Nonparametric Density Estimation under Total Positivity

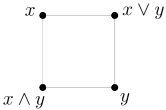
Elina Robeva
University of British Columbia

May 24, 2021

Totally Positive Distributions

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

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

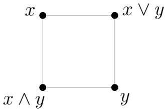


where $x \wedge y$ and $x \vee y$ are the componentwise minimum and maximum.

Totally Positive 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*:

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

Properties of MTP_2 distributions

Theorem (Fortuin Kasteleyn Ginibre, 1971)

MTP_2 implies positive association.

Definition

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

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

Theorem (Karlin and Rinot 1980; Lebowitz 1972)

If $X = (X_1, \ldots, X_d)$ is MTP_2, then

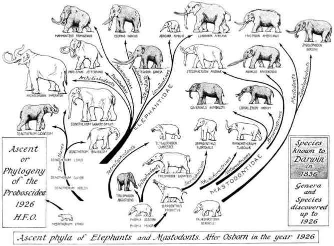
(i) *any marginal distribution is MTP_2,*

(ii) *any conditional distribution is MTP_2,*

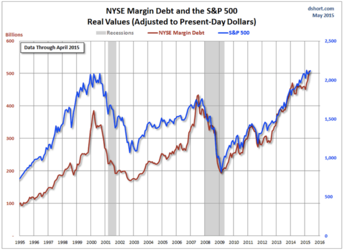
(iii) *X has the marginal independence structure*

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

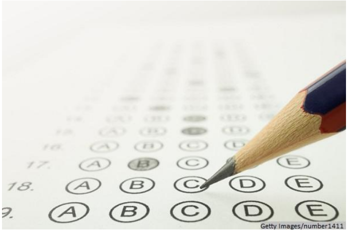
Examples of Totally Positive Distributions



[Felsenstein, 1973]



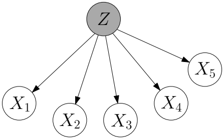
[Agrawal, Roy, and Uhler, 2019]



[Fallat, Lauritzen, Sadeghi, Uhler, Wermuth, and Zwiernik, 2016]

Examples of MTP_2 distributions

- The joint distribution of observed variables influenced by one hidden variable (in the Gaussian and binary case)



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

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

$$(\Sigma^{-1})_{ij} \leq 0, \quad \forall i \neq j.$$

Very rare: out of 100,000 uniformly sampled 5×5 correlation matrices none were MTP_2. [Fallat et al., 2016].

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

$$(\Sigma^{-1})_{ij} \leq 0, \quad \forall i \neq j.$$

Very rare: out of 100,000 uniformly sampled 5×5 correlation matrices none were MTP_2. [Fallat et al., 2016].

Data: grades of 88 students in 5 math subjects

[Mardia, Kent and Bibby, 1979, Fallat et al., 2016]

$$S = \begin{pmatrix} 305.7680 & 127.2226 & 101.5794 & 106.2727 & 117.4049 \\ 127.2226 & 172.8422 & 85.1573 & 94.6729 & 99.0120 \\ 101.5794 & 85.1573 & 112.8860 & 112.1134 & 121.8706 \\ 106.2727 & 94.6729 & 112.1134 & 220.3804 & 155.5355 \\ 117.4049 & 99.0120 & 121.8706 & 155.5355 & 297.7554 \end{pmatrix}$$

	mechanics	vectors	algebra	analysis	statistics
mechanics	305.7680	127.2226	101.5794	106.2727	117.4049
vectors	127.2226	172.8422	85.1573	94.6729	99.0120
algebra	101.5794	85.1573	112.8860	112.1134	121.8706
analysis	106.2727	94.6729	112.1134	220.3804	155.5355
statistics	117.4049	99.0120	121.8706	155.5355	297.7554

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

$$(\Sigma^{-1})_{ij} \leq 0, \quad \forall i \neq j.$$

Very rare: out of 100,000 uniformly sampled 5×5 correlation matrices none were MTP_2. [Fallat et al., 2016].

