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 $\lambda$ of a Poisson 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 $\lambda$ of a Poisson process
- Distribution functions in interval censoring models (non-regular or “inverse problem”)

Some Theory for Estimation – p. 5/60
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 $\lambda$ of a Poisson process
- Distribution functions in interval censoring models (non-regular or “inverse problem”)
- Distribution functions in direct observation models (regular or “forward problem”)

Some Theory for Estimation – p. 5/60
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 $\lambda$ 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
  - supremaum 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??
- . . .
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?

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.** $X \sim P_{\kappa}, \kappa \in K$, 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.** \( X \sim P_\kappa, \kappa \in K \), 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 \( f \downarrow 0 \).

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

\[
\hat{f}_n = \arg\max_{f \in \mathcal{M}_1} \left\{ \sum_{i=1}^{n} \log f(X_i) \right\} = \text{the MLE},
\]

\[
\tilde{f}_n = \arg\min_{f \in \mathcal{K}_1} \psi_n(f) = \text{the LSE}
\]

where \( \psi_n(f) \equiv \frac{1}{2} \int_0^\infty f^2(x) \, dx - \int_0^\infty f(x) \, dF_n(x) \).

In this particular case, \( \hat{f}_n = \tilde{f}_n \), i.e. LSE = MLE. (This is not true in general.)
Step 2. Characterization: the Fenchel conditions

\[ F_n(x) \leq \hat{F}_n(x) \equiv \int_0^x \hat{f}_n(t) \, dt \quad \text{for all } x \in [0, \infty), \quad \text{and} \]

\[ F_n(x) = \hat{F}_n(x) \quad \text{if and only if } \hat{f}_n(x-) > \hat{f}_n(x+). \]

The second of these is equivalent to

\[ \int_0^\infty (\hat{F}_n(x) - F_n(x)) \, d\hat{f}_n(x) = 0. \]

The geometric interpretation of these two conditions is

\[ \hat{f}_n(x) = \left\{ \begin{array}{l} \text{the left-derivative of the slope at } x \text{ of the least concave majorant } \hat{F}_n \text{ of } F_n \\ \equiv \partial I_1(F_n) \\ \equiv \text{Grenander estimator of } f. \end{array} \right\} \]
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

\[ \hat{s}_n(a) \equiv \text{argmax}_s \{ F_n(s) - as \}, \quad a > 0. \]

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

\[ \left\{ \hat{f}_n(t) \leq a \right\} = \left\{ \hat{s}_n(a) \leq t \right\}. \]

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 $x_0 \in (0, 1)$

$$\sqrt{n}(\hat{f}_n(x_0) - f(x_0)) \to_d S(x_0) = \partial I_1(U)(x_0)$$

where $S$ is the left derivative of the least concave majorant $I_1(U) = C$ of a standard Brownian bridge process $U$ on $[0, 1]$.

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

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

- 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

$$n^{1/3}(\hat{f}_n(x_0) - f(x_0)) \to_d \left(\frac{|f'(x_0)f(x_0)|}{2}\right)^{1/3}S(0)$$

where $S(0) = \partial I_1(Z)(0)$ 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
  $$Z_{a,b}(h) \equiv \sqrt{f(x_0)}W(h) + f'(x_0)h^2 \equiv aW(h) - bh^2.$$

- Process related to estimator maintaining Fenchel relations in the limit is $C$ and its slope process $C^{(1)} \equiv S$:
  $$C(h) \geq Z(h) \text{ for all } h \in (-\infty, \infty),$$
  $$C(h) = Z(h) \text{ if and only if } C^{(1)}(h-) > C^{(1)}(h+).$$

- 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),

$$n^{p/(2p+1)}(\hat{f}_n(x_0) - f(x_0)) \to_d (f(x_0)^p A)^{1/(2p+1)} S_p(0);$$

with $A = f^{(p)}(x_0)/(p + 1)!$. Here $S_p(0) = \partial I_1(Z)(0)$ is the slope at 0 of the least concave majorant of $Z(h) = W(h) - |h|^{p+1}$.

• “Driving process” is

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

• Process related to estimator maintaining Fenchel relations in the limit is $C_p \equiv I_1(Z_p)$ and its slope process

$$C_p^{(1)} \equiv S_p \partial I_1(Z_p):$$

$$C_p(h) \geq Z_p(h) \text{ for all } h \in (-\infty, \infty),$$

$$C_p(h) = Z_p(h) \text{ if and only if } C_p^{(1)}(h-) > C_p^{(1)}(h+).$$
• 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

$$
\sqrt{n}(\hat{f}_n(x_0) - f(x_0)) \rightarrow_{d} \frac{f(x_0)}{\sqrt{p}} \left\{ \sqrt{1 - p}Z + S\left(\frac{x_0 - a}{b - a}\right) \right\}
$$

where $p \equiv f(x_0)(b - a) = F(b) - F(a),$ $Z \sim N(0, 1),$ $S$ is the process of slopes of a Brownian bridge process $U$ as in case 1, and $Z$ and $S$ are independent.

