Smoothness of conditional independence models for discrete data

A. Forcina, Dipartimento di Economia, Finanza e Statistica,
University of Perugia, Italy

December 6, 2010

Abstract

We investigate the family of conditional independence models require constraints on complete but non hierarchical marginal log-linear parameters; by exploiting results on the mixed parameterization, we show that they are smooth when the jacobian of a reconstruction algorithm has spectral radius strictly less than 1. This condition is always satisfies in simple contexts where only two marginals are involved. For the general case, we describe an efficient algorithm for checking whether the condition is satisfied with high probability; this approach is applied to assess smoothness of several conditional independence models.

Keywords: Conditional Independence, Marginal Log-linear Models, Mixed Parameterizations, Smoothness.

1 Introduction

Roughly speaking, a model is smooth when the variety that it defines in the parameter space can be approximated everywhere by a linear space (see Drton, 2009); this is a relevant property for determining the asymptotic distribution of the likelihood ratio. Different approaches may be used to establish whether a given model is smooth: (i) check whether the model belongs to a family of smooth graphical models, (ii) check whether the model belongs to the family of complete and hierarchical marginal parameterization as described by Forcina et al (2010) or (iii) check whether the algebraic variety defined by the model contains no singularities. The problem with the first approach is that the class of independence models which can be represented by graphs is limited (Studeny, 2001, p. 51); algebraic methods are powerful, but can be used only on problems of limited size in terms of number of parameters. The problem with the approach based on Bergsma and Rudas (2002) class of hierarchical and complete marginal parameterizations is that, while checking completeness is a simple task, checking that the model is hierarchical is much more difficult. In addition, in case we established that the model is non hierarchical, this does not imply that it is non smooth (Forcina et al, 2010).

Bergsma and Rudas (2002, Theorem 3) show that for any model based on a set of marginal log-linear parameters which is not complete (i.e. the same interaction is defined in two different marginals) there is at least a point of singularity in the parameter space. Here is a known example where this happens.
Example 1. Suppose that $X_2 \independent X_4 \mid X_1$ and $X_1, \independent (X_2, X_4) \mid X_3$; Drton (2009) shows that, in the binary case, the model is non smooth. Here completeness is violated because the first independence require that the log-linear interaction $X_1 X_2 X_4$ must be set to 0 in the \{X_1, X_2, X_4\} marginal while the second independence requires that the same interaction must be set to 0 in the full joint distribution.

For the following example, due to Šimeček (2006), no hierarchical parameterization exists (see Forcina et al, 2010, Example 10). For the binary case, M. Drton (personal communication) has an algebraic argument to prove that the model is smooth.

Example 2. Let $X_1 \independent X_2 \mid X_3$, $X_2 \independent X_3 \mid X_4$ and $X_2 \independent X_4 \mid X_1$. The problem here is that the \{X_2, X_3\} interaction belongs to the \{X_1, X_2, X_3\} marginal but has to be constrained in the \{X_2, X_3, X_4\} marginal, the \{X_2, X_4\} interaction belongs to the \{X_2, X_3, X_4\} marginal but has to be constrained in the \{X_1, X_2, X_4\} marginal and the \{X_1, X_2\} interaction belongs to the \{X_1, X_2, X_3\} marginal but has to be constrained in the \{X_1, X_2, X_3\} marginal. Thus the three marginals are connected as in a loop: for any ordering of the three marginals each one contains an interaction which has to be constrained to 0 in the next one.

One way of determining whether models like Example 2 (additional instances will be given in Section 3) are smooth would be to show that, given a complete but non hierarchical marginal parameterization, there is an algorithm for reconstructing a compatible probability distribution which always converge to a solution which must be unique. By applying fixed point theorems (see for example Bulirsch and Stoer, 2002, p. 297), we propose such an algorithm and study its convergence properties; the algorithm requires updating a small set of key probabilities and is such that, once these probabilities are known, the joint distribution is uniquely determined. Because convergence and uniqueness depend on properties of the matrix of derivatives of these key probabilities in one step relative to the same probabilities in the previous step, we use exponential family results to derive an explicit expression for the jacobian matrix. Convergence to a unique solution is ensured when the spectral radius (maximum absolute value of the eigenvalues) of the jacobian is strictly less than 1 everywhere on the parameter space. In the artificial case of models involving only two marginals, we prove that any such algorithm always converges to a unique solution. With loops involving several marginals, the structure of the derivative matrix becomes too complex for determining an upper bound of the spectral radius; however we provide an efficient numerical method for checking that the derivative matrix has spectral radius less than 1 everywhere on the parameter space with probability arbitrarily close to 1; this numerical test is applied to a few interesting examples of complete non hierarchical models which, once properly implemented, turn out to be smooth.

