Adaptive estimation in symmetric location model under log-concavity constraint

Nilanjana Laha

Department of Biostatistics, Harvard University,
677 Huntington Ave,
Boston, MA 02115, U.S.A
e-mail: nlaha@hsph.harvard.edu

Abstract: We revisit the problem of estimating the center of symmetry θ of an unknown symmetric density f. Although Stone (1975), Van Eeden (1970), and Sacks (1975) constructed adaptive estimators of θ in this model, their estimators depend on external tuning parameters. In an effort to reduce the burden of tuning parameters, we impose an additional restriction of log-concavity on f. We construct truncated one-step estimators which are adaptive under the log-concavity assumption. Our simulations indicate that the untruncated version of the one step estimator, which is tuning parameter free, is also asymptotically efficient. We also study the maximum likelihood estimator (MLE) of θ in the shape-restricted model.

1. Introduction

In this paper, we revisit the symmetric location model with an additional shape-restriction of log-concavity. We 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 symmetric about θ. Then the symmetric location model $\mathcal{P}_s$ is given by

$$\mathcal{P}_s = \left\{ f \in \mathcal{P} \; \middle| \; f(x; \theta) = g(x - \theta), \; \theta \in \mathbb{R}, \; g \in \mathcal{S}_0, \; \mathcal{I}_f < \infty \right\}, \qquad (1)$$

where $\mathcal{I}_f$ is the Fisher information for location. It is well-established that (Huber, 1964, Theorem 3) $\mathcal{I}_f$ 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

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

Estimation of θ in $\mathcal{P}_s$ is an old semi-parametric problem, dating back to Stein (1956). From then on, the problem of estimating θ_0 in $\mathcal{P}_s$ 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 has pointed out, if f is totally unrestricted, θ is not identifiable. Second, the definition of location becomes unclear in the absence of symmetry (Takeuchi, 1975). The appeal of the above model lies in the fact that adaptive estimation of θ is possible in this model (Stone, 1975). In other words, there exist consistent estimators of θ in $\mathcal{P}_s$, whose asymptotic variance attains the parametric lower bound, which is $\mathcal{I}_f^{-1}$ in this case. See Sections 3.2, 3.3, and 6.3 of Bickel *et al.* (1998) for more discussion on adaptive estimation in $\mathcal{P}_s$.

There are general classes of nonparametric estimators, which, following some clever reconstruction, lead to adaptive estimators of θ in $\mathcal{P}_s$. Examples include the one step estimator used by Stone, and the Hodges-Lehmann rank estimator used by Van Eeden. Beran uses a linearized rank estimator introduced by Kraft and Van Eeden (1970), where Sacks uses a linear functional of order statistics. All these estimators involve various tuning parameters. The success of these type of nonparametric estimators generally depend crucially on the choice of the tuning parameters. (cf. Sacks (1975); see also Park (1990) for a thorough empirical study of similar estimators in a closely related nonparametric problem, the two-sample location problem.) However, no data-dependent method has been prescribed to choose these tuning parameters. Therefore, despite attractive theoretical properties, the implementation of the nonparametric estimators of θ_0 is not straightforward.

Although the tuning parameters stemming from different nonparametric approaches appear to be different, they generally fall in one of the following categories: (a) scaling parameter for approximating derivatives by quotient (e.g. Beran, Sacks and Van Eeden), (b) bandwidth selection parameter if kernels are used (e.g. Stone), (c) the number of basis functions (e.g. Beran), (d) parameters arising due to truncation (e.g. the estimators of Stone and Sacks) or data-partitioning (e.g., Van Eeden). We will elaborate a little bit on the first three type of tuning parameters. They arise solely because the adaptive estimators of θ_0 require estimating g, g' (e.g. Stone, Van Eeden, and Beran), and in some cases, higher derivatives (e.g. g'', Sacks). In fact, such tuning parameters are unavoidable in nonparametric estimation of the above quantities. Moreover, Hogg (1974) points out that in practice, nonparametric estimation of such functions may be too slow. This is precisely where semi-parammeric models can help because the additional structure can be exploited to construct computationally efficient estimators of g and g' without using tuning parameters.

If we impose an additional shape restriction of log-concavity on g, for instance, the task of estimating g' becomes much simple. The reason is, the class of log-concave densities is structurally rich enough to admit a maximum likelihood estimator (MLE) (Pal *et al.*, 2007; Dümbgen and Rufibach, 2009). Similar results hold for its subclasses, e.g. the class of all symmetric (about the origin) log-concave densities as well (Doss and Wellner, 2019). The log-concave MLE type

density estimators allow for computationally efficient estimation of the scores without any tuning parameters. These score estimates can readily be used to construct a one step estimator.

We show that under the log-concavity assumption, truncated versions of the above-mentioned one step estimator are adaptive provided the truncation parameter $\eta_n \to 0$ slowly enough. This truncation parameter is our only tuning parameter, which also is introduced purely due to technical reasons in the proof. Moreover, we empirically show that the efficiency of our estimators monotonously increases as $\eta_n \to 0$. In fact, the untruncated one step estimator attains the highest efficiency, and also performs reliably under varied settings. Thus, for practical implementation, the proposed estimator of this paper is the untruncated one-step estimator, which is fully tuning parameter free. We also touch upon another important tuning parameter free estimator of θ_0, namely the MLE. In particular, we establish its existence under the shape-constrained model. Our methods can be implemented using the R package `log.location` which can be accessed at https://github.com/nilanjanalaha/log.location.

The imposition of log-concavity on $\mathcal{P}_s$ may seem forced, but is not at all unnatural. The class of log-concave densities, $\mathcal{LC}$, is an important subclass of the class of unimodal densities. Many common symmetric unimodal densities, e.g. Gaussian, logistic, and Laplace, are log-concave. Unimodality is a reasonable assumption in context of location estimation of symmetric densities. As Takeuchi (1975) points out, in practice, multimodal densities generally result from unimodal mixtures. Separate procedures are available for the latter class. The difficulty with the unimodality shape restriction, however, stems from the fact that the corresponding density-class is still large, especially it is not structurally rich enough to admit an MLE (Birgé, 1997). Therefore unlike the log-concavity assumption, the unimodality assumption does not provide computational advantages. Hence, we impose the assumption of log-concavity on $\mathcal{P}_s$ instead of just unimodality.

Finally, this paper is an attempt towards bridging the gap between the symmetric location model and log-concavity. Although shape-constrained estimation has a rich history, so far there has been little to no use of shape-constraints in one-sample symmetric location problem. In fact, to the best of our knowledge, Van Eeden (1970) is the only one to incorporate shape-constraints in treating the problem considered here. Actually Van Eeden requires f to be log-concave although her paper does not mention log-concavity. She requires the function $f'(F^{-1}(u))/f(F^{-1}(u))$ to be non-increasing in $u \in (0,1)$, which is equivalent to f being log-concave (Bobkov, 1996, Proposition A.1). As made clear by our earlier discussion, Van Eeden does not use shape-restricted tools tailored for log-concave densities because they were not available at that time. We also want to mention Bhattacharyya and Bickel (2013), who consider both location and scale estimation in an elliptical symmetry model, which albeit bearing some resemblance, is different from the model considered in this paper. Also, Bhattacharyya and Bickel (2013)'s estimation procedure is completely different from ours.

1.1. Notation and terminology

For a concave function $\psi : \mathbb{R} \mapsto \mathbb{R}$, the domain $\mathrm{dom}(\psi)$ will be defined as in (Rockafellar, 1970, p. 40), that is, $\mathrm{dom}(\psi) = \{x \in \mathbb{R} : \psi(x) > -\infty\}$. For any concave function ψ, 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 $\mathcal{K}(\psi)$ 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. We denote the support of any density f by $\mathrm{supp}(f) = \{x \in \mathbb{R} : f(x) > 0\}$. We will denote by $\mathrm{int}(A)$ the interior of a set A.

For a distribution function F, we let $J(F)$ denote the set $\{x : 0 < F(x) < 1\}$. 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 $\lim_{n \to \infty} \int h dF_n = \int h dF$. For any real valued function $h : \mathbb{R} \mapsto \mathbb{R}$, we let $||h||_k$ denote its L_k norm, i.e.

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

For densities f_1 and f_2, the Hellinger distance $H(f_1, f_2)$ is defined by

$$H^2(f_1, f_2) = \frac{1}{2} \int_{-\infty}^{\infty} (f_1(x) - f_2(x))^2 dx.$$

We denote the order statistics of a random sample $(Y_1, \ldots, Y_n)$ by $(Y_{(1)}, \ldots, Y_{(n)})$.

As usual, we denote the set of natural numbers by $\mathbb{N}$. We denote by C an arbitrary constant which may vary from line to line. For two sequences x_n and y_n, we write $x_n \lesssim y_n$ if there exists a constant $C > 0$ so that $x_n \leq C y_n$.

1.2. Problem set up

To formalize the set up, first, let us define

$$\mathcal{C} := \left\{ \phi : \mathbb{R} \mapsto [-\infty, \infty) \;\middle|\; \phi \text{ is concave, closed, and proper} \right\}. \qquad (2)$$

We let $\mathcal{SC}_\theta = \mathcal{S}_\theta \cap \mathcal{C}$ denote the class of all closed and proper concave functions symmetric about $\theta \in \mathbb{R}$. Here a proper and closed concave function is as defined in Rockafellar (1970), page 24 and 50. Letting $\mathcal{LC}$ denote the class of log-concave densities

$$\mathcal{LC} := \left\{ f \in \mathcal{P} \;\middle|\; \phi = \log f \in \mathcal{C} \right\},$$

we set $\mathcal{SLC}_\theta = \mathcal{LC} \cap \mathcal{S}_\theta$. Suppose we observe n independent and identically distributed (i.i.d.) random variables $X \equiv X_1, \ldots, X_n$ with density $f_0 \equiv g_0(\cdot - \theta_0) \in \mathcal{P}_0$, where

$$\mathcal{P}_0 = \left\{ f \in \mathcal{P} \;\middle|\; f(x; \theta) = g(x - \theta), \; \theta \in \mathbb{R}, \; g \in \mathcal{SLC}_0, \; \mathcal{I}_f < \infty \right\} \qquad (3)$$

is the symmetric log-concave location model. Our aim is to estimate the location parameter θ_0.

Let us denote $\phi_0 = \log f_0$, and $\psi_0 = \log g_0$. We let F_0 and G_0 be 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.

We use the following convention throughout the paper while setting notations for the one step estimators and the MLE. We use a `hat` on the quantities related to the MLE. For example, we denote the MLE of θ_0 and g_0 by $\widehat{\theta}_n$ and $\widehat{g}_n$, respectively. The similar quantities in the one-step estimator context will use a `tilde`, e.g. $\tilde{\theta}_n$, $\tilde{g}_n$ etc. Some quantities like $\widehat{h}_n$, the MLE in $\mathcal{LC}$, or $\widehat{f}_\theta$, the MLE in $\mathcal{SLC}_\theta$, will be introduced in context of the one step estimator, but their notations use the `hat` instead of the `tilde` because they are MLEs.

The article is organized as follows. In Section 2, we introduce the one step estimator, and discuss its asymptotic properties. In Section 3, we explore the MLE of θ_0 in $\mathcal{P}_0$. We provide an empirical study in Section 4. The proofs are deferred to the appendix.

2. One step estimator

Let $\bar{\theta}_n$ be a preliminary estimator of θ_0. Had g_0 been known, a valid estimator of θ_0 would be readily given by the one step estimator (see p. 71 of Van der Vaart, 1998)

$$\bar{\theta}_n - \int_{-\infty}^{\infty} \frac{\psi_0'(x - \theta_0)}{\mathcal{I}_{f_0}} d\mathbb{F}_n(x). \tag{4}$$

In fact, the above estimator is $\sqrt{n}$ consistent with asymptotic variance $\mathcal{I}_{f_0}^{-1}$ (cf. Theorem 5.45 of Van der Vaart, 1998). Suppose $\tilde{g}_n \in \mathcal{S}_0$ is an estimator of g_0. Further suppose $\tilde{\psi}_n = \log \tilde{g}_n$ is directionally differentiable on the support of $\tilde{g}_n$. The latter always holds if $\tilde{g}_n \in \mathcal{LC}$ (Theorem 0.6.3, Hiriart-Urruty and Lemaréchal, 2004). Suppose $\tilde{\psi}_n'$ is the right derivative of $\tilde{\psi}_n$. Defining $\tilde{\psi}_n'$ to be zero outside $\operatorname{supp}(\tilde{g}_n)$, we can define an estimator of θ_0 along the lines of (4) as follows:

$$\widehat{\theta}_{n,0} = \bar{\theta}_n - \int_{-\infty}^{\infty} \frac{\tilde{\psi}_n'(x - \bar{\theta}_n)}{\widehat{\mathcal{I}}_n} d\mathbb{F}_n(x), \tag{5}$$

where

$$\widehat{\mathcal{I}}_n = \int_{-\infty}^{\infty} \tilde{\psi}_n'(x - \bar{\theta}_n)^2 d\mathbb{F}_n(x) \tag{6}$$

is an estimator of the Fisher information $\mathcal{I}_{f_0}$. We will refer to $\widehat{\theta}_{n,0}$ as the untruncated one step estimator.

The asymptotic behavior of $\tilde{\psi}_n'$ can be hard to control in the tails, which creates technical difficulties in the asymptotic analysis of $\widehat{\theta}_{n,0}$. As we already mentioned in the introduction, a common approach to tackle this problem is trimming the extreme observations, which leads to a truncated one step estimator similar to Stone.

We let η_n denote the truncation parameter, which is usually a small positive fraction. Denote by $\tilde{G}_n$ the distribution function corresponding to $\tilde{g}_n$. Letting ξ_n be the $(1-\eta_n)$-th quantile of $\tilde{G}_n$, we define the truncated one step estimator as follows:

$$\tilde{\theta}_n = \bar{\theta}_n - \int_{\bar{\theta}_n - \xi_n}^{\bar{\theta}_n + \xi_n} \frac{\tilde{\psi}'_n(x - \bar{\theta}_n)}{\widehat{\mathcal{I}}_n(\eta_n)} d\mathbb{F}_n(x). \tag{7}$$

Here $\widehat{\mathcal{I}}_n(\eta_n)$ is a truncated version of $\widehat{\mathcal{I}}_n$, given by

$$\widehat{\mathcal{I}}_n(\eta_n) = \int_{\bar{\theta}_n - \xi_n}^{\bar{\theta}_n + \xi_n} \tilde{\psi}'_n(x - \bar{\theta}_n)^2 d\mathbb{F}_n(x). \tag{8}$$

Note that the symmetry of $\tilde{g}_n$ about 0 implies that $-\xi_n = \tilde{G}_n^{-1}(\eta_n)$. Ideally, we should denote the one step estimator in (7) by $\tilde{\theta}_n(\eta_n)$ but here we suppress the dependence on η_n to avoid cumbersome notation.

$\mathcal{I}_{f_0}$ could also be estimated by a smoother version of $\widehat{\mathcal{I}}_n(\eta_n)$, namely,

$$\widehat{\mathcal{I}}_n^*(\eta_n) = \int_{\bar{\theta}_n - \xi_n}^{\bar{\theta}_n + \xi_n} \tilde{\psi}'_n(x - \bar{\theta}_n)^2 \tilde{g}_n(x - \bar{\theta}_n) dx.$$

However, our simulations indicate that the estimator $\widehat{\mathcal{I}}_n(\eta)$ yields a more efficient one-step estimator. Therefore, $\widehat{\mathcal{I}}_n(\eta)$ is our preferred estimator for the Fisher information.

2.1. Main result

The first main result of this paper states that if $\eta_n \to 0$ at a sufficiently slow rate, then the truncated one step estimator defined in (7) is adaptive for certain choices of $\tilde{g}_n$. However, we require a technical assumption on ψ_0 to prove this theorem.

Assumption A. There exists $\kappa > 0$ so that any directional derivative ψ'_0 of ψ_0 satisfies

$$|\psi'_0(x) - \psi'_0(y)| \le \kappa |x - y| \quad \text{ for all } x, y \in \text{int}(\text{dom}(\psi_0)).$$

Since ψ_0 is concave, it is directionally differentiable everywhere on $\text{int}(\text{dom}(\psi_0))$. If ψ_0 is twice differentiable on $\text{int}(\text{dom}(\psi_0))$, Assumption A interprets as $|\psi''_0| \le \kappa$. Control on the second derivative is often required in the analysis of shape-restricted models (cf. Kuchibhotla *et al.*, 2017; Mukherjee and Sen, 2019). Simple algebra shows that common symmetric log-concave densities like Gaussian, Laplace, and Logistic satisfy Assumption A. Later in Section 4, we consider an example where Assumption A is violated. Whether Assumption A is necessary is unknown to us, although Section 4 hints that the truncated one step estimators may still be adaptive even under the violation of Assumption A.

Now we state the requirements for $\tilde{g}_n$. Later in this section, we demonstrate how to build estimators which satisfy such conditions.

Condition 1. Let $y_n = o_p(1)$ be a random sequence. The density estimator $\tilde{g}_n$ satisfies the followings:

(A) $\|\tilde{g}_n - g_0\|_1 \to_p 0$ and $\sup_{x \in \mathbb{R}} |\tilde{g}_n(x + y_n) - g_0(x)| \to_p 0$.

(B) For any compact set $K \subset \text{int}(\text{dom}(\psi_0))$, $\sup_{x \in K} |\tilde{\psi}_n(x+y_n) - \psi_0(x)| \to_p 0$.

(C) Suppose $x \in \text{int}(\text{dom}(\psi_0))$ is a continuity point of ψ_0'. Then

$$\tilde{\psi}_n'(x + y_n) \to_p \psi_0'(x).$$

Condition 1 (A) implies $H(\tilde{g}_n, g_0) \to_p 0$ because $H(\tilde{g}_n, g_0) \lesssim \sqrt{\|\tilde{g}_n - g_0\|_1}$. However, we require stronger control over the rate of decay of the Hellinger error $H(\tilde{g}_n, g_0)$.

Condition 2. There exists $p \in (0, 1)$ so that $H(\tilde{g}_n, g_0) = O_p(n^{-p})$.

Now we are ready to state our main theorem. The proof of Theorem 1 can be found in Appendix B.

Theorem 1. *Suppose $f_0 \in \mathcal{P}_0$ satisfies Assumption A and $\bar{\theta}_n$ is a $\sqrt{n}$-consistent estimator of θ_0. Let $\tilde{g}_n \in \mathcal{SLC}_0$ be an estimator of g_0 which satisfies Conditions 1 and 2. Suppose $\eta_n = Cn^{-2p'/5}$, where $C > 0$ is any constant, and $p' \in (0, p]$, where p is as in Condition 2. Then the estimator $\tilde{\theta}_n$ defined in (7) satisfies*

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

A couple of remarks are in order. First, Theorem 1 requires $\tilde{g}_n \in \mathcal{SLC}_0$. This automatically rules out most nonparametric density estimators including the symmetrized kernel density estimator of Stone. Second, Theorem 1 requires $\bar{\theta}_n$ to be $\sqrt{n}$-consistent. Stone and Beran impose similar conditions on their preliminary estimators. The Z-estimator of the shift in the logistic location shift model is $\sqrt{n}$-consistent under minimal regularity conditions (cf. Example 5.40 and Theorem 5.23, Van der Vaart, 1998). When $f_0 \in \mathcal{P}_0$, the sample mean and the sample median also satisfy this requirement.

Now we give example of two $\tilde{g}_n$'s, which satisfy the conditions of Theorem 1.

Partial MLE estimator $\widehat{g}_{\bar{\theta}_n}$: For any $\theta \in \mathbb{R}$, the density class $\mathcal{SLC}_\theta$ admits an MLE (Theorem 2.1(C), Doss and Wellner, 2019). When $\theta = \bar{\theta}_n$, the MLE in the class $\mathcal{SLC}_{\bar{\theta}_n}$ is a legitimate estimator of f_0. We denote the corresponding density by $\widehat{f}_{\bar{\theta}_n}$. Then the centered density $\widehat{g}_{\bar{\theta}_n} = \widehat{f}_{\bar{\theta}_n}(\cdot + \bar{\theta}_n)$ is a potential choice for $\tilde{g}_n$ because $\widehat{g}_{\bar{\theta}_n} \in \mathcal{SLC}_0$. We call this estimator a *Partial MLE estimator* to distinguish it from the traditional MLE of g_0, which we will discuss in Section 3. From Doss and Wellner (2019) it follows that $\log \widehat{g}_{\bar{\theta}_n} = \widehat{\psi}_{\bar{\theta}_n}$ is a piecewise linear concave function with domain $[-a, a]$, where $a = |X|_{(n)}$.

Geometric mean type symmetrized estimator $\tilde{g}_n^{geo,sym}$: We denote by $\widehat{h}_n$ the MLE of f_0 among the class of all log-concave densities, which exists by Pal *et al.* (2007). The finite sample and asymptotic properties of $\widehat{h}_n$ are well-established (Dümbgen and Rufibach, 2009; Cule and Samworth, 2010). In particular, $\log \widehat{h}_n$ is piecewise linear with domain $[X_{(1)}, X_{(n)}]$. However, the estimator $\widehat{h}_n$ need not be symmetric about any $\theta \in \mathbb{R}$. A symmetrized version of

$\widehat{h}_n$ is given by

$$\tilde{g}_n^{geo,sym}(z) := C_n^{geo}\bigg(\widehat{h}_n(\bar{\theta}_n + z)\widehat{h}_n(\bar{\theta}_n - z)\bigg)^{1/2}, \quad z \in \mathbb{R} \tag{9}$$

where C_n^{geo} is a random normalizing constant. Here "geo" refers to the mode of symmetrization, which is the geometric mean in this case. Since addition preserves concavity, $\log\big(\widehat{h}_n(\bar{\theta}_n + z)\big) + \log\big(\widehat{h}_n(\bar{\theta}_n - z)\big)$ is concave, which entails that $\tilde{g}_n^{geo,sym} \in \mathcal{SLC}_0$. The support of $\tilde{g}_n^{geo,sym}$ takes the form $[-a, a]$, where

$$a = \min(X_{(n)} - \bar{\theta}_n, \bar{\theta}_n - X_{(1)}).$$

Observe that the support of $\tilde{g}_n^{geo,sym}$ is smaller than that of $\widehat{g}_{\bar{\theta}_n}$, and it may also exclude some data points. Simulations suggest that the performance of $\tilde{g}_n^{geo,sym}$ can suffer, especially in small samples, due to the exclusion of data points.

Proposition 1 states that $\widehat{g}_{\bar{\theta}_n}$ and $\tilde{g}_n^{geo,sym}$ satisfy Conditions 1 and 2, as postulated. The proof of Proposition 1 can be found in Appendix C.

Proposition 1. *Suppose $f_0 \in \mathcal{P}_0$. Then $\tilde{g}_n = \widehat{g}_{\bar{\theta}_n}$ and $\tilde{g}_n^{geo,sym}$ satisfy Condition 1 and Condition 2 with $p = 1/4$ and $2/5$, respectively.*

The $\tilde{\psi}_n$ corresponding to $\tilde{g}_n = \widehat{g}_{\bar{\theta}_n}$ and $\tilde{g}_n^{geo,sym}$ is non-smooth since $\tilde{\psi}_n$ is piecewise linear in both cases. Such an estimator may not be the best choice in small samples. Although a smoothed version of $\tilde{g}_n$ may perform better in small samples, tuning of the smoothing parameter in a data dependent way may be a non-trivial task. For the log-concave MLE $\widehat{h}_n$, however, Chen and Samworth (2013) construct a well-behaved smoothing parameter in a completely data-dependent way. This smoothing parameter is given by

$$\widehat{\lambda}_n := \sqrt{\widehat{s}_n^2 - \widehat{\sigma}_n^2}, \tag{10}$$

where $\widehat{s}_n^2$ is the sample variance and $\widehat{\sigma}_n^2$ is the variance corresponding to $\widehat{h}_n$, that is

$$\widehat{s}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \quad \text{and} \quad \widehat{\sigma}_n^2 = \int_{-\infty}^{\infty} z^2 \widehat{h}_n(z)dz - \bigg(\int_{-\infty}^{\infty} z\widehat{h}_n(z)dz\bigg)^2.$$

That the right hand side of (10) is positive follows from (2.1) of Chen and Samworth (2013). In light of the above, we construct a smooth $\tilde{g}_n$ which is symmetric about zero although it is not log-concave.

Smoothed symmetrized estimator $\tilde{g}_n^{sym,sm}$: Let us define the smoothed version of $\widehat{h}_n$ by

$$\widehat{h}_n^{sm}(z) = \frac{1}{\widehat{\lambda}_n}\int_{-\infty}^{\infty} \widehat{h}_n(z - t)\varphi(t/\widehat{\lambda}_n)dt, \quad z \in \mathbb{R}, \tag{11}$$

where φ is the standard normal density and $\widehat{\lambda}_n$ is as defined in (10). We define the smoothed symmetrized estimator by

$$\tilde{g}_n^{sym,sm}(z) = \frac{\widehat{h}_n^{sm}(\bar{\theta}_n + z) + \widehat{h}_n^{sm}(\bar{\theta}_n - z)}{2}. \tag{12}$$

It is natural to ask if similar data-dependent smoothing parameters exist for $\widehat{g}_{\bar{\theta}_n}$ and $\tilde{g}_n^{geo,sym}$ as well. Although a quantity analogous to $\widehat{\lambda}_n$ can be defined for these estimators, there is no guarantee that the former will be positive. Nevertheless, data dependent smoothing of $\widehat{g}_{\bar{\theta}_n}$ can be an interesting direction for future research.

It can be shown that $\tilde{g}_n^{sym,sm}$ satisfies Condition 1 and Condition 2 with $p = 1/5$. Moreover, although $\tilde{g}_n^{sym,sm}$ is not log-concave, it leads to an adaptive estimator of θ_0 for suitably chosen η_n.

Theorem 2. *Suppose $f_0 \in \mathcal{P}_0$ satisfies Assumption A, and $\bar{\theta}_n$ is a $\sqrt{n}$-consistent estimator of θ_0. Let $\tilde{g}_n = \tilde{g}_n^{sym,sm}$ and $\eta_n = Cn^{-2p'/5}$, where $C > 0$ and $p' \in (0, 1/5]$. Then the estimator $\tilde{\theta}_n$ defined in (7) satisfies*

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

The proof of Theorem 2 can be found in Appendix D.

Remark 1. We suspect that the rate of decay of the Hellinger error of the estimators $\widehat{g}_{\bar{\theta}_n}$ and $\tilde{g}_n^{sym,sm}$ is faster than our obtained rates, which are $O_p(-1/4)$ and $O_p(-1/5)$, respectively. Our guess is based on the fact that the geometric symmetrized estimator $\tilde{g}_n^{geo,sym}$ and the full MLE in $\mathcal{P}_0$ (see Theorem 5) are Hellinger consistent at the rate $O_p(n^{-2/5})$. The latter indicates that $H(\tilde{g}_n, g_0)$ is possibly $O_p(n^{-2/5})$ if $\tilde{g}_n$ is an equally good estimator of g_0. However, the knowledge of p does not contribute much in the tuning of η_n for practical implementation. Therefore, we do not pursue further theoretical investigation on the best possible rate of η_n in this paper.

For convenience, we list the key differences among our three main estimators of g_0 in Table 2.1.

TABLE 1

Comparison of different $\tilde{g}_n$'s: here $\widehat{h}_n$ is the log-concave MLE, $\widehat{h}_n^{sm}$ is the smoothed log-concave MLE as defined in (11), and C_n^{geo} is the normalizing constant in (9).

Estimator ($\tilde{g}_n$)	$\tilde{g}_n^{sym,sm}$	$\widehat{g}_{\bar{\theta}_n}$	$\tilde{g}_n^{geo,sym}$
Summary	Smmothed symmetrized	Partial MLE	GM type symmetrized
Formula of $\tilde{g}_n(z)$	$2^{-1}(\widehat{h}_n^{sm}(\bar{\theta}_n + z) + \widehat{h}_n^{sm}(\bar{\theta}_n - z))$	$\widehat{g}_{\bar{\theta}_n}(z)$	$C_n^{geo}(\widehat{h}_n(\bar{\theta}_n + z) \times \widehat{h}_n(\bar{\theta}_n - z))^{1/2}$
Log-concave	No	Yes	Yes
Smooth	Yes	No	No
Support	$\mathbb{R}$	$[-|X|_{(n)}, |X|_{(n)}]$	$[-a, a]$, $a = \min(X_{(n)} - \bar{\theta}_n, \bar{\theta}_n - X_{(1)})$

We close this section with a conjecture. It has previously been mentioned that the lack of control on $\tilde{\psi}_n'$ at the tails make asymptotic analysis of the untruncated estimator difficult. However, we conjecture that the untruncated estimator $\widehat{\theta}_{n,0}$ is also adaptive, i.e. $\sqrt{n}(\widehat{\theta}_{n,0} - \theta_0) \to_d N(0, \mathcal{I}_{f_0}^{-1})$. Our simulations in Section 4 does not refute this conjecture.

3. Maximum likelihood estimator (MLE)

In this section, we prove that the MLE of (θ_0, g_0) exists, and explore some of its properties. Before going into further details, we introduce some new terminologies. Recall that by our definition of $\mathcal{SC}_\theta$, the class $\mathcal{SC}_0$ consists of all proper closed concave functions symmetric about the origin. For $\psi \in \mathcal{SC}_0$ and $\theta \in \mathbb{R}$, following Dümbgen *et al.* (2011) and Xu and Samworth (2019), we define the criterion function for maximum likelihood estimation by

$$\Psi(\theta, \psi, F) = \int_{-\infty}^{\infty} \psi(x - \theta) dF(x) - \int_{-\infty}^{\infty} e^{\psi(x-\theta)} dx. \tag{13}$$

Following Silverman (1982), we included a Lagrange term to get rid of the normalizing constant involved in density estimation. This is a common device in log-concave density estimation literature (cf. Dümbgen and Rufibach, 2009; Doss and Wellner, 2019).

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,

$$\Psi_n(\theta, \psi) = \int_{-\infty}^{\infty} \psi(x - \theta) d\mathbb{F}_n(x) - \int_{-\infty}^{\infty} e^{\psi(x-\theta)} dx. \tag{14}$$

Let us denote the MLE of (θ_0, g_0) by $(\widehat{\theta}_n, \widehat{g}_n)$ when they exist. We also denote $\widehat{\psi}_n = \log \widehat{g}_n$. Observe that provided they exist, $(\widehat{\theta}_n, \widehat{\psi}_n)$ satisfies

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

For fixed $\theta \in \mathbb{R}$, denote by $\widehat{\psi}_\theta$ the maximizer of $\Psi_n(\theta, \psi)$ in $\psi \in \mathcal{SC}_0$. Theorem 2.1(C) of Doss and Wellner (2019) implies the maximizer $\widehat{\psi}_\theta$ exists, is unique, and that it satisfies

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

It is not hard to see that if the MLE $(\widehat{\theta}_n, \widehat{\psi}_n)$ exists, then

$$\widehat{\theta}_n = \underset{\theta \in \mathbb{R}}{\arg\max} \ \Psi_n(\theta, \widehat{\psi}_\theta) \quad \text{and} \quad \widehat{\psi}_n = \widehat{\psi}_{\widehat{\theta}_n}.$$

Note that $\widehat{g}_n = e^{\widehat{\psi}_n}$ is the MLE of g_0, and $\widehat{f}_n = \widehat{g}_n(\cdot - \widehat{\theta}_n)$ is the MLE of f_0. Theorem 3 implies that the the MLE $(\widehat{\theta}_n, \widehat{\psi}_n)$ exists when $\mathbb{F}_n$ is non-degenerate. The proof of Theorem 3 can be found in Appendix F.

Theorem 3. *When $\mathbb{F}_n$ is non-degenerate, the MLE $(\widehat{\theta}_n, \widehat{\psi}_n)$ of (θ_0, ψ_0) exists. If $\widehat{\theta}_n$ is unique, then $\widehat{\theta}_n \in [X_{(1)}, X_{(n)}]$. Otherwise, we can find at least one $\widehat{\theta}_n \in [X_{(1)}, X_{(n)}]$.*

Observe that Theorem 3 does not inform whether $\widehat{\theta}_n$ is unique. Since $\Psi_n(\theta, \psi)$ may not be jointly concave in θ and ψ, existence of a maximizer does not automatically imply its uniqueness. For a particular choice of $\widehat{\theta}_n$ however, the estimator $\widehat{\psi}_n = \widehat{\psi}_{\widehat{\theta}_n}$ is unique by Theorem 2.1(c) of Doss and Wellner (2019). Therefore, if (θ, ψ_1) and (θ, ψ_2) both are MLEs of (θ_0, ψ_0), we must have $\psi_1 = \psi_2$.

Although we can not theoretically prove the uniqueness of $\widehat{\theta}_n$, we are unaware of any set up which leads to non-unique MLE. Moreover, in all our simulations, $\widehat{\theta}_n$ turned out to be unique, even when the underlying density f_0 was skewed or non-log-concave. Considering this fact, in what follows, we refer to $\widehat{\theta}_n$ as "the MLE" instead of "an MLE". We must remark that even if $\widehat{\theta}_n$ is not unique, all our theorems still hold for each version of $\widehat{\theta}_n$.

When $\mathbb{F}_n$ is degenerate, Lemma 1 entails that the MLE does not exist. However, for distributions with a density, probability of $\mathbb{F}_n$ being degenerate is zero. Therefore we will not worry about this particular situation. The proof of Lemma 1 is delegated to Appendix E.

Lemma 1. *Suppose $\mathbb{F}_n$ is degenerate, i.e. $\mathbb{F}_n\{x_0\} = 1$ for some $x_0 \in \mathbb{R}$. Then the MLE of (θ_0, g_0) in $\mathcal{P}_0$ does not exist.*

The following theorem sheds some light on the structure of $\widehat{\psi}_n$. This theorem is a direct consequence of Theorem 2.1(c) of Doss and Wellner (2019), and hence we skip the proof.

Theorem 4. *Suppose $(\widehat{\theta}_n, \widehat{\psi}_n)$ is the MLE. For $\mathbb{F}_n$ non-degenerate, $\widehat{\psi}_n$ 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 $x \notin [-|X - \widehat{\theta}_n|_{(n)}, |X - \widehat{\theta}_n|_{(n)}]$, we have $\widehat{\psi}_n(x) = -\infty$. Moreover if $0 \notin \{\pm|X_1 - \widehat{\theta}_n|, \ldots, \pm|X_n - \widehat{\theta}_n|\}$, then $\widehat{\psi}'_n(0\pm) = 0$.*

The MLE can be computed using our R package `log.location`, which implements a grid search method to optimize $\Psi_n(\theta, \widehat{\psi}_\theta)$ in θ.

3.1. Asymptotic properties of the MLE

For $f_0 \in \mathcal{P}_0$, we showed that the one-step estimators are consistent. Theorem 5 (A) below shows that the MLE $\widehat{\theta}_n$ enjoys similar consistency property. In fact, $\widehat{\theta}_n$ is strongly consistent for θ_0. Part A of Theorem 5 also entails that $\widehat{g}_n$ and $\widehat{f}_n$ are strongly Hellinger consistent. Part B of Theorem 5 concerns the rate of convergences. The proof of Theorem 5 is delegated to Appendix G.

