Identifiability of Gaussian mixtures from sixth-order moments

Alexander Taveira Blomenhofer [1]

[1]Centrum Wiskunde & Informatica

Gaussian Mixtures

Gaussian mixture (GM) distributions are an ubiquitous model in Data Science, often used to describe clustered data.

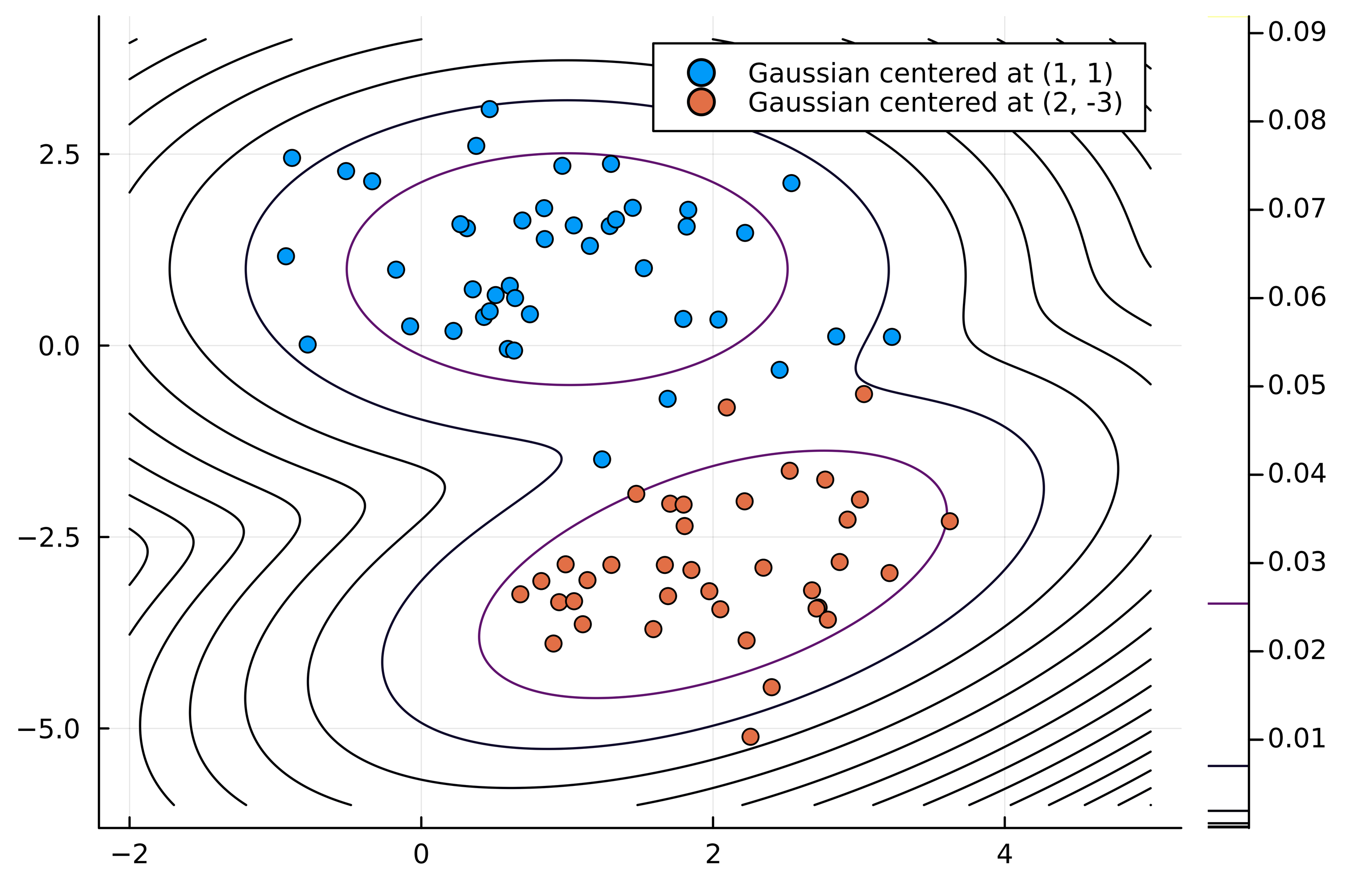
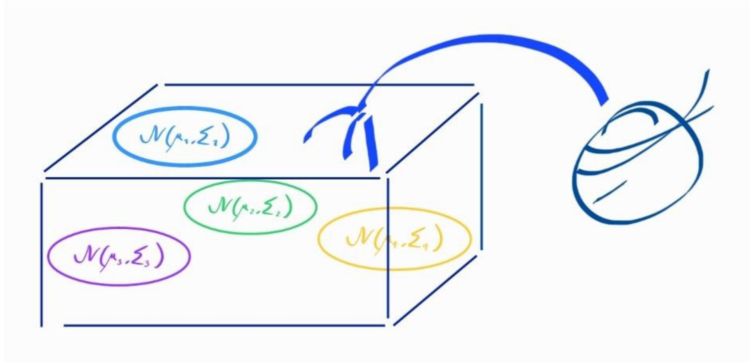


