

# Algebraic Statistics, Hierarchical Models and Markov Bases

Erik Stokes

(joint work with Sonja Petrović)

June 21, 2010

- A statistical model is a parametrized family of probability distributions.

- A statistical model is a parametrized family of probability distributions.

## Example

*The normal distributions are parametrized by  $(\mu, \nu)$ ,  $\nu \geq 0$ . (mean, variance)*

- A statistical model is a parametrized family of probability distributions.

## Example

*The normal distributions are parametrized by  $(\mu, \nu)$ ,  $\nu \geq 0$ . (mean, variance)*

- We are interested in families that are parametrized by algebraic objects (like polynomials).

## Example

- *Suppose we flip two (not necessarily fair) coins and get results  $x_1$  and  $x_2$ .*
- *Let  $p_i$  be the probability that  $x_1$  is in state  $i$  and  $q_j$  be likewise for  $x_2$  ( $i, j$  are  $H$  or  $T$ ).*
- *The probability that  $x_1 = H$  AND  $x_2 = T$  is  $p_H q_T$ . There are 4 other such monomials for the other 4 results.*
- *This gives a map  $\mathbb{R}^4 \rightarrow \mathbb{R}^4$ .*

$$(p_H, p_T, q_H, q_T) \mapsto (p_H q_H, p_H q_T, p_T q_H, p_T q_T)$$

- *Each choice of  $p_i$  and  $q_j$  gives a discrete distribution of probabilities for the 4 states.*

# Setup

- We are interested in *contingency tables* with dimensions  $d_1 \times d_2 \times \cdots \times d_n$ .

# Setup

- We are interested in *contingency tables* with dimensions  $d_1 \times d_2 \times \cdots \times d_n$ .
- For each subset  $\{1, \dots, n\}$  we can sum over the indices *not* in  $F$  to form the  $F$ -margin of the table.

# Setup

- We are interested in *contingency tables* with dimensions  $d_1 \times d_2 \times \cdots \times d_n$ .
- For each subset  $\{1, \dots, n\}$  we can sum over the indices *not* in  $F$  to form the  $F$ -margin of the table.
- We want to fix some subsets and look at all the tables sharing those margins.

# Setup

- We are interested in *contingency tables* with dimensions  $d_1 \times d_2 \times \cdots \times d_n$ .
- For each subset  $\{1, \dots, n\}$  we can sum over the indices *not* in  $F$  to form the  $F$ -margin of the table.
- We want to fix some subsets and look at all the tables sharing those margins.
- Each such collection of subsets describes a *hierarchical model*.

# Setup

- We are interested in *contingency tables* with dimensions  $d_1 \times d_2 \times \cdots \times d_n$ .
- For each subset  $\{1, \dots, n\}$  we can sum over the indices *not* in  $F$  to form the  $F$ -margin of the table.
- We want to fix some subsets and look at all the tables sharing those margins.
- Each such collection of subsets describes a *hierarchical model*.
- A special case is when we want to look at all the  $d_1 \times d_2$  tables with fixed row and column sums.

# Example: Flipping Coins

- We flip some coins and get the following data.

	H	T
H	8	10
T	15	10

# Example: Flipping Coins

- We flip some coins and get the following data.  
Are the coins really fair; are they independent?

	H	T
H	8	10
T	15	10

# Example: Flipping Coins

- We flip some coins and get the following data.  
Are the coins really fair; are they independent?

	H	T	
H	8	10	18
T	15	10	25
	23	20	

- Look at the total number of heads and tails for each coin.

# Example: Flipping Coins

- We flip some coins and get the following data. Are the coins really fair; are they independent?

	H	T
H	8	10
T	15	10

- Look at the total number of heads and tails for each coin.
- One way to test this is by randomly generating tables with the same row/column sums and comparing them (using  $\chi^2$ -statistics).

# Example: Flipping Coins

- We flip some coins and get the following data. Are the coins really fair; are they independent?

	H	T
H	8	10
T	15	10

- Look at the total number of heads and tails for each coin.
- One way to test this is by randomly generating tables with the same row/column sums and comparing them (using  $\chi^2$ -statistics).
- How can we generate such tables?

