Univariate log-concave density estimation with symmetry or modal constraints

Charles R. Doss* and **Jon A. Wellner†**

School of Statistics
University of Minnesota
Minneapolis, MN 55455
e-mail: cdoss@stat.umn.edu
url: http://users.stat.umn.edu/~cdoss/

Department of Statistics, Box 354322
University of Washington
Seattle, WA 98195-4322
e-mail: jaw@stat.washington.edu
url: http://www.stat.washington.edu/jaw/

Abstract: We study nonparametric maximum likelihood estimation of a log-concave density function f_0 which is known to satisfy further constraints, where either (a) the mode m of f_0 is known, or (b) f_0 is known to be symmetric about a fixed point m. We develop asymptotic theory for both constrained log-concave maximum likelihood estimators (MLE's), including consistency, global rates of convergence, and local limit distribution theory. In both cases, we find the MLE's pointwise limit distribution at m (either the known mode or the known center of symmetry) and at a point $x_0 \neq m$. Software to compute the constrained estimators is available in the R package `logcondens.mode`.

The symmetry-constrained MLE is particularly useful in contexts of location estimation. The mode-constrained MLE is useful for mode-regression. The mode-constrained MLE can also be used to form a likelihood ratio test for the location of the mode of f_0. These problems are studied in separate papers. In particular, in a separate paper we show that, under a curvature assumption, the likelihood ratio statistic for the location of the mode can be used for hypothesis tests or confidence intervals that do not depend on either tuning parameters or nuisance parameters.

AMS 2000 subject classifications: Primary 62G07; secondary 62G05, 62G20.
Keywords and phrases: Mode, consistency, convergence rate, empirical processes, convex optimization, log-concave, shape constraints, symmetric.

Received October 2018.

Contents

1. Introduction and overview

The classes of log-concave densities on $\mathbb{R}$ (and on $\mathbb{R}^d$) have great importance in statistics for a variety of reasons including their many natural closure properties, including closure under convolution, affine transformations, convergence in distribution, and marginalization. These classes are also unimodal and serve as important nonparametric generalizations of the class of Gaussian distributions.

Nonparametric estimation in the *unconstrained classes* of log-concave densities has developed rapidly in the past 10–15 years. Existence of maximum likelihood estimators for log-concave densities on $\mathbb{R}$ was provided by Walther (2002), while Pal, Woodroofe and Meyer (2007) established consistency. Dümbgen and Rufibach (2009) gave rates of convergence in certain uniform metrics, and pro-

vided efficient algorithms based on "active set" methods (see also Dümbgen and Rufibach (2011)). Balabdaoui, Rufibach and Wellner (2009) established pointwise limit distribution theory for the MLE's, while Doss and Wellner (2016) established rates of convergence of the MLE in the Hellinger metric. There has also been rapid progress in estimation of log-concave densities on $\mathbb{R}^d$; see e.g. Cule, Samworth and Stewart (2010), Cule and Samworth (2010), Dümbgen, Samworth and Schuhmacher (2011), Seregin and Wellner (2010), and Han and Wellner (2016).

Interesting uses of the unconstrained log-concave MLE's in more complicated models, mostly in mixture modeling and clustering, have been considered by Chang and Walther (2007), Eilers and Borgdorff (2007), Walther (2009), and Cule and Samworth (2010).

On the other hand, for a number of important statistical problems it is of great interest to understand estimation in several important *sub-classes* of the class of all log-concave densities on $\mathbb{R}$.

- For testing that a log-concave density on $\mathbb{R}$ is symmetric about a known point, for example 0, we need know how to estimate the log-concave density both *with* and *without* the constraint of symmetry.
- For the basic problem of estimation of location with a symmetric error density, it is important to know how to estimate a *symmetric* log-concave density with mode (and median and mean) equal to 0.
- For inference about the mode of a log-concave density it is necessary to understand how to estimate a log-concave density with a known mode m (but without the constraint of symmetry).

Once the properties of nonparametric estimators within these sub-classes is understood, then the estimators can be used to develop statistical methods with known properties for other more complex statistical problems. For example: the basic procedures we study here can be viewed as building blocks to be used for, among others:

(a) Testing the hypothesis of symmetry of a log-concave density.
(b) Estimation of the location of a symmetric log-concave density.
(c) Inference about the mode of a log-concave density.
(d) Nonparametric modal regression (as in Chen et al. (2016), but using log-concavity).
(e) Semiparametric estimation in mixture models based on symmetric log-concave distributions; see e.g. Balabdaoui and Doss (2018), Pu and Arias-Castro (2017), and Eilers and Borgdorff (2007).
(f) Modal clustering (as in Chacón (2018a), but using log-concavity).
(g) Estimation of a spherically symmetric multivariate log-concave density, which is pursued in Xu and Samworth (2017).
(h) Inference about the center of an elliptical multivariate distribution based on the assumption of a log-concave underlying shape.

Thus our focus here is on estimation of a log-concave density in two important sub-classes: Let $\mathcal{LC}$ denote the class of all log-concave densities on the real line $\mathbb{R}$.

The two subclasses we study here are:

(1) The class $\mathcal{LC}(m) = \mathcal{LC}_m$ of all log-concave densities with mode a fixed number m.
(2) The class $\mathcal{SLC}(0) = \mathcal{SLC}_0$ of all log-concave densities symmetric at 0.

We let $\hat{f}_n^0$ denote the maximum likelihood estimator of $f_0 \in \mathcal{LC}_m$ based on an i.i.d. sample $X_1, \ldots, X_n$ from f_0; and we let $\hat{g}_n^0$ denote the maximum likelihood estimator of $g_0 \in SLC_0$, based on an i.i.d. sample from g_0.

We rely on the methods and properties developed here for the subclass $\mathcal{LC}(m)$ to derive new inference procedures for the mode in Doss and Wellner (2019). The sub-class $\mathcal{SLC}(0)$ has already been used in Balabdaoui and Doss (2018) to study semiparametric mixture models. The methods developed here for $\mathcal{SLC}(0)$ are also being used in an on-going study by Laha (2019) of efficient estimation of a location parameter in the classical semiparametric symmetric location model with the (very natural) assumption of a symmetric log-concave error distribution. Methodology based on modes or local maxima of nonparametrically estimated functions has seen a resurgence in recent years; see, e.g., Chen et al. (2016), Chen, Genovese and Wasserman (2015), and Qiao and Polonik (2016). A recent survey on estimation and inference for the mode and on mode-based methodology is given by Chacón (2018b).

Thus our main goals here are the following:

(a) To show that the mode-constrained MLE's $\hat{f}_n^0 \in \mathcal{LC}(m)$ and $\hat{g}_n^0 \in \mathcal{SLC}(0)$ exist and to provide useful characterizations thereof.
(b) Establish useful finite-sample properties of $\hat{f}_n^0$ and $\hat{g}_n^0$.
(c) Establish consistency of the mode-constrained and symmetric mode-constrained MLE's with respect to the Hellinger metric.
(d) Establish local rates of convergence of the constrained estimators $\hat{f}_n^0$ and $\hat{g}_n^0$ and establish the (pointwise) asymptotic distributions of the constrained estimators.
(e) Establish global rates of convergence of the constrained estimators.

Here is a brief summary of the paper: In Section 2 we show that the constrained estimators exist and satisfy useful characterizations. Section 3 provides plots of the constrained estimators and provides comparisons to each other and to the unconstrained maximum likelihood estimators $\hat{f}_n \in \mathcal{LC}$. In Section 4 we summarize results concerning consistency and global rates of convergence, while Section 5 addresses local rates of convergence and limiting distributions at fixed points. Section 6 summarizes some problems and difficulties concerning extensions to higher dimensions. All the proofs are given in Sections 7 and 8.

Many of our theorems have parts labeled "A", "B", and "C." In general the "A parts" of results here have been proved by other authors (as noted in the theorem statements), the "B parts" were proved (for the most part) in the University of Washington Ph.D. dissertation of the first author, Doss (2013b). The "C parts" are new findings by the present authors, whose proofs are in some cases (as noted in text near the corresponding results) related to proofs developed by Balabdaoui and Doss (2018).

2. Maximum likelihood estimator finite sample properties: Unconstrained and mode-constrained

2.1. Notation and terminology

Several classes of concave functions will play a central role in this paper. In particular, we let

$$\mathcal{C} := \{\varphi : \ \mathbb{R} \to [-\infty, \infty) \mid \varphi \ \text{is concave, closed, and proper}\} \tag{2.1}$$

and, for any fixed $m \in \mathbb{R}$,

$$\mathcal{C}_m := \{\varphi \in \mathcal{C} \mid \varphi(m) \geq \varphi(x) \ \text{for all} \ x \in \mathbb{R}\} \tag{2.2}$$

is the class of concave functions on $\mathbb{R}$ with mode at m. We also let

$$\mathcal{SC}_0 := \{\varphi \in \mathcal{C}_0 \mid \varphi(-x) = \varphi(x) \ \text{for all} \ x \in \mathbb{R}\}. \tag{2.3}$$

Here proper and closed concave functions are as defined in Rockafellar (1970), pages 24 and 50. We will follow the convention that all concave functions φ are defined on all of $\mathbb{R}$ and take the value $-\infty$ off of their effective domains where $\operatorname{dom}(\varphi) := \{x \ : \ \varphi(x) > -\infty\}$ (Rockafellar (1970), page 40). The classes of unconstrained and constrained log-concave densities are then

$$\mathcal{LC} := \left\{ e^\varphi \ : \ \int e^\varphi d\lambda = 1, \ \ \varphi \in \mathcal{C} \right\},$$

$$\mathcal{LC}_m := \left\{ e^\varphi \ : \ \int e^\varphi d\lambda = 1, \ \ \varphi \in \mathcal{C}_m \right\}, \quad \text{and}$$

$$\mathcal{SLC}_0 := \left\{ e^\varphi \ : \ \int e^\varphi d\lambda = 1, \ \ \varphi \in \mathcal{SC}_0 \right\},$$

where λ is Lebesgue measure on $\mathbb{R}$. We let $X_1, \ldots, X_n$ be the observations, independent and identically distributed with density f_0 with respect to Lebesgue measure. Here we assume throughout that $f_0 \in \mathcal{LC}$ and frequently that $f_0 = e^{\varphi_0} \in \mathcal{LC}_m$ for some $m \in \mathbb{R}$ or $f_0 = e^{\varphi_0} \in \mathcal{SLC}_0$. We let $X_{(1)} < \cdots < X_{(n)}$ denote the order statistics of the X_i's, and write $|X|_{(1)} < \cdots < |X|_{(n)}$ for the order statistics of $|X_1|, \ldots, |X_n|$. We let $\mathbb{P}_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ denote the empirical measure, let $\mathbb{F}_n(x) = n^{-1} \sum_{i=1}^n 1_{(-\infty, x]}(X_i)$ denote the empirical distribution function, and let $\mathbb{G}_n(x) = n^{-1} \sum_{i=1}^n 1_{[0,x]}(|X_i|)$ denote the empirical distribution function of $|X_1|, \ldots, |X_n|$.

We define the log-likelihood criterion function $\Psi_n : \mathcal{C} \to \mathbb{R}$ by

$$\Psi_n(\varphi) = \frac{1}{n} \sum_{i=1}^n \varphi(X_i) - \int_{\mathbb{R}} e^{\varphi(x)} dx = \mathbb{P}_n \varphi - \int_{\mathbb{R}} e^\varphi d\lambda \tag{2.4}$$

where we have used the standard device of including the Lagrange term $\int_{\mathbb{R}} e^{\varphi(x)} dx$ in Ψ_n to avoid the normalization constraints involved in the classes

$\mathcal{LC}_m$ and $\mathcal{SLC}_0$. This is as in Silverman (1982), Dümbgen and Rufibach (2009), and other current literature.

We will denote the unconstrained MLE's of φ_0, f_0, and F_0 by $\widehat{\varphi}_n$, $\widehat{f}_n$, and $\widehat{F}_n$ respectively. The corresponding constrained estimators with mode m and symmetric estimators with mode at 0 will be denoted by $\widehat{\varphi}_n^0$, $\widehat{f}_n^0$, $\widehat{F}_n^0$, and $\widehat{\psi}_n^0$, $\widehat{g}_n^0$, $\widehat{G}_n^0$ respectively. Thus

$$\widehat{\varphi}_n \equiv \underset{\varphi \in \mathcal{C}}{\operatorname{argmax}}\, \Psi_n(\varphi), \quad \widehat{\varphi}_n^0 \equiv \underset{\varphi \in \mathcal{C}_m}{\operatorname{argmax}}\, \Psi_n(\varphi), \quad \text{and} \quad \widehat{\psi}_n^0 \equiv \underset{\psi \in \mathcal{SC}_0}{\operatorname{argmax}}\, \Psi_n(\psi).$$

Before proceeding to results concerning existence and uniqueness of the constrained estimators $\widehat{\varphi}_n^0$ and $\widehat{\psi}_n^0$, we first explain some undesirable properties of "naive" constrained estimators based on the unconstrained MLE's $\widehat{f}_n$ and $\widehat{\varphi}_n$.

2.2. Naive estimators

We can easily construct "naive" estimators under our two classes of constraints. For instance, a naive mode-constrained estimator based on the unconstrained log-concave MLE is $\tilde{f}_n^0(x) = \widehat{f}_n(x - (m - \hat{m}_n))$, where $\hat{m}_n$ is the mode of $\widehat{f}_n$. Then $\tilde{f}_n^0$ indeed has mode m. Let $\tilde{\varphi}_n^0 = \log \tilde{f}_n^0$. Unfortunately, these estimators have quite undesirable properties. For example, when $\varphi_0'(x_0) \neq 0$, we can see that

$$\begin{aligned} n^{2/5}\big(\tilde{\varphi}_n^0(x_0) - \varphi_0(x_0)\big) &= n^{2/5}\big(\widehat{\varphi}_n(x_0 + (\hat{m}_n - m)) - \varphi_0(x_0) - \varphi_0'(x_0)(\hat{m} - m)\big) \\ &\quad + n^{2/5}\varphi_0'(x_0)(\hat{m} - m) \\ &= O_p(1) + n^{1/5}O_p(1) \end{aligned}$$

since $\hat{m}_n - m = O_p(n^{-1/5})$ by Theorem 3.6 of Balabdaoui, Rufibach and Wellner (2009), so the first summand is $O_p(1)$ by Corollary 2.2 of Balabdaoui, Rufibach and Wellner (2009) (and its proof, see their (4.34)). Thus, away from the mode, this naive estimator in fact converges at a slower rate than $n^{-2/5}$.

Similarly, a naive 0-symmetric estimator can be constructed. Let $\tilde{g}_n^s(y) = (\widehat{f}_n(\hat{m}_n + y)\widehat{f}_n(\hat{m}_n - y))^{1/2}$. Then $\tilde{g}_n^s$ is symmetric about its mode 0. (It is not necessarily a bona fide density that integrates to 1, but its integral converges to 1.) Again, unfortunately, a similar analysis as above shows that if $x_0 \neq m$, then

$$n^{2/5}(\log \tilde{g}_n^s(x_0) - \varphi_0(x_0)) = O_p(1) + n^{2/5}\hat{m}_n\varphi_0'(x_0)$$

since $\varphi_0'(x_0) = -\varphi_0'(-x_0)$. Since $n^{1/5}\hat{m}_n = O_p(1)$ and $\varphi_0'(x_0) \neq 0$, we again see that the naive estimator converges at a slower rate than $n^{-2/5}$.

In summary, naive plug-in estimation for the mode and symmetry constraints does not work. The poor performance of these and other "naive" or "plug-in" estimators motivates study of the constrained MLE's, which we now pursue.

2.3. The unconstrained and the constrained MLE's

To develop theory for the mode-constrained estimators $\widehat{\varphi}_n^0$, $\widehat{f}_n^0$, and $\widehat{F}_n^0$ it will be helpful to consider *mode-augmented data* $Z_1, \ldots, Z_N$ with $N = n$ or $n + 1$ as follows:

(1) If $m = X_{(k)}$ for some $k \in \{1, \ldots, n\}$ then $Z_j \equiv X_{(j)}$ for $j \in \{1, \ldots, n\}$ and $N = n$.

(2) If $m \in (X_{(k-1)}, X_k)$ for some $k \in \{1, \ldots, n+1\}$ (where $X_{(0)} \equiv -\infty$ and $X_{(n+1)} \equiv +\infty$), then we define $Z_i \equiv X_{(i)}$ for $i \in \{1, \ldots, k-1\}$, $Z_k \equiv m$, and $Z_i \equiv X_{(i-1)}$ for $i \in \{k+1, \ldots, n+1\}$. In this case $\underline{Z} = (X_{(1)}, \ldots, X_{(k-1)}, m, X_{(k)}, \ldots, X_{(n)}) \in \mathbb{R}^{n+1}$ and $N = n + 1$.

Theorem 2.1. *The following statements hold almost surely when* $X_1, \ldots, X_n$ *are i.i.d. from a density on* $\mathbb{R}$.

A. *(Pal, Woodroofe and Meyer ([2007](#)), Rufibach ([2006](#))) For $n \geq 2$ the (un-constrained) nonparametric MLE $\widehat{\varphi}_n$ exists and is unique. It is linear on all intervals $[X_{(j)}, X_{(j+1)}]$, $j = 1, \ldots, n$. Moreover, $\widehat{\varphi}_n = -\infty$ and $\widehat{f}_n = 0$ on $\mathbb{R} \setminus [X_{(1)}, X_{(n)}]$.*

B. *(Doss ([2013b](#))) For $N \geq 2$ the mode-constrained MLE $\widehat{\varphi}_n^0$ exists and is unique. It is piecewise linear with knots at the Z_i's and domain $[Z_1, Z_N]$. If m is not a data point, then at least one of $(\widehat{\varphi}_n^0)'(m+)$ or $(\widehat{\varphi}_n^0)'(m-)$ is 0.*

C. *The constrained MLE $\widehat{\psi}_n^0 \in \mathcal{SC}_0$ exists for $n \geq 1$ and is unique. It is piecewise linear with knots contained in the set of $2n + 1$ points $-|X|_{(n)}, \ldots, -|X|_{(1)}, 0, |X|_{(1)}, \ldots, |X|_{(n)}$, and is $-\infty$ for $x \notin [-|X|_{(n)}, |X|_{(n)}]$. Furthermore, $(\widehat{\psi}_n^0)'(0\pm) = 0$.*

The previous result shows that the MLE's exist. Unfortunately, there is no closed form expression for the MLE's. However, since they are solutions to optimization problems, they satisfy certain optimality conditions. Thus, the next two theorems we present provide systems of inequalities and equalities that characterize the MLE's.

Theorem 2.2.

A. *(Rufibach ([2006](#)), Dümbgen and Rufibach ([2009](#))) Let $\widehat{\varphi}_n$ be a concave function such that $\{x : \widehat{\varphi}_n(x) > -\infty\} = [X_{(1)}, X_{(n)}]$. Then $\widehat{f}_n = e^{\widehat{\varphi}_n} \in \mathcal{LC}$ is the unconstrained MLE if and only if*

$$\int \Delta(x) d\mathbb{F}_n(x) \leq \int \Delta(x) \exp(\widehat{\varphi}_n(x)) dx = \int \Delta(x) d\widehat{F}_n(x)$$

for any function $\Delta : \mathbb{R} \to \mathbb{R}$ such that $\widehat{\varphi}_n + \lambda\Delta$ is concave for some $\lambda > 0$.

B. *(Doss ([2013b](#))) Suppose that $\widehat{f}_n^0 = e^{\widehat{\varphi}_n^0} \in \mathcal{LC}_m$. Then $\widehat{f}_n^0$ is the MLE over $\mathcal{LC}_m$ if and only if*

$$\int \Delta d\mathbb{F}_n \leq \int \Delta d\widehat{F}_n^0 \tag{2.5}$$

for all Δ such that $\widehat{\varphi}_n^0 + t\Delta \in \mathcal{C}_m$ for some $t > 0$.

C. Suppose that $\widehat{g}_n^0 = e^{\widehat{\psi}_n^0} \in \mathcal{SLC}_0$ and $\widehat{G}_n^0(x) \equiv \int_{-\infty}^x \widehat{g}_n(y)dy$. Then $\widehat{g}_n^0$ is the MLE over $\mathcal{SLC}_0$ if and only if

$$\int \Delta d\mathbb{F}_n \leq \int \Delta d\widehat{G}_n^0 \tag{2.6}$$

for all Δ such that $\widehat{\psi}_n^0 + t\Delta \in \mathcal{SC}_0$ for some $t > 0$.

To state the second characterization theorem for the MLE's, we first introduce some further notation and definitions. For a continuous and piecewise linear function $h : [A, B] \to \mathbb{R}$ we define its *knots* to be

$$\mathcal{S}_n(h) := \{t \in (A, B) : \ h'(t-) \neq h'(t+)\} \cup \{A, B\}.$$

Note that $\widehat{\varphi}_n$, $\widehat{\varphi}_n^0$, and $\widehat{\psi}_n^0$ are all continuous and piecewise linear functions (with $A = X_{(1)}$, $B = X_{(n)}$ in the case of $\widehat{\varphi}_n$ and $\widehat{\varphi}_n^0$, and with $A = -|X|_{(n)}$, $B = |X|_{(n)}$ in the case of $\widehat{\psi}_n$), and we have

$$\mathcal{S}_n(\widehat{\varphi}_n) \subset \{X_{(1)}, X_{(2)}, \ldots, X_{(n)}\},$$
$$\mathcal{S}_n(\widehat{\varphi}_n^0) \subset \{X_{(1)}, X_{(2)}, \ldots, X_{(n)}\},$$
$$\mathcal{S}_n(\widehat{\psi}_n^0) \subset \{-|X|_{(n)}, \ldots, |X|_{(n)}\}.$$

Now suppose that $\widehat{\varphi}_n^0$ is piecewise linear with knots at the (mode-augmented) data, let $m \in \mathbb{R}$, and assume that $\widehat{f}_n^0 \equiv \exp(\widehat{\varphi}_n^0) \in \mathcal{LC}_m$. For $t \in \mathbb{R}$ define

$$
\begin{aligned}
&\mathbb{F}_{n,L}(t) \equiv \int_{(-\infty, t]} d\mathbb{F}_n(y), &&\mathbb{F}_{n,R}(t) \equiv \int_{[t,\infty)} d\mathbb{F}_n(y), \\
&\mathbb{Y}_{n,L}(t) \equiv \int_{X_{(1)}}^t \mathbb{F}_{n,L}(x)dx, &&\mathbb{Y}_{n,R}(t) \equiv \int_t^{X_{(n)}} \mathbb{F}_{n,R}(x)dx, \\
&\widehat{F}_{n,L}^0(t) \equiv \int_{-\infty}^t \widehat{f}_n^0(y)dy, &&\widehat{F}_{n,R}^0(t) \equiv \int_t^{X_{(n)}} \widehat{f}_n^0(y)dy, \\
&\widehat{H}_{n,L}^0(t) \equiv \int_{X_{(1)}}^t \widehat{F}_{n,L}^0(x)dx, &&\widehat{H}_{n,R}^0(t) \equiv \int_t^{X_{(n)}} \widehat{F}_{n,R}^0(x)dx.
\end{aligned} \tag{2.7}
$$

Definition 2.3. With m considered as a possible knot of $\widehat{\varphi}_n^0$ we say that m is a *left knot* (or LK) if $(\widehat{\varphi}_n^0)'(m-) > 0$ and that m is a *right knot* (or RK) if $(\widehat{\varphi}_n^0)'(m+) < 0$. We say that m is *not a knot* (or NK) if $(\widehat{\varphi}_n^0)'(m) = 0$. All other knots are considered to be left knots (LKs) or right knots (RKs) depending on whether they are strictly smaller or strictly larger than m.

Theorem 2.4.

A. *(Rufibach ([2006](#)), Dümbgen and Rufibach ([2009](#))) Let $\widehat{F}_n(x) \equiv \int_{-\infty}^x e^{\widehat{\varphi}_n(y)}dy$, and assume further that $\widehat{f}_n = e^{\widehat{\varphi}_n} \in \mathcal{LC}$. Then $\widehat{f}_n$ is the MLE in $\mathcal{LC}$ if and only if*

$$\widehat{H}_n(t) \equiv \int_{X_{(1)}}^t \widehat{F}_n(y)dy \leq \int_{X_{(1)}}^t \mathbb{F}_n(y)dy \equiv \mathbb{Y}_n(t) \quad for \ all \ \ t \in \mathbb{R}$$

with equality if $t \in \mathcal{S}_n(\widehat{\varphi}_n)$.

B. *(Doss, 2013b) With the notation in (2.7), $\widehat{f}_n^0 = e^{\widehat{\varphi}_n^0}$ is the MLE of $f_0 \in \mathcal{LC}_m$ if and only if*

$$\widehat{H}_{n,L}^0(t) \leq \mathbb{Y}_{n,L}(t) \quad for \quad X_{(1)} \leq t \leq m \tag{2.8}$$

$$\widehat{H}_{n,R}^0(t) \leq \mathbb{Y}_{n,R}(t) \quad for \quad m \leq t \leq X_{(n)} \tag{2.9}$$

with equality in (2.8) if t is a left knot of $\widehat{\varphi}_n^0$ and equality in (2.9) if t is a right knot of $\widehat{\varphi}_n^0$.

C. *$\widehat{g}_n^0 = e^{\widehat{\psi}_n^0} \in \mathcal{SLC}_0$ is the MLE if and only if $\widehat{g}_n^+ \equiv 2\widehat{g}_n^0$ satisfies, with $\widehat{G}_{n,R}^+(x) \equiv \int_x^{|X|_{(n)}} \widehat{g}_n^+(y)dy$ and $\mathbb{F}_{n,R}^+(x) \equiv n^{-1}\sum_{i=1}^n 1\{|X|_{(i)} \geq x\}$,*

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

Remark 2.5. The conditions (2.8) and (2.9) only involve data from the left and right sides of m, and hence are separate characterizations in a sense. But they are coupled by way of the (global) constraint $\widehat{F}_n^0(X_{(n)}) = 1$ (or, equivalently, $\widehat{\varphi}_n^0 \in \mathcal{C}_m$) which involves the data on *both* sides of m.

Remark 2.6. The "C parts" of Theorems 2.1, 2.2, and 2.4 will be proved here in detail via methods similar to those introduced briefly in Balabdaoui and Doss (2018) in the course of a study of two-component mixture models based on symmetric log-concave components.

These characterization theorems have two important corollaries. (Recall that $\mathbb{G}_n$ denotes the empirical distribution function of the $|X_i|$'s.)

Corollary 2.7 (MLE's related to $\mathbb{F}_n$ at knot points)**.** *Each of the following holds almost surely.*

A. *$\mathbb{F}_n - n^{-1} \leq \widehat{F}_n \leq \mathbb{F}_n$ on $\mathcal{S}_n(\widehat{\varphi}_n)$.*
B. *$\mathbb{F}_n - n^{-1} \leq \widehat{F}_n^0 \leq \mathbb{F}_n$ on $\mathcal{S}_n(\widehat{\varphi}_n^0) \setminus \{m\}$.*
C. *$\mathbb{G}_n - n^{-1} \leq \widehat{G}_n^+ \leq \mathbb{G}_n$ on $\mathcal{S}_n(\widehat{\psi}_n^0) \cap [0, |X|_{(n)}]$.*

Now for any distribution function F on $\mathbb{R}$ let $\mu(F) \equiv \int x dF(x)$ and $\mathrm{Var}(F) = \int(x - \mu(F))^2 dF(x)$.

Corollary 2.8 (Mean and variance inequalities)**.**

A. *$\mu(\widehat{F}_n) = \mu(\mathbb{F}_n)$ and $\mathrm{Var}(\widehat{F}_n) \leq \mathrm{Var}(\mathbb{F}_n)$.*
B. *$\mu(\widehat{G}_n^0) = 0$ and $\mathrm{Var}(\widehat{G}_n^0) \leq \int x^2 d\mathbb{F}_n(x)$.*

Because $\Delta_{\pm}(x) = \pm x$ does not have mode m, and because $-(x - \mu)^2$ only has mode m if $\mu = m$, we cannot make comparisons between the mean and variances of $\mathbb{F}_n$ and $\widehat{F}_n^0$.

3. Hellinger consistency and rates

Pal, Woodroofe and Meyer (2007) showed that the unconstrained MLE's $\{\widehat{f}_n\}$ are a.s. consistent in the Hellinger metric H where $H^2(p, q) \equiv (1/2)\int\{\sqrt{p} - $

$\sqrt{q}\}^2 d\lambda$, and their methods also yield consistency for the MLE's over any sub-class $\mathcal{S} \subset \mathcal{LC}$ for which the MLE's $\{\widehat{g}_n\}$ exist and satisfy

$$\sup_n \sup_x \log \widehat{g}_n(x) < \infty \quad \text{a.s.}.$$

This nicely includes the subclass $\mathcal{S} = \mathcal{LC}_m$ when $f_0 \in \mathcal{LC}_m$; i.e. the mode m has been correctly specified. Further consistency results are due to Cule and Samworth (2010), Rufibach (2006), and Dümbgen and Rufibach (2009). To the best of our knowledge, this is the first treatment of the consistency and global rate properties of the constrained estimators.

Theorem 3.1 (Hellinger consistency and rates of convergence)**.**

A. *(Doss and Wellner (2016)) If $f_0 \in \mathcal{LC}$, then $H(\widehat{f}_n, f_0) = O_p(n^{-2/5})$.*
B. *(Doss (2013b)) If $f_0 \in \mathcal{LC}_m$, then $H(\widehat{f}_n^0, f_0) = O_p(n^{-2/5})$.*
C. *(Balabdaoui and Doss (2018)) If $f_0 \in \mathcal{SLC}_0$, then $H(\widehat{g}_n^0, f_0) = O_p(n^{-2/5})$.*

Remark 3.2. Kim and Samworth (2016) extend Part A of Theorem 3.1 by upper bounding the maximal risk of $\widehat{f}_n$: their Theorem 5 implies that $\sup_{f \in \mathcal{LC}} E_f H^2(\widehat{f}_n, f)$ is $O(n^{-4/5})$ (considering squared Hellinger rather than Hellinger distance). They also provide a matching lower bound: their Theorem 1 implies

$$\inf_{\tilde{f}_n} \sup_{f \in \mathcal{LC}} E_f H^2(\tilde{f}_n, f) \geq cn^{-4/5},$$

for some $c > 0$, where the infimum is over all (measurable) estimators $\tilde{f}_n$ of f. Neither upper nor lower bounds for the (Hellinger) minimax risk are known for either of the constrained density classes we consider in the present paper, although we conjecture that $n^{-4/5}$ is the minimax rate of convergence in both cases.

Remark 3.3. When $f_0 \in \mathcal{LC} \setminus \mathcal{LC}_m$, then we can show that $H^2(\widehat{f}_n^0, f_0^*) \to_{a.s.} 0$ where f_0^* satisfies

$$K(f_0, f_0^*) = \inf_{g \in \mathcal{LC}_m} K(f_0, g).$$

Similarly, when $f_0 \in \mathcal{LC} \setminus \mathcal{SLC}_0$, then we can show that $H^2(\widehat{g}_n^0, g_0^*) \to_{a.s.} 0$ where g_0^* satisfies

$$K(f_0, g_0^*) = \inf_{g \in \mathcal{SLC}_0} K(f_0, g);$$

but we will not pursue this here since our goal in this paper is to understand the null hypothesis (or correctly specified) behavior of the constrained estimators $\widehat{\varphi}_n^0$, $\widehat{f}_n^0$, and $\widehat{F}_n^0$. See Doss and Wellner (2019) for some initial steps concerning the power of a likelihood ratio test based on $2 \log \lambda_n$ when $f_0 \in \mathcal{LC} \setminus \mathcal{LC}_m$.

In addition to considering Hellinger distance, one can consider the sup norm (on compact sets) as a metric for global convergence. It turns out that the proofs in Doss and Wellner (2019) rely crucially on knowing the rate of sup-norm convergence for $\widehat{\varphi}_n^0$ (as well as for $\widehat{\varphi}_n$). Thus we study the sup-norm rate of convergence for $\widehat{\varphi}_n^0$ in that paper. In Theorem 4.1 of that paper we find, when

the true log-density satisfies a Hölder condition of order 2 and $\varphi_0^{(2)}(m) < 0$, that the rate of convergence is $(\log n/n)^{2/5}$ on compact sets interior to the support of f_0.

