Shape-Constrained Inference for Concave-Transformed Densities and their Modes

Charles R. Doss

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

Doctor of Philosophy

University of Washington

2013

Reading Committee:

Jon A. Wellner, Chair

Peter D. Hoff

Volodymyr Minin

Program Authorized to Offer Degree:
Statistics

University of Washington

Abstract

Shape-Constrained Inference for Concave-Transformed Densities and their Modes

Charles R. Doss

Chair of the Supervisory Committee:
Professor Jon A. Wellner
Statistics

We consider inference about functions estimated via shape constraints based on concavity. We consider log-concave densities and other "concave-transformed" densities on the real line, where a concave-transformed class is one given by applying a transformation (e.g. the logarithm or a power function) to concave functions. We expect our proofs and results to be relevant in other concavity-based settings. Concave functions are always unimodal, so concave-transformed densities can be used as surrogates for unimodal ones, and the mode is thus a natural parameter of interest. In nonparametric settings the mode is generally not estimable at a root-n rate and does not always have a normal limiting distribution, and current methods for testing or forming confidence intervals for the location of the mode are generally complicated. In the setting of log-concave density estimation we construct a likelihood ratio test for the location of the mode by comparing the log-concave maximum likelihood estimate (MLE) to the MLE over the constrained subclass of log-concave densities with a fixed mode. The test can be inverted to form a confidence set. We study the properties of the constrained MLE and the Wilks phenomenon of the likelihood ratio statistic. Proving global rates of convergence of $n^{2/5}$, for both the constrained and unconstrained MLEs, is an important step in understanding the likelihood ratio statistic and this result is also

of independent interest. These global rate results apply to Hellinger and total variation distance, as well as to the size of the likelihood ratio statistic, and they apply to many concave-transformed density classes beyond log-concave ones.

TABLE OF CONTENTS

Page

List of Figures . . . iii

Chapter 1: Introduction . . . 1
1.1 Review: Shape-Constrained Estimation . . . 4
1.2 Review: The Mode . . . 11
1.3 Likelihood Ratio Methods and the Testing Approach to Confidence Sets . . . 17
1.4 Overview of this Thesis . . . 18

Chapter 2: Finite Sample Theory . . . 21

Chapter 3: Asymptotics I: Global . . . 42
3.1 Consistency for the CMLE . . . 42
3.2 Global Rates of Convergence for the UMLE and CMLE . . . 49

Chapter 4: Asymptotics II: Local . . . 80
4.1 Lower Bounds for Estimation of the Location of the Mode . . . 80
4.2 Upper Bounds and Tightness for the CMLE at the Mode . . . 90

Chapter 5: Asymptotics III: Limit Distribution Theory . . . 107
5.1 Limit Characterizations . . . 108
5.2 Local Limit Distribution . . . 154
5.3 Inference for the Mode: Likelihood Ratio Methods . . . 195
5.4 Future Work . . . 204

Chapter 6: Computation, Implementation, and Applications . . . 209
6.1 Active Set Algorithm . . . 209
6.2 Simulation Results . . . 214
6.3 Real Data Results . . . 217

Appendix A: Calculations and Technical Results . . . 221
A.1 Calculations for the Gap Problem . . . 221

Appendix B: Skorokhod Topologies . . . 231

Appendix C: Code . . . 239

Bibliography . . . 266

LIST OF FIGURES

Figure Number Page

2.1 Density and log-density UMLE and CMLE 28
2.2 UMLE and CMLE processes 39
4.1 Gap Problem Perturbations 99
5.1 Perturbations for Limit Characterization 113
6.1 Wilks Phenomenon: Likelihood Ratio Asymptotic Distribution 215
6.2 Fixed Alternative 217
6.3 Brightstar rotational velocities 219
6.4 S&P 500 for years 2003-2006 220

ACKNOWLEDGMENTS

I give my heartfelt thanks to my advisor, Jon Wellner, who spent countless hours reading my work and answering my innumerable questions, and whose advice and guidance, especially when progress was slow, made this thesis possible. It was a pleasure to work with Jon the past several years; his genuine interest and curiosity towards statistics and mathematics are contagious to all those around him.

Thanks to Peter Hoff, Vladimir Minin, and Chris Hoffman, for their instruction during my time at the university and for taking the time to read my thesis or to listen to my examinations. Special thanks to Vladimir, whose friendliness and humor, as well as his advice and mentoring, I very much appreciated during our time working together.

Thanks to Tilmann Gneiting and the other faculty, students, and staff at the Applied Mathematics Institute at the University of Heidelberg, for their support, and for being warm and welcoming during my visit there.

Thanks to Cheryl Bissett, Vickie Graybeal, Mee-Ling Hon, Kate Reinking, Ellen Reynolds, and Connie Sugatan, who were always friendly and available to help during my time in the department. Thanks also to Kris Shaw and Matt Jay, with whom it was a pleasure to work on the department website, and who were always very responsive when I had computer-related questions. Thanks to Adrien Saumard, for his helpful proofreading of my thesis.

Finally, thanks to my fellow graduate students, with whom I commiserated about the unpleasant aspects and enjoyed the pleasant aspects of graduate school; your presence brightened the last five years immeasurably.

DEDICATION

To Mom, Dad, Natalie, and Mark,

to Grandma and Grandpa,

and to Grandmaman and Grandpapa,

for their support, love, and sacrifices

Chapter 1

INTRODUCTION

In the field of nonparametric function estimation based on independent and identically distributed (i.i.d.) data, fully automatic estimators, such as maximum likelihood estimators (MLEs), do not exist for many models. One class of estimators used to circumvent this problem can be described by their dependence on smoothness assumptions. These estimators include kernel density estimators, (see, e.g. Parzen (1962)), wavelet estimators (see, e.g. Donoho et al. (1996)), and estimators based on penalizing roughness (see, e.g. Silverman (1982)). The estimators in this very broad family all share a dependence upon tuning parameters. While there exist methods for choosing tuning parameters automatically, there is never a guarantee that these methods give rise to the optimal choice, and in most instances the performance of the estimator is highly dependent on the tuning parameter. In fact, different functionals of the true underlying function generally have different optimal choices of tuning parameters for their estimation, and the optimal choices often depend on second or third derivatives of the true function, which are estimated very poorly. Furthermore, the difficulty in choosing tuning parameters is exacerbated as the dimension d of the data increases, when the number of tuning parameters may be $O(d^2)$, as in the case of the kernel density estimator (with a fully general bandwidth matrix).

An alternative class of estimators can be described as "shape-constrained" estimators. An example arises from estimation of a regression function which is assumed to be concave. Such an assumption may arise naturally in many contexts (in contrast to smoothness assumptions); as noted by Hanson and Pledger (1976),

> In economics, utility functions are usually assumed to be concave; marginal util-

> ity is often assumed to be convex; and functions representing productivity, supply, and demand curves are often assumed to be either concave or convex.

It was the case of a concave production function that motivated Hildreth (1954) to first apply concave regression, which has since been studied by Hanson and Pledger (1976), Mammen (1991), and Seijo and Sen (2011), among others. Since the negative of a concave function is a convex function, studying concave regression is the same as studying convex regression.

Concavity is the shape constraint about which this thesis is concerned. We will use this shape constraint to carry out inference about the location of the mode of a nonparametrically estimated unimodal function on $\mathbb{R}$. We say that a function f is *unimodal* if there exists a (possibly not unique) point m, a *mode* of f, such that f is nondecreasing on $(-\infty, m]$ and nonincreasing on $[m, \infty)$. Concave functions are always unimodal, so we will develop tests and confidence intervals based on the assumption of concavity. Future work will include understanding how these tests perform in more general (misspecified) settings. Here are some examples of problems we have in mind.

Example 1.0.1 (Concave or Convex Regression)**.** Suppose that we observe i.i.d. random variables $(X_i, Y_i) \in \mathbb{R}^d \times \mathbb{R}$ where we assume that

$$Y_i = \varphi(X_i) + \epsilon_i$$

where ϵ_i are mean 0 random noise variables and φ is a concave function. This is the problem mentioned above, for which rates of convergence of the least squares estimator for this model are given by Mammen (1991) for $d = 1$, and consistency is given for general $d \geq 1$ by Seijo and Sen (2011). Note that all results also hold if φ is assumed to be convex instead of concave. Consider the problem of testing $H_0 \colon m(\varphi) = m_0$ versus $H_1 \colon m(\varphi) \neq m_0$, where $m(\varphi)$ is the componentwise infimum of $\arg\max \varphi$ and $m_0 \in \mathbb{R}^d$. The corresponding confidence region estimation problem is to find a $1 - \alpha$ confidence region for $m(\varphi)$ where $\alpha \in (0, 1)$.

Example 1.0.2 (Convex Hazard Function)**.** Suppose we observe i.i.d. $X_1, \ldots, X_n \in \mathbb{R}$ which are assumed to come from a distribution that has a convex (or "bathtub shaped") hazard function $\lambda = f/(1-F)$ where f and F are the density and cumulative distribution function (c.d.f.) of the X_i, respectively. Jankowski and Wellner (2009) find characterizations, show consistency, and, under certain conditions, find limiting distributions at a fixed point for $\widehat{\lambda}$, where $\widehat{\lambda}$ is the MLE of λ. Consider the problem of testing $H_0\colon m(\lambda) = m_0$ versus $H_1\colon m(\lambda) \neq m_0$ where m is the infimum of $\arg\min \lambda$ and $m_0 \in \mathbb{R}$. The corresponding interval estimation problem is to find a $1-\alpha$ confidence interval for $m(\lambda)$ where $\alpha \in (0,1)$.

Example 1.0.3 (Concave-transformed Density)**.** Suppose we observe i.i.d. $X_1, \ldots, X_n \in \mathbb{R}^d$ which are assumed to come from a density f that is of the form $h \circ \varphi$ where φ is concave and h is known and is either $h(y) = e^y$ or $h(y) = (-y)^{1/s}$ for $s \in (-1, 0)$. Since h is increasing and φ is concave and thus unimodal, f is also always unimodal. Estimation in these settings is studied in Koenker and Mizera (2010) and Seregin and Wellner (2010). Consider the problem of testing $H_0\colon m(f) = m_0$ versus $H_1\colon m(f) \neq m_0$ where m is the componentwise infimum of $\arg\max f$ and $m_0 \in \mathbb{R}^d$. The corresponding confidence region estimation problem is to find a $1-\alpha$ confidence region for $m(f)$ where $\alpha \in (0,1)$.

We consider all of these settings to be "nonparametric concavity-based function estimation problems" since, up to a fixed and known transformation, the function being estimated is concave. Such problems share many features, including having pointwise rates of convergence of $n^{2/5}$ (under a strict curvature assumption), global rates of convergence of at least $n^{2/5}$, and a universal limiting distribution for pointwise estimates (up to a scaling factor depending on the true distribution).

In this thesis, we focus our attention on the setting of log-concave densities on $\mathbb{R}$, as in Example 1.0.3 with $h(y) = e^y$ and $d = 1$. We construct a likelihood ratio test for testing the location of the mode or to be inverted to form confidence intervals. To form the test, we need to understand both the constrained log-concave MLE when the location of the mode is fixed and known as well as the unconstrained MLE over all log-concave densities. The unconstrained MLE has been well-studied already, see Section 1.1.1, so the majority of this thesis will be concerned with understanding the constrained MLE. We expect (i) that our methodology for inference about the mode of a unimodal density can be applied to problems in which the model is misspecified, i.e. concavity does not hold, and (ii) that our theoretical results will be useful in a variety of different concavity-based problems, such as the examples listed above.

1.1 Review: Shape-Constrained Estimation

Grenander (1956) initiated shape-constrained estimation by studying the (nonparametric) maximum likelihood estimator (which coincides with the least-squares estimator) of a monotone decreasing density, now generally known as the "Grenander estimator." Brunk (1958), Ayer et al. (1955), and Eeden (1956a,b) continued the field. The former estimated a monotone regression function and the latter two estimated sequences of probabilities known to be monotonically ordered. The limiting distribution for the Grenander estimator was found by Prakasa Rao (1969) and the limiting distribution for monotone regression was studied in Brunk (1970) and in Wright (1981). See Barlow et al. (1972) and Robertson et al. (1988) for the use of ordering constraints in a wide range of settings.

Hildreth (1954) studied the least-squares estimator of a concave regression function, motivated by *a priori* assumptions based on economic theory. Hanson and Pledger (1976) showed uniform consistency on compacta of the least-squares estimate of a concave regression function, and Mammen (1991) established the $n^{2/5}$ rate of convergence for estimation at a fixed point. It appears that Anevski (1994, 2003), motivated by data on bird migration, was the first to estimate a convex density function, which he did via maximum likelihood. The actual limiting distribution for estimation at a fixed point in the cases of the maximum likelihood estimator (MLE) and the least-squares estimator for a convex, decreasing density, and for the least-squares estimator for a convex or concave regression function were found in Groeneboom et al. (2001a,b), assuming the second derivative exists and is nonzero. To describe the limiting distribution, we first define an "invelope" process for Brownian motion, as studied in Groeneboom et al. (2001b). Let $Y(t) = \int_0^t W(t)\,dt - t^4$ where W is Brownian motion. Then H is the invelope if H satisfies

$$\begin{aligned} &H(t) \le Y(t) \text{ for all } t, \\ &\int (H(t) - Y(t))dH^{(3)}(t) = 0, \\ &H^{(2)}(t) \text{ is concave}, \end{aligned} \tag{1.1}$$

where $H^{(i)}$ is the ith derivative of H.[1] The non-Gaussian limiting distributions for each of the convex estimators described above is then a constant times $H^{(2)}(0)$. The constant depends on the second derivative of the estimand at the fixed point and, in the case of both density estimators, it depends also on the height of the density at the fixed point. These estimators were shown to be rate-minimax in Jongbloed (1995) (see also Groeneboom et al. (2001b)). The study of k-monotone densities, which are a generalization of monotone (1-monotone) and convex, decreasing (2-monotone) densities, is underway in Balabdaoui and Wellner (2007) and Balabdaoui and Wellner (2010).

1.1.1 *Log-Concavity*

The class of log-concave densities, denoted by $\mathcal{P}$ and defined by

$$\mathcal{P} := \left\{ f : \mathbb{R} \to \mathbb{R} \Big| \int f d\lambda = 1,\ f = e^{\varphi},\ \text{ and } \varphi \text{ is concave} \right\},$$

[1]We will also sometimes use the notation H' or H'' to signify the first or second derivatives, respectively, of a function H.

where λ is Lebesgue measure, is mathematically and statistically interesting. One of the most classic results on log-concavity is that of Ibragimov (1956), who proved that a density is in $\mathcal{P}$ if and only if it is strongly unimodal, i.e. its convolution with any unimodal density is again unimodal.

The work of Schoenberg (1951) relates log-concave functions to the classes of totally positive functions and Pólya frequency functions, showing that a function is log-concave if and only if it is a Pólya frequency function of order 2. For definitions and more on totally positive functions and Pólya frequency functions of all orders, Karlin (1968) is a good reference. Karlin (1968) showed additionally that $f_0 \in \mathcal{P}$ if and only if the family $f_0(x-\theta)$ for $\theta \in \mathbb{R}$ has monotone likelihood ratio, i.e. if and only if $f_0(x-\theta_1)/f_0(x-\theta_2)$ is monotone in x. Additionally, knowing $f_0 \in \mathcal{P}$ allows us to conclude that other functions possess some shape restriction. It is known that $f_0 \in \mathcal{P}$ implies that $F_0(x)$ and $1-F_0(x)$ are log-concave, where $F_0(x) = \int_{-\infty}^{x} f_0(u)du$ is the cumulative distribution function (cdf) of f_0 and $1-F_0$ is the survival function. We can also conclude that the hazard rate $f_0(x)/(1-F_0(x))$ is nondecreasing and the reverse hazard rate $f_0(x)/F_0(x)$ is nonincreasing. These results are given in, e.g., Propositions B.8 and B.8.a, pages 101-102, of Marshall and Olkin (2007). Log-concavity is also closed under taking weak limits, i.e. if F_n, for $n = 1, 2, \ldots$, are distributions with log-concave densities and $F_n \to_d F_0$, then F_0 has a log-concave density (Dharmadhikari and Joag-Dev, 1988).

"Convex geometric analysis" (or "analytical convex geometry", perhaps) is an area of research that ties together analysis and convex geometry, and yields results about log-concave densities as well as measures known as log-concave measures. One of the classical results in the area is known as the Brunn-Minkowski inequality, which can be interpreted as saying

> the function giving the nth root of the volumes of parallel hyperplane sections of an $n+1$-dimensional convex body is concave (Gardner, 2002, page 361).

The Prékopa-Leindler inequality is a generalization of the Brunn-Minkowski inequality and was originally shown in Prékopa (1971, 1973) and Leindler (1972). The Prékopa-Leindler inequality yields many results related to log-concave densities. For instance, it can be used to provide a proof of the above-stated result that log-concave densities have log-concave cdfs and survival functions. This follows from the fact that log-concave densities have "log-concave measures," see, e.g. page 378 of Gardner (2002). The Prékopa-Leindler inequality also shows that marginals of multi-dimensional log-concave functions are log-concave, i.e. if $f_0 \colon \mathbb{R}^{m+n} \to \mathbb{R}$ is log-concave then $x \mapsto \int_{\mathbb{R}^n} f_0(x,y)dy$ is log-concave. This can be used to see that convolutions of log-concave functions are log-concave (a fact which also follows from Ibragimov (1956)); see page 372–373 of Gardner (2002). The Prékopa-Leindler inequality has been further generalized by the Borell-Brascamp-Lieb inequality, proved in Theorem 3.3 of Brascamp and Lieb (1976) and also Theorem 3.1 of Borell (1975) (for more on the historical development, see page 375 of Gardner (2002)).

Brascamp and Lieb (1976) provide a moment inequality which was generalized by Hargé (2004): if we let $f_0 \colon \mathbb{R}^d \to \mathbb{R}$ be a log-concave density, take $Y \sim f_0\phi/(\int (f_0\phi)d\lambda)$ where ϕ is a normal density on $\mathbb{R}^d$ and λ is Lebesgue measure, let g be any convex function on $\mathbb{R}^d$, and let $X \sim \phi$ on $\mathbb{R}^d$, then we can conclude

$$Eg(Y - E(Y)) \leq Eg(X - E(X)).$$

Another area in which log-concavity finds uses is in results about "peakedness." In one dimension, a random variable Y is said to be more peaked about a point μ than a random variable X, which is written $Y \overset{p}{\geq} X$, if for all $\lambda > 0$,

$$P(|Y - \mu| \leq \lambda) \geq P(|X - \mu| \leq \lambda).$$

Now let us take $X_1, \ldots, X_n$ and $Y_1, \ldots Y_n$ to be drawn from log-concave and symmetric densities and to be independent. Then Sherman (1955) showed that $Y_i \overset{p}{\geq} X_i$ for $i = 1, \ldots, n$, implies that for any $c_i \in \mathbb{R}$,

$$\sum_{i=1}^{n} c_i Y_i \overset{p}{\geq} \sum_{i=1}^{n} c_i X_i.$$

Next we take a and b to be elements of R_+^n and assume that a *majorizes* b, i.e. $b \in \text{conv}\{\Pi a\}$ where conv denotes the convex hull, Π is the permutation group on n elements, and $\Pi a = \{\pi a | \pi \in \Pi\}$. Then with X_i as above but now also assumed to all come from an identical density, Proschan (1965) showed

$$\sum_{i=1}^{n} b_i X_i \overset{p}{\geq} \sum_{i=1}^{n} a_i X_i.$$

One application of this result is related to understanding the distribution of the average, $\overline{X}_n = \sum_i^n X_i/n$. As explained in the introduction of Proschan (1965), we know that $\overline{X}_n$ cannot deviate much from the population average, but we do not know, *a priori*, that the probability of a given deviation is decreasing in n. The above result on peakedness can be used to conclude that the probability of $\overline{X}_n$ deviating from its mean a given amount is strictly increasing as n increases, whenever the underlying distribution has a symmetric, log-concave density.

We have seen above that log-concavity has been well studied from the analytic and probabilistic viewpoint. Log-concavity has also enjoyed a recent resurgence in the statistical literature. Walther (2002) used a semiparametric model based on log-concavity to test in a multi-scale fashion whether a distribution was a mixture. Part of the motivation for this method is the idea of using log-concavity as a surrogate for unimodality since the class of log-concave densities is contained in the class of unimodal ones. The MLE of a unimodal density does not exist but the MLE of a log-concave density does.

The log-concave class is easily defined on $\mathbb{R}^d$ for $d \in \mathbb{N}$ since concavity (of, say, sets) is definable on any vector space; this is in contrast to the situation for other classes, such as those based on monotonicity. There is no universal definition of "monotonicity" of functions on $\mathbb{R}^d$. Cule et al. (2010) showed the existence of the log-concave MLE on $\mathbb{R}^d$ and defined and implemented an algorithm for computing the estimate. Cule and Samworth (2010) studied the log-concave MLE in $\mathbb{R}^d$ under model misspecification and Dümbgen et al. (2011) computed regression estimates in a setting in which the errors are specified to be log-concave.

Tails of log-concave densities are, by definition, sub-exponential. Thus the family excludes heavy-tailed distributions, such as Student's t_d distributions or Snedecor's F_{d_1,d_2} distributions. It does however includes the class of normal densities. It also includes the Gamma(α, β) for $\alpha \geq 1$, all uniform densities, and Beta(α, β) for $\alpha, \beta \geq 1$. Thus the log-concave family contains many familiar distributions and can thus be used as a stand-in or a generalization for them. Because of the lightness of log-concave tails, related classes have been explored by several authors. Koenker and Mizera (2010) studied r-concavity, a generalization of log-concavity allowing for polynomially decreasing tails, and Seregin and Wellner (2010) continued this study, showing the existence and consistency of the r-concave MLE in $\mathbb{R}^d$. Papers exploring recent statistical developments using log-concavity include Bagnoli and Bergstrom (2005) and Walther (2009). The former paper includes several possible applications of log-concavity to fields including statistics, economics, and political science, as well as a table of common log-concave distributions.

Walther (2002) showed the existence and Pal et al. (2007a) showed the consistency of the MLE over $\mathcal{P}$ (which, recall, is also known as the class of Pólya frequency functions of order 2). Dümbgen and Rufibach (2009) gave uniform consistency results including rates of convergence on compact sets, and Balabdaoui et al. (2009) found the limiting distribution of the log-concave MLE at a fixed point by showing that, up to a constant, the limiting distribution again depends on $H''(0)$, where H is the invelope defined in (1.1). In addition, they showed that over $\mathcal{P}$ the minimax convergence rate for estimation of a density's mode is the same as the rate computed by Has'minskii (1979) in the larger unimodal setting, namely $n^{1/5}$. Balabdaoui et al. (2009) found the limiting distribution of the mode of the log-concave MLE, and along the way showed that it is asymptotically rate-minimax. In particular, if f_0 is twice continuously differentiable at $m(f_0)$ and $f_0''(m(f_0)) < 0$, where $m(f)$ is (the infimum of) the mode of a function f, then letting $\hat{f}_n$ denote the log-concave MLE of an i.i.d. sample of size n from f_0, they showed

$$n^{1/5}(m(\hat{f}_n) - m(f_0)) \to_d K_{f_0} \operatorname{argmax}_t H^{(2)}(t), \tag{1.2}$$

where $K_{f_0} = \left(\frac{(4!)^2 f_0(m(f_0))}{f_0''(m(f_0))^2} \right)^{(1/5)}$ and H is the same invelope function defined above in (1.1).

Romano (1988b) gave a minimax lower bound for estimating the mode of a density over certain classes of functions, which are not necessarily unimodal but do have a unique maximizer, known to have $p \geq 2$ bounded derivatives in a neighborhood of the mode. The bound on the rate of convergence is then $n^{(p-1)/(2p+1)}$; note that this agrees with the rate of $n^{1/5}$ given by Has'minskii (1979) when $p = 2$. The results of Romano (1988b) do not directly apply to the case wherein there is a cusp at the mode. The cusp setting is very analogous to having "$p = 1$", in which case we expect rates of $n^{1/3}$ to be possible for estimation of the mode, as well as for the height of the density at its mode (see (1.7) and (1.8) in the next section). Log-concave density estimates have been shown to be adaptive to the underlying Hölder-continuity by Rufibach (2006). This gives us hope that mode estimates based on the log-concave MLE will achieve rates of up to $n^{1/3}$ if the underlying estimator has a cusp-shaped peak at $m(f_0)$.

1.2 Review: The Mode

Estimating location is one of the most classical problems in statistics. Staudte and Sheather (1990) defined a functional $\mu(F_X)$ as a measure of location if

(i). for all $a, b \in \mathbb{R}$, $\mu(aX + b) = a\mu(X) + b$

(ii). $X \geq 0$ implies $\mu(X) \geq 0$,

where X has a distribution function F_X which is assumed to belong to a family of absolutely continuous distribution functions. This definition of location parameters includes the (possibly trimmed) mean, linear combinations of two quantiles (such as the median), and the mode. Bickel and Frühwirth (2006) used this definition and study (robust) estimators of the mode as a measure of location. A mode of a density f is a possibly non-unique point m such that for all $x \in \mathbb{R}$,

$$f(m) \geq f(x). \tag{1.3}$$

To "estimate the mode" one first has to define a specific functional to be estimated, such as

$$m(f) := \inf \{t | f(t) \geq f(x) \text{ for all } x \in \mathbb{R}\}, \tag{1.4}$$

where f is a density function. Whether one defines m with an inf or a sup or something else is usually not particularly important. Note that we will abuse notation by allowing m to refer to either the functional or to its value at a specific f when f is clear from context. The above-defined functional m is well-defined for any density f, and thus any density estimate automatically yields an estimate of the mode (functional). The classical paper of Parzen (1962) estimates the density f_0 from which i.i.d. observations $X_1, \ldots, X_n$ are drawn via a kernel density estimator, given by

$$f_n(t) := \frac{1}{na_n} \sum_{i=1}^{n} K\left(\frac{t - X_i}{a_n}\right), \tag{1.5}$$

for some choice of kernel, K, usually taken to be a density, and some bandwidth sequence a_n, with $a_n \to 0$ and $na_n \to \infty$. Under conditions which imply at least 2-times everywhere-continuous-differentiability on both the density and kernel, Parzen (1962) showed the mode estimate is asymptotically normal. However, the asymptotic bias was forced to go to 0 faster than the asymptotic variance resulting in a non-optimal mean-squared error. Eddy (1980) continued the work of Parzen (1962), by modifying the asymptotics to allow the variance and bias to be balanced to give an optimal mean-squared error rate of $n^{2/7}$ and a Gaussian limiting distribution for $n^{2/7}(m(f_n) - m(f_0))$. To do this, it is assumed that the true density, f_0, is 4-times continuously differentiable everywhere, and the bandwidth is chosen appropriately, in addition to requirements on the user-chosen kernel. Romano (1988b) studied asymptotics of mode estimates based on KDEs, but allowed data-dependent bandwidths so that, for instance, the mode and KDE can be scale equivariant. Romano (1988b) also made his assumptions on f_0 local to the mode (in contrast to the global assumptions of Parzen (1962) and Eddy (1980)). Beyond assuming uniqueness of the mode and making some differentiability and moment assumptions on the kernel, Romano (1988b) assumed f_0 has three continuous derivatives in a neighborhood of $m(f_0)$ and that $f_0^{(3)}(m(f_0)) < 0$. When these hold, we can conclude that $n^{2/7}(m(f_n) - m(f_0))$ converges to a Gaussian distribution, with mean and variance computed in the paper, which depend on $f_0(m(f_0))$, $f_0^{(2)}(m(f_0))$, and $f_0^{(3)}(m(f_0))$, as well as on the kernel used.

Romano (1988b) also found minimax rate lower bounds, mentioned previously, for estimating the mode over classes with varying levels of differentiability. That is, he showed

$$\limsup_{n\to\infty} \inf_{T_n} \sup_{f\in N_n(\epsilon,p,f_0)} n^{(p-1)/(2p+1)} E_f \left| T_n - m(f) \right| > 0$$

where $N_n(\epsilon, p, f_0)$ are classes of functions local to f_0 all having some unique-mode assumptions, having $p \geq 2$ bounded derivatives in a neighborhood of $m(f_0)$, and having $f_0''(m(f_0)) < 0$. This shows that the mode estimator given by the KDE is minimax rate optimal for $p = 3$. The paper did not show additional rate results for KDEs of densities with higher-order derivative assumptions. For $p = 2$ the minimax rate was computed to be $n^{1/5}$ which agrees with the rate found by Has'minskii (1979) for estimation of the mode over classes with a second derivative assumption. Both Romano (1988b) and Has'minskii (1979) require that f_0 have a unique maximizer, which Has'minskii (1979) refers to as "unimodal," in contrast to our definition of unimodality (which is that f_0 is nondecreasing on $(-\infty, m(f_0)]$ and nonincreasing on $[m(f_0), \infty)$).

Chernoff (1964) studied the "naïve estimate" of the mode which is based on using a uniform kernel with some bandwidth choice, a, to estimate a density which has a strictly negative second derivative at the mode. Chernoff (1964) took $\hat{x}_a$ to be, say, the midpoint of an interval of length a with the most observations. Letting $\tilde{x}_a$ be

$$\arg\max_x \int_{-\infty}^{\infty} \frac{1}{a} 1_{[0,1]} \left(\frac{x-y}{a} \right) f_0(y) dy,$$

Chernoff (1964) studied $n^{1/3} \left(\hat{x}_a - \tilde{x}_a \right)$, ignoring the bias term. Due to the non-differentiability of the uniform kernel at the endpoints of its support, the asymptotic results of Parzen (1962) do not apply. A similar approach to Chernoff (1964) was taken by Venter (1967), who advocated picking an order-statistic bandwidth, $r \in \mathbb{N}$, and then searching for the two data points closest together with r order-statistics between them. Letting θ_n be the midpoint of this interval (i.e. the average of the two aforementioned order-statistics), Venter (1967) studied $n^{1/3}(\theta_n - m(f_0))$ when f_0 has a cusp at the mode, i.e. for some $\gamma_i > 0$ for $i = 1, 2, 3$, we have

$$f_0(x) = \gamma_0 - \gamma_1 (x - m(f_0))_+ - \gamma_2 (m(f_0) - x)_+ + o(|x - m(f_0)|) \quad \text{as} \;\; x \to m(f_0),$$

where $(x)_+ = \max(0, x)$. In both Chernoff (1964) and Venter (1967), after normalizing by a constant depending on f_0 and on the bandwidth chosen, the object of interest was shown to converge in distribution to

$$\arg\max_z W(z) - z^2 \tag{1.6}$$

where $W(z)$ is a mean-zero Gaussian process on $\mathbb{R}$. In the case of Chernoff (1964), $W(z)$ is a Brownian motion. For Venter (1967), it has a more complicated covariance structure.

Grenander (1965) suggested an arithmetically computable estimate of a density's mode, as opposed to all the mode estimates beforehand which required an intermediate KDE. Letting $X_{(i)}$, for $i = 1, \ldots, n$, be the order statistics of the observed data, we define

$$T_1 = \sum_{r=1}^{n-k} \frac{1}{(X_{(r+k)} - X_{(r)})^p}$$

and

$$T_2 = \sum_{r=1}^{n-k} \frac{\frac{1}{2}X_{(r)} + \frac{1}{2}X_{(r+k)}}{(X_{(r+k)} - X_r)^p},$$

where $k \in \mathbb{N}$ and $p > 0$. Then Grenander's estimator is T_2/T_1. The idea is that for n large, this quantity should be close to $\int_{-\infty}^{\infty} x f_0^{p+1}(x)dx / \int_{-\infty}^{\infty} f_0^{p+1}(x)dx$, which, for p large, should be close to $m(f_0)$. This estimate was shown to be consistent in Grenander (1965) and the rate of convergence and Gaussian limit distribution were given Hall (1982) under the assumption of four continuous derivatives locally about the mode. This estimator then achieves the same rate as the kernel density estimator of Eddy (1980), namely $n^{2/7}$, but the assumptions are local instead of global. Grenander's estimator apparently does not adapt to higher levels of differentiability; however, a number of authors have recommended its use, such as Dalenius (1965), Ekblom (1972), and Andriano et al. (1978).

Herrmann and Ziegler (2004) studied how smoothness conditions local to $m(f_0)$ affect estimation of $f_0(m(f_0))$ and $m(f_0)$; previously, most authors had assumed at least two derivatives, excluding cusp behavior at the mode (except for Venter (1967), as described above). Herrmann and Ziegler (2004) defined constants ρ and $\tilde{\rho}$ such that $|f_0(m(f_0)) - f_0(y)| \leq C_1|m(f_0)-y|^{\tilde{\rho}}$ and $f_0(m(f_0))-f_0(y) \geq C_2|m(f_0)-y|^{\rho}$ for all y in some neighborhood of $m(f_0)$. Then, letting f_n be a KDE based on a bounded and symmetric kernel with compact support, they concluded that

$$\left(\frac{n}{\log n}\right)^{(\tilde{\rho}/\rho)/(2\tilde{\rho}+1)} (m(f_n) - m(f_0)) = O(1), \tag{1.7}$$

almost surely, whereas for estimating the height at $m(f_0)$ we have

$$\left(\frac{n}{\log n}\right)^{(\tilde{\rho}^2/\rho)/(2\tilde{\rho}+1)} (f_n(m(f_n)) - f(m(f_0))) = O(1), \tag{1.8}$$

almost surely. Note that $\tilde{\rho} \leq \rho$ since for x and y close enough we must have $|x-y|^{\rho} < \frac{C_1}{C_2}|x-y|^{\tilde{\rho}}$, and if k_{nv} is the first non-vanishing derivative of f, then $\tilde{\rho} = k_{nv} = \rho$; taking $\tilde{\rho} = \rho = 2$ gives $n^{1/5}$ as the rate for estimating $m(f_0)$ and $n^{2/5}$ as the rate for estimating $f_0(m(f_0))$. When we take $\tilde{\rho} = \rho = 1$ we get $n^{1/3}$ which is between $n^{1/5}$ and $n^{2/5}$. Thus, this shows that sharper peaks are good for estimating $m(f_0)$ but bad for estimating $f_0(m(f_0))$ when we only have local smoothness assumptions. In fact, if f_0 looks like a translate of $-|x|^{\rho}$ locally at the mode, then $\tilde{\rho} = \rho$ and as $\rho \to 0$, i.e. the spikiness increases arbitrarily,

$(n/\log n)^{(\tilde{\rho}/\rho)/(2\tilde{\rho}+1)} \to n/\log n$. Other previously-quoted results for pointwise estimation do not have a log term in the rate; this may have to do with the fact that this is an almost sure result instead of an in-probability result. Note that these asymptotics are not directly comparable to the minimax calculations of Romano (1988b), because in the latter, f_0 was assumed to have 2 or more derivatives. Thus, in that setting, even as the total amount of (local) differentiability increases, the (non) spikiness stays the same.

As the number of derivatives increase (but the number vanishing at the mode stays constant), the rate for Eddy (1980) approaches $n^{1/2}$. Thus, more smoothness gives better estimates of the mode instead of worse ones, as opposed to the spirit of Herrmann and Ziegler (2004); however, Herrmann and Ziegler (2004) (and Romano (1988b) and Hall (1982)) are based on local smoothness conditions, whereas the smoothness conditions for Eddy (1980) are global conditions which are very stringent. Herrmann and Ziegler (2004) explained that these global conditions are only actually needed for controlling the "deterministic part" of estimating $f_0(x)$, i.e. for controlling $\sup_x |E(f_n(x)) - f_0(x)|$, which is actually not needed for estimating the mode. On the bottom of page 634, Romano (1988b) noted that as more odd-order derivatives are 0, the rate of convergence to $m(f_0)$ increases; however, this is not necessarily a decrease in spikiness since even order derivatives apparently can remain non-zero.

All of the mode estimates mentioned in this section require some choice of bandwidth parameter (including the direct estimates of Grenander (1965) and Venter (1967)). In fact, in order to optimize the KDE of Romano (1988b) $f_0^{(3)}(m(f_0))$ must be estimated, and it has a different optimal bandwidth than the optimal bandwidth for estimation of $f_0(m(f_0))$. Additionally, as is apparent from the results above, all the asymptotics of the kernel-based estimates are heavily dependent on the choice of kernel. Using a discontinuous kernel may result in wildly different asymptotics than a kernel with several derivatives. Global smoothness assumptions may be unnecessary for asymptotics to hold, but local smoothness or cusp assumptions are still important. This means that forming confidence intervals based on limiting distribution asymptotics is complicated and often requires estimation of several parameters. Additionally, Romano (1988a) showed that a straightforward application of the bootstrap implemented by resampling from the kernel density estimate yields invalid results, although modified bootstrap techniques can be made to work. Thus, inference on the mode at this point is, when possible, not straightforward.

1.3 Likelihood Ratio Methods and the Testing Approach to Confidence Sets

Using (1.2), one could form asymptotic confidence intervals for the parameter $m(f_0)$. We do not yet have any closed form information about the distribution of the random component, $\arg\max_t H^{(2)}(t)$, but its distribution can be estimated via simulation. However, to form confidence intervals, the parameter K_{f_0} would need to be estimated. This includes estimating $f_0''(m(f_0))$, and the log-concave MLE does not provide an automatic estimate thereof. Another approach to forming a confidence interval for a parameter is via a likelihood ratio statistic. We constrain our class of log-concave distributions to be those for which the parameter of interest, in this case the mode, is a known value denoted as m and we denote this subclass as $\mathcal{P}_m$. Letting $L_n(f) = L(f; X_1, \ldots, X_n) = \prod_{i=1}^n f(X_i)$ denote the likelihood of the data, the log-likelihood-ratio statistic is

$$2 \log \lambda_{n,m} := 2 \log \left(\frac{\sup_{f \in \mathcal{P}} L_n(f)}{\sup_{f \in \mathcal{P}_m} L_n(f)} \right).$$

It is often the case that likelihood ratio statistics converge to parameter-free distributions, a fact often referred to as the "Wilks Phenomenon" after Wilks (1938) (see Fan et al. (2001, 2002), and, for a general geometric approach to likelihood ratio statistics and the Wilks Phenomenon, see Fan et al. (2000)). In the parametric case the limiting distribution is chi-square for a wide variety of problems, independent of the underlying true distribution. A parameter-free asymptotically α-level test function $\phi_\alpha : \mathbb{R} \to \{0, 1\}$ can then be constructed based on the limiting distribution of $2 \log \lambda_{n,m}$: we set $\phi_\alpha(m) \equiv \phi_\alpha(m; X_1, \ldots, X_n) = 1$, i.e. we reject the null hypothesis, if $2 \log \lambda_{n,m} \equiv 2 \log \lambda_{n,m}(X_1, \ldots, X_n)$ is too large. Then to compute a confidence set, we can "invert" the above test as a function from the space of models to $\{0, 1\}$. That is, an asymptotically α-level confidence set would be

$$\phi_\alpha^{-1}(0) = \{m | \phi_\alpha(m; X_1, \ldots, X_n) = 0\} \subset \mathbb{R},$$

all the mode values we do not reject at the α level. In the nonparametric shape constrained setting, a test of this nature was constructed in Banerjee and Wellner (2001). The authors worked in the setting of monotone estimation, specifically estimating under the interval censoring model. They followed the program explained above to compute α-level confidence intervals for the distribution function of interest at a fixed point. Namely, they estimated under a restricted submodel and found the limiting distribution of the likelihood ratio statistic, which indeed turned out to be parameter-free, as expected.

1.4 Overview of this Thesis

Thus, the goal for this thesis is to proceed as in Banerjee and Wellner (2001), replacing the interval censoring model with the log-concave density model as our full model, and fixing the location of the mode rather than the value of the distribution function at a point in our constrained submodel. The full model is $\mathcal{P}$ defined in (2.3), below, and the submodel is $\mathcal{P}_m$ defined in (2.4), below. In Chapter 2 we study finite sample properties of the MLE over $\mathcal{P}_m$. We show its existence and uniqueness, and give two characterizations of the estimator. We use these characterizations to relate the constrained estimator and the unconstrained estimator.

In Chapter 3 we begin the study of the asymptotics of the MLEs, with a focus on global results. In particular, we show the global consistency in Hellinger distance of the constrained MLE, which also implies some other consistency results. We then show that when the i.i.d. data come from any log-concave density, the global rate of convergence of the constrained or unconstrained MLE in Hellinger distance is (at least) $n^{2/5}$.

In Chapter 4 we continue studying the asymptotics of the MLEs. We find the minimax lower bound for estimation of the mode of a log-concave density when the density has a cusp to be $n^{1/3}$. This extends the lower bound results of Balabdaoui et al. (2009) (which, recall, gave a rate of $n^{2/5}$ when $f_0''(m(f_0)) < 0$) to a non-differentiable case. We then find local rates of convergence for estimating the density's value at the mode, and use these to show uniform tightness of the constrained MLE in a shrinking neighborhood around the mode.

These results will be used in Chapter 5, wherein we finally find the joint limiting distribution of our estimators at the mode. To do so we first study a limiting process that will govern the limiting distribution of our estimators. The (random part of the) limiting process for problems based on underlying concavity depends on

$$dX(t) = adW(t) - bt^2 dt,$$

where $W(t)$ is a standard Brownian motion and a and b are positive constants. That is, $dX(t)$ is (heuristically) white noise about a "canonical" concave function, $-bt^2$. As in Banerjee and Wellner (2001), we expect the limiting distribution of our estimator to be given by "estimating" a canonical concave function $-bt^2$, in the sense of minimizing an objective function, depending on the "data" $X(t)$, over the constrained subclass of concave functions whose arg max is fixed to be 0. The unconstrained limiting estimator is studied and characterized already by Balabdaoui et al. (2009): it is given by the invelope process defined above in (1.1), studied extensively by Groeneboom et al. (2001a,b). Thus, in Chapter 5 we study the modally-constrained limiting estimator, which we use to find the joint limiting distribution of our (finite sample) MLEs. Along the way, we develop needed topological results related to a specific Skorokhod topology. Finally, we discuss the limiting distribution of our likelihood ratio statistic and the Wilks phenomenon, as well as directions for future research.

Chapter 6 is focused on aspects related to computing our constrained estimator, likelihood ratio test, and confidence intervals. Based on the algorithm of Dümbgen et al. (2007) for computing the unconstrained MLE, we develop an algorithm for computing the constrained MLE. Related code is given in Appendix C. We then present simulation results related to inference about the mode, as well as inference about the mode in a handful of real data examples.

Chapter 2

FINITE SAMPLE THEORY

Likelihood-based estimation and inference in nonparametric shape-constrained estimation settings generally requires several steps that are taken for granted in parametric problems. The first is to discover if the likelihood is even bounded; there are settings, e.g. unimodal density estimation, wherein it is unbounded, and direct maximization of the likelihood is not possible (Birgé, 1997). If the likelihood is indeed bounded, we need to know if the likelihood is maximized at a unique point or if there are many MLEs. If there is a unique MLE, we can proceed and study it. However, again in contrast to parametric estimators and to some nonparametric estimators, we often do not have a closed-form solution for the MLE. In parametric problems, closed-form solutions often arise as solutions to the likelihood equation(s), the values such that the score function (the derivative of the log-likelihood) is zero. By the nature of shape-constrained problems, the maximizers are generally on the boundary of the parameter space, which entails the possibility of non-zero derivatives of the likelihood. Of course, the likelihood necessarily decreases away from its maximizer. Thus, instead of equality constraints in the likelihood equations, we have inequality constraints. These inequality constraints are often still a "characterization" of the MLE, i.e. they hold for a parameter if and only if it is the MLE, much in the way that in many parametric problems, a parameter is the MLE if and only if it solves the likelihood equations. Thus developing these characterizations is often one of the first major steps in understanding a given shape-constrained problem. Such characterizations seem to be fundamental in developing pointwise asymptotic results, as in Groeneboom et al. (2001a,b), Jankowski and Wellner

(2009), Dümbgen and Rufibach (2009), and Groeneboom et al. (2008a,b), just to name a few. Consistency results are less dependent on the characterizations; Seregin and Wellner (2010) (and its predecessor, Pal et al. (2007a)) for instance, did not develop characterizations before proving consistency results, although others have.

In the case of the unconstrained MLE (UMLE), we know of uniqueness and existence already by Walther (2002) and Pal et al. (2007a), and we already have two characterization theorems, given by Theorems 2.2 and 2.4 of Dümbgen and Rufibach (2009). Our first step towards understanding the likelihood ratio statistic is to develop analogous results for the constrained MLE (CMLE).

Since concave functions will play a central role in this thesis, we will specify terminology related to concave functions now. We will mostly be interested in *proper* and *closed* concave functions. Rockafellar (1970) defines proper (page 24) and closed (page 52) convex functions, and a concave function is proper or closed if its negative is a proper or closed convex function, respectively. We will follow the convention that all concave functions φ are defined on all of $\mathbb{R}$ and take the value $-\infty$ off of their *effective domains*, $\operatorname{dom}\varphi := \{x : \varphi(x) > -\infty\}$. These conventions are motivated in (Rockafellar, 1970, pp. 40). The class of concave functions we will generally be using, thus, is

$$\mathcal{C} := \{\varphi \colon \mathbb{R} \to [-\infty, \infty) | \varphi \text{ is a closed, proper concave function}\}. \tag{2.1}$$

We will also be considering the modally-constrained class,

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

The unconstrained class of log-concave densities is then

$$\mathcal{P} := \left\{ f\colon \mathbb{R} \to [0,\infty) \middle| \int f \, d\lambda = 1, \; f = e^{\varphi}, \;\; \text{and } \varphi \in \mathcal{C} \right\}, \tag{2.3}$$

where λ is Lebesgue measure, and the class of log-concave densities constrained to have $m \in \mathbb{R}$ lie in its modal interval is

$$\begin{aligned} \mathcal{P}_m &:= \left\{ f \in \mathcal{P} \middle| \, f(m) \geq f(x) \text{ for all } x \in \mathbb{R} \right\} \\ &= \left\{ f\colon \mathbb{R} \to [0,\infty) \middle| \int f \, d\lambda = 1, \; f = e^{\varphi}, \;\; \text{and } \varphi \in \mathcal{C}_m \right\}. \end{aligned} \tag{2.4}$$

We will let X_i, $i = 1, \ldots, n$, be the data observations, taken to be distinct, and we let $X_{(1)} < \cdots < X_{(i)} < \cdots < X_{(n)}$ be the order statistics. In this chapter on finite sample results, we will not assert anything about the distribution of the X_i, except that we will assume that the X_i are all distinct. In fact, uniqueness is unnecessary if we allow weights to be associated to each data point, but we will not need to pursue that here. In later chapters, when we are studying (well-specified) probabilistic results, we will assert that X_i are i.i.d. from some log-concave density. We will also make use of "modally-augmented data," Z_i, $i = 1, \ldots, n_1$, whose definition depends on whether the mode is or is not a data point. (Note that once we assume the data are drawn from a density, with probability 1 the mode will not be a data point, but to form confidence intervals we will need to allow the possibility that the mode is a data point.) If the mode m is a data point define k to be the index such that $X_{(k)} = m$, and define $Z_i := X_{(i)}$, and then let $n_1 = n$. If m is not a data point then let k be the index in $\{1, \ldots, n+1\}$ such that $m \in (X_{(k-1)}, X_{(k)})$, where for the purposes of the above definition we let $X_{(0)}$ be $-\infty$ and let $X_{(n+1)}$ be ∞. Then for $i = 1, \ldots, k-1$, let $Z_i := X_{(i)}$ and for $i = k, \ldots, n+1$ let $Z_i := X_{(i-1)}$, and let $n_1 = n+1$. We let $\mathbb{P}_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ denote the empirical measure of the data and $\mathbb{F}_n(x) := \frac{1}{n}\sum_{i=1}^n 1_{(-\infty,x]}(X_i)$ denote the empirical cumulative distribution function (cdf) of the data. We let $\mathcal{K}$ be the $\varphi \in \mathcal{C}$ such that e^{φ} is a *density*, i.e. $\mathcal{K} := \log \circ \mathcal{P}$. Similarly, we let $\mathcal{K}_m$ denote the class of piecewise linear concave functions φ such that e^{φ} is a *density* with mode at m, i.e. $\mathcal{K}_m := \log \circ \mathcal{P}_m$. Let $\mathcal{C}_{n,m}$ denote the (random) class of piecewise linear concave functions φ with knots at the Z_i and $\operatorname{dom} \varphi = [Z_1, Z_{n_1}]$, and let $\mathcal{K}_{n,m}$ denote the class of concave functions φ with knots at the Z_i and where e^{φ} is a density and $\operatorname{dom} \varphi = [Z_1, Z_{n_1}]$. Here, by "piecewise linear," we mean that the functions are piecewise linear and continuous on the interior of their domain, (Z_1, Z_{n_1}).

We define the ("log-likelihood") objective functional $\Psi_n \colon \mathcal{C}_m \to \mathbb{R}$ to be

$$\Psi_n(\varphi) = \frac{1}{n} \sum_{i=1}^{n} \varphi(X_i) - \int_{\mathbb{R}} e^{\varphi(x)} dx.$$

We let $\hat{\varphi}_n$ be the UMLE, i.e. the maximizer of Ψ_n over $\mathcal{K}$. It exists and is unique for all $n \geq 2$ by Theorem 1 of Cule et al. (2010) or Proposition 1 of Walther (2002) (see also Theorem 2.14 of Seregin and Wellner (2010)). Our first goal will be to show an analogous result for the class $\mathcal{K}_m$.

Definition 2.0.1. A *cone* C is a subset of a real vector space V such that for all $c \in C$ and $\lambda \geq 0$ we have $\lambda c \in C$. C is a *convex* cone if it is a cone and a convex set, i.e. if $c_1, c_2 \in C$ then for any $\lambda \in [0,1]$, $\lambda c_1 + (1-\lambda) c_2 \in C$. Taking V to be finite dimensional, we say that C is *(finitely) generated* by a set $b_i \in C$, $i = 1, \ldots, k < \infty$ if for all $c \in C$ we can write $c = \sum_{i=1}^{k} \alpha_i b_i$ for some nonnegative numbers $\alpha_i \geq 0$.

Fact 2.0.2. *$\mathcal{C}_{n,m}$ is a convex cone with finite generating set given by*

$$\{(x - Z_i)_-\}_{2 \leq i \leq k} \bigcup \{(Z_i - x)_-\}_{k \leq i \leq n_1 - 1} \bigcup \{\pm 1\},$$

where $(x)_- = \min(x, 0)$ *and* $(x)_+ = \max(x, 0)$.

Proof. It is clear that $\mathcal{C}_n$ is a convex cone, and then it is also clear that $\mathcal{C}_{n,m}$ is a convex cone because the mode will be preserved under positive scaling. To see that $\mathcal{C}_{n,m}$ is generated by the given set, for $\varphi \in \mathcal{C}_{n,m}$ let a_i for $i = 2, \ldots, k$ and b_i, $i = k, \ldots, n_1 - 1$ be given by $\varphi'(Z_i-) =: \sum_{j=i}^k a_j$ and $\varphi'(Z_i+) =: -\sum_{j=k}^i b_j$. Let $C := \varphi(m)$. Then $\varphi(x) = C + \sum_{i=2}^k a_i(x - Z_i)_- + \sum_{i=k}^{n-1} b_i(Z_i - x)_-$. □

In the next proof we will identify piecewise linear functions φ with their values at the vector $Z = (Z_1, \ldots, Z_{n_1})$. Thus, for a function f defined at each of the components of $x \in \mathbb{R}^N$, $N \in \mathbb{N}$, we define $\mathrm{ev}_x f$ by $\mathrm{ev}_x f = (f(x_1), \ldots, f(x_N))$.

Lemma 2.0.3. For $n \geq 2$, the MLE $\hat{f}_n^0 = e^{\hat{\varphi}_n^0}$ over $\mathcal{K}_m$ exists, is unique, and $\hat{\varphi}_n^0$ is piecewise linear with knots at the Z_i's and domain equal to $[Z_1, Z_{n_1}]$. If m is not a data point then at least one of $(\hat{\varphi}_n^0)'(m+)$ or $(\hat{\varphi}_n^0)'(m-)$ is 0. Further $\hat{\varphi}_n^0$ is the unique maximizer of Ψ_n over $\mathcal{C}_m$.

Proof. This argument is similar to Theorems 2.1 and 2.2 from Rufibach (2006). It is clear that the solution is piecewise linear with knots at the Z_i's. This follows from comparing any two concave functions with the same values at the data points, because the line between $\hat{\varphi}_n^0(Z_i)$ and $\hat{\varphi}_n^0(Z_{i+1})$ falls below any other concave function with the same values. Analogously it follows that $\hat{\varphi}_n^0$ is flat either directly to the left of the mode or directly to the right as long as the mode is not a data point. The domain of $\hat{\varphi}_n^0$ is identically equal to the set $[Z_1, Z_{(n_1)}]$ because if it were smaller the log-likelihood would be $-\infty$ and it does not increase $\int \hat{\varphi}_n^0 \, d\mathbb{F}_n$ to make it larger. That the MLE exists is shown as in Rufibach (2006), by showing that as the norm of φ, thought of as a vector in $\mathbb{R}^{n_1}$, goes to infinity so does $\Psi_n(\varphi)$. Then we let $v := \mathrm{ev}_Z \varphi = (\varphi(Z_1), \ldots, \varphi(Z_{n_1}))$. We will write $\Psi_n(v)$ and $\Psi_n(\varphi)$ interchangeably. Now note that Ψ_n, thought of as a function on $\mathbb{R}^{n_1}$ (possibly ignoring one of its arguments) is continuous. The cone $\mathcal{C}_{n,m}$ is a closed set because pointwise limits of concave functions are concave, and the mode of the limit will still be m. Thus if we can look at Ψ_n on a bounded subset of $\mathcal{K}_{n,m}$, then that subset is compact and continuous functions achieve minima on compact sets, so the MLE exists. The method is to now show that as the distance from the origin to v goes to infinity so does $\Psi_n(v)$. Take a sequence v_j, $j = 1, \ldots,$ which has limit coordinates of γ_i, $i = 1, \ldots, n$, which might be $\pm\infty$. If one of the limiting coordinates is negative infinity then we are done. If one of them is positive infinity then the kth coordinate (corresponding to the mode) is infinite. But if m is distinct from the $X_{(i)}$ and $v_k = \varphi(m)$ goes to infinity then, by concavity, the other coordinates must go to negative infinity, in order to have a total integral of 1. So again, we are done.

(If the mode is equal to one of the $X_{(i)}$, then the proof in Rufibach (2006) holds verbatim. It will be repeated here for completeness. We consider a sequence of concave φ_k with $\int e^{\varphi_k(x)}dx = 1$ and $\varphi_k(X_{(j)}) \to \infty$. We can then say

$$
\begin{aligned}
1 &\geq \int_{X_{(j-1)}}^{X_{(j)}} \exp(\varphi_k(x))dx \\
&= (X_{(j)} - X_{(j-1)})e^{\varphi_k(X_j)} \frac{1 - e^{-\delta_k}}{\delta_k} \\
&\geq (X_{(j)} - X_{(j-1)})e^{\varphi_k(X_j)} \frac{1}{1 + \delta_k}
\end{aligned}
$$

where $\delta_k = \varphi_k(X_{(j)}) - \varphi_k(X_{(j-1)})$ and we use $\frac{1-e^{-x}}{x} \geq \frac{1}{1+x}$ for $x \geq 0$, which is equivalent to saying $e^x \geq 1 + x$ for $x \geq 0$. Thus we have for δ_k the bound $\delta_k \geq (X_{(j)} - X_{(j-1)})e^{\varphi_k(X_j)} - 1$ so that

$$
-\varphi_k(X_{(j)}) - \varphi_k(X_{(j-1)}) = -2\varphi_k(X_{(j)}) + \delta_k \geq -2\varphi_k(X_{(j)}) + (X_{(j)} - X_{(j-1)})e^{\varphi_k(X_j)} - 1,
$$

and certainly $-c_1 y + c_2 e^y \longrightarrow \infty$ as $y \to \infty$, so certainly the likelihood goes to $-\infty$ if $\varphi_k(X_{(j)}) \to \infty$. Thus we are done if $j \geq 2$, and an analogous proof holds if $j = 1$, so we are done.)

Finally, If $\hat{\varphi}$ maximizes Ψ_n over $\mathcal{K}_m$ then it maximizes Ψ_n over $\mathcal{C}_m$ because for $t \in \mathbb{R}$ and

$\varphi \in \mathcal{K}_m$, $\Psi_n(\varphi(x)+t) = \int (\varphi(x)+t)\, d\mathbb{F}_n(x) - \int e^{\varphi(x)+t} dx$ and this equals

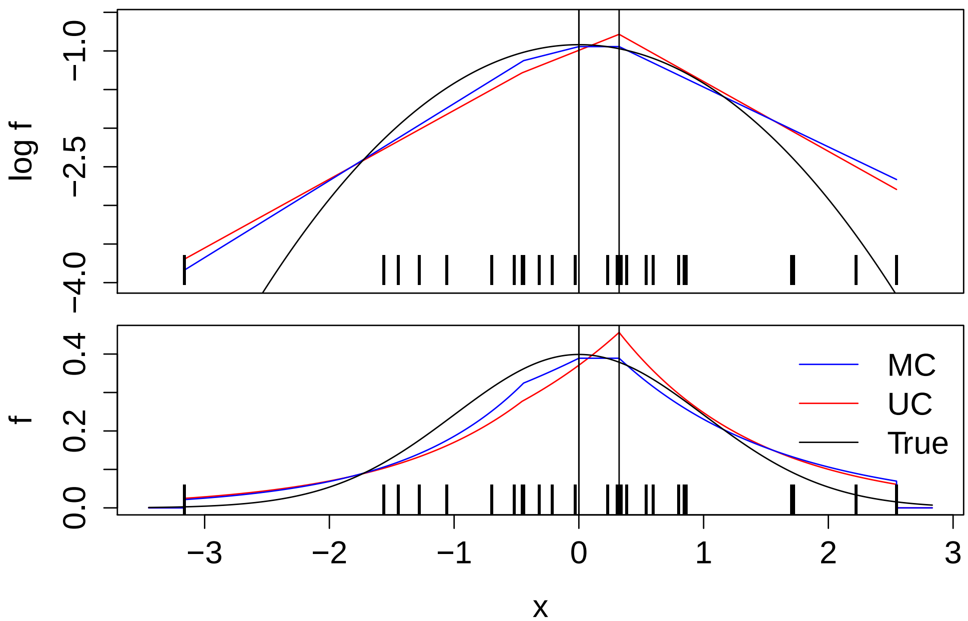
$$\int \varphi(x)\, d\mathbb{F}_n(x) + t - 1 + 1 - e^t 1 = \Psi_n(\varphi) - e^t + t + 1 < \Psi_n(\varphi),$$

since $e^t > 1 + t$ for all $t \in \mathbb{R}$. Uniqueness follows because Ψ is strictly concave, which follows from the linearity of $\varphi \mapsto \int \varphi(x)\, d\mathbb{F}_n(x)$ and the strict concavity of $x \mapsto -e^x$. □

We can compute both the CMLE and the UMLE via active set algorithms, which will be described in Chapter 6. Figure 2.1 shows the results of applying the algorithms to 25 data points drawn from a Normal$(0, 1)$ distribution with density $e^{-x^2/2}/\sqrt{2\pi}$ and log-density $-\log(\sqrt{2\pi}) - x^2/2$. We plot both the UMLE and the CMLE with mode constrained to be 0, the true mode, and we also plot the true underlying functions. Both estimators of the log-density are piecewise linear concave functions with knots at certain data points, and, for the CMLE, an LK at the mode.

Viewing m as a (possible) knot of the estimator $\hat{\varphi}_n^0$, we will not classify m as simply "a knot" but as a "left knot" (LK) or as a "right knot" (RK) or as "not a knot" (NK). We say m is an RK if $(\hat{\varphi}_n^0)'(m+) < 0$, we say m is an LK if $(\hat{\varphi}_n^0)'(m-) > 0$, and we say m is an NK if $(\hat{\varphi}_n^0)'(m) = 0$. All other knots are considered to be both LKs and RKs. Note that if m is not a data point it cannot be both an RK and an LK. If m is a data point, then it is both an RK and an LK if and only if $\hat{\varphi}_n^0$ coincides with the unconstrained MLE $\hat{\varphi}_n$. This follows from the characterization of the $\hat{\varphi}_n$ given in Dümbgen and Rufibach (2009). We denote the number of knots as l^0, where we count the mode m as a knot regardless of whether it is or is not. We let the knots have Z-indices $1 = j_1, \ldots, j_{l^0} = n_1$, and let $1 \le p \le l^0$ be such that $j_p = k$, i.e. $Z_{j_p} = m$. Now that we have our CMLE, $\hat{\varphi}_n^0$, in hand, our next step is to provide a characterization of it. This next result follows mainly from the definition of $\hat{\varphi}_n^0$ as a maximizer of the log-likelihood objective function over $\mathcal{P}_m$.

Figure 2.1: Densities f and $\log f$: CMLEs (MC), UMLEs (UC), and True underlying $N(0,1)$-based functions from which 25 i.i.d. data points were sampled. Vertical black lines at the true mode 0 and at the mode of the UMLE.



Theorem 2.0.4. (Characterization 0) *For a function $\hat{\varphi}_n^0 \in \mathcal{K}_m$, $\hat{\varphi}_n^0$ is the MLE (over $\mathcal{K}_m$) if and only if*

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

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

Proof. First, we know by Lemma 2.0.3 that $\hat{\varphi}_n^0$ is the maximizer over $\mathcal{C}_m$ if and only if it is the maximizer over the class of functions in $\mathcal{C}_m$ that have domain equal to $[Z_1, Z_{n_1}]$, i.e. are bounded on $[Z_1, Z_{n_1}]$. We will show that $\hat{\varphi}_n^0$ maximizes Ψ_n if and only if (2.5) holds for all Δ such that $\hat{\varphi}_n^0 + t\Delta \in \mathcal{C}_m$ and Δ is bounded on $[Z_1, Z_{n_1}]$. This then shows the statement of the theorem.

Now, we differentiate Ψ_n at a bounded function $\phi \in \mathcal{C}_m$ in the direction of a bounded Δ

to see

$$\begin{aligned} D_\Delta \Psi_n(\phi) &= \lim_{\epsilon \to 0+} \frac{1}{\epsilon} \left(\Psi_n(\phi + \epsilon\Delta) - \Psi_n(\phi) \right) \\ &= \int \frac{\partial}{\partial \epsilon} \left(\phi(x) + \epsilon \Delta(x) \right) |_{\epsilon=0} \, d\mathbb{F}_n(x) - \int \frac{\partial}{\partial \epsilon} (e^{\phi(x)+\epsilon\Delta(x)})|_{\epsilon=0} \, dx \\ &= \int \Delta(x) d\mathbb{F}_n(x) - \int \Delta(x) e^{\phi(x)} dx \end{aligned}$$

which holds for any Δ such that $\phi + \epsilon\Delta \in \mathcal{C}_m$ and $\phi + \epsilon\Delta$ is bounded, for some positive ϵ. Note that in the second equality, we can differentiate under the integral sign. This is because, for all $\epsilon < \delta$, $|\frac{\partial}{\partial \epsilon} e^{\hat{\varphi}_n^0(x)+\epsilon\Delta(x)}| = |e^{\hat{\varphi}_n^0(x)+\epsilon\Delta(x)} \Delta(x)|$, and this is bounded above by $e^{\hat{\varphi}_n^0(x)+\delta M} M$, which is integrable. Thus, the mean value theorem and dominated convergence theorem show that we can pass the derivative through the integral.

Now as noted above, Ψ_n is concave because $x \mapsto -e^x$ is. For all bounded $\Delta \in \mathcal{C}_m$ the above derivative is necessarily ≤ 0 at $\hat{\varphi}_n^0$ since $\hat{\varphi}_n^0$ maximizes Ψ_n, which shows the necessity of (2.5). The condition (2.5) is also sufficient because a sufficient condition for maximizing a convex function is that its derivative, as defined above, is everywhere nonpositive. □

We will now use Characterization 0 to find a new characterization of the MLE, which is analogous to the following characterization in the unconstrained case.

Theorem 2.0.5. (Theorem 2.4, (Dümbgen and Rufibach, 2009)) *Let $\hat{\varphi}_n$ be a concave function such that it is linear on all intervals $[X_{(i)}, X_{(i+1)}]$ for $1 \leq i \leq n-1$ and it is $-\infty$ on $\mathbb{R} \backslash [X_{(1)}, X_{(n)}]$. Define $\hat{F}_n(x) := \int_{-\infty}^x e^{\hat{\varphi}_n(u)} du$, $\hat{H}_n(x) = \int_{-\infty}^x \hat{F}_n(u) du$, and $\mathbb{Y}_n(x) = \int_{-\infty}^x \mathbb{F}_n(u) du$, and assume that $\hat{F}_n(X_{(n)}) = 1$. Then $\hat{\varphi}_n$ is the MLE if and only if for all $t \in [X_{(1)}, X_{(n)}]$,*

$$\hat{H}_n(t) \leq \mathbb{Y}_n(t), \tag{2.6}$$

and there is equality at the knots of $\hat{\varphi}_n$.

The analogous version below is similar, but has conditions for each side of the mode.

Theorem 2.0.6. (Characterization 1) *Let $\hat{\varphi}_n^0$ be piecewise linear with knots at the data points, let $m \in \mathbb{R}$, and assume that $\hat{\varphi}_n^0$ achieves its maximum at m. Let $\hat{f}_n^0 = e^{\hat{\varphi}_n^0}$ and for $t \in \mathbb{R}$ let*

$$\begin{aligned} \mathbb{F}_{n,L}(t) &= \textstyle\int_{(-\infty,t]} d\mathbb{F}_n(u) & \mathbb{F}_{n,R}(t) &= \textstyle\int_{[t,\infty)} d\mathbb{F}_n(u), \\ \mathbb{Y}_{n,L}(t) &= \textstyle\int_{X_{(1)}}^t \mathbb{F}_{n,L}(u)du & \mathbb{Y}_{n,R}(t) &= \textstyle\int_t^{X_{(n)}} \mathbb{F}_{n,R}(u)du, \\ \hat{F}_{n,L}^0(t) &= \textstyle\int_{-\infty}^t \hat{f}_n^0(u)du & \hat{F}_{n,R}^0(t) &= \textstyle\int_t^\infty \hat{f}_n^0(u)du \\ \hat{H}_{n,L}^0(t) &= \textstyle\int_{X_{(1)}}^t \hat{F}_{n,L}^0(u)du & \hat{H}_{n,R}^0(t) &= \textstyle\int_t^{X_{(n)}} \hat{F}_{n,R}^0(u)du. \end{aligned}$$

Then $\hat{\varphi}_n^0$ is the MLE of $\varphi \in \mathcal{K}_m$ if and only if

$$\mathbb{Y}_{n,L}(t) := \int_{X_{(1)}}^t \mathbb{F}_{n,L}(x)dx \geq \int_{X_{(1)}}^t \hat{F}_{n,L}^0(x)dx =: \hat{H}_{n,L}^0(t) \text{ for } X_{(1)} \leq t \leq m \tag{2.7}$$

and

$$\mathbb{Y}_{n,R}(t) := \int_t^{X_{(n)}} \mathbb{F}_{n,R}(x)dx \geq \int_t^{X_{(n)}} \hat{F}_{n,R}^0(x)dx =: \hat{H}_{n,R}^0(t) \text{ for } X_{(n)} \geq t \geq m \tag{2.8}$$

with equality in (2.7) if t is an LK and with equality in (2.8) if t is an RK. Note that if $m < X_{(1)}$ or $m > X_{(n)}$, (2.7) or (2.8) is an empty condition, respectively.

We can rewrite the above conditions as

$$\int_b^t \mathbb{F}_{n,L}(x)dx \geq \int_b^t \hat{F}_{n,L}^0(x)dx \quad \text{for } X_{(1)} \leq b \leq t \leq c \leq m, \tag{2.9}$$

and

$$\int_t^c \mathbb{F}_{n,R}(x)dx \geq \int_t^c \hat{F}_{n,R}^0(x)dx \quad \text{for } m \leq b \leq t \leq c \leq X_{(n)}, \tag{2.10}$$

where b and c are either knots or equal to m and whenever t is an LK or an RK in (2.9) or (2.10), respectively, we have equality.

The mode m can only be both LK and RK if m is a data point and m is the mode of the unconstrained log-concave estimator (which is an event with probability 0 if m is the true mode of f).

Proof. First we assume $\hat{\varphi}_n^0$ is the MLE and use (2.5) to show that (2.7) and (2.8) hold via integration by parts. For $X_{(1)} \le t \le m$, we choose $\Delta(x) = (x-t)_-$ (which is concave with m as a mode). Then for F equal to either $\mathbb{F}_n$ or $\hat{F}_n^0$, integration by parts yields (see, e.g., page 108, exercise 34, of Folland (1999))

$$\int_{[X_{(1)},t]} (x-t)_- \, dF(x) = (t-t)_- F(t+) - (X_{(1)}-t)F(X_{(1)}-) - \int_{[X_{(1)},t]} F(x)\, dx.$$

Since $\mathbb{F}_n(X_{(1)}-)$ and $\hat{F}_n^0(X_{(1)}-)$ are both 0, this yields $\int_{[X_{(1)},t]}(x-t)_- \, dF(x) = -\int_{[X_{(1)},t]} F(x)\, dx$. Thus, by our initial characterization (2.5), we get $-\int_{X_{(1)}}^t \mathbb{F}_n(x)\, dx \le -\int_{X_{(1)}}^t \hat{F}_n^0(x)\, dx$ which is (2.7) (since $\hat{F}_n^0 = \hat{F}_{n,L}^0$ and $\mathbb{F}_n = \mathbb{F}_{n,L}$).

Similarly, for $X_{(n)} \ge t \ge m$, let $\Delta(x) = (t-x)_-$; this yields

$$\begin{aligned}\left(t - X_{(n)}\right) \mathbb{F}_n(X_{(n)}+) - \int_t^{X_{(n)}} \mathbb{F}_n(x) d(-x) &= \int (t-x)_- \, d\mathbb{F}_n(x) \\ &\le \int (t-x)_- \, d\hat{F}_n^0(x) \\ &= \left(t - X_{(n)}\right) \hat{F}_n^0(X_{(n)}+) - \int_t^{X_{(n)}} \hat{F}_n^0(x) d(-x),\end{aligned}$$

and, recalling that we have already shown $\hat{F}_n^0(X_{(n)}) = 1$, this is equivalent to

$$-\int_t^{X_{(n)}} \mathbb{F}_n(x) d(-x) \le -\int_t^{X_{(n)}} \hat{F}_n^0(x) d(-x),$$

so we have (2.8). We get equality at some knot points also: set $\Delta(x) = (x-b)_+$ where $X_{(n)} \ge b \ge m$ is any RK. Then, by the definition of an RK, Δ is an allowable perturbation because $\hat{\varphi}_n^0(b+) - \hat{\varphi}_n^0(b-) > 0$ so for some δ small enough, $\hat{\varphi}_n^0 + \delta\Delta$ is still concave with mode at m. Using this Δ we get

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

to be bounded above by

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

so that $\int_b^{X_{(n)}} \mathbb{F}_n(x)dx \geq \int_b^{X_{(n)}} \hat{F}_n^0(x)dx$, and thus for any $X_{(n)} \geq b \geq m$ that is an RK we have the inequality both ways, $\int_b^{X_{(n)}} \mathbb{F}_n(x)dx = \int_b^{X_{(n)}} \hat{F}_n^0(x)dx$. (Note that, for instance, if the slope to the left of m is 0, then $\hat{\varphi}_n^0(x) + \epsilon\,(m-x)_+$ does not have m as a mode for any ϵ. This is why we divide into the cases of LK and of RK). An analogous argument holds for the $X_{(1)} \leq c \leq m$ that are LKs. We have thus shown that (2.7) and (2.8) hold with the appropriate equalities, so are done with this implication.

Now we will show the reverse implication. We assume (2.7) and (2.8) hold and consider Δ with domain equal to $[Z_{(1)}, Z_{(n_1)}]$ and that are piecewise linear with knots at the Z_i. This is justified by the proof of Theorem 2.0.4, or, rather, by the fact that a sufficient condition for $\hat{\varphi}_n^0$ to maximize Ψ_n over a convex set $\mathcal{S}$ is that the derivative in all directions Δ such that $\hat{\varphi}_n^0 + \epsilon\Delta \in \mathcal{S}$ for some $\epsilon > 0$ is nonpositive. By Lemma 2.0.3, we can take $\mathcal{S}$ to be the subset of $\mathcal{C}_m$ consisting of only functions that have domain equal to $[Z_{(1)}, Z_{(n_1)}]$ and are piecewise linear with knots at the Z_i. If we maximize over $\mathcal{S}$ then we maximize over $\mathcal{C}_m$. Note that if $\hat{\varphi}_n^0 + \epsilon\Delta$ is piecewise linear with knots at the Z_i and with domain equal to $[Z_{(1)}, Z_{(n_1)}]$, then Δ is piecewise linear with knots at the Z_i and domain containing $[Z_{(1)}, Z_{(n_1)}]$, and we can restrict Δ to have domain equal to $[Z_{(1)}, Z_{(n_1)}]$ with no loss of generality. We also need $\hat{\varphi}_n^0 + \epsilon\Delta$ to be concave with mode m.

Such Δ are not necessarily concave because they can have positive changes in slope at knots of $\hat{\varphi}_n^0$. Between two knots such Δ must have only negative changes. Recall we have defined the indices $1 = j_1, \ldots, j_{l^0} = n_1$ so that Z_{j_i} are the knots and p is defined such that there are $p-1$ knots strictly less than the mode. The key to the proof is to write

$$\begin{aligned}\Delta'(r-) = \sum_{i=2}^{p} &\left[C_i 1_{(Z_{j_{i-1}}, Z_{j_i}]}(r) + \sum_{j=j_{i-1}+1}^{j_i} \beta_j 1_{(Z_{j_{i-1}}, Z_j]}(r) \right] \\ + \sum_{i=p}^{l^0-1} &\left[D_i 1_{(Z_{j_i}, Z_{j_{i+1}}]}(r) + \sum_{j=j_i}^{j_{i+1}-1} \alpha_j 1_{(Z_j, Z_{j_{i+1}}]}(r) \right]\end{aligned} \tag{2.11}$$

with $\beta_j \geq 0$, $\alpha_j \leq 0$, $C_i \in \mathbb{R}$, and $D_i \in \mathbb{R}$. See Remark 2.0.7 below. We will consider only Δ that are 0 on $[m, \infty)$, i.e. we argue only on the left side of the mode. We can argue on the right side symmetrically, and we can shift by a constant because $\mathbb{F}_n(\infty) = \hat{F}_n^0(\infty) = 1$. Now, note that if m is not an LK then $C_p \geq 0$ (which refers to the value of $\Delta'(r-)$ on the interval $(Z_{j_p - 1}, m]$ since $m = Z_{j_p}$). If m is an LK then we have $\int_{Z_{j_{p-1}}}^{m} (\mathbb{F}_n(x) - \hat{F}_n^0(x))\, dx = 0$ by hypothesis. Either way, we have

$$C_p \int_{Z_{j_{p-1}}}^{m} (\mathbb{F}_n(x) - \hat{F}_n^0(x))\, dx \geq 0. \tag{2.12}$$

We thus have

$$\begin{aligned}\int \Delta d\mathbb{F}_n &= \Delta(m) - \sum_{i=2}^{p} \left[C_i \int_{Z_{j_{i-1}}}^{Z_{j_i}} \mathbb{F}_n(x)\, dx + \sum_{j=j_{i-1}+1}^{j_i} \beta_j \int_{Z_{j_{i-1}}}^{Z_j} \mathbb{F}_n(x)\, dx \right] \\ &\leq \Delta(m) - \sum_{i=2}^{p} \left[C_i \int_{Z_{j_{i-1}}}^{Z_{j_i}} \hat{F}_n^0(x)\, dx + \sum_{j=j_{i-1}+1}^{j_i} \beta_j \int_{Z_{j_{i-1}}}^{Z_j} \hat{F}_n^0(x)\, dx \right] \\ &= \int \Delta d\hat{F}_n^0\end{aligned}$$

as desired, where we only need $i \leq p$ since we consider Δ that is 0 on $[m, \infty)$, and where the inequality follows from noting $-\beta_j \int_{Z_{j_{i-1}}}^{Z_j} \mathbb{F}_n(x)\, dx \leq -\beta_j \int_{Z_{j_{i-1}}}^{Z_j} \hat{F}_n^0(x)\, dx$ by assumption and because $\beta_j \geq 0$. We also have $-C_i \int_{Z_{j_i-1}}^{Z_{j_i}} \mathbb{F}_n(x)\,dx = -C_i \int_{Z_{j_i-1}}^{Z_{j_i}} \hat{F}_n^0(x)\,dx$ for all i except for $i = p$, by the equality-at-knots assumption, and for $i = p$ we have (2.12). We can thus conclude by Theorem 2.0.4 that $\hat{\varphi}_n^0$ is the CMLE.

It is clear that (2.9) and (2.10) are equivalent to (2.7) and (2.8) (with all corresponding equalities), respectively. For instance, $\int_{X_{(1)}}^t \mathbb{F}_{n,L}(x)dx \geq \int_{X_{(1)}}^t \hat{F}_{n,L}^0(x)dx$ for all $X_{(1)} \leq t \leq m$, with equality when t is an LK, if and only if

$$\int_{X_{(1)}}^b \mathbb{F}_{n,L}(x)\,dx + \int_b^t \mathbb{F}_{n,L}(x)\,dx \geq \int_{X_{(1)}}^b \hat{F}_{n,L}^0(x)\,dx + \int_b^t \hat{F}_{n,L}^0(x)\,dx$$

holds for all $X_{(1)} \leq t \leq m$ and all knots b, with equality when t is an LK, and this is equivalent to (2.9).

