Nonparametric Estimation of Log – Concave Densities

Kaspar Rufibach
Biostatistics unit, ISPMZ
University of Zurich

Based on joint work with
Lutz Dümbgen (Bern), Fadoua Balabdaoui (Paris), and Jon A. Wellner (Seattle)

1. Estimation of log-concave densities
2. Uniform consistency and rate of convergence
3. Limit distribution at a fixed point
4. Estimation of the mode
5. Further results

f log-concave if

$$f(x) = \exp \varphi(x)$$

for some concave function $\varphi : \mathbb{R} \to [-\infty, \infty)$.

- Log-concave densities are unimodal.
- f, g log-concave $\Rightarrow$ $f * g$ log-concave (Prekopa, 1971).
- f unimodal, g log-concave $\Rightarrow$ $f * g$ unimodal (Ibragimov, 1956).
- $f * g$ unimodal for any unimodal g $\Rightarrow$ f log-concave (Ibragimov, 1956).
- $\int \exp(t_0|\varphi|) < \infty$ $(t_0 < 1)$ $\Rightarrow$ light tails.

Goal: Estimate log-concave f from i.i.d. data (non-parametrically)

Why estimate a log-concave density?

- No way of estimating a unimodal density via NPML! Log-concavity is a surrogate.
- Normal, Uniform, $\Gamma(r, \lambda)$ for $r \geq 1$, beta(a, b) for $a \geq 1$, $b \geq 1$, generalized Pareto, Gumbel, Fréchet, logistic or Laplace are log-concave.
- Tests for multimodality and mixing based on semiparametric model with densities of the form $f_{c,\varphi}(x) = \exp(\varphi(x) + cx^2)$, $c > 0$ (Walther, 2001, 2002).
- Fully automatic estimator: no choice of bandwidth, kernel, prior, tuning parameter, ...
- Properties of derived functions: f log-concave $\Rightarrow$ hazard $\nearrow$

Non-parametric maximum likelihood estimation

$X_1 < X_2 < \ldots < X_n$ i.i.d. rv with CDF F, log–density function φ

Maximize adjusted log–likelihood function to get $\widehat{\varphi}_n$:

$$
\begin{aligned}
\widehat{\varphi}_n \quad &:= \quad \underset{\varphi \text{ concave}, \int \exp(\varphi)=1}{\arg\max} \left\{ \log\left(\prod_{i=1}^{n} f(X_i) \right) \right\} \\
&= \quad \underset{\varphi \text{ concave}}{\arg\max} \left\{ \sum_{i=1}^{n} \log f(X_i) - \lambda \int \exp \varphi(x) \, \mathrm{d}x \right\} \\
&= \quad \underset{\varphi \text{ concave}}{\arg\max} \left\{ \underbrace{\sum_{i=1}^{n} \varphi(X_i)}_{\text{log-likelihood}} - \underbrace{n \int \exp \varphi(x) \, \mathrm{d}x}_{\text{Lagrange term}} \right\} \\
&=: \quad \underset{\varphi \text{ concave}}{\arg\max} \, \psi_n(\varphi)
\end{aligned}
$$

λ: Lagrange multiplier, $\lambda = n$ guarantees PDF ("Silverman's trick")

Density estimator: $\widehat{f}_n = \exp(\widehat{\varphi}_n)$

Theorem (Existence, uniqueness, form)

- $\widehat{\varphi}_n$ *exists and is unique.*
- $\widehat{\varphi}_n$ *is continuous and piecewise linear on* $[X_1, X_n]$*, has only knots in* $\{X_1, \ldots, X_n\}$.
- $\widehat{f}_n \equiv 0$ *on* $\mathbb{R} \setminus [X_1, X_n]$.

