation of (L_n^*, U_n^*).

$$
\begin{aligned}
&(L_n^*, U_n^*) \leftarrow (L_n, U_n)\\
&(l^0, u^0) \leftarrow ConcInt(g(L_n^*), g(U_n^*)\\
&(\tilde{L}_n^*, \tilde{U}_n^*) \leftarrow (h(l^0), h(u^0))\\
&(l^0, u^0) \leftarrow ConcInt(g(1-U_n^*), g(1-L_n^*))\\
&(\tilde{L}_n^*, \tilde{U}_n^*) \leftarrow (1-h(u^0), 1-h(l^0))\\
&\text{while } (\tilde{L}_n^*, \tilde{U}_n^*) \neq (L_n^*, U_n^*) \text{ do}\\
&\quad (L_n^*, U_n^*) \leftarrow (\tilde{L}_n, \tilde{U}_n)\\
&\quad (l^0, u^0) \leftarrow ConcInt(g(L_n^*), g(U_n^*)\\
&\quad (\tilde{L}_n^*, \tilde{U}_n^*) \leftarrow (h(l^0), h(u^0))\\
&\quad (l^0, u^0) \leftarrow ConcInt(g(1-U_n^*), g(1-L_n^*))\\
&\quad (\tilde{L}_n^*, \tilde{U}_n^*) \leftarrow (1-h(u^0), 1-h(l^0))\\
&\text{end while}
\end{aligned}
$$

In order to establish some properties of our new confidence bands, we first need the following lemma:

Lemma 2*. Let $s \in (0,1)$ or $s > 0$. For real numbers $a < b$ and $0 < u < v < 1$ define

$$\gamma_1 = \frac{v^{s^*} - u^{s^*}}{s^*(b-a)}, \qquad \gamma_2 = \frac{[(1-u)^{s^*} - (1-v)^{s^*}]}{s^*(b-a)}.$$

(i) If $L_n(a) \geq u$ and $U_n(b) \leq v$, then L_n^* and U_n^* are Lipschitz-continuous on $\mathbb{R}$ with Lipschitz constant $\max\{\gamma_1, \gamma_2\}$.

(ii) If $U_n(a) \leq u$ and $L_n(b) \geq v$, then

$$\begin{aligned} U_n^*(x) &\leq u\Big(1 + s^*\gamma_1 u^{-s^*}(x-a)\Big)^{1/s^*} \qquad \text{for } \; x \leq a, \quad \text{and} \\ 1 - L_n^*(x) &\leq (1-v)\Big(1 - s^*\gamma_2(1-v)^{-s^*}(x-b)\Big)^{1/s^*} \qquad \text{for } \; x \geq b. \end{aligned}$$

Proof. If $L_n^{s^*} \equiv 1$ and $U_n^{s^*} \equiv 0$, there is no $G \in \mathcal{F}_{b-s^*}$ lying between L_n and U_n. In this case the statements of Lemma 2^{s^*} are trivially true. Otherwise let the set $\mathcal{C}_{LU} = \{G \in |L_n \leq G \leq U_n\}$ be non-empty.

Proof of (i). Let $G \in \mathcal{C}_{LU}$ with density g. Notice that to prove the Lipschitz-continuity of L_n^* and U_n^*, it is sufficient to show that

$$g(x) \leq \max\{\gamma_1, \gamma_2\}$$

for any G and $x \in J(G)$. Now we will first consider the case when $s < 0$ or G^{s^*} is convex. Let $x \geq b$. Since $-s^* > 0$, concavity of $-G^{s^*}$ leads to

$$-s^* g(x) \leq -s^* G^{s^*-1}(x) g(x) \leq -s^* G^{s^*-1}(b) g(b) \leq \frac{G^{s^*}(a) - G^{s^*}(b)}{b-a}.$$

Taking s^* to the other side we obtain,

$$g(x) \leq \frac{G^{s^*}(b) - G^{s^*}(a)}{s^*(b-a)} \leq \frac{v^{s^*} - u^{s^*}}{s^*(b-a)} = \gamma_1.$$

since $G(a) \geq L_n(a) \geq u$ and $G(b) \leq U_n(b) \leq v$. For $x \leq a$ convexity of $(1-G)^{s^*}$ leads to

$$-s^* g(x) \leq -s^*(1-G(x))^{s^*-1} g(x) \leq -s^*(1-G(a))^{s^*-1} g(a) \leq \frac{(1-G(b))^{s^*} - (1-G(a))^{s^*}}{b-a}.$$

Therefore,

$$g(x) \leq \frac{(1-G(a))^{s^*} - (1-G(b))^{s^*}}{s^*(b-a)} \leq \frac{(1-u)^{s^*} - (1-v)^{s^*}}{s^*(b-a)} = \gamma_2.$$

For $a < x < b$, concavity of $-G^{s^*}$ implies that

$$-s^*g(x) = (-s^*g(x)G(x)^{s^*-1})G(x)^{1-s^*} \leq \frac{G(a)^{s^*} - G(x)^{s^*}}{x-a}G(x)^{1-s^*} \leq \frac{u^{s^*} - G(x)^{s^*}}{x-a}G(x)^{1-s^*} \tag{3.19}$$

where $(1-G)^{s^*}$ being convex implies

$$-s^*g(x) = (-s^*g(x)(1-G(x))^{s^*-1})(1-G(x))^{1-s^*} \leq \frac{(1-G(b))^{s^*} - (1-G(x))^{s^*}}{b-x}(1-G(x))^{1-s^*}$$
$$\frac{(1-v)^{s^*} - (1-G(x))^{s^*}}{b-x}(1-G(x))^{1-s^*} \tag{3.20}$$

Adding (3.19) and (3.20) multiplied by $x-a$ and $x-b$ respectively yields that

$$-s^*g(x) \leq G(x)^{1-s^*}(u^{s^*} - G(x)^{s^*}) + (1-G(x))^{1-s^*}((1-v)^{s^*} - (1-G(x))^{s^*}).$$

Now the function

$$\Psi(y) = y^{1-s^*}(u^{s^*} - y^{s^*}) + (1-y)^{1-s^*}((1-v)^{s^*} - (1-y)^{s^*})$$

is convex on $[0,1]$ because on $[0,1]$,

$$\Psi''(y) = -s^*(1-s^*)(1-v)^{s^*}(1-y)^{-(s^*+1)} \geq 0.$$

Now since $u \leq G(x) \leq v$, convexity of $\Psi(y)$ leads to

$$-s^*g(x) \leq \max\{\Psi(u), \Psi(v)\} \leq \max\{(1-v)^{s^*} - (1-u)^{s^*}, u^{s^*} - v^{s^*}\}.$$

Taking $-s^*$ to the other side we obtain that $g(x) \leq \max\{\gamma_1, \gamma_2\}$.

Proof of (ii) It is enough to show that for any $G \in \mathcal{C}_{LU}$,

$$G(x) \leq u\left(1 + s^*\gamma_1 u^{-s^*}(x-a)\right)_+^{1/s^*} \quad \text{for } x \leq a \tag{3.21}$$

and

$$1 - G(x) \leq (1-v)\left(1 - s^*\gamma_2(1-v)^{-s^*}(x-b)\right)_+^{1/s^*} \quad \text{for } x \geq b \tag{3.22}$$

Suppose $s < 0$. Let $x \leq a$. If $x \leq F^{-1}(0)$ then (3.21) is trivially true. For $x > F^{-1}(a)$ (3.3) can be applied since $x - a \geq F^{-1}(0) - a$. Also since G^{s^*} is convex, $G(a) \leq u$ and $G(b) \geq v$ we deduce that

$$\begin{aligned}
G(x) &\leq G(a)\left(1 + s^*\frac{g(a)}{G(a)}(x-a)\right)_+^{1/s^*} \\
&= G(a)\left(1 + G(a)^{-s^*}s^*g(a)G(a)^{s^*-1}(x-a)\right)_+^{1/s^*} \\
&\leq G(a)\left(1 + G(a)^{-s^*}\frac{G(b)^{s^*} - G(a)^{s^*}}{b-a}(x-a)\right)_+^{1/s^*} \\
&\leq u\left(1 + u^{-s^*}\frac{v^{s^*} - u^{s^*}}{b-a}(x-a)\right)_+^{1/s^*} \\
&= u\left(1 + u^{-s^*}\frac{v^{s^*} - u^{s^*}}{b-a}(x-a)\right)^{1/s^*}
\end{aligned}$$

which proves (3.21). Now let $x \geq b$. If $x \geq F^{-1}(1)$ then (3.22) is trivial. Hence, we will assume $x < F^{-1}(1)$. Using concavity of $-(1-G)^{s^*}$ and (3.4) we obtain that

$$\begin{aligned}
1 - G(x) \leq& (1-G(b))\left(1 - s^*\frac{g(b)}{1-G(b)}(x-b)\right)_+^{1/s^*} \\
=& (1-G(b))\left(1 - s^*g(b)(1-G(b))^{s^*-1}(1-G(b))^{-s^*}(x-b)\right)_+^{1/s^*} \\
\leq& (1-G(b))\left(1 - (1-G(b))^{-s^*}\frac{(1-G(a))^{s^*} - (1-G(b))^{s^*}}{b-a}(x-b)\right)_+^{1/s^*} \\
\leq& (1-G(b))\left(1 - (1-v)^{-s^*}\frac{(1-u)^{s^*} - (1-v)^{s^*}}{b-a}(x-b)\right)_+^{1/s^*} \\
=& (1-v)\left(1 - (1-v)^{-s^*}\frac{(1-u)^{s^*} - (1-v)^{s^*}}{b-a}(x-b)\right)^{1/s^*}
\end{aligned}$$

which proves (3.22). □

3.8 *Proof of Theorem 3.2 when* $s \in (0, \infty)$

Proof. Our proof of Theorem 3.2 for the case $s \in (0, \infty)$ closely parallels the proof for the case $s \in (-1, 0]$. The main difference is the proof of (iii) implies (iv). When $0 < s < \infty$, $s^* = s/(1+s) \in (0, 1)$, and hence $1 - 2s^* < 0$ for $s > 1$. This requires a slightly different argument in this range and results in different constants in the Lipschitz bounds.

Let us denote $\inf J(F)$ and $\sup J(F)$ by a and b respectively. Notice that $J(F) = (a, b)$ if F is continuous.

Proof of (i) implies (ii): Since F is bi-s^*-concave with $s^* > 0$, $\psi = F^{s^*}$ is concave on (a, ∞). Consequently ψ, and hence F also, is continuous on (a, ∞) by Lemma 6 of Dümbgen *et al.* (2017). Similarly, the s^*-concavity of $1 - F$ implies continuity of $1 - F$ on $(-\infty, b)$. Now if $a = b$, F would be degenerate. Hence, $a < b$ and F is continuous on $\mathbb{R}$. Therefore we can also conclude that $J(F) = (a, b)$.

Let $x \in (a, b)$. Concavity of ψ implies that

$$F'(x\pm) = \lim_{t \to 0, \pm t > 0} \frac{\psi^{1/s^*}(x+t) - \psi^{1/s^*}(t)}{t} = \frac{1}{s^*} \psi(x)^{1/s^* - 1} \psi'(x\pm)$$

exist and satisfy

$$F'(x+) \leq F'(x-).$$

Similarly, concavity of $(1 - F)^{s^*}$ yields

$$(1 - F)'(x-) \geq (1 - F)'(x+)$$

which implies that

$$-F'(x-) \geq -F'(x+).$$

Therefore $F'(x-) = F'(x+)$ which proves the differentiability of F. It also shows that $\psi'(x+) = \psi'(x-) = \psi'(x)$ on (a, b).

By Lemma 6 of Dümbgen *et al.* (2017) for each $x \in (a, b)$ and $c \in [\psi'(x+), \psi'(x-)]$ one has

$$\psi(x+t) - \psi(x) \leq ct \quad \text{for } t \in (a-x, \infty)$$

since ψ is concave on (a, ∞). Therefore for such x and t,

$$\psi(x+t) - \psi(x) \leq t\psi'(x).$$

Hence,

$$F^{s^*}(x+t) - F^{s^*}(x) \leq ts^* f(x) F(x)^{s^*-1}$$

or,

$$\frac{F^{s^*}(x+t)}{F^{s^*}(x)} \leq 1 + s^* \frac{f(x)}{F(x)} t.$$

Hence,

$$\frac{F(x+t)}{F(x)} \leq \left(1 + s^* \frac{f(x)}{F(x)} t\right)^{1/s^*}.$$

Analogously it follows that for $t \in (-\infty, b-x)$,

$$(1 - F(x+t))^{s^*} - (1 - F(x))^{s^*} \leq -ts^* f(x)(1 - F(x))^{s^*-1}$$

which yields

$$\left(\frac{1 - F(x+t)}{1 - F(x)}\right)^{s^*} \leq 1 - ts^* \frac{f(x)}{1 - F(x)}$$

or

$$F(x+t) \geq 1 - (1 - F(x)) \cdot \left(1 - ts^* \frac{f(x)}{1 - F(x)}\right)^{1/s^*}.$$

Hence (3.4) is proved. Notice that for $s^* < 0$ the inequalities in (3.3) hold for all t because if $s^* < 0$, unlike the present case, F^{s^*} and $(1-F)^{s^*}$ are convex on the entire real line.

Proof of (ii) implies (iii): Since (ii) holds, F is continuous and differentiable on $J(F)$ with derivative $f = F'$ and satisfies (3.4). Now let $x, y \in J(F)$ with $x < y$. Let

$$h = f/F^{1-s^*}. \tag{3.23}$$

Then applying (3.4) we obtain that

$$\frac{F^{s^*}(x)}{F^{s^*}(y)} \leq 1 + s^* \frac{f(y)}{F(y)}(x-y).$$

Hence,

$$\begin{aligned} F^{s^*}(x) &\leq F^{s^*}(y) + s^* \frac{f(y)}{F(y)^{1-s^*}}(x-y) \\ &= F^{s^*}(y) + s^* h(y)(x-y) \\ &\leq F^{s^*}(x) + s^* h(x)(y-x) + s^* h(y)(x-y). \end{aligned}$$

Therefore

$$s^*(x-y)(h(y) - h(x)) \geq 0$$

where $s^*(x-y) < 0$, implying that $h(y) \leq h(x)$. Therefore h is non-increasing. Now let

$$\tilde{h} = f/(1-F)^{1-s^*}. \tag{3.24}$$

From (3.4) we also obtain that

$$(1-F(x))^{s^*} - (1-F(y))^{s^*} \leq -ts^* \frac{f(y)}{(1-F(y))^{1-s^*}} = -ts^* \tilde{h}(y)$$

or

$$\begin{aligned} (1-F(x))^{s^*} &\leq (1-F(y))^{s^*} - (x-y)s^* \tilde{h}(y) \\ &= (1-F(x))^{s^*} - (y-x)s^* \tilde{h}(x) - (x-y)s^* \tilde{h}(y) \\ &= (1-F(x))^{s^*} - s^*(y-x)(\tilde{h}(y) - \tilde{h}(x)). \end{aligned}$$

Since $s^*(y-x) > 0$, the last inequality leads to

$$0 \leq \tilde{h}(y) - \tilde{h}(x),$$

implying that $\tilde{h}$ is non-decreasing.

Proof of (iii) implies (iv): If the conditions of (iii) hold, then it immediately follows that $f > 0$ on $J(F)$. If not, suppose that $f(x_0) = 0$ for some $x_0 \in J(F)$ where $J(F) = (a, b)$ since F is continuous. Then since $f(x)/F(x)^{1-s^*}$ is non-increasing, $f(x) = 0$ for $x \in [x_0, b)$. Similarly since $f(x)/(1-F(x))^{1-s^*}$ is non-decreasing we obtain $f(x) = 0$ for $x \in (a, x_0]$. Therefore, $F' = 0$ or F is constant on (a, b) or $J(F)$. Then F violates the continuity condition of (iii). Hence $f > 0$ on $J(F)$.

Suppose h and $\tilde{h}$ are as defined in (3.23) and (3.24). Then the monotonicities of h and $\tilde{h}$ imply that for any $x, x_0 \in J(F)$,

$$f(x) = \begin{cases} F^{1-s^*}(x)h(x) \leq h(x_0) & \text{if } \ x \geq x_0, \\ (1-F(x))^{1-s^*}\tilde{h}(x) \leq \tilde{h}(x_0) & \text{if } \ x \leq x_0. \end{cases}$$

Next, let $c, d \in J(F)$ with $c < d$. We will bound $(f(y) - f(x))/(y-x)$ for $x, y \in J(F)$ such that $x, y \in (c, d)$ with $x \neq y$. This will yield local Lipschitz-continuity of f on $J(F)$. To this end, note that

$$\begin{aligned} \frac{f(y)-f(x)}{y-x} &= \frac{F^{1-s^*}(y)h(y) - F^{1-s^*}(x)h(x)}{y-x} \\ &= h(y)\frac{F^{1-s^*}(y) - F^{1-s^*}(x)}{y-x} + F^{1-s^*}(x)\frac{h(y)-h(x)}{y-x} \\ &\leq h(c)\frac{F^{1-s^*}(y) - F^{1-s^*}(x)}{y-x} \\ &\to h(c)(1-s^*)f(x)F^{-s^*}(x) = (1-s^*)h(c)h(x)F^{1-2s^*}(x) \end{aligned}$$

as $y \to x$. Here the inequality followed from the fact that

$$\frac{h(y)-h(x)}{y-x} \leq 0$$

which holds since h is non-increasing, Now since $h(x) \leq h(c)$, $s^* > 0$, $1 - s^* > 0$, and $F(c) \leq F(x) \leq F(d)$, we find that

$$\limsup_{y \to x} \frac{f(y) - f(x)}{y - x} \leq (1 - s^*) h(c)^2 F^{1-s^*}(d) F^{-s^*}(c) \tag{3.25}$$

for all $x \in (c, d)$. Analogously with $\bar{F} = 1 - F$ we obtain that

$$\begin{aligned} \frac{f(y) - f(x)}{y - x} &= \frac{\bar{F}^{1-s^*}(y)\tilde{h}(y) - \bar{F}^{1-s^*}(x)\tilde{h}(x)}{y - x} \\ &= \tilde{h}(y) \frac{\bar{F}^{1-s^*}(y) - \bar{F}^{1-s^*}(x)}{y - x} + \bar{F}^{1-s^*}(x) \frac{\tilde{h}(y) - \tilde{h}(x)}{y - x} \\ &\geq \tilde{h}(y) \frac{\bar{F}^{1-s^*}(y) - \bar{F}^{1-s^*}(x)}{y - x} \end{aligned}$$

since, by the non-decreasing property of $\tilde{h}$, for any $x, y \in J(F)$,

$$\frac{\tilde{h}(y) - \tilde{h}(x)}{y - x} > 0.$$

Next observe that since $1 - s^* = 1/(1 + s) > 0$, and $\bar{F}(y) \leq \bar{F}(x)$ if $y \geq x$,

$$\tilde{h}(y) \frac{\bar{F}^{1-s^*}(y) - \bar{F}^{1-s^*}(x)}{y - x} \geq \tilde{h}(d) \frac{\bar{F}^{1-s^*}(y) - \bar{F}^{1-s^*}(x)}{y - x}.$$

Hence as $y \to x$ it follows that

$$\begin{aligned} \liminf_{y \to x} \frac{f(y) - f(x)}{y - x} &\geq -\tilde{h}(d)(1 - s^*) f(x) \bar{F}^{-s^*}(x) \\ &= -\tilde{h}(d)\tilde{h}(x)(1 - s^*) \bar{F}^{1-2s^*}(x). \end{aligned}$$

Therefore using the fact that $\tilde{h}(x) \leq \tilde{h}(d)$ and $1 - s^*, s^* > 0$ we conclude that

$$\liminf_{y \to x} \frac{f(y) - f(x)}{y - x} \geq -\tilde{h}(d)^2 (1 - s^*) \bar{F}^{1-s^*}(c) \bar{F}^{-s^*}(d). \tag{3.26}$$

Combining the above with (3.25) we find that f is Lipschitz-continuous on (c, d) with Lipschitz-constant

$$\max\{(1 - s^*) h(c)^2 F^{1-s^*}(d) F^{-s^*}(c), (1 - s^*) \tilde{h}(d)^2 \bar{F}^{1-s^*}(c) \bar{F}^{-s^*}(d)\}.$$

This proves that f is locally Lipschitz continuous on $J(F)$. Hence, f is also locally absolutely continuous with L^1-derivative f' such that

$$f(y) - f(x) = \int_x^y f'(t)dt \quad \text{for all } x, y \in J(F);$$

hence $f'(x)$ can be chosen so that

$$f'(x) \in \left[\liminf_{y \to x} \frac{f(y) - f(x)}{y - x}, \limsup_{y \to x} \frac{f(y) - f(x)}{y - x} \right].$$

However (3.25) and (3.26) imply that for $c < x < d$,

$$\begin{aligned} &\left[\liminf_{y \to x} \frac{f(y) - f(x)}{y - x}, \limsup_{y \to x} \frac{f(y) - f(x)}{y - x} \right] \\ &\quad \subset \left[-(1 - s^*)\tilde{h}(d)^2 \bar{F}^{1-s^*}(c) F^{-s^*}(d), (1 - s^*)h(c)^2 F^{1-s^*}(d) F^{-s^*}(c) \right] \end{aligned}$$

Now since f and F are continuous and $F > 0$ on $J(F)$, so are h and $\tilde{h}$. Therefore, letting $c, d \to x$ it follows that

$$\frac{-(1 - s^*)f(x)^2 \bar{F}^{1-2s^*}(x)}{\bar{F}^{2-2s^*}(x)} \leq f'(x) \leq (1 - s^*) \frac{f(x)^2 F^{1-2s^*}(x)}{F^{2-2s^*}(x)};$$

and this implies (3.5).

Proof of (iv) implies (i): Notice that the fact that (iii) implies (i) can be easily verified since f/F^{1-s^*} non-increasing on $J(F)$ implies that F^{s^*} is concave on $J(F)$. Since F is continuous, $J(F) = (a, b)$. Now $F^{s^*} \in (0, 1)$ on $J(F)$ and $F^{s^*}(x) = 1$ for $x \geq b$. Therefore F^{s^*} is concave on (a, ∞). Similarly one can show that $(1 - F)^{s^*}$ is concave on $(-\infty, b)$. Therefore F is bi-s^*-concave. Therefore it is enough to prove that (iv) implies (iii).

By Lemma 7 of Dümbgen *et al.* (2017) h is non-increasing on $J(F)$ if and only if for any $x \in J(F)$ the following holds:

$$\limsup_{y \to x} \frac{h(y) - h(x)}{y - x} \leq 0.$$

Suppose $x \neq y \in J(F)$ and $r := \min(x, y)$ and $s := \max(x, y)$. Then it follows that

$$\begin{aligned}
\frac{h(y) - h(x)}{y - x} &= \frac{f(y)/F^{1-s^*}(y) - f(x)/F^{1-s^*}(x)}{y - x} \\
&= \frac{1}{F^{1-s^*}(y)} \frac{f(y) - f(x)}{y - x} - \frac{f(x)}{F^{1-s^*}(x)F^{1-s^*}(y)} \frac{F^{1-s^*}(y) - F^{1-s^*}(x)}{y - x} \\
&= \frac{1}{F^{1-s^*}(y)} \frac{\int_r^s f'(t)dt}{s - r} - \frac{f(x)}{F^{1-s^*}(x)F^{1-s^*}(y)} \frac{F^{1-s^*}(y) - F^{1-s^*}(x)}{y - x} \\
&\leq \frac{(1 - s^*)}{F^{1-s^*}(y)(s - r)} \int_r^s \frac{f(t)^2}{F(t)} dt - \frac{f(x)}{F^{1-s^*}(x)F^{1-s^*}(y)} \frac{F^{1-s^*}(y) - F^{1-s^*}(x)}{y - x}
\end{aligned}$$

by (3.5). Since F is continuous by (iv), $J(F) = (a, b)$. Also since $x, y \in J(F)$, $[r, s] \subset J(F)$. Since f and F are continuous on $J(F)$ and $F > 0$ on $J(F)$, f^2/F is continuous and integrable on $J(F)$ and hence also on $[r, s]$. Letting $y \to x$ we obtain that

$$\limsup_{y \to x} \frac{h(y) - h(x)}{y - x} \leq \frac{(1 - s^*)f(x)^2}{F^{2-s^*}(x)} - \frac{(1 - s^*)f(x)^2}{F^{2-s^*}(x)} = 0.$$

Analogously, by Lemma 7 of Dümbgen *et al.* (2017), to show $\tilde{h}$ is non-decreasing it is enough to show that

$$\liminf_{y \to x} \frac{\tilde{h}(y) - \tilde{h}(x)}{y - x} \geq 0.$$

To verify this suppose $x \neq y \in J(F)$ and $r := \min(x, y)$ and $s := \max(x, y)$. As before we calculate

$$\begin{aligned}
\frac{\tilde{h}(y) - \tilde{h}(x)}{y - x} &= \frac{f(y)/\bar{F}^{1-s^*}(y) - f(x)/\bar{F}^{1-s^*}(x)}{y - x} \\
&= \frac{1}{F^{1-s^*}(y)} \frac{f(y) - f(x)}{y - x} - \frac{f(x)}{\bar{F}^{1-s^*}(x)\bar{F}^{1-s^*}(y)} \frac{\bar{F}^{1-s^*}(y) - \bar{F}^{1-s^*}(x)}{y - x} \\
&= \frac{1}{\bar{F}^{1-s^*}(y)} \frac{\int_r^s f'(t)dt}{s - r} - \frac{f(x)}{\bar{F}^{1-s^*}(x)\bar{F}^{1-s^*}(y)} \frac{\bar{F}^{1-s^*}(y) - \bar{F}^{1-s^*}(x)}{y - x} \\
&\geq -\frac{(1 - s^*)}{\bar{F}^{1-s^*}(y)(s - r)} \int_r^s \frac{f(t)^2}{\bar{F}(t)} dt - \frac{f(x)}{\bar{F}^{1-s^*}(x)\bar{F}^{1-s^*}(y)} \frac{\bar{F}^{1-s^*}(y) - \bar{F}^{1-s^*}(x)}{y - x}
\end{aligned}$$

by (3.5). Since f and $\bar{F}$ are continuous on $J(F)$, letting $y \to x$ it follows that

$$\liminf_{y \to x} \frac{\tilde{h}(y) - \tilde{h}(x)}{y - x} \geq -\frac{(1 - s^*)f(x)^2}{\bar{F}^{2-s^*}(x)} + \frac{(1 - s^*)f(x)^2}{\bar{F}^{2-s^*}(x)} = 0.$$

□

Chapter 4

Application of shape-constrained techniques to HIV vaccine trials

4.1 Introduction

To date, the RV144 trial conducted in Thailand is the only vaccine efficacy trial to show signal of efficacy (31%) against HIV infection (Haynes *et al.*, 2012a). RV144 inspired a phase 1b trial, named HVTN (HIV Vaccine Trials Network) 097 whose aim was to evaluate the safety and immunogenicity of the same regimen in a South African population (Gray *et al.*, 2014). Because the predominant subtype of HIV in South Africa, namely clade C, differs from the common clades in Thailand, namely clades B and E, scientists also ran another study (HVTN 100) of a modified regimen that was adapted to include HIV strains matched to the South African clade C infections, for increasing the potential for high efficacy against clade C infections (Bekker *et al.*, 2018). The phase 1/2 HVTN 100 trial assessed its safety and immunogenicity in South Africa. In this work, we seek to compare the immune response profiles of these two vaccines. 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.*, 2012b). 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) of the regimens from HVTN 097 and HVTN 100?

Q2 Does an ordering exist between the distributions of the immune responses in question Q1?

Q3 How can we measure the discrepancy between the densities underlying the two immune responses?