Data: grades of 88 students in 5 math subjects

[Mardia, Kent and Bibby, 1979, Fallat et al., 2016]

$$S^{-1} = 10^{-3} \cdot \begin{pmatrix} 5.2446 & -2.4351 & -2.7395 & 0.0116 & -0.1430 \\ -2.4351 & 10.4268 & -4.7078 & -0.7928 & -0.1660 \\ -2.7395 & -4.7078 & 26.9548 & -7.0486 & -4.7050 \\ 0.0116 & -0.7928 & -7.0486 & 9.8829 & -2.0184 \\ -0.1430 & -0.1660 & -4.7050 & -2.0184 & 6.4501 \end{pmatrix}$$

	mechanics	vectors	algebra	analysis	statistics
mechanics	5.2446	−2.4351	−2.7395	0.0116	−0.1430
vectors	−2.4351	10.4268	−4.7078	−0.7928	−0.1660
algebra	−2.7395	−4.7078	26.9548	−7.0486	−4.7050
analysis	0.0116	−0.7928	−7.0486	9.8829	−2.0184
statistics	−0.1430	−0.1660	−4.7050	−2.0184	6.4501

Examples of MTP_2 distributions

Data: 2016 monthly correlations of global stock markets

[*InvestmentFrontier.com*; Agrawal, Roy, and Uhler, 2019]

$$S = \begin{pmatrix} 1.000 & 0.606 & 0.731 & 0.618 & 0.613 \\ 0.606 & 1.000 & 0.550 & 0.661 & 0.598 \\ 0.731 & 0.550 & 1.000 & 0.644 & 0.569 \\ 0.618 & 0.661 & 0.644 & 1.000 & 0.615 \\ 0.613 & 0.598 & 0.569 & 0.615 & 1.000 \end{pmatrix}$$

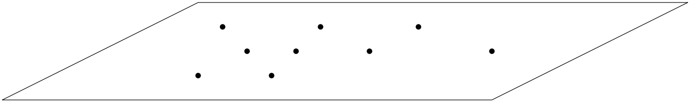
	Nasdaq	Canada	Europe	UK	Australia
Nasdaq	1.000	0.606	0.731	0.618	0.613
Canada	0.606	1.000	0.550	0.661	0.598
Europe	0.731	0.550	1.000	0.644	0.569
UK	0.618	0.661	0.644	1.000	0.615
Australia	0.613	0.598	0.569	0.615	1.000

$$S^{-1} = \begin{pmatrix} 2.629 & -0.480 & -1.249 & -0.202 & -0.490 \\ -0.480 & 2.109 & -0.039 & -0.790 & -0.459 \\ -1.249 & -0.039 & 2.491 & -0.675 & -0.213 \\ -0.202 & -0.790 & -0.675 & 2.378 & -0.482 \\ -0.490 & -0.459 & -0.213 & -0.482 & 1.992 \end{pmatrix}$$

	Nasdaq	Canada	Europe	UK	Australia
Nasdaq	2.629	−0.480	−1.249	−0.202	−0.490
Canada	−0.480	2.109	−0.039	−0.790	−0.459
Europe	−1.249	−0.039	2.491	−0.675	−0.213
UK	−0.202	−0.790	−0.675	2.378	−0.482
Australia	−0.490	−0.459	−0.213	−0.482	1.992

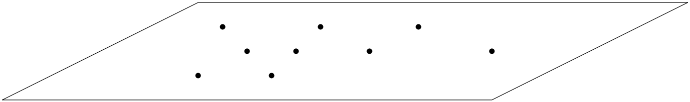
Density estimation

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?



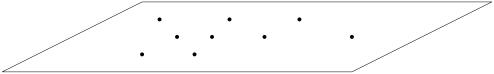
Density estimation

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?



- 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

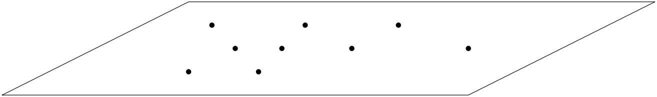
Shape-constrained density estimation



- monotonically decreasing densities: [Grenander 1956, Rao 1969]
- convex densities: [Anevski 1994, Groeneboom, Jongbloed, and Wellner 2001]
- **log-concave** densities: [Cule, Samworth, and Stewart 2008, Dümbgen et al. 2009, ...]
- generalized additive models with shape constraints: [Chen and Samworth 2016]

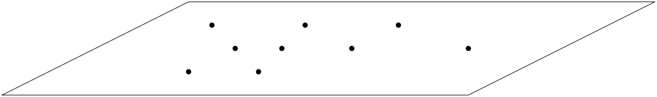
- **totally positive (and log-concave) densities**

Maximum Likelihood Estimation under MTP_2



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 under MTP_2

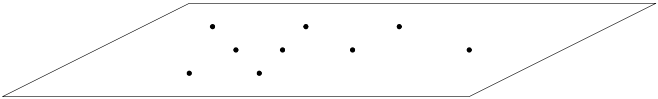


Given i.i.d. samples $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$ with weights $w = (w_1, \ldots, w_n)$ (where $w_1, \ldots, w_n \geq 0$, $\sum w_i = 1$) from a distribution p on $\mathbb{R}^d$, can we estimate p?