Figure 1. Samples from a mixture of two Gaussian distributions on $\mathbb{R}^2$: Each of the 80 samples was chosen by selecting one of the two Gaussians with probability $\frac{1}{2}$, then sampling the selected Gaussian. The contour lines are the level sets of the probability density function of the mixture.

A mixture of m Gaussians $\mathcal{N}(\mu_1, \Sigma_1), \ldots, \mathcal{N}(\mu_m, \Sigma_m)$ is sampled by choosing $i \in \{1, \ldots, m\}$ at random, then sampling the Gaussian with parameters (μ_i, Σ_i).



Identifying parameters from moments

In applications, an important problem is to estimate the parameters of a Gaussian mixture from empirical data. We look at the related question of *moment identifiability*: Which moments are necessary or sufficient to determine the general parameters $(\mu_1, \Sigma_1), \ldots, (\mu_m, \Sigma_m)$ from the (exact) knowledge of moments?

Summary

1. Moments of degree at most 4 never suffice, if $m \geq 2$.
2. There is computational evidence that moments of degree 5 determine the parameters up to finitely many possibilities, for $m = \Theta(n^3)$, but we lack a proof.
3. Moments of degree 6 uniquely determine the parameters for some $m = \Theta(n^4)$.

Our proof of statement 3 uses techniques from Algebraic Geometry, notably, the theory of *secant varieties*. Moments of a Gaussian mixture are sums, or, more precisely, convex combinations, of moments of Gaussian distributions.

Gaussian mixture moments

The degree-6 moments of a (uniformly weighted) Gaussian mixture attain the form

$$\frac{1}{m}\sum_{i=1}^{m} \ell_i^6 + 15q_i\ell_i^4 + 45q_i^2\ell_i^2 + 15q_i^3 \tag{1}$$

Here, $\ell_i = \mu_i^T X$ are linear forms, whose coefficients are the mean vectors. The quadratic forms $q_i := X^T\Sigma_i X$ encode the covariance matrices.

The Gaussian moment variety and its secants

The *Gaussian moment variety* $\mathrm{GM}_6(\mathbb{C}^n)$ is the Zariski closure of the set of all sextic forms

$$\ell^6 + 15q\ell^4 + 45q^2\ell^2 + 15q^3, \tag{2}$$

where ℓ is a linear form and q is a quadratic form on $\mathbb{C}^n$.

Its m-th secant variety $\sigma_m \mathrm{GM}_6(\mathbb{C}^n)$ is the closure of all m-fold sums of elements of $\mathrm{GM}_6(\mathbb{C}^n)$. Therefore, $\sigma_m \mathrm{GM}_6(\mathbb{C}^n)$ is the closure of the set of rank-m GM moments, cf. (1).

