Logarithmic Voronoi cells for Gaussian models

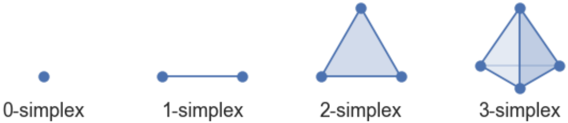
Yulia Alexandr (UC Berkeley)

Applied Combinatorics, Algebra, Topology & Statistics at KTH,
May 10, 2022

Logarithmic Voronoi cells for discrete models

- A *probability simplex* is defined as

$$\Delta_{n-1} = \{(p_1, \ldots, p_n) \,:\, p_1 + \cdots + p_n = 1, p_i \geq 0 \text{ for } i \in [n]\}.$$



- An *algebraic statistical model* is a subset $\mathcal{M} = \mathcal{V} \cap \Delta_{n-1}$ for some variety $\mathcal{V} \subseteq \mathbb{C}^n$.
- For an empirical data point $u = (u_1, \ldots, u_n) \in \Delta_{n-1}$, the *log-likelihood function* defined by u assuming distribution $p = (p_1, \ldots, p_n) \in \mathcal{M}$ is

$$\ell_u(p) = u_1 \log p_1 + u_2 \log p_2 + \cdots + u_n \log p_n + \log(c).$$

Maximum likelihood estimation

1. The maximum likelihood estimation problem (MLE):

Given a sampled empirical distribution $u \in \Delta_{n-1}$, which point $p \in \mathcal{M}$ did it most likely come from? In other words, we wish to maximize $\ell_u(p)$ over all points $p \in \mathcal{M}$.

Maximum likelihood estimation

1. The maximum likelihood estimation problem (MLE):

Given a sampled empirical distribution $u \in \Delta_{n-1}$, which point $p \in \mathcal{M}$ did it most likely come from? In other words, we wish to maximize $\ell_u(p)$ over all points $p \in \mathcal{M}$.

2. Computing logarithmic Voronoi cells:

Given a point $q \in \mathcal{M}$, what is the set of all points $u \in \Delta_{n-1}$ that have q as a global maximum when optimizing the function $\ell_u(p)$ over $\mathcal{M}$?

We call the set of all such elements $u \in \Delta_{n-1}$ above the *logarithmic Voronoi cell* at q.

Proposition (A., Heaton)

Logarithmic Voronoi cells are convex sets.

The *log-normal space* at q is the space of possible data points $u \in \mathbb{R}^n$ for which q is a critical point of $\ell_u(p)$. It is a *linear* space.

Intersecting this space with the simplex Δ_{n-1}, we obtain a polytope, which we call the *log-normal polytope* at q.

The log-normal polytope at q contains the logarithmic Voronoi cell at q.

Proposition (A., Heaton)

Logarithmic Voronoi cells are convex sets.

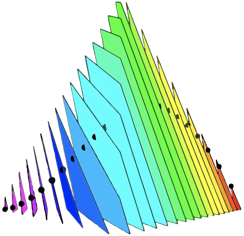
The *log-normal space* at q is the space of possible data points $u \in \mathbb{R}^n$ for which q is a critical point of $\ell_u(p)$. It is a *linear* space.

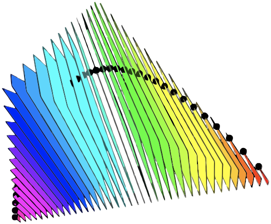
Intersecting this space with the simplex Δ_{n-1}, we obtain a polytope, which we call the *log-normal polytope* at q.

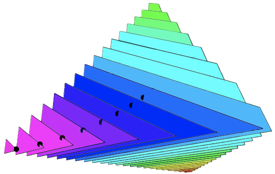
The log-normal polytope at q contains the logarithmic Voronoi cell at q.

Example (The twisted cubic.)

The curve is given by $p \mapsto \left(p^3, 3p^2(1-p), 3p(1-p)^2, (1-p)^3\right)$.







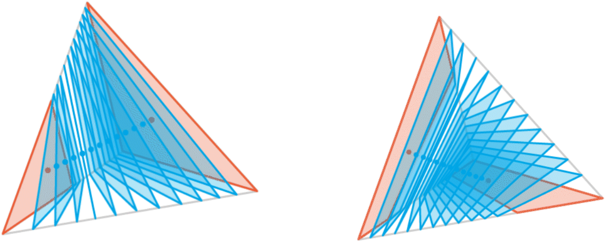
Polytopal cells

The *maximum likelihood degree* (ML degree) of $\mathcal{M}$ is the number of complex critical points when optimizing $\ell_u(x)$ over $\mathcal{M}$ for generic data u.

Theorem (A., Heaton)

If $\mathcal{M}$ is a finite model, a linear model, a toric model, or a model of ML degree 1, the logarithmic Voronoi cell at any point $p \in \mathcal{M}$ is equal to the log-normal polytope at p.