In Section 2 we introduce basic notations, review relevant results on exponential family and the mixed parameterization and derive new results concerning the derivative matrix for the mixed parameterization. In Section 3 we describe the reconstruction algorithm and give a formula for computing the one step ahead derivative and prove that, any conditional independence model involving two marginals is smooth. In Section 4 we illustrate the method for numerical assessment of smoothness with a few examples.
2 Preliminary results

2.1 Notations

Consider the joint distribution of \( d \) discrete random variables where \( X_j, j = 1, \ldots, d \), takes values in \((1, \ldots, r_j)\). For conciseness, we denote variables by their indices and use capitals to denote non-empty subsets of \( V = \{1, \ldots, d\} \); such subsets will determine the variables involved either in a marginal distribution or in an interaction term. The collection of all non-empty subsets of a set \( M \subseteq V \) will be denoted by \( \mathcal{P}(M) \). The distribution of variables in \( V \) is determined by the vector of joint probabilities \( p \), of dimension \( t = \prod_j r_j \) its entries, in lexicographic order, correspond to cell probabilities and are assumed to be strictly positive. For any \( M \in \mathcal{P}(V) \), let \( p(M) \) denote the vector of probabilities for the marginal distribution of the variables \( X_j \) for \( j \in M \), with entries in lexicographic order.

Assume, for convenience, that log-linear parameters for the joint distribution, as well as for an arbitrary margin, are based on adjacent contrasts (see for example Forcina et al, 2010, p. 2520). In short, for any \( I \in \mathcal{P}(M) \), the vector of log-linear interaction parameters will be denoted by \( \eta(I,M) \) and can be computed as

\[
\eta(I,M) = H(I,M) \log p(M),
\]

where \( H(I,M) \) is a matrix of row contrasts, see the Appendix for details. Let also \( \eta(M) = H(M) \log p(M) \) denote the vector obtained by stacking the \( \eta(I,M) \) components one below the other in lexicographic order relative to \( I \in \mathcal{P}(M) \). Under multinomial sampling, \( \eta(M) \) defines a vector of variation independent canonical parameters for \( p(M) \). Let \( G(M) \) be any right inverse of \( H(M) \), then (1) implies the reconstruction formula

\[
\frac{\exp[G(M)p(M)]}{\Gamma[\exp[G(M)p(M)]].}
\]

A convenient choice for \( G(M) \), described in the Appendix, has blocks of columns \( G(I,M) \) for all \( I \in \mathcal{P}(M) \) such that \( \mu(I) = G'(I,M)p(M) \) is the survival function within \( I \)

\[
\mu_I(x_I) = P(X_j > x_j, \ j \in I), \ \text{for all} \ I \in \mathcal{P}(M),
\]

where the vector \( x_I \) has elements \( x_j = 1, \ldots, r_j - 1, \ \forall j \in I \) (see Bartolucci, Colombi and Forcina, 2007, p. 699). Within the theory of exponential families (Barndorff-Nielsen, 1979, p. 121) the \( \mu \)'s are known as \textit{mean parameters} and there is a diffeomorphism between the vector \( \mu(M) \), obtained by stacking one below the other the \( \mu(I), I \in \mathcal{P}(M) \), and \( \eta(M) \). It is worth noting that \( \mu(I) \) does not depend on \( M \).

