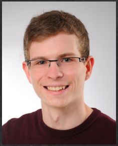
Invariant theory and scaling algorithms for maximum likelihood estimation

Kathlén Kohn
KTH Stockholm

joint with

Carlos Améndola
TU Munich

Philipp Reichenbach
TU Berlin

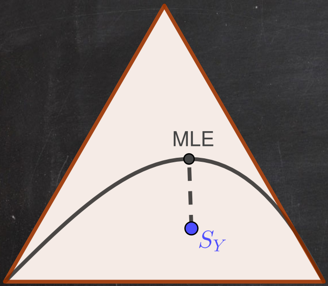


Anna Seigal
University of Oxford

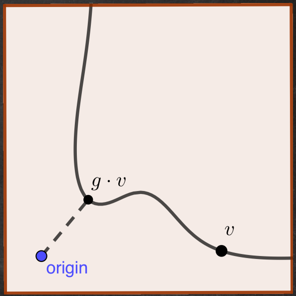
May 20, 2020

Global picture

Statistics

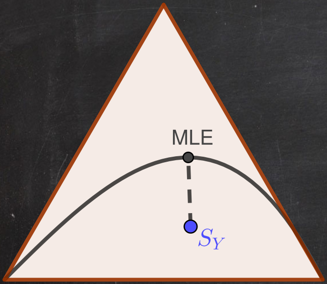


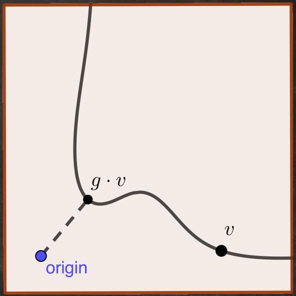
Invariant theory



Global picture

Statistics | **Invariant theory**

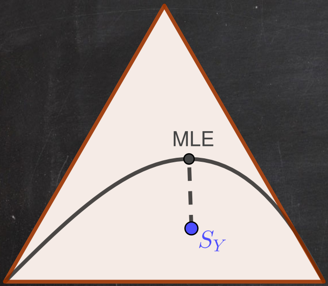




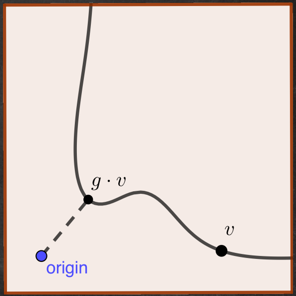
Given: statistical model
sample data S_Y

Global picture

Statistics



Invariant theory



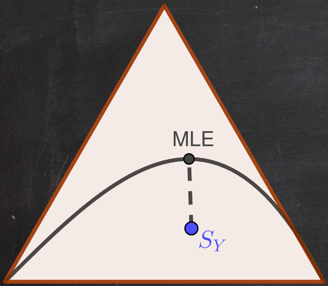
Given: statistical model
sample data S_Y

Task: find **maximum likelihood estimate (MLE)**

$=$ point in model that best fits S_Y

Global picture

Statistics

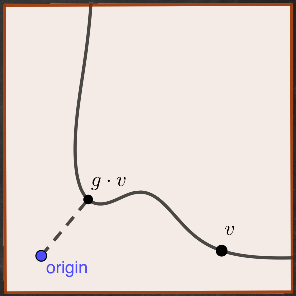


Given: statistical model
sample data S_Y

Task: find **maximum likelihood estimate (MLE)**

$=$ point in model that best fits S_Y

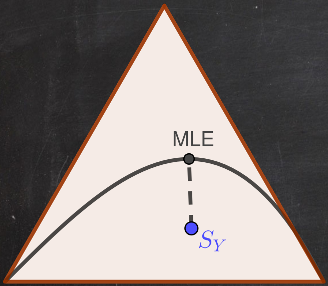
Invariant theory



Given: orbit $G \cdot v = \{g \cdot v \mid g \in G\}$

Global picture

Statistics

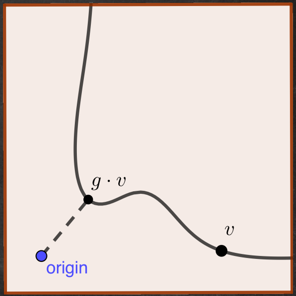


Given: statistical model
sample data S_Y

Task: find **maximum likelihood estimate (MLE)**

$=$ point in model that best fits S_Y

Invariant theory



Given: orbit $G \cdot v = \{g \cdot v \mid g \in G\}$

Task: compute **capacity**

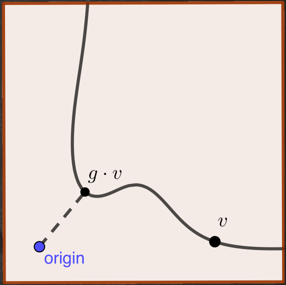
$=$ closest distance of orbit to origin

Invariant theory

Stability notions

The **orbit** of a vector v in a vector space V under an action by a group G is

$$G.v = \{g \cdot v \mid g \in G\} \subset V.$$

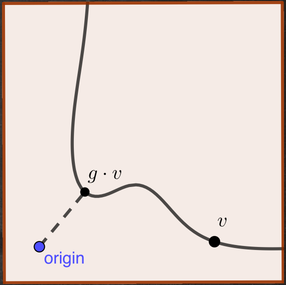


Invariant theory

Stability notions

The **orbit** of a vector v in a vector space V under an action by a group G is

$$G.v = \{g \cdot v \mid g \in G\} \subset V.$$



- v is **unstable** iff $0 \in \overline{G.v}$ (i.e. v can be scaled to 0 in the limit)

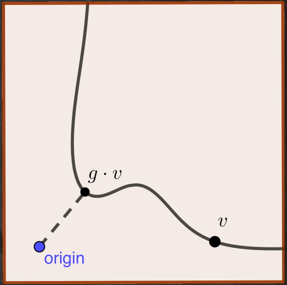
The **null cone** of the action by G is the set of unstable vectors v.

