The Annals of Statistics
2019, Vol. 47, No. 5, 2950–2976
https://doi.org/10.1214/18-AOS1770

INFERENCE FOR THE MODE OF A LOG-CONCAVE DENSITY

BY CHARLES R. DOSS[1] AND JON A. WELLNER[2]

University of Minnesota and University of Washington

We study a likelihood ratio test for the location of the mode of a log-concave density. Our test is based on comparison of the log-likelihoods corresponding to the unconstrained maximum likelihood estimator of a log-concave density and the constrained maximum likelihood estimator where the constraint is that the mode of the density is fixed, say at m. The constrained estimation problem is studied in detail in Doss and Wellner (2018). Here, the results of that paper are used to show that, under the null hypothesis (and strict curvature of $-\log f$ at the mode), the likelihood ratio statistic is asymptotically pivotal: that is, it converges in distribution to a limiting distribution which is free of nuisance parameters, thus playing the role of the χ_1^2 distribution in classical parametric statistical problems. By inverting this family of tests, we obtain new (likelihood ratio based) confidence intervals for the mode of a log-concave density f. These new intervals do not depend on any smoothing parameters. We study the new confidence intervals via Monte Carlo methods and illustrate them with two real data sets. The new intervals seem to have several advantages over existing procedures. Software implementing the test and confidence intervals is available in the R package `logcondens.mode`.

1. Introduction and overview: Inference for the mode. Let $\mathcal{P}$ denote the class of all log-concave densities f on $\mathbb{R}$. It is well known since Ibragimov (1956) that all log-concave densities f are strongly unimodal, and conversely; see Dharmadhikari and Joag-Dev (1988) for an exposition of the basic theory. Of course, "the mode" of a log-concave density f may not be a single point. It is, in general, the modal interval $\mathrm{MI}(f) \equiv \{x \in \mathbb{R} : f(x) = \sup_{y\in\mathbb{R}} f(y)\}$, and to describe "the mode" completely we need to choose a specific element of $\mathrm{MI}(f)$, for example, $M(f) \equiv \inf\{x \in \mathrm{MI}(f)\}$. For a large subclass of log-concave densities, the set reduces to a single point. Our focus here is on the latter case and, indeed, on inference concerning $M(f)$ based on i.i.d. observations $X_1, \dots, X_n$ with density $f_0 \in \mathcal{P}$. We have restricted to log-concave densities for several reasons:

Received December 2016; revised October 2018.

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

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

MSC2010 subject classifications. Primary 62G07; secondary 62G15, 62G10, 62G20.

Key words and phrases. Mode, empirical processes, likelihood ratio, pivot, convex optimization, log-concave, shape constraints.

(a) It is well known that the MLE over the class of all unimodal densities does not exist; see, for example, Birgé (1997).

(b) On the other hand, MLEs do exist for the class $\mathcal{P}$ of log-concave densities if $n \geq 2$; see, for example, Pal, Woodroofe and Meyer (2007), Rufibach (2006), Dümbgen and Rufibach (2009).

(c) Moreover, the MLEs for the class of log-concave densities have remarkable stability and continuity properties under model misspecification; see, for example, Dümbgen, Samworth and Schuhmacher (2011).

Before proceeding with our overview, it will be helpful to introduce some notation for derivatives. (Further notation and terminology will be given in Section 1.1.) In particular, we let f' denote the derivative of a differentiable function f, and we write f'' for the second derivative. We also use the notation $f^{(i)}$ for the ith derivative of f, particularly for higher derivatives.

Concerning estimation of the mode, Balabdaoui, Rufibach and Wellner (2009) showed that if $f_0 = e^{\varphi_0}$ where the concave function φ_0 has second derivative $\varphi_0^{(2)} \equiv \varphi_0''$ at the mode $m_0 = M(f_0)$ of f_0 satisfying $\varphi_0^{(2)}(m_0) < 0$, then the MLE $M(\widehat{f}_n)$ satisfies

$$n^{1/5}\big(M(\widehat{f}_n) - M(f_0)\big) \to_d \left(\frac{(4!)^2 f_0(m_0)}{f_0^{(2)}(m_0)^2}\right)^{1/5} M\big(H_2^{(2)}\big), \tag{1.1}$$

where $M(H_2^{(2)})$ has a universal distribution (not depending on f_0). Here, $\{H_2(t) : t \in \mathbb{R}\}$ is the "invelope" process on $\mathbb{R}$ defined in terms of the "driving process" $\{Y(t) : t \in \mathbb{R}\}$ defined by $Y(t) = -t^4 + \int_0^t W(s)\,ds$ for $t \in \mathbb{R}$. Thus with $X(t) \equiv Y^{(1)}(t) = -4t^3 + W(t)$,

$$dX(t) = g_0(t)\,dt + dW(t), \tag{1.2}$$

where W is two-sided Brownian motion on $\mathbb{R}$ and $g_0(t) \equiv -12t^2$. The process H_2 and its concave second derivative $H_2^{(2)}$ first appeared in Groeneboom, Jongbloed and Wellner (2001a, 2001b) in the study of other nonparametric estimation problems involving convex or concave functions; see also Balabdaoui, Rufibach and Wellner (2009).

The limit distribution (1.1) gives useful information about the behavior of $M(\widehat{f}_n)$, but it is somewhat difficult to use for inference because of the constant $((4!)^2 f_0(m_0)/f_0^{(2)}(m_0)^2)^{1/5}$ which involves the unknown density through the second derivative $f_0^{(2)}(m_0)$. This can be estimated via smoothing methods, but because we wish to avoid the consequent problem of choosing bandwidths or other tuning parameters, we take a different approach to inference here.

Instead, we first consider the following testing problem: test

$$H : M(f) = m \quad \text{versus} \quad K : M(f) \neq m,$$

where $m \in \mathbb{R}$ is fixed. To construct a likelihood ratio test of H versus K, we first need to construct both the unconstrained MLEs $\widehat{f}_n$ and the mode-constrained MLEs $\widehat{f}_n^0$. The unconstrained MLEs $\widehat{f}_n$ are available from the results of Pal, Woodroofe and Meyer (2007), Rufibach (2006) and Dümbgen and Rufibach (2009) cited above. Corresponding results concerning the existence and properties of the mode-constrained MLEs $\widehat{f}_n^0$ are given in the companion paper Doss and Wellner (2018). Global convergence rates for both estimators are given in Doss and Wellner (2016a). Once both the unconstrained estimators $\widehat{f}_n$ and the constrained estimators $\widehat{f}_n^0$ are available, then we can consider the natural likelihood ratio test of H versus K: reject the null hypothesis H if

$$2\log\lambda_n \equiv 2\log\lambda_n(m) \equiv 2n\mathbb{P}_n(\log\widehat{f}_n - \log\widehat{f}_n^0) = 2n\mathbb{P}_n(\widehat{\varphi}_n - \widehat{\varphi}_n^0)$$

is "too large" where $\widehat{f}_n = \exp(\widehat{\varphi}_n)$, $\widehat{f}_n^0 = \exp(\widehat{\varphi}_n^0)$, $\mathbb{P}_n = \sum_{i=1}^n \delta_{X_i}/n$ is the empirical measure, and $\mathbb{P}_n(g) = \int g\,d\mathbb{P}_n$. To carry out this test, we need to know how large is "too large"; that is, we need to know the (asymptotic) distribution of $2\log\lambda_n$ when H is true. Thus the primary goal of this paper is to prove the following theorem.

THEOREM 1.1. *If $X_1, \dots, X_n$ are i.i.d. $f_0 = e^{\varphi_0}$ with mode m where φ_0 is concave, twice continuously differentiable at m, and $\varphi_0^{(2)}(m) < 0$, then*

$$2\log\lambda_n \to_d \mathbb{D},$$

where $\mathbb{D}$ is a universal limiting distribution (not depending on f_0); thus $2\log\lambda_n$ is asymptotically pivotal under the assumption $\varphi_0^{(2)}(m) < 0$.

With Theorem 1.1 in hand, our likelihood ratio test with (asymptotic) size $\alpha \in (0,1)$ becomes: "reject H if $2\log\lambda_n > d_\alpha$" where d_α is chosen so that $P(\mathbb{D} > d_\alpha) = \alpha$. Furthermore, we can then form confidence intervals for m by inverting the family of likelihood ratio tests: let

$$J_{n,\alpha} \equiv \{m \in \mathbb{R} : 2\log\lambda_n(m) \le d_\alpha\}. \tag{1.3}$$

Then it follows that for $f_0 \in \mathcal{P}_m = \{f \in \mathcal{P} : M(f) = m\}$ with $(\log f_0)^{(2)}(m) < 0$, we have

$$P_{f_0}(m \in J_{n,\alpha}) \to P(\mathbb{D} \le d_\alpha) = 1 - \alpha.$$

This program is very much analogous to the methods for pointwise inference for nonparametric estimation of monotone increasing or decreasing functions developed by Banerjee and Wellner (2001) and Banerjee (2007). Those methods have recently been extended to include pointwise inference for nonparametric estimation of a monotone density by Groeneboom and Jongbloed (2015). Theorem 1.1 says that $2\log\lambda_n$ is (asymptotically) pivotal over the class of all log-concave densities f_0 satisfying $(\log f_0)^{(2)}(m) < 0$. (That log-likelihood ratios are frequently asymptotically pivotal is sometimes known as the "Wilks phenomenon" in honor of the classical result in this direction in regular parametric models by Wilks (1938).) We can specify more about the form of the limit random variable $\mathbb{D}$; see Remark 4.1.

A secondary goal of this paper is to begin a study of the likelihood ratio statistics $2\log\lambda_n$ under fixed alternatives. We leave the study of the log likelihood ratio statistic under local (contiguous) alternatives for future work. Our second theorem concerns the situation when $f \in \mathcal{P}$ has mode $M(f) \neq m$.

THEOREM 1.2. *Suppose that $f_0 \in \mathcal{P}$ with $m \notin \mathrm{MI}(f_0)$. Then*

$$\begin{aligned} \frac{2}{n}\log\lambda_n(m) &\to_p 2K\left(f_0, f_m^0\right) \\ &= 2\inf\left\{K(f_0, g) : g \in \mathcal{P}_m\right\} > 0, \end{aligned} \tag{1.4}$$

where $f_m^0 \in \mathcal{P}_m$ achieves the infimum in (1.4) and

$$K(f, g) \equiv \begin{cases} \displaystyle\int f(x)\log\frac{f(x)}{g(x)}\,dx & \text{if } f \prec\prec g, \\ \infty & \text{otherwise.} \end{cases}$$

Here, $f \prec\prec g$ means $f = 0$ whenever $g = 0$ except perhaps on a set of Lebesgue measure 0. The proof of Theorem 1.2 is given in Section 4.2, and relies on the methods used by Cule and Samworth (2010) and Dümbgen, Samworth and Schuhmacher (2011), in combination with the results of Doss and Wellner (2016a). Theorem 1.2 implies consistency of the likelihood ratio test based on the critical values from Theorem 1.1. That is, let d_α satisfy $P(\mathbb{D} > d_\alpha) = \alpha$ for $0 < \alpha < 1$, and suppose we reject $H : M(f) = m$ if $2\log\lambda_n(m) > d_\alpha$.

