# 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 5

- A: Some multivariate shape-constrained classes
  - ▷ Log-concave & $h-$convex on $\mathbb{R}^d$
  - ▷ Block-decreasing on $\mathbb{R}^d_+$
  - ▷ Scale mixtures of uniform on $\mathbb{R}^d_+$
  - ▷ $h-$convex on $\mathbb{R}^d_+$ with $h$ increasing.
- B: Review of available theory; MLEs for multivariate classes.
- C: Some alternative classes in $\mathbb{R}$:
  Bondesson's hyperbolically monotone classes.
- D: Open problems and questions: $\mathbb{R}^1$
- E: Some problems and questions: $\mathbb{R}^d$

# A. Some multivariate shape-constrained classes

- "Block decreasing" densities on $R^{+d} = [0, \infty)^d$
- Monotone with non-negative increments on rectangles (as for a multivariate d.f.)
- Convex and decreasing
- $k-$monotone; completely monotone
- log$-$concave; $s-$concave; $h-$ transform of convex (or concave)

# A. Some multivariate shape-constrained classes

**Block-decreasing densities on** $\mathbb{R}^{+d} = [0, \infty)^d$

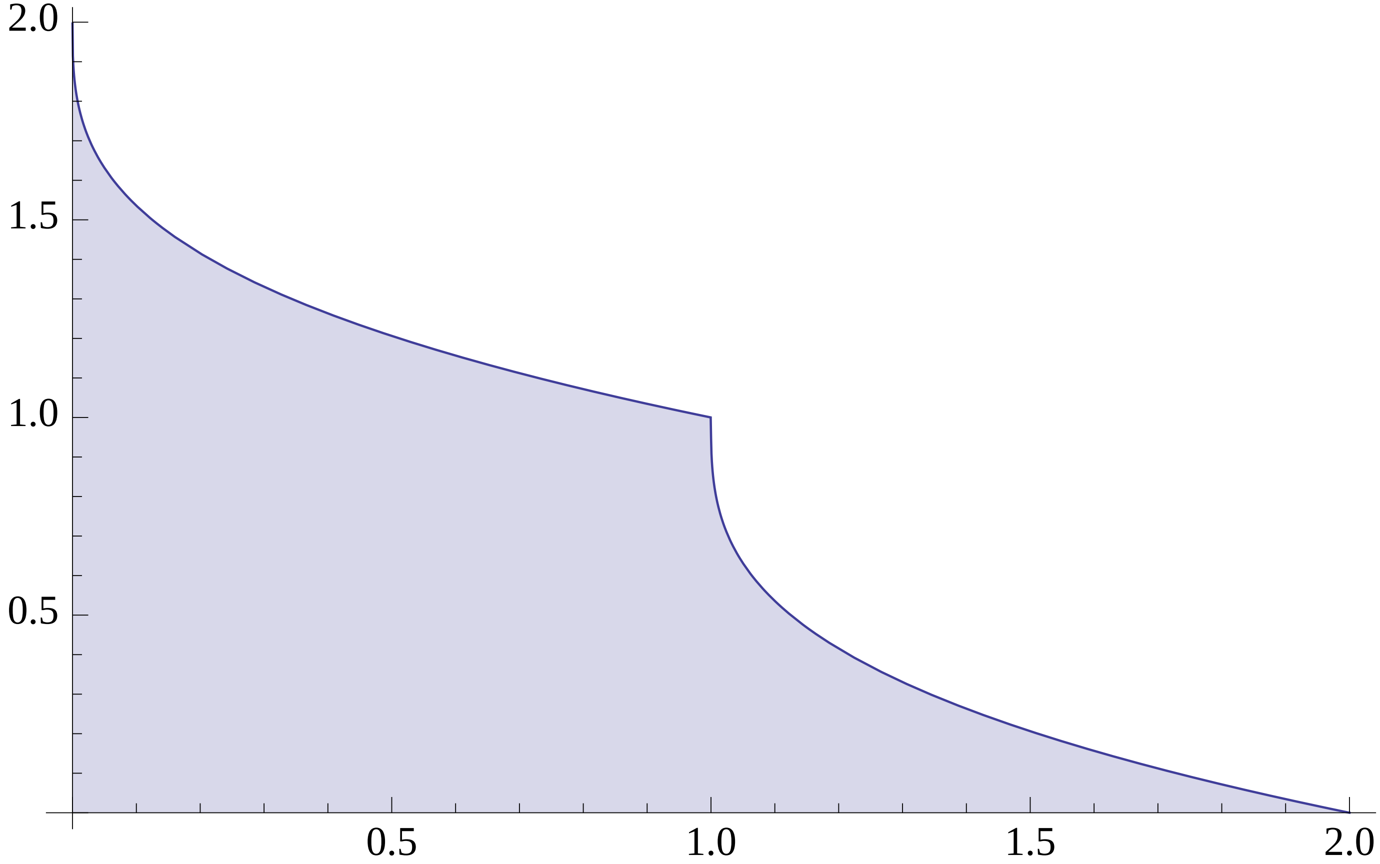
- $$\mathcal{BD}(\mathbb{R}^d) \;=\; \left\{ \begin{array}{l} f : [0,\infty)^d \to \mathbb{R}^+ \Big| \int f(\underline{x})d\underline{x} = 1, \\ f(\underline{x} + h\underline{e}_j) \le f(x) \\ \text{for all basis vectors } \underline{e}_j, j = 1, \ldots, d, h > 0 \end{array} \right\}.$$

