Smoothness of Binary Conditional Independence Models

Robin J. Evans
University of Washington
www.stat.washington.edu/∼rje42

WOGAS3, University of Warwick

April 2011

Outline

Outline

1. Introduction
2. Reducing the Problem
3. Context Specific Conditional Independences
4. Results

Acknowledgements

Thanks also to Mathias Drton, Antonio Forcina, Thomas Richardson and Alex Volfovsky for discussions, helpful comments and code.

Setting and Notation

We work over a set $N = \{1, \ldots, n\}$ indexing multivariate binary random variables $\boldsymbol{X} = (X_1, \ldots, X_n)$.

For $A \subseteq N$, use X_A to mean subvector $(X_i)_{i \in A}$.

For $i_A \in \{0, 1\}^{|A|}$, write $p_{i_A} \equiv P(X_A = i_A)$, and for small examples,

$$p_{010} \equiv P(X_1 = 0,\, X_2 = 1,\, X_3 = 0).$$

Algebraically we work in the polynomial ring

$$\mathbb{C}[p_{00\cdots 0}, p_{10\cdots 0}, \ldots, p_{11\cdots 1}].$$

Write $\Delta_k \subset \mathbb{R}^{k+1}$ for k-dimensional probability simplex, and Δ_k^+ for its interior.

Discrete Conditional Independence

Let $\boldsymbol{p} \in \mathbb{C}^{2^n}$, and $A, B, C \subseteq N$ disjoint. We write $\langle A \perp\!\!\!\perp B \,|\, C\rangle_{\boldsymbol{p}}$, if

$$p_{i_C} p_{i_A i_B i_C} - p_{i_A i_C} p_{i_B i_C} = 0 \qquad \forall (i_A, i_B, i_C) \in \{0,1\}^{|ABC|}.$$

If $\boldsymbol{p} \in \Delta_{2^n-1}$ this is conditional independence of X_A and X_B given X_C under $\boldsymbol{p}$.

A *relation* R is a collection of triples $(A, B \,|\, C)$. A relation defines an algebraic variety

$$\boldsymbol{V}(R) \equiv \{\boldsymbol{p} \,|\, \langle A \perp\!\!\!\perp B \,|\, C\rangle_{\boldsymbol{p}} \text{ for all } (A, B \,|\, C) \in R \text{ and } \sum p = 1\}.$$

Define the *model* associated with R as

$$\boldsymbol{V}_p(R) \equiv \boldsymbol{V}(R) \cap \Delta^+_{2^n-1}.$$

Smoothness

If a model is *smooth* everywhere (i.e. it is locally Euclidean) then it has nice statistical properties:

- asymptotic normality of MLE;
- likelihood ratio tests are asymptotically χ^2;
- model selection by BIC consistent.

However, not all CI models are smooth. Let $R \equiv \{(1,2\,|\,\emptyset),(1,2\,|\,3)\}$; then for binary random variables,

$$1 \perp\!\!\!\perp 2, \quad 1 \perp\!\!\!\perp 2\,|\,3 \quad \Longleftrightarrow \quad \left\{ \begin{array}{rc} & 1 \perp\!\!\!\perp 2,3 \\ \text{or} & 2 \perp\!\!\!\perp 1,3 \end{array} \right\};$$

the model is a union of two smooth hyper-surfaces, so at the intersection ($\perp\!\!\!\perp \{1,2,3\}$) model is not locally Euclidean.

Objective

Questions of interest:

- which models are smooth everywhere and which are not?
- which parts of non-smooth models are so?

In particular, we will attempt to answer these questions for (strictly positive) binary models on up to 4 random variables.

Much of our terminology and approach comes from Drton and Xiao (2010).

Outline

Completeness

We need only consider triples with singletons in first two entries: $(i, j \,|\, C)$.

This still leaves $2^{24} = 1.7 \times 10^7$ relations on 4 variables.

Some relations yield equivalent models. For example:

$$
\begin{aligned}
R_1 &= \{(1, 2 \,|\, \emptyset), (1, 3 \,|\, 2)\} \\
R_2 &= \{(1, 2 \,|\, \emptyset), (1, 3 \,|\, 2), (1, 3 \,|\, \emptyset), (1, 2 \,|\, 3)\},
\end{aligned}
$$

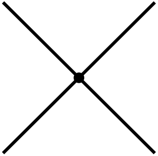
then $\boldsymbol{V}_p(R_1) = \boldsymbol{V}_p(R_2)$.

This defines an equivalence class on relations.

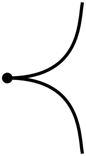
Can consider only *complete* relations: i.e. maximal with respect to equivalence of these models. In this case R_2 is complete.

