Nonparametric estimation under Shape Restrictions

Jon A. Wellner

University of Washington, Seattle

Statistical Seminar, Frejus, France

August 30 - September 3, 2010

Outline: Five Lectures on Shape Restrictions

- **L1: Monotone functions: maximum likelihood and least squares**
- L2: Optimality of the MLE of a monotone density (and comparisons?)
- L3: Estimation of convex and $k-$monotone density functions
- L4: Estimation of log-concave densities: $d = 1$ and beyond
- L5: More on higher dimensions and some open problems

Outline: Lecture 1

- A: Maximum likelihood and least squares estimators (and more?)
- B: Switching: a simple key result
- C: Limit theory via switching and argmax continuous mapping
- D: Complements: Pollard's localization method ??
- E: Other nonparametric function estimation problems ??

A. Maximum likelihood, monotone density

- Model: $\mathcal{D} \equiv$ all monotone decreasing densities (wrt Lebesgue measure) on $\mathbb{R}^+ = (0, \infty)$.
- Observations: $X_1, \ldots, X_n$ i.i.d. $f_0 \in \mathcal{D}$.
- MLE: $\widehat{f}_n \equiv \text{argmax}_{f \in \mathcal{D}} \left\{ \sum_{i=1}^n \log f(X_i) \right\}$
- LSE: $\widetilde{f}_n \equiv \text{argmin}_{f \in \mathcal{D}} \psi_n(f)$
  where

$$
\begin{aligned}
\psi_n(f) \quad &\equiv \quad \frac{1}{2} \int_0^\infty f^2(x)dx - \int_0^\infty f(x)d\mathbb{F}_n(x) \\
&=? \quad \frac{1}{2} \left\{ \int_0^\infty (f^2(x) - f_n(x))^2 dx - \int_0^\infty f_n^2(x)dx \right\}
\end{aligned}
$$

if $\mathbb{F}_n$ had density f_n (which it doesn't, of course!).

A. Maximum likelihood, monotone density

Theorem. (a) $\widehat{f}_n = \widetilde{f}_n$ exists and is unique. It is a piecewise constant function with jumps (down) only at the order statistics.
(b) The MLE $\widehat{f}_n$ is characterized by the "Fenchel" conditions

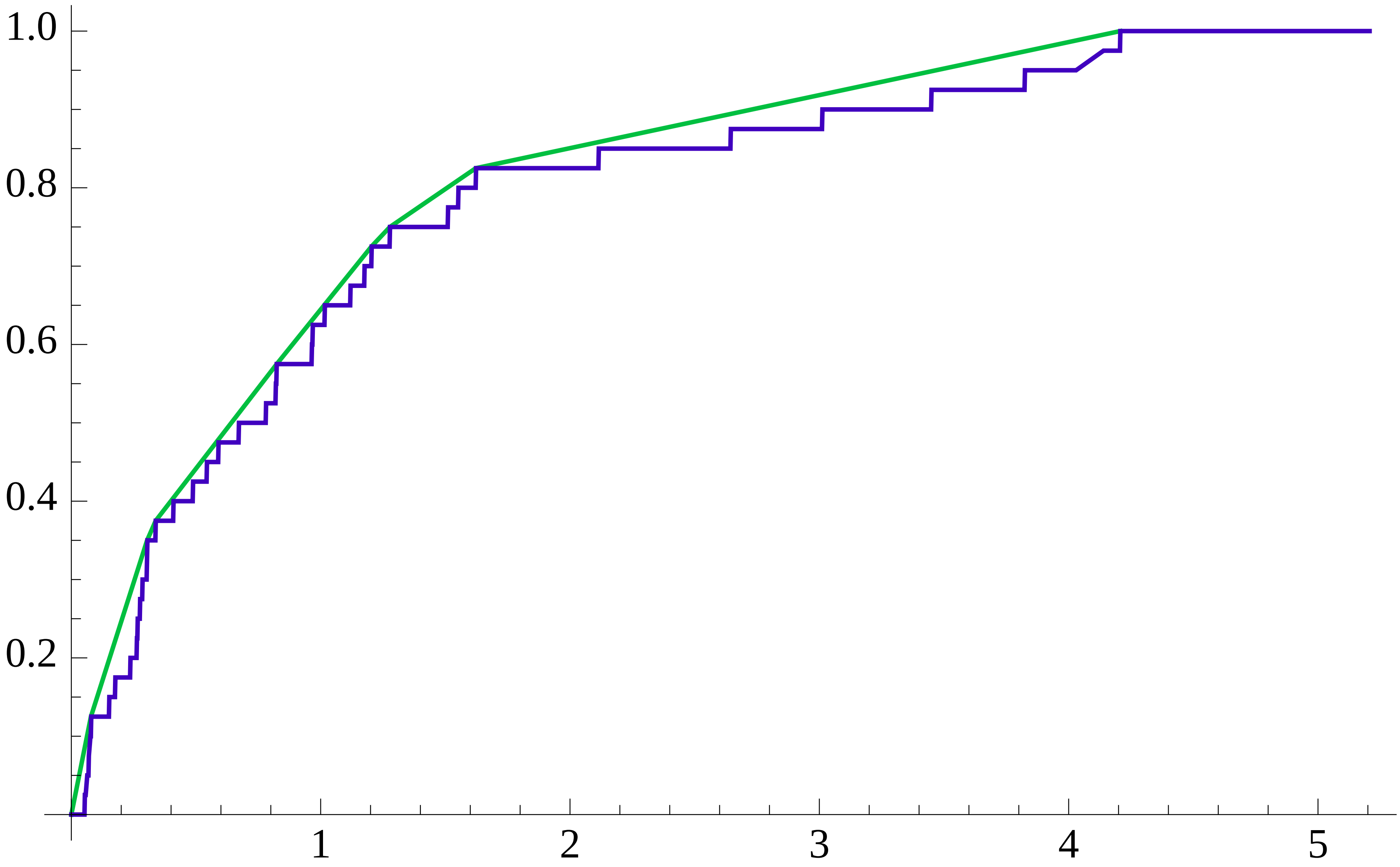
$$\mathbb{F}_n(x) \le \widehat{F}_n(x) \equiv \int_0^x \widehat{f}_n(t)dt \quad \text{for all} \quad x \ge 0, \text{ and}$$
$$\mathbb{F}_n(x) = \widehat{F}_n(x) \quad \text{if and only if } \widehat{f}_n(x-) > \widehat{f}_n(x+).$$