Theorem 5. *Suppose $f_0 \in \mathcal{P}_0$. Then the following assertions hold:*
(A) As $n \to \infty$, $\widehat{\theta}_n \to_{a.s.} \theta_0$, $H(\widehat{f}_n, f_0) \to_{a.s.} 0$, and $H(\widehat{g}_n, g_0) \to_{a.s.} 0$.
(B) Furthermore, $|\widehat{\theta}_n - \theta_0| = O_p(n^{-2/5})$, $H(\widehat{f}_n, f_0) = O_p(n^{-2/5})$, and $H(\widehat{g}_n, g_0) = O_p(n^{-2/5})$.

The rate of $H(\widehat{g}_n, g_0)$ as given by Theorem 5 is standard for log-concave density estimators. The MLEs in $\mathcal{SLC}_0$ and $\mathcal{LC}$ have the same rate of Hellinger

error decay (see Theorem 4.1(c) of Doss and Wellner, 2019). Moreover, this rate probably can not be improved by any other estimator of g_0. To see why, first note that Theorem 1 of Doss and Wellner (2016) proves that the minimax rate of Hellinger error decay in $\mathcal{LC}$ is $O_p(n^{-2/5})$. Remark 4.2 of Doss and Wellner (2019) conjectures that the minimax rate of estimation in the constrained class $\mathcal{SLC}_0$ stays the same. Since estimation of g_0 in $\mathcal{P}_0$ can not be easier than estimation in the smaller class $\mathcal{SLC}_0$, it is likely that the minimax rate of estimating g_0 in $\mathcal{P}_0$ is also $O_p(n^{-2/5})$.

However, the MLE $\widehat{\theta}_n$ probably convergences to θ_0 at a rate faster than $O_p(n^{-2/5})$. Our simulations suggest that $\widehat{\theta}_n$ is $\sqrt{n}$-consistent, based on which, we conjecture that $\widehat{\theta}_n$ is also an adaptive estimator of θ_0. In our model, the low dimensional parameter of interest, i.e. θ_0, is bundled with the infinite dimensional nuisance parameter. Obtaining the precise rate of convergence for the MLE in such semiparametric models is typically difficult (Murphy and Vaart, 2000). Nevertheless, since the MLE is tuning parameter free, finding its exact asymptotic distribution will be an interesting future research direction.
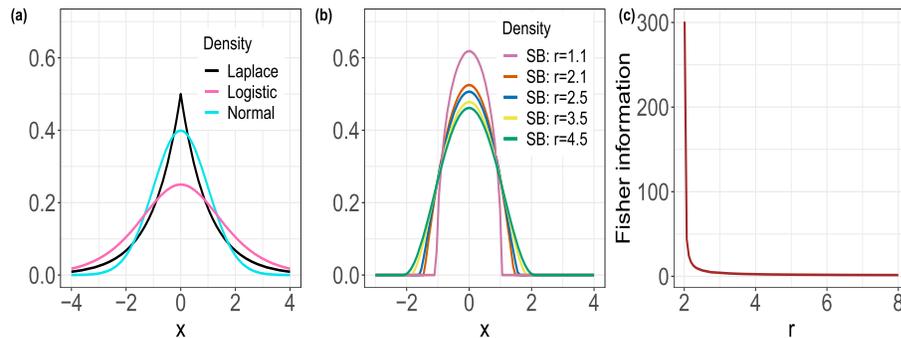
4. Simulation study

FIG 1. (a) Plot of the standard Laplace, standard normal and standard logistic densities. (b) Plot of the symmetrized beta density $f_{0,r}$, defined in (15), for different values of r. (c) Plot of Fisher information $\mathcal{I}_{f_{0,r}}$ vs r where $f_{0,r}$ is the symmetrized beta density.

This section compares the efficiency of our estimators and the coverage of the resulting confidence intervals with that of Stone and Beran. The general set-up of the simulation is as follows. We consider as g_0 the standard normal, standard logistic, and standard Laplace density. We also consider a fourth density, namely the symmetrized beta density, which is defined as follows:

$$f_0(x) \equiv f_{0,r}(x) = \frac{\Gamma\left((3+r)/2\right)}{\sqrt{\pi r}\Gamma(1+r/2)}\left(1 - \frac{x^2}{r}\right)^{r/2} 1_{[-\sqrt{r},\sqrt{r}]}(x), \quad r > 0. \quad (15)$$

Here Γ is the usual Gamma function. It is straightforward to verify that in this case

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

Some computation shows that $r \leq 2$ leads to $\mathcal{I}_{f_{0,r}} = \infty$. However for $r > 2$, $\mathcal{I}_{f_{0,r}} < \infty$, and $f_0 \in \mathcal{P}_0$. This is an example of a case where Assumption A fails to hold because ϕ_0'' is unbounded. We consider the symmetrized beta density with $r = 2.1$ and 4.5.

See Figure 1a and 1b for a pictorial representation of the above-mentioned densities. Figure 1c 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 $[-1, 1]$.

We set $\theta_0 = 0$, and generate 3000 samples of size $n = 40, 100, 200$, and 500 from each of the above-mentioned densities. We define the efficiency of an estimator θ_n by

$$\text{Efficiency}(\theta_n) = \frac{1/(n\mathcal{I}_{f_0})}{Var(\theta_n)}. \tag{16}$$

In practice, we replace $Var(\theta_n)$ by its Monte Carlo estimate.

The shape-constrained estimators:

Along with the MLE and the untruncated one step estimator defined in (5), we consider the truncated one step estimators with truncation level $\eta = 10^{-2}$, 10^{-3}, and 10^{-5}. We select the sample mean as the preliminary estimator $\bar{\theta}_n$ because it exhibited slightly better overall performance than other potential choices of $\bar{\theta}_n$, e.g. the median and the trimmed mean. We choose the partial MLE estimator and the smoothed symmetrized estimator $\tilde{g}_n^{sym,sm}$ as the estimator of g_0 because simulations suggest that they perform significantly better than $\tilde{g}_n^{geo,sym}$.

Comparators: Stone and Beran's estimators:

As mentioned earlier, Stone's estimator is a truncated one step estimator which uses symmetrized Gaussian kernels to estimate g_0. Similar to Stone, we let the corresponding truncation parameter and the kernel bandwidth parameter to be $d_n s_n$ and $t_n s_n$, respectively, where s_n is the median absolute deviation (MAD), and $d_n > 0$ and $t_n > 0$ are tuning parameters. Following Stone, we take the preliminary estimator to be the sample median.

As previously stated, Beran's estimator is a rank-based estimator which depends on the scores. Beran uses Fourier series expansion to estimate the scores, which requires choosing (a) the number of basis functions $(b_{c,n})$, and (b) a scaling parameter ρ_n, which is used to approximate a derivative term by quotients during the estimation of the Fourier coefficients. This estimator uses a preliminary estimator of θ_0, which we take to be the sample median following Beran's

suggestion. In this case, the sum of squares of the estimated Fourier coefficients is a consistent estimator of $\mathcal{I}_{f_0}$ (see (3.3) of Beran, 1974).

For sample size $n = 40$, Stone uses $d_n = 20$ and $t_n = 0.60$, but Beran does not give any demonstration on how to choose the tuning parameters. To choose some reliable values for the associated tuning parameters, we start with some pre-selected grids, and employ a grid search procedure (see Appendix I for more details). The selected tuning parameter is the maximizer of the estimated efficiency among the grid, where the efficiency is estimated using one hundred Monte Carlo replications. Of course, this procedure requires the knowledge of the unknown distribution, and hence, not implementable in practice. However, our procedure at least guarantees a reliable benchmark to compare the performance of our estimators. We refer to the resulting tuning parameters as "optimal" for the sake of simplicity. However, it should be kept in mind that these tuning parameters depend on the chosen grid, and therefore, may be different from the globally optimal tuning parameters if the grid selection is not accurate enough. This could have been overcome by an exhaustive search but that is beyond the scope of the current paper.

For each distribution and each sample size, we construct two versions of the nonparametric estimators. The first version is based on the aforementioned optimal tuning parameter, and the other version uses tuning parameters slightly away from the optimal region. For convenience, we will refer to the second set of tuning parameters as "non-optimal". See Appendix I for more details on these tuning parameters.

We should mention that neither Stone nor Beran construct confidence intervals. However, both estimators rely on consistent estimators of $\mathcal{I}_{f_0}$, namely, the estimator $\widehat{A}_n(r_n, c_n)$ of Stone (see (1.10) of Stone), and the squared L_2 norm of the estimated score in Beran. We use the above estimators of $\mathcal{I}_{f_0}$ to build the respective confidence intervals of Stone and Beran.

Results:

Figure 2 implies that Stone and Beran's estimators have high efficiency when they are equipped with the optimal tuning parameters. In fact Stone's estimator has better efficiency than all other estimators in case of logistic and normal distribution. However, even with the optimal tuning parameter, the coverage of Stone's confidence interval is quite low (see Figure 5). The coverage of Beran's confidence interval is comparatively better but still not as good as the shape-constrained estimators (see Figure 3). The poor coverage of the nonparametric confidence intervals is probably due to their smaller width, as shown by Figure 4. We suspect that for our tuning parameters, the nonparametric estimators of $\mathcal{I}_{f_0}$ overestimate $\mathcal{I}_{f_0}$, leading to narrow confidence intervals. When the tuning parameters are non-optimal, the nonparametric estimators suffer in terms of both efficiency and the coverage. This is most evident in large samples because in this case, their performance does not significantly improve with the sample size. Figure 2 and 3 entail that all estimators have markedly poor performance

in the symmetrized beta case when $r = 2.1$.

Let us turn our attention to the one step estimators now. Figure 2 underscore that the efficiency of the one step estimators monotonously decreases with the truncation, with the highest efficiency being observed at the truncation level zero. However, the difference becomes smaller as the truncation level decreases. In particular, at truncation level 10^{-5}, the difference almost vanishes. The one step estimators with lower truncation level, i.e. $\eta \leq 0.001$, exhibit satisfactory performance in terms of both efficiency and coverage (see Figure 2 and Figure 3). The estimators with higher truncation level lag in terms of efficiency as expected, although they exhibit superior coverage in some cases.

The additional gain in coverage that sometimes accompany higher levels of truncation is probably due to slightly wider confidence intervals (see Figure 4). Wider confidence intervals are expected with high levels of truncation since the length of the confidence intervals, which is a constant multiple of $\widehat{\mathcal{I}}_n(\eta)^{-1}$, increases in η. However, higher level of truncation may not always lead to a better coverage, especially since high truncation level can also result in significant loss of efficiency. See for instance the case of symmetrized beta with $r = 2.1$, where the one step estimators with truncation 0.01 lags behind the other one-step estimators in terms of both efficiency and coverage. This case clearly demonstrates that the one step estimators with higher level of truncation are not always reliable. In contrast, the one step estimators with low level of truncation, particularly the untruncated one step estimator, always exhibit satisfactory performance. In view of above, we propose the untruncated estimator for practical implementation.

Close inspection shows the smoothed symmetrized estimators have better overall performance than the partial MLE estimators with the obvious exception of Laplace distribution, which has a non-smooth density. Finally, we note that the one step estimators with lower truncation level have better efficiency than the MLE under all distributions except Laplace. However, when it comes to the coverage of the confidence intervals, the MLE can be competitive with the best one step estimators, especially in small samples.

In summary, the coverage of the nonparametric confidence intervals is not satisfactory for the tuning parameters considered here, and the efficiency of the nonparametric estimators depends crucially on the tuning parameters. For some choices of tuning parameter, these estimators may exhibit excellent efficiency but for other choices, they severely underperform. In contrast, our untruncated one step estimator and the resulting confidence interval perform reasonably well under all scenarios. The performance of the untruncated one-step estimator also speaks in favor of our conjecture that it is an adaptive estimator. Although we do not show the plots of the mean squared error (MSE) here, they depict the same patterns as the efficiency plots in Figure 2.

We close this section with a remark on the necessity of Assumption A. The symmetrized beta distributions do not satisfy Assumption A, but the one step estimators still seem to be efficient when $r = 4.5$. Although the one step estimators perform poorly in case of $r = 2.1$, they still perform better than Stone and Beran's estimators, whose asymptotically efficiency under this distribution is

theoretically validated. Thus, our simulations do not refute the possibility that Assumption A might be unnecessary.
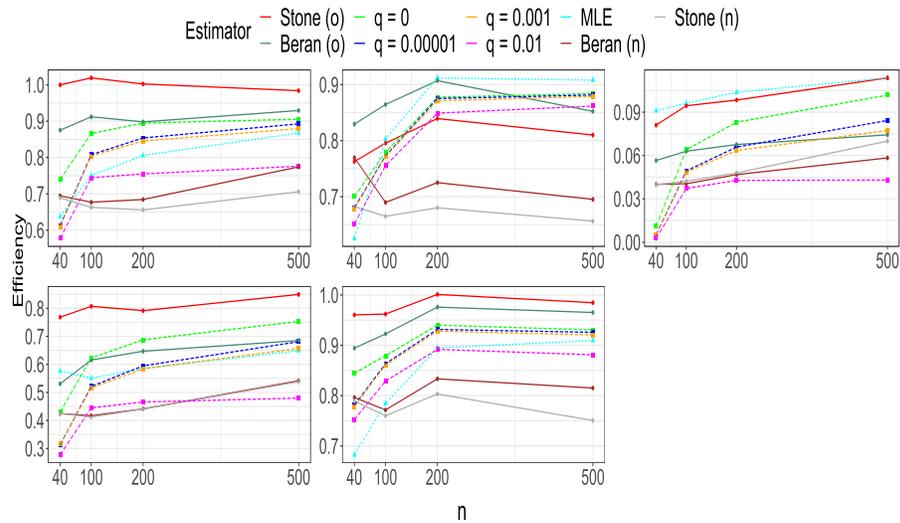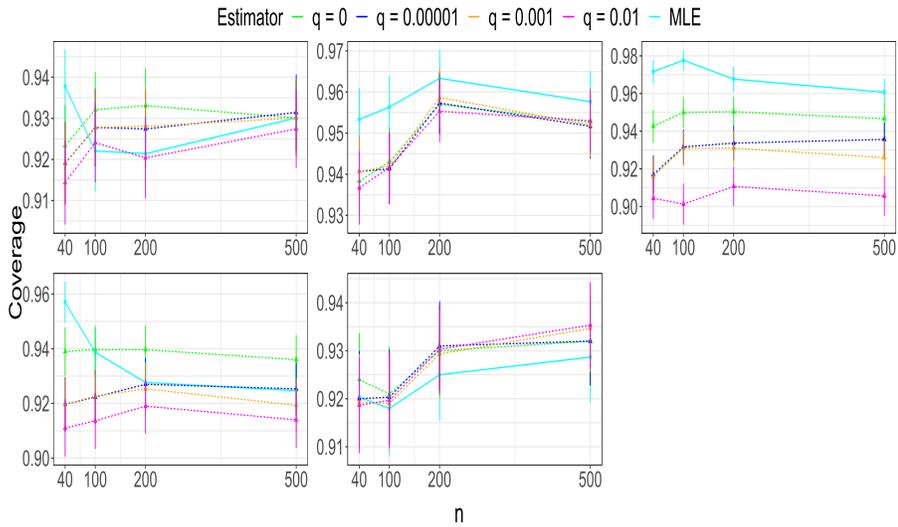

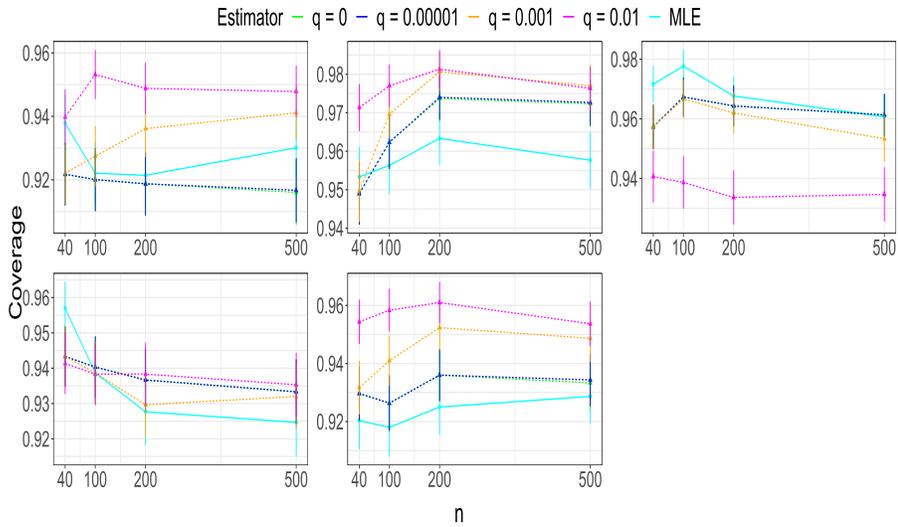(a) Comparison plot when the one step estimators are the Partial MLE estimator
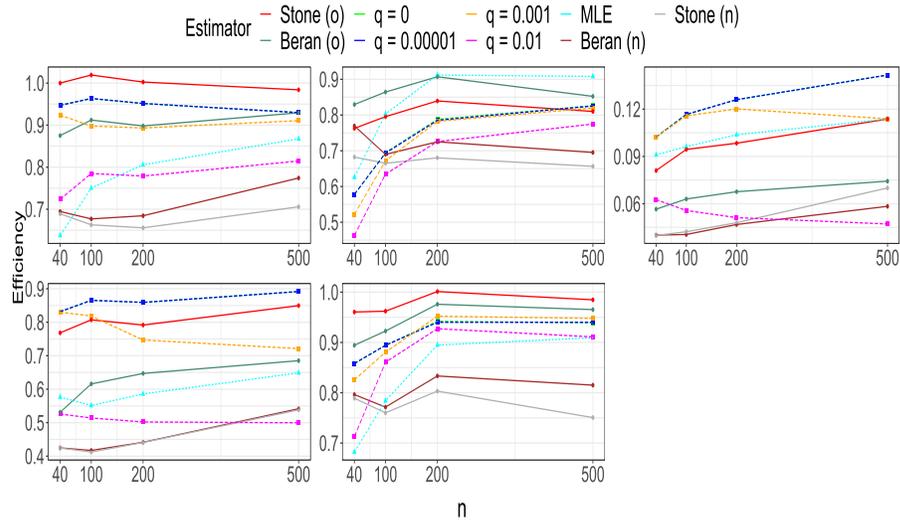
(b) Comparison plot when the one step estimators are the smoothed symmetrized estimator

FIG 2. *comparison of efficiency: the data-generating distributions are normal (topleft), Laplace (topmiddle), Symmetrized beta with r = 2.1 (topright), and r = 4.5 (bottomleft), and logistic (bottommiddle). For Stone's and Beran's estimators (in solid lines), (o) stands for the optimal tuning parameter, and (n) corresponds to the non-optimal tuning parameter. Here q stands for the truncation parameter η in our one-step estimators (in dashed lines).*
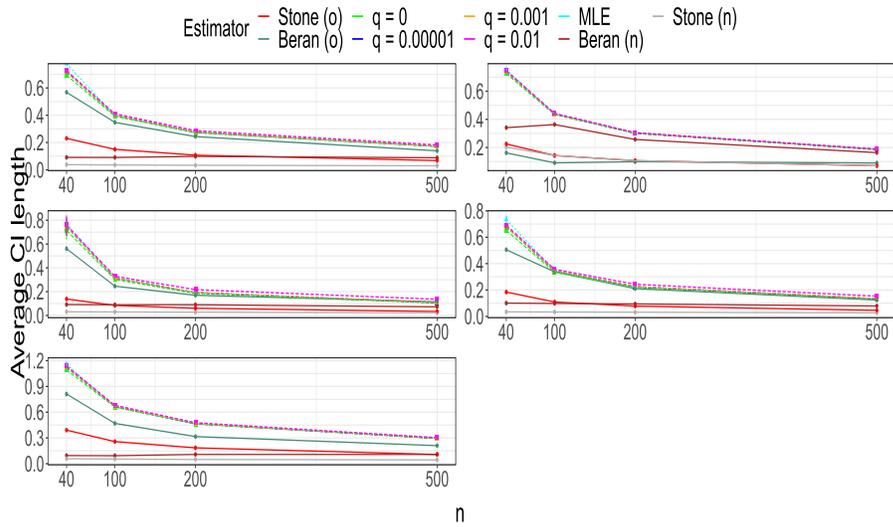
(a) Comparison plot when the one step estimators are the Partial MLE estimator
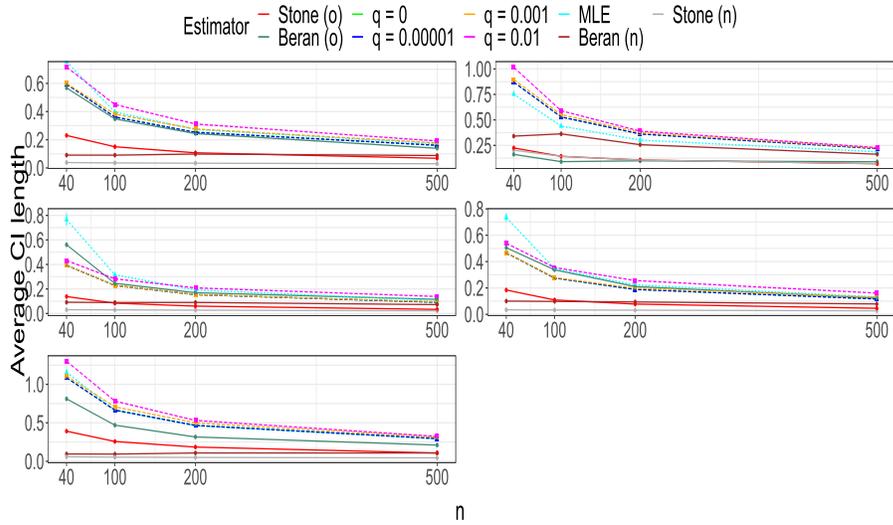

(b) Comparison plot when the one step estimators are the smoothed symmetrized estimator

FIG 3. *Comparison of the coverage of the 95% confidence intervals: the data-generating distributions are normal (topleft), Laplace (topmiddle), Symmetrized beta with $r = 2.1$ (topright), and $r = 4.5$ (bottomleft), and logistic (bottommiddle). Here q stands for the truncation parameter η in our one-step estimators. The errorbars are given by ± 2 standard deviation.*

(a) Comparison plot when the one step estimators are the Partial MLE estimator

(b) Comparison plot when the one step estimators are the smoothed symmetrized estimator

Fɪɢ 4. *Comparison of the average confidence interval length (averaged across the 3000 Monte Carlo samples): the data-generating distributions are normal (topleft), Laplace (topmiddle), Symmetrized beta with $r = 2.1$ (topright), and $r = 4.5$ (bottomleft), and logistic (bottommiddle). For Stone's and Beran's estimators (in solid lines), (o) stands for the optimal tuning parameter, and (n) corresponds to the non-optimal tuning parameter. Here q stands for the truncation parameter η in our one-step estimators (in dashed lines). The errorbars are given by ± 2 standard deviation.*
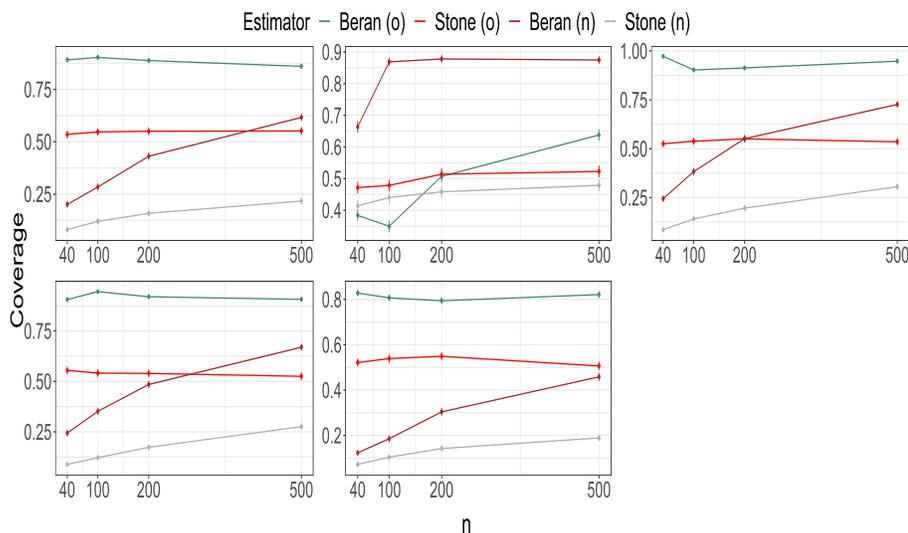
FIG 5. *Comparison of the coverage of the 95% confidence intervals for Beran's and Stone's estimators: the data-generating distributions are normal (topleft), Laplace (topmiddle), Symmetrized beta with $r = 2.1$ (topright), and $r = 4.5$ (bottomleft), and logistic (bottommiddle). Here (o) stands for the optimal tuning parameter, and (n) corresponds to the non-optimal tuning parameter. The errorbars are given by ± 2 standard deviation.*

5. Discussion

In this paper, we show that under the additional assumption of log-concavity, adaptive estimation of θ_0 is possible with only one tuning parameter. Our simulations suggest that the tuning parameter-free untruncated one step estimator may also be adaptive. This demonstrates the usefulness of log-concavity assumption in semiparametric models in facilitating a simplified estimation procedure. It is natural to ask what happens if the above shape restriction fails to hold. This question can be answered building on the log-concave projection theory developed by Dümbgen *et al.* (2011), Cule and Samworth (2010), Xu and Samworth (2019), and Barber and Samworth (2020). See Laha (2019) for discussion of the case when the log-concavity assumption is violated in our model $\mathcal{P}_0$. In particular, it can be shown that, even if $f \notin \mathcal{P}_0$, as long as f is symmetric about θ_0, the MLE and the truncated one step estimators are still consistent under mild conditions.

Acknowledgement

References

Bagnoli, M. and Bergstrom, T. (2005). Log-concave probability and its applications. *Econ. Theory*, **26**, 445–469. MR2213177

Barber, R. F. and Samworth, R. J. (2020). Local continuity of log-concave projection, with applications to estimation under model misspecification. https://arxiv.org/pdf/2002.06117.pdf.

Beran, R. (1974). Asymptotically efficient adaptive rank estimates in location models. *Ann. Statist.*, **2**, 63–74. MR0345295

Bhattacharyya, S. and Bickel, P. J. (2013). Adaptive estimation in elliptical distributions with extensions to high dimensions. http://sites.science.oregonstate.edu/~bhattash/Research_files/mixture_elliptic.pdf

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. MR1623559

Billingsley, P. (2012). *Probability and Measure*. Wiley, New York. Wiley. MR0534323

Billingsley, P. (2013). *Convergence of probability measures*. John Wiley & Sons. MR0233396

Birgé, L. (1997). Estimation of unimodal densities without smoothness assumptions. *Ann. Statist.*, **25**, 970–981. MR1447736

Bobkov, S. (1996). Extremal properties of half-spaces for log-concave distributions. *Ann. Probab.*, **24**, 35–48. MR1387625

Bobkov, S. and Ledoux, M. (2014). One-dimensional empirical measures, order statistics and Kantorovich transport distances. *preprint*. MR4028181

Chen, Y. and Samworth, R. J. (2013). Smoothed log-concave maximum likelihood estimation with applications. *Statist. Sinica*, **23**, 1373–1398. MR3114718

Cule, M. and Samworth, R. (2010). Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density. *Electron. J. Stat.*, **4**, 254–270. MR2645484

Devroye, L. (1987). *A course in density estimation*. Progress in probability and statistics. Birkhäuser. MR0891874

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. MR3485950

Doss, C. R. and Wellner, J. A. (2019). Univariate log-concave density estimation with symmetry or modal constraints. *Electron. J. Stat.*, **25**, 2391–2461. MR3983344

Dudley, R. M. (2003). *Real Analysis and Probability*. Cambridge Studies in Advanced Mathematics. Cambridge University Press. MR1932358

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. MR2546798

Dümbgen, L., Samworth, R., and Schuhmacher, D. (2011). Approximation by log-concave distributions, with applications to regression. *Ann. Statist.*, **39**, 702–730. MR2816336

Dümbgen, L., Kolesnyk, P., and Wilke, R. A. (2017). Bi-log-concave distribution

functions. *J. Statist. Plann. Inference*, **184**, 1–17. MR3600702

Hiriart-Urruty, J.-B. and Lemaréchal, C. (2004). *Fundamentals of convex analysis*. Springer Science & Business Media. MR1865628

Hogg, R. V. (1974). Adaptive robust procedures: A partial review and some suggestions for future applications and theory. *J Am Stat Assoc*, **69**(348), 909–923. MR0461779

Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.*, **35**, 73–101. MR0161415

Kraft, C. H. and Van Eeden, C. (1970). Efficient linearized estimates based on ranks. *Nonparametric techniques in Statistical Inference, (M. L. Puri, ed.)*, pages 267–273. MR0278466

Kuchibhotla, A. K., Patra, R. K., and Sen, B. (2017). Efficient estimation in convex single index models. `http://users.stat.ufl.edu/~rohitpatra/PapersandDraft/Cvx_SIMpaper.pdf`.

Laha, N. (2019). Location estimation for symmetric log-concave densities. `https://arxiv.org/pdf/1911.06225.pdf`.

Mukherjee, R. and Sen, B. (2019). On efficiency of the plug-in principle for estimating smooth integrated functionals of a nonincreasing density. *Electron. J. Stat.*, **13**, 4416–4448. MR4028511

Murphy, S. A. and Van der Vaart, A. W. (2000). On profile likelihood. *J Am Stat Assoc.*, **95**, 449–465. MR1803168

Pal, J. K., Woodroofe, M., and Meyer, M. (2007). Estimating a pólya frequency function$_2$. *Lecture Notes-Monograph Series*, **54**, 239–249. MR2459192

Park, B. U. (1990). Efficient estimation in the two-sample semiparametric location-scale models. *Probab. Theory Related Fields*, **86**, 21–39. MR1061946

Pratt, J. W. (1960). On interchanging limits and integrals. *Ann. Math. Statist.*, **31**, 74–77. MR0123673

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press. MR0274683

Sacks, J. (1975). An asymptotically efficient sequence of estimators of a location parameter. *Ann. Statist.*, **3**, 285–298. MR0359174

Shorack, G. R. (2000). *Probability for Statisticians*. Springer. MR1762415

Silverman, B. W. (1982). On the estimation of a probability density function by the maximum penalized likelihood method. *Ann. Statist.*, pages 795–810. MR0663433

Stein, C. (1956). Efficient nonparametric testing and estimation. *Proc. Third Berkley Symp. Math. Statist. Prob*, **1**, 187–196. MR0084921

Stone, C. J. (1975). Adaptive maximum likelihood estimators of a location parameter. *Ann. Statist.*, **3**, 267–284. MR0362669

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. MR0483173

Van der Vaart, A. W. (1998). *Asymptotic Statistics*. Asymptotic Statistics. Cambridge University Press. MR1652247

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York. MR1385671

Van der Vaart, A. W. and Wellner, J. A. (2007). Empirical processes indexed by estimated functions. *Asymptotics: Particles, Processes and Inverse Problems*, **55**, 234–252. MR2459942

Van Eeden, C. (1970). Efficiency-robust estimation of location. *Ann. Math. Statist.*, **41**, 172–181. MR0263194

Villani, C. (2003). *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society. MR1964483

Villani, C. (2009). *Optimal transport; Old and New*, volume 338 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin. MR2459454

Xu, M. and Samworth, R. J. (2019). High-dimensional nonparametric density estimation via symmetry and shape constraints. https://arxiv.org/abs/1903.06092v1. MR4255102

Appendix

The appendix is organized as follows. Appendices B, C, and D contain the proofs for the one step estimators, where Appendices E, F, and G contain the proofs for the MLE. The proof of the main theorem is presented first, followed by the auxiliary lemmas required for the proof. Some common technical facts, which are used repeatedly in the proofs, are listed at the end in Appendix H. Appendix I contains details on the selected tuning parameters for Stone and Beran's estimators.

Before proceeding any further, we introduce some new notations and terminologies. For $i = 1, \ldots, n$, consider the pseudo-observations $Z_i = X_i - \theta_0$. Note that, if the X_i's have density f_0, then the Z_i's have density g_0, and distribution function G_0. We will denote the log-densities corresponding to $\widehat{h}_n^{sm}$, $\tilde{g}_n^{sym,sm}$, $\widehat{g}_\theta$ and $\tilde{g}_n^{geo,sym}$ by $\tilde{\phi}_n^{sm}$, $\tilde{\psi}_n^{sym,sm}$, $\widehat{\psi}_\theta$ and $\tilde{\psi}_n^{geo,sym}$, respectively. As usual, $(\tilde{\phi}_n^{sm})'$, $(\tilde{\psi}_n^{sym,sm})'$, $\widehat{\psi}_\theta'$, and $(\tilde{\psi}_n^{geo,sym})'$ will denote the corresponding right derivatives. We remark in passing that there is nothing special about the right derivative, and any L_1 derivative would have worked. However, we fix one specific version to avoid future confusion. We denote the distribution functions of $\widehat{h}_n^{sm}$, $\tilde{g}_n^{sym,sm}$, and $\tilde{g}_n^{geo,sym}$ by $\widehat{H}_n^{sm}$, $\tilde{G}_n^{sym,sm}$, and $\tilde{G}_n^{geo,sym}$, respectively.

The empirical process of the X_i's will be denoted by $\mathbb{G}_n = \sqrt{n}(\mathbb{F}_n - F_0)$. For any function $h : \mathbb{R} \mapsto \mathbb{R}$, and a measure Q on $\mathbb{R}$, we write $Qh := \int_{\mathbb{R}} h\,dQ$ provided h is integrable with respect to Q. Suppose $\mathcal{H}$ is a class of Q-measurable functions. We denote by $\|Q\|_{\mathcal{H}}$ the supremum $\sup_{h \in \mathcal{H}} |Qh|$. For the sake of simplicity, we will denote $\tilde{\delta}_n = \theta_0 - \bar{\theta}_n$ and $\delta_n = \theta_0 - \widehat{\theta}_n$ in our proofs.

For a measure P on $\mathbb{R}$, we define the $L_{P,k}$ norm of the function h as

$$\|h\|_{P,k} = \left(\int_{-\infty}^{\infty} |h(x)|^k dP(x) \right)^{1/k}, \quad k \geq 1.$$

For any class of functions $\mathcal{H}$, we will denote

$$\|\mathcal{H}\|_{P,k} = \sup_{h \in \mathcal{H}} \|h\|_{P,k}.$$

For two distribution functions F_1 and F_2 with densities f_1 and f_2, the total variation distance between F_1 and F_2 is given by $d_{TV}(F_1, F_2) = \|f_1 - f_2\|_1/2$. We define the Wasserstein distance between two measures μ and ν on $\mathbb{R}$ by

$$d_W(\mu, \nu) = \int_{-\infty}^{\infty} |F(x) - G(x)| dx, \tag{17}$$

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), page 75. By an abuse of notation, sometime we will denote the above distance by $d_W(F, G)$ as well.

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. For any function h, $1_{[h(x) \leq C]}$ will denote the indicator function of the event $h(x) \leq C$. For any set A, we let $1_A(x)$ be the indicator function of the event $x \in A$. As usual, we denote by φ the standard Gaussian density.

In some of our proofs, we will replace $\tilde{g}_n^{sym,sm}$ by a more general mixture density which satisfies Condition 3.

