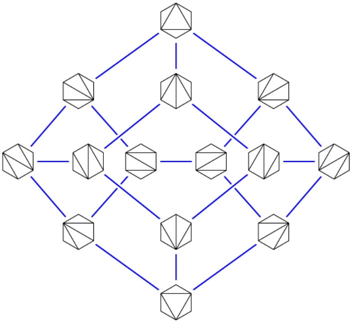
# Geometry of Log-Concave Density Estimation

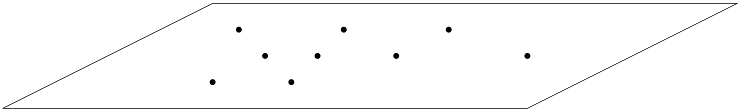
**Bernd Sturmfels**
*MPI Leipzig and UC Berkeley*



*joint paper with Elina Robeva and Caroline Uhler*

# Weighted Density Estimation

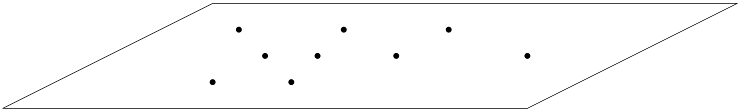
*Given data*: a point configuration $X = \{x_1, \ldots, x_n\} \in \mathbb{R}^d$ with weights $w = (w_1, \ldots, w_n)$, where $w_1, \ldots, w_n \geq 0$, $\sum w_i = 1$.



These are i.i.d. samples from an unknown probablity distribution $p$ on $\mathbb{R}^d$. How to estimate $p$?

# Weighted Density Estimation

*Given data*: a point configuration $X = \{x_1, \ldots, x_n\} \in \mathbb{R}^d$ with weights $w = (w_1, \ldots, w_n)$, where $w_1, \ldots, w_n \geq 0$, $\sum w_i = 1$.

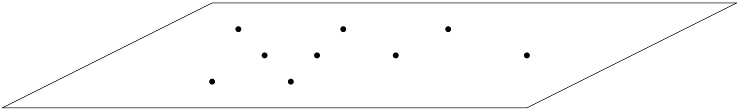


These are i.i.d. samples from an unknown probablity distribution $p$ on $\mathbb{R}^d$. How to estimate $p$? Use maximum likelihood estimation, i.e. maximize the logarithm of the probability of observing the data:

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

# Weighted Density Estimation

*Given data*: a point configuration $X = \{x_1, \dots, x_n\} \in \mathbb{R}^d$ with weights $w = (w_1, \dots, w_n)$, where $w_1, \dots, w_n \geq 0$, $\sum w_i = 1$.



These are i.i.d. samples from an unknown probablity distribution $p$ on $\mathbb{R}^d$. How to estimate $p$? Use maximum likelihood estimation, i.e. maximize the logarithm of the probability of observing the data:

$$\text{maximize}_p \quad \sum_{i=1}^{n} w_i \log(p(x_i))$$
$$\text{s.t.} \quad p \text{ is a density}$$

**Q**: *Does this optimization problem make sense?*

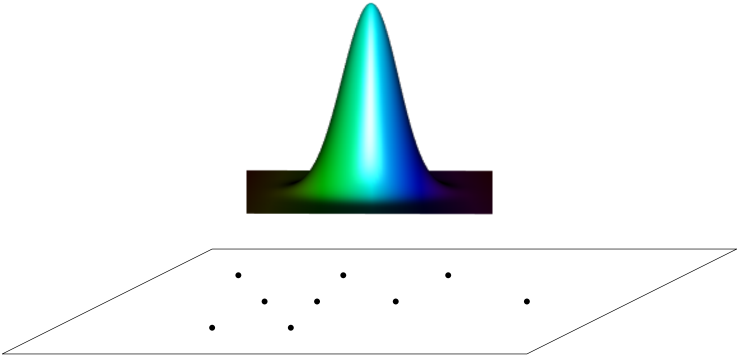
**A**: *No, because we can choose* $p$ *arbitrarily close to* $\sum_{i=1}^n w_i \delta_{x_i}$.

