LOG-CONCAVE DENSITY ESTIMATION WITH APPLICATIONS

Joint work with Y. Chen, M. Cule, L. Dümbgen R. Gramacy, D. Schuhmacher, M. Stewart, M. Yuan
The problem

Let $X_1, \ldots, X_n$ be a random sample from a density $f_0$ in $\mathbb{R}^d$.

How should we estimate $f_0$?

Two main alternatives:

• Parametric models: use e.g. MLE. Assumptions often too restrictive.

• Nonparametric models: use e.g. kernel density estimate. Choice of bandwidth difficult, particularly for $d > 1$. 
Is there a third way?


A density \( f \) is log-concave if \( \log f \) is concave.

- Univariante examples: normal, logistic, Gumbel densities, as well as Weibull, Gamma, Beta densities for certain parameter values.
Characterising log-concave densities

Cule, S. and Stewart (2010)

Let $X$ have density $f$ in $\mathbb{R}^d$. For a subspace $V$ of $\mathbb{R}^d$, let $P_V(x)$ denote the orthogonal projection of $x$ onto $V$. Then in order that $f$ be log-concave, it is:

1. necessary that for any subspace $V$, the marginal density of $P_V(X)$ is log-concave (Prékopa 1973), and the conditional density $f_{X|P_V(X)}(\cdot|t)$ of $X$ given $P_V(X) = t$ is log-concave for each $t$

2. sufficient that, for every $(d - 1)$-dimensional subspace $V$, the conditional density $f_{X|P_V(X)}(\cdot|t)$ of $X$ given $P_V(X) = t$ is log-concave for each $t$. 
Unbounded likelihood!

Consider maximizing the likelihood $L(f) = \prod_{i=1}^{n} f(X_i)$ over all densities $f$. 
Existence and uniqueness


Let $X_1, \ldots, X_n$ be independent with density $f_0$ in $\mathbb{R}^d$, and suppose that $n \geq d + 1$. Then, with probability one, a log-concave maximum likelihood estimator $\hat{f}_n$ exists and is unique.
Sketch of proof

Consider maximizing over all log-concave functions

\[ \psi_n(f) = \frac{1}{n} \sum_{i=1}^{n} \log f(X_i) - \int_{\mathbb{R}^d} f(x) \, dx. \]

Any maximizer \( f \) must satisfy:

1. \( f(x) > 0 \) iff \( x \in C_n \equiv \text{conv}(X_1, \ldots, X_n) \)

2. Fix \( y = (y_1, \ldots, y_n) \) and let \( \bar{h}_y : \mathbb{R}^d \to \mathbb{R} \) be the smallest concave function with \( \bar{h}_y(X_i) \geq y_i \) for all \( i \). Then \( \log f = \bar{h}_{y^*} \) for some \( y^* \)

3. \( \int_{\mathbb{R}^d} f(x) \, dx = 1. \)
Schematic diagram of MLE on log scale
Computation


First attempt: minimise

\[ \tau(y) = -\frac{1}{n} \sum_{i=1}^{n} h_y(X_i) + \int_{C_n} \exp\{h_y(x)\} \, dx. \]
Computation


First attempt: minimise

\[ \tau(y) = -\frac{1}{n} \sum_{i=1}^{n} \bar{h}_y(X_i) \int_{C_n} \exp\{\bar{h}_y(x)\} \, dx. \]

Better: minimise

\[ \sigma(y) = -\frac{1}{n} \sum_{i=1}^{n} y_i + \int_{C_n} \exp\{\bar{h}_y(x)\} \, dx. \]

Then \( \sigma \) has a unique minimum at \( y^* \), say, \( \log \hat{f}_n = \bar{h}_y^* \) and \( \sigma \) is convex ...
Computation


First attempt: minimise

$$\tau(y) = -\frac{1}{n} \sum_{i=1}^{n} \bar{h}_y(X_i) + \int_{C_n} \exp\{\bar{h}_y(x)\} \, dx.$$ 

Better: minimise

$$\sigma(y) = -\frac{1}{n} \sum_{i=1}^{n} y_i + \int_{C_n} \exp\{\bar{h}_y(x)\} \, dx.$$ 

Then $\sigma$ has a unique minimum at $y^*$, say, $\log \hat{f}_n = \bar{h}_{y^*}$ and $\sigma$ is convex ... but non-differentiable!
Log-concave projections


Let $\mathcal{P}$ be the set of all probability distributions $P$ on $\mathbb{R}^d$ with $P(H) < 1$ for all hyperplanes $H$. Let

$$\mathcal{P}_r = \left\{ P \in \mathcal{P} : \int_{\mathbb{R}^d} \|x\|^r P(dx) < \infty \right\}, \quad r = 1, 2.$$ 