If m is both an RK and an LK then m is simply a knot and $\hat{\varphi}_n^0$ coincides with the unconstrained MLE by the characterization of the unconstrained MLE in Dümbgen and Rufibach (2009). □

Remark 2.0.7. To see that the representation (2.11) holds, we only need to consider Δ that are concave between any two knots of $\hat{\varphi}_n^0$ because $\hat{\varphi}_n^0$ is linear between two knots. That is, such a Δ satisfies

$$\Delta'(r-) = \sum_{j=2}^{n_1} a_j 1_{[Z_{j-1} < r \leq Z_j]}.$$

Recall $1 = j_1, \ldots, j_{l^0} = n_1$ are indices of the knots, i.e. Z_{j_i} is a knot, and $Z_{j_p} = m$. Now, we will define for $2 \leq i \leq p$,

$$C_i := (a_{j_i})_-$$

$$\beta_{j_i} := (a_{j_i})_+$$

$$\beta_j := a_j - a_{j+1} \text{ for } j = j_{i-1}, \ldots, j_i - 1,$$

and for $p \leq i \leq l^0 - 1$,

$$\begin{aligned} D_i &:= (a_{j_i})_+ \\ \alpha_{j_i} &:= (a_{j_i})_- \\ \alpha_j &:= a_{j+1} - a_j \text{ for } j = j_i + 1, \ldots, j_{i+1}. \end{aligned}$$

Note that all α's are nonpositive, all β's are nonnegative, and for $2 \leq i \leq p$ and $\alpha \leq j_p$

$$\sum_{\gamma=\alpha}^{j_i - 1} \beta_\gamma = a_\alpha - a_{\alpha+1} + ... + a_{j_i - 1} - a_{j_i} = a_\alpha - a_{j_i},$$

and for $p \leq i \leq l^0 - 1$ and $\beta \geq j_p$

$$\sum_{\gamma=j_i+1}^{\beta-1} \alpha_\gamma = a_{\beta+1} - a_\beta + \ldots + a_{j_i+1} - a_{j_i} = a_\beta - a_{j_i},$$

and for $2 \leq i \leq p$, $C_i + \beta_{j_i} = a_{j_i}$, and for $p \leq i \leq l^0 - 1$, $D_i + \alpha_{j_i} = a_{j_i}$. Then we can see that if $Z_{\gamma-1} < r \leq Z_\gamma$ for $j_{i-1} < \gamma \leq j_i$ with $2 \leq i \leq p$, then

$$C_i + \sum_{j=\gamma}^{j_i} \beta_j = C_i + \beta_{j_i} + \alpha_\gamma - a_{j_i} = a_\gamma = \Delta'(r-),$$

as desired. Similarly, if $Z_{\gamma-1} < r \leq Z_\gamma$, and $j_{i-1} < \gamma \leq j_i$ then

$$D_i + \sum_{j=j_i}^{\gamma-1} \alpha_j = a_\gamma,$$

as desired. This shows that we can represent $\Delta'(r-)$ by

$$\Delta'(r-) = \sum_{i=2}^{p} \left[C_i 1_{(Z_{j_{i-1}}, Z_{j_i}]}(r) + \sum_{j=j_{i-1}+1}^{j_i} \beta_j 1_{(Z_{j_{i-1}}, Z_j]}(r) \right] + \sum_{i=p}^{l^0-1} \left[D_i 1_{(Z_{j_i}, Z_{j_{i+1}}]}(r) + \sum_{j=j_i}^{j_{i+1}-1} \alpha_j 1_{(Z_j, Z_{j_{i+1}}]}(r) \right],$$

as desired.

Remark 2.0.8. We will make an explanatory remark on notation here. In the characterization for our constrained estimator, we have left-side processes and right-side processes. Thus, because the proofs for the left-side and right-side processes are entirely symmetric, we choose to keep the notation for the two sides symmetric. For instance, despite the fact that $\mathbb{F}_{n,L} = \mathbb{F}_n$, when we are discussing the (characterization of the) CMLE, we will refer to this function as $\mathbb{F}_{n,L}$. On the other hand, when discussing the UMLE we will denote this function by $\mathbb{F}_n$ and, similarly, we will denote $\int_{-\infty}^{x} \mathbb{F}_n(u)du$ by $\mathbb{Y}_n(x)$, since left-side and right-side notation is irrelevant for the UMLE.

Note that (2.9) and (2.10) entail that for any Z_i that is an LK and an RK,

$$\int_{Z_i}^{t} \left(\mathbb{F}_n(u) - \hat{F}_n^0(u) \right) du \geq 0, \tag{2.13}$$

for all t on the same side of m as Z_i is, i.e. the display holds for all $X_{(1)} \leq t \leq m$ if $Z_i \leq m$, and it holds for all $m \leq t \leq X_{(n)}$ if $m \leq Z_i$. For $Z_i \leq t \leq m$ or $m \leq t \leq Z_i$, this is precisely what (2.9) and (2.10) say, respectively. If $X_{(1)} \leq t \leq Z_i \leq m$, then by (2.9) $\int_{X_{(1)}}^{Z_i} \left(\mathbb{F}_n(u) - \hat{F}_n^0(u) \right) du = 0$, and thus

$$0 \leq \int_{X_{(1)}}^{t} \left(\mathbb{F}_n(u) - \hat{F}_n^0(u) \right) du = - \int_{t}^{Z_i} \left(\mathbb{F}_n(u) - \hat{F}_n^0(u) \right) du.$$

Similarly, for $m \le Z_i \le t \le X_{(n)}$, by (2.10), $\int_{Z_i}^{X_{(n)}} \left(\mathbb{F}_{n,R}(u) - \hat{F}^0_{n,R}(u) \right) du = 0$, so

$$0 \le \int_t^{X_{(n)}} \left(\mathbb{F}_{n,R}(u) - \hat{F}^0_{n,R}(u) \right) du = -\int_{Z_i}^t \left(\mathbb{F}_{n,R}(u) - \hat{F}^0_{n,R}(u) \right) du.$$

(Similarly, (2.10) says directly that for $m \le t \le Z_i$, $0 \le -\int_{Z_i}^t \left(\mathbb{F}_{n,R}(u) - \hat{F}^0_{n,R}(u) \right) du$.) Now, since $\mathbb{F}_{n,R}(u) = 1 - \mathbb{F}_n(u)$ and $\hat{F}^0_{n,R}(u) = 1 - \hat{F}^0_n(u)$, this shows that

$$0 \le \int_{Z_i}^t -\left((1 - \mathbb{F}_n(u)) - (1 - \hat{F}^0_n(u)) \right) du = \int_{Z_i}^t (\mathbb{F}_n(u) - \hat{F}^0_n(u))\, du,$$

so we have shown (2.13). Next we use the characterization to show that at knots, the estimator cdf and the empirical cdf are very close to each other. Recall that the only thing we have assumed about our data X_i are that they are unique (and that assumption could be relaxed, but the $1/n$ below would be modified). Thus, the following is not a probabilistic statement.

Corollary 2.0.9. *For any knot Z_i excluding $Z_k = m$, we have*

$$\hat{F}^0_n(Z_i) \in \left[\mathbb{F}_n(Z_i) - \frac{1}{n}, \mathbb{F}_n(Z_i) \right], \tag{2.14}$$

deterministically.

Proof. For $i \ne k$ take Z_i to be a knot, and define the function $D_i(z) := \int_{Z_i}^z \left(\mathbb{F}_n(x) - \hat{F}^0_n(x) \right) dx$ for $z \in \mathbb{R}$. Then by (2.13), for z in a neighborhood of Z_i we have $D_i(z) \ge 0$ and $D_i(Z_i) = 0$, so Z_i is a local minimum of D_i. So using the Fundamental Theorem of Calculus (for right-and left-continuous integrands), we see

$$\begin{aligned}\mathbb{F}_n(Z_i+) - \hat{F}_n^0(Z_i+) &= \lim_{h\downarrow 0} \frac{\int_{Z_i}^{Z_i+h}\left(\mathbb{F}_n(x) - \hat{F}_n^0(x)\right)dx}{h} \geq 0 \\ \hat{F}_n^0(Z_i-) - \mathbb{F}_n(Z_i-) &= \lim_{h\downarrow 0} \frac{-\int_{Z_i-h}^{Z_i}\left(\mathbb{F}_n(x) - \hat{F}_n^0(x)\right)dx}{h} \geq 0.\end{aligned} \tag{2.15}$$

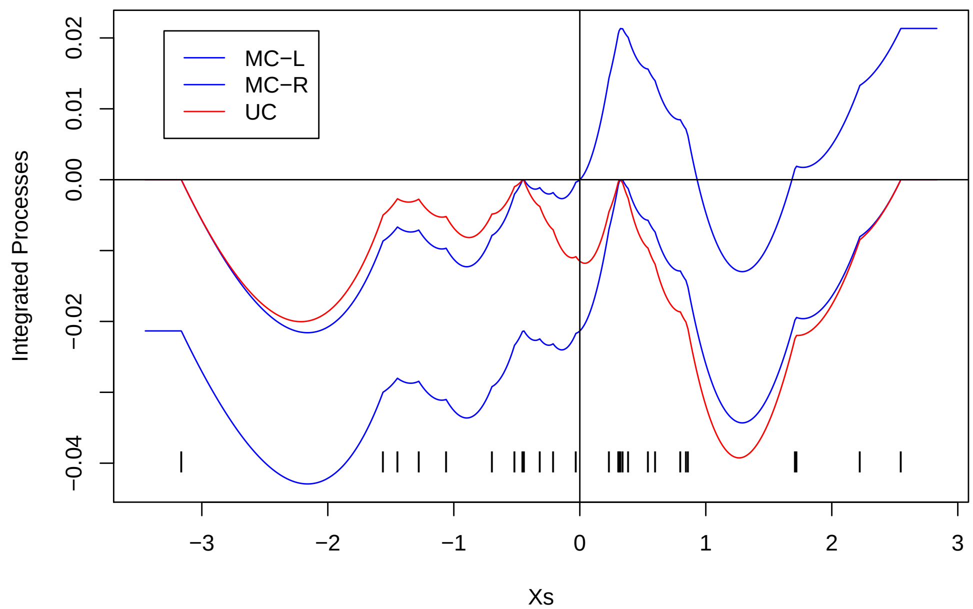
Since $\hat{F}_n^0$ is continuous and $\mathbb{F}_n(Z_i-) = \mathbb{F}_n(Z_i) - 1/n$, we have shown

$$\mathbb{F}_n(Z_i) \geq \hat{F}_n^0(Z_i) \geq \mathbb{F}_n(Z_i) - \frac{1}{n}.$$

□

We now will plot the processes from the preceding characterization theorem, as well as a corresponding process for the unconstrained estimator, for the same data and estimators in Figure 2.1. We plot $\int_{X_{(1)}}^t (\hat{F}_{n,L}^0(x) - \mathbb{F}_{n,L}(x))\,dx$, which is the left side modally-constrained process, labelled as "MC-L"; we plot $\int_t^{X_{(n)}} (\hat{F}_{n,R}^0(x) - \mathbb{F}_{n,R}(x))\,dx$, which is the right side modally-constrained process, labelled as "MC-R"; we also plot, $\int_{X_{(1)}}^t (\hat{F}_n(x) - \mathbb{F}_n(x))\,dx$, the unconstrained process, labelled as "UC". This process is used in Theorem 2.4 of Dümbgen and Rufibach (2009) to give a characterization of the UMLE. Recall that in the simulated data for this figure, the mode of the CMLE is constrained to be 0, which is also the true mode. We will refer to this as m here, rather than as 0, and, in this paragraph, when we say "below" we mean "not above". Notice that the left side process is always below 0 to the left of m, but that to the right of m the left side process is not necessarily below 0. The right side process is always below 0 to the right of m. In this case, it is also below 0 to the left of m, but that does not necessarily have to happen. Finally, the unconstrained process is below 0 everywhere. Notice that at the knots (visible in Figure 2.1), the corresponding processes touch 0. That is, the unconstrained process equals 0 at any knot of $\hat{\varphi}_n$. The left side process

Figure 2.2: Plot of processes at the once-integrated level: $\int_{X_{(1)}}^{t} (\hat{F}^0_{n,L}(x) - \mathbb{F}_{n,L}(x))\, dx$, the left side modally-constrained process (MC-L), $\int_{t}^{X_{(n)}} (\hat{F}^0_{n,R}(x) - \mathbb{F}_{n,R}(x))\, dx$, the right side modally-constrained process (MC-R), and $\int_{X_{(1)}}^{t} (\hat{F}_n(x) - \mathbb{F}_n(x))\, dx$, the unconstrained process (UC).



equals 0 at any knot of $\hat{\varphi}^0_n$ at or to the left of m and the right side process equals 0 at any knot of $\hat{\varphi}^0_n$ at or to the right of m. To more easily state the next results, we will denote the knots of the constrained estimator, $\hat{\varphi}^0_n$, as $\tau^0_{n,i}$, and the knots of the unconstrained estimator, $\hat{\varphi}_n$, as $\tau_{n,i}$. In both cases, we index such that if $i < 0$ then $\tau^0_{n,i} < m$ and $\tau_{n,i} < m$ and if $i > 0$ then $\tau^0_{n,i} > m$ and $\tau_{n,i} > m$. For the constrained estimator, we set $\tau^0_{n,0} = m$ with the caveat that this is either a one-sided knot or not a knot at all.

Proposition 2.0.10. *If we have $\tau_{n,i_1} < \tau^0_{n,i_2} < \tau_{n,i_3}$, or $\tau^0_{n,i_1} < \tau_{n,i_2} < \tau^0_{n,i_3}$ or $\tau^0_{n,i_1} = \tau_{n,i_2} < \tau^0_{n,i_3} = \tau_{n,i_4}$ and all indices are strictly less than 0 (i.e. the knots are on the same side of m), we can conclude there is a point $x \in (\tau_{n,i_1}, \tau_{n,i_3})$ such that $\hat{F}^0_n(x) - \hat{F}_n(x) = 0$. Similarly, if the above statement holds but all indices are strictly greater than 0 we can reach* *the same conclusion (there is a point* $x \in (\tau_{n,i_1}, \tau_{n,i_3})$ *such that* $\hat{F}_n^0(x) - \hat{F}_n(x) = 0$*).*

We can reach the same conclusion if m *is an LK and either of* $\tau^0_{n,i_1} < \tau_{n,i_2} < \tau^0_{n,i_3} = m$ *or* $\tau^0_{n,i_1} = \tau_{n,i_2} < \tau^0_{n,i_3} = \tau_{n,i_4} = m$ *holds, or if* m *is an RK and either of* $m = \tau^0_{n,i_1} < \tau_{n,i_2} < \tau^0_{n,i_3}$ *or* $m = \tau^0_{n,i_1} = \tau_{n,i_2} < \tau^0_{n,i_3} = \tau_{n,i_4}$ *holds.*

Proof. By Theorem 2.0.5 for the UMLE and by Theorem 2.0.6 for the CMLE, we can say that for $i < 0$,

$$\int_{-\infty}^{\tau_{n,i}} (\mathbb{F}_n(u) - \hat{F}_n^0(u))du \geq 0 = \int_{-\infty}^{\tau_{n,i}} (\mathbb{F}_n(u) - \hat{F}_n(u))du$$

and

$$\int_{-\infty}^{\tau^0_{n,i}} (\mathbb{F}_n(u) - \hat{F}_n(u))du \geq 0 = \int_{-\infty}^{\tau^0_{n,i}} (\mathbb{F}_n(u) - \hat{F}_n^0(u))du$$

Thus we conclude that for all $i < 0$ we have both

$$\int_{-\infty}^{\tau_{n,i}} (\hat{F}_n^0(u) - \hat{F}_n(u))du \leq 0$$

and

$$\int_{-\infty}^{\tau^0_{n,i}} (\hat{F}_n^0(u) - \hat{F}_n(u))du \geq 0.$$

Let $\hat{D}_n(x) = \hat{F}_n^0(u) - \hat{F}_n(u)$ and $\hat{C}_n(x) = \int_{-\infty}^{x} \hat{D}_n(u)du$. For $i_1, i_2, i_3 < 0$ consider $\tau_{n,i_1} < \tau^0_{n,i_2} < \tau_{n,i_3}$. Because $\hat{C}_n(\tau_{n,i_1}), \hat{C}_n(\tau_{n,i_3}) \leq 0$, $\hat{C}_n(\tau^0_{n,i_2}) \geq 0$, and the knots are distinct so that $\hat{C}_n$ crosses below 0 and then back above 0, the Intermediate Value Theorem implies there exist distinct points $\tau_{n,i_1} \leq x_1 < x_2 \leq \tau_{n,i_3}$ such that $\hat{C}_n(x_1) = \hat{C}_n(x_2) = 0$. Then, since $\hat{D}_n$ is continuous so by the Fundamental Theorem of Calculus $\hat{C}_n$ is differentiable, the Mean Value Theorem implies that there is a point $x_3 \in (x_1, x_2)$ such that $\hat{D}_n(x_3) = \frac{\hat{C}_n(x_2) - \hat{C}_n(x_1)}{x_2 - x_1} = 0$, i.e. there is a point $x_3 \in (\tau_{n,i_1}, \tau_{n,i_3})$ such that $\hat{F}_n^0(x_3) - \hat{F}_n(x_3) = 0$. Similarly, if we have distinct points $\tau^0_{n,i_1} < \tau_{n,i_2} < \tau^0_{n,i_3} < m$ or points $\tau^0_{n,i_1} = \tau_{n,i_2} < \tau^0_{n,i_3} = \tau_{n,i_4}$, we can conclude there is a point $x \in (\tau_{n,i_1}, \tau_{n,i_3})$ such that $\hat{F}_n^0(x) - \hat{F}_n(x) = 0$, because we will get distinct points at which $\hat{C}_n$ equals 0. □

Because consistency of our estimator will provide alternating sequences of knots for the constrained and unconstrained estimators, we will have sequences of points of equality for $\hat{D}_n = \hat{F}_n^0(u) - \hat{F}_n(u)$. Thus the following restatement of the Mean Value Theorem will be useful.

Proposition 2.0.11. *If $\hat{F}_n^0 - \hat{F}_n$ is 0 at both of $x_1 < x_2$ which are both points in $\operatorname{dom}\hat{\varphi}_n^0 = \operatorname{dom}\hat{\varphi}_n$, then there is a point $x \in (x_1, x_2)$ such that $\hat{f}_n^0(x) - \hat{f}_n(x) = 0$.*

Proof. This is by the Mean Value Theorem, like the previous proposition, because $\hat{F}_n^0(t) - \hat{F}_n(t) = \int_{-\infty}^{t} (\hat{f}_n^0(u) - \hat{f}_n(u)) du$. □

Chapter 3
ASYMPTOTICS I: GLOBAL

3.1 Consistency for the CMLE

The log-concave MLE on $\mathbb{R}$ was proved consistent by both Dümbgen and Rufibach (2009) and Pal et al. (2007a), the former proving consistency in total variation distance and the uniform norm and the latter proving consistency in the Hellinger metric, given by

$$H^2(g, f) := \frac{1}{2} \int \left(\sqrt{g(x)} - \sqrt{f(x)} \right)^2 dx. \tag{3.1}$$

Cule and Samworth (2010) and Seregin and Wellner (2010) prove consistency of the log-concave MLE on $\mathbb{R}^d$. Our proof follows the proofs of Pal et al. (2007a) and Seregin and Wellner (2010), so we begin by citing Theorem 3.1 from Pal et al. (2007a), which gives Hellinger-consistency of the MLE over any class of unimodal densities for which the modal height of the estimator grows slowly enough. Here is the theorem.

Theorem 3.1.1. *Let $\mathcal{U}$ be a class of unimodal densities for which the MLE exists and let $\hat{g}_n$ be the MLE for a sample of size n drawn from a density $f \in \mathcal{U}$. If*

$$\sup_x \log \hat{g}_n(x) = o\left(\frac{\sqrt{n}}{\log(n)} \right) \tag{3.2}$$

holds almost surely, then $H(\hat{g}_n, f) \longrightarrow_{a.s.} 0$ as $n \to \infty$.

Note that in the above theorem the hypothesis does not bound $\sup_x \log \hat{g}_n(x)$ from below. Thus to prove consistency of the unconstrained log-concave MLE, Pal et al. (2007a) show its mode is bounded above almost surely, and the same proof works for the constrained log-concave MLE. We will repeat it here. First we quote some results without proof. Here are Lemmas 3 and 4 on page 246 of Pal et al. (2007a).

Lemma 3.1.2 (Lemma 3, page 426, Pal et al. (2007a))**.** Let g be a log-concave density. Let $a, b \in \mathbb{R}$ and let $0 < g(a) < g(b)$. Then

$$g(b) \leq \frac{1}{|b-a|} \left(1 + \log(g(b)) - \log(g(a))\right). \tag{3.3}$$

Lemma 3.1.3 (Lemma 4, page 426, Pal et al. (2007a))**.** Let $a, b, x > 0$. If $x \leq a \log(x) + b$ then

$$x \leq 2a \log(2a) + 2b. \tag{3.4}$$

Next, we quote Lemma 3.4 from page 3764 of Seregin and Wellner (2010), where their "decreasing transformation" is $h(y) = e^{-y}$.

Lemma 3.1.4 (Lemma 3.4, page 3764, Seregin and Wellner (2010))**.** Let $\varphi(x)$ be concave such that $e^{\varphi(x)}$ is a density. Then

$$\int_{\mathbb{R}} |\varphi(x)| \, e^{\varphi(x)} \, du < \infty. \tag{3.5}$$

Now we prove the height of the CMLE is bounded. The proof is identical to the analogous proof for the UMLE from Pal et al. (2007a).

Theorem 3.1.5. *Let $\mathcal{S}$ be a submodel of the class of log-concave densities. We assume $X_i \sim_{iid} f_0 \in \mathcal{S}$, for $i = 1, \ldots, n$, and that the MLE of $f_0 \in \mathcal{S}$ exists, and we denote this MLE by $\hat{g}_n$. Then almost surely* $\sup_n \sup_x \log \hat{g}_n(x) < \infty$.

Proof. Let $m_n = \arg\max_x \hat{g}_n(x)$. Then we let q denote $[n/4] + 1$ if $m_n > X_{([n/2])}$ or $[3n/4]$ if $m_n \leq X_{([n/2])}$. Let K_n be q or $n - q$ if $m_n > X_{([n/2])}$ or $m_n \leq X_{([n/2])}$, respectively, so that $K_n \geq n/4$. Then from Lemma 3.1.2

$$\hat{g}_n(m_n) \leq \frac{1}{|m_n - X_{(q)}|}\left(1 + \log\left(\frac{\hat{g}_n(m_n)}{\hat{g}_n(X_{(q)})}\right)\right). \tag{3.6}$$

Define the log likelihood as $l_n(g) = \sum_{i=1}^n \log g(X_i)$. Then since $f_0 \in \mathcal{S}$,

$$l_n(f_0) \leq l_n(\hat{g}_n) \leq K_n \log(\hat{g}_n(X_{(q)})) + (n - K_n)\log(\hat{g}_n(m_n)). \tag{3.7}$$

Thus,

$$\begin{aligned} \log\left(\frac{\hat{g}_n(m_n)}{\hat{g}_n(X_{(q)})}\right) \leq & \frac{n}{K_n}\log(\hat{g}_n(m_n)) - \frac{1}{K_n} l_n(f_0) \\ \leq & 4\left(\log(\hat{g}_n(m_n)) - \frac{1}{n} l_n(f_0)\right). \end{aligned} \tag{3.8}$$

Combining (3.8) with (3.6), we see

$$\begin{aligned} \hat{g}_n(m_n) &\leq \frac{1}{|m_n - X_{(q)}|}\left(1 + 4\log(\hat{g}_n(m_n) - \frac{4}{n} l_n(f_0)\right) \\ &= A_n \log(\hat{g}_n(m_n)) + B_n, \end{aligned}$$

where

$$A_n = \frac{4}{|m_n - X_{(q)}|} \text{ and } B_n = \frac{1}{|m_n - X_{(q)}|}\left(1 - \frac{4}{n} l_n(f_0)\right).$$

Thus by Lemma 3.1.3

$$\hat{g}_n(m_n) \leq 2A_n \log(2A_n) + 2B_n.$$

By Lemma 3.1.4 and the strong law of large numbers $|\frac{1}{n} l_n(f_0)|$ is almost surely bounded. We have $|m_n - X_{(q)}| \geq |X_{([n/2])} - X_{(q)}|$, and because order statistics converge to appropriate quantiles and by the strict inequalities $F^{-1}(1/4) < F^{-1}(1/2) < F^{-1}(3/4)$, we conclude that $\sup_n A_n < \infty$ and $\sup_n B_n < \infty$ almost surely, and thus we are done. □

For our consistency results, we will make the following "null hypothesis" assumption.

Assumption A (Modally-Constrained Log-Concave Null Hypothesis)**.** *Let $f_0 \in \mathcal{P}_m$, and assume $X_1, \ldots, X_n$ are i.i.d. observations from f_0.*

Corollary 3.1.6. *Let Assumption A be satisfied. The MLE of $f_0 \in \mathcal{P}_m$, $\hat{f}_n^0$, is almost surely Hellinger-consistent for f_0.*

Proof. This follows from Theorem 3.1.1 and Theorem 3.1.5. □

Corollary 3.1.7. *Let Assumption A be satisfied and assume that f_0 is continuous. Then the MLE, $\hat{f}_n^0$, over $\mathcal{P}_m$ based on i.i.d. observations drawn from f_0 is consistent in the uniform norm, i.e.*

$$\sup_x |\hat{f}_n^0(x) - f_0(x)| \longrightarrow_{a.s.} 0 \tag{3.9}$$

and

$$\sup_x |\hat{F}_n^0(x) - F_0(x)| \longrightarrow_{a.s.} 0. \tag{3.10}$$

Proof. Recall the elementary inequality $d_{TV}(P,Q) \leq \sqrt{2}H(P,Q)$ where P and Q are measures and total variation distance is

$$d_{TV}(P,Q) = \frac{1}{2}\int_{\mathbb{R}} \left| \frac{dP}{d\mu} - \frac{dQ}{d\mu} \right| d\mu,$$

where $dP/d\mu$ and $dQ/d\mu$ are Radon-Nikodym derivatives with respect to some dominating measure μ. Also, recall the fact that $d_{TV}(P_n, Q) \to 0$ implies $P_n \to_d Q$, where $\to_d$ means convergence in distribution. See, e.g., pages 20-21 of DasGupta (2008). These facts allow us to conclude from Corollary 3.1.6 that $\hat{F}_n^0 \to_d F_0$ almost surely. Then Proposition 2 on page

4 of Cule and Samworth (2010) gives the result $\sup_x e^{|ax|}|\hat{f}_n^0(x) - f_0(x)| \longrightarrow_{a.s.} 0$ for certain values of $a > 0$, and this is much stronger than what we need to conclude that (3.9) holds. (3.10) also follows from consistency in total variation norm since for any $x \in \mathbb{R}$

$$|\hat{F}_n^0(x) - F_0(x)| \leq \int_{-\infty}^{x} |\hat{f}_n^0(u) - f_0(u)| du \leq 2 d_{TV}(\hat{f}_n^0, f_0).$$

In fact, convergence in Hellinger distance gives uniform consistency in more general settings than log-concavity. See the proof of Lemma 2.19 in Seregin and Wellner (2010) which shows this result for convex-transformed densities whose domain is $\mathbb{R}^d$ (and which satisfy certain regularity conditions). □

Now to prove uniform consistency on compacta (without assuming continuity of f_0), we first recall Lemma 3.14 from Seregin and Wellner (2010). The proof there shows the following slight rephrasing of the lemma.

Lemma 3.1.8. If φ_n and φ_0 are convex functions such that $\int_{\mathbb{R}} (e^{\varphi_n/2} - e^{\varphi_0/2})^2 \, d\lambda \to 0$, then $\varphi_n(x)$ converges to $\varphi_0(x)$ for all x in the interior of $\operatorname{dom} \varphi_0$. The convergence is uniform on compacta on the interior of $\operatorname{dom} \varphi_0$.

Proof. This is what is shown by the proof of Lemma 3.20 of Seregin and Wellner (2010), by setting $h(y) = e^{-y}$ and $g_n = \varphi_n$, $n = 0, \ldots, \infty$. □

Corollary 3.1.9. *Let Assumption A be satisfied. Let $K = [b, c]$ be contained in the interior of* $\operatorname{dom} \varphi_0$. *Then*

$$\sup_{u \in K} \left| \hat{\varphi}_n^0(u) - \varphi_0(u) \right| \longrightarrow 0 \quad \text{as } n \to \infty, \tag{3.11}$$

and also

$$\sup_{u \in K} \left| \hat{f}_n^0(u) - f_0(u) \right| \longrightarrow 0 \quad \text{as } n \to \infty, \tag{3.12}$$

almost surely.

Proof. (3.11) follows from almost sure Hellinger consistency of $\hat{f}_n^0$ (Corollary 3.1.6) and Lemma 3.1.8. Then (3.12) follows immediately from (3.11). □

Under the concavity shape constraint, consistency of the estimator implies consistency of its derivative by Groeneboom et al. (2001b). Recall that by Rockafellar (1970) both the positive and negative directional derivatives exist and are finite at all points in the interior of $\mathrm{dom}\,\varphi$ for a concave function φ.

Lemma 3.1.10 (Lemma 3.1, page 1675, of Groeneboom et al. (2001b))**.** If φ_n and φ are concave functions and $\varphi_n(u) \to \varphi(u)$ for all u in some closed set $C \subseteq \mathbb{R}$, then for all u in the interior of C, we have

$$\infty > \varphi'(u-) \geq \liminf_{n\to\infty} \varphi_n'(u-) \geq \liminf_{n\to\infty} \varphi_n'(u+) \geq \varphi'(u+) > -\infty. \tag{3.13}$$

Proof. The proof in Groeneboom et al. (2001b) gives the result for convex functions. There, they assume that the functions φ and φ_n are convex, decreasing densities, rather than simply being convex, but they do not use that the functions are decreasing densities. □

The appropriate topology to use for discontinuous functions such as these derivatives of concave functions is not the uniform topology but rather the Skorokhod topology. Instead of doing that now, we will simply note that for a sequence of monotone functions converging to a monotone limit, pointwise consistency implies uniform consistency on intervals on which the limit is continuous.

Corollary 3.1.11. *Let Assumption A be satisfied. Let $\hat{f}_n^0$ be the MLE of $f_0 \in \mathcal{P}_m$ and $\hat{\varphi}_n^0 = \log \hat{f}_n^0$, and let $[b,c]$ be a closed interval interior to $\mathrm{dom}\,\varphi_0$ and on which φ_0 is continuous. Then*

$$\sup_{u\in[b,c]} \left|(\hat{\varphi}_n^0)'(u) - \varphi_0'(u)\right| \longrightarrow 0 \quad \text{as } n \to \infty, \tag{3.14}$$

almost surely.

Proof. Because we have pointwise consistency of $\hat{\varphi}_n^0$ to φ_0 on $[b, c]$, we can apply Lemma 3.1.10 to get pointwise convergence of $(\hat{\varphi}_n^0)'$ to φ_0' at all points at which φ_0 is continuous. Then we are done, since pointwise convergence of a monotone function on a bounded interval implies uniform convergence on that interval if the limit is (monotone and) continuous. That is, taking $b = 0$ and $c = 1$ without loss of generality, and taking M to be a large integer, we have

$$\sup_{x\in[b,c]} |(\hat{\varphi}_n^0)'(x) - \varphi_0'(x)| = \max_{1\le j\le M} \sup_{(j-1)/M\le x\le j/M} |(\hat{\varphi}_n^0)'(x) - \varphi_0'(x)|,$$

which is equal to

$$\max_{1\le j\le M} \left(\sup_{(j-1)/M\le x\le j/M} (\hat{\varphi}_n^0)'(x) - \varphi_0'(x) \vee \sup_{(j-1)/M\le x\le j/M} \varphi_0'(x) - (\hat{\varphi}_n^0)'(x) \right),$$

which, by monotonicity, is bounded above by

$$\max_{1\le j\le M} \left((\hat{\varphi}_n^0)'((j-1)/M) - \varphi_0'(j/M) \vee \varphi_0'((j-1)/M) - (\hat{\varphi}_n^0)'(j/M) \right),$$

which for M large enough, by continuity of φ_0, is bounded above by

$$\max_{1\le j\le M} \left((\hat{\varphi}_n^0)'((j-1)/M) - \varphi_0'((j-1)/M) \vee \varphi_0'(j/M) - (\hat{\varphi}_n^0)'(j/M) \right) + \epsilon,$$

which converges almost surely to $0+\epsilon$ as n goes to infinity, by pointwise consistency of $(\hat{\varphi}_n^0)'$. Let ϵ go to 0 to complete the proof. ☐

3.2 Global Rates of Convergence for the UMLE and CMLE

3.2.1 Introduction and overview

We now turn our attention to global rates of convergence of log-concave maximum likelihood estimators, with focus on the Hellinger metric. In this section we will focus on statements of results while omitting most proofs, which can be found in Doss and Wellner (2013). A density p on $\mathbb{R}^d$ is *log-concave* if

$$p = e^{\varphi} \quad \text{where} \quad \varphi : \mathbb{R}^d \mapsto \mathbb{R} \quad \text{is concave.}$$

Log-concave densities are always unimodal and have convex level sets. Furthermore, log-concavity is preserved under marginalization and convolution; see e.g. Dharmadhikari and Joag-Dev (1988) chapter 2, pages 61-66. Thus the classes of log-concave densities can be viewed as natural nonparametric extensions of the class of Gaussian densities.

The classes of log-concave densities on $\mathbb{R}$ and $\mathbb{R}^d$ are special cases of the classes of s-concave densities as is nicely explained by Dharmadhikari and Joag-Dev (1988), pages 84-99. These classes are defined by the generalized means of order s as follows. Let

$$M_s(a, b; \theta) \equiv \begin{cases} ((1-\theta)a^s + \theta b^s)^{1/s}, & s \neq 0, \; a, b \geq 0, \\ a^{1-\theta} b^{\theta}, & s = 0, \\ \min(a, b), & s = -\infty. \end{cases}$$

Then $p \in \widetilde{\mathcal{P}}_{d,s}$, the class of s-concave densities on $C \subset \mathbb{R}^d$ if p satisfies

$$p((1-\theta)x_0 + \theta x_1) \geq M_s(p(x_0), p(x_1); \theta)$$

for all $x_0, x_1 \in C$ and $\theta \in (0,1)$. It is not hard to see that $\widetilde{\mathcal{P}}_{d,0}$ consists of densities of the form $p = e^{\varphi}$ where $\varphi \in [-\infty, \infty)$ is concave, and densities p in $\widetilde{\mathcal{P}}_{d,s}$ with $s < 0$ have the form $p = \varphi^{1/s}$ where $\varphi \in [0, \infty)$ is convex, and with $s > 0$ have the form $p = \varphi_+^{1/s}$ where φ is concave on C (and then we write $\widetilde{\mathcal{P}}_{d,s}(C)$); see for example Dharmadhikari and Joag-Dev (1988) page 86. These classes are nested since

$$\widetilde{\mathcal{P}}_{d,s}(C) \subset \widetilde{\mathcal{P}}_{d,0} \subset \widetilde{\mathcal{P}}_{d,r} \subset \widetilde{\mathcal{P}}_{d,-\infty}, \qquad \text{if} \quad -\infty < r < 0 < s < \infty. \tag{3.15}$$

Here we view the classes $\widetilde{\mathcal{P}}_{1,s}$ defined above for $d = 1$ in terms of the generalized means M_s as being obtained as increasing transforms h_s of the class of concave functions on $\mathbb{R}$ with

$$h_s(y) = \begin{cases} e^y, & s = 0, \\ (-y)_+^{1/s}, & s < 0, \\ y_+^{1/s}, & s > 0. \end{cases} \tag{3.16}$$

Thus we define

$$\begin{aligned} \mathcal{P}_{1,0} &= \{p = e^{\varphi} : \quad \varphi \ \text{ is concave}\}, \\ \mathcal{P}_{1,s} &= \{p = h_s(\varphi) : \quad \varphi \ \text{ is concave}\}, \quad s < 0, \\ \mathcal{P}_{1,s} &= \{p = h_s(\varphi) : \quad \varphi \ \text{ is concave}\}, \quad s > 0 \end{aligned}$$

where all the concave functions are assumed to be closed (i.e. upper semicontinuous), proper, and are viewed as concave functions on all of $\mathbb{R}$ rather than on a (possibly) specific set C. Thus we allow $\text{ran}(\varphi) \subset [-\infty, \infty)$. See (3.18) in Section 3.2.2. This view simplifies our treatment in much the same way as the treatment in Seregin and Wellner (2010), but with "increasing" transformations replacing the "decreasing" transformations of Seregin and Wellner, and "concave functions" here replacing the "convex functions" of Seregin and Wellner.

Nonparametric estimation of log-concave and s-concave densities has developed rapidly in the last decade: For log-concave densities on $\mathbb{R}$, Pal et al. (2007b) established existence of the Maximum Likelihood Estimator (MLE) $\widehat{p}_n$ of p_0, provided a method to compute it, and showed that it is Hellinger consistent: $H(\widehat{p}_n, p_0) \rightarrow_{a.s.} 0$ where $H^2(p,q) = (1/2)\int\{\sqrt{p}-\sqrt{q}\}^2 dx$ is the (squared) Hellinger distance. Dümbgen and Rufibach (2009) also discussed algorithms to compute $\widehat{p}_n$ and rates of convergence with respect to supremum metrics on compact subsets of the support of p_0 under Hölder smoothness assumptions on p_0. Balabdaoui et al. (2009) established limit distribution theory for the MLE of a log-concave density at fixed points under various differentiability assumptions and investigated the natural mode estimator associated with the MLE. Seregin and Wellner (2010) showed that the MLE exists and is consistent for the classes $\mathcal{P}_{d,s}$ with $s \in (-1,0) \cup (0,\infty)$. Although it has been conjectured that the MLE is Hellinger-consistent at rate $n^{-2/5}$ in the one-dimensional cases (see e.g. Seregin and Wellner (2010), pages 3378-3379), to the best of our knowledge this has not yet been proved.

The main difficulty in establishing global rates of convergence with respect to the Hellinger or other metrics has been to derive suitable bounds for the metric entropy with bracketing for appropriately large subclasses $\mathcal{P}$ of log-concave or s-concave densities. We seek bounds of the form

$$\log N_{[\,]}(\epsilon, \mathcal{P}, H) \lesssim K\epsilon^{-1/2}, \quad \epsilon \le \epsilon_0 \tag{3.17}$$

where $N_{[\,]}(\epsilon, \mathcal{P}, H)$ denotes the minimal number of $\epsilon-$brackets with respect to the Hellinger metric H needed to cover $\mathcal{P}$. We will establish such bounds in Section 3.2.5 using recent results of Dryanov (2009) for convex functions on $\mathbb{R}$ and Guntuboyina and Sen (2013) who extended the results of Dryanov from $\mathbb{R}$ to $\mathbb{R}^d$. These recent results build on earlier work by Bronšteĭn (1976) and Dudley (1984); see also Dudley (1999), pages 269-281. The main difficulty has been that the bounds of Bronšteĭn (1976) involve restrictions on the Lipschitz behavior of the convex functions involved as well as bounds on the supremum norm of the functions. The classes of log-concave functions to be considered must include the estimators $\widehat{p}_n$ (at least with arbitrarily high probability for large n). It is well-known that log-concave densities on $\mathbb{R}$ can have at most two discontinuities with these occurring at the endpoints of the support; see e.g. Schoenberg (1951), page 339. Since the estimators $\widehat{p}_n$ are discontinuous at the upper and lower ends of their support (which is contained in the support of the true density p_0), the supremum norm does not give control of the Lipschitz behavior of the estimators in neighborhoods of the end points of their support. Dryanov (2009) showed how to get rid of the constraint on Lipschitz behavior when moving from metric entropy with respect to supremum norms to metric entropies with respect to L_r norms. Furthermore, Guntuboyina and Sen (2013) showed how to extend Dryanov's results from $\mathbb{R}$ to $\mathbb{R}^d$. Here we show how the results of Dryanov (2009) and Guntuboyina and Sen (2013) can be strengthened from metric entropy with respect to L_r to bracketing entropy with respect to L_r, and we carry these results over to the class of log-concave densities by an argument which we call *dual induction*, since it involves keeping track of brackets for concave (or convex) functions, and their transforms by an exponential function, simultaneously.

Once bounds of the form (3.17) are available, then relatively standard tools from empirical process theory going back to Birgé and Massart (1993), van de Geer (1993), Wong and Shen (1995), and developed further in van de Geer (2000) and van der Vaart and Wellner (1996), become available.

Our focus here will be on global convergence rates for the MLE's $\widehat{p}_n$ of log-concave and s-concave densities p_0 on $\mathbb{R}$. This seems to be a natural first step toward rates of MLEs and regularized versions of MLEs in the more difficult log-concave and s-concave problems with $d \geq 2$.

3.2.2 *Basic definitions and notation*

We will restrict attention to the class of concave functions

$$\mathcal{C} := \{\varphi : \mathbb{R} \to [-\infty, \infty) | \varphi \text{ is a closed, proper concave function}\}, \tag{3.18}$$

where Rockafellar (1970) defines proper (page 24) and closed (page 52) convex functions. A concave function is proper or closed if its negative is a proper or closed convex function, respectively. We also follow the convention that all concave functions φ are defined on all of $\mathbb{R}$ and take the value $-\infty$ off of their *effective domains*, $\operatorname{dom}\varphi := \{x : \varphi(x) > -\infty\}$. These conventions are motivated in (Rockafellar, 1970, pp. 40). While we consider $\varphi \in \mathcal{C}$ to be defined on $\mathbb{R}$, we will still sometimes consider a function ψ which is the "restriction of φ to D" for some $D \subset \mathbb{R}$. By this, in keeping with the above-mentioned convention, we still mean that ψ is defined on $\mathbb{R}$, where if $x \notin D$ then $\psi(x) = -\infty$, and otherwise $\psi(x) = \varphi(x)$. We will let $\varphi|_D$ denote such restricted functions ψ. Additionally, when we want to speak about the range of any function f (not necessarily concave) we will use set notation, e.g. for $S \subseteq \mathbb{R}$, $f(S) := \{y : f(x) = y \text{ for some } x \in S\}$. We will call $\operatorname{ran}\varphi := \varphi(\operatorname{dom}\varphi)$ the *effective range* of φ. We will sometimes want to restrict not the domain of φ but, rather, the range of φ. We will thus let $\varphi|^I$ denote $\varphi|_{D_{\varphi,I}}$ for any interval $I \subset \mathbb{R}$, where $D_{\varphi,I} = \{x : \varphi(x) \in I\}$. Thus, for instance, for all intervals I containing $\operatorname{ran}\varphi$, we have $\varphi|^I \equiv \varphi$.

We will be considering classes of nonnegative concave-transformed functions of the type $h \circ \mathcal{C}$ for some transformation h where $h(-\infty) = 0$ and $h(\infty) = \infty$. For example, the class of log concave densities $\mathcal{P}_{1,0}$ is a subset of the class of non-negative functions $\mathcal{F}_{1,0} = \{h_0 \circ \varphi : \ \varphi \text{concave}\}$ where $h_0(u) = e^u$, and, for $s < 0$, the class of s-concave densities $\mathcal{P}_{1,s}$ is a subset of the class of non-negative functions $\mathcal{F}_{1,s} = \{h_s \circ \varphi : \ \varphi \text{ concave}\}$ where $h_1(u) = (-u)^{1/s}$, $u < 0$, $h_s(u) = +\infty$ for $u \geq 0$. We will elaborate on this in Section 3.2.6.

We will slightly abuse notation by allowing the dom and ran operators to apply to such concave-transformed functions. In this case, we let $\operatorname{dom} h \circ \varphi := \{x : h(\varphi(x)) > 0\}$ and $\operatorname{ran} h \circ \varphi := h(\varphi(\operatorname{dom} h \circ \varphi))$.

3.2.3 Maximum likelihood estimators: basic properties

We divide our treatment here according to $s = 0$ and then $s \in (-1, 0) \cup (0, \infty)$.

Log-concave densities: basic properties and consistency

Let $X_1, \ldots, X_n$ be i.i.d. with density $p_0 = e^{\varphi_0}$ where $\varphi_0 : \mathbb{R} \to [-\infty, \infty)$ is concave. Thus p_0 is log-concave. Write $\mathbb{P}_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ for the empirical measure of the X_i's. The maximum likelihood estimator $\widehat{p}_n = \exp(\widehat{\varphi}_n)$ of p_0 maximizes

$$\Psi_n(\varphi) \quad = \quad \mathbb{P}_n \log p - \int \log p(x) dx = \mathbb{P}_n \varphi - \int e^{\varphi(x)} dx$$

over all concave functions φ. From Walther (2002), Pal et al. (2007b) and Dümbgen and Rufibach (2009) (Theorem 2.1) we know that $\widehat{\varphi}_n$ exists and is unique. It is linear on all intervals $[X_{(j)}, X_{(j+1)}]$, $j = 1, \ldots, n-1$, where $X_{(1)} \leq \cdots \leq X_{(n)}$ denote the order statistics of the X_i's. Furthermore $\widehat{\varphi}_n = -\infty$ on $\mathbb{R} \setminus [X_{(1)}, X_{(n)}]$. Thus $\widehat{p}_n$ is upper semicontinuous with jump discontinuities (down to zero) at both $X_{(1)}$ and $X_{(n)}$.

The following lemma is basic.

Lemma 3.2.1. For any log-concave density p on $\mathbb{R}$, there exist $a > 0$ and $b \in \mathbb{R}$ such that

$$p(x) \leq e^{-a|x|+b} \;\; \text{for all} \;\; x \in \mathbb{R}.$$

This is a simplified version of Lemma A.1 of Dümbgen and Rufibach (2009); earlier results with a similar spirit were given by Schoenberg (1951) and Devroye (1984). An analogous result for log-concave densities on $\mathbb{R}^d$ is given in Lemma 1 of Cule and Samworth (2010).

Theorem 3.2.2. *(Consistency and boundedness of $\widehat{p}_n$ for $\mathcal{P}_{1,0}$)*

(i) $H(\widehat{p}_n, p_0) \to_{a.s.} 0$.

(ii) If S is a compact set strictly contained in the support of p_0,

$$\sup_{x \in S} |\widehat{p}_n(x) - p_0(x)| \to_{a.s.} 0.$$

(iii) If p_0 is continuous on $\mathbb{R}$, and $p_0(x) \le e^{-a_0|x|+b_0}$ (by Lemma 3.2.1), then for any $0 \le a < a_0$,

$$\sup_{x \in \mathbb{R}} e^{a|x|} |\widehat{p}_n(x) - p_0(x)| \to_{a.s.} 0.$$

(iv) For any p_0 log-concave with $p_0(x) \le e^{-a_0|x|+b_0}$ (by Lemma 3.2.1), then for any $0 \le a < a_0$,

$$\int_{\mathbb{R}} e^{a|x|} |\widehat{p}_n(x) - p_0(x)| dx \to_{a.s.} 0.$$

(v) $\limsup_{n\to\infty} \sup_x \widehat{p}_n(x) \le M(p_0) < \infty$ *almost surely.*

The first statement (i) is proved by Pal et al. (2007b); statement (ii) is a corollary of Theorem 4.1 of Dümbgen and Rufibach (2009); (iii) and (iv) are special cases of Theorem 4.1 of Cule and Samworth (2010), but (iv) with $a = 0$ is also given in Corollary 4.2 by Dümbgen and Rufibach (2009). (v) This is Theorem 3.2 of Pal et al. (2007b) and will be needed to handle cases in which the mode(s) of p_0 are in the boundary of the support.

s-concave densities: basic properties and consistency

Let $X_1, \ldots, X_n$ be i.i.d. with density $p_0 = \varphi_0^{1/s}$ where $\varphi_0 : \mathbb{R} \to [0, \infty)$ is convex. Thus p_0 is s-concave. Write $\mathbb{P}_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ for the empirical measure of the X_i's. The maximum likelihood estimator $\widehat{p}_n = \widehat{\varphi}_n^{1/s}$ of p_0 maximizes

$$\Psi_n(\varphi) \quad = \quad \mathbb{P}_n \varphi^{1/s}$$

over all convex functions φ for which $\int (\varphi)^{1/s}(x)dx = 1$. From Seregin and Wellner (2010) (Theorem 2.12, page 3757) we know that $\widehat{\varphi}_n$ exists if $n \geq n_1 = 1/(r-1)$ with $r \equiv -1/s > 1$ the case $s < 0$ and if $n \geq 2$ when $s > 0$. Seregin and Wellner (2010) page 3762 conjectured that $\widehat{\varphi}_n$ is unique when it exists. Note that the class $\mathcal{P}_{1,-\infty}$ corresponds to the class of all unimodal densities; see e.g. Dharmadhikari and Joag-Dev (1988) page 85, and for this class it is known that the MLE does not exist (see e.g. Birgé (1997)). In fact, it is easily seen that the MLE $\widehat{p}_n$ of $p_0 \in \mathcal{P}_{1,s}$ does not exist for any $s < -1$; see Proposition 8.1 of Doss and Wellner (2013).

Theorem 3.2.3. *(Consistency and boundedness of $\widehat{p}_n$ for $\mathcal{P}_{1,s}$, $s \in (-1,0) \cup (0,\infty)$)*

(i) $H(\widehat{p}_n, p_0) \rightarrow_{a.s.} 0$.

(ii) If C is a compact set strictly contained in the support of p_0,

$$\sup_{x \in C} |\widehat{p}_n(x) - p_0(x)| \rightarrow_{a.s.} 0.$$

(iii) If p_0 is continuous on $\mathbb{R}$, then

$$\sup_{x \in \mathbb{R}} |\widehat{p}_n(x) - p_0(x)| \rightarrow_{a.s.} 0.$$

(iv) $\limsup_{n\to\infty} \sup_x \widehat{p}_n(x) \leq M(p_0) < \infty$ *almost surely.*

The first statement (i) is Theorem 2.17 of Seregin and Wellner (2010); statements (ii) and (iii) are consequences of Theorem 2.18 of Seregin and Wellner (2010). (iv) is Lemma 3.17 in Seregin and Wellner (2010) and will be needed to handle cases in which the mode(s) of p_0 are in the boundary of the support (of p_0).

3.2.4 Bracketing entropy bounds and rates of convergence for log-concave and s-concave densities

Log-concave and s-concave densities: rates for the MLE

Our main goal is to establish rates of convergence for the Hellinger consistency given in (i) of Theorems 3.2.2 and 3.2.3. It suffices, without loss of generality, to suppose that the mode (or smallest mode to be more precise) m_0 of p_0 is 0. Although the MLE $\widehat{p}_n$ of p_0 will then have mode $\widehat{m}_n \to_{a.s.} m_0 = 0$, $\widehat{p}_n$ will usually not have mode exactly at 0. We therefore consider the following subclasses $\mathcal{P}_{1,M,s}$ of log-concave and s-concave densities which will contain both p_0 and $\widehat{p}_n$ with high probability for large n. Let m_p denote the (smallest) mode of p. Since all s-concave densities are unimodal, this is well-defined. Then, for $0 < M < \infty$, let

$$\mathcal{P}_{1,M,0} \equiv \left\{ \begin{array}{l} p = e^{\varphi} : \ \varphi \text{ is concave}, \varphi \in [-\infty, \infty), \\ \quad m_p \in [-1,1], \ 1/M \leq p(m_p) \leq M, \quad p(0) > 0 \end{array} \right\}, \tag{3.19}$$

$$\mathcal{P}_{1,M,s} \equiv \left\{ \begin{array}{l} p = \varphi_+^{1/s} : \ \varphi \text{ is convex}, \ m_p \in [-1,1], \\ \qquad\qquad 1/M \leq p(m_p) \leq M, \quad p(0) > 0 \end{array} \right\}, \quad s < 0, \tag{3.20}$$

and

$$\mathcal{P}_{1,M,s} \equiv \left\{ \begin{array}{l} p = \varphi_+^{1/s} : \ \varphi \text{ is concave}, \ m_p \in [-1,1], \\ \qquad\qquad 1/M \leq p(m_p) \leq M, \quad p(0) > 0 \end{array} \right\}, \quad s > 0. \tag{3.21}$$

The following lemma gives upper envelopes for the classes $\mathcal{P}_{1,M,s}$ with $-1 < s \leq 0$.

Lemma 3.2.4. For any $p \in \mathcal{P}_{1,M,0}$ and $0 < M < \infty$,

$$p(x) \le \left\{ \begin{array}{ll} Me\exp\left(-\frac{1}{M}(x-1)\right), & x \ge 1, \\ Me\exp\left(-\frac{1}{M}|x+1|\right), & x \le -1, \\ Me, & x \in [-1,+1], \end{array} \right\} \equiv p_{u,1,0}(x). \tag{3.22}$$

For any $p \in \mathcal{P}_{1,M,s}$ with $-1 < s < 0$ and $0 < M < \infty$

$$\begin{aligned} p(x) \quad &\le \quad \left\{ \begin{array}{ll} M\left(1+s-\frac{s}{M}(x-1)\right)^{1/s}, & x \ge 1, \\ M\left(1+s-\frac{s}{M}(x-1)\right)^{1/s}, & x \le -1, \\ M(1+s)^{1/s}, & x \in [-1,+1], \end{array} \right\} \\ &\equiv \quad p_{u,1,s-}(x). \end{aligned} \tag{3.23}$$

For any $p \in \mathcal{P}_{1,M,s}$ with $0 < s \le 1$ and $0 < M < \infty$

$$\begin{aligned} p(x) \quad &\le \quad \left\{ \begin{array}{ll} M\left(1+s-\frac{s}{M}(x-1)\right)_+^{1/s}, & x \ge 1, \\ M\left(1+s-\frac{s}{M}(x-1)\right)_+^{1/s}, & x \le -1, \\ M(1+s)^{1/s}, & x \in [-1,+1], \end{array} \right\} \\ &\equiv \quad p_{u,1,s+}(x). \end{aligned} \tag{3.24}$$

The above envelopes will later play an important role in the proof of Theorem 3.2.32. Now let the bracketing entropy of a class of functions $\mathcal{F}$ with respect to a semi-metric d on $\mathcal{F}$ be defined in the usual way; see e.g. Dudley (1999) page 234, van der Vaart and Wellner (1996), page 83, or van de Geer (2000), page 16. With this preparation we can state our main results as follows:

Theorem 3.2.5. *Suppose that* $s \in (-1/5, 1]$. *Then*

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

for all $\epsilon \le \epsilon_0$ *where the constant implied by* $\lesssim$ *and* ϵ_0 *depends only on* M *and* s

Theorem 3.2.5 is the main tool we need to obtain rates of convergence for the MLEs $\widehat{p}_n$. This is given in our second main theorem:

Theorem 3.2.6. *Suppose that* $s \in (-1/5, \infty)$ *and* $\widehat{p}_n$ *is the MLE of the* s*-concave density* p_0. *Then* $n^{2/5} H(\widehat{p}_n, p_0) = O_p(1)$.

Theorem 3.2.6 is a fairly straightforward consequence of Theorem 3.2.5 by applying van de Geer (2000), Theorem 7.4, page 99, or van der Vaart and Wellner (1996), Theorem 3.4.4 in conjunction with Theorem 3.4.1, pages 322-323. In addition, we have further consequences since the Hellinger metric dominates the total variation or L_1-metric and via van de Geer (2000), Corollary 7.5, page 100:

Corollary 3.2.7. *Suppose that* $s \in (-1/5, \infty)$ *and* $\widehat{p}_n$ *is the MLE of the* s*-concave density* p_0. *Then* $n^{2/5} \int_{\mathbb{R}} |\widehat{p}_n(x) - p_0(x)| dx = O_p(1)$.

Corollary 3.2.8. *Suppose that* $s \in (-1/5, \infty)$. *If* $\widehat{p}_n$ *is the MLE of the* s*-concave density* p_0*, then the log-likelihood ratio (divided by* n*)* $\mathbb{P}_n \log(\widehat{p}_n / p_0)$ *satisfies*

$$n^{4/5} \mathbb{P}_n \log\left(\frac{\widehat{p}_n}{p_0}\right) = O_p(1). \tag{3.25}$$

The above results also apply to the case of the log-concave MLE constrained to have known location of the mode.

Theorem 3.2.9. *Suppose* $\hat{p}_n^0$ *is the MLE of the log-concave density* p_0 *which has fixed and known mode. Then* $n^{2/5} H(\hat{p}_n^0, p_0) = O_p(1)$.

Corollary 3.2.10. *Suppose that $\hat{p}_n^0$ is the MLE of the log-concave density p_0 which has fixed and known mode. Then $n^{2/5} \int_{\mathbb{R}} |\hat{p}_n^0(x) - p_0(x)| dx = O_p(1)$.*

Corollary 3.2.11. *If $\hat{p}_n^0$ is the MLE of the log-concave density p_0 which has fixed and known mode, then the log-likelihood ratio (divided by n) $\mathbb{P}_n \log(\hat{p}_n^0/p_0)$ satisfies*

$$n^{4/5} \mathbb{P}_n \log \left(\frac{\hat{p}_n^0}{p_0} \right) = O_p(1). \tag{3.26}$$

The results (3.25) and (3.26) are of interest in connection with the study of the likelihood ratio statistic, $\mathbb{P}_n \log \left(\widehat{p}_n / \hat{p}_n^0 \right)$, for tests (and resulting confidence intervals) for the mode of p_0. This likelihood ratio statistic is also $O_p(n^{-4/5})$, by Corollary 3.2.8 and Corollary 3.2.11, and this will be of use in Section 5.3.

3.2.5 Bracketing entropy bounds: extending Guntuboyina and Sen (2013)

Control of the entropies of classes of concave (or convex) functions with respect to supremum metrics requires control of Lipschitz constants, which we do not have. Thus, we will use L_r with $r \geq 1$ and related distances instead. We are really interested in L_r distances of concave-transformed classes of functions. Thus, when we measure distances between concave functions, we will not use the L_r metrics themselves, but new metrics that will give us information about L_r distance on the transformed scale. We will define such distances in Section 3.2.6. First, we will define the classes of concave and concave-transformed functions which we will be studying.

Definition 3.2.12. (i) For fixed $0 \leq u \leq \infty$, $-\infty < b_1 < b_2 < \infty$, and $-\infty < B_1 < B_2 < \infty$, we let $\mathcal{C}([b_1, b_2], [B_1, B_2], u)$ be the class of all functions $\varphi \in \mathcal{C}$ satisfying:

(a) The domain of φ ends within u of the endpoints, i.e. letting $\mathrm{dom}(\varphi) := [d_{\varphi,1}, d_{\varphi,2}] \subseteq$

$[b_1, b_2]$, we have

$$d_{\varphi,1} - b_1 \le u \quad \text{and} \quad b_2 - d_{\varphi,2} \le u,$$

(b) $\operatorname{ran} \varphi \subseteq [B_1, B_2]$.

(ii) Similarly, for u, b_1, and b_2 as above, and for $0 \le B_1 \le B_2 < \infty$, we define $\mathcal{F}_h([b_1, b_2], [B_1, B_2], u)$ to be the class of all concave-transformed functions $f = h \circ \varphi$ satisfying

(a) $\operatorname{ran} f \subseteq [B_1, B_2]$,

(b) The domain of f ends within u of the endpoints, i.e. letting $\operatorname{dom}(f) := [d_{f,1}, d_{f,2}] \subseteq [b_1, b_2]$, we have

$$d_{f,1} - b_1 \le u \quad \text{and} \quad b_2 - d_{f,2} \le u.$$

Note that this means

$$\mathcal{F}_h([b_1, b_2], [B_1, B_2], u) = h \circ \mathcal{C}([b_1, b_2], [h^{-1}(B_1), h^{-1}(B_2)], u).$$

In the above, we take $u = \infty$ to mean that the domains may be any subinterval of $[b_1, b_2]$.
(iii) When $\mathcal{F}_h$ is a class of concave-transformed densities p, then we denote the class described in (ii) by $\mathcal{P}_h([b_1, b_2], [B_1, B_2], u)$.

The above classes are the classes whose bracketing entropy we will control via the methods of Guntuboyina and Sen (2013) and the new methods developed here in Section 3.2.6.

Proposition 3.2.13. *Let $r \ge 1$, $-\infty < b_1 < b_2 < \infty$, $-\infty < B_1 < B_2 < \infty$, and $0 < \epsilon \le \epsilon_0 (B_2 - B_1)(b_2 - b_1)^{1/r}$ for absolute constants $c_2 > 0$ and $\epsilon_0 > 0$. Then*

$$\log N_{[\,]}\left(\epsilon, \mathcal{C}([b_1, b_2], [B_1, B_2], c_2 \epsilon^r (b_2 - b_1)), L_r(\lambda)\right) \le c \left(\frac{(B_2 - B_1)(b_2 - b_1)^{1/r}}{\epsilon} \right)^{1/2}. \quad (3.27)$$

We may take all brackets $[l, u]$ such that $l(x) = B_1$ and $u(x) = B_2$ for all x such that $|x - b_i| < c_2\epsilon^r(b_2 - b_1)$, i.e. for all x in the set where the domains may end.

We also note that we can simply state Theorem 3.1 in Guntuboyina and Sen (2013) in terms of bracketing entropy instead of metric entropy. This yields:

Proposition 3.2.14 (Extension of Theorem 3.1 of Guntuboyina and Sen (2013))**.** *Let $r \geq 1$, $-\infty < b_1 < b_2 < \infty$, $-\infty < B_1 < B_2 < \infty$, and $0 < \epsilon \leq \epsilon_0(B_2 - B_1)(b_2 - b_1)^{1/r}$, where $\epsilon_0 > 0$ is an absolute constant. Then*

$$\log N_{[]}\left(\epsilon, \mathcal{C}([b_1, b_2], [B_1, B_2], 0), L_r(\lambda)\right) \leq c\left(\frac{(B_2 - B_1)(b_2 - b_1)^{1/r}}{\epsilon}\right)^{1/2}, \tag{3.28}$$

where c is a constant depending only on r

Now the above result gives us a bracketing number for a single ϵ value that governs the size of the window in which the domains of the concave functions can vary, for classes of bounded functions. However, we are really interested in classes who allow domain endpoints to vary throughout the interval $[b_1, b_2]$. In the next section we will use the above result which allows for small windows of varying domain to build up to bracketing entropy control for classes with varying domain over the entire interval.

3.2.6 Bracketing entropy bounds: the dual induction

The function classes in which we will be interested in the end are the classes $\mathcal{P}_{1,M,0}$ or $\mathcal{P}_{1,M,s}$ define in Section 3.2.4, or, more generally the classes $\mathcal{P}_{1,M,h}$ defined in Section 3.2.7, to which the MLEs belong (with high probability as sample size gets large). However, such classes contain functions that are arbitrarily close to or are equal to 0, and these correspond to concave functions that take unboundedly large (negative) values. Thus the corresponding concave classes do not have finite bracketing entropy for the L_r distance. To get around this difficulty, we will consider classes of truncated concave functions and the corresponding concave-transformed classes, and we will define a new metric for the concave functions that relates to the L_r distance of the corresponding concave-transformed class.

Definition 3.2.15. A *concave-function transformation*, h, is a nondecreasing function from $[-\infty, \infty]$ to $[0, \infty]$ such that $h(\infty) = \infty$ and $h(-\infty) = 0$. We define its limit points $\tilde{y}_0 < \tilde{y}_\infty$ by $\tilde{y}_0 = \inf\{y : h(y) > 0\}$ and $\tilde{y}_\infty = \sup\{y : h(y) < \infty\}$, we assume that $h(\tilde{y}_0) = 0$ and $h(\tilde{y}_\infty) = \infty$, and we define $h_0 = \lim_{y \searrow \tilde{y}_0} h(y)$. We assume h is continuously differentiable on $(\tilde{y}_0, \tilde{y}_\infty)$.

The transformation h is not necessarily continuous at $\tilde{y}_0$ if $\tilde{y}_0$ is not $-\infty$, so h_0 is the minimum value of h that is not 0.

Remark 3.2.16. These transformations correspond to "decreasing transformations" in the terminology of Seregin and Wellner (2010) In that paper, the transformations are applied to convex functions whereas here we apply our transformations to concave ones. Since negatives of convex functions are concave, and vice versa, each of our transformations h defines a decreasing transformation $\tilde{h}$ as defined in Seregin and Wellner (2010) via $\tilde{h}(y) = h(-y)$.

We will sometimes make the following assumptions.

Assumption B. *Assume that the transformation h satisfies:*

T.1 $h'(y) = o((-y)^{-(\alpha+1)})$ *as* $y \searrow \tilde{y}_0$ *for some* $\alpha > 1$.

T.2 if $\tilde{y}_0 > -\infty$, *then for all* $\tilde{y}_0 < c < \tilde{y}_\infty$, *there is an* $0 < M_c < \infty$ *such that* $h'(y) \le M_c$ *for all* $y \in (\tilde{y}_0, c]$;

T.3 if $\tilde{y}_\infty < \infty$ *then for some* $0 < c < C$, $c(y - \tilde{y}_\infty)^{-\beta} \le h(y) \le C(y - \tilde{y}_\infty)^{-\beta}$ *for some* $\beta > 1$ *and* y *in a neighborhood of* $\tilde{y}_\infty$;

T.4 if $\tilde{y}_\infty = \infty$ *then* $h(y)^\gamma h(-Cy) = o(1)$ *for some* $\gamma, C > 0$, *as* $y \to \infty$.

Note that Assumption (T.2) does not preclude h from being discontinuous at $\tilde{y}_0$ when $\tilde{y}_0 < \infty$. Additionally, notice this assumption holds automatically if $\tilde{y}_0 = -\infty$ when Assumption (T.1) holds.

Definition 3.2.17. For a decreasing vector $\underline{y}_k = (y_0, \ldots, y_k) \in \mathbb{R}^{k+1}$, i.e. a vector with $\tilde{y}_\infty \geq y_0 > y_1 > \ldots > y_k \geq \tilde{y}_0$, for $-\infty < b_1 < b_2 < \infty$, and letting $\mathcal{D}(b_1) := \{\varphi \in \mathcal{C} | \operatorname{dom}\varphi = [b_1, d_\varphi] \text{ or } \operatorname{dom}\varphi = \emptyset\}$ we define

$$\mathcal{C}_k \equiv \mathcal{C}([b_1, b_2], [y_k, y_0], \infty) \cap \mathcal{D}(b_1),$$

a class of concave functions whose domains have right endpoint which may vary freely in $[b_1, b_2]$. We also define the concave-transformed functions

$$\mathcal{F}_{k,h} = h \circ \mathcal{C}_k.$$

Example 3.2.18. The class of log-concave densities, as discussed in Section 3.2.4 is obtained by taking $h(y) = e^y \equiv h_0(y)$ for $y \in \mathbb{R}$. Then $\tilde{y}_0 = -\infty$ and $\tilde{y}_\infty = \infty$. Assumption (T.4) holds with any $\gamma > C > 0$, and Assumption (T.1) holds for any $\alpha > 1$.

Example 3.2.19. The classes of s-concave functions with $s \in (-1, 0)$, as discussed in Section 3.2.4 are obtained by taking $h(y) = (-y)^{1/s} \equiv h_s(y)$ for $s \in (-1, 0)$ and for $y < 0$. Here $\tilde{y}_0 = -\infty$ and $\tilde{y}_\infty = 0$. Assumption (T.3) holds for $\beta = -1/s$, and Assumption (T.1) holds for any $\alpha \in (1, -1/s)$.

Example 3.2.20. The classes of s-concave functions with $0 < s < \infty$, as discussed in Section 3.2.4 are obtained by taking $h(y) = (y)_+^{1/s} \equiv h_s(y)$ for $s \in (0, \infty)$. Here $\tilde{y}_0 = 0$ and $\tilde{y}_\infty = \infty$. Assumption (T.1) holds for any $\alpha > 1 > -1/s$, Assumption (T.2) fails if $s > 1$, and Assumption (T.4) holds for any $C, \gamma > 0$. These (small) classes $\mathcal{P}_h$ are covered by our Corollary 3.2.35.

Example 3.2.21. To connect the preceding two examples, consider $\tilde{h}_s(y) = (1 + sy)^{1/s}$ for $y \in (-\infty, -1/s)$ with $-1 < s < 0$. Note that $\tilde{h}_s(y) \to e^y$ as $s \nearrow 0$. Here $\tilde{y}_0 = -\infty$ and $\tilde{y}_\infty = -1/s$. Assumption (T.3) holds for $\beta = -1/s$, and Assumption (T.1) holds for any $\alpha \in (1, -1/s)$.

Example 3.2.22. To illustrate the possibilities further, consider $h(y) = \tilde{h}_s(y) = (1+sy)^{1/s}$ for $y \in [0, -1/s)$ with $-1 < s < 0$, and $h(y) = \tilde{h}_r(y)$ for $y \in (-\infty, 0)$ and $r \in (-1, 0]$. Here $\tilde{y}_0 = -\infty$ and $\tilde{y}_\infty = -1/s$. Assumption (T.3) holds for $\beta = -1/s$, and Assumption (T.1) holds for any $\alpha \in (1, -1/r)$. Note that this example fails to satisfy Assumption (T.4) when $r < 0$ and $s = 0$ (and then $\tilde{y}_\infty = \infty$).

We are interested in bracketing entropy of $\mathcal{F}_{k,h}$ with the $L_r(\lambda)$ norm but we can get control of bracketing entropy for $\mathcal{C}_k$ since it is a class of concave functions. Thus we define a metric on the latter space that relates to $L_r(\lambda)$ distance on $\mathcal{F}_{k,h}$.

Definition 3.2.23. Let h be a concave-function transformation and assume we have a decreasing sequence $\tilde{y}_\infty > y_0 > y_1 \cdots > y_k > \infty$, denoted $\underline{y}_k = (y_0, y_1, \ldots, y_k) \in \mathbb{R}^{k+1}$. If $\tilde{y}_0 > -\infty$ then if h is discontinuous at $\tilde{y}_0$, we also assume $y_k > \tilde{y}_0$, but if h is continuous at $\tilde{y}_0$ (i.e. approaches 0 from the right) we allow y_k to possibly be $\tilde{y}_0$. Set $w_j \equiv w_{j,h} = \sup_{y \in [y_j, y_{j-1}]} h'(y)$ for $j = 1, \ldots, k+1$. Then for $x \in \mathbb{R}$ define

$$W(x) \equiv W(x; \underline{y}_k, h) \equiv \sum_{j=1}^{k} w_j \lambda((y_k, x] \cap [y_j, y_{j-1}])$$

where λ denotes Lebesgue measure on $\mathbb{R}$. (Alternatively, for Borel subsets A of $\mathbb{R}$ define the measure W by

$$W(A) \equiv W(A; \underline{y}_k, h) \equiv \sum_{j=1}^{k} w_j \lambda(A \cap [y_j, y_{j-1}]).)$$

Then for $a, b \in \overline{\mathbb{R}} \equiv [-\infty, \infty]$ define the (weighted) distance $d_{k,h} \equiv d_{\underline{y}_{k+1},h}$ by

$$d_{k,h}(a,b) \equiv d_{\underline{y}_k,h}(a,b) \equiv \int_{a \wedge b}^{a \vee b} dW(x) = W(a \vee b) - W(a \wedge b).$$

Note that if we make Assumption (T.2) then by the definition of $\underline{y}_k$ and the fact that h' is continuous on $(\tilde{y}_0, \tilde{y}_\infty)$, then in all scenarios (i.e. if $\tilde{y}_0 = -\infty$ or if $\tilde{y}_0$ is finite and h is continuous or discontinuous) the weights w_j are finite.

Next, we define the $L_r(\lambda)$ generalization of the above metric by integrating.

Definition 3.2.24. Let λ denote Lebesgue measure on $\mathbb{R}$, let h be a concave-function transformation, and let $\underline{y}_k$ be as in Definition 3.2.23. For two functions φ_1 and φ_2 defined on $[b_1, b_2]$ we define

$$\begin{aligned} d_{r,k,h}(\varphi_1, \varphi_2) = d_{r,\underline{y}_k,h}(\varphi_1, \varphi_2) &= \left(\int_{x \in \mathrm{dom}(\varphi_1) \cup \mathrm{dom}(\varphi_2)} d_{\underline{y}_k,h}(\varphi_1(x), \varphi_2(x))^r \, dx \right)^{1/r} \\ &= \left(\int_{b_1}^{b_2} d_{\underline{y}_k,h}(\varphi_1(x), \varphi_2(x))^r \, dx \right)^{1/r}, \end{aligned}$$

where we take φ_1 or φ_2 to be $-\infty$ outside their respective domains.

Note that since dW only puts mass on $[y_k, y_0]$, the distance is always finite (since the two domains are bounded). $d_{r,k,h}(\cdot, \cdot)$ is indeed a metric; i.e. the triangle inequality holds; see Lemma 8.2 of Doss and Wellner (2013) for the proof.

We will generally apply $d_{r,k,h}$ to concave functions, but in some instances it will be useful to take $h(x) = \mathrm{Id}(x) := x$ and apply $d_{r,k,Id}$ to log-concave functions (or brackets thereof), which yields a truncated version of L_r distance. The next results provide the motivation for using these new metrics $d_{r,k,h}$.

Lemma 3.2.25. Let h be a concave-function transformation, let $\underline{y}_k$ be as in Definition 3.2.23, let $\varphi_i \in \mathcal{C}_k$, and take $r \geq 1$. Then

$$d_{r,(h(y_0),h(y_k)),\mathtt{Id}}(h \circ \varphi_1, h \circ \varphi_2) \leq d_{r,k,h}(\varphi_1, \varphi_2), \tag{3.29}$$

and

$$\|(h \circ \varphi_1 - h \circ \varphi_2) 1_{\mathrm{dom}\,\varphi_1 \cap \mathrm{dom}\,\varphi_2}\|_r \leq d_{r,k,h}(\varphi_1, \varphi_2). \tag{3.30}$$

This means we can control entropy of the classes $\mathcal{F}_{k,h}$ of log-concave functions in terms of entropy with the $d_{\underline{y},h}$ metric on concave functions.

Lemma 3.2.26. Fix $1 \leq r < \infty$. Let $-\infty < b_1 < b_2 < \infty$, $-\infty < B_1 < B_2 < \infty$, and let $[B_1, B_2] \subseteq (\tilde{y}_0, \tilde{y}_\infty)$. Then

$$\begin{aligned} N_{[\,]}(\epsilon, \mathcal{F}_{k,h}, L_r(\lambda)) &\leq N_{[\,]}(\epsilon, \mathcal{F}_{k,h}, d_{r,(h(y_0),h(y_k)),\mathtt{Id}}) \\ &\leq N_{[\,]}(\epsilon, \mathcal{C}_k, d_{r,k,h}), \end{aligned} \tag{3.31}$$

for all $\epsilon > 0$. Moreover, setting $w_1 = \sup_{y \in [B_1, B_2]} |h'(y)|$, we have

$$\begin{aligned} & N_{[\,]}(\epsilon, h \circ \mathcal{C}([b_1, b_2], [B_1, B_2], 0), L_r(\lambda)) \\ & \quad \leq N_{[\,]}(\epsilon, \mathcal{C}([b_1, b_2], [B_1, B_2], 0), d_{r,(B_2,B_1),h}) \\ & \quad \leq N_{[\,]}(\epsilon/w_1, \mathcal{C}([b_1, b_2], [B_1, B_2], 0), L_r(\lambda)) \\ & \quad \leq \exp\left(c \left(\frac{(B_2 - B_1)(b_2 - b_1)^{1/r}}{\epsilon/w_1} \right)^{1/2} \right), \end{aligned} \tag{3.32}$$

for all $\epsilon \leq \epsilon_0 w_1(B_2 - B_1)(b_2 - b_1)^{1/r}$ where ϵ_0 is the constant depending only on r taken from Proposition 3.2.14.

Remark 3.2.27. If we take functions f_1 and f_2 which take values in $[y_{k-1}, y_0]$, then from the definitions of $d_{r,k,h}$ and $d_{r,k-1,h}$, we can see that

$$d_{r,k,h}(f_1, f_2) = d_{r,k-1,h}(f_1, f_2). \tag{3.33}$$

This is because it is true pointwise, i.e. $d_{k,h}(p_1, p_2) = d_{k-1,h}(p_1, p_2)$ if p_1 and p_2 are in $[y_{k-1}, y_0]$, since the corresponding measures $w_{\underline{y}_k,h}$ and $w_{\underline{y}_{k-1},h}$ are the same on $[y_{k-1}, y_0]$.

Similarly, since the measure W corresponding to $d_{r,k,\mathtt{Id}}$ is Lebesgue measure λ on $[y_k, y_0]$, for two functions f_1 and f_2 that take values in $[y_k, y_0]$, we have

$$\|f_1 - f_2\|_r^r = d_{r,k,\mathtt{Id}}(f_1, f_2). \tag{3.34}$$

Remark 3.2.28. The space $(\mathcal{F}_{k,h}, d_{r,(h(y_0),h(y_k)),\mathtt{Id}})$ corresponds to the space $(\mathcal{C}_k, d_{r,\underline{y}_k,h})$ (for any decreasing vector $\underline{y}_k$). The metric on the former space is not quite the $L_r(\lambda)$ metric because distance is truncated when functions fall below the cutoff $h(y_k)$. Thus to get from $d_{r,(h(y_0),h(y_k)),\mathtt{Id}}(f_1, f_2)$ to $\|f_1 - f_2\|_r$, we need to also control the difference in the domains of f_1 and f_2.

We can now state the main technical proposition needed for the bracketing entropy bound on $\mathcal{F}_{k,h}$. For now, we do not make any of (T.1)–(T.4) of Assumption B on h, which may mean that some of the weights $w_{\gamma,h}$ are infinite; in such a case, the conclusion of the following proposition is tautological.

Proposition 3.2.29. *Let h be as in Definition 3.2.15, let $\underline{y}_k \in \mathbb{R}^{k+1}$ as in Definition 3.2.23, and let $d_{r,k,h}$ and its corresponding weight sequence $w_{\gamma,h}$ be as in Definition 3.2.24. Fix $\epsilon > 0$*

and assume that for all $1 \leq \gamma \leq k$, y_γ satisfy

$$\epsilon / w_{\gamma,h} \leq \epsilon_0 (b_2 - b_1)^{1/r} (y_{\gamma-1} - y_\gamma), \tag{3.35}$$

for $\epsilon_0 > 0$ a positive constant (not necessarily defined as in Proposition 3.2.14). Then we have for any $\zeta \in [0,1]$, $1 \leq r < \infty$,

$$\begin{aligned} \log N_{[]} & \left(\epsilon \cdot \left(1 + \sum_{\gamma=1}^{k} h(y_{\gamma-1})^{r(1-\zeta)} \right)^{1/r}, \mathcal{F}_{k,h}, L_r(\lambda) \right) \\ & \leq \; c \sum_{\gamma=1}^{k} \left\{ \left(\frac{(y_{\gamma-1} - y_\gamma)(b_2 - b_1)^{1/r}}{\epsilon / w_{\gamma,h}} \right)^{1/2} \right. \\ & \qquad \left. + \; 2 \; \log \left(1 + \frac{b_2 - b_1}{\epsilon^r / h(y_{\gamma-1})^{r\zeta}} \right) \right\}. \end{aligned} \tag{3.36}$$

The condition (3.35) is not fundamental. It is essentially keeping ϵ from being too large, which is an unimportant constraint. The proposition could be phrased without this condition, but we phrase it with the condition and then pick $\underline{y}_k$ sequences later that satisfy it.

3.2.7 Bracketing entropy bounds: putting the pieces together

Bracketing results

We use the above result to prove an actual bracketing entropy bound of the type in which we are interested. Recall, for $b_1 < b_2$ and $B > 0$, Definition 3.2.12 of $\mathcal{F}([b_1, b_2], [0, B], \infty)$, and recall the definition of $\mathcal{D}(b_1)$ in Definition 3.2.17.