Singular Points

How can a variety be non-smooth at $\boldsymbol{p}$?



intersection of varieties



Jacobian not of full rank

Set of points where either of these happen is the *singular locus*. All non-smooth points are singular (converse false).

Varieties can be uniquely decomposed into *irreducible* components:

$$V = V_1 \cup \cdots \cup V_k.$$

Reducible varieties can be dealt with via their (unique) irreducible components.

How do we find these components?

Representability

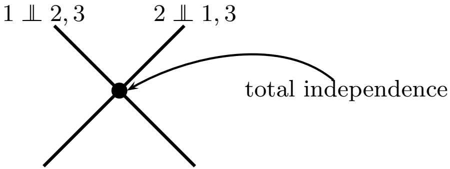
Consider the complete relation $R \equiv \{(1, 2 \,|\, \emptyset), (1, 2 \,|\, 3)\}$.

There is no binary probability distribution which satisfies precisely $1 \perp\!\!\!\perp 2$ and $1 \perp\!\!\!\perp 2 \,|\, 3$. We say that R is not *representable* (or there is no faithful distribution for R).

This is because (in the binary case)

$$1 \perp\!\!\!\perp 2 \;\text{ and }\; 1 \perp\!\!\!\perp 2 \,|\, 3 \quad \Longleftrightarrow \quad 1 \perp\!\!\!\perp 2, 3 \;\text{ or }\; 2 \perp\!\!\!\perp 1, 3.$$

That is, the variety is reducible into a union of two smaller varieties. The intersection is non-smooth.



In fact, all non-representable models induce reducible varieties; thus we need only consider representable models.

Classifying Representable Discrete Models

Work by Matúš and Studený (1994, 1999) classified all CI relations on 4 variables representable by any distribution.

Šimeček (2006) classified all 1098 (up to symmetry) discrete representable relations on 4 variables.

He reduces positive discrete case to 299 relations known representable, plus 57 unclassified.

We are able to show that 9 of these are not positive discrete representable. 3 appear to be positive binary representable (numerically).

Thus problem reduced to 302 plus 45, for a total of 347.

After removing models containing $1 \perp\!\!\!\perp 2$ and $1 \perp\!\!\!\perp 2 \,|\, 3$ but not $1 \perp\!\!\!\perp 2, 3$ or $2 \perp\!\!\!\perp 1, 3$ (or permutations thereof), we get down to at most 206 binary representable.

Outline

1. Introduction
2. Reducing the Problem
3. Context Specific Conditional Independences
4. Results

An Oddity

As mentioned above,

$$1 \perp\!\!\!\perp 2, \quad 1 \perp\!\!\!\perp 2 \,|\, 3$$

is not binary representable. However,

$$1 \perp\!\!\!\perp 2 \,|\, 4, \quad 1 \perp\!\!\!\perp 2 \,|\, 3, 4$$

is binary representable. We say then that representability of binary models is not *minor closed* (Šimeček, 2006).

A *minor* of a relation is what we get by marginalizing and conditioning.

The sets of general discrete and Gaussian representable relations are minor closed.

Forbidden minors are useful for characterizing representable relations in these classes.

What Goes Wrong

$$\begin{array}{rl} & 1 \perp\!\!\!\perp 2 \mid 4, \\ \text{and} & 1 \perp\!\!\!\perp 2 \mid 3, 4 \end{array} \quad \iff \quad \left\{ \begin{array}{rl} & 1 \perp\!\!\!\perp 2, 3 \mid 4 \\ \text{or} & 2 \perp\!\!\!\perp 1, 3 \mid 4 \\ & \\ & \end{array} \right\}$$

Instead define a relation as sets of quadruples $(A, B \mid C, i_C)$.

Lemma (Evans, 2011)

Under context specific relations, representability is minor closed.

Of our 206 remaining relations, 57 are not representable under the extended class.

What Goes Wrong

$$
\begin{array}{rl} & 1 \perp\!\!\!\perp 2 \,|\, 4, \\ \text{and} & 1 \perp\!\!\!\perp 2 \,|\, 3, 4 \end{array}
\quad \iff \quad
\left\{ \begin{array}{lcc} & \multicolumn{2}{c}{1 \perp\!\!\!\perp 2, 3 \,|\, 4} \\ \text{or} & \multicolumn{2}{c}{2 \perp\!\!\!\perp 1, 3 \,|\, 4} \\ \text{or} & 1 \perp\!\!\!\perp 2, 3 \,|\, 4 = 0, & 2 \perp\!\!\!\perp 1, 3 \,|\, 4 = 1 \\ \text{or} & 1 \perp\!\!\!\perp 2, 3 \,|\, 4 = 1, & 2 \perp\!\!\!\perp 1, 3 \,|\, 4 = 0 \end{array} \right\}
$$

Instead define a relation as sets of quadruples $(A, B \,|\, C, i_C)$.

