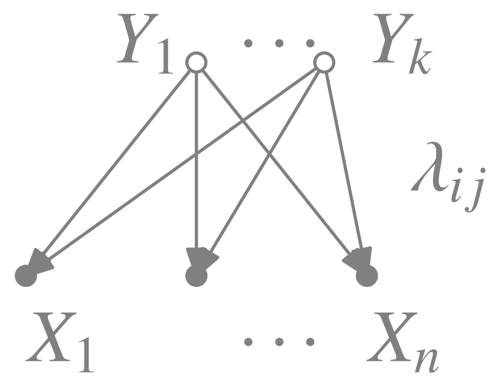
Applied algebraic geometry: algebraic statistics and optimisation

Jan Draisma, Universität Bern

Latent variables

Example (*k*-factor model)



$\Sigma_X = \Lambda^T D\lambda + D'$ where D, D' diagonal

Forget about positive-definiteness (think of the Λ, D, D' as variables).

$$M_{k,n} := \overline{\{\Sigma = \Lambda^T \Lambda + D \mid \Lambda \in \mathbb{C}^{k\times n}, D \text{ diagonal}\}}.$$

Question (Drton-Sturmfels-Sullivant)
Equations for $M_{k,n}$?

Equations for the k-factor model

Obvious equations: $(k+1) \times (k+1)$-subdeterminants outside the diagonal, as well as $\sigma_{ij} - \sigma_{ji}$.

For $k = 1$ these generate the ideal.

For $k = 2$ there is an additional type of equation: $\sum_{\pi \in S_5} sgn(\pi)\pi \cdot \sigma_{12}\sigma_{23}\sigma_{34}\sigma_{45}\sigma_{51}$, the *pentad*. Pentads and off-diagonal 3×3-determinants generate the ideal of $M_{2,n}$.

For general k, there is the following existence result:

Theorem

For each fixed k, there exists an n_0 such that for $n \geq n_0$ the variety $M_{k,n}$ is the zero set of the equations coming from M_{k,n_0} by simultaneously permuting rows and columns.

Noetherianity up to symmetry

Recall

A top space X is *Noetherian* if every descending chain $X \supseteq X_1 \supseteq \cdots$ of closed subsets stabilises, i.e., $X_n = X_{n+1}$ for $n \gg 0$.

Definition

Suppose a group G acts by homeomorphisms on a space X. Then call X G-Noetherian if every chain $X \supseteq X_1 \supseteq \cdots$ of G-stable closed subsets stabilises.

Fundamental example

The space $\mathbb{C}^{k\times\mathbb{N}}$ with the Zariski topology is $\mathrm{Sym}(\mathbb{N})$-Noetherian. (Follows from our earlier results on sequences of ideals.)

Constructions

G-stable subspaces, and G-equivariant images of G-Noetherian topological spaces are G-Noetherian. Also, if $Y \subseteq X$ is H-Noetherian for some $H \subseteq G$, then $\bigcup_{g\in G} gY$ is G-Noetherian.

Proof sketch

Pass to an infinite-dimensional limit:
$M_{k,\mathbb{N}} := \{\Sigma = \Lambda^T\Lambda + D \mid \Lambda \in \mathbb{C}^{k\times\mathbb{N}}, D \text{ diagonal}\} \subseteq \mathbb{C}^{\mathbb{N}\times\mathbb{N}}$. This is stable under $\mathrm{Sym}(\mathbb{N})$.

Let Q_k be the variety of $\mathbb{N}\times\mathbb{N}$-matrices all of whose off-diagonal $(k+1)\times(k+1)$-minors vanish. So $M_{k,\mathbb{N}} \subseteq Q_k$, and also Q_k is $\mathrm{Sym}(\mathbb{N})$-stable.