COROLLARY 1.3. *If the hypotheses of Theorem 1.2 hold, then the likelihood ratio test "reject H if $2\log\lambda_n(m) > d_\alpha$" is consistent: if $f \notin \mathcal{P}_m$, then*

$$P_f\left(2\log\lambda_n(m) > d_\alpha\right) \to 1.$$

Here is an explicit example.

EXAMPLE 1.4. Suppose that f is the Laplace density given by

$$f(x) = (1/2)\exp\left(-|x|\right).$$

First, we note that $M(f) = 0$ so that $f \notin \mathcal{P}_1$. Thus for testing $H : M(f) = 1$ versus $K : M(f) \neq 1$, the Laplace density f satisfies $f \in \mathcal{P}\setminus\mathcal{P}_1$. So we have (incorrectly) hypothesized that $M(f) = 1 \equiv m$. In this case, the constrained MLE $\widehat{f}_n^0$ satisfies $\int |\widehat{f}_n^0 - f^0|\,dx \to_{\text{a.s.}} 0$ where $f^0 \equiv g^* \in \mathcal{P}_1$ is determined by Theorem 4.2 which is the population analogue of Theorem 2.10 of Doss and Wellner (2018). It also satisfies (1.4) in Theorem 1.2. In the present case, $g^* = g_{a^*}$ where $\{g_a : a \in (0, 1]\}$ is the family of densities given by

$$g_a(x) = \begin{cases} (1/2)e^x & -\infty < x \le -a, \\ (1/2)e^{-a} & -a \le x \le 1, \\ (1/2)e^{-a}e^{-c(x-1)} & 1 \le x < \infty, \end{cases}$$

where $c \equiv c(a) = 1/(2e^a - (2 + a))$ is chosen so that $\int g_a(x)\, dx = 1$. Here, it is not hard to show that $a^* \approx 0.490151\ldots$ satisfies $c(a^*)^2 = \exp(-(a^* - 1))$, while $K(f, f^0) = K(f, g_{a^*}) \approx 0.03377\ldots$.

Although the basic approach here has points in common with the developments in Banerjee and Wellner (2001) and Banerjee (2007), the details of the proofs require several new tools and techniques due to the relative lack of development of theory for the mode-constrained log-concave MLEs. Furthermore, the proof of Theorem 1.1 is significantly more complicated than corresponding proofs in Banerjee and Wellner (2001), Banerjee (2007) or Groeneboom and Jongbloed (2015): in the present context, the mode-constrained estimator and the unconstrained estimator are not identically equal to each other away from the constraint, whereas in many monotonicity-based cases, the corresponding constrained and unconstrained estimators are indeed equal away from the constraint. In the case of monotone density estimation studied by Groeneboom and Jongbloed (2015), the constrained and unconstrained estimators are not identically equal away from the constraint, but the differences can be handled using the so-called min-max formula (see, e.g., Lemma 3.2 of Groeneboom and Jongbloed (2015)), which does not have an analog for concavity-based problems. Thus, beyond being interesting in its own right, the proof of Theorem 1.1 is useful for opening the door to the study of likelihood ratios in other concavity/convexity-based problems. These could be likelihood ratios for locations of extrema or likelihood ratios for the values (heights) of functions in concavity/convexity-based problems. We present some discussion of possible extensions in Section 6.

To prove Theorem 1.1, we first prepare the way by reviewing the local asymptotic distribution theory for the unconstrained estimators $\widehat{f}_n$ and $\widehat{\varphi}_n$ developed by Balabdaoui, Rufibach and Wellner (2009) and asymptotic theory for $\widehat{f}_n^0$ and $\widehat{\varphi}_n^0$ developed by Doss and Wellner (2018). These results are stated in Section 3.

Section 4 contains an outline of our proof of Theorem 1.1 and the full proof of Theorem 1.2. The complete details of the long proof of Theorem 1.1 are deferred to Sections A.1 and A.2 in the Supplementary Material Doss and Wellner (2019). In Section A.1, we treat remainder terms in a local neighborhood of the mode m, while remainder terms away from the mode are treated in Section A.2. Our proofs in Sections A.1 and A.2 rely heavily on the theory developed for the constrained estimators in Doss and Wellner (2018) and on the new uniform consistency results for the constrained estimator presented in Section 2 (with proofs in Section A.4).

In Section 5, we present Monte Carlo estimates of quantiles of the distribution of $\mathbb{D}$ and provide empirical evidence supporting the universality of the limit distribution (under the assumption that $\varphi_0^{(2)}(m) < 0$). We illustrate the likelihood ratio confidence sets with Monte Carlo evidence demonstrating the coverage probabilities of our proposed intervals are near the nominal levels. Further simulation studies and application to two data sets can be found in Doss and Wellner (2016b). Section 6 gives a brief description of further problems and potential developments. We also discuss connections with the results of Romano (1988a, 1988b), Donoho and Liu (1991) and Pfanzagl (1998, 2000). In Section 1.1, we discuss notation and terminology.

1.1. *Notation and terminology.* Several classes of concave functions will play a central role in this paper.

$$\mathcal{C} := \{\varphi : \mathbb{R} \to [-\infty, \infty) \mid \varphi \text{ is concave, closed and proper}\} \tag{1.5}$$

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

$$\mathcal{C}_m := \{\varphi \in \mathcal{C} \mid \varphi(m) \geq \varphi(x) \text{ for all } x \in \mathbb{R}\}. \tag{1.6}$$

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

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

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

$$\Psi_n(\varphi) = \frac{1}{n}\sum_{i=1}^n \varphi(X_i) - \int_{\mathbb{R}} e^{\varphi(x)}\, dx = \mathbb{P}_n\varphi - \int_{\mathbb{R}} e^{\varphi}\, d\lambda, \tag{1.7}$$

where we have used the standard device of including the Lagrange term $\int_{\mathbb{R}} e^{\varphi(x)}\, dx$ in Ψ_n so that we can maximize Ψ_n over all concave functions $\mathcal{C}$ or $\mathcal{C}_m$ (rather than maximizing over classes corresponding to density functions). This is as in

Silverman (1982). We will denote the unconstrained MLEs of φ_0, f_0 and F_0 by $\widehat{\varphi}_n$, $\widehat{f}_n$ and $\widehat{F}_n$, respectively. These exist uniquely by Proposition 1 of Walther (2002). The corresponding constrained estimators with mode m will be denoted by $\widehat{\varphi}_n^0$, $\widehat{f}_n^0$ and $\widehat{F}_n^0$. These exist uniquely by Theorem 2.6 of Doss and Wellner (2018) (or Lemma 2.0.3 of Doss (2013b)). Thus

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

2. Uniform consistency and rates. Here, we recall the uniform rate-consistency theorem of Dümbgen and Rufibach (2009), and give a partial analogue for the mode-constrained MLE. The new result, given in Theorem 2.1 Part B below, is of interest in its own right for describing the theoretical behavior of the mode-constrained MLE. Additionally, the proof of Theorem 1.1 relies on (both parts of) Theorem 2.1. It should be mentioned that Theorem 2.1 Part B is a nontrivial extension of the theorem of Dümbgen and Rufibach (2009), with a fairly difficult proof.

To state the uniform results, we define $\mathcal{H}^{\beta,L}(I)$ to be the collection of real-valued functions g on the closed interval I satisfying $|g(y) - g(x)| \le L|y - x|$ if $\beta = 1$ and $|g'(y) - g'(x)| \le L|y - x|^{\beta-1}$ if $\beta > 1$, for all $x, y \in I$. We let $\rho_n \equiv n^{-1} \log n$.

THEOREM 2.1 (Uniform consistency and rates of convergence). A. (*Dümbgen and Rufibach* (2009)) *Suppose that* $f_0 \in \mathcal{P}$. *If* $\varphi_0 \in \mathcal{H}^{\beta,L}(K)$ *for some* $1 \le \beta \le 2$, $L > 0$, *and* $K = [b, c] \subset \operatorname{int}(\{f_0 > 0\})$, *then*

$$\sup_{t \in K} (\widehat{\varphi}_n - \varphi_0)(t) = O_p\big(\rho_n^{\beta/(2\beta+1)}\big), \quad \text{and} \tag{2.1}$$

$$\sup_{t \in K_n} (\varphi_0 - \widehat{\varphi}_n)(t) = O_p\big(\rho_n^{\beta/(2\beta+1)}\big), \tag{2.2}$$

where $K_n \equiv [b + \rho_n^{1/(2\beta+1)}, c - \rho_n^{1/(2\beta+1)}]$. *These results remain true when* $\widehat{\varphi}_n$ *is replaced by* $\widehat{f}_n$ *and* φ_0 *by* f_0.

B. *Suppose that* $f_0 \in \mathcal{P}_m$, $\varphi_0 \in \mathcal{H}^{2,L}(K)$ *for some* $L > 0$, $\varphi_0^{(2)}(m) < 0$, *and* $K = [b, c] \subset \operatorname{int}(\{f_0 > 0\})$. *Then the results of Part* A *hold true with* $\beta = 2$, *with* $\widehat{\varphi}_n$ *replaced by* $\widehat{\varphi}_n^0$ *and with* $\widehat{f}_n$ *replaced by* $\widehat{f}_n^0$.

The proof of Theorem 2.1 is given in Section A.4 in Doss and Wellner (2019).

3. Unconstrained and constrained local limit processes. As can be seen from Theorem 3.1A below, $\widehat{\varphi}_n(x)$ and $\widehat{\varphi}_n^0(x)$ are asymptotically equivalent at fixed $x \ne m$ when $\varphi_0''(x) < 0$ and $M(f_0) = m$. Thus, it turns out that the limit distribution of $2 \log \lambda_n$, under the hypotheses of Theorem 1.1, depends on the joint distribution of $\widehat{\varphi}_n(x)$ and $\widehat{\varphi}_n^0(x)$ at points x near m, and, specifically, within $n^{-1/5}$-neighborhoods of m. Thus, in this section we recall the limit distributions of $\widehat{\varphi}_n$

and $\widehat{\varphi}_n^0$ from Theorem 2.1 of Balabdaoui, Rufibach and Wellner (2009) and Theorems 5.5 and 5.7 (see also Theorem 5.8) of Doss and Wellner (2018). The process giving the limit distribution of $\widehat{\varphi}_n$ was first studied by Groeneboom, Jongbloed and Wellner (2001a). Here are the assumptions we will need.

