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 2

- A: Local asymptotic minimax lower bounds
- B: Lower bounds for estimation of a monotone density
  Several scenarios
- C: Global lower bounds and upper bounds (briefly)
- D: Lower bounds for estimation of a convex density
- E: Lower bounds for estimation of a log-concave density

A. Local asymptotic minimax lower bounds

Proposition. (Two-point lower bound) Let $\mathcal{P}$ be a set of probability measures on a measurable space $(\mathbb{X}, \mathcal{A})$, and let ν be a real-valued function defined on $\mathcal{P}$. Moreover, let $l : [0, \infty) \to [0, \infty)$ be an increasing convex loss function with $l(0) = 0$. Then, for any $P_1, P_2 \in \mathcal{P}$ such that $H(P_1, P_2) < 1$ and with

$$\begin{aligned} E_{n,i} f(X_1, \ldots, X_n) = E_{n,i} f(X) &= \int f(x) dP_i^n(x) \\ &\equiv \int f(x_1, \ldots, x_n) dP_i(x_1) \cdots dP_i(x_n)\,, \end{aligned}$$

for $i = 1, 2$, it follows that

$$\begin{aligned} &\inf_{T_n} \max \left\{ E_{n,1} l(|T_n - \nu(P_1)|),\, E_{n,2} l(|T_n - \nu(P_2)|) \right\} \qquad (1) \\ &\geq\ l\left(\frac{1}{4} |\nu(P_1) - \nu(P_2)| \{1 - H^2(P_1, P_2)\}^{2n} \right). \end{aligned}$$

A. Local asymptotic minimax lower bounds

Proof. By Jensen's inequality

$$E_{n,i} l(|T_n - \nu(P_i)|) \ge l(E_{n,i}|T_n - \nu(P_i)|)\,, \qquad i = 1, 2\,,$$

and hence the left side of (1) is bounded below by

$$l\left(\inf_{T_n} \max\{E_{n,1}|T_n - \nu(P_1)|\,,\ E_{n,2}|T_n - \nu(P_2)|\right)\,.$$

Thus it suffices to prove the proposition for $l(x) = x$. Let $p_1 \equiv dP_1/(d(P_1 + P_2)$, $p_2 = dP_2/d(P_1 + P_2)$, and $\mu = P_1 + P_2$ (or let p_i be the density of P_i with respect to some other convenient dominating measure μ, $i = 1, 2$).

A. Local asymptotic minimax lower bounds

Two Facts:

Fact 1: Suppose P, Q abs. cont. wrt μ,

$$H^2(P,Q) \equiv 2^{-1} \int \{\sqrt{p} - \sqrt{q}\}^2 d\mu = 1 - \int \sqrt{pq} d\mu \equiv 1 - \rho(P,Q).$$

Then

$$(1 - H^2(P,Q))^2 \le 1 - \left\{1 - \int (p \wedge q)\, d\mu\right\}^2 \le 2 \int (p \wedge q)\, d\mu\,.$$

Fact 2: If P and Q are two probability measures on a measurable space $(\mathbb{X}, \mathcal{A})$ and P^n and Q^n denote the corresponding product measures on $(\mathbb{X}^n, \mathcal{A}_n)$ (of $X_1, \ldots, X_n$ i.i.d. as P or Q respectively), then $\rho(P,Q) \equiv \int \sqrt{pq} d\mu$ satisfies

$$\rho(P^n, Q^n) = \rho(P,Q)^n\,. \tag{2}$$

Exercise. Prove Fact 1.
Exercise. Prove Fact 2.

A. Local asymptotic minimax lower bounds

$$
\begin{aligned}
&\max\left\{E_{n,1}|T_n - \nu(P_1)|,\ E_{n,2}|T_n - \nu(P_2)|\right\} \\
&\quad \geq \frac{1}{2}\left\{E_{n,1}|T_n - \nu(P_1)| + E_{n,2}|T_n - \nu(P_2)|\right\} \\
&\quad = \frac{1}{2}\left\{\int |T_n(x) - \nu(P_1)| \prod_{i=1}^{n} p_1(x_i) d\mu(x_1)\cdots d\mu(x_n)\right. \\
&\qquad\qquad \left. + \int |T_n(x) - \nu(P_2)| \prod_{i=1}^{n} p_2(x_i) d\mu(x_1)\cdots d\mu(x_n)\right\} \\
&\quad \geq \frac{1}{2}\left\{\int [|T_n(x) - \nu(P_1)| + |T_n(x) - \nu(P_2)|] \prod_{i=1}^{n} p_1(x_i) \wedge \prod_{i=1}^{n} p_2(x_i) d\mu(x_1)\right. \\
&\quad \geq \frac{1}{2}|\nu(P_1) - \nu(P_2)| \int \prod_{i=1}^{n} p_1(x_i) \wedge \prod_{i=1}^{n} p_2(x_i) d\mu(x_1)\cdots d\mu(x_n) \\
&\quad \geq \frac{1}{4}|\nu(P_1) - \nu(P_2)|\{1 - H^2(P_1^n, P_2^n)\}^2 \qquad \text{by Fact 1} \\
&\quad = \frac{1}{4}|\nu(P_1) - \nu(P_2)|\{1 - H^2(P_1, P_2)\}^{2n} \qquad \text{by Fact 2}.
\end{aligned}
$$

B. Lower bounds, monotone density

Several scenarios, estimation of $f(x_0)$:

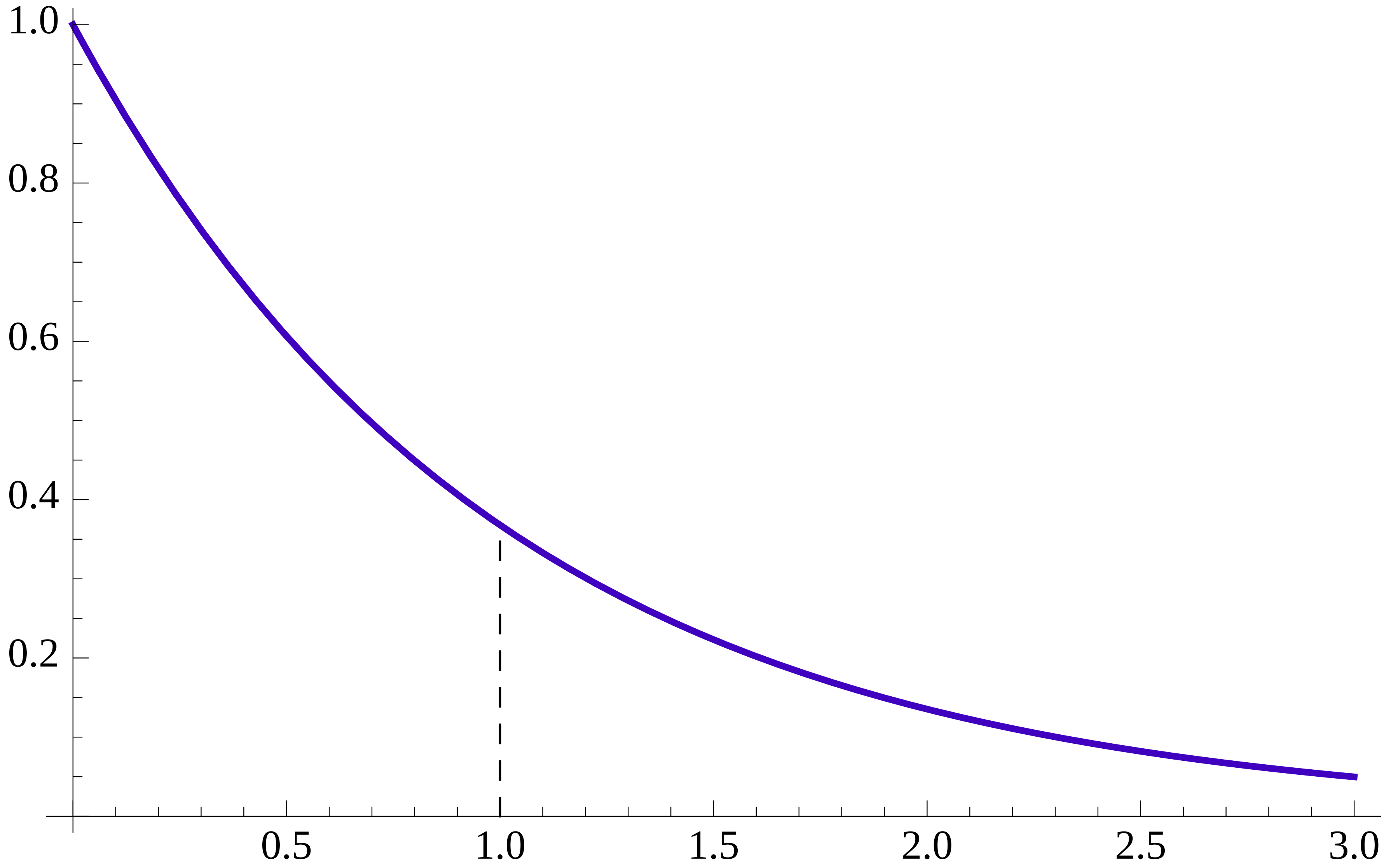
S1 When $f(x_0) > 0$, $f'(x_0) < 0$.

S2 When $x_0 \in (a, b)$ with $f(x)$ constant on (a, b).
In particular, $f(x) = 1_{[0,1]}(x)$, $x_0 \in (0, 1)$.

