Exact Solutions in Log-Concave Maximum Likelihood Estimation

Miruna-Ştefana SOREA

SISSA, Trieste, Italy

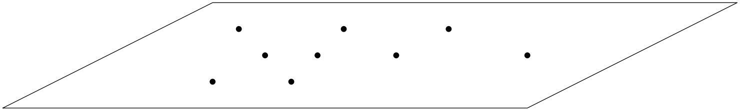
Workshop on Real Algebraic Geometry and Algorithms for Geometric Constraint Systems, The Fields Institute, June 14 - 18, 2021

Joint work with:

Alexandros Grosdos
Alexander Heaton
Kaie Kubjas
Olga Kuznetsova
Georgy Scholten

Weighted Density Estimation

Data set: $X = (x_1, x_2, \ldots, x_n)$ a point configuration in $\mathbb{R}^d$ with weights $w = (w_1, w_2, \ldots, w_n)$ such that $w_i \geq 0$ and $w_1 + w_2 + \cdots + w_n = 1$.


[1]

Goal: estimate an unknown probability distribution
How: MLE

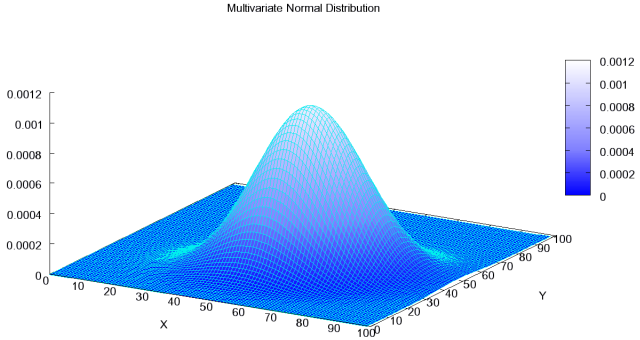
$$\hat{p} := \mathrm{argmax}_p \sum_{i=1}^{n} w_i \cdot \log(p(x_i)),$$

where p is a density

[1] Source: http://gac-school.imj-prg.fr/talks/sturmfels.pdf

Parametric statistics

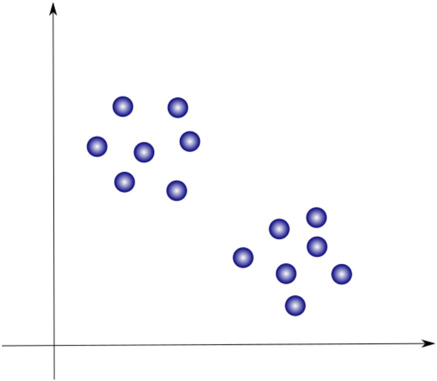
$$\max_{(\mu,\Sigma)} \sum_{i=1}^{n} w_i \cdot \log(p_{\mu,\Sigma}(x_i)). \quad (1)$$



2

2 Source: CCBY-SA3.0, https://commons.wikimedia.org/w/index.php?curid=1260349

Parametric statistics



[3]

[3] Source: https://towardsdatascience.com/gaussian-mixture-models-explained-6986aaf5a95

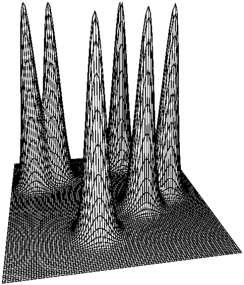
Nonparametric statistics

$$\max_{p} \sum_{i=1}^{n} w_i \cdot \log(p(x_i)) \text{ s.t. } \int_{\mathbb{R}^d} p(x)\mathrm{d}x = 1. \tag{2}$$

[4] Source: arXiv:0804.3989v1

Nonparametric statistics

$$\max_{p} \sum_{i=1}^{n} w_i \cdot \log(p(x_i)) \text{ s.t. } \int_{\mathbb{R}^d} p(x)\mathrm{d}x = 1. \qquad (2)$$

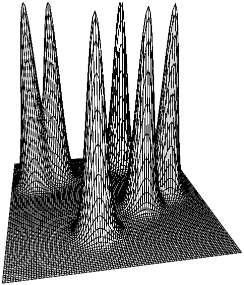


[4]

