Supplementary material for "Inference for the mode of a log-concave density"

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: In this supplement we present proofs for Doss and Wellner [2018a]. Equation, theorem, or assumption references made to the main document do not contain letters, and those to the supplement do.

AMS 2000 subject classifications: Primary 62G07; secondary 62G15, 62G10, 62G20.
Keywords and phrases: mode, empirical processes, likelihood ratio, pivot, convex optimization, log-concave, shape constraints.

Contents

[*]Supported in part by NSF Grants DMS-1712664, DMS-1104832 and a University of Minnesota Grant-In-Aid grant

[†]Supported in part by NSF Grants DMS-1104832 and DMS-1566514, NI-AID grant 2R01 AI291968-04, and by the Isaac Newton Institute for Mathematical Sciences program *Statistical Scalability*, EPSRC Grant Number LNAG/036 RG91310.

This supplement presents detailed arguments for the theorems and propositions of Doss and Wellner [2018a]. Appendices A.1–A.3 contain the proofs of the main result, Theorem 1.1. The global consistency Theorem 2.1 provides one of the main tools used in the proofs given in Appendices A.1–A.3. Its proof is given in Appendix A.4. Appendix B.1 reviews the local limit processes and the key scaling relations satisfied by these processes, while Appendix B.2 provides corrections for some typographical errors in Balabdaoui, Rufibach and Wellner [2009]. Finally, Appendix C summarizes key technical lemmas used in the proofs.

A. Proofs

We now deal with the remainder terms defined in (4.9) in the course of our "proof sketch" for Theorem 1.1. We first deal with the "local" remainder terms $R_{n,j},\ R^0_{n,j}$ with $j \in \{2,3\}$ in Subsection A.1. The analysis of these local remainder terms depends crucially on Theorem 2.1. Subsection A.2 is dedicated to the proofs for the "non-local" remainder terms.

For a function $f : \mathbb{R} \to \mathbb{R}$, we let $\|f\| := \sup_{x\in\mathbb{R}} |f(x)|$, and for a set $J \subset \mathbb{R}$ we let $\|f\|_J := \sup_{x\in J} |f(x)|$. Recall also the following two key assumptions from Section 3 of Doss and Wellner [2018a] where they appear as Assumption 1 and Assumption 2:

Assumption A.1. *(Curvature at m) Suppose that $X_1, \ldots, X_n$ are i.i.d. $f_0 = e^{\varphi_0} \in \mathcal{P}_m$ and that φ_0 is twice continuously differentiable at m with $\varphi_0''(m) < 0$.*

Assumption A.2. *(Curvature at $x_0 \neq m$) Suppose that $X_1, \ldots, X_n$ are i.i.d. $f_0 = e^{\varphi_0} \in \mathcal{P}_m$ and that φ_0 is twice continuously differentiable at $x_0 \neq m$ with $\varphi_0''(x_0) < 0$ and $f_0(x_0) > 0$.*

A.1. The local remainder terms $R_{n,j}, R^0_{n,j}$, $j \in \{2, 3\}$

We first deal with the (easy) local remainder terms.

Proposition A.1. *Let $t_{n,1} = m - Mn^{-1/5}$ and $t_{n,2} = m + Mn^{-1/5}$ for $M > 0$. Then the remainder terms $R_{n,2}$, $R^0_{n,2}$, $R_{n,3}$, and $R^0_{n,3}$ satisfy $nR_{n,j} = o_p(1)$ and $nR^0_{n,j} = o_p(1)$ for $j \in \{2, 3\}$.*

Proof. Recall that the remainder terms $R_{n,2}$, $R^0_{n,2}$, $R_{n,3}$, and $R^0_{n,3}$ given by (4.4), (4.5), (4.7), and (4.8) are all of the form a constant times

$$\tilde{R}_n \equiv \int_{D_n} e^{\tilde{x}_n(u)} (\widehat{\varphi}_n(u) - \varphi_0(m))^3 du, \quad \text{or}$$
$$\tilde{R}^0_n \equiv \int_{D_n} e^{\tilde{x}^0_n(u)} (\widehat{\varphi}^0_n(u) - \varphi_0(m))^3 du,$$

where D_n is a (possibly random) interval of length $O_p(n^{-1/5})$ and $\tilde{x}_{n,j}$ converges in probability, uniformly in $u \in D_n$, to zero. But by Assumption A.1 and by

Theorem 2.1 Part A it follows that for any $M > 0$ we have

$$\begin{aligned}
&\sup_{|t|\le M} |\widehat{\varphi}_n(m+n^{-1/5}t) - \varphi_0(m)| \\
&\le \sup_{|t|\le M} \left\{ |\widehat{\varphi}_n(m+n^{-1/5}t) - \varphi_0(m+n^{-1/5}t)| + |\varphi_0(m+n^{-1/5}t) - \varphi_0(m)| \right\} \\
&= O_p((n^{-1}\log n)^{2/5}) + O(n^{-2/5}) = O_p((n^{-1}\log n)^{2/5}),
\end{aligned}$$

and hence

$$n|\tilde{R}_n| = nO_p((n^{-1}\log n)^{6/5} \cdot n^{-1/5}) = O_p(n^{-2/5}(\log n)^{6/5}) = o_p(1),$$

By Assumption A.1 and by Theorem 2.1 Part B it follows that for any $M > 0$ we have

$$n|\tilde{R}_n^0| = nO_p((n^{-1}\log n)^{6/5} \cdot n^{-1/5}) = O_p(n^{-2/5}(\log n)^{6/5}) = o_p(1).$$

Note that the o_p terms do not depend on M, by Theorem 2.1. This completes the proof of negligibility of the local error terms $R_{n,j}, R_{n,j}^0$, $j \in \{2,3\}$. □

A.2. The global remainder terms $R_{n,1}$ and $R_{n,1}^c$

Recall that the remainder terms $R_{n,1}$ and $R_{n,1}^c$ are given by (4.2) and (4.3). Note that the integral in the definition of (4.3) is over $[X_{(1)}, X_{(n)}] \setminus D_n$, and hence this term in particular has a global character. We will see later that $R_{n,1}$ also can be seen as having a global nature.

Outline: From now on, we will focus our analysis on the portion of $R_{n,1,t_1,t_2}^c$ given by integrating over the left side, $[X_{(1)}, t_1]$. Arguments for the integral over $[t_2, X_{(n)}]$ are analogous. Thus, by a slight abuse of notation, define the one-sided counterpart to $R_{n,1,t_1,t_2}^c$ from (4.3) for any $t < m$ by

$$R_{n,1,t}^c \equiv \int_{[X_{(1)},t]} \widehat{\varphi}_n d\widehat{F}_n - \widehat{\varphi}_n^0 d\widehat{F}_n^0 - \int_{[X_{(1)},t]} (e^{\widehat{\varphi}_n} - e^{\widehat{\varphi}_n^0})\, d\lambda. \tag{A.1}$$

Here λ is Lebesgue measure (and is unrelated to the likelihood ratio λ_n). The analysis of $R_{n,1,t}^c$ is the greatest difficulty in understanding $2\log\lambda_n$. The proof that $R_{n,1,t_n}^c$ is $o_p(n^{-1})$ when $b \to \infty$ where $t_n = m - bn^{-1/5}$ is somewhat lengthy so we provide an outline here.

1. **Step 1, Decomposition of $R_{n,1,t}^c$:** Decompose $R_{n,1,t}^c$, to see that

$$R_{n,1,t}^c = A_{n,t}^1 + E_{n,t}^1 - T_{n,t}^1 = A_{n,t}^2 + E_{n,t}^2 + T_{n,t}^2, \tag{A.2}$$

   where the summands $A_{n,t}^i, E_{n,t}^i, T_{n,t}^i$ are defined below (see (A.11) and the preceding text).

2. **Step 2, Global $O_p(n^{-1})$ conclusion:** In this section we use the fact that away from the mode, the characterizations of $\widehat{\varphi}_n$ and $\widehat{\varphi}_n^0$ are identical to study T_n^i, $i = 1, 2$, which are related to $\int_{[X_{(1)},t]} (\widehat{\varphi}_n - \widehat{\varphi}_n^0)^2 \widehat{f}_n d\lambda$. and $\int_{[X_{(1)},t]} (\widehat{\varphi}_n - \widehat{\varphi}_n^0)^2 \widehat{f}_n^0 d\lambda$. We will show $T_n^i = O_p(n^{-1})$, $i = 1, 2$. Note $O_p(n^{-1})$ would be the size of the integral if it were over a local interval of length $O_p(n^{-1/5})$ (under our curvature assumptions), but here the integral is over an interval of constant length or larger, so this result is global in nature.
3. **Step 3, Convert global O_p to local o_p to global o_p:** Convert the global $O_p(n^{-1})$ conclusion over T_n^i into an $o_p(n^{-1})$ conclusion over a interval of length $O_p(n^{-1/5})$ local to m. Feed this result back into the argument in Step 2, yielding $T_{n,t}^i = o_p(n^{-1})$, $i = 1, 2$. Apply Lemma C.2 to show additionally that there exist knots of $\widehat{\varphi}_n$ and $\widehat{\varphi}_n^0$ that are $o_p(n^{-1/5})$ apart in an $O_p(n^{-1/5})$ length interval on which $\|\widehat{\varphi}_n^0 - \widehat{\varphi}_n\| = o_p(n^{-2/5})$, $\|\widehat{f}_n^0 - \widehat{f}_n\| = o_p(n^{-2/5})$, and $\|\widehat{F}_n^0 - \widehat{F}_n\| = o_p(n^{-3/5})$.
4. **Step 4, Concluding arguments:** Return to the decomposition of $R_{n,1,t}^c$ given in Step 1; the terms given there depend on $\widehat{\varphi}_n^0 - \widehat{\varphi}_n$, $\widehat{f}_n^0 - \widehat{f}_n$, and $\widehat{F}_n^0 - \widehat{F}_n$. Thus, using the results of Step 3 we can show $nR_{n,1,t}^c = o_p(1)$ as desired.

To finalize the argument, in Section A.3, we take $t_n = m - bn^{-1/5}$, but we also need to let $b \to \infty$. Thus, the O_p and o_p statements above need to hold uniformly in b.

A.2.1. Decomposition of $R_{n,1,t}^c$

We begin by decomposing $R_{n,1,t}^c$ for fixed $t < m$. By (C.2) with $\varphi_{1n} = \widehat{\varphi}_n$ and $\varphi_{2n} = \widehat{\varphi}_n^0$, we see that

$$
\begin{aligned}
R_{n,1,t}^c &= \int_{[X_{(1)},t]} \left(\widehat{\varphi}_n \widehat{f}_n - \widehat{\varphi}_n^0 \widehat{f}_n^0 - \left(\widehat{\varphi}_n - \widehat{\varphi}_n^0 + \frac{(\widehat{\varphi}_n - \widehat{\varphi}_n^0)^2}{2} e^{\varepsilon_n^1} \right) \widehat{f}_n^0 \right) d\lambda \\
&= \int_{[X_{(1)},t]} \left(\widehat{\varphi}_n \widehat{f}_n - \widehat{\varphi}_n \widehat{f}_n^0 - \frac{(\widehat{\varphi}_n - \widehat{\varphi}_n^0)^2}{2} e^{\varepsilon_n^1} \widehat{f}_n^0 \right) d\lambda \\
&= \int_{[X_{(1)},t]} \left(\widehat{\varphi}_n (\widehat{f}_n - \widehat{f}_n^0) - \frac{(\widehat{\varphi}_n - \widehat{\varphi}_n^0)^2}{2} e^{\varepsilon_n^1} \widehat{f}_n^0 \right) d\lambda, \qquad \text{(A.3)}
\end{aligned}
$$

where λ is Lebesgue measure and $\varepsilon_n^1(x)$ lies between 0 and $\widehat{\varphi}_n(x)-\widehat{\varphi}_n^0(x)$. Again applying (C.2) now with $\varphi_{1n}=\widehat{\varphi}_n^0$ and $\varphi_{2n}=\widehat{\varphi}_n$, we see that

$$\begin{aligned} R^c_{n,1,t} &= \int_{[X_{(1)},t]} \left(\widehat{\varphi}_n \widehat{f}_n - \widehat{\varphi}_n^0 \widehat{f}_n^0 + \left(e^{\widehat{\varphi}_n^0-\widehat{\varphi}_n} - 1\right) \widehat{f}_n \right) d\lambda \\ &= \int_{[X_{(1)},t]} \left(\widehat{\varphi}_n \widehat{f}_n - \widehat{\varphi}_n^0 \widehat{f}_n^0 + \left(\widehat{\varphi}_n^0 - \widehat{\varphi}_n + \frac{(\widehat{\varphi}_n^0-\widehat{\varphi}_n)^2}{2} e^{\varepsilon_n^2} \right) \widehat{f}_n \right) d\lambda \\ &= \int_{[X_{(1)},t]} \left(\widehat{\varphi}_n^0 (\widehat{f}_n - \widehat{f}_n^0) + \frac{(\widehat{\varphi}_n^0-\widehat{\varphi}_n)^2}{2} e^{\varepsilon_n^2} \widehat{f}_n \right) d\lambda, \end{aligned} \tag{A.4}$$

where ε_n^2 lies between 0 and $\widehat{\varphi}_n^0(x)-\widehat{\varphi}_n(x)$. For a function $f(x)$, recall the notation $f_s(x)=f(x)-f(s)$ for $x\le s$ and $f_s(x)=0$ for $x\ge s$. Now define $A_{n,t}^i \equiv A_n^i$, $i=1,2$ by

$$A_n^1 \equiv \int_{[X_{(1)},t]} \widehat{\varphi}_{n,t}\, d\left(\mathbb{F}_n - \widehat{F}_n^0\right) \quad \text{and} \quad A_n^2 \equiv \int_{[X_{(1)},t]} \widehat{\varphi}_{n,t}^0 d\left(\widehat{F}_n - \mathbb{F}_n\right) \tag{A.5}$$

and define $E_{n,t}^1 \equiv E_n^1$ to be

$$\begin{aligned} &\int_{(\tau,t]} \widehat{\varphi}_{n,t}\, d\left(\widehat{F}_n - \mathbb{F}_n\right) + \widehat{\varphi}_n(t)(\widehat{F}_n(t) - \widehat{F}_n^0(t)) \\ &\qquad + (\widehat{\varphi}_n(\tau) - \widehat{\varphi}_n(t))(\widehat{F}_n(\tau) - \mathbb{F}_n(\tau)) \end{aligned} \tag{A.6}$$

and $E_{n,t}^2 \equiv E_n^2$ to be

$$\begin{aligned} &\int_{(\tau^0,t]} \widehat{\varphi}_{n,t}^0 d\left(\mathbb{F}_n - \widehat{F}_n^0\right) + \widehat{\varphi}_n^0(t)(\widehat{F}_n(t) - \widehat{F}_n^0(t)) \\ &\qquad + (\widehat{\varphi}_n^0(\tau^0) - \widehat{\varphi}_n^0(t))(\mathbb{F}_n(\tau^0) - \widehat{F}_n^0(\tau^0)), \end{aligned} \tag{A.7}$$

where $\tau \equiv \tau_-(t) = \sup S(\widehat{\varphi}_n) \cap (-\infty,t]$ and $\tau^0 \equiv \tau_-^0(t) = \sup S(\widehat{\varphi}_n^0) \cap (-\infty,t]$. We will assume that

$$\tau \le \tau^0$$

without loss of generality, because the arguments are symmetric in $\widehat{\varphi}_n$ and $\widehat{\varphi}_n^0$, since we will be arguing entirely on one side of the mode.

Our next lemma will decompose the first terms in (A.3) and (A.4), into $A_n^i + E_n^i$, $i=1,2$. The crucial observation is that $A_n^1 \le 0$ and $A_n^2 \ge 0$, by taking $\Delta = \widehat{\varphi}_{n,t}$ and $\Delta = \widehat{\varphi}_{n,t}^0$ in the characterization theorems for the constrained and unconstrained MLEs, Theorem 2.2 A and B of Doss and Wellner [2018b]. Note that since $t \le m$, $\widehat{\varphi}_{n,t}$ has modal interval containing m.

