week 11 day 2

Graphical models: algebra
Algebraic & Geometric Methods in Statistics

Sonja Petrović
Created for Math/Stat 561

Mar 22, 2023.

Agenda

1. Code for generating the minors and the statements from a graph (directed or undirected)
2. Binomials? Markov bases? **Connection**
3. Computing the MLE of an example graph $\mapsto$ homework 5

(a)

(b)

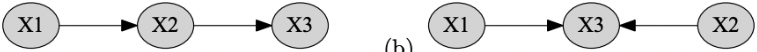


Figure 3.2.3: Directed graphs representing (a) $X_1 \perp\!\!\!\perp X_3 \mid X_2$ and (b) $X_1 \perp\!\!\!\perp X_2$.

Figure 1: Source: Oberwolfach lectures

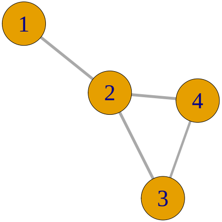
- Here is an incredible online resource: Maathuis, Drton, Lauritzen & Wainwright's Handbook of graphical models

Part One

Code for generating the minors and the statements from a graph (directed or undirected)

Global & Pairwise Markov proprety - algebra

$C_{pairs} = \{1 \perp\!\!\!\perp 3|(2,4), 1 \perp\!\!\!\perp 4|(2,3)\}$.
$C_{global} = C_{pairs} \cup \{1 \perp\!\!\!\perp (3,4)|2\}$.



Question (example)

How many polynomials generate the corresponding CI ideal?

- $M_1 = \begin{bmatrix} p_{0000} & p_{0001} & p_{0010} & p_{0011} \\ p_{1000} & p_{1001} & p_{1010} & p_{1011} \end{bmatrix}$
- $M_2 = \begin{bmatrix} p_{0100} & p_{0101} & p_{0110} & p_{0111} \\ p_{1100} & p_{1101} & p_{1110} & p_{1111} \end{bmatrix}$
- The conditional independence ideal for each statement is generated by two minors of M_1 and two minors of M_2
- The conditional independence ideal for each statement is generated by **all** 2×2 minors M_1 and of M_2

Code for generating the polynomials (minors)

```
R1 = QQ[p_(0,0,0,0)..p_(1,1,1,1)]
M1 = matrix{{p_(0,0,0,0),p_(0,0,0,1),p_(0,0,1,0),p_(0,0,1,1)},
  {p_(1,0,0,0),p_(1,0,0,1),p_(1,0,1,0),p_(1,0,1,1)}}
M2 = matrix{{p_(0,1,0,0),p_(0,1,0,1),p_(0,1,1,0),p_(0,1,1,1)},
  {p_(1,1,0,0),p_(1,1,0,1),p_(1,1,1,0),p_(1,1,1,1)}}

--pairwise Markov property
IP = ideal(det(M1_{0,2}),det(M1_{1,3}),det(M2_{0,2}),
  det(M2_{1,3}),det(M1_{0,1}),det(M1_{2,3}),det(M2_{0,1}),det(M2_{2
--global Markov property
IG = minors(2,M1) + minors(2,M2)
```

Task

Run this code. What is the output? Compare to next page.

- *Reminder:* Look at slides 10 and 11 of **lecture 4** – `M2` code for computing ideals (minors) of given CI statements.
- We can compute the ideal I_G of a graphical model as follows:

```
i97 : loadPackage "GraphicalModels"
o97 = GraphicalModels
i99 : G = graph({{1,2},{2,3},{3,4},{2,4}})
o99 = Graph{1 => {2}       }
            2 => {1, 3, 4}
            3 => {2, 4}
            4 => {2, 3}
i100 : pairMarkov G
o100 = {{{1}, {4}, {2, 3}}, {{1}, {3}, {4, 2}}}
i101 : globalMarkov G
o101 = {{{1}, {3, 4}, {2}}}
-- This method displays only non-redundant statements.
```

. . . package shortcuts!!

```
i103 : R=markovRing(2,2,2,2);
i104 : conditionalIndependenceIdeal (R, pairMarkov(G)) / print;
- p        p          + p        p
   1,1,1,2 2,1,1,1     1,1,1,1 2,1,1,2
- p        p          + p        p
   1,1,2,2 2,1,2,1     1,1,2,1 2,1,2,2
- p        p          + p        p
   1,2,1,2 2,2,1,1     1,2,1,1 2,2,1,2
- p        p          + p        p
   1,2,2,2 2,2,2,1     1,2,2,1 2,2,2,2
- p        p          + p        p
   1,1,2,1 2,1,1,1     1,1,1,1 2,1,2,1
- p        p          + p        p
   1,1,2,2 2,1,1,2     1,1,1,2 2,1,2,2
- p        p          + p        p
   1,2,2,1 2,2,1,1     1,2,1,1 2,2,2,1
- p        p          + p        p
   1,2,2,2 2,2,1,2     1,2,1,2 2,2,2,2
```