Theorem 3.2.30. *Let $r \geq 1$. Assume h is a concave-function transformation and that*

Assumptions (T.1) *and* (T.2) *hold, and let*

$$\mathcal{G} \equiv \mathcal{F}_h([b_1, b_2], [0, B], \infty) \cap h \circ \mathcal{D}(b_1).$$

Assume h *is continuous. For some* $\epsilon_0 > 0$ *and all* $\epsilon \leq \epsilon_0 B(b_2 - b_1)^{1/r}$, *we have*

$$\log N_{[\,]}(\epsilon, \mathcal{G}, L_r(\lambda)) \lesssim \left(\frac{B(b_2 - b_1)^{1/r}}{\epsilon} \right)^{1/2}, \tag{3.37}$$

where $\lesssim$ *means* $\leq$ *up to a constant depending only on* r *and* h. *Similarly,* ϵ_0 *is a constant depending on* r *and* h.

Rate results

We now use Theorem 3.2.30 to control the bracketing entropy for the log-concave classes in which we are really interested, i.e. those to which the Maximum Likelihood Estimator (MLE) $\hat{p}_n$ belongs with high probability, and to establish Hellinger rates of convergence.

Similarly to our previous definitions, we define

$$\mathcal{P}_h := \{h \circ \mathcal{C}\} \cap \left\{ p : \int p \, d\lambda = 1 \right\},$$

the class of h-concave-transformed densities, and we extend the definitions (3.19) and (3.38) to an arbitrary concave-function transformation h as follows.

$$\mathcal{P}_{1,M,h} \equiv \left\{ \begin{array}{rl} p \in \mathcal{P}_h : & m_p \in [-1, 1], \ 0 < p(0), \\ & 1/M \leq p(m_p) \leq M \end{array} \right\}. \tag{3.38}$$

As with the analogous classes of log-concave and the s-concave densities, the class $\mathcal{P}_{1,M,h}$ has an upper envelope, given in the following proposition.

Proposition 3.2.31. *Let h be a concave-function transformation such that Assumption* (T.1) *holds with exponent $\alpha = -1/t$ where $-1 < t \le 0$. Then for any $p \in \mathcal{P}_{1,M,h}$ with $0 < M < \infty$,*

$$
\begin{aligned}
p(x) &\le \left\{ \begin{array}{ll} D\left(-h^{-1}(M) + \frac{L}{2M}(x-1)\right)^{1/t}, & x \ge 2M+1, \\ D\left(-h^{-1}(M) + \frac{L}{2M}|x+1|\right)^{1/t}, & x \le -(2M+1), \\ M, & \textit{otherwise}, \end{array} \right\} \\
&\equiv p_{u,1,h}(x),
\end{aligned} \tag{3.39}
$$

where $0 < D, L < \infty$ are constants depending only on h and M.

For our asymptotic results, we make the following assumption:

Assumption C. *We assume that X_i, $i = 1, \ldots, n$ are i.i.d. random variables with distribution P_0 having density $p_0 = h \circ \varphi_0 \in \mathcal{P}_h$ with respect to Lebesgue measure, where φ_0 is concave.*

We can now state and prove our main theorem.

Theorem 3.2.32. *Assume that $h^{1/2}$ is a concave-function transformation and that Assumption B,* (T.1)–(T.4)*, hold for $\sqrt{h}$, where the exponent α in* (T.1) *satisfies $\alpha > 5/2$. Assume h is continuous. Suppose Assumption C holds and suppose that $\widehat{p}_n$ is the concave-transformed MLE of p_0. Then*

$$H(\widehat{p}_n, p_0) = O_p(n^{-2/5}). \tag{3.40}$$

The following corollaries connect the general Theorem 3.2.32 with Theorem 3.2.6 and Examples 3.2.18, 3.2.19, and 3.2.20.

Corollary 3.2.33. *Suppose that p_0 in Assumption C is log-concave; that is, $p_0 = h_0 \circ \varphi_0$ with $h_0(y) = e^y$ as in Example 3.2.18 and φ_0 concave. Then $H(\widehat{p}_n, p_0) = O_p(n^{-2/5})$.*

Corollary 3.2.34. *Suppose that p_0 in Assumption C is s-concave with $-1/5 < s < 0$; that is, $p_0 = h_s \circ \varphi_0$ with $h_s(y) = (-y)^{1/s}$ for $y < 0$ as in Example 3.2.19 with $-1/5 < s < 0$ and φ_0 concave. Then $H(\widehat{p}_n, p_0) = O_p(n^{-2/5})$.*

Corollary 3.2.35. *Suppose that p_0 in Assumption C is h-concave where h is a concave tranformation satisfying (T.1) and (T.4) of Assumption B, and that h satisfies $h = h_2 \circ \Psi$ where Ψ is concave and h_2 is a concave transformation for which $\sqrt{h_2}$ satisfies Assumption C, (T.1)–(T.4), where the exponent α in (T.1) satisfies $\alpha > 5/2$. Then, if $\widehat{p}_n$ is the concave transformed MLE of p_0, $H(\widehat{p}_n, p_0) = O_p(n^{-2/5})$. In particular the conclusion holds for $h = h_s$ given by $h_s(y) = y_+^{1/s}$ with $s > 0$.*

Theorem 3.2.32 has further corollaries, for example via Examples 3.2.21 and 3.2.22. We do not yet know if the hypothesis $s > -1/5$ in Corollary 3.2.34 can be improved to (say) $s > -1/2$ or $s > -1$.

Proof of Theorem 3.2.32. **Step 1: Reduction from $\mathcal{P}_h$ to $\mathcal{P}_{1,M,h}$.** We first show that we may assume, without loss of generality, that $p_0 \in \mathcal{P}_{1,M,h}$ for some $M > 0$. To see this, consider translating and rescaling the data: we let $\tilde{X}_i \equiv cX_i + b$ with $c > 0$ and $b \in \mathbb{R}$ so that each $\tilde{X}_i$ has density $\tilde{p}_0(x) = p_0((x - b)/c)/c$. To choose b and $c > 0$ so that $\tilde{p}_0 \in \mathcal{P}_{1,M,h}$ we argue as follows. For definiteness, let $m_p \equiv \inf\{m : \ m \text{ is a mode of } \ p\}$. Now $m_0 = m_{p_0}$ is in the support of p_0. (Note that m_0 in the boundary of the support of p_0 is possible.) Furthermore, there exists a point $a_0 \neq m_0$ in the interior of the support of p_0 (since otherwise the support of p_0 is degenerate and any such p_0 is not a density). Thus there exists a closed interval $[c_0, d_0]$ with $m_0, a_0 \in (c_0, d_0)$. Without loss of generality we may assume that $m_0 < a_0$ and we may take $d_0 = a_0 + 2(a_0 - m_0) = 3a_0 - 2m_0$, $c_0 = m_0 - (a_0 - m_0) = 2m_0 - a_0$. Thus $a_0 = (c_0 + d_0)/2$ and $m_0 = (3c_0 + d_0)/4$. Then it is easily seen that if $c = 1/(d_0 - a_0)$ and $b = -ca_0$ we have

$$\tilde{p}_0(0) = p_0(a_0)/c > 0, \qquad \tilde{p}_0(-1/2) = p_0(m_0)/c, \quad \text{and}$$
$$\tilde{p}_0(1) = p_0(d_0)/c > 0, \qquad \tilde{p}_0(-1) = p_0(c_0)/c,$$

so that $\tilde{m}_0 \equiv m_{\tilde{p}_0} = -1/2$ is the (smallest) mode of $\tilde{p}_0$.

For the rescaled data the MLE satisfies $\widehat{p}_n((x-b)/c; \underline{X})/c = \widehat{p}_n(x; \underline{\tilde{X}})$, and since the Hellinger metric is invariant under affine transformations, it follows that

$$H(\widehat{p}_n(\cdot; \underline{X}), p_0) = H(\widehat{p}_n(\cdot; \underline{\tilde{X}}), \tilde{p}_0).$$

Hence if (3.40) holds for $\tilde{p}_0$ and the transformed data, it also holds for p_0 and the original data. Thus we can henceforth assume that $p_0 \in \mathcal{P}_{1,M,h}$ for any $M > p_0(m_{p_0})$.

Now by the consistency results in Theorems 3.2.2 and 3.2.3 (and the general version of the latter in Theorem 2.17 of Seregin and Wellner (2010) which holds under their assumptions (D.1)–(D.4) and which are in turn implied by our (T.1)–(T.4)) for $g \equiv h^{1/2}$, it follows that $H(\hat{p}_n, p_0) \to_{a.s.} 0$, and we also have uniform convergence of $\widehat{p}_n$ to p_0 on compact subsets strictly contained in the support of p_0. Since $p_0 \in \mathcal{P}_{1,M,h}$ for any $M > p_0(m_{p_0})$, it follows that

$$P_0(\widehat{p}_n \in \mathcal{P}_{1,M,h}) \to 1 \quad \text{as} \quad n \to \infty$$

for all $M > M(p_0) \geq p_0(m_{p_0})$ where $M(p_0)$ is as defined in Theorem 3.2.2 part (v) when $s = 0$ and Theorem 3.2.3 part (iv). This completes step 1.

Step 2. Control of Hellinger bracketing entropy for $\mathcal{P}_{1,M,h}$ suffices. Step 2a: For $\delta > 0$, let

$$\overline{\mathcal{P}}_h(\delta) \equiv \{(p + p_0)/2 : \; p \in \mathcal{P}_h, \; H((p + p_0)/2, p_0) < \delta\}. \tag{3.41}$$

Suppose that we can show that

$$\log N_{[\,]}(\epsilon, \overline{\mathcal{P}}_h(\delta), H) \lesssim \epsilon^{-1/2} \tag{3.42}$$

for all $0 < \delta \le \delta_0$ for some $\delta_0 > 0$. Then it follows from van der Vaart and Wellner (1996), Theorems 3.4.1 and 3.4.4 (with $p_n = p_0$ in Theorem 3.4.4) or, alternatively, from van de Geer (2000), Theorem 7.4 and an inspection of her proofs, that any r_n satisfying

$$r_n^2 \Psi(1/r_n) \le \sqrt{n} \tag{3.43}$$

where

$$\Psi(\delta) \equiv J_{[\,]}(\delta, \overline{\mathcal{P}}_h(\delta), H)\left(1 + \frac{J_{[\,]}(\delta, \overline{\mathcal{P}}_h(\delta), H)}{\delta^2\sqrt{n}}\right)$$

and

$$J_{[\,]}(\delta, \overline{\mathcal{P}}_h(\delta), H) \equiv \int_0^\delta \sqrt{\log N_{[\,]}(\epsilon, \overline{\mathcal{P}}_h(\delta), H)}\,d\epsilon$$

gives a rate of convergence for $H(\widehat{p}_n, p_0)$. But it is easily seen that if (3.42) holds, then $r_n = n^{2/5}$ satisfies (3.43), and hence (3.40) holds.

Step 2b. Thus we want to show that (3.42) holds if we have an appropriate bracketing entropy bound for $\mathcal{P}_{1,M,h}$. First note that

$$N_{[\,]}(\epsilon, \overline{\mathcal{P}}_h(\delta), H) \le N_{[\,]}(\epsilon, \mathcal{P}_h(4\delta), H)$$

in view of van der Vaart and Wellner (1996), exercise 3.4.4 (or van de Geer (2000)), Lemma

4.2, page 48). Furthermore,

$$N_{[\,]}(\epsilon, \mathcal{P}_h(4\delta), H) \le N_{[\,]}(\epsilon, \mathcal{P}_{1,M,h}, H)$$

since $\mathcal{P}_h(4\delta) \subset \mathcal{P}_{1,M,h}$ for all $0 < \delta \le \delta_0$ with $\delta_0 > 0$ sufficiently small. This holds since Hellinger convergence implies pointwise convergence for concave transformed functions which in turn implies uniform convergence on compact subsets of the domain of p_0 via Rockafellar (1970), Theorem 10.8. See Lemma 8.1 of Doss and Wellner (2013) for details of the proofs.

Finally, note that

$$\begin{aligned} N_{[\,]}(\epsilon, \mathcal{P}_{1,M,h}, H) &= N_{[\,]}(\epsilon, \mathcal{P}^{1/2}_{1,M,h}, L_2(\lambda/2)) \\ &= N_{[\,]}(\epsilon, \mathcal{P}^{1/2}_{1,M,h}, L_2(\lambda)/\sqrt{2}) = N_{[\,]}(\epsilon/\sqrt{2}, \mathcal{P}^{1/2}_{1,M,h}, L_2(\lambda)) \end{aligned}$$

by the definition of H and $L_2(\lambda)$. Thus it suffices to show that

$$\log N_{[\,]}(\epsilon, \mathcal{P}^{1/2}_{1,M,h}, L_2(\lambda)) \lesssim \frac{1}{\epsilon^{1/2}} \tag{3.44}$$

where the constant involved depends only on M and h. This completes the proof of Step 2.

Step 3: Proof of the bracketing bound (3.44).

In fact we will show that

$$\log N_{[\,]}(\epsilon, \mathcal{P}^{1/2}_{1,M,h}, L_r(Q)) \lesssim \frac{1}{\epsilon^{1/2}}, \tag{3.45}$$

for any measure Q with Lebesgue-density q where q is bounded above by 1 and where the constants in $\lesssim$ depend only on r, M and h. (Note that $q \equiv 1$ works.) Now (3.45) holds if it holds when we replace $\mathcal{P}^{1/2}_1 \equiv \mathcal{P}^{1/2}_{1,M,h}$ by the two classes $\mathcal{P}^{1/2}_1|_{(-\infty,0]}$ and $\mathcal{P}^{1/2}_1|_{[0,\infty)}$.

This is because if we have $\exp(c_1/\epsilon^{1/2})$ brackets $[l_{n_1}, u_{n_1}]$ for $\mathcal{P}_1^{1/2}|_{(-\infty,0]}$ and $\exp(c_2/\epsilon^{1/2})$ brackets $[l_{n_2}, u_{n_2}]$ for $\mathcal{P}_1^{1/2}|_{[0,\infty)}$, where $c_i > 0$ are constants depending only on p_0 and on M, and where we let n_i range from 1 to the appropriate index, for $i = 1, 2$, then we can define functions

$$u_{n_1,n_2}(x) = u_{n_1}(x)1_{(-\infty,0)}(x) + u_{n_2}(x)1_{[0,\infty)}(x)$$

and similarly define l_{n_1,n_2} via l_{n_1} and l_{n_2}. These form brackets for $\mathcal{P}_1^{1/2}$; there are no more than $\exp((c_1 + c_2)/\epsilon^{1/2})$ of them; and their size is no larger than $2^{1/r}\epsilon$.

These two intervals are symmetric to each other, so we will only consider the restriction to $[0, \infty)$. We will use the method of Theorem 2.7.4 of van der Vaart and Wellner (1996) with Theorem 3.2.30. We first partition $[0, \infty)$ into intervals $I_j = [j, j+1]$ of length 1, $j \in \mathbb{N}$ and consider $\mathcal{F}_j^{1/2}$, the restriction of $\mathcal{P}_{1,M,h}^{1/2}$ to those intervals.

Thus by the envelope $p_{u,1,\sqrt{h}}$ for $\mathcal{P}_{1,M,\sqrt{h}}$ defined in (3.39), $\mathcal{F}_j^{1/2} \subset h^{1/2} \circ \mathcal{C}(I_j, [0, B_j], \infty)$ where $B_j := K(j-1)^{-\alpha}$ (with $\alpha = -1/t$) for $j \geq 2$ and B_0, B_1, and K are given by a constant depending on M and h. By definition, all $p \in \mathcal{P}_{1,M,h}$ satisfy $p(0) > 0$, which means $p|_{I_j} \in h \circ \mathcal{D}(j)$, so we can apply Theorem 3.2.30.

Now, we let a_j be any sequence of numbers in $(0, \infty]$. We fix $\epsilon > 0$ and for each $j \in \mathbb{N}$ take an ϵa_j-bracket $[l_{j,1}, u_{j,1}], \ldots, [l_{j,p_j}, u_{j,p_j}]$ for $\sqrt{h} \circ \mathcal{C}(I_j, [0, (\sqrt{h})^{-1}(B_j)], \infty)$ for the $L_r(\lambda)$ norm on I_j. By Theorem 3.2.30, the p_j's satisfy

$$\log p_j \lesssim \left(\frac{B_j \cdot 1}{\epsilon a_j} \right)^{1/2}. \tag{3.46}$$

If $\epsilon a_j \geq B_j$ then we can take $p_j = 1$. We form brackets

$$\left[\sum_j l_{j,i_j} 1_{I_j}, \sum_j u_{j,i_j} 1_{I_j} \right] \tag{3.47}$$

where i_j range over all possible values $1, \ldots, p_j$. The number of brackets is bounded by $\prod_j p_j$. Now, the $L_r(Q)$ size of a bracket $[l, u]$ defined above is $\|u - l\|_{r,Q}$ where

$$\begin{aligned}\|u - l\|_{r,Q}^r &= \int_{x_0}^{\infty} |u - l|^r q d\lambda \leq \sum_j \int_{I_j} \|q\|_{\infty,j} |u(x) - l(x)|^r dx \\ &= \sum_j \|q\|_{\infty,j} \|u - l\|_{r,\lambda}^r \leq \sum_j \|q\|_{\infty,j} a_j^r \epsilon^r \\ &\leq \epsilon^r \sum_j \|q\|_{\infty,j} a_j^r.\end{aligned}$$

Thus,

$$\log N_{[]} \left(\epsilon \left(\sum_j \|q\|_{j,\infty} a_j^r \right)^{1/r}, \mathcal{P}_{1,M,h}^{1/2}|_{[0,\infty)}, L_r(Q) \right) \lesssim \sum_j \left(\frac{B_j}{\epsilon a_j} \right)^{1/2}. \tag{3.48}$$

The choice

$$a_j^{r+1/2} = \frac{B_j^{1/2}}{\|q\|_{\infty,j}} =: \frac{B_j^{1/2}}{q_j}, \tag{3.49}$$

or $a_j = B_j^{1/(1+2r)}/q_j^{2/(1+2r)}$, reduces both sums in (3.48) to $\sum_j B_j^{r/(1+2r)} q_j^{1/(1+2r)}$, by the computations

$$a_j^r q_j = B_j^{r/(1+2r)} q_j^{1-2r/(1+2r)} = B_j^{r/(1+2r)} q_j^{1/(1+2r)}$$

and

$$\left(\frac{B_j}{a_j} \right)^{1/2} = \left(\frac{B_j}{B_j^{1/(1+2r)}/q_j^{2/(1+2r)}} \right)^{1/2} = \left(B_j^{2r/(1+2r)} q_j^{2/(1+2r)} \right)^{1/2} = B_j^{r/(1+2r)} q_j^{1/(1+2r)}.$$

Thus, for the moment denoting $\sum_j B_j^{r/(1+2r)} q_j^{1/(1+2r)}$ by the symbol S, (3.48) says

$$\log N_{[]} \left(\epsilon S^{1/r}, \mathcal{P}_{1,M,h}^{1/2}|_{[0,\infty)}, L_r(Q) \right) \lesssim \frac{1}{\epsilon^{1/2}} S, \tag{3.50}$$

so letting $v = \epsilon S^{1/r}$, we have

$$\log N_{[]}\left(v, \mathcal{P}^{1/2}_{1,M,h}|_{[0,\infty)}, L_r(Q)\right) \lesssim \frac{1}{v^{1/2}} S^{(1+2r)/(2r)}. \tag{3.51}$$

So we just need to show that $S = \sum_j B_j^{r/(1+2r)} q_j^{1/(1+2r)} < \infty$. Since we assumed $q_j \le 1$ and $B_j = K(j-1)^{-\alpha}$, the inequality $S < \infty$ holds where S depends only on r, M, and h, as long as $\alpha > (1+2r)/r$. This completes the proof of (3.45). Taking $r = 2$ and $Q = \lambda$ yields (3.44) for $\alpha > 5/2$ and completes the proof of the theorem. □

The following lemma allows us to extend our previous bracketing results, and so our rate results, to some cases where Assumption B does not hold for the transformation h_1, by allowing us to show that $\mathcal{P}_{h_1} \subseteq \mathcal{P}_{h_2}$ for a transformation h_2 such that Assumption B does hold for h_2.

Lemma 3.2.36. Let h_1 and h_2 be concave-function transformations. If Ψ is a concave function such that $h_1 = h_2 \circ \Psi$, then $\mathcal{P}_{h_1} \subseteq \mathcal{P}_{h_2}$.

Additionally, we can extend the above results to the case of the constrained log-concave MLE with fixed and known location of the mode. We stated the following theorem earlier as Theorem 3.2.9, along with the corollaries that follow, Corollary 3.2.10 and Corollary 3.2.11.

Theorem. *Suppose Assumption C holds where $p_0 = h_0 \circ \varphi_0$ and $h_0(y) = e^y$, and where the mode of p_0 is fixed and known. Let $\hat{p}_n^0$ be the (modally constrained) MLE of p_0. Then $H(\hat{p}_n^0, p_0) = O_p(n^{-2/5})$.*

Proof. The proof is based on the proof of Theorem 3.2.32. By Corollary 3.1.6, $\hat{p}_n^0$ is almost surely Hellinger consistent, and additionally we have almost sure uniform convergence of $\hat{p}_n^0$ on compact subsets strictly contained in the support of p_0, by Lemma 3.1.8. These are the only two properties needed for Step 1 of the proof of Theorem 3.2.32, which thus goes through.

Then, since $\sqrt{h_0}(y) = e^{y/2}$ satisfies the assumptions of Theorem 3.2.32, i.e. h_0 is continuous and Assumption B, (T.1)–(T.4), hold, where the exponent α in (T.1) can be any $\alpha > 5/2$, the remaining two steps of the proof of Theorem 3.2.32 go through without alteration. Thus, we are done. $\square$

Chapter 4

ASYMPTOTICS II: LOCAL

In the previous chapter we studied global (consistency) results. In this chapter our focus will be more local. In the Section 4.1, we will find lower bounds for estimation of the *location* of the mode, in the specific setting where φ_0 has a cusp at the mode. In Section 4.2 we will study our estimator at the mode. We will study both the behavior of the knots, finding that when $f_0''(m) < 0$ they are separated by $O_p(n^{-1/5})$, and use this to study the estimator $\hat{f}_n^0$ on neighborhoods shrinking towards the mode. In this chapter we will thus generally be continuing to operate under Assumption A, the null hypothesis, from page 45.

4.1 Lower Bounds for Estimation of the Location of the Mode

Has'minskii (1979) established a rate of $n^{-1/5}$ for an L_1 minimax lower bound for estimating the mode over unimodal densities, assuming the true density satisfies $f_0''(m) < 0$. Balabdaoui et al. (2009) establish a rate of $n^{-1/5}$ for an L_1 minimax lower bound for estimating the mode over log-concave densities, assuming $f_0''(m) < 0$. In addition, they demonstrate that over both log-concave and over unimodal densities, the bound depends on f_0 through the constant $\frac{f_0(m)}{f_0''(m)^2}$. We now establish an analogous bound for estimating the mode over log-concave densities, assuming that f_0 is cusp shaped. That is, we assume φ_0 is concave with unique mode at m, and

$$\varphi_0(x) = \varphi_0(m) - \gamma_+(x-m)_+ - \gamma_-(m-x)_+ + o(|x-m|), \tag{4.1}$$

where $\gamma_+ > 0$, $\gamma_- > 0$, and $f_0(x) = e^{\varphi_0(x)}$ is the corresponding density. To be precise, we will define two mode functionals by

$$M_I(f) = \inf\{x | f(x) \geq f(y) \text{ for all } y \in \mathbb{R}\} \tag{4.2}$$

$$M_S(f) = \sup\{x | f(x) \geq f(y) \text{ for all } y \in \mathbb{R}\}. \tag{4.3}$$

Recall that we defined Hellinger distance in (3.1). We will now use the local Hellinger balls about f_0 defined by

$$\mathcal{LC}_{n,\tau} := \left\{ g \in \mathcal{P} | H^2(g, f_0) \leq \tau/n \right\}. \tag{4.4}$$

Proposition 4.1.1. *Take $M_I(f)$, $M_S(f)$ and $\mathcal{LC}_{n,\tau}$ as defined in (4.2), (4.3), and (4.4), respectively. Assume (4.1) is satisfied by $f_0 \in \mathcal{P}_m$ and $M_I(f_0) = M_S(f_0) = m$. Then*

$$\sup_{\tau>0} \liminf_n \inf_{T_n} \sup_{f \in \mathcal{LC}_{n,\tau}} n^{1/3} E_f \left| T_n - M_I(f) \right| \geq \left(e \frac{4^3 \cdot 6}{12} \left(1 - \frac{\varphi_0'(m-)}{\varphi_0'(m+)} \right) \frac{f_0'(m-)^2}{f_0(m)} \right)^{-1/3} \tag{4.5}$$

and

$$\sup_{\tau>0} \liminf_n \inf_{T_n} \sup_{f \in \mathcal{LC}_{n,\tau}} n^{1/3} E_f \left| T_n - M_S(f) \right| \geq \left(e \frac{4^3 \cdot 6}{12} \left(1 - \frac{\varphi_0'(m+)}{\varphi_0'(m-)} \right) \frac{f_0'(m+)^2}{f_0(m)} \right)^{-1/3}, \tag{4.6}$$

where E_f denotes expectation taken with respect to n i.i.d. samples drawn from density f and where $\inf_{T_n}$ is the infimum over all estimators of the mode, i.e. all estimators of $M_I(f)$ or $M_S(f)$, respectively, based on $X_1, \ldots, X_n$.

The constant $\left(e \frac{4^3 \cdot 6}{12} \right)^{-1/3}$ is approximately equal to 0.179, and, note that by (4.1), we have $\varphi_0'(m+) = -\gamma_+$ and $\varphi_0'(m-) = \gamma_-$.

Proof. Without loss of generality we assume $m = 0$. Let us also assume for now that

$\gamma_- \geq \gamma_+ > 0$. We define a perturbation by flattening φ_0 on an interval about 0,

$$\varphi_{\epsilon,L}(x) = \begin{cases} \varphi_0(x) & \text{if } x \notin [-\epsilon\frac{\gamma_+}{\gamma_-}, \epsilon] \\ \varphi_0(\epsilon) & \text{if } x \in [-\epsilon\frac{\gamma_+}{\gamma_-}, \epsilon] \end{cases}.$$

Then define $h_{\epsilon,L} = e^{\varphi_{\epsilon,L}}$ and $f_{\epsilon,L}(x) = \frac{h_{\epsilon,L}(x)}{\int_{\mathbb{R}} h_{\epsilon,L}(u)du}$. We will use Lemma 4.1 in Groeneboom (1996) which concludes that

$$\inf_{T_n} \max\left\{E_{f_1}\left|T_n - T(f_1)\right|, E_{f_2}\left|T_n - T(f_2)\right|\right\} \geq \frac{1}{4}\left|T(f_1) - T(f_2)\right|\left(1 - H^2(f_1, f_2)\right)^{2n},$$

where f_1 and f_2 are densities and T is a functional (i.e. a parameter). We will first analyze the Hellinger distance and show that it satisfies

$$H^2(f_{\epsilon,L}, f_0) = f_0(0)\left(\frac{\gamma_+^3}{\gamma_-} + \gamma_+^2\right)\epsilon^3 + o(\epsilon^3)$$

as $\epsilon \to 0$. By Lemma 4.1.3, we only need to compute the Hellinger (squared) integral on the interval $[-\epsilon\frac{\gamma_+}{\gamma_-}, \epsilon]$. Now, by using Lemma 4.1.2 and the definition of $h_{\epsilon,L}$ for the second equality below, we see that as $\epsilon \to 0$,

$$\begin{aligned}\int_{-\epsilon\frac{\gamma_+}{\gamma_-}}^{\epsilon}\left(\sqrt{f_{\epsilon,L}(x)} - \sqrt{f_0(x)}\right)^2 dx &= \int_{-\epsilon\frac{\gamma_+}{\gamma_-}}^{\epsilon}\left(\sqrt{\frac{h_{\epsilon,L}(x)}{\int_{\mathbb{R}} h_{\epsilon,L}(u)du}} - \sqrt{f_0(x)}\right)^2 dx \\ &= \int_{-\epsilon\frac{\gamma_+}{\gamma_-}}^{0}\left(\sqrt{e^{\varphi_0(\epsilon)}}(1 + O(\epsilon^2)) - e^{\frac{1}{2}(\varphi_0(0) + x\gamma_- + o(x))}\right)^2 dx \\ &\quad + \int_0^{\epsilon}\left(\sqrt{e^{\varphi_0(\epsilon)}}(1 + O(\epsilon^2)) - e^{\frac{1}{2}(\varphi_0(0) - x\gamma_+ + o(x))}\right)^2 dx,\end{aligned}$$

which is equal to

$$\int_{-\epsilon\frac{\gamma_+}{\gamma_-}}^{0}\left(\sqrt{f_0(0)}e^{-\frac{1}{2}\epsilon\gamma_+}e^{o(\epsilon)}(1+O(\epsilon^2))-\sqrt{f_0(0)}e^{\frac{1}{2}x\gamma_-}e^{o(x)}\right)^2dx$$
$$+\int_0^{\epsilon}\left(\sqrt{f_0(0)}e^{-\frac{1}{2}\epsilon\gamma_+}e^{o(\epsilon)}(1+O(\epsilon^2))-\sqrt{f_0(0)}e^{-\frac{1}{2}x\gamma_+}e^{o(x)}\right)^2dx.$$

The above display then equals

$$f_0(0)e^{-\epsilon\gamma_+}e^{o(\epsilon)}\Bigg(\int_{-\epsilon\frac{\gamma_+}{\gamma_-}}^{0}\left(1+O(\epsilon^2)-e^{\frac{1}{2}(\epsilon\gamma_++x\gamma_-)}\right)^2dx$$
$$+\int_0^{\epsilon}\left(1+O(\epsilon^2)-e^{\frac{1}{2}(\epsilon\gamma_+-x\gamma_+)}\right)^2dx\Bigg),$$

which, by using the substitution $u=-x\frac{\gamma_-}{\gamma_+}$, equals

$$f_0(0)e^{-\epsilon\gamma_+}e^{o(\epsilon)}\Bigg(\int_0^{\epsilon}\left(1+O(\epsilon^2)-e^{\frac{1}{2}(\epsilon\gamma_+-u\gamma_+)}\right)^2\frac{\gamma_+}{\gamma_-}du$$
$$+\int_0^{\epsilon}\left(1+O(\epsilon^2)-e^{\frac{1}{2}(\epsilon\gamma_+-x\gamma_+)}\right)^2dx\Bigg),$$

which is

$$f_0(0)e^{-\epsilon\gamma_+}e^{o(\epsilon)}\left(\frac{\gamma_+}{\gamma_-}+1\right)\int_0^{\epsilon}\left(1+O(\epsilon^2)-e^{\frac{1}{2}(\epsilon\gamma_+-x\gamma_+)}\right)^2dx. \tag{4.7}$$

Now we analyze the integral, using the Taylor expansion

$$e^{\frac{1}{2}(\epsilon\gamma_+-x\gamma_+)}=1+\frac{1}{2}(\epsilon\gamma_+-x\gamma_+)+\frac{1}{4}e^{\xi_\epsilon}(\epsilon\gamma_+-x\gamma_+)^2,$$

where ξ_ϵ is between 0 and $\epsilon\gamma_+ - x\gamma_+$. We then can see

$$\begin{aligned}\int_0^\epsilon \left(1+O(\epsilon^2)-e^{\frac{1}{2}(\epsilon\gamma_+-x\gamma_+)}\right)^2 dx &= \int_0^\epsilon \left(O(\epsilon^2)-\left(\frac{\gamma_+}{2}(\epsilon-x)+\frac{\gamma_+^2}{4}e^{\xi_\epsilon}(\epsilon-x)^2\right)\right)^2 dx \\ &= \int_0^\epsilon \left(\left(\frac{\gamma_+}{2}\right)^2(x-\epsilon)^2+O(\epsilon^3)\right) dx \\ &= \frac{\gamma_+^2}{4\cdot 3}\epsilon^3+O(\epsilon^4),\end{aligned}$$

where, to get the second equality, we noted for ϵ small enough that $e^{\xi_\epsilon} \leq 2$ and did the elementary calculation

$$\begin{aligned}&\left(O(\epsilon^2)-\frac{\gamma_+}{2}(\epsilon-x)-\frac{\gamma_+^2}{4}e^{\xi_\epsilon}(\epsilon-x)^2\right)^2 \\ &\quad=\left(\frac{\gamma_+}{2}(\epsilon-x)\right)^2+\frac{2\gamma_+}{2}(\epsilon-x)\left(O(\epsilon^2)-\frac{\gamma_+^2}{4}e^{\xi_\epsilon}(\epsilon-x)^2\right) \\ &\quad\quad+\left(O(\epsilon^2)-\frac{\gamma_+^2}{4}e^{\xi_\epsilon}(\epsilon-x)^2\right)^2 \\ &\quad=\left(\frac{\gamma_+}{2}(\epsilon-x)\right)^2+O(\epsilon^3).\end{aligned}$$

Now, using $e^{O(\epsilon)} = 1 + O(\epsilon)$, we see that (4.7) equals

$$\begin{aligned}&f_0(0)e^{-\epsilon\gamma_+}e^{o(\epsilon)}\left(\frac{\gamma_+}{\gamma_-}+1\right)\int_0^\epsilon\left(1+O(\epsilon^2)-e^{\frac{1}{2}(\epsilon\gamma_+-x\gamma_+)}\right)^2 dx \\ &\quad=(1+O(\epsilon))\,f_0(0)\left(\frac{\gamma_+}{\gamma_-}+1\right)\left(\frac{\gamma_+^2}{4\cdot 3}\epsilon^3+O(\epsilon^4)\right) \\ &\quad=f_0(0)\left(\frac{\gamma_+}{\gamma_-}+1\right)\frac{\gamma_+^2}{12}\epsilon^3+O(\epsilon^4).\end{aligned}$$

Then, by Lemma 4.1.3, we conclude that

$$
\begin{aligned}
H^2(f_{\epsilon,L}, f_0) &= \int_{[-\epsilon\frac{\gamma_+}{\gamma_-},\epsilon]} \left(\sqrt{f_{\epsilon,L}(x)} - \sqrt{f_0(x)}\right)^2 dx + \int_{[-\epsilon\frac{\gamma_+}{\gamma_-},\epsilon]^c} \left(\sqrt{f_{\epsilon,L}(x)} - \sqrt{f_0(x)}\right)^2 dx \\
&= f_0(0)\left(\frac{\gamma_+}{\gamma_-} + 1\right)\frac{\gamma_+^2}{12}\epsilon^3 + O(\epsilon^4) + O(\epsilon^4) \\
&= f_0(0)\left(\frac{\gamma_+}{\gamma_-} + 1\right)\frac{\gamma_+^2}{12}\epsilon^3 + o(\epsilon^3),
\end{aligned}
$$

as we desired. To do this Hellinger distance computation we assumed $\gamma_- \geq \gamma_+$; if we assume the reverse, $\gamma_- \leq \gamma_+$, the argument is symmetric: we define

$$
\varphi_{\epsilon,R}(x) = \begin{cases} \varphi_0(x) & \text{if } x \notin [-\epsilon, \epsilon\frac{\gamma_-}{\gamma_+}] \\ \varphi_0(\epsilon) & \text{if } x \in [-\epsilon, \epsilon\frac{\gamma_-}{\gamma_+}] \end{cases},
$$

define $h_{\epsilon,R}$ and $f_{\epsilon,R}$ analogously, and proceed as above, since it was just an integral calculation, to get a symmetric result with the roles of γ_+ and γ_- reversed. Also, note that whether we define the mode to be the infimum or supremum of the modal interval obviously did not affect the integral calculation either. Thus, if we define

$$
f_\epsilon = \begin{cases} f_{\epsilon,L} & \text{if } \gamma_- \geq \gamma_+ \\ f_{\epsilon,R} & \text{if } \gamma_- \leq \gamma_+ \end{cases}, \tag{4.8}
$$

then we have shown

$$
H^2(f_\epsilon, f_0) = f_0(0)\left(\frac{\min(\gamma_+,\gamma_-)}{\max(\gamma_+,\gamma_-)} + 1\right)\frac{\min(\gamma_+,\gamma_-)^2}{12}\epsilon^3 + o(\epsilon^3). \tag{4.9}
$$

Now we note that

$$
|M_I(f_\epsilon) - M_I(f_0)| = \min\left\{\frac{\gamma_+}{\gamma_-}, 1\right\}\epsilon,
$$

and apply Lemma 4.1 from Groeneboom (1996) with $\epsilon = cn^{-1/3}$ for any $c > 0$ and $l(x) = |x|$ to see

$$\begin{aligned}
&\inf_{T_n} \max \left\{ E_{f_{cn^{-1/3}}} \left|T_n - M_I(f_{cn^{-1/3}})\right|, E_{f_0} \left|T_n - M_I(f_0)\right| \right\} \\
&\qquad \geq \frac{1}{4} \left|M_I(f_{cn^{-1/3}}) - M_I(f_0)\right| \left(1 - H^2(f_{cn^{-1/3}}, f_0)\right)^{2n} \\
&\qquad = \frac{1}{4} \min \left\{ \frac{\gamma_+}{\gamma_-}, 1 \right\} cn^{-1/3} \left(1 - H^2(f_{cn^{-1/3}}, f_0)\right)^{2n}.
\end{aligned}$$

Denote $\tilde{\rho} = f_0(0) \left(\frac{\min(\gamma_+,\gamma_-)}{\max(\gamma_+,\gamma_-)} + 1 \right) \frac{\min(\gamma_+,\gamma_-)^2}{12}$, and note that by (4.9),

$$\left(1 - H^2(f_{cn^{-1/3}}, f_0)\right)^{2n} = \left(1 - \tilde{\rho}\frac{c^3}{n} + o\left(\frac{1}{n}\right)\right)^{2n} \longrightarrow \exp\left\{-2\tilde{\rho}c^3\right\}$$

as $n \to \infty$. We have thus shown that

$$\begin{aligned}
&\liminf_n \inf_{T_n} \max n^{1/3} \left\{ E_{f_{cn^{-1/3}}} \left|T_n - M_I(f_{cn^{-1/3}})\right|, E_{f_0} \left|T_n - M_I(f_0)\right| \right\} \\
&\quad \geq \frac{1}{4} \min \left\{ \frac{\gamma_+}{\gamma_-}, 1 \right\} c e^{-2\tilde{\rho}c^3} \\
&\quad = \frac{1}{4} \min \left\{ \frac{\gamma_+}{\gamma_-}, 1 \right\} \left(\frac{1}{6\tilde{\rho}}\right)^{1/3} e^{-1/3}
\end{aligned} \tag{4.10}$$

where the last equality is true if we maximize over c by setting $c = \left(\frac{1}{6\tilde{\rho}}\right)^{1/3}$. (We find this value of c from the elementary set-derivative-to-0 calculation of $0 = e^{-2\tilde{\rho}c^3} + c(-6\tilde{\rho}c^2)e^{-2\tilde{\rho}c^3}$ which yields $6c^3\tilde{\rho} = 1$.) Now note that

$$\frac{1}{4} \min \left\{ \frac{\gamma_+}{\gamma_-}, 1 \right\} \left(\frac{1}{6\tilde{\rho}}\right)^{1/3} e^{-1/3} = \left(\frac{1}{4^3 \cdot 6 e f_0(0) \left(1 + \frac{\gamma_-}{\gamma_+}\right) \frac{\gamma_-^2}{12}} \right)^{1/3}. \tag{4.11}$$

This follows if $\gamma_- \geq \gamma_+$ by

$$\min\left\{\frac{\gamma_+}{\gamma_-},1\right\}\left(\left(\frac{\min(\gamma_+,\gamma_-)}{\max(\gamma_+,\gamma_-)}+1\right)\min(\gamma_+,\gamma_-)^2\right)^{-1/3} = \frac{\gamma_+}{\gamma_-}\left(\left(\frac{\gamma_+}{\gamma_-}+1\right)\gamma_+^2\right)^{-1/3} = \left(\left(1+\frac{\gamma_-}{\gamma_+}\right)\gamma_-^2\right)^{-1/3},$$

and if $\gamma_- \leq \gamma_+$ then it follows by

$$\min\left\{\frac{\gamma_+}{\gamma_-},1\right\}\left(\left(\frac{\min(\gamma_+,\gamma_-)}{\max(\gamma_+,\gamma_-)}+1\right)\min(\gamma_+,\gamma_-)^2\right)^{-1/3} = 1\cdot\left(\left(\frac{\gamma_-}{\gamma_+}+1\right)\gamma_-^2\right)^{-1/3}.$$

Now note that $f_0'(x-) = \varphi_0'(x-)e^{\varphi_0(x)}$, so

$$f_0'(0-)^2 = (\varphi_0(0-)f_0(x))^2 = \gamma_-^2 f_0(0)^2, \tag{4.12}$$

which says

$$\gamma_-^2 f_0(0) = \frac{f_0'(0-)^2}{f_0(0)}.$$

Thus, plugging (4.11) and (4.12) into (4.10), we have shown

$$\begin{aligned}
&\liminf_n \inf_{T_n} \max n^{1/3}\left\{E_{f_{cn^{-1/3}}}\left|T_n - M_I(f_{cn^{-1/3}})\right|, E_{f_0}\left|T_n - M_I(f_0)\right|\right\} \\
&\geq \left(4^3\cdot 6ef_0(0)\left(1+\frac{\gamma_-}{\gamma_+}\right)\frac{\gamma_-^2}{12}\right)^{-1/3} \\
&= \left(e\frac{4^3\cdot 6}{12}\left(1-\frac{\varphi_0'(0-)}{\varphi_0'(0+)}\right)\frac{f_0'(0-)^2}{f_0(0)}\right)^{-1/3}.
\end{aligned}$$

Thus we have shown (4.5). The argument to show (4.6) is entirely symmetric. □

Here are the lemmas we needed.

Lemma 4.1.2. Let h_ϵ be defined as $h_{\epsilon,L}$ if $\gamma_- \geq \gamma_+$, and as $h_{\epsilon,R}$ if $\gamma_- < \gamma_+$. Then

$$\int_{\mathbb{R}} h_\epsilon(x)dx = 1 + O(\epsilon^2) \tag{4.13}$$

and

$$\frac{1}{\sqrt{\int_{\mathbb{R}} h_\epsilon(x)dx}} = 1 + O(\epsilon^2). \tag{4.14}$$

Proof. We assume without loss of generality that $m = 0$, that $\gamma_- \geq \gamma_+$ and that thus h_ϵ is $h_{\epsilon,L}$, and we see that $\int_{\mathbb{R}} f_0(x)dx - \int_{\mathbb{R}} h_\epsilon(x)dx$ is equal to

$$\int_0^\epsilon \left(e^{\varphi_0(0)-x\gamma_+ + o(x)} - e^{\varphi_0(0)-\epsilon\gamma_+ + o(\epsilon)}\right)dx + \int_{-\epsilon\frac{\gamma_+}{\gamma_-}}^0 \left(e^{\varphi_0(0)+x\gamma_- + o(x)} - e^{\varphi_0(0)-\epsilon\gamma_+ + o(\epsilon)}\right)dx,$$

which is equal to

$$f_0(0)e^{o(\epsilon)}\left(\int_0^\epsilon \left(e^{-x\gamma_+} - e^{-\epsilon\gamma_+}\right)dx + \int_{-\epsilon\frac{\gamma_+}{\gamma_-}}^0 \left(e^{x\gamma_-} - e^{-\epsilon\gamma_+}\right)dx\right).$$

By computing the integrals, we see that this is equal to

$$f_0(0)e^{o(\epsilon)}\left(\frac{1}{\gamma_+}\left(1 - e^{-\epsilon\gamma_+}\right) - \epsilon e^{-\epsilon\gamma_+} + \frac{1}{\gamma_-}\left(1 - e^{-\epsilon\gamma_+}\right) - \epsilon\frac{\gamma_+}{\gamma_-}e^{-\epsilon\gamma_+}\right),$$

which is

$$f_0(0)e^{o(\epsilon)}\left(\frac{1}{\gamma_+} - \left(\frac{1}{\gamma_+} + \epsilon\right)e^{-\epsilon\gamma_+} + \frac{1}{\gamma_-} - \left(\frac{1}{\gamma_-} + \epsilon\frac{\gamma_+}{\gamma_-}\right)e^{-\epsilon\gamma_+}\right). \tag{4.15}$$

Then we expand $e^{-\epsilon\gamma_+}$ as $1 + -\epsilon\gamma_+ + e^{\xi}\frac{\epsilon^2\gamma_+^2}{2} = 1 + -\epsilon\gamma_+ + O(\epsilon^2)$ for some $\xi \leq 0$, and see that (4.15) is

$$f_0(0)e^{o(\epsilon)}\left(\frac{1}{\gamma_+}-\left(\frac{1}{\gamma_+}+\epsilon\right)\left(1+-\epsilon\gamma_++O(\epsilon^2)\right)+\frac{1}{\gamma_-}-\left(\frac{1}{\gamma_-}+\epsilon\frac{\gamma_+}{\gamma_-}\right)\left(1+-\epsilon\gamma_++O(\epsilon^2)\right)\right),$$

which is

$$\begin{aligned}f_0(0)e^{o(\epsilon)}&\left(\frac{1}{\gamma_+}-\left(\frac{1}{\gamma_+}-\epsilon+O(\epsilon^2)+\epsilon-\epsilon^2\gamma_++O(\epsilon^3)\right)\right.\\&\left.+\frac{1}{\gamma_-}-\left(\frac{1}{\gamma_-}-\epsilon\frac{\gamma_+}{\gamma_-}+O(\epsilon^2)+\epsilon\frac{\gamma_+}{\gamma_-}-\epsilon^2\frac{\gamma_+^2}{\gamma_-}+O(\epsilon^3)\right)\right),\end{aligned}$$

and this is equal to $f_0(0)e^{o(\epsilon)}O(\epsilon^2)=O(\epsilon^2)$. Thus $\int_{\mathbb{R}}h_\epsilon(x)dx=1-O(\epsilon^2)$, which is (4.13).

Next we will show that

$$\frac{1}{\sqrt{1+O(\epsilon^2)}}=1+O(\epsilon^2). \tag{4.16}$$

This follows again by Taylor expansions about 1. Let $g(x)=x^{-1/2}$, so $g^{(1)}(x)=\frac{-1}{2}x^{-3/2}$, and $g^{(2)}(x)=\frac{3}{4}x^{-5/2}$, so with $\xi\in[1-|O(\epsilon^2)|,1+|O(\epsilon^2)|]$,

$$\frac{1}{\sqrt{1+O(\epsilon^2)}}=1+\frac{-1}{2}O(\epsilon^2)+\frac{3}{4}\xi^{-5/2}\frac{O(\epsilon^4)}{2}=1+O(\epsilon^2).$$

Then (4.14) follows immediately from (4.13) and (4.16). □

Lemma 4.1.3. Let f_ϵ be as in (4.8). Then

$$\int_{(-\infty,a-\epsilon\frac{\gamma_+}{\gamma_-}]\cup[a+\epsilon,\infty)}\left(\sqrt{f_\epsilon(x)}-\sqrt{f_0(x)}\right)^2dx=O(\epsilon^4).$$

Proof. Denote $I_\epsilon=[a-\epsilon\frac{\gamma_+}{\gamma_-},a+\epsilon]$. Then

$$\int_{I_\epsilon^c}\left(\sqrt{f_\epsilon(x)}-\sqrt{f_0(x)}\right)^2dx=\int_{I_\epsilon^c}\left(\sqrt{\frac{h_\epsilon}{\int h_\epsilon(u)du}}-\sqrt{f_0(x)}\right)^2dx,$$

which by the definition of h_ϵ equals

$$\begin{aligned}\int_{I_\epsilon^c}\left(\sqrt{\frac{f_0(x)}{\int h_\epsilon(u)du}}-\sqrt{f_0(x)}\right)^2 dx &= \int_{I_\epsilon^c} f_0(x)\left(\sqrt{\frac{1}{\int h_\epsilon(u)du}}-1\right)^2 dx \\ &= \left(1+O(\epsilon^2)-1\right)^2\int_{I_\epsilon^c} f_0(x)dx \\ &= O(\epsilon^4),\end{aligned}$$

where the second-to-last equality is by Lemma 4.1.2 and the last because $0 \le \int_{I_\epsilon^c} f(x)dx \le 1$. □

4.2 Upper Bounds and Tightness for the CMLE at the Mode

In this section we study (upper bounds for) rates of convergence of the constrained MLE. We will give results both away from the mode and at the mode, although the main focus and the case requiring the most work will be the latter case. Before studying the value of the CMLE at a given point, we study what has been called "the Gap problem" (Balabdaoui and Wellner, 2007), which is the distance of the estimator's knots on either side of a fixed point. Control of the size of the gap between the knots can then be translated into control of the value of the CMLEs ($\hat{\varphi}_n^0$ and $\hat{f}_n^0$) both in relation to the unconstrained MLEs ($\hat{\varphi}_n$ and $\hat{f}_n$) and to the true functions (φ_0 and f_0). The results in this section will be of fundamental importance to the local asymptotic results given in Chapter 5.

As in the previous section, we will work in the setting in which Assumption A, the null hypothesis, is satisfied. In contrast to the previous section, in which we gave rate lower bounds in the setting wherein the true log-density has a cusp at the mode, in this section we will study (upper bounds in) the setting in which the second derivative of φ_0 (and thus f_0) is strictly negative at a given point. Thus, in addition to Assumption A, we will generally make the following assumption in this section.

Assumption D. *We assume that φ_0 is twice continuously differentiable in a neighborhood of m (and thus f_0 also is), and that $\varphi_0''(m) < 0$ (and thus $f_0''(m) < 0$). In particular, m is an interior point of the support of f_0.*

Note, of course, that since exp is continuously differentiable on $(-\infty, \infty)$, when φ_0 is twice continuously differentiable and bounded away from 0, f_0 is also twice continuously differentiable (and the converse holds, by consideration of log). Similarly, since $f_0''(x) = e^{\varphi_0(x)}(\varphi_0'(x))^2 + e^{\varphi_0(x)}\varphi_0''(x)$, and $\varphi_0'(m) = 0$, we have that $\varphi_0''(m) < 0$ if and only if $f_0''(m) < 0$.

4.2.1 The Gap Problem

In this section we study the gap problem. For a fixed point x_0, we define $\tau_{n,-}(x_0)$ to be the closest knot of $\hat{\varphi}_n$ to x_0 that is less than x_0, and, similarly, $\tau_{n,+}(x_0)$ to be the closest knot to x_0 that is larger than x_0. Similarly, we define $\tau^0_{n,-}(x_0)$ and $\tau^0_{n,+}(x_0)$ to be the closest knots of $\hat{\varphi}^0_n$ that are smaller than x_0 and larger than x_0, respectively. Here, our convention is that $\tau^0_{n,+}(x_0)$ and $\tau^0_{n,-}(x_0)$ do *not* equal m, which is somewhat different than our standard convention of treating the mode like a knot. The first result we give is for the UMLE.

Proposition 4.2.1 (The Unconstrained Gap Problem)**.** *Let X_i be i.i.d. from f_0, where $f_0 \in \mathcal{P}$. Also assume f_0 is twice continuously differentiable at the point x_0, and $f_0''(x_0) < 0$. Then*

$$\tau_{n,+}(\xi_n) - \tau_{n,-}(\xi_n) = O_p(n^{-1/5})$$

where ξ_n are any real random variables satisfying $\xi_n \to_p x_0$.

Proposition 4.2.2 (The Constrained Gap Problem, away from the mode)**.** *Let Assumption A hold. Also assume f_0 is twice continuously differentiable at the point x_0 and that*

$f_0''(x_0) < 0$, where $x_0 \neq m$. Then

$$\tau_{n,+}^0(\xi_n) - \tau_{n,-}^0(\xi_n) = O_p(n^{-1/5})$$

where ξ_n are any real random variables satisfying $\xi_n \to_p x_0$.

Proof of 4.2.1 and 4.2.2. The proofs of Propositions 4.2.1 and 4.2.2 are the same, and they are the same as the proof of Theorem 4.3 in Balabdaoui et al. (2009). Both propositions rely on noting that since $f''(x_0) < 0$, $\tau_{n,+}^0(\xi_n) - \tau_{n,-}^0(\xi_n) \to_p 0$ and $\tau_{n,+}(\xi_n) - \tau_{n,-}(\xi_n) \to_p 0$, otherwise our density estimators could not be consistent, i.e. could not satisfy Corollary 3.1.9.

The only difference in Proposition 4.2.1 and Theorem 4.3 in Balabdaoui et al. (2009) is that in the latter ξ_n are taken fixed and equal to x_0, whereas we allow ξ_n to be any real random variables satisfying $\xi_n \to_p x_0$. This is the formulation given for the analogous result in the convex density estimation problem by Lemma 4.2 (pages 1677-1678) of Groeneboom et al. (2001b). It does not change the proof (note that, of course, $\hat{\varphi}_n$ is still linear on $[\tau_{n,-}(\xi_n), \tau_{n,+}(\xi_n)]$ just as it is linear on $[\tau_{n,-}(x_0), \tau_{n,+}(x_0)]$, and this is the only way in which the differing definitions could affect the proof).

Because we consider the point $x_0 \neq m$, and because $\xi_n \to_p x_0$, with high probability as n gets large, $\tau_{n,+}^0(\xi_n)$ and $\tau_{n,-}^0(\xi_n)$ are both on the same side of m that x_0 is. Then the same triangular perturbation used in the unconstrained case is acceptable for the constrained estimator, and the same proof that gives Theorem 4.3 in Balabdaoui et al. (2009) also gives Proposition 4.2.2. □

Away from the mode, the UMLE and CMLE behave similarly and similar (or identical) proof methods work. On intervals which include the mode, the arguments must differ: acceptable Δ perturbations for the UMLE are not necessarily acceptable for the CMLE (i.e. $\hat{\varphi}_n^0 + \epsilon\Delta$ is not necessarily in $\mathcal{C}_m$, so Theorem 2.0.4 does not necessarily apply). To prove the following result we have to construct a different family of perturbations than those that are used in the argument for the unconstrained MLE (or for the CMLE away from the mode). Note that we state this result for m fixed, and later extend to consider $\tau^0_{n,+}(\xi_n) - \tau^0_{n,-}(\xi_n)$ for $\xi_n \to m$.

Theorem 4.2.3. *Under Assumptions A and D, we can conclude that*

$$\tau^0_{n,+}(m) - \tau^0_{n,-}(m) = O_p(n^{-1/5}).$$

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

$$\int_{\tau^0_{n,-}}^{\tau^0_{n,+}} \Delta(t)dt = 0, \tag{4.17}$$

$$\int_{\tau^0_{n,-}}^{0} t\Delta(t)dt = 0. \tag{4.18}$$

In (4.18), the reason that we need to consider the integral over $[\tau^0_{n,-}, 0]$ is because $\hat{\varphi}^0_n$ has 0 as an LK, as we will see in the calculations later. For plots of the perturbations we will construct, see Figure 4.1. For the case $\tau^0_{n,+} > -\tau^0_{n,-}$, we define

$$\Delta_{LK,0}(t) = \begin{cases} \frac{\tau^0_{n,+} + m_2 \cdot (-\frac{\tau^0_{n,-}}{4})}{\frac{\tau^0_{n,-}}{4} - \tau^0_{n,-}} \cdot (t - \tau^0_{n,-}) & \tau^0_{n,-} \le t \le \frac{\tau^0_{n,-}}{4} \\ \tau^0_{n,+} + m_2 \cdot (-t) & \frac{\tau^0_{n,-}}{4} \le t \le 0 \\ (\tau^0_{n,+} - t) & 0 \le t \le \tau^0_{n,+} \\ 0 & \text{otherwise} \end{cases} \tag{4.19}$$

where

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

When $\tau^0_{n,+} > -\tau^0_{n,-}$, the function defined in (4.19) has integral $(\tau^0_{n,+})(5\tau^0_{n,+} - \tau^0_{n,-})/(2(5\tau^0_{n,+} + \tau^0_{n,-}))$. For the case $\tau^0_{n,+} < -\tau^0_{n,-}$, we define

$$\Delta_{LK,0}(t) = \begin{cases} (t - \tau^0_{n,-}) & \tau^0_{n,-} \le t \le \frac{\tau^0_{n,-}}{2} \\ \frac{\tau^0_{n,-}}{2} - \tau^0_{n,-} + m_2 \cdot (\frac{\tau^0_{n,-}}{2} - t) & \frac{\tau^0_{n,-}}{2} \le t \le 0 \\ \left(\frac{-\frac{\tau^0_{n,-}}{2} + m_2 \frac{\tau^0_{n,-}}{2}}{\tau^0_{n,+}} \right) (\tau^0_{n,+} - t) & 0 \le t \le \tau^0_{n,+} \\ 0 & \text{otherwise} \end{cases}, \tag{4.20}$$

where

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

is defined so that (4.18) holds. When $\tau^0_{n,+} < -\tau^0_{n,-}$, the function defined in (4.20) has integral $-\tau^0_{n,-}(3\tau^0_{n,+} - \tau^0_{n,-})/(10\tau^0_{n,+} - 4\tau^0_{n,-})$. Then we define

$$\Delta_{LK,1}(t) = \Delta_{LK,0}(t) - M_{LK} 1_{[\tau^0_{n,-}, \tau^0_{n,+}]}(t), \tag{4.21}$$

where

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

is the appropriate shift so that (4.17) holds. M_{LK} is $o_p(1)$ since, when $\tau^0_{n,+} > -\tau^0_{n,-}$, the denominator of M_{LK} is $2(5\tau^0_{n,+} + \tau^0_{n,-}) > 2 \cdot 4\tau^0_{n,+}$. Thus in this case M_{LK} is

$$\frac{(\tau^0_{n,+})(5\tau^0_{n,+} - \tau^0_{n,-})}{2(5\tau^0_{n,+} + \tau^0_{n,-})} < \frac{5\tau^0_{n,+} - \tau^0_{n,-}}{8}, \tag{4.22}$$

which is $o_p(1)$ since $\tau^0_{n,+}$ and $\tau^0_{n,-}$ both converge in probability to 0. A similar analysis applies when $\tau^0_{n,+} \leq -\tau^0_{n,-}$. In this case, the denominator satisfies $10\tau^0_{n,+} - 4\tau^0_{n,-} > -4\tau^0_{n,+}$, so M_{LK} is

$$\frac{-\tau^0_{n,-}(3\tau^0_{n,+} - \tau^0_{n,-})}{10\tau^0_{n,+} - 4\tau^0_{n,-}} < \frac{-\tau^0_{n,-}(3\tau^0_{n,+} - \tau^0_{n,-})}{-4\tau^0_{n,-}} < \frac{(-3\tau^0_{n,-} - \tau^0_{n,-})}{4} = -\tau^0_{n,-} \tag{4.23}$$

Then $\Delta_{LK,0}$ is an acceptable perturbation for $\hat{\varphi}^0_n$, since we can have $m_2 > 1$ when 0 is an LK, and $\Delta_{LK,1}$ has the properties (4.17) and (4.18). We also define $\Delta_{RK,1}$ analogously as $\Delta_{LK,1}$, with analogous constant M_{RK}.

Now, consider the case in which 0 is an NK. When 0 was an LK, we required the integral in (4.18) to be over over $[\tau^0_{n,-}, 0]$. When 0 is an NK, i.e. $(\hat{\varphi}^0_n)'(t) = 0$ for all $t \in [\tau^0_{n,-}, \tau^0_{n,+}]$, we do not need a condition like (4.18) to hold at all, we only need Δ to satisfy (4.17). So, if $\tau^0_{n,+} > -\tau^0_{n,-}$ define

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

and otherwise define

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

Define $h_l = \max(\tau^0_{n,+}, -\tau^0_{n,-})$. We can see that $\int \Delta_{NK,0}(x)dx = h_l(\tau^0_{n,+} - \tau^0_{n,-})/2$, so set

$$\Delta_{NK,1}(x) := \Delta_{NK,0}(x) - \frac{h_l}{2} 1_{[\tau^0_{n,-}, \tau^0_{n,+}]}(x), \tag{4.24}$$

so that (4.17) is satisfied. Let

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

where

$$M_n = M_{RK}1_{[m \text{ is RK}]} + M_{LK}1_{[m \text{ is LK}]} + h_l 1_{[m \text{ is NK}]}.$$

Then set

$$\begin{aligned} \Delta_1(x) &= \Delta_{RK,1}(x)1_{[m \text{ is RK}]} + \Delta_{LK,1}(x)1_{[m \text{ is LK}]} + \Delta_{NK,1}(x)1_{[m \text{ is NK}]} \qquad (4.25) \\ &= \Delta_0(x) - M_n. \end{aligned}$$

Note that M_n is $o_p(1)$ since M_{RK}, M_{LK} and h_l all are $o_p(1)$. Then, since Δ_0 is an acceptable perturbation for the characterization (2.5), we see that

$$\begin{aligned}\int \Delta_1 d(\mathbb{F}_n - F_0) &= \int \Delta_1 d(\mathbb{F}_n - \hat{F}_n^0) + \int \Delta_1 d(\hat{F}_n^0 - F_0) \\ &= \int \Delta_0 d(\mathbb{F}_n - \hat{F}_n^0) - M_n \int_{\tau_{n,-}^0}^{\tau_{n,+}^0} d(\mathbb{F}_n - \hat{F}_n^0) + \int \Delta_1 d(\hat{F}_n^0 - F_0) \\ &\leq M_n \Big| \int_{\tau_{n,-}^0}^{\tau_{n,+}^0} d(\mathbb{F}_n - \hat{F}_n^0) \Big| + \int \Delta_1(x)(\hat{f}_n^0 - f_0)(x)dx \\ &\leq \frac{2M_n}{n} + \int \Delta_1(x)(\hat{f}_n^0 - f_0)(x)dx,\end{aligned}$$

where the last inequality is by Corollary 2.0.9. Then Lemma A.1.1 from page 221 in Appendix A.1 says

$$\left| \int \Delta_1 d(\mathbb{F}_n - F_0) \right| \leq O_p(n^{-4/5}) + \frac{K}{2} h_l^4, \tag{4.26}$$

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

for some $K > 0$ and where we picked ϵ from Lemma A.1.1 to be $K/2$. So, we have

$$0 \leq K(1 + o_p(1))h_l^4 \leq -\int \Delta_1(x)(\hat{f}_n^0 - f_0)(x)dx \leq \frac{2M_n}{n} - \int \Delta_1 d(\mathbb{F}_n - F_0)$$

which is less than or equal to

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

so that

$$0 \leq h_l^4 \leq \frac{2}{K} \frac{O_p(n^{-4/5})}{1 + o_p(1)} = O_p(n^{-4/5})$$

which means

$$0 \leq h_l = O_p(n^{-1/5}),$$

and $\tau^0_{n,+}$ and $-\tau^0_{n,-}$ are both always bounded by h_l, so we are done. □

The above proposition considered the gap between the knots about the fixed point m. We now extend the result to consider the gap between knots about variable points ξ_n which converge to m.

Corollary 4.2.4 (The Gap Problem-Mode)**.** *Under Assumptions A and D,*

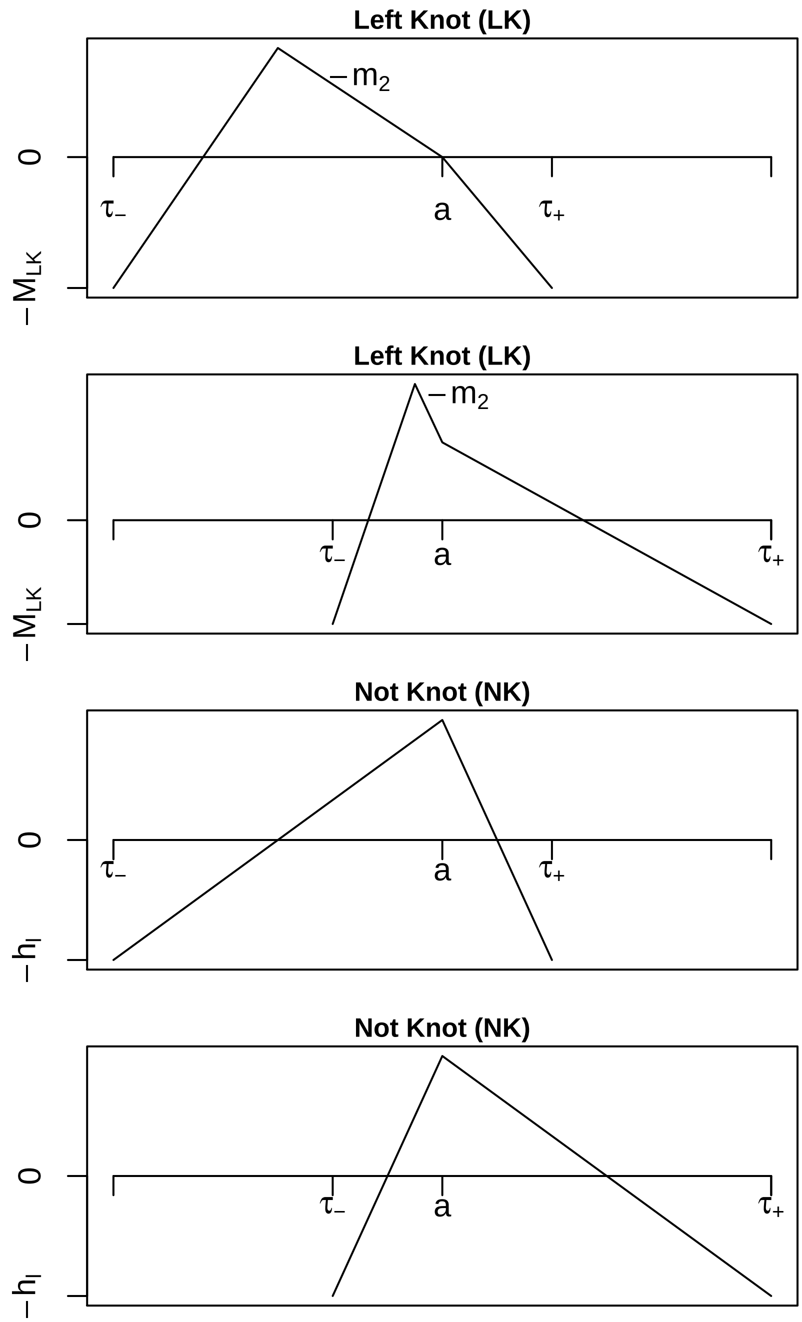
$$\tau^0_{n,+}(\xi_n) - \tau^0_{n,-}(\xi_n) = O_p(n^{-1/5})$$

where ξ_n are any real random variables satisfying $\xi_n \to_p m$.

Proof. This follows from considering the event $m \in [\tau^0_{n,-}(\xi_n), \tau^0_{n,+}(\xi_n)]$ and its complement separately. Note that by their definitions, $\tau^0_{n,+}(\xi_n)$ and $\tau^0_{n,-}(\xi_n)$, are never equal to m; either they equal $\tau^0_{n,+}(m)$ and $\tau^0_{n,-}(m)$, respectively, or they are equal to knots strictly on one side of the mode, such that the triangular perturbation used in Proposition 4.2.2 applies. So, we prove that $P\left(\frac{\tau^0_{n,+}(\xi_n)-\tau^0_{n,-}(\xi_n)}{n^{-1/5}} > M \text{ and } m \notin [\tau^0_{n,-}(\xi_n), \tau^0_{n,+}(\xi_n)]\right) \to 0$ as $M \to \infty$ using the same proof as for Proposition 4.2.2 (and Theorem 4.3 in Balabdaoui et al. (2009)), since the same Δ perturbation will be acceptable in this case. To use this perturbation requires that $\hat{\varphi}^0_n$ is not flat on the interval $[\tau^0_{n,-}(\xi_n), \tau^0_{n,+}(\xi_n)]$, which is true since we are on the event where there is a knot between ξ_n and m. See the proof-sketch of Proposition 4.2.2 and the proof of Theorem 4.3 in Balabdaoui et al. (2009). With that result in hand, we are done, since

$$\begin{aligned}
P\left(\frac{\tau^0_{n,+}(\xi_n) - \tau^0_{n,-}(\xi_n)}{n^{-1/5}} > M\right) =& P\left(\frac{\tau^0_{n,+}(\xi_n) - \tau^0_{n,-}(\xi_n)}{n^{-1/5}} > M, m \in [\tau^0_{n,-}(\xi_n), \tau^0_{n,+}(\xi_n)]\right) \\
&+ P\left(\frac{\tau^0_{n,+}(\xi_n) - \tau^0_{n,-}(\xi_n)}{n^{-1/5}} > M, m \notin [\tau^0_{n,-}(\xi_n), \tau^0_{n,+}(\xi_n)]\right)
\end{aligned}$$

Figure 4.1: Perturbations used for the Gap Problem



and both of the summands on the right go to 0 as $M \to \infty$, since

$$P\left(\frac{\tau_{n,+}^0(\xi_n) - \tau_{n,-}^0(\xi_n)}{n^{-1/5}} > M, m \in [\tau_{n,-}^0(\xi_n), \tau_{n,+}^0(\xi_n)]\right)$$
$$= P\left(\frac{\tau_{n,+}^0(m) - \tau_{n,-}^0(m)}{n^{-1/5}} > M, m \in [\tau_{n,-}^0(\xi_n), \tau_{n,+}^0(\xi_n)]\right)$$

which goes to 0 as M goes to ∞ by Theorem 4.2.3. □

We have now studied the order of magnitude of the gap between consecutive knots about the mode. We now proceed to using the rate of convergence of the gap to a rate of convergence for (the value of) $\hat{\varphi}_n^0$. Note that in the following proposition, we only assume that Assumption D holds if $x_0 = m$.

Proposition 4.2.5. *Let Assumption A hold. Let x_0 be a point in the support of f_0 (which can be equal or not equal to m) at which f_0 is twice continuously differentiable and at which $f_0''(x_0) < 0$. Let $M > 0$ be fixed. Then taking the derivative to be either the right- or left-derivative, we have*

$$\sup_{t,s \in [x_0 - Mn^{-1/5}, x_0 + Mn^{-1/5}]} \left|(\hat{\varphi}_n^0)'(t) - (\hat{\varphi}_n)'(s)\right| = O_p(n^{-1/5}), \tag{4.27}$$

and

$$\sup_{t,s \in [x_0 - Mn^{-1/5}, x_0 + Mn^{-1/5}]} \left|\hat{\varphi}_n^0(t) - \hat{\varphi}_n(s) - \hat{\varphi}_n'(s)(t-s)\right| = O_p(n^{-2/5}). \tag{4.28}$$

Proof. First we show (4.27). Let $\tau_{n,i}$ and $\tau_{n,j}^0$ be knots for $\hat{\varphi}_n$ and $\hat{\varphi}_n^0$, respectively, for $i = 1, \ldots, 4$ and $j = 1, 2, 3$, and such that they are the first knots satisfying $x_0 + Mn^{-1/5} < \tau_{n,1}$, and $\tau_{n,i} < \tau_{n,i}^0 < \tau_{n,i+1}$ for $i = 1, 2, 3$. These knots exist by consistency of the unconstrained and constrained estimators. Note that as long as n is large enough, regardless of whether

$x_0 = m$ or not, all the knots will be on the same side of m. Then by Proposition 2.0.10 on page 39 in Chapter 2, we have points $\tilde{s}_i \in (\tau_{n,i}, \tau_{n,i+1})$, for $i = 1, 2, 3$, such that $\hat{F}_n(\tilde{s}_i) = \hat{F}_n^0(\tilde{s}_i)$. Then by Proposition 2.0.11 (i.e. the Mean Value Theorem) from Chapter 2, we have points $s_i \in (\tilde{s}_i, \tilde{s}_{i+1})$ for $i = 1, 2$ such that $\hat{f}_n(s_i) = \hat{f}_n^0(s_i)$ and thus $\hat{\varphi}_n(s_i) = \hat{\varphi}_n^0(s_i)$. Similarly, there exist $s_{-2} < s_{-1} < x_0 - Mn^{-1/5}$ such that $\hat{\varphi}_n(s_i) = \hat{\varphi}_n^0(s_i)$ for $i = -1, -2$. Now, as in the proof of Lemma 4.4 on page 1682 of Groeneboom et al. (2001b), we note that for $t \in [x_0 - Mn^{-1/5}, x_0 + Mn^{-1/5}]$,

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

and similarly

$$(\hat{\varphi}_n^0)'(t+) \geq (\hat{\varphi}_n^0)'(t-) \geq \hat{\varphi}_n'(s_{-2}+).$$

Using monotonicity of $\hat{\varphi}_n'$ (again) we note for all $t \in [x_0 - Mn^{-1/5}, x_0 + Mn^{-1/5}]$ that $\hat{\varphi}_n'(s_{-2}+) \leq \hat{\varphi}_n'(t) \leq \hat{\varphi}_n'(s_2-)$. Thus, taking the derivative to be either the right- or left-derivative, we have shown

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

so that for all $s \in [x_0 - Mn^{-1/5}, x_0 + Mn^{-1/5}]$ we have

$$\hat{\varphi}_n'(s_{-2}) - \hat{\varphi}_n'(s) \leq (\hat{\varphi}_n^0)'(t) - \hat{\varphi}_n'(s) \leq \hat{\varphi}_n'(s_2) - \hat{\varphi}_n'(s).$$

The right- and left-hand sides are $O_p(n^{-1/5})$ by Lemma 4.5 on page 1319 of Balabdaoui et al. (2009), which yields

$$\sup_{|u| \leq 2M} \left| \hat{\varphi}_n'(x_0 + n^{-1/5}u) - \varphi_0'(x_0) \right| = O_p(n^{-1/5}), \tag{4.29}$$

and so in particular for any $s \in [x_0 - Mn^{-1/5}, x_0 + Mn^{-1/5}]$, both $\hat{\varphi}_n'(s_2) - \hat{\varphi}_n'(s)$ and

$|\hat{\varphi}_n'(s_{-2}) - \hat{\varphi}_n'(s)|$ are $O_p(n^{-1/5})$ by the triangle inequality. Note that we are using Proposition 4.2.2 and Corollary 4.2.4 here to imply that $s_{\pm 2} \in [x_0 - 2Mn^{-1/5}, x_0 + 2Mn^{-1/5}]$. Thus (4.27) holds.

Then (4.28) follows as in the proof of Lemma 4.4 of Groeneboom et al. (2001b), which relies only on having points of closeness or, in our case, equality of $\hat{f}_n^0$ and $\hat{f}_n$, from (4.27), and from concavity. Take $M > 0$ and the s_i as above. We know that $\hat{\varphi}_n^0(s_i) = \hat{\varphi}_n(s_i)$ and by (4.27) that for any $\epsilon > 0$ we can fix a $c > 0$ such that with probability greater than $1 - \epsilon$ for any $t \in [x_0 - Mn^{-1/5}, x_0 + Mn^{-1/5}]$ we have $\left|(\hat{\varphi}_n^0)'(s_i) - (\hat{\varphi}_n)'(t)\right| \leq cn^{-1/5}$. Hence for t and s in $[x_0 - Mn^{-1/5}, x_0 + Mn^{-1/5}]$ and n large, since $s_{-1} < x_0 - Mn^{-1/5}$, with high probability we have

$$\hat{\varphi}_n^0(t) \leq \hat{\varphi}_n^0(s_{-1}) + (\hat{\varphi}_n^0)'(s_{-1})(t - s_{-1}) \leq \hat{\varphi}_n(s_{-1}) + \left((\hat{\varphi}_n)'(s) + cn^{-1/5}\right)(t - s_{-1}). \quad (4.30)$$

We would like to Taylor expand at $\hat{\varphi}_n(s_{-1})$, but $\hat{\varphi}_n$ is not continuously differentiable. We will instead use concavity of $\hat{\varphi}_n$ to arrive at a "concave Taylor expansion" i.e a Taylor expansion that applies to concave functions. Taking $y, x \in \mathbb{R}$, we have $\hat{\varphi}_n(y) = \hat{\varphi}_n(x) + \int_x^y \hat{\varphi}_n'(u)du$ since $\hat{\varphi}_n$ is absolutely continuous (Royden, 1988, Corollary 15, page 110). Then since $\hat{\varphi}_n'$ is nonincreasing, $\hat{\varphi}_n'(x)(y - x) \geq \int_x^y \hat{\varphi}_n'(u)du \geq \hat{\varphi}_n'(y)(y - x)$, which means we can write $\hat{\varphi}_n(y) = \hat{\varphi}_n(x) + d(y - x)$ where $\hat{\varphi}_n'(x) \geq d \geq \hat{\varphi}_n'(y)$ (i.e. d is in the subdifferential of $\hat{\varphi}_n$ at some point ξ between x and y (Rockafellar, 1970)). Thus, abusing notation by writing $\hat{\varphi}_n'(\xi)$ in place of d, we have our "concave Taylor expansion" $\hat{\varphi}_n(y) = \hat{\varphi}_n(x) + \hat{\varphi}_n'(\xi)(y - x)$. Thus, for $\hat{\varphi}_n'(s_{-1}) \leq \hat{\varphi}_n'(\xi) \leq \hat{\varphi}_n'(s)$, (4.30) is equal to

$$\hat{\varphi}_n(s) + \hat{\varphi}_n'(\xi)(s_{-1} - s) + \hat{\varphi}_n'(s)(t - s_{-1}) + cn^{-1/5}(t - s_{-1}).$$

This, in turn, equals

$$\hat{\varphi}_n(s) + (\hat{\varphi}_n'(\xi) - \hat{\varphi}_n'(s))(s_{-1} - s) + \hat{\varphi}_n'(s)(s_{-1} - s) + \hat{\varphi}_n'(s)(t - s_{-1}) + cn^{-1/5}(t - s_{-1})$$

which is bounded above by

$$\hat{\varphi}_n(s) + \hat{\varphi}_n'(s)(t - s) + cn^{-1/5}(t - s_{-1}) \leq \hat{\varphi}_n(s) + \hat{\varphi}_n'(s)(t - s) + 2Mcn^{-2/5},$$

since $\hat{\varphi}_n'$ is monotone decreasing (so $(\hat{\varphi}_n'(\xi) - \hat{\varphi}_n'(s))(s_{-1} - s) < 0$). We have shown one half of the absolute value inequality (4.28). Now, we will show the reverse inequality, so that we can conclude that (4.28) holds. The argument is somewhat similar to the above one, again using the definition of concavity, the definition of s_i, our previously mentioned "concave Taylor expansion," the fact that all points t, s, and s_i are of order $n^{-1/5}$ apart, and the fact that

$$\sup_{t,s \in [x_0 - 2Mn^{-1/5}, x_0 + 2Mn^{-1/5}]} \left| \hat{\varphi}_n'(t) - \hat{\varphi}_n'(s) \right| \leq K_\epsilon n^{-1/5}$$

with probability $1 - \epsilon$ by Lemma 4.5, page 1319, of Balabdaoui et al. (2009). We have

$$\hat{\varphi}_n^0(t) \geq \hat{\varphi}_n^0(s_{-1}) + \frac{\hat{\varphi}_n^0(s_1) - \hat{\varphi}_n^0(s_{-1})}{s_1 - s_{-1}}(t - s_{-1}) = \hat{\varphi}_n(s_{-1}) + \frac{\hat{\varphi}_n(s_1) - \hat{\varphi}_n(s_{-1})}{s_1 - s_{-1}}(t - s_{-1}),$$

which equals

$$\hat{\varphi}_n(s) + \hat{\varphi}_n'(\xi_1)(s_{-1} - s) + \frac{t - s_{-1}}{s_1 - s_{-1}}\big(\hat{\varphi}_n(s) + \hat{\varphi}_n'(\xi_2)(s_1 - s) - \hat{\varphi}_n(s) - \hat{\varphi}_n'(\xi_1)(s_{-1} - s)\big),$$

which is greater than or equal to

$$\hat{\varphi}_n(s) + \hat{\varphi}_n'(s)(s_{-1} - s) + (\hat{\varphi}_n'(\xi_1) - \hat{\varphi}_n'(s))(s_{-1} - s)$$
$$+ \frac{t - s_{-1}}{s_1 - s_{-1}}\big(\hat{\varphi}_n'(s)(s_1 - s) + (\hat{\varphi}_n'(\xi_2) - \hat{\varphi}_n'(s))(s_1 - s) - \hat{\varphi}_n'(s)(s_{-1} - s)\big)$$

which is greater than or equal to

$$\hat{\varphi}_n(s) + \hat{\varphi}_n'(s)(s_{-1} - s) - 3MK_\epsilon n^{-2/5}$$
$$+ \frac{t - s_{-1}}{s_1 - s_{-1}}\big(\hat{\varphi}_n'(s)(s_1 - s) - 3MK_\epsilon n^{-2/5} - \hat{\varphi}_n'(s)(s_{-1} - s)\big),$$

which is in turn greater than or equal to

$$\hat{\varphi}_n(s) + \hat{\varphi}_n'(s)(s_{-1} - s) + \hat{\varphi}_n'(s)(t - s_{-1}) - 6MK_\epsilon n^{-2/5}$$
$$= \hat{\varphi}_n(s) + \hat{\varphi}_n'(s)(t - s) - 6MK_\epsilon n^{-2/5}.$$

Thus we have shown (4.28). □

Without much trouble we can translate the above proposition about order of magnitude of the log-density differences to the order of magnitude of the density differences.