The equality condition in the last display can be rewritten as

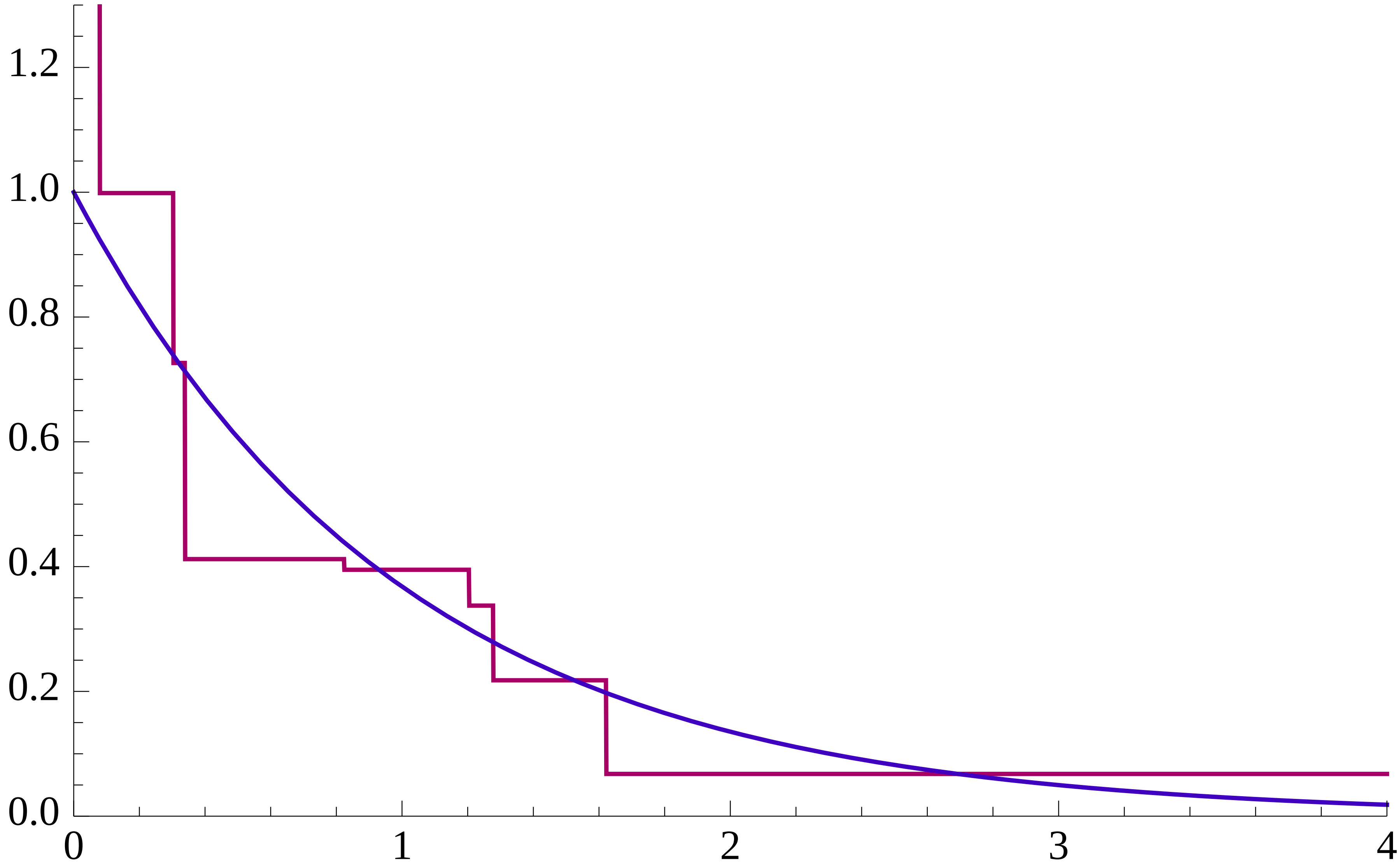
$$\int_0^\infty (\widehat{F}_n(x) - \mathbb{F}_n(x)) d\widehat{f}_n(x) = 0.$$

(c) Geometrically, $\widehat{f}_n$ is the left-derivative at x of the least concave majorant $\widehat{F}_n$ of $\mathbb{F}_n$.

A. Maximum likelihood, monotone density



A. Maximum likelihood, monotone density



A. Maximum likelihood, monotone density

Proof; Existence and Uniqueness: The log-likelihood function (divided by n) is $L_n(f) = \mathbb{P}_n \log f = n^{-1} \sum_{i=1}^n \log f(X_i)$. If we define $\breve{f}$ by $\breve{f}(x) = C \sum_{i=1}^n f(X_{(i)}) 1_{(X_{(i-1)}, X_{(i)}]}(x)$ where C is a normalizing constant to make $\int_0^\infty \breve{f}(x)dx = 1$, then

$$L_n(\breve{f}) = \log C + L_n(f) \ge L_n(f) \qquad \text{since}$$

$$1 = \int_0^\infty \breve{f}(x)dx = C \sum_{i=1}^n (X_{(i)} - X_{(i-1)}) f(X_{(i)}) \le C \int_0^{X_{(n)}} f(x)dx \le C.$$

Thus the MLE $\widehat{f}_n$ can be taken to be a histogram type estimator with breaks only at the order statistics.

Existence follows since we can restrict the maximization of L_n to the compact set

$$\mathcal{D}_M \equiv \{f \in \mathcal{D} : \ f \text{ a histogram}, \ f(0) \le M, \ f(M) = 0\}$$

for $M = \max\{1/X_{(1)}, 2X_{(n)}\}$.

