Estimation
Under Shape Constraints
Monotone, Convex, and Beyond

Jon A. Wellner University of Washington

based on joint work with Piet Groeneboom,
Geurt Jongbloed,
and Fadoua Balabdaoui

Talk to be given November 22, 2004
Paris

Email: jaw@stat.washington.edu

*http: //www.stat.washington.edu/
jaw/jaw.research.html*

Outline

1. Introduction:
Montone and convex densities as mixtures and the inverse problems

2. Questions, Problems, Issues: A-G

3. Current state, Problems A-G,
Monotone Case

4. Current state, Problems A-G,
Convex Case

5. Current state, Problems A-G,
$k-$monotone Case

6. Summary: Current States

7. Hermite Interpolation Problems
and Rates

1. Introduction: Montone and convex densities as mixtures

I. Monotone densities as mixtures of uniforms

$$\begin{aligned} f(x) &= \int_0^\infty \frac{1}{y} 1_{[0,y]}(x)\,dG(y) = \int_x^\infty \frac{1}{y}\,dG(y) \\ F(x) &= x f(x) + G(x) \end{aligned}$$

Inverting to get G yields:

$$G(x) = F(x) - x f(x)\,.$$

Alternatively, for G with finite mean

$$\mu(G) = \int_0^\infty x dG(x)\,,$$

$$\begin{aligned} f(x) &= \frac{1}{\mu(G)} \int_0^\infty 1_{[0,y]}(x) dG(y) = \frac{1 - G(x)}{\mu(G)} \\ F(x) &= \frac{1}{\mu(G)} \int_0^x (1 - G(y)) dy\,. \end{aligned}$$

Inverting to get G yields:

$$1 - G(x) = \frac{f(x)}{f(0)}\,.$$

II. Convex densities as mixtures of triangulars

$$
\begin{aligned}
f(x) &= \int_0^\infty \frac{2}{y^2}(y-x)1_{[0,y]}(x)\,dG(y) \\
&= \int_x^\infty \frac{2}{y}(y-x)\,dG(y) \\
F(x) &= xf(x) + x^2 \int_x^\infty \frac{1}{y^2}\,dG(y) + G(x)
\end{aligned}
$$

Inverting to get G yields:

$$
G(x) = F(x) - xf(x) + \frac{1}{2}x^2 f'(x)\,.
$$

Alternatively, for G with finite second moment

$$\mu_2(G) = \int_0^\infty x^2 dG(x)\,,$$

$$\begin{aligned} f(x) &= \frac{2}{\mu_2(G)} \int_0^\infty (y-x) 1_{[0,y]}(x) dG(y) \\ &= \frac{2}{\mu_2(G)} \int_x^\infty (1-G(y)) dy \\ F(x) &= \frac{2}{\mu_2(G)} \int_0^x \int_z^\infty (1-G(y)) dy dz\,. \end{aligned}$$

Inverting to get G yields:

$$1 - G(x) = \frac{f'(x)}{f'(0)}\,.$$

General story for $k \geq 1$:

$$
\begin{aligned}
f(x) &= \int_0^\infty \frac{k}{y^k}(y-x)^{k-1} 1_{[0,y]}(x)\,dG(y) \\
&= \int_x^\infty \frac{k}{y^k}(y-x)^{k-1}\,dG(y) \\
F(x) &= G(x) + \int_x^\infty [1-(1-x/y)^k]\,dG(y)
\end{aligned}
$$

Inverting to get G yields:

$$
\begin{aligned}
G(x) &= F(x) - xf(x) + \frac{1}{2}x^2 f'(x) \\
&\quad + \cdots + \frac{(-1)^k}{k!}x^k f^{(k-1)}(x)
\end{aligned}
$$

