Some Theory for Estimation with Shape Constraints

Jon A. Wellner

University of Washington

- Talks at YES-I Conference on
  Shape Restricted Inference
  Eurandom,
  The Netherlands, October 8-10, 2007
- *Email: jaw@stat.washington.edu*
  http: //www.stat.washington.edu/jaw/jaw.research.html
- Based on joint work with Piet Groeneboom, Geurt Jongbloed;
  former Ph.D. Students Jian Huang, Moulinath Banerjee, Fadoua Balabdaoui, Marloes Maathuis, and Shuguang Song;
  current Ph.D. student Marios Pavlides, current post-doc Hanna Jankowski;
  and the work of many others.

- 1987: Diaconis and Freedman paper entitled:

- 1987: Diaconis and Freedman paper entitled:

  “A dozen de Finetti-style theorems
  in search of a theory”

- 1987: Diaconis and Freedman paper entitled:

  “A dozen de Finetti-style theorems
  in search of a theory”

- Since then Diaconis has written at least a half dozen further papers on de Finetti theorems.

- 1987: Diaconis and Freedman paper entitled:

  "A dozen de Finetti-style theorems
  in search of a theory"

- Since then Diaconis has written at least a half dozen further papers on de Finetti theorems.

- Is there a theory now??

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 of normal (convolution - deconvolution)

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

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

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

- Monotone: blockwise decreasing

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

- Monotone: blockwise decreasing
- Monotone: mixtures of (anchored at zero rectangular) uniform densities

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

- Monotone: blockwise 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 decreasing
- Monotone: mixtures of (anchored at zero rectangular) uniform densities
- Monotone: distribution functions
- Convex

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

- Monotone: blockwise decreasing
- Monotone: mixtures of (anchored at zero rectangular) uniform densities
- Monotone: distribution functions
- Convex
- Mixtures (normal, . . .)

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

- Monotone: blockwise decreasing
- Monotone: mixtures of (anchored at zero rectangular) uniform densities
- Monotone: distribution functions
- Convex
- 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
- 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
- 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
- Hazard function h:

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

- Intensity function λ of a Poison process
- Distribution functions in interval censoring models

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

- Density function f
- Hazard function h:

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

- Intensity function λ of a Poison process
- Distribution functions in interval censoring models
- Spectral density function

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

- Density function f
- Hazard function h:

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

- Intensity function λ of a Poison process
- Distribution functions in interval censoring models
- Spectral density function
- $\ldots$

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
- . . .

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??
- ...

Current state of shape restricted inference

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

Focus in my three lectures:

- Maximum likelihood and least squares estimators

Current state of shape restricted inference

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

Focus in my three lectures:

- 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?

Focus in my three lectures:

- 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?

Focus in my three lectures:

- Maximum likelihood and least squares estimators
- Pointwise limit theory
- Local (pointwise) lower bounds
- Where does empirical process theory come in?

Current state of shape restricted inference

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

Focus in my three lectures:

- Maximum likelihood and least squares estimators
- Pointwise limit theory
- Local (pointwise) lower bounds
- Where does empirical process theory come in?

Outline, rest of Lecture 1

1. An outline (or pattern) for pointwise limit theory

Outline, rest of Lecture 1

1. An outline (or pattern) for pointwise limit theory
2. 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 $f'(x_0) < 0$

Outline, rest of Lecture 1

1. An outline (or pattern) for pointwise limit theory
2. 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 $f^{(j)}(x_0) = 0$, $j = 1, \ldots, k-1$, $f^{(k)}(x_0) \neq 0$.