We would like to

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

Maximum Likelihood Estimation under MTP_2



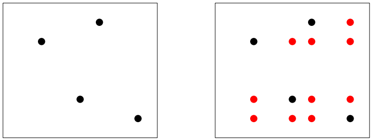
Given i.i.d. samples $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$ with weights $w = (w_1, \ldots, w_n)$ (where $w_1, \ldots, w_n \geq 0$, $\sum w_i = 1$) from a distribution p on $\mathbb{R}^d$, can we estimate p?

We would like to

$$\text{maximize}_p \quad \sum_{i=1}^{n} w_i \log(p(x_i))$$

$$\text{s.t.} \qquad p \text{ is an } \mathrm{MTP}_2 \text{ density.}$$

Problem: likelihood is unbounded, MTP_2 constraint is too weak for MLE to exist.



Maximum Likelihood Estimation under MTP_2

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

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

Maximum Likelihood Estimation under MTP_2

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

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

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.

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

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

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

Maximum Likelihood Estimation under Log-Concavity

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

Theorem (Cule, Samworth and Stewart 2008)

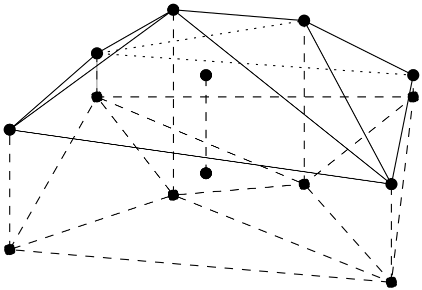
- *With probability 1, a log-concave maximum likelihood estimator $\hat{p}$ exists and is unique as long as there are at least $d+1$ samples.*

Maximum Likelihood Estimation under Log-Concavity

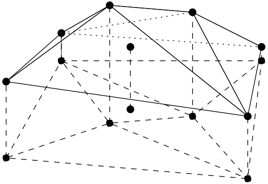
$$\begin{aligned} \text{maximize}_p \quad & \sum_{i=1}^{n} w_i \log(p(x_i)) \\ \text{s.t.} \quad & p \text{ is a density} \\ \text{and} \quad & p \text{ is log-concave.} \end{aligned}$$

Theorem (Cule, Samworth and Stewart 2008)

- *With probability 1, a log-concave maximum likelihood estimator $\hat{p}$ exists and is unique as long as there are at least $d+1$ samples.*
- *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

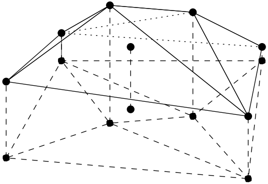


Given points $X = \{x_1, \dots, x_n\}$ and heights $y = (y_1, \dots, y_n) \in \mathbb{R}^n$, the *tent function*

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

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*

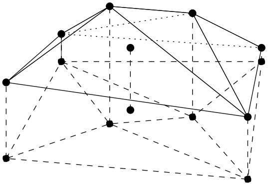
$$h_{X,y} : \mathbb{R}^d \to \mathbb{R}$$

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.

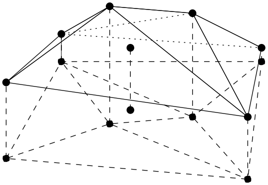
$$\begin{aligned} \text{maximize}_p \quad & \sum_{i=1}^{n} w_i \log(p(x_i)) \\ \text{s.t.} \quad & p \text{ is a density} \\ \text{and} \quad & p \text{ is log-concave.} \end{aligned}$$

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

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.

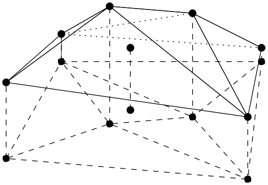
$$\begin{aligned} \text{maximize}_p \quad & \sum_{i=1}^n w_i \log(p(x_i)) \\ \text{s.t.} \quad & p \text{ is a density} \\ \text{and} \quad & p \text{ is log-concave.} \end{aligned}$$

INFINITE DIMENSIONAL

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

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.

