invariant theory for maximum likelihood estimation

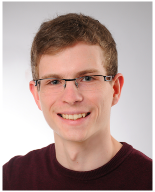
Anna Seigal
University of Oxford

joint work with

Carlos Améndola
TU Munich

Kathlén Kohn
KTH Stockholm

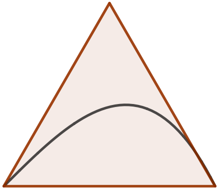
Philipp Reichenbach
TU Berlin



June 16, 2021

statistical setting

statistical model $\mathcal{M}$ in some space:



discrete: random variable X with m states

probability mass function: $p = (p_1, \dots, p_m)$, $p_j = \mathbb{P}(X = j)$

statistical model $\mathcal{M}$: subset of probability simplex $\Delta_{m-1} = \left\{ q \in \mathbb{R}^m \mid q_j \geq 0, \sum_{j=1}^m q_j = 1 \right\}$.

multivariate Gaussian: random vector $x \in \mathbb{R}^m$

probability density function: $\rho_\Psi(x) = \frac{1}{\sqrt{\det(2\pi\Psi^{-1})}} \exp\left(-\frac{1}{2}x^\mathsf{T}\Psi x\right)$, Ψ = concentration matrix

statistical model $\mathcal{M}$: subset of positive definite matrices

group actions

group: set G of elements such that $\mathrm{id} \in G, \quad gh \in G, \quad g^{-1} \in G$.
G acts linearly on space V means each $g \in G$ gives linear transformation of V.

Example

symmetric group → permutation matrices $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} v_2 \\ v_1 \\ v_3 \end{bmatrix}.$

orbit is $G \cdot v = \{g \cdot v \mid g \in G\}$ invariants are functions that are constant on orbits

Example

$v_1 + v_2 + v_3, \quad v_1 v_2 + v_1 v_3 + v_2 v_3, \quad v_1 v_2 v_3$
elementary symmetric polynomials generate ring of invariants.

groups in statistics

i.i.d. independent and identically distributed
$\rightarrow$ permuting observations doesn't change distribution
$\rightarrow$ invariant under group of permutation matrices

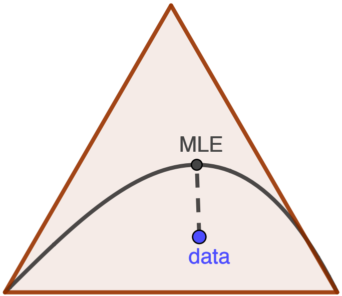
group symmetry models[1] [2]: covariance matrices invariant under some fixed group

transformation family[3]: statistical model on which a group acts transitively

[1] Steen Andersson. "Invariant normal models." Annals of Statistics (1975).
[2] Jesper Madsen. "Invariant normal models with recursive graphical Markov structure". Annals of Statistics (2000).
[3] Jan Draisma, Sonja Kuhnt, Piotr Zwiernik. "Groups acting on Gaussian graphical models." Annals of Statistics (2013).

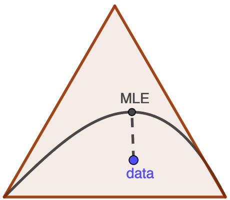
parameter estimation

have a statistical model $\mathcal{M}$ and some data.
want to use the data to estimate parameters in $\mathcal{M}$.



maximum likelihood estimate (MLE):
point in model that maximizes likelihood of observing data.

maximum likelihood estimation



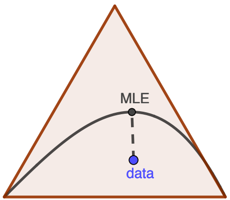
discrete: random variable with m states

- model: $\mathcal{M}$ a set of $p = (p_1, \ldots, p_m)$ in probability simplex Δ_{m-1}
- data: $u = (u_1, \ldots, u_m)$, u_j = fraction of times j occurs.
- likelihood: $L(p) = p_1^{u_1} \cdots p_m^{u_m}$, log-likelihood $\ell(p) = \sum_i u_i \log p_i$

$$MLE : \hat{p} = \underset{p \in \mathcal{M}}{\operatorname{argmax}}\, \ell(p).$$

Example:
$\mathcal{M} = \Delta_{m-1}$,
unique MLE is $\hat{p} = u$.

maximum likelihood estimation



multivariate Gaussian: random vector in $\mathbb{R}^m$

- model: $\mathcal{M}$ a set of possible concentration matrices Ψ in positive definite cone
- data: samples $Y_i \in \mathbb{R}^m$ summarized by sample covariance matrix $S_Y = \frac{1}{n}\sum_{i=1}^n Y_i Y_i^\mathsf{T}$
- log-likelihood: $\ell(\Psi) = \log\det(\Psi) - \mathrm{tr}(\Psi S_Y)$

$$MLE : \hat{\Psi} = \underset{\Psi \in \mathcal{M}}{\operatorname{argmax}}\, \ell(\Psi).$$

Example:
$\mathcal{M} = \mathrm{PD}_m$,
unique MLE $\hat{\Psi} = S_Y^{-1}$ ...does not exist if $\mathrm{rank}(S_Y) < m$

maximum likelihood thresholds

thresholds[4]: how many samples needed for:
(i) log-likelihood to be bounded
(ii) MLE to exist
(iii) MLE to be unique.

