Context-specific independencies in Hierarchical Multinomial Marginal Models

Federica Nicolussi, Manuela Cazzaro*

May 28, 2018

Abstract Marginal or conditional independencies are well known relationships among variables involved in a contingency table. In this paper we handle with categorical (ordinal) variables and we focus on the (in)dependence relationships under this marginal and conditional perspective in addition to context-specific point of view. The last statement concerns independencies holding only in a subspace of the outcome space. We take advantage from the Hierarchical Multinomial Marginal models environment and we provide several original results about the representation of context-specific independencies through these models. An application about the innovation degree of the Italian enterprises is provided.

Keywords Context-specific independencies, ordinal variables, Hierarchical Multinomial Marginal models.

1 Introduction

In this work we deal with categorical (ordinal) variables collected in a contingency table and we propose a model able to capture different kind of independence relationships involving ordinal variables. Different models have been proposed in the literature with the aim of describing (in)dependence relationships among the variables focusing on the independence or the dependence structure. In particular, we will refer to the Marginal models, Agresti (2013), that imposes constraints on marginal distributions of the tables. More specifically, we will focus on Colombi *et al.* (2014) Hierarchical Multinomial Marginal (HMM) models that are specified by a set of marginal distributions of the contingency table together with a set of interactions defined within different marginal distributions. Particular case of these models are the classical Log-Linear models, the Bergsma & Rudas (2002) Marginal models, even extended by Bartolucci *et al.* (2007) and Cazzaro & Colombi (2014) to more general types of interactions, the Glonek & McCullagh (1995) Multivariate Logistic models. In particular, in this work we take advantage of the possibility of using different interactions that are significant also when we handle with ordinal variables, Cazzaro & Colombi (2014).

In this environment, we will focus on the relationships among a set of categorical (ordinal) variables under the perspective of testing, simultaneously, marginal, conditional and context-specific (CS) independencies. The first two are well known relationships among variables involved in a contingency table, the CS independence, instead, is a conditional independence which holds only in a subspace of the outcome space. For instance, given 3 variables X_1, X_2 and X_3, we have $X_1 \perp X_2 | X_3 = 1$ and $X_1 \not\perp X_2 | X_3 \neq 1$. It is interesting to study this kind of independence as it

*Department of Statistics and Quantitative Methods, University of Milan-Bicocca. Via Bicocca degli Arcimboldi, 8, 20126, Milan, Italy. *Email:* federica.nicolussi@unimib.it, manuela.cazzaro@unimib.it.

allows us to focus on the modality(ies) which discriminate(s) and really affect(s) the connection among the variables.

The work follows this structure. At first we give an overview of the HMM models with a special attention to the representation of CS independence via HMM models, in Section 2. In this section, we reach out the same results of Nyman *et al.* (2016) by using a different approach concerning the variables coded with *baseline* logits. It is worthwhile to note that the known results in the literature are carried out limited to the classical log-linear models. Furthermore, in Subsection 3.2 and 3.3, we provide as new result, how it is possible to define CS independence by using appropriate interactions for ordinal variables. Finally, in Section 4 some applications to a real dataset on the innovation status of small and medium Italian firms are shown. The conclusion is reported in Section 5. All the proofs of the theorems lie in the Appendix A in order to make more flowing the paper.

2 Hierarchical Multinomial Marginal models

The Hierarchical Multinomial Marginal (HMM) models, defined by Colombi *et al.* (2014) and based on the work of Bartolucci *et al.* (2007) extended by Cazzaro & Colombi (2014), are used here in order to describe marginal, conditional and CS independence statements also when we deal with ordinal variables. It is worthwhile to highlight that HMM models are specified by a set of marginal distributions of the contingency table together with a set of interactions (the HMM parameters) defined within different marginal distributions according to the rules of *hierarchy* and *completeness*, see Bergsma & Rudas (2002) and Bartolucci *et al.* (2007). This means that every interaction is "complete" as it is uniquely defined in one marginal distribution and it satisfies the "hierarchy" condition because it is defined within the first marginal distribution which involves it.

Let us consider q categorical variables $\mathcal{Q} = (X_1, \dots, X_q)$ taking values $i_{\mathcal{Q}} \in (i_1, \dots, i_q)$ in the contingency table $\mathcal{I} = (I_1 \times \cdots \times I_q)$. Thus, the generic variable X_j takes values in $\{1, \dots, I_j\}$. Let us denote the probability of a generic cell, $i_{\mathcal{Q}}$ of $\mathcal{I}$, with $\pi(i_{\mathcal{Q}})$, thus the probability of the whole contingency table is represented by the vector π, obtained by stacking all the $\pi(i_{\mathcal{Q}})$ in the lexicographical order. A parameterization of the vector π will be defined through the function $\eta = h(\pi)$.

Similarly, by considering a subset of variables $\mathcal{M} \subseteq \mathcal{Q}$, which generate the contingency table $\mathcal{I}_{\mathcal{M}}$, the marginal probability of the generic cell $i_{\mathcal{M}}$ is $\pi_{\mathcal{M}}(i_{\mathcal{M}})$, obtained by summarizing respect to the variables $\mathcal{Q}\backslash\mathcal{M}$; the marginal probability distribution $\pi_{\mathcal{M}}$ is obtained by stacking all the $\pi_{\mathcal{M}}(i_{\mathcal{M}})$ in the lexicographical order.

The HMM parameters η are contrasts among the logarithms (of sums) of probabilities of disjoint subsets of cells on a marginal distribution $\mathcal{M}$ and they are characterized by the set $\mathcal{L}$, $\mathcal{L} \subseteq \mathcal{M}$, of variables involved and the marginal distribution $\mathcal{M}$ where they are defined. When there is only a single variable in the set $\mathcal{L}$, the corresponding set of parameters are logits while increasing the number of variables in the set $\mathcal{L}$, the parameters are contrasts of logits of increasing order. Here we take advantage from different types of aggregation criteria to catch the order inherent in the variable modalities. In particular, by considering a variable X_j, in addition to the classical *baseline* logit, $\log(\pi_j(I_j)) - \log(\pi_j(i_j))$ with $i_j = 1, \dots I_j - 1$, we use also the *local* logit, $\log(\pi_j(i_j+1)) - \log(\pi_j(i_j))$ with $i_j = 1, \dots I_j - 1$ and the *continuation* logit, $\log\left(\sum_{i=i_j+1}^{I_j} \pi_j(i)\right) - \log(\pi_j(i_j))$ with $i_j = 1, \dots I_j - 1$. While the *baseline* logits are useful for the nominal variables, the other two kinds of logits are typically used to capture some trend within the modalities of ordinal variables, see Cazzaro & Colombi (2014). In general, we refer to the HMM parameters with the symbol $\eta_{\mathcal{L}}^{\mathcal{M}}(i_{\mathcal{L}})$ where $\mathcal{L}$ is the interaction set, $\mathcal{M}$ is the marginal distribution where the parameter is evaluated and $i_{\mathcal{L}}$ is the vector of modalities associated to the variables in $\mathcal{L}$ which the parameter refers. In the probabilities involved by the parameters, the indexes associated to the variables $\mathcal{M}\backslash\mathcal{L}$, not involved in the interaction, are set to the "reference" modalities that in this work we set to the last one $I_{\mathcal{M}\backslash\mathcal{L}}$ without loss of generality.

By considering a variable X_j, the vector of logits evaluated in the marginal $\mathcal{M}$, according to the *baseline* criterion is

$$\eta_j^{\mathcal{M}} = \begin{pmatrix} \log\left(\frac{\pi_{\mathcal{M}}(I_{\mathcal{M}})}{\pi_{\mathcal{M}}(i_j=1, i_{\mathcal{M}\backslash j}=I_{\mathcal{M}\backslash j})}\right) \\ \log\left(\frac{\pi_{\mathcal{M}}(I_{\mathcal{M}})}{\pi_{\mathcal{M}}(i_j=2, i_{\mathcal{M}\backslash j}=I_{\mathcal{M}\backslash j})}\right) \\ \vdots \\ \log\left(\frac{\pi_{\mathcal{M}}(I_{\mathcal{M}})}{\pi_{\mathcal{M}}(i_j=I_j-1, i_{\mathcal{M}\backslash j}=I_{\mathcal{M}\backslash j})}\right) \end{pmatrix}; \tag{1}$$

according to the *local* criterion is

$$\eta_j^{\mathcal{M}} = \begin{pmatrix} \log\left(\frac{\pi_{\mathcal{M}}(i_j=2, i_{\mathcal{M}\backslash j}=I_{\mathcal{M}\backslash j})}{\pi_{\mathcal{M}}(i_j=1, i_{\mathcal{M}\backslash j}=I_{\mathcal{M}\backslash j})}\right) \\ \log\left(\frac{\pi_{\mathcal{M}}(i_j=3, i_{\mathcal{M}\backslash j}=I_{\mathcal{M}\backslash j})}{\pi_{\mathcal{M}}(i_j=2, i_{\mathcal{M}\backslash j}=I_{\mathcal{M}\backslash j})}\right) \\ \vdots \\ \log\left(\frac{\pi_{\mathcal{M}}(I_{\mathcal{M}})}{\pi_{\mathcal{M}}(i_j=I_j-1, i_{\mathcal{M}\backslash j}=I_{\mathcal{M}\backslash j})}\right) \end{pmatrix}; \tag{2}$$