- Polonik (1995, 1998): MLE exists and coincides with "silhouette estimator".

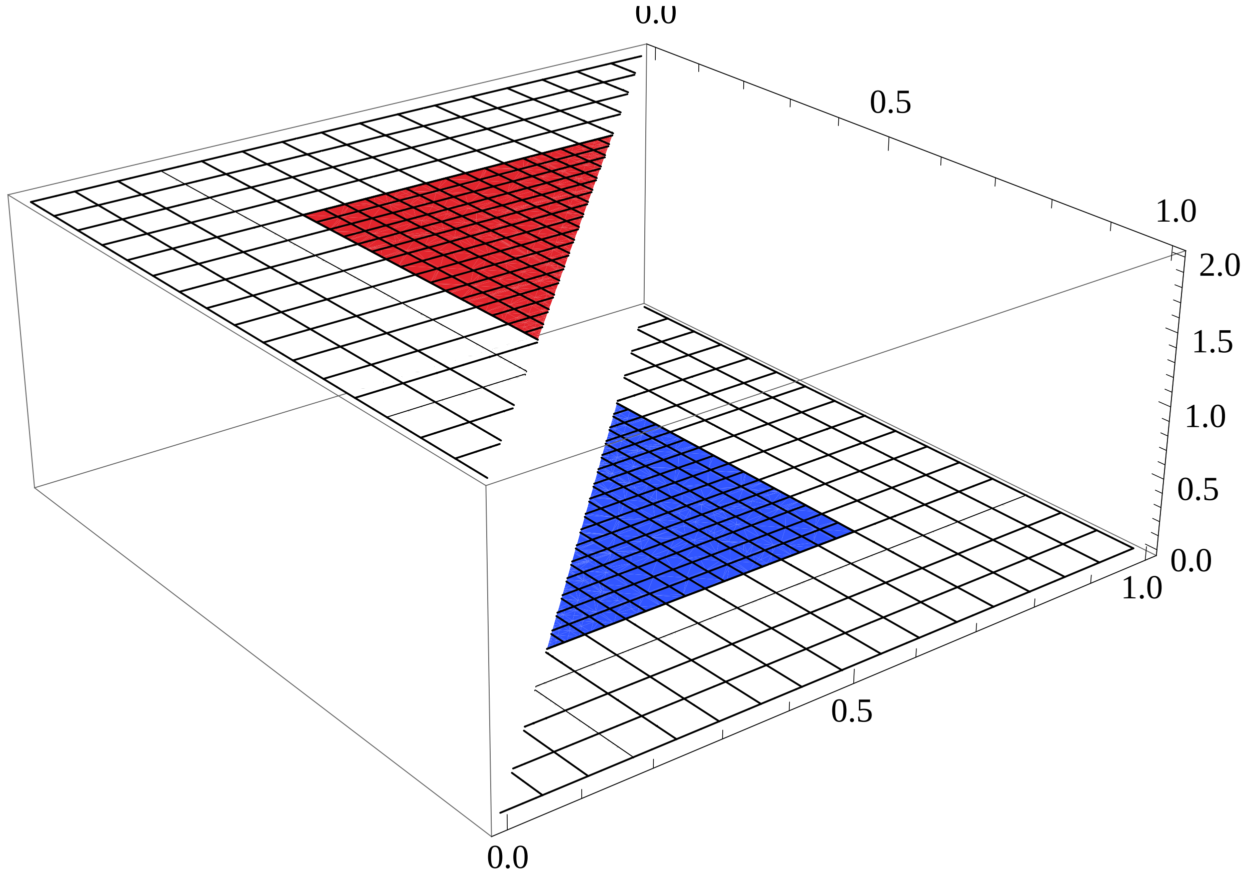
- Biau and Devroye (2003): global minimax lower bounds ... and showed that a generalization of Birgé's histogram estimator acheives the bounds.