This is much as in case 1, but with a twist or two.

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

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

  $C_{loc}(h) = \mathbb{Z}(h)$ \textbf{if and only if} $C_{loc}^{(1)}(h-) > C_{loc}^{(1)}(h+).$
• 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

\[
P(\hat{f}_n(x_0) - \bar{f}(x_0) \leq x) \rightarrow P(\arg\max\{\mathbb{N}_0(h) - \rho_{x+d/2,x-d/2}(h)\} \leq 0)
\]

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,

\[
\rho_{B,C}(h) \equiv \begin{cases} Bh, & h \geq 0 \\ -Ch, & h < 0. \end{cases}
\]

\[
\bar{f}(x_0) \equiv (f(x_0+) + f(x_0-))/2, \quad d \equiv f(x_0-) - f(x_0+).
\]

Furthermore, by switching again in the limit (Poisson) problem,

\[
\hat{f}_n(x_0) - \bar{f}(x_0) \rightarrow_d \mathbb{R}(0)
\]

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 $M(h) \equiv N_0(h) - (d/2)|h|$.

• Process related to estimator maintaining Fenchel relations in the limit is $K$ and its slope process $K^{(1)} \equiv \mathbb{R}$:

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

• 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 $f \downarrow 0$, $f$ 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

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

$$\tilde{f}_n = \arg\min_{f \in \mathcal{K}_2} \psi_n(f) = \text{the LSE}$$

where $\psi_n(f) \equiv \frac{1}{2} \int_0^\infty f^2(x) \, dx - \int_0^\infty f(x) \, dF_n(x)$. In this case, $\hat{f}_n \neq \tilde{f}_n$, i.e. LSE $\neq$ MLE.
Step 2. Characterization: the Fenchel conditions for $\tilde{f}_n$:

Let

$$\tilde{H}_n(x) \equiv \int_0^x \int_0^y \tilde{f}_n(t) dt dy$$

for all $x \in [0, \infty)$, and

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

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

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

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

$\tilde{H}_n$ has convex second derivative $\tilde{f}_n$. 
Step 3. Localization rate / tightness

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

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

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

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

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

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

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

$$\sup_{|t| \leq M} |\tilde{f}(x_0 + n^{-1/5}t) - f_0(x_0) - n^{-1/5}tf'(x_0)| = \mathcal{O}_p(n^{-2/5}),$$

and

$$\sup_{|t| \leq M} |\tilde{f}'(x_0 + n^{-1/5}t) - f'(x_0)| = \mathcal{O}_p(n^{-1/5}).$$

Step 4. Localize the Fenchel relations: define

$$\mathbb{Y}^{loc}_n(t) \equiv n^{4/5} \int_{x_0}^{x_0 + n^{-1/5}t} \left\{ F_n(v) - F_n(x_0) ight\} dv + \int_{x_0}^{v} \left( f(x_0) + (u - x_0)f(x_0) du \right) dv,$$
\[ \tilde{H}^{\text{loc}}_n(t) \equiv n^{4/5} \int_{x_0}^{x_0 + n^{-1/5}t} \int_{x_0}^{v} \{ \tilde{f}_n(u) - f(x_0) - (u - x_0)f'(x_0) \} \, du \, dv \]
\[ + \tilde{A}_n t + \tilde{B}_n. \]

Then
\[ \tilde{H}^{\text{loc}}_n(t) \geq \mathbb{Y}^{\text{loc}}_n(t) \]

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

Note that
\[ (\tilde{H}^{\text{loc}}_n)^{(2)}(t) = n^{2/5}(\tilde{f}_n(x_0 + n^{-1/5}t) - f(x_0) - n^{-1/5}tf'(x_0)), \]
\[ (\tilde{H}^{\text{loc}}_n)^{(3)}(t) = n^{1/5}(\tilde{f}'_n(x_0 + n^{-1/5}t) - f'(x_0)). \]
Step 5. Weak convergence of the (localized) driving process $\mathbb{Y}_n$ to a limit (Gaussian) driving process 

$$
\mathbb{Y}^{loc}_n(t) \overset{d}{=} \frac{n^{3/10}}{\int_{x_0}^{x_0+n^{-1/5}t}} \left\{ \mathbb{U}_n(F_0(v)) - \mathbb{U}_n(F(x_0)) \right\} dv + \frac{1}{24} f''(x_0) t^4 \ + o(1)
$$

$$
\Rightarrow \sqrt{f(x_0)} \int_0^t W(s) ds + \frac{1}{24} f''(x_0) t^4
$$

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

$$
= a \int_0^t W(s) ds + \sigma t^4
$$

$$
\equiv \mathbb{Y}(t) \equiv \mathbb{Y}_{a,\sigma}(t)
$$

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. $\text{Uniform}(0, 1)$, $a \equiv \sqrt{f(x_0)}$, $\sigma \equiv f''(x_0)/24$. 

