Moment and Cumulant Tensor Varieties of Gaussian Mixtures

Carlos Enrique Améndola Cerón
(TU Berlin)

AGATES: Tensors in statistics, optimization and machine learning

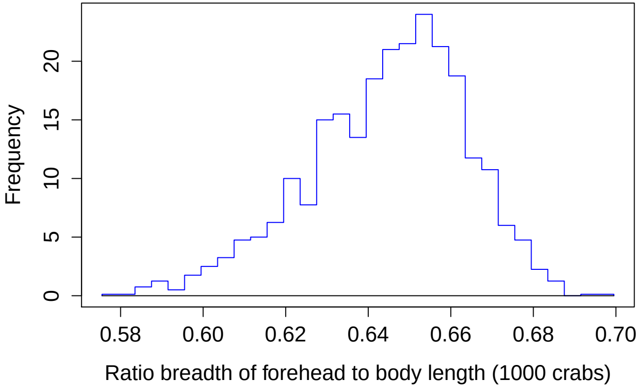
November 22, 2022

Some References / Personal History

- K. Pearson (1894), *Contributions to the Mathematical Theory of Evolution*, Philosophical Transactions of the Royal Society of London, 71-110.
- C. Améndola, M. Drton and B. Sturmfels (2015), *Maximum likelihood estimates for Gaussian mixtures are transcendental*, MACIS, Berlin, 579-590.
- C. Améndola, J.C. Faugère and B. Sturmfels (2016), *Moment Varieties of Gaussian Mixtures*, JAS 7, 14-28.
- C. Améndola (2017), *Algebraic Statistics of Gaussian Mixtures*, PhD Thesis, TU Berlin.
- C. Améndola, K. Ranestad and B. Sturmfels (2018), *Algebraic Identifiability of Gaussian Mixtures*, IMRN, 21, 6556-6580.
- D. Agostini, C. Améndola, K. Ranestad (2019), *Moment Identifiability of Homoscedastic Gaussian Mixtures*, FoCM 21, 695-724.
- C. Améndola, J. Lindberg, J. Rodriguez (2020), *Solving parametrized polynomial systems with decomposable projections*, arXiv: 1612.08807.
- J. Lindberg, C. Améndola, J. Rodriguez (2021), *Estimating Gaussian mixtures using sparse polynomial moment systems*, arXiv: 2106.15675

Origins

In 1893, the zoologist Raphael Weldon at University College London came to his stats colleague Karl Pearson with a histogram of measurements taken from a population of female Naples' crabs.



Origins

Unlike other measurements taken from the same population, the distribution did not appear to be Gaussian, but asymmetrical and possibly bimodal.

Pearson and Weldon reasoned this might be due to a *nonhomogeneous* sample where the individuals come from two distinct subpopulations, and they were excited to think they could be observing *evolution* unfold!

Multivariate Gaussian density in $\mathbb{R}^n$

$$f_X(x) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \mathrm{e}^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$$

for $x, \mu \in \mathbb{R}^n$ and Σ symmetric positive definite $n \times n$ matrix.
Parameters

μ: mean vector
Σ: covariance matrix

Gaussian Mixture Models

A mixture density f is a convex combination of densities $f_{X_1}, f_{X_2}, \dots, f_{X_k}$:

$$f_X(x) = \lambda_1 f_{X_1}(x) + \cdots + \lambda_k f_{X_k}(x)$$

where $\lambda_i \geq 0$ and $\lambda_1 + \cdots + \lambda_k = 1$.

Model for crabs data: mixture of two univariate Gaussians ($k = 2$).

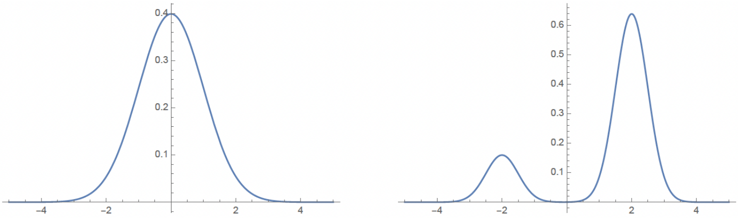


Figure: $\mathcal{N}(0, 1)$ density (left) and $0.2\mathcal{N}(-2, 0.5) + 0.8\mathcal{N}(2, 0.5)$ density (right).

Univariate Gaussian Mixture Model

If $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$, then the mixture density is

$$f_{\lambda,\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}} \left[\frac{\lambda}{\sigma_1} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2} \right) + \frac{1-\lambda}{\sigma_2} \exp\left(-\frac{(x-\mu_2)^2}{2\sigma_2^2} \right) \right]$$

with 5 parameters:
$\mu_1, \mu_2 \in \mathbb{R} \quad \sigma_1 > 0, \sigma_2 > 0, \quad \lambda \in [0,1]$.

