# Log-concave density estimation for $d \geq 2$

Kaspar Rufibach

Biostatistics Unit
Institute of Social and Preventive Medicine
University of Zurich

Meeting of the Royal Statistical Society
$12^{\text{th}}$ May 2010

# Log-concave density estimation on $\mathbb{R}$

$X_1, \ldots, X_n \in \mathbb{R}$ i.i.d. rv's with density $f : \mathbb{R} \mapsto \mathbb{R}$.

Goal: Maximize likelihood-function

$$
\begin{aligned}
L_n^1(\varphi) &= \frac{1}{n} \sum_{i=1}^{n} \log f(X_i) - \int_{\mathbb{R}} f(t)\,\mathrm{d}t \\
&= \frac{1}{n} \sum_{i=1}^{n} \varphi(X_i) - \int_{\mathbb{R}} \exp \varphi(t)\,\mathrm{d}t
\end{aligned}
$$

over all log-concave functions $f = \exp \varphi$ to get concave maximum likelihood log-density estimator $\widehat{\varphi}_n$.

# How can we maximize $L_n^1$?

1. Show existence and uniqueness of $\widehat{\varphi}_n$.
2. Show that $\widehat{\varphi}_n$ is a piecewise linear function on $[X_1, X_n]$ and has only knots in $\{X_1, \ldots, X_n\}$ $\Rightarrow$ restrict attention to class

$$\mathcal{P}_n \quad = \quad \Big\{\varphi \in \mathbb{R}^n \ : \ \frac{\varphi_{i+1} - \varphi_i}{X_{i+1} - X_i} \geq \frac{\varphi_i - \varphi_{i-1}}{X_i - X_{i-1}}, i = 2, \ldots, n-1\Big\}$$

   where $\varphi_i = \varphi(X_i)$.

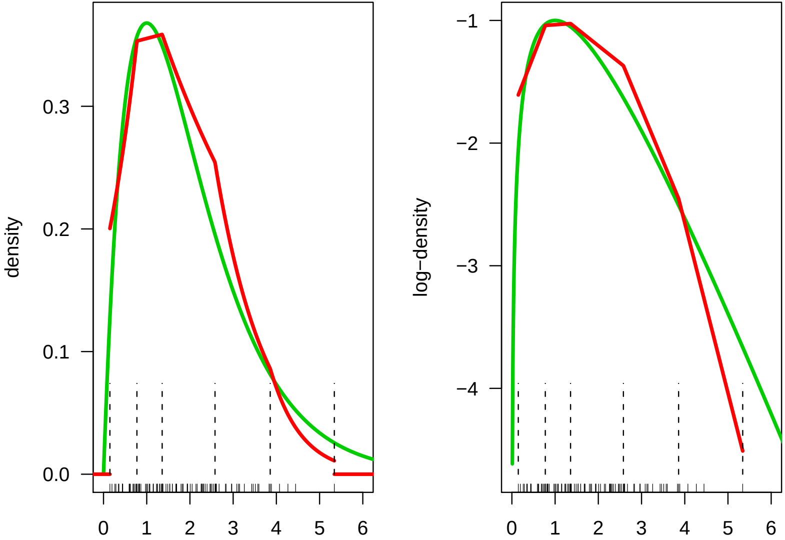
3. For $\varphi \in \mathcal{P}_n$ the log-likelihood function $L_n^1$ can be written as

$$L_n^1(\varphi_1, \ldots, \varphi_n) \quad = \quad \frac{1}{n}\sum_{i=1}^n \varphi_i \ - \ \sum_{i=1}^n \int_{X_{i-1}}^{X_i} \exp\Big(\frac{\varphi_i - \varphi_{i-1}}{X_i - X_{i-1}}(t - X_{i-1}) + \varphi_{i-1}\Big)\, \mathrm{d}t.$$

4. Maximizing $L_n^1 : \mathbb{R}^n \mapsto \mathbb{R}$ is then a constrained optimization problem with linear constraints:

$$\widehat{\varphi}_n \quad = \quad \max_{\varphi \in \mathcal{P}_n} L_n^1(\varphi_1, \ldots, \varphi_n).$$

# Example: $\Gamma(2,1)$ data, $n = 100$



# Why is $d \geq 2$ different from $d = 1$?

$\mathbf{X}_1, \ldots, \mathbf{X}_n \in \mathbb{R}^d$ i.i.d. rv's with density $f : \mathbb{R}^d \mapsto \mathbb{R}$, $n \geq d + 1$, $d \geq 2$.