Lemma A.2. *Let all terms be as defined above. We then have*

$$\int_{[X_{(1)},t]} \widehat{\varphi}_n (\widehat{f}_n - \widehat{f}_n^0) d\lambda = A_{n,t}^1 + E_{n,t}^1 \tag{A.8}$$

and

$$\int_{[X_{(1)},t]} \widehat{\varphi}_n^0 (\widehat{f}_n - \widehat{f}_n^0) d\lambda = A_{n,t}^2 + E_{n,t}^2. \tag{A.9}$$

Proof. We first show (A.8). We can see $\int_{[X_{(1)},t]} \widehat{\varphi}_n(\widehat{f}_n - \widehat{f}_n^0)$ equals

$$\int_{[X_{(1)},t]} (\widehat{\varphi}_{n,\tau_-} + \widehat{\varphi}_n - \widehat{\varphi}_{n,\tau_-})\widehat{f}_n d\lambda - \int_{[X_{(1)},t]} (\widehat{\varphi}_{n,t} + \widehat{\varphi}_n - \widehat{\varphi}_{n,t})\widehat{f}_n^0 d\lambda,$$

and since $\int \widehat{\varphi}_{n,\tau_-}\, d\left(\mathbb{F}_n - \widehat{F}_n\right) = 0$, this equals

$$\begin{aligned}
&\int_{[X_{(1)},\tau_-]} \widehat{\varphi}_{n,\tau_-} d\mathbb{F}_n + \int_{[X_{(1)},\tau_-]} (\widehat{\varphi}_n - \widehat{\varphi}_{n,\tau_-})\widehat{f}_n d\lambda + \int_{(\tau_-,t]} \widehat{\varphi}_n \widehat{f}_n d\lambda \\
&\quad - \left(\int_{[X_{(1)},t]} \widehat{\varphi}_{n,t} \widehat{f}_n^0 d\lambda + \widehat{\varphi}_n(t)\widehat{F}_n^0(t) \right) \\
&\quad = \int_{[X_{(1)},t]} \widehat{\varphi}_{n,t} d\mathbb{F}_n + \int_{[X_{(1)},\tau_-]} (\widehat{\varphi}_{n,\tau_-} - \widehat{\varphi}_{n,t}) d\mathbb{F}_n - \int_{(\tau_-,t]} \widehat{\varphi}_{n,t} d\mathbb{F}_n \\
&\quad\quad + \int_{[X_{(1)},\tau_-]} \widehat{\varphi}_n(\tau_-)\widehat{f}_n d\lambda + \int_{(\tau_-,t]} \widehat{\varphi}_n \widehat{f}_n d\lambda - \left(\int \widehat{\varphi}_{n,t} \widehat{f}_n^0 d\lambda + \widehat{\varphi}_n(t)\widehat{F}_n^0(t) \right) \\
&\quad = \int_{[X_{(1)},t]} \widehat{\varphi}_{n,t}\, d\left(\mathbb{F}_n - \widehat{F}_n^0\right) + (\widehat{\varphi}_n(t) - \widehat{\varphi}_n(\tau_-))\, \mathbb{F}_n(\tau_-) - \int_{(\tau_-,t]} \widehat{\varphi}_n d\mathbb{F}_n \\
&\quad\quad + \widehat{\varphi}_n(t)(\mathbb{F}_n(t) - \mathbb{F}_n(\tau_-)) + \widehat{\varphi}_n(\tau_-)\widehat{F}_n(\tau_-) + \int_{(\tau_-,t]} \widehat{\varphi}_n \widehat{f}_n d\lambda - \widehat{\varphi}_n(t)\widehat{F}_n^0(t),
\end{aligned}$$

which equals

$$\begin{aligned}
&\int \widehat{\varphi}_{n,t} d\left(\mathbb{F}_n - \widehat{F}_n^0\right) + \int_{(\tau_-,t]} \widehat{\varphi}_n d\left(\widehat{F}_n - \mathbb{F}_n\right) \\
&\quad + \widehat{\varphi}_n(t)(\mathbb{F}_n(t) - \widehat{F}_n^0(t)) + \widehat{\varphi}_n(\tau_-)(\widehat{F}_n(\tau_-) - \mathbb{F}_n(\tau_-))
\end{aligned}$$

which equals

$$\begin{aligned}
&\int \widehat{\varphi}_{n,t} d\left(\mathbb{F}_n - \widehat{F}_n^0\right) + \int_{(\tau_-,t]} \widehat{\varphi}_{n,t} d\left(\widehat{F}_n - \mathbb{F}_n\right) \\
&\quad + \widehat{\varphi}_n(t)(\widehat{F}_n(t) - \widehat{F}_n^0(t)) + (\widehat{\varphi}_n(\tau_-) - \widehat{\varphi}_n(t))(\widehat{F}_n(\tau_-) - \mathbb{F}_n(\tau_-)),
\end{aligned}$$

as desired.

Now we show (A.9). We see $\int_{[X_{(1)},t]} \widehat{\varphi}_n^0(\widehat{f}_n - \widehat{f}_n^0) d\lambda$ equals

$$\int_{[X_{(1)},t]} \left(\widehat{\varphi}_{n,t}^0 + \widehat{\varphi}_n^0 - \widehat{\varphi}_{n,t}\right) \widehat{f}_n d\lambda - \int_{[X_{(1)},t]} \left(\widehat{\varphi}_{n,\tau_-^0}^0 + \widehat{\varphi}_n^0 - \widehat{\varphi}_{n,\tau_-^0}^0\right) \widehat{f}_n^0 d\lambda$$

and since $\int \widehat{\varphi}_{n,\tau_-^0}^0\, d(\mathbb{F}_n - \widehat{F}_n^0) = 0$, this equals

$$\begin{aligned}
&\int_{[X_{(1)},t]} \widehat{\varphi}_{n,t}^0 \widehat{f}_n d\lambda + \widehat{\varphi}_n^0(t)\widehat{F}_n(t) - \left[\int \widehat{\varphi}_{n,\tau_-^0}^0 d\mathbb{F}_n + \int_{[X_{(1)},\tau_-^0]} \widehat{\varphi}_n^0(\tau_-^0)\widehat{f}_n^0 d\lambda \right. \\
&\quad\quad \left. + \int_{(\tau_-^0,t]} \widehat{\varphi}_n^0 \widehat{f}_n^0 d\lambda \right],
\end{aligned}$$

which equals

$$\int_{[X_{(1)},t]} \widehat{\varphi}^0_{n,t}\widehat{f}_n d\lambda + \widehat{\varphi}^0_n(t)\widehat{F}_n(t) - \Bigg[\int_{[X_{(1)},t]} \widehat{\varphi}^0_{n,t} d\mathbb{F}_n + \int_{[X_{(1)},t]} (\widehat{\varphi}^0_{n,\tau^0_-} - \widehat{\varphi}^0_{n,t}) d\mathbb{F}_n + \widehat{\varphi}^0_n(\tau^0_-)\widehat{F}^0_n(\tau^0_-) + \int_{(\tau^0_-,t]} \widehat{\varphi}^0_n \widehat{f}^0_n d\lambda \Bigg]$$

which equals

$$\int_{[X_{(1)},t]} \widehat{\varphi}^0_{n,t} d(\widehat{F}_n - \mathbb{F}_n) + \widehat{\varphi}^0_n(t)\widehat{F}_n(t) - \Bigg[\int_{[X_{(1)},\tau^0_-]} (\widehat{\varphi}^0_n(t) - \widehat{\varphi}^0_n(\tau^0_-)) d\mathbb{F}_n - \int_{(\tau^0,t]} \widehat{\varphi}^0_n d\mathbb{F}_n + \int_{(\tau^0_-,t]} \widehat{\varphi}^0_n(t) d\mathbb{F}_n + \widehat{\varphi}^0_n(\tau^0_-)\widehat{F}^0_n(\tau^0_-) + \int_{(\tau^0_-,t]} \widehat{\varphi}^0_n \widehat{f}^0_n d\lambda \Bigg]$$

which equals

$$\int_{[X_{(1)},t]} \widehat{\varphi}^0_{n,t} d(\widehat{F}_n - \mathbb{F}_n) + \int_{(\tau^0_-,t]} \widehat{\varphi}^0_n d(\mathbb{F}_n - \widehat{F}^0_n) + \widehat{\varphi}^0_n(t)\widehat{F}_n(t) + \widehat{\varphi}^0_n(\tau^0_-)\mathbb{F}_n(\tau^0_-) - \widehat{\varphi}^0_n(t)\mathbb{F}_n(\tau^0_-) - \widehat{\varphi}^0_n(t)(\mathbb{F}_n(t) - \mathbb{F}_n(\tau^0_-)) - \widehat{\varphi}^0_n(\tau^0_-)\widehat{F}^0_n(\tau^0_-)$$

which equals

$$\int_{[X_{(1)},t]} \widehat{\varphi}^0_{n,t} d(\widehat{F}_n - \mathbb{F}_n) + \int_{(\tau^0_-,t]} \widehat{\varphi}^0_n d(\mathbb{F}_n - \widehat{F}^0_n) + \widehat{\varphi}^0_n(t)(\widehat{F}_n(t) - \mathbb{F}_n(t)) + \widehat{\varphi}^0_n(\tau^0_-)(\mathbb{F}_n(\tau^0_-) - \widehat{F}^0_n(\tau^0_-))$$

which equals

$$\int_{[X_{(1)},t]} \widehat{\varphi}^0_{n,t} d\left(\widehat{F}_n - \mathbb{F}_n\right) + \int_{(\tau^0_-,t]} \widehat{\varphi}^0_{n,t} d\left(\mathbb{F}_n - \widehat{F}^0_n\right) + \widehat{\varphi}^0_n(t)(\widehat{F}_n(t) - \widehat{F}^0_n(t)) + (\widehat{\varphi}^0_n(\tau^0_-) - \widehat{\varphi}^0_n(t))(\mathbb{F}_n(\tau^0_-) - \widehat{F}^0_n(\tau^0_-)),$$

as desired. □

Define $T^i_{n,t} \equiv T^i_n$, $i = 1, 2$, by

$$T^1_n = \int_{[X_{(1)},t]} \frac{(\widehat{\varphi}_n - \widehat{\varphi}^0_n)^2}{2} e^{\varepsilon^1_n} \widehat{f}^0_n d\lambda \quad \text{and} \quad T^2_n = \int_{[X_{(1)},t]} \frac{(\widehat{\varphi}_n - \widehat{\varphi}^0_n)^2}{2} e^{\varepsilon^2_n} \widehat{f}_n d\lambda, \tag{A.10}$$

so that

$$R^c_{n,1,t} = A^1_{n,t} + E^1_{n,t} - T^1_{n,t} = A^2_{n,t} + E^2_{n,t} + T^2_{n,t}. \tag{A.11}$$

by (A.3) and (A.4). Recall (from page 5) that $A^1_n \le 0 \le A^2_n$. Thus

$$E^1_n - E^2_n \ge E^1_n - E^2_n - T^2_n - T^1_n = A^2_n - A^1_n \ge \begin{cases} A^2_n \ge 0, \\ -A^1_n \ge 0. \end{cases} \tag{A.12}$$

To see that $R^c_{n,1} = O_p(n^{-1})$ we need to see that E^i_n, A^i_n, and T^i_n are each $O_p(n^{-1})$ (for, say, $i = 1$). We can see already that $E^1_n - E^2_n = O_p(n^{-1})$ (by direct analysis of the terms in $E^1_n - E^2_n$ from (A.6) and (A.7)), which yields that A^1_n and A^2_n are both $O_p(n^{-1})$. However, it is clear that we also need to analyze $T^1_n + T^2_n$ to understand $R^c_{n,1}$. We need to show that $T^1_{n,t} + T^2_{n,t}$ is $O_p(n^{-1})$ to see that $R^c_{n,1,t} = O_p(n^{-1})$; but we will also be able to use that $T^1_{n,t} + T^2_{n,t} = O_p(n^{-1})$ to then find t^* values such that $T^1_{n,t^*} + T^2_{n,t^*} = o_p(n^{-1})$, which will allow us to argue in fact that $E^1_{n,t^*} + E^2_{n,t^*} = o_p(n^{-1})$ (rather than just $O_p(n^{-1})$), and thus that $R^c_{n,1,t^*} = o_p(n^{-1})$, as is eventually needed. Thus, we will now turn our attention to studying $T^1_n + T^2_n$. Afterwards, we will study

$$R^c_{n,1,t} = (A^1_{n,t} + E^1_{n,t} - T^1_{n,t} + A^2_{n,t} + E^2_{n,t} + T^2_{n,t})/2, \tag{A.13}$$

from (A.11). From seeing $T^1_{n,t^*} + T^2_{n,t^*} = o_p(n^{-1})$, we will be able to conclude that $A^1_{n,t^*} + A^2_{n,t^*}$ and $E^1_{n,t^*} + E^2_{n,t^*}$ are also $o_p(n^{-1})$, as desired. Then we can conclude $R^c_{n,1,t^*} = o_p(n^{-1})$.

A.2.2. Show $T^i_n = O_p(n^{-1})$, $i = 1, 2$

The next lemma shows that terms that are nearly identical to T^i_n are $O_p(n^{-1})$. The difference between the integrand in the terms in the lemma and the integrand defining T^i_n is that ε^i_n is replaced by a slightly different $\tilde\varepsilon^i_n$. Previously, we considered t to be fixed, whereas now we will have it vary with n.

Lemma A.3. *Let $t_n < m$ be a (potentially random) sequence such that*

$$t_n \le \max\left(S(\widehat\varphi_n) \cup S(\widehat\varphi^0_n)\right) \cap (-\infty, m). \tag{A.14}$$

Let

$$\tilde T^1_{n,t} = \int_{[X_{(1)}, t_n]} \left(\widehat\varphi^0_n - \widehat\varphi_n\right)^2 e^{\tilde\varepsilon^1_n} \widehat f^0_n d\lambda \quad \text{and} \quad \tilde T^2_{n,t} = \int_{[X_{(1)}, t_n]} \left(\widehat\varphi^0_n - \widehat\varphi_n\right)^2 e^{\tilde\varepsilon^2_n} \widehat f_n d\lambda, \tag{A.15}$$

where $\tilde\varepsilon^1_n(x)$ lies between $\widehat\varphi_n(x) - \widehat\varphi^0_n(x)$ and 0, and $\tilde\varepsilon^2_n(x)$ lies between $\widehat\varphi^0_n(x) - \widehat\varphi_n(x)$ and 0, and are defined in (A.23) in the proof. Then we have

$$\tilde T^i_{n,t_n} = O_p(n^{-1}) \quad \text{for } i = 1, 2. \tag{A.16}$$

Proof. For a function $f(x)$, recall the notation $f_s(x) = f(x) - f(s)$ for $x \le s$ and $f_s(x) = 0$ for $x \ge s$. Let $\tau \in S(\widehat\varphi_n)$ and $\tau^0 \in S(\widehat\varphi^0_n)$, and assume that

$$\tau \le \tau^0 < m. \tag{A.17}$$

(The argument is symmetric in $\widehat\varphi_n$ and $\widehat\varphi^0_n$, so we may assume this without loss of generality.) We will show the lemma holds for the case $t_n = \tau^0$, and then the general $t_n \le \tau^0$ case follows since the integral is increasing in t_n. Now, because

$\widehat{\varphi}^0_{n,\tau}$ is concave, for $\varepsilon \le 1$, the function $\widehat{\varphi}_n(x) + \varepsilon(\widehat{\varphi}^0_{n,\tau}(x) - \widehat{\varphi}_{n,\tau}(x))$ is concave. So, by Theorem 2.2, page 43, of Dümbgen and Rufibach [2009], we have

$$\int (\widehat{\varphi}^0_{n,\tau} - \widehat{\varphi}_{n,\tau}) d(\mathbb{F}_n - \widehat{F}_n) \le 0. \tag{A.18}$$

Similarly, if τ^0 is a knot of $\widehat{\varphi}^0_n$ and is less than the mode, then since $\widehat{\varphi}^0_n(x) + \varepsilon\left(\widehat{\varphi}_{n,\tau^0}(x) - \widehat{\varphi}^0_{n,\tau^0}(x)\right)$ is concave with mode at m for ε small (since $\widehat{\varphi}_{n,\tau^0}(x) - \widehat{\varphi}^0_{n,\tau^0}(x)$ is only nonzero on the left side of the mode), by the characterization Theorem 2.2 B of Doss and Wellner [2018b], we have

$$\int (\widehat{\varphi}_{n,\tau^0} - \widehat{\varphi}^0_{n,\tau^0}) d(\mathbb{F}_n - \widehat{F}^0_n) \le 0.$$

Then setting $II^L_{n,\tau^0} := \int_{[X_{(1)},\tau^0]} \left(\widehat{\varphi}_n - \widehat{\varphi}^0_n\right) d\left(\mathbb{F}_n - \widehat{F}^0_n\right)$, we have

$$\begin{aligned} 0 &\ge \int_{[X_{(1)},\tau^0]} \left(\widehat{\varphi}_{n,\tau^0} - \widehat{\varphi}^0_{n,\tau^0}\right) d\left(\mathbb{F}_n - \widehat{F}^0_n\right) && \text{(A.19)} \\ &= II^L_{n,\tau^0} - \left(\widehat{\varphi}_n(\tau^0) - \widehat{\varphi}^0_n(\tau^0)\right)\left(\mathbb{F}_n(\tau^0) - \widehat{F}^0_n(\tau^0)\right). && \text{(A.20)} \end{aligned}$$

And setting $I^L_{n,\tau^0} := \int_{[X_{(1)},\tau^0]} \left(\widehat{\varphi}^0_n - \widehat{\varphi}_n\right) d\left(\mathbb{F}_n - \widehat{F}_n\right)$, we have

$$\begin{aligned} I^L_{n,\tau^0} = &\int_{[X_{(1)},\tau]} \left(\widehat{\varphi}^0_n(u) - \widehat{\varphi}^0_n(\tau)\right) d\left(\mathbb{F}_n - \widehat{F}_n\right)(u) \\ &- \int_{[X_{(1)},\tau]} \left(\widehat{\varphi}_n(u) - \widehat{\varphi}_n(\tau)\right) d\left(\mathbb{F}_n - \widehat{F}_n\right)(u) \\ &+ \left(\widehat{\varphi}^0_n(\tau) - \widehat{\varphi}_n(\tau)\right) \int_{[X_{(1)},\tau]} d\left(\mathbb{F}_n - \widehat{F}_n\right) \\ &+ \int_{(\tau,\tau^0]} \left(\widehat{\varphi}^0_n - \widehat{\varphi}_n\right) d\left(\mathbb{F}_n - \widehat{F}_n\right), \end{aligned} \tag{A.21}$$

and, since the first two summands together yield the left hand side of (A.18), we have

$$I^L_{n,\tau^0} \le \left(\widehat{\varphi}^0_n(\tau) - \widehat{\varphi}_n(\tau)\right) \int_{[X_{(1)},\tau]} d\left(\mathbb{F}_n - \widehat{F}_n\right) + \int_{(\tau,\tau^0]} \left(\widehat{\varphi}^0_n - \widehat{\varphi}_n\right) d\left(\mathbb{F}_n - \widehat{F}_n\right). \tag{A.22}$$