Corollary 4.2.6. *Under the hypotheses of Proposition 4.2.5, we also have*

$$\sup_{t \in [x_0 - Mn^{-1/5}, x_0 + Mn^{-1/5}]} \left|(\hat{f}_n^0)'(t) - \hat{f}_n'(t)\right| = O_p(n^{-1/5}) \tag{4.31}$$

$$\sup_{t \in [x_0 - Mn^{-1/5}, x_0 + Mn^{-1/5}]} \left|\hat{f}_n^0(t) - \hat{f}_n(t)\right| = O_p(n^{-2/5}). \tag{4.32}$$

Proof. First we show (4.32), which follows, as usual, from a Taylor expansion and the analogous property at the log-level. For $t \in [x_0 - Mn^{-1/5}, x_0 + Mn^{-1/5}]$,

$$\left| \hat{f}_n^0(t) - \hat{f}_n(t) \right| = \hat{f}_n(t) \left| \exp\{\hat{\varphi}_n^0(t) - \hat{\varphi}_n(t)\} - 1 \right| = \hat{f}_n(t) e^{\xi_n(t)} \left| \hat{\varphi}_n^0(t) - \hat{\varphi}_n(t) \right|,$$

where $\xi_n(t)$ is between 0 and $\hat{\varphi}_n^0(t) - \hat{\varphi}_n(t)$. By taking the supremum over $t \in [x_0 - Mn^{-1/5}, x_0 + Mn^{-1/5}]$, applying uniform consistency on compacta of $\hat{f}_n$ (i.e. Theorem 4.1, page 47, of Dümbgen and Rufibach (2009)), and by applying (4.28) to both of the remaining terms (so that $e^{\xi_n(t)}$ is bounded by, say, 2, with high probability), we conclude that

$$\sup_{t \in [x_0 - Mn^{-1/5}, x_0 + Mn^{-1/5}]} \left| \hat{f}_n^0(t) - \hat{f}_n(t) \right| = O_p(n^{-2/5}),$$

so we have shown (4.32).

Then to show (4.31), we use what we just showed after writing

$$\begin{aligned} \left| (\hat{f}_n^0)'(t) - \hat{f}_n'(t) \right| &= \left| (\hat{\varphi}_n^0)'(t) \hat{f}_n^0(t) - \hat{\varphi}_n'(t) \hat{f}_n(t) \right| \\ &= \left| (\hat{\varphi}_n^0)'(t) \hat{f}_n(t) - \hat{\varphi}_n'(t) \hat{f}_n(t) + (\hat{\varphi}_n^0)'(t)(\hat{f}_n^0(t) - \hat{f}_n(t)) \right| \\ &\le \hat{f}_n(t) \left| (\hat{\varphi}_n^0)'(t) - \hat{\varphi}_n'(t) \right| + \left| (\hat{\varphi}_n^0)'(t) \right| \left| (\hat{f}_n^0(t) - \hat{f}_n(t)) \right|. \end{aligned}$$

Then, since φ_0 is continuous on $[x_0 - \delta, x_0 + \delta]$ for some $\delta > 0$, we can say for some $M_1 < \infty$ that with high probability $\sup_{t \in [x_0 - \delta, x_0 + \delta]} |(\hat{\varphi}_n^0)'(t)| \le M_1$ for n large enough by Corollary 3.1.11, and then similarly by uniform consistency of $\hat{f}_n$ (e.g. Theorem 4.1, page 47, of Dümbgen and Rufibach (2009)) we can say for some $M_2 < \infty$ that with high probability $\sup_{t \in [x_0 - \delta, x_0 + \delta]} |\hat{f}_n(t)| \le M_2$ for n large enough. Thus, we have

$$\sup_{t \in [x_0 - Mn^{-1/5}, x_0 + Mn^{-1/5}]} \left| (\hat{f}_n^0)'(t) - \hat{f}_n'(t) \right| \le M_2 \left| (\hat{\varphi}_n^0)'(t) - \hat{\varphi}_n'(t) \right| + M_1 \left| (\hat{f}_n^0(t) - \hat{f}_n(t)) \right|.$$

Then, for all n large enough, with high probability by (4.32) and by (4.27) the above display is $O_p(n^{-1/5}) + O_p(n^{-2/5}) = O_p(n^{-1/5})$, as desired. □

Proposition 4.2.5 and Corollary 4.2.6 yield rates of $O_p(n^{-2/5})$ and $O_p(n^{-1/5})$ at any point x_0 at which $f_0''(x_0) < 0$. However, when $x_0 \neq m$, we expect that the rates $O_p(n^{-2/5})$ and $O_p(n^{-1/5})$ can be replaced by $o_p(n^{-2/5})$ and $o_p(n^{-1/5})$. This improvement will follow from results in Chapter 5 that show that the estimators are asymptotically equivalent away from the mode. For now, having shown the tightness of the unconstrained and constrained estimators differences, we can translate results about the relationship between the unconstrained estimators and the true functions to results about the relationship of the constrained estimators to the true functions.

Corollary 4.2.7. *Under the hypotheses of Proposition 4.2.5 we have*

$$\sup_{|u|\le M} \left| (\hat{\varphi}_n^0)'(x_0 + n^{-1/5}u) - \varphi_0'(x_0) \right| = O_p(n^{-1/5}), \tag{4.33}$$

and

$$\sup_{|u|\le M} \left| \hat{\varphi}_n^0(x_0 + n^{-1/5}u) - \varphi_0(x_0) - n^{-1/5}u\varphi_0'(x_0) \right| = O_p(n^{-2/5}). \tag{4.34}$$

Proof. Equation (4.33) follows from $\sup_{|u|\le M} \left| (\hat{\varphi}_n^0)'(x_0 + n^{-1/5}u) - \hat{\varphi}_n'(x_0 + n^{-1/5}u) \right|$ and $\sup_{|u|\le M} \left| (\hat{\varphi}_n)'(x_0 + n^{-1/5}u) - \varphi_0'(x_0) \right|$ being $O_p(n^{-1/5})$, which are by (4.27) and Lemma 4.5 on page 1319 of Balabdaoui et al. (2009). Similarly, the (other conclusion from) the same Lemma 4.5 as well as (4.28) yield (4.34). □

Chapter 5

ASYMPTOTICS III: LIMIT DISTRIBUTION THEORY

The limit distribution of the UMLE $\hat{f}_n$ at a point x_0 such that $f_0''(x_0) < 0$ is studied in Balabdaoui et al. (2009). To discuss the limit, we first define the "invelope" process (mentioned in (1.1) in Chapter 1), studied first in Groeneboom et al. (2001b), on which the limit depends. Let $Y(t) = \int_0^t W(t)\,dt - t^4$. Then H is the invelope if H satisfies

$$\begin{aligned} &H(t) \le Y(t) \text{ for all } t, \\ &\int (H(t) - Y(t)) dH^{(3)}(t) = 0, \\ &H^{(2)}(t) \text{ is concave.} \end{aligned} \tag{5.1}$$

Then the limit distribution of the UMLE is given by the following theorem.

Theorem 5.0.8 (Theorem 2.1, page 1305, Balabdaoui et al. (2009)). *Assume that $f_0 \in \mathcal{P}$, that f_0 is twice continuously differentiable at $x_0 \in \mathbb{R}$, that $f_0(x_0) > 0$, and that $f_0''(x_0) > 0$. Then*

$$\begin{pmatrix} n^{2/5}(\hat{f}_n(x_0) - f_0(x_0)) \\ n^{1/5}(\hat{f}_n'(x_0) - f_0'(x_0)) \end{pmatrix} \overset{d}{\to} \begin{pmatrix} c_2(x_0, \varphi_0) H''(0) \\ d_2(x_0, \varphi_0) H'''(0) \end{pmatrix},$$

and

$$\begin{pmatrix} n^{2/5}(\hat{\varphi}_n(x_0) - \varphi_0(x_0)) \\ n^{1/5}(\hat{\varphi}_n'(x_0) - \varphi_0'(x_0)) \end{pmatrix} \overset{d}{\to} \begin{pmatrix} C_2(x_0, \varphi_0) H''(0) \\ D_2(x_0, \varphi_0) H'''(0) \end{pmatrix},$$

where the constants are

$$c(x_0, \varphi_0) = \left(\frac{f_0(x_0)^3 |\varphi_0^{(2)}(x_0)|}{4!} \right)^{1/5} \tag{5.2}$$

$$d(x_0, \varphi_0) = \left(\frac{f_0(x_0)^4 |\varphi_0^{(2)}(x_0)|^3}{(4!)^3} \right)^{1/5} \tag{5.3}$$

$$C(x_0, \varphi_0) = \left(\frac{|\varphi_0^{(2)}(x_0)|}{f_0(x_0)^2 (4!)^3} \right)^{1/5} \tag{5.4}$$

$$D(x_0, \varphi_0) = \left(\frac{|\varphi_0^{(2)}(x_0)|^3}{f_0(x_0)(4!)^3} \right)^{1/5}. \tag{5.5}$$

We would like to arrive at the joint limiting distribution of the UMLE and the CMLE. The limit distribution of the CMLE will be related to a "constrained invelope" process of Brownian motion, just as the limit distribution of the UMLE is related to an invelope process. Much in the same way that the characterization for the CMLE (given in Theorem 2.0.6) was a two-sided version of the characterization of the UMLE (given in Theorem 2.0.5), the constrained invelope process, which will give the limit distribution for the CMLE, will also be a two-sided version of the invelope given in (5.1).

5.1 Limit Characterizations

Theorem 5.1.13, in Section 5.1.3, will give the uniqueness of the two-sided constrained process on the whole real line. Before we get to that result, we will study analogous processes which are defined on compact intervals in Section 5.1.1 and Section 5.1.2, to gain intuition.

5.1.1 *First Limit Characterization, on* $[-c, c]$

Let W be a two-sided Brownian motion starting at 0. For $c > 0$, let

$$\begin{aligned} X(t) =& W(t) - 4t^3, \\ \tilde{X}_R(t) =& \int_t^c dX(u) = X(c) - X(t), \\ \tilde{X}_L(t) =& \int_{-c}^t dX(u) = X(t) - X(-c), \end{aligned} \tag{5.6}$$

and

$$\begin{aligned} \tilde{Y}_R(t) &= \int_t^c \tilde{X}_R(u)du, \\ \tilde{Y}_L(t) &= \int_{-c}^t \tilde{X}_L(u)du. \end{aligned} \tag{5.7}$$

These processes will play the role in the limit that our data observations play in finite samples. We define the "least-squares" objective functional to be minimized by

$$\phi_c(g) = \frac{1}{2}\int_{-c}^c g^2(t)\,dt - \int_{-c}^c g(t)\,dX(t). \tag{5.8}$$

We want to minimize ϕ_c over the class of convex functions with mode at 0, i.e. over

$$\mathcal{G}^0_{c,k} := \{g\colon [-c,c] \to \mathbb{R} | -g \in \mathcal{C}_0, g(-c) = k, g(c) = k\}.$$

We constrain the value of g at the endpoints so as to ensure that a minimum exists. If we consider the subset of $g \in \mathcal{G}^0_{c,k}$ such that $g \geq -M$ for some $0 < M < \infty$, then this subset is compact in $L_p(\lambda)$, by Theorem 3.1 on page 146 of Dryanov (2009). The argument used to prove Lemma 4.1 of Chen and Wellner (2013) shows that indeed there is a random variable $0 < M < \infty$ such that any minimizer of ϕ_c is larger than $-M$ almost surely, and thus we can almost surely restrict attention to a compact subset. Our first result gives an "outside-in" characterization of the minimizer of ϕ_c over $\mathcal{G}^0_{c,k}$. It is "outside-in" in that all of the integrated processes that appear in the theorem begin at the outside of the interval $[-c, c]$. This is the manner in which the characterization is given in Theorem 5.0.8. We will later find that an "inside-out" characterization is more natural for the constrained problem, and will study such a characterization in Section 5.1.2. The results of this latter characterization on $[-c, c]$ will motivate our study of the processes on $(-\infty, \infty)$ in Section 5.1.3.

Theorem 5.1.1 (Outside-in characterization). *Let $\hat{f}^0_c \colon [-c, c] \to \mathbb{R}$ be measurable, and set $\tau^0_{c,1} = \inf\{c \geq t \geq 0 : (\hat{f}^0_c)'(t+) > 0\}$, $\tau^0_{c,-1} = \sup\{-c \leq t \leq 0 : (\hat{f}^0_c)'(t-) < 0\}$ and let*

$$\tilde{F}_R(t) = \tilde{c}_R + \int_t^c \hat{f}^0_c(u)du, \quad \tilde{F}_L(t) = \tilde{c}_L + \int_{-c}^t \hat{f}^0_c(u)du,$$

$$\tilde{H}_R(t) = \int_t^c \tilde{F}_R(u)du, \quad \tilde{H}_L(t) = \int_{-c}^t \tilde{F}_L(u)du,$$

where $\tilde{c}_L$ and $\tilde{c}_R$ are chosen so that $\tilde{H}_R(\tau^0_{c,1}) = \tilde{Y}_R(\tau^0_{c,1})$ and $\tilde{H}_L(\tau^0_{c,-1}) = \tilde{Y}_L(\tau^0_{c,-1})$. Let ϕ_c be defined by (5.8) and assume its minimizer is a piecewise linear function whose points of jump, excluding $\pm c$, are isolated. Then $\hat{f}^0_c \in \mathcal{G}^0_{c,k}$ is the unique minimizer of ϕ_c over $\mathcal{G}^0_{c,k}$ if and only if

(i). We have

$$\tilde{c}_L + \tilde{c}_R + \int_{-c}^c \hat{f}^0_c(t)dt = \int_{-c}^c dX(t),$$

which means that for all $t \in [-c, c]$,

$$(\tilde{F}_L - \tilde{X}_L)(t) = -(\tilde{F}_R - \tilde{X}_R)(t). \tag{5.9}$$

(ii).

$$\tilde{H}_L(t) - \tilde{Y}_L(t) \geq 0 \qquad \textit{for} \qquad -c \leq t \leq 0 \tag{5.10}$$

$$\tilde{H}_R(t) - \tilde{Y}_R(t) \geq 0 \qquad \textit{for} \qquad c \geq t \geq 0. \tag{5.11}$$

(iii).

$$\int_{[-c,\tau^0_{c,-1}]} (\tilde{H}_L(u) - \tilde{Y}_L(u))\, d(\hat{f}^0_c)'(u) = 0 = \int_{[\tau^0_{c,1},c]} (\tilde{H}_R(u) - \tilde{Y}_R(u))\, d(\hat{f}^0_c)'(u). \tag{5.12}$$

Note that $(\hat{f}^0_c)'$ *is* 0 *on the interval* $(\tau^0_{c,-1}, \tau^0_{c,1})$ *so the bounds of one of the above integrals may be extendable if one of* $\tau^0_{c,-1}$ *or* $\tau^0_{c,1}$ *are not* 0*.*

We give an outline of the proof of the necessity of the theorem's conditions before jumping in. The proof of sufficiency is mostly just by unwinding the proof of necessity, given the above definitions. The argument for the proof of necessity is similar to the argument for the finite sample theorem, Theorem 2.0.6; we argue via perturbations Δ, keeping in mind that $\Delta(\pm c)$ needs to be 0. One way to see the importance of this condition is to think of minimizing an objective function like ϕ_c but with $c = \infty$; then any perturbation which does not vanish at infinity will lead to an infinite value of the objective function. Our proof outline is as follows.

(i). To show (5.10) and (5.11), we would like to use elbow functions (see the top row on the left of Figure 5.1 for elbow functions) as our perturbations. However, these functions are not acceptable perturbations because they are not 0 at $\pm c$. Thus, we add kinks at knot points to define acceptable elbow perturbations which are 0 at $\pm c$ (see the bottom row on the left of Figure 5.1 for elbow perturbations). We use (5.12) to show the kinks we added at the knots can be ignored.

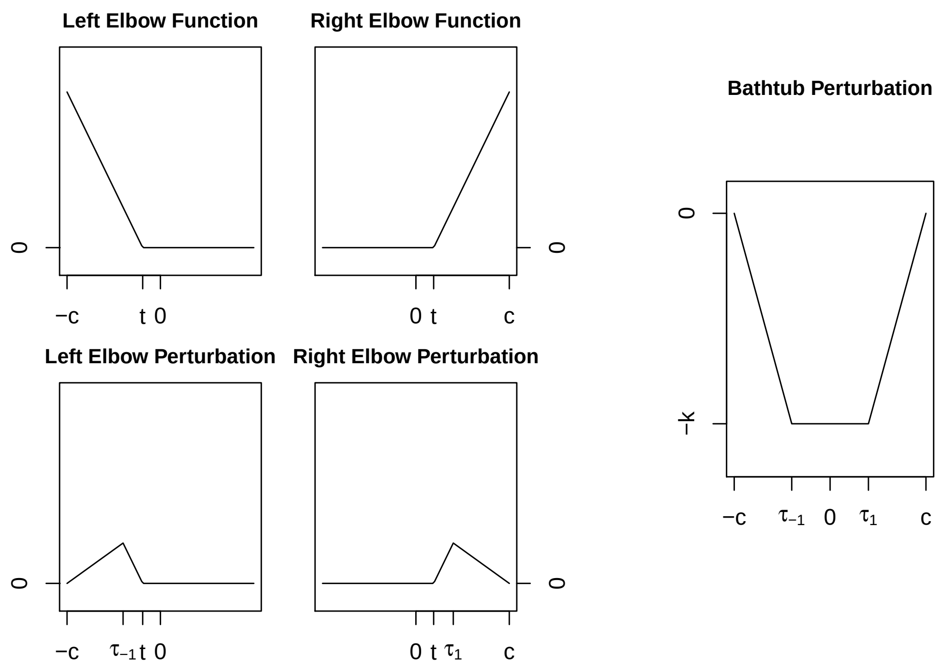


Figure 5.1: Perturbations for Limit Characterization

(ii). To show (5.9) we would like to use a constant perturbation, but that will not be 0 at $\pm c$. Instead we use a bathtub perturbation (on the right of Figure 5.1) which has kinks at knots so as to be 0 at $\pm c$. We again use (5.12) to show the kinks we added at the knots can be ignored.

(iii). However to show (5.12) we need (5.9). To resolve this circularity, we note that we have to make the choice of two degrees of freedom in defining the first and second integrals of $\hat{f}_c^0$ (the choices of where to start the two integrals). We can thus fix those starting locations (which is the same as modifying each level of integral by a constant) such that we know $(\tilde{H}_L - \tilde{Y}_L)(\tau^0_{c,-1}) = 0 = (\tilde{H}_R - \tilde{Y}_R)(\tau^0_{c,1})$; this is why we chose $\tilde{c}_L$ and $\tilde{c}_R$ above. This way we can prove (5.9) and then still expect that (5.12) is true with our choices of H's and Y's.

Now we actually prove the theorem.

Proof of Theorem 5.1.1. First, notice that

$$\lim_{\epsilon\to 0}\frac{\phi_c(\hat{f}_c^0+\epsilon\Delta)-\phi_c(\hat{f}_c^0)}{\epsilon} = \int_{-c}^{c}\Delta(u)d(\tilde{F}_L-\tilde{X}_L)(u)$$
$$= \int_{-c}^{0}\Delta(u)d(\tilde{F}_L-\tilde{X}_L)(u) - \int_0^c \Delta(u)d(\tilde{F}_R-\tilde{X}_R)(u), \quad (5.13)$$

because $d\tilde{F}_R = -d\tilde{F}_L$ and $d\tilde{X}_R = -d\tilde{X}_L$. Now we will first assume $\hat{f}_c^0$ minimizes ϕ_c over $\mathcal{G}_{c,k}^0$ and show that this implies the conditions (i)–(iii). Since $\hat{f}_c^0$ minimizes ϕ_c over $\mathcal{G}_{c,k}^0$, for any Δ such that there exists $\epsilon_0 > 0$ such that $\hat{f}_c^0 + \epsilon\Delta \in \mathcal{G}_{c,k}^0$, we automatically have the inequality

$$0 \le \lim_{\epsilon\to 0}\frac{\phi_c(\hat{f}_c^0+\epsilon\Delta)-\phi_c(\hat{f}_c^0)}{\epsilon} = \int_{-c}^{0}\Delta(u)d(\tilde{F}_L-\tilde{X}_L)(u) - \int_0^c \Delta(u)d(\tilde{F}_R-\tilde{X}_R)(u).$$

Next, define the family of elbow perturbation functions (see Figure 5.1)

$$\Delta_t(u) = \begin{cases} (u-t)_+ - \frac{c-t}{c-\tau_{c,1}^0}\left(u-\tau_{c,1}^0\right)_+ & \text{if } t\ge 0 \\ (t-u)_+ - \frac{t-c}{\tau_{c,-1}^0+c}\left(\tau_{c,-1}^0-u\right)_+ & \text{if } t\le 0 \end{cases},$$

and let $g_{t,\epsilon}(u) = \hat{f}_c^0(u) + \epsilon\Delta_t(u) \in \mathcal{G}_{c,k}^0$ be the perturbed $\hat{f}_c^0$. For $t \ge 0$, we have

$$0 \le \lim_{\epsilon\to 0}\frac{\phi_c(g_{t,\epsilon})-\phi_c(\hat{f}_c^0)}{\epsilon} = -\int_0^c\left((u-t)_+ - \frac{c-t}{c-\tau_{c,1}^0}\left(u-\tau_{c,1}^0\right)_+\right)d(\tilde{F}_R-\tilde{X}_R)(u).$$

Note that $\Delta_t(c) = 0 = \Delta_t(t)$. Integration by parts thus gives

$$0 \le -\int_t^c \left((u-t)_+ - \frac{c-t}{c-\tau_{c,1}^0} \left(u - \tau_{c,1}^0\right)_+ \right) d(\tilde{F}_R - \tilde{X}_R)(u)$$
$$= -0 + 0 + \int_t^c \left(\tilde{F}_R(u) - \tilde{X}_R(u)\right) du - \frac{c-t}{c-\tau_{c,1}^0} \int_{\tau_{c,1}^0}^c \left(\tilde{F}_R(u) - \tilde{X}_R(u)\right) du,$$

which equals

$$-\left(\tilde{H}_R(c) - \tilde{Y}_R(c) - (\tilde{H}_R(t) - \tilde{Y}_R(t))\right) + \frac{c-t}{c-\tau_{c,1}^0} \left(\tilde{H}_R(c) - \tilde{Y}_R(c) - (\tilde{H}_R(\tau_{c,1}^0) - \tilde{Y}_R(\tau_{c,1}^0))\right)$$
$$= \tilde{H}_R(t) - \tilde{Y}_R(t),$$

since $\tilde{H}_R'(t) = -\tilde{F}_R(t)$ and $\tilde{Y}_R'(t) = -\tilde{X}_R(t)$ for all $t \in \mathbb{R}$. So we have shown (5.11). Similarly for $t \le 0$,

$$0 \le \int_{-c}^t \left((t-u)_+ - \frac{t-c}{\tau_{c,-1}^0 - c} \left(\tau_{c,-1}^0 - u\right)_+ \right) d(\tilde{F}_L - \tilde{X}_L)(u)$$
$$= 0 - 0 - \int_{-c}^t (\tilde{F}_L - \tilde{X}_L)(u) d(-u) + \frac{t-c}{\tau_{c,-1}^0 + c} \int_{-c}^{\tau_{c,-1}^0} \left(\tilde{F}_L - \tilde{X}_L\right)(u) d(-u)$$
$$= \tilde{H}_L(t) - \tilde{Y}_L(t) - \left(\tilde{H}_L(-c) - \tilde{Y}_L(-c)\right)$$
$$- \frac{t-c}{\tau_{c,-1}^0 + c} \left(\tilde{H}_L(\tau_{c,-1}^0) - \tilde{Y}_L(\tau_{c,-1}^0) - (\tilde{H}_L(-c) - \tilde{Y}_L(-c))\right)$$
$$= \tilde{H}_L(t) - \tilde{Y}_L(t),$$

since $\tilde{H}_L'(t) = \tilde{F}_L(t)$ and $\tilde{Y}_L'(t) = \tilde{X}_L(t)$. So we have shown (5.10). Now setting $\pm\Delta(u)$ to be the bathtub function $\Delta_B(u) = -k + c_1 \left(\tau_{c,-1}^0 - u\right)_+ + c_2 \left(u - \tau_{c,1}^0\right)_+$, where $c_1 = k/(\tau_{c,-1}^0 + c)$

and $c_2 = k/(c - \tau_{c,1}^0)$ so that $\Delta_B(\pm c) = 0$ (see Figure 5.1), we see

$$\begin{aligned} 0 &= \int_{-c}^{\tau_{c,1}^0} \Delta_B(u) d(\tilde{F}_L - \tilde{X}_L)(u) - \int_{\tau_{c,1}^0}^{c} \Delta_B(u) d(\tilde{F}_R - \tilde{X}_R)(u) \\ &= \Delta_B(\tau_{c,1}^0)(\tilde{F}_L - \tilde{X}_L)(\tau_{c,1}^0) - 0 - 0 + \Delta_B(\tau_{c,1}^0)(\tilde{F}_R - \tilde{X}_R)(\tau_{c,1}^0) \\ &\quad - \int_{-c}^{\tau_{c,1}^0} (\tilde{F}_L - \tilde{X}_L)(u) d\Delta_B(u) + \int_{\tau_{c,1}^0}^{c} (\tilde{F}_R - \tilde{X}_R)(u) d\Delta_B(u), \end{aligned}$$

and since Δ_B' is 0 on $(\tau_{c,-1}^0, \tau_{c,1}^0)$, the above display equals

$$\begin{aligned} &(-k)(\tilde{F}_L - \tilde{X}_L)(\tau_{c,1}^0) + (-k)(\tilde{F}_R - \tilde{X}_R)(\tau_{c,1}^0) \\ &\quad - \int_{-c}^{\tau_{c,-1}^0} (\tilde{F}_L - \tilde{X}_L)(u) d\Delta_B(u) + \int_{\tau_{c,1}^0}^{c} (\tilde{F}_R - \tilde{X}_R)(u) d\Delta_B(u) \\ &\quad = (-k)(\tilde{F}_L - \tilde{X}_L)(\tau_{c,1}^0) + (-k)(\tilde{F}_R - \tilde{X}_R)(\tau_{c,1}^0), \end{aligned} \tag{5.14}$$

where the final equality is by the definitions in the theorem statement of $\tilde{c}_L$ and $\tilde{c}_R$, i.e. because we defined $\tilde{H}_L(\tau_{c,-1}^0) = \tilde{Y}_L(\tau_{c,-1}^0)$ and $\tilde{H}_R(\tau_{c,1}^0) = \tilde{Y}_R(\tau_{c,1}^0)$. Now, by definition, $(\tilde{F}_L - \tilde{X}_L)(u) = -(\tilde{F}_R - \tilde{X}_R)(u) + C$ for some $C \in \mathbb{R}$, but since $(\tilde{F}_L - \tilde{X}_L)(\tau_{c,1}^0) = -(\tilde{F}_R - \tilde{X}_R)(\tau_{c,1}^0)$, we conclude $C = 0$ and we have shown (5.9). Next, we will show (5.12). First, define

$$f_L(u) = \begin{cases} \hat{f}_c^0(u) - \hat{f}_c^0(0) & \text{if } u \le 0 \\ 0 & \text{if } u > 0 \end{cases},$$

and

$$f_R(u) = \begin{cases} \hat{f}_c^0(u) - \hat{f}_c^0(0) & \text{if } u \ge 0 \\ 0 & \text{if } u < 0 \end{cases},$$

noting that $\hat{f}_c^0(0) = \hat{f}_c^0(\tau_{\pm 1})$. Both f_L and f_R are convex with mode at 0 so certainly both $\hat{f}_c^0 + f_L$ and $\hat{f}_c^0 + f_R$ are in $\mathcal{G}_{c,k}^0$. In fact $\hat{f}_c^0(u) - f_L(u) = f_R(u) + \hat{f}_c^0(0) \in \mathcal{G}_{c,k}^0$ (so

$\hat{f}_c^0(u) - f_R(u) = f_L(u) + \hat{f}_c^0(0) \in \mathcal{G}_{c,k}^0$). Then we define new perturbation functions

$$\begin{aligned}\Delta_L(u) &= f_L(u) - \frac{\hat{f}_c^0(-c) - \hat{f}_c^0(\tau_{c,-1}^0)}{\tau_{c,-1}^0 + c}(\tau_{c,-1}^0 - u)_+ \\ \Delta_R(u) &= f_R(u) - \frac{\hat{f}_c^0(c) - \hat{f}_c^0(\tau_{c,1}^0)}{c - \tau_{c,1}^0}(u - \tau_{c,1}^0)_+.\end{aligned}$$

Note that $\Delta(\pm c) = 0 = \Delta(0)$, for Δ equal to Δ_L or Δ_R, and, for small ϵ, $\hat{f}_c^0 \pm \epsilon\Delta \in \mathcal{G}_{c,k}$, despite the fact that $\Delta \notin \mathcal{G}_{c,k}$. This is because, using f_R for example, $\hat{f}_c^0 + \epsilon f_R \in \mathcal{G}_{c,k}$ and $\hat{f}_c^0 + \epsilon f_R$ has an isolated knot at $\tau_{c,1}^0$ by definition. Thus, for ϵ small enough, we can add the concave function $-\epsilon\frac{\hat{f}_c^0(c)-\hat{f}_c^0(\tau_{c,1}^0)}{c-\tau_{c,1}^0}(u-\tau_{c,1}^0)_+$ and the overall function remains convex. We have

$$\begin{aligned}0 &= \int_{-c}^{c} \Delta_L(u)\, d(\tilde{F}_L - \tilde{X}_L)(u) \\ &= 0 - 0 - \int_{-c}^{\tau_{c,-1}^0} f_L'(u)\left(\tilde{F}_L - \tilde{X}_L\right)(u)du + \frac{\hat{f}_c^0(c) - \hat{f}_c^0(\tau_{c,1}^0)}{c - \tau_{c,1}^0}\int_{-c}^{\tau_{c,-1}^0}\left(\tilde{F}_L - \tilde{X}_L\right)(u)du\end{aligned}$$

which equals

$$-\int_{-c}^{\tau_{c,-1}^0}(\hat{f}_c^0)'(u)\left(\tilde{F}_L - \tilde{X}_L\right)(u)du + 0 = -0 + 0 + \int_{[-c,\tau_{c,-1}^0]}\left(\tilde{H}_L - \tilde{Y}_L\right)(u)d(\hat{f}_c^0)'(u). \tag{5.15}$$

Normally after the integration by parts for the last equality the bounds of integration would be $(-c, \tau_{c,-1}^0]$, but because $\tilde{H}_L(-c) - \tilde{Y}_L(-c) = 0$, we can extend the integral to include the point $-c$. A similar calculation to (5.15) yields

$$0 = \int_{[\tau_{c,1}^0,c]}\left(\tilde{H}_R - \tilde{Y}_R\right)(u)d(\hat{f}_c^0)'(u),$$

so we have showed (5.12). We have now proved the necessity of the conditions.

Now we prove sufficiency. Assume that $g \in \mathcal{G}_{c,k}^0$. Note that $g^2 - (\hat{f}_c^0)^2 = (g - \hat{f}_c^0)^2 +$

$2\hat{f}_c^0(g - \hat{f}_c^0) \geq 2\hat{f}_c^0(g - \hat{f}_c^0)$. Thus

$$\begin{aligned}\phi_c(g) - \phi_c(\hat{f}_c^0) &= \frac{1}{2}\int_{-c}^{c}\left(g^2(t) - (\hat{f}_c^0)^2(t)\right)dt - \int_{-c}^{c}\left(g(t) - \hat{f}_c^0(t)\right)dX(t) \\ &\geq \int_{-c}^{c}\left(g(t) - \hat{f}_c^0(t)\right)\hat{f}_c^0(t)\,dt - \int_{-c}^{c}\left(g(t) - \hat{f}_c^0(t)\right)dX(t),\end{aligned}$$

with strict inequality if g is not Lebesgue-almost-surely identical to $\hat{f}_c^0$. We rewrite the last expression as

$$\int_{-c}^{c}\left(g(t) - \hat{f}_c^0(t)\right)d\left(\tilde{F}_L - \tilde{X}_L\right)(t) = \int_{-c}^{0}(g - \hat{f}_c^0)d(\tilde{F}_L - \tilde{X}_L) - \int_{0}^{c}(g - \hat{f}_c^0)d(\tilde{F}_R - \tilde{X}_R).$$

Integration by parts lets us write the above display as

$$\begin{aligned}&(g - \hat{f}_c^0)(0)(\tilde{F}_L - \tilde{X}_L)(0) - 0 - \int_{-c}^{0}\left(\tilde{F}_L - \tilde{X}_L\right)(t)\left(g'(t) - (\hat{f}_c^0)'(t)\right)dt \\ &\quad - 0 + (g - \hat{f}_c^0)(0)(\tilde{F}_R - \tilde{X}_R)(0) + \int_{0}^{c}(\tilde{F}_R - \tilde{X}_R)(t)(g'(t) - (\hat{f}_c^0)'(t))dt \\ &\quad = -\int_{-c}^{0}\left(\tilde{F}_L - \tilde{X}_L\right)(t)\left(g'(t) - (\hat{f}_c^0)'(t)\right)dt + \int_{0}^{c}(\tilde{F}_R - \tilde{X}_R)(t)(g'(t) - (\hat{f}_c^0)'(t))dt,\end{aligned}$$

by using (5.9). Now, we deduce from (5.12), in similar fashion to display (5.15), that

$$\begin{aligned}&\int_{-c}^{0}\left(\tilde{F}_L - \tilde{X}_L\right)(t)(\hat{f}_c^0)'(t)dt - \int_{0}^{c}(\tilde{F}_R - \tilde{X}_R)(t)(\hat{f}_c^0)'(t)dt \\ &\quad = \left((\tilde{H}_L - \tilde{Y}_L)(0)(\hat{f}_c^0)'(0-) - (\tilde{H}_L - \tilde{Y}_L)(-c+)(\hat{f}_c^0)'(-c+)\right) \\ &\quad\quad - \left((\tilde{H}_R - \tilde{Y}_R)(c-)(\hat{f}_c^0)'(c-) - (\tilde{H}_R - \tilde{Y}_R)(0+)(\hat{f}_c^0)'(0+)\right) \\ &\quad\quad - \int_{(-c,0)}\left(\tilde{H}_L - \tilde{Y}_L\right)(t)d(\hat{f}_c^0)'(t) + \int_{(0,c)}(\tilde{H}_R - \tilde{Y}_R)(t)d(\hat{f}_c^0)'(t) \\ &\quad = 0,\end{aligned}$$

where the last equality follows from (5.12), because $(\hat{f}_c^0)'$ is 0 on $(\tau_{c,-1}^0, \tau_{c,1}^0)$, and because, considering the left side for instance, if $\tau^0_{c,-1} \neq 0$ then $(\hat{f}^0_c)'(0-) = 0$, and if $\tau^0_{c,-1} = 0$ then $\tilde{H}_L(0) - \tilde{Y}_L(0) = 0$, again by (5.12). Thus, we have shown

$$\phi_c(g) - \phi_c(\hat{f}^0_c) \geq -\int_{-c}^{0} \left(\tilde{F}_L - \tilde{X}_L\right)(t) g'(t)dt + \int_0^c (\tilde{F}_R - \tilde{X}_R)(t)g'(t)dt.$$

Now we start by considering g which is of the form $g(t) = a + \sum_{i \leq 0} \alpha_i(\eta_i - t)_+ + \sum_{i\geq 0} \beta_i(t - \eta_i)_+$ with $\eta_i \in \mathbb{R}$ an increasing sequence, $\eta_0 = 0$, $\alpha_i, \beta_i \geq 0$ and $a \in \mathbb{R}$. When g has this form, we see that the above display equals

$$\sum_{i\leq 0}(-\alpha_i) \int_{-c}^{\eta_i} -\left(\tilde{F}_L(t) - \tilde{X}_L(t)\right) dt + \sum_{i \geq 0} \beta_i \int_{\eta_i}^{c} \left(\tilde{F}_R(t) - \tilde{X}_R(t)\right) dt \geq 0,$$

where the inequality follows from (5.10) and (5.11) and because the α_i and β_i are positive. Thus, for g of the above type, $\phi_c(g) - \phi_c(\hat{f}^0_c) \geq 0$. Then, because the g of the above type approximate any function in $\mathcal{G}^0_{c,k}$ arbitrarily well in the uniform norm and because ϕ_c is certainly continuous with respect to that norm (by the Bounded Convergence Theorem), it follows that $\hat{f}^0_c$ minimizes ϕ_c over $\mathcal{G}^0_{c,k}$. □

From now on, we will let $\hat{f}^0_c$ be the above-described minimizer of ϕ_c over $\mathcal{G}^0_{c,k}$. Also, for $i = 1, \ldots, \infty$, we will define $\tau^0_{c,\pm i}$ in a fashion analogous to the definition of $\tau^0_{c,\pm 1}$. For $i > 1$, let $\tau^0_{c,i} := \inf\left\{\tau^0_{c,i-1} \leq t \leq c : (\hat{f}^0_c)'(t+) > (\hat{f}^0_c)'(\tau^0_{c,i-1}+)\right\}$ and for $i < -1$, let $\tau^0_{c,i} := \sup\left\{\tau^0_{c,i+1} \geq t \geq -c : (\hat{f}^0_c)'(\tau^0_{c,i+1}-) > (\hat{f}^0_c)'(t-)\right\}$. If, in the definition of τ^0_i for $i > 0$, the set over which the infimum is taken is equal to c then we assign τ^0_i to be c, and if the set is empty then we assign τ^0_i to be ∞, and similarly when $i < 0$, but with $-c$ and $-\infty$, respectively. This definition relies on the separation of the knots of $\hat{f}^0_c$ away from c. Now we will note a corollary for the limit case similar to Corollary 2.0.9 for the finite sample case. Note that neither $\tau = 0$ nor $\tau = \pm c$ is allowed in this corollary.

Corollary 5.1.2. *Let τ be a point of jump of $\hat{f}^0_c$. Then if $-c < \tau < 0$, then $\tilde{F}_L(\tau) = \tilde{X}_L(\tau)$*

If $c > \tau > 0$ then $\tilde{F}_R(\tau) = \tilde{X}_R(\tau)$.

The corollary follows immediately from the following fact.

Fact 5.1.3. *If H and Y are differentiable and $H \geq Y$ on a set $[a, b]$ then at any point $\tau \in (a, b)$ at which $H(\tau) = Y(\tau)$, we can conclude that $H'(\tau) = Y'(\tau)$.*

Proof. Since τ is an interior point to (a, b), $H(t) - Y(t) \geq 0$ for t in a neighborhood of τ. Since $H(\tau) - Y(\tau) = 0$, τ is a local minimum. Thus $H'(\tau) - Y'(\tau) = 0$. ☐

Notice that via Corollary 5.1.2 we conclude that $\int_{\tau^0_{c,i}}^{\tau^0_{c,i+1}} (\hat{f}^0_c(v)dv - dX(v)) = 0$ for knots $\tau^0_{c,i}, \tau^0_{c,i+1}$ with either $-c < \tau^0_{c,i}, \tau^0_{c,i+1} < 0$ or $0 < \tau^0_{c,i}, \tau^0_{c,i+1} < c$, and

$$\tilde{c}_L + \int_{-c}^{\tau_{m_L}} (\hat{f}^0_c(v)dv - dX(v)) = 0 = \tilde{c}_R + \int_{\tau_{m_R}}^{c} (\hat{f}^0_c(v)dv - dX(v)),$$

where τ_{m_L} is the infimum of the knots greater than $-c$ and τ_{m_R} is the supremum of the knots less than c. Thus, defining τ_L and τ_R by (5.16) below, the only aspect of condition (5.9) that is not implied by the other pieces of the characterization (since τ_L and τ_R are on opposite sides of 0) is that

$$\int_{\tau_L}^{\tau_R} (\hat{f}^0_c(v)dv - dX(v)) = 0.$$

This inspires our later characterization in which we replace (5.9) with (5.29). Furthermore, the fact that at a knot both "$F = X$" and "$H = Y$" inspires our re-characterization in which the processes have integrals starting at knots, with no extra affine translation.

Remark 5.1.4. Notice that $\tau^0_{c,\pm 1}$ may be equal to 0. (Almost surely, no more than one of them will be 0, though.) Regardless, (5.12) implies that, as mentioned in the proof outline on page 112, $(\tilde{H}_L - \tilde{Y}_L)(\tau^0_{c,-1}) = 0 = (\tilde{H}_R - \tilde{Y}_R)(\tau^0_{c,1})$. However, we cannot apply Corollary 5.1.2 to whichever of $\tau^0_{c,\pm 1}$ is 0. Thus, we will need to define τ_R and τ_L to be always different than 0 so that we can apply Corollary 5.1.2.

5.1.2 Second Limit Characterization, on $[-c, c]$

The characterization in Theorem 5.1.1 is "outside-in" in the sense that the integrals all begin at the "outside" (at $\pm c$) and move "in" towards 0. Here we rewrite the "outside-in" characterization in terms of an "inside-out" characterization. The integrals will not start at 0 but at τ_R and τ_L, which we define by

$$\tau_L = \sup\{\tau^0_{c,i} : \tau^0_{c,i} < 0\} \text{ and } \tau_R = \inf\{\tau^0_{c,i} : \tau^0_{c,i} > 0\}, \tag{5.16}$$

where $\tau^0_{c,i}$ are the knots of the piecewise linear function $\hat{f}^0_c$, where the endpoints $\pm c$ count as knots. Note that τ_L and τ_R are not necessarily equal to $\tau^0_{c,-1}$ and $\tau^0_{c,1}$, respectively, which we used in Section 5.1.1, since we guarantee $\tau_L < 0 < \tau_R$ (but $\tau^0_{c,\pm 1}$ may be equal to 0). There almost surely exists a large c such that the minimizer of ϕ_c has at least $m \in \mathbb{N}$ knots on either side of 0, so we will take c large enough that $\hat{f}^0_c$ has at least one knot strictly less than 0 and at least one knot strictly greater than 0. The following fact and its proof contain the bookkeeping we need.

Fact 5.1.5. *Let $d < e$ be real numbers. Let $L(t) = \int_d^t \int_d^u dm(v)du + k^o_1 + k^o_2(t-d)$ and let $R(t) = \int_t^e \int_u^e dm(v)du + k^i_1 + k^i_2(e-t)$ for some arbitrary measure m and constants $k^o_1, k^o_2, k^i_1, k^i_2 \in \mathbb{R}$. We can take $\int_a^b$ to mean the integral over $[a,b]$, $(a,b]$, $[a,b)$, or (a,b), as long as the notation is consistent throughout the statement (and proof) of the result. We can then conclude that there exists an affine function $A(t)$ such that $L(t) + A(t) = R(t)$ for all t. In particular, $L - R$ is affine, and if $L(d) = R(d)$ and $L(e) = R(e)$ then $L = R$.*

Proof. For a heuristic proof, we assume that L and R are differentiable: $L' - R'$ is a constant, call it s, so that $L(t)$ and $R(t) + st$ have the same derivative and so also differ by a constant, i.e. $L - R$ is affine, and we are done.

However, we will not assume we can differentiate L and R. Instead, we will be completely general and do everything by hand.

$$
\begin{aligned}
R(t) &= \int_d^e \int_u^e dm(v)du - \int_d^t \int_u^e dm(v)du + k_1^i + k_2^i(e-t) \\
&= \int_d^e \int_u^e dm(v)du + k_1^i - \int_d^t \left(\int_d^e dm(v) - \int_d^u dm(v) \right) du \\
&\quad + k_2^i(e-d+d-t) \\
&= \int_d^e \int_u^e dm(v)du + k_1^i + k_2^i(e-d) - (t-d)\int_d^e dm(v) + \int_d^t \int_d^u dm(v)du \\
&\quad + k_2^i(d-t),
\end{aligned}
$$

and, continuing, we see that the above equals

$$
\begin{aligned}
&\int_d^e \int_u^e dm(v)du + k_1^i + k_2^i(e-d) + (k_2^i + \int_d^e dm(v))(d-t) \\
&+ \int_d^t \int_d^u dm(v)du + k_1^o + k_2^o(t-d) - (k_1^o + k_2^o(t-d)) \\
= &\int_d^e \int_u^e dm(v)du + k_1^i - k_1^o + k_2^i(e-d) + \left(k_2^i - k_2^o + \int_d^e dm(v) \right)(d-t) \\
&+ L(t).
\end{aligned}
$$

Grouping the constants together, we can say

$$R(t) = k_1 + k_2(d-t) + L(t),$$

and we are done since we have equality at two points. That is, since $R(d) = L(d)$, we conclude $k_1 = 0$, and then since $R(e) = L(e)$, and $e \neq d$, we conclude $k_2 = 0$, so $R(t) = L(t)$ for all t.

The final statement of the proof, that there exists an affine function $A(t)$ such that $L + A = R$, follows simply from defining the affine function A by its values at two points, d and e: set $A(d) = R(d) - L(d)$ and $A(e) = R(e) - L(e)$. Then by the result above, since

$L + A$ is still of the form needed (i.e. double left-integral plus affine function), we conclude $L + A = R$. $\square$

The above calculation now applies to our H's and our Y's. We start on the left hand side, setting both $\tilde{H}_L$ and $\tilde{Y}_L$ as the L function in Fact 5.1.5, which then says that there exist affine functions, A_1 and A_2, such that $\tilde{H}_L(t) = \int_t^{\tau_L} \int_u^{\tau_L} \hat{f}_c^0(v)dvdu + A_1(t)$ and $\tilde{Y}_L(t) = \int_t^{\tau_L} \int_u^{\tau_L} dX(v)du + A_2(t)$. Thus by the characterization in Theorem 5.1.1 we have

$$\int_t^{\tau_L} \int_u^{\tau_L} \hat{f}_c^0(v)dvdu + A_1(t) = \tilde{H}_L(t) \geq \tilde{Y}_L(t) = \int_t^{\tau_L} \int_u^{\tau_L} dX(v)du + A_2(t) \qquad (5.17)$$

for $t < 0$, with equality at points of jump of $\hat{f}_c^0$ less than or equal 0. Using the equality of $\tilde{H}_L$ and $\tilde{Y}_L$ at the knot τ_L, and the fact that both $\int_{\tau_L}^{\tau_L} \int_u^{\tau_L} dX(v)du$ and $\int_{\tau_L}^{\tau_L} \int_u^{\tau_L} \hat{f}_c^0(v)dvdu$ are 0, we conclude that $A_1(\tau_L) = A_2(\tau_L)$. That is, we can write $A_1(t) - A_2(t)$ as $c_L(\tau_L - t)$, so that

$$\int_t^{\tau_L} \int_u^{\tau_L} \hat{f}_c^0(v)dvdu + c_L(\tau_L - t) = \tilde{H}_L(t) - A_2(t) \geq \tilde{Y}_L(t) - A_2(t) = \int_t^{\tau_L} \int_u^{\tau_L} dX(v)du$$

for $t < 0$, with equality at points of jump of $\hat{f}_c^0$ less than or equal to 0. Then, by Corollary 5.1.2, since we chose $\tau_L < 0$, we can conclude that the derivatives of each side of the above inequality *at* τ_L are equal, i.e.

$$c_L = \int_{\tau_L}^{\tau_L} \hat{f}_c^0(v)dv + c_L = \int_{\tau_L}^{\tau_L} dX(v) = 0,$$

and that

$$\int_t^{\tau_L} \int_u^{\tau_L} \hat{f}_c^0(v)dvdu \geq \int_t^{\tau_L} \int_u^{\tau_L} dX(v)du, \qquad (5.18)$$

for $t < 0$ with equality for any knot less than 0. The same argument works for the right hand side of 0 by setting $\tilde{H}_R$ and $\tilde{Y}_R$ to be R, writing $\tilde{H}_R(t) = \int_{\tau_R}^t \int_{\tau_R}^u \hat{f}_c^0(v)\,dvdu + A_3(t)$

and $\tilde{Y}_R(t) = \int_{\tau_R}^{t} \int_{\tau_R}^{u} dX(v)du + A_4(t)$ and using the equality of $\tilde{H}_R$ and $\tilde{Y}_R$ at the knot τ_R, to conclude

$$\int_{\tau_R}^{t} \int_{\tau_R}^{u} \hat{f}_c^0(v)dvdu + c_R(t-\tau_R) \geq \int_{\tau_R}^{t} \int_{\tau_R}^{u} dX(v)du, \tag{5.19}$$

and again argue that $c_R = 0$ by Corollary 5.1.2, because $\tau_R > 0$. We will now define our "inside-out" functions and summarize what we just showed. Note that the second equality in (5.20) and (5.21) depends on the argument above, i.e. on the fact that $c_R = 0$ and $c_L = 0$.

Definition 5.1.6. Define H_L, H_R, Y_L, Y_R, F_L, F_R, X_L, and X_R as follows. Let

$$H_L(t) := \tilde{H}_L(t) - A_2(t) = \int_{t}^{\tau_L} \int_{u}^{\tau_L} \hat{f}_c^0(v)dvdu =: \int_{t}^{\tau_L} F_L(u)du, \tag{5.20}$$

$$H_R(t) := \tilde{H}_R(t) - A_4(t) = \int_{\tau_R}^{t} \int_{\tau_R}^{u} \hat{f}_c^0(v)dvdu =: \int_{\tau_R}^{t} F_R(u)du, \tag{5.21}$$

and

$$Y_L(t) := \tilde{Y}_L(t) - A_2(t) = \int_{t}^{\tau_L} \int_{u}^{\tau_L} dX(v)du =: \int_{t}^{\tau_L} X_L(u)du, \tag{5.22}$$

$$Y_R(t) := \tilde{Y}_R(t) - A_4(t) = \int_{\tau_R}^{t} \int_{\tau_R}^{u} dX(v)du =: \int_{\tau_R}^{t} X_R(u)du, \tag{5.23}$$

where

$$F_L(u) = \int_{u}^{\tau_L} \hat{f}_c^0(v)dv \text{ and } F_R(u) = \int_{\tau_R}^{u} \hat{f}_c^0(v)dv, \tag{5.24}$$

$$X_L(u) = \int_{u}^{\tau_L} dX(v) \text{ and } X_R(u) = \int_{\tau_R}^{u} dX(v). \tag{5.25}$$

The above definitions trivially imply, for all $t \in [-c, c]$, that

$$H_L(t) - Y_L(t) = \tilde{H}_L(t) - \tilde{Y}_L(t) \;\text{ and }\; H_R(t) - Y_R(t) = \tilde{H}_R(t) - \tilde{Y}_R(t). \tag{5.26}$$

Additionally, the above shows what we already knew from the relationship $\tilde{c}_L + \int_{-c}^{\tau_L}(\hat{f}_c^0(v)dv -$

$dX(v)) = 0$, that since $H_L' - Y_L' = \tilde{F}_L - \tilde{X}_L$, $H_R' - Y_R' = -(\tilde{F}_R - \tilde{X}_R)$, and

$$H_L' - Y_L' = -(F_L - X_L) \text{ and } H_R' - Y_R' = (F_R - X_R) \tag{5.27}$$

that

$$-(F_L - X_L)(t) = (\tilde{F}_L - \tilde{X}_L)(t) = -(\tilde{F}_R - \tilde{X}_R)(t) = (F_R - X_R)(t) \tag{5.28}$$

(with the middle equality by (5.9)). With the above definitions in hand we can present the following inside-out version of Theorem 5.1.1.

Corollary 5.1.7 (Inside-out Characterization)**.** *Let* $\hat{f}_c^0 \colon [-c, c] \to \mathbb{R}$ *be measurable and define* τ_L *and* τ_R *by* (5.16). *Let* ϕ_c *be defined by* (5.8) *and assume its minimizer is a piecewise linear function whose points of jump, excluding* $\pm c$*, are isolated. Then* $\hat{f}_c^0 \in \mathcal{G}_{c,k}^0$ *is the unique minimizer of* ϕ_c *over* $\mathcal{G}_{c,k}^0$ *if and only if*

(i).

$$\int_{\tau_L}^{\tau_R} \hat{f}_c^0(v)dv = \int_{\tau_L}^{\tau_R} dX(v). \tag{5.29}$$

(ii).

$$H_L(t) - Y_L(t) \geq 0 \qquad \text{for} \qquad -c \leq t \leq 0 \tag{5.30}$$

$$H_R(t) - Y_R(t) \geq 0 \qquad \text{for} \qquad c \geq t \geq 0. \tag{5.31}$$

(iii).

$$\int_{[-c,\tau_{c,-1}^0]} (H_L(u) - Y_L(u))d(\hat{f}_c^0)'(u) = 0 = \int_{[\tau_{c,1}^0,c]} (H_R(u) - Y_R(u))d(\hat{f}_c^0)'(u). \tag{5.32}$$

Note that $(\hat{f}_c^0)'$ is 0 on the interval $(\tau_{c,-1}^0, \tau_{c,1}^0)$ so the bounds of one of the above integrals may be extendable if one of $\tau^0_{c,-1}$ or $\tau^0_{c,1}$ are not 0.

Proof. Conditions (ii) and (iii) follow from (5.26). By (5.26), it is trivial that (5.30), (5.31), and (5.32) hold if and only if (5.10), (5.11), and (5.12) also hold.

We may assume (5.10), (5.11), and (5.12) (equivalent to (5.30), (5.31), and (5.32)) since these are implied in the proof of necessity, when we assume $\hat{f}^0_c$ is the MLE (by Theorem 5.1.1), or assumed in the proof of sufficiency. Then we know that $\tilde{F}_L(\tau_L)-\tilde{X}_L(\tau_L) = 0 = \tilde{F}_R(\tau_R) - \tilde{X}_R(\tau_R)$ by Corollary 5.1.2 since $\tau_L < 0 < \tau_R$. Thus $(\tilde{F}_L - \tilde{X}_L)(0) = (\tilde{F}_L - \tilde{X}_L)(0) - (\tilde{F}_L - \tilde{X}_L)(\tau_L) = \int_{\tau_L}^0 (\hat{f}^0_c(v)dv - dX(v))$ and, similarly, $(\tilde{F}_R - \tilde{X}_R)(0) = \int_0^{\tau_R} (\hat{f}^0_c(v)dv - dX(v))$. Since $(\tilde{F}_R - \tilde{X}_R)(0) + (\tilde{F}_L - \tilde{X}_L)(0) = \tilde{c}_L + \tilde{c}_R + \int_{-c}^{c} (\hat{f}^0_c(v)dv - dX(v))$, we can conclude

$$\int_{\tau_L}^{\tau_R} \left(\hat{f}^0_c(v)dv - dX(v)\right) = \tilde{c}_L + \tilde{c}_R + \int_{-c}^{c} \left(\hat{f}^0_c(v)dv - dX(v)\right),$$

and so (5.29) is equivalent to (5.9) in the context of conditions (5.30), (5.31), and (5.32) or conditions (5.10), (5.11), and (5.12). That is, we have shown both necessity and sufficiency of the conditions (i), (ii) and (iii).

□

Remark 5.1.8. Note that the inside-out and outside-in *functions* are not equal, just their differences are. For example, $\tilde{H}_R(t) \neq \int_{\tau_R}^t \int_{\tau_R}^u \hat{f}^0_c(v)dvdu = H_R(t)$, $\tilde{H}_R(t) = \int_{\tau_R}^t \int_{\tau_R}^u \hat{f}^0_c(v)dvdu + A_1(t)$ but $H_R(t) = \int_{\tau_R}^t \int_{\tau_R}^u \hat{f}^0_c(v)dvdu + A_1(t) - A_2(t)$.

Remark 5.1.9. Note that by the equality of our inside-out $H - Y$ functions and our outside in $H - Y$ functions (i.e. by (5.26)) we have trivially shown that we can replace condition (5.9) in Theorem 5.1.1 with condition (5.29).

Remark 5.1.10. Recall that by (5.12), for $i < 0$ we have

$$0 = (\tilde{H}_L - \tilde{Y}_L)(\tau^0_{c,i}) = \tilde{c}_L(\tau^0_{c,i} + c) + \int_{-c}^{\tau^0_{c,i}} \int_{-c}^{u} (\hat{f}^0_c(v)dv - dX(v)).$$

Thus, we can conclude that

$$\frac{\int_{-c}^{\tau_{c,i}^0}\int_{-c}^{u}(\hat{f}_c^0(v)dv - dX(v))}{(\tau_{c,i}^0 + c)} = -\tilde{c}_L = \frac{\int_{-c}^{\tau_{c,j}^0}\int_{-c}^{u}(\hat{f}_c^0(v)dv - dX(v))}{(\tau_{c,j}^0 + c)}.$$

for $i, j < 0$. We can make similar conclusions based on $\tilde{H}_R - \tilde{Y}_R$.

Remark 5.1.11. We make one more remark, on the constraint (5.29) and how it relates to the finite sample case. In the finite sample case there is no precisely analogous constraint stated but we do have the fact that the estimated distribution function and the empirical distribution function are equal at $-\infty$ and at ∞. We will use Fact 5.1.5 (or its simple proof) repeatedly here. We note for the finite sample "outside-in" processes that we can write the left-processes as the right-processes:

$$\hat{H}_{n,R}^0(t) = (X_{(n)} - t) - (\hat{H}_{n,L}^0(X_{(n)}) - \hat{H}_{n,L}^0(t))$$
$$\mathbb{Y}_{n,R}(t) = (X_{(n)} - t) - (\mathbb{Y}_{n,L}(X_{(n)}) - \mathbb{Y}_{n,L}(t)),$$

for all $t \in \mathbb{R}$. Note that the slope term is the same in both instances, -1. This is because $\hat{F}_n^0(-\infty, \infty) = 1 = \mathbb{F}_n(-\infty, \infty)$,via the calculation

$$\int_{X_{(1)}}^{t}\int_{-\infty}^{v} = \left(\int_{X_{(1)}}^{X_{(n)}} - \int_{t}^{X_{(n)}}\right)\left(\int_{-\infty}^{\infty} - \int_{v}^{\infty}\right) = \int_{X_1}^{X_{(n)}}\int_{-\infty}^{v} - \int_{t}^{X_{(n)}}\int_{-\infty}^{\infty} + \int_{t}^{X_{(n)}}\int_{v}^{\infty},$$

leaving out the integrand. If we use either $d\hat{F}_n^0$ or $d\mathbb{F}_n$ as the integrand, then the $\int_{-\infty}^{\infty}$ term is 1 and so we get the $1 \cdot (X_{(n)} - t)$ term as stated. This means that $\mathbb{Y}_{n,R}(t) - \mathbb{Y}_{n,L}(t)$ is affine with the same slope term as $\hat{H}_{n,R}^0(t) - \hat{H}_{n,L}^0(t)$, or, rather, that $\hat{H}_{n,R}^0 - \mathbb{Y}_{n,R}$ equals $\hat{H}_{n,L}^0 - \mathbb{Y}_{n,L}$ minus a constant term. Now, so far we have discussed the outside-in processes for finite n. In the limit case we can't use outside-in processes so we use inside-out ones. So let's consider the inside-out ones for the finite sample case. Using the same calculation as above, we see with $C_n(u) = \mathbb{F}_n(u) - \hat{F}_n^0(u)$, that

$$\int_t^{\tau_{L,n}} \int_v^{\tau_{L,n}} dC_n(u) = \int_{X_{(1)}}^{\tau_{L,n}} \int_v^{\tau_{L,n}} dC_n(u) - \int_{X_{(1)}}^{t} \int_{-\infty}^{\tau_{L,n}} dC_n(u) + \int_{X_{(1)}}^{t} \int_{X_{(1)}}^{v} dC_n(u).$$

Now here the middle term is $\left(\int_{-\infty}^{\tau_{L,n}} dC_n(u)\right)\left(X_{(1)} - t\right)$. The slope, $\left(\int_{-\infty}^{\tau_{L,n}} dC_n(u)\right)$, is not 0 but is less than $1/n$ in absolute value. That is, if we were to define inside-out finite-sample processes, in addition to the double integral $\int_t^{\tau_{L,n}} \int_v^{\tau_{L,n}} dC_n(u)$, we would need to include the above adjustment slope term, of order $1/n$. This would then mean that the difference in the slope terms of our double inside-out integrals $\int_t^{\tau_{L,n}} \int_v^{\tau_{L,n}} dC_n(u)$ and $\int_{\tau_{R,n}}^{t} \int_{\tau_{R,n}}^{v} dC_n(u)$ would be equal to

$$\int_{-\infty}^{\infty} dC_n(u) - \int_{\tau_{L,n}}^{\tau_{R,n}} dC_n(u) = -\int_{\tau_{L,n}}^{\tau_{R,n}} dC_n(u),$$

which is of order $1/n$. Now in the limit case for the inside-out integrals, the constraint (5.29) can be rephrased as $(H_L - Y_L)' - (H_R - Y_R)' = 0$. That is, for outside-in integrals, having $\int_{-\infty}^{\infty} dC_n(u) = 0$ ensures that the difference in the slopes between the left- and right- $H - Y$ processes is 0. In the inside-out case (for finite samples), the term is $(\int_{-\infty}^{\infty} - \int_{\tau_{L,n}}^{\tau_{R,n}}) dC_n(u)$. As stated previously, in the limit case we do not have control of $\int_{\tau_L}^{\tau_R} (\hat{f}_c^0(u)du - dX(u))$ automatically from the other conditions, so forcing it to be 0 becomes a separate condition.

5.1.3 Limit Characterization on $(-\infty, \infty)$

In the previous section we characterized the minimizer of ϕ_c in terms of its once- and twice-integrated processes, restricting our attention to functions on compact sets. We now want an analogous characterization of the limiting process, but on all of $\mathbb{R}$. We can not simply minimize a functional analogous to ϕ_c, though, because for all functions $g(t)$ for which $\int_{-\infty}^{\infty} (g(t) - 12t^2)^2 dt = \infty$, we will get an infinite value for $\phi(g) = \frac{1}{2} \int_{-\infty}^{\infty} (g(t) -$

$12t^2)^2dt - \int_{-\infty}^{\infty}(g(t) - 12t^2)dW(t)$, which is one way to rewrite the objective function. (Note that this rewrite of the objective does not add any terms that depend on g, and also note that $\int_{-\infty}^{\infty} g\, dW$ is generally defined (and thus finite) for $g \in L^2(dt)$.) We do not want to restrict our attention to functions for which $\phi(g)$ is finite because this set is not closed and the function we are interested in (heuristically, the "minimizer" of ϕ) is not in that set. Rather, we will define our "minimizer" function via characterizing conditions in analogy to the previous sections. We will then show that such conditions, which are on the once- and twice-integrated processes, force uniqueness of any process that meets them. Later, when we show that such a process is the limit of the localized processes, we will automatically have proved its existence. First we will state the analogous result for the unconstrained case.

Theorem 5.1.12 (Groeneboom et al. (2001a))**.** *Let $W(t)$ be a two-sided Brownian motion starting from* 0*, let $X(t) = W(t) + 4t^3$, and let $Y(t) = \int_0^t W(s)ds + t^4$. Then there exists an almost surely uniquely defined random continuous function H satisfying the following conditions:*

(i). The function H is everywhere above Y,

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

(ii). H has a convex second derivative.

(iii). H satisfies

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

We think of H'' as the convex limit "estimator" of function $h_0(t) = 12t^2$, which is the limiting analog of $f_0(t)$, and is observed with added noise through dX where $dX(t) = dW(t) + 12t^2$, and we denote it by $\hat{f} := H''$. $\hat{f}$ is piecewise linear with separated knots, which we denote by τ_i for $i = \pm 1, \pm 2, ...$, where we take $\cdots < \tau_{-2} < \tau_{-1} < 0 < \tau_1 < \tau_2 < \cdots$. Next, we will give the uniqueness portion of the theorem for the modally-constrained case.

Theorem 5.1.13 (Uniqueness of Process on $\mathbb{R}$)**.** *Let $W(t)$ be a two-sided Brownian motion starting from 0, let $X(t) = W(t) - 4t^3$. Assume H_L and H_R are processes on $\mathbb{R}$ such that $H_L''(t) = H_R''(t)$ and this second derivative, denoted by $\hat{f}^0(t)$, is piecewise linear and convex with antimode at 0. We denote its knots as τ_i^0, which we take to be a (strictly) increasing sequence, with $\tau_i^0 \le 0$ if $i \le 0$ and $\tau_i^0 \ge 0$ if $i \ge 0$. We define τ_L and τ_R as in (5.16) and use them to define, F_L, F_R, X_L, and X_R by (5.24) and (5.25), and Y_L and Y_R by (5.22) and (5.23), extending the definitions for u and t outside of $[-c, c]$ (and substituting $\hat{f}^0$ for $\hat{f}_c^0$). If*

(i).

$$\int_{\tau_L}^{\tau_R} (\hat{f}^0(v)dv - dX(v)) = 0 \tag{5.33}$$

(ii).

$$H_L(t) - Y_L(t) \ge 0 \quad \textit{for} \quad t \le 0 \tag{5.34}$$

$$H_R(t) - Y_R(t) \ge 0 \quad \textit{for} \quad t \ge 0 \tag{5.35}$$

(iii).

$$\int_{(-\infty, \tau_{-1}^0]} (H_L(u) - Y_L(u)) d(\hat{f}^0)'(u) = 0 = \int_{[\tau_1^0, \infty)} (H_R(u) - Y_R(u)) d(\hat{f}^0)'(u). \tag{5.36}$$

then we may almost surely conclude that H_L and H_R are unique. We can conclude this uniqueness without assuming in advance that τ_L and τ_R are given, i.e., τ_L and τ_R are unique *as well.*

Much of the proof will be similar to the proof of sufficiency for Theorem 5.1.1. Note that since we defined F_L, F_R, X_L, X_R, Y_L, and Y_R with the same definitions (but different values) as in the $[-c, c]$ case (i.e. by (5.24), (5.25), (5.22), and (5.23)), we still have the same relationships that follow from those definitions, e.g. (5.27). Now, we will begin the proof of Theorem 5.1.13 by showing some Lemmas.

Lemma 5.1.14. Let the assumptions of Theorem 5.1.13 hold. Then, for any (fixed or random) $T \geq 0$, with probability 1 there are knots τ_+ and τ_- of $\hat{f}^0$ with $\tau_+ > T$ and $\tau_- < -T$.

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

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

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

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

for some new random coefficients, B_i (where only for $i = 0, 1$ are B_i not equal to A_i). Now, letting $\varphi(t) = \sqrt{\frac{2}{3}t^3 \log\log t}$, then by page 1714 of Lachal (1997) (or from page 238 of Watanabe (1970)), we know that almost surely $\limsup_{t\to\infty} \int_0^t W(u)du/\varphi(t) = 1$. Thus,

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

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

In proving Theorem 5.1.13, we will not be able to speak of $\hat{f}^0$ as a minimizer of an objective function, but we will instead show that $\hat{f}^0$ behaves as we would expect a minimizer to behave, i.e. for acceptable Δ perturbations we will see that $\int \Delta(t)(\hat{f}^0(t)dt - dX(t)) \geq 0$, analogously to $\hat{f}_n^0$ in Theorem 2.0.4 on page 28.

Proposition 5.1.15. *Let the assumptions of Theorem 5.1.13 hold. Let $\Delta \colon \mathbb{R} \to \mathbb{R}$ be convex with antimode at 0. If $-\infty < a < 0 < b < \infty$ are knots of $\hat{f}^0$ then*

$$\int_a^b \Delta(t)(\hat{f}^0(t)dt - dX(t)) \geq 0, \tag{5.37}$$

and thus, by Lemma 5.1.14,

$$\liminf_{a \to \infty} \int_{-a}^a \Delta(t)(\hat{f}^0(t)dt - dX(t)) \geq 0. \tag{5.38}$$

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

$$\begin{aligned}\int_a^b \Delta(t)(\hat{f}^0(t)dt - dX(t)) &= -\int_a^0 \Delta(t)(dF_L(t) - dX_L(t)) + \int_0^b \Delta(t)(dF_R(t) - dX_R(t)) \\ &= -\left[(\Delta(F_L - X_L))(a, 0] - \int_a^0 ((F_L - X_L)\Delta')(t)dt\right] \\ &\quad + (\Delta(F_R - X_R))(0, b] - \int_0^b ((F_R - X_R)\Delta')(t)dt.\end{aligned}$$

For a and b or any other knots which satisfy $a < 0 < b$ we have, by Fact 5.1.3 and hypotheses

(ii) and (iii), that $(F_R - X_R)(b) = 0 = (F_L - X_L)(a)$. Also recalling that $(H_L - Y_L)' = -(F_L - X_L)$ (i.e. (5.27)), we see that the above display equals

$$
\begin{aligned}
&- \Delta(0)((F_L - X_L)(0) + (F_R - X_R)(0)) \\
&\quad - \left[((H_R - Y_R)\Delta'(\cdot+))(0, b] - \int_0^b (H_R - Y_R)(t) d\Delta'(t) \right] \\
&\quad - \left[((H_L - Y_L)\Delta'(\cdot-))(a, 0] - \int_a^0 (H_L - Y_L)(t) d\Delta'(t) \right],
\end{aligned}
$$

which equals

$$
\begin{aligned}
&- \Delta(0) \left(\int_0^{\tau_L} (\hat{f}^0(t)dt - dX(t)) + \int_{\tau_R}^0 (\hat{f}^0(t)dt - dX(t)) \right) \\
&\quad + (H_R - Y_R)(0)\Delta'(0+) - (H_L - Y_L)(0)\Delta'(0-) \\
&\quad + \int_0^b (H_R - Y_R)(t) d\Delta'(t) + \int_a^0 (H_L - Y_L)(t) d\Delta'(t) \\
&\geq 0,
\end{aligned}
$$

where the inequality follows because each of the three lines in the final expression is greater than or equal to 0. The first line is equal to 0 by (5.33); the third line is ≥ 0 by (5.34) and (5.35), and the fact that Δ is convex so Δ' is monotonically increasing, i.e. yields a positive measure; the second line is ≥ 0 because Δ has antimode at 0, so that both $(H_R - Y_R)(0)$ and $\Delta'(0+)$ are nonnegative and both $-(H_L - Y_L)(0)$ and $\Delta'(0-)$ are nonpositive. □

The above proof actually applies to Δ such that $\hat{f}^0(t) + \epsilon\Delta(t) \in \mathcal{G}^0$, where $\mathcal{G}^0$ is the set of convex functions with antimode 0, but we will not need this per se. Rather, in the next result, we will express the same idea by showing $\int_a^b \hat{f}^0(t) \left(\hat{f}^0(t)dt - dX(t) \right) = 0$, and re-express this via integration by parts formulae.

Proposition 5.1.16. *Let the assumptions of Theorem 5.1.13 hold. We assume that a and* *b are knots such that $a < 0 < b$. Then*

$$\int_a^b \hat{f}^0(t)(\hat{f}^0(t)dt - dX(t)) = \int_a^0 ((F_L - X_L)(\hat{f}^0)')(t)dt - \int_0^b ((F_R - X_R)(\hat{f}^0)')(t)dt$$
$$= \int_{(a,\tau_L]} (H_L - Y_L)(t)d(\hat{f}^0)'(t) + \int_{[\tau_R,b)} (H_R - Y_R)(t)d(\hat{f}^0)'(t)$$
$$= 0.$$

Proof. We start by writing

$$\int_a^b \hat{f}^0(t)(\hat{f}^0(t)dt - dX(t)) = -\int_a^0 \hat{f}^0(t)d(F_L(t) - X_L(t)) + \int_0^b \hat{f}^0(t)d(F_R(t) - X_R(t)),$$

which, by integration by parts, equals

$$-\left[(\hat{f}^0(F_L - X_L))(a,0] - \int_a^0 ((F_L - X_L)(\hat{f}^0)')(t)dt\right]$$
$$+ (\hat{f}^0(F_R - X_R))(0,b] - \int_0^b ((F_R - X_R)(\hat{f}^0)')(t)dt,$$

and, as in the beginning of the proof of Proposition 5.1.15, we use the fact that $F_L(a) - X_L(a) = 0 = F_R(b) - X_R(b)$ to see that the above display equals

$$-\left[(\hat{f}^0(F_L - X_L))(0) - \int_a^0 ((F_L - X_L)(\hat{f}^0)')(t)dt\right]$$
$$- (\hat{f}^0(F_R - X_R))(0) - \int_0^b ((F_R - X_R)(\hat{f}^0)')(t)dt,$$

which, by (5.33), equals

$$\int_a^0 ((F_L - X_L)(\hat{f}^0)')(t)dt - \int_0^b ((F_R - X_R)(\hat{f}^0)')(t)dt.$$

Using that $\hat{f}^0$ is constant on (τ_L, τ_R), we can write the expression above as

$$\int_a^{\tau_L} ((F_L - X_L)(\hat{f}^0)')(t)dt - \int_{\tau_R}^b ((F_R - X_R)(\hat{f}^0)')(t)dt,$$

which shows the first equality of the proposition. By integration by parts and (5.27), this equals

$$-\left(((\hat{f}^0)'(\cdot+)(H_L - Y_L))(a, \tau_L] - \int_{(a,\tau_L]} (H_L - Y_L)(t)d(\hat{f}^0)'(t) \right)$$
$$- \left[((\hat{f}^0)'(\cdot-)(H_R - Y_R))(\tau_R, b] - \int_{[\tau_R,b)} (H_R - Y_R)(t)d(\hat{f}^0)'(t) \right],$$

which, by (5.36) and since a, τ_L, τ_R, and b are knots, equals

$$\int_{(a,\tau_L]} (H_L - Y_L)(t)d(\hat{f}^0)'(t) + \int_{[\tau_R,b)} (H_R - Y_R)(t)d(\hat{f}^0)'(t),$$

which shows the second equality of the proposition. The third, i.e. seeing that the above display is 0, follows again by (5.36). □

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

Lemma 5.1.17. Let the assumptions of Theorem 5.1.13 hold. Let τ_1, τ_2 be knots of $\hat{f}^0$ and let $t \in [\tau_1, \tau_2]$. For any function g, we define $\Delta g = g(\tau_2) - g(\tau_1)$ and $\bar{g} = g(\tau_1) + g(\tau_2)/2$. In this context we take τ to be the identity function, i.e. we define $\Delta\tau = \tau_2 - \tau_1$ and

$\bar{\tau} = (\tau_1 + \tau_2)/2$. Then if $0 < \tau_1 < \tau_2$, we can conclude

$$\begin{aligned} H_R(t) = &\frac{(Y_R(\tau_2)(t-\tau_1) + Y_R(\tau_1)(\tau_2 - t))}{\Delta\tau} \\ &- \frac{1}{2}\left(\frac{\Delta X_R}{\Delta\tau} + \frac{4}{(\Delta\tau)^3}(\bar{X}_R\Delta\tau - \Delta Y_R)(t-\bar{\tau})\right)(t-\tau_1)(\tau_2 - t), \end{aligned} \tag{5.39}$$

and thus

$$H_R(\bar{\tau}) = \bar{Y}_R - \frac{1}{8}\Delta X_R \Delta\tau. \tag{5.40}$$

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

$$\begin{aligned} H_L(t) = &\frac{Y_L(\tau_2)(t-\tau_1) + Y_L(\tau_1)(\tau_2 - t)}{\Delta\tau} \\ &- \frac{1}{2}\left(\frac{-\bar{X}_L}{\Delta\tau} + \frac{4}{(\Delta\tau)^3}(-\bar{X}_L\Delta\tau - \Delta Y_L)(t-\bar{\tau})\right)(t-\tau_1)(\tau_2 - t), \end{aligned} \tag{5.41}$$

and thus

$$H_L(\bar{\tau}) = \bar{Y}_L + \frac{1}{8}\Delta X_L \Delta\tau. \tag{5.42}$$

Proof. The proof idea is the same as for the unconstrained case, which is that between two knots $\hat{f}^0$ is linear and thus H_L and H_R are cubic polynomials. Thus, taking $\tau_1 < \tau_2 < 0$, H_L is defined by its values and its derivative's values at τ_i, for $i = 1, 2$. Thus, if we name the polynomial on the right hand side of (5.41) P_L, it suffices to check that $P_L(\tau_i)$ and $P_L'(\tau_i)$ equal $H_L(\tau_i)$ and $H_L'(\tau_i)$, respectively, for $i = 1, 2$, to conclude that $H_L(t) = P_L(t)$ for $t \in [\tau_1, \tau_2]$. We know that $H_L(\tau_i) = Y_L(\tau_i)$ by (5.36) and it is immediate that $P_L(\tau_i) = Y_L(\tau_i)$, so we only need to check the derivative values. To differentiate, we denote

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

so that

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

and

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

so that

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

This equals $H_L'(\tau_1)$, as desired, since $H_L'(\tau_1) = Y_L'(\tau_1)$ by Fact 5.1.3 since τ_1 is strictly less than 0, and $Y_L'(\tau_1) = -X_L(\tau_1)$. Similarly, $P_L'(\tau_2) = -X_L(\tau_2)$ and, letting P_R be the polynomial on the right hand side of (5.41), $P_R'(\tau_i) = X_R(\tau_i)$ and $P_R(\tau_i) = Y_R(\tau_i)$. Then (5.40) and (5.42) follow immediately. □

From the above result, as in the unconstrained case, one can derive a representation formula for the difference of $\hat{f}^0(t) - h_0(t)$ where $h_0(t) = 12t^2$. For example, for $0 < \tau_1 < \tau_2$ and for $V(t) = \int_0^t W(u)du$ and Δg and $\bar{g}$ as defined in the previous Lemma, we have

$$\hat{f}^0(t) - h_0(t) = (\Delta\tau)^2 + \frac{\Delta W}{\Delta\tau} + \frac{12}{(\Delta\tau)^3}\left(\bar{W}\Delta\tau - \Delta V\right)(t - \bar{\tau}) - 12(t-\bar{\tau})^2.$$

We will not need this formula or others like it, so we will not prove them. Rather, we will prove Lemmas analogous to the gap problem (i.e. Theorem 4.2.3) and tightness results (i.e. Corollary 4.2.7) in the finite sample case. Here, rather than showing that knot differences and function differences are tight as sample size varies, we show knot differences and function differences are tight as the location index t varies.

