Shape constrained estimation: a brief introduction

Jon A. Wellner

University of Washington, Seattle

Cowles Foundation, Yale

November 7, 2016

Econometrics lunch seminar, Yale

Based on joint work with:

- Fadoua Balabdaoui
- Piet Groeneboom
- Geurt Jongbloed
- Kaspar Rufibach
- Arseni Seregin
- Charles Doss

Outline

- 1. Some shape constrained models:
  - ▷ A: Monotone (and unimodal).
  - ▷ B: Convex (or concave):
  - ▷ C: $k-$monotone, completely monotone
  - ▷ D: log-concave; log-convex; s-concave; bi-logconcave
  - ▷ E: hyperbolically monotone,
    hyperbolically completely monotone
  - ▷ F: higher dimensional monotone, convex, log-concave.
- 2. Estimation methods:
  - ▷ A. maximum likelihood
  - ▷ B. least squares
  - ▷ C. minimum Rényi entropy
  - ▷ D. penalized likelihoods

- 3. Types of results?
  - A. Pointwise consistency; global consistency ?
  - B. Global consistency; rates of global consistency?
  - C. Global rates of convergence: Hellinger/ L1 ?
  - D. Local distribution theory / limiting distributions?
  - E. Risk bounds: lower; upper; adaptation to boundary?
- 4. Testing and confidence intervals
- 5. Algorithms
  - A. active set
  - B. interior point

1. Some shape constrained models:

A. Monotone (density):

$$\mathcal{F}_{mon} \equiv \{f \text{ a density on } \mathbb{R}^+,\ f(x) \le f(y) \text{ if } x \ge y\}.$$

f on $\mathbb{R}^+ = [0, \infty)$ is decreasing (i.e. non-increasing) if and only if it is a scale mixture of uniform densities: for $x > 0$,

$$f(x) = \int_0^\infty \frac{1}{y} 1_{[0,y)}(x) dG(y) = \int_{\{y>x\}} \frac{1}{y} dG(y)\,, \qquad (1)$$

for some distribution function G on $(0, \infty)$. [Schoenberg, 1941; Williamson, 1956; Feller, 1971].

Thus if $X \sim f \in \mathcal{F}_{mon}$, there is a random variable $Z \sim G$ on $\mathbb{R}^+$ such that

$$X \stackrel{d}{=} UZ$$

where $U \sim$ Uniform$[0, 1]$.

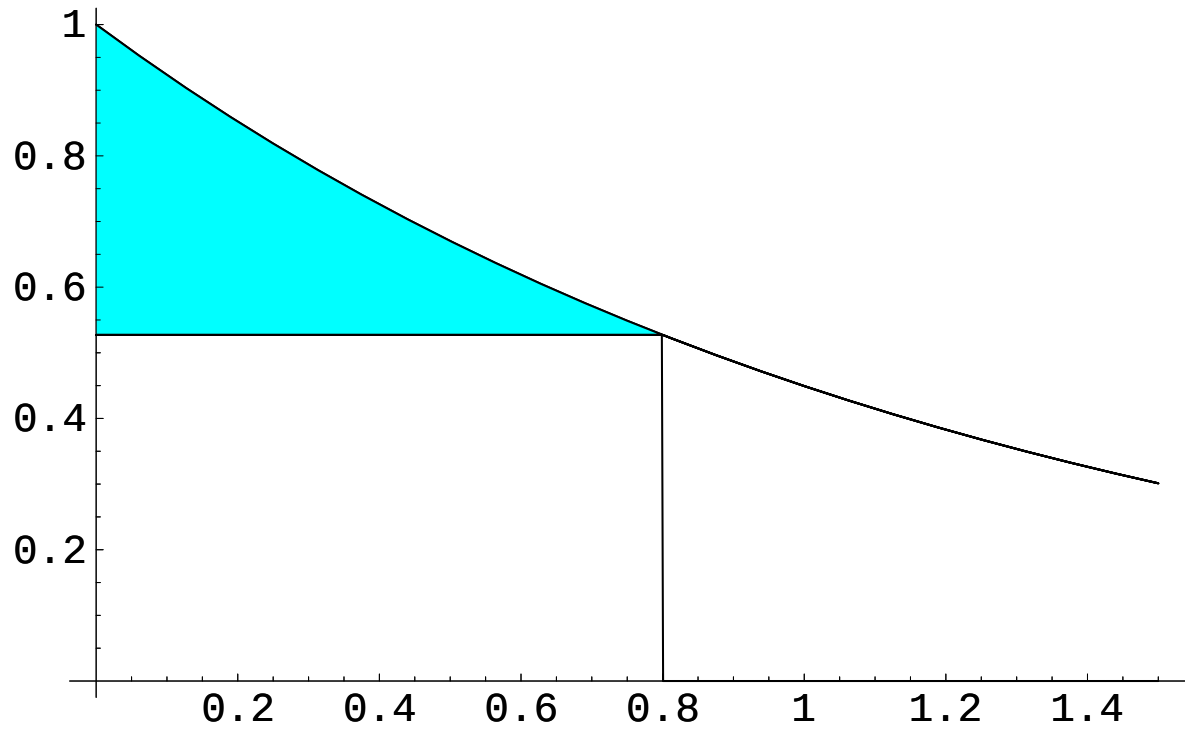
The relationship (1) implies that the corresponding distribution function F is given by

$$\begin{aligned} F(x) &= \int_0^\infty \frac{x}{y} 1_{[0,y)}(x) dG(y) + \int_0^\infty 1_{[y,\infty)}(x) dG(y) \\ &= xf(x) + G(x)\,, \end{aligned}$$

and this can be "inverted" to yield

$$G(x) = F(x) - xf(x)\,. \tag{2}$$

From Figure 1 we see that the function on the right side of (2) is non-negative and non-decreasing: the shaded area gives exactly the difference $F(x) - xf(x)$.



Inverting the monotone density mixture representation

Alternatively, if G is a distribution function with finite mean $\mu_G \equiv \int_0^\infty (1 - G(y))dy$, then

$$f(x; G) \equiv \frac{1}{\mu_G}(1 - G(x)) \tag{3}$$

is a monotone decreasing density on $\mathbb{R}^+$. This mixture representation yields the sub-class of monotone densities with $f(0+) < \infty$. This is the way that monotone densities often arise in applications, since the right side in (3) is the limiting distribution of both the forward and backward recurrence time distributions (or the distributions of "excess life" γ_t and "current life" δ_t) in renewal theory; see e.g. Karlin-Taylor (1975).

If we change the uniform density to a triangular density, the mixture representation yields convex decreasing densities $f \in \mathcal{F}_{conv}$:

$$f(x) = \int_0^\infty 2y^{-1}\left(1 - \frac{x}{y}\right)_+ dG(y).$$

C. $k-$monotone; completely monotone Since the triangular density $2(1-x)1_{[0,1]}(x)$ is just the Beta$(1,2)$ density, we easily find:

Proposition. (Williams, 1956). A density f is a $k-$monotone (completely monotone) density if and only if it can be represented as a scale mixture of Beta$(1,k)$ (exponential) densities; with $x_+ \equiv x1\{x \geq 0\}$,

$$f(x) = \begin{cases} \int_0^\infty y^{-1}\left(1 - \frac{x}{ky}\right)_+^{k-1} dG(y), & k \in \{1,2,\ldots\}, \\ \int_0^\infty y^{-1}\exp(-x/y)dG(y), & k = \infty\,, \end{cases} \tag{4}$$

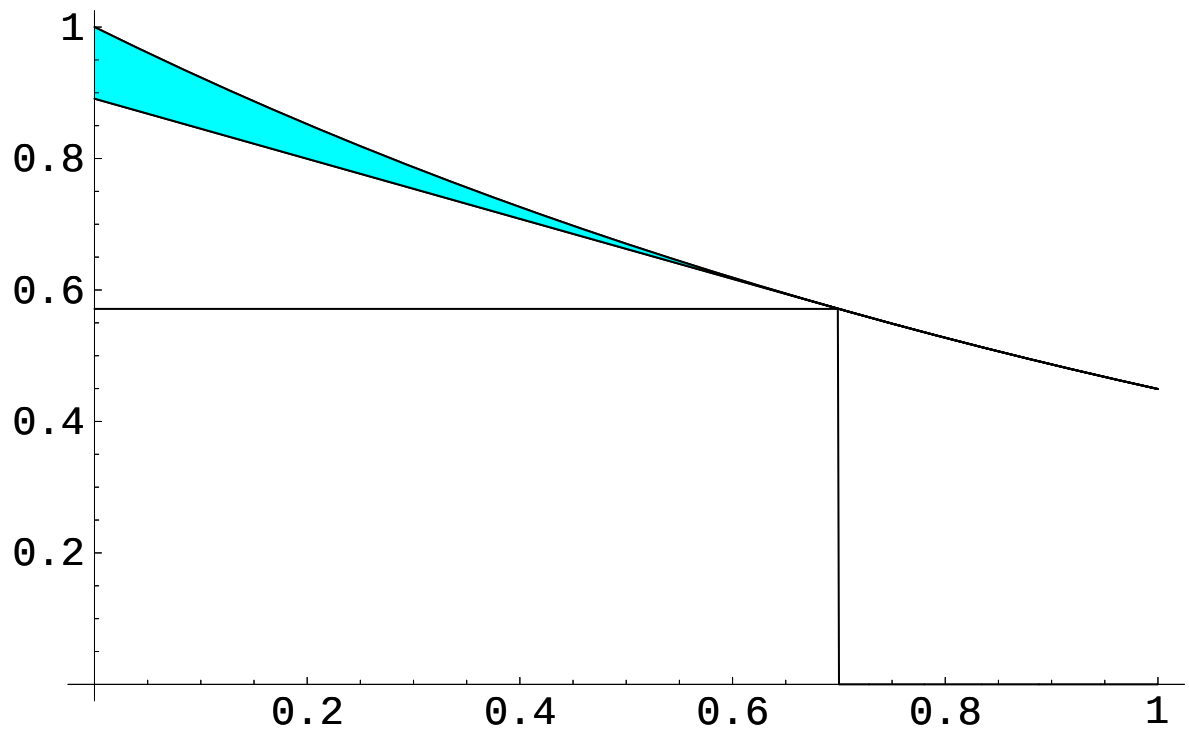
for some distribution function G on $(0,\infty)$.

The inversion formulas corresponding to these mixture representations are given in the following proposition.

Proposition 2. Suppose that f is a $k-$monotone density with distribution function F (so $F(x) = \int_0^x f(t)dt$). Then the distribution function $G = G_k$ of (4) is given at continuity points of G_k by

$$G_k(t) = \sum_{j=0}^{k} \frac{(-1)^j}{j!}(kt)^j F^{(j)}(kt)\,, \tag{5}$$

$$G_\infty(t) = \lim_{k\to\infty} G_k(t)\,. \tag{6}$$



Graphical view of the inversion formula, convex decreasing density

Note that for $3 \leq k < \infty$

$$\mathcal{F}_\infty \subset \mathcal{F}_k \subset \mathcal{F}_2 \subset \mathcal{F}_{mon}.$$

D. Log-concave; log-convex; $s-$concave; ...

- p is log-concave on $\mathbb{R}^d$ if $\log p \equiv \varphi$ is concave (on $\mathbb{R}^d$).
- p is log-convex on $\mathbb{R}^d$ if $\log p \equiv \varphi$ is convex (on $\mathbb{R}^d$).
- for $s > 0$, p is $s-$concave if $p = \varphi_+^{1/s}$ for some φ concave;
  for $s < 0$, p is $s-$concave if $p = \varphi_+^{1/s}$ for some φ convex.