A. Maximum likelihood, monotone density

Proof; Characterization: Let $\mathcal{M} = \{f : \ f(x) \ge 0$ for all $x \ge 0,\ f \searrow\}$. Then $\mathcal{D} \subset \mathcal{M}$ and $\mathcal{M}$ is a convex cone. We replace maximization of the log-likelihood

$$\mathbb{P}_n \log f = n^{-1} \sum_{i=1}^{n} \log f(X_i) = \int_0^\infty \log f(x) d\mathbb{F}_n(x)$$

over $\mathcal{D}$ by minimization of

$$\ell_n(f) \equiv -\mathbb{P}_n \log f + \int_0^\infty f(x) dx \quad \text{over} \quad \mathcal{M}.$$

Suppose $\widehat{f}_n$ minimizes $-\mathbb{P}_n \log f$ over $\mathcal{D}$. Then $\widehat{f}_n$ minimizes $\ell_n(f)$ over $\mathcal{M}$. To see this, let $g \in \mathcal{M}$ with $\int_0^\infty g(x)dx = c \in (0, \infty)$. Since $g/c \in \mathcal{D}$

$$\begin{aligned} \ell_n(g) - \ell_n(\widehat{f}_n) &= -\mathbb{P}_n \log(g/c) - \log c + c + \mathbb{P}_n \log \widehat{f}_n - 1 \\ &= \ell_n(g/c) - \ell_n(\widehat{f}_n) - \log c - 1 + c \\ &\ge 0 + 0 = 0 \end{aligned}$$

since $g/c \in \mathcal{D}$ and $c - 1 \ge \log c$. Equality holds if $g = \widehat{f}_n$. Thus $\widehat{f}_n$ maximizes ℓ_n over $\mathcal{M}$.

A. Maximum likelihood, monotone density

Now for $g \in \mathcal{M}$ and $\epsilon > 0$ consider

$$\ell_n(\widehat{f}_n + \epsilon g) \geq \ell_n(\widehat{f}_n).$$

Thus

$$\begin{aligned}
0 &\leq \lim_{\epsilon \downarrow 0} \frac{\ell_n(\widehat{f}_n + \epsilon g) - \ell_n(\widehat{f}_n)}{\epsilon} \\
&= -\int_0^\infty \frac{g}{\widehat{f}_n} d\mathbb{F}_n + \int_0^\infty g(x)dx \\
&= -\int_0^\infty \frac{1_{[0,y]}(x)}{\widehat{f}_n(x)} d\mathbb{F}_n(x) + y \quad \text{for all } y > 0 \\
&\qquad \text{by taking } \quad g(x) = 1_{[0,y]}(x) \\
&= y - \int_0^y \frac{1}{\widehat{f}_n(x)} d\mathbb{F}_n(x) \\
&= \int_0^y \frac{1}{\widehat{f}_n} d(\widehat{F}_n - \mathbb{F}_n). \qquad (1)
\end{aligned}$$

A. Maximum likelihood, monotone density

If y satisfies $\widehat{f}_n(y-) > \widehat{f}_n(y+)$, then the function $\widehat{f}_n + \epsilon 1_{[0,y]} \in \mathcal{M}$ for $\epsilon < 0$ and $|\epsilon|$ sufficiently small.

Repeating the argument for $\epsilon < 0$ and these values of y yields

$$0 = \int_0^y \frac{1}{\widehat{f}_n} d(\widehat{F}_n - \mathbb{F}_n) \quad \text{if} \quad \widehat{f}_n(y-) > \widehat{f}_n(y+). \tag{2}$$

Since $\widehat{f}_n$ is piecewise constant, the inequalities and equalities in (1) and (2) can be rewritten as claimed:

$$\mathbb{F}_n(x) \le \widehat{F}_n(x) \equiv \int_0^x \widehat{f}_n(t)dt \quad \text{for all} \quad x \ge 0, \text{ and}$$
$$\mathbb{F}_n(x) = \widehat{F}_n(x) \quad \text{if and only if } \widehat{f}_n(x-) > \widehat{f}_n(x+).$$

Now consider the LSE $\widetilde{f}_n$. Suppose that $\widetilde{f}_n$ minimizes

$$\psi_n(f) = \frac{1}{2}\int_0^\infty f^2(x)dx - \int_0^\infty f d\mathbb{F}_n$$

over $\mathcal{M}$.

A. Maximum likelihood, monotone density

Then for $g \in \mathcal{M}$ and $\epsilon > 0$ we have $\psi_n(\tilde{f}_n + \epsilon g) \geq \psi_n(\tilde{f}_n)$ and hence

$$
\begin{aligned}
0 \;&\leq\; \lim_{\epsilon \downarrow 0} \frac{\psi_n(\tilde{f}_n + \epsilon g) - \psi_n(\tilde{f}_n)}{\epsilon} \\
&= \int_0^\infty g(x)\tilde{f}_n(x)dx - \int_0^\infty g d\mathbb{F}_n = \int_0^\infty g d(\tilde{F}_n - \mathbb{F}_n) \\
&= \int_0^y d(\tilde{F}_n - \mathbb{F}_n) = \tilde{F}_n(y) - \mathbb{F}_n(y) \quad \text{for all } y > 0 \qquad (3)
\end{aligned}
$$

by choosing $g(x) = 1_{[0,y]}(x)$ for $x \geq 0$, $y > 0$. If $\tilde{f}_n(y-) > \tilde{f}_n(y+)$, then $\tilde{f}_n + \epsilon 1_{[0,y]} \in \mathcal{M}$ for $\epsilon < 0$ with $|\epsilon|$ small, so repeating the argument for $\epsilon < 0$ and these y's yields

$$\tilde{F}_n(y) - \mathbb{F}_n(y) = 0 \quad \text{if} \quad \tilde{f}_n(y-) > \tilde{f}_n(y+). \qquad (4)$$

But (3) and (4) give exactly the same characterization of $\tilde{f}_n$ derived above for $\hat{f}_n$. Thus $\tilde{f}_n = \hat{f}_n$ **in this case**.

B. Switching: a simple key result

- Groeneboom (1985), Prakasa Rao (1969)?
- Introduce first in the context of $\widehat{f}_n$
- More general version.