[4] Source: arXiv:0804.3989v1

Nonparametric statistics

$$\max_{p} \sum_{i=1}^{n} w_i \cdot \log(p(x_i)) \text{ s.t. } \int_{\mathbb{R}^d} p(x)\mathrm{d}x = 1. \qquad (2)$$



[4]

$\Rightarrow$ Impose constraints: shape constraints on the graph of p

[4] Source: arXiv:0804.3989v1

The log-concave **maximum likelihood estimation (MLE)** problem aims to find a Lebesgue density p that solves

$$\max_{p} \sum_{i=1}^{n} w_i \log(p(x_i)) \text{ s.t. } \log(p) \text{ is concave and } \int_{\mathbb{R}^d} p(x)\mathrm{d}x = 1. \quad (3)$$

In particular, can we find **exact solutions** to this problem?

The log-concave **maximum likelihood estimation (MLE)** problem aims to find a Lebesgue density p that solves

$$\max_{p} \sum_{i=1}^{n} w_i \log(p(x_i)) \text{ s.t. } \log(p) \text{ is concave and } \int_{\mathbb{R}^d} p(x)\mathrm{d}x = 1. \tag{3}$$

In particular, can we find **exact solutions** to this problem?

Our contribution:

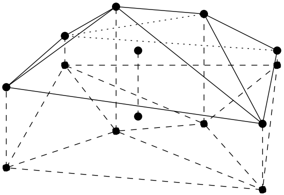
- In general, the MLE for log-concave densities is transcendental.
- For specific cases we found closed form solutions (one cell in one dimension).
- α-certify numerical solutions

Structure of the talk

1. Motivating example
2. Geometry of log-concave MLE
3. Transcendental in general
4. Closed-form solutions in special cases
5. Certifying solutions with Smale's α-theory
6. Conclusions

Motivating example

$\log(\hat{p}(\cdot))$ is a piecewise linear function with regions of linearity inducing a subdivision of the convex hull of X
(Cule, Samworth, and Stewart, 2010, generalized by Robeva, Sturmfels, and Uhler, 2019):


[5]

We can numerically compute the MLE with the LogConcDEAD package ([CGS09]).

[5] Source: arXiv:0804.3989v1

Motivating example

$$X = (0,1),(0,9),(1,4),(2,4),(2,6),(3,3),(5,5),(6,3),(6,9),(7,6),(7,8),(8,9),(9,5),(9,9).$$

with uniform weights. Consider the subdivision induced by the MLE. How many cells does it have?

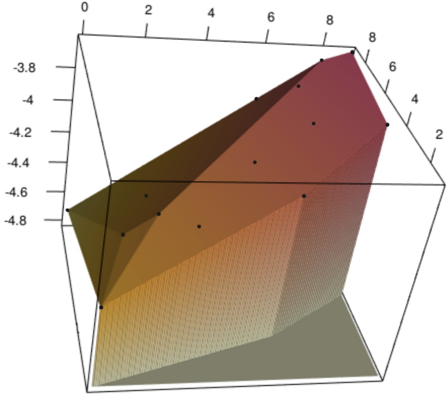
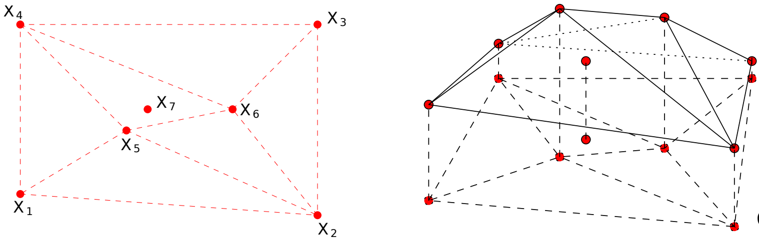


Figure: How do we find the correct number of linear pieces?