Now, we apply (C.1) of Lemma C.1 to see that

$$\begin{aligned} I^L_{n,\tau^0} + II^L_{n,\tau^0} &= \int_{[X_{(1)},\tau^0]} \left(\widehat{\varphi}^0_n - \widehat{\varphi}_n\right) d\left(\widehat{F}^0_n - \widehat{F}_n\right) \\ &= \begin{cases} \int_{X_{(1)}}^{\tau^0} \left(\widehat{\varphi}^0_n - \widehat{\varphi}_n\right)^2 e^{\tilde{\varepsilon}^1_n} \widehat{f}^0_n d\lambda \ge 0, \\ \int_{X_{(1)}}^{\tau^0} \left(\widehat{\varphi}^0_n - \widehat{\varphi}_n\right)^2 e^{\tilde{\varepsilon}^2_n} \widehat{f}_n d\lambda \ge 0, \end{cases} \end{aligned} \tag{A.23}$$

where $\tilde{\varepsilon}_n^2(x)$ lies between $\widehat{\varphi}_n^0(x) - \widehat{\varphi}_n(x)$ and 0 and $\tilde{\varepsilon}_n^1(x)$ lies between $\widehat{\varphi}_n(x) - \widehat{\varphi}_n^0(x)$ and 0. By (A.20) and (A.22), (A.23) is bounded above by

$$\begin{aligned}&\left(\widehat{\varphi}_n^0(\tau) - \widehat{\varphi}_n(\tau)\right) \int_{[X_{(1)},\tau]} d\left(\mathbb{F}_n - \widehat{F}_n\right) + \int_{(\tau,\tau^0]} \left(\widehat{\varphi}_n^0 - \widehat{\varphi}_n\right) d\left(\mathbb{F}_n - \widehat{F}_n\right) \\ &\quad + \left(\widehat{\varphi}_n(\tau^0) - \widehat{\varphi}_n^0(\tau^0)\right)\left(\mathbb{F}_n(\tau^0) - \widehat{F}_n^0(\tau^0)\right).\end{aligned} \tag{A.24}$$

By Proposition 7.1 of Doss and Wellner [2018b] and Lemma 4.5 of Balabdaoui, Rufibach and Wellner [2009], $\sup_{t\in[\tau,\tau^0]} \left|\widehat{\varphi}_n^0(t) - \widehat{\varphi}_n(t)\right| = O_p(n^{-2/5})$. By Corollary 2.5 of Dümbgen and Rufibach [2009],

$$\left|\int_{[X_{(1)},\tau]} d\left(\mathbb{F}_n - \widehat{F}_n\right)\right| \le 1/n,$$

so the first term in the above display is $O_p(n^{-7/5})$. Similarly, by Corollary 2.7B of Doss and Wellner [2018b], $\left|\mathbb{F}_n(\tau^0) - \widehat{F}_n^0(\tau^0)\right| \le 1/n$, so the last term in the previous display is $O_p(n^{-7/5})$. We can also see that the middle term in the previous display equals

$$\begin{aligned}&(\widehat{\varphi}_n^0 - \widehat{\varphi}_n)(\tau^0)(\mathbb{F}_n - \widehat{F}_n)(\tau^0) - (\widehat{\varphi}_n^0 - \widehat{\varphi}_n)(\tau)(\mathbb{F}_n - \widehat{F}_n)(\tau) \\ &\quad - \int_{(\tau,\tau^0]} (\mathbb{F}_n - \widehat{F}_n)(\widehat{\varphi}_n^0 - \widehat{\varphi}_n)' d\lambda.\end{aligned} \tag{A.25}$$

Now the middle term in the previous display is $O_p(n^{-7/5})$. For the last two terms, we apply Lemma C.4 taking $I = [\tau, \tau^0]$ to see that

$$\sup_{t\in(\tau,\tau^0]} n^{3/5} \left|\int_{(\tau,t]} d\left(\mathbb{F}_n - \widehat{F}_n\right)\right| = O_p(1).$$

Thus, using Proposition 7.1 of Doss and Wellner [2018b] and Lemma 4.5 of Balabdaoui, Rufibach and Wellner [2009], we have

$$\int_{(\tau,\tau^0]} (\mathbb{F}_n - \widehat{F}_n)(\widehat{\varphi}_n^0 - \widehat{\varphi}_n)' d\lambda = O_p(n^{-4/5}) \int_{(\tau,\tau^0]} d\lambda = O_p(n^{-1}), \tag{A.26}$$

so we have now shown that (A.25) is $O_p(n^{-1})$, so the middle term in (A.24) is $O_p(n^{-1})$. Thus, (A.24) is $O_p(n^{-1})$, and since (A.24) bounds (A.23) we can conclude that

$$\int_{X_{(1)}}^{\tau^0} \left(\widehat{\varphi}_n^0 - \widehat{\varphi}_n\right)^2 e^{\tilde{\varepsilon}_n^1} \widehat{f}_n^0 d\lambda = \int_{X_{(1)}}^{\tau^0} \left(\widehat{\varphi}_n^0 - \widehat{\varphi}_n\right)^2 e^{\tilde{\varepsilon}_n^2} \widehat{f}_n d\lambda = O_p(n^{-1}), \tag{A.27}$$

and so we are done. $\square$

Remark A.1. Note that if we computed the integrals in (A.27) over an interval of length $O_p(n^{-1/5})$, by using that the corresponding integrand is $O_p(n^{-4/5})$ (under smoothness/curvature assumptions), the integrals would be $O_p(n^{-1})$. However, (A.27) shows that the integrals are $O_p(n^{-1})$ over a larger interval whose length is constant or larger, with high probability. Thus we can use (A.27) to show that $\widehat{\varphi}_n^0 - \widehat{\varphi}_n$ must be of order smaller than $O_p(n^{-2/5})$ somewhere, and this line of reasoning will in fact show that $T_{n,t}^1$ and $T_{n,t}^2$ are $o_p(n^{-1})$ for certain t values.

Remark A.2. Having shown (A.16), it may seem that we can easily find a subinterval over which the corresponding integrals are $o_p(n^{-1})$ (or smaller), and that this should allow us to quickly finish up our proof. There is an additional difficulty, though, preventing us from naively letting $|t| \to \infty$: we need to control the corresponding integrals actually within small neighborhoods of m (of order $O_p(n^{-1/5})$), not just arbitrarily far away from m. This is because our asymptotic results for the limit distribution take place in $n^{-1/5}$ neighborhoods of m.

To connect the result about $\tilde{T}_n^i$ to the title of this section (which states $T_n^i = O_p(n^{-1})$), note that by Lemma C.5, $0 \le T_n^i \le 2\tilde{T}_n^i = O_p(n^{-1})$.

A.2.3. Local and Global $o_p(n^{-1})$ Conclusion

We will now find a subinterval I such that

$$\int_I \left(\widehat{\varphi}_n^0 - \widehat{\varphi}_n\right)^2 e^{\tilde{\varepsilon}_n^2} \widehat{f}_n d\lambda = o_p(n^{-1}).$$

We will argue by partitioning a larger interval over which the above integral is $O_p(n^{-1})$ into smaller subintervals. Let $\varepsilon > 0$. Let $L > 0$ be such that intervals of length $Ln^{-1/5}$ whose endpoints converge to m contain a knot from each of $\widehat{\varphi}_n$ and $\widehat{\varphi}_n^0$ with probability $1-\varepsilon$. Also let $\delta > 0$ and $\zeta = \delta/L$ which we take without loss of generality to be the reciprocal of an integer. By Proposition 7.3 of Doss and Wellner [2018b], fix $M \ge L$ large enough such that with probability $1 - \varepsilon\zeta$ for any random variable $\xi_n \to_p m$, $[\xi_n - Mn^{-1/5}, \xi_n + Mn^{-1/5}]$ contains knots of both $\widehat{\varphi}_n$ and of $\widehat{\varphi}_n^0$, when n is large enough. Now, each of the intervals

$$I_{jn} := (\tau^0 - Mjn^{-1/5}, \tau^0 - M(j-1)n^{-1/5}] \text{ for } j = 1, \dots, 1/\zeta$$

contains a knot of $\widehat{\varphi}_n$ and of $\widehat{\varphi}_n^0$ by taking ξ_n to be $\tau^0 - Mjn^{-1/5}$. There are $1/\zeta$ such intervals so the probability that all intervals contain a knot of both $\widehat{\varphi}_n$ and $\widehat{\varphi}_n^0$ is $1-\varepsilon$. Now, let $K = O_p(1)$ be such that $\int_{X_{(1)}}^{\tau^0} \left(\widehat{\varphi}_n^0 - \widehat{\varphi}_n\right)^2 e^{\tilde{\varepsilon}_n^2} \widehat{f}_n d\lambda \le Kn^{-1}$ for $\tau^0 < m$, by Lemma A.3. In particular,

$$\int_{I_{j^*}} \left(\widehat{\varphi}_n^0 - \widehat{\varphi}_n\right)^2 e^{\tilde{\varepsilon}_n^2} \widehat{f}_n d\lambda := \min_{j=1,\dots,1/\zeta} \int_{I_j} \left(\widehat{\varphi}_n^0 - \widehat{\varphi}_n\right)^2 e^{\tilde{\varepsilon}_n^2} \widehat{f}_n d\lambda \le \zeta K n^{-1}. \tag{A.28}$$

We next conclude by Lemma C.2, since $\zeta = \delta/L$, that there exists a subinterval $J^* \subset I_{j^*}$ containing knots $\eta \in S(\widehat{\varphi}_n)$ and $\eta^0 \in S(\widehat{\varphi}_n^0)$, such that

$$\sup_{x \in J^*} |\widehat{\varphi}_n(x) - \widehat{\varphi}_n^0(x)| \le c\delta_1 n^{-2/5} \quad \text{and} \quad |\eta^0 - \eta| \le c\delta_1 n^{-1/5} \tag{A.29}$$

for a universal constant $c > 0$ and where $\delta_1 \to 0$ as $\delta \to 0$.

We can now re-apply the proof of Lemma A.3, this time taking as our knots η and η^0, and again assuming without loss of generality $\eta \le \eta^0$. We again see that (A.23) is bounded above by (A.24), and the middle term of (A.24) is bounded by (A.25). Using (A.29), we can conclude by (A.26) that (A.25) is bounded by $\delta_2 O_p(n^{-1})$, so (A.24) is also, and so (A.23) is also, where $\delta_2 \to 0$ as $\delta \to 0$. We can conclude

$$\int_{X_{(1)}}^{\eta^0} \left(\widehat{\varphi}_n^0 - \widehat{\varphi}_n\right)^2 e^{\tilde{\varepsilon}_n^2} \widehat{f}_n d\lambda \le \tilde{\delta} n^{-1}.$$

Now $\eta^0 \ge \tau^0 - Mn^{-1/5}/\zeta$, the endpoint of $I_{1/\zeta,n}$. Thus, take $bn^{-1/5} \ge \tau^0 - Mn^{-1/5}/\zeta$, let $t_n = m - bn^{-1/5}$ and now let $J^* = [t_n - \tilde{L}n^{-1/5}, t_n]$ where $\tilde{L} = \max(L, 8D/\varphi_0^{(2)}(m))$, chosen so that we can apply Lemma C.2. Then

$$\int_{X_{(1)}}^{t_n} \left(\widehat{\varphi}_n^0 - \widehat{\varphi}_n\right)^2 e^{\tilde{\varepsilon}_n^2} \widehat{f}_n d\lambda \le \tilde{\delta} n^{-1} \tag{A.30}$$

with high probability. Analogously,

$$\int_{X_{(1)}}^{t_n} \left(\widehat{\varphi}_n^0 - \widehat{\varphi}_n\right)^2 e^{\tilde{\varepsilon}_n^1} \widehat{f}_n^0 d\lambda \le \tilde{\delta} n^{-1} \tag{A.31}$$

as $n \to \infty$ with high probability. And we can apply Lemma C.2 to the interval J^* to see

$$\|\widehat{\varphi}_n - \widehat{\varphi}_n^0\|_{J^*} = \delta O_p(n^{-2/5}), \qquad \|\widehat{f}_n - \widehat{f}_n^0\|_{J^*} \le \delta K n^{-2/5}, \tag{A.32}$$

$$\|\widehat{F}_n - \widehat{F}_n^0\|_{J^*} \le \delta K n^{-3/5}, \quad \text{and} \qquad |\tau - \tau^0| \le \delta K n^{-1/5}, \tag{A.33}$$

where $\tau \in S(\widehat{\varphi}_n) \cap J^*$ and $\tau^0 \in S(\widehat{\varphi}_n^0) \cap J^*$, and

$$\|(\widehat{\varphi}_n - \widehat{\varphi}_n^0)'\|_{[\max(\tau,\tau^0)+\delta O_p(n^{-1/5}), t_n - \delta O_p(n^{-1/5})]} = \delta K n^{-1/5}; \tag{A.34}$$

here, $K = O_p(1)$ and depends on ε and $\tilde{L} \equiv \tilde{L}_\varepsilon$, but not on δ or t_n. Thus when we eventually let $\tilde{\delta} \to 0$, so $b \equiv b_{\tilde{\delta}} \to \infty$, we can still conclude $\tilde{\delta} K \to 0$. We also continue to assume, without loss of generality, that

$$\tau \le \tau^0.$$

Thus, here is the sense in which we mean o_p, for the remainder of the proof: if we say, e.g., $E^1_{n,t_n} - E^2_{n,t_n} = o_p(n^{-1})$ we mean for any $\tilde{\delta} > 0$, we may set $t_n = m - bn^{-1/5}$ and choose b large enough that $|E^1_{n,t_n} - E^2_{n,t_n}| \le \tilde{\delta} K n^{-1}$ where K does not depend on t_n.

We can now conclude that

$$\tilde{T}^i_{n,t_n} = o_p(n^{-1}) \quad \text{for} \quad i = 1, 2, \tag{A.35}$$

The difference in the definitions of T_n^i (defined in (A.10)) and $\tilde{T}_n^i$ (defined in (A.15)), for $i = 1, 2$, is only in the $e^{\varepsilon_n^i}$'s and $e^{\tilde{\varepsilon}_n^i}$'s. These arise from Taylor expansions of the exponential function. The definition of T_n^i arises from the expansions of $R^c_{n,1,t}$ (see (A.3) and (A.4)). Thus, if we let $e^x = 1+x+2^{-1}x^2e^{\varepsilon(x)}$ we can see that $\varepsilon^1_n(x) = \varepsilon(\widehat{\varphi}_n(x) - \widehat{\varphi}^0_n(x))$ and $\varepsilon^2_n(x) = \varepsilon(\widehat{\varphi}^0_n(x) - \widehat{\varphi}_n(x))$. Let $e^x = 1 + xe^{\tilde{\varepsilon}(x)}$. Then we can see that $\tilde{\varepsilon}^1_n(x) = \tilde{\varepsilon}(\widehat{\varphi}_n(x) - \widehat{\varphi}^0_n(x))$ and $\tilde{\varepsilon}^2_n(x) = \tilde{\varepsilon}(\widehat{\varphi}^0_n(x) - \widehat{\varphi}_n(x))$. Now, by Lemma C.5, for all $x \in \mathbb{R}$, $e^{\varepsilon(x)} \le 2e^{\tilde{\varepsilon}(x)}$, so that

$$0 \le T^i_{n,t_n} \le 2\tilde{T}^i_{n,t_n} = o_p(n^{-1}), \text{ for } i = 1, 2, \tag{A.36}$$

by (A.35).

A.2.4. Return to $R^c_{n,1,t}$

We take t_n and J^* as defined at the end of the previous section. Now, if we could show that $E^1_{n,t_n} - E^2_{n,t_n} = o_p(n^{-1})$ then from (A.12) we could conclude that A^i_{n,t_n}, $i = 1, 2$, are both $o_p(n^{-1})$. If, in addition, we can show $E^1_{n,t_n} + E^2_{n,t_n} = o_p(n^{-1})$, then since

$$R^c_{n,1,t_n} = (E^1_{n,t_n} + E^2_{n,t_n} + A^1_{n,t_n} + A^2_{n,t_n} + T^2_{n,t_n} - T^1_{n,t_n})/2 \tag{A.37}$$

by (A.11), we could conclude $R^c_{n,1,t_n} = o_p(n^{-1})$. Unfortunately it is difficult to get any results about $E^1_{n,t_n} - E^2_{n,t_n}$. We can analyze $E^1_{n,t_n} + E^2_{n,t_n}$, though. The next lemma shows that the difficult terms in $E^1_{n,t_n} + E^2_{n,t_n}$ are $o_p(n^{-1})$.