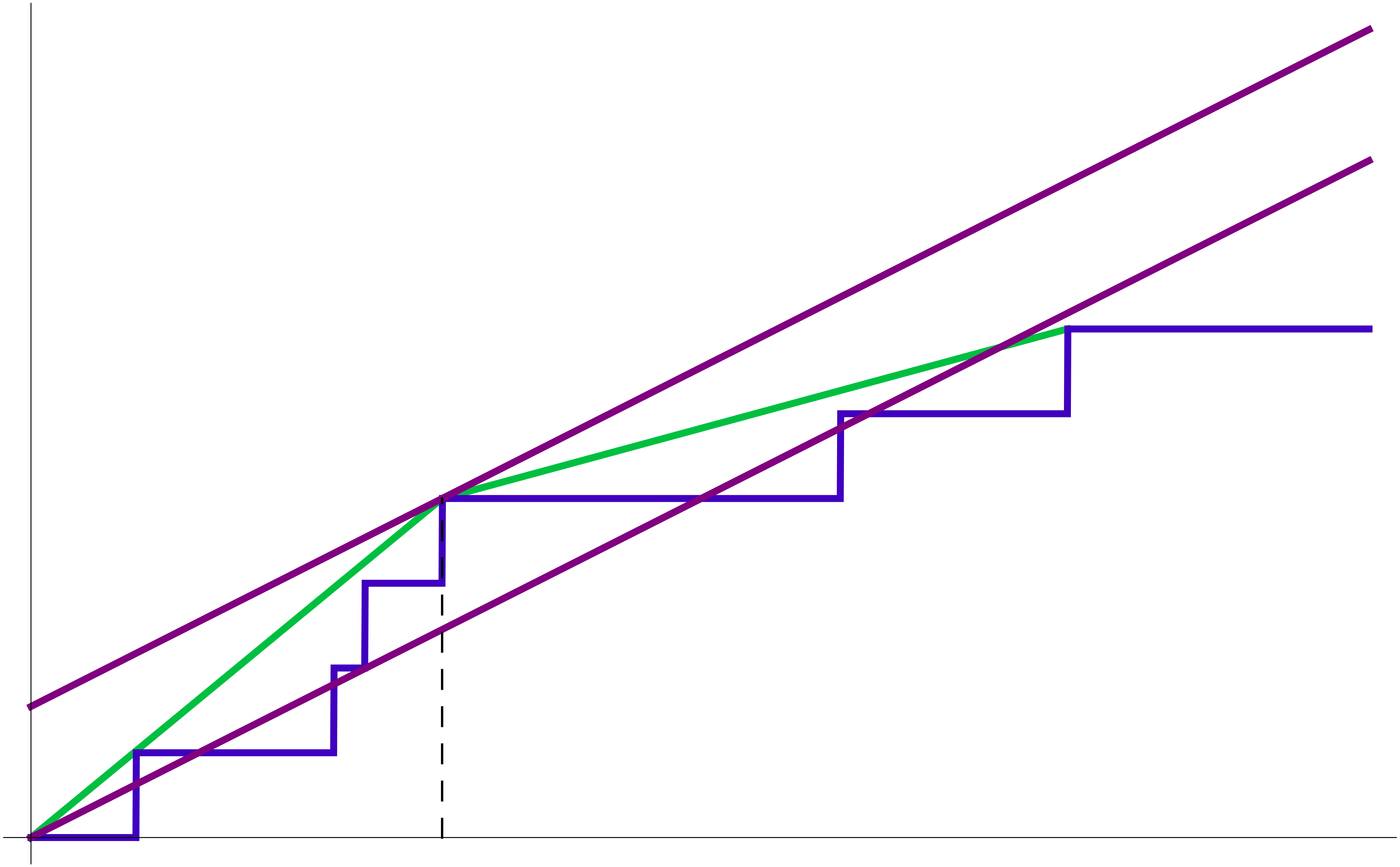
Switching for $\widehat{f}_n$: Define

$$\begin{aligned} \widehat{s}_n(a) &\equiv \text{argmax}_{s\ge 0}\{\mathbb{F}_n(s) - as\}, \quad a > 0 \\ &\equiv \sup\{s \ge 0 : \ \mathbb{F}_n(s) - as = \sup_{z\ge 0}(\mathbb{F}_n(z) - az)\}. \end{aligned}$$

Then for each fixed $t \in (0, \infty)$ and $a > 0$

$$\left\{\widehat{f}_n(t) < a\right\} = \{\widehat{s}_n(a) < t\}\,.$$

B. Switching: a simple key result



B. Switching: a simple key result

More general result: Suppose $\Phi : D \subset \mathbb{R} \to \mathbb{R}$ where D is closed. Let

$$\begin{aligned} \widehat{\Phi}(x) &\equiv \text{least concave majorant of } \Phi \\ &= \inf\{g(x)\Big| g : D \to \mathbb{R},\ g \text{ closed},\ g \text{ concave}, g \geq \Phi\}. \end{aligned}$$

Let $\widehat{\phi}_L$ and $\widehat{\phi}_R$ denote the left and right derivatives of $\widehat{\Phi}$.

Define

$$\begin{aligned} \kappa_L(y) &\equiv \text{argmax}_x^L\{\Phi(x) - yx\} \\ &= \inf\{x \in D : \ \Phi(x) - yx = \sup_{z \in D}(\Phi(z) - yz)\}, \\ \kappa_R(y) &\equiv \text{argmax}_x^R\{\Phi(x) - yx\} \\ &= \sup\{x \in D : \ \Phi(x) - yx = \sup_{z \in D}(\Phi(z) - yz)\}. \end{aligned}$$

B. Switching: a simple key result

Theorem. Suppose that Φ is a proper upper-semicontinuous real-valued function defined on a closed subset $D \subset \mathbb{R}$. Then $\widehat{\Phi}$ is proper if and only if $\Phi \le l$ for some linear function l on D. Furthermore, if conv(hypo(Φ)) is closed, then the functions κ_L and κ_R are well defined and the following switching relations hold:

$$\begin{aligned} \widehat{\phi}_L(x) < y \quad &\text{if and only if} \quad \kappa_R(y) < x; \\ \widehat{\phi}_R(x) \le y \quad &\text{if and only if} \quad \kappa_L(y) \le x. \end{aligned}$$

Proof. See Balabdaoui, Jankowski, Pavlides, Seregin, and W (2010) – which is based on Rockafellar (1970).

We will apply this theorem with Φ taken to be various random processes, including:

- $\Phi = \mathbb{U}$, a Brownian bridge process on $[0, 1]$.
- $\Phi = aW(h) - bh^2$ for $a, b > 0$ and W two-sided Brownian motion.

B. Switching: a simple key result

Reminder:

$$\text{hypo}(f) = \{(x, \alpha) \in \mathbb{R}^d \times R : \ \alpha \leq f(x)\},$$
$$\text{conv}(C) = \{\sum_{i=1}^{k} \lambda_i x_i : \ x_i \in C, \ \lambda_i \geq 0, \sum_{1}^{k} \lambda_i = 1, \ k \geq 0\}.$$

f is upper semicontinuous at all $x \in \mathbb{R}^d$ if and only if hypo(f) is closed.

C. Limit theory via switching and argmax CM

Two illustrative cases:

- Case 1: $f_0(x) = 1_{[0,1]}(x)$ (degenerate mixing, $G = \delta_1$).
- Case 2: f_0 with $f_0(x_0) > 0$, $f_0'(x_0) < 0$. (Strictly decreasing at x_0).