Invariant theory

Stability notions

The **orbit** of a vector v in a vector space V under an action by a group G is

$$G.v = \{g \cdot v \mid g \in G\} \subset V.$$



- v is **unstable** iff $0 \in \overline{G.v}$ (i.e. v can be scaled to 0 in the limit)
- v **semistable** iff $0 \notin \overline{G.v}$

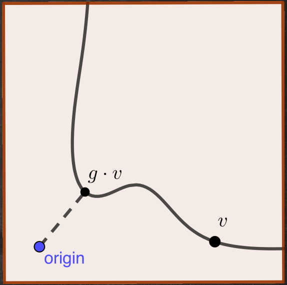
The **null cone** of the action by G is the set of unstable vectors v.

Invariant theory

Stability notions

The **orbit** of a vector v in a vector space V under an action by a group G is

$$G.v = \{g \cdot v \mid g \in G\} \subset V.$$



- v is **unstable** iff $0 \in \overline{G.v}$ (i.e. v can be scaled to 0 in the limit)
- v **semistable** iff $0 \notin \overline{G.v}$
- v **polystable** iff $v \neq 0$ and its orbit $G.v$ is closed

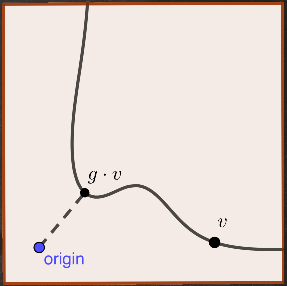
The **null cone** of the action by G is the set of unstable vectors v.

Invariant theory

Stability notions

The **orbit** of a vector v in a vector space V under an action by a group G is

$$G.v = \{g \cdot v \mid g \in G\} \subset V.$$



- v is **unstable** iff $0 \in \overline{G.v}$ (i.e. v can be scaled to 0 in the limit)
- v **semistable** iff $0 \notin \overline{G.v}$
- v **polystable** iff $v \neq 0$ and its orbit $G.v$ is closed
- v is **stable** iff v is polystable and its stabilizer is finite

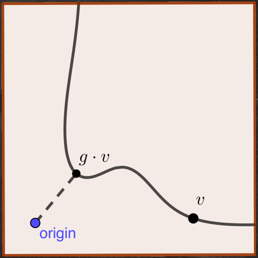
The **null cone** of the action by G is the set of unstable vectors v.

Invariant theory

Null cone membership testing

Classical and often hard question: Describe null cone
(essentially equivalent to finding generators for the ring of polynomial invariants)

Modern approach: Provide a test to determine if a vector v lies in null cone



Invariant theory

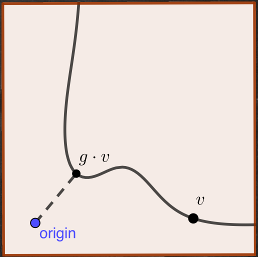
Null cone membership testing

Classical and often hard question: Describe null cone
(essentially equivalent to finding generators for the ring of polynomial invariants)

Modern approach: Provide a test to determine if a vector v lies in null cone

The **capacity** of v is

$$\mathrm{cap}_G(v) := \inf_{g \in G} \|g \cdot v\|_2^2.$$



Observation: $\mathrm{cap}_G(v) = 0$ iff v lies in null cone

Invariant theory

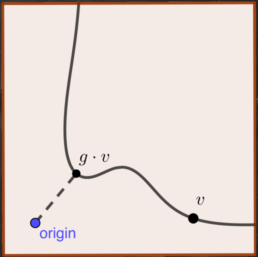
Null cone membership testing

Classical and often hard question: Describe null cone
(essentially equivalent to finding generators for the ring of polynomial invariants)

Modern approach: Provide a test to determine if a vector v lies in null cone

The **capacity** of v is

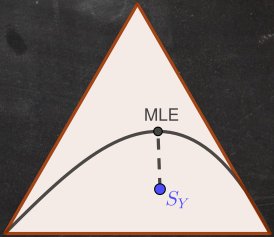
$$\mathrm{cap}_G(v) := \inf_{g \in G} \|g \cdot v\|_2^2.$$



Observation: $\mathrm{cap}_G(v) = 0$ iff v lies in null cone

Hence: Testing null cone membership is a minimization problem.
$\rightsquigarrow$ algorithms: [series of 3 papers in 2017 – 2019 by
Bürgisser, Franks, Garg, Oliveira, Walter, Wigderson]

Maximum likelihood estimation

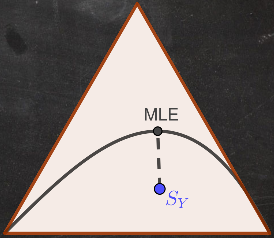


Maximum likelihood estimation

Given:

- $\mathcal{M}$: a statistical **model** $=$ a set of probability distributions
- $Y = (Y_1, \ldots, Y_n)$: n samples of observed **data**

Goal: find a distribution in the model $\mathcal{M}$ that best fits the empirical data Y



Maximum likelihood estimation

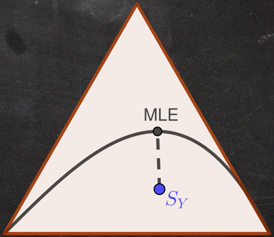
Given:

- $\mathcal{M}$: a statistical **model** = a set of probability distributions
- $Y = (Y_1, \ldots, Y_n)$: n samples of observed **data**

Goal: find a distribution in the model $\mathcal{M}$ that best fits the empirical data Y

Approach: maximize the **likelihood function**

$$L_Y(\rho) := \rho(Y_1) \cdots \rho(Y_n), \qquad \text{where } \rho \in \mathcal{M}.$$



A **maximum likelihood estimate (MLE)** is a distribution in the model $\mathcal{M}$ that maximizes the likelihood L_Y.

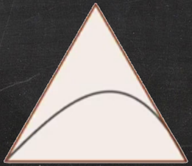
Discrete statistical models

A probability distribution on m states is determined by is **probability mass function** ρ, where ρ_j is the probability that the j-th state occurs.

ρ is a point in the **probability simplex**

$$\Delta_{m-1} = \left\{ q \in \mathbb{R}^m \mid q_j \geq 0 \text{ and } \sum q_j = 1 \right\}.$$

A **discrete statistical model** $\mathcal{M}$ is a subset of the simplex Δ_{m-1}.

