Nonparametric estimation under Shape Restrictions

Jon A. Wellner

University of Washington, Seattle

Statistical Seminar, Frejus, France
August 30 - September 3, 2010
Outline: Five Lectures on Shape Restrictions

• L1: Monotone functions: maximum likelihood and least squares

• L2: Optimality of the MLE of a monotone density (and comparisons?)

• L3: Estimation of convex and $k$-monotone density functions

• L4: Estimation of log-concave densities: $d = 1$ and beyond

• L5: More on higher dimensions and some open problems
Outline: Lecture 3

• A: Convex decreasing and $k$–monotone densities as mixtures
• B: Existence and uniqueness of MLE, $k$–monotone, $k \geq 2$
• C: Consistency, $k$–monotone, $2 \leq k \leq \infty$
• D: Global rates of convergence: $2 \leq k < \infty$
• E: Local rate of convergence: $k = 2$
• F: Limiting distributions at a fixed point: $k = 2$
A. Convex decreasing and $k$–monotone 
densities as mixtures

**Definition 1.** Let $k$ be an integer, $k \geq 2$. A density $f$ on $\mathbb{R}^+$ is said to be $k$–monotone if $f^{(j)}$ exists for $j = 1, \ldots, k - 2$ with $(-1)^j f^{(j)}(x) \geq 0$ and $f^{(k-2)}$ is convex. Let $\mathcal{D}_k$ denote the class of all $k$–monotone densities.

**Definition 2.** A density $f$ on $\mathbb{R}^+$ is said to be completely monotone if $f^{(j)}$ exists for $j = 1, \ldots$ with $(-1)^j f^{(j)}(x) \geq 0$ for all $j$. Let $\mathcal{D}_\infty$ denote the class of all completely monotone densities.

In part D of Lecture 2 it was noted that every monotone decreasing density $f$ on $\mathbb{R}^+$ is a scale mixture of uniform densities, and every decreasing convex density $f$ is a scale mixture of triangular (or Beta(1, 2)) densities. In fact this extends to the class of $k$–monotone densities.
A. Convex decreasing and $k$–monotone densities as mixtures

$f \in \mathcal{D}_k$ if and only if

$$f(x) = \int_0^\infty ky^{-1}(1 - x/y)_{+}^{k-1}dG(y) \quad \text{for some distribution } G.$$ while $f \in \mathcal{D}_\infty$ if and only if

$$f(x) = \int_0^\infty y^{-1}\exp(-x/y)dG(y) \quad \text{for some distribution } G.$$ It is convenient to recast this as follows:

**Proposition 1.** (Williamson, 1956; Lévy, 1962; Bernstein)

A density $f$ is a $k$–monotone (completely monotone) density if and only if it can be represented as a scale mixture of Beta$(1, k)$ (exponential) densities; i.e. with $x_+ \equiv x1\{x \geq 0\}$,

$$f(x) = \begin{cases} \int_0^\infty y^{-1}(1 - x/ky)^{k-1}dG(y), & k \in \{1, 2, \ldots\}, \\ \int_0^\infty y^{-1}\exp(-x/y)dG(y), & k = \infty, \end{cases}$$

(1)

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

Statistical Seminar, Fréjus
A. Convex decreasing and $k$–monotone densities as mixtures

The inversion formulas corresponding to these mixture representations are given in the following proposition.

**Proposition 2.** Suppose that $f$ is a $k$–monotone density with distribution function $F$ (so $F(x) = \int_0^x f(t)dt$). Then the distribution function $G = G_k$ of (1) is given at continuity points of $G_k$ by

$$G_k(t) = \sum_{j=0}^{k} \frac{(-1)^j}{j!} (kt)^j F(j)(kt), \quad (2)$$

and the distribution function $G = G_\infty$ of the $k = \infty$ part of (1) is given at continuity points of $G_\infty$ by

$$G_\infty(t) = \lim_{k \to \infty} G_k(t). \quad (3)$$
A. Convex decreasing and $k$–monotone densities as mixtures

It will be convenient to have notation for the classes of functions given by the mixing representations in (1) when the mixing measure $G$ is not require to have mass 1, and hence the resulting functions $f$ are not necessarily densities. We denote these classes by $\mathcal{M}_k$ for $1 \leq k \leq \infty$. 
B. Existence and uniqueness of MLE, $k$–monotone, $k \geq 2$