# Estimating Model Parameters

Assume that $p$ is a $d$-dimensional Gaussian distribution:

$$p(x) \;=\; \frac{1}{\big(2\pi\,\mathrm{det}(\Sigma)\big)^{d}}\cdot \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right).$$

Mean $\mu \in \mathbb{R}^d$ and covariance matrix $\Sigma \in \mathrm{Sym}_2(\mathbb{R}^d)$ are unknown.



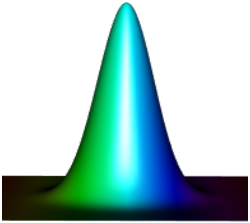
MLE is easy: $\hat{\mu} = \sum_{i=1}^{n} w_i x_i$ and $\hat{\Sigma} = \sum_{i=1}^{n} w_i(x-\hat{\mu})(x-\hat{\mu})^T$.

# Non-Parametric Statistics

We do **not** assume a model with finitely many parameters.
The fewer assumptions the better. *This leads to*

Shape-constrained maximum likelihood estimation

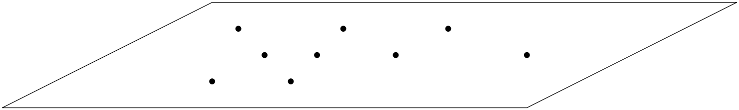
- monotonically decreasing densities: Grenander 1956, Rao 1969
- convex densities: Anevski 1994, Groeneboom, Jongbloed, and Wellner 2001
- log-concave densities: Cule, Samworth, and Stewart 2008
- generalized additive models with shape constraints: Chen and Samworth 2016



Gaussian densities are log-concave:
$-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)$ is a concave function

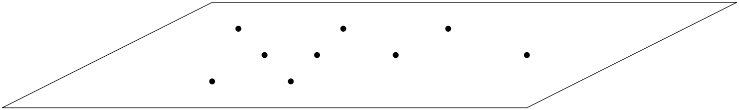
# Our Optimization Problem

Maximize the log-likelihood of the given sample $(X, w)$ over all integrable functions $p : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ such that $\log(p)$ is concave and $\int_{\mathbb{R}^d} p(x)dx = 1$.



# Our Optimization Problem

Maximize the log-likelihood of the given sample $(X, w)$
over all integrable functions $p : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ such that
$\log(p)$ is concave and $\int_{\mathbb{R}^d} p(x)dx = 1$.



This problem was solved for uniform weights $w = \frac{1}{n}(1, 1, \ldots, 1)$ by

M. Cule, R. Samworth and M. Stewart:
*Maximum likelihood estimation of a multi-dimensional log-concave density*,
J. R. Stat. Soc. Ser. B Stat. Methodol. **72** (2010) 545–607.

M. Cule, R.B. Gramacy and R. Samworth: *LogConcDEAD: an R package for maximum likelihood estimation of a multivariate log-concave density*,
J. Statist. Software **29** (2009) Issue 2.

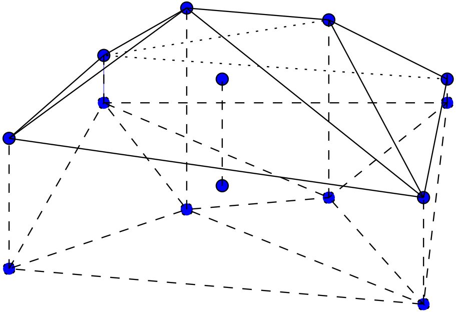
We extend to arbitrary $w$ and develop the link to geometric combinatorics:

J. De Loera, J. Rambau and F. Santos: *Triangulations. Structures for Algorithms and Applications*, Algorithms and Computation in Mathematics **25**, Springer Berlin, 2010.

