Theory for multivariate shape constraints:

partial results, open problems, and conjectures

Jon A. Wellner

University of Washington

- Talk at JSM, Vancouver
  August 1, 2010
- *Email: jaw@stat.washington.edu*
  http: //www.stat.washington.edu/jaw/jaw.research.html
- Based on joint work with former Ph.D. students and postdocs:
  Marios Pavlides, Arseni Seregin

Outline

1. Shape constraints in $\mathbb{R}^1$ and $\mathbb{R}^d$: some background (and limitations).
2. Goals for nonparametric estimation theory
3. Review of results for estimating a monotone density in $\mathbb{R}^1$
4. Review of recent shape-constrained estimation progress for $\mathbb{R}^d$
5. Summary: available theory for $\mathbb{R}^d$
6. Open problems and some conjectures

1. Shape constraints in $\mathbb{R}^1$ and $\mathbb{R}^d$: background

$\mathbb{R}^1$: Several types of shape constraints, and several types of target functions to estimate:

- type of shape constraint:
  - monotone
  - convex (or concave)
  - $k-$monotone
  - completely monotone
  - $\log-$concave and $s-$concave
  - hyperbolically monotone (Bondesson)

1. Shape constraints in $\mathbb{R}^1$ and $\mathbb{R}^d$: background

$\mathbb{R}^1$: Several types of shape constraints, and several types of target functions to estimate:

- type of shape constraint:
  - monotone
  - convex (or concave)
  - $k-$monotone
  - completely monotone
  - $\log -$concave and $s-$concave
  - hyperbolically monotone (Bondesson)
- Type of function:
  - regression function $r(x)$
  - density function $f(x)$
  - hazard functions, $\lambda(x)$

Focus here on density functions — and on methods related or connected to Maximum Likelihood.

$\mathbb{R}^d$: More types of shape constraints, and (more?) types of target functions to estimate:

- Types of shape constraint:
  - block monotone or coordinate-wise monotone
  - monotone with non-negative increments on rectangles (as for a multivariate d.f.)
  - convex (or concave) and decreasing
  - $k-$monotone; completely monotone
  - $\log -$concave and $s-$concave
  - $h-$ transform of convex (or concave)

Focus here on density functions — and on methods related or connected to Maximum Likelihood.
$\mathbb{R}^d$: More types of shape constraints, and (more?) types of target functions to estimate:

- Types of shape constraint:
    - block monotone or coordinate-wise monotone
    - monotone with non-negative increments on rectangles (as for a multivariate d.f.)
    - convex (or concave) and decreasing
    - $k-$monotone; completely monotone
    - $\log -$concave and $s-$concave
    - $h-$ transform of convex (or concave)
- Types of function: regression r, density f, hazard $\lambda \ldots$.

2. Goals for limit theory

- A. Consistency (pointwise or with respect to some metric or topology)?
- B. Local (at fixed points) rates of convergence? Upper and lower bounds?
- C. Global (summary or metric) rates of convergence? Upper and lower bounds?
- D. Rates for smooth functionals: upper and lower bounds?
- E. Rates and corrections for inconsistency at boundary points?
- F. Kiefer-Wolfowitz type theorems: at what rate does the integrated version of the estimator "look like the natural empirical estimator"?

My talk today: limited to A-C.

3. Review for estimating a monotone density function

A. Marshall's lemma. Consistency "easy" now via Pfanzagl (1989), van de Geer (1993), and empirical process methods.

B_L: Local asymptotic minimax lower bound, Groeneboom (198?):
if $f(x_0) > 0$, $f'(x_0) < 0$ and f' continuous at x_0, then for any estimator T_n of $f(x_0)$,

$$\liminf_{n\to\infty} \sup_{f:H(f,f_0)\le cn^{-1/2}} n^{1/3} E_f|T_n - f(x_0)| \ge \frac{e^{-1/3}}{4}(2|f'(x_0)|f(x_0))^{1/3}.$$

B^U: Local convergence theorem, Prakasa Rao (1970):
if $f(x_0) > 0$, $f'(x_0) < 0$, and f' is continuous at x_0, then

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

where $\mathbb{S}(0)$ is the (left-)slope at zero of the least concave majorant of $W(t) - t^2$, and W is two-sided BM starting at 0.

C$_L$. Birgé (1987). Let $\mathcal{F}$ denote the class of all decreasing densities f on $[0,1]$ satisfying $f \le M$ with $M > 1$. Then the minimax risk for $\mathcal{F}$ with respect to the L_1 metric $d_1(f,g) \equiv \int |f(x) - g(x)| dx$ based on n observations is

$$R_M(d_1, n) \equiv \inf_{\hat{f}_n} \sup_{f \in \mathcal{F}} E_f d_1(\hat{f}_n, f).$$

Then there is an absolute constant C such that

$$R_M(d_1, n) \ge C \left(\frac{\log M}{n} \right)^{1/3}.$$

C$_U$. Birgé (1989). Let $\hat{f}_n$ denote the Grenander estimator of $f \in \mathcal{F}$. Then