according to the *continuation* criterion is

$$\eta_j^{\mathcal{M}} = \begin{pmatrix} \log\left(\frac{\sum_{i=2}^{I_j}\pi_{\mathcal{M}}(i_j=i, i_{\mathcal{M}\backslash j}=I_{\mathcal{M}\backslash j})}{\pi_{\mathcal{M}}(i_j=1, i_{\mathcal{M}\backslash j}=I_{\mathcal{M}\backslash j})}\right) \\ \log\left(\frac{\sum_{i=3}^{I_j}\pi_{\mathcal{M}}(i_j=i, i_{\mathcal{M}\backslash j}=I_{\mathcal{M}\backslash j})}{\pi_{\mathcal{M}}(i_j=2, i_{\mathcal{M}\backslash j}=I_{\mathcal{M}\backslash j})}\right) \\ \vdots \\ \log\left(\frac{\pi_{\mathcal{M}}(I_{\mathcal{M}})}{\pi_{\mathcal{M}}(i_j=I_j-1, i_{\mathcal{M}\backslash j}=I_{\mathcal{M}\backslash j})}\right) \end{pmatrix}. \tag{3}$$

Generally speaking, within a given set of variables A, let us denote with i_A^* the particular cell (or group of cells) which identify the aggregation criterion reference modalities of the variables in A depending on the type of logits assigned to the variables on which the parameters is based. In particular, the index i_A^*, in *baseline* logit is I_A, in the *local* logit is i_j+1 for all $X_j \in A$ and in the *continuation* logit is $\sum_{i \geq i_j+1} i$ for all $X_j \in A$. Higher order parameters are obtained as contrast of logits and preserve the type of coding. Thus, the HMM parameters of the $\mathcal{L}$, defined in the marginal distribution $\mathcal{M}$ have the following form:

$$\eta_{\mathcal{L}}^{\mathcal{M}}(i_{\mathcal{L}}|I_{\mathcal{M}\backslash\mathcal{L}}) = \sum_{\mathcal{J}\subseteq\mathcal{L}} (-1)^{|\mathcal{L}\backslash\mathcal{J}|} \log \pi_{\mathcal{M}}\left(i_{\mathcal{L}\backslash\mathcal{J}}, i_{\mathcal{J}}^*, I_{\mathcal{M}\backslash\mathcal{L}}\right) \tag{4}$$

where $\mathcal{M} \subseteq \mathcal{Q}$ denotes the marginal table $\mathcal{I}_{\mathcal{M}}$ where the parameter is defined; $\mathcal{L} \subseteq \mathcal{M}$ is the subset of variables which the parameter refers. Note that the modalities $I_{\mathcal{M}\backslash\mathcal{L}}$ select the levels of the conditioning variables. In this context we can simply denote:

$$\eta_{\mathcal{L}}^{\mathcal{M}}(i_{\mathcal{L}}|I_{\mathcal{M}\backslash\mathcal{L}}) = \eta_{\mathcal{L}}^{\mathcal{M}}(i_{\mathcal{L}}). \tag{5}$$

Note that for each $\mathcal{L}$ the parameter $\eta_{\mathcal{L}}^{\mathcal{M}}(I_{\mathcal{L}})$ is trivially zero whatever the coding of the variables as ......

The following example considers a marginal set $\mathcal{M}$ composed of two variables and show the possible parameters that we can build within the marginal distribution $\pi_{\mathcal{M}}$ with the different aggregation criteria. Note that, for simplicity we consider all the variables coded with the same aggregation criterion, this however does not always true. In general we choose an aggregation criterion for any variable.

Example 2.1. *Let us consider two variables X_1, X_2 collected in a 3×3 contingency table. In Table 1 are the parameters (4) according to the different coding:*

type	$\eta_1^{12}(i_1)$	$\eta_2^{12}(i_2)$	$\eta_{12}^{12}(i_1 i_2)$
baseline	$\log\left(\frac{\pi_{33}}{\pi_{i_1 3}}\right)$	$\log\left(\frac{\pi_{33}}{\pi_{3 i_2}}\right)$	$\log\left(\frac{\pi_{i_1 i_2}\pi_{33}}{\pi_{i_1 3}\pi_{3 i_2}}\right)$
local	$\log\left(\frac{\pi_{(i_1+1)3}}{\pi_{i_1 3}}\right)$	$\log\left(\frac{\pi_{3(i_2+1)}}{\pi_{3 i_2}}\right)$	$\log\left(\frac{\pi_{i_1 i_2}\pi_{(i_1+1)(i_2+1)}}{\pi_{(i_1+1)i_2}\pi_{i_1(i_2+1)}}\right)$
cont	$\log\left(\frac{\sum_{i_1'>i_1}\pi_{(i_1')3}}{\pi_{i_1 3}}\right)$	$\log\left(\frac{\sum_{i_2'>i_2}\pi_{3(i_2')}}{\pi_{3 i_2}}\right)$	$\log\left(\frac{\pi_{i_1 i_2}\sum_{i_1'>i_1, i_2'>i_2}\pi_{(i_1')(i_2')}}{\sum_{i_1'>i_1}\pi_{(i_1')i_2}\sum_{i_2'>i_2}\pi_{i_1(i_2')}}\right)$

Table 1: Different coding for logits and contrasts of logits.

Note that the classical log-linear model is a particular case of HMM model where the parameters are all based on *baseline* logits and there are only one marginal set equal to the joint distribution $\mathcal{M} = \mathcal{Q}$. In general, by using the matrix notation, we can define a HMM model as follows.

Definition 1. *The HMM model is described by the vector of parameters η listed in the lexicographical order, obtained as:*

$$\eta = C \log(M\boldsymbol{\pi}) \tag{6}$$

where C is an adequate contrast matrix, M is the marginalization matrix and $\boldsymbol{\pi}$ is the vector with the probability in lexicographical order.

Note that when the matrix M is the identity matrix, the parameters in (6) denote the classical log-linear models. The matrix C can assume different forms according to the kind of "aggregation" criterion is chosen for the parameters. The construction of the matrix C according to the aggregation criterion is shifted on the Appendix.

In the environment of HMM models, conditional independencies among variables can be tested by imposing to appropriate HMM parameters to be zero. For instance, given three variables X_1, X_2 and X_3, in order to represent the conditional independence $X_1 \perp X_2 | X_3$ we have that $\eta_{12}^{123}(i_{12}) = \eta_{123}^{123}(i_{123}) = 0$ for any $i_{12} \in \mathcal{I}_{12}$ and $i_{123} \in \mathcal{I}_{123}$, where the numbers 1, 2 and 3 in the parameters refer to the variables X_1, X_2 and X_3, respectively.
Bergsma & Rudas (2002) and Bartolucci *et al.* (2007) proved that the above mentioned parameters provide a parameterization of the full joint probability function $\pi_{\mathcal{Q}}$ if and only if the property of hierarchy and completeness are satisfied. These two properties make sure of the smoothness of the parametrization that implies the existence of the maximum likelihood estimation.

3 Context-Specific independence in HMM models

Let us suppose we want to investigate a CS independence among the variables in the marginal set $\mathcal{M}$. Thus, by collecting the variables in the marginal set $\mathcal{M}$ in three subsets, supposing A,

B and C, let us say that we are interested in defining the following statement

$$A \perp B | (C = i'_C), \qquad i'_C \in \mathcal{K} \tag{7}$$

where $A \cup B \cup C = \mathcal{M}$, and i'_C is the vector of certain modalities of variables in C, taking values in $\mathcal{K} \subset \mathcal{I}_C$, for which the conditional independence holds, see among others Højsgaard (2004), Nyman *et al.* (2014), Nyman *et al.* (2016).

In the following subsections we handle the CS independencies in HMM models when the parameters are coded with different criteria.

3.1 Constraints on HMM parameters based on *baseline* logit

Nyman *et al.* (2016) provide the condition to define a CS independence in classical log-linear models. Next we will reach the same condition, in a new way, for the CS independencies on the HMM models with parameters based on baseline logit.

Theorem 3.1. *The CS independence in formula (7) holds if and only if the HMM parameters, based on baseline logits, satisfy the following constraints*

$$\sum_{c \in \mathcal{P}(C)} (-1)^{|C \setminus c|} \eta^{\mathcal{M}}_{vc}(i_v \cap i'_c) = 0 \qquad i_v \in \mathcal{I}_v \qquad i'_c \in (\mathcal{K} \cap \mathcal{I}_c)\,, \tag{8}$$

$\forall v \in \mathcal{V} = \{(\mathcal{P}(A) \setminus \emptyset) \cup (\mathcal{P}(B) \setminus \emptyset)\}$*, where* $\mathcal{P}(\cdot)$ *denotes the power set.*

The Example 3.1 shows step by step how to get the constraints in formula (8).