# Maximum Likelihood Estimation

## Theorem

*A log-concave maximum likelihood estimate $\hat{p}$ exists for all $(X, w)$. It is unique with probability $1$. The concave function $log(\hat{p})$ is a tent function supported on the convex polytope $P = conv(X)$.*

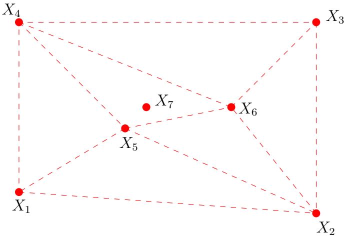


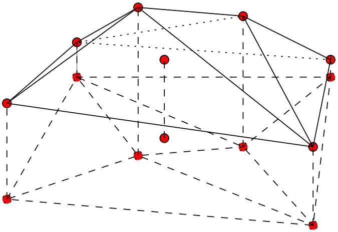
Tent function means: *piecewise linear and concave, supported on a regular polyhedral subdivision of the configuration $X$ of $n$ points in $\mathbb{R}^d$.*

# Tent Functions

Given points $X = \{x_1, \ldots, x_n\}$ and heights $y_1, \ldots, y_n$ at these 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 $i$. Thus,

$$\hat{p} = \exp(h_{X,y}) \qquad \text{for some height vector } y \in \mathbb{R}^n.$$



Two equivalent Optimization Problems:

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

INFINITE DIMENSIONAL

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

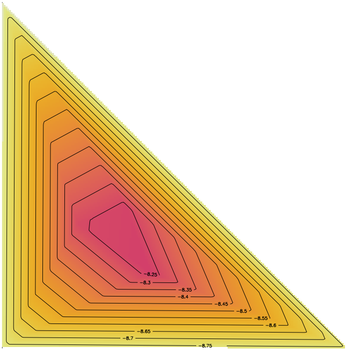
FINITE DIMENSIONAL

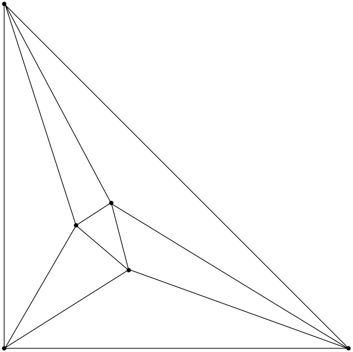
# LogConcDEAD

## Example

Let $d = 2$, $n = 6$, $w = \frac{1}{6}(1, 1, 1, 1, 1, 1)$, and fix the point configuration

$$X \;=\; \big(\,(0,0)\,,\,(100,0)\,,\,(0,100)\,,\,(22,37)\,,\,(43,22)\,,\,(36,41)\,\big).$$



The optimal log-concave density $\hat{p}$ for the six data points in $X$ with unit weights.

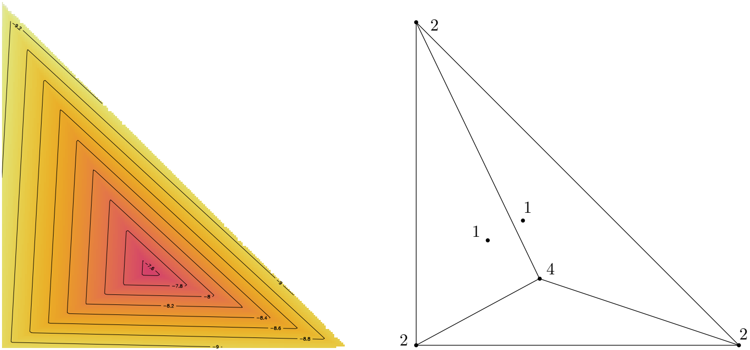
Computed with the **R package** of Cule, Gramacy and Samworth.

