MLE for tensor normal models

Visu Makam
Institute for Advanced Study and University of Melbourne

joint with Harm Derksen and Michael Walter

Statistical models

- A statistical model $\mathcal{M}$ is a collection of probability distributions on (say) $\mathbb{C}^d$.
- Given some empirical data, want to recover the "most likely" distribution from our statistical model.
- How many samples do you need for this "most likely" distribution to be unique (almost surely)?
- Almost surely $=$ data not pathological.

Likelihood function

- We consider Gaussian distributions on $\mathbb{C}^d$ (with mean 0)
- For a concentration matrix $\Psi \in \mathrm{PD}_d$, the probability density function is given by

$$f_\Psi(Y) = \det(\Psi/\pi)e^{-Y^\dagger \Psi Y}$$

- Suppose we have m samples $Y_1, \ldots, Y_m$. Each $Y_i \in \mathbb{C}^d$. The likelihood function is

$$L_Y(\Psi) = \prod_{i=1}^{m} f_\Psi(Y_i).$$

- A Gaussian model $\mathcal{M}$ is just a subset of PD_d.
- A maximum likelihood estimate (MLE) is a $\Psi \in \mathcal{M}$ which maximizes the likelihood function.

Gaussian group models

Let $G \subseteq \mathrm{GL}_d$ (i.e., G acts on $\mathbb{C}^d$) be a Zariski-closed subgroup that is closed under adjoint and scalar multiples.

We define the *Gaussian group model*

$$\mathcal{M}_G = \{g^\dagger g \mid g \in G\} \subseteq \mathrm{PD}_d.$$

Suppose we have m-samples $Y = (Y_1, \dots, Y_m) \in (\mathbb{C}^d)^m$.

Theorem (Améndola, Kohn, Reichenbach, Seigal 2020)

Let $G_{\mathrm{SL}} = G \cap \mathrm{SL}_d$. Then

- *L_Y is bounded if and only if Y is G_{SL}-semistable.*
- *An MLE exists if and only if Y is G_{SL}-polystable.*
- *An MLE exists uniquely if and only if Y is G_{SL}-stable.*

Warning: Over $\mathbb{R}$, MLE exists uniquely does not mean Y is G_{SL}-stable.

Invariant theory notions

Suppose an algebraic group G acts on $\mathbb{C}^n$, i.e., $\rho : G \to \mathrm{GL}_n$. Then $v \in \mathbb{C}^n$ is called

- unstable if $0 \in \overline{G \cdot v}$;
- semistable if $0 \notin \overline{G \cdot v}$;
- polystable if $v \neq 0$ and $G \cdot v$ is closed;
- stable if polystable and finite stabilizer (modulo Kernel of ρ).

Closure is in either Zariski or Euclidean topology.

When we work over $\mathbb{R}$, the closure is in Euclidean topology.

Sample size thresholds

Let $\mathcal{M}_G \subseteq \mathrm{PD}_d$ be a Gaussian group model.

Suppose we have m samples $Y = (Y_1, \ldots, Y_m) \in (\mathbb{C}^d)^m$

Problem (Sample size thresholds)

Do we have (almost surely)

- *boundedness of log-likelihood function;*
- *existence of MLE;*
- *existence of a unique MLE.*

This is equivalent to asking if for the action of G_{SL} on $(\mathbb{C}^d)^m$, we have

- generic semistability;
- generic polystability;
- generic stability (Over $\mathbb{R}$, this is not quite true).

Tensor normal models

Let

$$V(\mathbf{d}) = \mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2} \otimes \cdots \otimes \mathbb{C}^{d_k}.$$

Consider the group $G = \mathrm{GL}_{d_1} \times \mathrm{GL}_{d_2} \times \cdots \times \mathrm{GL}_{d_k} \subseteq \mathrm{GL}(V(\mathbf{d}))$.

Concretely

$$G = \{g_1 \otimes g_2 \otimes \cdots \otimes g_k \mid g_i \in \mathrm{GL}_{d_i}\}.$$

The Gaussian group model $\mathcal{M}_G$ is called a tensor normal model

We can take

$$G_{\mathrm{SL}} = \mathrm{SL}_{d_1} \times \mathrm{SL}_{d_2} \times \cdots \times \mathrm{SL}_{d_k}.$$

A purely algebro-geometric question

Problem

Given m *and* $\mathbf{d} = (d_1, \dots, d_k)$*, for the action of* G_{SL} *on* $V(\mathbf{d})^{\oplus m}$*, do we have generic semistability/polystability/stability?*

For example, is the $\mathrm{SL}_n \times \mathrm{SL}_n \times \mathrm{SL}_n$-orbit of a generic tensor in $\mathbb{C}^n \otimes \mathbb{C}^n \otimes \mathbb{C}^n$ closed?

An easy example

Suppose $m = 2$ and $\mathbf{d} = (2, 3, 5, 176)$.

So, we consider the action of

$$G_{\mathrm{SL}} = \mathrm{SL}_2 \times \mathrm{SL}_3 \times \mathrm{SL}_5 \times \mathrm{SL}_{176}$$

on

$$V(\mathbf{d})^{\oplus 2} = (\mathbb{C}^2 \otimes \mathbb{C}^3 \otimes \mathbb{C}^5 \otimes \mathbb{C}^{176})^{\oplus 2}$$

Just viewing as an SL_{176} representation, we get $V(\mathbf{d})^{\oplus 2} = (\mathbb{C}^{176})^{\oplus 60}$.

Easy fact: For the action of $G = \mathrm{SL}(W)$ on W, every $0 \in \overline{G \cdot w}$ for all $w \in W$, i.e., all points are unstable.

Fact: Same is true for the action of $G = \mathrm{SL}(W)$ on W^m if $m < \dim(W)$.