2.2 Results on the mixed parameterization

In the following a crucial role is played by results on the so called \textit{mixed parameterization} Barndorff-Nielsen (1979, p. 121-122). For an arbitrary partition of the collection of interactions \( I \in \mathcal{P}(M) \) into \( U \) and \( V = \mathcal{P}(M) \setminus U \), there is a diffeomorphism between the log-linear parameters \( \eta(M) \) (or the mean parameter \( \mu(M) \)) and the pair of vectors \( [\eta(U,M), \mu(V)] \) where \( \eta(U,M) = (\eta(I,M), I \in U) \) is composed by canonical parameters, and \( \mu(V) = [\mu(I), I \in V] \) is composed of mean parameters; in addition, the two components are variation independent.
It is easily verified that $(\partial \mu(M))/(\partial \eta(M))' = G(M)'\Omega(M)G(M)$. Let $F_{u,v}(M) = G_u(M)'\Omega(M)G_v(M)$, where $G_h(M)$ is the set of columns of $G(M)$ associated with $I \in \mathcal{H}$ and $\Omega(M) = \text{diag}[p(M)] - p(M)p(M)'$ is the derivative of $t = \exp(s)/\exp(s)$ with respect to $s$. Because $G(M)$ is a full rank design matrix whose columns are vectors of 0’s and 1’s, if we let $x, y$ be any two distinct columns of $G(M)$, they define two binary variables, say $X, Y$, with $E(X) = x'p(M)$, $E(Y) = y'p(M)$. In addition

$$x'\Omega(M)y = x'\text{diag}[p(M)]y - x'p(M)p(M)'y = E(XY) - E(X)E(Y).$$

It follows that $F(M) = G(M)'\Omega(M)G(M)$ is the variance covariance matrix of a collection of distinct binary variables, thus, if the elements of $p$ are all strictly positive, $F(M)$ is positive definite.

The following result implies that any block of $F(M)$ depends on the subset of interactions involved but not on the marginal distribution where they are defined. Recall that $G_h(M)$ is the design matrix within $M$ for the set of interactions $I \in \mathcal{H} \subseteq \mathcal{P}(M)$.

**Lemma 1.** Suppose that $\mathcal{U}, \mathcal{V} \subseteq \mathcal{P}(M)$, then

$$G_u(M)'\Omega(M)G_v(M) = G_u(V)'\Omega(V)G_v(V).$$

**Proof.** Let $L$ be the matrix such that $p(M) = Lp(V)$ and let $s$ be the ratio between the length of $p(V)$ and that of $p(M)$. Because $LL'/s$ is the projection matrix onto the space spanned by the columns of $G_u(V)$ with $M = \mathcal{P}(M)$, it follows that if $\mathcal{H} \subseteq \mathcal{P}(M)$, then $(LL'/s)G_h(V) = G_h(V)$; in addition, $G_h(M) = LG_h(V)/s$ and $\Omega(M) = L\Omega(V)L'$, then the result follows because

$$G_u(M)'\Omega(M)G_v(M) = G_u(V)'(LL'/s)\Omega(V)(LL'/s)G_v(V).$$

$\square$

Within a given marginal distribution it is possible to compute the derivative of the elements in $\mu_u$ with respect to those in $\mu_v$ for given $\eta(U, M)$.

**Lemma 2.** Let $\mathcal{U} \in \mathcal{P}(M)$ and $\mathcal{V} = \mathcal{P}(M)\setminus\mathcal{U}$; for fixed $\eta_u(M)$ we have

$$\frac{\partial \mu_u}{\partial \mu_v} = F_{uv}F_{vv}^{-1} \tag{3}$$

**Proof.** Recall that

$$\frac{\partial}{\partial (\eta_u' \eta_v')} \begin{pmatrix} \mu_u \\ \mu_v \end{pmatrix} = \begin{pmatrix} F_{uu} & F_{uv} \\ F_{vu} & F_{vv} \end{pmatrix}, \quad \frac{\partial}{\partial (\eta_u' \eta_v')} \begin{pmatrix} \eta_u \\ \eta_v \end{pmatrix} = \begin{pmatrix} I & 0 \\ F_{vu} & F_{vv} \end{pmatrix},$$

the equation on the left follows simply from the expression of $\mu_u$ and the chain rule while the one on the right uses the fact that $(\partial \eta_u)/(\partial \eta_v)$ is a matrix of 0’s because the two components, being canonical parameters, are variation independent.

By the derivative of the inverse function we then have

$$\frac{\partial}{\partial (\eta_u' \mu_v')} \begin{pmatrix} \mu_u \\ \mu_v \end{pmatrix} = \begin{pmatrix} I & 0 \\ -F_{vv}^{-1}F_{vu} & F_{vv}^{-1} \end{pmatrix} \begin{pmatrix} F_{uu} & F_{uv} \\ F_{vu} & F_{vv} \end{pmatrix}^{-1} = \begin{pmatrix} F_{uu} - F_{uv}F_{vv}^{-1}F_{vu} & F_{uv}F_{vv}^{-1} \\ 0 & F_{vv}^{-1}I \end{pmatrix}$$

$\square$
This result tells us how the mean parameters, reconstructed by a mixed parameterization, depend on the two components. Because $\mu_v$ is fixed, the corresponding component in the derivative is an identity matrix; on the other hand, because $\mu_v$ and $\eta_u$ are variation independent, the corresponding component of the derivative is a matrix of 0's. The component on the top right measures how $\mu_u$, the mean parameters reconstructed with the mixed parameterization, changes in response to local changes in $\mu_v$.

