Nonparametric estimation of $s$–concave and log-concave densities: alternatives to maximum likelihood

Jon A. Wellner

University of Washington, Seattle

Cemmap Seminar, London
October 27, 2015
Economics Seminar, University College London

Based on joint work with:

- Qiyang (Roy) Han
- Charles Doss
- Fadoua Balabdaoui
- Kaspar Rufibach
- Arseni Seregin
Outline

• A: Log-concave and $s$–concave densities on $\mathbb{R}$ and $\mathbb{R}^d$

• B: $s$–concave densities on $\mathbb{R}$ and $\mathbb{R}^d$

• C: Maximum Likelihood for log-concave and $s$–concave densities
  ▶ 1: Basics
  ▶ 2: On the model
  ▶ 3: Off the model

• D. An alternative to ML: Rényi divergence estimators
  ▶ 1. Basics
  ▶ 2. On the model
  ▶ 3. Off the model

• E. Summary: problems and open questions
A. Log-concave densities on $\mathbb{R}$ and $\mathbb{R}^d$

If a density $f$ on $\mathbb{R}^d$ is of the form

$$f(x) \equiv f_\varphi(x) = \exp(\varphi(x)) = \exp\left(-(-\varphi(x))\right)$$

where $\varphi$ is concave (so $-\varphi$ is convex), then $f$ is log-concave. The class of all densities $f$ on $\mathbb{R}^d$ of this form is called the class of log-concave densities, $\mathcal{P}_{\text{log-concave}} \equiv \mathcal{P}_0$.

Properties of log-concave densities:

- Every log-concave density $f$ is unimodal (quasi concave).
- $\mathcal{P}_0$ is closed under convolution.
- $\mathcal{P}_0$ is closed under marginalization.
- $\mathcal{P}_0$ is closed under weak limits.
- A density $f$ on $\mathbb{R}$ is log-concave if and only if its convolution with any unimodal density is again unimodal (Ibragimov, 1956).
• Many parametric families are log-concave, for example:
  ▶ Normal \((\mu, \sigma^2)\)
  ▶ Uniform\((a, b)\)
  ▶ Gamma\((r, \lambda)\) for \(r \geq 1\)
  ▶ Beta\((a, b)\) for \(a, b \geq 1\)

• \(t_r\) densities with \(r > 0\) are not log-concave.

• Tails of log-concave densities are necessarily sub-exponential.

• \(\mathcal{P}_{\text{log-concave}} = \) the class of “Polyá frequency functions of order 2”, \(PF_2\), in the terminology of Schoenberg (1951) and Karlin (1968). See Marshall and Olkin (1979), chapter 18, and Dharmadhikari and Joag-Dev (1988), page 150. for nice introductions.
B. $s$–concave densities on $\mathbb{R}$ and $\mathbb{R}^d$

If a density $f$ on $\mathbb{R}^d$ is of the form

$$f(x) \equiv f_\varphi(x) = \begin{cases} (\varphi(x))^{1/s}, & \varphi \text{ convex, if } s < 0 \\ \exp(-\varphi(x)), & \varphi \text{ convex, if } s = 0 \\ (\varphi(x))^{1/s}, & \varphi \text{ concave, if } s > 0, \end{cases}$$

then $f$ is $s$-concave.

The classes of all densities $f$ on $\mathbb{R}^d$ of these forms are called the classes of $s$–concave densities, $\mathcal{P}_s$. The following inclusions hold: if $-\infty < s < 0 < r < \infty$, then

$$\mathcal{P}_r \subset \mathcal{P}_0 \subset \mathcal{P}_s \subset \mathcal{P}_{-\infty}$$
Properties of \( s \)-concave densities:

- Every \( s \)-concave density \( f \) is quasi-concave.
- The Student \( t_\nu \) density, \( t_\nu \in \mathcal{P}_s \) for \( s \leq -1/(1+\nu) \). Thus the Cauchy density (\( = t_1 \)) is in \( \mathcal{P}_{-1/2} \subset \mathcal{P}_s \) for \( s \leq -1/2 \).
- The classes \( \mathcal{P}_s \) have interesting closure properties under convolution and marginalization which follow from the Borell-Brascamp-Lieb inequality: let \( 0 < \lambda < 1 \), \( -1/d \leq s \leq \infty \), and let \( f, g, h : \mathbb{R}^d \to [0, \infty) \) be integrable functions such that
  \[
  h((1 - \lambda)x + \lambda y) \geq M_s(f(x), g(x), \lambda) \quad \text{for all} \quad x, y \in \mathbb{R}^d
  \]
  where
  \[
  M_s(a, b, \lambda) = ((1 - \lambda)a^s + \lambda b^s)^{1/s}, \quad M_0(a, b, \lambda) = a^{1-\lambda} b^\lambda.
  \]
  Then
  \[
  \int_{\mathbb{R}^d} h(x) dx \geq M_s/(sd+1) \left( \int_{\mathbb{R}^d} f(x) dx, \int_{\mathbb{R}^d} g(x) dx, \lambda \right).
  \]
• If \( f \in \mathcal{P}_s \) and \( s > -1/(d+1) \), then \( E_f \parallel X \parallel < \infty \).

• If \( f \in \mathcal{P}_s \) and \( s > -1/d \), then \( \parallel f \parallel_\infty < \infty \).

• If \( f \in \mathcal{P}_0 \), then for some \( a > 0 \) and \( b \in \mathbb{R} \)

\[
f(x) \leq \exp(-a \parallel x \parallel + b) \text{ for all } x \in \mathbb{R}.
\]

• If \( f \in \mathcal{P}_s \) and \( s > -1/d \), then with \( r \equiv -1/s > d \), then for some \( a, b > 0 \)

\[
f(x) \leq (b + a \parallel x \parallel)^{-r} \text{ for all } x \in \mathbb{R}.
\]

• If \( s < -1/d \) then there exists \( f \in \mathcal{P}_s \) with \( \parallel f \parallel_\infty = \infty \).

• If \( -1/d < s \leq -1/(d+1) \), then there exists \( f \in \mathcal{P}_s \) with \( E_f \parallel X \parallel = \infty \).
\[ f(x) = \frac{1}{\sqrt{r} \text{Beta}(r/2, 1/2)} \left( \frac{r}{r + x^2} \right)^{(1+r)/2}, \]

with \( r = .05 \), so \( f \in \mathcal{P}_{-1/(1+.05)} \).
\[ f(x) = \frac{1}{2\Gamma(1/2)}|x|^{-1/2}\exp(-|x|), \quad \text{with} \quad f \in \mathcal{P}_{-2}. \]
C. Maximum Likelihood:

0-concave and $s$-concave densities

**MLE of $f$ and $\varphi$:** Let $\mathcal{C}$ denote the class of all concave functions $\varphi : \mathbb{R} \to [-\infty, \infty)$. The estimator $\hat{\varphi}_n$ based on $X_1, \ldots, X_n$ i.i.d. as $f_0$ is the maximizer of the “adjusted criterion function”

$$
\ell_n(\varphi) = \int \log f_{\varphi}(x) dF_n(x) - \int f_{\varphi}(x) dx
$$

$$
= \begin{cases} 
\int \varphi(x) dF_n(x) - \int e^{\varphi(x)} dx, & s = 0, \\
\int (1/s) \log(-\varphi(x)) + dF_n(x) - \int (-\varphi(x))^{1/s} dx, & s < 0,
\end{cases}
$$

over $\varphi \in \mathcal{C}$. 
1. Basics

- The MLE's for $\mathcal{P}_0$ exist and are unique when $n \geq d + 1$.
- The MLE's for $\mathcal{P}_s$ exist for $s \in (-1/d, 0)$ when

$$n \geq d \left( \frac{r}{r - d} \right)$$

where $r = -1/s$. Thus $n \to \infty$ as $-1/s \to r \searrow d$.
- Uniqueness of MLE's for $\mathcal{P}_s$?
- MLE $\hat{\phi}_n$ is piecewise affine for $-1/d < s \leq 0$.
- The MLE for $\mathcal{P}_s$ does not exist if $s < -1/d$. (Well known for $s = -\infty$ and $d = 1$.)
2. On the model

- The MLE’s are Hellinger and $L_1$– consistent.