Note: To yield an honest mixture: $0 < \lambda < 1$, $(\mu_1, \sigma_1) \neq (\mu_2, \sigma_2)$

Key Estimation problem

Given samples $y_1, \dots, y_N$ coming from a mixture of k Gaussians, how to recover parameters $\mu_i, \sigma_i^2, \lambda_i$?

Pearson's approach

How to find the unknown parameters? **Method of Moments**

Moment of *order* j of density $f(x)$:

$$m_j = \mathbb{E}[X^j] = \int_{\mathbb{R}} x^j f(x) dx$$

and sample moments

$$\hat{m}_j = \frac{1}{N} \sum_{i=1}^{N} y_i^j$$

The method of moments (MOM) matches sample moments to density's moments. This is a system of equations in the parameters. Nice property:

"The moments of a mixture are the mixture of the moments"

Moments as Polynomials

The first moments of a Gaussian $N(\mu, \sigma^2)$ are

$$\begin{aligned} m_1 &= \mu \\ m_2 &= \mu^2 + \sigma^2 \\ m_3 &= \mu^3 + 3\mu\sigma^2 \\ m_4 &= \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4 \\ m_5 &= \mu^5 + 10\mu^3\sigma^2 + 15\mu\sigma^4 \\ m_6 &= \mu^6 + 15\mu^4\sigma^2 + 45\mu^2\sigma^4 + 15\sigma^6. \end{aligned}$$

In general, m_j is a homogeneous polynomial of degree j in $\mathbb{Z}[\mu, \sigma]$.

Key observation:

Moments up to order d define a *Gaussian moment surface* $\mathcal{G}_{1,d} \subset \mathbb{R}^d$.

Furthermore,

The Gaussian mixture moments up to order d belong to the *secant variety* $\mathrm{Sec}_k(\mathcal{G}_{1,d}) \subset \mathbb{R}^d$.

Method of Moments

How many equations do we need? A mixture of k univariate Gaussians has

$$\dim \Theta_{1,k} = k + k + k - 1 \, = \, 3k - 1$$

independent parameters, so we need at least $d = 3k - 1$ moments.
Moment system of equations for $n = 1, k = 2$:

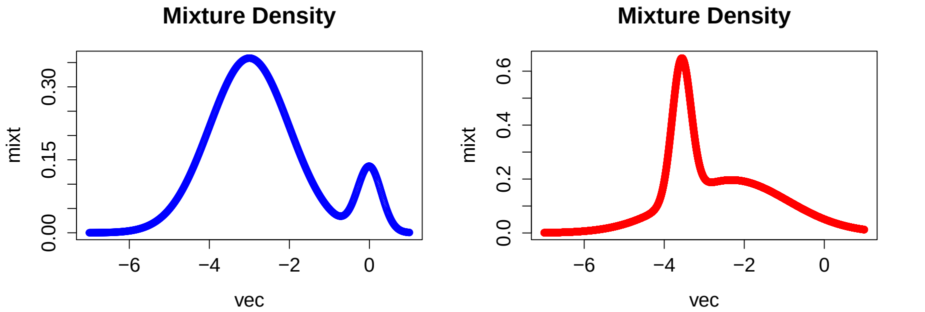
$$\begin{aligned}
\lambda\mu_1 + (1-\lambda)\mu_2 &= m_1 \\
\lambda(\mu_1^2 + \sigma_1^2) + (1-\lambda)(\mu_2^2 + \sigma_2^2) &= m_2 \\
\lambda(\mu_1^3 + 3\mu_1\sigma_1^2) + (1-\lambda)(\mu_2^3 + 3\mu_2\sigma_2^2) &= m_3 \\
\lambda(\mu_1^4 + 6\mu_1^2\sigma_1^2 + 3\sigma_1^4) + (1-\lambda)(\mu_2^4 + 6\mu_2^2\sigma_2^2 + 3\sigma_2^4) &= m_4 \\
\lambda(\mu_1^5 + 10\mu_1^3\sigma_1^2 + 15\mu_1\sigma_1^4) + (1-\lambda)(\mu_2^5 + 10\mu_2^3\sigma_2^2 + 15\mu_2\sigma_2^4) &= m_5
\end{aligned}$$

How to solve the system? Can it even be solved?

Algebraic Identifiability: Does the system have finitely many solutions?

Rational Identifiability: Does the system have a unique solution?