The condition $P_0 \in \mathcal{P}_1$ is necessary and sufficient for the existence of an a.e. unique log-concave density $f^*$ that maximises $\int_{\mathbb{R}^d} \log f \, dP_0$ over all log-concave densities.
Let $P_0 \in \mathcal{P}_1$ have distribution function $F_0$. Let

$$S(f^*) = \{x \in \mathbb{R} : \log f^*(x) > \frac{1}{2} \log f^*(x-\delta) + \frac{1}{2} \log f^*(x+\delta) \ \forall \delta > 0\}.$$ 

Then the distribution function $F^*$ of $f^*$ is characterised by

$$\int_{-\infty}^{x} \{F^*(t) - F_0(t)\} \, dt \begin{cases} \leq 0 & \text{for all } x \in \mathbb{R} \\ = 0 & \text{for all } x \in S(f^*) \cup \{\infty\}. \end{cases}$$
Example 1

Suppose $f_0(x) = \frac{1}{2} (1 + x^2)^{-3/2}$. Then $f^*(x) = \frac{1}{2} e^{-|x|}$. 
Example 2
Log-concave projections preserve independence  Chen and S. (2011)

Suppose $P \in \mathcal{P}_1$ can be written as $P = P_1 \otimes P_2$, where $P_1$ and $P_2$ are probability measures on $\mathbb{R}^{d_1}$ and $\mathbb{R}^{d_2}$, with $d_2 = d - d_1$. If $f^*$ is the log-concave projection of $P$ and $f^*_\ell$ is the projection of $P_\ell$ ($\ell = 1, 2$), then

$$f^*(x) = f^*_1(x_1)f^*_2(x_2)$$

for $x = (x_1^T, x_2^T)^T \in \mathbb{R}^d$.

This makes log-concave projections very attractive for independent component analysis (S. and Yuan, 2012).
Convergence of log-concave densities

Cule and S. (2010)

Let \((f_n)\) be a sequence of log-concave densities on \(\mathbb{R}^d\) with \(f_n \xrightarrow{d} f\) for some density \(f\). Then:

(a) \(f\) is log-concave

(b) \(f_n \to f\) almost everywhere

(c) Let \(a_0 > 0\) and \(b_0 \in \mathbb{R}\) be such that \(f(x) \leq e^{-a_0 \|x\|} + b_0\). If \(a < a_0\) then \(\int e^{a \|x\|} |f_n(x) - f(x)| \, dx \to 0\) and, if \(f\) is continuous, \(\sup_x e^{a \|x\|} |f_n(x) - f(x)| \to 0\).
Theoretical properties


Now let $X_1, \ldots, X_n \overset{iid}{\sim} P_0 \in \mathcal{P}_1$, and let $f^*$ denote the log-concave projection of $P_0$. Taking $a_0 > 0$ and $b_0 \in \mathbb{R}$ such that $f^*(x) \leq e^{-a_0\|x\|+b_0}$, we have for any $a < a_0$ that

$$\int_{\mathbb{R}^d} e^{a\|x\|} |\hat{f}_n(x) - f^*(x)| \, dx \xrightarrow{a.s.} 0,$$

and, if $f^*$ is continuous, $\sup_x e^{a\|x\|} |\hat{f}_n(x) - f^*(x)| \xrightarrow{a.s.} 0$. 

Fitting finite mixtures of log-concave densities


Now suppose

$$f(x) = \sum_{j=1}^{p} \pi_j f_j(x),$$

where the weights $\pi_j$ are positive and sum to one, and each $f_j$ is log-concave on $\mathbb{R}^d$.

We can combine the algorithm for finding the log-concave MLE with the EM algorithm to fit such a mixture.
Smoothed log-concave density estimator


Let

$$\tilde{f}_n = \hat{f}_n * \phi_{\hat{A}},$$

where $\phi_{\hat{A}}$ is a $d$-dimensional normal density with mean zero and covariance matrix $\hat{A} = \hat{\Sigma} - \tilde{\Sigma}$. Here, $\hat{\Sigma}$ is the sample covariance matrix and $\tilde{\Sigma}$ is the covariance matrix corresponding to $\hat{f}_n$.

Then $\tilde{f}_n$ is a smooth, fully automatic log-concave estimator supported on the whole of $\mathbb{R}^d$ which satisfies the same theoretical properties as $\hat{f}_n$.