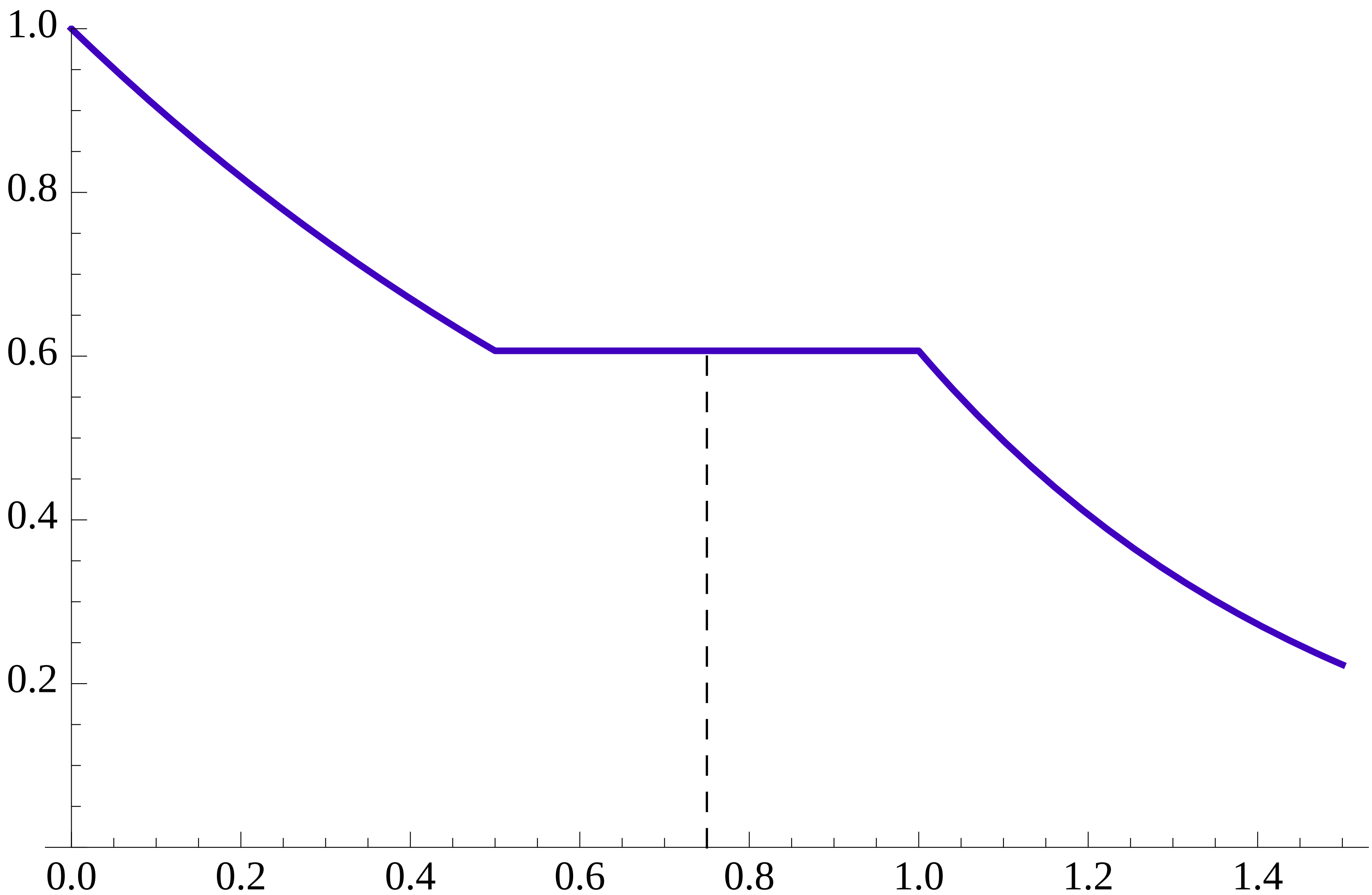
S3 When f is discontinuous at x_0.

S4 When $f^{(j)}(x_0) = 0$ for $j = 1, \ldots, k - 1$, $f^{(k)}(x_0) \neq 0$.

