Estimating Gaussian Mixtures Using Sparse Polynomial Moment Systems

Julia Lindberg

Online Machine Learning Seminar

UT Austin

March 15, 2023

Joint work with Carlos Améndola and Jose Rodriguez

Table of Contents

1. Problem Set Up
2. (Numerical) Algebraic Geometry Primer
3. Density Estimation for Gaussian Mixture Models
4. Applications in High Dimensional Statistics

Problem Set Up

Density Estimation

- A common problem studied in statistics is density estimation

Problem Set Up

Density Estimation

- A common problem studied in statistics is density estimation
- Given N samples from a distribution p (unknown), can we estimate p?

Problem Set Up

Density Estimation

- A common problem studied in statistics is density estimation
- Given N samples from a distribution p (unknown), can we estimate p?
- Need to assume p is from some family of distributions

Problem Set Up

Density Estimation

- A common problem studied in statistics is density estimation
- Given N samples from a distribution p (unknown), can we estimate p?
- Need to assume p is from some family of distributions

Theorem (Chapter 3 [GBC16])

A Gaussian mixture model is a universal approximator of densities, in the sense that any smooth density can be approximated with any specific nonzero amount of error by a Gaussian mixture model with enough components.

Gaussian Mixture Models

- A random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ is a *Gaussian* random variable if it has density

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

- X is distributed as a *mixture of k Gaussians* if it is the convex combination of k Gaussian densities

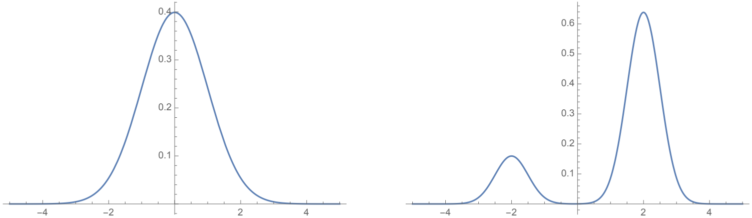

Figure: $\mathcal{N}(0,1)$ density (left) and $0.2\mathcal{N}(-2, 0.5) + 0.8\mathcal{N}(2, 0.5)$ density (right).

- Given iid samples, $y_1, \ldots, y_N$, distributed as the mixture of k Gaussians, how to recover parameters $\mu_i, \sigma_i^2, \lambda_i$?

Density Estimation

MLE

- Given iid samples, $y_1, \ldots, y_N$, distributed as the mixture of k Gaussians, how to recover parameters $\mu_i, \sigma_i^2, \lambda_i$?

- **Idea 1 :** Maximum likelihood estimation

$$\text{argmax}_{\mu,\sigma^2,\lambda} \prod_{j=1}^{N} \sum_{i=1}^{k} \lambda_i \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(- \frac{(y_j - \mu_i)^2}{2\sigma_i^2} \right)$$

- Given iid samples, $y_1, \ldots, y_N$, distributed as the mixture of k Gaussians, how to recover parameters $\mu_i, \sigma_i^2, \lambda_i$?

- **Idea 1 :** Maximum likelihood estimation

$$\text{argmax}_{\mu,\sigma^2,\lambda} \prod_{j=1}^{N} \sum_{i=1}^{k} \lambda_i \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(y_j-\mu_i)^2}{2\sigma_i^2}\right)$$

- Iterative algorithms (EM) can find local optima

- Given iid samples, $y_1, \ldots, y_N$, distributed as the mixture of k Gaussians, how to recover parameters $\mu_i, \sigma_i^2, \lambda_i$?

- **Idea 1 :** Maximum likelihood estimation

$$\text{argmax}_{\mu,\sigma^2,\lambda} \prod_{j=1}^{N} \sum_{i=1}^{k} \lambda_i \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(y_j-\mu_i)^2}{2\sigma_i^2}\right)$$

- Iterative algorithms (EM) can find local optima
  - Local optima can be arbitrarily bad and random initialization will converge to these bad points with probability $1 - e^{-\Omega(k)}$ [JZB+16]

Density Estimation

MLE

- Given iid samples, $y_1, \ldots, y_N$, distributed as the mixture of k Gaussians, how to recover parameters $\mu_i, \sigma_i^2, \lambda_i$?

- **Idea 1 :** Maximum likelihood estimation

$$\operatorname{argmax}_{\mu,\sigma^2,\lambda} \prod_{j=1}^{N} \sum_{i=1}^{k} \lambda_i \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(y_j-\mu_i)^2}{2\sigma_i^2}\right)$$

- Iterative algorithms (EM) can find local optima
  - Local optima can be arbitrarily bad and random initialization will converge to these bad points with probability $1 - e^{-\Omega(k)}$ [JZB+16]
  - No bound on number of critical points [AFS16]

- Given iid samples, $y_1, \ldots, y_N$, distributed as the mixture of k Gaussians, how to recover parameters $\mu_i, \sigma_i^2, \lambda_i$?

- **Idea 1 :** Maximum likelihood estimation

$$\text{argmax}_{\mu,\sigma^2,\lambda} \prod_{j=1}^{N} \sum_{i=1}^{k} \lambda_i \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(y_j-\mu_i)^2}{2\sigma_i^2}\right)$$

- Iterative algorithms (EM) can find local optima
  - Local optima can be arbitrarily bad and random initialization will converge to these bad points with probability $1 - e^{-\Omega(k)}$ [JZB+16]
  - No bound on number of critical points [AFS16]
  - Need to access all samples at each iteration

- **Idea 2 :** Method of moments

- **Idea 2 :** Method of moments
  - The method of moments estimator is consistent

- **Idea 2 :** Method of moments
  - The method of moments estimator is consistent
  - Gaussian mixture models are identifiable from their moments

- **Idea 2 :** Method of moments
  - The method of moments estimator is consistent
  - Gaussian mixture models are identifiable from their moments
  - **IF** you can solve the moment equations, then can recover exact parameters

Method of Moments

- For $i \geq 0$, the i*−th moment* of a random variable X with density f is

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

Method of Moments

- For $i \geq 0$, the $i\text{−}th$ *moment* of a random variable X with density f is

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

- For parameterized distributions, moments are functions of parameters

Method of Moments

- For $i \geq 0$, the *$i-th$ moment* of a random variable X with density f is

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

- For parameterized distributions, moments are functions of parameters
- Ex. The first few moments of a $\mathcal{N}(\mu, \sigma^2)$ random variable are:

$$m_1 = \mu, \qquad m_2 = \mu^2 + \sigma^2, \qquad m_3 = \mu^3 + 3\mu\sigma^2$$

Method of Moments

- Consider a statistical model with p unknown parameters, $\theta = (\theta_1, \ldots, \theta_p)$ and the moments up to order M as functions of θ

$$m_1 = g_1(\theta), \ldots, m_M = g_M(\theta)$$

Method of Moments

- Consider a statistical model with p unknown parameters, $\theta = (\theta_1, \ldots, \theta_p)$ and the moments up to order M as functions of θ

$$m_1 = g_1(\theta), \ldots, m_M = g_M(\theta)$$

- **Method of Moments:**
  1. Compute sample moments

$$\overline{m}_i = \frac{1}{N} \sum_{j=1}^{N} y_j^i$$

Method of Moments

- Consider a statistical model with p unknown parameters, $\theta = (\theta_1, \ldots, \theta_p)$ and the moments up to order M as functions of θ

$$m_1 = g_1(\theta), \ldots, m_M = g_M(\theta)$$

- **Method of Moments:**
  1. Compute sample moments
  $$\overline{m}_i = \frac{1}{N} \sum_{j=1}^{N} y_j^i$$
  2. Solve $g_i(\theta) = \overline{m}_i$ for $i = 1, \ldots, M$ to recover parameters

- The moments of the Gaussian distributions are $M_0(\mu,\sigma^2) = 1$, $M_1(\mu,\sigma^2) = \mu$,

$$M_\ell(\mu,\sigma^2) = \mu M_{\ell-1} + (\ell-1)\sigma^2 M_{\ell-2}, \qquad \ell \geq 2$$

- The moments of the Gaussian distributions are $M_0(\mu, \sigma^2) = 1$, $M_1(\mu, \sigma^2) = \mu$,

$$M_\ell(\mu, \sigma^2) = \mu M_{\ell-1} + (\ell - 1)\sigma^2 M_{\ell-2}, \qquad \ell \geq 2$$

- The moments of mixtures of k Gaussians are

$$m_\ell = \sum_{i=1}^{k} \lambda_i M_\ell(\mu_i, \sigma_i^2), \qquad \ell \geq 0$$

Method of Moments

$k = 1$

- When $k = 1$ this is just density estimation for $\mathcal{N}(\mu_1, \sigma_1^2)$