Lemma A.4. *Let all terms be as defined above. For any $t < m$ let $F^1_{n,t} = \int_{(\tau,t]} \widehat{\varphi}_{n,t} d(\widehat{F}_n - \mathbb{F}_n)$ and $F^2_{n,t} = \int_{(\tau^0,t]} \widehat{\varphi}^0_{n,t} d(\mathbb{F}_n - \widehat{F}^0_n)$. Then*

$$F^1_{n,t_n} + F^2_{n,t_n} = o_p(n^{-1}).$$

Proof. For the proof, denote $t \equiv t_n$ and recall that we assume $\tau \le \tau^0$. We see

$$\begin{aligned}
&\int_{(\tau,t]} \widehat{\varphi}_{n,t} d(\widehat{F}_n - \mathbb{F}_n) + \int_{(\tau^0,t]} \widehat{\varphi}^0_{n,t} d(\mathbb{F}_n - \widehat{F}^0_n) \\
&\quad = \int_{(\tau^0,t]} \widehat{\varphi}_{n,t} d(\widehat{F}_n - \mathbb{F}_n) + \int_{(\tau,\tau^0]} \widehat{\varphi}_{n,t} d(\widehat{F}_n - \mathbb{F}_n) \\
&\qquad - \left(\int_{(\tau^0,t]} \widehat{\varphi}_{n,t} d(\widehat{F}^0_n - \mathbb{F}_n) + \int_{(\tau^0,t]} (\widehat{\varphi}^0_{n,t} - \widehat{\varphi}_{n,t}) d(\widehat{F}^0_n - \mathbb{F}_n) \right)
\end{aligned}$$

which equals

$$\int_{(\tau^0,t]} \widehat{\varphi}_{n,t}(\widehat{f}_n - \widehat{f}^0_n) d\lambda - \int_{(\tau^0,t]} (\widehat{\varphi}^0_{n,t} - \widehat{\varphi}_{n,t}) d(\widehat{F}_n - \mathbb{F}_n) + \int_{(\tau,\tau^0]} \widehat{\varphi}_{n,t} d(\widehat{F}_n - \mathbb{F}_n). \tag{A.38}$$

Note $\|\widehat{\varphi}^0_{n,t}\|_{J^*} = O_p(n^{-2/5})$. This follows because $n^{1/5}(\tau^0 - t) = O_p(1)$ by Proposition 7.3 of Doss and Wellner [2018b], and because $\|(\widehat{\varphi}^0_n)'\|_{J^*} = n^{-1/5}O_p(1)$ by Corollary 7.1 of Doss and Wellner [2018b], since $\varphi_0'(m) = 0$. In both cases the

$O_p(1)$ does not depend on t_n. Thus, the first term in (A.38) is $o_p(n^{-1})$, since $\|\widehat{f}_n - \widehat{f}_n^0\|_{J^*} = o_p(n^{-2/5})$. We will rewrite the other two terms of (A.38) with integration by parts. The negative of the middle term, $\int_{(\tau^0,t]} (\widehat{\varphi}_{n,t}^0 - \widehat{\varphi}_{n,t}) d(\widehat{F}_n - \mathbb{F}_n)$, equals

$$((\widehat{F}_n - \mathbb{F}_n)(\widehat{\varphi}_{n,t}^0 - \widehat{\varphi}_{n,t}))(\tau^0, t] - \int_{(\tau^0,t]} (\widehat{F}_n - \mathbb{F}_n)((\widehat{\varphi}_n^0)' - \widehat{\varphi}_n') d\lambda. \tag{A.39}$$

Note $\|\widehat{F}_n - \mathbb{F}_n\|_{J^*} = O_p(n^{-3/5})$ by Lemma C.4. Thus the first term in (A.39) is $o_p(n^{-1})$ because $\|\widehat{\varphi}_n^0 - \widehat{\varphi}_n\|_{J^*} = o_p(n^{-2/5})$. The second term in (A.39) is $o_p(n^{-1})$ because (A.34) implies that $\int_{[\tau^0,t]} |(\widehat{\varphi}_n - \widehat{\varphi}_n)'| d\lambda = o_p(n^{-2/5})$, and, as already noted, $\|\widehat{F}_n - \mathbb{F}_n\|_{J^*} = O_p(n^{-3/5})$. Thus (A.39) is $o_p(n^{-1})$.

We have left the final term of (A.38). This can be bounded by

$$\left| (\widehat{F}_n - \mathbb{F}_n)(\widehat{\varphi}_{n,t})(\tau, \tau^0] - \int_{(\tau,\tau^0]} (\widehat{F}_n - \mathbb{F}_n) \widehat{\varphi}_n' d\lambda \right|, \tag{A.40}$$

and the second term above bounded by $\|\widehat{F}_n - \mathbb{F}_n\|_{[\tau,\tau^0]} \widehat{\varphi}_n'(\tau+)(\tau^0 - \tau) = o_p(n^{-1})$ because $\|\widehat{F}_n - \mathbb{F}_n\|_{J^*} = O_p(n^{-3/5})$, $\widehat{\varphi}_n'(\tau+) = O_p(n^{-1/5})$ (since $\varphi_0'(m) = 0$, and recall $\widehat{\varphi}_n$ is linear on $[\tau, \tau^0]$), and $(\tau^0 - \tau) = o_p(n^{-1/5})$. The first term of (A.40) is $o_p(n^{-1})$ because in fact $\|\widehat{F}_n - \mathbb{F}_n\|_{[\tau,\tau^0]} = o_p(n^{-3/5})$, by Lemma C.4 (since $|\tau - \tau^0| = o_p(n^{-1/5})$), and $\|\widehat{\varphi}_{n,t}\|_{[\tau,\tau^0]} = O_p(n^{-2/5})$. Thus (A.38) is $o_p(n^{-1})$ so we are done. $\square$

For $t < m$, define

$$G_{n,t}^1 = \widehat{\varphi}_n(t)(\widehat{F}_n(t) - \widehat{F}_n^0(t)) + (\widehat{\varphi}_n(\tau) - \widehat{\varphi}_n(t))(\widehat{F}_n(\tau) - \mathbb{F}_n(\tau))$$

and

$$G_{n,t}^2 = \widehat{\varphi}_n^0(t)(\widehat{F}_n(t) - \widehat{F}_n^0(t)) + (\widehat{\varphi}_n^0(\tau^0) - \widehat{\varphi}_n^0(t))(\mathbb{F}_n(\tau^0) - \widehat{F}_n^0(\tau^0)),$$

so that $G_{n,t_n}^i = E_{n,t_n}^i - F_{n,t_n}^i$ for $i = 1, 2$ (recalling the definitions of E_{n,t_n}^i in (A.6) and (A.7)). The key idea now is that the first term in $G_{n,t}^i$ matches up with $R_{n,1,t_1,t_2}$. To make this explicit, we need to define a one-sided version of $R_{n,1,t_1,t_2}$. Since both $\widehat{f}_n$ and $\widehat{f}_n^0$ integrate to 1, note for any $t_1 \le m \le t_2$, that

$$R_{n,1,t_1,t_2} = -\varphi_0(m) \int_{D_{n,t_1,t_2}^c} (\widehat{f}_n - \widehat{f}_n^0) d\lambda;$$

thus, define

$$R_{n,1,t_1} = -\varphi_0(m)(\widehat{F}_n(t_1) - \widehat{F}_n^0(t_1)). \tag{A.41}$$

The corresponding definition for the right side is $-\varphi_0(m) \int_{t_2}^{X_{(n)}} (\widehat{f}_n - \widehat{f}_n^0) d\lambda$, which when summed with (A.41) yields $R_{n,1,t_1,t_2}$.

Lemma A.5. *Let all terms be as defined above. We then have for $i = 1, 2$,*

$$G^i_{n,t_n} + R_{n,1,t_n} = o_p(n^{-1}).$$

Proof. The second terms in the definitions of G^i_{n,t_n}, $i = 1, 2$, are both $O_p(n^{-7/5})$ since $\widehat{\varphi}_n(\tau) - \widehat{\varphi}_n(t)$ and $\widehat{\varphi}^0_n(\tau^0) - \widehat{\varphi}^0_n(t)$ are both $O_p(n^{-2/5})$ by Lemma 4.5 of Balabdaoui, Rufibach and Wellner [2009] and Corollary 5.4 of Doss and Wellner [2018b], and the terms $\widehat{F}_n(\tau) - \mathbb{F}_n(\tau)$ and $\widehat{F}^0_n(\tau^0) - \mathbb{F}_n(\tau^0)$ are both $O_p(n^{-1})$ by Corollary 2.4 and Corollary 2.12 of Doss and Wellner [2018b].

Thus we consider the first terms of G^i_{n,t_n}, in sum with $R_{n,1,t_n}$. Consider the case $i = 1$; we see that

$$\widehat{\varphi}_n(t_n)(\widehat{F}_n(t_n) - \widehat{F}^0_n(t_n)) + R_{n,1,t_n} = (\widehat{\varphi}_n(t_n) - \varphi_0(m))(\widehat{F}_n(t_n) - \widehat{F}^0_n(t_n)), \quad (A.42)$$

and $(\widehat{\varphi}_n(t_n) - \varphi_0(m)) = O_p(n^{-2/5})$ by Lemma 4.5 of Balabdaoui, Rufibach and Wellner [2009] since $\varphi_0'(m) = 0$. Crucially, we are not making a claim that $\widehat{\varphi}_n$ is close to $\varphi_0(m)$ uniformly over an interval, just a claim at the point t_n satisfying $t_n \to m$, so the O_p statement does not depend on b. Since $\widehat{F}_n(t_n) - \widehat{F}^0_n(t_n) = o_p(n^{-3/5})$ by (A.33), we conclude that (A.42) is $o_p(n^{-1})$. Identical reasoning applies to the case $i = 2$, using Corollary 5.4 of Doss and Wellner [2018b]. Thus we are done. □

Thus by Lemmas A.4 and A.5,

$$R_{n,1,t_n} + (E^1_{n,t_n} + E^2_{n,t_n})/2 = o_p(n^{-1}). \quad (A.43)$$

Now we decompose the A^i_{n,t_n} terms. Let

$$B^1_{n,t_n} = \int \widehat{\varphi}_{n,\tau}\, d(\mathbb{F}_n - \widehat{F}^0_n),$$
$$B^2_{n,t_n} = \int \widehat{\varphi}^0_{n,\tau^0}\, d(\widehat{F}_n - \mathbb{F}_n).$$

where τ and τ^0 are as previously defined (on page 12) and

$$C^1_{n,t_n} = \int \widehat{\varphi}_n'(t_n-)(x - t_n)_-\, d(\mathbb{F}_n - \widehat{F}^0_n)$$
$$C^2_{n,t_n} = \int (\widehat{\varphi}^0_n)'(t_n-)(x - t_n)_-\, d(\widehat{F}_n - \mathbb{F}_n).$$

Then, for $i = 1, 2$, by the definitions of τ and τ^0,

$$A^i_n = B^i_n + C^i_n,$$

and note that

$$B^1_{n,t_n} = \int (\widehat{\varphi}_{n,\tau} - \widehat{\varphi}^0_{n,\tau^0})\, d(\mathbb{F}_n - \widehat{F}^0_n) \quad \text{and} \quad B^2_{n,t_n} = \int (\widehat{\varphi}^0_{n,\tau^0} - \widehat{\varphi}_{n,\tau})\, d(\widehat{F}_n - \mathbb{F}_n), \quad (A.44)$$

by the characterization theorems, Theorem 2.2 of Doss and Wellner [2018b] with $\Delta = \pm\widehat{\varphi}^0_{n,\tau^0}$ and Theorem 2.8 of Doss and Wellner [2018b] with $\Delta = \pm\widehat{\varphi}_{n,\tau}$. Perhaps strangely, it seems it is easier to analyze $B_n^1 - B_n^2$ than $B_n^1 + B_n^2$, and $C_n^1 + C_n^2$ rather than $C_n^1 - C_n^2$. Perhaps more strangely, this will suffice. Again by Theorem 2.2 of Doss and Wellner [2018b] with $\Delta = \widehat{\varphi}_{n,\tau}$ and Theorem 2.8 of Doss and Wellner [2018b] with $\Delta = \widehat{\varphi}^0_{n,\tau^0}$, $B^1_{n,t_n} \le 0$ and $B^2_{n,t_n} \ge 0$, so

$$B^1_{n,t_n} - B^2_{n,t_n} \le \begin{cases} -B^2_{n,t_n} \le 0 \\ B^1_{n,t_n} \le 0. \end{cases} \tag{A.45}$$

Thus, if we can show $B^1_{n,t_n} - B^2_{n,t_n} = o_p(n^{-1})$ then $B^i_{n,t_n} = o_p(n^{-1})$, $i = 1, 2$, so $B^1_{n,t_n} + B^2_{n,t_n} = o_p(n^{-1})$. We do this in the following lemma.

Lemma A.6. *With all terms as defined above,*

$$B^1_{n,t_n} - B^2_{n,t_n} = o_p(n^{-1}),$$

and thus

$$B^1_{n,t_n} + B^2_{n,t_n} = o_p(n^{-1}).$$

Proof. Now by (A.44)

$$B^1_{n,t_n} - B^2_{n,t_n} = \int (\widehat{\varphi}_{n,\tau} - \widehat{\varphi}^0_{n,\tau}) d(\widehat{F}_n - \widehat{F}^0_n) + \int (\widehat{\varphi}^0_{n,\tau} - \widehat{\varphi}^0_{n,\tau^0}) d(\widehat{F}_n - \widehat{F}^0_n)$$

which equals

$$\int (\widehat{\varphi}_{n,\tau} - \widehat{\varphi}^0_{n,\tau}) d(\widehat{F}_n - \widehat{F}^0_n) + (\widehat{\varphi}^0_n(\tau^0) - \widehat{\varphi}^0_n(\tau))(\widehat{F}_n - \widehat{F}^0_n)(\tau) + \int_\tau^{\tau^0} \widehat{\varphi}^0_{n,\tau^0} d(\widehat{F}_n - \widehat{F}^0_n). \tag{A.46}$$

The first term in (A.46) equals, applying (C.1),

$$\int_{X_{(1)}}^{\tau} (\widehat{\varphi}_n - \widehat{\varphi}^0_n)^2 e^{\tilde{\varepsilon}^1_n} \widehat{f}^0_n d\lambda - ((\widehat{\varphi}_n - \widehat{\varphi}^0_n)(\widehat{F}_n - \widehat{F}^0_n))(\tau), \tag{A.47}$$

where $\tilde{\varepsilon}^1_n$ is identical to $\tilde{\varepsilon}^1_n$ in Lemma A.3, and thus by (A.30) the first term in (A.47) is $o_p(n^{-1})$. The second term is also $o_p(n^{-1})$ since $\|\widehat{\varphi}_n - \widehat{\varphi}^0_n\|_{J^*} = o_p(n^{-2/5})$ and $\|\widehat{F}_n - \widehat{F}^0_n\|_{J^*} = o_p(n^{-3/5})$. This also shows that the middle terms in (A.46) is $o_p(n^{-1})$. To see the last term is $o_p(n^{-1})$, recall $\|\widehat{\varphi}^0_{n,\tau^0}\|_{J^*} = O_p(n^{-2/5})$ by Corollary 5.4 of Doss and Wellner [2018b], using that $\varphi_0'(m) = 0$. Since $|\tau^0 - \tau| = o_p(n^{-1/5})$ and $\|\widehat{f}_n - \widehat{f}^0_n\|_{J^*} = o_p(n^{-2/5})$ we see the last term of (A.46) is $o_p(n^{-1})$, so (A.46) is $o_p(n^{-1})$, so $B^1_{n,t_n} - B^2_{n,t_n} = o_p(n^{-1})$. By (A.45), $B^1_{n,t_n} + B^2_{n,t_n} = o_p(n^{-1})$, so we are done. □

We now turn our attention to $C^1_{n,t_n} + C^2_{n,t_n}$.

Lemma A.7. *With all terms as defined above,*

$$C^1_{n,t_n} + C^2_{n,t_n} = o_p(n^{-1}).$$

Proof. Note that $C^1_{n,t_n} + C^2_{n,t_n}$ equals

$$\int (\widehat{\varphi}'_n(t_n-) - (\widehat{\varphi}^0_n)'(t_n-))(x-t_n)_- \, d(\mathbb{F}_n - \widehat{F}^0_n) \\ + \int (\widehat{\varphi}^0_n)'(t_n-)(x-t_n)_- \, d(\mathbb{F}_n - \widehat{F}^0_n) + \int (\widehat{\varphi}^0_n)'(t_n-)(x-t_n)_- \, d(\widehat{F}_n - \mathbb{F}_n)$$

which equals

$$\int (\widehat{\varphi}'_n(t_n-) - (\widehat{\varphi}^0_n)'(t_n-))(x-t_n)_- \, d(\mathbb{F}_n - \widehat{F}^0_n) + \int (\widehat{\varphi}^0_n)'(t_n-)(x-t_n)_- \, d(\widehat{F}_n - \widehat{F}^0_n). \tag{A.48}$$

Since $(\widehat{F}_n - \widehat{F}^0_n)(X_{(1)}) = 0$, the second term in (A.48) equals

$$-(\widehat{\varphi}^0_n)'(t_n-) \int_{-\infty}^{t_n} (\widehat{F}_n - \widehat{F}^0_n) d\lambda = -(\widehat{\varphi}^0_n)'(t_n-) \int_{\upsilon}^{t_n} (\widehat{F}_n - \widehat{F}^0_n) d\lambda \tag{A.49}$$

for a point $\upsilon \in [\tau, \tau^0]$ which exists by the proof of Proposition 2.13 of Doss and Wellner [2018b]. By (A.33), since $t_n - \upsilon = O_p(n^{-1/5})$,

$$\int_{\upsilon}^{t_n} (\widehat{F}_n - \widehat{F}^0_n) d\lambda = o_p(n^{-4/5}). \tag{A.50}$$

Since $t_n \to m$, by Corollary 5.4 of Doss and Wellner [2018b], $(\widehat{\varphi}^0_n)'(t_n-) = O_p(1) n^{-1/5}$. As in previous cases, by taking $t_n = \xi_n$ and $C = 0$ in that corollary, the $O_p(1)$ does not depend on t_n. Thus we have shown (A.49) is $o_p(n^{-1})$.