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

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

THEOREM 3.1 (Balabdaoui, Rufibach and Wellner (2009), Doss and Wellner (2018)). A. (*At a point $x_0 \neq m$.*) *Suppose that φ_0 and f_0 satisfy Assumption* 2. *Then*

$$\begin{pmatrix} n^{2/5}(\widehat{\varphi}_n(x_0) - \varphi_0(x_0)) \\ n^{2/5}(\widehat{\varphi}_n^0(x_0) - \varphi_0(x_0)) \end{pmatrix} \to_d \begin{pmatrix} \mathbb{V} \\ \mathbb{V} \end{pmatrix},$$

where $\mathbb{V} \equiv C(x_0, \varphi_0) H^{(2)}(0)$, where H is described in Theorem 5.1 *of Doss and Wellner* (2018), *and where $C(x_0, \varphi_0)$ is as given in* (B.6) (*but with m replaced by x_0*):

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

Consequently,

$$n^{2/5}(\widehat{\varphi}_n(x_0) - \widehat{\varphi}_n^0(x_0)) \to_p 0.$$

B. (*In $n^{-1/5}$—neighborhoods of m.*) *Suppose φ_0 and f_0 satisfy Assumption* 1. *Define processes $\mathbb{X}_n$ and $\mathbb{X}_n^0$ by*

$$\mathbb{X}_n(t) \equiv n^{2/5}(\widehat{\varphi}_n(m + n^{-1/5}t) - \varphi_0(m)),$$
$$\mathbb{X}_n^0(t) \equiv n^{2/5}(\widehat{\varphi}_n^0(m + n^{-1/5}t) - \varphi_0(m)).$$

Then the finite-dimensional distributions of $(\mathbb{X}_n(t), \mathbb{X}_n^0(t))$ converge in distribution to the finite-dimensional distributions of the processes

$$(\widehat{\varphi}_{a,\sigma}(t), \widehat{\varphi}_{a,\sigma}^0(t)) \stackrel{d}{=} \frac{1}{\gamma_1\gamma_2^2}(\widehat{\varphi}(t/\gamma_2), \widehat{\varphi}^0(t/\gamma_2)) \equiv (\mathbb{X}(t), \mathbb{X}^0(t)),$$

where $H, H^0, H^{(2)} = \widehat{\varphi}$ and $(H^0)^{(2)} = \widehat{\varphi}^0$ are as described in Theorems 5.1 *and* 5.2 *of Doss and Wellner* (2018), *and γ_i, $i = 1, 2$, is described in Section* B.1. *Furthermore, for $p \geq 1$*

$$(\mathbb{X}_n(t), \mathbb{X}_n^0(t)) \to_d (\mathbb{X}(t), \mathbb{X}^0(t))$$

in $\mathcal{L}^p[-K, K] \times \mathcal{L}^p[-K, K]$ for each $K > 0$.

4. Proof sketches for Theorems 1.1 and 1.2. Now we present proof sketches for our two main theorems (and make use of the results in the previous two sections).

4.1. *Proof sketch for Theorem* 1.1. To begin our sketch of the proof of Theorem 1.1 we first give the basic decomposition we will use. For the entire proof, we argue on the event where $m \in (X_{(1)}, X_{(n)})$, which has probability approaching 1 as $n \to \infty$ (so our convergence in distribution conclusion is not affected by our restricting attention to this event). We begin by using $\int_{\mathbb{R}} \widehat{f}_n(u)\, du = 1 = \int_{\mathbb{R}} \widehat{f}_n^0(u)\, du$ to write

$$
\begin{aligned}
2\log\lambda_n &= 2n\mathbb{P}_n(\widehat{\varphi}_n - \widehat{\varphi}_n^0) \\
&= 2n\int_{\mathbb{R}} (\widehat{\varphi}_n - \widehat{\varphi}_n^0)\, d\mathbb{F}_n - \int_{\mathbb{R}} (\widehat{f}_n(u) - \widehat{f}_n^0(u))\, du \\
&= 2n\int_{[X_{(1)}, X_{(n)}]} (\widehat{\varphi}_n\, d\widehat{F}_n - \widehat{\varphi}_n^0\, d\widehat{F}_n^0) \\
&\qquad - 2n\int_{[X_{(1)}, X_{(n)}]} (e^{\widehat{\varphi}_n(u)} - e^{\widehat{\varphi}_n^0(u)})\, du,
\end{aligned} \tag{4.1}
$$

where we have used the characterization Theorems 2.2 and 2.8 of Doss and Wellner (2018) with $\Delta = \pm\widehat{\varphi}_n$ and $\Delta = \pm\widehat{\varphi}_n^0$, respectively. As we will see, inclusion of the second term in (4.1) will be of considerable help in the analysis.

Now we split the integrals in (4.1) into two regions: let $D_n \equiv [t_1, t_2]$ for some $t_1 < m < t_2$ and then let $D_n^c \equiv [X_{(1)}, X_{(n)}] \setminus D_n$. The set D_n is the region containing the mode m; here the unconstrained estimator $\widehat{\varphi}_n$ and the constrained estimator $\widehat{\varphi}_n^0$ tend to differ. On the other hand, D_n^c is the union of two sets away from the mode, and on both of these sets the unconstrained estimator $\widehat{\varphi}_n$ and the constrained estimator $\widehat{\varphi}_n^0$ are asymptotically equivalent (or at least nearly so). Sometimes we will take the t_i, $i = 1, 2$, to be constant in n, sometimes to be fixed or random sequences approaching m as $n \to \infty$. We will sometimes suppress the dependence of $D_n \equiv D_{n,t_1,t_2}$ on t_i, and will emphasize it when it is important. Now, from (4.1), we can write

$$
\begin{aligned}
2\log\lambda_n &= 2n\biggl\{\int_{D_n} (\widehat{\varphi}_n\, d\widehat{F}_n - \widehat{\varphi}_n^0\, d\widehat{F}_n^0) - \int_{D_n} (e^{\widehat{\varphi}_n(u)} - e^{\widehat{\varphi}_n^0(u)})\, du \\
&\qquad + \int_{D_n^c} (\widehat{\varphi}_n\, d\widehat{F}_n - \widehat{\varphi}_n^0\, d\widehat{F}_n^0) - \int_{D_n^c} (e^{\widehat{\varphi}_n(u)} - e^{\widehat{\varphi}_n^0(u)})\, du\biggr\} \\
&= 2n\biggl\{\int_{D_n} ((\widehat{\varphi}_n - \varphi_0(m))\, d\widehat{F}_n - (\widehat{\varphi}_n^0 - \varphi_0(m))\, d\widehat{F}_n^0) \\
&\qquad - \int_{D_n} ((e^{\widehat{\varphi}_n(u)} - e^{\varphi_0(m)}) - (e^{\widehat{\varphi}_n^0(u)} - e^{\varphi_0(m)}))\, du\biggr\} \\
&\qquad + 2n(R_{n,1} + R_{n,1}^c),
\end{aligned}
$$

where

$$R_{n,t_1,t_2} \equiv R_{n,1} \equiv \int_{D_n} \varphi_0(m)\big(\widehat{f}_n(x) - \widehat{f}_n^0(x)\big)\,dx, \tag{4.2}$$

$$R_{n,t_1,t_2}^c \equiv R_{n,1}^c \equiv \int_{D_n^c} \big(\widehat{\varphi}_n\,d\widehat{F}_n - \widehat{\varphi}_n^0\,d\widehat{F}_n^0\big) - \int_{D_n^c} \big(e^{\widehat{\varphi}_n(u)} - e^{\widehat{\varphi}_n^0(u)}\big)\,du. \tag{4.3}$$

Now we use an expansion of the exponential function to rewrite the second part of the main term: since

$$e^b - e^a = e^a\{e^{b-a} - 1\} = e^a\left\{(b-a) + \frac{1}{2}(b-a)^2 + \frac{1}{6}e^{a^*}(b-a)^3\right\},$$

where $|a^*| \le |b-a|$, we have

$$\begin{aligned}
&\int_{D_n} \big(\big(e^{\widehat{\varphi}_n(u)} - e^{\varphi_0(m)}\big) - \big(e^{\widehat{\varphi}_n^0(u)} - e^{\varphi_0(m)}\big)\big)\,du \\
&\qquad = \int_{D_n} e^{\varphi_0(m)}\left(\big(\widehat{\varphi}_n(u) - \varphi_0(m)\big) + \frac{1}{2}\big(\widehat{\varphi}_n(u) - \varphi_0(m)\big)^2\right)\,du + R_{n,2} \\
&\qquad\quad - \int_{D_n} e^{\varphi_0(m)}\left(\big(\widehat{\varphi}_n^0(u) - \varphi_0(m)\big) + \frac{1}{2}\big(\widehat{\varphi}_n^0(u) - \varphi_0(m)\big)^2\right)\,du - R_{n,2}^0,
\end{aligned}$$

where

$$R_{n,2} \equiv \int_{D_n} \frac{1}{6} f_0(m) e^{\tilde{x}_{n,2}(u)}\big(\widehat{\varphi}_n(u) - \varphi_0(m)\big)^3\,du, \tag{4.4}$$

$$R_{n,2}^0 \equiv \int_{D_n} \frac{1}{6} f_0(m) e^{\tilde{x}_{n,2}^0(u)}\big(\widehat{\varphi}_n^0(u) - \varphi_0(m)\big)^3\,du. \tag{4.5}$$

Thus

$$\begin{aligned}
&2\log\lambda_n \\
&\quad = n\left\{2\int_{D_n}\left(\big(\widehat{\varphi}_n(u) - \varphi_0(m)\big)\big(d\widehat{F}_n(u) - f_0(m)\,du\big)\right.\right. \\
&\qquad\quad \left. - 2\int_{D_n}\big(\widehat{\varphi}_n^0(u) - \varphi_0(m)\big)\big(d\widehat{F}_n^0(u) - f_0(m)\,du\big)\right) \\
&\qquad\quad \left. - \int_{D_n}\big(\big(\widehat{\varphi}_n(u) - \varphi_0(m)\big)^2 - \big(\widehat{\varphi}_n^0(u) - \varphi_0(m)\big)^2\big) f_0(m)\,du\right\} \\
&\qquad\quad + 2n\big(R_{n,1} + R_{n,1}^c + R_{n,2} - R_{n,2}^0\big).
\end{aligned} \tag{4.6}$$

Now we expand the first two terms in the last display, again using a two term expansion, $e^b - e^a = (e^{b-a} - 1)e^a = e^a\{(b-a) + \frac{1}{2}(b-a)^2 e^{a^*}\}$, to find that

$$\int_{D_n} \big(\widehat{\varphi}_n(u) - \varphi_0(m)\big)\big(d\widehat{F}_n(u) - f_0(m)\,du\big)$$

$$= \int_{D_n} (\widehat{\varphi}_n(u) - \varphi_0(m))(e^{\widehat{\varphi}_n(u)-\varphi_0(u)} - 1) f_0(m)\, du$$

$$= \int_{D_n} (\widehat{\varphi}_n(u) - \varphi_0(m))\Big((\widehat{\varphi}_n(u) - \varphi_0(m)) + e^{\tilde{x}_{n,3}(u)} \frac{1}{2}(\widehat{\varphi}_n(u) - \varphi_0(m))^2\Big) \times f_0(m)\, du$$

$$= \int_{D_n} (\widehat{\varphi}_n(u) - \varphi_0(m))^2 f_0(m)\, du + R_{n,3},$$

where

$$R_{n,3,t_1,t_2} \equiv R_{n,3} = \int_{D_n} \frac{1}{2} f_0(m) e^{\tilde{x}_{n,3}(u)} (\widehat{\varphi}_n(u) - \varphi_0(m))^3\, du. \tag{4.7}$$