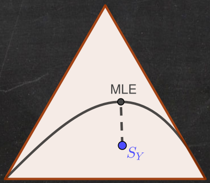


Discrete statistical models

maximum likelihood estimation

Given data is a **vector of counts** $Y \in \mathbb{Z}^m_{\geq 0}$,
where Y_j is the number of times the j-th state occurs.

The **empirical distribution** is $S_Y = \frac{1}{n} Y \in \Delta_{m-1}$, where $n = Y_1 + \ldots + Y_m$.



Discrete statistical models

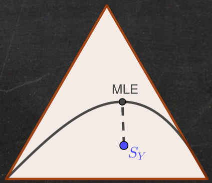
maximum likelihood estimation

Given data is a **vector of counts** $Y \in \mathbb{Z}^m_{\geq 0}$,
where Y_j is the number of times the j-th state occurs.

The **empirical distribution** is $S_Y = \frac{1}{n} Y \in \Delta_{m-1}$, where $n = Y_1 + \ldots + Y_m$.

The **likelihood function** takes the form $\quad L_Y(\rho) = \rho_1^{Y_1} \cdots \rho_m^{Y_m}, \quad$ where $\rho \in \mathcal{M}$.

An **MLE** is a point in model $\mathcal{M}$ that maximizes the likelihood L_Y of observing Y.



Log-linear models

$=$ set of distributions whose logarithms lie in a fixed linear space.

Let $A \in \mathbb{Z}^{d \times m}$, and define

$$\mathcal{M}_A = \{\rho \in \Delta_{m-1} \mid \log \rho \in \mathrm{rowspan}(A)\}\,.$$

We assume that $\mathbb{1} := (1, \ldots, 1) \in \mathrm{rowspan}(A)$ (i.e., uniform distribution in $\mathcal{M}_A$).

Log-linear models

$=$ set of distributions whose logarithms lie in a fixed linear space.

Let $A \in \mathbb{Z}^{d \times m}$, and define

$$\mathcal{M}_A = \{\rho \in \Delta_{m-1} \mid \log \rho \in \mathrm{rowspan}(A)\}\,.$$

We assume that $\mathbb{1} := (1, \ldots, 1) \in \mathrm{rowspan}(A)$ (i.e., uniform distribution in $\mathcal{M}_A$).

Matrix $A = [a_1 \mid a_2 \mid \ldots \mid a_m]$ also defines an **action by the torus**

- $\mathrm{GT}_d =$ group of complex, diagonal, invertible $d \times d$ matrices
- on $\mathbb{C}^m$:
- $g \in \mathrm{GT}_d$ acts on $x \in \mathbb{C}^m$ by left multiplication with

$$\begin{bmatrix} g^{a_1} & & \\ & \ddots & \\ & & g^{a_m} \end{bmatrix}, \qquad \text{where } g^{a_j} = g_1^{a_{1j}} \ldots g_d^{a_{dj}}\,.$$

Log-linear models

= set of distributions whose logarithms lie in a fixed linear space.

Let $A \in \mathbb{Z}^{d \times m}$, and define

$$\mathcal{M}_A = \{\rho \in \Delta_{m-1} \mid \log \rho \in \mathrm{rowspan}(A)\}\,.$$

We assume that $\mathbb{1} := (1, \ldots, 1) \in \mathrm{rowspan}(A)$ (i.e., uniform distribution in $\mathcal{M}_A$).

Matrix $A = [a_1 \mid a_2 \mid \ldots \mid a_m]$ also defines an **action by the torus**

- GT_d = group of complex, diagonal, invertible $d \times d$ matrices
- on $\mathbb{C}^m$:
- $g \in \mathrm{GT}_d$ acts on $x \in \mathbb{C}^m$ by left multiplication with

$$\begin{bmatrix} g^{a_1} & & \\ & \ddots & \\ & & g^{a_m} \end{bmatrix}, \qquad \text{where } g^{a_j} = g_1^{a_{1j}} \ldots g_d^{a_{dj}}\,.$$

$\mathcal{M}_A$ is the orbit of the uniform distribution in $\Delta_{m-1} \cap \mathbb{R}^m_{>0}$.

Example

$\mathcal{M}_A = \{\rho \in \Delta_{m-1} \mid \log \rho \in \mathrm{rowspan}(A)\}$. $\quad A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix}$

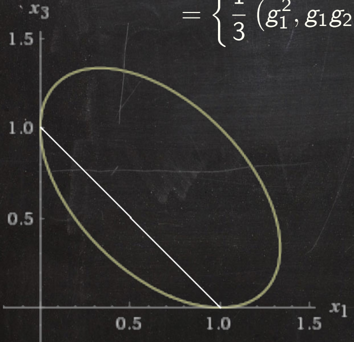
$g \in \mathrm{GT}_2$ acts on $x \in \mathbb{C}^3$ by $\begin{bmatrix} g^{a_1} & & \\ & g^{a_2} & \\ & & g^{a_3} \end{bmatrix} = \begin{bmatrix} g_1^2 & & \\ & g_1 g_2 & \\ & & g_2^2 \end{bmatrix}.$

Example

$$\mathcal{M}_A = \{\rho \in \Delta_{m-1} \mid \log \rho \in \mathrm{rowspan}(A)\}. \qquad A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix}$$

$g \in \mathrm{GT}_2$ acts on $x \in \mathbb{C}^3$ by $\begin{bmatrix} g^{a_1} & & \\ & g^{a_2} & \\ & & g^{a_3} \end{bmatrix} = \begin{bmatrix} g_1^2 & & \\ & g_1 g_2 & \\ & & g_2^2 \end{bmatrix}.$

$$\begin{aligned} \mathcal{M}_A &= (\mathrm{GT}_2 \cdot \tfrac{1}{3}\mathbb{1}) \cap \Delta_2 \cap \mathbb{R}^3_{>0} \\ &= \left\{ \frac{1}{3}\left(g_1^2, g_1 g_2, g_2^2\right) \mid g_1, g_2 > 0,\ g_1^2 + g_1 g_2 + g_2^2 = 3 \right\} \end{aligned}$$