Answering Q1 is the first step to our analysis. Apart from that, the density estimators have an importance of their own as well, because they can help the vaccine developers in designing subsequent vaccines. Bekker *et al.* (2018) provided statistical evidence that the IgG binding immune response to the V1V2 region in HVTN 100 was lower than for the earlier RV144 regimen. In exploring Q2 and Q3, we aim to provide further statistical verification of this finding, leveraging shape information regarding the immune response densities that may lead to more efficient inferences.

There are numerous nonparametric methods that we can implement in an effort to answer the above questions. However, when a trial is conducted on a homogeneous population like ours, in many cases, the underlying densities of the associated variables are unimodal. Furthermore, because many commonly encountered unimodal densities are log-concave, the assumption of log-concavity is not unreasonable either. Exploratory analysis based on our data also supports the assumption of these shape constraints.

Recently, shape constraints have gained much attention in density estimation. Although shape-constrained techniques have been applied in circuit design (Hannah and Dunson, 2012),

economics and operational research (Johnson *et al.*, 2018), there is little to no literature on application of shape-restricted tools in the context of vaccine trials. The advantage of shape-constrained density estimation is that they do not need tuning parameters at all (the log-concave maximum likelihood estimator (MLE) of Dümbgen and Rufibach, 2009), or need a very few (the unimodal density estimator of Birgé, 1997, needs only one). Also, shape-constrained density estimation methods do not require so many smoothness assumptions as the nonparametric methods for asymptotic consistency. We refer the interested readers to Samworth and Sen (2018) and Johnson *et al.* (2018) for more details on the applicability and recent developments in shape-constrained methods.

The central theme of our paper is to investigate whether the incorporation of shape constraints results in more efficient and simpler estimation and testing procedures for our data. In this process, we develop some general shape-constrained techniques, which can be applied to other studies, including, but not limited to, vaccine trials. Now we briefly discuss our methods and the main contributions of our study.

Estimating the underlying densities:

In vaccine trials, traditionally a kernel density estimator (KDE) is used for the purpose of density estimation (cf. Miladinovic *et al.*, 2014). However, using a cross-validation, we show that the log-concave density estimators of Dümbgen and Rufibach (2009) and Chen and Samworth (2013) minimize the risk among a class of shape-constrained estimators and the KDE. Our finding supports the use of shape constraints to estimate the density of vaccine immune responses.

Shape-constrained tests of stochastic dominance:

To answer Q2, we investigate if one immune response stochastically dominantes the other. A random variable X is said to stochastically dominate another random variable Y (in

first order) if the corresponding distribution functions F and G satisfy $G(x) \geq F(x)$ for all $x \in \mathbb{R}$. The dominance is regarded as "strict", if in addition, there exists $x \in \mathbb{R}$, such that $G(x) > F(x)$ for some $x \in \mathbb{R}$. If F does not strictly stochastically dominate G, then this event is defined as the non-dominance of F over G (Whang, 2019, p. 25).

While a test of stochastic dominance can be formulated in many ways, Davidson and Duclos (2013) and Álvarez-Esteban *et al.* (2016) advocate testing the null of non-dominance against the alternative of stochastic dominance, if our aim is to investigate whether one distribution dominates over the other. However, in the context of continuous distribution functions, unrestricted stochastic dominance is almost impossible to establish via hypothesis testing (Davidson and Duclos, 2013; Álvarez-Esteban *et al.*, 2016; Whang, 2019) because the distribution functions always overlap at the tails. Therefore, following Kaur *et al.* (1994), Davidson and Duclos (2013), and Chang and McKeague (2016), we test instead for the restricted stochastic dominance.

To the best of our knowledge, we are the first to introduce, and study, the use of shape constraints (we use Dümbgen and Rufibach (2009) and Birgé (1997)'s density estimators) in the context of testing the null of non-dominance against stochastic dominance. Moreover, some of our nonparametric test statistics as well were not studied in the context of testing the null of non-dominance. We show that, if the shape constraints are satisfied, asymptotically, our tests control the type I error at any null configuration, and are unbiased. We also show that our tests are consistent against the alternatives lying in the interior of the class of alternative distributions. We apply these tests to our data, and also perform simulation studies to compare the performance of our nonparametric and shape constrained tests.

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

To answer Q3, we rely on the Hellinger distance as a measure of discrepancy between two densities, where, for densities f and g, the Hellinger distance $H^2(f, g)$ is defined by

$$H^2(f, g) = \int_{-\infty}^{\infty} \left(\sqrt{f(x)} - \sqrt{g(x)} \right)^2 dx.$$

We use the log-concave density estimators of Dümbgen and Rufibach (2009), Chen and Samworth (2013), and the unimodal density estimator of Birgé (1997) to construct plug-in estimators of the Hellinger distance. The close connection of these shape-constrained density estimators with the MLE mitigates the need for further bias correction. This is where the shape-constrained estimates have an edge over the KDE, because, unless bias-corrected, plug-in estimators based on the latter are biased (cf. Section 2 of Robins *et al.*, 2009).

We establish that, when the shape constraints are satisfied, our plug-in estimators are asymptotically unbiased. We also show that, under some regularity conditions, the plug-in estimator based on the unimodal density estimators is asymptotically normal. The asymptotic variance of this estimator matches the asymptotic variance of the nonparametric plug-in estimators of Kandasamy *et al.* (2015), which encompass the plug-in estimators based on the KDE, and their improved bias corrected versions as well. The asymptotic results, that we establish for the plug-in estimators based on Birgé (1997)'s estimator, are not specific to the Hellinger distance, and can be extended to other plug-in estimators as well. See Section 4.8.2 for more discussion on the general set-up. Finally, we perform a simulation study to compare the performance of our estimators with the corresponding plug-in estimators based on the KDE (both bias corrected and the naïve versions).

The rest of the paper is organized as follows. In Section 4.2, we provide background on the HVTN trials. Section 4.3 discusses a cross validation procedure for selecting optimal density estimators for our trials. Section 4.4 proposes our tests of stochastic dominance. Section 4.4.3 discusses their asymptotic properties. In Section 4.4.4, we present a simulation study

to assess the performance of these tests, and Section 4.4.5 analyses the application of these tests to our data. Section 4.5 introduces the plug-in shape-constrained estimators of the Hellinger distance between two densities, discusses its asymptotic properties, and presents a simulation study on these estimators. All the proofs are deferred to the Proofs section (Sections 4.8.1 and 4.8.2).

4.2 Background: HVTN 097 and HVTN 100

In this section, first we describe our data, and then present an exploratory analysis that hints at the possibility of an ordering between the immune responses of the regimens studied in HVTN 097 and HVTN 100.

Table 4.1 provides a comparison between the two trials. 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). The magnitude of the responses is measured in net MFI (mean fluorescence intensity) units, whose logarithm is the variable of interest.

In this study, we only consider the responses measured 2 weeks after the last vaccination, which corresponds to the peak immune response. A participant is referred to as per-protocol if they have completed all stages of vaccination and provided samples at two weeks after vaccination at 6 months. Our analysis are based on vaccinated per-protocol participants only. From Table 4.1, we observe that HVTN 097 and HVTN 100 have 73 and 184 vaccinated per-protocol participants, respectively. Among them, 68 and 180 vaccinees developed immune response for at least one of the seven clade C V1V2 antigens under consideration. We only include these respondents in our study. 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_{100} and f_{097} denote the densities of the immune responses in the HVTN 097 and HVTN 100 trials respectively. Denote by F_{100} and F_{097} the

corresponding distribution functions.

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	5:1	5:1
Per protocol vaccinees	73	185
Positive respondents	68	180
Age-range	18-40	18-40
Time period	June 2013- Apr 2016	Jan 2015-Jan 2017
Clade of HIV-1 insert strains used in vaccines	B and E	C
Products used	AIDSVAX and ALVAC	ALVAC and gp120

Table 4.1: 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.

Now we provide some exploratory analysis of our data, which supports the findings of Bekker *et al.* (2018) regarding the dominance of the HVTN 097 trial in the context of IgG binding response. To see the connection, first, from Figure 4.1i, we observe that the empirical distribution function of the immune response from HVTN 100 trial is always greater than that of HVTN 097 except at the tails, suggesting the stochastic dominance of the HVTN

097 immune response over the HVTN 100 immune response. The histogram in Figure 4.1ii and the boxplot in Figure 4.1iii also suggest that the HVTN 097 trial induces higher immune response. The plot of the KDEs in Figure 4.2 agrees with the above. 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 it appears to be reasonable to compare the immune responses to the HVTN 100 regimen to those to the HVTN 097 regimen because the trials were conducted conducted on similar populations. In this context, observe that Figure 4.1 also suggests that f_{097} and f_{100} have the same support.

4.3 Density estimation

In this section, we explore different estimators of f_{097} and f_{100}, and choose the "best" among them by a 10-fold least square cross-validation. Apart from the least square cross-validation, we also consider the log-likelihood functions of the density estimators as a criterion for choosing the best density estimator. Since our study includes some unimodal and log-concave density estimators, in the following two paragraphs, we present some observations in support of the shape-restriction assumptions. However, keep in mind that an exploratory analysis can not validate the assumption of shape restrictions. It can at most provide some support.

We already mentioned in the introduction that, since each trial was conducted on homogenous populations in South Africa, there is a possibility that the underlying densities are unimodal. The histogram in Figure 4.1ii supports this claim. Figure 4.2, which displays the KDEs with the least square cross validation (LSCV) bandwidth (Bowman, 1984; Rudemo, 1982), also agrees with the same.

It is difficult to provide any visual evidence for the assumption of log-concavity. However, the density plot in Figure 4.1 does not depict any departure from log-concavity either. Using the test of log-concavity in Chen and Samworth (2013), we test the null of log-concavity against the alternative of violation of log-concavity for both densities. 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.

Now we list below the density estimators that we consider in this section.

Unimodal density estimators:

Grenander-type density estimator (Birgé, 1997) This density estimator is a piecewise constant function supported on the range of the data. When the mode is known, the MLE among all unimodal densities exists (Rao, 1969), which is based on the Grenander estimator of monotone densities. Using this mode-constrained density estimator, Birgé (1997) constructs an estimator for the general case, when the mode is unknown. His estimation procedure depends on a tuning parameter $\eta \in (0, 1)$, which quantifies the error associated with the approximation of the mode. The selection procedure of this η is more straightforward than the bandwidth of a KDE, because a smaller η always leads to more accurate estimation (cf. Definition 3 of Birgé, 1997). Indeed, if we choose the parameter $\eta = o(n^{-1})$, then Theorem 1 of Birgé (1997) implies that the total variation distance between Birgé (1997)'s estimator and the MLE among unimodal densities, based on the true mode, is $o_p(n^{-1/2})$.

In light of this, we choose η to be the inverse of the combined sample size of the two trials.

When the estimation of a smooth density is of primary interest, we do not recommend using this density estimator considering its discontinuous nature. This gives little to no reason for comparing this density estimator against other estimators which are better suited for estimating smooth densities. When the unimodality constraint is satisfied, however, among all of our unimodal density estimators, only this estimator is shown to possess desirable asymptotic properties, which we exploit during the construction of the asymptotically consistent tests in Section 4.4, and the plug-in estimator in Section 4.5. Therefore, we are interested to see how this estimator compares against other density estimators. Figure 4.4b displays

Birgé (1997)'s density estimators for the IgG binding immune responses.

Bernstein polynomial approach This density estimator was suggested by Turnbull and Ghosh (2014), who use Bernstein polynomials to estimate a unimodal density. This density estimator is smooth because it is a polynomial. However, this estimation procedure relies on tuning parameter (degree of the polynomial), though data dependent method exists for choosing the tuning parameter. We used the condition number (CN) approach developed in Turnbull and Ghosh (2014) to obtain the optimal degree of the polynomial. Figure 4.4a illustrates the density estimators for our data.

Apart from the above two estimators, we also consider the unimodal density estimators of Wolters (2012), Wolters and Braun (2018) and Hall and Huang (2002).

Log-concave density estimators

We consider two types of log-concave density estimators. The first one is the log-concave density estimator of Dümbgen and Rufibach (2009), which is the maximum likelihood estimator (MLE) among the class of all log-concave densities. However, this density estimator lacks smoothness because the corresponding log-density is piecewise affine. Chen and Samworth (2013) proposes a smoothed version of the log-concave MLE using Gaussian smoothing kernels with a data dependent bandwidth. We choose this density estimator as our second log-concave density estimator. Both these density estimators are tuning parameter free and possess desirable asymptotic properties

Kernel density estimators

Finally, we also include the KDE in our study as a comparator. 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).

Least square cross-validation (LSCV)

We consider a ten-fold cross-validation where we split our data randomly into ten equal folds, where the splitting procedure is independent of the observations. For $i = 1, \ldots, 10$, let $S_{1,i}$ and $S_{0,i}$ denote the i-th fold (test set) and its complement, respectively. Let $\hat{f}_{n,i}$ be an estimator of the true density f based on the observations in the training set $S_{0,i}$. We will denote the cardinality of a set A by $|A|$.

The LSCV focuses on minimizing the MISE, which is the integrated squared error between the estimated density $\hat{f}_{n,i}$ and the true density f. Thus, the risk function to be minimized is

$$\text{MISE} = 10^{-1} \sum_{i=1}^{10} E\bigg[\int_{-\infty}^{\infty} \Big(\hat{f}_{n,i}(x) - f(x) \Big)^2 dx \bigg].$$

Since the true density f is unavailable, instead we minimize the quantity

$$\text{ISE} = 10^{-1} \sum_{i=1}^{10} \text{ISE}_i, \tag{4.1}$$

where

$$\text{ISE}_i = \int_{-\infty}^{\infty} \hat{f}_{n,i}(x)^2 dx - 2|S_{1,i}|^{-1} \sum_{X_j \in S_{1,i}} \hat{f}_{n,i}(X_j). \tag{4.2}$$

Kullback-Leibler (KL) divergence

The Kullback-Leibler (KL) divergence between two densities f and h is given by

$$D_{KL}(h|f) = \int_{-\infty}^{\infty} \Big(\log f(x) - \log h(x) \Big) f(x) dx,$$

whose second term can be approximated by

$$-\, l_n = n^{-1} \sum_{i=1}^{n} \log h(X_i), \tag{4.3}$$

which is also the log-likelihood function of the density h. Therefore, when h is an estimator of f, we can use $-l_n$ as a loss-function as well.

Optimal density estimator

Table 4.2 lists the values of the ISE (obtained from LSCV) and $-l_n$ for different density estimators, for each trial. Based on the ISE criterion, the smoothed log-concave MLE estimator is the optimal density estimator for the HVTN 097 trial, where Birgé (1997)'s estimator is the worst. For the HVTN 100 trial though, the ISE of Birgé (1997)'s estimator is smaller than all other estimators by a large margin. When we look at the criterion $-l_n$, Birgé (1997)'s estimator, followed by the log-concave MLE estimators (both smoothed and non-smoothed), secures the minimum value for the HVTN 097 trial. Birgé (1997)'s estimator is also the minimizer of $-l_n$ for the HVTN 100 trial, though the difference with the other estimators decreases in this case. In view of the spike of Birgé (1997)'s estimator at the mode for HVTN 100, this finding is not surprising. The reason is that the density at the mode contributes a large positive term to the expression l_n and ISE.

Because Birgé (1997)'s estimator lacks smoothness, we recommend either the smoothed log-concave MLE, or the log-concave MLE for estimating the density of our immune responses.

Table 4.2: Table of the estimated risks for different density estimators of the aggregated immune response in trials HVTN 097 and HVTN 100

	HVTN 097		HVTN 100	
Estimators	ISE	$-l_n$	ISE	$-l_n$
Smoothed log-concave MLE	-0.196	1.758	-0.129	2.127
Log-concave MLE	-0.193	1.758	-0.130	2.127
Unimodal(Grenander)	-0.172	1.662	-0.126	2.071
Unimodal (Bernstein polynomial)	-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
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

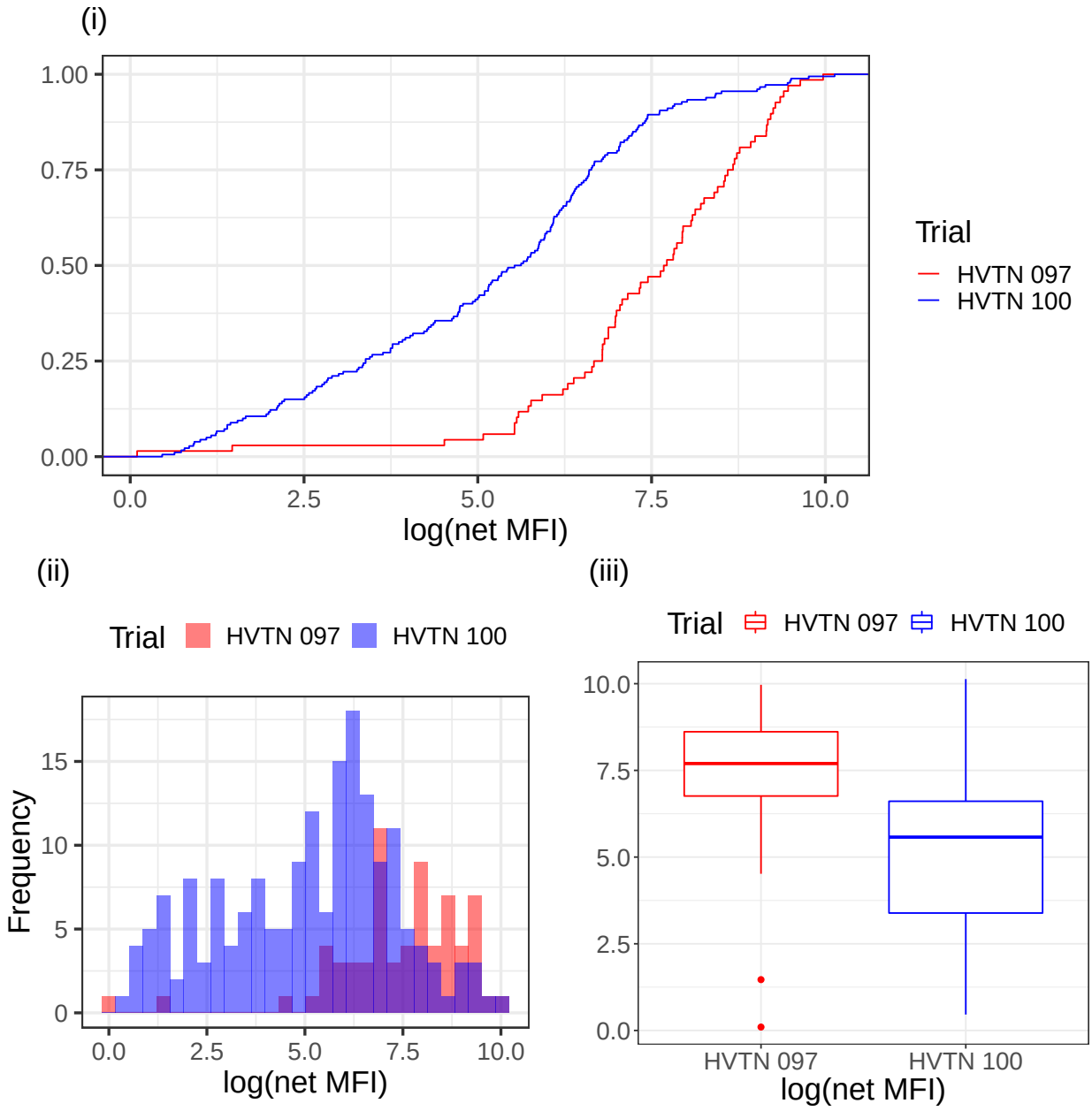


Figure 4.1: Plots (i), (ii), and (iii) display the empirical distribution functions, the histogram, and the boxplot of the average IgG binding responses (measured in log (net MFI)) corresponding to the HVTN 097 and HVTN 100 regimens.

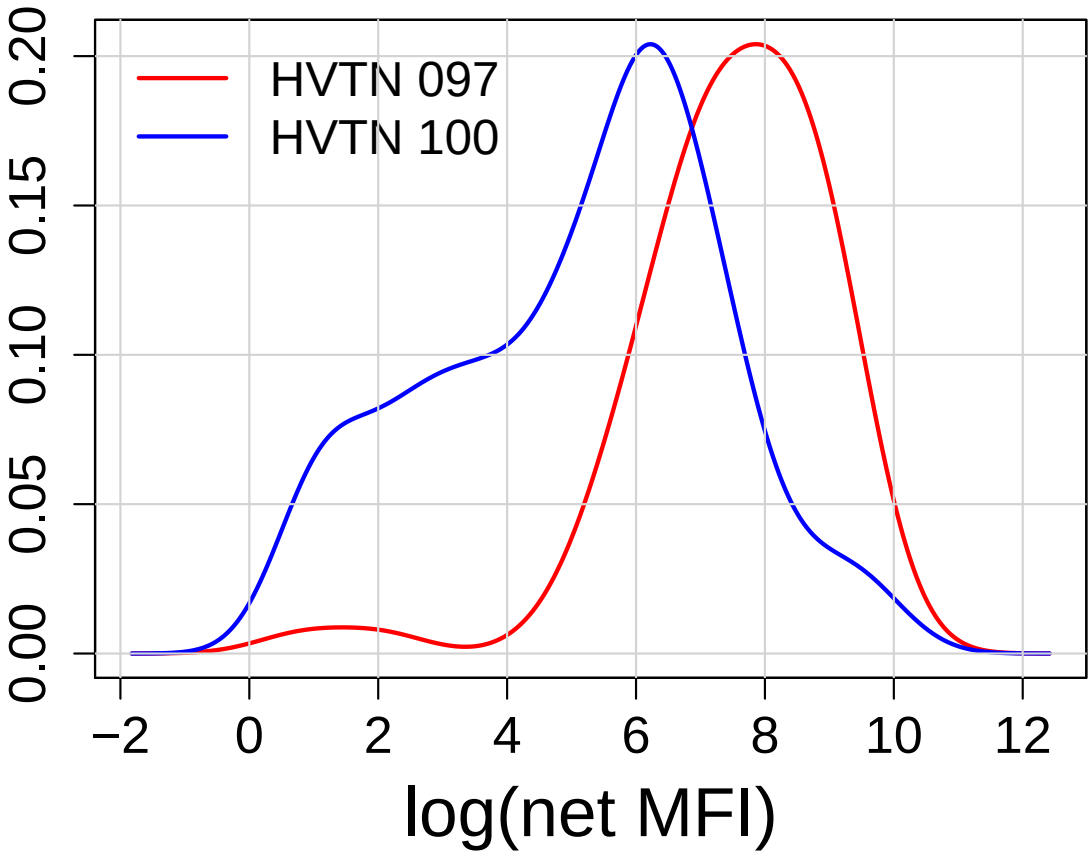
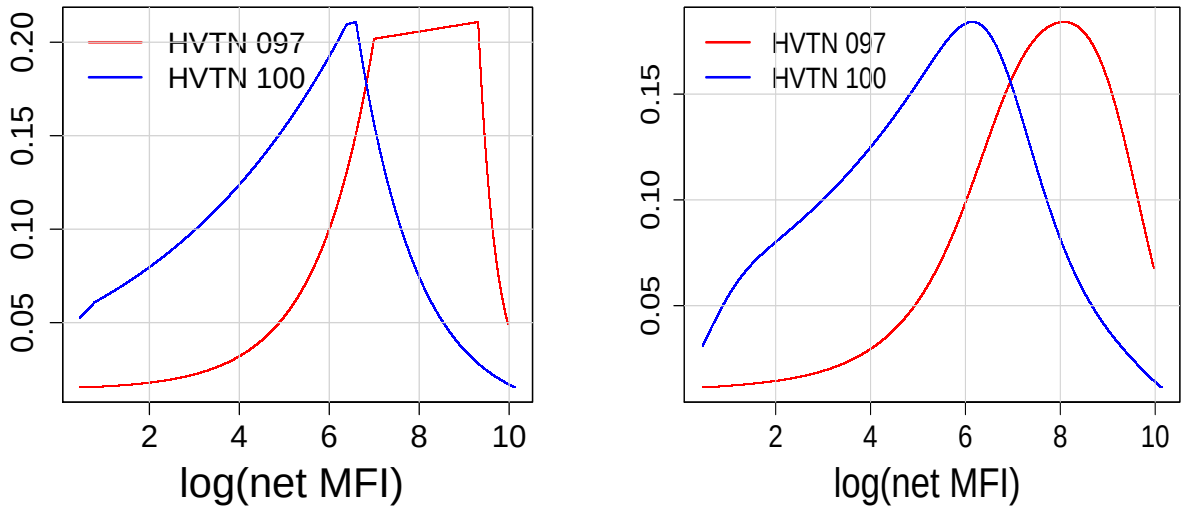


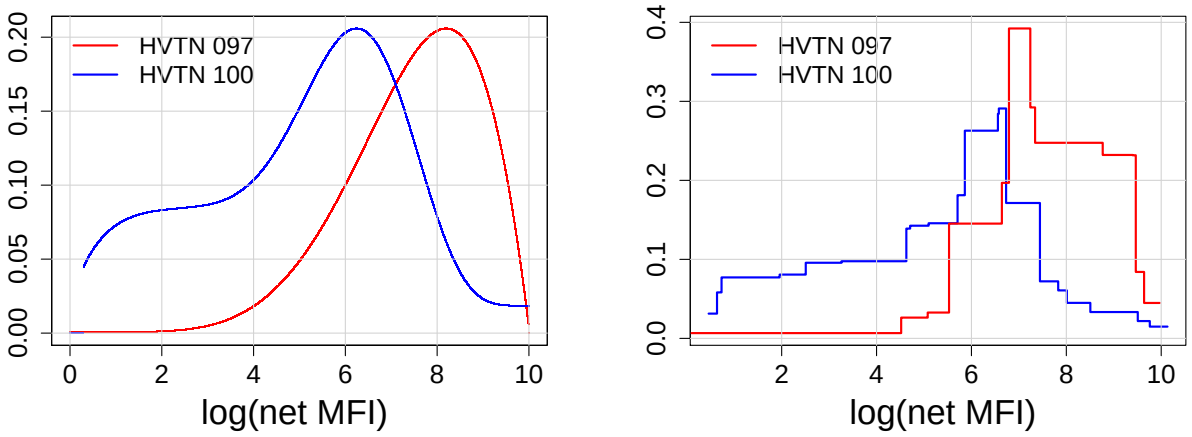
Figure 4.2: The KDEs corresponding to the immune responses from the trials HVTN 097 and HVTN 100, where the bandwidths were chosen using the LSCV selector (Bowman, 1984; Rudemo, 1982).



(a) Log-concave MLE

(b) Smoothed log-concave MLE

Figure 4.3: Log-concave density estimators based on the immune responses in the HVTN 097 and the HVTN 100 trials.



(a) Turnbull and Ghosh (2014)'s estimators

(b) Birgé (1997)'s estimators

Figure 4.4: Unimodal density estimators based on the immune responses in the HVTN 097 and the HVTN 100 trials.

4.4 *Test of stochastic dominance*

To provide an answer to the query made by Q2, we construct tests of stochastic dominance. To this end, we use the distribution functions of the log-concave estimators of Dümbgen and Rufibach (2009), Chen and Samworth (2013), and the unimodal estimator of Birgé (1997). We compare the resulting shape-constrained tests with their nonparametric counterparts.

4.4.1 Notations and terminologies

Before proceeding further, let us set up some notations that we will use 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 sample size $N = m + n$. We denote its order statistics by $(Z_{(1)}, \ldots, Z_{(N)})$, and write $\mathbb{H}_N$ for the corresponding empirical distribution function.

Our shape-restricted methods rely on the unimodal and log-concave density estimators of f and g. We denote the unimodal density estimators of f and g based on Birgé (1997) by $\widehat{f}_m$ and $\widehat{g}_n$, respectively. Recall from Section 4.3 that the construction of these unimodal density estimators require 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 MLE of f and g (Dümbgen and Rufibach, 2009), and write $\tilde{f}_m^{sm}$ and $\tilde{g}_n^{sm}$ for its 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.