4. Local limit processes and limiting distributions at fixed points

Our goal in this section is to describe the limiting distributions of our estimators, both unconstrained and constrained, at fixed points x_0 (and m and 0) at which the true density f_0 satisfies a curvature condition. We also want to compare and contrast the behavior of the three different estimators.

4.1. The limit processes, unconstrained and constrained

We first need to introduce the local limit proceses which are needed to treat the local (at a single point or in a neighorhood of a point) limiting distributions of the estimators, unconstrained and constrained. For all of our estimators (including the unconstrained and the two different mode-constrained estimators), the limit distributions are not Gaussian. Rather, they are defined in terms of so-called *invelope* processes of integrated Brownian motion. We first recall the invelope process related to the limit distribution for the unconstrained estimators; this process was first presented and studied in Groeneboom, Jongbloed and Wellner (2001a) (and shown to yield the limit distribution in several convex function estimation problems in Groeneboom, Jongbloed and Wellner (2001b)). Let W be a two-sided standard Brownian motion starting at 0 and for any $t \in \mathbb{R}$ let

$$X(t) = W(t) - 4t^3, \quad Y(t) = \int_0^t X(s)ds = \int_0^t W(s)\,ds - t^4. \quad (4.1)$$

Theorem 4.1 (Groeneboom, Jongbloed and Wellner (2001a)). *Let W, X, and Y be as in* (4.1). *Then there exists an almost surely uniquely defined random continuous function H satisfying the following conditions:*

(i) *The function H is everywhere below Y:*

$$H(t) \leq Y(t) \quad \text{for all } t \in \mathbb{R}.$$

(ii) *H has a concave second derivative.*
(iii) *H satisfies*

$$\int_{-\infty}^\infty (H(t) - Y(t))d(H^{(3)})(t) = 0.$$

The random variables $H^{(2)}(0)$ and $H^{(3)}(0)$ give the universal component of the limit distribution of $\widehat{f}_n(x_0)$ and $(\widehat{f}_n)'(x_0)$; see Theorem 4.5, below.

Theorem 4.1 concerns a process H, related to the unconstrained concave estimation problem. In the mode constrained estimation problem, $f_0 \in \mathcal{LC}_m$, instead of having one process we have two, one for the left-hand side of 0 (negative

axis) and one for the right-hand side of 0 (positive axis). (Here, 0 corresponds to the mode m, by a translation.) The definitions of the left- and right-hand processes depend on a random starting point for the corresponding integrals involved, which we will eventually denote τ_L and τ_R (this is made clear in (4.3)–(4.5), below). To define τ_L and τ_R, we must define rigorously the possible 'bend points' of $\widehat{\varphi}^0$. To describe the situation exactly, we also will define 'bend points' τ_+^0 and τ_-^0, satisfying $\tau_+^0 \leq \tau_R$ and $\tau_-^0 \geq \tau_L$, where the inequality may or may not be strict; these bend points arise in (4.10) below. For a concave function g, we let $g'(\cdot-)$ and $g'(\cdot+)$ be the left and right derivatives, respectively (which are always well defined).

Theorem 4.2. *Assume that $\{H_L(t) : t \leq 0\}$ and $\{H_R(t) : t \geq 0\}$ are random processes with concave second derivatives so that $\widehat{\varphi}^0(t) \equiv H_L^{(2)}(t)1_{(-\infty,0)}(t) + H_R^{(2)}(t)1_{[0,\infty)}(t)$ satisfies $\widehat{\varphi}^0 \in \mathcal{C}_0$. Define the 'bend points' $\widehat{S}^0$ by*

$$(\widehat{S}^0(\widehat{\varphi}^0))^c \equiv (\widehat{S}^0)^c := \left\{ t \in \mathbb{R} : (\widehat{\varphi}^0)^{(2)}(t\pm) = 0 \right\}. \tag{4.2}$$

Next, define

$$\tau_-^0(\widehat{\varphi}^0) \equiv \tau_-^0 = \sup \left\{ t \in \widehat{S}^0 : (\widehat{\varphi}^0)'((t-\varepsilon)-) > 0 \text{ for all } \varepsilon > 0 \right\}, \tag{4.3}$$

$$\tau_+^0(\widehat{\varphi}^0) \equiv \tau_+^0 = \inf \left\{ t \in \widehat{S}^0 : (\widehat{\varphi}^0)'((t+\varepsilon)+) < 0 \text{ for all } \varepsilon > 0 \right\}, \tag{4.4}$$

$$\tau_L = \sup \left(\widehat{S}^0 \cap (-\infty, 0) \right) \text{ and } \tau_R = \inf \left(\widehat{S}^0 \cap (0, \infty) \right). \tag{4.5}$$

Let W be a standard two-sided Brownian motion with $W(0) = 0$, and for $t \in \mathbb{R}$ let

$$X(t) = W(t) - 4t^3,$$

$$Y_L(t) = \int_t^{\tau_L} \int_u^{\tau_L} dX(v)du \quad \text{and} \quad Y_R(t) = \int_{\tau_R}^t \int_{\tau_R}^u dX(v)du. \tag{4.6}$$

With these definitions, we assume that:

(i) $-\infty < \tau_L \leq 0$ *and* $0 \leq \tau_R < \infty$ *and*

$$\int_{\tau_L}^{\tau_R} (\widehat{\varphi}^0(v)dv - dX(v)) = 0. \tag{4.7}$$

(ii)

$$H_L(t) - Y_L(t) \leq 0 \text{ for } t \leq 0, \tag{4.8}$$

$$H_R(t) - Y_R(t) \leq 0 \text{ for } t \geq 0, \tag{4.9}$$

(iii)

$$\int_{(-\infty,\tau_-^0]} (H_L(u) - Y_L(u))d(\widehat{\varphi}^0)'(u) = 0 = \int_{[\tau_+^0,\infty)} (H_R(u) - Y_R(u))d(\widehat{\varphi}^0)'(u). \tag{4.10}$$

Then, H_L and H_R are unique, as are τ_L and τ_R.

Theorem 4.2 shows that processes with the given properties are unique; that they exist follows from the proofs of Theorem 4.8 and Theorem 4.9, which show that H_L and H_R exist since they are limit versions of certain finite sample processes $(\widehat{H}^{\varphi}_{n,L}, \widehat{H}^{\varphi}_{n,R})$.

If $\widehat{\varphi}^0$ is, in fact, piecewise linear, then τ^0_- is just the last knot point τ of $\widehat{\varphi}^0$ with $(\widehat{\varphi}^0)'(\tau-) > 0$. By Theorem 23.1 of Rockafellar (1970), a finite, concave function on $\mathbb{R}$ such as $\widehat{\varphi}^0$ has well-defined right- and left-derivatives at all of $\mathbb{R}$; the specification of left- and right- derivatives in the definitions of τ^0_- and τ^0_+ are for concreteness but not necessary since we consider all $\varepsilon > 0$.

The distinction between τ_R and τ^0_+ depends only on the behavior of $\widehat{\varphi}^0$ at 0, and can be understood by considering the case where the infima in (4.4) and in (4.5) are actually minima (the infima are attained). In that case, we see that τ^0_+ can be thought of as the smallest "right-knot" in the sense that $\widehat{\varphi}^0$ has a strictly negative slope to the right of τ^0_+. And τ_R can be thought of as the smallest positive knot. Note that (by concavity) all positive knots are right-knots, so that $\tau_R \geq \tau^0_+$. Note that the infimum defining τ_R in (4.5) is taken over knots that are strictly larger than 0, so that (when the infima are attained) we have $\tau_R > 0$. On the other hand, if 0 is a right-knot then $\tau^0_+ = 0$, so that then τ_R and τ^0_+ are distinct. If 0 is not a right-knot, then we will have $\tau^0_+ = \tau_R$. These statements are slightly complicated by the fact that τ^0_+ and τ_R are defined as infima rather than minima, but the intuitive differences are captured by the previous description. Corresponding statements hold for τ_L and τ^0_-.

The distinction between the two sets of knots pairs is important because many of our arguments depend on constructing "perturbations" of $\widehat{\varphi}^0$, and we can use different types of perturbations at each pair. This means that the different knot pairs have different properties: if we replace τ_L, τ_R by τ^0_-, τ^0_+ in (4.7), then that display may not hold, and similarly, if we replace τ^0_-, τ^0_+ by τ_L, τ_R in (4.10), then that display may not hold. The following lemma holds for τ^0_-, τ^0_+ but not necessarily for τ_L, τ_R.

Lemma 4.3. *With the definitions and assumptions as in Theorem 4.2,*

$$(\widehat{\varphi}^0)'(t) = 0 \text{ for } t \in (\tau^0_-, \tau^0_+). \tag{4.11}$$

Now we introduce the appropriate limit processes for the symmetric about 0 mode-constrained estimators. The characterization is similar to that for the mode-constrained (but not symmetric) processes, but since it is defined only on $[0, \infty)$ the processes are not the same.

Theorem 4.4. *Assume H^+ is a random process on $[0, \infty)$, and assume that $\widehat{\psi}^0 \equiv (H^+)^{(2)} \in \mathcal{C}_0$. Define*

$$(\widehat{S}^+)^c \equiv S(\widehat{\psi}^0)^c = \left\{ t \geq 0 : (\widehat{\psi}^0)^{(2)}(t\pm) = 0 \right\} \setminus \{0\},$$

$$\tau^+_+ = \inf \left\{ t \in \widehat{S}^+ : (\widehat{\psi}^0)'(t + \varepsilon+) < 0 \text{ for all } \varepsilon > 0 \right\}, \qquad and$$

$$\tau^+_R = \inf \widehat{S}^+ \cap (0, \infty).$$

For $t \geq 0$, let $W(t)$ be a one-sided standard Brownian motion with $W(0) = 0$, let $X(t) = W(t) - 4t^3$, and let

$$Y^+(t) = \int_{\tau_R^+}^t \int_{\tau_R^+}^u dX(v)du.$$

Suppose that $\tau_R^+ < \infty$ and

(i)

$$\int_0^{\tau_R^+} (\widehat{\psi}^0(u)du - dX(u)) = 0, \qquad (4.12)$$

(ii)

$$H^+(t) - Y^+(t) \leq 0 \quad \text{for all } t \geq 0, \qquad (4.13)$$

(iii)

$$\int_{[\tau_+^+, \infty)} (H^+ - Y^+)d(\widehat{\psi}^0)' = 0. \qquad (4.14)$$

Then H^+ is unique.

4.2. Unconstrained and constrained pointwise limit theory at $x_0 \neq m$

The two main limit theorems below will concern the limiting distributions of our estimators and their derivatives. Recall, we assume that $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f_0 = e^{\varphi_0}$, where f_0 is a non-degenerate density on $\mathbb{R}$. The three sets of estimators of f_0, φ_0, f_0', and φ_0' to be considered are:

A $\widehat{f}_n$, $\widehat{\varphi}_n$, $(\widehat{f}_n)'$, and $(\widehat{\varphi}_n)'$.
B $\widehat{f}_n^0$, $\widehat{\varphi}_n^0$, $(\widehat{f}_n^0)'$, and $(\widehat{\varphi}_n^0)'$.
C $\widehat{g}_n^0$, $\widehat{\psi}_n^0$, $(\widehat{g}_n^0)'$, and $(\widehat{\psi}_n^0)'$.

Then the corresponding curvature assumptions are:

Curvature Assumption 1. $\varphi_0''(x_0) < 0$, where $x_0 \in \text{int } \{x : f_0(x) > 0\}$.
Curvature Assumption 2a. $\varphi_0''(x_0) < 0$ with $x_0 \neq m$ and $x_0 \in \text{int } \{x : f_0(x) > 0\}$.
Curvature Assumption 2b. $\varphi_0''(m) < 0$.

Note that Hall ([1984](#)) shows that Curvature Assumption 2b holds for the class of symmetric α-stable densities on $\mathbb{R}$ for all $0 < \alpha < 2$. Assumption 1 will be used for the estimators in A, whereas for the estimators in B and C we will use Assumptions 2a and 2b. In all three cases we assume $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f_0$.

To state our theorem we first define some constants as follows:

$$c(x_0, \varphi_0) = \left(\frac{f_0(x_0)^3 |\varphi_0^{(2)}(x_0)|}{4!} \right)^{1/5}, \qquad d(x_0, \varphi_0) = \left(\frac{f_0(x_0)^4 |\varphi_0^{(2)}(x_0)|^3}{(4!)^3} \right)^{1/5}$$

$$C(x_0, \varphi_0) = \left(\frac{|\varphi_0^{(2)}(x_0)|}{f_0(x_0)^2 4!} \right)^{1/5}, \qquad D(x_0, \varphi_0) = \left(\frac{|\varphi_0^{(2)}(x_0)|^3}{f_0(x_0)(4!)^3} \right)^{1/5}.$$

Theorem 4.5 (Limiting distributions at a fixed point $x_0 \neq m$)**.**

A. *Balabdaoui, Rufibach and Wellner (2009). Suppose that $f_0 \in \mathcal{LC}$ and that the Curvature Assumption 1 holds at x_0. Then with H as in Theorem 4.1 and $\widehat{\varphi} \equiv H''$,*

$$\begin{pmatrix} n^{2/5}(\widehat{f}_n(x_0) - f_0(x_0)) \\ n^{1/5}(\widehat{f}'_n(x_0) - f'_0(x_0)) \\ n^{2/5}(\widehat{\varphi}_n(x_0) - \varphi_0(x_0)) \\ n^{1/5}(\widehat{\varphi}'_n(x_0) - \varphi'_0(x_0)) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} c(x_0, \varphi_0)\widehat{\varphi}(0) \\ d(x_0, \varphi_0)\widehat{\varphi}'(0) \\ C(x_0, \varphi_0)\widehat{\varphi}(0) \\ D(x_0, \varphi_0)\widehat{\varphi}'(0) \end{pmatrix}, \qquad (4.15)$$

B. *Suppose that $f_0 \in \mathcal{LC}_m$ and that the Curvature Assumption 2a holds at x_0 with $x_0 \neq m$. Then (4.15) continues to hold with $\widehat{f}_n$, $\widehat{\varphi}_n$, $(\widehat{f}_n)'$, and $(\widehat{\varphi}_n)'$ replaced by $\widehat{f}_n^0$, $\widehat{\varphi}_n^0$, $(\widehat{f}_n^0)'$, and $(\widehat{\varphi}_n^0)'$.*
C. *Suppose that $f_0 \in \mathcal{SLC}_0$ and that the Curvature Assumption 2a holds at x_0 with $x_0 \neq 0$. Then (4.15) continues to hold with $\widehat{f}_n$, $\widehat{\varphi}_n$, $(\widehat{f}_n)'$, and $(\widehat{\varphi}_n)'$ replaced by $\widehat{g}_n^0$, $\widehat{\psi}_n^0$, $(\widehat{g}_n^0)'$, and $(\widehat{\psi}_n^0)'$; and $c(x_0, \varphi_0)$, $d(x_0, \varphi_0)$, $C(x_0, \varphi_0)$, and $D(x_0, \varphi_0)$ replaced by $2^{-2/5}c(x_0, \varphi_0)$, $2^{-1/5}d(x_0, \varphi_0)$, $2^{-2/5}C(x_0, \varphi_0)$, and $2^{-1/5}D(x_0, \varphi_0)$.*

Remark 4.6. (i) Comparing the MLE's for $\mathcal{LC}$ and $\mathcal{LC}_m$: Note that the limiting distributions in A and B at a point $x_0 \neq m$ are the same. At a fixed point $x_0 \neq m$, the constraint that the mode is known does not help in estimating the function at x_0. As we will see below, this picture changes when $x_0 = m$.

(ii) Note that the rate of convergence of $\widehat{f}_n^0$ and $\widehat{g}_n^0$ as $x_0 \neq m$ (or $x_0 \neq 0$ in the case of $\widehat{g}_n^0$) is $n^{-2/5}$ in contrast to the $n^{-1/5}$ rate achieved by the naive estimators discussed in Subsection 2.2.

(iii) Comparing the MLE's for $\mathcal{LC}$ and $\mathcal{LC}_m$ with the MLE for $\mathcal{SLC}_m$: The limiting distributions for the symmetric log-concave class $\mathcal{SLC}$ in C are smaller than the limiting distributions of the MLE's for the possibly asymmetric log-concave classes $\mathcal{LC}$ and $\mathcal{LC}_m$ by a factor of $2^{-2/5} \approx .757858\ldots$ for the functions themselves and by a factor of $2^{-1/5} \approx .870551\ldots$ for the derivatives of the functions. Thus the symmetry constraint substantially reduces the variability of the estimators (see also Figure 3).

4.3. Mode-constrained and symmetry-constrained pointwise limit theory at $x_0 = m$

The limit distribution of the mode-constrained estimators at a point x_0 depends on whether $x_0 = m$ or $x_0 \neq m$. In the latter case the asymptotics are the same as the unconstrained estimator, but in the former case they depend on the mode-constrained limit process.

Theorem 4.7 (Limiting distributions at m).

B. *Let Curvature Assumption 2b hold. Let constants $c(x_0, \varphi_0), d(x_0, \varphi_0)$, $C(x_0, \varphi_0)$, and $D(x_0, \varphi_0)$ be as given in Theorem 4.5. Let $\widehat{\varphi}^0$ be as in Theorem 4.2. Then*

$$\begin{pmatrix} n^{2/5}(\widehat{f}_n^0(m) - f_0(m)) \\ n^{1/5}((\widehat{f}_n^0)'(m) - f_0'(m)) \\ n^{2/5}(\widehat{\varphi}_n^0(m) - \varphi_0(m)) \\ n^{1/5}((\widehat{\varphi}_n^0)'(m) - \varphi_0'(m)) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} c(m, \varphi_0)\widehat{\varphi}^0(0) \\ d(m, \varphi_0)(\widehat{\varphi}^0)'(0) \\ C(m, \varphi_0)\widehat{\varphi}^0(0) \\ D(m, \varphi_0)(\widehat{\varphi}^0)'(0) \end{pmatrix}. \qquad (4.16)$$

C. *Suppose that $f_0 \in \mathcal{SLC}_0$ and that the Curvature Assumption 2b holds at $m = 0$. Then*

$$\begin{pmatrix} n^{2/5}(\widehat{g}_n^0(0) - f_0(0)) \\ n^{2/5}(\widehat{\psi}_n^0(0) - \psi_0(0)) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} 2^{-2/5}c(0, \varphi_0)\widehat{\psi}^0(0) \\ 2^{-2/5}C(0, \varphi_0)\widehat{\psi}^0(0) \end{pmatrix} \qquad (4.17)$$

where $\widehat{\psi}^0$ is given by Theorem 4.4.

We label the parts of Theorem 4.7 as "B" and "C" to be in parallel with the labeling in Theorem 4.5. Notice that for the symmetric MLE, $((\widehat{g}_n^0)'(0)$ and $((\widehat{\psi}_n^0)'(0)$ are always both equal to 0, the value of $f_0'(0)$ and of $\varphi_0'(0)$, so we do not state a limit theorem for these estimators. Theorems 4.5 and 4.7 follow from more general theorems about the estimators not just at x_0 but in local $n^{-1/5}$ neighborhoods of x_0, stated below as Theorems 4.8 and 4.9. The "local neighborhood" Theorem 4.8 is the version from which we can derive the limit distribution of the mode likelihood ratio statistic $2 \log \lambda_n$ studied in Doss and Wellner (2019). Monte Carlo estimates of the distribution functions of $\widehat{\varphi}'(0)$ and of $(\widehat{\varphi}^0)'(0)$ are presented in Figure 1 (left plot). Note that $(\widehat{\varphi}^0)'(0)$ is stochastically smaller than $\widehat{\varphi}'(0)$.

4.4. *Local process limit theory, mode- and symmetry-constrained*

Here we state the local process limit theorems behind Theorem 4.7 B, where $x_0 = m$. In this subsection we will formulate a more general version of that theorem which applies to our estimators in $n^{-1/5}$ neighborhoods of m.

Recall the definition of Y in (4.1). Now, for positive numbers a and σ, Let

$$Y_{a,\sigma}(t) \equiv \sigma \int_0^t W(s)ds - at^4 \stackrel{d}{=} \sigma(\sigma/a)^{3/5}Y((a/\sigma)^{2/5}t), \qquad (4.18)$$

$$Y_{a,\sigma}^{(1)}(t) = \sigma W(t) - 4at^3 \stackrel{d}{=} \sigma(a/\sigma)^{1/5}Y^{(1)}((a/\sigma)^{2/5}t). \qquad (4.19)$$

Let $H_{a,\sigma}, H_{L,a,\sigma}$, and $H_{R,a,\sigma}$ denote the unconstrained and mode-constrained left- and right-processes for $Y_{a,\sigma}$. Then

$$H_{a,\sigma}(t) \stackrel{d}{=} \sigma(\sigma/a)^{3/5}H((a/\sigma)^{2/5}t),$$

$$H_{a,\sigma}^{(1)}(t) \stackrel{d}{=} \sigma(\sigma/a)^{1/5}H^{(1)}((a/\sigma)^{2/5}t),$$

and

$$\widehat{\varphi}_{a,\sigma} = H_{a,\sigma}^{(2)} \stackrel{d}{=} \sigma^{4/5}a^{1/5}H^{(2)}((a/\sigma)^{2/5}\cdot) \tag{4.20}$$

Identical scaling relationships hold for $H_{L,a,\sigma}, H_{R,a,\sigma}$, and the corresponding derivatives, including $\widehat{\varphi}_{a,\sigma}^0 \equiv H_{R,a,\sigma}^{(2)}$:

$$\widehat{\varphi}_{a,\sigma}^0 = H_{R,a,\sigma}^{(2)} \stackrel{d}{=} \sigma^{4/5}a^{1/5}H_R^{(2)}((a/\sigma)^{2/5}\cdot). \tag{4.21}$$

Theorem 4.8. *Let Curvature Assumption 2b hold. Let H be as in Theorem 4.1 and let $\widehat{\varphi} \equiv H''$ and let $\widehat{\varphi}^0$ be as in Theorem 4.2. Let $\sigma \equiv 1/\sqrt{f_0(m)}$ and $a = |\varphi_0^{(2)}(m)|/4!$. Then*

$$\begin{pmatrix} n^{2/5}(\widehat{\varphi}_n(m + n^{-1/5}t) - \varphi_0(m)) \\ n^{1/5}(\widehat{\varphi}_n'(m + n^{-1/5}t) - \varphi_0'(m)) \\ n^{2/5}(\widehat{\varphi}_n^0(m + n^{-1/5}t) - \varphi_0(m)) \\ n^{1/5}((\widehat{\varphi}_n^0)'(m + n^{-1/5}t) - \varphi_0'(m)) \end{pmatrix} \stackrel{d}{\to} \begin{pmatrix} \widehat{\varphi}_{a,\sigma}(t) \\ \widehat{\varphi}_{a,\sigma}'(t) \\ \widehat{\varphi}_{a,\sigma}^0(t) \\ (\widehat{\varphi}_{a,\sigma}^0)'(t) \end{pmatrix}$$

as processes in $(\mathcal{C}_\infty \times \mathcal{D}_\infty)^2$, where $\mathcal{C}_\infty$ is the set of continuous functions on $(-\infty, \infty)$ with the topology of uniform convergence on compact sets and $\mathcal{D}_\infty$ is the set of right-continuous with limits from the left ("cadlag") functions on $(-\infty, \infty)$ with the topology of M_1 convergence on compacta. (The M_1 topology is discussed in detail in Subsection 8.2.4).

A similar useful result for the symmetry constrained problem which generalizes or extends Theorem 4.7 part C is as follows.

Theorem 4.9. *Suppose that the Curvature Assumption 2b holds and $f_0 \in \mathcal{SLC}_0$. Let $\sigma \equiv 1/\sqrt{2f_0(0)}$, and $a \equiv |\varphi_0^{(2)}(0)|/4!$. We let $Y_{a,\sigma}$ and $Y_{a,\sigma}^{(1)}$ be as in (4.18) and (4.19), and $H_{a,\sigma}^+(|t|)$ be the corresponding symmetry-constrained process for $t \in \mathbb{R}$ as in Theorem 4.4, and let $\widehat{\psi}_{a,\sigma}^0(|t|) \equiv (H_{a,\sigma}^+)''(|t|)$. Then*

$$\begin{pmatrix} n^{2/5}(\widehat{\psi}_n^0(n^{-1/5}t) - \varphi_0(0)) \\ n^{1/5}((\widehat{\psi}_n^0)'(n^{-1/5}t) - \varphi_0'(0)) \end{pmatrix} \to_d \begin{pmatrix} \widehat{\psi}_{a,\sigma}^0(|t|) \\ (\widehat{\psi}_{a,\sigma}^0)'(|t|) \end{pmatrix} \tag{4.22}$$

as processes in $\mathcal{C}_\infty \times \mathcal{D}_\infty$ with the topology of uniform convergence on compacts on $\mathcal{C}_\infty$ and the M_1 topology on $\mathcal{D}_\infty$.

Remark 4.10. To this point we have focused on the case in which the point of symmetry m is known (and equal to 0). If m is unknown (and possibly different from 0) and $f_0 \in \mathcal{SLC}_m$, then it is well-known that m is also the mean and median of f_0 and hence it can be estimated in several different ways by estimators $\hat{m}$ satisfying $\sqrt{n}(\hat{m} - m) = O_p(1)$. For example, we could take $\hat{m} = \overline{X}_n$ or $\hat{m} = \mathbb{F}_n^{-1}(1/2)$, the sample median. Then we can proceed by assuming that $f_0 \in \mathcal{SLC}_{\hat{m}}$ and carrying out the estimation as described above with the X_i's shifted by $\hat{m}$. Denote the resulting estimators of g and ψ by $\hat{\hat{g}}_n^0$ and $\hat{\hat{\psi}}_n^0$. Then since $n^{-1/2} = o(n^{-2/5})$ it is easily seen that that Theorems 4.5 C, 4.7 C and 4.9 continue to hold with $\hat{g}_n^0$ replaced by $\hat{\hat{g}}_n^0$ and $\hat{\psi}_n^0$ replaced by $\hat{\hat{\psi}}_n^0$.
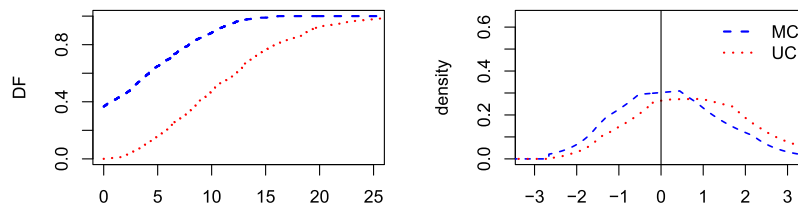
Fig 1. *Monte Carlo estimates of limit distributions for estimation of the derivative and the maximum. The left plot gives Monte Carlo estimates of distribution functions of $|\widehat{\varphi}'(0)|$ and $|(\widehat{\varphi}^0)'(0)|$. The right plot gives density estimates of $N(\widehat{\varphi})$ and $N(\widehat{\varphi}^0)$. The number of Monte Carlos used is 10^3, each with sample size 10^4. Here, we sampled from $N(0,1)$ (although this is inconsequential).*

4.4.1. Asymptotics for the maximum

We now consider the asymptotic distribution of estimators of the maximum functional $N(f) \equiv \sup_{x \in \mathbb{R}} f(x)$. The maximum functional is of interest, for instance, in Polonik (1995) (see page 872). An estimate of $N(f_0)$ is needed for estimation of functionals of the form $\int f_0^k(x)dx$. In the mode constrained case where $f_0 \in \mathcal{LC}_0$, $n^{2/5}(N(\widehat{f}_n^0) - N(f_0)) = n^{2/5}(\widehat{f}_n^0(m) - f_0(m))$. In the symmetry-constrained case with $f_0 \in \mathcal{SLC}_0$, $n^{2/5}(N(\widehat{g}_n^0) - N(f_0)) = n^{2/5}(\widehat{g}_n^0(0) - f_0(0))$. Thus the asymptotic distribution in those two cases is given by Theorem 4.7. We present the asymptotic distribution in the unconstrained case here. In fact, in the unconstrained case, we can present a somewhat stronger result, where we allow the possibility of increasingly flat modal regions.

Theorem 4.11. *Let W denote two-sided Brownian motion starting at 0, and define $Y_k(t) = \int_0^t W(s)ds - t^{k+2}$ for $k \geq 2$ an even integer. Let H_k be the lower invelope of Y_k, as defined in Theorem 2.1 of Balabdaoui, Rufibach and Wellner (2009), and let $\widehat{\varphi}_k \equiv H_k^{(2)}$. Suppose $f_0 \in \mathcal{LC}$ and that $\varphi_0^{(j)}(m) = 0$ for $j = 2, \ldots, k-1$, $\varphi_0^{(k)}(m) < 0$ and $\varphi_0^{(k)}$ is continuous in a neighborhood of m. Then*

$$\begin{pmatrix} n^{1/(2k+1)}(\widehat{M}_n - M(f_0)) \\ n^{k/(2k+1)}(\widehat{N}_n - N(f_0)) \end{pmatrix} \to_d \begin{pmatrix} e_k(\varphi_0)M(\widehat{\varphi}_k) \\ c_k(\varphi_0)N(\widehat{\varphi}_k) \end{pmatrix}$$

where

$$e_k(\varphi_0)^{2k+1} = \left(\frac{(k+2)!^2}{f_0(m)|\varphi_0^{(k)}(m)|^2} \right), \quad c_k(\varphi_0)^{2k+1} = \left(\frac{f_0(m)^{k+1}|\varphi_0^{(k)}(m)|}{(k+2)!} \right).$$

We compared, by Monte Carlo, the densities of $N(\widehat{\varphi})$ (note $\widehat{\varphi} \equiv \widehat{\varphi}_2$) and of $N(\widehat{\varphi}^0)$. See estimates in Figure 1 (right plot); those estimates are log-concave MLE's (based on Monte Carlo simulations, as described in the caption). Additionally, simulations not presented here indicate that $P(|N(\widehat{\varphi}^0)| \leq t) \geq P(|N(\widehat{\varphi})| \leq t)$ for all $t \geq 0$ (i.e., $|N(\widehat{\varphi}^0)|$ is stochastically smaller than $|N(\widehat{\varphi})|$).

The estimated density of $N(\widehat{\varphi})$ in Figure 1 should be compared with the estimated density of $\widehat{\varphi}(0)$ given in Figure 1 of Azadbakhsh, Jankowski and Gao (2014), noting that their $\mathbb{C}(0)$ is our $\widehat{\varphi}(0)$.

Remark 4.12. As is well known, one can see from Theorem 4.11 that the rate of convergence of the mode decreases as k increases, while the rate of convergence of the maximum increases (and gets closer to $n^{1/2}$). One can also check that $e_k(\log f) \searrow 1$ and $c_k(\log f) \nearrow 1$ when $f(x) = C_k \exp(-|x|^k/k)$ where k is an even integer.

5. Simulation results

Software to compute the mode-constrained estimator, and also to implement the likelihood ratio test and corresponding confidence intervals studied in Doss and Wellner (2019), is available in the package `logcondens.mode` (Doss, 2013a) in R (R Core Team, 2016). Here we illustrate the existence and characterization results on simulated data in Figure 2. There are two columns of four plots. The left column includes the mode-constrained log-concave MLE. The right column includes the 0-symmetric log-concave MLE. The data points are represented by vertical hash lines along the bottom of each plot. The density, log density, and distribution function are plotted in the top three rows, with the unconstrained log-concave MLE in red and the true (unknown) function in black. On the left the mode-constrained MLE is in blue, and on the right the 0-symmetric MLE is in blue. The empirical df $\mathbb{F}_n$ is plotted in green in the third row. In the last row, we plot $\mathbb{Y}_{n,L} - \widehat{H}_{n,L}^0$ (blue) and $\mathbb{Y}_{n,R} - \widehat{H}_{n,R}^0$ (purple) to illustrate Theorem 2.4 B (left plot), the corresponding symmetry-constrained process (blue) to illustrate Theorem 2.4 C (right plot), and $\mathbb{Y}_n - \widehat{H}_n$ in red (both plots) to illustrate Theorem 2.4 A. In all the plots, dashed vertical red lines give $S_n(\widehat{\varphi}_n)$ and dashed vertical blue lines give knots of the constrained estimator (which frequently overlap). The solid blue line is the specified mode value for the mode-constrained MLE.