Example 3.1. *Let us consider four variables collected in the marginal* $\mathcal{I}_{\mathcal{M}}$ *of dimension* $3 \times 3 \times 3 \times 3$ *and let us consider the CS independence* $X_1 \perp X_2 | (X_3 X_4) = (1, 1)$*. The HMM parameter* $\eta^{1234}_{1234}(1111)$ *based on baseline logit can be decomposed as follows*

$$\begin{aligned}
\eta^{1234}_{1234}(1111) &= \log\left(\tfrac{\pi_{3333}\pi_{1133}\pi_{1313}\pi_{3113}\pi_{1331}\pi_{3131}\pi_{3311}\pi_{1111}}{\pi_{1333}\pi_{3133}\pi_{3313}\pi_{1113}\pi_{3331}\pi_{3111}\pi_{1311}\pi_{1131}}\right) = \\
&= \log\left(\tfrac{\pi_{3333}\pi_{1133}\pi_{1313}\pi_{3113}}{\pi_{1333}\pi_{3133}\pi_{3313}\pi_{1113}}\right) + \log\left(\tfrac{\pi_{3333}\pi_{1133}\pi_{1331}\pi_{3131}}{\pi_{1333}\pi_{3133}\pi_{3331}\pi_{1131}}\right) + \\
&\quad - \log\left(\tfrac{\pi_{3333}\pi_{1133}}{\pi_{1333}\pi_{3133}}\right) + \log\left(\tfrac{\pi_{3311}\pi_{1111}}{\pi_{3111}\pi_{1311}}\right) = \\
&= (-1)^{|\{4\}|+1}\eta^{1234}_{123}(111) + (-1)^{|\{3\}|+1}\eta^{1234}_{124}(111) + \\
&\quad (-1)^{|\{3,4\}|+1}\eta^{1234}_{12}(11) + (-1)^{|\{3,4\}|}\eta^{1234}_{12}(11|11).
\end{aligned}$$

From the CS independence we have that $\eta^{1234}_{12}(11|11) = 0$ *and by shifting the right hand side on the left we get:*

$$\eta^{1234}_{1234}(1111) - \eta^{1234}_{123}(111) - \eta^{1234}_{124}(111) + \eta^{1234}_{12}(11) = 0$$

The same equivalence holds for $\eta^{1234}_{1234}(1211)$*,* $\eta^{1234}_{1234}(2111)$ *and* $\eta^{1234}_{1234}(2211)$*.*

Note that, having the CS independence $X_1 \perp X_2 | (X_3 X_4) = (1, 3)$*, the constraints involving the variables* X_4 *at the fourth modality are zero by definition, thus formula (8) becomes*

$$-\eta^{1234}_{123}(i_{123}) + \eta^{1234}_{12}(i_{12}) = 0$$

where $i_{123} \in \{(111), (121), (211), (221)\}$ *and* $i_{12} \in \{(11), (12), (21), (22)\}$*.*

Remark 1. *If in the CS statement in formula (7) $\mathcal{K} = \mathcal{I}_C$, then the constraints in formula (8) satisfy the conditional independence $A \perp B|C$.*

From Remark 1 comes that the CS independence $A \perp B|D(C = i'_C)$, for $i'_C \in \mathcal{K}$, matches with the CS independence $A \perp B|(DC = i'_{DC})$, where $i'_{DC} = i'_D \cap i'_C$,, for $i'_C \in \mathcal{K}$ and $i'_D \in \mathcal{I}_D$. Henceforth, this situations will be described as $A \perp B|(DC) = (*, i'_C)$ where the asterisk denotes we refer to all modalities.

It is appropriate to remember that not all list of independencies are representable throw HMM models. Theorem 3.2 gives the sufficient condition.

First of all, let us define with the symbol $\mathcal{M}(\mathcal{L})$ the marginal set that first, in the partially ordered class of marginals, contains the subset of variables $\mathcal{L}$.

Theorem 3.2. *Given a list of CS independencies such as in formula (7), if it is possible to build a class of marginal distribution $\mathcal{M}_j$, $j = 1, \ldots, s$ such that for each CS independence holds that*

$$C \subseteq \mathcal{M}(\mathcal{L}) \subseteq (A \cup B \cup C) \tag{9}$$

for all interaction set $\mathcal{L} = vc$, with $v \in \mathcal{V}$ and $c \in \mathcal{P}(C)$, that are involved in the constraints according to formula (8), then the list of independencies are representable via HMM models and thus there exist a smooth parametrization associated.

In literature, this topic is deep discussed, see among other Drton *et al.* (2009); Rudas *et al.* (2010); Forcina (2012); Colombi & Forcina (2014). In particular, in the last one it is shown how lists of CS independencies can be parametrized through HMM models while there is no HMM parametrization able to describe the list obtained by the conditional independencies associated to the CS indepencencies ones.

Remark 2. *Given a CS statement as in formula (7), the number of constraints imposed at a saturated log-linear model are $\left[\left(\prod_{j\in(A\cup B)} I_j\right) - 1\right] \times |\mathcal{K}|$.*

As mentioned before, the aim of this work is to provide a model able to represent the CS independence statements by considering also ordinal variables. When we handle with ordinal variables, *baseline* logits are no longer appropriate. The *local*, *continuation* or *reverse* approaches are more suitable. The following subsections deal with these logits.

3.2 Constraints on HMM parameters based on *local* logit

Let us suppose that the conditional set in (7) is composed only of ordinal variables and we use parameters based on *local* logits to code these ones, then the CS independence can be described by Theorem 3.3.

Theorem 3.3. *The CS independence in formula (7) holds if and only if the HMM parameters based on local logits satisfy the following constraints*

$$\sum_{c\in\mathcal{P}(C)} (-1)^{|C\setminus c|} \sum_{i_c \geq i'_c} \eta^{\mathcal{M}}_{vc}(i_{vc}) = 0 \tag{10}$$

$\forall v \in \mathcal{V}$ where $\mathcal{V} = \{(\mathcal{P}(A) \setminus \emptyset) \cup (\mathcal{P}(B) \setminus \emptyset)\}$, $i_{vc} = i_v \cap i_c$, $\forall i_v \in \mathcal{I}_v$ and $\forall i'_c \in (\mathcal{K} \cap \mathcal{I}_c)$.

Example 3.2. *Let us consider the case of three variables collected in a $2 \times 2 \times 4$ contingency table. If we want to consider the CS independence $X_1 \perp X_2 | X_3 = 2$ where all the variables are coded with local approach in the parameters we consider the decomposition in formula (10):*

$$e^{\left(\left(\eta_{123}^{123}(112)+\eta_{123}^{123}(113)\right)-\eta_{12}^{123}(11)\right)} = \left(\frac{\pi_{223}\pi_{113}\pi_{122}\pi_{212}}{\pi_{123}\pi_{213}\pi_{222}\pi_{112}}\right)\left(\frac{\pi_{224}\pi_{114}\pi_{123}\pi_{213}}{\pi_{124}\pi_{214}\pi_{223}\pi_{113}}\right)\left(\frac{\pi_{124}\pi_{214}}{\pi_{224}\pi_{114}}\right) =$$
$$= \frac{\pi_{122}\pi_{212}}{\pi_{222}\pi_{112}}$$

that becomes equal to 1 *when the CS independence holds, thus the* log *of the previuos fraction is equal to* 0.

Until now we consider the CS independence like in formula (7), but when we handle with ordinal variables a more interesting specification of CS independence is

$$A \perp B | C \geq i'_C, \qquad i'_C \in \mathcal{K} \tag{11}$$

or

$$A \perp B | C \leq i'_C, \qquad i'_C \in \mathcal{K} \tag{12}$$

where in this case the class $\mathcal{K}$ is composed of only one cell i'_C and the CS independence must hold for all modalities of variables in C greater(lower) than or equal to the cell in $\mathcal{K}$. Obviously, if the constraints in Theorem 3.3 are satisfied for each $i_C \geq i'_C$ ($i'_C \leq i_C$), then the (11) (or (12)) holds too. But in the case of *local* parameters, there is a easiest way to define the CS independence in formula (11), as shown in Corollary 1.

Corollary 1. *The CS independence in formula (11) holds if and only if the HMM parameters based on local logits satisfy the following constraints:*

$$\eta_{vc}^{\mathcal{M}}(i_{vc}) = 0 \qquad i_{vc} = i_v \cap i_c \qquad i_c \geq i'_c \qquad i'_c \in (\mathcal{K} \cap \mathcal{I}_c) \qquad i_v \in \mathcal{I}_v \tag{13}$$

$\forall v \in \mathcal{V}$ *where* $\mathcal{V} = \{(\mathcal{P}(A) \setminus \emptyset) \cup (\mathcal{P}(B) \setminus \emptyset)\}$, *and* $\forall c \in \mathcal{P}(C)$ *with* $c \neq \emptyset$.