Outline, rest of Lecture 1

1. An outline (or pattern) for pointwise limit theory
2. 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).

Outline, rest of Lecture 1

1. An outline (or pattern) for pointwise limit theory
2. 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.

Outline, rest of Lecture 1

1. An outline (or pattern) for pointwise limit theory
2. 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.

Outline, rest of Lecture 1

1. An outline (or pattern) for pointwise limit theory
2. 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.

Outline, Lecture 2

- Illustration of the pointwise limit theory pattern:
  convex densities

Outline, Lecture 2

- Illustration of the pointwise limit theory pattern: convex densities
- Illustration of the pointwise limit theory pattern: log-concave densities

Outline, Lecture 2

- Illustration of the pointwise limit theory pattern: convex densities
- Illustration of the pointwise limit theory pattern: log-concave densities
- Illustration of the pointwise limit theory pattern: convex hazard functions

Outline, Lecture 2

- Illustration of the pointwise limit theory pattern:
  convex densities
- Illustration of the pointwise limit theory pattern:
  log-concave densities
- Illustration of the pointwise limit theory pattern:
  convex hazard functions
- A functional of interest:
  estimation of the mode

Outline, Lecture 3

- Illustration of the pointwise limit theory pattern: competing risks with current status data

Outline, Lecture 3

- Illustration of the pointwise limit theory pattern: competing risks with current status data
- Partial illustration of the pointwise limit theory pattern: $k-$monotone densities

Outline, Lecture 3

- Illustration of the pointwise limit theory pattern: competing risks with current status data
- Partial illustration of the pointwise limit theory pattern: $k-$monotone densities
- Partial illustration of the pointwise limit theory pattern: distribution functions and monotone densities on $\mathbb{R}^2$

Outline, Lecture 3

- Illustration of the pointwise limit theory pattern: competing risks with current status data
- Partial illustration of the pointwise limit theory pattern: $k-$monotone densities
- Partial illustration of the pointwise limit theory pattern: distribution functions and monotone densities on $\mathbb{R}^2$
- Problems and directions ...

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: bracketing CLT with functions dependent on n.

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: bracketing CLT with functions dependent on n.
- **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

Global analogues:

- Global rate result via Birgé & Massart, Wong & Shen global rate theorem (van der Vaart and Wellner (1996), Theorems 3.2.5 or 3.4.4).

- **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 provided by Groeneboom's lemma (Donoho & Liu, Le Cam).

Global analogues:

- Global rate result via Birgé & Massart, Wong & Shen global rate theorem (van der Vaart and Wellner (1996), Theorems 3.2.5 or 3.4.4).
- Global minimax lower bounds (Assouad's lemma or Fano's lemma).