Condition 3 (Condition for $\tilde{g}_n^{sym,sm}$). The density $\tilde{g}_n$ is symmetric about zero and satisfies $\tilde{g}_n(x) = (g_{1n}(x) + g_{2n}(x))/2$, where g_{1n} and g_{2n} are log-concave densities. The densities $\tilde{g}_n$, g_{1n}, and g_{2n} satisfy Conditions 1 and 2. Moreover, $\text{supp}(g_{1n}) = \text{supp}(g_{2n}) = \mathbb{R}$, and the p in Condition 2 is the same for $\tilde{g}_n$, g_{1n}, and g_{2n}.

We will later show in Lemma C.1 that $\widehat{h}_n^{sm}(\bar{\theta}_n \pm \cdot)$ and $\tilde{g}_n^{sym,sm}$ satisfy Condition 1, and in Lemma D.1, we will show that these densities satisfy Condition 2 with $p = 1/5$. Since $\widehat{h}_n^{sm}(\bar{\theta}_n \pm \cdot)$ is the convolution of two log-concave densities, it is log-concave. That $\tilde{g}_n^{sym,sm}$ satisfies Condition 3 follows immediately from the above results.

We will frequently use the fact that if f is a log-concave density, then $f > 0$ on $J(F) = \{0 < F < 1\}$ (cf. Theorem 1(iv) of Dümbgen *et al.*, 2017). Therefore, $\text{int}(\text{dom}(\log f)) = \text{int}(\text{supp}(f)) = \text{int}(J(F))$. As a consequence, F^{-1} is strictly increasing, and differentiable with derivative $1/f(F^{-1}(t))$ on $(0, 1)$ (see Fact 8). Also, a log-concave f is continuous on $\text{int}(\text{dom}(\log f))$, and thus $\text{int}(J(F))$, by Theorem 10.1 of Rockafellar (1970). When $f_0 \in \mathcal{P}_0$, furthermore, f_0 and g_0 are absolutely continuous on $\mathbb{R}$ by Theorem 3 of Huber (1964). We state some more useful facts about log-concave densities below.

Fact 1 (Lemma 1 of Cule and Samworth (2010)). *If f is a univariate log-concave density, then there exists $\alpha > 0$ and $\beta \in \mathbb{R}$ so that $f(x) \leq e^{-\alpha|x|+\beta}$.*

Fact 2. *If f is log-concave, then f/F is non-increasing on $J(F)$ and $f/(1-F)$ is non-decreasing on $J(F)$.*

Proof. This is a well known fact about log-concave densities. See for reference Theorem 1 and Corollary 2 of Bagnoli and Bergstrom (2005). □

The following two facts will be very useful to lower bound $\tilde{g}_n$ on $[-\xi_n, \xi_n]$.

Fact 3. *Suppose $\tilde{g}_n$ is a log-concave density satisfying Condition 1. Then for any $x \in \mathbb{R}$,*

$$\tilde{g}_n(x) \geq \omega_n \min(\tilde{G}_n(x), 1 - \tilde{G}_n(x)),$$

where $\omega_n \geq 0$ satisfies $\omega_n \to_p \omega_0 > 0$. Here $\omega_0 > 0$ is a constant depending only on g_0.

Proof. If $\tilde{G}_n(x)$ is zero or one, then the statement trivially holds. Therefore, we assume $x \in J(\tilde{G}_n)$, i.e. $0 < \tilde{G}_n(x) < 1$. For $q \in (0, 1/2)$, Fact 2 implies

$$\tilde{g}_n(\tilde{G}_n^{-1}(q)) \geq 2q\Big(\tilde{g}_n(\tilde{G}_n^{-1}(1/2))\Big) \geq 2\min(q, 1 - q)\tilde{g}_n(\tilde{G}_n^{-1}(1/2)).$$

For $q \in (1/2, 1)$, on the other hand, Fact 2 implies

$$\tilde{g}_n(\tilde{G}_n^{-1}(q)) \geq 2(1 - q)\tilde{g}_n(\tilde{G}_n^{-1}(1/2)) \geq 2\min(q, 1 - q)\tilde{g}_n(\tilde{G}_n^{-1}(1/2)).$$

Because $\tilde{G}_n(x) \in (0, 1)$, replacing q by $\tilde{G}_n(x)$ we obtain that

$$\tilde{g}_n(x) \overset{(a)}{=} \tilde{g}_n(\tilde{G}_n^{-1}(\tilde{G}_n(x))) \geq 2\tilde{g}_n(\tilde{G}_n^{-1}(1/2)) \min\Big(\tilde{G}_n(x), 1 - \tilde{G}_n(x)\Big). \quad (18)$$

Here (a) uses the fact that $\tilde{g}_n(x) > 0$ which follows since $x \in \text{int}(J(\tilde{G}_n)) = \text{int}(\text{dom}(\tilde{\psi}_n))$. The rest of the proof follows setting $\omega_n = 2\tilde{g}_n(\tilde{G}_n^{-1}(1/2))$, which converges in probability to $\omega_0 = 2g_0(0)$ by Condition 1 and Fact 11. □

Fact 4. *Suppose either $\tilde{g}_n$ satisfies Condition 3, or $\tilde{g}_n$ is a log-concave density satisfying Condition 1. Then the assertions of Fact 3 hold.*

Proof. If $\tilde{g}_n \in \mathcal{LC}$, the proof follows from Fact 3. Therefore we consider the case when $\tilde{g}_n$ satisfies Condition 3. Since the component densities g_{1n} and g_{2n} in Condition 3 are log-concave, Fact 3 applies to them. Denote by G_{1n} and G_{2n} the corresponding distribution functions. Equation 18 in the proof of Fact 3 implies

$$g_{1n}(x) + g_{2n}(x) \geq b_n\Big\{G_{1n}(x)\Big(1 - G_{1n}(x)\Big) + G_{2n}(x)\Big(1 - G_{2n}(x)\Big)\Big\},$$

where

$$b_n = 2\min\Big(g_{1n}(G_{1n}^{-1}(1/2)), g_{2n}(G_{2n}^{-1}(1/2))\Big) \to_p 2g_0(0)$$

by the fact that g_{1n} and g_{2n} satisfy Condition 1 and Fact 11. Since the function $(x, y) \mapsto xy$ is convex on $\{(x, y) : x, y \geq 0\}$, it follows that

$$\tilde{g}_n(x) \geq b_n\frac{G_{1n}(x) + G_{2n}(x)}{2}\left(1 - \frac{G_{1n}(x) + G_{2n}(x)}{2}\right) = b_n\tilde{G}_n(x)(1 - \tilde{G}_n(x)).$$

The proof follows noting $x(1 - x) \geq \min(x, 1 - x)/2$ for $x \in (0, 1)$. □

Appendix B: Proof of Theorem 1

We first argue that it suffices to prove the theorem only for the case when η_n equals $Cn^{-2p/5}$. In the latter case, we would show $\sqrt{n}(\tilde{\theta}_n - \theta_0) \to_d N(0, \mathcal{I}_{f_0}^{-1})$ for $\eta_n = Cn^{-2p/5}$ when $H(\tilde{g}_n, g_0) = O_p(n^{-p})$. Note that for any $p' \in (0, p]$,

$H(\tilde{g}_n, g_0) = O_p(n^{-p'})$ trivially holds since $H(\tilde{g}_n, g_0) = O_p(n^{-p})$. Therefore, replacing p by p' in what we just proved, $\sqrt{n}(\tilde{\theta}_n - \theta_0) \to_d N(0, \mathcal{I}_{f_0}^{-1})$ would follow identically for $\eta_n = Cn^{-2p'/5}$. Thus, it is enough to consider the case when $\eta_n = Cn^{-2p/5}$.

From (7) we obtain that

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

Denoting $\tilde{\delta}_n = \theta_0 - \bar{\theta}_n$, we observe that the above expression writes as

$$\underbrace{\int_{-\xi_n}^{\xi_n} \frac{\tilde{\psi}_n'(z) - \psi_0'(z - \tilde{\delta}_n)}{\widehat{\mathcal{I}}_n(\eta_n)} d(\mathbb{F}_n(z + \bar{\theta}_n) - F_0(z + \bar{\theta}_n))}_{T_{1n}}$$

$$+ \underbrace{\int_{-\xi_n}^{\xi_n} \frac{\tilde{\psi}_n'(z)}{\widehat{\mathcal{I}}_n(\eta_n)} \left(f_0(z + \bar{\theta}_n) - g_0(z) \right) dz}_{T_{2n}} + \underbrace{\int_{-\xi_n}^{\xi_n} \frac{\tilde{\psi}_n'(z) - \psi_0'(z)}{\widehat{\mathcal{I}}_n(\eta_n)} g_0(z) dz}_{T_{3n}}$$

$$+ \underbrace{\int_{-\xi_n}^{\xi_n} \frac{\psi_0'(z)}{\widehat{\mathcal{I}}_n(\eta_n)} g_0(z) dz}_{T_{4n}} + \underbrace{\int_{-\xi_n}^{\xi_n} \frac{\psi_0'(z - \tilde{\delta}_n)}{\widehat{\mathcal{I}}_n(\eta_n)} d(\mathbb{F}_n(z + \bar{\theta}_n) - F_0(z + \bar{\theta}_n))}_{T_{5n}}$$

$$(19)$$

Observe that T_{3n} and T_{4n} vanish since $\tilde{\psi}_n'$ and ψ_0' are odd functions while g_0 is an even function.

The proof of Theorem 1 has three main steps. The first step uses Donsker Theorem to show that the empirical process term T_{1n} is $o_p(n^{-1/2})$. The term T_{2n} accounts for the bias due to the use of $\bar{\theta}_n$ instead of the true center θ_0 in the construction of the scores. The second step of the proof shows that the order of T_{2n} is same as $\tilde{\delta}_n = \theta_0 - \bar{\theta}_n$. In particular, we will show that $T_{2n} = -\tilde{\delta}_n(1 + o_p(1))$. Since $\tilde{\delta}_n = O_p(n^{-1/2})$, the above two steps lead to

$$\sqrt{n}(\bar{\theta}_n - \tilde{\theta}_n) = o_p(1) + \sqrt{n}(\bar{\theta}_n - \theta_0) + \sqrt{n}T_{5n}.$$

The third step of the proof shows that the term $\sqrt{n}T_{5n}$ is asymptotically normal with variance $\mathcal{I}_{f_0}^{-1}$. A rearranging of the terms in the above display then establishes the desired asymptotic convergence of $\sqrt{n}(\tilde{\theta}_n - \theta_0)$. The rest of the proof is devoted towards proving the above-mentioned three steps.

First step: asymptotic negligibility of $\sqrt{n}T_{1n}$:

First, let us denote $\mathcal{T}_n = [\bar{\theta}_n - \xi_n, \bar{\theta}_n + \xi_n]$. Recall that in Section 1.1 we denoted the empirical process $\sqrt{n}(\mathbb{F}_n - F_0)$ by $\mathbb{G}_n$. Note that $\sqrt{n}T_{1n}$ also writes as

$$\sqrt{n}T_{1n} = \sqrt{n}\int_{\bar{\theta}_n - \xi_n}^{\bar{\theta}_n + \xi_n} \frac{\tilde{\psi}_n'(x - \bar{\theta}_n) - \phi_0'(x)}{\widehat{\mathcal{I}}_n(\eta_n)} d(\mathbb{F}_n - F_0)(x) = \int_{-\infty}^{\infty} \frac{h_n(x)}{\widehat{\mathcal{I}}_n(\eta_n)} d\mathbb{G}_n(x),$$

$$(20)$$

where by h_n we denote the function

$$h_n(x) = (\tilde{\psi}'_n(x - \bar{\theta}_n) - \phi'_0(x))1_{\mathcal{T}_n}(x), \quad x \in \mathbb{R}. \tag{21}$$

Because $\eta_n = O(n^{-2p/5})$, Lemma B.15 implies

$$\sup_{x \in \mathcal{T}_n} |\tilde{\psi}'_n(x - \bar{\theta}_n)| = O_p(n^{p/5}). \tag{22}$$

Thus $\tilde{\psi}'_n$ restricted to the compact set $\mathcal{T}_n$ is bounded. We can extend the function $x \mapsto \tilde{\psi}'_n(x)1_{\mathcal{T}_n}(x)$ to $\mathbb{R}$ in a way such that the resulting function $\hat{u}_n$ is still monotone and has the same bound. This can be done by setting $\hat{u}_n$ to be $\tilde{\psi}'_n(\bar{\theta}_n - \xi_n)$ and $\tilde{\psi}'_n(\bar{\theta}_n + \xi_n)$ on the intervals $(-\infty, \bar{\theta}_n - \xi_n]$ and $[\bar{\theta}_n + \xi_n, \infty)$, respectively. Note also that we can replace $\tilde{\psi}'_n$ by $\hat{u}_n$ in the definition of h_n, i.e.

$$h_n(x) = (\hat{u}_n(x - \bar{\theta}_n) - \phi'_0(x))1_{\mathcal{T}_n}(x).$$

Let us denote $M_n = Cn^{p/5}$ for some $C > 0$ and define

$$\mathcal{U}_n(M_n) = \left\{ u : \mathbb{R} \mapsto [-M_n, M_n] \ \middle| \ u \text{ is non-increasing} \right\}. \tag{23}$$

Since $\|\hat{u}_n\|_\infty = O_p(n^{p/5})$, for sufficiently large C, $\hat{u}_n(\cdot - \bar{\theta}_n) \in \mathcal{U}_n(M_n)$ with high probability. Now define the class $\mathcal{H}_n(C)$ by

$$\mathcal{H}_n(C) = \left\{ h : \mathbb{R} \mapsto \mathbb{R} \ \middle| \ h(x) = (u(x) - \phi'_0(x))1_{[r_1, r_2]}(x), \ u \in \mathcal{U}_n(M_n), \right.$$
$$\|h\|_{P_0,2} \le Cn^{-2p/5}(\log n)^{3/2}, \quad \|h\|_\infty \le M_n,$$
$$\left. [r_1, r_2] \subset [\theta_0 - C \log n, \theta_0 + C \log n] \cap \text{int}(\text{dom}(\phi_0)) \right\}.$$

The notation $\mathcal{H}_n(C)$ does not depend on M_n because $M_n = Cn^{p/5}$ is also a function of C.

We want to show that $h_n \in \mathcal{H}_n(C)$ with high probability for large n. Note that

$$\sup_{x \in \mathcal{T}_n} |\phi'_0(x)| = \sup_{x \in [-\delta_n - \xi_n, -\delta_n + \xi_n]} |\psi'_0(x)|.$$

Lemma B.10 in conjunction with the fact that $\eta_n = O(n^{-2p/5})$ implies

$$\sup_{x \in \mathcal{T}_n} |\phi'_0(x)| = O_p(\log n). \tag{24}$$

Thus (22) and (24) imply $\|h_n\|_\infty = O_p(n^{p/5})$. Lemma B.16 bounds the $L_{P_0,2}$ norm of h_n entailing $\|h_n\|_{P_0,2} = O_p(n^{-2p/5}(\log n)^{3/2})$. Lemma B.6 implies, on the other hand,

$$\lim_{n \to \infty} P\Big(\mathcal{T}_n \subset [\theta_0 - C \log n, \theta_0 + C \log n] \cap \text{int}(\text{dom}(\phi_0))\Big) = 1.$$

Therefore, we conclude that given $t > 0$, we can choose $C > 0$ so large such that $P(h_n \in \mathcal{H}_n(C)) > 1 - t$.

Theorem 2.7.5 of Van der Vaart and Wellner (1996) (pp. 159) states that there exists an absolute constant $C' > 0$ so that for any $\epsilon > 0$ and any probability measure $\mathbb{R}$ on the real line,

$$\log N_{[\,]}(\epsilon, \mathcal{U}_n(M_n), L_2(Q)) \le C' M_n \epsilon^{-1}. \tag{25}$$

On the other hand, using Theorem 2.7.5 of Van der Vaart and Wellner (1996), it can also be shown that the class $\mathcal{F}_I$ of all indicator functions of the form $1_{[z_1,z_2]}$, where $z_1 \le z_2$ with $z_1, z_2 \in \mathbb{R}$, satisfies

$$\log N_{[\,]}(\epsilon, \mathcal{F}_I, L_2(Q)) \le C'2\epsilon^{-1}. \tag{26}$$

Using (25) and (26) we derive that

$$\log N_{[\,]}(\epsilon, \mathcal{H}_n(C), L_2(P_0)) \lesssim M_n \epsilon^{-1}.$$

For $x < 1$, the bracketing integral

$$\mathcal{J}_{[\,]}(x, \mathcal{H}_n(C), L_2(P_0)) = \int_0^x \sqrt{1 + \log N_{[\,]}(\epsilon, \mathcal{H}_n(C), L_2(P_0))} d\epsilon$$

$$\lesssim 2M_n \int_0^{x/M_n} \epsilon^{-1/2} d\epsilon,$$

which equals $\sqrt{xM_n}$. Let us also denote $K_n = Cn^{-2p/5}(\log n)^{3/2}$. Note that

$$\|\mathcal{H}_n(C)\|_{P_0,2} = \sup_{h \in \mathcal{H}_n(C)} \|h\|_{P_0,2} = K_n.$$

Then from Fact 9 it follows that

$$E\|\mathbb{G}_n\|_{\mathcal{H}_n(C)} \lesssim \mathcal{J}_{[\,]}(K_n, \mathcal{H}_n(C), L_2(P_0))\Big(1 + \frac{\mathcal{J}_{[\,]}(K_n, \mathcal{H}_n(C), L_2(P_0))}{K_n^2\sqrt{n}} M_n\Big)$$

which is bounded by a constant multiple of

$$\sqrt{K_n M_n} + K_n^{-1} M_n^2 n^{-1/2} \lesssim (\log n)^{3/4} n^{-p/10} + (\log n)^{-3/2} n^{-1/2}.$$

Now fix $t' > 0$ and $\xi > 0$. We can choose C so large such that $P(h_n \notin \mathcal{H}_n(C)) < \xi/2$. Therefore

$$P\Big(\int_{-\infty}^{\infty} h_n(x) d\mathbb{G}_n(x) > t'\Big)$$

$$\le P\Big(\int_{-\infty}^{\infty} h_n(x) d\mathbb{G}_n(x) > t', h_n \in \mathcal{H}_n(C)\Big) + P\Big(h_n \notin \mathcal{H}_n(C)\Big)$$

$$\overset{(a)}{\le} E\Big[\sup_{h \in \mathcal{H}_n(C)} \Big|\int_{-\infty}^{\infty} h(x) d\mathbb{G}_n(x)\Big|\Big]/t' + \xi/2$$

$$= O_p((\log n)^{3/4} n^{-p/10})/t' + \xi/2,$$

which is less than ξ for sufficiently large n. Here (a) follows from Markov inequality. Since t' and ξ are arbitrary, we conclude that $\int h_n d\mathbb{G}_n$ is $o_p(1)$. Finally an application of Lemma B.17 leads to $\widehat{\mathcal{I}}_n(\eta_n) \to_p \mathcal{I}_{f_0}$, and thus from (20), $\sqrt{n}T_{1n} = o_p(1)$ follows.

Second step: asymptotic limit of $T_{2n}/\tilde{\delta}_n$:

Let us define $\mathcal{A}_n = [-\xi_n, \xi_n - \tilde{\delta}_n]$, Observe that $T_{2n}/\tilde{\delta}_n$ can be written as

$$\int_{-\xi_n}^{\xi_n} \frac{\tilde{\psi}_n'(z)}{\widehat{\mathcal{I}}_n(\eta_n)} \frac{\Big(g_0(z - \tilde{\delta}_n) - g_0(z)\Big)}{\tilde{\delta}_n} dz = \int_{-\xi_n}^{\xi_n} \frac{\tilde{\psi}_n'(z)}{\widehat{\mathcal{I}}_n(\eta_n)} \frac{\int_z^{z - \tilde{\delta}_n} g_0'(t) dt}{\tilde{\delta}_n} dz \tag{27}$$

$$= -\int_{\mathbb{R}} \underbrace{1_{\mathcal{A}_n}(t)g_0'(t)\frac{\int_t^{t+\tilde{\delta}_n}\tilde{\psi}_n'(z)dz}{\tilde{\delta}_n\widehat{\mathcal{I}}_n(\eta_n)}}_{b_n(t)} \, dt,$$

where the last equality follows by Fubini's Theorem since g_0 is absolutely continuous.

Note that (27) implies $T_{2n} = -\int_{\mathbb{R}} b_n(t)dt$. The following lemma, which is proved in Appendix B.1, establishes $T_{2n} \to_p -1$, thus completing the proof of the second step.

Lemma B.1. *Under the set up of Theorem 1,* $\mathbb{Y}_n \equiv \int_{-\infty}^{\infty} b_n(t)dt \to_p 1$ *where* $b_n(t)$ *is as defined in* (27).

Third step: showing the asymptotic normality of T_{5n}***:***

A change of variable leads to

$$\sqrt{n}T_{5n} = \sqrt{n}\int_{\bar{\theta}_n-\xi_n}^{\bar{\theta}_n+\xi_n} \frac{\psi_0'(x-\theta_0)}{\widehat{\mathcal{I}}_n(\eta_n)}d(\mathbb{F}_n - F_0)(x)$$

$$= \int_{-\infty}^{\infty} \frac{\phi_0'(x)}{\widehat{\mathcal{I}}_n(\eta_n)}d\mathbb{G}_n(x) - \int_{C_n} \frac{\phi_0'(x)}{\widehat{\mathcal{I}}_n(\eta_n)}d\mathbb{G}_n(x), \qquad (28)$$

where $C_n = (-\infty, \bar{\theta}_n - \xi_n] \cup [\bar{\theta}_n + \xi_n, \infty]$. The central limit theorem yields

$$\int_{-\infty}^{\infty} \phi_0'(x)d\mathbb{G}_n(x) = \sum_{i=1}^{n} \frac{\phi_0'(X_i) - E[\phi_0'(X_i)]}{\sqrt{n}} \to_d N(0, \mathcal{I}_{f_0}).$$

Then from Lemma B.17 and Slutsky's theorem it follows that

$$\int_{-\infty}^{\infty} \frac{\phi_0'(x)}{\widehat{\mathcal{I}}_n(\eta_n)}d\mathbb{G}_n(x) \to_d N(0, \mathcal{I}_{f_0}^{-1}).$$

Thus it suffices to show that the second term on the right hand side of (28) is $o_p(1)$. To that end, observe that $1 - 1_{C_n} = 1_{C_n^c}$ belongs to the class of all indicator functions of the form $1_{[z_1,z_2]}$, where $z_1 \leq z_2$ with $z_1, z_2 \in \mathbb{R}$. Since the latter class is Donsker by (26), Theorem 2.1 of Van der Vaart and Wellner (2007) entails that the second term on the right hand side of (28) is of order $o_p(1)$ provided

$$\widehat{\mathcal{I}}_n(\eta_n)^{-2}\int_{-\infty}^{\infty} 1_{C_n}(x)\phi_0'(x)^2 f_0(x)dx \to_p 0.$$

Since $\widehat{\mathcal{I}}_n(\eta_n) \to_p \mathcal{I}_{f_0} > 0$ by Lemma B.17, we only need to show that the integral in the last display is $o_p(1)$. Because $\mathcal{I}_{f_0} < \infty$, Fact 12 implies that given any $\epsilon > 0$, there exists $\sigma > 0$ so that $P_0(\mathcal{B}) < \sigma$ implies $\int_{\mathcal{B}} \phi_0'^2(x)f_0(x)dx < \epsilon$ for any P_0-measurable set $\mathcal{B} \subset \mathbb{R}$. Thus the proof follows if we can show that

$\int_{\mathcal{C}_n} f_0(x)dx = o_p(1)$. To that end, observe that

$$\int_{C_n} f_0(x)dx = 1 - F_0(\bar{\theta}_n + \xi_n) + F_0(\bar{\theta}_n - \xi_n)$$

$$\to_p 1 - F_0(\theta_0 + G_0^{-1}(1)) + F_0(\theta_0 + G_0^{-1}(0))$$

by continuous mapping theorem because (a) $\bar{\theta}_n \to_p \theta_0$, (b) $\xi_n \to_p G_0^{-1}(1)$ by Lemma B.3, and (c) F_0 is continuous. Since $\theta_0 + G_0^{-1}(1) = F_0^{-1}(1)$ and $\theta_0 + G_0^{-1}(0) = F_0^{-1}(0)$, the proof follows. $\qquad\square$

B.1. Proof of key lemmas for Theorem 1

Proof of Lemma B.1. Recall that we defined $\mathcal{A}_n = [-\xi_n, \xi_n - \tilde{\delta}_n]$ in the proof of Theorem 1. Let us define $\mathcal{A}'_n = [-\xi_n - 2|\tilde{\delta}_n|, \xi_n + 2|\tilde{\delta}_n|]$. We also denote

$$\mathcal{I}_{1n} = \int_{\mathcal{A}_n} \tilde{\psi}'_{n_k}(t)^2 \tilde{g}_{n_k}(t)dt \quad \text{and} \quad \mathcal{I}_{2n} = \int_{\mathcal{A}_n + \tilde{\delta}_n} \tilde{\psi}'_{n_k}(t)^2 \tilde{g}_{n_k}(t)dt.$$

First we will show that it suffices to consider almost sure convergence of Y_n along some suitably chosen subsequence. We claim that given any subsequence of $\{n\}$, we can always obtain a further subsequence $\{n_k\}_{k \geq 1}$ so that the set

$$\mathcal{M} = \Bigg\{ \bar{\theta}_{n_k} \to_k \theta_0, \ \widehat{I}_{n_k}(\eta_{n_k}) \to_k \mathcal{I}_{f_0}, \ \xi_{n_k} \to_k G_0^{-1}(1), \ \omega_{n_k} \to_k \omega_0,$$

$$\frac{(\log n_k)^2 H(\tilde{g}_{n_k}, g_0)^2}{\inf_{x \in \mathcal{A}'_n} \tilde{g}_{n_k}(x)} \to_k 0, \ \lim_{k \to \infty} \mathcal{I}_{in_k} = \mathcal{I}_{f_0} \text{ for } i = 1, 2, \tag{29}$$

$$\|\tilde{g}_{n_k} - g_0\|_\infty \to_k 0, \ \mathcal{A}'_{n_k} \subset \text{int}(\text{dom}(\psi_0)) \text{ for all sufficiently large } n_k \Bigg\}$$

has probability one, where ω_n and ω_0 are as in Fact 4. The claim follows directly by Fact 6 noting

(a) $\bar{\theta}_n \to_p \theta_0$.
(b) $\widehat{\mathcal{I}}_n(\eta_n) \to_p \mathcal{I}_{f_0}$ by Lemma B.17.
(c) $\xi_n \to_p G_0^{-1}(1)$ by Lemma B.3.
(d) $\omega_n \to_p \omega_0$ by Fact 4.
(e) Suppose $\tilde{\xi}_n = \tilde{G}_n^{-1}(1 - \eta_n/2)$. Then Lemma B.7 implies $\sup_{x \in [-\tilde{\xi}_n, \tilde{\xi}_n]}(\tilde{g}_{n_k}(x))^{-1}$ is $O_p(n^{2p/5})$. However, $\mathcal{A}'_n \subset [-\tilde{\xi}_n, \tilde{\xi}_n]$ by Lemma B.5 with probability tending to one since $\tilde{\delta}_n = O_p(n^{-1/2})$. Since $H(\tilde{g}_n, g_0) = O_p(n^{-p})$ and $p \in (0, 1)$, it follows that

$$\frac{(\log n)^2 H(\tilde{g}_n, g_0)^2}{\inf_{x \in \mathcal{A}'_n} \tilde{g}_n(x)} \to_p 0.$$

(f) Suppose $\tilde{\xi}_n = (\tilde{G}_n)^{-1}(1 - \eta_n/2)$. Lemma B.5 implies that with probability tending to one, $\xi_n + 2|\tilde{\delta}_n| \leq \tilde{\xi}_n$, which implies

$$\mathcal{I}_{1n} \leq \mathcal{I}_{2n} \leq \int_{-\tilde{\xi}_n}^{\tilde{\xi}_n} \tilde{\psi}'_n(z)^2 \tilde{g}_n(z)dz \overset{(a)}{\to}_p \mathcal{I}_{f_0},$$

where (a) follows from Lemma B.14 noting $\tilde{g}_n \in \mathcal{SLC}_0$. On the other hand,

Fatou's lemma and Condition 1 indicates that

$$\liminf_n \mathcal{I}_{2n} \geq \mathcal{I}_{1n} \geq \int_{-\infty}^{\infty} \liminf_n \left(1_{\mathcal{A}_n}(z)\tilde{\psi}_n'(z)^2 \tilde{g}_n(z)\right) dz = \mathcal{I}_{f_0}.$$

Therefore, $\mathcal{I}_{1n}, \mathcal{I}_{2n} \to_p \mathcal{I}_{f_0}$.

(g) $\|\tilde{g}_n - g_0\|_\infty \to_p 0$ by Condition 1.

(h) Since $\tilde{\delta}_n = O_p(n^{-1/2})$, Lemma B.6 yields $P(\mathcal{A}_n' \subset \text{int}(\text{dom}(\psi_0))) \to_n 1$.

Suppose we can show that as $k \to \infty$, $\mathbb{Y}_{n_k} \to 1$ on $\mathcal{M}$. Then it would establish that every subsequence of n has a further subsequence n_k along which $\mathbb{Y}_{n_k} \to_{a.s.} 1$. Then Fact 7 would yield $\mathbb{Y}_n \to_p 1$, as desired. For the sake of simplicity, we will drop k from the subscript from the definitions of $\mathcal{M}$ and $\mathbb{Y}_{n_k}$.

Now we derive some useful inequalities which hold on $\mathcal{M}$. First, Lemma B.4 implies that

$$\xi_n \leq \frac{-5 \log 2 + 2p \log n}{5 w_n}.$$

Because $w_n \to_n w_0 > 0$, there exists $C > 0$ so that $\xi_n \leq C \log n$ for all sufficiently large n on $\mathcal{M}$. Equation 40, on the other hand, implies that $|\psi_0'(\xi_n + |\tilde{\delta}_n|)|$ is of the order of ξ_n. Therefore for large enough C,

$$\limsup_n \sup_{t \in \mathcal{A}_n'} \psi_0'(t) = |\psi_0'(\xi_n + |\tilde{\delta}_n|)| \leq C \log n \quad \text{on } \mathcal{M}. \tag{30}$$

Here the monotonicity of ψ_0' was used to obtain the last equality. Also note that because $\mathcal{A}_n' \subset \text{int}(\text{dom}(\psi_0))$ for all sufficiently large n on $\mathcal{M}$, we can apply Lemma B.11 on g_0 to obtain

$$\limsup_n \sup_{t \in \mathcal{A}_n} \frac{g_0(t)}{g_0(t + \tilde{\delta}_n)} \leq \limsup_n e^{O(|\tilde{\delta}_n|\xi_n)} \overset{(a)}{=} 1 \quad \text{on } \mathcal{M}, \tag{31}$$

where (a) follows because $\tilde{\delta}_n \xi_n \to_n 0$ on $\mathcal{M}$.

Next we will establish the pointwise convergence of $b_n(t)$ on $\mathcal{M}$. Since $\|\tilde{g}_n - g_0\|_\infty \to 0$ on $\mathcal{M}$, Lemma B.8(B) holds on $\mathcal{M}$. Using Lemma B.8(B) and the mean value Theorem, we can show that on $\mathcal{M}$, $(\tilde{\psi}_n(t + \tilde{\delta}_n) - \tilde{\psi}_n(t))/\tilde{\delta}_n \to \psi_0'(t)$ for any $t \in \text{int}(\text{dom}(\psi_0))$ that is a continuity point of ψ_0'. Because ψ_0 is concave, ψ_0' is continuous Lebesgue almost everywhere on $\text{int}(\text{dom}(\psi_0))$ (see Corollary 25.5.1 and Theorem 25.5 of Rockafellar, 1970). Also noting $\widehat{\mathcal{I}}_n(\eta_n) \to_n \mathcal{I}_{f_0}$ on $\mathcal{M}$, we obtain that

$$g_0'(t) \frac{\int_t^{t+\tilde{\delta}_n} \tilde{\psi}_n'(z) dz}{\tilde{\delta}_n \widehat{\mathcal{I}}_n(\eta_n)} \to_n \frac{g_0'(t)\psi_0'(t)}{\mathcal{I}_{f_0}}, \quad \text{Lebesgue a.e. } t \in \text{int}(\text{dom}(\psi_0)) \text{ on } \mathcal{M}.$$

Since $\mathcal{A}_n \subset \text{int}(\text{dom}(\psi_0))$ for sufficiently large n, and $\xi_n \to G_0^{-1}(1)$ on $\mathcal{M}$, it follows that $1_{\mathcal{A}_n}(t)$ converges to $1_{\text{int}(\text{dom}(\psi_0))}(t)$ pointwise on $\mathcal{M}$ as well. Noting $g_0'(t) = \psi_0'(t)g_0(t)$ for all $t \in \text{int}(\text{dom}(\psi_0))$, we then obtain that on $\mathcal{M}$,

$$b_n(t) = 1_{\mathcal{A}_n}(t)g_0'(t) \frac{\int_t^{t+\tilde{\delta}_n} \tilde{\psi}_n'(z) dz}{\tilde{\delta}_n \widehat{\mathcal{I}}_n(\eta_n)} \to_n 1_{\text{int}(\text{dom}(\psi_0))}(t) \frac{\psi_0'(t)^2 g_0(t)}{\mathcal{I}_{f_0}} \tag{32}$$

for all $t \in \mathbb{R}$ except a set of Lebesgue measure zero. The concavity of $\tilde{\psi}_n$ implies that its right derivative $\tilde{\psi}'_n$ is non-increasing. Hence, for any $t \in \text{int}(\text{dom}(\psi_0))$, we have,

$$\min\left\{\tilde{\psi}'_n(t), \tilde{\psi}'_n(t + \tilde{\delta}_n)\right\} \le \frac{\displaystyle\int_t^{t+\tilde{\delta}_n} \tilde{\psi}'_n(z)dz}{\tilde{\delta}_n} \le \max\left\{\tilde{\psi}'_n(t), \tilde{\psi}'_n(t + \tilde{\delta}_n)\right\}, \quad (33)$$

yielding
$$\frac{\left|\displaystyle\int_t^{t+\tilde{\delta}_n} \tilde{\psi}'_n(z)dz\right|}{|\tilde{\delta}_n|} \le |\tilde{\psi}'_n(t)| + |\tilde{\psi}'_n(t + \tilde{\delta}_n)|. \quad (34)$$

Using (34), we can bound $|b_n(t)|$ noting

$$|b_n(t)| \le \underbrace{1[t \in \mathcal{A}_n]|\psi'_0(t)|g_0(t)\frac{|\tilde{\psi}'_n(t)|}{\widehat{\mathcal{I}}_n(\eta_n)}}_{b_{1n}(t)} + \underbrace{1[t \in \mathcal{A}_n]|\psi'_0(t)|g_0(t)\frac{|\tilde{\psi}'_n(t + \tilde{\delta}_n)|}{\widehat{\mathcal{I}}_n(\eta_n)}}_{b_{2n}(t)}.$$

Now defining

$$T_{21,n}(t) = 1[t \in \mathcal{A}_n]|\psi'_0(t)|\sqrt{g_0(t)}\left(\sqrt{g_0(t)} - \sqrt{\tilde{g}_n(t)}\right)\frac{|\tilde{\psi}'_n(t)|}{\widehat{\mathcal{I}}_n(\eta_n)},$$

$$T_{22,n}(t) = 1[t \in \mathcal{A}_n]|\psi'_0(t)|\sqrt{g_0(t)}\frac{|\tilde{\psi}'_n(t)|\sqrt{\tilde{g}_n(t)}}{\widehat{\mathcal{I}}_n(\eta_n)},$$

$$T_{23,n}(t) = 1[t \in \mathcal{A}_n]|\psi'_0(t)|\frac{g_0(t)}{\sqrt{g_0(t + \tilde{\delta}_n)}}\left(\sqrt{g_0(t + \tilde{\delta}_n)} - \sqrt{\tilde{g}_n(t + \tilde{\delta}_n)}\right)\frac{|\tilde{\psi}'_n(t + \tilde{\delta}_n)|}{\widehat{\mathcal{I}}_n(\eta_n)},$$

$$T_{24,n}(t) = 1[t \in \mathcal{A}_n]|\psi'_0(t)|\frac{g_0(t)}{\sqrt{g_0(t + \tilde{\delta}_n)}}\frac{|\tilde{\psi}'_n(t + \tilde{\delta}_n)|\sqrt{\tilde{g}_n(t + \tilde{\delta}_n)}}{\widehat{\mathcal{I}}_n(\eta_n)},$$

we note that

$$b_{1n}(t) = T_{21,n}(t) + T_{22,n}(t) \quad \text{and} \quad b_{2n}(t) = T_{23,n}(t) + T_{24,n}(t).$$

Thus we can upper bound $|b_n(t)|$ by $c_n(t)$ where

$$c_n(t) = T_{21,n}(t) + T_{22,n}(t) + T_{23,n}(t) + T_{24,n}(t).$$

Our aim is to apply Fact 10 (Pratt's Lemma) with $a_n = 0$ to prove the current Lemma. To this end, we first show that the following assertions hold on $\mathcal{M}$:

A1. $T_{21,n} \to_n 0$ and $T_{23,n} \to_n 0$ Lebesgue almost everywhere on $\mathbb{R}$.
A2. $\int_{\mathbb{R}} T_{21,n}(t)dt \to_n 0$ and $\int_{\mathbb{R}} T_{23,n}(t)dt \to_n 0$.
A3. There are functions $t_{22} : \mathbb{R} \mapsto \mathbb{R}$ and $t_{24} : \mathbb{R} \mapsto \mathbb{R}$ so that $T_{22} \to_n t_{22}$ and $T_{24} \to_n t_{24}(t)$ Lebesgue almost everywhere on $\mathbb{R}$.
A4. The functions t_{22} and t_{24} in A3 are integrable. Moreover, $\int_{\mathbb{R}} T_{22,n}(t)dt \to_n \int_{\mathbb{R}} t_{22}(t)dt$ and $\int_{\mathbb{R}} T_{24,n}(t)dt \to_n \int_{\mathbb{R}} t_{24}(t)dt$.