Part Two

Binomials? Markov bases? **Connection**

The model of independence is a graphical model

Example 1.2.6 (Independence)**.** Let $\Gamma = [1][2]$. Then the hierarchical model consists of all positive probability matrices $(p_{i_1 i_2})$

$$p_{i_1 i_2} = \frac{1}{Z(\theta)} \theta^{(1)}_{i_1} \theta^{(2)}_{i_2}$$

where $\theta^{(j)} \in (0, \infty)^{r_j}$, $j = 1, 2$. That is, the model consists of all positive rank one matrices. It is the positive part of the model of independence $\mathcal{M}_{X \perp\!\!\!\perp Y}$, or in algebraic geometric language, the positive part of the Segre variety. $\square$

Figure 2: Oberwolfach Lectures

Example 3.1.10. Let X_1, X_2, X_3, X_4 be binary random variables, and consider the conditional independence model

$$\mathcal{C} = \{1 \perp\!\!\!\perp 3 \mid \{2,4\}, 2 \perp\!\!\!\perp 4 \mid \{1,3\}\}.$$

These are the conditional independence statements that hold for the graphical model associated to the four cycle graph with edges $\{12, 23, 34, 14\}$; see Section 3.2. The conditional independence ideal is generated by eight quadratic binomials:

$$\begin{aligned} I_{\mathcal{C}} &= I_{1 \perp\!\!\!\perp 3 \mid \{2,4\}} + I_{2 \perp\!\!\!\perp 4 \mid \{1,3\}} \\ &= \langle p_{1111}p_{2121} - p_{1121}p_{2111}, p_{1112}p_{2122} - p_{1122}p_{2112}, \\ &\quad\; p_{1211}p_{2221} - p_{1221}p_{2211}, p_{1212}p_{2222} - p_{1222}p_{2212}, \\ &\quad\; p_{1111}p_{1212} - p_{1112}p_{1211}, p_{1121}p_{1222} - p_{1122}p_{1221}, \\ &\quad\; p_{2111}p_{2212} - p_{2112}p_{2211}, p_{2121}p_{2222} - p_{2122}p_{2221} \rangle . \end{aligned}$$

The ideal $I_{\mathcal{C}}$ is radical and has nine minimal primes. One of these is a toric ideal I_Γ, namely the vanishing ideal of the hierarchical (and graphical) model associated to the simplicial complex $\Gamma = [12][23][34][14]$. The other eight components are linear ideals whose varieties all lie on the boundary of the probability simplex. In particular, all the irreducible components of the variety $V(I_{\mathcal{C}})$ are unirational. □

Figure 3: Oberwolfach Lectures

Hierarchical log-linear models

Definition [Simplicial complex]

Definition 9.3.1. For a set S, let 2^S denote its power set, that is, the set of all of its subsets. A *simplicial complex* with ground set S is a set $\Gamma \subseteq 2^S$ such that if $F \in \Gamma$ and $F' \subseteq F$, then $F' \in \Gamma$. The elements of Γ are called the *faces* of Γ and the inclusion maximal faces are the *facets* of Γ.

- We will use the bracket notation from the theory of hierarchical log-linear models
- $\Gamma = [12][13][23]$ is the bracket notation for the simplicial complex $\Gamma = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}\}$. The geometric realization of Γ is the boundary of a triangle.
- *Hierarchical models are log-linear models*, so they can be described as $\mathcal{M}_A$ for a suitable matrix A associated to the simplicial complex Γ. *Notation...*

Example 9.3.3 (Independence). Let $\Gamma = [1][2]$. Then the hierarchical model consists of all positive probability matrices $(p_{i_1 i_2})$,

$$p_{i_1 i_2} = \frac{1}{Z(\theta)} \theta^{(1)}_{i_1} \theta^{(2)}_{i_2},$$

where $\theta^{(j)} \in (0, \infty)^{r_j}$, $j = 1, 2$. That is, the model consists of all positive rank one matrices. It is the positive part of the model of independence $\mathcal{M}_{X \perp\!\!\!\perp Y}$, or, in algebraic geometric language, the positive part of the Segre variety. The normalizing constant is

$$Z(\theta) = \sum_{i_1=1}^{r_1} \sum_{i_2=1}^{r_2} \theta_{i_1} \theta_{i_2}.$$

In this case, the normalizing constant factorizes as

$$Z(\theta) = \left(\sum_{i_1=1}^{r_1} \theta_{i_1} \right) \left(\sum_{i_2=1}^{r_2} \theta_{i_2} \right).$$

Complete factorization of the normalizing constant as in this example is a rare phenomenon.