Method of Moments

$k = 1$

- When $k = 1$ this is just density estimation for $\mathcal{N}(\mu_1, \sigma_1^2)$
- The moment equations are

$$
\begin{aligned}
1 &= \lambda_1 \\
\overline{m}_1 &= \lambda_1 \mu_1 \\
\overline{m}_2 &= \lambda_1(\mu_1^2 + \sigma_1^2)
\end{aligned}
$$

Method of Moments

$k = 1$

- When $k = 1$ this is just density estimation for $\mathcal{N}(\mu_1, \sigma_1^2)$
- The moment equations are

$$\begin{aligned} 1 &= \lambda_1 \\ \overline{m}_1 &= \lambda_1 \mu_1 \\ \overline{m}_2 &= \lambda_1(\mu_1^2 + \sigma_1^2) \end{aligned}$$

- There is a unique solution given by

$$\lambda_1 = 1, \qquad \mu_1 = \overline{m}_1, \qquad \sigma_1^2 = \overline{m}_2 - \overline{m}_1^2$$

- When $k = 2$, the first 6 moment equations are

$$
\begin{aligned}
1 &= \lambda_1 + \lambda_2 \\
\overline{m}_1 &= \lambda_1\mu_1 + \lambda_2\mu_2 \\
\overline{m}_2 &= \lambda_1(\mu_1^2 + \sigma_1^2) + \lambda_2(\mu_2^2 + \sigma_2^2) \\
\overline{m}_3 &= \lambda_1(\mu_1^3 + 3\mu_1\sigma_1^2) + \lambda_2(\mu_2^3 + 3\mu_2\sigma_2^2) \\
\overline{m}_4 &= \lambda_1(\mu_1^4 + 6\mu_1^2\sigma_1^2 + 3\sigma_1^4) + \lambda_2(\mu_2^4 + 6\mu_2^2\sigma_2^2 + 3\sigma_2^4) \\
\overline{m}_5 &= \lambda_1(\mu_1^5 + 10\mu_1^3\sigma_1^2 + 15\mu_1\sigma_1^4) + \lambda_2(\mu_2^5 + 10\mu_2^3\sigma_2^2 + 15\mu_2\sigma_2^4)
\end{aligned}
$$

- When $k = 2$, the first 6 moment equations are

$$\begin{aligned}
1 &= \lambda_1 + \lambda_2 \\
\overline{m}_1 &= \lambda_1\mu_1 + \lambda_2\mu_2 \\
\overline{m}_2 &= \lambda_1(\mu_1^2 + \sigma_1^2) + \lambda_2(\mu_2^2 + \sigma_2^2) \\
\overline{m}_3 &= \lambda_1(\mu_1^3 + 3\mu_1\sigma_1^2) + \lambda_2(\mu_2^3 + 3\mu_2\sigma_2^2) \\
\overline{m}_4 &= \lambda_1(\mu_1^4 + 6\mu_1^2\sigma_1^2 + 3\sigma_1^4) + \lambda_2(\mu_2^4 + 6\mu_2^2\sigma_2^2 + 3\sigma_2^4) \\
\overline{m}_5 &= \lambda_1(\mu_1^5 + 10\mu_1^3\sigma_1^2 + 15\mu_1\sigma_1^4) + \lambda_2(\mu_2^5 + 10\mu_2^3\sigma_2^2 + 15\mu_2\sigma_2^4)
\end{aligned}$$

- **Obervation:** If $(\lambda_1, \mu_1, \sigma_1^2, \lambda_2, \mu_2, \sigma_2^2)$ is a solution, so is $(\lambda_2, \mu_2, \sigma_2^2, \lambda_1, \mu_1, \sigma_1^2)$

- When $k = 2$, the first 6 moment equations are

$$
\begin{aligned}
1 &= \lambda_1 + \lambda_2 \\
\overline{m}_1 &= \lambda_1\mu_1 + \lambda_2\mu_2 \\
\overline{m}_2 &= \lambda_1(\mu_1^2 + \sigma_1^2) + \lambda_2(\mu_2^2 + \sigma_2^2) \\
\overline{m}_3 &= \lambda_1(\mu_1^3 + 3\mu_1\sigma_1^2) + \lambda_2(\mu_2^3 + 3\mu_2\sigma_2^2) \\
\overline{m}_4 &= \lambda_1(\mu_1^4 + 6\mu_1^2\sigma_1^2 + 3\sigma_1^4) + \lambda_2(\mu_2^4 + 6\mu_2^2\sigma_2^2 + 3\sigma_2^4) \\
\overline{m}_5 &= \lambda_1(\mu_1^5 + 10\mu_1^3\sigma_1^2 + 15\mu_1\sigma_1^4) + \lambda_2(\mu_2^5 + 10\mu_2^3\sigma_2^2 + 15\mu_2\sigma_2^4)
\end{aligned}
$$

- **Obervation:** If $(\lambda_1, \mu_1, \sigma_1^2, \lambda_2, \mu_2, \sigma_2^2)$ is a solution, so is $(\lambda_2, \mu_2, \sigma_2^2, \lambda_1, \mu_1, \sigma_1^2)$
  - This symmetry is called *label swapping*

- When $k = 2$, the first 6 moment equations are

$$
\begin{aligned}
1 &= \lambda_1 + \lambda_2 \\
\overline{m}_1 &= \lambda_1\mu_1 + \lambda_2\mu_2 \\
\overline{m}_2 &= \lambda_1(\mu_1^2 + \sigma_1^2) + \lambda_2(\mu_2^2 + \sigma_2^2) \\
\overline{m}_3 &= \lambda_1(\mu_1^3 + 3\mu_1\sigma_1^2) + \lambda_2(\mu_2^3 + 3\mu_2\sigma_2^2) \\
\overline{m}_4 &= \lambda_1(\mu_1^4 + 6\mu_1^2\sigma_1^2 + 3\sigma_1^4) + \lambda_2(\mu_2^4 + 6\mu_2^2\sigma_2^2 + 3\sigma_2^4) \\
\overline{m}_5 &= \lambda_1(\mu_1^5 + 10\mu_1^3\sigma_1^2 + 15\mu_1\sigma_1^4) + \lambda_2(\mu_2^5 + 10\mu_2^3\sigma_2^2 + 15\mu_2\sigma_2^4)
\end{aligned}
$$

- **Obervation:** If $(\lambda_1, \mu_1, \sigma_1^2, \lambda_2, \mu_2, \sigma_2^2)$ is a solution, so is $(\lambda_2, \mu_2, \sigma_2^2, \lambda_1, \mu_1, \sigma_1^2)$
    - This symmetry is called *label swapping*
    - For a k mixture model, solutions will come in groups of k!

Method of Moments
History Detour

- The study of mixtures of Gaussians dates back to Karl Pearson in 1894 studying measurements of Naples crab populations [Pea94]

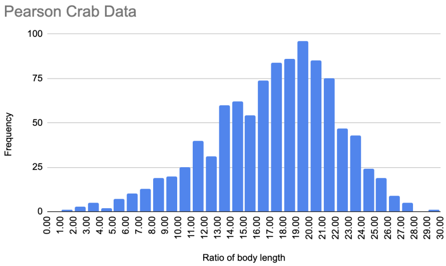

Figure: Pearson's crab data

Method of Moments

History Detour

- The study of mixtures of Gaussians dates back to Karl Pearson in 1894 studying measurements of Naples crab populations [Pea94]

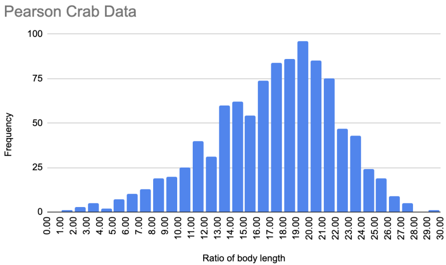

Figure: Pearson's crab data

- Pearson reduced this to finding roots of degree 9 polynomial in the variable $x = \mu_1 \mu_2$

Method of Moments

History Detour

- The study of mixtures of Gaussians dates back to Karl Pearson in 1894 studying measurements of Naples crab populations [Pea94]

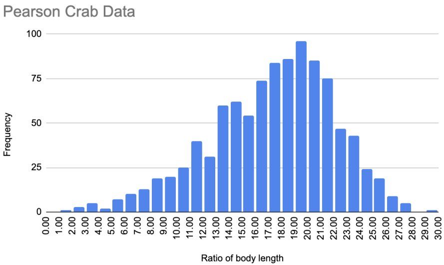

Figure: Pearson's crab data

- Pearson reduced this to finding roots of degree 9 polynomial in the variable $x = \mu_1\mu_2$
- **Framework:** Solve square polynomial system to get finitely many potential densities then select one closest to the next sample moments