For linear models, logarithmic Voronoi cells at all interior points on the model have the same combinatorial type.



Gaussian models

Let X be an m-dimensional random vector, which has the density function

$$p_{\mu,\Sigma}(x) = \frac{1}{(2\pi)^{m/2}(\det \Sigma)^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}, \quad x \in \mathbb{R}^m$$

with respect to the parameters $\mu \in \mathbb{R}^m$ and $\Sigma \in \mathrm{PD}_m$.

Such X is distributed according to the *multivariate normal distribution*, also called the *Gaussian distribution* $\mathcal{N}(\mu, \Sigma)$.

For $\Theta \subseteq \mathbb{R}^m \times \mathrm{PD}_m$, the statistical model

$$\mathcal{P}_\Theta = \{\mathcal{N}(\mu, \Sigma) : \theta = (\mu, \Sigma) \in \Theta\}$$

is called a *Gaussian model*. We identify the Gaussian model $\mathcal{P}_\Theta$ with its parameter space Θ.

Gaussian models

For a sampled data consisting of n vectors $X^{(1)}, \cdots, X^{(n)} \in \mathbb{R}^m$, we define the *sample mean* and *sample covariance* as

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X^{(i)} \qquad \text{and} \qquad S = \frac{1}{n}\sum_{i=1}^{n}(X^{(i)} - \bar{X})(X^{(i)} - \bar{X})^T,$$

respectively. The *log-likelihood function* is defined as

$$\ell_n(\mu, \Sigma) = -\frac{n}{2}\log\det\Sigma - \frac{1}{2}\operatorname{tr}\left(S\Sigma^{-1}\right) - \frac{n}{2}(\bar{X} - \mu)^T\Sigma^{-1}(\bar{X} - \mu).$$

The problem of maximizing $\ell_n(\mu, \Sigma)$ over Θ is *maximum likelihood estimation*.

The *logarithmic Voronoi cell* of $\theta = (\mu, \Sigma) \in \Theta$, is the set of all multivariate distributions $(\bar{X}, S)$ for which ℓ_n is maximized at θ.

Gaussian models

In practice, we will only consider models given by parameter spaces of the form $\Theta = \mathbb{R}^m \times \Theta_2$ where $\Theta_2 \subseteq \mathrm{PD}_m$. **Thus, a Gaussian model is a subset of** PD_m. The log-likelihood function is then

$$\ell_n(\Sigma, S) = -\frac{n}{2}\log\det\Sigma - \frac{n}{2}\operatorname{tr}(S\Sigma^{-1}).$$

For $\Sigma \in \Theta_2$, the *log-normal matrix space* $\mathcal{N}_\Sigma\Theta_2$ at Σ is the set of $S \in \mathrm{Sym}_m(\mathbb{R})$ such that Σ appears as a critical point of $\ell_n(\Sigma, S)$. The intersection $\mathrm{PD}_m \cap \mathcal{N}_\Sigma\Theta_2$ is the *log-normal spectrahedron* $\mathcal{K}_\Theta\Sigma$ at Σ.

If Σ is a covariance matrix, its inverse Σ^{-1} is a *concentration matrix*.

Discrete vs. Gaussian

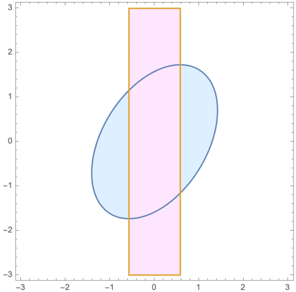
$$\begin{aligned}
\text{Simplex } \Delta_{n-1} &\longleftrightarrow \text{Cone } \mathrm{PD}_m \\
\text{Model } \mathcal{M} \subseteq \Delta_{n-1} &\longleftrightarrow \text{Model } \Theta \subseteq \mathrm{PD}_m \\
\sum_{i=1}^{n} u_i \log p_i &\longleftrightarrow \log\det\Sigma - \mathrm{tr}(S\Sigma^{-1}) \\
\text{Log-normal space } &\longleftrightarrow \text{Log-normal matrix space} \\
\text{Log-normal polytope} &\longleftrightarrow \text{Log-normal spectrahedron}
\end{aligned}$$

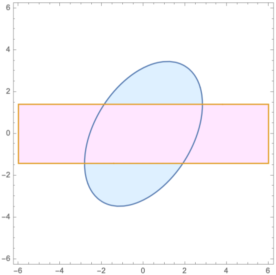
Example: conditional independence model

Consider the model Θ given by the conditional independence statements $X_1 \perp\!\!\!\perp X_3$ and $X_1 \perp\!\!\!\perp X_3 | X_2$. Parametrically,

$$\Theta = \{\Sigma = (\sigma_{ij}) \in \mathrm{PD}_3 : \sigma_{13} = 0 \text{ and } \sigma_{12}\sigma_{23} - \sigma_{22}\sigma_{13} = 0\}.$$

