Maximum Likelihood Density Estimation under Total Positivity

Elina Robeva
MIT

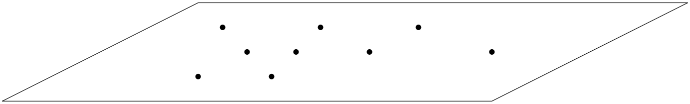
joint work with Bernd Sturmfels, Ngoc Tran, and Caroline Uhler
arXiv:1806.10120

ICERM Workshop on Nonlinear Algebra in Applications

November 12, 2018

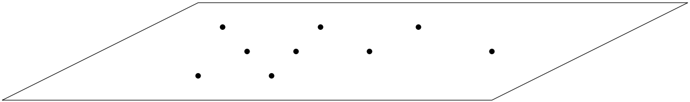
Density estimation

Given i.i.d. samples $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$ from an unknown distribution on $\mathbb{R}^d$ with density p, can we estimate p?



Density estimation

Given i.i.d. samples $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$ from an unknown distribution on $\mathbb{R}^d$ with density p, can we estimate p?



- parametric: assume that p lies in some parametric family, and estimate parameters
  - finite-dimensional problem
  - too restrictive; the real-world distribution might not lie in the specified parametric family
- non-parametric: assume that p lies in a non-parametric family, e.g. impose shape-constraints on p (convex, log-concave, monotone, etc.)
  - infinite-dimensional problem
  - need constraints that are:
    - strong enough so that there is no spiky behavior
    - weak enough so that function class is large

Shape-constrained density estimation

- monotonically decreasing densities: [Grenander 1956, Rao 1969]
- convex densities: [Anevski 1994, Groeneboom, Jongbloed, and Wellner 2001]
- **log-concave** densities: [Cule, Samworth, and Stewart 2008]
- generalized additive models with shape constraints: [Chen and Samworth 2016]

- this talk: **totally positive and log-concave** densities

MTP_2 distributions

- A distribution with density p on $\mathcal{X} \subseteq \mathbb{R}^d$ is *multivariate totally positive of order 2* (or MTP_2) if

$$p(x)p(y) \leq p(x \wedge y)p(x \vee y) \quad \text{for all } x, y \in \mathcal{X},$$

  where $x \wedge y$ and $x \vee y$ are the componentwise minimum and maximum.

MTP_2 distributions

- A distribution with density p on $\mathcal{X} \subseteq \mathbb{R}^d$ is *multivariate totally positive of order 2* (or MTP_2) if

$$p(x)p(y) \leq p(x \wedge y)p(x \vee y) \quad \text{for all } x, y \in \mathcal{X},$$

where $x \wedge y$ and $x \vee y$ are the componentwise minimum and maximum.

- MTP_2 is the same as *log-supermodular*:

$$\log(p(x))+\log(p(y)) \leq \log(p(x \wedge y))+\log(p(x \vee y)) \quad \text{for all } x, y \in \mathcal{X}.$$

MTP_2 distributions

- A distribution with density p on $\mathcal{X} \subseteq \mathbb{R}^d$ is *multivariate totally positive of order 2* (or *MTP_2*) if

$$p(x)p(y) \le p(x \wedge y)p(x \vee y) \quad \text{for all } x, y \in \mathcal{X},$$

  where $x \wedge y$ and $x \vee y$ are the componentwise minimum and maximum.

- MTP_2 is the same as *log-supermodular*:

$$\log(p(x)) + \log(p(y)) \le \log(p(x \wedge y)) + \log(p(x \vee y)) \quad \text{for all } x, y \in \mathcal{X}.$$

- A random vector X taking values in $\mathbb{R}^d$ is *positively associated* if for any non-decreasing functions $\phi, \psi : \mathbb{R}^d \to \mathbb{R}$

$$\text{cov}(\phi(X), \psi(X)) \ge 0.$$

- MTP_2 implies positive association (Fortuin Kasteleyn Ginibre inequality, 1971).

Properties of MTP_2 distributions

Theorem (Fallat, Lauritzen, Sadeghi, Uhler, Wermuth and Zwiernik, 2015)
If $X = (X_1, \ldots, X_d)$ is MTP_2, then

(i) *any marginal distribution is MTP_2,*

(ii) *any conditional distribution is MTP_2,*

(iii) *X has the marginal independence structure*

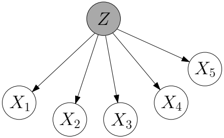
$$X_i \perp\!\!\!\perp X_j \iff cov(X_i, X_j) = 0.$$

Theorem (Karlin and Rinott, 1980)
If $p(x) > 0$ and p is MTP_2 for any pair of coordinates when the others are held constant, then p is MTP_2.

Examples of MTP_2 distributions

- A Gaussian random variable $X \sim \mathcal{N}(\mu, \Sigma)$ is MTP_2 whenever Σ^{-1} is an M-matrix, i.e. its off-diagonal entries are nonpositive.

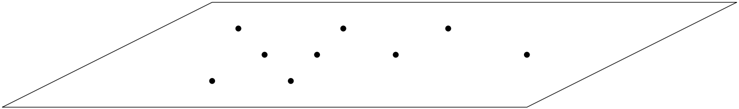
- The joint distribution of observed variables influenced by one hidden variable



- Very common in real data: e.g. IQ test scores, phylogenetics data, financial econometrics data, and others

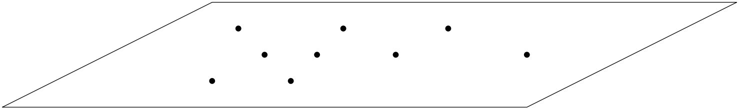
- Many models imply MTP_2:
  - Ferromagnetic Ising models
  - Order statistics of i.i.d. variables
  - Brownian motion tree models
  - Latent tree models (e.g. single factor analysis models)

Maximum Likelihood Estimation



Given i.i.d. samples $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$ with weights $w = (w_1, \ldots, w_n)$ (where $w_1, \ldots, w_n \geq 0$, $\sum w_i = 1$) from a distribution p on $\mathbb{R}^d$, can we estimate p?

Maximum Likelihood Estimation

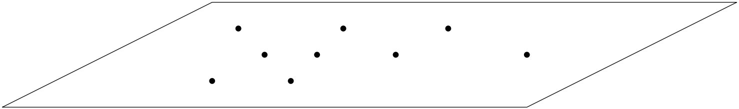


Given i.i.d. samples $X = \{x_1, \dots, x_n\} \subset \mathbb{R}^d$ with weights $w = (w_1, \dots, w_n)$ (where $w_1, \dots, w_n \geq 0$, $\sum w_i = 1$) from a distribution p on $\mathbb{R}^d$, can we estimate p?