Now suppose that $X_1, \ldots, X_n$ are i.i.d. $f_0 \in D_k$ for some $k \in \{2, \ldots, \infty\}$. The MLE $\hat{f}_n \equiv \hat{f}_{n,k}$ is defined by

$$\hat{f}_n = \arg\max\{\mathbb{P}_n \log f : f \in D_k\}.$$ 

The LSE $\tilde{f}_n \equiv \tilde{f}_{n,k}$ of $f_0$ is defined by

$$\tilde{f}_n \equiv \arg\min\{\psi_n(f) : f \in M_k \cap L_2(\lambda)\}$$

where

$$\psi_n(f) \equiv \frac{1}{2} \int_0^\infty f^2(x) \, dx - \int_0^\infty f(x) \, d\mathbb{F}_n(x).$$
B. Existence and uniqueness of MLE, $k$–monotone, $k \geq 2$

Theorem. ($k = 2$: Groeneboom, Jongbloed, W (2001);

2 < $k < \infty$: Balabdaoui (200x);

2 $\leq$ $k$ < $\infty$: Seregin (2010);

$k = \infty$: Jewell (1982))

(a) For 2 $\leq$ $k$ $\leq$ $\infty$ the MLE $\hat{f}_n$ exists and is unique.

(b) For 2 $\leq$ $k$ < $\infty$ the LSE $\tilde{f}_n$ exists and is unique.

(c) $\tilde{f}_{n,k} \neq \hat{f}_{n,k}$ for all $k \geq 2$.

Proof. Methods:

- Nonparametric estimation in mixtures: Lindsay (1983a,b); Lindsay (1995); Lindsay and Roeder (1993).

- Positivity / total positivity: Schoenberg and Whitney (1953); Polya and Szegő (1925); Karlin (1968).
B. Existence and uniqueness of MLE, $k$–monotone, $k \geq 2$

Theorem. $\hat{f}_{n,k}$ is characterized by:

(a) $2 \leq k < \infty$: The “Fenchel” conditions hold:

$$\int_{0}^{\infty} \frac{k(y-x)^{k-1}}{y^k \hat{f}_{n,k}(x)} d\hat{F}_n(x) \leq 1$$

for all $y > 0$ with equality if and only if $y \in \text{supp}(\hat{G}_{n,k})$.

(b) $k = \infty$: The “Fenchel” conditions hold:

$$\int_{0}^{\infty} \frac{\exp(-x/y)}{y \hat{f}_{n,\infty}(x)} d\hat{F}_n(x) \leq 1$$

for all $y > 0$ with equality if and only if $y \in \text{supp}(\hat{G}_{n,\infty})$. 
B. Existence and uniqueness of MLE,  
\( k - \text{monotone}, \ k \geq 2 \)

To state the characterization of the LSE \( \tilde{f}_n \) we define \( Y_{n,k} \) and \( \tilde{H}_{n,k} \) by:

\[
Y_{n,k}(x) \equiv \int_0^x \int_0^{x_{k-1}} \cdots \int_0^{x_2} \int_0^{x_1} F_n(x_1) dx_1 dx_2 \cdots dx_{k-1},
\]

\[
\tilde{H}_{n,k}(x) \equiv \int_0^x \int_0^{x_{k-1}} \cdots \int_0^{x_2} \int_0^{x_1} \tilde{f}_{n,k}(x_0) dx_0 dx_2 \cdots dx_{k-1}
\]

for \( x \geq 0 \).

**Theorem.** (a) \( \tilde{f}_{n,k} \) is characterized by:

\[
\tilde{H}_{n,k}(x) \geq Y_{n,k}(x) \quad \text{for all} \quad x \geq 0.
\] (4)

with equality holding if and only if \( x \in \text{supp}(\tilde{G}_{n,k}) \).

(b) The equality conditions can be expressed as

\[
\int_0^\infty (\tilde{H}_{n,k}(y) - Y_{n,k}(y)) d\tilde{H}_{n,k}^{(2k-1)}(y) = 0.
\]
It is not hard to see that

\[ Y_{n,k}(y) = \int_0 y \frac{(y - x)^{k-1}}{(k-1)!} dF_n(x) = \int_0^\infty \frac{(y - x)^{k-1}}{(k-1)!} dF_n(x), \]

\[ \tilde{H}_{n,k}(y) = \int_0 y \frac{(y - x)^{k-1}}{(k-1)!} d\tilde{F}_{n,k}(x) = \int_0^\infty \frac{(y - x)^{k-1}}{(k-1)!} d\tilde{F}_{n,k}(x) \]

Thus the inequality part of the second theorem can be rewritten as

\[ \int_0^\infty \frac{(y - x)^{k-1}}{(k-1)!} d\left(\tilde{F}_{n,k}(x) - F_n(x)\right) \geq 0 \quad \text{for all } y > 0. \]