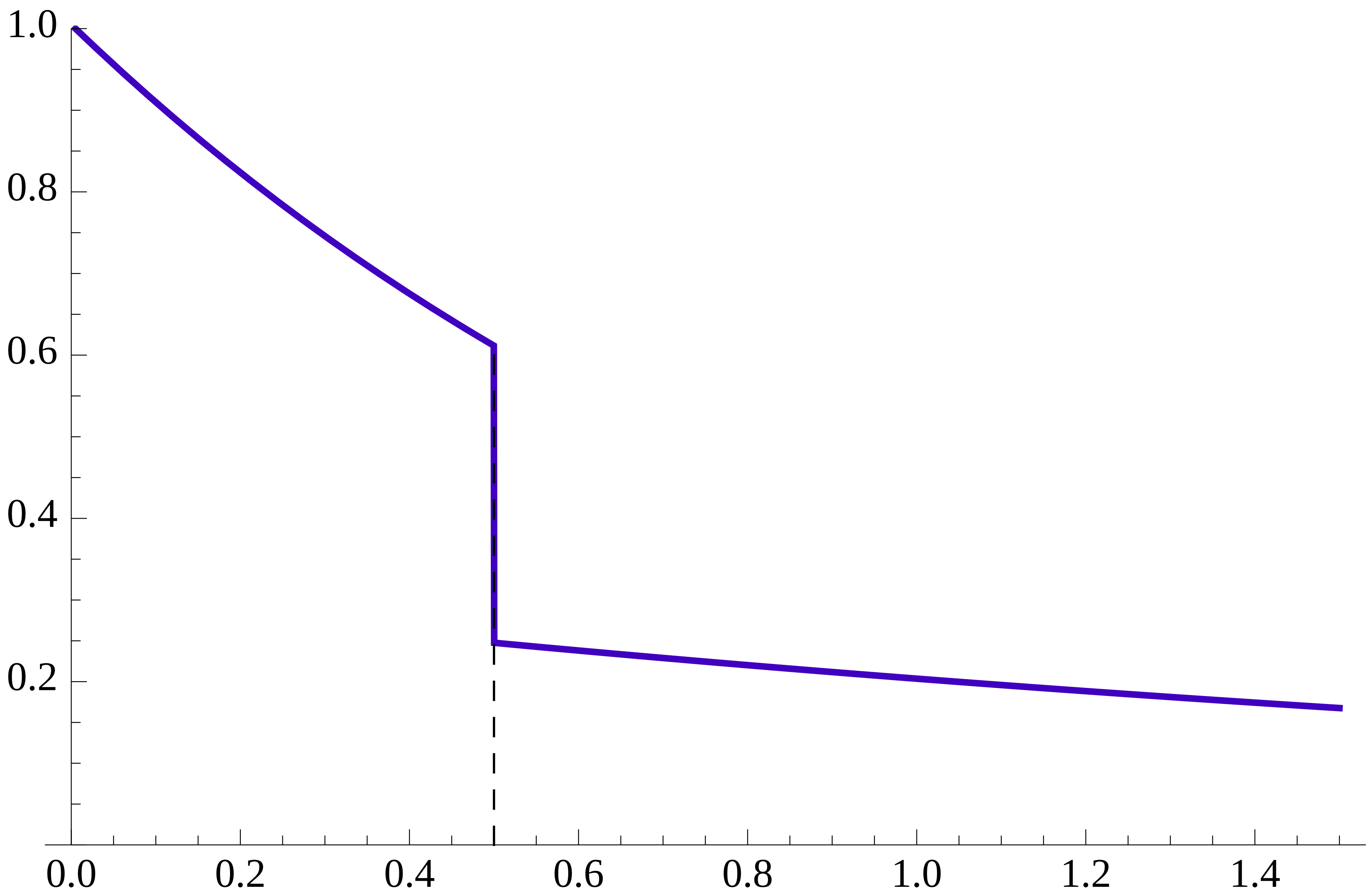
B. Lower bounds, monotone density



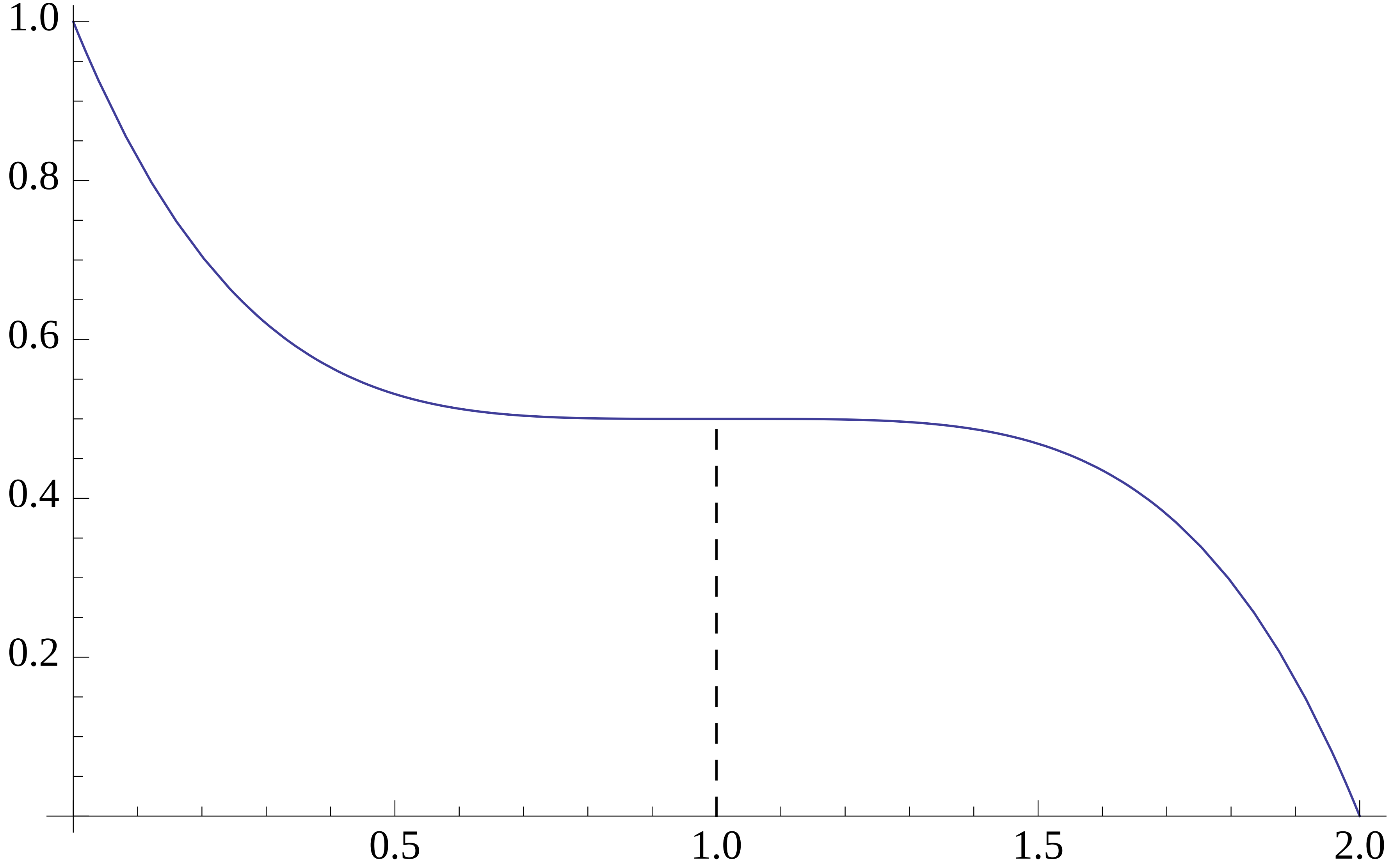
B. Lower bounds, monotone density



B. Lower bounds, monotone density



B. Lower bounds, monotone density



B. Lower bounds, monotone density

S1: $f_0(x_0) > 0$, $f_0'(x_0) < 0$. Suppose that we want to estimate $\nu(f) = f(x_0)$ for a fixed Let f_0 be the density corresponding to P_0, and suppose that $f_0'(x_0) < 0$. To apply our two-point lower bound Proposition we need to construct a sequence of densities f_n that are "near" f_0 in the sense that

$$nH^2(f_n, f_0) \to \ A$$

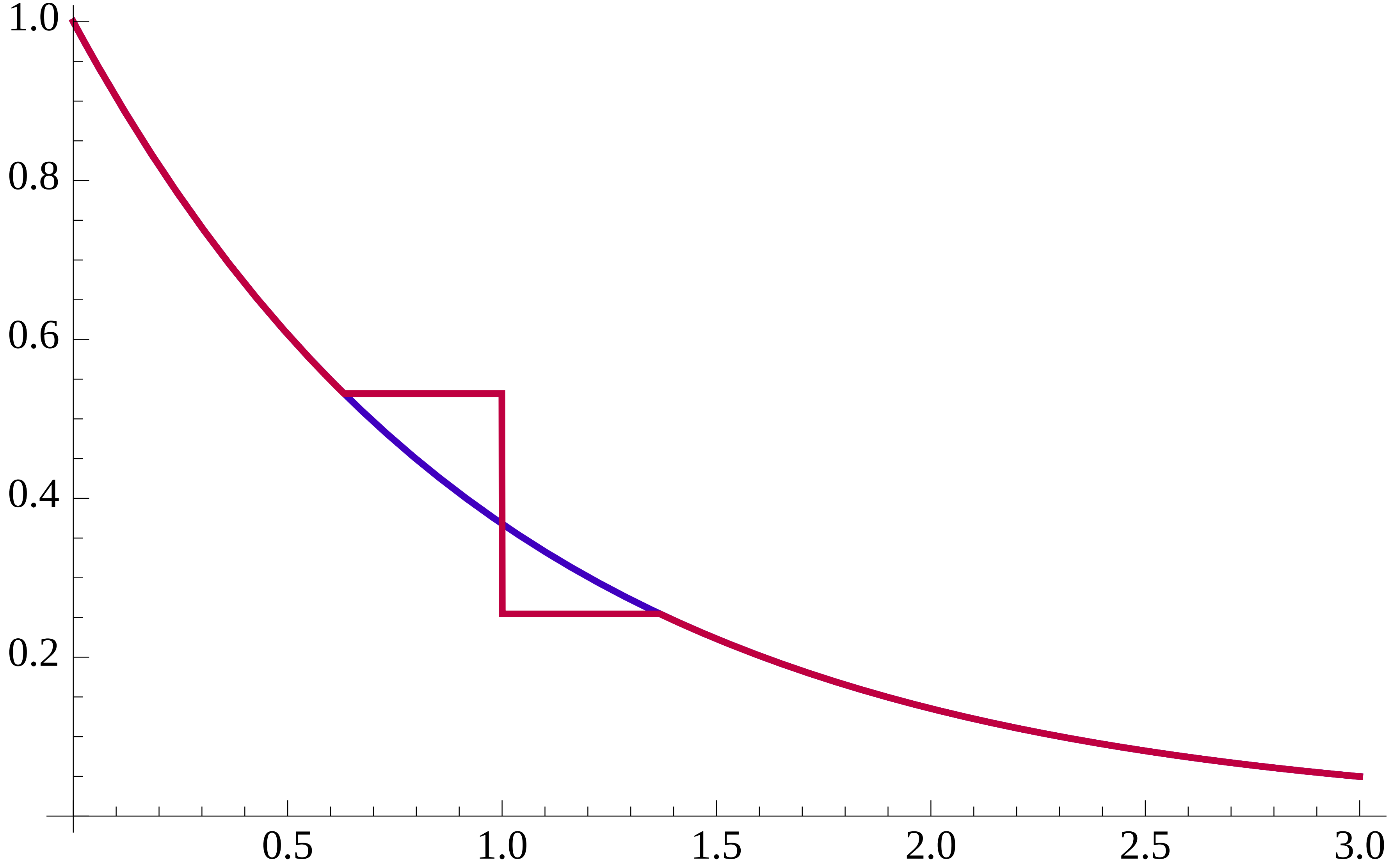
for some constant A, and

$$|\nu(f_n) - \nu(f_0)| = b_n^{-1}$$

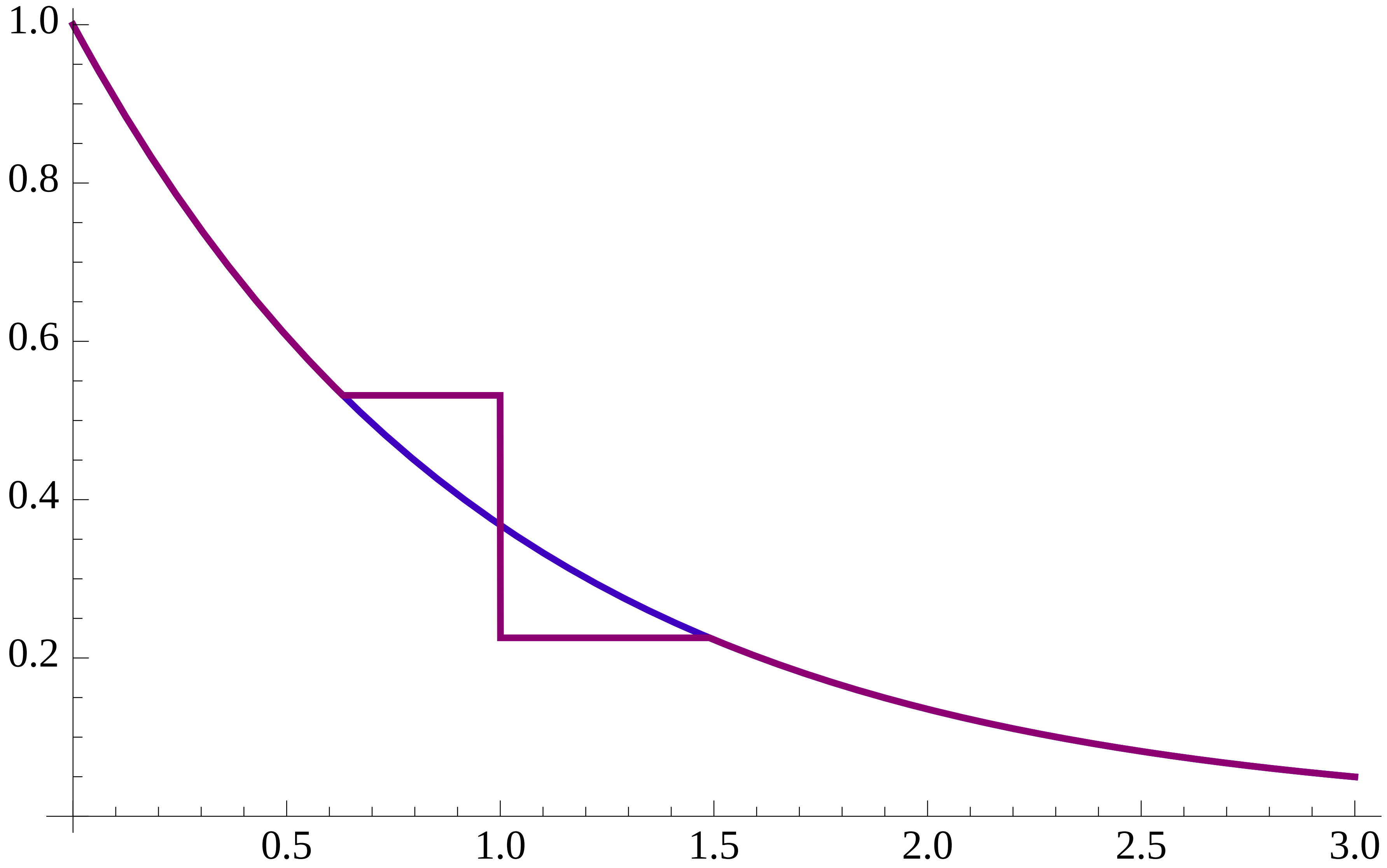
where $b_n \to \infty$. Hence we will try the following choice of f_n. For $c > 0$, define

$$f_n(x) = \begin{cases} f_0(x) & \text{if} \ \ x \le x_0 - cn^{-1/3} \ \ \text{or} \ \ x > x_0 + cn^{-1/3}, \\ f_0(x_0 - cn^{-1/3}) & \text{if} \ \ x_0 - cn^{-1/3} < x \le x_0, \\ f_0(x_0 + cn^{-1/3}) & \text{if} \ \ x_0 < x \le x_0 + cn^{-1/3}. \end{cases}$$