Since $\mathbb{F}_n(-\infty) - \widehat{F}^0_n(-\infty) = 0$, the first term in (A.48) equals

$$-(\widehat{\varphi}_n - \widehat{\varphi}^0_n)'(t_n-) \int_{X_{(1)}}^{t_n} (\mathbb{F}_n - \widehat{F}^0_n) d\lambda,$$

which equals

$$-(\widehat{\varphi}_n - \widehat{\varphi}^0_n)'(t_n-) \int_{\tau^0}^{t_n} (\mathbb{F}_n - \widehat{F}^0_n) d\lambda \tag{A.51}$$

because $\widehat{H}^0_{n,L}(\tau^0) = \mathbb{Y}_{n,L}(\tau^0)$ by Theorem 2.10 of Doss and Wellner [2018b]. The absolute value of (A.51) is bounded above by

$$|\widehat{\varphi}'_n(t_n-) - (\widehat{\varphi}^0_n)'(t_n-)|(t_n - \tau^0) \sup_{u \in [\tau^0, t_n]} |\mathbb{F}_n(u) - \widehat{F}^0_n(u)|. \tag{A.52}$$

We know $|\widehat{\varphi}'_n(t_n-) - (\widehat{\varphi}^0_n)'(t_n-)| = O_p(n^{-1/5})$ but unfortunately it is not necessarily $o_p(n^{-1/5})$. However,

$$|\widehat{\varphi}'_n(t_n-) - (\widehat{\varphi}^0_n)'(t_n-)|(t_n - \tau^0) = |(\widehat{\varphi}_n - \widehat{\varphi}^0_n)(t_n) - (\widehat{\varphi}_n - \widehat{\varphi}^0_n)(\tau^0)| = o_p(n^{-2/5}),$$

so (A.52) is $o_p(n^{-1})$, and thus so also is (A.48); that is, $C^1_{n,t_n} + C^2_{n,t_n} = o_p(n^{-1})$. □

Thus we have shown that $C^1_{n,t_n} + C^2_{n,t_n} = o_p(n^{-1})$ and $B^1_{n,t_n} + B^2_{n,t_n} = o_p(n^{-1})$, so we can conclude

$$A^1_{n,t_n} + A^2_{n,t_n} = o_p(n^{-1}).$$

Together with (A.43), (A.36), and (A.37), we can conclude that

$$R_{n,t_n} + R^c_{n,1,t_n} = o_p(n^{-1}). \tag{A.53}$$

A.3. Proof completion / details: the main result

The preceding one-sided arguments apply symmetrically to the error terms on the right side of m. Thus, we now return to handling simultaneously the two-sided error terms. We have thus shown for any $\delta > 0$ we can find a $b \equiv b_\delta$, such that, letting $t_{n,1} = m - bn^{-1/5}$, $t_{n,2} = m + bn^{-1/5}$, we have

$$|R_{n,1,t_{n,1},t_{n,2}} + R^c_{n,1,t_{n,1},t_{n,2}}| \le \delta K n^{-1} \tag{A.54}$$

where $K = O_p(1)$ does not depend on b (i.e., on δ). Now by Proposition A.1

$$|R_{n,2,t_{n,1},t_{n,2}}| + |R^0_{n,2,t_{n,1},t_{n,2}}| + |R_{n,3,t_{n,1},t_{n,2}}| + |R^0_{n,3,t_{n,1},t_{n,2}}| = o_p(n^{-1}).$$

Let

$$\begin{aligned} R_{n,t_{n,1},t_{n,2}} \equiv 2n(&R_{n,1,t_{n,1},t_{n,2}} + R^c_{n,1,t_{n,1},t_{n,2}} + R_{n,2,t_{n,1},t_{n,2}} \\ &- R^0_{n,2,t_{n,1},t_{n,2}} + R_{n,3,t_{n,1},t_{n,2}} - R^0_{n,3,t_{n,1},t_{n,2}}). \end{aligned}$$

Then by (4.9), write $2\log\lambda_n = \mathbb{D}_{n,t_{n,1},t_{n,2}} + R_{n,t_{n,1},t_{n,2}}$ (slightly modifying the form of the subscripts). Now fixing any subsequence of $\{n\}_{n=1}^\infty$, we can find a subsubsequence such that $R_{n,t_{n,1},t_{n,2}} \to_d \delta R$ for a tight random variable R by (A.54). For $b > 0$ let $\mathbb{D}_b \equiv \int_{-b/\gamma_2}^{b/\gamma_2} (\widehat{\varphi}^2(u) - (\widehat{\varphi}^0)^2(u))du$, as in (4.11), which lets us conclude that

$$2\log\lambda_n = \mathbb{D}_{n,t_{n,1},t_{n,2}} + R_{n,t_{n,1},t_{n,2}} \to_d \mathbb{D}_b + \delta R \tag{A.55}$$

along the subsubsequence. Taking say $\delta = 1$ shows that there exists a (tight) limit random variable, which we denote by $\mathbb{D}$. Then, since R does not depend on δ, we can let $\delta \searrow 0$ so $b_\delta \equiv b \nearrow \infty$, and see that $\lim_{b\to\infty} \mathbb{D}_b = \mathbb{D}$, which can now be seen to be pivotal. Thus along this subsubsequence, $2\log\lambda_n \to_d \mathbb{D}$. This was true for an arbitrary subsubsequence, and so the convergence in distribution holds along the original sequence. Thus,

$$2\log\lambda_n \to_d \mathbb{D} \quad \text{as} \quad n \to \infty.$$

A.4. Proofs for global consistency

Proof of Theorem 2.1 Part B. We now indicate the changes to the arguments of Dümbgen and Rufibach [2009] which are needed to prove an analog of Theorem 4.1 of Dümbgen and Rufibach [2009] for $\widehat{\varphi}_n^0$. Note that our Theorem 2.1 part B is only a partial analogue of Theorem 2.1 part A since we only consider the case $\beta = 2$ and require m to be unique. We assume $f_0 \in \mathcal{P}_m := \{e^{\varphi} : \int e^{\varphi(x)} dx = 1, \varphi \in \mathcal{C}_m\}$ where for m fixed $\mathcal{C}_m$ is the class of concave, closed, proper functions with m as a maximum. We need to study the allowed 'caricatures' of the Lemmas A.4 and A.5 of Dümbgen and Rufibach [2009], which differ for $\widehat{\varphi}_n^0$ from those for $\widehat{\varphi}_n$. Let $\rho_n \equiv n^{-1} \log n$. Note, we define here a function Δ to be "piecewise linear (with q knots)" to mean that

$$\begin{aligned}&\mathbb{R} \text{ may be partitioned into } q+1 \text{ non-degenerate intervals} \\ &\text{on each of which } \Delta \text{ is linear.}\end{aligned} \tag{A.56}$$

In particular, Δ may be discontinuous. Let $\mathcal{D}_k$ be the family of piecewise linear functions on $\mathbb{R}$ with at most k knots. Let

$$\widehat{M} := \overline{\{x \in \mathbb{R} : (\widehat{\varphi}_n^0)'(x) = 0\}} \qquad \text{and} \qquad \widehat{N} := [\tau_L, \tau_R], \tag{A.57}$$

where τ_L is the greatest knot of $\widehat{\varphi}_n^0$ strictly smaller than m and τ_R is the smallest knot of $\widehat{\varphi}_n^0$ strictly larger than m. Note that $\widehat{M}$, the (closed) modal interval of $\widehat{\varphi}_n^0$, is contained in $\widehat{N}$, and may or may not be strictly contained in $\widehat{N}$. Let $\mathcal{S}_n(\widehat{\varphi}_n^0)$ denote the set of knots of $\widehat{\varphi}_n^0$. For a function f, let $f(x+) = \lim_{t \searrow x} f(t)$ and $f(x-) = \lim_{t \nearrow x} f(t)$ when these limits exist.

Lemma A.8. *Let $\widehat{M}$ be as in (A.57). Let $\Delta : \mathbb{R} \to \mathbb{R}$ be piecewise linear in the sense of (A.56), such that*

$$\Delta 1_{\widehat{M}} + (-\infty) \times 1_{\widehat{M}^c} \text{ is concave with mode at } m, \tag{A.58}$$

and assume for each knot q of Δ that one of the following holds:

$$q \in \mathcal{S}_n(\widehat{\varphi}_n^0) \setminus \{m\} \qquad \text{and} \qquad \Delta(q) = \liminf_{x \to q} \Delta(x) \tag{A.59}$$

$$\Delta(q) = \lim_{r \to q} \Delta(r) \qquad \text{and} \qquad \Delta'(q-) \geq \Delta'(q+) \tag{A.60}$$

$$q = m, \Delta(q) = \lim_{r \in \widehat{M}, r \to q} \Delta(r) \qquad \text{and} \qquad \Delta(q) = \liminf_{x \to q} \Delta(x). \tag{A.61}$$

Then

$$\int \Delta d\mathbb{F}_n \leq \int \Delta d\widehat{F}_n^0. \tag{A.62}$$

Note that if m is not a knot of $\widehat{\varphi}_n^0$, so is interior to $\widehat{M}$, then Δ must be continuous at m and (A.58) implies that m is a local mode of Δ. If $m \in S_n(\widehat{\varphi}_n^0)$ such that $(\widehat{\varphi}_n^0)'(m+) < 0$, then (A.61) allows Δ to be discontinuous at m but forces $\Delta(m-) \leq \Delta(m+)$.

Proof. We show we can construct a sequence Δ_k converging pointwise to Δ, with $|\Delta_k| \le |\Delta|$, and such that $\widehat{\varphi}_n^0 + \varepsilon\Delta_k$ is concave with mode at m for small enough ε. We first show this holds on the interval $\widehat{N}$.

If m is not a knot of $\widehat{\varphi}_n^0$ then by (A.58) $\widehat{\varphi}_n^0 + \varepsilon\Delta$ is concave with mode at m on $\widehat{N}$. Now if m is a knot of $\widehat{\varphi}_n^0$, either (A.60) holds or (A.61) holds. In the former case, again for $\varepsilon > 0$ small enough, $\widehat{\varphi}_n^0 + \varepsilon\Delta$ is concave with mode at m on $\widehat{N}$.

Thus assume (A.61) holds. For concreteness, assume $m \in S_n(\widehat{\varphi}_n^0)$ is such that $(\widehat{\varphi}_n^0)'(m-) < 0 = (\widehat{\varphi}_n^0)'(m+)$. For $x \in [m-1/k, m]$ define $\Delta_k(x)$ to be the linear function connecting $\Delta(m)$ to $\Delta(m-1/k)$ for $k = 1, \ldots,$ and let $\Delta_k(x) = \Delta(x)$ for $x \in \widehat{N} \setminus [m-1/k, m]$. Then for k large,

$$\widehat{\varphi}_n^0 + \varepsilon\Delta_k \text{ is concave with mode } m \tag{A.63}$$

on $\widehat{N}$, and $\Delta_k(x)$ is monotonically increasing to $\Delta(x)$ (again, for $k \ge$ some K), by (A.61).

For knots q of Δ with $q \ne m$, similar arguments can be made; one can define $\Delta_k(x)$ such that $|\Delta_k(x)| \le |\Delta(x)|$ where the knots q_k of Δ_k are either knots of $\widehat{\varphi}_n^0$ or satisfy $\Delta_k'(q_k-) > \Delta_k'(q+)$ so that for $\varepsilon > 0$ small (A.63) holds globally. Thus, by the dominated convergence theorem, and the characterization theorem for $\widehat{F}_n^0$,

$$\int \Delta d\mathbb{F}_n = \lim_{k\to\infty} \int \Delta_k d\mathbb{F}_n \le \lim_k \int \Delta_k d\widehat{F}_n^0 = \int \Delta d\widehat{F}_n^0.$$

□

For the next lemma, we define for a function $\Delta : \mathbb{R} \to \mathbb{R}$,

$$W(\Delta) = \sup_{x\in\mathbb{R}} \frac{|\Delta(x)|}{1 \vee |\varphi_0(x)|} \qquad \text{and} \qquad \sigma^2(\Delta) = \int_{\mathbb{R}} \Delta^2(x) dF_0(x).$$

Also, for a point $x \in \mathbb{R}$, let $\tau_+^0(x) = \min S_n(\widehat{\varphi}_n^0) \cap [x, \infty)$ and $\tau_-^0(x) = \max S_n(\widehat{\varphi}_n^0) \cap (-\infty, x]$.

Lemma A.9. *Let $T = [A, B]$ be a compact subinterval strictly contained in $\{f_0 > 0\}$. Let $\varphi_0 - \widehat{\varphi}_n^0 \ge \varepsilon$ or $\widehat{\varphi}_n^0 - \varphi_0 \ge \varepsilon$ on some interval $[c, c+\delta] \subset T$ with length $\delta > 0$ and suppose $X_{(1)} < c$ and $X_{(n)} > c + \delta$. Suppose $[\tau_-^0(c), \tau_+^0(c+\delta)] \cap \hat{N} = \emptyset$. Then there exists a piecewise linear function Δ with at most three knots, each of which satisfies one of conditions (A.59) or (A.60) and a positive constant $K' = K'(f_0, T)$ such that*

$$|\varphi_0 - \widehat{\varphi}^0| \ge \varepsilon |\Delta|, \tag{A.64}$$

$$\Delta(\varphi_0 - \widehat{\varphi}^0) \ge 0, \tag{A.65}$$

$$\Delta \le 1, \tag{A.66}$$

$$\int_c^{c+\delta} \Delta^2(x) dx \ge \delta/3, \tag{A.67}$$

$$W(\Delta) \le K' \delta^{-1/2} \sigma(\Delta). \tag{A.68}$$

Proof. The proof is identical to the proof of Lemma A.5 in Dümbgen and Rufibach [2009]; the condition $[\tau_-^0(c), \tau_+^0(c+\delta)] \cap \hat{N} = \emptyset$ allows us to use identical perturbations for $\widehat{\varphi}_n^0$ that one can use for $\widehat{\varphi}_n$. $\square$

Next, we need an adaptation of the above lemma for the more difficult case where we have to accomodate the modal constraint. We assume here that the length of $\widehat{N}$ is shorter than δ, which will be true with high probability when we apply the lemma to the case where $\widehat{N}$ is of order $n^{-1/5}$ and δ of order $(\log n/n)^{1/5}$.

Lemma A.10. *Let $T = [A, B]$ be a compact interval strictly contained in $\{f_0 > 0\}$. Let $\varphi_0 - \widehat{\varphi}_n^0 \geq \varepsilon$ or $\widehat{\varphi}_n^0 - \varphi_0 \geq \varepsilon$ on some interval $[c, c+\delta] \subset T$ with length $\delta > 0$ and suppose that $X_{(1)} < c$ and $X_{(n)} > c + \delta$. Suppose also that $[\tau_-^0(c), \tau_+^0(c+\delta)] \cap \hat{N} \neq \emptyset$ and $|\widehat{N}|$, the length of $\widehat{N}$, is no larger than $\delta/4$, and suppose $T \setminus [c, \infty)$ and $T \setminus (-\infty, c+\delta]$ both contain a knot of $\widehat{\varphi}_n^0$. Then there exists a piecewise linear (in the sense of (A.56)) function Δ with at most 4 knots, satisfying the conditions of Lemma A.8, and there exists a positive $K' \equiv K'(f_0, T)$ such that*

$$|\varphi_0 - \widehat{\varphi}_n^0| \geq \varepsilon |\Delta|, \tag{A.69}$$

$$\Delta(\varphi_0 - \widehat{\varphi}_n^0) \geq 0, \tag{A.70}$$

$$\Delta \leq 1 \tag{A.71}$$

$$\int_c^{c+\delta} \Delta^2(x)dx \geq \delta/6, \tag{A.72}$$

$$W(\Delta) \leq K'\delta^{-1/2}\sigma(\Delta). \tag{A.73}$$

Proof. We argue by several different cases. We focus only on the cases where $\widehat{N}$ is near to $[c, c+\delta]$ in the sense that we now assume that either $\widehat{N} \cap [c, c+\delta] \neq \emptyset$ or there are no knots of $\widehat{\varphi}_n^0$ between $\widehat{N}$ and $[c, c+\delta]$. In any other case, the proof Lemma A.5 of Dümbgen and Rufibach [2009] applies without modification.

We begin with the cases where $\widehat{\varphi}_n^0 - \varphi_0 \geq \varepsilon$ on $[c, c+\delta]$. There are separate subcases depending on how $\widehat{N}$ relates to $[c, c+\delta]$ and the (non-)existence of other knots in $[c, c+\delta]$. In all cases, we will first verify conditions (A.69)–(A.72) and put off verifying (A.73) until later.

Case 1.1 Assume $\widehat{\varphi}_n^0 - \varphi_0 \geq \varepsilon$ on $[c, c+\delta]$ and $\widehat{N} \subset [c, c+\delta]$. Let $\Delta \in \mathcal{D}_4$ be continuous (and piecewise linear) with knots at c, τ_L, τ_R, and $c+\delta$, and let Δ be equal to -1 on $\widehat{N}$ and 0 on $[c, c+\delta]^c$. Thus Δ satisfies conditions (A.58) and (A.60) of Lemma A.8. Then $|\Delta| \leq 1$ on $[c, c+\delta]$ and is 0 on $[c, c+\delta]^c$, so (A.69) is satisfied, and so is (A.70). Since Δ is always nonpositive, (A.71) is trivially satisfied. We see that

$$\begin{aligned} \int_c^{c+\delta} \Delta^2(x)dx &\geq \int_0^{x_0} \left(\frac{x}{x_0}\right)^2 dx + \int_{x_0}^{\delta} \left(\frac{\delta - x}{\delta - x_0}\right)^2 dx \\ &= \frac{x_0}{3} + \frac{\delta - x_0}{3} = \frac{\delta}{3}, \end{aligned} \tag{A.74}$$

so (A.72) is satisfied.

The next two cases assume $\widehat{\varphi}_n^0 - \varphi_0 \geq \varepsilon$ on $[c, c+\delta]$ and $\widehat{N} \not\subset [c, c+\delta]$. Recall $\widehat{N} = [\tau_L, \tau_R]$.

