Nonparametric estimation of log-concave densities

Jon A. Wellner

University of Washington, Seattle

Northwestern University

November 5, 2010

Conference on Shape Restrictions

in Non- and Semi-Parametric Estimation

of Econometric Models

Based on joint work with:

- Fadoua Balabdaoui
- Kaspar Rufibach
- Arseni Seregin

Outline

- A: Log-concave densities on $\mathbb{R}^1$
- B: Nonparametric estimation, log-concave on $\mathbb{R}$
- C: Limit theory at a fixed point in $\mathbb{R}$
- D: Estimation of the mode, log-concave density on $\mathbb{R}$
- E: Generalizations: $s-$concave densities on $\mathbb{R}$ and $\mathbb{R}^d$
- F: Summary; problems and open questions

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:
  - $\triangleright$ Normal (μ, σ^2)
  - $\triangleright$ Uniform(a, b)
  - $\triangleright$ Gamma(r, λ) for $r \geq 1$
  - $\triangleright$ 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", PFF_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.

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 $H(\widehat{f}_n, f_0) \rightarrow_{a.s.} 0$.

- (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 $\mathbb{R}$

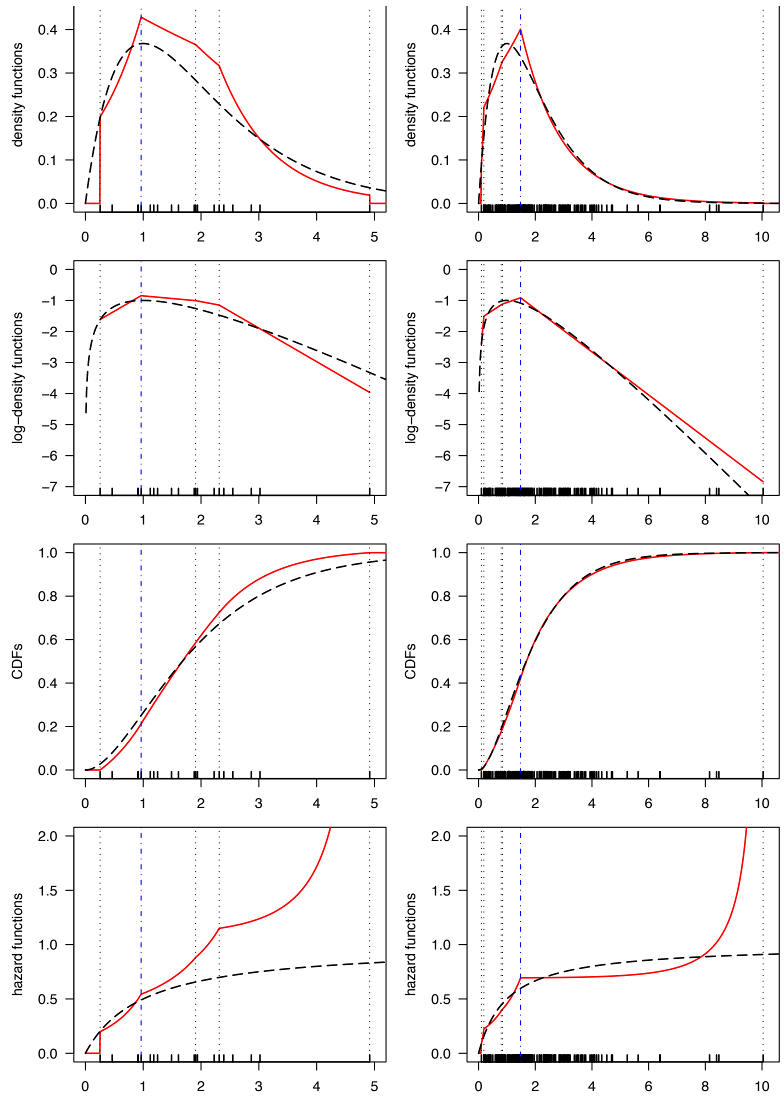
- If $\varphi_0 \in \mathcal{H}^{\beta,L}(T)$ as above and, with $\varphi_0' = \varphi_0(\cdot -)$ or $\varphi_0'(\cdot +)$, $\varphi_0'(x) - \varphi_0'(y) \ge C(y-x)$ for some $C > 0$ and all $A \le x < y \le B$, then