Alternatively,

$$f(x) = \frac{k}{\mu_k(G)} \int_x^\infty (y-x)^{k-1} dG(y)\,,$$

and the corresponding inversion formula for G is given by

$$1 - G(x) = \frac{f^{(k-1)}(x)}{f^{(k-1)}(0)}\,.$$

2. Questions, Problems, and Issues

A. Nonparametric ML Estimation of
$f \in \mathcal{F}_1$, $\mathcal{F}_2$, or $\mathcal{F}_k$?

B. Nonparametric ML Estimation of G or G_A?

C. Asymptotic distribution theory for nonparametric ML estimators in A and B?

D. Likelihood ratio tests for $f(x_0) = f_0(x_0)$?

E. Behavior of global functionals such as

$$\int |\widehat{f}_n(x) - f(x)|^r dx$$

or

$$\sup_x |\widehat{f}_n(x) - f(x)| \, ?$$

F. Attainment of Minimax Bounds for f?
For G?

G. [LR-based] Confidence bands for f?
For G?

3. Current State, Problems A-F monontone case

A. Nonparametric ML Estimator of f:

Grenander (1956)

$$\widehat{f}_n = \text{left derivative of the least concave majorant of } \mathbb{F}_n, \text{ the empirical distribution of } X_1, \ldots, X_n \text{ i.i.d. } F$$

Other related work:
Ayer, Brunk, Ewing, Reid, Silverman (1955);
van Eeden (1956), (1957).

B. Inversion to estimate G and G_A?

Estimation of G:

$$\widehat{G}_n(x) = \widehat{F}_n(x) - x\widehat{f}_n(x)\,,$$

Estimation of G_A:
Woodroofe and Sun (1993), Sun and Woodroofe (1996). Penalized version of the Grenander estimator – and hence an estimator of $f(0)$. This yields an estimator of G_A:

$$1 - \widehat{G}_A(x) = \frac{\widehat{f}_n^{WS}(x)}{\widehat{f}_n^{WS}(0)}\,.$$

Kulikov (2002): use $\widehat{f}_n(n^{-\alpha})$ to estimate $f(0)$, $\alpha \geq 1/3$.

C. Asymptotic distribution theory at fixed points:

Prakasa Rao (1969)
Groeneboom (1985), (1988)

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

$$\begin{aligned} 2Z &= \text{slope at 0 of the least concave majorant of } W(t) - t^2 \\ &\stackrel{d}{=} \text{slope at 0 of the greatest convex minorant of } W(t) + t^2 \end{aligned}$$

Crux: Understanding limit Gaussian estimation problem!

$$X(t) = t^2 + W(t)$$

$$dX(t) = 2t\,dt + dW(t)$$

D. Likelihood ratio tests for $f(x_0) = f_0(x_0)$?

Banerjee and Wellner (2001)
- different monotone function model

Banerjee (2003?)

E. Behavior of global functionals such as

$$\int |\widehat{f}_n(x) - f(x)|^r dx$$

$r = 1$: Groeneboom (1985),
Groeneboom, Hooghiemstra,
and Lopuhaä (1999).

$$\sup_x |\widehat{f}_n(x) - f(x)|\,?$$

start: Hooghiemstra and Lopuhaä (1999)

$$n^{1/6}\left\{ n^{1/3}\int_0^1 |\widehat{f}_n(x) - f(x)|\,dx - \mu_1 \right\} \rightarrow_d N(0, \sigma_1^2)$$

where

$$\begin{aligned} \mu_1 &= 2E|V(0)|\int_0^1 \left|\frac{1}{2}f(x)f'(x)\right|^{1/3} dx \\ \sigma_1^2 &= 8\int_0^\infty Cov(|V(0)|, |V(c)-c|)\,dc\,. \end{aligned}$$

$$V(c) = \sup\{t : W(t) - (t-c)^2 \ \text{ is maximal}\}\,.$$

Kulikov (2002): for $1 \le r < 5/2$

$$n^{1/6}\left\{ n^{1/3}\left\{ \int_0^1 |\widehat{f}_n(x) - f(x)|^r \, dx \right\}^{1/r} - \mu_r \right\}$$
$$\to_d N(0, \sigma_r^2)$$

The restriction $r < 5/2$ is related to the inconsistency of $\widehat{f}_n(0)$.

F. Attainment of Minimax Bounds?

- For f: Birgé (1986), (1987), (1989)
- For G: corollary of Birgé
- For G_A ? bounds?