The *log-likelihood* of observing $X = \{x_1, \dots, x_n\}$ with weights $w = (w_1, \dots, w_n)$ if they are drawn i.i.d. from p is (up to an additive constant)

$$\ell_p(X, w) := \sum_{i=1}^{n} w_i \log(p(x_i)).$$

Maximum Likelihood Estimation



Given i.i.d. samples $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$ with weights $w = (w_1, \ldots, w_n)$ (where $w_1, \ldots, w_n \geq 0$, $\sum w_i = 1$) from a distribution p on $\mathbb{R}^d$, can we estimate p?

The *log-likelihood* of observing $X = \{x_1, \ldots, x_n\}$ with weights $w = (w_1, \ldots, w_n)$ if they are drawn i.i.d. from p is (up to an additive constant)

$$\ell_p(X, w) := \sum_{i=1}^{n} w_i \log(p(x_i)).$$

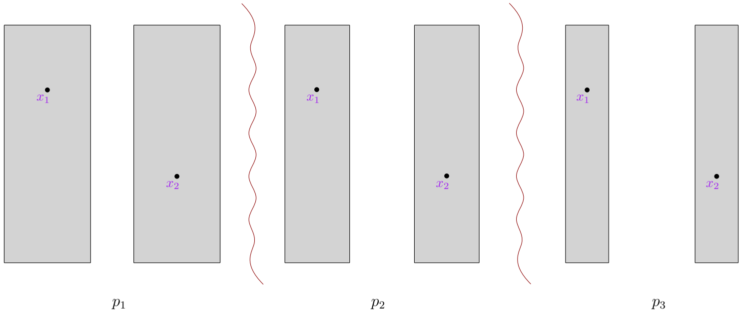
We would like to

$$\begin{aligned} \text{maximize}_p \quad & \sum_{i=1}^{n} w_i \log(p(x_i)) \\ \text{s.t.} \quad & p \text{ is an } \mathrm{MTP}_2 \text{ density.} \end{aligned}$$

Maximum Likelihood Estimation under MTP_2

Suppose we observe two points: $X = \{x_1, x_2\} \subset \mathbb{R}^2$. We can find a sequence of MTP_2 densities $p_1, p_2, p_3, \ldots$ such that

$$\ell_{p_n}(X) \to \infty \quad \text{as } n \to \infty.$$

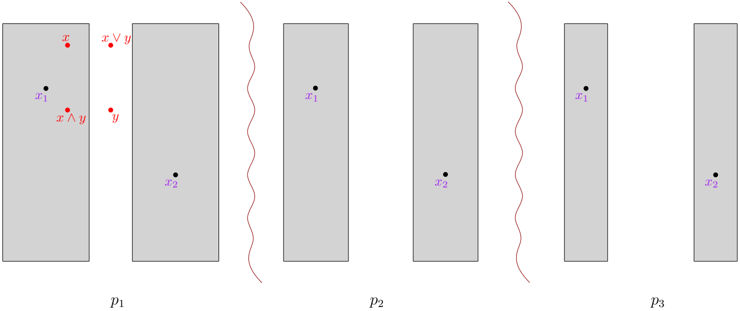


Thus, the MLE doesn't exist.

Maximum Likelihood Estimation under MTP_2

Suppose we observe two points: $X = \{x_1, x_2\} \subset \mathbb{R}^2$. We can find a sequence of MTP_2 densities $p_1, p_2, p_3, \ldots$ such that

$$\ell_{p_n}(X) \to \infty \quad \text{as } n \to \infty.$$



Thus, the MLE doesn't exist.

Maximum Likelihood Estimation under MTP_2

To ensure that the likelihood function is bounded, we impose the condition that p is log-concave.

$$\begin{aligned} \text{maximize}_p \quad & \sum_{i=1}^{n} w_i \log(p(x_i)) \\ \text{s.t.} \quad & p \text{ is an } \mathrm{MTP}_2 \text{ density}, \end{aligned}$$

Maximum Likelihood Estimation under MTP_2

To ensure that the likelihood function is bounded, we impose the condition that p is log-concave.

$$\begin{aligned} \text{maximize}_p \quad & \sum_{i=1}^{n} w_i \log(p(x_i)) \\ \text{s.t.} \quad & p \text{ is an } \mathrm{MTP}_2 \text{ density}, \\ \text{and} \quad & p \text{ is log-concave.} \end{aligned}$$

A function $f : \mathbb{R}^d \to \mathbb{R}$ is *log-concave* if its logarithm is concave.

Maximum Likelihood Estimation under MTP_2

To ensure that the likelihood function is bounded, we impose the condition that p is log-concave.

$$\begin{aligned} \text{maximize}_p \quad & \sum_{i=1}^{n} w_i \log(p(x_i)) \\ \text{s.t.} \quad & p \text{ is an } \mathrm{MTP}_2 \text{ density,} \\ \text{and} \quad & p \text{ is log-concave.} \end{aligned}$$

A function $f : \mathbb{R}^d \to \mathbb{R}$ is *log-concave* if its logarithm is concave.

- Log-concavity is a natural assumption because it ensures the density is continuous and includes many known families of parametric distributions.
- Log-concave families:
  - Gaussian; Uniform(a, b); Gamma(k, θ) for $k \geq 1$; Beta(a, b) for $a, b \geq 1$.
- Maximum likelihood estimation under log-concavity is a well-studied problem (Cule et al. 2008, Dümbgen et al. 2009, Schuhmacher et al. 2010, ...).

Maximum Likelihood Estimation under Log-Concavity

$$\begin{aligned} \text{maximize}_p \quad & \sum_{i=1}^{n} w_i \log(p(x_i)) \\ \text{s.t.} \quad & p \text{ is a density} \\ \text{and} \quad & p \text{ is log-concave.} \end{aligned}$$

Maximum Likelihood Estimation under Log-Concavity

$$\begin{aligned} \text{maximize}_p \quad & \sum_{i=1}^{n} w_i \log(p(x_i)) \\ \text{s.t.} \quad & p \text{ is a density} \\ \text{and} \quad & p \text{ is log-concave.} \end{aligned}$$

Theorem (Cule, Samworth and Stewart 2008)

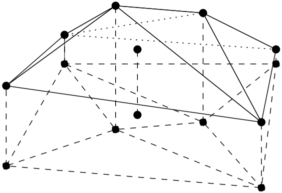
- *With probability 1, a log-concave maximum likelihood estimator $\hat{p}$ exists and is unique.*