Case 1: Groeneboom (1983), Groeneboom and Pyke (1983). If $f_0(x) = 1_{[0,1]}(x)$, then for $0 < x_0 < 1$,

$$\mathbb{S}_n(x_0) \equiv \sqrt{n}(\widehat{f}_n(x_0) - f_0(x_0)) \to_d \mathbb{S}(x_0)$$

where $\mathbb{S}$ is the left-derivative of the least concave majorant $\mathbb{C}$ of a standard Brownian bridge process $\mathbb{U}$ on $[0, 1]$.

C. Limit theory via switching and argmax CM

Proof, Case 1: By the switching relation

$$
\begin{aligned}
&P(\sqrt{n}(\widehat{f}_n(x_0) - f_0(x_0)) < t) \\
&\quad = P(\widehat{f}_n(x_0) < f_0(x_0) + n^{-1/2}t) \\
&\quad = P(\hat{s}_n(f_0(x_0) + n^{-1/2}t) < x_0) \\
&\quad = P(\mathrm{argmax}_h\{\mathbb{F}_n(x_0+h) - (f_0(x_0) + n^{-1/2}t)(x_0+h)\} < 0) \\
&\quad = P(\mathrm{argmax}_h \mathbb{Z}_n(h) < 0) \qquad (5)
\end{aligned}
$$

where, since $f_0(x_0) = 1$ implies that $xf_0(x_0) = x_0 = F(x_0)$,

$$
\begin{aligned}
\mathbb{Z}_n(h) &\equiv n^{1/2}(\mathbb{F}_n(x_0+h) - F(x_0) - hf_0(x_0) - t(x_0+h)n^{-1/2}) \\
&= n^{1/2}(\mathbb{F}_n(x_0+h) - F(x_0+h)) \\
&\qquad + n^{1/2}(F(x_0+h) - F(x_0) - hf_0(x_0)) - t(x_0+h) \\
&= \mathbb{U}_n(x_0+h) - t(x_0+h) \\
&\rightsquigarrow \mathbb{U}(x_0+h) - t(x_0+h)
\end{aligned}
$$

where $\mathbb{U}_n \equiv \sqrt{n}(\mathbb{F}_n - F)$ denotes the uniform empirical process and $\mathbb{U}$ denotes a Brownian bridge process.

C. Limit theory via switching and argmax CM

Thus by the (argmax) continuous mapping theorem it follows that the right side of (5) converges to

$$
\begin{aligned}
& P(\mathrm{argmax}_h\{\mathbb{U}(x_0+h) - t(X_0+h)\} < 0) \\
& \quad = P(\mathrm{argmax}_s\{\mathbb{U}(s) - ts\} < x_0) \\
& \quad = P(\mathbb{S}(x_0) < t)
\end{aligned}
$$

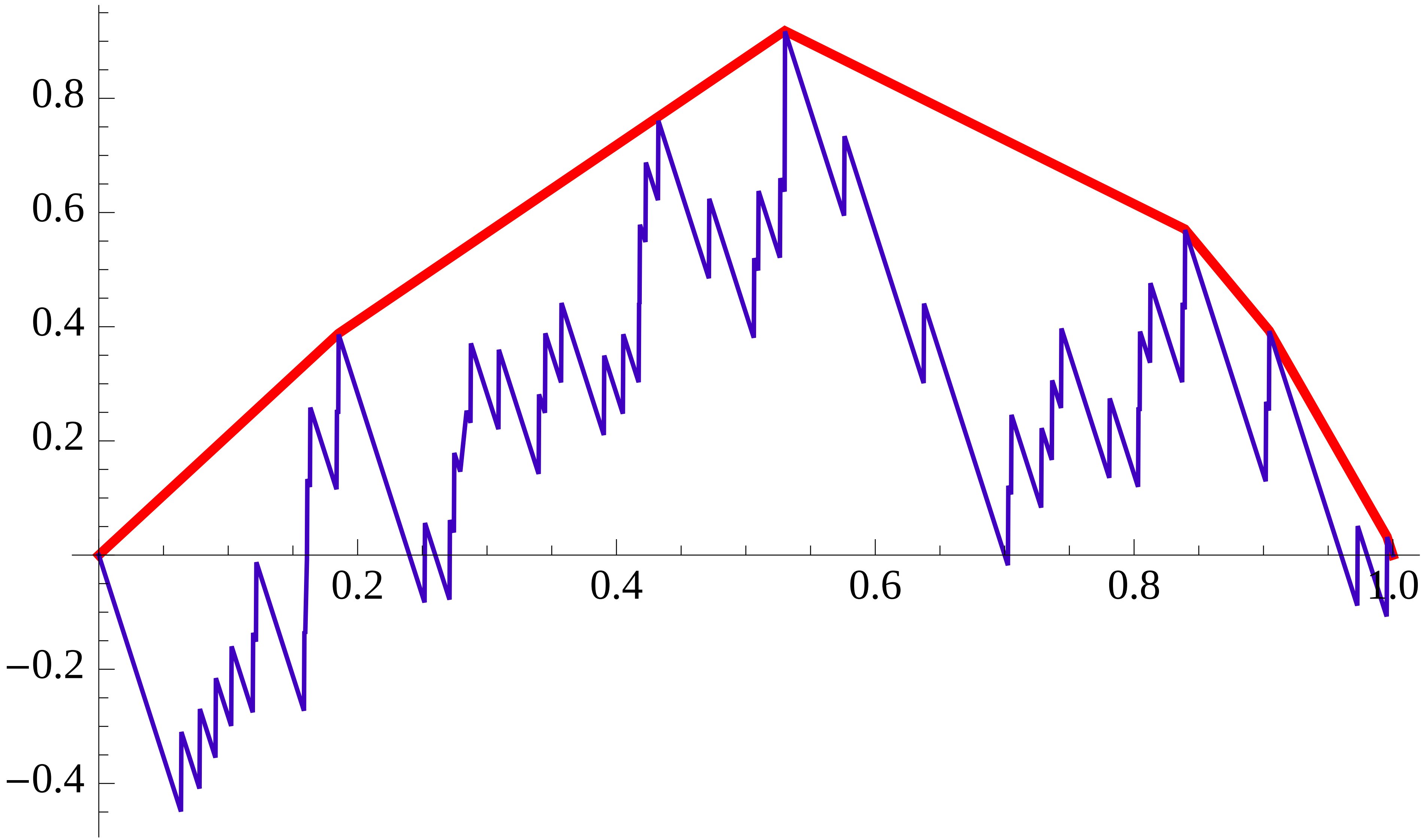
by the general version of the switching relation. Hence

$$\sqrt{n}(\widehat{f}_n(x_0) - f_0(x_0)) \to_d \mathbb{S}(x_0). \qquad \square$$