2.3 Complete and hierarchical marginal parameterizations

A marginal log-linear parameterization of a discrete distribution may be defined (Bergsma and Rudas, 2002) by a non decreasing sequence of marginals, $M_1, \ldots, M_s$ and the collection of interactions $\mathcal{J}_m$ defined within $M_m$, $m = 1, \ldots, s$; let $\eta$ be the vector obtained by stacking one below the other the vectors $\eta(\mathcal{J}_m, M_m)$ for $m = 1, \ldots, s$. We recall that:

Definition 1. The vector of marginal parameters $\eta$ is called complete if (i) $\bigcup_1^s \mathcal{J}_m = \mathcal{P}(V)$ and (ii) $\mathcal{J}_h \cap \mathcal{J}_k = \emptyset$ for all $h \neq k$.

In words, any possible interaction is defined in one and only one marginal and $M_s = V$.

Definition 2. The vector of marginal parameters $\eta$ is called hierarchical if, in addition to being complete, it is such that, for $m = 2, \ldots, s$,

$$\mathcal{J}_m = \mathcal{P}(M_m) \setminus \bigcup_1^{m-1} \mathcal{J}_m. \quad (4)$$

In words, the interactions that belong to $\mathcal{P}(M_m)$ but are not defined within $M_m$ must be defined in one of the previous marginals. This condition is violated when, like in Example 2, there is a sub-collection of marginals $M_h, \ldots, M_{h+t}$, such that some of the interactions which belong to $\mathcal{P}(M_{h+i})$, $i = 1, \ldots, t-1$, have to be defined in $M_{h+i+1}$ and some elements of $\mathcal{P}(M_{h+t})$ have to be defined in $M_h$. Such collections of marginals will be called loops.

3 A reconstruction algorithm for marginals in a loop

Let $\mathcal{D}_r = \bigcup_1^r \mathcal{J}_i$; suppose that a parameterization is complete and that (4) holds for a given $m$, then Bergsma and Rudas (2002) basic argument can be used to prove the following:

Lemma 3. Suppose we have proved that there is a diffeomorphism between $\eta(\mathcal{D}_{m-1})$ and $\mu(\mathcal{D}_{m-1})$ and that (4) holds, then there is a diffeomorphism between $\eta(\mathcal{D}_m)$ and $\mu(\mathcal{D}_m)$.

One way to prove that an independence model which require a non hierarchical parameterization is smooth is to define an algorithm which reconstruct the mean parameters of the marginals within a loop and show that the algorithm always converges to a unique solution, then the mapping between $\eta$ and $\mu$ is a diffeomorphism and any model defined by constraining to 0 elements of $\eta$ is smooth. The structure of such an algorithm is described below.

Suppose we have a loop $M_h, \ldots, M_{h+t}$; for $m = h + 1, \ldots, h + t$, let $\mathcal{K}_m = \mathcal{P}(M_{m-1}) \cap \mathcal{J}_m$ denote the collection of interactions defined within $M_m$ which are contained in the previous marginal; let also $\mathcal{K} = \bigcup_{h+1}^{h+t} \mathcal{K}_m$. The mean parameters in $\mu(\mathcal{K})$ are kind of key parameters in the sense that, once a set of compatible values for $\mu(\mathcal{K})$ is given, the marginal distributions in $M_h, \ldots, M_{h+t}$ is uniquely determined by the mixed parameterization. To fix ideas, consider the following example which is an extended version of Example 2.
Example 3. Consider the model defined by the collection of independencies: $1 \perp 3 \mid 2$, $1 \perp 2 \mid 5$, $1 \perp 5 \mid 4$, $1 \perp 4 \mid 3$ and note that the relevant marginals: $M_h = \{1, 2, 3\}$, $M_{h+1} = \{1, 2, 5\}$, $M_{h+2} = \{1, 4, 5\}$, $M_{h+3} = \{1, 3, 4\}$ constitute a loop because the last element in the 4th column in Table 1 has to be defined in the next marginal. It is convenient to set $D_{h-1} = \{1, 2, 3, 4, 5, 23, 25, 34, 45\}$, the collection of interaction not involved in the independence statements to be defined in marginals before $M_h$.