Identifiability from Moments



$0.9N(-3,1) + 0.1N(0,0.3)$ $\quad$ $0.7N(-2.32,2) + 0.3N(-3.57,0.06)$

Both have:

$$m_1 = -2.7,\ \ m_2 = 9.03,\ \ m_3 = -32.4,\ \ m_4 = 124.227,\ \ m_5 = -502.2$$

Pearson's Computations

We may assume the data is centered: $m_1 = 0$.
Pearson applies cleverly a series of variable substitutions and eliminations, until he obtains a single equation in the variable $x = \mu_1\mu_2$.

Pearson's Polynomial

$$24x^9 - 28\kappa_4 x^7 + 36m_3^2 x^6 - (24m_3\kappa_5 - 10\kappa_4^2)x^5 -$$
$$(148m_3^2\kappa_4 + 2\kappa_5^2)x^4 + (288m_3^4 - 12\kappa_4\kappa_5 m_3 - \kappa_4^3)x^3$$
$$+(24m_3^3\kappa_5 - 7m_3^2\kappa_4^2)x^2 + 32m_3^4\kappa_4 x - 24m_3^6 = 0$$

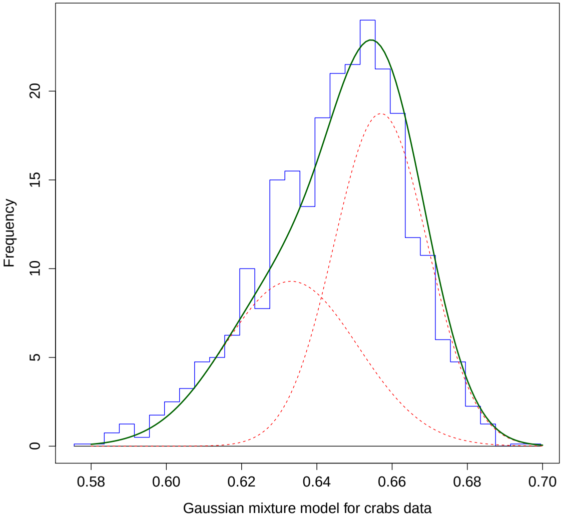
where $\kappa_4 = 9m_2^2 - 3m_4$ and $\kappa_5 = 30m_2m_3 - 3m_5$ (multiples of cumulants).

From which one can recover $y = \mu_1 + \mu_2$, and from there σ_1, σ_2 and λ.

For the original crabs data, one has

$m_1 = 16.799$, $m_2 = 304.923$, $m_3 = 5831.759$, $m_4 = 116061.435$, $m_5 = 2385609.719$

Pearson's Solution



Reaction to Pearson's MOM

"...the analytical difficulties, even for the case of $k = 2$*, are so considerable, that it may be questioned whether the general theory could ever be applied in practice to any numerical case."*- K. Pearson (1894)

"The solution of an equation of the ninth degree, where almost all powers, to the ninth, of the unknown quantity are existing, is, however, a very laborious task. Mr. Pearson has indeed possessed the energy to perform this heroic task in some instances...but I fear he will have few successors,..."
- Charlier (1906)

Alternative: Ronald Fisher introduces the method of *Maximum Likelihood*.
The *Expectation-Maximization* (EM) algorithm easily applies.

However, many problems arise... (likelihood function can even be *unbounded!*)

Reaching Transcendence

MLEs for Gaussian mixture models are transcendental functions of the data!

Theorem [Am., Drton, Sturmfels (2016)]

There exist samples $y_1, y_2, \ldots, y_N$ in $\mathbb{Q}^n$ whose MLEs for the mixture of two Gaussians are not algebraic numbers over $\mathbb{Q}$. In fact, there exist rational samples such that every non-trivial critical point $(\hat{\alpha}, \hat{\mu}_1, \hat{\mu}_2, \hat{\Sigma}_1, \hat{\Sigma}_2)$ of the log-likelihood function ℓ has at least one transcendental coordinate.

As a consequence of the theorem, there is no notion of ML degree for the Gaussian Mixture model.

Problem

Does there exist a universal bound on the number of non-trivial critical points for the log-likelihood function of the mixture of two Gaussians?

Search for ML Upper Bound

Theorem [Am., Drton, Sturmfels (2016)]

There exist sequences of samples on $\mathbb{R}$ such that the number of non-trivial critical points for the log-likelihood function of the mixture of two Gaussians increases beyond any bound.

- Even *worse*: Srebro's conjecture that there are no local maxima that are not global in the infinite sample limit when sampling from a Gaussian mixture distribution is *false* [Chi, Zhang, Balakrishnan, Wainwright, Jordan (2016)]
- Alternative: go back to MOM. Recent revival in Computer Science and Machine Learning by [Belkin, Sinha (2010)], [Moitra, Kalai, Valiant (2012)], [Hardt, Price (2015)], [Hsu, Huang, Ge, Kakade (2015)], ...
- But essential feasibility of the MOM approach sometimes overlooked!