Example 3.3. *From Example 3.2 let us consider a marginal set $\mathcal{M} = (X_1, X_2, X_3)$. The CS independence $X_1 \perp X_2 | X_3 \geq 2$ holds if*

$$\eta_{12}^{123}(11) = 0 \qquad \eta_{123}^{123}(112) = 0 \qquad \eta_{123}^{123}(113) = 0.$$

3.3 Constraints on parameters based on *continuation* logit

As it is shown in Table 1, the parameters based on *continuation* logits involve also sum of probabilities. This make impossible to explicit constraints to define the CS independence as defined in formula (7). However, since this kind of parametrization is adopted when the variables are ordinal, it is helpful also to consider the particular cases displayed in formula (11) and (12). In this section we deal with these questions.

Theorem 3.4. *The CS independence in formula (11) holds if and only if the HMM parameters based on continuation logits satisfy the following constraints:*

$$\eta_{vc}^{\mathcal{M}}(i_{vc}) = 0 \qquad i_{vc} = i_v \cap i_c \qquad i_c \geq i'_c \qquad i'_c \in (\mathcal{K} \cap \mathcal{I}_c) \qquad i_v \in \mathcal{I}_v \tag{14}$$

$\forall v \in \mathcal{V}$, *where* $\mathcal{V} = \{(\mathcal{P}(A) \setminus \emptyset) \cup (\mathcal{P}(B) \setminus \emptyset)\}$, *and* $\forall c \in \mathcal{P}(C)$ *with* $c \neq \emptyset$.

Example 3.4. *Let us consider the situation described in the Example 3.3 but with parameters based on continuation logits. The parameters involved in Theorem 3.4 are* $\eta_{12}^{123}(11)$, $\eta_{123}^{123}(112)$ *and* $\eta_{123}^{123}(113)$*. In particular, the first is*

$$\eta_{12}^{123}(11) = \log\left(\frac{\pi_{114}\pi_{224}}{\pi_{124}\pi_{214}}\right).$$

Note that, $X_1 \perp X_2|X_3 \geq 2$ *implies* $X_1 \perp X_2|X_3 = 4$*. Then the previous parameter is equal to zero.*

About the second parameter, we have:

$$\eta_{123}^{123}(112) = \log\left(\frac{(\pi_{223}+\pi_{224})(\pi_{113}+\pi_{114})(\pi_{122})(\pi_{212})}{(\pi_{123}+\pi_{124})(\pi_{213}+\pi_{214})(\pi_{222})(\pi_{112})}\right).$$

Since the variable X_3 *appears only with modalities* $2, 3$ *and* 4 *for which the CS independence holds, then we get that even this parameter is null. In the same way we progress for the third parameter that is equal to zero.*

Remark 3. *When we are interested in defining a CS independence as expressed in formula (12), we can proceed in an analogous way previously sorting in a descending order the modalities of the interest variable. This corresponds to the reverse continuation coding of the variable.*

Thus, if, for instance, we are interested in checking if a CS independence between two variables holds when the population is young or adult against old, we can sort the modalities of the variable *Age* in the reverse order $\{Old, Adult, Young\}$ and then consider the CS independence in formula (11).

In general, we can decide to codify the variables heterogeneously, with different kinds of logits, in order to suit the nature of the variables. However, as it is shown in this section, the constraints required to define CS independence statements depend on the type of logits used to code the variables in the conditional set. Here we present an example in order to show how to apply the different theorems when we handle with variables coded with different type of logits.

Example 3.5. *Let us consider a marginal set* $\mathcal{M}$ *composed of 4 variables collected in a* $2\times2\times4\times4$ *contingency table* $\mathcal{I}_{\mathcal{M}}$*. We codify the variables with baseline, baseline, local and continuation logits, respectively. We are interested in checking the CS independence* $X_1 \perp X_2|X_3X_4 \geq (2,2)$ *that means that the CS independence must hold when the variables* X_3 *and* X_4 *assume, respectively, the values* $X_3 \geq 2$ *and* $X_4 \geq 2$ *that is the levels* $\{(2,2); (2,3); (3,2); (3,3)\}$*. In this case, noting that the variables in the conditioning set are coded with the local and the continuation logits, the results due to Corollary 1 and Theorem 3.4 imply that the following parameters, involving the* *conditioning variables with values greater or equal to* $(2,2)$*, have to be zero, how effectively is:*

$$\eta_{1234}(1122) = \log\left(\frac{(\pi_{1122})(\pi_{2222})(\pi_{2132})(\pi_{2123}+\pi_{2124})(\pi_{1232})(\pi_{1223}+\pi_{1224})(\pi_{1133}+\pi_{1134})(\pi_{2233}+\pi_{2234})}{(\pi_{2122})(\pi_{1222})(\pi_{1132})(\pi_{1123}+\pi_{1124})(\pi_{2232})(\pi_{2223}+\pi_{2224})(\pi_{2133}+\pi_{2134})(\pi_{1233}+\pi_{1234})}\right) = 0$$

$$\eta_{1234}(1132) = \log\left(\frac{(\pi_{1132})(\pi_{2232})(\pi_{2142})(\pi_{2133}+\pi_{2134})(\pi_{1242})(\pi_{1233}+\pi_{1234})(\pi_{1143}+\pi_{1144})(\pi_{2243}+\pi_{2244})}{(\pi_{2132})(\pi_{1232})(\pi_{1142})(\pi_{1133}+\pi_{1134})(\pi_{2242})(\pi_{2233}+\pi_{2234})(\pi_{2143}+\pi_{2144})(\pi_{1243}+\pi_{1244})}\right) = 0$$

$$\eta_{1234}(1123) = \log\left(\frac{(\pi_{1123})(\pi_{2223})(\pi_{2133})(\pi_{2124})(\pi_{1233})(\pi_{1224})(\pi_{1134})(\pi_{2234})}{(\pi_{2123})(\pi_{1223})(\pi_{1133})(\pi_{1124})(\pi_{2233})(\pi_{2224})(\pi_{2134})(\pi_{1234})}\right) = 0$$

$$\eta_{1234}(1133) = \log\left(\frac{(\pi_{1133})(\pi_{2233})(\pi_{2143})(\pi_{2134})(\pi_{1243})(\pi_{1234})(\pi_{1144})(\pi_{2244})}{(\pi_{2133})(\pi_{1233})(\pi_{1143})(\pi_{1134})(\pi_{2243})(\pi_{2234})(\pi_{2144})(\pi_{1244})}\right) = 0$$

$$\eta_{123}(113) = \log\left(\frac{(\pi_{2134})(\pi_{1234})(\pi_{1144})(\pi_{2244})}{(\pi_{1134})(\pi_{2234})(\pi_{2144})(\pi_{1244})}\right) = 0$$

$$\eta_{124}(112) = \log\left(\frac{(\pi_{2142})(\pi_{1242})(\pi_{1143}+\pi_{1144})(\pi_{2243}+\pi_{2244})}{(\pi_{1142})(\pi_{2242})(\pi_{2143}+\pi_{2144})(\pi_{1243}+\pi_{1244})}\right) = 0$$

$$\eta_{124}(113) = \log\left(\frac{(\pi_{2143})(\pi_{1243})(\pi_{1144})(\pi_{2244})}{(\pi_{1143})(\pi_{2243})(\pi_{2144})(\pi_{1244})}\right) = 0$$

$$\eta_{12}(11) = \log\left(\frac{(\pi_{1144})(\pi_{2244})}{(\pi_{2144})(\pi_{1244})}\right) = 0$$

The same holds for the remaining modalities of X_1X_2*.*

4 Application

In this section we study the relationships among a set of variables by using the HMM model to represent list of independencies, among which CSIs, as presented in Section 3. At first we select a number if marginal set according to the nature of the variables. Several models were tested and in each of them the likelihood ratio test G^2 is carried out. The G^2 compares the model under investigation with the saturated (unconstrained) one; under the null hypothesis the G^2 follows the χ^2 distribution with *df* equal to the difference between the free parameters in the two models. We reject all models with a *p-value* lower than 0.05. Among the non rejected models, we choose the one with greatest Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).
Since to testing all possible models, particularly when we handle with CSIs, is computationally expensive, we implement a three steps procedure to achieve the best model. At first, we carried out an exploratory phase where we test all paired conditional independencies in order to the have an overview of the weakest relationships. Then, we consider as *reduced* model all the paired independencies that have lead to a *p-value* greater than 0.05 in the previous step. Starting from the *reduced* models we discard one by one all paired independencies. We choose the HMM model with greatest AIC and BIC.
A further simplification of the HMM model is obtained evaluating the model with the highest order parameters constrained to zero.
Finally, once obtained the best HMM we move on to further simplification by testing the CSIs by simplifying the conditional ones that have lead to reject the model.