B. Lower bounds, monotone density



B. Lower bounds, monotone density



B. Lower bounds, monotone density

It is easy to see that

$$\begin{aligned} n^{1/3}|\nu(f_n)-\nu(f_0)| &= |n^{1/3}(f_0(x_0-cn^{-1/3})-f_0(x_0))| \\ &\to |f_0'(x_0)|c \qquad (3) \end{aligned}$$

On the other hand some calculation shows that

$$\begin{aligned} H^2(p_n,p_0) &= \frac{1}{2}\int_0^\infty [\sqrt{f_n(x)}-\sqrt{f_0(x)}]^2\,dx \\ &= \frac{1}{2}\int_0^\infty \frac{[\sqrt{f_n(x)}-\sqrt{f_0(x)}]^2[\sqrt{f_n(x)}+\sqrt{f_0(x)}]^2}{[\sqrt{f_n(x)}+\sqrt{f_0(x)}]^2}\,dx \\ &= \frac{1}{2}\int_{x_0-cn^{-1/3}}^{x_0+cn^{-1/3}} \frac{[f_n(x)-f_0(x)]^2}{[\sqrt{f_n(x)}+\sqrt{f_0(x)}]^2}\,dx \\ &\sim \frac{f_0'(x_0)^2}{4f_0(x_0)}\frac{c^3}{3n}. \end{aligned}$$

B. Lower bounds, monotone density

Now we can combine these two pieces with our two-point lower bound Proposition to find that, for any estimator T_n of $\nu(f) = f(x_0)$ and the loss function $l(x) = |x|$ we have

$$\inf_{T_n} \max \left\{ E_n n^{1/3} |T_n - \nu(f_n)|, \, E_0 n^{1/3} |T_n - \nu(f_0)| \right\}$$

$$\geq \frac{1}{4} |n^{1/3}(\nu(f_n) - \nu(f_0))| \left\{ 1 - \frac{nH^2(f_n, f_0)}{n} \right\}^{2n}$$

$$= \frac{1}{4} |n^{1/3}(f_0(x_0 - cn^{-1/3}) - f_0(x_0))| \left\{ 1 - \frac{nH^2(f_n, f_0)}{n} \right\}^{2n}$$

$$\to \frac{1}{4} |f_0'(x_0)| c \exp\left(-2 \frac{f_0'(x_0)^2}{12 f_0(x_0)} c^3 \right) = \frac{1}{4} |f_0'(x_0)| c \exp\left(- \frac{f_0'(x_0)^2}{6 f_0(x_0)} c^3 \right)$$

B. Lower bounds, monotone density

We now choose c to maximize the quantity on the right side. It is easily seen that the maximum is achieved when

$$c = c_0 \equiv \left(\frac{2 f_0(x_0)}{f_0'(x_0)^2} \right)^{1/3} .$$

This yields

$$\liminf_{n\to\infty} \inf_{T_n} \max \left\{ E_n n^{1/3} |T_n - \nu(f_n)|,\ E_0 n^{1/3} |T_n - \nu(f_0)| \right\} \geq \frac{e^{-1/3}}{4} \left(2 |f_0'(x_0)| f_0(x_0) \right)^{1/3} .$$

This lower bound has the appropriate structure in the sense that the (nonparametric) MLE of f, $\widehat{f}_n(x_0)$ converges at rate $n^{1/3}$ and it has the same dependence on $f_0(x_0)$ and $f_0'(x_0)$ as does the MLE.

B. Lower bounds, monotone density

Furthermore, note that for n sufficiently large

$$\sup_{f: H(f,f_0) \le C n^{-1/2}} E_f |T_n - \nu(f)|$$
$$\ge \max \left\{ E_n n^{1/3} |T_n - \nu(f_n)|,\ E_0 n^{1/3} |T_n - \nu(f_0)| \right\}$$

if $C^2 > 2A \equiv 2 f_0'(x_0)^2 c_0^3 / (12 f_0(x_0))$, and hence we conclude that

$$\liminf_{n\to\infty} \inf_{T_n} \sup_{f: H(f,f_0) \le C n^{-1/2}} E_f |T_n - \nu(f)|$$
$$\ge \frac{e^{-1/3}}{4} \left(2 |f_0'(x_0)| f_0(x_0) \right)^{1/3}$$
$$= \frac{e^{-1/3}}{4^{2/3}} \left(2^{-1} |f_0'(x_0)| f_0(x_0) \right)^{1/3}$$

for all C sufficiently large.

Comparison of $E|\mathbb{S}(0)|$ with $\frac{e^{-1/3}}{4^{2/3}} = 0.284356$? From Groeneboom and Wellner (2001), $E|\mathbb{S}(0)| = 2E|Z| = 2(.41273655) = 0.825473$.

B. Lower bounds, monotone density

S2: $x_0 \in (a,b)$ with $f_0(x) = f_0(x_0) > 0$ for all $x \in (a,b)$. To apply our two-point lower bound Proposition we again need to construct a sequence of densities f_n that are "near" f_0 in the sense that $nH^2(f_n, f_0) \rightarrow \ A$ for some constant A, and $|\nu(f_n) - \nu(f_0)| = b_n^{-1}$ where $b_n \rightarrow \infty$. In this scenario we define a sequence of densities $\{f_n\}$ by

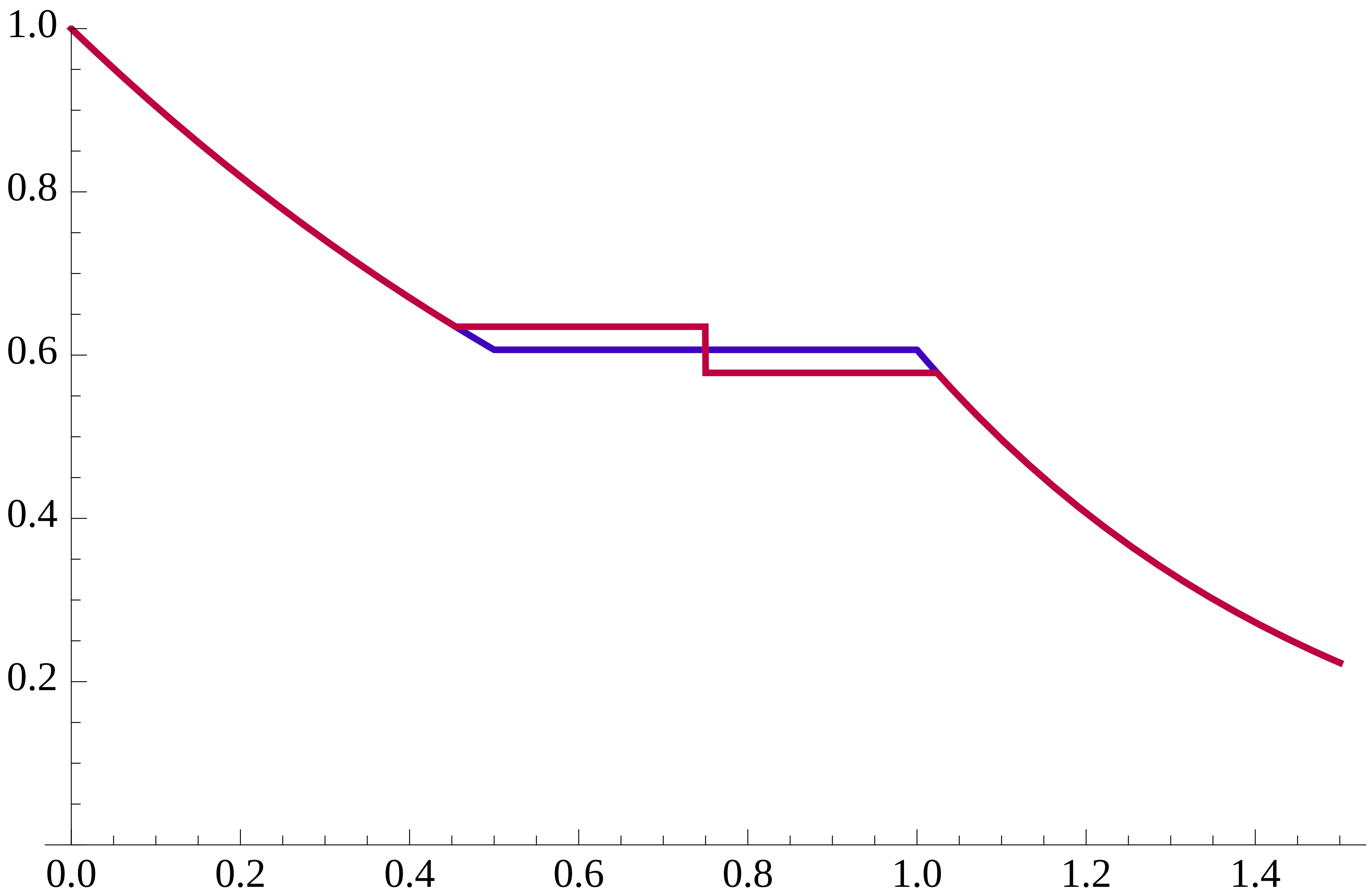
$$f_n(x) = \begin{cases} f_0(x), & x \le a_n \\ f_0(x) + \frac{c}{\sqrt{n}} \frac{b-a}{x_0 - a}\ , & a_n < x \le x_0 \\ f_0(x) - \frac{c}{\sqrt{n}} \frac{b-a}{b-x_0} & x_0 < x < \tilde{b}_n \\ f_0(x), & b \ge b_n . \end{cases}$$

where

$$a_n \equiv \sup\{x : f_0(x) \ge f_0(x_0) + cn^{-1/2}(b-a)/(x_0 - a)\}$$
$$b_n \equiv \inf\{x : f_0(x) < f_0(x_0) - cn^{-1/2}(b-a)/(b - x_0)\}.$$

The intervals (a_n, a) and $(b, \tilde{b}_n)$ may be empty if $f(a-) > f(a+)$ and/or $f(b+) < f(b-)$ and n is large.

