Hierarchical Multinomial Marginal Models

Modelli gerarchici marginali per variabili casuali multinomiali

Roberto Colombi
Dipartimento di Ingegneria dell'Informazione e Metodi Matematici,
Università di Bergamo
e-mail: colombi@unibg.it

Riassunto: Questo lavoro descrive i modelli marginali gerarchici per tabelle di contingenza multidimensionali basati su una parametrizzazione delle probabilità congiunte proposta da Bartolucci *et al.* (2007). Questa classe di modelli include come casi particolari molti modelli per tabelle di contingenza, introdotti come alternative ai modelli log-lineari per ovviare alle ben note limitazioni di questi ultimi nel parametrizzare distribuzioni marginali e nel trattare in modo appropriato le variabili ordinali. L'utilità dei modelli presentati è illustrata nel contesto della parametrizzazione di modelli ricorsivi a blocchi specificati dalle proprietà markoviane di Andersson, Madigan e Perlman.

Keywords: block recursive models, generalized odds ratios, marginal models, monotone dependence, multinomial-poisson homogeneous models

1. Introduction

In the log-linear parametrization all the interactions are contrasts of logarithms of joint probabilities and this is the main reason why this parametrization is not convenient to express hypotheses on marginal distributions or to model ordered categorical data. On the contrary *Hierarchical Multinomial Marginal models* (HMM) (Bartolucci *et al.* 2007) are based on parameters, called *generalized marginal interactions*, which are contrasts of logarithms of sums of probabilities. HMM models allow great flexibility in choosing the marginal distributions, within which the interactions are defined, and they are a useful tool for modeling marginal distributions and for taking into proper account the presence of ordinal categorical variables. For example only base-line and local logits together with local and base-line log-odds ratios, defined on joint probabilities, are log-linear parameters while all the known logits (base-line, local, global, continuation etc.) and all types of log-odds ratios (base-line, local, global, continuation, etc.) defined on joint or marginal probability functions are generalized marginal interactions. The HMM models are based on an ordered family of marginal probability functions such that every probability function of the family is not a marginal distribution of any of its predecessors. The parameters of an HMM model are interactions which are defined within the marginal distributions of the previous family. Several models proposed in the literature are special cases of HMM models. *Log-linear Models* are HMM models where the interactions are defined within the joint probability function. The Bergsma and Rudas (2002) *Marginal Models* are HMM models where the interactions are of log-linear type but are defined in different marginal distributions. The Glonek and McCullagh (1995) *Multivariate Logit Models* are HMM models where the parameters are the highest order interactions that can be defined within each of the marginal distributions. HMM models are introduced in section two and in section three the usefulness of the HMM parameterizations in the context of block recursive multivariate logit models is examined.

2. Basic concepts on HMM models

In this section we show that *generalized marginal interactions* are standard log-linear interactions which are computed in tables obtained by marginalizing with respect to some variables and by aggregating the categories of some other variables. Secondly we show that every generalized marginal interaction can be seen as a contrast of well known types of generalized logits and log-odds ratios. We consider q categorical variables A_j, $j = 1, ..., q$, where A_j has categories in the set $\mathcal{A}_j = \{a_{ji_j}, i_j = 1, 2, ..., r_j\}$. For ordinal variables the numbering of the categories is assumed to be coherent with their order. The vector of the $c = \prod_1^q r_j$ joint probabilities is denoted by $\boldsymbol{\pi}$ and is assumed to be strictly positive. The set of variables that defines a given marginal distribution is denoted by the set $\mathcal{M}$ of indices of the corresponding variables. The set $\mathcal{M}$ is called *marginal set* and the distribution associated with it is called $\mathcal{M}$*-marginal distribution.* The set $\mathcal{Q} = \{1, \ldots, q\}$ refers to the joint distribution.

2.1. Generalized marginal interactions

Any generalized marginal interaction is defined by the interaction set $\mathcal{I}$ of the variables interacting with one another, by the $\mathcal{M}$-marginal distribution where it is defined, $\mathcal{M} \supseteq \mathcal{I}$, and by the logit type assigned to each variable of $\mathcal{M}$. At first we examine the problem of allocating the interaction sets among the marginal sets within which they may be defined.