Geometry of log-concave MLE

Definition (Tent function)

*Given points $X = (x_1, x_2, \ldots, x_n) \subset \mathbb{R}^d$ and heights $y_1, \ldots, y_n$ at the points, the **tent function** $h_{X,y} : \mathbb{R}^d \to \mathbb{R}$ is the smallest concave function such that $h_{X,y}(x_i) \geq y_i$, for all i.*



[6]

[6] **Source**: `http://gac-school.imj-prg.fr/talks/sturmfels.pdf`

Geometry of log-concave MLE

Cule, Samworth, and Stewart, 2010 show for uniform weights and Robeva, Sturmfels, and Uhler, 2019 in general that

$$\max_p \sum_{i=1}^{n} w_i \log(p(x_i)) \text{ s.t. } \log(p) \text{ is concave and } \int_{\mathbb{R}^d} p(x)\mathrm{d}x = 1$$

$$\Updownarrow$$

$$\max_{y \in \mathbb{R}^n} \quad w \cdot y - \int_P \exp\left(h_{X,y}(t)\right) dt. \qquad (4)$$

We can recover the MLE $\hat{p}(\cdot) = \exp(h_{X,y^*}(\cdot))$ from optimal y^*.

Transcendentality

Theorem (Grosdos, Heaton, Kubjas, Kuznetsova, Scholten, and S., 2020)

Assume that $X \subset \mathbb{Q}^d$. If $\mathrm{vol}(\mathrm{conv}(X)) \neq 1$ then at least one coordinate of the optimal height vector y^* is transcendental. If $\mathrm{vol}(\mathrm{conv}(X)) = 1$, then all coordinates of y^* are algebraic if and only if w is in the cone over the secondary polytope $\Sigma(X)$.

Proof sketch

Remember Cule, Samworth, and Stewart, 2010's reformulation

$$\max_{y \in \mathbb{R}^n} \quad w \cdot y - \int_P \exp\left(h_{X,y}(t)\right) dt. \tag{4}$$

Proposition (Robeva, Sturmfels, and Uhler, 2019)

To find the optimal height vector y^* in (4) is to maximize the following rational-exponential objective function over $y \in \mathbb{R}^n$:

$$S(y_1, \ldots, y_n) = \quad w \cdot y \; - \sum_{\sigma \in \Delta} \sum_{i \in \sigma} \frac{\mathrm{vol}(\sigma) \cdot \exp(y_i)}{\prod_{\alpha \in \sigma \setminus i}(y_i - y_\alpha)}, \tag{5}$$

where Δ is any regular triangulation that refines the subdivision induced by the tent function $h_{X,y}$.

Proof sketch

Proposition (Grosdos, Heaton, Kubjas, Kuznetsova, Scholten, and S., 2020)

Fix a maximal triangulation Δ. *Then the critical equations from Robeva, Sturmfels, and Uhler, 2019 can be written in the form* $Ae^y = w$, *with* A *invertible,*

Proof sketch

Proposition (Grosdos, Heaton, Kubjas, Kuznetsova, Scholten, and S., 2020)

Fix a maximal triangulation Δ. Then the critical equations from Robeva, Sturmfels, and Uhler, 2019 can be written in the form $Ae^y = w$, with A invertible, such that:

$$\begin{aligned} \exp(y_1) &= f_1(y_1, y_2, \dots, y_n) \\ \exp(y_2) &= f_2(y_1, y_2, \dots, y_n) \\ &\vdots \\ \exp(y_n) &= f_n(y_1, y_2, \dots, y_n) \end{aligned} \tag{6}$$

where $f_1, \dots, f_n \in \mathbb{R}(y_1, \dots, y_n)$. If $x_1, \dots, x_n \in \mathbb{Q}^d$, then $f_1, \dots, f_n \in \mathbb{Q}(y_1, \dots, y_n)$.