Lemma 5.1.18. Let the assumptions of Theorem 5.1.13 hold. For $t \in \mathbb{R}$, we let $\tau_+^0(t)$ and $\tau_-^0(t)$ be the first knot of $\hat{f}^0$ at or after the point t and before t, respectively. Then, for all $\epsilon > 0$ there exists M_ϵ such that for all $t > 0$,

$$P(\tau_+^0(t) > t + M_\epsilon) < \epsilon, \tag{5.43}$$

$$P(\tau_-^0(-t) < -t - M_\epsilon) < \epsilon, \tag{5.44}$$

$$P((t - M_\epsilon) \vee 0 \leq \tau_-^0(t) \vee 0) > 1 - \epsilon, \tag{5.45}$$

$$P(\tau_+^0(-t) \wedge 0 \leq (-t + M_\epsilon) \wedge 0) > 1 - \epsilon, \tag{5.46}$$

where M_ϵ is independent of t.

Proof. We will show for all $t, \epsilon > 0$ there exists $M = M_\epsilon$ such that $P(\tau_+^0(t) > t+M) < \epsilon$. The statement for $\tau_-^0(-t)$ is analogous. By Lemma 5.1.14, for any t we can find a knot larger than t; similarly, we can take t_ϵ large enough such that with probability $1 - \epsilon$ there exists a knot $0 < \tau_-^0(t) < t$. To match notation up with Lemma 5.1.17, we will rename $\tau_-^0(t)$ and $\tau_+^0(t)$ as τ_1 and τ_2 and define Δg and $\bar{g}$, for any function g, as in the lemma. Now, once we have $0 < \tau_1 < \tau_2$, Lemma 5.1.17 allows us to conclude that $Y_R(\bar{\tau}) \leq H_R(\bar{\tau}) = \bar{Y}_R - \Delta X_R \Delta\tau/8$ which is if and only if

$$Y(\bar{\tau}) \leq \bar{Y} - \frac{1}{8}\Delta X \Delta\tau, \tag{5.47}$$

where $Y(t) = \int_0^t X(u)du = V(t) + t^4$ and $V(t) = \int_0^t W(u)du$. The "if and only if" follows because $Y_R(t) = Y(t) + A(t)$ where $A(t)$ is a random affine function. Since for any affine function

$$A(\bar{\tau}) = A\left(\frac{\tau_1 + \tau_2}{2}\right) = \frac{A(\tau_1) + A(\tau_2)}{2} =: \bar{A},$$

we see that

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

Since ΔX trivially equals ΔX_R, we have shown (5.47). Now, by our choice of t_ϵ, for $M_\epsilon > 0$ we can say that

$$\begin{aligned} P(\tau_2 > t_\epsilon + M_\epsilon) &\leq P(\tau_1 \leq 0) + P(0 < \tau_1 < t_\epsilon, \tau_2 > t_\epsilon + M_\epsilon) \\ &\leq \epsilon + P\left(Y_R(\bar{\tau}) \leq H_R(\bar{\tau}) = \bar{Y}_R - \frac{1}{8}\Delta X_R \Delta\tau, 0 < \tau_1 < t_\epsilon, \tau_2 > t_\epsilon + M_\epsilon\right) \\ &\leq 2\epsilon, \end{aligned} \tag{5.48}$$

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

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

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

$$\begin{aligned} &P\left(Y_R(\bar{\tau}) \leq H_R(\bar{\tau}) = \bar{Y}_R - \frac{1}{8}\Delta X_R \Delta\tau, 0 < \tau_1 < t, \tau_2 > t + M_\epsilon\right) \\ &\quad = P\left(V(\bar{\tau}) - \bar{V} + \frac{1}{8}\Delta W \Delta\tau \leq -\left(\frac{\Delta\tau}{2}\right)^4, 0 < \tau_1 < t, \tau_2 > t + M_\epsilon\right). \end{aligned}$$

Groeneboom et al. (2001a) show that

$$P\left(V(\bar{\tau}) - \bar{V} + \frac{1}{8}\Delta W \Delta\tau \leq -\left(\frac{\Delta\tau}{2}\right)^4, \tau_1 < M_\epsilon, \tau_2 > M_\epsilon\right) < \epsilon, \tag{5.49}$$

and thus that

$$P\left(V(\bar{\tau})-\bar{V}+\frac{1}{8}\Delta W\Delta\tau\leq-\left(\frac{\Delta\tau}{2}\right)^{4},\tau_1<t-M_\epsilon,\tau_2>t+M_\epsilon\right)<\epsilon. \tag{5.50}$$

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

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

where $\bar{s}=(s_1+s_2)/2$, $\bar{V}=(V(s_1)+V(s_2))/2$, $\Delta W=W(s_2)-W(s_1)$, and $\Delta s=s_2-s_2$ and $r_i=s_i-t$ for $i=1,2$. This shows that the left hand sides of both of (5.49) and (5.50) are, regardless of t, bounded by

$$P\left(V(\bar{s})-\bar{V}+\frac{1}{8}\Delta W\Delta s\leq-\left(\frac{\Delta s}{2}\right)^{4},\ \text{for some}\ \ s_1<-M_\epsilon,s_2>M_\epsilon\right). \tag{5.51}$$

This probability is defined in (2.27) on page 1633 of Groeneboom et al. (2001a), and is shown to be less than ϵ at the top of page 1634, so, using this fact, we have now shown (5.49) and (5.50).

The probability we consider in (5.48) is on the event $\{0<\tau_1<t_\epsilon,\tau_2>t_\epsilon+M_\epsilon\}$ rather than $\{0<\tau_1<t_\epsilon-M_\epsilon,\tau_2>t_\epsilon+M_\epsilon\}$. The only cost for this is we need to double our M_ϵ for this to correspond with the probability in (5.50). Thus (5.48) holds, but we do not yet have independence from t because of the t_ϵ in the expression. We easily circumvent this by replacing M_ϵ by $t_\epsilon + M_\epsilon$. Now we have shown (5.43) holds with M_ϵ independent of t.

Now we show (5.44). Note that we can write an analogous version of (5.48) for $t > M_\epsilon$ as

$$\begin{aligned}
&P(0 \le \tau_1 \le t - M_\epsilon) \\
&\quad \le P(\tau_2 > t + M_\epsilon) + P(0 \le \tau_1 \le t - M_\epsilon, \tau_2 \le t + M_\epsilon) \\
&\quad \le \epsilon + P\left(Y_R(\bar{\tau}) \le H_R(\bar{\tau}) = \bar{Y}_R - \frac{1}{8}\Delta X_R \Delta\tau, 0 < \tau_1 < t_\epsilon - M_\epsilon, \tau_2 \le t_\epsilon + M_\epsilon\right) \\
&\quad \le 2\epsilon
\end{aligned}$$

because, by the argument we just went through, the probability in the third line is again bounded by (5.51). Note we have t in place of t_ϵ, so the above statement is already independent of t as long as $t > M_\epsilon$. Thus we have shown $P((t-M) \vee 0 \le \tau_-^0(t) \vee 0 \le t) > 1-\epsilon$, since if $t < M_\epsilon$ this probability is trivially 1.

Showing the analogous statements for the left side, the existence of M independent of t such that $P(\tau_-^0(t) < -t - M) < \epsilon$ and $P(-t \le \tau_+^0(-t) \wedge 0 \le (-t+M) \wedge 0) > 1-\epsilon$, can be done analogously. □

The next result will relate the unconstrained and constrained limit estimators. Recall the definitions of Y and H from Theorem 5.1.12, and recall that we let $\hat{f} = H''$ be the unconstrained limit estimator.

Corollary 5.1.19. *Let the assumptions of Theorem 5.1.13 hold. For any $t \in \mathbb{R}$, define*

$$s^+(t) = \inf\{s \in [t,\infty) | \hat{f}^0(s) = \hat{f}(s)\} \tag{5.52}$$

$$s^-(t) = \sup\{s \in (-\infty,t] | \hat{f}^0(s) = \hat{f}(s)\}. \tag{5.53}$$

Then we can say that for all $\epsilon > 0$, there exists M_ϵ such that

$$P(t - s^-(t) > M_\epsilon) < \epsilon \tag{5.54}$$

$$P(s^+(t) - t > M_\epsilon) < \epsilon, \tag{5.55}$$

where M_ϵ is independent of $t \in \mathbb{R}$.

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

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

where ν_i are knots for $\hat{f}$ and $\nu^0 i$ are knots for $\hat{f}^0$. Similarly, define a left-side sequence of knots

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

Then we argue by the Intermediate Value Theorem and the Mean Value Theorem.

First, we assume we are given such a sequence, without loss of generality take it to be a right-side sequence (on the probability 1 set on which Theorem 5.1.12 holds). Then we can say, by our hypotheses, that

$$(H_R - Y_R)(\nu_i^0) = 0 \leq (H - Y)(\nu_i^0) \quad \text{for } i = 1, 2 \tag{5.56}$$

$$(H_R - Y_R)(\nu_i) \geq 0 = (H - Y)(\nu_i) \quad \text{for } i = 1, 2, 3. \tag{5.57}$$

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

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

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

$$F_R(t_i) - F(t_i) - A'(t_i) = 0,$$

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

Now we will construct right-side sequences or left-side sequences of knots and be done. Note that by Lemma 5.1.18 and the analogous lemma for the unconstrained case, Lemma 2.7, page 1638, of Groeneboom et al. (2001a), there exists a large $M > 0$ such that with probability $1-\epsilon$ there exists a right-side sequence of knots contained in any interval of length $\geq M$ that lies in $[0, \infty)$ and a left-side sequence of knots in any interval of length $\geq M$ that lies in $(-\infty, 0]$. For any $t > 0$, note that the interval $[t - 2M, t]$ contains an interval of length at least M which lies either entirely in $(-\infty, 0]$ or entirely in $[0, \infty)$. Thus, $[t-2M, t]$ contains a one-sided sequence of knots, and thus an $s < t$ such that $\hat{f}^0(s) = \hat{f}(s)$, with probability $1-\epsilon$. Similarly, there exists a one-sided sequence of knots in $[t, t+M]$, and thus an $s > t$ such that $\hat{f}^0(s) = \hat{f}(s)$, with probability $1-\epsilon$. Thus, for $t > 0$, we have shown (5.54) and (5.55). Similarly, for $t < 0$, we consider intervals $[t-M, t]$ and $[t, t+2M]$ in which there exist one-sided sequences of knots, which allows us to conclude that $\hat{f}^0(s) = \hat{f}(s)$ for an $s > t$ and an $s < t$. □

Showing that $\hat{f}$ almost surely has knots going towards $\pm\infty$ would be another way to construct one-sided sequences of knots to finish the previous proof. Now we show a basic fact which will help us control the difference of two convex functions.

Fact 5.1.20. *Let g_1, g_0 on $[a,b]$ be two convex functions, and let $t \in [a,b]$. Then*

$$\begin{aligned} g_1(t) - g_0(t) \leq \frac{t-a}{b-a}(g_1(b) - g_0(b)) + \frac{b-t}{b-a}(g_1(a) - g_0(a)) \\ + \frac{(b-t)(t-a)}{b-a}(g_0^-(b) - g_0^+(a)). \end{aligned} \tag{5.58}$$

Proof. Let $\lambda = (t-a)/(b-a)$, which means that $1 - \lambda = (b-t)/(b-a)$. Then since $\lambda b + (1-\lambda)a = t$, by convexity we have

$$g_1(t) - g_0(t) \leq \lambda g_1(b) + (1-\lambda)g_1(a) - (\lambda g_0(t) + (1-\lambda)g_0(t)),$$

and, denoting the right and left derivatives as g^+ and g^-, respectively, we also have

$$g_0(t) \geq g_0(a) + (t-a)g_0^+(a) \quad \text{and} \quad g_0(t) \geq g_0(b) - (b-t)g_0^-(b).$$

Combining the above, we get

$$g_1(t) - g_0(t) \leq \lambda g_1(b) + (1-\lambda)g_1(a) - (\lambda g_0(t) + (1-\lambda)g_0(t))$$

which is bounded above by

$$\lambda g_1(b) + (1-\lambda)g_1(a) - (\lambda(g_0(b) - (b-t)g_0^-(b))) + (1-\lambda)(g_0(a) + (t-a)g_0^+(a)),$$

which equals

$$\begin{aligned} &\lambda(g_1(b) - g_0(b)) + (1-\lambda)(g_1(a) - g_0(a)) + (\lambda(b-t)g_0^-(b) - (1-\lambda)(t-a)g_0^+(a) \\ &\quad = \lambda(g_1(b) - g_0(b)) + (1-\lambda)(g_1(a) - g_0(a)) + \frac{(b-t)(t-a)}{b-a}\left(g_0^-(b) - g_0^+(a)\right), \end{aligned}$$

so we have shown (5.58). $\square$

Now we translate the result about the distribution of knots into one about the distribution of differences between our modally constrained limit estimator, $\hat{f}^0$, and the unconstrained limit estimator, $\hat{f}$.

Lemma 5.1.21. Let the assumptions of Theorem 5.1.13 hold. For a function g, let g^+ and g^- denote its right and left derivatives, respectively. For all $\epsilon > 0$ there exists M_ϵ such that

$$P\left(|\hat{f}^0(t) - \hat{f}(t)| > M_\epsilon\right) < \epsilon \tag{5.59}$$

$$P\left(|(\hat{f}^0)^+(t) - \hat{f}^+(t)| > M_\epsilon\right) < \epsilon \tag{5.60}$$

$$P\left(|(\hat{f}^0)^-(t) - \hat{f}^-(t)| > M_\epsilon\right) < \epsilon, \tag{5.61}$$

where M_ϵ is independent of t.

Proof. This follows from Lemma 5.1.18 and an argument similar to the finite sample tightness results. First we show (5.60) and (5.61), exactly as in Proposition 4.2.5. We can pick, by Corollary 5.1.19, $t - 2M < s_{-2} < s_{-1} < t < s_1 < s_2 < t + 2M$ where $\hat{f}^0(s_i) = \hat{f}(s_i)$ for $i = -2, -1, 1, 2$, with probability $1 - \epsilon$ for M appropriately large. Then

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

and, similarly,

$$(\hat{f}^0)'(t) \geq \hat{f}'(s_{-2})$$

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

$$(\hat{f}^0)'(t) - \hat{f}'(t) \leq \hat{f}'(s_2) - \hat{f}'(t) \leq \hat{f}'(s_2) - \hat{f}'(s_{-2})$$

and

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

that is,

$$|(\hat{f}^0)'(t) - \hat{f}'(t)| \leq \hat{f}'(s_2) - \hat{f}'(s_{-2}). \tag{5.62}$$

We can bound this last quantity with high probability by the standard argument

$$\begin{aligned}\hat{f}'(s_2) - \hat{f}'(s_{-2}) &\leq \hat{f}'(t+2M) - \hat{f}'(t-2M) \\ &\leq |\hat{f}'(t+2M) - h_0'(t+2M)| + |h_0'(t+2M) - h_0'(t-2M)| \\ &\quad + |h_0'(t-2M) - \hat{f}'(t-2M)|,\end{aligned} \tag{5.63}$$

which is less than $M + M + 24 \cdot 2M$ with probability $1-\epsilon$, independently of t, by (2.36) or (2.37) of Lemma 2.7 on page 1638 of Groeneboom et al. (2001a). Thus we have shown (5.60) and (5.61), which we will now use to show (5.59).

We first apply Fact 5.1.20 to the difference $|\hat{f}^0(t) - \hat{f}(t)|$ by applying (5.58) to both $\hat{f}^0 - \hat{f}$ and to $\hat{f} - \hat{f}^0$, using the points s_{-1} and s_1 as a and b, respectively. Then (5.58) tells us that we can bound both of these differences if we can bound both

$$(\hat{f}^0)'(s_1) - (\hat{f}^0)'(s_{-1}) \leq (\hat{f}^0)'(t+M) - (\hat{f}^0)'(t-M) \tag{5.64}$$

and

$$\hat{f}'(s_1) - \hat{f}'(s_{-1}) \leq \hat{f}'(t+M) - \hat{f}'(t-M), \tag{5.65}$$

since all the other terms are 0 by the definition of the s_i. Here we can take the derivatives to be either left or right derivatives. As in (5.63), we can bound $(\hat{f}^0)'(t+M) - (\hat{f}^0)'(t-M)$

from above by

$$|(\hat{f}^0)'(t+M) - \hat{f}'(t+M)| + |\hat{f}'(t+M) - \hat{f}'(t-M)| + |\hat{f}'(t-M) - (\hat{f}^0)'(t-M)|. \tag{5.66}$$

The first and last terms are bounded by (5.60) or (5.61). The middle term is shown to be bounded by (5.63). The middle term also bounds (5.65), the other term we needed to bound. All of this is with probability $1-\epsilon$ and independently of t, so we are done. □

Now that we have tightness of the difference of the constrained and unconstrained estimators, we can trivially translate this into tightness of the constrained estimator and the "truth", $h_0(t) = 12t^2$, since we have such tightness for the unconstrained estimator already.

Lemma 5.1.22. Let the assumptions of Theorem 5.1.13 hold. Let $(\hat{f}^0)^+(t)$ and $(\hat{f}^0)^-(t)$ denote the right and left derivatives of $\hat{f}^0$, respectively. Then, for all $\epsilon > 0$ there exists M_ϵ such that

$$P\left(|\hat{f}^0(t) - h_0(t)| > M_\epsilon\right) < \epsilon \tag{5.67}$$

$$P\left(|(\hat{f}^0)^+(t) - h_0'(t)| > M_\epsilon\right) < \epsilon \tag{5.68}$$

$$P\left(|(\hat{f}^0)^-(t) - h_0'(t)| > M_\epsilon\right) < \epsilon, \tag{5.69}$$

where M_ϵ is independent of t.

Proof. This is automatic from Lemma 2.7, page 1638, of Groeneboom et al. (2001a) and our previous result, Lemma 5.1.21. □

Now we prove the uniqueness theorem.

Proof of Theorem 5.1.13. We define new objective functions with variable bounds of integration,

$$\phi_{a,b}(g) = \frac{1}{2}\int_a^b g^2(t)dt - \int_a^b f(t)dX(t), \tag{5.70}$$

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

$$f_i = H''_{L,i} = H''_{R,i}. \tag{5.71}$$

We also for convenience will occasionally use the notation $F_i(t) = \int_0^t f_i(t)dt$, always in the form $dF_i(t) = f_i(t)dt$. Now, using that $f_1^2 - f_2^2 = (f_1 - f_2)^2 + 2(f_1 - f_2)f_2$, we see that

$$\begin{aligned}\phi_{a,b}(f_1) - \phi_{a,b}(f_2) =& \frac{1}{2}\int_a^b (f_1(t) - f_2(t))^2\, dt + \int_a^b (f_1(t) - f_2(t))\, f_2(t)dt - \int_a^b (f_1(t) - f_2(t))\, dX(t)\\ =& \frac{1}{2}\int_a^b (f_1(t) - f_2(t))^2\, dt + \int_a^b (f_1(t) - f_2(t))\, d(F_2(t) - X(t)).\end{aligned}$$

Now, we specify that a_n^i and b_n^i are knots for f_i, and, using Lemma 5.1.14, we take $a_n^2 < a_n^1 < -n < 0 < n < b_n^1 < b_n^2$. Then

$$\phi_{a_n^2,b_n^2}(f_1) - \phi_{a_n^2,b_n^2}(f_2) \geq \frac{1}{2}\int_{a_n^2}^{b_n^2} (f_1(t) - f_2(t))^2\, dt \geq \frac{1}{2}\int_{-n}^{n} (f_1(t) - f_2(t))^2\, dt$$

by Propositions 5.1.15 and 5.1.16, and, similarly, $\phi_{a_n^1,b_n^1}(f_2) - \phi_{a_n^1,b_n^1}(f_1) \geq \frac{1}{2}\int_{-n}^{n} (f_2(t) - f_1(t))^2\, dt$. Now, we see directly from (5.70) that

$$\begin{aligned}\phi_{a_n^2,b_n^2}(f_1) - \phi_{a_n^1,b_n^1}(f_1) =& \frac{1}{2}\int_{a_n^2}^{b_n^2} f_1^2(t)dt - \frac{1}{2}\int_{a_n^1}^{b_n^1} f_1^2(t)dt - \left(\int_{a_n^2}^{b_n^2} f_1(t)dX(t) - \int_{a_n^1}^{b_n^1} f_1(t)dX(t)\right)\\ =& \frac{1}{2}\int_{A_n} f_1^2(t)dt - \int_{A_n} f_1(t)dX(t),\end{aligned}$$

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

$$\int_{-n}^{n} (f_1(t) - f_2(t))^2\, dt \leq \phi_{a_n^2, b_n^2}(f_1) - \phi_{a_n^1, b_n^1}(f_1) - (\phi_{a_n^2, b_n^2}(f_2) - \phi_{a_n^1, b_n^1}(f_2))$$
$$= \frac{1}{2}\int_{A_n} \left(f_1(t)^2 - f_2(t)^2\right) dt - \int_{A_n} (f_1(t) - f_2(t))\, dX(t).$$

Recalling $h_0(t) = 12t^2$, we rewrite the above with the following algebra:

$$\begin{aligned}
&\frac{1}{2}\left(f_1^2 - f_2^2\right) d\lambda - (f_1 - f_2)\, dX \\
&\quad = \left(\frac{1}{2}f_1^2 - f_1 h_0 + \frac{1}{2}h_0^2 - \frac{1}{2}h_0^2 - (\frac{1}{2}f_2^2 - f_2 h_0 + \frac{1}{2}h_0^2 - \frac{1}{2}h_0^2\right) d\lambda - (f_1 - f_2)\, dW \\
&\quad = \frac{1}{2}\left((f_1 - h_0)^2 - (f_2 - h_0)^2\right) d\lambda - (f_1 - f_2)\, dW.
\end{aligned}$$

Thus we can conclude that

$$\begin{aligned}
0 &\leq \lim_n \int_{-n}^{n} (f_1 - f_2)^2\, d\lambda \\
&\leq \liminf_n \left(\frac{1}{2}\int_{A_n} \left((f_1 - h_0)^2 - (f_2 - h_0)^2\right) d\lambda - \int_{A_n} (f_1 - f_2)\, dW\right).
\end{aligned} \tag{5.72}$$

The proof will now proceed as follows. We first will show that the right hand side of (5.72) is finite. For a function $g\colon \mathbb{R} \to \mathbb{R}$, we introduce the notation

$$\|g\|_a^b = \sup_{t\in[a,b]} |g(t)| \quad \text{and} \quad \|g\|_a^\infty = \sup_{t\in[a,\infty)} |g(t)|. \tag{5.73}$$

Then once we know that $\int_{-\infty}^{\infty} (f_1 - f_2)^2\, d\lambda$ is finite we will also be able to say that $||f_1(t) - f_2(t)||_n^\infty \to 0$ as $n \to \infty$. We then will revisit our earlier argument which showed the right hand side of (5.72) was finite and use this new fact to show that it is, in fact, 0, and we will then be done.

Thus, our next step is to show that $\int_{-\infty}^{\infty} (f_1(t) - f_2(t))^2\, dt < \infty$. Note that we only need to control the $\liminf_n$ of the right hand side of (5.72) since $\int_{-n}^{n}(f_1 - f_2)^2\, d\lambda$ is non-negative and non-decreasing in n. We will first show that $\int_{b_n^1}^{b_n^2}(f_1 - f_2)\, dW < \infty$. The argument will be identical for $\int_{a_n^1}^{a_n^2}(f_1 - f_2) dW < \infty$. By integration by parts,

$$\begin{aligned}\int_{b_n^1}^{b_n^2} (f_1 - f_2)(u) dW(u) &= \int_{b_n^1}^{b_n^2} (f_1 - f_2)(u) d(W(u) - W(b_n^1)) \\ &= \left(W(b_n^2) - W(b_n^1)\right)\left(f_1(b_n^2) - f_2(b_n^2)\right) \\ &\quad - \int_{b_n^1}^{b_n^2} \left(W(u) - W(b_n^1)\right)\left(f_1'(u) - f_2'(u)\right) du\end{aligned} \tag{5.74}$$

where we can take f_i' to be the right-derivative, but this choice is inconsequential because of the almost sure continuity of W. Thus, by (5.79) and (5.80), for all n, with probability $1 - \epsilon$, we can conclude that

$$\left| \int_{b_n^1}^{b_n^2} (f_1 - f_2)(u) dW(u) \right| \le K_\epsilon \left(\left| W(b_n^2) - W(b_n^1) \right| + \left| \int_{b_n^1}^{b_n^2} (W(u) - W(b_n^1)) du \right| \right).$$

Lemma 5.1.23 shows for $i = 1, 2$ that $\int_{b_n^1}^{b_n^2}(f_i - h_0)^2 d\lambda < K_{\epsilon,2}$ with probability $1 - \epsilon$. Thus since, by (5.78), $\left(\left| W(b_n^2) - W(b_n^1) \right| + \left| \int_{b_n^1}^{b_n^2} (W(u) - W(b_n^1)) du \right| \right)$ is $O_p(1)$, and since this argument is perfectly symmetrical and applies to the interval $[a_n^1, a_n^2]$, we have now shown that the right hand side of (5.72) is $O_p(1)$ and thus finite almost surely, as desired.

Now that we have shown that $\int_{-\infty}^{\infty}(f_1 - f_2)^2 d\lambda < \infty$ almost surely, we can conclude that

$$\|f_1 - f_2\|_n^\infty \to 0 \tag{5.75}$$

almost surely as $n \to \infty$. We now use this new fact together with arguments similar to those we have already used to show that $\int_{-\infty}^{\infty} (f_1 - f_2)^2\, d\lambda = 0$ almost surely. Whereas we initially used Lemma 5.1.23 to conclude that $||f_1' - h_0'||_{b_n^1}^{b_n^2} = O_p(1)$, now by (5.75), Lemma 5.1.23 allows us to conclude that almost surely $\int_{b_n^1}^{b_n^2} |f_1' - h_0'| d\lambda \to 0$. Thus we can reexamine

(5.74) and, using Hölder's inequality with $p = 1$ and $q = \infty$, we see that

$$\left| \left(W(b_n^2) - W(b_n^1)\right) \left(f_1(b_n^2) - f_2(b_n^2)\right) - \int_{b_n^1}^{b_n^2} \left(W(u) - W(b_n^1)\right) \left(f_1'(u) - f_2'(u)\right) du \right|$$
$$\leq \left|W(b_n^2) - W(b_n^1)\right| \left|f_1(b_n^2) - f_2(b_n^2)\right| + ||W(\cdot) - W(b_n^1)||_{b_n^1}^{b_n^2} \int_{b_n^1}^{b_n^2} |f_1'(u) - f_2'(u)| du$$
$$\leq \epsilon \left(\left|W(b_n^2) - W(b_n^1)\right| + ||W(\cdot) - W(b_n^1)||_{b_n^1}^{b_n^2} \right),$$

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

$$\left| \int_{b_n^1}^{b_n^2} (f_1 - f_2)(u) dW(u) \right| \leq \epsilon. \tag{5.76}$$

Next we show that the other term in (5.72), $\int_{A_n} \left((f_1 - h_0)^2 - (f_2 - h_0)^2 \right) d\lambda/2$, is small. We have already shown that for any $\epsilon > 0$ we may pick an M_ϵ such that both $|\int_{b_n^1}^{b_n^2} (f_1 - h_0)\, d\lambda|$ and $b_n^2 - b_n^1$ are bounded by M_ϵ with probability $1 - \epsilon$. Thus, defining $\epsilon_2 = \epsilon / M_\epsilon$ we take n large enough such that with probability $1 - \epsilon$ we have $||f_1 - f_2||_n^\infty < \epsilon_2$. Then let $\delta(t) = f_1(t) - f_2(t)$ and conclude that

$$\int_{b_n^1}^{b_n^2} (f_1 - h_0)^2 d\lambda = \int_{b_n^1}^{b_n^2} (f_2 - h_0 + \delta)^2 d\lambda$$
$$\leq \int_{b_n^1}^{b_n^2} (f_2 - h_0)^2 d\lambda + 2\epsilon_2 \int_{b_n^1}^{b_n^2} |f_2 - h_0| d\lambda + \epsilon_2^2 (b_n^2 - b_n^1),$$

and that the above display is bounded above by

$$\int_{b_n^1}^{b_n^2} (f_2 - h_0)^2\, d\lambda + \frac{\epsilon}{M_\epsilon} M_\epsilon + \left(\frac{\epsilon}{M_\epsilon} \right)^2 M_\epsilon \leq \int_{b_n^1}^{b_n^2} (f_2 - h_0)^2\, d\lambda + 2\epsilon,$$

with probability $1-2\epsilon$ and n large enough. Similarly, $\int_{b_n^1}^{b_n^2}(f_2-h_0)^2d\lambda \le \int_{b_n^1}^{b_n^2}(f_1-h_0)^2d\lambda+2\epsilon$, and thus

$$\left|\frac{1}{2}\int_{A_n}\left((f_1-h_0)^2-(f_2-h_0)^2\right)d\lambda\right| \le \epsilon \tag{5.77}$$

with probability $1-2\epsilon$. Thus we have shown that with probability approaching 1 both terms in (5.72) are bounded by ϵ as n goes to infinity. Thus since $\int_{-n}^{n}(f_1-f_2)^2d\lambda$ is non decreasing in n, $\int_{-n}^{n}(f_1-f_2)^2d\lambda < \epsilon$ with probability $1-\epsilon$ and thus it must be 0 almost surely. □

The following lemma translates Lemma 5.1.22 into a more direct tightness result.

Lemma 5.1.23. Let the assumptions of Theorem 5.1.13 hold. Let f_i, $i=1,2$, be as in (5.71) and a_n^i and b_n^i as defined on page 147. We then have

$$b_n^2-b_n^1=O_p(1). \tag{5.78}$$

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

$$||f_i-h_0||_{b_n^1}^{b_n^2} < K_\epsilon \tag{5.79}$$

$$||f_i'-h_0'||_{b_n^1}^{b_n^2} < K_\epsilon, \tag{5.80}$$

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

$$\int_{b_n^1}^{b_n^2}|f_i-h_0|^k d\lambda < K_{\epsilon,k}, \tag{5.81}$$

where K_ϵ and $K_{\epsilon,k}$ do not depend on n. Further, if for a fixed random outcome ω we have that $\|f_1^\omega - f_2^\omega\|_n^\infty \to 0$ as $n \to \infty$ then we can conclude that

$$\int_{b_n^1}^{b_n^2} |(f_i^\omega)' - h_0'| \, d\lambda \to 0 \tag{5.82}$$

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

Proof. We will start by showing (5.78), which follows immediately from Lemma 5.1.18. This is because for any $\epsilon > 0$ we can pick an M_ϵ such that $P(b_n^1 - n < M_\epsilon)$ with probability $1-\epsilon/2$. Then using the same fact but applied to f_2, we can say that the probability that the first knot of f_2 after $n + M_\epsilon$ is after $n + 2M_\epsilon$ is less than $\epsilon/2$. Thus for any n,

$$P(b_n^2 - n > 2M_{\epsilon/2}) < P(b_n^1 - n > M_{\epsilon/2}) + P(b_n^1 - n \leq M_{\epsilon/2}, b_n^2 - n - M_{\epsilon/2} > M_{\epsilon/2}),$$

and this is less than ϵ.

Next we will show (5.79) and (5.80). Let g_1 and g_0 be monotone functions. Then for any $t \in [a, b]$, we have that

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

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

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

By monotonicity and Lemma 5.1.22, we can say

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

where f_i' refers to either the left or the right derivative. This is independent of n thanks to the linearity of h_0'. Thus we have shown (5.80).

Now we use Fact 5.1.20, about differences of convex functions, to establish (5.79). We allow f_i, $i = 1, 2$, and h_0 to g_1 and g_0. Regardless of the choice of which is g_1 and which is g_0, we know we can bound the first two terms in (5.58), the weighted differences $\lambda(g_1(n+M) - g_0(n+M)) + (1-\lambda)(g_1(n) - g_0(n))$, by $2M_\epsilon$ with probability $1-\epsilon$, independently of n. If h_0 is g_0 then for the third term of (5.58) we have to bound $(g_0^-(n+M) - g_0^+(n)) = 24M$ which is independent of n. If f_i is g_0, then for the third term of (5.58) we have

$$|f_i^-(n+M) - f_i^+(n))| \leq |f_i^-(n+M) - h_0'(n+M)| + |h_0'(n) - f_i^+(n)| + h_0'(n+M) - h_0'(n)$$

which we can again bound independently of n by the linearity of h_0' and Lemma 5.1.22 with probability $1-\epsilon$. Since $[b_n^1, b_n^2] \subset [n, n+M]$ with probability $1-\epsilon$ for appropriately large M, and since $(n+M-t)(t-n)/M \leq (M/2)^2/M$ which is independent of n and t, the bound is independent of n or t. Thus we have shown (5.79).

Next, (5.81) follows immediately from (5.79) and (5.78), since we can bound

$$\int_{b_n^1}^{b_n^2} |f_i - h_0|^k d\lambda \leq \int_{b_n^1}^{b_n^2} K_\epsilon^k d\lambda \leq K_\epsilon^k \cdot K_\epsilon,$$

with probability $1-\epsilon$.

Finally, we show that if for a random outcome ω, $||f_1^\omega - f_2^\omega||_n^\infty \to 0$ as $n \to \infty$ then (5.82) follows. First, note for any a, b that if $\epsilon < \int_a^b ((f_i^\omega)' - h_0')d\lambda = (f_i^\omega - h_0)(b) - (f_i^\omega - h_0)(a)$, and if $(f_i^\omega - h_0)(b) > \epsilon/2$ then $(f_i^\omega - h_0)(a) < -\epsilon/2$. Similarly, if $-\epsilon > \int_a^b ((f_i^\omega)' - h_0')d\lambda$ we can conclude that $(f_i^\omega - h_0)$ at a or at b is larger than $\epsilon/2$ in absolute value. Since we can take n large enough that $|f_i^\omega - h_0|$ is less than $\epsilon/2$ at any $a, b > n$, by contradiction we have that $|\int_a^b ((f_i^\omega)' - h_0')d\lambda| < \epsilon$ for such a and b. Now, since $\{t \in [b_n^1, b_n^2) | (f_1^\omega)'(t) > (f_2^\omega)'(t)\}$ and $\{t \in [b_n^1, b_n^2) | (f_1^\omega)'(t) \leq (f_2^\omega)'(t)\}$ are both intervals by monotonicity of $(f_1^\omega)'$ and linearity of

$(f_2^\omega)'$ on $[b_n^1, b_n^2]$, we can conclude that $\int_a^b |(f_i^\omega)' - h_0'| d\lambda < \epsilon$ as desired. □

5.2 Local Limit Distribution

We will now study processes localized at the mode, m, and will use tightness results from Section 4.2 and the results of Section 5.1 to find the process limit distributions. Finding the process limit distributions is the goal of this section, and this goal is accomplished in Theorem 5.2.13, on page 177. We begin by defining two versions of the localized process, one at the density level and one at the log-density level. The latter is more fundamental since it is where the concavity shape constraint really comes forth. The former, of course, yields the asymptotics for the density estimators. The two are related via identities given below in Definition 5.2.17 and Lemma 5.2.18. In this section we will have several definitions and redefinitions of our processes. We have collected many of these into a summary section, Section 5.2.1, beginning on page 189, which is intended to be printed as a reference to aid in reading. In the localized expressions we define below, we will leave in the slope terms (with coefficient $f_0'(m)$ or $\varphi_0'(m)$) despite the fact that the slopes are 0, to make them more transparent and to be in line with the analysis in Balabdaoui et al. (2009) on page 1313. This will make it easier to distinguish when we are relying on having 0 slope and when we are not.

Definition 5.2.1. (Localized-at-the-Mode Processes): For $t \in \mathbb{R}$, define

$$m_n(t) := m + tn^{-1/5}, \text{ so that } m_n^{-1}(s) = n^{1/5}(s - m),$$

and define τ_n^L to be the first knot strictly less than m and τ_n^R to be the first knot strictly greater than m. We will define

$$\tilde{f}_0(u) = f_0(m) + f_0'(m)(u-m) = f_0(m),$$

$$\tilde{\varphi}_0(u) = \varphi_0(m) + \varphi_0'(m)(u-m) = \varphi_0(m),$$

where the notation $\tilde{f}_0(u)$ is used in place of $f_0(m)$ simply to remind ourselves that we are approximating $f_0(u)$ with a one term Taylor expansion where the derivative happens to be 0. Next, we define

$$\begin{aligned}
\mathbb{Y}_n^f(t) &:= n^{4/5} \int_m^{m_n(t)} \left(\int_m^v d\mathbb{F}_n(u) - \int_m^v \tilde{f}_0(u)du \right) dv \\
\hat{H}_n^f(t) &:= n^{4/5} \int_m^{m_n(t)} \left(\int_m^v \hat{f}_n(u)du - \int_m^v \tilde{f}_0(u)du \right) dv + A_n t + B_n \\
\mathbb{Y}_{n,L}^f(t) &:= n^{4/5} \int_{m_n(t)}^{\tau_n^L} \left(\int_v^{\tau_n^L} d\mathbb{F}_n(u) - \int_v^{\tau_n^L} \tilde{f}_0(u)du \right) dv \\
\mathbb{Y}_{n,R}^f(t) &:= n^{4/5} \int_{\tau_n^R}^{m_n(t)} \left(\int_{\tau_n^R}^v d\mathbb{F}_n(u) - \int_{\tau_n^R}^v \tilde{f}_0(u)du \right) dv \\
\hat{H}_{n,L}^f(t) &:= n^{4/5} \int_{m_n(t)}^{\tau_n^L} \left(\int_v^{\tau_n^L} \hat{f}_n^0(u)du - \int_v^{\tau_n^L} \tilde{f}_0(u)du \right) dv + A_n^L(m_n^{-1}(\tau_n^L) - t) + B_n^L \\
\hat{H}_{n,R}^f(t) &:= n^{4/5} \int_{\tau_n^R}^{m_n(t)} \left(\int_{\tau_n^R}^v \hat{f}_n^0(u)du - \int_{\tau_n^R}^v \tilde{f}_0(u)du \right) dv + A_n^R(t - m_n^{-1}(\tau_n^R)) + B_n^R
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{Y}_n^{\varphi}(t) &:= \frac{\mathbb{Y}_n^f(t)}{f_0(m)} - n^{4/5}\int_m^{m_n(t)}\int_m^v (D(u)+R(u))\,dudv \\
\hat{H}_n^{\varphi}(t) &:= n^{4/5}\int_m^{m_n(t)}\int_m^v (\hat{\varphi}_n(u)-\tilde{\varphi}_0(u))\,dudv + \frac{A_n t + B_n}{f_0(m)} \\
\mathbb{Y}_{n,L}^{\varphi}(t) &:= \frac{\mathbb{Y}_{n,L}^f(t)}{f_0(m)} - n^{4/5}\int_{m_n(t)}^{\tau_n^L}\int_v^{\tau_n^L} \left(D_n^0(u)+R_n^0(u)\right)dudv \\
\mathbb{Y}_{n,R}^{\varphi}(t) &:= \frac{\mathbb{Y}_{n,R}^f(t)}{f_0(m)} - n^{4/5}\int_{\tau_n^R}^{m_n(t)}\int_{\tau_n^R}^v \left(D_n^0(u)+R_n^0(u)\right)dudv \\
\hat{H}_{n,L}^{\varphi}(t) &:= n^{4/5}\int_{m_n(t)}^{\tau_n^L}\int_v^{\tau_n^L} \left(\hat{\varphi}_n^0(u)-\tilde{\varphi}_0(u)\right)dudv + \frac{A_n^L(m_n^{-1}(\tau_n^L)-t)+B_n^L}{f_0(m)} \\
\hat{H}_{n,R}^{\varphi}(t) &:= n^{4/5}\int_{\tau_n^R}^{m_n(t)}\int_{\tau_n^R}^v \left(\hat{\varphi}_n^0(u)-\tilde{\varphi}_0(u)\right)dudv + \frac{A_n^R(t-m_n^{-1}(\tau_n^R))+B_n^R}{f_0(m)}
\end{aligned}$$

where

$$\begin{aligned}
A_n = n^{3/5}\left(\hat{F}_n(m)-\mathbb{F}_n(m)\right) \quad &\text{and} \quad B_n = n^{4/5}\left(\hat{H}_n(m)-\mathbb{Y}_n(m)\right), \\
A_n^L = n^{3/5}\left(\mathbb{F}_{n,L}(\tau_n^L)-\hat{F}_{n,L}^0(\tau_n^L)\right) \quad &\text{and} \quad B_n^L = n^{4/5}\left(\hat{H}_{n,L}^0(\tau_n^L)-\mathbb{Y}_{n,L}(\tau_n^L)\right)=0, \\
A_n^R = n^{3/5}\left(\mathbb{F}_{n,R}(\tau_n^R)-\hat{F}_{n,R}^0(\tau_n^R)\right) \quad &\text{and} \quad B_n^R = n^{4/5}\left(\hat{H}_{n,R}^0(\tau_n^R)-\mathbb{Y}_{n,R}(\tau_n^R)\right)=0
\end{aligned}$$

and

$$\begin{aligned}
D_n(u) = \frac{1}{2}\left(\hat{\varphi}_n(u)-\varphi_0(m)\right)^2 \quad &\text{and} \quad R_n(u) = \sum_{j=3}^{\infty}\frac{1}{j!}\left(\hat{\varphi}_n(u)-\varphi_0(m)\right)^j \\
D_n^0(u) = \frac{1}{2}\left(\hat{\varphi}_n^0(u)-\varphi_0(m)\right)^2 \quad &\text{and} \quad R_n^0(u) = \sum_{j=3}^{\infty}\frac{1}{j!}\left(\hat{\varphi}_n^0(u)-\varphi_0(m)\right)^j.
\end{aligned}$$

Note that in Balabdaoui et al. (2009), $\mathbb{Y}_n^f$ is denoted by $\mathbb{Y}_n^{loc}$, $\mathbb{Y}_n^{\varphi}$ is denoted by $\mathbb{Y}_n^{locmod}$, and similarly for $\hat{H}_n^f$ and $\hat{H}_n^{\varphi}$. The definitions of A_n, B_n, A_n^L, B_n^L, A_n^R, and B_n^R are motivated by (5.95) below. Note that A_n^L and A_n^R appear to be different than A_n by a sign change.

This comes from the definitions of our left- and right-processes, which entails that, e.g., $(\hat{H}^0_{n,R} - \mathbb{Y}_{n,R})'(t) = -(\hat{F}^0_{n,R} - \mathbb{F}_{n,R})(t)$, which accounts for the apparent difference. Also, if we had expanded at a more general point than m, then $D(u)$ would be a term containing drift whereas $R(u)$ would still be a negligible remainder term; however, since $\varphi_0'(m) = 0$, they are both negligible. Note that $D(u) + R(u)$ is denoted by $\Psi_{n,2}(u)$ in Balabdaoui et al. (2009). Finally, notice that we have *defined* the relationship between $\mathbb{Y}_n^f$ and $\mathbb{Y}_n^\varphi$, via the terms D and R, whereas for $\hat{H}_n^f$ and $\hat{H}_n^\varphi$, we will *show* that a similar relationship holds. Thus, the D and R terms actually arise as the difference between $\hat{H}_n^f$ and $\hat{H}_n^\varphi$ and we build them into the definitions of $\mathbb{Y}_n^\varphi$ (and, similarly for the left- and right-side processes). We will now see the just-mentioned explanation for the D and R processes, as well as some other identities, in the following lemma.

Lemma 5.2.2 (Identities)**.** Under Assumptions A and D, we have the following identities. First,

$$f_0(m)^{-1}\left(\hat{f}_n(u) - \tilde{f}_0(u)\right) = \hat{\varphi}_n(u) - \tilde{\varphi}_0(u) + D_n(u) + R_n(u). \tag{5.83}$$

Next, the f-processes and the φ-processes are related by

$$\hat{H}_n^\varphi(t) = \frac{\hat{H}_n^f(t)}{f_0(m)} - n^{4/5}\int_m^{m_n(t)}\int_m^v (D_n(u) + R_n(u))\,dudv \tag{5.84}$$

$$\hat{H}_{n,L}^\varphi(t) = \frac{\hat{H}_{n,L}^f(t)}{f_0(m)} - n^{4/5}\int_{m_n(t)}^{\tau_n^L}\int_v^{\tau_n^L} \left(D_n^0(u) + R_n^0(u)\right) dudv \tag{5.85}$$

$$\hat{H}_{n,R}^\varphi(t) = \frac{\hat{H}_{n,R}^f(t)}{f_0(m)} - n^{4/5}\int_{\tau_n^R}^{m_n(t)}\int_{\tau_n^R}^v \left(D_n^0(u) + R_n^0(u)\right) dudv. \tag{5.86}$$

(Compare these identities to the definitions of $\mathbb{Y}_n^\varphi(t), \mathbb{Y}_{n,L}^\varphi(t)$, and $\mathbb{Y}_{n,R}^\varphi(t)$.) Now, taking $u \in [m - Mn^{-1/5}, m + Mn^{-1/5}]$ for any $M > 0$, we have

$$R_n(u) - o_p(n^{-2/5}) = R_n^0(u) - o_p(n^{-2/5}) = 0, \tag{5.87}$$

$$D_n(u) - o_p(n^{-2/5}) = D_n^0(u) - o_p(n^{-2/5}) = \frac{1}{2}\left(\varphi_0'(m)(u-m)\right)^2 = 0, \tag{5.88}$$

all uniformly for $u \in [m - Mn^{-1/5}, m + Mn^{-1/5}]$.

Additionally, recalling the notation $g(a,b] = g(b) - g(a)$, we have

$$\left|\left((\hat{H}_{n,R}^{\varphi})' - (\mathbb{Y}_{n,R}^{\varphi})'\right)\left(a_n^{-1}(\tau_n^L), a_n^{-1}(\tau_n^R)\right]\right| \le \frac{4}{f_0(m)} n^{-2/5} \tag{5.89}$$

with high probability for large n, and

$$\mathbb{Y}_{n,L}^{\varphi}(t) - \hat{H}_{n,L}^{\varphi}(t) \ge 0, \tag{5.90}$$

$$\mathbb{Y}_{n,R}^{\varphi}(t) - \hat{H}_{n,R}^{\varphi}(t) \ge 0, \tag{5.91}$$

and

$$\int_{\infty}^{\tau_n^{-1}} (\mathbb{Y}_{n,L}^{\varphi}(t) - \hat{H}_{n,L}^{\varphi}(t)) d(\hat{H}_{n,L}^{\varphi})^{(3)}(t) = 0 = \int_{\tau_n^1}^{\infty} \left(\mathbb{Y}_{n,R}^{\varphi}(t) - \hat{H}_{n,R}^{\varphi}(t)\right) d(\hat{H}_{n,R}^{\varphi})^{(3)}(t) \tag{5.92}$$

where τ_n^1 is the first right-knot of $(\hat{H}_{n,R}^{\varphi})''$ greater or equal to m and τ_n^{-1} is the first left-knot less than or equal to m.

Proof. We first show (5.83). We start by the exponential series expansion, which we compute about the density at m,

$$\hat{g}(u) - f_0(m) = f_0(m)(\exp\{\hat{\varphi}(u) - \varphi_0(m)\} - 1) = f_0(m) \sum_{j=1}^{\infty} \frac{1}{j!} (\hat{\varphi}(u) - \varphi_0(m))^j, \tag{5.93}$$

where $\hat{g}$ is either $\hat{f}_n$ or $\hat{f}_n^0$, and $\hat{\varphi}$ is either $\hat{\varphi}_n$ or $\hat{\varphi}_n^0$, respectively. Now (5.84), (5.85), and (5.86) follow directly from (5.83) and the definitions of the processes. For example, the definition of $\hat{H}^f_{n,L}$ is

$$\hat{H}^f_{n,L}(t) = n^{4/5} \int_{m_n(t)}^{\tau_n^L} \int_v^{\tau_n^L} (\hat{f}_n^0(u) - \tilde{f}_0(u)) du dv + A_n^L(m_n^{-1}(\tau_n^L) - t) + B_n^L,$$

which by (5.93) is

$$n^{4/5} f_0(m) \int_{m_n(t)}^{\tau_n^L} \int_v^{\tau_n^L} (\hat{\varphi}_n^0(u) - \tilde{\varphi}_0(u) + D_n^0(u) + R_n^0(u)) du dv + A_n^L(m_n^{-1}(\tau_n^L) - t) + B_n^L,$$

as desired.

The proofs for (5.87) and (5.88) follow from writing

$$\begin{aligned}(\hat{\varphi}(u) - \varphi_0(m))^j &= \left(\hat{\varphi}(u) - \varphi_0(m) - \varphi_0'(m)(u-m) + \varphi_0'(m)(u-m)\right)^j \\ &= \sum_{i=0}^{j} \binom{j}{i} \left(\hat{\varphi}(u) - \varphi_0(m) - \varphi_0'(m)(u-m)\right)^{j-i} \left(\varphi_0'(m)(u-m)\right)^i,\end{aligned} \tag{5.94}$$

where we take $\hat{\varphi}$ to be either $\hat{\varphi}_n$ or $\hat{\varphi}_n^0$. Then if we take $j = 2$, the above display is $O_p(n^{-4/5}) + O_p(n^{-3/5}) + (\varphi_0'(m)(u-m))^2$, by (4.34) for $\hat{\varphi} = \hat{\varphi}_n^0$, and the analogous (4.17) on page 1319 of Balabdaoui et al. (2009) for $\hat{\varphi} = \hat{\varphi}_n$ (and by the fact $u - m = O_p(n^{-1/5})$, which is by definition). We did not use that $\varphi_0'(m) = 0$ here. Next, using (5.94) for $j \geq 3$, we see that $(\hat{\varphi}(u) - \varphi_0(m))^j \leq K n^{-j/5}$ (again without using that $\varphi_0'(m) = 0$) for some $K = O_p(1)$, and thus $|R_n(u)|$ and $|R_n^0(u)|$ are bounded by

$$\sum_{j=3}^{\infty} K n^{-j/5} = \frac{K n^{-3/5}}{1 - n^{-1/5}} = O_p(n^{-3/5}) = o_p(n^{-2/5}),$$

so we have shown (5.87).

Now we show (5.89). We have

$$(\hat{H}^f_{n,R})'(t) - (\mathbb{Y}^f_{n,R})'(t) = n^{3/5} \int_{\tau_n^R}^{m_n(t)} (\hat{f}^0_n(u)du - d\mathbb{F}_n(u)) + A_n^R$$

so that

$$\begin{aligned}(\hat{H}^f_{n,R})'(m_n^{-1}(\tau_n^R)) - (\mathbb{Y}^f_{n,R})'(m_n^{-1}(\tau_n^R)) - \Big((\hat{H}^f_{n,R})'(m_n^{-1}(\tau_n^L)) - (\mathbb{Y}^f_{n,R})'(m_n^{-1}(\tau_n^L))\Big) \\ = n^{3/5} \int_{\tau_n^L}^{\tau_n^R} (\hat{f}^0_n(u)du - d\mathbb{F}_n(u)).\end{aligned}$$

Thus, similarly,

$$\Big((\hat{H}^\varphi_{n,R})' - (\mathbb{Y}^\varphi_{n,R})'\Big)\left(m_n^{-1}(\tau_n^L), m_n^{-1}(\tau_n^R)\right] = \frac{1}{f_0(m)} n^{3/5} \int_{\tau_n^L}^{\tau_n^R} (\hat{f}^0_n(u)du - d\mathbb{F}_n(u))$$

which is bounded by $4n^{-2/5}/f_0(m)$ with high probability, by applying Corollary 2.0.9 twice (recall that we defined τ_n^R and τ_n^L to be not equal to m, which we did precisely so we as to be able to apply Corollary 2.0.9). Thus we have shown (5.89).

Last, we show (5.90) and (5.91). First, we use the definitions of A_n^L, A_n^R, B_n^L, and B_n^R to show inequalities analogous to (5.90) and (5.91) but for the f-processes. We have

$$\mathbb{Y}^f_{n,R}(t) - \hat{H}^f_{n,R}(t) = n^{4/5} \int_{\tau_n^R}^{m_n(t)} \left(\int_{\tau_n^R}^v d\mathbb{F}_n(u) - \int_{\tau_n^R}^v \hat{f}^0_n(u)du \right) dv - (t - m_n^{-1}(\tau_n^R))A_n^R - B_n^R,$$

which, by the definition of A_n^R and m_n^{-1}, equals

$$\begin{aligned} n^{4/5} \int_{\tau_n^R}^{m_n(t)} \left(\int_{\tau_n^R}^{X_{(n)}} (d\mathbb{F}_n(u) - \hat{f}^0_n(u)du) - \int_v^{X_{(n)}} (d\mathbb{F}_n(u) - \hat{f}^0_n(u)du) \right) dv \\ - n^{4/5} \left(\frac{t}{n^{1/5}} - (\tau_n^R - m) \right) \int_{\tau_n^R}^{X_{(n)}} (d\mathbb{F}_n(u) - \hat{f}^0_n(u)du) - B_n^R, \end{aligned}$$

which equals

$$-n^{4/5}\int_{\tau_n^R}^{m_n(t)}\left(\int_v^{X_{(n)}}(d\mathbb{F}_n(u)-\hat{f}_n^0(u)du)\right)dv - B_n^R,$$

which equals

$$n^{4/5}\int_{m_n(t)}^{X_{(n)}}\left(\int_v^{X_{(n)}}(d\mathbb{F}_n(u)-\hat{f}_n^0(u)du)\right)dv, \tag{5.95}$$

by the definition of B_n^R, which is $-B_n^R = n^{4/5}\int_{\tau_n^R}^{X_{(n)}}\left(\int_v^{X_{(n)}}(d\mathbb{F}_n(u)-\hat{f}_n^0(u)du)\right)dv$. (Note that the fact that $\mathbb{Y}_{n,R}^f(t)-\hat{H}_{n,R}^f(t)$ equals (5.95) provides the motivation for the definitions of A_n^L, A_n^R, B_n^L, and B_n^R.) Thus,

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

for all $t\geq 0$, with equality if $m_n(t)$ is a right-knot, by Theorem 2.0.6. Now since by definition and by (5.86) we have

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

we can conclude for $t\geq 0$ that

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

with equality if $m_n(t)$ is a right-knot, as desired. We have thus shown (5.91) and the right side of (5.92), and (5.90) and the left side of (5.92) are similar. □

Next we will note that the terms A_n^L and A_n^R, defined in Definition 5.2.17, are asymptotically negligible. As in Corollary 2.0.9, here the only thing we use about our data observations X_i are that they are unique (an assumption which could be relaxed by introducing weights). Thus, the following is not a probabilistic statement.

Lemma 5.2.3. We have

$$|A_n^L| \to 0 \;\text{ and }\; |A_n^R| \to 0, \tag{5.96}$$

as $n \to \infty$, deterministically.

Proof. $|A_n^L| = n^{3/5} \left| \mathbb{F}_{n,L}(\tau_n^L) - \hat{F}^0_{n,L}(\tau_n^L) \right| \leq n^{3/5} n^{-1} \to 0$, since we can apply Corollary 2.0.9 because τ_n^L is strictly less than 0. Similarly, since $\mathbb{F}_{n,L}(X_{(n)}) = 1 = \mathbb{F}_n(X_{(n)})$, by the same corollary, $|A_n^R| \to 0$. □

We will use the following metric spaces for our results about convergence in distribution.

Definition 5.2.4. For $0 < c \leq \infty$, define

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

and

$$\mathcal{F}_{(-c,c),M} := \{f \in \mathcal{D}_c | f \text{ is non-decreasing and } \|f\| \leq M\}. \tag{5.97}$$

where "cadlag" means right-continuous functions which have limits from the left, and if $c = \infty$ then we take $\mathcal{C}_\infty$ to be continuous functions h defined on $(-\infty, \infty)$ (i.e. we do not assert anything about $h(\pm\infty)$). We let $\|f\|$ be the supremum of f over its domain, and this is the distance we use in $\mathcal{C}_c$ when $c < \infty$. When $c = \infty$ we use the topology of uniform convergence on compacta, see Whitt (1970). In $\mathcal{D}_c$ we will use what is known as the M_1 Skorokhod norm, which we discuss in detail in Appendix B, and which is defined in Section 12.3, (3.7), page 395, of Whitt (2002) when $c < \infty$ and in Section 12.9 there when $c = \infty$.

We do not subscript the norms by c. We need to consider the space $\mathcal{D}_c$ because it is within this space that the third derivatives of our H processes will lie. Since we will be letting

c go to infinity, the choice to consider a closed interval or open interval for the domain of the functions in $\mathcal{D}_c$ is somewhat arbitrary. Since we are considering derivatives, it makes the most sense to exclude the endpoints of the interval. An alternative option is to consider $(-c, c]$ or $[-c, c)$, and consider the left- or right-derivative, respectively, but we will just exclude both endpoints. Now, the derivatives we will consider are always non-decreasing, since our estimators are convex. Thus, the subset of $\mathcal{D}_c$ in which they will lie is (B.3). The important fact that we state next about $\mathcal{F}_{(-c,c),M}$ is that with the M_1 topology $\mathcal{F}_{(-c,c),M}$ is precompact, see page 237 in Appendix B for the proof.

Lemma 5.2.5. Let $M, c < \infty$. Then $\mathcal{F}_{(-c,c),M}$ is precompact in $M_{1,(-c,c)}$.

Precompactness is fundamental for tightness arguments, which we will make next about our finite sample processes. Tightness will allow us to use the uniqueness theorem, Theorem 5.1.13, to find the limiting distributions of our processes.

Lemma 5.2.6. Under Assumptions A and D, the processes $(\hat{H}^f_{n,L})'''$, $(\hat{H}^f_{n,L})''$, $(\hat{H}^f_{n,L})'$ and $\hat{H}^f_{n,L}$ are all tight in $\mathcal{D}_c \times \mathcal{C}_c^3$ when $0 < c < \infty$. Similarly, $(\hat{H}^\varphi_{n,L})'''$, $(\hat{H}^\varphi_{n,L})''$, $(\hat{H}^\varphi_{n,L})'$ and $\hat{H}^\varphi_{n,L}$ are tight in the same space. The same tightness holds if we replace the L-processes by the $R-$processes.

Proof. We will discuss the tightness for the left-side processes. The argument for the right-side processes is analogous. Under Assumptions A and D, Corollary 4.2.7, on page 106, shows that for any ϵ, we can take $M > 0$ large enough that $(\hat{H}^\varphi_{n,L})'''$ lies in $\mathcal{F}_{(-c,c),M}$ with probability $1 - \epsilon$. Since $\mathcal{F}_{(-c,c),M}$ is precompact in $\mathcal{D}_c$ by Lemma 5.2.5, $(\hat{H}^\varphi_{n,L})'''$ is tight, by the definition of tightness. Then $(\hat{H}^\varphi_{n,L})''$ is uniformly bounded by Corollary 4.2.7, and we just argued that its derivative is uniformly bounded, so, since the set of functions with their values as well as the values of their derivatives uniformly bounded by M is compact in $\mathcal{C}_c$ (via the Arzela-Ascoli theorem, see e.g. Royden (1988)), we can conclude that $(\hat{H}^\varphi_{n,L})''$

is tight in $\mathcal{F}_{(-c,c),M}$. Similarly, since integrals on bounded intervals of uniformly bounded functions are also uniformly bounded (and by (5.96) together with the fact that $m_n^{-1}(\tau_n^L) - t$ is $O_p(1)$ by Corollary 4.2.4, page 98) we know that $(\hat{H}^{\varphi}_{n,L})'$ and $\hat{H}^{\varphi}_{n,L}$ are uniformly bounded, and their respective derivatives are uniformly bounded, so we can again conclude that they are tight. An identical argument works for the right-side processes. $\square$

Now we analyze the random noise processes. We first add new notation for the derivatives of the Y-processes, to help in our analysis of the convergence of the empirical process to Brownian motion.

Definition 5.2.7 (Derivatives of the Y-processes)**.** Let

$$\mathbb{X}_n^f(v) := (\mathbb{Y}_n^f)'(v) = n^{3/5} \int_m^{m_n(v)} \left(d\mathbb{F}_n(u) - \tilde{f}_0(u)du \right),$$

$$\mathbb{X}_{n,L}^f(v) := -(\mathbb{Y}_{n,L}^f)'(v) = n^{3/5} \int_{m_n(v)}^{\tau_n^L} \left(d\mathbb{F}_n(u) - \tilde{f}_0(u)du \right),$$

$$\mathbb{X}_{n,R}^f(v) := (\mathbb{Y}_{n,R}^f)'(v) = n^{3/5} \int_{\tau_n^R}^{m_n(v)} \left(d\mathbb{F}_n(u) - \tilde{f}_0(u)du \right),$$

and let

$$\mathbb{X}_n^{\varphi}(v) := (\mathbb{Y}_n^{\varphi})'(v) = \frac{\mathbb{X}_n^f(v)}{f_0(m)} - n^{3/5} \int_m^{m_n(v)} \left(D_n(u) + R_n(u) \right) du,$$

$$\mathbb{X}_{n,L}^{\varphi}(v) := -(\mathbb{Y}_{n,L}^{\varphi})'(v) = \frac{\mathbb{X}_{n,L}^f(v)}{f_0(m)} - n^{3/5} \int_{m_n(v)}^{\tau_n^L} \left(D_n^0(u) + R_n^0(u) \right) du,$$

$$\mathbb{X}_{n,R}^{\varphi}(v) := (\mathbb{Y}_{n,R}^{\varphi})'(v) = \frac{\mathbb{X}_{n,R}^f(v)}{f_0(m)} - n^{3/5} \int_{\tau_n^R}^{m_n(v)} \left(D_n^0(u) + R_n^0(u) \right) du.$$

The second equalities in the above definitions follow directly from differentiating the Y-processes. We will first show convergence of the unconstrained processes. We let "$A_n \Rightarrow A$" mean that A_n converges weakly to A in a space which will be specified in each context (van der Vaart and Wellner, 1996).

Lemma 5.2.8. Under Assumptions A and D,

$$\left(\mathbb{X}_n^f(t), \mathbb{Y}_n^f(t)\right) \Rightarrow \left(\sqrt{f_0(m)}W(t) + \frac{f_0''(m)}{6}t^3, \sqrt{f_0(m)}\int_0^t W(v)dv + \frac{f_0''(m)}{24}t^4\right), \quad (5.98)$$

where W is a standard Brownian motion and the convergence is in $\mathcal{D}_c \times \mathcal{C}_c$ with $0 < c < \infty$.

Proof. Note that $F_0(u) = F_0(m) + f_0(m)(u-m) + \frac{f_0'(m)}{2}(u-m)^2 + \frac{f_0''(\xi)}{6}(u-m)^3$, i.e.

$$\int_m^v \tilde{f}_0(u)du = F_0(v) - F_0(m) - \frac{f_0''(\xi)}{6}(v-m)^3 \quad (5.99)$$

for some $\xi \in (m, v)$. Thus,

$$n^{3/5}\int_m^{m_n(v)}\left(d\mathbb{F}_n(u) - \tilde{f}_0(u)du\right)$$
$$= n^{3/5}\left(\mathbb{F}_n(m_n(v)) - F_0(m_n(v)) - (\mathbb{F}_n(m) - F_0(m))\right) + \frac{n^{3/5}}{6}f_0''(\xi_n)(m_n(v)-m)^3, \quad (5.100)$$

and this is equal in distribution to

$$n^{1/10}\left(\mathbb{U}_n(F_0(m_n(v))) - \mathbb{U}_n(F_0(m))\right) + \frac{f_0''(m)}{6}v^3 + o_p(1), \quad (5.101)$$

where $\mathbb{F}_n^*(t)$ is the empirical c.d.f. for n i.i.d. uniform random variables and $\mathbb{U}_n(t)$ is the corresponding empirical process, $\mathbb{U}_n(t) = \sqrt{n}(\mathbb{F}_n^*(t) - t)$. We know by Theorem 12.3.4 on page 502 of Shorack and Wellner (1986) that there exist a sequence of Brownian bridge processes B_n such that $\|\mathbb{U}_n - B_n\| = O(\log(n)n^{-1/2})$ almost surely. Thus, continuing, we have that (5.101) is equal to

$$n^{1/10}(B_n(F_0(m_n(v))) - B_n(F_0(m))) + \frac{n^{1/10}\log(n)}{\sqrt{n}}M_n(v) + \frac{f_0''(m)}{6}v^3 + o_p(1), \quad (5.102)$$

where $0 < M_n(v) \leq M < \infty$ almost surely for all $|v| \leq c$. Next we use that $B_n(t) = W_n(t) - tW_n(1)$ where $W_n(t) = B_n(t) + tN$ is a Brownian motion and N is a standard Normal random variable. (We do not assert anything about the joint distribution of the B_n so can use a single N to form our Brownian motions.) Thus (5.102) equals in distribution

$$n^{1/10}\left(W_n(F_0(m_n(v))) - W_n(F_0(m)) - (F_0(m_n(v)) - F_0(m))W_n(1)\right) + \frac{f_0''(m)}{6}v^3 + o_p(1),$$

which is equal to

$$\begin{aligned}
&W_n(v)\sqrt{n^{1/5}(F_0(m_n(v)) - F_0(m))/v} - W_n(1)n^{1/10}f_0(m)vn^{-1/5} + \frac{f_0''(m)}{6}v^3 + o_p(1) \\
&\quad = W_n(v)\sqrt{f_0(m)} + \frac{f_0''(m)}{6}v^3 + o_p(1).
\end{aligned}$$

This gives the first component of (5.98). Using this, we see next that the second component of (5.98), $n^{4/5}\int_m^{m_n(t)}\int_m^v (d\mathbb{F}_n(u) - \tilde{f}_0(u)du)$, equals[1]

$$n^{1/5}\int_m^{m_n(t)}\left(\sqrt{f_0(m)}W_n(a_n^{-1}(v)) + \frac{f_0''(m)}{6}a_n^{-1}(v)^3 + o_p(1)\right)dv,$$

which is equal to

$$\begin{aligned}
&\int_0^t\left(\sqrt{f_0(m)}W_n(v) + \frac{f_0(m)}{6}v^3 + o_p(1)\right)n^{1/5}n^{-1/5}dv \\
&\quad = \sqrt{f_0(m)}\int_0^t W_n(v)dv + \frac{f_0''(m)}{24}t^4 + o_p(1),
\end{aligned}$$

with the $o_p(1)$ error still uniform in $|t| \leq c$ (and where we take $\mathbb{F}_n = \mathbb{U}_n(F_0)$). Thus we have shown the joint equality in distribution desired for the first two terms of the vector of interest. □

[1] Note for indexing: if $\mathbb{X}_n(m_n(v)) = \mathbb{Z}_n(v) + o_p(1)$ then $\mathbb{X}_n(m_n \circ m_n^{-1}(w)) = \mathbb{Z}_n(m_n^{-1}(w)) + o_p(1)$.

Lemma 5.2.8 shows that the limit of the processes $\left(\mathbb{X}_n^f(t), \mathbb{Y}_n^f(t)\right)$ is a Gaussian term plus a drift term. The same analysis of the empirical process as in the proof of Lemma 5.2.8 applies to the corresponding constrained processes, and this allows us to conclude, letting $\mathbb{D}_n(u) := (\mathbb{F}_n - F_0)(u)$, that

$$\begin{pmatrix} n^{3/5} \int_0^{m_n(t)} d\mathbb{D}_n(u) \\ n^{4/5} \int_0^{m_n(t)} \int_0^v d\mathbb{D}_n(u) \\ n^{3/5} \int_{m_n(t)}^{\tau_n^L} d\mathbb{D}_n(u) \\ n^{4/5} \int_{m_n(t)}^{\tau_n^L} \int_v^{\tau_n^L} d\mathbb{D}_n(u) dv \\ n^{3/5} \int_{\tau_n^R}^{m_n(t)} d\mathbb{D}_n(u) \\ n^{4/5} \int_{\tau_n^R}^{m_n(t)} \int_{\tau_n^R}^v d\mathbb{D}_n(u) dv \end{pmatrix} =_d \sqrt{f_0(m)} \begin{pmatrix} W(t) \\ \int_0^t W(u)\, du \\ \left(W(m_n^{-1}(\tau_n^L)) - W(t)\right) \\ \int_t^{m_n^{-1}(\tau_n^L)} \left(W(m_n^{-1}(\tau_n^L)) - W(v)\right) dv \\ \left(W(t) - W(m_n^{-1}(\tau_n^R))\right) \\ \int_{m_n^{-1}(\tau_n^R)}^t \left(W(v) - W(m_n^{-1}(\tau_n^R))\right) dv \end{pmatrix} + \epsilon_n(t),$$

$$=: \sqrt{f_0(m)} G_n(t) + \epsilon_n(t), \tag{5.103}$$

where the Gaussian vector G_n is defined by the above equality and where $\epsilon_n(t)$ is a vector of error terms where, letting $|\cdot|$ denote Euclidean distance, $|\epsilon_n(t)| = o_p(1)$ uniformly in $t \in [-c, c]$, i.e. $\epsilon_n \Rightarrow 0$ in $(\mathcal{D}_c \times \mathcal{C}_c)^3$. In Lemma 5.2.8 we calculated the drift terms which, when added to the first two components of G_n give the limit distribution of $\left(\mathbb{X}_n^f(t), \mathbb{Y}_n^f(t)\right)$. We thus just need to calculate the drift terms for the constrained processes, i.e. corresponding to the last four terms in the above display, which we will do to prove the following lemma.

Lemma 5.2.9. Let P_n be the vector of drift terms

$$P_n(t) := \begin{pmatrix} \frac{1}{6}t^3 \\ \frac{1}{24}t^4 \\ n^{3/5}\left(\frac{1}{2}(\tau_n^L - m)^2(\tau_n^L - m_n(t)) - \frac{1}{2}(\tau_n^L - m)(\tau_n^L - m_n(t))^2 + \frac{1}{6}(\tau_n^L - m_n(t))^3\right) \\ n^{4/5}\left(\frac{1}{4}(m - \tau_n^L)^2(\tau_n^L - m_n(t))^2 + \frac{1}{6}(m - \tau_n^L)(\tau_n^L - m_n(t))^3 + \frac{1}{24}(\tau_n^L - m_n(t))^4\right) \\ n^{3/5}\left(\frac{1}{2}(\tau_n^R - m)^2(m_n(t) - \tau_n^R) + \frac{1}{2}(\tau_n^R - m)(m_n(t) - \tau_n^R)^2 + \frac{1}{6}(m_n(t) - \tau_n^R)^3\right) \\ n^{4/5}\left(\frac{1}{4}(\tau_n^R - m)^2(m_n(t) - \tau_n^R)^2 + \frac{1}{6}(\tau_n^R - m)(m_n(t) - \tau_n^R)^3 + \frac{1}{24}(m_n(t) - \tau_n^R)^4\right) \end{pmatrix},$$

and let G_n be the Gaussian vector defined in (5.103). Then under Assumptions A and D, the process $\left(\mathbb{X}_n^f, \mathbb{Y}_n^f, \mathbb{X}_{n,L}^f, \mathbb{Y}_{n,L}^f, \mathbb{X}_{n,R}^f, \mathbb{Y}_{n,R}^f\right)$ is equal in distribution in the space $(\mathcal{D}_c \times \mathcal{C}_c)^3$, with $0 < c < \infty$, to

$$\sqrt{f_0(m)}G_n + f_0''(m)P_n + \epsilon_n, \tag{5.104}$$

where $\epsilon_n(t)$ is a vector of error terms where $|\epsilon_n(t)| \to 0$ uniformly for $t \in [-c, c]$.

Proof. Lemma 5.2.8 gives the first two components of the convergence in (5.104), and (5.103) gives all the random mean-0 Gaussian limits. We thus just need to compute the drift terms for the four constrained processes. We start with $\mathbb{X}_{n,R}^f$. Now, by definition, $n^{-3/5}\mathbb{X}_{n,R}^f(v)$ is

$$\int_{\tau_n^R}^{m_n(v)} \left(d\mathbb{F}_n(u) - \tilde{f}_0(u)du\right) = \int_{\tau_n^R}^{m_n(v)} d\mathbb{D}_n(u) + \int_{\tau_n^R}^{m_n(v)} \left(dF_0(u) - \tilde{f}_0(u)du\right), \tag{5.105}$$

where we now analyze the latter term, which will contain the drift. Using the Taylor expansion

$$F_0(v) - F_0(\tau_n^R) = f_0(\tau_n^R)(v - \tau_n^R) + \frac{f_0'(\tau_n^R)}{2}(v - \tau_n^R)^2 + \frac{f_0''(\xi_{1,n})}{6}(v - \tau_n^R)^3,$$

we see that $\int_{\tau_n^R}^v \left(dF_0(u) - \tilde{f}_0(u)du\right) = \int_{\tau_n^R}^v (dF_0(u) - (f_0(m) + f_0'(m)(u-m))du)$ which equals

$$\begin{aligned}
&f_0(\tau_n^R)(v-\tau_n^R) + \frac{f_0'(\tau_n^R)}{2}(v-\tau_n^R)^2 + \frac{f_0''(\xi_{1,n})}{6}(v-\tau_n^R)^3 \\
&\quad - \left(f_0(m)(v-\tau_n^R) + \frac{f_0'(m)}{2}((v-m)^2 - (\tau_n^R - m)^2)\right) \\
&\quad = (f_0(\tau_n^R) - f_0(m))(v-\tau_n^R) + \frac{f_0'(\tau_n^R)}{2}(v-\tau_n^R)^2 \\
&\quad\quad - \frac{f_0'(m)}{2}((v-\tau_n^R)^2 + 2(v-\tau_n^R)(\tau_n^R - m)) + \frac{f_0''(\xi_{1,n})}{6}(v-\tau_n^R)^3,
\end{aligned}$$

using $(v-m)^2 = (v-\tau_n^R)^2 + 2(v-\tau_n^R)(\tau_n^R - m) + (\tau_n^R - m)^2$. We are not dropping the $f_0'(m)$ term even though it is 0, so that at the end we can notice that we actually need $f_0'(m)$ to be 0 here, otherwise this calculation does not work. Continuing, we see that the above quantity equals

$$\begin{aligned}
&\left(f_0'(m)(\tau_n^R - m) + \frac{f_0''(\xi_{2,n})}{2}(\tau_n^R - m)^2\right)(v-\tau_n^R) + \frac{1}{2}(f_0'(\tau_n^R) - f_0'(m))(v-\tau_n^R)^2 \\
&\quad - \frac{f_0'(m)}{2} 2(v-\tau_n^R)(\tau_n^R - m) + \frac{f_0''(\xi_{1,n})}{6}(v-\tau_n^R)^3,
\end{aligned}$$

which equals

$$\begin{aligned}
&\frac{1}{2} f_0''(\xi_{2,n})(\tau_n^R - m)^2(v-\tau_n^R) + \frac{1}{2}(f_0''(\xi_{3,n})(\tau_n^R - m)(v-\tau_n^R)^2) \\
&\quad - \frac{1}{2} f_0'(m)(v-\tau_n^R)(\tau_n^R - m) + \frac{1}{6} f_0''(\xi_{1,n})(v-\tau_n^R)^3,
\end{aligned}$$

which, since $f_0'(m) = 0$, equals

$$f_0''(m)\left(\frac{1}{2}(\tau_n^R - m)^2(v-\tau_n^R) + \frac{1}{2}(\tau_n^R - m)(v-\tau_n^R)^2 + \frac{1}{6}(v-\tau_n^R)^3\right) + o_p(n^{-3/5})c^3, \quad (5.106)$$

where we take $v \in [m - cn^{-1/5}, m + cn^{-1/5}]$. Note that if $f_0'(m)$ were not 0 here this expression would be $O_p(n^{-2/5})$ not $O_p(n^{-3/5})$. Note $n^{3/5}$ times (5.106) is the fifth component of $P_n(m_n^{-1}(v))$ plus an $o_p(1)$ term. Thus, by (5.103), (5.105), and (5.106), we have shown that $\mathbb{X}^f_{n,R}$ is equal in distribution to the fifth component of G_n plus the fifth component of P_n plus an $o_p(1)$ term.