Then, for $-\infty < s < 0 < r < \infty$

$$\mathcal{P}_{-\infty} \supset \mathcal{P}_s \supset \mathcal{P}_0 \supset \mathcal{P}_r \supset \mathcal{P}_\infty.$$

E. Hyperbolically $k-$monotone and completely monotone

A non-negative function $f(x)$ on $(0, \infty)$ is said to be hyperbolically monotone of order k (or HM_k) if for each fixed u the function

$$H(w) \equiv f(uv)f(u/v), \quad w \equiv 2^{-1}(v + 1/v)$$

satisfies $(-1)^j H^{(j)}(w) \geq 0$, $j = 0, 1, \ldots, k-1$ and $(-1)^{(k-1)} H^{(k-1)}(w)$ is right-continuous and decreasing. If f is HM_k for every k, then f is hyperbolically completely monotone, or HM_∞.

- Any uniform density is HM_1.
- The half-normal densities $f(x) = 2\phi(x/\sigma)/\sigma$ for $x \geq 0$ are HM_1, but not HM_2.
- The log-normal distribution is HM_∞.

Proposition 1. (Bondesson, 1991) $X \sim HM_1$ if and only if $\log X$ has a log-concave density on $\mathbb{R}$.

Proposition 2. (Bondesson, 1991) If $X \sim HM_k$ and $Y \sim HM_k$ are independent then $XY \sim HM_k$ and $X/Y \sim HM_k$.

Thus:

$$\text{log-concave} = \log HM_1 \subset \log HM_k \subset \log HM_\infty.$$

2. Estimation methods:

- Maximum likelihood:
  - Suppose that $X_1, \ldots, X_n$ are i.i.d. $p_0 \in \mathcal{P}$.
  - Then with $\mathbb{P}_n \equiv n^{-1} \sum_1^n \delta_{X_i}$, the log-likelihood is defined by

$$L_n(p) \equiv \sum_{i=1}^n \log p(X_i) \equiv n\mathbb{P}_n(\log p).$$

  - If $\widehat{p}_n$ satisfies

$$L_n(\widehat{p}_n) = \max_{p \in \mathcal{P}} L_n(p),$$

  then $\widehat{p}_n$ is the Maximum Likelihood Estimator of $p_0 \in \mathcal{P}$.
  - The "adjusted log-likelihood" Ψ_n is defined by

$$\Psi_n(p) \equiv L_n(p) - \int_{\mathbb{R}^d} p(x)dx.$$

- Least squares:
  - $X_1, \ldots, X_n$ i.i.d. $p_0 \in \mathcal{P}$.
  - The least squares contrast function is defined by

$$\begin{aligned}\Psi_n(p) &= \frac{1}{2}\int p^2(x)dx - \int p(x)d\mathbb{P}_n(x) \\ &\approx \frac{1}{2}\int (p(x) - \overline{p}_n(x))^2 dx - \frac{1}{2}\int \overline{p}_n^2(x)dx\end{aligned}$$

  if $\mathbb{P}_n$ had density $\overline{p}_n$; but the neglected term depends only on the data, not on p.
  - If $\widehat{p}_n$ satisfies

$$\Psi_n(\widehat{p}_n) = \max_{p \in \mathcal{P}} \Psi_n(p),$$

  then $\widehat{p}_n$ is the least squares estimator of $p_0 \in \mathcal{P}$.

- Other methods:
  - Minimum Rényi entropy (or divergence).
  - Penalized versions of maximum likelihood, LS, ...
  - Least squares for regression models ...
  - L_1 for regression models?

Example 1. Monotone densities.

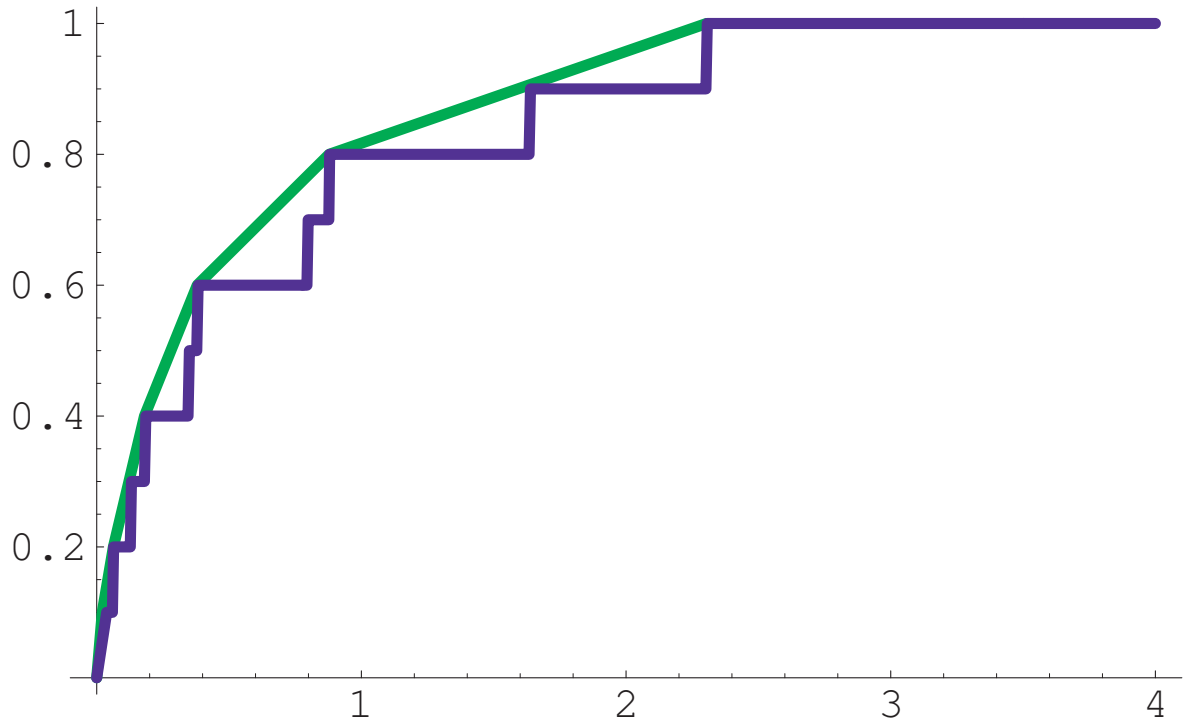
Grenander (1956) showed that the maximum likelihood estimator $\widehat{f}_n$ of f is the (left-) derivative of the least concave majorant of the empirical distribution function $\mathbb{F}_n$

$$\widehat{f}_n \;=\; \text{left derivative of the least concave majorant of } \mathbb{F}_n,$$
$$\text{the empirical distribution of } \quad X_1, \ldots, X_n \;\; \text{i.i.d. } F$$

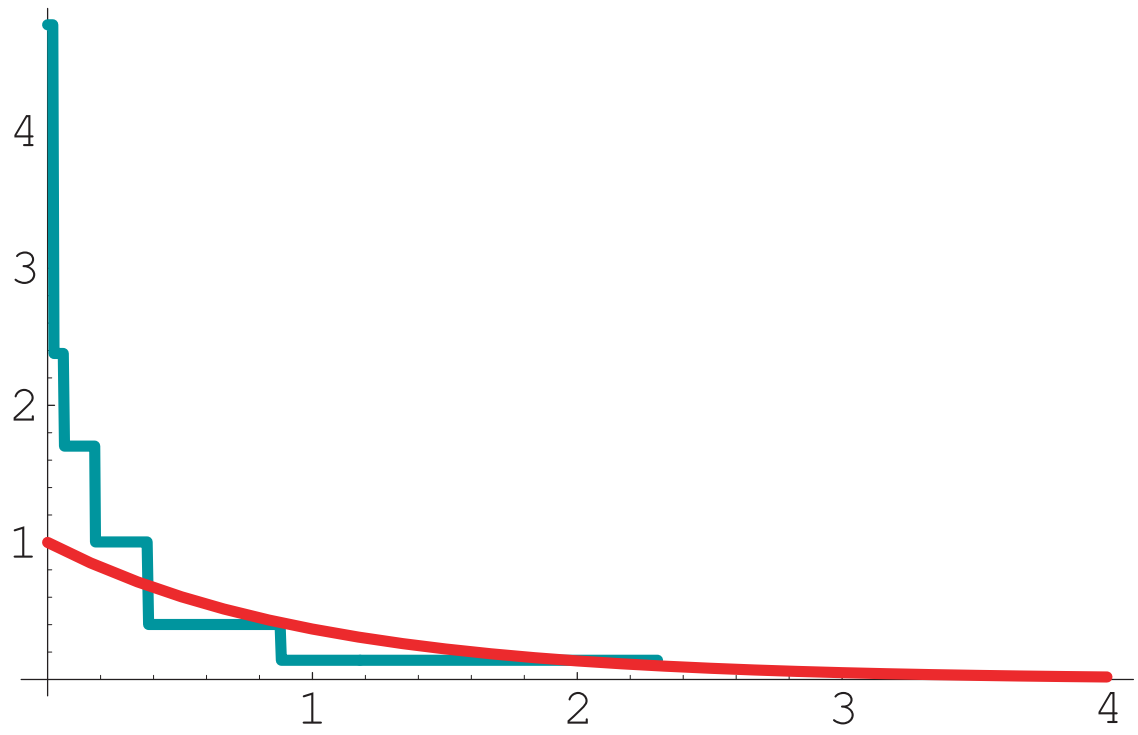
For the class of monotone densities on $[0, \infty)$,

the MLE and LSE are identical.

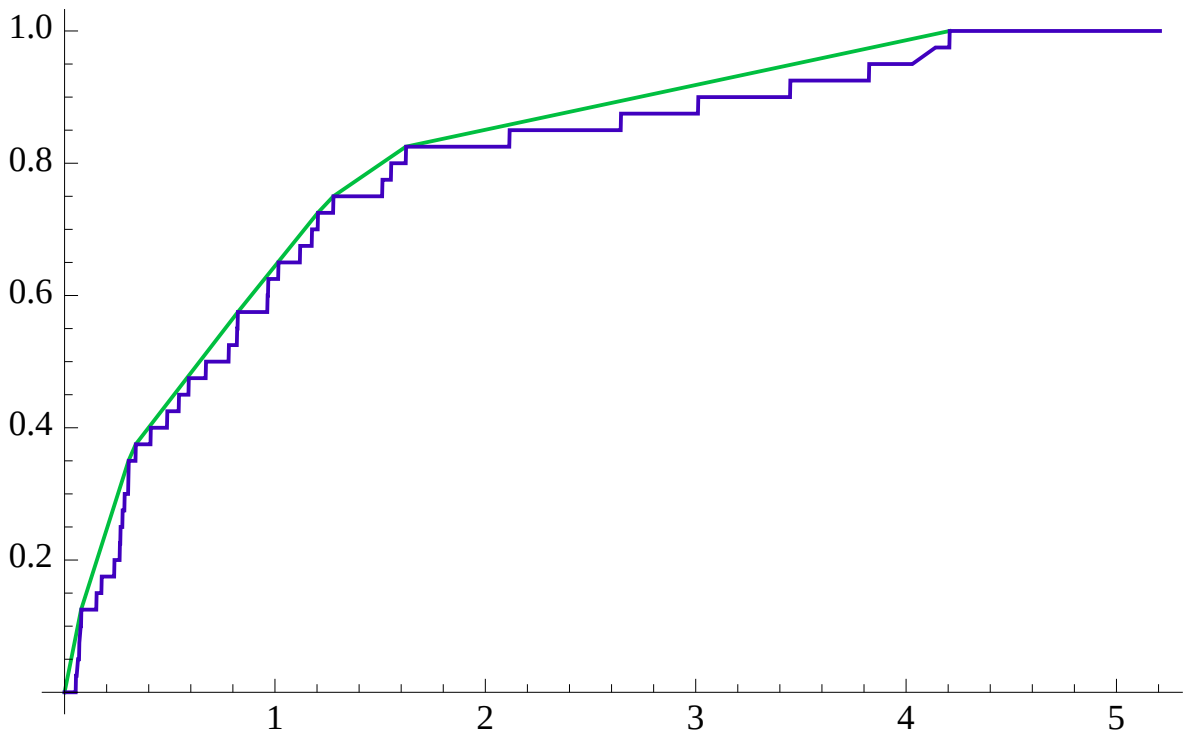
This is *not* true in general.