# LogConcDEAD

## Example

Let $d = 2$, $n = 6$, $w = \frac{1}{12}(2, 2, 2, 1, 4, 1)$, and fix the point configuration

$$X \;=\; \big(\,(0,0)\,,\,(100,0)\,,\,(0,100)\,,\,(22,37)\,,\,(43,22)\,,\,(36,41)\,\big).$$

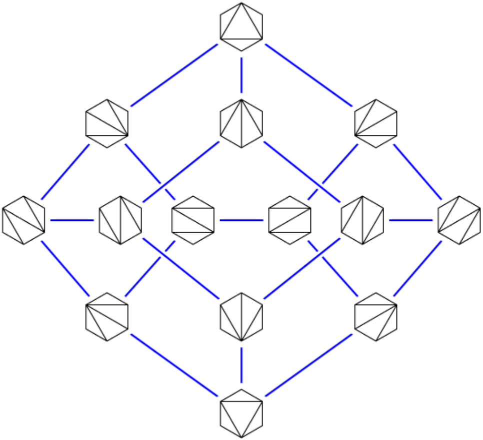



The optimal log-concave density $\hat{p}$ for the six data points in $X$ with non-unit weights.

Computed with the **R package** of Cule, Gramacy and Samworth.

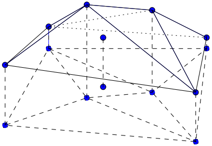
# Secondary Polytope

The *secondary polytope* $\Sigma(X)$ has dimension $n-d-1$ but lives in $\mathbb{R}^n$. Its faces are in bijection with the regular subdivisions of $X$. The vertices of $\Sigma(X)$ correspond to regular triangulations of $X$.



I.M. Gel'fand, M.M. Kapranov and A.V. Zelevinsky: *Discriminants, Resultants and Multidimensional Determinants*, Birkhäuser, Boston, 1994.

# Samworth Body



The **support function** of the secondary polytope $\Sigma(X)$ is the p.l. function that measures the volume under the tent:

$$\mathbb{R}^n \to \mathbb{R}, \quad y \mapsto \int_P h_{X,y}(t)dt.$$

The convex polyhedron dual to the secondary is unbounded:

$$\Sigma(X)^* \quad = \quad \{\, y \in \mathbb{R}^n \, : \, \int_P h_{X,y}(t)dt \, \leq \, 1 \,\}.$$

The *Samworth body* is the following continuous analogue:

$$\mathcal{S}(X) \quad = \quad \{\, y \in \mathbb{R}^n \, : \, \int_P \exp(h_{X,y}(t))dt \, \leq \, 1 \,\}.$$

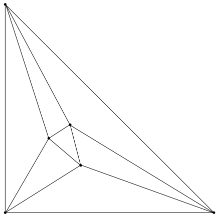
## Proposition

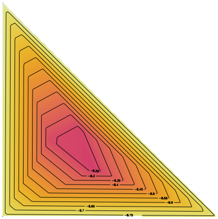
*The Samworth body $\mathcal{S}(X)$ is a full-dimensional closed convex set in $\mathbb{R}^n$.*

# Log-Concave Density Estimation

.... is Linear Programming over the Samworth body:

Maximize $w \cdot y$ subject to $y \in \mathcal{S}(X)$.





## Proposition

*This is equivalent to the unconstrained optimization problem*

$$\text{Maximize} \quad w \cdot y - \int_P \exp(h_{X,y}(t))dt \quad \textit{over} \quad \textit{all} \quad y \in \mathbb{R}^n$$

**Interpretation**: the optimal value function of our convex optimization problem is the *Legendre-Fenchel transform* of the convex function $y \mapsto \int_P \exp(h_{X,y}(t))dt$.