Moment Identifiability

- Moment Problem: Does a sequence of moments $m_1, m_2, \ldots$ specify uniquely a probability distribution? Yes (under mild conditions)
- Do the moments up to some finite order D: $m_1, m_2, \ldots, m_D$ specify uniquely a mixture of Gaussians? Yes! *Hilbert's Basis Theorem*
- How many moments do we need to algebraically identify a mixture of Gaussians? (map from the model parameters to the moments is generically finite-to-one).
- What is the minimal D for rational identifiability?
- For $n = 1, k = 2$, then $d = 5$ is enough for algebraic identifiability [Pearson (1894)] and $d = 6$ is enough for rational identifiability [Lazard (2004)]
- Thus, for any given $m_1, m_2, m_3, m_4, m_5, m_6$, there is at most one mixture of two Gaussians with such moments.
- For $n = 1$, rational identifiability is achieved for $D \leq 4k - 2$ [Kalai, Moitra, Valiant (2012)]

Geometric Interpretation of Identifiability

- The parameter space $\Theta_{n,k}$ of a mixture of k Gaussians in $\mathbb{R}^n$ has

$$\dim \Theta_{n,k} = k \cdot n + k \cdot \frac{n(n+1)}{2} + k - 1$$

- On the other hand, there are $\binom{n+d}{d} - 1$ nontrivial moments up to order d, and they give a *polynomial* map to the moment space

$$M : \Theta_{n,k} \longrightarrow \mathbb{A}^M_{n,d} := \mathbb{R}^{\binom{n+d}{d}-1}$$

- For $k = 1$, this is precisely the *Gaussian moment variety* $\mathcal{G}_{n,d}$:

$$M : \Theta_{n,1} \longrightarrow \mathcal{G}_{n,d} \subset \mathbb{A}^M_{n,d}$$

- Moments of k mixtures correspond to the k-th secant of this variety:

$$M(\Theta_{n,k}) = \mathrm{Sec}_k(\mathcal{G}_{n,d})$$

  This is the (closure of) the union of all the linear spaces spanned by collections of k points on $\mathcal{G}_{n,d}$.
- In other words, $\mathrm{Sec}_k(\mathcal{G}_{n,d})$ are *moment varieties of Gaussian mixtures*.

Moment Identifiability and Defectivity

- Moment map

$$M : \Theta_{n,k} \longrightarrow \mathbb{A}^M_{n,d} := \mathbb{R}^{\binom{n+d}{d}-1}$$

- The method of moments consists exactly in computing the fibers of the map M. We have generic *algebraic identifiability* precisely when

$$\dim M(\Theta_{n,k}) = \dim \Theta_{n,k}$$

- Secant variety $\mathrm{Sec}_k(\mathcal{G}_{n,d}) = M(\Theta_{n,k})$
- The following bound always holds

$$\dim \mathrm{Sec}_k(\mathcal{G}_{n,d}) \leq \min \left\{ \dim \Theta_{n,k} \, , \, \dim \mathbb{A}^M_{n,d} \right\}$$

- When the inequality is strict, the variety $\mathcal{G}_{n,d}$ is k-*defective*.
- Study of defectivity is a classical topic in algebraic geometry!

Case $n = 1, k \geq 3$

Theorem [Am., Ranestad, Sturmfels (2017)]

"The mixture model of k univariate Gaussians is algebraically identifiable by the moments of order $\leq 3k - 1$."

Proof.

We prove that the *Gaussian moment surface* $\mathcal{G}_{1,d}$ is never *defective*, that is, all of its secant varieties $\mathrm{Sec}_k(\mathcal{G}_{1,d})$ always have the expected dimension. □

Conjecture

"This model is rationally identifiable by the moments of order $\leq 3k$."

A bit more on the Gaussian surface

Theorem [Am., Faugère, Sturmfels (2016)]

"Let $d \geq 3$. The Gaussian moment surface $\mathcal{G}_{1,d}$ has degree $\binom{d}{2}$ in $\mathbb{P}^d$ and its homogeneous prime ideal is minimally generated by $\binom{d}{3}$ cubics. These are the 3×3-minors of the $3 \times d$-matrix

$$H_d \;=\; \begin{pmatrix} 0 & m_0 & 2m_1 & 3m_2 & 4m_3 & \cdots & (d-1)m_{d-2} \\ m_0 & m_1 & m_2 & m_3 & m_4 & \cdots & m_{d-1} \\ m_1 & m_2 & m_3 & m_4 & m_5 & \cdots & m_d \end{pmatrix}.$$

The 3×3-minors of the matrix H_d form a Gröbner basis with respect to the reverse lexicographic term order.