Figure 3 gives plots of $\sqrt{n}(\widehat{G}_n^0 - F_0)$ ("SC"), $\sqrt{n}(\widehat{F}_n^0 - F_0)$ ("MC"), $\sqrt{n}(\widehat{F}_n - F_0)$ ("UC"), and $\sqrt{n}(\mathbb{F}_n - F_0)$ ("E"). The left and right plot are each one simulation with sample size $n = 200$ and $n = 2000$, respectively, from a $N(0, 1)$ distribution. The plots show improvements by $\widehat{G}_n^0$ and $\widehat{F}_n^0$ over $\widehat{F}_n$. The MC and UC lines are indistinguishable when $n = 2000$ since one needs to plot locally to the mode 0 to see differences between $\widehat{F}_n^0$ and $\widehat{F}_n$ when n is large.

6. Outlook and further problems

Motivated by likelihood ratio test considerations as well as potential uses in several semiparametric settings, we have introduced estimators for a log-concave density known to satisfy a further constraint of either having a known mode or of being symmetric (about 0). Our estimators are based on the maximum likelihood principle. The constrained MLE's that we develop are more challenging to compute and to study theoretically than certain naïve estimators discussed
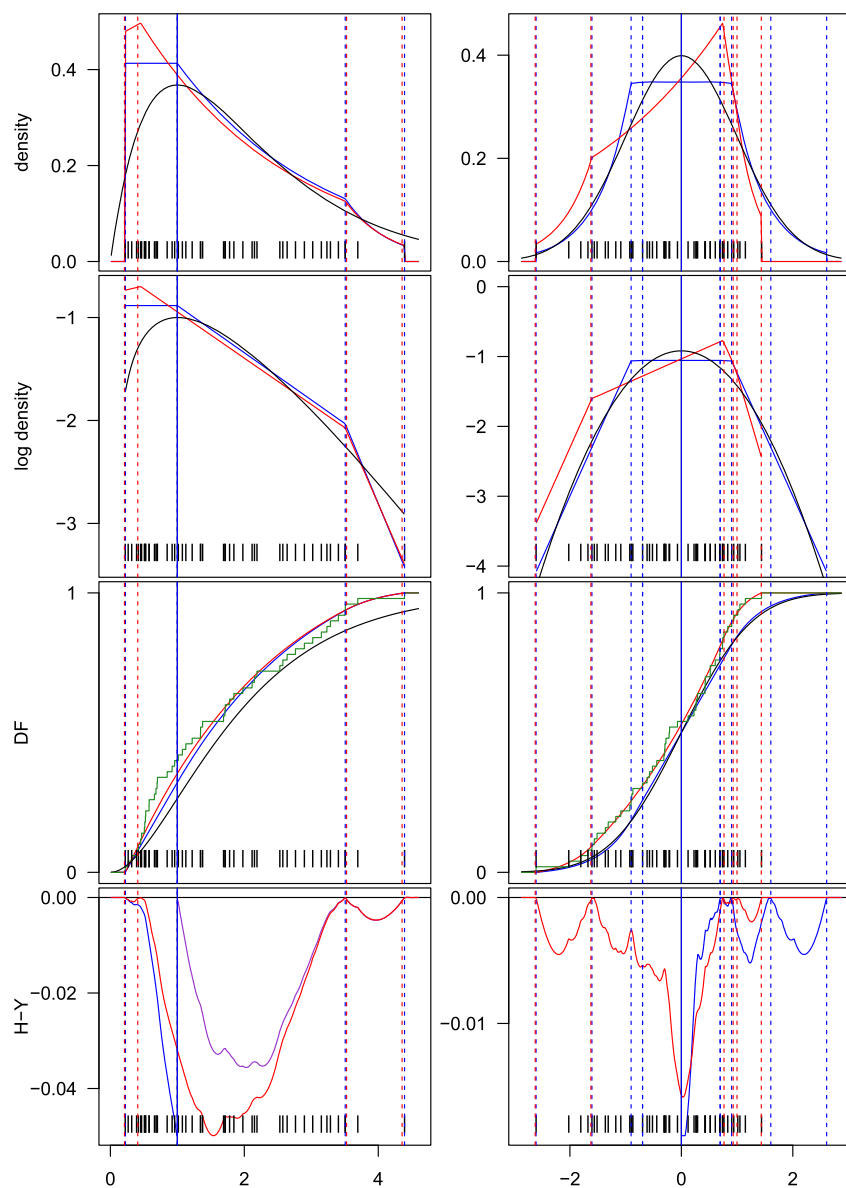
FIG 2. *Left: Log-concave MLE and mode-constrained MLE of Gamma with density xe^{-x}, $x \geq 0$, with $n = 50$ and $m = 1$ well-specified. Right: Log-concave MLE and 0-symmetric MLE of $N(0,1)$ with $n = 50$.*

in Subsection 2.2, but have much better theoretical behavior. We developed a fast algorithm for computation of the estimators which is made available in the `logcondens.mode` package for the R programming language. We found that the
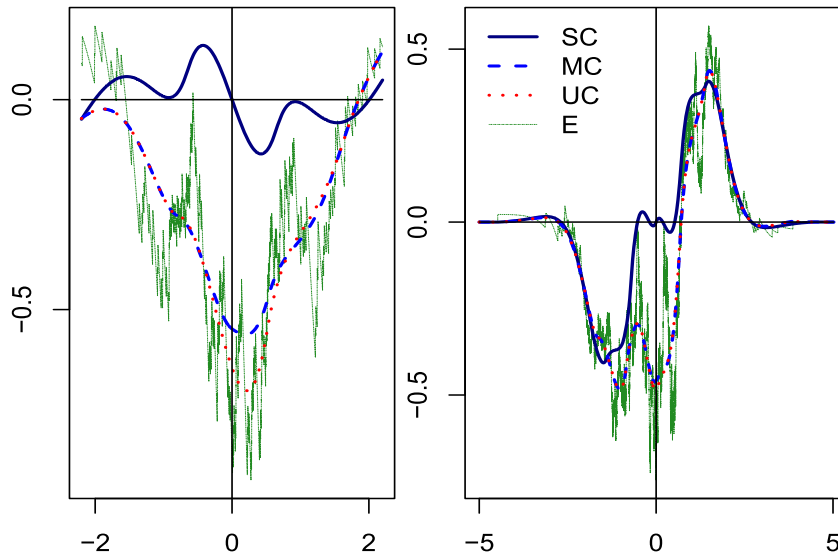
FIG 3. *Empirical processes for the symmetry constrained (SC) log-concave MLE, mode constrained (MC) log-concave MLE, unconstrained (UC) log-concave MLE, and empirical (E) distribution. We used a sample of $n = 200$ (left) and $n = 2000$ (right) from a $N(0, 1)$ distribution.*

constrained MLE's are consistent and indeed we presented the $n^{-2/5}$ rate of convergence, globally and locally (with some proofs given in Doss (2013b)). We found the pointwise asymptotic distribution of the MLE's, in fact; this necessitated studying and characterizing certain limit processes that govern the limit distributions of the MLE's. Studying the limit processes in the constrained cases seems to be somewhat more challenging than in the unconstrained case (i.e., than in the case given in Theorem 4.1), because the definitions and characterizing conditions for the constrained limit processes depend on certain knots (of the limit process) in complicated ways. Nonetheless, our proofs of Theorem 4.2 and Theorem 4.4 are different and shorter than the proof of Theorem 4.1; for the latter, Groeneboom, Jongbloed and Wellner (2001a) initially characterized the process on an interval $[-c, c]$ and then through further tightness-type arguments they showed that one can let $c \to \infty$. In the proofs of Theorem 4.2 and Theorem 4.4, we argue directly about a process on $(-\infty, \infty)$, skipping the step of considering the process on $[-c, c]$ and allowing for more direct proofs of the results.

The following are interesting questions beyond those already posed in the introduction that are motivated by the present work.

(a) One motivation for the study of $\widehat{f}_n^0$ given here has been the likelihood ratio tests and confidence intervals for the mode introduced in Doss and Wellner (2019). But the constrained estimators may be of interest for the study of semi-parametric two- and k−sample problems with (constrained) log-concave errors. For example suppose that $X_1, \ldots, X_m$ are i.i.d. with $X_i \overset{d}{=} \mu + \varepsilon_i$, $i = 1, \ldots, m$,

while $Y_1, \ldots, Y_n$ are i.i.d. with $Y_j \overset{d}{=} \nu + \delta_j$ where $\mu, \nu \in \mathbb{R}$ and ε_i, δ_j are i.i.d. with log-concave density f with mode at 0. Other variants of this problem might involve constraining f to be log-concave with mean or median at 0 rather than mode at 0. Constraining f to be symmetric about its mode of 0 and log-concave, as in Balabdaoui and Doss (2018), is also of interest.

(b) In Balabdaoui and Doss (2018), a mixture density $g_0(x) = \sum_{i=1}^k \pi_i f_0(x - \mu_i)$ is estimated where f_0 is constrained to be in $\mathcal{SLC}_0$, $k = 2$, $\mu_1 < \mu_2$, and $\pi_1 = 1 - \pi_2 \notin \{0, 1/2, 1\}$. Restriction to the case $k = 2$ is made for identifiability reasons. Can the asymptotic distribution theory we developed in the present paper be extended to the MLE of f_0 (and of g_0) in the semiparametric mixture setting? Extensions to the case $k > 2$ could also be possible and would certainly be interesting.

7. Proof sketches and outlines

In this section we give some outlines of the proofs of the results in Section 4. Full proofs are given in Section 8. Here we outline the material in each subsection of Section 8.

Subsection 8.2.1:

The main goal of Subsection 8.2.1 is to show the following proposition.

Proposition 7.1. *Suppose either Assumption 2a holds (at $x_0 \neq m$) or Assumption 2b holds (at $x_0 = m$). Let $C > 0$ and let $I_n = [x_0 - Cn^{-1/5}, x_0 + Cn^{-1/5}]$. Then*

$$\sup_{t \in I_n} |(\widehat{\varphi}_n^0)'(t) - \varphi_0'(x_0)| = O_p(n^{-1/5}), \tag{7.1}$$

where $(\widehat{\varphi}_n^0)'$ denotes either the right or left derivative, and

$$\sup_{t \in I_n} |\widehat{\varphi}_n^0(t) - \varphi_0(x_0) - (t - x_0)\varphi_0'(x_0)| = O_p(n^{-2/5}). \tag{7.2}$$

We may replace the interval I_n by $[\xi_n - Cn^{-1/5}, \xi_n + Cn^{-1/5}]$ for any $\xi_n \to m$. Then the random variables implied by the O_p upper bounds depend on C but not on ξ_n.

This proposition is of crucial importance for showing the results in Subsection 4.3 (Theorems 4.7, 4.8, and 4.9). The proof of Proposition 7.1 depends on the following two propositions.

Proposition 7.2. *Let $K = [b, c]$ be a closed interval contained in the interior of the support of $f_0 \in \mathcal{LC}_m$. Then*

$$\sup_{t \in K} |\widehat{\varphi}_n^0(t) - \varphi_0(t)| \to_{a.s.} 0 \quad and \quad \sup_{t \in K} |\widehat{f}_n^0(t) - f_0(t)| \to_{a.s.} 0 \quad as \quad n \to \infty.$$

Proposition 7.3. *Suppose that Curvature Assumption 2b holds so $\varphi_0''(m) < 0$. Let $\tau_+^0(\xi_n)$ denote the smallest knot of $\widehat{\varphi}_n^0$ strictly greater than ξ_n, and let $\tau_-^0(\xi_n)$*

denote the largest knot of $\widehat{\varphi}_n^0$ strictly smaller than ξ_n. Then for all $\varepsilon > 0$ there exists $C^0 > 0$ such that for any random variables $\xi_n \to_p m$

$$P(n^{1/5}(\tau_+^0(\xi_n) - \tau_-^0(\xi_n)) \geq C^0) \leq \varepsilon$$

for $n \geq N^0$ for some N^0. The integer N^0 may depend on ξ_m but C^0 does not depend on ξ_m.

Proposition 7.2 is proved in Doss (2013b). It is needed to show Proposition 7.3, which is needed to prove Proposition 7.1. The proof of Proposition 7.1 depends on Proposition 7.3, finding points of closeness of $\widehat{\varphi}_n$ and $\widehat{\varphi}_n^0$, and properties of convex functions. The full proof of Proposition 7.1 is given in Doss (2013b) (see Corollary 4.2.7 there). Thus the main goal of Subsection 8.2.1 is to prove Proposition 7.3 about the "gap problem" (a term coined in Balabdaoui and Wellner (2007)) for the constrained MLE near m. The proof depends on constructing certain (somewhat complicated) classes of perturbation functions which can be related to $\tau_+^0(\xi_n) - \tau_-^0(\xi_n)$, and then applying an argument pioneered by Kim and Pollard (1990) to these perturbations (see Lemma 8.2).

Subsections 8.2.1 and 8.2.3:

Subsection 8.2.1 is devoted to the proof of Theorem 4.2. The proof proceeds by defining an "objective function"

$$\phi_{a,b}(g) = \frac{1}{2}\int_a^b g^2(t)dt - \int_a^b g(t)dX(t) \tag{7.3}$$

for $a < 0 < b$. Now, $\widehat{\varphi}^0$ is not a minimizer of this objective function (which has finite bounds of integration) but we can show that $\widehat{\varphi}^0$ behaves in a sense as if it were a minimizer of $\phi_{a,b}$ over a the space of concave functions with mode at 0. We show that, for certain values $a < 0 < b$, directional derivatives of $\phi_{a,b}$, $\int_a^b \Delta(t)(\widehat{\varphi}^0(t)dt - dX(t))$, in the direction of a function Δ (assumed to be concave with mode at 0), are either nonnegative or in some cases are zero (see Propositions 8.7 and 8.8 for exact statements). This allows us to argue as follows. We want to show the uniqueness of any φ satisfying the characterizing conditions of Theorem 4.2. We assume there exist φ_1 and φ_2 both satisfying the characterizing conditions. We examine $\phi_{a_2,b_2}(\varphi_1) - \phi_{a_2,b_2}(\varphi_2)$ and $\phi_{a_1,b_1}(\varphi_2) - \phi_{a_1,b_1}(\varphi_1)$ where a_i, b_i are certain knot points for φ_i. By Propositions 8.7 and 8.8, we are able to show that both of these differences are no smaller than $\int_{-n}^n (\varphi_1(t) - \varphi_2(t))^2 dt > 0$ for $n > 0$ related to the knot points a_i, b_i. On the other hand, after deriving results about the knots a_i, b_i and relating the processes φ_i to the "observed" process Y (see Lemma 8.9), we are able to show that $\phi_{a_2,b_2}(\varphi_1) - \phi_{a_2,b_2}(\varphi_2)$ and $\phi_{a_1,b_1}(\varphi_2) - \phi_{a_1,b_1}(\varphi_1)$ are also nonpositive, by using properties of Brownian motion. Thus the difference $\int_{-n}^n (\varphi_1(t) - \varphi_2(t))^2 dt = 0$. By letting $n \to \infty$, we see $\varphi_1 = \varphi_2$ so the proof is complete. The proof is somewhat complicated by the fact that the "knots" of the concave function $\widehat{\varphi}^0$ are not separated but rather could be a complicated "Cantor-type" set, as described in Sinai (1992), and so their behavior requires careful analysis. The proof of Theorem 4.4 in Subsection 8.2.3 follows a similar argument.

Subsection 8.2.4:

Theorem 4.5 A about the unconstrained estimator is Theorem 2.1 of Balabdaoui, Rufibach and Wellner (2009). Part B of the theorem is then proved in an identical fashion as that theorem, because for n large enough, in an $n^{-1/5}$ neighborhood of $x_0 \neq m$, the constrained and unconstrained estimators satisfy the same characterization. Part C then follows from Part B. The main focus of Subsection 8.2.4 is to show Theorem 4.8. Theorem 4.9 follows in a similar fashion. Theorem 4.8 proves Theorem 4.7 B, which we show now. A similar argument shows that Theorem 4.9 proves Theorem 4.7 C.

Proof of Theorem 4.7 B. Note that, now with $\sigma = 1/\sqrt{f_0(m)}$ and $a = |\varphi_0^{(2)}(m)|/4!$, (4.21) equals

$$\frac{1}{\gamma_1\gamma_2^2}H_R^{(2)}(\cdot/\gamma_2) \equiv \frac{1}{\gamma_1\gamma_2^2}\widehat{\varphi}^0(\cdot/\gamma_2), \tag{7.4}$$

where

$$\gamma_1 = \left(\frac{f_0(m)^4|\varphi_0^{(2)}(m)|^3}{(4!)^3}\right)^{1/5} = \frac{1}{\sigma}\left(\frac{a}{\sigma}\right)^{3/5}, \tag{7.5}$$

$$\gamma_2 = \left(\frac{(4!)^2}{f_0(m)|\varphi_0^{(2)}(m)|^2}\right)^{1/5} = \left(\frac{\sigma}{a}\right)^{2/5}, \tag{7.6}$$

and we note that

$$\gamma_1\gamma_2^{3/2} = \sigma^{-1} = \sqrt{f_0(m)}, \quad \gamma_1\gamma_2^4 = a^{-1} = \frac{4!}{|\varphi_0^{(2)}(m)|}, \tag{7.7}$$

$$\frac{1}{\gamma_1\gamma_2^2} = C(m,\varphi_0) \equiv \left(\frac{4!f_0(m)^2}{|\varphi_0^{(2)}(m)|}\right)^{-1/5}. \tag{7.8}$$

This gives the constant in the limit distribution for $\widehat{\varphi}_n^0(m)$. For $(\widehat{\varphi}_n^0)'(m)$, we see from (7.4) that

$$(\widehat{\varphi}_{a,\sigma}^0)' =_d \frac{1}{\gamma_1\gamma_2^3}(\widehat{\varphi}^0)'(\cdot/\gamma_2), \tag{7.9}$$

and $1/\gamma_1\gamma_2^3 = D(m,\varphi_0)$. Thus Theorem 4.7 B follows from Theorem 4.8. $\quad\square$

Next we briefly outline the proof of Theorem 4.8. We define two sets of localized processes. We need to define left-side and right-side processes; for ease of exposition, here we only discuss right-side processes. We let $t_{n,b} \equiv m+bn^{-1/5}$. We let $s_{n,R}$ be any knot (sequence) of $\widehat{\varphi}_n^0$ strictly larger than m satisfying $n^{1/5}(s_{n,R}-m) = O_p(1)$. The first set of localized processes is the "f-processes" (written in terms of the densities and the empirical distribution):

$$\mathbb{Y}_{n,R}^f(b) \equiv n^{4/5}\int_{s_{n,R}}^{t_{n,b}}\left(\int_{s_{n,R}}^v (d\mathbb{F}_n - f_0(m)d\lambda)\right)dv,$$

$$\widehat{H}_{n,R}^{f}(b) \equiv n^{4/5} \int_{s_{n,R}}^{t_{n,b}} \left(\int_{s_{n,R}}^{v} (\widehat{f}_{n}^{0} - f_{0}(m)) d\lambda \right) dv + A_{n,R} n^{1/5} (t_{n,b} - s_{n,R}),$$

where $A_{n,R} = n^{3/5} \left(\mathbb{F}_{n,R}(s_{n,R}) - \widehat{F}_{n,R}^{0}(s_{n,R}) \right)$, and $A_{n,R} n^{1/5} (t_{n,b} - s_{n,R})$ is shown to be asymptotically negligible. These processes are needed because we can show that

$$\mathbb{Y}_{n,R}^{f}(\cdot) \Rightarrow \sqrt{f_{0}(m)} \int_{\nu_{n,R}}^{\cdot} \int_{\nu_{n,R}}^{v} dW \, dv - f_{0}^{(2)}(m) \int_{\nu_{n,R}}^{\cdot} \int_{\nu_{n,R}}^{v} \frac{u^{2}}{2} du \, dv \qquad (7.10)$$

where $\nu_{n,R} = n^{1/5}(s_{n,R} - m)$ and W is a standard Brownian motion with $W(0) = 0$ (see Lemma 8.17). Furthermore, the characterization from Theorem 2.4 B applies to $\mathbb{Y}_{n,R}^{f}$ and $\widehat{H}_{n,R}^{f}$, so $\mathbb{Y}_{n,R}^{f}(b) - \widehat{H}_{n,R}^{f}(b) \geq 0$ for $b > 0$, with equality at certain points (see Lemma 8.18).

On the other hand, the tightness proposition from above, Proposition 7.1, applies to the log-densities. Thus we define the second set of processes, the "φ-processes", written in terms of log-densities:

$$\mathbb{Y}_{n,R}^{\varphi}(b) = \frac{\mathbb{Y}_{n,R}^{f}(b)}{f_{0}(m)} - n^{4/5} \int_{s_{n,R}}^{t_{n,b}} \int_{s_{n,R}}^{v} R_{n}^{0}(u) du \, dv,$$

$$\widehat{H}_{n,R}^{\varphi}(b) = n^{4/5} \int_{s_{n,R}}^{t_{n,b}} \int_{s_{n,R}}^{v} \left(\widehat{\varphi}_{n}^{0}(u) - \varphi_{0}(m) \right) du \, dv + \frac{A_{n,R} n^{1/5} (t_{n,b} - s_{n,R})}{f_{0}(m)}$$

where R_{n}^{0} is a remainder term. These are related to the f-processes by a Taylor expansion (the delta method). One can translate the characterizing inequalities from the f-processes to the φ-processes, to see that $\mathbb{Y}_{n,R}^{\varphi}(b) - \widehat{H}_{n,R}^{\varphi}(b) \geq 0$ for $b \geq 0$ with equality at certain points. By (7.10), $\mathbb{Y}_{n,R}^{\varphi}$ can be shown to converge to a limiting Gaussian process. Furthermore, we can apply Proposition 7.1 and the Arzela-Ascoli theorem (after analyzing various remainder terms) to see that $\widehat{H}_{n,R}^{\varphi}$ is tight (Lemma 8.24).

Finally, we make a subsequence argument using tightness. By tightness, Prohorov's theorem, and the Skorokhod construction, for any subsequence we can find a further subsubsequence that converges almost surely to a limit process. By using the characterization (Lemma 8.21, and by analysis of various remainder terms), we show that the limit process must satisfy the unique characterizing conditions given by Theorem 4.2. This shows the limit is the same along all subsequences and so the unique process $\widehat{\varphi}^{0}$ given in Theorem 4.2 is the limit, which completes the proof. The argument showing that the characterizing conditions pass from the finite sample processes to the limit process is somewhat complicated by the fact that the integrands in question are defined to begin at random knot points.

Another issue of note is that $(\widehat{\varphi}_{n}^{0})'$ is a discontinuous function. We must choose or find an appropriate metric space in which to study its convergence; the metric we choose is the so-called M_{1} Skorokhod metric, which differs from the (perhaps more standard) J_{1} Skorokhod metric (referred to as "the" Skorokhod

metric in chapter 12 of Billingsley (1999)). The J_1 metric unfortunately does not allow multiple jumps to approximate a single jump, whereas the M_1 metric does. Since we do not have a proof that multiple jumps of $(\widehat{\varphi}_n^0)'$ do not approximate a single jump in the limit, we must use the M_1 metric. See Lemma 8.23 and the preceding text for further discussion.

8. Proofs

8.1. Proofs for Section 2

Proofs of Theorems 2.1 A, 2.2 A, 2.4 A and Corollaries 2.7 A and 2.8 A may be found in Pal, Woodroofe and Meyer (2007), Rufibach (2006), and Dümbgen and Rufibach (2009). Proofs of parts B and C of Corollary 2.7 and part B of Corollary 2.8 follow from the corresponding parts of Theorems 2.2 and 2.4, much as in the unconstrained case, A. It remains to prove parts B and C of Theorems 2.1, 2.2, and 2.4.

In the following proofs, we let $\mathcal{C}_{n,m}$ denote the (random) class of concave functions with knots at the Z_i's and support on $[X_{(1)}, X_{(n)}]$, and let $\mathcal{K}_{n,m}$ denote the class of concave functions φ with knots at the Z_i's and where e^φ is a density with support $[X_{(1)}, X_{(n)}]$.

Proof of Theorem 2.1. Proof of Theorem 2.1 B: As in the unconstrained case (Theorem 2.1 of Dümbgen and Rufibach (2009)), it is easy to argue that the solution is piecewise linear with knots at the Z_i's, and that $\widehat{\varphi}_n^0$ is flat either directly to the left of the mode or directly to the right of the mode as long as the mode is not a data point. If the mode is equal to one of the X_i's then the proof given by Rufibach (2006) for the unconstrained MLE existence applies directly. Thus assume the mode is not one of the X_i. Consider a sequence $\{v_j\}$ which has limit coordinates $\gamma = (\gamma_1, \ldots, \gamma_N)$ which may be $\pm\infty$. Let φ_γ be the piecewise linear function given by linearly interpolating γ. Then φ_γ has a flat modal region on $[Z_i, Z_j]$ for some $i < j$. Since $\int e^{\varphi_\gamma(x)}dx = 1$, we have $\varphi_\gamma(m) \leq \log(1/(Z_j - Z_i))$. Since m is the mode of e^{φ_γ} no coordinate of γ is positive infinity, if one of the coordinates is $-\infty$ then $\Psi_n(\varphi_\gamma) = -\infty$. This shows we can consider the continuous function Ψ_n on a compact set so it achieves a maximum. The proof that if $\widehat{\psi}_n$ maximizes Ψ_n over $\mathcal{C}_m$ then $\int e^{\widehat{\psi}_n(x)}dx = 1$ as well as the proof that $\widehat{\psi}_n$ is unique are as in the unconstrained case (Rufibach (2006)).

For Theorem 2.1 C, note that

$$\operatorname*{argmax}_{g \in \mathcal{SLC}_0} \frac{1}{n} \sum_{i=1}^n \log g(X_i) = \operatorname*{argmax}_{g \in \mathcal{SLC}_0} \frac{1}{n} \sum_{i=1}^n \log g(|X_i|) = \operatorname*{argmax} \frac{1}{n} \sum_{i=1}^n \log g(|X_i|)$$

where the last argmax is over log-concave functions with mode at 0 and which integrate to $1/2$ on $[0, \infty)$. $\qquad\square$

The proof of Theorem 2.2 B is standard. See Doss (2013b). For the proof of Theorem 2.4 B we need to introduce a certain *cone* $C \subseteq \mathbb{R}^d$ (defined, e.g., on page 13 of Rockafellar (1970)). We say the cone C is *(finitely) generated* by a set $\{b_i \in C : i = 1, \ldots, k < \infty\}$ if for all $c \in C$ we can write $c = \sum_{i=1}^{k} \alpha_i b_i$ for some nonnegative numbers $\alpha_i \geq 0$. Let $(x)_- = \min(x, 0)$ and $(x)_+ = \max(x, 0)$.

Proposition 8.1. $\mathcal{C}_{n,m}$ *is a convex cone with finite generating set given by*

$$\{(x - Z_i)_-\}_{2 \leq i \leq k} \bigcup \{(Z_i - x)_-\}_{k \leq i \leq N-1} \bigcup \{\pm 1\}.$$

Proof. It is clear that $\mathcal{C}_{n,m}$ is a cone because concavity and the mode are preserved under positive scaling. For $\varphi \in \mathcal{C}_{n,m}$, with $\varphi'(Z_i-) = \sum_{j=i}^{k} a_j$ for $i \leq k$, with, $\varphi'(Z_i+) = \sum_{j=k}^{i} b_j$ for $i \geq k$, and with $\varphi(m) = C$, we can write $\varphi(x) = C + \sum_{i=2}^{k} a_i(x - Z_i)_- + \sum_{i=k}^{n-1} b_i(Z_i - x)_-$. $\square$

Proof of Theorem 2.4. Proof of Part B: First we assume $\widehat{\varphi}_n^0$ is the MLE and use (2.5) to show that (2.8) and (2.9) hold. Using the generating functions described in Proposition 8.1 as our Δ yields equations (2.8) and (2.9) via integration by parts. That is, for $t \leq m$, we choose $\Delta(x) = (x - t)_-$ (which is concave with m as a mode). Then integration by parts yields $-\int_{X_{(1)}}^{t} F(x)dx = \int_{X_{(1)}}^{X_{(n)}} (x-t)_- dF(x)$, by Lemma 9.1 B, for F equal to either $\mathbb{F}_n$ or $\widehat{F}_n^0$ since $\mathbb{F}_n(X_{(1)}-) = \widehat{F}_n^0(X_{(1)}-) = 0$. Thus, by our initial characterization (2.5), we get (2.8). Similarly, for $t \geq m$, let $\Delta(x) = (t - x)_-$; this yields

$$\left(t - X_{(n)}\right) \mathbb{F}_n(X_{(n)}) - \int_{t}^{X_{(n)}} \mathbb{F}_n(x)d(-x)$$
$$\leq \left(t - X_{(n)}\right) \widehat{F}_n^0(X_{(n)}) - \int_{t}^{X_{(n)}} \widehat{F}_n^0(x)d(-x),$$

and, recalling that we have already shown $\widehat{F}_n^0(X_{(n)}) = 1$, this is equivalent to $\int_{t}^{X_{(n)}} \mathbb{F}_n(x)dx \leq \int_{t}^{X_{(n)}} \widehat{F}_n^0(x)dx$, so we have (2.9). We get equality at some knot points also: set $\Delta(x) = (x - b)_+$ where $b \geq m$ is any RK. Then, by the definition of a RK, Δ is an allowable perturbation because $\widehat{\varphi}_n^0(b+) - \widehat{\varphi}_n^0(b-) > 0$ so for some δ small enough, $\widehat{\varphi}_n^0 + \delta\Delta$ is still concave with mode at m. Using this Δ we get

$$\mathbb{F}_n(X_{(n)}) \left(X_{(n)} - b\right) - \int_{b}^{X_{(n)}} \mathbb{F}_n(x)dx$$
$$= \int_{X_{(1)}}^{X_{(n)}} (x - b)_+ d\mathbb{F}_n(x) \leq \int_{X_{(1)}}^{X_{(n)}} (x - b)_+ d\widehat{F}_n^0(x)$$
$$= \widehat{F}_n^0(X_{(n)}) \left(X_{(n)} - b\right) - \int_{b}^{X_{(n)}} \widehat{F}_n^0(x)dx,$$

so that $\int_{b}^{X_{(n)}} \mathbb{F}_n(x)dx \geq \int_{b}^{X_{(n)}} \widehat{F}_n^0(x)dx$, and thus for any $b \geq m$ that is a RK we have the inequality both ways, $\int_{b}^{X_{(n)}} \mathbb{F}_n(x)dx = \int_{b}^{X_{(n)}} \widehat{F}_n^0(x)dx$. We have thus shown that (2.8) and (2.9) hold with the appropriate equalities.