4.1 Innovation Study Survey 2010-2012

In this section we analyzed the dataset on firm's innovations concerning the period starting from 2009 to the 2012. there are 7 variables of interest, concerning different environments of the firm's life. The first kind of variables are the firm's featuring: the *enterprise size*, **DIM** (1= Small, 2= Medium), the *percentage of graduate employers*, **DEG** (1= 0% $\vdash$ 10%, 2= 10% $\vdash$ 50%, 3=50% $\vdash$ 100%) and the *main market (in revenue terms)*, **MRKT** (A= Regional, B= National, C= International), henceforth denoted as variables **1**, **2** and **3**, respectively. Then there are the variables concerning the innovation in some aspect of the enterprise: *innovation in marketing strategies*, **IMAR** (Yes, No), *innovation in organization system*, **IOR** (Yes, No) and *innovation in products or services or production line or investment in R&D*, **IPR** (Yes, No), henceforth denoted as variables **4**, **5** and **6**, respectively. Finally, we consider the *revenue growth* variable in 2012, **GROW** (Yes, No) henceforth denoted as variable **7**.

The survey covers 18697 firms, collected in a $2 \times 2 \times 2 \times 2 \times 3 \times 3 \times 2$ contingency table. Figure 1 report the mosaic plot for each paired variables, where the distribution of frequencies of the marginal contingency table of the two variables are highlighted. The mosaic plot is a graphical representation of categorical and ordinal variables and it is instructive on the multivariate distribution especially when the number of variables involved is low. In the plot, we have a rectangular for each cell of the contingency table which dimension is proportional to the frequency. In situation of independencies the plot is "regular"????, see for instance Kateri (2014).

A study on the relationships between paired variables is lead by evaluating the *equivalent Normal correlation coefficient* given by $\rho(X,Y) = (1 - exp\{-2I(X,Y)\})^{\frac{1}{2}}$ where $I(X,Y)$ is the Kullblack-Leibler information divergence between the observed joint distribution respect to the theoretical situation of independence, see for instance Whittaker (2009). More the real situation is far from the theoretical situation of (marginal) independence, greater is the distance $I(X,Y)$. Similarly, to high values of $I(X,Y)$ corresponds values of $\rho(X,Y)$, near to 1. In figure 2 is represented this index, through a correlation matrix plot while, in table 2, are reported the value. These indexes suggest that the connection between the paired variables, considered marginally respect the other, is not strong.

	DIM	DEG	MRKT	IMAR	IOR	IPR	GROW
DIM	1.00	0.17	0.22	0.07	0.11	0.12	0.10
DEG	0.17	1.00	0.15	0.09	0.12	0.10	0.08
MRKT	0.22	0.15	1.00	0.09	0.08	0.16	0.11
IMAR	0.07	0.09	0.09	1.00	0.30	0.23	0.07
IOR	0.11	0.12	0.08	0.30	1.00	0.30	0.07
IPR	0.12	0.10	0.16	0.23	0.30	1.00	0.07
GROW	0.10	0.08	0.11	0.07	0.07	0.07	1.00

Table 2: The equivalent Normal correlation coefficient based on the Kullback-Leibler information divergence.

On order to build a HMM model to the dataset considered, we have first to decide how to build the marginal sets. Usually, the decision of the marginal sets $\mathcal{M}$ is took in concordance with the aim of the analysis and/or by supporting the nature of the variables. In this case, we have collect three types of variables, so it makes sense to maintain this division and add one type of variables at time. Thus, as first marginal we consider only the firm's features variables $(1,2,3)$, the we add also the innovations variables $(4,5,6)$ and finally we consider also the revenue growth variable (7).The marginal sets that we define thus are $\{(1,2,3);(1,2,3,4,5,6);(1,2,3,4,5,6,7)\}$. The we build the saturated model by defining the HMM parameters by respecting the *com-*

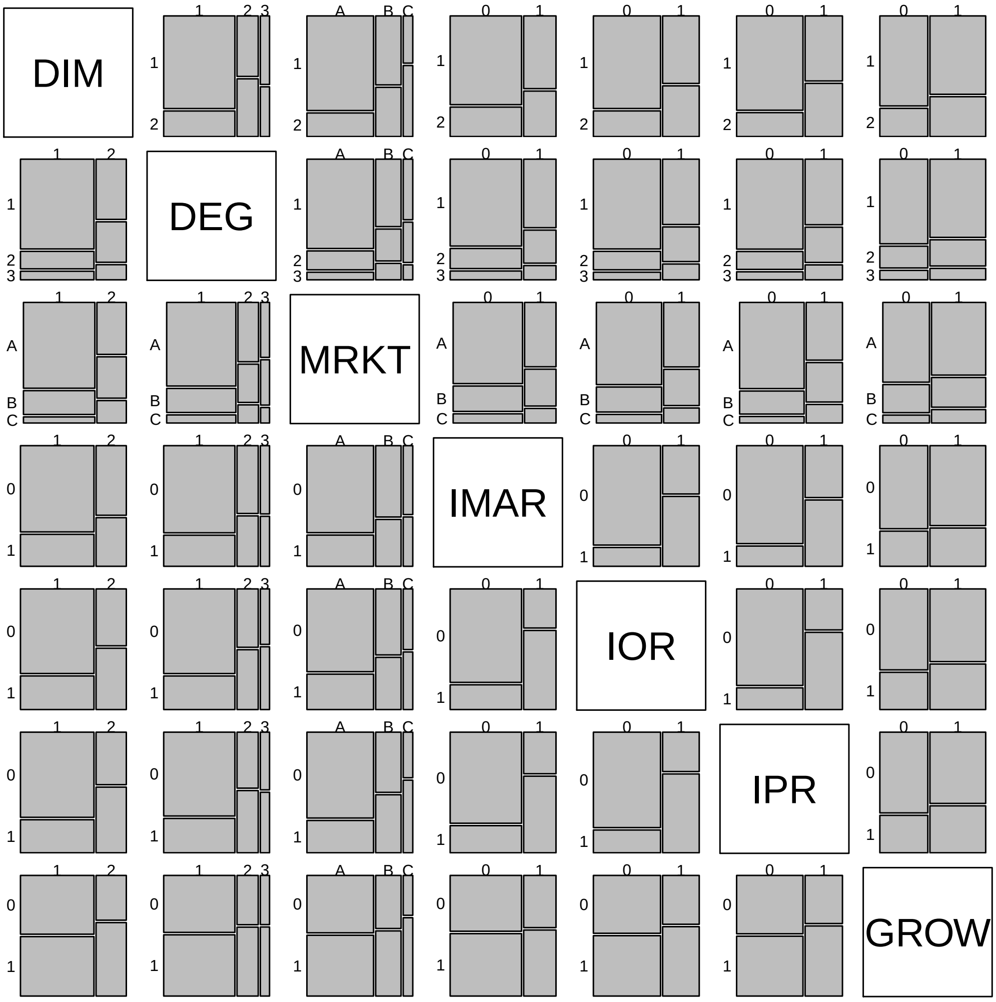

Figure 1: Mosaic plot of paried variables.

pleteness and the *hierarchical* properties, i.e. each parameters in the first possible marginal contingency table. Then we test several independence models by constraining to zero certain parameters (or sum of these). In order to use parameters coherent with the kind of variables, we coded the dichotomous variables (1,4,5,6,7) with *baseline* logits, while, the variables with three modalities (2,3) with the *local* logits.

As first we tested all the HMM models associated with only one conditional independence between two variables. We found four eligible conditional independencies:

(a) $4 \perp 7|12356$ defined in the marginal $\mathcal{M} = (1,2,3,4,5,6,7)$;

(b) $6 \perp 7|12345$ defined in the marginal $\mathcal{M} = (1,2,3,4,5,6,7)$;

(c) $1 \perp 4|2356$ defined in the marginal $\mathcal{M} = (1,2,3,4,5,6)$;

(c) $3 \perp 5|1246$ defined in the marginal $\mathcal{M} = (1,2,3,4,5,6)$.

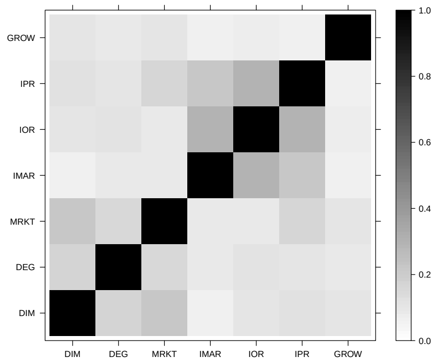

Figure 2: Representation of the equivalent Normal correlation coefficient of the 7 variables throw a correlation matrix plot.

In Table 3 are reported teh HMM models satisfying the combination of these independencies. The only admissible models are the ones with the asterisk near the *p-value* since it is greater than 0.05. Among these we prefer the HMM model characterized only by the **(c)** conditional independence because of its AIC is the lowest among the admissible models.

However, from table 3, there are some clues that the other independencies could be too. Thus, among the **(c)**, we took into account also the independencies **(a)**, **(b)** and **(d)** but under the CSI profile. We test all possible models. A selection of admissible models is reported in table **??**. In order to reduce the CSI to test, it is possible to take advantage from the mosaic plot evaluated in all the conditional distribution, just to have an idea of where the link between the variables in strongest. In Figure are reported the mosaic plot for the CSI selected from table **??**. However, it is necessary to keep in mind that all this "tricks" need to have an idea of the connections between the variables, but do not consider all the independence relationships at the same time as the HMM models do.