This model is the union of two linear three-dimensional planes. It has ML degree 2. The log-normal spectrahedron of each point $\Sigma \in \Theta$ is an ellipse. Each log-Voronoi cell is given as:





Concentration models

Let $G = (V, E)$ be a simple undirected graph with $|V(G)| = m$. A *concentration model* of G is the model

$$\Theta = \{\Sigma \in \mathrm{PD}_m : (\Sigma)^{-1}_{ij} = 0 \text{ if } ij \notin E(G) \text{ and } i \neq j\}.$$

Proposition (A., Hoșten)

Let Θ be a concentration model of some graph G. For a point $\Sigma \in \Theta$, its logaritmic Voronoi cell is equal to its log-normal spectrahedron.

In fact, we can describe $\log \mathrm{Vor}_\Theta(\Sigma)$ as:

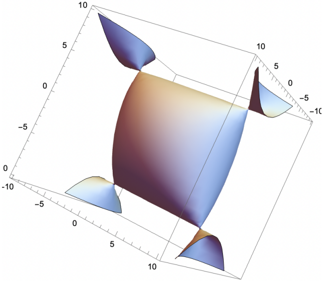
$$\log \mathrm{Vor}_\Theta(\Sigma) = \{S \in \mathrm{PD}_m : \Sigma_{ij} = S_{ij} \text{ for all } ij \in E(G) \text{ and } i = j\}.$$

Example

The concentration model of $\overset{1}{\bullet}\!-\!\overset{2}{\bullet}\!-\!\overset{3}{\bullet}\!-\!\overset{4}{\bullet}$ is defined by

$$\Theta = \{\Sigma = (\sigma_{ij}) \in \mathrm{PD}_4 : (\Sigma^{-1})_{13} = (\Sigma^{-1})_{14} = (\Sigma^{-1})_{24} = 0\}.$$

Let $\Sigma = \begin{pmatrix} 6 & 1 & \frac{1}{7} & \frac{1}{28} \\ 1 & 7 & 1 & \frac{1}{4} \\ \frac{1}{7} & 1 & 8 & 2 \\ \frac{1}{28} & \frac{1}{4} & 2 & 9 \end{pmatrix}$.



Then $\log \mathrm{Vor}_\Theta(\Sigma) = \left\{ (x,y,z) : \begin{pmatrix} 6 & 1 & x & y \\ 1 & 7 & 1 & z \\ x & 1 & 8 & 2 \\ y & z & 2 & 9 \end{pmatrix} \succ 0 \right\}$.

Directed graphical models

Directed graphical models are Gaussian models defined by directed acyclic graphs (DAGs). Each vertex j defines a random variable such that

$$X_j = \sum_{k \in \mathrm{pa}(j)} \lambda_{kj} X_k + \varepsilon_j .$$

Theorem (A., Hoşten)

For Gaussian models of ML degree one, logarithmic Voronoi cells and log-normal spectrahedra coinside.

Corollary

Logarithmic Voronoi cells of directed graphical models are equal to log-normal spectrahedra.

Covariance models and the bivariate correlation model

Let $A \in \mathrm{PD}_m$ and let $\mathcal{L}$ be a linear subspace of $\mathrm{Sym}_m(\mathbb{R})$. Then $A + \mathcal{L}$ is an affine subspace of $\mathrm{Sym}(\mathbb{R}^m)$. Models defined by $\Theta = (A + \mathcal{L}) \cap \mathrm{PD}_m$ are called *covariance models*.

The *bivariate correlation model* is the covariance model

$$\Theta = \left\{ \Sigma_x = \begin{pmatrix} 1 & x \\ x & 1 \end{pmatrix} : x \in (-1, 1) \right\}.$$

This model has ML degree 3. For a general matrix $S = (S_{ij}) \in \mathrm{PD}_2$, the critical points are given by the roots of the polynomial

$$f(x) = x^3 - bx^2 - x(1 - 2a) - b,$$

where $b = S_{12}$ and $a = (S_{11} + S_{22})/2$ [Améndola and Zwiernik].