Similarly, the equality part of the first theorem can be rewritten as

\[ \int_0^\infty \frac{k(y - x)^{k-1}}{y^k \hat{f}_{n,k}(x)} d\left(\hat{F}_{n,k}(x) - F_n(x)\right) \geq 0 \quad \text{for all } y > 0. \]
Suppose that $X_1, \ldots, X_n$ are i.i.d. $P_0$ with density $p_0 \in \mathcal{P}$, a convex class of densities with respect to a $\sigma$–finite measure $\mu$ on a measurable space $(\mathcal{X}, \mathcal{A})$. Let

$$\hat{p}_n \equiv \arg\max_{p \in \mathcal{P}} \mathbb{P}_n \log(p).$$

For $0 < \alpha \leq 1$, let $\varphi_{\alpha}(t) = (t^\alpha - 1)/(t^\alpha + 1)$ for $t \geq 0$, $\varphi_{\alpha}(t) = -1$ for $t < 0$. Then $\varphi_{\alpha}$ is bounded and continuous for each $\alpha \in (0, 1]$. For $0 < \beta < 1$ define

$$h_\beta^2(p, q) \equiv 1 - \int p^\beta q^{1-\beta} d\mu.$$

Note that $h_{1/2}(p, q) \equiv H(p, q)$ is the Hellinger distance between $p$ and $q$, and by Hölder’s inequality, $h_\beta(p, q) \geq 0$ with equality if and only if $p = q$ a.e. $\mu$. 
**C: Consistency, $k$–monotone, $2 \leq k \leq \infty$**

**Proposition:** (Pfanzagl; van de Geer) Suppose that $\mathcal{P}$ is convex. Then

$$h_{1-\alpha/2}^2(\hat{p}_n, p_0) \leq (\mathbb{P}_n - P_0) \left( \varphi_{\alpha} \left( \frac{\hat{p}_n}{p_0} \right) \right).$$

In particular, when $\alpha = 1$ we have, with $\varphi \equiv \varphi_1$,

$$H^2(\hat{p}_n, p_0) \leq (\mathbb{P}_n - P_0) \left( \varphi \left( \frac{\hat{p}_n}{p_0} \right) \right) = (\mathbb{P}_n - P_0) \left( \frac{2\hat{p}_n}{\hat{p}_n + p_0} \right).$$

**Corollary:** (Pfanzagl (1988); van de Geer, (1993, 1996)) Suppose that $\{\varphi_{\alpha}(p/p_0) : p \in \mathcal{P}\}$ is a $P_0$ Glivenko-Cantelli class. Then for each $0 < \alpha \leq 1$, $h_{1-\alpha/2}(\hat{p}_n, p_0) \rightarrow_{a.s.} 0$. 
C: Consistency, \( k \)-monotone, \( 2 \leq k \leq \infty \)

**Proof.** Since \( \mathcal{P} \) is convex and \( \hat{p}_n \) maximizes \( P_n \log p \) over \( \mathcal{P} \), it follows that

\[
P_n \log \frac{\hat{p}_n}{(1-t)\hat{p}_n + tp_1} \geq 0
\]

for all \( 0 \leq t \leq 1 \) and every \( p_1 \in \mathcal{P} \); this holds in particular for \( p_1 = p_0 \). Note that equality holds if \( t = 0 \). Differentiation of the left side with respect to \( t \) at \( t = 0 \) yields

\[
P_n \frac{p_1}{\hat{p}_n} \leq 1 \quad \text{for every} \quad p_1 \in \mathcal{P}.
\]

If \( L : (0, \infty) \to \mathbb{R} \) is increasing and \( t \mapsto L(1/t) \) is convex, then Jensen’s inequality yields

\[
P_n L \left( \frac{\hat{p}_n}{p_1} \right) \geq L \left( \frac{1}{P_n (p_1/\hat{p}_n)} \right) \geq L(1) = P_n L \left( \frac{p_1}{p_1} \right).
\]

Choosing \( L = \varphi_\alpha \) and \( p_1 = p_0 \) in this last inequality and noting that \( L(1) = 0 \), it follows that

\[
0 \leq P_n \varphi_\alpha(\hat{p}_n/p_0) = (P_n - P_0) \varphi_\alpha(\hat{p}_n/p_0) + P_0 \varphi_\alpha(\hat{p}_n/p_0); \quad (5)
\]
C: Consistency, $k$–monotone, $2 \leq k \leq \infty$

see van der Vaart and Wellner (1996) page 330, and Pfanzagl (1988), pages 141 - 143. Now we show that

$$P_0 \varphi_\alpha(p/p_0) = \int \frac{p^\alpha - p_0^\alpha}{p^\alpha + p_0^\alpha} dP_0 \leq - \left( 1 - \int p_0^\beta p^{1-\beta} d\mu \right)$$