- Pavlides (2008, 2009): asymptotic minimax lower bounds for estimation of $f(x_0)$

# A. Some multivariate shape-constrained classes



# A. Some multivariate shape-constrained classes



# A. Some multivariate shape-constrained classes

**Monotone with non-negative increments on rectangles (as for a multivariate d.f.)**

= "Scale mixtures of uniform densities" on $\mathbb{R}^{+d}$

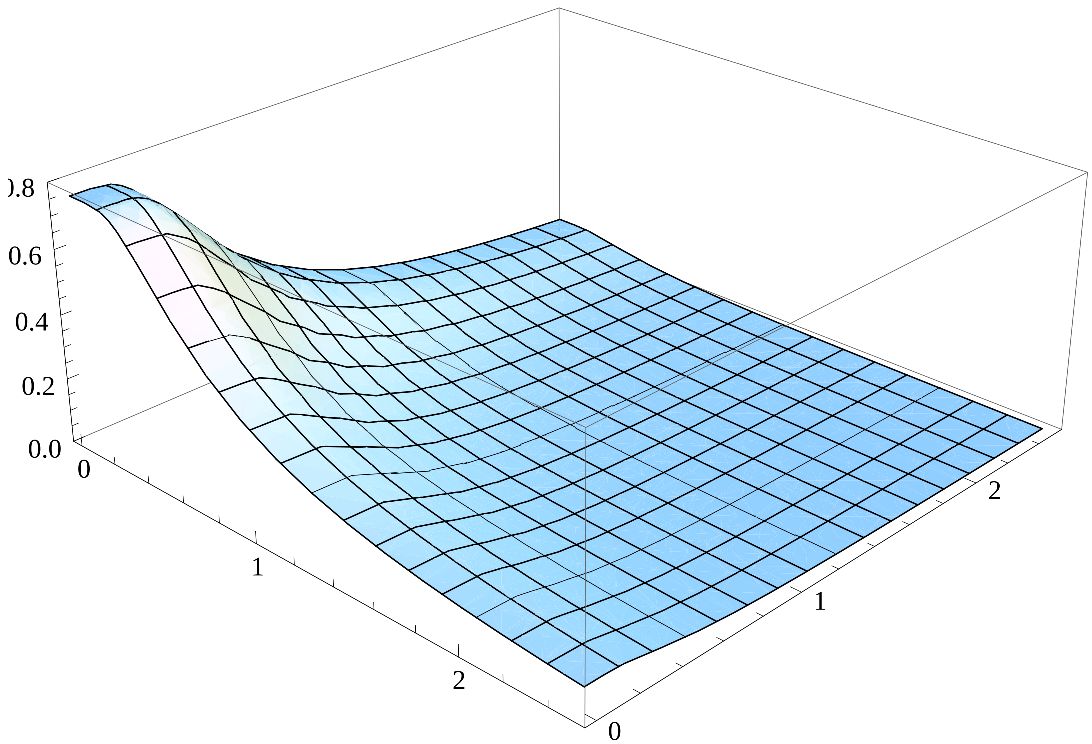
$$f(\underline{x}) = \int_{\mathbb{R}^{+d}} \frac{1}{\prod_{j=1}^{d} y_j} 1_{[\underline{0},\underline{y}]}(\underline{x}) dG(\underline{y})$$