It offers potential improvements for small sample sizes.
Smoothed estimator in classification
Smoothed log-concave MLE theory

Chen and S. (2011)

Let \( X_1, \ldots, X_n \overset{iid}{\sim} P_0 \in \mathcal{P}_2 \), let \( \mu = \int x \; dP_0(x) \), and
\[ \Sigma = \int (x - \mu)(x - \mu)^T \; dP_0(x). \]

Let \( f^{**} = f^* * \phi_{A^*} \), where \( A^* = \Sigma - \Sigma^* \) with
\[ \Sigma^* = \int (x - \mu)(x - \mu)^T f^*(x) \; dx. \] Taking \( a_0 > 0 \) and \( b_0 \in \mathbb{R} \)
such that \( f^{**}(x) \leq e^{-a_0\|x\|} + b_0 \), we have for all \( a < a_0 \) that
\[
\int_{\mathbb{R}^d} e^{a\|x\|} |\tilde{f}_n(x) - f^{**}(x)| \xrightarrow{a.s.} 0.
\]
Suppose $P_0 \in \mathcal{P}_1$. Then $\text{tr}(A^*) = 0$ if and only if $P_0$ has a log-concave density.

We can therefore use $\hat{A}$ as a test statistic, and generate a critical value from bootstrap samples drawn from $\hat{f}_n$.

This test is consistent: if $P_0$ is not log-concave, then the power converges to 1 as $n \to \infty$. 

Classification problems

Let \((X, Y), (X_1, Y_1), \ldots, (X_n, Y_n)\) be i.i.d. pairs in \(\mathbb{R}^d \times \{1, \ldots, K\}\), with \(\mathbb{P}(Y = r) = \pi_r\) and \((X | Y = r) \sim P_r\), for \(r = 1, \ldots, K\).

A classifier is a function \(C : \mathbb{R}^d \rightarrow \{1, \ldots, K\}\).

We aim to minimise the \textit{misclassification error rate} or \textit{risk}:

\[
\text{Risk}(C) = \mathbb{P}(C(X) \neq Y).
\]
Smoothed log-concave and Bayes classifiers

Suppose each class distribution $P_r$ has a (Lebesgue) density $f_r$. The smoothed log-concave classifier is

$$\hat{C}_n^{SLC}(x) = \arg\max_{r \in \{1, \ldots, K\}} N_r \hat{f}_{n,r}(x),$$

where $N_r = \sum_{i=1}^n 1_{\{Y_i = r\}}$ and $\hat{f}_{n,r}$ is the smoothed log-concave estimate based on $\{X_i : Y_i = r\}$.

The Bayes classifier is

$$C^{Bayes}(x) = \arg\max_{r \in \{1, \ldots, K\}} \pi_r f_r(x).$$

Its risk is optimal...
Smoothed log-concave and Bayes classifiers

Suppose each class distribution $P_r$ has a (Lebesgue) density $f_r$. The smoothed log-concave classifier is

$$\hat{C}_{n}^{SLC}(x) = \arg\max_{r \in \{1, \ldots, K\}} N_r \tilde{f}_{n,r}(x),$$

where $N_r = \sum_{i=1}^{n} 1\{Y_i = r\}$ and $\tilde{f}_{n,r}$ is the smoothed log-concave estimate based on $\{X_i : Y_i = r\}$.

The Bayes classifier is

$$C^{\text{Bayes}}(x) = \arg\max_{r \in \{1, \ldots, K\}} \pi_r f_r(x).$$

Its risk is optimal. . .*but it can’t be used in practice!*
Theory for smoothed log-concave classifiers

Chen and S. (2011)

The smoothed log-concave Bayes classifier is

\[ C^{SLCBayes}(x) = \arg \max_{r \in \{1, \ldots, K\}} \pi_r f_r^{**}(x). \]

Let \( X^{**} = \{ x \in \mathbb{R}^d : |\arg \max_r \pi_r f_r^{**}(x)| = 1 \} \). Then

\[ \hat{C}_n^{SLC}(x) \xrightarrow{a.s.} C^{SLCBayes}(x) \]

for almost all \( x \in X^{**} \), and

\[ \text{Risk}(\hat{C}_n^{SLC}) \rightarrow \text{Risk}(C^{SLCBayes}). \]
Breast cancer data
Classification boundaries
Consider the regression model

\[ Y_i = \mu(x_i) + \epsilon_i, \quad i = 1, \ldots, n, \]

where \( \epsilon_1, \ldots, \epsilon_n \) are i.i.d., log-concave and \( \mathbb{E}(\epsilon_i) = 0 \). In both of the cases i) \( \mu \) is linear and ii) \( \mu \) is isotonic, we can jointly estimate \( \mu \) and the distribution of \( \epsilon_i \).

Significant improvements are obtainable over usual methods when errors are non-normal.
Summary

- The log-concave MLE is a fully automatic, nonparametric density estimator.
- It has several extensions which can be used in a wide variety of applications, e.g. classification, clustering, functional estimation and regression problems.
- Many challenges remain: faster algorithms, dependent data, further theoretical results, other applications and constraints,...
References