(6) for $\beta = 1 - \alpha/2$. Note that this holds if and only if

$$-1 + 2 \int \frac{p^\alpha}{p_0^\alpha + p^\alpha} d\mu \leq -1 + \int p_0^\beta p^{1-\beta} d\mu,$$

or

$$\int p_0^\beta p^{1-\beta} d\mu \geq 2 \int \frac{p^\alpha}{p_0^\alpha + p^\alpha} p_0 d\mu.$$

But his holds if

$$p_0^\beta p^{1-\beta} \geq 2 \frac{p^\alpha p_0}{p_0^\alpha + p^\alpha}.$$
With $\beta = 1 - \alpha/2$, this becomes

$$
\frac{1}{2}(p_0^\alpha + p^\alpha) \geq p_0^{\alpha/2} p^{\alpha/2} = \sqrt{p_0^\alpha p^\alpha},
$$

and this holds by the arithmetic mean - geometric mean inequality. Thus (6) holds. Combining (6) with (5) yields the claim of the proposition. The corollary follows by noting that $\varphi(t) = (t - 1)/(t + 1) = 2t/(t + 1) - 1$. 

Statistical Seminar, Fréjus
C: Consistency, \( k \)-monotone, \( 2 \leq k \leq \infty \)

To apply this to the MLEs \( \hat{f}_{n,k} \in D_k \), we take \( \mathcal{P} = D_k \), which is convex in view of the mixture representation.

We first show that the map \( G \mapsto f_G(x) \) is continuous with respect to the topology of vague convergence for distributions \( G \). This follows easily since for each fixed \( x > 0 \) the kernels

\[
y \mapsto ky^{-1}(1 - x/y)^{k-1} \equiv m_k(x, y)
\]

for this mixing family are bounded, continuous, and satisfy \( m_k(x, y) \to 0 \) as \( y \to 0 \) or \( \infty \) for every \( x > 0 \). Since vague convergence of distribution functions implies that integrals of bounded continuous functions vanishing at infinity converge, it follows that \( G \mapsto f_G(x) \) is continuous with respect to the vague topology for every \( x > 0 \). This implies, that the family

\[
\mathcal{F}_k = \left\{ \frac{f_G}{f_G + f_0} : G \text{ a d.f. on } \mathbb{R}^+ \right\}
\]
is pointwise, for a.e. \( x \), continuous in \( G \) wrt the vague topology. Since the family of sub-distribution functions \( G \) on \( \mathbb{R} \) is compact for the vague topology (Bauer (1972), p. 241), and the family of functions \( F_k \) is uniformly bounded by 1, we conclude from the argument of Wald (1949) that

\[
N[\epsilon, F_k, L_1(P_0)] < \infty \quad \text{for every } \epsilon > 0.
\]

Thus \( F_k \) is \( P_0 \)-Glivenko-Cantelli and we conclude that \( \hat{f}_{n,k} = f_{\hat{G}_n} \) satisfies

\[
H(\hat{f}_{n,k}, f_0) \to_{a.s.} 0.
\]

The same argument works for \( k = \infty \) and yields a different proof of a result of Jewell (1982).
Based on the bound

\[ f(x) \leq \frac{1}{x} \left( 1 - \frac{1}{k} \right)^{k-1} \quad \text{for all } x > 0, \quad f \in D_k \]

and subsequence arguments, it follows that for each \( c > 0 \)

\[
\sup_{x \geq c} |\hat{f}_{n,k}(x) - f_{0,k}(x)| \to 0 \quad \text{as } n \to \infty,
\]

\[
\sup_{x \geq c} |\hat{f}^{(j)}_{n,k}(x) - f_{0,k}^{(j)}(x)| \to 0 \quad \text{as } n \to \infty, 1 \leq j \leq k - 1, \quad \text{and}
\]

\[
\hat{f}^{(k-1)}_{n}(x) \to_{a.s.} f_{0,k}^{(k-1)}(x) \quad \text{if the derivative } f_{0,k}^{(k-1)}(x) \text{ exists.}
\]

What about rates of convergence?
D: Global Rates, $2 \leq k < \infty$

Based on:

- Empirical process fluctuation bound: Birgé & Massart; van der Vaart & W
- Rate of convergence result: Birgé & Massart; van der Vaart & W (1996)
- Entropy bound for bounded sub-classes of $\mathcal{D}_k$: Gao & W (2009)
- If $f_0(0) < \infty$, then $\hat{f}_{n,k}(0) = O_p(1)$. Gao & W (2009)
Empirical process result:
Suppose that $\mathcal{P}$ is a collection of densities, $\mathcal{P}_0 \subset \mathcal{P}$

**Theorem.** (Thm 3.2.5, vdV & W, simplified) Suppose that $X_1, \ldots, X_n$ are i.i.d. $P_0$ with density $p_0 \in \mathcal{P}_0$. Let $H$ be the Hellinger distance between densities, and let $m_p$ be defined, for $p \in \mathcal{P}$, by

$$m_p(x) = \log \left( \frac{(p(x) + p_0(x))}{(2p_0(x))} \right).$$

Then

$$M(p) - M(p_0) \equiv P_0(m_p - m_{p_0}) \lesssim -H^2(p, p_0).$$

Furthermore, with $\mathcal{M}_\delta = \{m_p - m_{p_0} : H(p, p_0) \leq \delta, \ p \in \mathcal{P}_0\}$, we also have

$$E_{P_0}^*\|G_n\|_{\mathcal{M}_\delta} \lesssim \tilde{J}_[](\delta, \mathcal{P}_0, H) \left( 1 + \frac{\tilde{J}_[](\delta, \mathcal{P}_0, H)}{\delta^2 \sqrt{n}} \right) \equiv \phi_n(\delta, \mathcal{P}_0), \quad (7)$$

where

$$\tilde{J}_[](\delta, \mathcal{P}_0, H) = \int_0^\delta \sqrt{1 + \log N[]}(\epsilon, \mathcal{P}_0, H) \, d\epsilon.$$
D: Global Rates, $2 \leq k < \infty$

**Entropy bound for bounded sub-classes of $\mathcal{D}_k$:** Let

$$\mathcal{P}_0 \equiv \mathcal{D}_k^B([0, A]) \equiv \{ f \in \mathcal{D}_k : f(0) \leq B, f(x) = 0 \text{ if } x > A \}.$$  

Gao & W (2009) show that for $\epsilon > 0$

$$\log N_{[\cdot]}(\epsilon, \mathcal{D}_k^B([0, A]), H) \leq C\epsilon^{-1/k}$$

where $C = C_k(A, B)$.

**If** $f_0(0) < \infty$, **then** $\hat{f}_{n,k}(0) = O_p(1)$. By the characterization of $\hat{f}_{n,k}$,

$$1 \geq \int_0^y \frac{k (y - x)^{k-1}}{y^k \hat{f}_{n,k}(x)} d\mathbb{F}_n(x) \quad \text{for all } y > 0$$

with equality if $y \in \{\tau_1, \ldots, \tau_m\} \equiv \text{supp}(\hat{G}_{n,k})$ where $0 < \tau_1 < \cdots < \tau_m < \infty$. Thus for $y = \tau_1$ and $0 \leq x \leq \tau_1$,

$$1 = \frac{k}{\tau_1} \int_0^{\tau_1} \frac{(1 - x/\tau_1)^{k-1}}{\hat{f}_{n,k}(x)} d\mathbb{F}_n(x)$$
D: Global Rates, $2 \leq k < \infty$

where

$$\hat{f}_{n,k}(x) = \int_{0}^{\infty} \frac{k}{y} \left(1 - \frac{x}{y}\right)^{k-1} d\widehat{G}_{n,k}(y)$$

$$\geq \left(1 - \frac{x}{\tau_1}\right)^{k-1} + \int_{0}^{\infty} \frac{k}{y} d\widehat{G}_{n,k}(y) = (1 - x/\tau_1)^{k-1} \hat{f}_{n,k}(0).$$

Hence

$$1 \leq \frac{k}{\tau_1} \int_{0}^{\tau_1} \frac{(1 - x/\tau_1)^{k-1}}{\hat{f}_{n,k}(0)(1 - x/\tau_1)^{k-1}} d\overline{F}_n(x) = \frac{k}{\tau_1 \hat{f}_{n,k}(0)} \overline{F}_n(\tau_1),$$

which yields

$$\hat{f}_{n,k}(0) \leq k \frac{\overline{F}_n(\tau_1)}{\tau_1} \leq k \sup_{t > 0} \frac{\overline{F}_n(t)}{t} \leq k \sup_{t > 0} \frac{\overline{F}_n(t)}{F_0(t)} f_0(0) = O_p(1).$$

Statistical Seminar, Fréjus 3.23
Combining these facts proves:

**Theorem. (Gao & W, 2009)** Suppose that $f_0 \in \mathcal{D}_k^B([0, A])$ for some $0 < A, B < \infty$. Then $\{\hat{f}_{n,k}\}$ satisfies

$$H(\hat{f}_{n,k}, f_0) = O_p(n^{-\frac{k}{2k+1}}).$$

**Questions:**

- What is the rate for $\hat{f}_{n,\infty} \in \mathcal{D}_\infty$?
- Can we go beyond $\mathcal{D}_k^B([0, A])$?
- Is $n^{-1/(2k+1)}$ the rate of convergence of $d_{BL}(\hat{G}_{n,k}, G_0)$?
E: Rates of convergence: local results, $k = 2$

- Difficulty: no switching relation! Study LSE as first step.

- Proceed by localizing the Fenchel conditions
  - Step 1: localization rate or tightness result
    Empirical process theory: Kim-Pollard type lemmas
  - Step 2: Weak convergence of the localized driving process to a limit Gaussian driving process
    Empirical process theory: bracketing CLT with functions dependent on $n$.
  - Step 3: Preservation of (localized) characterizing relations in the limit.
  - Step 4: Establishing uniqueness of the limiting (Gaussian world) estimator resulting from the Fenchel relations.

- Cross check limit distributions with lower bound theory.
Step 1: Localization:

- Fenchel characterization implies midpoint properties.
- Midpoint properties + Kim-Pollard type lemma implies gap rate.
- Gap rate $\tau_n^+ - \tau_n^- = O_p(n^{-1/5})$ yields tightness.
Mid-point properties: Recall the Fenchel characterization of the LSE, $k = 2$:

$$\tilde{H}_n(x) \geq Y_n(x) \text{ for all } x \geq 0.$$  \hspace{1cm} (8)

with equality holding if and only if $x \in \text{supp}(\tilde{G}_n)$.

(b) The equality conditions can be expressed as

$$\int_0^{\infty} (\tilde{H}_n(y) - Y_n(y)) \, d\tilde{H}_n(3)(y) = 0.$$

It follows that $\tilde{H}_n$ is piecewise cubic: for $\tau_1 < \tau_2$, with $\tau_1, \tau_2 \in \text{supp}(\tilde{G}_{n,2})$ two successive touch points,

$$\tilde{H}_n(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 \text{ on } [\tau_1, \tau_2]$$

where $a_0, a_1, a_2, a_3$ are determined by

$$\tilde{H}_n(\tau_j) = Y_n(\tau_j), \quad j = 1, 2, \quad \text{and}$$

$$\tilde{F}_n(\tau_j) = F_n(\tau_j), \quad j = 1, 2.$$
Upshot: for $x \in [\tau_1, \tau_2]$

\[
\tilde{H}_n(x) = \left\{ Y_n(\tau_2)(x - \tau_1) + Y_n(\tau_1)(\tau_2 - x) \right\}/\Delta \tau \\
- \frac{1}{2} \left\{ \frac{\Delta F_n}{\Delta \tau} + \frac{4(\bar{F}_n \Delta \tau - \Delta Y_n)(x - \bar{\tau})}{(\Delta \tau)^3} \right\} (x - \tau_1)(x - \tau_2),
\]

so, with $\bar{\tau} \equiv (\tau_2 + \tau_1)/2$ and $\Delta \tau \equiv \tau_2 - \tau_1$,

\[
\tilde{H}_n(\bar{\tau}) = \bar{Y}_n - \frac{1}{8} \Delta F_n \Delta \tau
\]

where

\[
\Delta Y_n \equiv Y_n(\tau_2) - Y_n(\tau_1), \quad \Delta F_n \equiv \bar{F}_n(\tau_2) - \bar{F}_n(\tau_1), \\
\bar{Y}_n \equiv (Y_n(\tau_2) + Y_n(\tau_1))/2, \quad \bar{F}_n \equiv (\bar{F}_n(\tau_2) + \bar{F}_n(\tau_1))/2.
\]

Now we can rewrite $\tilde{H}_n(\bar{\tau}) \geq \bar{Y}_n(\bar{\tau})$ as

\[
\bar{Y}_n - \frac{1}{8} \Delta F_n \Delta \tau \geq \bar{Y}_n(\bar{\tau}),
\]
Now let $x_0$ with $f_0^{(2)}(x_0) > 0$ be fixed, let $\xi_n \to x_0$, and take

$$\tau_1 \equiv \tau_n^- \equiv \max\{t \in \text{supp}(\tilde{G}_n) : t \leq \xi_n\},$$

$$\tau_2 \equiv \tau_n^+ \equiv \min\{t \in \text{supp}(\tilde{G}_n) : t > \xi_n\}.$$  