Now we will show the reverse implication. We assume (2.8) and (2.9) hold and consider Δ with support $[X_{(1)}, X_{(n)}]$ and piecewise linear with knots at the Z_i. These are all the Δ's we need to consider, since the rest were ruled out previously by elementary considerations. We also need $\widehat{\varphi}_n^0 + \varepsilon\Delta$ to be concave with mode m. Now, we do not know if m will be a NK or a LK or a RK, so we argue by cases. If m is a knot for $\widehat{\varphi}_n^0$ in one direction, without loss of generality, we can say that m is a RK and we have $\int_m^c \mathbb{F}_n(x)dx = \int_m^c \widehat{F}_n^0(x)dx$ for any $c > m$ that is also a knot. Recall that we have defined the indices $1 = j_1, \ldots, j_{l^0} = N$ so that Z_{j_i} are the knots. We write

$$\Delta'(r-) = \sum_{i=2}^{l} -C_i 1_{[Z_{j_i-1} < r \le Z_{j_i}]} + \sum_{j=j_{i-1}+1}^{j_i} \beta_j 1_{[Z_{j_i-1} < r \le Z_j]} \qquad (8.1)$$

with $\beta_j \ge 0$ and all $C_i \ge 0$. Since m is a RK, m is not also a LK (otherwise m is simply a knot and $\widehat{\varphi}_n^0$ coincides with the unconstrained MLE and the characterization of the unconstrained MLE in (Dümbgen and Rufibach, 2009) implies we are done). This forces $C_p = 0$ (which refers to the interval $(Z_{j_p-1}, m = Z_{j_p}]$). We thus have

$$\int \Delta d\mathbb{F}_n = \Delta(X_{(n)}) - \left[\sum_{i=2}^{l} -C_i \int_{Z_{j_i-1}}^{Z_{j_i}} \mathbb{F}_n(x)\,dx + \sum_{j=j_{i-1}+1}^{j_i} \beta_j \int_{Z_{j_i-1}}^{Z_j} \mathbb{F}_n(x)\,dx\right]$$

$$\le \Delta(X_{(n)}) - \left[\sum_{i=2}^{l} -C_i \int_{Z_{j_i-1}}^{Z_{j_i}} \widehat{F}_n^0(x)\,dx + \sum_{j=j_{i-1}+1}^{j_i} \beta_j \int_{Z_{j_i-1}}^{Z_j} \widehat{F}_n^0(x)\,dx\right]$$

$$= \int \Delta d\widehat{F}_n^0$$

as desired, where the inequality follows from noting that $-\beta_j \int_{Z_{j_i-1}}^{Z_j} \mathbb{F}_n(x)dx \le -\beta_j \int_{Z_{j_i-1}}^{Z_j} \widehat{F}_n^0(x)dx$ by assumption and because $\beta_j \ge 0$. We also have

$$C_i \int_{Z_{j_i-1}}^{Z_{j_i}} \mathbb{F}_n(x)\,dx = C_i \int_{Z_{j_i-1}}^{Z_{j_i}} \widehat{F}_n^0(x)\,dx$$

for all i except for $i = p$, by the equality-at-knots assumption. However, for $i = p$, we have $C_i = 0$. An analogous argument holds for the case where m is an LK and for the case where m is neither an LK nor an RK.

Part C follows as in the proof of Theorem 2.1 C; $\widehat{g}_n^+$ is the MLE over $\mathcal{LC}_0$ of $|X_1|, \ldots, |X_n|$. Thus we apply the result of Part B. Note that $0 \in S_n(\widehat{\psi}_n^0) \cap [0, |X|_{(n)}]$ only if $(\widehat{\psi}_n^0)'(0+) < 0$ (so that 0 is a right knot). This is only possible if 0 is an observed data point.   □

Proof of Corollary 2.7 B and C follow as in the unconstrained case (Corollary 2.5, Dümbgen and Rufibach (2009)).

8.2. Proofs for local rates of convergence, limit processes, and limit theory

8.2.1. Proofs for local rates of convergence

This section is devoted to showing, Proposition 7.1, stated above, which is needed for proving Theorems 4.8, and 4.9. See 7 for a discussion of the proof of Proposition 7.1, the full details of which are given in Doss (2013b) (Corollary 4.2.7). Our main goal here is to prove Proposition 7.3.

Proof of Proposition 7.3. We first consider the case $\xi_n = m$. For ease of notation and without loss of generality, we assume $m = 0$ and abbreviate $\tau_{n,+}^0(0)$ by $\tau_{n,+}^0$ and $\tau_{n,-}^0$ by $\tau_{n,-}^0$. We will argue via a family of perturbations which can be separated into subfamilies, depending on whether 0 is a left-knot (LK), 0 is a right-knot (RK), or 0 is not a knot (NK). If 0 is a one-sided knot (LK or RK), we have different perturbation subfamilies depending on whether $\tau_{n,+}^0 > -\tau_{n,-}^0$ or not. We will start with the case in which 0 is a LK and we construct Δ that has the two properties

$$\int_{\tau_{n,-}^0}^{\tau_{n,+}^0} \Delta(t)dt = 0, \tag{8.2}$$

$$\int_{\tau_{n,-}^0}^{0} t\Delta(t)dt = 0. \tag{8.3}$$

Whereas in the unconstrained case construction of such an acceptable perturbation function was straightforward (Balabdaoui, Rufibach and Wellner, 2009), in the constrained case, construction of such a Δ that is an acceptable perturbation (i.e. keeps the mode fixed) is much less straightforward. We consider several cases separately.

Case 1 $\tau_{n,+}^0 > -\tau_{n,-}^0$. In this case we define

$$\Delta_{LK,0}(t) = \begin{cases} \frac{\tau_{n,+}^0 + m_2 \cdot (-\frac{\tau_{n,-}^0}{4})}{\frac{\tau_{n,-}^0}{4} - \tau_{n,-}^0} \cdot (t - \tau_{n,-}^0), & \tau_{n,-}^0 \le t \le \frac{\tau_{n,-}^0}{4} \\ \tau_{n,+}^0 + m_2 \cdot (-t), & \frac{\tau_{n,-}^0}{4} \le t \le 0 \\ (\tau_{n,+}^0 - t), & 0 \le t \le \tau_{n,+}^0 \\ 0, & \text{otherwise,} \end{cases} \tag{8.4}$$

where

$$m_2 := m_2(\tau_{n,-}^0, \tau_{n,+}^0) = \left(\frac{-9 - 3\frac{\tau_{n,+}^0}{-\tau_{n,-}^0}}{1 - 5\frac{\tau_{n,+}^0}{-\tau_{n,-}^0}} \right) \left(\frac{\tau_{n,+}^0}{-\tau_{n,-}^0} \right).$$

This function has integral

$$\frac{(\tau_{n,+}^0)(5\tau_{n,+}^0 - \tau_{n,-}^0)(\tau_{n,+}^0 - \tau_{n,-}^0)}{2(5\tau_{n,+}^0 + \tau_{n,-}^0)} \equiv M_{LK,case1}(\tau_{n,+}^0 - \tau_{n,-}^0).$$

Case 2 $\tau_{n,+}^0 < -\tau_{n,-}^0$. In this case we define

$$\Delta_{LK,0}(t) = \begin{cases} (t - \tau_{n,-}^0), & \tau_{n,-}^0 \leq t \leq \frac{\tau_{n,-}^0}{2} \\ \frac{\tau_{n,-}^0}{2} - \tau_{n,-}^0 + m_2 \cdot (\frac{\tau_{n,-}^0}{2} - t), & \frac{\tau_{n,-}^0}{2} \leq t \leq 0 \\ \left(\frac{-\frac{\tau_{n,-}^0}{2} + m_2 \frac{\tau_{n,-}^0}{2}}{\tau_{n,+}^0} \right) (\tau_{n,+}^0 - t), & 0 \leq t \leq \tau_{n,+}^0 \\ 0, & \text{otherwise,} \end{cases}$$ (8.5)

where

$$m_2 := m_2(\tau_{n,-}^0, \tau_{n,+}^0) = \frac{2\tau_{n,-}^0 + \tau_{n,+}^0}{2\tau_{n,-}^0 - 5\tau_{n,+}^0}$$

is defined so that (8.3) holds. This function has integral

$$\frac{-\tau_{n,-}^0(3\tau_{n,+}^0 - \tau_{n,-}^0)(\tau_{n,+}^0 - \tau_{n,-}^0)}{10\tau_{n,+}^0 - 4\tau_{n,-}^0} \equiv M_{LK,case2}(\tau_{n,+}^0 - \tau_{n,-}^0).$$

Then we define

$$\Delta_{LK,1}(t) = \Delta_{LK,0}(t) - M_{LK} 1_{[\tau_{n,-}^0, \tau_{n,+}^0]}(t),$$ (8.6)

where

$$M_{LK} = \frac{(\tau_{n,+}^0)(5\tau_{n,+}^0 - \tau_{n,-}^0)}{2(5\tau_{n,+}^0 + \tau_{n,-}^0)} 1_{[\tau_{n,+}^0 > \tau_{n,-}^0]} + \frac{-\tau_{n,-}^0(3\tau_{n,+}^0 - \tau_{n,-}^0)}{10\tau_{n,+}^0 - 4\tau_{n,-}^0} 1_{[\tau_{n,+}^0 \leq \tau_{n,-}^0]},$$

is $o_p(1)$ by uniform consistency of $\widehat{\varphi}_n^0$ and is the appropriate shift so that (8.2) holds. Then $\Delta_{LK,0}$ is an acceptable perturbation for $\widehat{\varphi}_n^0$, since we can have $m_2 > 1$ when 0 is a LK, and $\Delta_{LK,1}$ has the properties (8.2) and (8.3). We also define $\Delta_{RK,1}$ analogously as $\Delta_{LK,1}$, with analogous constant M_{RK}.

Now consider the case in which 0 is not a knot (NK). In this case, because $(\widehat{\varphi}_n^0)'(t) = 0$ for all $t \in (\tau_{n,-}^0, \tau_{n,+}^0)$, we only need Δ to have the property

$$\int_{\tau_{n,-}^0}^{\tau_{n,+}^0} \Delta(t)dt = 0.$$ (8.7)

So, if $\tau_{n,+}^0 > -\tau_{n,-}^0$ define

$$\Delta_{NK,0}(t) := \begin{cases} \frac{\tau_{n,+}^0}{-\tau_{n,-}^0}(t - \tau_{n,-}^0), & \text{for } t \in [\tau_{n,-}^0, 0] \\ (\tau_{n,+}^0 - t), & \text{for } t \in [0, \tau_{n,+}^0] \\ 0, & \text{otherwise,} \end{cases}$$ (8.8)

and if $\tau_{n,+}^0 < -\tau_{n,-}^0$, define

$$\Delta_{NK,0}(t) := \begin{cases} (t - \tau_{n,-}^0), & \text{for } t \in [\tau_{n,-}^0, 0] \\ \frac{-\tau_{n,-}^0}{\tau_{n,+}^0}(\tau_{n,+}^0 - t), & \text{for } t \in [0, \tau_{n,+}^0] \\ 0, & \text{otherwise.} \end{cases}$$ (8.9)

Denote $h_l = \max(\tau^0_{n,+}, -\tau^0_{n,-})$ and $h_s = \min(\tau^0_{n,+}, -\tau^0_{n,-})$. We then compute $\int \Delta_{NK,0}(x) dx = h_l(\tau^0_{n,+} - \tau^0_{n,-})/2$, so set

$$\Delta_{NK,1}(x) := \Delta_{NK,0}(x) - \frac{h_l}{2} 1_{[\tau^0_{n,-}, \tau^0_{n,+}]}(x), \tag{8.10}$$

so that (8.7) is satisfied. Let

$$\Delta_0(x) = \Delta_{RK,0}(x) 1_{[m \text{ is RK}]} + \Delta_{LK,0}(x) 1_{[m \text{ is LK}]} + \Delta_{NK,0}(x) 1_{[m \text{ is NK}]},$$
$$M_n = M_{RK} 1_{[m \text{ is RK}]} + M_{LK} 1_{[m \text{ is LK}]} + h_l 1_{[m \text{ is NK}]},$$

and let

$$\Delta_1(x) = \Delta_{RK,1}(x) 1_{[m \text{ is RK}]} + \Delta_{LK,1}(x) 1_{[m \text{ is LK}]} + \Delta_{NK,1}(x) 1_{[m \text{ is NK}]}$$
$$= \Delta_0(x) - M_n. \tag{8.11}$$

Note that M_n is $o_p(1)$ by uniform consistency in a neighborhood of m. By the characterization Theorem 2.2B and Corollary 2.7, we see that

$$\int \Delta_1 d(\mathbb{F}_n - F_0) = \int \Delta_1 d(\mathbb{F}_n - \widehat{F}^0_n) + \int \Delta_1 d(\widehat{F}^0_n - F_0)$$

$$= \int \Delta_0 d(\mathbb{F}_n - \widehat{F}^0_n) - M_n \int_{\tau^0_{n,-}}^{\tau^0_{n,+}} d(\mathbb{F}_n - \widehat{F}^0_n) + \int \Delta_1 d(\widehat{F}^0_n - F_0)$$

$$\leq M_n \left| \int_{\tau^0_{n,-}}^{\tau^0_{n,+}} d(\mathbb{F}_n - \widehat{F}^0_n) \right| + \int \Delta_1(x)(\widehat{f}^0_n - f_0)(x) dx$$

$$\leq \frac{2M_n}{n} + \int \Delta_1(x)(\widehat{f}^0_n - f_0)(x) dx. \tag{8.12}$$

Then Lemma 8.2 yields

$$\left| \int \Delta_1 d(\mathbb{F}_n - F_0) \right| \leq O_p(n^{-4/5}) + \frac{K}{2} h_l^4, \tag{8.13}$$

$$\int \Delta_1(x)(\widehat{f}^0_n - f_0)(x) dx \leq -K h_l^4 + o_p(h_l^4),$$

for some $K > 0$ and where we picked ε from Lemma 8.2 to be $K/2$. So, by rearranging (8.12), we have

$$K(1 + o_p(1)) h_l^4 \leq -\int \Delta_1(x)(\widehat{f}_n - f_0)(x) dx$$

$$\leq \frac{2M_n}{n} - \int \Delta_1 d(\mathbb{F}_n - F_0) \leq \frac{2M_n}{n} + \left| \int \Delta_1 d(\mathbb{F}_n - F_0) \right|$$

$$\leq O_p(n^{-4/5}) + \frac{K}{2} h_l^4,$$

and hence

$$(K/2 + o_p(1)) h_l^4 \leq O_p(n^{-4/5})$$

which yields $0 \le h_l = O_p(n^{-1/5})$. Since $\tau_{n,+}^0$ and $-\tau_{n,-}^0$ are both bounded by h_l, we are done for the case $\xi_n = m$.

Extending to the case where $\xi_n \to m$ rather than being fixed and equal to m follows from considering the event $m \in [\tau_{n,-}^0(\xi_n), \tau_{n,+}^0(\xi_n)]$ and its complement separately, and then by showing on the event $m \notin [\tau_{n,-}^0(\xi_n), \tau_{n,+}^0(\xi_n)]$ that $\tau_{n,+}^0(\xi_n) - \tau_{n,-}^0(\xi_n) = O_p(n^{-1/5})$. This is straightforward. $\qquad\square$

The following lemmas were used.

Lemma 8.2. *We continue with the setup of Proposition 7.3. That is, we define* $h_l = \max(\tau_{n,+}^0 - m, m - \tau_{n,-}^0)$, $h_s = \min(\tau_{n,+}^0 - m, m - \tau_{n,-}^0)$ *and* Δ_1 *as in* (8.11). *Then for all* $\varepsilon > 0$,

$$\left| \int \Delta_1 d(\mathbb{F}_n - F_0) \right| \le \varepsilon h_l^4 + O_p(n^{-4/5}) \tag{8.14}$$

and

$$\int \Delta_1(x)(\widehat{f}_n^0 - f_0)(x)dx \le -\frac{f_0(0)|\varphi_0''(0)|}{2} K h_l^4 + o_p(h_l^4), \tag{8.15}$$

where $K > 0$ *is from Lemma 8.3, and does not depend on* f_0.

Proof. We examine $\int \Delta_1(x)(\widehat{f}_n^0 - f_0)(x)dx$ by repeated Taylor expansions, where we let Δ_1 be $\Delta_{LK,1}$ or $\Delta_{NK,1}$, and we will expand at m, which we again take to be 0, without loss of generality. Write $(\widehat{f}_n^0 - f_0) = f_0((\widehat{f}_n^0/f_0) - 1) = f_0(\exp\{\widehat{\varphi}_n^0 - \varphi_0\} - 1)$. Then write $d_n = \widehat{\varphi}_n^0 - \varphi_0$ so we can expand

$$\exp(d_n(t)) - 1 = \sum_{i=1}^{1} \frac{d_n(t)^i}{i!} + e^{\xi_{1,n,t}} \frac{d_n(t)^2}{2!},$$

$$f_0(t) = \sum_{i=0}^{1} \frac{f_0^{(i)}(0)t^i}{i!} + \frac{f_0^{(2)}(\xi_{2,n,t})t^2}{2!}$$

for $t \in [\tau_{n,-}^0, \tau_{n,+}^0]$, where $\xi_{1,n,t}$ is between 0 and $d_n(t)$ and $\xi_{2,n,t}$ is between 0 and t. So, writing $\|\cdot\|_\infty$ as the uniform norm over $[\tau_{n,-}^0, \tau_{n,+}^0]$, we can see that $f_0(t)(e^{d_n(t)} - 1)$ equals

$$\left(\sum_{i=0}^{1} \frac{(f_0)^{(i)}(0)t^i}{i!} + \frac{(f_0)^{(2)}(\xi_{2,n,t})t^2}{2!} \right) \left(\sum_{i=1}^{1} \frac{d_n(t)^i}{i!} + e^{\xi_{1,n,t}} \frac{d_n(t)^2}{2!} \right)$$

$$= f_0(0)d_n(t) + o_p(\|d_n(t)\|_\infty)$$

since $f_0(0) > 0$, $f_0^{(i)}$ is continuous and thus bounded on a neighborhood of 0 for $i \in \{0, 1, 2\}$, and, by uniform consistency, $d_n(t)^i$ and t^i both go to 0 uniformly in a neighborhood of 0. Then we examine

$$\int_{\tau_{n,-}^0}^{\tau_{n,+}^0} \Delta_1(t)d_n(t)dt = \sum_{i=0}^{1} \frac{d_n^{(i)}(0)}{i!} \int_{\tau_{n,-}^0}^{\tau_{n,+}^0} t^i \Delta_1(t)dt + \int_{\tau_{n,-}^0}^{\tau_{n,+}^0} \frac{d_n^{(2)}(\xi_{3,n,t})}{2!} t^2 \Delta_1(t)dt$$

where $\xi_{3,n,t} \in [\tau_{n,-}^0, \tau_{n,+}^0]$. Note that $d_n^{(2)}(t) = -\varphi_0^{(2)}(t)$, and that since $d_n^{(2)}$ is continuous *at* 0 we can write $d_n^{(2)}(\xi_{3,n,t}) = d_n^{(2)}(0) + \varepsilon_n(t)$ where $\|\varepsilon_n(t)\|_\infty \to_p 0$ since $\tau_{n,+}^0 - \tau_{n,-}^0 \to_p 0$. Now we consider the different possible forms Δ_1 may take.

If 0 is NK, then $(\widehat{\varphi}_n^0)^{(2)}(0) = 0 = \varphi_0'(0)$, so

$$
\begin{aligned}
\int_{\tau_{n,-}^0}^{\tau_{n,+}^0} & \Delta_{NK,1}(t)d_n(t)dt \\
&= \sum_{i=0}^2 \frac{d_n^{(i)}(0)}{i!} \int_{\tau_{n,-}^0}^{\tau_{n,+}^0} t^i \Delta_{NK,1}(t)dt + \int_{\tau_{n,-}^0}^{\tau_{n,+}^0} \frac{\varepsilon_n(t)}{2!} t^2 \Delta_{NK,1}(t)dt \\
&= -\frac{\varphi_0^{(2)}(0)}{2!} \int_{\tau_{n,-}^0}^{\tau_{n,+}^0} t^2 \Delta_{NK,1}(t)dt + \int_{\tau_{n,-}^0}^{\tau_{n,+}^0} \frac{\varepsilon_n(t)}{2!} t^2 \Delta_{NK,1}(t)dt. \quad (8.16)
\end{aligned}
$$

Now we show that we get the same expansion if 0 is a LK. Note that for $t \in [0, \tau_{n,+}^0]$, $\widehat{\varphi}_n^0(t) = \widehat{\varphi}_n^0(0)$ and for $t \in [\tau_{n,-}^0, 0]$, $\widehat{\varphi}_n^0(t) = \widehat{\varphi}_n^0(0) + (\widehat{\varphi}_n^0)'(0-)t$. Thus

$$
\begin{aligned}
\int_{\tau_{n,-}^0}^{\tau_{n,+}^0} \Delta_{LK,1}(t)\widehat{\varphi}_n^0(t)dt &= \int_{\tau_{n,-}^0}^0 \Delta_{LK,1}(t)\widehat{\varphi}_n^0(t)dt + \int_0^{\tau_{n,+}^0} \Delta_{LK,1}(t)\widehat{\varphi}_n^0(t)dt \\
&= \widehat{\varphi}_n^0(0) \int_{\tau_{n,-}^0}^0 \Delta_{LK,1}(t)dt + (\widehat{\varphi}_n^0)'(0-) \int_{\tau_{n,-}^0}^0 t\Delta_{LK,1}(t)dt \\
&\quad + \widehat{\varphi}_n^0(0) \int_0^{\tau_{n,+}^0} \Delta_{LK,1}(t)dt \\
&= 0.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\int_{\tau_{n,-}^0}^{\tau_{n,+}^0} & \Delta_{LK,1}(t)d_n(t)dt \\
&= \int_{\tau_{n,-}^0}^{\tau_{n,+}^0} \Delta_{LK,1}(t)\widehat{\varphi}_n^0(t)dt - \int_{\tau_{n,-}^0}^{\tau_{n,+}^0} \Delta_{LK,1}(t)\varphi_0(t)dt \\
&= -\int_{\tau_{n,-}^0}^{\tau_{n,+}^0} \Delta_{LK,1}(t)\varphi_0(t)dt \\
&= -\sum_{i=0}^2 \frac{\varphi_0^{(i)}(0)}{i!} \int_{\tau_{n,-}^0}^{\tau_{n,+}^0} t^i \Delta_{LK,1}(t)dt + \int_{\tau_{n,-}^0}^{\tau_{n,+}^0} \frac{\varepsilon_n(t)}{2!} t^2 \Delta_{LK,1}(t)dt \\
&= -\frac{\varphi_0^{(2)}(0)}{2!} \int_{\tau_{n,-}^0}^{\tau_{n,+}^0} t^2 \Delta_{LK,1}(t)dt + \int_{\tau_{n,-}^0}^{\tau_{n,+}^0} \frac{\varepsilon_n(t)}{2!} t^2 \Delta_{LK,1}(t)dt. \quad (8.17)
\end{aligned}
$$

Since an analogous statement holds for $\Delta_{RK,1}$, we have shown

$$
\int_{\tau_{n,-}^0}^{\tau_{n,+}^0} \Delta_1(t)d_n(t)dt = -\frac{\varphi_0^{(2)}(0)}{2!} \int_{\tau_{n,-}^0}^{\tau_{n,+}^0} t^2 \Delta_1(t)dt + \int_{\tau_{n,-}^0}^{\tau_{n,+}^0} \frac{\varepsilon_n(t)}{2!} t^2 \Delta_1(t)dt
$$

where $\|\varepsilon_n\|_\infty \to_p 0$. Thus

$$
\begin{aligned}
\int \Delta_1(x)(\widehat{f}_n^0 - f_0)(x)dx &= f_0(0)(1 + o_p(1)) \int \Delta_1(t)d_n(t) \\
&= f_0(0)(1 + o_p(1))^2 \frac{-\varphi_0(2)(0)}{2!} \int_{\tau_{n,-}^0}^{\tau_{n,+}^0} t^2 \Delta_1(t)dt.
\end{aligned}
$$

Lemma 8.3 shows that

$$
\int_{\tau_{n,-}^0}^{\tau_{n,+}^0} t^2 \Delta_1(t)dt \le -Kh_l^4 \tag{8.18}
$$

which yields (8.15), our desired conclusion:

$$
\int_{\tau_{n,-}^0}^{\tau_{n,+}^0} \Delta_1(t) \left(\widehat{f}_n^0 - f_0 \right)(t)dt \le -\frac{f_0(0)|\varphi_0^{(2)}(0)|}{2}Kh_l^4 + o(h_l^4).
$$

Now we show (8.14). First for a fixed $\delta > 0$, we know that $\tau_{n,+}^0$ and $\tau_{n,-}^0$ will be in $[-\delta, \delta]$ eventually, with high probability. Now we consider three families of functions, analogous to $\Delta_{LK,1}$, $\Delta_{RK,1}$, and $\Delta_{NK,1}$, respectively. Define $\Delta_{LK,1,b,c}$ by replacing $\tau_{n,-}^0$ with b and $\tau_{n,+}^0$ with c in (8.6). Define $\Delta_{NK,1,b,c}$ by replacing $\tau_{n,-}^0$ with b and $\tau_{n,+}^0$ with c in (8.10). Define $\mathcal{F}_{LK,b,R} := \{\Delta_{LK,b,y}|b < a < y, \ 0 \le y - b \le R\}$ and $\mathcal{F}_{NK,b,R} := \{\Delta_{NK,b,y}|b < a < y, \ 0 \le y - b \le R\}$ for $R > -b$. Define $\mathcal{F}_{RK,b,R}$ analogously to $\mathcal{F}_{LK,b,R}$. Let $\mathcal{F} = \mathcal{F}_{LK,b,R} \cup \mathcal{F}_{RK,b,R} \cup \mathcal{F}_{NK,b,R}$, and note that $\mathcal{F}$ is VC-class with VC-index of 4. Thus Theorem 2.6.7 on page 141 of van der Vaart and Wellner (1996) shows that the entropy bound condition in Lemma A.1 on page 2560 of Balabdaoui and Wellner (2007) holds for $\mathcal{F}$. Then the function $F_{b,R}(x)$, defined to be constant equal to $(7/4)R$ on $[b, b + R]$ and 0 otherwise, is an envelope for $\mathcal{F}$. That $F_{b,R}$ is an envelope is immediate for $\Delta \in \mathcal{F}_{NK,b,R}$ and for the setting where $\tau_{n,+}^0 < -\tau_{n,-}^0$ and $\Delta \in \mathcal{F}_{LK,b,R}$ (and analogously when $\tau_{n,+}^0 > -\tau_{n,-}^0$ and $\Delta \in \mathcal{F}_{RK,b,R}$). (For $\Delta \in \mathcal{F}_{NK,b,R}$, the longer interval has slope ± 1 and the other interval has opposite sign slope. For the case $\tau_{n,+}^0 < -\tau_{n,-}^0$ and $\Delta \in \mathcal{F}_{LK,b,R}$, the interval $[\tau_{n,-}^0, \tau_{n,-}^0/2]$ has slope 1 and the slope on the rest is opposite sign (and analogously for $\mathcal{F}_{RK,b,R}$).) For the case $\tau_{n,+}^0 \ge -\tau_{n,-}^0$ and $\Delta \in \mathcal{F}_{LK,b,R}$, we need only note

$$
0 \le m_2 = \frac{\tau_{n,+}^0}{-\tau_{n,-}^0} \left(\frac{-9 - 3\frac{\tau_{n,+}^0}{-\tau_{n,-}^0}}{1 - 5\frac{\tau_{n,+}^0}{-\tau_{n,-}^0}} \right) \le 3,
$$

so that $(-\tau_{n,-}^0/4)m_2 \le (3/4)\tau_{n,+}^0$. Next, we compute the integral of the envelope squared

$$
EF_{b,R}^2(X) = \int_b^{b+R} R^2 f_0(x)dx \le \|f_0\|_\infty R^3,
$$

where $\|f_0\|_\infty$ is the supremum over $\mathbb{R}$ of the (log-concave) density f_0, and is thus universal across b and R. Thus, we can conclude from Lemma A.1 on page 2560 of Balabdaoui and Wellner (2007) that for $\varepsilon > 0$ and with $s = 2$ and $d = 2$,

$$
\begin{aligned}
\left| \int \Delta_1 d(\mathbb{F}_n - F_0) \right| &\leq \varepsilon (\tau_{n,+}^0 - \tau_{n,-}^0)^4 + O_p(n^{-4/5}), \\
&\leq \varepsilon 2^4 h_l^4 + O_p(n^{-4/5})
\end{aligned}
$$

as desired. $\qquad\square$

Lemma 8.3. *For $h_l = \max(\tau_{n,+}^0 - m, m - \tau_{n,-}^0)$, $h_s = \min(\tau_{n,+}^0 - m, m - \tau_{n,-}^0)$ and Δ_1 defined by (8.11), we have*

$$
\int \Delta_1(t)(t-m)^2 dt \leq -K h_l^4,
$$

for some $K > 0$.

Proof. First, we consider $\Delta_{LK,1}$, assume without loss of generality that $m = 0$, and assume $\tau_{n,+}^0 > -\tau_{n,-}^0$. Direct computation shows

$$
\begin{aligned}
\int_{\tau_{n,-}^0}^{\tau_{n,+}^0} t \Delta_{LK,1}(t) dt &= -\frac{5}{12} \left(\frac{(\tau_{n,+}^0 - \tau_{n,-}^0)(\tau_{n,+}^0)^3}{5\tau_{n,+}^0 + \tau_{n,-}^0} \right) \\
&\leq -\frac{5}{12} \left(\frac{(\tau_{n,+}^0 - \tau_{n,-}^0)(\tau_{n,+}^0)^3}{5(\tau_{n,+}^0)} \right)
\end{aligned}
$$

and this is bounded above by

$$
-\frac{5}{12} \frac{(\tau_{n,+}^0)^4}{5\tau_{n,+}^0} = -\frac{(\tau_{n,+}^0)^3}{12} = -\frac{h_l^3}{12},
$$

since $\tau_{n,+}^0 > -\tau_{n,-}^0$.

We break $\int_{\tau_{n,-}^0}^{\tau_{n,+}^0} t^2 \Delta_{LK,1}(t) dt$ into two pieces. First we see

$$
\int_{\tau_{n,-}^0}^{0} t^2 \Delta_{LK,1}(t) dt = \frac{-3(\tau_{n,-}^0)^4 \tau_{n,+}^0 - 19(-(\tau_{n,-}^0))^3 (\tau_{n,+}^0)^2}{96(5\tau_{n,+}^0 + \tau_{n,-}^0)} < 0.
$$

Next we see that

$$
\begin{aligned}
\int_{0}^{\tau_{n,+}^0} t^2 \Delta_{LK,1}(t) dt &= \frac{-3(-\tau_{n,-}^0)(\tau_{n,+}^0)^4 - 5(\tau_{n,+}^0)^5}{12(5\tau_{n,+}^0 + \tau_{n,-}^0)} \\
&\leq -\frac{3(-\tau_{n,-}^0)(\tau_{n,+}^0)^4 + 5(\tau_{n,+}^0)^5}{12 \cdot 5(\tau_{n,+}^0)} \leq -\frac{(\tau_{n,+}^0)^4}{12}.
\end{aligned}
$$

Thus,

$$
\int_{\tau_{n,-}^0}^{\tau_{n,+}^0} t^2 \Delta_{LK,1}(t) dt \leq -\frac{(\tau_{n,+}^0)^4}{12} = -\frac{h_l^4}{12},
$$

as desired.

Now we consider $\Delta_{LK,1}$ when $\tau_{n,+}^0 < -\tau_{n,-}^0$, and again split the computation into two pieces. First we see

$$\int_0^{\tau_{n,+}^0} t^2 \Delta_{LK,1}(t)dt = -\frac{2(\tau_{n,-}^0)^2(\tau_{n,+}^0)^3 - 3(\tau_{n,+}^0)^5}{12(5\tau_{n,+}^0 - 2\tau_{n,-}^0)}.$$

Next we find that

$$\int_{\tau_{n,-}^0}^a t^2 \Delta_{LK,1}(t)dt = -\frac{(-\tau_{n,-}^0)^5 + 5(\tau_{n,-}^0)^4\tau_{n,+}^0}{48(5\tau_{n,+}^0 - 2\tau_{n,-}^0)}$$

so that

$$\int_{\tau_{n,-}^0}^{\tau_{n,+}^0} t^2 \Delta_{LK,1}(t)dt = \frac{(\tau_{n,-}^0)^5 - 5(\tau_{n,-}^0)^4\tau_{n,+}^0 - 8(\tau_{n,-}^0)^3(\tau_{n,+}^0)^3 + 12(\tau_{n,+}^0)^5}{48(5\tau_{n,+}^0 - 2\tau_{n,-}^0)}$$

$$\leq -\frac{(-\tau_{n,-}^0)^4}{48 \cdot 7} \tag{8.19}$$

and it remains only to prove the last inequality. To do this, let $a \equiv \tau_{n,-}^0 < 0 < \tau_{n,+}^0 \equiv b$ where $-a \geq b$. Thus we want to show that

$$K(a,b) = -5a^5 + 30a^4b + 56a^2b^3 - 84b^5 \geq 0$$

for all $a < 0 < b$ with $-a \geq b$. But this holds if and only if $J(v,c) \equiv K(-v, cv) \geq 0$ holds for all $v \geq 0$ and $0 \leq c \leq 1$. But

$$J(v,c) = K(-v, cv) \geq v^5(5c^5 + 30c^5 + 56c^5 - 84c^5) = v^5 c^5 \cdot 7 \geq 0.$$

Thus (8.19) holds.