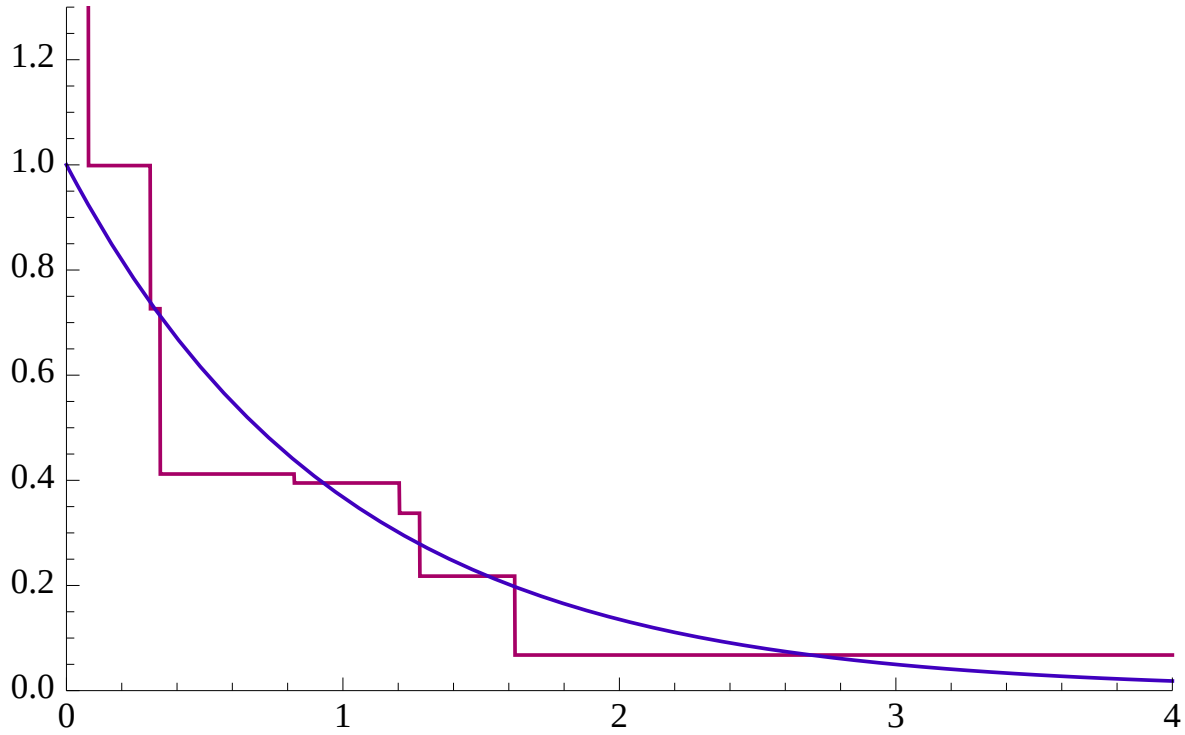
Least Concave Majorant and Empirical $n = 10$



Grenander Estimator and Exp(1) density, $n = 10$



Least Concave Majorant and Empirical $n = 40$



Grenander Estimator and Exp(1) density, $n = 40$

Properties of Grenander's estimator:

What do we know about $\widehat{f}_n$?

- Consistency:
  - $\widehat{f}_n(x) \to_{a.s.} f_0(x)$ for each $x > 0$.
  - If f_0 is continuous on $(0, \infty)$, $\sup_{x \geq c} |\widehat{f}_n(x) - f_0(x)| \to_{a.s.} 0$.
  - If $f_0(0) < \infty$, then $\widehat{f}_n(0) \to_d f_0(0) U^{-1}$ where $U \sim \text{Uniform}(0, 1)$.
  - (Groeneboom, Birgé, van de Geer) Let

$$H^2(f, g) \equiv 2^{-1} \int \{\sqrt{f(x)} - \sqrt{g(x)}\}^2 dx$$

be the Hellinger distance between f and g.
If $\int_{f_0 < 1} f_0^{1-\alpha}(x) dx < \infty$ and $\int_{[f_0 > 1]} f_0^{1+\alpha}(x) dx < \infty$, then

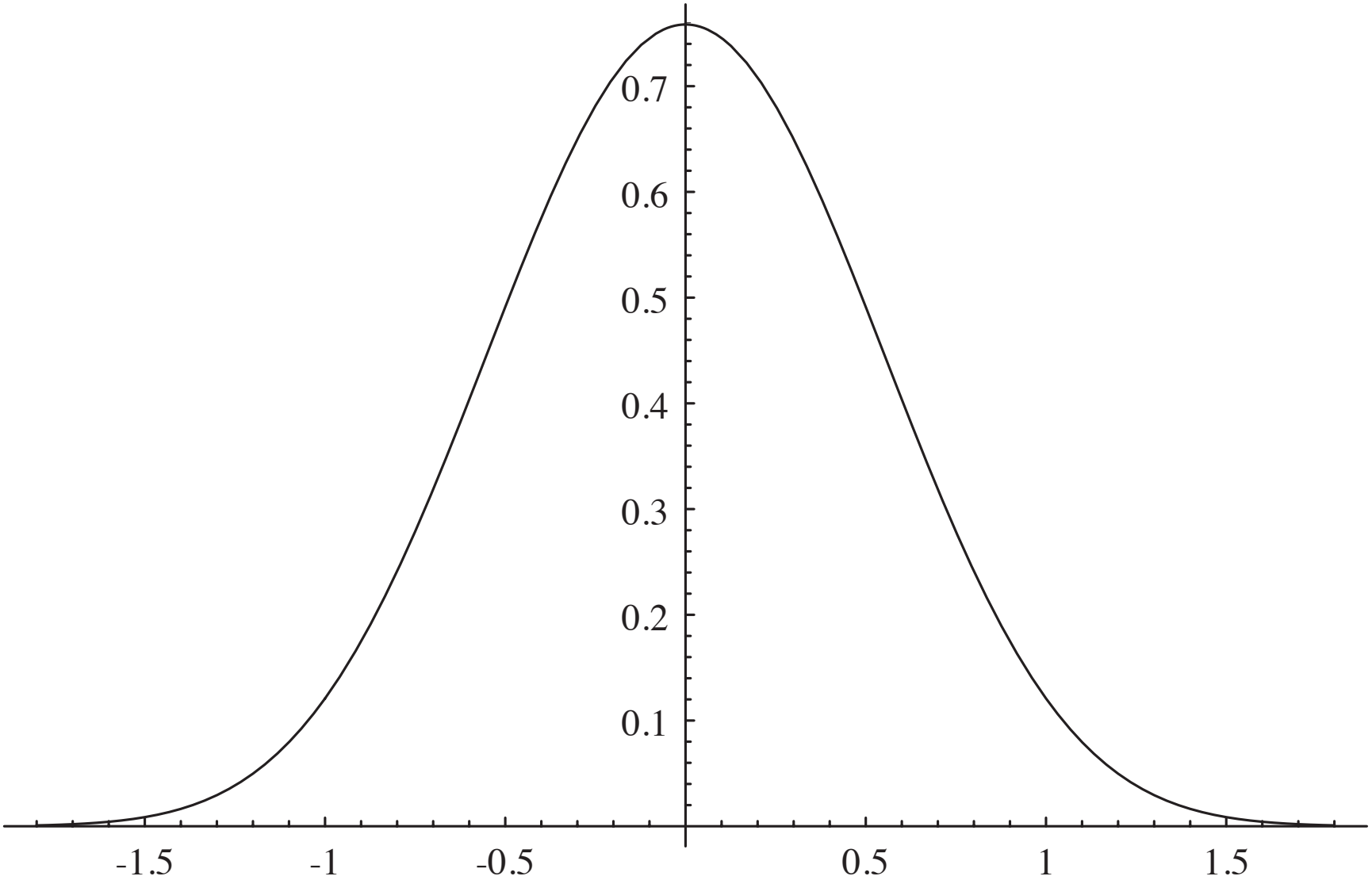
$$H(\widehat{f}_n, f_0) = O_p(n^{-1/3}).$$

- Limit distribution theory:
  - $f_0'(x) < 0$, $f_0(x) > 0$, and f_0' is continuous in a neighborhood of f_0, then with $C \equiv (2^{-1} f_0(x)|f_0'(x)|)^{1/3}$,
  - $$n^{1/3}(\widehat{f}_n(x) - f_0(x)) \to_d C\mathbb{S}(0) \stackrel{d}{=} C2Z$$

    where

    $$\begin{aligned} \mathbb{S}(0) &\equiv \text{slope at zero of the least concave majorant} \\ &\qquad \text{of } W(h) - h^2, \\ Z &\equiv \text{argmax}_h\{W(h) - h^2\}; \end{aligned}$$

    here W is two-sided Brownian motion started at 0.
  - The distribution of Z is called Chernoff's distribution.
  - From Groeneboom (1989), lots is known about f_Z.



Chernoff's density $f_Z(x) = (1/2)g(x)g(-x)$.

- Marshall's lemma: $\|\widehat{F}_n - F_0\|_\infty \leq \|\mathbb{F}_n - F_0\|_\infty$.

- Kiefer-Wolfowitz theorem:
  - If $\alpha_1(F) \equiv \inf\{t : \ F_0(t) = 1\} < \infty$,
    $\beta_1(F) \equiv \sup_{t<\alpha_1}(-f_0'(t)/f_0^2(t)) > 0$,
    and $\gamma_1(F) \equiv \sup_{t<\alpha_1}(-f'(t))/\inf_{t<\alpha_1} f^2(t) < \infty$,
  - ... then

$$\|\widehat{F}_n - \mathbb{F}_n\|_\infty = O((n^{-1}\log n)^{2/3}) \quad \text{a.s.}$$

and hence

$$\sqrt{n}\|\widehat{F}_n - \mathbb{F}_n\|_\infty = O(n^{-1/6}(\log n)^{2/3}) \quad \text{a.s..}$$

- Model miss-specification: If f is not monotone (or even if F does not have a density), then

$$\int |\widehat{f}_n(x) - f_0^*(x)| dx \to_{a.s.} 0$$

where

$$K(f, f_0^*) = \inf\{K(f, g) : \ g \in \mathcal{F}_{mon}\}.$$

(Pfanzagl (1990), Patilea (2001), Jankowski (2014), Dümbgen, Samworth, and Schumacher (2011)).