Then $\tilde{H}_n(\bar{\tau}_n) \geq Y_n(\bar{\tau}_n)$ can be rewritten as

$$\frac{1}{2} \left( Y_n(\tau_n^+) + Y_n(\tau_n^-) \right) - \frac{1}{8} \left\{ F_n(\tau_n^+) - F_n(\tau_n^-) \right\} (\tau_n^+ - \tau_n^-) \geq Y_n(\bar{\tau}_n).$$  

Replacing $Y_n$ and $F_n$ by their deterministic counterparts and then expanding the integrands at $\bar{\tau}_n$ yields

$$\int_{\tau_n^-}^{\tau_n^+} (\tau_n^+ - x) f_0(x) dx + \int_{\tau_n^-}^{\tau_n^+} (x - \tau_n^-) f_0(x) dx - \frac{1}{4} (\tau_n^+ - \tau_n^-) \int_{\tau_n^-}^{\tau_n^+} f_0(x) dx$$

$$= \int_{[\tau_n^-, \tau_n^+]} \left\{ \frac{1}{2} (\bar{\tau}_n + \tau_n^+) - x \right\} f_0(x) dx + \int_{[\tau_n^-, \bar{\tau}_n]} \left\{ x - \frac{1}{2} (\tau_n^- + \bar{\tau}_n) \right\} f_0(x) dx$$

$$= -\frac{1}{192} f_0(\bar{\tau}_n)(\tau_n^+ - \tau_n^-)^4 + o_P(\tau_n^+ - \tau_n^-)^4,$$

by using consistency of $\tilde{f}_n$ to ensure that $\bar{\tau}_n$ belongs to a sufficiently small neighborhood of $x_0$.  

Statistical Seminar, Fréjus 3.29
E: Rates of convergence: local results, $k = 2$
The difference between (9) and the deterministic version is

\[
\int_{[\tau_n^-, \bar{\tau}_n]} (z - (\tau_n^- + \bar{\tau}_n)/2)d(\mathbb{F}_n(z) - F_0(z)) \\
+ \int_{[\bar{\tau}_n, \tau_n^+]} ((\tau_n^+ + \bar{\tau}_n)/2 - z)d(\mathbb{F}_n(z) - F_0(z))
\]

\[\equiv U_n(\tau_n^-, \bar{\tau}_n) - U_n(\bar{\tau}_n, \tau_n^+) \quad \text{where}
\]

\[
U_n(x, y) \equiv \int_{[x, y]} (z - (x + y)/2)d(\mathbb{F}_n(z) - F_0(z)).
\]

By an empirical process argument – as in Kim and Pollard (1991), there exist constants \( \delta > 0 \) and \( c_0 > 0 \) such that, for each \( \epsilon > 0 \) and each \( x \) satisfying \( |x - x_0| < \delta \),

\[
|U_n(x, y)| \leq \epsilon|y - x|^4 + O_p(n^{-4/5}), \quad \text{for all } 0 \leq y - x \leq c_0.
\]

This implies that

\[
|U_n(\tau_n^-, \bar{\tau}_n) - U_n(\bar{\tau}_n, \tau_n^+)| \leq \epsilon(\tau_n^+ - \tau_n^-)^4 + O_p(n^{-4/5}).
\]
E: Rates of convergence: local results, $k = 2$

Putting the pieces together by choosing $\epsilon = f_0^{(2)}(x_0)/384$ it follows that

$$-rac{1}{192} f_0^{(2)}(x_0)(\tau^+_n - \tau^-_n)^4 + o_p(\tau^+_n - \tau^-_n)^4$$

$$+ \frac{1}{384} f_0^{(2)}(x_0)(\tau^+_n - \tau^-_n)^4 + O_p(n^{-4/5}) \geq 0,$$

and hence

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

This leads to:

**Proposition:** Suppose that $f_0'(x_0) < 0$, $f_0^{(2)}(x_0) > 0$ and $f_0^{(2)}$ continuous in a neighborhood of $x_0$. Then

$$\sup_{|t| \leq M} |\tilde{f}_n(x_0 + n^{-1/5}t) - f_0(x_0) - n^{-1/5}tf_0'(x_0)| = O_p(n^{-2/5}),$$

$$\sup_{|t| \leq M} |\tilde{f}'_n(x_0 + n^{-1/5}t) - f_0'(x_0)| = O_p(n^{-2/5}).$$

... and a corresponding result for the MLE $\hat{f}_n$. 