Next we will compute the corresponding drift term at the Y-level, i.e. the drift term for $\mathbb{Y}^f_{n,R}$. Recall that $n^{-4/5}\mathbb{Y}^f_{n,R}$ is equal to $\int_{\tau_n^R}^{m_n(t)}\int_{\tau_n^R}^{v}\left(d\mathbb{F}_n(u)-\tilde{f}_0(u)du\right)$, which, by (5.105) is equal to

$$\int_{\tau_n^R}^{m_n(t)}\int_{\tau_n^R}^{v} d\mathbb{D}_n(u)+\int_{\tau_n^R}^{m_n(t)}\int_{\tau_n^R}^{v}\left(dF_0(u)-\tilde{f}_0(u)\right). \tag{5.107}$$

So, the drift term we need to compute, minus the integral of the $o_p(n^{-3/5})$ term in (5.106), is

$$\begin{aligned}\int_{\tau_n^R}^{m_n(t)}&\int_{\tau_n^R}^{v}(dF_0(u)-\tilde{f}_0(u)du)-o_p(n^{-4/5})c^4\\ &=f_0''(m)\int_{\tau_n^R}^{m_n(t)}\left(\frac{1}{2}(\tau_n^R-m)^2(v-\tau_n^R)+\frac{1}{2}(\tau_n^R-m)(v-\tau_n^R)^2+\frac{1}{6}(v-\tau_n^R)^3\right)dv,\end{aligned}$$

which equals

$$f_0''(m)\left(\frac{1}{2}(\tau_n^R-m)^2(m_n(t)-\tau_n^R)^2+\frac{1}{6}(\tau_n^R-m)(m_n(t)-\tau_n^R)^3+\frac{1}{24}(m_n(t)-\tau_n^R)^4\right). \tag{5.108}$$

Thus, by (5.107), (5.103), and (5.108) we have shown that $\mathbb{Y}^f_{n,R}$ is equal in distribution to the sixth component of G_n plus the sixth component of P_n plus an $o_p(1)$ term.

We now do similar calculations for the left-hand side. We will leave off the $f_0'(m)$ term here. By writing

$$F_0(v)-F_0(\tau_n^L)=f_0(\tau_n^L)(\tau_n^L-v)-\frac{1}{2}f_0'(\tau_n^L)(\tau_n^L-v)^2+\frac{1}{6}f_0''(\xi_{1,n})(\tau_n^L-v)^3,$$

we see that the left-hand side drift term is $\int_v^{\tau_n^L}\left(dF_0(u)-\tilde{f}_0(u)du\right) = F_0(\tau_n^L)-F_0(v)-$

$f_0(m)(\tau_n^L - v)$ which, since $f_0'(m) = 0$, equals

$$
\begin{aligned}
&(f_0(\tau_n^L) - f_0(m))(\tau_n^L - v) + \frac{1}{6} f_0''(\xi_{1,n})(\tau_n^L - v)^3 \\
&\quad = \left(+ \frac{1}{2} f_0''(\xi_{2,n})(\tau_n^L - m)^2 \right) (\tau_n^L - v) - \frac{1}{2} \left(+ f_0''(\xi_{3,n})(\tau_n^L - m) \right) (\tau_n^L - v)^2 \\
&\qquad + \frac{1}{6} f_0''(\xi_{1,n})(\tau_n^L - v)^3 \\
&\quad = \frac{1}{2} f_0''(\xi_{2,n})(\tau_n^L - m)^2(\tau_n^L - v) - \frac{1}{2} f_0''(\xi_{3,n})(\tau_n^L - m)(\tau_n^L - v)^2 + \frac{1}{6} f_0''(\xi_{1,n})(\tau_n^L - v)^3,
\end{aligned}
$$

which equals

$$
f_0''(m) \left(\frac{1}{2}(\tau_n^L - m)^2(\tau_n^L - v) - \frac{1}{2}(\tau_n^L - m)(\tau_n^L - v)^2 + \frac{1}{6}(\tau_n^L - v)^3 \right) + o_p(n^{-3/5})c^3 \quad (5.109)
$$

(which is identical to what we got for $\mathbb{X}_{n,R}^f$, up to replacing τ_n^R with τ_n^L). We have now shown that $\mathbb{X}_{n,L}^f$ is equal in distribution to the third component of G_n plus the third component of P_n plus an $o_p(1)$ term.

Next we compute the drift term in $\mathbb{Y}_{n,L}^f$, minus the integrated negligible error term from above. This, by (5.109), is

$$
\begin{aligned}
&\int_{m_n(t)}^{\tau_n^L} \left(\int_v^{\tau_n^L} dF_0(u) - \tilde{f}_0(u) du \right) dv - o_p(n^{-4/5})c^4 \\
&\quad = f_0''(m) \int_{\tau_n^L}^{m_n(t)} \left(\frac{1}{2}(m - \tau_n^L)^2(v - \tau_n^L) - \frac{1}{2}(m - \tau_n^L)(v - \tau_n^L)^2 + \frac{1}{6}(v - \tau_n^L)^3 \right) dv \\
&\quad = f_0''(m) \Big|_{\tau_n^L}^{m_n(t)} \left(\frac{1}{4}(m - \tau_n^L)^2(v - \tau_n^L)^2 - \frac{1}{6}(m - \tau_n^L)(v - \tau_n^L)^3 + \frac{1}{24}(v - \tau_n^L)^4 \right) dv \\
&\quad = f_0''(m) \left(\frac{1}{4}(m - \tau_n^L)^2(m_n(t) - \tau_n^L)^2 - \frac{1}{6}(m - \tau_n^L)(m_n(t) - \tau_n^L)^3 + \frac{1}{24}(m_n(t) - \tau_n^L)^4 \right) dv,
\end{aligned}
$$

and this is equal to

$$f_0''(m)\left(\frac{1}{4}(m-\tau_n^L)^2(\tau_n^L-m_n(t))^2+\frac{1}{6}(m-\tau_n^L)(\tau_n^L-m_n(t))^3+\frac{1}{24}(\tau_n^L-m_n(t))^4\right)dv \tag{5.110}$$

(which is again identical to the term for $\mathbb{Y}_{n,R}^f$ up to switching τ_n^R and τ_n^L). We have now shown that $\mathbb{Y}_{n,L}^f$ is equal in distribution to the fourth component of G_n plus the fourth component of P_n plus an $o_p(1)$ term. Thus we have shown that

$$\begin{pmatrix} \mathbb{X}_{n,L}^f(t) \\ \mathbb{Y}_{n,L}^f(t) \\ \mathbb{X}_{n,R}^f(t) \\ \mathbb{Y}_{n,R}^f(t) \end{pmatrix} =_d \sqrt{f_0(m)} \begin{pmatrix} W(m_n^{-1}(\tau_n^L) - W(t)) \\ \int_t^{m_n^{-1}(\tau_n^L)} \left(W(m_n^{-1}(\tau_n^L)) - W(v)\right) dv \\ W(t) - W(m_n^{-1}(\tau_n^R)) \\ \int_{m_n^{-1}(\tau_n^R)}^t \left(W(v) - W(m_n^{-1}(\tau_n^R))\right) dv \end{pmatrix} + f_0''(m)P_n(t) + \epsilon_n(t),$$

with (5.106), (5.108), (5.109), and (5.110) showing that P_n is as defined in the statement of the lemma. Since the first two components of (5.104) were given by Lemma 5.2.8, we are done. □

Next, we convert the above lemma at the f-level to a result at the φ-level.

Lemma 5.2.10. Under Assumptions A and D, $\left(\mathbb{X}_n^\varphi, \mathbb{Y}_n^\varphi, \mathbb{X}_{n,L}^\varphi, \mathbb{Y}_{n,L}^\varphi, \mathbb{X}_{n,R}^\varphi, \mathbb{Y}_{n,R}^\varphi\right)$ is equal in distribution in the space $(\mathcal{D}_c \times \mathcal{C}_c)^3$, with $0 < c < \infty$, to

$$\frac{1}{\sqrt{f_0(m)}} G_n + \varphi_0''(m)P_n + \epsilon_n, \tag{5.111}$$

where G_n and P_n are as in Lemma 5.2.9 and $\epsilon_n(t)$ is a vector of error terms where $|\epsilon_n(t)| \to 0$ uniformly for $t \in [-c, c]$,

Proof. Since $\varphi_0''(m) = f_0''(m)/f_0(m)$, and by (5.88) and (5.87), we can conclude that (5.111) holds. □

We still need to modify our processes so that the constants in the Y-processes line up with those in our limiting versions. We define the same constants as on page 1324 Balabdaoui et al. (2009), although we flip signs since our limiting result is stated via convexity not concavity. So, we define γ_1 and γ_2 such that

$$|\gamma_1|\gamma_2^{3/2} = \sqrt{f_0(m)} \qquad \text{and} \qquad |\gamma_1|\gamma_2^4 = \frac{4!}{|\varphi_0''(m)|}, \tag{5.112}$$

and $\gamma_1 = -|\gamma_1|$; this has the solution for γ_1 and γ_2 that we give below, along with one final re-definition of our processes.

Definition 5.2.11. Let

$$\gamma_1 = -\left(\frac{f_0(m)^{4/3}|\varphi_0''(m)|}{4!}\right)^{3/5} \qquad \text{and} \qquad \gamma_2 = \left(\frac{4!}{|\varphi_0''(m)|\sqrt{f_0(m)}}\right)^{2/5}.$$

Let

$$\begin{aligned}
&\tilde{H}_n(t) = \gamma_1 \hat{H}_n^{\varphi}(\gamma_2 t), && \tilde{H}_{n,L}(t) = \gamma_1 \hat{H}_{n,L}^{\varphi}(\gamma_2 t), \quad \tilde{H}_{n,R}(t) = \gamma_1 \hat{H}_{n,R}^{\varphi}(\gamma_2 t),\\
&\tilde{\mathbb{Y}}_n(t) = \gamma_1 \mathbb{Y}_n^{\varphi}(\gamma_2 t), && \tilde{\mathbb{Y}}_{n,L}(t) = \gamma_1 \mathbb{Y}_{n,L}^{\varphi}(\gamma_2 t), \quad \tilde{\mathbb{Y}}_{n,R}(t) = \gamma_1 \mathbb{Y}_{n,R}^{\varphi}(\gamma_2 t),\\
&\tilde{\mathbb{X}}_n(t) = \gamma_1\gamma_2 \mathbb{X}_n^{\varphi}(\gamma_2 t), && \tilde{\mathbb{X}}_{n,L}(t) = \gamma_1\gamma_2 \mathbb{X}_{n,L}^{\varphi}(\gamma_2 t), \quad \tilde{\mathbb{X}}_{n,R}(t) = \gamma_1\gamma_2 \mathbb{X}_{n,R}^{\varphi}(\gamma_2 t),
\end{aligned}$$

and we also let the new change-of-scale function be defined as

$$m_{n,\gamma_2}(t) := m_n(\gamma_2 t) = m + \frac{\gamma_2}{n^{1/5}} t, \quad \text{so that} \quad m_{n,\gamma_2}^{-1}(s) = \frac{m_n^{-1}(s)}{\gamma_2} = \frac{n^{1/5}}{\gamma_2}(t - m).$$

With these definitions, using $W(\cdot) =_d -W(\cdot)$, we see that

$$\begin{pmatrix} \tilde{\mathbb{Y}}_n(t) \\ \tilde{\mathbb{X}}_n(t) \end{pmatrix} =_d \begin{pmatrix} \int_0^t W(v)dv + t^4 \\ W(t) + 4t^3 \end{pmatrix} + \epsilon_n(t). \tag{5.113}$$

by the definitions of γ_1 and γ_2. We will conclude later that $m_{n,\gamma_2}^{-1}(\tau_n^R) = (\tau_n^R - m)n^{1/5}\gamma_2^{-1}$ converges in probability to τ_R, and that $m_{n,\gamma_2}^{-1}(\tau_n^L)$ converges in probability to τ_L, by arguing along subsequences. If we assume that to be the case, we can see that our constrained tilde-processes have the weak limits we want them to have.

Lemma 5.2.12. For any subsequence along which $m_{n,\gamma_2}^{-1}(\tau_n^R) \to_p \tau_R$ and $m_{n,\gamma_2}^{-1}(\tau_n^L) \to_p \tau_L$, we can conclude that $\left(\tilde{\mathbb{Y}}_n(t), \tilde{\mathbb{X}}_n(t), \tilde{\mathbb{Y}}_{n,L}(t), \tilde{\mathbb{X}}_{n,L}(t), \tilde{\mathbb{Y}}_{n,R}(t), \tilde{\mathbb{X}}_{n,R}(t)\right)$ converges weakly to $(Y(t), X(t), Y_L(t), X_L(t), Y_R(t), X_R(t))$ in the space $(\mathcal{C}_c \times \mathcal{D}_c)^3$. Here, $Y(t)$ and $X(t)$ are as in Theorem 5.1.12 and $Y_L(t)$, $X_L(t)$, $Y_R(t)$, and $X_R(t)$ are as in Theorem 5.1.13 (on page 129), all restricted to $t \in [-c, c]$, for $0 < c < \infty$.

Proof. The convergence of the first two components is given by (5.113). Thus, we will show the convergence for the last four (constrained) components. If $G_{n,i}$ and $P_{n,i}$ are the ith entry in the vectors G_n and P_n, respectively, then $\tilde{\mathbb{Y}}_{n,R}$ is equal in distribution to

$$\gamma_1 \left(\frac{1}{\sqrt{f_0(m)}} G_{n,6}(\gamma_2 t) + \varphi_0''(m) P_{n,6}(\gamma_2 t) + \epsilon_{n,6}(\gamma_2 t) \right). \tag{5.114}$$

The drift term is given by $\gamma_1 \varphi_0''(m) P_{n,6}(\gamma_2 t)$, which equals

$$n^{4/5}\gamma_1\varphi_0''(m) \left(\frac{1}{4}(\tau_n^R - m)^2 (m_{n,\gamma_2}(t) - \tau_n^R)^2 + \frac{1}{6}(\tau_n^R - m)(m_{n,\gamma_2}(t) - \tau_n^R)^3 + \frac{1}{24}(m_{n,\gamma_2}(t) - \tau_n^R)^4 \right),$$

and this equals

$$\begin{aligned} n^{4/5}|\gamma_1\varphi_0''(m)|\gamma_2^4 \bigg(&\frac{1}{4} m_{n,\gamma_2}^{-1}(\tau_n^R)^2 (t - m_{n,\gamma_2}^{-1}(\tau_n^R))^2 + \frac{1}{6} m_{n,\gamma_2}^{-1}(\tau_n^R)(t - m_{n,\gamma_2}^{-1}(\tau_n^R))^3 \\ &+ \frac{1}{24}(t - m_{n,\gamma_2}^{-1}(\tau_n^R))^4 \bigg), \end{aligned} \tag{5.115}$$

since

$$m_{n,\gamma_2}(t) - \tau_n^R = \left(t - n^{1/5}\gamma_2^{-1}(\tau_n^R - m)\right) n^{-1/5}\gamma_2 = \left(t - m_{n,\gamma_2}^{-1}(\tau_n^R)\right) n^{-1/5}\gamma_2.$$

Since we have assumed that $m_{n,\gamma_2}^{-1}(\tau_n^R) = (\tau_n^R - m)n^{1/5}\gamma_2^{-1}$ converges in probability to τ_R, we can conclude that (5.115) converges in probability uniformly for $t \in [-c, c]$ to

$$|\gamma_1\varphi_0''(m)|\gamma_2^4 \left(\frac{1}{4}\tau_R^2(t-\tau_R)^2 + \frac{1}{6}\tau_R(t-\tau_R)^3 + \frac{1}{24}(t-\tau_R)^4\right),$$

which is equal to

$$t^4 - 4t\tau_R^3 + 3\tau_R^4 = t^4 - 4\tau_R^3(t-\tau_R) - \tau_R^4 = \int_{\tau_R}^t \int_{\tau_R}^v 12u^2 du dv,$$

by the definitions of γ_1 and γ_2. That is, this drift term $\gamma_1\varphi_0''(m)P_{n,6}(\gamma_2 t)$ converges to the drift term for the corresponding process, Y_R, in Theorem 5.1.13. Similar calculations show that the drift term for $\tilde{\mathbb{X}}_{n,R}$, $\varphi_0''(m)\gamma_1\gamma_2 P_{n,6}(\gamma_2 t)$, converges to the drift term for X_R, $\int_{\tau_R}^t 12u^2\, du$. The calculations for the left-hand side process drift terms are identical to those for the right-hand side process drift terms. Thus we have concluded that the drift terms

$$\varphi_0''(m)\left(\gamma_1 P_{n,3}(\gamma_2 t), \gamma_1\gamma_2 P_{n,4}(\gamma_2 t), \gamma_1 P_{n,5}(\gamma_2 t), \gamma_1\gamma_2 P_{n,6}(\gamma_2 t)\right) \tag{5.116}$$

converge in probability uniformly for $t \in [-c, c]$, so converge weakly, to

$$\left(\int_t^{\tau_L}\int_v^{\tau_L} 12u^2 du dv,\ \int_t^{\tau_L} 12u^2 du,\ \int_{\tau_R}^t\int_{\tau_R}^v 12u^2 du dv,\ \int_{\tau_R}^t 12u^2 du\right). \tag{5.117}$$

Next we analyze the Gaussian terms. The Gaussian term in $\tilde{\mathbb{Y}}_{n,R}(t)$ is $\gamma_1 G_{n,6}(\gamma_2 t)/\sqrt{f_0(m)}$.

This is equal to

$$\frac{\gamma_1}{\sqrt{f_0(m)}}\int_{m_n^{-1}(\tau_n^R)}^{\gamma_2 t}\left(W(v)-W(m_n^{-1}(\tau_n^R))\right)dv,$$

which, by the change of variable $w = v/\gamma_2$, equals

$$\frac{\gamma_1}{\sqrt{f_0(m)}}\int_{m_n^{-1}(\tau_n^R)/\gamma_2}^{t}\left(W(\gamma_2 w)-W(m_n^{-1}(\tau_n^R))\right)\gamma_2 dw,$$

which, since $W(\cdot) =_d -W(\cdot)$, is equal in distribution to

$$\frac{|\gamma_1|}{\sqrt{f_0(m)}}\int_{m_{n,\gamma_2}^{-1}(\tau_n^R)}^{t}\left(\sqrt{\gamma_2}W(w)-\sqrt{\gamma_2}W(m_{n,\gamma_2}^{-1}(\tau_n^R))\right)\gamma_2 dw.$$

Since we defined γ_1 and γ_2 such that $|\gamma_1|\gamma_2^{3/2} = \sqrt{f_0(m)}$, the above expression is equal to

$$\int_{m_{n,\gamma_2}^{-1}(\tau_n^R)}^{t}\left(W(v)-W(m_{n,\gamma_2}^{-1}(\tau_n^R))\right)dv. \tag{5.118}$$

Since we assumed that $m_{n,\gamma_2}^{-1}(\tau_n^R) \to_p \tau_R$ and $m_{n,\gamma_2}^{-1}(\tau_n^L) \to_p \tau_L$, we can conclude that (5.118) converges weakly in $\mathcal{C}_c$ to $\int_{\tau_R}^t (W(v) - W(\tau_R))\,dv$. Similar calculations apply to $\tilde{\mathbb{X}}_{n,R}$ and to $\tilde{\mathbb{Y}}_{n,L}$ and $\tilde{\mathbb{X}}_{n,L}$, so that we can conclude that the Gaussian terms

$$\left(\frac{\gamma_1}{\sqrt{f_0(m)}}G_{n,3}(\gamma_2 t), \frac{\gamma_1\gamma_2}{\sqrt{f_0(m)}}G_{n,4}(\gamma_2 t), \frac{\gamma_1}{\sqrt{f_0(m)}}G_{n,5}(\gamma_2 t), \frac{\gamma_1\gamma_2}{\sqrt{f_0(m)}}G_{n,6}(\gamma_2 t)\right) \tag{5.119}$$

converge weakly in $(\mathcal{C}_c \times \mathcal{D}_c)^2$ to

$$\left(\int_t^{\tau_L}(W(\tau_L)-W(v))\,dv,\ W(\tau_L)-W(t),\ \int_{\tau_R}^{t}(W(v)-W(\tau_R))\,dv,\ W(t)-W(\tau_R)\right). \tag{5.120}$$

Since, by Lemma 5.2.10 and by definition, $\left(\tilde{\mathbb{Y}}_{n,L}(t), \tilde{\mathbb{X}}_{n,L}(t), \tilde{\mathbb{Y}}_{n,R}(t), \tilde{\mathbb{X}}_{n,R}(t)\right)$ is equal in distribution to (5.119) plus (5.116) plus an error term given by $\epsilon_n(t)$, we can conclude that

$\left(\tilde{\mathbb{Y}}_{n,L}(t), \tilde{\mathbb{X}}_{n,L}(t), \tilde{\mathbb{Y}}_{n,R}(t), \tilde{\mathbb{X}}_{n,R}(t)\right)$ converges weakly in $(\mathcal{C}_c \times \mathcal{D}_c)^2$ to (5.120) plus (5.117), as desired. We already found the limit of $\tilde{\mathbb{X}}_n$ and $\tilde{\mathbb{Y}}_n$. Thus we are done with the lemma. □

For our main theorem, we will let

$$E_c = (\mathcal{D}_c \times \mathcal{C}_c^3)^3 \times (\mathcal{D}_c \times \mathcal{C}_c)^3 \times \mathbb{R}^4,$$

for $0 < c \le \infty$. On the component spaces use the uniform and $M_{1,(-c,c)}$ topologies, as discussed in Definition 5.2.4 and in Appendix B. We take the product topology on E_c. The main theorem of this section is the following.

Theorem 5.2.13. *Define $Z_n \in E_\infty$ by*

$$\{Z_n\} := \Big\{(\tilde{H}'''_{n,L}, \tilde{H}''_{n,L}, \tilde{H}'_{n,L}, \tilde{H}_{n,L}, \tilde{H}'''_{n,R}, \tilde{H}''_{n,R}, \tilde{H}'_{n,R}, \tilde{H}_{n,R}, \tilde{H}'''_{n}, \tilde{H}''_{n}, \tilde{H}'_{n}, \tilde{H}_{n},$$
$$\tilde{\mathbb{Y}}'_{n,L}, \tilde{\mathbb{Y}}_{n,L}, \tilde{\mathbb{Y}}'_{n,R}, \tilde{\mathbb{Y}}_{n,R}, \tilde{\mathbb{Y}}'_{n}, \tilde{\mathbb{Y}}_{n}, m^{-1}_{n,\gamma_2}(\tau^L_n), m^{-1}_{n,\gamma_2}(\tau^R_n), m^{-1}_{n,\gamma_2}(\tau^0_{n,-1}), m^{-1}_{n,\gamma_2}(\tau^0_{n,1})\Big\},$$

where the component processes are defined in Definition 5.2.20 on page 193. Then we can say that a process, Z_0, that satisfies the hypotheses of Theorem 5.1.13 and Theorem 2.1 of Groeneboom et al. (2001a) exists in E_∞. That is, its twelvth and eighteenth coordinates are the H and Y processes from Theorem 2.1 of Groeneboom et al. (2001a). The ninth through eleventh coordinates are the third through first derivatives of the H process, the seventeenth coordinate is the derivative of the Y process. The first and fifth coordinates are equal, and the first through third coordinates are the third through first derivatives of the fourth coordinate. Similarly for the fifth through eighth coordinates. The second and fifth coordinates are piecewise linear. The nineteenth and twenty-first coordinate are the first knot of the second coordinate process, where the latter is possibly 0 and the former is strictly less than 0. Analogous definitions hold for the twentieth and twenty-second coordinate on *the right-side of* 0*. To state that* Z_0 *satisfies the hypotheses of the uniqueness theorem, we define, for* $z = (z_1, \ldots, z_{22})$,

$$\phi_1(z) = \big(z_7 - z_{15}\big)(z_{19}, z_{20}],$$

$$\phi_2(z) = \inf_{t\in(-\infty,0]} (z_4(t) - z_{14}(t)) \wedge 0 \quad \textit{and} \quad \phi_3(z) = \inf_{t\in[0,\infty)} (z_8(t) - z_{16}(t)) \wedge 0,$$

$$\phi_4(z) = \int_{[z_{22},\infty)} (z_8(t) - z_{16}(t)) dz_5(t) \quad \textit{and} \quad \phi_5(z) = \int_{(\infty,z_{21})} (z_4(t) - z_{14}(t)) dz_1(t).$$

Then we have that

$$\phi_i(Z_0) = 0 \textit{ for } i = 1, 2, 3, 4, 5,$$

and, under Assumptions A and D, we have that

$$Z_n \Rightarrow Z_0$$

in E_∞*.*

As a reminder about the last four coordinates of Z_n: recall that $\tau^0_{n,-1}$ and $\tau^0_{n,1}$ are the first knots of $(\hat{H}^f_{n,R})''$ which are less than or equal to m or greater than or equal to m, respectively. They differ in definition from τ^L_n and τ^R_n because the latter are not allowed to be equal to m. Note that a functional that reads off a given knot of $(\hat{H}^f_{n,R})''$, such as $\tau^0_{n,-1}$, is not continuous with respect to the uniform norm. If H''_R is piecewise linear, has a knot at τ^0_1 with $0 < \tau^0_1$ and no knots on $(0, \tau^0_1)$, then it can be uniformly approximated by piecewise linear convex functions with knots arbitrarily close to 0. Of course, any uniform approximation of H''_R has knots approaching τ^0_1. That is to say, the map $H''_R \mapsto \tau^0_1$ is upper semicontinuous but not continuous at such H''_R. This is the reason for the last four coordinates in our space E_c: even if we can conclude that, e.g., $\tilde{H}''_{n,R} \Rightarrow H''_R$, we cannot directly apply the continuous mapping theorem to say that the knots of $\tilde{H}''_{n,R}$ converge in probability to the knots of H''_R.

Proof. We begin by noting that Z_n, when its first 18 components are restricted to $[-c, c]$, is tight in E_c for any $0 < c < \infty$. The first 18 components of Z_n are defined on $(-\infty, \infty)$, but we will also be considering their restrictions to compact domains, so we will let $R_{m_1} : E_{m_2} \to E_{m_1}$ be such that for a function g on $[-m_2, m_2]$ or $(-m_2, m_2)$, $R_{m_1}(g)$ on $[-m_1, m_1]$ is defined by restricting g to $[-m_1, m_1]$ or $(-m_1, m_1)$, respectively. We slightly abuse notation by only subscripting by m_1 and not m_2, since R_{m_1} applied to g defined on $[-m_2, m_2]$, for any $m_2 > m_1$, yields the same output, i.e. $R_{m_1} = R_{m_1} \circ R_{m_2}$, for any $m_2 > m_1$. We also abuse notation by defining R_{m_1} on a vector to be applied componentwise, unless that vector has real-valued components, which are then ignored; i.e. for $z \in E_{m_2}$ by $R_{m_1}(z) := (R_{m_1}(z_1), \ldots, R_{m_1}(z_{18}), z_{19}, \ldots, z_{22})$. Now, under Assumptions A and D, Lemma 5.2.6 yields that both the left- and right- constrained H-processes and their first, second, and third derivatives are tight in $\mathcal{D}_c \times \mathcal{C}_c^3$, and from Lemma 5.2.12 we know that $R_c(\tilde{\mathbb{Y}}'_{n,L}, \tilde{\mathbb{Y}}_{n,L}, \tilde{\mathbb{Y}}'_{n,R}, \tilde{\mathbb{Y}}_{n,R}, \tilde{\mathbb{Y}}'_n, \tilde{\mathbb{Y}}_n,)$ is tight in $\mathcal{C}_c^6$. Similarly, from Balabdaoui et al. (2009) we know that $R_c(\tilde{H}'''_n, \tilde{H}''_n, \tilde{H}'_n, \tilde{H}_n)$ is tight in $\mathcal{D}_c \times \mathcal{C}_c^3$. Thus, by the definition of tightness, $\{R(Z_n)\}$ is tight in E_c since its components are tight in the component spaces of the product space, E_c, for $0 < c < \infty$. Now for a sequence $R_c(Z_n)$ in a metric space such as E_c, if for all subsequences $R_c(Z_{n_i})$, $i = 1, \ldots, \infty$, of $R_c(Z_n)$, there is a subsubsequence n_{i_j}, $j = 1, \ldots, \infty$, such that $R_c(Z_{n_{i_j}})$ converges weakly to the same limit, then that same limit is the weak limit of the original sequence $R_c(Z_n)$. This is the fact we will eventually use to conclude that Z_n converges weakly to Z_0 in E_∞.

On the other hand, if a sequence of elements Z_n is tight, then along any subsequence n_i, $i = 1, \ldots, \infty$ there is a subsubsequence n_{i_j}, $j = 1, \ldots, \infty$ such that $Z_{n_{i_j}}$ has a weak limit. So far, we have only argued that our sequences Z_n are tight when restricted to compact domains, i.e. $R_m(Z_n)$ is tight in E_m. However, having restrictions which are tight in E_m for all $0 < m < \infty$ turns out to be turns out to be equivalent to being tight in E_∞. For C_m and C_∞, this is Corollary 5 of Whitt (1970). For D_m and D_∞ with the J_1 topology this is the Corollary on page 120 of Lindvall (1973). For other topologies on D_m and D_∞, such as the M_1 topology that we consider, the proofs should also go through (since there are results of similar spirit for the J_i and M_i, $i = 1, 2$, toplogies, e.g. see Theorem 12.9.3 in Whitt (2002)), but we will just prove what we need here since it is straightforward. That is, we will prove that along any subsequence of Z_n there is a subsubsequence which has a limit in E_∞. (Later we will show that limit is always the same.) Because Z_n is tight in E_m for *each* $m \in \mathbb{N}$, we can use a Cantor diagonalization argument to show that a limit Z_0 exists in E_∞. Thus, we start with an arbitrary subsequence n_i. For $m = 1$, there is a subsubsequence n_{i_j} such that $R_1(Z_{n_{i_j}})$ converges to a weak limit $Z_{1,0} \in E_1$.[2] Similarly, for $m = 2$, there is a subsequence $n_{i_{j_k}}$ of n_{i_j} such that $R_2(Z_{n_{i_{j_k}}})$ has a weak limit $Z_{2,0} \in E_2$, and, by continuing, we can see that there is a weak limit $Z_{m,0} \in E_m$ for all $m \in \mathbb{N}$. Now, we take a diagonal sequence, d_i, defined by $d_1 = n_1$, $d_2 = n_{i_2}$, $d_3 = n_{i_{j_3}}$, and so on, which has the property that it is contained in the original subsequence, and in fact after its mth element it is contained in the mth subsequence. Thus, $R_m(Z_{d_i})$ converges weakly to $Z_{m,0}$ as $i \to \infty$ in the space E_m for all $m \in \mathbb{N}$. Furthermore, for $m_1 < m_2$, $R_{m_1}(Z_{m_2,0})$ is equal in distribution to $Z_{m_1,0}$. Then for a bounded continuous function h_{m_1} on E_{m_1}, we can construct a bounded continuous function $\tilde{h}_{m_2} := h_{m_1} \circ R_{m_1}$ on E_{m_2}. We know $\tilde{h}_{m_2}$ is continuous because by construction, distance in E_{m_2} is no smaller than distance in E_{m_1}, so if $z_i \to z$ in E_{m_2} then certainly $R_{m_1}(z_i)$ converges to $R_{m_1}(z)$. Thus by the continuous mapping theorem or the definition of weak convergence, since $R_{m_2}(Z_n) \Rightarrow Z_{m_2,}$, $E(\tilde{h}_{m_2}(R_{m_2}(Z_n))) \to E(\tilde{h}_{m_2}(Z_{m_2,0}))$ and since $R_{m_1}(Z_n) \Rightarrow Z_{m_1,0}$, $E(h_{m_1}(R_{m_1}(Z_n))) \to E(h_{m_1}(Z_{m_1,0}))$. Since $h_{m_1} \circ R_{m_1} = \tilde{h}_{m_2} \circ R_{m_2}$, it is trivial that the two limiting means are the same. Thus for all continuous bounded functions

[2] A note on notation: $Z_{m,0}$ (for any $m > 0$) is an element of E_m, whereas $Z_{0,i}$ (for $i = 1, \ldots, 22$) is the ith component of Z_0.

h_{m_1} on E_{m_1}, $E(h_{m_1}(R_{m_1}(Z_{m_2,0}))) = E(h_{m_1}(Z_{m_1,0}))$, which implies that the restriction of $Z_{m_2,0}$ to $[-m_1, m_1]$, $R_{m_1}(Z_{m_2,0})$, has the same distribution as $Z_{m_1,0}$ (see, e.g., chapter 1 of Billingsley (1999)). Thus, by the Kolmogorov Extension Theorem (see, e.g., page 11 of Øksendal (2003)) we can define Z_0 on $\mathbb{R}$ by giving it the (consistent) marginals of $Z_{m,0}$ as $m \to \infty$.

So far, we have shown that any subsequence of Z_n has a further subsubsequence that converges weakly to Z_0 in E_∞. It remains to show, via the continuous mapping theorem, that Z_0 always satisfies the uniqueness criteria of Theorem 5.1.13 (regardless of the subsequence or subsubsequence chosen), so that the limit along all subsubsequences is the same. Then we can conclude that the original sequence Z_n converges weakly to Z_0 in E_∞ (van der Vaart and Wellner, 1996). Thus, we now argue about Z_0, the limit of our arbitrary subsubsequence. Our unconstrained processes $(\tilde{H}_n''', \tilde{H}_n'', \tilde{H}_n', \tilde{H}_n)$ are identical to the processes defined in Balabdaoui et al. (2009), so we know their limit is unique. For our constrained processes we first note that $\tilde{H}''_{n,L}(t) = \tilde{H}''_{n,R}(t)$ for all $t \in \mathbb{R}$, and that this function is convex with antimode at 0, so this necessarily holds in the limit (for any subsubsequence) as well. We assume that the limit is piecewise linear. Now we can define τ_L and τ_R as the limits of the subsubsequences of $m_{n,\gamma_2}^{-1}(\tau_n^L)$ and $m_{n,\gamma_2}^{-1}(\tau_n^R)$, and use them to simultaneously define Y_L and Y_R. Then we have shown above in Lemma 5.2.12 that the limit of the Y-processes and their derivatives are the corresponding processes from Theorem 5.1.12 and from Theorem 5.1.13 (on page 129).

We have, by (5.89),

$$\left| \left((\hat{H}_{n,R}^{\varphi})' - (\mathbb{Y}_{n,R}^{\varphi})' \right) \left(a_n^{-1}(\tau_n^L), a_n^{-1}(\tau_n^R) \right] \right| \le \frac{4}{f_0(m)} n^{-2/5} \tag{5.121}$$

(Note that $(\hat{H}_{n,R}^{\varphi})' - (\mathbb{Y}_{n,R}^{\varphi})'$ is equivalent to $(\hat{H}_{n,L}^{\varphi})' - (\mathbb{Y}_{n,L}^{\varphi})'$, but we only state the result for one of them.) Now, $((\tilde{H}_{n,R})' - (\tilde{\mathbb{Y}}_{n,R})')(t) = \gamma_1\gamma_2((\hat{H}_{n,R}^{\varphi})' - (\mathbb{Y}_{n,R}^{\varphi})')(\gamma_2 t)$, which means that

$$\left|\left((\tilde{H}_{n,R})' - (\tilde{\mathbb{Y}}_{n,R})'\right)\left(m_{n,\gamma_2}^{-1}(\tau_n^L), m_{n,\gamma_2}^{-1}(\tau_n^R)\right]\right| \le \frac{4\gamma_1\gamma_2}{f_0(m)} n^{-2/5} \tag{5.122}$$

Thus, for the continuous functional $\phi : E_c \to \mathbb{R}$ defined by

$$\phi(z_1, \ldots, z_{22}) = \big(z_7 - z_{15}\big)(z_{19}, z_{20}],$$

we have $\phi(Z_n) \to 0$ almost surely and $\phi(Z_n) \to \phi(Z_0)$ in probability along any subsubsequence where $n \to \infty$. This implies that, in fact, $\phi(Z_0) = 0$ almost surely, since $P(|\phi(Z_0)| > \epsilon) < \epsilon$ for all $\epsilon > 0$. Thus we have shown (5.33).

Now recall (5.90) and (5.91), which say that

$$\begin{aligned} \mathbb{Y}^{\varphi}_{n,L}(t) - \hat{H}^{\varphi}_{n,L}(t) &\ge 0 \quad \text{for } t \le 0, \\ \mathbb{Y}^{\varphi}_{n,R}(t) - \hat{H}^{\varphi}_{n,R}(t) &\ge 0 \quad \text{for } t \ge 0, \end{aligned}$$

with equalities for t such that $m_n(t)$ is left- or right-knot, respectively. This immediately implies, recalling $\gamma_1 < 0$, that

$$\begin{aligned} \tilde{\mathbb{Y}}_{n,L}(t) - \tilde{H}_{n,L}(t) &\le 0 \quad \text{for } t \le 0, \\ \tilde{\mathbb{Y}}_{n,R}(t) - \tilde{H}_{n,R}(t) &\le 0 \quad \text{for } t \ge 0, \end{aligned} \tag{5.123}$$

with equality if $m_{n,\gamma_2}(t)$ is a knot. That is, the functional for the right side, $\phi : E_c \to \mathbb{R}$, defined by

$$\phi(z_1, \ldots, z_{22}) = \inf_{t \in [0,c]} (z_8(t) - z_{16}(t)) \wedge 0$$

is clearly continuous so since $\phi(Z_n) = 0$ almost surely for all n, we can conclude that the limit Z_0 of the subsubsequence satisfies $\phi(Z_0) = 0$ almost surely also, with the same argument as above to go from convergence in probability to an almost sure statement in the limit. We can conclude identically for the corresponding functional for the left side, $\phi(z_1, \ldots, z_{22}) = \inf_{t \in [-c,0]}(z_4(t) - z_{14}(t)) \wedge 0$. Since these conclusions hold for all $c > 0$, we have shown (5.34) and (5.35).

Now, considering the right side, we have by the Lebesgue-Stieltjes integral change of variables $w = u/\gamma_2$, that

$$\int_{m_n^{-1}(\tau_{n,1}^0)}^{\infty} (\hat{H}_{n,R}^{\varphi} - \mathbb{Y}_{n,R}^{\varphi})(u) d(\hat{H}_{n,R}^{\varphi})'''(u) = \int_{m_{n,\gamma_2}^{-1}(\tau_{n,1}^0)}^{\infty} (\hat{H}_{n,R}^{\varphi} - \mathbb{Y}_{n,R}^{\varphi})(\gamma_2 w)(\gamma_2 d(\hat{H}_{n,R}^{\varphi})''')(\gamma_2 w)$$

where, recall, $\tau_{n,1}^0$ is the smallest knot of $(\hat{H}_{n,R}^f)''$ greater than or equal to m. Using the equality for t such that $m_{n,\gamma_2}(t)$ is a knot in (5.123), this shows that

$$\int_{m_{n,\gamma_2}^{-1}(\tau_{n,1}^0)}^{\infty} (\tilde{H}_{n,R} - \tilde{\mathbb{Y}}_{n,R})(u) d(\tilde{H}_{n,R})'''(u) = 0, \tag{5.124}$$

since the above two displays are only different by a constant factor of $\gamma_1 \gamma_2^3$. Now, similarly to Groeneboom et al. (2001b) we need that $\phi_R(Z_0) = 0$ for the functional

$$\phi_R(z) := \int_{[\tau_1^0(z), c]} (z_8(t) - z_{16}(t))\, dz_5(t),$$

where $\tau_1^0(z)$ is the infimum of the knots of z_6 (or the points of jump of z_5) in $[0, \infty)$. Note that $\tau_1^0(z)$ is not necessarily equal to z_{22}, but, as argued previously (on page 178), we know that $z_{22} \leq \tau_1^0(z)$. Now, by Lemma 5.2.14 below and by the continuous mapping theorem, since $\phi_{R,c}(Z_n) = 0$ for all n and $c > 0$ (by (5.124)), we can conclude that $\phi_{R,c}(Z_0) = 0$ for all $c > 0$, where $\phi_{R,c}(z) = \int_{[z_{22}, \infty)} d_z(t)\, d(1_{[z_{22}, c]}(t) z_5(t))$. This allows us to conclude that $\phi_R(Z_0) = 0$, as follows. First, we let $c \to \infty$, to see that

$$\int_{-\infty}^{\infty} (Z_{0,8}(t) - Z_{0,16}(t))\, d(1_{[Z_{0,22}, \infty)}(t) Z_{0,5})(t) = 0. \tag{5.125}$$

Next since $Z_{0,22} \leq \tau_1^0(Z_0)$ (and since we just showed $Z_{0,8}(t) \geq Z_{0,16}(t)$), we can see that $\int_{[\tau_1^0(Z_0),\infty)}(Z_{0,8}(t) - Z_{0,16}(t))\, d(1_{[Z_{0,22},\infty)}(t)Z_{0,5}(t)) = 0$. Now, if $Z_{0,8}(\tau_1^0(Z_0)) - Z_{0,16}(\tau_1^0(Z_0))$ were strictly positive, then since $Z_{0,5}$ is non-decreasing and (by the definition of $\tau_1^0(Z_0)$) has positive mass at $\tau_1^0(Z_0)$, since we know $Z_{0,8}(t) \geq Z_{0,16}(t)$ for all $t > 0$, and since $Z_{0,22} \leq \tau_1^0(Z_0)$, we can see that the integral $\phi_{R,c}(Z_0) = \int_{[Z_{0,22},\infty)}(Z_{0,8}(t) - Z_{0,16}(t))\, dZ_{0,5}(t)$ would also be strictly positive. Thus, since this integral is 0, we conclude that $Z_{0,8}(\tau_1^0(Z_0)) - Z_{0,16}(\tau_1^0(Z_0)) = 0$. Finally: if $\tau_1^0(Z_0) \neq Z_{0,22}$, then we can immediately drop the indicator in the measure in (5.125) to see that $\int_{[\tau_1^0(Z_0),\infty)}(Z_{0,8}(t) - Z_{0,16}(t))\, dZ_{0,5}(t) = 0$; if $\tau_1^0(Z_0) = Z_{0,22}$ then, because $Z_{0,8}(\tau_1^0(Z_0)) - Z_{0,16}(\tau_1^0(Z_0)) = 0$, we also have $\int_{[\tau_1^0(Z_0),\infty)}(Z_{0,8}(t) - Z_{0,16}(t))\, dZ_{0,5}(t) = 0$. This gives the right side of (5.36). The left side is identical.

We have now shown all of the conditions of the uniqueness theorem and so we can conclude that the limiting element, Z_0, is unique, i.e. that it is the same regardless of our subsubsequence, and so we can conclude that

$$Z_n \Rightarrow Z_0.$$

□

Lemma 5.2.14. Let z_n be a sequence converging to z in E_∞ such that z_5 and z_1 are non-decreasing, piecewise constant, and have separated points of jump. Define $\phi_{R,c}$ and $\phi_{L,c}$ by

$$\phi_{R,c}(z) = \int_{-\infty}^{\infty} (z_8(t) - z_{16}(t))\, d(1_{[z_{22},c]}(t)z_5(t)),$$
$$\phi_{L,c}(z) = \int_{-\infty}^{\infty} (z_4(t) - z_{14}(t))\, d(1_{[-c,z_{21}]}(t)z_1(t)).$$

Then for all $c > 0$, $\phi_{R,c}(z_n) \to \phi_{R,c}(z)$ and $\phi_{L,c}(z_n) \to \phi_{L,c}(z)$ as $n \to \infty$.

Proof. For convenience, we let $d_z(t) := z_8(t) - z_{16}(t)$ in this proof, and we let $z_n = (z_{n,1}, \ldots, z_{n,22})$ and $z = (z_1, \ldots, z_{22})$ satisfy the hypotheses of the lemma statement. Now since $z_{n,5}$ converges to z_5 in the $M_{1,(-\infty,\infty)}$ topology, $z_{n,5}$ converges weakly to z_5 in the sense that for all $t \in (-\infty, \infty)$ that are continuity points of z_5, $z_{n,5}(t) \to z_5(t)$ as $n \to \infty$. This follows from $M_{1,(-\infty,\infty)}$ convergence by Lemma 12.5.1 on page 405 of Whitt (2002). Next, we define $\phi_{R,c}(z) = \int_{(-\infty,\infty)} d_z(t)\, d(1_{[z_{22},c]}(t) z_5(t))$. Then we see that $\phi_{R,c}(z_n) - \phi_{R,c}(z)$ equals $\int_{-\infty}^{\infty} d_{z_n}(t)\, d(1_{[(z_{n,22}),c]}(t) z_{n,5}(t)) - \int_{-\infty}^{\infty} d_z(t)\, d(1_{[z_{22},c]}(t) z_5(t))$, which equals

$$\int_{-\infty}^{\infty} (d_{z_n}(t) - d_z(t))\, d(1_{[z_{n,22},c]}(t) z_{n,5}(t)) - \int_{-\infty}^{\infty} d_z(t)\, d(1_{[z_{22},c]}(t) z_5(t) - 1_{[(z_{n,22}),c]}(t) z_{n,5}(t)).$$

Since $1_{[(z_{n,22}),c]}(t) z_{n,5}(t))$ converges weakly (in the sense mentioned just above) to $1_{[z_{22},c]}(t) z_5(t)$ and $d_z(t)$ is a bounded continuous function on $[z_{22}, c]$, we can conclude that the second term in the above display converges to 0 as $n \to \infty$. Then since $(d_{z_n}(t) - d_z(t))$ converges uniformly to 0 as $n \to \infty$ and $1_{[z_{n,22},c]}(t) z_{n,5}(t)$ has total mass bounded by $z_5(c+\epsilon) + \epsilon - (z_5(-\epsilon) - \epsilon)$ (since convergence in $M_{1,(-\infty,\infty)}$ implies convergence at points of continuity of the limit, z_5, which is piecewise constant with separated knots, so that there is a point of continuity within ϵ of c and of 0), we can conclude that the first term also converges to 0 as $n \to \infty$ (see, e.g., page 275 of Royden (1988)), so we are done with the statement for $\phi_{R,c}$. For $\phi_{L,c}$, the proof is identical. □

We have now shown the main result we wanted. We will end this section by unraveling definitions to relate the processes that will appear in our study of the likelihood ratio statistic to the processes for which we just showed convergence. We also will show a corollary giving our estimator limit distributions at the mode m.

Corollary 5.2.15. *Under Assumptions A and D, for $t \in \mathbb{R}$,*

$$n^{2/5}(\hat{\varphi}_n(m_n(t)) - \tilde{\varphi}_0(m_n(t))) = \frac{1}{\gamma_1\gamma_2^2}\tilde{H}_n''\left(\frac{t}{\gamma_2}\right), \tag{5.126}$$
$$n^{2/5}(\hat{\varphi}_n^0(m_n(t)) - \tilde{\varphi}_0(m_n(t))) = \frac{1}{\gamma_1\gamma_2^2}\tilde{H}_{n,R}''\left(\frac{t}{\gamma_2}\right), \tag{5.127}$$

and

$$n^{3/5}\int_m^{m_n(v)}(d\mathbb{F}_n(u) - \tilde{f}_0(u)du) = \mathbb{X}_n^f(v) = (\mathbb{Y}_n^f)'(v)$$
$$n^{3/5}\int_{\tau_n^R}^{m_n(v)}(d\mathbb{F}_n(u) - \tilde{f}_0(u)du) = \mathbb{X}_{n,R}^f(v) = (\mathbb{Y}_{n,R}^f)'(v) = \frac{f_0(m)}{\gamma_1\gamma_2}\tilde{\mathbb{X}}_{n,R}\left(\frac{v}{\gamma_2}\right) + o_p(1),$$
$$n^{3/5}\int_{m_n(v)}^{\tau_n^L}(d\mathbb{F}_n(u) - \tilde{f}_0(u)du) = \mathbb{X}_{n,L}^f(v) = -(\mathbb{Y}_{n,L}^f)'(v) = \frac{f_0(m)}{\gamma_1\gamma_2}\tilde{\mathbb{X}}_{n,L}\left(\frac{v}{\gamma_2}\right) + o_p(1).$$

Proof. All the statements effectively follow from unraveling our definitions. For (5.126), we have

$$n^{4/5}\int_m^{m_n(t)}\int_m^v(\hat{\varphi}_n(u) - \tilde{\varphi}_0(u))dudv = \hat{H}_n^{\varphi}(t) - \frac{A_n t + B_n}{f_0(m)} = \frac{1}{\gamma_1}\tilde{H}_n(\frac{t}{\gamma_2}) - \frac{A_n t + B_n}{f_0(m)}$$

so that we differentiate twice to see

$$n^{3/5}\int_m^{m_n(t)}(\hat{\varphi}_n(u) - \tilde{\varphi}_0(u))du = \frac{1}{\gamma_1\gamma_2}\tilde{H}_n'(\frac{t}{\gamma_2}) - \frac{A_n}{f_0(m)},$$
$$n^{2/5}(\hat{\varphi}_n(m_n(t)) - \tilde{\varphi}_0(m_n(t))) = \frac{1}{\gamma_1\gamma_2^2}\tilde{H}_n''(\frac{t}{\gamma_2}).$$

Similarly, for (5.127),

$$n^{4/5}\int_m^{m_n(t)}\int_m^v(\hat{\varphi}_n^0(u) - \tilde{\varphi}_0(u))dudv = \hat{H}_{n,R}^{\varphi}(t) - \frac{A_n^R(t - m_n^{-1}(\tau_n^R)) + B_n^R}{f_0(m)}$$
$$= \frac{1}{\gamma_1}\tilde{H}_{n,R}(\frac{t}{\gamma_2}) - \frac{A_n^R(t - m_n^{-1}(\tau_n^R)) + B_n^R}{f_0(m)}$$

so that we differentiate twice to see

$$n^{3/5}\int_m^{m_n(t)}(\hat\varphi_n^0(u)-\tilde\varphi_0(u))du = \frac{1}{\gamma_1\gamma_2}\tilde H'_{n,R}(\frac{t}{\gamma_2}) - \frac{A_n}{f_0(m)},$$
$$n^{2/5}(\hat\varphi_n^0(m_n(t))-\tilde\varphi_0(m_n(t))) = \frac{1}{\gamma_1\gamma_2^2}\tilde H''_{n,R}(\frac{t}{\gamma_2}).$$

We will not record the identical calculation for the left-side. Now the latter three identities in this Corollary follow from the equalities

$$n^{3/5}\int_m^{m_n(v)}(d\mathbb{F}_n(u)-\tilde f_0(u)du) = f_0(m)\left(\int_m^{m_n(v)}(D_n(u)+R_n(u))du + \mathbb{X}_n^{\varphi}(v)\right),$$
$$n^{3/5}\int_{\tau_n^R}^{m_n(v)}(d\mathbb{F}_n(u)-\tilde f_0(u)du) = f_0(m)\left(\int_{\tau_n^R}^{m_n(v)}(D_n^0(u)+R_n^0(u))du + \mathbb{X}_{n,R}^{\varphi}(v)\right),$$
$$n^{3/5}\int_{m_n(v)}^{\tau_n^L}(d\mathbb{F}_n(u)-\tilde f_0(u)du) = f_0(m)\left(\int_{m_n(v)}^{\tau_n^L}(D_n^0(u)+R_n^0(u))du + \mathbb{X}_{n,L}^{\varphi}(v)\right),$$

which, up to $o_p(1)$ error (and by recalling (5.88) and (5.87)) equal $f_0(m)\tilde{\mathbb{X}}_n\left(v/\gamma_2\right)/(\gamma_1\gamma_2)$, $f_0(m)\tilde{\mathbb{X}}_{n,R}\left(v/\gamma_2\right)/(\gamma_1\gamma_2)$, and $f_0(m)\tilde{\mathbb{X}}_{n,L}\left(v/\gamma_2\right)/(\gamma_1\gamma_2)$, respectively. □

Our last result gives the joint distributions of our estimator values at the fixed point m. For the convergence of the functions at the exponential level, we could unravel the f-processes that we used above, but we will instead just consider the fixed point m and use the delta method on the φ-processes (or, rather, the tilde-processes).

Corollary 5.2.16. *Under Assumptions A and D, with H and H_R defined as in Theorem 2.1* *of Groeneboom et al. (2001a) and our Theorem 5.1.13, respectively, we can conclude that*

$$n^{2/5}\begin{pmatrix}\hat{\varphi}_n(m)-\varphi_0(m)\\ \hat{\varphi}_n^0(m)-\varphi_0(m)\\ \hat{f}_n(m)-f_0(m)\\ \hat{f}_n^0(m)-f_0(m)\end{pmatrix} \to_d \frac{1}{\gamma_1\gamma_2^2}\begin{pmatrix}H''(0)\\ H_R''(0)\\ f_0(m)H''(0)\\ f_0(m)H_R''(0)\end{pmatrix},$$

where the constants are given by

$$\frac{1}{\gamma_1\gamma_2^2} = -\left(\frac{|\varphi_0''(m)|}{f_0(m)^2 4!}\right)^{1/5} \quad \text{and} \quad \frac{f_0(m)}{\gamma_1\gamma_2^2} = -\left(\frac{|\varphi_0''(m)|f_0(m)^3}{4!}\right)^{1/5}.$$

Proof. When φ_n is either $\hat{\varphi}_n$ or $\hat{\varphi}_n^0$, we can write

$$n^{2/5}(e^{\varphi_n(m)}-e^{\varphi_0(m)}) = \frac{e^{\varphi_n(m)}-e^{\varphi_0(m)}}{\varphi_n(m)-\varphi_0(m)} n^{2/5}(\varphi_n(m)-\varphi_0(m)).$$

Then, in either case, using the definition of the derivative, as n gets large $(e^{\varphi_n(m)} - e^{\varphi_0(m)})/(\varphi_n(m)-\varphi_0(m))$ converges in probability to $f_0(m)$. Then the result follows by Theorem 5.1.13 and Corollary 5.2.15. □

5.2.1 Summary Section

This section contains previously given definitions, identities, and basic lemmas, all in one place, to serve as a reference during reading of Section 5.2.

Definition 5.2.17. (Localized-at-the-Mode Processes): For $t \in \mathbb{R}$, define

$$m_n(t) := m + tn^{-1/5}, \quad \text{so that} \quad m_n^{-1}(s) = n^{1/5}(s - m),$$

and define τ_n^L to be the first knot strictly less than m and τ_n^R to be the first knot strictly greater than m. We will define

$$\tilde{f}_0(u) = f_0(m) + f_0'(m)(u - m) = f_0(m),$$
$$\tilde{\varphi}_0(u) = \varphi_0(m) + \varphi_0'(m)(u - m) = \varphi_0(m),$$

where the notation $\tilde{f}_0(u)$ is used in place of $f_0(m)$ simply to remind ourselves that we are approximating $f_0(u)$ with a one term Taylor expansion where the derivative happens to be 0. Next, we define, at the f-level,

$$\mathbb{Y}_n^f(t) := n^{4/5} \int_m^{m_n(t)} \left(\int_m^v d\mathbb{F}_n(u) - \int_m^v \tilde{f}_0(u) du \right) dv$$
$$\hat{H}_n^f(t) := n^{4/5} \int_m^{m_n(t)} \left(\int_m^v \hat{f}_n(u) du - \int_m^v \tilde{f}_0(u) du \right) dv + A_n t + B_n,$$

and

$$\mathbb{Y}^f_{n,L}(t) := n^{4/5} \int_{m_n(t)}^{\tau_n^L} \left(\int_v^{\tau_n^L} d\mathbb{F}_n(u) - \int_v^{\tau_n^L} \tilde{f}_0(u) du \right) dv$$

$$\mathbb{Y}^f_{n,R}(t) := n^{4/5} \int_{\tau_n^R}^{m_n(t)} \left(\int_{\tau_n^R}^{v} d\mathbb{F}_n(u) - \int_{\tau_n^R}^{v} \tilde{f}_0(u) du \right) dv$$

$$\hat{H}^f_{n,L}(t) := n^{4/5} \int_{m_n(t)}^{\tau_n^L} \left(\int_v^{\tau_n^L} \hat{f}_n^0(u) du - \int_v^{\tau_n^L} \tilde{f}_0(u) du \right) dv + A_n^L(m_n^{-1}(\tau_n^L) - t) + B_n^L$$

$$\hat{H}^f_{n,R}(t) := n^{4/5} \int_{\tau_n^R}^{m_n(t)} \left(\int_{\tau_n^R}^{v} \hat{f}_n^0(u) du - \int_{\tau_n^R}^{v} \tilde{f}_0(u) du \right) dv + A_n^R(t - m_n^{-1}(\tau_n^R)) + B_n^R$$

and we define, at the φ-level,

$$\mathbb{Y}^\varphi_n(t) := \frac{\mathbb{Y}^f_n(t)}{f_0(m)} - n^{4/5} \int_m^{m_n(t)} \int_m^v (D(u) + R(u))\, du dv$$

$$\hat{H}^\varphi_n(t) := n^{4/5} \int_m^{m_n(t)} \int_m^v (\hat{\varphi}_n(u) - \tilde{\varphi}_0(u))\, du dv + \frac{A_n t + B_n}{f_0(m)},$$

and

$$\mathbb{Y}^\varphi_{n,L}(t) := \frac{\mathbb{Y}^f_{n,L}(t)}{f_0(m)} - n^{4/5} \int_{m_n(t)}^{\tau_n^L} \int_v^{\tau_n^L} \left(D_n^0(u) + R_n^0(u)\right) du dv$$

$$\mathbb{Y}^\varphi_{n,R}(t) := \frac{\mathbb{Y}^f_{n,R}(t)}{f_0(m)} - n^{4/5} \int_{\tau_n^R}^{m_n(t)} \int_{\tau_n^R}^{v} \left(D_n^0(u) + R_n^0(u)\right) du dv$$

$$\hat{H}^\varphi_{n,L}(t) := n^{4/5} \int_{m_n(t)}^{\tau_n^L} \int_v^{\tau_n^L} \left(\hat{\varphi}_n^0(u) - \tilde{\varphi}_0(u)\right) du dv + \frac{A_n^L(m_n^{-1}(\tau_n^L) - t) + B_n^L}{f_0(m)}$$

$$\hat{H}^\varphi_{n,R}(t) := n^{4/5} \int_{\tau_n^R}^{m_n(t)} \int_{\tau_n^R}^{v} \left(\hat{\varphi}_n^0(u) - \tilde{\varphi}_0(u)\right) du dv + \frac{A_n^R(t - m_n^{-1}(\tau_n^R)) + B_n^R}{f_0(m)},$$

where

$$A_n = n^{3/5}\left(\hat{F}_n(m) - \mathbb{F}_n(m)\right) \quad \text{and} \quad B_n = n^{4/5}\left(\hat{H}_n(m) - \mathbb{Y}_n(m)\right),$$
$$A_n^L = n^{3/5}\left(\mathbb{F}_{n,L}(\tau_n^L) - \hat{F}_{n,L}^0(\tau_n^L)\right) \quad \text{and} \quad B_n^L = n^{4/5}\left(\hat{H}_{n,L}^0(\tau_n^L) - \mathbb{Y}_{n,L}(\tau_n^L)\right) = 0,$$
$$A_n^R = n^{3/5}\left(\mathbb{F}_{n,R}(\tau_n^R) - \hat{F}_{n,R}^0(\tau_n^R)\right) \quad \text{and} \quad B_n^R = n^{4/5}\left(\hat{H}_{n,R}^0(\tau_n^R) - \mathbb{Y}_{n,R}(\tau_n^R)\right) = 0$$

and

$$D_n(u) = \frac{1}{2}\left(\hat{\varphi}_n(u) - \varphi_0(m)\right)^2 \quad \text{and} \quad R_n(u) = \sum_{j=3}^{\infty}\frac{1}{j!}\left(\hat{\varphi}_n(u) - \varphi_0(m)\right)^j$$
$$D_n^0(u) = \frac{1}{2}\left(\hat{\varphi}_n^0(u) - \varphi_0(m)\right)^2 \quad \text{and} \quad R_n^0(u) = \sum_{j=3}^{\infty}\frac{1}{j!}\left(\hat{\varphi}_n^0(u) - \varphi_0(m)\right)^j.$$

Lemma 5.2.18 (Identities)**.** Under Assumptions A and D, we have the following identities. First,

$$f_0(m)^{-1}\left(\hat{f}_n(u) - \tilde{f}_0(u)\right) = \hat{\varphi}_n(u) - \tilde{\varphi}_0(u) + D_n(u) + R_n(u).$$

Next, the f-processes and the φ-processes are related by

$$\hat{H}_n^{\varphi}(t) = \frac{\hat{H}_n^f(t)}{f_0(m)} - n^{4/5}\int_m^{m_n(t)}\int_m^v \left(D_n(u) + R_n(u)\right) du dv$$
$$\hat{H}_{n,L}^{\varphi}(t) = \frac{\hat{H}_{n,L}^f(t)}{f_0(m)} - n^{4/5}\int_{m_n(t)}^{\tau_n^L}\int_v^{\tau_n^L} \left(D_n^0(u) + R_n^0(u)\right) du dv$$
$$\hat{H}_{n,R}^{\varphi}(t) = \frac{\hat{H}_{n,R}^f(t)}{f_0(m)} - n^{4/5}\int_{\tau_n^R}^{m_n(t)}\int_{\tau_n^R}^v \left(D_n^0(u) + R_n^0(u)\right) du dv.$$

(Compare these identities to the definitions of $\mathbb{Y}_n^{\varphi}(t), \mathbb{Y}_{n,L}^{\varphi}(t)$, and $\mathbb{Y}_{n,R}^{\varphi}(t)$.) Now, taking $u \in [m - Mn^{-1/5}, m + Mn^{-1/5}]$ for any $M > 0$, we have

$$R_n(u) - o_p(n^{-2/5}) = R_n^0(u) - o_p(n^{-2/5}) = 0,$$

$$D_n(u) - o_p(n^{-2/5}) = D_n^0(u) - o_p(n^{-2/5}) = \frac{1}{2}\left(\varphi_0'(m)(u-m)\right)^2 = 0,$$

all uniformly for $u \in [m - Mn^{-1/5}, m + Mn^{-1/5}]$. Additionally, recalling the notation $g(a,b] = g(b) - g(a)$, we have

$$\left|\left((\hat{H}_{n,R}^{\varphi})' - (\mathbb{Y}_{n,R}^{\varphi})'\right)\left(a_n^{-1}(\tau_n^L), a_n^{-1}(\tau_n^R)\right]\right| \leq \frac{4}{f_0(m)} n^{-2/5}$$

with high probability for large n, and

$$\mathbb{Y}_{n,L}^{\varphi}(t) - \hat{H}_{n,L}^{\varphi}(t) \geq 0,$$
$$\mathbb{Y}_{n,R}^{\varphi}(t) - \hat{H}_{n,R}^{\varphi}(t) \geq 0,$$

and

$$\int_{\infty}^{\tau_n^{-1}} (\mathbb{Y}_{n,L}^{\varphi}(t) - \hat{H}_{n,L}^{\varphi}(t)) d(\hat{H}_{n,L}^{\varphi})^{(3)}(t) = 0 = \int_{\tau_n^1}^{\infty} \left(\mathbb{Y}_{n,R}^{\varphi}(t) - \hat{H}_{n,R}^{\varphi}(t)\right) d(\hat{H}_{n,R}^{\varphi})^{(3)}(t)$$

where τ_n^1 is the first right-knot of $(\hat{H}_{n,R}^{\varphi})''$ greater or equal to m and τ_n^{-1} is the first left-knot less than or equal to m.

Definition 5.2.19 (Derivatives of the Y-processes)**.**

$$\mathbb{X}_n^f(v) := (\mathbb{Y}_n^f)'(v) = n^{3/5} \int_m^{m_n(v)} \left(d\mathbb{F}_n(u) - \tilde{f}_0(u)du\right),$$
$$\mathbb{X}_{n,L}^f(v) := -(\mathbb{Y}_{n,L}^f)'(v) = n^{3/5} \int_{m_n(v)}^{\tau_n^L} \left(d\mathbb{F}_n(u) - \tilde{f}_0(u)du\right),$$
$$\mathbb{X}_{n,R}^f(v) := (\mathbb{Y}_{n,R}^f)'(v) = n^{3/5} \int_{\tau_n^R}^{m_n(v)} \left(d\mathbb{F}_n(u) - \tilde{f}_0(u)du\right),$$

and

$$\mathbb{X}_n^{\varphi}(v) := (\mathbb{Y}_n^{\varphi})'(v) = \frac{\mathbb{X}_n^f(v)}{f_0(m)} - n^{3/5}\int_m^{m_n(v)} (D_n(u) + R_n(u))\, du,$$
$$\mathbb{X}_{n,L}^{\varphi}(v) := -(\mathbb{Y}_{n,L}^{\varphi})'(v) = \frac{\mathbb{X}_{n,L}^f(v)}{f_0(m)} - n^{3/5}\int_{m_n(v)}^{\tau_n^L} \left(D_n^0(u) + R_n^0(u)\right) du,$$
$$\mathbb{X}_{n,R}^{\varphi}(v) := (\mathbb{Y}_{n,R}^{\varphi})'(v) = \frac{\mathbb{X}_{n,R}^f(v)}{f_0(m)} - n^{3/5}\int_{\tau_n^R}^{m_n(v)} \left(D_n^0(u) + R_n^0(u)\right) du.$$

The definitions that follow make γ_1 and γ_2 satisfy

$$|\gamma_1|\gamma_2^{3/2} = \sqrt{f_0(m)} \qquad \text{and} \qquad |\gamma_1|\gamma_2^4 = \frac{4!}{|\varphi_0''(m)|}.$$

Definition 5.2.20. Let

$$\gamma_1 = -\left(\frac{f_0(m)^{4/3}|\varphi_0''(m)|}{4!}\right)^{3/5} \qquad \text{and} \qquad \gamma_2 = \left(\frac{4!}{|\varphi_0''(m)|\sqrt{f_0(m)}}\right)^{2/5}.$$

Let

$$\tilde{H}_n(t) = \gamma_1 \hat{H}_n^{\varphi}(\gamma_2 t), \qquad \tilde{H}_{n,L}(t) = \gamma_1 \hat{H}_{n,L}^{\varphi}(\gamma_2 t), \quad \tilde{H}_{n,R}(t) = \gamma_1 \hat{H}_{n,R}^{\varphi}(\gamma_2 t),$$
$$\tilde{\mathbb{Y}}_n(t) = \gamma_1 \mathbb{Y}_n^{\varphi}(\gamma_2 t), \qquad \tilde{\mathbb{Y}}_{n,L}(t) = \gamma_1 \mathbb{Y}_{n,L}^{\varphi}(\gamma_2 t), \quad \tilde{\mathbb{Y}}_{n,R}(t) = \gamma_1 \mathbb{Y}_{n,R}^{\varphi}(\gamma_2 t),$$
$$\tilde{\mathbb{X}}_n(t) = \gamma_1\gamma_2 \mathbb{X}_n^{\varphi}(\gamma_2 t), \qquad \tilde{\mathbb{X}}_{n,L}(t) = \gamma_1\gamma_2 \mathbb{X}_{n,L}^{\varphi}(\gamma_2 t), \quad \tilde{\mathbb{X}}_{n,R}(t) = \gamma_1\gamma_2 \mathbb{X}_{n,R}^{\varphi}(\gamma_2 t),$$

and we also let the new change-of-scale function be defined as

$$m_{n,\gamma_2}(t) := m_n(\gamma_2 t) = m + \frac{\gamma_2}{n^{1/5}} t, \quad \text{so that} \quad m_{n,\gamma_2}^{-1}(s) = \frac{m_n^{-1}(s)}{\gamma_2} = \frac{n^{1/5}}{\gamma_2}(t - m).$$

We study asymptotics in the space

$$E_c = (\mathcal{D}_c \times \mathcal{C}_c^3)^3 \times (\mathcal{D}_c \times \mathcal{C}_c)^3 \times \mathbb{R}^4,$$

for $0 < c \leq \infty$. The local process we study in E_∞ is

$$\{Z_n\} := \Big\{(\tilde{H}'''_{n,L}, \tilde{H}''_{n,L}, \tilde{H}'_{n,L}, \tilde{H}_{n,L}, \tilde{H}'''_{n,R}, \tilde{H}''_{n,R}, \tilde{H}'_{n,R}, \tilde{H}_{n,R}, \tilde{H}'''_n, \tilde{H}''_n, \tilde{H}'_n, \tilde{H}_n,$$
$$\tilde{\mathbb{Y}}'_{n,L}, \tilde{\mathbb{Y}}_{n,L}, \tilde{\mathbb{Y}}'_{n,R}, \tilde{\mathbb{Y}}_{n,R}, \tilde{\mathbb{Y}}'_n, \tilde{\mathbb{Y}}_n, m^{-1}_{n,\gamma_2}(\tau^L_n), m^{-1}_{n,\gamma_2}(\tau^R_n), m^{-1}_{n,\gamma_2}(\tau^0_{n,-1}), m^{-1}_{n,\gamma_2}(\tau^0_{n,1})\Big\}.$$

5.3 Inference for the Mode: Likelihood Ratio Methods

In the previous section we found the joint limit distribution of the UMLE $\hat{f}_n$ and the CMLE $\hat{f}_n^0$. We can now use that joint limit distribution to study the limit distribution of the main statistic of interest, the log likelihood ratio statistic, $2\log\lambda_n$. As mentioned in Section 1.3, one benefit to likelihood ratio statistics is that they are often asymptotically pivotal, much as in the classical parametric case, and hence using the likelihood ratio statistic (to test or to form confidence intervals) is straightforward. This phenomenon is sometimes called the "Wilks phenomenon" after Wilks (1938). An alternative approach to testing or forming asymptotic confidence intervals for the mode is to use a Wald type test statistic based on $\hat{m}_n := m(\hat{f}_n)$, which involves estimating the constant $K_{f_0} = ((4!)^2 f_0(m)/f_0''(m)^2)^{1/5}$ appearing in (1.2). We do not need to deal with this extra complication when we use the log likelihood ratio, since it is asymptotically pivotal. Thus, our first interest in $2\log\lambda_n$, equal to $2n\int_{-\infty}^{\infty}\log\left(\hat{f}_n/\hat{f}_n^0\right)\,d\mathbb{F}_n$, is to establish its limit distribution and see that it does not depend on any parameters of the underlying true density. We will consider this log likelihood integral over two separate sets: over a local neighborhood of the mode m, and over the complement of that neighborhood. Our main unproven conjecture is as follows.

Conjecture 5.3.1. *Let $\hat{f}(x)$ be $H''(x)$, where H is defined in Theorem 5.1.12, and let $\hat{f}^0(x)$ be $H_R''(x)$ where H_R is defined in Theorem 5.1.13. Then under Assumptions A and D,*

$$2\log\lambda_n \longrightarrow_d \int_{-\infty}^{\infty}(\hat{f}^2(u) - (\hat{f}^0)^2(u))\,du =: \mathbb{D}. \tag{5.128}$$

If this conjecture is true then it implies that the Wilks phenomenon holds for log-concave densities with $f_0''(m) < 0$. We could then test or form an (asymptotic) confidence interval for the mode based on the log likelihood ratio statistic. In order to calibrate the test or interval, we need to understand the limit distribution on the right of (5.128), $\mathbb{D}$. We could estimate $\mathbb{D}$ by picking any log-concave density f_0 with $f_0''(m) < 0$ and computing $2\log\lambda_n$, where n is large, a large number of times. By Conjecture 5.3.1, this would give an estimate of the limit distribution of $2\log\lambda_n$, and thus an estimate of limit $(1-\alpha)$-quantiles $q_{1-\alpha}$, for any f_0 with $f_0''(m) < 0$. Thus, for $f_0 \in \mathcal{P}$ with $f_0''(m) < 0$, we can consider the hypothesis test

$$H_0 : m = m_0 \quad \text{versus} \quad H_1 : m \neq m_0 \tag{5.129}$$

where $m_0 \in \mathbb{R}$ is fixed. We could use $2\log\lambda_n(m_0)$, the log likelihood ratio given by using m_0 as the constrained mode for $\hat{f}_n^0$, to form the asymptotically level-α test $\phi_{\alpha,n}(m_0)$ given by

$$\phi_{\alpha,n}(m_0) = \begin{cases} 1 & \text{if } \; 2\log\lambda_n(m_0) > q_{1-\alpha} \\ 0 & \text{if } \; 2\log\lambda_n(m_0) \leq q_{1-\alpha} \end{cases}, \tag{5.130}$$

where 1 means to reject. Then to compute a confidence set, we can invert the above test as a function from $\mathbb{R}$, the space of all modes, to $\{0,1\}$. That is, an asymptotically α-level confidence set would be

$$\phi_{\alpha,n}^{-1}(0) = \{m | \phi_{\alpha,n}(m) = 0\} = \{m | 2\log\lambda_n(m) \leq q_{1-\alpha}\} \subset \mathbb{R}, \tag{5.131}$$

all the mode values m we do not reject at the α level.

Note that the conjectured form of the limit distribution in Conjecture 5.3.1 is the same as in Theorem 2.5 of Banerjee and Wellner (2001), which is about constraining the value of monotone functions at a fixed point in the domain, t_0. In that work, away from an interval shrinking towards the location of the constraint, t_0, the constrained and unconstrained MLEs are identical. Thus, the likelihood ratio there only depends on a neighborhood about t_0. In our case, the UMLE $\hat{f}_n$ and the CMLE $\hat{f}_n^0$ are not identically equal away from m (which is analogous to t_0), so we have a remainder term to handle. Simulations indicate that away from the mode the two estimators are indeed much closer to each other than they are near to the mode. See Section 6.2 for simulation results, which include simulations providing evidence that the Wilks phenomenon holds. In this section, we will give the (mathematical) indications that Conjecture 5.3.1 holds and make clear the difficulties remaining.

We will begin by giving a rapid sketch of the type of argument we are making and of possible approaches. We can write our statistic as

$$\begin{aligned}
2\log\lambda_n &= 2n\int_{[X_{(1)},X_{(n)}]}(\hat{\varphi}_n(u)-\hat{\varphi}_n^0(u))d\mathbb{F}_n(u)\\
&= 2n\int_{[X_{(1)},X_{(n)}]}(\hat{\varphi}_n(u)-\hat{\varphi}_n^0(u))d\mathbb{F}_n(u)-2n\int_{[X_{(1)},X_{(n)}]}(e^{\hat{\varphi}_n(u)}-e^{\hat{\varphi}_n^0(u)})du.
\end{aligned}$$

Adding the second term, equal to 0, is key to understanding the statistic. Next, we Taylor expand the exponential terms we just added in, which will leave us with an expression that looks like

$$2n\int((\hat{\varphi}_n-\varphi_0)-(\hat{\varphi}_n^0-\varphi_0))d(\mathbb{F}_n-F_0)-n\int((\hat{\varphi}_n-\varphi_0)^2-(\hat{\varphi}_n^0-\varphi_0)^2)dF_0,$$

plus some remainder terms that depend on the Taylor expansion of the exponential function. We could attempt to take this to the limit and perhaps arrive at

$$2\int_{-c}^{c}(\hat{f}-\hat{f}^0)dX-\int_{-c}^{c}(\hat{f}^2-(\hat{f}^0)^2)d\lambda, \tag{5.132}$$

where λ is Lebesgue measure, dX is as in Theorem 5.1.12 or Theorem 5.1.13, and $\hat{f}$ and $\hat{f}^0$ are the unconstrained and constrained limiting estimators from those theorems, respectively. We would then want to convert this expression to

$$\int_{-\infty}^{\infty}(\hat{f}^2-(\hat{f}^0)^2)d\lambda,$$

which is the form in which we would like to write the limit. This would require doing an integration by parts using properties of the limiting estimators as well as letting c in (5.132) go to ∞. Rather than do the integration by parts conversion in the limit, we do it in the finite sample world. This turns out to take care of both problems at once. It leaves us, however, with several remainder terms, as previously mentioned. We will now start a more rigorous analysis of the statistic. We have

$$\begin{aligned} 2\log\lambda_n &= 2n\int_{[X_{(1)},X_{(n)}]}(\hat{\varphi}_n-\hat{\varphi}_n^0)d\mathbb{F}_n - 2n\int_{[X_{(1)},X_{(n)}]}(e^{\hat{\varphi}_n(u)}-e^{\hat{\varphi}_n^0(u)})du \\ &= 2n\int_{[X_{(1)},X_{(n)}]}\hat{\varphi}_n d\hat{F}_n - \hat{\varphi}_n^0 d\hat{F}_n^0 - 2n\int_{[X_{(1)},X_{(n)}]}(e^{\hat{\varphi}_n(u)}-e^{\hat{\varphi}_n^0(u)})du, \end{aligned}$$

by taking $\Delta = \hat{\varphi}_n$ and $\Delta = \hat{\varphi}_n^0$ in Theorem 2.2, page 43, of Balabdaoui et al. (2009) and our analogous Theorem 2.0.4 for the constrained estimator. Next, we break these integrals up onto two domains. We define

$$D_n := [m - b_n, m + b_n] \quad \text{and} \quad D_n^c := [X_{(1)}, X_{(n)}] \setminus D_n, \tag{5.133}$$

for some sequence $b_n \to 0$ as $n \to \infty$. Then we can break our statistic up into integrals on D_n and on D_n^c:

$$\begin{aligned} 2\log\lambda_n = 2n&\int_{D_n}(\hat{\varphi}_n d\hat{F}_n - \hat{\varphi}_n^0 d\hat{F}_n^0) - 2n\int_{D_n}(e^{\hat{\varphi}_n(u)}-e^{\hat{\varphi}_n^0(u)})du \\ &+ 2n\int_{D_n^c}(\hat{\varphi}_n d\hat{F}_n - \hat{\varphi}_n^0 d\hat{F}_n^0) - 2n\int_{D_n^c}(e^{\hat{\varphi}_n(u)}-e^{\hat{\varphi}_n^0(u)})du, \end{aligned}$$

which equals

$$2n\Bigg[\int_{D_n} (\hat{\varphi}_n(u) - \varphi_0(m))d\hat{F}_n(u) - (\hat{\varphi}_n^0(u) - \varphi_0(m))d\hat{F}_n^0(u) \tag{5.134}$$
$$- \int_{D_n} ((e^{\hat{\varphi}_n(u)} - e^{\varphi_0(m)}) - (e^{\hat{\varphi}_n^0(u)} - e^{\varphi_0(m)}))du$$
$$+ \int_{D_n^c} (\hat{\varphi}_n(u) - \varphi_0(u))d\hat{F}_n(u) - (\hat{\varphi}_n^0(u) - \varphi_0(u))d\hat{F}_n^0(u)$$
$$- \int_{D_n^c} ((e^{\hat{\varphi}_n(u)} - e^{\varphi_0(u)}) - (e^{\hat{\varphi}_n^0(u)} - e^{\varphi_0(u)}))du$$
$$+ R_{n,1}\Bigg],$$

where

$$R_{n,1} := \int_{[X_{(1)},X_{(n)}]} R_{n,1}(x)dx := \int_{[X_{(1)},X_{(n)}]} (\varphi_0(m)1_{D_n}(x) + \varphi_0(x)1_{D_n^c}(x)) \left(\hat{f}_n(x) - \hat{f}_n^0(x)\right) dx.$$

We will not formally analyze $R_{n,1}$ yet, since it will depend on the precise definition of D_n, but we make a heuristic comment about why we would expect it to be small. On D_n^c, we expect $\left(\hat{f}_n(x) - \hat{f}_n^0(x)\right)$ to be small. We expect the term $\int_{[X_{(1)},X_{(n)}]} (\varphi_0(m)1_{D_n}(x)) \left(\hat{f}_n(x) - \hat{f}_n^0(x)\right) dx$ to be small because of Corollary 2.0.9 and the analogous result, Corollary 2.5, page 44, of Dümbgen and Rufibach (2009). This intuition is based on the idea that we choose the endpoints of D_n to be knots shared by both estimators, in which case $\int_{D_n} \left(\hat{f}_n(x) - \hat{f}_n^0(x)\right) dx \leq 2/n$. We do not yet know if the CMLE and the UMLE share knots, but, even if the two estimators do not share knots exactly, perhaps they will have knots nearby and this remainder term will still be small. Continuing from (5.134), we can expand the two normalizing integrals as

$$\int_{D_n} ((e^{\hat{\varphi}_n(u)} - e^{\varphi_0(m)}) - (e^{\hat{\varphi}_n^0(u)} - e^{\varphi_0(m)}))du + \int_{D_n^c} ((e^{\hat{\varphi}_n(u)} - e^{\varphi_0(u)}) - (e^{\hat{\varphi}_n^0(u)} - e^{\varphi_0(u)}))du,$$

which is equal to

$$\int_{D_n}\left(e^{\varphi_0(m)}\left((\hat\varphi_n(u)-\varphi_0(m))+\frac{1}{2}(\hat\varphi_n(u)-\varphi_0(m))^2\right)+R_{n,2}(u)\right)du$$
$$-\int_{D_n}\left(e^{\varphi_0(m)}\left((\hat\varphi_n^0(u)-\varphi_0(m))+\frac{1}{2}(\hat\varphi_n^0(u)-\varphi_0(m))^2\right)+R_{n,2}^0(u)\right)du$$
$$+\int_{D_n^c}\left(e^{\varphi_0(u)}\left((\hat\varphi_n(u)-\varphi_0(u))+\frac{1}{2}(\hat\varphi_n(u)-\varphi_0(u))^2\right)+R_{n,2}(u)\right)du$$
$$-\int_{D_n^c}\left(e^{\varphi_0(u)}\left((\hat\varphi_n^0(u)-\varphi_0(u))+\frac{1}{2}(\hat\varphi_n^0(u)-\varphi_0(u))^2\right)+R_{n,2}^0(u)\right)du,$$

where the remainder terms are

$$R_{n,2}(u)=\frac{f_0(m)}{6}e^{\tilde x_n(u)}(\hat\varphi_n(u)-\varphi_0(m))^3 1_{D_n}(u)+\frac{f_0(u)}{6}e^{\tilde x_n(u)}(\hat\varphi_n(u)-\varphi_0(u))^3 1_{D_n^c}(u)$$
$$R_{n,2}^0(u)=\frac{f_0(m)}{6}e^{\tilde x_n^0(u)}(\hat\varphi_n^0(u)-\varphi_0(m))^3 1_{D_n}(u)+\frac{f_0(u)}{6}e^{\tilde x_n^0(u)}(\hat\varphi_n^0(u)-\varphi_0(u))^3 1_{D_n^c}(u),$$

and $\tilde x_n(u)$ is between 0 and $\hat\varphi_n(u)-\varphi_0(m)$ for $u\in D_n$, and between 0 and $\hat\varphi_n(u)-\varphi_0(u)$ for $u\in D_n^c$.[3] Similarly, $\tilde x_n^0(u)$ is between 0 and $\hat\varphi_n^0(u)-\varphi_0(m)$ for $u\in D_n$, and between 0 and $\hat\varphi_n^0(u)-\varphi_0(u)$ for $u\in D_n^c$. Then we have, removing for now the remainder terms,

$$2\log\lambda_n-n\left(R_{n,1}-\int_{[X_{(1)},X_{(n)}]}(R_{n,2}(u)-R_{n,2}^0(u))du\right)$$
$$=n\Big[2\int_{D_n}(\hat\varphi_n(u)-\varphi_0(m))(d\hat F_n(u)-f_0(m)du)-2\int_{D_n}(\hat\varphi_n^0(u)-\varphi_0(m))(d\hat F_n^0(u)-f_0(m)du)$$
$$-\int_{D_n}\left((\hat\varphi_n(u)-\varphi_0(m))^2-(\hat\varphi_n^0(u)-\varphi_0(m))^2\right)f_0(m)du$$
$$+2\int_{D_n^c}(\hat\varphi_n(u)-\varphi_0(u))(d\hat F_n(u)-f_0(u)du)-2\int_{D_n^c}(\hat\varphi_n^0(u)-\varphi_0(u))(d\hat F_n^0(u)-f_0(u)du)$$
$$-\int_{D_n^c}\left((\hat\varphi_n(u)-\varphi_0(u))^2-(\hat\varphi_n^0(u)-\varphi_0(u))^2\right)f_0(u)du\Big].$$

---

[3]Note that the remainder terms are measurable since they can each be written as the difference between two measurable functions.