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

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

where

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

Step 2. Characterization: the Fenchel conditions

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

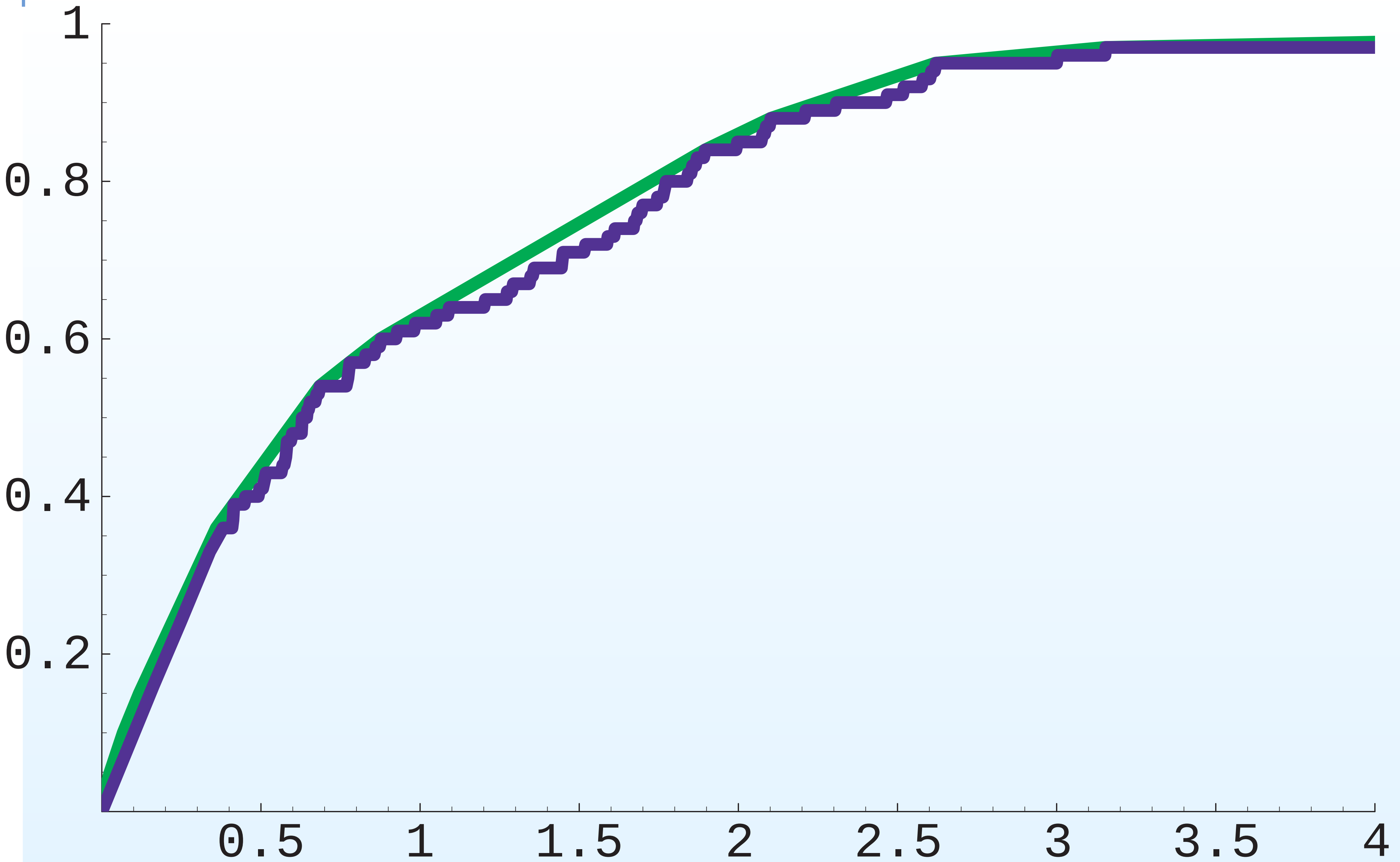
$$\mathbb{F}_n(x) = \widehat{F}_n(x) \quad \text{if and only if } \widehat{f}_n(x-) > \widehat{f}_n(x+).$$