Identical calculations hold for $\Delta_{RK,1}$. Now we examine $\Delta_{NK,1}$. Direct computation shows

$$\int_{\tau_{n,-}^0}^{\tau_{n,+}^0} t^2 \Delta_{NK,1}(t)dt = -\frac{1}{12}(h_s^3 h_l + h_l^4) \leq -\frac{1}{12}h_l^4. \qquad \square$$

8.2.2. Proofs for mode-constrained limit process

This entire section is devoted to the proof of Theorem 4.2, and thus throughout this section we take the definitions and assumptions as given in that theorem.

Proof of Lemma 4.3. If $0 \in \widehat{S}^0$ then one of τ_+^0 or τ_-^0 is 0, because then $(\widehat{\varphi}^0)^{(2)}$ is not constant-equal-to-0 on any open neighborhood of 0; if it is not constant-equal-to-0 on $[0, \delta)$ for any $0 < \delta$ then $(\widehat{\varphi}^0)'$ is also not constant on $[0, \delta)$, so since $0 \geq (\widehat{\varphi}^0)'(0+)$, we have $(\widehat{\varphi}^0)'(\delta+) < 0$ for all $\delta > 0$, i.e. $\tau_+^0 = 0$ (of course $\tau_+^0 \geq 0$ since 0 is the mode of $\widehat{\varphi}^0$). Similarly, if $(\widehat{\varphi}^0)^{(2)}$ is not 0 on any neighborhood below 0, then $\tau_-^0 = 0$.

Now, if $0 \notin \widehat{S}^0$ then $\tau_-^0 = \sup \widehat{S}^0 \cap (-\infty, 0]$ and $\tau_+^0 = \inf \widehat{S}^0 \cap [0, \infty)$. If $0 \in \widehat{S}^0$ then we have shown that one of $\tau_\pm^0$ equals 0. Thus, in either case, it is clear that $(\widehat{\varphi}^0)^{(2)}(t\pm) = 0$ for $0 \leq t < \tau_+^0$ and for $\tau_-^0 < t \leq 0$. Thus, regardless of whether $0 \in \widehat{S}^0$, $(\widehat{\varphi}^0)^{(2)}$ and so $(\widehat{\varphi}^0)'$ are constant and equal to 0 on (τ_-^0, τ_+^0), i.e. (4.11) holds (and so (τ_-^0, τ_+^0) is the modal interval of $\widehat{\varphi}^0$). $\qquad\square$

The next lemma gives the sense in which $\widehat{\varphi}^0$ is piecewise affine.

Lemma 8.4. *Assume that H_L, H_R, and $\widehat{\varphi}^0$ are as in Theorem 4.2. Define*

$$S_L = \{t \leq 0 : H_L(t) = Y_L(t), H_L'(t) = Y_L'(t)\},$$
$$S_R = \{t \geq 0 : H_R(t) = Y_R(t), H_R'(t) = Y_R'(t)\},$$
$$\text{and } S^0 = S_L \cup S_R \cup \{0\}. \tag{8.20}$$

Then $(\widehat{\varphi}^0)'$ is a monotonically nonincreasing function and the 'bend points' of $\widehat{\varphi}^0$, $\widehat{S}^0$, defined in (4.2), satisfy $\widehat{S}^0 \subset S^0$. Additionally, with probability 1 the following statements hold. The sets S_L, S_R, and S^0 are all closed and have Lebesgue measure 0. For any fixed $t \neq 0$, $t \notin S^0$ and so $(H^0)^{(3)}(t)$ is well-defined.

The lemma says that for any knot $\tau \leq \tau_+^0$, if $\tau < 0$ then $\tau \in S_L$. Similarly if $\tau \geq \tau_-^0$ and $\tau > 0$ then $\tau \in S_R$. It is possible but not guaranteed that 0 is a knot and lies in either S_L or S_R.

Proof. By Theorem 4.2, displays (4.8) and (4.9), $H^0 - Y^0 \leq 0$, which allows us to conclude

$$\{t < 0 : H_L(t) = Y_L(t)\} = \{t < 0 : H_L(t) = Y_L(t), H_L'(t) = Y_L'(t)\},$$
$$\{t > 0 : H_R(t) = Y_R(t)\} = \{t > 0 : H_R(t) = Y_R(t), H_R'(t) = Y_R'(t)\}; \tag{8.21}$$

the first line follows since $H_L - Y_L$ is differentiable on $(-\infty, 0)$, and a differentiable function has derivative 0 at a local maximum (see, e.g., Dieudonné (1969), page 153, Problem 3, part (a)). The same argument applies to the second line of (8.21).

Now, the following argument holds with probability 1 and for any fixed $c > 0$. On $[0, c]$, H_R has a bounded second derivative, so that there exists a constant $a > 0$ such that $\tilde{H}_R(t) := H_R(t) + at^2$ is convex on $[0, c]$. Let $\tilde{Y}_R(t) := Y_R(t) + at^2$. Now, $Y_R + A = Y$ for an affine function A and where $Y(t) = \int_0^t \int_0^u dX(z) du$. Let $\tilde{Y}(t) = Y(t) + at^2 = \tilde{Y}_R(t) + A(t)$ so that

$$\{x \in [0, c] : H_R(x) = Y_R(x)\} = \{x \in [0, c] : \tilde{H}_R(x) + A(x) = \tilde{Y}_R(x) + A(x)\}$$
$$= \{x \in [0, c] : \tilde{H}_R(x) + A(x) = \tilde{Y}(x)\} \tag{8.22}$$

We also have $\tilde{H}_R + A \leq \tilde{Y}_R + A = \tilde{Y}$ by (4.9), so that, letting $M_{\tilde{Y}}$ be the greatest convex minorant of $\tilde{Y}$ on $[0, c]$, we have $\tilde{H}_R + A \leq M_{\tilde{Y}} \leq \tilde{Y}$, since $\tilde{H}_R + A$ is convex and below $\tilde{Y}$.

Let $T = \{x \in [0, c] : M_{\tilde{Y}}(x) = \tilde{Y}(x)\}$. By the proof of Corollary 2.1 of Groeneboom, Jongbloed and Wellner (2001a) (see also Definition 1 and Theorem 1 of Sinai (1992)), T is a (Cantor-type) set which has Lebesgue measure 0;

(8.22) is contained in T, and thus (by (8.21) and) by letting $c \to \infty$, we see that S_R is contained in a set which has Lebesgue measure 0. Finally, S_R is closed because $H_R - Y_R$ and $(H_R - Y_R)'$ are both continuous functions. By an analogous argument, we can conclude that S_L is closed and has Lebesgue measure 0 and thus also S^0 is closed and has Lebesgue measure 0. By (4.10),

$$\int_{\{t \in (-\infty, \tau_-^0] : H_L(t) \neq Y_L(t)\}} d(\widehat{\varphi}^0)'(t) = 0 = \int_{\{t \in [\tau_+^0, \infty) : H_R(t) \neq Y_R(t)\}} d(\widehat{\varphi}^0)'(t),$$

(8.23)

(regardless of whether one of τ_+^0 or τ_-^0 is 0 or not).

Thus we now conclude that $\widehat{S}^0 \subset S^0$ as follows. If τ is an element of the set $\widehat{S}^0$ then for any $\varepsilon > 0$ $(\widehat{\varphi}^0)'(\tau - \varepsilon, \tau + \varepsilon) < 0$ (here $(\widehat{\varphi}^0)'$ refers to the signed measure corresponding to $(\widehat{\varphi}^0)'$). This is by the definition of derivative; since $(\widehat{\varphi}^0)'$ is nonincreasing, $\delta \mapsto (\widehat{\varphi}^0)'(\tau + \delta) - (\widehat{\varphi}^0)'(\tau - \delta) \leq 0$ is nonincreasing. Thus if

$$\frac{(\widehat{\varphi}^0)'(\tau + \delta) - (\widehat{\varphi}^0)'(\tau - \delta)}{2\delta}$$

(8.24)

does not converge to 0 as $\delta \searrow 0$, then for all $\varepsilon > 0$ there is $0 < \delta < \varepsilon$ such that (8.24) is $<$ const., i.e.,

$$(\widehat{\varphi}^0)'(\tau + \varepsilon) - (\widehat{\varphi}^0)'(\tau - \varepsilon) \leq (\widehat{\varphi}^0)'(\tau + \delta) - (\widehat{\varphi}^0)'(\tau - \delta) < 0. \qquad (8.25)$$

Since $H_R - Y_R$ is continuous on all of $\mathbb{R}$, if $H_R(\tau) - Y_R(\tau) < 0$ for $\tau \geq \tau_+^0$, then on a neighborhood $(\tau - \varepsilon, \tau + \varepsilon)$ for some $\varepsilon > 0$ we have $H_R - Y_R < 0$ and so the integral on the right-hand side of (8.23) is strictly less than 0. Thus if $\tau \geq \tau_+^0$, then $(H_R - Y_R)(\tau) = 0$. For $\tau > 0$, this implies that $\tau \in S_R$ by (8.21). Similarly, if $\tau \leq \tau_-^0$ and $\tau < 0$, then $(H_L - Y_L)(\tau) = 0$ and $\tau \in S_L$. We have $(\tau_-^0, \tau_+^0) \cap \widehat{S}^0 = \emptyset$ by Lemma 4.3, so we have shown that if $\tau \in \widehat{S}^0$, then τ is either 0 (if one of τ_-^0 or τ_+^0 is 0) or is in S_L or S_R, i.e. $\tau \in S^0$.

Now, by the proof of Theorem 1 of Sinai (1992), any fixed point $t \geq 0$ belongs to $T \supset S_R$ with probability zero. An analogous statement holds for $t \leq 0$ and S_L. Thus, if $t \neq 0$, $t \notin S^0$ and so $(H_L)^{(2)}$ is concave and affine in a neighborhood of t, so $(H_L)^{(3)}(t)$ is well-defined. $\qquad \square$

By Lemma 8.4,

$$H_L(\tau_L) = Y_L(\tau_L) = 0, \text{ and } H_L'(\tau_L) = Y_L'(\tau_L) = 0. \qquad (8.26)$$

This is because, by its definition, either $\tau_L < 0$, in which case $\tau_L \in S_L$, or there is a sequence $\{\tau_{L,n}^0\} \subset S_L$ with $\tau_{L,n}^0 < 0$ for all n. In this latter case, since $H_L - Y_L$ and $H_L' - Y_L'$ are both continuous, we still conclude that $H_L(\tau_L) = Y_L(\tau_L) = 0$, and $H_L'(\tau_L) = Y_L'(\tau_L) = 0$. Analogously, $H_R(\tau_R) = Y_R(\tau_R)$ and $H_R'(\tau_R) = Y_R'(\tau_R)$. This suggests the following definitions:

$$F_L(u) = \int_u^{\tau_L} \widehat{\varphi}^0(v) dv \quad \text{and} \quad F_R(u) = \int_{\tau_R}^u \widehat{\varphi}^0(v) dv, \qquad (8.27)$$

$$X_L(u) = \int_u^{\tau_L} dX(v) \quad \text{and} \quad X_R(u) = \int_{\tau_R}^u dX(v). \qquad (8.28)$$

It is then true by definition that

$$Y_L(t) = \int_t^{\tau_L} X_L(u)du, \quad \text{and} \quad Y_R(t) = \int_{\tau_R}^t X_R(u)du. \tag{8.29}$$

It is furthermore true that

$$H_L(t) = \int_t^{\tau_L} \int_u^{\tau_L} \widehat{\varphi}^0(v)dvdu = \int_t^{\tau_L} F_L(u)du, \tag{8.30}$$

$$H_R(t) = \int_{\tau_R}^t \int_{\tau_R}^u \widehat{\varphi}^0(v)dvdu = \int_{\tau_R}^t F_R(u)du, \tag{8.31}$$

since $H_L(t) = \int_t^{\tau_L} \int_u^{\tau_L} \widehat{\varphi}^0(v)dvdu + A_L(t)$ where $A_L(t)$ is an affine function, and we just verified in (8.26) that $A_L(t) \equiv 0$. An analogous statement holds for H_R.

Remark 8.5. Note that since S^0 is closed, τ_+, τ_- are both elements of S^0. Also, it is possible for $H_L^{(3)}$ to have a point of increase at 0 without $H_L'(0) = Y_L'(0)$, since the inequality (4.8) only holds on $(-\infty, 0]$, so not in an open neighborhood of 0. Similarly for $H_R^{(3)}$. This is why we add the point 0 to S^0.

Lemma 8.4 suggests that we can think of $\widehat{\varphi}^0$ as being piecewise affine (with a potentially uncountable number of knot points), because with probability 1, the union of the open intervals on which $\widehat{\varphi}^0$ is affine has full Lebesgue measure on the real line (meaning its complement has Lebesgue measure 0). For $t \in \mathbb{R}$, we let $\tau_+^0(t)$ be the first knot larger than t, and analogously for $\tau_-^0(t)$,

$$\tau_+^0(t) = \inf\left(\widehat{S}^0 \cap [t, \infty)\right) \quad \text{and} \quad \tau_-^0(t) = \sup\left(\widehat{S}^0 \cap (-\infty, t]\right). \tag{8.32}$$

Lemma 8.6. *We again assume the full setup of Theorem 4.2. Then, for any (fixed or random) $T \geq 0$, with probability 1 there are 'knot points' $\tau_+^0(T)$ and $\tau_-^0(T)$ in $\widehat{S}^0$.*

Proof. We fix $T \geq 0$, and we will show that there exists $\tau_+(T) \in \widehat{S}^0$ with $\tau_+(T) > T \geq 0$. We assume for contradiction that $\widehat{\varphi}^0$ has no knots on (T, ∞), and thus is linear. Thus H_R is cubic on $[T, \infty)$, so can be written as $H_R(t) = \sum_{i=0}^3 A_i(t - T)^i$ for some random A_i. By definition, we have

$$Y_R(t) = \int_{\tau_R}^t X_R(u)du = \int_{\tau_R}^t (X(u) - X(\tau_R))du$$

$$= \int_0^t X(u)du - \int_0^{\tau_R} X(u)du - (t - \tau_R)X(\tau_R).$$

In other words, $Y_R(t)$ is $\int_0^t X(u)du = V(t) + t^4$ plus a random affine function, where $V(t) = \int_0^t W(u)du$. Thus we can write

$$Y_R(t) - H_R(t) = V(t) + t^4 - \sum_{i=0}^3 B_i(t - T)^i,$$

for some new random coefficients, B_i (where only for $i = 0, 1$ are B_i not equal to A_i). Now, let $\varphi(t) = \sqrt{\frac{2}{3}t^3 \log\log t}$. Then by page 1714 of Lachal (1997) (or

from page 238 of Watanabe (1970)), we know that almost surely

$$\limsup_{t \to \infty} \frac{\int_0^t W(u)du}{\varphi(t)} = 1.$$

Thus,

$$\frac{Y_R(t) - H_R(t)}{\varphi(t)} = \frac{V(t)}{\varphi(t)} + \frac{t^4 - \sum_{i=0}^3 B_i(t-T)^i}{\varphi(t)},$$

which gets larger than 0 for t large enough, as it is almost surely bounded below by a quadratic polynomial (with positive first coefficient) minus 1. This contradicts the fact that $Y_R(t) - H_R(t) \leq 0$ for all t, so we are done. Our argument applies with probability 1 to any $T \geq 0$, and thus to the entire sample space of any random $T \geq 0$. An identical argument works for showing there exists a knot less than $-T$. $\qquad \square$

We will not speak of $\widehat{\varphi}^0$ as a minimizer of an objective function, but we will instead show that for acceptable Δ perturbations that $\int \Delta(t)(\widehat{\varphi}^0(t)dt - dX(t)) \geq 0$, i.e. $\widehat{\varphi}^0$ behaves as we would expect a minimizer to behave.

Proposition 8.7. *We assume the full setup of Theorem 4.2. Let $\Delta : \mathbb{R} \to \mathbb{R}$ be concave with maximum at 0. If $a, b \in \widehat{S}^0$ with $-\infty < a < 0 < b < \infty$ then*

$$\int_a^b \Delta(t)(\widehat{\varphi}^0(t)dt - dX(t)) \geq 0, \tag{8.33}$$

and thus, by Lemma 8.6, $\limsup_{a \to \infty} \int_{-a}^a \Delta(t)(\widehat{\varphi}^0(t)dt - dX(t)) \geq 0$.

Proof. We use the notation $g(a, b] = g(b) - g(a)$ here. We have

$$\int_a^b \Delta(t)(\widehat{\varphi}^0(t)dt - dX(t))$$
$$= -\int_a^0 \Delta(t)(dF_L(t) - dX_L(t)) + \int_0^b \Delta(t)(dF_R(t) - dX_R(t))$$
$$= -\left[(\Delta(F_L - X_L))(a, 0] - \int_a^0 ((F_L - X_L)\Delta')(t)dt\right]$$
$$+ (\Delta(F_R - X_R))(0, b] - \int_0^b ((F_R - X_R)\Delta')(t)dt.$$

By Lemma 8.4, $a \in S_L$, $b \in S_R$, and since neither a nor b is 0, we have $(F_R - X_R)(b) = 0 = (F_L - X_L)(a)$ and we recall (4.10). Also recalling that $(H_L - Y_L)' = -(F_L - X_L)$, we see that the above display equals

$$-\Delta(0)((F_L - X_L)(0) + (F_R - X_R)(0))$$
$$- \left[((H_R - Y_R)\Delta'(\cdot+))(0, b] - \int_0^b (H_R - Y_R)(t)d\Delta'(t)\right]$$

$$- \left[((H_L - Y_L)\Delta'(\cdot -))(a, 0] - \int_a^0 (H_L - Y_L)(t)d\Delta'(t) \right]$$

$$= -\Delta(0) \left(\int_0^{\tau_L} (\widehat{\varphi}^0(t)dt - dX(t)) + \int_{\tau_R}^0 (\widehat{\varphi}^0(t)dt - dX(t)) \right)$$

$$+ (H_R - Y_R)(0)\Delta'(0+) - (H_L - Y_L)(0)\Delta'(0-)$$

$$+ \int_0^b (H_R - Y_R)(t)d\Delta'(t) + \int_a^0 (H_L - Y_L)(t)d\Delta'(t)$$

$$\geq 0,$$

where the inequality follows because each of the three lines in the final expression is ≤ 0, as follows. The first line is equal to 0 by (4.7); the third line is ≥ 0 by (4.8) and (4.9), and the fact that Δ is concave so Δ' is monotonically nonincreasing so $d\Delta'$ is a nonpositive measure; similarly, the second line is ≥ 0 because Δ has maximum at 0, so that $(H_R - Y_R)(0)$, $\Delta'(0+)$, $(H_L - Y_L)(0)$, and $-\Delta'(0-)$ are nonpositive. $\qquad\square$

The above proof can be extended to Δ such that $\widehat{\varphi}^0(t) + \varepsilon\Delta(t) \in \mathcal{G}^0$, where $\mathcal{G}^0$ is the set of concave functions with maximum at 0, but we will not need this per se. Rather, in the next result we will express the same idea by showing for knots $a < 0 < b$ that $\int_a^b \widehat{\varphi}^0(t) \left(\widehat{\varphi}^0(t)dt - dX(t) \right) = 0$, and re-express this via integration by parts formulae.

Proposition 8.8. *We again assume the full setup of Theorem 4.2 and assume that $a, b \in \widehat{S}^0$, and $a < 0 < b$. Then*

$$\int_a^b \widehat{\varphi}^0(t)(\widehat{\varphi}^0(t)dt - dX(t))$$

$$= \int_a^0 ((F_L - X_L)(\widehat{\varphi}^0)')(t)dt - \int_0^b ((F_R - X_R)(\widehat{\varphi}^0)')(t)dt$$

$$= \int_a^{\tau_L} (H_L - Y_L)(t)d(\widehat{\varphi}^0)'(t) + \int_{\tau_R}^b (H_R - Y_R)(t)d(\widehat{\varphi}^0)'(t)$$

$$= 0.$$

Proof. Since $a < 0 < b$, we again have $F_L(a) - X_L(a) = 0 = F_R(b) - X_R(b)$ by Lemma 8.4, so

$$\int_a^b \widehat{\varphi}^0(t)(\widehat{\varphi}^0(t)dt - dX(t))$$

$$= -\int_a^0 \widehat{\varphi}^0(t)d(F_L(t) - X_L(t)) + \int_0^b \widehat{\varphi}^0(t)d(F_R(t) - X_R(t))$$

which equals

$$-\left[(\widehat{\varphi}^0(F_L - X_L))(a, 0] - \int_a^0 ((F_L - X_L)(\widehat{\varphi}^0)')(t)dt \right]$$

$$+ (\widehat{\varphi}^0(F_R - X_R))(0, b] - \int_0^b ((F_R - X_R)(\widehat{\varphi}^0)')(t)dt$$

which equals

$$- \left[(\widehat{\varphi}^0(F_L - X_L))(0) - \int_a^0 ((F_L - X_L)(\widehat{\varphi}^0)')(t)dt \right]$$

$$- (\widehat{\varphi}^0(F_R - X_R))(0) - \int_0^b ((F_R - X_R)(\widehat{\varphi}^0)')(t)dt$$

$$= \int_a^0 ((F_L - X_L)(\widehat{\varphi}^0)')(t)dt - \int_0^b ((F_R - X_R)(\widehat{\varphi}^0)')(t)dt,$$

where the last equality is by (4.7). Since $\widehat{\varphi}^0$ is constant on (τ_-^0, τ_+^0), and using the notation $g[c, d] = g(d) - g(c-)$, we can write the last expression above as

$$\int_a^{\tau_-^0} ((F_L - X_L)(\widehat{\varphi}^0)')(t)dt - \int_{\tau_+^0}^b ((F_R - X_R)(\widehat{\varphi}^0)')(t)dt$$

$$= ((\widehat{\varphi}^0)'(H_L - Y_L))[a, \tau_-^0] - \int_{[a,\tau_-^0]} (H_L - Y_L)(t)d(\widehat{\varphi}^0)'(t)$$

$$- \left[((\widehat{\varphi}^0)'(H_R - Y_R))[\tau_+^0, b] - \int_{[\tau_+^0,b]} (H_R - Y_R)(t)d(\widehat{\varphi}^0)'(t) \right]$$

which equals

$$- \int_{[a,\tau_-^0]} (H_L - Y_L)(t)d(\widehat{\varphi}^0)'(t) + \int_{[\tau_+^0,b]} (H_R - Y_R)(t)d(\widehat{\varphi}^0)'(t) = 0,$$

using integration by parts formula (see Lemma 9.1) for the first equality since $H_L - Y_L$ and $H_R - Y_R$ are continuous, and using (4.10) for the third equality. The second equality follows from Lemma 8.4, since a knot a and limit of knots τ_- are elements of S_L, and similarly b, τ_+ are elements of S_R. □

Next we prove a representation lemma, analogous to the midpoint result for the unconstrained (and compact support) case in Lemma 2.3 on page 1631 of Groeneboom, Jongbloed and Wellner (2001a).

Lemma 8.9. *We again assume the full setup of Theorem 4.2. Let $\tau_1, \tau_2 \in \widehat{S}^0$ be such that $\widehat{\varphi}^0$ is affine on $[\tau_1, \tau_2]$, and let $t \in [\tau_1, \tau_2]$. For any function g, we define $\Delta g = g(\tau_2) - g(\tau_1)$ and $\bar{g} = \frac{g(\tau_1)+g(\tau_2)}{2}$, including in particular, $\Delta\tau = \tau_2 - \tau_1$ and $\bar{\tau} = (\tau_1 + \tau_2)/2$. Then if $0 < \tau_1 < \tau_2$, we can conclude*

$$H_R(t) = \frac{(Y_R(\tau_2)(t - \tau_1) + Y_R(\tau_1)(\tau_2 - t))}{\Delta\tau}$$
$$- \frac{1}{2}\left(\frac{\Delta X_R}{\Delta\tau} + \frac{4}{(\Delta\tau)^3}(\bar{X}_R\Delta\tau - \Delta Y_R)(t - \bar{\tau}) \right)(t - \tau_1)(\tau_2 - t), \tag{8.34}$$

and thus

$$H_R(\bar{\tau}) = \bar{Y}_R - \frac{1}{8}\Delta X_R \Delta\tau. \tag{8.35}$$

If $\tau_1 < \tau_2 < 0$, we can conclude

$$\begin{aligned}
H_L(t) = {} & \frac{Y_L(\tau_2)(t - \tau_1) + Y_L(\tau_1)(\tau_2 - t)}{\Delta\tau} \\
& - \frac{1}{2}\left(\frac{-\bar{X}_L}{\Delta\tau} + \frac{4}{(\Delta\tau)^3}(-\bar{X}_L \Delta\tau - \Delta Y_L)(t - \bar{\tau})\right)(t - \tau_1)(\tau_2 - t),
\end{aligned} \tag{8.36}$$

and thus

$$H_L(\bar{\tau}) = \bar{Y}_L + \frac{1}{8}\Delta X_L \Delta\tau. \tag{8.37}$$

Proof. We assume that on $[\tau_1, \tau_2]$, that $\widehat{\varphi}^0$ is linear and thus H_L and H_R are cubic polynomials. Thus, taking $\tau_1 < \tau_2 < 0$, H_L is defined by its values and its derivative's values at τ_i, for $i = 1, 2$. Thus, if we name the polynomial on the right hand side of (8.36) P_L, it suffices to check that $P_L(\tau_i)$ and $P'_L(\tau_i)$ equal $H_L(\tau_i)$ and $H'_L(\tau_i)$, respectively, for $i = 1, 2$, to conclude that $H_L(t) = P_L(t)$ for $t \in [\tau_1, \tau_2]$. We know that $H_L(\tau_i) = Y_L(\tau_i)$ by (4.10) and it is immediate that $P_L(\tau_i) = Y_L(\tau_i)$, so we only need to check the derivative values. To differentiate, we denote

$$A(t) = \frac{1}{2}\left(\frac{-\bar{X}_L}{\Delta\tau} + \frac{4}{(\Delta\tau)^3}(-\bar{X}_L \Delta\tau - \Delta Y_L)(t - \bar{\tau})\right),$$

so that

$$P_L(t) = \frac{Y_L(\tau_2)(t - \tau_1) + Y_L(\tau_1)(\tau_2 - t)}{\Delta\tau} - A(t)(t - \tau_1)(\tau_2 - t),$$

and

$$P'_L(t) = \frac{Y_L(\tau_2) - Y_L(\tau_1)}{\Delta\tau} - A'(t)(t - \tau_1)(\tau_2 - t) - A(t)((\tau_2 - t) - (t - \tau_1)),$$

so that

$$\begin{aligned}
P'_L(\tau_1) &= \frac{\Delta Y_L}{\Delta\tau} - A(\tau_1)\Delta\tau \\
&= \frac{\Delta Y_L}{\Delta\tau} - \frac{1}{2}\left\{\frac{-\Delta X_L}{\Delta\tau} + \frac{4}{(\Delta\tau)^3}(-\bar{X}_L \Delta\tau - \Delta Y_L)\left(\frac{-\Delta\tau}{2}\right)\right\}\Delta\tau \\
&= -X_L(\tau_1).
\end{aligned}$$

This equals $H'_L(\tau_1)$, as desired, since $H'_L(\tau_1) = Y'_L(\tau_1)$ by Lemma 8.4 since τ_1 is strictly less than 0, and $Y'_L(\tau_1) = -X_L(\tau_1)$. Similarly, $P'_L(\tau_2) = -X_L(\tau_2)$ and, letting P_R be the polynomial on the right hand side of (8.36), $P'_R(\tau_i) = X_R(\tau_i)$ and $P_R(\tau_i) = Y_R(\tau_i)$. Then (8.35) and (8.37) follow immediately. $\square$

Next, we show a tightness-type of results for the bend points. Recall the definition (8.32) of $\tau_-^0(t)$ and $\tau_+^0(t)$.

Lemma 8.10. *Let the assumptions of Theorem 4.2 hold. Then, for all $\varepsilon > 0$ there exists M_ε such that for all $t > 0$,*

$$P(\tau_+^0(t) > t + M_\varepsilon) < \varepsilon, \tag{8.38}$$

$$P(\tau_-^0(-t) < -t - M_\varepsilon) < \varepsilon, \tag{8.39}$$

$$P((t - M_\varepsilon) \vee 0 \leq \tau_-^0(t) \vee 0) > 1 - \varepsilon, \tag{8.40}$$

$$P(\tau_+^0(-t) \wedge 0 \leq (-t + M_\varepsilon) \wedge 0) > 1 - \varepsilon, \tag{8.41}$$

where M_ε does not depend on t.

Proof. We will show for all $t, \varepsilon > 0$ there exists $M = M_\varepsilon$ such that $P(\tau_+^0(t) > t + M) < \varepsilon$. The statement for $\tau_-^0(-t)$ is analogous. By Lemma 8.6, for any t we can find $\tau_2 \in \widehat{S}^0$ where $\tau_2 < \infty$ is taken to be $\tau_+^0(t)$; similarly, we can take $t \equiv t_\varepsilon$ large enough such that with probability $1 - \varepsilon$ there exists a knot $0 < \tau_1 < t$. To match notation up with Lemma 8.9, we will define Δg and $\bar{g}$, for any function g, as in the lemma. Since $\widehat{\varphi}^0$ is affine on $[\tau_1, \tau_2]$, Lemma 8.9 allows us to conclude that $Y_R(\bar{\tau}) \leq H_R(\bar{\tau}) = \bar{Y}_R - \Delta X_R \Delta \tau / 8$ which is if and only if

$$Y(\bar{\tau}) \leq \bar{Y} - \frac{1}{8} \Delta X \Delta \tau, \tag{8.42}$$

where $Y(t) = \int_0^t X(u) du = V(t) + t^4$ and $V(t) = \int_0^t W(u) du$. The "if and only if" follows because $Y_R(t) = Y(t) + A(t)$ where $A(t)$ is a random affine function. Since for any affine function $A(\bar{\tau}) =: \bar{A}$, we see that

$$Y_R(\bar{\tau}) - \bar{Y}_R = Y(\bar{\tau}) + A(\bar{\tau}) - (\bar{Y} + \bar{A}) = Y(\bar{\tau}) - \bar{Y}.$$

Since ΔX trivially equals ΔX_R, we have shown (8.42). Let $M_\varepsilon > 0$ and let B_t be the event $\{0 < \tau_1 < t, \tau_2 > t + M_\varepsilon\}$. We then see

$$P(\tau_2 > t_\varepsilon + M_\varepsilon) \leq P(\tau_1 \leq 0) + P(B_{t_\varepsilon})$$
$$\leq \varepsilon + P\left(Y_R(\bar{\tau}) \leq H_R(\bar{\tau}) = \bar{Y}_R - \frac{1}{8} \Delta X_R \Delta \tau, B_{t_\varepsilon}\right)$$
$$\leq 2\varepsilon, \tag{8.43}$$

where we now show that the last inequality follows from page 1633 in the proof of Lemma 2.4 in Groeneboom, Jongbloed and Wellner (2001a). We have already noted that $Y_R(\bar{\tau}) \leq \bar{Y}_R - \frac{1}{8} \Delta X_R \Delta \tau$ if and only if $Y(\bar{\tau}) \leq \bar{Y} - \frac{1}{8} \Delta X \Delta \tau$. Then Groeneboom, Jongbloed and Wellner (2001a) show algebraically that this inequality can be rewritten as

$$V(\bar{\tau}) - \bar{V} + \frac{1}{8} \Delta W \Delta \tau \leq -\left(\frac{\Delta \tau}{2}\right)^4.$$

Thus we have shown for any $t > 0$,

$$P\left(Y_R(\bar{\tau}) \leq H_R(\bar{\tau}) = \bar{Y}_R - \frac{1}{8}\Delta X_R \Delta \tau, B_t\right)$$

$$= P\left(V(\bar{\tau}) - \bar{V} + \frac{1}{8}\Delta W \Delta \tau \leq -\left(\frac{\Delta \tau}{2}\right)^4, B_t\right).$$

Groeneboom, Jongbloed and Wellner (2001a) show that

$$P\left(V(\bar{\tau}) - \bar{V} + \frac{1}{8}\Delta W \Delta \tau \leq -\left(\frac{\Delta \tau}{2}\right)^4, \tau_1 < M_\varepsilon, \tau_2 > M_\varepsilon\right) < \varepsilon, \qquad (8.44)$$

and thus that

$$P\left(V(\bar{\tau}) - \bar{V} + \frac{1}{8}\Delta W \Delta \tau \leq -\left(\frac{\Delta \tau}{2}\right)^4, \tau_1 < t - M_\varepsilon, \tau_2 > t + M_\varepsilon\right) < \varepsilon.$$
$$(8.45)$$

This independence from t follows because

$$\{(W(s) - W(t), V(s) - V(t) - (s - t)W(t))\}_{s \in \mathbb{R}}$$

is equal in distribution to $\left\{W(u - t), \int_t^u W(u - t)\,du\right\}_{u \in \mathbb{R}}$, since $\int_t^s W(u)du = \int_t^s (W(u) - W(t))du + (s - t)W(t)$, and thus $V(\bar{s}) - \bar{V} + \frac{1}{8}\Delta W \Delta s$ equals

$$V(\bar{s}) - V(t) - W(t)(\bar{s} - t)$$
$$- \left(\frac{1}{2}(V(s_1) - V(t) - W(t)(s_1 - t)) + \frac{1}{2}(V(s_2) - V(t) - W(t)(s_2 - t))\right)$$
$$+ \frac{1}{8}(W(s_2) - W(t) - (W(s_1) - W(t)))(s_2 - t - (s_1 - t))$$
$$=_d V\left(\frac{r_1 + r_2}{2}\right) - \frac{V(r_1) + V(r_2)}{2} + \frac{1}{8}(W(r_2) - W(r_1))(r_2 - r_1)$$

where $\bar{s} = (s_1 + s_2)/2$, $\bar{V} = (V(s_1) + V(s_2))/2$, $\Delta W = W(s_2) - W(s_1)$, and $\Delta s = s_2 - s_2$ and $r_i = s_i - t$ for $i = 1, 2$. This shows that the left hand sides of both of (8.44) and (8.45) are, regardless of t, bounded by

$$P\left(V(\bar{s}) - \bar{V} + \frac{1}{8}\Delta W \Delta s \leq -\left(\frac{\Delta s}{2}\right)^4, \text{ for some } s_1 < -M_\varepsilon, s_2 > M_\varepsilon\right).$$
$$(8.46)$$

This probability is defined in (2.27) on page 1633 of Groeneboom, Jongbloed and Wellner (2001a), and is shown to be less than ε at the top of page 1634, so, using this fact, we have now shown (8.44) and (8.45).