Proposition [Am., Ranestad, Sturmfels (2017)]

The singular locus of $\mathcal{G}_{1,d}$ is the line $\langle m_0, m_1, \ldots, m_{d-2} \rangle$.

A bit more on the proof...

Main geometric tool is the classification of all k-defective surfaces:

Theorem (Terracini, 1921)

Let $X \subset \mathbb{P}^N$ be a reduced, irreducible, non-degenerate projective surface that is k-defective. Then $k \geq 2$ and either

(1) X is the *quadratic Veronese embedding* of a rational normal surface Y in $\mathbb{P}^k$; or

(2) X is contained in a *cone over a curve*, with apex a linear space of dimension $\leq k-2$.

Furthermore, for general points $x_1, \ldots, x_k$ on X there is a hyperplane section tangent along a curve C that passes through these points.
In case (1), the curve C is irreducible;
in case (2), the curve C decomposes into k algebraically equivalent curves $C_1, \ldots, C_k$ with $x_i \in C_i$.

The Pearson hypersurface

"The complete solution depends in ...numerical equation of the ninth order. It is possible that a simpler solution may be found, but the method adopted has only been chosen after many trials and failures."- K. Pearson (1894)

Theorem [Am., Faugère, Sturmfels (2016)]

The defining homogeneous polynomial of $\mathrm{Sec}_2(\mathcal{G}_{1,6})$ *in* $\mathbb{P}^6$ *is a sum of* 31154 *monomials of degree* 39. *This polynomial has degrees* $25, 33, 32, 23, 17, 12, 9$ *in* $m_0, m_1, m_2, m_3, m_4, m_5, m_6$ *respectively.*

- We see in particular that m_6 can be recovered from m_1, m_2, m_3, m_4 and m_5 by solving a nonic.
- We recover Pearson's algebraic identifiability degree 9.

Case $n = 1, k \geq 3$

Theorem [Am., Faugère, Sturmfels (2016)]

$\mathrm{Sec}_3(\mathcal{G}_{1,8})$ has the expected dimension. The mixture model of $k = 3$ univariate Gaussians is algebraically identifiable from its first 8 moments with algebraic degree of identifiability equal to 225.

Concerning the number of solutions, we have observed (counting relabeling) $1 = 1! \times 1^2$, $18 = 2! \times 3^2$, $1350 = 3! \times 15^2$...

Conjecture

The number of solutions up to symmetry equals $\big((2k-1)!!\big)^2$.

If true, the expected number of complex solutions for the 11 moment equations when $k = 4$ would be $4! \times 105^2 = 264600$. *However,*

Theorem [Am., Lindberg, Rodriguez (2020)]

For $k = 4$, there are $4! \times 10350 = 248400$ complex solutions.

Open Problem: How many solutions for $k \geq 5$?

Gaussian Moment Tensors

Let $n = 2$:

$$X \sim N\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{12} & \sigma_{22}\end{pmatrix}\right).$$

Then

- moments of order 1 (*vector* (M_1, M_2)):
  $m_{10} = \mu_1$, $m_{01} = \mu_2$
- moments of order 2 (*matrix* $M_{11}, M_{12} = M_{21}, M_{22}$):
  $m_{20} = \mu_1^2 + \sigma_{11}$, $m_{11} = \mu_1\mu_2 + \sigma_{12}$, $m_{02} = \mu_2^2 + \sigma_{22}$
- moments of order 3 (*tensor* $M_{111}, M_{112} = M_{121} = M_{211}, M_{122} = M_{212} = M_{221}, M_{222}$):
  $m_{30} = \mu_1^3 + 3\mu_1\sigma_{11}$, $m_{21} = \mu_1^2\mu_2 + 2\mu_1\sigma_{12} + \mu_2\sigma_{11}$,
  $m_{12} = \mu_1\mu_2^2 + 2\mu_2\sigma_{12} + \mu_1\sigma_{22}$, $m_{03} = \mu_2^3 + 3\mu_2\sigma_{22}$

In general, the dth order Gaussian moments are entries of a *symmetric tensor* of order d with format $n \times n \times \cdots \times n$.

Method of Moments for $n > 1$

Example ($n = 2, k = 6$)

- Consider a mixture model of $k = 6$ Gaussian components in the plane: $n = 2$.
- The number of parameters is

$$6 \times (2 + 3) + 5 \;=\; 35$$

- To recover the model parameters from the mixture moments, we may measure all moments up to order $d = 7$.
- In fact, the number of moments up to order $d = 7$ is

$$2 + 3 + 4 + 5 + 6 + 7 + 8 \;=\; 35$$

- Since $\mathrm{Sec}_6(\mathcal{G}_{2,7})$ is *not* defective, we have algebraic identifiability.