With this theorem:

$$\widehat{\varphi}_n = \underset{\varphi\in\mathcal{C}}{\arg\max}\Big\{\sum_{i=1}^{n}\varphi(X_i) - n\sum_{k=1}^{n-1}\Delta X_{k+1}J(\varphi(X_k),\varphi(X_{k+1}))\Big\}$$

where $\Delta X_k = X_k - X_{k-1}$ and

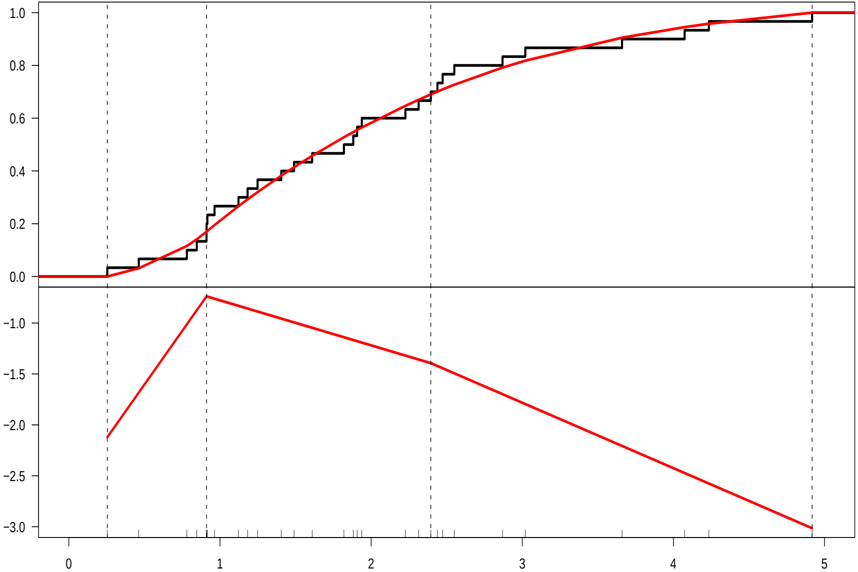
$$J(r,s) = \int_0^1 \exp\Big((1-t)r + ts\Big)\,\mathrm{d}t$$

$$\mathcal{C} = \Big\{\boldsymbol{\varphi}\in\mathbb{R}^n \;:\; \frac{\Delta\varphi_i}{\Delta X_i}\geq 0 \quad \text{for all } i = 2,\ldots,n\Big\}.$$

$\Rightarrow$ finite-dimensional optimization problem with $n-1$ linear constraints!

Computation of $\widehat{\varphi}_n$: iterative convex minorant algorithm, active set algorithm

$\Gamma(2,1)$-sample of size $n = 30$



Characterizations and further properties of $\widehat{\varphi}_n$

Reminder: Concave functional

$$\psi_n(\widehat{\varphi}_n) \quad = \quad \int \varphi(x)\,\mathrm{d}\mathbb{F}_n(x) - \int \exp\varphi(x)\,\mathrm{d}x.$$

Considering

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \psi_n(\widehat{\varphi}_n + t\Delta) \quad \le \quad 0$$

for all t such that $\widehat{\varphi}_n + t\Delta$ concave yields

$$\int \Delta\,\mathrm{d}\mathbb{F}_n \quad \le \quad \int \Delta\,\mathrm{d}\widehat{F}_n$$

where $\mathbb{F}_n$ is the ECDF of $X_1, \ldots, X_n$ and

$$\widehat{F}_n(x) \quad := \quad \int_{X_1}^{x} \widehat{f}_n(t)\,\mathrm{d}t.$$

Choices of $\Delta(x) = \pm x,\ \Delta(x) = -x^2$ yields

$$\begin{aligned} \mathrm{Mean}(\widehat{F}_n) &= \mathrm{Mean}(\mathbb{F}_n) \\ \mathrm{Var}(\widehat{F}_n) &\leq \mathrm{Var}(\mathbb{F}_n). \end{aligned}$$

Set of knots of $\widehat{\varphi}_n$:

$$\mathrm{knots}(\widehat{\varphi}_\mathrm{n}) \;:=\; \{t \;:\; \widehat{\varphi}_n'(t-) > \widehat{\varphi}_n'(t+)\} \begin{cases} \supset \{X_1, X_n\} \\ \subset \{X_1, \ldots, X_n\}. \end{cases}$$

If Δ is continuous, piecewise linear, knots only in $\mathrm{knots}(\widehat{\varphi}_\mathrm{n})$ we even get

$$\int \Delta(x)\, \mathrm{d}\mathbb{F}_n(x) \;=\; \int \Delta(x) \widehat{f}_n(x)\, \mathrm{d}x.$$