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

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

- 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 $\mathbb{R}$

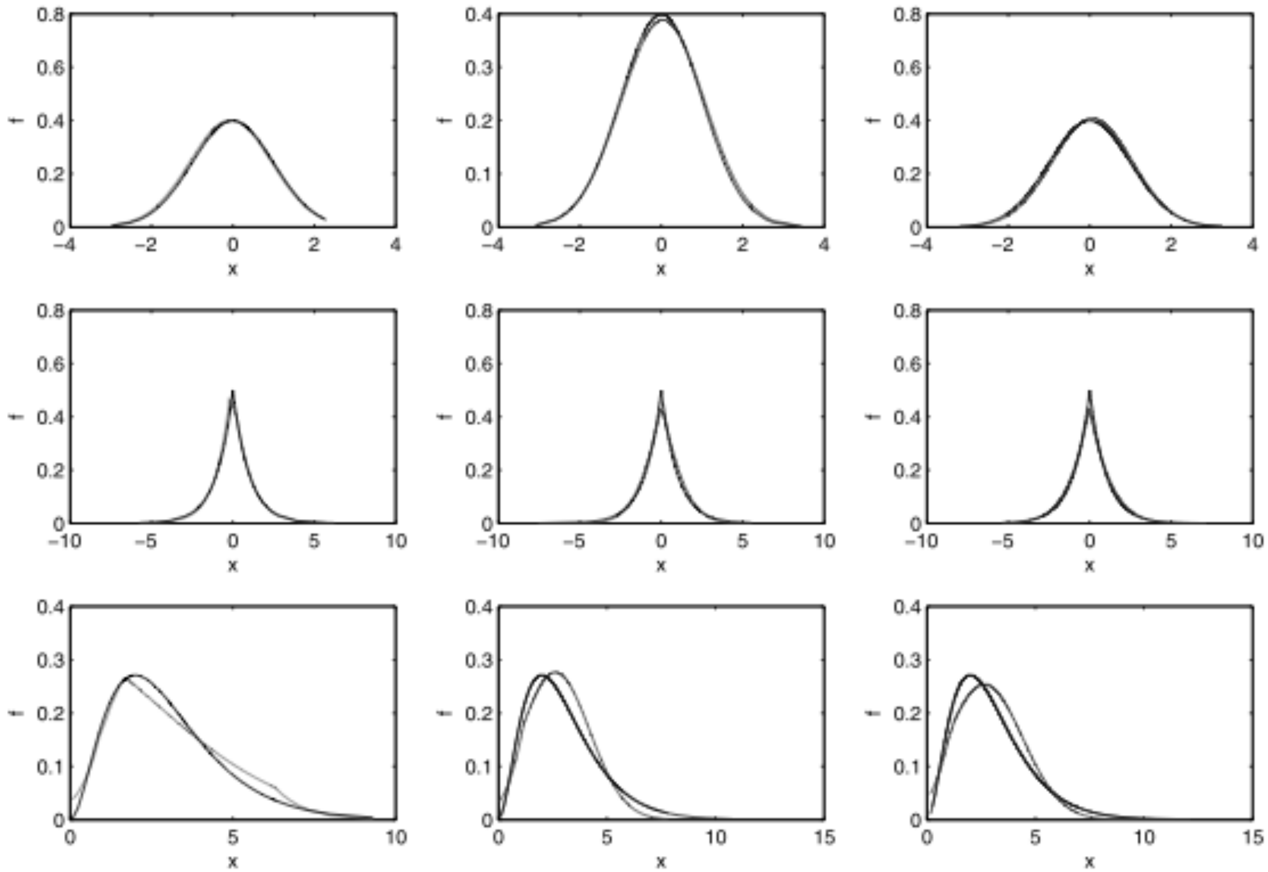


FIG 2. *The estimated log-concave density for different simulation examples. The sample sizes are 50,100 and 200 respectively for first, second and third columns. The three rows correspond to simulations from a Normal(0,1), a double-exponential and a Gamma(3,2) density. The bold one corresponds to the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