# Barvinok meets Samworth

## Lemma (Barvinok 1993)

*Fix linear function $\ell : \mathbb{R}^d \to \mathbb{R}$ and a d-simplex $\sigma$. Then*

$$\int_\sigma \exp(\ell(t))dt \;=\; vol(\sigma) \sum_{i=0}^{d} \exp(y_i) \prod_{j \neq i} (y_i - y_j)^{-1},$$

*where $y_0, y_1, \ldots, y_d$ are the values of $\ell$ at the vertices of $\sigma$.*

## Theorem (Cule, Samworth, Stewart 2008)

*Let $y \in \mathbb{R}^n$ such that $h_{X,y}$ induces a triangulation $\Delta$ of $X$. Then*

$$\int_C \exp(h_{X,y}(t))dt \;=\; \sum_{\sigma \in \Delta} \sum_{i \in \sigma} \frac{vol(\sigma)\exp(y_i)}{\prod_{j \in \sigma \setminus i} (y_i - y_j)}.$$

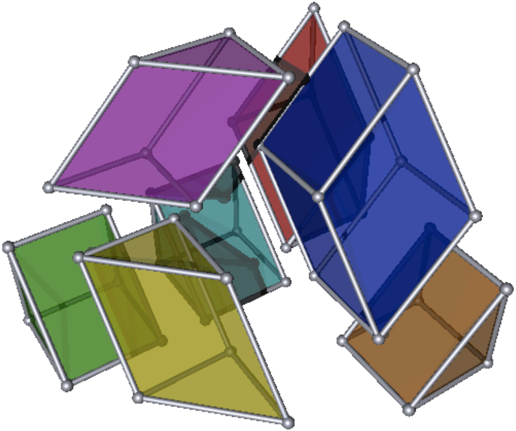
## Corollary

*On the secondary cone of a fixed triangulation $\Delta$, the Samworth body $\mathcal{S}(X)$ consists of all $y$ such that the right hand side is $\leq 1$.*

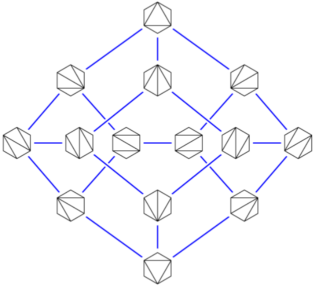
# Every Regular Subdivision Arises

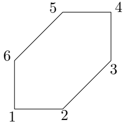
**Theorem**

*For every regular subdivision $\Delta$ of $X$, there exists an open subset $\mathcal{U}_\Delta \subset \mathbb{R}^n$ such that, for every $w \in \mathcal{U}_\Delta$, the optimal solution $\hat{p}$ to the optimization problem for $(X, w)$ gives rise to the subdivision $\Delta$.*



# Six Points in the Plane






Fix the configuration
$X = \{(0,0),(1,0),(2,1),(2,2),(1,2),(0,1)\}$

We sampled 100,000 vectors uniformly from the simplex $\{w \in \mathbb{R}^6_{\geq 0} : \sum_{i=1}^6 = 1\}$.
For each $w$, we computed the optimal $y$ and the subdivision it induces:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ∅ | 35 | 46 | 24 | 15 | 13 | 26 | 25 | 14 | 36 |
| 30.5 | 5.95 | 5.85 | 5.84 | 5.83 | 5.75 | 5.70 | 3.91 | 3.90 | 3.87 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 13 15 | 26 46 | 15 35 | 13 35 | 24 26 | 24 46 | 13 14 | 35 36 | 14 24 | 26 36 | 14 46 | 25 35 | 15 25 |
| 1.23 | 1.21 | 1.21 | 1.20 | 1.16 | 1.14 | 0.96 | 0.92 | 0.92 | 0.92 | 0.92 | 0.90 | 0.90 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 25 26 | 14 15 | 36 46 | 24 25 | 13 36 | 13 46 | 26 35 | 15 24 | 13 14 15 | 13 15 35 | 14 24 46 | 24 26 46 |
| 0.89 | 0.89 | 0.87 | 0.87 | 0.84 | 0.82 | 0.77 | 0.70 | 0.25 | 0.24 | 0.23 | 0.22 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 15 25 35 | 26 36 46 | 13 35 36 | 24 25 26 | 13 36 46 | 25 26 35 | 15 24 25 | 14 15 24 | 13 14 46 | 26 35 36 |
| 0.22 | 0.21 | 0.20 | 0.18 | 0.18 | 0.16 | 0.15 | 0.15 | 0.15 | 0.14 |