Case 1.2 Assume $\widehat{\varphi}_n^0 - \varphi_0 \geq \varepsilon$ on $[c, c+\delta]$ and $\widehat{N} \not\subset [c, c+\delta]$. Additionally, assume there exists $\tau \in S_n(\widehat{\varphi}_n^0) \cap ([c, c+\delta] \setminus \widehat{N})$. We now again let $\Delta \in \mathcal{D}_3 \subset \mathcal{D}_4$ be continuous, now with $\Delta(\tau) = -1$. If $\tau_R < c+\delta$, then set the knots of Δ at $c \vee \tau_R, \tau$, and $c+\delta$, and set Δ to be 0 on $[c \vee \tau_R, c+\delta]^c$. If $\tau_L > c$, then set the knots at c, τ, and $(c+\delta) \wedge \tau_L$ and set Δ to be 0 on $[c, (c+\delta) \wedge \tau_L]^c$. Consider the case where $\tau_R < c+\delta$, and the other case is identical. Since $\widehat{N} \not\subset [c, c+\delta]$ and $|\widehat{N}| \leq \delta/4$, $\tau_R - c < \delta/4$. Again, Δ satisfies conditions (A.58) and (A.60) of Lemma A.8. Conditions (A.69)–(A.71) can be immediately verified, as before. Condition (A.72) can be verified as in the previous case, replacing δ by $3\delta/4$, since $\tau_R - c < \delta/4$, and this yields

$$\int_c^{c+\delta} \Delta^2(x)dx \geq \delta/4. \tag{A.75}$$

Case 1.3 Assume $\widehat{\varphi}_n^0 - \varphi_0 \geq \varepsilon$ on $[c, c+\delta]$ and $\widehat{N} \not\subset [c, c+\delta]$. Additionally, assume that $S_n(\widehat{\varphi}_n^0) \cap ([c, c+\delta] \setminus \widehat{N}) = \emptyset$. We define $\tilde{\Delta}$ to be an affine function either with $\tilde{\Delta}(c) = -\varepsilon$ and $\tilde{\Delta}$ nonincreasing or $\tilde{\Delta}(c+\delta) = -\varepsilon$ and $\tilde{\Delta}$ nondecreasing. Thus, $\tilde{\Delta} \leq -\varepsilon$ on $[c, c+\delta]$. We take $\tilde{\Delta}$ to be tangent to $\varphi_0 - \widehat{\varphi}_n^0$ (but this is not essential). Next let $(c_1, d_1) := \{\tilde{\Delta} < 0\} \cap (c_0, d_0)$ where $[c_0, d_0] \supset ([c, c+\delta] \setminus \widehat{N})$ is defined to be the maximal interval on which $\varphi_0 - \widehat{\varphi}_n^0$ is concave, so $\widehat{\varphi}_n^0$ is linear. Define $\Delta \in \mathcal{D}_2 \subset \mathcal{D}_4$ via

$$\Delta(x) := \begin{cases} 0 & x \notin [c_1, d_1], \\ \tilde{\Delta}(x)/\varepsilon & x \in [c_1, d_1]. \end{cases}$$

Now, (A.58) of Lemma A.8 is seen to be satisfied since Δ is 0 on $\widehat{N}$, and since by definition $\tau_L \neq m \neq \tau_R$, (A.59) is satisfied at c_1 and d_1. Since $\tilde{\Delta}$ is tangent to $\varphi_0 - \widehat{\varphi}_n^0$, condition (A.69) is verified, and (A.70) and (A.71) are also seen to be verified. Condition (A.72) holds easily since in fact $\Delta \leq -1$ on $[c, c+\delta]$.

Case 2 Now assume $\varphi_0 - \widehat{\varphi}_n^0 \geq \varepsilon$ on $[c, c+\delta]$.

Case 2.1 Assume $\widehat{N} \subset (c, c+\delta)$. Then if $c+\delta/2 \leq m \leq c+\delta$ then set $c_0 = \tau_{-1}(c)$, the largest knot of $\widehat{\varphi}_n^0$ not larger than c, set $x_0 = m$, and set $d_0 = \tau_R$. If $c \leq m < c+\delta/2$, set $c_0 = \tau_L$, set $x_0 = m$, and let $d_0 = \tau_1(c+\delta)$, the smallest knot of $\widehat{\varphi}_n^0$ not smaller than $c+\delta$.

Case 2.2 Assume $\widehat{N} \not\subset (c, c+\delta)$. Then $(c, c+\delta) \setminus \widehat{N}$ is an interval, and we set x_0 to be the midpoint of this interval; if $m \geq c+\delta/2$ then set $c_0 = \tau_{-1}(c)$ and and set $d_0 = \tau_1(c+\delta) \wedge \tau_L$. Similarly, if $m < c+\delta/2$, set $c_0 = \tau_{-1}(c) \vee \tau_R$ and set $d_0 = \tau_1(c+\delta)$. Since $|\widehat{N}| \leq \delta/4$, $c_0 - c \leq \delta/4$ and $c+\delta - d_0 < \delta/4$ (where only one of the previous inequalities is relevant, depending on whether $m < c+\delta/4$ or $m > c+3\delta/4$).

For both Case 2.1 and 2.2 we then define $\Delta \in \mathcal{D}_3 \subset \mathcal{D}_4$ by

$$\Delta(x) := \begin{cases} 0, & x \in [c_0, d_0]^c, \\ 1 + \beta_1(x - x_0) & x \in [c_0, x_0], \\ 1 + \beta_2(x - x_0), & x \in [x_0, d_0], \end{cases} \tag{A.76}$$

where $\beta_1 \geq 0$ is chosen such that if

$$(\varphi_0 - \widehat{\varphi}_n^0)(c_0) \geq 0 \qquad \text{then} \qquad \Delta(c_0) = 0, \text{ and if} \tag{A.77}$$

$$(\varphi_0 - \widehat{\varphi}_n^0)(c_0) < 0 \qquad \text{then} \qquad \operatorname{sign}(\Delta) = \operatorname{sign}(\varphi_0 - \widehat{\varphi}_n^0) \text{ on } [c_0, x_0], \tag{A.78}$$

where $\operatorname{sign}(y)$ is 1 if $y \geq 0$ and -1 if $y < 0$. Similarly, $\beta_2 \leq 0$ is chosen such that if

$$(\varphi_0 - \widehat{\varphi}_n^0)(d_0) \geq 0 \qquad \text{then} \qquad \Delta(d_0) = 0, \text{ and if} \tag{A.79}$$

$$(\varphi_0 - \widehat{\varphi}_n^0)(d_0) < 0 \qquad \text{then} \qquad \operatorname{sign}(\Delta) = \operatorname{sign}(\varphi_0 - \widehat{\varphi}_n^0) \text{ on } [x_0, d_0]. \tag{A.80}$$

That is, Δ is defined to be 1 at x_0 and, if $\varphi_0 - \widehat{\varphi}_n^0$ crosses below 0 on $[c_0, c] \cup [c + \delta, d_0]$ at potential points $\tilde{c}$ or $\tilde{d}$, then Δ crosses below 0 at the same point(s). We note also for future reference in Case 2.1 that if $c + \delta/2 \leq m$ then $(\varphi_0 - \widehat{\varphi}_n^0)(d_0) = 0$ since $d_0 = \tau_R \leq c + \delta$, so we are in case (A.79) and $\Delta(d_0) = 0$. Thus $W(\Delta) = W(\Delta 1_{[c_0, x_0]})$, because we have thus forced $W(\Delta 1_{[x_0, d_0]}) = 1$ (and $1 \leq W(\Delta 1_{[c_0, x_0]})$). Similarly, if $m < c + \delta/2$ then we are in case (A.77), and $W(\Delta) = W(\Delta 1_{[x_0, d_0]})$. Now we check that the conditions of Lemma A.8 hold.

Case 2.1 (continued) If $m \in \widehat{N} \subset [c, c + \delta]$ then Δ is continuous at m (so (A.61) holds), (A.58) holds, and at c_0 and d_0 (A.59) holds (possibly with one discontinuity) since these are both knots.

Case 2.2 (continued) Note, if $\widehat{N} \not\subset [c, c + \delta]$, then Δ is 0 on $\widehat{N} \supseteq \widehat{M}$: if $\widehat{N} \cap [c, c + \delta] = \emptyset$ then this is immediate (since the endpoint of $\widehat{N}$ is the nearest knot to $[c, c + \delta]$), and if one of τ_L or τ_R lies in $(c, c + \delta)$, then $\varphi_0 - \widehat{\varphi}_n^0$ is greater or equal to ε at that point, so Δ will be 0 at that point. Now (A.59) holds at c_0, d_0, and Δ is 0 on $\widehat{N} \supseteq \widehat{M}$ so (A.61) holds.

Now we check the remaining conditions. Conditions (A.70) and (A.71) hold by construction for both Case 2.1 and 2.2. We check Condition (A.72) holds for the two cases.

Case 2.1 (continued) Define $\Delta_*(x)$ to be the triangle function with $\Delta_*(x_0) = 1$ and $\Delta_*(c) = \Delta_*(c + \delta) = 0$. We assume without loss of generality that $m \geq c + \delta/2$. Then, $\Delta 1_{[c, x_0]} \geq \Delta_* 1_{[c, x_0]}$, so by (A.74), $\int \Delta^2(x)dx \geq \int_c^m \Delta_*^2(x)dx \geq (m - c)/3 \geq \delta/6$.

Case 2.2 (continued) Define Δ_* to be the triangle function with $\Delta_*(x_0) = 1$, $\Delta_*(c_0 \vee c) = 0 = \Delta_*(d_0 \wedge (c + \delta))$, and $\Delta_*(x) = 0$ for $x \notin [c, c + \delta]$. Then again $\Delta 1_{[c_0 \vee c, d_0 \wedge (c + \delta)]} \geq \Delta_* 1_{[c_0 \vee c, d_0 \wedge (c + \delta)]}$. Since $c_0 - c \leq \delta/4$ and $c + \delta - d_0 < \delta/4$, $d_0 \wedge (c + \delta) - (c_0 \vee c) \geq 3\delta/4$. Thus, as in (A.75), $\int_c^{c+\delta} \Delta(x)^2 dx \geq \int_c^{c+\delta} \Delta_*(x)^2 dx \geq \delta/4$.

Next we check (A.69) for both Case 2.1 and 2.2. If $\varphi_0 - \widehat{\varphi}_n^0 \geq 0$ on $[c_0, d_0]$ there is nothing to check (since then $|\Delta| = \Delta \leq 1$). Assume that there is thus a point $\tilde{d}$ with $c+\delta < \tilde{d} < d_0$ such that $\varphi_0 - \widehat{\varphi}_n^0 \le 0$ on $[\tilde{d}, d_0]$. (An analogous argument holds for a point $\tilde{c} < c$). By construction, $(\varphi_0 - \widehat{\varphi}_n^0)(c+\delta) \ge \varepsilon \ge \varepsilon\Delta(c+\delta) \ge 0$, and $(\varphi_0 - \widehat{\varphi}_n^0)(\tilde{d}) = \varepsilon\Delta(\tilde{d}) = 0$; on $[c+\delta, d_0]$, $\varphi_0 - \widehat{\varphi}_n^0$ is concave by the definition of d_0. Thus, $|(\varphi_0 - \widehat{\varphi}_n^0)'(x+)| \ge \varepsilon|\beta_2|$ for any $x \in [\tilde{d}, d_0]$. Thus $(\varphi_0 - \widehat{\varphi}_n^0)(x) \le \varepsilon\Delta(x) \le 0$ for $x \in [\tilde{d}, d_0]$. Thus we have shown (A.69).

Lastly, we check (A.73) in all cases. Note that since T is a compact interval strictly contained in $\{f > 0\}$, there exists a constant C_0 such that $f(x) \ge C_0$ for $x \in T$. Now, in Case 1.1, $W(\Delta) \le \|\Delta\| = 1$ where $\|\Delta\| = \sup_{x\in\mathbb{R}} |\Delta(x)|$. And we have $\sigma(\Delta)^2 \ge C_0 \int_{\mathbb{R}} \Delta(x)^2 dx \ge C_0\delta/3$ by (A.74). So let $K' \ge (3/C_0)^{1/2}$, and then (A.73) holds.

Similarly, in Case 1.2, $W(\Delta) \le 1$ and $\sigma(\Delta)^2 \ge C_0\delta/4$ by (A.75), so let $K' \ge (4/C_0)^{1/2}$ and then (A.73) holds.

To handle the remaining cases, we consider $h(x)$ defined by $h(x) = 1_Q(x)(\alpha + \gamma x)$ for $\alpha, \gamma \in \mathbb{R}$ where $Q = [x_0, y_0]$ is a nondegenerate interval, $Q \subseteq T$. We always have

$$W(h) \le \|h\| \qquad \text{and} \qquad \sigma(h)^2 \ge C_0 \int_{x_0}^{y_0} h(x)^2 dx.$$

Now $\int_{\mathbb{R}} h(x)^2 dx$ is invariant under translations of h, sign changes of h, and replacing h by $h(-x)$. Thus, we assume that $h(y_0) > 0$, by replacing h by $-h$ if necessary. If $\min_{y\in Q} h(y) \ge 0$ then let $[x_0, y_0] = [0, y_0]$, taking $x_0 = 0$ by translation. Otherwise, take $x_0 < 0 < y_0$ and $h(0) = 0$, by translation. Furthermore, we assume $h(y_0) = \|h\|$ by replacing $h(x)$ by $-h(-x)$ if $h(x_0) < -h(y_0) < 0$ (so $h(y_0) > 0$, still), or by $h(-x)$ if $h(x_0) > h(y_0) > 0$. Note that we have forced h to be nondecreasing so $\gamma \ge 0$.

Now if we are in the case $\inf_{y\in Q} h(y) = h(0) = \alpha > 0$ with $x_0 = 0$, then

$$\begin{aligned}
\int_0^{y_0} (\alpha + \gamma x)^2 dx &= \frac{1}{3\gamma}\left((\alpha + \gamma y_0)^3 - \alpha^3\right) \\
&= \frac{1}{3} y_0 \left((\alpha + \gamma y_0)^2 + \alpha(\alpha + \gamma y_0) + \alpha^2\right) = \frac{1}{3}(y_0 - x_0)\|h\|^2,
\end{aligned}$$

since $\|h\| = \alpha + \gamma y_0$ in this case. If we are in the case $h(x_0) < 0 < h(y_0)$ with $x_0 < 0 < y_0$, then

$$\int_{x_0}^{y_0} (\gamma x)^2 dx = \frac{\gamma^2}{3} y_0^3 - \frac{\gamma^2}{3} x_0^3 = \frac{1}{3}(y_0 - x_0)\gamma^2(y_0^2 - y_0 x_0 + x_0^2) \ge \frac{1}{3}(y_0 - x_0)\|h\|^2$$

since $\|h\|^2 = \gamma^2 y_0^2$ in this case. Thus, by (C.6),

$$\left(\frac{3}{C_0(y_0 - x_0)}\right)^{1/2} \sigma(h) \ge \|h\|. \tag{A.81}$$

Now we apply these computations to the remaining cases. In Case 1.3, Δ is of the form of h defined above and the corresponding x_0, y_0 satisfy $y_0 - x_0 \ge 3\delta/4$

since $[c_1, d_1] \supseteq [c, c+\delta] \setminus \widehat{N}$ and $|\widehat{N}| \le \delta/4$. Note that we can take $Q \subset T$ by the assumption that there are knots of $\widehat{\varphi}_n^0$ above and below $[c, c+\delta]$, and these bound the support of Δ. Thus this case is complete since $\sigma(\Delta)^2 \ge (C_0/3)(y_0 - x_0)W(\Delta)^2$.

For Case 2, Δ equals $h_1 + h_2$ where h_1, h_2 are of the type considered above and have disjoint support, where both supports are contained in T again by the assumption that $\widehat{\varphi}_n^0$ has knots above and below $[c, c+\delta]$.

Case 2.1 (continued) Assume without loss of generality that $m \ge c + \delta/2$. Then, as noted after display (A.80), $W(\Delta) = W(\Delta 1_{[c_0, x_0]}) \equiv W(h_1)$. For h_1, the corresponding x_0, y_0 satisfy $y_0 - x_0 \ge \delta/2$. Thus,

$$W(\Delta) = W(h_1) \le \frac{6^{1/2}}{C_0^{1/2}\delta^{1/2}}\sigma(h_1) \le \frac{6^{1/2}}{C_0^{1/2}\delta^{1/2}}\sigma(\Delta).$$

Case 2.2 (continued) In this case, for both h_1, h_2, the corresponding x_0, y_0 satisfy $y_0 - x_0 \ge 3\delta/8$ (again using $|\widehat{N}| \le \delta/4$). Thus,

$$W(\Delta) = \max(W(h_1), W(h_2)) \le \frac{8^{1/2}}{C_0^{1/2}\delta^{1/2}} \max(\sigma(h_1), \sigma(h_2)) \le \frac{8^{1/2}}{C_0^{1/2}\delta^{1/2}}\sigma(\Delta).$$

This completes the proof. □

Now we complete the proof of Theorem 4.7.B. We treat the case $m \in K$. We can always enlarge K so this holds. Now, since $\varphi_0''(m) < 0$, there is an interval K^0 containing m such that $\widehat{\varphi}_n^0$ has knots above and below K^0 with high probability for large n since $\widehat{\varphi}_n^0$ is uniformly consistent by Proposition 7.2 of Doss and Wellner [2018b]. Thus, K^0 satisfies the condition needed for Lemma A.10. Now suppose that

$$\sup_{t\in K}(\widehat{\varphi}_n^0 - \varphi_0)(t) \ge C\varepsilon_n \quad \text{ or } \quad \sup_{t\in[A+\delta_n, B-\delta_n]}(\varphi_0 - \widehat{\varphi}_n^0(t) \ge C\varepsilon_n$$

for some $C > 0$ where $\varepsilon_n = \rho_n^{2/5}$ and $\delta_n = \rho_n^{1/5} = \varepsilon_n^{1/2}$. By Lemma A.3 of Dümbgen and Rufibach [2009] (stated below as Lemma A.11 for convenience) with $\varepsilon = C\varepsilon_n$, if $C \ge K(2, L)^{-2}$ and n is large it follows that there is a random interval $[c_n, c_n + \delta_n]$ either contained in K^0 or contained in $K \setminus K^0$ on which either $\widehat{\varphi}_n^0 - \varphi_0 \ge C\varepsilon_n/4$ or $\varphi_0 - \widehat{\varphi}_n^0 \ge C\varepsilon_n/4$. In the case $[c_n, c_n + \delta_n] \subset K^0$, then since $\tau_R - \tau_L = O_p(n^{-1/5})$ by Proposition 7.3 of Doss and Wellner [2018b]. (since we assumed $\varphi_0''(m) < 0$) so for n large

$$|\widehat{N}| = \tau_R - \tau_L \le \delta_n/4 = (\log n/n)^{1/5}/4,$$

so we can find a random function Δ_n with no more than four knots which satisfies the conditions of Lemma A.10. If $[c_n, c_n+\delta_n] \subset K \setminus K^0$ then we can find a random function Δ_n with no more than three knots satisfying the conditions of Lemma A.9. Now, calculating as in the proof of Theorem 4.1 of Dümbgen and Rufibach [2009], we can see that for a constant G_0,

$$C^2 \le \frac{16G_0^2(1+o(1))\varepsilon_n^{-2}\rho_n}{\sigma^2(\Delta_n)} = \frac{16G_0^2(1+o(1))}{\delta_n^{-1}\sigma^2(\Delta_n)} \le \frac{48G_0^2(1+o(1))}{\inf_{t\in K} f_0(t)}.$$

But if we choose C strictly larger than the constant on the right side we find that the set is empty, and hence has probability 0 on an event with probability increasing to 1.