Goal: Maximize likelihood-function

$$L_n^d(\varphi) \quad = \quad \frac{1}{n}\sum_{i=1}^n \varphi(\mathbf{X}_i) \; - \; \int_{\mathbb{R}^d} \exp\varphi(\mathbf{t})\, \mathrm{d}\mathbf{t}$$

over all log-concave functions $f = \exp\varphi$ to get concave maximum likelihood log-density estimator $\widehat{\varphi}_n$.

# Why is $d \geq 2$ different from $d = 1$?

$\mathbf{X}_1, \ldots, \mathbf{X}_n \in \mathbb{R}^d$ i.i.d. rv's with density $f : \mathbb{R}^d \mapsto \mathbb{R}$, $n \geq d+1$, $d \geq 2$.

Goal: Maximize likelihood-function

$$L_n^d(\varphi) \quad = \quad \frac{1}{n}\sum_{i=1}^n \varphi(\mathbf{X}_i) \; - \; \int_{\mathbb{R}^d} \exp\varphi(\mathbf{t})\,\mathrm{d}\mathbf{t}$$

over all log-concave functions $f = \exp\varphi$ to get concave maximum likelihood log-density estimator $\widehat{\varphi}_n$.

Show that $\widehat{\varphi}_n$ is a tent function $\bar{h}_\mathbf{y}$ on $\mathcal{C}_n = \mathrm{conv}(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ for a vector $\mathbf{y} \in \mathbb{R}^n$

# Why is $d \geq 2$ different from $d = 1$?

$\mathbf{X}_1, \ldots, \mathbf{X}_n \in \mathbb{R}^d$ i.i.d. rv's with density $f : \mathbb{R}^d \mapsto \mathbb{R}$, $n \geq d + 1$, $d \geq 2$.

Goal: Maximize likelihood-function

$$L_n^d(\mathbf{y}) \;=\; \frac{1}{n}\sum_{i=1}^n \bar{h}_{\mathbf{y}}(\mathbf{X}_i) \;-\; \int_{\mathcal{C}_n} \exp \bar{h}_{\mathbf{y}}(\mathbf{t})\,\mathrm{d}\mathbf{t}$$

over all log-concave functions $f = \exp \varphi$ to get concave maximum likelihood log-density estimator $\widehat{\varphi}_n$.

Show that $\widehat{\varphi}_n$ is a tent function $\bar{h}_{\mathbf{y}}$ on $\mathcal{C}_n = \mathrm{conv}(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ for a vector $\mathbf{y} \in \mathbb{R}^n \Rightarrow$ restrict attention to this class of functions, write $L_n^d(\mathbf{y}) : \mathbb{R}^n \mapsto \mathbb{R}$.

# Why is $d \geq 2$ different from $d = 1$?

$\mathbf{X}_1, \ldots, \mathbf{X}_n \in \mathbb{R}^d$ i.i.d. rv's with density $f : \mathbb{R}^d \mapsto \mathbb{R}$, $n \geq d + 1$, $d \geq 2$.

Goal: Maximize likelihood-function

$$L_n^d(\mathbf{y}) \quad = \quad \frac{1}{n} \sum_{i=1}^n \bar{h}_{\mathbf{y}}(\mathbf{X}_i) \; - \; \int_{\mathcal{C}_n} \exp \bar{h}_{\mathbf{y}}(\mathbf{t}) \, \mathrm{d}\mathbf{t}$$

over all log-concave functions $f = \exp \varphi$ to get concave maximum likelihood log-density estimator $\widehat{\varphi}_n$.

Show that $\widehat{\varphi}_n$ is a tent function $\bar{h}_{\mathbf{y}}$ on $\mathcal{C}_n = \mathrm{conv}(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ for a vector $\mathbf{y} \in \mathbb{R}^n \Rightarrow$ restrict attention to this class of functions, write $L_n^d(\mathbf{y}) : \mathbb{R}^n \mapsto \mathbb{R}$.

Two problems here:

$\Rightarrow$ $L_n^d$ as a function of the the tent pole heights $\mathbf{y}$ is not convex!
$\Rightarrow$ How can we enforce concavity restriction on tent function in $\mathbb{R}^d$, i.e. parametrize the tent functions in $\mathbb{R}^d$?

We do not parametrize!