Let us denote $c(t) = t_{22}(t) + t_{24}(t)$. Then A1-A4 imply that on $\mathcal{M}$, $c_n(t) \to_n c(t)$ Lebesgue almost everywhere, c is integrable, and $\int_{\mathbb{R}} c_n(t)dt \to_n \int_{\mathbb{R}} c(t)dt$. Since $|b_n(t)| \le c_n(t)$ and (32) holds, Pratt's Lemma (see Fact 10) yields

$$\int_{-\infty}^{\infty} b_n(t)dt \to_n \int_{G_0^{-1}(0)}^{G_0^{-1}(1)} \frac{g_0'(t)^2 \psi_0'(t)dt}{\mathcal{I}_{f_0}} = 1 \quad \text{on } \mathcal{M},$$

which completes the proof of Lemma B.1.

Proof of A1 and A3:

Since $\|\tilde{g}_n - g_0\|_\infty \to_n 0$ on $\mathcal{M}$, Lemma B.8 implies that on $\mathcal{M}$, the functions $\tilde{g}_n, \tilde{g}_n(\cdot + \tilde{\delta}_n)$ converge pointwise to g_0, and $\tilde{\psi}_n', \tilde{\psi}_n'(\cdot + \tilde{\delta}_n)$ converge to ψ_0' Lebesgue almost everywhere on $\text{int}(\text{dom}(\psi_0))$. Continuity of g_0 implies $g_0(t + \tilde{\delta}_n) \to_n g_0(t)$ for all $t \in \mathbb{R}$. Using the above, it can be shown that

$$T_{21,n}, T_{23,n} \to_n 0, \quad T_{22,n}, T_{24,n} \to_n 1_{\text{int}(\text{dom}(\psi_0))} \psi_0'^2 g_0 / \mathcal{I}_{f_0} \quad \text{a.e. Lebesgue on } \mathcal{M}.$$

Proof of A2:

Using Cauchy-Schwarz inequality, the bound on g_0 from Fact 1, and the bound on ψ_0' from (30), we can show that there exists $C > 0$ such that the following holds for all sufficiently large n on $\mathcal{M}$:

$$\int_{-\infty}^{\infty} |T_{21,n}(t)|dt$$

$$\le C \log n \left(\int_{-\xi_n - \tilde{\delta}_n}^{\xi_n} \frac{\left(\sqrt{g_0(t)} - \sqrt{\tilde{g}_n(t)} \right)^2}{\tilde{g}_n(t)} \right)^{1/2} \left(\int_{\mathcal{A}_n} \frac{|\tilde{\psi}_n'(t)|^2 \tilde{g}_n(t)dt}{\widehat{\mathcal{I}}_n(\eta_n)^2} \right)^{1/2}$$

$$\le C \log n \frac{H(\tilde{g}_n, g_0)}{\left(\inf_{x \in \mathcal{A}_n'} \tilde{g}_n(x) \right)^{1/2}} \frac{\sqrt{\mathcal{I}_{1n}}}{\widehat{\mathcal{I}}_n(\eta_n)},$$

which approaches zero as $n \to \infty$ because

$$\frac{C(\log n) H(g_0, \tilde{g}_n)}{\left(\inf_{x \in \mathcal{A}_n'} \tilde{g}_n(x) \right)^{1/2}} \to_n 0, \quad \mathcal{I}_{1n}, \ \widehat{\mathcal{I}}_n(\eta_n) \to_n \mathcal{I}_{f_0} \quad \text{on } \mathcal{M}. \tag{35}$$

by (29). The proof for $T_{23,n}$ is similar. An application of the Cauchy-Schwarz inequality, the bound on ψ_0' by (30) and the bound in (31) imply that the following holds for all large n on $\mathcal{M}$:

$$\int_{-\infty}^{\infty} T_{23,n}(t)dt \le \frac{C(\log n) H(g_0, \tilde{g}_n)}{\left(\inf_{x \in \mathcal{A}_n'} \tilde{g}_n(x) \right)^{1/2}} \frac{\left(\int_{-\xi_n + \tilde{\delta}_n}^{\xi_n} \tilde{\psi}_n'(t)^2 \tilde{g}_n(t)dt \right)^{1/2}}{\widehat{\mathcal{I}}_n(\eta_n)}$$

$$= \frac{C(\log n)H(g_0, \tilde{g}_n)}{\left(\inf_{x \in \mathcal{A}'_n} \tilde{g}_n(x)\right)^{1/2}} \frac{\sqrt{\mathcal{I}_{2n}}}{\widehat{\mathcal{I}}_n(\eta_n)},$$

which converges to zero by (35) and the fact that $\mathcal{I}_{2n} \to_n \mathcal{I}_{f_0}$ on $\mathcal{M}$, thus completing the proof of A2.

Proof of A4:

Let us define

$$\mathcal{T}_{22,n}(t) = |T_{22,n}(t)|/g_0(t) \quad \text{and} \quad \mathcal{T}_{24,n}(t) = |T_{24,n}(t)|/g_0(t).$$

We will show that on $\mathcal{M}$, for each $\epsilon > 0$, there exists $\sigma > 0$ so that the following bounds are true for any G_0-measurable set $\mathcal{B} \subset \mathbb{R}$ satisfying $\int_{\mathcal{B}} g_0(t)dt < \sigma$:

$$\limsup_n \int_{\mathcal{B}} \mathcal{T}_{22,n}(t)g_0(t)dt < \epsilon \quad \text{and} \quad \limsup_n \int_{\mathcal{B}} \mathcal{T}_{24,n}(t)g_0(t)dt < \epsilon. \quad (36)$$

Next, we will show that

$$\limsup_n \int_{-\infty}^{\infty} \mathcal{T}_{22,n}(t)g_0(t)dt < \infty \text{ and } \limsup_n \int_{-\infty}^{\infty} \mathcal{T}_{24,n}(t)g_0(t)dt < \infty. \quad (37)$$

If (36) and (37) hold, Fact 13 underscores that the sequences $(\mathcal{T}_{22,n})_{n \geq 1}$ and $(\mathcal{T}_{24,n})_{n \geq 1}$ are uniform integrable with respect to the measure induced by G_0. Then A4 follows from A3 and Theorem 16.13 of Billingsley (2012) (Vitali convergence Theorem). Thus it suffices to show that (36) and (37) hold.

Note that since $\mathcal{I}_{f_0} < \infty$, by Fact 12, given any $\epsilon > 0$, we can choose $\sigma > 0$ so that for any G_0-measurable set $\mathcal{B} \subset \mathbb{R}$ satisfying $\int_{\mathcal{B}} g_0(t)dt < \sigma$, the integral $\int_{\mathcal{B}} \psi_0(x)^2 g_0(x) < \epsilon^2 \mathcal{I}_{f_0} = \epsilon'$ (say). It will soon be clear why this choice of ϵ' works. Using the Cauchy-Schwarz inequality in the third step, we calculate

$$\int_{\mathcal{B}} \mathcal{T}_{22,n}(t)g_0(t)dt = \int_{\mathcal{B}} T_{22,n}(t)dt = \int_{\mathcal{B} \cap \mathcal{A}_n} |\psi'_0(t)| \sqrt{g_0(t)} \frac{|\tilde{\psi}'_n(t)| \sqrt{\tilde{g}_n(t)}}{\widehat{\mathcal{I}}_n(\eta_n)}$$

$$\leq \left(\int_{\mathcal{B}} \psi'_0(t)^2 g_0(t)dt \right)^{1/2} \left(\int_{\mathcal{A}_n} \frac{\tilde{\psi}'_n(t)^2 \tilde{g}_n(t)dt}{\widehat{\mathcal{I}}_n(\eta_n)^2} \right)^{1/2},$$

which is bounded by $\sqrt{\epsilon' \mathcal{I}_{1n}}/\widehat{\mathcal{I}}_n(\eta_n)$. Noting $\mathcal{I}_{1n} \to_n \mathcal{I}_{f_0}$ and $\widehat{\mathcal{I}}_n(\eta_n) \to_n \mathcal{I}_{f_0}$ on $\mathcal{M}$, we obtain

$$\limsup_n \int_{\mathcal{B}} \mathcal{T}_{22,n}(t)g_0(t)dt \leq \sqrt{\epsilon'/\mathcal{I}_{f_0}} = \epsilon \quad \text{on } \mathcal{M}.$$

Letting $\mathcal{B} = \mathbb{R}$, and repeating the above steps, we can show that

$$\limsup_n \int_{\mathbb{R}} \mathcal{T}_{22,n}(t)g_0(t)dt < 1 \quad \text{on } \mathcal{M}.$$

For $\mathcal{T}_{24,n}$, the Cauchy-Schwarz inequality yields

$$\int_{\mathcal{B}} \mathcal{T}_{24,n}(t)g_0(t)dt \leq \left(\int_{\mathcal{B} \cap \mathcal{A}_n} \frac{g_0(t)^2 \psi'_0(t)^2}{g_0(t + \tilde{\delta}_n)} dt \right)^{1/2} \left(\int_{-\xi_n + \tilde{\delta}_n}^{\xi_n} \frac{\tilde{\psi}'_n(t)^2 \tilde{g}_n(t)dt}{\widehat{\mathcal{I}}_n(\eta_n)^2} \right)^{1/2}$$

$$\overset{(a)}{\leq} \sqrt{\epsilon'} \sup_{t \in \mathcal{A}_n} \left(\frac{g_0(t)}{g_0(t + \tilde{\delta}_n)} \right)^{1/2} \frac{\mathcal{I}_{2n}^{1/2}}{\widehat{\mathcal{I}}_n(\eta_n)}. \tag{38}$$

Here (a) follows because $\int_{\mathcal{B}} \psi_0'(t)^2 g_0(t) dt < \epsilon'$. The fact that $\mathcal{I}_{1n}, \widehat{\mathcal{I}}_n(\eta_n) \to_n \mathcal{I}_{f_0}$ on $\mathcal{M}$, in conjunction with the bound in (31), implies

$$\limsup_n \int_{\mathcal{B}} \mathcal{T}_{24,n}(t) g_0(t) dt \leq \sqrt{\epsilon'/\mathcal{I}_{f_0}} = \epsilon \quad \text{on } \mathcal{M}.$$

Thus (36) is proved. Letting $\mathcal{B} = \mathbb{R}$ leads to $\limsup_n \int_{\mathbb{R}} \mathcal{T}_{24,n}(t) g_0(t) dt < 1$ on $\mathcal{M}$, thus finishing the proof of (37). □

B.2. Auxilliary lemmas for the proof of Theorem 1

B.2.1. Lemmas on ξ_n:

Unless otherwise mentioned, for all the lemmas on ξ_n, ξ_n will denote $\tilde{G}_n^{-1}(1-\eta_n)$, where the choice of $\tilde{g}_n$ should be clear from the context.

Lemma B.2. *Suppose $\tilde{g}_n \in \mathcal{S}_0$ satisfies Conditions 1 and 2. Let $\eta_n = Cn^{-2p/5}$, where $C > 0$ and p is as in Condition 2. Then for $\xi_n = \tilde{G}_n^{-1}(1 - \eta_n)$, we have*

$$P\Big([-\xi_n, \xi_n] \subset int(\mathrm{dom}(\psi_0)) \Big) \to 1.$$

Proof of Lemma B.2. Using Fact 5 in step (a) we obtain that

$$|G_0(-\xi_n) - \tilde{G}_n(-\xi_n)| \leq d_{TV}(G_0, \tilde{G}_n) \overset{(a)}{\leq} \sqrt{2} H(\tilde{g}_n, g_0) = O_p(n^{-p})$$

by Condition 2. Therefore $G_0(-\xi_n) \geq \tilde{G}_n(-\xi_n) + o_p(n^{-p}) \geq \eta_n + o_p(n^{-p})$ because $F(F^{-1}(q)) \geq q$ for any distribution function F, and $q \in (0,1)$. Since $\eta_n = Cn^{-2p/5} \gg n^{-p}$, it follows that $P(G_0(-\xi_n) \geq \eta_n/2) \to 1$. Thus $P(-\xi_n \in int(dom(\psi_0))) \to_n 1$. Since $\psi_0 \in \mathcal{SC}_0$, $int(dom(\psi_0))$ is an interval of the form $(-a,a)$ for some $a > 0$. Noting $-\xi_n \in (-a,a)$ implies $[-\xi_n, \xi_n] \subset (-a,a)$, the proof follows. □

Lemma B.3. *Consider the set up of Lemma B.2. Then $\xi_n \to_p G_0^{-1}(1)$ as $\eta_n \to 0$.*

Proof of Lemma B.3. Note that $G_0^{-1}(1)$ takes value in $\mathbb{R} \cup \{\infty\}$. Lemma B.2 implies $P(\xi_n \leq G_0^{-1}(1)) \to 1$. Suppose, if possible, $\xi_n \to_p G_0^{-1}(1)$ does not hold. Then, since G_0^{-1} is continuous on $(0,1)$, there exists $t \in (0,1)$ so that

$$\limsup_n P(\xi_n < G_0^{-1}(t)) > 0.$$

However, because $\eta_n \to 0$, $1 - \eta_n \geq 2t$ for sufficiently large n, which yields $\xi_n \geq \tilde{G}_n^{-1}(2t)$. Now by Fact 11 and Condition 1, $G_n^{-1}(2t) \to_p G_0^{-1}(2t)$. However, $G_0^{-1}(2t) > G_0^{-1}(t)$, where the strict inequality follows because g_0 being log-concave, is positive $J(G_0)$. Therefore the proof follows by contradiction. □

Lemma B.4. *Suppose either $\tilde{g}_n \in \mathcal{SLC}_0$ is a density satisfying Condition 1 and Condition 2, or $\tilde{g}_n$ satisfies Condition 3. Let $\eta_n = Cn^{-2p/5}$ where p is as*

in Condition 2 (or Condition 3) and $C > 0$. Then

$$\xi_n \le \frac{-\log 2 + 2p(\log n)/5}{\omega_n}$$

where ω_n is as in Fact 4. In fact, $|\xi_n| = O_p(\log n)$.

Proof of Lemma B.4. Observe that if $\tilde{g}_n \in \mathcal{SLC}_0$, then $\tilde{g}_n(\tilde{G}_n^{-1}(z)) > 0$ for $z \in (0,1)$. If $\tilde{g}_n$ satisfies Condition 3, then also the above holds because by Condition 3, $\tilde{g}_n > 0$ on $\mathbb{R}$. Since $\tilde{g}_n$ is symmetric about zero, $0 = \tilde{G}_n^{-1}(1/2)$. Noting $\xi_n = -\tilde{G}_n^{-1}(\eta_n)$, we therefore derive that

$$\xi_n = \tilde{G}_n^{-1}(1/2) - \tilde{G}_n^{-1}(\eta_n) = \int_{\eta_n}^{1/2} \frac{dz}{\tilde{g}_n(\tilde{G}_n^{-1}(z))} \le \int_{\eta_n}^{1/2} \frac{dz}{\omega_n \tilde{G}_n(\tilde{G}_n^{-1}(z))},$$

where ω_n is as in Fact 4. Because $\tilde{g}_n > 0$ on $J(\tilde{G}_n)$, it follows that $\tilde{G}_n$ is continuous on $J(\tilde{G}_n)$. Therefore, we have $\tilde{G}_n(\tilde{G}_n^{-1}(z)) = z$, implying

$$\xi_n \le \frac{\log(1/2) - \log(\eta_n)}{\omega_n} = \frac{-\log 2 + 2p(\log n)/5}{\omega_n}.$$

Since $\omega_n \to_p \omega_0 > 0$ by Fact 4, the proof follows. $\qquad\square$

Lemma B.5. *Consider the set up of Lemma B.4. Let $\tilde{\xi}_n = \tilde{G}_n^{-1}(\eta_n/2)$. Suppose y_n is a sequence of non-negative random variables so that $P(y_n < \eta_n/(2g_0(0))) \to 1$. Then*

$$P([-\xi_n - y_n, \xi_n + y_n] \subset [-\tilde{\xi}_n, \tilde{\xi}_n]) \to 1. \tag{39}$$

Proof of Lemma B.5. Under our set up, $\tilde{g}_n$ is positive on the set $J(\tilde{G}_n)$. Therefore the function $\tilde{G}_n^{-1}$ is continuous on $(0,1)$. Hence the mean value theorem implies

$$\tilde{G}_n^{-1}(\eta_n) - \tilde{G}_n^{-1}(\eta_n/2) = \frac{\eta_n}{2\tilde{g}_n(\tilde{G}_n^{-1}(t))} \ge \frac{\eta_n}{2\|\tilde{g}_n\|_\infty}$$

for some $t \in [\eta_n/2, \eta_n]$. Condition 1 implies that $\|\tilde{g}_n\|_\infty \to_p \|g_0\|_\infty = g_0(0)$. Therefore, as $n \to \infty$,

$$P\left(\liminf_n \frac{\tilde{G}_n^{-1}(\eta_n) - \tilde{G}_n^{-1}(\eta_n/2)}{\eta_n} \ge \frac{1}{2g_0(0)} \right) \to 1.$$

Hence if $y_n < \eta_n/(2g_0(0))$, then $\tilde{G}_n^{-1}(\eta_n) - y_n \ge \tilde{G}_n^{-1}(\eta_n/2)$ with probability tending to one. Since $\tilde{g}_n$ is symmetric about zero, we obtain that $\tilde{G}_n^{-1}(1 - \eta_n) + y_n < \tilde{G}_n^{-1}(1 - \eta_n/2)$ with probability tending to one. Since $\xi_n = \tilde{G}_n^{-1}(1 - \eta_n)$ and $\tilde{\xi}_n = \tilde{G}_n^{-1}(1 - \eta_n/2)$, the proof follows. $\qquad\square$

Lemma B.6. *Consider the set up of Lemma B.4. Then for $y_n = o_p(\eta_n)$, we have*

$$P\Big([-\xi_n - |y_n|, \xi_n + |y_n|] \subset int(\mathrm{dom}(\psi_0)) \Big) \to 1.$$

Proof of Lemma B.6. Letting $\tilde{\xi}_n = -\tilde{G}_n^{-1}(\eta_n/2)$, and applying Lemma B.2, we obtain $P([-\tilde{\xi}_n, \tilde{\xi}_n] \subset int(\mathrm{dom}(\psi_0))) \to 1$. Then the result follows from Lemma B.5. $\qquad\square$

*B.2.2. **lemmas on** $\tilde{g}_n$ **and** g_0:*

Lemma B.7. *Suppose $\xi_n = \tilde{G}_n^{-1}(1-\eta_n)$ where $\tilde{g}_n \in \mathcal{SLC}_0$ satisfies Condition 1. Then*

  A. $\sup_{x \in [-\xi_n, \xi_n]} \tilde{g}_n(x) = O_p(1)$ and $\sup_{x \in [-\xi_n, \xi_n]} \tilde{g}_n(x)^{-1} = O_p(\eta_n^{-1})$.
  B. $\tilde{\psi}_n = \log \tilde{g}_n$ *satisfies* $\sup_{x \in [-\xi_n, \xi_n]} \tilde{\psi}_n(x) = O_p(1)$. *For* $\eta_n = Cn^{-2p/5}$ *with*
      $p \in (0,1)$ *and* $C > 0$, *we have* $\sup_{x \in [-\xi_n, \xi_n]} (-\tilde{\psi}_n(x)) = O_p(\log n)$.

Proof of Lemma B.7. The upper bound on $\tilde{g}_n$ follows from Fact 1 and Condition 1. For the upper bound on $\tilde{g}_n^{-1}$, note that Fact 4 implies that

$$\tilde{g}_n(x) \geq w_n \min(\tilde{G}_n(x), 1 - \tilde{G}_n(x)), \quad \text{for all } x \in \mathbb{R}.$$

Since $\tilde{G}_n$ is a non-decreasing and $1 - \tilde{G}_n$ is a non-increasing function, any $x \in [-\xi_n, \xi_n]$ satisfies

$$\tilde{g}_n(x) \geq \omega_n \min(\tilde{G}_n(-\xi_n), 1 - \tilde{G}_n(\xi_n)) = \omega_n \eta_n$$

because $\tilde{G}_n(-\xi_n) = \eta_n$. Since the random variable $\omega_n \to_p \omega_0 > 0$ by Fact 4, part A of the current lemma follows. Part B follows directly from Part A. $\quad\square$

Lemma B.8. *Assume $f_0 \in \mathcal{P}_0$. Suppose $\{\tilde{g}_n\}_{n \geq 1}$ is a sequence of log-concave densities satisfying $\|\tilde{g}_n - g_0\|_\infty \to_n 0$. Then the followings hold for any $y_n \to_n 0$:*
  (A) *Let $\tilde{\psi}_n = \log \tilde{g}_n$. Then $\tilde{\psi}_n(\cdot + y_n) \to_n \psi_0$ everywhere on $\mathrm{int}(\mathrm{dom}(\psi_0))$.*
  (B) *$\tilde{\psi}_n'(\cdot + y_n) \to_n \psi_0'$ Lebesgue almost everywhere on $\mathrm{int}(\mathrm{dom}(\psi_0))$. In particular, if x is a continuity point of ψ_0', then $\tilde{\psi}_n'(x + y_n) \to_n \psi_0'$.*

Proof of Lemma B.8. Since g_0 is continuous, it follows that $\sup_{x \in \mathbb{R}} |\tilde{g}_n(x+y_n) - g_0(x)| \to_n 0$. Since for each $x \in \mathrm{int}(\mathrm{dom}(\psi_0))$, there exists an open neighborhood around x where $|\psi_0| < \infty$, $\tilde{\psi}_n(x + y_n) \to_n \psi_0(x)$ for each $x \in \mathrm{int}(\mathrm{dom}(\psi_0))$. Therefore part (A) follows. For part (B), first note that if $x \in \mathrm{int}(\mathrm{dom}(\psi_0))$ is a continuity point of ψ_0', then $\tilde{\psi}_n'(x+y_n) \to_n \tilde{\psi}_n'(x)$ by Theorem 25.7 of Rockafellar (1970). Now since ψ_0 is concave, ψ_0 is continuously differentiable at x if it is differentiable at x (Rockafellar, 1970, Corollary 25.5.1). However, a concave ψ_0 is differentiable Lebesgue almost everywhere on $\mathrm{dom}(\phi_0)$ (Rockafellar, 1970, Theorem 25.5). Therefore, the lemma follows. $\quad\square$

*B.2.3. **Lemmas on** g_0:*

Lemma B.9. *Suppose g_0 satisfies Assumption A. Let κ be as in Assumption A. Then for any $x \in \mathrm{dom}(\psi_0)$, ψ_0 satisfies*

$$|\psi_0(x)| \leq |\psi_0(0)| + \kappa x^2.$$

In particular, if $\eta_n = Cn^{-p}$ for $p \in (0,1)$ and $C > 0$, then under the set up of Lemma B.4, $\xi_n = \tilde{G}_n^{-1}(1 - \eta_n)$ satisfies

$$\sup_{x \in [-\xi_n, \xi_n]} |\psi_0(x)| = O_p((\log n)^2).$$

proof of Lemma B.9. Because $\psi_0 \in \mathcal{SC}_0$, zero is the mode of ψ_0. Therefore, the upper bound on ψ_0 follows since $\psi_0(x) < \psi_0(0)$. For the lower bound, first note that the concavity of ψ_0 indicates that if $x \geq 0$ and $x \in \mathrm{dom}(\psi_0)$, then

$$\psi_0(x) \geq \psi_0(0) + \psi_0'(x-)x.$$

By our notation, $\psi_0'(x+) = \psi_0'(x)$. Noting Assumption A implies $\psi_0'(x-) \geq \psi_0'(0-) - \kappa x$, we derive

$$\psi_0(x) \geq \psi_0(0) + \psi_0'(0-)x - \kappa x^2.$$

Since $\psi_0'(0-) \geq 0$, the above yields $\psi_0(x) \geq \psi_0(0) - \kappa x^2$ for all $x \geq 0$. Since ψ_0 is symmetric about zero, we derive that $\psi_0(x) \geq \psi_0(0) - \kappa x^2$ for all $x \in \mathbb{R}$. In conjunction with the fact that $\psi_0(x) \leq \psi_0(0)$, the latter implies $|\psi_0(x)| \leq |\psi_0(0)| + \kappa x^2$ for all $x \in \mathbb{R}$. Since $P([-\xi_n, \xi_n] \subset \mathrm{dom}(\psi_0)) \to 1$ by Lemma B.2, the rest of the proof follows noting $\xi_n = O_p(\log n)$ for $\eta_n = n^{-2p/5}$ by Lemma B.4. $\qquad\square$

Lemma B.10. *Suppose $\psi_0 \in \mathcal{SC}_0$ satisfies Assumption A. Further suppose η_n is as in Lemma B.9 and $y_n > 0$ satisfies $y_n = o_p(\eta_n)$. Then*

$$\sup_{x \in [-\xi_n - y_n, \xi_n + y_n]} |\psi_0'(x)| = O_p(-\log(\eta_n)).$$

Proof of Lemma B.10. Since $\psi_0 \in \mathcal{SC}_0$, ψ_0' attains its maxima on any interval at the endpoints. Lemma B.6 implies $[-\xi_n - y_n, \xi_n + y_n] \subset \mathrm{dom}(\psi_0)$ with probability approaching one. Therefore Assumption A implies

$$\psi_0'(-\xi_n - y_n) \leq |\psi_0'(0)| + \kappa(|\xi_n| + y_n). \tag{40}$$

Rest of the proof follows from Lemma B.4 and the fact that $y_n = o_p(1)$. $\qquad\square$

Lemma B.11. *Under the set up of Theorem 1, there exists $C > 0$ so that if $b > 0$ satisfies $[-b - |\tilde{\delta}_n|, b + |\tilde{\delta}_n|] \subset int(\mathrm{dom}(\psi_0))$, then*

$$\sup_{t \in [-b,b]} \frac{g_0(t)}{g_0(t + \tilde{\delta}_n)} \leq e^{|\tilde{\delta}_n|(C + \kappa b + \kappa|\tilde{\delta}_n|)}.$$

Proof of Lemma B.11. Recalling that we use ψ_0' to denote the right derivative of ψ_0, we obtain

$$\frac{g_0(t)}{g_0(t + \tilde{\delta}_n)} = \exp\Big(\psi_0(t) - \psi_0(t + \tilde{\delta}_n)\Big)$$

$$\overset{(a)}{\leq} \exp\Big(|\tilde{\delta}_n| \max\{|\psi_0'(t)|, |\psi_0'(t + \tilde{\delta}_n)|\}\Big)$$

where (a) follows from (33). If t, $t - \tilde{\delta}_n \in int(\mathrm{dom}(\psi_0))$, by Assummption A, it also holds that $|\psi_0'(t)| \leq \psi_0'(0-) + \kappa|t|$ and $|\psi_0'(t + \tilde{\delta}_n)| \leq \psi_0'(0-) + \kappa|t| + \kappa|\tilde{\delta}_n|$. Thus for $C = \psi_0'(0-)$, we obtain that

$$\frac{g_0(t)}{g_0(t + \tilde{\delta}_n)} \leq e^{|\tilde{\delta}_n|(C + \kappa|t| + \kappa|\tilde{\delta}_n|)},$$

from which, the result follows. $\qquad\square$

B.2.4. **Lemmas regarding** $\tilde{\psi}_n$**:**

Lemma B.12. *Suppose* $\tilde{g}_n$ *satisfies Condition 1 and Condition 2 with* $p \in (0,1)$. *Further suppose* a_n *and* $\tilde{g}_n$ *satisfies*

$$\sup_{x \in [-a_n, a_n]} (|\psi_0(x)| + |\tilde{\psi}_n(x)|) = O_p((\log n)^2). \tag{41}$$

Then

$$\int_{-a_n}^{a_n} (\tilde{\psi}_n(x) - \psi_0(x))^2 g_0(x) dx = O_p((\log n)^2 n^{-2p}),$$

$$\int_{-a_n}^{a_n} (\tilde{\psi}_n(x) - \psi_0(x))^2 \tilde{g}_n(x) dx = O_p((\log n)^2 n^{-2p}).$$

In particular, if $\tilde{g}_n \in \mathcal{SLC}_0$*, then* (41) *holds with* $a_n = \xi_n(\tilde{G}_n) = \tilde{G}_n^{-1}(1 - \eta_n)$*, where* $\eta_n = C n^{-2p/5}$ *for some* $C > 0$.

Proof of Lemma B.12. We first invoke an algebraic fact. For any $x, y > 0$,

$$(\sqrt{x} - \sqrt{y})^2 = \min(x, y)\left(\sqrt{\frac{\max(x,y)}{\min(x,y)}} - 1\right)^2 = \min(x,y)\left(e^{|\log x - \log y|/2} - 1\right)^2.$$

Since for any $z > 0$, z and $z^2/2$ are bounded above by $e^z - 1$, it follows that

$$\left(e^{|\log x - \log y|/2} - 1\right)^2 \geq (\log x - \log y)^2/4, (\log x - \log y)^4/8^2.$$

Thus

$$4\int_{-a_n}^{a_n} (\sqrt{\tilde{g}_n(x)} - \sqrt{g_0(x)})^2 dx \geq \int_{-a_n}^{a_n} \min(\tilde{g}_n(x), g_0(x))(\tilde{\psi}_n(x) - \psi_0(x))^2 dx$$

$$8^2 \int_{-a_n}^{a_n} (\sqrt{\tilde{g}_n(x)} - \sqrt{g_0(x)})^2 dx \geq \int_{-a_n}^{a_n} \min(\tilde{g}_n(x), g_0(x))(\tilde{\psi}_n(x) - \psi_0(x))^4 dx. \tag{42}$$

Therefore,

$$\int_{-a_n}^{a_n} (\tilde{\psi}_n(x) - \psi_0(x))^2 g_0(x) dx$$

$$= \int_{-a_n}^{a_n} (\tilde{\psi}_n(x) - \psi_0(x))^2 \Big(g_0(x) - \min(g_0(x), \tilde{g}_n(x))\Big) dx$$

$$+ \int_{-a_n}^{a_n} (\tilde{\psi}_n(x) - \psi_0(x))^2 \min(g_0(x), \tilde{g}_n(x)) dx$$

$$\overset{(a)}{=} \int_{-a_n}^{a_n} (\tilde{\psi}_n(x) - \psi_0(x))^2 (g_0(x) - \tilde{g}_n(x)) 1_{[\tilde{g}_n < g_0]} dx + O_p(n^{-2p})$$

$$= \underbrace{\int_{-a_n}^{a_n} (\tilde{\psi}_n(x) - \psi_0(x))^2 (\sqrt{g_0(x)} - \sqrt{\tilde{g}_n(x)})^2 1_{[\tilde{g}_n < g_0]} dx}_{T_1}$$

$$-2\underbrace{\int_{-a_n}^{a_n}(\tilde{\psi}_n(x)-\psi_0(x))^2\sqrt{\tilde{g}_n(x)}(\sqrt{\tilde{g}_n(x)}-\sqrt{g_0(x)})1_{[\tilde{g}_n<g_0]}dx}_{T_2}+O_p(n^{-2p})$$

where (a) follows from (42) and Condition 2. We can upper bound $|\psi_0(x)-\tilde{\psi}_n(x)|$ noting

$$\sup_{x\in[-a_n,a_n]}|\psi_0(x)-\tilde{\psi}_n(x)|\leq\sup_{x\in[-a_n,a_n]}(|\psi_0(x)|+|\tilde{\psi}_n(x)|)=O_p((\log n)^2)$$

by (41). Therefore $T_1\leq O_p((\log n)^2)H(\tilde{g}_n,g_0)^2$, which is $O_p(n^{-2p})$. On the other hand, noting T_2 can be written as

$$T_2=\int_{-a_n}^{a_n}(\tilde{\psi}_n(x)-\psi_0(x))^2\sqrt{\min(\tilde{g}_n(x),g_0(x))}(\sqrt{\tilde{g}_n(x)}-\sqrt{g_0(x)})1_{[\tilde{g}_n<g_0]}dx,$$

by an application of the Cauchy-Schwarz inequality, we derive

$$|T_2|\leq\left(\int_{-a_n}^{a_n}(\tilde{\psi}_n(x)-\psi_0(x))^4\min(\tilde{g}_n(x),g_0(x))dx\right)^{1/2}H(\tilde{g}_n,g_0),$$

which, by (42) and Condition 2, is $O_p(n^{-2p})$, thus completing the proof of the first part.