The bivariate correlation model

Fix $c \in (-1, 1)$ so $\Sigma_c \in \Theta$. The log-normal spectrahedron of Σ_c is

$$\begin{aligned}\mathcal{K}_\Theta(\Sigma_c) &= \{S \in \mathrm{PD}_2 : f(c) = 0\} \\ &= \{S \in \mathrm{PD}_2 : a = (bc^2 - c^3 + b + c)/2c\} \\ &= \left\{ S_{b,k} = \begin{pmatrix} k & b \\ b & 2a - k \end{pmatrix} \succ 0 : \substack{0 \le k \le 2a, \\ a = (bc^2 - c^3 + b + c)/2c} \right\}.\end{aligned}$$

Theorem (A., Hoșten)

Let Θ be the bivariate correlation model and let $\Sigma_c \in \Theta$. If $c > 0$, then

$$\log \mathrm{Vor}_\Theta(\Sigma_c) = \{S_{b,k} \in \mathcal{K}_\Theta(\Sigma_c) : b \ge 0\}.$$

If $c < 0$, then

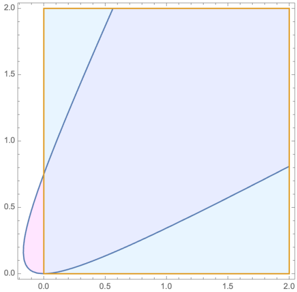
$$\log \mathrm{Vor}_\Theta(\Sigma_c) = \{S_{b,k} \in \mathcal{K}_\Theta(\Sigma_c) : b \le 0\}.$$

The bivariate correlation model

Important things to note:

- The log-Voronoi cell of Σ_c is strictly contained in the log-normal spectrahedron of Σ_c.
- Logarithmic Voronoi cells of Θ are semi-algebraic sets! **This is extremely surprising!**

The logarithmic Voronoi cell and the log-normal spectrahedron at $c = 1/2$:



Equicorrelation models

An *equicorrelation model* is given by the parameter space

$\Theta_m = \{\Sigma_x \in \mathrm{Sym}(\mathbb{R}^m) : \Sigma_{ii} = 1, \Sigma_{ij} = x \text{ for } i \neq j, i, j \in [m], x \in \mathbb{R}\} \cap \mathrm{PD}_m$.

Fix $c \in \mathbb{R}$ such that $-\frac{1}{m-1} < c < 1$. For $S \in \mathrm{PD}_m$, we define the *symmetrized sample covariance matrix* to be the matrix

$$\bar{S} = \frac{1}{m!} \sum_{P \in S_m} PSP^T.$$

Let $\mathcal{N}$ denote the space of all symmetrized sample covariance matrices.
Note:

- $\bar{S}_{ii} = a$ and $\bar{S}_{ij} = b$ whenever $i \neq j$,
- $\langle S, \Sigma_c^{-1} \rangle = \langle \bar{S}, \Sigma_c^{-1} \rangle$.

Equicorrelation models

Every equicorrelation model has ML degree 3 with

$$f_m(x) = (m-1)x^3 + ((m-2)(a-1) - (m-1)b)x^2 + (2a-1)x - b.$$

Setting $f_m(c) = 0$, we get the relationship $b = g(a)$. Note:

$$\mathcal{K}_{\Theta_m}(\Sigma_c) \cap \mathcal{N} = \{\bar{S}_b \in \mathrm{PD}_m : b = g(a)\}.$$

Let c_1 and c_2 denote the other two roots of f_m. Then:

$$\log \mathrm{Vor}_{\Theta_m}(\Sigma_c) \cap \mathcal{N} = \{\bar{S}_b \in \mathcal{K}_{\Theta_m}(\Sigma_c) \cap \mathcal{N} : \ell_n(\Sigma_c, \bar{S}_b) \geq \ell_n(\Sigma_{c_i}, \bar{S}_b), i = 1, 2\}$$

Theorem (A., Hoșten)

Let $\Sigma_c \in \Theta_m$. The logarithmic Voronoi cell at Σ_c is given as

$$\log \mathrm{Vor}_{\Theta_m}(\Sigma_c) = \{S \in \mathrm{PD}_m : \psi(S) = \bar{S}, \bar{S} \in \log \mathrm{Vor}_{\Theta_m}(\Sigma_c) \cap \mathcal{N}\},$$

where $\psi : \mathrm{PD}_m \to \mathcal{N} : S \mapsto \bar{S}$.

The boundary: transcendental? algebraic?

Given a Gaussian model Θ and $\Sigma \in \Theta$, the matrix $S \in \mathrm{PD}_m$ is on the boundary of $\log \mathrm{Vor}_\Theta(\Sigma)$ if $S \in \log \mathrm{Vor}_\Theta(\Sigma)$ and there is some $\Sigma' \in \Theta$ such that $\ell(\Sigma, S) = \ell(\Sigma', S)$.

The bivariate correlation models fit into a larger class of models known as *unrestricted correlation models*. Such a model is given by the parameter space

$$\Theta = \{\Sigma \in \mathrm{Sym}(\mathbb{R}^m) : \Sigma_{ii} = 1,\ i \in [m]\} \cap \mathrm{PD}_m \,.$$

When $m = 3$, the model is a compact spectrahedron known as the elliptope. Its ML degree is 15.

Conjecture

The logarithmic Voronoi cells for general points on the elliptope are not semi-algebraic; in other words, their boundary is defined by transcendental functions.

Thanks!