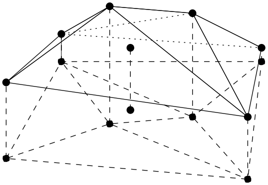
$$\begin{aligned} \text{maximize}_p \quad & \sum_{i=1}^{n} w_i \log(p(x_i)) \\ \text{s.t.} \quad & p \text{ is a density} \\ \text{and} \quad & p \text{ is log-concave.} \end{aligned}$$

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

INFINITE DIMENSIONAL

FINITE DIMENSIONAL

Optimizing over Tent Functions



Given points $X = \{x_1, \dots, x_n\}$ and heights $y = (y_1, \dots, y_n) \in \mathbb{R}^n$, 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.

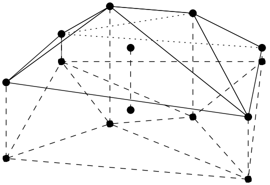
$$\begin{aligned} \text{maximize}_p \quad & \sum_{i=1}^{n} w_i \log(p(x_i)) \\ \text{s.t.} \quad & p \text{ is a density} \\ \text{and} \quad & p \text{ is log-concave.} \end{aligned}$$

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.

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

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

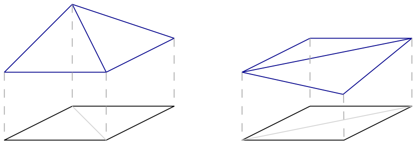
INFINITE DIMENSIONAL

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?
5. Sample complexity?



Existence and Uniqueness of the MLE

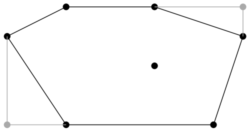
Theorem (R., Sturmfels, Tran, Uhler, 2018)

- *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.*
- *Furthermore, the MLE is a consistent estimator.*

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

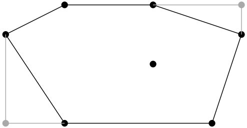
The Support of the MLE

Support of MLE = "Min-max convex hull" of X.



The Support of the MLE

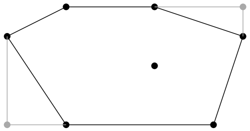
Support of MLE = "Min-max convex hull" of X.



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

The Support of the MLE

Support of MLE = "Min-max convex hull" of X.

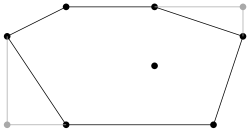


- *MMconv*(X) = smallest *min-max closed & convex* set containing X.
- *MM*(X) = smallest *min-max closed* set S containing X, i.e. $x, y \in S \Rightarrow x \wedge y, x \vee y \in S$

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

The Support of the MLE

Support of MLE = "Min-max convex hull" of X.

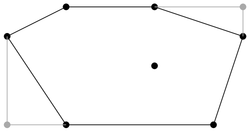


- $MMconv(X)$ = smallest *min-max closed & convex* set containing X.
- $MM(X)$ = smallest *min-max closed* set S containing X, i.e. $x, y \in S \Rightarrow x \wedge y, x \vee y \in S$

Is it always true that $\mathrm{MMconv}(X) = \mathrm{conv}(\mathrm{MM}(X))$? No, not always!

The Support of the MLE

Support of MLE = "Min-max convex hull" of X.



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

Is it always true that $\text{MMconv}(X) = \text{conv}(\text{MM}(X))$? No, not always!

Lemma

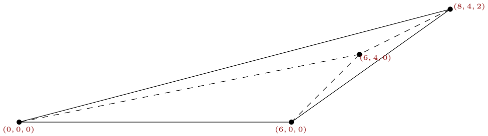
Let $X = \{x_1, \ldots, x_n\}$. If $X \subseteq \mathbb{R}^2$ or $X \subseteq \prod_{i=1}^{d} \{a_i, b_i\}$, then,

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

$\Longrightarrow$ we can compute $\text{MMconv}(X)$ for $X \subseteq \mathbb{R}^2$ and $X \subseteq \prod_{i=1}^{d} \{a_i, b_i\}$.

