Some Theory for Estimation with Shape Constraints

Jon A. Wellner

University of Washington

- Talk at meeting on
  Nonsmooth Inference, Analysis, and Dependence
  Nya Varvet, Göteborg,
  Sweden, June 10, 2008

- Talk at meeting on
  Nonsmooth Inference, Analysis, and Dependence
  Nya Varvet, Göteborg,
  Sweden, June 10, 2008
- *Email: jaw@stat.washington.edu*
  http: //www.stat.washington.edu/jaw/jaw.research.html

- Talk at meeting on
  Nonsmooth Inference, Analysis, and Dependence
  Nya Varvet, Göteborg,
  Sweden, June 10, 2008
- *Email: jaw@stat.washington.edu*
  http: //www.stat.washington.edu/jaw/jaw.research.html
- Based on joint work with:

- Talk at meeting on
  Nonsmooth Inference, Analysis, and Dependence
  Nya Varvet, Göteborg,
  Sweden, June 10, 2008
- *Email: jaw@stat.washington.edu*
  http: //www.stat.washington.edu/jaw/jaw.research.html
- Based on joint work with:
  - Piet Groeneboom and Geurt Jongbloed;

- Talk at meeting on
  Nonsmooth Inference, Analysis, and Dependence
  Nya Varvet, Göteborg,
  Sweden, June 10, 2008
- *Email: jaw@stat.washington.edu*
  http: //www.stat.washington.edu/jaw/jaw.research.html
- Based on joint work with:
  - Piet Groeneboom and Geurt Jongbloed;
  - former Ph.D. Students Jian Huang, Moulinath Banerjee, Fadoua Balabdaoui, Marloes Maathuis, Marios Pavlides; and Shuguang Song;

- Talk at meeting on
  Nonsmooth Inference, Analysis, and Dependence
  Nya Varvet, Göteborg,
  Sweden, June 10, 2008
- *Email: jaw@stat.washington.edu*
  http: //www.stat.washington.edu/jaw/jaw.research.html
- Based on joint work with:
  - Piet Groeneboom and Geurt Jongbloed;
  - former Ph.D. Students Jian Huang, Moulinath Banerjee, Fadoua Balabdaoui, Marloes Maathuis, Marios Pavlides; and Shuguang Song;
  - former post-doc Hanna Jankowski;

- Talk at meeting on
  Nonsmooth Inference, Analysis, and Dependence
  Nya Varvet, Göteborg,
  Sweden, June 10, 2008
- *Email: jaw@stat.washington.edu*
  http: //www.stat.washington.edu/jaw/jaw.research.html
- Based on joint work with:
  - Piet Groeneboom and Geurt Jongbloed;
  - former Ph.D. Students Jian Huang, Moulinath Banerjee, Fadoua Balabdaoui, Marloes Maathuis, Marios Pavlides; and Shuguang Song;
  - former post-doc Hanna Jankowski;
  - and the work of many others...

Types of shape restrictions for functions on $\mathbb{R}$

- Monotone

Types of shape restrictions for functions on $\mathbb{R}$

- Monotone
- Unimodal, antimodal, piecewise monotone

Types of shape restrictions for functions on $\mathbb{R}$

- Monotone
- Unimodal, antimodal, piecewise monotone
- Convex, concave, log-concave

Types of shape restrictions for functions on $\mathbb{R}$

- Monotone
- Unimodal, antimodal, piecewise monotone
- Convex, concave, log-concave
- $k-$monotone, completely monotone

Types of shape restrictions for functions on $\mathbb{R}$

- Monotone
- Unimodal, antimodal, piecewise monotone
- Convex, concave, log-concave
- $k-$monotone, completely monotone
- mixtures

Types of shape restrictions for functions on $\mathbb{R}$

- Monotone
- Unimodal, antimodal, piecewise monotone
- Convex, concave, log-concave
- $k-$monotone, completely monotone
- mixtures
  - normal (convolution - deconvolution)

Types of shape restrictions for functions on $\mathbb{R}^d$

- Monotone: blockwise (or weakly) increasing / decreasing

Types of shape restrictions for functions on $\mathbb{R}^d$

- Monotone: blockwise (or weakly) increasing / decreasing
- Monotone: mixtures of (anchored at zero rectangular) uniform densities

Types of shape restrictions for functions on $\mathbb{R}^d$

- Monotone: blockwise (or weakly) increasing / decreasing
- Monotone: mixtures of (anchored at zero rectangular) uniform densities
- Monotone: distribution functions

Types of shape restrictions for functions on $\mathbb{R}^d$

- Monotone: blockwise (or weakly) increasing / decreasing
- Monotone: mixtures of (anchored at zero rectangular) uniform densities
- Monotone: distribution functions
- Convex, concave, log-concave

Types of shape restrictions for functions on $\mathbb{R}^d$

- Monotone: blockwise (or weakly) increasing / decreasing
- Monotone: mixtures of (anchored at zero rectangular) uniform densities
- Monotone: distribution functions
- Convex, concave, log-concave
- Mixtures (normal, . . .)

Types of shape restrictions for functions on $\mathbb{R}^d$

- Monotone: blockwise (or weakly) increasing / decreasing
- Monotone: mixtures of (anchored at zero rectangular) uniform densities
- Monotone: distribution functions
- Convex, concave, log-concave
- Mixtures (normal, . . .)
- . . .

Types of functions to be estimated on $\mathbb{R}$ and $\mathbb{R}^d$

- Density function f

Types of functions to be estimated on $\mathbb{R}$ and $\mathbb{R}^d$

- Density function f
- Regression function, $r(x) = E(Y|X = x)$

Types of functions to be estimated on $\mathbb{R}$ and $\mathbb{R}^d$

- Density function f
- Regression function, $r(x) = E(Y|X = x)$
- Hazard function h:

$$h(x) = \frac{f(x)}{1 - F(x)}$$

Types of functions to be estimated on $\mathbb{R}$ and $\mathbb{R}^d$

- Density function f
- Regression function, $r(x) = E(Y|X = x)$
- Hazard function h:

$$h(x) = \frac{f(x)}{1 - F(x)}$$

- Intensity function λ of a Poison process

Types of functions to be estimated on $\mathbb{R}$ and $\mathbb{R}^d$

- Density function f
- Regression function, $r(x) = E(Y|X = x)$
- Hazard function h:

$$h(x) = \frac{f(x)}{1 - F(x)}$$

- Intensity function λ of a Poison process
- Distribution functions in interval censoring models (non-regular or "inverse problem")

Types of functions to be estimated on $\mathbb{R}$ and $\mathbb{R}^d$

- Density function f
- Regression function, $r(x) = E(Y|X = x)$
- Hazard function h:

$$h(x) = \frac{f(x)}{1 - F(x)}$$

- Intensity function λ of a Poison process
- Distribution functions in interval censoring models (non-regular or “inverse problem”)
- Distribution functions in direct observation models (regular or “forward problem”)

Types of functions to be estimated on $\mathbb{R}$ and $\mathbb{R}^d$

- Density function f
- Regression function, $r(x) = E(Y|X = x)$
- Hazard function h:

$$h(x) = \frac{f(x)}{1 - F(x)}$$

- Intensity function λ of a Poison process
- Distribution functions in interval censoring models (non-regular or “inverse problem”)
- Distribution functions in direct observation models (regular or “forward problem”)
- Spectral density function

What kind of theory?

- Local (pointwise), or

What kind of theory?

- Local (pointwise), or
- Global (with some global measure of deviations):
  - Hellinger distance (densities or intensities)
  - L_1 or L_2-distances
  - supremum metrics

What kind of theory?

- Local (pointwise), or
- Global (with some global measure of deviations):
  - Hellinger distance (densities or intensities)
  - L_1 or L_2-distances
  - supremum metrics
- Lower bounds, or

What kind of theory?

- Local (pointwise), or
- Global (with some global measure of deviations):
  - Hellinger distance (densities or intensities)
  - L_1 or L_2-distances
  - supremum metrics
- Lower bounds, or
- Upper bounds.

What kind of theory?

- Local (pointwise), or
- Global (with some global measure of deviations):
  - Hellinger distance (densities or intensities)
  - L_1 or L_2-distances
  - supremum metrics
- Lower bounds, or
- Upper bounds.
- Estimation, or

What kind of theory?

- Local (pointwise), or
- Global (with some global measure of deviations):
  - Hellinger distance (densities or intensities)
  - L_1 or L_2-distances
  - supremum metrics
- Lower bounds, or
- Upper bounds.
- Estimation, or
- Testing (within a shape constrained family)

What kind of theory?

- Local (pointwise), or
- Global (with some global measure of deviations):
  - Hellinger distance (densities or intensities)
  - L_1 or L_2-distances
  - supremum metrics
- Lower bounds, or
- Upper bounds.
- Estimation, or
- Testing (within a shape constrained family)
- Confidence sets?

What kind of theory?

- Local (pointwise), or
- Global (with some global measure of deviations):
  - Hellinger distance (densities or intensities)
  - L_1 or L_2-distances
  - supremum metrics
- Lower bounds, or
- Upper bounds.
- Estimation, or
- Testing (within a shape constrained family)
- Confidence sets?
- Assuming shape constraint, or

What kind of theory?

- Local (pointwise), or
- Global (with some global measure of deviations):
  - Hellinger distance (densities or intensities)
  - L_1 or L_2-distances
  - supremum metrics
- Lower bounds, or
- Upper bounds.
- Estimation, or
- Testing (within a shape constrained family)
- Confidence sets?
- Assuming shape constraint, or
- Testing to see if a shape constraint is true?

What type of estimator?

- Maximum likelihood.

What type of estimator?

- Maximum likelihood.
- Least squares (or minimum contrast) estimator

What type of estimator?

- Maximum likelihood.
- Least squares (or minimum contrast) estimator
- Bayes (nonparametric or shape restricted)

What type of estimator?

- Maximum likelihood.
- Least squares (or minimum contrast) estimator
- Bayes (nonparametric or shape restricted)
- Isotonization of kernel or other smooth estimators

What type of estimator?

- Maximum likelihood.
- Least squares (or minimum contrast) estimator
- Bayes (nonparametric or shape restricted)
- Isotonization of kernel or other smooth estimators
- Minimum distance

What type of estimator?

- Maximum likelihood.
- Least squares (or minimum contrast) estimator
- Bayes (nonparametric or shape restricted)
- Isotonization of kernel or other smooth estimators
- Minimum distance
- Rearrangement methods

What type of estimator?

- Maximum likelihood.
- Least squares (or minimum contrast) estimator
- Bayes (nonparametric or shape restricted)
- Isotonization of kernel or other smooth estimators
- Minimum distance
- Rearrangement methods
- ...

Model or observational setting?

- Censoring models and partial observation schemes

Model or observational setting?

- Censoring models and partial observation schemes
  - Current status data

Model or observational setting?

- Censoring models and partial observation schemes
  - Current status data
  - Interval censoring case 2, case k, mixed case

Model or observational setting?

- Censoring models and partial observation schemes
  - Current status data
  - Interval censoring case 2, case k, mixed case
  - Panel count data

Model or observational setting?

- Censoring models and partial observation schemes
  - Current status data
  - Interval censoring case 2, case k, mixed case
  - Panel count data
  - Competing risks

Model or observational setting?

- Censoring models and partial observation schemes
  - Current status data
  - Interval censoring case 2, case k, mixed case
  - Panel count data
  - Competing risks
- Semiparametric model? (heavy-tailed or unknown error distribution)

Model or observational setting?

- Censoring models and partial observation schemes
  - Current status data
  - Interval censoring case 2, case k, mixed case
  - Panel count data
  - Competing risks
- Semiparametric model? (heavy-tailed or unknown error distribution)
- Dependence??

Model or observational setting?

- Censoring models and partial observation schemes
  - Current status data
  - Interval censoring case 2, case k, mixed case
  - Panel count data
  - Competing risks
- Semiparametric model? (heavy-tailed or unknown error distribution)
- Dependence??
- $\ldots$

Model or observational setting?

- Censoring models and partial observation schemes
  - Current status data
  - Interval censoring case 2, case k, mixed case
  - Panel count data
  - Competing risks
- Semiparametric model? (heavy-tailed or unknown error distribution)
- Dependence??
- $\ldots$

Current state of shape restricted inference

A collection of several dozen results in search of a theory?

Main topics in my lecture:

- Maximum likelihood and least squares estimators

Current state of shape restricted inference

A collection of several dozen results in search of a theory?

Main topics in my lecture:

- Maximum likelihood and least squares estimators
- Pointwise limit theory

Current state of shape restricted inference

A collection of several dozen results in search of a theory?

Main topics in my lecture:

- Maximum likelihood and least squares estimators
- Pointwise limit theory
- Local (pointwise) lower bounds

Current state of shape restricted inference

A collection of several dozen results in search of a theory?

Main topics in my lecture:

- Maximum likelihood and least squares estimators
- Pointwise limit theory
- Local (pointwise) lower bounds
- Adaptation to “local smoothness” (or lack thereof).

Current state of shape restricted inference

A collection of several dozen results in search of a theory?

Main topics in my lecture:

- Maximum likelihood and least squares estimators
- Pointwise limit theory
- Local (pointwise) lower bounds
- Adaptation to “local smoothness” (or lack thereof).
- Some comparisons of maximum likelihood (and “canonical least squares”) estimators to rearrangement type estimators

1.1 An outline (or pattern) for pointwise limit theory

- **Step 0.** $X \sim P_\kappa$, $\kappa \in \mathcal{K}$, a set of shape-restricted functions

1.1 An outline (or pattern) for pointwise limit theory

- **Step 0.** $X \sim P_\kappa$, $\kappa \in \mathcal{K}$, a set of shape-restricted functions
- **Step 1.** Optimization criterion determining estimator:

1.1 An outline (or pattern) for pointwise limit theory

- **Step 0.** $X \sim P_\kappa$, $\kappa \in \mathcal{K}$, a set of shape-restricted functions
- **Step 1.** Optimization criterion determining estimator:
  - Log-likelihood.

1.1 An outline (or pattern) for pointwise limit theory

- **Step 0.** $X \sim P_\kappa$, $\kappa \in \mathcal{K}$, a set of shape-restricted functions
- **Step 1.** Optimization criterion determining estimator:
  - Log-likelihood.
  - Least squares contrast function.