Statistical Seminar, Fréjus 3.32
Step 2: Localize the Fenchel conditions

Define

\[ Y_{n}^{loc}(t) \equiv n^{4/5} \int_{x_0}^{x_0+n^{-1/5}t} \left\{ F_n(v) - F_n(x_0) \right\} dv \]

\[ - \int_{x_0}^{v} (f_0(x_0) + (u - x_0)f'_0(x_0))du \}

\[ \equiv d n^{3/10} \int_{x_0}^{x_0+n^{-1/5}} \left\{ U_n(F_0(v)) - U_n(F(x_0)) \right\} dv \]

\[ + \frac{f_0^{(2)}(x_0)}{4!} t^4 + o(1) \]

\[ \sim \sqrt{f_0(x_0)} \int_0^t W(s)ds + \frac{f_0^{(2)}(x_0)}{4!} t^4 \text{ by KMT} \]

or by theorem 2.11.22 or 2.11.23, vdV & W (1996)

\[ \equiv a \int_0^t W(s)ds + bt^4 \equiv Y_{a,b}(t). \]
F: Limiting distributions at a fixed point: \( k = 2 \)

Similarly, define

\[
\tilde{H}_n^{loc}(t) \equiv n^{4/5} \int_{x_0}^{x_0 + n^{-1/5}t} \int_{x_0}^{v} \{\tilde{f}_n(u) - f_0(x_0) - (u - x_0)f_0'(x_0)\} \, du \, dv \\
+ \tilde{B}_n t + \tilde{A}_n
\]

where

\[
\tilde{A}_n \equiv n^{4/5}(\tilde{H}_n(x_0) - \Upsilon_n(x_0)) = O_p(1) \\
\tilde{B}_n \equiv n^{3/5}(\tilde{F}_n(x_0) - \mathbb{F}_n(x_0)) = O_p(1).
\]

Furthermore

\[
\tilde{H}_n^{loc}(t) - \Upsilon_n^{loc}(t) = n^{4/5}\{\tilde{H}_n(x_0 + n^{-1/5}t) - \Upsilon_n(x_0 + n^{-1/5}t)\} \geq 0
\]
**F: Limiting distributions at a fixed point: \( k = 2 \)**

**Step 3: Preservation of (localized) characterizing relations in the limit**

- \( \{ (\tilde{H}_{loc}^1, \tilde{H}_{loc}^2, \tilde{H}_{loc}^3) \}_{n \geq 1} \) is tight.
- \( Y_{loc} \xrightarrow{\text{a.s.}} Y_{a,b} \).
- Fenchel relations satisfied:
  - \( \tilde{H}_{loc}^n(t) \geq Y_{loc}^n(t) \) for all \( t \)
  - \( \int_{-\infty}^{\infty} (\tilde{H}_{loc}^n(t) - Y_{loc}^n(t)) d\tilde{H}_{loc}^{(3)}(t) = 0 \)
- Any limit process \( H \) for a subsequence \( \{ \tilde{H}_{loc}^n \} \) must satisfy
  - \( H(t) \geq Y_{a,b}(t) \) for all \( t \).
  - \( \int_{-\infty}^{\infty} (H(t) - Y_{a,b}(t)) dH^{(3)}(t) = 0 \).
- Show the process \( H \) characterized by these two conditions is unique!
**F: Limiting distributions at a fixed point: \( k = 2 \)**

Upshot after rescaling to \( \mathbb{Y}_{1,1} \equiv \mathbb{Y} \):

**Theorem.** (Groeneboom, Jongbloed & W (2001)) If \( f \in \mathcal{D}_2, f_0(x_0) > 0, f_0^{(2)}(x_0) > 0 \), and \( f_0^{(2)} \) continuous in a neighborhood of \( x_0 \), then

\[
\left( \begin{array}{c}
\frac{n^{2/5}}{5}(\tilde{f}_n(x_0) - f(x_0)) \\
\frac{n^{1/5}}{5}(\tilde{f}'_n(x_0) - f'(x_0))
\end{array} \right) \rightarrow_d \left( \begin{array}{c}
c_1(f)H^{(2)}(0) \\
c_2(f)H^{(3)}(0)
\end{array} \right)
\]

where

\[
c_1(f) \equiv \left( \frac{f^2(x_0)f''(x_0)}{4!} \right)^{1/5}, \quad c_2(f) \equiv \left( \frac{f(x_0)f''(x_0)^3}{4!^3} \right)^{1/5}.
\]
Limiting distributions at a fixed point: $k = 2$
**F:** Limiting distributions at a fixed point: $k = 2$
F: Limiting distributions at a fixed point: $k = 2$
F: Limiting distributions at a fixed point: $k = 2$