for some probability distribution $G$ on $\mathbb{R}^{+d}$.

Example: $dG(y_1, y_2) = (y_1 y_2)^{-2} g(1/y_1, 1/y_2, \theta) dy_1 dy_2$ with

$$g(u, v, \theta) = \{(1 + \theta u)(1 + \theta v) - \theta\} \text{exp}(-u - v - \theta uv), \qquad \theta = .4,$$

# A. Some multivariate shape-constrained classes



# A. Some multivariate shape-constrained classes

**Convex and decreasing on $\mathbb{R}^d$**

- Seregin (2010)'s increasing convex transformed classes with $h(x) = x$, so that $f(x) = \varphi(x)$ with $\varphi$ convex (or convex and decreasing).

Example: $f(x) = \exp(-|x|) 1_{(0,\infty)^d}(x)$.

# A. Some multivariate shape-constrained classes

**Log-concave densities on $\mathbb{R}^d$**

- $$f(\underline{x}) = \exp(\varphi(\underline{x})) = \exp(-(-\varphi(\underline{x}))$$

  where $\varphi : \mathbb{R}^d \mapsto \mathbb{R}$ is concave (so $-\varphi$ is convex).

- Exponentially decaying tails; does not include multivariate $t-$densities.

# A. Some multivariate shape-constrained classes

- $s-$convex densities and $h-$ convex densities
(Koenker and Mizera; Seregin, Seregin and Wellner)

$$f(\underline{x}) = h(\varphi(\underline{x}))$$

where $\varphi : \mathbb{R}^d \mapsto \mathbb{R}$ is convex, $h : \mathbb{R} \mapsto \mathbb{R}^+$ is decreasing and continuous; e.g. $h_s(u) \equiv (1 + u/s)^{-s}$ with $s > d$.
Larger classes than log-concave: includes multivariate $t_n$ for $d < s \leq n + d$.

# B: Review of available theory:

# MLEs for multivariate classes

Block decreasing densities on $\mathbb{R}^{+d}$

| Problem | Lower Bound | Upper Bound / MLE |
|---|---|---|
| A<br>(consist) | | Polonik (1995,1998)<br>Pavlides (2008?) |
| B<br>(local) | Pavlides (2008 & 2009)<br>rate: $n^{1/(d+2)}$<br>$\left\{\prod_{j=1}^{d} \frac{\partial f}{\partial x_j}(x) f(x)\right\}^{1/(d+2)}$ | ?<br>rate: $n^{1/(d+2)}$ ??<br>const: ?? |
| C<br>(global) | Biau and Devroye (2003)<br>rate: $n^{1/(d+2)}$ | Biau and Devroye (2003)<br>analogues of Birgé's<br>histogram estimators |

# B: Review of available theory:

# MLEs for multivariate classes

Scale mixtures of uniform on $\mathbb{R}^{+d}$

| Problem | Lower Bound | Upper Bound / MLE |
|---|---|---|
| A<br>(consist) | | Pavlides (2008)<br>Pavlides & W (2010) |
| B<br>(local) | Pavlides (2008)<br>rate: $n^{1/3}$ (all $d$)<br>$\left\{ \frac{\partial^d f}{\partial x_1 \cdots \partial x_d}(x) f(x) \right\}^{1/3}$ | ?<br>Pavlides (2008),<br>partial results |
| C<br>(global) | ??<br>(hints from entropy bounds<br>of Blei, Gao, and Li)<br>?? | ??<br>??<br>?? |