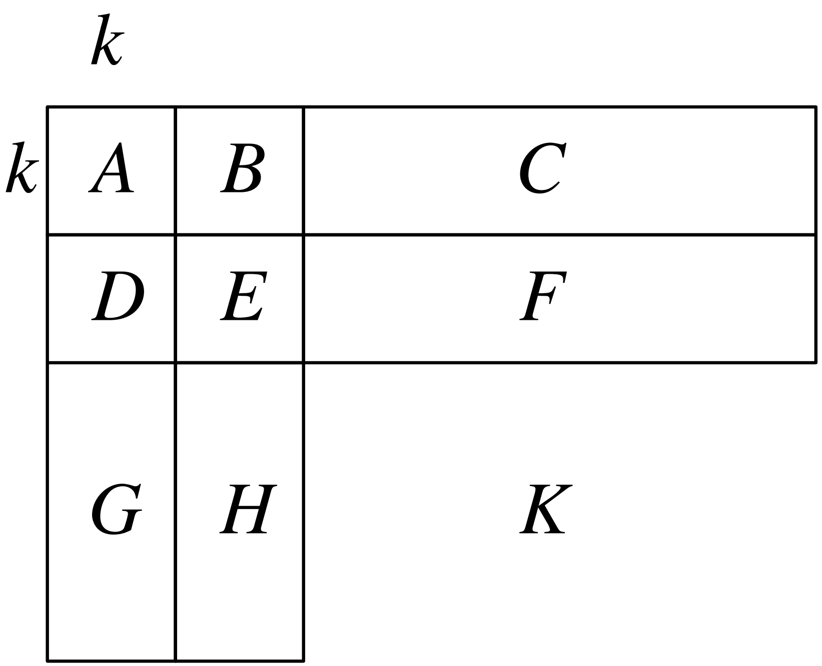
By induction on k we prove that Q_k is $\mathrm{Sym}(\mathbb{N})$-Noetherian.

For $k = 0$, $Q_{0,n}$ is the space of diagonal matrices, and the result follows from Noetherianity of $\mathbb{C}^{1\times\mathbb{N}}$.

Equations for the k-factor model

Suppose the claim is true for $k-1$. Write $Q_{k,n} = Q_{k-1,n} \cup Z$ where Z is the open subset with at least one off-diagonal $k \times k$-subdeterminant $\neq 0$. Then $Z = \bigcup_{g \in \mathrm{Sym}(\mathbb{N})} gZ'$ where Z' is the open subset where the $\{1, \ldots, k\} \times \{k+1, \ldots, 2k\}$-subdet is nonzero.

Up to permutations, elements of Z' are of the following form:



where $\det(B) \neq 0$, and each off-diagonal element of K is determined by B, C, H.

The space with coordinates A, B, C, D, E, F, G and the diagonal elements of H is Noetherian under Sym(N) where $N := \mathbb{N} \setminus \{1, \ldots, k\}$. (It is $\mathbb{C}^{(4k+1)\times N}$ times a finite-dimensional space.)

It follows that Z' is Sym(N)-Noetherian, and Z is Sym($\mathbb{N}$)-Noetherian, and so is Q_k. $\square$

(The corresponding *ideal-theoretic* statement is widely open!
And so is the value of n_0, which could be as small as $2k + 2$.)

New topic: Optimisation

Typical goal: maximise $f(x)$ subject to $x \in X$ (1)

Here $X \subseteq \mathbb{R}^n$ is a compact set given by polynomial weak inequalities (and equalities) and f is a polynomial function.

We'll discuss an approach due to Lasserre.

Write $R = \mathbb{C}[x_1, \ldots, x_n]$ and write R_d for poly of degree $\leq d$.

Simple idea: think of points in X as linear functions on R. This linear function $\mathrm{ev}_x : R \to \mathbb{R}, h \mapsto h(x)$ satisfies:
1. $\mathrm{ev}(1) = 1$
2. $\mathrm{ev}_x(h) \geq 0$ for all $h \in R$ that are nonnegative everywhere on X.

We find that the maximum we're trying to find has the *upper bound* the following maximum:

Relaxation

$\sup \ell(f)$ where $\ell : R \to \mathbb{R}$ is linear and satisfies $\ell(1) = 1$, $\ell(h) \geq 0$ for all h nonnegative on X. (2)

Proposition

In fact, this sup is attained at ev_x for some $x \in X$.

Proof: First, note that ℓ is automatically continuous in the ∞-norm: if $|h - h'|_\infty < \epsilon$ then $\epsilon - l(h - h') = l(\epsilon - (h - h')) \geq 0$, etc. By Stone-Weierstrass, since R separates points of X, R is dense in $C^0(X, R)$. Hence ℓ extends uniquely to a continuous linear functional $\ell : C^0(X, R) \to \mathbb{R}$ with the same properties. (If h is continous and nonnegative on X, approximate $k := h + \epsilon 1$ by a polynomial p with $|p - k|_\infty \leq \epsilon$ and so that $|\ell(p - k)| \leq \epsilon$. Then $\ell(p) \geq 0$ implies $\ell(h) \geq -2\epsilon$.)