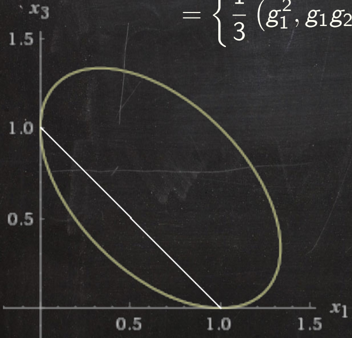
Example

$$\mathcal{M}_A = \{\rho \in \Delta_{m-1} \mid \log \rho \in \mathrm{rowspan}(A)\}. \qquad A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix}$$

$g \in \mathrm{GT}_2$ acts on $x \in \mathbb{C}^3$ by $\begin{bmatrix} g^{a_1} & & \\ & g^{a_2} & \\ & & g^{a_3} \end{bmatrix} = \begin{bmatrix} g_1^2 & & \\ & g_1 g_2 & \\ & & g_2^2 \end{bmatrix}$.

$$\mathcal{M}_A = (\mathrm{GT}_2 \cdot \frac{1}{3}\mathbb{1}) \cap \Delta_2 \cap \mathbb{R}^3_{>0}$$

$$= \left\{ \frac{1}{3}\left(g_1^2, g_1 g_2, g_2^2\right) \mid g_1, g_2 > 0,\ g_1^2 + g_1 g_2 + g_2^2 = 3 \right\}$$



$$= \left\{ \rho \in \mathbb{R}^3_{>0} \mid \rho_2^2 = \rho_1 \rho_3,\ \rho_1 + \rho_2 + \rho_3 = 1 \right\}$$

Example

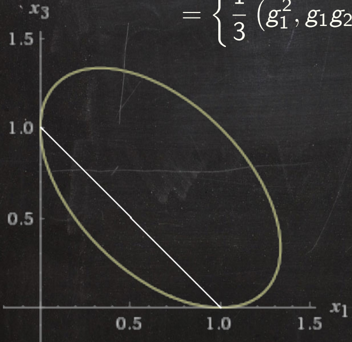
$$\mathcal{M}_A = \{\rho \in \Delta_{m-1} \mid \log \rho \in \mathrm{rowspan}(A)\}. \qquad A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix}$$

$g \in \mathrm{GT}_2$ acts on $x \in \mathbb{C}^3$ by $\begin{bmatrix} g^{a_1} & & \\ & g^{a_2} & \\ & & g^{a_3} \end{bmatrix} = \begin{bmatrix} g_1^2 & & \\ & g_1 g_2 & \\ & & g_2^2 \end{bmatrix}$.

$$\mathcal{M}_A = (\mathrm{GT}_2 \cdot \frac{1}{3}\mathbb{1}) \cap \Delta_2 \cap \mathbb{R}^3_{>0}$$

$$= \left\{ \frac{1}{3}\,(g_1^2, g_1 g_2, g_2^2) \mid g_1, g_2 > 0,\ g_1^2 + g_1 g_2 + g_2^2 = 3 \right\}$$

$$= \{\rho \in \mathbb{R}^3_{>0} \mid \rho_2^2 = \rho_1 \rho_3,\ \rho_1 + \rho_2 + \rho_3 = 1\}$$



other examples: independence model, graphical models, hierarchical models, . . .

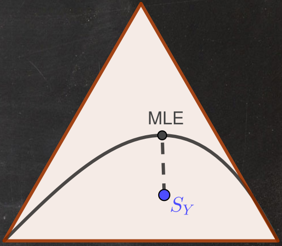
Maximum likelihood estimation

for log-linear models

An MLE in $\mathcal{M}_A$ given data Y is a point $\hat{\rho}$ in the model such that

$$A\hat{\rho} = AS_Y, \qquad \text{where } S_Y = \frac{1}{n}Y.$$

The MLE is unique **if it exists**!

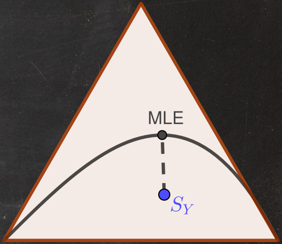


Maximum likelihood estimation

for log-linear models

An MLE in $\mathcal{M}_A$ given data Y is a point $\hat{\rho}$ in the model such that

$$A\hat{\rho} = AS_Y, \qquad \text{where } S_Y = \frac{1}{n}Y.$$



The MLE is unique **if it exists**!

Model $\mathcal{M}_A$ is not closed: MLE may not exist if S_Y has zeroes. True maximizer could be on boundary of model.

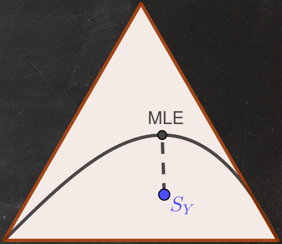
Maximum likelihood estimation

for log-linear models

An MLE in $\mathcal{M}_A$ given data Y is a point $\hat{\rho}$ in the model such that

$$A\hat{\rho} = AS_Y, \qquad \text{where } S_Y = \frac{1}{n}Y.$$

The MLE is unique **if it exists**!



Model $\mathcal{M}_A$ is not closed: MLE may not exist if S_Y has zeroes.
True maximizer could be on boundary of model.

polyhedral condition for MLE existence:
For $A = [a_1 \mid a_2 \mid \dots \mid a_m] \in \mathbb{Z}^{d \times m}$, we define

$$P(A) = \mathrm{conv}\,\{a_1, a_2, \dots, a_m\} \subset \mathbb{R}^d.$$

Theorem (Eriksson, Fienberg, Rinaldo, Sullivant '06)
MLE given Y exists in $\mathcal{M}_A$ iff AS_Y is in relative interior of $P(A)$.

Stability for torus actions

The action of the torus GT_d given by the matrix $A \in \mathbb{Z}^{d \times m}$ is in fact well-defined on projective space $\mathbb{P}^{m-1}$.

Stability for torus actions

The action of the torus GT_d given by the matrix $A \in \mathbb{Z}^{d\times m}$ is in fact well-defined on projective space $\mathbb{P}^{m-1}$.