An ordered family $\mathcal{H} = \{\mathcal{M}_1, ..., \mathcal{M}_s\}$ of distinct marginal sets is called *hierarchical family of marginal sets* if $\mathcal{M}_k$ is not a subset of $\mathcal{M}_h$ for every $h < k$, $k = 2, ..., s$. Let $\mathcal{F}_k$ be the family of interaction sets allocated within the $\mathcal{M}_k$-marginal distribution and let $\mathcal{P}_k = \mathcal{P}(\mathcal{M}_k)$ be the family of all non-empty subsets of $\mathcal{M}_k$.

Given a hierarchical family $\mathcal{H}$ of marginal sets a family of interaction sets is called *complete hierarchical family of interaction sets* if (i) every interaction set $\mathcal{I}, \mathcal{I} \in \mathcal{P}(Q)$, is assigned to one $\mathcal{M}_k$-marginal distribution in $\mathcal{H}$, (ii) $\mathcal{F}_1 = \mathcal{P}_1$ and $\mathcal{F}_k = \mathcal{P}_k \setminus \bigcup_{h<k} \mathcal{F}_h$.

We now introduce Bartolucci, Colombi and Forcina (Bartolucci *et al.* 2007) generalized marginal interactions that include the well known types of logits: local, baseline, global, continuation and reverse-continuation, the types of generalized log-odds ratios discussed by Douglas *et al.* (1990) and the recursive or nested logits and log-odds ratios introduced by Cazzaro and Colombi (2006b). We start from logits defined on marginal distributions and log-odds ratios defined on bivariate distributions which are the simplest type of generalized marginal interactions.

Given $r_j - 1$ pairs $\mathcal{B}_j(m_j, 0)$, $\mathcal{B}_j(m_j, 1)$, $m_j = 1, 2, ..., r_j - 1$, of disjoint subsets of $\mathcal{A}_j$, the logits, defined on a marginal distribution, are the log-probability odds: $\log \frac{P(A_j \in B_j(m_j,1))}{P(A_j \in B_j(m_j,0))}$.

The sets $\mathcal{B}_j(m_j, 0)$ are equal to $\{a_{jm_j}\}$ for local and continuation logits, to $\{a_{ji_j} : i_j = 1, \ldots, m_j\}$ for global and reverse continuation logits, $m_j = 1, 2, ..., r_j - 1$. The sets $\mathcal{B}_j(m_j, 1)$ are equal to $\{a_{j(m_j+1)}\}$ for local and reverse continuation logits, to $\{a_{ji_j} : i_j = m_j + 1, \ldots, r_j\}$ for global and continuation logits, $m_j = 1, 2, ..., r_j - 1$. Base-line logits are defined by setting the sets $\mathcal{B}_j(m_j, 0)$ to be equal to $\{a_{j1}\}$ and the sets

$\mathcal{B}_j(m_j, 1)$ to be equal to $\{a_{j(m_j+1)}\}$ for any $m_j < r_j$.

For recursive or nested logits the sets $\mathcal{B}_j(m_j, 0)$ and $\mathcal{B}_j(m_j, 1)$ define a *Coherent Complete Hierarchy of Sets* as specified in details in Cazzaro and Colombi (2006b).

Once logit types for the categorical variables A_1 and A_2 are specified, it is easy to obtain the probabilities: $p_{\{1,2\}}(h_1, h_2; m_1, m_2) = pr(A_1 \in \mathcal{B}_1(m_1, h_1), A_2 \in \mathcal{B}_2(m_2, h_2))$, $m_1 = 1, 2, ..., r_1 - 1, m_2 = 1, 2, ..., r_2 - 1, h_1 = 0, 1, h_2 = 0, 1$. A family of generalized log-odds ratios, defined on the bivariate distribution of A_1, A_2, is composed by the standard log-odds ratios $\ln \frac{p_{\{1,2\}}(1,1;m_1,m_2)p_{\{1,2\}}(0,0;m_1,m_2)}{p_{\{1,2\}}(0,1;m_1,m_2)p_{\{1,2\}}(1,0;m_1,m_2)}$, $m_1 = 1, 2, ..., r_1 - 1$, $m_2 = 1, 2, ..., r_2 - 1$.