Hence $V(\mathbf{d})^{\oplus 2}$ generically unstable for SL_{176} action, and hence for G_{SL} action as well.

If $md_1 d_2 \dots d_{k-1} < d_k$, then generically unstable.

Some more examples:

Suppose $m = 1$,

- $\mathbf{d} = (2, 3, 1)$ – generically unstable
- $\mathbf{d} = (2, 3, 2)$ – generically polystable, but not generically stable
- $\mathbf{d} = (2, 3, 3)$ – generically polystable, but not generically stable
- $\mathbf{d} = (2, 3, 4)$ – generically polystable, but not generically stable
- $\mathbf{d} = (2, 3, 5)$ – generically unstable
- $\mathbf{d} = (2, 3, 6)$ – generically polystable, but not generically stable
- $\mathbf{d} = (2, 3, 7)$ – generically unstable

Two step strategy:

- The property of being generically semistable/polystable/stable is invariant under castling transforms [Sato, Kimura, Elashvili]
- Understand the minimal elements in each castling equivalence class.

Castling transforms

Suppose H acts on W and $\dim W = n$.

We have an action

$$H \times \mathrm{SL}_k \circlearrowright W \otimes \mathbb{C}^k$$

Suppose $0 < k < n$, we also have

$$H \times \mathrm{SL}_{n-k} \circlearrowright W^* \otimes \mathbb{C}^{n-k}.$$

$H \times \mathrm{SL}_k$-orbits on $W \otimes \mathbb{C}^k \leftrightarrow H \times \mathrm{SL}_{n-k}$-orbits on $W^* \otimes \mathbb{C}^{n-k}$.

Stabilizers are preserved by this correspondence

Applying to tensors

We have m and $\mathbf{d} = (d_1, \dots, d_k)$. Want to understand action of $G(\mathbf{d}) = \prod_{i=1}^{k} \mathrm{SL}_{d_i}$ on $V(\mathbf{d})^{\oplus m} = (\mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_k})^{\oplus m}$

Suppose $d_k < N := m d_1 d_2 \cdots d_{k-1}$.

Take $H = \prod_{i=1}^{k-1} \mathrm{SL}_{d_i}$ and take $W = (\mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2} \otimes \cdots \otimes \mathbb{C}^{d_{k-1}})^{\oplus m}$.

The action we are interested in:

$$H \times \mathrm{SL}_{d_k} \circlearrowright W \otimes \mathbb{C}^{d_k}.$$

Its castling transform:

$$H \times \mathrm{SL}_{N-d_k} \circlearrowright W \otimes \mathbb{C}^{N-d_k}.$$

$$(m, (d_1, \dots, d_k)) \longleftrightarrow (m, (d_1, \dots, d_{k-1}, N - d_k)).$$

Example

Consider

$$m = 1, \mathbf{d} = (2,3,4)$$

$N = 1 \cdot 2 \cdot 3 = 6 > d_k = 4$

$$(2,3,4) \longleftrightarrow (2,3,2) = (2,2,3)$$

$$(2,2,3) \longleftrightarrow (2,2,1) = (2,2)$$

Hence, we are down to looking at $\mathrm{SL}_2 \times \mathrm{SL}_2$ acting on $\mathrm{Mat}_{2,2}$ by left-right action.

Easy to check that it is generically polystable but not generically stable.

Warning: $(1,(2,3,4)) \equiv (1,(2,2))$ but $(2,(2,3,4)) \not\equiv (2,(2,2))$.

Minimal Castling equivalence classes

Consider $(m, \mathbf{d})$ (w.l.o.g., assume $d_1 \leq d_2 \leq \cdots \leq d_k$). Let $N = md_1 d_2 \ldots d_{k-1}$.

Castling: $(m, (d_1, \ldots, d_k)) \longleftrightarrow (m, (d_1, \ldots, d_{k-1}, N - d_k))$

Then $(m, \mathbf{d})$ is minimal if one of the following holds:

1. $d_k > N$
2. $d_k = N$
3. $d_k \leq \frac{1}{2} N$

We already argued (1) in example → generically unstable

Case (2) is generically polystable → easy to compute stabilizer in general position: you get $\mathrm{SL}_{d_1} \times \mathrm{SL}_{d_2} \times \cdots \times \mathrm{SL}_{d_{k-1}}$.

Case (3) is generically stable with a couple of exceptions. Deduce from results of Andreev, Elashvili and Vinberg.

Final answer:

$$R(d_1, \dots, d_k; m) := m \prod_{i=1}^{k} d_i + \sum_{n=1}^{k} (-1)^n \sum_{1 \le i_1 < \dots < i_n \le k} \gcd(d_{i_1}, \dots, d_{i_n})^2$$

$$\operatorname{gmax}(d_1, \dots, d_k) = \max_{i<j} \gcd(d_i, d_j)$$

$$\Delta(d_1, \dots, d_k) = m d_1 \cdots d_k - 1 - \sum_{i=1}^{k} (d_i^2 - 1).$$

Theorem (Derksen, M. , Walter)

1. *If $R > 0$, then almost surely an MLE exists. Further,*
    - *$m \ge 2$: In this case, we always have $R \ge \operatorname{gmax}^2$ and almost surely an MLE exists uniquely if and only if $R > \operatorname{gmax}^2$ or $\operatorname{gmax} = 1$.*
    - *$m = 1$: In this case, we always have $\Delta \ge -2$ and almost surely an MLE exists uniquely if and only if $\Delta \ge -1$.*
2. *If $R = 0$, then almost surely an MLE exists. Further, almost surely an MLE exists uniquely if and only if $\operatorname{gmax} = 1$.*
3. *If $R < 0$, then likelihood function is unbounded.*