- To study the model we look at the *Markov basis*:

- To study the model we look at the *Markov basis*:  
First, regard each table as a vector and add them term-by-term.

- To study the model we look at the *Markov basis*:  
First, regard each table as a vector and add them term-by-term.

Think of this as “moving” from one table to another.

- To study the model we look at the *Markov basis*:  
First, regard each table as a vector and add them term-by-term.

Think of this as “moving” from one table to another.

A Markov basis is a set of moves that can move between any 2 tables with the same margins.

- To study the model we look at the *Markov basis*:  
First, regard each table as a vector and add them term-by-term.

Think of this as “moving” from one table to another.

A Markov basis is a set of moves that can move between any 2 tables with the same margins.

The elements of a Markov basis have margins that are all 0.

- To study the model we look at the *Markov basis*:  
First, regard each table as a vector and add them term-by-term.

Think of this as “moving” from one table to another.

A Markov basis is a set of moves that can move between any 2 tables with the same margins.

The elements of a Markov basis have margins that are all 0.

- A Markov basis for the previous example is

$$\left\{ \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \right\}$$

- Question: Do Markov bases exist?

- Question: Do Markov bases exist?

Answer: Yes, the set of all tables whose margins are all 0 is a Markov basis.

- Question: Do Markov bases exist?

Answer: Yes, the set of all tables whose margins are all 0 is a Markov basis.

- Better question: Do *finite* Markov basis exist?

- Question: Do Markov bases exist?

Answer: Yes, the set of all tables whose margins are all 0 is a Markov basis.

- Better question: Do *finite* Markov basis exist?

Better answer: Still yes thanks to [Diaconis-Sturmfels 1998]

Unfold each table into a vector and split it into its positive and negative parts.

$$\begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \leftrightarrow (2, -1, -1, 1) = (2, 0, 0, 1) - (0, 1, 1, 0) \\ = x_1^2 x_4 - x_2 x_3$$

Each table is thus associated to a binomial.

Unfold each table into a vector and split it into its positive and negative parts.

$$\begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \leftrightarrow (2, -1, -1, 1) = (2, 0, 0, 1) - (0, 1, 1, 0) \\ = x_1^2 x_4 - x_2 x_3$$

Each table is thus associated to a binomial.

### Theorem (Diaconis-Sturmfels 1998)

*Any set of binomials,  $\mathcal{M}$  such that every binomial corresponding to a table whose margins are all 0 can be written as a combination of the binomials in  $\mathcal{M}$  is a Markov basis.*

Unfold each table into a vector and split it into its positive and negative parts.

$$\begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \leftrightarrow (2, -1, -1, 1) = (2, 0, 0, 1) - (0, 1, 1, 0) \\ = x_1^2 x_4 - x_2 x_3$$

Each table is thus associated to a binomial.

### Theorem (Diaconis-Sturmfels 1998)

*Any set of binomials,  $\mathcal{M}$  such that every binomial corresponding to a table whose margins are all 0 can be written as a combination of the binomials in  $\mathcal{M}$  is a Markov basis.*

Hilbert's Basis Theorem tells us that a finite  $\mathcal{M}$  exists.

Unfold each table into a vector and split it into its positive and negative parts.

$$\begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \leftrightarrow (2, -1, -1, 1) = (2, 0, 0, 1) - (0, 1, 1, 0) \\ = x_1^2 x_4 - x_2 x_3$$

Each table is thus associated to a binomial.

### Theorem (Diaconis-Sturmfels 1998)

*Any set of binomials,  $\mathcal{M}$  such that every binomial corresponding to a table whose margins are all 0 can be written as a combination of the binomials in  $\mathcal{M}$  is a Markov basis.*

Hilbert's Basis Theorem tells us that a finite  $\mathcal{M}$  exists. This is the algebra in algebraic statistics.

# Graphs and More

- A graph is a collection of *vertices* with *edges* between them.
- If the margins of our model all have size 2 then we can associate a graph,  $\Delta$ , to the model.

# Graphs and More

- A graph is a collection of *vertices* with *edges* between them.
- If the margins of our model all have size 2 then we can associate a graph,  $\Delta$ , to the model.
- If some margins have more than 2 elements, then we get a *simplicial complex* of a *hypergraph* instead of a graph.