Maximum Likelihood Estimation under Log-Concavity

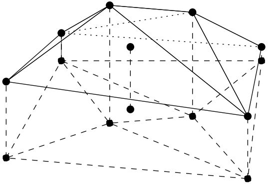
$$\begin{aligned} \text{maximize}_p \quad & \sum_{i=1}^{n} w_i \log(p(x_i)) \\ \text{s.t.} \quad & p \text{ is a density} \\ \text{and} \quad & p \text{ is log-concave.} \end{aligned}$$

Theorem (Cule, Samworth and Stewart 2008)

- *With probability 1, a log-concave maximum likelihood estimator $\hat{p}$ exists and is unique.*
- *Moreover, $\log(\hat{p})$ is a 'tent-function' supported on the convex hull of the data $P(X) = conv(x_1, \ldots, x_n)$.*



Optimizing over Tent Functions

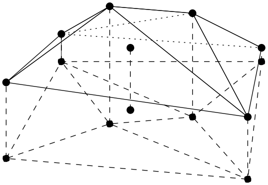


Given points $X = \{x_1, \ldots, x_n\}$ and heights $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$, the *tent function*

$$h_{X,y} : \mathbb{R}^d \to \mathbb{R}$$

is the smallest concave function such that $h_{X,y}(x_i) \geq y_i$ for all i. Thus, $\hat{p} = \exp(h_{X,y})$ for some y.

Optimizing over Tent Functions



Given points $X = \{x_1, \ldots, x_n\}$ and heights $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$, the *tent function*

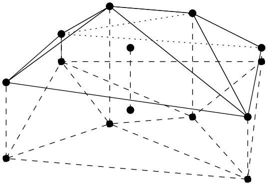
$$h_{X,y} : \mathbb{R}^d \to \mathbb{R}$$

is the smallest concave function such that $h_{X,y}(x_i) \geq y_i$ for all i. Thus, $\hat{p} = \exp(h_{X,y})$ for some y.

$$\begin{aligned} \text{maximize}_p \quad & \sum_{i=1}^{n} w_i \log(p(x_i)) \\ \text{s.t.} \quad & p \text{ is a density} \\ \text{and} \quad & p \text{ is log-concave.} \end{aligned}$$

INFINITE DIMENSIONAL

Optimizing over Tent Functions



Given points $X = \{x_1, \ldots, x_n\}$ and heights $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$, 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. Thus, $\hat{p} = \exp(h_{X,y})$ for some y.

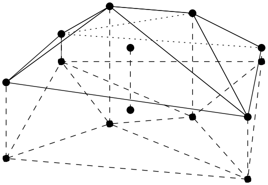
$$\begin{aligned} \text{maximize}_p \quad & \sum_{i=1}^{n} w_i \log(p(x_i)) \\ \text{s.t.} \quad & p \text{ is a density} \\ \text{and} \quad & p \text{ is log-concave.} \end{aligned}$$

$$\begin{aligned} \text{maximize}_{y \in \mathbb{R}^n} \quad & \sum_{i=1}^{n} w_i \log \exp(h_{X,y}(x_i)) \\ \text{s.t.} \quad & \exp(h_{X,y}) \text{ is a density.} \end{aligned}$$

INFINITE DIMENSIONAL

FINITE DIMENSIONAL

Optimizing over Tent Functions



Given points $X = \{x_1, \ldots, x_n\}$ and heights $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$, 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. Thus, $\hat{p} = \exp(h_{X,y})$ for some y.

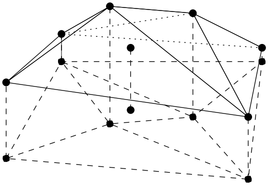
$$\begin{aligned} \text{maximize}_p \quad & \sum_{i=1}^n w_i \log(p(x_i)) \\ \text{s.t.} \quad & p \text{ is a density} \\ \text{and} \quad & p \text{ is log-concave.} \end{aligned}$$

$$\begin{aligned} \text{maximize}_{y \in \mathbb{R}^n} \quad & \sum_{i=1}^n w_i h_{X,y}(x_i) \\ \text{s.t.} \quad & \exp(h_{X,y}) \text{ is a density.} \end{aligned}$$

INFINITE DIMENSIONAL

FINITE DIMENSIONAL

Optimizing over Tent Functions



Given points $X = \{x_1, \ldots, x_n\}$ and heights $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$, 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. Thus, $\hat{p} = \exp(h_{X,y})$ for some y.

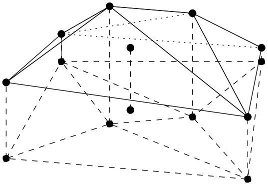
$$\begin{aligned} \text{maximize}_p \quad & \sum_{i=1}^{n} w_i \log(p(x_i)) \\ \text{s.t.} \quad & p \text{ is a density} \\ \text{and} \quad & p \text{ is log-concave.} \end{aligned}$$

INFINITE DIMENSIONAL

$$\begin{aligned} \text{maximize}_{y \in \mathbb{R}^n} \quad & \sum_{i=1}^{n} w_i y_i \\ \text{s.t.} \quad & \exp(h_{X,y}) \text{ is a density.} \end{aligned}$$

FINITE DIMENSIONAL

Optimizing over Tent Functions



Given points $X = \{x_1, \ldots, x_n\}$ and heights $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$, 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. Thus, $\hat{p} = \exp(h_{X,y})$ for some y.

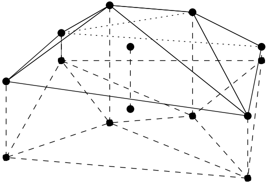
$$\begin{aligned} \text{maximize}_p \quad & \sum_{i=1}^{n} w_i \log(p(x_i)) \\ \text{s.t.} \quad & p \text{ is a density} \\ \text{and} \quad & p \text{ is log-concave.} \end{aligned}$$

INFINITE DIMENSIONAL

$$\begin{aligned} \text{maximize}_{y \in \mathbb{R}^n} \quad & \sum_{i=1}^{n} w_i y_i \\ \text{s.t.} \quad & \int \exp(h_{X,y}(t))dt = 1 \end{aligned}$$

FINITE DIMENSIONAL

Optimizing over Tent Functions



Given points $X = \{x_1, \ldots, x_n\}$ and heights $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$, 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. Thus, $\hat{p} = \exp(h_{X,y})$ for some y.

$$\begin{array}{ll} \text{maximize}_p & \displaystyle\sum_{i=1}^{n} w_i \log(p(x_i)) \\ \text{s.t.} & p \text{ is a density} \\ \text{and} & p \text{ is log-concave.} \end{array}$$

$$\max_{y \in \mathbb{R}^n} \quad \sum_{i=1}^{n} w_i y_i - \int \exp(h_{X,y}(t))dt$$

INFINITE DIMENSIONAL

FINITE DIMENSIONAL

Maximum Likelihood Estimation under Log-concavity and MTP_2

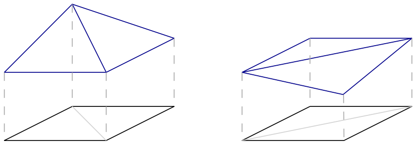
Questions:

1. Does the MLE under log-concavity and MTP_2 exist with probability 1 and, if so, is it unique?
2. What is the shape of the MLE under log-concavity and MTP_2?
   2.1 What is the support of the MLE?
   2.2 Is the MLE always exp(tent function)?
3. Which tent functions are allowed?
4. Can we compute the MLE?