Identifiability

Different notions of identifiability based on fiber of map:

$$\begin{aligned}\Phi_M \ : \ \Delta_{k-1} \times \mathbb{R}^k \times \mathbb{R}^k_{>0} &\to \mathbb{R}^M \\ (\lambda, \mu, \sigma^2) &\mapsto (m_0, \ldots, m_M)\end{aligned}$$

Identifiability

Different notions of identifiability based on fiber of map:

$$\begin{aligned} \Phi_M \ : \ \Delta_{k-1} \times \mathbb{R}^k \times \mathbb{R}^k_{>0} &\to \mathbb{R}^M \\ (\lambda, \mu, \sigma^2) &\mapsto (m_0, \ldots, m_M) \end{aligned}$$

1. Algebraic: For what M is $|\Phi_M^{-1}(m)| < \infty$ for **almost all** $m \in \mathrm{Im}(\Phi_M)$?
   - $3k - 1$ [ARS18]

Identifiability

Different notions of identifiability based on fiber of map:

$$\Phi_M \; : \; \Delta_{k-1} \times \mathbb{R}^k \times \mathbb{R}^k_{>0} \to \mathbb{R}^M$$
$$(\lambda, \mu, \sigma^2) \mapsto (m_0, \ldots, m_M)$$

1. Algebraic: For what M is $|\Phi_M^{-1}(m)| < \infty$ for **almost all** $m \in \mathrm{Im}(\Phi_M)$?
   - $3k - 1$ [ARS18]
2. Statistical: For what M does $|\Phi_M^{-1}(m)| = k!$ for **all** $m \in \mathrm{Im}(\Phi_M)$?
   - $4k - 2$ [KMV12]

Identifiability

Different notions of identifiability based on fiber of map:

$$\begin{aligned}\Phi_M \; : \; \Delta_{k-1} \times \mathbb{R}^k \times \mathbb{R}^k_{>0} &\to \mathbb{R}^M \\ (\lambda, \mu, \sigma^2) &\mapsto (m_0, \ldots, m_M)\end{aligned}$$

1. Algebraic: For what M is $|\Phi_M^{-1}(m)| < \infty$ for **almost all** $m \in \text{Im}(\Phi_M)$?
   - $3k-1$ [ARS18]
2. Statistical: For what M does $|\Phi_M^{-1}(m)| = k!$ for **all** $m \in \text{Im}(\Phi_M)$?
   - $4k-2$ [KMV12]
3. Rational: For what M is $|\Phi_M^{-1}(m)| = k!$ for **almost all** $m \in \text{Im}(\Phi_M)$?

Identifiability

Different notions of identifiability based on fiber of map:

$$\Phi_M \ : \ \Delta_{k-1} \times \mathbb{R}^k \times \mathbb{R}^k_{>0} \to \mathbb{R}^M$$
$$(\lambda, \mu, \sigma^2) \mapsto (m_0, \ldots, m_M)$$

1. Algebraic: For what M is $|\Phi_M^{-1}(m)| < \infty$ for **almost all** $m \in \mathrm{Im}(\Phi_M)$?
   - $3k-1$ [ARS18]
2. Statistical: For what M does $|\Phi_M^{-1}(m)| = k!$ for **all** $m \in \mathrm{Im}(\Phi_M)$?
   - $4k-2$ [KMV12]
3. Rational: For what M is $|\Phi_M^{-1}(m)| = k!$ for **almost all** $m \in \mathrm{Im}(\Phi_M)$?

Theorem (L., Améndola, Rodriguez)

Mixtures of k univariate Gaussians are rationally identifiable from moments $m_1, \ldots, m_{3k+2}$.

Identifiability

Different notions of identifiability based on fiber of map:

$$\Phi_M \; : \; \Delta_{k-1} \times \mathbb{R}^k \times \mathbb{R}^k_{>0} \to \mathbb{R}^M$$
$$(\lambda, \mu, \sigma^2) \mapsto (m_0, \dots, m_M)$$

1. Algebraic: For what M is $|\Phi_M^{-1}(m)| < \infty$ for **almost all** $m \in \mathrm{Im}(\Phi_M)$?
   - $3k - 1$ [ARS18]
2. Statistical: For what M does $|\Phi_M^{-1}(m)| = k!$ for **all** $m \in \mathrm{Im}(\Phi_M)$?
   - $4k - 2$ [KMV12]
3. Rational: For what M is $|\Phi_M^{-1}(m)| = k!$ for **almost all** $m \in \mathrm{Im}(\Phi_M)$?

Theorem (L., Améndola, Rodriguez)

Mixtures of k univariate Gaussians are rationally identifiable from moments $m_1, \dots, m_{3k+2}$.

- **Conjecture:** Gaussian mixture models are rationally identifiable from $m_1, \dots, m_{3k}$

Method of Moments Framework

1. Solve moment equations

$$
\begin{aligned}
1 &= m_0 \\
\overline{m}_1 &= m_1 \\
&\vdots \\
\overline{m}_{3k-1} &= m_{3k-1}
\end{aligned}
$$

over the complex numbers to get finitely many *complex* solutions

Method of Moments Framework

1. Solve moment equations

$$
\begin{aligned}
1 &= m_0 \\
\overline{m}_1 &= m_1 \\
&\vdots \\
\overline{m}_{3k-1} &= m_{3k-1}
\end{aligned}
$$

   over the complex numbers to get finitely many *complex* solutions

2. Filter out statistically meaningful solutions (real solutions with $\lambda_i \geq 0, \sigma_i^2 > 0$)

Method of Moments Framework

1. Solve moment equations

$$\begin{aligned} 1 &= m_0 \\ \overline{m}_1 &= m_1 \\ &\vdots \\ \overline{m}_{3k-1} &= m_{3k-1} \end{aligned}$$

   over the complex numbers to get finitely many *complex* solutions

2. Filter out statistically meaningful solutions (real solutions with $\lambda_i \geq 0, \sigma_i^2 > 0$)

3. Select statistically meaningful solution agreeing with moments $\overline{m}_{3k}, \overline{m}_{3k+1}, \overline{m}_{3k+2}$

Method of Moments Framework

1. Solve moment equations

$$
\begin{aligned}
1 &= m_0 \\
\overline{m}_1 &= m_1 \\
&\vdots \\
\overline{m}_{3k-1} &= m_{3k-1}
\end{aligned}
$$

   over the complex numbers to get finitely many *complex* solutions

2. Filter out statistically meaningful solutions (real solutions with $\lambda_i \geq 0, \sigma_i^2 > 0$)

3. Select statistically meaningful solution agreeing with moments $\overline{m}_{3k}, \overline{m}_{3k+1}, \overline{m}_{3k+2}$

   Question: How do I solve a square system of polynomial equations?

Table of Contents

1. Problem Set Up
2. (Numerical) Algebraic Geometry Primer
3. Density Estimation for Gaussian Mixture Models
4. Applications in High Dimensional Statistics

Algebraic Geometry Primer

- Let $f_1, \dots, f_m \in \mathbb{R}[x_1, \dots, x_n]$. The (complex) *variety* of $F = \langle f_1, \dots, f_m \rangle$ is

$$\mathcal{V}(F) = \{x \in \mathbb{C}^n : f_1(x) = 0, \dots, f_m(x) = 0\}$$

Algebraic Geometry Primer

- Let $f_1, \ldots, f_m \in \mathbb{R}[x_1, \ldots, x_n]$. The (complex) *variety* of $F = \langle f_1, \ldots, f_m \rangle$ is

$$\mathcal{V}(F) = \{x \in \mathbb{C}^n : f_1(x) = 0, \ldots, f_m(x) = 0\}$$

- Interested in case when $n = m$ and $|\mathcal{V}(F)| < \infty$

- Consider $|\mathcal{V}(F)| < \infty$. **Question:** How big is $|\mathcal{V}(F)|$?

- Consider $|\mathcal{V}(F)| < \infty$. **Question:** How big is $|\mathcal{V}(F)|$?

Theorem (Bezout)

$|\mathcal{V}(F)| \leq d_1 \cdots d_n$ *where* $d_i = \deg(f_i)$

- Consider $|\mathcal{V}(F)| < \infty$. **Question:** How big is $|\mathcal{V}(F)|$?

Theorem (Bezout)

$|\mathcal{V}(F)| \leq d_1 \cdots d_n$ *where* $d_i = \deg(f_i)$

- Can be strict upper bound when f_i are sparse

- Consider $|\mathcal{V}(F)| < \infty$. **Question:** How big is $|\mathcal{V}(F)|$?