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

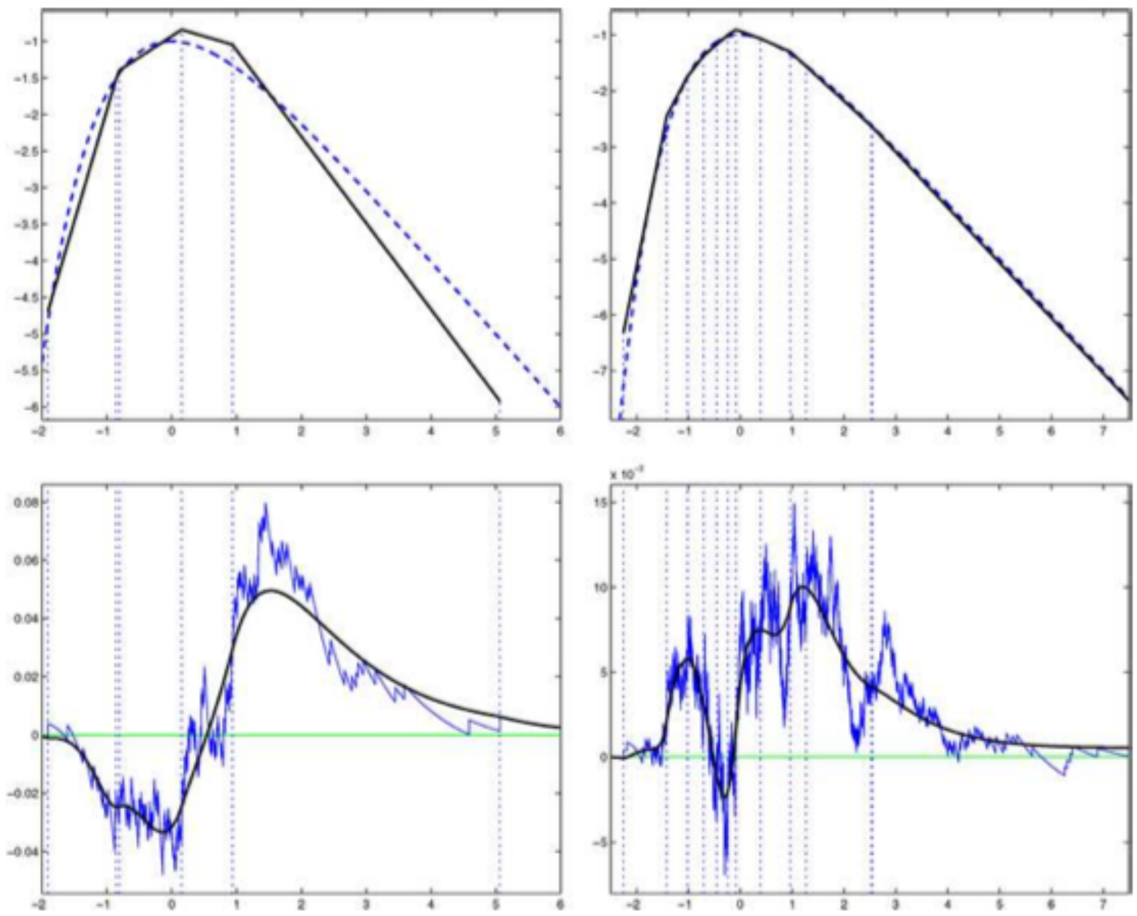


Figure 3. Density functions and empirical processes for Gumbel samples of size $n = 200$ and $n = 2000$.

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

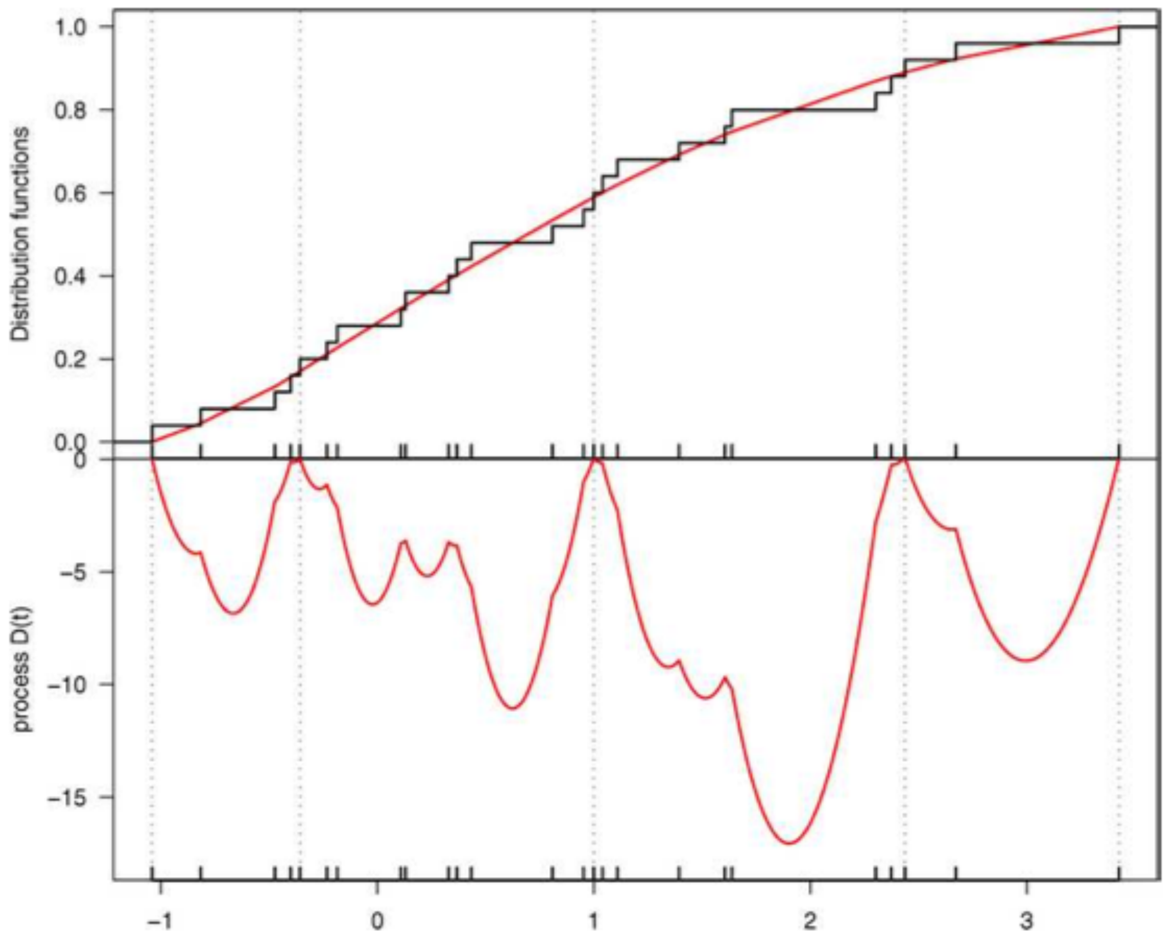


Figure 1. Distribution functions and the process $D(t)$ for a Gumbel sample.