Maximum Likelihood Estimation under Log-concavity and MTP_2

Questions:

1. Does the MLE under log-concavity and MTP_2 exist with probability 1 and, if so, is it unique?
2. What is the shape of the MLE under log-concavity and MTP_2?
   2.1 What is the support of the MLE?
   2.2 Is the MLE always exp(tent function)?
3. Which tent functions are allowed?
4. Can we compute the MLE?



Recall: p is MTP_2 if and only if $\log(p)$ is *supermodular*, i.e.

$$\log p(x) + \log p(y) \leq \log p(x \wedge y) + \log p(x \vee y), \text{ for all } x, y.$$

Existence and Uniqueness of the MLE

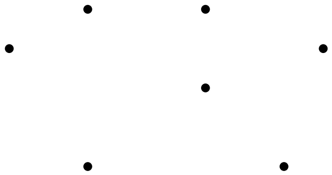
Theorem (R., Sturmfels, Tran, Uhler)

The maximum likelihood estimator under log-concavity and MTP_2 exists and is unique with probability 1 as long as there are at least 3 samples.

Proof uses convergence properties for log-concave distributions, and does not shed light on the shape of the MLE.

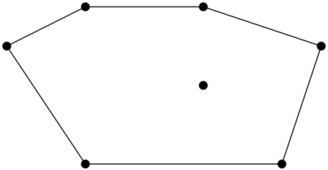
The Support of the MLE

Consider the following samples:



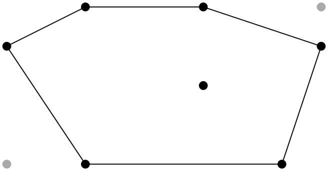
The Support of the MLE

Under log-concavity, the support of the MLE is the convex hull:



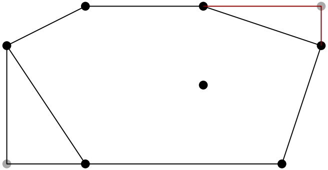
The Support of the MLE

Under log-concavity and MTP_2 we need the density to be nonzero at more points:



The Support of the MLE

Under log-concavity and MTP_2 we need the density to be nonzero at more points:



and we need the convex hull of all of these points.

Support of the MLE = "min-max convex hull" of X.

The Min-Max Convex Hull

Definition

$$\begin{aligned} \mathrm{MM}(X) &= \text{smallest } \textit{min-max closed} \text{ set } S \text{ containing } X, \text{ i.e. } x, y \in S \Rightarrow x \wedge y, x \vee y \in S \\ \mathrm{MMconv}(X) &= \text{smallest } \textit{min-max closed and convex} \text{ set containing } X. \end{aligned}$$

The Min-Max Convex Hull

Definition

$$\begin{aligned} \mathrm{MM}(X) &= \text{smallest } \textit{min-max closed} \text{ set } S \text{ containing } X, \text{ i.e. } x, y \in S \Rightarrow x \wedge y, x \vee y \in S \\ \mathrm{MMconv}(X) &= \text{smallest } \textit{min-max closed and convex} \text{ set containing } X. \end{aligned}$$

- How can we find MMconv(X) for $X = \{x_1, \dots, x_n\} \subseteq \mathbb{R}^d$?

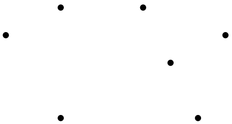
The Min-Max Convex Hull

Definition

MM(X) $=$ smallest *min-max closed* set S containing X, i.e. $x, y \in S \Rightarrow x \wedge y, x \vee y \in S$

MMconv(X) $=$ smallest *min-max closed and convex* set containing X.

- How can we find MMconv(X) for $X = \{x_1, \dots, x_n\} \subseteq \mathbb{R}^d$?
- Intuitive first proposal:



Start with X.

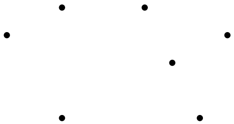
Add points to X
until we get MM(X).

The Min-Max Convex Hull

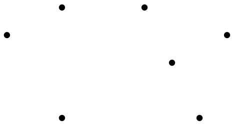
Definition

$$\mathrm{MM}(X) = \text{smallest } \textit{min-max closed} \text{ set } S \text{ containing } X, \text{ i.e. } x, y \in S \Rightarrow x \wedge y, x \vee y \in S$$
$$\mathrm{MMconv}(X) = \text{smallest } \textit{min-max closed and convex} \text{ set containing } X.$$

- How can we find $\mathrm{MMconv}(X)$ for $X = \{x_1, \ldots, x_n\} \subseteq \mathbb{R}^d$?
- Intuitive first proposal:



Start with X.



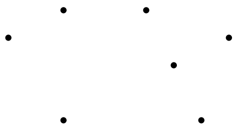
Add points to X
until we get $\mathrm{MM}(X)$.

The Min-Max Convex Hull

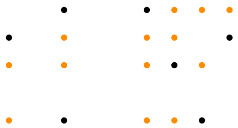
Definition

$$\begin{aligned} \mathrm{MM}(X) &= \text{smallest } \textit{min-max closed} \text{ set } S \text{ containing } X\text{, i.e. } x, y \in S \Rightarrow x \wedge y, x \vee y \in S \\ \mathrm{MMconv}(X) &= \text{smallest } \textit{min-max closed and convex} \text{ set containing } X. \end{aligned}$$

- How can we find $\mathrm{MMconv}(X)$ for $X = \{x_1, \dots, x_n\} \subseteq \mathbb{R}^d$?
- Intuitive first proposal:



Start with X.



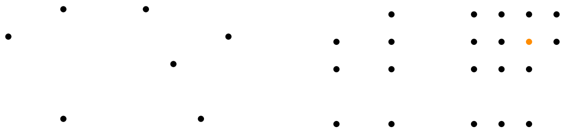
Add points to X
until we get $\mathrm{MM}(X)$.