Similarly,

$$\int_{D_n} (\widehat{\varphi}_n^0(u) - \varphi_0(m))(d\widehat{F}_n^0(u) - f_0(m)\, du) = \int_{D_n} (\widehat{\varphi}_n^0(u) - \varphi_0(m))^2 f_0(m)\, du + R_{n,3}^0,$$

where

$$R_{n,3,t_1,t_2}^0 \equiv R_{n,3}^0 = \int_{D_n} \frac{1}{2} f_0(m) e^{\tilde{x}_{n,3}^0(u)} (\widehat{\varphi}_n^0(u) - \varphi_0(m))^3\, du. \tag{4.8}$$

If we let $t_1 = t_{n,1} = m - bn^{-1/5}$ and $t_2 = t_{n,2} = m + bn^{-1/5}$ for $b > 0$, then from (4.6) we now have

$$\begin{aligned} 2 \log \lambda_n &= n \int_{D_n} f_0(m)\{(\widehat{\varphi}_n(u) - \varphi_0(m))^2 - (\widehat{\varphi}_n^0(u) - \varphi_0(m))^2\}\, du \\ &\quad + 2n(R_{n,1} + R_{n,1}^c + R_{n,2} - R_{n,2}^0 + R_{n,3} - R_{n,3}^0) \\ &\equiv \mathbb{D}_{n,b} + R_n. \end{aligned} \tag{4.9}$$

Now we sketch the behavior of $\mathbb{D}_{n,b}$. Let $u = m + vn^{-1/5}$; with this change of variables and the definition of $t_{n,i}$, $i = 1, 2$, we can rewrite $\mathbb{D}_{n,b}$ as

$$\mathbb{D}_{n,b} = f_0(m) n^{4/5} \int_{-b}^{b} \{(\widehat{\varphi}_n(m + n^{-1/5}v) - \varphi_0(m))^2 - (\widehat{\varphi}_n^0(m + n^{-1/5}v) - \varphi_0(m))^2\}\, dv.$$

By Theorem 3.1B this converges in distribution to

$$f_0(m) \int_{-b}^{b} \{(\widehat{\varphi}_{a,\sigma}(v))^2 - (\widehat{\varphi}_{a,\sigma}^0(v))^2\}\, dv, \tag{4.10}$$

where the processes $(\widehat{\varphi}_{a,\sigma}, \widehat{\varphi}^0_{a,\sigma})$ are related to $(\widehat{\varphi}, \widehat{\varphi}^0)$ by the scaling relations (B.1) and (B.2). We conclude that the limiting random variable in (4.10) is equal in distribution to

$$
\begin{aligned}
& f_0(m) \int_{-b}^{b} \left\{ \left(\frac{1}{\gamma_1 \gamma_2^2} \widehat{\varphi}(v/\gamma_2) \right)^2 - \left(\frac{1}{\gamma_1 \gamma_2^2} \widehat{\varphi}^0(v/\gamma_2) \right)^2 \right\} dv \\
&\quad = \frac{f_0(m)}{\gamma_1^2 \gamma_2^3} \frac{1}{\gamma_2} \int_{-b}^{b} \{ \widehat{\varphi}(v/\gamma_2)^2 - \widehat{\varphi}^0(v/\gamma_2)^2 \} \, dv \\
&\quad = \int_{-b/\gamma_2}^{b/\gamma_2} \{ \widehat{\varphi}(s)^2 - \widehat{\varphi}^0(s)^2 \} \, ds \qquad (4.11)
\end{aligned}
$$

in view of (B.5). This is not yet free of the parameter γ_2, but it will become so if we let $b \to \infty$. If we show that this is permissible and we show that the remainder term R_n in (4.9) is negligible, then the proof of Theorem 1.1 will be complete. For details, see Section A of Doss and Wellner (2019).

REMARK 4.1. As is suggested by (4.11) (and proved in Section A of Doss and Wellner (2019)), the form of the random variable $\mathbb{D}$ from Theorem 1.1 is

$$
\mathbb{D} = \int_{-\infty}^{\infty} \{ \widehat{\varphi}(u)^2 - \widehat{\varphi}^0(u)^2 \} \, du.
$$

The form of this random variable is the same as that found in Banerjee and Wellner (2001) and Banerjee (2007), if we replace our $\widehat{\varphi}$ and $\widehat{\varphi}^0$ with the corresponding random functions studied in the monotone case.

4.2. *Proof of Theorem* 1.2. Recall $\mathcal{P}_m = \{ f \in \mathcal{P} : M(f) = m \}$. We now assume that $f \in \mathcal{P} \setminus \mathcal{P}_m$. Let λ be Lebesgue measure and let

$$
\begin{aligned}
f_m^0 &= \operatorname*{argmin}_{g \in \mathcal{P}_m} \left\{ - \int \log g f_0 \, d\lambda \right\} \\
&= \operatorname*{argmin}_{g \in \mathcal{P}_m} \int f_0 (\log f_0 - \log g) \, d\lambda = \operatorname*{argmin}_{g \in \mathcal{P}_m} K(f_0, g), \qquad (4.12)
\end{aligned}
$$

where we will make (4.12) rigorous later, in Theorem 4.1. Let $\mathbb{P}_n = \sum_{i=1}^n \delta_{X_i}/n$ be the empirical measure and for a function g let $\mathbb{P}_n(g) = \int g \, d\mathbb{P}_n$. We now have

$$
\begin{aligned}
n^{-1} \log \lambda_n(m) &= \mathbb{P}_n \big(\log \widehat{f}_n / \widehat{f}_n^0 \big) = \mathbb{P}_n \left\{ \log \frac{\widehat{f}_n}{f_0} \cdot \frac{f_0}{f_m^0} \cdot \frac{f_m^0}{\widehat{f}_n^0} \right\} \\
&= \mathbb{P}_n \left\{ \log \frac{\widehat{f}_n}{f_0} \right\} + \mathbb{P}_n \left\{ \log \frac{f_0}{f_m^0} \right\} - \mathbb{P}_n \left\{ \log \frac{\widehat{f}_n^0}{f_m^0} \right\}. \qquad (4.13)
\end{aligned}
$$

From this, we will conclude that as $n \to \infty$,

$$
n^{-1} 2 \log \lambda_n(m) = O_p(n^{-4/5}) + 2\mathbb{P}_n \left\{ \log \frac{f_0}{f_m^0} \right\} - o_p(1) \to_p 2K(f_0, f_m^0).
$$

That $\mathbb{P}_n\{\log \frac{\widehat{f}_n}{f_0}\} = O_p(n^{-4/5})$ follows from Doss and Wellner (2016a), Corollary 3.2, page 962. The convergence of $2\mathbb{P}_n\{\log \frac{f_0}{f_m^0}\}$ to $K(f_0, f_m^0)$ follows from the weak law of large numbers. The indicated negligibility of the third term in (4.13) follows from Theorem 4.3 below (which is a constrained analogue of Theorem 2.15 of Dümbgen, Samworth and Schuhmacher (2011)).

It remains to justify the definition given in (4.12), and to show that the third term of (4.13) is $o_p(1)$, under the assumptions of Theorem 1.2. We first state three theorems. These are mode-constrained analogues of Theorems 2.2, 2.7 and 2.15 of Dümbgen, Samworth and Schuhmacher (2011), and are proved with methods similar to the methods used in Dümbgen, Samworth and Schuhmacher (2011).

Now we set

$$L(\varphi, Q) \equiv \int \varphi \, dQ - \int e^{\varphi} \, d\lambda + 1$$

and define

$$L_m(Q) \equiv \sup_{\varphi \in \mathcal{C}_m} L(\varphi, Q),$$

where

$$\mathcal{C}_m \equiv \{\varphi : m \in \mathrm{MI}(\varphi), \varphi \text{ concave}\}$$

and recall $\mathrm{MI}(\varphi) \equiv \{x \in \mathbb{R} : \varphi(x) = \sup_{y \in \mathbb{R}} \varphi(y)\}$. If for fixed Q, there exists $\psi_m \in \mathcal{C}_m$ such that

$$L(\psi_m, Q) = L_m(Q) \in \mathbb{R},$$

then ψ_m will automatically satisfy $\int \exp(\psi_m(x)) \, dx = 1$: note that $\phi + c \in \mathcal{C}_m$ for any fixed $\phi \in \mathcal{C}_m$ and $c \in \mathbb{R}$. On the other hand,

$$\begin{aligned}\frac{\partial}{\partial c} L(\mathcal{C} + c, Q) &= \frac{\partial}{\partial c}\left\{\int (\phi + c) \, dQ - e^c \int e^{\phi} \, d\lambda + 1\right\} \\ &= 1 - e^c \int e^{\phi} \, d\lambda\end{aligned}$$

if $L(\phi, Q) \in \mathbb{R}$. Thus $L(\phi + c, Q)$ is maximal for $c = -\log \int e^{\phi} \, d\lambda$.

For the next theorem, we need to define

$$\mathrm{csupp}_m(Q) = \bigcap_{C \in \mathfrak{C}} C$$

$$\text{where } \mathfrak{C} = \{C \subseteq \mathbb{R}^d : C \text{ closed, convex, } Q(C) = 1, m \in C\}.$$

THEOREM 4.1. *Let Q be a measure on $\mathbb{R}^d$. The value of $L_m(Q)$ is real if and only if $\int |x| \, dQ(x) < \infty$ and* $\mathrm{int}(\mathrm{csupp}_m(Q)) \neq \varnothing$. *In that case, there exists a unique $\psi_m \equiv \psi_m(\cdot|Q) \in \mathrm{argmax}_{\varphi \in \mathcal{C}_m} L(\varphi, Q)$. This function ψ_m satisfies $\int e^{\psi_m} \, d\lambda = 1$ and*

$$\mathrm{int}(\mathrm{csupp}_m(Q)) \subseteq \mathrm{dom}(\psi_m) \subseteq \mathrm{csupp}_m(Q),$$

where $\mathrm{dom}(\psi_m) \equiv \{x \in \mathbb{R}^d : \psi_m(x) > -\infty\}$.

Theorem 4.1 justifies rigorously the definition given in (4.12), since $f_0 \in \mathcal{P}$ has a finite mean and satisfies $\mathrm{int}(\mathrm{csupp}_m(P_0)) \neq \emptyset$, where P_0 is the measure corresponding to f_0. Now, for $\psi_m \in \mathcal{C}_m$, let

$$\mathcal{S}(\psi_m) = \{x \in \mathrm{dom}(\psi_m) : \psi_m(x) > 2^{-1}(\psi_m(x-\delta) + \psi_m(x+\delta)) \text{ for all } \delta > 0\},$$

and

$$\mathcal{S}_L(\psi_m) = \{x \in \mathcal{S}(\psi_m) : \psi_m'(x-) > 0\} \subseteq (-\infty, m],$$
$$\mathcal{S}_R(\psi_m) = \{x \in \mathcal{S}(\psi_m) : \psi_m'(x+) < 0\} \subseteq [m, \infty).$$

It is possible for m to be an element of either of the sets $\mathcal{S}_L(\psi_m)$ and $\mathcal{S}_R(\psi_m)$ without being a member of the other. The following theorem is the population analogue of Theorem 2.10 of Doss and Wellner (2018).