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 $\widehat{\varphi}_n(x_0)$:

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

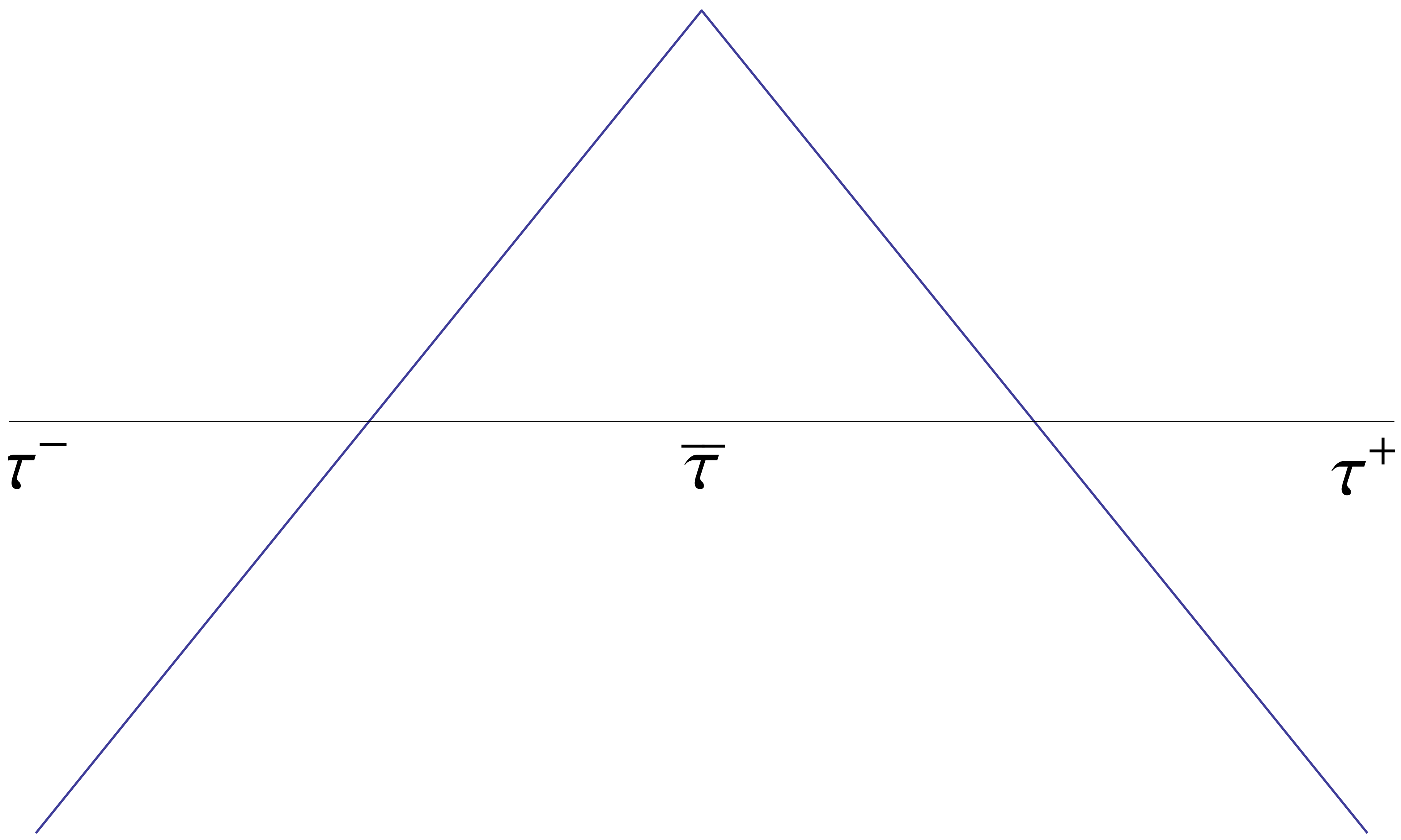
where

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

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

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

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



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