# B: Review of available theory:

# MLEs for multivariate classes

Log concave densities on $\mathbb{R}^d$

| Problem | Lower Bound | Upper Bound / MLE |
| --- | --- | --- |
| A<br>(consist) | consistency with<br>misspecification!<br><br>computation: | Cule and Samworth (2010a)<br>Schumacher, Rufibach,<br>Samworth (2010)<br>Cule, Samworth, Stewart<br>(2010) |
| B<br>(local) | Seregin (2010)<br>rate: $n^{2/(d+4)}$<br>$\left\{ f^{d+2}(x) \text{curv}_x(\varphi) \right\}^{1/(d+4)}$<br>$\text{curv}_x(\varphi) = \det \nabla^2 \varphi(x)$ | ??<br>??<br>?? |
| C<br>(global) | ??<br>conjectures:<br>Seregin and W (2010) | ??<br>??<br>?? |

# B: Review of available theory:
# MLEs for multivariate classes

$s-$convex and $h-$convex densities on $\mathbb{R}^d$

| Problem | Lower Bound | Upper Bound / MLE |
|---|---|---|
| A (consist) | computation: (related estimators) (convex regression) | Seregin & W (2010) Koenker & Mizera (2010) Seijo and Sen (2010) |
| B (local) | Seregin (2010) rate: $n^{2/(d+4)}$ $\left\{\frac{f(x)\text{curv}_x(\varphi)}{h'(\varphi(x))^4}\right\}^{1/(d+4)}$ | MLE & LSE rate inefficient $d > 4$ ?? ?? |
| C (global) | ?? ?? ?? | ?? LSE rate inefficient, $d > 4$?? or $d \geq 4$?? |

## C: Some alternative classes on $\mathbb{R}$: Bondesson's hyperbolically monotone classes

Let $k \geq 1$ be an integer. $f : (0, \infty) \to \mathbb{R}^+$ is *hyperbolically monotone* of order $k$ ($HM_k$) if, for each fixed $u > 0$, the function

$$H(w) \equiv f(uv)f(u/v), \quad w \equiv \frac{1}{2}\left(v + \frac{1}{v}\right) \geq 1,$$

is such that

$$\begin{aligned} &(-1)^j H^{(j)}(w) \geq 0, \quad \text{for } j = 0, \ldots, k-1, \\ &(-1)^k H^{(k-1)}(w) \quad \text{is right continuous and decreasing.} \end{aligned}$$

If $f$ is hyperbolically monotone for all $k$, $f$ is said to be *hyperbolically completely monotone* (HCM or $HM_\infty$).

Note that $f(uv)f(u/v)$ is always a function of $w$ by symmetry, and for $v \geq 1$ $v = w + \sqrt{w^2 - 1}$.

# C: Some alternative classes on $\mathbb{R}$:

# Bondesson's hyperbolically monotone classes

**Example 1. Half-normal distributions**

$$f(x) = \sqrt{\frac{2}{\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) 1_{(0,\infty)}(x)$$

has

$$\begin{aligned} H(w) &= f(uv)f(u/v) = \frac{2}{\pi\sigma^2} \exp\left(-\frac{u^2}{2\sigma^2}\left(v + \frac{1}{v}\right)^2 + \frac{u^2}{\sigma^2}\right) \\ &= \frac{2}{\pi\sigma^2} \exp\left(-\frac{u^2}{2\sigma^2} w^2 + \frac{u^2}{\sigma^2}\right). \end{aligned}$$

For fixed $u > 0$ $w \mapsto H(w)$ is decreasing, but $-H'$ is not. Thus $f \in HM_1$ while $f \notin HM_2$.