1.1 An outline (or pattern) for pointwise limit theory

- **Step 0.** $X \sim P_\kappa$, $\kappa \in \mathcal{K}$, a set of shape-restricted functions
- **Step 1.** Optimization criterion determining estimator:
  - Log-likelihood.
  - Least squares contrast function.
- **Step 2.** Fenchel conditions characterizing the solution of the optimization problem

1.1 An outline (or pattern) for pointwise limit theory

- **Step 0.** $X \sim P_\kappa$, $\kappa \in \mathcal{K}$, a set of shape-restricted functions
- **Step 1.** Optimization criterion determining estimator:
  - Log-likelihood.
  - Least squares contrast function.
- **Step 2.** Fenchel conditions characterizing the solution of the optimization problem
- **Step 3.** Localization rate or tightness result
  empirical process theory: Kim-Pollard type lemmas.

1.1 An outline (or pattern) for pointwise limit theory

- **Step 0.** $X \sim P_\kappa$, $\kappa \in \mathcal{K}$, a set of shape-restricted functions
- **Step 1.** Optimization criterion determining estimator:
  - Log-likelihood.
  - Least squares contrast function.
- **Step 2.** Fenchel conditions characterizing the solution of the optimization problem
- **Step 3.** Localization rate or tightness result
  empirical process theory: Kim-Pollard type lemmas.
- **Step 4.** Localization of the Fenchel conditions.

1.1 An outline (or pattern) for pointwise limit theory

- **Step 0.** $X \sim P_\kappa$, $\kappa \in \mathcal{K}$, a set of shape-restricted functions
- **Step 1.** Optimization criterion determining estimator:
  - Log-likelihood.
  - Least squares contrast function.
- **Step 2.** Fenchel conditions characterizing the solution of the optimization problem
- **Step 3.** Localization rate or tightness result
  empirical process theory: Kim-Pollard type lemmas.
- **Step 4.** Localization of the Fenchel conditions.
- **Step 5.** Weak convergence of the (localized) driving process to a limit (Gaussian) driving process
  empirical process theory: CLT's with functions dependent on n.

- **Step 6.** Preservation of (localized) Fenchel relations in the limit.

- **Step 6.** Preservation of (localized) Fenchel relations in the limit.
- **Step 7.** Unique (Gaussian world) estimator resulting from localized limit processes and limit Fenchel relations

- **Step 6.** Preservation of (localized) Fenchel relations in the limit.
- **Step 7.** Unique (Gaussian world) estimator resulting from localized limit processes and limit Fenchel relations
- **Step 8** Cross-check/compare limiting result with local pointwise lower bound theory (Le Cam, Donoho & Liu, Groeneboom).

1.2 Illustration of the pattern: the Grenander estimator

Step 0. $X \sim f$ on $[0, \infty)$ with $f \searrow 0$.

Step 1. Optimization criterion: log-likelihood or least squares

$$\widehat{f}_n = \text{argmax}_{f \in \mathcal{M}_1} \left\{ \sum_{i=1}^n \log f(X_i) \right\} = \text{the MLE},$$

$$\widetilde{f}_n = \text{argmin}_{f \in \mathcal{K}_1} \psi_n(f) = \text{the LSE}$$

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

In this particular case, $\widehat{f}_n = \widetilde{f}_n$, i.e. LSE = MLE.

(This is not true in general.)

Step 2. Characterization: the Fenchel conditions

$$\mathbb{F}_n(x) \le \widehat{F}_n(x) \equiv \int_0^x \widehat{f}_n(t)dt \qquad \text{for all } x \in [0,\infty), \quad \text{and}$$

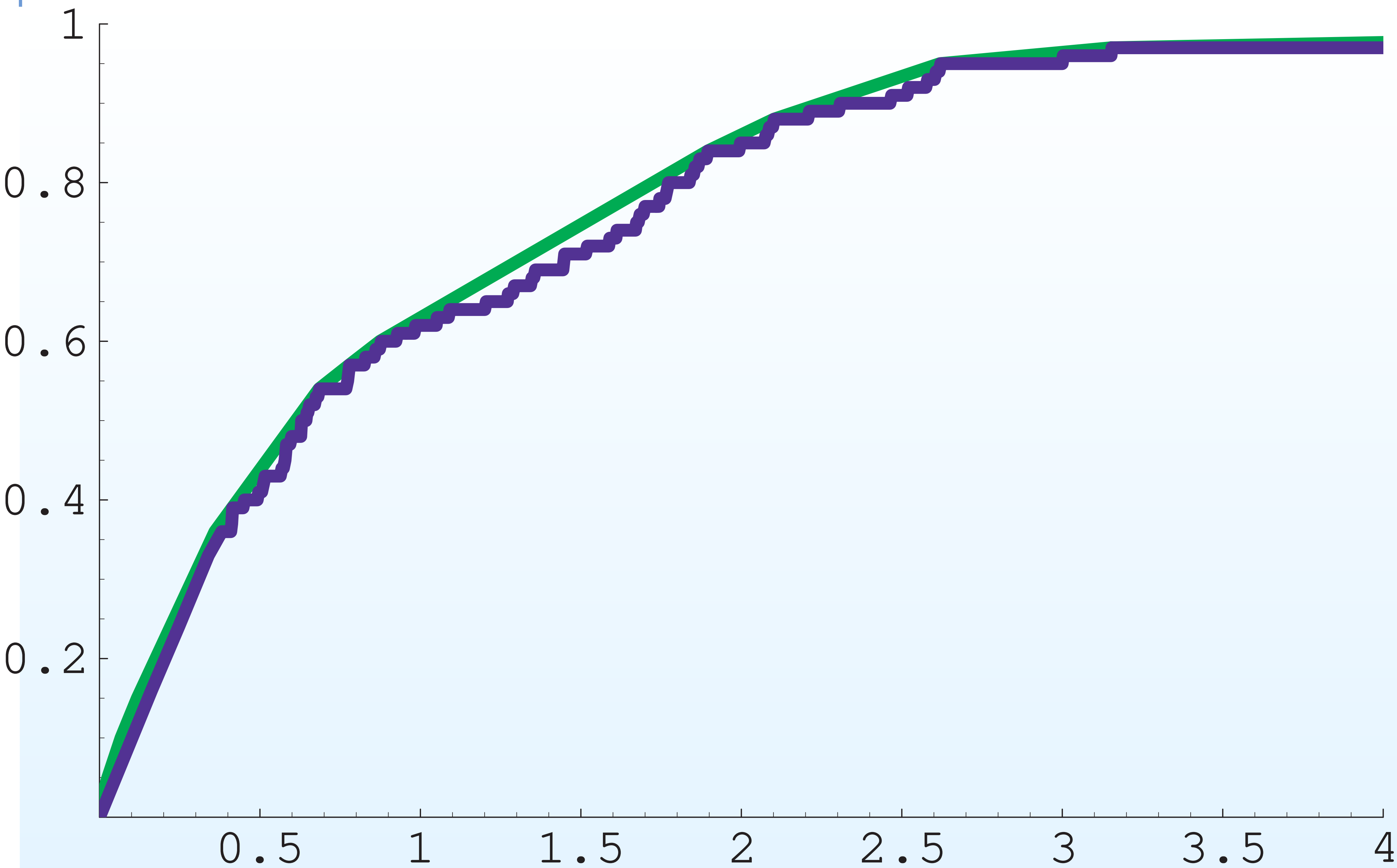
$$\mathbb{F}_n(x) = \widehat{F}_n(x) \quad \text{if and only if } \widehat{f}_n(x-) > \widehat{f}_n(x+).$$

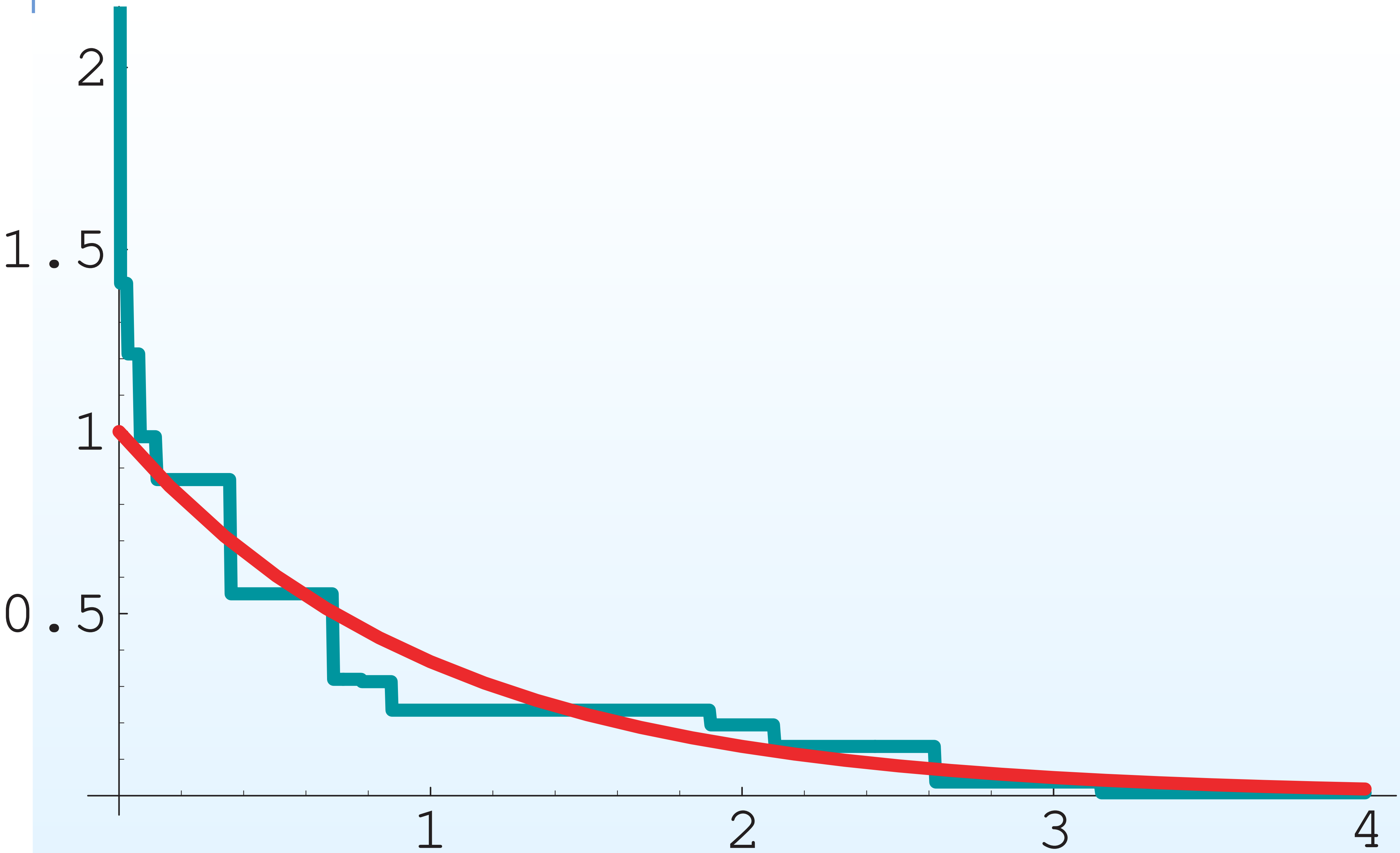
The second of these is equivalent to

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

The geometric interpretation of these two conditions is

$$\begin{aligned} \widehat{f}_n(x) &= \left\{ \begin{array}{l} \text{the left-derivative of the slope at } x \text{ of the} \\ \text{least concave majorant } \widehat{F}_n \text{ of } \mathbb{F}_n \end{array} \right\} \\ &\equiv \partial \mathcal{I}_1(\mathbb{F}_n) \\ &\equiv \text{Grenander estimator of } f. \end{aligned}$$


1
0.8
0.6
0.4
0.2
0.5
1
1.5
2
2.5
3
3.5
4


2
1.5
1
0.5
1
2
3
4

1. Illustration of the pattern via the Grenander estimator of a monotone density when:
    - **Case 1.** $f(x) = 1_{[0,1]}(x)$; uniform density (or degenerate mixing distribution)

1. Illustration of the pattern via the Grenander estimator of a monotone density when:
   - **Case 1.** $f(x) = 1_{[0,1]}(x)$; uniform density (or degenerate mixing distribution)
   - **Case 2.** At a point x_0 with $f'(x_0) < 0$

1. Illustration of the pattern via the Grenander estimator of a monotone density when:
    - **Case 1.** $f(x) = 1_{[0,1]}(x)$; uniform density (or degenerate mixing distribution)
    - **Case 2.** At a point x_0 with $f'(x_0) < 0$
    - **Case 3.** At a point x_0 with $f^{(j)}(x_0) = 0$, $j = 1, \ldots, k-1$, $f^{(k)}(x_0) \neq 0$.

1. Illustration of the pattern via the Grenander estimator of a monotone density when:
   - **Case 1.** $f(x) = 1_{[0,1]}(x)$; uniform density (or degenerate mixing distribution)
   - **Case 2.** At a point x_0 with $f'(x_0) < 0$
   - **Case 3.** At a point x_0 with $f^{(j)}(x_0) = 0$, $j = 1, \ldots, k-1$, $f^{(k)}(x_0) \neq 0$.
   - **Case 4.** At a point $x_0 \in (a, b)$ with $f(x)$ constant on (a, b).

1. Illustration of the pattern via the Grenander estimator of a monotone density when:
   - **Case 1.** $f(x) = 1_{[0,1]}(x)$; uniform density (or degenerate mixing distribution)
   - **Case 2.** At a point x_0 with $f'(x_0) < 0$
   - **Case 3.** At a point x_0 with $f^{(j)}(x_0) = 0$, $j = 1, \ldots, k-1$, $f^{(k)}(x_0) \neq 0$.
   - **Case 4.** At a point $x_0 \in (a, b)$ with $f(x)$ constant on (a, b).
   - **Case 5.** At a point x_0 where f is discontinuous.

1. Illustration of the pattern via the Grenander estimator of a monotone density when:
   - **Case 1.** $f(x) = 1_{[0,1]}(x)$; uniform density (or degenerate mixing distribution)
   - **Case 2.** At a point x_0 with $f'(x_0) < 0$
   - **Case 3.** At a point x_0 with $f^{(j)}(x_0) = 0$, $j = 1, \ldots, k-1$, $f^{(k)}(x_0) \neq 0$.
   - **Case 4.** At a point $x_0 \in (a,b)$ with $f(x)$ constant on (a,b).
   - **Case 5.** At a point x_0 where f is discontinuous.