Theorem (Bezout)

$|\mathcal{V}(F)| \leq d_1 \cdots d_n$ *where* $d_i = \deg(f_i)$

- Can be strict upper bound when f_i are sparse

Theorem (BKK Bound [Ber75, Kho78, Kou76])

$|\mathcal{V}(F) \cap (\mathbb{C}^*)^n| \leq \mathrm{MVol}(\mathrm{Newt}(f_1), \ldots, \mathrm{Newt}(f_n))$

- Consider $|\mathcal{V}(F)| < \infty$. **Question:** How big is $|\mathcal{V}(F)|$?

Theorem (Bezout)

$|\mathcal{V}(F)| \leq d_1 \cdots d_n$ *where* $d_i = \deg(f_i)$

- Can be strict upper bound when f_i are sparse

Theorem (BKK Bound [Ber75, Kho78, Kou76])

$|\mathcal{V}(F) \cap (\mathbb{C}^*)^n| \leq \mathrm{MVol}(\mathrm{Newt}(f_1), \ldots, \mathrm{Newt}(f_n))$

- In general, not easy to compute the mixed volume (#P hard)

Finding All Complex Solutions

Homotopy Continuation

- Idea: Solving most polynomial systems is hard, but some are easy

Finding All Complex Solutions

Homotopy Continuation

- Idea: Solving most polynomial systems is hard, but some are easy

$$H_T = \begin{cases} 2(x_2x_3 - x_1x_4) + 3x_3 = 0 \\ 2(x_1x_4 - x_2x_3) + 4x_4 = 0 \\ x_1^2 + x_3^2 = 1 \\ x_2^2 + x_4^2 = 1 \end{cases} \qquad H_S = \begin{cases} x_1^2 = 1 \\ x_2^2 = 1 \\ x_3^2 = 1 \\ x_4^2 = 1 \end{cases}$$

- Idea: Solving most polynomial systems is hard, but some are easy

$$H_T = \begin{cases} 2(x_2x_3 - x_1x_4) + 3x_3 = 0 \\ 2(x_1x_4 - x_2x_3) + 4x_4 = 0 \\ x_1^2 + x_3^2 = 1 \\ x_2^2 + x_4^2 = 1 \end{cases} \qquad H_S = \begin{cases} x_1^2 = 1 \\ x_2^2 = 1 \\ x_3^2 = 1 \\ x_4^2 = 1 \end{cases}$$

- Can I map my solutions from H_S to H_T?

- Idea: Solving most polynomial systems is hard, but some are easy

$$H_T = \begin{cases} 2(x_2x_3 - x_1x_4) + 3x_3 = 0 \\ 2(x_1x_4 - x_2x_3) + 4x_4 = 0 \\ x_1^2 + x_3^2 = 1 \\ x_2^2 + x_4^2 = 1 \end{cases} \qquad H_S = \begin{cases} x_1^2 = 1 \\ x_2^2 = 1 \\ x_3^2 = 1 \\ x_4^2 = 1 \end{cases}$$

- Can I map my solutions from H_S to H_T?
- Define $H_t := (1-t)H_S + tH_T$ and compute H_t as $t \to 1$
  - Called following *homotopy paths*

Finding All Complex Solutions

Homotopy Continuation

- Idea: Solving most polynomial systems is hard, but some are easy

$$H_T = \begin{cases} 2(x_2x_3 - x_1x_4) + 3x_3 = 0 \\ 2(x_1x_4 - x_2x_3) + 4x_4 = 0 \\ x_1^2 + x_3^2 = 1 \\ x_2^2 + x_4^2 = 1 \end{cases} \qquad H_S = \begin{cases} x_1^2 = 1 \\ x_2^2 = 1 \\ x_3^2 = 1 \\ x_4^2 = 1 \end{cases}$$

- Can I map my solutions from H_S to H_T?
- Define $H_t := (1-t)H_S + tH_T$ and compute H_t as $t \to 1$
  - Called following *homotopy paths*
- Typically use predictor-corrector methods
  - Predict: Take step along tangent direction at a point
  - Correct: Use Newton's method

Homotopy Continuation Visual

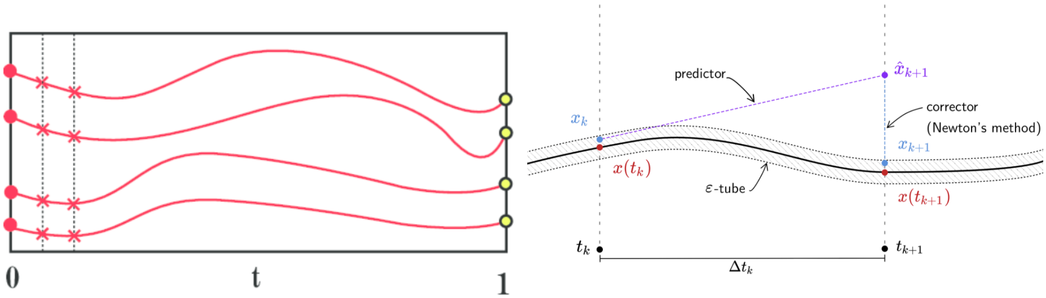

Figure: The homotopy $H_t = (1-t)H_S + tH_T$ (left)[KW14] and the predictor corrector step (right) [BT18]

- Want to pick a start system, H_S, such that
  1. The solutions of H_S are easy to find
  2. The number of solutions to $H_S \approx$ the number of solutions to H_T

- Want to pick a start system, H_S, such that
  1. The solutions of H_S are easy to find
  2. The number of solutions to $H_S \approx$ the number of solutions to H_T

- If $|\mathcal{V}(F)| \approx d_1 \cdots d_n$ then a **total degree** start system is suitable. i.e.

$$H_S = \langle x_1^{d_1} - 1, \ldots, x_n^{d_n} - 1 \rangle$$

- Want to pick a start system, H_S, such that
  1. The solutions of H_S are easy to find
  2. The number of solutions to $H_S \approx$ the number of solutions to H_T

- If $|\mathcal{V}(F)| \approx d_1 \cdots d_n$ then a **total degree** start system is suitable. i.e.

$$H_S = \langle x_1^{d_1} - 1, \ldots, x_n^{d_n} - 1 \rangle$$

- If $\mathrm{MVol}(\mathrm{Newt}(f_1), \ldots, \mathrm{Newt}(f_n)) \ll d_1 \cdots d_n$ then a **polyhedral** start system is suitable

- Want to pick a start system, H_S, such that
  1. The solutions of H_S are easy to find
  2. The number of solutions to $H_S \approx$ the number of solutions to H_T

- If $|\mathcal{V}(F)| \approx d_1 \cdots d_n$ then a **total degree** start system is suitable. i.e.

$$H_S = \langle x_1^{d_1} - 1, \ldots, x_n^{d_n} - 1 \rangle$$

- If $\mathrm{MVol}(\mathrm{Newt}(f_1), \ldots, \mathrm{Newt}(f_n)) \ll d_1 \cdots d_n$ then a **polyhedral** start system is suitable

- There exists an algorithm that finds this binomial start system [HS95]

Examples of Start Systems

$$F = \langle x^2 - 3x + 2,\ 2xy + y - 1 \rangle$$

Total degree: $\langle x^2 - 1,\ y^2 - 1 \rangle$

Polyhedral: $\langle x^2 + 2,\ y - 1 \rangle$

Table of Contents

1. Problem Set Up
2. (Numerical) Algebraic Geometry Primer
3. Density Estimation for Gaussian Mixture Models
4. Applications in High Dimensional Statistics

Back to Gaussian Mixture Models

- There are three special cases of Gaussian mixture models commonly studied in the statistics literature:

Back to Gaussian Mixture Models

- There are three special cases of Gaussian mixture models commonly studied in the statistics literature:
  1. The mixing coefficients are known
  2. The mixing coefficients are known and the variances are equal
  3. Only the means are unknown

Main Result

Theorem (L., Améndola, Rodriguez [LAR21])

In all cases, Gaussian mixture models are algebraically identifiable using moment equations of lowest degree. Moreover, the mixed volume of each of set of equations is given below.