The Min-Max Convex Hull

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



It turns out that

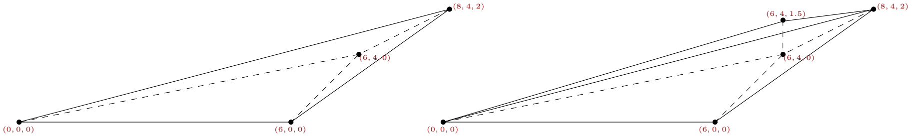
$$\mathrm{MM}(X) = X.$$

But

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

The Min-Max Convex Hull

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

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

This is because:

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

Therefore,

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

Computing $MMconv(X)$

Theorem (The 2-D Projections Theorem)

For any finite subset $X \subseteq \mathbb{R}^d$,

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

$$\pi_{ij} : \mathbb{R}^d \to \mathbb{R}^2, \quad x \mapsto (x_i, x_j).$$

Computing $MMconv(X)$

Theorem (The 2-D Projections Theorem)

For any finite subset $X \subseteq \mathbb{R}^d$,

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

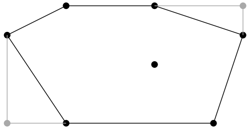
$$\pi_{ij} : \mathbb{R}^d \to \mathbb{R}^2, \quad x \mapsto (x_i, x_j).$$

Corollary (Queyranne and Tardella, 2006)

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

A linear inequality 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?
5. Sample complexity?

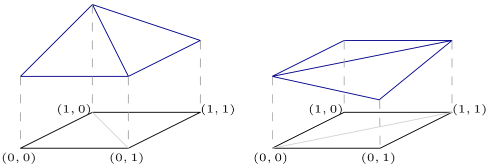
Supermodular Tent Functions

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, 2018)

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***.*



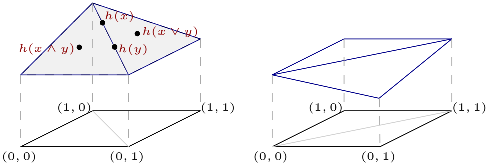
Supermodular Tent Functions

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

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

Theorem (R., Sturmfels, Tran, Uhler, 2018)

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

To find the best $h_{X,y}$, need to optimize over heights y that induce bimonotone subdivisions.

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

Maximum Likelihood Estimation under Log-concavity and MTP_2

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?
5. Sample complexity?

When is the MLE the exponential of a tent function?

Theorem (R., Sturmfels, Tran, Uhler 2018)

If $X \subseteq \mathbb{R}^2$ or $X \subseteq \prod_{i=1}^{d} \{a_i, b_i\}$, then the MLE is the exponential of a tent function

$$p^* = \exp(h_{\tilde{X},y}),$$

where $\tilde{X} = MM(X)$.

Finding the MLE

Theorem (R., Sturmfels, Tran, Uhler 2018)

If $X \subseteq \mathbb{R}^2$ or $X \subseteq \prod_i \{a_i, b_i\}$, then the set of heights for which $\exp(h_{\tilde{X},y})$ is MTP_2 is a convex polytope $\mathcal{S}$.
Therefore, we can use, e.g. the conditional gradient method, to find y^*.

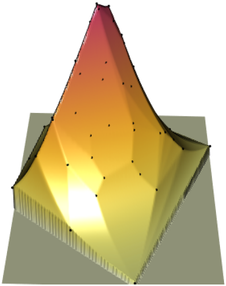
$$\begin{array}{ll} \text{maximize}_y & \displaystyle\sum_{i=1}^{n} w_i y_i - \int \exp(h_{\tilde{X},y}) \\ \text{s.t.} & y \in \mathcal{S}. \end{array}$$

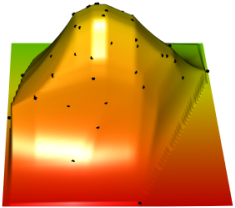
Finding the MLE

Theorem (R., Sturmfels, Tran, Uhler 2018)

If $X \subseteq \mathbb{R}^2$ or $X \subseteq \prod_i \{a_i, b_i\}$, then the set of heights for which $\exp(h_{\tilde{X},y})$ is MTP_2 is a convex polytope $\mathcal{S}$.
Therefore, we can use, e.g. the conditional gradient method, to find y^*.