G. [LR-based] Confidence bands for f?
For G?

Hengartner and Stark (1999);
Dümbgen (1998)

nothing on LR-based confidence bands

4. Current State, Problems A-G convex case

A. Nonparametric ML Estimator of f:

Hampel (1987), Anevski (1994), Jongbloed (1995). The MLE $\widehat{f}_n$ is piecewise linear, convex, and characterized by:

$$\widehat{H}_n(t, \widehat{f}_n) \begin{cases} \leq t^2/2, & t \geq 0 \\ = t^2/2, & \widehat{f}_n'(t-) < \widehat{f}_n'(t+) \end{cases}$$

where

$$\widehat{H}_n(t, f) = \int_0^t \frac{t-u}{f(u)} d\mathbb{F}_n(u)\,.$$

Groeneboom, Jongbloed, Wellner (2001)

B. Inversion to estimate G and G_A?

Estimation of G:

$$\widehat{G}_n(x) = \widehat{F}_n(x) - x\widehat{f}_n(x) + \frac{1}{2}x^2 \widehat{f}_n'(x)\,.$$

Estimation of G_A: nothing yet!!

C. Asymptotic distribution theory at fixed points:

Groeneboom, Jongbloed, Wellner (2001)

$$n^{2/5}(\widehat{f}_n(x) - f(x)) \rightarrow_d \left|\frac{1}{24}f^2(x)f''(x)\right|^{1/5} H''(0)$$
$$n^{1/5}(\widehat{f}_n'(x) - f'(x)) \rightarrow_d \left|\frac{1}{24^3}f(x)f''(x)^3\right|^{1/5} H^{(3)}(0)$$

$$\begin{aligned} H &= \text{“invelope” of the process} \\ & \quad Y(t) = \int_0^t W(s)ds + t^4 \end{aligned}$$

$$\begin{aligned} X(t) &= Y'(t) = 4t^3 + W(t) \\ dX(t) &= 12t^2 dt + dW(t)\,. \end{aligned}$$

Crux: Understanding limit Gaussian estimation problem!

D. Likelihood ratio tests for $f(x_0) = f_0(x_0)$?

nothing yet!

E. Behavior of global functionals such as

$$\int |\widehat{f}_n(x) - f(x)|^r dx$$

or

$$\sup_x |\widehat{f}_n(x) - f(x)| \, ?$$

nothing yet

F. Attainment of global minimax bounds?

nothing yet ...

G. [LR-based] Confidence bands for f?
For G?

nothing yet ...

5. Current state, Problems A-G, $k-$monotone Case

A. Nonparametric ML Estimator of f:

Balabdaoui (2004):
For $k \geq 2$ the MLE $\widehat{f}_n$ is piecewise polynomial of degree $k-1$, with $(-1)^k \widehat{f}_n^{(k-2)}$ convex, and characterized by:

$$\widehat{H}_n(t,\widehat{f}) \begin{cases} \le t^k/k, & t \ge 0 \\ = t^k/k, & (-1)^{k-1}\widehat{f}_n^{(k-1)}(t-) \\ & < (-1)^{k-1}\widehat{f}_n^{(k-1)}(t+) \end{cases}$$

where

$$\widehat{H}_n(t,f) = \int_0^t \frac{(t-x)_+^{k-1}}{f(x)} d\mathbb{F}_n(x)\,.$$

Consistency of $\widehat{f}_n$:

Balabdaoui (2004);
Jewell (1982) for $k=\infty$;
Pfanzagl (1990); van de Geer (1993)?

B. Least Squares Estimator $\tilde{f}_n$ of f:
Minimize

$$\Phi_n(f) = \frac{1}{2}\int_0^\infty f^2(t)dt - \int_0^\infty f(t)d\mathbb{F}_n(t)$$

over the class of square integrable $k-$monotone functions on $(0,\infty)$ where $\mathbb{F}_n$ is the empirical distribution function of $X_1,\ldots,X_n$.