A. Log-concave densities on $\mathbb{R}^1$

Suppose that

$$f(x) \equiv f_\varphi(x) = \exp(\varphi(x)) = \exp\left(-(-\varphi(x))\right)$$

where φ is concave (and $-\varphi$ is convex). The class of all densities f on $\mathbb{R}$ of this form is called the class of *log-concave* densities, $\mathcal{P}_{log-concave} \equiv \mathcal{P}_0$.

Properties of log-concave densities:

- A density f on $\mathbb{R}$ is log-concave if and only if its convolution with any unimodal density is again unimodal (Ibragimov, 1956).
- Every log-concave density f is unimodal (but need not be symmetric).
- $\mathcal{P}_0$ is closed under convolution.

A. Log-concave densities on $\mathbb{R}^1$

- Many parametric families are log-concave, for example:
  - ▷ Normal (μ, σ^2); Uniform(a, b);
  - ▷ Gamma(r, λ) for $r \geq 1$; Beta(a, b) for $a, b \geq 1$
- t_r densities with $r > 0$ are not log-concave
- Tails of log-concave densities are necessarily sub-exponential
- $\mathcal{P}_{log-concave}$ = the class of "Polyá frequency functions of order 2", PF_2, in the terminology of Schoenberg (1951) and Karlin (1968). See Marshall and Olkin (1979), chapter 18, and Dharmadhikari and Joag-Dev (1988), page 150. for nice introductions.
- Thus:

$$\text{log-concave} = PF_2 = \text{strongly uni-modal}$$

B. Nonparametric estimation, log-concave on $\mathbb{R}$

- The (nonparametric) MLE $\widehat{f}_n$ exists (Rufibach, Dümbgen and Rufibach).
- $\widehat{f}_n$ can be computed: R-package "logcondens" (Dümbgen and Rufibach)
- In contrast, the (nonparametric) MLE for the class of unimodal densities on $\mathbb{R}^1$ does not exist. Birgé (1997) and Bickel and Fan (1996) consider alternatives to maximum likelihood for the class of unimodal densities.
- Consistency and rates of convergence for $\widehat{f}_n$:
  Dümbgen and Rufibach, (2009); Pal, Woodroofe and Meyer (2007).
- Pointwise limit theory? Yes! Balabdaoui, Rufibach, and W (2009).

B. Nonparametric estimation, log-concave on $\mathbb{R}$

MLE of f and φ: Let $\mathcal{C}$ denote the class of all concave function $\varphi : \mathbb{R} \to [-\infty, \infty)$. The estimator $\widehat{\varphi}_n$ based on $X_1, \ldots, X_n$ i.i.d. as f_0 is the maximizer of the "adjusted criterion function"

$$\begin{aligned} \ell_n(\varphi) &= \int \log f_\varphi(x) d\mathbb{F}_n(x) - \int f_\varphi(x) dx \\ &= \int \varphi(x) d\mathbb{F}_n(x) - \int e^{\varphi(x)} dx \end{aligned}$$

over $\varphi \in \mathcal{C}$.

Properties of $\widehat{f}_n$, $\widehat{\varphi}_n$: (Dümbgen & Rufibach, 2009)

- $\widehat{\varphi}_n$ is piecewise linear.
- $\widehat{\varphi}_n = -\infty$ on $\mathbb{R} \setminus [X_{(1)}, X_{(n)}]$.
- The knots (or kinks) of $\widehat{\varphi}_n$ occur at a subset of the order statistics $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$.
- Characterized by ...

B. Nonparametric estimation, log-concave on $\mathbb{R}$

... $\widehat{\varphi}_n$ is the MLE of $\log f_0 = \varphi_0$ if and only if

$$\widehat{H}_n(x) \begin{cases} \leq \mathbb{H}_n(x), & \text{for all } x > X_{(1)}, \\ = \mathbb{H}_n(x), & \text{if } x \text{ is a knot.} \end{cases}$$

where

$$\widehat{F}_n(x) = \int_{X_{(1)}}^{x} \widehat{f}_n(y)dy, \qquad \widehat{H}_n(x) = \int_{X_{(1)}}^{x} \widehat{F}_n(y)dy,$$

$$\mathbb{H}_n(x) = \int_{-\infty}^{x} \mathbb{F}_n(y)dy.$$

Furthermore, for every function Δ such that $\widehat{\varphi}_n + t\Delta$ is concave for t small enough,

$$\int_{\mathbb{R}} \Delta(x)d\mathbb{F}_n(x) \leq \int_{\mathbb{R}} \Delta(x)d\widehat{F}_n(x).$$

B. Nonparametric estimation, log-concave on $\mathbb{R}$

Consistency of $\widehat{f}_n$ and $\widehat{\varphi}_n$:

- (Pal, Woodroofe, & Meyer, 2007):
  If $f_0 \in \mathcal{P}_0$, then $H(\widehat{f}_n, f_0) \to_{a.s.} 0$.

- (Dümbgen & Rufibach, 2009):
  If $f_0 \in \mathcal{P}_0$ and $\varphi_0 \in \mathcal{H}^{\beta,L}(T)$ for some compact $T = [A, B] \subset \{x : \ f_0(x) > 0\}^\circ$, $M < \infty$, and $1 \le \beta \le 2$. Then

$$\sup_{t \in T}(\widehat{\varphi}_n(t) - \varphi_0(t)) = O_p\left(\left(\frac{\log n}{n}\right)^{\beta/(2\beta+1)}\right), \quad \text{and}$$

$$\sup_{t \in T_n}(\varphi_0(t) - \widehat{\varphi}_n(t)) = O_p\left(\left(\frac{\log n}{n}\right)^{\beta/(2\beta+1)}\right)$$

  where $T_n \equiv [A + (\log n/n)^{\beta/(2\beta+1)}, B - (\log n/n)^{\beta/(2\beta+1)}]$ and $\beta/(2\beta+1) \in [1/3, 2/5]$ for $1 \le \beta \le 2$.

- The same remains true if $\widehat{\varphi}_n$, φ_0 are replaced by $\widehat{f}_n$, f_0.

B. Nonparametric estimation, log-concave on $\mathbb{R}$

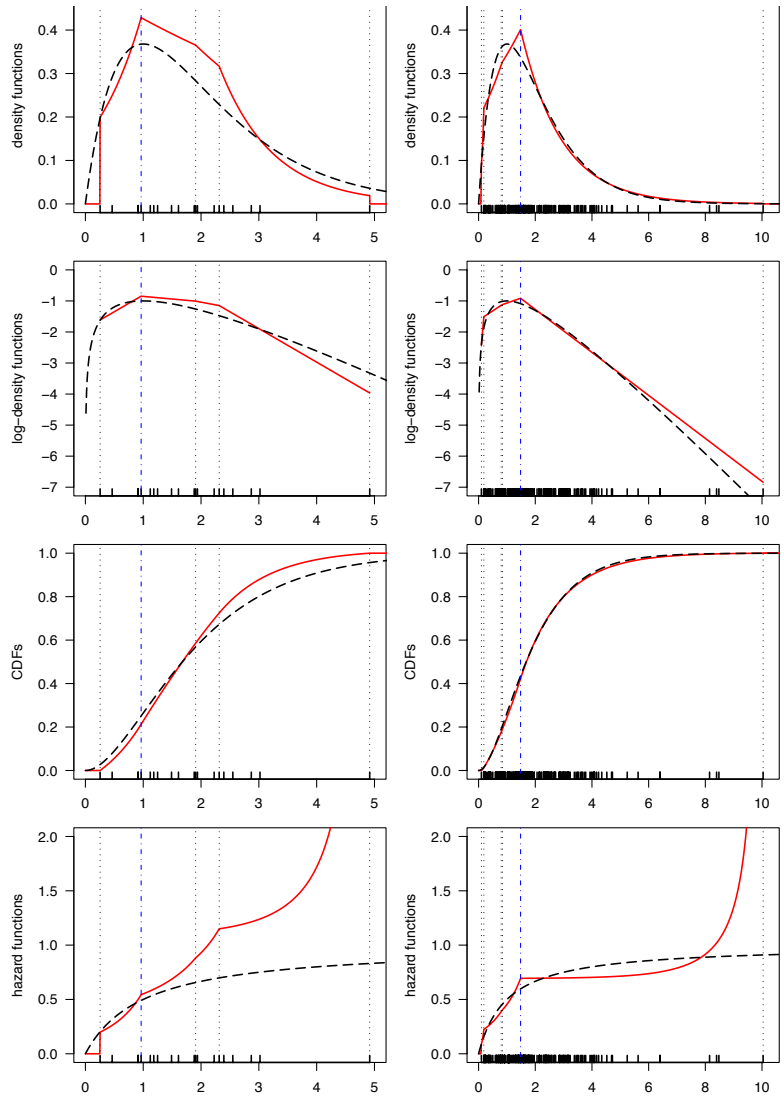
- If $\varphi_0 \in \mathcal{H}^{\beta,L}(T)$ as above and, with $\varphi_0' = \varphi_0(\cdot-)$ or $\varphi_0'(\cdot+)$, $\varphi_0'(x) - \varphi_0'(y) \geq C(y-x)$ for some $C > 0$ and all $A \leq x < y \leq B$, then

$$\sup_{t \in T_n} |\widehat{F}_n(t) - \mathbb{F}_n(t)| = O_p\left(\left(\frac{\log n}{n}\right)^{3\beta/(4\beta+2)}\right).$$

  where $3\beta/(2\beta+4) \in [1/2, 3/5] = [.5, .6]$ for $1 \leq \beta \leq 2$.

- If $\beta > 1$, this implies $\sup_{t \in T_n} |\widehat{F}_n(t) - \mathbb{F}_n(t)| = o_p(n^{-1/2})$.