Now we can repeat our Taylor expansions in the same fashion as previously on the first and third lines of the above expression. For the first line, we see that

$$\begin{aligned}
2\int_{D_n} & (\hat{\varphi}_n(u)-\varphi_0(m))(e^{\hat{\varphi}_n(u)-\varphi_0(m)}-1)f_0(m)du - 2\int_{D_n}(\hat{\varphi}_n^0(u)-\varphi_0(m))(e^{\hat{\varphi}_n^0(u)-\varphi_0(m)}-1)f_0(m)du \\
&= 2\int_{D_n}(\hat{\varphi}_n(u)-\varphi_0(m))\left((\hat{\varphi}_n(u)-\varphi_0(m))+e^{\tilde{x}_{n,2}(u)}\frac{1}{2}(\hat{\varphi}_n(u)-\varphi_0(m))^2\right)f_0(m)du \\
&\quad - 2\int_{D_n}(\hat{\varphi}_n^0(u)-\varphi_0(m))\left((\hat{\varphi}_n(u)-\varphi_0(m))+e^{\tilde{x}_{n,2}(u)}\frac{1}{2}(\hat{\varphi}_n(u)-\varphi_0(m))^2\right)f_0(m)du \\
&= 2\int_{D_n}\big((\hat{\varphi}_n(u)-\varphi_0(m))^2 f_0(m)+R_{n,3}(u)\big)\,du \\
&\quad - 2\int_{D_n}\big((\hat{\varphi}_n^0(u)-\varphi_0(m))^2 f_0(m)+R_{n,3}^0(u)\big)\,du,
\end{aligned}$$

and, similarly, we see that

$$\begin{aligned}
2\int_{D_n^c} & (\hat{\varphi}_n(u)-\varphi_0(u))(e^{\hat{\varphi}_n(u)-\varphi_0(u)}-1)f_0(u)du - 2\int_{D_n}(\hat{\varphi}_n^0(u)-\varphi_0(u))(e^{\hat{\varphi}_n^0(u)-\varphi_0(u)}-1)f_0(u)du \\
&= 2\int_{D_n^c}\big((\hat{\varphi}_n(u)-\varphi_0(u))^2 f_0(u)+R_{n,3}(u)\big)\,du \\
&\quad - 2\int_{D_n^c}\big((\hat{\varphi}_n^0(u)-\varphi_0(u))^2 f_0(u)+R_{n,3}^0(u)\big)\,du,
\end{aligned}$$

where the remainder terms are nearly identical to previous remainder terms in form (but not in value),

$$\begin{aligned}
R_{n,3}(u) &= \frac{f_0(m)}{2}e^{\tilde{x}_{n,3}(u)}(\hat{\varphi}_n(u)-\varphi_0(m))^3 1_{D_n}(u)+\frac{f_0(u)}{2}e^{\tilde{x}_{n,3}(u)}(\hat{\varphi}_n(u)-\varphi_0(u))^3 1_{D_n^c}(u) \\
R_{n,3}^0(u) &= \frac{f_0(m)}{2}e^{\tilde{x}_{n,3}^0(u)}(\hat{\varphi}_n^0(u)-\varphi_0(m))^3 1_{D_n}(u)+\frac{f_0(u)}{2}e^{\tilde{x}_{n,3}^0(u)}(\hat{\varphi}_n^0(u)-\varphi_0(u))^3 1_{D_n^c}(u).
\end{aligned}$$

Thus, what we have shown is

$$
\begin{aligned}
&2\log\lambda_n - n\int_{[X_{(1)},X_{(n)}]}(R_{n,1}(u)+R_{n,2}(u)-R^0_{n,2}(u)+R_{n,3}(u)-R^0_{n,3}(u))du \\
&\quad = n\int_{D_n}\left((\hat\varphi_n(u)-\varphi_0(m))^2-(\hat\varphi^0_n(u)-\varphi_0(m))^2\right)f_0(m)du \qquad (5.135)\\
&\qquad + n\int_{D_n^c}\left((\hat\varphi_n(u)-\varphi_0(u))^2-(\hat\varphi^0_n(u)-\varphi_0(u))^2\right)f_0(u)du.
\end{aligned}
$$

Now we will analyze the first term on the right side of the above display, and see that, with an appropriate definition of D_n, it converges to the limit distribution that we expect. Let us define D_n as

$$D_n := [m - b_{n,L}n^{-1/5}, m + b_{n,R}n^{-1/5}],$$

where $b_{n,L}, b_{n,R} \in [M_1, M_2]$, for $0 < M_1 < M_2$. Then using the change of variables $u = m_n(t) = m + tn^{-1/5}$, we can write the first term on the right side of (5.135) as

$$
\begin{aligned}
&n\int_{m-b_{n,L}n^{-1/5}}^{m+b_{n,R}n^{-1/5}}\left[(\hat\varphi_n(u)-\varphi_0(m))^2-\left(\hat\varphi^0_n(u)-\varphi_0(m)\right)^2\right]f_0(m)\,du \\
&\quad = f_0(m)n^{1/5}\int_{b_{n,L}}^{b_{n,R}} n^{4/5}\left[(\hat\varphi_n(m_n(t))-\varphi_0(m))^2-\left(\hat\varphi^0_n(m_n(t))-\varphi_0(m)\right)^2\right]n^{-1/5}dt.
\end{aligned}
$$

Now, if $b_{n,L}$ and $b_{n,R}$ converge in probability to b_L and b_R, respectively, then by Theorem 5.2.13, and by (5.126) and (5.127), the above expression converges in distribution to

$$f_0(m)\int_{b_L}^{b_R}\left[\left(\frac{1}{\gamma_1\gamma_2^2}H''\left(\frac{t}{\gamma_2}\right)\right)^2-\left(\frac{1}{\gamma_1\gamma_2^2}H''_R\left(\frac{t}{\gamma_2}\right)\right)^2\right]dt \qquad (5.136)$$

where H and H''_R are defined as in Theorem 5.1.12 and Theorem 5.1.13 respectively. Now, by the definition of γ_1 and γ_2 in (5.112), the above display is equal to

$$\frac{1}{\gamma_2}\int_{b_L}^{b_R}\left[H''\left(\frac{t}{\gamma_2}\right)^2-H''_R\left(\frac{t}{\gamma_2}\right)^2\right]dt = \int_{b_L/\gamma_2}^{b_R/\gamma_2}\left[H''(s)^2-H''_R(s)^2\right]ds. \qquad (5.137)$$

If we let b_L and b_R converge to $-\infty$ and ∞, respectively, then this gives the limit distribution in Conjecture 5.3.1. Thus, what remains is to show that all the remainder terms in (5.135), including the difference-of-squares integral on D_n^c, are negligible. Recall from Corollary 3.2.8 and Corollary 3.2.11 that $\int_{-\infty}^{\infty} \log\left(\hat{f}_n/\hat{f}_n^0\right) d\mathbb{F}_n = O_p(n^{-4/5})$. If we restrict this integral to D_n^c, then we need to improve the rate by a factor of $n^{-1/5}$ and convert from O_p to o_p to be done. Alternatively, many of the remainder terms involve quantities such as $n\int_{D_n^c}\left((\hat{\varphi}_n(u)-\varphi_0(u))^2-(\hat{\varphi}_n^0(u)-\varphi_0(u))^2\right) f_0(u)du$. Recall from Theorem 3.2.6 and Theorem 3.2.9 that $\int_{-\infty}^{\infty}\left(e^{\hat{\varphi}_n/2}-e^{\hat{\varphi}_n^0/2}\right)^2 d\lambda = O_p(n^{-4/5})$. If this rate of convergence can be improved when the integral is restricted to D_n^c (by $n^{-1/5}$ and from O_p to o_p) and can give a rate of convergence for $(\hat{\varphi}_n(u)-\varphi_0(u))^2-(\hat{\varphi}_n^0(u)-\varphi_0(u))^2$, we will similarly be done.

5.4 *Future Work*

There are many additional questions which are beyond the scope of this thesis. One major aim is to finish the proof of Conjecture 5.3.1. Many other questions remain, too. After we find the limiting distribution of our likelihood ratio statistic under the null hypothesis, we will be interested in the behavior of the likelihood ratio statistc under (both fixed and local) alternative hypotheses wherein m is not equal to the hypothesized m_0 in the hypothesis test (5.129). We conjecture that in this scenario,

$$n^{-1} 2 \log \lambda_n \to_p 2 \inf K(f_0, f_0^*),$$

where $K(f_0, f_0^*) := E_{f_0}(\log(f_0/f_0^*))$ is the Kullback-Leibler divergence, and the infimum is taken over all $f_0^* \in \mathcal{P}$ that have mode equal to m_0. We further conjecture that the infimum is achieved by some fixed f_0^* that minimizes the Kullback-Leibler divergence. Note that since $K(f_0, f_0^*) > 0$ for any two densities f_0 and f_0^* that are not Lebesgue-almost-everywhere equal, this would imply that the likelihood ratio test has power asymptotically equal to 1.

Understanding the behavior of $2 \log \lambda_n$ under local (contiguous) alternative hypotheses is another goal. Knowing the limit distribution of $2 \log \lambda_n$ under local alternatives allows us to perform (asymptotically correct) power calculations. We conjecture that when the data come from a sequence of contiguous alternative distributions, $2 \log \lambda_n$ will have a limit in distribution of similar form as $\mathbb{D}$ in (5.128), but a different actual distribution. We expect that $\hat{f}$ and $\hat{f}^0$ in (5.128) will be replaced by analogous convex and modally-constrained convex functions that will be defined by minimizing an objective like (5.8), but wherein X in (5.6) is replaced by a process $X(t) = W(t) - 4t^3 - \Psi(t)$ where $\Psi(t)$ will depend on the sequence of contiguous alternatives.

We may also consider the "misspecified" case wherein the true density f_0 is not in $\mathcal{P}$, i.e. is not log-concave. Cule et al. (2010) show that for any density f_0 with a finite mean and satisfying $\int_{\mathbb{R}} f_0 \log_+ f_0 \, d\lambda < \infty$, there is a log-concave density f_0^* that minimizes the Kullback-Leibler distance from f_0 to the class of log-concave densities. Here, $\log_+ x = \max(\log x, 0)$. Thus, it would be interesting to consider how our likelihood ratio statistic behaves when we hypothesize the correct mode m_0 but the true density is not log-concave. Does our likelihood ratio statistic still provide a testing procedure and confidence interval?

The methods here may be extendable to similar settings such as, e.g., the setting of a concave regression function (Seijo and Sen, 2011) or of a convex hazard function (Jankowski and Wellner, 2009) (see Chapter 1, page 2). Additionally, enforcing convexity-based shape-constraints may be useful in semiparametric settings; for instance, one could constrain the hazard function in the proportional hazards model (Cox and Oakes, 1984) to be convex. Dümbgen et al. (2011) study the regression setting in which errors are specified to be log-concave and the regression functions are assumed to lie in some parametric or nonparametric class. They propose an iterative algorithm for computing the estimates. The algorithm requires estimating the error distribution and recentering it since it will not be centered at 0. One alternative is to consider "modal regression," wherein we use the log-concave estimator constrained to have its mode at 0 for the error distribution. With this approach, no recentering would be necessary.

Another possibility would be to model the errors in regression as having log-concave densities that are symmetric about 0. Symmetry will imply that the mean, median, and mode all coincide for these densities, so recentering will certainly be unnecessary. Symmetry in combination with log-concavity is also a very appealing constraint in its own right. To our knowledge, concave shape-constraints have not yet been combined with symmetry in density estimation problems. In $\mathbb{R}^d$, there are multiple possible symmetries to enforce, so which is most useful would need to be explored.

One motivation for forming confidence intervals for the mode via inversion of likelihood ratio tests is the poor performance of the bootstrap in both shape-constrained problems and in mode estimation problems, as in Sen et al. (2010) and Romano (1988a), respectively. Sen et al. (2010) considers the bootstrap in the case of problems based on monotonicity. However, we do not yet know the performance of the bootstrap in cases with (underlying) concavity such as the classes of concave regression functions or of log-concave density functions, for estimation of the distributions of estimators at a fixed point or of estimators of the mode. The standard n-of-n bootstrap, the m-of-n bootstrap, or a bootstrap based on smoothing the MLE may have varying levels of success.

Log-concave densities necessarily have sub-exponential tails. Such light tails are inappropriate for some scenarios. It would be useful to extend the current methodology to problems where light-tailedness is not realistic, as well as to $\mathbb{R}^d$ for $d > 1$. Seregin and Wellner (2010) have shown consistency for their "convex-transformed" estimators in $\mathbb{R}^d$ which can have polynomial tails. In Section 3.2, for the case $d = 1$, we found the rate of convergence ($n^{-2/5}$) of these transformed estimators. (There, we referred to these as "concave-transformed" estimators, which is how we will refer to them now. The transformations h governing these estimators can be very general. It would be interesting to explore more fully the sorts of possibilities that these transformations allow.

For instance, the most common transformations chosen are $h \equiv h_s$, with h_s defined in (3.16) on page 50. These choices yield s-concave classes (with $s = 0$ corresponding to log-concave and $s = -\infty$ corresponding to all unimodal densities), where the densities are $h_s \circ \varphi$ where φ is concave. Recall that when we take $s = -\infty$, i.e. consider unimodal density estimation, the maximum likelihood estimate does not exist. In order to arrive at an estimate, one often bounds or penalizes the height at the mode (Woodroofe and Sun, 1993). Certain appropriate choices of transformation h may provide an alternative set of approaches: if $h(y)$ is chosen to be similar to $h_{s_1}(y)$ as $y \to \infty$ for s_1 relatively large, and $h(y)$ is chosen to be similar to $h_{s_2}(y)$ as $y \to -\infty$ for s_2 relatively small, then the class of h-transformed densities may be a useful surrogate for the class of unimodal ones. We described hybrid h of this sort in Example 3.2.22, page 65. How to choose s_1 and s_2 appropriately will require further study. From Theorem 2.13 of Seregin and Wellner (2010), we know that if $s_2 > -1$ then when n is large enough the s_2-concave MLE exists. It is not yet clear in a hybrid h as just described whether we would need $s_2 > -1$ or whether we could take s_2 closer to $-\infty$ to better approximate the class of unimodal densities. In Example 3.2.22, we note that taking $s_1 = 0$ fails the assumptions needed for existence, uniqueness, and consistency in Seregin and Wellner (2010). This is somewhat noninuitive, since as s_1 increases, the classes seemingly get smaller (see (3.15), page 50), so another question is whether the hypotheses in Seregin and Wellner (2010) can be weakened while maintaining consistency for a hybrid h. This approach to unimodal density estimation, using h-transformed classes with a hybrid-h, is related to the approach taken by Meyer and Woodroofe (2004). The authors of that work consider classes of densities that are unimodal except on an interval about their mode, on which the densities are forced to be concave.

There are still many things we do not know for concave-transformed density estimation. Algorithms for computing the estimators have not yet been developed. Furthermore for $d > 1$, we do not have asymptotic distribution theory (pointwise or global) for any classes of concave-transformed densities, including log-concave ones, at this point. In addition, it is suspected that rates of convergence for, say, the log-concave MLE for $d \geq 4$ will be suboptimal (see page 15 of Seregin and Wellner (2010)), motivating the study of sieved or penalized estimators. Cule et al. (2010) find an algorithm for computing the log-concave MLE in $\mathbb{R}^d$, implemented in the `LogConcDEAD` package available on CRAN, but for d much larger than 4 it can be prohibitively slow. Thus it will be important for penalized estimators to be developed in tandem with fast algorithms for computing the estimators.

Chapter 6

COMPUTATION, IMPLEMENTATION, AND APPLICATIONS

In this chapter we discuss computational aspects related to the CMLE. In Section 6.1 we extend the active set algorithm of Dümbgen et al. (2007) from the case of the UMLE to the case of the CMLE. Using the active set algorithm, in Section 6.2 we study $2\log\lambda_n$ and the resulting tests and confidence sets via simulation. In Section 6.3 we apply our procedures to some data examples.

6.1 Active Set Algorithm

In this section we will adopt the notation of Dümbgen et al. (2007) to make reference between that work and this one as simple as possible. Section 3 of Dümbgen et al. (2007) gives details of a general active set algorithm for maximizing an arbitrary continuous and concave function $L\colon \mathbb{R}^M \to [-\infty, \infty)$, $M \in \mathbb{N}$, that is coercive in the sense that $L(\psi) \to -\infty$ as $|\psi| \to \infty$, and that is continuously differentiable on $\operatorname{dom} L := \{\psi \in \mathbb{R}^M | L(\psi) > -\infty\}$. We want to maximize L over a closed convex set given by

$$\mathcal{K} := \left\{\psi \in \mathbb{R}^M | v_i'\psi \le c_i, i = 1, \ldots, q\right\}, \tag{6.1}$$

where $v_1, \ldots, v_q$ are linearly independent vectors in $\mathbb{R}^M$ and $c_1, \ldots, c_q \in \mathbb{R}$ are such that $\mathcal{K} \cap \operatorname{dom} L \neq \emptyset$. Here for a vector $v \in \mathbb{R}^M$, v' denotes its transpose. For any index set $A \subseteq \{1, \ldots, q\}$ we let

$$\mathcal{V}(A) := \left\{\psi \in \mathbb{R}^M | v_a'\psi = c_a \text{ for all } a \in A\right\}.$$

We want to compute (an element of)

$$\mathcal{K}_* := \arg\max_{\psi \in \mathcal{K}} L(\psi).$$

It is more simple to compute (an element of)

$$\mathcal{V}_*(A) := \arg\max_{\psi \in \mathcal{V}(A)} L(\psi),$$

so the idea is to vary A until we find that $\tilde{\psi}(A) \in \mathcal{V}_*(A)$ also belongs to K_*.

The active set algorithm depends on two ingredients. The first is an algorithm for computing a point $\tilde{\psi}(A) \in \mathcal{V}_*(A)$. This works the same for the CMLE as for the UMLE: a Newton-Raphson algorithm can be used, since we are maximizing a concave function over some (unconstrained) Euclidean space (whose dimension depends on how many constraints are active). The second is their Theorem 3.1. That theorem is as follows, where, for a vector $\psi \in \mathbb{R}^M$, we let

$$A(\psi) := \left\{ i \in \{1, \ldots, q\} | v_i' \psi \geq c_i \right\}.$$

Theorem 6.1.1 (Theorem 3.1, page 6, of Dümbgen et al. (2007))**.** *Let $b_1, \ldots, b_M$ be a basis of $\mathbb{R}^M$ such that*

$$v_i' b_j \begin{cases} > 0 & \text{if } \ i = j \leq q \\ = 0 & \text{otherwise} \end{cases}. \tag{6.2}$$

(i). A vector $\psi \in \mathcal{K} \cap \operatorname{dom} L$ belongs to $\mathcal{K}_*$ if and only if

$$b_i' \nabla L(\psi) \begin{cases} = 0 & \text{for all } \ i \in \{1, \ldots, M\} \setminus A(\psi) \\ \geq 0 & \text{for all } \ i \in A(\psi) \end{cases}. \tag{6.3}$$

(ii). For any given set $A \subseteq \{1, \ldots, q\}$, a vector $\psi \in \mathcal{V}(A) \cap \operatorname{dom} L$ belongs to $\mathcal{V}_*(A)$ if and

only if

$$b_i' \nabla L(\psi) = 0 \text{ for all } i \in \{1, \ldots, M\} \setminus A.$$

To use Theorem 6.1.1 we only need basis vectors $b_1, \ldots, b_M$ of $\mathbb{R}^M$ such that (6.2) holds, where the constraint vectors $v_1, \ldots, v_q$ are given by (6.1) (and in our case will specify concavity of the functions ψ). In Section 3.2 of Dümbgen et al. (2007) the values $b_1, \ldots, b_M$ and $v_1, \ldots, v_q$ are specified so that the the active set algorithm can be applied to the case of maximizing their objective function L, i.e. computing the UMLE. We will now specify acceptable vectors that satisfy (6.2) for the case of the CMLE, which will allow us to run the algorithm to compute the CMLE. We also define a slightly modified objective function L. Recall the definition of n_1 and $Z_1, \ldots, Z_{n_1}$, on page 23. The Z_i's were defined to be equal to the order statistics $X_{(i)}$, with the exception that Z_k is set to be equal to the mode m regardless of whether it is a data point or not. Now as in Dümbgen et al. (2007), we will allow $X_1, \ldots, X_n$ to come with weights $\tilde{p}_1, \ldots, \tilde{p}_n$. (These weights could arise from binning observations. They add no complexity to the algorithm, and are in fact necessary for the algorithm given in Section 3 of Dümbgen et al. (2007).) We define corresponding weights $p_1, \ldots, p_{n_1}$, for $Z_1, \ldots, Z_{n_1}$. These are defined so that $p_i = \tilde{p}_i$ for all $i \neq k$, and we define p_k equal to $\tilde{p}_k$ if $Z_k = X_k = m$, i.e. if m is a data point, and we define $p_k = 0$ if $Z_k = m$ is not a data point. Recall the log likelihood objective functional defined on page 24, which we want to maximize over the set of all concave functions on $\mathbb{R}$ with mode at m, i.e. over $\mathcal{C}_m$. We replace the weights of $1/n$ with the weights p_i to arrive at a generalized log likelihood function

$$\Psi_n(\varphi) := \sum_{i=1}^{n_1} p_i \varphi(Z_i) - \int_{\mathbb{R}} e^{\varphi} \, d\lambda.$$

where, recall that $p_k = 0$ if Z_k is not a data point. Now we only wish to consider $\varphi \in \mathcal{C}_m$ such that φ is also in $\mathcal{L}$, the class of piecewise linear functions with knots at $Z_1, \ldots, Z_{n_1}$.

The map ev_Z (for $Z = (Z_1, \ldots, Z_{n_1})$) is a bijection between $\mathcal{L}$ and $\mathbb{R}^{n_1}$. Thus we will now restrict attention to $\psi \in \mathbb{R}^{n_1}$ and define the corresponding objective function $L\colon \mathbb{R}^{n_1} \to \mathbb{R}$ by

$$L(\psi) := \sum_{i=1}^{n_1} p_i \psi_i - \int_{\mathbb{R}} e^{\psi} \, d\lambda, \tag{6.4}$$

so that $L(\operatorname{ev}_Z \varphi) = \Psi_n(\varphi)$. As noted in Section 3 of Dümbgen et al. (2007), a function $\varphi \in \mathcal{L}$ lies in $\mathcal{C}$ if and only if $\psi := \operatorname{ev}_Z \varphi \in \mathbb{R}^{n_1}$ satisfies

$$\frac{\psi_{j+1} - \psi_j}{\delta_j} - \frac{\psi_j - \psi_{j-1}}{\delta_{j-1}} =: \tilde{v}_j' \Psi \le 0 =: c_j, \qquad \text{for } j - 2, \ldots, n_1 - 1,$$

where we have defined $c_j := 0$, $\delta_j = Z_{j+1} - Z_j$, and $\tilde{v}_j = (\tilde{v}_{i,j})_{i=1}^{n_1}$ to have three nonzero components

$$\tilde{v}_{j-1,j} := \frac{1}{\delta_{j-1}}, \qquad \tilde{v}_{j,j} := -\frac{\delta_{j-1} + \delta_j}{\delta_{j-1}\delta_j}, \qquad \tilde{v}_{j+1,j} := \frac{1}{\delta_j}.$$

Now, it is also clear that $\varphi(m)$ is a local maximum for $\varphi \in \mathcal{L}$ when $X_{(1)} < m = Z_k < X_{(n)}$ if and only if $\psi := \operatorname{ev}_Z \varphi \in \mathbb{R}^{n_1}$ satisfies

$$Z_{k+j} - Z_{k-1+j} =: w_j' \psi \le 0 =: d_j, \qquad \text{for } j = 0, 1,$$

where we have defined $d_j := 0$, $w_0 := (0, \ldots, 0, 1, -1, 0, \ldots, 0)'$ and $w_1 := (0 \ldots, 0, 0, -1, 1, \ldots, 0)'$, where both -1's are in the kth position. Thus, since for concave functions local maxima are global maxima, we define the v_i sequence of constraints in (6.1) by

$$v_1 = \tilde{v}_2, \ldots, v_{k-2} = \tilde{v}_{k-1}, v_{k-1} = w_0, v_k = w_1, v_{k+1} = \tilde{v}_{k+1}, \ldots, v_{n_1-1} = \tilde{v}_{n_1-1}.$$

If $Z_k \le X_{(1)}$ then we exclude w_0, and if $Z_k \ge X_{(n)}$ then we exclude w_1. If $X_{(1)} < m = Z_k < X_{(n)}$ then there are $n_1 - 3$ vectors $\tilde{v}_i$ and 2 vectors w_i. If $Z_k \le X_{(1)}$ then there are $n_1 - 2$ vectors $\tilde{v}_i$ and 1 vector w_1, and similarly if $Z_k > X_{(n)}$. Thus in all cases, we have $n_1 - 1$

vectors v_i.

Next, we define the basis vectors for $\mathbb{R}^{n_1}$. Let $b_{n_1} = (1, \ldots, 1) \in \mathbb{R}^{n_1}$ and let

$$b_{i-1} = d_i = (\min(0, Z_i - Z_j))_{j=1}^{n_1} \text{ for } 2 \leq i \leq k \tag{6.5}$$

$$b_i = a_i = (\min(0, Z_j - Z_i))_{j=1}^{n_1} \text{ for } k \leq i \leq n_1 - 1, \tag{6.6}$$

which defines $(k-1) + (n_1 - k) + 1 = n_1$ linearly independent vectors in $\mathbb{R}^{n_1}$, a basis. Now we check that they satisfy the conditions (6.2) of Theorem 6.1.1. First, note that $v_i' b_{n_1} = 0$ for $1 \leq i \leq n_1 - 1$. Now we will consider the left side basis and constraints. For $1 \leq i \leq k-1$, $v_{i-1}' b_{i-1} > 0$ because for $2 \leq k-1$, $\tilde{v}_i d_i = \tilde{v}_{i,i} d_{i,i} > 0$, and $w_0' b_k = Z_k - Z_{k-1} > 0$. Next, we take $1 \leq i, j \leq k-1$ and show $v_i' b_j = 0$. This is immediate if $i - j \geq 1$. If $i - j = -1$, then

$$v_i' b_{i+1} = (Z_{i+1} - Z_{i-1}) \frac{1}{Z_i - Z_{i-1}} - (Z_{i+1} - Z_i) \left(\frac{1}{Z_i - Z_{i-1}} + \frac{1}{Z_{i+1} - Z_i} \right),$$

which equals

$$(\delta_i + \delta_{i-1}) \frac{1}{\delta_{i-1}} - \delta_i \left(\frac{1}{\delta_{i-1}} + \frac{1}{\delta_i} \right) = \frac{\delta_i}{\delta_{i-1}} + 1 - \left(\frac{\delta_i}{\delta_{i-1}} + 1 \right) = 0.$$

If $i - j \leq -2$, then $i \leq j - 2 \leq k - 3$, so $v_i' b_j = \tilde{v}_{i+1} d_{j+1}$, which equals

$$\tilde{v}_{i,i+1} d_{i,j+1} + \tilde{v}_{i+1,i+1} d_{i+1,j+1} + \tilde{v}_{i+2,i+1} d_{i+2,j+1},$$

which, since $Z_{j+1} - Z_i = \delta_j + \cdots + \delta_i$, is

$$\frac{1}{\delta_i}(\delta_j + \cdots + \delta_i) - \left(\frac{1}{\delta_{i+1}} + \frac{1}{\delta_i} \right)(\delta_j + \cdot + \delta_{i+1}) + \frac{1}{\delta_{i+1}}(\delta_j + \cdots + \delta_{i+2}),$$

which equals

$$\frac{1}{\delta_1}(\delta_j+\cdots+\delta_{i+1})+1-\frac{1}{\delta_i}(\delta_j+\cdots+\delta_{i+1})-\frac{1}{\delta_{i+1}}(\delta_j+\cdots+\delta_{i+2})-1+\frac{1}{\delta_{i+1}}(\delta_j+\cdots+\delta_{i+2})=0.$$

Thus, we have shown that the conditions (6.2) hold, and so we can use the active set algorithm to compute the CMLE.

6.2 Simulation Results

In this section we study the performance of our testing and confidence set procedures via simulation. We begin by studying the limit distribution of our statistic $2\log\lambda_n$, and presenting evidence that the Wilks phenomenon holds. We studied the limit distribution of $2\log\lambda_n$ from data generated by three separate true densities f_0 with $f_0''(m) < 0$: Gamma$(3,1)$ with density $f_0(x) = x^2 e^{-x}/\Gamma(3)$ on $(0,\infty)$, Beta$(2,3)$ with density $f_0(x) = x(1-x)^2/B(2,3)$ on $[0,1]$, Weibull$(1.5,1)$ with density $f_0(x) = 1.5x^{.5}e^{-x^{1.5}}$ on $[0,\infty)$, where $\Gamma(3)$, $B(2,3)$, and 1.5^{-1} are the appropriate normalizing constants, respectively. For each simulation we used a sample size of 1200 for computing the statistic and 10^4 Monte Carlo simulations to estimate the (limit) distribution. Figure 6.1 contains the plots of the empirical distribution functions from the 10^4 points. The dashed line at the top is the 95% confidence line. For all three of these densities with curvature at m the estimated distribution functions are nearly identical. Now, based on (1.2) on page 10, we might expect the limit distributions to depend on $f_0(m)$ and on $f_0''(m)$, if the Wilks phenomenon did not hold. We selected distributions that have very different values for these parameters, as can be seen in Table 6.1. For comparison, we also added a chi-squared distribution with 1 degree of freedom, denoted χ_1^2, to Figure 6.1. This is the limit distribution for regular parametric likelihood ratio statistics for one parameter. (Note that we did *not* plot the estimated distribution function of $2\log\lambda_n$ based on data generated from a χ_1^2 distribution; we plotted the actual χ_1^2 distribution itself.)

	Weibull(1.5, 1)	Gamma(3, 1)	Beta(2, 3)
$f_0(m)$	0.75	0.27	1.8
$f_0''(m)$	−2.4	−0.14	−24

Table 6.1: Parameters governing the limit distribution of log-concave mode estimate

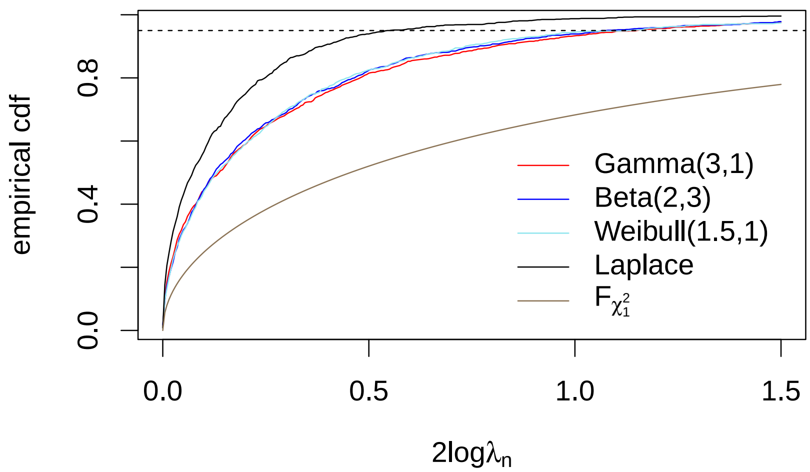


Figure 6.1: Monte Carlo'd limit distribution of the log likelihood ratio statistic and Wilks phenomenon

The apparent limit distribution for $2 \log \lambda_n$ appears to be distinct from that of the χ_1^2. This is as expected, since this is a non regular problem with non normal limit distributions.

Finally, we also plot the limit distribution of $2 \log \lambda_n$ when the true density is the Laplace density $f_0(x) = e^{-|x|}/2$. Note that this density is not twice differentiable at $m = 0$, rather, it has a cusp at $m = 0$. The lower bound we computed in Section 4.1 indicates that perhaps a different rate is achievable at a cusp than the $n^{1/5}$ lower bound that Balabdaoui et al. (2009) achieved when $f_0''(m) < 0$. If the rate is different then we would also expect that the limit distribution would be different, and we see in Figure 6.1 that the limit corresponding to the Laplace density is different than that for the densities with $f_0''(m) < 0$.

Now, using the estimated distribution for $2 \log \lambda_n$ when $f_0''(m) < 0$ given in Figure 6.1, we

	N(0,1)		Γ(3,1)	
Sample Size	50	300	100	400
Coverage Estimate	.94	.95	.94	.94

Table 6.2: Estimates of nominally 95% confidence set coverage under H_0

can find $(1-\alpha)$-quantiles $q_{1-\alpha}$ and use the procedure (5.130) for the hypothesis test (5.129), or form the confidence set given by (5.131). We have implemented these procedures, and we will now check by simulation how well they perform. Table 6.2 shows how the actual coverage, estimated via Monte Carlo, compares to the nominal (or, asymptotic) coverage. We use two different true log-concave densities, a Gamma$(3,1)$, which was used above, and a Normal$(0,1)$. For each, we use two different sample sizes n: 100 and 400 for the gamma and 50 and 300 for the normal. In all scenarios, the true coverage is within .01 of the nominal .95 coverage, even when sample size is only 100 or 50 for the gamma and normal, respectively. Note that the coverage of a confidence set that is given by inverting a test $\phi_{\alpha,n}$ is the same as the size of that test. Thus our confidence sets and tests seem to have nearly correct coverage even when sample size is not too large.

We can next check that our test $\phi_{\alpha,n}$ has power converging to 1 under a fixed alternative. We consider the test

$$H_0 : m = 2 \quad \text{versus} \quad H_1 : m \neq 2,$$

and simulate from Gamma$(2,1)$. Its mode is 1, so H_0 does not hold; see Figure 6.2 for a plot of the true density, its mode (the vertical line in black) and the hypothesis mode (the vertical line in red). Table 6.3 gives the power estimate for sample sizes $n = 100, 500, 1000,$ and 3000, and shows that the power indeed converges to 1 as sample size gets very large.

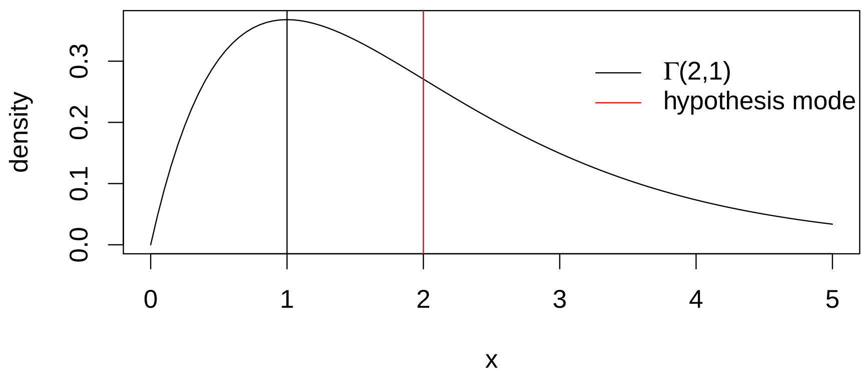


Figure 6.2: Hypothesis test under a fixed alternative

Sample Size	100	500	1000	3000
Power Estimate	.75	.95	.99	.999

Table 6.3: Estimates of power under a fixed alternative $\Gamma(2, 1)$

6.3 Real Data Results

In the previous section, we used simulations to demonstrate that we can form effective confidence sets for the mode. We will now examine a few real examples. We first consider the rotational velocities of 3933 stars from the Bright Star Catalogue which lists all stars of stellar magnitude 6.5 or brighter (which is approximately all stars visible from earth) (Hoffleit and Warren, 1991). From Owen (2001), page 8,

> Stars rotate around the center of our galaxy, with a velocity that depends in part, upon their distance from the center. The radial velocity of a star is the speed with which it appears to be moving away from us, with negative values for stars getting closer. The rotational velocity of a star is its velocity, perpendicular to the line connecting it to the sun.

A stars rotational velocity is affected by other physical quantities in a galaxy, and the distribution of rotational velocities can provide evidence for or against various models of galaxy behavior. We are thus interested in the distribution of the rotational velocities. See Binney and Merrifield (1998) for information on the movement of stars. In Figure 6.3, we present the data, along the bottom of the plot, and some density estimates. We plot a kernel density estimate with bandwidth chosen by Silverman's rule-of-thumb (Silverman, 1986, page 48) to be 14.3, the log-concave UMLE, the log-concave CMLE with mode fixed at 16, and the confidence interval for the mode given by our likelihood ratio statistic, i.e. (5.131). Using kernel density estimates when the support is bounded (especially when the density is large near the support boundary) is more complicated than otherwise, and methods that do not account for the boundary behavior do not perform well (Jones, 1993). For instance, the bandwidth given by Sheather and Jones (1991) chooses a small bandwidth to allow for a large decrease in the density at 0, and thus undersmooths the rest of the density. This is why we chose the rule-of-thumb bandwidth. The 95% confidence interval given by (5.131), and plotted in red in Figure 6.3, is $(0, 19.6)$. The interval is somewhat large for a sample size this large. This is caused by the fact that the UMLE is very flat on $[0, 16]$. This relatively flat interval is the reason we also plotted the CMLE with mode fixed at 16. The UMLE and CMLE are visually quite similar, so it is plausible that the true density indeed has a flat modal region.

Next, we consider the 1006 daily log returns for the S&P 500 stock market index from January 1, 2003 to December 29, 2006. In Figure 6.4 we plot the data (along the bottom), a kernel density estimate with bandwidth chosen by the method of Sheather and Jones (1991) to be .13, the log-concave UMLE, and the 95% confidence interval for the mode given by our likelihood ratio statistic. The log-concave mode estimate is 0.17, and the 95% confidence interval is $(0.10, 0.21)$. Since one important question is whether the distribution of the log

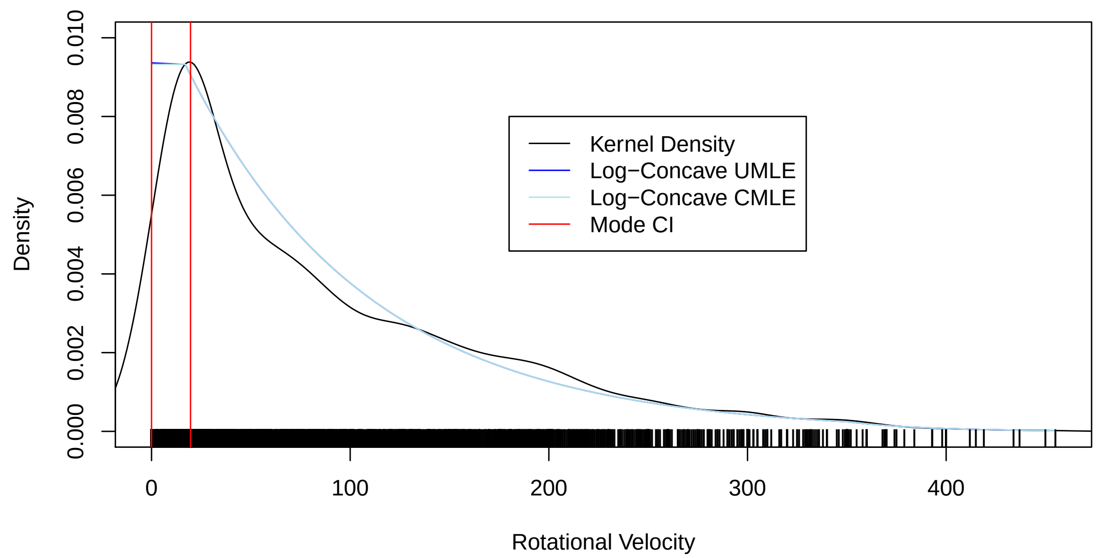


Figure 6.3: Rotational velocity of 3933 stars with stellar magnitude brighter than 6.5.

retuns is Gaussian, we also plot the maximum likelihood Gaussian density estimate, for comparison. The mean of the data (i.e. the Gaussian density MLE for the mean, median, and mode) is 0.04 with a Wald normal-approximation confidence interval of $(-0.004, 0.09)$. This latter confidence interval is plotted as two green diamonds along the bottom of the plot. Note that our confidence interval for the mode excludes 0. It also does not intersect with the confidence interval for the mean of the data. Thus, our procedure highlights some interesting features of the data and provide evidence for its non normality.

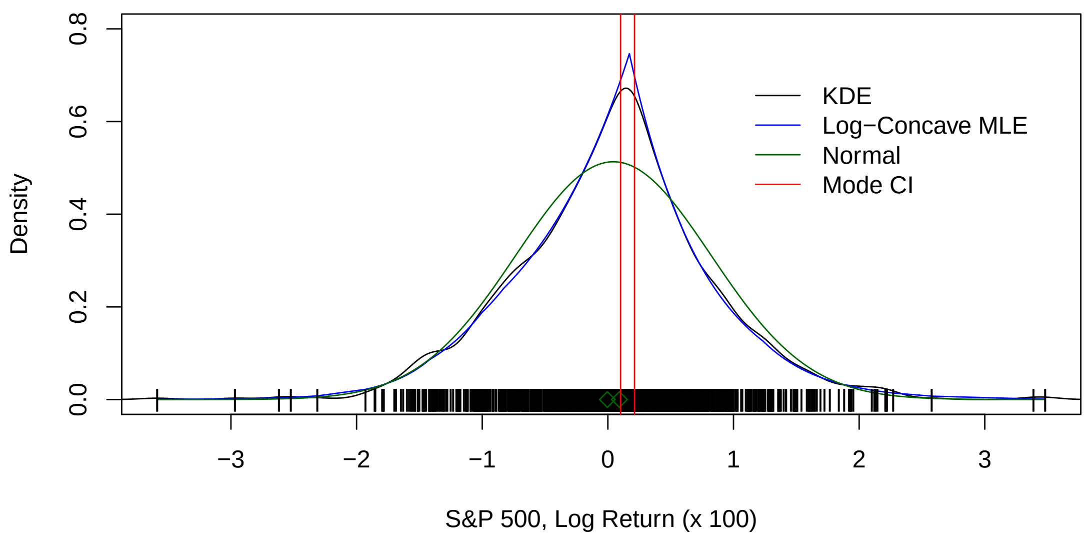


Figure 6.4: 1006 S&P 500 daily log returns for the years 2003-2006

Appendix A
CALCULATIONS AND TECHNICAL RESULTS

A.1 Calculations for the Gap Problem

Here are additional proofs and calculations needed for the proof of Theorem 4.2.3. The following lemma is similar to Lemma 4.4 of Balabdaoui et al. (2009).

Lemma A.1.1. Let Assumptions A and D hold. Define $h_l = \max(\tau^0_{n,+}(m) - a, a - \tau^0_{n,-}(m))$ and Δ_1 as in (4.25). Then we have for all $\epsilon > 0$,

$$\left| \int \Delta_1 d(\mathbb{F}_n - F_0) \right| \le \epsilon h_l^4 + M_n n^{-4/5}, \tag{A.1}$$

$$\int \Delta_1(x)(\hat{f}^0_n - f_0)(x)dx \le -\frac{f_0(0)\varphi''_0(0)}{2} K h_l^4 + o_p(h_l^4), \tag{A.2}$$

where $K < 0$ is from Lemma A.1.2, and does not depend on f, and $M_n = O_p(1)$ but its distribution depends on ϵ.

Proof. As in the proof of Theorem 4.2.3, take m to be 0, without loss of generality. We examine $\int \Delta_1(x)(\hat{f}^0_n - f_0)(x)dx$ by repeated Taylor expansions at 0, where we let Δ_1 be $\Delta_{LK,1}$ or $\Delta_{NK,1}$. The proof when Δ_1 is $\Delta_{RK,1}$ is symmetric to the case when it is $\Delta_{LK,1}$. Write $(\hat{f}^0_n - f_0) = f_0(\frac{\hat{f}^0_n}{f_0} - 1) = f_0(\exp\{\hat{\varphi}^0_n - \varphi_0\} - 1)$. Then write $d_n = \hat{\varphi}^0_n - \varphi_0$ and expand to see

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

and

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

for $t \in [\tau^0_{n,-}, \tau^0_{n,+}]$, where $\xi_{1,n,t}$ is between 0 and $d_n(t)$ and $\xi_{2,n,t}$ is between 0 and t. For the moment, define $\|\cdot\|_{n,\infty}$ to be the uniform norm over $[\tau^0_{n,-}, \tau^0_{n,+}]$, so that we can write

$$\begin{aligned} f_0(t)(e^{d_n(t)} - 1) &= \left(\sum_{i=0}^{1} \frac{(f_0)^{(i)}(0) t^i}{i!} + \frac{(f_0)^{(2)}(\xi_{2,n,t}) t^2}{2!} \right) \left(\sum_{i=1}^{1} \frac{d_n(t)^i}{i!} + e^{\xi_{1,n,t}} \frac{d_n(t)^2}{2!} \right) \\ &= f_0(0) d_n(t) + o_p(||d_n(t)||_{n,\infty}), \end{aligned} \tag{A.3}$$

since $f_0^{(i)}$ is continuous and thus bounded on a neighborhood of 0 for $0 \le i \le 2$ by Assumption D, and, by uniform consistency of $\hat{\varphi}^0_n$ (i.e. Corollary 3.1.9), $d_n(t)^i$ and t^i both go to 0 uniformly in a neighborhood of 0. So, we have written the second product in the integrand in (A.2) as the above display. Thus, we next consider $\int_{\tau^0_{n,-}}^{\tau^0_{n,+}} \Delta_1(t) d_n(t)\, dt$ and we consider the different possible forms Δ_1 may take. First we consider the case wherein 0 is an LK so $\Delta_1 = \Delta_{LK,1}$. Note that for $t \in [0, \tau^0_{n,+}]$, $\hat{\varphi}^0_n(t) = \hat{\varphi}^0_n(0)$ and for $t \in [\tau^0_{n,-}, 0]$, $\hat{\varphi}^0_n(t) = \hat{\varphi}^0_n(0) + (\hat{\varphi}^0_n)'(0-)t$. Thus

$$\int_{\tau^0_{n,-}}^{\tau^0_{n,+}} \Delta_{LK,1}(t) \hat{\varphi}^0_n(t) dt = \int_{\tau^0_{n,-}}^{0} \Delta_{LK,1}(t) \hat{\varphi}^0_n(t) dt + \int_{0}^{\tau^0_{n,+}} \Delta_{LK,1}(t) \hat{\varphi}^0_n(t) dt \tag{A.4}$$

$$= \hat{\varphi}^0_n(0) \int_{\tau^0_{n,-}}^{0} \Delta_{LK,1}(t) dt + (\hat{\varphi}^0_n)'(0-) \int_{\tau^0_{n,-}}^{0} t \Delta_{LK,1}(t) dt \tag{A.5}$$

$$+ \hat{\varphi}^0_n(0) \int_{0}^{\tau^0_{n,+}} \Delta_{LK,1}(t) dt \tag{A.6}$$

$$= 0, \tag{A.7}$$

by (4.17) and (4.18). Hence,

$$\begin{aligned} \int_{\tau^0_{n,-}}^{\tau^0_{n,+}} \Delta_{LK,1}(t) d_n(t) dt &= \int_{\tau^0_{n,-}}^{\tau^0_{n,+}} \Delta_{LK,1}(t) \hat{\varphi}^0_n(t) dt - \int_{\tau^0_{n,-}}^{\tau^0_{n,+}} \Delta_{LK,1}(t) \varphi_0(t) dt \\ &= - \int_{\tau^0_{n,-}}^{\tau^0_{n,+}} \Delta_{LK,1}(t) \varphi_0(t) dt. \end{aligned} \tag{A.8}$$

By continuity of $\varphi_0^{(2)}$ at 0 (i.e. Assumption D), we can write

$$\varphi_0(t) = \sum_{i=0}^{1} \frac{\varphi_0^{(i)}(0)}{i!} t^i + \frac{\varphi_0^{(2)}(\xi_{3,n,t})}{2} t^2 = \sum_{i=0}^{2} \frac{\varphi_0^{(i)}(0)}{i!} t^i + \frac{\epsilon_n(t)}{2} t^2, \tag{A.9}$$

for some $\epsilon_n(t)$ with $\|\epsilon_n\|_{n,\infty} = o_p(1)$, so that (A.8) is equal to

$$\begin{aligned} -\sum_{i=0}^{2} \frac{\varphi_0^{(i)}(0)}{i!} \int_{\tau_{n,-}^0}^{\tau_{n,+}^0} t^i \Delta_{LK,1}(t)dt + \int_{\tau_{n,-}^0}^{\tau_{n,+}^0} \frac{\epsilon_n(t)}{2!} t^2 \Delta_{LK,1}(t)dt \\ = -\frac{\varphi_0^{(2)}(0)}{2!} \int_{\tau_{n,-}^0}^{\tau_{n,+}^0} t^2 \Delta_{LK,1}(t)dt + \int_{\tau_{n,-}^0}^{\tau_{n,+}^0} \frac{\epsilon_n(t)}{2!} t^2 \Delta_{LK,1}(t)dt, \end{aligned} \tag{A.10}$$

by (4.17) and because $\varphi_0'(0) = 0$. (A.10) also holds with an analogous proof if we replace $\Delta_{LK,1}$ by $\Delta_{RK,1}$. Now, if 0 is an NK, then $\Delta_1 = \Delta_{NK,1}$, and $\hat{\varphi}_n^0$ and φ_0 are both twice differentiable so we can write

$$\int_{\tau_{n,-}^0}^{\tau_{n,+}^0} \Delta_1(t) d_n(t)\, dt = \sum_{i=0}^{1} \frac{d_n^{(i)}(0)}{i!} \int_{\tau_{n,-}^0}^{\tau_{n,+}^0} t^i \Delta_1(t)\, dt + \int_{\tau_{n,-}^0}^{\tau_{n,+}^0} \frac{d_n^{(2)}(\xi_{3,n,t})}{2!} t^2 \Delta_1(t)\, dt, \tag{A.11}$$

where $\xi_{3,n,t} \in [\tau_{n,-}^0, \tau_{n,+}^0]$. Note that and $(\hat{\varphi}_n^0)^{(1)}(0)$, $(\hat{\varphi}_n^0)^{(2)}(0)$, and $\varphi_0'(0)$ are all 0, and since $d_n^{(2)}$ is continuous at 0 we can write $d_n^{(2)}(\xi_{3,n,t}) = d_n^{(2)}(0) + \epsilon_n(t)$ where $||\epsilon_n(t)\|_{n,\infty} \to_p 0$ since $\tau_{n,+}^0 - \tau_{n,-}^0 \to_p 0$. Thus

$$\begin{aligned} \int_{\tau_{n,-}^0}^{\tau_{n,+}^0} \Delta_{NK,1}(t) d_n(t)dt = \sum_{i=0}^{2} \frac{d_n^{(i)}(0)}{i!} \int_{\tau_{n,-}^0}^{\tau_{n,+}^0} t^i \Delta_{NK,1}(t)dt + \int_{\tau_{n,-}^0}^{\tau_{n,+}^0} \frac{\epsilon_n(t)}{2!} t^2 \Delta_{NK,1}(t)dt \\ = -\frac{\varphi_0^{(2)}(0)}{2!} \int_{\tau_{n,-}^0}^{\tau_{n,+}^0} t^2 \Delta_{NK,1}(t)dt + \int_{\tau_{n,-}^0}^{\tau_{n,+}^0} \frac{\epsilon_n(t)}{2!} t^2 \Delta_{NK,1}(t)dt, \end{aligned} \tag{A.12}$$

by (4.17) and since $d_n^{(1)}(0) = 0$ and $d_n^{(2)}(0) = -\varphi_0^{(2)}(0)$. Thus, by (A.12) and (A.10) (and the analogous statement when Δ_1 is $\Delta_{RK,1}$), we have shown

$$\int_{\tau^0_{n,-}}^{\tau^0_{n,+}} \Delta_1(t) d_n(t) dt = -\frac{\varphi_0^{(2)}(0)}{2!} \int_{\tau^0_{n,-}}^{\tau^0_{n,+}} t^2 \Delta_1(t) dt + \int_{\tau^0_{n,-}}^{\tau^0_{n,+}} \frac{\epsilon_n(t)}{2!} t^2 \Delta_1(t) dt \tag{A.13}$$

where $\|\epsilon_n(t)\|_{n,\infty} \to_p 0$. Thus, by (A.3) and by (A.13),

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

Lemma A.1.2 shows that

$$\int_{\tau^0_{n,-}}^{\tau^0_{n,+}} t^2 \Delta_1(t) dt \le K h_l^4 \tag{A.14}$$

which yields our desired conclusion (A.2):

$$\int_{\tau^0_{n,-}}^{\tau^0_{n,+}} \Delta_1(t) \left(\hat{f}^0_n - f_0 \right)(t) dt \le -\frac{f_0(0)\varphi_0^{(2)}(0)}{2} K h_l^4 + o(h_l^4).$$

Now we show (A.1). First for a fixed $\delta > 0$, we know that $\tau^0_{n,+}$ and $\tau^0_{n,-}$ will be in $[-\delta, \delta]$ eventually, with high probability. Now we consider three families of functions, analogous to $\Delta_{LK,1}$, $\Delta_{RK,1}$, and $\Delta_{NK,1}$, respectively. For $b < 0 < c$, define $\Delta_{LK,b,c}$ by replacing $\tau^0_{n,-}$ with b and $\tau^0_{n,+}$ with c in (4.21) (and in the definitions of the quantities in that display). Define $\Delta_{NK,b,c}$ by replacing $\tau^0_{n,-}$ with b and $\tau^0_{n,+}$ with c in (4.24) (and in the definitions of the quantities in that display). For $b < 0 < -b < R$, define $\mathcal{F}_{LK,b,R} := \{\Delta_{LK,b,y} | b < 0 < y,\ 0 \le y - b \le R\}$ and $\mathcal{F}_{NK,b,R} := \{\Delta_{NK,b,y} | b < 0 < y,\ 0 \le y - b \le R\}$. Define $\mathcal{F}_{RK,b,R}$ analogously to $\mathcal{F}_{LK,b,R}$. Let $\mathcal{F}_{b,R} = \mathcal{F}_{LK,b,R} \cup \mathcal{F}_{RK,b,R} \cup \mathcal{F}_{NK,b,R}$, and note $\mathcal{F}_{b,R}$ is a VC-class (van der Vaart and Wellner, 1996) with VC-index of 4. Thus Theorem A.1.7 shows that the condition (A.15) holds for $\mathcal{F}_{b,R}$. By Lemma A.1.3, the function $F_{b,R}(x) = 1_{[b,b+R]}(x) \cdot 7/4$ is an envelope for $\mathcal{F}_{b,R}$. Next, we compute the integral of the envelope squared, which is

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

where $\|f_0\|$ is the supremum over $\mathbb{R}$ of the density f_0, and is thus universal across b and R. Thus, we can conclude from Lemma A.1.5 with $s = 2$ and $d = 2$, that for $\epsilon > 0$ there exists an $M_n = O_p(1)$ (whose distribution is dependent on ϵ) such that

$$\left| \int \Delta_1 d(\mathbb{F}_n - F_0) \right| \leq \epsilon(\tau_{n,+}^0 - \tau_{n,-}^0)^4 + M_n n^{-4/5} \leq \epsilon 2^4 h_l^4 + M_n n^{-4/5},$$

as desired. □

Lemma A.1.2. Let Assumptions A and D hold. Define $h_l = \max(\tau_{n,+}^0(m) - a, a - \tau_{n,-}^0(m))$ and Δ_1 as in (4.25). Then, for some $K < 0$,

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

Proof. Assume without loss of generality that $m = 0$. First, we consider $\Delta_{LK,1}$, and assume $\tau_{n,+}^0 > -\tau_{n,-}^0$. We break $\int_{\tau_{n,-}^0}^{\tau_{n,+}^0} t^2 \Delta_{LK,1}(t)dt$ into two pieces. Direct computation shows

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

Next we see that

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

Thus, as desired,

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

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

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

Next we see

$$\int_{\tau_{n,-}^0}^{a} t^2 \Delta_{LK,1}(t)dt = -\frac{(-\tau_{n,-}^0)^5 + 5(\tau_{n,-}^0)^4\tau_{n,+}^0}{48(5\tau_{n,+}^0 - 2\tau_{n,-}^0)} \leq -\frac{(-\tau_{n,-}^0)^5 + 5(\tau_{n,-}^0)^4\tau_{n,+}^0}{-48 \cdot 7\tau_{n,-}^0},$$

which is less than or equal to

$$-\frac{(\tau_{n,-}^0)^4 + 5(-\tau_{n,-}^0)^3\tau_{n,+}^0}{48 \cdot 7} \leq -\frac{(\tau_{n,-}^0)^4}{48 \cdot 7},$$

so that

$$\int_{\tau_{n,-}^0}^{\tau_{n,+}^0} t^2 \Delta_{LK,1}(t)dt \leq -\frac{(\tau_{n,-}^0)^4}{48 \cdot 7} = -\frac{h_l^4}{48 \cdot 7},$$

as desired. Identical calculations hold for $\Delta_{RK,1}$.

Now we examine $\Delta_{NK,1}(t)$. Direct computation shows

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

□

Lemma A.1.3. Let $\mathcal{F}_{b,R}$ be as defined on page 224. Then the function $F_{b,R}(x) = 1_{[b,b+R]}(x) \cdot 7/4$ is an envelope for the class $\mathcal{F}_{b,R}$.

Proof. We will show that $F_{b,R}$ is an envelope for the three classes $\mathcal{F}_{NK,b,R}$, $\mathcal{F}_{LK,b,R}$, and $\mathcal{F}_{RK,b,R}$. That $F_{b,R}$ is an envelope is immediate for $\Delta_{NK,b,y} \in \mathcal{F}_{NK,b,R}$. See definition of $\Delta_{NK,1}$ (and of $\Delta_{NK,0}$) on page 96 (and, recall, we have substituted b and y for $\tau^0_{n,-}$ and $\tau^0_{n,+}$) and Figure 4.1. $\Delta_{NK,b,y}$ is triangular shaped and the longer interval, i.e. the interval of length $\max(\tau^0_{n,+} - m, m - \tau^0_{n,-})$, it has largest slope, of absolute value 1. Thus the increase from its minimum to its maximum is no larger than $y - b \leq R$. Its minimum is below 0. Thus, we just need to check that the minimum value is also not larger in absolute value than $7R/4$. Its minimum has absolute value $\max(-b, y)/2 \leq R$, so we have shown that $F_{b,R}$ is an envelope for $\mathcal{F}_{RK,b,R}$.

That $F_{b,R}$ is an envelope is also immediate for the setting where $y < -b$ and $\Delta_{LK,b,y} \in \mathcal{F}_{LK,b,R}$ (and analogously when $y > -b$ and $\Delta_{RK,b,y} \in \mathcal{F}_{RK,b,R}$). Recall the definition of $\Delta_{LK,1}$ (and of $\Delta_{LK,0}$) on page 94. In this setting, $\Delta_{LK,b,y}$ takes its maximum at $b/2$, see Figure 4.1, and the interval $[b, b/2]$ has slope 1, so that the increase from minimum to maximum is bounded by $|b|/2 < R/2$. Thus, again, we just need to check that the minimum value is also not larger in absolute value than $7R/4$. In this case, by (4.23), the minimum value is not larger in absolute value than $-b < R$.

For the case $y \geq -b$ and $\Delta_{LK,b,y} \in \mathcal{F}_{LK,b,R}$, we need only note

$$0 \leq m_2 = \frac{y}{-b}\left(\frac{-9 - 3\frac{y}{-b}}{1 - 5\frac{y}{-b}}\right) \leq 3,$$

so that $\frac{-b}{4} m_2 \leq \frac{3}{4} y$. In this case, $\Delta_{LK,b,y}$ takes its maximum at $b/4$, see Figure 4.1. Then from the (second line of the) definition of $\Delta_{LK,0}$ in (4.19), the increase from minimum to maximum of $\Delta_{LK,b,y}$ (which is the same as the maximum value of $\Delta_{LK,0}$) is $y + m_2(-b)/4 \leq 7R/4$. By (4.22), the minimum value is not larger in absolute value than $3y/4 \leq 3R/4$, so $F_{b,R}$ is an envelope for $\Delta_{LK,b,y}$ in this case as well, and we are now done. $\square$

A.1.1 Empirical Process Results

Here we will rephrase Lemma A.1 from page 2560 of Balabdaoui and Wellner (2007), which is an extension of Lemma 4.1 of Kim and Pollard (1990). Here we phrase it by assuming the existence of a constant C that is universal across the $\mathcal{F}_{x,R}$ classes, as x and R vary; the original proof from Balabdaoui and Wellner (2007) proves this modified statement. This constant exists automatically for VC-classes. For the lemma we will consider families of functions of the following type.

Definition A.1.4 (R-monotone families)**.** Let $\mathcal{F}_{x,R}$ be families of functions, let $\mathcal{F} := \cup_x \mathcal{F}_{x,R}$, and let S be a map from $\mathcal{F}$ to $\{(x,y)|y \geq x\}$. If $S(f) := (S_1(f), S_2(f)) = (x,y)$, we denote this by writing $f = f_{x,y}$, and we require that $f_{x,c} \in \mathcal{F}_{x,R}$ if $x < c \leq x + R$.

Thus, for example, $\mathcal{F}_{x,R_1} \subseteq \mathcal{F}_{x,R_2}$ if $R_2 \geq R_1$. Note we do not require monotonicity in the first index, i.e. we may have $\mathcal{F}_{x+\epsilon,R-\epsilon} \subsetneq \mathcal{F}_{x_2,R}$ (although one may have such monotonicity in practice). The lemma does not require this monotonicity, even though that may seem nonintuitive, because of the uniform entropy bound across the classes.

Lemma A.1.5. Let $x_0 \in \mathbb{R}$ and $\delta > 0$. For all $x \in [x_0 - \delta, x_0 + \delta]$ and $0 < R \leq R_0 \in \mathbb{R}$ let $\mathcal{F}_{x,R}$ be R-monotone families of functions, with $\mathcal{F} := \cup_x \mathcal{F}_{x,R}$. Assume that each $\mathcal{F}_{x,R}$ admits an envelope $F_{x,R}$ such that for $0 < R \leq R_0$

$$EF_{x,R}^2(X) \leq KR^{2d-1},$$

for some $d \geq 1/2$ and K not depending on x or R. Assume further for all x, $R \leq R_0$, that

$$\sup_{x,R} \sup_{Q} \int_0^1 \sqrt{\log N(\eta \|F_{x,R}\|_{Q,2}, \mathcal{F}_{x,R}, L_2(Q))}\, d\eta < C. \tag{A.15}$$

Then, for all $\epsilon > 0$, there exist $O_p(1)$ random variables M_n such that

$$|\int fd(\mathbb{F}_n - F_0)| \leq \epsilon|S_2(f) - S_1(f)|^{s+d} + n^{-(s+d)/(2s+1)}M_n,$$

for all $f \in \mathcal{F}$ with $|S_2(f) - S_1(f)| < R_0$ and for any $s > 0$, where M_n does not depend on $S(f)$ but does depend on ϵ; that is, the right hand side depends only on f through $S_2(f) - S_1(f)$.

Proof. The proof of Lemma A.1 on page 2560 of Balabdaoui and Wellner (2007) as written actually proves this statement. We do not need for the functions to be defined by their support, except to force the nested nature of the $\mathcal{F}_{x,R}$, which is used in the equality

$$\sum_{1\leq j\leq j_n} n^{2(k+d)/(2k+1)} \frac{E\{\sup_{y:0\leq y-x<jn^{-1/(2k+1)}} |(\mathbb{P}_n - P_0)(f_{x,y})|\}^2}{(\epsilon(j-1)^{k+d} + m)^2}$$
$$= \sum_{1\leq j\leq j_n} n^{2(k+d)/(2k+1)} \frac{E\{\sup_{f_{x,y}\in\mathcal{F}_{x,jn^{-1/(2k+1)}}} |(\mathbb{P}_n - P_0)(f_{x,y})|\}^2}{(\epsilon(j-1)^{k+d} + m)^2},$$

which is the fourth in the string on page 2561. (It may seem nonintuitive that one does not need control of the support, given the "$y - x$" in the conclusion, but this comes from the control of the envelope function, which is somewhat analogous to a constraint on the support of the families.) Similarly, the next line in the proof (the fifth), which asserts

$$E\{\sup_{f_{x,y}\in\mathcal{F}_{x,jn^{-1/(2k+1)}}} |(\mathbb{P}_n - P_0)(f_{x,y})|\}^2 \leq Cn^{-1}n^{-(2d-1)/(2k+1)}j^{2d-1},$$

for some constant C that depends only on x_0 and δ, follows trivially from (A.15) (since the bound in their (A.3) depends on the entropy integral). This saves us from needing to make any argument about the entropy condition being uniformly bounded as $\|F_{x,R}\|$ changes. □

Remark A.1.6. The condition (A.15) is stronger than what is in Balabdaoui and Wellner

(2007). We do not know if it is implied by the condition in Balabdaoui and Wellner (2007). Both conditions are satisfied if all $\mathcal{F}_{x,R}$ are VC and have a uniformly bounded VC-index, because of the following theorem from van der Vaart and Wellner (1996):

Theorem A.1.7 (Theorem 2.6.7, page 141, of van der Vaart and Wellner (1996))**.** *For any VC-class $\mathcal{F}$ with VC-index $V(\mathcal{F})$ and with measurable envelope F, $r \geq 1$ and any probability measure Q with $\|F\|_{Q,r} > 0$, we have*

$$N(\epsilon\|F\|_{Q,r}, \mathcal{F}, L_r(Q)) \leq KV(\mathcal{F})(16e)^{V(\mathcal{F})} \left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{F})-1)}.$$

Appendix B

SKOROKHOD TOPOLOGIES

In spaces of continuous functions, perhaps the most standard candidate to serve as a metric is the uniform distance. However, when one needs to study spaces of discontinuous functions, as we do, the uniform distance usually is not a very good candidate, and it is often replaced by one of the Skorokhod topologies. There are a handful of different Skorokhod topologies and metrics. Chapters 11 and 12 of Whitt (2002) provide a good exposition of them, and will be the basis for our definitions here. We will use what is called the M_1 Skorokhod topology. We do not work with what is perhaps the most classical Skorokhod topology, the J_1 Skorokhod topology, for reasons that are explained in Remark B.0.17 below. The graphic on page 384 of Whitt (2002) may be useful in gaining some intuition about the differences between J_1, M_1, and the other Skorokhod topologies. Note that if the image space of our functions were $\mathbb{R}^k$ with $k > 1$, then we would need to choose between the strong M_1 (SM_1) toplogy and the weak M_1 (WM_1) topology, but when $k = 1$ they coincide. (This is clear from the fact that their definitions only differ in the usage of the sets $[a, b]$ rather than $[[a, b]]$, where these two sets are defined on page 394 of Whitt (2002), and are identical when $k = 1$.) This is only of concern to us in that we will choose to use the terminology of the SM_1 topology, but could have used the terminology for WM_1. Also, since we are not keeping track of a multitude of topologies, we could simplify the notation from Whitt (2002), but we will choose rather to keep the identical notation in order to make reference between our work and Whitt (2002) simple.

The functions for which we need to do this extra topological work are the derivatives of our convex functions, i.e. the (monotonically increasing) third derivatives of our processes. We would like to show they are tight, i.e. with high probability they are in a compact subspace of $\mathcal{D}_c$. One candidate to do this is the compactness Theorem 12.12.2 on page 425 of Whitt (2002). However, this theorem requires that we have uniform control at the endpoints of the interval, $\pm c$. We do not have any such uniform control yet, so we have two possible alternative approaches. We could show that, with high probability, the functions we are interested in will lie in a slightly smaller class of functions constrained so that they are constant in a small uniform neighborhood of $\pm c$. Rather than doing that, we will pursue the second alternative, and relax the compactness theorem slightly to ignore the behavior at $\pm c$, since we are letting c go to ∞ anyway. The method to do this is to study functions defined on open sets rather than on closed sets.

Definition B.0.8. For an interval $I \subseteq \mathbb{R}$, which may have one or both endpoints finite or infinite and either endpoint may be closed or open, we define

$$\mathcal{D}_I := \{h | h : I \to \mathbb{R}, h \text{ is cadlag and bounded}\}, \tag{B.1}$$

where “cadlag” means left-continuous and with limits to the right. Letting S be one of the topologies J_1, J_2, M_1, M_2, we will define S_I as the topological space

$$S_I := (\mathcal{D}_I, S), \tag{B.2}$$

where the topologies are defined on a general D_I in Section 12.9 of Whitt (2002). The class of functions in which we are interested is:

$$\mathcal{F}_{I,M} := \{f : I \to \mathbb{R} | f \text{ is non-decreasing, and } \|f\| \leq M\}, \tag{B.3}$$

where, recall, $\|f\|$ is the supremeum of f over its domain.

We will need the $\mathcal{D}_I$ notation in this section, but outside of it we will only refer to $\mathcal{D}_c$, wherein I is taken to be $(-c, c)$. Note that since the weak convergence of the empirical processes to a Brownian bridge occurs with the J_1 topology on D_c, this convergence also holds for the weaker M_1 topology. Rather than using the M_1 topology for our empirical processes, we could use the J_1 topology there and just use the M_1 for the third derivatives of our processes. We will choose the former, but solely for notational convenience; it will not affect our results, since both topologies are strong enough for the results we need.

We will not define the metric inducing the M_1 topology on $\mathcal{D}_{[-c,c]}$ here, since the definition is somewhat lengthy; rather, we refer the reader to Whitt (2002). Discussion of this topology can be found in chapter 12 section 3, and the definition of the metric is in equation (3.7) there. We are more interested in extending results from the M_1 topology on $\mathcal{D}_{[-c,c]}$ to the M_1 topology on $\mathcal{D}_{(-c,c)}$. Thus, we will define the latter topology, in terms of the former. As in Whitt (2002) chapter 12 section 9, we say that for x_n and x in $\mathcal{D}_{(-c,c)}$, $x_n \to x$ in $M_{1,(-c,c)}$ if $x_n \to x$ in $M_{1,[\alpha,\beta]}$ for all α and β that are interior points of $(-c, c)$ and continuity points of x. We are not done yet, though, because to use standard weak convergence theory on metric spaces we indeed need to know that $M_{1,(-c,c)}$ with the above definition is a metric space. Luckily, we have the following result from Whitt (1980).

Theorem B.0.9 (Thm 2.6, Whitt (1980))**.** *Let $I \subseteq \mathbb{R}$ be an interval where either endpoint can be finite or infinite and, if an endpoint is finite, it can be open or closed, and if infinite, it is open. Then the space $M_{1,I}$ is metrizable as a complete and separable metric space.*

Note that the theorem as first stated in Whitt (1980) on page 73 is for the J_1 metric, but part of the content of section 6 of that paper is that the theorem holds for the M_1 metric as well (see page 80). Now, knowing that we are working in a metric space, we can continue with standard probability theory on metric spaces, including showing (pre-)compactness via sequential compactness. First, we define the modulus of continuity for the M_1 topology. (Actually in Whitt (2002) it is defined for the SM_1 topology, hence the "s"-subscript.)

Definition B.0.10 (pg 381, ch 11, (5.3), Whitt (2002))**.** For a set $A \subseteq \mathbb{R}$, by $\|x - A\|$ we mean

$$\|x - A\| := \inf_{y \in A} \|x - y\|. \tag{B.4}$$

Note that we have also taken $\|f\|$ to be the supremum of f over its domain. It will be clear from context, i.e. from whether the argument is a function or a set, which usage of $\|\cdot\|$ is intended.

Definition B.0.11 (pg 402, Whitt (2002))**.** Let

$$w_s(x, \delta) = \sup \|x(t_2) - [x(t_1), x(t_3)]\|,$$

where the sup is taken over t_1, t_2, and t_3 such that $-c \vee (t_2 - \delta) \le t_1 < t_2 < t_3 \le c \wedge (t_2 + \delta)$. Note that w_s coincides with the definition of Δ_{M_1} in Skorokhod (1956).

Now, to get to sequential compactness, we will extend Lemmas 2.3.5 and 2.4.1 of Skorokhod (1956). These lemmas are as follows.

Lemma B.0.12 (Lemma 2.3.5, Skorokhod (1956))**.** Assume that $\lim_{n\to\infty} x_n(t)$ exists for $t \in N$, where N is a dense subset of the open interval $I = [a, b]$ and has $a, b \in N$. Define $x(t) := \lim_{n\to\infty} x_n(t)$, but note that we do not make assumptions on x (i.e. we do not assume anything about $x(t)$ for $t \notin N$, and so, e.g., cannot assume $x \in \mathcal{D}_I$). If we also assume that

$$\lim_{\delta\to 0}\limsup_{n\to\infty} w_s(x_n,\delta)=0, \tag{B.5}$$

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\sup_{b-\delta<t<b}|x_n(b)-x_n(t)|=0, \tag{B.6}$$

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\sup_{a<t<a+\delta}|x_n(t)-x_n(a)|=0, \tag{B.7}$$

then we can conclude that there exists $\bar{x}\in\mathcal{D}_I$ such that $x_n\to\bar{x}$ in $M_{2,I}$. Specifically,

$$\bar{x}(t)=\lim_{t^*\searrow t,t^*\in N} x(t^*) \text{ for } t\in[a,b) \tag{B.8}$$

$$\bar{x}(b)=\lim_{t^*\nearrow b,t^*\in N} x(t^*). \tag{B.9}$$

Note that the above lemma yields a result for the M_2 topology, not the M_1, but it will be converted later to a result about the M_1 topology.

Lemma B.0.13 (Lemma 2.4.1, Skorokhod (1956))**.** If we assume that x_n and x are in $\mathcal{D}_I$, then $x_n\to x$ in $M_{1,I}$ if and only if $x(t)=\lim_{n\to\infty}x_n(t)$ for all t in a dense subset N of a closed interval I with a and b in N, and

$$\lim_{\delta\to 0}\limsup_{n\to\infty} w_s(x_n,\delta)=0. \tag{B.10}$$

The above lemmas imply the following rephrased version of Lemma B.0.12.

Lemma B.0.14. Assume that $\lim_{n\to\infty}x_n(t)$ exists for $t\in N$, where N is a dense subset of the open interval $I=(a,b)$. If we also assume that

$$\lim_{\delta\to 0}\limsup_{n\to\infty} w_s(x_n,\delta)=0, \tag{B.11}$$

then we can conclude that there exists $\bar{x}\in\mathcal{D}_I$ such that $x_n\to\bar{x}$ in $M_{1,I}$.