For $k \geq 1$, we use the notation $||\cdot||_k$ to denote the usual L_k norm defined by

$$||\mu||_k = \left(\int_{-\infty}^{\infty} |\mu(x)|^k dx \right)^{1/k},$$

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 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 define

$$\text{dist}(A, B) = \min\{|x - y| \ : \ x \in A, y \in B\}.$$

Since we let both m and n approach infinity, we require some condition on the joint rate. Towards that end, we assume that, as $m, n \to \infty$,

$$m/N \to \lambda \in (0, 1). \tag{4.4}$$

Letting $H = \lambda F + (1 - \lambda)G$, for $p \in (0, 1/2)$, we denote by $D_p(F, G)$ the set

$$D_p(F, G) \equiv D_p = [H^{-1}(p), H^{-1}(1 - p)], \tag{4.5}$$

and by $D_{p,m,n}$, the set

$$D_{p,m,n} = [\mathbb{H}_N^{-1}(p), \mathbb{H}_N^{-1}(1 - p)]. \tag{4.6}$$

Construction of tests

We have already clarified in the introduction that we are interested in testing the null of non-dominance against restricted stochastic dominance, which can be formulated as

$$H_0 : F(z) = G(z) \text{ for all } z \in D_p, \text{ or } F(z) > G(z) \text{ for some } z \in D_p,$$

and

$$H_1 : F(z) \leq G(z) \text{ for all } z \in D_p, \text{ and } F(z) < G(z) \text{ for some } z \in D_p, \tag{4.7}$$

where D_p was defined in (4.5). Note that, to satisfy our H_1, we require F to strictly dominate G only over the restricted region D_p. Here, we could relax H_1 by letting the restricted set to be any compact set $D \subset \{0 < H < 1\}$ instead of D_p. To this end, observe that, in the event of strict stochastic dominance, such a D is likely to satisfy

$$\inf_{z\in D} \Big(G(z) - F(z) \Big) > 0,$$

so that for large m and n,

$$\inf_{z\in D}\left(\mathbb{G}_n(z)-\mathbb{F}_m(z)\right)>0,$$

which is the the main motivation behind the restriction. However, there is an advantage of choosing $D=D_p$. To see this, observe that, for large m and n, $\text{dist}(D_{p,m,n},D_p)$ becomes small almost surely. Therefore, during construction of the test statistics, we can replace the unknown set D_p by $D_{p,m,n}$, which, regardless of the forms of F and G, always allows us to utilize $100(1-2p)\%$ of the combined data. Large values of p make the rejection of H_0 easier, as a result of which, the test becomes more powerful. Here we fix the value of p large enough so that $D_{p,m,n}$ at least excludes the minimum and maximum observations from both samples.

Figure 4.5 exemplifies three different scenarios associated with our hypotheses. Figure 4.5i gives an example of the equality of distributions, and Figure 4.5ii illustrates the crossing of two distribution functions, which is also a case of non-dominance. Figure 4.5iii displays strict stochastic dominance over the region D_p.

Now we will introduce our test statistics. Our first test rejects the H_0 for large values of $T_{1,m,n}(\widehat{F}_m,\widehat{G}_n)$, where for distribution functions F_1 and F_2, we define $T_{1,m,n}(F_1,F_2)$ by

$$T_{1,m,n}(F_1,F_2)=\sqrt{\frac{mn}{N}}\inf_{z\in[p,1-p]}\left(F_2(\mathbb{H}_N^{-1}(z))-F_1(\mathbb{H}_N^{-1}(z))\right). \tag{4.8}$$

Our second test rejects the H_0 for large values of $T_{2,m,n}(\widehat{F}_m,\widehat{G}_n)$, where for distribution functions F_1 and F_2, $T_{2,m,n}$ is defined by

$$T_{2,m,n}(F_1,F_2)=\sqrt{\frac{mn}{N}}\inf_{z\in[p,1-p]}\frac{F_2(\mathbb{H}_N^{-1}(z))-F_1(\mathbb{H}_N^{-1}(z))}{\sqrt{z(1-z)}}. \tag{4.9}$$

Note that $T_{2,m,n}$ is a scaled version of $T_{1,m,n}$.

Similar to $T_{1,m,n}(\widehat{F}_m,\widehat{G}_n)$ and $T_{2,m,n}(\widehat{F}_m,\widehat{G}_n)$, we can construct log-concave tests which reject H_0 for large values of $T_{i,m,n}(\tilde{F}_m,\tilde{G}_n)$ for $i=1,2$. We compare the power of our shape-constrained tests with that of the the nonparametric tests which reject H_0 for large values of $T_{1,m,n}(\mathbb{F}_m,\mathbb{G}_n)$ and $T_{2.m,n}(\mathbb{F}_m,\mathbb{G}_n)$. We do not base our tests on $\tilde{F}_m^{sm}$ and $\tilde{G}_n^{sm}$ because

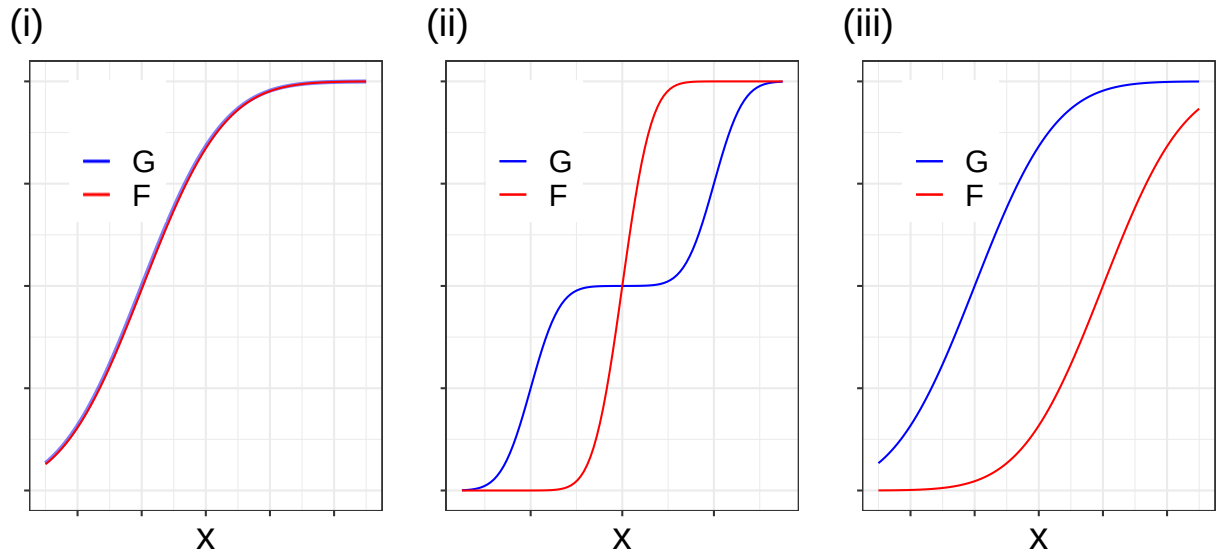


Figure 4.5: This figure displays plots of two distribution functions F and G. The range of x in these plots correspond to the respective $D_p \equiv D_p(F, G)$.

(i) This plot corresponds to the case when $F = G$ on D_p.

(ii) In this case, F and G cross each other on D_p.

(iii) In this case, F strictly stochastically dominates G. In fact $G(x) > F(x)$ for all $x \in D_p$.

we do not have any result on the asymptotic type I and type II errors of the resulting tests, since we are not yet aware of any result on the rate of $\sup_{x \in D_p} |\tilde{F}_m^{sm}(x) - \mathbb{F}_m(x)|$. The latter enables us to infer on the asymptotics of $T_{1,m,n}(\tilde{F}_m^{sm}, \tilde{G}_n^{sm})$ and $T_{2,m,n}(\tilde{F}_m^{sm}, \tilde{G}_n^{sm})$.

Ledwina and Wyλupek (2013) used a test statistic similar to $T_{2,m,n}(\mathbb{F}_m, \mathbb{G}_n)$ (the second test statistic in Section 2.2 of Ledwina and Wyλupek, 2013) for testing the null of stochastic dominance against the null of non-dominance. Kaur *et al.* (1994) also used a infimum based test statistic for testing the null of non-dominance against stochastic dominance. However,

we are not aware of any existing test, which uses either $T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n)$ or $T_{2,m,n}(\mathbb{F}_m, \mathbb{G}_n)$, for testing non-dominance against stochastic dominance.

The critical values of all of our tests will be selected to ensure that the probability of rejection of H_0 is asymptotically no greater than the nominal level α when $F(x) = G(x)$ for all $x \in D_p$. Here we emphasize that, for small samples, this need not guarantee the control of the type I error at all null configurations. Therefore, we need to provide the reasons behind this selection:

1. In section 4.4.3, we show that, for large m and n, under some regularity conditions, critical values chosen in this way provide control of the type I error at any fixed null configuration (F, G), for the tests we discussed so far.

2. For shape-constrained tests, in small samples, bootstrap or sub-sampling is necessary for finding critical values. We generate the bootstrap samples from the shape-constrained estimators of f and g. (See section 4.4.2 for more details.) This can be easily done under the restriction $F = G$. However, we are not aware of any method for computing the joint log-concave MLE of f and g under different null configurations. Similarly, there is no analogue of Birgé (1997)'s unimodal estimator for null configurations except that of $F = G$.

When $F(x) = G(x)$ for all $x \in D_p$, and the common distribution function is continuous, $T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n)$ and $T_{2,m,n}(\mathbb{F}_m, \mathbb{G}_n)$ are distribution free. To see this, note that we can write $\mathbb{F}_m = U_m \circ F$ and $\mathbb{G}_n = V_n \circ G$, where U_m and V_n are empirical distributions based on m and n independent uniform random variables, respectively. If $F(x) = G(x)$, then $\mathbb{H}_N(x) = \mathbb{W}_n \circ F(x)$, where $\mathbb{W}_n$ is the empirical distribution corresponding to the pooled uniform sample. These relations, combined with (4.8) and (4.9), imply that $T_{i,m,n}(\mathbb{F}_m, \mathbb{G}_n) = T_{i.m,n}(U_m, V_n)$ for $i = 1, 2$. Setting $\kappa'_{1,m,n,\alpha}$ and $\kappa'_{2,m,n,\alpha}$ to be the $(1-\alpha)$-th quantiles of $T_{1.m,n}(U_m, V_n)$ and $T_{2,m,n}(U_m, V_n)$, respectively, we let our nonparametric tests reject H_0 when $T_{i,m,n}(U_m, V_n) \geq$

$\kappa'_{i,m,n,\alpha}$ for $i = 1, 2$.

Apart from the above tests, we also study the shape-constrained versions of the one-sided Wilcoxon rank-sum (WRS) test. The one-sided WRS test is asymptotically the most powerful nonparametric test (cf. Example 25.46 of Van der Vaart, 1998) for testing

$$H_0^a : \int_{-\infty}^{\infty} G(z)dF(z) \geq 1/2 \quad vs \quad H_1^a : \int_{-\infty}^{\infty} G(z)dF(z) < 1/2. \tag{4.10}$$

Yet we consider this test for testing our hypotheses. The reasons are two-folds. First, this test has seen use in the context of the comparison of two vaccines (cf. Miladinovic *et al.*, 2014). Second, it is a popular choice for testing the null of equality of F of G (Lee and Wolfe, 1976).

The one-sided WRS test rejects H_0 for large values of $T_{3,m,n}(\widehat{F}_m, \widehat{G}_n)$, where, for distribution functions F_1 and F_2,

$$T_{3,m,n}(F_1, F_2) = \sqrt{\frac{12mn}{N+1}}\left(\int_{-\infty}^{\infty} F_2(x)dF_1(x) - 1/2\right). \tag{4.11}$$

Plugging in shape-constrained estimators of F and G in (4.11), we derive the shape-restricted test statistics, namely $T_{3,m,n}(\widehat{F}_m, \widehat{G}_n)$ and $T_{3,m,n}(\tilde{F}_m, \tilde{G}_n)$. We do not use $\tilde{F}_m^{sm}$ and $\tilde{G}_n^{sm}$ here for the same reason we did not use them in constructing the last two tests. Throughout this document, we refer to the tests based on $T_{3,m,n}$ as "WRS-type tests".

When $F = G$, $T_{3,m,n}(\mathbb{F}_m, \mathbb{G}_n)$ is distribution free. Therefore, we use the $(1-\alpha)$-th quantile of the resulting distribution as a critical value for the nonparametric test. Note that, this critical value may not control the type I error for all of our null configurations.

4.4.2 Bootstrap

Although Section 4.4.3 provides asymptotic critical values for the shape-constrained tests developed in the current section, simulations suggest that, unless the sample sizes are as large as 500, these critical values do not control the type I error for the shape-constrained tests based on $T_{1,m,n}$ and $T_{2,m,n}$, even when $F = G$. When the sample size is small or moderate,

Davidson and Duclos (2013) advocates using the bootstrap or subsampling methods to obtain critical values rather than relying on the asymptotic critical values for tests of stochastic dominance. Because the combined sample size N in our motivating example is less than 250, we use the bootstrap to implement the shape-constrained tests to our data.

We carry out the bootstrap procedure for our shape-constrained tests in the following steps:

1. We estimate the density of the pooled sample $(X_1, \ldots, X_m, Y_1, \ldots, Y_n)$.

2. We generate independent samples of size m and n from the estimated density, and calculate the shape-constrained test statistics. We repeat this step B times.

3. We reject the test only if the test statistic computed from the original data exceeds the $(1-\alpha)$-th quantile of the sample of B bootstrap test statistics.

In Section 4.3, we chose the log-concave MLE (Dümbgen and Rufibach, 2009) and its smoothed version (Chen and Samworth, 2013) for estimating the underlying densities. Therefore, we can use either of these estimators to generate observations from f and g in the first step.

Also observe that our bootstrap data generating procedure (DGP) uses $F(x) = G(x)$ for all $x \in \mathbb{R}$, which may be too restrictive, because, ideally, one should use the restriction that $F(x) = G(x)$ for all $x \in D_p$. However, we do not know the shape-constrained MLEs of F and G under the latter restriction. The main advantage of using MLEs is that they do not require tuning parameters. Also, as we show in our upcoming Theorem 4.1, Lemma 4.1, and Theorem 4.2, both restrictions lead to the same asymptotic distributions for the statistics $T_{1,m,n}$ and $T_{2,m,n}$.

4.4.3 Asymptotic critical values

In this section, we explore the asymptotic distribution of some of the test statistics discussed in Section 4.4. In particular, we establish that, for the tests based on $T_{1,m,n}$ and $T_{2,m,n}$, critical values can be selected so that the following assertions hold:

1. These tests control the type I error asymptotically for every (F, G) satisfying the H_0.

2. These tests are consistent for pairs of distribution functions (F, G) forming the alternative

$$H_1^* : G(x) > F(x) \quad \text{for all } x \in D_p. \tag{4.12}$$

3. These tests are asymptotically unbiased for the alternative H_1.

We show that the WRS-type tests (except the one based on $T_{3,m,n}(\tilde{F}_m, \tilde{G}_n)$) control the type I error for distributions in H_0^a, and are consistent for H_1^a. This is unsurprising in view of the fact that these tests are designed in such way. We exclude $T_{3,m,n}(\tilde{F}_m, \tilde{G}_n)$ from our discussion because we do not know whether $\sqrt{m}||\tilde{F}_m - \mathbb{F}_m||_\infty$ is small. Without this information, we are unable to infer on the asymptotic limit of $T_{3,m,n}(\tilde{F}_m, \tilde{G}_n)$. However, our simulations indicate that $T_{3,m,n}(\tilde{F}_m, \tilde{G}_n)$ has the same asymptotic distribution as $T_{3,m,n}(\mathbb{F}_m, \mathbb{G}_n)$. In the rest of this section, when we mention shape-constrained $T_{i,m,n}$, we only refer to the statistics $T_{i,m,n}(\widehat{F}_m, \widehat{G}_n)$ for $i = 1, 2, 3$, and $T_{i,m,n}(\tilde{F}_m, \tilde{G}_n)$ for $i = 1, 2$.

A key result of this section is that the distance between the shape-constrained test statistics and their nonparametric counterparts is $o_p(N^{-1/2})$. Therefore, the shape-constrained test statistics have the same asymptotic distributions as their nonparametric counterparts. Hence, the asymptotic critical values of the nonparametric tests can also serve as the asymptotic critical values of the shape-constrained tests. Thus, our first task is to find the asymptotic distributions of the nonparametric tests.

Suppose $(\Omega, \mathcal{A}, \mathcal{P})$ 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 $\to_{a.s.}$ refer 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, 2009a), we can show that there exist two independent Brownian bridges $\mathbb{V}_1$ and $\mathbb{V}_2$ on $(\Omega, \mathcal{A}, \mathcal{P})$ such that

$$||\sqrt{m}(\mathbb{F}_m - F) - \mathbb{V}_1 \circ F||_\infty \to_{a.s.} 0 \quad \text{as } m \to \infty, \tag{4.13}$$

and

$$||\sqrt{n}(\mathbb{G}_n - G) - \mathbb{V}_2 \circ G||_\infty \to_{a.s.} 0 \quad \text{as } n \to \infty. \tag{4.14}$$

Let us denote

$$\mathbb{U} = \lambda^{1/2}\mathbb{V}_2 - (1-\lambda)^{1/2}\mathbb{V}_1, \tag{4.15}$$

where λ is as in (4.4). Note that $\mathbb{U}$ is also distributed as a Brownian bridge. We will show that the asymptotic distributions of our test statistics depend on $\mathbb{U}$.

Asymptotic critical values of the nonparametric tests

All of our theorems for the nonparametric tests require that the underlying distributions F and G be continuous. This type of condition is necessary for the weak convergence of the empirical processes to Brownian bridges. We begin our discussion with the test statistics $T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n)$ and $T_{2,m,n}(\mathbb{F}_m, \mathbb{G}_n)$. We will first describe the asymptotic results, and then discuss their implications for the nonparametric tests.

Our first theorem says that $T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n)$ and $T_{2,m,n}(\mathbb{F}_m, \mathbb{G}_n)$ diverge to negative infinity under any configuration in H_0 except the case when $F = G$ on D_p. Hence it follows that, tests that reject the H_0 for high values of these test statistics, control the type I error at any null configuration, provided they also control the type I error when F and G agree on D_p. Part (B) of Theorem 4.1 implies consistency of these tests against configurations in H_1^* in (4.12).

Theorem 4.1. *Let F and G be two continuous distribution functions. Suppose that m and n satisfy (4.4). Then the following assertions hold for $i = 1, 2$:*

(A) Suppose that $F(x) > G(x)$ for some $x \in D_p$. Then, as $m, n \to \infty$,

$$T_{i,m,n}(\mathbb{F}_m, \mathbb{G}_n) \to_{a.s.} -\infty, \qquad \text{as } i = 1, 2.$$

(B) Suppose that $G(x) > F(x)$ for each $x \in D_p$. Then, as $m, n \to \infty$,

$$T_{i,m,n}(\mathbb{F}_m, \mathbb{G}_n) \to_{a.s.} \infty, \qquad \text{as } i = 1, 2.$$

Next, we consider the asymptotic distributions of $T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n)$, and its scaled version $T_{2,m,n}(\mathbb{F}_m, \mathbb{G}_n)$ under F, G satisfying $F(x) = G(x)$ for all $x \in D_p$. The asymptotic distributions of these test statistics can be used to construct asymptotic critical values.

Lemma 4.1. *Let F and G be continuous distribution functions satisfying $F(x) = G(x)$ for all $x \in D_p$, which was defined in (4.5). Suppose that m and n satisfy (4.4) and $p \in (0, 1/2)$. Then as $m, n \to \infty$, we have*

$$T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n) \to_{a.s.} \inf_{t \in [p, 1-p]} \mathbb{U}(t),$$

and

$$T_{2,m,n}(\mathbb{F}_m, \mathbb{G}_n) \to_{a.s.} \inf_{t \in [p, 1-p]} \frac{\mathbb{U}(t)}{\sqrt{t(1-t)}},$$

where $\mathbb{U}$ is as defined in (4.15).

Observe that Theorem 4.1 does not apply to alternative distributions F and G that agree at some $x \in D_p$. Theorem 4.2 discloses the asymptotic distribution of $T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n)$ and $T_{2,m,n}(\mathbb{F}_m, \mathbb{G}_n)$ under these special configurations. To formalize the setup, we define the contact set C by

$$C = \{x \in D_p \ : \ F(x) = G(x)\}. \tag{4.16}$$

Also define

$$\{H(C) = t \in [p, 1-p] \ : \ H^{-1}(t) \in C\}. \tag{4.17}$$

Observe that the set C contains the quantiles of the members of $H(C)$. Define $\kappa_{1,\alpha}$ to be the $(1-\alpha)$-th quantile of $\inf_{t\in[p,1-p]} \mathbb{U}(t)$, and denote by $\kappa_{2,\alpha}$ the $(1-\alpha)$-th quantile of

$$\inf_{t\in[p,1-p]} \frac{\mathbb{U}(t)}{\sqrt{t(1-t)}}.$$

Theorem 4.2. *Suppose that the pair of continuous distributions (F, G) satisfy the alternative hypothesis H_1 but the contact set C defined in (4.16) is non-empty. Further suppose that m and n satisfy (4.4) and $p \in (0, 1/2)$. Then*

(A)

$$T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n) \to_{a.s.} \inf_{x\in C} \mathbb{U} \circ F(x), \tag{4.18}$$

where $\mathbb{U}$ is as defined in (4.15). Also,

$$\inf_{x\in C} \mathbb{U} \circ F(x) \geq \inf_{t\in[p,1-p]} \mathbb{U}(t), \tag{4.19}$$

and,

$$\liminf_{m,n\uparrow\infty} P(T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n) \geq \kappa_{1,\alpha}) \geq \alpha. \tag{4.20}$$

(B) Suppose that F and G have continuous densities f and g, respectively. Further suppose that $h(x) = \lambda f(x) + (1-\lambda)g(x)$ is bounded away from 0 for all $x \in D_p$, where λ is defined in (4.4). Then,

$$T_{2,m,n}(\mathbb{F}_m, \mathbb{G}_n) \to_p \inf_{t\in H(C)} \frac{\mathbb{U}(t)}{\sqrt{t(1-t)}}. \tag{4.21}$$

Moreover,

$$\liminf_{m,n\uparrow\infty} P(T_{2,m,n}(\mathbb{F}_m, \mathbb{G}_n) \geq \kappa_{2,\alpha}) \geq \alpha. \tag{4.22}$$

Observe that part (B) of Theorem 4.2 requires some extra conditions, which, though being mild, require more than just that F and G be continuous. The positivity of h ensures the continuity of H^{-1}, which is used in the proof of part (B) of Theorem 4.2.

Consider the tests that reject H_0 when $T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n) > \kappa_{1,\alpha}$, and when $T_{2,m,n}(\mathbb{F}_m, \mathbb{G}_n) > \kappa_{2,\alpha}$. To see the implication of theorem 4.2, we note that, (4.20) and (4.22), combined with Lemma 4.1, imply that the above two tests are asymptotically unbiased for the alternative H_1. Theorem 4.1 implies that these tests control the type I error at all (F, G) satisfying the H_0. From Subsection 4.4.1, recall the critical values $\kappa'_{1,m,n,\alpha}$ and $\kappa'_{2,m,n,\alpha}$ that were used for the nonparametric tests. Since they are the exact critical values under the configuration (F, G) satisfying $F = G$ on D_p, tests using these critical values enjoy the same asymptotic properties as those using $\kappa_{i,\alpha}$. This follows as a corollary to Lemma 4.1, Theorem 4.1, and Theorem 4.2.. However, the tests based on $\kappa_{i,\alpha}$ or $\kappa_{i,m,n,\alpha}$ are conservative because the asymptotic type I error equals 0 for all null configurations except for the configuration in which F and G are equal on D_p.

Theorem 4.2 has another implication. The alternative distributions F and G, that agree only at some $x \in D_p$, form the boundary of the H_0 (Davidson and Duclos, 2013). Theorem 4.2 implies that under this configuration, $T_{2,m,n}$ is asymptotically distributed as a standard Gaussian random variable. Therefore, the corresponding test based on $\kappa_{2,\alpha}$ rejects the H_0 at a probability higher than α. Similarly, using Theorem 4.2, and noting $\mathbb{U} \circ F(x) > \inf_{t \in [p,1-p]} \mathbb{U}(t)$ with probability one, we arrive at the same conclusion for $T_{1,m,n}$. In view of the above, it is likely that, even though our tests based on $T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n)$ and $T_{2,m,n}(\mathbb{F}_m, \mathbb{G}_n)$ control the type I error at each null configuration, they do not control the size. Hence, even though our test

Now let us consider our WRS test statistic, which is based on the statistic $T_{3,m,n}$ defined in (4.11). Suppose that F and G are continuous distribution functions. In that case, it is

well known that (Dwass, 1956), when $F = G$,

$$T_{3,m,n}(\mathbb{F}_m, \mathbb{G}_n) \to_d N(0,1). \tag{4.23}$$

From the asymptotic normality of $T_{3,m,n}(\mathbb{F}_m, \mathbb{G}_n)$(Dwass, 1956), it also follows that, for (F, G) satisfying H_0^a,

$$T_{3,m,n}(\mathbb{F}_m, \mathbb{G}_n) \to_p -\infty, \tag{4.24}$$

and for (F, G) satisfying H_1^a,

$$T_{3,m,n}(\mathbb{F}_m, \mathbb{G}_n) \to_p \infty. \tag{4.25}$$

Asymptotic critical values of the shape-constrained tests

To show that the shape-constrained tests enjoy the same asymptotic properties as the non-parametric tests, we show that the difference between the respective test statistics is $o_p(1)$. However, to prove this, we need to impose some extra conditions on the densities. First of all, we need the corresponding shape restrictions to hold. Also, the underlying densities are required to possess smoothness or curvature to certain degree, depending on the type of shape constraint being used. Therefore, we will discuss the extra conditions for the unimodal and the log-concave densities separately.

We begin our discussion with the tests based on $\widehat{F}_m$ and $\widehat{G}_n$. In this case, apart from being unimodal, we also require f and g to be nowhere flat inside their respective domains. In other words, we need F and G to be strictly convex to the left of their respective modes, and strictly concave to the right of it. We state this regularity condition as Condition A below.

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

Condition A is a mild curvature condition. The distribution functions F and G satisfying Condition A are at the interior of the class of distribution functions with unimodal densities.

Condition A is required for Theorem 4.3, which shows that the differences $\sqrt{m}||\widehat{F}_m - \mathbb{F}_m||_\infty$ and $\sqrt{n}||\widehat{G}_n - \mathbb{G}_n||_\infty$ approach 0 almost surely as $m, n \to \infty$. Though we use Theorem 4.3 only in the context of our tests, it may be of independent interest as well.

We state Theorem 4.3 in terms of f for the sake of simplicity.

Theorem 4.3. *Suppose that f is a unimodal density satisfying Condition A. Let $\widehat{f}_m$ be the unimodal density estimator of Birgé (1997), 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 4.3. 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)

$$\sqrt{m}\int_{-\infty}^{\infty} |\widehat{f}_m(x) - \widehat{f}_m^0(x)|dx \to_{a.s.} 0.$$

(B)

$$\sqrt{m}||\widehat{F}_m - \mathbb{F}_m||_\infty \to_{a.s.} 0.$$

(C)

$$||\sqrt{m}(\widehat{F}_m - F) - \mathbb{U} \circ F||_\infty \to_{a.s.} 0,$$

where $\mathbb{U}$ is as defined in (4.15).

If one of the distributions, say F, violates Condition A, the difference between $\mathbb{F}_m$ and $\widehat{F}_m$ becomes non-negligible in $||\cdot||_\infty$. In fact, $\sqrt{m}(\widehat{F}_m - F)$ will have a different weak limit, which is no longer a Gaussian process (Beare *et al.*, 2017).