Theorem (Characterization of $\widehat{F}_n$)

For $a < t < b$ with $a, b \in \mathrm{knots}(\widehat{\varphi}_\mathrm{n})$,

$$\begin{aligned} \int_a^t \mathbb{F}_n(x)\,\mathrm{d}x &\geq \int_a^t \widehat{F}_n(x)\,\mathrm{d}x \\ \int_t^b \mathbb{F}_n(x)\,\mathrm{d}x &\leq \int_t^b \widehat{F}_n(x)\,\mathrm{d}x \\ \int_a^b \mathbb{F}_n(x)\,\mathrm{d}x &= \int_a^b \widehat{F}_n(x)\,\mathrm{d}x. \end{aligned}$$

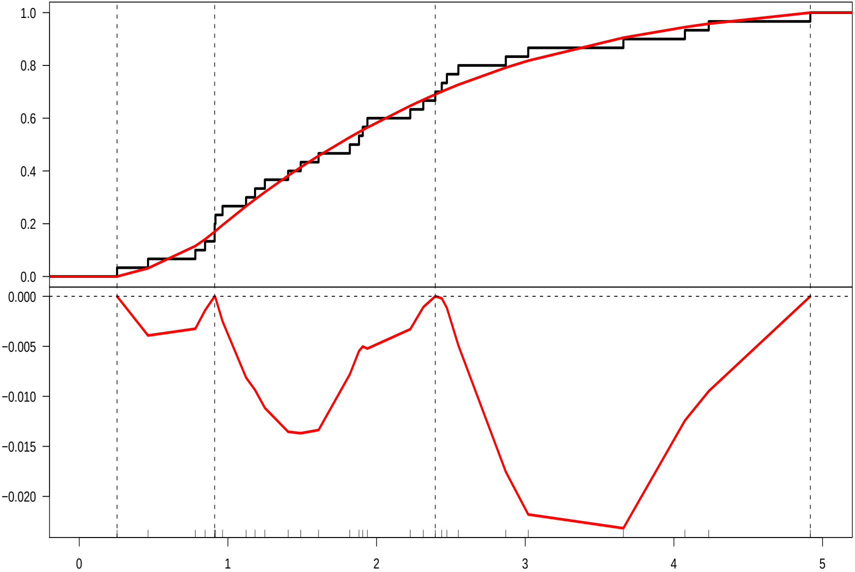
Corollary

This theorem implies $\widehat{F}_n(X_1) = 0$, $\widehat{F}_n(X_n) = 1$ and

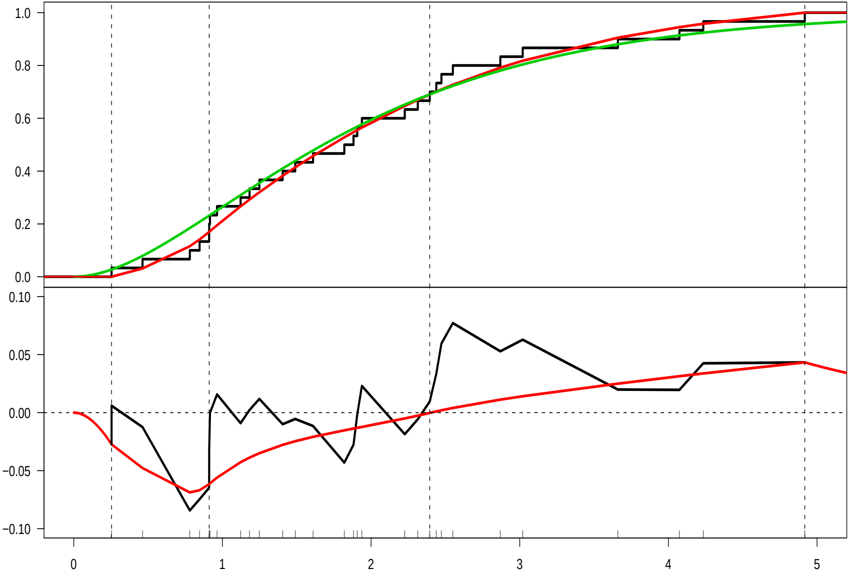
$$\mathbb{F}_n - \frac{1}{n} \leq \widehat{F}_n \leq \mathbb{F}_n$$

on the set $\mathrm{knots}(\widehat{\varphi}_\mathrm{n})$.