The LSE $\tilde{f}_n$ is a $k-$monotone spline of order $k-1$ that is $k-2$ times continuously differentiable at its (simple) knots. The number of knots is at most n and there exists a discrete measure $\tilde{G}$ with masses $c_1, \ldots, c_m > 0$ at $a_1, \ldots, a_m$ (all depending on $X_1, \ldots, X_n$) such that

$$\tilde{f}_n(x) = \int_0^\infty \frac{k(y-x)_+^{k-1}}{y^k} d\tilde{G}_n(y)\,.$$

Furthermore $\tilde{f}_n$ is the LSE if and only if

$$H_n(x) \begin{cases} \geq \mathbb{Y}_n(x), & \text{for all } x > 0 \\ = \mathbb{Y}_n(x), & \text{iff } x \in \{a_1, \cdots, a_m\} \end{cases} \tag{1}$$

where

$$\begin{aligned} \mathbb{Y}_n(x) &= \int_0^x \int_0^{t_{k-1}} \cdots \int_0^{t_2} \mathbb{F}_n(t_1) dt_1 \ldots dt_{k-1} \\ &= \int_0^x \frac{(x-t)^{k-1}}{(k-1)!} d\mathbb{F}_n(t) \end{aligned}$$

and

$$\begin{aligned} \tilde{H}_n(x) &= \int_0^x \int_0^{t_k} \cdots \int_0^{t_2} \tilde{f}_n(t_1) dt_1 \cdots dt_k \\ &= \int_0^x \frac{(x-t)^{k-1}}{(k-1)!} \tilde{f}_n(t) dt. \end{aligned}$$

It follows from the characterization (1) that

$$H_n(a_j) = \mathbb{Y}_n(a_j) \qquad \text{and}$$
$$H_n'(a_j) = \mathbb{Y}_n'(a_j),$$

for all $j = 1, \ldots, m$.

C. Asymptotic distribution at a fixed point?

Step 1: Asymptotic Minimax Lower Bounds:

- Fix $k \geq 2$ and $j \in \{0, \ldots, k-1\}$.
- Suppose $f^{(k)}(x_0) > 0$.

For any estimator $\widehat{T}_{n,j}$ of $f^{(j)}(x_0)$:

$$\sup_{\tau} \liminf_{n\to\infty} \sup_{f \in \mathcal{F}_{n,\tau}} n^{(k-j)/(2k+1)} E_f |\widehat{T}_{n,j} - f^{(j)}(x_0)| \geq d_{k,j} \left\{ |f^{(k)}(x_0)|^{2j+1} f(x_0)^{k-j} \right\}^{1/(2k+1)}$$

where $d_{k,j} > 0$, $j = 0, \ldots, k-1$.

Step 2: Conjecture:

$$\begin{pmatrix} n^{k/(2k+1)}(\widehat{f}_n - f^{(0)})(x_0) \\ n^{(k-1)/(2k+1)}(\widehat{f}_n^{(1)} - f^{(1)})(x_0) \\ \cdot \\ \cdot \\ \cdot \\ n^{1/(2k+1)}(\widehat{f}_n^{(k-1)} - f^{(k-1)})(x_0) \end{pmatrix}$$

$$\rightarrow_d \begin{pmatrix} A_{k,0}(f) H_k^{(k)}(0) \\ A_{k,1}(f) H_k^{(k+1)}(0) \\ \cdot \\ \cdot \\ \cdot \\ A_{k,k-1}(f) H_k^{(2k-1)}(0) \end{pmatrix}$$

where

$$A_{k,k-j} = \frac{\left(((-1)^k f^{(k)}(x_0))^{2j+1} f(x_0)^{k-j}\right)^{1/(2k+1)}}{((2k)!)^{(2k-1)/(2k+1)}}$$

and H_k is a piecewise polynomial function of degree $2k-1$ which satisfies

$$H_k(t) \geq Y_k(t) \quad \text{for all} \quad t \in R,$$
$$\int (H_k(t) - Y_k(t)) dH_k^{(2k-1)}(t) = 0\,,$$

and, for $t \geq 0$,

$$\begin{aligned} Y_k(t) \;&=\; \int_0^t \int_0^{s_{k-1}} \cdots \int_0^{s_1} W(s_1) ds_1 \ldots ds_{k-1} \\ &\quad + (-1)^k \frac{k!}{(2k)!} t^{2k} \end{aligned}$$

where W is a two-sided Brownian motion process starting from 0.