B. Lower bounds, monotone density



B. Lower bounds, monotone density

It is easy to see that

$$\sqrt{n}|\nu(f_n) - \nu(f_0)| = \sqrt{n}|f_n(x_0) - f_0(x_0)| = c\frac{b-a}{x_0 - a} \tag{4}$$

On the other hand some calculation shows that

$$\begin{aligned} H^2(f_n, f_0) &\sim \frac{c^2(b-a)^2}{4nf_0(x_0)}\left\{\frac{1}{x_0 - a} + \frac{1}{b - x_0}\right\} \\ &= \frac{c^2(b-a)^3}{4nf_0(x_0)(x_0 - a)(b - x_0)}. \end{aligned}$$

B. Lower bounds, monotone density

Combining these two pieces with the two-point lower bound Proposition we find that, in scenario 2, for any estimator T_n of $\nu(f) = f(x_0)$ and the loss function $l(x) = |x|$ we have

$$\begin{aligned}
&\inf_{T_n} \max\left\{E_n\sqrt{n}|T_n - \nu(f_n)|,\, E_0\sqrt{n}|T_n - \nu(f_0)|\right\} \\
&\geq \frac{1}{4}|\sqrt{n}(\nu(f_n) - \nu(f_0))| \left\{1 - \frac{nH^2(f_n, f_0)}{n}\right\}^{2n} \\
&= \frac{1}{4}c\frac{b-a}{x_0-a}\left\{1 - \frac{nH^2(f_n, f_0)}{n}\right\}^{2n} \\
&\to \frac{1}{4}c\frac{b-a}{x_0-a}\exp\left(-\frac{c^2(b-a)^3}{2f_0(x_0)(x_0-a)(b-x_0)}\right) \\
&\equiv Ac\,\exp(-Bc^2)
\end{aligned}$$

B. Lower bounds, monotone density

We now choose c to maximize the quantity on the right side. It is easily seen that the maximum is achieved when $c = c_0 \equiv 1/\sqrt{2B}$, with $Ac_0 \exp(-Bc_0^2) = Ac_0 \exp(-1/2)$ and

$$c_0 = \left(\frac{f_0(x_0)}{(x_0 - a)(b - x_0)} (b-a)^3 \right)^{1/2} .$$

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

Repeating this argument with the right-continuous version of the sequence $\{f_n\}$ yields a similar bound, but with the factor $\sqrt{(b - x_0)/(x_0 - a)}$ replaced by $\sqrt{(x_0 - a)/(b - x_0)}$.

B. Lower bounds, monotone density

By taking the maximum of the two lower bounds yields the last display with the right side replaced by

$$\frac{e^{-1/2}}{4}\sqrt{\frac{f_0(x_0)}{b-a}}\,\mathrm{max}\left\{\sqrt{\frac{b-x_0}{x_0-a}},\sqrt{\frac{x_0-a}{b-x_0}}\right\}$$

$$\geq \frac{e^{-1/2}}{4}\sqrt{\frac{f_0(x_0)}{b-a}}\left\{\sqrt{\frac{b-x_0}{x_0-a}}\cdot\frac{b-x_0}{b-a}+\sqrt{\frac{x_0-a}{b-x_0}}\cdot\frac{x_0-a}{b-a}\right\}.$$

This lower bound has the appropriate structure in the sense that the MLE of f, $\widehat{f}_n(x_0)$ converges at rate $n^{1/2}$ and the limiting behavior of the MLE has exactly the same dependence on $f_0(x_0)$, $b-a$, x_0-a, and $b-x_0$.

B. Lower bounds, monotone density

Theorem. (Carolan and Dykstra, 1999) If f_0 is decreasing with f_0 constant on (a, b), the maximal open interval containing x_0, then, with $p \equiv f_0(x_0)(b-a) = P_0(a < X < b)$,

$$\sqrt{n}(\widehat{f}_n(x_0) - f_0(x_0)) \to_d \sqrt{\frac{f_0(x_0)}{b-a}} \left\{ \sqrt{1-p}Z + \mathbb{S}\left(\frac{x_0 - a}{b-a}\right) \right\}$$

where $Z \sim N(0,1)$ and $\mathbb{S}$ is the process of left-derivatives of the least concave majorant $\widehat{\mathbb{U}}$ of a Brownian bridge process $\mathbb{U}$ independent of Z.

Note that by using Groeneboom (1983)

$$\begin{aligned}
&E\left|\sqrt{\frac{f_0(x_0)}{b-a}} \left\{ \sqrt{1-p}Z + \mathbb{S}\left(\frac{x_0 - a}{b-a}\right) \right\}\right| \\
&\geq \sqrt{\frac{f_0(x_0)}{b-a}} E\left|\mathbb{S}\left(\frac{x_0 - a}{b-a}\right)\right| \\
&= \sqrt{\frac{f_0(x_0)}{b-a}} 2\sqrt{\frac{2}{\pi(b-a)}} \left\{ \frac{(b-x_0)^{3/2}}{(x_0-a)^{1/2}} + \frac{(x_0-a)^{3/2}}{(b-x_0)^{1/2}} \right\}.
\end{aligned}$$

B. Lower bounds, monotone density

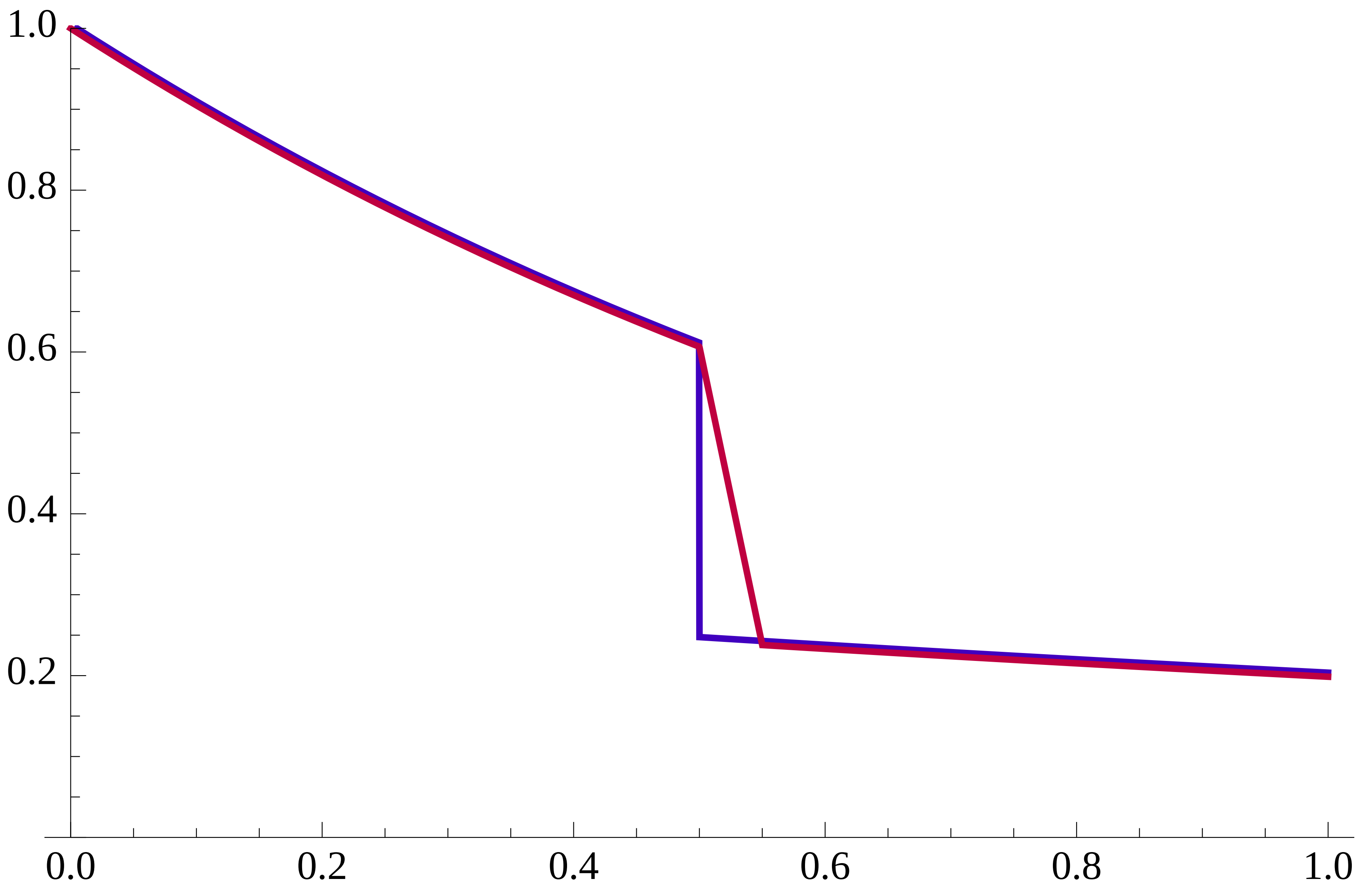
S3: $f_0(x_0-) > f_0(x_0+)$. In this case we consider estimation of the functional $\nu(f) = (f(x_0+) + f(x_0-))/2 \equiv \overline{f}(x_0)$. To apply our two-point lower bound Proposition, consider the following choice of f_n: for $c > 0$, define

$$\tilde{f}_n(x) = \begin{cases} f_0(x) & \text{if} \quad x \le x_0 \quad \text{or} \quad x > b_n, \\ f_0(x_0) & \\ +(x - x_0)\frac{f_0(b_n) - f_0(x_0)}{c/n} & \text{if} \quad x_0 < x \le b_n. \end{cases}$$

where $b_n \equiv x_0 + c/n$. Then define $f_n = \tilde{f}_n / \int_0^\infty \tilde{f}_n(y)dy$.
In this case

$$\begin{aligned} \nu(f_n) - \nu(f_0) &= f_n(x_0) - f_0(x_0-) = \frac{\tilde{f}_n(x_0)}{1 + o(1)} - \frac{f_0(x_0+) + f_0(x_0-)}{2} \\ &= \frac{1}{2}(f_0(x_0-) - f_0(x_0+)) + o(1) \equiv d + o(1). \end{aligned}$$