Goal: Maximize likelihood-function

$$L_n^d(\mathbf{y}) \quad = \quad \frac{1}{n}\sum_{i=1}^{n} \bar{h}_{\mathbf{y}}(\mathbf{X}_i) \;-\; \int_{\mathcal{C}_n} \exp \bar{h}_{\mathbf{y}}(\mathbf{t})\,\mathrm{d}\mathbf{t}$$

C., S., S. realized: The function

$$\sigma_n^d(\mathbf{y}) \quad = \quad \frac{1}{n}\sum_{i=1}^{n} y_i \;-\; \int_{\mathcal{C}_n} \exp \bar{h}_{\mathbf{y}}(\mathbf{t})\,\mathrm{d}\mathbf{t}$$

- is convex but not differentiable anymore as a function of $\mathbf{y}$,
- has the unique maximum $\mathbf{y}^*$ such that $\widehat{\varphi}_n = \bar{h}_{\mathbf{y}^*}$ when maximized over $\mathbb{R}^d$.

C., S., S. implemented maximization of the non-differentiable function $\sigma_n^d$ via Shor's ($r$-)algorithm.

# Consistency

C., S., S., prove: Under some weak conditions on a density $f_0$ there exists a log-concave $f^*$ such that

$$f^* = \min_{f \text{ log-concave}} d_{KL}(f, f_0),$$

$$\int_{\mathbb{R}^d} e^{a\|\mathbf{x}\|} |\widehat{f}_n(\mathbf{x}) - f^*(\mathbf{x})| \, \mathrm{d}\mathbf{x} \quad \xrightarrow{\text{as}} \quad 0.$$

Interpretation:

- $\widehat{f}_n$ is consistent w.r.t. to a strong norm.
- Robustness: If $f_0$ is not "too much" non-log-concave, $\widehat{f}_n$ is still sensible.

More details on the structure of possible deviations: Dümbgen et al (2010).

# Rate of convergence – conjecture

If the true density $f_0$ is Hölder with exponent $\alpha$, then for a fixed $\mathbf{x}$

$$|\widehat{f}_n(\mathbf{x}) - f_0(\mathbf{x})| \quad = \quad \begin{cases} O_{\mathrm{p}}\left(n^{-\alpha/(2\alpha+d)}\right) & \text{if } d \leq 2\alpha, \\ O_{\mathrm{p}}\left(n^{-\alpha/(2d)}\right) & \text{if } d > 2\alpha. \end{cases}$$

See Birgé & Massart (1993) for general results on rate of convergence for minimum contrast estimators.

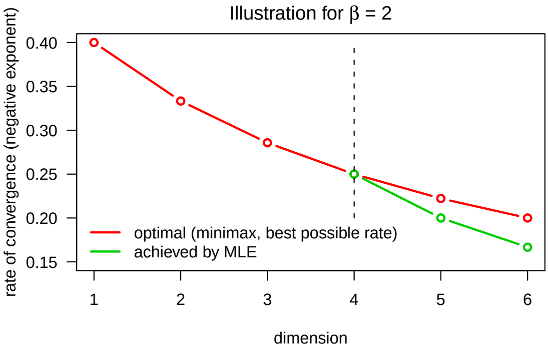
$d = 1$: established for $\alpha \in [1, 2]$:

- If $\alpha = 1 \Rightarrow$ rate $= n^{-1/3}$. Extrapolated from result on sup-norm
- If $\alpha = 2 \Rightarrow$ rate $= n^{-2/5}$. Even limiting distribution available

These results entail asymptotic equivalence of $\mathbb{F}_n$ and $\widehat{F}_n$ (if $\beta > 1$):

$$\max_{t \in T_n} |\mathbb{F}_n(t) - \widehat{F}_n(t)| \quad = \quad o_{\mathrm{p}}(n^{-1/2}).$$

# Rate of convergence – conjecture




- MLE $\widehat{f}_n(x)$ "only" rate efficient up to dimension 4?
- Log-concavity is "not enough" for $d > 4$: To get optimal rates for MLE
  - should we penalize?
  - Or put more constraints on the class of functions?
  - What additional constraints?

# Summary

C., S., S. achieved:

- Introduction of log-concave density estimator for $d \geq 2$.
- Existence, uniqueness.
- Transformation of initial maximization problem into a tractable one.
- Application of Shor's algorithm to find $\widehat{\varphi}_n$.
- Consistency of $\widehat{f}_n$ also under misspecification.
- Assessment of finite-$n$ performance.
- "Estimation" of mixtures in dimension $d \geq 2$ ("" due to identifiability issues).
- Plug-in estimation of functionals.
- Sketch of a test to assess log-concavity.

Thank you for your attention.