THEOREM 4.2. *Let Q be a distribution on $\mathbb{R}$ with $\mathrm{int}(\mathrm{csupp}_m(Q)) \neq \emptyset$, with finite first moment, and with distribution function G. Let F_m be a distribution function with log-density $\varphi_m \in \mathcal{C}_m$. Then $\varphi_m = \psi_m(\cdot|Q)$ if and only if*

$$\int_{-\infty}^{x} F_m(y)\,dy \le \int_{-\infty}^{x} G(y)\,dy \qquad \text{for all } x \le m, \quad \text{and} \tag{4.14}$$

$$\int_{x}^{\infty} (1 - F_m(y))\,dy \le \int_{x}^{\infty} (1 - G(y))\,dy \qquad \text{for all } x \ge m, \tag{4.15}$$

with equality in (4.14) if $x \in \mathcal{S}_L(\varphi_m)$ and equality in (4.15) if $x \in \mathcal{S}_R(\varphi_m)$.

Thus, again much as in Dümbgen, Samworth and Schuhmacher (2011), for $x \in \mathcal{S}(\psi_m(\cdot|Q))$, $x \le m$, and (small) $\delta > 0$,

$$0 \le \frac{1}{\delta}\int_{x-\delta}^{x} (F_m(y) - G(y))\,dy \to F_m(x) - G(x-) \qquad \text{as } \delta \searrow 0,$$
$$0 \ge \frac{1}{\delta}\int_{x}^{x+\delta} F_m(y) - G(y))\,dy \to F_m(x) - G(x) \qquad \text{as } \delta \searrow 0,$$

and hence $G(x-) \le F_m(x) \le G(x)$ for all $x \in \mathbb{R}$.

Now we need to understand the properties of the maps $Q \mapsto L_m(Q)$ and $Q \mapsto \psi_m(\cdot|Q)$ on $\mathcal{Q}_1 \cap \mathcal{Q}_{0,m}$, where we let $\mathcal{Q}_1 = \{Q : \int |x|\,dQ < \infty\}$ and $\mathcal{Q}_{0,m} = \{Q : \mathrm{int}(\mathrm{csupp}_m(Q)) \neq \emptyset\}$. As in Dümbgen, Samworth and Schuhmacher (2011), we show that these are both continuous with respect to Mallows distance D_1:

$$D_1(Q, Q') \equiv \inf_{(X,X')} E|X - X'|,$$

where the infimum is taken over all pairs (X, X') of random variables $X \sim Q$ and $X' \sim Q'$ on a common probability space. Convergence of Q_n to Q in Mallows distance is equivalent to having $\int |x|\,dQ_n \to \int |x|\,dQ$ and $Q_n \Rightarrow Q$ (Mallows (1972)).

THEOREM 4.3. *Let $\{Q_n\}$ be a sequence of distributions on $\mathbb{R}^d$ such that $D_1(Q_n, Q) \to 0$ for some $Q \in \mathcal{Q}_1$. Then*

$$L_m(Q_n) \to L_m(Q).$$

If $Q \in \mathcal{Q}_{0,m} \cap \mathcal{Q}_1$, the probability densities $f^0 \equiv \exp(\psi_m(\cdot|Q))$ and $f_n^0 \equiv \exp(\psi_m(\cdot|Q_n))$ are well defined for large n and satisfy

$$\lim_{n\to\infty, x\to y} f_n^0(x) = f^0(y) \qquad \text{for all } y \in \mathbb{R}^d \setminus \partial\{f^0 > 0\},$$

$$\limsup_{n\to\infty, x\to y} f_n^0(x) \le f^0(y) \qquad \text{for all } y \in \partial\{f^0 > 0\},$$

$$\int |f_n^0(x) - f^0(x)|\, dx \to 0.$$

We can now show that the third term of (4.13) is $o_p(1)$ under the assumptions of Theorem 1.2. First, note that $\mathbb{P}_n$ converges weakly to P, the measure corresponding to $f_0 \in \mathcal{P}$, with probability 1 and $\mathbb{P}_n(|x|) = n^{-1}\sum_{i=1}^n |X_i| \to_{\text{a.s.}} \int |x|\, dP(x)$ by the strong law of large numbers. Thus $D_1(\mathbb{P}_n, P) \to_{\text{a.s.}} 0$. It follows from Theorem 4.3 that

$$\log \widehat{f}_n^0 = \operatorname*{argmax}_{\varphi\in\mathcal{C}_m}\left\{\mathbb{P}_n\varphi - \int e^{\varphi}\, d\lambda + 1\right\}$$
$$= \psi_m(\cdot|\mathbb{P}_n) = \widehat{\varphi}_n^0,$$

where, by the last part of Theorem 4.3, $\int |\widehat{f}_n^0 - f_m^0|\, d\lambda \to_{\text{a.s.}} 0$ and by the continuity

$$L_m(\mathbb{P}_n) = L(\psi_m(\cdot|\mathbb{P}_n), \mathbb{P}_n) = \mathbb{P}_n \log \widehat{f}_n^0$$
$$\to_{\text{a.s.}} L(\psi_m(\cdot|P), P) = P(\log f_m^0). \tag{4.16}$$

But then

$$\mathbb{P}_n \log \frac{\widehat{f}_n^0}{f_m^0} = \mathbb{P}_n \log \widehat{f}_n^0 - P \log f_m^0 - (\mathbb{P}_n \log f_m^0 - P \log f_m^0)$$
$$\to_{\text{a.s.}} 0 - 0 = 0$$

by (4.16) for the first term and the strong law of large numbers for the second term (using that $-\infty < L_m(P) < \infty$ by Theorem 4.1).

5. Simulations: Some comparisons and examples.

5.1. *Monte Carlo estimates of the distribution of* $\mathbb{D}$. To implement our likelihood ratio test and the corresponding new confidence intervals, we first conducted

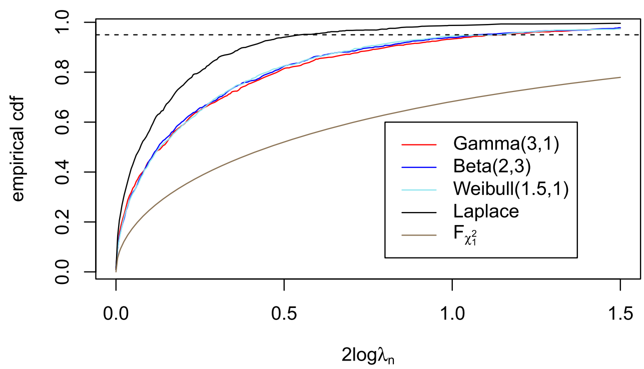

FIG. 1. *Monte Carlo distributions of* $2\log\lambda_n$ *for four distributions*, $n = 10^4$, $M = 5 \times 10^3$ *replications*, *together with the exact distribution function of* χ_1^2.

a Monte Carlo study of the null-hypothesis limit random variable, $\mathbb{D}$. We did this by simulating $M = 5 \times 10^3$ samples of $n = 10^4$ from the following distributions satisfying the key hypothesis ($\varphi_0''(m) < 0$) of Theorem 1.1: Gamma(3, 1), Beta(2, 3), Weibull(3/2, 1). The results are shown in Figure 1. Figure 1 also includes: (i) a plot of the known d.f. of a chi-square random variable with 1 degree of freedom, (which is the limiting distribution of the likelihood ratio test statistic for testing a one-dimensional parameter in a regular parametric model); (ii) a plot of the empirical distribution of $2\log\lambda_n$ for $M = 5 \times 10^3$ samples of size $n = 10^4$ drawn drawn from the Laplace density $2^{-1}\exp(-|x|)$ for which the assumption of Theorem 1 fails. In keeping with Theorem 1.1, the empirical results for all the distributions satisfying Theorem 1.1 are tightly clustered and in fact are almost visually indistinguishable, in spite of the fact that the various constants associated with these distributions are quite different, as shown in Table 1; in the next to last column $C(f_0) \equiv ((4!)^2 f_0(m)/(f_0''(m))^2))^{1/5}$ from (1.1), and in the last column SLC stands for "strongly log-concave" (see, e.g., Saumard and Wellner (2014)).

TABLE 1
Numerical characteristics of the distributions in the null hypothesis Monte Carlo study

Distribution	m	$f_0(m)$	$f_0''(m)$	$\varphi_0''(m)$	$C(f_0)$	SLC
N(0, 1)	0	$(2\pi)^{-1/2} = 0.3989\ldots$	$-(2\pi)^{-1/2}$	-1	4.28	Y
Gamma(3, 1)	2	$2e^{-2}$	$-e^{-2}$	$-1/2$	6.109	Y
Weibull(3/2, 1)	$3^{-2/3}$	$\frac{3^{2/3}}{2e^{1/3}}$	$-\frac{27}{8e^{1/3}}$	$-\frac{9\cdot 3^{1/4}}{4}$	2.36	N
Beta(2, 3)	3^{-1}	$\frac{16}{9}$	-24	$-\frac{27}{2}$	1.12	Y
Logistic	0	$1/4$	$-1/8$	$-1/2$	6.207	N
Gumbel	0	e^{-1}	$-e^{-1}$	-1	4.3545	N
χ_4^2	2	$\frac{1}{2e}$	$-\frac{1}{8e}$	$-\frac{1}{4}$	8.7091	N

TABLE 2
Estimated critical values d_α

α	0.25	0.20	0.15	0.10	0.05	0.01
d_α	0.40	0.49	0.61	0.79	1.11	1.92

Now for $\alpha \in (0, 1)$ let d_α satisfy $P(\mathbb{D} > d_\alpha) = \alpha$. Table 2 gives a few estimated values for d_α: These are based on 350,000 Monte Carlo simulations each based on simulating 1×10^6 observations from a standard normal. These values, and the simulated critical values for all $\alpha \in (0, 1)$, are available in the `logcondens.mode` package (Doss (2013a)) in R (R Core Team (2016)).

Banerjee and Wellner (2001) study a likelihood ratio test in the context of constraints based on monotonicity, and find a universal limiting distribution for their likelihood ratio test. Let $\mathbb{D}_{\text{mono}}$ be a random variable with this distribution. Comparison of the values in Table 2 with Table 2 of Banerjee and Wellner (2001) (particularly Method 2 in column 3 of that table) suggest, perhaps surprisingly, that $P(2\mathbb{D} \le t) \approx P(\mathbb{D}_{\text{mono}} \le t)$ for $t \in \mathbb{R}$. It would be quite remarkable if this held exactly. We do not have any explanation for this observed phenomenon.