B. Lower bounds, monotone density



B. Lower bounds, monotone density

Some calculation shows that

$$H^2(f_n, f_0) = \frac{cr^2}{n}(1 + o(1)) \quad \text{where}$$

$$r^2 = \frac{\{\sqrt{f_0(x_0-)} - \sqrt{f_0(x_0+)}\}^2\{3\sqrt{f_0(x_0-)} + \sqrt{f_0(x_0+)}\}}{\sqrt{f_0(x_0-)} + \sqrt{f_0(x_0+)}}.$$

Combining these pieces with the two-point lower bound yields

$$\begin{aligned}
&\inf_{T_n} \max \{E_n|T_n - \nu(f_n)|,\, E_0|T_n - \nu(f_0)|\} \\
&\quad \geq \frac{1}{4}|\nu(f_n) - \nu(f_0)| \left\{1 - \frac{nH^2(f_n, f_0)}{n}\right\}^{2n} \\
&\quad = \frac{1}{8}(f_0(x_0-) - f_0(x_0+))(1 + o(1)) \left\{1 - \frac{cr^2(1 + o(1))}{n}\right\}^{2n} \\
&\quad \to \frac{d}{4} \exp\left(-cr^2\right) = \frac{d}{4e} \qquad \text{by choosing} \quad c = 1/r^2.
\end{aligned}$$

B. Lower bounds, monotone density

This corresponds to the following theorem for the MLE $\widehat{f}_n$:

Theorem. (Anevski and Hössjer, 2002; W, 2007) If x_0 is a discontinuity point of f_0, $d \equiv (f_0(x_0-) - f_0(x_0+))/2$ with $f_0(x_0+) > 0$ and $\overline{f}(x_0) \equiv (f_0(x_0) + f_0(x_0-))/2$, then

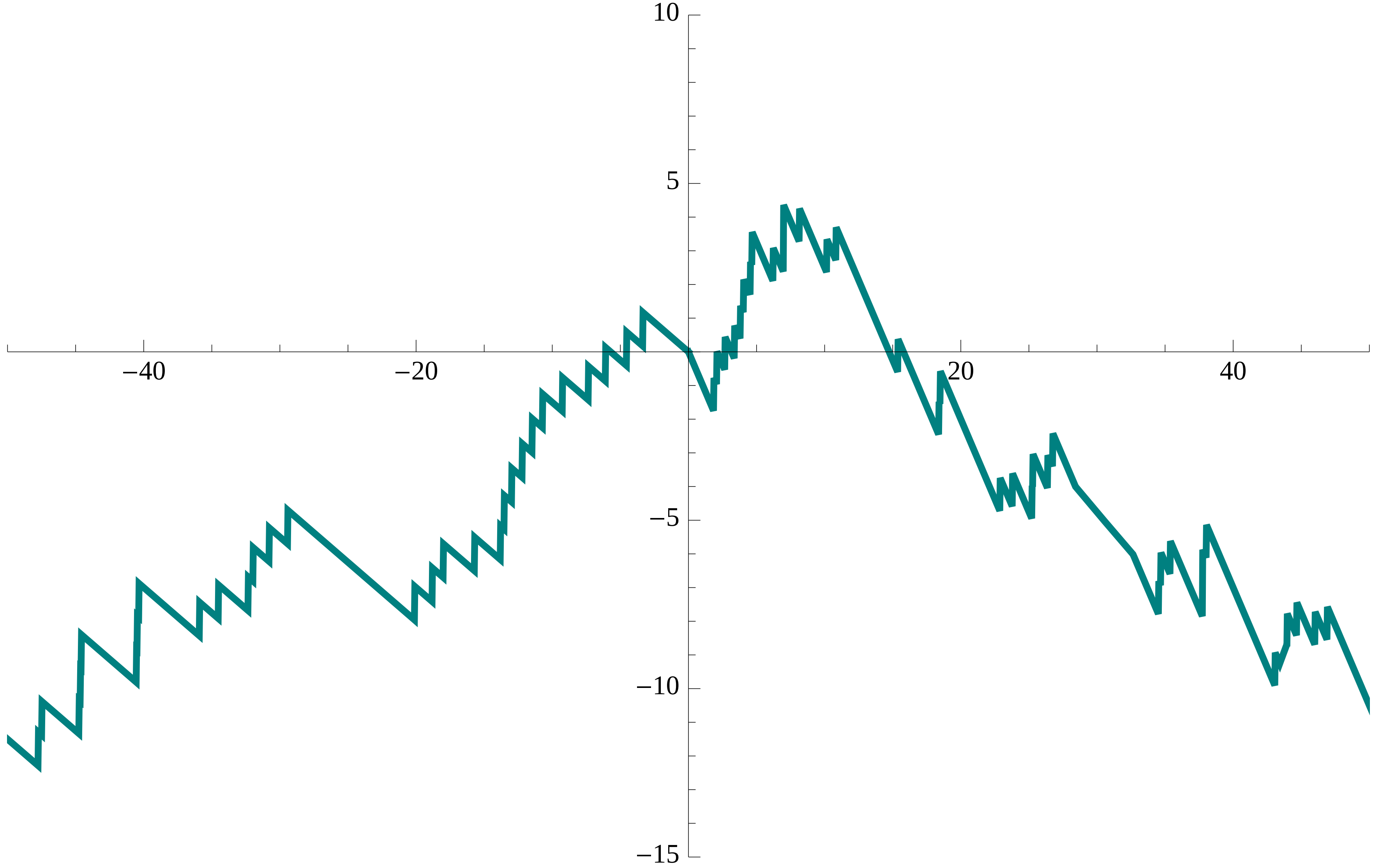
$$\widehat{f}_n(x_0) - \overline{f}_0(x_0) \to_d \mathbb{R}(0)$$

where $h \mapsto \mathbb{R}(h)$ is the process of left-derivatives of the least concave majorant $\widehat{\mathbb{M}}$ of the process $\mathbb{M}$ defined by

$$\mathbb{M}(h) \;=\; \mathbb{N}_0(h) - d|h| \equiv \begin{cases} \mathbb{N}(f_0(x_0+)h) - f_0(x_0+)h - dh, & h \geq 0 \\ -\mathbb{N}(f_0(x_0-)h) - f_0(x_0-)h + dh, & h < 0 \end{cases}$$

where $\mathbb{N}$ is a standard (rate 1) two-sided Poisson process on $\mathbb{R}$.

B. Lower bounds, monotone density



B. Lower bounds, monotone density

S4: $f_0(x_0) > 0$, $f_0^{(j)}(x_0) = 0$, $j = 1, 2, \ldots, p-1$, and $f_0^{(p)}(x_0) \neq 0$. In this case, consider the perturbation f_ϵ of f_0 given for $\epsilon > 0$ by

$$f_\epsilon(x) = \begin{cases} f_0(x) & \text{if } \; x \le x_0 - \epsilon \;\; \text{or} \;\; x > x_0 + \epsilon, \\ f_0(x_0 - \epsilon) & \text{if } \; x_0 - \epsilon < x \le x_0 \\ f_0(x_0 + \epsilon) & \text{if } \; x_0 < x \le x_0 + \epsilon. \end{cases}$$

Then for $\nu(f) = f(x_0)$

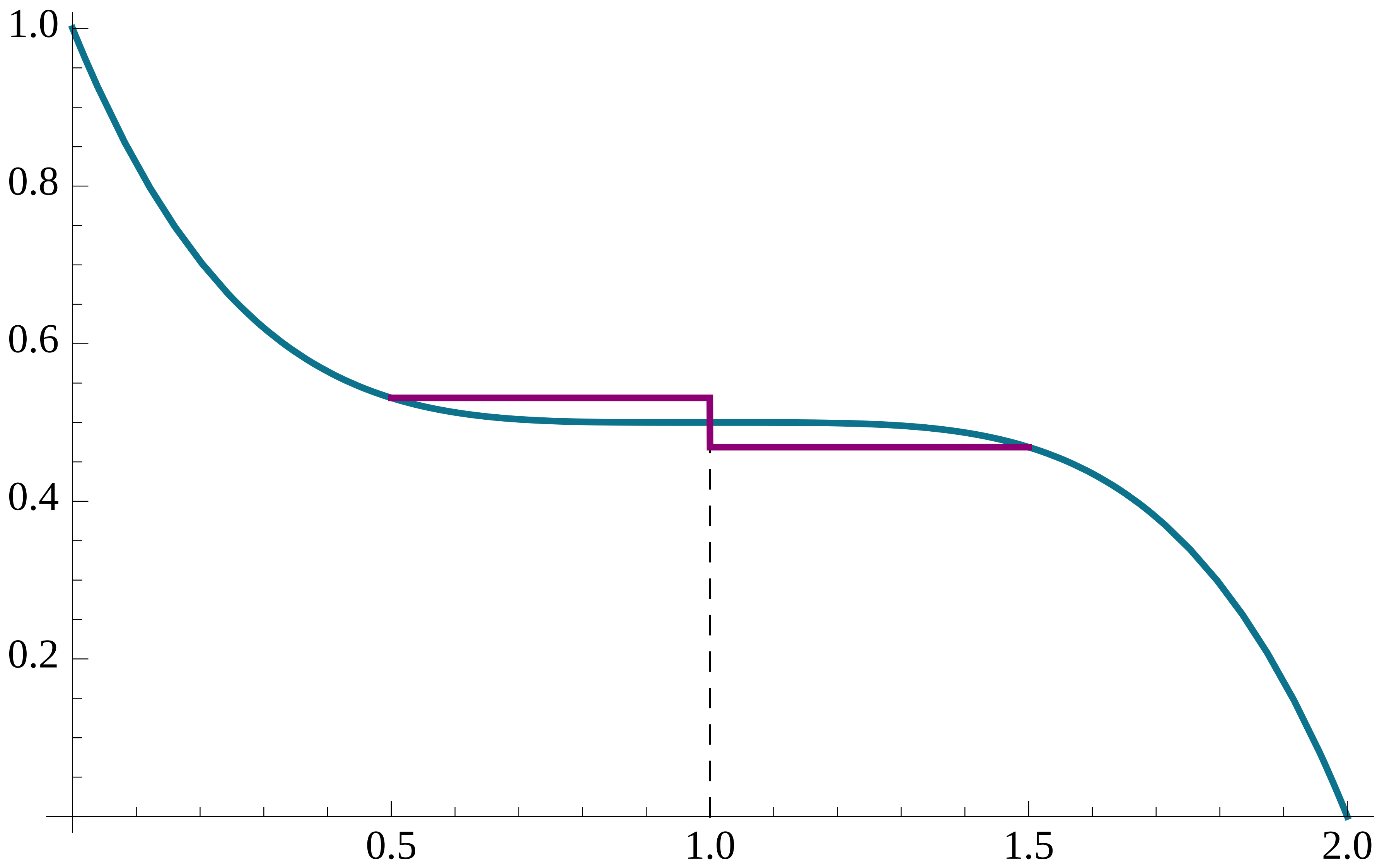
$$\nu(f_\epsilon) - \nu(f_0) \sim \frac{|f_0^{(p)}(x_0)|}{p!} \epsilon^p,$$

