Complete quadrics: Schubert calculus for Gaussian models and semidefinite programming

Mateusz Michałek

University of Konstanz

Joint work with: Laurent Manivel, Leonid Monin, Tim Seynnaeve, Martin Vodicka and Jarosław Wiśniewski

False coin

False coin: 60% vs. 40%
How to detect?

False coin

False coin: 60% vs. 40%
How to detect?
Throw 100 times: 57 heads, 43 tails

False coin

False coin: 60% vs. 40%
How to detect?
Throw 100 times: 57 heads, 43 tails
What is the probability of this result for two cases?
$1.75 \cdot 10^{-30} \cdot \binom{100}{57}$ vs. $6 \cdot 10^{-33} \cdot \binom{100}{57}$

False coin

False coin: 60% vs. 40%

How to detect?

Throw 100 times: 57 heads, 43 tails

What is the probability of this result for two cases?

$1.75 \cdot 10^{-30} \cdot \binom{100}{57}$ vs. $6 \cdot 10^{-33} \cdot \binom{100}{57}$

We maximize the likelihood to find the correct probability distribution

False coin

False coin: 60% vs. 40%
How to detect?
Throw 100 times: 57 heads, 43 tails
What is the probability of this result for two cases?
$1.75 \cdot 10^{-30} \cdot \binom{100}{57}$ vs. $6 \cdot 10^{-33} \cdot \binom{100}{57}$
We maximize the likelihood to find the correct probability distribution

Example$\rightarrow$ statistical model
two points$\rightarrow$ variety

Gaussian model

- one dimensional Gaussian distribution: mean μ and variance $\Sigma > 0$

Gaussian model

- one dimensional Gaussian distribution: mean μ and variance $\Sigma > 0$
- Multivariate Gaussian distribution on $\mathbb{R}^n$:

$$f_{\mu,\Sigma}(x) := \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)},$$

Here $\mu \in \mathbb{R}^n$ and Σ symmetric PD matrix

Gaussian model

- one dimensional Gaussian distribution: mean μ and variance $\Sigma > 0$
- Multivariate Gaussian distribution on $\mathbb{R}^n$:

$$f_{\mu,\Sigma}(x) := \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)},$$

Here $\mu \in \mathbb{R}^n$ and Σ symmetric PD matrix

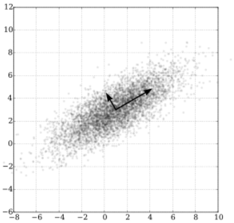

image from wiki

Linear concentration models

Concentration matrix $K = \Sigma^{-1}$

Linear concentration models

Concentration matrix $K = \Sigma^{-1}$

- Anderson 1970: model $K \in \mathcal{L} \subset S^2\mathbb{R}^n = S^2V$ for a linear space

Linear concentration models

Concentration matrix $K = \Sigma^{-1}$

- Anderson 1970: model $K \in \mathcal{L} \subset S^2\mathbb{R}^n = S^2V$ for a linear space

Our variety/model:

$$\mathcal{L}^{-1} = \{K^{-1} : K \in \mathcal{L}\} \subset S^2V^*$$

Linear concentration models

Concentration matrix $K = \Sigma^{-1}$

- Anderson 1970: model $K \in \mathcal{L} \subset S^2\mathbb{R}^n = S^2V$ for a linear space

Our variety/model:

$$\mathcal{L}^{-1} = \{K^{-1} : K \in \mathcal{L}\} \subset S^2V^*$$

Data gives: sample covariance matrix Σ_0

Linear concentration models

Concentration matrix $K = \Sigma^{-1}$

- Anderson 1970: model $K \in \mathcal{L} \subset S^2\mathbb{R}^n = S^2V$ for a linear space

Our variety/model:

$$\mathcal{L}^{-1} = \{K^{-1} : K \in \mathcal{L}\} \subset S^2V^*$$

Data gives: sample covariance matrix Σ_0
Main question: Which $\Sigma \in \mathcal{L}^{-1}$ maximizes the likelihood?

