# Lecture 2: 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.

• Illustration of the pointwise limit theory pattern: convex densities

- Illustration of the pointwise limit theory pattern: convex densities
- Illustration of the pointwise limit theory pattern: convex hazards

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

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

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

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

- 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  $\mathbb{R}^d$

- 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  $\mathbb{R}^d$

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

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

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

Log-likelihood

- 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 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 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 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 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 limit Fenchel relations

- Step 6. Preservation of (localized) Fenchel relations in the limit.
- Step 7. Unique (Gaussian world) estimator resulting from limit Fenchel relations
- Step 8 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).

2.1 Illustration: convex decreasing densities

Step 0.  $X \sim f$  on  $[0, \infty)$  with  $f \searrow 0$  and convex ( $\mathcal{K}$ ) Step 1. Optimization criteron: log-likelihood or least squares

$$\widehat{f}_n = \operatorname{argmax}_{f \in \mathcal{K}} \left\{ \sum_{i=1}^n \log f(X_i) \right\}$$

$$\widetilde{f}_n = \operatorname{argmin}_{f \in \mathcal{K}} \psi_n(f)$$

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 for  $\widetilde{f_n}$ : let

$$\widetilde{H}_{n}(x) \equiv \int_{0}^{x} \int_{0}^{y} \widetilde{f}_{n}(t) dt dy \quad \text{for all } x \in [0, \infty),$$
$$\mathbb{Y}_{n}(x) = \int_{0}^{x} \mathbb{F}_{n}(y) dy$$

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

$$\begin{split} \widetilde{H}_n(x) &\geq \mathbb{Y}_n(x) \quad \text{ for all } x > 0, \\ \int_0^\infty (\widetilde{H}_n(x) - \mathbb{Y}_n(x)) d\widetilde{H}_n^{(3)}(x) = 0, \\ \widetilde{H}_n \text{ has convex second derivative } \widetilde{f}_n \end{split}$$

and

#### 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

 $\tau_n^- \equiv \max\{t \le \xi_n : \widetilde{f}_n^{(3)} \text{discontinuous at } t\} \quad \tau_n^+ \equiv \min\{t > \xi_n : \widetilde{f}_n^{(3)} \text{discontinuous at } t\}$ 

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

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

Step 4. Localize the Fenchel relations: define

$$\mathbb{Y}_{n}^{loc}(t) \equiv n^{4/5} \int_{x_{0}}^{x_{0}+n^{-1/5}t} \{\mathbb{F}_{n}(v) - \mathbb{F}_{n}(x_{0}) + \int_{x_{0}}^{v} (f(x_{0}) + (u - x_{0})f(x_{0})du\} dv$$

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

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

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

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

 $\mathbb{Y}_{n}^{loc}(t)$  $\stackrel{d}{=} n^{3/10} \int_{x}^{x_0 + n^{-1/5}t} \{ \mathbb{U}_n(F_0(v)) - \mathbb{U}_n(F(x_0)) \} dv + \frac{1}{24} f''(x_0) t^4 + o(1) \} dv$  $\Rightarrow \sqrt{f(x_0)} \int_0^t W(s) ds + \frac{1}{24} f''(x_0) t^4$ 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. 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. • $\{(\widetilde{H}_n^{loc}, \widetilde{H}_n^{loc,(1)}, \widetilde{H}_n^{loc,(2)}, \widetilde{H}_n^{loc,(3)})\}_{n \ge 1}$ is tight.

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

• 
$$\mathbb{Y}_n^{loc} \Rightarrow \mathbb{Y}$$

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

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

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

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

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

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

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

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

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

- $\{(\widetilde{H}_n^{loc}, \widetilde{H}_n^{loc,(1)}, \widetilde{H}_n^{loc,(2)}, \widetilde{H}_n^{loc,(3)})\}_{n \ge 1}$  is tight.
- $\mathbb{Y}_n^{loc} \Rightarrow \mathbb{Y}$
- Fenchel relations satisfied:
  - $\circ \ \widetilde{H}_n^{loc}(x) \geq \mathbb{Y}_n^{loc}(x) \text{ for all } x$
  - $\circ \int_{-\infty}^{\infty} (\widetilde{H}_n^{loc}(x) \mathbb{Y}_n^{loc}(x)) d\widetilde{H}_n^{loc,(3)}(x) = 0.$
- Any limit process H for a subsequence  $\{\widetilde{H}_{n'}^{loc}\}$  must satisfy
  - $\circ H(x) \geq \mathbb{Y}(x)$  for all x.
  - $\circ \int_{-\infty}^{\infty} (H(x) \mathbb{Y}(x)) dH^{(3)}(x) = 0.$
  - $\circ$   $H^{(2)}$  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  $f \in C$ ,  $f(x_0) > 0$ ,  $f''(x_0) > 0$ , and f'' continuous in a neighborhood of  $x_0$ , then

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

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  $f_{\epsilon}$  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

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

$$n^{2/5} \inf_{T_n} \max \{ E_{n,P_n} | T_n - f_{\epsilon_n}(x_0) |, E_{n,P} | T_n - f(x_0) | \}$$
  
$$\geq \frac{1}{4} \left( \frac{3}{e\sqrt{2}} \right)^{1/5} \cdot c_1(f),$$

$$n^{2/5} \inf_{T_n} \max \{ E_{n,P_n} | T_n - f_{\epsilon_n}(x_0) |, E_{n,P} | T_n - f(x_0) | \}$$
  
$$\geq \frac{1}{4} \left( \frac{3}{e\sqrt{2}} \right)^{1/5} \cdot c_1(f),$$

$$n^{2/5} \inf_{T_n} \max \{ E_{n,P_n} | T_n - f_{\epsilon_n}(x_0) |, E_{n,P} | T_n - f(x_0) | \}$$
  
$$\geq \frac{1}{4} \left( \frac{3}{e\sqrt{2}} \right)^{1/5} \cdot c_1(f),$$

$$n^{2/5} \inf_{T_n} \max \{ E_{n,P_n} | T_n - f_{\epsilon_n}(x_0) |, E_{n,P} | T_n - f(x_0) | \}$$
  
$$\geq \frac{1}{4} \left( \frac{3}{e\sqrt{2}} \right)^{1/5} \cdot c_1(f),$$

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

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

- 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 "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  $H^{(3)}$