Example 9.3.4 [no-3-factor interaction]

Γ=[12][13][23].
The hierarchical model $\mathcal{M}_\Gamma$ consists of all $r_1 \times r_2 \times r_3$ tables $(p_{i_1 i_2 i_3})$ with:

$$p_{i_1 i_2 i_3} = \frac{1}{Z(\theta)} \theta^{(12)}_{i_1 i_2} \theta^{(13)}_{i_1 i_3} \theta^{(23)}_{i_2 i_3},$$

for some positive real tables $\theta^{(12)} \in (0, \infty)^{r_1 \times r_2}$, $\theta^{(13)} \in (0, \infty)^{r_1 \times r_3}$, and $\theta^{(23)} \in (0, \infty)^{r_2 \times r_3}$.

- In the case of binary random variables, its implicit representation is given by the equation:

$$p_{111} p_{122} p_{212} p_{221} = p_{112} p_{121} p_{211} p_{222}.$$

The log-linear model consists of all positive probability distributions that satisfy this quartic equation.

Example - by hand.

Where are the "A" matrices??

Example 9.3.8 (Sufficient statistics of hierarchical models)**.** Returning to our examples above, for $\Gamma = [1][2]$ corresponding to the model of independence, the minimal sufficient statistics are the row and column sums of $u \in \mathbb{N}^{r_1 \times r_2}$. That is,

$$A_{[1][2]}u \;=\; (u|_1, u|_2).$$

- The recipe is the same as it was for other models on contingency tables! Columns are joint probabilities and rows are parameters:

Example 9.3.9 (Marginals of a 4-way table)**.** Let $\Gamma = [12][14][23]$ and $r_1 = r_2 = r_3 = r_4 = 2$. Then A_Γ is the matrix

$$\begin{array}{c} \\ 11\cdot\cdot \\ 12\cdot\cdot \\ 21\cdot\cdot \\ 22\cdot\cdot \\ 1\cdot\cdot1 \\ 1\cdot\cdot2 \\ 2\cdot\cdot1 \\ 2\cdot\cdot2 \end{array}
\begin{array}{c}
\begin{array}{cccccccccccccccc} 1111 & 1112 & 1121 & 1122 & 1211 & 1212 & 1221 & 1222 & 2111 & 2112 & 2121 & 2122 & 2211 & 2212 & 2221 & 2222 \end{array} \\
\begin{pmatrix}
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1
\end{pmatrix}
\end{array},$$

These matrices are **huge**.

How are we ever going to compute anything, like a Markov basis for exact testing?!??

Citing Dobra 2003:

The statistical theory on graphical models (Madigan and York 1995; Whittaker 1990; Lauritzen 1996) shows that the conditional dependencies induced by a set of fixed marginals among the variables cross-classified in a table of counts can be visualized by means of an independence graph. In particular, a lot of attention has been given to **decomposable** graphs (Lauritzen 1996):

- a special class of graphs that can be '*broken*' into components such that
  1. every *component* is associated with *exactly one fixed marginal*, and
  2. *no information is lost* in the decomposition process, that is, no marginal is 'split' between two components.

Decomposable complexes

- Notation: $|\Gamma|$ = ground set of the complex Γ (the union of all faces).

Definition [decomposable complex] defn. 9.3.11.

A simplicial complex Γ is **reducible**, with reducible **decomposition** Γ_1, S, Γ_2 and **separator** $S \subset |\Gamma|$ if it satisfies $\Gamma = \Gamma_1 \cup \Gamma_2$ and $\Gamma_1 \cap \Gamma_2 = S$.
Furthermore, we assume here that neither Γ_1 nor Γ_2 is $= S$.
A simplicial complex is **decomposable** if it is reducible and Γ_1 and Γ_2 are decomposable or simplices.

Examples

- $[1][2] =$

Decomposable complexes

- Notation: $|\Gamma|$ = ground set of the complex Γ (the union of all faces).

Definition [decomposable complex] defn. 9.3.11.

A simplicial complex Γ is **reducible**, with reducible **decomposition** Γ_1, S, Γ_2 and **separator** $S \subset |\Gamma|$ if it satisfies $\Gamma = \Gamma_1 \cup \Gamma_2$ and $\Gamma_1 \cap \Gamma_2 = S$.
Furthermore, we assume here that neither Γ_1 nor Γ_2 is $= S$.
A simplicial complex is **decomposable** if it is reducible and Γ_1 and Γ_2 are decomposable or simplices.