The preferred model is described by the conditional independence **(c)** and by the CSIs ...... $1 \perp 2|34567 = (1, *, 3, *, 1)$ that is when there are no innovation in **IORG**, when the innovation **IMAR** assume any modality, when the firm works in an international market, when the percentage of degree employers is whatever and when the firm is small. In correspondence of this model we have `df=121`, `Gsq=141.83`, `p-val=0.09`, `AIC=-192.17`, `BIC=1116.46`.

In tables **??** and **??** are reported the first order (logits) and the second order (contrasts of logit) parameters respectively. A brief consideration on the model is here discussed starting from these parameters.

All the analysis were carried out with the statistical software R R Core Team (2014), with the help of specific packages. In particular the package `hmmm`, Colombi *et al.* (2014) for test the HMM models and the package `vcd`, Meyer *et al.* (2017) for the graphical representation of the categorical variables.

Table 3: HMM models satisfying the combination of conditional independencies (a)$4 \perp 7|123456$, (b) $6 \perp 7|12345$, (c) $1 \perp 4|2356$ and (d) $3 \perp 5|1246$.

	G^2	df	pvalue	AIC
(a)	81.70	72	0.20*	-350.30
(b)	88.55	72	0.09*	-343.45
(c)	49.66	36	0.06*	**-454.34**
(d)	64.12	48	0.06*	-415.88
(a), (b)	139.74	108	0.02	-220.26
(a), (c)	131.35	108	0.06*	**-228.65**
(a), (d)	145.81	120	0.05*	-190.19
(b), (c)	138.20	108	0.03	-221.80
(b), (d)	152.66	120	0.02	-183.34
(c), (d)	105.03	72	0.01	-326.97
(a), (b), (c)	189.39	144	0.01	-98.61
(a), (b), (d)	203.85	156	0.01	-60.15
(a), (c), (d)	186.73	144	0.01	-101.27
(b), (c), (d)	193.58	144	0.00	-94.42
(a), (b), (c), (d)	244.77	180	0.00	28.77

G^2 is the statistic test of the LR test against the saturated model; *df* are the degree of freedom obtained as difference between the total number of parameters and the free (unconstrained) ones; the *p-value* refers to the LR test; AIC is the Akaike Information Criterion.

5 Conclusion

In this work we provide several results in the environment of context-specific independences. At first, we focus on the problem to handle with ordinal variables where it is more useful to use parameters based on the *local* or *continuation* logits compares to the classical ones based on the *baseline* logits. In this case, not only we confirm the results on *baseline* logits such as provided in Nyman *et al.* (2016), even if in the marginal models, but we provide the results in the case of *local* and *continuation* parameters.

The application shows a small part of the potentiality of this work. About the problem of the possibility to represent a list of CSIs through HMM models, a used method is to take advantage from the graphical models which is well know when correspond to HMM models, see Nicolussi *et al.* (2017). In **??** is deepen this topic.

Appendices

A Matrix C

The contrasts matrix C in formula 6 selects the probabilities from the vector π to build the vector of parameters. It is a block diagonal matrix with blocks C_J, with $J \in \mathcal{P}(\mathcal{Q})$ and $C_J = \otimes_{j\in\mathcal{M}} C_{J,j}$. The development of $C_{J,j}$ depends on the kind of aggregation criterion we chose. Let us define the identity matrix of order r with the symbol Id_r, the column vector of 1s with r elements with the symbol $\mathbf{1}_r$, with $\mathbf{0}'_r$ the row vector of 0s of dimension $(1 \times r)$ and with $1^U_{r\times r}$ the upper squared triangular matrix of 1s, with 0s on the diagonal.

In the case of parameters based on *baseline* logit the matrix assumes the following form (see

	Gsq	df	pval	AIC	BIC
1, 1, 1, 1	133.15	110.00	0.07	-222.85	1171.98
1, 1, 1, 2	134.38	110.00	0.06	-221.62	1173.21
1, 1, 2, 1	138.84	110.00	0.03	-217.16	1177.67
1, 1, 2, 2	136.31	110.00	0.05	-219.69	1175.14
1, 1, 3, 1	137.55	110.00	0.04	-218.45	1176.38
1, 1, 3, 2	143.65	110.00	0.02	-212.35	1182.48
1, 2, 1, 1	131.60	110.00	0.08	-224.40	1170.43
1, 2, 1, 2	133.37	110.00	0.06	-222.63	1172.19
1, 2, 2, 1	131.77	110.00	0.08	-224.23	1170.60
1, 2, 2, 2	131.53	110.00	0.08	-224.47	1170.36
1, 2, 3, 1	131.39	110.00	0.08	-224.61	1170.22
1, 2, 3, 2	133.70	110.00	0.06	-222.30	1172.53
2, 1, 1, 1	132.02	110.00	0.07	-223.98	1170.85
2, 1, 1, 2	133.14	110.00	0.07	-222.86	1171.97
2, 1, 2, 1	134.13	110.00	0.06	-221.87	1172.96
2, 1, 2, 2	131.44	110.00	0.08	-224.56	1170.27
2, 1, 3, 1	133.31	110.00	0.06	-222.69	1172.14
2, 1, 3, 2	132.50	110.00	0.07	-223.50	1171.33
2, 2, 1, 1	131.49	110.00	0.08	-224.51	1170.32
2, 2, 1, 2	140.86	110.00	0.03	-215.14	1179.69
2, 2, 2, 1	132.73	110.00	0.07	-223.27	1171.56
2, 2, 2, 2	138.39	110.00	0.03	-217.61	1177.22
2, 2, 3, 1	131.88	110.00	0.08	-224.12	1170.71
2, 2, 3, 2	135.25	110.00	0.05	-220.75	1174.08

Table 4: Equality constraints

Bartolucci *et al.* (2007)):

$$C_{J,j} = \begin{cases} \left(-\mathbf{1}_{I_j-1}, Id_{I_j-1}\right) & \text{if } j \in J \\ \left(1, \mathbf{0}'_{I_j-1}\right) & \text{oterhwise .} \end{cases} \tag{15}$$

When the parameters are based on *local* logits we have (see Forcina *et al.* (2010))::

$$C_{J,j} = \begin{cases} \left(\mathbf{0}_{I_j-1}, Id_{I_j-1}\right) - \left(Id_{I_j-1}, \mathbf{0}_{I_j-1}\right) & \text{if } j \in J \\ \left(1, \mathbf{0}'_{I_j-1}\right) & \text{oterhwise .} \end{cases} \tag{16}$$

Finally, for *continuation* logits we have:

$$C_{J,j} = \begin{cases} \left(\mathbf{0}_{I_j-1}, Id_{I_j-1}\right) - \left(1^T_{(I_j-1)\times(I_j-1)}, \mathbf{0}_{I_j-1}\right) & \text{if } j \in J \\ \left(1, \mathbf{0}'_{I_j-1}\right) & \text{oterhwise .} \end{cases} \tag{17}$$

	Gsq	df	pval	AIC	BIC
1, 1, 1, 1	133.15	110	0.07	-222.854	1171.975
1, 1, 1, 2	135.16	112	0.07	-216.836	1162.321
1, 1, 2, 1	140.63	112	0.03	-211.368	1167.789
1, 1, 2, 2	144.38	116	0.04	-199.619	1148.193
1, 1, 3, 1	146.83	114	0.02	-201.170	1162.314
1, 1, 3, 2	158.77	120	0.01	-177.229	1139.239
1, 2, 1, 1	133.30	112	0.08	-218.703	1160.454
1, 2, 1, 2	136.58	114	0.07	-211.418	1152.067
1, 2, 2, 1	141.74	116	0.05	-202.256	1145.556
1, 2, 2, 2	146.14	120	0.05	-189.859	1126.609
1, 2, 3, 1	150.09	120	0.03	-185.915	1130.553
1, 2, 3, 2	161.34	126	0.02	-162.657	1106.794
2, 1, 1, 1	133.82	112	0.08	-218.182	1160.975
2, 1, 1, 2	140.15	116	0.06	-203.852	1143.960
2, 1, 2, 1	144.08	116	0.04	-199.922	1147.890
2, 1, 2, 2	153.90	124	0.04	-174.097	1111.026
2, 1, 3, 1	152.23	120	0.02	-183.771	1132.697
2, 1, 3, 2	170.73	132	0.01	-141.275	1081.160
2, 2, 1, 1	134.33	116	0.12	-209.674	1138.139
2, 2, 1, 2	146.91	120	0.05	-189.088	1127.380
2, 2, 2, 1	147.19	124	0.08	-180.806	1104.318
2, 2, 2, 2	165.64	132	0.03	-146.362	1076.072
2, 2, 3, 1	160.24	132	0.05	-151.763	1070.671
2, 2, 3, 2	186.73	144	0.01	-101.269	1027.132

Table 5: Disequality constraints: $\geq$ *modality*

B Proofs and further results

Lemma 1. *Given a HMM parameter $\eta^{\mathcal{M}}_{\mathcal{L}C}(i_{\mathcal{L}C})$, where the set $\mathcal{L}C$ is the union of two sets of variables belonging in $\mathcal{M}$, it can be decomposed as follow*

$$\eta^{\mathcal{M}}_{\mathcal{L}C}(i_{\mathcal{L}C}) = \sum_{\substack{J\subseteq C\\ J\neq\emptyset}} (-1)^{|J|+1}\eta^{\mathcal{M}}_{\mathcal{L}(C\setminus J)}(i_{\mathcal{L}(C\setminus J)}|i^*_J) + (-1)^{|C|}\eta^{\mathcal{M}}_{\mathcal{L}}(i_{\mathcal{L}}|i_C) \quad (18)$$

Independencies	Gsq	df	pval	AIC	BIC
(a), **(b)**	139.74	108	0.02	-220.26	1190.24
(a), **(c)**	168.57	120	0.00	-167.42	1149.04
(b), **(c)**	141.34	120	**0.09**	-194.66	1121.81
(a), **(b)** ,**(c)**	180.97	132	0.00	-131.03	1091.40

Table 6: HMMM which combining the three independencies **(a)** $1 \perp 2|34567$, **(b)**$1 \perp 4|23567$ and **(c)** $1 \perp 6|23457$.