5.2. *Comparisons via simulations.* Code to compute the mode-constrained log-concave MLE, implement a corresponding test, and invert the family of tests to form confidence intervals is available in the `logcondens.mode` package (Doss (2013a)). We can thus test our procedure and compare it to alternatives.

Romano (1988a) proposed and investigated two methods of forming confidence intervals for the mode of a unimodal density. His estimators of the mode and confidence intervals were based on the classical kernel density estimators of the density f going back to Parzen (1962). One method, which Romano called the "normal approximation method," is based on the limiting normality of the kernel density estimator of the mode, together with a plug-in estimator of the asymptotic variance. Romano's second method involved bootstrapping the mode estimator, and involved the choice of two bandwidths, one for the initial estimator to determine the mode, and a second (larger) bandwidth for the bootstrap sampling. The abstract of Romano (1988a) states: "In summary, the results are negative in the sense that a straightforward application of a naive bootstrap yields invalid inferences. In particular, the bootstrap fails if resampling is done from the kernel density estimate." That is, one must use a second (larger) bandwidth for the bootstrap resampling to achieve valid inference. This thus necessitates selection of two tuning parameters for the bootstrap procedure. Romano (1988a) notes in summarizing his simulation results:

> ... but the problem of constructing a confidence interval for the mode for smaller sample sizes remains a challenging one. In summary, the simulations reinforce the idea that generally automatic methods like the bootstrap need mathematical and numerical justification before their use can be recommended.

The bootstrap simulations that Romano (1988a) refers to in the previous quote are based on an underlying $N(0,1)$ or a χ_4^2 distribution with a sample size of $n = 100$. Romano (1988a) also performs simulations for the normal approximation method for the same underlying distributions and based on the same sample size. For the normal approximation method, a grid of bandwidths h are used for the simulation. For the bootstrap, a matrix of bandwidth pairs (h, b) (one for estimation, one for resampling) are used. Monte Carlo estimates of coverage probabilities are presented in Tables 1–4 of Romano (1988a).

In Figure 2, we consider the case of a true underlying χ_4^2 distribution, and we plot all the estimated coverage probabilities of Romano's bootstrap CI's (blue; these are from Table 4 of Romano (1988a)) together with the target (ideal) coverage (green line) and the estimated coverage probabilities of our likelihood ratio (LR) based CI's (magenta). As can be seen, the estimated coverage probabilities of our LR-based procedure are reasonably close to the target values without requiring any bandwidth choice.

Corresponding comparison plots based on Tables 1, 2 and 3 of Romano (1988a), as well as tables of the simulated coverage probabilities, are given in Doss and Wellner (2016b). We do not include them here due to space constraints. Doss and Wellner (2016b) also includes a Monte Carlo simulation study of lengths of the CIs in some settings.

Methods of bandwidth selection for various problems have received considerable attention in the period since Romano (1988a); see especially Léger and Romano (1990), Mammen and Park (1997), Härdle, Marron and Wand (1990), Hall

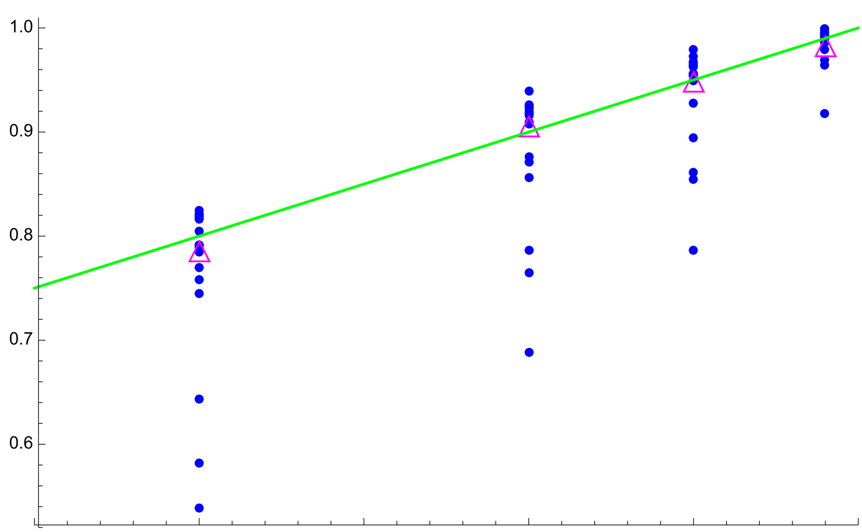

FIG. 2. *Coverage probabilities, Romano's Table 4 compared with LR coverage probabilities,* χ_4^2 *data, Bootstrap confidence intervals (blue dots); LR confidence intervals (magenta triangles). The green line is the nominal level.*

and Johnstone (1992), Ziegler (2001), Hazelton (1996a, 1996b) and Samworth and Wand (2010). Although bandwidth selection in connection with mode estimation is mentioned briefly by Léger and Romano (1990) (see their last paragraph, page 734), we are not aware of any specific proposal or detailed study of bandwidth selection methods in the problem of confidence intervals for the mode of a unimodal density. For this reason, we have not undertaken a full comparative study of possible methods here. Further comparisons of our LR based confidence intervals with methods based on kernel density estimates of the type studied by Romano (1988a) but incorporating current state of the art bandwidth selection procedures will be of interest.

5.3. *Comparisons via data examples*. We used our procedure for formation of modal confidence intervals (CIs) on two real data sets, the rotational velocities of stars from the Bright Star Catalogue (Hoffleit and Warren (1991)) and daily log returns for the S&P 500 stock market index. To see the former, see Section 5.3 of Doss and Wellner (2016b). Here, we discuss the 1006 daily log returns for the S&P 500 stock market index from January 1, 2003, to December 29, 2006. In Figure 3, we plot the data, a kernel density estimate with bandwidth 0.13 (Sheather and Jones (1991)), the log-concave MLE, and the 95% confidence interval for the mode given by our likelihood ratio statistic. We also plot the maximum likelihood Gaussian density estimate, for comparison. The sample mean is 0.04, the sample median is 0.081 and the log-concave mode estimate is 0.17. A 95% CIs for the mean is $[-0.004, 0.09]$ and a 95% CI for the median is $[0.037, 0.122]$ (Hogg and Craig (1970), pages 539–540). Our likelihood ratio CI for the mode is $[0.10, 0.21]$. Note that our confidence interval for the mode excludes 0 and does not intersect with the CI for the mean. Thus, our procedure highlights some interesting features of the data and provides evidence for its nonnormality. Also note that the lengths of the mean, median and our LR-based mode CI are 0.094, 0.085 and 0.11. Thus, despite the fact that our mode estimator does not generally have a $n^{-1/2}$ rate of convergence, the three confidence intervals are of fairly similar length on a data

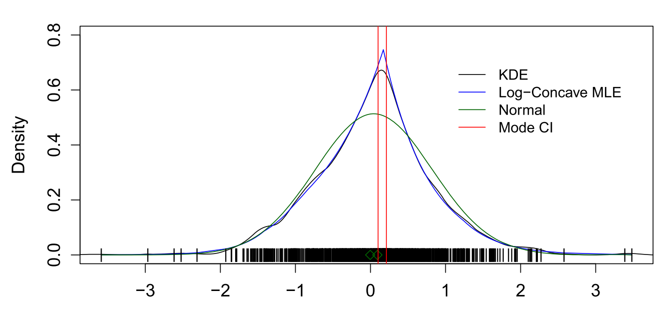

FIG. 3. 1006 *S&P* 500 *daily log returns for the years* 2003–2006.

set with 1006 observations, which is encouraging for our mode CI procedure and for any future extensions (e.g., mode regression CIs).

6. Further problems and potential developments.

6.1. *Uniformity and rates of convergence.* There is a long line of research giving negative results concerning nonparametric estimation, starting with Bahadur and Savage (1956), Blum and Rosenblatt (1966), Singh (1963, 1968) and continuing with Donoho (1988) and Pfanzagl (1998, 2000). In particular, Pfanzagl (2000) considers a general setting involving estimators or confidence limits with optimal convergence rate $n^{-\rho}$ with $0 < \rho < 1/2$. He shows, under weak additional conditions, that: (i) there do not exist estimators which converge locally uniformly to a limit distribution; and (ii) there are no confidence limits with locally uniform asymptotic coverage probability. As an example, he considers the mode of probability distributions P on $\mathbb{R}$ with corresponding densities p having a unique mode $M(p)$ and continuous second derivative in a neighborhood of $M(p)$. Pfanzagl (2000) also reproves the result of Has'minskii (1979) to the effect that the optimal rate of convergence of a mode estimator for such a class is $n^{-1/5}$. In this respect, we note that Balabdaoui, Rufibach and Wellner (2009) established a comparable lower bound for estimation of the mode in the class of log-concave densities with continuous second derivative at the mode; they obtained a constant which matches (up to absolute constants) the pointwise (fixed P) behavior of the plug-in log-concave MLE of the mode. Romano (1988b) gives a detailed treatment of minimax lower bounds for estimation of the mode under smoothness and curvature assumptions: assuming a bounded derivative of order p in a neighborhood of the mode $M(f_0)$, Romano shows in his Theorem 3.1 that the minimax rate for estimation of $M(f_0)$ is n^{-r} where $r = (p-1)/(2p+1)$. He also shows that when $p = 3$, the rate $n^{-2/7}$ can be achieved by a kernel density estimator.

Our approach here has been to construct reasonable confidence intervals with pointwise (in P or density p) correct asymptotic coverage without proof of any local uniformity properties. In view of the recent uniform rate results of Kim and Samworth (2016), we suspect that our new confidence intervals *will* (eventually) be shown to have some uniformity of convergence in their coverage probabilities over appropriate subclasses of the class of log-concave densities, but we leave the uniformity issues to future work.

6.2. *Some further directions and open questions.* We now turn to discussion of some difficulties and potential for further work.

6.2.1. *Relaxing the second derivative assumption.* As noted in the previous subsection, most of the available research concerning inference for $M(f)$ has assumed $f \in C^2(m, \mathrm{loc})$ and $f^{(2)}(M(f)) < 0$. Second derivative-type assumptions of this type are made in Parzen (1962), Has'minskii (1979), Eddy (1980), Donoho and Liu (1991), Romano (1988a, 1988b) and Pfanzagl (2000). Exceptions include Müller (1989), Ehm (1996), Herrmann and Ziegler (2004), Balabdaoui, Rufibach and Wellner (2009).