Table 1: Key interactions and their dependence structure for Example 3.

<table>
<thead>
<tr>
<th>$M_m$</th>
<th>$J_m$</th>
<th>$K_m$</th>
<th>$\mathcal{P}(M_m) \setminus J_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_h$</td>
<td>${13, 123}$</td>
<td>-</td>
<td>${3, 2, 23, 1, 12}$</td>
</tr>
<tr>
<td>$M_{h+1}$</td>
<td>${12, 125}$</td>
<td>${12}$</td>
<td>${5, 2, 25, 1, 15}$</td>
</tr>
<tr>
<td>$M_{h+2}$</td>
<td>${15, 145}$</td>
<td>${15}$</td>
<td>${5, 4, 45, 1, 14}$</td>
</tr>
<tr>
<td>$M_{h+3}$</td>
<td>${14, 134}$</td>
<td>${14}$</td>
<td>${4, 3, 34, 1, 13}$</td>
</tr>
</tbody>
</table>

A reconstruction algorithm within a loop may be based on the following steps:

**Initialization step** For $m = h, \ldots, h + t - 1$ partition $\mathcal{P}(M_m)$ into $\mathcal{U}_m = J_m \cup K_{m+1}$ and $\mathcal{V}_m = \mathcal{P}(M_m) \setminus \mathcal{J}_m$; set $\eta(K_{m+1}, M_m) = 0$ and determine a first guess for $\mu(\mathcal{P}(M_m))$ by a mixed parameterization which combines $\eta(\mathcal{U}_m, M_m)$ and $\mu(\mathcal{V}_m)$.

**Iteration step** Once each marginal in the loop has been visited at least once so that an initial guess for $\mu(K_m)$, $m = h + 1, \ldots, h + t$, is available, set $\mathcal{U}_m = J_m$, $\mathcal{V}_m = \mathcal{P}(M_m) \setminus J_m$ and compute an updated guess for the marginal distribution in $M_m$ by a mixed parameterization that combines $\eta(\mathcal{U}_m, M_m)$ and $\mu(\mathcal{V}_m)$. Note however that, while for $\mu(\mathcal{V}_m \setminus K_{m+1})$ the true value is available from marginals before the loop, for the elements of $\mu(K_{m+1})$ the available estimate is the one produced in the previous cycle.

A full cycle of the algorithm consists in updating $\hat{\mu}(K)^{s-1}$ into $\hat{\mu}(K)^s$. If we assume that $\eta$ is a compatible vector of marginal log-linear parameters, the algorithm defines a fixed-point function that satisfies

$$x = \Psi(x) \tag{5}$$

for some $x \in \mathcal{X}$, the space where $\mu(K)$ can vary.

**Remark 1.** In principle, more general structures of non hierarchical parameterizations are possible and an extended version of the above algorithm could be designed, but a more complex notation would be required.

### 3.1 Convergence of the algorithm

We first recall that a function is a contraction if there exists a constant $k < 1$ such that

$$\|\Psi(x_a) - \Psi(x_b)\| \leq k\|x_a - x_b\| \tag{6}$$

for any $x_a, x_b \in \mathcal{X}$. For a square matrix $D$ let $\rho(D)$ denote the spectral radius (maximum of the absolute value of its eigenvalues) of $D$. Recall that
Lemma 4. A sufficient condition for $\Psi$ to be a contraction is that
$$\rho(D) = \rho\left(\frac{\partial \Psi(x)}{\partial x'}\right) < 1.$$ 

Proof. Follows from the mean value theorem applied to (6), the multiplication property of a matrix norm and the fact that the spectral radius is the greatest lower bound for any matrix norms of $D$ (see Horn and Johnson, 2009, pp. 290, 297).

Theorem 1. Let $x \in \mathcal{X}$; if the marginal parameters are compatible, so that (5) has at least a solution in $\mathcal{X}$ and $\rho(D) < 1$, then the solution must be unique.