- The log-concave MLE’s $\hat{f}_{n,0}$ satisfy

$$\int e^{a|x|} |\hat{f}_{n,0}(x) - f_0(x)| dx \rightarrow a.s. 0.$$  

for $a < a_0$ where $f_0(x) \leq \exp(-a_0|x| + b_0)$.

- The $s$–concave MLE’s are computationally awkward; log is “too aggressive” a transform for an $s$–concave density. [Note that ML has difficulties even for location $t$– families: multiple roots of the likelihood equations.]

- Pointwise distribution theory for $\hat{f}_{n,0}$ when $d = 1$; no pointwise distribution theory for $\hat{f}_{n,s}$ when $d = 1$;

  no pointwise distribution theory for $\hat{f}_{n,0}$ or $\hat{f}_{n,s}$ when $d > 1$.

- Global rates? $H(\hat{f}_{n,s}, f_0) = O_p(n^{-2/5})$ for $-1 < s \leq 0$, $d = 1$. 

3. Off the model

Now suppose that $Q$ is an arbitrary probability measure on $\mathbb{R}^d$ with density $q$ and $X_1, \ldots, X_n$ are i.i.d. $q$.

- The MLE $\hat{f}_n$ for $\mathcal{P}_0$ satisfies:
  \[ \int_{\mathbb{R}^d} |\hat{f}_n(x) - f^*(x)| \, dx \to_{a.s.} 0 \]
  where, for the Kullback-Leibler divergence
  \[ K(q, f) = \int q \log(q/f) \, d\lambda, \]
  \[ f^* = \arg\min_{f \in \mathcal{P}_0(\mathbb{R}^d)} K(q, f) \]
  is the “pseudo-true” density in $\mathcal{P}_0(\mathbb{R}^d)$ corresponding to $q$.

In fact:
  \[ \int_{\mathbb{R}^d} e^{a \|x\|} |\hat{f}_n(x) - f^*(x)| \, dx \to_{a.s.} 0 \]
  for any $a < a_0$ where $f^*(x) \leq \exp(-a_0 \|x\| + b_0)$. 
The MLE $\hat{f}_n$ for $\mathcal{P}_s$ does not behave well off the model. Retracing the basic arguments of Cule and Samworth (2010) leads to negative conclusions. (How negative remains to be pinned down!)

**Conclusion:** Investigate alternative methods for estimation in the larger classes $\mathcal{P}_s$ with $s < 0$! This leads to the proposals by Koenker and Mizera (2010).
D. An alternative to ML: Rényi divergence estimators

0. Notation and Definitions

- $\beta = 1 + 1/s < 0$, $\alpha^{-1} + \beta^{-1} = 1$.
- $\mathcal{C}(X)$ = all continuous functions on $\text{conv}(X)$.
- $\mathcal{C}^*(X)$ = all signed Radon measures on $\mathcal{C}(X) = \text{dual space of } \mathcal{C}(X)$.
- $\mathcal{G}(X)$ = all closed convex (lower s.c.) functions on $\text{conv}(X)$.
- $\mathcal{G}(X)^\circ = \{G \in \mathcal{C}^*(X) : \int gdG \leq 0 \text{ for all } g \in \mathcal{G}(X)\}$, the polar (or dual) cone of $\mathcal{G}(X)$. 
Primal problems: $\mathcal{P}_0$ and $\mathcal{P}_s$:

• $\mathcal{P}_0$: $\min_{g \in g(X)} L_0(g, \mathbb{P}_n)$ where
  \[ L_0(g, \mathbb{P}_n) = \mathbb{P}_n g + \int_{\mathbb{R}^d} \exp(-g(x)) \, dx. \]

• $\mathcal{P}_s$: $\min_{g \in g(X)} L_s(g, \mathbb{P}_n)$ where
  \[ L_s(g, \mathbb{P}_n) = \mathbb{P}_n g + \frac{1}{|\beta|} \int_{\mathbb{R}^d} g(x)^\beta \, dx. \]
Dual problems: $\mathcal{P}_0$ and $\mathcal{P}_s$:

- $\mathcal{D}_0$: \[ \max_f \left\{ - \int f(y) \log f(y) \, dy \right\} \text{ subject to } f(y) = \frac{d(\mathbb{P}_n - G)}{dy} \text{ for some } G \in \mathcal{G}(X)^\circ. \]

- $\mathcal{D}_s$: \[ \max_f \int \frac{f(y)^\alpha}{\alpha} \, dy \text{ subject to } f(y) = \frac{d(\mathbb{P}_n - G)}{dy} \text{ for some } G \in \mathcal{G}(X)^\circ. \]
Why do these make sense?

- Population version of $\mathcal{P}_0$: \( \min_{g \in G} L_0(g, f_0) \) where
  \[
  L_0(g, f_0) = \int \{g(x)f_0(x) + e^{-g(x)}\}dx.
  \]
  Minimizing the integrand pointwise in \( g = g(x) \) for fixed \( f_0(x) \) yields
  \( f_0(x) - e^{-g(x)} = 0 \) if \( e^{-g(x)} = f_0(x) \).

- Population version of $\mathcal{P}_s$: \( \min_{g \in G} L_s(g, f_0) \) where
  \[
  L_s(g, f_0) = \int \{g(x)f_0(x) + \frac{1}{|\beta|}g^\beta(x)\}dx.
  \]
  Minimizing the integrand pointwise in \( g = g(x) \) for fixed \( f_0(x) \) yields
  \( f_0(x) + (\beta/|\beta|)g^{\beta-1}(x) = 0 \), and hence
  \( g^{1/s} = g^{1/s}(x) = f_0(x) \).
1. Basics for the Rényi divergence estimators:

- (Koenker and Mizera, 2010) If $\text{conv}(X)$ has non-empty interior, then strong duality between $\mathcal{P}_s$ and $\mathcal{D}_s$ holds. The dual optimal solution exists, is unique, and $\hat{f}_n = \hat{g}_n^{1/s}$.

- (Koenker and Mizera, 2010) The solution $f = g^{1/s}$ in the population version of the problem when $Q = P_0$ has density $p_0 \in \mathcal{P}_s$ is Fisher-consistent; i.e. $f = p_0$. 
2. Off the model: Han & W (2015)

Let

\[ Q_1 = \{Q \text{ on } (\mathbb{R}^d, \mathcal{B}^d) : \int \|x\| dQ(x) < \infty \}, \]
\[ Q_0 = \{Q \text{ on } (\mathbb{R}^d, \mathcal{B}^d) : \text{int}(\text{csupp}(Q)) \neq \emptyset \}. \]

- Theorem (Han & W, 2015): If \(-1/(d + 1) < s < 0\) and \(Q \in Q_0 \cap Q_1\), then the primal problem \(\mathcal{P}_s(Q)\) has a unique solution \(\tilde{g} \in \mathcal{G}\) which satisfies \(\tilde{f} = \tilde{g}^{1/s}\) where \(\tilde{g}\) is bounded away from 0 and \(\tilde{f}\) is a bounded density.