Note that

$$d^k Y_k(t) = (-1)^k t^k dt + dW(t)\,.$$

6. Summary current states
Problems A-G

Problem/k	1	2	k	∞
A	Y	Y	Y	Y
B	Y	Y	Y	Y
B (alt)	Y	N	N	N
C	Y	Y	N	N
D	Y	N	N	N
E	Y	N	N	N
F	Y	N	N	N
G	N	N	N	N

Problem/k	1	2
A: MLE of $f \in \mathcal{F}_k$	Grenander 1956	GJW 2001
B: MLE of B1: G B2: G_A	Woodroofe Sun (1993,’96)	GJW,’01 Hampel ’87 Anevski ’94 Jongbloed ’
C. Limit distrib. C2:	P.Rao 1969 Groeneboom ’85, ’89 WS ’93, ’96	GJW 2001 Balabdaoui 2004(?)
D: LR tests	Banerjee -Wellner ’01	
E: global functionals	Groeneboom ’85 GHL ’99 HL ’02	
F: minimax attained?	Birgé ’87 Birgé ’89	
G: conf. bands	Hengartner - Stark ’95 Dümbgen ’98	

7. The Hermite interpolation problem(s).

When $k = 2$, the processes H_n corresponding to the least squares estimator can be written down explicitly as a cubic spline in terms of the "knots": If τ_1 and τ_2 are two successive points of touch of H_n and $\mathbb{Y}_n$, then, letting $\overline{\tau} = (\tau_1 + \tau_2)/2$, $\overline{\mathbb{Y}'_n} = (\mathbb{Y}'_n(\tau_1) + \mathbb{Y}'_n(\tau_2))/2$, $\overline{\mathbb{Y}_n} = (\mathbb{Y}_n(\tau_1) + \mathbb{Y}_n(\tau_2))/2$, and
$\Delta\mathbb{Y}'_n = \mathbb{Y}'_n(\tau_2) - \mathbb{Y}'_n(\tau_1)$,
$\Delta\mathbb{Y}_n = \mathbb{Y}_n(\tau_2) - \mathbb{Y}_n(\tau_1)$,

$$
\begin{aligned}
H_n(t) &= \frac{\mathbb{Y}_n(\tau_2)(t-\tau_1) + \mathbb{Y}_n(\tau_1)(\tau_2 - t)}{\Delta\tau} \\
&\quad - \frac{1}{2}\left\{ \frac{\Delta\mathbb{Y}'_n}{\Delta\tau} + \frac{4(\overline{\mathbb{Y}'_n}\Delta\tau - \Delta\mathbb{Y}_n)(t-\overline{\tau})}{(\Delta\tau)^3} \right\} \\
&\quad \cdot (t-\tau_1)(\tau_2 - t)\,.
\end{aligned}
$$

What plays the role of this formula for general k?

Let $\tau_0 < \tau_1 < \cdots < \tau_{2k-3}$ be $2k-2$ successive jump points of $H_n^{(2k-1)}$; these are exactly the points of touch of H_n and $\mathbb{Y}_n$. Using the consequence (2) of the characterization of $\tilde{f}_n$, it turns out via the theorems of

Schoenberg and Whitney (1953) and
Karlin and Ziegler (1966)

that H_n is the unique spline of degree $2k-1$ with simple knots $\tau_0, \tau_1, \cdots, \tau_{2k-3}$ that solves the Hermite problem

$$H_n(\tau_j) = \mathbb{Y}_n(\tau_j), \text{ and } H_n'(\tau_j) = \mathbb{Y}_n'(\tau_j),$$

for $j = 0, \cdots, 2k-3$. By standard theory for Hermite interpolation problems (see e.g. Nürnberger (1989), pages 106 - 112), we can express the interpolating spline H_n as

$$H_n(t)$$
$$= \sum_{i=-(2k-1)}^{2k-4} \left(\sum_{j=0}^{2k-3} (a_{ij}\mathbb{Y}_n(\tau_j) + \ b_{ij}\mathbb{Y}_n'(\tau_j)) \right) B_i^{2k-1}(t)$$

where $\{B_i^{2k-1}: \ i \in \{-(2k-1), \ldots, 2k-4\}\}$ is the B-spline basis for the space of splines of degree $2k-1$ with simple knots $\tau_1, \ldots \tau_{2k-4}$, $A = (a_{ij})_{ij}$ and $B = (b_{ij})_{ij}$ are both $(4k-4) \times (k-1)$ sub-matrices obtained by extracting the odd and even columns of the inverse of the matrix obtained in the Hermite problem.