# Questions

## Question

*How can we translate information (combinatorial, algebraic topological) about  $\Delta$  to information about the Markov basis?*

## Question

*How can we translate information (combinatorial, algebraic topological) about  $\Delta$  to information about the Markov basis?*

- Can we find:
  - upper bound for the degrees?
  - lower bound for the degrees?
  - any interesting/useful properties (e.g. squarefree, Cohen-Macaulay)?

## Question

*How can we translate information (combinatorial, algebraic topological) about  $\Delta$  to information about the Markov basis?*

- Can we find:
  - upper bound for the degrees?
  - lower bound for the degrees?
  - any interesting/useful properties (e.g. squarefree, Cohen-Macaulay)?
- In terms of:
  - dimension (size of the margins)
  - connectedness
  - number of “loops”

# Answers

- We have some answers:

- We have some answers:
  - All binary  $(2 \times 2 \cdots \times 2)$  models up to  $n = 5$  have known Markov bases [Develin-Sullivant 2003]

- We have some answers:
  - All binary ( $2 \times 2 \cdots \times 2$ ) models up to  $n = 5$  have known Markov bases [Develin-Sullivant 2003]
  - *Decomposable* models (in particular trees) have quadratic Markov bases. [Dobra 2000]

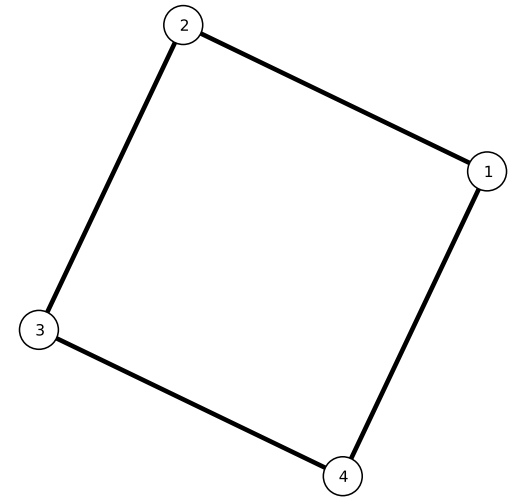
- We have some answers:
  - All binary ( $2 \times 2 \cdots \times 2$ ) models up to  $n = 5$  have known Markov bases [Develin-Sullivant 2003]
  - *Decomposable* models (in particular trees) have quadratic Markov bases. [Dobra 2000]
  - $K_4$ -minor free graphs have binary Markov bases that live in degree at most 4 [Král'-Norine-Pangrác 2008].

Let  $\Delta = \{F_1, \dots, F_k\}$  be a collection of subset of  $\{1, \dots, n\}$ .

### Theorem (Petrovic-S.,Kahle)

*If  $d$  is the size of the smallest element not in some element of  $\Delta$  then the smallest degree of any element in a Markov basis is  $2^{d-1}$ .*

The smallest elements  
not in some element of  $\Delta$  have size 2.



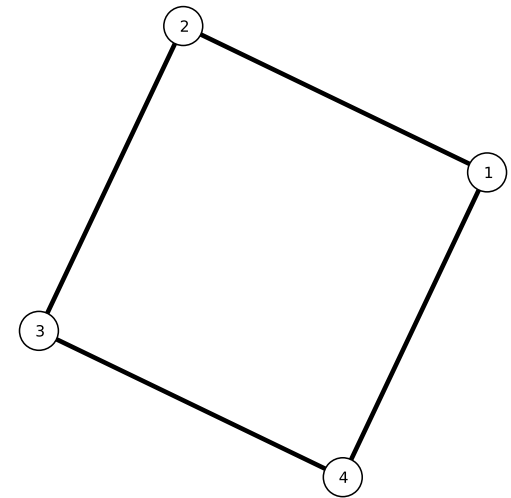
Let  $\Delta = \{F_1, \dots, F_k\}$  be a collection of subset of  $\{1, \dots, n\}$ .

### Theorem (Petrovic-S.,Kahle)

*If  $d$  is the size of the smallest element not in some element of  $\Delta$  then the smallest degree of any element in a Markov basis is  $2^{d-1}$ .*

The smallest elements  
not in some element of  $\Delta$  have size 2.

