Context-Specific independencies embedded in Chain Graph Models of type I

Federica Nicolussi, Manuela Cazzaro

Abstract For a set of variables collected in a contingency table, we focus on a particular kind of relationships such as the context-specific independencies. These are conditional independencies that hold for particular values of the conditioning set. Given the advantages of the graphical models, we use them to represent different relationships among the variables, including the context-specific independencies. In particular, we enrich Chain Graph models with labelled arcs. Furthermore, we consider the well-known relationships between Chain Graph models and Hierarchical Multinomial Marginal models and we introduce new constraints on parameters in order to describe the context-specific relationship. Finally, we provide an application to the study of innovation in Italy by comparing two different periods.

Keywords: context-specific independencies; categorical variables; ordinal variables; stratified chain graph models.

1 Introduction

A Context-Specific independence (CSI) is a particular relationship that focuses on certain value(s) of conditioning variables. Indeed, it is not rare to observe phenomena that are independent under particular conditions, but, under other circumstances, they have on the contrary a strong connection. In this case, stating that there is conditional independence between the two phenomena is not true, but not considering the lack of "partial" connection could be inaccurate. In this work we consider a set of categorical variables and we study different kind of relationships, among which the CSIs, that lie between them. In the literature, marginal and conditional indepen-

Federica Nicolussi

University of Milano-Bicocca, Via Bicocca degli Arcimboldi 8 e-mail: federica.nicolussi@unimib.it

Manuela Cazzaro

University of Milano-Bicocca, Via Bicocca degli Arcimboldi 8 e-mail: manuela.cazzaro@unimib.it

dencies get more attention and are deeply studied. For instance, given two variables, say X_1 and X_2, it is usual to investigate if they are marginally independent ($X_1 \perp X_2$) or conditionally independent given a third variable X_3 ($X_1 \perp X_2|X_3$). The CSI statement establishes that the variables X_1 and X_2 are independent given $X_3 = i_3$ while the same statement does not hold when $X_3 \neq i_3$; see among others Boutilier [2]. Indeed, these CSIs were mainly examined to study problems concerning latent variables; see for instance [13].

Our aim is to incorporate the CSI conditions in graphical models that are suitable to represent different kind of relationships. Nyman et al., [11, 12] analyse CSIs in graphical models based on undirected graphs, or on directed acyclic graphs, using the classical log-linear parametrization. In both papers they adapt these kind of graphs with labelled arcs in order to take into account the CSIs. In this work we follow the same approach and we enrich chain graphs with labelled arcs in order to display also the CSIs. Furthermore, we take advantage of Hierarchical Multinomial Marginal (HMM) parametrization [1, 4], as a generalization of the log-linear models, to represent the dependence relationships. A further advantage to consider the CSIs lies also in the possibility of reducing the number of HMM parameters.

The paper has the following structure. In Section 2.1 we give an overview of HMM models by considering also the case when we deal with ordinal variables. About this, we propose the constraints on HMM models able to satisfy the CSIs. The representation through (Stratified) Chain Graph models is debated in Section 2.2. In Section 3 an application on the study of the trend of innovation degree on Italian enterprises is provided. Finally, Section 4 is dedicated to a conclusion.

2 Methodology

Let us consider q categorical (ordinal) variables $V = \{X_1, \ldots, X_q\}$ taking values in the contingency table $\mathscr{I} = (\mathscr{I}_1 \times \ldots \times \mathscr{I}_q)$, where $\mathscr{I}_j = \{1, \ldots, n_j\}$, with $j = 1, \ldots, q$, such that $i_j \in \mathscr{I}_j$ is the generic value of the variable X_j. Note that $(i_1, \ldots, i_q)$ identifies a particular cell of the contingency table $\mathscr{I}$ and henceforth we refer to it with the shortcut $(i_{1 \ldots q})$. In the following subsection we describe the methodology able to define a system of independencies (marginal, conditional and context specific) that reveals the relationships among all the variables involved in the contingency table.