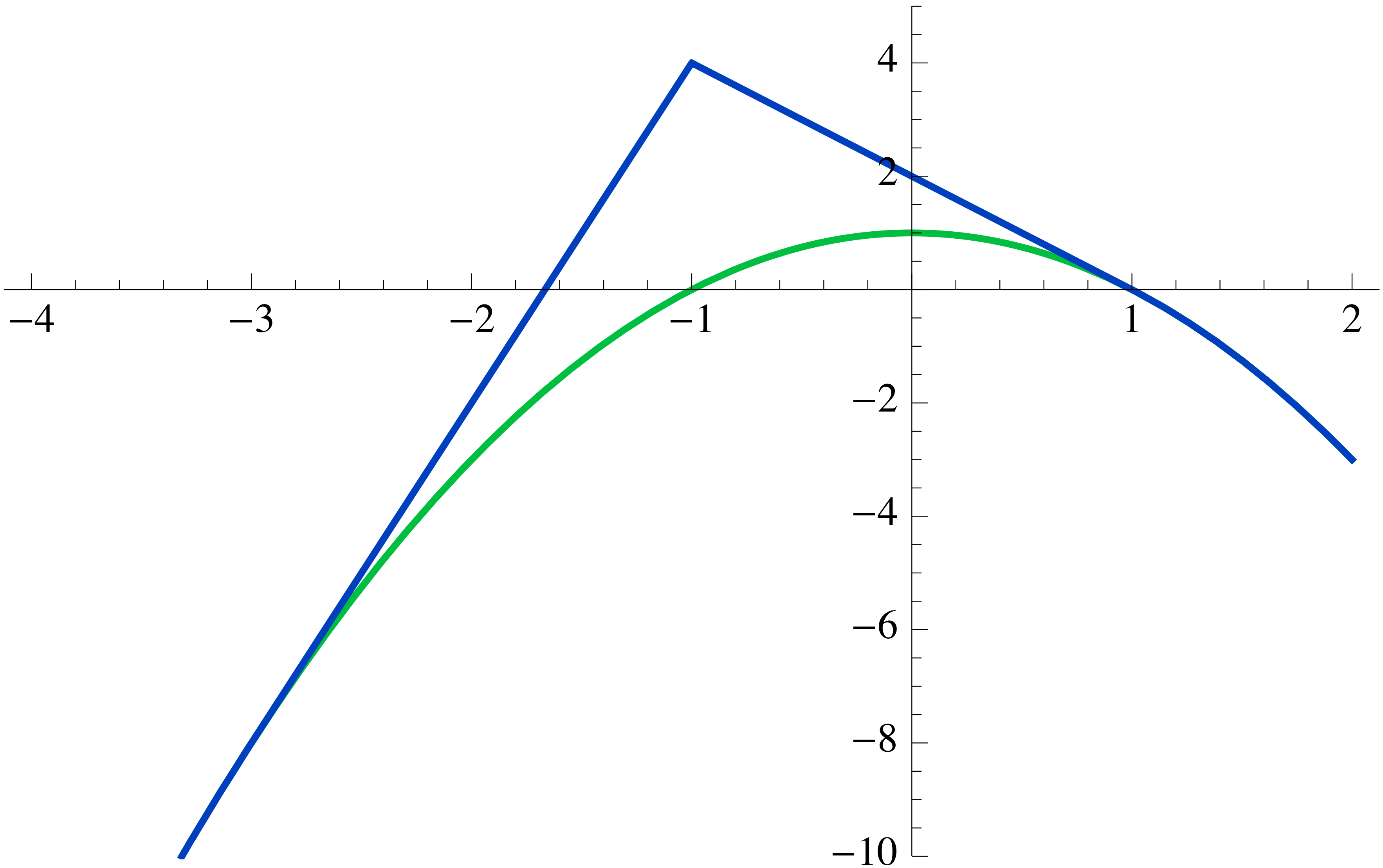
Let $x_0 = M(f_0)$ be the *mode* of the log-concave density f_0, recalling that $\mathcal{P}_0 \subset \mathcal{P}_{unimodal}$. Lower bound calculations using Jongbloed's perturbation φ_ϵ of φ_0 yields:

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 $\nu \equiv 2f_0''(x_0)^2/(5f_0(x_0))$,

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

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.