When the same logit type is used for A_1 and A_2, a family of symmetric odds ratios is defined, the family is asymmetric otherwise. A family of odds ratios is denoted by the name of the logit type used for A_1 and by the name of the logit type used for A_2 (local-global o.r., local-continuation o.r., etc.). If the same logit type is used for both variables the name is not repeated (local o.r., global o.r., continuation o.r., etc.).

Given a vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_q)'$ of q components, $\boldsymbol{x}_{\mathcal{M}}$ is the vector with components $x_j : j \in \mathcal{M}$. The notation $\mathbf{1}_{\mathcal{M}}$ indicates a vector of ones of dimension equal to the cardinality $|\mathcal{M}|$ of $\mathcal{M}$, the dimension is not specified when it is clear from the context. If $\boldsymbol{x}_{\mathcal{M}\cup\mathcal{I}}$ is a vector such that: $\boldsymbol{x}_{\mathcal{M}} = \boldsymbol{h}_{\mathcal{M}}, \boldsymbol{x}_{\mathcal{I}} = \boldsymbol{k}_{\mathcal{I}}$ we write $\boldsymbol{x}_{\mathcal{M}\cup\mathcal{I}} = (\boldsymbol{h}_{\mathcal{M}}, \boldsymbol{k}_{\mathcal{I}})$.

In order to introduce in full generality the generalized marginal interactions we define the marginal probabilities: $p_{\mathcal{M}}(\boldsymbol{h}_{\mathcal{M}}; \boldsymbol{m}_{\mathcal{M}}) = P(A_j \in \mathcal{B}_j(m_j, h_j), \forall j \in \mathcal{M})$, where $\boldsymbol{m}_{\mathcal{M}}$ is a row vector of integers $m_j,\ 1 \le m_j < r_j,\ j \in \mathcal{M}$, and $\boldsymbol{h}_{\mathcal{M}}$ is a row vector whose elements, $h_j,\ j \in \mathcal{M}$, are equal to zero or to one; these probabilities are probabilities in a contingency table where the variables $A_j, \forall j \in \mathcal{M}$, have been dichotomized according to the categories $\mathcal{B}_j(m_j, 0)$ and $\mathcal{B}_j(m_j, 1)$ and where the variables $A_j, \forall j \notin \mathcal{M}$, have been marginalized. Note that different $\boldsymbol{m}_{\mathcal{M}}$ denote different tables while different $\boldsymbol{h}_{\mathcal{M}}$ denote different probabilities within the same table. In general $\mathcal{B}_j(m_j, 0) \cup \mathcal{B}_j(m_j, 1) \subset \mathcal{A}_j$, thus the probabilities of the above mentioned tables do not always sum to one.

The generalized marginal interactions $\eta_{\mathcal{I};\mathcal{M}}(\boldsymbol{m}_{\mathcal{I}})$ are standard baseline log-linear interactions defined in the previous marginalized and aggregated tables. A formal definition of the generalized marginal interactions is:

$$\eta_{\mathcal{I};\mathcal{M}}(\boldsymbol{m}_{\mathcal{I}}) = \sum_{\mathcal{K}\subseteq\mathcal{I}} (-1)^{|\mathcal{I}\setminus\mathcal{K}|} \log p_{\mathcal{M}}(\mathbf{0}_{\mathcal{M}\setminus\mathcal{K}}, \mathbf{1}_{\mathcal{K}}; \boldsymbol{m}_{\mathcal{I}}, \mathbf{1}_{\mathcal{M}\setminus\mathcal{I}}). \tag{1}$$

The type of logits adopted for each variable should carry over when defining higher order interactions within the same marginal distribution, but not necessarily between different marginal distributions.