By the Riesz representation theorem, there is a probability measure μ on X such that $\ell(h) = \int_X h(x)\mathrm{d}\mu(x)$ for all $h \in C^0(X, \mathbb{R})$. Then, $\ell(f) = \int_X f(x)d\mu(x) \leq \max_{x \in X} f(x)$. $\square$

For each $d \geq \deg(f)$ we get the following further:

Relaxation

$\max \ell(f)$ such that $\ell : R_d \to \mathbb{R}$ linear, $\ell(1) = 1$ and $\ell(h) \geq 0$ for all $h \in R_d$ nonnegative on X. $(3)_d$

This is a finite-dimensional linear program, but still with infinitely many, hard-to-handle constraints. Note that the maximum is attained, since for every monomial m the functions $c1 - m$ and $c1 + m$ is nonnegative for $c \gg 0$, so the domain over which ℓ runs is compact. Using Tychonov, find that the max in $(3)_d$ converges to the sup/max in (2).

How to certify positivity?

There are polynomials that are clearly positive everywhere, namely *sums of squares*: $p_1^2 + \ldots + p_k^2$ where the p_i are polynomials. Let $S \subseteq R$ be the set of these.

Now assume $X = \{x \in \mathbb{R}^n \mid g_1(x), \ldots, g_r(x) \geq 0\}$. Set $g_0 := 1$. Define the *quadratic module* generated by the g_i as $M := \sum_{i=0}^{m} S\, g_r$. These polynomials are clearly nonnegative on X.

So it is natural to replace the condition "$\ell(h) \geq 0$ for $h \in R_d$ nonnegative on X" by "$\ell(h) \geq 0$ for $h \in M_d$".

These conditions are equivalent to the symmetric bilinear form β_i on $R_{\lfloor (d-\deg g_i)/2 \rfloor} =: V_{i,d}$ defined by $\beta_i(p, q) := \ell(p g_i q)$ being positive semidefinite for each i.

Infinite LP relaxation

Relaxation (for each $d \geq \deg(f)$) $\max \ell(f)$ where $\ell : R_d \to \mathbb{R}$ linear, such that $\ell(1) = 1$ and $\ell(h) \geq 0$ for all $h \in M_d$. $(4)_d$

Archimedean assumption

Assume that there exists a $u \in M$ such that $\{x \in \mathbb{R}^n \mid u(x) \geq 0\}$ is compact.

Theorem (Schmüdgen, Putinar, Jacobi)

Under this assumption, each polynomial p *strictly positive* on X lies in M.

It follows that the optimal value of $(4)_d$ for $d \to \infty$ still converges to the actual optimal value of (1).

But the number of linear conditions on ℓ is still infinite.

The condition $\ell(h) \geq 0$ for each $h \in M_d$ is equivalent to $\ell(p^2 g_i) \geq 0$ for each $p \in R_{\lfloor (d-\deg(g_i))/2 \rfloor} =: P_{i,d}$ and each $i = 1, \ldots, r$.

Let β_i be the symmetric bilinear form on $P_{i,d}$ defined by $\beta_i(p, q) := \ell(pg_iq)$. Then the above is equivalent to β_i being *positive semidefinite*.

This leads to the following optimisation problem:

Lasserre's hierarchy

(For any $d \geq 2 \deg f$): $\max \beta_0(f, 1)$ where $\beta_0, \ldots, \beta_r$ are symmetric bilinear forms on $P_{0,d}, \ldots, P_{r,d}$ satisfying $(4)_d$

1. $\beta_0(1, 1) = 1$ and
2. $\beta_i(p, q) = \beta_j(p', q')$ for all i, j, p, q such that $pg_iq = p'g_jq'$ and
3. β_i is positive semidefinite.

Conditions 2,3 imply that there is a linear $\ell : R_d \to \mathbb{R}$ with $\beta_i(p, q) = \ell(pg_iq)$ for all i and $p, q \in P_{i,d}$ and with $\ell(M_d) \subseteq \mathbb{R}_{\geq 0}$.

This is a *semidefinite program*: except for the positive semidefiniteness condition, the conditions on the β_i are affine-linear, and we're optimising a linear function. The general form of a semidefinite program is as follows: *max* $\ell(t)$ *s.t.* $A_0 + t_1 A_1 + \ldots + A_m t_m$ *is positive semidefinite*. Here ℓ is linear and $A_0, \ldots, A_m$ are real symmetric matrices. The above can be put in this form by parameterising the affine space given by the linear conditions on the β_i.

Under the Archimedean assumption, the optimal value of $(4)_d$ converges to the optimal value of (1) as $d \to \infty$.

Remark

One can enforce this assumption by adding the constraint $g_{r+1}(x) := a - \|x\|^2 \geq 0$ for some sufficiently large a so that X is contained in this set.