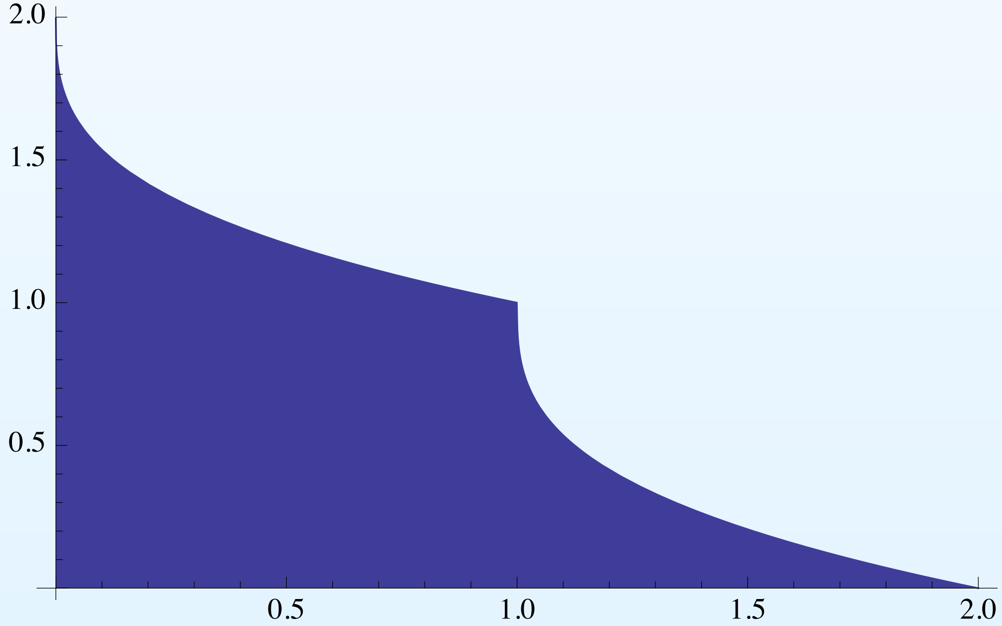
$$\sup_{f \in \mathcal{F}_M} E_f d_1(\hat{f}_n, f) \le 4.75 \left(\frac{\log M}{n} \right)^{1/3}.$$

Table 1: Montone density on $\mathbb{R}^+$

Problem	Lower Bound	Upper Bound / MLE
A (consist)		Grenander? Marshall? ? or Prakasa Rao
B (local theory)	Groeneboom (?) rate: $n^{1/3}$ const: $(\lvert f'(x)\rvert f(x))^{1/3}$	Prakasa Rao (1969) rate: $n^{1/3}$ const: $(\lvert f'(x)\rvert f(x))^{1/3}$
C (global theory)	Birgé (1987) Groeneboom (1986)	Birgé (1989)

4. Shape-constrained estimation on $\mathbb{R}^d$

- "Block decreasing" densities on $\mathbb{R}^{+d} = [0, \infty)^d$
  (Polonik; Biau and Devroye; Pavlides)
  Decreasing along along all lines parallel to coordinate axes.
  Can be viewed as the convex hull of densities which are uniform on (compact) lower layers:



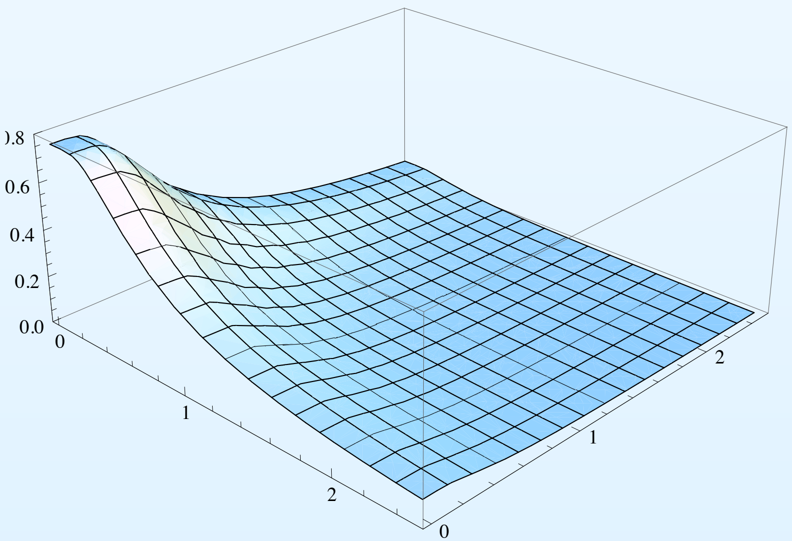
- “Scale mixtures of uniform densities” on $\mathbb{R}^{+d}$: (Pavlides; Pavlides and Wellner):

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

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

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



- Log-concave densities on $\mathbb{R}^d$
  (Cule, Samworth, Stewart; Koenker and Mizera; Seregin and Wellner)

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

  where $\varphi : \mathbb{R}^d \mapsto \mathbb{R}$ is concave (so $-\varphi$ is convex).
  Necessarily exponentially decaying tails; does not include multivariate $t-$densities.

- $s-$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 $d < s \leq n + d$.

Table 2: Block decreasing densities on $\mathbb{R}^{+d}$

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

Table 3: Scale mixtures of uniform on $\mathbb{R}^{+d}$

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

Table 4: Log concave densities on $\mathbb{R}^d$

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

Table 5: $s-$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: $n^{2/(d+4)}$ $\left\{ \frac{f(x)\mathrm{curv}_x(\varphi)}{h'(\varphi(x))^4} \right\}^{1/(d+4)}$	?? LSE rate inefficient $d > 4$?? ??
C (global)	?? ?? ??	?? LSE rate inefficient, $d > 4$?? or $d \geq 4$??

5. Open problems and some conjectures

- 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 $d > 4$.
- 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?
  Conjecture 3: Yes for $d = 1$ via the hyperbolic $k-$monotone classes of Bondesson?