2.1 Hierarchical Multinomial Marginal models for context-specific independencies

The HMM model is a generalization of the classical log-linear model which allows to represent conditional and marginal independencies in the same model. Instead of considering only the joint distribution, this model takes into account also marginal

distributions and, on these, define the log-linear parameters by respecting certain properties of completeness and hierarchy. These new parameters are contrasts (of sum) of logarithms of probabilities and henceforth we refer to them as HMM parameters.

For instance, let us suppose to consider three variables, X_1, X_2 and X_3. We are interested in describing that variables X_1 and X_2 are independent given by X_3, jointly considered, $X_1 \perp X_2 | X_3$, and that X_2 is marginally independent of X_3, $X_2 \perp X_3$. To this aim, we consider the marginal distribution of $\{X_2, X_3\}$ and the joint distribution. We refer to these by defining the class of marginal distributions $\{\{2,3\}; \{1,2,3\}\}$ where $\{2,3\}$ and $\{1,2,3\}$ are a shortcut for $\{X_2, X_3\}$ and $\{X_1, X_2, X_3\}$. Then we define the classical log-linear parameters on the marginal contingency table $\mathscr{I}_{23}$ concerning the variables $\{2,3\}$ and the remaining parameters on the contingency table $\mathscr{I}$. Let us define the HMM parameters with the caption $\eta_{\mathscr{L}}^{\mathscr{M}}(i_{\mathscr{L}})$ where $\mathscr{M}$ refers to the marginal distribution, $\mathscr{L}$ denotes the subset of variables to which the parameter pertains and $i_{\mathscr{L}}$, in parenthesis, represents the values of the variable selected in $\mathscr{L}$ (when the parenthesis are omitted, it means that the parameters refer to each $i_{\mathscr{L}} \in \mathscr{I}_{\mathscr{L}}$). Finally, in order to test the marginal and conditional independencies, we have to constrain to zero the parameters $\eta_{2,3}^{\{2,3\}}, \eta_{1,2}^{\{1,2,3\}}$ and $\eta_{1,2,3}^{\{1,2,3\}}$.

Let us consider the following statement of CSI where the conditional independence holds only in a subset of variables. For instance

$$\begin{cases} X_1 \perp X_2 | X_3 = i_3, \ i_3 \in \mathscr{K} \\ X_1 \not\perp X_2 | X_3 = i_3, \ i_3 \notin \mathscr{K} \end{cases} \tag{1}$$

where $\mathscr{K} \subseteq \mathscr{I}_3$ is a subset of the values i_3 of X_3 for which the conditional independence holds.

One main goal of this work is the definition of the constraints on HMM parameters in order to satisfy the CSI in formula (1). In [11], Nyman et al. deal with the log-linear parameters defined on the joint distribution. Here, as first improvement, we take into account the HMM parameters defined also on marginal distributions, see [10]. Thus, as before, we proceed to define the class of marginal distributions (the same mentioned above) and to specify the parameters evaluated on suitable marginal distributions. The constraints satisfying the CSI in formula (1) are

$$\eta_{1,2}^{\{1,2,3\}}(i_{12}) + \eta_{1,2,3}^{\{1,2,3\}}(i_{12}, i_3) = 0 \qquad i_{12} \in \mathscr{I}_{12} \quad i_3 \in \mathscr{K}, \tag{2}$$

where $\mathscr{I}_{12}$ is the marginal contingency table concerning the variables $\{1,2\}$. Another important aspect of this work is to consider the possible presence of ordinal variables. The classical log-linear models, in fact, look poor when we want to focus on the interpretation of the effects among the variables, in particular, when we take into account ordinal variables; see for instance [3]. For this reason we choose different criteria for coding the variables through the parameters. In fact, beyond the classical *baseline* criterion, we take advantage of the *local* criterion that is more suitable for ordinal variables. By adopting the *local* criterion for coding the conditioning variable, as it is shown in [10], the constraints in formula (2) become:

$$\eta_{1,2}^{\{1,2,3\}}(i_{12}) + \sum_{i_3^*=1}^{i_3} \eta_{1,2,3}^{\{1,2,3\}}(i_{12}, i_3^*) = 0 \qquad i_1 \in \mathscr{I}_{12} \quad i_3 \in \mathscr{K}. \qquad (3)$$

It is worthwhile to note that, when we deal with local parameters, if the CSI is presented in the following different statement: $X_1 \perp X_2 | X_3 \geq i_3$, $i_3 \in \mathscr{K}$, the constraints in formula (3) are equal to the ones in formula (2). More details are given in [10].