# Every Tent Function Arises

### Lemma

*Let $\Delta$ be a regular triangulation, given by $h_{X,y^*}$ for some $y^* \in \partial\mathcal{S}(X)$. There exist weights $w \in \mathbb{R}^n_{\geq 0}$ that induce $y^*$.*

**Proof**: The vector $y^*$ is the global maximizer of the function

$$\sum_{i=1}^{n} w_i y_i \;-\int exp(h_{X,y}(t))dt.$$

By taking the partial derivative with respect to $y_i$, we find

$$w_i \;=\; \frac{\partial}{\partial y_i}\int exp(h_{X,y^*}(t))dt$$

$$=\sum_{\substack{\sigma\in\Delta:\\ i\in\sigma}} \mathrm{vol}(\sigma)exp(y_i^*)H(y_j^* - y_i^*,\, j\in\sigma\backslash i),$$

where $H(u_1,\ldots,u_d)$ is a certain explicit function of $d$ arguments.

# A Symmetric Function

## Proposition

*The following expressions define the same function $H : \mathbb{R}^d \to \mathbb{R}$:*

- $$H = (-1)^d \frac{1 + u_1^{-1} + \cdots + u_d^{-1}}{u_1 u_2 \cdots u_d} + \sum_{j=1}^{d} \frac{e^{u_j}}{u_j^2 \prod_{k \neq j}(u_j - u_k)}$$
- $$H = \sum_{r=0}^{\infty} \frac{h_r(u_1, \ldots, u_d)}{(r+d+1)!}$$
- $$H = \int_{\Sigma_d} \left(1 - \sum_{i=1}^{d} t_i\right) \exp\left(\sum_{i=1}^{d} u_i t_i\right) dt_1 \ldots dt_d .$$

*This function is positive, increasing in each argument, and convex.*

Here $h_r$ is the homogeneous symmetric function, and $\Sigma_d$ is the standard simplex.

# Every Tent Function Arises

We characterize the normal cones of the Samworth body:

**Theorem**

*Fix a vector $y \in \partial\mathcal{S}(X)$, let $\Delta$ be the regular subdivision of $X$ that is induced by $h_{X,y}$ and $\Delta_1, \ldots, \Delta_m$ all regular triangulations of $X$ which refine $\Delta$. Write $w^{\Delta_1}, \ldots, w^{\Delta_m}$ for their weight vectors in $\mathbb{R}^n$*

*with $i$-th coordinates seen two slides ago.*

*A vector of weights $w \in \mathbb{R}^n_{>0}$ induces the heights $y$ if and only if*

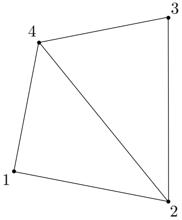
$$w \in Cone\,(w^{\Delta_1}, \ldots, w^{\Delta_m}).$$

**Corollary**

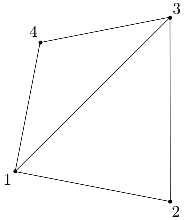
*Fix $y^* = (c, c, \ldots, c)$, where $c = -\log(\mathrm{vol}(P))$, so that $\exp(h_{X,y})$ is a probability density. Then $w^{\Delta_i}$ is precisely the vertex of the secondary polytope $\Sigma(X)$ given by the regular triangulation $\Delta_i$.*

# Four Points in the Plane

Let $x_1, x_2, x_3, x_4 \in \mathbb{R}^2$ be in convex position. Then $X$ has two triangulations:




$\Delta_1 = \{124, 234\}$



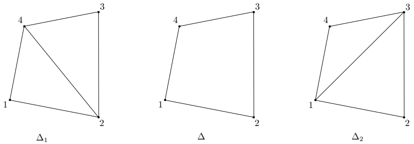

$\Delta_2 = \{123, 134\}$

Pick $y \in \mathbb{R}^4$. If $h_{X,y}$ induces $\Delta_1$, then the weight vector $w^{\Delta_1}$ has coordinates

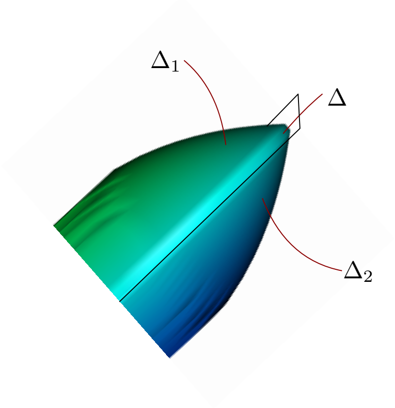
$$\begin{aligned}
w_1^{\Delta_1} &= v_{124} e^{y_1} H(y_2 - y_1, y_4 - y_1) \\
w_2^{\Delta_1} &= v_{124} e^{y_2} H(y_1 - y_2, y_4 - y_2) + v_{234} e^{y_2} H(y_3 - y_2, y_4 - y_2) \\
w_3^{\Delta_1} &= v_{234} e^{y_3} H(y_2 - y_3, y_4 - y_3) \\
w_4^{\Delta_1} &= v_{124} e^{y_4} H(y_1 - y_4, y_2 - y_4) + v_{234} e^{y_4} H(y_2 - y_4, y_3 - y_4).
\end{aligned}$$

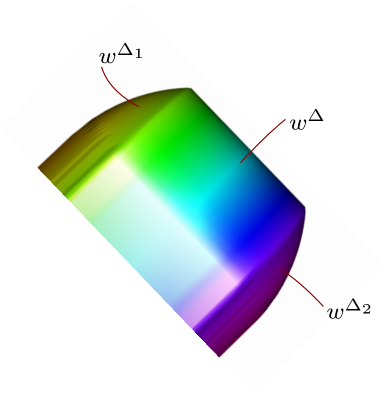
There is an analogous vector $w^{\Delta_2}$ for the other triangulation. If $h_{X,y}$ induces the flat subdivision $\Delta$ then $w$ can be any positive linear combination of $w^{\Delta_1}$ and $w^{\Delta_2}$.

# Convex Bodies




**Samworth body** and **its dual**




---

The **secondary fan** of $X$ and the **secondary polytope** of $X$.

# Unit weights

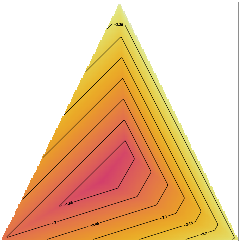
## Theorem

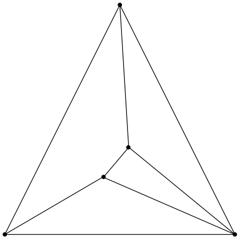
*Let $X$ be a configuration of $n = d + 2$ points that span $\mathbb{R}^d$. If $w = \frac{1}{n}(1, \ldots, 1)$, then the optimal density $\hat{p}$ is log-linear, and the optimal subdivision is trivial.*

## Example

Unit weights on the following configuration of five points

$$X = \{(0,0), (40,0), (20,40), (17,10), (21,15)\}$$



## Theorem

*For any integer $d \geq 2$, there exists a configuration of $n = d + 3$ points in $\mathbb{R}^d$ for which the optimal subdivision with respect to unit weights is non-trivial.*