Linear concentration models

Concentration matrix $K = \Sigma^{-1}$

- Anderson 1970: model $K \in \mathcal{L} \subset S^2\mathbb{R}^n = S^2V$ for a linear space

Our variety/model:

$$\mathcal{L}^{-1} = \{K^{-1} : K \in \mathcal{L}\} \subset S^2V^*$$

Data gives: sample covariance matrix Σ_0
Main question: Which $\Sigma \in \mathcal{L}^{-1}$ maximizes the likelihood?

Theorem

Let $\pi : \mathbb{P}(S^2V^*) \dashrightarrow \mathbb{P}\left(S^2V^*/(\mathcal{L}^\perp)\right)$.
There is a unique PD $\Sigma \in \mathcal{L}^{-1}$ such that $\pi(\Sigma) = \pi(\Sigma_0)$. This is the MLE.

Geometric setting

$$\mathbb{P}(S^2V) \supset \mathcal{L} \dashrightarrow \mathcal{L}^{-1} \subset \mathbb{P}(S^2V^*) \dashrightarrow \mathbb{P}\left(S^2V^*/(\mathcal{L}^{\perp})\right)$$

Our interest: fibers of $\pi_{|\mathcal{L}^{-1}}$

Definition

The maximum likelihood degree is the degree of the (finite) map $\pi_{\mathcal{L}^{-1}}$. For $\mathcal{L}$ general, the ML-degree depends on: $d = \dim \mathcal{L}$ and n. It is denoted by $\phi(n, d)$.

Geometric setting

$$\mathbb{P}(S^2V) \supset \mathcal{L} \dashrightarrow \mathcal{L}^{-1} \subset \mathbb{P}(S^2V^*) \dashrightarrow \mathbb{P}\left(S^2V^*/(\mathcal{L}^\perp)\right)$$

Our interest: fibers of $\pi_{|\mathcal{L}^{-1}}$

Definition

The maximum likelihood degree is the degree of the (finite) map $\pi_{\mathcal{L}^{-1}}$. For $\mathcal{L}$ general, the ML-degree depends on: $d = \dim \mathcal{L}$ and n. It is denoted by $\phi(n, d)$.

Theorem (Teissier)

For general $\mathcal{L}$ we have $\mathcal{L}^{-1} \cap \mathcal{L}^\perp = \emptyset$.

Geometric setting

$$\mathbb{P}(S^2V) \supset \mathcal{L} \dashrightarrow \mathcal{L}^{-1} \subset \mathbb{P}(S^2V^*) \dashrightarrow \mathbb{P}\left(S^2V^*/(\mathcal{L}^\perp)\right)$$

Our interest: fibers of $\pi_{|\mathcal{L}^{-1}}$

Definition

The maximum likelihood degree is the degree of the (finite) map $\pi_{\mathcal{L}^{-1}}$. For $\mathcal{L}$ general, the ML-degree depends on: $d = \dim \mathcal{L}$ and n. It is denoted by $\phi(n, d)$.

Theorem (Teissier)

For general $\mathcal{L}$ we have $\mathcal{L}^{-1} \cap \mathcal{L}^\perp = \emptyset$.

Corollary

$\phi(n, d) = \deg \mathcal{L}^{-1}$

$$
\begin{tikzcd}
 & CQ \arrow[dl] \arrow[d] \arrow[dr] & \\
 & \Delta \subset \mathbb{P}(S^2V) \times \mathbb{P}(S^2V^*) \arrow[dl] \arrow[dr] & \\
\mathbb{P}(S^2V) \arrow[rr, dashed] & & \mathbb{P}(S^2V^*)
\end{tikzcd}
$$

understand $\phi(n, \cdot) \Leftrightarrow$ understand the cohomology class $[\Delta]$

$$
\begin{tikzcd}
 & CQ \arrow[dl] \arrow[d] \arrow[dr] & \\
 & \Delta \subset \mathbb{P}(S^2V) \times \mathbb{P}(S^2V^*) \arrow[dl] \arrow[dr] & \\
\mathbb{P}(S^2V) \arrow[rr, dashed] & & \mathbb{P}(S^2V^*)
\end{tikzcd}
$$

understand $\phi(n,\cdot) \Leftrightarrow$ understand the cohomology class $[\Delta]$
Idea: look at a resolution CQ

Complete Quadrics

CQ: closure of the image of the set of invertible matrices under the map

$$\varphi : \mathbb{P}(S^2V) \dashrightarrow \mathbb{P}\left(S^2V\right) \times \mathbb{P}\left(S^2(\bigwedge^2 V)\right) \times \cdots \times \mathbb{P}\left(S^2(\bigwedge^{n-1} V)\right),$$

sending a matrix A to $(A, \bigwedge^2 A, \ldots, \bigwedge^{n-1} A)$.