The probability we consider in (8.43) is on the event $B_{t_\varepsilon} = \{0 < \tau_1 < t_\varepsilon, \tau_2 > t_\varepsilon + M_\varepsilon\}$ rather than $C_{t_\varepsilon} \equiv \{0 < \tau_1 < t_\varepsilon - M_\varepsilon, \tau_2 > t_\varepsilon + M_\varepsilon\}$. The only cost for this is we need to double our M_ε for this to correspond with the probability in

(8.45). Thus (8.43) holds, but we do not yet have independence from t because of the t_ε in the expression. We easily circumvent this by replacing M_ε by $t_\varepsilon + M_\varepsilon$. Now we have shown (8.38) holds with M_ε independent of t.

Now we show (8.39). Note that we can write an analogous version of (8.43) for $t > M_\varepsilon$ as

$$P(0 \le \tau_1 \le t - M_\varepsilon) \le P(\tau_2 > t + M_\varepsilon) + P(C_t) \tag{8.47}$$

$$\le \varepsilon + P\left(Y_R(\bar{\tau}) \le H_R(\bar{\tau}) = \bar{Y}_R - \frac{1}{8}\Delta X_R \Delta \tau, C_{t_\varepsilon}\right)$$

$$\le 2\varepsilon$$

because, by the argument we just went through, the probability in the third line is again bounded by (8.46). Note we have t in place of t_ε in (8.47), so the above statement is already independent of t as long as $t > M_\varepsilon$. Thus we have shown $P((t - M) \vee 0 \le \tau_-^0(t) \vee 0 \le t) > 1 - \varepsilon$, since if $t < M_\varepsilon$ this probability is trivially 1. Showing the analogous statements for the left side, the existence of M, not depending on t, such that $P(\tau_-^0(-t) < -t - M) < \varepsilon$ and $P(-t \le \tau_+^0(-t) \wedge 0 \le (-t + M) \wedge 0) > 1 - \varepsilon$, can be done analogously. $\square$

The next result will relate the unconstrained and constrained limit estimators in the Gaussian setting.

Corollary 8.11. *Let the definitions and assumptions of Theorem 4.2 hold, and let Y and H be as in Theorem 4.1. Let $\widehat{\varphi} \equiv H''$. For any $t \in \mathbb{R}$, define*

$$s^+(t) = \inf\{s \in [t, \infty) : \widehat{\varphi}^0(s) = \widehat{\varphi}(s)\} \tag{8.48}$$

$$s^-(t) = \sup\{s \in (-\infty, t] : \widehat{\varphi}^0(s) = \widehat{\varphi}(s)\}. \tag{8.49}$$

Then we can say that for all $\varepsilon > 0$, there exists M_ε, not depending on t, such that

$$P(t - s^-(t) > M_\varepsilon) < \varepsilon \tag{8.50}$$

$$P(s^+(t) - t > M_\varepsilon) < \varepsilon. \tag{8.51}$$

Proof. Define a right-side sequence of knots to be a sequence of points

$$0 < \nu_1 < \nu_1^0 < \nu_2 < \nu_2^0 < \nu_3,$$

where ν_i are knots for $\widehat{\varphi}$ and ν_i^0 are knots for $\widehat{\varphi}^0$. Similarly, define a left-side sequence of knots

$$\nu_{-3} < \nu_{-2}^0 < \nu_{-2} < \nu_{-1}^0 < \nu_{-1} < 0.$$

Then we argue by the Intermediate Value Theorem and the Mean Value Theorem. First, we assume we are given such a sequence, without loss of generality take it to be a right-side sequence (on the probability 1 event on which Theo-

rem 4.1 holds). Then we can say, by our hypotheses, that

$$(H_R - Y_R)(\nu_i^0) = 0 \leq (H - Y)(\nu_i^0) \quad \text{for } i = 1, 2 \tag{8.52}$$

$$(H_R - Y_R)(\nu_i) \geq 0 = (H - Y)(\nu_i) \quad \text{for } i = 1, 2, 3. \tag{8.53}$$

By the Intermediate Value Theorem we can pick points $x_1 \in [\nu_1, \nu_1^0], x_2 \in [\nu_1^0, \nu_2], x_3 \in [\nu_2, \nu_2^0]$ such that $(H_R - Y_R)(x_i) = (H - Y)(x_i)$ for $i = 1, 2, 3$. Since $Y_R(t) - Y(t) = A(t)$ is a (random) affine function, we can conclude for $i = 1, 2, 3$ that

$$H_R(x_i) - H(x_i) - A(x_i) = 0.$$

We apply the Mean Value Theorem and get $t_i \in (x_i, x_{i+1})$ for $i = 1, 2$ such that

$$F_R(t_i) - H'(t_i) - A'(t_i) = 0.$$

Again applying the Mean Value Theorem, we get $s \in (t_1, t_2) \subset (x_1, x_3) \subset [\nu_1, \nu_2^0] \subset [\nu_1, \nu_3]$.

Now we will construct right-side sequences or left-side sequences of knots and be done. Note that by Lemma 8.10 and the analogous lemma for the unconstrained case, Lemma 2.7, page 1638, of Groeneboom, Jongbloed and Wellner (2001a), there exists a large $M > 0$ such that with probability $1 - \varepsilon$ there exists a right-side sequence of knots contained in any interval of length $\geq M$ that lies in $[0, \infty)$ and a left-side sequence of knots in any interval of length $\geq M$ that lies in $(-\infty, 0]$. For any $t > 0$, note that the interval $[t - 2M, t]$ contains an interval of length at least M which lies either entirely in $(-\infty, 0]$ or entirely in $[0, \infty)$. Thus, $[t - 2M, t]$ contains a one-sided sequence of knots, and thus an $s < t$ such that $\widehat{\varphi}^0(s) = \widehat{\varphi}(s)$, with probability $1 - \varepsilon$. Similarly, there exists a one-sided sequence of knots in $[t, t + M]$, and thus an $s > t$ such that $\widehat{\varphi}^0(s) = \widehat{\varphi}(s)$, with probability $1 - \varepsilon$. Thus, for $t > 0$, we have shown (8.50) and (8.51). Similarly, for $t < 0$, we consider intervals $[t - M, t]$ and $[t, t + 2M]$ in which there exist one-sided sequences of knots, which allows us to conclude that $\widehat{\varphi}^0(s) = \widehat{\varphi}(s)$ for an $s > t$ and an $s < t$. $\square$

Lemma 8.12. *Let the assumptions of Theorem 4.2 hold. For all $\varepsilon > 0$ there exists M_ε, not depending on t, such that*

$$P\left(|\widehat{\varphi}^0(t) - \widehat{\varphi}(t)| > M_\varepsilon\right) < \varepsilon \text{ and } P\left(|(\widehat{\varphi}^0)'(t) - \widehat{\varphi}'(t)| > M_\varepsilon\right) < \varepsilon, \tag{8.54}$$

where the derivatives can be taken to be right or left derivatives.

Proof. This follows from Lemma 8.10 and an argument similar to the finite sample tightness results. We can pick, by Corollary 8.11, $t - 2M < s_{-2} < s_{-1} < t < s_1 < s_2 < t + 2M$ where $\widehat{\varphi}^0(s_i) = \widehat{\varphi}(s_i)$ for $i = -2, -1, 1, 2$, with probability $1 - \varepsilon$ for M appropriately large. Then

$$(\widehat{\varphi}^0)'(t) \leq \frac{\widehat{\varphi}^0(s_2) - \widehat{\varphi}^0(s_1)}{s_2 - s_1} = \frac{\widehat{\varphi}(s_2) - \widehat{\varphi}(s_1)}{s_2 - s_1} \leq \widehat{\varphi}'(s_2),$$

and, similarly,

$$(\widehat{\varphi}^0)'(t) \geq \widehat{\varphi}'(s_{-2})$$

where $\widehat{\varphi}'$ and $(\widehat{\varphi}^0)'$ can be either the left or right derivatives. Thus

$$(\widehat{\varphi}^0)'(t) - \widehat{\varphi}'(t) \leq \widehat{\varphi}'(s_2) - \widehat{\varphi}'(t) \leq \widehat{\varphi}'(s_2) - \widehat{\varphi}'(s_{-2})$$

and

$$(\widehat{\varphi}^0)'(t) - \widehat{\varphi}'(t) \geq \widehat{\varphi}'(s_{-2}) - \widehat{\varphi}'(t) \geq \widehat{\varphi}'(s_{-2}) - \widehat{\varphi}'(s_2),$$

that is,

$$|(\widehat{\varphi}^0)'(t) - \widehat{\varphi}'(t)| \leq \widehat{\varphi}'(s_2) - \widehat{\varphi}'(s_{-2}). \tag{8.55}$$

Let $h_0(t) = -12t^2$. With high probability, the right side of (8.55) is bounded by

$$\begin{aligned}
\widehat{\varphi}'(t + 2M) &- \widehat{\varphi}'(t - 2M) \\
&\leq |\widehat{\varphi}'(t + 2M) - h_0'(t + 2M)| + |h_0'(t + 2M) - h_0'(t - 2M)| \\
&\quad + |h_0'(t - 2M) - \widehat{\varphi}'(t - 2M)|,
\end{aligned} \tag{8.56}$$

which is less than $M + M + 24 \cdot 2M$ with probability $1 - \varepsilon$, independently of t, by (2.36) or (2.37) of Lemma 2.7 on page 1638 of Groeneboom, Jongbloed and Wellner (2001a). Thus we have shown the second statement in (8.54), which we will now use to show the first statement in (8.54).

We first apply Lemma 9.2 to the difference $|\widehat{\varphi}^0(t) - \widehat{\varphi}(t)|$ by applying (9.1) to both $\widehat{\varphi}^0 - \widehat{\varphi}$ and to $\widehat{\varphi} - \widehat{\varphi}^0$, using the points s_{-1} and s_1 as a and b, respectively. Then by (9.1), we can bound both of these differences if we can bound both

$$(\widehat{\varphi}^0)'(s_1) - (\widehat{\varphi}^0)'(s_{-1}) \leq (\widehat{\varphi}^0)'(t + M) - (\widehat{\varphi}^0)'(t - M) \tag{8.57}$$

and

$$\widehat{\varphi}'(s_1) - \widehat{\varphi}'(s_{-1}) \leq \widehat{\varphi}'(t + M) - \widehat{\varphi}'(t - M), \tag{8.58}$$

since all the other terms are 0 by the definition of the s_i. Here we can take the derivatives to be either left or right derivatives. As in (8.56), we can bound $(\widehat{\varphi}^0)'(t + M) - (\widehat{\varphi}^0)'(t - M)$ from above by

$$\begin{aligned}
|(\widehat{\varphi}^0)'(t + M) &- \widehat{\varphi}'(t + M)| + |\widehat{\varphi}'(t + M) - \widehat{\varphi}'(t - M)| \\
&+ |\widehat{\varphi}'(t - M) - (\widehat{\varphi}^0)'(t - M)|.
\end{aligned}$$

The first and last terms are bounded by the second statement in (8.54). The middle term is shown to be bounded by (8.56). The middle term also bounds (8.58). All of this is with probability $1 - \varepsilon$ and uniformly in t, so we are done. □

For the next lemma, let $h_0(t) = -12t^2$ be the "true" concave function.

Lemma 8.13. *Let the definitions and assumptions of Theorem 4.2 hold. Then, for all $\varepsilon > 0$ there exists M_ε, independent of t, such that*

$$P\left(|\widehat{\varphi}^0(t) - h_0(t)| > M_\varepsilon\right) < \varepsilon \tag{8.59}$$

$$P\left(|(\widehat{\varphi}^0)'(t) - h_0'(t)| > M_\varepsilon\right) < \varepsilon. \tag{8.60}$$

where the derivatives can be right or left derivatives.

Proof. This is immediate from Lemma 2.7, page 1638, of Groeneboom, Jongbloed and Wellner (2001a) and Lemma 8.12. □

We are now in a position to prove Theorem 4.2.

Proof of Theorem 4.2. We define objective functions with variable bounds of integration,

$$\phi_{a,b}(g) = \frac{1}{2} \int_a^b g^2(t)dt - \int_a^b g(t)dX(t), \tag{8.61}$$

where we will always take $a < 0 < b$. For $i = 1, 2$, we will take $H_{L,i}$ and $H_{R,i}$ to satisfy the hypotheses stated in the theorem, and we need to show $H_{L,1} \equiv H_{L,2}$ and $H_{R,1} \equiv H_{R,2}$ almost surely. We will denote $F_{L,i} = -H'_{L,i}$ and $F_{R,i} = H'_{R,i}$ and

$$\varphi_i = H''_{L,i} = H''_{R,i}. \tag{8.62}$$

We also will use the notation $dF_i(t) = \varphi_i(t)dt$. Now, using that $\varphi_1^2 - \varphi_2^2 = (\varphi_1 - \varphi_2)^2 + 2(\varphi_1 - \varphi_2)\varphi_2$, we see that

$$\phi_{a,b}(\varphi_1) - \phi_{a,b}(\varphi_2) = \frac{1}{2} \int_a^b (\varphi_1 - \varphi_2)^2 \, d\lambda + \int_a^b (\varphi_1(t) - \varphi_2(t)) \, d(F_2(t) - X(t))$$

where λ is Lebesgue measure. Now, we specify that a_n^i and b_n^i are knots for φ_i, and, using Lemma 8.6, we take $a_n^2 < a_n^1 < -n < 0 < n < b_n^1 < b_n^2$. Then

$$\phi_{a_n^2, b_n^2}(\varphi_1) - \phi_{a_n^2, b_n^2}(\varphi_2) \geq \frac{1}{2} \int_{a_n^2}^{b_n^2} (\varphi_1 - \varphi_2)^2 \, d\lambda \geq \frac{1}{2} \int_{-n}^{n} (\varphi_1 - \varphi_2)^2 \, d\lambda$$

by Propositions 8.7 and 8.8, and, similarly,

$$\phi_{a_n^1, b_n^1}(\varphi_2) - \phi_{a_n^1, b_n^1}(\varphi_1) \geq \frac{1}{2} \int_{-n}^{n} (\varphi_2(t) - \varphi_1(t))^2 \, dt.$$

Now, we see directly from (8.61) that $\phi_{a_n^2, b_n^2}(\varphi_1) - \phi_{a_n^1, b_n^1}(\varphi_1)$ equals

$$\frac{1}{2} \int_{a_n^2}^{b_n^2} \varphi_1^2(t)dt - \frac{1}{2} \int_{a_n^1}^{b_n^1} \varphi_1^2(t)dt - \left(\int_{a_n^2}^{b_n^2} \varphi_1(t)dX(t) - \int_{a_n^1}^{b_n^1} \varphi_1(t)dX(t) \right)$$

$$= \frac{1}{2} \int_{A_n} \varphi_1^2(t)dt - \int_{A_n} \varphi_1(t)dX(t),$$

where $A_n = [a_n^2, a_n^1] \cup [b_n^1, b_n^2]$. Thus we have

$$\int_{-n}^{n} (\varphi_1 - \varphi_2)^2 \, d\lambda \leq \phi_{a_n^2, b_n^2}(\varphi_1) - \phi_{a_n^1, b_n^1}(\varphi_1) - (\phi_{a_n^2, b_n^2}(\varphi_2) - \phi_{a_n^1, b_n^1}(\varphi_2))$$

$$= \frac{1}{2} \int_{A_n} (\varphi_1^2 - \varphi_2^2) \, d\lambda - \int_{A_n} (\varphi_1(t) - \varphi_2(t)) \, dX(t).$$

Recalling $h_0(t) = 12t^2$, we see the previous display equals

$$\int_{A_n} \frac{1}{2} \left((\varphi_1 - h_0)^2 - (\varphi_2 - h_0)^2 \right) d\lambda - (\varphi_1 - \varphi_2) \, dW.$$

Thus we can conclude that

$$0 \leq \lim_n \int_{-n}^n (\varphi_1 - \varphi_2)^2 \, d\lambda$$
$$\leq \liminf_n \left(\frac{1}{2} \int_{A_n} \left((\varphi_1 - h_0)^2 - (\varphi_2 - h_0)^2 \right) d\lambda - \int_{A_n} (\varphi_1 - \varphi_2) \, dW \right).$$
$$(8.63)$$

The proof will now proceed as follows. We first will show that the right hand side of (8.63) is finite. For a function $g \colon \mathbb{R} \to \mathbb{R}$, we let $\|g\|_a^b = \sup_{t \in [a,b]} |g(t)|$ and $\|g\|_a^\infty = \sup_{t \in [a,\infty)} |g(t)|$. Using that $\int_{-\infty}^\infty (\varphi_1 - \varphi_2)^2 \, d\lambda$ is finite we will then conclude that $\|\varphi_1(t) - \varphi_2(t)\|_n^\infty \to 0$ as $n \to \infty$. We then will revisit our earlier argument which showed the right hand side of (8.63) was finite and use this new fact to show that (8.63) is, in fact, 0. This will finish the proof.

Thus, our next step is to show that $\int_{-\infty}^\infty (\varphi_1(t) - \varphi_2(t))^2 \, dt < \infty$. Note that we only need to control the $\liminf_n$ of the right hand side of (8.63) since $\int_{-n}^n (\varphi_1 - \varphi_2)^2 \, d\lambda$ is non-negative and non-decreasing in n. We will first show that $\int_{b_n^1}^{b_n^2} (\varphi_1 - \varphi_2) \, dW < \infty$. An identical argument shows $\int_{a_n^1}^{a_n^2} (\varphi_1 - \varphi_2) dW < \infty$. By integration by parts,

$$\int_{b_n^1}^{b_n^2} (\varphi_1 - \varphi_2)(u) dW(u) = \int_{b_n^1}^{b_n^2} (\varphi_1 - \varphi_2)(u) d(W(u) - W(b_n^1))$$
$$= \left(W(b_n^2) - W(b_n^1) \right) \left(\varphi_1(b_n^2) - \varphi_2(b_n^2) \right) \qquad (8.64)$$
$$- \int_{b_n^1}^{b_n^2} \left(W(u) - W(b_n^1) \right) (\varphi_1'(u) - \varphi_2'(u)) \, du$$

where we can take φ_i' to be the right-derivative, but this choice is inconsequential because of the almost sure continuity of W. Thus, by (8.69) and (8.70), for all n, with probability $1 - \varepsilon$, we can conclude that $\left| \int_{b_n^1}^{b_n^2} (\varphi_1 - \varphi_2)(u) dW(u) \right|$ is bounded above by

$$K_\varepsilon \left(\left| W(b_n^2) - W(b_n^1) \right| + \left| \int_{b_n^1}^{b_n^2} (W(u) - W(b_n^1)) du \right| \right).$$

Lemma 8.14 shows for $i = 1, 2$ that $\int_{b_n^1}^{b_n^2} (\varphi_i - h_0)^2 d\lambda < K_{\varepsilon,2}$ with probability $1 - \varepsilon$. Thus since, by (8.68), $\left(\left| W(b_n^2) - W(b_n^1) \right| + \left| \int_{b_n^1}^{b_n^2} (W(u) - W(b_n^1)) du \right| \right)$ is $O_p(1)$, and since this argument is perfectly symmetrical and applies to the interval $[a_n^1, a_n^2]$, we have now shown that the right hand side of (8.63) is $O_p(1)$ and thus finite almost surely, as desired.

Now that we have shown that $\int_{-\infty}^{\infty} (\varphi_1 - \varphi_2)^2 d\lambda < \infty$ almost surely, we can conclude that

$$\|\varphi_1 - \varphi_2\|_n^\infty \to 0 \tag{8.65}$$

almost surely as $n \to \infty$, and now using (8.65) with arguments similar to those used above, we will show $\int_{-\infty}^{\infty} (\varphi_1 - \varphi_2)^2 d\lambda = 0$ almost surely. By (8.65), Lemma 8.14 below allows us to conclude that almost surely $\int_{b_n^1}^{b_n^2} |\varphi_1' - h_0'| d\lambda \to 0$. Thus we can reexamine (8.64) and see that the right side is bounded above by

$$\left| W(b_n^2) - W(b_n^1) \right| \left| \varphi_1(b_n^2) - \varphi_2(b_n^2) \right| + \|W(\cdot) - W(b_n^1)\|_{b_n^1}^{b_n^2} \int_{b_n^1}^{b_n^2} |\varphi_1'(u) - \varphi_2'(u)| du$$

$$\leq \varepsilon \left(\left| W(b_n^2) - W(b_n^1) \right| + \|W(\cdot) - W(b_n^1)\|_{b_n^1}^{b_n^2} \right),$$

where we may choose n large enough to make the inequality occur with probability $1 - \varepsilon$ for any positive ε. Thus, since $\left(\left| W(b_n^2) - W(b_n^1) \right| + \|W(\cdot) - W(b_n^1)\|_{b_n^1}^{b_n^2} \right) = O_p(1)$, we have shown that we may choose n large enough that with probability $1 - \varepsilon$

$$\left| \int_{b_n^1}^{b_n^2} (\varphi_1(u) - \varphi_2(u)) \, dW(u) \right| \leq \varepsilon. \tag{8.66}$$

Next we show that the other term in (8.63), $\int_{A_n} \left((\varphi_1 - h_0)^2 - (\varphi_2 - h_0)^2 \right) d\lambda / 2$, is small. By Lemma 8.14, for any $\varepsilon > 0$ we may pick an M_ε such that both $| \int_{b_n^1}^{b_n^2} (\varphi_1 - h_0) \, d\lambda |$ and $b_n^2 - b_n^1$ are bounded by M_ε with probability $1 - \varepsilon$. Thus, defining $\varepsilon_2 = \varepsilon / M_\varepsilon$ we take n large enough such that with probability $1 - \varepsilon$ we have $\|\varphi_1 - \varphi_2\|_n^\infty < \varepsilon_2$. Then let $\delta(t) = \varphi_1(t) - \varphi_2(t)$ and conclude that

$$\int_{b_n^1}^{b_n^2} (\varphi_1 - h_0)^2 d\lambda = \int_{b_n^1}^{b_n^2} (\varphi_2 - h_0 + \delta)^2 d\lambda$$

$$\leq \int_{b_n^1}^{b_n^2} (\varphi_2 - h_0)^2 d\lambda + 2\varepsilon_2 \int_{b_n^1}^{b_n^2} |\varphi_2 - h_0| d\lambda + \varepsilon_2^2 (b_n^2 - b_n^1),$$

and that the above display is bounded above by

$$\int_{b_n^1}^{b_n^2} (\varphi_2 - h_0)^2 \, d\lambda + \varepsilon + \left(\frac{\varepsilon}{M_\varepsilon} \right)^2 M_\varepsilon \leq \int_{b_n^1}^{b_n^2} (\varphi_2 - h_0)^2 \, d\lambda + 2\varepsilon,$$

with probability $1 - 2\varepsilon$ and n large enough. Similarly, $\int_{b_n^1}^{b_n^2} (\varphi_2 - h_0)^2 d\lambda \leq \int_{b_n^1}^{b_n^2} (\varphi_1 - h_0)^2 d\lambda + 2\varepsilon$, and thus

$$\left| \frac{1}{2} \int_{A_n} \left((\varphi_1 - h_0)^2 - (\varphi_2 - h_0)^2 \right) d\lambda \right| \leq \varepsilon \tag{8.67}$$

with probability $1 - 2\varepsilon$. Thus we have shown that with probability approaching 1 both terms in (8.63) are bounded by ε as n goes to infinity. Thus since $\int_{-n}^n (\varphi_1 -$

$\varphi_2)^2 d\lambda$ is non decreasing in n, $\int_{-n}^{n} (\varphi_1 - \varphi_2)^2 d\lambda < \varepsilon$ with probability $1 - \varepsilon$ and thus it must be 0 almost surely. $\qquad\square$

The following lemma translates Lemma 8.13 into a more direct tightness result.

Lemma 8.14. *Let the definitions and assumptions of Theorem 4.2 hold and let $h_0(t) = -12t^2$. Let φ_i, $i = 1, 2$, be as in (8.62) and a_n^i and b_n^i as defined on page 2439. We then have*

$$b_n^2 - b_n^1 = O_p(1). \tag{8.68}$$

Furthermore, for $i = 1, 2$ and any $\varepsilon > 0$ and $k > 0$, there exist $K_\varepsilon, K_{\varepsilon,k} > 0$ such that with probability greater than $1 - \varepsilon$ we have

$$||\varphi_i - h_0||_{b_n^1}^{b_n^2} < K_\varepsilon \tag{8.69}$$

$$||\varphi_i' - h_0'||_{b_n^1}^{b_n^2} < K_\varepsilon, \tag{8.70}$$

(in which we take φ_i' to be either the right of the left derivative) and thus that

$$\int_{b_n^1}^{b_n^2} |\varphi_i - h_0|^k d\lambda < K_{\varepsilon,k}, \tag{8.71}$$

where K_ε and $K_{\varepsilon,k}$ do not depend on n. Further, if almost surely $\|\varphi_1 - \varphi_2\|_n^\infty \to 0$ as $n \to \infty$ then we can conclude that almost surely

$$\int_{b_n^1}^{b_n^2} |(\varphi_i)' - h_0'| \, d\lambda \to 0 \tag{8.72}$$

as $n \to \infty$, for $i = 1, 2$. The statements also hold if we replace b_n^1 by a_n^2 and b_n^2 by a_n^1.

Proof. (8.68) follows immediately from Lemma 8.10.

Next we will show (8.69) and (8.70). Let g_1 and g_0 be monotone functions. Then for any $t \in [a, b]$, we have that

$$g_1(t) - g_0(t) \le g_1(b) - g_0(a) = g_1(b) - g_0(b) + g_0(b) - g_0(a)$$

and similarly $g_0(t) - g_1(t) \le g_0(b) - g_0(a) + g_1(a) - g_0(a)$. Thus

$$|(g_1 - g_0)(t)| \le |(g_1 - g_0)(b)| + |(g_1 - g_0)(a)| + g_0(b) - g_0(a).$$

By monotonicity and Lemma 8.13, we can say

$$||\varphi_i' - h_0'||_{b_n^1}^{b_n^2} < 2M_\varepsilon + h_0'(n + M_\varepsilon) - h_0'(n) = 2M_\varepsilon + 24M_\varepsilon,$$

where φ_i' refers to either the left or the right derivative. This is independent of n thanks to the linearity of h_0'. Thus we have shown (8.70).

Now we establish (8.69). Fix $i \in \{1, 2\}$. We will apply Lemma 9.2 twice, with φ_i as g_1 and h_0 as g_0, and then with the reverse assignments. We let

$[a, b] = [n, n + M]$. Regardless of the choice of which is g_1 and which is g_0, we can bound the first two terms in (9.1), the weighted differences $\lambda(g_1(n + M) - g_0(n+M)) + (1-\lambda)(g_1(n) - g_0(n))$, by $2M_\varepsilon$ with probability $1 - \varepsilon$, independently of n, by Lemma 8.13. If h_0 is g_0 then for the third term of (9.1) we have to bound $(h'_0(n+) - h'_0(n + M-)) = 24M$ which is independent of n. If φ_i is g_0, then for the third term of (9.1) we have that $|\varphi'_i(n+M-) - \varphi'_i(n+)|$ is bounded above by

$$|\varphi'_i(n + M-) - h'_0(n + M)| + |h'_0(n) - \varphi'_i(n+)| + h'_0(n + M) - h'_0(n)$$

which we can again bound independently of n by the linearity of h'_0 and Lemma 8.13 with probability $1 - \varepsilon$. Since $[b_n^1, b_n^2] \subset [n, n + M]$ with probability $1 - \varepsilon$ for appropriately large M, and since $(n + M - t)(t - n)/M \leq (M/2)^2/M$ which is independent of n and t, the bound is independent of n or t. Thus we have shown (8.69). Then (8.71) follows immediately from (8.69) and (8.68), since we can bound $\int_{b_n^1}^{b_n^2} |\varphi_i - h_0|^k d\lambda \leq \int_{b_n^1}^{b_n^2} K_\varepsilon^k d\lambda \leq K_\varepsilon^k \cdot K_\varepsilon$, with probability $1 - \varepsilon$.

Finally, we show that if for a random outcome ω, $||\varphi_1^\omega - \varphi_2^\omega||_n^\infty \to 0$ as $n \to \infty$ then (8.72) follows. First, note for any a, b that if $\varepsilon < \int_a^b ((\varphi_i^\omega)' - h'_0) d\lambda = (\varphi_i^\omega - h_0)(b) - (\varphi_i^\omega - h_0)(a)$, and if $(\varphi_i^\omega - h_0)(b) > \varepsilon/2$ then $(\varphi_i^\omega - h_0)(a) < -\varepsilon/2$. Similarly, if $-\varepsilon > \int_a^b ((\varphi_i^\omega)' - h'_0) d\lambda$ we can conclude that $(\varphi_i^\omega - h_0)$ at a or at b is larger than $\varepsilon/2$ in absolute value. Since we can take n large enough that $|\varphi_i^\omega - h_0|$ is less than $\varepsilon/2$ at any $a, b > n$, by contradiction we have that $|\int_a^b ((\varphi_i^\omega)' - h'_0) d\lambda| < \varepsilon$ for such a and b. Now, since $\{t \in [b_n^1, b_n^2) : (\varphi_1^\omega)'(t) > (\varphi_2^\omega)'(t)\}$ and $\{t \in [b_n^1, b_n^2) : (\varphi_1^\omega)'(t) \leq (\varphi_2^\omega)'(t)\}$ are both intervals by monotonicity of $(\varphi_1^\omega)'$ and linearity of $(\varphi_2^\omega)'$ on $[b_n^1, b_n^2]$, we can conclude that $\int_a^b |(\varphi_i^\omega)' - h'_0| d\lambda < \varepsilon$ as desired. □

8.2.3. Proof for symmetric limit process

Proof of Theorem 4.4. The proof follows as in the proof of Theorem 4.2 in the previous section, where we replace H_R by H^+, τ_+^0 by τ_+^+, τ_R^0 by τ_R^+, and we take H_L to be 0, and also replace τ_L and τ_+^0 by 0.