1. Illustration of the pattern via the Grenander estimator of a monotone density when:
   - **Case 1.** $f(x) = 1_{[0,1]}(x)$; uniform density (or degenerate mixing distribution)
   - **Case 2.** At a point x_0 with $f'(x_0) < 0$
   - **Case 3.** At a point x_0 with $f^{(j)}(x_0) = 0$, $j = 1, \ldots, k-1$, $f^{(k)}(x_0) \neq 0$.
   - **Case 4.** At a point $x_0 \in (a, b)$ with $f(x)$ constant on (a, b).
   - **Case 5.** At a point x_0 where f is discontinuous.

Special feature:
Grenander and other monotone function problems.

Switching

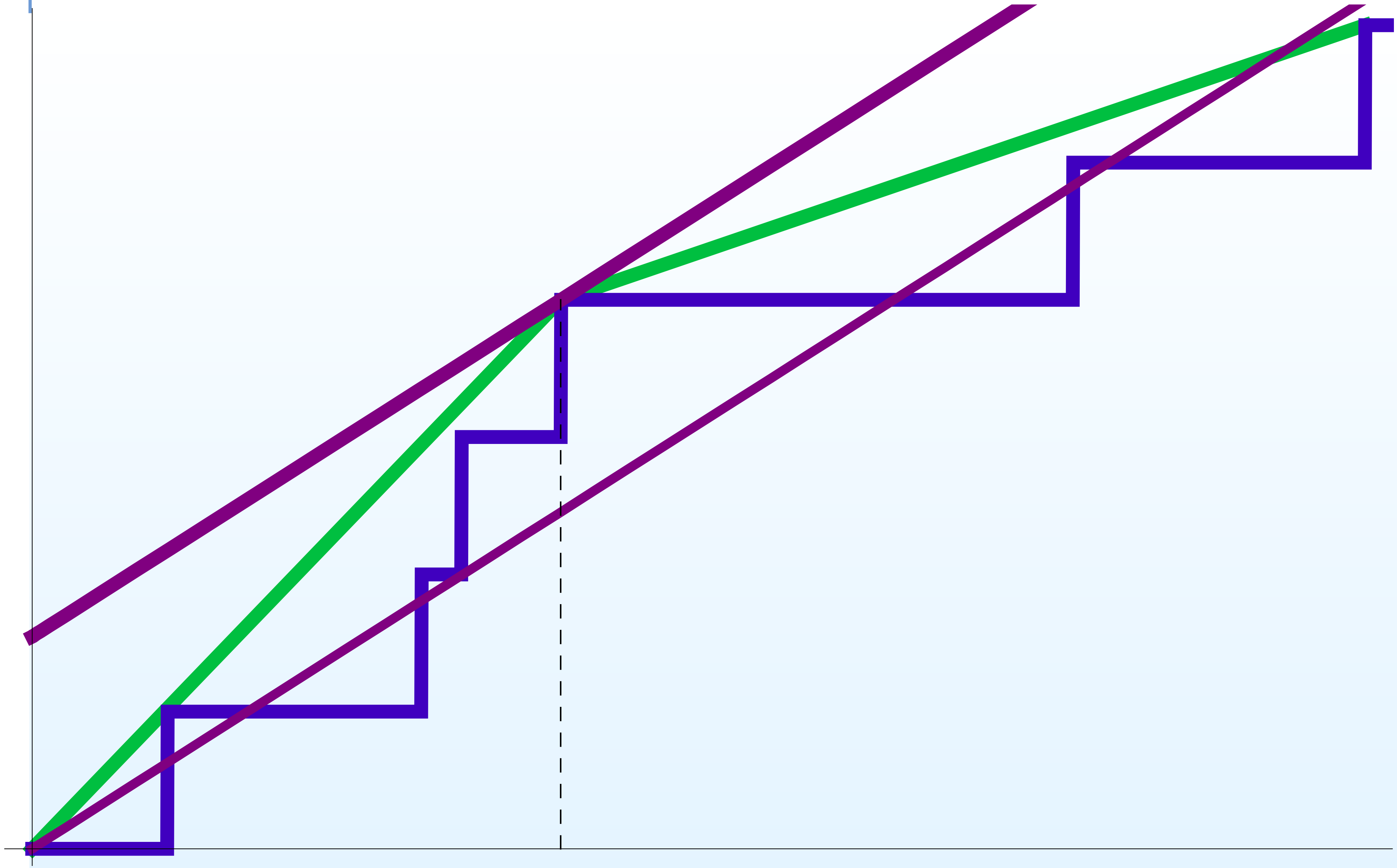
Let

$$\widehat{s}_n(a) \equiv \text{argmax}_s\{\mathbb{F}_n(s) - as\}, \qquad a > 0.$$

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

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

Warning: Nothing similar (yet?) for other shape constraints.



Steps 3-8 in Case 1. When f is the Uniform density on $[0,1]$, Groeneboom and Pyke (1983) show that for each $x_0 \in (0,1)$

$$\sqrt{n}(\widehat{f}_n(x_0) - f(x_0)) \to_d \mathbb{S}(x_0) = \partial \mathcal{I}_1(\mathbb{U})(x_0)$$

where $\mathbb{S}$ is the left derivative of the least concave majorant $\mathcal{I}_1(\mathbb{U}) = \mathbb{C}$ of a standard Brownian bridge process $\mathbb{U}$ on $[0,1]$.

- "Driving process" is $\mathbb{U}$.
- Process related to estimator maintaining Fenchel relations in the limit is $\mathbb{C}$ and its slope process $\mathbb{C}^{(1)} \equiv \mathbb{S}$:

$$\begin{aligned} \mathbb{C}(t) \ge \mathbb{U}(t) \quad & \text{for all } t \in (0,1), \\ \mathbb{C}(t) = \mathbb{U}(t) \quad & \text{if and only if } \mathbb{C}^{(1)}(t-) > \mathbb{C}^{(1)}(t+). \end{aligned}$$

- No localization in this case!
- From lower bound theory: $\hat{f}_n$ is (locally minimax) rate optimal; no estimator can achieve a better rate.

Steps 3-7 in Case 2. When f satisfies $f'(x_0) < 0$, $f(x_0) > 0$ and f' is continuous in a neighborhood of x_0, then Prakasa-Rao (1970) (see also Groeneboom (1985), Kim and Pollard (1990)) showed

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

where $\mathbb{S}(0) = \partial\mathcal{I}_1(Z)(0)$ is the slope at 0 of the least concave majorant of $Z(h) \equiv W(h) - h^2$ for a two-sided Brownian motion process W.

- "Driving process" is $\mathbb{Z}_{a,b}(h) \equiv \sqrt{f(x_0)}W(h) + f'(x_0)h^2 \equiv aW(h) - bh^2$.
- Process related to estimator maintaining Fenchel relations in the limit is $\mathbb{C}$ and its slope process $\mathbb{C}^{(1)} \equiv \mathbb{S}$:

$$\begin{aligned} \mathbb{C}(h) \geq \mathbb{Z}(h) \ &\text{ for all } h \in (-\infty, \infty), \\ \mathbb{C}(h) = \mathbb{Z}(h) \ &\text{ if and only if } \mathbb{C}^{(1)}(h-) > \mathbb{C}^{(1)}(h+). \end{aligned}$$

- Localization rate is $n^{-1/3}$

- From lower bound theory: $\hat{f}_n$ is (locally minimax) rate optimal in this scenario; no estimator can achieve a better minimax pointwise rate of convergence when $f'(x_0) < 0$.
- Moreover, the dependence of the limit distribution on f via $(|f'(x_0)f(x_0)|/2)^{1/3}$ is also optimal.
- For all the lower bound results noted here, see

  http://www.stat.washington.edu/jaw/RESEARCH/TALKS/MonAltHyp.pdf

under the entry for
Young European Statisticians Workshop (YES-I)
on Shape Restricted Inference

Steps 3-8 in Case 3. If $f^{(j)}(x_0) = 0$, $j = 1, \ldots, p-1$, $f^{(p)}(x_0) \neq 0$, then from the methods of Wright (1981) and Leurgans (1982),

$$n^{p/(2p+1)}(\widehat{f}_n(x_0) - f(x_0)) \to_d (f(x_0)^p A)^{1/(2p+1)} \mathbb{S}_p(0);$$

with $A = f^{(p)}(x_0)/(p+1)!$. Here $\mathbb{S}_p(0) = \partial\mathcal{I}_1(\mathbb{Z})(0)$ is the slope at 0 of the least concave majorant of $\mathbb{Z}(h) = W(h) - |h|^{p+1}$.

- "Driving process" is

$$\mathbb{Z}_p(h) \equiv \mathbb{Z}_{p,a,b}(h) \equiv \sqrt{f(x_0)}W(h) - A|h|^{p+1} \equiv aW(h) - b|h|^{p+1}.$$

- Process related to estimator maintaining Fenchel relations in the limit is $\mathbb{C}_p \equiv \mathcal{I}_1(\mathbb{Z}_p)$ and its slope process $\mathbb{C}_p^{(1)} \equiv \mathbb{S}_p \partial\mathcal{I}_1(\mathbb{Z}_p)$:

$$\mathbb{C}_p(h) \geq \mathbb{Z}_p(h) \;\text{ for all } h \in (-\infty, \infty),$$
$$\mathbb{C}_p(h) = \mathbb{Z}_p(h) \;\text{ if and only if } \mathbb{C}_p^{(1)}(h-) > \mathbb{C}_p^{(1)}(h+).$$

- Localization rate is $n^{-1/(2p+1)}$
- From lower bound theory: $\hat{f}_n$ is (locally minimax) rate optimal in this scenario; no estimator can achieve a better minimax pointwise rate of convergence when $f^{(j)}(x_0) = 0$, $j = 1, \ldots, p-1$, $f^{(p)}(x_0) \neq 0$.
- Moreover, the dependence of the limit distribution on f via $(|f^{(p)}(x_0)f(x_0)^p)^{1/(2p+1)}$ is also optimal.

Steps 3-8 in Case 4. If $x_0 \in (a,b)$ with $f(x)$ constant on (a,b), then Carolan and Dykstra (1999) showed that

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

where $p \equiv f(x_0)(b-a) = F(b) - F(a)$, $Z \sim N(0,1)$, $\mathbb{S}$ is the process of slopes of a Brownian bridge process $\mathbb{U}$ as in case 1, and Z and $\mathbb{S}$ are independent.
This is much as in case 1, but with a twist or two.

- "Driving process" is $\mathbb{Z}(h) \equiv \mathbb{U}(F(a+h)) - \mathbb{U}(F(a))$.
- Process related to estimator maintaining Fenchel relations in the limit is $\mathbb{C}_{loc} \equiv \mathcal{I}_1(\mathbb{Z})$ and its slope process $\mathbb{C}_{loc}^{(1)} \equiv \mathbb{S}_{loc} \equiv \partial \mathcal{I}_1(\mathbb{Z})$:

$$\mathbb{C}_{loc}(h) \geq \mathbb{Z}(h) \ \text{ for all } h \in [0, b-a],$$

$$\mathbb{C}_{loc}(h) = \mathbb{Z}(h) \ \text{ if and only if } \mathbb{C}_{loc}^{(1)}(h-) > \mathbb{C}_{loc}^{(1)}(h+).$$

- Localization only to the interval $[a, b]$.
- From lower bound theory: $\hat{f}_n$ is (locally minimax) rate optimal in this scenario; no estimator can achieve a better minimax pointwise rate of convergence when f is flat in a neighborhood of x_0.

Steps 3-8 in Case 5. If f is discontinuous at x_0, then Anevski and Hössjer (2002) show that

$$P(\widehat{f}_n(x_0)-\bar{f}(x_0) \le x) \to P(\text{argmax}\{\mathbb{N}_0(h)-\rho_{x+d/2,x-d/2}(h)\} \le 0)$$

where $\mathbb{N}_0$ is a two-sided, centered Poisson process with rates $f(x_0+)$ and $f(x_0-)$ to the right and left of 0 respectively,

$$\rho_{B,C}(h) \equiv \left\{ \begin{array}{ll} Bh, & h \ge 0 \\ -Ch, & h < 0. \end{array} \right\},$$

$\bar{f}(x_0) \equiv (f(x_0+) + f(x_0-))/2$, $d \equiv f(x_0-) - f(x_0+)$.
Furthermore, by switching again in the limit (Poisson) problem,

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

where $\mathbb{R}(h)$ is the process of slopes (left derivatives) of the least concave majorant of the process

$$\mathbb{M}(h) \equiv \mathbb{N}_0(h) - (d/2)|h|.$$

- "Driving process" is $\mathbb{M}(h) \equiv \mathbb{N}_0(h) - (d/2)|h|$.
- Process related to estimator maintaining Fenchel relations in the limit is $\mathbb{K}$ and its slope process $\mathbb{K}^{(1)} \equiv \mathbb{R}$:

$$\begin{aligned} \mathbb{K}(h) &\geq \mathbb{M}(h) \ \text{ for all } h \in R, \\ \mathbb{K}(h) &= \mathbb{M}(h) \ \text{ if and only if } \mathbb{K}^{(1)}(h-) > \mathbb{K}^{(1)}(h+). \end{aligned}$$

- Localization rate is n^{-1}!

2. Illustration of the pattern:

the MLE of a convex decreasing density

Step 0. $X \sim f$ on $[0, \infty)$ with $f \searrow 0$, f convex.

$$f(x) = \int_0^\infty \frac{2}{y^2}(y-x)_+ dG(y), \quad G \text{ a distribution function}$$

Step 1. Optimization criterion: log-likelihood or least squares

$$\widehat{f}_n = \text{argmax}_{f \in \mathcal{M}_2} \left\{ \sum_{i=1}^n \log f(X_i) \right\} = \text{the MLE},$$

$$\widetilde{f}_n = \text{argmin}_{f \in \mathcal{K}_2} \psi_n(f) = \text{the LSE}$$

where $\psi_n(f) \equiv \frac{1}{2} \int_0^\infty f^2(x)dx - \int_0^\infty f(x)d\mathbb{F}_n(x)$.
In this case, $\widehat{f}_n \neq \widetilde{f}_n$, i.e. LSE $\neq$ MLE.

Step 2. Characterization: the Fenchel conditions for $\widetilde{f}_n$:
let

$$\widetilde{H}_n(x) \equiv \int_0^x \int_0^y \widetilde{f}_n(t)\,dt\,dy \qquad \text{for all } x \in [0,\infty), \quad \text{and}$$

$$\mathbb{Y}_n(x) = \int_0^x \mathbb{F}_n(y)\,dy$$

Then $\widetilde{f}_n \in \mathcal{K}$ is the LSE if and only if

$$\widetilde{H}_n(x) \geq \mathbb{Y}_n(x) \qquad \text{for all } x > 0,$$

$$\int_0^\infty (\widetilde{H}_n(x) - \mathbb{Y}_n(x))\,d\widetilde{H}_n^{(3)}(x) = 0,$$

$\widetilde{H}_n$ has convex second derivative $\widetilde{f}_n$.