$\square$

Lemma A.11 (Lemma A.3, Dümbgen and Rufibach [2009]). *For any $\beta \in [1,2]$ and $L > 0$ there exists a constant $K = K(\beta, L) \in (0,1]$ with the following property: suppose that g and $\hat{g}$ are concave and real-valued functions on $T = [A, B]$ where $g \in \mathcal{H}^{\beta,L}(T)$. Let $\varepsilon > 0$ and $0 < \delta < K \min(B - A, \varepsilon^{1/\beta})$. Then*

$$\sup_{t \in T} (\hat{g} - g)(t) \geq \varepsilon \quad \text{or} \quad \sup_{t \in [A+\delta, B-\delta]} (g - \hat{g})(t) \geq \varepsilon$$

implies that for some $c \in [A, B - \delta]$

$$\inf_{t \in [c, c+\delta]} (\hat{g} - g)(t) \geq \varepsilon/4 \quad \text{or} \quad \inf_{t \in [c, c+\delta]} (g - \hat{g})(t) \geq \varepsilon/4.$$

B. Local asymptotic distribution theory near the mode

B.1. Limit processes and scaling relations

From Theorems 5.1 and 5.2 of Doss and Wellner [2018b], we know that the processes H and $H^{(2)} = \widehat{\varphi}$ and H^0 and $(H^0)^{(2)} = \widehat{\varphi}^0$ exist and are unique in the limiting Gaussian white noise problem described by (1.2). We now introduce further notation and basic scaling results that are needed in the proof of Theorem 1.1. As in Groeneboom, Jongbloed and Wellner [2001] Appendix A, Proposition A.1, and Theorem 4.6 of Balabdaoui, Rufibach and Wellner [2009] (noting the corrections indicated in Subsection B.2 below), $\sigma \equiv 1/\sqrt{f_0(m)}$, $a = |\varphi_0^{(2)}(m)|/4!$, and let

$$\begin{aligned}
Y_{a,\sigma}(t) &\equiv \sigma \int_0^t W(s)ds - at^4 \stackrel{d}{=} \sigma(\sigma/a)^{3/5} Y((a/\sigma)^{2/5} t), \\
Y_{a,\sigma}^{(1)}(t) &= \sigma W(t) - 4at^3 \stackrel{d}{=} \sigma(\sigma/a)^{1/5} Y^{(1)}((a/\sigma)^{2/5} t),
\end{aligned}$$

where $Y \equiv Y_{1,1}$. These processes arise as the limits of appropriate (integrated) localized empirical processes. Similar relations are satisfied by the unconstrained and constrained invelope processes $H_{a,\sigma}$, $H_{a,\sigma}^0$, and their derivatives: with $H \equiv H_{1,1}$ and $H^0 \equiv H_{1,1}^0$, where H^0 can be either H_L or H_R,

$$\begin{aligned}
H_{a,\sigma}(t) &\stackrel{d}{=} \sigma(\sigma/a)^{3/5} H((a/\sigma)^{2/5} t), \\
H_{a,\sigma}^0(t) &\stackrel{d}{=} \sigma(\sigma/a)^{3/5} H^0((a/\sigma)^{2/5} t), \\
H_{a,\sigma}^{(1)}(t) &\stackrel{d}{=} \sigma(\sigma/a)^{1/5} H^{(1)}((a/\sigma)^{2/5} t), \\
(H_{a,\sigma}^0)^{(1)}(t) &\stackrel{d}{=} \sigma(\sigma/a)^{1/5} (H^0)^{(1)}((a/\sigma)^{2/5} t),
\end{aligned}$$

and

$$\begin{aligned}\widehat{\varphi}_{a,\sigma} &= H_{a,\sigma}^{(2)} \stackrel{d}{=} \sigma^{4/5} a^{1/5} H^{(2)}((a/\sigma)^{2/5}\cdot) \\ &= \frac{1}{\gamma_1\gamma_2^2} H^{(2)}(\cdot/\gamma_2) \equiv \frac{1}{\gamma_1\gamma_2^2}\widehat{\varphi}(\cdot/\gamma_2), \end{aligned} \tag{B.1}$$

and, similarly,

$$\begin{aligned}\widehat{\varphi}_{a,\sigma}^0 &= (H_{a,\sigma}^0)^{(2)} \stackrel{d}{=} \sigma^{4/5} a^{1/5} (H^0)^{(2)}((a/\sigma)^{2/5}\cdot) \\ &= \frac{1}{\gamma_1\gamma_2^2} (H^0)^{(2)}(\cdot/\gamma_2) \equiv \frac{1}{\gamma_1\gamma_2^2}\widehat{\varphi}^0(\cdot/\gamma_2), \end{aligned} \tag{B.2}$$

Here

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

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

and we note that

$$\gamma_1\gamma_2^{3/2} = \sigma^{-1} = \sqrt{f_0(m)}, \quad \gamma_1\gamma_2^4 = a^{-1} = \frac{4!}{|\varphi_0^{(2)}(m)|}, \tag{B.5}$$

$$\gamma_1\gamma_2^2 = \frac{1}{C(m,\varphi_0)} \equiv \left(\frac{4! f_0(m)^2}{|\varphi_0^{(2)}(m)|} \right)^{1/5}. \tag{B.6}$$

B.2. Corrections for Balabdaoui, Rufibach and Wellner [2009]

In (4.25) of Balabdaoui, Rufibach and Wellner [2009], replace

$$Y_{k,a,\sigma}(t) := a\int_0^t W(s)ds - \sigma t^{k+2}$$

by

$$Y_{k,a,\sigma}(t) := \sigma\int_0^t W(s)ds - a t^{k+2}$$

to accord with Groeneboom, Jongbloed and Wellner [2001], page 1649, line -4, when $k=2$. In (4.22) of Balabdaoui, Rufibach and Wellner [2009], page 1321, replace the definition of γ_1 by

$$\gamma_1 = \left(\frac{f_0(x_0)^{k+2}|\varphi_0(x_0)|^3}{(4!)^3} \right)^{1/(2k+1)}.$$

In (4.23) of Balabdaoui, Rufibach and Wellner [2009], page 1321, replace the definition of γ_2 by

$$\gamma_2 = \left(\frac{((k+2)!)^2}{f_0(x_0)|\varphi_0^{(k)}(x_0)|^2} \right)^{1/(2k+1)}.$$

When $k = 2$ and $x_0 = m$, these definitions of γ_1, γ_2 reduce to γ_1 and γ_2 as given in (B.4). One line above (4.25) of Balabdaoui, Rufibach and Wellner [2009], page 1321, change $Y_{a,k,\sigma}$ to $Y_{k,a,\sigma}$.

C. Lemmas

Below are some useful lemmas.

Lemma C.1. *Let $f_{in} = e^{\varphi_{in}}$ for $i = 1, 2$, and let x be such that $|f_{1n}(x) - f_{2n}(x)| \to 0$ as $n \to \infty$. Then*

$$f_{1n}(x) - f_{2n}(x) = (\varphi_{1n}(x) - \varphi_{2n}(x))\, e^{\tilde{\varepsilon}_n(x)} e^{\varphi_{2n}(x)} \tag{C.1}$$

$$= \left(\varphi_{1n}(x) - \varphi_{2n}(x) + \frac{(\varphi_{1n}(x) - \varphi_{2n}(x))^2}{2} e^{\varepsilon_n(x)} \right) e^{\varphi_{2n}(x)} \tag{C.2}$$

where $\tilde{\varepsilon}_n(x)$ and $\varepsilon_n(x)$ both lie between 0 and $\varphi_{1n}(x) - \varphi_{2n}(x)$, and thus converge to 0 as $n \to \infty$.

Proof. Taylor expansion shows

$$f_{1n}(x) - f_{2n}(x) = e^{\varphi_{1n}(x)} - e^{\varphi_{2n}(x)} = \left(e^{\varphi_{1n}(x) - \varphi_{2n}(x)} - 1 \right) e^{\varphi_{2n}(x)}$$
$$= (\varphi_{1n}(x) - \varphi_{2n}(x))\, e^{\varepsilon_n(x)} e^{\varphi_{2n}(x)},$$

yielding (C.1). The second expression, (C.2), follows from a similar (two-term) expansion. □

Lemma C.2. *Assume φ_0 is twice continuously differentiable in a neighborhood of m and $\varphi_0''(m) < 0$. Let I be a random interval whose endpoints are in an $O_p(n^{-1/5})$ neighborhood of m. Let D be such that for any $\xi_n \to m$, $|(\widehat{\varphi}_n^0)'(\xi_n) - \varphi_0'(\xi_n)| \le Dn^{-1/5}$ with probability $1 - \varepsilon$ for large n by Corollary 5.4 of Doss and Wellner [2018b]. for $\varepsilon > 0$. Assume $n^{1/5}\lambda(I) \ge 8D/\varphi_0^{(2)}(m)$. Let $L > 0$, $\varepsilon > 0$, and $\check{\delta} > 0$. Suppose there exists $\check{K} > 0$ such that*

$$\int_I (\widehat{\varphi}_n^0 - \widehat{\varphi}_n)^2 d\lambda \le \frac{\check{\delta}}{L} \check{K} n^{-1} \tag{C.3}$$

with probability $1 - \varepsilon$ for n large. Then for any interval $J \subset I$ where $\lambda(J) = Ln^{-1/5}$, we have with probability $1 - \varepsilon$ for n large

(A) $\|\widehat{\varphi}_n - \widehat{\varphi}_n^0\|_J \le \delta O_p(n^{-2/5})$,

(B) $\|\widehat{f}_n - \widehat{f}_n^0\|_J \le \delta O_p(n^{-2/5})$,
(C) $\|\widehat{F}_n - \widehat{F}_n^0\|_J \le \delta O_p(n^{-3/5})$, *and*
(D) *there exist knots* $\eta \in S(\widehat{\varphi}_n) \cap J$ *and* $\eta^0 \in S(\widehat{\varphi}_n^0) \cap J$ *such that* $|\eta - \eta^0| \le \delta O_p(n^{-1/5})$,
(E) *and, letting* $K = [\max(\eta, \eta^0) + \delta O_p(n^{-1/5}), \sup J - \delta O_p(n^{-1/5})]$, *we have* $\|\widehat{\varphi}_n' - (\widehat{\varphi}_n^0)'\|_K \le \delta O_p(n^{-1/5})$,

where $\delta \to 0$ *as* $\check{\delta} \to 0$.

Proof. First we prove the first three statements. By Taylor expansion, $\varphi_0'(x) = \varphi_0^{(2)}(\xi)(x-m)$ where ξ is between x and m. Let $J = [j_1, j_2] \subseteq I$. Then $\varphi_0'(j_2) - \varphi_0'(j_1) \ge |\varphi_0^{(2)}(m)| L n^{-1/5} 2$ for n large enough since $\varphi_0^{(2)}$ is continuous near m. Note with probability $1-\varepsilon$, $|(\widehat{\varphi}_n^0)'(j_1) - \varphi_0'(j_1)|$ and $|(\widehat{\varphi}_n^0)'(j_2) - \varphi_0'(j_2)|$ are less than $n^{-1/5} D$ by applying Corollary 5.4 of Doss and Wellner [2018b] twice taking $\xi_n = j_1$ and $\xi_n = j_2$, and $C = 0$ (not $C = \lambda(J)$). Here $(\widehat{\varphi}_n^0)'$ may be the right or left derivative. Now apply Lemma C.3 (taking I in that lemma to be our J) with $\varphi_U' - \varphi_L' = n^{-1/5}\left(2D + |\varphi_0^{(2)}(m)| L/2\right)$ and $\varepsilon = \check{\delta}\check{K} n^{-1}/L$. Then for small enough $\check{\delta}$,

$$\left(\frac{\check{\delta}\check{K}}{L(2D + |\varphi_0^{(2)}(m)|L/2)^2}\right)^{1/3} n^{-1/5} \le L n^{-1/5} = \lambda(J),$$

as needed. Thus, Lemma C.3 allows us to conclude

$$\|\widehat{\varphi}_n^0 - \widehat{\varphi}_n\| \le \left(8 \frac{n^{-6/5}\check{\delta}\check{K}}{L}\left(2D + \varphi_0^{(2)}(m) L/2\right)\right)^{1/3}.$$

Thus taking $\check{\delta}$ so that $\check{\delta}\check{K}D \to 0$, we see that Lemma C.2 (A) holds. Then (B) follows by the delta method (or Taylor expansion of exp). Note that $\check{K}$ and D depend only on ε, not on I.

We show (C) and (D) next. Note that (C) follows from (D) and (B). This is because

$$\widehat{F}_n(x) - \widehat{F}_n^0(x) = \widehat{F}_n(\eta) - \widehat{F}_n^0(\eta) + \int_\eta^x (\widehat{f}_n(x) - \widehat{f}_n^0(x)) dx.$$

By (B), the second term above is $\delta O_p(n^{-3/5})$ since $x \in J$ satisfies $|x - \eta| \le L n^{-1/5}$. We can next see that the first term in the previous display is $\delta^{1/2} O_p(n^{-3/5})$. Notice that $\sup_{t \in [\eta, \eta^0]} |\int_\eta^t d(\widehat{F}_n(u) - \mathbb{F}_n(u))| = \delta^{1/2} O_p(n^{-3/5})$ by Lemma C.4, where the random variable implicit in the O_p statement depends on J only through L. Since $|\widehat{F}_n(\eta) - \mathbb{F}_n(\eta)| \le 1/n$ by Corollary 2.4 of Doss and Wellner [2018b], we see that $\sup_{t \in [\eta, \eta^0]} |\widehat{F}_n(t) - \mathbb{F}_n(t)| = \delta^{1/2} O_p(n^{-3/5})$. Similarly, since $|\widehat{F}_n^0(\eta^0) - \mathbb{F}_n(\eta^0)| \le 1/n$ by Corollary 2.12 of Doss and Wellner [2018b], $\sup_{t \in [\eta, \eta^0]} |\widehat{F}_n^0(t) - \mathbb{F}_n(t)| = \delta^{1/2} O_p(n^{-3/5})$ by analogous computations. Together, these let us conclude that $\sup_{t \in [\eta, \eta^0]} |\widehat{F}_n^0(t) - \widehat{F}_n(t)| = \delta^{1/2} O_p(n^{-3/5})$.

Now we prove (D). Let $\delta > 0$ and define i_1 and i_2 by $I = [i_1 - \delta n^{-1/5}, i_2 + \delta n^{-1/5}]$, taking δ small enough that $i_1 < i_2$. Let $i_2 - i_1 = \tilde{M} n^{-1/5}$. Then, by the Taylor expansion of φ_0' (see the beginning of this proof), $\varphi_0'(i_2) - \varphi_0'(i_1) \geq |\varphi_0^{(2)}(m)|\tilde{M} n^{-1/5}/2$ for δ small and n large enough, since $\varphi_0^{(2)}$ is continuous. Additionally, by applying Corollary 7.1 of Doss and Wellner [2018b] twice, taking $\xi_n = i_1$ and $\xi_n = i_2$ and $C = 0$ (not $C = \tilde{M}$), we find D (independent of $\lambda(I)$) such that $|(\widehat{\varphi}_n^0)'(i_1) - \varphi_0'(i_1)|$ and $|(\widehat{\varphi}_n^0)'(i_2) - \varphi_0'(i_2)|$ are, with probability $1-\varepsilon$, less than $n^{-1/5}D$. Here $(\widehat{\varphi}_n^0)'$ may be the right or left derivative. For δ small enough, by assumption $D < |\varphi_0^{(2)}(m)|\tilde{M}/8$, and we then have $(\widehat{\varphi}_n^0)'(i_1+) - (\widehat{\varphi}_n^0)'(i_2-) \geq |\varphi_0^{(2)}(m)|\tilde{M} n^{-1/5}/4$. We do not know a priori how much $(\widehat{\varphi}_n^0)'$ decreases at any specific knot in $S(\widehat{\varphi}_n^0)$, but by partitioning $[i_1, i_2]$ into intervals of a fixed length, we can find one such interval on which $(\widehat{\varphi}_n^0)'$ decreases by the corresponding average amount. That is, there exists a subinterval of $[i_1, i_2]$, denoted $J^* = [l^*, r^*]$, of length $2\delta n^{-1/5}$, such that

$$(\widehat{\varphi}_n^0)'(l^*+) - (\widehat{\varphi}_n^0)'(r^*-) \geq \frac{|\varphi_0^{(2)}(m)|\tilde{M} n^{-1/5}/4}{(i_2 - i_1)/2\delta n^{-1/5}} = \frac{|\varphi_0^{(2)}(m)|\delta n^{-1/5}}{2} \equiv K n^{-1/5}$$