An important corollary to Theorem 4.3 is the approximation of $T_i(\widehat{F}_m, \widehat{G}_n)$ by $T_i(\mathbb{F}_m, \mathbb{G}_n)$ for $i = 1, 2, 3$, which is formalized by Lemma 4.2.

Lemma 4.2. *Suppose that f and g are unimodal densities satisfying Condition A. Further suppose that m, n satisfy (4.4). Then*

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

Now, we focus on the test statistics $T_{1,m,n}(\tilde{F}_m, \tilde{G}_n)$ and $T_{2,m,n}(\tilde{F}_m, \tilde{G}_n)$, which are connected to the log-concave density estimators. First, we will impose a smoothness condition on f and g. Before stating this condition, which we refer to as Condition B1, we need to define the Hölder class. For a compact set $K \subset \mathbb{R}$, a function μ is said to be in the Hölder class $\mathcal{H}^{\beta,L}(K)$ with exponent $\beta > 1$, and constant $L > 0$, if for all $x, y \in K$, we have

$$|\mu'(x) - \mu'(y)| \leq L|x - y|^{\beta-1}.$$

Condition B1. *The density μ satisfies $\log \mu \in \mathcal{H}^{\beta,L}(K)$, where the exponent $\beta \in [1, 2]$, the constant $L > 0$, and $K \subset \mathbb{R}$ is a compact set.*

Condition B1 is required for the application of Theorem 4.4 of Dümbgen and Rufibach (2009), which is essential to proving the asymptotic equivalence of $T_i(\tilde{F}_m, \tilde{G}_n)$ and $T_i(\mathbb{F}_m, \mathbb{G}_n)$ for $i = 1, 2$. In addition to Condition B1, we also require f and g to satisfy Condition B2, which concerns the curvature of the densities. Similar to Condition B1, Condition B2 also appears in Theorem 4.4 of Dümbgen and Rufibach (2009).

Condition B2. *μ is a log-concave density such that for some $C > 0$, the concave function $\phi = \log \mu$ satisfies*

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

for all $x, y \in K$ such that $x < y$, where K is any compact set satisfying $K \subset dom(\phi)$. Here ϕ' is either 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 $x \in K$.

Now we are in a situation to state Lemma 4.3, which establishes the asymptotic equivalence of $T_i(\mathbb{F}_m, \mathbb{G}_n)$ and $T_i(\tilde{F}_m, \tilde{G}_n)$ upto a $o_p(1)$ term.

Lemma 4.3. *Suppose that f and g are log-concave densities. Suppose, further, that there exists $p' < p$ such that*

$$D_{p'} \subset int(dom(f_X)) \cap int(dom(f_Y)). \tag{4.26}$$

Then it follows that

$$|T_i(\mathbb{F}_m, \mathbb{G}_n) - T_i(\tilde{F}_m, \tilde{G}_n)| = o_p(1), \quad \text{for } i = 1, 2.$$

Lemma 4.2 and Lemma 4.3 establish the asymptotic properties of the shape constrained tests. First, it follows that the conclusions of Theorem 4.1, Theorem 4.2, and Lemma 4.1 still hold if we replace $T_{i,m,n}(\mathbb{F}_m, \mathbb{G}_n)$ by $T_{i,m,n}(\widehat{F}_m, \widehat{G}_n)$, or $T_{i,m,n}(\tilde{F}_m, \tilde{G}_n)$, for $i = 1, 2$. Second, we see that, if we replace $T_{3,m,n}(\mathbb{F}_m, \mathbb{G}_n)$ by $T_{3,m,n}(\widehat{F}_m, \widehat{G}_n)$ in (4.23), (4.24), and (4.25), the convergences are still satisfied.

4.4.4 Simulation

In this section, we compare the performances of the tests designed in Sections 4.4.1. Our simulation scheme involves a parameter γ varying over the range $[\gamma_0, \gamma_1]$. Here γ quantifies the difference between the data generating distribution functions F_γ and G_γ. Generally, when $\gamma = \gamma_0$, either $F_{\gamma_0} = G_{\gamma_0}$, or F_{γ_0} does not stochastically dominate G_{γ_0}. When $\gamma = \gamma_1$, F_{γ_1} stochastically dominates G_{γ_1} over $D_p(F_{\gamma_1}, G_{\gamma_1})$. For intermediate values of γ, $(F_{\gamma_0}, G_{\gamma_0})$ is in between these two extreme cases. We will denote the corresponding densities by f_γ and g_γ, respectively. Since the simulations are expensive, we generally choose only ten equally spaced γs in the range $[\gamma_0, \gamma_1]$. However, the chosen γs always include γ_0 and γ_1.

For our simulation study, we consider the following cases:

(a) $F_\gamma \sim Gamma(3 + \gamma, 1)$, and $G_\gamma \sim Gamma(3, 1)$, where $Gamma(a, b)$ is a Gamma random variable with shape parameter a and scale parameter b. Here we let $\gamma \in [0, 1]$.

(b) $F_\gamma \sim N(\gamma, 1)$, and $G_\gamma \sim N(0, 1)$, where γ varies as in case (a).

(c) $F_\gamma \sim N(0, 1)$ and $G_\gamma \sim N(-3 - \gamma, 1)/2 + N(3 - \gamma, 1)/2$, where $\gamma \in [0, 2.7]$.

(d) $F_\gamma \sim N(\gamma, 1)$, and $G_\gamma \sim N(.5, 2)$, where $\gamma \in [0, 2.5]$.

(e) $F_\gamma \sim Gamma(2, \gamma)$ and $G_\gamma \sim Gamma(1, 0.5)$, where $\gamma \in [0.1, 0.5]$.

First, we check which of the above cases satisfy the shape constraints. To this end, Figure 4.6 shows that the densities in case (a) and case (b) are unimodal. Though it is not evident from the plots, these densities are also log-concave. The density g_γ in case (c), however, is bimodal for each $\gamma \in [0, 3]$ (Figure 4.6 illustrates the subcase $\gamma = 3$). Because log-concavity implies unimodality, this g_γ is not log-concave either. Since the densities in cases (d) and (e) are either normal or gamma, they also satisfy both the log-concavity and the unimodality constraints.

Now we investigate when the H_0 is satisfied. In the first three cases, it is not hard to see that $F_0 = G_0$, and $F_\gamma(z) < G_\gamma(z)$ for $z \in D_p(F_\gamma, G_\gamma)$ for $\gamma > 0$. In the last two three cases, F_γ and G_γ cross each other on D_p at $\gamma = \gamma_0$. However, as γ increases, F_γ and G_γ eventually satisfy H_1. Figure 4.7 displays the plots of F_γ and G_γ for several γs in cases (c), (d) and (e).

To summarize, Cases (a) and (b) consider the null when $F = G$. Cases (c), (d) and (e) consist of null configurations where the distribution functions cross each other on D_p. Case (c) is special because it explores the case when the shape constraints are not satisfied.

Now we describe the common simulation design for all the above cases. For each γ, we generate 100 observations each from the distributions F_γ and G_γ. Hence $m = n$, and the λ in (4.4) is 0.5 in our setup. We chose $n = 100$ because it is a moderate sample size. Simulations showed us that $n = 100$ is not large enough for the asymptotic critical values to work for the shape constrained tests.

We repeat the above process for 10000 times, and calculate the empirical power of our tests from the 10000 Monte Carlo samples. We denote the power corresponding to the configuration (F_γ, G_γ) by $\nu(\gamma)$. For the shape-constrained tests, we use the bootstrap with the smoothed log-concave density estimator (Chen and Samworth, 2013) as the data generating density. In this step, using the data generating mechanism in Cule *et al.* (2010), we generate $B = 200$ bootstrap samples. We set the p in (4.5) to be 0.05. The tests based

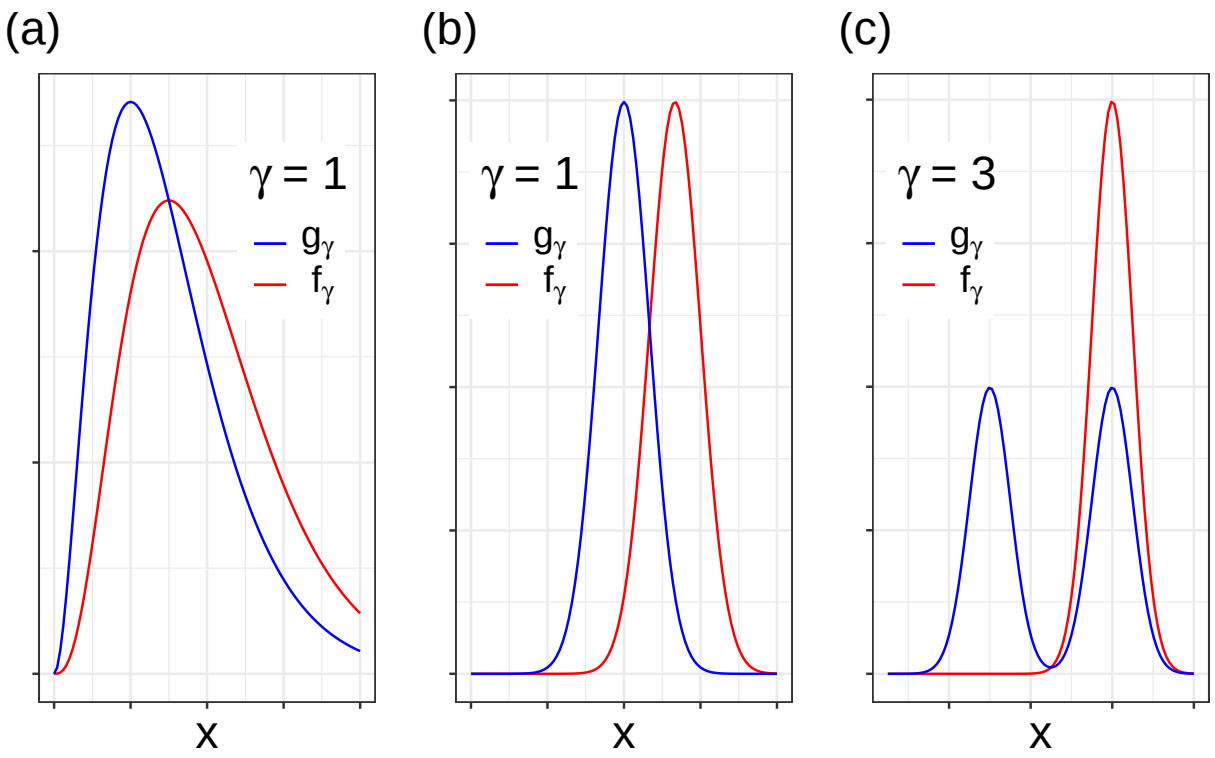


Figure 4.6: Plots of the densities f_γ and g_γ corresponding to the the cases (a), (b), and (c), for $\gamma = 1, 1$, and 3, respectively.

on the Empirical CDF will be referred as NP (nonparametric) tests. We will refer to the tests based on $\widehat{F}_m$ and $\widehat{G}_n$ as UM (unimodal) tests. Similarly, by LC(log-concave) tests, we will refer to the tests based on $\tilde{F}_m$ and $\tilde{G}_n$. We will denote the tests based on $T_{1,m,n}$ and $T_{2,m,n}$ by "RSD: unscaled" and "RSD: scaled" , respectively, where RSD stands for restrictive stochastic dominance.

Cases (a) and (b)

Figure 4.9, which gives the power curve $\gamma \mapsto \nu(\gamma)$ for cases (a) and (b), exhibit similar patterns in these two cases. First, Figure 4.9 indicates that, in both cases, the powers of

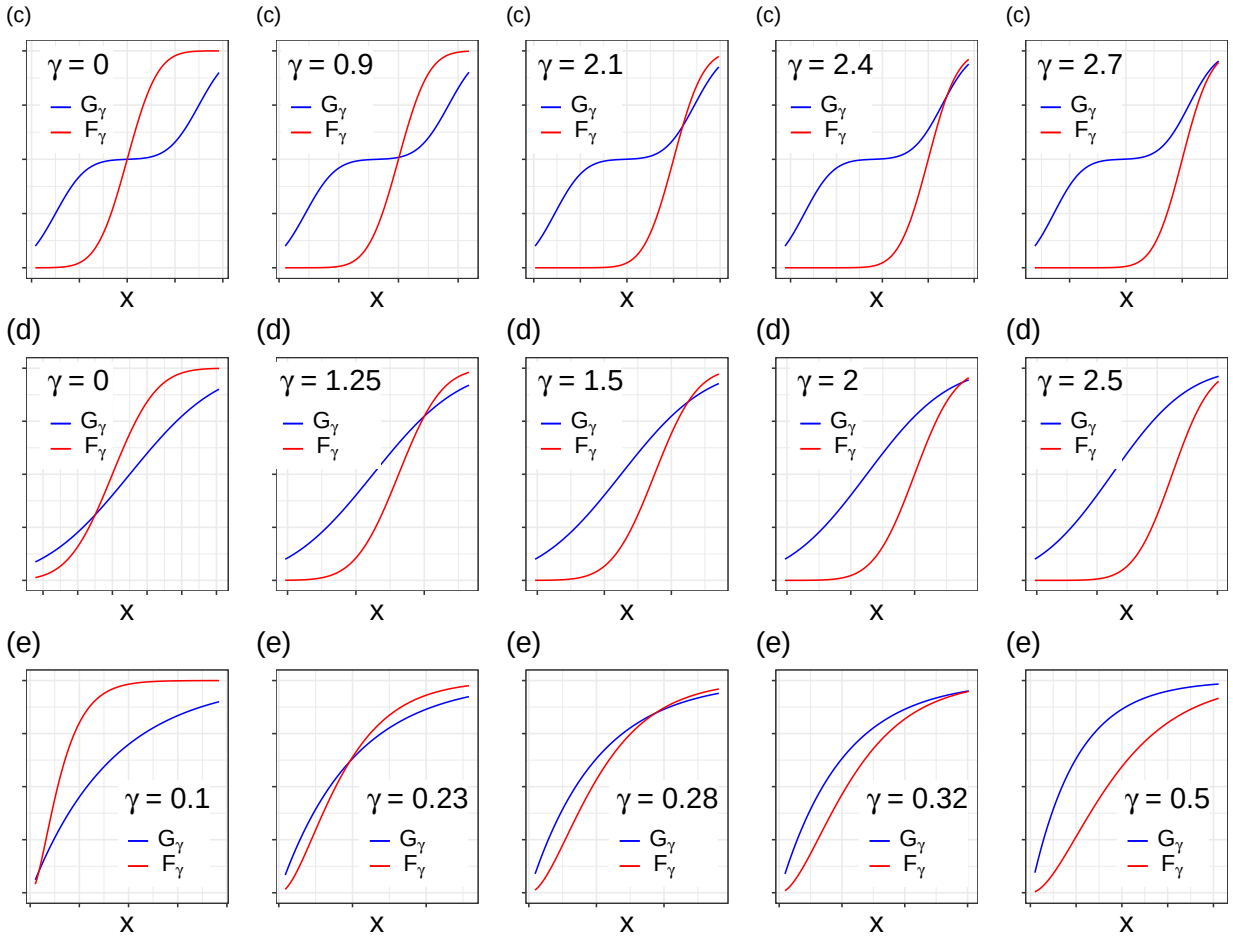


Figure 4.7: Plots of the distribution functions F_γ and G_γ for some values of γ, in cases (c), (d), and (e). In all the plots, the distribution functions are shown only on the region $D_p(F_\gamma, G_\gamma)$.

the NP, the UM, and the LC tests are very similar. However, a close observation unveils that the power of the NP tests are slightly higher. Second, Figure 4.9 illustrates that the WRS-type tests outperforms all other tests in terms of power. This is not totally unexpected since the WRS test is known to be specially suitable for location-shift type alternatives (Lee and Wolfe, 1976), as in case (a) and case (b). Also, it turns out that, when $\gamma = 0$, all the tests have type I error below or equal to 0.05. This indicates that our tests control the type I error under the equality of distributions. This is expected because our tests are designed so as to provide the latter.

Cases (c), (d) and (e)

Cases (c), (d) and (e) allow us to infer on the type I error when the null distribution functions cross each other. In these three cases, let γ^* denote the highest value of γ on the chosen grid so that (F_γ, G_γ) is still a null configuration. For cases (c), (d), and (e), γ^* take values 1.8, 2, and 0.28, respectively. At γ^*, the quantity $\inf_{x\in D_p}(G_\gamma(x) - F_\gamma(x))$ takes the values -0.13, -0.01, and -0.02, respectively, in cases (c), (d), and (e). (See Figure 4.7 for the corresponding plots of F_γ and G_γ.)

Figure 4.7 and 4.9 underscores that our tests fail to control the type I error if $\gamma < \gamma^*$ is close to γ^*. These are the cases where $\inf_{x\in D_p}(G_\gamma(x) - F_\gamma(x))$ is a small negative number, but $G_\gamma(x) > F_\gamma(x)$ for most $x \in D_p$. These cases are near the boundary of H_0. Therefore, our tests do not have the correct size, which was indicated in Subsection 4.4.3. Among all the tests, the WRS-type tests display the highest size.

Except for case (c), the tests based on $T_i(\mathbb{F}_m, \mathbb{G}_n)$ and $T_i(\widehat{F}_m, \widehat{G}_n)$ consistently exhibit the lowest rejection probability among all the tests, closely followed by the tests based on $T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n)$ and $T_{2,m,n}(\mathbb{F}_m, \mathbb{G}_n)$. The aberration in case (c) is expected because one density in this case is bimodal. Table 4.3, which gives the type I error of the RDS-type tests at γ^*, agrees with this observation.

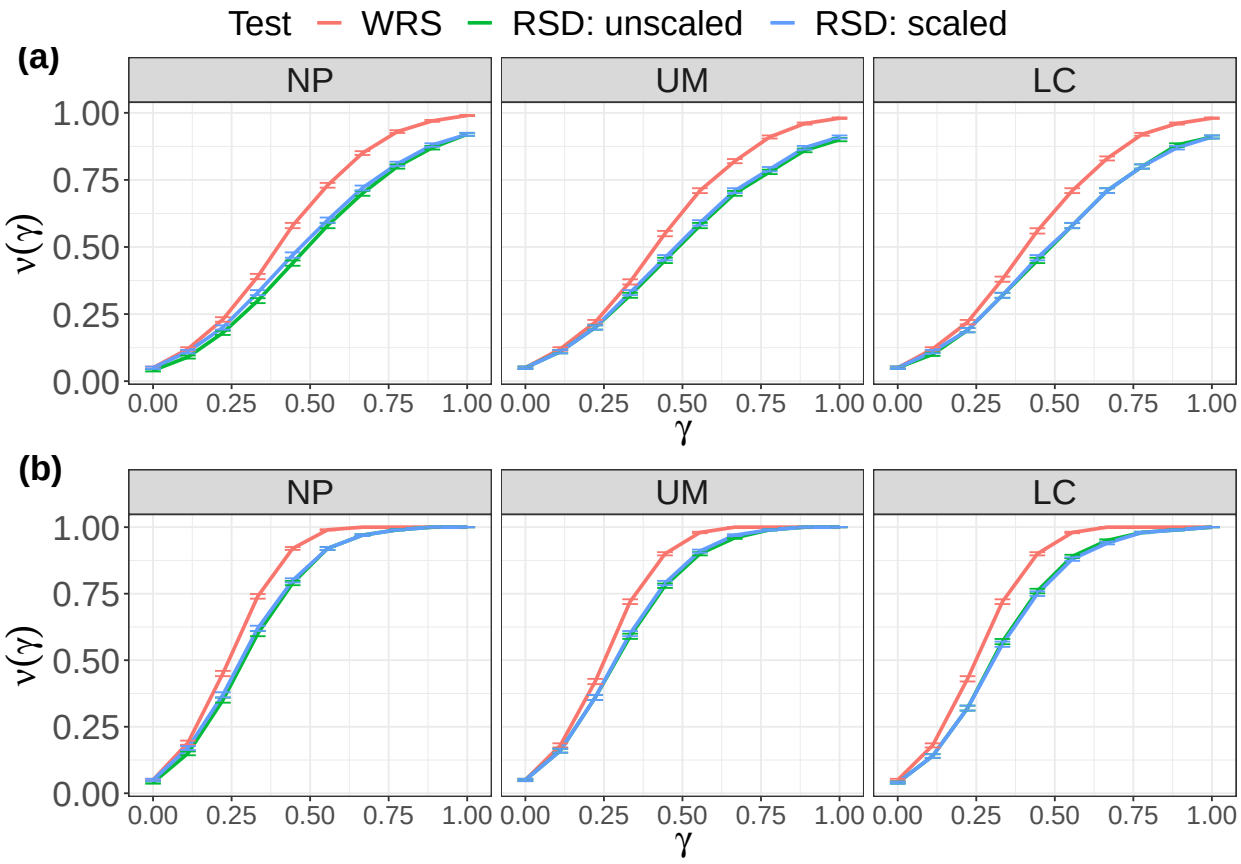


Figure 4.8: This plot compares the powers of different tests for case (a) (top), and case (b) (bottom). The NP, UM, and LC tests are given by the left, middle, and the right panels, respectively. In each of the above panels, the power curves of WRS-type, unscaled RSD ($T_{1,m,n}$ based), and the scaled RSD ($T_{2,m,n}$ based) tests are drawn in red, green, and blue, respectively. The errorbars (± 2 SD) corresponding to these tests are drawn in the same colours.

Figure 4.7 and 4.9 show that, however, the RDS-type NP and UM tests control the type I error at null configurations, where the associated γ is not too close to γ^*. These configurations are away from the boundary of H_0. In cases (d) and (e), the tests based on $T_{1,m,n}(\widehat{F}_m, \widehat{G}_n)$ and $T_{2,m,n}(\widehat{F}_m, \widehat{G}_n)$ control the type I error for $\gamma = 1.50$ and 0.23, respectively. Figure 4.7 shows that crossing of F and G occurs under both these scenarios. The term $\inf_{x \in D_p}(G_\gamma(x) - F_\gamma(x))$ equals -0.04 and -0.05 in these two cases, respectively.

Case	γ^*	$T_{1,m,n}$			$T_{2,m,n}$		
		NP	UM	LC	NP	UM	LC
(c)	1.8	0.09	0.14	0.36	0.10	0.15	0.43
(d)	2	0.38	0.33	0.81	0.38	0.31	0.88
(e)	0.28	0.14	0.14	0.75	0.16	0.13	0.78

Table 4.3: Estimated type I error of the RDS-type tests at $(F_{\gamma^*}, G_{\gamma^*})$ for cases (c), (d), and (e)

The test based on $T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n)$ has the best type I error control in case (c). In cases (d) and (e), the test based on $T_{2,m,n}(\widehat{F}_m, \widehat{G}_n)$ has the best type I error control. That being said, the difference between NP and UM tests in these two cases are almost negligible.

4.4.5 Application to HIV vaccine trial data

When we test the non-dominance against the dominance of F_{097} over F_{100}, all of our tests reject the H_0, supporting the claim that the IgG binding response rate was lower for the HVTN 100 regimen than for the HVTN 097 regimen. Now we want to assess the power of our tests when observations X_i's and Y_j's are generated from distributions (F, G) lying in a neighborhood of (F_{097}, F_{100}). Because (F_{097}, F_{100}) is unknown, we analyze the power in

a neighborhood of $(\tilde{F}^{sm}_{097}, \tilde{F}^{sm}_{100})$ instead, where $\tilde{F}^{sm}_{097}$ and $\tilde{F}^{sm}_{100}$ 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 $\tilde{f}^{0}_{m,n}$. Letting $\tilde{F}^{0}_{m,n}$ denote the corresponding distribution function, we consider the mixture distributions

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

and

$$\tilde{F}^{sm}_{100}(\gamma) = (1-\gamma)\tilde{F}_n m, n^0 + \gamma\tilde{F}^{sm}_{100}, \tag{4.27}$$

where $\gamma \in [0,1]$. Note that, as in section 4.4.4, here also γ quantifies the departure of the configuration $(\tilde{F}^{sm}_{097}(\gamma), \tilde{F}^{sm}_{100}(\gamma))$ from the null of equality of distributions. In other words, γ parametrizes the distance between $\tilde{F}^{sm}_{097}(\gamma)$ and $\tilde{F}^{sm}_{100}(\gamma)$, which increases as $\gamma \uparrow 1$. We denote the densities of $\tilde{F}^{sm}_{097}(\gamma)$ and $\tilde{F}^{sm}_{100}(\gamma)$ by $\tilde{f}^{sm}_{097}(\gamma)$ and $\tilde{f}^{sm}_{100}(\gamma)$, respectively.

Now observe that the mixture densities may not be log-concave or even unimodal. Hence, we will obtain the log-concave projections (Dümbgen *et al.*, 2011) of $\tilde{f}^{sm}_{097}(\gamma)$ and $\tilde{f}^{sm}_{100}(\gamma)$, respectively. Log-concave projection of a density f is the log-concave density closest to f in Kullback-Leibler (KL) distance. We adopt a two step approach to compute the log-concave projections. In the first step, we simulate 1000 observations from each of $\tilde{f}^{sm}_{097}(\gamma)$ and $\tilde{f}^{sm}_{100}(\gamma)$. In the second step, we calculate the smoothed log-concave MLE density estimators (Chen and Samworth, 2013) based on these two simulated samples, which are the approximate smoothed log-concave projections of $\tilde{f}^{sm}_{097}(\gamma)$ and $\tilde{f}^{sm}_{100}(\gamma)$. Finally, we generate two samples of size 68 and 180 from the log-concave projections using the methods in Dümbgen and Rufibach (2010). Then, taking $p = 0.05$, we follow the simulation design of Section 4.4.4, and denote the power at configuration $(\tilde{F}^{sm}_{097}(\gamma), \tilde{F}^{sm}_{100}(\gamma))$ by $\nu(\gamma)$.

Figure 4.10 gives the plots of $\nu(\gamma)$ versus γ for our the tests. We observe that the results are very similar to the cases (a) and (b) in Section 4.4.4, probably because in all these cases,

the null configuration corresponds to the equality of distribution functions. It turns out that the NP tests, i.e. tests based on $T_i(\mathbb{F}_m, \mathbb{G}_n)$, yield slightly greater power than the shape-constrained tests (LC and UM). Also, the WRS-type tests have the highest power, which is consistent with what we observed in all the cases in Section 4.4.4. Although its failure to control the type I error in cases (c), (d), and (e) in Section 4.4.4 puts its reliability in doubt, the other tests also have impressive power for higher γ, and have power almost one at $\gamma = 1$. These observations strengthens the claim that F_{097} stochastically dominates F_{100}, at least on the set $D_p(F_{097}, F_{100})$.