Step 3. Localization rate / tightness

Proposition. Let x_0 be an interior point of the support of f. For $0 < x \le y$, define $U_n(x, y)$ by

$$U_n(x, y) \equiv \int_{[x,y]} \{z - (x + y)/2\} d(\mathbb{F}_n - F)(y).$$

Then there exist $\delta > 0$ and $c_0 > 0$ so that, for each $\epsilon > 0$ and x with $|x - x_0| < \delta$,

$$|U_n(x, y)| \le \epsilon(y - x)^4 + O_p(n^{-4/5}), \qquad 0 \le y - x_0 \le c_0.$$

Proposition. Let x_0 and f satisfy $f''(x_0) > 0$ and f'' continuous at x_0. Let $\xi_n \to x_0$, and let

$$\tau_n^- \equiv \max\{t \le \xi_n : \ \widetilde{f}_n^{(1)} \text{discontinuous at } t\} \quad \tau_n^+ \equiv \min\{t > \xi_n : \widetilde{f}_n^{(1)} \text{disco}$$

Then $\tau_n^+ - \tau_n^- = O_p(n^{-1/5})$.

Proposition. Suppose $f'(x_0) < 0$, $f''(x_0) > 0$ and f'' continuous in a nbhd. of x_0. Then

$$\sup_{|t| \le M} |\widetilde{f}(x_0 + n^{-1/5}t) - f_0(x_0) - n^{-1/5} t f'(x_0)| = O_p(n^{-2/5}),$$

and

$$\sup_{|t| \le M} |\widetilde{f}'(x_0 + n^{-1/5}t) - f'(x_0)| = O_p(n^{-1/5}).$$

Step 4. Localize the Fenchel relations: define

$$\begin{aligned} \mathbb{Y}_n^{loc}(t) \quad &\equiv \quad n^{4/5} \int_{x_0}^{x_0 + n^{-1/5}t} \{\mathbb{F}_n(v) - \mathbb{F}_n(x_0) \\ &\qquad\qquad + \int_{x_0}^{v} (f(x_0) + (u - x_0) f(x_0) du \Big\} dv, \end{aligned}$$

$$\begin{aligned}\widetilde{H}_n^{loc}(t) &\equiv n^{4/5}\int_{x_0}^{x_0+n^{-1/5}t}\int_{x_0}^{v}\{\widetilde{f}_n(u)-f(x_0)-(u-x_0)f'(x_0)\}du dv\\ &\quad +\widetilde{A}_n t+\widetilde{B}_n.\end{aligned}$$

Then

$$\widetilde{H}_n^{loc}(t)\geq \mathbb{Y}_n^{loc}(t)$$

with equality if and only if $x_0+n^{-1/5}t$ is a jump point of $\widetilde{H}_n^{(3)}$.

Note that

$$\begin{aligned}(\widetilde{H}_n^{loc})^{(2)}(t) &= n^{2/5}(\widetilde{f}_n(x_0+n^{-1/5}t)-f(x_0)-n^{-1/5}tf'(x_0)),\\ (\widetilde{H}_n^{loc})^{(3)}(t) &= n^{1/5}(\tilde{f}_n'(x_0+n^{-1/5}t)-f'(x_0)).\end{aligned}$$

Step 5. Weak convergence of the (localized) driving process $\mathbb{Y}_n$ to a limit (Gaussian) driving process

$$
\begin{aligned}
&\mathbb{Y}_n^{loc}(t) \\
&\quad \stackrel{d}{=} \quad n^{3/10} \int_{x_0}^{x_0 + n^{-1/5} t} \{\mathbb{U}_n(F_0(v)) - \mathbb{U}_n(F(x_0))\} dv + \frac{1}{24} f''(x_0) t^4 + o(1) \\
&\quad \Rightarrow \quad \sqrt{f(x_0)} \int_0^t W(s) ds + \frac{1}{24} f''(x_0) t^4
\end{aligned}
$$

by KMT or theorems 2.11.22 or 2.11.23, VdV & W (1996)

$$
\begin{aligned}
&\quad = \quad a \int_0^t W(s) ds + \sigma t^4 \\
&\quad \equiv \quad \mathbb{Y}(t) \equiv \mathbb{Y}_{a,\sigma}(t)
\end{aligned}
$$

where $\mathbb{U}_n(t) \equiv \sqrt{n}(\mathbb{G}_n(t) - t)$ is the empirical process of $\xi_1, \ldots, \xi_n$ i.i.d. Uniform$(0,1)$, $a \equiv \sqrt{f(x_0}$, $\sigma \equiv f''(x_0)/24$.

Step 6. Preservation of (localized) Fenchel relations in the limit.

- $\{(\widetilde{H}_n^{loc}, \widetilde{H}_n^{loc,(1)}, \widetilde{H}_n^{loc,(2)}, \widetilde{H}_n^{loc,(3)})\}_{n\geq 1}$ is tight.

Step 6. Preservation of (localized) Fenchel relations in the limit.

- $\{(\widetilde{H}_n^{loc}, \widetilde{H}_n^{loc,(1)}, \widetilde{H}_n^{loc,(2)}, \widetilde{H}_n^{loc,(3)})\}_{n\geq 1}$ is tight.
- $\mathbb{Y}_n^{loc} \Rightarrow \mathbb{Y}$

Step 6. Preservation of (localized) Fenchel relations in the limit.

- $\{(\widetilde{H}_n^{loc}, \widetilde{H}_n^{loc,(1)}, \widetilde{H}_n^{loc,(2)}, \widetilde{H}_n^{loc,(3)})\}_{n\geq 1}$ is tight.
- $\mathbb{Y}_n^{loc} \Rightarrow \mathbb{Y}$
- Fenchel relations satisfied:

Step 6. Preservation of (localized) Fenchel relations in the limit.

- $\{(\widetilde{H}_n^{loc}, \widetilde{H}_n^{loc,(1)}, \widetilde{H}_n^{loc,(2)}, \widetilde{H}_n^{loc,(3)})\}_{n\geq 1}$ is tight.
- $\mathbb{Y}_n^{loc} \Rightarrow \mathbb{Y}$
- Fenchel relations satisfied:
  - $\widetilde{H}_n^{loc}(x) \geq \mathbb{Y}_n^{loc}(x)$ for all x

Step 6. Preservation of (localized) Fenchel relations in the limit.

- $\{(\widetilde{H}_n^{loc}, \widetilde{H}_n^{loc,(1)}, \widetilde{H}_n^{loc,(2)}, \widetilde{H}_n^{loc,(3)})\}_{n\geq 1}$ is tight.
- $\mathbb{Y}_n^{loc} \Rightarrow \mathbb{Y}$
- Fenchel relations satisfied:
  - $\widetilde{H}_n^{loc}(x) \geq \mathbb{Y}_n^{loc}(x)$ for all x
  - $\int_{-\infty}^{\infty}(\widetilde{H}_n^{loc}(x) - \mathbb{Y}_n^{loc}(x))d\widetilde{H}_n^{loc,(3)}(x) = 0.$

Step 6. Preservation of (localized) Fenchel relations in the limit.

- $\{(\widetilde{H}_n^{loc}, \widetilde{H}_n^{loc,(1)}, \widetilde{H}_n^{loc,(2)}, \widetilde{H}_n^{loc,(3)})\}_{n\geq 1}$ is tight.
- $\mathbb{Y}_n^{loc} \Rightarrow \mathbb{Y}$
- Fenchel relations satisfied:
  - $\widetilde{H}_n^{loc}(x) \geq \mathbb{Y}_n^{loc}(x)$ for all x
  - $\int_{-\infty}^{\infty} (\widetilde{H}_n^{loc}(x) - \mathbb{Y}_n^{loc}(x)) d\widetilde{H}_n^{loc,(3)}(x) = 0$.
- Any limit process H for a subsequence $\{\widetilde{H}_{n'}^{loc}\}$ must satisfy

Step 6. Preservation of (localized) Fenchel relations in the limit.

- $\{(\widetilde{H}_n^{loc}, \widetilde{H}_n^{loc,(1)}, \widetilde{H}_n^{loc,(2)}, \widetilde{H}_n^{loc,(3)})\}_{n\geq 1}$ is tight.
- $\mathbb{Y}_n^{loc} \Rightarrow \mathbb{Y}$
- Fenchel relations satisfied:
  - $\widetilde{H}_n^{loc}(x) \geq \mathbb{Y}_n^{loc}(x)$ for all x
  - $\int_{-\infty}^{\infty}(\widetilde{H}_n^{loc}(x) - \mathbb{Y}_n^{loc}(x))d\widetilde{H}_n^{loc,(3)}(x) = 0$.
- Any limit process H for a subsequence $\{\widetilde{H}_{n'}^{loc}\}$ must satisfy
  - $H(x) \geq \mathbb{Y}(x)$ for all x.

Step 6. Preservation of (localized) Fenchel relations in the limit.

- $\{(\widetilde{H}_n^{loc}, \widetilde{H}_n^{loc,(1)}, \widetilde{H}_n^{loc,(2)}, \widetilde{H}_n^{loc,(3)})\}_{n\geq 1}$ is tight.
- $\mathbb{Y}_n^{loc} \Rightarrow \mathbb{Y}$
- Fenchel relations satisfied:
  - $\widetilde{H}_n^{loc}(x) \geq \mathbb{Y}_n^{loc}(x)$ for all x
  - $\int_{-\infty}^{\infty} (\widetilde{H}_n^{loc}(x) - \mathbb{Y}_n^{loc}(x)) d\widetilde{H}_n^{loc,(3)}(x) = 0$.
- Any limit process H for a subsequence $\{\widetilde{H}_{n'}^{loc}\}$ must satisfy
  - $H(x) \geq \mathbb{Y}(x)$ for all x.
  - $\int_{-\infty}^{\infty} (H(x) - \mathbb{Y}(x)) dH^{(3)}(x) = 0$.

Step 6. Preservation of (localized) Fenchel relations in the limit.

- $\{(\widetilde{H}_n^{loc}, \widetilde{H}_n^{loc,(1)}, \widetilde{H}_n^{loc,(2)}, \widetilde{H}_n^{loc,(3)})\}_{n\geq 1}$ is tight.
- $\mathbb{Y}_n^{loc} \Rightarrow \mathbb{Y}$
- Fenchel relations satisfied:
  - $\widetilde{H}_n^{loc}(x) \geq \mathbb{Y}_n^{loc}(x)$ for all x
  - $\int_{-\infty}^{\infty} (\widetilde{H}_n^{loc}(x) - \mathbb{Y}_n^{loc}(x)) d\widetilde{H}_n^{loc,(3)}(x) = 0$.
- Any limit process H for a subsequence $\{\widetilde{H}_{n'}^{loc}\}$ must satisfy
  - $H(x) \geq \mathbb{Y}(x)$ for all x.
  - $\int_{-\infty}^{\infty} (H(x) - \mathbb{Y}(x)) dH^{(3)}(x) = 0$.
  - $H^{(2)}$ is convex.

Step 6. Preservation of (localized) Fenchel relations in the limit.

- $\{(\widetilde{H}_n^{loc}, \widetilde{H}_n^{loc,(1)}, \widetilde{H}_n^{loc,(2)}, \widetilde{H}_n^{loc,(3)})\}_{n\geq 1}$ is tight.
- $\mathbb{Y}_n^{loc} \Rightarrow \mathbb{Y}$
- Fenchel relations satisfied:
  - $\widetilde{H}_n^{loc}(x) \geq \mathbb{Y}_n^{loc}(x)$ for all x
  - $\int_{-\infty}^{\infty}(\widetilde{H}_n^{loc}(x) - \mathbb{Y}_n^{loc}(x))d\widetilde{H}_n^{loc,(3)}(x) = 0$.
- Any limit process H for a subsequence $\{\widetilde{H}_{n'}^{loc}\}$ must satisfy
  - $H(x) \geq \mathbb{Y}(x)$ for all x.
  - $\int_{-\infty}^{\infty}(H(x) - \mathbb{Y}(x))dH^{(3)}(x) = 0$.
  - $H^{(2)}$ is convex.
- Is there a unique such process $H = H_{a,\sigma}$? If so, done!

Step 7. Unique (Gaussian world) estimator resulting from limit Fenchel relations! (Proof: suppose there are two such processes, H_1 and H_2. Then GJW (2001) showed $H_1 = H_2 \equiv H$.)

Upshot: after rescaling to universal ($a = 1$, $\sigma = 1$) limit:

Theorem. If $f \in \mathcal{C}$, $f(x_0) > 0$, $f''(x_0) > 0$, and f'' continuous in a neighborhood of x_0, then

$$\begin{pmatrix} n^{2/5}(\widetilde{f}_n(x_0) - f(x_0)) \\ n^{1/5}(\widetilde{f}'_n(x_0) - f'(x_0)) \end{pmatrix} \to_d \begin{pmatrix} c_1(f) H^{(2)}(0) \\ c_2(f) H^{(3)}(0) \end{pmatrix}$$

where

$$c_1(f) \equiv \left(\frac{f^2(x_0) f''(x_0)}{24} \right)^{1/5}, \qquad c_2(f) \equiv \left(\frac{f(x_0) f''(x_0)^3}{24^3} \right)^{1/5}.$$

Step 8 (or $0'$). Cross-check/compare limiting result with local pointwise lower bound theory.
Use Groeneboom's lower bound lemma (relative of results of Donoho & Liu, Le Cam).