2.2. Bartolucci Colombi Forcina main result

Generalizing a previous result of Bergsma and Rudas (2002), Bartolucci, Colombi and Forcina (Bartolucci *et al.* 2007) have proved that the interactions $\eta_{\mathcal{I};\mathcal{M}_k}(\boldsymbol{m}_{\mathcal{I}}),\ \forall \mathcal{I} \in \mathcal{F}_k,\ \forall \mathcal{M}_k \in \mathcal{H}$, where $\boldsymbol{m}_{\mathcal{I}}$ is a row vector of integers $m_j,\ m_j = 1, 2, ...r_j - 1,\ \forall j \in \mathcal{I}$, parameterize the joint distribution of the q categorical variables. Any parametrization in function of a family of generalized marginal interactions associated to a complete hierarchical family of interactions is a *Hierarchical Multinomial Marginal model* HMM .

When the data come from S different strata, a vector $\boldsymbol{\eta}_s$, $s = 1, 2, ..., S$, of generalized marginal interactions is defined within each stratum and then the differences between strata are described by the linear model $\boldsymbol{\eta}_s = \boldsymbol{X}_s\boldsymbol{\beta}, \quad s = 1, 2, ...S$, where $\boldsymbol{X}_s$ is a matrix of covariates that describes the $s-th$ stratum. It is easy to see that HMM models are a special case of the HLP models introduced by Lang (2005). Altough HMM models have been developed regardless of the HLP models of Lang (2005) looking at them as special cases of HLP models allows us to consider sample sizes of some strata as random and enables us to resort to the general asymptotic and computational results of Lang (2004, 2005) (see also Cazzaro Colombi, 2008). Note however that the main difference with HLP models is that in HMM models the link function is invertible.

3. Parameterization of block recursive multinomial models

Examples of block recursive models have been examined in Bartolucci *et al.* (2007 sec. 2.4) because they represent an interesting setting to show how hierarchical family of marginal sets, complete hierarchical family of interaction sets and the associated generalized marginal interactions can be defined in practice. An application of these models to real data can also be found in Colombi e Forcina (2001). In this section we take a more general approach and we show the usefulness of the HMM models to parameterize the block recursive multinomial models specified by the Lauritzen, Wermuth and Frydenberg (LWF) Markov properties (Frydenberg, 1990) or by the Andersson, Madigan and Perlam (AMP) Markov properties (Andersson *et al.*, 2001) associated to a chain graph. As in section 2, sets of variables together with their joint distribution will be denoted by sets of integers. Let G be a chain graph having the sets of vertices $\mathcal{Q} = \{i : i = 1, 2, ..., q\}$ and with chain components denoted by $\tau_m, m = 1, 2, ..., s$. Given a subset $\mathcal{M}$ of vertices of a graph G, $pa_G(\mathcal{M})$, $nd_G(\mathcal{M})$, $nb_G(\mathcal{M})$, $cl_G(\mathcal{M})$ will denote the sets of parents, non descendants, neighbours and the closure of $\mathcal{M}$ respectively. Let G_m be the subgraph induced by the chain component τ_m. When G_m is not complete, $\mathcal{C}_m$ denotes the family of complete subsets and $\mathcal{D}_m$ the family of connected subsets of τ_m. Furthermore $\mathcal{K}$ is the directed acyclic graph having the chain components τ_k as vertices. In this graph, τ_h is a child of τ_k if $\exists i \in \tau_h : pa_G(i) \cap \tau_k \neq \emptyset$. It is assumed that the numbering of the chain components is such that the number m of a parent τ_m is smaller than the ones of its children and finally let it be $\mathcal{M}_m = \cup_{i=1}^{m}\tau_i$, $m = 1, \ldots, s$, $\mathcal{M}_0 = \emptyset$. For a less concise review of this graph-terminology see Andersson *et al.* (2001) and the bibliography herein quoted. A joint probability function $\boldsymbol{\pi}$ is a LWF block recursive model associated to the graph G iff it satisfies the following conditional independencies (Andersson *et al.*, 2001):