What happens if the second derivative curvature assumption does not hold, but instead is replaced by something either stronger or weaker, such as

$$f(m) - f(x) \le C|x - m|^r$$

for some C where $1 \le r < 2$ (in the "stronger" case) or $2 < r < \infty$ (in the "weaker" case)? It is natural to expect that it is easier to form confidence intervals for m when $1 \le r < 2$ holds, but that it is harder to form confidence intervals for m when $2 < r < \infty$. In fact, Balabdaoui, Rufibach and Wellner (2009), page 1313, gives the following result: if $f = \exp(\varphi)$ with φ concave and where $\varphi^{(j)}(m) = 0$ for $j = 2, \dots, k-1$ but $\varphi^{(k)}$ exists and is continuous in a neighborhood of m with $\varphi^{(k)}(m) \neq 0$, then

$$n^{1/(2k+1)}(\widehat{M}_n - m) \to_d C_k\big(f(m), \varphi^{(k)}(m)\big) M\big(H_k^{(2)}\big).$$

Thus the convergence rate of the log-concave MLE of the mode is slower as k increases. [On the other hand, by Theorem 2.1 of Balabdaoui, Rufibach and Wellner (2009), page 1305, the convergence rate of the MLE $\widehat{f}_n$ of f at m (and in a local neighborhood of m) is *faster*:

$$n^{k/(2k+1)}\big(\widehat{f}_n(m) - f(m)\big) \to_d c_k(m, f) H_k^{(2)}(0).]$$

Furthermore, the sketch of the proof of the limiting distribution of the likelihood ratio statistic in Section 4.1 (ignoring any remainder terms) together with the results of Balabdaoui, Rufibach and Wellner (2009), suggest that $2 \log \lambda_n \to_d \mathbb{D}_k$ under $f \in \mathcal{P}_m \cap \mathcal{Z}_k$ where

$$\mathcal{Z}_k = \{f \in \mathcal{P} : \varphi^{(j)}(m) = 0, j = 2, \dots, k-1, \varphi^{(k)}(m) \neq 0, \varphi \in C^k(m, \mathrm{loc})\}$$

and where with φ_k and φ_k^0 denoting the local limit processes in the white noise model (1.2) with drift term $g_0(t) = -12t^2$ replaced by $-(k+2)(k+1)|t|^k$,

$$\mathbb{D}_k \equiv \int \big\{\big(\widehat{\phi}_k(v)\big)^2 - \big(\widehat{\phi}_k^0(v)\big)^2\big\}\, dv.$$

We provide Monte Carlo evidence in support of this conjecture, by simulating $2 \log \lambda_n$ based on some parent distributions with $k \neq 2$. The results are given in Figure 4. Figure 4 contains empirical distributions of $2 \log \lambda_n$ (with $n = 10^4$ and $M = 5 \times 10^3$) for 9 parent distributions, as well as a plot of the df of a χ_1^2 random variable; all of the curves from Figure 1 are present, including $F_{\chi_1^2}$, the Laplace (with $k = 1$), the standard normal, Gamma$(3, 1)$, Beta$(2, 3)$ and Weibull$(1.5, 1)$ (all four having $k = 2$). We also add four parent distributions with $k > 2$. We include parent densities proportional to $\exp\{-|x|^j/j\}$ for $x \in \mathbb{R}$, labeled "Subbotin(j)," $j = 3, 4$ (having $k = j$). We also include parent densities proportional to $1 - |x|^j$ for $x \in [-1, 1]$, labeled "Bump(j)," $j = 3, 4$ (with $k = j$). The

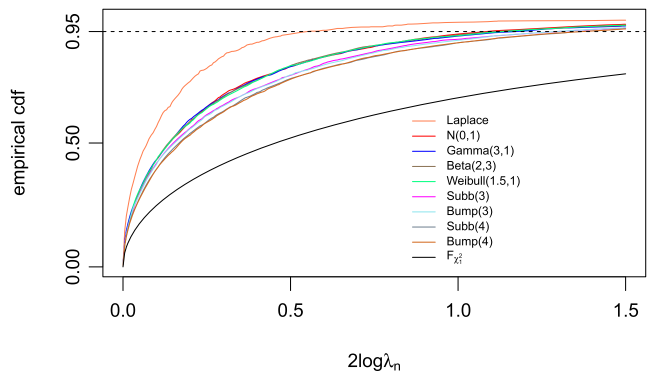

FIG. 4. *Empirical distributions of* $2\log\lambda_n$ *for* $f \in \mathcal{P}_m \cap \mathcal{Z}_k$, $k \in \{2, 3, 4\}$ *with* $n = 10^4$ *and* $M = 5 \times 10^3$ *replications.*

(Monte Carlo estimators of) df's based on the parent distributions with $k = 3$ (estimating $\mathbb{D}_3$) are grouped together, and the df's based on the parent distributions with $k = 4$ (estimating $\mathbb{D}_4$) are similarly grouped together. Note that the (Monte Carlo estimator of) the distribution of $\mathbb{D}_3$ seems to be stochastically larger than the (Monte Carlo estimator) of the distribution of $\mathbb{D}_2 \equiv \mathbb{D}$, and that the distribution of $\mathbb{D}_4$ is apparently stochastically larger than that of $\mathbb{D}_3$. This raises several possibilities:

Option 1: It seems likely that by choosing a critical value from the distribution of $\mathbb{D}_6$ (say), that the resulting confidence intervals will have correct coverage for $f \in \mathcal{P} \cap \mathcal{Z}_6$ with conservative coverage if we happen to have $f \in \mathcal{P} \cap \mathcal{Z}_2$ (in which case critical points from $\mathbb{D} = \mathbb{D}_2$ would have sufficed), and anticonservative coverage if the true f belongs to $\mathcal{P} \cap (\mathcal{Z}_k \setminus \mathcal{Z}_6)$ for some $k \geq 8$.

Option 2: Try to construct an adaptive procedure which first estimates k (the degree of "flatness") of the true $f \in \mathcal{P}$ (by $\hat{k}$ say), and then choose a critical point from the distribution of $\mathbb{D}_{\hat{k}}$.

We leave the investigation of both of these possibilities to future work.

6.2.2. *Relaxing the assumption of log-concavity.* It would also be of interest to relax the assumption of log-concavity used in the developments here. It would be very desirable to allow f_0 to be a completely arbitrary unimodal density, and allow the smoothness at the mode $M(f_0)$ to vary as noted in the previous subsection. As a more realistic replacement for this ambitious goal, we might instead consider enlarging from the class of log-concave densities to some class of $s-$concave densities, $\mathcal{P}_s$ with $-1 < s < 0$; that is, densities of the form $f_0 = \phi_0^{1/s}$ with $\phi_0 > 0$ convex; see, for example, Koenker and Mizera (2010), Doss and Wellner (2016a), and Han and Wellner (2016). Extensions in this direction will likely require further study of the Rényi divergence estimators studied in Han and Wellner (2016) and mode-constrained versions thereof. An interesting possible connection is that for the classes of $\alpha-$stable densities $\mathcal{S}_\alpha$ with $0 < \alpha \le 2$, we know that $f_0 \in \mathcal{S}_\alpha$ is unimodal. Moreover, it is also known from Hall (1984) that for the symmetric $\alpha-$stable distributions f_0'' exists in a neighborhood of the mode $m = M(f_0)$, and $f_0''(m) < 0$. It is apparently not known if the $\alpha-$stable densities are $s-$concave for some $s \in (-1, 0]$, even though this obviously holds in the (few) examples for which an explicit formula for the density $f_0 \in \mathcal{S}_\alpha$ is available: for example, for $f_0 =$ Cauchy, $f_0 \in \mathcal{S}_1 \cap \mathcal{P}_{-1/2}$, while for Lévy's completely asymmetric stable law, $f_0 \in \mathcal{S}_{1/2} \cap \mathcal{P}_{-2/3}$, and of course, for $f_0 =$ Gaussian, $f_0 \in \mathcal{S}_2 \cap \mathcal{P}_0$.

6.2.3. *Mode inference in other contexts.* The methods developed in this paper raise several questions about other settings in which inference about a convex function may be of interest.

(a) Can we do inference for the maxima or minima in the contexts of estimation of intensity functions, of (bathtub-shaped) hazard functions (Jankowski and Wellner (2009)), or of regression functions? For instance, let $Y_i = r(x_i) + \varepsilon_i$ where ε_i are mean zero i.i.d. observations and x_i are fixed numbers in $\mathbb{R}$. If we assume r to be convex, then much is known about the least-squares estimator $\widehat{r}_n$ of r; see, for example, Hildreth (1954), Hanson and Pledger (1976), Mammen (1991) and Groeneboom, Jongbloed and Wellner (2001a, 2001b). Can an argmin-constrained estimator $\widehat{r}_n^0$ be developed, in analogy with the estimator $\widehat{f}_n^0$, and used to develop likelihood ratio-based (or rather, residual sum-of-squares) tests and intervals for the location of the minimum of r? In such a problem, we conjecture that the universal component of the limit distribution of $\widehat{r}_n^0(m)$ will be the same as that studied in Theorem 1.1.

(b) Can the techniques used here be applied to form tests and intervals for the *value* (or height) of a concave function, f_0, rather than argmax? Here, f_0 could be a log-concave density or a concave regression function (and other settings could be of interest). That is, can we develop an estimator $\widehat{f}_n^0$ based on the constraint that f satisfies $f(x_0) = y_0$ for x_0, y_0 fixed and use $\widehat{f}_n^0$ with an unconstrained estimator $\widehat{f}_n$ to form a likelihood ratio test for $f_0(x_0)$? In the case where f_0 is a concave regression function, such a program has recently been studied by Doss (2018). Can this be extended to the density case, where f_0 is log-concave?

(c) Can inference for the mode be extended to semiparametric settings? For example can we form tests/intervals for the location of the minimum of an unknown convex "link" function m_0 in a single index model, $Y = m_0(\theta_0' X) + \varepsilon$, where $X \in \mathbb{R}^d$, $Y \in \mathbb{R}$, $E(\varepsilon|X) = 0$ and m_0 is assumed to be convex (Chen and Samworth (2016), Kuchibhotla, Patra and Sen (2017))? Can we form tests/intervals for a modal regression function, that is, for m_0 where $Y = m_0(X) + \varepsilon$ where ε has *mode* 0?

6.2.4. *Beyond dimension* $d = 1$. It seems natural to consider generalizations of the present methods to the case of multivariate log-concave and $s-$concave densities. While there is a considerable amount of work on estimation of multivariate modes, mostly via kernel density estimation, much less seems to be available in terms of confidence sets or other inference tools. For some of this see, for example, Tsybakov (1990), Abraham, Biau and Cadre (2003), Kim (1994), Klemelä (2005), Konakov (1973), Sager (1978), Samanta (1973). On the other hand, apparently very little is known about the multivariate mode estimator $M(\widehat{f}_n)$ where $\widehat{f}_n$ is the log-concave density estimator for f on $\mathbb{R}^d$ studied by Cule, Samworth and Stewart (2010) and Cule and Samworth (2010). Further study of this estimator will very likely require considerable development of new methods for study of the pointwise and local properties of the log-concave density estimator $\widehat{f}_n$ when $d \geq 2$.

Acknowledgments. We owe thanks to Lutz Dümbgen for several helpful conversations. This work was partially supported by a grant to the second author from the Simons Foundation and was carried out in part during a visit to the Isaac Newton Institute.