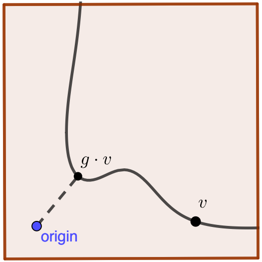
Example: $\mathcal{M} = \mathrm{PD}_m$, unique MLE $\hat{\Psi} = S_Y^{-1}$ $\quad$ $S_Y = \frac{1}{n}\sum_{i=1}^{n} Y_i Y_i^{\mathsf{T}}$
...thresholds (i), (ii), (iii) are all m.

- for $n \geq m$ samples, the MLE generically exists and is unique
- for $n < m$ samples the log-likelihood is unbounded

smaller models need fewer samples
e.g. for a graphical model, thresholds given in terms of graph structure[5]
recently: connections to rigidity theory

[4] Buhl, S. L. 'On the existence of maximum likelihood estimators for graphical Gaussian models' (1993)
[5] Caroline Uhler: Geometry of maximum likelihood estimation in Gaussian graphical models (2012)

stability



group G acts linearly on space V.

orbit is $G \cdot v = \{g \cdot v \mid g \in G\}$,

capacity is $\mathrm{cap}(v) = \inf_{g \in G} \|g \cdot v\|^2$.

Definition

(a) *unstable* if $\mathrm{cap}(v) = 0$ [unstable points $=$ *null cone*, where all invariants vanish]

(b) *semistable* if $\mathrm{cap}(v) > 0$

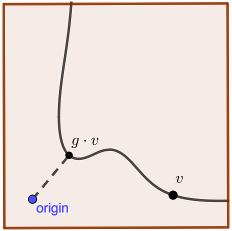
(c) *polystable* if $v \neq 0$ and $G \cdot v$ is closed

(d) *stable* if v polystable and stabilizer $\{g \in G \mid g \cdot v = v\}$ is finite

Examples: finite group, GL_m on $\mathbb{C}^m$, $\mathrm{SL}_{m_1} \times \mathrm{SL}_{m_2}$ on $\mathbb{C}^{m_1 \times m_2}$.

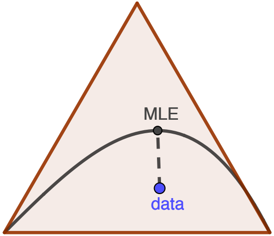
invariant theory and maximum likelihood estimation[6] [7]

invariant theory



$$\left\{ \begin{array}{c} \text{unstable} \\ \text{semistable} \\ \text{polystable} \\ \text{stable} \end{array} \right\}$$

statistics



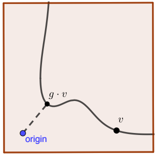
$$\left\{ \begin{array}{c} \text{likelihood unbounded from above} \\ \text{likelihood bounded from above} \\ \text{MLE exists} \\ \text{MLE exists uniquely} \end{array} \right\}$$

[6] C. Améndola, K. Kohn, P. Reichenbach, AS, Invariant theory and scaling algorithms for maximum likelihood estimation, to appear in SIAM Journal on Applied Algebra and Geometry (2021).

[7] ——–, Toric invariant theory for maximum likelihood estimation in log-linear models, arXiv:2012.07793 (2020)

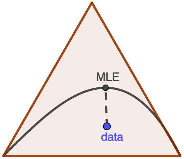
stability and MLE existence

invariant theory



torus action by matrix A

statistics



log-linear model $\mathcal{M}_A$

unstable		does not happen
semistable	$\Leftrightarrow$	extended MLE exists and is unique
polystable	$\Leftrightarrow$	MLE exists and is unique
stable		does not happen

matrix $A \in \mathbb{Z}^{d \times m}$ with $(1, \ldots, 1) \in \mathbb{C}^m$ in rowspan. data $u \in \mathbb{Z}^m_{\geq 0}$ with $u_1 + \cdots + u_m = n$.
stability under complex torus GT_d, action given by matrix nA with linearization $b = Au \in \mathbb{Z}^d$.

Proof idea. polyhedral conditions for MLE existence[8] relate to Hilbert-Mumford criterion.

[8] Eriksson N, Fienberg SE, Rinaldo A, Sullivant S: "Polyhedral conditions for the nonexistence of the MLE for hierarchical log-linear models." (2006).

log-linear models

log-linear models: distributions whose logarithm lies in fixed linear space.