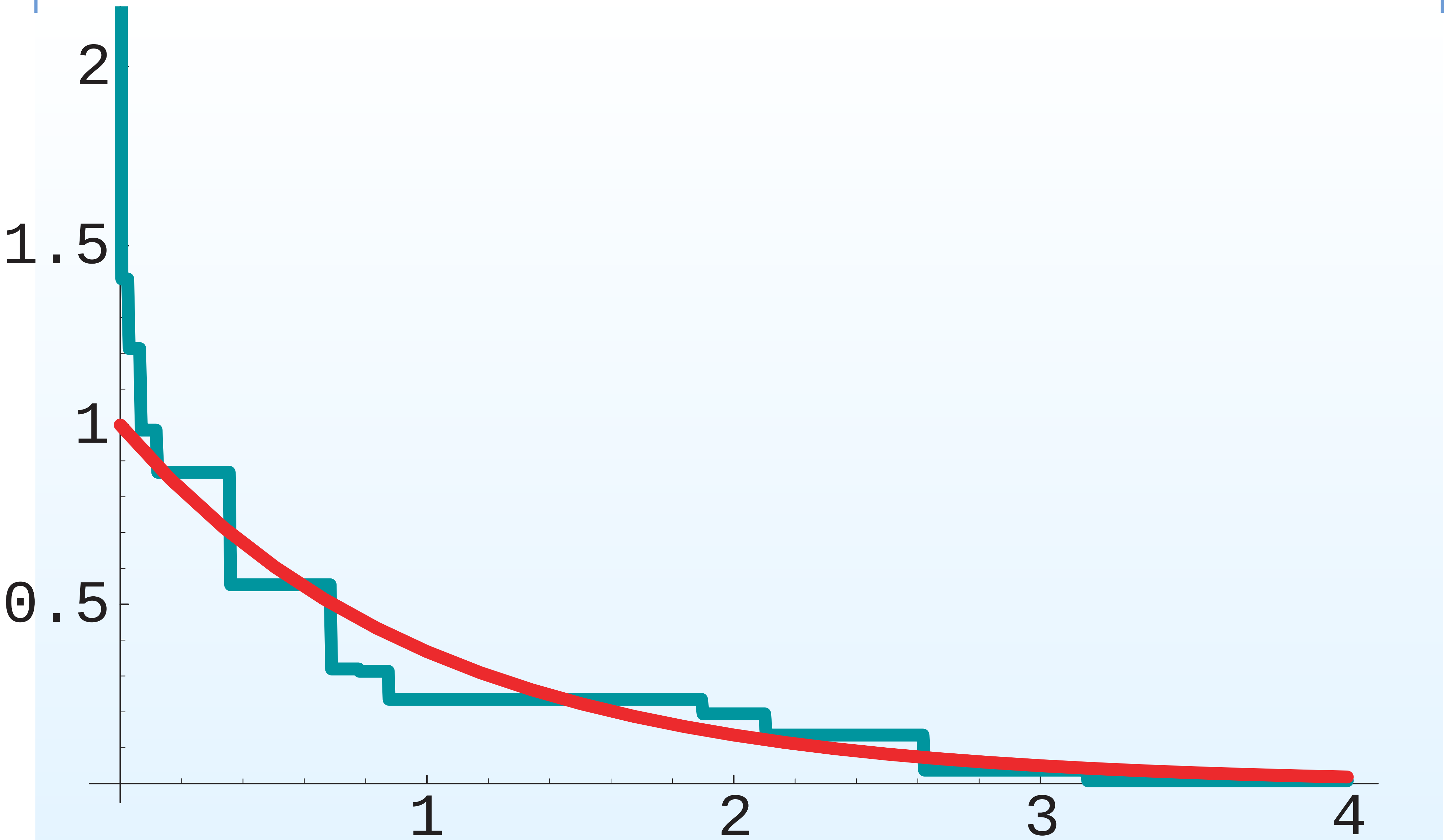
The second of these is equivalent to

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

The geometric interpretation of these two conditions is

$\widehat{f}_n(x) =$ the left-derivative of the slope at x of the least concave majorant $\widehat{F}_n$ of $\mathbb{F}_n$


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

Special feature:
Grenander and other monotone function problems.