The Min-Max Convex Hull

Definition

$$\mathrm{MM}(X) = \text{smallest } \textit{min-max closed} \text{ set } S \text{ containing } X, \text{ i.e. } x, y \in S \Rightarrow x \wedge y, x \vee y \in S$$
$$\mathrm{MMconv}(X) = \text{smallest } \textit{min-max closed and convex} \text{ set containing } X.$$

- How can we find MMconv(X) for $X = \{x_1, \ldots, x_n\} \subseteq \mathbb{R}^d$?
- Intuitive first proposal:



Start with X.

Add points to X until we get MM(X).

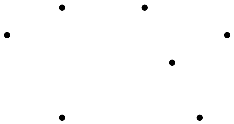
The Min-Max Convex Hull

Definition

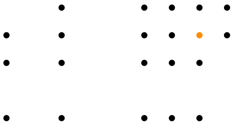
$$\mathrm{MM}(X) = \text{smallest } \textit{min-max closed} \text{ set } S \text{ containing } X\text{, i.e. } x, y \in S \Rightarrow x \wedge y, x \vee y \in S$$

$$\mathrm{MMconv}(X) = \text{smallest } \textit{min-max closed and convex} \text{ set containing } X.$$

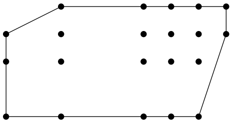
- How can we find $\mathrm{MMconv}(X)$ for $X = \{x_1, \ldots, x_n\} \subseteq \mathbb{R}^d$?
- Intuitive first proposal:



Start with X.



Add points to X until we get $\mathrm{MM}(X)$.



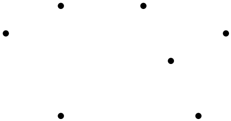
Take $\mathrm{conv}(MM(X))$.

The Min-Max Convex Hull

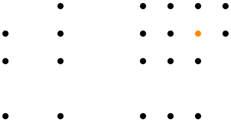
Definition

$$\mathrm{MM}(X) = \text{smallest } \textit{min-max closed} \text{ set } S \text{ containing } X\text{, i.e. } x, y \in S \Rightarrow x \wedge y, x \vee y \in S$$
$$\mathrm{MMconv}(X) = \text{smallest } \textit{min-max closed and convex} \text{ set containing } X.$$

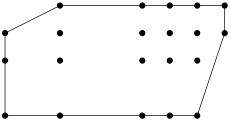
- How can we find $\mathrm{MMconv}(X)$ for $X = \{x_1, \ldots, x_n\} \subseteq \mathbb{R}^d$?
- Intuitive first proposal:



Start with X.



Add points to X until we get $\mathrm{MM}(X)$.



Take $\mathrm{conv}(MM(X))$.

- Is it always true that $\mathrm{MMconv}(X) = \mathrm{conv}(\mathrm{MM}(X))$?

The Min-Max Convex Hull

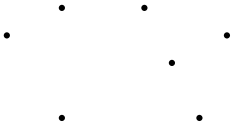
Definition

$\mathrm{MM}(X)$ = smallest *min-max closed* set S containing X, i.e. $x, y \in S \Rightarrow x \wedge y, x \vee y \in S$

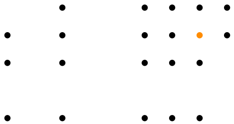
$\mathrm{MMconv}(X)$ = smallest *min-max closed and convex* set containing X.

- How can we find $\mathrm{MMconv}(X)$ for $X = \{x_1, \dots, x_n\} \subseteq \mathbb{R}^d$?

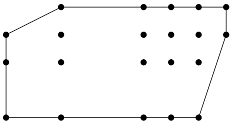
- Intuitive first proposal:



Start with X.



Add points to X until we get $\mathrm{MM}(X)$.



Take $\mathrm{conv}(MM(X))$.

- Is it always true that $\mathrm{MMconv}(X) = \mathrm{conv}(\mathrm{MM}(X))$? No!

The Min-Max Convex Hull

Lemma

Let $X = \{x_1, \ldots, x_n\}$. *If* $X \subseteq \mathbb{R}^2$ *or* $X \subseteq \{0,1\}^d$, *then,*

$$MMconv(X) = conv(MM(X)).$$

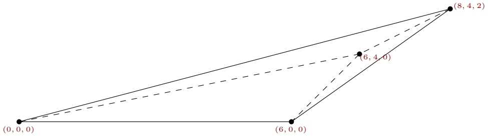
The Min-Max Convex Hull

Lemma

Let $X = \{x_1, \ldots, x_n\}$. If $X \subseteq \mathbb{R}^2$ or $X \subseteq \{0,1\}^d$, then,

$$MMconv(X) = conv(MM(X)).$$

Now, consider $X = \{(0,0,0), (6,0,0), (6,4,0), (8,4,2)\} \subseteq \mathbb{R}^3$.



It turns out that

$$\mathrm{MM}(X) = X.$$

But

$$\mathrm{conv}(\mathrm{MM}(X)) \text{ is } \textbf{not min-max closed}!$$

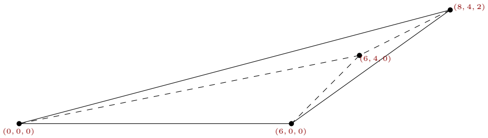
The Min-Max Convex Hull

Lemma

Let $X = \{x_1, \ldots, x_n\}$. If $X \subseteq \mathbb{R}^2$ or $X \subseteq \{0,1\}^d$, then,

$$MMconv(X) = conv(MM(X)).$$

Now, consider $X = \{(0,0,0), (6,0,0), (6,4,0), (8,4,2)\} \subseteq \mathbb{R}^3$.



It turns out that

$$\mathrm{MM}(X) = X.$$

But

conv(MM(X)) is **not min-max closed**!

This is because:

$$(6, 4, \frac{3}{2}) = \max\{(6,4,0), (6,3,\frac{3}{2})\} \notin \mathrm{conv}(\mathrm{MM}(X)).$$

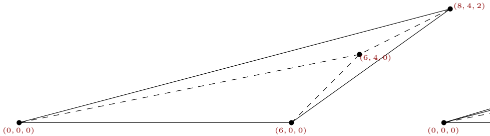
The Min-Max Convex Hull

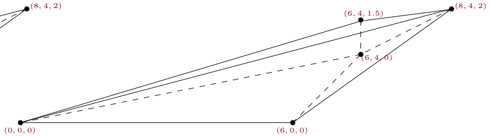
Lemma

Let $X = \{x_1, \dots, x_n\}$. If $X \subseteq \mathbb{R}^2$ or $X \subseteq \{0,1\}^d$, then,

$$MMconv(X) = conv(MM(X)).$$

Now, consider $X = \{(0,0,0), (6,0,0), (6,4,0), (8,4,2)\} \subseteq \mathbb{R}^3$.