Define f_ϵ by renormalizing (or linearly correcting) $\tilde{f}_\epsilon$ defined by

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

where c_ϵ is chosen so that $\tilde{f}_\epsilon$ is continuous at $x_0 - \epsilon$. Let P_n be defined by $f_{\epsilon_n} \equiv f_{\nu n^{-1/5}}$ where

$$\nu \equiv \frac{2 f''(x_0)^2}{5 f(x_0)}.$$

Proposition. If $f(x_0) > 0$, $f''(x_0) > 0$, and f'' is continuous in a neighborhood of x_0, for any estimators T_n of $f(x_0)$ and any estimators $\widetilde{T}_n$ of $f'(x_0)$,

$$n^{2/5} \inf_{T_n} \max \{E_{n,P_n}|T_n - f_{\epsilon_n}(x_0)|, E_{n,P}|T_n - f(x_0)|\}$$

$$\geq \frac{1}{4} \left(\frac{3}{e\sqrt{2}} \right)^{1/5} \cdot c_1(f),$$

$$n^{1/5} \inf_{T_n} \max \left\{ E_{n,P_n}|\widetilde{T}_n - f'_{\epsilon_n}(x_0)|, E_{n,P}|\widetilde{T}_n - f'(x_0)| \right\}$$

$$\geq \frac{1}{4} \left(\frac{6 \cdot 24^2}{e} \right)^{1/5} \cdot c_2(f)$$

The following pages show: (from Groeneboom, Jongbloed, and Wellner (2001))

- the "invelope process" H, and the driving process Y

The following pages show: (from Groeneboom, Jongbloed, and Wellner (2001))

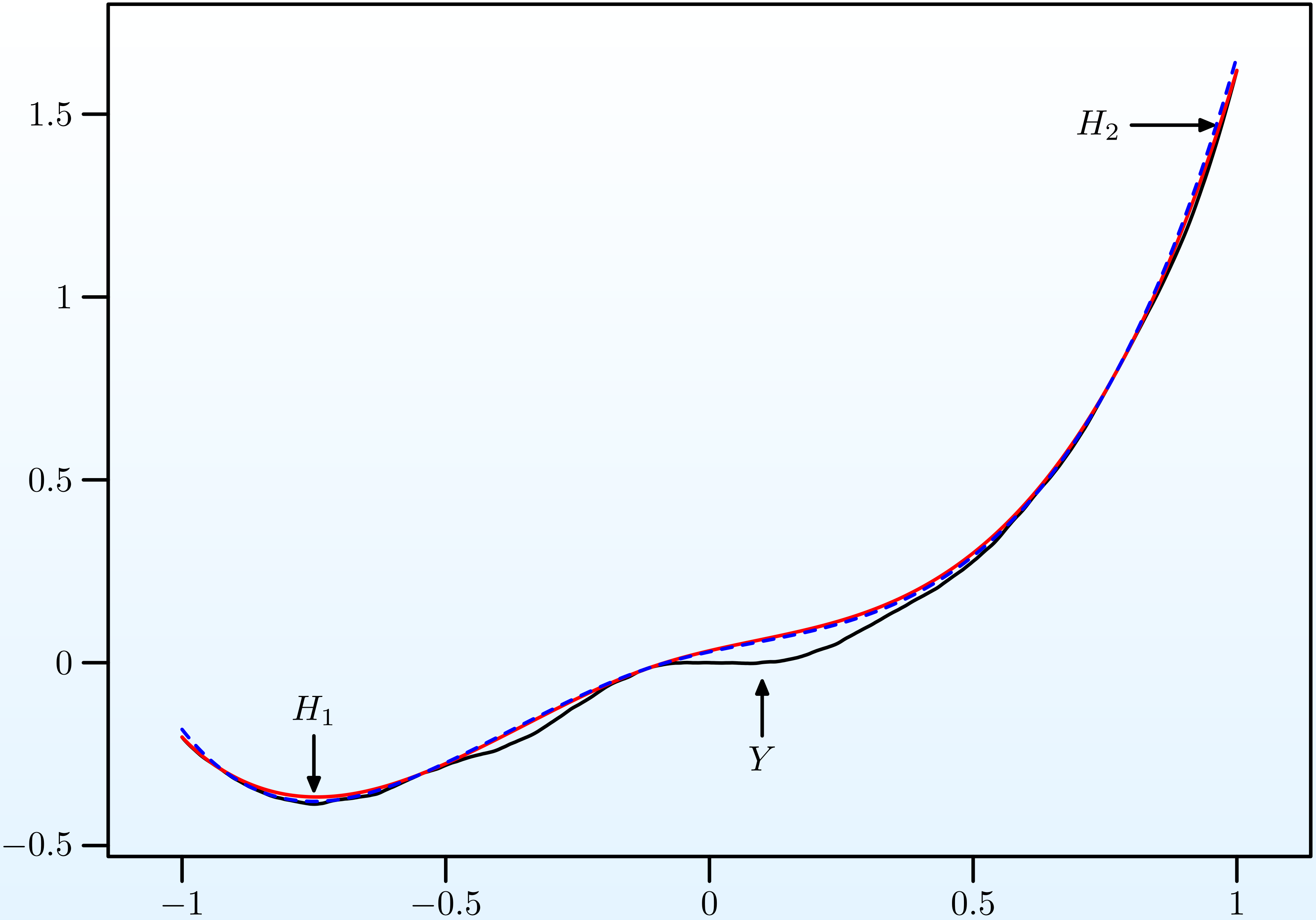
- the “invelope process” H, and the driving process Y
- the derivative process $H^{(1)}$, and the process $Y^{(1)}$

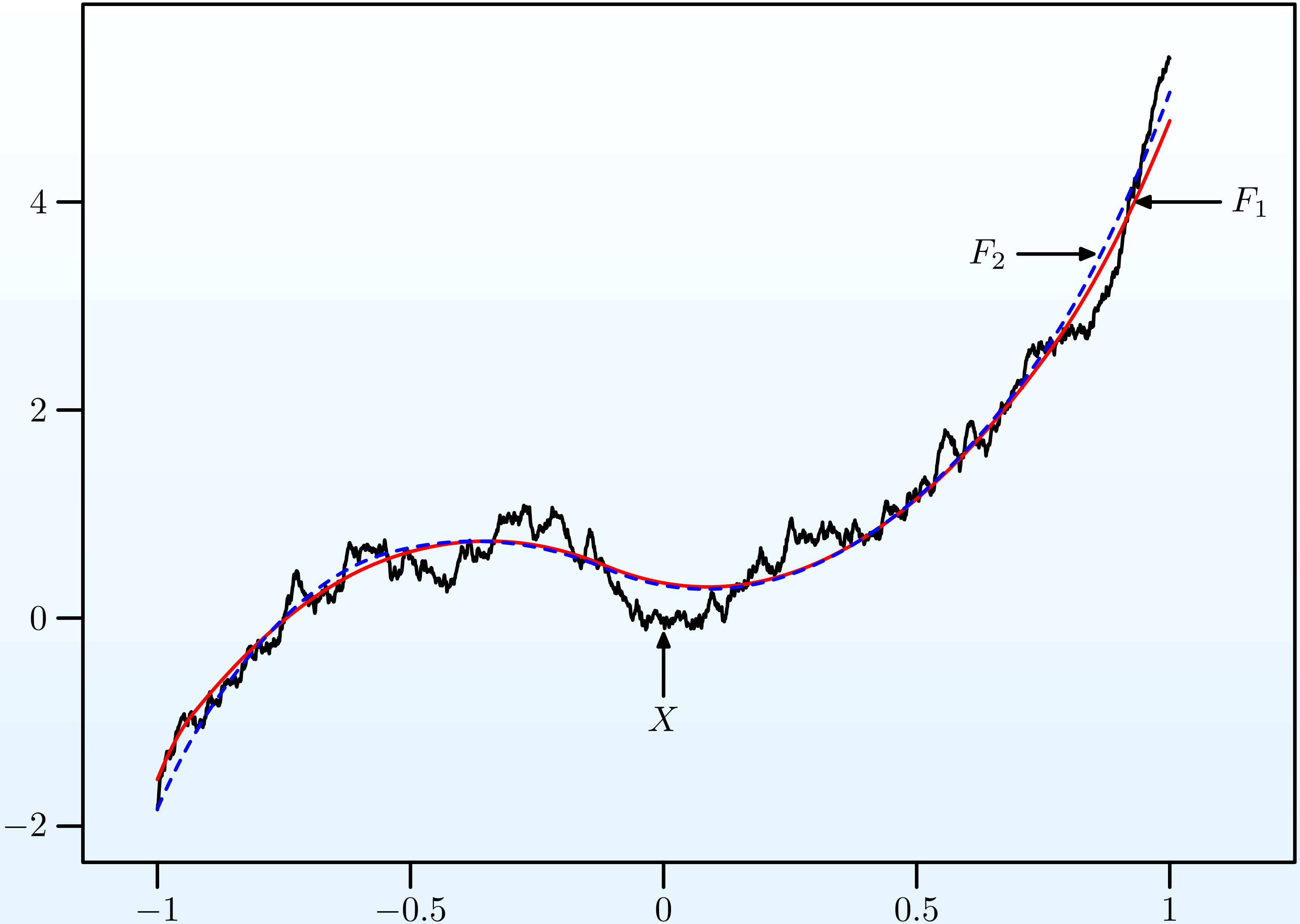
The following pages show: (from Groeneboom, Jongbloed, and Wellner (2001))

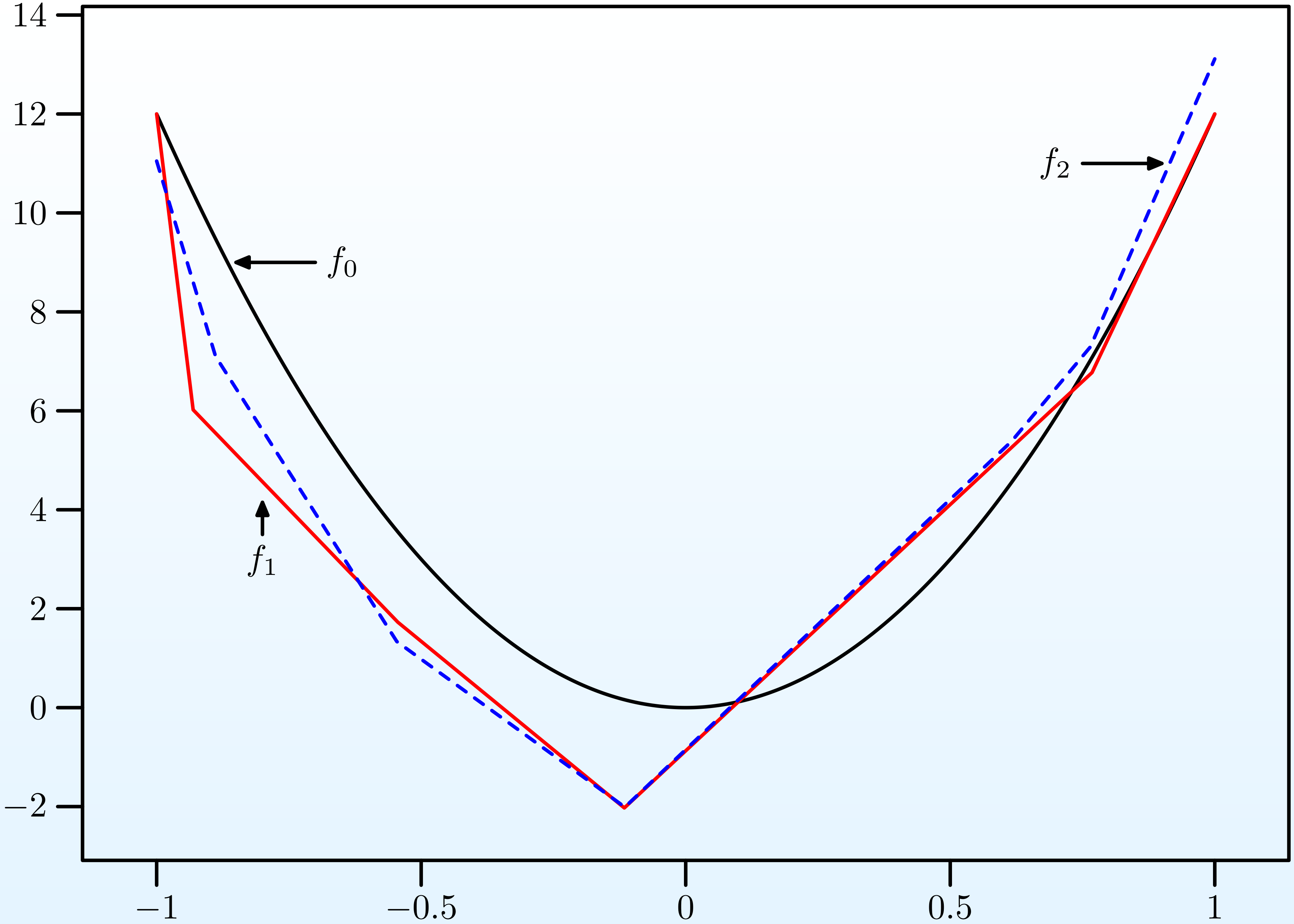
- the "invelope process" H, and the driving process Y
- the derivative process $H^{(1)}$, and the process $Y^{(1)}$
- the concave (limit world estimator of $12t^2$) process $H^{(2)}$

The following pages show: (from Groeneboom, Jongbloed, and Wellner (2001))