Closed forms in one dimension (one cell)

$$\exp(y_1) = \frac{1}{\mathrm{vol}(\sigma)} \left((1 + y_1 - y_2)\, w_1 + w_2 \right)$$
$$\exp(y_2) = \frac{1}{\mathrm{vol}(\sigma)} \left(w_1 + (1 - y_1 + y_2)\, w_2 \right)$$

Set $\rho := w_1/w_2$.

$$\exp(y_{12}) = -\rho \frac{y_{12} + \rho^{-1} + 1}{y_{12} - \rho - 1}. \qquad (7)$$

Tools: W-Lambert function and number of solutions (branches)

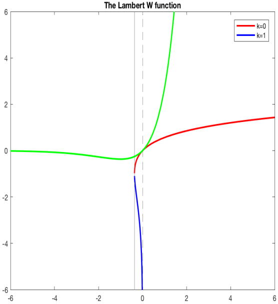
Closed forms in one dimension (one cell)

Definition 1 (W-Lambert function)

For $x \geq -\frac{1}{e}; y \in \mathbb{R}$, *the W-Lambert function* is the inverse of $x\exp(x)$:

$$y = W(y)\exp(W(y)).$$

It is not a one-to-one function and it has two branches.



Lambert function

Definition 2 (generalized W-Lambert function)

For $x, t_i, s_j \in \mathbb{R}$, consider $\frac{(x-t_1)(x-t_2)\ldots(x-t_n)}{(x-s_1)(x-s_2)\ldots(x-s_m)} \exp(x)$.
Its (generally multi-valued) inverse function at the point $a \in \mathbb{R}$ is $W(t_1, t_2, \ldots, t_n; s_1, s_2, \ldots, s_m; a)$, *the generalized W-Lambert function*.[7] We have $W(;;a) = \log(a)$.

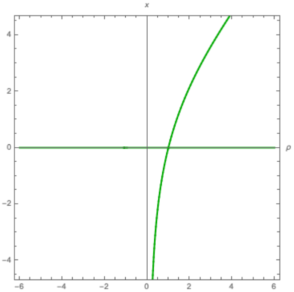
[7] (Mező and Baricz, 2017)

Closed forms in one dimension (one cell)

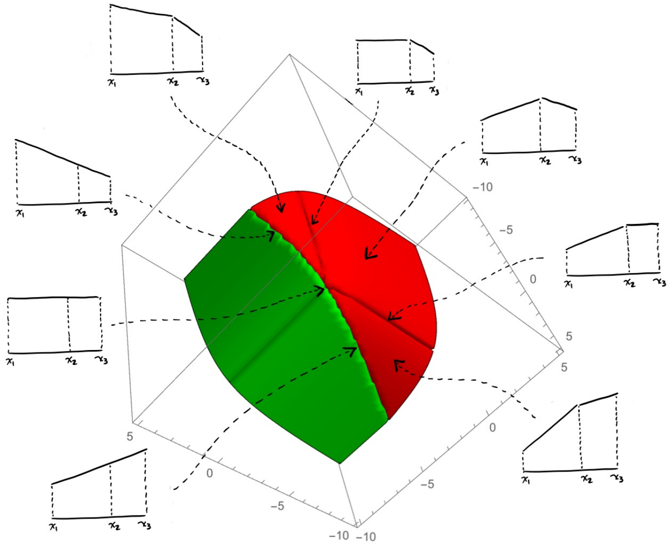
Proposition 3 (Grosdos, Heaton, Kubjas, Kuznetsova, Scholten, and S., 2020)

The tent poles corresponding to a single-cell triangulation in 1 dimension are given by:

$$y_1 = \log(w_1 W(\rho + 1; -\rho^{-1} - 1; -\rho) + w_1 + w_2) - \log(\mathrm{vol}(\sigma)),$$
$$y_2 = \log(-w_2 W(\rho + 1; -\rho^{-1} - 1; -\rho) + w_1 + w_2) - \log(\mathrm{vol}(\sigma)),$$



Closed forms in one dimension (two cells)



Closed forms in one dimension (two cells)

$$\exp(y_{12}) = \frac{(-(1+y_{12})(1+y_{23}) + \frac{v_2}{v_1}y_{12}^2)w_1 + (-1-y_{23})w_2 - w_3}{(-1-y_{23})w_1 + (-1+y_{12})(1+y_{23})w_2 + (-1+y_{12})w_3},$$

$$\exp(y_{23}) = \frac{(-1-y_{23})w_1 + (-1+y_{12})(1+y_{23})w_2 + (-1+y_{12})w_3}{-w_1 + (y_{12}-1)w_2 - ((y_{12}-1)(y_{23}-1) + \frac{v_1}{v_2}y_{23}^2)w_3}. \quad (8)$$

Polynomial-exponential systems: (Maignan, 1998) an algorithm giving the number of solutions of such a system is provided, where all the solutions are contained in a generalized open rectangle of type $I_1 \times I_2 \subset \mathbb{R}^2$, under the hypothesis that at least one of the intervals I_1 or I_2 is bounded.