So the smallest degree  
of anything in the Markov basis is  $2^{2-1} = 2$ .



Let  $\Delta = \{F_1, \dots, F_k\}$  be a collection of subset of  $\{1, \dots, n\}$ .

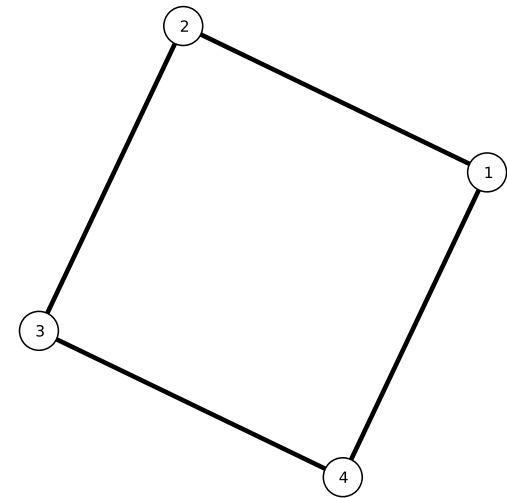
### Theorem (Petrovic-S.,Kahle)

*If  $d$  is the size of the smallest element not in some element of  $\Delta$  then the smallest degree of any element in a Markov basis is  $2^{d-1}$ .*

The smallest elements  
not in some element of  $\Delta$  have size 2.

So the smallest degree  
of anything in the Markov basis is  $2^{2-1} = 2$ .

In fact, this model  
has a Markov basis with elements only in  
degrees 2 and 4.



Let  $\Delta = \{F_1, \dots, F_k\}$  be a collection of subset of  $\{1, \dots, n\}$ .

### Theorem (Petrovic-S.,Kahle)

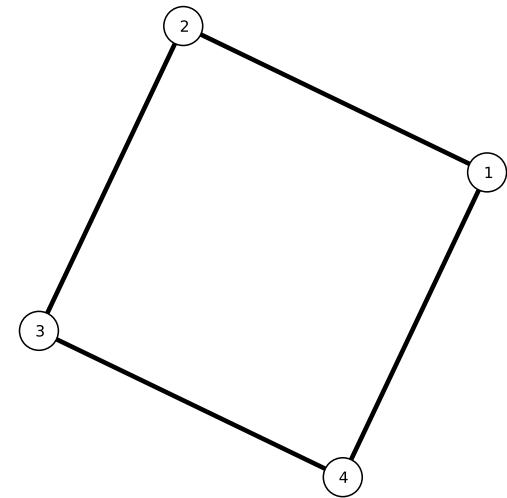
*If  $d$  is the size of the smallest element not in some element of  $\Delta$  then the smallest degree of any element in a Markov basis is  $2^{d-1}$ .*

The smallest elements  
not in some element of  $\Delta$  have size 2.

So the smallest degree  
of anything in the Markov basis is  $2^{2-1} = 2$ .

In fact, this model  
has a Markov basis with elements only in  
degrees 2 and 4.





A more general result shows that the fact  
that  $\Delta$  contains a cycle implies that there is  
a degree 4 element.



# Open Problems

- Can we find an *upper* bound on the degrees of the Markov basis?
- Many models have more than one Markov basis; is there a way to tell when the Markov basis is unique (theoretically or computationally)?
- What if we also fix the sum of some other subset (the subtable problem) or replace the equalities with inequalities (the bounded table problem).

# References

-  Persi Diaconis and Bernd Sturmfels, *Algebraic algorithms for sampling from conditional distributions*, Ann. Statist. **26** (1998), no. 1, 363–397. MR 1608156 (99j:62137)
-  Adrian Dobra and Stephen E. Fienberg, *Bounds for cell entries in contingency tables given marginal totals and decomposable graphs*, Proc. Natl. Acad. Sci. USA **97** (2000), no. 22, 11885–11892 (electronic). MR 1789526 (2001g:62038)
-  Daniel Král', Serguei Norine, and Ondrej Pangrác, *Markov bases of binary graph models of  $K_4$ -minor free graphs*, preprint arXiv:0810.1979v1.
-  Sonja Petrović and Erik Stokes, *Markov degrees of hierarchical models and betti numbers of stanley-reisner ideals*, preprint arXiv:0910.1610v1.