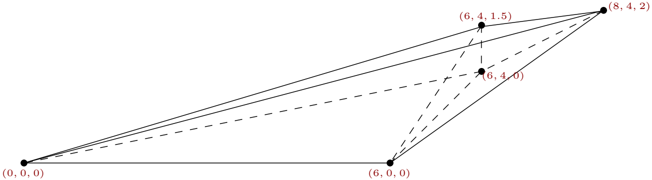
$$\begin{array}{ll} \textit{maximize}_y & \sum_{i=1}^{n} w_i y_i - \int \exp(h_{\tilde{X},y}) \\ \textit{s.t.} & y \in \mathcal{S}. \end{array}$$





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

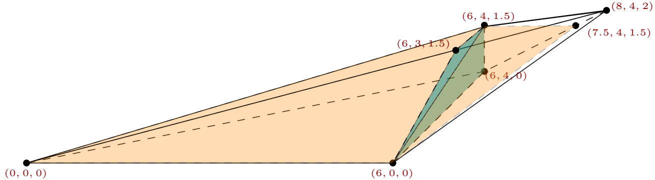
Let $X = \{(0,0,0), (6,0,0), (6,4,0), (8,4,2), (6,4,\frac{3}{2})\}$, $w = \frac{1}{28}(15,1,1,1,10)$. The log-concave MLE is not MTP_2.



The MTP_2 and log-concave MLE is exp(tent function) on $X \cup \{(6,3,\frac{3}{2}), (7.5,4,\frac{3}{2})\}$ with subdivision as above.

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

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.



The MTP_2 and log-concave MLE is exp(tent function) on $X \cup \{(6,3,\frac{3}{2}),(7.5,4,\frac{3}{2})\}$ with subdivision as above.

Sample complexity for log-concave density estimation

- $\mathcal{F}_d$ = set of log-concave and MTP_2 densities in $\mathbb{R}^d$
- $h^2(f,g) = \int(\sqrt{f} - \sqrt{g})^2$ Hellinger distance
- $\hat{p}_n$ = log-concave MLE from n samples

Theorem (Kim and Samworth, 2014)

(a). *Upper bound for $d = 2, 3$:*

$$\sup_{p_0 \in \mathcal{F}_2} \mathbb{E}_{p_0}[h^2(p_0, \hat{p}_n)] = \mathcal{O}(n^{-\frac{2}{d+1}} \log n).$$

(b). *Lower bound for $d \geq 2$:*

$$\inf_{\tilde{p}_n} \sup_{p_0 \in \mathcal{F}_2} \mathbb{E}_{p_0}[h^2(p_0, \tilde{p}_n)] \geq c_2 n^{-\frac{2}{d+1}}.$$

MLE is minimax optimal up to log factors
for $d = 2, 3$.

Sample complexity for log-concave density estimation

- $\mathcal{F}_d$ = set of log-concave and MTP_2 densities in $\mathbb{R}^d$
- $h^2(f,g) = \int(\sqrt{f}-\sqrt{g})^2$ Hellinger distance
- $\hat{p}_n$ = log-concave MLE from n samples

Theorem (Kim and Samworth, 2014)

(a). *Upper bound for $d = 2, 3$:*

$$\sup_{p_0 \in \mathcal{F}_2} \mathbb{E}_{p_0}[h^2(p_0, \hat{p}_n)] = \mathcal{O}(n^{-\frac{2}{d+1}} \log n).$$

(b). *Lower bound for $d \geq 2$:*

$$\inf_{\tilde{p}_n} \sup_{p_0 \in \mathcal{F}_2} \mathbb{E}_{p_0}[h^2(p_0, \tilde{p}_n)] \geq c_2 n^{-\frac{2}{d+1}}.$$

MLE is minimax optimal up to log factors for $d = 2, 3$.

Theorem (Carpenter et al., 2018)

When $d \geq 4$:

(a). *Upper bound:*

$$\sup_{p_0 \in \mathcal{F}_d} \mathbb{E}_{p_0}[h^2(p_0, \hat{p}_n)] = \tilde{\mathcal{O}}(n^{-\frac{2}{d+3}})$$

(b). *Lower bound:*

$$\inf_{\tilde{p}_n} \sup_{p_0 \in \mathcal{F}_d} \mathbb{E}_{p_0}[h^2(p_0, \tilde{p}_n)] \geq c_2 n^{-\frac{2}{d+1}}$$

Sample complexity for log-concave and MTP_2

- $\mathcal{F}_d$ = set of log-concave and MTP_2 densities in $\mathbb{R}^d$
- $h^2(f,g) = \int(\sqrt{f} - \sqrt{g})^2$ Hellinger distance
- $\hat{p}_n$ = MTP_2 and log-concave MLE from n samples