B. Nonparametric estimation, log-concave on $\mathbb{R}$



B. Nonparametric estimation, log-concave on $\mathbb{R}$

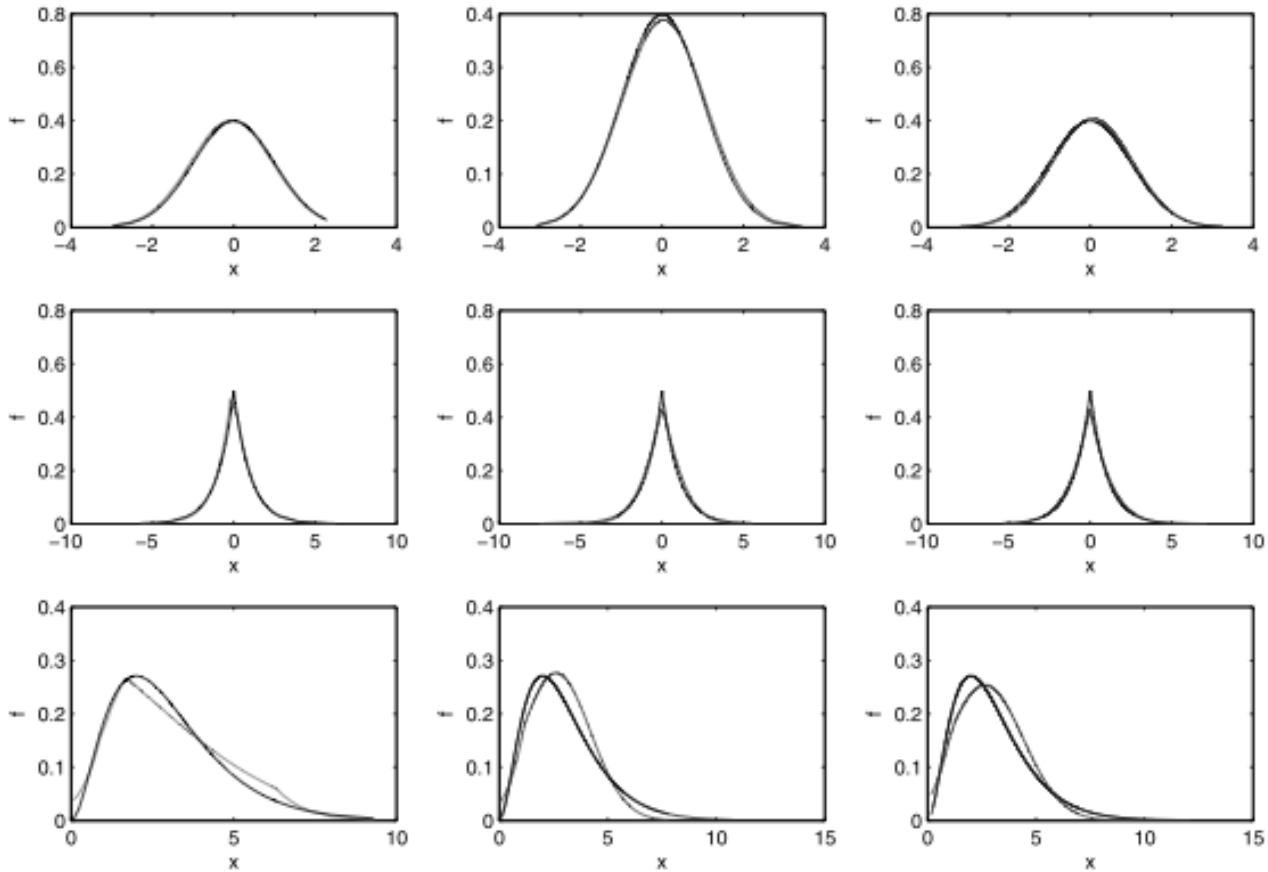


FIG 2. *The estimated log-concave density for different simulation examples. The sample sizes are 50,100 and 200 respectively for first, second and third columns. The three rows correspond to simulations from a Normal(0,1), a double-exponential and a Gamma(3,2) density. The bold one corresponds to the true density and the dotted one is the estimator.*

B. Nonparametric estimation, log-concave on $\mathbb{R}$

50 *L. Dümbgen and K. Rufibach*

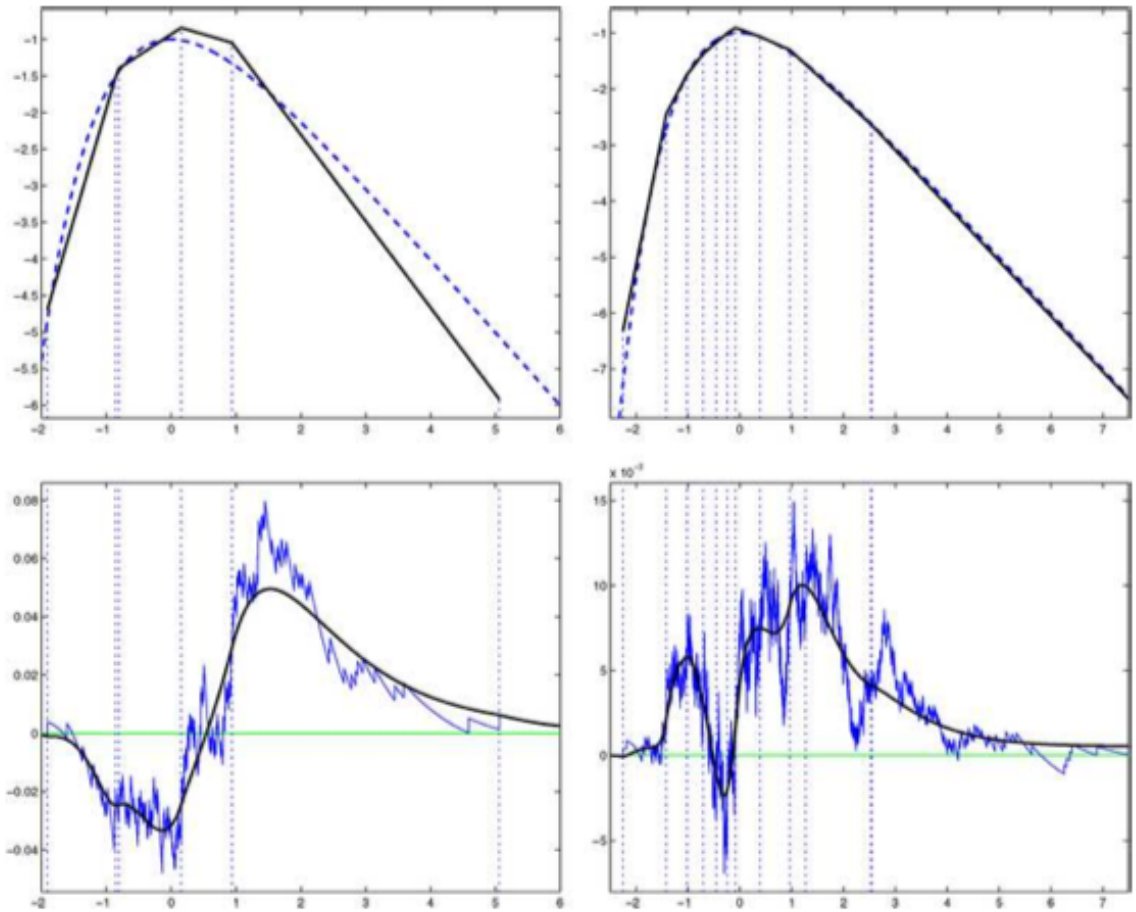


Figure 3. Density functions and empirical processes for Gumbel samples of size $n = 200$ and $n = 2000$.

Top panel: Estimated log-densities. Bottom panel: Empirical processes

B. Nonparametric estimation, log-concave on $\mathbb{R}$

Estimating log-concave densities 45

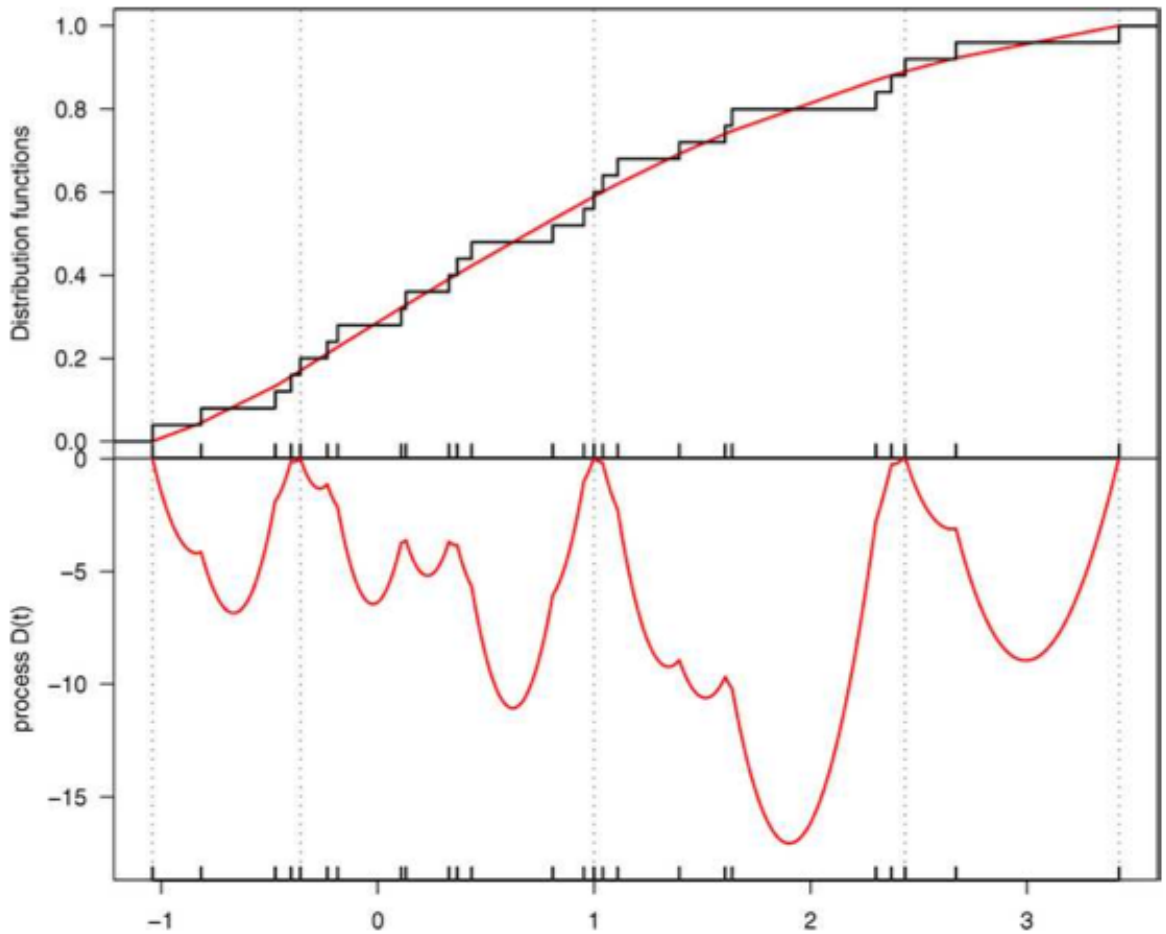


Figure 1. Distribution functions and the process $D(t)$ for a Gumbel sample.

C: Limit theory at a fixed point in $\mathbb{R}$

Assumptions:
- f_0 is log-concave, $f_0(x_0) > 0$.
- If $\varphi_0''(x_0) \neq 0$, then $k = 2$;
  otherwise, k is the smallest integer such that
  $\varphi_0^{(j)}(x_0) = 0$, $j = 2, \ldots, k-1$, $\varphi_0^{(k)}(x_0) \neq 0$.
- $\varphi_0^{(k)}$ is continuous in a neighborhood of x_0.

Example: $f_0(x) = C\exp(-x^4)$ with $C = \sqrt{2}\Gamma(3/4)/\pi$: $k = 4$.

Driving process: $Y_k(t) = \int_0^t W(s)ds - t^{k+2}$, W standard 2-sided Brownian motion.

Invelope process: H_k determined by limit Fenchel relations:

- $H_k(t) \leq Y_k(t)$ for all $t \in \mathbb{R}$
- $\int_{\mathbb{R}} (H_k(t) - Y_k(t)) dH_k^{(3)}(t) = 0$.
- $H_k^{(2)}$ is concave.