$\Gamma(2,1)$-sample of size $n = 30$



$\Gamma(2,1)$-sample of size $n = 30$



Assumptions

Let φ be Hölder-continuous on $T := [A, B] \subset \{f > 0\}$ with exponent $\beta \in [1, 2]$, i.e. for some constant $L > 0$

$$\begin{array}{rcll} |\varphi(x) - \varphi(y)| & \leq & L|x - y| & \text{if } \beta = 1 \\ |\varphi'(x) - \varphi'(y)| & \leq & L|x - y|^{\beta - 1} & \text{if } \beta > 1 \end{array}$$

for all $x, y \in T$.

We constrain attention to a fixed interval $T = [A, B]$.

Theorem (Consistency of $\widehat{\varphi}_n$)

Then, as $n \to \infty$,

$$\sup_{t \in [A+\delta_n, B-\delta_n]} |\widehat{\varphi}_n - \varphi|(t) \quad = \quad O_{\mathrm{p}}\left(\left(\frac{\log(n)}{n}\right)^{\beta/(2\beta+1)}\right),$$

where $\delta_n := (\log(n)/n)^{1/(2\beta+1)} \to 0$.

Result also holds for $\widehat{f}_n - f$.

Rate adaptivity:

$\beta = 1 \;\Rightarrow\;$ rate of $(\log(n)/n)^{1/3}$ (Grenander)

$\beta = 2 \;\Rightarrow\;$ rate of $(\log(n)/n)^{2/5}$. (Convex decreasing, $\in C^2[0, \infty)$)

Corollary (Uniform consistency on $\mathbb{R}$)

We immediately get

$$\int \left|\widehat{f}_n(x) - f(x)\right| \mathrm{d}x \ \to_{\mathrm{p}} \ 0 \quad \textit{and} \quad \|\widehat{F}_n - F\|_\infty^{\mathbb{R}} \ \to_{\mathrm{p}} \ 0.$$

Use modulus of continuity of uniform empirical process (Stute, 1982):

Theorem ($\widehat{F}_n$ and $\mathbb{F}_n$ asymptotically equivalent)

Under the above assumptions and $\varphi'(x) - \varphi'(y) \geq C(y - x)$*:*

$$\sup_{t \in [A+\delta_n, B-\delta_n]} |\widehat{F}_n - \mathbb{F}_n|(t) = O_{\mathrm{p}}\left(\left(\frac{\log(n)}{n}\right)^{3\beta/(4\beta+2)}\right).$$

$\beta = 1 : \Rightarrow \ O_{\mathrm{p}}((\log(n)/n)^{1/2})$

$\beta > 1 : \Rightarrow o_{\mathrm{p}}(n^{-1/2})$.

Theorem (Consistency of $\widehat{F}_n$)

If $\beta > 1$ *then*

$$\sup_{t \in [A+\delta_n, B-\delta_n]} |\widehat{F}_n - F|(t) = O_{\mathrm{p}}(n^{-1/2}).$$

No adaptivity? Rate determined by uniform distance from $\mathbb{F}_n$ to F.

Non-negative kernel, bandwidth $O_{\mathrm{p}}(n^{-1/5}) \;\Rightarrow$ rate of $O_{\mathrm{p}}(n^{-2/5})$.

Consistency of $\widehat{\varphi}_n$: "Method of caricatures"