	Known mixing coefficients	Known mixing coefficients + equal variances	Unknown means
Moment equations	$m_1, \ldots, m_{2k}$	$m_1, \ldots, m_{k+1}$	$m_1, \ldots, m_k$
Unknowns	μ_i, σ_i^2	μ_i, σ^2	μ_i
Mixed volume	$(2k-1)!!k!$	$\frac{(k+1)!}{2}$	$k!$
Mixed volume tight	Yes for $k \leq 8$	Yes for $k \leq 8$	Yes

	Mixed Volume	Bezout Bound
Known mixing coefficients	$(2k-1)!!k!$	$(2k)!$
Known mixing coefficients + equal variances	$\frac{(k+1)!}{2}$	$(k+1)!$
Unknown means	$k!$	$k!$

	Mixed Volume	Bezout Bound
Known mixing coefficients	$(2k-1)!!k!$	$(2k)!$
Known mixing coefficients + equal variances	$\frac{(k+1)!}{2}$	$(k+1)!$
Unknown means	$k!$	$k!$

- Our proofs of the mixed volume in the first two cases give a start system that tracks mixed volume number of paths

	Mixed Volume	Bezout Bound
Known mixing coefficients	$(2k-1)!!k!$	$(2k)!$
Known mixing coefficients + equal variances	$\frac{(k+1)!}{2}$	$(k+1)!$
Unknown means	$k!$	$k!$

- Our proofs of the mixed volume in the first two cases give a start system that tracks mixed volume number of paths
- In the final case if $\lambda_i = \frac{1}{k}$ and σ_i^2 are equal, there is a unique solution up to symmetry

Table of Contents

1. Problem Set Up
2. (Numerical) Algebraic Geometry Primer
3. Density Estimation for Gaussian Mixture Models
4. Applications in High Dimensional Statistics

Gaussian Mixture Models

In high dimensions

- A random variable $X \in \mathbb{R}^n$ is distributed as a *multivariate Gaussian* with mean $\mu \in \mathbb{R}^n$ and covariance $\Sigma \in \mathbb{R}^{n \times n}$, $\Sigma \succ 0$, if it has density

$$f_X(x_1, \ldots, x_n | \mu, \Sigma) = ((2\pi)^n \det(\Sigma))^{-1/2} \exp\Big(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\Big)$$

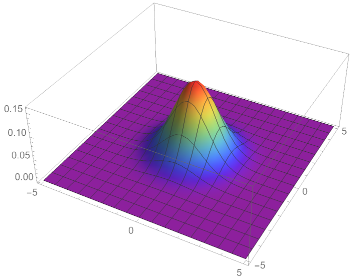

Figure: Gaussian density in $\mathbb{R}^2$ with mean $\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and covariance $\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Example

$k = n = 2$

Suppose $X \sim \lambda_1 \mathcal{N}(\mu_1, \Sigma_1) + \lambda_2 \mathcal{N}(\mu_2, \Sigma_2)$ where

$$\mu_1 = \begin{pmatrix} \mu_{11} \\ \mu_{12} \end{pmatrix}, \qquad \Sigma_1 = \begin{pmatrix} \sigma_{111} & \sigma_{112} \\ \sigma_{112} & \sigma_{122} \end{pmatrix}$$

$$\mu_2 = \begin{pmatrix} \mu_{21} \\ \mu_{21} \end{pmatrix}, \qquad \Sigma_2 = \begin{pmatrix} \sigma_{211} & \sigma_{212} \\ \sigma_{212} & \sigma_{222} \end{pmatrix}.$$

The moment equations up to order 3 are

$$\begin{aligned}
m_{00} &= \lambda_1 + \lambda_2 \\
m_{10} &= \lambda_1 \mu_{11} + \lambda_2 \mu_{21} \\
m_{01} &= \lambda_1 \mu_{12} + \lambda_2 \mu_{22} \\
m_{20} &= \lambda_1(\mu_{11}^2 + \sigma_{111}) + \lambda_2(\mu_{21}^2 + \sigma_{211}) \\
m_{11} &= \lambda_1(\mu_{11}\mu_{12} + \sigma_{112}) + \lambda_2(\mu_{21}\mu_{22} + \sigma_{212}) \\
m_{02} &= \lambda_1(\mu_{12}^2 + \sigma_{122}) + \lambda_2(\mu_{22}^2 + \sigma_{222}) \\
m_{30} &= \lambda_1(\mu_{11}^3 + 3\mu_{11}\sigma_{111}) + \lambda_2(\mu_{21}^3 + 3\mu_{21}\sigma_{211}) \\
m_{21} &= \lambda_1(\mu_{11}^2\mu_{12} + 2\mu_{11}\sigma_{112} + \mu_{12}\sigma_{111}) + \lambda_2(\mu_{21}^2\mu_{22} + 2\mu_{21}\sigma_{212} + \mu_{22}\sigma_{211}) \\
m_{12} &= \lambda_1(\mu_{11}\mu_{12}^2 + \mu_{11}\sigma_{122} + 2\mu_{12}\sigma_{112}) + \lambda_2(\mu_{21}\mu_{22}^2 + \mu_{21}\sigma_{222} + 2\mu_{22}\sigma_{212}) \\
m_{03} &= \lambda_1(\mu_{12}^3 + 3\mu_{12}\sigma_{122}) + \lambda_2(\mu_{22}^3 + 3\mu_{22}\sigma_{222})
\end{aligned}$$

Example

$k = n = 2$

Suppose $X \sim \lambda_1 \mathcal{N}(\mu_1, \Sigma_1) + \lambda_2 \mathcal{N}(\mu_2, \Sigma_2)$ where

$$\mu_1 = \begin{pmatrix} \mu_{11} \\ \mu_{12} \end{pmatrix}, \qquad \Sigma_1 = \begin{pmatrix} \sigma_{111} & \sigma_{112} \\ \sigma_{112} & \sigma_{122} \end{pmatrix}$$

$$\mu_2 = \begin{pmatrix} \mu_{21} \\ \mu_{21} \end{pmatrix}, \qquad \Sigma_2 = \begin{pmatrix} \sigma_{211} & \sigma_{212} \\ \sigma_{212} & \sigma_{222} \end{pmatrix}.$$

The moment equations up to order 3 are

$$\begin{aligned}
m_{00} &= \lambda_1 + \lambda_2 \\
m_{10} &= \lambda_1 \mu_{11} + \lambda_2 \mu_{21} \\
m_{01} &= \lambda_1 \mu_{12} + \lambda_2 \mu_{22} \\
m_{20} &= \lambda_1(\mu_{11}^2 + \sigma_{111}) + \lambda_2(\mu_{21}^2 + \sigma_{211}) \\
m_{11} &= \lambda_1(\mu_{11}\mu_{12} + \sigma_{112}) + \lambda_2(\mu_{21}\mu_{22} + \sigma_{212}) \\
m_{02} &= \lambda_1(\mu_{12}^2 + \sigma_{122}) + \lambda_2(\mu_{22}^2 + \sigma_{222}) \\
m_{30} &= \lambda_1(\mu_{11}^3 + 3\mu_{11}\sigma_{111}) + \lambda_2(\mu_{21}^3 + 3\mu_{21}\sigma_{211}) \\
m_{21} &= \lambda_1(\mu_{11}^2\mu_{12} + 2\mu_{11}\sigma_{112} + \mu_{12}\sigma_{111}) + \lambda_2(\mu_{21}^2\mu_{22} + 2\mu_{21}\sigma_{212} + \mu_{22}\sigma_{211}) \\
m_{12} &= \lambda_1(\mu_{11}\mu_{12}^2 + \mu_{11}\sigma_{122} + 2\mu_{12}\sigma_{112}) + \lambda_2(\mu_{21}\mu_{22}^2 + \mu_{21}\sigma_{222} + 2\mu_{22}\sigma_{212}) \\
m_{03} &= \lambda_1(\mu_{12}^3 + 3\mu_{12}\sigma_{122}) + \lambda_2(\mu_{22}^3 + 3\mu_{22}\sigma_{222})
\end{aligned}$$

- **Key Observation:** The $m_{0,0,\ldots,0,i_t,0,\ldots 0}-$th moment is the same as the i_t-th order moment for the univariate Gaussian mixture model $\sum_{\ell=1}^{k} \lambda_\ell \mathcal{N}(\mu_{\ell t}, \sigma_{\ell tt})$

- **Key Observation:** The $m_{0,0,\dots,0,i_t,0,\dots0}-$th moment is the same as the i_t-th order moment for the univariate Gaussian mixture model $\sum_{\ell=1}^{k} \lambda_\ell \mathcal{N}(\mu_{\ell t}, \sigma_{\ell tt})$
- Density estimation for high dimensional Gaussian mixture models becomes multiple instances of one dimensional problems

Higher Order Moments

Application of Univariate Results

- **Key Observation:** The $m_{0,0,\ldots,0,i_t,0,\ldots 0}-$th moment is the same as the i_t-th order moment for the univariate Gaussian mixture model $\sum_{\ell=1}^{k} \lambda_\ell \mathcal{N}(\mu_{\ell t}, \sigma_{\ell tt})$
- Density estimation for high dimensional Gaussian mixture models becomes multiple instances of one dimensional problems
- **Advantage:** Only track the best statistically meaningful solution