It remains to show that (41) holds when $\tilde{g}_n\in\mathcal{SLC}_0$ and $a_n=\tilde{G}_n^{-1}(1-\eta_n)$. Lemma B.9 entails that this a_n satisfies

$$\sup_{x\in[-a_n,a_n]}|\psi_0(x)|=O_p((\log n)^2).\tag{43}$$

The proof of the current lemma then follows noting Lemma B.7 implies

$$\sup_{x\in[-\xi_n,\xi_n]}|\tilde{\psi}_n(x)|=O_p(\log n).\qquad\qquad\square$$

B.2.5. **Lemmas regarding $\tilde{\psi}_n'$:**

Lemma B.13. *Let $\rho_n=\eta_n/\log n$. Suppose $\tilde{g}_n$ is a log-concave density satisfying Condition 1 and Condition 2. Let a_n be a positive sequence satisfying (41) such that $a_n=O_p(\log n)$,*

$$[-a_n-\rho_n,a_n+\rho_n]\subset int(\mathrm{dom}(\psi_0))\cap int(\mathrm{dom}(\tilde{\psi}_n)),\tag{44}$$

$$P(\tilde{G}_n(-a_n)>\eta_n/4,1-\tilde{G}_n(a_n)>\eta_n/4)\to1.\tag{45}$$

Then

$$\int_{-a_n}^{a_n}\left(\tilde{\psi}_n'(z)-\psi_0'(z)\right)^2dz=O_p((\log n)^3n^{-4p/5}),$$

$$\int_{-a_n}^{a_n}\left(\tilde{\psi}_n'(z)-\psi_0'(z)\right)^2\mu_n(z)dz=O_p((\log n)^3n^{-4p/5})),$$

for any density μ_n such that $\|\mu_n\|_\infty=O_p(1)$, where p is as in Condition 2. In particular, the lemma holds if $\tilde{g}_n\in\mathcal{SLC}_0$ and $a_n=\xi_n(\tilde{G}_n)=\tilde{G}_n^{-1}(1-\eta_n)$, where $\eta_n=Cn^{-2p/5}$ for some $C>0$.

Proof of Lemma B.13. Since $\tilde{\psi}_n$ is concave and $\rho_n > 0$, any $z \in \mathrm{dom}(\tilde{\psi}_n)$ satisfies

$$\frac{\tilde{\psi}_n(z+\rho_n) - \tilde{\psi}_n(z)}{\rho_n} \leq \tilde{\psi}'_n(z+) \leq \tilde{\psi}'_n(z-) \leq \frac{\tilde{\psi}_n(z) - \tilde{\psi}_n(z-\rho_n)}{\rho_n}. \qquad (46)$$

Now suppose (44) holds. Then the quantities

$$\Delta_n^+(z) = \frac{\tilde{\psi}_n(z+\rho_n) - \tilde{\psi}_n(z)}{\rho_n} - \frac{\psi_0(z+\rho_n) - \psi_0(z)}{\rho_n},$$

and

$$\Delta_n^-(z) = \frac{\tilde{\psi}_n(z) - \tilde{\psi}_n(z-\rho_n)}{\rho_n} - \frac{\psi_0(z) - \psi_0(z-\rho_n)}{\rho_n}$$

are well defined for all $z \in [-a_n, a_n]$. Recalling $\tilde{\psi}'_n(z) = \tilde{\psi}'_n(z+)$ and $\psi'_0(z) = \psi'_0(z+)$ by our notation, we can then show that under (44),

$$\tilde{\psi}'_n(z) - \psi'_0(z) \overset{(a)}{\leq} \frac{\tilde{\psi}_n(z) - \tilde{\psi}_n(z-\rho_n)}{\rho_n} - \frac{\psi_0(z) - \psi_0(z-\rho_n)}{\rho_n}$$
$$+ \left(\frac{\psi_0(z) - \psi_0(z-\rho_n)}{\rho_n} - \psi'_0(z) \right) \overset{(b)}{\leq} \Delta_n^-(z) + \kappa\rho_n$$

for all $z \in [-a_n, a_n]$, where (a) follows by (46), and (b) follows because

$$\frac{\psi_0(z) - \psi_0(z-\rho_n)}{\rho_n} - \psi'_0(z) \leq \rho_n^{-1} \left| \int_{z-\rho_n}^z \left(\psi'_0(t) - \psi'_0(z) \right) dt \right| \leq \kappa\rho_n/2$$

since Assumption A applies on the set $[z - \rho_n, z] \subset \mathrm{int}(\mathrm{dom}(\psi_0))$. Similarly, we can show that

$$\Delta_n^+(z) - \kappa\rho_n \leq \tilde{\psi}'_n(z) - \psi'_0(z) \quad \text{for all } z \in [-a_n, a_n],$$

provided (44) holds. Thus we have established that

$$|\tilde{\psi}'_n(z) - \psi'_0(z)| \leq \max\{\Delta_n^+(z), \Delta_n^-(z)\} + \kappa\rho_n \qquad (47)$$

whenever (44) holds. Now observe that the integral $\int_{-a_n}^{a_n} \Delta_n^+(z)^2 dz$ is well defined under (44), and equals

$$\int_{-a_n}^{a_n} \left(\frac{\tilde{\psi}_n(z+\rho_n) - \tilde{\psi}_n(z)}{\rho_n} - \frac{\psi_0(z+\rho_n) - \psi_0(z)}{\rho_n} \right)^2 dz$$

$$\leq 2 \int_{-a_n}^{a_n} \left(\frac{\tilde{\psi}_n(z+\rho_n) - \psi_0(z+\rho_n)}{\rho_n} \right)^2 dz + 2 \int_{-a_n}^{a_n} \left(\frac{\tilde{\psi}_n(z) - \psi_0(z)}{\rho_n} \right)^2 dz$$

$$\overset{(a)}{\leq} \frac{2}{\rho_n^2 \min(\tilde{g}_n(a_n+\rho_n), \tilde{g}_n(-a_n+\rho_n))} \int_{-a_n}^{a_n} (\tilde{\psi}_n(z+\rho_n)$$
$$- \psi_0(z+\rho_n))^2 \tilde{g}_n(z+\rho_n) dz$$
$$+ \frac{2}{\rho_n^2 \min(\tilde{g}_n(a_n), \tilde{g}_n(-a_n))} \int_{-a_n}^{a_n} (\tilde{\psi}_n(z) - \psi_0(z))^2 \tilde{g}_n(z) dz$$

$$\overset{(b)}{=} \frac{O_p((\log n)^2 n^{-2p})}{\rho_n^2 \min(\tilde{g}_n(a_n+\rho_n), \tilde{g}_n(-a_n-\rho_n))},$$

where (a) follows because $\tilde{g}_n$ being log-concave, and hence unimodal, attains

minimum over an interval at either of the endpoints; and (b) follows from Lemma B.12 and the fact that a_n and ρ_n are positive. Let us define

$$\epsilon_n(\rho_n) = \frac{(\log n)^2 n^{-2p}}{\rho_n^2 \min(\tilde{g}_n(a_n + \rho_n), \tilde{g}_n(-a_n - \rho_n))}. \tag{48}$$

Since (44) holds with probability tending to one by our assumption, we can write $\int_{-a_n}^{a_n} \Delta_n^+(z)^2 dz = O_p(\epsilon_n(\rho_n))$. Similarly, we can show that $\int_{-a_n}^{a_n} \Delta_n^-(z) dz$ is $O_p(\epsilon_n(\rho_n))$. The above, combined with (47), leads to

$$\int_{-a_n}^{a_n} (\tilde{\psi}_n'(z) - \psi_0'(z))^2 dz \leq 2 \int_{-a_n}^{a_n} \Delta_n^-(z)^2 dz + 2 \int_{-a_n}^{a_n} \Delta_n^+(z)^2 dz + 4\kappa^2 \rho_n^2 a_n$$

$$= O_p(\epsilon_n(\rho_n)) + O_p(\rho_n^2 a_n).$$

Note that $\rho_n^2 a_n = O(n^{-4p/5}/\log n)$ because $\rho_n = \eta_n/\log n$ and $a_n = O_p(\log n)$ by our assumption. Also, $\int_{-a_n}^{a_n} (\tilde{\psi}_n'(z) - \psi_0'(z))^2 \mu_n(z) dz$ can be bounded by

$$2\|f\|_\infty \left(\int_{-a_n}^{a_n} \Delta_n^-(z)^2 dz + \int_{-a_n}^{a_n} \Delta_n^+(z)^2 dz \right) + 2\kappa^2 \rho_n^2$$

which is $O_p(\epsilon_n(\rho_n)) + O_p(n^{-4/5}/(\log n)^2)$ because $\|\mu_n\|_\infty$ is $O_p(1)$ and ρ_n equals $\eta_n/\log n$. To prove the first part of the lemma, it only remains to show that

$$\epsilon_n(\rho_n) = O_p((\log n)^3 n^{-4p/5}). \tag{49}$$

Since $a_n - \rho_n \in \text{int}(\text{dom}(\tilde{\psi}_n))$ under (44), Fact 4 implies

$$\tilde{g}_n(-a_n - \rho_n) \geq \omega_n \tilde{G}_n(-a_n - \rho_n) = \omega_n \tilde{G}_n(-a_n) - \omega_n \int_{-a_n-\rho_n}^{-a_n} \tilde{g}_n(z) dz$$

under (44), where ω_n is as in fact 4. Note that

$$\left| \int_{-a_n-\rho_n}^{-a_n} \tilde{g}_n(z) dz \right| \leq \rho_n \|\tilde{g}_n\|_\infty.$$

Also since $P(\tilde{G}_n(-a_n) > \eta_n/4) \to 1$ by our assumption, the followings hold with probability tending to one,

$$\tilde{g}_n(-a_n - \rho_n) \geq \omega_n(\eta_n/4 - \rho_n \|\tilde{g}_n\|_\infty) \overset{(a)}{\geq} \omega_n \eta_n/8,$$

where (a) follows because $\rho_n = o(\eta_n)$ and $\|\tilde{g}_n\|_\infty = O_p(1)$ by Condition 1 and Fact 1. However, since $\omega_n \to_p \omega_0$ by Fact 4, the last display implies $\tilde{g}_n(-a_n - \rho_n)^{-1}$ is $O_p(1/\eta_n)$. Similarly, we can show that $\tilde{g}_n(a_n + \rho_n)^{-1}$ is $O_p(1/\eta_n)$. Thus

$$\epsilon_n(\rho_n) = \frac{O_p((\log n)^2 n^{-2p})}{\eta_n \rho_n^2}$$

follows. Since $\eta_n = n^{-2p/5}$ and $\rho_n = \eta_n/\log n$, (49) follows, thus completing the proof of the first part of Lemma B.13.

Now suppose $\tilde{g}_n \in \mathcal{SLC}_0$ and $a_n = \xi_n(\tilde{G}_n) = \tilde{G}_n^{-1}(1 - \eta_n)$. They satisfy (41) by Lemma B.12. Also $\xi_n = O_p(\log n)$ by Lemma B.4. Noting $\rho_n = o(\eta_n)$, (44) follows from Lemma B.5 and Lemma B.6. Since (45) trivially holds, second part of Lemma B.13 also follows. $\square$

Lemma B.14. *Suppose $\tilde{g}_n$ satisfies Condition 1 and Condition 2. Let a_n be a sequence of positive random variables satisfying*

$$a_n = O_p(\log n), \quad P\Big(a_n \in int(\mathrm{dom}(\psi_0))\Big) \to 1, \quad and \quad a_n \to_p G_0^{-1}(1). \quad (50)$$

Further suppose that $\tilde{g}_n$ and a_n satisfy

$$\int_{-a_n}^{a_n} (\tilde{\psi}_n'(z) - \psi_0'(z))^2 \tilde{g}_n(z) dz = O_p((\log n)^3 n^{-4p/5}), \quad (51)$$

where p is as in Condition 2. Then $\int_{-a_n}^{a_n} \tilde{\psi}_n'(z)^2 \tilde{g}_n(z) dz \to_p \mathcal{I}_{f_0}$. In addition, if $\tilde{g}_n \in \mathcal{SLC}_0$, then $a_n = \xi_n(\tilde{G}_n) = (\tilde{G}_n)^{-1}(1 - \eta_n)$ where $\eta_n = Cn^{-2p/5}$ for some $C > 0$.

Proof of Lemma B.14. Note that

$$\int_{-a_n}^{a_n} \Big(\tilde{\psi}_n'(z)^2 \tilde{g}_n(z) - \psi_0'(z)^2 g_0(z)\Big) dz$$

$$= \int_{-a_n}^{a_n} \Big((\tilde{\psi}_n'(z) - \psi_0'(z))^2 \tilde{g}_n(z) + 2\psi_0'(z)\tilde{\psi}_n'(z)\tilde{g}_n(z)$$

$$- \psi_0'(z)^2 \tilde{g}_n(z) - \psi_0'(z)^2 g_0(z)\Big) dz$$

$$= \underbrace{\int_{-a_n}^{a_n} (\tilde{\psi}_n'(z) - \psi_0'(z))^2 \tilde{g}_n(z) dz}_{T_1} + 2\underbrace{\int_{-a_n}^{a_n} \psi_0'(z)(\tilde{\psi}_n'(z) - \psi_0'(z))\tilde{g}_n(z) dz}_{T_2}$$

$$+ \underbrace{\int_{-a_n}^{a_n} \psi_0'(z)^2 (\tilde{g}_n(z) - g_0(z)) dz}_{T_3},$$

It is clear that by our assumption, $T_1 = O_p((\log n)^3 n^{-4p/5})$, which is $o_p(1)$.

Because ψ_0' is a non-increasing odd function, on any interval, $|\psi_0'|$ attains its maximum at both end points. Therefore,

$$|T_2| \le |\psi_0'(a_n)| \int_{-a_n}^{a_n} |\tilde{\psi}_n'(z) - \psi_0'(z)|\tilde{g}_n(z) dz$$

$$\overset{(a)}{\le} |\psi_0'(a_n)|\left(\int_{-a_n}^{a_n} (\tilde{\psi}_n'(z) - \psi_0'(z))^2 \tilde{g}_n(z) dz\right)^{1/2},$$

where (a) follows by the Cauchy-Schwarz inequality. Thus, $|T_2| \le |\psi_0'(a_n)|\sqrt{T_1}$. However, Assumption A implies that $|\psi_0'(a_n)| \le |\psi_0'(0)| + \kappa a_n = O_p(\log n)$ provided $a_n \in \mathrm{int}(\mathrm{dom}(\psi_0))$. By our assumption on a_n, the latter holds with probability tending to one. Hence,

$$T_2 = O_p(\log n)\sqrt{T_1} = O_p((\log n)^{5/2} n^{-2p/5}) = o_p(1).$$

Finally, using the fact $|\psi_0'(a_n)| = O_p(\log n)$ again, we bound $|T_3|$ by

$$O_p((\log n)^2) d_{TV}(\tilde{g}_n, g_0) \overset{(a)}{\le} O_p((\log n)^2) H(\tilde{g}_n, g_0) \overset{(b)}{=} O_p((\log n)^2 n^{-p}) = o_p(1),$$

where (a) follows from Fact 5 and (b) follows noting $H(\tilde{g}_n, g_0) = O_p(n^{-p})$ by

Condition [2]. Thus we have shown that

$$\int_{-a_n}^{a_n} \left(\tilde{\psi}_n'(z)^2 \tilde{g}_n(z) - \psi_0'(z)^2 g_0(z) \right) dz = o_p(1).$$

Since $a_n \to_p G_0^{-1}(1)$ by our assumption, noting $G_0^{-1}(0) = -G_0^{-1}(1)$, we also obtain $-a_n \to_p G_0^{-1}(0)$. Hence,

$$\int_{-a_n}^{a_n} \psi_0'(z)^2 g_0(z) dz \to_p \int_{G_0^{-1}(0)}^{G_0^{-1}(1)} \psi_0'(z)^2 g_0(z) dz = \mathcal{I}_{f_0},$$

which completes the proof of the first part of Lemma B.14. Second part of Lemma B.14 follows noting $\xi_n = O_p(\log n)$ by Lemma B.4, $\xi_n \in \mathrm{int}(\mathrm{dom}(\psi_0))$ with probability tending to one by Lemma B.2, $\xi_n \to_p G_0^{-1}(1)$ by Lemma B.3, and $\tilde{g}_n \in \mathcal{SLC}_0$ satisfies (51) by Lemma B.13. □

Lemma B.15. *Consider the set up of Theorem 1. Suppose $\eta_n = Cn^{-2p/5}$, where $C > 0$ and p is as in Condition 2. Let y_n be a sequence of positive random variables such that $P(|y_n| \le \eta_n/(2g_0(0))) \to 1$. Then*

$$\sup_{x \in [-\xi_n - y_n, \xi_n + y_n]} |\tilde{\psi}_n'(x)| = O_p(\eta_n^{-1/2}) = O_p(n^{p/5}).$$

Proof of Lemma B.15. Let $q \in (0, 1/2)$. Since $\tilde{g}_n$, being log-concave, is positive on $\mathrm{int}(J(\tilde{G}_n))$, using Fact 8 we obtain that

$$\int_{\tilde{G}_n^{-1}(q/2)}^{\tilde{G}_n^{-1}(q)} \tilde{\psi}_n'(x)^2 \tilde{g}_n(x) dx = \int_{q/2}^{q} \tilde{\psi}_n'(\tilde{G}_n^{-1}(z))^2 dz.$$

Note that $\tilde{\psi}_n'$ is non-increasing and positive on $(-\infty, -x]$, and $\tilde{g}_n(-x)$ is positive and non-decreasing on $(-\infty, -x]$. Thus $\tilde{\psi}_n' \circ \tilde{G}_n^{-1}$ is non-increasing. Therefore

$$q\tilde{\psi}_n'(\tilde{G}_n^{-1}(q))^2/2 \le \int_{\tilde{G}_n^{-1}(q/2)}^{\tilde{G}_n^{-1}(q)} \tilde{\psi}_n'(x)^2 \tilde{g}_n(x) dx \overset{(a)}{\le} \int_{\tilde{G}_n^{-1}(q/2)}^{\tilde{G}_n^{-1}(1-q/2)} \tilde{\psi}_n'(x)^2 \tilde{g}_n(x) dx$$

where (a) follows because $q < 1 - q/2$ for all $q \in (0, 1/2)$. Suppose $q = \eta_n/2$. Note that $\tilde{\xi}_n = -\tilde{G}_n^{-1}(\eta_n/2)$ and $|\tilde{\psi}_n'|$ is symmetric about zero. Then the last display leads to

$$\tilde{\psi}_n'(\tilde{\xi}_n)^2 \le \frac{4 \int_{\tilde{G}_n^{-1}(\eta_n/4)}^{\tilde{G}_n^{-1}(1-\eta_n/4)} \tilde{\psi}_n'(x)^2 \tilde{g}_n(x) dx}{\eta_n}.$$

From Lemma B.14 it follows that the integral converges in probability to $\mathcal{I}_{f_0}$. Therefore, $|\tilde{\psi}_n'(\tilde{\xi}_n)| = O_p(\eta_n^{-1/2})$ which implies

$$\sup_{x \in [-\tilde{\xi}_n, \tilde{\xi}_n]} |\tilde{\psi}_n'(x)| = O_p(\eta_n^{-1/2}). \tag{52}$$

The rest of the proof follows from (52) and Lemma B.5. □

Lemma B.16. *Consider the set up of Theorem 1. Then $\|h_n\|_{P_0,2}^2 = O_p(n^{-4p/5}(\log n)^3)$, where h_n is as defined in (21).*

Proof of Lemma B.16. Note that $\|h_n\|_{P_0,2}^2$ equals

$$\int_{-\xi_n}^{\xi_n} (\tilde{\psi}_n'(z) - \psi_0'(z - \tilde{\delta}_n))^2 f_0(z + \bar{\theta}_n) dz$$

$$\leq 2 \int_{-\xi_n}^{\xi_n} \left(\tilde{\psi}_n'(z) - \psi_0'(z) \right)^2 f_0(z + \bar{\theta}_n) dz$$

$$+ 2 \int_{-\xi_n}^{\xi_n} \left(\psi_0'(z - \tilde{\delta}_n) - \psi_0'(z) \right)^2 f_0(z + \bar{\theta}_n) dz$$

$$\leq 2 \underbrace{\int_{-\xi_n}^{\xi_n} \left(\tilde{\psi}_n'(z) - \psi_0'(z) \right)^2 g_0(z - \tilde{\delta}_n) dz}_{T_1}$$

$$+ 2 \underbrace{\int_{-\xi_n}^{\xi_n} \left(\psi_0'(z - \tilde{\delta}_n) - \psi_0'(z) \right)^2 g_0(z - \tilde{\delta}_n) dz}_{T_2}.$$

T_1 is $O_p(n^{-4p/5}(\log n)^3)$ by Lemma B.13. Since Assumption A implies ψ_0' is Lipschitz with constant κ on its domain, Lemma B.6 entails that T_2 is bounded by

$$2\kappa^2 \tilde{\delta}_n^2 \int_{-\xi_n}^{\xi_n} g_0(z - \tilde{\delta}_n) dz$$

which is $O_p(n^{-1})$ since $\tilde{\delta}_n = O_p(n^{-1/2})$. Hence, the proof follows. $\qquad\square$

B.2.6. *Lemma on consistency of Fisher information $\widehat{\mathcal{I}}_n(\eta_n)$:*

Lemma B.17. *Under the set up of Theorem 1, $\widehat{\mathcal{I}}_n(\eta_n) \to_p \mathcal{I}_{f_0}$ where $\widehat{\mathcal{I}}_n(\eta_n)$ is as defined in (8).*

of Lemma B.17 . Denoting $\tilde{\delta}_n = \theta_0 - \bar{\theta}_n$, we observe that

$$|\widehat{\mathcal{I}}_n(\eta_n) - \tilde{\mathcal{I}}_{f_0}(\eta_n)|$$

$$\leq \underbrace{\left| \int_{\bar{\theta}_n - \xi_n}^{\bar{\theta}_n + \xi_n} \tilde{\psi}_n'(x - \bar{\theta}_n)^2 d(\mathbb{F}_n - F_0)(x) \right|}_{T_1}$$

$$+ \underbrace{\left| \int_{-\xi_n}^{\xi_n} \tilde{\psi}_n'(x)^2 \Big(g_0(x - \tilde{\delta}_n) - g_0(x) \Big) dx \right|}_{T_2}$$

$$+ \underbrace{\left| \int_{-\xi_n}^{\xi_n} \tilde{\psi}_n'(x)^2 \Big(g_0(x) - \tilde{g}_n(x) \Big) dx \right|}_{T_3} + \underbrace{\left| \int_{-\xi_n}^{\xi_n} \tilde{\psi}_n'(x)^2 \tilde{g}_n(x) dx - \mathcal{I}_{f_0} \right|}_{T_4} \quad (53)$$

Let us consider the term T_1 first. Denoting $M_n = Cn^{p/5}$ as in the the proof of the first step of Theorem 1, we recall the class of functions $\mathcal{U}_n(M_n)$ defined

in (23).

In the same way we showed that $h_n \in \mathcal{H}_n(C)$ with high probability in the proof of the first step of Theorem 1, we can show that the function defined by

$$\tilde{h}_n(x) = \tilde{\psi}'_n(x - \bar{\theta}_n)^2 1_{[\bar{\theta}_n - \xi_n, \bar{\theta}_n + \xi_n]}(x)$$

is a member of the class

$$\mathcal{V}_n(C) = \left\{ h : \mathbb{R} \mapsto \mathbb{R} \; \middle| \; h(x) = u(x)^2 1_{[r_1, r_2]}(x), \; u \in \mathcal{U}_n(M_n), \right.$$

$$\left. [r_1, r_2] \subset [\theta_0 - C \log n, \theta_0 + C \log n] \cap \mathrm{int}(\mathrm{dom}(\phi_0)) \right\}$$

with high probability for all large n provided $C > 0$ is sufficiently large. Using (25), (26) and following some standard calculations, we can show that

$$\sup_Q \log N_{[\,]}(\epsilon, \mathcal{V}_n(C), L_2(Q)) \lesssim M_n^2 \epsilon^{-1},$$

where the supremum is over all probability measures on $\mathbb{R}$. Because bracketing number is larger than covering number, it also follows that

$$\sup_Q \log N(\epsilon, \mathcal{V}_n(C), L_2(Q)) \lesssim M_n^2 \epsilon^{-1}.$$

The definition of $\mathcal{U}_n(M_n)$ in (23) implies that the functions in $\mathcal{V}_n(C)$ are uniformly bounded by M_n^2. Since for any fixed $\epsilon > 0$,

$$M_n^2 \sup_Q \log N(\epsilon, \mathcal{V}_n(C), L_2(Q)) \lesssim M_n^4 \epsilon^{-1} = O(n^{-4p/5}),$$

Fact 14 leads to $E\|\mathbb{F}_n - F_0\|_{\mathcal{V}_n(C)} = o(1)$. Thus Markov inequality yields that $\|\mathbb{F}_n - F_0\|_{\mathcal{V}_n(C)} = o_p(1)$. Since for large C, $P(\tilde{h}_n \in \mathcal{V}_n(C))$ with high probability, it can be shown that $\int \tilde{h}_n d(\mathbb{F}_n - F_0) \to_p 0$, which establishes $T_1 = o_p(1)$.

Since the supremum of $|\tilde{\psi}'_n|$ over $[-\xi_n, \xi_n]$ is $O_p(n^{p/5})$ by Lemma B.15, we obtain that

$$T_2 \le O_p(n^{2p/5}) d_{TV}(g_0(\cdot - \tilde{\delta}_n), g_0) \overset{(a)}{\le} O_p(n^{2p/5}) H(g_0(\cdot - \tilde{\delta}_n), g_0) \overset{(b)}{=} O_p(\tilde{\delta}_n n^{2p/5}),$$

which is $o_p(1)$ because $\tilde{\delta}_n = O_p(n^{-1/2})$ and $p \in (0, 1)$. Here (a) and (b) follow from Fact 5 and Fact 15, respectively. In a similar way we can show that

$$T_3 \le O_p(n^{2p/5}) d_{TV}(\tilde{g}_n, g_0) \le O_p(n^{2p/5}) H(\tilde{g}_n, g_0),$$

which is $O_p(n^{-3p/5})$ by Condition 2. Finally, noting T_4 is also $o_p(1)$ by Lemma B.14, the proof follows from (53). □

Appendix C: Proof of proposition 1

We will first show that $\widehat{g}_{\bar{\theta}_n}$ and $\tilde{g}_n^{geo, sym}$ satisfy Condition 1. Then using this result, we will show in Lemma C.2 and Lemma C.3 that $\widehat{g}_{\bar{\theta}_n}$ and $\tilde{g}_n^{geo, sym}$ satisfy Condition 2, respectively. To show that Condition 1 holds for these two densities, we prove a general Proposition which states that Condition 1 holds for all the density estimators of g_0 we have discussed so far.

Proposition 2. *Suppose $f_0 \in \mathcal{P}_1$ and $\bar{\theta}_n$ is a consistent estimator of θ_0. Then $\tilde{g}_n = \widehat{g}_{\bar{\theta}_n}, \widehat{h}_n(\bar{\theta}_n \pm \cdot), \tilde{g}_n^{geo,sym}, \widehat{h}_n^{sm}(\bar{\theta}_n \pm \cdot),$ and $\tilde{g}_n^{sym,sm}$ satisfy Condition 1.*

The key step in proving Proposition 2 is showing that the L_1 consistency in Condition 1(A) holds, which is established by Lemma C.1. The proof of Lemma C.1 can be found in Appendix C.1.

Lemma C.1. *Suppose $\bar{\theta}_n \to_p \theta_0$, and $\tilde{g}_n$ is one among $\widehat{h}_n^{sm}(\bar{\theta}_n \pm \cdot), \widehat{h}_n(\bar{\theta}_n \pm \cdot),$ $\tilde{g}_n^{sym,sm}, \tilde{g}_n^{geo,sym},$ and $\widehat{g}_{\bar{\theta}_n}$. Then $\|\tilde{g}_n - g_0\|_1 \to_p 0$.*

Now we are ready to prove Proposition 2.

Proof of Proposition 2. As in the proof of Lemma C.1, one can show that it suffices to prove Proposition 2 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 we will verify Condition 1 when $\tilde{g}_n \in \mathcal{LC}$. Note that, this covers the case of $\widehat{g}_{\bar{\theta}_n}, \widehat{h}_n(\bar{\theta}_n \pm \cdot),$ $\tilde{g}_n^{geo,sym},$ and $\widehat{h}_n^{sm}(\bar{\theta}_n \pm \cdot)$. We will consider the case of $\tilde{g}_n^{sym,sm}$ separately because the latter is not log-concave.

Assuming $\tilde{g}_n \in \mathcal{LC}$, to verify part A of Condition 1, we first note that

$$\|\tilde{g}_n(\cdot + y_n) - g_0\|_1 \leq \|\tilde{g}_n(\cdot + y_n) - g_0(\cdot + y_n)\|_1 + \|g_0(\cdot + y_n) - g_0\|_1,$$

whose first term converges to zero almost surely by Lemma C.1. Also, since g_0 is continuous, $g_0(x + y_n)$ converges to $g_0(x)$ for each $x \in \mathbb{R}$. Therefore the second term also converges to zero almost surely by Glick's Theorem (Theorem 2.6, Devroye, 1987). Thus we obtain that $\|\tilde{g}_n(\cdot + y_n) - g_0\|_1 \to_{a.s.} 0$. Since $\tilde{g}_n(\cdot + y_n)$ is log-concave, the above, combined with Proposition 2(c) of Cule and Samworth (2010), yields that $\|\tilde{g}_n(\cdot + y_n) - g_0\|_\infty \to_{a.s.} 0$ which completes the verification of part A of Condition 1. As a consequence,

$$\tilde{\psi}_n(x + y_n) = \log(\tilde{g}_n(x + y_n)) \to_{a.s.} \psi_0(x), \quad \text{for each } x \in \text{int}(\text{dom}(\psi_0)).$$

Since $\tilde{\psi}_n$ is concave for $\tilde{g}_n \in \mathcal{LC}$, Theorem 10.8 of Rockafellar (1970) entails that the above pointwise convergence translates to uniform convergence on all compact sets inside $\text{int}(\text{dom}(\psi_0))$, which leads to

$$\sup_{x \in K} |\tilde{\psi}_n(x + y_n) - \psi_0(x)| \to_{a.s.} 0,$$

proving part B of Condition 1. Since $\tilde{\psi}_n$ is concave, Part C follows directly from Part B by Theorem 25.7 of Rockafellar (1970). Thus we have established Condition 1 for $\tilde{g}_n = \widehat{g}_{\bar{\theta}_n}, \widehat{h}_n(\bar{\theta}_n \pm \cdot), \tilde{g}_n^{geo,sym},$ and $\widehat{h}_n^{sm}(\bar{\theta}_n \pm \cdot)$.

Now we verify Condition 1 for $\tilde{g}_n^{sym,sm}$. Part A of Condition 1 can be verified noting (12) implies

$$2 \sup_{x \in \mathbb{R}} |\tilde{g}_n^{sym,sm}(x + y_n) - g_0(x)|$$
$$\leq \sup_{x \in \mathbb{R}} |\widehat{h}_n^{sm}(\bar{\theta}_n + x + y_n) - g_0(x)| + \sup_{x \in \mathbb{R}} |\widehat{h}_n^{sm}(\bar{\theta}_n - x + y_n) - g_0(x)|,$$

which converges to zero almost surely because, as we have already shown, the log-concave density $\widehat{h}_n^{sm}$ satisfies Condition 1(A).

To prove part B, we observe that

$$\sup_{x \in K} \left| \tilde{\psi}_n^{sym,sm}(x + y_n) - \psi_0(x) \right| \leq \frac{\sup_{x \in K} |\tilde{g}_n^{sym,sm}(x + y_n) - g_0(x)|}{\min\left(\inf_{x \in K} \tilde{g}_n^{sym,sm}(x + y_n), \inf_{x \in K} g_0(x) \right)},$$

whose numerator converges to zero almost surely by part A of Condition 1. Thus, to verify part B of Condition 1 for $\tilde{g}_n^{sym,sm}$, we only need to show that the denominator of the term on the right hand side of last display is bounded away from zero. To this end, notice that

$$\inf_{x \in K} \tilde{g}_n^{sym,sm}(x + y_n) \geq \inf_{x \in K} \min(\widehat{h}_n^{sm}(\bar{\theta}_n + x + y_n),$$

$$\widehat{h}_n^{sm}(\bar{\theta}_n - x - y_n)) \xrightarrow[a.s.]{(a)} \inf_{x \in K} g_0(x),$$

where (a) follows because we just showed that $\widehat{h}_n^{sm}(\bar{\theta}_n \pm \cdot)$ satisfy Condition 1. Now $\inf_{x \in K} g_0(x) > 0$ because K is a subset of $\mathrm{int}(\mathrm{dom}(\psi_0))$. Thus we have verified part B of Condition 1 for $\tilde{g}_n^{sym,sm}$.

Next note that $\widehat{h}_n^{sm}$ is a smooth function, and it is also positive on $\mathbb{R}$. Therefore $\tilde{\psi}_n^{sym,sm}$ and $\tilde{\phi}_n^{sm}$ are differentiable on $\mathbb{R}$. Therefore, for any $x \in \mathbb{R}$,

$$(\tilde{\psi}_n^{sym,sm})'(x) = \varrho_n(x)\left((\tilde{\phi}_n^{sm})'(\bar{\theta}_n + x) \right) - (1 - \varrho_n(x))\left((\tilde{\phi}_n^{sm})'(\bar{\theta}_n - x) \right), \quad (54)$$

where $\varrho_n(x) = \widehat{h}_n^{sm}(\bar{\theta}_n + x)/2\tilde{g}_n^{sym,sm}(x) < 1$. Thus

$$|(\tilde{\psi}_n^{sym,sm})'(x + y_n) - \psi_0'(x)| \leq \varrho_n(x + y_n)|(\tilde{\phi}_n^{sm})'(\bar{\theta}_n + x + y_n) - \psi_0'(x)|$$

$$+ (1 - \varrho_n(x + y_n))|(\tilde{\phi}_n^{sm})'(\bar{\theta}_n - x - y_n) - \psi_0'(-x)|.$$

Since ϱ_n is uniformly bounded by one, Condition 1(C) applied on the concave function $\tilde{\phi}_n^{sm}(\bar{\theta}_n \pm \cdot)$ completes the verification of part C for $\tilde{\psi}_n^{sym,sm}$. $\qquad \square$

C.1. Auxiliary lemmas for the proof of proposition 1

Proof of Lemma C.1. First we show that it suffices to prove the current lemma when $\bar{\theta}_n \to_{a.s.} \theta_0$. Suppose we are able to prove $\|\tilde{g}_n - g_0\|_1$ is strongly consistent when $\bar{\theta}_n \to_{a.s.} 0$. Since $\bar{\theta}_n \to_p 0$, Fact 6 implies 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 g_0 approaches zero almost surely. In that case, Fact 7 implies that $\|\tilde{g}_n - g_0\|_1$ converges in probability to zero. Therefore, in what follows, we assume that $\bar{\theta}_n \to_{a.s.} \theta_0$.

We begin with the case of $\widehat{h}_n^{sm}$. Theorem 1 of Chen and Samworth (2013) implies that when f_0 has finite second central moment, we have

$$\int_{-\infty}^{\infty} |\widehat{h}_n^{sm}(x) - f_0(x)| dx \to_{a.s.} 0. \quad (55)$$

That f_0 has second central moment is immediate by Fact 1. Note that

$$\|\widehat{h}_n^{sm}(\bar{\theta}_n + \cdot) - g_0\|_1 \leq \|\widehat{h}_n^{sm} - f_0\|_1 + \|f_0(\bar{\theta}_n + \cdot) - g_0\|_1,$$

whose first term converges to zero almost surely by (55), and the second term

$$\|f_0(\bar{\theta}_n + \cdot) - g_0\|_1 = \|g_0(-\tilde{\delta}_n + \cdot) - g_0\|_1 \overset{(a)}{\leq} \sqrt{2}H(g_0(-\tilde{\delta}_n + \cdot), g_0) \overset{(b)}{\lesssim} \tilde{\delta}_n \to_{a.s.} 0$$

where (a) and (b) follow from Fact 5 and Fact 15, respectively. Thus we have established that $\|\widehat{h}_n^{sm}(\bar{\theta}_n + \cdot) - g_0\|_1 \to_{a.s.} 0$. Since g_0 is symmetric about zero, $\|\widehat{h}_n^{sm}(\bar{\theta}_n - \cdot) - g_0\|_1 \to_{a.s.} 0$ follows. Because $\|\widehat{h}_n - f_0\|_1 \to_{a.s.} 0$ by Theorem 4 of Cule and Samworth (2010), the proof of $\|\widehat{h}_n(\bar{\theta}_n \pm \cdot) - g_0\|_1 \to_{a.s.} 0$ follows in the same way.