Theorem (R., Uhler, Wang, 2018)

When $d = 2$:

(a). *Upper bound:*

$$\sup_{p_0 \in \mathcal{F}_2} \mathbb{E}_{p_0}[h^2(p_0, \hat{p}_n)] = \mathcal{O}(n^{-\frac{2}{3}} \log n).$$

(b). *Lower bound:*

$$\inf_{\tilde{p}_n} \sup_{p_0 \in \mathcal{F}_2} \mathbb{E}_{p_0}[h^2(p_0, \tilde{p}_n)] \geq c_2 n^{-\frac{2}{3}}.$$

MLE is minimax optimal up to log factors.

Sample complexity for log-concave and MTP$_2$

- $\mathcal{F}_d$ = set of log-concave and MTP$_2$ densities in $\mathbb{R}^d$
- $h^2(f,g) = \int(\sqrt{f} - \sqrt{g})^2$ Hellinger distance
- $\hat{p}_n$ = MTP$_2$ and log-concave MLE from n samples

Theorem (R., Uhler, Wang, 2018)

When $d = 2$:

(a). *Upper bound:*

$$\sup_{p_0 \in \mathcal{F}_2} \mathbb{E}_{p_0}[h^2(p_0, \hat{p}_n)] = \mathcal{O}(n^{-\frac{2}{3}} \log n).$$

(b). *Lower bound:*

$$\inf_{\tilde{p}_n} \sup_{p_0 \in \mathcal{F}_2} \mathbb{E}_{p_0}[h^2(p_0, \tilde{p}_n)] \geq c_2 n^{-\frac{2}{3}}.$$

MLE is minimax optimal up to log factors.

Problem

When $d \geq 3$:

(a). *Upper bound?*

$$\sup_{p_0 \in \mathcal{F}_d} \mathbb{E}_{p_0}[h^2(p_0, \hat{p}_n)] = \mathcal{O}(n^{-\frac{2}{3}} \log n)$$

(b). *Lower bound:*

$$\inf_{\tilde{p}_n} \sup_{p_0 \in \mathcal{F}_d} \mathbb{E}_{p_0}[h^2(p_0, \tilde{p}_n)] \geq c_2 poly(d) n^{-\frac{2}{3}}$$

No curse of dimensionality?

Conclusion

MLE

- Is the MLE always exp(tent function)?
- Rates for $d \geq 3$?
- Faster algorithms for finding the MLE?

Conclusion

MLE

- Is the MLE always exp(tent function)?
- Rates for $d \geq 3$?
- Faster algorithms for finding the MLE?

Kernel Density Estimation

- Min-max closure of sample, $\mathrm{MM}(X)$, convolved with standard Gaussian is MTP_2.
  [with Ali Zartash, 2019]
- Consistency and rates?

Conclusion

MLE

- Is the MLE always exp(tent function)?
- Rates for $d \geq 3$?
- Faster algorithms for finding the MLE?

Kernel Density Estimation

- Min-max closure of sample, $\mathrm{MM}(X)$, convolved with standard Gaussian is MTP_2.
  [with Ali Zartash, 2019]
- Consistency and rates?

- Test statistics for MTP_2 distributions?
- Multivariate total positivity of higher orders (i.e. MTP_k)?

Thank you!

References

M. Cule, R. Samworth, M. Stuart. *Maximum likelihood estimation of a multidimensional log-concave density.* Statistical Methodology Series B 72:5 (2010), 545-607

J.-C. Hütter, C. Mao, P. Rigollet, and E. Robeva. *Optimal Rates for Estimation of Two-dimensional Totally Positive Distributions.*

E. Robeva, B. Sturmfels, and C. Uhler. *Geometry of Log-concave Density Estimation.* arXiv:1704.01910

E. Robeva, B. Sturmfels, N. Tran, and C. Uhler. *Maximum Likelihood Estimation for Totally Positive Log-Concave Densities.* arXiv:1806.10120

E. Robeva and M. Sun. *Bimonotone Subdivisions of Point Configurations in the Plane*. In preparation

E. Robeva, C. Uhler, and Y. Wang. *Rates of Estimation of Log-Concave and Totally Positive Densities.* In preparation

A. Zartash and **E. Robeva**. *Kernel density estimation for totally positive random vectors.* arXiv:1910:02345