2.2 Stratified Chain Graph Models

A *Chain Graph* (CG) is a graph with both directed and undirected arcs and without any directed or semi-directed cycle. The vertices of a CG can be grouped in so-called *Chain Components*, denoted by $T_1, ..., T_s$, that are the connected undirected components. Intuitively, Chain Graph Models (CGM) are graphical models which take advantage of chain graphs; see [5]. The structure of relationships among variables which follow an inherent order is well represented from these models. In particular, we can distinguish variables linked by symmetric relationships and variables linked by unilateral dependence. In this case we follow this order for collecting them in chain components.

In the literature, the representation of independencies through CGs is not unique, a deep dissertation is discussed in [5]; in this work we adopt the point of view of Lauritzen and Wermuth, [7], also known as chain graph models of type I, that is a subclass of the HMM models; see [9] and [14]. These CGMs are the natural extension of the graphical models based on undirected graph and directed acyclic graph. They interpret the lack of (un)directed arcs conditionally with respect to the remaining variables in the same component. In addition, all the systems of independencies representable through these graphical models benefit from the existence of a smooth likelihood function. In order to take into account the CSIs, we propose Stratified Chain Graph Models (SCGMs) as extension of Stratified Graphical Models (SGMs) introduced by Nyman et al., [11]. Similarly to SGM, we denote the CSIs through labelled arcs. Figure 1 depicts an example of a SCGM. In this case the lack of the directed arcs between the nodes X_1 and X_5, X_2 and X_5 and finally between X_2 and X_3 represents the conditional independencies $X_1 X_2 \perp X_5 | X_3 X_4$ and $X_2 \perp X_3 | X_1 X_4 X_5$. Then the labelled arc between the nodes X_3 and X_4 represents the CSI $X_3 \perp X_4 | X_1 X_2 X_5 = (i_1, *, i_5)$ where the asterisk is a symbol for referring to all the values of the variable X_2 in this case.

3 Application

In the next subsection we implement the presented model with an application to a real dataset. At first, we select the variables and we define the marginal distributions to take into account, according to the focus of the analysis. In order to find the best

fitting model, we proceed with a three steps algorithm where each model is tested by using the likelihood ratio test G^2. The algorithm is explained below.

Step 1: We test the CGMs associated to all possible CGs obtained by deleting only one arc (at time) from the complete graph. Among these models, we select the ones with a *p-value* of the likelihood ratio test greater than 0.01.

Step 2: Similarly to Step 1, we test the SCGMs associated to all possible SCGs obtained by replacing only one arc (at time) with a labelled arc with all possible labels considered one at time. Among these models, we select the ones with a *p-value* of the likelihood ratio test greater than 0.1.

Step 3: From all admissible models selected in the previous two steps, we test all possible combinations of marginal, conditional independencies and CSIs and we maintain the one with lower AIC (Akaike Information Criterion) between the models with a *p-value* higher than 0.05.