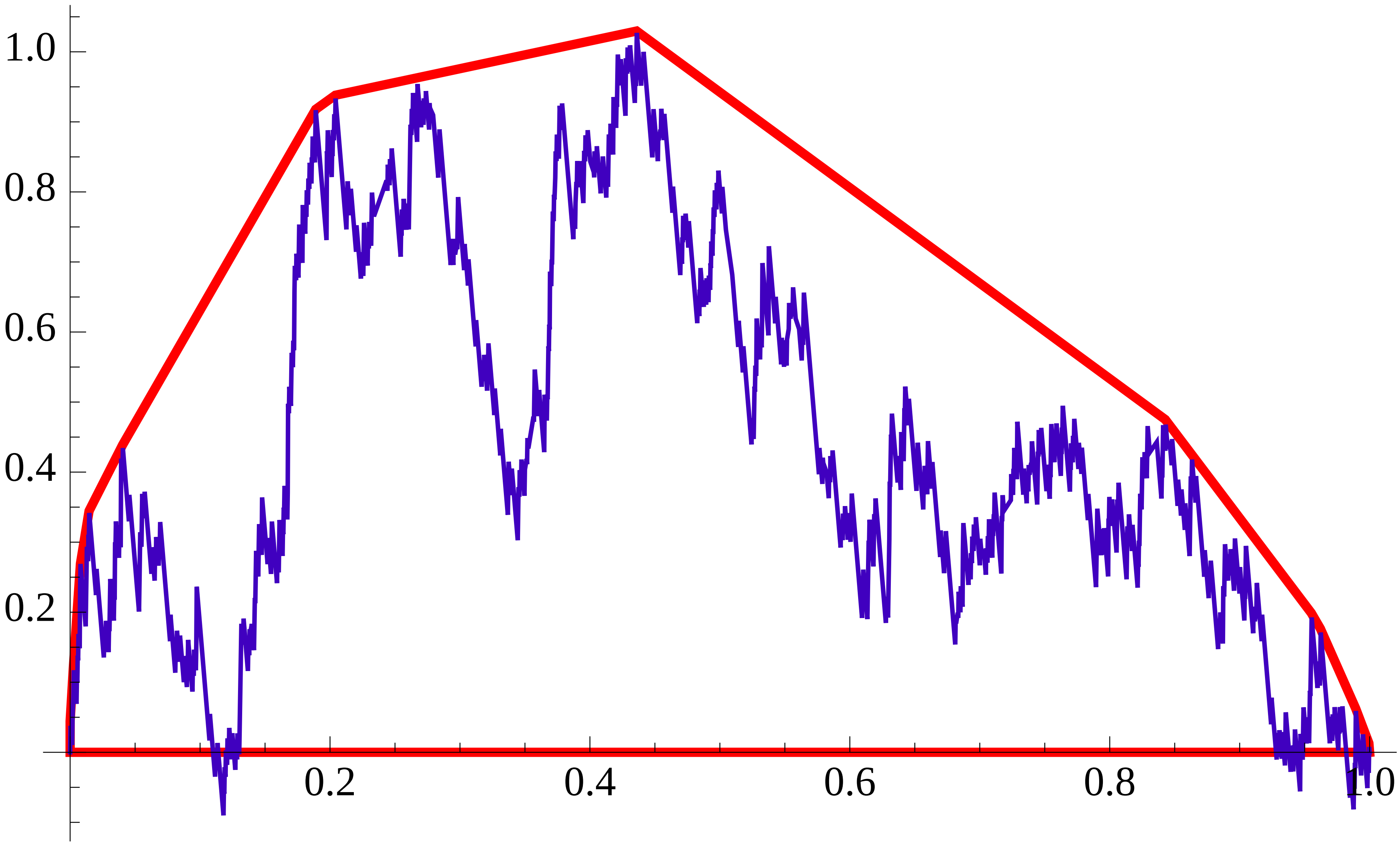
This one-dimensional convergence extends straightforwardly to convergence of the finite-dimensional distributions, and (by monotonicity) to convergence in the Skorokhod topology on $D[a, 1-a]$ for each fixed $a \in (0, 1/2)$.

Exercise 1. $\mathbb{S}_n \rightsquigarrow \mathbb{S}$ in $L_1([0,1],\lambda)$ with $\lambda =$Lebesgue measure; this also holds in $L_p([0,1],\lambda)$ for $1 \le p < 2$, but not in $L_2([0,1],\lambda)$.

C. Limit theory via switching and argmax CM



C. Limit theory via switching and argmax CM



C. Limit theory via switching and argmax CM

Case 2: Prakasa Rao (1969), Groeneboom (1985). If $f_0(x_0) > 0$, $f_0'(x_0) < 0$, and f_0' is continuous at x_0, then

$$\begin{aligned} \mathbb{S}_n(x_0, t) &\equiv n^{1/3}(\widehat{f}_n(x_0 + n^{-1/3} c_0 t) - f_0(x_0)) \\ &\to_d (2^{-1} f_0(x_0)|f_0'(x_0)|)^{1/3} \mathbb{S}(t) \end{aligned}$$

where $\mathbb{S}$ is the left-derivative of the least concave majorant $\mathbb{C}$ of $W(t) - t^2$, W is a standard two-sided Brownian motion process starting at 0, and $c_0 \equiv 4 f_0(x_0)/(f_0'(x_0))^2)^{1/3}$. In particular:

$$\mathbb{S}_n(x_0) \equiv n^{1/3}(\widehat{f}_n(x_0) - f_0(x_0)) \to_d (2^{-1} f_0(x_0)|f_0'(x_0)|)^{1/3} \mathbb{S}(0).$$

Proof, Case 2: By the switching relation

C. Limit theory via switching and argmax CM

$$
\begin{aligned}
&P(n^{1/3}(\hat{f}_n(x_0+n^{-1/3}t)-f(x_0))<y)\\
&\quad= P(\hat{f}_n(x_0+n^{-1/3}t)<f(x_0)+yn^{-1/3}),\\
&\quad= P(\hat{s}_n(f(x_0)+yn^{-1/3})<x_0+n^{-1/3}t)\\
&\quad= P(\text{argmax}_v\{\mathbb{F}_n(v)-(f(x_0)+n^{-1/3}y)v\}<x_0+n^{-1/3}t)
\end{aligned}
$$