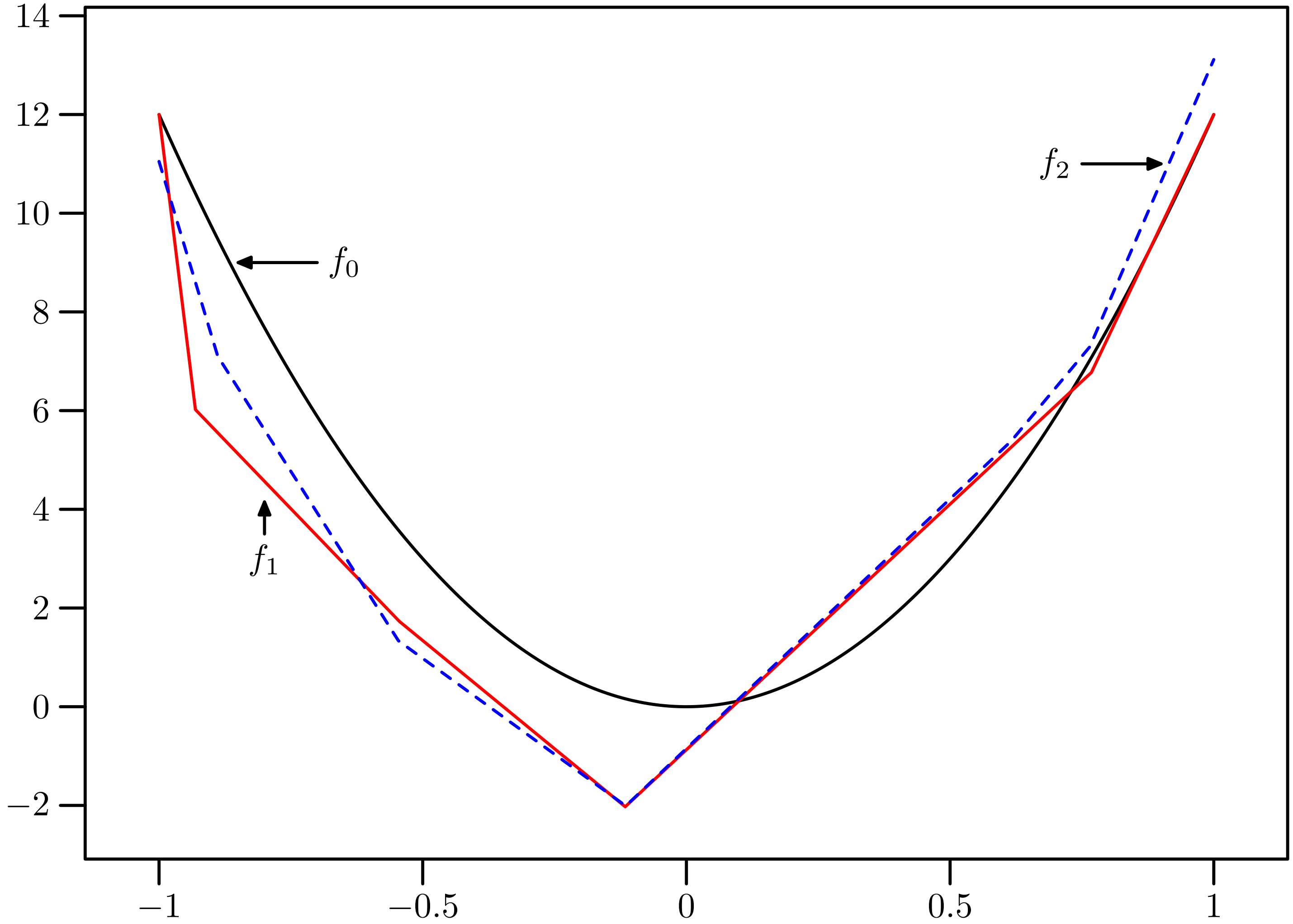
Then $\widehat{M}_n \equiv M(\widehat{f}_n) \equiv \min\{u : \widehat{f}_n(u) = \sup_t \widehat{f}_n(t)\}$, satisfies

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

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 $\mathbb{R}$

When $f_0 = \phi$, the standard normal density, $M(f_0) = 0$, $f_0(0) = (2\pi)^{-1/2}$, $f_0''(0) = -(2\pi)^{-1/2}$, and hence

$$\left(\frac{((4)!)^2 f_0(0)}{f_0^{(2)}(x_0)^2}\right)^{1/5} = \left(\frac{24^2(2\pi)^{-1/2}}{(2\pi)^{-1}}\right)^{1/5} = 4.28452\ldots$$

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 $\mathbb{R}^d$:

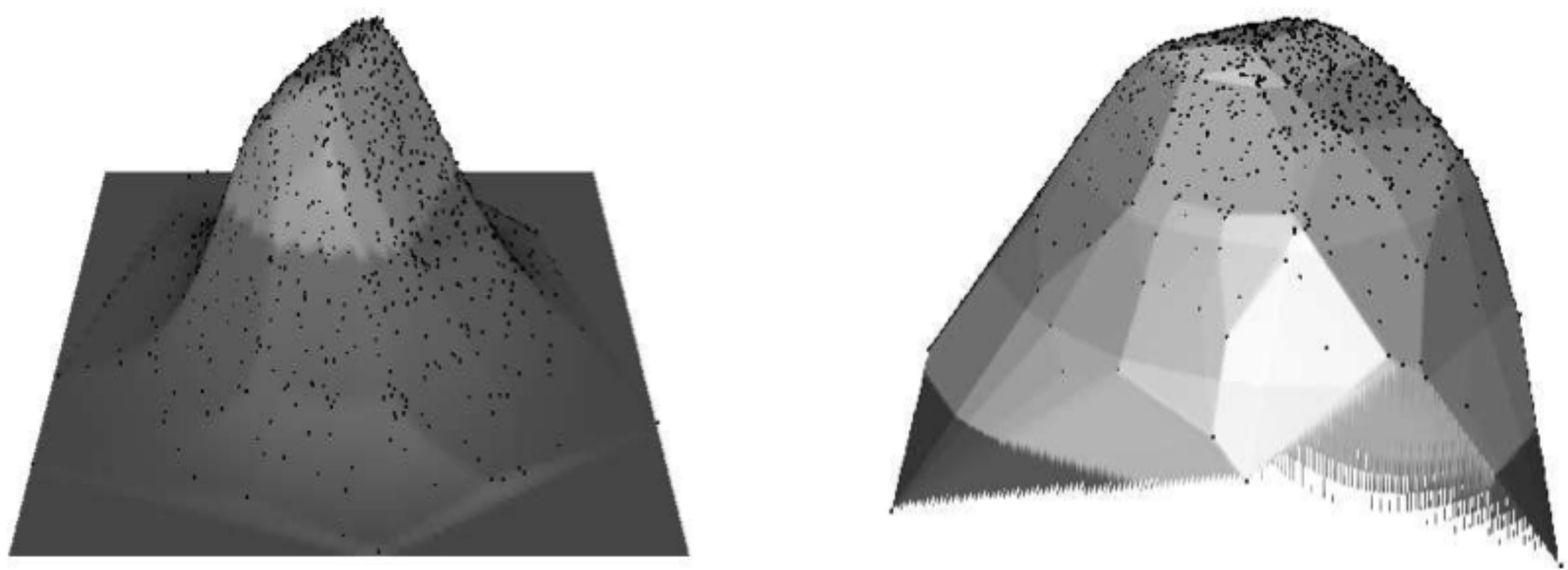
- A density f on $\mathbb{R}^d$ is log-concave if $f(x) = \exp(\varphi(x))$ with φ concave.
- Some properties:
  - $\triangleright$ Any log−concave f is unimodal
  - $\triangleright$ The level sets of f are closed convex sets
  - $\triangleright$ Convolutions of log-concave distributions are log-concave.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