Some Theory for Estimation – p. 35/60
Step 6. Preservation of (localized) Fenchel relations in the limit.

- \( \{(\tilde{H}_n^{loc}, \tilde{H}_n^{loc,(1)}, \tilde{H}_n^{loc,(2)}, \tilde{H}_n^{loc,(3)})\}_{n \geq 1} \) is tight.
Step 6. Preservation of (localized) Fenchel relations in the limit.

- \[ \{(\tilde{H}^{loc}_{n}, \tilde{H}^{loc,(1)}_{n}, \tilde{H}^{loc,(2)}_{n}, \tilde{H}^{loc,(3)}_{n})\}_{n \geq 1} \text{ is tight.} \]

- \[ Y^{loc}_{n} \Rightarrow Y \]
Step 6. Preservation of (localized) Fenchel relations in the limit.

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

- \( Y_{loc}^{(n)} \Rightarrow Y \)

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

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

- \[ Y^{loc}_n \Rightarrow Y \]

- Fenchel relations satisfied:
  - \[ \widetilde{H}^{loc}_n(x) \geq Y^{loc}_n(x) \text{ for all } x \]
Step 6. Preservation of (localized) Fenchel relations in the limit.

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

- $Y^{loc}_{n} \Rightarrow Y$

- Fenchel relations satisfied:
  - $\tilde{H}^{loc}_{n}(x) \geq Y^{loc}_{n}(x)$ for all $x$
  - $\int_{-\infty}^{\infty} (\tilde{H}^{loc}_{n}(x) - Y^{loc}_{n}(x)) d\tilde{H}^{loc,(3)}_{n}(x) = 0.$
Step 6. Preservation of (localized) Fenchel relations in the limit.

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

- \{Y^{loc}_{n} \Rightarrow Y\}

- Fenchel relations satisfied:
  - \(\tilde{H}^{loc}_{n}(x) \geq Y^{loc}_{n}(x)\) for all \(x\)
  - \(\int_{-\infty}^{\infty} (\tilde{H}^{loc}_{n}(x) - Y^{loc}_{n}(x)) d\tilde{H}^{loc,(3)}_{n}(x) = 0\).

- Any limit process \(H\) for a subsequence \(\{\tilde{H}^{loc}_{n'}\}\) must satisfy
Step 6. Preservation of (localized) Fenchel relations in the limit.

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

- Fenchel relations satisfied:
  - $\tilde{H}_{n}^{loc}(x) \geq \mathbb{Y}_{n}^{loc}(x)$ for all $x$
  - $\int_{-\infty}^{\infty} (\tilde{H}_{n}^{loc}(x) - \mathbb{Y}_{n}^{loc}(x)) d\tilde{H}_{n}^{loc,(3)}(x) = 0$.

- Any limit process $H$ for a subsequence $\{\tilde{H}_{n'}^{loc}\}$ must satisfy
  - $H(x) \geq \mathbb{Y}(x)$ for all $x$. 

Some Theory for Estimation – p. 36/60
Step 6. Preservation of (localized) Fenchel relations in the limit.

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

Fenchel relations satisfied:
- \( \tilde{H}^{loc}_n(x) \geq Y^{loc}_n(x) \) for all \( x \)
- \( \int_{-\infty}^{\infty} (\tilde{H}^{loc}_n(x) - Y^{loc}_n(x))d\tilde{H}^{loc,(3)}_n(x) = 0. \)

Any limit process \( H \) for a subsequence \( \{ \tilde{H}^{loc}_{n'} \} \) must satisfy
- \( H(x) \geq Y(x) \) for all \( x \).
- \( \int_{-\infty}^{\infty} (H(x) - Y(x))dH^{(3)}(x) = 0. \)
Step 6. Preservation of (localized) Fenchel relations in the limit.

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

- \( Y_{loc}^{(n)} \Rightarrow Y \)

- Fenchel relations satisfied:
  - \( \widehat{H}_{loc}^{(n)}(x) \geq Y_{loc}^{(n)}(x) \) for all \( x \)
  - \( \int_{-\infty}^{\infty} (\widehat{H}_{loc}^{(n)}(x) - Y_{loc}^{(n)}(x)) d\widehat{H}_{loc}^{(3)(n)}(x) = 0. \)

- Any limit process \( H \) for a subsequence \( \{\widehat{H}_{loc}^{(n')}\} \) must satisfy
  - \( H(x) \geq Y(x) \) for all \( x \).
  - \( \int_{-\infty}^{\infty} (H(x) - Y(x)) dH^{(3)}(x) = 0. \)
  - \( H^{(2)} \) is convex.
Step 6. Preservation of (localized) Fenchel relations in the limit.