It turns out that

$$\mathrm{MM}(X) = X.$$

But

$$\text{conv}(\text{MM}(X)) \text{ is } \textbf{not min-max closed}!$$

This is because:

$$(6, 4, \frac{3}{2}) = \max\{(6,4,0), (6,3,\frac{3}{2})\} \notin \text{conv}(\text{MM}(X)).$$

Therefore,

$$\text{conv}(\text{MM}(X)) \subsetneq \text{MMconv}(X).$$

The 2-D Projections Theorem

Theorem (The 2-D Projections Theorem)

For any finite subset $X \subseteq \mathbb{R}^d$. Then we have

$$MMconv(X) = \bigcap_{1 \le i < j \le d} \pi_{ij}^{-1} \left(conv(\pi_{ij}(MM(X)) \right).$$

$$\pi_{ij} : \mathbb{R}^d \to \mathbb{R}, \quad x \mapsto (x_i, x_j).$$

The 2-D Projections Theorem

Theorem (The 2-D Projections Theorem)

For any finite subset $X \subseteq \mathbb{R}^d$. Then we have

$$MMconv(X) = \bigcap_{1 \le i < j \le d} \pi_{ij}^{-1}\left(conv(\pi_{ij}(MM(X))\right).$$

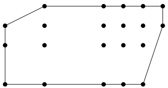
$$\pi_{ij} : \mathbb{R}^d \to \mathbb{R}, \quad x \mapsto (x_i, x_j).$$

Corollary (Queyranne and Tardella, 2006)

A subset C in $\mathbb{R}^d$ is a min-max closed convex polytope if and only if it is defined by a finite collection of bimonotone linear inequalities.

A linear inequality $a \cdot x + b \le 0$ is *bimonotone* if it has the form

$$a_i x_i + a_j x_j + b \le 0, \quad \text{where} \quad a_i a_j \le 0.$$



Back to Log-concave and MTP_2 Maximum Likelihood Estimation

1. Does the MLE under log-concavity and MTP_2 exist with probability 1 and, if so, is it unique? Yes.
2. What is the shape of the MLE under log-concavity and MTP_2?
   2.1 What is the support of the MLE? MMconv(X); We can compute it.
   2.2 Is the MLE always exp(tent function)?
3. Which tent functions are allowed?
4. Can we compute the MLE?

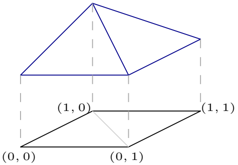
Supermodular Tent Functions

Recall that $p = \exp(h)$ is $\textbf{MTP}_2$ if and only if h is **supermodular**, i.e.

$$h(x) + h(y) \leq h(x \wedge y) + h(x \vee y), \qquad \text{for all } x, y \in \mathbb{R}^d.$$

Theorem (R., Sturmfels, Tran, Uhler)

Let $X \subset \mathbb{R}^d$ be a finite set of points. A tent function h is supermodular if and only if all of the walls of the subdivision h induces are ***bimonotone****.*



Remark

If we want to find the best supermodular $h_{X,y}$, we need to optimize over the set of heights y that induce bimonotone subdivisions.

- *In general not convex.*
- *Example: $X = \{0, 1\} \times \{0, 1\} \times \{0, 1, 2\}$.*

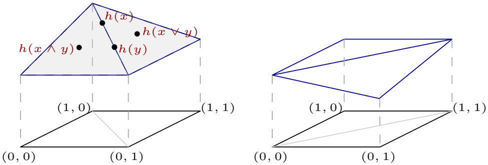
Supermodular Tent Functions

Recall that $p = \exp(h)$ is $\mathbf{MTP}_2$ if and only if h is **supermodular**, i.e.

$$h(x) + h(y) \leq h(x \wedge y) + h(x \vee y), \qquad \text{for all } x, y \in \mathbb{R}^d.$$

Theorem (R., Sturmfels, Tran, Uhler)

Let $X \subset \mathbb{R}^d$ be a finite set of points. A tent function h is supermodular if and only if all of the walls of the subdivision h induces are **bimonotone**.



Remark

If we want to find the best supermodular $h_{X,y}$, we need to optimize over the set of heights y that induce bimonotone subdivisions.

- *In general not convex.*
- *Example: $X = \{0,1\} \times \{0,1\} \times \{0,1,2\}$.*

Is the MLE is the exponential of a tent function?

1. Does the MLE under log-concavity and MTP_2 exist with probability 1 and, if so, is it unique? Yes.
2. What is the shape of the MLE under log-concavity and MTP_2?
   2.1 What is the support of the MLE? MMconv(X); We can compute it.
   2.2 Is the MLE always exp(tent function)?
3. Which tent functions are allowed? Bimonotone tent functions.
4. Can we compute the MLE?

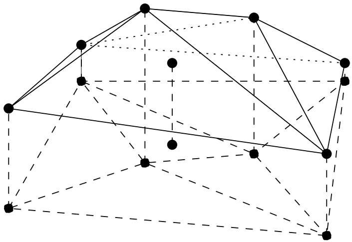
Why is the Log-concave MLE the exponential of a tent function?

Recall:

$$\begin{aligned} \text{maximize}_p \quad & \sum_{i=1}^{n} w_i \log(p(x_i)) \\ \text{s.t.} \quad & p \text{ is a density} \\ \text{and} \quad & p \text{ is log-concave.} \end{aligned}$$

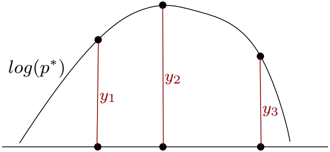
Theorem (Cule, Samworth and Stewart 2008)

- *With probability 1, a log-concave maximum likelihood estimator $\hat{p}$ exists and is unique.*
- *Moreover, $\log(\hat{p})$ is a 'tent-function' supported on the convex hull of the data $P(X) = conv(x_1, \ldots, x_n)$.*



Why is the Log-concave MLE the exponential of a tent function?

$$\begin{array}{rl} \text{maximize}_p & \displaystyle\sum_{i=1}^{n} w_i \log(p(x_i)) \\ \text{s.t.} & p \text{ is a density} \\ \text{and} & p \text{ is log-concave.} \end{array}$$

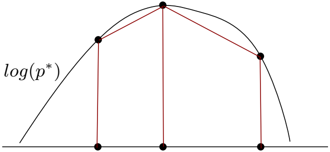


Proof of theorem:

- Suppose that p^* is the MLE and that $\log p^*$ is not a tent function.
- Let $y_i = \log p^*(x_i), i = 1, \ldots, n$.
- Consider $p = \exp(h_{X,y})$. It gives a higher objective value than p^*.
- Thus, p^* has to be a tent function.