A recent theorem due to Massarenti and Mella reduces secant identifiability to computations.

Theorem (Massarenti-Mella, [2])

Let $V \subseteq \mathbb{C}^N$ an irreducible variety of dimension n, not contained in a proper subspace and with nondegenerate Gauss map. Assume that for $m+1$ general points $x_1, \ldots, x_{m+1} \in V$, the tangent spaces

$$T_{x_1}V + \ldots + T_{x_{m+1}}V = T_{x_1}V \oplus \ldots \oplus T_{x_{m+1}}V \tag{3}$$

form a direct sum, and that $(m+1)n + m \leq N$. Then, a sum

$$t = x_1 + \ldots + x_m \tag{4}$$

of m general elements of V has no representation as a sum of m elements of V, other than (4).

Secant identifiability for $\mathrm{GM}_6(\mathbb{C}^n)$

The Gaussian moment variety $\mathrm{GM}_6(\mathbb{C}^n)$ does indeed have a nondegenerate Gauss map, for $n \geq 2$. Expressions for the tangent spaces may be obtained from deriving curves. Therefore, it suffices to verify that

$$0 = \sum_{i=1}^{m} (\ell_i^5 + 10q_i\ell_i^3 + 15q_i^2\ell_i)h_i + \sum_{i=1}^{m} (\ell_i^4 + 6q_i\ell_i^2 + 3q_i^2)p_i \tag{5}$$

implies $\ell_1 = \ldots = \ell_m = 0 = q_1 = \ldots = q_m$.

Proof Sketch.

We use a *variable splitting trick*: Rewrite the variables as $(X, Y) = (X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2})$, with $n_1 + n_2 = n$, and assume that all $q_i \in \mathbb{R}[X]$, while $\ell_i \in \mathbb{R}[Y]$. Then, (5) splits into a system of 7 equations, which correspond to the parts of degree $0, \ldots, 6$ in X. At the same time, the $h_i = h_{i,X} + h_{i,Y}$ split into an X-part and a Y-part, while $p_i = p_{i,X} + p_{i,X,Y} + p_{i,Y}$ split into a pure X-part, a pure Y-part, and a part bilinear in (X, Y). Looking at the degree-6 part in X, one gets

$$0 = \sum_{i=1}^{m} q_i^2 p_{i,X}. \tag{6}$$

Known results about the Hilbert series of the ideal $(q_1^2, \ldots, q_m^2)$, cf. [3], allow to conclude from (6) that $p_{1,X} = \ldots = p_{m,X} = 0$, if $m = \Theta(n_1^4)$. Looking at the part of degree 5 in X, which is

$$0 = 15\sum_{i=1}^{m} q_i^2\ell_i h_{i,Y} + 3\sum_{i=1}^{m} q_i^2 p_{i,Y}, \tag{7}$$

a similar argument allows to conclude $-5\ell_i h_{i,Y} = p_{i,Y}$, if $m = \Theta(n_1^4)$. We continue by plugging in this identity into the part of degree 1 in X. One obtains

$$0 = \sum_{i=1}^{m} (\ell_i^5 h_{i,Y} + \ell_i^4 p_{i,Y}) = -4\sum_{i=1}^{m} \ell_i^5 h_{i,Y} \tag{8}$$

It follows $h_{i,Y} = 0$, if $m = \Theta(n_2^5)$, by the Alexander-Hirschowitz theorem. Thus, also $p_{i,Y} = 0$ for all i. With similar arguments, one may resolve the rest of this system and see that the remaining X-homogeneous parts of the h_i and p_i must be zero:

Indeed, the two parts of degree 5 and 1 in X yield equations

$$0 = \sum_{i=1}^{m} q_i^2(15\ell_i h_{i,X} + 3p_{i,X,Y}), \quad 0 = \sum_{i=1}^{m} \ell_i^4(\ell_i h_{i,X} + p_{i,X,Y}) \tag{9}$$