C: Limit theory at a fixed point in $\mathbb{R}$

Theorem. (Balabdaoui, Rufibach, & W, 2009)

- Pointwise limit theorem for $\widehat{f}_n(x_0)$:

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

where

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

$$d_k \equiv \left(\frac{f_0(x_0)^{k+2} |\varphi_0^{(k)}(x_0)|^3}{[(k+2)!]^3} \right)^{1/(2k+1)}.$$

C: Limit theory at a fixed point in $\mathbb{R}$

- Pointwise limit theorem for $\widehat{\varphi}_n(x_0)$:

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

where

$$C_k \equiv \left(\frac{|\varphi_0^{(k)}(x_0)|}{f_0(x_0)^k (k+2)!} \right)^{1/(2k+1)},$$

$$D_k \equiv \left(\frac{|\varphi_0^{(k)}(x_0)|^3}{f_0(x_0)^{k-1}[(k+2)!]^3} \right)^{1/(2k+1)}.$$

- Proof: Use the same perturbation as for convex - decreasing density proof with perturbation version of characterization:

D: Mode estimation, log-concave density on $\mathbb{R}$

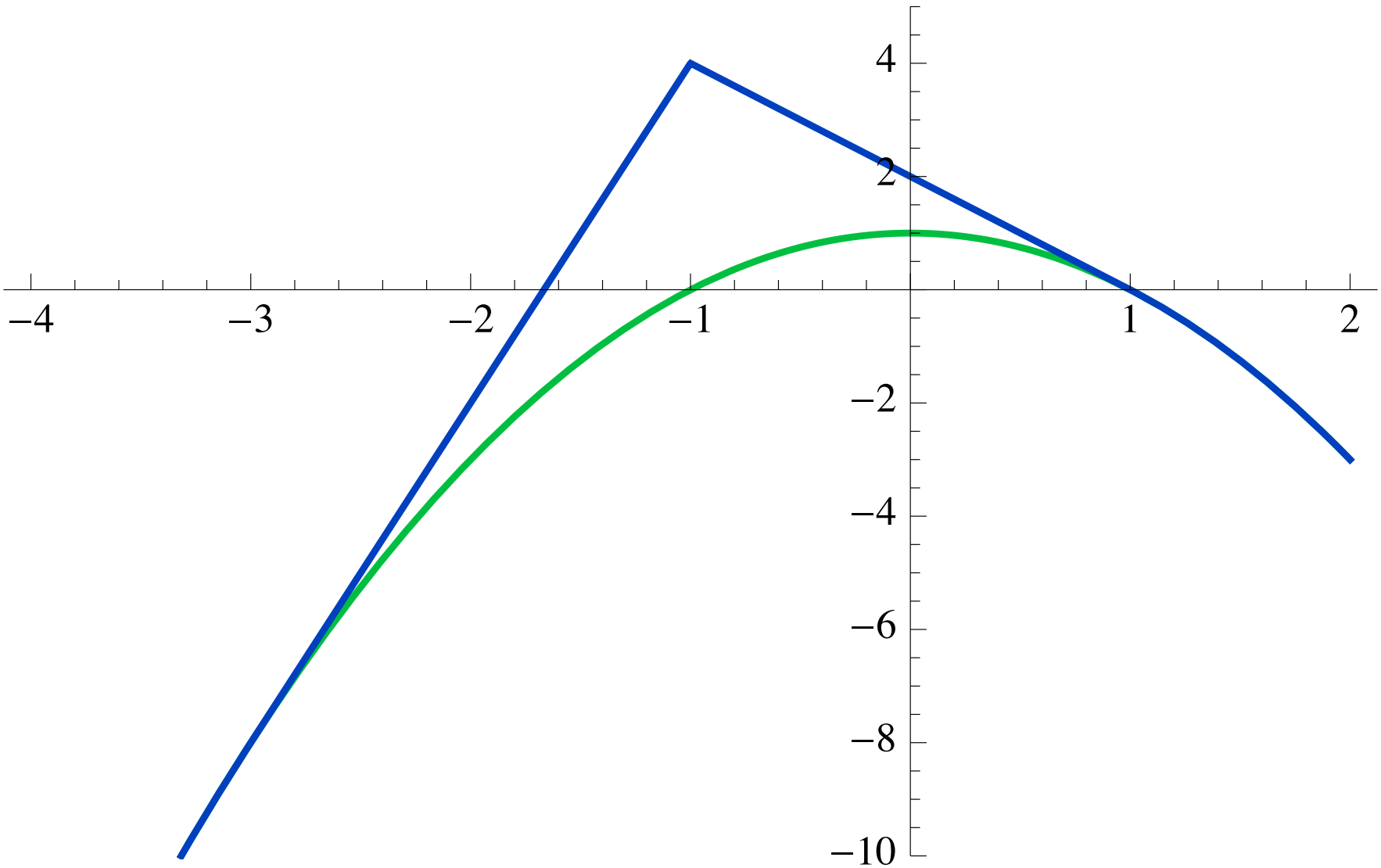
Let $x_0 = M(f_0)$ be the *mode* of the log-concave density f_0, recalling that $\mathcal{P}_0 \subset \mathcal{P}_{unimodal}$. Lower bound calculations using Jongbloed's perturbation φ_ϵ of φ_0 yields:

Proposition. If $f_0 \in \mathcal{P}_0$ satisfies $f_0(x_0) > 0$, $f_0''(x_0) < 0$, and f_0'' is continuous in a neighborhood of x_0, and T_n is any estimator of the mode $x_0 \equiv M(f_0)$, then $f_n \equiv \exp(\varphi_{\epsilon_n})$ with $\epsilon_n \equiv \nu n^{-1/5}$ and $\nu \equiv 2f_0''(x_0)^2/(5f_0(x_0))$,

$$\liminf_{n\to\infty} n^{1/5} \inf_{T_n} \max\{E_n|T_n - M(f_n)|, E_0|T_n - M(f_0)|\}$$
$$\geq \frac{1}{4}\left(\frac{5/2}{10e}\right)^{1/5}\left(\frac{f_0(x_0)}{f_0''(x_0)^2}\right)^{1/5}.$$

Does the MLE $M(\widehat{f}_n)$ achieve this?

D: Mode estimation, log-concave density on $\mathbb{R}$



D: Mode estimation, log-concave density on $\mathbb{R}$

Proposition. (Balabdaoui, Rufibach, & W, 2009)
Suppose that $f_0 \in \mathcal{P}_0$ satisfies:

- $\varphi_0^{(j)}(x_0) = 0$, $j = 2, \ldots, k-1$,
- $\varphi_0^{(k)}(x_0) \neq 0$, and
- $\varphi_0^{(k)}$ is continuous in a neighborhood of x_0.

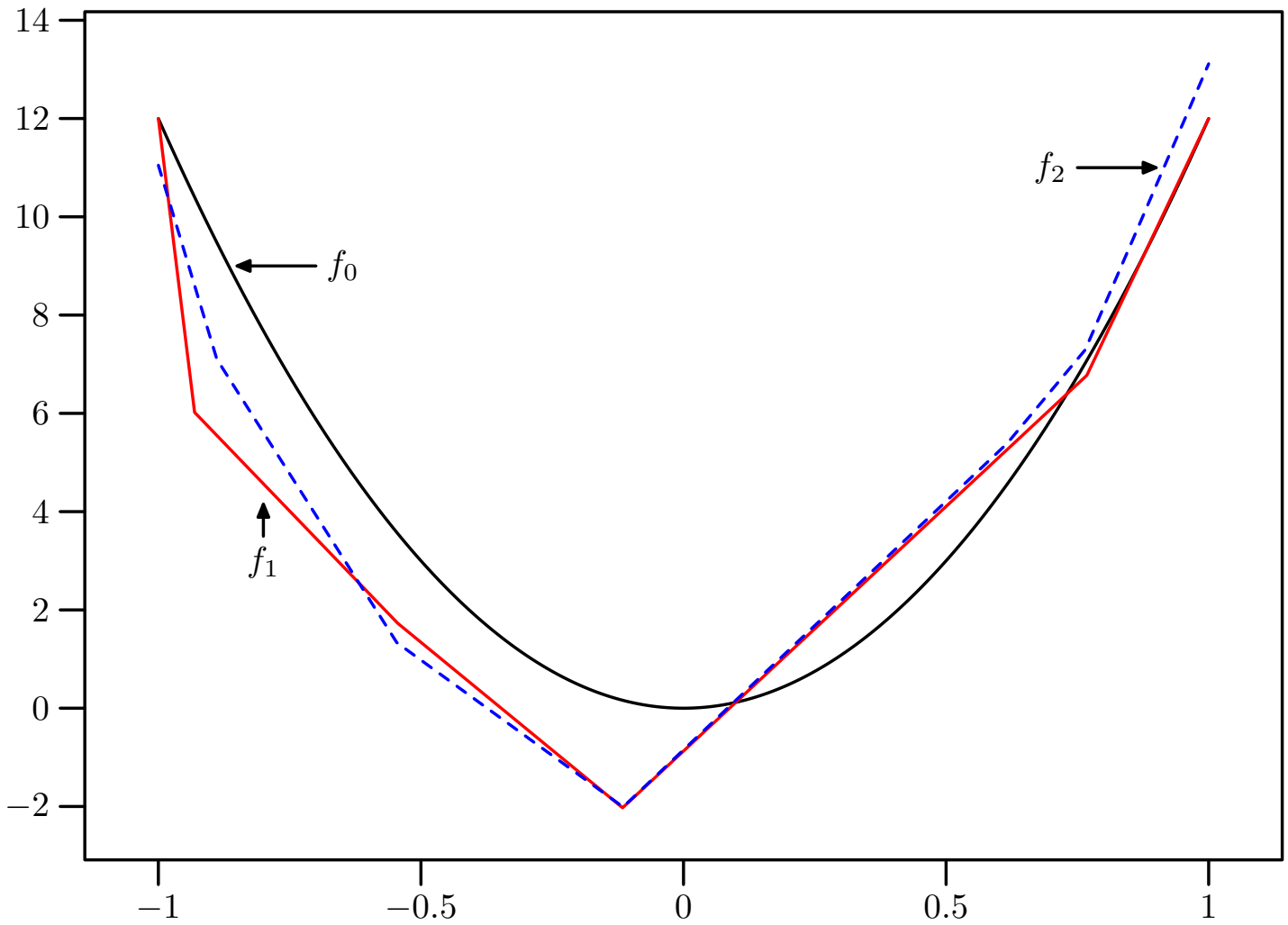
Then $\widehat{M}_n \equiv M(\widehat{f}_n) \equiv \min\{u : \widehat{f}_n(u) = \sup_t \widehat{f}_n(t)\}$, satisfies

$$n^{1/(2k+1)}(\widehat{M}_n - M(f_0)) \to_d \left(\frac{((k+2)!)^2 f_0(x_0)}{f_0^{(k)}(x_0)^2}\right)^{1/(2k+1)} M(H_k^{(2)})$$