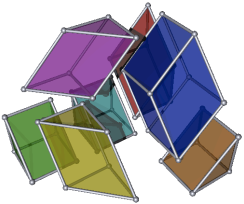
# Experiments

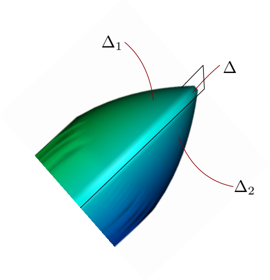
We sampled six i.i.d. points in $\mathbb{R}^2$ from four different distributions:

- Gaussian $\mathcal{N}(0,1)$
- Uniform
- Circular: $(U_1^a \cos(2\pi U_2), U_1^a \sin(2\pi U_2))$, where $U_1, U_2$ are i.i.d uniform on $[0,1]$, and $a = 0.3$
- Circular: $(U_1^a \cos(2\pi U_2), U_1^a \sin(2\pi U_2))$, where $U_1, U_2$ are i.i.d uniform on $[0,1]$, and $a = 0.1$

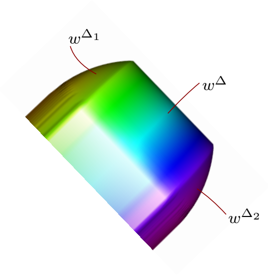
| Subdivision: number of | | | | Convex | Gaussian | Uniform | Circular | Circular |
|---|---|---|---|---|---|---|---|---|
| 3-gons | 4-gons | 5-gons | 6-gons | hull | $\mathcal{N}(0,1)$ | $a = 0.5$ | $a = 0.3$ | $a = 0.1$ |
| 1 | 0 | 0 | 0 | 3 | 948 | 533 | 257 | 34 |
| 0 | 1 | 0 | 0 | 4 | 8781 | 6719 | 4596 | 1507 |
| 0 | 0 | 1 | 0 | 5 | 8209 | 9743 | 10554 | 8504 |
| 0 | 0 | 0 | 1 | 6 | 1475 | 2805 | 4495 | 9887 |
| 2 | 0 | 0 | 0 | 4 | 8 | 3 | 6 | 7 |
| 1 | 1 | 0 | 0 | 5 | 1 | 2 | 1 | 2 |
| 3 | 0 | 0 | 0 | 3 | 6 | 2 | 2 | 1 |
| 2 | 1 | 0 | 0 | 4 | 39 | 16 | 4 | 7 |
| 2 | 0 | 1 | 0 | 5 | 1 | 1 | 0 | 1 |
| 1 | 2 | 0 | 0 | 5 | 1 | 0 | 1 | 6 |
| 4 | 0 | 0 | 0 | 4 | 1 | 0 | 0 | 0 |
| 3 | 1 | 0 | 0 | 3 | 114 | 38 | 10 | 1 |
| 3 | 0 | 1 | 0 | 4 | 39 | 20 | 9 | 2 |
| 2 | 2 | 0 | 0 | 4 | 59 | 19 | 16 | 9 |
| 5 | 0 | 0 | 0 | 3 | 3 | 0 | 0 | 0 |
| 4 | 1 | 0 | 0 | 4 | 1 | 0 | 0 | 0 |
| 4 | 0 | 1 | 0 | 3 | 90 | 27 | 8 | 1 |
| 3 | 2 | 0 | 0 | 3 | 120 | 32 | 11 | 0 |
| 5 | 1 | 0 | 0 | 3 | 50 | 11 | 3 | 0 |
| 7 | 0 | 0 | 0 | 3 | 2 | 1 | 0 | 0 |

# Open Problems











- Design a **test statistic** for log-concavity based on optimal $\Delta$.
- What is the smallest size $n$ of a configuration $X$ in $\mathbb{R}^d$ whose optimal subdivision with unit weights has at least $c$ cells?
  (e.g. we showed $n = d + 3$ for $c = 2, d \geq 2$.)
- Which subdivisions are realizable by points with unit weights?
- For a fixed $w$ and a fixed combinatorial type of subdivision $\Delta$, study the semianalytic set of all configurations $X$ such that $\Delta$ is the optimal subdivision for the data $(X, w)$.