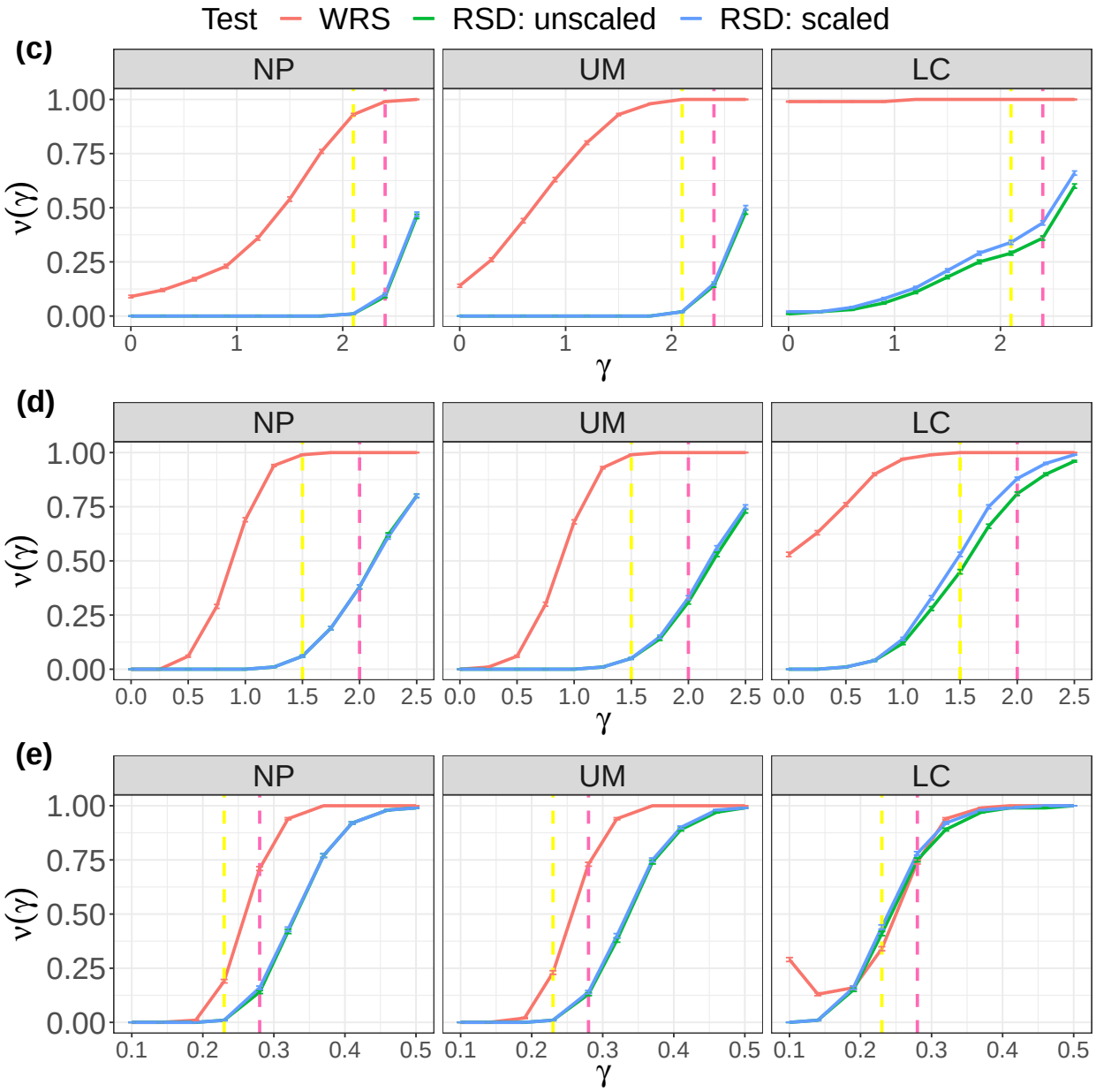


Figure 4.9: This plot compares the powers of different tests for case (c) (top), case (d) (middle), and case (e) (bottom). The NP, UM, and LC tests are given by the left, middle, and the right panels, respectively. In each of the above panels, the power curves of WRS-type ($T_{3,m,n}$ based), unscaled RSD ($T_{1,m,n}$ based), and the scaled RSD ($T_{2,m,n}$ based) tests are drawn in red, green, and blue, respectively. The errorbars ($\pm 2SD$) corresponding to these tests are drawn in the same colours. The pink line represents γ^*. The yellow line represents the largest γ for which the RSD based UM test satisfies $\nu(\gamma) \leq 0.05$ (γ =2.1, 1.50 and 0.23 for cases (c), (d) and (e), respectively).

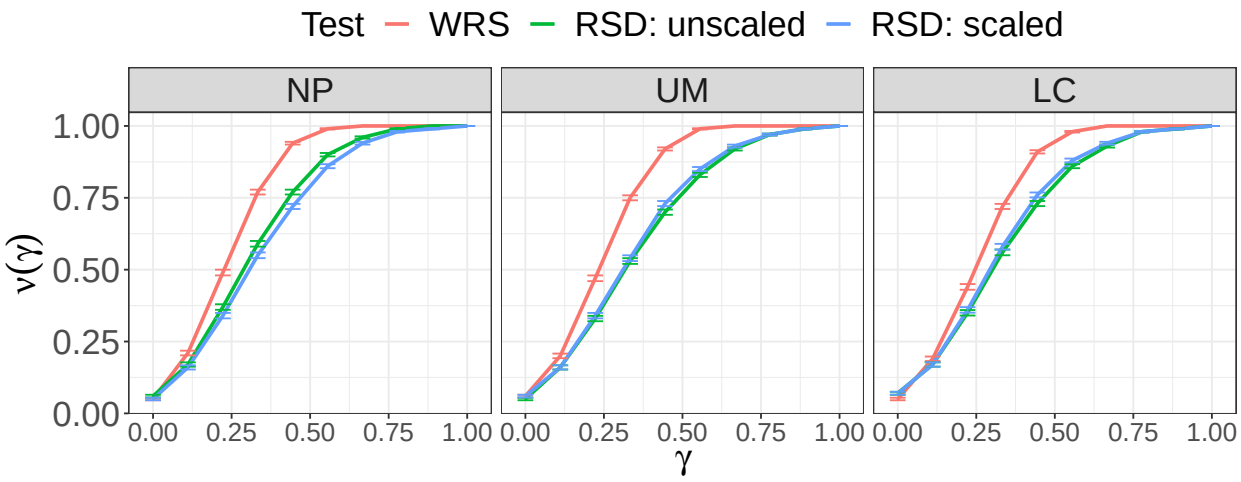


Figure 4.10: The power of the NP, the UM, and the LC tests are given by the left, middle, and the right panels, respectively. In each of the above panels, the power curves of WRS-type, unscaled RSD ($T_{1,m,n}$ based), and the scaled RSD ($T_{2,m,n}$ based) tests are drawn in red, green, and blue, respectively. The errorbars ($\pm$2SD) corresponding to these tests are drawn in the same colours.

4.5 *Measures of discrepancy*

The rejection of the tests of stochastic dominance in Section 4.4 raises new questions. If we believe that F and G are ordered, then it is natural to ask how different f and g actually are. We will quantify this difference in terms of the Hellinger distance.

We now briefly discuss the reasons that led us to study the Hellinger distance. The Hellinger distance, also known as the Bhattacharya distance, is an f-divergence (Sason, 2018). Members of the class of f-divergences are widely used to measure the similarity or dissimilarity between two probability measures. The class of f-divergences also includes popular measures of discrepancies such as the KL divergence or the χ^2 divergence (Nielsen and Nock, 2014). Unlike the Hellinger distance, these divergences are asymmetric in their arguments. As symmetry is a natural property for a measure of discrepancy to have, the Hellinger distance may be preferable in this respect. Another crucial advantage of the Hellinger distance relative to these other two divergences is that its value is finite for every pair of densities. 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, in connection to our present setting, the Hellinger distance appeals to us more than does 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). In view of the above, we will study the Hellinger distance as a measure of discrepancy between f and g in this work.

4.5.1 Shape-constrained estimators of $H^2(f, g)$

In this section, we keep using the notations developed in Subsection 4.4.1. We estimate $H^2(f, g)$ 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 the unimodal density estimator of Birgé (1997). If the respective shape constraints are satisfied, then these estimators are all related to the MLE; in particular, they are either identical to the MLE (log-concave MLE), equal to its smoothed version (smoothed log-concave MLE), or closely connected to it (the unimodal estimator).

Recalling the definition of the density estimators $\tilde{f}_m$, $\tilde{g}_n$, $\widehat{f}_m$, and $\widehat{g}_n$ from Section 4.4.1, we propose the plug-in estimators $H^2(\tilde{f}_m, \tilde{g}_n)$ and $H^2(\widehat{f}_m, \widehat{g}_n)$ for the purpose of estimating $H^2(f, g)$. Our simulations in section 4.5.3 support the above claim.

Asymptotic properties of $H^2(\widehat{f}_m, \widehat{g}_n)$

We now investigate the asymptotic properties of the unimodal estimator $H^2(\widehat{f}_m, \widehat{g}_n)$. In fact, we will show that, under some regularity conditions, $H^2(\widehat{f}_m, \widehat{g}_n)$ is asymptotically normal.

Before stating the asymptotic distribution of $H^2(\widehat{f}_m, \widehat{g}_n)$, we need to provide some background 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. Then the influence functions of T in a nonparametric model are defined as the (respectively F- and G-almost surely unique) functions $x \mapsto \psi_f(x; f, g)$ and $x \mapsto \psi_g(x; f, g)$ that satisfy the following display for all f_1 and g_1 in $\mathcal{P}$:

$$\frac{\partial}{\partial t} T(F + t(F_1 - F), G)\bigg|_{t=0} = \int_{-\infty}^{\infty} \psi_f(x; f, g) f_1(x) dx, \tag{4.28}$$

$$\frac{\partial}{\partial t} T(F, G + t(G_1 - G))\bigg|_{t=0} = \int_{-\infty}^{\infty} \psi_g(x; f, g) g_1(x) dx, \tag{4.29}$$

where above F_1 and G_1 represent the cumulative distribution functions corresponding to f_1 and g_1. In our case, $T(f, g) = H^2(f, g)$, and it follows that

$$\psi_f(x; f, g) = 2^{-1}\left(1 - \sqrt{\frac{g(x)}{f(x)}} - H^2(f, g)\right) 1_{\mathrm{supp}(f)}(x), \tag{4.30}$$

and

$$\psi_g(x; f, g) = 2^{-1}\left(1 - \sqrt{\frac{f(x)}{g(x)}} - H^2(f, g)\right)1_{\text{supp}(g)}(x). \tag{4.31}$$

Suppose that $\widehat{f}_m^0$ is the Grenander estimator of f based on the true mode of f. Theorem 4.3 established the asymptotic equivalence of $\widehat{f}_m$ and $\widehat{f}_m^0$ in terms of the Kolmogorov-Smirnov distance and the L_1 distance. Our next lemma establishes another form of asymptotic equivalence between these estimators, showing that $H^2(\widehat{f}_m, \widehat{g}_n)$ and $H^2(\widehat{f}_m^0, \widehat{g}_n^0)$ differ by an $o_p(n^{-1/2})$ term, where $\widehat{g}_n^0$ is the Grenander estimator based on the true mode of g. The proof of Lemma 4.4 relies on the results on unimodal regularizations discussed in Lemma 1 of Birgé (1997).

Lemma 4.4. *Consider two unimodal densities f and g with modes M and M', respectively. Suppose that $\widehat{f}_m$ and $\widehat{g}_n$ are Birgé (1997)'s unimodal density estimators as defined in Section 4.3, based on samples of size m and n drawn from distributions with densities f and g, respectively. Denote by $\widehat{f}_m^0$ and $\widehat{g}_n^0$ the Grenander estimators corresponding to f and g based on their true modes. If f and g satisfy the conditions of Theorem 4.3, then*

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

Letting $b, B > 0$, 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,

$$\mathcal{P}(b, B) = \left\{f \in \mathcal{P} \ : \ b \le f(x) \le B, \quad \text{for } x \in \text{supp}(f)\right\}. \tag{4.32}$$

We now present the main theorem of this subsection.

Theorem 4.4. *Suppose that f, $g \in \mathcal{P}(b, B)$, where b, $B > 0$. Further suppose that f and g satisfy condition A. Let $\widehat{f}_m$ and $\widehat{g}_n$ be Birgé (1997)'s estimators of f and g, defined in Section 4.3, based on samples of size m and n, respectively. Further suppose that $m/N \to \lambda$, where $N = m + n$. Then,*

$$\sqrt{N}[H^2(\widehat{f}_m, \widehat{g}_n) - H^2(f, g)] \to_d N(0, \sigma_{f,g}^2).$$

Here,

$$\sigma_{f,g}^2 = \frac{1}{\lambda}\int_{-\infty}^{\infty} \psi_f^2(x; f, g) f(x) dx + \frac{1}{1-\lambda}\int_{-\infty}^{\infty} \psi_g^2(y; f, g) g(y) dy,$$

where ψ_f and ψ_g are defined in (4.30) and (4.31), respectively.

The asymptotic variance $\sigma_{f,g}^2$ equals the lower bound for asymptotic variance of a regular estimator under the nonparametric model (Kandasamy *et al.*, 2015). See Van der Vaart (1998); Birgé and Massart (1995) for more detail on the lower bound.

Lemma 4.4 is pivotal to the proof of Theorem 4.4. Another key step in proving Theorem 4.4 is the Von Mises expansion (VME) (Fernholz, 2012) of $H^2(f, g)$, which has the same essence as the Taylor series expansion (Kandasamy *et al.*, 2015). See Birgé and Massart (1995) and Kandasamy *et al.* (2015) for more discussions of application of the VME in estimation of functionals of densities.

Asymptotic properties of the log-concave estimators

We now discuss the asymptotic properties of the estimators $H^2(\tilde{f}_m, \tilde{g}_n)$ and $H^2(\tilde{f}_m^{sm}, \tilde{g}_n^{sm})$. Specifically, we will show that, when f and g are log-concave, these estimators converge to $H^2(f, g)$ almost surely. We prove this as follows. First, we note that, with probability one, both the pairs $\tilde{f}_m$ and $\tilde{g}_n$, or $\tilde{f}_m^{sm}$ and $\tilde{g}_n^{sm}$, converge pointwise to f and g, respectively, almost everywhere with respect to the Lebesgue measure. The above fact follows from Proposition 2(b) of Cule and Samworth (2010) combined with Theorem 4 of Cule and Samworth (2010) (for $\tilde{f}_m$ and $\tilde{g}_n$), or Theorem 1 of Chen and Samworth (2013) (for $\tilde{f}_m^{sm}$ and $\tilde{g}_n^{sm}$). The last two theorems require the first moments of f and g to be finite, which is trivially satisfied if f and g are log-concave (cf. Lemma 1 Cule and Samworth, 2010). Next. an application of Scheffé's theorem shows that the pointwise convergence of $\tilde{f}_m$ and $\tilde{g}_n$ leads to

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

as $m, n \to \infty$. Similarly, we can show that, as $m, n \to \infty$,

$$H^2(\tilde{f}_m^{sm}, \tilde{g}_n^{sm}) \to_{a.s.} 1 - \int_{-\infty}^{\infty} \Big(f(x)g(x) \Big)^{1/2} dx.$$

We summarize the above finding in the following Lemma.

Lemma 4.5. *Suppose that the densities f and g are log-concave. Then as $m, n \to \infty$, $H^2(\tilde{f}_m, \tilde{g}_n) \to_{a.s.} H^2(f, g)$ and $H^2(\tilde{f}_m^{sm}, \tilde{g}_n^{sm}) \to_{a.s.} H^2(f, g)$.*

Even though we do not currently have any rate result for $H^2(\tilde{f}_m, \tilde{g}_n)$ and $H^2(\tilde{f}_m^{sm}, \tilde{g}_n^{sm})$, we conjecture that

$$\sqrt{N}[H^2(\tilde{f}_m, \tilde{g}_n) - H^2(f, g)] \to_d N(0, \sigma_{f,g}^2),$$

where $\sigma_{f,g}^2$ is as in Theorem 4.4.

Remark 3. The case of model misspecification is of natural interest for shape constrained estimators. Suppose that f and g are not log-concave, but they have finite first moments and satisfy

$$\int_{-\infty}^{\infty} f(x) \log_+ f(x) dx < \infty \quad \text{and} \quad \int_{-\infty}^{\infty} g(x) \log_+ g(x) dx < \infty,$$

where $\log_+ x$ denotes $\log(\max\{x, 0\})$. Then one can show that, when $m, n \to \infty$, $H^2(\tilde{f}_m, \tilde{g}_n)$ and $H^2(\tilde{f}_m^{sm}, \tilde{g}_n^{sm})$ converge almost surely to certain limits. The above follows from Proposition 2(b) of Cule and Samworth (2010) in conjunction with Theorem 4 of Cule and Samworth (2010) (in the case of $\tilde{f}_m$ and $\tilde{g}_n$), and Theorem 1 of Chen and Samworth (2013) (in the case of $\tilde{f}_m^{sm}$ and $\tilde{g}_n^{sm}$). The almost sure limit of $H^2(\tilde{f}_m, \tilde{g}_n)$ turns out to be $H^2(f^*, g^*)$. Here f^* and g^* are the log-concave approximations of f and g, respectively, in the sense of Dümbgen *et al.* (2011). On the other hand, the almost sure limit of $H^2(\tilde{f}_m^{sm}, \tilde{g}_n^{sm})$ changes to $H^2(f^{**}, g^{**})$. Here f^{**} can be expressed as the convolution of f^* with some centered Gaussian density. The variance of the latter can be expressed as the difference between the second central moments of f and f^*. The density g^{**} has the similar representation in terms of g and g^*.

4.5.2 Comparators

In our upcoming simulation study, we will consider several natural comparators for the shape-constrained estimators $H^2(\widehat{f}_m, \widehat{g}_n)$ and $H^2(\tilde{f}_m, \tilde{g}_n)$. These comparators make use of Gaussian kernel density estimators of the densities f and g. As in Section 4.3, the corresponding bandwidth is chosen using the univariate least square cross-validation (LSCV) selector of Bowman (1984) and Rudemo (1982). Let $\widehat{f}_{m,k}$ and $\widehat{g}_{n,k}$ be the corresponding KDEs of f and g.

A naïve estimator based on f and g is given by the simple plug-in estimator $H^2(\widehat{f}_{m,k}, \widehat{g}_{n,k})$. Though simple to implement, and therefore natural to consider as a comparator, it is known that the bias of this estimator can decrease to 0 at a rate slower than $n^{-1/2}$, thereby leading to a suboptimal estimator. We direct the interested readers to Section 2 of Robins *et al.* (2009) for more detail on the suboptimality of kernel density estimator-based plug-in estimators.

To improve the performance of the naïve plug-in estimator, we also consider its bias-corrected version. For the bias correction, we use the one-step Newton-Raphson procedure (cf. Van der Vaart, 1998; Pfanzagl and Wefelmeyer, 1985), which involves the influence functions $\psi_f(x; f, g)$ and $\psi_g(x; f, g)$ defined in (4.30) and (4.31), respectively. The corresponding bias-corrected estimators take the form

$$\begin{aligned}(H^2)^*(\widehat{f}_{m,k}, \widehat{g}_{n,k}) = \; & H^2(\widehat{f}_{m,k}, \widehat{g}_{n,k}) \\ & + \int_{-\infty}^{\infty} \psi_f(x; \widehat{f}_{m,k}, \widehat{g}_{n,k}) d\mathbb{F}_m(x) + \int_{-\infty}^{\infty} \psi_g(y; \widehat{f}_{m,k}, \widehat{g}_n) d\mathbb{G}_n(y),\end{aligned}$$

which, following the expansion of ψ_f and ψ_g, rewrites as

$$1 - \frac{1}{2}\left(\int_{-\infty}^{\infty} \sqrt{g(x)/f(x)} d\mathbb{F}_m(x) + \int_{-\infty}^{\infty} \sqrt{f(y)/g(y)} d\mathbb{G}_n(y) \right).$$

Note that the construction of this bias-corrected estimator requires 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,

it is necessary to carry out some extra analytical calculations that depend on the functional of interest. The shape-constrained plug-in estimators do not require the bias correction for asymptotic normality, and so do not require that these analytical calculations be performed. Shape constrained estimators are therefore desirable in that, once the shape-constrained estimates of the underlying densities have been obtained, an asymptotically normal estimator of any sufficiently smooth parameter will automatically be available; specifically, these estimators are given by the plug-in estimators of the shape-constrained density estimators.

4.5.3 Simulations:

Methods

We now describe the simulation experiments that we conducted to assess the performance of the shape-constrained estimators $H^2(\widehat{f}_m, \widehat{g}_n)$ (unimodal), $H^2(\tilde{f}_m, \tilde{g}_n)$ (log-concave), and $H^2(\tilde{f}_m^{sm}, \tilde{g}_n^{sm})$ (smoothed LC), and their comparators, namely the nonparametric estimator $H^2(\widehat{f}_{m,k}, \widehat{g}_{n,k})$, which we refer to as KDE, and its bias corrected version $(H^2)^*(\widehat{f}_{m,k}, \widehat{g}_{n,k})$, which we refer to as KDE(BC). Here we keep using the notations developed in Section 4.4.1. To keep things simple, we choose $m = n$. We let n vary 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 generate observations from the following combinations of (f, g). Our first three cases constitute the configurations (a), (b), and (c) in Section 4.4.1 with $\gamma = 1, 1$, and 3, respectively — see Figure 4.6 for the plots of the corresponding densities. In addition to these, we also consider the following three cases:

(d) $f \sim N(0, 1)$, and $g \sim Gamma(3.61, 0.71)$ (see Figure 4.11).

(e) f is standard Laplace density and $g = f(\cdot - 2)$ (see Figure 4.11).

(f) $f = \tilde{f}^{sm}_{100}$ and $g = \tilde{f}^{sm}_{097}$, where $\tilde{f}^{sm}_{100}$ and $\tilde{f}^{sm}_{097}$ are the smoothed log-concave MLEs of f_{100} and f_{097}, respectively, based on our data (see Figure 4.3b).

Thus case (c) is the only case where the shape constraints are not satisfied. For $n = 50, \ldots, 500$, we generate 10,000 pairs of samples with common size n from each of the six combinations mentioned above. We estimate the bias and the MSE of our estimators using these Monte Carlo samples. We also calculate the average coverage probability of the 95% confidence interval of $H^2(f, g)$ based on the unimodal estimator $H^2(\widehat{f}_m, \widehat{g}_n)$ and its asymptotic distribution given by Theorem 4.4.

Results

Figures 4.12 and 4.13 plot the absolute bias and the MSE as a function of n. For the ease of comparison among different methods, we scaled up the bias and the MSE by a factor of $\sqrt{n}$ and n, respectively. First of all, note that, in case (c), where the shape constraints are not satisfied, KDE and KDE(BC) perform far better than the shape-constrained estimators, whose bias and MSE increase sharply with n in this particular case.

We now consider the cases in which the shape constraint is satisfied, that is, all of the cases except for the already-considered case (c). Figures 4.12 and 4.13 indicate that, in all cases except (a), the absolute bias and the MSE of KDE increases with n, which agrees with our previous knowledge on the KDE plug-in estimators (see Section 4.5.2). Especially, for case (d), the MSE of KDE is exceptionally high. Its bias corrected version, namely KDE(BC), however, performs favorably relative to the shape-constrained estimators in terms of both absolute bias and the MSE, even when the shape constraints are satisfied, especially for cases (b), (e), and (f). In case of (b) and (c), the bias corrected estimator has the lowest bias and lowest MSE uniformly over all n.

Figures 4.12 and 4.13 show that, among the shape-constrained estimators of $H^2(f, g)$, the unimodal estimator $H^2(\widehat{f}_m, \widehat{g}_n)$ generally exhibits higher MSE and absolute bias. The

log-concave estimator $H^2(\tilde{f}_m, \tilde{g}_n)$ displays better performance. In particular, this estimator has the lowest absolute bias and MSE for case (e) except for very small samples. The smoothed log-concave estimator $H^2(\tilde{f}_m^{sm}, \tilde{g}_n^{sm})$, however, outperforms $H^2(\tilde{f}_m, \tilde{g}_n)$ in many cases, especially for small samples. For case (d), $H^2(\tilde{f}_m^{sm}, \tilde{g}_n^{sm})$ has the lowest absolute bias and MSE uniformly over all sample sizes. We conclude that the KDE(BC) estimator and the smoothed log-concave estimators have the best overall performance. Recall that, in case (f), the samples are generated from smoothed log-concave estimators $\tilde{f}_{100}^{sm}$ and $\tilde{f}_{097}^{sm}$. Therefore, as expected, the plug-in estimators based on the log-concave densities have the lowest absolute bias in this case.

Figure 4.14 displays the plots of the coverage probabilities of the 95% confidence intervals based on $H^2(\widehat{f}_m, \widehat{g}_n)$. For case (c), we see that the coverage probability decays as n increases. This is unsurprising in view of $H^2(\widehat{f}_m, \widehat{g}_n)$'s high bias and high MSE in case (c), as indicated by Figures 4.12 and 4.13. Observe that the coverage probability is noticeably poor for case (d). This can be explained from Figure 4.12, as it indicates that, in this particular case, the absolute bias is quite high for $H^2(\widehat{f}_m, \widehat{g}_n)$. Figure 4.14 implies that the coverage probabilities are around 0.60 for small samples. Although this probability gradually increases as n grows, even at $n = 500$, the probability does not reach 0.90 for any of the cases. This observation aligns with the disappointing performance of $H^2(\widehat{f}_m, \widehat{g}_n)$ in terms of bias and variance. We deduce that the usual sample sizes of phase 1b and phase 2 trials are insufficient for the asymptotics in Theorem 4.4 to kick in here.

Application to HVTN 097 and HVTN 100 data

The estimates of $H^2(f_{097}, f_{100})$, given by the log-concave, smoothed-log-concave, and the unimodal plug-in estimators, are 0.18, 0.15, and 0.21, respectively. The 95% confidence interval based on the unimodal estimators is given by (0.13 0.30). The KDE plug-in estimator and its bias-corrected version estimate the Hellinger distance to be 0.16 and 0.19, respectively.

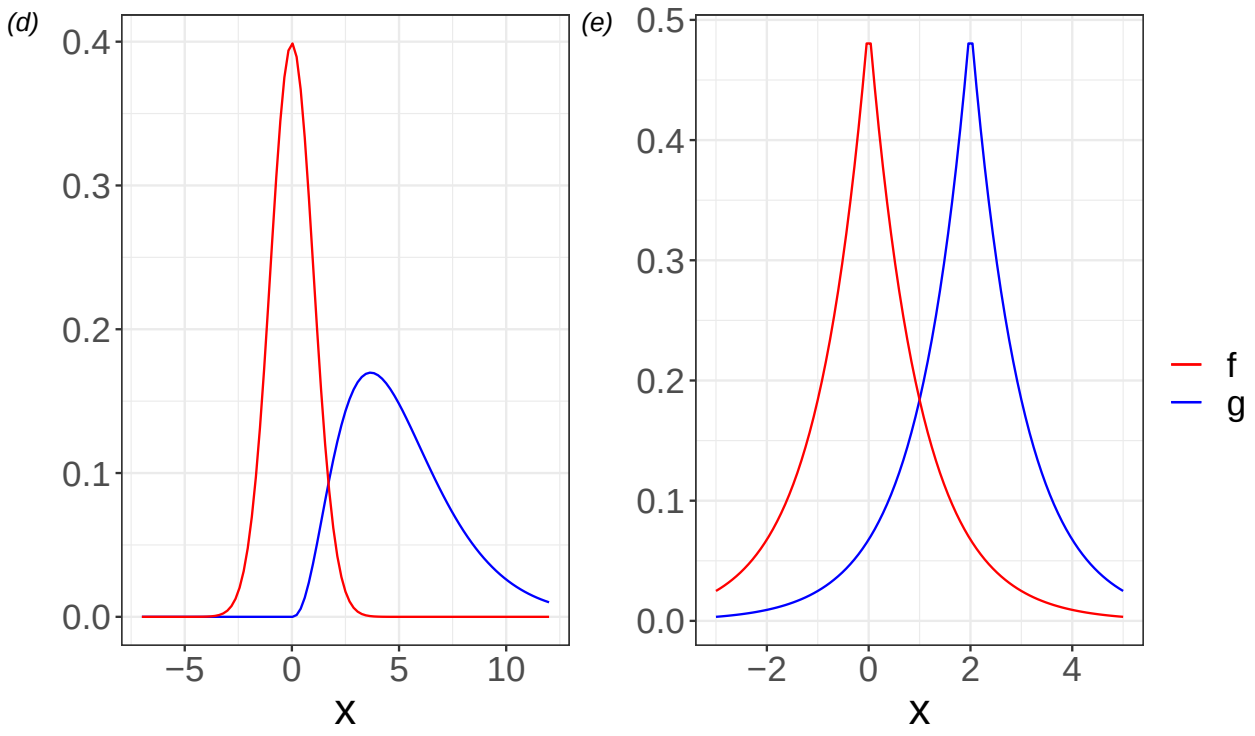


Figure 4.11: Plots of the densities f and g for the cases (d) and (e).