$\Rightarrow$ Approximate $\varphi - \widehat{\varphi}_n$ near $\arg\max|\varphi - \widehat{\varphi}_n|$ by Δ's in the class

$$\mathcal{D} = \{\Delta \ : \ \text{piecewise linear with} \# knots \leq 3\}.$$

$\Rightarrow$ Use exponential inequalities and bracketing to show that

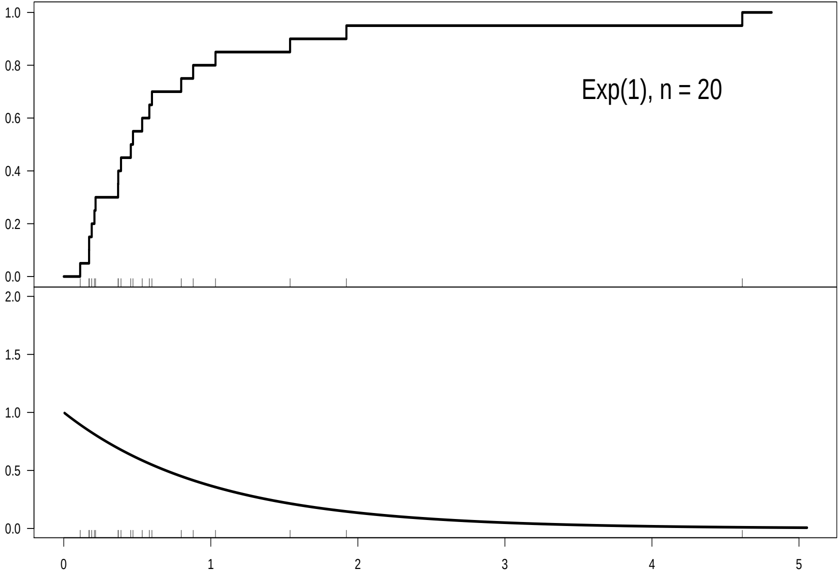
$$\sup_{\Delta \in \mathcal{D}} \frac{|\int \Delta \, \mathrm{d}(\mathbb{F}_n - F)|}{\sigma(\Delta) + \mathrm{peanuts}} \quad = \quad O_{\mathrm{p}}((\log(n)/n)^{1/2})$$

where $\sigma(\Delta)^2 = \int \Delta^2 \, \mathrm{d}F$.

$\Rightarrow$ Crucial (to get rid of $\widehat{F}_n$): Characterization of $\widehat{F}_n$ via integrals.

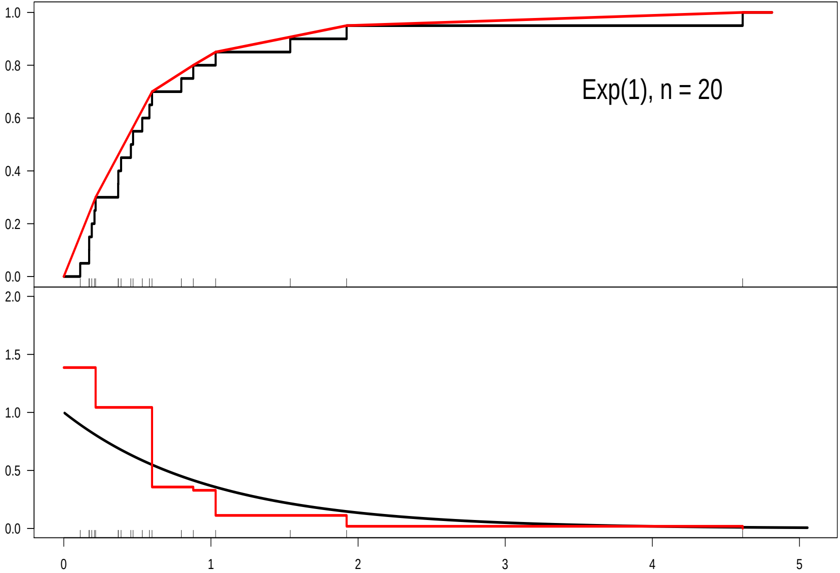
Marshall's lemma

Estimation of monotone decreasing density on $[0, \infty)$: $\widehat{F}_n = \mathrm{CM}(\mathbb{F}_n)$



Marshall's lemma

Estimation of monotone decreasing density on $[0, \infty)$: $\widehat{F}_n = \mathrm{CM}(\mathbb{F}_n)$



In this (monotone) case one has:

$$\|\widehat{F}_n - F\|_\infty^{\mathbb{R}} \quad \leq \|\mathbb{F}_n - F\|_\infty^{\mathbb{R}}$$

for all $n > 1$.

$\Rightarrow$ non-asymptotic result!