Higher Dimensional Mixtures

Example $n = 3, k = 2$

Consider the $1 + 2 \times (3 + 6) = 19$ unknown parameters: $0 < \lambda < 1$ and

$$X_1 \sim N\left(\begin{pmatrix}\mu_1\\ \mu_2\\ \mu_3\end{pmatrix}, \begin{pmatrix}\sigma_{11} & \sigma_{12} & \sigma_{13}\\ \sigma_{12} & \sigma_{22} & \sigma_{23}\\ \sigma_{13} & \sigma_{23} & \sigma_{33}\end{pmatrix}\right)$$

$$X_2 \sim N\left(\begin{pmatrix}\nu_1\\ \nu_2\\ \nu_3\end{pmatrix}, \begin{pmatrix}\tau_{11} & \tau_{12} & \tau_{13}\\ \tau_{12} & \tau_{22} & \tau_{23}\\ \tau_{13} & \tau_{23} & \tau_{33}\end{pmatrix}\right)$$

There are $3 + 6 + 10 = 19$ moments up to degree $d = 3$: $m_{100}, m_{010}, m_{001}$
$m_{200}, m_{020}, m_{002}, m_{110}, m_{101}, m_{011}$
$m_{300}, m_{030}, m_{003}, m_{210}, m_{201}, m_{102}, m_{021}, m_{012}, m_{111}$

$$
\begin{aligned}
m_{100} &= \lambda\mu_1 + (1-\lambda)\nu_1 \\
m_{010} &= \lambda\mu_2 + (1-\lambda)\nu_2 \\
m_{001} &= \lambda\mu_3 + (1-\lambda)\nu_3 \\
m_{200} &= \lambda(\mu_1^2 + \sigma_{11}) + (1-\lambda)(\nu_1^2 + \tau_{11}) \\
m_{020} &= \lambda(\mu_2^2 + \sigma_{22}) + (1-\lambda)(\nu_2^2 + \tau_{22}) \\
m_{002} &= \lambda(\mu_3^2 + \sigma_{33}) + (1-\lambda)(\nu_3^2 + \tau_{33}) \\
m_{110} &= \lambda(\mu_1\mu_2 + \sigma_{12}) + (1-\lambda)(\nu_1\nu_2 + \tau_{12}) \\
m_{101} &= \lambda(\mu_1\mu_3 + \sigma_{13}) + (1-\lambda)(\nu_1\nu_3 + \tau_{13}) \\
m_{011} &= \lambda(\mu_2\mu_3 + \sigma_{23}) + (1-\lambda)(\nu_2\nu_3 + \tau_{23}) \\
m_{300} &= \lambda(\mu_1^3 + 3\sigma_{11}\mu_1) + (1-\lambda)(\nu_1^3 + 3\tau_{11}\nu_1) \\
m_{030} &= \lambda(\mu_2^3 + 3\sigma_{22}\mu_2) + (1-\lambda)(\nu_2^3 + 3\tau_{22}\nu_2) \\
m_{003} &= \lambda(\mu_3^3 + 3\sigma_{33}\mu_3) + (1-\lambda)(\nu_3^3 + 3\tau_{33}\nu_3) \\
m_{210} &= \lambda(\mu_1^2\mu_2 + \sigma_{11}\mu_2 + 2\sigma_{12}\mu_1) + (1-\lambda)(\nu_1^2\nu_2 + \tau_{11}\nu_2 + 2\tau_{12}\nu_1) \\
m_{201} &= \lambda(\mu_1^2\mu_3 + \sigma_{11}\mu_3 + 2\sigma_{13}\mu_1) + (1-\lambda)(\nu_1^2\nu_3 + \tau_{11}\nu_3 + 2\tau_{13}\nu_1) \\
m_{120} &= \lambda(\mu_1\mu_2^2 + \sigma_{22}\mu_1 + 2\sigma_{12}\mu_2) + (1-\lambda)(\nu_1\nu_2^2 + \tau_{22}\nu_1 + 2\tau_{12}\nu_2) \\
m_{102} &= \lambda(\mu_1\mu_3^2 + \sigma_{33}\mu_1 + 2\sigma_{13}\mu_3) + (1-\lambda)(\nu_1\nu_3^2 + \tau_{33}\nu_1 + 2\tau_{13}\nu_3) \\
m_{021} &= \lambda(\mu_2^2\mu_3 + \sigma_{22}\mu_3 + 2\sigma_{23}\mu_2) + (1-\lambda)(\nu_2^2\nu_3 + \tau_{22}\nu_3 + 2\tau_{23}\nu_2) \\
m_{012} &= \lambda(\mu_2\mu_3^2 + \sigma_{33}\mu_2 + 2\sigma_{23}\mu_3) + (1-\lambda)(\nu_2\nu_3^2 + \tau_{33}\nu_2 + 2\tau_{23}\nu_3) \\
m_{111} &= \lambda\left(\mu_1\mu_2\mu_3 + \sigma_{12}\mu_3 + \sigma_{13}\mu_2 + \sigma_{23}\mu_1\right) \\
&\quad + (1-\lambda)(\nu_1\nu_2\nu_3 + \tau_{12}\nu_3 + \tau_{13}\nu_2 + \tau_{23}\nu_1)
\end{aligned}
$$