- Theorem (Han & W, 2015): Let \(d = 1\). If \(\hat{f}_{n,s}\) denotes the solution to the primal problem \(\mathcal{P}_s\) and \(\hat{f}_{n,0}\) denotes the solution to the primal problem \(\mathcal{P}_0\), then for any \(\kappa > 0, p \geq 1\),

\[ \int (1 + |x|)^\kappa |\hat{f}_{n,s}(x) - \hat{f}_{n,0}(x)|^p dx \to 0 \text{ as } s \uparrow 0. \]
Theorem (Han & W, 2015): Suppose that:
(i) $d \geq 1$,
(ii) $-1/(d + 1) < s < 0$, and
(iii) $Q \in Q_0 \cap Q_1$.
If $f_{Q,s}$ denotes the (pseudo-true) solution to the primal problem $P_s(Q)$, then for any $\kappa < r - d = (-1/s) - d$,

$$\int (1 + |x|)^\kappa |\hat{f}_{n,s}(x) - f_{Q,s}(x)| dx \to a.s. 0 \text{ as } n \to \infty.$$
3. On the model: $Q$ has density $f \in \mathcal{P}_{s'}$; $f = g^{1/s'}$ for some $g$ convex.

- Consistency: Suppose that: (i) $d \geq 1$ and $-1/d < s < 0$ and $s' > s$ if $s \leq -1/(d+1)$, $s' = s$ if $s > -1/(d+1)$. Then for any $\kappa < r - d = (-1/s) - d$,

$$
\int (1 + |x|)^\kappa |\hat{f}_{n,s}(x) - f(x)| dx \rightarrow_{a.s.} 0 \text{ as } n \rightarrow \infty.
$$

Thus $H(\hat{f}_{n,s}, f) \rightarrow_{a.s.} 0$ as well.

- Pointwise limit theory: (paralleling the results of Balabdaoui, Rufibach, and W (2009) for $s = 0$)

Assumptions:

\begin{itemize}
  \item (A1) $g_0 \in \mathcal{G}$ and $f_0 \in \mathcal{P}_s(\mathbb{R})$ with $-1/2 < s < 0$.
  \item (A2) $f_0(x_0) > 0$.
\end{itemize}
(A3) $g_0$ is locally $C^2$ in a neighborhood of $x_0$ with $g_0''(x_0) > 0$. 
Theorem 1. (Pointwise limit theorem; Han & W (2015))

Under assumptions (A1)-(A3), we have

\[
\begin{pmatrix}
(n_5^2 \left( \widehat{g}_n(x_0) - g_0(x_0) \right)) \\
(n_5^1 \left( \widehat{g}'_n(x_0) - g'_0(x_0) \right))
\end{pmatrix}
\overset{d}{\to}
\begin{pmatrix}
- \left( \frac{g_0^4(x_0)g_0^{(2)}(x_0)}{r^4 f_0(x_0)^2(4)!} \right)^{1/5} H_2^{(2)}(0) \\
- \left( \frac{g_0^2(x_0)g_0^{(2)}(x_0)^3}{r^2 f_0(x_0)^3(4)!} \right)^{1/5} H_2^{(3)}(0)
\end{pmatrix},
\]

and ...
... furthermore
\[
\left( n^{\frac{2}{5}}(\tilde{f}_n(x_0) - f_0(x_0)) \right) \rightarrow_d \left( n^{\frac{1}{5}}(\tilde{f}_n(x_0) - f'_0(x_0)) \right) \rightarrow_d \begin{pmatrix}
\left( \frac{r f_0(x_0)^3 g_0^{(2)}(x_0)}{g_0(x_0)(4)!} \right)^{1/5} H_2^{(2)}(0) \\
\left( \frac{r^3 f_0(x_0)^4 (g_0^{(2)}(x_0))^3}{g_0(x_0)^3 [4)!^3} \right)^{1/5} H_2^{(3)}(0)
\end{pmatrix},
\]
where \( H_2 \) is the unique lower envelope of the process \( Y_2 \) satisfying

1. \( H_2(t) \leq Y_2(t) \) for all \( t \in \mathbb{R} \);
2. \( H_k^{(2)} \) is concave;
3. \( H_2(t) = Y_2(t) \) if the slope of \( H_2^{(2)} \) decreases strictly at \( t \).
4. \( Y_2(t) = \int_0^t W(s)ds - t^4, t \in \mathbb{R} \) where \( W \) is two-sided Brownian motion started at 0.
• Estimation of the mode for $d = 1$.