In going from Lemma 2.3.5 of Skorokhod (1956) (i.e. Lemma B.0.12) to Lemma B.0.14, we changed I to an open interval from a closed one, dropped the endpoint conditions (B.6) and (B.7), and also used the M_1 topology instead of the M_2 topology.

Proof of Lemma B.0.14. First we define $\bar{x}$. For any α, t, β such that $a < \alpha < t < \beta < b$, and via Lemma B.0.12 applied to the closed interval $[\alpha, \beta]$, we know we can define $\bar{x}$ as in (B.8). This allows us to define $\bar{x}(t)$ for any $t \in (a, b)$. Since we only apply the definition to t that are interior to $[\alpha, \beta]$ we do not need to verify the endpoint conditions (B.6) and (B.7), which have no bearing on such t.

Now the definition of x_n converging to $\bar{x}$ in $M_{1,(a,b)}$ is that the restrictions of x_n converge to the restriction of $\bar{x}$ in $M_{1,[\alpha,\beta]}$ for all $a < \alpha < \beta < b$ such that α and β are continuity points of $\bar{x}$. Since (B.11) trivially implies (B.10) and we have shown $\bar{x}$ exists and is a cadlag function, we can conclude that $x_n \to \bar{x}$ in $M_{1,[\alpha,\beta]}$ and thus $x_n \to x$ in $M_{1,(a,b)}$ as desired. □

Next, we extend 2.72 on page 278 of Skorokhod (1956) to say:

Lemma B.0.15. Let $A \subset \mathcal{D}_I$ where $I = (a, b)$. Then A is pre-compact in $M_{1,I}$ if

$$\sup_{x \in A} \|x\| < \infty, \tag{B.12}$$

$$\lim_{\delta \to 0} \sup_{x \in A} w_s(x, \delta) = 0. \tag{B.13}$$

Proof. The only content here is in using (B.12) to conclude that $\lim_{n\to\infty} x_n(t)$ exists for a dense subset of I, exactly as in the proof of 2.72 on page 278 of Skorokhod (1956). We can take a countable dense subset N of I, and for each $t \in N$, we can pick a subsequence of the (bounded) sequence of real numbers $x_n(t)$ such that the subsequence has a limit, by (B.12). Using Cantor diagonalization, we can pick a single subsequence n_k such that $x_{n_k}(t)$ has a limit for all $t \in N$. This, together with (B.13) are the hypotheses for Lemma B.0.14, so we can conclude that there exists $\bar{x} \in \mathcal{D}_I$ such that $x_n \to \bar{x}$ in $M_{1,I}$. Since for each sequence x_n in A we found a subsequence with a limit in $\mathcal{D}_I$, we have shown A is precompact. □

Finally, we conclude with the precompactness of the class in which we are actually interested.

Lemma B.0.16. For $I = (-c, c)$ and $M < \infty$, define $\mathcal{F}_{I,M}$ by (B.3). We can conclude that $\mathcal{F}_{I,M}$ is precompact in $M_{1,I}$.

Proof. By the definition of $\mathcal{F}_{I,M}$ with $M < \infty$, (B.12) is satisfied. The second condition, (B.13) is satisfied for classes of increasing functions on a bounded domain, as is stated in the proof of Corollary 12.5.1 in Whitt (2002). Thus $\mathcal{F}_{I,M}$ is precompact in $M_{1,M}$ as desired. □

Remark B.0.17 ($\mathcal{F}_{[-c,c]}$ is not relatively compact in J_1). Taking $\mathcal{F}_{[-c,c]} \subset \mathcal{D}_c$ as above, and taking the J_1 Skorokhod topology defined by the metric d° from equation (12.16) on page 125 of Billingsley (1999) or by the metric d from equation (12.13) of page 124 (which yield the same topology), then $\mathcal{F}_{[-c,c]}$ is not relatively compact in $\mathcal{D}_c$. This Skorokhod topology is referred to as the J_1 toplogy in Whitt (2002). This non relative compactness follows from Theorem 12.4 on page 132 of Billingsley (1999) which characterizes relative compactness in this topology. Specifically, if we define

$$w''(f, \delta) := \sup_{t_1 < t_2 < t_3} \min \{|f(t_3) - f(t_2)|, f(t_2) - f(t_1)|\},$$

where $t_3 - t_1 \leq \delta$, as in the theorem from Billingsley (1999), then for all $\delta > 0$ it is easy to construct $f_\delta \in \mathcal{F}_{[-c,c]}$ such that $w''(f_\delta, \delta) \geq M$, so that $\lim_{\delta \to 0} \sup_{f \in \mathcal{F}} w''(f, \delta)$ does not converge to 0. Having this latter convergence to 0 is a necessary condition for relative compactness, so we will have proved that $\mathcal{F}_{[-c,c]}$ is not relatively compact. The f_δ are constructed to be continuous approximations to a jump at, say, 0: we set f_δ to be constant and equal to $-M$ on $[-c, -\delta/2]$, to be equal to M on $[\delta/2, c]$, and to linearly interpolate between $-M$ and M on $[-\delta/2, \delta/2]$. Then taking $t_3 = \delta/2$, $t_2 = 0$, and $t_1 = -\delta/2$, we see $w''(f_\delta, \delta) \geq M$, as claimed.

We can see that the J_1 topology does not accomodate the approximation of jump functions by continuous functions. While the increasing functions in which we are interested are piecewise constant jump functions, the class of such functions is not closed (in any Skorokhod topology). This class is also not even compact in the J_1 topology; see the bottom left of the diagram on page 384 of Whitt (2002), which indicates how multiple jumps can not approximate a single jump in the J_1 topology. Thus, the J_1 topology is not useful for us in this context. The M_1 topology does allow multiple jumps (in monotone functions) to approximate a single jump. The M topologies allow continuous functions to approximate discontinuous ones, which we may not need, so it is possible that the J_2 topology would be the weakest topology (least number of open sets, largest distances) that we could use, but the M_1 will also suffice.

Appendix C

CODE

```
getLR <- function(rdist=rnorm,aval=0, N.MC=1e2, n.SS=1e4,
                  prec=10^-10,
                  seedVal=NULL,
                  debugging=NULL){
  if (!is.null(seedVal)) set.seed(seedVal)
  if (is.character(debugging)) {

    sink(paste(debugging, "_log.txt", sep=""))
    currTime <- prevTime <- 0
    print(paste("Parameters are: N.MC=", N.MC,
                "// n.SS=", n.SS, "// aval=", aval,
                " and you can load rdist from the rsav file, ",
                debugging, ".rsav.",
                sep=""))
  }
  LRs <- TLLRs <- vector(length=N.MC)

  for (i in 1:N.MC){
    myxx <- sort(rdist(n.SS))

    myxx.uniq <- rle(myxx)
    if (is.character(debugging)) {
      prevTime <- currTime
      currTime <- Sys.time()
      save(file=paste(debugging, ".rsav", sep=""),
           myxx,myxx.uniq,rdist)
      print(paste("Starting the ", i, "th iteration.",
                  "Memory usage is ", sum(gc()[1:2,2]),
                  " Printing the time for the last iteration ",
                  "and the current time.",
                  sep=""))
      print(currTime-prevTime)
      print(currTime);
    }
    myxx <- myxx.uniq$values
    myww <- myxx.uniq$lengths / n.SS
    res.UC <- activeSetLogCon(x=myxx,w=myww, prec=prec,print=F)
    res.MC <- activeSetLogCon.mode(x=myxx,aval=aval,w=myww,
                                   prec =prec,
                                   print=F)
```

```
    if (res.MC$dlcMode$isx) {
      print("getLR Warning: res.MC$dlcMode$idx is true. This is not standard.")
      TLLRs[i] <- 2*(sum(res.UC$phi - res.MC$phi))
    }
    else {TLLRs[i] <- 2*(sum(res.UC$phi - res.MC$phi[-res.MC$dlcMode$idx]))}
    if (is.character(debugging)){
      if (TLLRs[i] <= 0 || TLLRs[i] > 10e4){
        print("getLR ERROR: 2log-LR returned value <=0 or > 10e4");
        tmpfile <- paste(debugging,"tmpxxs", i, ".rsav", sep="");
        save(myxx,myww,aval, file=tmpfile)
        print(paste("Saved myxxs to ", getwd(),  tmpfile, sep=""))
      }
    }

  }
  if (is.character(debugging)) sink()

  return(list(LRs=exp(TLLRs/2), TLLRs=TLLRs))
}
J10 <- function (x, y)
{
    m <- length(x)
    z <- exp(x)
    d <- y - x
    LARGE <- (abs(d) > 0.01)
    SMALL <- !LARGE
    II <- (1:m)[LARGE]
    z[II] <-  (exp(y[II]) - z[II] - d[II]*z[II])/(d[II]^2)
    II <- (1:m)[SMALL]
    z[II] <- z[II] * (1/2 + d[II] * (1/6 + d[II] * (1/24 + d[II] *
        (1/120 + d[II]/720))))
    return(z)
}
J20 <- function (x, y)
{
    m <- length(x)
    z <- exp(x)
    d <- y - x
    LARGE <- (abs(d)>0.02)
    SMALL <- !LARGE
    II <- (1:m)[LARGE]
    z[II] <- 2* (exp(y[II]) - z[II] - z[II]*d[II] - z[II]*d[II]^2/2 ) / (d[II]^3)
    II <- (1:m)[SMALL]
    z[II] <- z[II] * (1/3 + d[II] * (1/12 + d[II] * (1/60 + d[II] *
        (1/360 + d[II]/2520))))
    return(z)
}
J11 <-   function (x, y)
{
```

```
  m <- length(x)
  z <- exp(x)
  d <- y - x
  LARGE <- (abs(d)>.02)
  SMALL <- !LARGE
  II <- (1:m)[LARGE]
  z[II] <- (d[II]*(exp(y[II])+z[II]) - 2*(exp(y[II])-z[II])) / (d[II]^3)

  II <- (1:m)[SMALL]
  z[II] <- z[II] * (1/6 + d[II] * (1/12 + d[II] * (1/40 + d[II] *
                                                  (1/180 + d[II]/1008))))
  return(z)
}
J00 <- function (x, y, v = 1)
{
    m <- length(x)
    z <- exp(x)
    d <- y - x
    ed <- exp(v*d)
    LARGE <- (abs(d)>0.005)
    SMALL <- !LARGE
    II <- (1:m)[LARGE]
    z[II] <- (exp(v*y[II]+(1-v)*x[II]) - z[II]) / d[II]
    II <- (1:m)[SMALL]
    z[II] <- z[II] * (v + d[II] * (v/2 + d[II] * (v/6 + d[II] *
        (v/24 + d[II] * v/120))))
    return(z)
}
Local_LL_all <- function (x, w, phi)
{
  n <- length(x)
  dx <- diff(x)
  ll <- sum(w * phi) - sum(dx * J00(phi[1:(n - 1)], phi[2:n]))
  LIs <- (1:n)[exp(phi)==Inf]
  if (length(LIs) > 0 ) stop("Local_LL_all Error: We have not yet
 accounted for extraordinarily steep/large phi.  This is probably an error.")
  grad <- matrix(w, ncol = 1)
  grad[1:(n - 1)] <- grad[1:(n - 1)] - (dx * J10(phi[1:(n -  1)], phi[2:n]))
  grad[2:n] <- grad[2:n] - (dx * J10(phi[2:n], phi[1:(n - 1)]))
  tmp <- c(dx * J20(phi[1:(n - 1)], phi[2:n]), 0) +
    c(0, dx *  J20(phi[2:n], phi[1:(n - 1)]))
  tmp <- tmp + mean(tmp) * 1e-12
  mhess2 <- matrix(0, nrow = n, ncol = n)
  mhess3 <- mhess2
  mhess1 <- tmp
  tmp <- c(0, dx * J11(phi[1:(n - 1)], phi[2:n]))
  tmp.up <- diag(tmp[2:n], nrow = n - 1, ncol = n - 1)
  mhess2[1:(n - 1), 2:n] <- tmp.up
  mhess3[2:n, 1:(n - 1)] <- diag(tmp[2:n], nrow = n - 1, ncol = n -  1)
  mhess <- diag(mhess1) + mhess2 + mhess3
```

```
  phi_new <- phi + solve(mhess) %*% grad
  dirderiv <- t(grad) %*% (phi_new - phi)
  return(list(ll = ll, phi_new = phi_new, dirderiv = dirderiv))
}
LocalExtend <- function (x, IsKnot, x2, phi2, constr=NULL) {
    n <- length(x)
    K <- (1:n) * IsKnot
    K <- K[K > 0]
    phi <- 1:n * 0
    phi[K] <- phi2
    for (k in 1:(length(K) - 1)) {
        if (K[k + 1] > (K[k] + 1)) {
            ind <- (K[k] + 1):(K[k + 1] - 1)
            lambda <- (x[ind] - x2[k])/(x2[k + 1] - x2[k])
            phi[ind] <- (1 - lambda) * phi2[k] + lambda * phi2[k +
                1]
        }
      }
    if (length(constr)>1 && constr[1] != constr[2]){
      ind <- K[constr[1]]:K[constr[2]]
      phi[ind] <- rep(phi2[constr[1]], length=length(ind))
    }
    return(matrix(phi, ncol = 1))
}
intF <- function (s, x, phi, Fhat,
                  prec=1e-10)
{

  n <- length(x)
  lows <- s<x[1]
  upps <- s>x[n]
  dx <- c(NA, diff(x))
  dphi <- c(NA, diff(phi))
  f <- exp(phi)
  F <- Fhat
  intF.xi <- c(0, rep(NA, n - 1))
  for (i in 2:n) {
    if (abs(dphi[i]) < prec){
    intF.xi[i] <- dx[i] * (F[i-1] + f[i-1]*dx[i]/2)
    }
    else{
      intF.xi[i] <- dx[i] * (F[i - 1] + dx[i]/dphi[i] *
                             (J00(phi[i - 1], phi[i], 1) - f[i - 1]))
    }
  }
  intF.xi <- cumsum(intF.xi)
  intF.s <- rep(0, length(s))
  for (k in 1:length(s)) {
    if (!lows[k]){
      extra <- 0
```

```
      if (!upps[k]) mys <- s[k]
      else {
        extra <- s[k]-x[n]
        mys <- x[n]
      }
      j <- max((1:n)[x <= mys])
      j <- min(j, n - 1)
      Fj <- F[j]
      ds <- (mys-x[j])
      if (abs(dphi[j+1]) < prec){
        intF.s[k] <- intF.xi[j] +  ds * (Fj  + ds*f[j+1]/2)
      }
      else{
        intF.s[k] <- intF.xi[j] + ds * Fj + dx[j+1] *
          (dx[j + 1]/dphi[j + 1] *
           J00(phi[j], phi[j + 1], ds/(dx[j + 1])) -
           ds/(dphi[j + 1]) * f[j])
      }
      intF.s[k] <- intF.s[k]+extra
    }
  }


  return(intF.s)
}
intECDF <- function (s, x)
{
  n <- length(x)
  lows <- s<x[1]
  upps <- s>x[n]
  n <- length(x)
  x <- sort(x)
  dx <- c(0, diff(x))
  intED.xi <- cumsum(dx * ((0:(n - 1))/n))
  intED.s <- rep(0, length(s))
  for (k in 1:length(s)) {
    if ( !lows[k]){
      mys <- s[k]
      if (upps[k]) mys <- x[n]
      j <- max((1:n)[x <= mys])
      j <- min(j, n - 1)
      xj <- x[j]
      intED.s[k] <- intED.xi[j] + (mys - xj) * j/n
    }
  }

  intED.s[upps] <- intED.s[upps]+s[upps]-x[n]
  return(intED.s)
}
activeSetLogCon <- function(x, xgrid=NULL, w = NA, prec=10^-10, print = FALSE, logfile=NULL)
```

```
{
  if (!is.null(logfile) && !is.na(logfile)) sink(logfile)
  xn <- sort(x)
  if ((!identical(xgrid, NULL) & (!identical(w, NA)))) {
    stop("If w != NA then xgrid must be NULL!\n")
  }
  if (identical(w,NA)){
    tmp <- preProcess(x,xgrid=xgrid)
    x <- tmp$x
    w <- tmp$w
    sig <- tmp$sig
  }
  else {
    if (abs(sum(w) - 1) > prec) stop("activeSetLogCon Error: weights w do not sum to 1.")
    tmp <- cbind(x, w)
    tmp <- tmp[order(x), ]
    x <- tmp[, 1]
    w <- tmp[, 2]
    est.m <- sum(w * x)
    est.sd <- sum(w * (x - est.m)^2)
    est.sd <- sqrt(est.sd * length(x)/(length(x) - 1))
    sig <- est.sd
  }
  n <- length(x)
  phi <- LocalNormalize(x, 1:n * 0)
  IsKnot <- 1:n * 0
  IsKnot[c(1, n)] <- 1
  res1 <- LocalMLE(x, w, IsKnot, phi, prec)
  phi <- res1$phi
  L <- res1$L
  conv <- res1$conv
  H <- res1$H
  iter1 <- 1
  while ((iter1 < 500) & (max(H) > prec * mean(abs(H)))) {
    IsKnot_old <- IsKnot
    iter1 <- iter1 + 1
    tmp <- max(H)
    k <- (1:n) * (H == tmp)
    k <- min(k[k > 0])
    IsKnot[k] <- 1
    res2 <- LocalMLE(x, w, IsKnot, phi, prec)
    phi_new <- res2$phi
    L <- res2$L
    conv_new <- res2$conv
    H <- res2$H
    while ((max(conv_new) > prec * max(abs(conv_new)))) {
      JJ <- (1:n) * (conv_new > 0)
      JJ <- JJ[JJ > 0]
      tmp <- conv[JJ]/(conv[JJ] - conv_new[JJ])
      lambda <- min(tmp)
```

```
    KK <- (1:length(JJ)) * (tmp == lambda)
    KK <- KK[KK > 0]
    IsKnot[JJ[KK]] <- 0
    phi <- (1 - lambda) * phi + lambda * phi_new
    conv <- pmin(c(LocalConvexity(x, phi), 0))
    res3 <- LocalMLE(x, w, IsKnot, phi, prec)
    phi_new <- res3$phi
    L <- res3$L
    conv_new <- res3$conv
    H <- res3$H
    if (print == TRUE) {
      print(paste("iter1=", iter1, " / L=", round(L, 4),
                  " / max(H)=", round(max(H), 4),
                  " / #knots = ", sum(IsKnot),
                  sep = ""))
    }
  }
  phi <- phi_new
  conv <- conv_new
  if (sum(IsKnot != IsKnot_old) == 0) {
    break
  }
  if (print == TRUE) {
    print(paste("iter1=", iter1, " / L=", round(L, 4),
                " / max(H)=", round(max(H), 4),
                " / #knots = ", sum(IsKnot),
                sep = ""))
  }
}

KK <- (1:n)[as.logical(IsKnot)]
Fhat <- LocalF(x, phi)
phia <- max(phi)
aidx <- which(phi==phia)
aval <- x[aidx]
dlcMode <- list(val=aval,idx=aidx, isx=TRUE)
class(dlcMode) <- "dlc.mode"

res1 <- list(xn=xn,
             x=x,
             w=w,
             IsKnot = IsKnot,
             L = L,
             knots=x[IsKnot==1],
             phi = as.vector(phi),
             fhat=as.vector(exp(phi)),
             Fhat = as.vector(Fhat),
             H = as.vector(H),

             n=length(xn),
```

```
               m=n,
               mode=aval,


               dlcMode=dlcMode,
               sig=sig)
  phi.f <- function(x0){



    evaluateLogConDens(x0,res=res1,which=1)[,2]
  }
  fhat.f <- function(x0){


    evaluateLogConDens(x0,res=res1,which=2)[,3]
  }
  Fhat.f <- function(x0){


    evaluateLogConDens(x0,res=res1,which=3)[,4]
  }
  E.f <- intFfn(x,phi,Fhat)
  {
    phiK <- phi[KK]
    slopes <- diff(phiK)/diff(x[KK])
    phiPR.f <- stepfun(x=x[KK], c(slopes[1],slopes,tail(slopes,1)), right=FALSE,f=0)
    phiPL.f <- stepfun(x=x[KK], c(slopes[1],slopes,tail(slopes,1)), right=TRUE,f=1)
    phiPL <- as.vector(phiPL.f(x))
    phiPR <- as.vector(phiPR.f(x))
  }
  if (!is.null(logfile) && !is.na(logfile)) sink(NULL)
  return(c(res1,
           list(phi.f=phi.f, fhat.f=fhat.f, Fhat.f=Fhat.f, E.f=E.f,
           phiPL=phiPL,phiPR=phiPR,
           phiPL.f=phiPL.f,phiPR.f=phiPR.f)))
}
logConDens <- function (x, xgrid = NULL, smoothed = TRUE, print = FALSE, gam = NULL,
    xs = NULL, prec=1e-10, logfile=NULL)
{
    res1 <- activeSetLogCon(x, xgrid = xgrid, print = print, prec=prec, logfile=logfile)
    if (identical(smoothed, FALSE)) {
        res <- c(res1)
    }
    if (identical(smoothed, TRUE)) {
        if (identical(xs, NULL)) {
            r <- diff(range(x))
            xs <- seq(min(x) - 0.1 * r, max(x) + 0.1 * r, length = 500)
        }
        smo <- evaluateLogConDens(xs, res1, which = 4:5, gam = gam,
```

```
            print = print)
        f.smoothed <- smo[, "smooth.density"]
        F.smoothed <- smo[, "smooth.CDF"]
        mode <- xs[f.smoothed == max(f.smoothed)]
        res2 <- list(f.smoothed = f.smoothed, F.smoothed = F.smoothed,
            gam = gam, xs = xs, mode = mode)
        res <- c(res1, res2)
    }
    res$smoothed <- smoothed
    class(res) <- "dlc"
    return(res)
}
LCLRCImode <- function(x,

                       w=NA,
                       alpha=.05,
                       prec=1e-10,
                       CIprec=1e-4,
                       print=F){
  x <- sort(x)
  nn <- length(x)
  myLRmodeTest <- function(mm){LRmodeTest(mode=mm, x=x,w=w,alpha=alpha,prec=prec,print=print)}
  MLE.UC <- activeSetLogCon(x=x,w=w,

                            prec=prec,
                            print=print)
  mhat <- MLE.UC$dlcMode$val
  if (!myLRmodeTest(mhat)) return(numeric(0))
  if (myLRmodeTest(x[1])) L <- x[1]
  else{
    L <- x[1]
    R <- mhat
    while(R-L > CIprec){
      mid <- (R+L)/2
      if (myLRmodeTest(mid)) R <- mid
      else L <- mid
    }
  }
  Lpt <- L

  if (myLRmodeTest(x[nn])) R <- x[nn]
  else{
    L <- mhat
    R <- x[nn]
    while (R-L > CIprec){
      mid <- (R+L)/2
      if (myLRmodeTest(mid)) L <- mid
      else R <- mid
    }
  }
```

```
  Rpt <- R
  return(c(Lpt,Rpt))
}
LRmodeTest <- function(mode, x,w,alpha,prec,print){
  res.UC <- activeSetLogCon(x=x,w=w,

                            prec=prec,print=print)
  res.MC <- activeSetLogCon.mode(x=x,

                                 w=w,
                                 aval=mode,
                                 prec =prec,
                                 print=print)


  if (res.MC$dlcMode$isx) LL <- 2*(sum(res.UC$phi - res.MC$phi))
  else  LL <- 2*(sum(res.UC$phi - res.MC$phi[-res.MC$dlcMode$idx]))
  CC <- LCTLLRdistn@q(1-alpha)
  if (LL <= CC) TRUE
  else FALSE
}

LocalConvexity.mode <- function (z, phi, k=NULL){

  n <- length(z)
  dphi <- diff(phi)
  dz <- diff(z)
  deriv <- dphi/dz
  conv <- rep(0,n);
  conv[2:(n - 1)] <- diff(deriv[1:(n - 1)])
  conv <- matrix(conv,ncol=1);


  if (k>=2 && k<n){

    mono <- c(-dphi[k-1], dphi[k])
  }
  else if (k==1){

    mono <- c(0,dphi[1])
  }
  else if (k==n){
    mono <- c(dphi[n-1],0)
  }
  shape <- list(conv=conv,mono=matrix(mono,ncol=1))
  return(shape)
}
LocalMLE.mode <- function (z, w, IsKnot, IsMIC, a, phi_o, prec, print=FALSE)
{
  n <- length(z)
```

```
k <- a$idx;
r1 <- LocalCoarsen.mode(z, w, IsKnot, IsMIC,a);
IsKnot <- syncIsKnot(IsKnot,IsMIC,k);

K <- (1:n)[IsKnot>0];

res2 <- MLE.mode(r1$y,r1$constr, r1$w2, phi_o[K], print=print)

phi <- LocalExtend(z, IsKnot, r1$y, res2$phi,
                   r1$constr)
shape <- LocalConvexity.mode(z, phi,k=k)
shape$conv <- shape$conv * IsKnot
shape$mono <- shape$mono * IsMIC
H <- rep(0,n)
HR <- rep(0,n)
H.m <- matrix(nrow=2,ncol=1)


JJ <- (1:n) * IsKnot

JJ <- JJ[JJ > 0]
p <- sum(JJ<k)
p <- p+1
m <- length(JJ)
if (is.null(a)){
  for (i in 1:(m - 1)) {
    if (JJ[i + 1] > JJ[i] + 1) {
      dtmp <- z[JJ[i + 1]] - z[JJ[i]]
      ind <- (JJ[i] + 1):(JJ[i + 1] - 1)
      mtmp <- length(ind)
      ztmp <- (z[ind] - z[JJ[i]])
      dstmp <- c(z[JJ[i]+1]-z[JJ[i]],diff(z[ind]),
                 z[JJ[i+1]]-z[JJ[i+1]-1])
      wtmp <- w[ind]
      J01s <- J10(phi[ind],phi[ind-1]) * dstmp[1:mtmp]


      J10s <- J10(phi[ind],phi[ind+1]) * dstmp[2:(mtmp+1)]
      H[ind] <- cumsum(wtmp * ztmp) - ztmp * cumsum(wtmp) +
        ztmp * sum(wtmp * (1-ztmp/dtmp))
      jtmp <- - ztmp*cumsum(J01s+J10s) + cumsum(ztmp * (J01s + J10s)) +
        sum((J01s+J10s)*(1-ztmp/dtmp))*ztmp
      H[ind] <- H[ind] - jtmp
      H[IsKnot] <- 0;
    }
  }
}
else{

  if (k!=1) {
```

```
    for (i in 1:(p-1)) {
      if (JJ[i + 1] > JJ[i] + 1) {

        cond1 <- r1$constr[1]!=r1$constr[2] && r1$constr[1]==1 && i==1
        cond2 <- r1$constr[1]!=r1$constr[2] && r1$constr[1]==i+1
&& JJ[i+1]>JJ[i]+1 && JJ[i+2]>JJ[i+1]+1
        if (cond1){
          LK <- 1
          RK <- JJ[i+1]
        }
        else if (cond2){
          LK <- JJ[i+1];
          RK <- JJ[i+2];
        }
        else{
          LK <- JJ[i+1]
          RK <- JJ[i+1]
        }
        ind <- (JJ[i] + 1):(RK - 1)
        mtmp <- length(ind)
        ztmp <- (z[ind] - z[JJ[i]])
        dstmp <- c(z[JJ[i]+1]-z[JJ[i]],diff(z[ind]),
                   z[RK]-z[RK-1])
        wtmp <- w[ind]
        J01s <- J10(phi[ind],phi[ind-1]) * dstmp[1:mtmp]
        J10s <- J10(phi[ind],phi[ind+1]) * dstmp[2:(mtmp+1)]
        H[ind] <- cumsum(wtmp * ztmp) - ztmp * cumsum(wtmp)
        jtmp <- - ztmp*cumsum(J01s+J10s) + cumsum(ztmp * (J01s + J10s))
        if (cond1){
          H0 <- -ztmp*w[1]
          J0 <- -ztmp*(J10(phi[1],phi[2])*(z[2]-z[1]))
        }
        else if (cond2){

          dtmp <- z[LK] - z[JJ[i]]
          tmptmp <- (JJ[i] + 1):(LK - 1)
          ind.lin <- (1:length(tmptmp))
          H0 <- ztmp*sum(wtmp[ind.lin] * (1-ztmp[ind.lin]/dtmp))
          J0 <- ztmp*sum((J01s[ind.lin]+J10s[ind.lin])*(1-ztmp[ind.lin]/dtmp))


        }
        else{
          dtmp <- z[LK] - z[JJ[i]]
          H0 <- ztmp * sum(wtmp * (1-ztmp/dtmp))
          J0 <- sum((J01s+J10s)*(1-ztmp/dtmp))*ztmp
        }
        H[ind] <- H[ind] + H0  - jtmp - J0
        H[IsKnot] <- 0;
        if (cond2)    break
```

```
        }
      }
      p2 <- p-1

    }
    else{
      p2 <- 1
    }
    for (i in p2:(m-1)) {
      if (JJ[i + 1] > JJ[i] + 1) {

        cond1 <- r1$constr[1]!=r1$constr[2] && r1$constr[1]==m-1 && i==m-1
        cond2 <- r1$constr[1]!=r1$constr[2] && r1$constr[1]==i
&& JJ[i+1]>JJ[i]+1 && JJ[i+2]>JJ[i+1]

        if (cond1){
          LK <- JJ[i+1]
          RK <- JJ[i+1]
        }
        else if (cond2){
          LK <- JJ[i+1];
          RK <- JJ[i+2];
        }
        else{
          LK <- JJ[i+1]
          RK <- JJ[i+1]
        }

        ind <- (JJ[i] + 1):(RK - 1)
        mtmp <- length(ind)


        ztmp <- rev(z[RK] - z[ind])
        dstmp <- c(z[JJ[i]+1]-z[JJ[i]],diff(z[ind]),
                   z[RK]-z[RK-1])
        wtmp <- rev(w[ind])
        J01s <- rev(J10(phi[ind],phi[ind-1]) * dstmp[1:mtmp])
        J10s <- rev(J10(phi[ind],phi[ind+1]) * dstmp[2:(mtmp+1)])
        HR[ind] <- rev(cumsum(wtmp * ztmp) - ztmp * cumsum(wtmp) )
        jtmp <- rev( -ztmp*cumsum(J01s+J10s) + cumsum(ztmp * (J01s + J10s)) )
        if (cond1){
          HR0 <- -rev(ztmp*w[n])
          JR0 <- -rev(ztmp*(J10(phi[n],phi[n-1])*(z[n]-z[n-1])))

        }
        else if (cond2){

          dtmp <- z[RK] - z[LK]

          lentmp <- (RK-1) - (LK+1) +1
```

```
          ind.lin <- 1:lentmp
          HR0 <- rev(ztmp*sum(wtmp[ind.lin] * (1-ztmp[ind.lin]/dtmp)))
          JR0 <- rev(ztmp*sum((J01s[ind.lin]+J10s[ind.lin])*(1-ztmp[ind.lin]/dtmp)))


        }
        else{
          dtmp <- z[RK] - z[JJ[i]]
          HR0 <- rev( ztmp * sum(wtmp * (1-ztmp/dtmp)) )
          JR0 <- rev( sum((J01s+J10s)*(1-ztmp/dtmp))*ztmp )
        }
        HR[ind] <- HR[ind] + HR0  - jtmp - JR0
        HR[IsKnot] <- 0;
      }
    }
    if (k==n){
      H.m[1] <- H[k]
      H.m[2] <- 0

    }
    else if (k==1){
      H.m[1] <- 0
      H.m[2] <- HR[k]
      H <- HR
    }
    else{
      H.m[1] <- H[k]
      H.m[2] <- HR[k]
      H[k:(n-1)] <- HR[k:(n-1)]
    }

    H[k] <- H[n] <- H[1] <- 0
  }
  res <- list(phi = matrix(phi, ncol = 1), L = res2$L,
              shape=shape, H = matrix(H, ncol = 1),H.m=matrix(H.m,ncol=1))
  return(res)
}
getConstraintVecs <- function(z, a=NULL){
  dz <- diff(z);
  n <- length(z)
  invdz <- 1/dz;
  V <- matrix(0,nrow=n,ncol=n)
  for (i in 1:n){
    for (j in 2:(n-1)){
      if (i==j-1) V[i,j] <- invdz[i]
      else if (i==j) V[i,j] <- -(invdz[j-1]+invdz[j])
      else if (i==j+1) V[i,j] <- invdz[j]
    }
```

```
  }
  if (is.null(a)){

    return(list(V1=V, V2=V,W=NULL))
  }
  else{
    k <- a$idx;
    W <- matrix(0,nrow=n,ncol=2);
    W[k,1] <- W[k,2] <- -1;
    W[k-1,1] <- W[k+1,2] <- 1;
    if (k==1) W[,1] <- rep(0,n)
    else if (k==n) W[,2] <- rep(0,n)
    return(list(V1=V[,1:(k-1)],V2=V[,(k+1):n],W=W))
  }
}
getBasisVecs <- function(z){
  n <- length(z)
  B <- -matrix(z,nrow=n,ncol=n,byrow=FALSE) +
    matrix(z,nrow=n,ncol=n,byrow=TRUE)
  A <- -B
  B[,1] <- rep(1,n)
  B <- apply(B, c(1,2), function(x){max(x,0)})
  A <- apply(A, c(1,2), function(x){max(x,0)})
  A[,n] <- rep(1,n)
  return(list(B=B,A=A))


}
MLE.mode <- function (y,constr, w = NA, phi_o = NA, prec = 10^(-7), print = FALSE)
{
    n <- length(y)
    if (sum(y[2:n] <= y[1:n - 1]) > 0) {
        cat("We need strictly increasing numbers y(i)!\n")
        stop("Exiting because of bad ys")
    }
    if (max(is.na(w)) == 1) {
        w <- rep(1/n, n)
    }
    if (sum(w < 0) > 0) {
      cat("We need nonnegative weights w(i) !\n")
      stop("exiting because of bad ws")
    }
    ww <- w/sum(w)
    if (max(is.na(phi_o)) == 1) {
        m <- sum(ww * y)
        s2 <- sum(ww * (y - m)^2)
        phi <- LocalNormalize(y, -(y - m)^2/(2 * s2))
    }
    else {
        phi <- LocalNormalize(y, phi_o)
```

```
}
iter0 <- 0
r1 <- LocalLLall.mode(y, ww, phi,constr)
L <- r1$ll
phi_new <- r1$phi.new
dirderiv <- r1$dirderiv

while ((dirderiv >= prec) && (iter0 < 100)) {
    iter0 <- iter0 + 1
    L_new <- Local_LL(y, ww, phi_new)
    iter1 <- 0

    if (print == TRUE) {
      print(paste("mle.mode:outer1: iter0=", iter0, " / iter1=", iter1,
                  " / L_new=", round(L_new, 4),
                  " / L=", round(L,4),
                  " / L_new < L=", L_new<L,
                  sep = ""))
    }
    while ((L_new < L) && (iter1 < 20)) {



        iter1 <- iter1 + 1
        phi_new <- 0.5 * (phi + phi_new)
        L_new <- Local_LL(y, ww, phi_new)
        dirderiv <- 0.5 * dirderiv
        if (print == TRUE) {
            print(paste("mle.mode:inner1: iter0=", iter0, " / iter1=", iter1,
              " / L_new=", round(L_new, 4), " / dirderiv=", round(dirderiv,
                4), sep = ""))
        }
    }





    if (L_new >= L) {
        tstar <- max((L_new - L)/dirderiv)
        if (tstar >= 0.5) {
            phi <- LocalNormalize(y, phi_new)
        }
        else {
            tstar <- max(0.5/(1 - tstar))
            phi <- LocalNormalize(y, (1 - tstar) * phi +
              tstar * phi_new)
        }
        r1 <- LocalLLall.mode(y, ww, phi,constr)
        L <- r1$ll
        phi_new <- r1$phi.new
```

```
            dirderiv <- r1$dirderiv
        }
        else {
            dirderiv <- 0
        }
        if (print == TRUE) {
            print(paste("mle.mode:outer2: iter0=", iter0, " / iter1=", iter1,
                " / L=", round(L, 4), " / dirderiv=", round(dirderiv,
                  4), sep = ""))
        }
    }
    r1 <- list(phi = matrix(phi, ncol = 1), L = L,
               Fhat = matrix(LocalF(y,  phi), ncol = 1))
    return(r1)
}
getm <- function(x,phi,a){
  print("getm: careful using this function: m is called p now.")
  k <- a$idx;
  phim <- max(phi[k-1],phi[k])
  m <- k-2 + which(phim==phi[(k-1):k]);
}
insert <- function(val,vec,idx){
  n <- length(vec)
  if (idx<=1)
    return(c(val,vec))
  else if (idx>=n+1)
    return(c(vec,val))
 else
    return(c(vec[1:(idx-1)],val,vec[idx:n]));
}
delete <- function(vec,idcs){
  keep <- rep(TRUE,length(vec))
  keep[idcs] <- FALSE
  vec[keep];
}
FK2CK <- function(y,w,phi,constr){
  ll <- length(y);
  o <- length(constr)
  if (o<=1 || constr[1]<1 || constr[o] > ll || constr[1]==constr[2]){
    res <- list(y=y,w=w,phi=phi,constr=constr)
  }
  else if (o == 2 && constr[1]==constr[2]-1){
    m <- min(constr);
    wghtsum <- sum(w[m:(m+1)])
    w <- delete(w,m);
    w[m] <- wghtsum;
    v <- delete(y,m);
    phi <- delete(phi,m);
    res <- list(y=v, w=w,phi=phi,constr=constr)
  }
```

```
  else {
    print(paste("FK2CK ... : bad constraint vec passed. ",
                "len should be 2.  constr is"));
    print(constr);
  }
  return(res);
}
LocalCoarsen.mode <- function (z, w, IsKnot,IsMIC,a)
{
  n <- length(z)

  idcs <- seq(from=1,by=1,length=a$idx-1)
  KL <- idcs * IsKnot[idcs]
  KL <- KL[KL>0]
  p <- length(KL)+1
  constr <- rep(p,length=2)
  aIsKnot <- NULL;
  if (identical(IsMIC,c(0,0))){
    aIsKnot <- FALSE
    if (a$idx==1) constr <- c(p,p+1)
    else constr <- c(p-1,p)
  }
  else{

    aIsKnot <- TRUE
    if (a$idx != 1) constr[1] <- c(p-1,p)[IsMIC[1]+1]
    if (a$idx != n) constr[2] <- c(p+1,p)[IsMIC[2]+1]
  }

  KR <- ((a$idx):n) * IsKnot[(a$idx):n]
  KR <- KR[KR>0]
  K <- c(KL,KR)
  x2 <- z[K]
  w2 <- w[K]
  for (k in 1:(length(K) - 1)){
    if (K[k + 1] > (K[k] + 1)){
      ind <- (K[k] + 1):(K[k + 1] - 1)
      lambda <- (z[ind] - x2[k])/(x2[k + 1] - x2[k])
      w2[k] <- w2[k] + sum(w[ind] * (1 - lambda))
      w2[k + 1] <- w2[k + 1] + sum(w[ind] * lambda)
    }
  }
  w2 <- w2/sum(w2)
  return(list(y = matrix(x2, ncol = 1), w2 = matrix(w2, ncol = 1),
             constr=constr,  p=p, aIsKnot=aIsKnot))
}
syncIsKnot <- function(IsKnot,IsMIC,k){
  if (max(is.na(IsMIC))==1)
    print("syncIsKnot:IsMIC should have negative or NULL if k==1 or n, not NA")
  if (k==1) IsKnot[k] <- 1
```

```
  else if (k==length(IsKnot)) IsKnot[k] <- 1
  else if (max(IsMIC)==1) IsKnot[k] <- 1
  else IsKnot[k] <- 0
  return(IsKnot)
}
getGrad <- function(y,w,phi, constr=0){
  ll <- length(y);
  o <- length(constr)
  if (o<=1 || constr[1]<1 || constr[o] > ll || constr[1]==constr[2]){
    grad <- matrix(0,ncol=1,nrow=ll);
    grad <- t(getGrad.uncstr(x=y,w=w,phi=phi));
  }
  else if (o==2 && constr[1]==constr[2]-1){

    om1 <- o-1
    grad <- matrix(0,ncol=1,nrow=ll-om1);
    m <- min(constr);
    if (m==1){
      endseq <- seq(from=m+o, by=1, length=ll-m-o+1)
      phi <- c(rep(phi[m],o), phi[endseq]);

    }
    else if (m==ll-1)
      {phi <- c(phi[1:(m-1)], rep(phi[m],o));}
    else
      {phi <- c(phi[1:(m-1)], rep(phi[m],o), phi[(m+o):ll]);}
    xtraRow <- rep(0,ll-1); xtraRow[m] <- 1;
    projmat <- diag(rep(1,ll-1))
    if (m<ll-1) {projmat <- rbind(projmat[1:m,],xtraRow,projmat[(m+1):(ll-1),])}
    else {projmat <- rbind(projmat[1:m,],xtraRow)}
    grad.tmp <- getGrad.uncstr(y,w,phi);
    grad <- t(grad.tmp) %*% projmat
  }
  else {
    print(paste("getGrad ... : bad constraint vec passed. ",
                "len should be 2.  constr is"));
    print(constr);
  }
  grad;
}
reducePhi <- function(phi,constr){
  ll <- length(phi)
  o <- length(constr)
  if (o<=1 || constr[1]<1 || constr[o] > ll || constr[1]==constr[2]){
    return(phi)
  }
  else if (o==2 && constr[1]==constr[2]-1){
    m <- min(constr);
    phi <- delete(phi,m+1)
  }
```

```
  else {
    print(paste("reducePhi ... : bad constraint vec passed. ",
                "len should be 2.  constr is"));
    print(constr);
  }
  return(phi)
}
unreducePhi <- function(phi.red,constr){
  ll <- length(phi.red)
  o <- length(constr)
  if (o<=1 || constr[1]<1 || constr[1]==constr[2] || constr[o] > ll+1){
    return(phi.red)
  }
  else if (o==2 && constr[1]==constr[2]-1){
    m <- min(constr);
    if (m==1){
      endseq <- seq(from=m+o-1, by=1, length=ll+1-m-o+1)
      phi.red <- c(rep(phi.red[m],o), phi.red[endseq]);
    }
    else if (m==ll)
      {phi.red <- c(phi.red[1:(m-1)], rep(phi.red[m],o));}
    else
      {phi.red <- c(phi.red[1:(m-1)], rep(phi.red[m],o), phi.red[(m+1):ll]);}
  }
  else {
    print(paste("unreducePhi ... : bad constraint vec passed. ",
                "len should be 2.  constr is"));
    print(constr);
  }
  return(phi.red)
}
getHess <- function(y,w,phi,constr=0){
  ll <- length(y);
  o <- length(constr)
  if (o<1 || constr[1]<1 || constr[o] > ll || constr[1]==constr[2]){
    hess <- matrix(0,ncol=ll,nrow=ll);
    hess <- getHess.uncstr(x=y,w=w,phi=phi);
  }
  else if (o==2 && constr[1]==constr[2]-1){
    om1 <- o-1
    hess <- matrix(0,ncol=ll-om1,nrow=ll-om1);
    m <- min(constr);


    if (m==1) {
      endseq <- seq(from=m+o, by=1, length=ll-m-o+1)
      phi <- c(rep(phi[m],o), phi[endseq]);
    }
    else if (m==ll-1)
      {phi <- c(phi[1:(m-1)], rep(phi[m],o));}
```

```
    else
      {phi <- c(phi[1:(m-1)], rep(phi[m],o), phi[(m+o):ll]);}



    xtraRow <- rep(0,ll-1); xtraRow[m] <- 1;
    projmat <- diag(rep(1,ll-1))
    if (m<ll-1) {projmat <- rbind(projmat[1:m,],xtraRow,projmat[(m+1):(ll-1),])}
    else {projmat <- rbind(projmat[1:m,],xtraRow)}
    hess.tmp <- getHess.uncstr(y,w,phi);
    hess <- t(projmat) %*% hess.tmp %*% projmat;
  }
  else {
    print(paste("getHess ... : bad constraint vec passed. ",
                "len should be 2.  constr is"));
    print(constr);
  }
  hess;
}
Local_LL.mode <- function(y,w,phi,constr=0){
  ll <- length(y);
  o <- length(constr)
  if (o<1 || constr[1]<1 || constr[o] > ll || constr[1]==constr[2]){
    return(Local_LL(y,w,phi));
  }
  else if (o == 2 && constr[1]==constr[2]-1){


    m <- min(constr);
    phi[m+1] <- phi[m]
    return(Local_LL(y,w,phi));
  }
  else {
    print(paste("Local_LL.mode ... : bad constraint vec passed. ",
                "len should be 2.  constr is"));
    print(constr);
  }
  -1;
}
getGrad.uncstr <- function(x,w,phi){
  grad <- matrix(w, ncol = 1)
  n <- length(phi);
  dx <- diff(x)
  J10s <- J10(phi[1:(n - 1)], phi[2:n])
  J01s <- J10(phi[2:n], phi[1:(n - 1)])
  grad[1:(n - 1)] <- grad[1:(n - 1)] - (dx * J10s)
  grad[2:n] <- grad[2:n] - (dx * J01s);
  grad
}
LocalLLall.mode <- function(y,w,phi,constr=NULL){
```

```
  ll <- Local_LL.mode(y=y,w=w,phi=phi,constr=constr);
  LIs <- (1:length(phi))[exp(phi)==Inf]
  if (length(LIs) > 0 ) stop("LocalLLall.mode Error: We have not yet
 accounted for extraordinarily steep/large phi.  This is probably an error.")
  grad <- getGrad(y=y,w=w,phi=phi,constr=constr)
  hess <- getHess(y=y,w=w,phi=phi,constr=constr)

  phi.red <- reducePhi(phi,constr)
  phi.red.new <- phi.red - solve(hess) %*% t(grad)
  dirderiv <- grad %*% (phi.red.new - phi.red)
  phi.new <- unreducePhi(phi.red.new,constr)
  return(list(ll=ll, phi.new=phi.new, dirderiv=dirderiv));
}
getHess.uncstr <- function(x,w,phi){
  dx <- diff(x);
  n <- length(x);
  tmp <- c(dx * J20(phi[1:(n - 1)], phi[2:n]), 0) +
    c(0, dx * J20(phi[2:n], phi[1:(n - 1)]))
  tmp <- tmp + mean(tmp) * 10^(-12)
  mhess2 <- matrix(0, nrow = n, ncol = n)
  mhess3 <- mhess2
  mhess1 <- tmp
  tmp <- c(0, dx * J11(phi[1:(n - 1)], phi[2:n]))
  tmp.up <- diag(tmp[2:n], nrow = n - 1, ncol = n - 1)
  mhess2[1:(n - 1), 2:n] <- tmp.up
  mhess3[2:n, 1:(n - 1)] <- diag(tmp[2:n], nrow=n-1, ncol=n-1)
  mhess <- diag(mhess1) + mhess2 + mhess3;

  -mhess;
}
getLambda <- function(shape.new,shape,IsKnot,IsMIC, k){
  conv.new <- shape.new$conv; conv <- shape$conv
  mono.new <- shape.new$mono; mono <- shape$mono
  n <- length(conv);
  JJ1 <- (1:n) * (conv.new > 0)
  JJ1 <- JJ1[JJ1 > 0]
  JJ2 <- (1:2) * (mono.new>0)
  JJ2 <- JJ2[JJ2>0]
  tmp1 <- conv[JJ1]/(conv[JJ1] - conv.new[JJ1])
  tmp2 <- mono[JJ2]/(mono[JJ2] - mono.new[JJ2])
  lambda <- min(c(tmp1,tmp2))
  if (!is.null(IsMIC) && !is.null(k)){
    KK1 <- (1:length(JJ1)) * (tmp1 == lambda)
    KK1 <- KK1[KK1 > 0]
    IsKnot[JJ1[KK1]] <- 0
    KK2 <- (1:length(JJ2)) * (tmp2 ==lambda)
    KK2 <- KK2[KK2>0]
    IsMIC[JJ2[KK2]] <- 0
    if (k==1) IsMIC[1] <- -1
    else if (k==n) IsMIC[2] <- -1
```

```
    IsKnot <- syncIsKnot(IsKnot,IsMIC,k)
  }
  else{


    KK <- (1:length(JJ1)) * (tmp1 == lambda)
    KK <- KK[KK > 0]


    IsKnot[JJ1[KK]] <- 0
  }
  return(list(lambda=lambda, IsKnot=IsKnot, IsMIC=IsMIC));
}
inactivate <- function(H,H.m, IsKnot, IsMIC, k){
  tmp <- max(c(H,H.m))
  j1 <- (1:length(H)) * (H == tmp)
  j2 <- (1:2) * (H.m==tmp)
  j1 <- j1[j1>0]
  j2 <- j2[j2>0]


  if ( length(j2) > 0 )
    IsMIC[j2[1]] <- 1
  else
    IsKnot[j1[1]] <- 1
  IsKnot <- syncIsKnot(IsKnot,IsMIC,k=k)
  return(list(IsKnot=IsKnot, IsMIC=IsMIC))
}
activeSetLogCon.mode <- function (x,xgrid=NULL,
                                  aval=x[1], w = NA, print = FALSE,
                                  logfile=NULL,
                                  prec=10^-10) {

  if (!is.null(logfile) && !is.na(logfile)) sink(logfile)
  xn <- sort(x)
  if ((!identical(xgrid, NULL) & (!identical(w, NA)))) {
    stop("If w != NA then xgrid must be NULL!\n")
  }
  if (identical(w,NA)){
    tmp <- preProcess(x,xgrid=xgrid)
    x <- tmp$x
    w <- tmp$w
    sig <- tmp$sig
  }
  else {
    if (abs(sum(w) - 1) > prec) stop("activeSetLogCon.mode Error: weights w do not sum to 1.")
    tmp <- cbind(x, w)
    tmp <- tmp[order(x), ]
    x <- tmp[, 1]
    w <- tmp[, 2]
```

```
    est.m <- sum(w * x)
    est.sd <- sum(w * (x - est.m)^2)
    est.sd <- sqrt(est.sd * length(x)/(length(x) - 1))
    sig <- est.sd
  }
  n1 <- length(x)
  phi <- LocalNormalize(x, 1:n1 * 0)
  r1 <- x2z(x=x,w=w,phi=phi,aval=aval);
  z <- r1$z; w <- r1$w.a;  phi <- r1$phi; a <- r1$a
  n <- length(z);
  IsKnot <- 1:n * 0
  IsMIC <- c(0,0)

  IsKnot[c(1, n)] <- 1
  IsKnot <- syncIsKnot(IsKnot,IsMIC,a$idx);

  res1 <- LocalMLE.mode(z, w, IsKnot, IsMIC, a=a, phi_o=phi, prec, print=print)
  phi <- res1$phi
  L <- res1$L
  shape <- res1$shape
  H <- res1$H
  H.m <- res1$H.m
  iter1 <- 1



  while ((iter1 < 500) &&
         ((max(H) > prec * mean(abs(H))) || (max(H.m) > prec * mean(abs(H.m))))) {
    IsKnot_old <- IsKnot
    IsMIC_old <- IsMIC
    iter1 <- iter1 + 1
    IC <- inactivate(H,H.m,IsKnot,IsMIC, k=a$idx)
    IsKnot <- IC$IsKnot;
    IsMIC <- IC$IsMIC

    res2 <- LocalMLE.mode(z, w, IsKnot,IsMIC, a=a, phi_o=phi, prec,print=print)
    phi_new <- res2$phi
    L <- res2$L
    shape_new <- res2$shape
    H <- res2$H
    H.m <- res2$H.m
    if (print == TRUE) {
      Kidx <- (1:n)[as.logical(IsKnot-IsKnot_old)]


      print(paste("ASLCM:Proc2: ", "iter1=", iter1, " / L=", round(L, 4),
                  " / max(H)=", round(max(H), 4),
                  " / max(Hm)=", round(max(H.m), 4),
                  " / IsKnot idx=", Kidx,
```

```
              sep = ""))
  save(H,H.m, file="activesetlogconmode.print.rsav");

}
while (max(shape_new$conv) > prec * max(abs(shape_new$conv)) ||
       max(shape_new$mono) > prec * max(abs(shape_new$mono))   ){
  IsKnot_old2 <- IsKnot
  IsMIC_old2 <- IsMIC
  s1 <- getLambda(shape_new, shape, IsKnot, IsMIC, k=a$idx)
  lambda <- s1$lambda;
  IsKnot <- s1$IsKnot;
  IsMIC <- s1$IsMIC;
  phi <- (1 - lambda) * phi + lambda * phi_new
  shape <- LocalConvexity.mode(z=z, phi, k=a$idx)
  shape$conv <- pmax(shape$conv,0)
  shape$mono <- pmax(shape$mono,0)

  res3 <- LocalMLE.mode(z, w, IsKnot,IsMIC, a=a,phi, prec,print=print)
  phi_new <- res3$phi
  L <- res3$L
  shape_new <- res3$shape
  H <- res3$H
  H.m <- res3$H.m
  H <- H * (1-IsKnot)
  H.m <- H.m * (1-IsMIC)
  if (print == TRUE) {
    Kidx <- (1:n)[as.logical(IsKnot-IsKnot_old2)]

    print(paste("ASLCM:Proc1: ","iter1=", iter1, " / L=", round(L, 4),
                " / max(H)=", round(max(H), 4),
                " / max(Hm)=", round(max(H.m), 4),
                " / max(convnew)=", round(max(shape_new$conv), 4),
                " / max(mononew)=", round(max(shape_new$mono), 4),
                " / IsKnot idx=", Kidx,

                sep = ""))
    print(paste("IsMic")); print(IsMIC)
  }
}


phi <- phi_new
shape <- shape_new;

if (sum(IsKnot != IsKnot_old) == 0 && sum(IsMIC!=IsMIC_old)==0) {
  if (print==TRUE){
    print("No change in constraints. Ending.")
  }
  break
}
```

```
}
Fhat <- LocalF(z, phi)
tmp <- LocalCoarsen.mode(z,w,IsKnot,IsMIC,a)
MI <- z[IsKnot>0][tmp$constr]
KK <- (1:n)[as.logical(IsKnot)]

  res1 <- list(xn=xn,
             x=x,
             z=z,
             w=w,
             L=L,
             MI=MI,
             IsKnot=IsKnot,
             IsMIC=IsMIC,
             constr=tmp$constr,
             knots= z[KK],
             phi=as.vector(phi),
             fhat=as.vector(exp(phi)),
             Fhat=as.vector(Fhat),
             H=as.vector(H),
             H.m=as.vector(H.m),
             n=length(xn),
             m=length(z),
             m1=n1,

             dlcMode=a,
             sig=sig);
  res1.tmp <- list(xn=xn,
                   x=z,
             z=z,
             w=w,
             L=L,
             MI=MI,
             IsKnot=IsKnot,
             IsMIC=IsMIC,
             constr=tmp$constr,
             knots= z[KK],
             phi=as.vector(phi),
             fhat=as.vector(exp(phi)),
             Fhat=as.vector(Fhat),
             H=as.vector(H),
             H.m=as.vector(H.m),
             n=length(xn),
             m=length(z),
             m1=n1,

             dlcMode=a,
             sig=sig);

class(res1) <- "dlc";
```

```
  class(res1.tmp) <- "dlc";
  phi.f <- function(x0){

    evaluateLogConDens(xs=x0, res=res1.tmp, which=1)[,2]
  }
  fhat.f <- function(x0){

    evaluateLogConDens(xs=x0, res=res1.tmp, which=2)[,3]
  }
  Fhat.f <- function(x0){

    evaluateLogConDens(xs=x0, res=res1.tmp, which=3)[,4]
  }
  E.f <- intFfn(z,phi,Fhat)


  HL.f <- E.f
  HR.f <- intFfn(z,phi,Fhat,side="right")

  {
    phiK <- phi[as.logical(IsKnot)]
    slopes <- diff(phiK)/diff(z[KK])
    phiPL.f <- stepfun(x=z[KK], c(slopes[1],slopes,tail(slopes,1)), right=TRUE,f=1)
    phiPR.f <- stepfun(x=z[KK], c(slopes[1],slopes,tail(slopes,1)), right=FALSE,f=0)
    phiPL <- phiPL.f(z)
    phiPR <- phiPR.f(z)



  }
  if (!is.null(logfile) && !is.na(logfile)) sink(NULL)
  return(c(res1,
           list(phi.f=phi.f, fhat.f=fhat.f, Fhat.f=Fhat.f,
                E.f=E.f, HL.f=HL.f,HR.f=HR.f,
                phiPL=phiPL,phiPR=phiPR,
                phiPL.f=phiPL.f,phiPR.f=phiPR.f)))
}
```

BIBLIOGRAPHY

Andriano, K., Gentle, J., and Sposito, V. (1978), "Comparison of some estimators of the mode," in *Proceedings of the Social Statistcs Section of the American Statistics Association*, pp. 760–764.

Anevski, D. (1994), "Estimating the derivative of a convex density," Tech. Rep. 8, University of Lund.

— (2003), "Estimating the derivative of a convex density," *Statistica Neerlandica*, 57, 245–257.

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T., and Silverman, E. (1955), "An Empirical Distribution Function for Sampling with Incomplete Information," *Annals of Mathematical Statistics*, 26, 641–647.

Bagnoli, M. and Bergstrom, T. (2005), "Log-concave Probability and its Applications," *Economic Theory*, 26, 445–469.

Balabdaoui, F., Rufibach, K., and Wellner, J. A. (2009), "Limit distribution theory for maximum likelihood estimation of a log-concave density," *The Annals of Statistics*, 37, 1299–1331.

Balabdaoui, F. and Wellner, J. (2010), "Estimation of a k-monotone density: characterizations, consistency and minimax lower bounds," *Statistica Neerlandica*, 64, 45–70.

Balabdaoui, F. and Wellner, J. A. (2007), "Estimation of a k-monotone density: limit distribution theory and the spline connection," *The Annals of Statistics*, 35, 2536–2564.

Banerjee, M. and Wellner, J. (2001), "Likelihood ratio tests for monotone functions," *Annals of statistics*, 29, 1699–1731.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D. (1972), *Statistical inference under order restrictions. The theory and application of isotonic regression*, John Wiley & Sons, London-New York-Sydney, wiley Series in Probability and Mathematical Statistics.

Bickel, D. R. and Frühwirth, R. (2006), "On a fast, robust estimator of the mode: comparisons to other robust estimators with applications," *Computational Statistics & Data Analysis*, 50, 3500–3530.

Billingsley, P. (1999), *Convergence of probability measures*, Wiley Series in Probability and Statistics: Probability and Statistics, New York: John Wiley & Sons Inc., 2nd ed., a Wiley-Interscience Publication.

Binney, J. J. and Merrifield, M. (1998), *Galactic Astronomy*, Princeton, NJ: Princeton University Press.

Birgé, L. (1997), "Estimation of unimodal densities without smoothness assumptions," *The Annals of Statistics*, 25, 970–981.

Birgé, L. and Massart, P. (1993), "Rates of convergence for minimum contrast estimators," *Probab. Theory Related Fields*, 97, 113–150.

Borell, C. (1975), "Convex set functions in d-space," *Periodica Mathematica Hungarica. Journal of the János Bolyai Mathematical Society*, 6, 111–136.

Brascamp, H. J. and Lieb, E. H. (1976), "On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation," *Journal of Functional Analysis*, 22, 366–389.

Bronšteĭn, E. M. (1976), "ε-entropy of convex sets and functions," *Sibirsk. Mat. Ž.*, 17, 508–514, 715.

Brunk, H. (1958), "On the estimation of parameters restricted by inequalities," *The Annals of Mathematical Statistics*, 29, 437–454.

Brunk, H. D. (1970), "Estimation of isotonic regression," in *Nonparametric Techniques in Statistical Inference (Proc. Sympos., Indiana Univ., Bloomington, Ind., 1969)*, London: Cambridge Univ. Press, pp. 177–197.

Chen, Y. and Wellner, J. A. (2013), "The Convex Density Estimators," Tech. rep., University of Cambridge, in preparation.

Chernoff, H. (1964), "Estimation of the mode," *Annals of the Institute of Statistical Mathematics*, 16, 31–41.

Cox, D. R. and Oakes, D. (1984), *Analysis of survival data*, Monographs on Statistics and Applied Probability, London: Chapman & Hall.

Cule, M. and Samworth, R. (2010), "Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density," *Electronic Journal of Statistics*, 4, 254–270.

Cule, M., Samworth, R., and Stewart, M. (2010), "Maximum likelihood estimation of a multi-dimensional log-concave density," *Journal of the Royal Statistical Society: Series B*, 72, 545–607.

Dalenius, T. (1965), "The mode—a neglected statistical parameter," *Journal of the Royal Statistical Society. Series A*, 128, 110–117.

DasGupta, A. (2008), *Asymptotic theory of statistics and probability*, Springer Texts in Statistics, New York: Springer.

Devroye, L. (1984), "A simple algorithm for generating random variates with a log-concave density," *Computing*, 33, 247–257.

Dharmadhikari, S. and Joag-Dev, K. (1988), *Unimodality, convexity, and applications*, Probability and Mathematical Statistics, Boston, MA: Academic Press Inc.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., and Picard, D. (1996), "Density Estimation by Wavelet Thresholding," *The Annals of Statistics*, 24, 508–539.

Doss, C. R. and Wellner, J. A. (2013), "Global rates of convergence of the MLEs of log-concave and s-concave densities." Tech. rep., University of Washington, available at http://arxiv.org/abs/1306.1438.

Dryanov, D. (2009), "Kolmogorov entropy for classes of convex functions," *Constr. Approx.*, 30, 137–153.

Dudley, R. M. (1984), "A course on empirical processes," in *École d'été de probabilités de Saint-Flour, XII—1982*, Berlin: Springer, vol. 1097 of *Lecture Notes in Math.*, pp. 1–142.

— (1999), *Uniform central limit theorems*, vol. 63 of *Cambridge Studies in Advanced Mathematics*, Cambridge: Cambridge University Press.

Dümbgen, L., Hüsler, A., and Rufibach, K. (2007), "Active set and EM algorithms for log-concave densities based on complete and censored data." Tech. rep., University of Bern, available at arXiv: 0707.4643.

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

Dümbgen, L., Samworth, R., and Schuhmacher, D. (2011), "Approximation by log-concave distributions, with applications to regression," *The Annals of Statistics*, 39, 702–730.

Eddy, W. F. (1980), "Optimum Kernel Estimators of the Mode," *The Annals of Statistics*, 8, pp. 870–882.

Eeden, C. (1956a), "Maximum likelihood estimation of ordered probabilities, 2," Tech. rep., Stichting Mathematisch Centrum.

— (1956b), "Maximum likelihood estimation of partially or completely ordered parameters," Tech. rep., Stichting Mathematisch Centrum.

Ekblom, H. (1972), "A Monte Carlo investigation of mode estimators in small samples," *Applied Statistics*, 21, 177–184.

Fan, J., Hung, H.-N., and Wong, W.-H. (2000), "Geometric understanding of likelihood ratio statistics," *Journal of the American Statistical Association*, 95, 836–841.

Fan, J., Zhang, C., and Zhang, J. (2001), "Generalized likelihood ratio statistics and Wilks phenomenon," *The Annals of Statistics*, 29, 153–193.

— (2002), "Correction: "Generalized likelihood ratio statistics and Wilks phenomenon" [Ann. Statist. **29** (2001), no. 1, 153–193; MR1833962 (2002e:62034)]," *The Annals of Statistics*, 30, 1811.

Folland, G. B. (1999), *Real analysis*, Pure and Applied Mathematics (New York), New York: John Wiley & Sons Inc., 2nd ed., modern techniques and their applications, A Wiley-Interscience Publication.

Gardner, R. J. (2002), "The Brunn-Minkowski inequality," *American Mathematical Society. Bulletin. New Series*, 39, 355–405.

Grenander, U. (1956), "On the theory of mortality measurement. II," *Skandinavisk Aktuarietidskrift*, 39, 125–153.

— (1965), "Some direct estimates of the mode," *The Annals of Mathematical Statistics*, 36, 131–138.

Groeneboom, P. (1996), "Lectures on inverse problems," in *Lectures on probability theory and statistics (Saint-Flour, 1994)*, Berlin: Springer, vol. 1648 of *Lecture Notes in Math.*, pp. 67–164.

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

— (2001b), “Estimation of a convex function: characterizations and asymptotic theory,” *The Annals of Statistics*, 29, 1653–1698.

Groeneboom, P., Maathuis, M. H., and Wellner, J. A. (2008a), “Current status data with competing risks: consistency and rates of convergence of the MLE,” *The Annals of Statistics*, 36, 1031–1063.

— (2008b), “Current status data with competing risks: limiting distribution of the MLE,” *The Annals of Statistics*, 36, 1064–1089.

Guntuboyina, A. and Sen, B. (2013), “Covering numbers for convex functions,” *IEEE Trans. Inform. Theor.*, 59, 1957–1965.

Hall, P. (1982), “Asymptotic theory of Grenander’s mode estimator,” *Probability Theory and Related Fields*, 60, 315–334.

Hanson, D. and Pledger, G. (1976), “Consistency in concave regression,” *The Annals of Statistics*, 4, 1038–1050.

Hargé, G. (2004), “A convex/log-concave correlation inequality for Gaussian measure and an application to abstract Wiener spaces,” *Probability theory and related fields*, 130, 415–440.

Has’minskii, R. (1979), “Lower bound for the risks of nonparametric estimates of the mode,” *Contribution to Statistics: Jaroslav Hájek Memorial Volume*, 91–97.

Herrmann, E. and Ziegler, K. (2004), “Rates of consistency for nonparametric estimation of the mode in absence of smoothness assumptions,” *Statistics and Probability Letters*, 68, 359 – 368.

Hildreth, C. (1954), “Point estimates of ordinates of concave functions,” *Journal of the American Statistical Association*, 49, 598–619.

Hoffleit, D. and Warren, Jr., W. H. (1991), *Bright Star Catalogue*, Yale University Observatory, New Haven. http://adsabs.harvard.edu/abs/1995yCat.5050....0H.

Ibragimov, I. (1956), “On the composition of unimodal distributions,” *Theory of Probability and its Applications*, 1, 255.

Jankowski, H. and Wellner, J. (2009), “Nonparametric estimation of a convex bathtub-shaped hazard function,” *Bernoulli*, 15, 1010.

Jones, M. C. (1993), “Simple boundary correction for kernel density estimation,” *Statistics and Computing*, 3, 135–146.

Jongbloed, G. (1995), "Three Statistical Inverse Problems: estimators-algorithms-asymptotics," Ph.D. thesis.

Karlin, S. (1968), *Total Positivity: Vol.: 1*, Stanford University Press.

Kim, J. and Pollard, D. (1990), "Cube root asymptotics," *The Annals of Statistics*, 191–219.

Koenker, R. and Mizera, I. (2010), "Quasi-concave density estimation," *The Annals of Statistics*, 38, 2998–3027.

Lachal, A. (1997), "Local asymptotic classes for the successive primitives of Brownian motion," *The Annals of Probability*, 25, 1712–1734.

Leindler, L. (1972), "On a certain converse of Hölder's inequality. II," *Acta Universitatis Szegediensis. Acta Scientiarum Mathematicarum*, 33, 217–223.

Lindvall, T. (1973), "Weak Convergence of Probability Measures and Random Functions in the Function Space D[0,Inf)," *Journal of Applied Probability*, 10, 109–121.

Mammen, E. (1991), "Nonparametric regression under qualitative smoothness assumptions," *The Annals of Statistics*, 19, 741–759.

Marshall, A. W. and Olkin, I. (2007), *Life distributions*, Springer Series in Statistics, New York: Springer.

Meyer, M. C. and Woodroofe, M. (2004), "Consistent maximum likelihood estimation of a unimodal density using shape restrictions," *The Canadian Journal of Statistics*, 32, 85–100.

Øksendal, B. (2003), *Stochastic differential equations*, Universitext, Berlin: Springer-Verlag, sixth ed., an introduction with applications.

Owen, A. B. (2001), *Empirical likelihood*, Chapman and Hall.

Pal, J. K., Woodroofe, M., and Meyer, M. (2007a), "Estimating a Polya frequency function$_2$," in *Complex datasets and inverse problems*, Beachwood, OH: Inst. Math. Statist., vol. 54 of *IMS Lecture Notes Monogr. Ser.*, pp. 239–249.

— (2007b), "Estimating a Polya frequency function$_2$," in *Complex datasets and inverse problems*, Beachwood, OH: Inst. Math. Statist., vol. 54 of *IMS Lecture Notes Monogr. Ser.*, pp. 239–249.

Parzen, E. (1962), "On estimation of a probability density function and mode," *The Annals of Mathematical Statistics*, 33, 1065–1076.

Prakasa Rao, B. L. S. (1969), "Estimation of a unimodal density," *Sankhyā, Series A*, 31, 23–36.

Prékopa, A. (1971), "Logarithmic concave measures with application to stochastic programming," *Acta Universitatis Szegediensis. Acta Scientiarum Mathematicarum*, 32, 301–316.

— (1973), "On logarithmic concave measures and functions," *Acta Universitatis Szegediensis. Acta Scientiarum Mathematicarum*, 34, 335–343.

Proschan, F. (1965), "Peakedness of distributions of convex combinations," *The Annals of Mathematical Statistics*, 36, 1703–1706.

Robertson, T., Wright, F. T., and Dykstra, R. L. (1988), *Order restricted statistical inference*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, Chichester: John Wiley & Sons Ltd.

Rockafellar, R. T. (1970), *Convex analysis*, Princeton Mathematical Series, No. 28, Princeton, N.J.: Princeton University Press.

Romano, J. P. (1988a), "Bootstrapping the mode," *Annals of the Institute of Statistical Mathematics*, 40, 565–586.

— (1988b), "On weak convergence and optimality of kernel density estimates of the mode," *The Annals of Statistics*, 16, 629–647.

Royden, H. L. (1988), *Real Analysis*, New York: Macmillan Publishing Company, 3rd ed.

Rufibach, K. (2006), "Log-concave density estimation and bump hunting for IID observations," Ph.D. thesis, Univ. Bern and Gottingen.

Schoenberg, I. J. (1951), "On Pólya frequency functions. I. The totally positive functions and their Laplace transforms," *Journal d'Analyse Mathématique*, 1, 331–374.

Seijo, E. and Sen, B. (2011), "Nonparametric least squares estimation of a multivariate convex regression function," *The Annals of Statistics*, 39, 1633–1657.

Sen, B., Banerjee, M., and Woodroofe, M. (2010), "Inconsistency of Bootstrap: The Grenander Estimator," *The Annals of Statistics*, 38, 1953–1977.

Seregin, A. and Wellner, J. A. (2010), "Nonparametric estimation of multivariate convex-transformed densities," *The Annals of Statistics*, 38, 3751–3781, with supplementary material available online.

Sheather, S. J. and Jones, M. C. (1991), "A reliable data-based bandwidth selection method for kernel density estimation," *Journal of the Royal Statistical Society. Series B (Methodological)*, 53, 683–690.

Sherman, S. (1955), "A theorem on convex sets with applications," *The Annals of Mathematical Statistics*, 26, 763–767.

Shorack, G. R. and Wellner, J. A. (1986), *Empirical Processes with Applications to Statistics*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, New York: John Wiley & Sons Inc.

Silverman, B. (1982), "On the estimation of a probability density function by the maximum penalized likelihood method," *The Annals of Statistics*, 795–810.

— (1986), *Density estimation for statistics and data analysis*, Monographs on Statistics and Applied Probability, London: Chapman & Hall.

Skorokhod, A. V. (1956), "Limit Theorems for Stochastic Processes," *Theory of Probability and its Applications*, 1, 261.

Staudte, R. and Sheather, S. (1990), *Robust estimation and testing*, Wiley New York.

van de Geer, S. (1993), "Hellinger-consistency of certain nonparametric maximum likelihood estimators," *The Annals of Statistics*, 21, 14–44.

van de Geer, S. A. (2000), *Applications of Empirical Process Theory*, vol. 6 of *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge: Cambridge University Press.

van der Vaart, A. W. and Wellner, J. A. (1996), *Weak convergence and empirical processes*, Springer Series in Statistics, New York: Springer-Verlag, with applications to statistics.

Venter, J. (1967), "On estimation of the mode," *The Annals of Mathematical Statistics*, 38, 1446–1455.

Walther, G. (2002), "Detecting the presence of mixing with multiscale maximum likelihood," *Journal of the American Statistical Association*, 97, 508–513.

— (2009), "Inference and modeling with log-concave distributions," *Statistical Science*, 24, 319–327.

Watanabe, H. (1970), "An Asymptotic property of Gaussian Processes," *Transactions of the American Mathematical Society*, 148, 233.

Whitt, W. (1970), "Weak Convergence of Probability Measures on the Function Space C[0, Inf)," *The Annals of Mathematical Statistics*, 41, 939–944.

— (1980), "Some useful functions for functional limit theorems," *Mathematics of Operations Research*, 5, 67–85.

— (2002), *Stochastic-process limits*, Springer Series in Operations Research, New York: Springer-Verlag, an introduction to stochastic-process limits and their application to queues.

Wilks, S. (1938), "The large-sample distribution of the likelihood ratio for testing composite hypotheses," *The Annals of Mathematical Statistics*, 9, 60–62.

Wong, W. H. and Shen, X. (1995), "Probability inequalities for likelihood ratios and convergence rates of sieve MLEs," *The Annals of Statistics*, 23, 339–362.

Woodroofe, M. and Sun, J. (1993), "A Penalized Maximum Likelihood Estimate of $f(0+)$ when f is Nonincreasing," *Statistica Sinica*, 3, 501–515.

Wright, F. (1981), "The asymptotic behavior of monotone regression estimates," *The Annals of Statistics*, 9, 443–448.

$\mathcal{P}$	Class of log-concave densities	23
m	Fixed mode (regardless of true data distribution)	23
$\mathcal{P}_m$	Class of log-concave densities with fixed mode at m	23
$\mathcal{C}_m$	Cone of concave functions with mode at m	23
λ	Lebesgue measure	23
n	Number of data points (Xs)	23
n_1	Number of Zs	23
k	$Z_k = m$	23
$\mathbb{P}_n, \mathbb{F}_n$	Empirical measure, $\frac{1}{n}\sum_{i=1}^n \delta_{X_i}$, and empirical cdf, $\frac{1}{n}\sum_{i=1}^n 1_{(-\infty,x]}(X_i)$	23
$\mathcal{C}_{n,m}$	Sub-cone of $\mathcal{C}_m$ which is piecewise linear with knots determined by data	24
$\mathcal{K}_m$	Cone of convex functions φ, with mode at m and $\int e^{\varphi(x)}dx = 1$	24
$\mathcal{K}_{n,m}$	Sub-cone of $\mathcal{K}_m$ which is piecewise linear with knots determined by data	24
$(x)_-, (x)_+$	$\min(x,0)$ and $\max(x,0)$, respectively	24
$\mathrm{ev}_x f$	$(f(x_1),\ldots,f(x_n))$ for $x \in \mathbb{R}^n$	25
LK,RK,NK	Left Knot, Right Knot, or Not a Knot	27
l^0	Number of knots in $\hat{\varphi}_n^0$	27
p	$Z_{j_p} = m$	27
$\hat{H}_{n,L}^0, \hat{H}_{n,R}^0$	$\int_{X_{(1)}}^t \hat{F}_n^0(x)dx$ and $\int_t^{X_{(n)}} \hat{F}_{n,R}^0(x)dx$	30
$\mathbb{Y}_{n,L}, \mathbb{Y}_{n,R}$	$\int_{X_{(1)}}^t \mathbb{F}_n(x)dx$ and $\int_t^{X_{(n)}} \mathbb{F}_{n,R}(x)dx$	30
$\hat{H}_n$		29
$\mathbb{Y}_n$	$\int_{X_{(1)}}^t \mathbb{F}_n(x)dx$	29
$\mathbb{F}_{n,L}, \mathbb{F}_{n,R}$		30
$\hat{F}_{n,L}^0, \hat{F}_{n,R}^0$		30
$\tau_{n,i}^0, \tau_{n,i}$	ith knot for unconstrained or constrained estimator, $\hat{f}_n$ or $\hat{f}_n^0$, respectively	39
$\tau_{c,i}^0$	ith knot for constrained limiting estimator, $\hat{f}_c^0$	118

h_0	$h_0(t) = 12t^2$, the limiting "truth"	128
$\hat{f}$	unconstrained limiting estimator	128
$\hat{f}^0$	constrained limiting estimator	129
τ_i	ith knot for unconstrained limiting estimator, $\hat{f}^0$	129
τ_i^0	ith knot for constrained limiting estimator, $\hat{f}$	129
$\mathcal{G}_{c,k}^0$		110
τ_L	Largest knot strictly less than 0 for $\hat{f}_c^0$	128
τ_R	Smallest knot strictly greater than 0 for $\hat{f}_c^0$	128
$g(a,b]$	$g(a,b] := g(b) - g(a)$	131
$\mathcal{G}^0$		132
$\Vert g\Vert_a^b, \Vert g\Vert_a^\infty$	$\sup_{t\in[a,b]} \lvert g(t)\rvert$, $\sup_{t\in[a,\infty)} \lvert g(t)\rvert$	148
$\tau_{n,+}(x), \tau_{n,+}^0(x)$	First knot larger than the point x for $\hat{f}_n$ or $\hat{f}_n^0$, respectively	91
$\tau_{n,-}(x), \tau_{n,-}^0(x)$	First knot smaller than the point x for $\hat{f}_n$ or $\hat{f}_n^0$, respectively	91
$\mathbb{Y}_n^f, \mathbb{Y}_{n,L}^f, \mathbb{Y}_{n,R}^f$		158
$\mathbb{Y}_n^\varphi, \mathbb{Y}_{n,L}^\varphi, \mathbb{Y}_{n,R}^\varphi$		158
$\hat{H}_n^f, \hat{H}_{n,L}^f, \hat{H}_{n,R}^f$		158
$\hat{H}_n^\varphi, \hat{H}_{n,L}^\varphi, \hat{H}_{n,R}^\varphi$		158
$\Vert f\Vert$	$\Vert f\Vert$ is the supremum of f over its domain	162
$\Rightarrow$	weak convergence	164
$\mathbb{D}_n$	$\mathbb{F}_n - F_0$	167
$\lvert \cdot \rvert$	Euclidean distance	167

VITA

Charles Doss was born in Tallahassee, Florida in 1985, and moved to Columbus, Ohio in 1994. He received a Bachelor of Science with Honors in mathematics in 2007 from the University of Chicago. He worked as a research associate at the Federal Reserve Bank in Chicago before beginning his graduate studies in statistics at the University of Washington. He is an avid runner. In August 2013, he became a Doctor of Philosophy.