The L_1 consistency of $\tilde{g}_n^{sym,sm}$ also follows noting (12) implies

$$2\|\tilde{g}_n^{sym,sm} - g_0\|_1 \leq \|\widehat{h}_n^{sm}(\bar{\theta}_n + \cdot) - g_0\|_1 + \|\widehat{h}_n^{sm}(\bar{\theta}_n - \cdot) - g_0\|_1 \to_{a.s.} 0. \quad (56)$$

Next, we consider the geometric mean estimator $\widehat{f}_n^{geo,sym}$. We have already established

$$\int_{-\infty}^{\infty} |\widehat{h}_n(\bar{\theta}_n \pm x) - g_0(x)|dx \to_{a.s.} 0, \quad (57)$$

which entails that the distribution functions of $\widehat{h}_n(\bar{\theta}_n \pm \cdot)$ converge weakly to G_0. The above, combined with Proposition 2(b) of Cule and Samworth (2010) shows that (57) leads to almost sure convergence of $\widehat{h}_n(\bar{\theta}_n \pm x)$ to $g_0(x)$ almost everywhere on $\mathbb{R}$ with respect to the Lebesgue measure. As a consequence, it follows that

$$\widehat{f}_n^{geo,sym}(x)C_n^{geo} \to_{a.s.} \sqrt{g_0(x)g_0(-x)} = g_0(x) \quad a.e. \quad x.$$

Recall from (9) that

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

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 almost everywhere to g_0 almost surely. Therefore, $\tilde{G}_n^{sym,sm}$ converges weakly to G_0 almost surely. The desired strong L_1 consistency then follows from Proposition 2(c) of Cule and Samworth (2010).

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 (2019). According to this theory, $\widehat{G}_{\bar{\theta}_n}$ can be interpreted as the the projection (w.r.t. Kullback-Leibler divergence) of $\mathbb{F}_{n,X-\bar{\theta}_n}$, the empirical distribution function of the $X_i - \bar{\theta}_n$'s, onto the space of the distribution functions with density in $\mathcal{SLC}_0$. This projection operator has some continuity properties. In particular, if we can show that

$$d_W(\mathbb{F}_{n,X-\bar{\theta}_n}, G_0) \to_{a.s.} 0, \quad (58)$$

the desired L_1 consistency $\|\widehat{g}_{\bar{\theta}_n} - g_0\|_1 \to_{a.s.} 0$ follows from Proposition 6 of Xu and Samworth (2019) provided G_0 is non-degenerate and it has first finite moment. The non-degeneracy is trivial and the existence of first moment follows from Fact 1. Hence, it is enough to prove (58) holds, for which, by Theorem 6.9 of Villani (2009), it suffices to show

$$\int_{-\infty}^{\infty} |x|d\mathbb{F}_{n,X-\bar{\theta}_n}(x) \to_{a.s.} \int_{-\infty}^{\infty} |x|dG_0(x), \quad (59)$$

and that $\mathbb{F}_{n,X-\bar{\theta}_n}$ converges to G_0 weakly with probability one. Since $\bar{\theta}_n$ is strongly consistent for θ_0, and

$$\int_{-\infty}^{\infty} |x| d\mathbb{F}_{n,X-\bar{\theta}_n}(x) = \int_{-\infty}^{\infty} |x - \bar{\theta}_n| d\mathbb{F}_n(x),$$

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 $|x - \bar{\theta}_n| \le |x - \theta_0| + 1$ with probability one for all sufficiently large n, where the latter is integrable with respect to f_0. 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 (59).

Our next step is to prove the weak convergence of $\mathbb{F}_{n,X-\bar{\theta}_n}$ to G_0. To this end, we note that

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

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 establishes (58), which proves the strong L_1 consistency of $\widehat{g}_{\bar{\theta}_n}$, thus finishing the proof of the current lemma. $\qquad\square$

C.1.1. *Lemmas on Hellinger error of $\widehat{g}_{\bar{\theta}_n}$ and $\tilde{g}_n^{geo,sym}$:*

Lemma C.2. *Suppose $f_0 = g_0(\cdot - \theta_0)$ where $g_0 \in \mathcal{SLC}_0$ and $\bar{\theta}_n - \theta_0 = O_p(n^{-1/2})$. Then $H(\widehat{g}_{\bar{\theta}_n}, g_0) = O_p(n^{-1/4})$.*

Proof of Lemma C.2. From Theorem 4.1 of Doss and Wellner (2019) it follows that $H(\widehat{g}_{\theta_0}, g_0) = O_p(n^{-2/5})$. The result will therefore follow by triangle inequality if we can show that $H(\widehat{g}_{\theta_0}, \widehat{g}_{\bar{\theta}_n}) = O_p(n^{-1/4})$. To that end, for any function $\phi : \mathbb{R} \mapsto \mathbb{R}$, and distribution function G, we define the functional $\Phi : (\phi, G) \mapsto \mathbb{R}$ by

$$\Phi(\phi, G) = \int_{-\infty}^{\infty} \phi(x) dG(x) - \int_{-\infty}^{\infty} e^{\phi(x)} dx. \tag{60}$$

Recall that we defined $\widehat{\psi}_\theta$ to be $\log \widehat{g}_\theta$ for any $\theta > 0$. Denoting $\mathbb{F}_{n,Y}$ to be the empirical distribution function of random variables $Y_1, \ldots, Y_n$, we observe that

for any $\theta > 0$, $\widehat{\psi}_\theta$ writes as (see (2.4) of Doss and Wellner, 2019)

$$\widehat{\psi}_\theta = \arg\max_{\phi \in \mathcal{SC}_0} \Phi(\phi, \mathbb{F}_{n,X-\theta}) = \arg\max_{\phi \in \mathcal{SC}_0} \Phi(\phi, \mathbb{F}_{n,Z+\theta_0-\theta}),$$

where $Z = X - \theta_0$. Let us denote $\delta = \theta_0 - \theta$. Using Lemma C.4 we obtain that

$$\arg\max_{\phi \in \mathcal{SC}_0} \Phi(\phi, \mathbb{F}_{n,Z+\delta}) = \arg\max_{\phi \in \mathcal{C}} \Phi(\phi, \mathbb{F}_{n,Z+\delta}^{sym}), \quad \text{where}$$

$$\mathbb{F}_{n,Z+\delta}^{sym}(x) = \frac{\mathbb{F}_{n,Z+\delta}(x) + 1 - \mathbb{F}_{n,Z+\delta}(-x)}{2} \overset{(a)}{=} \frac{\mathbb{F}_{n,Z}(x-\delta) + 1 - \mathbb{F}_{n,Z}(-x-\delta)}{2}$$

is the symmetrized version of $\mathbb{F}_{n,Z+\delta}$. Here (a) follows because $\mathbb{F}_{n,Z+\delta}(x)$ equals $\mathbb{F}_{n,Z}(x-\delta)$. In particular, the choice $\theta = \theta_0$ yields $\delta = 0$, which leads to $\widehat{\psi}_{\theta_0} = \arg\max_{\phi \in \mathcal{C}} \Phi(\phi, \mathbb{F}_{n,Z}^{sym})$, where $\mathbb{F}_{n,Z}^{sym}(x) = (\mathbb{F}_{n,Z}(x) + 1 - \mathbb{F}_{n,Z}(-x))/2$. When $\theta = \bar{\theta}_n$, on the other hand, $\delta = \tilde{\delta}_n$, which yields

$$\widehat{\psi}_{\bar{\theta}_n} = \arg\max_{\phi \in \mathcal{C}} \Phi(\phi, \mathbb{F}_{n,Z+\tilde{\delta}_n}^{sym}), \quad \mathbb{F}_{n,Z+\tilde{\delta}_n}^{sym}(x) = \frac{\mathbb{F}_{n,Z}(x-\tilde{\delta}_n) + 1 - \mathbb{F}_{n,Z}(-x-\tilde{\delta}_n)}{2}.$$

If we can show that $\mathbb{F}_{n,Z}^{sym}(x)$ and $\mathbb{F}_{n,Z+\tilde{\delta}_n}^{sym}$ are non-degenerate with finite first moment, then Theorem 1 of Barber and Samworth (2020) yields

$$E[H(\widehat{g}_{\bar{\theta}_n}, \widehat{g}_{\theta_0})] \lesssim \left(\frac{d_W(\mathbb{F}_{n,Z}^{sym}, \mathbb{F}_{n,Z+\tilde{\delta}_n}^{sym})}{\epsilon_{\mathbb{F}_{n,Z}^{sym}}} \right)^{1/2}, \tag{61}$$

where for any distribution function F, ϵ_F is defined by

$$\epsilon_F = E_F[|Y - E_F[Y]|].$$

Here E_F is the expectation with respect to F. Since G_0 is non-degenerate, $\mathbb{F}_{n,Z}$ is non-degenerate with probability one. Therefore both $\mathbb{F}_{n,Z}^{sym}(x)$ and $\mathbb{F}_{n,Z+\tilde{\delta}_n}^{sym}$ are non-degenerate with probabilty one. Also, because $\mathbb{F}_{n,Z}$ has finite first moment for all $n \geq 1$, both $\mathbb{F}_{n,Z}^{sym}(x)$ and $\mathbb{F}_{n,Z+\tilde{\delta}_n}^{sym}$ have finite first moment for all $n \geq 1$. Therefore (61) holds.

Next, we show that $\epsilon_{\mathbb{F}_{n,Z}^{sym}}$ is bounded away from zero almost surely. To that end, we first prove the side result that $d_W(\mathbb{F}_{n,Z}^{sym}, G_0) \to_{a.s.} 0$. By (17),

$$d_W(\mathbb{F}_{n,Z}^{sym}, G_0) = \int_{-\infty}^{\infty} |\mathbb{F}_{n,Z}^{sym}(x) - G_0(x)| dx$$

$$= \int_{-\infty}^{\infty} \left| \frac{\mathbb{F}_{n,Z}(x) + 1 - \mathbb{F}_{n,Z}(-x)}{2} - G_0(x) \right|,$$

which, due to the symmetry of g_0 about the origin, equals

$$\int_{-\infty}^{\infty} \left| \frac{\mathbb{F}_{n,Z}(x) + 1 - \mathbb{F}_{n,Z}(-x)}{2} - \frac{G_0(x) + 1 - G_0(-x)}{2} \right|$$

$$\leq \frac{1}{2} \left(\int_{-\infty}^{\infty} |\mathbb{F}_{n,Z}(x) - G_0(x)| \, dx + \int_{-\infty}^{\infty} |\mathbb{F}_{n,Z}(-x) - G_0(-x)| \, dx \right)$$

$$\leq \int_{-\infty}^{\infty} |\mathbb{F}_{n,Z}(x) - G_0(x)| \, dx,$$

which equals $d_W(\mathbb{F}_{n,Z}, G_0)$. The latter converges to zero almost surely by Vara-darajan's Theorem (Dudley, 2003, Theorem 11.4.1) and the strong law of large numbers. Therefore, $d_W(\mathbb{F}^{sym}_{n,Z}, G_0) \to_{a.s.} 0$ follows.

Proposition 1 of Barber and Samworth (2020) implies that if F and F' are distribution functions with finite first moment, then $\epsilon_F > 0$, and $|\epsilon_F - \epsilon_{F'}|$ is bounded by $2d_W(F, F')$. Now G_0 being log-concave, has finite first moment. Therefore we have $\epsilon_{G_0} > 0$. Also since G_0 and $\mathbb{F}^{sym}_{n,Z}$ are non-degenerate, it follows that

$$\left| \epsilon_{\mathbb{F}^{sym}_{n,Z}} - \epsilon_{G_0} \right| \le d_W(\mathbb{F}^{sym}_{n,Z}, G_0),$$

which implies $\epsilon_{\mathbb{F}^{sym}_{n,Z}} > \epsilon_{G_0} - 2d_W(\mathbb{F}^{sym}_{n,Z}, G_0)$. We have just shown $d_W(\mathbb{F}^{sym}_{n,Z}, G_0)$ converges to zero almost surely. Therefore $\epsilon_{\mathbb{F}^{sym}_{n,Z}} > \epsilon_{G_0}/2$ for sufficiently large n almost surely.

If we can show that $d_W(\mathbb{F}^{sym}_{n,Z}, \mathbb{F}^{sym}_{n,Z+\tilde{\delta}_n}) = O_p(\tilde{\delta}_n)$, the proof of lemma C.2 follows from (61) because $\tilde{\delta}_n = O_p(n^{-1/2})$. To that end, we use an alternative representation of d_W which is due to the Kantorovich-Rubinstein duality theo-rem (cf. Theorem 2.5 Bobkov and Ledoux, 2014). For distribution functions F_1 and F_2 with finite first moment, it holds that

$$d_W(F_1, F_2) = \sup_{h \in \mathrm{Lip}_1} \left| \int_{-\infty}^{\infty} h(x) d(F_1 - F_2) \right|$$

where Lip_1 is the set of all real-valued functions $h : \mathbb{R} \mapsto \mathbb{R}$ with Lipschitz constant one. Therefore,

$$
\begin{aligned}
d_W(\mathbb{F}^{sym}_{n,Z}, \mathbb{F}^{sym}_{n,Z+\tilde{\delta}_n}) &= \sup_{h \in \mathrm{Lip}_1} \frac{1}{2} \left| \int_{-\infty}^{\infty} h(x) d(\mathbb{F}_{n,Z}(x - \tilde{\delta}_n) + \mathbb{F}_{n,Z}(-x - \tilde{\delta}_n)) \right. \\
&\quad \left. - \int_{-\infty}^{\infty} h(x) d(\mathbb{F}_{n,Z}(x) + \mathbb{G}_n(-x)) \right| \\
&= \sup_{h \in \mathrm{Lip}_1} \frac{1}{2} \left| \int_{-\infty}^{\infty} \left(h(x + \tilde{\delta}_n) - h(x) \right) d(\mathbb{F}_{n,Z}(x) + \mathbb{G}_n(-x)) \right| \\
&\le \sup_{h \in \mathrm{Lip}_1} \frac{1}{2} \int_{-\infty}^{\infty} |h(x + \tilde{\delta}_n) - h(x)| d(\mathbb{F}_{n,Z}(x) + \mathbb{F}_{n,Z}(-x)) \\
&\overset{(a)}{\le} |\tilde{\delta}_n| \int_{-\infty}^{\infty} d(\mathbb{F}_{n,Z}(x) + \mathbb{F}_{n,Z}(-x)) dx/2,
\end{aligned}
$$

which equals $|\tilde{\delta}_n|$. Here (a) uses the fact that h is Lipschitz with Lipschitz constant one. Therefore, the proof follows. $\square$

Lemma C.3. *Suppose $f_0 \in \mathcal{P}_0$ and $\bar{\theta}_n - \theta_0 = O_p(n^{-1/2})$. Then the geometric mean estimator $\tilde{g}_n(x) = \sqrt{\widehat{h}_n(\bar{\theta}_n + x)\widehat{h}_n(\bar{\theta}_n - x)}/C^{geo}_n$ satisfies $H(\tilde{g}_n, g_0) = O_p(n^{-2/5})$.*

Proof of Lemma C.3. We first decompose $H(\tilde{g}_n, g_0)^2$ as follows:

$$H(\tilde{g}_n, g_0)^2 \leq \int_{-\infty}^{\infty} \left(\left(\sqrt{\widehat{h}_n(\bar{\theta}_n + x)\widehat{h}_n(\bar{\theta}_n - x)}/C_n^{geo} \right)^{1/2} - \sqrt{g_0(x)} \right)^2 dx$$

$$\leq 2(C_n^{geo})^{-1} \underbrace{\int_{-\infty}^{\infty} \left(\left(\widehat{h}_n(\bar{\theta}_n + x)\widehat{h}_n(\bar{\theta}_n - x) \right)^{1/4} - \sqrt{g_0(x)} \right)^2 dx}_{T_1}$$

$$+ 2(C_n^{geo})^{-1} \underbrace{(\sqrt{C_n^{geo}} - 1)^2}_{T_2}. \tag{62}$$

We focus on T_1 first. Note that

$$T_1 = \int_{-\infty}^{\infty} \left(\left(\widehat{h}_n(\bar{\theta}_n + x)\widehat{h}_n(\bar{\theta}_n - x) \right)^{1/4} - \sqrt{g_0(x)} \right)^2 dx$$

$$= \int_{-\infty}^{\infty} \left(\left(\frac{\widehat{h}_n(\bar{\theta}_n + x)}{\widehat{h}_n(\bar{\theta}_n - x)} \right)^{1/4} \sqrt{\widehat{h}_n(\bar{\theta}_n - x)} - \sqrt{g_0(x)} \right)^2 \tag{63}$$

$$\lesssim \underbrace{H(\widehat{h}_n, f_0(\cdot + \tilde{\delta}_n))^2}_{T_{11}} + \underbrace{\int_{-\infty}^{\infty} \left(\left(\frac{\widehat{h}_n(\bar{\theta}_n + x)}{\widehat{h}_n(\bar{\theta}_n - x)} \right)^{1/4} - 1 \right)^2 \widehat{h}_n(\bar{\theta}_n - x) dx}_{T_{12}}.$$

T_{11} is bounded by $2H(\widehat{h}_n, f_0)^2 + 2H(f_0, f_0(\cdot + \tilde{\delta}_n))^2$. Thus $T_{11} = O_p(n^{-4/5})$ follows noting (a) $H(\widehat{h}_n, f_0) = O_p(n^{-2/5})$ by Theorem 3.2 of Doss and Wellner (2016), and (b) $H(f_0, f_0(\cdot + \tilde{\delta}_n)) = O_p(\tilde{\delta}_n)$ by Fact 15. Since $(x-1)^2 \leq (x^2-1)^2$ for $x > 0$, the term T_{12} can be bounded by

$$\int_{-\infty}^{\infty} \left(\sqrt{\widehat{h}_n(\bar{\theta}_n + x)} - \sqrt{\widehat{h}_n(\bar{\theta}_n - x)} \right)^2 dx = \int_{-\infty}^{\infty} \left(\sqrt{\widehat{h}_n(x)} - \sqrt{\widehat{h}_n(2\bar{\theta}_n - x)} \right)^2 dx.$$

Since $f_0(2\theta_0 - x) = f_0(x)$, we can further bound T_{12} by a constant multiple of

$$\int_{-\infty}^{\infty} \left(\sqrt{\widehat{h}_n(x)} - \sqrt{f_0(x)} \right)^2 dx + \int_{-\infty}^{\infty} \left(\sqrt{\widehat{h}_n(2\bar{\theta}_n - x)} - \sqrt{f_0(2\bar{\theta}_n - x)} \right)^2 dx$$

$$+ \int_{-\infty}^{\infty} \left(\sqrt{f_0(2\bar{\theta}_n - x)} - \sqrt{f_0(2\theta_0 - x)} \right)^2 dx$$

$$\leq 4H(\widehat{h}_n, f_0)^2 + \int_{-\infty}^{\infty} \left(\sqrt{f_0(x - 2\tilde{\delta}_n)} - \sqrt{f_0(x)} \right)^2 dx$$

$$= 8H(\widehat{h}_n, f_0)^2 + 4H(f_0(\cdot + 2\tilde{\delta}_n), f_0)^2,$$

whose first term is $O_p(n^{-4/5})$, and the second term, by Fact 15, is of order $O_p(\tilde{\delta}_n^2)$. Thus similar to T_{11}, T_{12} is $O_p(n^{-4/5})$ as well. Therefore from (63) it follows that $T_1 = O_p(n^{-4/5})$.

Using the fact that $(x-1)^2 \leq (x^2-1)^2$ for non-negative x, we obtain that

$T_2 \leq (C_n^{geo} - 1)^2$, which equals

$$\left(\int_{-\infty}^{\infty} \sqrt{\widehat{h}_n(\bar{\theta}_n + x)\widehat{h}_n(\bar{\theta}_n - x)}dx - 1 \right)^2$$

$$= \left(\int_{-\infty}^{\infty} \sqrt{\widehat{h}_n(\bar{\theta}_n + x)} \left(\sqrt{\widehat{h}_n(\bar{\theta}_n - x)} - \sqrt{\widehat{h}_n(\bar{\theta}_n + x)} \right) dx \right)^2.$$

The Cauchy-Schwarz inequality implies that the term on the right hand side of the above display is bounded by

$$\int_{-\infty}^{\infty} \left(\sqrt{\widehat{h}_n(\bar{\theta}_n - x)} - \sqrt{\widehat{h}_n(\bar{\theta}_n + x)} \right)^2 dx$$

$$\lesssim \int_{-\infty}^{\infty} \left(\sqrt{\widehat{h}_n(\bar{\theta}_n - x)} - \sqrt{f_0(\bar{\theta}_n - x)} \right)^2 dx$$

$$+ \int_{-\infty}^{\infty} \left(\sqrt{\widehat{h}_n(\bar{\theta}_n + x)} - \sqrt{f_0(\bar{\theta}_n + x)} \right)^2 dx$$

$$+ \int_{-\infty}^{\infty} \left(\sqrt{f_0(\bar{\theta}_n - x)} - \sqrt{f_0(\bar{\theta}_n + x)} \right)^2 dx.$$

Clearly, the first two terms equal $4H(\widehat{h}_n, f_0)^2$, which is $O_p(n^{-4/5})$. Since f_0 is symmetric about θ_0, we can show that $f_0(\bar{\theta}_n - x) = f_0(2\tilde{\delta}_n + \bar{\theta}_n + x)$, which implies the third term equals $H(f_0, f_0(\cdot + 2\tilde{\delta}_n))^2$, which, by Fact 15, is of order $O_p(\tilde{\delta}_n^2)$. Thus we have established that T_2 is $O_p(n^{-4/5})$ as well, which also implies that $C_n^{geo} \to_p 1$. Therefore, by Slutskey's Theorem and (62), the proof follows. □

Lemma C.4. *Suppose F is non-degenerate and F has finite first moment. Define*

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

Then it follows that $\arg\max_{\phi \in SC_\theta} \Phi(\phi, F) = \arg\max_{\phi \in C} \Phi(\phi, F_\theta^{sym})$ where Φ is as defined in (60).

Proof of Lemma C.4. First we will show that

$$\arg\max_{\phi \in SC_\theta} \Phi(\phi, F) = \arg\max_{\phi \in SC_\theta} \Phi(\phi, F_\theta^{sym}). \tag{64}$$

Recall the definition of Ψ from (13). For any distribution function F and $\psi \in SC_0$, the following holds:

$$\Psi(0, \psi, F) = \int_{-\infty}^{0} \psi(x)dF(x) + \int_{0}^{\infty} \psi(x)dF(x) - \int_{-\infty}^{\infty} e^{\psi(x)}dx$$

$$= -\int_{0}^{\infty} \psi(-x)dF(-x) + \int_{0}^{\infty} \psi(x)dF(x) - \int_{-\infty}^{\infty} e^{\psi(x)}dx$$

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

where the last step uses $\psi(x) = \psi(-x)$. By symmetry, 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, $\quad \Psi(0, \psi, F) = 2^{-1} \int_{-\infty}^\infty \psi(x) d(F(x) - F(-x)) - \int_{-\infty}^\infty e^{\psi(x)} dx.$

Equation 13 implies $\Psi(\theta, \psi, F) = \Psi(\theta, \psi, F(\cdot + \theta))$. Therefore,

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

$$\overset{(a)}{=} 2^{-1} \int_{-\infty}^\infty \psi(z - \theta) d(F(z) - F(2\theta - z)) - \int_{-\infty}^\infty e^{\psi(z)} dz$$

$$= \int_{-\infty}^\infty \psi(z - \theta) dF_\theta^{sym}(z) - \int_{-\infty}^\infty e^{\psi(z)} dz = \Psi(\theta, \psi, F_\theta^{sym}) \quad (65)$$

where (a) follows substituting $\theta + x$ by z. Suppose $\psi \in \mathcal{SC}_0$ and $\phi = \psi(\cdot - \theta)$. Equation 60 implies that for any $\phi \in \mathcal{SC}_\theta$, $\Phi(\phi, F) = \Psi(\theta, \psi, F)$, where $\psi = \phi(\cdot + \theta)$. This, in conjunction with (65), yields that $\Phi(\phi, F) = \Phi(\phi, F_\theta^{sym})$ for any $\phi \in \mathcal{SC}_\theta$. Therefore, (64) follows.

Proposition 4(iii) of Xu and Samworth (2019) entails that $\arg\max_{\phi \in \mathcal{SC}_\theta} \Phi(\phi, F)$ exists and is unique for a degenerate F with finite first moment. Under similar conditions on F, $\arg\max_{\phi \in \mathcal{C}} \Phi(\phi, F_\theta^{sym})$ also exists and it is unique by Theorem 2.7 of Dümbgen *et al.* (2011). Therefore, it suffices to prove $\arg\max_{\phi \in \mathcal{C}} \Phi(\phi, F_\theta^{sym})$ is in $\mathcal{SC}_\theta$ because the latter implies

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

which, in conjuction with (64), completes the proof of the current lemma.

Without loss of generality, we will assume $\theta = 0$. In that case, $F_0^{sym}(x) = (F(x) + 1 - F(-x))/2$, which implies

$$dF_0^{sym}(x) = (dF(x) - dF(-x))/2 = -dF_0^{sym}(-x). \quad (66)$$

For any concave function $\phi \in \mathcal{C}$, note that $\Phi(\phi, F_0^{sym})$ can be written as

$$\int_{-\infty}^0 \phi(x) dF_0^{sym}(x) + \int_0^\infty \phi(x) dF_0^{sym}(x) - \int_{-\infty}^\infty e^{\phi(x)} dx$$

$$= -\int_0^\infty \phi(-x) dF_0^{sym}(-x) + \int_0^\infty \phi(x) dF_0^{sym}(x) - \int_{-\infty}^\infty e^{\phi(x)} dx$$

$$\overset{(a)}{=} \int_0^\infty \phi(-x) dF_0^{sym}(x) + \int_0^\infty \phi(x) dF_0^{sym}(x) - \int_{-\infty}^\infty e^{\phi(x)} dx$$

$$\overset{(b)}{=} \int_{-\infty}^\infty \frac{\phi(-x) + \phi(x)}{2} dF_0^{sym}(x) - \int_{-\infty}^\infty e^{\phi(x)} dx,$$

where (a) uses (66), and (b) follows since

$$\int_0^\infty \frac{\phi(-x) + \phi(x)}{2} dF_0^{sym}(x) = \int_{-\infty}^0 \frac{\phi(-x) + \phi(x)}{2} dF_0^{sym}(x)$$

by symmetry. Moreover, since exponential function is convex, we obtain

$$\Phi(\phi, F_0^{sym}) \le \int_{-\infty}^{\infty} \frac{\phi(x) + \phi(-x)}{2} dF_0^{sym}(x) - \int_{-\infty}^{\infty} e^{(\phi(x)+\phi(-x))/2} dx,$$

which proves that a $\phi \in \mathcal{SC}_0$ maximizes $\Phi(\phi, F_0^{sym})$ over $\mathcal{C}$, as speculated. Therefore, the proof follows. $\quad\square$

Appendix D: Proof of Theorem 2

Similar to Theorem 1, we can argue that it suffices to prove Theorem 2 for the case when η_n is $Cn^{-2p/5}$. For the rest of the proof, we will denote $\xi_n = (\tilde{G}_n^{sym,sm})^{-1}(1 - \eta_n)$. First of all note that $\widehat{h}_n^{sm}(\bar{\theta}_n \pm \cdot)$ and $\tilde{g}_n^{sym,sm}$ satisfy Condition 1 by Proposition 2. Lemma D.1 in Appendix D.1 implies that these densities also satisfy Condition 2 with $p = 1/5$. Since the proof of Theorem 2 closely follows the proof of Theorem 1, we will only highlight the differences. Following the arguments in Theorem 1, we can represent $-(\hat{\theta}_n - \bar{\theta}_n)$ as the sum of the three terms T_{1n}, T_{2n}, and T_{5n}, where

$$T_{1n} = \int_{-\xi_n}^{\xi_n} \frac{(\tilde{\psi}_n^{sym,sm})'(z) - \psi_0'(z - \tilde{\delta}_n)}{\widehat{\mathcal{I}}_n(\eta_n)} d(\mathbb{F}_n(z + \bar{\theta}_n) - F_0(z + \bar{\theta}_n)),$$

$$T_{2n} = \int_{-\xi_n}^{\xi_n} \frac{(\tilde{\psi}_n^{sym,sm})'(z)}{\widehat{\mathcal{I}}_n(\eta_n)} \left(f_0(z + \bar{\theta}_n) - g_0(z) \right) dz,$$

and T_{5n} is as in (19). The treatment of T_{5n} in this case will be identical to that in Theorem 1. Hence it suffices to redo step one and step two of Theorem 1 only in context of $\tilde{g}_n^{sym,sm}$.

Step one: showing $T_{1n} = o_p(1)$***:***

The main difference in the analysis of T_{1n} between Theorem 1 and here stems from the fact that $(\tilde{\psi}_n^{sym,sm})'$ is no longer guaranteed to be monotone since $\tilde{g}_n^{sym,sm}$ is not log-concave. So one needs to be more careful before applying the Donsker theorem to control the T_{1n} term here. By construction, $\widehat{h}_n^{sm}$ and $\tilde{g}_n^{sym,sm}$ are positive on the entire real line, and differentiable everywhere. Using (54), we obtain the formula

$$(\tilde{\psi}_n^{sym,sm})'(x) = \varrho_n(x) \left((\tilde{\phi}_n^{sm})'(\bar{\theta}_n + x) \right) - (1 - \varrho_n(x)) \left((\tilde{\phi}_n^{sm})'(\bar{\theta}_n - x) \right),$$

where $\varrho_n(x) = \widehat{h}_n^{sm}(\bar{\theta}_n + x)/2\tilde{g}_n^{sym,sm}(x)$. Note that $(\tilde{\phi}_n^{sm})'(\bar{\theta}_n \pm \cdot)$ is non-increasing because $\widehat{h}_n^{sm}$ is log-concave. On the other hand, because $\widehat{h}_n^{sm}$ is smooth, and $\widehat{h}_n^{sm} > 0$ on $\mathbb{R}$, ϱ_n is differentiable with derivative

$$\varrho_n'(x) = \frac{(\widehat{h}_n^{sm})'(\bar{\theta}_n - x)\widehat{h}_n^{sm}(\bar{\theta}_n + x) + \widehat{h}_n^{sm}(\bar{\theta}_n - x)(\widehat{h}_n^{sm})'(\bar{\theta}_n + x)}{(\widehat{h}_n^{sm}(\bar{\theta}_n - x) + \widehat{h}_n^{sm}(\bar{\theta}_n + x))^2},$$

which is less than $|(\tilde{\phi}_n^{sm})'(\bar{\theta}_n - x)| + |(\tilde{\phi}_n^{sm})'(\bar{\theta}_n + x)|$ in absolute value. However, Lemma D.6 implies that

$$\sup_{x \in [-\xi_n, \xi_n]} \left(|(\tilde{\phi}_n^{sm})'(\bar{\theta}_n - x)| + |(\tilde{\phi}_n^{sm})'(\bar{\theta}_n + x)| \right) = O_p(n^{p/5}). \qquad (67)$$

Therefore, on $[-\xi_n, \xi_n]$, the derivative of ϱ_n is uniformly bounded by an $O_p(n^{p/5})$ term. The same bound can be proved for $1 - \varrho_n$ as well. Noting ϱ_n is a fraction, we also deduce that $\|\varrho_n\|_\infty$ and $\|1 - \varrho_n\|_\infty$ are bounded by one. For a convex set $\mathcal{X} \subset \mathbb{R}$ and a number $M > 0$, define the class of functions $\mathcal{D}_{n,M}(\mathcal{X})$ by

$$\mathcal{D}_{n,M}(\mathcal{X}) = \left\{ h : \mathcal{X} \mapsto \mathbb{R} \; \middle| \; h \text{ is differentiable on } \mathcal{X}, \; \sup_{x \in \mathcal{X}} |h(x)| + \sup_{x \in \mathcal{X}} |h'(x)| \leq M \right\}$$

As in the proof of Theorem 1, we let $M_n = Cn^{p/5}$ where $C > 0$ is a constant. Our earlier discussion on ϱ_n indicates that for sufficiently large $C > 0$, ϱ_n and $1 - \varrho_n$ restricted to $[-\xi_n, \xi_n]$ belongs to $\mathcal{D}_{n,M_n}([-\xi_n, \xi_n])$ with high probability as $n \to \infty$. Note also that (67) implies $(\tilde{\phi}_n^{sm})'(\bar{\theta}_n \pm \cdot) \in \mathcal{U}_n(M_n)$ with high probability for sufficiently large $C > 0$, where $\mathcal{U}_n(M_n)$ is as defined in (23). Therefore it is not hard to see that for sufficiently large $C > 0$, $(\tilde{\psi}_n^{sym,sm})' 1_{[-\xi_n, \xi_n]} \in \mathcal{U}_n^{sym}(M_n, -\xi_n, \xi_n)$ with high probability as $n \to \infty$, where for $-\infty \leq r_1 < r_2 \leq \infty$ and $C > 0$, the class $\mathcal{U}_n^{sym}(M_n, r_1, r_2)$ is defined by

$$\mathcal{U}_n^{sym}(M_n, r_1, r_2) = \left\{ h : \mathbb{R} \mapsto [-M_n, M_n] \; \middle| \; h(x) = q_1(x)f_1(x) + q_2(x)f_2(x) \text{ for} \right.$$

$$\left. x \in [-r, r], \text{ and } 0 \text{ o.w. where } q_1, q_2 \in \mathcal{D}_{n,M_n}([r_1, r_2]), \; f_1, f_2 \in \mathcal{U}_n(M_n) \right\}$$
$$(68)$$

It must be noted that in case of Theorem 1, we had $\tilde{\psi}_n' \in \mathcal{U}_n(M_n)$. Thus in Theorem 2, $\mathcal{U}_n(M_n)$ is replaced by $\mathcal{U}_n^{sym}(M_n, -\xi_n, \xi_n)$.

