On the identification of discrete graphical models with hidden nodes

E. Stanghellini (joint with B. Vantaggi)

April 6, 2010

Introduction

- The focus is on discrete graphical models with one latent binary variable
- Conditions of identification of these models are established
- They are based on the structure of the conditional independence graph

The model I

Let $G^K = (K, E)$ be an undirected graph with:

- node set $K = \{0, 1, \ldots, n\}$
- edge set $E = \{(i, j)\}$, whenever vertices i and j are adjacent in G^K

Partition $K = \{0, O\}$, with $O = \{1, \ldots, n\}$.

The model II

To each node v is associated a discrete random variable A_v. A discrete undirected graphical model is:

- a family of joint distributions of the variables A_v, $v \in K$, satisfying the Markov property with respect to G^K,
- namely that $A_v \perp\!\!\!\perp A_u \mid \text{rest}$ whenever u and v are not adjacent in G^K.

Let A_0 be a binary unobserved variable. Let A_v with $v \in O$ be the observable random variables.

We consider a multidimensional contingency table obtained by the cross classification of N objects according to A_v, $v \in K$.

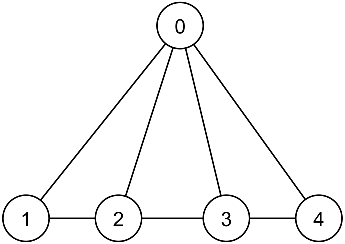


Figure: An example of c.i. graph

Let X be the $2\ell \times 1$, $\ell = \prod_{v=1}^{n} l_v$, vector of the counts of the contingency table, with A_0 running the slowest.

The model III

We assume

- that the elements of X are independent Poisson random variables with $E(X) = \mu_X$.
- $\log \mu_X = Z\beta$, where Z is a $2\ell \times p$ design matrix such that the joint distribution of A_v, $v \in K$, factorizes according to G^K
- β is a p-dimensional vector of unknown parameters. We adopt the corner point parametrization that takes as first level the cell with $A_v = 0$, for all $v \in K$.

The multinomial case can be addressed by assuming N fixed.

The model III

We assume

- that the elements of X are independent Poisson random variables with $E(X) = \mu_X$.
- $\log \mu_X = Z\beta$, where Z is a $2\ell \times p$ design matrix such that the joint distribution of A_v, $v \in K$, factorizes according to G^K
- β is a p-dimensional vector of unknown parameters. We adopt the corner point parametrization that takes as first level the cell with $A_v = 0$, for all $v \in K$.

The multinomial case can be addressed by assuming N fixed.

The model III

We assume

- that the elements of X are independent Poisson random variables with $E(X) = \mu_X$.
- $\log \mu_X = Z\beta$, where Z is a $2\ell \times p$ design matrix such that the joint distribution of A_v, $v \in K$, factorizes according to G^K
- β is a p-dimensional vector of unknown parameters. We adopt the corner point parametrization that takes as first level the cell with $A_v = 0$, for all $v \in K$.

The multinomial case can be addressed by assuming N fixed.

The observable r.v.'s

Let Y the $\ell \times 1$ vector of the counts in the marginal table, obtained by the cross classification of the N objects according to the observed variables only.

- Y may be written as $Y = LX$, with $L = (1,1) \otimes I_\ell$.
- each element of Y is Poisson, with $E(Y) = \mu_Y = Le^{Z\beta}$

For now, we assume that also the observable variables are binary (can be removed).

Note that the typical element of μ_Y is

$$e^{a} + e^{a+b}$$

For instance (with reference to the previous example):

$$\begin{array}{rcccc} \mu[1\,0\,0\,0] = & \underbrace{e^{\mu_0+\beta_1}} & + & \underbrace{e^{\mu_0+\beta_1+\beta_0+\beta_{01}}} \\ & \mu[0\,1\,0\,0\,0] & + & \mu[1\,1\,0\,0\,0] \end{array}$$

An example

We will assume that all observable r.v.'s are connected to the latent one (if not take the subgraph of the relevant var's).

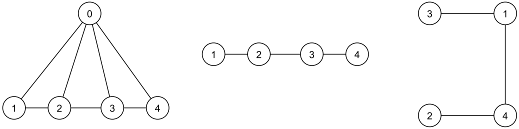


Figure: Example of (a) a G^K graph (b) the induced subgraph G^O and (c) $\bar{G}^O$

Local identification - I

By the inverse function theorem, a model is locally identified if the rank of the transformation from the natural parameters μ_Y to the new parameters β is full. This is equivalent to the rank of following derivative matrix

$$D(\beta)^T = \frac{\partial \mu_Y^T}{\partial \beta} = \frac{\partial (Le^{Z\beta})^T}{\partial \beta} = (LRZ)^T \tag{1}$$

being full, where $R = \mathrm{diag}(\mu_X)$. Note that the (i,j)-th element of $D(\beta)$ is the partial derivative of the i-th component of μ_Y with respect to β_j the j-th element of β.

Local identification - II

Note that, by setting $t_j = e^{\beta_j}$ for any parameter β_j, the parametrization map turns into a polynomial one.