A **linearization** is a consistent action on $\mathbb{C}^m$, given by a character $b \in \mathbb{Z}^d$:

$$g \in \mathrm{GT}_d \text{ acts on } x \in \mathbb{C}^m \text{ by } \begin{bmatrix} g^{a_1-b} & & \\ & \ddots & \\ & & g^{a_m-b} \end{bmatrix}.$$

Stability for torus actions

The action of the torus GT_d given by the matrix $A \in \mathbb{Z}^{d \times m}$ is in fact well-defined on projective space $\mathbb{P}^{m-1}$.

A **linearization** is a consistent action on $\mathbb{C}^m$, given by a character $b \in \mathbb{Z}^d$:

$$g \in \mathrm{GT}_d \text{ acts on } x \in \mathbb{C}^m \text{ by } \begin{bmatrix} g^{a_1 - b} & & \\ & \ddots & \\ & & g^{a_m - b} \end{bmatrix}.$$

polyhedral conditions for stability:

Define sub-polytopes of $P(A) = \mathrm{conv}\{a_1, a_2, \ldots, a_m\}$ that depend on $x \in \mathbb{C}^m$:

$$P_x(A) = \mathrm{conv}\left\{a_j \mid j \in \mathrm{supp}(x)\right\}.$$

Theorem (standard, proof via Hilbert-Mumford criterion)

Consider the action of GT_d given by matrix $A \in \mathbb{Z}^{d \times m}$ with linearization $b \in \mathbb{Z}^d$.

(a)	x unstable	$\Leftrightarrow$	$b \notin P_x(A)$	can be scaled to 0 in the limit
(b)	x semistable	$\Leftrightarrow$	$b \in P_x(A)$	cannot be scaled to 0 in the limit
(c)	x polystable	$\Leftrightarrow$	$b \in \mathrm{relint}\, P_x(A)$	closed orbit
(d)	x stable	$\Leftrightarrow$	$b \in \mathrm{int}\, P_x(A)$	finite stabilizer

Stability for torus actions

The action of the torus GT_d given by the matrix $A \in \mathbb{Z}^{d\times m}$ is in fact well-defined on projective space $\mathbb{P}^{m-1}$.
A **linearization** is a consistent action on $\mathbb{C}^m$, given by a character $b \in \mathbb{Z}^d$:

$$g \in \mathrm{GT}_d \text{ acts on } x \in \mathbb{C}^m \text{ by } \begin{bmatrix} g^{a_1 - b} & & \\ & \ddots & \\ & & g^{a_m - b} \end{bmatrix}.$$

polyhedral conditions for stability:
Define sub-polytopes of $P(A) = \mathrm{conv}\{a_1, a_2, \ldots, a_m\}$ that depend on $x \in \mathbb{C}^m$:

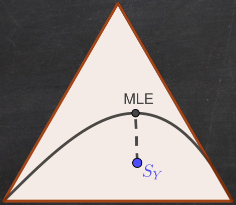
$$P_x(A) = \mathrm{conv}\,\{a_j \mid j \in \mathrm{supp}(x)\}\,.$$

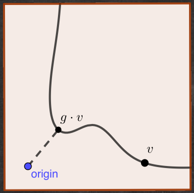
Theorem (standard, proof via Hilbert-Mumford criterion)
Consider the action of GT_d given by matrix $A \in \mathbb{Z}^{d\times m}$ with linearization $b \in \mathbb{Z}^d$.

(a)	x unstable	$\Leftrightarrow$	$b \notin P_x(A)$	can be scaled to 0 in the limit
(b)	x semistable	$\Leftrightarrow$	$b \in P_x(A)$	cannot be scaled to 0 in the limit
(c)	x **polystable**	$\Leftrightarrow$	$b \in \mathrm{relint}\, P_x(A)$	closed orbit
(d)	x stable	$\Leftrightarrow$	$b \in \mathrm{int}\, P_x(A)$	finite stabilizer

Combining both worlds

Theorem (Améndola, Kohn, Reichenbach, Seigal)
Consider a vector of counts $Y \in \mathbb{Z}^m$ with $n = \sum Y_j$, matrix $A \in \mathbb{Z}^{d \times m}$, and $b = AY \in \mathbb{Z}^d$. The MLE given Y in $\mathcal{M}_A$ exists iff $\mathbb{1} \in \mathbb{C}^m$ is polystable under the action of GT_d given by matrix nA with linearization b.

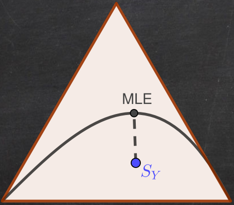


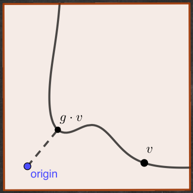


attains its maximum $\Leftrightarrow$ attains its minimum

Combining both worlds

Theorem (Améndola, Kohn, Reichenbach, Seigal)
Consider a vector of counts $Y \in \mathbb{Z}^m$ with $n = \sum Y_j$, matrix $A \in \mathbb{Z}^{d \times m}$, and $b = AY \in \mathbb{Z}^d$. The MLE given Y in $\mathcal{M}_A$ exists iff $\mathbb{1} \in \mathbb{C}^m$ is polystable under the action of GT_d given by matrix nA with linearization b.





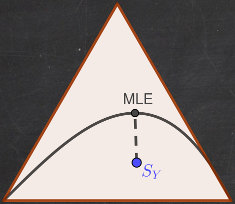
attains its maximum $\Leftrightarrow$ attains its minimum

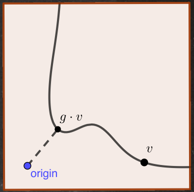
How are the two optimal points related?