Corollary 2.7.2 of Van der Vaart and Wellner (1996) implies

$$\sup_Q \log N_{[\,]}(\epsilon, \mathcal{D}_{n,M_n}([r_1, r_2]), L_2(Q)) \lesssim \frac{(r_2 - r_1)M_n}{\epsilon},$$

where the supremum is over all probability measure Q on real line. On the other hand, (25) implies $\sup_Q \log N_{[\,]}(\epsilon, \mathcal{U}_n(M_n), L_2(Q)) \lesssim M_n/\epsilon$. Furthermore, (26) entails that the bracketing entropy of the function-class $\mathcal{F}_I$, consisting of indicator functions of the form $1_{[r_1, r_2]}$, is of the order ϵ^{-1}. Therefore we can show that

$$\sup_Q \log N_{[\,]}(\epsilon, \mathcal{U}_n^{sym}(M_n, r_1, r_2), L_2(Q)) \lesssim \frac{(r_2 - r_1)M_n}{\epsilon}. \qquad (69)$$

Next, we replace the class $\mathcal{H}_n(C)$ in the proof of Theorem 1 by the class

$$\mathcal{H}_n^{sym}(C) = \left\{ h : \mathbb{R} \mapsto \mathbb{R} \; \middle| \; h(x) = (u(x) - \phi_0'(x))1_{[r_1, r_2]}(x), \; u \in \mathcal{U}_n^{sym}(M_n, r_1, r_2), \right.$$

$$\|h\|_{P_0,2} \leq Cn^{-2p/5}(\log n)^{3/2}, \quad \|h\|_\infty \leq M_n,$$

$$\left. [r_1, r_2] \subset [\theta_0 - C\log n, \theta_0 + C\log n] \cap \text{int}(\text{dom}(\phi_0)) \right\},$$

where we substituted the class $\mathcal{U}_n(C)$ in $\mathcal{H}_n(C)$ by the class $\mathcal{U}_n^{sym}(C, r_1, r_2)$. Although the dependence of M_n on C is suppressed by its notation, the former is a function of C and n. This validates that the set $\mathcal{H}_n^{sym}(C)$ depends only on C and n, as indicated by the notation. Note that $[-\xi_n, \xi_n] \subset \text{int}(\text{dom}(\psi_0))$ by Lemma B.2, and ξ_n is $O_p(\log n)$ by Lemma B.4. Therefore proceeding as in Theorem 1, but replacing Lemma B.16 by Lemma D.7, we can also show that the function

$$h_n(x) = ((\tilde{\psi}_n^{sym,sm})'(x - \bar{\theta}_n) - \phi_0'(x))1_{[\bar{\theta}_n - \xi_n, \bar{\theta}_n + \xi_n]}(x), \quad x \in \mathbb{R}, \tag{70}$$

is a member of $\mathcal{H}_n^{sym}(C)$ with high probability for sufficiently large n. Using (69) in conjuction with (26) we can show that

$$\sup_Q \log N_{[\,]}(\epsilon, \mathcal{H}_n^{sym}(C), L_2(Q)) \lesssim \frac{C(\log n)M_n}{\epsilon}. \tag{71}$$

Since the bracketing entropy of $\mathcal{H}_n^{sym}(C)$ differs from that of $\mathcal{H}_n(C)$ only by a poly-log term, so does the entropy integral. Also, noting $\tilde{g}_n^{sym,sm}$ yields a consistent $\widehat{\mathcal{I}}_n(\eta_n)$ (see Lemma D.5) analogous to the log-concave $\tilde{g}_n$'s, rest of the proof of $T_1 = o_p(1)$ follows in a similar fashion as that of Theorem 1.

Step two: showing $T_{2n} \to_p -1$:

Recall the function b_n defined in (27). Because $T_{2n} = -\int_{\mathbb{R}} b_n(t)dt$, it suffices to show that $\mathbb{Y}_n = \int_{\mathbb{R}} b_n(t)dt \to_{a.s.} 1$. The proof is not much different from the proof of Lemma B.1. We will only point out where the current proof differs from the proof of Lemma B.1. Suppose $\mathcal{A}_n$ and $\mathcal{A}_n'$ are as defined in the proof of Lemma B.1. Let us also introduce the integrals

$$\mathcal{I}_{1n}^+ = \int_{\mathcal{A}_n} \tilde{\phi}_n^{sm}(\bar{\theta}_n + t)^2 \widehat{h}_n^{sm}(\bar{\theta}_n + t)dt, \quad \mathcal{I}_{2n}^+ = \int_{\mathcal{A}_n + \tilde{\delta}_n} \tilde{\phi}_n^{sm}(\bar{\theta}_n + t)^2 \widehat{h}_n^{sm}(\bar{\theta}_n + t)dt,$$

$$\mathcal{I}_{1n}^- = \int_{\mathcal{A}_n} \tilde{\phi}_n^{sm}(\bar{\theta}_n - t)^2 \widehat{h}_n^{sm}(\bar{\theta}_n - t)dt, \quad \mathcal{I}_{2n}^- = \int_{\mathcal{A}_n + \tilde{\delta}_n} \tilde{\phi}_n^{sm}(\bar{\theta}_n - t)^2 \widehat{h}_n^{sm}(\bar{\theta}_n - t)dt.$$

The above integrals replace the integrals $\mathcal{I}_{1n}$ and $\mathcal{I}_{2n}$ in the proof of Lemma B.1. We also define

$$\mathcal{J}_n^+ = \frac{(\log n)^2 H(\widehat{h}_n^{sm}(\bar{\theta}_n + \cdot), g_0)^2}{\inf_{x \in \mathcal{A}_n'} \widehat{h}_n^{sm}(\bar{\theta}_n + x)}, \quad \mathcal{J}_n^- = \frac{(\log n)^2 H(\widehat{h}_n^{sm}(\bar{\theta}_n - \cdot), g_0)^2}{\inf_{x \in \mathcal{A}_n'} \widehat{h}_n^{sm}(\bar{\theta}_n - x)}$$

Similar to Lemma B.1, it can be shown that it suffices to show that every subsequence has a further subsequence n_k, along which, $\mathbb{Y}_{n_k} \to_{a.s.} 1$. We claim that given any sequence, there exists a subsequence n_k such that the set $\mathcal{M}^{sym}$ has probability one, where we define $\mathcal{M}^{sym}$ to be the set on which the followings hold:
(a) $\bar{\theta}_{n_k} \to_k \theta_0$, (b) $\widehat{\mathcal{I}}_{n_k}(\eta_{n_k}) \to_k \mathcal{I}_{f_0}$, (c) $\xi_{n_k} \to_k G_0^{-1}(1)$, (d) $\omega_{n_k} \to \omega_0$, (e) $\mathcal{J}_{n_k}^+$, $\mathcal{J}_{n_k}^- \to_k 0$, (f) $\mathcal{I}_{in_k}^+, \mathcal{I}_{in_k}^- \to_k \mathcal{I}_{f_0}$ for $i = 1, 2$, (g) $\|\widehat{h}_{n_k}^{sm}(\bar{\theta}_{n_k} \pm \cdot) - g_0\|_\infty \to 0$, (h) $\|\tilde{g}_{n_k}^{sym,sm} - g_0\|_\infty \to 0$, (i) $\mathcal{A}_{n_k}' \subset \text{int}(\text{dom}(\psi_0))$ for all sufficiently large k.

Note that $\mathcal{M}^{sym}$ is similar to the good set $\mathcal{M}$ in the proof of Lemma B.1. The claim that there exists a sequence n_k so that $P(\mathcal{M}^{sym}) = 1$ can be verified using Fact 6 in the same way we verified a similar claim for $\mathcal{M}$. The only difference is that here we require Lemma D.5 for (b), Lemma D.6 instead of Lemma B.7 for (e), and Lemma D.5 instead of Lemma B.14 for (f). As in Lemma B.1, we will show that $\mathbb{Y}_{n_k} \to_k 1$ on $\mathcal{M}^{sym}$. For the sake of simplicity, we drop k from the subscripts.

The pointwise converges of b_n can be proved along the lines of (32). However, Lemma B.8 can not be directly applied this time because $\tilde{g}_n^{sym,sm}$ is not log-concave. On the other hand, Lemma B.8 does apply to $\widehat{h}_n^{sm}(\bar{\theta}_n \pm \cdot)$, because the latter is log-concave. Exploiting the connection between $(\tilde{\psi}_n^{sym,sm})'$ and $(\tilde{\phi}_n^{sm})'$ as given by (54), and arguing as in the proof of Proposition 2, we can show that the assertions of Lemma B.8 still hold for $\tilde{g}_n^{sym,sm}$ on $\mathcal{M}^{sym}$. Thus (32) holds for b_n in case of $\tilde{g}_n^{sym,sm}$.

However, we can not bound b_n using (33) because $(\tilde{\psi}_n^{sym,sm})'$ is not monotone. However, using (54), we can still bound

$$\left| \int_t^{t+\tilde{\delta}_n} (\tilde{\psi}_n^{sym,sm})'(z)dz \right| \leq \int_t^{t+\tilde{\delta}_n} \left(|(\tilde{\phi}_n^{sm})'(\bar{\theta}_n + z)| + |(\tilde{\phi}_n^{sm})'(\bar{\theta}_n - z)| \right) dz$$
$$\leq \tilde{\delta}_n \Big(\max\{|(\tilde{\phi}_n^{sm})'(\bar{\theta}_n + t + \tilde{\delta}_n)|, |(\tilde{\phi}_n^{sm})'(\bar{\theta}_n + t)|\}$$
$$+ \max\{|(\tilde{\phi}_n^{sm})'(\bar{\theta}_n - t - \tilde{\delta}_n)|, |(\tilde{\phi}_n^{sm})'(\bar{\theta}_n - t)|\} \Big).$$

Using the above, it can be shown that $|b_n(t)| \leq |b_n^+(t)| + |b_n^-(t)|$, where

$$b_n^+(t) = 1_{A_n}(t)|\psi_0'(t)|g_0(t)\Big(|(\tilde{\phi}_n^{sm})'(\bar{\theta}_n + t)| + |(\tilde{\phi}_n^{sm})'(\bar{\theta}_n + t + \tilde{\delta}_n)| \Big)/\widehat{\mathcal{I}}_n(\eta_n),$$
$$b_n^-(t) = 1_{A_n}(t)|\psi_0'(t)|g_0(t)\Big(|(\tilde{\phi}_n^{sm})'(\bar{\theta}_n - t)| + |(\tilde{\phi}_n^{sm})'(\bar{\theta}_n - t - \tilde{\delta}_n)| \Big)/\widehat{\mathcal{I}}_n(\eta_n).$$

The proof will be complete by Pratt's Lemma (Fact 10) if we can show that there exists integrable functions c_n^+, c_n^-, c^+ and c^- so that $|b_n^+| \leq c_n^+$, $|b_n^-| \leq c_n^-$, $\int_{\mathbb{R}} c_n^+(t)dt \to_n \int_{\mathbb{R}} c^+(t)dt$, $\int_{\mathbb{R}} c_n^-(t)dt \to_n \int_{\mathbb{R}} c^-(t)dt$, and $c_n^+ \to_n c^+$ and $c_n^- \to_n c^-$ almost everywhere Lebesgue on $\mathcal{M}^{sym}$. The functions c_n^+ and c_n^- can be constructed in the same way we constructed c_n for bounding b_n in the proof of Lemma B.1. Since the proof follows in a similar manner by replacing $\mathcal{M}$ by $\mathcal{M}^{sym}$, and $\tilde{g}_n$ by $\widehat{h}_n^{sm}(\bar{\theta}_n \pm \cdot)$, it is skipped. $\square$

D.1. Auxiliary lemma for Theorem 2

In this subsection, ξ_n will generally refer to $\xi_n(\tilde{G}_n^{sym,sm}) \equiv (\tilde{G}_n^{sym,sm})^{-1}(1-\eta_n)$. Although $\tilde{g}_n$ can be either $\widehat{h}_n^{sm}(\bar{\theta}_n \pm \cdot)$ or $\tilde{g}_n^{sym,sm}$, its definition should be clear from the context.

D.1.1. Lemmas on Hellinger error of $\tilde{g}_n^{sym,sm}$:

Lemma D.1. *Under the conditions of Theorem 2, $H(\widehat{h}_n^{sm}(\bar{\theta}_n\pm\cdot), g_0)=O_p(n^{-1/5})$ and $H(\tilde{g}_n^{sym,sm}, g_0) = O_p(n^{-1/5})$.*

Proof of Lemma D.1. First of all note that $2H^2(\widehat{h}_n^{sm}, f_0)^2$ is bounded by $\|\widehat{h}_n^{sm} - f_0\|_1$ which is not larger than

$$\widehat{\lambda}_n^{-1} \int_{-\infty}^{\infty} \left| \int_{-\infty}^{\infty} \Big(\widehat{h}_n(x-t) - f_0(x-t)\Big)\varphi(t/\widehat{\lambda}_n)dt\right|dx$$

$$+ \widehat{\lambda}_n^{-1} \int_{-\infty}^{\infty} \left| \int_{-\infty}^{\infty} \Big(f_0(x) - f_0(x-t)\Big)\varphi(t/\widehat{\lambda}_n)dt\right|dx$$

$$\leq \|\widehat{h}_n - f_0\|_1 + \widehat{\lambda}_n^{-1} \int_{-\infty}^{\infty} \left| \int_{-\infty}^{\infty} \varphi(t/\widehat{\lambda}_n) \int_{x-t}^{x} f_0'(z)dzdt\right|dx,$$

whose first term can be bounded using Fact 5, which yields

$$\|\widehat{h}_n - f_0\|_1 \leq \sqrt{2}H(\widehat{h}_n, f_0) = O_p(n^{-2/5})$$

by Theorem 3.2 of Doss and Wellner (2016). The second term

$$\widehat{\lambda}_n^{-1} \int_{-\infty}^{\infty} \left| \int_{-\infty}^{\infty} \varphi(t/\widehat{\lambda}_n) \int_{x-t}^{x} f_0'(z)dzdt\right|dx$$

$$\leq \widehat{\lambda}_n^{-1} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{0} \int_{x}^{x-t} \varphi(t/\widehat{\lambda}_n)|f_0'(z)|dzdt \right.$$

$$+ \int_{0}^{\infty} \int_{x-t}^{x} \varphi(t/\widehat{\lambda}_n)|f_0'(z)|dzdt \Bigg) dx$$

$$= \widehat{\lambda}_n^{-1} \int_{-\infty}^{\infty} \left(\int_{0}^{\infty} \int_{x}^{x+t} \varphi(-t/\widehat{\lambda}_n)|f_0'(z)|dzdt \right.$$

$$+ \int_{0}^{\infty} \int_{x-t}^{x} \varphi(t/\widehat{\lambda}_n)|f_0'(z)|dzdt \Bigg) dx$$

$$\leq \widehat{\lambda}_n^{-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{x-|t|}^{x+|t|} \varphi(t/\widehat{\lambda}_n)|f_0'(z)|dzdtdx$$

$$= \widehat{\lambda}_n^{-1} \int_{-\infty}^{\infty} \varphi(t/\widehat{\lambda}_n) \int_{-\infty}^{\infty} \int_{x-|t|}^{x+|t|} |f_0'(z)|dzdxdt$$

$$= 2 \int_{-\infty}^{\infty} |t|\widehat{\lambda}_n^{-1}\varphi(t/\widehat{\lambda}_n) \int_{-\infty}^{\infty} |f_0'(z)|dzdt$$

$$= 4\widehat{\lambda}_n E[|\mathbb{Z}|]f_0(\theta_0)$$

where $\mathbb{Z} \sim N(0,1)$. In the last step we used the fact $\int_{\mathbb{R}} |f_0'(z)|dz = 2f_0(\theta_0)$ which follows because $f_0 \in \mathcal{P}_0$. Thus

$$H^2(\widehat{h}_n^{sm}, f_0) = O_p(n^{-2/5}) + O_p(\widehat{\lambda}_n). \tag{72}$$

Our next step is finding the rate of $\widehat{\lambda}_n$. To that end, note that because $\int_R x\widehat{h}_n(x)dx$ is the sample average (Corollary 2.3 of Dümbgen and Rufibach, 2009), (10) implies that $\widehat{\lambda}_n^2 = \int_{\mathbb{R}} z^2 d(\mathbb{F}_n - \widehat{H}_n)$ where $\widehat{H}_n$ is the distribution function of $\widehat{h}_n$. Therefore,

$$\widehat{b}_n^2 \leq \left| \int_{-\infty}^{\infty} z^2 d(\mathbb{F}_n - F_0)\right| + \left| \int_{-\infty}^{\infty} z^2 d(\widehat{H}_n - F_0)\right|,$$

whose first term is $O_p(n^{-1/2})$ by the central limit theorem because F_0 has finite second central moment. On the other hand, since the second term equals

$$\left| \int_{-\infty}^{\infty} z^2 (\sqrt{\widehat{h}_n(z)} - \sqrt{f_0(z)})(\sqrt{\widehat{h}_n(z)} + \sqrt{f_0(z)})dz \right|,$$

The Cauchy-Schwarz inequality indicates that its square is bounded by

$$4H(\widehat{h}_n, f_0)^2 \int_{-\infty}^{\infty} z^4 (\widehat{h}_n(z) + f_0(z))dz.$$

The fourth moment of f_0 is finite by Fact 1. On the other hand, Theorem 4 of Cule and Samworth (2010) implies that there exists $a > 0$ so that

$$\int_{-\infty}^{\infty} e^{a|z|} |\widehat{h}_n(z) - f_0(z)|dz \to_{a.s.} 0.$$

Therefore it follows that $\int_{\mathbb{R}} z^4 \widehat{h}_n(z)dz = O_p(1)$. Thus, we conclude $\widehat{\lambda}_n$ is $O_p(1)H(\widehat{h}_n, f_0)$, which is $O_p(n^{-1/5})$. Therefore, (72) yields that $H(\widehat{h}_n^{sm}, f_0)$ is $O_p(n^{-1/5})$. Since the Hellinger distance is translation invariant,

$$H(\widehat{h}_n^{sm}(\bar{\theta}_n + \cdot), g_0) = H(\widehat{h}_n^{sm}, g_0(-\bar{\theta}_n + \cdot)) \le H(\widehat{h}_n^{sm}, f_0) + H(g_0(-\bar{\theta}_n + \cdot), f_0),$$

whose first term is $O_p(n^{-1/5})$, and second term is $O_p(|\bar{\theta}_n - \theta_0|)$ by Fact 15. Because $\bar{\theta}_n - \theta_0 = O_p(n^{-1/2})$, $H(\widehat{h}_n^{sm}(\bar{\theta}_n + \cdot), g_0) = O_p(n^{-1/5})$ follows. Since g_0 is symmetric about zero, we can show that $H(\widehat{h}_n^{sm}(\bar{\theta}_n - \cdot), g_0) = O_p(n^{-1/5})$ as well. Since $2g_0(x) = g_0(x) + g_0(-x)$, and

$$(\sqrt{a+b} - \sqrt{c+d})^2 \lesssim (\sqrt{a} - \sqrt{c})^2 + (\sqrt{b} - \sqrt{d})^2 \quad \text{for } a, b, c, d > 0,$$

it follows that

$$H(\tilde{g}_n^{sym,sm}, g_0) \lesssim H(\widehat{h}_n(\bar{\theta}_n + \cdot), g_0) + H(\widehat{h}_n(\bar{\theta}_n - \cdot), g_0) = O_p(n^{-1/5}). \qquad \square$$

D.1.2. *Lemmas on distance between $\tilde{G}_n^{sym,sm}$ and $\widehat{H}_n^{sm}(\bar{\theta}_n \pm \cdot)$:*

Lemma D.2. *Under the set up of Theorem 2,*

$$(A) \quad \|\tilde{G}_n^{sym,sm} - \widehat{H}_n^{sm}(\bar{\theta}_n \pm \cdot)\|_\infty = O_p(n^{-p})$$

$$(B) \quad \sup_{x \in [-\xi_n, \xi_n]} \widehat{h}_n^{sm}(\bar{\theta}_n \pm x)^{-1} = O_p(\eta_n^{-1}),$$

where $p = 1/5$ and $\xi_n = (\tilde{G}_n^{sym,sm})^{-1}(1 - \eta_n)$.

Proof of Lemma D.2. From the definition of total variation distance, it follows that
$\|\tilde{G}_n^{sym,sm} - \widehat{H}_n^{sm}(\bar{\theta}_n \pm \cdot)\|_\infty \le d_{TV}(\tilde{G}_n^{sym,sm}, \widehat{H}_n^{sm}(\bar{\theta}_n \pm \cdot))$, which equals

$$2\|\tilde{g}_n^{sym,sm} - \widehat{h}_n^{sm}(\bar{\theta}_n \pm \cdot)\|_1 \le 2\sqrt{2}H(\tilde{g}_n^{sym,sm}, \widehat{h}_n^{sm}(\bar{\theta}_n \pm \cdot)),$$

where the last step follows by Fact 5. The proof of part (A) then follows noting

$$H(\tilde{g}_n^{sym,sm}, \widehat{h}_n^{sm}(\bar{\theta}_n \pm \cdot)) \le H(\tilde{g}_n^{sym,sm}, g_0) + H(\widehat{h}_n^{sm}(\bar{\theta}_n \pm \cdot), g_0) = O_p(n^{-p})$$

by Lemma D.1.

For the proof of part (B), note that since $\widehat{h}_n^{sm}$ is log-concave, it attains its minimum on any interval at one of the endpoints. Therefore

$$\inf_{x\in[-\xi_n,\xi_n]} \widehat{h}_n^{sm}(\bar{\theta}_n \pm x) = \min(\widehat{h}_n^{sm}(\bar{\theta}_n + \xi_n), \widehat{h}_n^{sm}(\bar{\theta}_n - \xi_n)).$$

Using Fact 4 in step (a), and part A of the current lemma in step (b), we can show that

$$\widehat{h}_n^{sm}(\bar{\theta}_n - \xi_n) \overset{(a)}{\geq} \omega_n \widehat{H}_n^{sm}(\bar{\theta}_n - \xi_n) \overset{(b)}{\geq} \omega_n(\tilde{G}_n^{sym,sm}(-\xi_n) - O_p(n^{-p})),$$

which, by definition of ξ_n, equals $\omega_n \eta_n - O_p(n^{-p})$. Since $\omega_n \to_p \omega_0 > 0$ by Fact 4, and $\eta_n = Cn^{-2p/5}$, it follows that $\widehat{h}_n^{sm}(\bar{\theta}_n - \xi_n)^{-1} = O_p(\eta_n^{-1})$. In a similar way, it can be shown that $\widehat{h}_n^{sm}(\bar{\theta}_n + \xi_n)^{-1} = O_p(\eta_n^{-1})$. Therefore, the proof follows. □

D.1.3. *Lemmas on $(\tilde{\phi}_n^{sm})'$ and $(\tilde{\psi}_n^{sym,sm})'$:*

Lemma D.3. *The conditions of Lemma B.12 and Lemma B.13 hold for*

$$a_n = \xi_n \equiv \xi_n(\tilde{G}_n^{sym,sm}) = (\tilde{G}_n^{sym,sm})^{-1}(1 - \eta_n) \text{ and } \tilde{g}_n = \widehat{h}_n^{sm}(\bar{\theta}_n \pm \cdot).$$

Proof of Lemma D.3. Note that since $\tilde{g}_n^{sym,sm}$ satisfies Condition 3, Lemma B.9 entails that $a_n = \xi_n$ satisfies (43). Moreover, Lemma D.2 (B) indicates that the supremum of $-\tilde{\phi}_n^{sm}(\bar{\theta}_n \pm \cdot)$ over $[-a_n, a_n]$ is $O_p(\log n)$ for the above choice of a_n. Also because $\widehat{h}_n^{sm}$ satisfies Condition 1, $\tilde{\phi}_n^{sm}$ is bounded above. Therefore we obtain that the supremum of $|\tilde{\phi}_n^{sm}(\bar{\theta}_n \pm \cdot)|$ on $[-a_n, a_n]$ is $O_p(\log n)$. Thus, we conclude that (41) holds for our choice of a_n and $\tilde{g}_n$. As a result, this $(a_n, \tilde{g}_n)$ pair satisfies the conditions of Lemma B.12.

For Lemma B.13, first note that $a_n = O_p(\log n)$ by Lemma B.4. Noting $\text{dom}(\tilde{\phi}_n^{sm}) = \text{supp}(\widehat{h}_n^{sm}) = \mathbb{R}$, we also obtain that (44) holds with probability tending to one because

$$P\Big([-\xi_n - \eta_n/\log n, \xi_n + \eta_n/\log_n] \subset \text{int}(\text{dom}(\psi_0))\Big) \to 1$$

by Lemma B.6. Thus the conditions of Lemma B.13 are satisfied if (45) holds for $(\xi_n, \widehat{h}_n^{sm}(\bar{\theta}_n \pm \cdot))$. Now by Lemma D.2, with probability tending to one,

$$\widehat{H}_n^{sm}(\bar{\theta}_n - a_n) \geq \tilde{G}_n^{sym,sm}(-a_n) - o_p(n^{-p}) = \tilde{G}_n^{sym,sm}((\tilde{G}_n^{sym,sm})^{-1}(\eta_n)) - o_p(n^{-p}),$$

which is $\eta_n - o_p(n^{-p})$. Since η_n is $O(n^{-2p/5})$, it follows that

$$P\Big(\widehat{H}_n^{sm}(\bar{\theta}_n - a_n) \geq \eta_n/4\Big) \to 1.$$

Similarly we can show that

$$P\Big(1 - \widehat{H}_n^{sm}(\bar{\theta}_n + a_n) \geq \eta_n/4\Big) \to 1,$$

which implies (45) holds for $a_n = \xi_n$ when $\tilde{g}_n = \widehat{h}_n^{sm}(\bar{\theta}_n + \cdot)$. The proof for $\tilde{g}_n = \widehat{h}_n^{sm}(\bar{\theta}_n - \cdot)$ follows in a similar way, which completes the proof of the current lemma. □

Lemma D.4. *Suppose μ_n is a density so that $\|\mu_n\|_\infty = O_p(1)$. Let $\xi_n = (\tilde{G}_n^{sym,sm})^{-1}(1 - \eta_n)$. Then $p = 1/5$ satisfies*

$$\int_{-\xi_n}^{\xi_n} ((\tilde{\psi}_n^{sym,sm})'(x) - \psi_0'(x))^2 \mu_n(x) dx = O_p((\log n)^3 n^{-4p/5}).$$

Proof of lemma D.4. Using the representation of $(\tilde{\psi}_n^{sym,sm})'$ given by (54), we obtain that

$$\int_{-\xi_n}^{\xi_n} \left((\tilde{\psi}_n^{sym,sm})'(x) - \psi_0'(x) \right)^2 \mu_n(x) dx$$

$$= \int_{-\xi_n}^{\xi_n} \left((\tilde{\psi}_n^{sym,sm})'(x) - \varrho_n(x)\psi_0'(x) + (1 - \varrho_n(x))\psi_0'(-x) \right)^2 \mu_n(x) dx$$

$$\leq 2 \int_{-\xi_n}^{\xi_n} \varrho_n(x)^2 \left((\tilde{\phi}_n^{sm})'(\bar{\theta}_n + x) - \psi_0'(x) \right)^2 \mu_n(x) dx$$

$$+ 2 \int_{-\xi_n}^{\xi_n} (1 - \varrho_n(x))^2 \left((\tilde{\phi}_n^{sm})'(\bar{\theta}_n - x) - \psi_0'(-x) \right)^2 \mu_n(x) dx$$

$$\overset{(a)}{\leq} 2 \int_{-\xi_n}^{\xi_n} \left\{ \left((\tilde{\phi}_n^{sm})'(\bar{\theta}_n + x) - \psi_0'(x) \right)^2 \right.$$

$$\left. + \left((\tilde{\phi}_n^{sm})'(\bar{\theta}_n - x) - \psi_0'(-x) \right)^2 \right\} \mu_n(x) dx$$

which is $O_p((\log n)^3 n^{-4p/5})$ by Lemma D.3 and Lemma B.13. Here (a) follows because ϱ_n is a fraction. $\qquad\square$

Lemma D.5. *Suppose $\eta_n = Cn^{-2p/5}$, where $p = 1/5$ and $C > 0$. Let $\xi_n = (\tilde{G}_n^{sym,sm})^{-1}(1 - \eta_n)$. Then under the set up of Theorem 2,*

$$\int_{-\xi_n}^{\xi_n} (\tilde{\phi}_n^{sm})'(\bar{\theta}_n \pm x)^2 \hat{h}_n^{sm}(\bar{\theta}_n \pm x) \to_p \mathcal{I}_{f_0}, \tag{73}$$

and

$$\int_{-\xi_n}^{\xi_n} (\tilde{\psi}_n^{sym,sm})'(x)^2 \tilde{g}_n^{sym,sm}(\bar{\theta}_n \pm x) \to_p \mathcal{I}_{f_0}. \tag{74}$$

Proof of Lemma D.5. It suffices to show that the pairs $(\xi_n, \hat{h}_n^{sm}(\bar{\theta}_n \pm \cdot))$ and $(\xi_n, \tilde{g}_n^{sym,sm})$ satisfy the conditions of Lemma B.14. By Lemma D.3, $a_n = \xi_n$ satisfies the conditions of Lemma B.13, which entails that (a) a_n is $O_p(\log n)$ and (b) (44) holds for a_n with probability tending to one, where (44) implies $[-a_n, a_n] \subset \text{int}(\text{dom}(\psi_0))$. Next, the condition $a_n = \xi_n \to_p G_0^{-1}(1)$ holds by Lemma B.3. Finally, (51) holds for $\hat{h}_n^{sm}(\bar{\theta}_n \pm \cdot)$ and $\tilde{g}_n^{sym,sm}$ by Lemma D.3 and Lemma D.4, respectively. Therefore the proof follows from Lemma B.14. $\qquad\square$

Lemma D.6. *Let $\xi_n = (\tilde{G}_n^{sym,sm})^{-1}(1 - \eta_n)$ where $\eta_n = Cn^{-2p/5}$ for some $C > 0$ and $p = 1/5$. Suppose y_n is a sequence of random variables such that*

$P(|y_n| \leq \eta_n/(2g_0(0))) \to 1$. *Then under the conditions of Theorem* 2, *we have*

$$\sup_{x \in [-\xi_n - y_n, \xi_n + y_n]} \left\{ |(\tilde{\psi}_n^{sym,sm})'(x)| + |(\tilde{\phi}_n^{sm})'(\bar{\theta}_n \pm x)| \right\} = O_p(n^{p/5}).$$

Proof of Lemma D.6. Note that (54) implies

$$|(\tilde{\psi}_n^{sym,sm})'(x)| \leq \max \left\{ |(\tilde{\phi}_n^{sm})'(\bar{\theta}_n + x)|, |(\tilde{\phi}_n^{sm})'(\bar{\theta}_n - x)| \right\}.$$

Therefore it suffices to bound $|(\tilde{\phi}_n^{sm})'(\bar{\theta}_n \pm)|$ only. Since the proof of $(\tilde{\phi}_n^{sm})'(\bar{\theta}_n + \cdot)$ and $(\tilde{\phi}_n^{sm})'(\bar{\theta}_n - \cdot)$ are similar, we only show the proof for $(\tilde{\phi}_n^{sm})'(\bar{\theta}_n + \cdot)$. Denoting $\tilde{g}_n = \hat{h}_n^{sm}(\bar{\theta}_n + \cdot)$, $\tilde{\psi}_n = (\tilde{\phi}_n^{sm})'(\bar{\theta}_n + \cdot)$, and $\tilde{G}_n = \hat{H}_n^{sm}(\bar{\theta}_n + \cdot)$, note that the following holds for any $q \in (0, 1/2)$ by Fact 8 because $\tilde{g}_n$ is positive on $J(\tilde{G}_n)$:

$$\int_{(\tilde{G}_n^{sym,sm})^{-1}(q/2)}^{(\tilde{G}_n^{sym,sm})^{-1}(q)} \tilde{\psi}_n'(x)^2 \tilde{g}_n(x) dx = \int_{\tilde{G}_n((\tilde{G}_n^{sym,sm})^{-1}(q/2))}^{\tilde{G}_n((\tilde{G}_n^{sym,sm})^{-1}(q))} \tilde{\psi}_n'(\tilde{G}_n^{-1}(z))^2 dz.$$

Because $(\tilde{\psi}_n')^2$ is non-increasing on $(-\infty, 0]$, the above yields

$$\tilde{\psi}_n'((\tilde{G}_n^{sym,sm})^{-1}(q))^2 \leq \frac{\displaystyle\int_{(\tilde{G}_n^{sym,sm})^{-1}(q/2)}^{(\tilde{G}_n^{sym,sm})^{-1}(q)} \tilde{\psi}_n'(x)^2 \tilde{g}_n(x) dx}{\tilde{G}_n((\tilde{G}_n^{sym,sm})^{-1}(q)) - \tilde{G}_n((\tilde{G}_n^{sym,sm})^{-1}(q/2))}.$$

Letting $q = \eta_n/2$, and denoting $\tilde{\xi}_n = (\tilde{G}_n^{sym,sm})^{-1}(1 - \eta_n/2)$, we obtain that

$$\tilde{\psi}_n'(\tilde{\xi}_n)^2 \leq \int_{(\tilde{G}_n^{sym,sm})^{-1}(\eta_n/4)}^{(\tilde{G}_n^{sym,sm})^{-1}(\eta_n/2)} \tilde{\psi}_n'(x)^2 \tilde{g}_n(x) dx \bigg/ \left(\eta_n/4 - 2\|\tilde{G}_n - \tilde{G}_n^{sym,sm}\|_\infty \right).$$

Now Lemma D.2 implies

$$\eta_n/4 - 2\|\tilde{G}_n - \tilde{G}_n^{sym,sm}\|_\infty = \eta_n/4 - O_p(n^{-p}),$$

whose dominating term is $\eta_n/4$. Also

$$\int_{(\tilde{G}_n^{sym,sm})^{-1}(\eta_n/4)}^{(\tilde{G}_n^{sym,sm})^{-1}(\eta_n/2)} \tilde{\psi}_n'(x)^2 \tilde{g}_n(x) dx \leq \int_{(\tilde{G}_n^{sym,sm})^{-1}(\eta_n/4)}^{(\tilde{G}_n^{sym,sm})^{-1}(1-\eta_n/4)} \tilde{\psi}_n'(x)^2 \tilde{g}_n(x) dx$$

which converges in probability to $\mathcal{I}_{f_0}$ by Lemma D.5. Therefore $\tilde{\psi}_n'(\tilde{\xi}_n)$ is $O_p(\eta_n^{-1/2})$. The rest of the proof follows similar to the proof of Lemma B.15. □

Lemma D.7. *Suppose $\tilde{g}_n$ satisfies Condition* 2. *Then under the set up of Theorem* 2,

$$\int_{-\xi_n}^{\xi_n} ((\tilde{\psi}_n^{sym,sm})'(x) - \psi_0'(z - \tilde{\delta}_n))^2 f_0(z + \bar{\theta}_n) dz = O_p(n^{-4p/5}(\log n)^3).$$

Proof of Lemma D.7. The proof is similar to the proof of Lemma B.16. The only difference is that one needs to use use Lemma D.4 instead of Lemma B.13 to bound T_1 □

D.1.4. *Lemmas on consistency of Fisher information:*

Lemma D.8. *Under the set up of Theorem* 1, $\hat{\mathcal{I}}_n(\eta_n) \to_p \mathcal{I}_{f_0}$.

Proof of Lemma D.8. The proof follows in the same way as Lemma B.17 by replacing Lemma B.14 and Lemma B.15 by Lemma D.5 Lemma D.6, respectively, and replacing the class of monotone functions $\mathcal{U}_n(M_n)$ by the class $\mathcal{U}_n^{sym}(M_n, r_1, r_2)$ defined in (68). □

Appendix E: Proof of Lemma 1

Proof of Lemma 1. For $k \geq 1$, we denote A_k to be the set $[-1/(2k), 1/(2k)]$, and consider the sequence of functions $\{\psi_k\}_{k \geq 1} \in \mathcal{SC}_0$ defined by

$$\psi_k(x) = \begin{cases} \log k, & x \in A_k \\ -\infty, & \text{o.w.} \end{cases}$$

Observe that

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

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

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

leading to $\widehat{\psi}_n(x_0 - \widehat{\theta}_n) = \infty$, which contradicts the fact that $\widehat{\psi}_n$ is a proper concave function. Hence, we conclude that the MLE of (θ_0, ψ_0) does not exist when $\mathbb{F}_n$ is degenerate. □

Appendix F: Proof of Theorem 3

To prove Theorem 3, it will be beneficial to prove a general result first. We begin by stating a condition.