$$\begin{aligned} \tau_m \quad &\perp\!\!\!\perp \quad \mathcal{M}_{m-1} \setminus pa_{\mathcal{K}}(\tau_m)|pa_{\mathcal{K}}(\tau_m)[\boldsymbol{\pi}] && (2)\\ \forall m : \quad & m = 2, \ldots, s, \\ S \quad &\perp\!\!\!\perp \quad pa_{\mathcal{K}}(\tau_m) \setminus pa_G(S)|(pa_G(S) \cup nb_G(S))[\boldsymbol{\pi}] && (3)\\ \forall m : \quad & m = 2, \ldots, s,\ \forall S \subseteq \tau_m, \\ S \quad &\perp\!\!\!\perp \quad (\tau_m) \setminus cl_{G_m}(S)|pa_{\mathcal{K}}(\tau_m) \cup nb_G(S)[\boldsymbol{\pi}] && (4)\\ \forall m : \quad & m = 1, \ldots, s,\ \forall S \subseteq \tau_m. \end{aligned}$$

A joint probability function $\boldsymbol{\pi}$ of $\mathcal{Q}$ is an AMP block recursive model associated to the graph G iff it satisfies (2,4) and the following conditional independencies (Andersson *et* *al.*, 2001):

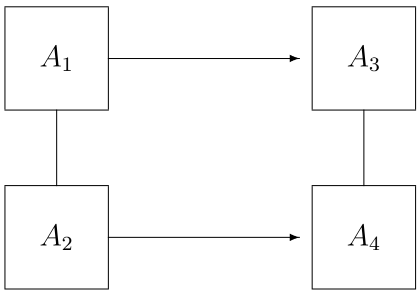
$$S \perp\!\!\!\perp pa_{\mathcal{K}}(\tau_m) \setminus pa_G(S)|pa_G(S)[\boldsymbol{\pi}],\ \forall m : m = 2, \ldots, s,\ \forall S \subseteq \tau_m. \quad (5)$$

At first we consider the case where every G_m is complete that is when $\tau_m \setminus cl_{G_m}(S) = \emptyset$, $\forall S \subseteq \tau_m$, $m = 1, 2, \ldots, s$, so that (4) is empty. To specify HMM models, for which the LWF conditional independencies (2,3) are equivalent to the nullity of some generalized marginal interactions, let us introduce the following family of marginal sets $P_{m,\tau_m} = pa_{\mathcal{K}}(\tau_m) \cup \tau_m$, and $\mathcal{M}_m$, $m = 1, \ldots, s$. The previous marginal sets must be ordered by increasing the index m. Moreover for a given m the marginal set P_{m,τ_m}, must precede the marginal set $\mathcal{M}_m$. If $\mathcal{M}_{m-1} = pa_{\mathcal{K}}(\tau_m)$, $\mathcal{M}_m$ must not be considered because this marginal set is a duplicated of $P_{m,\tau_m} = pa_{\mathcal{K}}(\tau_m) \cup \tau_m$. It is easy to see that this order satisfies the definition of section 2 and so the previous marginal sets can be used to define a complete hierarchical family of interaction sets in the following way: i) to P_{m,τ_m} is associated the family of interaction sets $\mathcal{F}^*_{m,\tau_m} = \{\mathcal{I} : \mathcal{I} = P \cup S,\ \forall P \subseteq pa_{\mathcal{K}}(\tau_m), \forall S \in \mathcal{P}(\tau_m)\}$, $m = 1, 2, ..., s$, ii) to every marginal $\mathcal{M}_m$ is associated the family of the interactions $\mathcal{F}_m = \{\mathcal{I} : \mathcal{I} = P \cup S, \forall P \in \mathcal{P}(\mathcal{M}_{m-1}) \setminus \mathcal{P}(pa_{\mathcal{K}}(\tau_m)), \forall S \in \mathcal{P}(\tau_m)\}$. To express the hypotheses (2) it is convenient to use logits of log-linear type (local or base-line) within every interaction family $\mathcal{F}_m$. Moreover having in mind the hypotheses (3), it is useful to define the interactions within every $\mathcal{F}^*_{m,\tau_m}$ by using local or base-line logits for the variables not belonging to τ_m and logits of any type for the variables in τ_m. From standard results on graphical modeling it follows that (2) is equivalent to the nullity of the interactions for every interaction set $\mathcal{I}$ belonging to $\mathcal{F}_m$ and that (3) is equivalent to the nullity of the interactions defined in $\mathcal{F}^*_{m,\tau_m}$ for every interaction set not belonging to the subset $\{\mathcal{I} : \mathcal{I} = P \cup S,\ \forall S \in \mathcal{P}(\tau_m),\ \forall P \subseteqq \bigcap_{s \in S} pa_G(\{s\})\}$.