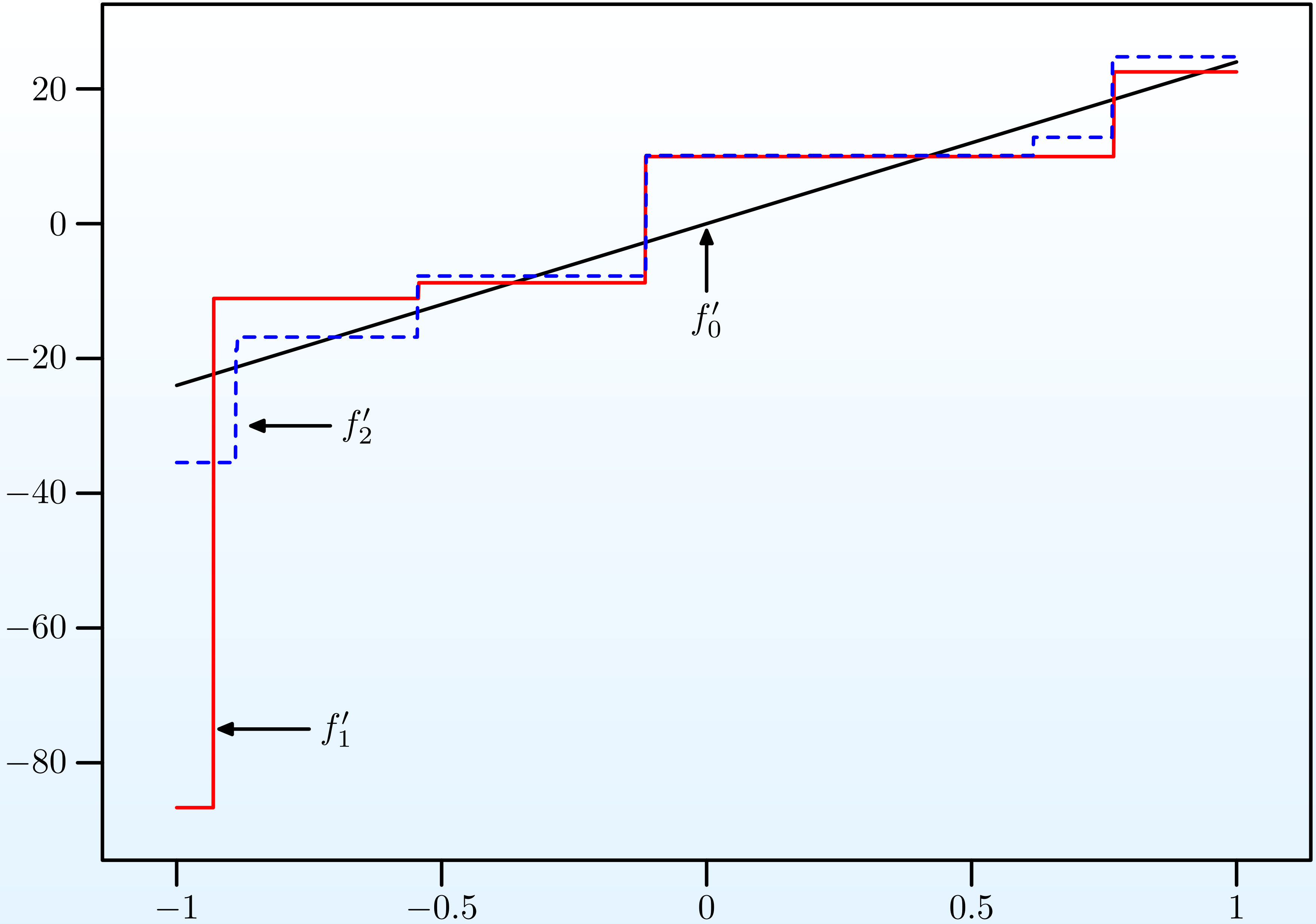
- the "invelope process" H, and the driving process Y
- the derivative process $H^{(1)}$, and the process $Y^{(1)}$
- the concave (limit world estimator of $12t^2$) process $H^{(2)}$
- the piecewise (limit world estimator of $24t$) process $H^{(3)}$


1.5
1
0.5
0
−0.5
−1
−0.5
0
0.5
1
H_2
H_1
Y


4
2
0
−2
−1
−0.5
0
0.5
1
F_1
F_2
X



14
12
10
8
6
4
2
0
−2
−1
−0.5
0
0.5
1
f_0
f_1
f_2


20
0
−20
−40
−60
−80
−1
−0.5
0
0.5
1
f'_0
f'_1
f'_2

3. Some Comparisons: MLE / LSE

versus Rearrangements

Monotone

- Monotone rearrangement, continuous case:
  $f^{mon-rearr} \equiv R(f)$ where

$$Z_f(s) = \lambda\{x : \ f(x) \ge s\}, \qquad R(f)(x) = Z_f^{-1}(x).$$

3. Some Comparisons: MLE / LSE

versus Rearrangements

Monotone

- Monotone rearrangement, continuous case: $f^{mon-rearr} \equiv R(f)$ where

$$Z_f(s) = \lambda\{x : \ f(x) \geq s\}, \qquad R(f)(x) = Z_f^{-1}(x).$$

- Monotone rearrangement, discrete case: $f^{mon-rearr} \equiv R(f)$ where

$$Z_f(s) = \#\{i \in \mathbb{Z}^+ : \ f(i) \geq s\}, \qquad R(f)(i) = Z_f^{-1}(i).$$

- Monotone Least Squares, continuous case: (Mammen)

$$f^{LSE} \equiv LS(f) = \partial \mathcal{I}_1 \left(\int_0^{\cdot} f du \right)$$

where $\mathcal{I}_1$ = Least Concave Majorant operator.

- Monotone Least Squares, continuous case: (Mammen)

$$f^{LSE} \equiv LS(f) = \partial \mathcal{I}_1 \left(\int_0^{\cdot} f du \right)$$

where $\mathcal{I}_1 =$ Least Concave Majorant operator.

- Empirical (or canonical Least Squares, continuous case:

$$f^{LSE-empirical} = f^{MLE} = \partial \mathcal{I}_1(F).$$

- Monotone Least Squares, continuous case: (Mammen)

$$f^{LSE} \equiv LS(f) = \partial \mathcal{I}_1 \left(\int_0^{\cdot} f\, du \right)$$

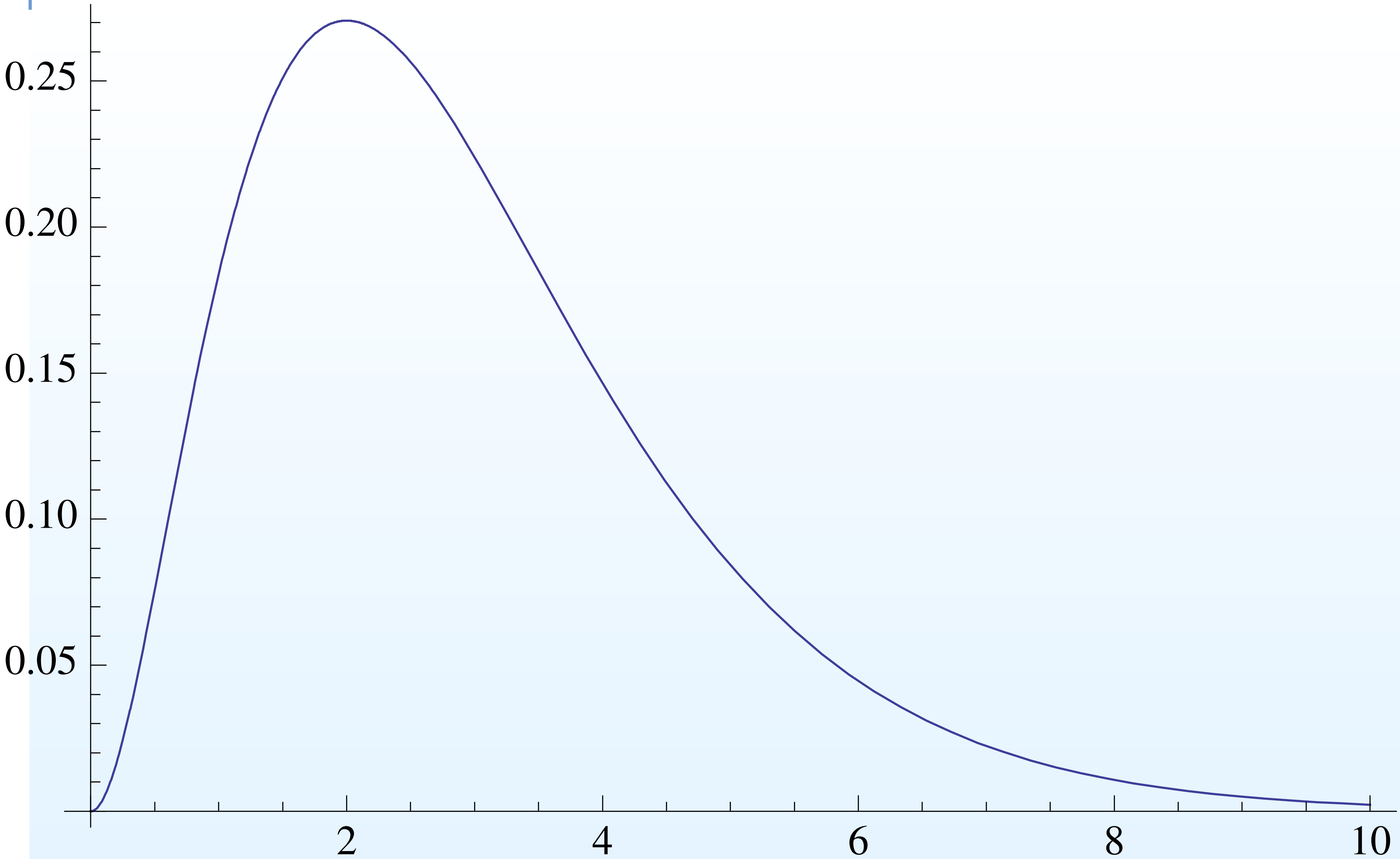
where $\mathcal{I}_1 =$ Least Concave Majorant operator.

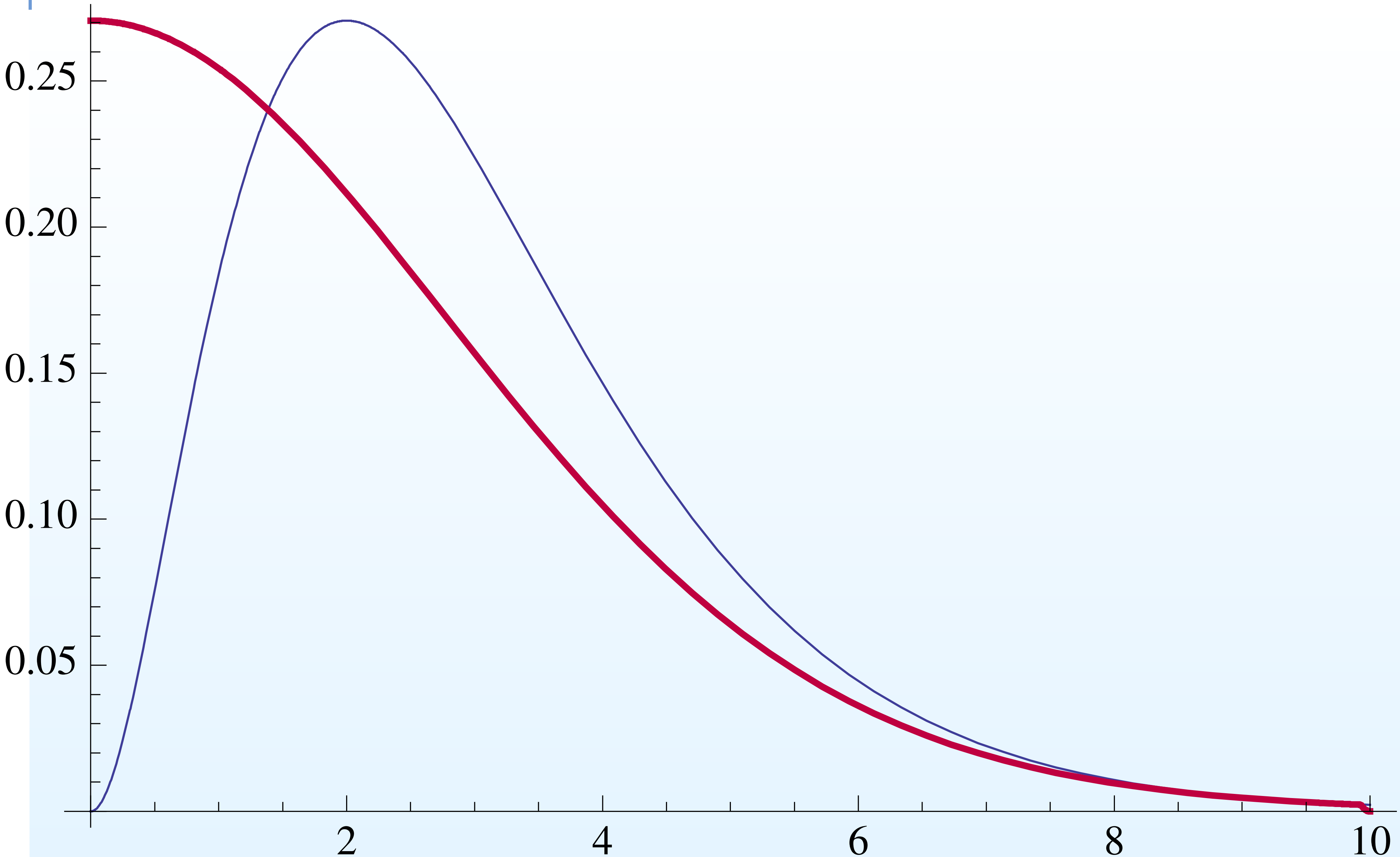
- Empirical (or canonical Least Squares, continuous case:

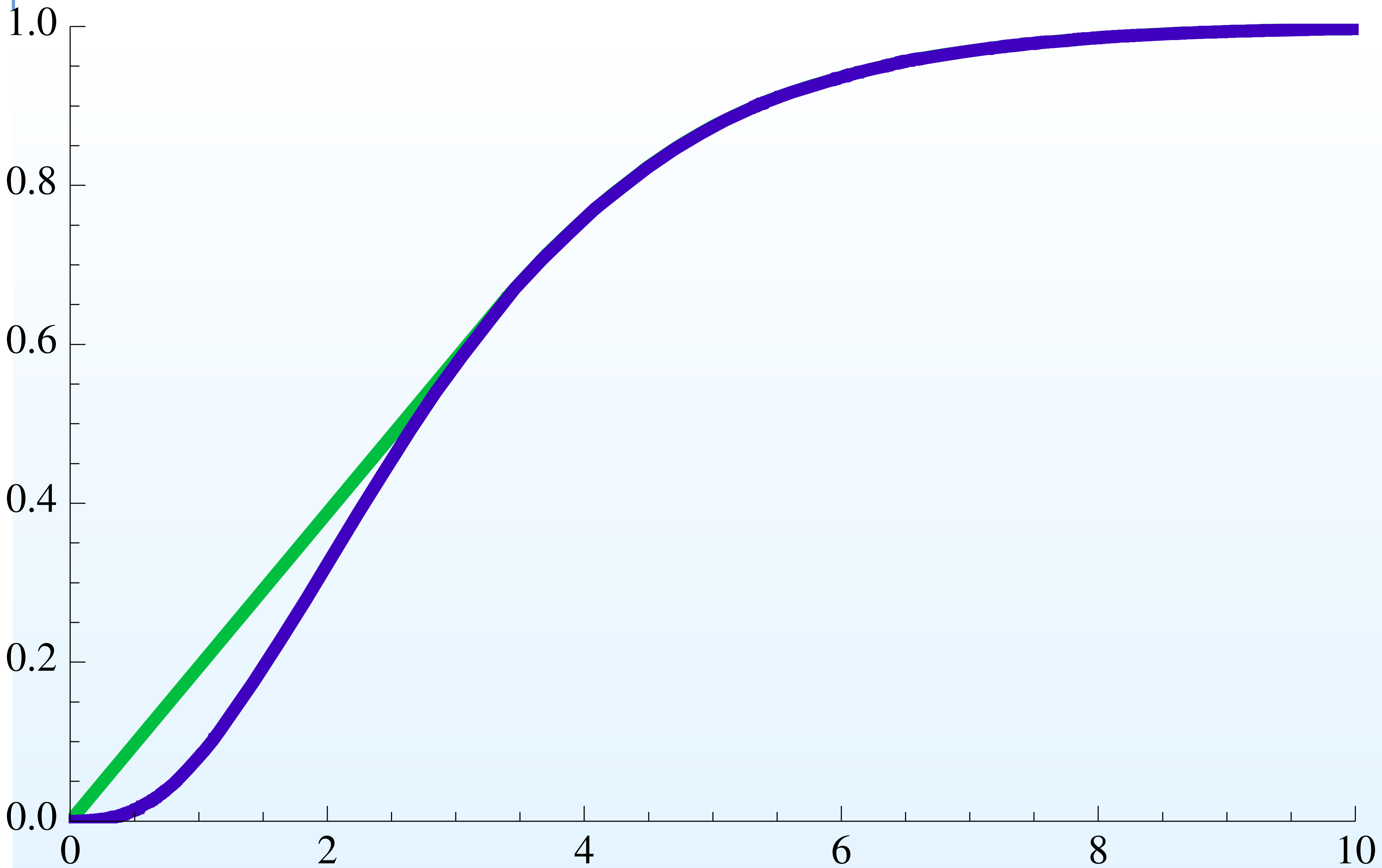
$$f^{LSE-empirical} = f^{MLE} = \partial \mathcal{I}_1(F).$$