Theorem (cont'd)
If $x \in \mathbb{C}^m$ is a point of minimal norm in the orbit $\mathrm{GT}_d \cdot \mathbb{1}$, then the MLE is

$$\frac{x^{(2)}}{\|x\|^2}, \qquad \text{where } x^{(2)} \text{ is the vector with } j\text{-th entry } |x_j|^2.$$

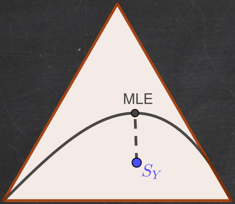
Algorithmic consequences

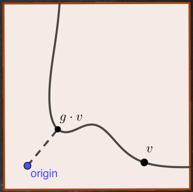




algorithms for finding MLE, e.g. iterative proportional scaling (IPS) $\leftrightarrow$ scaling algorithms to compute capacity

Algorithmic consequences



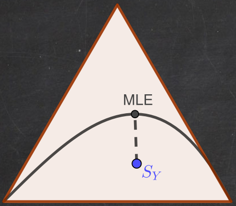


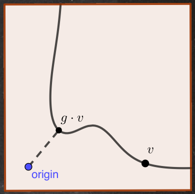
algorithms for finding MLE, e.g. iterative proportional scaling (IPS) $\leftrightarrow$ scaling algorithms to compute capacity

maximize likelihood $\Leftrightarrow$ minimize KL divergence

minimize ℓ_2-norm

Algorithmic consequences





algorithms for finding MLE, e.g. iterative proportional scaling (IPS) $\leftrightarrow$ scaling algorithms to compute capacity

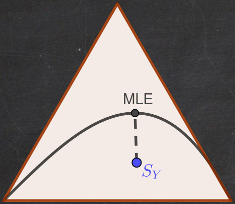
maximize likelihood $\Leftrightarrow$ minimize KL divergence

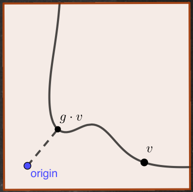
minimize ℓ_2-norm

model lives in $\Delta_{m-1} \cap \mathbb{R}^m_{>0}$

orbit lives in $\mathbb{C}^m$

Algorithmic consequences





algorithms for finding MLE, e.g. iterative proportional scaling (IPS) $\leftrightarrow$ scaling algorithms to compute capacity

maximize likelihood $\Leftrightarrow$ minimize KL divergence

minimize ℓ_2-norm

model lives in $\Delta_{m-1} \cap \mathbb{R}^m_{>0}$

orbit lives in $\mathbb{C}^m$

trivial linearization $b = 0$
(defines model and steps of IPS)

linearization $b = AY$

Gaussian statistical models

The density function of an m-dimensional Gaussian with mean zero and covariance matrix $\Sigma \in \mathbb{R}^{m \times m}$ is

$$\rho_\Sigma(y) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\left(-\frac{1}{2} y^T \Sigma^{-1} y\right), \quad \text{where } y \in \mathbb{R}^m.$$

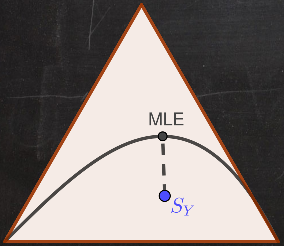
The **concentration matrix** $\Psi = \Sigma^{-1}$ is positive definite.

Gaussian statistical models

The density function of an m-dimensional Gaussian with mean zero and covariance matrix $\Sigma \in \mathbb{R}^{m \times m}$ is

$$\rho_\Sigma(y) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\left(-\frac{1}{2} y^T \Sigma^{-1} y\right), \qquad \text{where } y \in \mathbb{R}^m.$$

The **concentration matrix** $\Psi = \Sigma^{-1}$ is positive definite.
A **Gaussian model** $\mathcal{M}$ is a set of concentration matrices, i.e. a subset of the cone of $m \times m$ positive definite matrices.



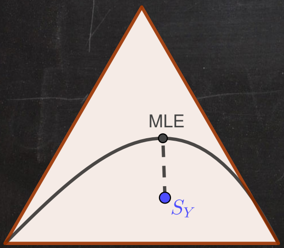
Gaussian statistical models

The density function of an m-dimensional Gaussian with mean zero and covariance matrix $\Sigma \in \mathbb{R}^{m\times m}$ is

$$\rho_\Sigma(y) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\left(-\frac{1}{2} y^T \Sigma^{-1} y\right), \quad \text{where } y \in \mathbb{R}^m.$$

The **concentration matrix** $\Psi = \Sigma^{-1}$ is positive definite.
A **Gaussian model** $\mathcal{M}$ is a set of concentration matrices, i.e. a subset of the cone of $m \times m$ positive definite matrices. Given data $Y = (Y_1, \ldots, Y_n)$, the likelihood is

$$L_Y(\Psi) = \rho_{\Psi^{-1}}(Y_1) \cdots \rho_{\Psi^{-1}}(Y_n), \quad \text{where } \Psi \in \mathcal{M}.$$



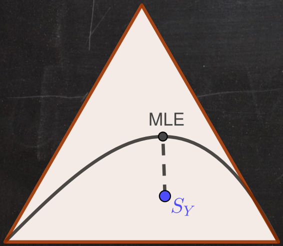
Gaussian statistical models

The density function of an m-dimensional Gaussian with mean zero and covariance matrix $\Sigma \in \mathbb{R}^{m \times m}$ is

$$\rho_\Sigma(y) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\left(-\frac{1}{2} y^T \Sigma^{-1} y\right), \quad \text{where } y \in \mathbb{R}^m.$$

The **concentration matrix** $\Psi = \Sigma^{-1}$ is positive definite.
A **Gaussian model** $\mathcal{M}$ is a set of concentration matrices, i.e. a subset of the cone of $m \times m$ positive definite matrices. Given data $Y = (Y_1, \ldots, Y_n)$, the likelihood is

$$L_Y(\Psi) = \rho_{\Psi^{-1}}(Y_1) \cdots \rho_{\Psi^{-1}}(Y_n), \quad \text{where } \Psi \in \mathcal{M}.$$



likelihood L_Y can be unbounded from above

MLE might not exist

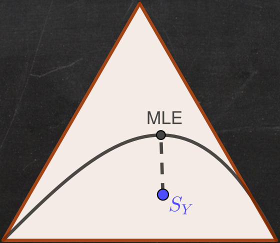
MLE might not be unique

Gaussian group model

The **Gaussian group model** of a group G with a representation $G \xrightarrow{\varphi} \mathrm{GL}_m$ on $\mathbb{R}^m$ is

$$\mathcal{M}_G := \left\{ \Psi_g = \varphi(g)^T \varphi(g) \mid g \in G \right\}.$$