We now define a HMM model for which the AMP conditional independencies (2,5) are equivalent to the nullity of some generalized marginal interactions. To this end let us consider the following family of marginal sets: $P_{m,S} = pa_{\mathcal{K}}(\tau_m) \cup S$, $\forall S \in \mathcal{P}(\tau_m)$, $\mathcal{M}_m$, $m = 1, \ldots, s$. The previous marginal sets must be ordered by increasing the index m. Moreover for a given m all the marginal sets $P_{m,S}$, ordered coherently with the partial order of inclusion, must precede the marginal set $\mathcal{M}_m$. If $\mathcal{M}_{m-1} = pa_{\mathcal{K}}(\tau_m)$, $\mathcal{M}_m$ must be omitted. It is easy to see that this order satisfies the definition of section 2 and so the previous marginal sets can be used to define a complete hierarchical family of interaction sets in the following way: i) to every $P_{m,S}$ is associated the family of the interactions $\mathcal{F}_{m,S} = \{\mathcal{I} : \mathcal{I} = P \cup S,\ \forall P \subseteq pa_{\mathcal{K}}(\tau_m)\}$, ii) to every marginal $\mathcal{M}_m$ is associated the same family of interactions $\mathcal{F}_m = \{\mathcal{I} : \mathcal{I} = S \cup P,\ \forall S \in \mathcal{P}(\tau_m),\ \forall\ P \in \mathcal{P}(\mathcal{M}_{m-1}) \setminus \mathcal{P}(pa_{\mathcal{K}}(\tau_m))\}$ introduced for the LWF models. To express the hypotheses (5) it is convenient to define the interactions within every $\mathcal{F}_{m,S}$ by assigning local or base-line logits to the variables not belonging to S and logits of any type to the variables in S. Note that in this way local or base-line logits are always assigned to variables not in $\mathcal{I}$. For every m, the interactions defined within the families $\mathcal{F}_{m,S}$ define a Glonek and McCullagh (1995) multivariate logit model for the joint distribution of the variables in τ_m conditioned by the variables in $pa_{\mathcal{K}}(\tau_m)$. We have already seen that (2) is equivalent to the nullity of the interactions for every $\mathcal{I}$ belonging to $\mathcal{F}_m$. Moreover from standard results on graphical modelling it follows that (5) is equivalent to the nullity of the interactions for every interaction set not belonging to the subset $\{\mathcal{I} : \mathcal{I} = P \cup S,\ \forall P \subseteq pa_G(S)\}$ of $\mathcal{F}_{m,S}$, $\forall S \in \mathcal{P}(\tau_m)$, $m = 1, \ldots, s$.

When the G_m's are not complete, to satisfy the further set of conditional independencies (4), the interactions associated to the sets $\{\mathcal{I} : \mathcal{I} = P \cup S, \forall P \subseteq pa_{\mathcal{K}}(\tau_m), \forall S \in \mathcal{P}(\tau_m) \setminus \mathcal{C}_m\}$ of F^*_{m,τ_m} must be equal to zero. These interactions are parameters of the LWF models but, in the case of AMP models, they are not because these interactions are already contained in the sets $F_{m,S}$, $S \in \mathcal{P}(\tau_m)$. Thus the previous constraints do not involve interactions associated to the complete hierarchical family of interaction sets $F_{m,S}$, $\forall S \in \mathcal{P}(\tau_m), F_m, m = 1, 2, ...s$ introduced to parameterize the AMP model. The previous drawback of the AMP models is avoided if the undirected graph Markov property (4) is replaced by the bi-directed graph Markov property $S \perp\!\!\!\perp (\tau_m) \setminus cl_{G_m}(S)|pa_{\mathcal{K}}(\tau_m)[\boldsymbol{\pi}]$, $\forall S \in \mathcal{D}_m$, $m = 1, \ldots, s$. From a result due to Lupparelli *et al.* (2008), it follows that the previous Markov property is equivalent to the nullity of the AMP interactions that are associated to the sets belonging to $\mathcal{F}_{m,S}$, $\forall S \notin \mathcal{D}_m$, $m = 1, \ldots, s$. All the constraints so far specified for LWF models and AMP models can be expressed in the form $\boldsymbol{C} \ln(\boldsymbol{M\pi}) = \boldsymbol{0}$ as shown by Bartolucci, Colombi and Forcina (2007) and, under these constraints, the multinomial log-likelihood can be maximized as shown by Lang (2004,2005). To clarify the previous definitions for AMP models we consider the seemingly unrelated logit regressions of Cox and Wermuth (1996), represented by the graph of figure 1.