Switching

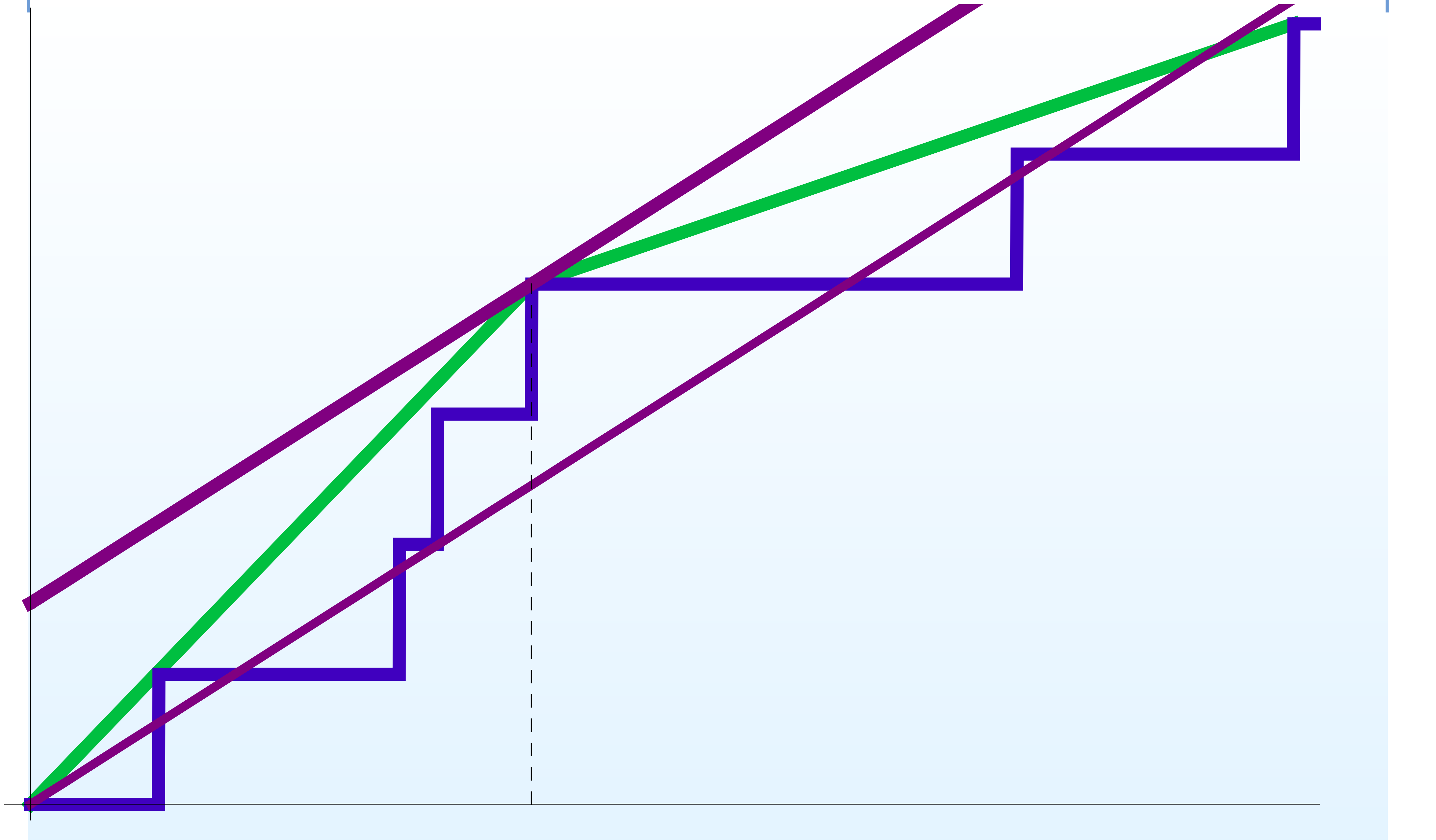
Let

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

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: Not available (yet?) for other models.



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

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

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

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

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

- No localization in this case!

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) showed that

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

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

Proof: See van der Vaart and Wellner (1996), pages 296 - 297.

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

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

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

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) that

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

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

- "Driving process" is $\mathbb{Z}(h) \equiv \sqrt{f(x_0)} W(h) - A|h|^{p+1}$.
- Process related to estimator maintaining Fenchel relations in the limit is $\mathbb{C}_p$ and its slope process $\mathbb{C}_p^{(1)} \equiv \mathbb{S}_p$:

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

- Localization rate is $n^{-1/(2p+1)}$

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

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

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; see the handout.

- "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 $\mathbb{C}_{loc}$ and its slope process $\mathbb{C}_{loc}^{(1)} \equiv \mathbb{S}_{loc}$:

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

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

- Localization only to the interval $[a,b]$

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,

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

$\bar{f}(x_0) \equiv (f(x_0+) + f(x_0-))/2$, $d \equiv f(x_0-) - f(x_0+)$. Somewhat more naturally,

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

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 $\mathbb{K}(1) \equiv \mathbb{R}$:

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

- Localization rate is n^{-1}!