Proof of Lemma 1 From Proposition (1) of Bartolucci *et al.* (2007) any parameter $\eta^{\mathcal{M}}_{\mathcal{L}C}(i_{\mathcal{L}C})$ can be rewritten as

$$\eta^{\mathcal{M}}_{\mathcal{L}C}(i_{\mathcal{L}C}) = \sum_{J\subseteq C}(-1)^{|C\backslash J|}\eta^{\mathcal{M}}_{\mathcal{L}}(i_{\mathcal{L}}|i_{C\backslash J}i^*_J) \tag{19}$$

where $\eta^{\mathcal{M}}_{\mathcal{L}}(i_{\mathcal{L}}|i_{C\backslash J}i^*_J)$ is the HMM parameter $\eta^{\mathcal{M}}_{\mathcal{L}}$ evaluated in the conditional distribution where the variables in $C\backslash J$ assume values $i_{C\backslash J}$ and the variables in J are set to a reference modality i^*_J.

When the set C is only one variable, $C = \gamma_1$, the decomposition in formula (19) becomes

$$\eta^{\mathcal{M}}_{\mathcal{L}C}(i_{\mathcal{L}C}) = \eta^{\mathcal{M}}_{\mathcal{L}}(i_{\mathcal{L}}|i^*_{\gamma_1}) - \eta^{\mathcal{M}}_{\mathcal{L}}(i_{\mathcal{L}}|i_{\gamma_1}) \tag{20}$$

that corresponds to formula (18).

When two variables belong to the set C, $C = \{\gamma_1, \gamma_2\}$, by applying the formula (19) only to γ_1 we get

$$\eta^{\mathcal{M}}_{\mathcal{L}C}(i_{\mathcal{L}C}) = \eta^{\mathcal{M}}_{\mathcal{L}\gamma_2}(i_{\mathcal{L}\gamma_2}|i^*_{\gamma_1}) - \eta^{\mathcal{M}}_{\mathcal{L}\gamma_2}(i_{\mathcal{L}\gamma_2}|i_{\gamma_1}); \tag{21}$$

the second addend on the right hand side, can be further decomposed by using the (19) as:

$$\eta^{\mathcal{M}}_{\mathcal{L}\gamma_2}(i_{\mathcal{L}\gamma_2}|i_{\gamma_1}) = \eta^{\mathcal{M}}_{\mathcal{L}}(i_{\mathcal{L}}|i_{\gamma_1}i^*_{\gamma_2}) - \eta^{\mathcal{M}}_{\mathcal{L}}(i_{\mathcal{L}}|i_{\gamma_1\gamma_2}). \tag{22}$$

Now, by considering the HMM parameter $\eta^{\mathcal{M}}_{\mathcal{L}\gamma_1}(i_{\mathcal{L}\gamma_1}|i^*_{\gamma_2})$ and by applying the formula (19), we get

$$\eta^{\mathcal{M}}_{\mathcal{L}\gamma_1}(i_{\mathcal{L}\gamma_1}|i^*_{\gamma_2}) = \eta^{\mathcal{M}}_{\mathcal{L}}(i_{\mathcal{L}}|i^*_{\gamma_1\gamma_2}) - \eta^{\mathcal{M}}_{\mathcal{L}}(i_{\mathcal{L}}|i_{\gamma_1}i^*_{\gamma_2}). \tag{23}$$

Note that the last addend on the right hand side of the (23) is exactly the first addend on the right hand side of (22). Thus, by replacing the (22) and (23) in the (21) we get:

$$\eta^{\mathcal{M}}_{\mathcal{L}C}(i_{\mathcal{L}C}) = \eta^{\mathcal{M}}_{\mathcal{L}\gamma_2}(i_{\mathcal{L}\gamma_2}|i^*_{\gamma_1}) - \eta^{\mathcal{M}}_{\mathcal{L}}(i_{\mathcal{L}}|i^*_{\gamma_1\gamma_2}) + \eta^{\mathcal{M}}_{\mathcal{L}\gamma_1}(i_{\mathcal{L}\gamma_1}|i^*_{\gamma_2}) + \eta^{\mathcal{M}}_{\mathcal{L}}(i_{\mathcal{L}}|i_{\gamma_1\gamma_2}) \tag{24}$$

that again corresponds to formula (18).

In general, when the set C is composed of k variables, $C = \{\gamma_1, \dots, \gamma_k\}$, we apply formula (19) recursively, focusing on only one variable of C each time, to any parameter in the formula without any index i^* in the conditioning set.

$$\eta^{\mathcal{M}}_{\mathcal{L}C}(i_{\mathcal{L}C}) = \sum_{j=1}^{k}(-1)^{j+1}\eta^{\mathcal{M}}_{\mathcal{L}C\backslash(\gamma_{jp}\gamma_j)}(i_{\mathcal{L}C\backslash(\gamma_{jp}\gamma_j)}|i^*_{\gamma_j}i_{\gamma_{jp}}) + (-1)^{|C|}\eta^{\mathcal{M}}_{\mathcal{L}}(i_{\mathcal{L}}|i_C). \tag{25}$$

where $\gamma_{jp} = \sum_{i=1}^{j-1}\gamma_i$.

Now, we take into account all the parameters having both i and i^* in the conditioning set. Let us denote it as $\eta^{\mathcal{M}}_{\mathcal{L}}(i_{\mathcal{L}}|i_A i^*_B)$. We can recognise this term in the last term of the right hand side of the decomposition 26 obtained applying the 19 to $\eta^{\mathcal{M}}_{\mathcal{L}A}(i_{\mathcal{L}A}|i^*_B)$:

$$\eta^{\mathcal{M}}_{\mathcal{L}A}(i_{\mathcal{L}A}|i^*_B) = \sum_{\substack{J\subseteq A\\ J\neq\emptyset}}(-1)^{|A\backslash J|}\eta^{\mathcal{M}}_{\mathcal{L}}(i_{\mathcal{L}}|i^*_{BJ}i_{A\backslash J}) + \eta^{\mathcal{M}}_{\mathcal{L}}(i_{\mathcal{L}}|i^*_B i_A) \tag{26}$$

By replacing in formula (25) each addend like $\eta^{\mathcal{M}}_{\mathcal{L}}(i_{\mathcal{L}}|i_A i^*_B)$ with the expression learned from formula (26), and applying this procedure recursively to any addend like $\eta^{\mathcal{M}}_{\mathcal{L}}(i_{\mathcal{L}}|i_A i^*_B)$, we finally obtain exactly the formula 18.

Corollary 2. *A parameter $\eta^{\mathcal{M}}_{\mathcal{L}}$ can be decomposed as the sum of greater order parameters as follows:*

$$\eta^{\mathcal{M}}_{\mathcal{L}}(i_{\mathcal{L}}|i_C) = \sum_{J\subseteq C}(-1)^{|C\backslash J|}\eta^{\mathcal{M}}_{\mathcal{L}J}(i_{\mathcal{L}J}|i^*_{C\backslash J}) \tag{27}$$

Proof of Corollary 2 From formula (18), we isolate the last right term having

$$\begin{aligned}(-1)^{|C|}\eta^{\mathcal{M}}_{\mathcal{L}}(i_{\mathcal{L}}|i_C) &= \textstyle\sum_{J\subseteq C}(-1)^{|J|}\eta^{\mathcal{M}}_{\mathcal{L}(C\setminus J)}(i_{\mathcal{L}(C\setminus J)}|i^*_J)\\ \eta^{\mathcal{M}}_{\mathcal{L}}(i_{\mathcal{L}}|i_C) &= \textstyle\sum_{J\subseteq C}(-1)^{|C\setminus J|}\eta^{\mathcal{M}}_{\mathcal{L}(C\setminus J)}(i_{\mathcal{L}(C\setminus J)}|i^*_J)\end{aligned} \tag{28}$$

By replacing $C\backslash J$ with J in the left side, we get exactly the decomposition in formula (27).