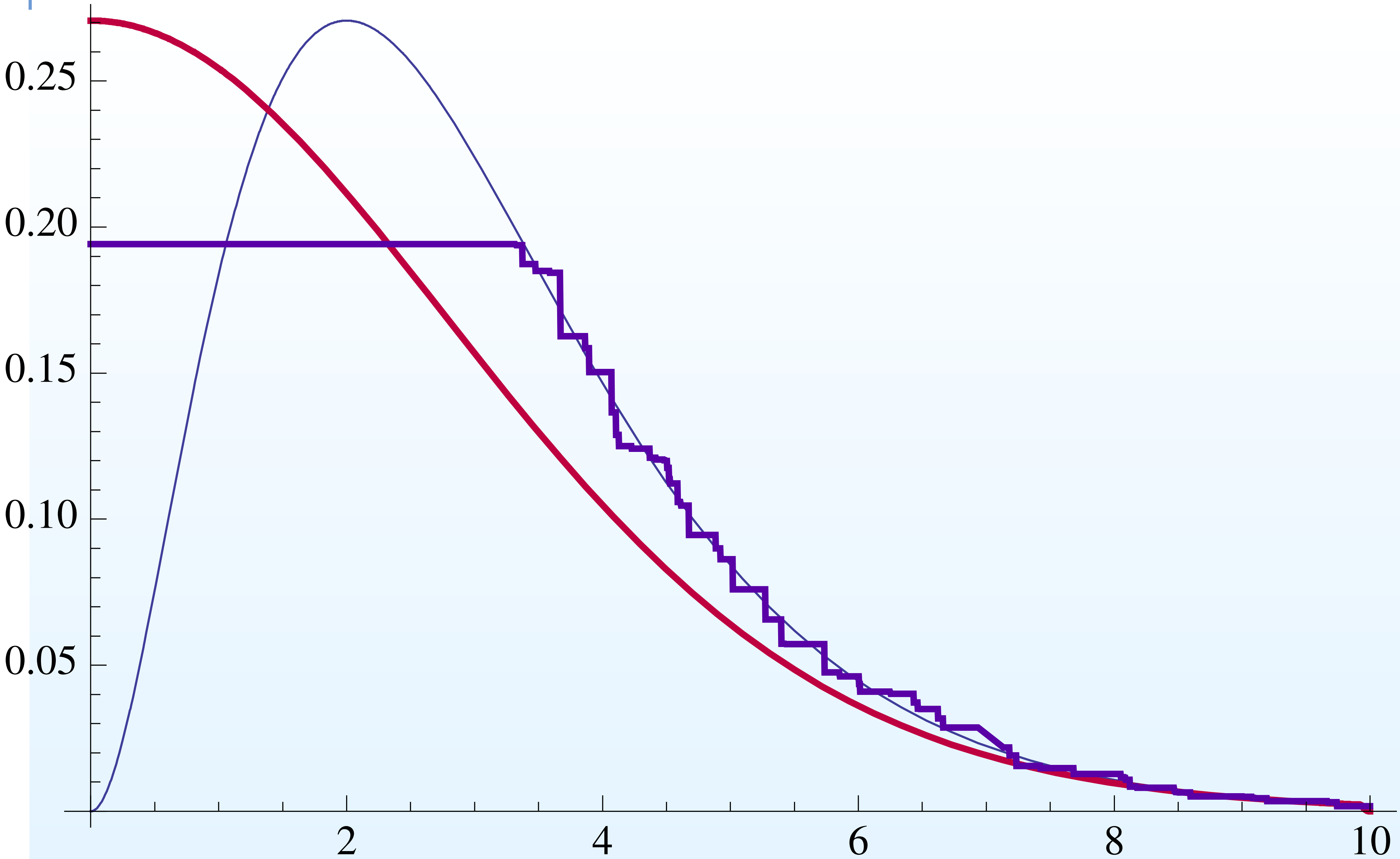
- Monotone Least Squares, discrete case:

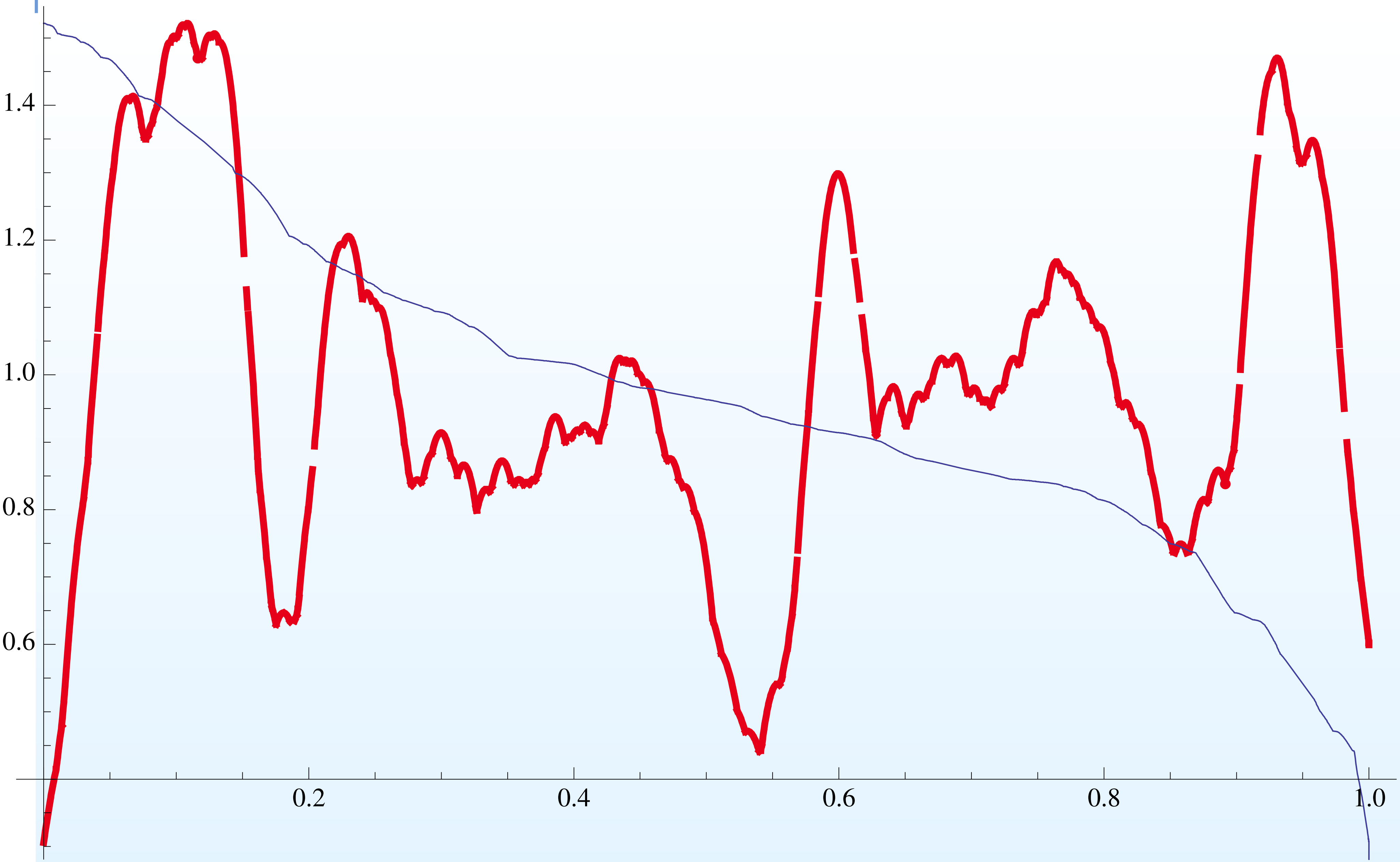
$$f^{LSE} = \partial \mathcal{I}_1 \left(\sum_0^{\cdot} f_i \right).$$