$$H^2(f_\epsilon, f_0) \sim A_p \frac{|f_0^{(p)}(x_0)|^2}{f_0(x_0)} \epsilon^{2p+1} \equiv B_p \epsilon^{2p+1}$$

where

$$A_p \equiv \frac{2p^2}{(2p!)^2(2p^2 + 3p + 1)}.$$

B. Lower bounds, monotone density



B. Lower bounds, monotone density

Choosing $\epsilon = cn^{-1/(2p+1)}$, plugging into our two-point bound, and optimizing with respect to c yields

$$\inf_{T_n} \max \left\{ n^{p/(2p+1)} E_n |T_n - \nu(f_n)|,\, n^{p/(2p+1)} E_0 |T_n - \nu(f_0)| \right\}$$

$$\geq \frac{1}{4} |\nu(f_n) - \nu(f_0)| \left\{ 1 - \frac{nH^2(f_n, f_0)}{n} \right\}^{2n}$$

$$\to \frac{1}{4} \frac{|f_0^{(p)}(x_0)|}{p!} c^p \exp\left(-2B_p c^{2p+1} \right)$$

$$= D_p \left(|f_0^{(p)}(x_0)| f_0(x_0)^p \right)^{1/(2p+1)} \quad \text{taking} \quad c = \left(\frac{p}{(2p+1)B_p} \right)^{1/(2p+1)}$$

with

$$D_p \equiv \frac{1}{4p!} \cdot \left(\frac{p^p}{(2p+1)A_p^p} \right)^{1/(2p+1)} \exp(-p/(2p+1)).$$

B. Lower bounds, monotone density

The resulting lower bound corresponds to the following theorem for $\widehat{f}_n$:

Theorem. (Wright (1981); Leurgans (1982); Anevski and Hössjer (2002)) Suppose that $f_0^{(j)}(x_0) = 0$ for $j = 1, \ldots, p-1$, $f_0^{(p)}(x_0) \neq 0$, and $f_0^{(p)}$ is continuous at x_0. Then

$$n^{p/(2p+1)}(\widehat{f}_n(x_0 + n^{-1/(2p+1)}t) - f_0(x_0)) \rightarrow_d C_p \mathbb{S}_p(t)$$

where $\mathbb{S}_p$ is the process given by the left-derivatives of the least concave majorant $\widehat{\mathbb{Y}}_p$ of $\mathbb{Y}_p(t) \equiv W(t) - |t|^{p+1}$, and where

$$C_p = \left(f_0(x_0)^p |f_0^{(p)}(x_0)| / (p+1)! \right)^{1/(2p+1)} .$$

In particular

$$n^{p/(2p+1)}(\widehat{f}_n(x_0) - f_0(x_0)) \rightarrow_d C_p \mathbb{S}_p(0)$$

Proof. Switching + (argmax-)continuous mapping theorem.

B. Lower bounds, monotone density

Summary: The MLE $\widehat{f}_n$ is *locally adaptive* to f_0, at least in scenarios 1-4.

S1: rate $n^{1/3}$; localization $n^{-1/3}$; constants agree with minimax lower bound.

S2: rate $n^{1/2}$; localization $n^0 = 1$, *none*; constants agree with minimax bound.

S3: rate $n^0 = 1$; localization n^{-1}; constants agree(?).

S4: rate $n^{p/(2p+1)}$; localization $n^{-1/(2p+1)}$; constants agree.

C: Global lower and upper bounds (briefly)

Birgé (1986, 1989) expresses the global optimality of $\widehat{f}_n$ in terms of its L_1-risks as follows:

Lower bound: Birgé (1987). Let $\mathcal{F}$ denote the class of all decreasing densities f on $[0, 1]$ satisfying $f \le M$ with $M > 1$. Then the minimax risk for $\mathcal{F}$ with respect to the L_1 metric $d_1(f, g) \equiv \int |f(x) - g(x)| dx$ based on n observations is

$$R_M(d_1, n) \equiv \inf_{\widehat{f}_n} \sup_{f \in \mathcal{F}} E_f d_1(\widehat{f}_n, f).$$

Then there is an absolute constant C such that

$$R_M(d_1, n) \ge C \left(\frac{\log M}{n} \right)^{1/3}.$$

Upper bound, Grenander: Birgé (1989). Let $\widehat{f}_n$ denote the Grenander estimator of $f \in \mathcal{F}$. Then

$$\sup_{f \in \mathcal{F}_M} E_f d_1(\widehat{f}_n, f) \le 4.75 \left(\frac{\log M}{n} \right)^{1/3}.$$

C: Global lower and upper bounds (briefly)

Birgé's bounds are complemented by the remarkable results of Groeneboom (1985), Groeneboom, Hooghiemstra, and Lopuhaa (1999). Set

$$V(t) \equiv \sup\{s : \; W(s) - (s-t)^2 \;\text{ is maximal}\}$$

where W is a standard two-sided Brownian motion process starting from 0.

Theorem. (Groeneboom (1985), GHL (1999)) Suppose that f is a decreasing density on $[0,1]$ satisfying:

- A1. $0 < f(1) \le f(y) \le f(x) \le f(0) < \infty$ for $0 \le x \le y \le 1$.
- A2. $0 < \inf_{0<x<1} |f'(x)| \le \sup_{0<x<1} |f'(x)| < \infty$.
- A3. $\sup_{0<x<1} |f''(x)| < \infty$.

Then, with $\mu = 2E|V(0)| \int_0^1 |\frac{1}{2}f'(x)f(x)|^{1/3} dx$,

$$n^{1/6}\left\{ n^{1/3} \int_0^1 |\widehat{f}_n(x) - f(x)| dx - \mu \right\} \to_d \sigma Z \sim N(0, \sigma^2)$$

where $\sigma^2 = 8 \int_0^\infty Cov(|V(0)|, |V(t) - t|) dt$.

D: Lower bounds: convex decreasing density

Now consider estimation of a *convex decreasing density* f on $[0, \infty)$. (Original motivation: Hampel's (1987) bird-migration problem.) Since f' exists almost everywhere, we are now interested in in estimation of $\nu_1(f) = f(x_0)$ *and* $\nu_2(f) = f'(x_0)$.

We let $\mathcal{D}_2$ denote the class of all convex decreasing densities on $\mathbb{R}^+$. Note that every $f \in \mathcal{D}_2$ can be written as a scale mixture of the triangular (or Beta$(1,2)$) density: if $f \in \mathcal{D}_2$, then

$$f(x) = \int_0^\infty 2y^{-1}(1 - x/y)_+ dG(y)$$

for some (mixing) distribution G on $[0, \infty)$. This corresponds to the fact that monotone decreasing density $f \in \mathcal{D} \equiv \mathcal{D}_1$ can be written as a scale mixture of the Uniform$(0,1)$ (or Beta$(1,1)$) density: if $f \in \mathcal{D}_1$, then

$$f(x) = \int_0^\infty y^{-1} 1_{[0,y]}(x) dG(y)$$

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

D: Lower bounds: convex decreasing density

Scenario 1: Suppose that $f_0 \in \mathcal{D}_2$ and $x_0 \in (0, \infty)$ satisfy $f_0(x_0) > 0$, $f_0''(x_0) > 0$, and f_0'' is continuous at x_0.