(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 $f^*(x) \leq \exp(-a_0\|x\| + b_0)$.

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

$r-$concave and $h-$ transformed convex densities on $\mathbb{R}^d$: (Seregin, 2010; Seregin &, 2010)

Generalization to $s-$concave densities: A density f on $\mathbb{R}^d$ is $r-$concave on $C \subset \mathbb{R}^d$ if

$$f(\lambda x + (1-\lambda)y) \geq M_r(f(x), f(y); \lambda)$$

for all $x, y \in C$ and $0 < \lambda < 1$ where

$$M_r(a,b;\lambda) = \begin{cases} ((1-\lambda)a^r + \lambda b^r)^{1/r}, & r \neq 0, a, b > 0, \\ 0, & r < 0, ab = 0 \\ a^{1-\lambda} b^{\lambda}, & r = 0. \end{cases}$$

Let $\mathcal{P}_r$ denote the class of all $r-$concave densities on C. For $r \leq 0$ it suffices to consider $C = \mathbb{R}^d$, and it is almost immediate from the definitions that if $f \in \mathcal{P}_r$ for some $r \leq 0$, then

$$f(x) = \left\{ \begin{array}{ll} g(x)^{1/r}, & r < 0 \\ \exp(-g(x)), & r = 0 \end{array} \right\} \qquad \text{for } \; g \; \text{ convex.}$$

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

- Long history: Avriel (1972), Prékopa (1973), Borell (1975), Rinott (1976), Brascamp and Lieb (1976)
- Nice connections to $t-$concave measures: (Borell, 1975)
- Known now in math-analysis as the Borell, Brascamp, Lieb inequality
- One way to get heavier tails than log-concave!

  Example: Multivariate $t-$density with $p-$degrees of freedom: if

$$f(x) = f(x; p, d) = \frac{\Gamma((d+p)/2)}{\Gamma(p/2)(p\pi)^{d/2}} \frac{1}{\left(1 + \frac{\|x\|^2}{p}\right)^{(d+p)/2}}$$

then $f \in \mathcal{P}_{-1/s}$ for $s \in (d, d+p]$; i.e. $f \in \mathcal{P}_r(\mathbb{R}^d)$ for $-1/(d+p) \le r < -1/d$.

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

A measure μ on $(\mathbb{R}, \mathcal{B})$ is called $t-$*concave* if for all $A, B \in \mathcal{B}$ and $0 \le \lambda \le 1$

$$\mu(\lambda A + (1-\lambda)B) \ge M_t(\mu(A), \mu(B), \lambda).$$

Theorem. (Borell, 1975) If $f \in \mathcal{P}_r$ with $-1/d \le r \le \infty$, then the measure $P = P_f$ defined by $P(A) = \int_A f(x)dx$ for Borel subsets A of $\mathbb{R}^d$ is $t-$concave with

$$t = \begin{cases} \frac{r}{1+dr}, & \text{if } -1/d < r < \infty, \\ -\infty, & \text{if } r = -1/d, \\ 1/d, & \text{if } r = \infty, \end{cases}$$

and conversely.

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

$h-$ **convex densities:** Seregin (2010), Seregin & W (2010))