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

where $A \in \mathbb{Z}^{d \times m}$ and we assume $(1, \ldots, 1) \in \mathrm{rowspan}(A)$.

e.g. independence model, graphical models, hierarchical models...

$$A = \begin{array}{c} \\ \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \mu_1 \\ \mu_2 \\ \mu_3 \end{array} \begin{array}{c} \begin{array}{ccc|ccc|ccc} p_{11} & p_{21} & p_{31} & p_{12} & p_{22} & p_{32} & p_{13} & p_{23} & p_{33} \end{array} \\ \left(\begin{array}{ccc|ccc|ccc} 1 & & & 1 & & & 1 & & \\ & 1 & & & 1 & & & 1 & \\ & & 1 & & & 1 & & & 1 \\ \hline 1 & 1 & 1 & & & & & & \\ & & & 1 & 1 & 1 & & & \\ & & & & & & 1 & 1 & 1 \end{array}\right) \end{array} \in \mathbb{Z}^{6 \times 9}$$

$$\log(p_{ij}) = \lambda_i + \mu_j$$

log-linear models and torus actions

log-linear model $\mathcal{M}_A = \{p \in \Delta_{m-1} \mid \log p \in \mathrm{rowspan}(A)\}$
parametrise $\mathcal{M}_A$:

$$\begin{aligned} \mathbb{R}^d_{>0} &\longrightarrow \Delta_{m-1} \\ \theta &\longmapsto \frac{1}{Z(\theta)}\theta^{a_j}, \qquad \theta^{a_j} = \theta_1^{a_{1j}} \cdots \theta_d^{a_{dj}}, \qquad Z(\theta) \text{ normalisation constant.} \end{aligned}$$

torus action: GT_d = complex, diagonal, invertible $d \times d$ matrices.
matrix $A \in \mathbb{Z}^{d\times m}$ gives action on $\mathbb{P}^{m-1}$: multiply by

$$\begin{bmatrix} \theta^{a_1} & & \\ & \ddots & \\ & & \theta^{a_m} \end{bmatrix}$$

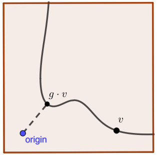
linearisation: action on $\mathbb{C}^m$ given by $b \in \mathbb{Z}^d$. multiply by

$$\begin{bmatrix} \theta^{a_1-b} & & \\ & \ddots & \\ & & \theta^{a_m-b} \end{bmatrix}$$

Gaussian group models

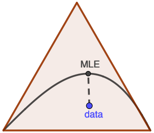
given a group G of $m \times m$ real matrices, Gaussian group model is $\mathcal{M}_G = \{g^{\mathsf{T}} g \mid g \in G\}$.

invariant theory



group G

statistics



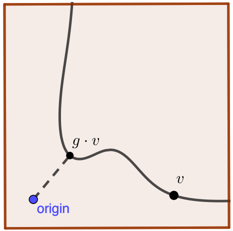
Gaussian group model $\mathcal{M}_G$

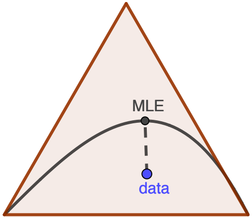
unstable	$\Leftrightarrow$	likelihood unbounded from above		
semistable	$\Leftrightarrow$	likelihood bounded from above		
polystable	$\Leftrightarrow$	MLE exists		
stable	$\Rightarrow$	finitely many MLEs	$\Leftrightarrow$	unique MLE

stability under $G \cap \mathrm{SL}_m$.   for complex G, get equivalence of all four conditions.

Proof idea. log-likelihood is $\log\det(\Psi) - \mathrm{tr}(\Psi S_Y) = \log\det(g^{\mathsf{T}} g) - \frac{1}{n} \|g \cdot Y\|^2$.

find MLE using group



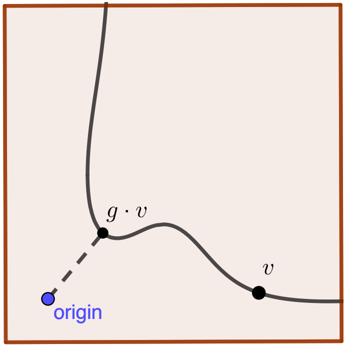


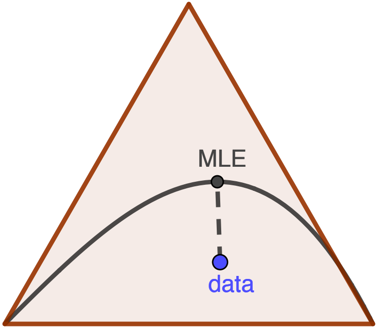
the point of minimal norm in the orbit gives the MLE.

- discrete: given $v \in \mathbb{C}^m$, MLE is $\hat{p}_i = \frac{|v_i|^2}{\|v\|^2}$
- Gaussian: given $g \in G$, MLE is $\hat{\Psi} = g^{\mathsf{T}} g$ (up to scale)

can use geodesically convex algorithms for norm minimization on orbit[9] to find MLE

[9] Bürgisser P, Franks C, Garg A, Oliveira R, Walter M, Wigderson A. Towards a theory of non-commutative optimization: Geodesic 1st and 2nd order methods for moment maps and polytopes (2019).





Thank you!