To establish lower bounds, consider the perturbations $\tilde{f}_\epsilon$ of f_0 given by

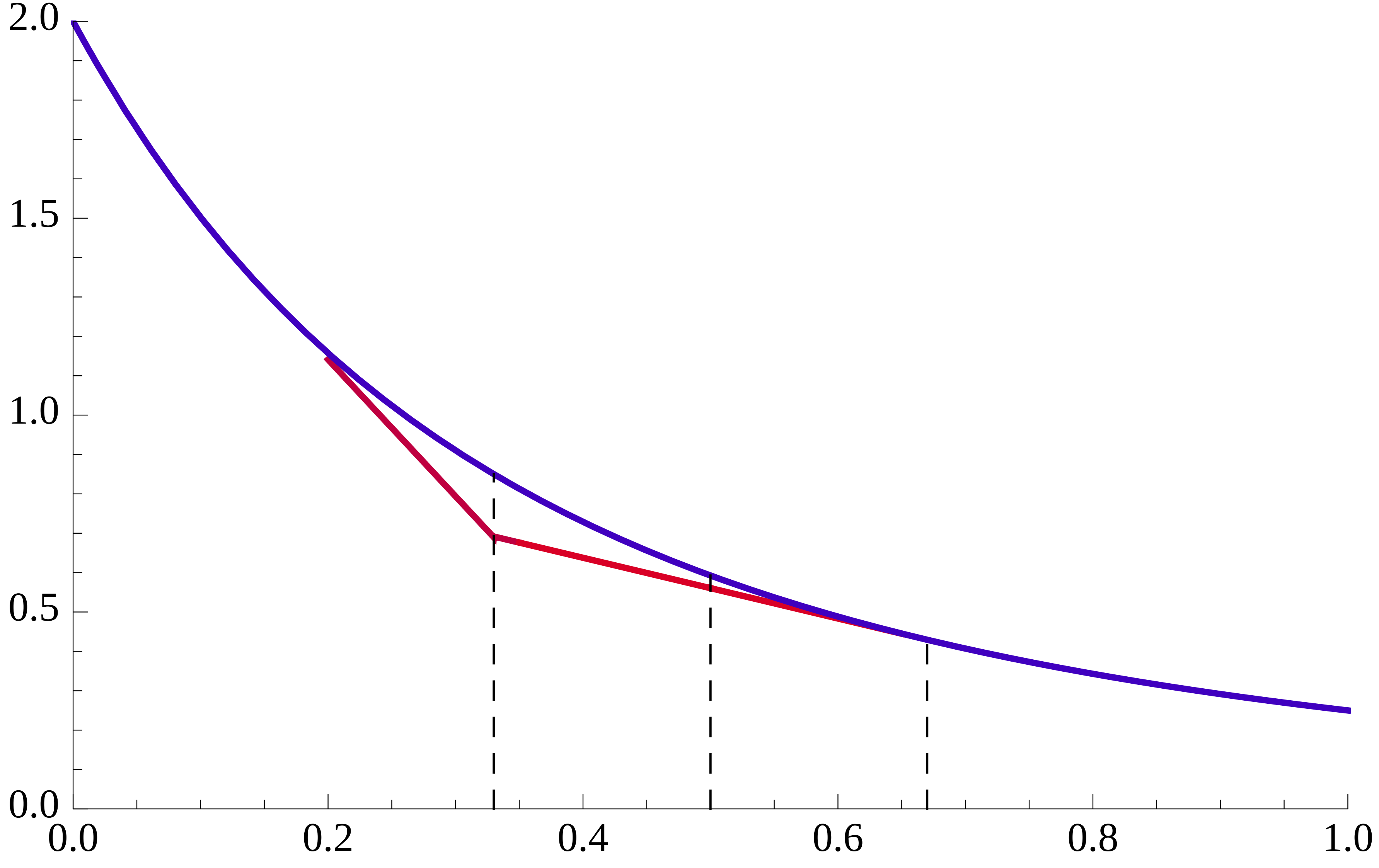
$$
\tilde{f}_\epsilon(x) = \begin{cases} f_0(x_0 - \epsilon c_\epsilon) + (x - x_0 + \epsilon c_\epsilon) f_0'(x_0 - \epsilon c_\epsilon), & x \in (x_0 - \epsilon c_\epsilon, x_0 - \epsilon), \\ f_0(x_0 + \epsilon) + (x - x_0 - \epsilon) f_0'(x_0 + \epsilon), & x \in (x_0 - \epsilon, x_0 + \epsilon), \\ f_0(x), & \text{elsewhere}; \end{cases}
$$

here c_ϵ is chosen so that $\tilde{f}_\epsilon$ is continuous at $x_0 - \epsilon$. Now define f_ϵ by

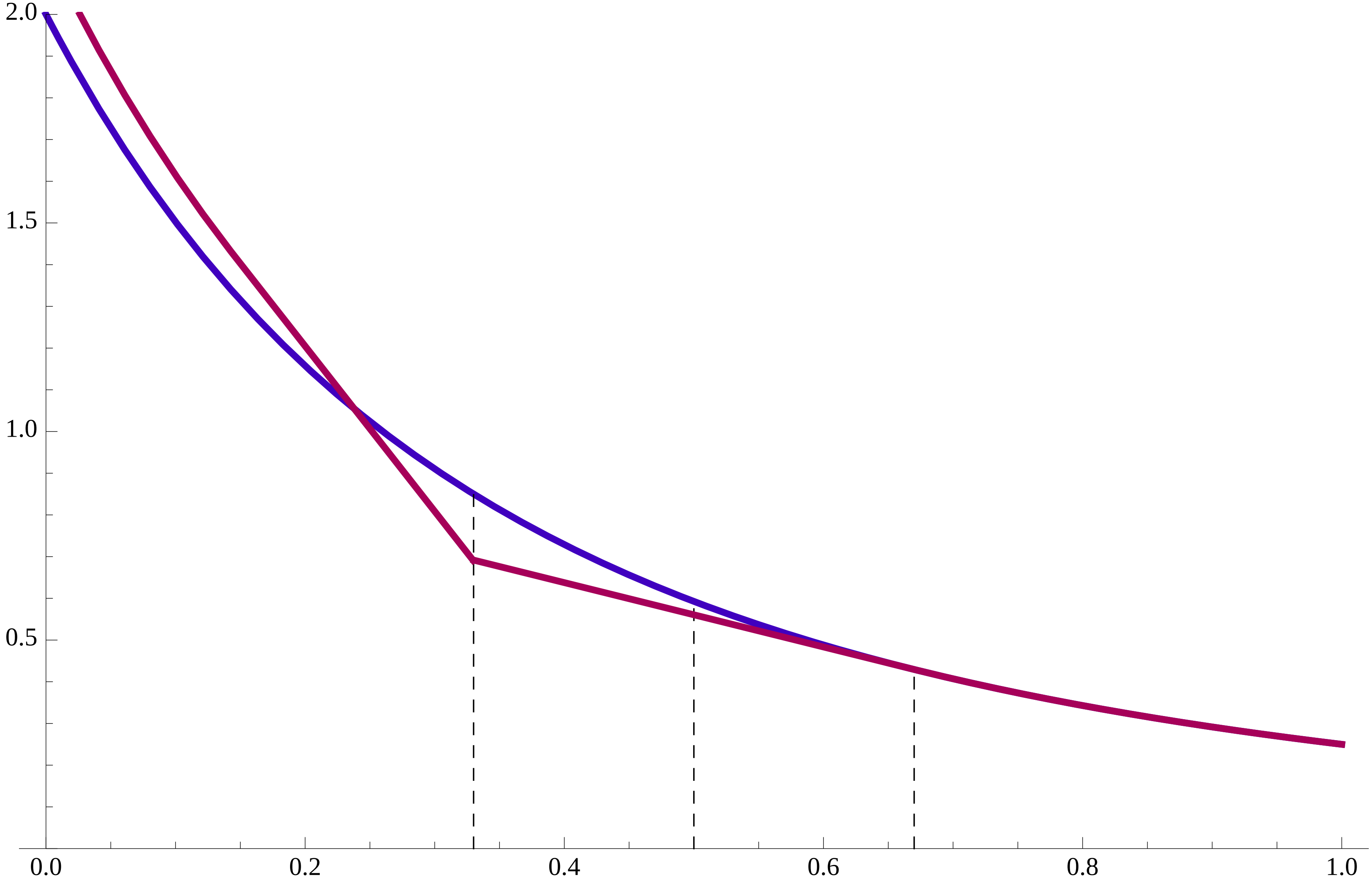
$$
f_\epsilon(x) = \tilde{f}_\epsilon(x) + \tau_\epsilon(x_0 - \epsilon - x) 1_{[0, x_0 - \epsilon]}(x)
$$

with τ_ϵ chosen so that f_ϵ integrates to 1.

D: Lower bounds: convex decreasing density



D: Lower bounds: convex decreasing density



D: Lower bounds: convex decreasing density

Now

$$|\nu_1(f_\epsilon) - \nu_1(f_0)| = |f_\epsilon(x_0) - f_0(x_0)| \sim \frac{1}{2} f_0^{(2)}(x_0)\epsilon^2(1 + o(1)),$$

$$|\nu_2(f_\epsilon) - \nu_2(f_0)| = |f_\epsilon'(x_0) - f_0'(x_0)| \sim f_0^{(2)}(x_0)\epsilon(1 + o(1)),$$

and some further computation (Jongbloed (1995), (2000)) shows that

$$H^2(f_\epsilon, f_0) = \frac{2 f_0^{(2)}(x_0)^2}{5 f_0(x_0)} \epsilon^5 (1 + o(1)).$$

Thus taking $\epsilon \equiv \epsilon_n = cn^{-1/5}$, writing f_n for f_{ϵ_n}, and using our two-point lower bound proposition yields

$$\liminf_{n\to\infty} \inf_{T_n} \max\left\{E_n n^{2/5}|T_n - \nu_1(f_n)|,\ E_0 n^{2/5}|T_n - \nu_1(f_0)|\right\}$$

$$\geq \frac{1}{4}\left(\frac{f_0^2(x_0) f_0^{(2)}(x_0)}{2 \cdot 8^2 e^2}\right)^{1/5},$$

and ...

D: Lower bounds: convex decreasing density

$$\liminf_{n\to\infty} \inf_{T_n} \max\left\{E_n n^{1/5}|T_n - \nu_2(f_n)|,\ E_0 n^{1/5}|T_n - \nu_2(f_0)|\right\}$$
$$\geq \frac{1}{4}\left(\frac{f_0(x_0) f_0^{(2)}(x_0)^3}{4e}\right)^{1/5} .$$

We will see that the MLE achieves these rates and that the limiting distributions involve exactly these constants tomorrow.

Other Scenarios?

S2: f_0 triangular on $[0,1]$?
(Degenerate mixing distribution at 1.)

S3: $x_0 \in (a,b)$ where f_0 is linear on (a,b)?

S4: x_0 a "bend" or "kink" point for x_0: $f_0'(x_0-) < f_0'(x_0+)$?

S5: $\cdots$?