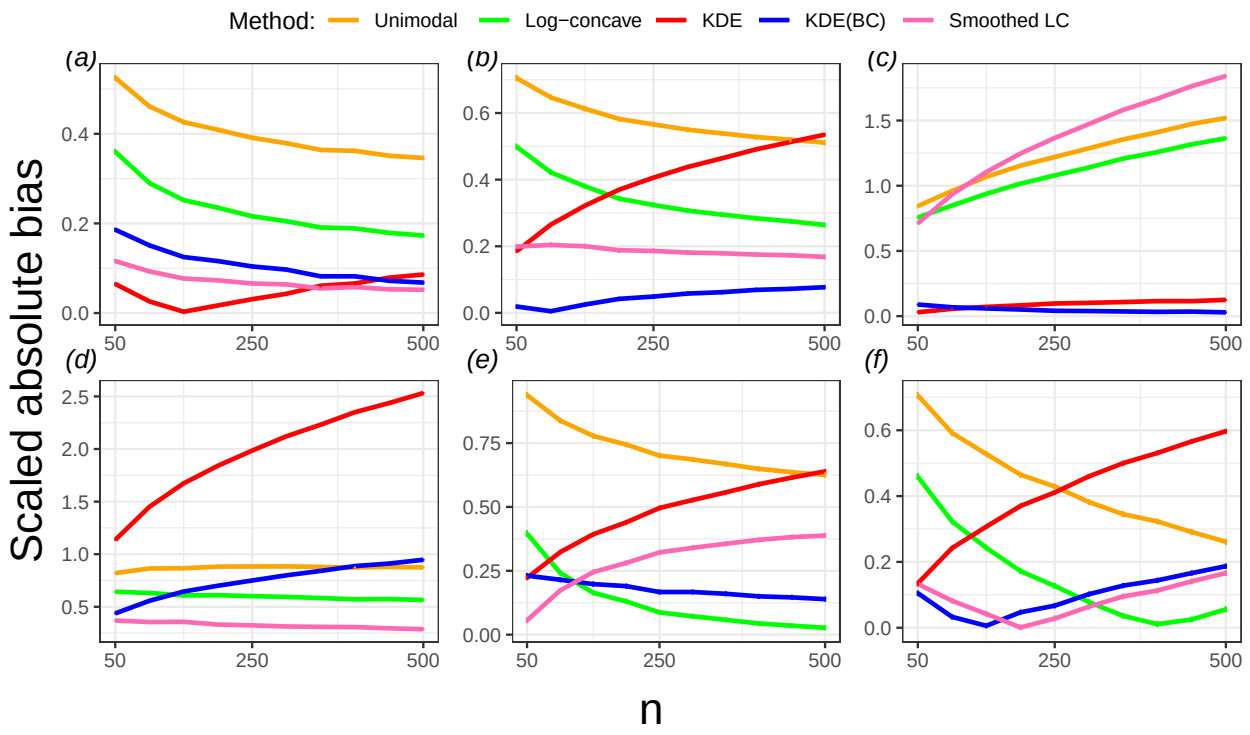


Figure 4.12: Plots of the absolute values of the biases for the cases (a)–(f). For each estimator, the errorbars are given by ±2 SD. Here SD denotes the standard deviation of the estimator, which is estimated from the 10000 Monte Carlo samples.

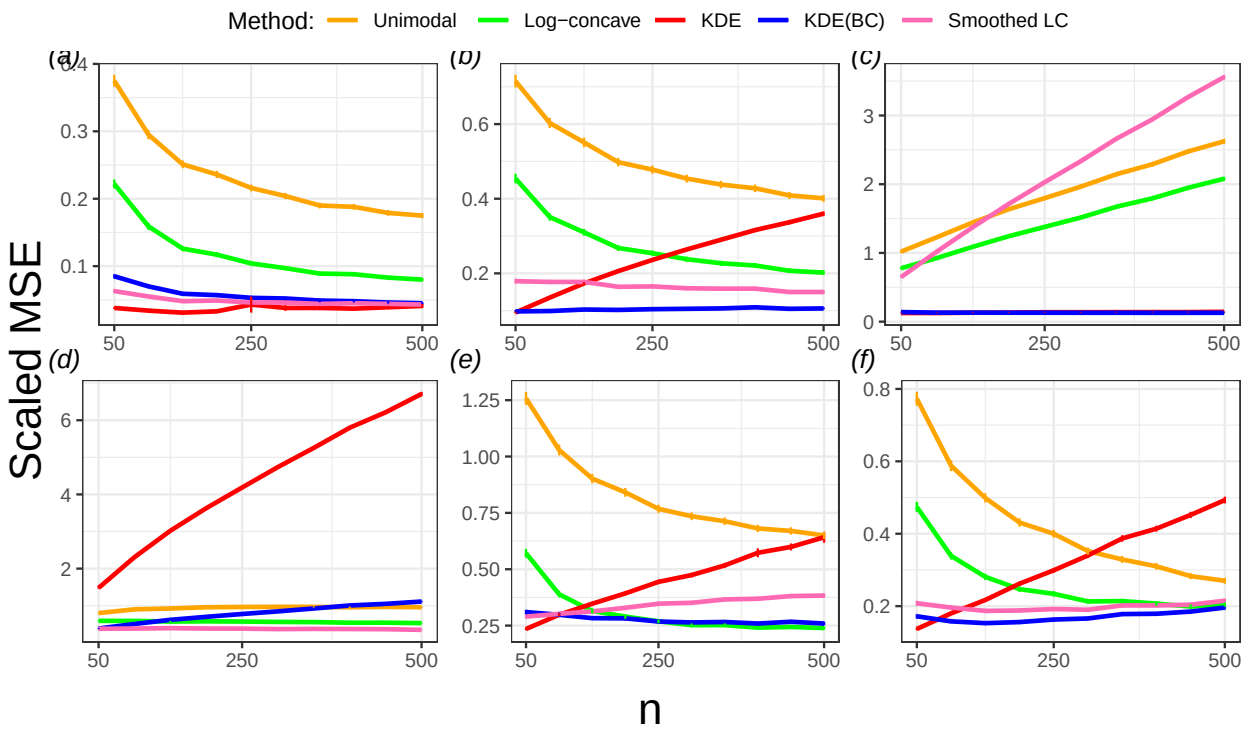


Figure 4.13: Plots of the MSE for cases (a)–(f). For each estimator, the errorbars are given by ±2 SD. Here SD denotes the standard deviation of the estimator, which is estimated from the 10000 Monte Carlo samples.

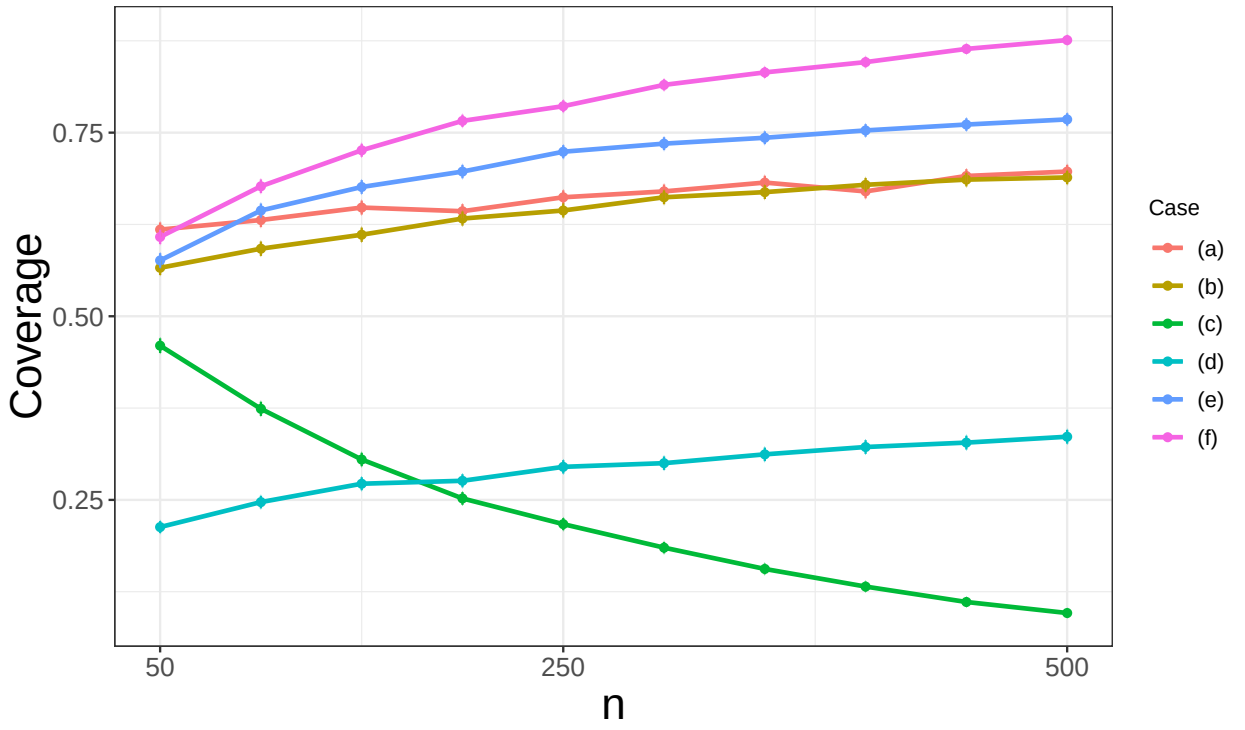


Figure 4.14: Estimated coverage probability of the 95% confidence intervals based on the estimator $H^2(\widehat{f}_m, \widehat{g}_n)$ and Theorem 4.4 for cases (a)–(f). The errorbars are given by $\pm 2SD$. Here SD, i.e. the standard deviation of the estimated probability $\widehat{p}$, is given by $(\widehat{p}(1-\widehat{p})/n)^{1/2}$.

4.6 Conclusion

The first contribution of our work is a novel analysis of the data from the HVTN 097 and HVTN 100 trials. Our study supports the findings in Bekker *et al.* (2018) that the average magnitude of IgG binding to V1V2 antigens for the HVTN 100 regimen is higher than for the HVTN 097 regimen. Indeed, 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 Hellinger distance between the corresponding densities. We found that the Hellinger distance between these two densities was approximately 0.20 (95% CI 0.13-0.30). This difference may be attributable to antigenic differences of the particular HIV-1 strains used in the two vaccines Bekker *et al.* (2018).

Our finding regarding the different immune response profiles of the HVTN 097 and 100 has the potential to impact future HIV vaccine research. As we described in the introduction, the RV144 trial in Thailand estimated the efficacy of the HVTN 097 regimen to be 31%. Though the efficacy of the HVTN 100 regimen is yet unknown, an ongoing Phase 3 trial (HVTN 702) to evaluate its efficacy is currently underway in South Africa. If the estimated efficacy of the HVTN 100 regimen turns out to be less than that of the HVTN 097 regimen, then researchers may infer that maintaining IgG binding to V1V2 antigens at the level induced by the HVTN 097 vaccine is important when trying to improve upon this regimen. However, if the outcome favors the HVTN 100 regimen, then the conclusion may be that IgG binding to V1V2 antigens is not critical for developing an efficacious HIV vaccine regimen. Instead, the immune responses that were boosted by the HVTN 100 regimen, such as responses related to the CD4+ T-cells or the gp120 antibody (Bekker *et al.*, 2018), may be given more emphasis when designing subsequent vaccines. Thus, in either case, our conclusion can help to generate a hypothesis about the association between IgG binding response and clade C antigens.

The second 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. Further investigation is required to validate this claim.

The third contribution of our work is methodological in nature. This contribution is two-fold. In Section 4.4, we introduce three novel shape-constrained tests. Though these tests have desirable asymptotic properties, our simulations illustrate that the shape-constrained tests do not exhibit much improvement over the nonparametric tests in terms of the power. Indeed, their overall performance is similar to that of the nonparametric tests. This observation hints that the knowledge of the shape of the density does not lead to more efficient (in the sense of p. 367, Van der Vaart, 1998) statistical inferences. To shed some light on the potential reason, we note that, in the nonparametric model, unbiased estimators or tests based on the empirical distribution functions are asymptotically efficient under appropriate conditions (cf. p.386, Van der Vaart, 1998). This is true because, in a nonparametric model, the empirical distribution function $\mathbb{F}_m$ is an asymptotically efficient estimator of the underlying distribution function F (see Example 25.24, Van der Vaart, 1998). We conjecture that, as long as F is sufficiently smooth, $\mathbb{F}_m$, as well as the shape-constrained distribution function $\tilde{F}_m$ (or $\widehat{F}_m$), is an asymptotically efficient estimator of F in the log-concavity (or unimodality) constrained model. Consequently, if the parameter of interest is a smooth functional of distribution functions, then an additional assumption of log-concavity or unimodality, despite allowing a simpler estimation procedure, may not facilitate a more efficient estimator. The same conclusion may hold for testing hypotheses that are formulated using smooth functionals of distribution functions only. Further investigation in this direction is needed to verify these claims.

We also introduce shape-constrained plug-in estimators of the Hellinger distance between

two densities and provide asymptotic consistency and distributional results. Our simulations show that, when the shape constraints are satisfied, the shape-constrained plug-in estimators have lower MSE and absolute bias than a KDE plug-in estimator. The performances of the log-concave plug-in estimators are comparable with that of the bias-corrected KDE plug-in estimator in most settings. Unlike the bias-corrected KDE plug-in estimator, the log-concave plug-in estimator requires neither selecting a tuning parameter nor carrying out the analytic calculations needed to derive the bias-correction term. Therefore, the log-concave plug-in estimator may be preferred in settings where this shape constraint is plausible.

4.7 Acknowledgment

The authors are grateful to the HIV Vaccine Trials Network for providing data from the HVTN 097 and HVTN 100 trials for this analysis. The authors are also grateful to the participants, investigators, and sponsors of the HVTN 097 and HVTN 100 trials. Research reported in this publication was partly supported by the National Institute Of Allergy And Infectious Diseases of the National Institutes of Health under Award Number UM1AI068635. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

4.8 Proofs

Before proceeding any further, we introduce some new notations. We let $l^\infty(T)$ denote the collection of all bounded functions on a set $T \subset \mathbb{R}$. The class $l^\infty(\mathbb{R})$ will be denoted by l^∞, 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

$$||\mu||_a^b = \sup_{x \in [a,b]} |\mu(x)|.$$

Suppose that C_0 is the collection of all compactly supported continuous functions. We denote the boundary of a set A by $\text{bd}(A)$. We let $I : \mathbb{R} \mapsto \mathbb{R}$ denote the identity map. To define the

LCM operator $\mathcal{M}_T$ over any convex set $T \subset \mathbb{R}$, we use the following definition from Beare *et al.* (2017).

Definition 2. Suppose that T is a nonempty convex set $T \subset \mathbb{R}^+$. The LCM over T is the operator $\mathcal{M}_T : l^\infty(T) \mapsto l^\infty(T)$ that maps each $\theta \in l^\infty(T)$ to the function

$$\mathcal{M}_T\theta(x) = \inf\{\mu(x) : \mu \in l^\infty(T), \mu \text{ is concave, and } \theta \leq \mu \text{ on } T\}, \quad x \in T.$$

Denote by $\mathcal{M}$ the LCM operator on $\mathbb{R}^+ = [0, \infty)$, i.e. $\mathcal{M} = \mathcal{M}_{\mathbb{R}^+}$.

4.8.1 Proofs of Section 4.4

Proof of Theorem 4.1

Note that we can write

$$\begin{aligned} T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n) &= \sqrt{\frac{mn}{N}} \inf_{t\in[p,1-p]} \Big(\mathbb{G}_n(\mathbb{H}_N^{-1}(t)) - \mathbb{F}_m(\mathbb{H}_N^{-1}(t)) \Big) \\ &= \sqrt{\frac{mn}{N}} \inf_{x\in D_{p,m,n}} \Big(\mathbb{G}_n(x) - \mathbb{F}_m(x) \Big). \end{aligned}$$

Now

$$\begin{aligned} &\sqrt{\frac{mn}{N}} \Big(\mathbb{G}_n(x) - \mathbb{F}_m(x) \Big) \\ &\quad = \sqrt{\frac{mn}{N}} \Big(\mathbb{G}_n(x) - G(x) \Big) - \sqrt{\frac{mn}{N}} \Big(\mathbb{F}_m(x) - F(x) \Big) \\ &\qquad + \sqrt{\frac{mn}{N}} \Big(G(x) - F(x) \Big). \end{aligned} \tag{4.33}$$

Suppose that (F, G) satisfies the null hypothesis but $F \neq G$. Then there exists $x \in D_p$ such that $G(x) - F(x) \leq -2\delta$ for some $\delta > 0$. Recall the definition of D_p from (4.5). Since H is continuous, we can find $p' > p$ such that $\inf_{x\in D_{p'}}(G(x)-F(x)) \leq -\delta$, and $H^{-1}(p') > H^{-1}(p)$. Note that, because $G(x)-F(x)$ is continuous, its infimum over the compact set $D_{p'}$ is attained by some $x_0 \in D_{p'}$. Recalling the definition of $D_{p,m,n}$ from (4.6), we observe that $x_0 \in D_{p,m,n}$

almost surely for all sufficiently large m and n. The latter follows because for all sufficiently large m and n, $D_{p'} \subset D_{p,m,n}$ almost surely as $\mathbb{H}_N^{-1}(t) \to_{a.s.} H^{-1}(t)$ for any $t \in (0,1)$.

As a consequence, for all sufficiently large m and n,

$$\sqrt{\frac{mn}{N}} \inf_{x \in D_{p,m,n}} \left(\mathbb{G}_n(x) - \mathbb{F}_m(x) \right) \leq \sqrt{\frac{mn}{N}} \left(\mathbb{G}_n(x_0) - \mathbb{F}_m(x_0) \right) \ a.s.$$

However, since $(m/N)^{1/2}$ and $(n/N)^{1/2}$ are bounded by (4.4), (4.33) indicates that

$$\sqrt{\frac{mn}{N}} \left(\mathbb{G}_n(x) - \mathbb{F}_m(x) \right) \leq O(1) \left(\sqrt{n} ||\mathbb{F}_m - F||_\infty + \sqrt{n} ||\mathbb{G}_n - G||_\infty \right) - O(\sqrt{N})\delta.$$

Observe that (4.13) and (4.14) imply that

$$\sqrt{n} ||\mathbb{F}_m - F||_\infty + \sqrt{n} ||\mathbb{G}_n - G||_\infty = ||\mathbb{V}_1||_\infty + ||\mathbb{V}_2||_\infty,$$

where $\mathbb{V}_1$ and $\mathbb{V}_2$ are two Brownain bridges. Since $\mathbb{V}_1$ and $\mathbb{V}_2$ are continuous almost surely, they are also bounded on $D_{p,m,n}$ with probability one. Thus, we obtain that, as m and n approach ∞,

$$T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n) = \inf_{x \in D_{p,m,n}} \sqrt{\frac{mn}{N}} \left(\mathbb{G}_n(x) - \mathbb{F}_m(x) \right) \to_{a.s.} -\infty.$$

Next we show that $T_{2,m,n}(\mathbb{F}_m, \mathbb{G}_n) \to_p -\infty$. To this end, we first note that, if $T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n) < 0$, it follows that $T_{2,m,n}(\mathbb{F}_m, \mathbb{G}_n) < 0$. Therefore, (4.8) and (4.9) imply that, as $m, n \to \infty$,

$$T_{2,m,n}(\mathbb{F}_m, \mathbb{G}_n) < T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n) \sup_{t \in [p, 1-p]} [t(1-t)]^{-1/2} \to -\infty.$$

This completes the proof of part (A) of Theorem 4.1.

Now suppose that $G(x) > F(x)$ 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

$$\inf_{x \in D_p} (G(x) - F(x)) > 2\delta$$

for some $\delta > 0$. By continuity of H, it also follows that we can choose $p' < p$ such that $H^{-1}(p') < H^{-1}(p)$, and

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

For sufficiently large m, n, the following holds with probability one:

$$D_{p,m,n} \subset [H^{-1}(p'), H^{-1}(1-p')].$$

Therefore,

$$\liminf_{m,n\uparrow\infty} \inf_{x\in D_{p,m,n}]} \Big(G(x) - F(x)\Big) \geq \delta \quad a.s. \tag{4.34}$$

On the other hand, (4.4) and (4.33) lead to

$$\begin{aligned} & T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n) \\ & \geq \inf_{x\in\mathbb{R}} \sqrt{\frac{mn}{N}}\Big(\mathbb{G}_n(x) - G(x)\Big) - \sup_{x\in\mathbb{R}} \sqrt{\frac{mn}{N}}\Big(\mathbb{F}_m(x) - F(x)\Big) \\ & \quad + \sqrt{\frac{mn}{N}} \inf_{x\in D_{p,m,n}} \Big(G(x) - F(x)\Big). \end{aligned}$$

Note that, as $m, n \to \infty$, the above lower bound approaches infinity almost surely by (4.13), (4.14) and (4.34). Thus, we have established that $T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n) \to_{a.s.} \infty$ as $m, n \to \infty$ satisfy (4.4).

Now we will show that $T_{2,m,n}(\mathbb{F}_m, \mathbb{G}_n) \to_{a.s.} \infty$ as well. Note that the positivity of $T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n) > 0$ implies the positivity of $T_{2,m,n}(\mathbb{F}_m, \mathbb{G}_n) > 0$. Hence,

$$T_{2,m,n}(\mathbb{F}_m, \mathbb{G}_n) > T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n) \inf_{t\in[p,1-p]} (t(1-t))^{-1/2} \to_p \infty$$

as $m, n \to \infty$. Therefore, the proof of part (B) of Theorem 4.1 also follows.

Proof of Theorem 4.2

Before proving Theorem 4.2, we introduce some notation and two lemmas.

Define

$$\omega_{m,n}(x) = \sqrt{\frac{mn}{N}}\Big(\mathbb{G}_n(x) - \mathbb{F}_m(x)\Big).$$

Recall that the corresponding probabilty space is $(\Omega, \mathcal{F}, P)$. Let $A \subset \Omega$ be such that $P(A) = 1$, and as $m, n \to \infty$, the following assertions hold on A:

$$||\mathbb{F}_m - F||_\infty \to 0, \quad ||\mathbb{G}_n - G||_\infty \to 0,$$

and

$$||\sqrt{m}(\mathbb{F}_m - F) - \mathbb{V}_1||_\infty \to 0, \quad \sqrt{n}(\widehat{G}_n - G) - \mathbb{V}_2||_\infty \to 0.$$

Here $\mathbb{V}_1$ and $\mathbb{V}_2$ are as defined in (4.13) and (4.14), respectively. Also define $\omega_1 = \sqrt{\lambda}\mathbb{V}_2(G) - \sqrt{1-\lambda}\mathbb{V}_1(F)$. Several times we will use the fact that (4.4) implies the following on A:

$$\lim_{m,n\uparrow\infty} \left|\left| \omega_{m,n} - \omega_1 - \sqrt{\frac{mn}{N}}(G - F) \right|\right|_\infty = 0. \tag{4.35}$$

We now state the first lemma.

Lemma 4.6. *Suppose that the conditions of Theorem 4.2 hold. Then, the following assertion holds on A:*

$$\lim_{m,n\uparrow\infty} \inf_{x \in D_{p,m,n} \cap C} \omega_{m,n}(x) = \inf_{x \in C} \mathbb{U} \circ F(x),$$

where $\mathbb{U}$ is as defined in (4.15).

Proof of Lemma 4.6. Note that (4.35) implies that

$$\lim_{m,n\uparrow\infty} \left| \inf_{x \in D_{p,m,n} \cap C} \omega_{m,n}(x) - \inf_{x \in D_{p,m,n} \cap C} \omega_1(x) \right| = 0.$$

Also observe that $\text{dist}(D_{p,m,n}, D_p) \to 0$ as $m, n \to \infty$ on A. Also noting that ω_1 is continuous on A, we derive that $\inf_{x \in D_{p,m,n} \cap C} \omega_1(x) \to \inf_{x \in D_p \cap C} \omega_1(x)$ on A as $m, n \to \infty$. Moreover, using that $C \subset D_p$ and combining this with the above, we see that $\inf_{x \in D_{p,m,n} \cap C} \omega_{m,n}(x) \to \inf_{x \in C} \omega_1(x)$ on A as $m, n \to \infty$. The result follows from (4.15) observing the fact that, for $x \in C$, $\omega_1(x) = \mathbb{U} \circ F(x)$. □

The second lemma 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

$$\mathbb{L}_0(t) = (1-\lambda)\Big(\lambda^{-1/2} s_2(t) \mathbb{V}_1(F \circ H^{-1}(t)) - (1-\lambda)^{-1/2} s_1(t) \mathbb{V}_2(G \circ H^{-1}(t))\Big). \tag{4.36}$$

We now state the second lemma.

Lemma 4.7. *In the setting of Theorem 4.2,*

$$\sup_{t \in [p', 1-p']} \left| \sqrt{\frac{mN}{n}} \frac{\mathbb{F}_m \circ \mathbb{H}_N^{-1}(t) - t}{[t(1-t)]^{1/2}} - \sqrt{\frac{m}{n}} \frac{\mathbb{L}_0(t)}{[t(1-t)]^{1/2}} \right| \to_p 0. \tag{4.37}$$

Moreover, $\mathbb{L}_0$ is continuous almost surely.

Proof of Lemma 4.7. Theorem 4.1 and Corollary 4.1 of Pyke and Shorack (1968) indicate (4.37), where here we emphasize that (4.36) represents the corrected formula for $\mathbb{L}_0$, given by (30) of Ledwina and Wyλupek (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, H is strictly increasing, which implies that H^{-1} is a continuous function. Also note that

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

Since f, g and H^{-1} are continuous, it follows that s_1 is continuous. Similarly we can show that s_2 is continuous. Therefore, from (4.36) it is not hard to see that $\mathbb{L}_0$ is continuous almost surely. □

We now prove Theorem 4.2.

Because F and G are continuous and satisfy the alternative hypothesis, there exists an open interval inside D_p such that $G - F$ is bounded away from 0 on this interval. Therefore, $D_p \setminus C$ contains at least one interval. Also there exists a $t_0 > 0$, such that for all $t \leq t_0$, $D_p \setminus C_t$ is non-empty. In fact, the latter set also contains at least one interval.

Because H is continuous, there exists a $p' \in (0,p)$ such that $H^{-1}(p') < H^{-1}(p)$, and $G - F$ is bounded away from 0 on $D_{p'}$. For any $t \in \mathbb{R}$, denote by C_t the set $\{x \in D_{p'} : \text{dist}(x, C) < t\}$. Our discussion in the last paragraph implies that, for $t \leq t_0$, the set $D_{p'} \setminus C_t$ is non-empty. Note that this set is also closed and bounded, hence compact. Therefore, the continuous function $G - F$ attains its minima on $D_p \setminus C_t$. Hence for all $t \leq t_0$, we can find $\delta_t > 0$ such that

$$\inf_{x \in D_{p'} \setminus C_t} \Big(G(x) - F(x) \Big) > \delta_t. \tag{4.38}$$

Fix $\epsilon > 0$. Note that $\omega_1 = \sqrt{\lambda}\mathbb{V}_2(G) - \sqrt{1-\lambda}\mathbb{V}_1(F)$ is a continuous function because G, F, $\mathbb{V}_1$ and $\mathbb{V}_2$ are continuous. Note that ω_1 is also uniformly continuous on $D_{p'}$ because the latter is a compact set. Hence, we can find $\sigma_\epsilon > 0$ such that, for $x, y \in D_{p'}$, $|x - y| \leq \sigma_\epsilon$ implies that $|\omega_1(x) - \omega_1(y)| < \epsilon/2$. Moreover, we can choose σ_ϵ so that $\sigma_\epsilon < t_0$.