Complete Quadrics

CQ: closure of the image of the set of invertible matrices under the map

$$\varphi : \mathbb{P}(S^2V) \dashrightarrow \mathbb{P}\left(S^2V\right) \times \mathbb{P}\left(S^2(\bigwedge^2 V)\right) \times \cdots \times \mathbb{P}\left(S^2(\bigwedge^{n-1} V)\right),$$

sending a matrix A to $(A, \bigwedge^2 A, \ldots, \bigwedge^{n-1} A)$.
Equivalently: blow-up of rank 1,2,$\ldots$, $n-2$ matrices

Complete Quadrics

CQ: closure of the image of the set of invertible matrices under the map

$$\varphi : \mathbb{P}(S^2V) \dashrightarrow \mathbb{P}\left(S^2V\right) \times \mathbb{P}\left(S^2(\bigwedge^2 V)\right) \times \cdots \times \mathbb{P}\left(S^2(\bigwedge^{n-1} V)\right),$$

sending a matrix A to $(A, \bigwedge^2 A, \ldots, \bigwedge^{n-1} A)$.
Equivalently: blow-up of rank 1,2,..., $n-2$ matrices
Two types of divisors:

- L_i: pull-back of $H_i \subset \mathbb{P}\left(S^2(\bigwedge^i V)\right)$
- S_i: exceptional divisor from rank i matrices

Complete Quadrics

CQ: closure of the image of the set of invertible matrices under the map

$$\varphi : \mathbb{P}(S^2V) \dashrightarrow \mathbb{P}\left(S^2V\right) \times \mathbb{P}\left(S^2(\bigwedge^2 V)\right) \times \cdots \times \mathbb{P}\left(S^2(\bigwedge^{n-1} V)\right),$$

sending a matrix A to $(A, \bigwedge^2 A, \dots, \bigwedge^{n-1} A)$.
Equivalently: blow-up of rank 1,2,..., $n-2$ matrices
Two types of divisors:

- L_i: pull-back of $H_i \subset \mathbb{P}\left(S^2(\bigwedge^i V)\right)$
- S_i: exceptional divisor from rank i matrices

$$\phi(n,d) = L_1^{\binom{n+1}{2}-d-1} L_{n-1}^d$$

Theorem (Schubert)

The classes $L_1, \ldots, L_{n-1}$ *form a basis of* $\mathrm{Pic}(CQ(V))$, *in which the classes* $S_1, \ldots, S_{n-1}$ *are given by the relations*

$$S_i = -L_{i-1} + 2L_i - L_{i+1},$$

with $L_0 = L_n := 0$.

- L_i: twice the fundamental roots
- S_i: twice the simple positive roots

$$\begin{aligned} \phi(n,d) &= L_1^{\binom{n+1}{2}-d-1} L_{n-1}^d \\ &= \frac{1}{n} \sum_{s=1}^{n-1} L_1^{\binom{n+1}{2}-d-2} L_{n-1}^d S_{n-s} \end{aligned}$$

$\mathbb{P}(S^2\mathcal{U}) \times_{G(r,n)} \mathbb{P}(S^2\mathcal{Q}^*)$ model of S_r

$L_1^{\binom{n+1}{2}-d-2} L_{n-1}^d S_{n-s} \to$
intersection theory on $\mathbb{P}(S^2\mathcal{U}) \times_{G(r,n)} \mathbb{P}(S^2\mathcal{Q}^*) \to$
intersection theory on $G(r,n)$

$$\mathbb{P}(S^2\mathcal{U}) \times_{G(r,n)} \mathbb{P}(S^2\mathcal{Q}^*) \text{ model of } S_r$$