SUPPLEMENTARY MATERIAL

Supplement to "Inference for the mode of a log-concave density" (DOI: 10.1214/18-AOS1770SUPP; .pdf). In the supplement, we provide additional proofs and technical details that were omitted from the main paper.

REFERENCES

ABRAHAM, C., BIAU, G. and CADRE, B. (2003). Simple estimation of the mode of a multivariate density. *Canad. J. Statist.* **31** 23–34. MR1985502

BAHADUR, R. R. and SAVAGE, L. J. (1956). The nonexistence of certain statistical procedures in nonparametric problems. *Ann. Math. Stat.* **27** 1115–1122. MR0084241

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

BANERJEE, M. (2007). Likelihood based inference for monotone response models. *Ann. Statist.* **35** 931–956. MR2341693

BANERJEE, M. and WELLNER, J. A. (2001). Likelihood ratio tests for monotone functions. *Ann. Statist.* **29** 1699–1731. MR1891743

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

BLUM, J. R. and ROSENBLATT, J. (1966). On some statistical problems requiring purely sequential sampling schemes. *Ann. Inst. Statist. Math.* **18** 351–355. MR0205414

CHEN, Y. and SAMWORTH, R. J. (2016). Generalized additive and index models with shape constraints. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 729–754. MR3534348

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

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

DHARMADHIKARI, S. and JOAG-DEV, K. (1988). *Unimodality, Convexity, and Applications. Probability and Mathematical Statistics*. Academic Press, Boston, MA. MR0954608

DONOHO, D. L. (1988). One-sided inference about functionals of a density. *Ann. Statist.* **16** 1390–1420. MR0964930

DONOHO, D. L. and LIU, R. C. (1991). Geometrizing rates of convergence. II, III. *Ann. Statist.* **19** 633–667, 668–701. MR1105839

DOSS, C. R. (2013a). logcondens.mode: Compute MLE of log-concave density on R with fixed mode, and perform inference for the mode. R package version 1.0.1. Available at http://CRAN.R-project.org/package=logcondens.mode.

DOSS, C. R. (2013b). Shape-constrained inference for concave-transformed densities and their modes. Ph.D. thesis, Univ. Washington, Seattle, WA.

DOSS, C. R. (2018). Concave regression: Value-constrained estimation and likelihood ratio-based inference. *Math. Program.* To appear. DOI:10.1007/s10107-018-1338-5.

DOSS, C. R. and WELLNER, J. A. (2016a). Global rates of convergence of the MLEs of log-concave and s-concave densities. *Ann. Statist.* **44** 954–981. MR3485950

DOSS, C. R. and WELLNER, J. A. (2016b). Inference for the mode of a log-concave density. Available at arXiv:1611.10348v2.

DOSS, C. R. and WELLNER, J. A. (2018). Log-concave density estimation with symmetry or modal constraints. Available at arXiv:1611.10335v3.

DOSS, C. R. and WELLNER, J. A. (2019). Supplement to "Inference for the mode of a log-concave density." DOI:10.1214/18-AOS1770SUPP.

DÜMBGEN, L. and RUFIBACH, K. (2009). Maximum likelihood estimation of a log-concave density and its distribution function: Basic properties and uniform consistency. *Bernoulli* **15** 40–68. MR2546798

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

EDDY, W. F. (1980). Optimum kernel estimators of the mode. *Ann. Statist.* **8** 870–882. MR0572631

EHM, W. (1996). Adaptive kernel estimation of a cusp-shaped mode. In *Applied Mathematics and Parallel Computing* 109–120. Physica, Heidelberg. MR1469264

GROENEBOOM, P. and JONGBLOED, G. (2015). Nonparametric confidence intervals for monotone functions. *Ann. Statist.* **43** 2019–2054. MR3375875

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

GROENEBOOM, P., JONGBLOED, G. and WELLNER, J. A. (2001b). Estimation of a convex function: Characterizations and asymptotic theory. *Ann. Statist.* **29** 1653–1698. MR1891742

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

HALL, P. and JOHNSTONE, I. (1992). Empirical functional and efficient smoothing parameter selection. *J. Roy. Statist. Soc. Ser. B* **54** 475–530. With discussion and a reply by the authors. MR1160479

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

HANSON, D. L. and PLEDGER, G. (1976). Consistency in concave regression. *Ann. Statist.* **4** 1038–1050. MR0426273

HÄRDLE, W., MARRON, J. S. and WAND, M. P. (1990). Bandwidth choice for density derivatives. *J. Roy. Statist. Soc. Ser. B* **52** 223–232. MR1049312

HAS'MINSKII, R. Z. (1979). Lower bound for the risks of nonparametric estimates of the mode. In *Contribution to Statistics: Jaroslav Hájek Memorial Volume* 91–97. Reidel, Dordrecht. MR0561262

HAZELTON, M. (1996a). Bandwidth selection for local density estimators. *Scand. J. Stat.* **23** 221–232. MR1394655

HAZELTON, M. L. (1996b). Optimal rates for local bandwidth selection. *J. Nonparametr. Stat.* **7** 57–66. MR1618191

HERRMANN, E. and ZIEGLER, K. (2004). Rates on consistency for nonparametric estimation of the mode in absence of smoothness assumptions. *Statist. Probab. Lett.* **68** 359–368. MR2081114

HILDRETH, C. (1954). Point estimates of ordinates of concave functions. *J. Amer. Statist. Assoc.* **49** 598–619. MR0065093

HOFFLEIT, D. and WARREN, W. H. JR. (1991). *Bright Star Catalogue*. Yale Univ. Observatory, New Haven, CT. Available at http://adsabs.harvard.edu/abs/1995yCat.5050....0H.

HOGG, R. V. and CRAIG, A. T. (1970). *Introduction to Mathematical Statistics*. 3rd ed. Macmillan, New York. MR0251823

IBRAGIMOV, I. A. (1956). On the composition of unimodal distributions. *Theory Probab. Appl.* **1** 225–260. MR0087249

JANKOWSKI, H. K. and WELLNER, J. A. (2009). Nonparametric estimation of a convex bathtub-shaped hazard function. *Bernoulli* **15** 1010–1035. MR2597581

KIM, J. (1994). Asymptotic theory for multi-dimensional mode estimator. *J. Korean Statist. Soc.* **23** 251–269. MR1327618

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

KLEMELÄ, J. (2005). Adaptive estimation of the mode of a multivariate density. *J. Nonparametr. Stat.* **17** 83–105. MR2112688

KOENKER, R. and MIZERA, I. (2010). Quasi-concave density estimation. *Ann. Statist.* **38** 2998–3027. MR2722462

KONAKOV, V. D. (1973). The asymptotic normality of the mode of multivariate distributions. *Teor. Veroyatn. Primen.* **18** 836–842. MR0336874

KUCHIBHOTLA, A. K., PATRA, R. K. and SEN, B. (2017). Efficient estimation in convex single index models. Available at arXiv:1708.00145v2.

LÉGER, C. and ROMANO, J. P. (1990). Bootstrap choice of tuning parameters. *Ann. Inst. Statist. Math.* **42** 709–735. MR1089472

MALLOWS, C. L. (1972). A note on asymptotic joint normality. *Ann. Math. Stat.* **43** 508–515. MR0298812

MAMMEN, E. (1991). Nonparametric regression under qualitative smoothness assumptions. *Ann. Statist.* **19** 741–759. MR1105842

MAMMEN, E. and PARK, B. U. (1997). Optimal smoothing in adaptive location estimation. *J. Statist. Plann. Inference* **58** 333–348. MR1450020

MÜLLER, H.-G. (1989). Adaptive nonparametric peak estimation. *Ann. Statist.* **17** 1053–1069. MR1015137

PAL, J. K., WOODROOFE, M. and MEYER, M. (2007). Estimating a Pólya frequency function$_2$. In *Complex Datasets and Inverse Problems*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **54** 239–249. IMS, Beachwood, OH. MR2459192

PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Stat.* **33** 1065–1076. MR0143282

PFANZAGL, J. (1998). The nonexistence of confidence sets for discontinuous functionals. *J. Statist. Plann. Inference* **75** 9–20. MR1671670

PFANZAGL, J. (2000). On local uniformity for estimators and confidence limits. *J. Statist. Plann. Inference* **84** 27–53. MR1747496

R CORE TEAM (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

ROCKAFELLAR, R. T. (1970). *Convex Analysis*. *Princeton Mathematical Series* **28**. Princeton Univ. Press, Princeton, NJ. MR0274683

ROMANO, J. P. (1988a). Bootstrapping the mode. *Ann. Inst. Statist. Math.* **40** 565–586. MR0964293

ROMANO, J. P. (1988b). On weak convergence and optimality of kernel density estimates of the mode. *Ann. Statist.* **16** 629–647. MR0947566

RUFIBACH, K. (2006). Log-concave density estimation and bump hunting for IID observations. Ph.D. thesis, Univ. Bern and Gottingen.

SAGER, T. W. (1978). Estimation of a multivariate mode. *Ann. Statist.* **6** 802–812. MR0491553

SAMANTA, M. (1973). Nonparametric estimation of the mode of a multivariate density. *South African Statist. J.* **7** 109–117. MR0331618

SAMWORTH, R. J. and WAND, M. P. (2010). Asymptotics and optimal bandwidth selection for highest density region estimation. *Ann. Statist.* **38** 1767–1792. MR2662359

SAUMARD, A. and WELLNER, J. A. (2014). Log-concavity and strong log-concavity: A review. *Stat. Surv.* **8** 45–114. MR3290441

SHEATHER, S. J. and JONES, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *J. Roy. Statist. Soc. Ser. B* **53** 683–690. MR1125725

SILVERMAN, B. W. (1982). On the estimation of a probability density function by the maximum penalized likelihood method. *Ann. Statist.* **10** 795–810. MR0663433

SINGH, R. (1963). Existence of bounded length confidence intervals. *Ann. Math. Stat.* **34** 1474–1485. MR0155394

SINGH, R. (1968). Non-existence of estimates of prescribed accuracy in fixed sample size. *Canad. Math. Bull.* **11** 135–139. MR0232487

TSYBAKOV, A. B. (1990). Recurrent estimation of the mode of a multidimensional distribution. *Problemy Peredachi Informatsii* **26** 38–45. MR1051586

WALTHER, G. (2002). Detecting the presence of mixing with multiscale maximum likelihood. *J. Amer. Statist. Assoc.* **97** 508–513. MR1941467

WILKS, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Stat.* **9** 60–62.

ZIEGLER, K. (2001). On bootstrapping the mode in the nonparametric regression model with random design. *Metrika* **53** 141–170. MR1839003

SCHOOL OF STATISTICS
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MINNESOTA 55455
USA
E-MAIL: cdoss@stat.umn.edu
URL: http://users.stat.umn.edu/~cdoss/

DEPARTMENT OF STATISTICS
UNIVERSITY OF WASHINGTON
BOX 354322
SEATTLE, WASHINGTON 98195-4322
USA
E-MAIL: jaw@stat.washington.edu
URL: http://www.stat.washington.edu/jaw/