Adaptation in log-concave density estimation

Oliver Feng

University of Cambridge

24 May 2018

Ongoing work with Arlene Kim (Sungshin Women's University),
Adityanand Guntuboyina (UC Berkeley) and
Richard J. Samworth (University of Cambridge)

Traditional approaches to density estimation

- Task: estimate an unknown density function $f_0 \colon \mathbb{R}^d \to [0, \infty)$ based on a random sample $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} f_0$.
- Parametric approaches (e.g. maximum likelihood estimation for Gaussian densities): often produce estimators that are analytically tractable and/or easy to compute, but parametric assumptions can be too restrictive.
- Non-parametric approaches (e.g. kernel density estimation): flexible, but often require the choice of one or more tuning parameters (e.g. kernel bandwidths).

Log-concave densities

- $f : \mathbb{R}^d \to [0, \infty)$ is said to be *log-concave* if $\log f$ is concave.
- When $d = 1$: unimodal with exponentially decaying tails.
- Examples: Gaussian, logistic, Gumbel and uniform densities.
- Log-concave density estimation: best of the parametric and non-parametric worlds – modelling flexibility and fully automatic procedures.

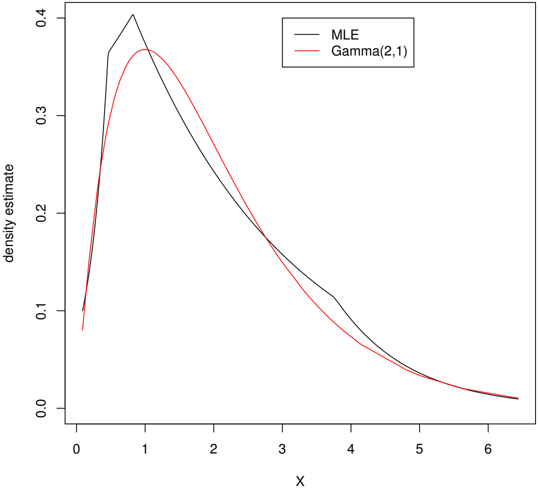
The log-concave maximum likelihood estimator

- $\hat{f}_n$ maximises the log-likelihood function $\sum_{i=1}^n \log f(X_i)$ over log-concave densities f.

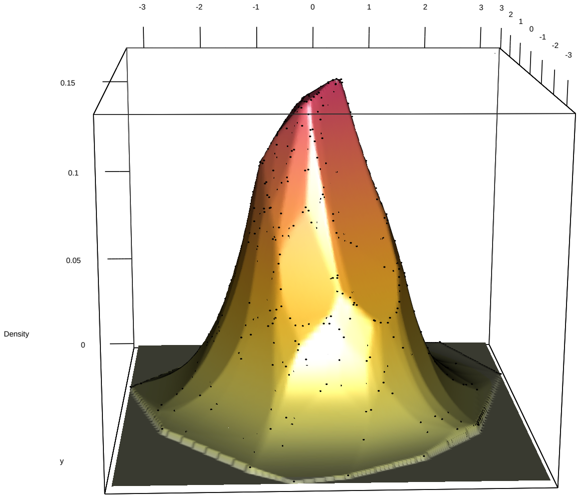
- Cule, Samworth and Stewart (2010): $\hat{f}_n$ exists and is unique with probability 1 if f_0 is log-concave and $n \geq d + 1$.

- $\log \hat{f}_n$ is a 'tent pole function' with tent poles at the X_i; $\hat{f}_n$ is supported on the convex hull of the data.

Plot when $d = 1$



Plot when $d = 2$



Performance of $\hat{f}_n$

- Introduce a global loss function: squared Hellinger distance $d_{\mathrm{H}}^2(f,g) := \int_{\mathbb{R}^d} (\sqrt{f} - \sqrt{g})^2$.

- Worst-case risk bounds for $\hat{f}_n$ (Kim and Samworth, 2016):

$$\sup_{f_0 \text{ log-concave}} \mathbb{E}_{f_0}\{d_{\mathrm{H}}^2(\hat{f}_n, f_0)\} \lesssim_d \begin{cases} n^{-4/5} & d = 1; \\ n^{-2/3} \log n & d = 2; \\ n^{-1/2} \log n & d = 3. \end{cases}$$

- In dimensions $1, 2, 3$, $\hat{f}_n$ has essentially the best possible worst-case performance (achievable by *any* estimator); it is (almost) minimax optimal.

Adaptation properties

- Motivation: $\log \hat{f}_n$ is concave and piecewise affine, so it is natural to expect more accuracy when $\log f_0$ is also close to piecewise affine.
- Kim, Guntuboyina and Samworth (2018) showed that if $d = 1$ and f_0 is (close to) log k-affine, then

$$\mathbb{E}_{f_0}\{d_{\mathrm{H}}^2(\hat{f}_n, f_0)\} \lesssim \frac{k}{n} \log^{5/4} n.$$

- New result: if $d = 2, 3$ and f_0 is (close to) a log k-affine function supported on a polyhedral set with at most m facets, then

$$\mathbb{E}_{f_0}\{d_{\mathrm{H}}^2(\hat{f}_n, f_0)\} \lesssim_d \frac{k(k+m)}{n} \log^{\gamma_d} n,$$

where $\gamma_2 = 9/2$ and $\gamma_3 = 8$.