Proof. Because we can assume that there is at least a compatible solution inside the parameter space, uniqueness follows easily (see for instance Agarwal et al, 2001, Theorem 1.1).

3.2 The one step ahead jacobian

An analytic expression for $D$ may be obtained by noting that $\mathcal{J}_m$, $m = (h + 1, \ldots, h + t)$, is updated within $M_m$ as a function of $\mathcal{J}_{m+1}$ computed within $M_{m+1}$ in the previous cycle when $m < h + t$; $\mathcal{J}_{h+1}$, determined in $M_{h+1}$, depends on $\mathcal{J}_{h}$, $h + 1, \ldots, h + 4$ determined in the current cycle and these depend on $\mathcal{J}_{m}$, $m = h, \ldots, h + 4$ uniquely determined by assigning to each marginal all the interactions involved in the corresponding independence statement. The following example should help clarify the procedure.

Example 4. Consider the model: 1 $\perp \perp$ 2 $\mid$ 3, 2 $\perp \perp$ 3 $\mid$ 4, 3 $\perp \perp$ 4 $\mid$ 5, 4 $\perp \perp$ 5 $\mid$ 1 and 5 $\perp \perp$ 1 $\mid$ 2.

The sequence of margins $M_h = \{1, 2, 3\}$, $M_{h+1} = \{2, 3, 4\}$, $M_{h+2} = \{3, 4, 5\}$, $M_{h+3} = \{1, 4, 5\}$, $M_{h+4} = \{1, 2, 5\}$ constitute a loop. In this case it is convenient to set $D_{h-1} = \{1, 2, 3, 4, 5, 13, 24, 35, 14, 25\}$, then the structure of the $\mathcal{J}_m$, $m = h, \ldots, h + 4$ is uniquely determined by assigning to each marginal all the interactions involved in the corresponding independence statement. The following table provides a quick reference tool for computing the jacobian by listing the elements of $\mathcal{J}_m$, $K_m$ and those of $A_m = \mathcal{P}(M_m) \setminus (D_{h-1} \cup J_m)$, where, in the mixed parameterization, the updated value of $\mathcal{J}_m$ is a function of $\mu(K_m)$ determined in the next marginal in the loop.

Table 2: Key interactions and their dependence structure for Example 4.

<table>
<thead>
<tr>
<th>$M_m$</th>
<th>$\mathcal{J}_m$</th>
<th>$K_m$</th>
<th>$A_m$</th>
<th>$M_m$</th>
<th>$\mathcal{J}_m$</th>
<th>$K_m$</th>
<th>$A_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_h$</td>
<td>{12, 123}</td>
<td>-</td>
<td>23</td>
<td>$M_{h+3}$</td>
<td>{45, 145}</td>
<td>45</td>
<td>15</td>
</tr>
<tr>
<td>$M_{h+1}$</td>
<td>{23, 234}</td>
<td>23</td>
<td>34</td>
<td>$M_{h+4}$</td>
<td>{15, 125}</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>$M_{h+2}$</td>
<td>{34, 345}</td>
<td>34</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The crucial result is that each component function involved in the updating of $\mu(K_m)$, $m = h + 1, \ldots, h + t$ may be expressed as a function which determines a vector $\mu(U)$ from a vector $\mu(V)$ in a mixed parameterization, it follows that the corresponding derivative may be computed by applying (3).

The following example involves a smaller number of marginals but has a more complex structure.
Example 5. Consider the model defined by $1 \perp \perp 2 \mid (3,5)$, $1 \perp \perp 3 \mid (4,5)$ and $4 \perp \perp 5 \mid (1,2)$. These independencies may be implemented by parameterizing the marginals $M_h = \{1,2,3,5\}$, $M_{h+1} = \{1,3,4,5\}$, $M_{h+2} = \{1,2,4,5\}$ which constitute a loop as detailed in the table below where it is convenient to set $D_{h-1} = \{1,2,3,4,5,14,15,23,24,25,34,35,124,235,345\}$.