Note that $T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n) = \inf_{x \in D_{p,m,n}} \omega_{m,n}(x)$. Letting $B^{(1)}_{m,n,\epsilon}$ and $B^{(2)}_{m,n,\epsilon}$ denote $B^{(1)}_{m,n,\epsilon}$ and $D_{p,m,n} \cap C_{\sigma_\epsilon} \setminus C$, respectively, we see that $D_{p,m,n} = B^{(1)}_{m,n,\epsilon} \cup B^{(2)}_{m,n,\epsilon} \cup [D_{p,m,n} \cap C]$. Hence,

$$T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n) = \min \left\{ \inf_{x \in B^{(1)}_{m,n,\epsilon}} \omega_{m,n}(x), \inf_{x \in B^{(2)}_{m,n,\epsilon}} \omega_{m,n}(x), \inf_{x \in D_{p,m,n} \cap C} \omega_{m,n}(x) \right\}. \tag{4.39}$$

We will now study the asymptotic behavior of the three infimum expressions appearing in the minimum above. We will show that, as $m, n \to \infty$, the first expression diverges to infinity, and the second expression is, up to a term that tends to zero with ϵ, lower bounded by the third expression for all m, n large enough. Therefore, by the above, we will see that the behavior of $\inf_{x \in D_{p,m,n}} \omega_{m,n}(x)$ is well described by the behavior of $\inf_{x \in D_{p,m,n} \cap C} \omega_{m,n}(x)$ for all m, n large enough. The remainder of the proof will then follow from the asymptotic distribution of this quantity.

We start by showing that $\inf_{x \in B^{(1)}_{m,n,\epsilon}} \omega_{m,n}(x)$ diverges to infinity when m, n satisfy (4.4). Using (4.38) and (4.35), we deduce from (4.35) that, on A,

$$\inf_{x \in D_{p'} \setminus C_{\sigma_\epsilon}} \omega_{m,n}(x) \geq \inf_{x \in D_{p'} \setminus C_{\sigma_\epsilon}} \omega_1(x) + \sqrt{\frac{mn}{N}} \delta_{\sigma_\epsilon} - \epsilon/4$$

for all sufficiently large m and n. Noting that the continuous function ω_1 is bounded on $D_{p'}$, we deduce that $\liminf_{m,n\uparrow\infty} \inf_{x\in D_{p'}\setminus C_{\sigma_\epsilon}} \omega_{m,n}(x) = \infty$ on A. Because $D_{p,m,n} \subset D_{p'}$ for all sufficiently large m and n on A, this implies that

$$\liminf_{m,n\uparrow\infty} \inf_{x\in B^{(1)}_{m,n,\epsilon}} \omega_{m,n}(x) = \infty. \tag{4.40}$$

Now we will find the limit of the infimum of $\omega_{m,n}$ among $x \in B^{(2)}_{m,n,\epsilon}$. Let us denote the boundary of the set C by $\mathrm{bd}(C)$. To this end, note that (4.35) implies that, for any $x \in C_{\sigma_\epsilon} \setminus C$, and $y \in \mathrm{bd}(C)$, we have

$$\begin{aligned}&\left|\omega_{m,n}(x) - \omega_{m,n}(y) - \sqrt{\frac{mn}{N}}\Big(G(x) - F(x)\Big)\right| \\ &\qquad \le \epsilon/2 + \sup_{|x-y|<\sigma_\epsilon,\ x,y\in D_{p'}} |\omega_1(x) - \omega_1(y)|,\end{aligned}$$

which, by our choice of σ_ϵ, is not larger than ϵ . The above leads to

$$\inf_{x\in B^{(2)}_{m,n,\epsilon}} \omega_{m,n}(x) \ge \sqrt{\frac{mn}{N}} \inf_{x\in B^{(2)}_{m,n,\epsilon}} \Big(G(x) - F(x)\Big) + \inf_{y\in D_{p,m,n}\cap \mathrm{bd}(C)} \omega_{m,n}(y) - \epsilon.$$

Note that, for all sufficiently large m and n, $D_{p,m,n} \subset D_{p'}$ on A. Therefore, (4.38) implies that $\inf_{x\in B^{(2)}_{m,n,\epsilon}} [G(x) - F(x)] \ge 0$, which indicates that, on A,

$$\inf_{x\in B^{(2)}_{m,n,\epsilon}} \omega_{m,n}(x) \ge \inf_{y\in D_{p,m,n}\cap C} \omega_{m,n}(y) - \epsilon$$

for all sufficiently large m and n. Combining the above with (4.40), the display in (4.39) shows that, on A,

$$\inf_{y\in D_{p,m,n}\cap C} \omega_{m,n}(y) - \epsilon \le T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n) \le \inf_{y\in D_{p,m,n}\cap C} \omega_{m,n}(y)$$

for all sufficiently large m and n. As $\epsilon > 0$ was arbitrary, on A it holds that

$$\begin{aligned}\liminf_{m,n\uparrow\infty} \inf_{y\in D_{p,m,n}\cap C} \omega_{m,n}(y) &\le \liminf_{m,n\uparrow\infty} T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n) \\ &\le \limsup_{m,n\uparrow\infty} T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n) \le \limsup_{m,n\uparrow\infty} \inf_{y\in D_{p,m,n}\cap C} \omega_{m,n}(y).\end{aligned}$$

By Lemma 4.6, on A it holds that $\lim_{m,n\uparrow\infty} \inf_{y\in D_{p,m,n}\cap C} \omega_{m,n}(y)$ exists on A and is, in particular, given by $\inf_{x\in C} \mathbb{U}\circ F(x)$. Consequently, the above shows that, on A, $\lim_{m,n\uparrow\infty} T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n) = \inf_{x\in C} \mathbb{U}\circ F(x)$. Because $P(A) = 1$, it follows that (4.18) from part (A) holds.

We now show that (4.19) from part (A) holds. Because $C \subset D_p$, $x \in C$ implies that $p \leq H(x) \leq 1-p$. Because $F(x) = H(x)$ for all $x \in C$, it also holds that $F(x) \in [p, 1-p]$ for $x \in C$. Thus, $\{F(x) : x \in C\} \subseteq [p, 1-p]$, which implies that $\inf_{x\in C} \mathbb{U}\circ F(x) \geq \inf_{t\in[p,1-p]} \mathbb{U}(t)$, that is, that (4.19) holds.

Equation 4.20 from part (A) follows readily from (4.18) and (4.19), combined with the fact that convergence in distribution implies convergence of cumulative distribution functions at all continuity points.

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

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

Note that $T_{2,m,n}(\mathbb{F}_m, \mathbb{G}_n) = \inf_{t\in[p,1-p]} \nu_{m,n}(t)$.

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

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

Combining the fact that $\sup_{t\in[0,1]} \left|\mathbb{H}_N \circ \mathbb{H}_N^{-1}(t) - t\right| \leq 1/N$ (p. 762 of Pyke and Shorack, 1968) with the fact that $\sqrt{m/n} = O(1)$ by (4.4), we see that

$$\sup_{t\in[0,1]} \left|\sqrt{\frac{mn}{N}}\left(\mathbb{G}_n \circ \mathbb{H}_N^{-1}(t) - \mathbb{F}_m \circ \mathbb{H}_N^{-1}(t)\right) + \sqrt{\frac{mN}{n}}\left(\mathbb{F}_m \circ \mathbb{H}_N^{-1}(t) - t\right)\right| = o(1). \quad (4.41)$$

The above readily shows that

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

Combining the above with (4.37), 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

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

the preceding limit reduces to

$$\sup_{t\in[p,1-p]}\left|\nu_{m,n}(t)+\sqrt{\frac{\lambda}{1-\lambda}}\frac{\mathbb{L}_0(t)}{[t(1-t)]^{1/2}}-(1-\lambda)\sqrt{\frac{mN}{n}}\nu(t)\right|\to_p 0, \tag{4.42}$$

where $\nu(t)=[G\circ H^{-1}(t)-F\circ H^{-1}(t)]/[t(1-t)]^{1/2}$. 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 $T_{2,m,n}(\mathbb{F}_m,\mathbb{G}_n)-\inf_{t\in H(C)}\mathbb{U}(t)/[t(1-t)]^{1/2}$ 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 (4.42) holds almost surely and set out to prove that, as $m',n'\to\infty$,

$$T_{2,m',n'}(\mathbb{F}_{m'},\mathbb{G}_{n'})-\inf_{t\in H(C)}\frac{\mathbb{U}(t)}{\sqrt{t(1-t)}}\to_{a.s.}0. \tag{4.43}$$

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

$$\sup_{t\in[p,1-p]}\left|\nu_{m',n'}(t)+\sqrt{\frac{\lambda}{1-\lambda}}\frac{\mathbb{L}_0(t)}{[t(1-t)]^{1/2}}-(1-\lambda)\sqrt{\frac{m'N'}{n'}}\nu(t)\right|\to 0 \tag{4.44}$$

on A', where $N'=m'+n'$. We choose A' so that $\mathbb{L}_0$ is also continuous on A', which Lemma 4.7 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 1 from part (A) by the denominator of $[t(1-t)]^{1/2}$ for part (B) changes little since this new denominator is 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',

$$\sqrt{\frac{\lambda}{1-\lambda}}\left|\frac{\mathbb{L}_0(t)}{[t(1-t)]^{1/2}} - \frac{\mathbb{L}_0(t')}{(t'(1-t'))^{1/2}}\right| < \epsilon/2. \tag{4.45}$$

Note that such a $\sigma_{\epsilon'}$ exists because $[p, 1-p]$ is a compact set, and the function $\mathbb{L}_0$ is continuous on A'.

Let $b = \sup_{z \in [H^{-1}(p), H^{-1}(1-p)]} h(z)$, and note that clearly b is nonnegative. Because f and g are continuous, h is also continuous, and therefore $b < \infty$. Taking $l_\epsilon = \sigma'_\epsilon / b$, we observe that $[p, 1-p]$ can be written as the union of the following three sets:

$$\begin{aligned}
\tilde{B}^{(1)}_\epsilon &= \{t \in [p, 1-p] \ : \ H^{-1}(t) \in C^c_{l_\epsilon}\}, \\
\tilde{B}^{(2)}_\epsilon &= \{t \in [p, 1-p] \ : \ H^{-1}(t) \in C_{l_\epsilon} \setminus C\}, \\
H(C) &= \{t \in [p, 1-p] \ : \ H^{-1}(t) \in C\}.
\end{aligned}$$

The quantity l_ϵ was chosen to ensure that

$$|t - t'| \leq \sigma'_\epsilon \quad \text{for all } t \in \tilde{B}^{(2)}_\epsilon \text{ and } t' \in \mathrm{bd}(H(C)). \tag{4.46}$$

We now that the above display is indeed valid. Fix $t \in \tilde{B}^{(2)}_\epsilon$ and $t' \in \mathrm{bd}(H(C))$. Because H^{-1} is continuous, it holds that $H^{-1}(t) \in \mathrm{bd}(C)$. Thus, $|H^{-1}(t) - H^{-1}(t')| \leq l_\epsilon$. Note that the continuity of H also implies that $H(t_0) = t_0$ 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, $|t - t'| \leq b|H^{-1}(t) - H^{-1}(t')|$. Combining this display

with the fact that $|H^{-1}(t) - H^{-1}(t')| \leq l_\epsilon$ and plugging in the choice of l_ϵ shows that (4.46) indeed holds.

Because $\nu(t) \geq 0$, for $t \in [p, 1-p]$, one can show that the infimum of ν over $\tilde{B}_\epsilon^{(1)}$ is bounded away from 0. Therefore, using (4.44) and imitating the proof of (4.40), we can show that, as $m', n' \to \infty$,

$$\inf_{t \in \tilde{B}_\epsilon^{(1)}} \nu_{m',n'}(t) \to \infty \quad \text{on } A'. \tag{4.47}$$

Next let us consider $t \in \tilde{B}_\epsilon^{(2)}$ and $t' \in \mathrm{bd}(H(C))$. Note that (4.42) and (4.46) imply that the following holds on A':

$$\begin{aligned} &\left| \nu_{m',n'}(t) - \nu_{m',n'}(t') - (1-\lambda)\sqrt{\frac{m'N'}{n'}}\nu(t) \right| \\ &\quad \leq o(1) - \sqrt{\frac{\lambda}{1-\lambda}} \sup_{t,t' \in S_t} \left| \frac{\mathbb{L}_0(t)}{[t(1-t)]^{1/2}} - \frac{\mathbb{L}_0(t')}{(t'(1-t'))^{1/2}} \right|, \end{aligned}$$

where $S_t = \{t, t' \in [p, 1-p] \ : \ |t - t'| < \sigma'_\epsilon\}$. Equation 4.45 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 $t \in \tilde{B}_\epsilon^{(2)}$ and $t' \in \mathrm{bd}(H(C))$, we obtain that

$$\inf_{t \in \tilde{B}_\epsilon^{(2)}} \nu_{m',n'}(t) \geq \inf_{t \in H(C)} \nu_{m',n'}(t) + \sqrt{\frac{m'N'}{n'}} \inf_{t' \in \tilde{B}_\epsilon^{(2)}} \nu(t') - \epsilon.$$

The fact that ν is non-negative on $[p, 1-p]$ yields that $\inf_{t' \in \tilde{B}_\epsilon^{(2)}} \nu(t') \geq 0$. Thus, the above shows that $\inf_{t \in \tilde{B}_\epsilon^{(2)}} \nu_{m',n'}(t) \geq \inf_{t \in H(C)} \nu_{m',n'}(t) - \epsilon$. Therefore, using (4.47), we derive that, for sufficiently large m' and n',

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

As $\epsilon > 0$ was arbitrary, the above shows that $\inf_{t\in[p,1-p]} \nu_{m',n'}(t) - \inf_{t\in H(C)} \nu_{m',n'}(t)$ converges to zero as $m', n' \to \infty$ on A'. Finally, (4.36) and (4.44) yield that

$$\sup_{t\in H(C)} \left| \nu_{m',n'}(t) - \frac{\mathbb{U}(t)}{\sqrt{t(1-t)}} \right| \to 0 \quad \text{on } A'.$$

Recall that A' was chosen to satisfy $P(A') = 1$. Thus, the above convergence holds with probability one, from which, (4.43) follows. As was discussed above (4.43), the fact that this equation holds implies that (4.21) from part (B) is satisfied. It is straightforward to see that the limiting term on the right hand side of the last display is stochastically greater than $\inf_{t\in[p,1-p]} \mathbb{U}(t)/\sqrt{t(1-t)}$, from which (4.22) follows.

Proof of Lemma 4.1

By similar arguments to those used to prove Lemma 4.7, it holds that

$$\sup_{t\in[1/N,1]} \left| \sqrt{\frac{mN}{n}} \left(\mathbb{F}_m \circ \mathbb{H}_N^{-1}(t) - F \circ H^{-1} \right) - \sqrt{\frac{m}{n}} \mathbb{L}_0(t) \right| \to_p 0,$$

where $\mathbb{L}_0$ was defined in (4.36). Also recall that Lemma 4.7 showed that $\mathbb{L}_0$ is continuous. Also, from (4.36) it follows that, in our case, $\mathbb{L}_0(t)$ takes the form

$$(1-\lambda) \left(\lambda^{-1/2} \mathbb{V}_1(t) - (1-\lambda)^{-1/2} \mathbb{V}_2(t) \right),$$

where $\mathbb{V}_1$ and $\mathbb{V}_2$ are two independent Brownian bridges. Thus, (4.8.1) reduces to

$$\sup_{t\in[1/N,1]} \left| \sqrt{\frac{mN}{n}} \left(\mathbb{F}_m \circ \mathbb{H}_N^{-1}(t) - t \right) + \mathbb{U}(t) \right| \to_p 0. \tag{4.48}$$

Observe that the following holds for any two real functions f_1 and f_2 and any set $A \subset \mathbb{R}$:

$$\inf_{x\in A} f_1(x) - \inf_{x\in A} f_2(x) \leq \sup_{x\in A} (f_2(x) - f_1(x)). \tag{4.49}$$

The above inequality, combined with (4.41) and (4.48), establishes that, as $m, n \to \infty$, $T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n)$ converges to $\inf_{t\in[p,1-p]} \mathbb{U}(t)$ in probability. To show the same for $T_{2,m,n}(\mathbb{F}_m, \mathbb{G}_n)$,

we note that, for $p \in (0,1)$,

$$\sup_{t\in[p,1-p]} \left| t^{-1/2}(1-t)^{-1/2} \left\{ \sqrt{\frac{mn}{N}} \left(\mathbb{G}_n \circ \mathbb{H}_N^{-1}(t) - \mathbb{F}_m \circ \mathbb{H}_N^{-1}(t) \right) - \mathbb{U}(t) \right\} \right|$$
$$\leq p^{-1} \sup_{t\in[p,1-p]} \left| \sqrt{\frac{mn}{N}} \left(\mathbb{G}_n \circ \mathbb{H}_N^{-1}(t) - \mathbb{F}_m \circ \mathbb{H}_N^{-1}(t) \right) - \mathbb{U}(t) \right| \to_p 0$$

as $m, n \to \infty$. Therefore, it follows that

$$T_{2,m,n}(\mathbb{F}_m, \mathbb{G}_n) \to_p \inf_{t\in[p,1-p]} \frac{\mathbb{U}(t)}{(t(1-t))^{1/2}},$$

which completes the proof.

Proof of Theorem 4.3

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. Let us denote the distribution function of $\widehat{f}_m^0$ by $\widehat{F}_m^0$. Note that $\widehat{F}_m^0$ can be expressed as

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

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 is 0. Hence, there is no ambiguity in the above definition of $\widehat{F}_m^0$.

Note that we can write $\widehat{F}_m^{0,-} = \mathcal{M}(\mathbb{F}_m^-)$, where $\mathcal{M}$ is the LCM operator on $\mathbb{R}$ (see Definition 1). From (4.13), we recall that $\sqrt{m}(\mathbb{F}_m^- - F^-)$ converges to the Brownian bridge

process $\mathbb{V}_1 \circ F$ almost surely. Using the above, we can expand $\sqrt{m}||\widehat{F}_m^{0,-} - \mathbb{F}_m^-||_\infty$ as follows:

$$\begin{aligned}
&\sqrt{m}||\widehat{F}_m^{0,-} - \mathbb{F}_m^-||_\infty \\
&= \left|\left|\left(\sqrt{m}\left(\mathcal{M}(\mathbb{F}_m^-) - F^-\right) - \mathbb{V}_1 \circ F^-\right) - \left(\sqrt{m}(\mathbb{F}_m^- - F^-) - \mathbb{V}_1 \circ F^-\right)\right|\right|_\infty \\
&\leq \left|\left|\frac{\mathcal{M}(F^- + m^{-1/2}\sqrt{m}(\mathbb{F}_m^- - F^-)) - F^-}{m^{-1/2}} - \mathbb{V}_1 \circ F^-\right|\right|_\infty \\
&\quad + ||\sqrt{m}(\mathbb{F}_m^- - F^-) - \mathbb{V}_1 \circ F^-||_\infty \\
&= \left|\left|\frac{\mathcal{M}(F^- + m^{-1/2}\sqrt{m}(\mathbb{F}_m^- - F^-)) - F^-}{m^{-1/2}} - \mathbb{V}_1 \circ F^-\right|\right|_\infty + o(1) \qquad (4.50)
\end{aligned}$$

almost surely, where the last equality follows by (4.13). In order to show that the right hand side of (4.50) approaches 0 almost surely, we invoke Proposition 2.2 of Beare *et al.* (2017). This proposition states that $\mathcal{M}$ is Hadamard directionally differentiable at F^-, tangentially to C_0, with respect to the $||\cdot||_\infty$ norm. Because F^- is strictly concave by Condition A, it also follows that the corresponding directional derivative $\dot{\mathcal{M}}_{F^-}$ at F^- is the identity map on $\mathcal{C}_0$, i.e. $\dot{\mathcal{M}}_{F^-}(\mu) = \mu$ for all functions $\mu \in C_0$ (Beare *et al.*, 2017, Remark 2.5 of). Noting that $\sqrt{m}(\mathbb{F}_m^- - F^-) = O_p(1)$, and applying Proposition 2.2 of Beare *et al.* (2017), we derive that, as $m \to \infty$,

$$\left|\left|\frac{\mathcal{M}(F^- + n^{-1/2}\sqrt{m}(\widehat{F}_m^{0,-} - F^-)) - \mathcal{M}(F^-)}{m^{-1/2}} - \mathbb{V}_1 \circ F^-\right|\right|_\infty \to_{a.s.} 0.$$

Now since F is concave, we have $\mathcal{M}(F^-) = F^-$. Thus the last display combined with (4.50), leads to

$$\sqrt{m}||\widehat{F}_m^{0,-} - \mathbb{F}_m^-||_\infty \to_{a.s.} 0.$$

Similarly one can also show that $\sqrt{m}||\widehat{F}_m^{0,+} - \mathbb{F}_m^+||_\infty \to_{a.s.} 0$ as $m \to \infty$.

Now note that the empirical distribution of the X_i's can be expressed as

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

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

$$\sqrt{m}||\widehat{F}_m^0 - \mathbb{F}_m||_\infty \leq \hat{\alpha}_m\sqrt{m}||\widehat{F}_m^{0,-} - \mathbb{F}_m^-||_\infty + (1 - \hat{\alpha}_m)\sqrt{m}||\widehat{F}_m^{0,+} - \mathbb{F}_m^+||_\infty \to_{a.s.} 0. \qquad (4.51)$$

To prove part (A) of the current theorem, now we invoke Theorem 1 of Birgé (1997), which states that

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

where η is as defined in Section 4.3, which implies that, in our case, $\eta=O(1/m)$. This, combined with (4.51), proves that the right hand side of the above display approaches 0 almost surely. Thus part (A) of the current theorem is proved.

Now note that since

$$\sqrt{m}||\widehat{F}_m-\mathbb{F}_m||_\infty\leq\sqrt{m}||\widehat{F}_m-\widehat{F}_m^0||_\infty+\sqrt{m}||\widehat{F}_m^0-\mathbb{F}_m||_\infty,$$

and

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

part (B) of the current theorem follows by part (A) and (4.51).

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

$$\begin{aligned}&\left|\left|\sqrt{m}(\widehat{F}_m-F)-\mathbb{V}_1\circ F\right|\right|_\infty\\&\quad\leq\left|\left|\sqrt{m}(\widehat{F}_m-\mathbb{F}_m)\right|\right|_\infty+\left|\left|\sqrt{m}(\mathbb{F}_m-F)-\mathbb{V}_1\circ F\right|\right|_\infty,\end{aligned}$$

which converges to 0 as $m\to\infty$ by part (B) of the current theorem and (4.13).

Proof of Lemma 4.2

Note that (4.49) implies that

$$\begin{aligned}&T_{1,m,n}(\widehat{F}_m,\widehat{G}_n)-T_{1,m,n}(\mathbb{F}_m,\mathbb{G}_n)\\&\leq\left(\frac{mn}{N}\right)^{1/2}\sup_{z\in[p,1-p]}\Big(\widehat{F}_m(\mathbb{H}_N^{-1}(z))-\mathbb{F}_m(\mathbb{H}_N^{-1}(z))\\&\quad-\widehat{G}_n(\mathbb{H}_N^{-1}(z))+\mathbb{G}_n(\mathbb{H}_N^{-1}(z))\Big)\\&\leq\left(\frac{mn}{N}\right)^{1/2}\Big(\sup_{x\in\mathbb{R}}|\widehat{F}_m(x)-\mathbb{F}_m(x)|+\sup_{x\in\mathbb{R}}|\widehat{G}_n(x)-\mathbb{G}_n(x)|\Big),\end{aligned}$$

which converges to 0 almost surely by part (B) of Theorem 4.3 and (4.4). Similarly one can show that $T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n) - T_{1,m,n}(\widehat{F}_m, \widehat{G}_n) \to_{a.s.} 0$, implying

$$|T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n) - T_{1,m,n}(\widehat{F}_m, \widehat{G}_n)| \to_{a.s.} 0.$$

Using part (B) of Theorem 4.3 and (4.4) in the second step, we also deduce that

$$\begin{aligned}
&|T_{2,m,n}(\widehat{F}_m, \widehat{G}_n) - T_{2,m,n}(\mathbb{F}_m, \mathbb{G}_n)| \\
&\quad \leq \left(\frac{mn}{N}\right)^{1/2} \frac{\sup_{x\in\mathbb{R}} |\widehat{F}_m(x) - \mathbb{F}_m(x)| + \sup_{x\in\mathbb{R}} |\widehat{G}_n(x) - \mathbb{G}_n(x)|}{\inf\limits_{z\in[p,1-p]} \sqrt{z(1-z)}} \\
&\quad \leq \frac{o(1)}{\sqrt{p(1-p)}} \quad a.s.
\end{aligned}$$

It remains to prove that $|T_{3,m,n}(\widehat{F}_m, \widehat{G}_n) - T_{3,m,n}(\mathbb{F}_m, \mathbb{G}_n)|$ converges to 0 almost surely. To this end, we first note that $\xi(F, G) = \int F dG$ is a Hadamard differentiable function with respect to the norm $||\cdot||_\infty$ for every pair of distribution functions (F, G) (see Section 5, pages 362 - 371 Lehmann, 1975), with the derivative

$$\dot{\xi}(\mu_X, \mu_Y; F, G) = \int \mu_X dG - \int \mu_Y dF.$$

Observe that we can write

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

where

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

Note that part (C) of Theorem 4.3 and (4.14) imply that, as $N \to \infty$,

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

and

$$\left(\frac{mn}{N(N+1)}\right)^{1/2} \to \sqrt{\lambda(1-\lambda)}.$$

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 (4.13), (4.14), and (4.4), 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 $T_{3,m,n}(\widehat{F}_m, \widehat{G}_n)$ follows noting

$$\begin{aligned} &12^{-1/2}|T_{3,m,n}(\widehat{F}_m, \widehat{G}_n) - T_{3,m,n}(\mathbb{F}_m, \mathbb{G}_n)| \\ &\quad \le \left|\left(\frac{mn}{N+1}\right)^{1/2}\left(\xi(\widehat{F}_m, \widehat{G}_n) - \xi(F,G)\right) - \mathbb{Y}\right| \\ &\qquad + \left|\left(\frac{mn}{N+1}\right)^{1/2}\left(\xi(\mathbb{F}_m, \mathbb{G}_n) - \xi(F,G)\right) - \mathbb{Y}\right|, \end{aligned}$$

which converges to 0 almost surely.

Proof of Lemma 4.3

Theorem 4.4 of Dümbgen and Rufibach (2009), 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 (4.6). We note that, for sufficiently large N, $D_{p,m,n} \subset D'_p$ with probability one. Therefore, (4.26) implies that

$$D_{p,m,n} \subset \text{int}(\text{dom}(\log f)) \cap \text{int}(\text{dom}(\log g)) \quad a.s.,$$

which indicates that the following holds with probability one,

$$\sup_{u\in D_{p,m,n}} |\tilde{F}_m(u) - \mathbb{F}_m(u)| + \sup_{x\in D_{p,m,n}} |\tilde{G}_n(x) - \mathbb{G}_n(x)| = o_p(m^{-1/2}) + o_p(n^{-1/2}), \tag{4.52}$$

which is $o_p(N^{-1/2})$ by (4.4). Now using (4.49), we derive

$$\begin{aligned}
&T_{1,m,n}(\tilde{F}_m, \tilde{G}_n) - T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n) \\
&\leq \left(\frac{mn}{m+n}\right)^{1/2} \sup_{z\in[p',1-p']} \Big(\tilde{F}_m(\mathbb{H}_N^{-1}(z)) - \mathbb{F}_m(\mathbb{H}_N^{-1}(z)) \\
&\quad - \tilde{G}_n(\mathbb{H}_N^{-1}(z)) + \mathbb{G}_n(\mathbb{H}_N^{-1}(z))\Big) \\
&\leq \left(\frac{mn}{m+n}\right)^{1/2} \Big(\sup_{x\in D_{p,m,n}} |\tilde{F}_m(x) - \mathbb{F}_m(x)| + \sup_{x\in D_{p,m,n}} |\tilde{G}_n(x) - \mathbb{G}_n(x)|\Big),
\end{aligned}$$

which is $o_p(1)$ by (4.52). Similarly one can show that $T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n) - T_{1,m,n}(\tilde{F}_m, \tilde{G}_n) = o_p(1)$, which implies $|T_{1,m,n}(\tilde{F}_m, \tilde{G}_n) - T_{1,m,n}(\mathbb{F}_m, \mathbb{G}_n)|$ converges to 0 in probability. Analogously, we can show that

$$|T_{2,m,n}(\tilde{F}_m, \tilde{G}_n) - T_{2,m,n}(\mathbb{F}_m, \mathbb{G}_n)| \to_p 0.$$

4.8.2 Proofs of Section 4.5

Although the aim of the current section is to derive the asymptotic distribution of $H^2(\widehat{f}_m, \widehat{g}_n)$, we embark on proving a more general result on plug-in estimators of integrated functionals. Theorem 4.4 then follows as a special case.