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

- \( Y_{n}^{loc} \Rightarrow Y \)

- Fenchel relations satisfied:
  - \( \tilde{H}_{n}^{loc}(x) \geq Y_{n}^{loc}(x) \) for all \( x \)
  - \( \int_{-\infty}^{\infty} (\tilde{H}_{n}^{loc}(x) - Y_{n}^{loc}(x)) d\tilde{H}_{n}^{loc,(3)}(x) = 0. \)

- Any limit process \( H \) for a subsequence \( \{\tilde{H}_{n}^{loc}\} \) must satisfy
  - \( H(x) \geq Y(x) \) for all \( x \).
  - \( \int_{-\infty}^{\infty} (H(x) - Y(x)) dH^{(3)}(x) = 0. \)
  - \( H^{(2)} \) 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

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

where

$$
c_1(f) \equiv \left( \frac{f^2(x_0)f''(x_0)}{24} \right)^{1/5}, \quad c_2(f) \equiv \left( \frac{f(x_0)f''(x_0)^3}{24^3} \right)^{1/5}.
$$
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 $f_\epsilon$ 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 $c_\epsilon$ 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 \( \tilde{T}_n \) of \( f'(x_0), \)

\[
\begin{align*}
n^{2/5} \inf_{T_n} \max \{ E_{n,P} |T_n - f_{\epsilon_n}(x_0)|, E_{n,P} |T_n - f(x_0)| \} & \\
& \geq \frac{1}{4} \left( \frac{3}{e\sqrt{2}} \right)^{1/5} \cdot c_1(f), \\
n^{1/5} \inf_{\tilde{T}_n} \max \{ E_{n,P} |\tilde{T}_n - f'_{\epsilon_n}(x_0)|, E_{n,P} |\tilde{T}_n - f'(x_0)| \} & \\
& \geq \frac{1}{4} \left( \frac{6 \cdot 24^2}{e} \right)^{1/5} \cdot c_2(f)
\end{align*}
\]
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 “envelope 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 $H^{(3)}$
Some Theory for Estimation – p. 42/60
Some Theory for Estimation – p. 43/60
3. Some Comparisons: MLE / LSE

versus Rearrangements

Monotone

• Monotone rearrangement, continuous case:

\[ f^{\text{mon-rearr}} \equiv R(f) \text{ where} \]

\[ Z_f(s) = \lambda \{ x : f(x) \geq s \} , \quad R(f)(x) = Z_f^{-1}(x). \]
3. Some Comparisons: MLE / LSE

versus Rearrangements

Monotone

• Monotone rearrangement, continuous case:

\[ f_{\text{mon-rearr}} = R(f) \]

Where

\[
Z_f(s) = \lambda \{ x : f(x) \geq s \}, \quad R(f)(x) = Z_f^{-1}(x).
\]

• Monotone rearrangement, discrete case:

\[ f_{\text{mon-rearr}} = R(f) \]

Where

\[
Z_f(s) = \# \{ i \in \mathbb{Z}^+ : f(i) \geq s \}, \quad R(f)(i) = Z_f^{-1}(i).
\]
Monotone Least Squares, continuous case: (Mammen)

\[ f^{LSE} \equiv LS(f) = \partial I_1 \left( \int_0^\cdot f \, du \right) \]

where \( I_1 = \) Least Concave Majorant operator.
• Monotone Least Squares, continuous case: (Mammen)

\[ f^{LSE} \equiv LS(f) = \partial I_1 \left( \int_0^\cdot f \, du \right) \]

where \( I_1 = \) Least Concave Majorant operator.

• Empirical (or canonical Least Squares, continuous case):

\[ f^{LSE-empirical} = f^{MLE} = \partial I_1(F). \]
• Monotone Least Squares, continuous case: (Mammen)

\[ f^{LS\text{E}} \equiv LS(f) = \partial I_1 \left( \int_0^\cdot f \, du \right) \]

where \( I_1 = \text{Least Concave Majorant operator} \).

• Empirical (or canonical Least Squares, continuous case):

\[ f^{LS\text{E}-empirical} = f^{MLE} = \partial I_1(F) \]

• Monotone Least Squares, discrete case:

\[ f^{LS\text{E}} = \partial I_1 \left( \sum_0^\cdot f_i \right) \]
Skiing toward the Nisqually Glacier