(depends only on image of G in GL_m, hence may assume $G \subseteq \mathrm{GL}_m$)



Gaussian group model

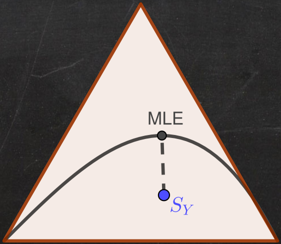
The **Gaussian group model** of a group G with a representation $G \xrightarrow{\varphi} \mathrm{GL}_m$ on $\mathbb{R}^m$ is

$$\mathcal{M}_G := \left\{ \Psi_g = \varphi(g)^T \varphi(g) \mid g \in G \right\}.$$

(depends only on image of G in GL_m, hence may assume $G \subseteq \mathrm{GL}_m$)

We want to find an MLE, i.e. a maximizer of

$$L_Y(\Psi_g)$$



Gaussian group model

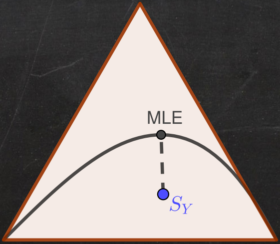
The **Gaussian group model** of a group G with a representation $G \xrightarrow{\varphi} \mathrm{GL}_m$ on $\mathbb{R}^m$ is

$$\mathcal{M}_G := \left\{ \Psi_g = \varphi(g)^T \varphi(g) \mid g \in G \right\}.$$

(depends only on image of G in GL_m, hence may assume $G \subseteq \mathrm{GL}_m$)

We want to find an MLE, i.e. a maximizer of

$$\log L_Y(\Psi_g) = \frac{1}{2} \underbrace{\left(n \log\det \Psi_g - \|g \cdot Y\|_2^2 \right)}_{\ell_Y(\Psi_g)} - \frac{nm}{2} \log(2\pi) \qquad \text{for } g \in G.$$



XIV - XVII

Combining both worlds

$$\sup_{g\in G} \ell_Y(\Psi_g) = -\inf_{\tau\in\mathbb{R}_{>0}} \left(\tau\left(\inf_{h\in G\cap \mathrm{SL_m}} \|h\cdot Y\|_2^2\right) - nm\log\tau\right).$$

Combining both worlds

Invariant theory classically over $\mathbb{C}$ – can also define Gaussian (group) models over $\mathbb{C}$

Proposition (Améndola, Kohn, Reichenbach, Seigal)
For $Y = (Y_1, \ldots, Y_n)$ with $Y_i \in \mathbb{C}^m$ and a group $G \subset \mathrm{GL}_m(\mathbb{C})$ closed under non-zero scalar multiples (i.e., $g \in G, \lambda \in \mathbb{C}, \lambda \neq 0 \Rightarrow \lambda g \in G$),

$$\sup_{g \in G} \ell_Y(\Psi_g) = - \inf_{\tau \in \mathbb{R}_{>0}} \left(\tau \left(\inf_{h \in G \cap \mathrm{SL_m}} \|h \cdot Y\|_2^2 \right) - nm \log \tau \right).$$

Combining both worlds

Invariant theory classically over $\mathbb{C}$ – can also define Gaussian (group) models over $\mathbb{C}$

Proposition (Améndola, Kohn, Reichenbach, Seigal)
For $Y = (Y_1, \ldots, Y_n)$ with $Y_i \in \mathbb{C}^m$ and a group $G \subset \mathrm{GL}_m(\mathbb{C})$ closed under non-zero scalar multiples (i.e., $g \in G, \lambda \in \mathbb{C}, \lambda \neq 0 \Rightarrow \lambda g \in G$),

$$\sup_{g \in G} \ell_Y(\Psi_g) = - \inf_{\tau \in \mathbb{R}_{>0}} \left(\tau \left(\inf_{h \in G \cap \mathrm{SL_m}} \|h \cdot Y\|_2^2 \right) - nm \log \tau \right).$$

If $h \cdot Y$ is a point of minimal norm in the $G \cap \mathrm{SL}_m$-orbit of Y, then an MLE for the Gaussian group model $\mathcal{M}_G$ is

$$\tau h^* h, \text{ where } \tau \text{ is the unique value minimizing } \tau \|h \cdot Y\|_2^2 - nm \log \tau.$$

Combining both worlds

Invariant theory classically over $\mathbb{C}$ – can also define Gaussian (group) models over $\mathbb{C}$

Proposition (Améndola, Kohn, Reichenbach, Seigal)
For $Y = (Y_1, \dots, Y_n)$ with $Y_i \in \mathbb{C}^m$ and a group $G \subset \mathrm{GL}_m(\mathbb{C})$ closed under non-zero scalar multiples (i.e., $g \in G, \lambda \in \mathbb{C}, \lambda \neq 0 \Rightarrow \lambda g \in G$),

$$\sup_{g \in G} \ell_Y(\Psi_g) = - \inf_{\tau \in \mathbb{R}_{>0}} \left(\tau \left(\inf_{h \in G \cap \mathrm{SL_m}} \|h \cdot Y\|_2^2 \right) - nm \log \tau \right).$$

If $h \cdot Y$ is a point of minimal norm in the $G \cap \mathrm{SL}_m$-orbit of Y, then an MLE for the Gaussian group model $\mathcal{M}_G$ is

$$\tau h^* h, \text{ where } \tau \text{ is the unique value minimizing } \tau \|h \cdot Y\|_2^2 - nm \log \tau.$$

Theorem (Améndola, Kohn, Reichenbach, Seigal)
Let Y and G as above. If G is linearly reductive,
ML estimation for $\mathcal{M}_G$ relates to the action by $G \cap \mathrm{SL}_m(\mathbb{C})$ as follows:

(a) Y unstable $\Leftrightarrow$ ℓ_Y not bounded from above
(b) Y semistable $\Leftrightarrow$ ℓ_Y bounded from above
(c) Y polystable $\Leftrightarrow$ MLE exists
(d) Y stable $\Leftrightarrow$ finitely many MLEs exist $\Leftrightarrow$ unique MLE