Alpha certify

Can we guarantee that a numerical solution is "good enough"?

Alpha certify

Can we guarantee that a numerical solution is "good enough"?

Smale's α-theory: makes mathematically rigorous the idea of approximate zeros in the sense of quadratic convergence of Newton iterations (Smale, 1986).

Given y, do local searches until we can find α-certified y' that converges to y^* when using Newton iterations.

Algorithm 1: Testing certifiability by digit refinement

Input: A system $\nabla S_i = 0$ coming from the ith candidate subdivision and the candidate, approximate solution $y^* = (y_1, \ldots, y_n)$.

Result: A refinement of the heights $y^* + \varepsilon$ along with alpha certification of the system, or inability to certify.

1 Let p be the number of trusted significant digits (in binary) of the approximate solution y^*.

2 Expressing y^* in binary, compute the α-value for all 3^n points $y_i + \epsilon_i 2^{-p}, \epsilon_i \in \{-1, 0, 1\}$. Keep the point with the lowest alpha value, and set this as the new y_i.

3 If the alpha value is below 0.157671 stop and return the solution. If it has decreased between steps or remained the same, increase p by 1 and go to step 2. If there is no improvement for several loops in a row, stop and declare inability to certify the system.

Challenges for **Nonparametric Algebraic Statistics**

Check out our preprint at
`https://arxiv.org/abs/2003.04840`

Thank you for your attention!

Bibliography:

- Cule, Madeleine, Richard Samworth, and Michael Stewart (2010). "Maximum likelihood estimation of a multi-dimensional log-concave density". In: *J. R. Stat. Soc. Ser. B Stat. Methodol.* 72.5, pp. 545–607. issn: 1369-7412. doi: 10.1111/j.1467-9868.2010.00753.x. url: https://doi.org/10.1111/j.1467-9868.2010.00753.x (cit. on pp. 12, 15, 17).
- Grosdos, Alexandros, Alexander Heaton, Kaie Kubjas, Olga Kuznetsova, Georgy Scholten, and S. (2020). "Exact Solutions in Log-Concave Maximum Likelihood Estimation". In: url: https://arxiv.org/abs/2003.04840 (cit. on pp. 16, 18, 19, 23).
- Maignan, Aude (1998). "Solving One and Two-dimensional Exponential Polynomial Systems". In: *Proc. of the 1998 Int. Symp. Symb. and Algebr. Comput.* ISSAC '98. Rostock, Germany: ACM, pp. 215–221. isbn: 1-58113-002-3. doi: 10.1145/281508.281616. url: http://doi.acm.org/10.1145/281508.281616 (cit. on p. 25).
- Mező, István and Árpád Baricz (2017). "On the generalization of the Lambert W function". In: *Trans. Amer. Math. Soc.* 369.11, pp. 7917–7934. issn: 0002-9947. doi: 10.1090/tran/6911. url: https://doi.org/10.1090/tran/6911 (cit. on p. 22).
- Robeva, Elina, Bernd Sturmfels, and Caroline Uhler (2019). "Geometry of log-concave density estimation". In: *Discrete Comput. Geom.* 61.1, pp. 136–160. issn: 0179-5376. doi: 10.1007/s00454-018-0024-y. url: https://doi.org/10.1007/s00454-018-0024-y (cit. on pp. 12, 15, 17–19).
- Smale, Steve (1986). "Newton's method estimates from data at one point". In: *The merging of disciplines: new directions in pure, applied, and computational mathematics.* Springer, New York, pp. 185–196 (cit. on pp. 26, 27).