$L_1^{\binom{n+1}{2}-d-2} L_{n-1}^d S_{n-s} \rightarrow$
intersection theory on $\mathbb{P}(S^2\mathcal{U}) \times_{G(r,n)} \mathbb{P}(S^2\mathcal{Q}^*) \rightarrow$
intersection theory on $G(r,n)$
Main advantage: we may use a very well developed theory of Pragacz, Lascoux and others
Crucial role: Segre class of symmetric square of a bundle

Definition

For I be a set of integers of cardinality r let:

$$s_{(d)}(\{x_i + x_j \mid 1 \leq i \leq j \leq r\}) = \sum_{\substack{\lambda(I) \vdash d \\ \#I = r}} \psi_I s_{\lambda(I)}(x_1, \dots, x_r).$$

Equivalently:

$$Seg_d(S^2\mathcal{U}) = \sum_{\substack{\lambda(I) \vdash d \\ \#I = r}} \psi_I \sigma_{\lambda(I)},$$

where σ_λ denote the Schubert classes in the Chow ring of the Grassmannian.
In both

$$\lambda(I) := (i_r - (r-1), i_{r-1} - (r-2), \dots, i_2 - 1, i_1).$$

Example

Consider $G(2,4)$. $\mathcal{U}$ has two Chern roots x_1, x_2.

$$x_1 + x_2 = -\ydiagram{1}, \quad x_1 \cdot x_2 = \ydiagram{1,1}$$

Chern roots of $S^2\mathcal{U}$ are $2x_1, x_1 + x_2, 2x_2$. Three respective Chern classes:

$$-3\,\ydiagram{1}, \quad 2\,\ydiagram{2} + 6\,\ydiagram{1,1}, \quad -4\,\ydiagram{2,1}.$$

By inverting the Chern polynomial we obtain the Segre classes:

$$3\,\ydiagram{1}, \quad \mathbf{7}\,\ydiagram{2} + 3\,\ydiagram{1,1}, \quad 10\,\ydiagram{2,1}, \quad 10\,\ydiagram{2,2}.$$

Their coefficients are the Lascoux coefficients, namely:

$$\psi_{0,2} = 3, \quad \psi_{0,3} = \mathbf{7}, \quad \psi_{1,2} = 3, \quad \psi_{1,3} = 10, \quad \psi_{2,3} = 10.$$

Example

Equivalently:

$$s_{(2)}(2x_1, x_1 + x_2, 2x_2) = 7x_1^2 + 7x_2^2 + 10x_1x_2 =$$

$$= 7(x_1^2 + x_1x_2 + x_2^2) + 3x_1x_2 = \mathbf{7}s_{(2,0)}(x_1, x_2) + 3s_{(1,1)}(x_1, x_2).$$

Central results

- Pfaffian formulas for ψ_I by Pragacz
- Recursive formulas for ψ_I by Pragacz and Laksov, Lascoux, Thorup

Theorem (Bothmer, Ranestad)

$$S_r L_1^{\binom{n+1}{2}-m-1} L_{n-1}^{m-1} = \sum_{\substack{I\subset[n] \\ \#I=n-r \\ \sum I=m-n+r}} \psi_I \psi_{[n]\setminus I}$$

Lemma

$\psi_{[n]\setminus I}$ *is a polynomial in* n

Corollary (Conjectured by Sturmfels and Uhler)

For any d *the function* $\phi(n,d)$ *is a polynomial in* n.

$\phi(n, 1) = n - 1$
$\phi(n, 2) =$
$\phi(n, 3) =$

$$\phi(n,1) = n-1$$
$$\phi(n,2) = (n-1)^2$$
$$\phi(n,3) =$$

$$\phi(n,1) = n-1$$
$$\phi(n,2) = (n-1)^2$$
$$\phi(n,3) = (n-1)^3$$

$\phi(n,1) = n-1$
$\phi(n,2) = (n-1)^2$
$\phi(n,3) = (n-1)^3 - \deg$ base locus$= \frac{1}{6}(5n-3)(n-1)(n-2)$