The left equation in (9) yields that $p_{i,X,Y} = -5\ell_i h_{i,X}$ for all i, as long as $m = \Theta(n_1^4)$. Plugged into the right equation, it resolves to $h_{i,X} = 0 = p_{i,X}$, as long as $m = \Theta(n_2^5)$. This concludes the proof. Hence, the tangent spaces are skew. □

Conclusion

Let $X = (X_1, \ldots, X_n)$. For $\mu \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n\times n}$ positive definite, write

$$s_6(\mu, \Sigma) := (\mu^T X)^6 + 15(X^T\Sigma X)(\mu^T X)^4 + 45(X^T\Sigma X)^2(\mu^T X)^2 + 15(X^T\Sigma X)^3 \tag{10}$$

for the sextic form of degree-6 moments of the Gaussian distribution $\mathcal{N}(\mu, \Sigma)$. With the aforementioned techniques, we obtain the following identifiability theorem for Gaussian mixtures from their degree-6 moments.

Theorem

For some $m = \Theta(n^4)$, general linear forms $\mu_1, \ldots, \mu_m \in \mathbb{R}^n$ and general positive definite covariance matrices $\Sigma_1, \ldots, \Sigma_m \in \mathbb{R}^{n\times n}$, denote by

$$f = \frac{1}{m}s_6(\mu_1, \Sigma_1) + \ldots + \frac{1}{m}s_6(\mu_m, \Sigma_m) \tag{11}$$

the form of degree-6 moments of the Gaussian mixture parametrized by the μ_i and Σ_i.

Then, there is only one mixture of (at most) m Gaussians, that has f as their degree-6 moments. Precisely, if $\nu_1, \ldots, \nu_m \in \mathbb{R}^n$ and $T_1, \ldots, T_m \in \mathbb{R}^{n\times n}$ are symmetric matrices such that

$$f = \frac{1}{m}s_6(\nu_1, T_1) + \ldots + \frac{1}{m}s_6(\nu_m, T_m) \tag{12}$$

then, up to permutation, $\mu_i = \nu_i$ and $\Sigma_i = T_i$, for all $i \in \{1, \ldots, m\}$.