Figure 1: *AMP Markov conditional independencies graph for the Cox, Wermuth seemingly unrelated logit regressions*



The variables A_1 and A_2 are explanatory for the variables A_3 and A_4 and the four variables are assumed to be ordinal. In this simple example there are no conditions (2,4) because the graph has only two complete chain components. The conditions (5) state that A_4 is independent of A_1 given A_2 and that A_3 is independent of A_2 given A_1. These conditional independence hypotheses can be expressed by linear equality constraints on the parameters of the HMM model reported in Table 1.

In the column *inter. sets* the variables involved in the interactions are reported and the column *marg. sets* describes the marginal distribution where the interactions are defined. For every interaction the *type* column describes the logits, the log-odds ratios and the higher order interactions used. The labels have the following meanings: *g* is a global logit, *gg* is a global log-o.r., *lg* is a local-global log-o.r., the other symbols are generalized interactions of higher order described in details in Colombi and Forcina, (2001). In particular, *llg* are contrasts of four global logits and *llgg* are contrasts of four global log-odds ratios. In Table 1 (=) denotes the interactions constrained to be null by the conditional independence hypotheses.

Table 1: *A parametrization of the Cox, Wermuth seemingly unrelated logit regressions model*

$inter.sets$	$marg.sets$	$type$	$inter.sets$	$marg.sets$	$type$
A_1	A_1, A_2	*g*	A_4	A_1, A_2, A_4	*g*
A_2	A_1, A_2	*g*	A_1, A_4 (=)	A_1, A_2, A_4	*lg*
A_1, A_2	A_1, A_2	*gg*	A_2, A_4 (+)	A_1, A_2, A_4	*lg*
			A_1, A_2, A_4 (=)	A_1, A_2, A_4	*llg*
A_3	A_1, A_2, A_3	*g*	A_3, A_4	A_1, A_2, A_3, A_4	*gg*
A_1, A_3 (+)	A_1, A_2, A_3	*lg*	A_1, A_3, A_4	A_1, A_2, A_3, A_4	*lgg*
A_2, A_3 (=)	A_1, A_2, A_3	*lg*	A_2, A_3, A_4	A_1, A_2, A_3, A_4	*lgg*
A_1, A_2, A_3 (=)	A_1, A_2, A_3	*llg*	A_1, A_2, A_3, A_4	A_1, A_2, A_3, A_4	*llgg*

4. Other applications of HMM models

As shown by Bartolucci *et al.* (2007) and Colombi and Forcina (2000, 2001) among many others, inequality constraints are a fundamental tool to specify hypotheses of stochastic dominance, monotone dependence and positive association in contingency tables. The great flexibility of HMM models, in modeling such type of hypotheses, is highlighted by Bartolucci, *et al.* (2007), Colombi and Forcina (2001), Bartolucci *et al.* (2001), Cazzaro and Colombi (2006a, 2006b). These works have in common the problem that the likelihood ratio statistics to test inequality constrained models has a non standard asymptotic distribution, called chi-bar squared, which is a mixture of chi squared distributions. The book of Silvapulle and Sen (2005) is an updated survey on testing inequality constraints. For example in the seemingly unrelated logit regressions example of Cox, Wermuth (+) in Table 1 denotes the interactions constrained to be non-negative under the hypothesis of monotone positive dependence of A_3 on A_1 and of A_4 on A_2. More recently Cazzaro *et al.* (2007), Colombi and Giordano (2006, 2008) have shown the usefulness of HMM models for modeling hypotheses of Granger-non causality and lumpability in multivariate Markov chains and Hidden Markov Models.