Example $n = 3, k = 2$

Consider the $1 + 2 \times (3 + 6) = 19$ unknown parameters: $0 < \lambda < 1$ and

$$X_1 \sim N\left(\begin{pmatrix}\mu_1\\ \mu_2\\ \mu_3\end{pmatrix}, \begin{pmatrix}\sigma_{11} & \sigma_{12} & \sigma_{13}\\ \sigma_{12} & \sigma_{22} & \sigma_{23}\\ \sigma_{13} & \sigma_{23} & \sigma_{33}\end{pmatrix}\right)$$

$$X_2 \sim N\left(\begin{pmatrix}\nu_1\\ \nu_2\\ \nu_3\end{pmatrix}, \begin{pmatrix}\tau_{11} & \tau_{12} & \tau_{13}\\ \tau_{12} & \tau_{22} & \tau_{23}\\ \tau_{13} & \tau_{23} & \tau_{33}\end{pmatrix}\right)$$

There are $3 + 6 + 10 = 19$ moments up to degree $d = 3$: $m_{100}, m_{010}, m_{001}$
$m_{200}, m_{020}, m_{002}, m_{110}, m_{101}, m_{011}$
$m_{300}, m_{030}, m_{003}, m_{210}, m_{201}, m_{102}, m_{021}, m_{012}, m_{111}$

The moment map $M : \Theta_{3,2} \longrightarrow \mathbb{A}^M_{3,3}$ with $\dim \Theta_{3,2} = \dim \mathbb{A}^M_{3,3} = 19$ is *defective*, since

$$\dim \mathrm{Sec}_2(\mathcal{G}_{3,3}) = 17 < 19.$$

Moment identifiability *fails* in this case!

Another Defective Case

Example ($n = 8, k = 11$)

- Consider a mixture model of $k = 11$ Gaussian components in $n = 8$ dimensional space.
- The number of parameters is

$$11 \times (8 + 36) + 10 = 494$$

- The number of moments up to order $d = 4$ is

$$8 + 36 + 120 + 330 = 494$$

- However, $\dim \mathrm{Sec}_{11}(\mathcal{G}_{8,4}) = 493 < 494$

Homoscedastic Gaussian mixtures

- If all the covariance matrices are equal $\Sigma_1 = \Sigma_2 = \cdots = \Sigma_k = \Sigma$ then we say the mixture is *homoscedastic*.
- The case $n = 1$ has been well studied in the statistics literature. [Bruce G. Lindsay. *Moment matrices: applications in mixtures. Annals of Statistics*, 722–740, 1989.]
- *Rational* identifiability recovery is very similar to Prony's method.

Theorem (Agostini, Am., Ranestad (2019))

Consider a generic homoscedastic mixture of k Gaussians in $\mathbb{R}^n$ and $k \leq n+1$. Then

- *if $k \geq 5$ there is algebraic identifiability from moments up to order 3.*
- *if $k < 5$ there is algebraic identifiability from moments up to order 4.*

Proof.

Study defectivity of homoscedastic secant $\mathrm{Sec}_k^H(\mathcal{G}_{n,d}) \subset \mathrm{Sec}_k(\mathcal{G}_{n,d})$ □

Example ($n = 2, k = 2, d = 3$)

The parameters for the homoscedastic mixture are two mean vectors $\mu_1 = \begin{pmatrix} \mu_{11} \\ \mu_{12} \end{pmatrix}$ and $\mu_2 = \begin{pmatrix} \mu_{21} \\ \mu_{22} \end{pmatrix}$, the common covariance $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$ and the mixture weight λ, in total $2 \times 2 + 3 + 1 = 8$ parameters. On the other hand, there are 9 bivariate moments up to order 3. The system is