Now we change variables $v=x_0+n^{-1/3}h$ in the argument of $\mathbb{F}_n$ and center and scale to find that the right side in the last display equals

$$
\begin{aligned}
&P(\text{argmax}_h\{\mathbb{F}_n(x_0+n^{-1/3}h)-(f(x_0)+n^{-1/3}y)(x_0+n^{-1/3}h)\}<t)\\
&= P\Big(\text{argmax}_h\{\mathbb{F}_n(x_0+n^{-1/3}h)-\mathbb{F}_n(x_0)-(F(x_0+n^{-1/3}h)-F(x_0))\\
&\qquad + F(x_0+n^{-1/3}h)-F(x_0)-f(x_0)n^{-1/3}h-n^{-2/3}yh\}<t\Big).
\end{aligned}
\tag{6}
$$

Now the stochastic term in (6) satisfies

C. Limit theory via switching and argmax CM

$$
\begin{aligned}
& n^{2/3}\left\{\mathbb{F}_n(x_0+n^{-1/3}h)-\mathbb{F}_n(x_0)-(F(x_0+n^{-1/3}h)-F(x_0))\right\} \\
& \stackrel{d}{=} \; n^{2/3-1/2}\left\{\mathbb{U}_n(F(x_0+n^{-1/3}h))-\mathbb{U}_n(F(x_0))\right\} \\
& = \; n^{1/(2\cdot 3}\left\{\mathbb{U}(F(x_0+n^{-1/3}h))-\mathbb{U}(F(x_0))\right\}+o_p(1) \qquad \text{by KMT} \\
& \qquad \text{or by Theorems 2.11.22 or 2.11.23} \\
& \stackrel{d}{=} \; n^{1/6}W(f(x_0)n^{-1/3}h)+o_p(1) \\
& \stackrel{d}{=} \; \sqrt{f(x_0)}W(h)+o_p(1)
\end{aligned}
$$

where W is a standard two-sided Brownian motion process starting from 0. On the other hand, with $\delta_n \equiv n^{-1/3}$,

$$
\begin{aligned}
& n^{2/3}\left(F(x_0+n^{-1/3})-F(x_0)-f(x_0)n^{-1/3}h\right) \\
& = \; \delta_n^{-2}(F(x_0+\delta_n h)-F(x_0)-f(x_0)\delta_n h) \\
& \to \; -b|h|^2 \qquad \text{with} \;\; b=|f'(x_0)|/2
\end{aligned}
$$

by our hypotheses, while $n^{2/3}n^{-1/3}n^{-1/3}h=n^0h=h$.

C. Limit theory via switching and argmax CM

Thus it follows that the last probability above converges to

$$P\left(\text{argmax}_h \left\{\sqrt{f(x_0)}W(h) - b|h|^2 - yh\right\} < t\right)$$
$$= P(\mathbb{S}_{a,b}(t) < y) \qquad \text{by switching again}$$

where

$$\begin{aligned} \mathbb{S}_{a,b}(t) &= \text{slope at } t \text{ of the least concave majorant of} \\ &\qquad aW(h) - bh^2 \equiv \sqrt{f_0(x_0)}W(h) - |f_0'(x_0)||h|^2/2 \\ &\stackrel{d}{=} |2^{-1} f_0(x_0) f_0'(x_0)| \mathbb{S}(t/c_0). \end{aligned}$$

Exercise 2. Prove the equality in distribution in the last display.

C. Limit theory via switching and argmax CM

Exercise 3. Let

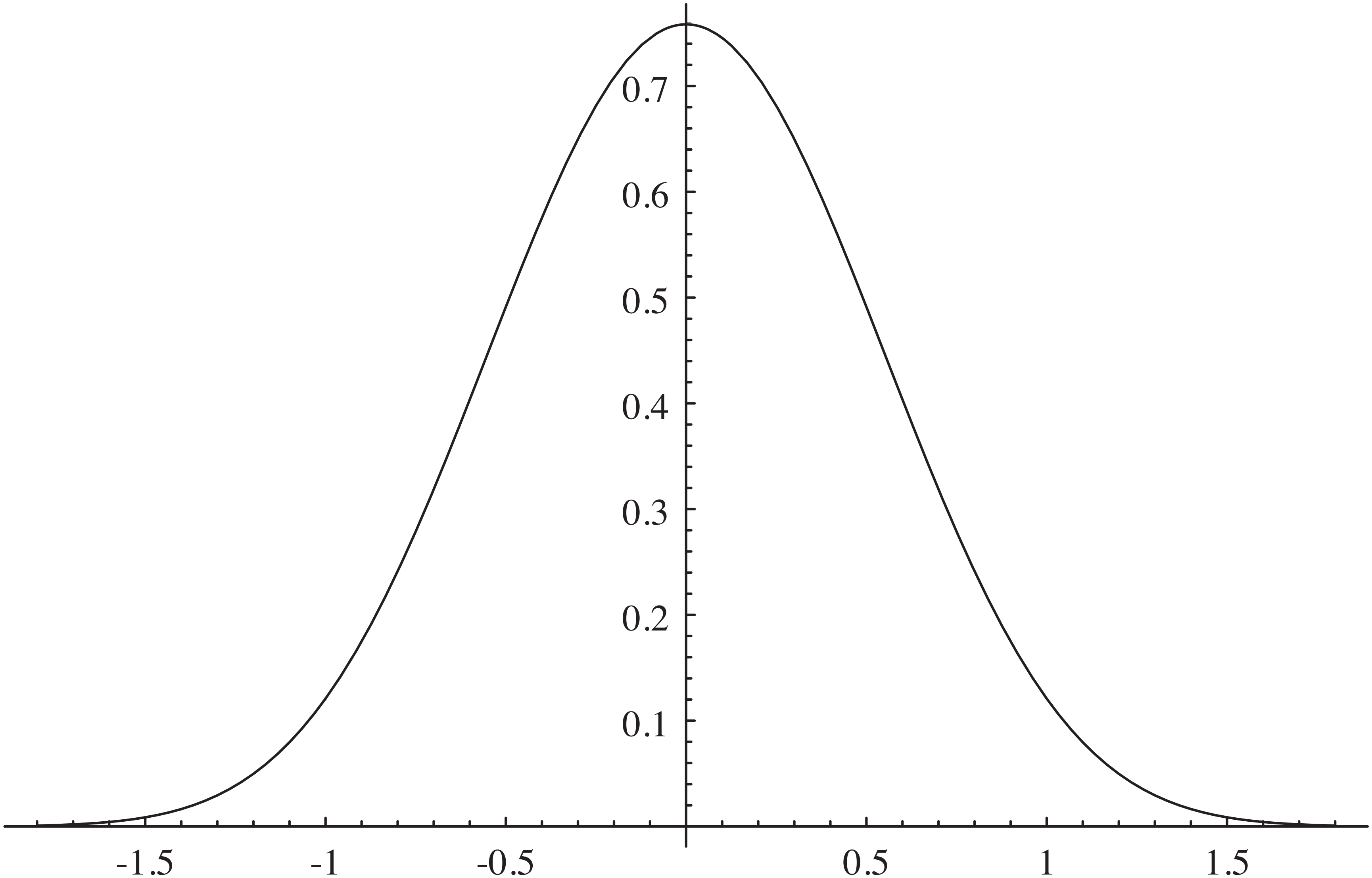
$$\mathbb{S}_n(x_0,t) \equiv n^{1/3}(\widehat{f}_n(x_0+n^{-1/3}t) - f(x_0)).$$