References

Andersson S.A., Madigan D., Perlman M.D. (2001) Alternative Markov properties for chain graphs, *Scandinavian Journal of Statististics*, 28, 33-85.

Agresti A., Coull B.A. (1998) Order-restricted inference for monotone trend alternatives in contingency tables, *Computational Statistics & Data Analysis*, 28, 139-155.

Bartolucci F., Dardanoni V., Forcina A. (2001) Positive quadrant dependence and marginal modelling in two-way tables with ordered margins, *Journal of the American Statistical Association*, 96, 1497-1505.

Bartolucci F., Colombi R., Forcina A. (2007) An extended class of marginal link functions for modelling contingency tables by equality and inequality constraints, *Statistica Sinica*, 17, 691-711.

Bergsma W.P., Rudas T. (2002) Marginal models for categorical data, *Annals of Statistics*, 30, 140-159.

Cazzaro M., Colombi R. (2006a) Maximum likelihood inference for log-linear models subject to constraints of double monotone dependence, *Statistical Methods and Applications*, 15, 177-190.

Cazzaro M., Colombi R. (2006b) Modelling two way contingency tables with recursive logits and odds ratios, *Statistical Methods and Applications*, to appear.

Cazzaro M., Colombi R. (2008) Multinomial-Poisson models subject to inequality constraints, submitted.

Cazzaro M., Colombi R., Giordano S. (2007) Testing Markov chain lumpability, *Proceedings of the 22nd International Workshop on Statistical Modelling*, Barcelona, 2-6 July, 158-163.

Colombi R., Forcina A. (2000) Modelling discrete data by equality and inequality constraints, *Statistica*, 60, 195-214.

Colombi R., Forcina A. (2001) Marginal regression models for the analysis of positive association of ordinal response variables, *Biometrika*, 88, 1007-1019.

Colombi R., Giordano S. (2006) Alcune indipendenze condizionali nelle serie storiche categoriali bivariate, *Statistica*, 1, 19-38

Colombi R., Giordano S. (2008) Lumpability for discrete hidden Markov models, submitted.

Cox D.R., Wermuth N. (1996) *Multivariate Dependencies: Models, Analysis and Interpretation*, Chapman & Hall, London.

Dardanoni V., Forcina A. (1998) A unified approach to likelihood inference on stochastic orderings in a nonparametric context, *Journal of the American Statistical Association*, 93, 1112-1123.

Douglas R., Fienberg S.E., Lee M.T., Sampson A.R., Whitaker L.R. (1990) Positive dependence concepts for ordinal contingency tables, in: *Topics in Statistical Dependence*, Block H.W., Sampson A.R. & Sanits T.H. (Eds.), Institute of Mathematical Statistics, Lecture Notes, Monograph Series, Haywar California, 189-202.

Frydenberg M. (1990) The chain graph Markov property, *Scandinavian Journal of Statististics*, 17, 333-353.

Glonek G.F.V., McCullagh P. (1995) Multivariate logistic models for contingency tables, *Journal of the Royal Statistical Society, B*, 57, 533-546.

Lang J.B. (2004) Multinomial Poisson homogeneous models for contingency tables, *The Annals of Statistics*, 32, 340-383.

Lang J.B. (2005) Homogeneous linear predictor models for contingency tables, *Journal of the American Statistical Association*, 100, 121-134.

Lupparelli M., Marchetti G. M., Bergsma W. P. (2008) Parameterizations and fitting of bi-directed graph models to categorical data, submitted *arXiv:0801.1440*.

Silvapulle M.J., Sen P.K. (2005) *Constrained Statistical Inference,* Wiley, New York.