Log-concave case: For any $C \geq 1$,

$$\|\widehat{F}_n - F\|_\infty^{\mathbb{R}} \quad \not\leq \quad C\|\mathbb{F}_n - F\|_\infty^{\mathbb{R}},$$

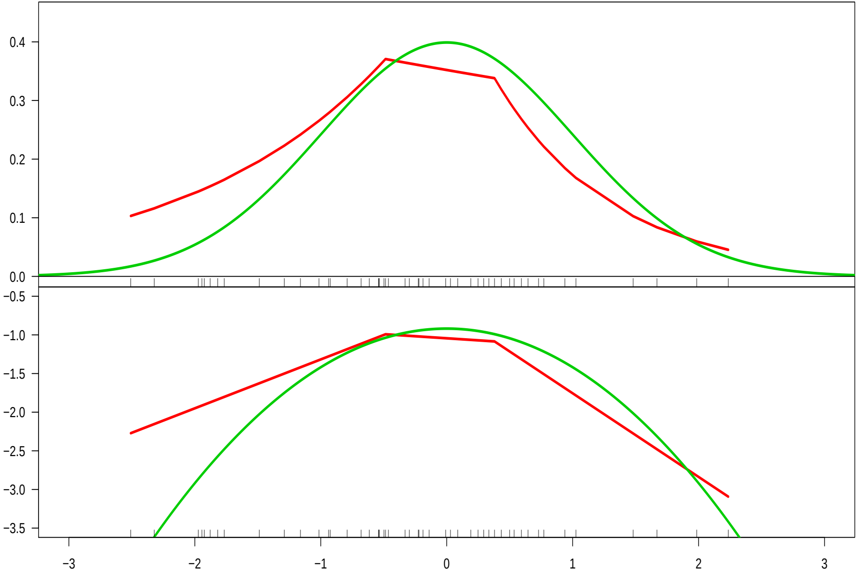
in general. $\exists$ pathological counterexamples!

Conjecture:

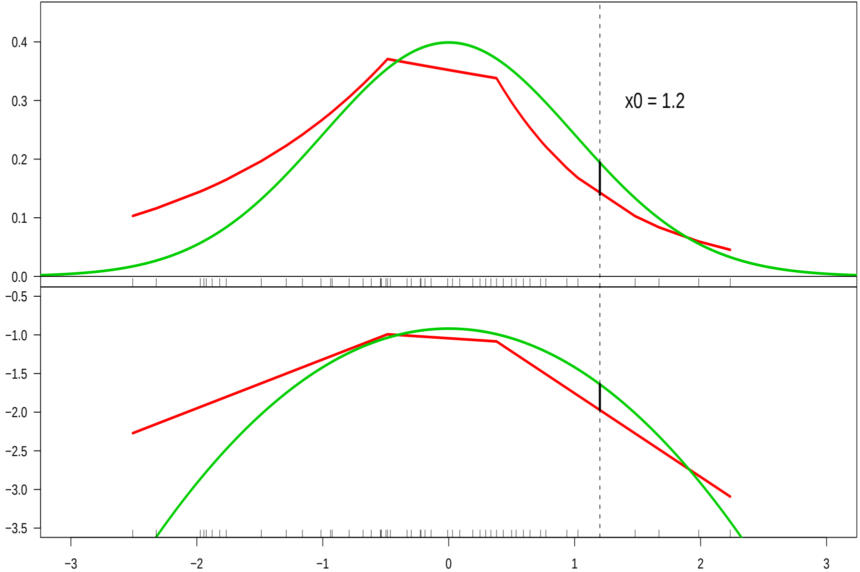
$$P\Big(\|\widehat{F}_n - F\|_\infty^{\mathbb{R}} \leq \|\mathbb{F}_n - F\|_\infty^{\mathbb{R}}\Big) \quad \rightarrow \quad 1.$$

Proof attempts yielded Marshall's Lemma for convex decreasing density estimation...

Normal density and log-density, $n = 50$.



Normal density and log-density, $n = 50$.



Assumptions

(A1) Fix $x_0 \in \mathbb{R}$, the "true" density f_0 is log-concave.

(A2) $f_0(x_0) > 0$.