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

$$T(f,g) = \int_{-\infty}^{\infty} v\Big(f(x), g(x)\Big)dx, \tag{4.53}$$

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

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

We have already mentioned in Section 4.5 that the Von Mises Expansion (VME) plays a critical role in the proofs of this section. So now we will elaborate a little bit on the VME of T. 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 (4.28)) 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 $f_1, f_2, g_1, g_2 \in \mathcal{P}_1$:

$$\begin{aligned} T(f_2,g_2) = T(f_1,g_1) + \int_{-\infty}^{\infty} \psi_f(x; f_1, g_1) f_2(x)dx + \int_{-\infty}^{\infty} \psi_g(x; f_1, g_1) g_2(x)dx \\ + O(||f_1 - f_2||_2^2) + O(||g_1 - g_2||_2^2). \end{aligned} \tag{4.54}$$

The first order VME implies that T can be written as a linear term plus second order bias term.

Let $\bar{f}_m$ and $\bar{g}_n$ be estimators of f and g based on samples of size m and n, respectively. We denote the corresponding distribution functions by $\bar{F}_m$ and $\bar{G}_n$. We aim to show that if $\bar{f}_m$ and $\bar{g}_n$ satisfy some regularity conditions, then the plug-in estimator $T(\bar{f}_m, \bar{g}_n)$ is $\sqrt{N}$-consistent for estimating $T(f,g)$. The first condition we require is related to the weak convergence of $\sqrt{m}(\bar{F}_m - F)$ and $\sqrt{n}(\bar{G}_n - G)$ to Brownian bridge processes.

Condition C1. *The distribution functions $\bar{F}_m$ and $\bar{G}_n$ corresponding to density estimators $\bar{f}_m$ and $\bar{g}_n$ satisfy*

$$\begin{aligned} \sqrt{m}(\bar{F}_m - F) &\to_d \mathbb{V}_1(F), \\ \sqrt{n}(\bar{G}_n - G) &\to_d \mathbb{V}_2(G), \end{aligned}$$

where $\mathbb{V}_1$ *and* $\mathbb{V}_2$ *are Brownian bridges.*

Grenander estimator of monotone and unimodal densities (Beare *et al.*, 2017) satisfy the weak convergences in Condition C1. Theorem 4.3 implies that the unimodal density estimator of Birgé (1997) also satisfies this condition. Kernel density estimators with either the fixed bandwidth, or the optimal bandwidth for minimizing MISE (using the least square cross validation or the plug-in bandwidth method) also satisfies Condition C1 (Giné and Nickl, 2008, Theorem 2(a) of).

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

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

$$O_p(||\bar{f}_m - f||_2^2) = o_p(m^{-1/2}) \quad \textit{and} \quad O_p(||\bar{g}_n - g||_2^2) = o_p(n^{-1/2}). \tag{4.55}$$

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

$$O_p(||\bar{f}_m - f||_2^2) = H^2(\bar{f}_m, f)O_p(1).$$

If $\bar{f}_m$ also satisfies $H^2(\bar{f}_m, f) = o_p(m^{-1/2})$, Condition C2 follows. For example, the log-concave MLE density estimators (Doss and Wellner, 2016) and the Grenander estimators of monotone density (Theorem 7.12 of van de Geer, 2000) satisfy Condition C2. When f is in Hölder class of order 2, loosely speaking, which is same as f having bounded second derivative, Condition C2 is satisfied with a KDE with bandwidth $O(n^{-1/5})$. The cross validated and the plug-in optimal bandwidth satisfies decays at the rate $O(n^{-1/5})$.

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 the 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 Ambrosio *et al.*, 2000), which means

$$\int_{-\infty}^{\infty} \mu(x)\varrho'(x)dx = \int_{-\infty}^{\infty} \mu'(x)\varrho(x)dx$$

for all continuously differentiable function ϱ.

Condition I. *The maps*

$$x \mapsto \psi_f(x; f, g) \quad \text{and} \quad x \mapsto \psi_g(x; f, g)$$

are of bounded total variation on $\mathbb{R}$.

Condition I is stronger than ψ_f and ψ_g being bounded, but this condition ensures the existence of some Lebesgue-Stieltjes integrals required in the proof of Theorem 4.5. A similar condition was used by Mukherjee and Sen (2018) who also considered a plug-in estimator to estimate smooth integrated functionals. Condition I also circumvents the need of imposing further conditions on the influence functions. For example, we do not require any Lipschitz-type condition similar to Assumption 4 of Kandasamy *et al.* (2015).

Suppose that $\bar{f}_m$ and $\bar{g}_n$ satisfy Conditions C1 and C2, and the influence functions ψ_f and ψ_g satisfy Condition I. Under this scenario, our next theorem obtains the asymptotic distribution of the estimator $T(\bar{f}_m, \bar{g}_n)$, which is a centered gaussian random variable. Related literature (cf. Kandasamy *et al.*, 2015; Van der Vaart, 1998) implies that the variance $\sigma^2_{f,g}$ agrees with the asymptotic lower bound for this case under the nonparametric model.

Theorem 4.5. *Suppose that* $\mathcal{P}_1 \subset \mathcal{P}$. *Let* $T : \mathcal{P}_1^2 \mapsto \mathbb{R}$ *be a functional of the form* (4.53) *satisfying the first order VME in* (4.54). *Consider* $f, g \in \mathcal{P}_1$. *We assume that the influence functions* ψ_f *and* ψ_g *defined in* (4.28) *satisfy Condition I . Suppose that* $\bar{f}_m$ *and* $\bar{g}_n$ *are some estimators of* f *and* g *based on samples of size* m *and* n, *respectively. Let us denote* $N = m + n$. *Suppose that* m *and* n *satisfy* (4.4). *Further suppose that* $\bar{f}_m$ *and* $\bar{g}_n$ *satisfy Conditions C1 and C2, and*

$$\lim_{m\to\infty} P\left(\bar{f}_m \in \mathcal{P}_1\right) = 1, \quad \text{and} \quad \lim_{n\to\infty} P\left(\bar{g}_n \in \mathcal{P}_1\right) = 1. \tag{4.56}$$

Then we have

$$\sqrt{N}\Big(T(\bar{f}_m, \bar{g}_n) - T(f,g)\Big) \to_d N(0, \sigma^2_{f,g}),$$

where

$$\sigma^2_{f,g} = \lambda^{-1}\int_{-\infty}^{\infty}\psi_f(x;f,g)^2 f(x)dx + (1-\lambda)^{-1}\int_{-\infty}^{\infty}\psi_g(x;f,g)^2 g(x)dx.$$

Proof. We observe that (4.56) implies

$$P\Big(\bar{f}_m, \bar{g}_n \in \mathcal{P}_1\Big) \to 1, \qquad \text{as } m,n \to \infty.$$

Since T satisfies the first order VME, (4.54) indicates that

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

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

$$\int_{-\infty}^{\infty}\psi_f(x;f,g)f(x)dx = 0.$$

hence we can write

$$\int_{-\infty}^{\infty}\psi_f(x;f,g)\bar{f}_m(x)dx = \int_{-\infty}^{\infty}\psi_f(x;f,g)d(\bar{F}_m(x) - F(x)).$$

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

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

since $\psi_f(\cdot; f, g)$ is of bounded total variation, it is also bounded, leading to

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

Therefore, we deduce that

$$\int_{-\infty}^{\infty} \psi_f(x; f, g) \bar{f}_m(x) dx = -\int_{-\infty}^{\infty} (\bar{F}_m(x) - F(x)) d\psi_f(x; f, g).$$

Similarly we can show that

$$\int_{-\infty}^{\infty} \psi_g(x; f, g) \bar{g}_n(x) dx = -\int_{-\infty}^{\infty} (\bar{G}_n(x) - G(x)) d\psi_g(x; f, g).$$

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

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

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^∞ is the set of all bounded functions on $\mathbb{R}$. Since $m/N \to \lambda$, Slutsky's Theorem yields

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

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

$$(\mu_1, \mu_2) \mapsto \int_{-\infty}^{\infty} \mu_1(x)\ d\psi_f(x; f, g) + \int_{-\infty}^{\infty} \mu_2(x)\ d\psi_g(x; f, g)$$

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

$$\begin{aligned}
&\int_{-\infty}^{\infty} \sqrt{N}(\bar{F}_m(x) - F(x)) d\psi_f(x; f, g) \\
&\qquad + \int_{-\infty}^{\infty} \sqrt{N}(\bar{G}_n(x) - G(x)) d\psi_g(x; f, g) \\
&\to_d (1-\lambda)^{-1/2} \int_{-\infty}^{\infty} \mathbb{U}(F(x)) d\psi_f(x; f, g) + \lambda^{-1/2} \int_{-\infty}^{\infty} \mathbb{U}(G(x)) d\psi_g(x; f, g).
\end{aligned}$$

Now for any continuous distribution function F, and for any function μ with finite total variation, the random variable

$$\mathbb{Y} = \int_{-\infty}^{\infty} \mathbb{U}(F(x))dh(x) \sim N(0, \sigma_\mu^2),$$

where

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

The above follows from the proof of Theorem 2.3 of Mukherjee and Sen (2018). Therefore,

$$(1-\lambda)^{-1/2} \int_{-\infty}^{\infty} \mathbb{U}(F) d\psi_f(x; f, g) + \lambda^{-1/2} \int_{-\infty}^{\infty} \mathbb{U}(G) d\psi_g(x; f, g) \stackrel{d}{=} N(0, \sigma_{f,g}^2),$$

which completes the proof. □

Now we focus on the special case at hand, i.e. $T(f,g) = H^2(f,g)$. Towards this end, our first task is to establish the existence of the first order VME. Recall the definition of $\mathcal{P}(b, B)$ from (4.32). Our next lemma ensures the existence of the first order VME for $T(f,g) = H^2(f,g)$ when $\mathcal{P}_1 = \mathcal{P}(b, B)$. Lemma 4.8 follows from Lemma 10 of Kandasamy *et al.* (2015), which establishes the existence of the first order VME for a large class of divergences including the Hellinger distance.

Lemma 4.8. *Let $0 < b < B < \infty$. Define the map $T : \mathcal{P}(b, B)^2 \mapsto \mathbb{R}$ by*

$$T(f,g) = H^2(f,g).$$

Then the first order VME in (4.54) holds for T for any $b, B > 0$.

Our next step is to obtain the asymptotic distribution of $H^2(\bar{f}_m, \bar{g}_n)$ when $\bar{f}_m$ and $\bar{g}_n$ are $\widehat{f}_m^0$ and $\widehat{g}_n^0$, i.e. the Grenander estimators of f and g based on the true modes. When $f, g \in \mathcal{P}(b, B)$ are continuous and unimodal, Corollary 6 shows that the conditions of Theorem 4.5 are satisfied for the above case. Observe that Corollary 6 combined with Lemma 4.4 implies Theorem 4.4, and thus establishes the asymptotic distribution of $H^2(\widehat{f}_m, \widehat{g}_n)$ as well.

Corollary 6. *Let f and g be continuous unimodal densities in $\mathcal{P}(b, B)$ for some $b, B > 0$. 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 that $m/(m+n) \to \lambda$. Then*

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

where

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

with ψ_f and ψ_g as in (4.30) and (4.31), respectively.

Proof. First note that, since f is continuous, it follows that (cf. Balabdaoui *et al.*, 2011)

$$||\widehat{f}_m^0 - f||_\infty \to_{a.s.} 0, \quad \text{as } m \to \infty,$$

which implies that

$$P\left(\widehat{f}_m^0 \in \mathcal{P}(b-\epsilon, B+\epsilon)\right) \to 1, \quad \text{for any } \epsilon \in (0, b).$$

Similarly we can show that

$$P\left(\widehat{g}_n^0 \in \mathcal{P}(b-\epsilon, B+\epsilon)\right) \to 1, \quad \text{for any } \epsilon \in (0, b).$$

Fix $\epsilon \in (0, b)$. We will show that the conditions of Theorem 4.5 are satisfied for $\mathcal{P}_1 = \mathcal{P}(b-\epsilon, B+\epsilon)$. The last two displays indicate (4.56) for $\widehat{f}_m^0$ and $\widehat{g}_n^0$. Also notice that Lemma 4.8 implies the first order VME of the function $H^2 : \mathcal{P}_1(b-\epsilon, B+\epsilon)^2 \mapsto \mathbb{R}$ for any $\epsilon \in (0, b)$. Condition I also follows in a straightforward way once we note that, when $f, g \in \mathcal{P}(b-\epsilon, B+\epsilon)$, (4.30) and (4.31) indicate that $\psi_f(\cdot; f, g)$ and $\psi_g(\cdot; f, g)$ are differentiable functions with integrable derivatives. Condition C1 follows from (4.13), (4.14), and (4.51).

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 exactly similar. Observe that

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

Using Theorem 7.12 of van de Geer (2000) one can show that $H^2(\widehat{f}_m^0, f) = O_p(n^{-2/3})$. Since we have already established (4.56) for $\mathcal{P}_1 = \mathcal{P}(b-\epsilon, B+\epsilon)$, we also have $||\widehat{f}_m^0||_\infty = O_p(1)$. Thus Condition $C2$ also follows. □

Finally, we will prove Lemma 4.4, which will complete the proof of Theorem 4.4.

Proof of Lemma 4.4

Proof. First we will show that, to prove Lemma 4.4, it suffices to show that $H^2(\widehat{f}_m, \widehat{f}_m^0)$ and $H^2(\widehat{g}_n, \widehat{g}_n^0)$ are $o_p(n^{-1/2})$. To this end, observe that the triangle inequality of the Hellinger metric indicates

$$\left| H(\widehat{f}_m, \widehat{g}_n) - H(\widehat{f}_m^0, \widehat{g}_n^0) \right| \leq H(\widehat{f}_m, \widehat{f}_m^0) + H(\widehat{g}_n, \widehat{g}_n^0) = o_p(n^{-1/4}), \tag{4.57}$$

which leads to

$$\left| H(\widehat{f}_m, \widehat{g}_n)^2 - H(\widehat{f}_m^0, \widehat{g}_n^0)^2 \right| \leq \left(H(\widehat{f}_m, \widehat{g}_n) + H(\widehat{f}_m^0, \widehat{g}_n^0) \right) o_p(n^{-1/4}).$$

However, Theorem 4.5 entails that $H(\widehat{f}_m^0, \widehat{g}_n^0)$ is $O_p(n^{-1/4})$, which, combined with (4.57) implies that $H(\widehat{f}_m, \widehat{g}_n)$ is also $O_p(n^{-1/4})$. Therefore, the proof of Lemma 4.4 follows from the above display. Thus it is enough to show that $H^2(\widehat{f}_m, \widehat{f}_m^0)$ and $H^2(\widehat{g}_n, \widehat{g}_n^0)$ are $o_p(n^{-1/2})$. We will prove this fact only for $H^2(\widehat{f}_m, \widehat{f}_m^0)$ 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 (1997) entails that, in this case, there exist $a, b, c \in \mathbb{R}$ such that

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

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

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 $x \in (c, b)$.

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

$$\begin{aligned}
2H^2(\widehat{f}_m, \widehat{f}^0_m) &= \int_{-\infty}^{\infty} \left(\sqrt{\widehat{f}_m(x)} - \sqrt{\widehat{f}^0_m(x)} \right)^2 dx \\
&= \int_a^b \left(\sqrt{\widehat{f}_m(x)} - \sqrt{\widehat{f}^0_m(x)} \right)^2 dx \\
&= \int_a^b \widehat{f}_m(x)dx + \int_a^b \widehat{f}^0_m(x)dx - 2\int_a^b \sqrt{\widehat{f}_m(x)\widehat{f}^0_m(x)}dx \\
&\leq \widehat{F}_m(b) - \widehat{F}_m(a) + \widehat{F}^0_m(b) - \widehat{F}^0_m(a) - 2\int_a^c \widehat{f}_m(x)dx - 2\int_c^b \widehat{f}^0_m(x)dx \\
&= \widehat{F}_m(b) - \widehat{F}_m(a) + \widehat{F}^0_m(b) - \widehat{F}^0_m(a) - 2\widehat{F}_m(c) \\
&\quad + 2\widehat{F}_m(a) - 2\widehat{F}^0_m(b) + 2\widehat{F}^0_m(c) \\
&= 2\left(\widehat{F}^0_m(c) - \widehat{F}_m(c) \right),
\end{aligned}$$

where the last step follows because $\widehat{F}^0_m(x) = \widehat{F}_m(x)$ for $x = a, b$. Hence, we observe that

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

which is $o_p(m^{-1/2})$ by Theorem 4.3(A). Hence the proof of Lemma 4.4 follows for this case.

Now suppose that $\widehat{M}_m < M$. Then from the proof of Lemma 1 in Birgé (1997), one can prove the existence of $a, b, c \in \mathbb{R}$ such that

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

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

3. $\widehat{f}_m(x) \geq \widehat{f}^0_m(x)$ for $x \in (a, c)$, and $\widehat{f}_m(x) \leq \widehat{f}^0_m(x)$ for $x \in (c, b)$.

Then in the same way as in the case of $\widehat{M}_m > M$, we can show that

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

which completes the proof of the current lemma. □

Proof of Theorem 4.3

Theorem 4.3 follows from Lemma 4.4 and Theorem 4.5.

BIBLIOGRAPHY

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

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

An, M. Y. (1998). Logconcavity versus logconvexity: a complete characterization. *J. Econom. Theory*, **80**(2), 350–369.

Bagnoli, M. and Bergstrom, T. (2005). Log-concave probability and its applications. *Econom. Theory*, **26**(2), 445–469.

Balabdaoui, F., Jankowski, H., Pavlides, M., Seregin, A., and Wellner, J. (2011). On the grenander estimator at zero. *Statist. Sinica*, **21**, 873.

Barlow, R. E. and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing*. Holt, Rinehart and Winston, Inc., New York-Montreal, Que.-London. Probability models, International Series in Decision Processes, Series in Quantitative Methods for Decision Making.

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.

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

Beran, R. (1974). Asymptotically efficient adaptive rank estimates in location models. *Ann. Statist.*, **2**, 63–74.

Bhattacharyya, S. (2013). *A Study of High-dimensional Clustering and Statistical Inference on Networks*. Ph.D. thesis, UC Berkeley.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1998). *Efficient and Adaptive Estimation for Semiparametric Models*. Springer-Verlag, New York.

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

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

Bobkov, S. and Ledoux, M. (2017). *One-dimensional empirical measures, order statistics, and Kantorovich transport distances*. Memoirs of the American Mathematical Society. American Mathematical Society.

Borell, C. (1975). Convex set functions in d-space. *Period. Math. Hungar.*, **6**(2), 111–136.

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

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press, Cambridge.

Brascamp, H. J. and Lieb, E. H. (1976). On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *J. Functional Analysis*, **22**(4), 366–389.

Chang, H. W. and McKeague, I. W. (2016). Empirical likelihood based tests for stochastic ordering under right censorship. *Electron. J. Stat.*, **10**, 2511–2536.

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

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.

Csörgő, M. and Révész, P. (1978). Strong approximations of the quantile process. *Ann. Statist.*, **6**(4), 882–894.

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.

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.

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

del Barrio, E., Giné, E., and Utzet, F. (2005). Asymptotics for L_2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances. *Bernoulli*, **11**(1), 131–189.

Devroye, L. (1987). *A course in density estimation*. Progress in probability and statistics. Birkhäuser.

Dharmadhikari, S. and Joag-Dev, K. (1988). *Unimodality, Convexity, and Applications*. Probability and Mathematical Statistics. Academic Press, Inc., Boston, MA.

Dierker, E. (1991). Competition for consumers. In *Equilibrium theory and applications*, pages 393 – 402. Cambridge University Press.

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.

Doss, C. R. and Wellner, J. A. (2019). Univariate log-concave density estimation with symmetry or modal constraints. *Electron. J. Stat.*, **13**, 2391–2461.

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.

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

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

Dümbgen, L., Kolesnyk, P., and Wilke, R. (2017). Bi-log-concave distribution functions. *J. Statist. Planning and Inference*, **184**, 1–17.

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

Erdős, P. and Stone, A. H. (1970). On the sum of two Borel sets. *Proc. Amer. Math. Soc.*, **25**, 304–306.

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

Gardner, R. J. (2002). The Brunn-Minkowski inequality. *Bull. Amer. Math. Soc. (N.S.)*, **39**(3), 355–405.

Giné, E. and Nickl, R. (2008). Uniform central limit theorems for kernel density estimators. *Probability Theory and Related Fields*, **141**, 333–387.

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.

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

Gray, G. E., Andersen-Nissen, E., Grunenberg, N., Huang, Y., Roux, S., Laher, F., Innes, C., Gu, N., DiazGranados, C., Phogat, S., Lee, C., Swann, E., Kim, J., O'Connell, R., Michael, N., Flach, B., DeRosa, S., Frahm, N., Morris, L., Montefiori, D., Gilbert, P., Tomaras, G., McElrath, J., and Corey, L. (2014). HVTN 097: evaluation of the RV144 vaccine regimen in hiv uninfected south african adults. *AIDS Res. Hum. Retroviruses*, **30**, A33–A34.

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

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

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. (2012a). Immune-correlates analysis of an HIV-1 vaccine efficacy trial. *New England Journal of Medicine*, **366**, 1275–1286.

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. (2012b). Immune-correlates analysis of an HIV-1 vaccine efficacy trial. *N Engl J Med*, **366**, 1275–1286.

Hendrickx, K. and Groeneboom, P. (2017). Current status linear regression. *arXiv:1601.00202*.

Hodges, J. L. and Lehmann, E. L. (1963). Estimates of location based on rank tests. *Ann. Math. Statist.*, **34**, 598–611.

Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.*, **35**, 73–101.

Johnson, A. L., Jiang, D. R., *et al.* (2018). Shape constraints in economics and operations research. *Statist. Sci.*, **33**, 527–546.

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 *Advances in Neural Information Processing Systems*, pages 397–405.

Kaur, A., Prakasa Rao, B., and Singh, H. (1994). Testing for second-order stochastic dominance of two distributions. *Econometric Theory*, **10**, 849–866.

Kim, A. K. H. and Samworth, R. J. (2016). Global rates of convergence in log-concave density estimation. *Ann. Statist.*, **44**, 2756–2779.

Kuchibhotla, A. K., Patra, R. K., and Sen, B. (2017). Efficient estimation in convex single index models. *arXiv:1708.00145v2*.

Ledwina, T. and Wyłupek, G. (2012). Two-sample test against one-sided alternatives. *Scand. J. Stat.*, **39**, 358–381.

Ledwina, T. and Wyłupek, G. (2013). Tests for first-order stochastic dominance. *preprint*.

Lee, Y. J. and Wolfe, D. A. (1976). A distribution-free test for stochastic ordering. *J. Amer. Statist. Assoc.*, **71**, 722–727.

Lehmann, E. L. (1975). *Nonparametrics: Statistical methods based on ranks*. San Francisco : Holden-Day.

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.

Mukherjee, R. and Sen, B. (2018). Estimation of integrated functionals of a monotone density. *arXiv:1808.07915*.

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.

Pal, J. K., Woodroofe, M., and Meyer, M. (2007). Estimating a pólya frequency function$_2$. *Lecture Notes-Monograph Series*, **54**, 239–249.

Pfanzagl, J. and Wefelmeyer, W. (1985). *Asymptotic expansions for general statistical models*. Lecture notes in statistics. Springer-Verlag.

Pratt, J. W. (1960). On interchanging limits and integrals. *Ann. Math. Statist.*, **31**, 74–77.

Prekopa, A. (1973). On logarithmic concave measures and functions. *Acta Sci. Math. (Szeged)*, **34**, 335–343.

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.

Rao, B. P. (1969). Estimation of a unimodal density. *Sankhya Ser. A*, **31**, 23–36.

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**.

Rinott, Y. (1976). On convexity of measures. *Ann. Probability*, **4**(6), 1020–1026.

Robins, J., Li, L., Tchetgen, E., and van der Vaart, A. W. (2009). Quadratic semiparametric von mises calculus. *Metrika*, **69**, 227–247.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press.

Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scand. J. Stat.*, **9**, 65–78.

Sacks, J. (1975). An asymptotically efficient sequence of estimators of a location parameter. *Ann. Statist.*, **3**, 285–298.

Samworth, R. J. and Sen, B. (2018). Editorial: Special issue on "nonparametric inference under shape constraints". *Statist. Sci.*, **33**, 469–472.

Sason, I. (2018). On f-divergences: Integral representations, local behavior, and inequalities. *Entropy*, **20**, 383.

Seijo, E. and Sen, B. (2011). Nonparametric least squares estimation of a multivariate convex regression function. *Ann. Statist.*, **39**, 1633–1657.

Shorack, G. and Wellner, J. (2009a). *Empirical Processes with Applications to Statistics*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104).

Shorack, G. R. (1984). Empirical and rank processes of observations and residuals. *Canad. J. Statist.*, **12**, 319–332.

Shorack, G. R. (2000). *Probability for Statisticians*. Springer.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York.

Shorack, G. R. and Wellner, J. A. (2009b). *Empirical Processes with Applications to Statistics*, volume 59 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. Reprint of the 1986 original.

Stein, C. (1956). Efficient nonparametric testing and estimation. *Proc. Third Berkley Symp. Math. Statist. Prob*, **1**, 187–196.

Stone, C. J. (1975). Adaptive maximum likelihood estimators of a location parameter. *Ann. Statist.*, **3**, 267–284.

Takeuchi, K. (1975). A survey of robust estimation of location: models and procedures, especially in case of measurement of a physical quantity. *Bull. Inst. Internat. Statist.*, **46**, 336–348. With discussion.

Turnbull, B. C. and Ghosh, S. K. (2014). Unimodal density estimation using bernstein polynomials. *Comput. Statist. Data Anal.*, **72**, 13 – 29.

van de Geer, S. (2000). *Empirical Processes in M-Estimation*. Cambridge University Press.

Van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge University Press.

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

van der Vaart, A. W. and Wellner, J. A. W. (2007). Empirical processes indexed by estimated functions. *Asymptotics: Particles, Processes and Inverse Problems*, **55**, 234–252.

Van Eeden, C. (1970). Efficiency-robust estimation of location. *Ann. Math. Statist.*, **41**, 172–181.

Villani, C. (2003). *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI.

Villani, C. (2009). *Optimal transport; Old and New*, volume 338 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin. Old and new.

Wand, M. P. and Jones, M. C. (1994). Multivariate plug-in bandwidth selection. *Comput. Statist.*, **9**, 97–116.

Whang, Y. J. (2019). *Econometric Analysis of Stochastic Dominance: Concepts, Methods, Tools, and Applications*. Cambridge University Press.

Wolters, M. (2012). A greedy algorithm for unimodal kernel density estimation by data sharpening. *J. Stat. Softw.*, **47**, 1–26.

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.

Xu, M. and Samworth, R. J. (2017). High-dimensional nonparametric density estimation via symmetry and shape constraints. *http://www. statslab. cam. ac. uk/˜ rjs57/Research. html*. Tech. rep.

Yamanishi, K., Takeuchi, J. I., Williams, G., and Milne, P. (2004). On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms. *Data Min. Knowl. Discov.*, **8**, 275–300.