This implies that if there exists a point in the parameter space of t_j at which the Jacobian has full rank, then the rank is full almost everywhere.

Therefore, either (a) there is no point in the parameter space where the rank is full or (b) the rank is full almost everywhere.

Object of the paper:

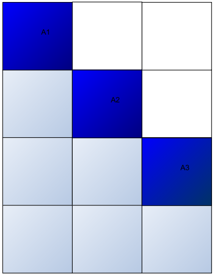
(a) to establish a necessary and sufficient condition on the graphical model to be full-rank almost everywhere in the parameter space

(b) to give the expression of the subspace where identifiability breaks down.

The general idea

Conditions are established such that (a) there is an ordering of rows and columns of $D(\beta)$, such that:

$\mathrm{D}(\beta)=$



and (b) blocks A_i are full-rank.

Condition (a): the existence of the ordering

It requires that the set O of the observed variables contains at least three variables.

Moreover, the complementary graph of the sub-graph G^O is connected, with an m-clique, $m \geq 3$.

Condition (b): the blocks A_i are full rank

Consider $I \subseteq O$, let μ_I be element of μ_Y associated to the entry of the contingency table having value zero for all variables except the variables in I.

Let β_I be the term of interactions among the variables in I.

Let d_I be the row of the matrix $D(\beta)$ corresponding to the first order partial derivative of μ_I with respect of β.

The generic A_i

Let S be a complete subgraph of G^O and $S' \supset S$:

$$
\begin{array}{c|cc}
 & \beta_S & \beta_{\{0,S\}} \\
\hline
\mathrm{d}_S & e^a(1+e^b) & e^{a+b} \\
\mathrm{d}_{S'} & e^{a+a'}(1+e^{b+b'}) & e^{a+a'+b+b'}
\end{array}
\tag{2}
$$

A_i is not full rank if and only if $b' = 0$. Therefore, conditions to assure that $b' \neq 0$ are given (see the paper).

What are the possible expressions of b'? Here are some example:

1. $b' = \beta_{\{0,v\}}$ with $v \in O$ a single node.
2. $b' = \beta_{\{0,l\}}$ with $\{0, l\}$ complete in G^K.
3. b' is a sum of different elements of β (e.g $\beta_0 + \beta_{01}$).

If we assume that the graph is *faithfull*, then 1. and 2. lead to a full rank sub-matrix A_i. Conditions are given to avoid 3., based on the c.i. graph (see the paper).

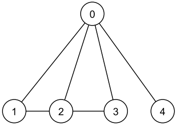
Violations of the conditions (a) and (b)

Violation of condition (b) leads to the expression of the subspace of null measure where identifiability breaks down.

Violation of condition (a) leads to a non full-rank model everywhere in the parameter space.

Here inspection of the matrix $D(\beta)$ is necessary to reveal (a) the rank of $D(\beta)$ and (b) which columns of $D(\beta)$ are linearly dependent.

An example of identified model



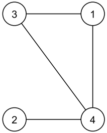


Figure: Full rank model, rank of $D(\beta) = 13$

The LRT against the saturated model has $\chi^2(2)$ as asymptotic distribution.

An example of not identified model

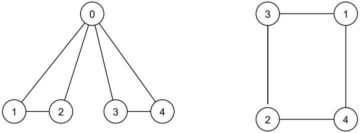


Figure: Violates condition (a), rank of $D(\beta)$=11

The true model dimension is 11.

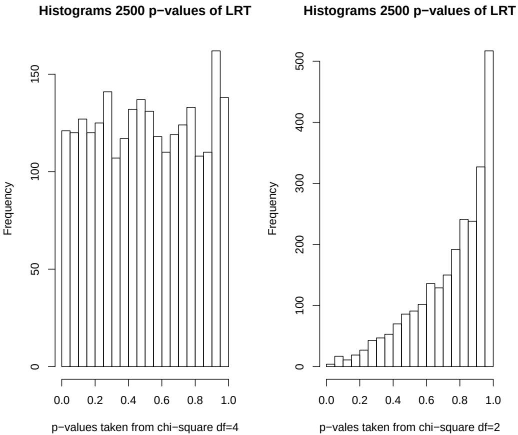
Focus on the model

Inspection of $D(\beta)$ leads to the following conclusions:

- columns $\beta_{12}, \beta_{012}, \beta_{34}, \beta_{034}$ are aliased (and so are other four parameters);
- if we impose $\beta_{012} = \beta_{034} = 0$ the model has full rank.

The LRT against the saturated model seems to have a $\chi^2(4)$ as asymptotic distribution (instead of a $\chi^2(2)$).

Histograms of p-value



Conclusions

- We have derived conditions for the proposed model to be locally identified almost everywhere.
- Expression of the null measure non identified subspace has been derived.
- For non full rank models, a way to compute the true model dimension has been established.
- This leads to compute correctly the asymptotic distribution of the LRT against the saturated model (work in progress, with A. Forcina).

Research question

Will algebraic statistics make may life easier?