Numerical experiments

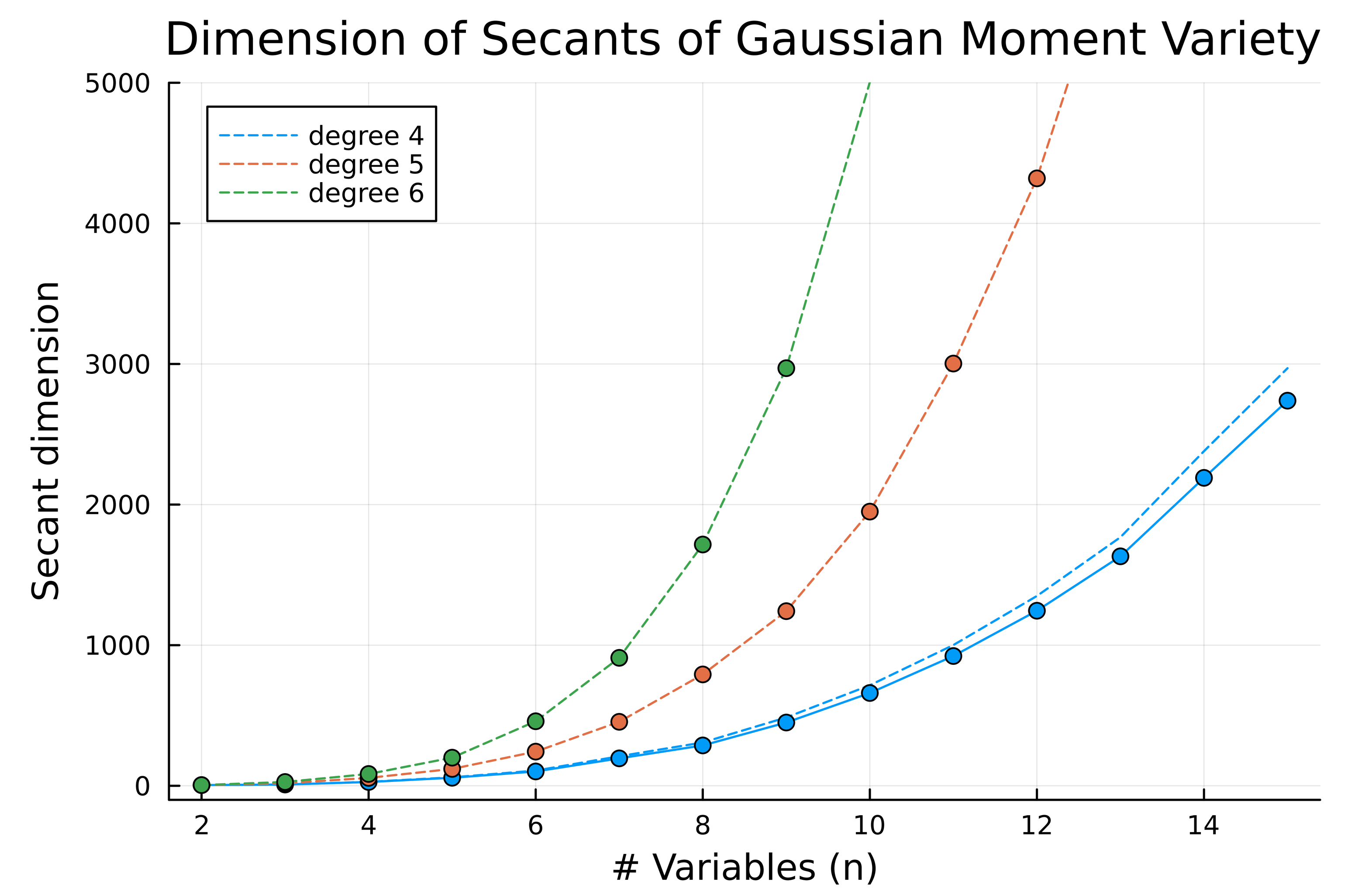


Figure 2. The blue, red, green points correspond to the dimensions of the secant variety $\sigma_m \mathrm{GM}_d(\mathbb{C}^n)$ for $d = 4, 5, 6$, respectively, and several small values of n. We always choose $m = \lfloor \frac{\dim \mathbb{C}[X]_d}{\binom{n+1}{2}+n} \rfloor$. This is the bound for identifiability obtained from counting parameters. The dashed lines correspond to the expected dimensions.

Numerical results, collected in Figure 2, suggest that secants to the Gaussian moment varieties of degree 5 and 6 are always nondefective, up to the rank bound obtained from counting parameters. The defect of $\sigma_m \mathrm{GM}_4(\mathbb{C}^n)$ is at least $\binom{m}{2}$, due to certain Koszul syzygies in the expression for the tangent space. This defect is the difference between the blue dashed line and the blue non-dashed line. In particular, all nontrivial secants of $\mathrm{GM}_4(\mathbb{C}^n)$ are defective.

References

[1] Alexander Taveira Blomenhofer. *Gaussian Mixture Separation and Denoising on Parameterized Varieties*. Doctoral thesis, Universität Konstanz, Konstanz, 2022.

[2] Alex Massarenti and Massimiliano Mella. Bronowski's conjecture and the identifiability of projective varieties, 2022.

[3] Gleb Nenashev. A note on Fröberg's conjecture for forms of equal degrees. *Comptes Rendus Mathematique*, 355(3):272–276, 2017.

[4] Alexander Taveira Blomenhofer. Identifiability of Gaussian mixtures from sixth order moments. WIP.

Author supported by NWO grant OCENW.GROOT.2019.015 (OPTIMAL)