Input: A set of sample moments $\mathbf{m}$[1]

[1] Sample moments need to be in the same cell as the moments of the true density

Input: A set of sample moments $\mathbf{m}$[1]

Output: Parameters $\lambda_\ell \in \mathbb{R}$, $\mu_\ell \in \mathbb{R}^n$, $\Sigma_\ell \succ 0$ for $\ell \in [k]$ such that $\mathbf{m}$ are the moments of distribution $\sum_{\ell=1}^{k} \lambda_\ell \mathcal{N}(\mu_\ell, \Sigma_\ell)$

[1] Sample moments need to be in the same cell as the moments of the true density

Input: A set of sample moments $\mathbf{m}$[1]

Output: Parameters $\lambda_\ell \in \mathbb{R}$, $\mu_\ell \in \mathbb{R}^n$, $\Sigma_\ell \succ 0$ for $\ell \in [k]$ such that $\mathbf{m}$ are the moments of distribution $\sum_{\ell=1}^{k} \lambda_\ell \mathcal{N}(\mu_\ell, \Sigma_\ell)$

1. Solve the general univariate case using sample moments $\overline{m}_{0,\dots,0,1}, \dots, \overline{m}_{0,\dots,0,3k-1}$ to get parameters λ_ℓ, $\mu_{\ell,1}$ and $\sigma_{\ell,1,1}$

[1] Sample moments need to be in the same cell as the moments of the true density

Algorithm

Density Estimation for High Dimensional Gaussian Mixture Models

Input: A set of sample moments $\mathbf{m}$[1]

Output: Parameters $\lambda_\ell \in \mathbb{R}$, $\mu_\ell \in \mathbb{R}^n$, $\Sigma_\ell \succ 0$ for $\ell \in [k]$ such that $\mathbf{m}$ are the moments of distribution $\sum_{\ell=1}^{k} \lambda_\ell \mathcal{N}(\mu_\ell, \Sigma_\ell)$

1. Solve the general univariate case using sample moments $\overline{m}_{0,\dots,0,1}, \dots, \overline{m}_{0,\dots,0,3k-1}$ to get parameters λ_ℓ, $\mu_{\ell,1}$ and $\sigma_{\ell,1,1}$
2. Select statistically meaningful solution with moments $\overline{m}_{0,\dots,0,3k}, \overline{m}_{0,\dots,0,3k+1}, \overline{m}_{0,\dots,0,3k+2}$

[1] Sample moments need to be in the same cell as the moments of the true density

Input: A set of sample moments $\mathbf{m}$[1]

Output: Parameters $\lambda_\ell \in \mathbb{R}$, $\mu_\ell \in \mathbb{R}^n$, $\Sigma_\ell \succ 0$ for $\ell \in [k]$ such that $\mathbf{m}$ are the moments of distribution $\sum_{\ell=1}^{k} \lambda_\ell \mathcal{N}(\mu_\ell, \Sigma_\ell)$

1. Solve the general univariate case using sample moments $\overline{m}_{0,\dots,0,1}, \dots, \overline{m}_{0,\dots,0,3k-1}$ to get parameters λ_ℓ, $\mu_{\ell,1}$ and $\sigma_{\ell,1,1}$
2. Select statistically meaningful solution with moments $\overline{m}_{0,\dots,0,3k}, \overline{m}_{0,\dots,0,3k+1}, \overline{m}_{0,\dots,0,3k+2}$
3. Using the mixing coefficients λ_ℓ solve the known mixing coefficients case $n-1$ times to obtain the remaining means and variances

[1]Sample moments need to be in the same cell as the moments of the true density

Input: A set of sample moments $\mathbf{m}$[1]

Output: Parameters $\lambda_\ell \in \mathbb{R}$, $\mu_\ell \in \mathbb{R}^n$, $\Sigma_\ell \succ 0$ for $\ell \in [k]$ such that $\mathbf{m}$ are the moments of distribution $\sum_{\ell=1}^k \lambda_\ell \mathcal{N}(\mu_\ell, \Sigma_\ell)$

1. Solve the general univariate case using sample moments $\overline{m}_{0,\ldots,0,1}, \ldots, \overline{m}_{0,\ldots,0,3k-1}$ to get parameters λ_ℓ, $\mu_{\ell,1}$ and $\sigma_{\ell,1,1}$
2. Select statistically meaningful solution with moments $\overline{m}_{0,\ldots,0,3k}, \overline{m}_{0,\ldots,0,3k+1}, \overline{m}_{0,\ldots,0,3k+2}$
3. Using the mixing coefficients λ_ℓ solve the known mixing coefficients case $n-1$ times to obtain the remaining means and variances
4. Select the statistically meaningful solution closest to next sample moments

[1] Sample moments need to be in the same cell as the moments of the true density

Input: A set of sample moments $\mathbf{m}$[1]

Output: Parameters $\lambda_\ell \in \mathbb{R}$, $\mu_\ell \in \mathbb{R}^n$, $\Sigma_\ell \succ 0$ for $\ell \in [k]$ such that $\mathbf{m}$ are the moments of distribution $\sum_{\ell=1}^{k} \lambda_\ell \mathcal{N}(\mu_\ell, \Sigma_\ell)$

1. Solve the general univariate case using sample moments $\overline{m}_{0,\ldots,0,1}, \ldots, \overline{m}_{0,\ldots,0,3k-1}$ to get parameters λ_ℓ, $\mu_{\ell,1}$ and $\sigma_{\ell,1,1}$
2. Select statistically meaningful solution with moments $\overline{m}_{0,\ldots,0,3k}, \overline{m}_{0,\ldots,0,3k+1}, \overline{m}_{0,\ldots,0,3k+2}$
3. Using the mixing coefficients λ_ℓ solve the known mixing coefficients case $n-1$ times to obtain the remaining means and variances
4. Select the statistically meaningful solution closest to next sample moments
5. The covariances are linear in the other entries, solve this linear system

[1] Sample moments need to be in the same cell as the moments of the true density

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

- Suppose $X \sim \lambda_1 \mathcal{N}(\mu_1, \Sigma_1) + \lambda_2 \mathcal{N}(\mu_2, \Sigma_2)$ where

$$\mu_1 = \begin{pmatrix} \mu_{11} \\ \mu_{12} \end{pmatrix}, \qquad \Sigma_1 = \begin{pmatrix} \sigma_{111}^2 & \sigma_{112} \\ \sigma_{112} & \sigma_{122}^2 \end{pmatrix}$$

$$\mu_2 = \begin{pmatrix} \mu_{21} \\ \mu_{21} \end{pmatrix}, \qquad \Sigma_2 = \begin{pmatrix} \sigma_{211}^2 & \sigma_{212} \\ \sigma_{212} & \sigma_{222}^2 \end{pmatrix}.$$

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

- Suppose $X \sim \lambda_1 \mathcal{N}(\mu_1, \Sigma_1) + \lambda_2 \mathcal{N}(\mu_2, \Sigma_2)$ where

$$\mu_1 = \begin{pmatrix} \mu_{11} \\ \mu_{12} \end{pmatrix}, \qquad \Sigma_1 = \begin{pmatrix} \sigma_{111}^2 & \sigma_{112} \\ \sigma_{112} & \sigma_{122}^2 \end{pmatrix}$$

$$\mu_2 = \begin{pmatrix} \mu_{21} \\ \mu_{21} \end{pmatrix}, \qquad \Sigma_2 = \begin{pmatrix} \sigma_{211}^2 & \sigma_{212} \\ \sigma_{212} & \sigma_{222}^2 \end{pmatrix}.$$

- Given sample moments

$$[\overline{m}_{10}, \overline{m}_{20}, \overline{m}_{30}, \overline{m}_{40}, \overline{m}_{50}, \overline{m}_{60}] = [-0.25,\ 2.75,\ -1.0,\ 22.75,\ -6.5,\ 322.75]$$

$$[\overline{m}_{01}, \overline{m}_{02}, \overline{m}_{03}, \overline{m}_{04}, \overline{m}_{05}] = [2.5,\ 16.125,\ 74.5,\ 490.5625,\ 2921.25]$$

$$[\overline{m}_{11}, \overline{m}_{21}] = [0.8125,\ 7.75]$$

Example (cont.)