<table>
<thead>
<tr>
<th>$M_m$</th>
<th>$J_m$</th>
<th>$K_m$</th>
<th>$A_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_h$</td>
<td>${12,125,123,1235}$</td>
<td>-</td>
<td>${13,135}$</td>
</tr>
<tr>
<td>$M_{h+1}$</td>
<td>${13,135,134,1345}$</td>
<td>${13,135}$</td>
<td>${45,145}$</td>
</tr>
<tr>
<td>$M_{h+2}$</td>
<td>${45,145,245,1245}$</td>
<td>${45,145}$</td>
<td>${12,125}$</td>
</tr>
</tbody>
</table>

### 3.3 Loops of two margins

When a loop is made of two margins only, the jacobian matrix has a particularly simple form because the only set of key parameters are those missing from the first marginal and updated in the second, say $\mu(K)$; these are functions of certain mean parameters updated in the first marginal, say $\mu(H)$. Then, by using (3) twice, the one-step ahead jacobian has the simple form

$$D = \frac{\partial \mu(K)}{\partial \mu(H)} \frac{\partial \mu(H)}{\partial \mu(K)} = F_{kh} F_{hh}^{-1} F_{hk} F_{kk}^{-1},$$

and we can prove the following result:

**Lemma 5.** The reconstruction algorithm involving only two marginals has a jacobian with spectral radius strictly less than 1.

**Proof.** Being covariance matrices, $F_{hh}$, $F_{kk}$ are both positive definite; let $A = F_{kk}$ and $B = F_{hh} F_{hh}^{-1} F_{hk}$ and note that $A - B$, being the residual variance in a linear regression, is also positive definite. Then Theorem 7.7.3 in Horn and Johnson (2009) implies that $\rho(D) = \rho(BA^{-1}) < 1$. □

### 4 Numerical investigations

An explicit expression for the jacobian is easily computed even in the general case, but it is difficult to determine an upper bound for $\rho(D)$ for any $p$. However, the numerical value of $\rho(D)$ is easily computed even for moderately large distributions and it is possible to sample a large set of points in the parameter space very quickly and possibly over-sample areas where the spectral radius tends to be larger. Numerical investigations on the main examples are summarized in Table 4 where $U$ denotes a uniform distribution. For each example, 10 thousand points from the parameter space of $p$ were selected with different sampling schemes: (i) $U$, (ii) $U^{1/5}$, (iii) $U^8$. Computations for each example take few seconds in the binary case and less than a minute with 3 categories.

The results in Table 4 indicate clearly that in all cases the spectral radius of the jacobian is well below 1 inside of the parameter space and is slowly approaching 1 only when we get extremely close to the boundary. Evidence that the procedure can easily detect points where
Table 4: *Estimated maximum spectral radius of the jacobian matrix out of 10 thousands draws with different sampling methods*

<table>
<thead>
<tr>
<th>Model</th>
<th>Binary case</th>
<th>3 levels per variable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>U</td>
<td>U^{1/5}</td>
</tr>
<tr>
<td>Example 2</td>
<td>0.3032</td>
<td>0.6547</td>
</tr>
<tr>
<td>Example 3</td>
<td>0.3078</td>
<td>0.4517</td>
</tr>
<tr>
<td>Example 4</td>
<td>0.3174</td>
<td>0.4092</td>
</tr>
<tr>
<td>Example 5</td>
<td>0.4049</td>
<td>0.6745</td>
</tr>
</tbody>
</table>

the condition is violated, is provided by the following. In Example 4 let $h = 1$, $D_0 = \emptyset$ and define the interactions $\{1, 2, 3, 4, 5, 13, 24, 35, 14, 25\}$ within the four main marginals; though by sampling from $U$ it is unlikely to find points where the condition is violated, with $U^{8}$ one quickly detects points where the spectral radius exceeds 1.

**Appendix**

Omit for simplicity specification of the reference marginal, then $H(I) = \bigotimes_{j \in M} H(I, j)$ where

$$H(I, j) = \begin{cases} 
0_{r_j-1} I_{r_j-1} - (I_{r_j-1} 0_{r_j-1}) & \text{if } j \in I \\
1 0_{r_j-1}' & \text{otherwise,}
\end{cases}$$

As concerns $G(I)$, let

$$G(I) = \bigotimes_{j \in M} G(I, j), \quad G(I, j) = \begin{cases} 
T_j & \text{if } j \in I \\
1_{r_j} & \text{otherwise,}
\end{cases}$$

where $T_j$ is the matrix obtained by removing the first columns from a $r_j \times r_j$ lower triangular matrix of 1’s.

**Acknowledgment**

I would like to thank I. Gerace from Perugia University and Y. Rinott from the Hebrew University for suggestions and T. Rudas from Budapest University for discussions.

**References**

**References**