In particular, we can see that analogs of the characterizing equalities and inequalities hold, and that an analog of Lemma 4.3 holds. The analog of Lemma 4.3 can immediately be seen to hold since by definition $(\widehat{\psi}^0)^{(2)}(t) = 0$ for $t \in (0, \tau_+^+)$. Proofs similar to the proofs of Proposition 8.8 and Proposition 8.7 show the following characterizing proposition holds.

Proposition 8.15. *If $\Delta \colon [0, \infty) \to \mathbb{R}$ is concave with maximum at 0 and if $0 \leq b \in \widehat{S}^+$, then*

$$\int_0^b \Delta(t) \left(\widehat{\psi}^0(t) dt - dX(t) \right) \geq 0.$$

And if $\Delta = \widehat{\psi}^0$ then the inequality is an equality.

In the proof of this proposition we define $F_R^+(u) = \int_{\tau_R^+}^{u} \widehat{\psi}^0(v) dv$ and $X_R^+(u) = \int_{\tau_R^+}^{u} dX(v)$ and then use the fact that by (4.12), $(F_R^+ - X_R^+)(0) = 0$. The proof of Theorem 4.4 then proceeds analogously to the proof of Theorem 4.2. This completes the proof of Theorem 4.4. $\qquad\square$

8.2.4. Proofs for pointwise limit theory

Theorem 4.5 A about the unconstrained estimator is Theorem 2.1 of Balabdaoui, Rufibach and Wellner (2009). Part B of the theorem is then proved in an identical fashion as that theorem, because for n large enough, in an $n^{-1/5}$ neighborhood of $x_0 \neq m$, the constrained and unconstrained estimators satisfy the same characterization. Part C then follows from Part B. Thus we focus first on proving Theorem 4.7 B, where $x_0 = m$, which follows from Theorem 4.8.

The statements in (4.16) about the limit distribution of $\widehat{f}_n^0$ and $(\widehat{f}_n^0)'$ now follow from the delta method (Taylor expansion). We devote the entire remainder of this subsection to proving Theorem 4.8. All the lemmas contained in the proof use the same notation, and have the same hypotheses as the theorem. An outline of the structure of the proof can be found in Section 7.

Proof of Theorem 4.8. Let $b \in \mathbb{R}$ denote our "local" parameter and let

$$t_{n,b} \equiv m + bn^{-1/5} \tag{8.73}$$

be the "global" parameter. We also let $s_{n,L}$ be any knot (sequence) of $\widehat{\varphi}_n^0$ strictly less than m satisfying $n^{1/5}(s_{n,L} - m) = O_p(1)$ and let $s_{n,R}$ be any knot (sequence) of $\widehat{\varphi}_n^0$ strictly larger than m satisfying $n^{1/5}(s_{n,R} - m) = O_p(1)$. Recall λ is Lebesgue measure, and define

$$\mathbb{Y}_n^f(b) \equiv n^{4/5} \int_m^{t_{n,b}} \left(\int_m^v (d\mathbb{F}_n - f_0(m)d\lambda) \right) dv,$$

$$\widehat{H}_n^f(b) \equiv n^{4/5} \int_m^{t_{n,b}} \left(\int_m^v (\widehat{f}_n - f_0(m))d\lambda \right) dv + A_n b + B_n,$$

and

$$\mathbb{Y}_{n,L}^f(b) \equiv n^{4/5} \int_{t_{n,b}}^{s_{n,L}} \left(\int_v^{s_{n,L}} (d\mathbb{F}_n - f_0(m)d\lambda) \right) dv,$$

$$\mathbb{Y}_{n,R}^f(b) \equiv n^{4/5} \int_{s_{n,R}}^{t_{n,b}} \left(\int_{s_{n,R}}^v (d\mathbb{F}_n - f_0(m)d\lambda) \right) dv,$$

$$\widehat{H}_{n,L}^f(b) \equiv n^{4/5} \int_{t_{n,b}}^{s_{n,L}} \left(\int_v^{s_{n,L}} (\widehat{f}_n^0 - f_0(m))d\lambda \right) dv + A_{n,L} n^{1/5}(s_{n,L} - t_{n,b}),$$

$$\widehat{H}_{n,R}^f(b) \equiv n^{4/5} \int_{s_{n,R}}^{t_{n,b}} \left(\int_{s_{n,R}}^v (\widehat{f}_n^0 - f_0(m))d\lambda \right) dv + A_{n,R} n^{1/5}(t_{n,b} - s_{n,R}),$$

where

$$A_n = n^{3/5}\left(\widehat{F}_n(m) - \mathbb{F}_n(m)\right) \quad \text{and} \quad B_n = n^{4/5}\left(\widehat{H}_n(m) - \mathbb{Y}_n(m)\right),$$

and

$$A_{n,L} = n^{3/5}\left(\mathbb{F}_{n,L}(s_{n,L}) - \widehat{F}^0_{n,L}(s_{n,L})\right), \tag{8.74}$$

$$A_{n,R} = n^{3/5}\left(\mathbb{F}_{n,R}(s_{n,R}) - \widehat{F}^0_{n,R}(s_{n,R})\right) \tag{8.75}$$

(recalling the definitions of $\mathbb{F}_{n,L}$ and $\mathbb{F}_{n,R}$ from (2.7)). Additionally, let

$$\mathbb{X}^f_n(b) := (\mathbb{Y}^f_n)'(b) = n^{3/5}\int_m^{t_{n,b}} \left(d\mathbb{F}_n - f_0(m)d\lambda\right),$$

$$\mathbb{X}^f_{n,L}(b) := -(\mathbb{Y}^f_{n,L})'(b) = n^{3/5}\int_{t_{n,b}}^{s_{n,L}} \left(d\mathbb{F}_n - f_0(m)d\lambda\right),$$

$$\mathbb{X}^f_{n,R}(b) := (\mathbb{Y}^f_{n,R})'(b) = n^{3/5}\int_{s_{n,R}}^{t_{n,b}} \left(d\mathbb{F}_n - f_0(m)d\lambda\right).$$

The terms for the constrained processes which would correspond to B_n turn out to be 0. Also, $A_{n,L}$ and $A_{n,R}$ appear to be off by a sign change when compared with A_n: this is because of the definitions of our left- and right-processes, which entails, e.g., $(\widehat{H}^0_{n,R} - \mathbb{Y}_{n,R})'(t) = -(\widehat{F}^0_{n,R} - \mathbb{F}_{n,R})(t)$. Note that in Balabdaoui, Rufibach and Wellner (2009), $\mathbb{Y}^f_n$ is denoted by $\mathbb{Y}^{loc}_n$ and similarly for $\widehat{H}^f_n$.

The proof proceeds as follows. We will derive the limit distribution for the empirical process-type $\mathbb{Y}$ and $\mathbb{X}$ terms. We will show that the estimator-type H terms (and appropriate derivatives) are tight, and also satisfy characterizations analogous to those given in Theorem 4.1 and Theorem 4.2. We argue then (by a continuous mapping argument) that a characterization must hold in the limit (along subsequences, using tightness of the H processes) and then apply Theorem 4.1 and Theorem 4.2 to conclude that the limit is as desired.

For $0 < c \leq \infty$, define

$$\mathcal{C}_c = \{h | h : [-c, c] \to \mathbb{R}, h \text{ is continuous}\}$$
$$\mathcal{D}_c = \{h | h : (-c, c) \to \mathbb{R}, h \text{ is cadlag and bounded}\},$$

where "cadlag" means right-continuous functions which have limits from the left. If $c = \infty$ then we we interpret the definition of $\mathcal{C}_\infty$ to mean continuous functions h defined on $(-\infty, \infty)$. We let $\|f\|$ be the supremum of f over its domain, and this is the distance we use in $\mathcal{C}_c$ when $c < \infty$. When $c = \infty$ we use the topology of convergence on all compacta (see Whitt (1970)). For $\mathcal{D}_c$ the uniform norm is too strong, so generally one uses a Skorokhod norm (Skorokhod (1956), see also Billingsley (1999)). We endow, for the moment, $\mathcal{D}_c$ with the J_1 Skorokhod norm (referred to as "the" Skorokhod topology in chapter 12 of Billingsley (1999)). When we come to proving tightness of our H-type processes we will further discuss topological details. Now we focus on the

empirical processes. We let "$A_n \Rightarrow A$" mean that A_n converges weakly to A in a space that will be specified in each context (Billingsley, 1999). The proof of the following lemma is standard but we will refer to it several times so we provide it here. Recall $t_{n,b} = m + bn^{-1/5}$, for $b \in \mathbb{R}$, and recall that $s_{n,L}, s_{n,R}$ are chosen to be of order $n^{-1/5}$ from m in probability. Let $\mathbb{D}_n = \mathbb{F}_n - F_0$, and define processes $\underline{\mathbb{A}}_n(b) \equiv (\mathbb{A}_{n,1}(b), \mathbb{A}_{n,2}(b), \mathbb{A}_{n,3}(b))$ and $\underline{\mathbb{B}}_n(b) \equiv (\mathbb{B}_{n,1}(b), \mathbb{B}_{n,2}(b), \mathbb{B}_{n,3}(b))$ for $b \in \mathbb{R}$ by

$$\underline{\mathbb{A}}_n(b) \equiv \frac{n^{3/5}}{\sqrt{f_0(m)}} \left(\int_m^{t_{n,b}} d\mathbb{D}_n, \int_{t_{n,b}}^{s_{n,L}} d\mathbb{D}_n, \int_{s_{n,R}}^{t_{n,b}} d\mathbb{D}_n \right),$$

and

$$\underline{\mathbb{B}}_n(b) \equiv \frac{n^{4/5}}{\sqrt{f_0(m)}} \left(\int_m^{t_{n,b}} \int_m^v d\mathbb{D}_n dv, \int_{t_{n,b}}^{s_{n,L}} \int_v^{s_{n,L}} d\mathbb{D}_n dv, \int_{s_{n,R}}^{t_{n,b}} \int_{s_{n,R}}^v d\mathbb{D}_n dv \right).$$

Let $\nu_{n,R} = n^{1/5}(s_{n,R} - m)$ and $\nu_{n,L} = n^{1/5}(s_{n,L} - m)$. For a (sequence of) Brownian motion processes $W \equiv W_n$ on $\mathbb{R}$ we also define corresponding approximating (nearly Gaussian) processes $\underline{G}_n \equiv (G_{n,1}, \ldots, G_{n,6})$ by

$$(G_{n,1}, G_{n,2}, G_{n,3}) = \left(\int_0^b dW, \int_b^{\nu_{n,L}} dW, \int_{\nu_{n,R}}^b dW \right),$$

and

$$(G_{n,4}, G_{n,5}, G_{n,6}) = \left(\int_0^b \int_0^v dW\, dv, \int_b^{\nu_{n,L}} \int_v^{\nu_{n,L}} dW\, dv, \int_{\nu_{n,R}}^b \int_{\nu_{n,R}}^v dW\, dv \right).$$

Lemma 8.16. *The vector of processes $(\mathbb{A}_n(b), \mathbb{B}_n(b))$ can be defined on a common probability space with a sequence of Brownian motion processes $W \equiv W_n$ so that*

$$\sup_{b \in [-c,c]} |(\mathbb{A}_n(b), \mathbb{B}_n(b)) - \underline{G}_n(b)| \to_p 0.$$

Proof. We prove that the difference of $\mathbb{A}_{n,1}(b) = n^{3/5} \int_m^{t_{n,b}} d\mathbb{D}_n$ and $G_{n,1}$ converges to 0 in probability uniformly in $|b| \leq c$ and that the difference of $\mathbb{B}_{n,1}(b) = n^{4/5} \int_m^{t_{n,b}} \int_m^v d\mathbb{D}_n dv$ and $G_{n,4}(b)$ converges to 0 in probability uniformly in $|b| \leq c$. The proofs for the other components are analogous. We can see $n^{3/5} \int_m^{t_{n,b}} d\mathbb{D}_n$ is equal in distribution to

$$n^{1/10} \left(\mathbb{U}_n(F_0(t_{n,b})) - \mathbb{U}_n(F_0(m)) \right) \tag{8.76}$$

where $\mathbb{U}_n(t) \equiv \sqrt{n}(\mathbb{F}_n^*(t) - t)$ is the empirical process corresponding to $\mathbb{F}_n^*(t)$, the empirical d.f. for n i.i.d. uniform random variables. By a Skorokhod construction (see e.g. Theorem 12.3.4, page 502, of Shorack and Wellner (2009); or see Mason and van Zwet (1987)) there exist a sequence of Brownian bridge processes B_n such that $\|\mathbb{U}_n - B_n\| = O(\log(n)n^{-1/2})$ almost surely. Thus, (8.76) is equal to

$$n^{1/10}(B_n(F_0(t_{n,b})) - B_n(F_0(m))) + n^{-4/10} \log(n) M_n(b), \tag{8.77}$$

where for all $|b| \le c$, $0 \le M_n(b) \le M = O(1)$ almost surely. Next we use that $B_n(t) = W_n(t) - tW_n(1)$ where $W_n(t) = B_n(t) + tN$ is a Brownian motion and N is a standard Normal random variable. Thus (8.77) equals

$$n^{1/10} \left(W_n(F_0(t_{n,b})) - W_n(F_0(m)) - (F_0(t_{n,b}) - F_0(m))W_n(1) \right) + o_p(1),$$

which is equal to

$$W_n(b)\sqrt{n^{1/5}(F_0(t_{n,b}) - F_0(m))/b} - W_n(1)f_0(m)bn^{-1/10} + o_p(1)$$
$$= W_n(b)\sqrt{f_0(m)} + o_p(1).$$

This shows that $\sup_{b \in [-c,c]} |\mathbb{A}_{n,1}(b) - G_{n,1}(b)| = o_p(1)$. Using this, we see that the process $n^{4/5} \int_m^{t_{n,b}} \int_m^v d\mathbb{D}_n dv$ defined on this probability space equals

$$n^{1/5} \int_m^{t_{n,b}} \left(\sqrt{f_0(m)}W_n(n^{1/5}(v-m)) + o_p(1) \right) dv = \sqrt{f_0(m)} \int_0^b W_n(v)dv + o_p(1)$$

with the $o_p(1)$ error still uniform in $|b| \le c$. Thus $\sup_{b \in [-c,c]} |\mathbb{B}_{n,1}(b) - G_{n,4}(b)| = o_p(1)$. This completes the proof for two of the terms and the other four are analogous. $\qquad\square$

Lemma 8.17. *Let* $\underline{P}_n \equiv (P_{n,1}, \ldots, P_{n,6})$ *be a vector of drift terms where we let*

$$P_{n,1}(b) = \frac{1}{6}b^3, \qquad\qquad P_{n,2}(b) = \int_b^{n^{1/5}(s_{n,L}-m)} \frac{1}{2}u^2 du,$$

$$P_{n,3}(b) = \int_{n^{1/5}(s_{n,R}-m)}^b \frac{1}{2}u^2 du, \qquad P_{n,4}(b) = \frac{1}{24}b^4,$$

and

$$P_{n,5}(b) = \int_b^{n^{1/5}(s_{n,L}-m)} \int_v^{n^{1/5}(s_{n,L}-m)} \frac{1}{2}u^2 du\, dv,$$

$$P_{n,6}(b) = \int_{n^{1/5}(s_{n,R}-m)}^b \int_{n^{1/5}(s_{n,R}-m)}^v \frac{1}{2}u^2 du\, dv.$$

Then the vector of processes $(\underline{\mathbb{X}}_n^f, \underline{\mathbb{Y}}_n^f) \equiv \left(\mathbb{X}_n^f, \mathbb{X}_{n,L}^f, \mathbb{X}_{n,R}^f, \mathbb{Y}_n^f, \mathbb{Y}_{n,L}^f, \mathbb{Y}_{n,R}^f \right)$ *can be defined on a common probability space with a sequence of Brownian motion processes* $W \equiv W_n$ *such that for* $0 < c < \infty$

$$\sup_{b \in [-c,c]} |(\mathbb{X}_n^f(b), \mathbb{Y}_n^f(b)) - \sqrt{f_0(m)}\underline{G}_n(b) - f_0''(m)\underline{P}_n(b)| \to_p 0 \quad \text{as } n \to \infty,$$

where $\underline{G}_n$ *is as in Lemma 8.16.*

Proof. We will show that the statement holds for $\mathbb{X}_{n,R}^f$. The proof for the other $\mathbb{X}$ terms and for all the $\mathbb{Y}$ terms are similar. Now, $n^{-3/5}\mathbb{X}_{n,R}^f(b)$ equals $\int_{s_{n,R}}^{t_{n,b}} d\mathbb{D}_n +$

$\int_{s_{n,R}}^{t_{n,b}} (f_0 - f_0(m)) d\lambda$; by a second-order Taylor expansion of f_0, since $f_0'(m) = 0$, the second term equals

$$\int_{s_{n,R}}^{t_{n,b}} (f_0^{(2)}(m) + o(1)) \frac{(w-m)^2}{2} dw = n^{-3/5} (f_0^{(2)}(m) + o(1)) \int_{n^{1/5}(s_{n,R}-m)}^{b} \frac{u^2}{2} du$$

where the $o(1)$ is uniform in $|b| \le c$. By Lemma 8.16, we have shown $\mathbb{X}_{n,R}^f$ has the desired limit. □

The next lemma shows that the localized processes still satisfy the characterizing system of equalities/inequalities of the global estimators; once these characterizing equalities/inequalities are carried over to the limit they will allow us to identify the limit distribution, via Theorems 4.1 and 4.2.

Lemma 8.18. *For* $b \in \mathbb{R}$,

$$\mathbb{Y}_{n,L}^f(b) - \widehat{H}_{n,L}^f(b) \ge 0, \tag{8.78}$$

$$\mathbb{Y}_{n,R}^f(b) - \widehat{H}_{n,R}^f(b) \ge 0, \tag{8.79}$$

and

$$\int_{\infty}^{\tau_{n,-}^0} \left(\mathbb{Y}_{n,L}^f(b) - \widehat{H}_{n,L}^f(b) \right) d(\widehat{H}_{n,L}^f)^{(3)}(b) = 0, \tag{8.80}$$

$$\int_{\tau_{n,+}^0}^{\infty} \left(\mathbb{Y}_{n,R}^f(b) - \widehat{H}_{n,R}^f(b) \right) d(\widehat{H}_{n,R}^f)^{(3)}(b) = 0 \tag{8.81}$$

where $\tau_{n,-}^0$ *is the largest left-knot of* $\widehat{\varphi}_n^0$ *(no larger than* m*) and* $\tau_{n,+}^0$ *is the smallest right-knot of* $\widehat{\varphi}_n^0$ *(no smaller than* m*).*

Proof. We consider the right-side process, and the left-side ones are analogous. The difference $\mathbb{Y}_{n,R}^f(b) - \widehat{H}_{n,R}^f(b)$ equals

$$n^{4/5} \int_{s_{n,R}}^{t_{n,b}} \left(\int_{s_{n,R}}^{v} d\mathbb{F}_n - \widehat{f}_n^0 d\lambda \right) dv - (b - n^{1/5}(s_{n,R} - m)) A_{n,R},$$

which, by the definition of $A_{n,R}$, equals

$$n^{4/5} \int_{s_{n,R}}^{t_{n,b}} \left(\int_{s_{n,R}}^{X_{(n)}} (d\mathbb{F}_n(u) - \widehat{f}_n^0(u)du) - \int_{v}^{X_{(n)}} (d\mathbb{F}_n(u) - \widehat{f}_n^0(u)du) \right) dv$$

$$- n^{4/5} (t_{n,b} - s_{n,R}) \int_{s_{n,R}}^{X_{(n)}} (d\mathbb{F}_n(u) - \widehat{f}_n^0(u)du),$$

which equals

$$-n^{4/5} \int_{s_{n,R}}^{t_{n,b}} \left(\int_{v}^{X_{(n)}} (d\mathbb{F}_n(u) - \widehat{f}_n^0(u)du) \right) dv,$$

which equals

$$n^{4/5} \int_{t_{n,b}}^{X_{(n)}} \left(\int_v^{X_{(n)}} (d\mathbb{F}_n(u) - \widehat{f}_n^0(u)du) \right) dv, \qquad (8.82)$$

since $\widehat{H}_{n,R}^0(s_{n,R}) - \mathbb{Y}_{n,R}(s_{n,R}) = 0$ by Theorem 2.4 B Thus,

$$\mathbb{Y}_{n,R}^f(b) - \widehat{H}_{n,R}^f(b) = n^{4/5} \left(\mathbb{Y}_{n,R}(t_{n,b}) - \widehat{H}_{n,R}^0(t_{n,b}) \right) \geq 0$$

for all $b \geq 0$, with equality if $t_{n,b}$ is a right-knot, by Theorem 2.4 B We have thus shown (8.79) and (8.81). Showing (8.78) and (8.80) is analogous. □

In order to show tightness of our H-processes, we want to apply Proposition 7.1, which is at the log-density level. To do this, we need to translate from the f-processes to processes defined at the log level, which we will refer to as φ-processes. Let

$$\begin{aligned}
\mathbb{Y}_n^\varphi(b) &= \frac{\mathbb{Y}_n^f(b)}{f_0(m)} - n^{4/5} \int_m^{t_{n,b}} \int_m^v R(u)dudv, \\
\widehat{H}_n^\varphi(b) &= \frac{\widehat{H}_n^f(b)}{f_0(m)} - n^{4/5} \int_m^{t_{n,b}} \int_m^v R_n(u)dudv, \\
\mathbb{Y}_{n,L}^\varphi(b) &= \frac{\mathbb{Y}_{n,L}^f(b)}{f_0(m)} - n^{4/5} \int_{t_{n,b}}^{s_{n,L}} \int_v^{s_{n,L}} R_n^0(u)dudv, \\
\mathbb{Y}_{n,R}^\varphi(b) &= \frac{\mathbb{Y}_{n,R}^f(b)}{f_0(m)} - n^{4/5} \int_{s_{n,R}}^{t_{n,b}} \int_{s_{n,R}}^v R_n^0(u)dudv, \\
\widehat{H}_{n,L}^\varphi(b) &= \frac{\widehat{H}_{n,L}^f(b)}{f_0(m)} - n^{4/5} \int_{t_{n,b}}^{s_{n,L}} \int_v^{s_{n,L}} R_n^0(u)dudv, \\
\widehat{H}_{n,R}^\varphi(b) &= \frac{\widehat{H}_{n,R}^f(b)}{f_0(m)} - n^{4/5} \int_{s_{n,R}}^{t_{n,b}} \int_{s_{n,R}}^v R_n^0(u)dudv,
\end{aligned} \qquad (8.83)$$

where

$$R_n(u) = \sum_{j=2}^\infty \frac{1}{j!} \left(\widehat{\varphi}_n(u) - \varphi_0(m) \right)^j, \quad \text{and} \quad R_n^0(u) = \sum_{j=2}^\infty \frac{1}{j!} \left(\widehat{\varphi}_n^0(u) - \varphi_0(m) \right)^j.$$

Also let

$$\begin{aligned}
\mathbb{X}_n^\varphi(v) &= (\mathbb{Y}_n^\varphi)'(v) = \frac{\mathbb{X}_n^f(v)}{f_0(m)} - n^{3/5} \int_m^{t_{n,v}} R_n(u)du, \\
\mathbb{X}_{n,L}^\varphi(v) &= -(\mathbb{Y}_{n,L}^\varphi)'(v) = \frac{\mathbb{X}_{n,L}^f(v)}{f_0(m)} - n^{3/5} \int_{t_{n,v}}^{s_{n,L}} R_n^0(u)du, \\
\mathbb{X}_{n,R}^\varphi(v) &= (\mathbb{Y}_{n,R}^\varphi)'(v) = \frac{\mathbb{X}_{n,R}^f(v)}{f_0(m)} - n^{3/5} \int_{s_{n,R}}^{t_{n,v}} R_n^0(u)du.
\end{aligned}$$

The above definitions are motivated by the following identities.

Lemma 8.19 (*f* to φ identities). *We have*

$$f_0(m)^{-1}\left(\widehat{f}_n(u) - f_0(m)\right) = \widehat{\varphi}_n(u) - \varphi_0(m) + R_n(u) \qquad (8.84)$$

$$f_0(m)^{-1}\left(\widehat{f}_n^0(u) - f_0(m)\right) = \widehat{\varphi}_n^0(u) - \varphi_0(m) + R_n^0(u). \qquad (8.85)$$

The φ-processes thus satisfy

$$\widehat{H}_n^\varphi(b) = n^{4/5}\int_m^{t_{n,b}}\int_m^v (\widehat{\varphi}_n(u) - \varphi_0(m))\,du\,dv + \frac{A_n b + B_n}{f_0(m)} \qquad (8.86)$$

$$\widehat{H}_{n,L}^\varphi(b) = n^{4/5}\int_{t_{n,b}}^{s_{n,L}}\int_v^{s_{n,L}} \left(\widehat{\varphi}_n^0(u) - \varphi_0(m)\right)\,du\,dv + \frac{A_{n,L}n^{1/5}(s_{n,L} - t_{n,b})}{f_0(m)}$$
$$\qquad (8.87)$$

$$\widehat{H}_{n,R}^\varphi(b) = n^{4/5}\int_{s_{n,R}}^{t_{n,b}}\int_{s_{n,R}}^v \left(\widehat{\varphi}_n^0(u) - \varphi_0(m)\right)\,du\,dv + \frac{A_{n,R}n^{1/5}(t_{n,b} - s_{n,R})}{f_0(m)}$$
$$\qquad (8.88)$$

Proof. The identities (8.84) and (8.85) are just the exponential series expansion about the density at m,

$$\hat{g}(u) - f_0(m) = f_0(m)(e^{\{\hat{\varphi}(u) - \varphi_0(m)\}} - 1) = f_0(m)\sum_{j=1}^\infty \frac{1}{j!}(\hat{\varphi}(u) - \varphi_0(m))^j, \quad (8.89)$$

where $\hat{g}$ is either $\widehat{f}_n$ or $\widehat{f}_n^0$, and $\hat{\varphi}$ is either $\widehat{\varphi}_n$ or $\widehat{\varphi}_n^0$, respectively. Now (8.86), (8.87), and (8.88) follow directly from either (8.84) or (8.85) and the definitions of the processes. □

Lemma 8.20. *Taking $b \in [m - cn^{-1/5}, m + cn^{-1/5}]$ for any $c > 0$, we have*

$$R_n(b) = o_p(n^{-2/5}) \quad and \quad R_n^0(b) = o_p(n^{-2/5}), \qquad (8.90)$$

uniformly for $b \in [m - cn^{-1/5}, m + cn^{-1/5}]$.

Proof. Note $(\hat{\varphi}(u) - \varphi_0(m))^j = O_p(n^{-2j/5})$, for $j \geq 2$, from by (7.2) for $\hat{\varphi} = \widehat{\varphi}_n^0$, and the analogous (4.17) on page 1319 of Balabdaoui, Rufibach and Wellner (2009) for $\hat{\varphi} = \widehat{\varphi}_n$, since $\varphi_0'(m) = 0$. Thus, we have shown (8.90). □

Lemma 8.21. *We have*

$$\mathbb{Y}_{n,L}^\varphi(b) - \widehat{H}_{n,L}^\varphi(b) \geq 0, \quad for\ b \leq 0, \qquad (8.91)$$

$$\mathbb{Y}_{n,R}^\varphi(b) - \widehat{H}_{n,R}^\varphi(b) \geq 0, \quad for\ b \geq 0, \qquad (8.92)$$

and

$$\int_\infty^{\tau_{n,-}^0} \left(\mathbb{Y}_{n,L}^\varphi(b) - \widehat{H}_{n,L}^\varphi(b)\right) d(\widehat{H}_{n,L}^\varphi)^{(3)}(b) = 0, \qquad (8.93)$$

$$\int_{\tau_{n,+}^0}^{\infty} \left(\mathbb{Y}_{n,R}^{\varphi}(b) - \widehat{H}_{n,R}^{\varphi}(b) \right) d(\widehat{H}_{n,R}^{\varphi})^{(3)}(b) = 0 \qquad (8.94)$$

where $\tau_{n,-}^0$ is the largest left-knot of $\widehat{\varphi}_n^0$ (no larger than m) and $\tau_{n,+}^0$ is the smallest right-knot of $\widehat{\varphi}_n^0$ (no smaller than m).

Proof. By the process definitions,

$$\mathbb{Y}_{n,R}^{\varphi}(b) - \widehat{H}_{n,R}^{\varphi}(b) = \frac{1}{f_0(m)} \left(\mathbb{Y}_{n,R}^{f}(b) - \widehat{H}_{n,R}^{f}(b) \right),$$

so by Lemma 8.18 we can conclude for $b \geq 0$ that

$$\mathbb{Y}_{n,R}^{\varphi}(b) - \widehat{H}_{n,R}^{\varphi}(b) \geq 0,$$

with equality if $t_{n,b}$ is a right-knot, as desired. We have thus shown (8.92) and (8.94), and (8.91) and (8.93) are similar. □

Lemma 8.22. *The vector of processes* $(\underline{\mathbb{X}}_n^{\varphi}, \underline{\mathbb{Y}}_n^{\varphi}) \equiv \left(\mathbb{X}_n^{\varphi}, \mathbb{X}_{n,L}^{\varphi}, \mathbb{X}_{n,R}^{\varphi}, \mathbb{Y}_n^{\varphi}, \mathbb{Y}_{n,L}^{\varphi}, \mathbb{Y}_{n,R}^{\varphi} \right)$ *can be defined on a common probability space with a sequence of Brownian motion processes* $W \equiv W_n$ *such that for* $0 < c < \infty$

$$\sup_{b \in [-c,c]} \left| (\underline{\mathbb{X}}_n^{\varphi}(b), \underline{\mathbb{Y}}_n^{\varphi}(b)) - \frac{1}{\sqrt{f_0(m)}} \underline{G}_n(b) - \varphi_0''(m) \underline{P}_n(b) \right| \to_p 0 \qquad as \ n \to \infty,$$
$$(8.95)$$

where $\underline{G}_n$ *and* $\underline{P}_n$ *are as in Lemma 8.16 and Lemma 8.17.*

Proof. By (8.90), since $\varphi_0''(m) = f_0''(m)/f_0(m)$, we can conclude that (8.95) holds. □

We have established the appropriate characterizing properties of the φ-processes, and the limit distribution of the $\mathbb{Y}^{\varphi}$ processes. It remains to prove tightness of the H^{φ}-processes. To begin, we discuss the spaces in which our convergences will occur. For $0 < c \leq \infty$, define

$$\mathcal{F}_{c,M} = \{ f \in \mathcal{D}_c | f \text{ is non-increasing and } \|f\| \leq M \},$$

where "cadlag" means right-continuous functions which have limits from the left. Our H, $H^{(1)}$, and $H^{(2)}$ functions (for the constrained and unconstrained estimators) are continuous and the uniform norm is appropriate for them. The $H^{(3)}$ type functions lie in $\mathcal{F}_{c,M}$, and the uniform norm is too strong. For the convergence of the $\mathbb{Y}$-processes we used the J_1 Skorokhod metric. Unfortunately this also is too strong as it does not allow multiple jumps to approximate a single jump in $\mathcal{F}_{c,M}$ (see Remark B.0.10 in Appendix B of Doss (2013b)). Thus, we will use the so-called M_1 Skorokhod metric on $\mathcal{F}_{c,M}$. This is defined in Section 12.3 of Whitt (2002), and discussed in the following sections. We give a brief introduction here. The M_1 metric is defined as follows. For a set $A \subseteq \mathbb{R}$, let $\|x-A\| := \inf_{y \in A} |x-y|$. (Note that we have also taken $\|f\|$ to be the supremum

of a function f over its domain. It will be clear from context which usage is intended.) For a function x and $\delta > 0$, let $w_s(x, \delta) = \sup \|x(t_2) - [x(t_1), x(t_3)]\|$, where the sup is taken over t_1, t_2, and t_3 such that $-c \vee (t_2 - \delta) \le t_1 < t_2 < t_3 \le c \wedge (t_2 + \delta)$.[1] Note that since sequences that converge in the J_1 topology also converge in the M_1 topology (Whitt (2002)), the weak convergences proved for the empirical processes in Lemma 8.22 still hold when we use the M_1 topology. By Whitt (1980) (see Theorem B.0.2 of Doss (2013b)), $\mathcal{F}_{c,M}$ is a complete, separable metric space. And furthermore we have the following.