Examples

- [1][2] = decomposable
- [12][23][345] =

Decomposable complexes

- Notation: $|\Gamma|$ = ground set of the complex Γ (the union of all faces).

Definition [decomposable complex] defn. 9.3.11.

A simplicial complex Γ is **reducible**, with reducible **decomposition** Γ_1, S, Γ_2 and **separator** $S \subset |\Gamma|$ if it satisfies $\Gamma = \Gamma_1 \cup \Gamma_2$ and $\Gamma_1 \cap \Gamma_2 = S$.
Furthermore, we assume here that neither Γ_1 nor Γ_2 is $= S$.
A simplicial complex is **decomposable** if it is reducible and Γ_1 and Γ_2 are decomposable or simplices.

Examples

- $[1][2]$ = decomposable
- $[12][23][345]$ = decomposable
- $[12][13][23]$ =

Decomposable complexes

- Notation: $|\Gamma|$ = ground set of the complex Γ (the union of all faces).

Definition [decomposable complex] defn. 9.3.11.

A simplicial complex Γ is **reducible**, with reducible **decomposition** Γ_1, S, Γ_2 and **separator** $S \subset |\Gamma|$ if it satisfies $\Gamma = \Gamma_1 \cup \Gamma_2$ and $\Gamma_1 \cap \Gamma_2 = S$.
Furthermore, we assume here that neither Γ_1 nor Γ_2 is $= S$.
A simplicial complex is **decomposable** if it is reducible and Γ_1 and Γ_2 are decomposable or simplices.

Examples

- $[1][2]$ = decomposable
- $[12][23][345]$ = decomposable
- $[12][13][23]$ = not reducible.
- $\Gamma = [12][13][23][345]$ is

Decomposable complexes

- Notation: $|\Gamma|$ = ground set of the complex Γ (the union of all faces).

Definition [decomposable complex] defn. 9.3.11.

A simplicial complex Γ is **reducible**, with reducible **decomposition** Γ_1, S, Γ_2 and **separator** $S \subset |\Gamma|$ if it satisfies $\Gamma = \Gamma_1 \cup \Gamma_2$ and $\Gamma_1 \cap \Gamma_2 = S$.
Furthermore, we assume here that neither Γ_1 nor Γ_2 is $= S$.
A simplicial complex is **decomposable** if it is reducible and Γ_1 and Γ_2 are decomposable or simplices.

Examples

- [1][2] = decomposable
- [12][23][345] = decomposable
- [12][13][23] = not reducible.
- Γ = [12][13][23][345] is reducible but not decomposable, with decomposition ([12][13][23], $\{3\}$, [345]).
- Any complex with only *two* facets is decomposable.

Markov bases of decomposable models

- If Γ is decomposable, then the Markov bases can be computed using a divide-and-conquer algorithm (via the decomposition).
  - The upshot is that they are all *quadratic* - degree $= 2$!
  - See Corollary 9.3.18, Example 9.3.19., but notation :(:(

Adrian Dobra 2003: *We show that primitive data swaps or moves are the only moves that have to be included in a Markov basis that links all the contingency tables having a set of fixed marginals when this set of marginals induces a decomposable independence graph. We give formulae that fully identify such Markov bases and show how to use these formulae to dynamically generate random moves.*

Markov bases of decomposable models

- If Γ is decomposable, then the Markov bases can be computed using a divide-and-conquer algorithm (via the decomposition).
  - The upshot is that they are all *quadratic* - degree $= 2$!
  - See Corollary 9.3.18, Example 9.3.19., but notation :(:(

Adrian Dobra 2003: *We show that primitive data swaps or moves are the only moves that have to be included in a Markov basis that links all the contingency tables having a set of fixed marginals when this set of marginals induces a decomposable independence graph. We give formulae that fully identify such Markov bases and show how to use these formulae to dynamically generate random moves.*

- Good/bad news:
  - What do you think about the quartic from Example 9.3.4:

  Γ=[12][13][23] has the following implicit description:

  $p_{111}p_{122}p_{212}p_{221} = p_{112}p_{121}p_{211}p_{222}$.

Why is this degree > 2? ... Is this model decomposable?

Question to ponder.

Why is [12][23][13] not a cycle?

How are complexes and graphs related?

Part Three - most likely in week 12!

Computing the MLE of an example graph $\mapsto$ homework 5

pages 2-13

The usual... license

This document is created for Math/Stat 561, Spring 2023.

Sources: textbook, *Kaie Kubjas*' Algebraic Statistics course at Aalto University, and Carlos Enrique Améndola Cerón's slides from a M2 workshop in GA Tech 2017.