Algorithm in Action

- **Step 1:** Solve general case to obtain $\lambda_\ell, \mu_{\ell 1}, \sigma^2_{\ell 11}$ for $\ell = 1, 2$

$$
\begin{aligned}
1 &= \lambda_1 + \lambda_2 \\
-0.25 &= \lambda_1 \mu_{11} + \lambda_2 \mu_{21} \\
2.75 &= \lambda_1(\mu_{11}^2 + \sigma_{111}^2) + \lambda_2(\mu_{21}^2 + \sigma_{211}^2) \\
-1 &= \lambda_1(\mu_{11}^3 + 3\mu_{11}\sigma_{111}^2) + \lambda_2(\mu_{21}^3 + 3\mu_{21}\sigma_{211}^2) \\
22.75 &= \lambda_1(\mu_{11}^4 + 6\mu_{11}^2\sigma_{111}^2 + 3\sigma_{111}^4) + \lambda_2(\mu_{21}^4 + 6\mu_{21}^2\sigma_{211}^2 + 3\sigma_{211}^4) \\
-6.5 &= \lambda_1(\mu_{11}^5 + 10\mu_{11}^3\sigma_{111}^2 + 15\mu_{11}\sigma_{111}^4) + \lambda_2(\mu_{21}^5 + 10\mu_{21}^3\sigma_{211}^2 + 15\mu_{21}\sigma_{211}^4)
\end{aligned}
$$

Example (cont.)

Algorithm in Action

- **Step 1:** Solve general case to obtain $\lambda_\ell, \mu_{\ell 1}, \sigma^2_{\ell 11}$ for $\ell = 1, 2$

$$
\begin{aligned}
1 &= \lambda_1 + \lambda_2 \\
-0.25 &= \lambda_1\mu_{11} + \lambda_2\mu_{21} \\
2.75 &= \lambda_1(\mu_{11}^2 + \sigma_{111}^2) + \lambda_2(\mu_{21}^2 + \sigma_{211}^2) \\
-1 &= \lambda_1(\mu_{11}^3 + 3\mu_{11}\sigma_{111}^2) + \lambda_2(\mu_{21}^3 + 3\mu_{21}\sigma_{211}^2) \\
22.75 &= \lambda_1(\mu_{11}^4 + 6\mu_{11}^2\sigma_{111}^2 + 3\sigma_{111}^4) + \lambda_2(\mu_{21}^4 + 6\mu_{21}^2\sigma_{211}^2 + 3\sigma_{211}^4) \\
-6.5 &= \lambda_1(\mu_{11}^5 + 10\mu_{11}^3\sigma_{111}^2 + 15\mu_{11}\sigma_{111}^4) + \lambda_2(\mu_{21}^5 + 10\mu_{21}^3\sigma_{211}^2 + 15\mu_{21}\sigma_{211}^4)
\end{aligned}
$$

- (Up to symmetry) two statistically meaningful solutions:

$$
\begin{aligned}
(\lambda_1, \lambda_2, \mu_{11}, \mu_{21}, \sigma_{111}^2, \sigma_{211}^2) &= (0.25,\ 0.75,\ 0, -1,\ 3,\ 1) \\
(\lambda_1, \lambda_2, \mu_{11}, \mu_{21}, \sigma_{111}^2, \sigma_{211}^2) &= (0.967,\ 0.033,\ -0.378,\ 3.493,\ 2.272,\ 0.396)
\end{aligned}
$$

- **Step 1:** Solve general case to obtain $\lambda_\ell, \mu_{\ell 1}, \sigma^2_{\ell 11}$ for $\ell = 1, 2$

$$\begin{aligned}
1 &= \lambda_1 + \lambda_2 \\
-0.25 &= \lambda_1 \mu_{11} + \lambda_2 \mu_{21} \\
2.75 &= \lambda_1(\mu_{11}^2 + \sigma_{111}^2) + \lambda_2(\mu_{21}^2 + \sigma_{211}^2) \\
-1 &= \lambda_1(\mu_{11}^3 + 3\mu_{11}\sigma_{111}^2) + \lambda_2(\mu_{21}^3 + 3\mu_{21}\sigma_{211}^2) \\
22.75 &= \lambda_1(\mu_{11}^4 + 6\mu_{11}^2\sigma_{111}^2 + 3\sigma_{111}^4) + \lambda_2(\mu_{21}^4 + 6\mu_{21}^2\sigma_{211}^2 + 3\sigma_{211}^4) \\
-6.5 &= \lambda_1(\mu_{11}^5 + 10\mu_{11}^3\sigma_{111}^2 + 15\mu_{11}\sigma_{111}^4) + \lambda_2(\mu_{21}^5 + 10\mu_{21}^3\sigma_{211}^2 + 15\mu_{21}\sigma_{211}^4)
\end{aligned}$$

- (Up to symmetry) two statistically meaningful solutions:

$$\begin{aligned}
(\lambda_1, \lambda_2, \mu_{11}, \mu_{21}, \sigma_{111}^2, \sigma_{211}^2) &= (0.25,\ 0.75,\ 0, -1,\ 3,\ 1) \\
(\lambda_1, \lambda_2, \mu_{11}, \mu_{21}, \sigma_{111}^2, \sigma_{211}^2) &= (0.967,\ 0.033,\ -0.378,\ 3.493,\ 2.272,\ 0.396)
\end{aligned}$$

- **Step 2:** First solution has $m_{60} = 322.75$, second has $m_{60} = 294.686$

Example (cont.)

Algorithm in Action

- **Step 1:** Solve general case to obtain $\lambda_\ell, \mu_{\ell 1}, \sigma^2_{\ell 11}$ for $\ell = 1, 2$

$$
\begin{aligned}
1 &= \lambda_1 + \lambda_2 \\
-0.25 &= \lambda_1 \mu_{11} + \lambda_2 \mu_{21} \\
2.75 &= \lambda_1(\mu_{11}^2 + \sigma_{111}^2) + \lambda_2(\mu_{21}^2 + \sigma_{211}^2) \\
-1 &= \lambda_1(\mu_{11}^3 + 3\mu_{11}\sigma_{111}^2) + \lambda_2(\mu_{21}^3 + 3\mu_{21}\sigma_{211}^2) \\
22.75 &= \lambda_1(\mu_{11}^4 + 6\mu_{11}^2\sigma_{111}^2 + 3\sigma_{111}^4) + \lambda_2(\mu_{21}^4 + 6\mu_{21}^2\sigma_{211}^2 + 3\sigma_{211}^4) \\
-6.5 &= \lambda_1(\mu_{11}^5 + 10\mu_{11}^3\sigma_{111}^2 + 15\mu_{11}\sigma_{111}^4) + \lambda_2(\mu_{21}^5 + 10\mu_{21}^3\sigma_{211}^2 + 15\mu_{21}\sigma_{211}^4)
\end{aligned}
$$

- (Up to symmetry) two statistically meaningful solutions:

$$
\begin{aligned}
(\lambda_1, \lambda_2, \mu_{11}, \mu_{21}, \sigma_{111}^2, \sigma_{211}^2) &= (0.25,\ 0.75,\ 0, -1,\ 3,\ 1) \\
(\lambda_1, \lambda_2, \mu_{11}, \mu_{21}, \sigma_{111}^2, \sigma_{211}^2) &= (0.967,\ 0.033,\ -0.378,\ 3.493,\ 2.272,\ 0.396)
\end{aligned}
$$

- **Step 2:** First solution has $m_{60} = 322.75$, second has $m_{60} = 294.686$
- Select first solution

Example (cont.)
Algorithm in Action

- **Step 3:** Using $\lambda_1 = 0.25$, $\lambda_2 = 0.75$ solve

$$\begin{aligned}
2.5 &= 0.25 \cdot \mu_{12} + 0.75 \cdot \mu_{22} \\
16.125 &= 0.25 \cdot (\mu_{12}^2 + \sigma_{122}^2) + 0.75 \cdot (\mu_{22}^2 + \sigma_{222}^2) \\
74.5 &= 0.25 \cdot (\mu_{12}^3 + 3\mu_{12}\sigma_{122}^2) + 0.75 \cdot (\mu_{22}^3 + 3\mu_{22}\sigma_{222}^2) \\
490.5625 &= 0.25 \cdot (\mu_{12}^4 + 6\mu_{12}^2\sigma_{122}^2 + 3\sigma_{122}^4) + 0.75 \cdot (\mu_{22}^4 + 6\mu_{22}^2\sigma_{222}^2 + 3\sigma_{222}^4)
\end{aligned}$$

Example (cont.)