Proof of Theroem 3.1 When the CS independence in formula (7) holds, let us consider the parameters $\eta^{\mathcal{M}}_{\mathcal{L}}$ when $\mathcal{L} = A \cup B \cup C \subseteq \mathcal{M}$. From Lemma 1 we can decompose it as

$$\eta^{\mathcal{M}}_{ABC}(i_{ABC}) = \sum_{\substack{J\subseteq C\\ J\neq\emptyset}} (-1)^{|J+1|}\eta^{\mathcal{M}}_{AB(C\setminus J)}(i_{AB(C\setminus J)}|i^*_J) + (-1)^{|C|}\eta^{\mathcal{M}}_{AB}(i_{AB}|i_C) \tag{29}$$

where $\eta^{ABC}_{AB}(i_{AB}|i_C)$ is the marginal parameter η^{ABC}_{AB} evaluated in the conditional distribution $(AB|C = i_C)$. The term $\eta^{ABC}_{AB}(i_{AB}|i_C)$ is equal to zero if and only if the CS independence in formula (7) holds. Thus,

$$\begin{aligned}\eta^{ABC}_{ABC}(i_{ABC}) - \textstyle\sum_{\substack{J\subseteq C\\ J\neq\emptyset}}(-1)^{|J+1|}\eta^{ABC}_{AB(C\setminus J)}(i_{AB(C\setminus J)}|i^*_J) &= 0\\ \eta^{ABC}_{ABC}(i_{ABC}) + \textstyle\sum_{\substack{J\subseteq C\\ J\neq\emptyset}}(-1)^{|J|}\eta^{ABC}_{AB(C\setminus J)}(i_{AB(C\setminus J)}|i^*_J) &= 0\\ \textstyle\sum_{J\subseteq C}(-1)^{|J|}\eta^{ABC}_{AB(C\setminus J)}(i_{AB(C\setminus J)}|i^*_J) &= 0\\ \textstyle\sum_{c\in\mathcal{P}(C)}(-1)^{|C\setminus c|}\eta^{ABC}_{ABc}(i_{ABc}|i^*_{C\setminus c}) &= 0\end{aligned} \tag{30}$$

Note that in the case of *baseline* coding, the cell $i^*_{C\setminus c}$ is equivalent to $I_{C\setminus c}$ thus the parameters $\eta^{ABC}_{ABc}(i_{ABc}|i^*_{C\setminus c})$ is $\eta^{ABC}_{ABc}(i_{ABc})$.
Finally, by considering that the previous decomposition holds for each set $v \in \mathcal{V} = \{(\mathcal{P}(A) \setminus \emptyset) \cup (\mathcal{P}(B) \setminus \emptyset)\}$, the formula (8) comes.

Proof of Theorem 3.2 The proof follows by theorem 1, Rudas *et al.* (2010).

Proof of Theorem 3.3 From the proof of Theorem 3.1, the decomposition in formula (30) still holds. However, by using the *local* logits $i^*_{C\setminus c} \neq I_{C\setminus c}$ and the identity $\eta^{ABC}_{ABc}(i_{ABc}|i^*_{C\setminus c}) = \eta^{ABC}_{ABc}(i_{ABc})$ does not hold any more because in *local* logits $i^*_{C\setminus c}$ is equal to $\cap_{j\in C\setminus c}(i_j+1)$ while the parameter $\eta^{ABC}_{AB}(i_{AB})$ is built in the conditional distribution where the variables in C assume the reference value I_C. Note that $\eta^{ABC}_{ABc}(i_{ABc}|i_{C\setminus c}+1)$ does not belong to this parametrization. Now we remark that between the *baseline* parameters, η_b, and the *local* parameters η_l, the following relationship holds:

$$\eta^{\mathcal{M}}_{b\mathcal{L}}(i_{\mathcal{L}}) = \sum_{i'_{\mathcal{L}}\geq i_{\mathcal{L}}} \eta^{\mathcal{M}}_{l\mathcal{L}}(i'_{\mathcal{L}}). \tag{31}$$

When the variables in the conditioning set C are based on *local* logits, it is enough to apply the decomposition in (31) only on the variables in the parameter $\eta^{ABC}_{ABC}(i_{ABC})$ in order to have a *baseline* approach in C. Thus we can rewrite (30) as:

$$\sum_{c\in\mathcal{P}(C)} (-1)^{|C\setminus c|} \sum_{i'_c\geq i_c} \eta^{ABC}_{ABc}(i_{AB}i'_c|I_{C\setminus c}) = 0 \tag{32}$$

where η_{ABC}^{ABC} are the *local* parameters and are exactly the same of formula (10). As in proof of the Theorem 3.1, the previous equivalence must hold for each subset v of $A \cup B$ with at least one element in A and one element in B.

Proof of Corollary 1 When i_C in $\mathcal{K}$ is equal to the last modalities I_C, for each $c \subseteq C$ and $c \neq \emptyset$, the parameters $\eta_{vC}^{\mathcal{M}}(i_{vC}) = 0$ by definition, thus formula (10) in Theorem 3.3 becomes $\eta_v^{\mathcal{M}} = 0\ \forall v \in \mathcal{V}$. When $\mathcal{K} = \{(I_{C\setminus j}) \cap (I_j - 1)\}$, that is the modality of every variable is equal to the last but the variable j assumes the modality $I_j - 1$, the constraints become $\eta_v^{\mathcal{M}}(i_v) = 0$ and $\eta_{vj}^{\mathcal{M}}(i_{vj}) = 0$. Applying this procedure recursively for each i'_C we obtain the constraints in formula (13).

Proof of Theorem 3.4 For a $c \in \mathcal{P}(C)$ and a $v \in \mathcal{V}$ we consider the parameters $\eta_{vc}^{\mathcal{M}}(i_{vc})$. Note that, each variable X_j in C assumes value in i_j or in $((i_j + 1) + \cdots + I_j)$ when it drop in the reference modality. But since in each of these distributions the CS independence (11) holds the parameters are equal to zero.

References

Agresti, Alan. 2013. *Categorical data analysis*. John Wiley & Sons.

Bartolucci, Francesco, Colombi, Roberto, & Forcina, Antonio. 2007. An extended class of marginal link functions for modelling contingency tables by equality and inequality constraints. *Statistica Sinica*, **17**, 691–711.

Bergsma, Wicher P, & Rudas, Tamas. 2002. Marginal models for categorical data. *Annals of Statistics*, 140–159.

Cazzaro, Manuela, & Colombi, Roberto. 2014. Marginal nested interactions for contingency tables. *Communications in Statistics-Theory and Methods*, **43**(13), 2799–2814.

Colombi, Roberto, & Forcina, Antonio. 2014. A class of smooth models satisfying marginal and context specific conditional independencies. *Journal of Multivariate Analysis*, **126**, 75–85.

Colombi, Roberto, Giordano, Sabrina, & Cazzaro, Manuela. 2014. hmmm: An R Package for Hierarchical Multinomial Marginal Models. *Journal of Statistical Software*, **59**(11), 1–25.

Drton, Mathias, *et al.* 2009. Likelihood ratio tests and singularities. *The Annals of Statistics*, **37**(2), 979–1012.

Forcina, Antonio. 2012. Smoothness of conditional independence models for discrete data. *Journal of Multivariate Analysis*, **106**, 49–56.

Forcina, Antonio, Lupparelli, Monia, & Marchetti, Giovanni M. 2010. Marginal parameterizations of discrete models defined by a set of conditional independencies. *Journal of Multivariate Analysis*, **101**(10), 2519–2527.

Glonek, Garique FV, & McCullagh, Peter. 1995. Multivariate logistic models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 533–546.

Højsgaard, Søren. 2004. Statistical inference in context specific interaction models for contingency tables. *Scandinavian Journal of Statistics*, **31**(1), 143–158.

Kateri, Maria. 2014. *Contingency table analysis.* Springer.

Meyer, David, Zeileis, Achim, & Hornik, Kurt. 2017. vcd: Visualizing Categorical Data. *R package version 1.4-4.*

Nicolussi, Federica, Colombi, Roberto, *et al.* 2017. Type II chain graph models for categorical data: A smooth subclass. *Bernoulli*, **23**(2), 863–883.

Nyman, Henrik, Pensar, Johan, Koski, Timo, Corander, Jukka, *et al.* 2014. Stratified graphical models-context-specific independence in graphical models. *Bayesian analysis*, **9**(4), 883–908.

Nyman, Henrik, Pensar, Johan, Koski, Timo, & Corander, Jukka. 2016. Context-specific independence in graphical log-linear models. *Computational Statistics*, **31**(4), 1493–1512.

R Core Team. 2014. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Rudas, Tamás, Bergsma, Wicher P, & Németh, Renáta. 2010. Marginal log-linear parameterization of conditional independence models. *Biometrika*, **97**(4), 1006–1012.

Whittaker, Joe. 2009. *Graphical models in applied multivariate statistics.* Wiley Publishing.