Condition 4. (Existence of log-concave projection.) F is a non-degenerate distribution function with finite first moment.

Any F satisfying Condition 4 has a well-defined log-concave projection, i.e. its projection (with respect to the KL divergence) onto the space of all distributions with density in $\mathcal{LC}$ is a unique distribution function (Theorem 2.2, Dümbgen *et al.*, 2011). Note that $\mathbb{F}_n$ satisfies Condition 4 with probability one. We will show that for any distribution function F satisfying Condition 4,

$$(\theta^*(F), \psi^*(F)) = \underset{\theta \in \mathbb{R}, \psi \in \mathcal{SC}_0}{\arg\max} \Psi(\theta, \psi, F) \tag{75}$$

exists where Ψ is the criterion function defined in (13).

Proposition 3. *If F satisfies condition 4, then $\theta^*(F)$ and $\psi^*(F)$ exist.*

Observe that Proposition 3 implies the first part of Theorem 3 because if $F = \mathbb{F}_n$, $(\theta^*(F), \psi^*(F))$ corresponds to the MLE $(\widehat{\theta}_n, \widehat{g}_n)$. The second part of Theorem 3 follows from Lemma F.1, which is proved in Appendix F.1.

Lemma F.1. *Suppose F is such that $J(F) = (a, b)$ where $a, b \in \mathbb{R}$, and $J(F) = \{0 < F < 1\}$. Then Under condition 4, there exists at least one $\theta^*(F)$ so that $\theta^*(F) \in [a, b]$.*

Thus it remains to prove Proposition 3. To that end, we will need a continuity result on the partially maximized criterion function

$$L(\theta; F) = \sup_{\psi \in \mathcal{SC}_0} \Psi(\theta, \psi, F). \tag{76}$$

Lemma F.2. *Suppose the distribution function F satisfies condition 4. Then the map $\theta \mapsto L(\theta; F)$ is continuous on $\mathbb{R}$, where $L(\theta; F)$ is as defined in (76).*

The proof of Lemma F.2 can be found in Appendix F.1. Now we are ready to prove Proposition 3.

Proof of Proposition 3. Let us define

$$L(F) = \sup_{\theta \in \mathbb{R}, \psi \in \mathcal{SC}_0} \Psi(\theta, \psi, F). \tag{77}$$

Our first step is to show that $L(F)$ is finite. From the definition of Ψ in (13), 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 4, 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 note that the map $x \mapsto -|x| \in \mathcal{SC}_0$. Therefore, (13) and (77) lead to

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

which follows from condition 4. Hence, 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 Lemma F.2, 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 $L(\theta; F)$. Now we invoke Proposition 4(iii) of Xu and Samworth (2019), which states that for each $\theta \in \mathbb{R}$, there exists a unique log-density ψ_θ, which maximizes $\Psi(\theta, \psi, F)$ in $\psi \in \mathcal{SC}_0$ provided F satisfies condition 4. It is not hard to see that $(\theta', \psi_{\theta'})$ will be 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 $\theta_k \to_k \pm\infty$ leads to $L(\theta_k; F) \to_k -\infty$, which contradicts the fact that $L(\theta_k; F) \to_k L(F) \in \mathbb{R}$, thus completing the proof.

Consider $\theta_k \to_k \pm\infty$. By Proposition 4(iii) of Xu and Samworth (2019), for each θ_k, there exists a log-density $\psi_{\theta_k} \in \mathcal{SC}_0$ such that $L(\theta_k; F) = \Psi(\theta, \psi_{\theta_k}, F)$. Now note that if $e^\psi \in \mathcal{SLC}_0$, then ψ satisfies

$$2xe^{\psi(x)} \leq \int_{-x}^{x} e^{\psi(z)}dz \leq 1 \quad \text{for any } x \geq 0,$$

which implies $|\psi(x)| \leq -\log|2x|$. Noting $\psi_{\theta_k} \in \mathcal{SC}_0$ for each $k \geq 1$, we obtain that

$$\Psi(\theta, \psi_{\theta_k}, F) = \int_{-\infty}^{\infty} \psi_{\theta_k}(x - \theta)dF(x) - 1 \leq -\int_{-\infty}^{\infty} \log\left(2|x - \theta_k|\right)dF(x) - 1.$$

Now if $\theta_k \to_k \pm\infty$, using Fatou's Lemma, we derive that

$$\limsup_{k \to \infty} L(\theta_k; F) \leq -\int_{-\infty}^{\infty} \liminf_{k \to \infty}\left(\log|x - \theta_k|\right)dF(x) - (\log 2 + 1),$$

which is $-\infty$. This leads to the desired contradiction, which completes the proof. □

F.1. Auxilliary Lemmas for Theorem 3

Proof of Lemma F.2. Observe that (13) implies $\Psi(\theta, \psi, F)$ can also be written as $\Psi(\theta, \psi, F) = \Psi(0, \psi, F(\cdot + \theta))$. Hence, to prove Lemma F.2, it suffices to show that as $\theta_k \to_k \theta \in \mathbb{R}$,

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

Proposition 6 of Xu and Samworth (2019) implies that under condition 4, the convergence in the above display holds if the Wasserstein distance

$$d_W(F(\cdot + \theta_k), F(\cdot + \theta)) \to_k 0. \tag{78}$$

Now by Theorem 6.9 of Villani (2009) (see also Theorem 7.12 of Villani, 2003), (78) follows if (a) $F(\cdot + \theta_k)$ converges weakly to $F(\cdot + \theta)$ as $k \to \infty$, and

$$(b) \qquad \int_{-\infty}^{\infty} |x|dF(x + \theta_k) \to_k \int_{-\infty}^{\infty} |x|dF(x + \theta).$$

Now (a) follows noting that for any bounded continuous function h,

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

by the dominated convergence theorem since $\theta_k \to_k \theta$. For proving (b), first notice that F has finite first moment by condition 4. Therefore, another application of the dominated convergence yields that as $\theta_k \to_k \theta$,

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

which proves (b), and thus completes the proof. □

Proof of Lemma F.1. We will show that if $J(F) = (a, b)$, the functional $\theta \mapsto L(\theta; F)$ defined in (76) is non-decreasing in θ on $(-\infty, a]$, and non-increasing in θ on $[b, \infty)$. Suppose the above claim holds. Then clearly either $L(\theta, F)$ attains its maximum in $[a, b]$ or $L(\theta, F) = L(F)$ over an interval with nonempty overlap with $[a, b]$. Here $L(F)$ is as defined in (77). In either cases, one can find a $\theta^*(F) \in [a, b]$, which completes the proof of Lemma F.1

To show that $L(\theta; F)$ is non-decreasing in θ on $(-\infty, a]$, we first note that for $\theta < \theta' \leq a$, and $\psi \in \mathcal{SC}_0$,

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

since ψ is non-increasing on $[0, \infty)$, and $0 \leq x - \theta' < x - \theta$ for $x \geq a$. Therefore, from (76), it is not hard to see that $L(\theta; F) \leq L(\theta'; F)$. Similarly we can show that for $\theta > \theta' \geq b$,

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

since ψ is non-decreasing on $(-\infty, 0]$, and $x - \theta < x - \theta' \leq 0$ for $x \leq b$. Therefore, $L(\theta; F) \leq L(\theta'; F)$, which completes the proof. □

Appendix G: Proof of Theorem 5

Before going into the proof, we will introduce some new notations and state some lemmas that will be required later in the proof. We let $\widehat{G}_n$ and $\widehat{F}_n$ denote the distribution functions corresponding to $\widehat{g}_n$ and $\widehat{f}_n$, respectively. Also, we let $\widehat{\phi}_n$ denote the log-density $\log \widehat{f}_n$. Also, we let $\delta_n = \theta_0 - \widehat{\theta}_n$.

Now we state a lemma which basically says that $\log \widehat{f}_n$ is uniformly bounded above for sufficiently large n with probability one. This lemma is proved in Appendix G.1.

Lemma G.1. *Under the hypotheses of Theorem 5,*

$$P\left(\sup_n \sup_{x \in \mathbb{R}} \log \widehat{f}_n(0) < \infty \right) = 1.$$

We first show that $\widehat{\theta}_n \to_{a.s.} \theta_0$. 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) satisfies

$$\sum_{i=1}^n \log f_n(X_i) \geq \sum_{i=1}^n \log f_0(X_i) \tag{79}$$

with probability one, 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 $\sup_x \log \widehat{f}_n(x) = \widehat{\phi}_n(\widehat{\theta}_n) = \widehat{\psi}_n(0)$. Lemma G.1 entails that $P(\limsup_n \widehat{\psi}_n(0) < \infty) = 1$. Also, note that being the MLE of f_0,

$\widehat{f}_n$ automatically satisfies (79), which implies

$$H(\widehat{f}_n, f_0) \to_{a.s.} 0. \tag{80}$$

Denote by $\widehat{F}_n$ the distribution of $\widehat{f}_n$. Because $d_{TV}(\widehat{F}_n, F_0) \leq \sqrt{2}H(\widehat{f}_n, f_0)$ by Fact 15, (80) implies $d_{TV}(\widehat{F}_n, F_0) \to_{a.s.} 0$, which indicates $F_n \to_d F_0$ almost surely. In that case, Proposition 2 of Cule and Samworth (2010) implies that there exists $\alpha > 0$ so that

$$\int_{-\infty}^{\infty} e^{\alpha |x|} \Big| \widehat{f}_n(x) - f_0(x) \Big| dx \to_{a.s.} 0.$$

Therefore the moments of $\widehat{f}_n$ converges almost surely to that of f_0. Notably, the first moment of f_0 is θ_0, and because $\widehat{g}_n$ is symmetric about zero, we also have

$$\int_{-\infty}^{\infty} x\widehat{f}_n(x)dx = \widehat{\theta}_n + \int_{-\infty}^{\infty} (x - \widehat{\theta}_n)\widehat{g}_n(x - \widehat{\theta}_n)dx = \widehat{\theta}_n.$$

Thus $\widehat{\theta}_n \to_{a.s.} \theta_0$ follows. Since $\mathcal{I}_{f_0} < \infty$, the density f_0 is absolutely continuous (Theorem 3, Huber, 1964). Because f_0 is continuous, Proposition 2 of Cule and Samworth (2010) yields another useful result which will be required later:

$$\sup_{x \in \mathbb{R}} \Big| \widehat{f}_n(x) - f_0(x) \Big| \to_{a.s.} 0. \tag{81}$$

Next we show that $H(\widehat{g}_n, g_0) \to_{a.s.} 0$, which completes the proof of part A. To that end, note that

$$2H^2(\widehat{g}_n, g_0) = \int_{-\infty}^{\infty} \left(\sqrt{\widehat{g}_n(z - \widehat{\theta}_n)} - \sqrt{g_0(z - \widehat{\theta}_n)} \right)^2 dz$$

$$\leq 4H(\widehat{f}_n, f_0)^2 + 2\int_{-\infty}^{\infty} \left(\sqrt{g_0(z - \widehat{\theta}_n)} - \sqrt{g_0(z - \theta_0)} \right)^2 dz,$$

where the first term on the right hand side of the last display approaches zero almost surely by (80). The integrand in the second term is also bounded above by a constant multiple of $g_0(z - \widehat{\theta}_n) + g_0(z - \theta_0)$, which converges to $2f_0(z)$, and is integrable. Therefore, using Pratt's lemma (Fact 10), we deduce that the second term also converge to zero almost surely. Hence $H^2(\widehat{g}_n, g_0) \to_{a.s.} 0$ follows.

Now we turn to the proof of pat B, where we first establish that $H(\widehat{f}_n, f_0) = O_p(n^{-4/5})$. To that end, we first introduce the class of functions

$$\mathcal{P}_{M,0} = \left\{ f \in \mathcal{LC} \;\middle|\; \sup_{x \in \mathbb{R}} f(x) < M, \inf_{|x| > 1} f(x) > 1/M, \; \text{supp}(f) \subset \text{supp}(f_0) \right\}.$$

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, $\mathrm{int}(\mathrm{dom}(f_0))$ 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

$$f_0((x - \beta)/\alpha) > \alpha/M, \quad \text{for } x = \pm 1, \quad \text{yielding} \quad \min(\tilde{f}_0(-1), \tilde{f}_0(1)) > 1/M.$$

The above implies $\inf_{x \in [-1,1]} \tilde{f}_0(x) > 1/M$, 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, (81) implies

$$\limsup_n \sup_{x \in \mathbb{R}} \widehat{f}_n(x) < M, \quad \text{and} \quad \lim_n \widehat{f}_n(\pm 1) > 1/M.$$

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 for $H(\widehat{f}_n, f_0) = O_p(n^{-2/5})$ now follows from an application of Theorem 3.4.1 and 3.4.4 of Van der Vaart and Wellner (1996).

Now we turn to establishing the rate of convergences of $\widehat{\theta}_n$ and $\widehat{g}_n$. If $x - \theta_0$ is a continuity point of g_0', using the fact that $\widehat{\theta}_n \to_{a.s.} \theta_0$, 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

$$\liminf_n \frac{\int_{-\infty}^{\infty} \left(\sqrt{g_0(x - \widehat{\theta}_n)} - \sqrt{g_0(x - \theta_0)} \right)^2 dx}{(\widehat{\theta}_n - \theta_0)^2} \geq \left(\frac{g_0'(x - \theta_0)}{2\sqrt{g_0(x - \theta_0)}} \right)^2 dx = \frac{\mathcal{I}_{f_0}}{4} \tag{82}$$

with probability one. Now observe that

$$2H(\widehat{f}_n, f_0)^2 = \int_{-\infty}^{\infty} \left(\sqrt{\widehat{g}_n(x - \widehat{\theta}_n)} - \sqrt{g_0(x - \theta_0)} \right)^2 dx$$

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

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 (82) 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

$$\frac{|T_c|}{|\widehat{\theta}_n - \theta_0|^2 + H(\widehat{g}_n, g_0)^2} = o_p(1). \tag{83}$$

Suppose (83) holds. Then it follows that

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

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

Hence, it remains to prove (83). To this end, notice that T_c can be written as

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

Recalling $\delta_n = \theta_0 - \widehat{\theta}_n$, and noting g_0 is absolutely continuous because $f_0 \in \mathcal{P}_0$, 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

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

yielding

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

Using the Cauchy-Schwarz inequality, we obtain that

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

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,

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

which, noting
$$t \mapsto \frac{g_0'(x+t)}{2\sqrt{g_0(x+t)}} - \frac{g_0'(x-t)}{2\sqrt{g_0(x-t)}}$$
is an even function for each $x > 0$, can be bounded above by
$$\int_0^\infty |\delta_n| \left(\int_0^{|\delta_n|} \frac{1}{(\delta_n)^2} \left(\frac{g_0'(x+t)}{2\sqrt{g_0(x+t)}} - \frac{g_0'(x-t)}{2\sqrt{g_0(x-t)}} \right)^2 dt \right) dx$$
using the Cauchy-Schwarz inequality. Therefore, we obtain
$$\frac{T_c^2}{2H(\widehat{g}_n, g_0)^2 \delta_n^2} \leq \frac{1}{|\delta_n|} \int_0^{|\delta_n|} \left[\int_0^\infty \left(\frac{g_0'(x+t)}{\sqrt{g_0(x+t)}} \right)^2 dx + \int_0^\infty \left(\frac{g_0'(x-t)}{\sqrt{g_0(x-t)}} \right)^2 dx \right.$$
$$\left. - 2 \int_0^\infty \frac{g_0'(x-t)}{\sqrt{g_0(x-t)}} \frac{g_0'(x+t)}{\sqrt{g_0(x+t)}} dx \right] dt. \tag{84}$$

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 $z \in (-|\delta_n|, 0)$,
$$|g_0'(z)/\sqrt{g_0(z)}| = |\psi_0'(z)|\sqrt{g_0(z)} \leq |\psi_0'(\delta_n)|\sqrt{g_0(0)} = O_p(1), \tag{85}$$
since $\psi_0 \in \mathcal{SC}_0$, and $\delta_n \to_{a.s.} 0$. Hence, for $t \in (0, |\delta_n|)$,
$$\int_0^\infty \left(\frac{g_0'(x-t)}{\sqrt{g_0(x-t)}} \right)^2 dx = \int_{-t}^\infty \left(\frac{g_0'(z)}{\sqrt{g_0(z)}} \right)^2 dz$$
$$= \int_{-t}^0 \left(\frac{g_0'(x)}{\sqrt{g_0(z)}} \right)^2 dz + \int_0^\infty \left(\frac{g_0'(z)}{\sqrt{g_0(z)}} \right)^2 dz$$
$$\leq |\delta_n| \psi_0'(\delta_n)^2 \sqrt{g_0(0)} + \mathcal{I}_{f_0}/2$$
$$= |\delta_n| O_p(1) + \mathcal{I}_{f_0}/2,$$
where the last step follows from (85). Hence, for any $t \in (0, |\delta_n|)$,
$$\int_0^\infty \left(\frac{g_0'(x+t)}{\sqrt{g_0(x+t)}} \right)^2 dx + \int_0^\infty \left(\frac{g_0'(x-t)}{\sqrt{g_0(x-t)}} \right)^2 dx = |\delta_n| O_p(1) + \mathcal{I}_{f_0}. \tag{86}$$

Our objective is to apply Fatou's lemma on the third term on the right hand side of (84). 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
$$g_0'(x-t)g_0'(x+t) \geq 0. \tag{87}$$
Keeping that in mind, we partition the term
$$- \int_0^\infty \frac{g_0'(x-t)}{\sqrt{g_0(x-t)}} \frac{g_0'(x+t)}{\sqrt{g_0(x+t)}} dx$$
$$= - \int_{|\delta_n|}^\infty \frac{g_0'(x-t)}{\sqrt{g_0(x-t)}} \frac{g_0'(x+t)}{\sqrt{g_0(x+t)}} dx - \int_0^{|\delta_n|} \frac{g_0'(x-t)}{\sqrt{g_0(x-t)}} \frac{g_0'(x+t)}{\sqrt{g_0(x+t)}} dx$$

$$\leq -\int_{|\delta_n|}^{\infty} \frac{g_0'(x-t)}{\sqrt{g_0(x-t)}} \frac{g_0'(x+t)}{\sqrt{g_0(x+t)}} dx + |\delta_n| O_p(1),$$

where the last step follows from (85). The above combined with (84) and (86) leads to

$$\limsup_n \frac{T_c^2}{2H(\widehat{g}_n, g_0)^2 \delta_n^2}$$

$$\leq \limsup_n \frac{1}{|\delta_n|} \int_0^{|\delta_n|} \left[|\delta_n| O_p(1) + \mathcal{I}_{f_0} - 2\int_{|\delta_n|}^{\infty} \frac{g_0'(x-t)}{\sqrt{g_0(x-t)}} \frac{g_0'(x+t)}{\sqrt{g_0(x+t)}} dx \right] dt$$

$$= O_p(1) \limsup_n |\delta_n| + \mathcal{I}_{f_0}$$

$$\quad - 2\liminf_n \frac{1}{|\delta_n|} \int_0^{|\delta_n|} \int_{|\delta_n|}^{\infty} \frac{g_0'(x-t)}{\sqrt{g_0(x-t)}} \frac{g_0'(x+t)}{\sqrt{g_0(x+t)}} dx dt$$

$$= 0 + \mathcal{I}_{f_0} - 2\liminf_n \int_{|\delta_n|}^{\infty} \frac{\int_0^{|\delta_n|} \frac{g_0'(x+t)}{\sqrt{g_0(x+t)}} \frac{g_0'(x-t)}{\sqrt{g_0(x-t)}} dt}{|\delta_n|} dx. \tag{88}$$

Therefore, an application of Fatou's Lemma and (87) yield

$$\liminf_n \int_{|\delta_n|}^{\infty} \frac{\int_0^{|\delta_n|} \frac{g_0'(x+t)}{\sqrt{g_0(x+t)}} \frac{g_0'(x-t)}{\sqrt{g_0(x-t)}} dt}{|\delta_n|} dx \geq \int_0^{\infty} \frac{g_0'(x)^2}{g_0(x)} dx = \frac{\mathcal{I}_{f_0}}{2}.$$

Thus (88) 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 (83) and thus completes the proof of part B of Theorem 5. $\square$

G.1. Auxilliary lemmas for Theorem 5

Proof of Lemma G.1. The proof is similar to the proof of Theorem 3.2 of Pal *et al.* (2007). Since $\widehat{f}_n$ is piecewise linear, $\widehat{f}_n$ attains its maxima at some order statistic, say $X_{(m)}$. If $m > n/2$, set $m_q = [n/4]$ where $[x]$ is the greatest integer less than or equal to x. For $m \leq n/2$, we let $m_q = [3n/4] + 1$. Set $K_n = m_q$ or $n - m_q$, accordingly as $m > n/2$ or $\leq n/2$. It is easy to see that $n/K_n \to 4$ as $n \to \infty$. Also,

$$\widehat{f}_n(X_{(m)}) \leq \frac{1}{|X_{(m)} - X_{(m_q)}|} \left(1 + \log \frac{\widehat{f}_n(X_{(m)})}{\widehat{f}_n(X_{(m_q)})} \right) \tag{89}$$

by Lemma 3 of Pal *et al.* (2007) (see our Lemma G.2). Now since

$$\sum_{i=1}^{n} \psi_0(X_i - \theta_0) \leq \sum_{i=1}^{n} \widehat{\psi}_n(X_i - \widehat{\theta}_n)$$

$$\leq K_n \log\Big(\widehat{f}_n(X_{(m_q)})\Big) + (n - K_n) \log\Big(\widehat{f}_n(X_{(m)})\Big),$$

$$K_n \log \frac{\widehat{f}_n(X_{(m)})}{\widehat{f}_n(X_{(m_q)})} \leq n\bigg(\log \widehat{f}_n(X_{(m)}) - l_n(\theta_0, \psi_0)/n\bigg). \tag{90}$$

Combining (89) and (90) we obtain that

$$\widehat{f}_n(X_{(m)}) \leq \frac{1}{|X_{(m)} - X_{(m_q)}|}\left(1 + \frac{n}{K_n}\bigg(\log \widehat{f}_n(X_{(m)}) - l_n(\theta_0, \psi_0)/n\bigg)\right)$$

$$= \frac{n/K_n}{|X_{(m)} - X_{(m_q)}|} \log \widehat{f}_n(X_{(m)}) + \frac{1}{|X_{(m)} - X_{(m_q)}|}\left(1 - \frac{l_n(\theta_0, \psi_0)}{K_n}\right).$$

Therefore by Lemma 4 of Pal *et al.* (2007) (see our Lemma G.3),

$$\widehat{f}_n(X_{(m)}) \leq \frac{2n/K_n}{|X_{(m)} - X_{(m_q)}|} \log\left(\frac{2n/K_n}{|X_{(m)} - X_{(m_q)}|}\right)$$

$$+ \frac{2}{|X_{(m)} - X_{(m_q)}|}\left(1 - \frac{l_n(\theta_0, \psi_0)}{K_n}\right)$$

which is finite by our choices of m, m_q and K_n. $\qquad\square$

The following lemmas appear in Pal *et al.* (2007) as Lemma 3 and 4 respectively.

Lemma G.2. *Suppose f is a log-concave density. If $0 < f(x) \leq f(y)$ for $x, y \in \mathbb{R}$, then*

$$f(y) \leq \frac{1 + \log(f(y)/f(x))}{|y - x|}.$$

Lemma G.3. *If $x, c_1, c_2 > 0$ and $x \leq c_1 \log x + c_2$, then $x \leq 2c_1 \log(2c_1) + 2c_2$.*

Appendix H: Technical facts

Below we list some facts which have been used repeatedly in our proofs. We begin with a well-known fact on total variation distance.

Fact 5. *Suppose F and G are two distribution functions with densities f and g, respectively. Then $d_{T,V}(F, G) \leq \sqrt{2}H(f, g)$.*

Fact 6 (Theorem 5.7 (ii) of Shorack (2000)). *Suppose $\{X_n\}_{n \geq 1}$ is a random sequence. If X_n satisfies $X_n \to_p X$ for some random variable X, then there exists a subsequence n_k such that $X_{n_k} \to_{a.s.} X$.*

Fact 7 (Theorem 5.7 (vii) of Shorack (2000)). *Suppose X_n is a sequence of random variables. Then for some random variable X, $X_n \to_p X$ if and only if every subsequence $\{n_k\}_{k \geq 1}$ contains a further subsequence $\{n_r\}_{r \geq 1}$ for which $X_{n_r} \to_{a.s.} X$.*

Fact 8 (Proposition A.18 of Bobkov and Ledoux (2014))**.** *Suppose the density f is supported on an open interval (possibly unbounded). Then F^{-1} is strictly increasing, and $F^{-1}(q_2) - F^{-1}(q_1) = \int_{q_1}^{q_2} dt/f(F^{-1}(t))$ for all $0 < q_1 < q_2 < 1$.*

Fact 9. *Suppose $\mathcal{F}$ is a class of measurable functions h such that $\int h^2 dP_0 < \epsilon^2$ where $\|h\|_\infty \leq M$ for some constant $M > 0$. Then*

$$E\|\mathbb{G}_n\|_{\mathcal{F}} \lesssim J_{[\ ]}(\epsilon, \mathcal{F}, L_2(P_0))\left(1 + \frac{MJ_{[\ ]}(\epsilon, \mathcal{F}, L_2(P_0))}{\epsilon^2\sqrt{n}}\right),$$

where

$$J_{[\ ]}(\epsilon, \mathcal{F}, L_2(P_0)) = \int_0^\epsilon \sqrt{1 + \log N_{[\ ]}(\epsilon', \mathcal{F}, L_2(P_0))}d\epsilon'.$$

Proof. Follows from Theorem 1.7.6 of Van der Vaart and Wellner (1996). $\qquad\square$

The next fact is Pratt's lemma (Pratt, 1960, Theorem 1). We state it here for convenience.

Fact 10. *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 $\int b_n d\mu \to \int b d\mu$.*

Fact 11. *Suppose $(F_n)_{n \geq 1}$ and F are distribution functions satisfying $\|F_n - F\|_\infty \to 0$. Further suppose F has density f and $t \in int(supp(f))$. Then $|F_n^{-1}(t) - F^{-1}(t)| \to 0$.*

Proof. Since F^{-1} is continuous at t, this is essentially Lemma A.5 of Bobkov and Ledoux (2014). $\qquad\square$

The following is a property of integrable functions.

Fact 12 (Exercise 16.18 of Billingsley (2013))**.** *Suppose P is a finite measure on $\mathbb{R}$ and $\int_{\mathbb{R}} |h|dP < \infty$ for some measurable function h. Then for each $\epsilon > 0$, there exists $\sigma > 0$ so that any P-measurable set $\mathcal{B}$ with $P(\mathcal{B}) < \sigma$ satisfies $\int_{\mathcal{B}} |h|dP < \epsilon$.*

The following is a sufficient (and necessary) condition for uniform integrability.

Fact 13 (Exercise 16.19 of Billingsley (2013))**.** *Suppose P is a finite measure on $\mathbb{R}$ and $(h_n)_{n \geq 1}$ is a sequence of P-measurable functions. Then $(h_n)_{n \geq 1}$ is uniformly integrable if and only if (i) $\sup_{n \geq 1} \int |h_n|dP < \infty$ (ii) given any $\epsilon > 0$, there exists $\sigma > 0$ so that any P-measurable set $\mathcal{B}$ with $P(\mathcal{B}) < \sigma$ satisfies $\sup_{n \geq 1} \int_{\mathcal{B}} |h_n|dP < \epsilon$.*

The following fact is a Glivenko cantelli type result for a class of functions $\mathcal{F}_n$ changing with n.

Fact 14. *Suppose $\mathcal{F}_n$ is a class of functions such that $\sup_{f \in \mathcal{F}_n} \|f\|_\infty \leq M_n$. Further suppose for any fixed $\epsilon > 0$, $M_n^2 \sup_Q \log N(\epsilon, \mathcal{F}_n, L_2(Q)) = o(n)$ where the supremum is over all probability measures on $\mathbb{R}$. Then $E\|\mathbb{P}_n - P\|_{\mathcal{F}_n} \to 0$ as $n \to \infty$.*

Proof of Fact 14. The proof is similar to the proof of Theorem 2.4.3 of Van der Vaart and Wellner (1996). Therefore we only highlight the differences. Suppose $X_1, \ldots, X_n \overset{iid}{\sim} P$. Consider also n independent Rademacher random variables $\epsilon_1, \ldots, \epsilon_n$. Using the symmetrization inequality (cf. Lemma 2.3.1 of Van der Vaart and Wellner, 1996,) and Fubini's theorem, one can show that

$$E\|\mathbb{P}_n - P\|_{\mathcal{F}_n} \leq 2E_X \underbrace{E_\epsilon \left\| \frac{1}{n} \sum_{i=1}^n \epsilon_i f(X_i) \right\|_{\mathcal{F}_n}}_{\mathbb{Y}_n(X) \equiv \mathbb{Y}_n(X_1, \ldots, X_n)},$$

where E_X and E_ϵ denote the expectations with respect to P and the law of ϵ_1, respectively. Fixing $\delta > 0$, and using the argument in the proof of Theorem 2.4.3 of Van der Vaart and Wellner (1996), we can show that

$$\mathbb{Y}_n(X) \leq (1 + \log N(\delta, \mathcal{F}_n, L_2(\mathbb{F}_n))^{1/2} M_n \sqrt{6/n} + \delta \qquad (91)$$

where $\mathbb{F}_n$ is the empirical distribution function of $X_1, \ldots, X_n$. Taking $\delta = 1/2$, for sufficiently large n, we have $\mathbb{Y}_n(X) \leq 1$ for any realizations of $X_1, \ldots, X_n$. Therefore $\mathbb{Y}_n(X)$ is a bounded sequence. For any $\delta > 0$, (91) also implies that $\lim_{n \to \infty} \mathbb{Y}_n(X) \leq \delta$. Since δ is arbitrary, this implies $\mathbb{Y}_n(X) \to 0$ as $n \to \infty$ for any realization of $X \equiv X_1, \ldots, X_n$. Therefore, using dominated convergence theorem we conclude that $E_X[\mathbb{Y}_n(X)] \to_n 0$. $\qquad \square$

Fact 15. *Suppose f_0 is a log-concave density with $\mathcal{I}_{f_0} < \infty$. Then $H(f_0(\cdot + y), f_0) = O(|y|)$.*

Proof of Fact 15. Note that

$$H(f_0(\cdot + y), f_0)^2 = \int_{-\infty}^\infty (\sqrt{f_0(x+y)} - \sqrt{f_0(x)})^2 dx$$
$$\leq \int_{-\infty}^\infty \left(\int_{x-|y|}^{x+|y|} \frac{|f_0'(z)|}{2\sqrt{f_0(z)}} dz \right)^2 dx,$$

which, by the Cauchy-Schwarz inequality, is bounded above by

$$\frac{|y|}{4} \int_{-\infty}^\infty \int_{x-|y|}^{x+|y|} \frac{f_0'(z)^2}{f_0(z)} dz dx \overset{(a)}{=} \frac{|y|^2}{2} \int_{-\infty}^\infty \phi_0'(z)^2 f_0(z) dz = |y|^2 \mathcal{I}_{f_0}/2,$$

where (a) follows by Fubini's Theorem. Since $\mathcal{I}_{f_0} < \infty$, the above is of order $O(|y|^2)$. $\qquad \square$

Appendix I: Tuning parameters for Stone and Beran's estimators

Stone's estimator has two tuning parameters d_n and t_n. To find the optimal (d_n, t_n) pair, we implement a grid search on a two dimensional grid. Each point on the grid is of the form (d, t) where $d \in \{10, 20, 30, \ldots, 80\}$, and $t_n \in \{0.10, 0.20, \ldots, 0.60\}$. For each distribution and each sample size, we estimate the efficiency of each pair using one hundred Monte Carlo samples. The optimal pair is the one that maximizes the estimated efficiency. Since Beran's estimator also uses two tuning parameters $b_{c,n}$ and ρ_n, we repeat the same procedure for

finding its optimal tuning parameters. The only difference is that in this case, the scaling parameter is chosen from the grid $\{0.10, 0.20, \ldots, 1.50\}$, and the number of basis functions is allowed to vary within the set $\{10, 20, \ldots, 50\}$. Table 2 and 3 tabulate the optimal tuning parameters that we obtained following the above-mentioned procedure.

TABLE 2
The optimal (d_n, t_n) pair for Stone's estimattor

n	Gaussian	Laplace	Symmetric beta $(r = 2.1)$	Symmetric beta $(r = 4.5)$	Logistic
40	(10, 0.80)	(20, 0.60)	(20, 0.60)	(40, 0.80)	(10, 0.80)
100	(50, 0.80)	(20, 0.50)	(40, 0.50)	(30, 0.60)	(10, 0.80)
200	(50, 0.80)	(20, 0.50)	(40, 0.50)	(50, 0.60)	(10, 0.80)
500	(60, 0.80)	(10, 0.50)	(20, 0.30)	(30, 0.40)	(30, 0.50)

TABLE 3
The optimal $(b_{c,n}, \rho_n)$ pair for Beran's estimattor

n	Gaussian	Laplace	Symmetric beta $(r = 2.1)$	Symmetric beta $(r = 4.5)$	Logistic
40	(10, 1.00)	(40, 0.40)	(10, 0.80)	(40, 1.40)	(10, 1.40)
100	(10, 1.00)	(40, 0.20)	(10, 0.40)	(40, 1.20)	(20, 1.40)
200	(10, 1.00)	(40, 0.20)	(40, 0.60)	(40, 1.00)	(25, 1.00)
500	(10, 0.60)	(40, 0.20)	(40, 0.60)	(35, 0.80)	(30, 1.00)

As mentioned previously, we consider another set of tuning parameters for these nonparametric estimators. These tuning prameters, i.e. the non-optimal tuning parameters, are provided in Table 4 and 5.

TABLE 4
The non-optimal (d_n, t_n) pair for Stone's estimattor

n	Gaussian	Laplace	Symmetric beta $(r = 2.1)$	Symmetric beta $(r = 4.5)$	Logistic
40	(30, 0.50)	(50, 0.50)	(40, 0.50)	(50, 0.50)	(50, 0.50)
100	(30, 0.50)	(50, 0.50)	(50, 0.50)	(50, 0.50)	(50, 0.50)
200	(30, 0.50)	(50, 0.50)	(50, 0.50)	(50, 0.50)	(50, 0.50)
500	(30, 0.50)	(50, 0.50)	(40, 0.50)	(50, 0.50)	(50, 0.50)

TABLE 5
The non-optimal $(b_{c,n}, \rho_n)$ pair for Beran's estimattor

n	Gaussian	Laplace	Symmetric beta $(r = 2.1)$	Symmetric beta $(r = 4.5)$	Logistic
40	(40, 0.20)	(10, 0.40)	(40, 0.20)	(30, 0.20)	(40, 0.20)
100	(40, 0.20)	(10, 1.20)	(40, 0.20)	(35, 0.20)	(40, 0.20)
200	(40, 0.20)	(10, 1.20)	(40, 0.20)	(40, 0.20)	(40, 0.20)
500	(40, 0.20)	(10, 1.20)	(40, 0.20)	(40, 0.20)	(40, 0.20)