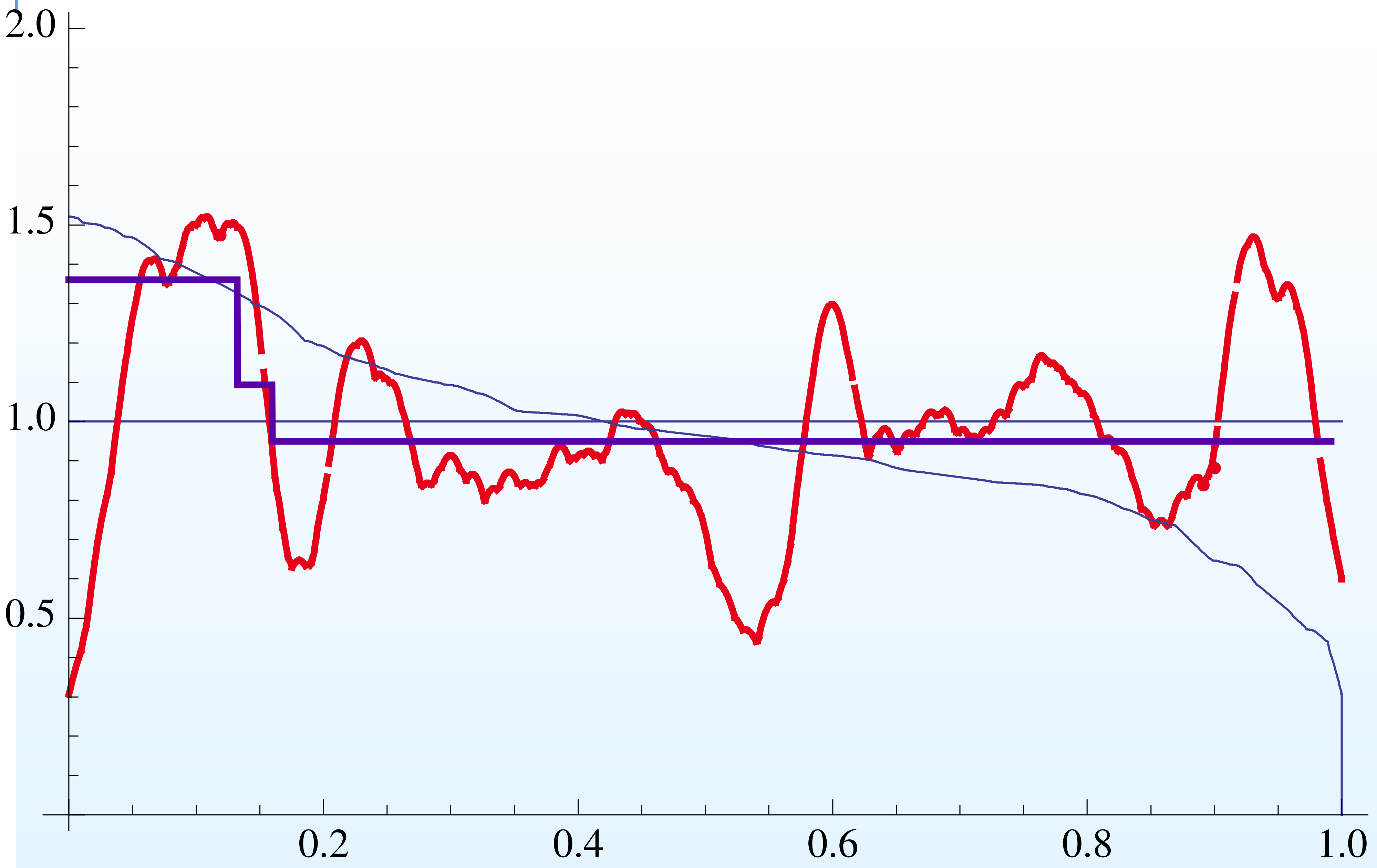

0.25
0.20
0.15
0.10
0.05
2
4
6
8
10

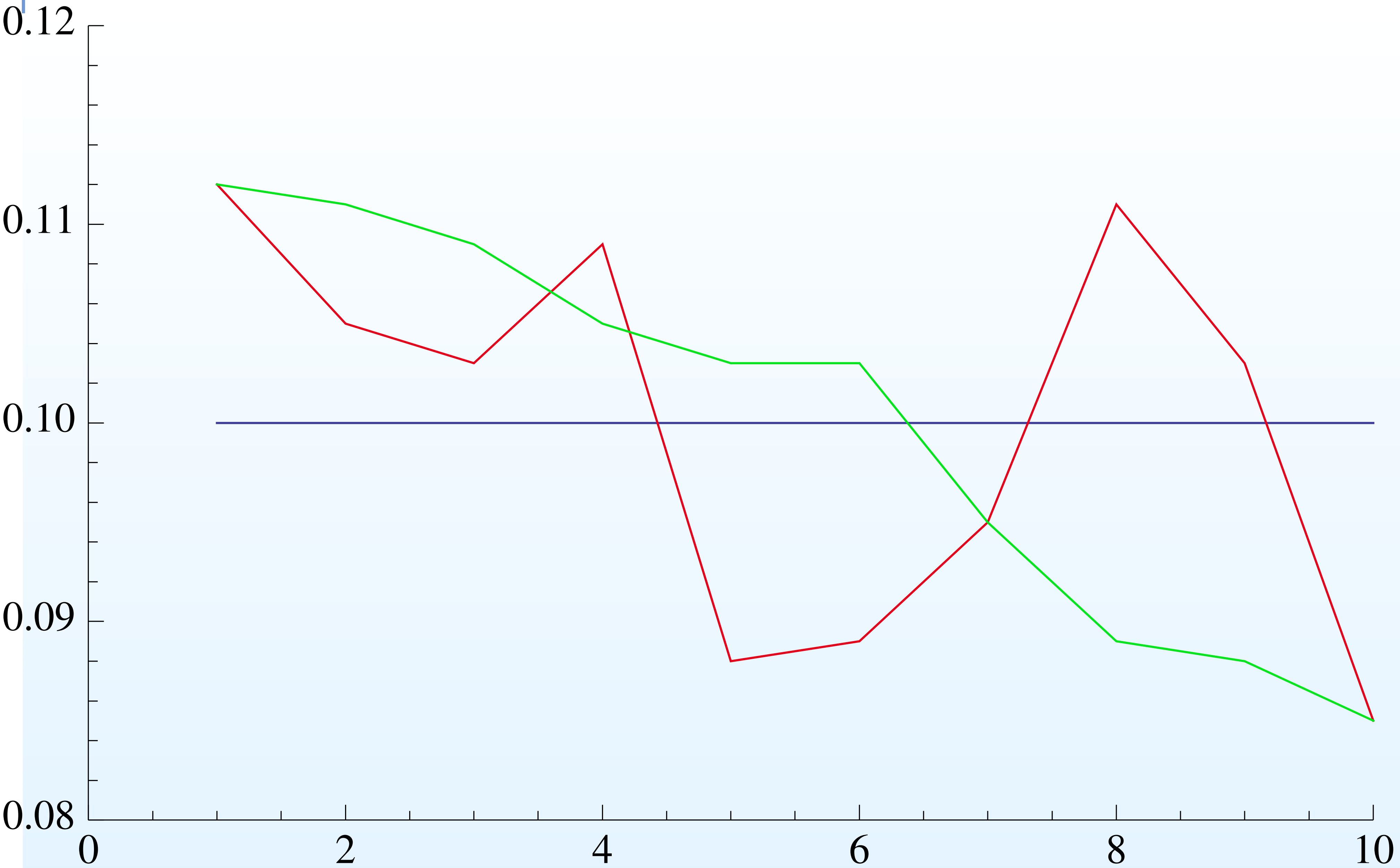

0.25
0.20
0.15
0.10
0.05
2
4
6
8
10

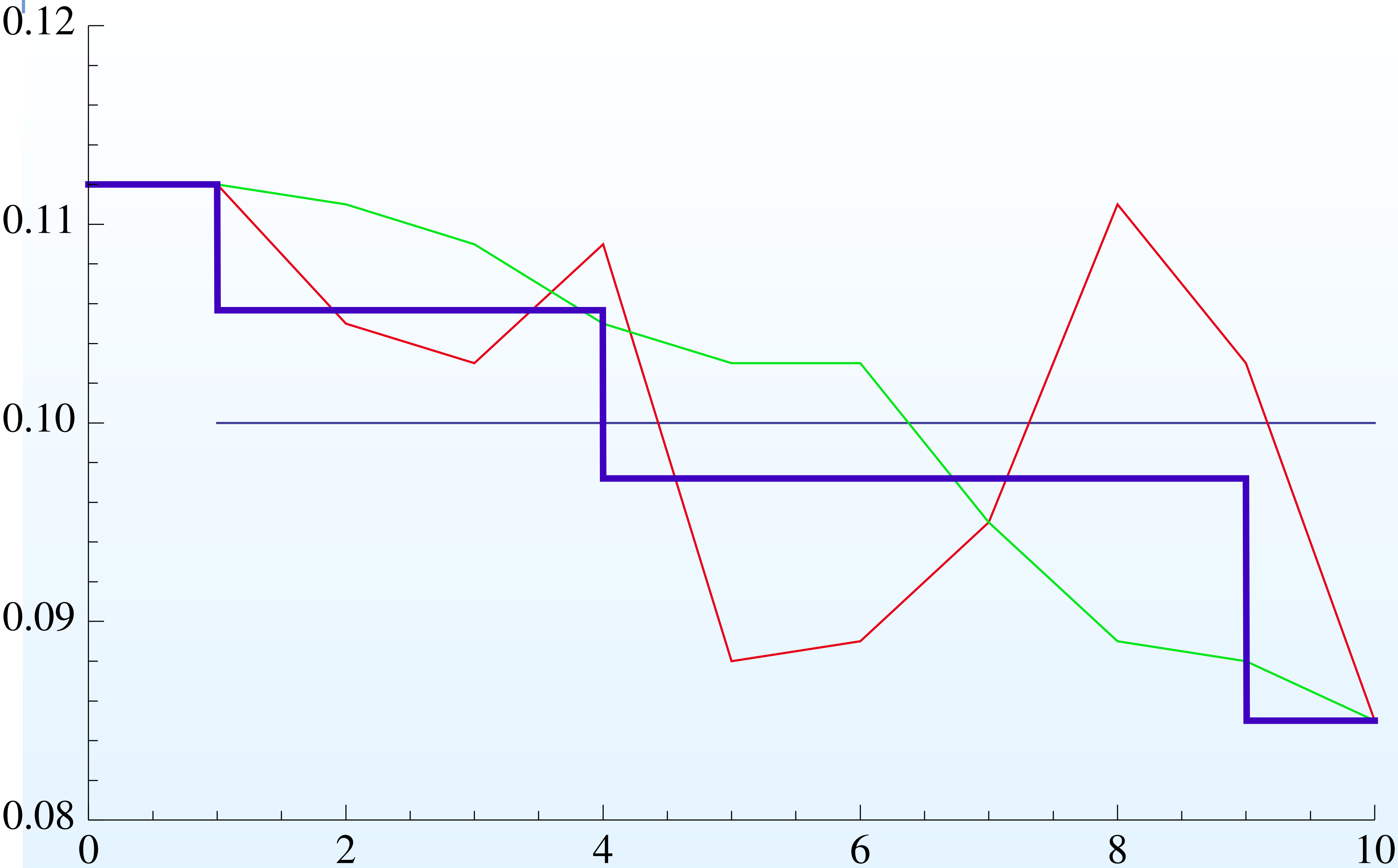

1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10

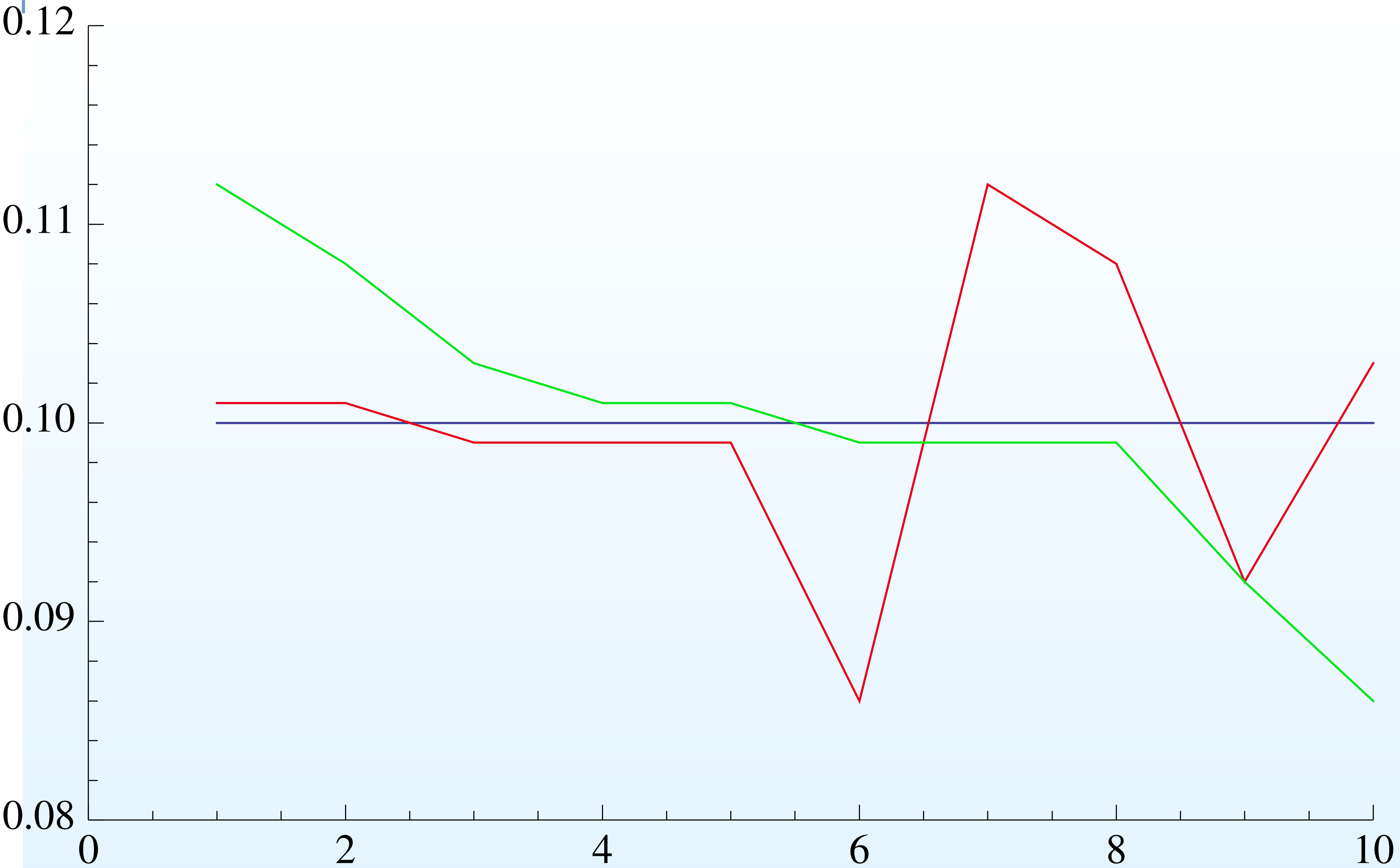

0.25
0.20
0.15
0.10
0.05
2
4
6
8
10

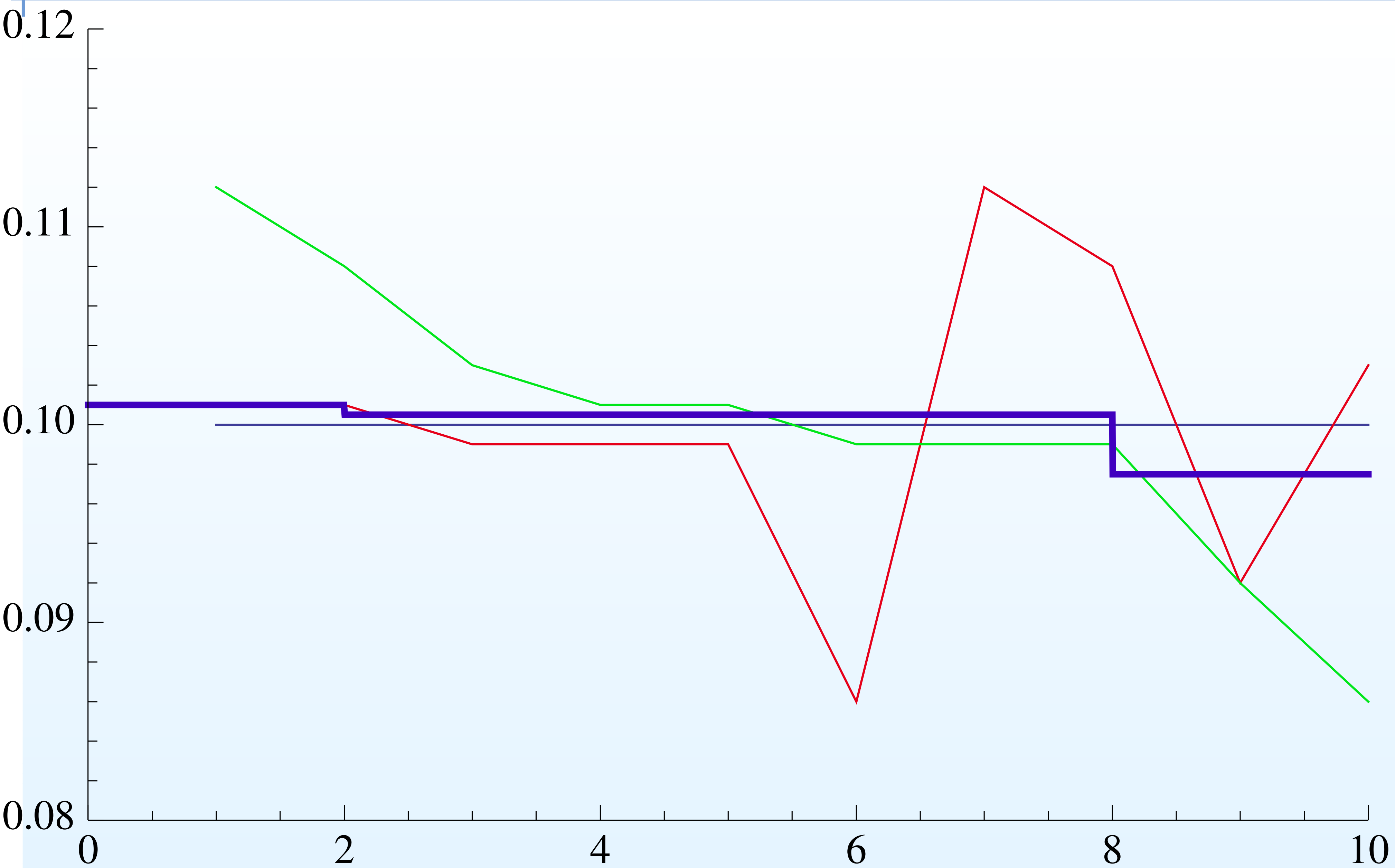

1.4
1.2
1.0
0.8
0.6
0.2
0.4
0.6
0.8
1.0


2.0
1.5
1.0
0.5
0.2
0.4
0.6
0.8
1.0


0.12
0.11
0.10
0.09
0.08
0
2
4
6
8
10


0.12
0.11
0.10
0.09
0.08
0
2
4
6
8
10


0.12
0.11
0.10
0.09
0.08
0
2
4
6
8
10


0.12
0.11
0.10
0.09
0.08
0
2
4
6
8
10

Skiing toward the Nisqually Glacier


© 2008 Amar Andalkar
www.skimountaineer.com