Proposition 8.23 (Lemma B.0.9 of Doss (2013b)). *$\mathcal{F}_{c,M}$ is precompact, meaning that every sequence in $\mathcal{F}_{c,M}$ has a convergent subsequence (not necessarily lying in $\mathcal{F}_{c,M}$).*

This is the fundamental property we need for tightness arguments, to which we now proceed.

Lemma 8.24. *The processes $(\widehat{H}_{n,L}^{\varphi})''', (\widehat{H}_{n,L}^{\varphi})'', (\widehat{H}_{n,L}^{\varphi})'$ and $\widehat{H}_{n,L}^{\varphi}$ are tight in $\mathcal{D}_c \times \mathcal{C}_c^3$ when $0 < c < \infty$. The same tightness holds if we replace the L-processes by the R$-$processes.*

Proof. We will discuss the tightness for the left-side processes. The argument for the right-side processes is analogous. Proposition 7.1 shows that for any ε, we can take $M > 0$ large enough that $(\widehat{H}_{n,L}^{\varphi})'''$ lies in $\mathcal{F}_{c,M}$ with probability $1 - \varepsilon$. Since $\mathcal{F}_{c,M}$ is precompact in $\mathcal{D}_c$ by Proposition 8.23, $(\widehat{H}_{n,L}^{\varphi})'''$ is tight. Then $(\widehat{H}_{n,L}^{\varphi})''$ is uniformly bounded by Proposition 7.1, and since its derivative is uniformly bounded, and since the set of functions with their values as well as the values of their derivatives uniformly bounded by M is compact in $\mathcal{C}_c$ (via the Arzela-Ascoli theorem, see e.g. Royden (1988)), we can conclude that $(\widehat{H}_{n,L}^{\varphi})''$ is tight in $\mathcal{F}_{c,M}$. Similarly, since integrals on bounded intervals of uniformly bounded functions are also uniformly bounded, and by Lemma 8.25 below, together with the fact that $n^{1/5}(s_{n,L} - b)$ is $O_p(1)$ by assumption we see that $(\widehat{H}_{n,L}^{\varphi})'$ and $\widehat{H}_{n,L}^{\varphi}$ are uniformly bounded, and their respective derivatives are uniformly bounded, so we can again conclude that they are tight. An identical argument works for the right-side processes. □

We will want to consider our processes in $\mathcal{C}_\infty$ and in $\mathcal{D}_\infty$. For the continuous processes in $\mathcal{C}_\infty$, Corollary 5 of Whitt (1970) says that processes that are tight in $\mathcal{C}_c$ for all $0 < c < \infty$ are then tight in $\mathcal{C}_\infty$. By Theorem 12.9.3 of Whitt (2002) (with Prohorov's theorem, e.g. van der Vaart and Wellner (1996) page 21), processes that are tight in $\mathcal{D}_c$, $0 < c < \infty$, are tight in $\mathcal{D}_\infty$. For the next lemma, recall the definitions of $A_{n,L}, A_{n,R}$ in (8.74) and (8.75).

Lemma 8.25. *As $n \to \infty$,*

$$|A_{n,L}| \to 0 \quad and \quad |A_{n,R}| \to 0, \ almost \ surely. \tag{8.96}$$

---

[1]Note that w_s coincides with the definition of Δ_{M_1} in Skorokhod (1956).

Proof. Because $s_{n,L}$ is strictly less than m, we can apply Corollary 2.7 B, so that $|A_{n,L}| = n^{3/5} \left| \mathbb{F}_{n,L}(s_{n,L}) - \widehat{F}_{n,L}^0(s_{n,L}) \right| \leq n^{-2/5} \to 0$ almost surely. Similarly, since $\mathbb{F}_{n,L}(X_{(n)}) = 1 = \mathbb{F}_n(X_{(n)})$, by the same corollary, $|A_{n,R}| \to 0$ almost surely. $\square$

With Lemmas 8.21, 8.22, and 8.24 in hand, we can now finish the proof of the theorem. Fix a subsequence n'. Let

$$Z_{n,L} = \left((\widehat{H}_{n,L}^\varphi)^{(3)}, (\widehat{H}_{n,L}^\varphi)^{(2)}, (\widehat{H}_{n,L}^\varphi)^{(1)}, \widehat{H}_{n,L}^\varphi, \mathbb{X}_{n,L}^\varphi, \mathbb{Y}_{n,L}^\varphi \right),$$

$$Z_{n,R} = \left((\widehat{H}_{n,R}^\varphi)^{(3)}, (\widehat{H}_{n,R}^\varphi)^{(2)}, (\widehat{H}_{n,R}^\varphi)^{(1)}, \widehat{H}_{n,R}^\varphi, \mathbb{X}_{n,R}^\varphi, \mathbb{Y}_{n,R}^\varphi \right),$$

By Lemmas 8.22 and 8.24, $Z_{n,R}$ and $Z_{n,L}$ are both tight in the space $E_c \equiv \mathcal{D}_c \times \mathcal{C}_c^3 \times \mathcal{D}_c \times \mathcal{C}_c$ with $0 < c < \infty$. This means they are also tight in E_∞, by the discussion after Lemma 8.24. Thus there exists a subsubsequence n'' such that $Z_{n'',R}$ and $Z_{n'',L}$ converge weakly. By the Skorokhod construction (see e.g., Chapter 14 of Shorack (2000)), we may assume that the convergence is almost sure (a.s.). Let $(Z_{0,L}, Z_{0,R})$ be the limit and let $Z_{0,L} = (H_L^{(3)}, H_L^{(2)}, H_L^{(1)}, H_L, X_L, Y_L)$, and $Z_{0,R} = (H_R^{(3)}, H_R^{(2)}, H_R^{(1)}, H_R, X_R, Y_R)$. Note that $(\widehat{H}_{n,R}^\varphi)^{(2)} = (\widehat{H}_{n,L}^\varphi)^{(2)}$ and this function is of course concave with mode at 0, so $H_R^{(2)} = H_L^{(2)}$ must also be concave with mode at 0. Let

$$\tau_L = \sup \left(\widehat{S}^0(H_R^{(2)}) \cap (-\infty, 0) \right) \quad \text{and} \quad \tau_R = \inf \left(\widehat{S}^0(H_R^{(2)}) \cap (0, \infty) \right) \quad (8.97)$$

with $\widehat{S}^0(H_R^{(2)})$ defined as in (4.2). There must be a sequence of knots $\tau_{n'',R} \in (S_n(\widehat{\varphi}_n^0) \cap (m, \infty))$ such that $n^{1/5}(\tau_{n'',R} - m) \to \tau_R$ a.s. To see that we can take $\tau_{n'',R}$ strictly greater than m, by (8.97) we see the only way $\tau_R = 0$ is if there exists a sequence of points of $\widehat{S}^0(H_R^{(2)})$ strictly greater than 0 and converging to 0. Similarly there is a sequence $\tau_{n'',L} \in (S_n(\widehat{\varphi}_n^0) \cap (-\infty, m))$ such that $n^{1/5}(\tau_{n'',L} \to \tau_L$ a.s. In our definitions of the (f- and φ-) processes, $s_{n,R}$ was any knot strictly greater than m satisfying $n^{1/5}(s_{n,R} - m) = O_p(1)$, and analogously for $s_{n,L}$. Take $s_{n'',R} = \tau_{n'',R}$ and $s_{n'',L} = \tau_{n'',L}$. Then let

$$s_{n''} = (n'')^{1/5} \left(\tau_{n'',L} - m, \tau_{n'',R} - m, \tau_{n'',-}^0 - m, \tau_{n'',+}^0 - m \right),$$

let $s_{n''} \to (\tau_L, \tau_R, \tau_-, \tau_+)$ and let $Z_0 = (Z_{0,L}, Z_{0,R}, \tau_L, \tau_R, \tau_-, \tau_+)$, where τ_- and τ_+ are the limits of the corresponding terms again by tightness from Proposition 7.3. By Lemma 8.22, if we let $Y \equiv Y_{a,\sigma}$ as defined in (4.18) with $a = |\varphi_0^{(2)}(m)|/4!$ and $\sigma = 1/\sqrt{f_0(m)}$, then

$$Y_R(b) = \int_{\tau_R}^b \int_{\tau_R}^v dY'(v) dv, \qquad X_R(b) = Y_R'(b) = \int_{\tau_R}^b dY',$$

$$Y_L(b) = \int_b^{\tau_L} \int_v^{\tau_L} dY'(v) dv, \qquad X_L(b) = -Y_L'(b) = \int_b^{\tau_L} dY'.$$

Let $\phi : E_c^2 \times \mathbb{R}^4 \to \mathbb{R}$ be defined by

$$\phi(z_1, \ldots, z_{16}) = \int_{z_{13}}^{z_{14}} z_8 d\lambda - dz_{11}$$

giving

$$\phi(Z_n) = \int_{n^{1/5}(\tau_{n,L} - m)}^{n^{1/5}(\tau_{n,R} - m)} (\widehat{H}_{n,R}^\varphi)^{(2)} d\lambda - d(\mathbb{Y}_{n,R}^\varphi)^{(1)}$$
$$= (\widehat{H}_{n,R}^\varphi - \mathbb{Y}_{n,R}^\varphi)'(n^{1/5}(\tau_{n,R} - m)) - (\widehat{H}_{n,R}^\varphi - \mathbb{Y}_{n,R}^\varphi)'(n^{1/5}(\tau_{n,L} - m)).$$

By Lemma 8.26 below, $\phi(Z_{n''}) \to 0$ a.s., and $Z_{n''} \to Z_0$ a.s., and ϕ is continuous at z such that z_8 and z_{11} are continuous functions, so a.s. $\phi(Z_0) = 0$, i.e.

$$\int_{\tau_L}^{\tau_R} H_R^{(2)} d\lambda - d(Y_R^{(1)}) = 0 \quad \text{a.s.,}$$

so condition (4.7) of Theorem 4.2 holds. Now, let $\phi_c : E_c^2 \times \mathbb{R}^4 \to \mathbb{R}$ be defined by $\phi_c(z) = \inf_{b \in [0,c]}(z_4(b) - z_6(b)) \wedge 0$ giving $\phi_c(Z_n) = \inf_{b \in [0,c]}(\widehat{H}_{n,L}^\varphi(b) - \mathbb{Y}_{n,L}^\varphi(b)) \wedge 0$. This ϕ_c is continuous since z_4 and z_6 are continuous, and $\phi_c(Z_n) = 0$ a.s. by Lemma 8.21, so $\phi_c(Z_0) = 0$ a.s., for all c. An analogous argument holds for the right-side processes. Thus,

$$(H_L - Y_L)(b) \le 0 \quad \text{for } b \le 0,$$
$$(H_R - Y_R)(b) \le 0 \quad \text{for } b \ge 0,$$

so (4.8) and (4.9) hold. Now, let $\phi_{R,c}(z) = \int \mathbb{1}_{[z_{16},c]}(z_{10} - z_{12}) dz_7$ (where z_{16} corresponds to τ_+, z_{10} to H_R, z_{12} to Y_R, and z_7 to $H_R^{(3)}$). Let

$$\phi_R(z) = \int_{[\tau_+^0(z_2), \infty)} (z_{10} - z_{12}) dz_7,$$

where $\tau_+^0(z_2)$ is defined as in (4.4) (and thus $\widehat{S}^0(z_2)$ is defined as in (4.2)). We want to show $\phi_R(Z_0) = 0$ a.s. Note that $\tau_+ \le \tau_+^0(Z_{0,2})$, although a priori we may not have equality. This is because a linear function may be well approximated by a nonlinear function, but the reverse is not true. Thus $\widehat{\varphi}_n^0$ could potentially have knots strictly between the limit knot $\tau_+^0(Z_{0,2})$ and m (on an $n^{-1/5}$ scale) so τ_+ could be smaller than $\tau_+^0(Z_{0,2})$, but $\widehat{\varphi}_n^0$ must have knots approaching $\tau_+^0(Z_{0,2})$ (on an $n^{-1/5}$ scale), so τ_+ cannot be larger than $\tau_+^0(Z_{0,2})$. By Lemma 8.21, $\phi_{R,c}(Z_n) = 0$ a.s., and by Lemma 8.27 below we can conclude $\phi_{R,c}(Z_0) = 0$ a.s. Now, let $c \to \infty$ to see

$$\int \mathbb{1}_{[Z_{0,16}, \infty)}(Z_{0,10} - Z_{0,12}) dZ_{0,7} = 0$$

and since the integrand is nonpositive and the integrating measure is nonpositive, this implies

$$\int \mathbb{1}_{[\tau_+^0(Z_{0,2}), \infty)}(Z_{0,10} - Z_{0,12}) dZ_{0,7} = 0$$

as desired. An analogous argument holds for the functional

$$\phi_L(z) = \int_{(-\infty, \tau_-^0(z_2)]} (z_4 - z_6) dz_1$$

with $\tau_-^0(z_2)$ defined as in (4.3). Thus we can a.s. conclude $\phi_R(Z_0) = 0$ and $\phi_L(Z_0) = 0$, so we have shown condition (4.10) of Theorem 4.2 holds. We have shown that as $n'' \to \infty$, $Z_{n'',0}$ converge a.s., so weakly, to Z_0 which satisfies the uniqueness criteria of Theorem 4.2. Thus we conclude that the limit does not depend on the choice of subsequence, and so can conclude $Z_{n,L} \Rightarrow Z_{0,L}$ and $Z_{n,R} \Rightarrow Z_{0,R}$ both in E_∞, as desired. This ends the proof of Theorem 4.8. □

Here are the two remaining lemmas we used in the proof of Theorem 4.8.

Lemma 8.26. *Let* $\nu_{n,R} = n^{1/5}(s_{n,R} - m)$ *and* $\nu_{n,L} = n^{1/5}(s_{n,L} - m)$. *Then almost surely, recalling the notation* $g(a, b] = g(b) - g(a)$, *we have*

$$\left| \left(\widehat{H}_{n,R}^\varphi - \mathbb{Y}_{n,R}^\varphi \right)' (\nu_{n,L}, \nu_{n,R}] \right| \leq \frac{4}{f_0(m)} n^{-2/5},$$

Proof. We have

$$(\widehat{H}_{n,R}^f - \mathbb{Y}_{n,R}^f)'(b) = n^{3/5} \int_{s_{n,R}}^{t_{n,b}} (\widehat{f}_n^0(u) du - d\mathbb{F}_n(u)) + A_{n,R}$$

so that

$$(\widehat{H}_{n,R}^f - \mathbb{Y}_{n,R}^f)'(\nu_{n,R}) - (\widehat{H}_{n,R}^f - \mathbb{Y}_{n,R}^f)'(\nu_{n,L})$$
$$= n^{3/5} \int_{s_{n,L}}^{s_{n,R}} (\widehat{f}_n^0(u) du - d\mathbb{F}_n(u)).$$

Thus,

$$\left(\widehat{H}_{n,R}^\varphi - \mathbb{Y}_{n,R}^\varphi \right)' (\nu_{n,L}, \nu_{n,R}] = \frac{n^{3/5}}{f_0(m)} \int_{s_{n,L}}^{s_{n,R}} (\widehat{f}_n^0(u) du - d\mathbb{F}_n(u))$$

which is bounded in absolute value by $2n^{-2/5}/f_0(m)$ almost surely, by applying Corollary 2.7 (recall, by definition, $s_{n,R}$ and $s_{n,L}$ are not equal to m). □

In the definition of $\phi_{R,c}$ in the lemma below, z_{16} corresponds to τ_+, z_{10} to H_R, z_{12} to Y_R, and z_7 to $H_R^{(3)}$. The variables defining $\phi_{L,c}$ are the analogous left-side terms.

Lemma 8.27. *For* $n \geq 0$, *assume* $z_n \in E_c^2 \times \mathbb{R}^4$ *is such that* z_n *converges to* z_0. *Assume* $z_{0,1}$ *and* $z_{0,7}$ *are nonincreasing. Let* $\phi_{R,c}(z) = \int \mathbb{1}_{[z_{16},c]}(z_{10} - z_{12}) dz_7$ *and* $\phi_{L,c}(z) = \int \mathbb{1}_{[-c, z_{15}]}(z_4 - z_6) dz_1$. *Assume further that* $\phi_{R,c}(z_n) = 0$, $\phi_{L,c}(z_n) = 0$, $z_{n,10} - z_{n,12} \leq 0$ *and* $z_{n,4} - z_{n,6} \leq 0$. *Then for* $c > 0$, $\phi_{R,c}(z_0) = 0$ *if* c *is not a discontinuity point of* $z_{0,7}$ *and* $\phi_{L,c}(z_0) = 0$ *if* $-c$ *is not a discontinuity point of* $z_{0,1}$.

Proof. We consider $\phi_{R,c}$, the proof for $\phi_{L,c}$ is analogous. For convenience, let $h_c(z) = \mathbb{1}_{[z_{16},c]}(z_{10} - z_{12})$, and then $\int h_c(z_n)dz_{n,7} - \int h_c(z_0)dz_{0,7}$ equals

$$\int (h_c(z_n) - h_c(z_0))\, dz_{n,7} - \int h_c(z_0)d(z_{0,7} - z_{n,7}). \qquad (8.98)$$

The integrand of the first term in (8.98) is uniformly converging to 0. If we let x_1 and x_2 be any fixed continuity points of z_0 satisfying $x_1 \leq z_{0,22} \leq c \leq x_2$ then, for large enough n the measure $z_{n,7}$ has total mass bounded by $z_{0,7}(x_1) - z_{0,7}(x_2) + 2 < \infty$. Thus the first term in (8.98) converges to 0 as $n \to \infty$ since $dz_{n,7}$ is a nonpositive measure (Royden (1988), Chapter 11.5). Now since $z_{n,5}$ converges to $z_{0,5}$ in the M_1 topology, $z_{n,5}$ converges weakly to z_5, in the sense that for all $t \in (-c,c)$ that are continuity points of $z_{0,5}$, $z_{n,5}(t) \to z_{0,5}(t)$ as $n \to \infty$ (Lemma 12.5.1, Whitt (2002)). The integrand of the second term in (8.98) is uniformly bounded and has discontinuity points at $z_{0,22}$ and c, so $\int \mathbb{1}_{(z_{0,16},c)}h_c(z_0)d(z_{0,7} - z_{n,7})$ converges to 0. By assumption c is not a discontinuity point of $z_{0,7}$ so $\int \mathbb{1}_{\{c\}}h_c(z_0)d(z_{0,7} - z_{n,7})$ goes to 0 as $n \to \infty$. If $z_{0,16}$ is not a discontinuity point of $z_{0,7}$ then similarly $\int \mathbb{1}_{\{z_{0,16}\}}h_c(z_0)d(z_{0,7} - z_{n,7})$ goes to 0. Thus, assume $z_{0,16}$ is a discontinuity point of $z_{0,7}$. Then by the M_1 convergence of $z_{n,7}$ to $z_{0,7}$, there exists a sequence x_n of discontinuity points of $z_{n,7}$ such that $x_n \to z_{0,16}$. By the assumption that $\phi_{R,c}(z_n) = 0$, we know $(z_{n,10} - z_{n,12}(x_n) = 0$ (since $z_{n,10} - z_{n,12} \leq 0$), and by uniform convergence, we see that $(z_{0,10} - z_{0,12})(z_{0,16}) = 0$. Thus $\int \mathbb{1}_{\{z_{0,16}\}}h_c(z_0)d(z_{0,7} - z_{n,7})$ converges to 0, and so (8.98) converges to 0 as $n \to \infty$, so we are done. □

Next we prove Theorem 4.9.

Proof of Theorem 4.9. We argue by considering $|X_1|, \ldots, |X_n| \overset{iid}{\sim} 2F_0 - 1$ on $[0,\infty)$ with density $g_0^+ = 2f_0\mathbb{1}_{[0,\infty)}$. Note that g_0^+ is log-concave with mode known to be at 0 and so we can use the results based on the mode-constrained MLE. Let $\psi_0^+ = \log g_0^+ = \varphi_0 + \log 2$ on $[0,\infty)$. Recall that

$$\mathbb{F}_{n,R}^+(x) = n^{-1}\sum_{i=1}^{n} \mathbb{1}_{\{|X_i| \geq x\}} \quad \text{and} \quad \widehat{G}_{n,R}^+(x) = \int_x^{|X|_{(n)}} \widehat{g}_n^+(u)du$$

where $\widehat{g}_n^+ = 2\widehat{g}_n^0\mathbb{1}_{[0,\infty)}$. Let $\widehat{\psi}_n^+ = \log \widehat{g}_n^+$. Let $s_{n,R} > 0$ be a knot of $\widehat{\psi}_n^+$ satisfying $n^{1/5}s_{n,R} = O_p(1)$. The proof proceeds as in the proof of Theorem 4.8, except we replace $\widehat{F}_{n,R}^0$ by $\widehat{G}_{n,R}^+$, $\mathbb{F}_{n,R}$ by $\mathbb{F}_{n,R}^+$, and f_0 by g_0^+, and we consider only "right-side" processes on $[0,\infty)$. To make things line up fully, we can take $s_{n,L}$ to be 0 and define left side analogs of $\widehat{G}_{n,R}^+$ and $\mathbb{F}_{n,R}^+$ to be identically 0. □

See the scaling relations in (7.5) and (7.6), and the three following displays, to see how Theorem 4.9 proves Theorem 4.7 C. Note that in Theorem 4.9, $\sigma = 1/\sqrt{2f_0(0)}$, whereas in (7.8) and (7.9), σ is $1/\sqrt{f_0(0)}$. This modification yields the factors of $2^{-2/5}$ and $2^{-1/5}$ appearing (twice each) on the right side of (4.15), as explained in Theorem 4.5 C.

8.2.5. Proof for the maximum functional limit theory

Proof of Theorem 4.11. If $\{F_n\}$ is a concave function sequence on $\mathbb{R}$ that converges uniformly on compacta to a function F, where F has a unique maximizer, then $N(F_n) \to N(F)$. Thus N is continuous on the subset of convex functions in $\mathcal{C}_\infty$ with the topology of uniform convergence on compacta. Let $\sigma = 1/\sqrt{f_0(m)}$, $a = |\varphi_0^{(2)}(m)|/4!$, $F_n(t) = n^{2/5}(\widehat{\varphi}_n(m + n^{-1/5}t) - \varphi_0(m))$, and $F(t) = \widehat{\varphi}_{a,\sigma}(t)$. Then by Theorem 4.8, F_n converges weakly to F so by the continuous mapping theorem $N(F_n) \to_d N(F)$. Now, by the scaling relationship (4.20), $N(F) =_d \sigma^{4/5}a^{1/5}N(\widehat{\varphi}((a/\sigma)^{2/5}\cdot)) = \sigma^{4/5}a^{1/5}N(\widehat{\varphi})$. We can check that $\sigma^{4/5}a^{1/5} = C(m, \varphi_0)$. Thus we have shown

$$n^{2/5}(\log N(\widehat{f}_n) - \log N(f_0)) \to_d C(m, \varphi_0)N(\widehat{\varphi})$$

since $N(F_n)$ is the left side of the above display. Applying the delta rule, we see also

$$n^{2/5}(N(\widehat{f}_n) - N(f_0)) \to_d c(m, \varphi_0)N(\widehat{\varphi}). \qquad \square$$

9. Technical lemmas

Here is a statement of the general integration by parts formulas for functions of bounded variation, used in our proof of Proposition 8.8. See, e.g., page 102 of Folland (1999) for the definition of bounded variation.

Lemma 9.1 (Folland (1999)). *Assume that F and G are of bounded variation on a set $[a, b]$ where $-\infty < a < b < \infty$*

A. *If at least one of F and G is continuous, then*

$$\int_{(a,b]} F dG + \int_{(a,b]} G dF = F(b)G(b) - F(a)G(a).$$

B. *If there are no points in $[a, b]$ where F and G are both discontinuous, then*

$$\int_{[a,b]} F dG + \int_{[a,b]} G dF = F(b)G(b) - F(a-)G(a-).$$

The next lemma is proved in Doss (2013b), page 143, for convex rather than concave functions.

Lemma 9.2 (Doss (2013b)). *Let g_1 and g_0 be concave functions on $[a, b]$, and let $t \in [a, b]$. Then*

$$\begin{aligned}
g_0(t) - g_1(t) &\le \frac{b-t}{b-a}(g_0(a) - g_1(a)) + \frac{t-a}{b-a}(g_0(b) - g_1(b)) \\
&\quad + \frac{(b-t)(t-a)}{b-a}(g_0'(a+) - g_0'(b-)).
\end{aligned} \qquad (9.1)$$

Acknowledgements

References

AZADBAKHSH, M., JANKOWSKI, H. and GAO, X. (2014). Computing confidence intervals for log-concave densities. *Comput. Statist. Data Anal.* **75** 248–264. MR3178372

BALABDAOUI, F. and DOSS, C. R. (2018). Inference for a two-component mixture of symmetric distributions under log-concavity. *Bernoulli* **24** 1053–1071. MR3706787

BALABDAOUI, F., RUFIBACH, K. and WELLNER, J. A. (2009). Limit distribution theory for maximum likelihood estimation of a log-concave density. *Ann. Statist.* **37** 1299–1331. MR2509075

BALABDAOUI, F. and WELLNER, J. A. (2007). Estimation of a k-monotone density: limit distribution theory and the spline connection. *Ann. Statist.* **35** 2536–2564. MR2382657

BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, second ed. *Wiley Series in Probability and Statistics: Probability and Statistics*. John Wiley & Sons Inc., New York. A Wiley-Interscience Publication. MR1700749

CHACÓN, J. E. (2018a). Mixture model modal clustering. *Advances in Data Analysis and Classification* **12** 1–26. MR3954514

CHACÓN, J. E. (2018b). The Modal Age of Statistics. *arXiv:1807.02789.*

CHANG, G. T. and WALTHER, G. (2007). Clustering with mixtures of log-concave distributions. *Comput. Statist. Data Anal.* **51** 6242–6251. MR2408591

CHEN, Y.-C., GENOVESE, C. R. and WASSERMAN, L. (2015). Asymptotic theory for density ridges. *Ann. Statist.* **43** 1896–1928. MR3375871

CHEN, Y.-C., GENOVESE, C. R., TIBSHIRANI, R. J. and WASSERMAN, L. (2016). Nonparametric modal regression. *Ann. Statist.* **44** 489–514. MR3476607

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

CULE, M., SAMWORTH, R. and STEWART, M. (2010). Maximum likelihood estimation of a multi-dimensional log-concave density. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** 545–607. MR2758237

DIEUDONNÉ, J. (1969). *Foundations of Modern Analysis*. Academic Press, New York-London. MR0349288

DOSS, C. R. (2013a). logcondens.mode: Compute MLE of Log-Concave Density on R with Fixed Mode, and Perform Inference for the Mode. R package ver-

sion 1.0.1. Available from http://CRAN.R-project.org/package=logcondens.mode.

Doss, C. R. (2013b). Shape-constrained inference for concave-transformed densities and their modes PhD thesis, University of Washington. MR3193124

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. With supplementary material available online. MR3485950

Doss, C. R. and Wellner, J. A. (2019). Inference for the mode of a log-concave density. *Annals of Statistics* **47** to appear. arXiv:1611.10348.

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. and Rufibach, K. (2011). logcondens: computations related to univariate log-concave density estimation. *Journal of Statistical Software* **39** 1–28.

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

Eilers, P. H. C. and Borgdorff, M. W. (2007). Non-parametric log-concave mixtures. *Comput. Statist. Data Anal.* **51** 5444–5451. MR2370883

Folland, G. B. (1999). *Real Analysis*, second ed. *Pure and Applied Mathematics (New York)*. John Wiley & Sons Inc., New York. Modern techniques and their applications, A Wiley-Interscience Publication. MR1681462

Groeneboom, P., Jongbloed, G. and Wellner, J. A. (2001a). A canonical process for estimation of convex functions: the "invelope" of integrated Brownian motion $+t^4$. *Ann. Statist.* **29** 1620–1652. MR1891741

Groeneboom, P., Jongbloed, G. and Wellner, J. A. (2001b). Estimation of a convex function: characterizations and asymptotic theory. *Ann. Statist.* **29** 1653–1698. MR1891742

Hall, P. (1984). On unimodality and rates of convergence for stable laws. *J. London Math. Soc. (2)* **30** 371–384. MR0771431

Han, Q. and Wellner, J. A. (2016). Approximation and estimation of *s*-concave densities via Rényi divergences. *Ann. Statist.* **44** 1332–1359. MR3485962

Kim, J. and Pollard, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191–219. MR1041391

Kim, A. K. H. and Samworth, R. J. (2016). Global rates of convergence in log-concave density estimation. *Ann. Statist.* **44** 2756–2779. MR3576560

Lachal, A. (1997). Local asymptotic classes for the successive primitives of Brownian motion. *Ann. Probab.* **25** 1712–1734. MR1487433

Laha, N. (2019). Some semiparametric models under log-concavity. PhD thesis in progress.

Mason, D. M. and van Zwet, W. R. (1987). A refinement of the KMT inequality for the uniform empirical process. *Ann. Probab.* **15** 871–884. MR0893903

Pal, J. K., Woodroofe, M. and Meyer, M. (2007). Estimating a Polya

frequency function$_2$. In *Complex datasets and inverse problems. IMS Lecture Notes Monogr. Ser.* **54** 239–249. Inst. Math. Statist., Beachwood, OH. MR2459192

Polonik, W. (1995). Measuring mass concentrations and estimating density contour clusters—an excess mass approach. *Ann. Statist.* **23** 855–881. MR1345204

Pu, X. and Arias-Castro, E. (2017). Semiparametric estimation of symmetric mixture models with monotone and log-concave densities. *arXiv:1702.08897.*

Qiao, W. and Polonik, W. (2016). Theoretical analysis of nonparametric filament estimation. *Ann. Statist.* **44** 1269–1297. MR3485960

R Core Team, (2016). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rockafellar, R. T. (1970). *Convex Analysis. Princeton Mathematical Series, No. 28.* Princeton University Press, Princeton, N.J. MR0274683

Romano, J. P. (1987). On bootstrapping the joint distribution of the location and size of the mode PhD thesis, University of California, Berkeley.

Royden, H. L. (1988). *Real Analysis*, Third ed. Macmillan Publishing Company, New York. MR1013117

Rufibach, K. (2006). Log-concave density estimation and bump hunting for IID observations. PhD thesis, Univ. Bern and Gottingen.

Seregin, A. and Wellner, J. A. (2010). Nonparametric estimation of multivariate convex-transformed densities. *Ann. Statist.* **38** 3751–3781. With supplementary material available online. MR2766867

Shorack, G. R. (2000). *Probability for Statisticians. Springer Texts in Statistics.* Springer-Verlag, New York. MR1762415

Shorack, G. R. and Wellner, J. A. (2009). *Empirical Processes with Applications to Statistics. Classics in Applied Mathematics* **59**. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. MR3396731

Silverman, B. W. (1982). On the estimation of a probability density function by the maximum penalized likelihood method. *Ann. Statist.* **10** 795–810. MR0663433

Sinai, Y. G. (1992). Statistics of shocks in solutions of inviscid Burgers equation. *Comm. Math. Phys.* **148** 601–621. MR1181071

Skorokhod, A. V. (1956). Limit theorems for stochastic processes. *Theory Probab. Appl.* **1** 261. MR0084897

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes. Springer Series in Statistics.* Springer-Verlag, New York. With applications to statistics. MR1385671

Walther, G. (2002). Detecting the presence of mixing with multiscale maximum likelihood. *J. Amer. Statist. Assoc.* **97** 508–513. MR1941467

Walther, G. (2009). Inference and modeling with log-concave distributions. *Statist. Sci.* **24** 319–327. MR2757433

Watanabe, H. (1970). An asymptotic property of Gaussian processes. *Trans. Amer. Math. Soc.* **148** 233. MR0256478

Whitt, W. (1970). Weak convergence of probability measures on the function

space $C[0, \infty)$. *Ann. Math. Stat.* **41** 939–944. MR0261646

Whitt, W. (1980). Some useful functions for functional limit theorems. *Math. Oper. Res.* **5** 67–85. MR0561155

Whitt, W. (2002). *Stochastic-Process Limits. Springer Series in Operations Research.* Springer-Verlag, New York. MR1876437

Xu, M. and Samworth, R. J. (2017). High-dimensional nonparametric density estimation via symmetry and shape constraints. Technical Report, Available at http://www.statslab.cam.ac.uk/~rjs57/Research.html.