$$f(\underline{x}) = h(\varphi(\underline{x})) \tag{1}$$

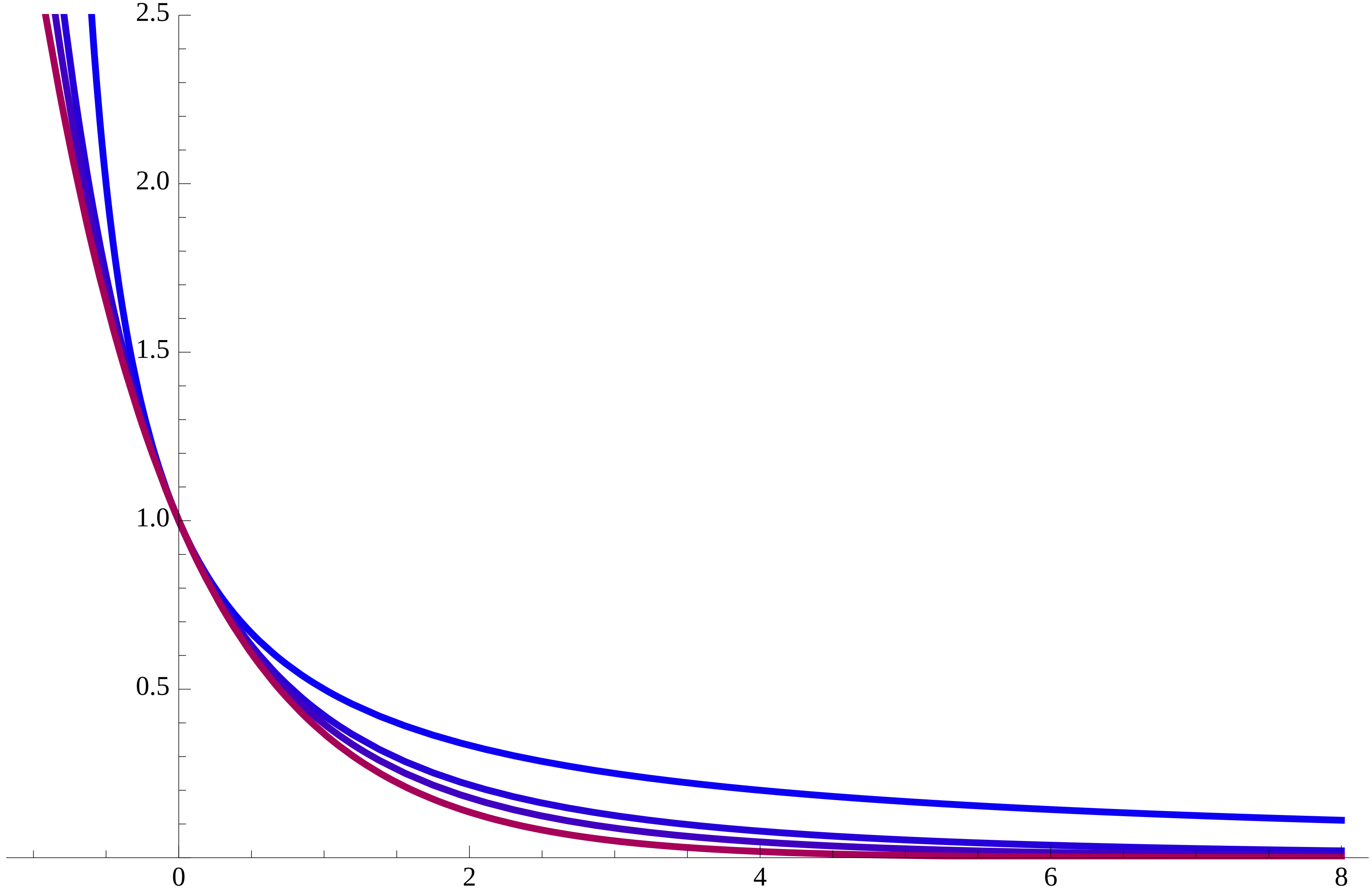
where $\varphi : \mathbb{R}^d \mapsto \mathbb{R}$ is convex, $h : \mathbb{R} \mapsto \mathbb{R}^+$ is decreasing and continuous; e.g. $h_s(u) \equiv (1 + u/s)^{-s}$ with $s > d$.

This motivates the following definition:

Definition. Say that $h : \mathbb{R} \to \mathbb{R}^+$ is a decreasing transformation if, with $y_0 \equiv \sup\{y : \ h(y) > 0\}$, $y_\infty \equiv \inf\{y : h(y) < \infty\}$,

- $h(y) = o(y^{-\alpha})$ for some $\alpha > d$ as $y \to \infty$.
- If $y_\infty > -\infty$, then $h(y) \asymp (y - y_\infty)^{-\beta}$ for some $\beta > d$ as $y \searrow y_\infty$.
- If $y_\infty = -\infty$, then $h(y)^\gamma h(-Cy) = o(1)$ as $y \to -\infty$ for some $\gamma, C > 0$.
- h is continuously differentiable on (y_∞, y_0).

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



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

Let $\mathcal{P}_h$ denote the collection of all densities on $\mathbb{R}^d$ of the form $f = h \circ \varphi$ for a fixed decreasing transformation h and φ convex, and let

$$\widehat{f}_n \equiv \text{argmax}_{f \in \mathcal{P}_h} \mathbb{P}_n \log f, \quad \text{the MLE.}$$

Theorem. $\widehat{f}_n \in \mathcal{P}_h$ exists if $n \ge \lceil n_d \rceil$ where

$$\begin{aligned} n_d &\equiv d + d\gamma 1\{y_\infty = -\infty\} + \frac{\beta d^2}{\alpha(\beta - d)} 1\{y_\infty > -\infty\} \\ &= \begin{cases} d+1, & \text{if } h(y) = e^{-y}, \\ d\left(\frac{s}{s-d}\right), & \text{if } h(y) = y^{-s}, \ s > d. \end{cases} \end{aligned}$$

Theorem. If h is a decreasing transformation as defined above, and $f_0 \in \mathcal{P}_h$, then

$$H(\widehat{f}_n, f_0) \rightarrow_{a.s.} 0.$$

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

Questions:

- Rates of convergence?
- 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 $\widehat{f}_n \in \mathcal{P}_h$?