Algorithm in Action

- **Step 3:** Using $\lambda_1 = 0.25$, $\lambda_2 = 0.75$ solve

$$
\begin{aligned}
2.5 &= 0.25 \cdot \mu_{12} + 0.75 \cdot \mu_{22} \\
16.125 &= 0.25 \cdot (\mu_{12}^2 + \sigma_{122}^2) + 0.75 \cdot (\mu_{22}^2 + \sigma_{222}^2) \\
74.5 &= 0.25 \cdot (\mu_{12}^3 + 3\mu_{12}\sigma_{122}^2) + 0.75 \cdot (\mu_{22}^3 + 3\mu_{22}\sigma_{222}^2) \\
490.5625 &= 0.25 \cdot (\mu_{12}^4 + 6\mu_{12}^2\sigma_{122}^2 + 3\sigma_{122}^4) + 0.75 \cdot (\mu_{22}^4 + 6\mu_{22}^2\sigma_{222}^2 + 3\sigma_{222}^4)
\end{aligned}
$$

- One statistically meaningful solution

$$(\mu_{12}, \mu_{22}, \sigma_{122}^2, \sigma_{222}^2) = (-2,\ 4,\ 2,\ 3.5\ )$$

Example (cont.)

Algorithm in Action

- **Step 3:** Using $\lambda_1 = 0.25$, $\lambda_2 = 0.75$ solve

$$
\begin{aligned}
2.5 &= 0.25 \cdot \mu_{12} + 0.75 \cdot \mu_{22} \\
16.125 &= 0.25 \cdot (\mu_{12}^2 + \sigma_{122}^2) + 0.75 \cdot (\mu_{22}^2 + \sigma_{222}^2) \\
74.5 &= 0.25 \cdot (\mu_{12}^3 + 3\mu_{12}\sigma_{122}^2) + 0.75 \cdot (\mu_{22}^3 + 3\mu_{22}\sigma_{222}^2) \\
490.5625 &= 0.25 \cdot (\mu_{12}^4 + 6\mu_{12}^2\sigma_{122}^2 + 3\sigma_{122}^4) + 0.75 \cdot (\mu_{22}^4 + 6\mu_{22}^2\sigma_{222}^2 + 3\sigma_{222}^4)
\end{aligned}
$$

- One statistically meaningful solution

$$(\mu_{12}, \mu_{22}, \sigma_{122}^2, \sigma_{222}^2) = (-2,\ 4,\ 2,\ 3.5\ )$$

- **Step 4:** Choose only statistically meaningful solution

Example (cont.)

Algorithm in Action

- **Step 5:** Solve the linear system

$$0.8125 = 0.25 \cdot (2 + \sigma_{112}) + 0.75 \cdot \sigma_{212}$$
$$7.75 = 0.25 \cdot (-4 + 2 \cdot \sigma_{112}) + 9$$

Example (cont.)

Algorithm in Action

- **Step 5:** Solve the linear system

$$0.8125 = 0.25 \cdot (2 + \sigma_{112}) + 0.75 \cdot \sigma_{212}$$
$$7.75 = 0.25 \cdot (-4 + 2 \cdot \sigma_{112}) + 9$$

- There is one solution

$$(\sigma_{112}, \sigma_{212}) = (0.5,\ 0.25)$$

- **Step 5:** Solve the linear system

$$0.8125 = 0.25 \cdot (2 + \sigma_{112}) + 0.75 \cdot \sigma_{212}$$
$$7.75 = 0.25 \cdot (-4 + 2 \cdot \sigma_{112}) + 9$$

- There is one solution

$$(\sigma_{112}, \sigma_{212}) = (0.5,\ 0.25)$$

- Estimate that our samples came from density

$$0.25 \cdot \mathcal{N}\left(\begin{bmatrix} -1 \\ -2 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 2 \end{bmatrix} \right) + 0.75 \cdot \mathcal{N}\left(\begin{bmatrix} 0 \\ 4 \end{bmatrix}, \begin{bmatrix} 3 & 0.25 \\ 0.25 & 3.5 \end{bmatrix} \right)$$

- Steps 3 and 4 can be run in parallel

- Steps 3 and 4 can be run in parallel
- Need to track $N_k + (2k-1)!!k! \cdot (n-1)$ homotopy paths where $N_k = \#$ of paths needed for a general k mixture model

- Steps 3 and 4 can be run in parallel
- Need to track $N_k + (2k-1)!!k! \cdot (n-1)$ homotopy paths where $N_k = \#$ of paths needed for a general k mixture model
- Number of homotopy paths is linear in n

- Steps 3 and 4 can be run in parallel
- Need to track $N_k + (2k-1)!!k! \cdot (n-1)$ homotopy paths where $N_k = \#$ of paths needed for a general k mixture model
- Number of homotopy paths is linear in n
- Even simpler in cases where some of the parameters are known

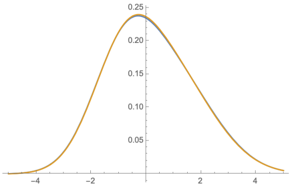

Figure: Two Gaussian mixture densities with $k = 3$ components and the same first eight moments.

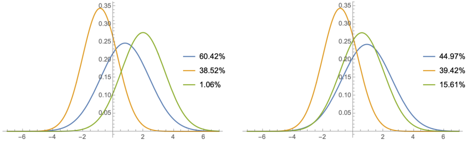

Figure: Individual components of two Gaussian mixture models with similar mixture densities.

- We perform the method of moments on the mixture of 2 Gaussians in $\mathbb{R}^n$ with diagonal covariance matrices

n	10	100	1,000	10,000	100,000
Time (s)	0.17	0.71	6.17	62.05	650.96
Error	7.8×10^{-15}	4.1×10^{-13}	5.7×10^{-13}	3.0×10^{-11}	1.8×10^{-9}
Normalized Error	1.9×10^{-16}	1.0×10^{-15}	1.4×10^{-16}	7.3×10^{-16}	4.5×10^{-15}

Table: Average running time and numerical error for a mixture of 2 Gaussians in $\mathbb{R}^n$

Conclusion

- Gave new rational and algebraic identifiability results for Gaussian mixture models
- Gave upper bound for number of solutions to univariate Gaussian k mixture moment systems in three cases
- Applied these results to efficiently do density estimation in high dimensions

Thank you! Questions?

Paper: 'Estimating Gaussian mixture models using sparse polynomial moment systems'

arXiv:2106.15675

References I

[AFS16] Carlos Améndola, Jean-Charles Faugère, and Bernd Sturmfels, *Moment varieties of Gaussian mixtures*, J. Algebr. Stat. **7** (2016), no. 1, 14–28. MR 3529332

[ARS18] Carlos Améndola, Kristian Ranestad, and Bernd Sturmfels, *Algebraic identifiability of Gaussian mixtures*, Int. Math. Res. Not. IMRN (2018), no. 21, 6556–6580. MR 3873537

[Ber75] David N. Bernstein, *The number of roots of a system of equations*, Funkcional. Anal. i Priložen. **9** (1975), no. 3, 1–4. MR 0435072

[BT18] Paul Breiding and Sascha Timme, *Homotopycontinuation.jl: A package for homotopy continuation in Julia*, Mathematical Software – ICMS 2018, Springer International Publishing, 2018, pp. 458–465.

[GBC16] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep learning*, MIT Press, 2016, `http://www.deeplearningbook.org`.

[HS95] Birkett Huber and Bernd Sturmfels, *A polyhedral method for solving sparse polynomial systems*, Math. Comp. **64** (1995), no. 212, 1541–1555. MR 1297471

[JZB+16] Chi Jin, Yuchen Zhang, Sivaraman Balakrishnan, Martin J. Wainwright, and Michael I. Jordan, *Local maxima in the likelihood of gaussian mixture models: Structural results and algorithmic consequences*, 30th Annual Conference on Neural Information Processing Systems, NIPS 2016, vol. 29, 2016, pp. 4116–4124.

[Kho78] Askold G. Khovanskii, *Newton polyhedra, and the genus of complete intersections*, Funktsional. Anal. i Prilozhen. **12** (1978), no. 1, 51–61. MR 487230

[KMV12] Adam Tauman Kalai, Ankur Moitra, and Gregory Valiant, *Disentangling Gaussians*, Communications of The ACM **55** (2012), no. 2, 113–120.

[Kou76] Anatoli G. Kouchnirenko, *Polyèdres de Newton et nombres de Milnor*, Invent. Math. **32** (1976), no. 1, 1–31. MR 419433

[KW14] Kei Kobayashi and Henry P. Wynn, *Computational algebraic methods in efficient estimation*, Geometric theory of information, Signals Commun. Technol., Springer, Cham, 2014, pp. 119–140. MR 3329739

[LAR21] Julia Lindberg, Carlos Améndola, and Jose Israel Rodriguez, *Estimating gaussian mixtures using sparse polynomial moment systems.*, arXiv preprint arXiv:2106.15675 (2021).

References II

[Pea94] Karl Pearson, *Contributions to the mathematical theory of evolution*, Philosophical Transactions of the Royal Society A **185** (1894), 71–110.