Why is the Log-concave MLE the exponential of a tent function?

$$
\begin{aligned}
\text{maximize}_p \quad & \sum_{i=1}^{n} w_i \log(p(x_i)) \\
\text{s.t.} \quad & p \text{ is a density} \\
\text{and} \quad & p \text{ is log-concave.}
\end{aligned}
$$

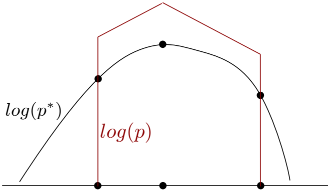


Proof of theorem:

- Suppose that p^* is the MLE and that $\log p^*$ is not a tent function.
- Let $y_i = \log p^*(x_i), i = 1, \dots, n$.
- Consider $p = \exp(h_{X,y})$. It gives a higher objective value than p^*.
- Thus, p^* has to be a tent function.

Why is the Log-concave MLE the exponential of a tent function?

$$\begin{aligned} \text{maximize}_p \quad & \sum_{i=1}^{n} w_i \log(p(x_i)) \\ \text{s.t.} \quad & p \text{ is a density} \\ \text{and} \quad & p \text{ is log-concave.} \end{aligned}$$

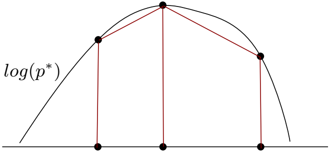


Proof of theorem:

- Suppose that p^* is the MLE and that $\log p^*$ is not a tent function.
- Let $y_i = \log p^*(x_i), i = 1, \ldots, n$.
- Consider $p = \exp(h_{X,y})$. It gives a higher objective value than p^*.
- Thus, p^* has to be a tent function.

Proving that the Log-concave MTP_2 MLE is the exponential of a tent function

$$
\begin{array}{lll}
\text{maximize}_p & \displaystyle\sum_{i=1}^{n} w_i \log(p(x_i)) \\
\text{s.t.} & p \text{ is a log-concave density} \\
\text{and} & p \text{ is } \text{MTP}_2.
\end{array}
$$

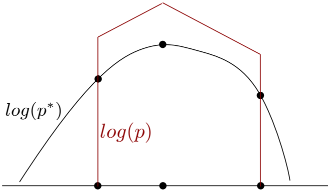


Proof that the MLE is a tent function:

- Suppose that p^* is the MLE and that $\log p^*$ is not a tent function
- Let $y_i = \log p^*(x_i), i = 1, \ldots, n$.
- Consider $p = \exp(h_{X,y})$. It gives a higher objective value than p^*.
- Thus, p^* has to be a tent function.

Proving that the Log-concave MTP$_2$ MLE is the exponential of a tent function

$$\begin{array}{lll} \text{maximize}_p & \displaystyle\sum_{i=1}^{n} w_i \log(p(x_i)) \\ \text{s.t.} & p \text{ is a log-concave density} \\ \text{and} & p \text{ is MTP}_2. \end{array}$$



Proof that the MLE is a tent function:

- Suppose that p^* is the MLE and that $\log p^*$ is not a tent function
- Let $y_i = \log p^*(x_i), i = 1, \ldots, n$.
- Consider $p = \exp(h_{X,y})$. It gives a higher objective value than p^*.
  - **Problem: is $p = \exp(h_{X,y})$ always MTP$_2$ assuming that p^* is MTP$_2$?**
- Thus, p^* has to be a tent function.

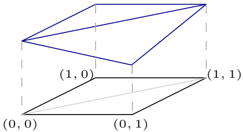
When is the MLE the exponential of a tent function?

Definition

Let $X = \{x_1, \ldots, x_n\} \subseteq \mathbb{R}^d$ be a *min-max closed* configuration. Then X is **tidy** if

The restriction of $h_{X,y}$ to X is supermodular $\iff$ The whole function $h_{X,y}$ is supermodular.

Example



If $X = \{(0,0),(0,1),(1,0),(1,1)\}$, then X is tidy because

$$y_{(0,0)} + y_{(1,1)} \geq y_{(0,1)} + y_{(1,0)} \implies h_{(X,y)} \text{ is supermodular.}$$

Example

Consider again

$$X = \{(0,0,0),(6,0,0),(6,4,0),(8,4,2),(6,4,\frac{3}{2})\}.$$

- The restriction of *any* $h_{X,y}$ to X is supermodular.
- But not all $h_{X,y}$ are supermodular! $\implies$ Not tidy.

When is the MLE the exponential of a tent function?

Theorem (R., Sturmfels, Tran, Uhler)

Let $X \subseteq \mathbb{R}^d$ be min-max closed such that conv$(X) =$ MMconv(X). Then, X is tidy if

- $X \subseteq \mathbb{R}^2$, *or*
- $X \subseteq \{0,1\}^d$.

Therefore, the MLE for configurations in $\mathbb{R}^2$ and in $\{0,1\}^d$ is always a tent function.

When is the MLE the exponential of a tent function?

Theorem (R., Sturmfels, Tran, Uhler)

Let $X \subseteq \mathbb{R}^d$ be min-max closed such that conv$(X) =$ MMconv(X). Then, X is tidy if

- $X \subseteq \mathbb{R}^2$, *or*
- $X \subseteq \{0, 1\}^d$.

Therefore, the MLE for configurations in $\mathbb{R}^2$ and in $\{0,1\}^d$ is always a tent function.

Conjecture

These are the only tidy configurations.