(A3) Log-density φ_0 is at least twice continuously differentiable in a neighborhood of x_0.

(A4) If $\varphi_0''(x_0) \neq 0$, then $k = 2$. Otherwise, suppose that k is the smallest integer such that $\varphi_0^{(j)}(x_0) = 0, j = 2, \ldots, k-1$, $\varphi_0^{(k)}(x_0) \neq 0$, and $\varphi_0^{(k)}$ is continuous in a neighborhood of x_0.

φ_0 concave, A3 and A4 $\Rightarrow$ k is even, $\varphi_0^{(k)}(x_0) < 0$.

Number of vanishing derivatives

(A4) If $\varphi_0''(x_0) \neq 0$, then $k = 2$. Otherwise, suppose that k is the smallest integer such that $\varphi_0^{(j)}(x_0) = 0, j = 2, \ldots, k-1$, $\varphi_0^{(k)}(x_0) \neq 0$, and $\varphi_0^{(k)}$ is continuous in a neighborhood of x_0.

Consider

$$f_0(x) \;=\; \sqrt{2}\,\frac{\Gamma(3/4)}{\pi}\exp(-x^4),\;\; x \in \mathbb{R}.$$

$\Rightarrow \varphi_0^{(j)}(x_0) = 0,\; j = 1, 2, 3$ for $x_0 = 0$, and $\varphi_0^{(4)}(x_0) \neq 0$.

Can construct examples with $\varphi_0''(x_1) = 0$, but x_1 not the mode.

Theorem

Suppose that A1 - A4 hold. Then

$$\begin{pmatrix} n^{k/(2k+1)}\left(\widehat{f}_n(x_0) - f_0(x_0)\right) \\ n^{(k-1)/(2k+1)}\left(\widehat{f}'_n(x_0) - f'_0(x_0)\right) \end{pmatrix} \to_d \begin{pmatrix} c_k(x_0,\varphi_0)\ H_k^{(2)}(0) \\ d_k(x_0,\varphi_0)\ H_k^{(3)}(0) \end{pmatrix}$$

and

$$\begin{pmatrix} n^{k/(2k+1)}\left(\widehat{\varphi}_n(x_0) - \varphi_0(x_0)\right) \\ n^{(k-1)/(2k+1)}\left(\widehat{\varphi}'_n(x_0) - \varphi'_0(x_0)\right) \end{pmatrix} \to_d \begin{pmatrix} C_k(x_0,\varphi_0)\ H_k^{(2)}(0) \\ D_k(x_0,\varphi_0)\ H_k^{(3)}(0) \end{pmatrix}$$

where H_k is the "lower invelope" of the process

$$Y_k(t) = \begin{cases} \int_0^t W(s)ds - t^{k+2} & \text{if } t \geq 0 \\ \int_t^0 W(s)ds - t^{k+2} & \text{if } t < 0. \end{cases}$$

for W two-sided Brownian motion, starting at 0.

Constants (d_k, C_k, D_k similar):

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

What is the "lower invelope" of Y_k?

A a.s. uniquely defined stochastic process H_k such that

- $H_k(t) \leq Y_k(t)$ for all $t \in \mathbb{R}$.
- $H_k^{(2)}$ is concave.
- $H_k(t) = Y_k(t)$ if the slope of $H_k^{(2)}$ decreases strictly at t.

"Be lucky...": H_k is basically the same process (apart from k) as in convex density estimation problem $\Rightarrow$ existence, uniqueness, ...

"...or not so lucky:" k is unfortunately not known.

Dependence on smoothness via k: Estimation is harder for higher k.

Sketch of proof

Show that $\tau_+ - \tau_- = O_{\mathrm{p}}(n^{-1/(2k+1)})$ for τ_-, τ_+ two consecutive knots of $\widehat{\varphi}_n$.