**Example 2. Uniform**$(a, b)$ If $f(x) = (b-a)^{-1} 1_{(a,b)}(x)$ with $0 \le a < b$, then for $v \ge 1$

$$H(w) = f(uv)f(u/v) = (b-a)^{-2} 1\{a < u/v \le uv < b\}$$

so $H(w) = 1\{1 \le v < V_u \equiv \min\{u/a, b/u\}\}$. Thus $H$ is decreasing and $f \in HM_1$.

# C: Some alternative classes on $\mathbb{R}$:

# Bondesson's hyperbolically monotone classes

**Exercise 1.** Show that $f(x) = Cx^{\beta-1}(1+cx)^{-\gamma}1_{(0,\infty}(x)$, with $\beta, \gamma, c \geq 0$ and $C$ a normalizing constant, satisfies $f \in HM_\infty$.

**Exercise 2.** (log-normal) $f(x) = C\exp(-(\log x - \mu)^2/2\sigma^2)$ satisfies $f \in HM_\infty$.

**Exercise 3.** $f(x) = (a-x)_+^{\gamma-1}1_{(0,\infty)}(x)$ satisfies $f \in HM_{\lfloor\gamma\rfloor}$.

**Theorem 1. (Bondesson, 1997)** If $X$ and $Y$ are independent random variables such that $X \sim f \in HM_k$ and $Y \sim g \in HM_k$, then $XY \sim HM_k$ and $X/Y \sim HM_k$.

**Theorem 2. (Bondesson, 1992)** $X \sim f \in HM_1$ if and only if $\log X \sim e^x f(e^x)$ is log-concave.

# C: Some alternative classes on $\mathbb{R}$:

# Bondesson's hyperbolically monotone classes

**Putting these two results together:**

- Transform the hyperbolically monotone classes from $\mathbb{R}_+$ to $\mathbb{R}$:

$$\begin{aligned} \mathcal{HM}_k \circ \exp &\equiv \{g(x) = e^x f(e^x) : \quad f \in HM_k\} \\ &\equiv \text{log-hyperbolically } k\text{−monotone} \end{aligned}$$

- $\mathcal{HM}_k \circ \exp$ is closed under convolution.
- $\mathcal{HM}_k \circ \exp$ have the same degree of smoothness as the $k+1$-monotone densities
- $\mathcal{HM}_\infty \circ \exp$ contains the Gaussian distributions (by Exercise 2).

**Conclusion:** The classes $\mathcal{HM}_k \circ \exp \searrow \mathcal{HM}_\infty \circ \exp$ provide a nice analogue of the $k-$monotone classes on $\mathbb{R}^+$ for $\mathbb{R}$ with nice closure properties.

# D: Open problems and questions: $\mathbb{R}^1$

- Are there "natural" switching relations for the $k-$monotone MLE's and / or LSE's?
- More connections to convexity theory?
- Pointwise rates of convergence for the $k-$monotone MLE's?
- Rates of convergence under degenerate mixing, $G = \delta_1$?
- Rates of convergence for the MLE's of $G$ (inverse problems)?
- Global rates of convergence in $L_1$ and Hellinger metrics, log-concave classes?
- Theory for natural discrete shape-constrained classes? (Monotone, convex-decreasing, completely monotone, .... )
- MLE's for Bondesson's $HM_k$ classes?

# E: Open problems and questions: $\mathbb{R}^d$

- local rates and global rates for shape constrained estimators in $\mathbb{R}^d$?

- Local (pointwise) limiting distribution theory for MLE's and other natural divergence-based estimators?

- When are the MLE's rate (in-)efficient?
  Conjecture 1: Block decreasing: inefficient for $d > 2$.
  Conjecture 2: Log-concave and $s-$concave: inefficient for $d > 4$.

- How to penalize or sieve or ... to obtain rate efficient estimators in these classes for higher dimensions?

- Do there exist natural shape-constraints with smoothness $> 2$ for which MLE's are rate-efficient and which have natural preservations properties under convolution, marginalization, and so forth?

# E: Open problems and questions: $\mathbb{R}^d$

- Faster and more efficient algorithms?

- Faster and more efficient algorithms?!

Je vous remerci!