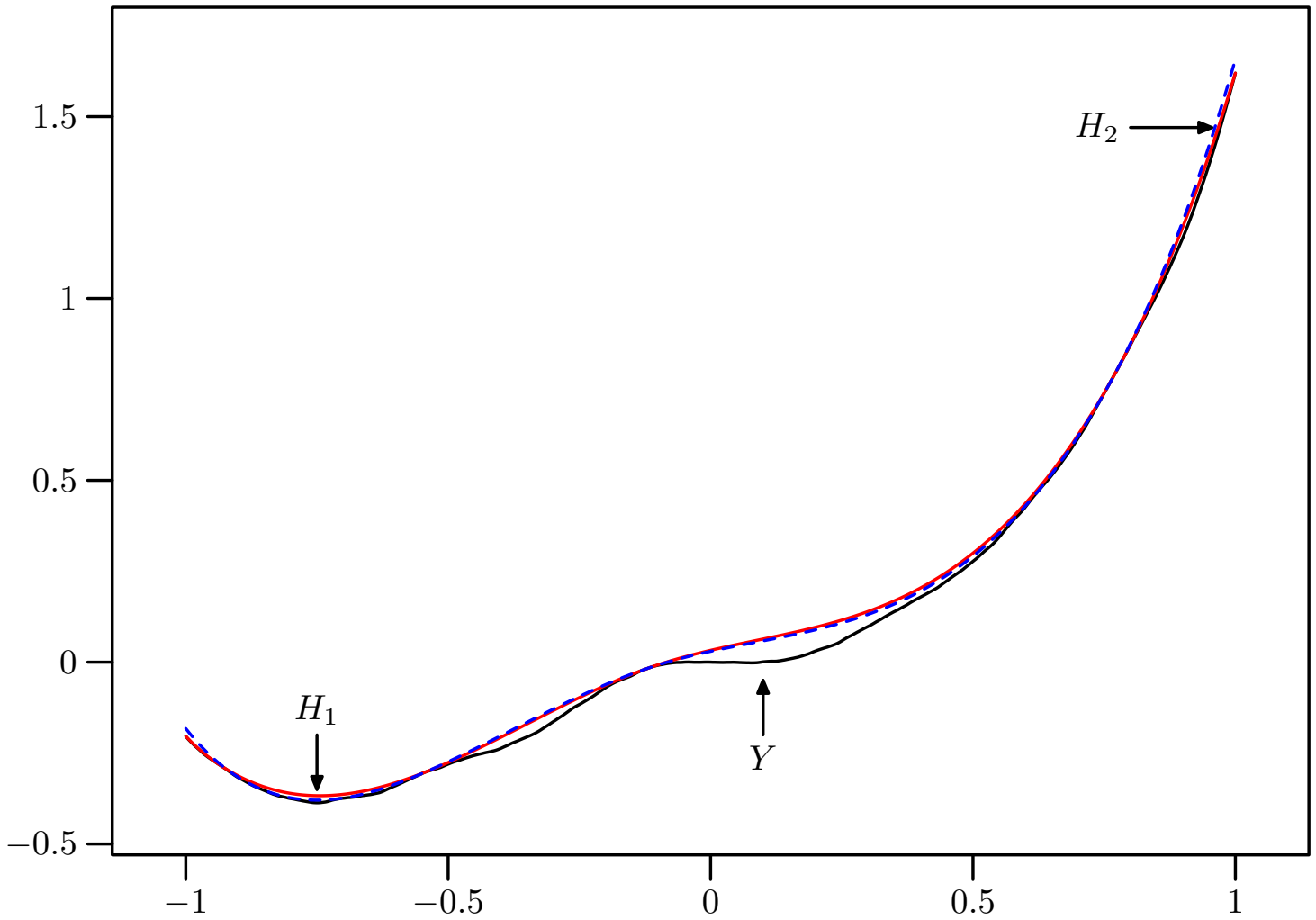
where $M(H_k^{(2)}) = \text{argmax}(H_k^{(2)})$.

Note that when $k = 2$ this agrees with the lower bound calculation, at least up to absolute constants.

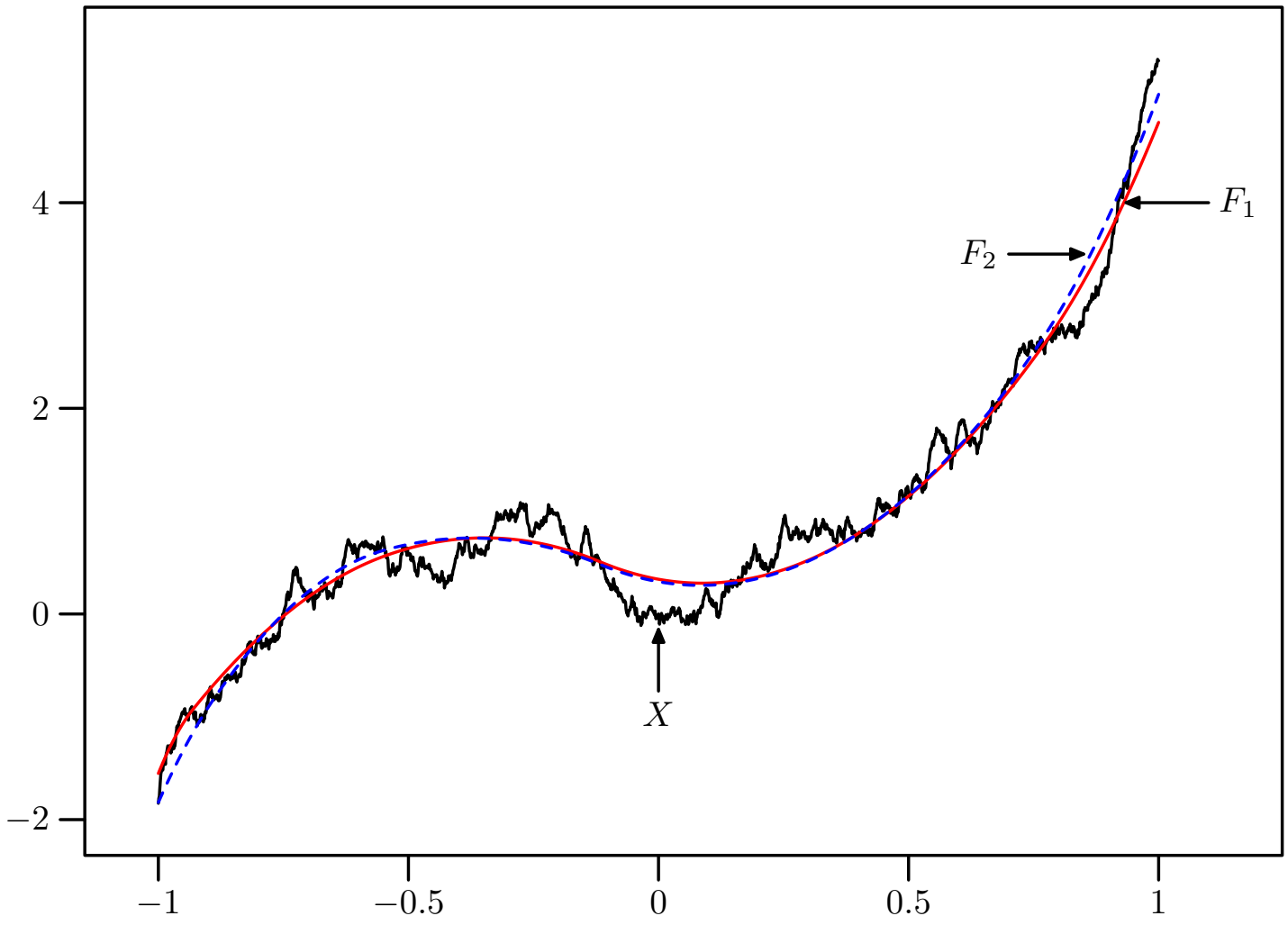
D: Mode estimation, log-concave density on $\mathbb{R}$



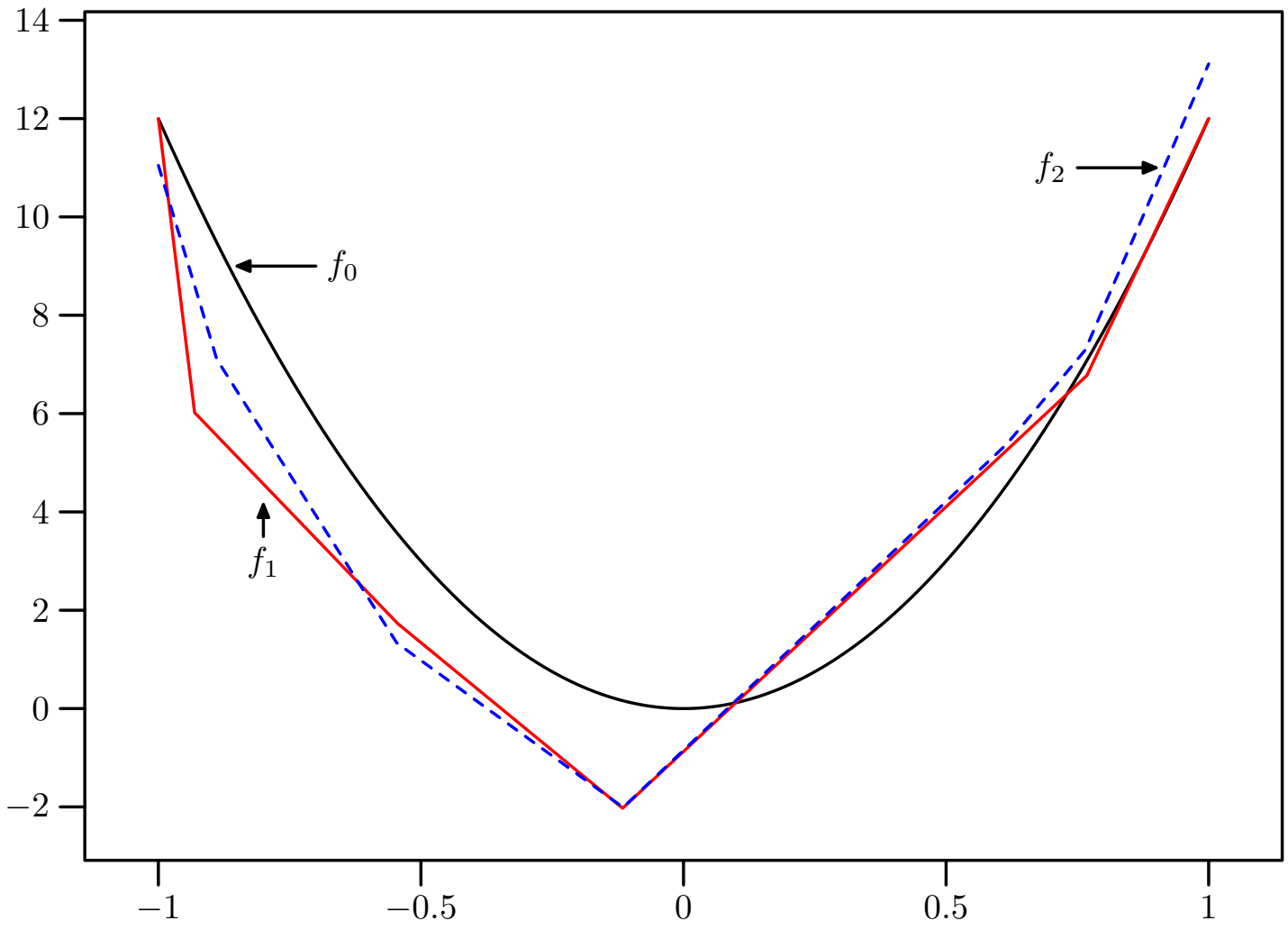
D: Mode estimation, log-concave density on $\mathbb{R}$