3.1 The Italian Innovation Survey

We analyse two datasets, concerning the Italian Innovation Survey, pertinent each to a three years period: the first 2008-2010 and the second 2010-2012; [6]. The two datasets involve 16531 and 18697 small and medium sized Italian firms, respectively. We evaluate the revenue growth between the considered years, X_1 (1= No, 2= Yes). Then, we consider different factors that contribute to the innovation status of an enterprise: *innovation in products or services or production line or investment in R&D*, X_2 (1=No, 2=Yes); *innovation in organization system*, X_3 (1=No, 2=Yes) and *innovation in marketing strategies*, X_4 (1=No, 2=Yes). Another type of variables we consider concerns the firm's features: the *main market (in revenue terms)*, X_5 (A= Regional, B= National, C= International); the *percentage of graduate employers*, X_6 (1= 0% ⊢ 10%, 2= 10% ⊢ 50%, 3= 50% ⊢ 100%) and the *enterprise size*, X_7 (1= Small, 2= Medium). We consider three marginal distributions. First, let us define the marginal distribution $\{5,6,7\}$ in order to study the symmetric relationships among
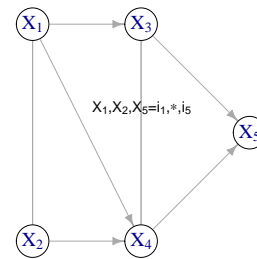
Fig. 1 SCGM with the labelled arc $X_3 - X_4$ referring to value i_1 of X_1, value i_5 of X_5 and all values of variable X_2.

the firm features; the second distribution $\{2,3,4,5,6,7\}$ to highlight possible influences of the firm features on the innovation variables; finally we consider the joint distribution $\{1,2,3,4,5,6,7\}$ in order to point out the effect of all variables on the revenue growth.

Following the three steps algorithm proposed in Section 3, in Figure 2 and Figure 3, we report the best fitting SCGM for the period 2008-2010 and 2010-2012, respectively. Note that the CSIs are represented by red arcs.

The list of independencies underlying the two SCGMs, together with the likelihood ratio test G^2, the corresponding *p-value* and the AIC value, are reported in Table 1.

Table 1 Values of the statistic tests of the selected HMMMs corresponding to the SCGMs in Figure 2 (period 2008-2010) and Figure 3 (period 2010-2012) with the list of independencies that they represent.

Period	Independencies	G^2	df	p-value	AIC
2008-2010	$X_1 \perp X_2\|X_3X_4X_5X_6X_7$ $X_4 \perp X_7\|X_2X_3X_5X_6$ $X_4 \perp X_6\|X_2X_3X_5X_7 = i_{2357}$ $X_3 \perp X_5\|X_2X_4X_6X_7 = i_{2467}$	126.02	112	0.17	-225.98
2010-2012	$X_1 \perp X_4\|X_2X_3X_5X_6X_7$ $X_3 \perp X_5\|X_2X_4X_6X_7$ $X_4 \perp X_7\|X_2X_3X_5X_6 = i_{2356}$	145.93	123	0.08	-184.06

where $i_{2357} = (2,2,3,*)$, $i_{2467} = (1,2,3,2)$ and $i_{2356} = (2,2,*,3)$.

Note that, in the two figures, despite the same structure of the undirected arcs, the presence (absence) of the directed arcs changes a little bit. In particular, in terms of innovation, the variable X_3 affects the growth X_1 in both models, while the influence of the other two innovation variables, X_2 and X_4, interchanges. Furthermore, the dependence relationships between the variable X_7 and X_4 or between X_5 and X_3 result weak or null. In fact, these independencies are present in both models, under the conditional or the CSI point of view. In the first period we may found the additional CSI between the X_6 and the X_4.

Focusing on the CSIs, we recognise in the first model that the percentage of graduate employers (X_6) does not affect the innovation in marketing strategies (X_4) when there is an innovation in products or services ($X_2 = 2$) and in the organization system ($X_3 = 2$) and when, whatever the size of the company ($X_7 = *$), the firm works mainly in an international market ($X_5 = 3$). Again, we can recognize that the type of the main market where the firm operates (X_5) does not affect the innovation in the organization system (X_3) when there is no innovation in products and services ($X_2 = 1$) but there is innovation in marketing strategies ($X_4 = 2$), the percentage of graduate employers is high ($X_6 = 3$) and the enterprise size is medium ($X_7 = 2$). On the other hand, in Figure 3 we can see that the size of the firm (X_7) does not affect the innovation in marketing strategies (X_4) when there is innovation in both products and services ($X_2 = 2$) and organization system ($X_3 = 2$), for any kind of market