Show that with $y_0 \neq x_0$ and the hypotheses of Case 2 satisfied at both x_0 and y_0, we have

$$\left(\begin{array}{c} \mathbb{S}_n(x_0,\cdot) \\ \mathbb{S}_n(y_0,\cdot) \end{array}\right) \rightsquigarrow \left(\begin{array}{c} \mathbb{S}_{a,b} \\ \widetilde{\mathbb{S}}_{\tilde{a},\tilde{b}} \end{array}\right) \quad \text{in} \quad D[-M,M]^2$$

for every $M > 0$ where $a = \sqrt{f(x_0)}$, $\tilde{a} = \sqrt{f(y_0)}$, $b = |f'(x_0)|/2$, $\tilde{b} = |f'(y_0)|/2$, and $\mathbb{S}_{a,b}$, $\widetilde{\mathbb{S}}_{\tilde{a},\tilde{b}}$ are the left-derivatives of the least concave majorant of $aW(h) - bh^2$ and $\tilde{a}\widetilde{W} - \tilde{b}h^2$ and where W and $\widetilde{W}$ are independent two-sided Brownian motion processes.

C. Limit theory via switching and argmax CM



E. Other monotone function problems

- Monotone hazard (rate) function
- Regression function
- Distribution function for interval censoring model
- Cumulative mean function, panel count data
- Sub-distribution functions, competing risks with interval censored data

Monotone hazard function:

- Model: $\mathcal{H} \equiv$ all monotone increasing (or decreasing) hazard rates (wrt Lebesgue measure) on $\mathbb{R}^+ = (0, \infty)$.

$$h(t) = \frac{f(t)}{1 - F(t)}; \quad f(t) = h(t)\exp\left(-\int_0^t h(s)ds\right) \equiv h(t)\exp\left(-H(t)\right)$$

- Observations: $X_1, \ldots, X_n$ i.i.d. f_0 with $h_0 \in \mathcal{H}$.
- MLE: $\widehat{f}_n \equiv \text{argmax}_{h \in \mathcal{H}} \left\{\sum_{i=1}^n \{\log h(X_i) - H(X_i)\}\right\}$

E. Other monotone function problems

Monotone regression:

- Model: $Y = r(x) + \epsilon$ where

  $r \in \mathcal{M} \equiv \{$all monotone (increasing) functions from D to $\mathbb{R}\}$

  $E(\epsilon) = 0$, $Var(\epsilon) < \infty$.

- Observations: $\{(x_{n,i}, Y_{n,i}) : i = 1, \ldots, n\}$ where $Y_{n,i} = r_0(x_{n,i}) + \epsilon_{n,i}$ for some $r_0 \in \mathcal{M}$ and $x_{n,1} \le \ldots \le x_{n,n}$.

- LSE (=MLE for Gaussian ϵ's):

$$\widehat{r}_n \equiv \mathrm{argmin}_{r \in \mathcal{M}_n} \frac{1}{2} \sum_{i=1}^{n} (Y_{n,i} - r(x_{n,i}))^2$$

  where $\mathcal{M}_n \subset \mathcal{M}$ is the subclass of monotone functions which are linear between successive $x_{n,i}$'s and the left and right of the range of the $x_{n,i}$'s.

E. Other monotone function problems

Interval censoring case 1 = Current status data:

- Model: $X \sim F$ on $\mathbb{R}^+$, $Y \sim G$ on $\mathbb{R}^+$ independent, $F \in \mathcal{F} \equiv \{\text{all distribution functions on } \mathbb{R}^+\}$.
  Observe $(Y, \Delta) \equiv (Y, 1_{[X \le Y]})$, so that
  $$(\Delta | Y) \sim \text{Bernoulli}(F(Y)).$$
  Thus the density of (Y, Δ) with respect to $G \times$ counting measure on$\{0, 1\}$ is
  $$p(y, \delta; F) = F(y)^{\delta}(1 - F(y))^{1-\delta}.$$
- Observations: $\{(Y_i, \Delta_i) : \ i = 1, \ldots, n\}$ i.i.d. as (Y, Δ).
- MLE:
  $$\widehat{F}_n = \text{argmax}_{F \in \mathcal{F}} \left\{ \mathbb{P}_n (\Delta \log F + (1 - \Delta) \log(1 - F) \right\}.$$

E. Other monotone function problems

Panel count data:

See Zhang and W (2000), (2007)

Competing risks data with current status observations:

See Groeneboom, Maathuis and W (2008a, 2008b)

F. Other properties of $\widehat{f}_n$

- (a) $\widehat{f}_n$ is not consistent at zero; general limit behavior at zero.
- (b) connections to unimodal density estimators
- (c) L_1 metric behavior: Groeneboom (1985), GHL (1999)
- (d) global upper bounds,
  L_1 & Hellinger: Birgé/Groeneboom/van de Geer
- (e) linear functionals
- (f) Marshall's lemma and Kiefer - Wolfowitz theory

Outline: (tomorrow)

- L1: Monotone functions: maximum likelihood and least squares
- **L2: Optimality of the MLE of a monotone density**
- L3: Estimation of convex and $k-$monotone density functions
- L4: Estimation of log-concave densities: $d = 1$ and beyond
- L5: More on higher dimensions and some open problems