since $i_2 - i_1 = \tilde{M} n^{-1/5}$. Let $x_l^* = \sup\left\{x \in J^* : (\widehat{\varphi}_n^0 - \widehat{\varphi}_n)'(x) \geq 0\right\}$, and let $x_l^* = l^*$ if the set is empty, and let $x_r^* = \inf\left\{x \in J^* : (\widehat{\varphi}_n^0 - \widehat{\varphi}_n)'(x) \leq 0\right\}$, and $x_r^* = x_l^* = r^*$ if the set is empty. Now $(\widehat{\varphi}_n^0 - \widehat{\varphi}_n)'$ decreases by at least $K n^{-1/5}/2$ on either $[l^*, x_l^*]$ or on $[x_r^*, r^*]$ (since $(\widehat{\varphi}_n^0 - \widehat{\varphi}_n)'$ is constant on $[x_l^*, x_r^*]$). Let $\eta_l^0 = \inf S(\widehat{\varphi}_n^0) \cap J^*$ and $\eta_r^0 = \sup S(\widehat{\varphi}_n^0) \cap J^*$. Now by assumption $\widehat{\varphi}_n$ is linear on $[\eta_l^0 - \delta n^{-1/5}, \eta_r^0 + \delta n^{-1/5}]$ (since $\widehat{\varphi}_n$ is linear on J^* and within $\delta n^{-1/5}$ of any knot of $\widehat{\varphi}_n^0$), and so

$$(\widehat{\varphi}_n^0 - \widehat{\varphi}_n)'(u) \geq \frac{K n^{-1/5}}{2} \qquad \text{for } u \in [\eta_l^0 - \delta n^{-1/5}, \eta_l^0], \quad \text{or} \tag{C.4}$$

$$(\widehat{\varphi}_n^0 - \widehat{\varphi}_n)'(u) \leq -\frac{K n^{-1/5}}{2} \qquad \text{for } u \in [\eta_r^0, \eta_r^0 + \delta n^{-1/5}], \tag{C.5}$$

depending on whether $(\widehat{\varphi}_n^0 - \widehat{\varphi}_n)'$ decreases by $K n^{-1/5}/2$ on $[l^*, x_l^*]$ (in which case $\eta_l^0 \leq x_l^*$) or on $[x_r^*, r^*]$ (in which case $x_r^* \leq \eta_r^0$). (In the former case, (C.4) holds because $(\widehat{\varphi}_n^0 - \widehat{\varphi}_n)'$ is nonincreasing and its decrease to 0 on $[l^*, x_l^*]$ actually happens on $[\eta_l^0, x_l^*]$ since η_l^0 is the last knot of $\widehat{\varphi}_n^0$ in J^*. Similar reasoning in the latter case yields (C.5).) If (C.4) holds then

$$\sup_{u \in [\eta_l^0 - \delta n^{-1/5}, \eta_l^0]} |\widehat{\varphi}_n^0(u) - \widehat{\varphi}_n(u)| \geq (K n^{-1/5}/2)(\delta n^{-1/5}/2) = |\varphi_0^{(2)}(m)| n^{-2/5}\delta^2/8,$$

and if (C.5) holds then $\sup_{u \in [\eta_r^0, \eta_r^0 + \delta n^{-1/5}]} |\widehat{\varphi}_n^0(u) - \widehat{\varphi}_n(u)| \geq |\varphi_0^{(2)}(m)| n^{-2/5}\delta^2/8$. This allows us to lower bound $\int_I (\widehat{\varphi}_n^0(u) - \widehat{\varphi}_n(u))^2$, to attain a contradiction with (C.3).

Assume that (C.4) holds. The case where (C.5) holds is shown analogously. Let $z = \operatorname{argmin}_{u \in [i_1, i_2]} |\widehat{\varphi}_n^0(u) - \widehat{\varphi}_n(u)|$. Let L be the affine function such that

$L(z) = \widehat{\varphi}_n^0(z) - \widehat{\varphi}_n(z)$ and L has slope $Kn^{-1/5}/2$. Then for $x \in [\eta_l^0 - \delta n^{-1/5}, \eta_l^0]$, $|L(x)| \le |\widehat{\varphi}_n^0(x) - \widehat{\varphi}_n(x)|$. For $z \le x \le \eta_l^0$, this is because

$$\begin{aligned}|\widehat{\varphi}_n^0(x) - \widehat{\varphi}_n(x)| = \widehat{\varphi}_n^0(x) - \widehat{\varphi}_n(x) &= \widehat{\varphi}_n^0(z) - \widehat{\varphi}_n(z) + \int_z^x (\widehat{\varphi}_n^0 - \widehat{\varphi}_n)' d\lambda \\ &\ge L(z) + \int_z^x L' d\lambda = |L(x)|,\end{aligned}$$

where the first equality holds since $\widehat{\varphi}_n^0 - \widehat{\varphi}_n$ is increasing on $[\eta_l^0 - \delta n^{-1/5}, \eta_l^0]$ (by (C.4)), so by the definition of z, $\widehat{\varphi}_n^0(x) - \widehat{\varphi}_n(x) \ge 0$ for $x \ge z$, and the last is similar. For $\eta_l^0 - \delta n^{-1/5} \le x \le z$,

$$\begin{aligned}-|\widehat{\varphi}_n^0(x) - \widehat{\varphi}_n(x)| = (\widehat{\varphi}_n^0(x) - \widehat{\varphi}_n)(x) &= (\widehat{\varphi}_n^0 - \widehat{\varphi}_n)(z) - \int_x^z (\widehat{\varphi}_n^0 - \widehat{\varphi}_n)' d\lambda \\ &\le L(z) - \int_x^z L' d\lambda = -|L(x)|,\end{aligned}$$

where the first and last equalities follow because for $\eta_l^0 \le x \le z$, since $\widehat{\varphi}_n^0 - \widehat{\varphi}_n$ is increasing it must be negative by the definition of z. Thus, $|L(u)| \le |\widehat{\varphi}_n^0(u) - \widehat{\varphi}_n(u)|$ on $[\eta_l^0 - \delta n^{-1/5}, \eta_l^0]$ so

$$\frac{\delta^3 K^2 n^{-1}}{3 \cdot 2^5} \le \int_{[\eta_l^0 - \delta n^{-1/5}, \eta_l^0]} L^2 d\lambda \le \int_{[\eta_l^0 - \delta n^{-1/5}, \eta_l^0]} (\widehat{\varphi}_n^0 - \widehat{\varphi}_n)^2 d\lambda \le \int_I (\widehat{\varphi}_n^0 - \widehat{\varphi}_n)^2 d\lambda,$$

where the quantity on the far left is $\int_0^{\delta n^{-1/5}/2} (xKn^{-1/5}/2)^2 dx$. This is a contradiction if δ is fixed and we let $\check{\delta} \to 0$, since $[\eta_l^0 - \delta n^{-1/5}, \eta_r^0 + \delta n^{-1/5}] \subset I$ by the definition of $[i_1, i_2]$, and then $\int_I (\widehat{\varphi}_n^0 - \widehat{\varphi}_n)^2 d\lambda \le \check{\delta} \check{K} n^{-1}$. A similar inequality can be derived if (C.5) holds. Thus $\delta \to 0$ as $\check{\delta} \to 0$.

Finally, we show (E) holds with similar logic. Let $\xi_1 < \xi_2$ be points such that $\widehat{\varphi}_n$ is linear on $[\xi_1, \xi_2]$, and let $\delta > 0$. Then if all knots ξ^0 of $\widehat{\varphi}_n^0$ satisfy $|\xi_i - \xi^0| > \sqrt{\delta} n^{-1/5}$, $i = 1, 2$, then we can see that $\|\widehat{\varphi}_n' - (\widehat{\varphi}_n^0)'\|_{[\xi_1, \xi_2]} \le \sqrt{\delta} n^{-1/5}$. This is because at any $x \in [\xi_1 + \sqrt{\delta} n^{-1/5}, \xi_2 - \sqrt{\delta} n^{-1/5}]$, if $(\widehat{\varphi}_n^0 - \widehat{\varphi}_n)'(x) > \sqrt{\delta} n^{-1/5}$, then $(\widehat{\varphi}_n^0 - \widehat{\varphi}_n)' > \sqrt{\delta} n^{-1/5}$ on $[x - \sqrt{\delta} n^{-1/5}, x]$ (since $\widehat{\varphi}_n$ is linear on a $\sqrt{\delta} n^{-1/5}$ neighborhood of x), so $(\widehat{\varphi}_n^0 - \widehat{\varphi}_n)(x - \sqrt{\delta} n^{-1/5}, x] > \delta n^{-2/5}$, a contradiction. Here we use the notation $g(a, b] = g(b) - g(a)$. Similarly if $(\widehat{\varphi}_n^0 - \widehat{\varphi}_n)'(x) < \sqrt{\delta} n^{-1/5}$ then $(\widehat{\varphi}_n^0 - \widehat{\varphi}_n)(x, x + \sqrt{\delta} n^{-1/5}] < -\delta n^{-2/5}$, a contradiction. We have thus shown that if we take η and η^0 to be the largest knot pair within $\sqrt{\delta} n^{-1/5}$ (meaning $\max(\eta, \eta^0)$ is largest among such pairs) then $\|(\widehat{\varphi}_n - \widehat{\varphi}_n^0)'\|_{[\max(\eta, \eta^0), \sup J - \sqrt{\delta} n^{-1/5}]} \le \sqrt{\delta} n^{-1/5}$, by Part ((A)) and by partitioning $[\max(\eta, \eta^0), \sup J]$ into intervals on which $\widehat{\varphi}_n$ is linear. □

The proofs of Parts (D) and (E) in the previous lemma could also be completed with the roles of $\widehat{\varphi}_n^0$ and $\widehat{\varphi}_n$ reversed. The next lemma provides the calculation used in Lemma C.2 to translate an upper bound on $\int_I \left(\widehat{\varphi}_n^0 - \widehat{\varphi}_n\right)^2 d\lambda$ into an upper bound on $\sup_{x \in I} \left(\widehat{\varphi}_n^0(x) - \widehat{\varphi}_n(x)\right)$ for an appropriate interval I.

Lemma C.3. *Let $\varepsilon > 0$. Assume φ_i, $i = 1, 2$, are functions on an interval I, where*

$$\varphi_L' \le \varphi_i'(x) \le \varphi_U' \ \text{ for } x \in I, i = 1, 2, \tag{C.6}$$

where φ_i' refers to either the left or right derivative and φ_L', φ_U' are real numbers. Assume I is of length no smaller than $\left(\varepsilon / \left(\varphi_U' - \varphi_L'\right)^2\right)^{1/3}$. Assume that $\int_I \left(\varphi_1(x) - \varphi_2(x)\right)^2 dx \le \varepsilon$. Then

$$\sup_{x \in I} |\varphi_1(x) - \varphi_2(x)| \le \left(8\varepsilon \left(\varphi_U' - \varphi_L'\right)\right)^{1/3}.$$

Proof. Assume that x is such that $\varphi_1(x) - \varphi_2(x) = \delta$, and without loss of generality, $\delta > 0$. Then by (C.6), if y is such that $|x - y| \le (\delta/2)/(\varphi_U' - \varphi_L')$ then $\varphi_1(y) - \varphi_2(y) \ge \delta/2$. Thus, if I is an interval whose length is no smaller than $(\delta/2)/(\varphi_U' - \varphi_L')$ then if $\varphi_1(x) - \varphi_2(x) \ge \delta$ for any $x \in J \subseteq I$ where the length of J is equal to $(\delta/2)/(\varphi_U' - \varphi_L')$, then

$$\int_I \left(\varphi_1(x) - \varphi_2(x)\right)^2 dx \ge \int_J \left(\varphi_1(x) - \varphi_2(x)\right)^2 dx \ge \frac{\delta^2}{2^2}\lambda(J) = \frac{1}{8}\frac{\delta^3}{\varphi_U' - \varphi_L'}.$$

Thus, substituting $\varepsilon = \frac{1}{8}\frac{\delta^3}{\varphi_U' - \varphi_L'}$, we see that for $x \in I$, $|\varphi_1(x) - \varphi_2(x)| \le \left(8\varepsilon\left(\varphi_U' - \varphi_L'\right)\right)^{1/3}$, as desired. ☐

Lemma C.4. *Let either Assumption A.1 hold at $x_0 = m$ or Assumption A.2 hold at $x_0 \ne m$, and let $\widehat{F}_n^0$ and $\widehat{F}_n$ be the log-concave mode-constrained and unconstrained MLEs of F_0. Let $I = [v_1, v_2]$ be a random interval whose dependence on n is suppressed and such that $n^{1/5}(v_j - x_0) = O_p(1)$, $j = 1, 2$. Then*

$$\left\{ \begin{array}{l} \sup_{t \in I} \left| \int_{[v_1, t]} d(\widehat{F}_n(u) - \mathbb{F}_n(u)) \right| \\ \\ \sup_{t \in I} \left| \int_{[v_1, t]} d(\widehat{F}_n^0(u) - \mathbb{F}_n(u)) \right| \end{array} \right\} = \sqrt{\lambda(I)} O_p(n^{-2/5}). \tag{C.7}$$

The random variables implicit in the O_p statements in (C.7) depend on I through its length (in which they are increasing) and not the location of its endpoints.

Proof. We analyze $\sup_{t \in I} \left| \int_{v_1}^t d(\widehat{F}_n(u) - \mathbb{F}_n(u)) \right|$ first. Note

$$\begin{aligned} \sup_{t \in I} & \left| \int_{[v_1, t]} d\left(\mathbb{F}_n - \widehat{F}_n\right) \right| \\ & \le \sup_{t \in I} \left| \int_{v_1}^t \left(\widehat{f}_n(u) - f_0(v_1) - (u - v_1) f_0'(v_1)\right) du \right. \\ & \qquad - \int_{v_1}^t \left(f_0(u) - f_0(v_1) - (u - v_1) f_0'(v_1)\right) du \\ & \qquad \left. - \int_{[v_1, t]} d\left(\mathbb{F}_n - F_0\right) \right|. \end{aligned} \tag{C.8}$$

By (the proof of) Lemma 8.16 of Doss and Wellner [2018b],

$$\sup_{t\in I}\left|\int_{[v_1,t]} d\left(\mathbb{F}_n - F_0\right)\right| = \sqrt{\lambda(I)}O_p(n^{-2/5}).$$

The supremum over the middle term in (C.8) is $\lambda(I)O_p(n^{-2/5})$ by a Taylor expansion of f_0, and applying in addition Lemma 4.5 of Balabdaoui, Rufibach and Wellner [2009], we see that the supremum over the first term in (C.8) is also $\lambda(I)O_p(n^{-2/5})$.

The same analysis, using Proposition 5.3 or Corollary 5.4 of Doss and Wellner [2018b], applies to $\sup_{t\in I}\left|\int_{[v_1,t]} d(\widehat{F}_n^0(u) - \mathbb{F}_n(u))\right|$. Note in all cases that the random variables implicit in the O_p statements depend on I only through its length (and they are increasing in the length) and not the location of its endpoints, since $\|f_0\| = f_0(m) < \infty$ and since $f_0^{(2)}$ is continuous and so uniformly bounded in a neighborhood of x_0. □

When we apply the previous lemma, the length of I will depend on ε which gives the probability bound implied by our o_p statements whereas its endpoints will depend on δ, which gives the size bound implied by our o_p statements.

Lemma C.5. *Let $\varepsilon(x)$ and $\tilde{\varepsilon}(x)$ be defined by $e^x = 1 + x + 2^{-1}x^2e^{\varepsilon(x)}$ and $e^x = 1 + xe^{\tilde{\varepsilon}(x)}$. Then*

$$e^{\varepsilon(x)} \le 2e^{\tilde{\varepsilon}(x)}. \tag{C.9}$$

Proof. We can see

$$e^{\varepsilon(x)} = \frac{e^x - 1 - x}{(x^2/2)} = \frac{2}{x^2}\sum_{k=2}^{\infty}\frac{x^k}{k!} = \sum_{k=0}^{\infty}\frac{2x^k}{(k+2)!}.$$

Similarly,

$$e^{\tilde{\varepsilon}(x)} = \frac{e^x - 1}{x} = \frac{1}{x}\sum_{k=1}^{\infty}\frac{x^k}{k!} = \sum_{k=0}^{\infty}\frac{x^k}{(k+1)!}.$$

Comparing coeffients in the two series, we see that

$$\frac{2}{(k+2)!} \le \frac{1}{(k+1)!} \quad \text{for all } k \ge 0$$

since $k + 2 \ge 2$ for $k \ge 0$. It follows that $e^{\varepsilon(x)} \le e^{\tilde{\varepsilon}(x)}$ for all $x \ge 0$. This implies that $\varepsilon(x) \le \tilde{\varepsilon}(x)$ for all $x \ge 0$.

Now for $x \le 0$ we have

$$\begin{aligned} 2e^{\tilde{\varepsilon}(x)} - e^{\varepsilon(x)} &= \frac{2}{x}\left\{e^x - 1 - \frac{1}{x}(e^x - 1 - x)\right\} \\ &= \frac{2}{x}\left\{\sum_{k=1}^{\infty}\frac{x^k}{k!} - \sum_{k=1}^{\infty}\frac{x^k}{(k+1)!}\right\} = \frac{2}{x}\sum_{k=1}^{\infty}\frac{k}{(k+1)!}x^k \end{aligned}$$

where the infinite sum is negative for all $x < 0$ since the first term is negative. It follows that

$$2e^{\tilde{\varepsilon}(x)} - e^{\varepsilon(x)} \geq 0$$

for all $x \leq 0$. Combined with the result for $x \geq 0$ the claimed result holds: $e^{\varepsilon(x)} \leq 2e^{\tilde{\varepsilon}(x)}$ for all $x \in \mathbb{R}$. $\square$

References

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 (2010h:62290)

Doss, C. R. and Wellner, J. A. (2018a). Inference for the mode of a log-concave density. Submitted to the *Annals of Statistics*. arXiv:1611.10348.

Doss, C. R. and Wellner, J. A. (2018b). Log-concave density estimation with symmetry or modal constraints. arXiv:1611.10335v3.

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

Groeneboom, P., Jongbloed, G. and Wellner, J. A. (2001). A canonical process for estimation of convex functions: the "invelope" of integrated Brownian motion $+t^4$. *Ann. Statist.* **29** 1620–1652. MR1891741 (2003c:62075)