**Theorem 2.** (Estimation of the mode) Assume (A1)-(A4) hold. Then

$$n^{1/5} (\hat{m}_n - m_0) \xrightarrow{d} d \left( \frac{g_0(m_0)^2 (4)!^2}{r^2 f_0(m_0) g_0^{(2)}(m_0)^2} \right)^{1/5} M(H_2^{(2)}),$$

(1)

where $\hat{m}_n = M(\hat{f}_n), m_0 = M(f_0)$.

• What is the price of assuming $s < 0$ when the truth $f \in \mathcal{P}_0$?

Assume $-1/2 < s < 0$ and $k = 2$. Let $f_0 = \exp(\varphi_0)$ be a log-concave density where $\varphi_0 : \mathbb{R} \to \mathbb{R}$ is the underlying concave function. Then $f_0$ is also $s$-concave.
Let $g_s := f_0^{-1/r} = \exp(-\varphi_0/r)$ be the underlying convex function when $f_0$ is viewed as an $s$-concave density. Calculation yields

$$g_s^{(2)}(x_0) = \frac{1}{r^2} g_s(x_0) \left( \varphi'_0(x_0)^2 - r\varphi''_0(x_0) \right).$$

Hence the constant before $H_2^{(2)}(0)$ appearing in the limit distribution for $\hat{f}_n$ becomes

$$\left( \frac{f_0(x_0)^3\varphi'_0(x_0)^2}{4!r} + \frac{f_0(x_0)^3|\varphi''_0(x_0)|}{4!} \right)^{1/5}.$$

The second term is the constant involved in the limiting distribution when $f_0(x_0)$ is estimated via the log-concave MLE: (2.2), page 1305 in Balabdaoui, Rufibach, & W (2009). The ratio of the two constants (or asymptotic relative efficiency) is shown for $f_0$ standard normal (blue) and logistic (magenta) in the figure:
• The first term is non-negative and is the price we pay by estimating a true log-concave density via the Rényi divergence estimator over a larger class of $s$-concave densities.

• Note that the first term vanishes as $r \to \infty$ (or $s \uparrow 0$).

• Note that the ratio is 1 at the mode of $f_0$.

• For estimation of the mode, the ratio of constants is always 1: nothing is lost by enlarging the class from $s = 0$ to $s < 0$!
E. Summary: problems and open questions

- Global rates of convergence?
- Limiting distribution(s) for $d > 1$? ($n^r$ with $r = 2/(4 + d)$?)
- MLE (rate-) inefficient for $d \geq 4$ (or perhaps $d \geq 3$)? How to penalize to get efficient rates?
- Can we go below $s = -1/(d + 1)$ with other methods?
- Multivariate classes with nice preservation/closure properties and smoother than log-concave?
- Algorithms for computing $\hat{f}_n \in \mathcal{P}_s$? (Non-smooth and convex; or non-smooth and non-convex?)
Guvenen et al (2014) have estimated models of income dynamics using very large (10 percent) samples of U.S. Social Security records linked to W2 data. The density is not log-concave, but an $s$–concave density with $s = -1/2$ fits well:

---

Courtesy Roger Koenker

Cemmap Seminar, UCL, October 27, 2015 1.32
F. Selected references

- Dümbgen and Rufibach (2009).
- Cule, Samworth, and Stewart (2010).
- Cule and Samworth (2010).
- Guntuboyina and Sen (xxxx)
Many thanks!