$(X_5 = *)$ and when the employers are high specialized (high degree of graduated employers, $X_6 = 3$).
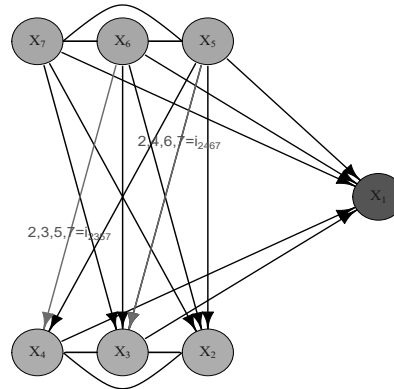
Fig. 2 Best fitting SCGM for the period from 2008 to 2010. $i_{2357} = (2,2,3,*)$ and $i_{2467} = (1,2,3,2)$.
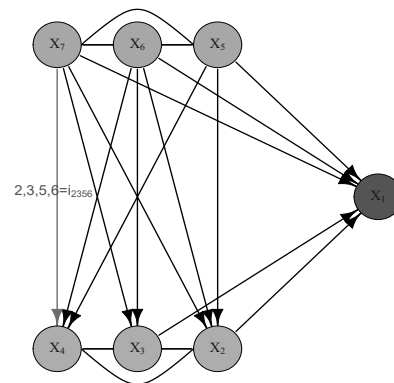
Fig. 3 Best fitting SCGM for the period from 2010 to 2012. $i_{2356} = (2,2,*,3)$.

All the analysis are carried out with the statistical software R, and the `package hmmm`, [4].

4 Conclusions

The representation of relationships among categorical variables increases by considering the CSIs. Graphical models have shown useful properties in the representation of complex structure of dependencies and, also in this case, they reveal suitable features. On the other hand, the study of CSIs allow us to study the values of the variables that really discriminate among dependence and independence structure by neglecting the unnecessary parameters. For this reason it is possible to develop strategies concerning the values of the conditioning variable where the independence does not hold. Further developments on these models may regard the multivariate regression models associated, similarly to the approach of [8].

References

1. Bartolucci, F., Colombi, R., Forcina, A.: An extended class of marginal link functions for modelling contingency tables by equality and inequality constraints. Statistica Sinica, **17**, 691–711 (2007)
2. Boutilier, C.: Context-specific independence in Bayesian networks. Proceedings of the Twelfth international conference on Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc., (1996)
3. Cazzaro, M., Colombi, R.: Marginal nested interactions for contingency tables. Communications in Statistics- Theory and Methods, **43**(13), 2799–2814 (2014)
4. Colombi, R., Giordano, S., Cazzaro, M.: hmmm: an R package for hierarchical multinomial marginal models. Journal of Statistical Software, **59**(11), 1–25 (2014)
5. Drton, M.: Discrete chain graph models. Bernoulli, **15**(3), 736–753 (2009)
6. ISTAT: Italian innovation survey 2002-2012. (2015) https://www.istat.it/it/archivio/87533
7. Lauritzen, S., Wermuth, N.: Graphical Models for Associations between Variables, some of which are Qualitative and some Quantitative. The Annals of Statistics, **17**(1), 31-57 (1989).
8. Marchetti, G., Lupparelli, M.: Chain graph models of multivariate regression type for categorical data. Bernoulli, **17**(3), 827–844 (2011)
9. Nicolussi, F.: Marginal parametrization for conditional independence models and graphical models for categorical data. PhD thesis (2013)
10. Nicolussi, F., Cazzaro, M.: Context-specific independencies for ordinal variables in chain regression models. arXiv:1712.05229 (2017)
11. Nyman, H., Pensar, J., Koski,T., Corander, J.: Context-specific independence in graphical log-linear models. Computational Statistics, **31**(4), 1493–1512 (2016)
12. Nyman, H., Pensar, J., Corander, J.: Context-specific and local independence in Markovian dependence structures. In Dependence Logic, (pp. 219-234). Springer International Publishing, (2016)
13. Roverato, A., La Rocca, L.: Log-mean linear models for binary data. Biometrika, **100**(1), 485–494 (2013)
14. Rudas, T., Bergsma, W. P., Németh, R.: Marginal log-linear parameterization of conditional independence models. Biometrika, **97**(4), 1006–1012 (2010)