Consider the processes ($r_n = n^{(k+2)/(2k+1)},\ x_n(t) = x_0 + n^{-1/(2k+1)}t$)

$$\mathbb{Y}_n^{loc}(t) := r_n \int_{x_0}^{x_n(t)} \Big(\mathbb{F}_n(v) - \mathbb{F}_n(x_0) - \int_{x_0}^{v} \Big(\sum_{j=0}^{k-1} \frac{f_0^{(j)}(x_0)(u - x_0)^j}{j!} \Big) du \Big) dv$$

$$\widehat{H}_n^{loc}(t) := r_n \int_{x_0}^{x_n(t)} \int_{x_0}^{v} \Big(\widehat{f}_n(u) - \sum_{j=0}^{k-1} \frac{f_0^{(j)}(x_0)}{j!}(u - x_0)^j \Big) du dv + \widehat{A}_n t + \widehat{B}_n.$$

$\Rightarrow$ Modify to be on "log-density" level, show weak convergence and $\widehat{H}_n^{loc}(t) \leq \mathbb{Y}_n^{loc}(t)$, then take suitable derivatives.

Remember: Log-concave densities are always unimodal.

Mode functional: for a (bounded) function g

$$M(g) \;:=\; \min\{t \;:\; g(t) \;=\; \max_{u \in \mathbb{R}} g(u)\}.$$

From

$$n^{k/(2k+1)}(\widehat{f}_n(x_0) - f_0(x_0)) \;\;\to_d\;\; c_k(x_0, \varphi_0)\; H_k^{(2)}(0)$$

one can deduce (non-trivially!)

Theorem

Suppose $f_0^{(k)}$ is continuous in a neighborhood of $m_0 = M(f_0)$. Then, for $\widehat{M}_n = M(\widehat{f}_n)$,

$$n^{1/(2k+1)}(\widehat{M}_n - m_0) \;\;\to_d\;\; \left(\frac{(k+2)!^2}{f_0(m_0)|\varphi_0^{(k)}(m_0)|^2}\right)^{1/(2k+1)} M(H_k^{(2)}).$$

$\Rightarrow$ Local asymptotic minimax lower bounds (L_1-risk):

$$\sup_{\tau>0} \limsup_{n\to\infty} n^{1/5} \inf_{T_n} \sup_{f\in\mathcal{LC}_{n,\tau}} E_f|T_n - M(f)| \quad \geq \quad .15512\left(\frac{f_0(m_0)}{f_0''(m_0)^2}\right)^{1/5}$$

$\Rightarrow$ Local asymptotic minimax lower bounds (L_1-risk):

$$\sup_{\tau>0} \limsup_{n\to\infty} n^{1/5} \inf_{T_n} \sup_{f\in \mathcal{LC}_{n,\tau}} E_f|T_n - M(f)| \geq .15512 \left(\frac{f_0(m_0)}{f_0''(m_0)^2} \right)^{1/5}$$

$$\sup_{\tau>0} \limsup_{n\to\infty} n^{1/5} \inf_{T_n} \sup_{f\in \mathcal{U}_{n,\tau}} E_f|T_n - M(f)| \geq .19784 \left(\frac{f_0(m_0)}{f_0''(m_0)^2} \right)^{1/5}.$$

$\Rightarrow$ Local asymptotic minimax lower bounds (L_1-risk):

$$\sup_{\tau>0} \limsup_{n\to\infty} n^{1/5} \inf_{T_n} \sup_{f\in\mathcal{LC}_{n,\tau}} E_f|T_n - M(f)| \;\geq\; .15512\left(\frac{f_0(m_0)}{f_0''(m_0)^2}\right)^{1/5}$$

$$\sup_{\tau>0} \limsup_{n\to\infty} n^{1/5} \inf_{T_n} \sup_{f\in\ \mathcal{U}_{n,\tau}} E_f|T_n - M(f)| \;\geq\; .19784\left(\frac{f_0(m_0)}{f_0''(m_0)^2}\right)^{1/5}.$$

$\Rightarrow$ Estimate f from censored data: Dümbgen, Hüsler, & R. (2007).

$\Rightarrow$ Estimation of mixtures: Chang & Walther (2007).

$\Rightarrow$ Log-concave densities in $\mathbb{R}^d$: Cule et al (2007).

$\Rightarrow$ $\widehat{f}_n$ as "smoother" in tail index estimation: Müller & R. (2007a, b).

$\Rightarrow$ Software: `logcondens`, on `CRAN`: R. and Dümbgen (2006).