8. Selected References.

Balabdaoui, F. (2004). Nonparametric estimation of a k-monotone density: A new asymptotic distribution theory. Ph.D. dissertation, University of Washington, Department of Statistics.

Balabdaoui, F. and Wellner, J. A. (2004a). Estimation of a k-monotone density, part 1: characterizations, consistency, and minimax lower bounds. *Technical Report* **459**, Department of Statistics, University of Washington. Available at: *http://www.stat.washington.edu/www/research/repo*

Balabdaoui, F. and Wellner, J. A. (2004b). Estimation of a k-monotone density, part 2: algorithms for computation and numerical

results. *Technical Report* **460**, Department of Statistics, University of Washington. Available at:
http://www.stat.washington.edu/www/research/repo

Balabdaoui, F. and Wellner, J. A. (2004c). Estimation of a k-monotone density, part 3: limiting Gaussian versions of the problem; invelopes and envelopes Technical Report 461, Department of Statistics, University of Washington. Available at:
http://www.stat.washington.edu/www/research/repo

Balabdaoui, F. and Wellner, J. A. (2004d). Estimation of a k-monotone density, part 4: limit distribution theory and the spline connection. Technical Report 462, Department of Statistics, University of Washington. Available at:
http://www.stat.washington.edu/www/research/repo

Banerjee, M. and Wellner, J. A. (2000). Likelihood ratio tests for monotone functions. *Ann. Statist.* **29**, 1699-1731.

Birgé, L. (1987). Estimating a density under order restrictions: nonasymptotic minimax risk. *Ann. Statist.* **15**, 995 - 1012.

Birgé, L. (1989). The Grenander estimator: a nonasymptotic approach. *Ann. Statist.* **17**, 1532 - 1549.

Groeneboom, P. (1985). Estimating a monotone density. *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer,* Vol. II, Lucien M. LeCam and Richard A. Olshen eds.

Groeneboom, P. (1989). Brownian motion with a parabolic drift and Airy functions. *Probability Theory and Related Fields* **81**, 79 - 109.

Groeneboom, P. and Jongbloed, G. (1995). Isotonic estimation and rates of convergence in Wicksell's problem. *Ann. Statist.* **23**, 1518-1542.

Groeneboom, P., Jongbloed, G., and Wellner, J. A. (2001a). A canonical process for estimation of convex functions: The "invelope" of integrated Brownian motion $+t^4$. *Ann. Statist.* **29**, 1620-1652.

Groeneboom, P., Jongbloed, G., and Wellner, J. A. (2001b). Estimation of convex functions: characterizations and asymptotic theory. *Ann. Statist.* **29**, 1653-1698.

Groeneboom, P. and Wellner, J. A. (2001). Computing Chernoff's distribution. *J. Computational and Graphical Statistics* **10**, 388 - 400.

Hooghiemstra, G. and Lopuhaä, H. P. (1998). A extremal limit theorem for the argmax process of Brownian motion minus a parabolic drift. *Extremes* **1**, 215-240.

Sun, J. and Woodroofe, M. (1996). Adaptive smoothing for a penalized NPMLE of a non-increasing density. *J. Statist. Planning and Inference* **52**, 153-159.

Wellner, J. A. (2003). Gaussian white noise models: some results for monotone functions. *Crossing Boundaries: Statistical Essays in Honor of Jack Hall*, IMS Lecture Notes - Monograph Series, Vol. 43, 87-104. J.E. Kolassa and D. Oakes, editors.

Woodroofe, M. and Sun, J. (1993). A penalized maximum likelihood estimate of $f(0+)$ when f is non-increasing. *Statistica Sinica* **3**, 501-515.