Combining both worlds

Real examples

Combining both worlds

Real examples

Theorem (Améndola, Kohn, Reichenbach, Seigal)
Let $Y = (Y_1, \ldots, Y_n)$ with $Y_i \in \mathbb{R}^m$, and let $G \subset \mathrm{GL}_m(\mathbb{R})$ be a linearly reductive group which is closed under non-zero scalar multiples.
ML estimation for $\mathcal{M}_G$ relates to the action by $G \cap \mathrm{SL}_m(\mathbb{R})$ as follows:

(a)	Y unstable	$\Leftrightarrow$	ℓ_Y not bounded from above		
(b)	Y semistable	$\Leftrightarrow$	ℓ_Y bounded from above		
(c)	Y polystable	$\Leftrightarrow$	MLE exists		
(d)	Y stable	$\Rightarrow$	finitely many MLEs exist	$\Leftrightarrow$	unique MLE

Combining both worlds

Real examples

Theorem (Améndola, Kohn, Reichenbach, Seigal)
Let $Y = (Y_1, \ldots, Y_n)$ with $Y_i \in \mathbb{R}^m$, and let $G \subset \mathrm{GL}_m(\mathbb{R})$ be a linearly reductive group which is closed under non-zero scalar multiples.
ML estimation for $\mathcal{M}_G$ relates to the action by $G \cap \mathrm{SL}_m(\mathbb{R})$ as follows:

(a) Y unstable $\Leftrightarrow$ ℓ_Y not bounded from above
(b) Y semistable $\Leftrightarrow$ ℓ_Y bounded from above
(c) Y polystable $\Leftrightarrow$ MLE exists
(d) Y stable $\Rightarrow$ finitely many MLEs exist $\Leftrightarrow$ unique MLE

Examples: **full Gaussian model, independence model, matrix normal model**

Combining both worlds

Real examples

Theorem (Améndola, Kohn, Reichenbach, Seigal)
Let $Y = (Y_1, \dots, Y_n)$ with $Y_i \in \mathbb{R}^m$, and let $G \subset \mathrm{GL}_m(\mathbb{R})$ be a linearly reductive group which is closed under non-zero scalar multiples.
ML estimation for $\mathcal{M}_G$ relates to the action by $G \cap \mathrm{SL}_m(\mathbb{R})$ as follows:

(a)	Y unstable	$\Leftrightarrow$	ℓ_Y not bounded from above			
(b)	Y semistable	$\Leftrightarrow$	ℓ_Y bounded from above			
(c)	Y polystable	$\Leftrightarrow$	MLE exists			
(d)	Y stable	$\Rightarrow$	finitely many MLEs exist	$\Leftrightarrow$	unique MLE	

Examples: **full Gaussian model, independence model, matrix normal model**

Theorem (Améndola, Kohn, Reichenbach, Seigal)
Let $Y = (Y_1, \dots, Y_n)$ with $Y_i \in \mathbb{R}^m$, and let $G \subset \mathrm{GL}_m(\mathbb{R})$ be a group which is closed under non-zero scalar multiples, but not necessarily linearly reductive.
ML estimation for $\mathcal{M}_G$ relates to the action by $G \cap \mathrm{SL}_m^{\pm}(\mathbb{R})$ as follows:

(a)	Y unstable	$\Leftrightarrow$	ℓ_Y not bounded from above
(b)	Y semistable	$\Leftrightarrow$	ℓ_Y bounded from above
(c)	Y polystable	$\Rightarrow$	MLE exists

Combining both worlds

Real examples

Theorem (Améndola, Kohn, Reichenbach, Seigal)
Let $Y = (Y_1, \ldots, Y_n)$ with $Y_i \in \mathbb{R}^m$, and let $G \subset \mathrm{GL}_m(\mathbb{R})$ be a linearly reductive group which is closed under non-zero scalar multiples.
ML estimation for $\mathcal{M}_G$ relates to the action by $G \cap \mathrm{SL}_m(\mathbb{R})$ as follows:

(a)	Y unstable	$\Leftrightarrow$	ℓ_Y not bounded from above		
(b)	Y semistable	$\Leftrightarrow$	ℓ_Y bounded from above		
(c)	Y polystable	$\Leftrightarrow$	MLE exists		
(d)	Y stable	$\Rightarrow$	finitely many MLEs exist	$\Leftrightarrow$	unique MLE

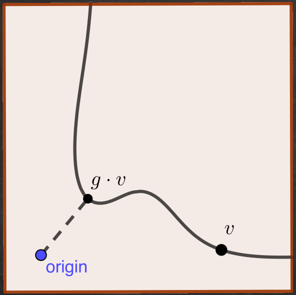
Examples: full Gaussian model, independence model, matrix normal model

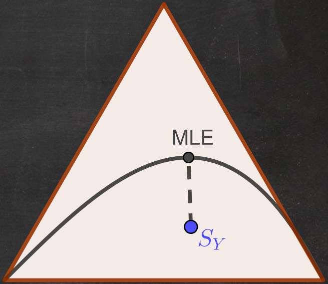
Theorem (Améndola, Kohn, Reichenbach, Seigal)
Let $Y = (Y_1, \ldots, Y_n)$ with $Y_i \in \mathbb{R}^m$, and let $G \subset \mathrm{GL}_m(\mathbb{R})$ be a group which is closed under non-zero scalar multiples, but not necessarily linearly reductive.
ML estimation for $\mathcal{M}_G$ relates to the action by $G \cap \mathrm{SL}_m^{\pm}(\mathbb{R})$ as follows:

(a)	Y unstable	$\Leftrightarrow$	ℓ_Y not bounded from above
(b)	Y semistable	$\Leftrightarrow$	ℓ_Y bounded from above
(c)	Y polystable	$\Rightarrow$	MLE exists

Example: Gaussian graphical models

Summary





	Invariant theory	**Statistics**
historical progression ↓	describe null cone	algorithms to find MLE
	algorithmic null cone membership testing	convergence analysis