Lemma (Evans, 2011)

Under context specific relations, representability is minor closed.

Of our 206 remaining relations, 57 are not representable under the extended class.

Outline

Results (I)

Easiest way is to prove smoothness is to demonstrate membership of a smooth model class.

Bergsma and Rudas (2002) give large and flexible class of smooth models, using marginal log-linear (MLL) parameters.

Šimeček Total	**356**
Not Discrete Rep.	9
Not Binary Rep.	141
Not Binary Rep. (ext)	54
BR Smooth	98
Other	54

The Hard Work

Remaining 54 models can be classified using computational algebra.

Using `Singular` one can calculate the singular locus, and saturate the ideal to get rid of non-positive points.

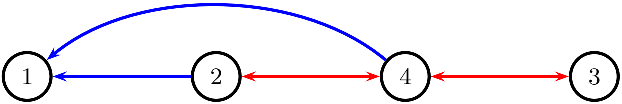
With 16 parameters and k constraints, tries to check $\binom{16}{k}$ minors of the Jacobian. For some k this takes too long. $\binom{16}{8} = 12,870$.

A better way: embed in smallest possible model with algebraic parametrization.

Example

$$1 \perp\!\!\!\perp 3, \quad 1 \perp\!\!\!\perp 3 \,|\, 2, 4, \quad 2 \perp\!\!\!\perp 3, \quad 3 \perp\!\!\!\perp 4.$$

Could parametrize with 16 params and 8 constraints, but this won't work. Instead consider graphical model:



Has an algebraic parametrization, and encodes $1 \perp\!\!\!\perp 3 \,|\, 2, 4$ and $2 \perp\!\!\!\perp 3$. (Richardson, 2009)

Can also easily include $3 \perp\!\!\!\perp 4$ (non-graphical). This leaves 9 parameters and 1 constraint ($1 \perp\!\!\!\perp 3$).

Example (cont.)

$$1 \perp\!\!\!\perp 3, \quad 1 \perp\!\!\!\perp 3 \,|\, 2, 4, \quad 2 \perp\!\!\!\perp 3, \quad 3 \perp\!\!\!\perp 4.$$

In this case we find the model is irreducible.

Singular when $3 \perp\!\!\!\perp 1, 2, 4$ and

$$\begin{aligned} &P(X_1 = 0 \,|\, X_2 = 0, X_4 = 0) - P(X_1 = 0 \,|\, X_2 = 1, X_4 = 0) \\ &- P(X_1 = 0 \,|\, X_2 = 0, X_4 = 1) + P(X_1 = 0 \,|\, X_2 = 1, X_4 = 1) = 0. \end{aligned}$$

Results (II)

33 models found to have non-trivial singularities.

21 models too slow to calculate the singular locus or saturate.

Open to ideas on how to deal with these! An interesting example:

$$1 \perp\!\!\!\perp 2 \,|\, 3, \qquad 1 \perp\!\!\!\perp 2 \,|\, 4.$$

No non-MLL model was found to be smooth.

Summary

We have

- reduced the number of models whose positive discrete representability is unknown from 57 to 45;
- shown that extending to include context specific conditional independences gives us relations which minor-closed under representability;
- gone some way to classifying the smoothness of binary CI models on four variables.

Thank you!

Non-Representable Positive Discrete Models

Consider a positive probability distribution over 4 discrete random variables. Then

$$\left\{ \begin{array}{c} 1 \perp\!\!\!\perp 3 \,|\, 2, 4 \\ 2 \perp\!\!\!\perp 4 \,|\, 1, 3 \\ 1 \perp\!\!\!\perp 2 \\ 1 \perp\!\!\!\perp 4 \end{array} \right\} \quad \Longrightarrow \quad 1 \perp\!\!\!\perp 2, 3, 4.$$

Proof.

Undirected 4-cycle model has factorization

$$\psi(1,2) \cdot \psi(2,3) \cdot \psi(3,4) \cdot \psi(1,4);$$

integrating out 3 gives

$$\psi(1,2) \cdot \psi(2,4) \cdot \psi(1,4),$$

which means no 3-way interaction 124. Then $1 \perp\!\!\!\perp 2$ and $1 \perp\!\!\!\perp 4$ implies $1 \perp\!\!\!\perp 2, 4$. $\square$