$\phi(n, 1) = n - 1$
$\phi(n, 2) = (n - 1)^2$
$\phi(n, 3) = (n - 1)^3 -$ deg base locus$= \frac{1}{6}(5n - 3)(n - 1)(n - 2)$

Theorem (Stückrad/ Chardin, Eisenbud, Ulrich)

$\phi(n, 4) = \frac{1}{12}(n - 1)(n - 2)(7n^2 - 19n + 6)$

$\phi(n,1) = n-1$
$\phi(n,2) = (n-1)^2$
$\phi(n,3) = (n-1)^3 -$ deg base locus $= \frac{1}{6}(5n-3)(n-1)(n-2)$

Theorem (Stückrad/ Chardin, Eisenbud, Ulrich)

$\phi(n,4) = \frac{1}{12}(n-1)(n-2)(7n^2-19n+6)$

1	2	4	4	2	1	0	0		
1	3	9	17	21	21	17	$\cdots$		
1	4	16	44	86	137	188	212	188	$\cdots$

Consequence:
How many quadrics in n variables pass through d (general) points and are tangent to $\binom{n+1}{2} - d - 1$ (general) hyperplanes?

Consequence:
How many quadrics in n variables pass through d (general) points and are tangent to $\binom{n+1}{2} - d - 1$ (general) hyperplanes?
For fixed d the answer is a polynomial in n.

$$\phi(n,17) = \frac{1}{355687428096000}(n-5)(n-4)(n-3)(n-2)(n-1)$$
$$(3024902557n^{12} - 111489409997n^{11} + 1862235028288n^{10} -$$
$$18676382506290n^{9} + 125446336704681n^{8} - 594987544526781n^{7} +$$
$$2047718727437714n^{6} - 5214795516381220n^{5} + 10138037306327912n^{4}$$
$$- 15696938913831072n^{3} + 18622763914779648n^{2}$$
$$- 12286614789872640n + 2964061900800)$$

Further results

- Types A and D
- Proof of NRS conjecture

$$S_{n-s}L_1^{\binom{n+1}{2}-m-1}L_{n-1}^{m-1} =$$

$$\sum_{\sum I \leq m-s, |I|=s} (-1)^{m-s-\sum I} \psi_I b_I(n) \binom{m-1}{m-s-\sum I}$$

- Explicit formulas in terms of dimensions of representations

Future directions

- Does the log-concavity of the coefficients of $\phi(\cdot, d)$ suggest some cohomology theory on infinite dimensional algebraic varieties?
- What is the degree of the dual variety to: matrices of fixed rank intersected with a space of fixed dimension?
- What if we intersect other cohomology classes with $L_{n-1}^{\binom{n+1}{2}-d}$?
- Can we define noncommutative matroids?
- What about graphical Gaussian models?
- What about linear covariance models?

Thank you!

Bibliography

- Michałek, Monin and Wiśniewski. Maximum likelihood degree and space of orbits of a $\mathbb{C}^*$ action. arXiv:2004.07735 (2020).
- Manivel, Michałek, Monin, Seynnaeve, Vodicka. Complete quadrics: Schubert calculus for Gaussian models and semidefinite programming. arXiv:2011.08791 (2020).

Bibliography

- Graf von Bothmer, Ranestad. A general formula for the algebraic degree in semidefinite programming. Bulletin of the London Mathematical Society 41.2 (2009): 193-197.
- Laksov, Lascoux, and Thorup. On Giambelli's theorem on complete correlations. Acta Mathematica 162.1 (1989): 143-199.
- Nie, Ranestad, and Sturmfels. The algebraic degree of semidefinite programming. Mathematical Programming 122.2 (2010): 379-405.
- Pragacz. Enumerative geometry of degeneracy loci. Annales scientifiques de l'École Normale Supérieure. Vol. 21. No. 3. 1988.
- Pragacz. Symmetric polynomials and divided differences in formulas of intersection theory. Banach Center Publications 36.1 (1996): 125-177.
- Sturmfels, Uhler. Multivariate Gaussians, semidefinite matrix completion, and convex algebraic geometry. Annals of the Institute of Statistical Mathematics 62.4 (2010): 603-638.