$$
\begin{aligned}
m_{10} &= \lambda\mu_{11} + (1-\lambda)\mu_{21} \\
m_{01} &= \lambda\mu_{12} + (1-\lambda)\mu_{22} \\
m_{20} &= \lambda(\mu_{11}^2 + \sigma_{11}) + (1-\lambda)(\mu_{21}^2 + \sigma_{11}) \\
m_{02} &= \lambda(\mu_{12}^2 + \sigma_{22}) + (1-\lambda)(\mu_{22}^2 + \sigma_{22}) \\
m_{11} &= \lambda(\mu_{11}\mu_{12} + \sigma_{12}) + (1-\lambda)(\mu_{21}\mu_{22} + \sigma_{12}) \\
m_{30} &= \lambda(\mu_{11}^3 + 3\sigma_{11}\mu_{11}) + (1-\lambda)(\mu_{21}^3 + 3\sigma_{11}\mu_{21}) \\
m_{03} &= \lambda(\mu_{12}^3 + 3\sigma_{22}\mu_{12}) + (1-\lambda)(\mu_{22}^3 + 3\sigma_{22}\mu_{22}) \\
m_{21} &= \lambda(\mu_{11}^2\mu_{12} + \sigma_{11}\mu_{12} + 2\sigma_{12}\mu_{11}) + (1-\lambda)(\mu_{21}^2\mu_{22} + \sigma_{11}\mu_{22} + 2\sigma_{12}\mu_{21}) \\
m_{12} &= \lambda(\mu_{11}\mu_{12}^2 + \sigma_{22}\mu_{11} + 2\sigma_{12}\mu_{12}) + (1-\lambda)(\mu_{21}\mu_{22}^2 + \sigma_{22}\mu_{21} + 2\sigma_{12}\mu_{22})
\end{aligned}
$$

Defective, of $\dim \mathrm{Sec}_2^H(\mathcal{G}_{2,3}) = 7 < 8$. Need $d = 4$ for identifiability.

Homoscedastic $n = 2, k = 2, d = 3$

What is $\mathrm{Sec}_2^H(\mathcal{G}_{2,3}) \subset \mathbb{A}^M_{2,3} = \mathbb{A}^9$?
Defined by the vanishing of all maximal minors of the 6×5-matrix

$$\begin{pmatrix} 0 & 0 & m_{00} & m_{10} & m_{01} \\ 0 & m_{10} & m_{20} & m_{30} & m_{21} \\ m_{01} & 0 & m_{02} & m_{12} & m_{03} \\ 0 & m_{00} & 2m_{10} & 2m_{20} & 2m_{11} \\ m_{00} & 0 & 2m_{01} & 2m_{11} & 2m_{02} \\ m_{10} & m_{01} & 2m_{11} & 2m_{21} & 2m_{12} \end{pmatrix}.$$

Even better, transforming to cumulant coordinates, the *cumulant variety* is

$$\log(\mathrm{Sec}_2^H(\mathcal{G}_{2,3})) = \left\{ \mathrm{rank} \begin{pmatrix} k_{30} & k_{21} & k_{12} \\ k_{21} & k_{12} & k_{03} \end{pmatrix} \leq 1 \right\}.$$

This is a *cone* over the twisted cubic curve!

Homoscedastic mixtures of two Gaussians

We can completely characterize the case of the mixture of $k = 2$ Gaussians.

Theorem (Agostini, Am., Ranestad (2019))

A generic homoscedastic mixture with $k = 2$ is algebraically identifiable from the moments or order up to $d = 4$ and is rationally identifiable from the moments of order up to 5.

- Actually, in the statistically significant cases, we can recover the parameters from the moments of order up to four.
- There are up to 3 real solutions but only *one* statistically significant.
- First case where this consideration affects the order of identifiability!
- We give a very explicit algorithm, which actually requires only n cumulants of order 3 and one cumulant of order 4.

Homoscedastic with known covariance

Complete classification of algebraic identifiability for homoscedastic Gaussian mixtures with *known* covariance (for all n, k, d)!

Theorem (Am. (2017))

Consider the statistical model given by the homoscedastic mixture of $k > 1$ Gaussians in $\mathbb{R}^n$ with known covariance Σ. Let $d > 1$ such that the number of moments is at least the number of parameters: $\binom{n+d}{d} \geq (n+1)k$. Then the model is always algebraically identifiable from the moments up to order d, except in the following cases:

- $d = 2$
- $d = 3,\ n = 4,\ k = 7$
- $d = 4,\ n = 2,\ k = 5$
- $d = 4,\ n = 4,\ k = 14$

Proof.

Use the Alexander-Hirschowitz Theorem □

Conclusion

Several challenges and research directions.

- For general $n, k \geq 1$, what is the minimum order $D(n, k)$ of moments that we need to identify the mixture model?
- Can we classify all Gaussian mixture models that are *defective*?
- Homoscedastic mixture case: cases $k > n + 1 > 2$ still open.
- Careful study of other submodels: equal means submodel and spherical Gaussian mixtures.
- Study of the *real* admissible solutions to the polynomial moment systems.
- Numerical considerations, combination with existing methods (EM algorithm, tensor decomposition, ...)

Dziękuję!