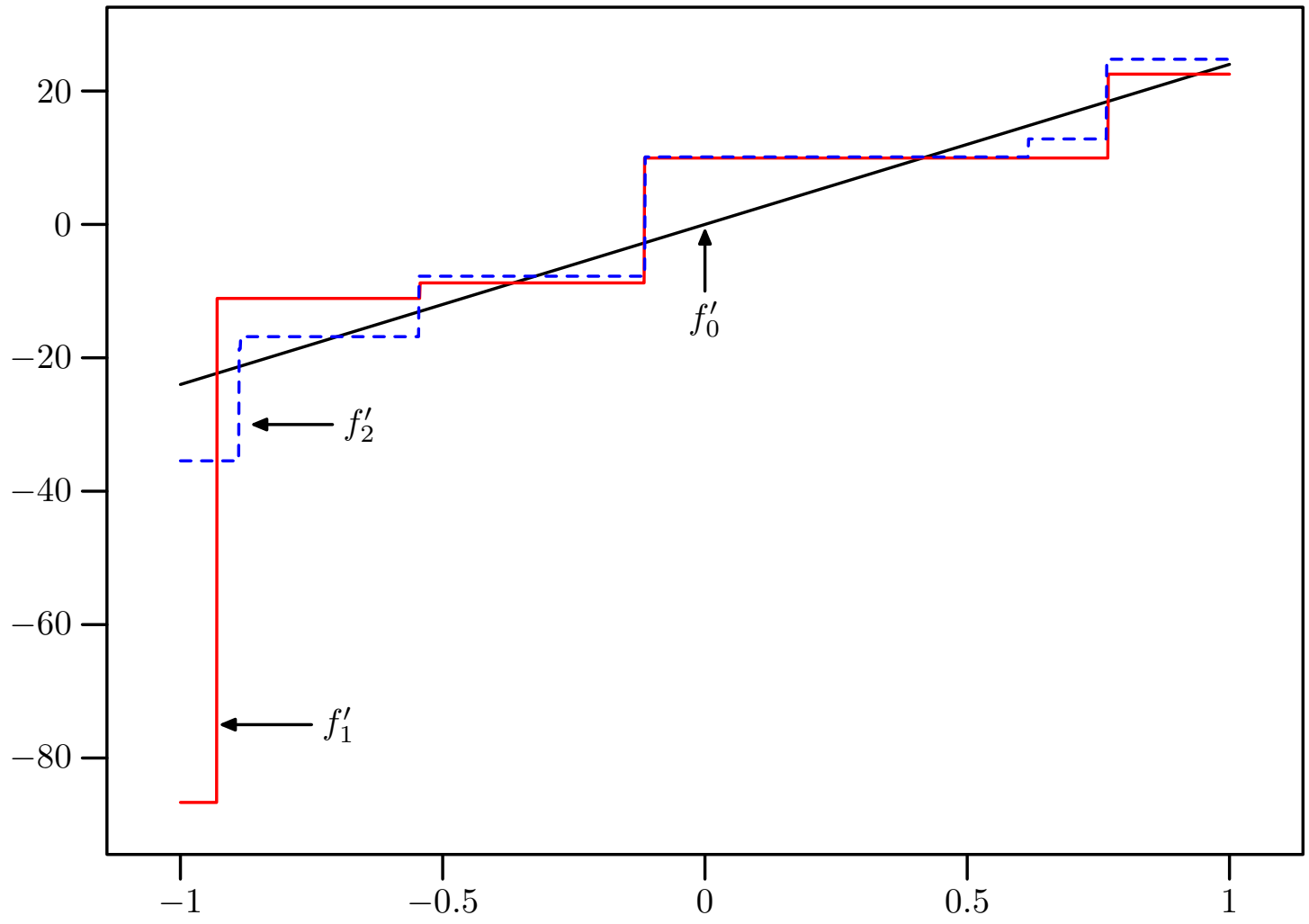
D: Mode estimation, log-concave density on $\mathbb{R}$



D: Mode estimation, log-concave density on $\mathbb{R}$



D: Mode estimation, log-concave density on $\mathbb{R}$



E: Generalizations of log-concave to $\mathbb{R}$ and $\mathbb{R}^d$:

Three generalizations:

- log$-$concave densities on $\mathbb{R}^d$
  (Cule, Samworth, and Stewart, 2010)
- $s-$concave and $h-$ transformed convex densities on $\mathbb{R}^d$
  (Seregin, 2010)
- Hyperbolically $k-$monotone and completely monotone densities on $\mathbb{R}$; (Bondesson, 1981, 1992)

E: Generalizations of log-concave to $\mathbb{R}$ and $\mathbb{R}^d$:

Log-concave densities on $\mathbb{R}^d$:

- A density f on $\mathbb{R}^d$ is log-concave if $f(x) = \exp(\varphi(x))$ with φ concave.
- Some properties:
  - ▷ Any log$-$concave f is unimodal
  - ▷ The level sets of f are closed convex sets
  - ▷ Convolutions of log-concave distributions are log-concave.
  - ▷ Marginals of log-concave distributions are log-concave.

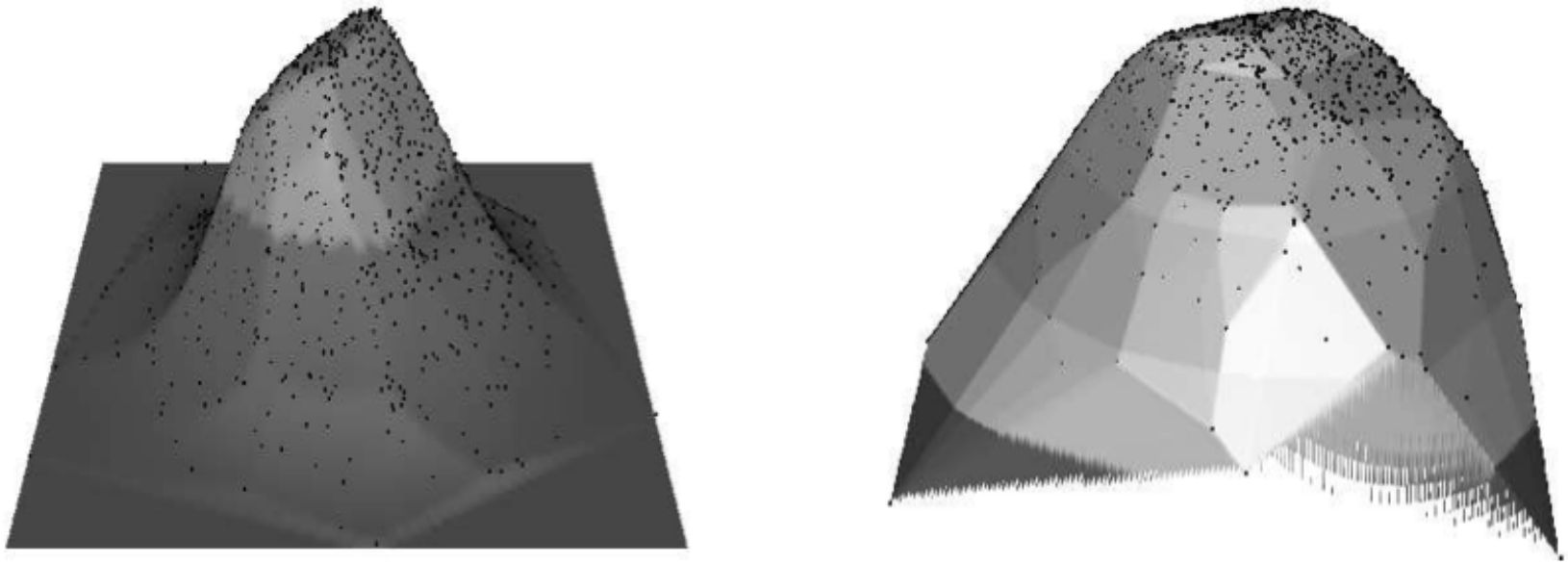
E: Generalizations of log-concave to $\mathbb{R}$ and $\mathbb{R}^d$:

MLE of $f \in \mathcal{P}_0(\mathbb{R}^d)$: (Cule, Samworth, Stewart, 2010)

- MLE $\widehat{f}_n = \text{argmax}_{f \in \mathcal{P}_0(\mathbb{R}^d)} \mathbb{P}_n \log f$ exists and is unique if $n \geq d + 1$.
- The estimator $\widehat{\varphi}_n$ of φ_0 is a "taut tent" stretched over "tent poles" of certain heights at a subset of the observations.
- Computable via non-differentiable convex optimization methods: Shor's (1985) $r-$algorithm: $R-$package LogConcDEAD (Cule, Samworth, Stewart, 2008).

E: Generalizations of log-concave to $\mathbb{R}$ and $\mathbb{R}^d$:

Log-concave density estimation 3



(a) Density **(b)** Log-density

Fig. 3. Log-concave maximum likelihood estimates based on 1000 observations (plotted as dots) from a standard bivariate normal distribution.

E: Generalizations of log-concave to $\mathbb{R}$ and $\mathbb{R}^d$:

- If f_0 is *any* density on $\mathbb{R}^d$ with $\int_{\mathbb{R}^d} \|x\| f_0(x)dx < \infty$, $\int_{\mathbb{R}^d} f_0(x)\log f_0(x)dx < \infty$, and $\{x \in \mathbb{R}^d : f_0(x) > 0\}^\circ = \text{int}(\text{supp}(f_0)) \neq \emptyset$, then $\widehat{f}_n$ satisfies:

$$\int_{\mathbb{R}^d} |\widehat{f}_n(x) - f^*(x)|dx \to_{a.s.} 0$$

where, for the Kullback-Leibler divergence

$$K(f_0, f) = \int f_0 \log(f_0/f)d\mu,$$

$$f^* = \text{argmin}_{f \in \mathcal{P}_0(\mathbb{R}^d)} K(f_0, f)$$

is the "pseudo-true" density in $\mathcal{P}_0(\mathbb{R}^d)$ corresponding to f_0. In fact:

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

for any $a < a_0$ where $f^*(x) \leq \exp(-a_0\|x\| + b_0)$.

Questions and Problems:

Questions:

- Rates of convergence? Limiting distribution(s) for $d > 1$? (n^r with $r = 2/(4+d)$?)
- MLE (rate-) inefficient for $d \geq 4$? How to penalize to get efficient rates?
- Multivariate classes with nice preservation/closure properties and smoother than log-concave?
- Can we treat $\widehat{f}_n \in \mathcal{P}_h$ with miss-specification: $f_0 \notin \mathcal{P}_h$?
- Algorithms for computing $\widehat{f}_n \in \mathcal{P}_h$?
- Related results for convex regression on $\mathbb{R}^d$: Seijo and Sen, *Ann. Statist.* (2011).

Further information:

- Other Jon W talks : www.stat.washington.edu/jaw/RESEARCH/TALKS/talks.html
- Jon W Probability Seminar: www.stat.washington.edu/jaw/RESEARCH/TALKS/UW-Prob-October2016.pdf
- Jon W Frejus lectures:
  - jaw/RESEARCH/TALKS/FrejusDay1-small.pdf
  - jaw/RESEARCH/TALKS/FrejusDay2-small.pdf
  - jaw/RESEARCH/TALKS/FrejusDay3-small.pdf
  - jaw/RESEARCH/TALKS/FrejusDay4-small.pdf
  - jaw/RESEARCH/TALKS/FrejusDay5-small.pdf

- Groeneboom & Jongbloed (2014). *Nonparametric Estimation under Shape Constraints*. Cambridge U. Press.
- Kim, A. K., Guntuboyina, A., and Samworth (2016). Adaptation in log-concave density estimation. arXiv:1609.00861.
- Chatterjee, S., Guntuboyina, A., and Sen, B. (2015). On risk bounds in isotonic and other shape restricted regression problems. *Ann. Statist.* **43**, 1774-1800.
- Guntuboyina, A. and Sen, B. (2015). Global risk bounds and adaptation in univariate convex regression. *Prob. Theor. Rel. Fields* **163**, 379-411.

Thanks!

Skiing toward the Nisqually Glacier