Optimization Problem in the Tidy Case

Theorem (R., Sturmfels, Tran, Uhler)

If $X \subseteq \mathbb{R}^d$ is a tidy configuration, then,

- *The MLE p^* is the exponential of a $p^* = \exp(h_{X,y^*})$, and*
- *The set of heights for which $\exp(h_{X,y})$ is MTP_2 is a convex polytope $\mathcal{S}$.*

Therefore, we can use, e.g. projected gradient descent or the conditional gradient method, to find the best heights y^*.

$$\begin{array}{ll} \text{maximize}_y & \displaystyle\sum_{i=1}^{n} w_i y_i - \int \exp(h_{X,y}) \\ \text{s.t.} & y \in \mathcal{S}. \end{array}$$

Optimization Problem in the Tidy Case

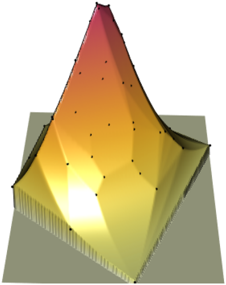
Theorem (R., Sturmfels, Tran, Uhler)

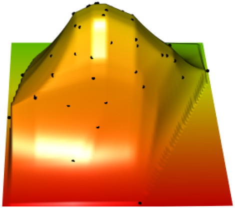
If $X \subseteq \mathbb{R}^d$ is a tidy configuration, then,

- *The MLE p^* is the exponential of a $p^* = \exp(h_{X,y^*})$, and*
- *The set of heights for which $\exp(h_{X,y})$ is MTP_2 is a convex polytope $\mathcal{S}$.*

Therefore, we can use, e.g. projected gradient descent or the conditional gradient method, to find the best heights y^*.

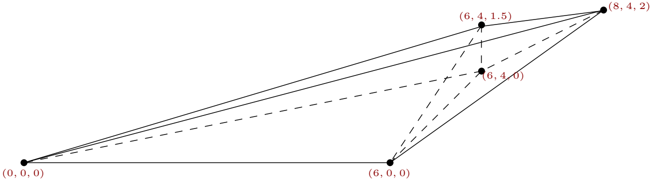
$$\begin{array}{ll} \text{maximize}_y & \sum_{i=1}^{n} w_i y_i - \int \exp(h_{X,y}) \\ \text{s.t.} & y \in \mathcal{S}. \end{array}$$





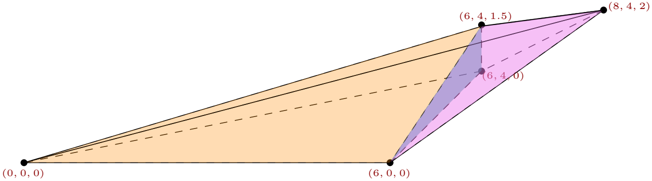
What is the shape of the MLE in the general case?

- In $\mathbb{R}^2$ and $\{0,1\}^d$ the MLE is the exponential of a tent function.
- If the log-concave MLE ϕ is a supermodular tent function, then ϕ is also the MTP_2 log-concave MLE.
- Let $X = \{(0,0,0),(6,0,0),(6,4,0),(8,4,2),(6,4,\frac{3}{2})\}$, $w = \frac{1}{28}(15,1,1,1,10)$. The log-concave MLE ϕ is not supermodular.



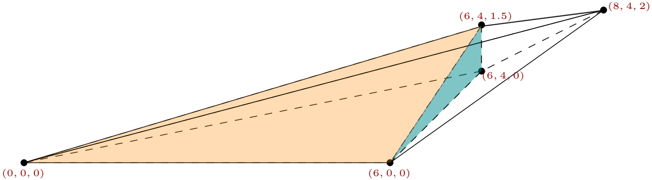
What is the shape of the MLE in the general case?

- In $\mathbb{R}^2$ and $\{0,1\}^d$ the MLE is the exponential of a tent function.
- If the log-concave MLE ϕ is a bimonotone tent function, then ϕ is also the MTP_2 log-concave MLE.
- Let $X = \{(0,0,0),(6,0,0),(6,4,0),(8,4,2),(6,4,\frac{3}{2})\}$, $w = \frac{1}{28}(15,1,1,1,10)$. The log-concave MLE ϕ is not bimonotone.



What is the shape of the MLE in the general case?

- In $\mathbb{R}^2$ and $\{0,1\}^d$ the MLE is the exponential of a tent function.
- If the log-concave MLE ϕ is a bimonotone tent function, then ϕ is also the MTP_2 log-concave MLE.
- Let $X = \{(0,0,0),(6,0,0),(6,4,0),(8,4,2),(6,4,\frac{3}{2})\}$, $w = \frac{1}{28}(15,1,1,1,10)$. The log-concave MLE ϕ is not bimonotone.

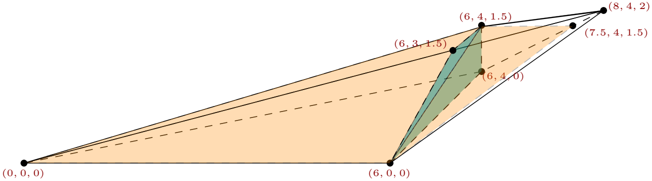


What is the shape of the MLE in the general case?

- In $\mathbb{R}^2$ and $\{0,1\}^d$ the MLE is the exponential of a tent function.

- If the log-concave MLE ϕ is a bimonotone tent function, then ϕ is also the MTP_2 log-concave MLE.

- Let $X = \{(0,0,0),(6,0,0),(6,4,0),(8,4,2),(6,4,\frac{3}{2})\}$, $w = \frac{1}{28}(15,1,1,1,10)$. The log-concave MLE ϕ is not bimonotone.



the MLE is a tent function on $X \cup \{(6,3,\frac{3}{2}),(7.5,4,\frac{3}{2})\}$ with subdivision as above.

Conjecture

Let $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$ be a point configuration, and let $w \in \mathbb{R}^n$ be the corresponding set of weights. Let $\phi : \mathbb{R}^d \to \mathbb{R}$ be the log-concave maximum likelihood estimator (which is a tent function above X), and let Δ be the subdivision it induces.

1. *If Δ is a bimonotone subdivision, then ϕ is also the MTP_2 log-concave MLE.*
2. *If Δ is not bimonotone, consider the hyperplanes spanned by each of the bimonotone codimension 1 cells of Δ, and intersect conv(X) with them. Call this new subdivision Δ'. The MTP_2 log-concave maximum likelihood estimator is a piecewise linear function whose underlying subdivision is Δ' or any subdivision refined by Δ'.*

Summary and Remaining Questions

Summary:

- We showed that the MLE under log-concavity and MTP_2 exists and is unique with probability one.
- We showed that in some cases it is the exponential of a tent function, and we can compute it using convex optimization over a finite-dimensional convex set.
- We saw which tent functions are supermodular, i.e. are candidates for the MLE.

Remaining questions and future work

- Characterize the shape of the MLE in the general case.
- Study the sample complexity of solving the problem.
- Design and analyze algorithms for finding the MLE.

Announcement

Applied Algebra Day
Saturday, Nov 17
9:30AM - 5PM
MIT, E17-304

Announcement

Applied Algebra Day
Saturday, Nov 17
9:30AM - 5PM
MIT, E17-304

Thank you!